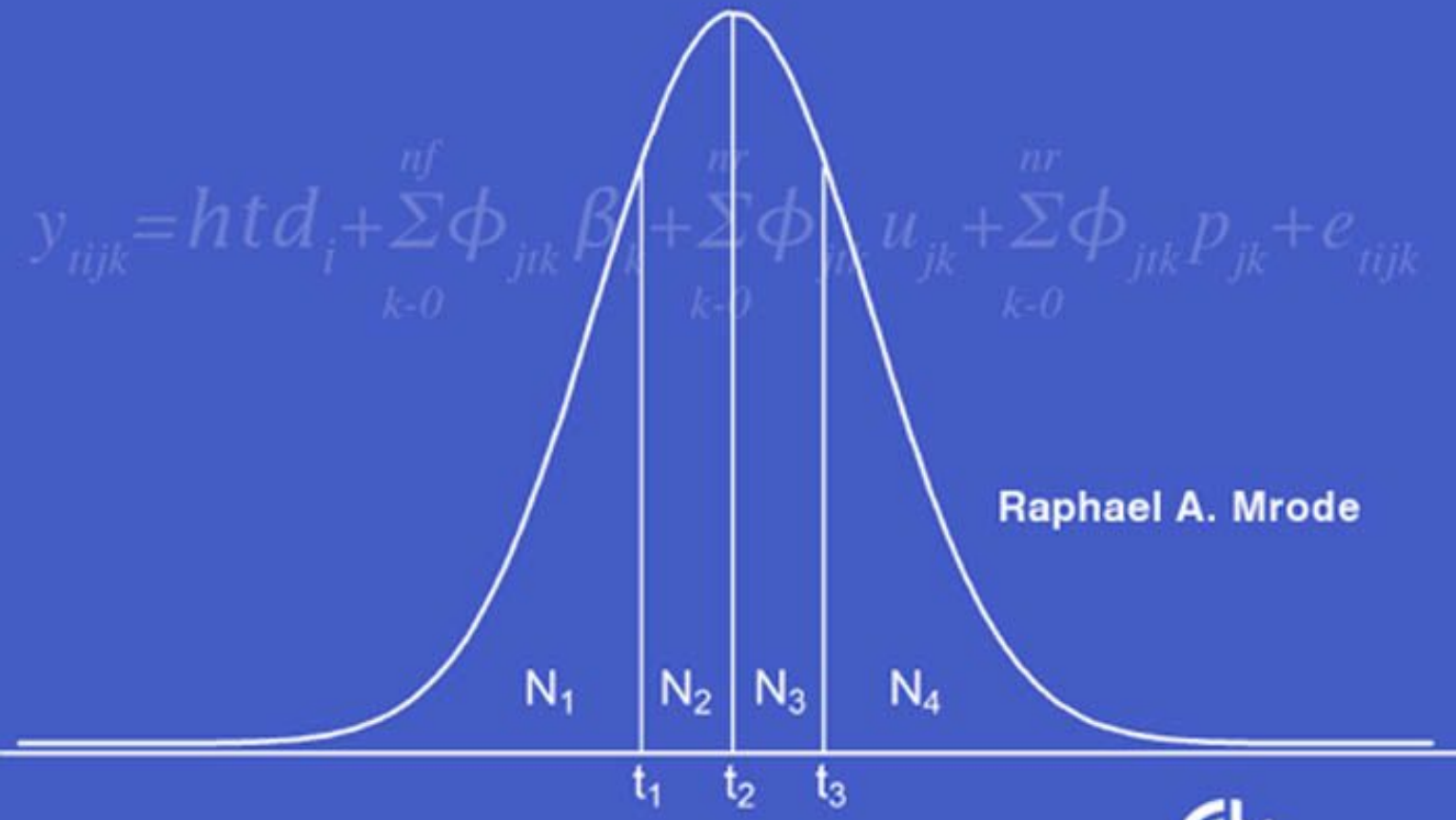
Linear Models for the Prediction of Animal Breeding Values

3rd Edition

$$y_{ijk} = htd_i + \sum_{k=0}^{nf} \phi_{jtk} \beta_k + \sum_{k=0}^{nr} \phi_{jtk} u_{jk} + \sum_{k=0}^{nr} \phi_{jtk} p_{jk} + e_{ijk}$$

Raphael A. Mrode



www.cabi.org

Linear Models for the Prediction of Animal Breeding Values

3rd Edition

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Linear Models for the Prediction of Animal Breeding Values

3rd Edition

Raphael A. Mrode
Scotland's Rural College, Edinburgh, UK
(Former Scottish Agricultural College)

With Chapter 15 contributed by
Robin Thompson
Rothamsted Research, Harpenden, UK

www.cabi.org

Downloaded From Vet-eBooks.com

CABI is a trading name of CAB International

CABI
Nosworthy Way
Wallingford
Oxfordshire OX10 8DE
UK

Tel: +44 (0)1491 832111
Fax: +44 (0)1491 833508
E-mail: info@cabi.org
Website: www.cabi.org

CABI
38 Chauncey Street
Suite 1002
Boston, MA 02111
USA

Tel: +1 800 552 3083 (toll free)
Tel: +1 (0)617 395 4051
E-mail: cabi-nao@cabi.org

© R.A. Mrode, 2014. All rights reserved. No part of this publication may be reproduced in any form or by any means, electronically, mechanically, by photocopying, recording or otherwise, without the prior permission of the copyright owners.

A catalogue record for this book is available from the British Library, London, UK.

Library of Congress Cataloging-in-Publication Data

Mrode, R. A.
Linear models for the prediction of animal breeding values / Raphael A. Mrode. -- 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84593-981-6 (pbk. : alk. paper) -- ISBN 978-1-78064-391-5 (hardback : alk. paper) 1. Livestock--Breeding--Mathematical models. 2. Livestock--Breeding--Statistical methods. 3. Livestock--Genetics--Mathematical models. 4. Livestock--Genetics--Statistical methods. I. Title.

SF105.M77 2013
636--dc23

2013025685

ISBN-13: 978 1 84593 981 6 (Pbk)
978 1 78064 391 5 (Hbk)

Commissioning editor: Sarah Hulbert
Editorial assistant: Alexandra Lainsbury
Production editor: Lauren Povey

Typeset by SPi, Pondicherry, India.
Printed and bound by Gutenberg Press Ltd, Tarxien, Malta.

Downloaded From Vet-eBooks.com

Contents

Preface xiii

Abbreviations xv

1. Genetic Evaluation with Different Sources of Records 1
- 1.1. Introduction 1
- 1.2. The Basic Model 1
- 1.3. Breeding Value Prediction from the Animal's Own Performance 2
  - 1.3.1. Single record 2
  - 1.3.2. Repeated records 3
- 1.4. Breeding Value Prediction from Progeny Records 6
- 1.5. Breeding Value Prediction from Pedigree 9
- 1.6. Breeding Value Prediction for One Trait from Another 10
- 1.7. Selection Index 11
  - 1.7.1. Accuracy of index 12
  - 1.7.2. Examples of selection indices using different sources of information 14
  - 1.7.3. Prediction of aggregate genotype 16
  - 1.7.4. Overall economic indices using predicted genetic merit 18
  - 1.7.5. Restricted selection index 19
  - 1.7.6. Index combining breeding values from phenotype and genetic marker information 21

2. Genetic Covariance Between Relatives 22
- 2.1. Introduction 22
- 2.2. The Numerator Relationship Matrix 22
- 2.3. Decomposing the Relationship Matrix 23
- 2.4. Computing the Inverse of the Relationship Matrix 25
  - 2.4.1. Inverse of the numerator relationship matrix ignoring inbreeding 26
  - 2.4.2. Inverse of the numerator relationship matrix accounting for inbreeding 28
- 2.5. Inverse of the Relationship Matrix for Sires and Maternal Grandsires 30
- 2.6. An Example of the Inverse of a Sire and Maternal Grandsire Relationship Matrix 32

3. Best Linear Unbiased Prediction of Breeding Value: Univariate Models with One Random Effect 34
- 3.1. Introduction 34
- 3.2. Brief Theoretical Background 35
- 3.3. A Model for an Animal Evaluation (Animal Model) 37

Downloaded From Vet-eBooks.com

3.3.1. Constructing the mixed model equations 38
3.3.2. Progeny (daughter) yield deviation 42
3.3.3. Accuracy of evaluations 44
3.4. A Sire Model 46
3.4.1. An illustration 46
3.5. Reduced Animal Model 49
3.5.1. Defining the model 49
3.5.2. An illustration 51
3.5.3. An alternative approach 54
3.6. Animal Model with Groups 54
3.6.1. An illustration 56

4. Best Linear Unbiased Prediction of Breeding Value: Models with Random Environmental Effects 61
4.1. Introduction 61
4.2. Repeatability Model 61
4.2.1. Defining the model 62
4.2.2. An illustration 62
4.2.3. Calculating daughter yield deviations 66
4.3. Model with Common Environmental Effects 66
4.3.1. Defining the model 67
4.3.2. An illustration 67

5. Best Linear Unbiased Prediction of Breeding Value: Multivariate Animal Models 70
5.1. Introduction 70
5.2. Equal Design Matrices and No Missing Records 71
5.2.1. Defining the model 71
5.2.2. An illustration 72
5.2.3. Partitioning animal evaluations from multivariate analysis 74
5.2.4. Accuracy of multivariate evaluations 76
5.2.5. Calculating daughter yield deviations in multivariate models 77
5.3. Equal Design Matrices with Missing Records 78
5.3.1. An illustration 78
5.4. Unequal Design Matrices 80
5.4.1. Numerical example 80
5.4.2. Illustrating the computation of DYD from a multivariate model 82
5.5. Multivariate Models with No Environmental Covariance 84
5.5.1. Different traits recorded on relatives 84
5.5.2. The multi-trait across-country evaluations (MACE) 86

6. Methods to Reduce the Dimension of Multivariate Models 95
6.1. Introduction 95
6.2. Canonical Transformation 95
6.2.1. The model 96
6.2.2. An illustration 97

Downloaded From Vet-eBooks.com

6.3. Cholesky Transformation 98
6.3.1. Calculating the transformation matrix and defining the model 98
6.3.2. An illustration 99
6.4. Factor and Principal Component Analysis 101
6.4.1. Factor analysis 102
6.4.2. Principal component analysis 105
6.4.3. Analysis with reduced rank PC model 106

7. Maternal Trait Models: Animal and Reduced Animal Models 109
7.1. Introduction 109
7.2. Animal Model for a Maternal Trait 110
7.2.1. An illustration 110
7.3. Reduced Animal Model with Maternal Effects 115
7.3.1. An illustration 116
7.4. Sire and Maternal Grandsire Model 119

8. Social Interaction Models 121
8.1. Introduction 121
8.2. Animal Model with Social Interaction Effects 123
8.2.1. Illustration of a model with social interaction 125
8.3. Partitioning Evaluations from Associative Models 127
8.4. Analysis Using Correlated Error Structure 128

9. Analysis of Longitudinal Data 130
9.1. Introduction 130
9.2. Fixed Regression Model 131
9.2.1. An illustration 132
9.3. Random Regression Model 136
9.3.1. Numerical application 138
9.3.2. Partitioning animal solutions from random regression model 142
9.3.3. Calculating daughter yield deviations 145
9.3.4. Reliability of breeding values 146
9.3.5. Random regression models with spline function 147
9.3.6. Random regression model for maternal traits 148
9.4. Covariance Functions 149
9.4.1. Fitting a reduced order covariance function 151
9.5. Equivalence of the Random Regression Model to the Covariance Function 155

10. Use of Genetic Markers in Breeding Value Prediction 156
10.1. Introduction 156
10.2. Defining a Model with Marker Information 156
10.3. Calculating the Covariance Matrix (G_v) for MQTL Effects 157
10.3.1. Numerical application 158
10.4. An Alternative Approach for Calculating G_v 160
10.5. Calculating the Inverse of G_v 161
10.6. Prediction of Breeding Values with Marker Information 165
10.6.1. An illustration 165

Downloaded From Vet-eBooks.com

10.7. Directly Predicting the Additive Genetic Merit at the MQTL 167
10.7.1. An illustration 168
10.8. Predicting Total Additive Genetic Merit 169
10.8.1. Numerical application 169
10.9. Analysis of Data with QTL Bracketed by Two Markers 171
10.9.1. Basic model 171
10.9.2. Calculating the covariance matrix, G 172
10.9.3. An illustration 174

11. Computation of Genomic Breeding Values and Genomic Selection 177
11.1. Introduction 177
11.2. General Linear Model 178
11.3. Coding and Scaling Genotypes 178
11.4. Fixed Effect Model for SNP Effects 179
11.5. Mixed Linear Model for Computing SNP Effects 182
11.5.1. SNP-BLUP model 183
11.5.2. Equivalent models: GBLUP 184
11.5.3. Equivalent models: selection index approach 187
11.6. Mixed Linear Models with Polygenic Effects 188
11.7. Single-step Approach 190
11.8. Bayesian Methods for Computing SNP Effects 193
11.8.1. BayesA 194
11.8.2. BayesB 197
11.8.3. BayesC 199
11.8.4. BayesCπ 201
11.9. Cross-validation and Genomic Reliabilities 202
11.10. Understanding SNP Solutions from the Various Models 202

12. Non-additive Animal Models 204
12.1. Introduction 204
12.2. Dominance Relationship Matrix 204
12.3. Animal Model with Dominance Effect 205
12.3.1. Solving for animal and dominance genetic effects separately 206
12.3.2. Solving for total genetic merit directly 208
12.4. Method for Rapid Inversion of the Dominance Matrix 209
12.4.1. Inverse of the relationship matrix of subclass effects 210
12.4.2. Prediction of dominance effects 211
12.4.3. Calculating the inverse of the relationship matrix among dominance and subclass effects for example data 212
12.5. Epistasis 215
12.5.1. Rules for the inverse of the relationship matrix for epistatic and subclass effects 216
12.5.2. Calculating the inverse relationship matrix for epistasis and the subclass matrix for an example pedigree 217

Downloaded From Vet-eBooks.com

13. Analysis of Ordered Categorical Traits 219
13.1. Introduction 219
13.2. The Threshold Model 220
13.2.1. Defining some functions of the normal distribution 220
13.2.2. Data organization and the threshold model 221
13.2.3. Numerical example 223
13.3. Joint Analysis of Quantitative and Binary Traits 230
13.3.1. Data and model definition 230
13.3.2. Numerical application 234

14. Survival Analysis 240
14.1. Introduction 240
14.2. Functional Survival 240
14.3. Censoring 240
14.4. Models for Analysis of Survival 241
14.4.1. Linear models 241
14.4.2. Random regression models for survival 241
14.4.3. Proportional hazard models 243
14.4.4. Non-parametric estimation of the survival function 245
14.4.5. Regression survival models 246
14.4.6. Mixed survival models 247
14.4.7. Group data survival model 250

15. Estimation of Genetic Parameters 251
Robin Thompson

15.1. Introduction 251
15.2. Univariate Sire Model 251
15.3. Numerical Example of Sire Model 252
15.4. Extended Model 253
15.5. Numerical Example 254
15.6. Animal Model 255
15.7. Numerical Example 257

16. Use of Gibbs Sampling in Variance Component Estimation and Breeding Value Prediction 260
16.1. Introduction 260
16.2. Univariate Animal Model 261
16.2.1. Prior distributions 261
16.2.2. Joint and full conditional distributions 262
16.2.3. Inferences from the Gibbs sampling output 264
16.2.4. Numerical application 265
16.3. Multivariate Animal Model 266
16.3.1. Prior distributions 267
16.3.2. Conditional probabilities 267
16.3.3. Numerical illustration 269

17. Solving Linear Equations 271
17.1. Introduction 271
17.2. Direct Inversion 271

Downloaded From Vet-eBooks.com

17.3. Iteration on the Mixed Model Equations 271
17.3.1. Jacobi iteration 272
17.3.2. Gauss–Seidel iteration 275
17.4. Iterating on the Data 276
17.4.1. Animal model without groups 278
17.4.2. Animal model with groups 282
17.4.3. Reduced animal model with maternal effects 284
17.5. Preconditioned Conjugate Gradient Algorithm 292
17.5.1. Computation strategy 293
17.5.2. Numerical application 294

Appendix A: Introduction to Matrix Algebra 299
A.1 Matrix: A Definition 299
A.2 Special Matrices 300
A.2.1 Square matrix 300
A.2.2 Diagonal matrix 300
A.2.3 Triangular matrix 300
A.2.4 Symmetric matrix 301
A.3 Basic Matrix Operations 301
A.3.1 Transpose of a matrix 301
A.3.2 Matrix addition and subtraction 301
A.3.3 Matrix multiplication 302
A.3.4 Direct product of matrices 302
A.3.5 Matrix inversion 303
A.3.6 Rank of a matrix 304
A.3.7 Generalized inverses 305
A.3.8 Eigenvalues and eigenvectors 305

Appendix B: Fast Algorithms for Calculating Inbreeding Based on the L Matrix 306
B.1 Meuwissen and Luo Algorithm 306
B.1.1 Illustration of the algorithm 307
B.2 Modified Meuwissen and Luo Algorithm 308
B.2.1 Illustration of the algorithm 309

Appendix C 311
C.1 Outline of the Derivation of the Best Linear Unbiased Prediction (BLUP) 311
C.2 Proof that $\hat{b}$ and $\hat{a}$ from MME are the GLS of b and BLUP of a, Respectively 312
C.3 Deriving the Equation for Progeny Contribution (PC) 313

Appendix D: Methods for Obtaining Approximate Reliability for Genetic Evaluations 314
D.1 Computing Approximate Reliabilities for an Animal Model 314
D.2 Computing Approximate Reliabilities for Random Regression Models 316
D.2.1 Determine value of observation for an animal 316

Downloaded From Vet-eBooks.com

D.2.2 Value of records on descendants 316
D.2.3 Value of records on ancestors 317

Appendix E 318
E.1 Canonical Transformation: Procedure to Calculate the Transformation Matrix and its Inverse 318
E.2 Canonical Transformation with Missing Records and Same Incidence Matrices 319
E.2.1 Illustration 320
E.3 Cholesky Decomposition 322

Appendix F: Procedure for Computing Deregressed Breeding Values 323

Appendix G: Calculating Φ, a Matrix of Legendre Polynomials Evaluated at Different Ages or Time Periods 325

References 327

Index 337

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Preface

Best linear unbiased prediction (BLUP) has become the most widely accepted method for genetic evaluation of domestic livestock. Since the method was first published by Henderson (1949) it has evolved in terms of its application in models for genetic evaluation, from sire, sire and maternal grandsire models in the early years, followed by univariate and multivariate animal models, random regression models for the analysis of longitudinal data and more recently, for the analysis of the genomic data (SNP-BLUP or GBLUP). Advances in computational methods and computing power have enhanced this development. Currently, most national genetic evaluation systems for several domestic livestock species are based on animal or random regression models using BLUP.

In view of these developments and the proliferation of information in the literature, there is no simple and straightforward text on the application of linear models to the prediction of breeding values. Moreover, in developing countries, where access to journals is limited, there is a basic lack of practical information on the subject. This book has been written with a good balance of theory and application to fill this gap. It places at the hand of the reader the application of BLUP in modelling several genetic situations in a single text. The book has been compiled from various publications and experience gained from several colleagues in the subject area and from involvement in several national evaluation schemes over the last 14 years. Relevant references are included to indicate sources of some of the materials.

Initially, in Chapter 1, the basic model and assumptions governing genetic evaluation are presented, together with simple situations involving prediction of breeding values from the records of an individual. This is followed by the introduction and use of selection indices to predict genetic merit combining information on several traits and individuals. Then the general framework on the application of BLUP in genetic evaluation in a univariate and multivariate situations is presented in Chapters 3 to 5. The simplification of multivariate evaluations by means of several transformations is also examined, followed by maternal trait and social interaction models. Random regression models for the analysis of longitudinal data are discussed in Chapter 9, followed by a chapter on incorporating genetic marker information into genetic evaluations in the context of marker-assisted selection and then genomic selection. Non-additive genetic animal models are discussed with methods for rapidly computing the inverse of the relationship matrices for dominance and epistasis effects. Next, threshold and survival models are discussed. In Chapters 15 and 16, the basic concepts for variance component estimation are introduced, followed by the application of the Gibbs sampler in estimation of genetic parameter and evaluations for univariate and multivariate models. Finally, computing strategies for solving mixed model equations are examined, with a presentation of the several formulae governing iterative procedures on the data. A knowledge of basic matrix algebra is needed to understand the principles of genetic evaluation discussed in the text. For the benefit of those not

Downloaded From Vet-eBooks.com

familiar with matrix algebra, a section on introductory matrix algebra has been incorporated as Appendix A. It is also assumed that the reader is familiar with the basic principles of quantitative genetics.

Several examples have been used to illustrate the various models for genetic evaluation covered in the text, and attempts have been made to present formulae that explain how the solutions for random and fixed effects in the models were obtained from the mixed model equations. This illustrates to the reader how the various pieces of information are weighted to obtain the genetic merit of an animal under various models.

Every attempt has been made to ensure the accuracy of the text, but in the event of errors being discovered, please inform the author.

Professor Robin Thompson contributed the chapter on estimation of variance components despite his busy schedule and reviewed the manuscript of earlier editions. His contribution is immensely acknowledged. The chapter on genomic selection was reviewed by Drs Ben Hayes, Ricardo Pong-Wong and Professor John A. Woolliams and I am grateful for their valuable input. Drs Gabor Mészáros and Sue Brotherstone reviewed the chapter on survival analysis within a very tight schedule and I acknowledge their contribution. I am grateful to Professor Denny Cruz and Dr Victor Olori for reviewing the chapters on social interaction and on reducing the dimension of multivariate analysis. I am greatly indebted to Professors W.G. Hill and Mr G. Swanson for reviewing the manuscript of earlier editions; their comments and suggestions resulted in substantial improvements in the text. Drs Martin Lidauer and Ismo Stranden read specific chapters or sections; I acknowledge their useful suggestions. The assistance of Dr Sebastian Mucha in preparing the graphs in the text is greatly acknowledged. In addition, experience gained from working with Dr Mike Coffey and the late Professors C. Smith and B.W. Kennedy has been valuable in writing this book. I also wish to express my thanks to Prof. R.L. Quaas for permission to use information from his unpublished note on inbreeding algorithm; Animal Genetics and Breeding Unit, University of New England, Australia, for allowing me to adopt some materials from BLUP Handbook for Chapter 2 of the text; and Prof. Fernando R.L. to use some his material from the Iowa State University 2010 summer course. My sincere gratitude to my wife, Doris, for her immense support, and for typing part of the manuscript. Special thanks to Kevwe, Joshua and Esther for their cooperation, especially when I had to take time off to prepare the manuscript, and to many dear friends who were of great encouragement. Finally, to God be all the glory.

R.A. Mrode
Scotland's Rural College
(Former Scottish Agricultural College)
Edinburgh

Downloaded From Vet-eBooks.com

Abbreviations

ADG	average daily gain
ARHS	adjusted right-hand side
BFAT	backfat thickness
BLUE	best linear unbiased estimator
BLUP	best linear unbiased prediction
BV	breeding value
BW	birth weight
CF	covariance function
CR	correlated response
DBV	direct breeding value
DGV	direct genomic breeding values
DIM	days in milk
DRB	deregressed breeding values
DRP	deregressed proofs
DSP	durable performance sum
DYD	daughter yield deviation
EBV	estimated breeding value
EDC	effective daughter contributions
EM	expectation maximization
ETA	estimated transmitting ability
FA	factor analysis
GEBV	genomic breeding values
GLS	generalized least squares
GR	growth rate
HTD	herd–test–day
HYS	herd–year–season
IBD	identical by descent
IGE	indirect genetic effects
INET	index net
LD	linkage disequilibrium
LGR	lean growth rate
LP	lean per cent
LPL	length of productive life
LS	lifespan
LSE	least squares equations
MACE	multi-trait across-country evaluation
MAS	marker-assisted selection
MBLUP	multivariate best linear unbiased prediction
MCMC	Markov chain Monte Carlo
MGD	maternal granddams

Downloaded From Vet-eBooks.com

MGS	maternal grandsire
ML	marker locus
MME	mixed model equations
MQTL	quantitative trait locus (linked to a marker locus)
MS	muscle score
PC	progeny contribution
PCG	preconditioned conjugate gradient
PEC	prediction error covariance
PEV	prediction error variance
PIN	production index
PLI	profitable life index
PPA	probable producing ability
PTA	predicted transmitting ability
PWG	post-weaning gain
PYD	progeny yield deviation
QTL	quantitative trait locus
RAM	reduced animal model
REML	restricted (or residual) maximum likelihood
RHS	right-hand side
RP	residual polygenic
RR	random regression
RRM	random regression model
RRS	risk ratios
SBV	associative breeding value
SCC	somatic cell count
SEP	standard error of prediction
SNP	single nucleotide polymorphism
SSR	simple sequence repeats
TBV	total breed value
WWG	pre-weaning gain
XFA	extended factor analysis
YSP	year–season–parity

Downloaded From Vet-eBooks.com

1 Genetic Evaluation with Different Sources of Records

1.1 Introduction

The prediction of breeding values constitutes an integral part of most breeding programmes for genetic improvement. Crucial to the accurate prediction of breeding value is the availability of records. In a population, data available at the initial stages are usually on individual animals, which may or may not be related, and later on offspring and other relatives. Thus initially, the prediction of breeding values may be based on the records of individuals and few relatives. In this chapter, the use of individual records and information from other related sources in the prediction of breeding value is addressed. Also, the principles for the calculation of selection indices combining information from different sources and relatives are discussed.

1.2 The Basic Model

Every phenotypic observation on an animal is determined by environmental and genetic factors and may be defined by the following model:

> Phenotypic observation = environmental effects + genetic effects + residual effects

or

$$y_{ij} = \mu_i + g_i + e_{ij} \tag{1.1}$$

where y_{ij} is the record j of the ith animal; μ_i refers to the identifiable non-random (fixed) environmental effects such as herd management, year of birth or sex of the ith animal; g_i is the sum of the additive (g_a), dominance (g_d) and epistatic (g_e) genetic values of the genotype of animal i; and e_{ij} is the sum of random environmental effects affecting animal i.

The additive genetic value in the g term above represents the average additive effects of genes an individual receives from both parents and is called the breeding value. Each parent contributes a sample half of its genes to its progeny. The average effect of the sample half of genes that a parent passes to its progeny is called the transmitting ability of the parent and corresponds to one-half of its additive genetic value. The breeding value of the progeny therefore is the sum of the transmitting abilities of both the parents. Since the additive genetic value is a function of the genes transmitted from parents to progeny, it is the only component that can be selected for and therefore the main component of interest. In most cases, dominance and

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

epistasis, which represent intra-locus and inter-loci interactions respectively, are assumed to be of little significance and are included in the e_{ij} term of the model as:

$$y_{ij} = \mu_i + g_{ai} + e^*_{ij} \tag{1.2}$$

with e^*_{ij} being the sum of the random environmental effects, dominance and epistatic genetic values. Equation 1.2 constitutes the linear model usually employed in most problems of breeding value prediction in animal breeding. Usually it is assumed that **y** follows a multivariate normal distribution, implying that traits are determined by infinitely many additive genes of infinitesimal effect at unlinked loci, the so-called infinitesimal model (Fisher, 1918; Bulmer, 1980). Also, it is assumed that var(**y**), var(**g**) and var($\mathbf{e}^*$) are known and that there is no correlation between **g** and $\mathbf{e}^*$ (cov(**g**, $\mathbf{e}^*$) = 0) nor is there any correlation among mates (cov($\mathbf{e}^*$, $\mathbf{e}^*$) = 0). Also μ, which is used subsequently in this chapter to represent the mean performance of animals in the same management group, for instance animals reared under the same management system, of the same age and sex, is assumed known. From Eqn 1.2, the problem of predicting breeding value reduces to that of adjusting phenotypic observations for identifiable non-random environmental effects and appropriately weighting the records of animals and their available relatives.

From the earlier explanation, if a_i is the breeding value of animal i, then:

$$a_i = g_{ai} = \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + m_i$$

where a_s and a_d are the breeding values of the sire and dam, respectively, and m_i is the deviation of the breeding value of animal i from the average breeding value for both parents, that is, Mendelian sampling. The sampling nature of inheritance implies that each parent passes only a sample one-half of their genes to their progeny. There is, therefore, genetic variation between offspring of the same parents since all offspring do not receive exactly the same genes. Mendelian sampling could be regarded as the deviation of the average effects of additive genes an individual receives from both parents from the average effects of genes from the parents common to all offspring.

The accurate prediction of breeding value constitutes an important component of any breeding programme since genetic improvement through selection depends on correctly identifying individuals with the highest true breeding value. The method used to predict breeding value depends on the type and amount of information available on candidates for selection. The next section discusses the prediction of breeding value using different sources of information. It should be noted that many applications of genetic evaluation deal with the prediction of transmitting ability, usually referred to as predicted transmitting ability (PTA) or estimated transmitting ability (ETA), which is one-half of the predicted breeding value.

1.3 Breeding Value Prediction from the Animal's Own Performance

1.3.1 Single record

When one phenotypic record is the only available information on each animal, the estimated breeding value (EBV) (a_i) for animal i can be calculated as:

$$\hat{a}_i = b(y_i - \mu) \tag{1.3}$$

Downloaded From Vet-eBooks.com

where b is the regression of true breeding value on phenotypic performance and μ, as indicated earlier, is the mean performance of animals in the same management group and is assumed to be known. Thus:

$$\begin{aligned} b &= \text{cov}(a, y)/\text{var}(y) = \text{cov}(a, a + e)/\text{var}(y) \\ &= \sigma_a^2/\sigma_y^2 \\ &= h^2 \end{aligned}$$

The prediction is simply the adjusted record multiplied by the heritability (h^2). The correlation between the selection criterion, in this case the phenotypic value, and the true breeding value is known as the accuracy of prediction. It provides a means of evaluating different selection criteria because the higher the correlation, the better the criterion as a predictor of breeding value. In some cases, the accuracy of evaluations is reported in terms of reliability or repeatability (r^2), which is the squared correlation between the selection criterion and the true breeding value. With a single record per animal, the accuracy is:

$$\begin{aligned} r_{a,y} &= \text{cov}(a, y)/(\sigma_a \sigma_y) \\ &= \sigma_a^2/(\sigma_a \sigma_y) \\ &= h \end{aligned}$$

and reliability equals h^2.

Expected response (R) to selection on the basis of a single record per individual (Falconer and Mackay, 1996) is:

$$R = ir_{a,y}^2 \sigma_y = ih^2 \sigma_y$$

where i, the intensity of selection, refers to the superiority of selected individuals above population average expressed in phenotypic standard deviation.

The variance of EBV ($\text{var}(\hat{a}_i)$) is:

$$\begin{aligned} \text{var}(\hat{a}_i) &= \text{var}(by) = \text{var}(h^2 y) \\ &= h^4 \sigma_y^2 \\ &= r_{a,y}^2 h^2 \sigma_y^2 = r_{a,y}^2 \sigma_a^2 \end{aligned} \tag{1.4}$$

Example 1.1

Given that the yearling weight of a heifer is 320 kg in a herd with a mean of 250 kg, predict her breeding value and its accuracy if the heritability of yearling weight is 0.45.

From Eqn 1.3:

$$\hat{a} = 0.45(320 - 250) = 31.50 \text{ kg}$$

and:

$$r_{a,y} = \sqrt{0.45} = 0.67$$

1.3.2 Repeated records

When multiple measurements on the same trait, such as milk yield in successive lactations, are recorded on an animal, its breeding value may be predicted from the

Downloaded From Vet-eBooks.com

mean of these records. With repeated measurements it is assumed that there is additional resemblance between records of an individual due to environmental factors or circumstances that affect the records of the individual permanently. In other words, there is an additional covariance between records of an individual due to non-genetic permanent environmental effects. Thus the between-individual variance is partly genetic and partly environmental (permanent environmental effect). The within-individual variance is attributed to differences between successive measurements of the individual arising from temporary environmental variations from one parity to the other. The variance of observations (var(y)) could therefore be partitioned as:

$$\text{var}(y) = \text{var}(g) + \text{var}(pe) + \text{var}(te)$$

where var(g) = genetic variance including additive and non-additive, var(pe) = variance due to permanent environmental effect, and var(te) = variance due to random temporary environmental effect.

The intra-class correlation (t), which is the ratio of the between-individual variance to the phenotypic:

$$t = (\text{var}(g) + \text{var}(pe))/\text{var}(y) \tag{1.5}$$

is usually called the repeatability and measures the correlation between the records of an individual. From Eqn (1.5):

$$\text{var}(te)/\text{var}(y) = 1 - t \tag{1.6}$$

With this model, it is always usually assumed that the repeated records on the individual measure the same trait, that is, there is a genetic correlation of 1 between all pairs of records. Also, it is assumed that all records have equal variance and that the environmental correlations between all pairs of records are equal. Let $\tilde{y}$ represent the mean of n records on animal i. The breeding value may be predicted as:

$$\hat{a}_i = b(\tilde{y}_i - \mu) \tag{1.7}$$

where:

$$b = \text{cov}(a, \tilde{y})/\text{var}(\tilde{y})$$

Now:

$$\text{cov}(a, \tilde{y}) = \text{cov}(a, g + pe + \Sigma te/n) = \sigma_a^2$$

and:

$$\text{var}(\tilde{y}) = \text{var}(g) + \text{var}(pe) + \text{var}(te)/n$$

Expressing the items in terms of Eqns 1.5 and 1.6:

$$\text{var}(t) = [t + (1 - t)/n]\sigma_y^2$$

Therefore:

$$\begin{aligned} b &= \sigma_a^2/[t + (1 - t)/n]\sigma_y^2 \\ &= nh^2/[1 + (n - 1)t] \end{aligned}$$

Note that b now depends on heritability, repeatability and the number of records.

Downloaded From Vet-eBooks.com

As mentioned earlier, the difference between repeated records of an individual is assumed to be due to temporary environmental differences between successive performances. However, if successive records are known to be affected by factors that influence performance, these must be corrected for. For instance, differences in age at calving in first and second lactations may influence milk yield in first and second lactations. Such age differences should be adjusted for before using the means of both lactations for breeding-value prediction.

The accuracy of the EBV is:

$$\begin{aligned} r_{a,y} &= \mathrm{cov}(a,\tilde{y}) / (\sigma_a \sigma_y) \\ &= \sigma_a^2 / (\sigma_a \sqrt{[t + (1-t)/n]}\sigma_y^2) \\ &= h\sqrt{[n/(1+(n-1)t)]} \\ &= \sqrt{[nh^2/(1+(n-1)t)]} = \sqrt{b} \end{aligned}$$

Compared with single records, there is a gain in the accuracy of prediction with repeated records from the above equation, which is dependent on the value of repeatability and the number of records. This gain in accuracy results mainly from the reduction in temporary environmental variance (within-individual variance) as the number of records increases. When t is low, this gain is substantial as the number of records increases. When t is high, there is little gain in accuracy with repeated records compared with using only single records. The gain in accuracy from repeated records compared with selection on single records can be obtained as the ratio of accuracy from repeated records (r_n) to that from single records (r_k):

$$\frac{r_n}{r_k} = \frac{\sqrt{\dfrac{h^2}{t + \dfrac{(1-t)}{n}}}}{h} = \sqrt{\frac{1}{t + \dfrac{(1-t)}{n}}}$$

Using the above equation, the gain in accuracy from repeated records compared with selection on single records is given in Table 1.1. The increase in accuracy with four measurements at a low t value of 0.4 was 35%, but this dropped to only 8% when t equalled 0.8. In general, the rate of increase dropped rapidly as the number of records exceeded four, and it is seldom necessary to record more than four measurements.

Table 1.1. Percentage increase in accuracy of prediction with repeated records compared with single records at a heritability of 0.35.

	Number of records				
t-values	2	4	6	8	10
0.4	20	35	41	45	47
0.6	12	20	22	24	25
0.8	5	8	10	10	10

Downloaded From Vet-eBooks.com

Expected response to selection on the basis of mean of repeated records is:

$$R = ih^2_{a,y}\sigma_y\sqrt{[t + (1 - t)/n]}$$

Example 1.2

Assume that a cow has a mean yield of 8000 kg of milk for first and second lactations. If the phenotypic standard deviation and heritability of milk yield in the first two lactations are 600 kg and 0.30, respectively, and the correlation between first and second lactation yields is 0.5, predict the breeding value of the cow for milk yield in the first two lactations and its accuracy. Assume that the herd mean for first and second lactations is 6000 kg.

From Eqn 1.7:

$$\hat{a}_{cow} = b(8000 - 6000)$$

with:

$$b = 2(0.3)/(1 + (2 - 1)0.5)) = 0.4$$

Therefore:

$$\hat{a}_{cow} = 0.4(8000 - 6000) = 800 \text{ kg}$$

and:

$$r_{a,\hat{a}} = \sqrt{0.4} = 0.632$$

1.4 Breeding Value Prediction from Progeny Records

For traits where records can be obtained only on females, the prediction of breeding values for sires is usually based on the mean of their progeny. This is typical of the dairy cattle situation, where bulls are evaluated on the basis of their daughters. Let $\tilde{y}_i$ be the mean of single records of n progeny of sire i and assume that the progeny are only related through the sire (paternal half-sibs), and so the breeding value of sire i is:

$$\hat{a}_i = b(\tilde{y}_i - \mu) \tag{1.8}$$

where:

$$b = \text{cov}(a, \tilde{y})/\text{var}(\tilde{y})$$

Now:

$$\text{cov}(a, \tilde{y}) = \text{cov}(a, \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + \Sigma e/n)$$

where a_s is the sire breeding value and a_d represents the breeding value for the dams. Therefore:

$$\text{cov}(a, \tilde{y}) = \tfrac{1}{2}\text{cov}(a, a_s) = \tfrac{1}{2}\sigma^2_a$$

Using the same principles in as in Section 1.3.2:

$$\text{var}(\tilde{y}) = [t + (1 - t)/n]\sigma^2_y$$

Downloaded From Vet-eBooks.com

assuming there is no environmental covariance between the half-sib records and t, the intra-class correlation between half-sibs, is $\frac{1}{4}\sigma_a^2/\sigma_y^2 = \frac{1}{4}h^2$.

Therefore:

$$\begin{aligned} b &= \tfrac{1}{2}\sigma_a^2/[t + (1 - t)/n]\sigma_y^2 \\ &= \tfrac{1}{2}h^2\sigma_y^2/[\tfrac{1}{4}h^2 + (1 - \tfrac{1}{4}h^2)/n]\sigma_y^2 \\ &= 2nh^2/(nh^2 + (4 - h^2)) \\ &= 2n/(n + (4 - h^2)/h^2) \\ &= 2n/n + k \end{aligned}$$

with:

$$k = (4 - h^2)/h^2$$

The term k is constant for any assumed heritability. The weight (b) depends on the heritability and number of progeny and approaches 2 as the number of daughters increases.

The accuracy of the EBV is:

$$r_{a,y} = \mathrm{cov}(a, \tilde{y}) / \sqrt{(\mathrm{var}(a)\ \mathrm{var}(\tilde{y}))}$$

From the above calculations, this could be expressed as:

$$\begin{aligned} r_{a,y} &= \frac{\frac{1}{2}h^2\sigma_y^2}{\sqrt{h^2\sigma_y^2\left(\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}\right)\sigma_y^2}} = \frac{\frac{1}{2}h}{\sqrt{\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}}} \\ &= \sqrt{\frac{nh^2}{nh^2 + (4-h^2)}} \\ &= \sqrt{\frac{n}{n+k}} \end{aligned}$$

which approaches unity (1) as the number of daughters becomes large. Reliability of the predicted breeding value therefore equals $n/(n + k)$.

The equation for expected response when selection is based on the mean of half-sibs is the same as that given in Section 1.3.2 for the mean of repeated records but with t now referring to the intra-class correlation between half-sibs.

The performance of any future daughters of the sire can be predicted from the mean performance of the present daughters. The breeding value of a future daughter ($\hat{a}_{daugh.}$) of the sire can be predicted as:

$$\hat{a}_{daugh.} = b(\tilde{y} - \mu)$$

with $\tilde{y}$ and μ as defined in Eqn 1.8, respectively, and:

$$b = \mathrm{cov}(a_{daugh.}, \tilde{y})/\mathrm{var}(\tilde{y})$$

Now:

$$\mathrm{cov}(a_{daugh.}, \tilde{y}) = \mathrm{cov}(\tfrac{1}{2}a_s + \tfrac{1}{2}a_{d*}, \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + \Sigma e/n)$$

where the subscript $d*$ refers to the dam of the future daughter, which is assumed to be unrelated to dams (d) of present daughters. Therefore:

$$\mathrm{cov}(a_{daugh.}, \tilde{y}) = \mathrm{cov}(\tfrac{1}{2}a_s, \tfrac{1}{2}a_s) = \tfrac{1}{4}\mathrm{cov}(a_s, a_s) = \tfrac{1}{4}\sigma_a^2$$

Downloaded From Vet-Books.com

Therefore:

$$b = \tfrac{1}{4}\sigma_a^2/[t + (1 - t)/n]\sigma_y^2$$

Using the same calculations for obtaining b in Eqn 1.8:

$$b = n/(n + k)$$

The b value is half of the value of b in Eqn 1.8, thus the predicted breeding value of a future daughter of the sire is equal to half the EBV of the sire. The performance of a future daughter of the sire can be predicted as:

$$y = M + \hat{a}_{daugh.}$$

where M is the management mean.

The accuracy of the predicted breeding value of the future daughter is:

$$r_{a,y} = \text{cov}(a_{daugh.}, \tilde{y}) / \sqrt{(\text{var}(a)\ \text{var}(\tilde{y}))}$$

This could be expressed as:

$$r_{a_{daugh.},\,y} = \frac{\tfrac{1}{4}h^2\sigma_y^2}{\sqrt{h^2\sigma_y^2\left(\tfrac{1}{4}h^2 + \dfrac{(1-\tfrac{1}{4}h^2)}{n}\right)\sigma_y^2}} = \frac{\tfrac{1}{4}h}{\sqrt{\tfrac{1}{4}h^2 + \dfrac{(1-\tfrac{1}{4}h^2)}{n}}}$$

$$= \tfrac{1}{2}\sqrt{\frac{n}{n+k}}$$

which is equal to half of the accuracy of the predicted breeding value of the sire. Reliability of the predicted breeding value equals $\frac{1}{4}n/(n + k)$, which is one-quarter of the reliability of the bull proof.

Example 1.3

Suppose the fat yield of 25 half-sib progeny of a bull averaged 250 kg in the first lactation. Assuming a heritability of 0.30 and herd mean of 200 kg, predict the breeding value of the bull for fat yield and its accuracy. Also predict the performance of a future daughter of this bull for fat yield in this herd.

From Eqn 1.6:

$$\hat{a}_{bull} = b(250 - 200)$$

with:

$$b = 2n/(n + (4 - h^2)/h^2) = 2(25)/(25 + (4 - 0.3)/0.3) = 1.34$$

$$\hat{a}_{bull} = 1.34(250 - 200) = 67 \text{ kg}$$

$$r_{a,\bar{y}} = \sqrt{\left(n/\left(n+k\right)\right)} = \sqrt{\left[25/\left(25+\left(4-0.3\right)/0.3\right)\right]} = 0.82$$

The future performance of the daughter of the bull is:

$$y = (0.5)a_{bull} + \text{herd mean}$$
$$= 33.5 + 200 = 233.5 \text{ kg}$$

Downloaded From Vet-eBooks.com

1.5 Breeding Value Prediction from Pedigree

When an animal has no record, its breeding value can be predicted from the evaluations of its sire (s) and dam (d). Each parent contributes half of its genes to their progeny, and so the predicted breeding value of progeny (o) is:

$$\hat{a}_o = (\hat{a}_s + \hat{a}_d)/2 \tag{1.9}$$

Let $f = (\hat{a}_s + \hat{a}_d)/2$, then the accuracy of the predicted breeding value is:

$$r_{\hat{a}_o,f} = \frac{\text{cov}(a_o, \frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)}{\sqrt{\sigma_a^2 \,\text{var}(\frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)}}$$

Now:

$$\text{cov}(a_o, \tfrac{1}{2}\hat{a}_s + \tfrac{1}{2}\hat{a}_d) = \text{cov}(a_o, \tfrac{1}{2}\hat{a}_s) + \text{cov}(a_o, \tfrac{1}{2}\hat{a}_d)$$
$$= \text{cov}(\tfrac{1}{2}a_s + \tfrac{1}{2}a_d, \tfrac{1}{2}\hat{a}_s) + \text{cov}(\tfrac{1}{2}a_s + \tfrac{1}{2}a_d, \tfrac{1}{2}\hat{a}_d)$$

Assuming sire and dam are unrelated:

$$\text{cov}(a_o, \tfrac{1}{2}\hat{a}_s + \tfrac{1}{2}\hat{a}_d) = \tfrac{1}{4}\text{cov}(a_s, \hat{a}_s) + \tfrac{1}{4}\text{cov}(a_d, \hat{a}_d)$$
$$= \tfrac{1}{4}\text{var}(\hat{a}_s) + \tfrac{1}{4}\text{var}(\hat{a}_d)$$

Substituting the solution for the variance of EBV in Eqn 1.4:

$$\text{cov}(a_o, \tfrac{1}{2}\hat{a}_s + \tfrac{1}{2}\hat{a}_d) = \tfrac{1}{4}(r_s^2 + r_d^2)\sigma_a^2$$

From the calculation above, the term $\text{var}(\frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)$ in the denominator of Eqn 1.9 is also equal to $\frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2$, assuming random mating and the absence of joint information in the sire and dam proofs. Therefore:

$$r_{\hat{a}_o,f} = \frac{\frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2}{\sqrt{\sigma_a^2 \,\frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2}} = \frac{\sigma_f}{\sigma_a} = \tfrac{1}{2}\sqrt{(r_s^2 + r_d^2)}$$

where:

$$\sigma_f = \sqrt{\left[\text{var}\left(\tfrac{1}{2}a_s + \tfrac{1}{2}a_d\right)\right]}$$

From the above equation, the upper limit for r when prediction is from pedigree is $\frac{1}{2}\sqrt{2} = 0.7$; that is, the accuracy of the proof of each parent is unity. Note that when the prediction is only from the sire proof, for instance, then:

$$r_{\hat{a}_o, \frac{1}{2}\hat{a}_s} = \tfrac{1}{2}\sqrt{r_s^2} = \tfrac{1}{2}\sqrt{n/n + k}$$

the accuracy of the predicted breeding value of a future daughter of the sire as shown in Section 1.4.

Expected response to selection on the basis of average proof of parents is:

$$R = ir_{\hat{a}o,f}\sigma_a$$

Substituting σ_f/σ_a for r:

$$R = i\sigma_f$$

Downloaded From Vet-eBooks.com

Example 1.4

Suppose that the EBVs for the sire and dam of a heifer are 180 and 70 kg for yearling body weight, respectively. Given that the accuracy of the proofs are 0.97 for the sire and 0.77 for the dam, predict the breeding value of the heifer and its accuracy for body weight at 12 months of age.

From Eqn 1.9:

$$\hat{a}_{heifer} = 0.5(180 + 70) = 125 \text{ kg}$$

The accuracy is:

$$r_{\hat{a},a} = 0.5\sqrt{\left(0.97^2 + 0.77^2\right)} = 0.62$$

1.6 Breeding Value Prediction for One Trait from Another

The breeding value of one trait may be predicted from the observation on another trait if the traits are genetically correlated. If y is the observation on animal i from one trait, its breeding value for another trait x is:

$$\hat{a}_{ix} = b(y - \mu) \tag{1.10}$$

with:

$$b = \text{cov}(a_x, \text{measurement on } y)/\text{var}(\text{measurement on } y) \tag{1.11}$$

The genetic correlation between traits x and y (r_{axy}) is:

$$r_{axy} = \text{cov}(a_x, a_y)/(\sigma_{ax}\sigma_{ay})$$

Therefore:

$$\text{cov}(a_x, a_y) = r_{axy}\sigma_{ax}\sigma_{ay} \tag{1.12}$$

Substituting Eqn 1.12 into Eqn 1.11:

$$b = r_{axy}\sigma_{ay}\sigma_{ax}/\sigma_y^2 \tag{1.13}$$

If the additive genetic standard deviations for x and y in Eqn 1.13 are expressed as the product of the square root of their individual heritabilities and phenotypic variances, then:

$$\begin{aligned} b &= r_{axy}\sigma_y\sigma_x h_x h_y/\sigma_y^2 \\ &= r_{axy}h_x h_y\sigma_x/\sigma_y \end{aligned} \tag{1.14}$$

The weight depends on the genetic correlation between the two traits, their heritabilities and phenotypic standard deviations.

The accuracy of the predicted breeding value is:

$$\begin{aligned} r_{ax,ay} &= \text{cov}(a_x, \text{measurement on } y)/\sigma_{ax}\sigma_y \\ &= r_{axy}\sigma_{ay}\sigma_{ax}/(\sigma_{ax}\sigma_y) \\ &= r_{axy}h_y \end{aligned}$$

The accuracy depends on the genetic correlation between the two traits and heritability of the recorded trait.

Downloaded From Vet-eBooks.com

Correlated response (CR) in trait x as a result of direct selection on y (Falconer and Mckay, 1996) is:

$$CRx = ih_x h_y r_{axy} \sigma_y$$

Example 1.5

Suppose the standard deviation for growth rate (GR) (g/day) to 400 days in a population of beef cattle was 80, with a heritability of 0.43. The standard deviation for lean growth rate (LGR) (g/day) for the same population was 32, with a heritability of 0.45. If the genetic correlation between both traits is 0.95 and the population mean for growth rate is 887 g/day, predict the breeding value for LGR for an animal with a GR of 945 g/day.

Using Eqn 1.10:

$$\hat{a}_{LGR} = b(945 - 887)$$

with:

$$b = \text{cov(GR, LGR)/var(GR)}$$

From Eqn 1.13:

$$b = (0.95(0.656)(0.671)(32))/80 = 0.167$$
$$\hat{a}_{LGR} = 0.167(945 - 887) = 9.686$$

The accuracy of the prediction is:

$$r = 0.95\left(\sqrt{0.43}\right) = 0.623$$

1.7 Selection Index

The selection index is a method for estimating the breeding value of an animal combining all information available on the animal and its relatives. It is the best linear prediction of an individual breeding value. The numerical value obtained for each animal is referred to as the index (I) and it is the basis on which animals are ranked for selection. Suppose y_1, y_2 and y_3 are phenotypic values for animal i and its sire and dam, then the index for this animal using this information would be:

$$I_i = \hat{a}_i = b_1(y_1 - \mu_1) + b_2(y_2 - \mu_2) + b_3(y_3 - \mu_3) \quad (1.15)$$

where b_1, b_2, b_3 are the factors by which each measurement is weighted. The determination of the appropriate weights for the several sources of information is the main concern of the selection index procedure. In Eqn 1.15, the index is an estimate of the true breeding value of animal i.

Properties of a selection index are:

1. It minimizes the average square prediction error, that is, it minimizes the average of all $(a_i - \hat{a}_i)^2$.
2. It maximizes the correlation $(r_{a,\hat{a}})$ between the true breeding value and the index. The correlation is often called the accuracy of prediction.
3. The probability of correctly ranking pairs of animals on their breeding value is maximized.

Downloaded From Vet-eBooks.com

The **b** values in Eqn 1.15 are obtained by minimizing $(\mathbf{a} - \mathbf{I})^2$, which is equivalent to maximizing r_{aI}. This is the same procedure employed in obtaining the regression coefficients in multiple linear regression. Thus the **b** values could be regarded as partial regression coefficients of the individual's breeding value on each measurement. The minimization results in a set of simultaneous equations similar to the normal equations of multiple linear regression, which are solved to obtain the **b** values. The set of equations to be solved for the **b** values is:

$$\begin{aligned} b_1p_{11} + b_2p_{12} + \cdots + b_mp_{1m} &= g_{11} \\ b_1p_{21} + b_2p_{22} + \cdots + b_mp_{2m} &= g_{12} \\ \vdots \qquad\qquad & \vdots \\ b_1p_{m1} + b_2p_{m2} + \cdots + b_mp_{mm} &= g_{1m} \end{aligned} \tag{1.16}$$

where p_{mm} and g_{mm} are the phenotypic and genetic variances, respectively, for individual or trait m; p_{mn} and g_{mn} are the phenotypic and genetic covariances, respectively, between individuals or traits m and n.

In matrix form, Eqn 1.16 is:

$$\mathbf{Pb} = \mathbf{G}$$

and:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$$

where **P** is the variance and covariance matrix for observations, and **G** is the covariance matrix between observations and breeding value to be predicted.

Therefore the selection index equation is:

$$\mathbf{I} = \hat{\mathbf{a}} = (\mathbf{P}^{-1}\mathbf{G})(\mathbf{y} - \boldsymbol{\mu}) \tag{1.17}$$

$$= \mathbf{b}(\mathbf{y} - \boldsymbol{\mu}) \tag{1.18}$$

where $\boldsymbol{\mu}$ refers to estimates of environmental influences on observations, assumed to be known without error. The application of the selection index to some data therefore involves setting up Eqn 1.17. From Eqn 1.18 it is obvious that the previous methods for predicting breeding values discussed in Sections 1.3 to 1.6 are no different from a selection index and they could be expressed as in Eqn 1.17.

1.7.1 Accuracy of index

As before, the accuracy ($r_{a,I}$) of an index is the correlation between the true breeding value and the index. The higher the correlation, the better the index as a predictor of breeding value. It provides a means of evaluating different indices based on different observations, to find out, for instance, whether a particular observation is worth including in an index or not.

From the definition above:

$$r_{a,I} = \text{cov}(a, I)/(\sigma_a\sigma_I)$$

First we need to calculate σ_I^2 and $\text{cov}(a, I)$ in the above equation. Using the formula for the variance of predicted breeding value in Section 1.3.1:

$$\sigma_I^2 = \text{var}(b_1y_1 + \text{var}(b_2y_2 + \ldots + 2b_1b_2\text{cov}(y_1, y_2) + \ldots$$

Downloaded From Vet-eBooks.com

$$= b_1^2 \text{var}(y_1) + b_2^2 \text{var}(y_2) + \ldots + 2b_1 b_2 \text{cov}(y_1, y_2) + \ldots$$
$$\sigma_I^2 = b_1^2 p_{11} + b_2^2 p_{22} + \ldots + 2b_1 b_2 p_{12} + \ldots$$

or in general:

$$\sigma_I^2 = \sum_{i=1}^{m} b_i^2 p_{ii} + \left(\sum_{i=1}^{m} \sum_{j=1}^{m} b_i b_j p_{ij}; i \neq j \right)$$

where m is the number of traits or individuals in the index.

In matrix notation:

$$\sigma_I^2 = \mathbf{b'Pb}$$

Now $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$; substituting this value for **b**:

$$\sigma_I^2 = \mathbf{G'P}^{-1}\mathbf{G} \tag{1.19}$$

The covariance between the true breeding value for trait or individual i and index is:

$$\text{cov}(a_i, I) = \text{cov}(a_i, b_1 y_1) + \text{cov}(a_i, b_2 y_2) + \ldots + \text{cov}(a_i, b_j y_j)$$
$$= b_1 \text{cov}(a_i, y_1) + b_2 \text{cov}(a_i, y_2) + \ldots + b_j \text{cov}(a_i, y_j)$$

or in general:

$$\text{cov}(a_i, I) = \sum_{j=1}^{m} b_j g_{ij} \tag{1.20}$$

where g_{ij} is the genetic covariance between traits or individuals i and j, and m is the number of traits or individuals in the index.

In matrix notation:

$$\text{cov}(a_i, I) = \mathbf{b'G}$$

Substituting $\mathbf{P}^{-1}\mathbf{G}$ for **b**:

$$\text{cov}(a_i, I) = \mathbf{G'P}^{-1}\mathbf{G}$$
$$= \sigma_I^2$$

Thus, as previously, the regression of breeding value on predicted breeding values is unity. Therefore:

$$r_{a,I} = \sigma_I^2 / (\sigma_a \sigma_I) = \sigma_I / \sigma_a$$

For calculation purposes, r is better expressed as:

$$r_{a,I} = \sqrt{\frac{\sum_{j=1}^{m} b_j g_{ij}}{\sigma_a^2}} \tag{1.21}$$

Response to selection on the basis of an index is:

$$R = i r_{a,I} \sigma_a$$
$$= i \sigma_I$$

Downloaded From Vet-eBooks.com

1.7.2 Examples of selection indices using different sources of information

Data available on correlated traits

Example 1.6

Assume the following parameters were obtained for average daily gain (ADG) from birth to 400 days and lean per cent (LP) at the same age in a group of beef calves:

	Heritability	Standard deviation
ADG (g/day)	0.43	80.0
LP (%)	0.30	7.2

The genetic and phenotypic correlations (r_g and r_p) between ADG and LP are 0.30 and −0.10, respectively. Construct an index to improve growth rate in the beef calves. Assuming ADG as trait 1 and LP as trait 2, then from the given parameters:

$$p_{11} = 80^2 = 6400$$
$$p_{22} = 7.2^2 = 51.84$$
$$p_{12} = rp\sqrt{(p_{11})(p_{22})} = -0.1\sqrt{(6400)(51.84)} = -57.6$$
$$g_{11} = h^2(p_{11}) = 0.43(6400) = 2752$$
$$g_{22} = h^2(p_{22}) = 0.30(51.84) = 15.552$$
$$g_{12} = rg\sqrt{(g_{11})(g_{22})} = 62.064$$

The index equations to be solved are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{-1} \begin{bmatrix} g_{11} \\ g_{21} \end{bmatrix}$$

Inserting appropriate values gives:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 6400.00 & -57.60 \\ -57.60 & 51.84 \end{bmatrix}^{-1} \begin{bmatrix} 2752.000 \\ 62.064 \end{bmatrix}$$

The solutions are $b_1 = 0.445$ and $b_2 = 1.692$.

The index therefore is:

$$I = 0.445(\text{ADG} - \mu_{\text{ADG}}) + 1.692(\text{LP} - \mu_{\text{LP}})$$

where μ_{ADG} and μ_{LP} are herd averages for ADG and LP. Using Eqn 1.21:

$$r = \sqrt{\left[\left(0.445(2752) + 1.692(62.064)\right)/2752\right]} = 0.695$$

Using single records on individual and relatives

Example 1.7

Suppose the ADG for a bull calf (y_1) is 900 g/day and the ADG for his sire (y_2) and dam (y_3) are 800 g/day and 450 g/day, respectively. Assuming all observations were obtained

Downloaded From Vet-eBooks.com

in the same herd and using the same parameters as in Example 1.6, predict the breeding value of the bull calf for ADG and its accuracy.

From the parameters given:

$p_{11} = p_{22} = p_{33} = \sigma_y^2 = 6400$
$p_{12} = \text{cov}(y_1, y_2) = \frac{1}{2}\sigma_a^2 = \frac{1}{2}(2752) = 1376$
$p_{13} = p_{12} = 1376$
$p_{23} = 0$
$g_{11} = \sigma_a^2 = 2752$
$g_{12} = g_{13} = \frac{1}{2}\sigma_a^2 = 1376$

The index equations are:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 6400 & 1376 & 1376 \\ 1376 & 6400 & 0000 \\ 1376 & 0000 & 6400 \end{bmatrix}^{-1} = \begin{bmatrix} 2752 \\ 1376 \\ 1376 \end{bmatrix}$$

Solutions to the above equations are $b_1 = 0.372$, $b_2 = 0.135$ and $b_3 = 0.135$. The index is:

$$I = 0.372(900 - \mu) + 0.135(800 - \mu) + 0.135(450 - \mu)$$

where μ is the herd average. The accuracy is:

$$r = \sqrt{[(0.372(2752) + 0.135(176) + 0.135(176))/2752]} = 0.712$$

The high accuracy is due to the inclusion of information from both parents.

Using means of records from animal and relatives

Example 1.8

It is given that average protein yield for the first two lactations for a cow ($\tilde{y}_1$) called Zena is 230 kg and the mean protein yield of five other cows ($\tilde{y}_2$), each with two lactations, is 300 kg. If all cows are all daughters of the same bull and no other relationship exists among them, predict the breeding value of Zena, assuming a heritability of 0.25, a repeatability (t) of 0.5, standard deviation of 34 kg and herd average of 250 kg for protein yield in the first two lactations.

From the given parameters:

$$g_{11} = \sigma_a^2 = h^2\sigma_y^2 = 0.25(34^2) = 289$$

and:

$$g_{12} = \text{covariance between half-sibs} = \tfrac{1}{4}(\sigma_a^2) = \tfrac{1}{4}(289) = 72.25$$

From calculations in Section 1.3.2:

$$p_{11} = \text{var}(\tilde{y}_1) = \left(t + \frac{(1-t)}{n}\right)\sigma_y^2$$

$$= (0.5 + (1 - 0.5)/2)34^2 = 867$$

Downloaded From Vet-eBooks.com

Using similar arguments:

$$p_{22} = \text{var}(\tilde{y}_2) = \sigma_B^2 + 1/n(\sigma_W^2)$$

where σ_B^2 is the between-cow variance and $1/n(\sigma_W^2)$ is the mean of the within-cow variance. From Section 1.4:

$$\sigma_B^2 = \tfrac{1}{4}\sigma_a^2$$

and for cow i in the group of five cows:

$$\sigma_W^2 = \text{var}(\tilde{y}_{2i} - \sigma_B^2)$$

where $\tilde{y}_{2i}$ is the mean of the first two lactations for cow i. Since all five cows each have two records like Zena:

$$\sigma_W^2 = (p_{11} - \tfrac{1}{4}\sigma_a^2)$$

and:

$$1/n(\sigma_W^2) = 1/n(p_{11} - \tfrac{1}{4}\sigma_a^2)$$

Therefore:

$$p_{22} = \tfrac{1}{4}\sigma_a^2 + 1/n(p_{11} - \tfrac{1}{4}\sigma_a^2)$$
$$= \tfrac{1}{4}(289) + (\tfrac{1}{5})(867 - \tfrac{1}{4}(289)) = 231.2$$

The index equations are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 867 & 72.25 \\ 72.25 & 231.2 \end{bmatrix}^{-1} \begin{bmatrix} 289 \\ 72.25 \end{bmatrix}$$

The solutions are $b_1 = 0.316$ and $b_2 = 0.213$ and the index is:

$$I = 0.316(230 - 250) + 0.213(300 - 250)$$

The accuracy of the index is:

$$r = \sqrt{\left[\left(0.316(289) + 0.213(72.5)/289\right)\right]} = 0.608$$

1.7.3 Prediction of aggregate genotype

At times, the aim is not just to predict the breeding value of a single trait but that of a composite of several traits evaluated in economic terms. The aggregate breeding value (H) or merit for such several or m traits can be represented as:

$$H = w_1a_1 + w_2a_2 + \ldots + w_ma_m$$

where a_i is the breeding value of the ith trait and w_i the weighting factor, which expresses the relative economic importance associated with the ith trait. The construction of an index to predict or improve H is based on the same principles as those discussed earlier except that it includes the relative economic weight for each trait.

Thus:

$$I = \mathbf{P}^{-1}\mathbf{Gw}(\mathbf{y} - \boldsymbol{\mu}) \qquad (1.22)$$

Downloaded From Vet-eBooks.com

where **w** is the vector of economic weights and all other terms are as defined in Eqn 1.17. The equations to be solved to get the weights (**b** values) to be used in the index are:

$$b_1p_{11} + b_2p_{12} + \cdots + b_mp_{1m} = w_1g_{11} + w_2g_{12} + \cdots w_mg_{1m}$$
$$b_1p_{21} + b_2p_{22} + \cdots + b_mp_{2m} = w_1g_{21} + w_2g_{22} + \cdots + w_mg_{m1}$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$b_1p_{m1} + b_2p_{m2} + \cdots + b_mp_{mm} = w_1g_{m1} + w_2g_{m2} + \cdots + w_mg_{mm}$$

In matrix notation these equations are:

$$\mathbf{Pb} = \mathbf{Gw}$$
$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$$

It should be noted that it is possible there are some traits in the index that are not in the aggregate breeding value but may be correlated with other traits in *H*. Conversely, some traits in the aggregate breeding value may be difficult to measure or occur late in life and may therefore not be in the index. Such traits may be replaced in the index with other highly correlated traits that are easily measurable or occur early in life. Consequently, the vector of economic weights may not necessarily be of the same dimension as traits in the index, as indicated in the equations for *b* above. Each trait in the index is weighted by the economic weight relevant to the breeding value of the trait it is predicting in the aggregate breeding value.

The index calculated using Eqn 1.22 implies that the same economic weights are applied to the traits in the aggregate genotype across the whole population. A change in the economic weight for one of the traits would imply recalculating the index. An alternative formulation of Eqn 1.22 involves calculating a sub-index for each trait in *H* without the economic weights. The final index in Eqn 1.23 is obtained by summing the sub-indices for each trait weighted by their respective economic weights. Thus:

$$I = \sum_{i=1}^{m} I_i w_i \tag{1.23}$$

where $I_i = \mathbf{P}^{-1}\mathbf{G}_i(\mathbf{y} - \boldsymbol{\mu})$, the sub-index for trait *i* in *H* and w_i = economic weight for trait *i*.

With Eqn 1.23, a change in the economic weights of any of the traits in the index can easily be implemented without recalculating the index.

To demonstrate that Eqns 1.22 and 1.23 are equivalent, assume that there are two traits in *H*, then Eqn 1.23 becomes:

$$I = I_1w_1 + I_2w_2$$
$$= \mathbf{P}^{-1}\mathbf{G}_1w_1(\mathbf{y} - \boldsymbol{\mu}) + \mathbf{P}^{-1}\mathbf{G}_2w_2(\mathbf{y} - \boldsymbol{\mu})$$

where $\mathbf{G}_i$ is the covariance matrix between trait *i* and all traits in the index. Thus:

$$I = \mathbf{P}^{-1}(\mathbf{G}_1w_1 + \mathbf{G}_2w_2)(\mathbf{y} - \boldsymbol{\mu})$$
$$= \mathbf{P}^{-1}\mathbf{Gw}(\mathbf{y} - \boldsymbol{\mu})$$

which is the same as Eqn 1.22.

Example 1.9

Assume the economic weights for ADG and LP are £1.5 and £0.5 per an increase of 1 kg in ADG and 1% increase in LP, respectively. Using the genetic parameters

Downloaded From Vet-eBooks.com

in Example 1.6, construct an index to select fast-growing lean beef calves using Eqn 1.22. Repeat the analysis using Eqn 1.23.

Using Eqn 1.22, index equations are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{-1} \begin{bmatrix} w_1 g_{11} + w_2 g_{12} \\ w_1 g_{21} + w_2 g_{22} \end{bmatrix}$$

Inserting the appropriate values:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 6400.00 & -57.60 \\ -57.60 & 51.84 \end{bmatrix}^{-1} \begin{bmatrix} 1.5(2752) + 0.5(62.064) \\ 1.5(62.064 + 0.5(15.552) \end{bmatrix}$$

Solutions for b_1 and b_2 from the above equations are 0.674 and 2.695, respectively. The index therefore is:

$$I = 0.674(\text{ADG} - \mu_{\text{ADG}}) + 2.694(\text{LP} - \mu_{\text{LP}})$$

Applying Eqn 1.23, the sub-index for ADG is the same as that calculated in Example 1.6 with $b_1 = 0.445$ and $b_2 = 1.692$.The sub-index for LP is:

$$b_1 p_{11} + b_2 p_{12} = g_{12}$$
$$b_1 p_{21} + b_2 p_{22} = g_{22}$$

which gives:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 6400.00 & -57.60 \\ -57.60 & 51.84 \end{bmatrix}^{-1} \begin{bmatrix} 62.064 \\ 15.552 \end{bmatrix}$$

The solutions are $b_1 = 0.0125$ and $b_2 = 0.314$. Multiplying the sub-indices by their respective weights gives:

$$I_{\text{ADG}} = 0.445(1.5)(\text{ADG} - \mu_{\text{ADG}}) + 1.692(1.5)(\text{LP} - \mu_{\text{LP}})$$
$$= 0.668(\text{ADG} - \mu_{\text{ADG}}) + 2.538(\text{LP} - \mu_{\text{LP}})$$

and:

$$I_{\text{LP}} = 0.0125(0.5)(\text{ADG} - \mu_{\text{ADG}}) + 0.314(0.5)(\text{LP} - \mu_{\text{LP}})$$
$$= 0.006(\text{ADG} - \mu_{\text{ADG}}) + 0.157(\text{LP} - \mu_{\text{LP}})$$

Summing the b terms from the two sub-indices, the final b terms are:

$$b_1 = 0.668 + 0.006 = 0.674$$
$$b_2 = 2.538 + 0.157 = 2.695$$

Therefore the final index is:

$$I = 0.675(\text{ADG} - \mu_{\text{ADG}}) + 2.695(\text{LP} - \mu_{\text{LP}})$$

which is the same as calculated using Eqn 1.22.

1.7.4 Overall economic indices using predicted genetic merit

Overall economic indices that combine (PTAs) or estimated breeding values (EBVs) calculated by best linear unbiased prediction (BLUP, see Chapter 3) have become

Downloaded From Vet-eBooks.com

very popular in the last decade. In addition to the recognition that more than one trait contributes to profitability, the broadening of selection goals has also been due to the need to incorporate health and welfare traits to accommodate public concerns. Examples of indices constructed with PTAs or BVs of several traits and used in genetic improvement of dairy cattle include production index (PIN), combining PTAs for milk, fat and protein in the UK, production life index (PLI), which is PIN plus PTAs for longevity and somatic cell count in the UK; and in the Netherlands, index net (INET), which combines BVs for milk, fat and protein and durable performance sum (DPS), which is INET plus durability (Interbull, 2000). The principles for calculating these indices are similar to those outlined in previous sections. Given that the PTAs or BVs are from a complete multivariate analysis, the optimal index weights (**b**) are the sum of the partial regression coefficients of each goal trait on each index trait, weighted by the economic value of the goal trait (Veerkamp *et al*., 1995). Thus given *m* traits in the selection goal and *n* traits in the index, then the partial regressions can be calculated as:

$$\mathbf{R} = \mathbf{G}^{-1}\mathbf{G}_{ig}$$

and:

$$\mathbf{b} = \mathbf{R}\mathbf{w}$$

where **R** is a matrix of partial genetic regression, $\mathbf{G}_{ig}$ is the matrix of genetic covariance between *m* goal and *n* index traits, **G** is the genetic covariance matrix between the index traits, and **w** is the vector of economic weights. It is obvious that when goal and index traits are the same, $\mathbf{G}_{ig} = \mathbf{G}$ and $\mathbf{b} = \mathbf{w}$. In the case where the index and goal traits are not the same, **R** can be estimated directly from a regression of phenotype on the EBVs for the index traits (Brotherstone and Hill, 1991). However, if PTAs or BVs are from a univariate analysis, rather than from a multivariate analysis, the use of **b** above results only in minimal loss of efficiency in the index (Veerkamp *et al*., 1995).

Selection based on breeding values from BLUP is usually associated with an increased rate of inbreeding as it favours the selection of closely related individuals. Quadratic indices can be used to optimize the rate of genetic gain and inbreeding. This does not fall within the main subject area of this text and interested readers should see the work by Meuwissen (1997) and Grundy *et al*. (1998).

1.7.5 Restricted selection index

Restricted selection index is used when the aim is to maximize selection for a given aggregate genotype, subject to the restriction that no genetic change is desired in one or more of the traits in the index for *H*. This is achieved by the usual index procedure and setting the covariance between the index and the breeding value ($\mathrm{cov}(I, a_i)$ for the *i*th trait specified not to change to zero. It was Kempthorne and Nordskog (1959) who introduced the idea of imposing restrictions on the general index procedure.

For instance consider the aggregate genotype composed of two traits:

$$H = w_1 a_1 + w_2 a_2$$

Downloaded From Vet-eBooks.com

However, it is desired that there should be no genetic change in trait 2; thus effectively:

$$H = w_1 a_1$$

and the index to predict H is:

$$I = b_1 y_1 + b_2 y_2$$

To ensure that there is no genetic change in trait 2, $\text{cov}(I, a_2)$ must be equal to zero. From Eqn 1.20:

$$\text{cov}(I, a_2) = b_1 g_{12} + b_2 g_{22} = 0$$

This is included as an extra equation to the normal equations for the b values, and a dummy unknown, the so-called Lagrange multiplier, is added to the vector of solutions for the index weights (Ronningen and Van Vleck, 1985). The equations for the index therefore are:

$$\begin{bmatrix} b_1 \\ b_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & g_{12} \\ p_{21} & p_{22} & g_{22} \\ g_{12} & g_{22} & 0 \end{bmatrix}^{-1} \begin{bmatrix} g_{11} \\ g_{12} \\ 0 \end{bmatrix} \tag{1.24}$$

Example 1.10

Using the same data and parameters as in Example 1.6, construct an index to improve the aggregate genotype for fast-growing lean cattle using an index consisting of GR and LP but with no genetic change in LP.

From Eqn 1.23 the index equations are:

$$\begin{bmatrix} 6400 & -57.60 & 62.064 \\ 57.60 & 51.80 & 15.552 \\ 62.064 & 15.552 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 2752 \\ 62.064 \\ 0 \end{bmatrix}$$

The solutions for b_1 and b_2 from solving the above equations are 0.325 and –1.303. Therefore the index is:

$$I = 0.325(\text{ADG} - \mu_{\text{ADG}}) + (-1.303(\text{LP} - \mu_{\text{LP}}))$$

The accuracy of this index (Eqn 1.21) is:

$$r = \sqrt{\left[(0.325)(2752) + \left(-1.303(62.064)\right)/2752\right]} = 0.544$$

which is lower than the accuracy for the equivalent index in Example 1.6, but with no restriction on LP, and is also lower than the accuracy of prediction of breeding value for ADG on the basis of single records. The imposition of a restriction on any trait in the index will never increase the efficiency of the index but usually reduces it unless $I_i = 0$ for the constrained trait.

Downloaded From Vet-eBooks.com

1.7.6 Index combining breeding values from phenotype and genetic marker information

Consider a situation in which one or more genes affecting a trait with a large impact on profit have been identified to be linked to a genetic marker (see Chapter 10). If genetic prediction based only on marker information is available in addition to the conventional BV estimated without marker information, then both sources of information can be combined into an index (Goddard, 1999). It is also possible that the conventional BV is based on a subset of traits in the breeding goal and marker information is available on other traits that are not routinely measured, such as meat quality traits.

A selection index could be used to combine both sources of information and the increase in accuracy from including marker information could be computed (Goddard, 1999). Given r as the accuracy of the conventional breeding BV and d as the proportion of genetic variance explained by the marker information, then the covariance between the two sources of information is dr^2. If m is the BV based on marker information and a the BV from phenotypic information, then:

$$\text{var}\begin{pmatrix} m \\ a \end{pmatrix} = \begin{pmatrix} d & dr^2 \\ dr^2 & r^2 \end{pmatrix}$$

Let g be the true breeding value to be predicted, then $\text{cov}(g, m) = d$ and $\text{cov}(g, a) = r^2$. The normal index equations are:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} d & dr^2 \\ dr^2 & r^2 \end{pmatrix}^{-1} \begin{pmatrix} d \\ r^2 \end{pmatrix}$$

Solving the above equations gives the following index weights:

$$b_1 = 1 - r^2/(1 - dr^2) \text{ and } b_2 = 1 - d/(1 - dr^2)$$

The variance of the index = reliability (r_I^2) is:

$$r_I^2 = [(1 - r^2)d + (1 - d)r^2]/(1 - dr^2)$$

The increase in reliability (r_{inc}^2) from incorporating marker information therefore is:

$$r_{inc}^2 = (r_I^2 - r^2) = d/(1 - dr^2)[(1 - r^2)^2]$$

For example, given that r^2 of the conventional BV is 0.34 and marker information accounts for 25% of the genetic variance, then r_I^2 is 0.459, an increase in reliability of 0.12. However, if r^2 is 0.81, then r_I^2 is 0.83 and r_{inc}^2 is only 0.02. Thus the usefulness of marker information is greater when reliability is low, such as in traits of low heritability and also traits that cannot be measured in young animals such as carcass traits (Goddard and Hayes, 2002).

Downloaded From Vet-eBooks.com

2 Genetic Covariance Between Relatives

2.1 Introduction

Of fundamental importance in the prediction of breeding values is the genetic relationship among individuals. From Chapter 1, it was found that use of the selection index to predict breeding values requires the genetic covariance between individuals to construct the genetic covariance matrix. Genetic evaluation using best linear unbiased prediction (BLUP), the subject of the next chapter, is heavily dependent on the genetic covariance among individuals, both for higher accuracy and for unbiased results. The genetic covariance among individuals is comprised of three components: the additive genetic variance, the dominance variance and the epistatic variance. This chapter addresses the calculation of the additive genetic relationship among individuals and how to determine the level of inbreeding. Dominance and epistasis genetic relationships are considered in Chapter 12, which deals with non-additive models.

2.2 The Numerator Relationship Matrix

The probability of identical genes by descent occurring in two individuals is termed the coancestry or the coefficient of kinship (Falconer and Mackay, 1996) and the additive genetic relationship between two individuals is twice their coancestry. The matrix that indicates the additive genetic relationship among individuals is called the numerator relationship matrix (**A**). It is symmetric and its diagonal element for animal i (a_{ii}) is equal to $1 + F_i$, where F_i is the inbreeding coefficient of animal i (Wright, 1922). The diagonal element represents twice the probability that two gametes taken at random from animal i will carry identical alleles by descent. The off-diagonal element, a_{ij}, equals the numerator of the coefficient of relationship (Wright, 1922) between animals i and j. When multiplied by the additive genetic variance (σ_u^2), $\mathbf{A}\sigma_u^2$ is the covariance among breeding values. Thus if u_i is the breeding value for animal i, $\text{var}(u_i) = a_{ii}\sigma_u^2 = (1 + F_i)\sigma_u^2$. The matrix **A** can be computed using path coefficients, but a recursive method that is suitable for computerization was described by Henderson (1976). Initially, animals in the pedigree are coded 1 to n and ordered such that parents precede their progeny. The following rules are then employed to compute **A**.

If both parents (s and d) of animal i are known:

$$a_{ji} = a_{ij} = 0.5(a_{js} + a_{jd}); \quad j = 1 \text{ to } (i - 1)$$
$$a_{ii} = 1 + 0.5(a_{sd})$$

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

Table 2.1. Pedigree for six animals.

Calf	Sire	Dam
3	1	2
4	1	Unknown
5	4	3
6	5	2

If only one parent s is known and assumed unrelated to the mate:

$a_{ji} = a_{ij} = 0.5(a_{js})$; $j = 1$ to $(i - 1)$
$a_{ii} = 1$

If both parents are unknown and are assumed unrelated:

$a_{ji} = a_{ij} = 0$; $j = 1$ to $(i - 1)$
$a_{ii} = 1$

For example, assume that the data in Table 2.1 are the pedigree for six animals. The numerator relationship matrix for the example pedigree is:

	1	2	3	4	5	6
1	1.00	0.00	0.50	0.50	0.50	0.25
2	0.00	1.00	0.50	0.00	0.25	0.625
3	0.50	0.50	1.00	0.25	0.625	0.563
4	0.50	0.00	0.25	1.00	0.625	0.313
5	0.50	0.25	0.625	0.625	1.125	0.688
6	0.25	0.625	0.563	0.313	0.688	1.125

For instance:

$a_{11} = 1 + 0 = 1$
$a_{12} = 0.5(0 + 0) = 0 = a_{21}$
$a_{22} = 1 + 0 = 1$
$a_{13} = 0.5(a_{11} + a_{12}) = 0.5(1.0 + 0) = 0.5 = a_{31}$
$a_{23} = 0.5(a_{12} + a_{22}) = 0.5(0 + 1.0) = 0.5 = a_{32}$
$\vdots$
$a_{34} = 0.5(a_{13}) = 0.5(0.5 + 0) = 0.25 = a_{43}$
$\vdots$
$a_{66} = 1 + 0.5(a_{52}) = 1 + 0.5(0.25) = 1.125$

From the above calculation, the inbreeding coefficient for calf 6 is 0.125.

2.3 Decomposing the Relationship Matrix

The relationship matrix can be expressed (Thompson, 1977a), as:

$$\mathbf{A} = \mathbf{TDT}' \tag{2.1}$$

Downloaded From Vet-eBooks.com

where **T** is a lower triangular matrix and **D** is a diagonal matrix. This relationship has been used to develop rules for obtaining the inverse of **A**. A non-zero element of the matrix **T**, say t_{ij}, is the coefficient of relationship between animals i and j, if i and j are direct relatives or $i = j$ and it is assumed that there is no inbreeding. Thus the matrix **T** traces the flow of genes from one generation to the other; in other words, it accounts only for direct (parent–offspring) relationships. It can easily be computed applying the following rules.

For the ith animal:

$t_{ii} = 1$

If both parents (s and d) are known:

$t_{ij} = 0.5(t_{sj} + t_{dj})$

If only one parent (s) is known:

$t_{ij} = 0.5(t_{sj})$

If neither parent is known:

$t_{ij} = 0$

The diagonal matrix **D** is the variance and covariance matrix for Mendelian sampling. The Mendelian sampling (m) for an animal i with breeding value u_i and u_s and u_d as breeding values for its sire and dam, respectively, is:

$$m_i = u_i - 0.5(u_s + u_d) \tag{2.2}$$

D has a simple structure and can easily be calculated. From Eqn 2.2, if both parents of animal i are known, then:

$$\begin{aligned}\text{var}(m_i) &= \text{var}(u_i) - \text{var}(0.5u_s + 0.5u_d)\\ &= \text{var}(u_i) - \text{var}(0.5u_s) - \text{var}(0.5u_d) + 2\text{cov}(0.5u_s, 0.5u_d)\\ &= (1 + F_i)\sigma_u^2 - 0.25a_{ss}\sigma_u^2 - 0.25a_{dd}\sigma_u^2 - 0.5a_{sd}\sigma_u^2\end{aligned}$$

where a_{ss}, a_{dd} and a_{sd} are elements of the relationship matrix **A**, and F_i is the inbreeding coefficient of animal i.

$$\text{var}(m_i)/\sigma_u^2 = d_{ii} = (1 + F_i) - 0.25a_{ss} - 0.25a_{dd} - 0.5a_{sd}$$

Since $F_i = 0.5a_{sd}$

$$\begin{aligned}d_{ii} &= 1 - 0.25(1 + F_s) - 0.25(1 + F_d)\\ &= 0.5 - 0.25(F_s + F_d)\end{aligned}$$

where F_s and F_d are the inbreeding coefficients of its sire and dam, respectively. If only one parent (s) is known, the diagonal element is:

$$\begin{aligned}d_{ii} &= 1 - 0.25(1 + F_s)\\ &= 0.75 - 0.25(F_s)\end{aligned}$$

and if no parent is known:

$d_{ii} = 1$

Downloaded From Vet-eBooks.com

For the pedigree in Table 2.1, the matrix **T** is:

	1	2	3	4	5	6
1	1.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0
3	0.5	0.5	1.0	0.0	0.0	0.0
4	0.5	0.0	0.0	1.0	0.0	0.0
5	0.5	0.25	0.5	0.5	1.0	0.0
6	0.25	0.625	0.25	0.25	0.5	1.0

and **D** is:

$$\mathbf{D} = \text{diag}(1.0, 1.0, 0.5, 0.75, 0.5, 0.469)$$

For instance, animal 4 has only the sire known, which is not inbred, therefore:

$$d_{44} = 0.75 - 0 = 0.75$$

and:

$$d_{66} = 0.5 - 0.25(0.125 + 0) = 0.469$$

because both parents are known and the sire has an inbreeding coefficient of 0.125.

2.4 Computing the Inverse of the Relationship Matrix

The prediction of breeding value requires the inverse of the relationship matrix, $\mathbf{A}^{-1}$. This could be obtained by setting up **A** by the recursive method and inverting it. This is, however, not computationally feasible when evaluating a large number of animals. In 1976, Henderson presented a simple procedure for calculating $\mathbf{A}^{-1}$ without setting up **A**. The procedure and its principles are described below.

From Eqn 2.1 the inverse of **A** can be written as:

$$\mathbf{A}^{-1} = (\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1} \tag{2.3}$$

The matrix $\mathbf{D}^{-1}$ is easy to obtain because **D** is a diagonal matrix. The diagonal elements of $\mathbf{D}^{-1}$ are simply the reciprocals of the diagonal elements of **D** computed in Section 2.3. $\mathbf{T}^{-1}$ is a lower triangular matrix with ones in the diagonals and the only non-zero elements to the left of the diagonal in the row for the animal i are –0.5 for columns corresponding to the known parents. It can be derived as **I** – **M**, where **I** is an identity matrix of the order of animals on the pedigree and **M** is a matrix of the contribution of gametes from parents to progeny (Kennedy, 1989). Since progeny i receives half of its genes from each parent, the only non-zero elements in row i of **M** are 0.5, corresponding to columns of known parents. Thus if both parents of progeny

Downloaded From Vet-eBooks.com

i are unknown, all elements of row *i* are zero. For the pedigree in Table 2.1, $\mathbf{T}^{-1}$ can be calculated as:

$$\underset{(\mathbf{I})}{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}} - \underset{(\mathbf{M})}{\begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.0 & 0.5 & 0.0 \end{bmatrix}}$$

$$= \underset{(\mathbf{T}^{-1})}{\begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & -0.5 & 1.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.5 & -0.5 & 1.0 & 0.0 \\ 0.0 & -0.5 & 0.0 & 0.0 & -0.5 & 1.0 \end{bmatrix}}$$

and:

$\mathbf{D}^{-1}$ = diag(1, 1, 2, 1.333, 2, 2.133)

2.4.1 Inverse of the numerator relationship matrix ignoring inbreeding

The relationship shown in Eqn 2.3 was used by Henderson (1976) to derive simple rules for obtaining $\mathbf{A}^{-1}$ without accounting for inbreeding. With inbreeding ignored, the diagonal elements of $\mathbf{D}^{-1}$ are either 2, or $\frac{4}{3}$ or 1 if both or one or no parents are known, respectively. Let α_i represent the diagonal element of $\mathbf{D}^{-1}$ for animal *i*. Initially set $\mathbf{A}^{-1}$ to zero and apply the following rules.

If both parents of the *i*th animal are known, add:

α_i to the (*i*,*i*) element

$-\alpha_i/2$ to the (*s*,*i*), (*i*,*s*), (*d*,*i*) and (*i*,*d*) elements

$\alpha_i/4$ to the (*s*,*s*), (*s*,*d*), (*d*,*s*) and (*d*,*d*) elements

If only one parent (*s*) of the *i*th animal is known, add:

α_i to the (*i*,*i*) element

$-\alpha_i/2$ to the (*s*,*i*) and (*i*,*s*) elements

$\alpha_i/4$ to the (*s*,*s*) element

If neither parent of the *i*th animal is known, add:

α_i to the (*i*,*i*) element

Downloaded From Vet-eBooks.com

As an illustration, the inverse of the relationship matrix in Section 2.2 can be calculated as below. Initially list all animals in the pedigree:

Calf	Sire	Dam
1	Unknown	Unknown
2	Unknown	Unknown
3	1	2
4	1	Unknown
5	4	3
6	5	2

Then set up a 6 × 6 table for the animals. For animals 1 and 2, both parents are unknown, therefore $\alpha_1 = \alpha_2 = 1$. Add 1 to their diagonal elements (1,1 and 2,2). For animal 3, both parents are known therefore $\alpha_3 = 2$. Add 2 to the 3,3 element, –1 to the (3,1), (1,3), (3,2) and (2,3) elements and $\frac{1}{2}$ to the (1,1), (1,2), (2,1) and (2,2) elements. For animal 4, only one parent is known, therefore $\alpha_4 = \frac{4}{3}$. Add $\frac{4}{3}$ to the (4,4) element, $-\frac{2}{3}$ to the (4,1) and (1,4) elements and $\frac{1}{3}$ to the (1,1) element. After the first four animals, the table is:

	1	2	3	4	5	6
1	$1 + \frac{1}{2} + \frac{1}{3}$	$\frac{1}{2}$	–1	$-\frac{2}{3}$		
2	$\frac{1}{2}$	$1 + \frac{1}{2}$	–1			
3	–1	–1	2			
4	$-\frac{2}{3}$			$\frac{4}{3}$		
5						
6						

After applying the relevant rules to animals 5 and 6, the inverse of **A** then is:

	1	2	3	4	5	6
1	1.83	0.5	–1.0	–0.67	0.0	0.0
2	0.5	2.0	–1.0	0.0	0.5	–1.0
3	–1.0	–1.0	2.5	0.5	–1.0	0.0
4	–0.67	0.0	0.5	1.83	–1.0	0.0
5	0.0	0.5	–1.0	–1.0	2.5	–1.0
6	0.0	–1.0	0.0	0.0	–1.0	2.0

Using Eqn 2.3, the inverse of **A** can be calculated directly. If inbreeding is ignored, **D** for the pedigree is:

$\mathbf{D}$ = diag(1.0, 1.0, 0.5, 0.75, 0.5, 0.5)

and:

$\mathbf{D}^{-1}$ = diag(1, 1, 2, 1.33, 2,2)

Downloaded From Vet-eBooks.com

Therefore the inverse of the relationship matrix using Eqn 2.3 is:

$$\underset{(\mathbf{T}^{-1})'}{\begin{bmatrix} 1.0 & 0.0 & -0.5 & -0.5 & 0.0 & 0.0 \\ 0.0 & 1.0 & -0.5 & 0.0 & 0.0 & -0.5 \\ 0.0 & 0.0 & 1.0 & 0.0 & -0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & -0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & -0.5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}} \underset{\mathbf{D}^{-1}}{\begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 2.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.33 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 2.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 2.00 \end{bmatrix}}$$

$$\underset{(\mathbf{T}^{-1})}{\begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & -0.5 & 1.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.5 & -0.5 & 1.0 & 0.0 \\ 0.0 & -0.5 & 0.0 & 0.0 & -0.5 & 1.0 \end{bmatrix}}$$

$$= \underset{\mathbf{A}^{-1}}{\begin{bmatrix} 1.83 & 0.50 & -1.00 & -0.67 & 0.00 & 0.00 \\ 0.50 & 2.00 & -1.00 & 0.00 & 0.50 & -1.00 \\ -1.00 & -1.00 & 2.50 & 0.50 & -1.00 & 0.00 \\ -0.67 & 0.00 & 0.50 & 1.83 & -1.00 & 0.00 \\ 0.00 & 0.50 & -1.00 & -1.00 & 2.50 & -1.00 \\ 0.00 & -1.00 & 0.00 & 0.00 & -1.00 & 2.00 \end{bmatrix}}$$

which is the same inverse obtained previously by applying the rules.

2.4.2 Inverse of the numerator relationship matrix accounting for inbreeding

The calculation of $\mathbf{A}^{-1}$ with inbreeding accounted for involves the application of the same rules outlined in Section 2.4.1 but $\mathbf{D}$ and therefore $\mathbf{D}^{-1}$ in Eqn 2.3 are calculated using the inbreeding coefficients of sires and dams (see Section 2.3). This implies that the diagonal elements of the relationship matrix are needed for $\mathbf{A}^{-1}$ to be properly calculated. This could be achieved by initially calculating the $\mathbf{A}$ for the group of animals and writing the diagonal elements to a file. The diagonal elements could be read from the file while computing $\mathbf{A}^{-1}$. For a large pedigree, this approach would require a large amount of memory for storage and be computationally demanding. However, Quaas (1976) presented a strategy for obtaining the diagonal elements of $\mathbf{A}$ while computing $\mathbf{A}^{-1}$ without setting up the relationship matrix.

Downloaded From Vet-eBooks.com

Recall from Section 2.3 that **A** can be expressed as:

$\mathbf{A} = \mathbf{TDT}'$
If $\mathbf{L} = \mathbf{T}\sqrt{\mathbf{D}}$
$\mathbf{A} = \mathbf{LL}'$ (2.4)

where **L** is a lower triangular matrix and, since **D** is diagonal, $\sqrt{\mathbf{D}}$ refers to a matrix obtained by calculating the square root of the diagonal elements of **D**. Equation 2.4 implies that the diagonal element of **A** for animal *i* is:

$$a_{ii} = \sum_{m=1}^{i} l_{im}^2 \quad (2.5)$$

Thus for a pedigree consisting of *m* animals:

$a_{11} = l_{11}^2$
$a_{22} = l_{21}^2 + l_{22}^2$
$a_{33} = l_{31}^2 + l_{32}^2 + l_{33}^2$
$\vdots$
$a_{mm} = l_{m1}^2 + l_{m2}^2 + l_{m3}^2 + \ldots + l_{mm}^2$

From the above, all the diagonal elements of **A** can be computed by calculating **L** one column at a time (Quaas, 1984). Only two vectors of dimension equal to the number of animals for storage will be required: one to store the column of **L** being computed and the second to accumulate the sum of squares of the elements of **L** for each animal. The matrices **L** and $\mathbf{A}^{-1}$ can be computed using the following procedure:

From Eqn 2.4 the diagonal element of **L** for animal *i* is:

$$l_{ii} = \sqrt{d_i}$$

$$l_{ii} = \sqrt{[0.5 - 0.25(F_s + F_d)]}$$

$$l_{ii} = \sqrt{[1.0 - 0.25(a_{ss} + a_{dd})]}; \quad \text{with } a_{ss} = 1 + F_{ss} \text{ and } a_{dd} = 1 + F_{dd}$$

Using equation [2.5]:

$$l_{ii} = \sqrt{\left[1.0 - 0.25\left(\sum_{m=1}^{s} l_{sm}^2 + \sum_{m=1}^{d} l_{dm}^2\right)\right]}$$

To set up $\mathbf{A}^{-1}$ at the same time, calculate the diagonal element of $\mathbf{D}^{-1}(\alpha_i)$ for animal *i* as $\alpha_i = 1/l_{ii}^2$. Then compute the contribution of animal *i* to $\mathbf{A}^{-1}$, applying the usual rules for computing $\mathbf{A}^{-1}$ (see Section 2.4.1).

The off-diagonal elements of **L** to the left of the diagonal for animal *i* are calculated as:

$l_{ij} = 0.5(l_{sj} + l_{dj})$; *s* and *d* equal to or greater than *j*

For the example pedigree used in Section 2.4.1 the **L** matrix is:

	1	2	3	4	5	6
1	1.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0
3	0.5	0.5	0.707	0.0	0.0	0.0
4	0.5	0.0	0.0	0.866	0.0	0.0
5	0.5	0.25	0.354	0.433	0.707	0.0
6	0.25	0.625	0.177	0.217	0.354	0.685

Downloaded From Vet-eBooks.com

and $\mathbf{A}^{-1}$ with inbreeding accounted for is:

	1	2	3	4	5	6
1	1.833	0.5	–1.0	–0.667	0.0	0.0
2	0.5	2.033	–1.0	0.0	0.533	–1.067
3	–1.0	–1.0	2.5	0.5	–1.0	0.0
4	–0.667	0.0	0.5	1.833	–1.0	0.0
5	0.0	0.533	–1.0	–1.0	2.533	–1.067
6	0.0	1.067	0.0	0.0	–1.067	2.133

The calculation columns of $\mathbf{L}$ and α_i for the first three animals are illustrated below:

$l_{11} = \sqrt{[1 - 0.25(0 + 0)]} = 1$

α_1 = 1 and its contribution to $\mathbf{A}^{-1}$ is computed using the rules in Section 2.4.1

$l_{21} = 0$

$l_{31} = 0.5(l_{11} + l_{21}) = 0.5(1 + 0) = 0.5$

$l_{41} = 0.5(l_{11}) = 0.5$

$l_{51} = 0.5(l_{41} + l_{31}) = 0.5(0.5 + 0.5) = 0.5$

$l_{61} = 0.5(l_{51} + l_{21}) = 0.5(0.5 + 0) = 0.25$

$l_{22} = \sqrt{[1 - 0.25(0 + 0)]} = 1$

α_2 = 1 and its contribution to $\mathbf{A}^{-1}$ is computed using the rules in Section 2.4.1

$l_{32} = 0.5(l_{12} + l_{22}) = 0.5(0 + 1) = 0.5$

$l_{42} = 0.5(l_{12}) = 0.5(0) = 0$

$l_{52} = 0.5(l_{42} + l_{32}) = 0.5(0 + 0.5) = 0.25$

$l_{62} = 0.5(l_{52} + l_{22}) = 0.5(0.25 + 1.0) = 0.625$

$$l_{33} = \sqrt{[1 - 0.25(l_{11}^2) - 0.25(l_{21}^2 + l_{22}^2)]}$$
$$= \sqrt{[1 - 0.25(1) - 0.25(0 + 1)]} = 0.707$$

$\alpha_3 = 1/(0.707)^2 = 2.0$ and its contribution to $\mathbf{A}^{-1}$ is computed using the usual rules

$l_{43} = 0.5(l_{13}) = 0.5(0) = 0$

$l_{53} = 0.5(l_{43} + l_{33}) = 0.5(0 + 0.707) = 0.354$

$l_{63} = 0.5(l_{53} + l_{23}) = 0.5(0.354 + 0) = 0.177$

Faster algorithms for computing the inverse of $\mathbf{A}$ accounting for inbreeding based on the $\mathbf{L}$ matrix have been published by Meuwissen and Luo (1992) and Quaas (unpublished note, 1995), and these are presented in Appendix B.

2.5 Inverse of the Relationship Matrix for Sires and Maternal Grandsires

In some cases, the prediction of breeding value is only for sires and maternal grandsires, the so-called sire and maternal grandsire (MGS) model. In such cases, the $\mathbf{A}^{-1}$ to be incorporated in the mixed model equations (MME) involves only sire and maternal grandsires and the rules for calculating $\mathbf{A}^{-1}$ are different from those discussed in the previous sections relating to pedigrees with individuals, sires and dams. With the

Downloaded From Vet-eBooks.com

MGS model, the relationship matrix **A** required pertains to males and can be approximated (Quaas, 1984) as:

$$a_{ii} = 1 + 0.25a_{sk} \tag{2.6}$$

$$a_{ij} = 0.5a_{sj} + 0.25a_{kj} \tag{2.7}$$

where s and k are the sires and maternal grandsires, respectively, for sire i. When all maternal granddams are unrelated (base animals) and there are no maternal half-sibs, the above will yield the exact **A**.

The inverse of approximate **A** can be calculated from a list of sires and maternal grandsires, applying Eqn 2.3. In this case, $\mathbf{T}^{-1}$ is a lower triangular matrix with ones in the diagonal and the only non-zero elements to the left of the diagonal in the row for the ith animal are –0.5 and –0.25 for the columns corresponding to the sire and maternal grandsire, respectively. The elements of **D** and therefore $\mathbf{D}^{-1}$ can be calculated in a manner similar to that described in Sections 2.3 and 2.4. The diagonal elements of **D** (d_{ii}) for animal i are calculated by the following rules.

If both sire (s) and maternal grandsire (k) are known:

$$d_{ii} = [\text{var}(u_i) - \text{var}(\tfrac{1}{2}u_s + \tfrac{1}{4}u_k)]/\sigma_u^2$$

where the u terms are breeding values. Following the same arguments as in Section 2.3:

$$d_{ii} = \tfrac{11}{16} - \tfrac{1}{4}F_s - \tfrac{1}{16}F_k$$

where F_s and F_k are inbreeding coefficients for sire and maternal grandsire, respectively.

When only the maternal grandsire is known:

$$d_{ii} = [\text{var}(u_i) - \text{var}(\tfrac{1}{4}u_k)]/\sigma_u^2$$

$$d_{ii} = \tfrac{15}{16} - \tfrac{1}{16}F_k$$

When only the sire is known or no parents are known, d_{ii} is as calculated in Section 2.3.

The elements of $\mathbf{D}^{-1}$ are reciprocals of **D**, calculated above. Using Eqn 2.3, $\mathbf{A}^{-1}$ can be calculated on the basis of $\mathbf{T}^{-1}$ and $\mathbf{D}^{-1}$, defined above, as follows:

Initially, set $\mathbf{A}^{-1}$ to zero.

If both sire (s) and maternal grandsire (k) of animal i are known, add:

d_{ii}^{-1} to the (i,i) element

$-d_{ii}^{-1}/2$ to the (s,i) and (i,s) elements

$-d_{ii}^{-1}/4$ to the (k,i) and (i,k) elements

$d_{ii}^{-1}/4$ to the (s,s) element

$d_{ii}^{-1}/8$ to the (s,k) and (k,s) elements

$d_{ii}^{-1}/16$ to the (k,k) element

Downloaded From Vet-eBooks.com

Without inbreeding, $d_{ii}^{-1} = \frac{16}{11}$.

If only the maternal grandsire (k) of animal i is known, add:

d_{ii}^{-1} to the (i,i) element

$-d_{ii}^{-1}/4$ to the (k,i) and (i,k) elements

$d_{ii}^{-1}/16$ to the (k,k) element

Without inbreeding, $d_{ii}^{-1} = \frac{16}{15}$.

If only the sire (s) of animal i is known, add:

d_{ii}^{-1} to the (i,i) element

$-d_{ii}^{-1}/2$ to the (s,i) and (i,s) elements

$d_{ii}^{-1}/4$ to the (s,s) element

Without inbreeding, $d_{ii}^{-1} = \frac{4}{3}$ in this situation, as in Section 2.4.1.

When s and k are unknown, add:

d_{ii}^{-1} to the (i,i) element

and $d_{ii}^{-1} = 1$.

2.6 An Example of the Inverse of a Sire and Maternal Grandsire Relationsip Matrix

A pedigree consisting of sires and maternal grandsires set up from the pedigree in Table 2.1 is:

Sire	Sire of sire	Maternal grandsire of sire
1	Unknown	Unknown
4	1	Unknown
5	4	1

Recoding sires 1 to n, the pedigree becomes:

1	Unknown	Unknown
2	1	Unknown
3	2	1

Using Eqns 2.6 and 2.7, $\mathbf{A} = \begin{bmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.625 \\ 0.5 & 0.625 & 1.125 \end{bmatrix}$

Note that the relationship among sires is the same as in **A** calculated from the full pedigree in Section 2.2.

The $\mathbf{T}^{-1}$ matrix for the pedigree is: $\mathbf{T}^{-1} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ -0.5 & 1.0 & 0.0 \\ -0.25 & -0.5 & 1.0 \end{bmatrix}$

Downloaded From Vet-eBooks.com

and:

$$\mathbf{D}^{-1} = \text{diag}(1, \tfrac{4}{3}, \tfrac{16}{11})$$

Applying Eqn 2.3, $\mathbf{A}^{-1}$ is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.0 & -0.5 & -0.25 \\ 0.0 & 1.0 & -0.5 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{3} & 0 \\ 0 & 0 & \frac{16}{11} \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ -0.5 & 1.0 & 0.0 \\ -0.25 & -0.5 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 1.424 & -0.485 & -0.364 \\ -0.485 & 1.697 & -0.727 \\ -0.364 & -0.727 & 1.455 \end{bmatrix}$$

To calculate the inverse of the sire and maternal grandsire relationship matrix, applying the rules given earlier, initially set $\mathbf{A}^{-1}$ to zero. The elements of $\mathbf{D}^{-1}$ have already been given above. Processing the first animal, add 1 (d^{-1}_{11}) to the diagonal element (1,1) of $\mathbf{A}^{-1}$. For the second animal, add $\frac{4}{3}$ (d^{-1}_{22}) to the diagonal element (2,2) of $\mathbf{A}^{-1}$, $\frac{1}{3}$ to the (1,1) element and $-\frac{2}{3}$ to the (1,2) and (2,1) elements. Finally processing the third animal, add $\frac{16}{11}$(d^{-1}_{33}) to the (3,3) element of $\mathbf{A}^{-1}$, $-\frac{16}{11}$ to the (3,4) and (4,3) elements, $-\frac{16}{22}$ to the (1,3) and (3,1) elements, $\frac{16}{44}$ to the (4,4) element, $\frac{16}{88}$ to the (1,4) and (4,1) elements and $\frac{16}{176}$ to the (1,1) element. This gives the same $\mathbf{A}^{-1}$ as previously calculated using Eqn 2.3.

In the next chapter, the incorporation of $\mathbf{A}^{-1}$ in the MME for the prediction of breeding value using BLUP is addressed.

Downloaded From Vet-eBooks.com

3 Best Linear Unbiased Prediction of Breeding Value: Univariate Models with One Random Effect

3.1 Introduction

In Chapter 1, the use of the selection index (best linear prediction) for genetic evaluation was examined; however, it is associated with some major disadvantages. First, records may have to be pre-adjusted for fixed or environmental factors and these are assumed to be known. These are not usually known, especially when no prior data exist for new sub-classes of fixed effect or new environmental factors. Second, solutions to the index equations require the inverse of the covariance matrix for observations and this may not be computationally feasible for large data sets.

Henderson (1949) developed a methodology called best linear unbiased prediction (BLUP), by which fixed effects and breeding values can be simultaneously estimated. The properties of the methodology are similar to those of a selection index and the methodology reduces to selection indices when no adjustments for environmental factors are needed. The properties of BLUP are more or less incorporated in the name:

- Best – means it maximizes the correlation between true (a) and predicted breeding value ($\hat{a}$) or minimizes prediction error variance (PEV) (var($a - \hat{a}$)).
- Linear – predictors are linear functions of observations.
- Unbiased – estimation of realized values for a random variable such as animal breeding values, and of estimable functions of fixed effects are unbiased (E($a = \hat{a}$)).
- Prediction – involves prediction of true breeding value.

BLUP has found widespread usage in genetic evaluation of domestic animals because of its desirable statistical properties. This has been enhanced by the steady increase in computing power and has evolved in terms of its application to simple models, such as the sire model, in its early years, to more complex models such as the animal, maternal, multivariate and random regression models, in recent years. Several general purpose computer packages for BLUP evaluations such as PEST (Groeneveld *et al*., 1990), BREEDPLAN, Mix 99 (Lidauer *et al*., 2011) and a host of others have been written and made available. In this chapter, BLUP's theoretical background is briefly presented, considering a univariate animal model, and its application to several univariate models in genetic evaluation is illustrated.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

3.2 Brief Theoretical Background

Consider the following equation for a mixed linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \tag{3.1}$$

where:

$\mathbf{y}$ = $n \times 1$ vector of observations; n = number of records
$\mathbf{b}$ = $p \times 1$ vector of fixed effects; p = number of levels for fixed effects
$\mathbf{a}$ = $q \times 1$ vector of random animal effects; q = number of levels for random effects
$\mathbf{e}$ = $n \times 1$ vector of random residual effects
$\mathbf{X}$ = design matrix of order $n \times p$, which relates records to fixed effects
$\mathbf{Z}$ = design matrix of order $n \times q$, which relates records to random animal effects

Both $\mathbf{X}$ and $\mathbf{Z}$ are termed incidence matrices.

It is assumed that the expectations (E) of the variables are:

$$E(\mathbf{y}) = \mathbf{Xb}; \quad E(\mathbf{a}) = E(\mathbf{e}) = 0$$

and it is assumed that residual effects, which include random environmental and non-additive genetic effects, are independently distributed with variance σ_e^2; therefore, $var(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}$; $var(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G}$ and $cov(\mathbf{a}, \mathbf{e}) = cov(\mathbf{e}, \mathbf{a}) = 0$, where $\mathbf{A}$ is the numerator relationship matrix.

Then:

$$\begin{aligned} var(\mathbf{y}) = \mathbf{V} &= var(\mathbf{Za} + \mathbf{e}) \\ &= \mathbf{Z}var(\mathbf{a})\mathbf{Z}' + var(\mathbf{e}) + cov(\mathbf{Za}, \mathbf{e}) + cov(\mathbf{e}, \mathbf{Za}) \\ &= \mathbf{ZGZ}' + \mathbf{R} + \mathbf{Z}cov(\mathbf{a}, \mathbf{e}) + cov(\mathbf{e}, \mathbf{a})\mathbf{Z}' \end{aligned}$$

Since $cov(\mathbf{a}, \mathbf{e}) = cov(\mathbf{e}, \mathbf{a}) = 0$, then:

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R} \tag{3.2}$$

$$\begin{aligned} cov(\mathbf{y}, \mathbf{a}) &= cov(\mathbf{Za} + \mathbf{e}, \mathbf{a}) \\ &= cov(\mathbf{Za}, \mathbf{a}) + cov(\mathbf{e}, \mathbf{a}) \\ &= \mathbf{Z}cov(\mathbf{a}, \mathbf{a}) \\ &= \mathbf{ZG} \end{aligned}$$

and:

$$\begin{aligned} cov(\mathbf{y}, \mathbf{e}) &= cov(\mathbf{Za} + \mathbf{e}, \mathbf{e}) \\ &= cov(\mathbf{Za}, \mathbf{e}) + cov(\mathbf{e}, \mathbf{e}) \\ &= \mathbf{Z}cov(\mathbf{a}, \mathbf{e}) + cov(\mathbf{e}, \mathbf{e}) \\ &= \mathbf{R} \end{aligned}$$

The general problem with respect to Eqn 3.1 is to predict a linear function of $\mathbf{b}$ and $\mathbf{a}$, that is, $\mathbf{k}'\mathbf{b} + \mathbf{a}$ (predictand), using a linear function of $\mathbf{y}$, say $\mathbf{L}'\mathbf{y}$ (predictor), given that $\mathbf{k}'\mathbf{b}$ is estimable. The predictor $\mathbf{L}'\mathbf{y}$ is chosen such that it is unbiased (i.e. its expected value is equal to the expected value of the predictand) and PEV is minimized. This minimization leads to the BLUP of $\mathbf{a}$ (Henderson, 1973) as:

$$\hat{\mathbf{a}} = BLUP(\mathbf{a}) = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{3.3}$$

and:

$$\mathbf{L}'\mathbf{y} = \mathbf{k}'\hat{\mathbf{b}} + \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

Downloaded From Vet-eBooks.com

where $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, the generalized least square solution (GLS) for **b**, and $\mathbf{k}'\hat{\mathbf{b}}$ is the best linear unbiased estimator (BLUE) of $\mathbf{k}'\mathbf{b}$, given that $\mathbf{k}'\mathbf{b}$ is estimable. BLUE is similar in meaning and properties to BLUP but relates to estimates of linear functions of fixed effects. It is an estimator of the estimable functions of fixed effects that has minimum sampling variance, is unbiased and is based on the linear function of the data (Henderson, 1984). An outline for the derivation of Eqn 3.3 and the equation for $\mathbf{L}'\mathbf{y}$ above are given in Appendix C, Section C.1.

As mentioned in Section C.1, BLUP is equivalent to the selection index with the GLS of $\hat{\mathbf{b}}$ substituted for **b** in Eqn 3.3. Alternatively, this could simply be illustrated (W.G. Hill, Edinburgh, 1995, personal communication) by considering the index to compute breeding values for a group of individuals with relationship matrix **A**, which have records with known mean. From Eqn 1.17, the relevant matrices are then:

$$\mathbf{P} = \mathbf{I}\sigma_e^2 + \mathbf{A}\sigma_a^2 \quad \text{and} \quad \mathbf{G} = \mathbf{A}\sigma_a^2$$

with:

$$\alpha = \sigma_e^2/\sigma_a^2 \quad \text{or} \quad (1 - h^2)/h^2$$

Hence:

$$I = \mathbf{P}^{-1}\mathbf{G}\mathbf{y} = (\mathbf{I} + \alpha\mathbf{A}^{-1})^{-1}\mathbf{y}$$

which is similar to the BLUP (Eqn 3.3) assuming fixed effects are absent and with $\mathbf{Z} = \mathbf{I}$.

The solutions for **a** and **b** in Eqn 3.3 require $\mathbf{V}^{-1}$, which is not always computationally feasible. However, Henderson (1950) presented the mixed model equations (MME) to estimate solutions **b** (fixed effects solutions) and predict solutions for random effects (**a**) simultaneously without the need for computing $\mathbf{V}^{-1}$. The proof that solutions for **b** and **a** from MME are the GLS of **b** and the BLUP of **a** is given in Appendix C, Section C.2. The MME for Eqn 3.1 are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

assuming that **R** and **G** are non-singular. Since $\mathbf{R}^{-1}$ is an identity matrix from the earlier definition of **R** in this section, it can be factored out from both sides of the equation to give:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \tag{3.4}$$

Note that the MME may not be of full rank, usually due to dependency in the coefficient matrix for fixed environmental effects. It may be necessary to set certain levels of fixed effects to zero when there is dependency to obtain solutions to the MME (see Section 3.6). However, the equations for **a** (Eqn 3.3) are usually of full rank since **V** is usually positive definite and **Xb** is invariant to the choice of constraint.

Some of the basic assumptions of the linear model for the prediction of breeding value were given in Section 1.2. The solutions to the MME give the BLUE of $\mathbf{k}'\mathbf{b}$ and

Downloaded From Vet-eBooks.com

BLUP of **a** under certain assumptions, especially when data span several generations and may be subject to selection. These assumptions are:

1. Distributions of **y**, **u** and **e** are assumed to be multivariate normal, implying that traits are determined by many additive genes of infinitesimal effects at many infinitely unlinked loci (infinitesimal model, see Section 1.2). With the infinitesimal model, changes in genetic variance resulting from selection, such as gametic disequilibrium (negative covariance between frequencies of genes at different loci), or from inbreeding and genetic drift, are accounted for in the MME through the inclusion of the relationship matrix (Sorensen and Kennedy, 1983), as well as assortative mating (Kemp, 1985).
2. The variances and covariances (**R** and **G**) for the base population are assumed to be known or at least known to proportionality. In practice, variances and covariances of the base population are never known exactly but, assuming the infinitesimal model, these can be estimated by restricted (or residual) maximum likelihood (REML) if data include information on which selection is based.
3. The MME can account for selection if based on a linear function of **y** (Henderson, 1975) and there is no selection on information not included in the data.

The use of these MME for the prediction of breeding values and estimation of fixed effects under an animal model is presented in the next section.

3.3 A Model for an Animal Evaluation (Animal Model)

Example 3.1
Consider the data set in Table 3.1 for the pre-weaning gain (WWG) of beef calves (calves assumed to be reared under the same management conditions).

The objective is to estimate the effects of sex and predict breeding values for all animals. Assume that $\sigma^2_a = 20$ and $\sigma^2_e = 40$, therefore $\alpha = \frac{40}{20} = 2$.

The model to describe the observations is:

$$y_{ij} = p_i + a_j + e_{ij}$$

where: y_{ij} = the WWG of the jth calf of the ith sex; p_i = the fixed effect of the ith sex; a_j = random effect of the jth calf; and e_{ij} = random error effect. In matrix notation the model is the same as that described in Eqn 3.1.

Table 3.1. Pre-weaning gain (kg) for five beef calves.

Calves	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

Downloaded From Vet-eBooks.com

3.3.1 Constructing the mixed model equations

The matrix **X** in the MME relates records to fixed (sex) effects. For the example data set, its transpose is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

The first row indicates that the first, fourth and fifth observations are from male calves and the second row shows the second and third records are from female calves.

The **Z** matrix relates records to all animals – those with or without yield records. In this case, animals 1 to 3 are parents with no records and animals 4 to 8 are recorded. Thus for the example data, **Z** is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that the first three columns of **Z** are zeros and these correspond to the animals 1 to 3, which are parents without records.

The vector **y** is simply the vector of the observations. For the data set under consideration, it is:

$$\mathbf{y}' = [4.5 \quad 2.9 \quad 3.9 \quad 3.5 \quad 5.0]$$

Having set up the matrices **X**, **Z** and **y**, the other matrices in the MME, such as **X′Z**, **Z′X**, **X′y** and **Z′y** are easily obtained by matrix multiplication. In practice, these matrices are not calculated through multiplication from the design matrices and vector of observations but are usually set up or computed directly. However, for the example data set, these matrices are:

$$\mathbf{X}'\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{Z}'\mathbf{X} \text{ is the transpose of } \mathbf{X}'\mathbf{Z}$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 13.0 \\ 6.8 \end{pmatrix} \text{ and the transpose of } \mathbf{Z}'\mathbf{y} \text{ is } (0 \quad 0 \quad 0 \quad 4.5 \quad 2.9 \quad 3.9 \quad 3.5 \quad 5.0)$$

The matrix **Z′Z** is a diagonal matrix, with the first three diagonal elements zeros and the next five elements all ones.

The various matrices in the MME have been calculated, apart from $\mathbf{A}^{-1}\alpha$. With these matrices, we can set up what are known as the least squares equations (LSE) as:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Downloaded From Vet-eBooks.com

For the example data set, the LSE are:

$$\begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 13.0 \\ 6.8 \\ 0 \\ 0 \\ 0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{bmatrix}$$

The addition of $\mathbf{A}^{-1}\alpha$ to $\mathbf{Z}'\mathbf{Z}$ in the LSE yields the MME. Using the rules outlined in Section 2.4.1, $\mathbf{A}^{-1}$ for the example data is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 & 0.000 & 0.000 \\ 0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 & 0.000 & 0.000 \\ 0.000 & 0.500 & 2.000 & 0.000 & -1.000 & 0.500 & 0.000 & -1.000 \\ -0.667 & 0.000 & 0.000 & 1.833 & 0.500 & 0.000 & -1.000 & 0.000 \\ 0.000 & -1.000 & -1.000 & 0.500 & 2.500 & 0.000 & -1.000 & 0.000 \\ -1.000 & -1.000 & 0.500 & 0.000 & 0.000 & 2.500 & 0.000 & -1.000 \\ 0.000 & 0.000 & 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & -1.000 & 0.000 & 2.000 \end{bmatrix}$$

and $\mathbf{A}^{-1}\alpha$ is easily obtained by multiplying every element of $\mathbf{A}^{-1}$ by 2, the value of α. Adding $\mathbf{A}^{-1}\alpha$ to $\mathbf{Z}'\mathbf{Z}$, the MME for the example data are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_1 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000 \end{bmatrix}^{-1} \begin{bmatrix} 13.0 \\ 6.8 \\ 0.0 \\ 0.0 \\ 0.0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{bmatrix}$$

Solving the MME by direct inversion of the coefficient matrix gives the following solutions:

Sex effects		Animals							
Males	Females	1	2	3	4	5	6	7	8
4.358	3.404	0.098	−0.019	−0.041	−0.009	−0.186	0.177	−0.249	0.183

Downloaded From Vet-eBooks.com

The solutions indicate that male calves have a higher rate of gain up to weaning than females calves, which is consistent with the raw averages for males and females. From the first row in the MME (Eqn 3.4), the equations for sex effect are:

$$(\mathbf{X'X})\hat{\mathbf{b}} = \mathbf{X'y} - (\mathbf{X'Z})\hat{\mathbf{a}}$$
$$\hat{\mathbf{b}} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{y} - \mathbf{Z}\hat{\mathbf{a}})$$

Thus the solution for the ith level of sex effect may be written as:

$$\hat{b}_i = \left(\sum_j y_{ij} - \sum_j \hat{a}_{ij}\right) \Big/ \text{diag}_i \tag{3.5}$$

where y_{ij} is the record and $\hat{a}_{ij}$ is the solution of the jth animals within the sex subclass i and diag_i is the sum of observations for the sex subclass i. For instance, the solution for male calves is:

$$b_1 = [(4.5 + 3.5 + 5.0) - (-0.009 + -0.249 + 0.183)]/3 = 4.358$$

The equations for animal effects from the second row of Eqn 3.4 are:

$$(\mathbf{Z'Z} + \mathbf{A}^{-1}\alpha)\hat{\mathbf{a}} = \mathbf{Z'y} - (\mathbf{Z'X})\hat{\mathbf{b}}$$
$$(\mathbf{Z'Z} + \mathbf{A}^{-1}\alpha)\hat{\mathbf{a}} = \mathbf{Z'}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$
$$(\mathbf{Z'Z} + \mathbf{A}^{-1}\alpha)\hat{\mathbf{a}} = (\mathbf{Z'Z})\mathbf{YD} \tag{3.6}$$

with $\mathbf{YD} = (\mathbf{Z'Z})^{-1}\mathbf{Z'}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$, where **YD** is the vector of yield deviations (YDs) and represents the yields of the animal adjusted for all effects other than genetic merit and error. The matrix $\mathbf{A}^{-1}$ has non-zero off-diagonals only for the animal's parents, progeny and mates (see Section 2.4), transferring off-diagonal terms to the right-hand side of Eqn 3.6 gives the equation for animal i with k progeny as:

$$(\mathbf{Z'Z} + u_{ii}\alpha)\hat{a}_i = \alpha u_{ip}(\hat{a}_s + \hat{a}_d) + (\mathbf{Z'Z})YD + \alpha\sum_k u_{im}(\hat{a}_{anim} - 0.5\hat{a}_m)$$

where u_{ip} is the element of the $\mathbf{A}^{-1}$ between animal i and its parents with the sign reversed, and u_{im} is the element of $\mathbf{A}^{-1}$ between the animal and the dam of the kth progeny.

Therefore:

$$(\mathbf{Z'Z} + u_{ii}\alpha)\hat{a}_i = \alpha u_{par}(PA) + (\mathbf{Z'Z})YD + 0.5\alpha\sum_k u_{prog}(2\hat{a}_{anim} - \hat{a}_m) \tag{3.7}$$

where PA is the parent average, $u_{par} = 2(u_{ip})$, with u_{ip} equal to 1, $\frac{2}{3}$; or $\frac{1}{2}$ if both, one or neither parents are known and $u_{prog} = u_{im}$, with u_{im} equal to 1 when the mate of animal i is known or $\frac{2}{3}$ when the mate is not known.

Multiplying both sides of the equation by $(\mathbf{Z'Z} + u_{ii}\alpha)^{-1}$ (VanRaden and Wiggans, 1991) gives:

$$a_i = n_1(PA) + n_2(YD) + n_3(PC) \tag{3.8}$$

where:

$$PC = \sum_k u_{prog}(2\hat{a}_{anim} - \hat{a}_m)/\sum_k u_{prog}$$

is regarded as the progeny contribution and n_1, n_2 and n_3 are weights that sum to one. The derivation of the equation for PC is given in Appendix C, Section C.3. The numerators of n_1, n_2 and n_3 are αu_{par}, $\mathbf{Z'Z}$ (number of records the animal has) and $0.5\alpha\Sigma_k u_{prog}$, respectively. The denominator of all three n terms is the sum of the three numerators.

Downloaded From Vet-eBooks.com

From Eqn 3.8, the breeding value for an animal is dependent on the amount of information available on that animal. For base animals, *YD* in the equation does not exist and $\hat{a}_s$ and $\hat{a}_d$ are zeros with no genetic groups in the model; therefore, the solutions for these animals are a function of the contributions from their progeny breeding values adjusted for the mate solutions (*PC*). For instance, the proof for sire 1 in Example 3.1 can be calculated from the contributions from its progeny (calves 4 and 6) using Eqn 3.8 as:

$$\hat{a}_1 = n_1(0) + n_3\left[\left(\tfrac{2}{3}\right)(2\hat{a}_4) + (1)(2\hat{a}_6 - \hat{a}_2)\right]\Big/\left(\tfrac{2}{3}+1\right)$$
$$\hat{a}_1 = n_1(0) + n_3\left[\left(\tfrac{2}{3}\right)(-0.018) + (1)(0.354 - (-0.019))\right]\Big/\left(\tfrac{2}{3}+1\right)$$
$$\hat{a}_1 = n_3(0.2166) = 0.098$$

with $n_1 = \frac{\alpha}{3.667}$ and $n_3 = 0.5\alpha(\frac{2}{3} + 1)/3.667$ and 3.667 is the sum of the numerators of n_1 and n_3. The higher breeding value for sire 1 compared with sire 3 is due to the fact that the progeny of sire 1 have higher proofs after correcting for the solutions of the mates.

The solutions for an animal with a record but with no progeny depend on the average contributions from its parents and its yield deviation. Equation 3.8 reduces to:

$$a_i = n_1(PA) + n_2(YD)$$

Thus for progeny 8, its EBV can be calculated as:

$$a = n_1(\hat{a}_3 + \hat{a}_6)/2 + n_2(y_8 - b_1)$$
$$= n_1(0.068) + n_2(5.0 - 4.358) = 0.183$$

with $n_1 = \frac{2\alpha}{5}$, $n_2 = \frac{1}{5}$ and 5 is the sum of the numerators of n_1 and n_2.

It can also be demonstrated that for an animal with a record but with no progeny its solution is a function of an estimate of Mendelian sampling (*m*) and parent average. From Equation c.8 in Appendix C, Section C.3, the solution for calf *i* can be written as:

$$(1 + u_{ii}\alpha)\hat{a}_i + \alpha u_{cs}\hat{a}_s + \alpha u_{cd}\hat{a}_d = y_i$$

Therefore:

$$\hat{a}_i = (1 + u_{ii}\alpha)^{-1}[y_i - \alpha u_{is}\hat{a}_s - \alpha u_{id}\hat{a}_d]$$

If there is no inbreeding, $u_{is} = u_{id} = -0.5u_{ii}$. Therefore:

$$\hat{a}_i = (1 + u_{ii}\alpha)^{-1}[y_i + 0.5u_{ii}\alpha(\hat{a}_s + \hat{a}_d)]$$
$$= (1 + u_{ii}\alpha)^{-1}[(y_i - 0.5(\hat{a}_s + \hat{a}_d)) + 0.5(1 + u_{ii}\alpha)(\hat{a}_s + \hat{a}_d)]$$
$$= (1 + u_{ii}\alpha)^{-1}(y_i - 0.5(\hat{a}_s + \hat{a}_d)) + 0.5(\hat{a}_s + \hat{a}_d)$$
$$\hat{a}_i = 0.5(a_s + a_d) + m_i \qquad (3.9)$$

where $m_i = k(y_i - 0.5\hat{a}_s - 0.5\hat{a}_d)$ is an estimate of Mendelian sampling, and $k = 1/(1 + d^{-1}\alpha)$, with $d = \frac{1}{2}$ if both parents of animal *i* are known or $\frac{3}{4}$ if only one parent is known. Alternatively, the weight (*k*) can also be derived as:

$$k = \text{cov}(m, y_c)/\text{var}(y_c) = \text{cov}(m, m + e)/(\text{var}(m) + \text{var}(e))$$

where y_c is the yield record corrected for fixed effects and parent average.

$$k = \text{var}(m)/(\text{var}(m)+\text{var}(e))$$
$$= d\sigma_a^2/(d\sigma_a^2+\sigma_e^2)$$
$$= dh^2/(dh^2 + (1- h^2))$$

Downloaded From Vet-eBooks.com

where d, as defined earlier, equals $\frac{1}{2}$, $\frac{3}{4}$ or 1 if both, one or no parents are known, respectively. Using the parameters for Example 3.1 and assuming both parents known, $k = 10/(10 + 40) = 0.2$.

Thus for progeny 8, its EBV can be calculated as:

$$\begin{aligned}\hat{a}_8 &= 0.5(\hat{a}_3 + \hat{a}_6) + k(y_3 - b_1 - 0.5(\hat{a}_3 + \hat{a}_6)) \\ &= 0.5(-0.041 + 0.177) + 0.2(5.0 - 4.358 - 0.5(-0.041 + 0.177)) \\ &= 0.183\end{aligned}$$

Compared with calf 7, the proof of calf 8 is higher because it has a higher parent average solution and higher estimate of Mendelian sampling.

In the case of an animal with records and having progeny, there is an additional contribution from its offspring to its breeding value. Thus the breeding values of progeny 4 and 6 using Eqn 3.8 are:

$$\begin{aligned}\hat{a}_4 &= n_1(\hat{a}_1/2) + n_2(y_4 - b_1) + n_3(2(\hat{a}_7) - \hat{a}_5) \\ &= n_1(0.098/2) + n_2(4.5 - 4.358) + n_3(2(-0.249) - (-0.186)) = -0.009\end{aligned}$$

with $n_1 = 2\alpha(\frac{2}{3})/4.667$, $n_2 = 1/4.667$ and $n_3 = 0.5\alpha/4.667$; 4.676 = the sum of the numerators of n_1, n_2 and n_3; and:

$$\begin{aligned}\hat{a}_6 &= n_1((\hat{a}_1 + \hat{a}_2)/2) + n_2(y_6 - b_2) + n_3(2(\hat{a}_8) - \hat{a}_3) \\ &= n_1((0.098 + -0.019)/2) + n_2(3.9 - 3.404) + n_3(2(0.183) - (-0.041)) \\ &= 0.177\end{aligned}$$

with $n_1 = \frac{2\alpha}{6}$, $n_2 = \frac{1}{6}$ and $n_3 = \frac{0.5\alpha}{6}$; 6 = the sum of the numerators of n_1, n_2 and n_3.

Although contributions from parent average to both calves are similar, differences in progeny contributions resulted in a higher breeding value for calf 6, accounting for about 75% of the difference in the predicted breeding values between both calves.

3.3.2 Progeny (daughter) yield deviation

The yield deviation of a progeny contributes indirectly to the breeding value of its sire after it has been combined with information from parents and the offspring of the progeny (see Eqn 3.8). Thus progeny contribution is a regressed measure and it is not an independent measure of progeny performance as information from parents and the progeny's offspring is included. VanRaden and Wiggans (1991) indicated that a more independent and unregressed measure of progeny performance is progeny yield deviation (PYD). However, they called it daughter yield deviation (DYD) as they were dealing with the dairy cattle situation and records were only available for daughters of bulls. PYD or DYD can simply be defined as a weighted average of corrected yield deviation of all progeny of a sire; the correction is for all fixed effects and the breeding values of the mates of the sire.

DYD has been used for various purposes in dairy cattle evaluation and research. It was used in the early 1990s for the calculation of conversion equations to convert bull evaluations across several countries (Goddard, 1985). It was initially the variable of choice for international evaluations of dairy bulls by Interbull, but, due to the inability of several countries to calculate DYD, deregressed proofs were used (Sigurdsson and Banos, 1995). In addition, Interbull methods for the validation of genetic trends in national evaluations prior to acceptance for international evaluations utilize DYDs (Boichard *et al*., 1995). DYDs are also commonly employed in

Downloaded From Vet-eBooks.com

dairy cattle studies aimed at detecting quantitative trait loci using the granddaughter design (Weller, 2001). The equation for calculating DYD from univariate animal model evaluations was presented by VanRaden and Wiggans (1991) and its derivation is briefly outlined here.

For the progeny (*prog*) of a bull *i* that has no offspring of her own, Eqn 3.8 becomes:

$$\hat{a}_{prog} = n_{1prog}PA + n_{2prog}YD \tag{3.10}$$

Substituting Eqn 3.10 into the equation for PC in Eqn 3.8 gives:

$$\begin{aligned} PC &= \sum_k u_{prog}[2(n_{1prog}PA + n_{2prog}YD) - \hat{a}_{mi}] \Big/ \sum_k u_{prog} \\ &= \sum_k u_{prog}[n_{1prog}(\hat{a}_i + \hat{a}_{mi}) + n_{2prog}2YD - \hat{a}_{mi}] \Big/ \sum_k u_{prog} \end{aligned}$$

where n_{1prog} and n_{2prog} are the n_1 and n_2 of progeny. Since these progeny have no offspring of their own, n_{3prog} equals zero; therefore n_{1prog} equals 1 – n_{2prog}. Then:

$$\begin{aligned} PC &= \sum_k u_{prog}[(1 - n_{2prog})(\hat{a}_i + \hat{a}_{mi}) + n_{2prog}2YD - \hat{a}_{mi}] \Big/ \sum_k u_{prog} \\ &= \sum_k u_{prog}[(1 - n_{2prog})\hat{a}_i + n_{2prog}(2YD - \hat{a}_{mi})] \Big/ \sum_k u_{prog} \\ &= \hat{a}_i + \sum_k u_{prog}[n_{2prog}(-\hat{a}_i + 2YD - \hat{a}_{mi})] \Big/ \sum_k u_{prog} \end{aligned} \tag{3.11}$$

Substituting Eqn 3.11 into Eqn 3.8 and accumulating all terms involving $\hat{a}_i$ to the left side gives:

$$\begin{aligned} &\hat{a}_i - n_3\hat{a}_i + n_3 \sum_k u_{prog}n_{2prog}\hat{a}_i \Big/ \sum_k u_{prog} \\ &\quad = n_1PA + n_2YD + n_3 \sum_k u_{prog}n_{2prog}(2YD - \hat{a}_{mi}) \Big/ \sum_k u_{prog} \end{aligned}$$

Therefore:

$$\begin{aligned} &\left(1 - n_3 + n_3 \sum_k u_{prog}n_{2prog} \Big/ \sum_k u_{prog}\right)\hat{a}_i \\ &\quad = n_1PA + n_2YD + n_3 \sum_k u_{prog}n_{2prog}(2YD - \hat{a}_m) \Big/ \sum_k u_{prog} \end{aligned}$$

Substituting $(n_1 + n_2)$ for 1 – n_3 and removing the common denominator of the *n* terms from both sides of the equation, with DYD as:

$$\text{DYD or PYD} = \sum_k u_{prog}n_{2prog}(2YD - \hat{a}_m) \Big/ \sum_k u_{prog}n_{2prog} \tag{3.12}$$

the breeding value of animal *i* can be expressed as:

$$\hat{a}_i = w_1PA + w_2YD + w_3\text{DYD} \tag{3.13}$$

where the weights w_1, w_2 and w_3 sum to unity. The numerators of w_1 and w_2 are equal to those of n_1 and n_2 in Eqn 3.8. The numerator of:

$$w_3 = 0.5\alpha\sum\nolimits_k u_{prog}n_{2prog}$$

which is derived as n_3 times:

$$\sum\nolimits_k u_{prog}n_{2prog} \Big/ \sum\nolimits_k u_{prog}$$

Downloaded From Vet-eBooks.com

As VanRaden and Wiggans (1991) indicated, w_3 is always less than unity and therefore less than n_3, which reflects that PYD or DYD is an unregressed measure of progeny performance. Note that, for bulls with granddaughters, PYD or DYD does not include information from these granddaughters. Also, in the dairy cattle situation, the information from sons is not included in the calculation of DYD.

Illustrating the calculation of PYD or DYD

The computation of DYD is usually carried out in dairy cattle evaluations and it is illustrated later for a dairy data set in Example 4.1. Using the beef data in Example 3.1, the calculation of PYD is briefly illustrated for animal 3, using information on both female and male progeny, since observations are available on both sexes.

First, the YDs for both progeny of sire 3 are calculated:

$$YD_5 = (y_5 - b_2) = (2.9 - 3.404) = -0.504$$
$$YD_8 = (y_8 - b_1) = (5.0 - 4.358) = 0.642$$

Therefore, using Eqn 3.12:

$$\begin{aligned} \text{PYD}_3 &= n_{2(5)}u_{(5)}(2YD_5 - \hat{a}_2) + n_{2(8)}u_{(8)}(2YD_8 - \hat{a}_6)/(n_{2(5)}u_{(5)} + n_{2(8)}u_{(8)}) \\ &= 0.2(1)(-1.008 - (-0.019) + 0.2(1)(1.284 - 0.177)/(0.2(1) + 0.2(1)) \\ &= 0.059 \end{aligned}$$

where $n_{2(j)}$ and $u_{(j)}$ are the n_2 and u for the jth progeny. Note that in calculating $n_{2(j)}$, it has been assumed that progeny j has no offspring. Thus $n_{2(5)} - 1/(1 + 2\alpha(1)) = 0.2$.

Using Eqn 3.12 to calculate the breeding value of sire 3 gives the value of 0.0098, with $w_1 = 0.833$ and $w_2 = 0.167$. This is different from the breeding value reported from solving the MME as the granddaughter information (calf 7) has not been included.

3.3.3 Accuracy of evaluations

The accuracy (r) of predictions is the correlation between true and predicted breeding values. However, in dairy cattle evaluations, the accuracy of evaluations is usually expressed in terms of reliability, which is the squared correlation between true and predicted breeding values (r^2). The calculation for r or r^2 requires the diagonal elements of the inverse of the MME, as shown by Henderson (1975).

If the coefficient matrix of the MME in Eqn 3.4 is represented as:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \text{and a generalized inverse of the coefficient matrix as:} \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

Henderson (1975) showed that:

$$\text{PEV} = \text{var}(a - \hat{a}) = C^{22}\sigma_e^2 \tag{3.14}$$

Thus the diagonal elements of the coefficient matrix for animal equations are needed to calculate PEV for animals. The PEV could be regarded as the fraction of additive genetic variance not accounted for by the prediction. Therefore, for animal i, it could be expressed as:

$$\text{PEV}_i = C_i^{22}\sigma_e^2 = (1 - r^2)\sigma_a^2$$

Downloaded From Vet-eBooks.com

where r^2 is the squared correlation between the true and EBVs. Thus:

$$d_i\sigma_e^2 = (1 - r^2)\sigma_a^2$$

where d_i is the ith diagonal element of $\mathbf{C}^{22}$.

$$d_i\sigma_e^2/\sigma_a^2 = 1 - r^2$$
$$r^2 = 1 - d_i\alpha$$

and the accuracy (r) is just the square root of reliability.

From Eqn 3.14 the standard error of prediction (SEP) is:

$$\text{SEP} = \sqrt{\text{var}\,(a - \hat{a})}$$
$$= \sqrt{d_i\sigma_e^2} \text{ for animal } i$$

Note also that:

$$r^2 = 1 - (\text{SEP}^2/\sigma_a^2)$$

The inverse of the coefficient matrix for Example 3.1 is:

$$\begin{bmatrix}
0.596 & 0.157 & -0.164 & -0.084 & -0.131 & -0.265 & -0.148 & -0.166 & -0.284 & -0.238 \\
0.157 & 0.802 & -0.133 & -0.241 & -0.112 & -0.087 & -0.299 & -0.306 & -0.186 & -0.199 \\
-0.164 & -0.133 & 0.471 & 0.007 & 0.033 & 0.220 & 0.045 & 0.221 & 0.139 & 0.134 \\
-0.084 & -0.241 & 0.007 & 0.492 & -0.010 & 0.020 & 0.237 & 0.245 & 0.120 & 0.111 \\
-0.131 & -0.112 & 0.033 & -0.010 & 0.456 & 0.048 & 0.201 & 0.023 & 0.126 & 0.218 \\
-0.265 & -0.087 & 0.220 & 0.020 & 0.048 & 0.428 & 0.047 & 0.128 & 0.243 & 0.123 \\
-0.148 & -0.299 & 0.045 & 0.237 & 0.201 & 0.047 & 0.428 & 0.170 & 0.220 & 0.178 \\
-0.166 & -0.306 & 0.221 & 0.245 & 0.023 & 0.128 & 0.170 & 0.442 & 0.152 & 0.219 \\
-0.284 & -0.186 & 0.139 & 0.120 & 0.126 & 0.243 & 0.220 & 0.152 & 0.442 & 0.168 \\
-0.238 & -0.199 & 0.134 & 0.111 & 0.218 & 0.123 & 0.178 & 0.219 & 0.168 & 0.422
\end{bmatrix}$$

The r^2, r and SEP for animals in Example 3.1 are:

Animal	Diagonals of inverse	r^2	r	SEP
1	0.471	0.058	0.241	4.341
2	0.492	0.016	0.126	4.436
3	0.456	0.088	0.297	4.271
4	0.428	0.144	0.379	4.138
5	0.428	0.144	0.379	4.138
6	0.442	0.116	0.341	4.205
7	0.442	0.116	0.341	4.205
8	0.422	0.156	0.395	4.109

In the example, the reliabilities of animals with records are generally higher than those of ancestors since each has only two progeny. The two calves in the female sex subclass are progeny of dam 2 and this may explain the very low reliability for this ancestor as the effective number of daughters is reduced. The amount of information

Downloaded From Vet-eBooks.com

on calves 4 and 5 is very similar; each has a record, a common sire and parents of the same progeny, hence they have the same reliability. Calf 8 has the highest reliability and this is due to the information from the parents (its sire has another progeny and the dam has both parents known) and its record. The standard errors are large due to the small size of the data set but follow the same pattern as the reliabilities.

In practice, obtaining the inverse of the MME for large populations is not feasible and various methods have been used to approximate the diagonal element of the inverse. A methodology published by Meyer (1989) is presented in Appendix D and was used in the national dairy evaluation programme in Canada (Wiggans *et al.*, 1992) in the 1990s.

3.4 A Sire Model

The application of a sire model implies that only sires are being evaluated using progeny records. Most early applications of BLUP for the prediction of breeding values, especially in dairy cattle, were based on a sire model. The main advantage with a sire model is that the number of equations is reduced compared with an animal model since only sires are evaluated. However, with a sire model, the genetic merit of the mate (dam of progeny) is not accounted for. It is assumed that all mates are of similar genetic merit and this can result in bias in the predicted breeding values if there is preferential mating.

The sire model in matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e} \tag{3.15}$$

All terms in Eqn 3.15 are as defined for Eqn 3.1 and **s** is the vector of random sire effects, **Z** now relates records to sires and:

$$\text{var}(\mathbf{s}) = \mathbf{A}\sigma_s^2$$
$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_s^2 + \mathbf{R}$$

where **A** is the numerator relationship matrix for sires, $\sigma_s^2 = 0.25\sigma_a^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$. The MME are exactly the same as in Eqn 3.4 except that $\alpha = \sigma_e^2/\sigma_s^2 = (4 - h^2)/h^2$.

3.4.1 An illustration

Example 3.2

An application of a sire model is illustrated below using the same data as for the animal model evaluation in Table 3.1. Assigning records to sires, and including the pedigree for sires, the data can be presented as:

Sex of progeny	Sire	Sire of sire	Dam of sire	WWG (kg)
Male	1	–	–	4.5
Female	3	–	–	2.9
Female	1	–	–	3.9
Male	4	1	–	3.5
Male	3	–	–	5.0

Downloaded From Vet-eBooks.com

The objective is to estimate sex effects and predict breeding values for sires 1, 3 and 4. Using the same parameters as in Section 3.3, $\sigma_s^2 = 0.25(20) = 5$ and $\sigma_e^2 = 60 - 5 = 55$, therefore $\alpha = 55/5 = 11$.

Setting up the design matrices and MME

The design matrix **X** relating records to sex is as defined in Section 3.3.1. However, **Z** is different and its transpose is:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

indicating that sires 1 and 3 have two records each while sire 4 has only one record. The vector of observations **y** is as defined in Section 3.3.1. The matrices **X′X**, **X′Z**, **Z′X**, **Z′Z**, **X′y** and **Z′y** in the MME can easily be calculated through matrix multiplication. Thus:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{X}'\mathbf{Z} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{Z}'\mathbf{Z} = \text{diag}(2,2,1), \ \mathbf{X}'\mathbf{y} \text{ is as in Section 3.3.1}$$

and the transpose of $\mathbf{Z}'\mathbf{y} = (\mathbf{Z}'\mathbf{y})' = [8.4 \quad 7.9 \quad 3.5]$

The LSE are:

$$\begin{bmatrix} 3 & 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{s}_1 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \begin{bmatrix} 13.00 \\ 6.80 \\ 8.40 \\ 7.90 \\ 3.50 \end{bmatrix}$$

Apart from the fact that sire 4 is the son of sire 1, no other relationships exist among the three sires. Therefore $\mathbf{A}^{-1}$ for the three sires is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.333 & 0.0 & -0.667 \\ 0.000 & 1.0 & 0.000 \\ -0.667 & 0.0 & 1.333 \end{bmatrix}$$

The MME obtained after adding $\mathbf{A}^{-1}\alpha$ to $\mathbf{Z}'\mathbf{Z}$ in the LSE are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{s}_1 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} \begin{bmatrix} 3.000 & 0.000 & 1.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 1.000 & 1.000 & 0.000 \\ 1.000 & 1.000 & 16.666 & 0.000 & -7.334 \\ 1.000 & 1.000 & 0.000 & 13.000 & 0.000 \\ 1.000 & 0.000 & -7.334 & 0.000 & 15.666 \end{bmatrix}^{-1} = \begin{bmatrix} 13.00 \\ 6.80 \\ 8.40 \\ 7.90 \\ 3.50 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

The solutions to the MME by direct inversion of the coefficient matrix are:

Sex effects		Sires		
Males	Females	1	3	4
4.336	3.382	0.022	0.014	–0.043

The difference between solutions for sex subclasses, **L′b**, where **L** is [1 – 1], is the same as in the animal model. However, sire proofs and differences between sire proofs ($s_i - s_j$) are different from those from the animal model, although the ranking for the three sires is the same in both models. The differences in the proofs are due to the lack of adjustment for breeding values of mates in the sire model and differences in progeny contributions under both models. In this example, most of the differences in sire solutions under both models are due to differences in progeny contributions. The proofs for these sires under the animal model are based on their progeny contributions, since their parents are unknown. This contribution from progeny includes information from progeny yields and those of grand-offspring of the sires. However, in the sire model, progeny contributions include information from only male grand-offspring of the sires in addition to progeny yields. The effect of this difference on sire proofs under the two models is illustrated for two bulls below.

From the calculations in Section 3.3.1, the proportionate contribution of calves 4 and 6 to the proof of sire 1 in the animal model are –0.003 and 0.102, respectively. Using Eqn 3.8, the contribution of information from the different yield records to sire 1 under the sire model are as follows.

Contributions (CONT) from yields for calves 4 and 6 are:

$$\text{CONT}_4 = n_2(0.082) = 0.010$$
$$\text{CONT}_6 = n_2(0.259) = 0.031$$

where $n_2 = 2/16.667$.

Contributions from yield record for male grand-progeny (calf 7) through animal 4 (progeny) is:

$$\text{CONT}_7 = n_3(-0.086) = -0.019$$

where $n_3 = 3.667/16.667$.
Therefore:

$$s_1 = \text{CONT}_4 + \text{CONT}_6 + \text{CONT}_7 = 0.022$$

In the sire model the sum of CONT_4 and CONT_7 is equivalent to the contribution from calf 4 to the sire proof in the animal model. Thus the main difference in the proof for sire 1 in the two models is due largely to the lower contribution of calf 6 in the sire model. This lower contribution arises from the fact the contribution is only from the yield record in the sire model while it is from the yield and the progeny of calf 6 in the animal model.

Similar calculations for sire 3 indicate that the proportionate contributions from its progeny are –0.088 for calf 5 and 0.047 for calf 8 in the animal model. However, in the sire model the contributions are –0.037 and 0.051, respectively, from the yield of these calves. Again, the major difference here is due to the contribution from calf 5, which contains information from her offspring (calf 7) in the animal model.

Downloaded From Vet-eBooks.com

The similarity of the contributions of calf 8 to the proof of sire 3 in both models is because it is a non-parent and the contribution is slightly higher under the sire model due to the lack of adjustment for the breeding value of the mate.

3.5 Reduced Animal Model

In Section 3.2, the BLUP of breeding value involved setting up equations for every animal, that is, all parents and progeny. Thus the order of the animal equations was equal to the number of animals being evaluated. If equations were set up only for parents, this would greatly reduce the number of equations to be solved, especially since the number of parents is usually less than the number of progeny in most data sets. Breeding values of progeny can be obtained by back-solving from the predicted parental breeding values. Quaas and Pollak (1980) developed the reduced animal model (RAM), which allowed equations to be set up only for parents in the MME, and breeding values of progeny are obtained by back-solving from the predicted parental breeding values. This section presents the theoretical background for the RAM and illustrates its use for the prediction of breeding values.

3.5.1 Defining the model

The application of a RAM involves setting up animal equations for parents only and representing the breeding values of non-parents in terms of parental breeding value. Thus for the non-parent i, its breeding value can be expressed as:

$$a_i = \tfrac{1}{2}(a_s + a_d) + m_i \tag{3.16}$$

where a_s and a_d are the breeding values of sire and dam and m_i is the Mendelian sampling. It was shown in Section 2.3 that:

$$\text{var}(m_i) = (0.5 - 0.25(F_s + F_d))\sigma_a^2$$

Let $F = (F_s + F_d)/2$, then:

$$\begin{aligned}\text{var}(m_i) &= (0.5 - 0.5(F))\sigma_a^2 \\ &= 0.5(1 - F)\sigma_a^2\end{aligned} \tag{3.17}$$

The animal model applied in Section 3.3 was:

$$y_{ijk} = p_i + a_j + e_{ij} \tag{3.18}$$

In matrix notation:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \tag{3.19}$$

The terms in the above equations have been defined in Section 3.3.

Using Eqn 3.18, Eqn 3.19 can be expressed as:

$$y_{ijk} = p_i + \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + m_j + e_{ijk} \tag{3.20}$$

For non-parents, the terms m_j and e_{ijk} can be combined to form a single residual term e^*_{ijk} as:

$$e^*_{ijk} = m_j + e_{ijk} \tag{3.21}$$

Downloaded From Vet-eBooks.com

and:

$$\text{var}(e^*_{ijk}) = \text{var}(m_j) + \text{var}(e_{ijk})$$

Using Eqn 3.19:

$$\text{var}(e^*_{ijk}) = \tfrac{1}{2}(1 - F)\sigma^2_a + \sigma^2_e$$

In general:

$$\text{var}(m_j) = d_j(1 - F_j)\sigma^2_a \tag{3.22}$$

where d_j equals $\frac{1}{2}$ or $\frac{3}{4}$ or 1 if both, one or no parents are known, respectively, and F_j is the average inbreeding for both parents or, if only one parent is known, it is the inbreeding coefficient of the known parent. F_j equals zero when no parent is known. Ignoring inbreeding:

$$\text{var}(e^*_{ijk}) = \sigma^2_e + d_j\sigma^2_a = (1 + d_j\alpha^{-1})\sigma^2_e$$

Equation 3.20 can be expressed in matrix notation as:

$$\mathbf{y} = \mathbf{X}_n\mathbf{b} + \mathbf{Z}_1\mathbf{a}_p + \mathbf{e}^* \tag{3.23}$$

where $\mathbf{X}_n$ is the incidence matrix that relates non-parents' records to fixed effects, $\mathbf{Z}_1$ is an incidence matrix of zeros and halves identifying the parents of animals, and $\mathbf{a}_p$ is a vector of breeding values of parents.

The application of RAM involves applying the model:

$$\mathbf{y}_p = \mathbf{X}_p\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

for parents and the model:

$$\mathbf{y}_n = \mathbf{X}_n\mathbf{b} + \mathbf{Z}_1\mathbf{a}_p + \mathbf{e}^*$$

for non-parents.

From the above two equations, the model for RAM analysis can be written as:

$$\begin{bmatrix}\mathbf{y}_p\\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{bmatrix}\mathbf{b} + \begin{bmatrix}\mathbf{Z}\\ \mathbf{Z}_1\end{bmatrix}\mathbf{a}_p + \begin{bmatrix}\mathbf{e}\\ \mathbf{e}^*\end{bmatrix}$$

If:

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix}\mathbf{Z}\\ \mathbf{Z}_1\end{bmatrix} \quad \text{and} \quad \mathbf{R} = \begin{bmatrix}\mathbf{R}_p\\ \mathbf{R}_n\end{bmatrix} = \begin{bmatrix}\mathbf{I}\sigma^2_e & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma^{2*}_e\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}+\mathbf{D}\alpha^{-1}\end{bmatrix}\sigma^2_e$$

Then:

$$\text{var}(\mathbf{y}) = \mathbf{W}\mathbf{A}_p\mathbf{W}'\sigma^2_e + \mathbf{R}$$
$$\text{var}(\mathbf{a}_p) = \mathbf{A}_p\sigma^2_a$$

where $\mathbf{A}_p$ is the relationship matrix among parents and $\mathbf{D}$ above is a diagonal matrix with elements as defined for d_j in Eqn 3.22.

The MME to be solved are:

$$\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{a}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W}\\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{A}^{-1}1/\sigma^2_a\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix} \tag{3.24}$$

Downloaded From Vet-eBooks.com

Equation 3.24 can also be written as:

$$\begin{bmatrix} \mathbf{X}_p'\mathbf{R}_p^{-1}\mathbf{X}_p + \mathbf{X}_n'\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{X}_p'\mathbf{R}_p^{-1}\mathbf{Z} + \mathbf{X}_n'\mathbf{R}_n^{-1}\mathbf{Z}_1 \\ \mathbf{Z}'\mathbf{R}_p^{-1}\mathbf{X}_p + \mathbf{Z}_1'\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}'\mathbf{R}_p^{-1}\mathbf{Z} + \mathbf{Z}_1'\mathbf{R}_n^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}1/\sigma_a^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p'\mathbf{R}_p^{-1}\mathbf{y}_p + \mathbf{X}_n'\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}'\mathbf{R}_p^{-1}\mathbf{y}_p + \mathbf{Z}_1'\mathbf{R}_n^{-1}\mathbf{y}_n \end{bmatrix}$$

Multiplying the equations above by $\mathbf{R}_p$ gives:

$$\begin{bmatrix} \mathbf{X}_p'\mathbf{X}_p + \mathbf{X}_n'\mathbf{R}_v^{-1}\mathbf{X}_n & \mathbf{X}_p'\mathbf{Z} + \mathbf{X}_n'\mathbf{R}_v^{-1}\mathbf{Z}_1 \\ \mathbf{Z}'\mathbf{X}_p + \mathbf{Z}_1'\mathbf{R}_v^{-1}\mathbf{X}_n & \mathbf{Z}'\mathbf{Z} + \mathbf{Z}_1'\mathbf{R}_v^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p'\mathbf{y}_p + \mathbf{X}_n'\mathbf{R}_v^{-1}\mathbf{y}_n \\ \mathbf{Z}'\mathbf{y}_p + \mathbf{Z}_1'\mathbf{R}_v^{-1}\mathbf{y}_n \end{bmatrix} \tag{3.25}$$

where $\mathbf{R}_v^{-1}$equals $1/(1 + \mathbf{D}\alpha^{-1})$.

3.5.2 An illustration

Example 3.3

The application of RAM using Eqn 3.24 for the prediction of breeding values is illustrated below with the same data set (Table 3.1) as in Example 3.1 for the animal model evaluation. The genetic parameters are $\sigma_a^2 = 20.0$ and $\sigma_e^2 = 40.0$.

Constructing the MME

First we need to set up **R**, the matrix of residual variances and its inverse. In the example data set, animals 4, 5 and 6 are parents; therefore the diagonal elements in **R** corresponding to these animals are equal to σ_e^2, that is, 40.0.

Calves 7 and 8 are non-parents, therefore the diagonal elements for these animals in **R** are equal to $\sigma_e^2 + d_i\sigma_a^2$, assuming that the average inbreeding coefficients of the parents of these animals equal zero. For each calf, d_i equals $\frac{1}{2}$ because both their parents are known, therefore $r_{77} = r_{88} = 40 + \frac{1}{2}(20) = 50$.

The matrix **R** for animals with records is:

$$\mathbf{R} = \text{diag}(40, 40, 40, 50, 50)$$

and:

$$\mathbf{R}^{-1} = \text{diag}(0.025, 0.025, 0.025, 0.020, 0.020)$$

The matrix X is the same as in Section 3.3.1 and relates records to sex effects.

Therefore:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} 0.065 & 0.000 \\ 0.000 & 0.050 \end{bmatrix}$$

For the matrix **W**, the rows for parents with records (animals 4, 5 and 6) consist of zeros except for the columns corresponding to these animals, which contain ones, indicating that they have records. However, the rows for non-parents with records (animals 7 and 8) contain halves in the columns that correspond to their parents, and otherwise zeros. Thus:

Downloaded From Vet-eBooks.com

$$\mathbf{W} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.5 \end{bmatrix}$$

and:

$$\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.005 & 0.0 & 0.0 & 0.005 \\ 0.0 & 0.0 & 0.0 & 0.03 & 0.005 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.005 & 0.03 & 0.0 \\ 0.0 & 0.0 & 0.005 & 0.0 & 0.0 & 0.03 \end{bmatrix}$$

The transpose of the vector of observations, $\mathbf{y}$, is as defined in Section 3.3.1. The remaining matrices, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{W}$, $\mathbf{W}'\mathbf{R}^{-1}\mathbf{X}$, $\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$ and $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$ can easily be calculated through matrix multiplication since $\mathbf{X}$, $\mathbf{R}^{-1}$, $\mathbf{W}$ and $\mathbf{y}$ have been set up. Therefore:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{W} = \begin{bmatrix} 0.000 & 0.000 & 0.010 & 0.035 & 0.010 & 0.010 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.025 & 0.025 \end{bmatrix}$$

The matrix $\mathbf{W}'\mathbf{R}^{-1}\mathbf{X}$ is the transpose of $\mathbf{X}'\mathbf{R}^{-1}\mathbf{W}$.

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} = \begin{bmatrix} 0.282 \\ 0.170 \end{bmatrix} \quad \text{and} \quad \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} = \begin{bmatrix} 0.000 \\ 0.000 \\ 0.050 \\ 0.148 \\ 0.107 \\ 0.148 \end{bmatrix}$$

The LSE are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 0.065 & 0.000 & 0.000 & 0.000 & 0.010 & 0.035 & 0.010 & 0.010 \\ 0.000 & 0.050 & 0.000 & 0.000 & 0.000 & 0.000 & 0.025 & 0.025 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 & 0.005 & 0.000 & 0.000 & 0.005 \\ 0.035 & 0.000 & 0.000 & 0.000 & 0.000 & 0.030 & 0.005 & 0.000 \\ 0.010 & 0.025 & 0.000 & 0.000 & 0.000 & 0.005 & 0.030 & 0.000 \\ 0.010 & 0.025 & 0.000 & 0.000 & 0.005 & 0.000 & 0.000 & 0.03 \end{bmatrix}^{-1} \begin{bmatrix} 0.282 \\ 0.170 \\ 0.000 \\ 0.000 \\ 0.050 \\ 0.148 \\ 0.107 \\ 0.148 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

The relationship matrix is only for parents, that is, animals 1 to 6. Thus:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 \\ 0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 \\ 0.000 & 0.500 & 1.500 & 0.000 & -1.000 & 0.000 \\ -0.667 & 0.000 & 0.000 & 1.333 & 0.000 & 0.000 \\ 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\ -1.000 & -1.000 & 0.000 & 0.000 & 0.000 & 2.000 \end{bmatrix}$$

Adding $\mathbf{A}^{-1}1/\sigma_a^2$ to the $\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}$ of the LSE gives the MME, which are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 0.065 & 0.000 & 0.000 & 0.000 & 0.010 & 0.035 & 0.010 & 0.010 \\ 0.000 & 0.050 & 0.000 & 0.000 & 0.000 & 0.000 & 0.025 & 0.025 \\ 0.000 & 0.000 & 0.092 & 0.025 & 0.000 & -0.033 & 0.000 & -0.050 \\ 0.000 & 0.000 & 0.025 & 0.100 & 0.025 & 0.000 & -0.050 & -0.050 \\ 0.010 & 0.000 & 0.000 & 0.025 & 0.080 & 0.000 & -0.050 & 0.005 \\ 0.035 & 0.000 & -0.033 & 0.000 & 0.000 & 0.097 & 0.005 & 0.000 \\ 0.010 & 0.025 & 0.000 & -0.050 & -0.050 & 0.005 & 0.130 & 0.000 \\ 0.010 & 0.025 & -0.050 & -0.050 & 0.005 & 0.000 & 0.000 & 0.130 \end{bmatrix}^{-1} \begin{bmatrix} 0.282 \\ 0.170 \\ 0.000 \\ 0.000 \\ 0.050 \\ 0.148 \\ 0.107 \\ 0.148 \end{bmatrix}$$

The solutions are:

Sex effects		Animals					
Males	Females	1	2	3	4	5	6
4.358	3.404	0.098	–0.019	–0.041	–0.009	–0.186	0.177

The solutions for sex effects and proofs for parents are exactly as obtained using the animal model in Example 3.1. However, the number of non-zero elements in the coefficient matrix is 38 compared with 46 for an animal model in Section 3.3 on the same data set. This difference will be more marked in large data sets or in data sets where the number of progeny far exceeds the number of parents. This is one of the main advantages of the reduced animal model, as the number of equations and therefore non-zero elements to be stored are reduced. The solutions for non-parents can be obtained by back-solving, as discussed in the next section.

Solutions for non-parents

With the reduced animal model, solutions for non-parents are obtained by back-solving, using the solutions for the fixed effects and parents. Equation 3.9, derived earlier from the MME for an animal with its parents, can be used to back-solve for non-parent solutions. However, the $\mathbf{R}^{-1}$ has not been factored out of the MME in Eqn 3.25, and so the k term in Eqn 3.9 now equals:

$$k = r^{11}/r^{11} + d_i^{-1}g^{-1} \tag{3.26}$$

Downloaded From Vet-eBooks.com

Solutions for non-parents in Example 3.3 can be solved using Eqn 3.9 but with k expressed as in Eqn 3.26. However, because there is a fixed effect in the model, $m_i = k(y_c - b_j - 0.5a_s - 0.5a_d)$. In Example 3.3, both parents of non-parents (animals 7 and 8) are known, therefore:

$k = 0.025/(0.025 + (2)0.05) = 0.20$

Solution for calves 7 and 8 are:

$$\hat{a}_7 = 0.5(-0.009 + -0.186) + 0.20(3.5 - 4.358 - 0.5(-0.009 + -0.186)) = -0.249$$
$$\hat{a}_8 = 0.5(-0.041 + 0.177) + 0.20(5.0 - 4.358 - 0.5(-0.041 + 0.177)) = 0.183$$

Again, these solutions are the same for these animals as under the animal model.

3.5.3 An alternative approach

Note that, if the example data had been analysed using Eqn 3.25, the design matrices would be of the following form:

$$\mathbf{X}'_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{X}'_n = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$\mathbf{Z}$ including ancestors is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Z}_1 = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

The remaining matrices can be calculated through matrix multiplication. The MME then are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_4 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 2.600 & 0.000 & 0.000 & 0.000 & 0.400 & 1.400 & 0.400 & 0.400 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 \\ 0.400 & 0.000 & 0.000 & 1.000 & 3.200 & 0.000 & -2.000 & 0.200 \\ 1.400 & 0.000 & -1.333 & 0.000 & 0.000 & 3.867 & 0.200 & 0.000 \\ 0.400 & 1.000 & 0.000 & -2.000 & -2.000 & 0.200 & 5.200 & 0.000 \\ 0.400 & 1.000 & -2.000 & -2.000 & 0.200 & 0.000 & 0.000 & 5.200 \end{bmatrix}^{-1} \begin{bmatrix} 11.300 \\ 6.800 \\ 0.000 \\ 0.000 \\ 2.000 \\ 5.900 \\ 4.300 \\ 5.900 \end{bmatrix}$$

and these give the same solutions as obtained from Eqn 3.24.

3.6 Animal Model with Groups

In Example 3.1 there were animals in the pedigree with unknown parents, usually called base population animals. The use of the relationship matrix in animal model evaluation assumes that these animals were sampled from a single population with average breeding value of

Downloaded From Vet-eBooks.com

zero and common variance σ_a^2. The breeding values of animals in subsequent generations are usually expressed relative to those of the base animals. However, if it is known that base animals were actually from populations that differ in genetic means, for instance, sires from different countries, this must then be accounted for in the model. In the dairy cattle situation, due to differences in selection intensity, the genetic means for sires of bulls, sires of cows, dams of bulls and dams of cows may all be different. These various sub-population structures should be accounted for in the model to avoid bias in the prediction of breeding values. This can be achieved through a proper grouping of base animals using available information.

Westell and Van Vleck (1987) presented a procedure for grouping, which has generally been adopted. For instance, if sires have been imported from several countries over a period of time and their ancestors are unknown, these sires could be assigned to groups on the basis of the expected year of birth of the ancestors and the country of origin. The sires born within a similar time period in a particular foreign country are assumed to come from ancestors of similar genetic merit. Thus each sire with one or both parents unknown is initially assigned phantom parents. Phantom parents are assumed to have had only one progeny each. Within each of the foreign countries, the phantom parents are grouped by the year of birth of their progeny and any other factor, such as sex of progeny. In addition, for the dairy cattle situation, the four selection paths – sire of sires, sire of dams, dam of sires and dam of dams – are usually assumed to be of different genetic merit and this is accounted for in the grouping strategy.

With groups, the model (Thompson, 1979) is:

$$y_{ij} = h_j + a_i + \sum_{k=1}^{n} t_{ik} g_k + e_{ij} \tag{3.27}$$

where h_j = effect of the jth herd, a_i = random effect of animal i, g_k = fixed group effect containing the kth ancestor, t_{ik} = the additive genetic relationship between the kth and ith animals and the summation is over all n ancestors of animal i, and e_{ij} = random environmental effect. From the model, it can be seen that the contribution of the group to the observation is weighted by the proportion of genes the ancestors in the group passed on to the animal with a record.

In matrix notation, the model can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{ZQg} + \mathbf{Za} + \mathbf{e} \tag{3.28}$$

where:

$$\mathbf{Q} = \mathbf{TQ}^*$$

$\mathbf{Q}^*$ assigns unidentified ancestors to groups and $\mathbf{T}$, a lower triangular matrix, is obtained from $\mathbf{A} = \mathbf{TDT}'$ (see Section 2.3). With this model the breeding value of an animal k is:

$$a_{k^*} = Q\hat{g} + \hat{a}_k$$

The MME are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{ZQ} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha & \mathbf{Z}'\mathbf{ZQ} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{ZQ} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Solving the MME above will yield vectors of solutions for $\mathbf{a}$ and $\mathbf{g}$ but the ranking criterion (breeding value) is $\hat{\mathbf{a}}_{k^*} = \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{a}}_k$ for animal k. Modification of the MME

Downloaded From Vet-eBooks.com

(Quaas and Pollak, 1981) and absorption of the group equations gave the following set of equations, which are usually solved to obtain $\hat{\mathbf{a}}^*$ directly (Westell *et al.*, 1988).

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}_{nn}\alpha & \mathbf{A}^{-1}_{np}\alpha \\ \mathbf{0} & \mathbf{A}^{-1}_{pn}\alpha & \mathbf{A}^{-1}_{pp}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} + \hat{\mathbf{Q}}\mathbf{g} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix} \tag{3.29}$$

where n is the number of animals and p the number of groups.

Let

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{-1}_{nn} & \mathbf{A}^{-1}_{np} \\ \mathbf{A}^{-1}_{pn} & \mathbf{A}^{-1}_{pp} \end{bmatrix}$$

The matrix $\mathbf{A}^{-1}$ is obtained by the usual rules for obtaining the inverse of the relationship matrix outlined in Section 2.4.1. A list of pedigrees, consisting of only actual animals but with unknown ancestors assigned to groups, is set up. For the ith animal calculate the inverse (b_i) of the variance of Mendelian sampling as:

b_i = 4/(2 + number of parents of animal i assigned to groups)

Then add:

b_i to the (i,i) element of $\mathbf{A}^{-1}$
$-\frac{b_i}{2}$ to the (i,s), (i,d), (s,i) and (d,i) elements of $\mathbf{A}^{-1}$
$\frac{b_i}{4}$ to the (s,s), (s,d), (d,s) and (d,d) elements of $\mathbf{A}^{-1}$

Thus for an animal i with both parents assigned to groups:

b_i = 4/(2 + 2) = 1

Then add:

1 to the (i,i) element of $\mathbf{A}^{-1}$
$-\frac{1}{2}$ to the (i,s), (i,d), (s,i) and (d,i) elements of $\mathbf{A}^{-1}$
$\frac{1}{4}$ to the (s,s), (s,d), (d,s) and (d,d) elements of $\mathbf{A}^{-1}$

3.6.1 An illustration

Example 3.4

An animal model evaluation with groups is illustrated below using the same data set and genetic parameters as in Example 3.1. The aim is to estimate sex effects and predict breeding values for animals and phantom parents (groups). The model in Eqn 3.28 and the MME in Eqn 3.29 are used for the analysis. The pedigree file for the data set is:

Calf	Sire	Dam
1	Unknown	Unknown
2	Unknown	Unknown
3	Unknown	Unknown
4	1	Uunknown
5	3	2
6	1	2
7	4	5
8	3	6

Downloaded From Vet-eBooks.com

Assuming that males are of different genetic merit compared to females, the unknown sires can be assigned to one group (G1) and unknown dams to another group (G2). The pedigree file now becomes:

Calf	Sire	Dam
1	G1	G2
2	G1	G2
3	G1	G2
4	1	G2
5	3	2
6	1	2
7	4	5
8	3	6

Recoding G1 as 9 and G2 as 10:

Calf	Sire	Dam
1	9	10
2	9	10
3	9	10
4	1	10
5	3	2
6	1	2
7	4	5
8	3	6

Setting up the design matrices and MME

The design matrices **X** and **Z**, and the matrices **X′X**, **X′Z**, **Z′X**, **X′y** and **Z′y** in the MME are exactly as in Example 3.1. The MME without addition of the inverse of the relationship matrix for animals and groups are:

$$\begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \\ \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} = \begin{bmatrix} 13.0 \\ 6.8 \\ 0 \\ 0 \\ 0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \\ 0 \\ 0 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

Using the procedure outlined above, $\mathbf{A}^{-1}$ for the example data is:

	1	2	3	4	5	6	7	8	9	10
1	1.83	0.50	0.00	−0.67	0.00	−1.00	0.00	0.00	−0.50	−0.17
2	0.50	2.00	0.50	0.00	−1.00	−1.00	0.00	0.00	−0.50	−0.50
3	0.00	0.50	2.00	0.00	−1.00	0.50	0.00	−1.00	−0.50	−0.50
4	−0.67	0.00	0.00	1.83	0.50	0.00	−1.00	0.00	0.00	−0.67
5	0.00	−1.00	−1.00	0.50	2.50	0.00	−1.00	0.00	0.00	0.00
6	−1.00	−1.00	0.50	0.00	0.00	2.50	0.00	−1.00	0.00	0.00
7	0.00	0.00	0.00	−1.00	−1.00	0.00	2.00	0.00	0.00	0.00
8	0.00	0.00	−1.00	0.00	0.00	−1.00	0.00	2.00	0.00	0.00
9	−0.50	−0.50	−0.50	0.00	0.00	0.00	0.00	0.00	0.75	0.75
10	−0.17	−0.50	−0.50	−0.67	0.00	0.00	0.00	0.00	0.75	1.08

and $\mathbf{A}^{-1}\alpha$ is easily obtained by multiplying every element of $\mathbf{A}^{-1}$ by 2, the value of α. The matrix $\mathbf{A}^{-1}\alpha$ is added to equations for animal and group to obtain the MME, which are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \\ \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} = \begin{bmatrix} 3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 & -1.000 & -0.333 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 & -1.000 & -1.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 & -1.000 & -1.000 \\ 1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 & 0.000 & -1.333 \\ 0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & -1.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.500 & 1.500 \\ 0.000 & 0.000 & -0.333 & -1.000 & -1.000 & -1.333 & 0.000 & 0.000 & 0.000 & 0.000 & 1.500 & 2.167 \end{bmatrix}^{-1} \begin{bmatrix} 13.0 \\ 6.8 \\ 0.00 \\ 0.00 \\ 0.00 \\ 4.50 \\ 2.90 \\ 3.90 \\ 3.50 \\ 5.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

There is dependency in the equations, that is, all effects cannot be estimated; therefore, the equation for the first group has been set to zero to obtain the following solutions:

Sex effects	
Males	5.474
Females	4.327
Animals	
1	−0.780
2	−0.936
3	−0.977
4	−1.287
5	−1.113
6	−0.741
7	−1.354
8	−0.782
Groups	
9	0.000
10	−1.795

Downloaded From Vet-eBooks.com

The animal proofs above are generally lower than those from Example 3.1, the model without groups. In addition, the ranking for animals is also different. However, the relationship between the two sets of solutions can be shown by recalculating the vector of solutions for animals using the group solutions ($\hat{\mathbf{g}}$) above and the estimated breeding values ($\hat{\mathbf{a}}$) from Example 3.1 as:

$$\hat{\mathbf{a}}_* = \hat{\mathbf{a}} + \mathbf{Qg}$$

where $\mathbf{Q} = \mathbf{TQ}^*$, as defined earlier.

Assigning phantom parents (M1 to M7) to animals with unknown ancestors, the pedigree for the example data can be written as:

Calf	Sire	Dam
1	M1	M2
2	M3	M4
3	M5	M6
4	1	M7
5	3	2
6	1	2
7	4	5
8	3	6

and the matrix **T** for the pedigree is:

	M1	M2	M3	M4	M5	M6	M7	1	2	3	4	5	6	7	8
M1	1.000	0.000	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M2	0.000	1.000	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M3	0.000	0.000	1.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M4	0.000	0.000	0.000	1.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M5	0.000	0.000	0.000	0.000	1.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M6	0.000	0.000	0.000	0.000	0.000	1.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M7	0.000	0.000	0.000	0.000	0.000	0.000	1.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
1	0.500	0.500	0.000	0.000	0.000	0.000	0.00	1.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
2	0.000	0.000	0.500	0.500	0.000	0.000	0.00	0.00	1.00	0.00	0.0	0.0	0.0	0.0	0.0
3	0.000	0.000	0.000	0.000	0.500	0.500	0.00	0.00	0.00	1.00	0.0	0.0	0.0	0.0	0.0
4	0.250	0.250	0.000	0.000	0.000	0.000	0.50	0.50	0.00	0.00	1.0	0.0	0.0	0.0	0.0
5	0.000	0.000	0.250	0.250	0.250	0.250	0.00	0.00	0.50	0.50	0.0	1.0	0.0	0.0	0.0
6	0.250	0.250	0.250	0.250	0.000	0.000	0.00	0.50	0.50	0.00	0.0	0.0	1.0	0.0	0.0
7	0.125	0.125	0.125	0.125	0.125	0.125	0.25	0.25	0.25	0.25	0.5	0.5	0.0	1.0	0.0
8	0.125	0.125	0.125	0.125	0.250	0.250	0.00	0.25	0.25	0.50	0.0	0.0	0.5	0.0	1.0

The transpose of the matrix $\mathbf{Q}^*$, which assigns phantom parents to groups is:

$$\mathbf{Q}^{*\prime} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Downloaded From Vet-eBooks.com

and the transpose of **Q** ($\mathbf{Q} = \mathbf{T}\mathbf{Q}^*$) is:

$$\mathbf{Q}' = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.25 & 0.5 & 0.5 & 0.375 & 0.5 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0.5 & 0.5 & 0.5 & 0.75 & 0.5 & 0.5 & 0.625 & 0.5 \end{pmatrix}$$

Therefore the vector of solutions using the EBVs from Example 3.1 is:

$$\hat{\mathbf{a}}_* = \hat{\mathbf{a}} + \mathbf{Q}\hat{\mathbf{g}} = \begin{bmatrix} 0.098 \\ -0.019 \\ -0.041 \\ -0.009 \\ -0.186 \\ 0.177 \\ -0.249 \\ 0.183 \end{bmatrix} + \begin{bmatrix} -0.898 \\ -0.898 \\ -0.898 \\ -1.346 \\ -0.898 \\ -0.898 \\ -1.122 \\ -0.898 \end{bmatrix} = \begin{bmatrix} -0.800 \\ -0.917 \\ -0.939 \\ -1.355 \\ -1.084 \\ -0.721 \\ -1.371 \\ -0.715 \end{bmatrix}$$

These solutions are similar to those obtained in the model with groups. The slight differences are due to differences in sex solutions in the two examples and this is explained later. This indicates that, when the solutions from the model without groups are expressed relative to the group solutions, similar solutions are obtained to those in the model with groups. Thus the differences between the solutions in Examples 3.1 and 3.4 are due to the fact that the solutions in the former are expressed relative to base animals assumed to have an average breeding value of zero, while in the latter solutions are relative to the group solutions, one of which is lower than zero.

The inclusion of groups also resulted in a larger sex difference compared with Example 3.1. The solution for sex effect *i* can be calculated using Eqn 3.5. For instance, the solution for male calves in Example 3.4 is:

$$b_1 = [(4.5 + 3.5 + 5.0) - (-1.287 + -1.354 + -0.782)]/3 = 5.474$$

Since $\sum_j y_{ij}$ in Eqn 3.5 is the same in both examples, differences in $\sum_j \hat{a}_{ij}$ between the sexes in both models would result in differences in the linear function of **b**. The difference between average breeding values of male and females calves is –0.02 and –0.214, respectively, in Examples 3.1 and 3.4. The larger difference in the latter accounted for the higher sex difference in Example 3.4. Males had a lower breeding value in Example 3.4 due to the higher proportionate contribution of group two to their solutions (see the matrix **Q** above).

The basic principles involved in the application of BLUP for genetic evaluations and the main assumptions have been covered in this chapter, and its application to more complex models involves an extension of these principles. Equation 3.1 is a very general model and **a** could include random animal effects for several traits (multivariate model), random environmental effects, such as common environmental effects affecting animals that are reared together, maternal effects (maternal model), non-additive genetic effects, such as dominance and epistasis (non-additive models), and repeated data on individuals (random regression model). The extension of the principles discussed in this chapter under these various models constitutes the main subject area of the subsequent chapters in the text.

Downloaded From Vet-eBooks.com

4 Best Linear Unbiased Prediction of Breeding Value: Models with Random Environmental Effects

4.1 Introduction

In some circumstances, environmental factors constitute an important component of the covariance between individuals such as members of a family reared together (common environmental effects) or between the records of an individual (permanent environmental effects). Such environmental effects are usually accounted for in the model to ensure accurate prediction of breeding values. This chapter deals with models that account these two main types of environmental effects in genetic evaluations.

4.2 Repeatability Model

The repeatability model has been employed for the analysis of data when multiple measurements on the same trait are recorded on an individual, such as litter size in successive pregnancies or milk yield in successive lactations (Interbull, 2000). The details of the assumptions and the components of the phenotypic variance have been given in Section 1.3.2. Briefly, the phenotypic variance comprises the genetic (additive and non-additive) variance, permanent environmental variance and temporary environmental variance. For an animal, the model usually assumes a genetic correlation of unity between all pairs of records, equal variance for all records and equal environmental correlation between all pairs of records. In practice, some of these assumptions do not hold in the analysis of real data. A more appropriate way of handling repeated measurements over time is by fitting a random regression model or a covariance function, and this is discussed in Chapter 9. This section has therefore been included to help illustrate the evolution of the model for the analysis of repeated records over time. The phenotypic structure for three observations of an individual under this model could be written (Quaas, 1984) as:

$$\text{var}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \sigma^2_{t1} + \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g \\ \sigma^2_{pe} + \sigma^2_g & \sigma^2_{t2} + \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g \\ \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g & \sigma^2_{t3} + \sigma^2_{pe} + \sigma^2_g \end{bmatrix}$$

with: σ^2_{ti} = temporary environmental variance specific to record i; σ^2_{pe} = covariance due to permanent environmental effects (variances and covariances are equal); and σ^2_g = genetic covariance (variances and covariances are equal). The correlation

Downloaded From Vet-eBooks.com

between records of an individual is referred to as repeatability and is $(\sigma_g^2 + \sigma_{pe}^2)/\sigma_y^2$. Genetic evaluation under this model is concerned not only with predicting breeding values but also permanent environmental effects.

4.2.1 Defining the model

The repeatability model is usually of the form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e} \tag{4.1}$$

where **y** = vector of observations, **b** = vector of fixed effects, **a** = vector of random animal effects, **pe** = vector of random permanent environmental effects and non-additive genetic effects, and **e** = vector of random residual effect. **X**, **Z** and **W** are incidence matrices relating records to fixed animal and permanent environmental effects, respectively.

Note that the vector **a** only includes additive random animal effects; consequently, non-additive genetic effects are included in the **pe** term. It is assumed that the permanent environmental effects and residual effects are independently distributed with means of zero and variance σ_{pe}^2 and σ_e^2, respectively. Therefore:

$$\text{var}(\mathbf{pe}) = \mathbf{I}\sigma_{pe}^2$$
$$\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}$$
$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2$$

and:

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_a^2 + \mathbf{WI}\sigma_{pe}^2\mathbf{W}' + \mathbf{R}$$

The MME for the BLUE of estimable functions of **b** and for the BLUP of **a** and **pe** are:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1}1/\sigma_a^2 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{I}(1/\sigma_{pe}^2) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

However, the MME with $\mathbf{R}^{-1}$ factored out from the above equations give the following equations, which are easier to set up:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \mathbf{I}\alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \tag{4.2}$$

where $\alpha_1 = \sigma_e^2/\sigma_a^2$ and $\alpha_2 = \sigma_e^2/\sigma_{pe}^2$

4.2.2 An illustration

Example 4.1

For illustrative purposes, assume a single dairy herd with the following data structure for five cows:

Downloaded From Vet-eBooks.com

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season.

It is assumed that $\sigma^2_a = 20.0$, $\sigma^2_e = 28.0$ and $\sigma^2_{pe} = 12.0$, giving a phenotypic variance (σ^2_y) of 60. From the given parameters, $\alpha_1 = 1.40$, $\alpha_2 = 2.333$ and repeatability is $(\sigma^2_a + \sigma^2_{pe})/\sigma^2_y = (20 + 12)/60 = 0.53$. The aim is to estimate the effects of lactation number and predict breeding values for all animals and permanent environmental effects for cows with records. The above genetic parameters are proportional to estimates reported by Visscher (1991) for fat yield for Holstein Friesians in the UK for the first two lactations using a repeatability model. Later, in Section 5.4, this data set is reanalysed using a multivariate model assuming an unequal design with different herd–year–season (HYS) effects defined for each lactation using corresponding multivariate genetic parameter estimates of Visscher (1991).

Setting up the design matrices

The transpose of the matrix **X**, which relates records to HYS and parity is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

The first four rows of $\mathbf{X}'$ relate records to HYS effects and the last two rows to parity effects. Considering only animals with records, $\mathbf{Z}'$ and $\mathbf{W}'$ are equal and for the example data set:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Each row of $\mathbf{Z}'$ corresponds to each cow with records. The matrices $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{W}'\mathbf{W}$ are both diagonal and equal and $\mathbf{Z}'\mathbf{Z}$ is:

$\mathbf{Z}'\mathbf{Z} = \text{diag}(2, 2, 2, 2, 2)$

Downloaded From Vet-eBooks.com

Note, however, that it is necessary to augment **Z′Z** by three columns and rows of zeros to account for animals 1 to 3, which are ancestors. The remaining matrices in the MME apart from $\mathbf{A}^{-1}$ can easily be calculated through matrix multiplication. The inverse of the relationship matrix ($\mathbf{A}^{-1}$) is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 2.50 & 0.50 & 0.00 & -1.00 & 0.50 & -1.00 & 0.50 & -1.00 \\ 0.50 & 1.50 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.83 & 0.50 & -0.67 & 0.00 & -1.00 & 0.00 \\ -1.00 & -1.00 & 0.50 & 2.50 & 0.00 & 0.00 & -1.00 & 0.00 \\ 0.50 & 0.00 & -0.67 & 0.00 & 1.83 & -1.00 & 0.00 & 0.00 \\ -1.00 & 0.00 & 0.00 & 0.00 & -1.00 & 2.00 & 0.00 & 0.00 \\ 0.50 & 0.00 & -1.00 & -1.00 & 0.00 & 0.00 & 2.50 & -1.00 \\ -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 2.00 \end{bmatrix}$$

and $\mathbf{A}^{-1}\alpha_1$ is added to the **Z′Z** to obtain the MME.

The MME are too large to be shown. There is dependency in the MME because the sum of equations for HYS 1 and 2 equals that of parity 1 and the sum of HYS 3 and 4 equals that for parity 2. The equations for HYS 1 and 3 were set to zero to obtain the following solutions by direct inversion of the coefficient matrix:

Effects	Solutions
HYS	
1	0.000
2	44.065
3	0.000
4	0.013
Parity	
1	175.472
2	241.893
Animal	
1	10.148
2	–3.084
3	–7.063
4	13.581
5	–18.207
6	–18.387
7	9.328
8	24.194
Permanent environment	
4	8.417
5	–7.146
6	–17.229
7	–1.390
8	17.347

Downloaded From Vet-eBooks.com

The fixed-effect solutions for parity indicate that yield at second lactation is higher than that at first, which is consistent with the raw averages. From the MME, the solution for level *i* of the *n*th fixed effect can be calculated as:

$$\hat{b}_{in} = \sum_{f=1}^{\text{diag}_{in}} y_{inf} - \sum_{j} \hat{b}_{inj} - \sum_{k} \hat{a}_{ink} - \sum_{l} \hat{pe}_{inl} \,/\, \text{diag}_{in} \qquad (4.3)$$

where y_{inf} is the record for animal *f* in level *i* of the *n*th fixed effect, diag_{in} is the number of observations for level *i* of the *n*th fixed effect, b_{inj}, $\hat{a}_{ink}$ and pe_{inl} are solutions for levels *j*, *k* and *l* of any other fixed effect, random animal and permanent environmental effects, respectively, within level *i* of the *n*th fixed effect. Thus the solution for level two of HYS effect is:

$$\begin{aligned}\hat{b}_{21} &= [445 - (2\hat{b}_{12}) - (\hat{a}_6 + \hat{a}_8) - (\hat{pe}_6 + \hat{pe}_8)]/2 \\ &= [445 - 2(175.472) - 5.807 - (0.118)]/2 \\ &= 44.065\end{aligned}$$

Breeding values for animals with a repeatability model can also be calculated using Eqn 3.8, except that YD is now yield corrected for the appropriate fixed effects, permanent environmental effect and averaged. Thus for animal 4:

$$\begin{aligned}\hat{a}_4 = {} & n_1[(\hat{a}_1 + \hat{a}_2)/2] + n_2[((y_{41} - \hat{b}_1 - \hat{b}_5 - \hat{pe}_4) + (y_{42} - \hat{b}_3 - \hat{b}_6 - \hat{pe}_4))/2] \\ & + n_3(2\hat{a}_7 - \hat{a}_3)\end{aligned}$$

where y_{ji} is yield for cow *j* in lactation *i*, $n_1 = 2.8/5.5$, $n_2 = 2/5.5$ and $n_3 = 0.7/5.5$ and 5.5 = the sum of the numerator of n_1, n_2 and n_3.

$$\begin{aligned}\hat{a}_4 = {} & n_1(3.532) + n_2[((201 - 0.0 - 175.472 - 8.417) \\ & + (280 - 0.0 - 241.893 - 8.147))/2] + n_3(18.656 - (-7.063)) \\ = {} & 13.581\end{aligned}$$

The higher breeding value for sire 1 compared with sire 3 is due to the fact that on average the daughters of sire 1 were of higher genetic merit after adjusting for the breeding values of mates. The very high breeding value for cow 8 results from the high parent average breeding value and she has the highest yield in the herd, resulting in a large YD.

The estimate of *pe* for animal *i* could be calculated as:

$$\hat{pe}_i = \left[\left(\sum_{f}^{mi} Y_{if} - \sum_{j} \hat{b}_{ij} - \sum_{k} \hat{a}_{ik}\right)\right] \Big/ (m_i + \alpha_2) \qquad (4.4)$$

where m_i is the number of records for animal *i* $\alpha_2 = \sigma_e^2/\sigma_{pe}^2$ and other terms are as defined in Eqn 4.3. Thus for animal 4:

$$\begin{aligned}\hat{pe}_4 = {} & [(201 - 0.0 - 175.472 - 13.581) \\ & + (280 - 0.0 - 241.893 - 13.581)]/(2 + 2.333) \\ = {} & 8.417\end{aligned}$$

The estimate of permanent environment effect for an animal represents environmental influences and non-additive genetic effect, which are peculiar to the animal and affect its performance for life. These environmental influences could either be

Downloaded From Vet-Books.com

favourable – for instance, animal 8 has the highest estimates of **pe** and this is reflected by her high average yield – or could reduce performance (for example, cow 6 has a very negative estimate of **pe** and low average yield). A practical example of such permanent environment effect could be the loss of a teat by a cow early in life due to infection. Thus differences in estimates of **pe** represent permanent environmental differences between animals and could help the farmer, in addition to the breeding value, in selecting animals for future performance in the same herd. The sum of breeding value and permanent environment effect ($\hat{a}_i + pe_i$) for animal i is termed the probable producing ability (PPA) and represents an estimate of the future performance of the animal in the same herd. If the estimate of the management level (M) for animal i is known, its future record (y_i) can be predicted as:

$$y_i = M + \text{PPA}$$

This could be used as a culling guide.

4.2.3 Calculating daughter yield deviations

As indicated in Section 3.3.3, daughter yield deviation (DYD) is commonly calculated for sires in dairy cattle evaluations. The calculation of DYD for sire 1 in Example 4.1 is hereby illustrated.

First, the yield deviations for the daughters (cows 4, 6 and 8) of sire 1 are calculated. Thus for cow i, $YD_i = (\mathbf{Z'Z})^{-1}\mathbf{Z'}(\mathbf{y}_i - \mathbf{Xb} - \mathbf{Wpe})$. Therefore:

$$\begin{aligned}
YD_4 &= \tfrac{1}{2}[(201 - 175.472 - 0 - 8.417) + (280 - 241.893 - 0 - 8417)] = 23.4005\\
YD_6 &= \tfrac{1}{2}[(160 - 175.472 - 44.065 - (-17.229))\\
&\quad + (190 - 241.893 - 0 - (-17.229)] = -38.486\\
YD_8 &= \tfrac{1}{2}[(285 - 175.472 - 44.065 - 17.347)\\
&\quad + (300 - 241.893 - 0.013 - 17.347)] = 44.432
\end{aligned}$$

Both parents of these daughters are known, therefore $n_{2prog} = 2/(2 + 2\alpha_1) = 0.4167$ and $u_{prog} = 1$ for each daughter. Using Eqn 3.12, DYD for sire 1 is:

$$\begin{aligned}
\text{DYD}_1 &= [u_{(4)}n_{2(4)}(2YD_4 - \hat{a}_2) + u_{(5)}n_{2(6)}(2YD_6 - \hat{a}_5)\\
&\quad + u_{(8)}n_{2(8)}(2YD_8 - \hat{a}_7)] \Big/ \left(\sum_3 (u_{prog} + n_{2prog})\right)\\
\text{DYD}_1 &= [(1)(0.4167)(2(23.4005) - (-3.084)) + (1)0.4167(2(-38.486) - (-18.207))\\
&\quad + (1)0.4167(2(44.432) - 9.328)]/(3(1)(0.4167))\\
&= 23.552
\end{aligned}$$

Calculating the proof of sire 1 using Eqn 3.13 and a DYD of 23.552 gives a breeding value of 9.058. It is slightly lower than the breeding value of 10.148 from solving the MME, as the contribution of the granddaughter through cow 4 is not included.

4.3 Model with Common Environmental Effects

Apart from the resemblance between records of an individual due to permanent environmental conditions, discussed in Section 4.2, environmental circumstances can also

Downloaded From Vet-eBooks.com

contribute to the resemblance between relatives. When members of a family are reared together, such as litters of pigs, they share a common environment and this contributes to the similarity between members of the family. Thus there is an additional covariance between members of a family due to the common environment they share and this increases the variance between different families. The environmental variance may be partitioned therefore into the between-family or group component (σ_c^2), usually termed the common environment, which causes resemblance between members of a family, and the within-family or within-group variance (σ_e^2). Sources of common environmental variance between families may be due to factors such as nutrition and/or climatic conditions. All sorts of relatives are subject to an environmental source of resemblance, but most analyses concerned with this type of variation in animal breeding tend to account for the common environment effects associated with full-sibs or maternal half-sibs, especially in pig and chicken studies.

4.3.1 Defining the model

Genetic evaluation under this model is concerned with prediction of breeding values and common environmental effects and the phenotypic variance may be partitioned into:

1. Additive genetic effects resulting from additive genes from parents.
2. Common environmental effects affecting full-sibs or all offspring of the same dam. In the case of full-sibs, it may be confounded with dominance effects peculiar to offspring of the same parents. Further explanation is given later on the components of the common environmental effect.
3. Random environmental effects.

In matrix notation, the model, which is exactly the same as in Eqn 4.1, is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{e}$$

where all terms are as given in Eqn 4.1 except **c**, which is the vector of common environmental effects and **W** now relates records to common environmental effects.

It is assumed that common environmental and residual effects are independently distributed with means of zero and variance σ_c^2 and σ_e^2, respectively. Thus var(**c**) = $\mathbf{I}\sigma_c^2$, var(**e**) = $\mathbf{I}\sigma_e^2$ and var(**a**) = $\mathbf{A}\sigma_a^2$.

The MME for the BLUP of **a** and **c** and BLUE of estimable functions of **b** are exactly the same as Eqn 4.2 but with $\alpha_1 = \sigma_e^2/\sigma_a^2$ and $\alpha_2 = \sigma_e^2/\sigma_c^2$.

4.3.2 An illustration

Example 4.2

Consider the following data set on the weaning weight of piglets, which are progeny of three sows mated to two boars:

Downloaded From Vet-eBooks.com

Piglet	Sire	Dam	Sex	Weaning weight (kg)
6	1	2	Male	90
7	1	2	Female	70
8	1	2	Female	65
9	3	4	Female	98
10	3	4	Male	106
11	3	4	Female	60
12	3	4	Female	80
13	1	5	Male	100
14	1	5	Female	85
15	1	5	Male	68

The objective is to predict breeding values for all animals and common environmental effects for full-sibs. Given that $\sigma_a^2 = 20$, $\sigma_c^2 = 15$ and $\sigma_e^2 = 65$, then $\sigma_y^2 = 100$, $\alpha_1 = 3.25$ and $\alpha_2 = 4.333$.

The model for the analysis is that presented in Eqn 4.5 and, as mentioned earlier, the MME for the BLUP of **a** and **c** and BLUE of estimable functions of **b** are as given in Eqn 4.2, using α_1 and α_2 defined above.

Setting up the design matrices

The transpose of the matrix **X**, which relates records to sex effects in this example is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

and $\mathbf{Z} = \mathbf{I}$, excluding parents. The transpose of matrix **W** that relates records to full-sibs is:

$$\mathbf{W}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The MME can be set up as discussed in Example 4.1. The solutions to the MME by direct inversion of the coefficient matrix are:

Effects	Solutions
Sex	
Male	91.493
Female	75.764
Animals	
1	–1.441
2	–1.175
3	1.441
4	1.441
5	–0.266
6	–1.098

Continued

Downloaded From Vet-eBooks.com

(*Continued*)

Effects	Solutions
7	–1.667
8	–2.334
9	3.925
10	2.895
11	–1.141
12	1.525
13	0.448
14	0.545
15	–3.819
Common environment	
2	–1.762
4	2.161
5	–0.399

The equation for the solution of the i level of fixed, animal and common environmental effects under this model are the same as those given for fixed (Eqn 4.3), animal and permanent environmental effects (Eqn 4.4), respectively, in Example 4.1. The inclusion of common environmental effects in the model allows for accurate prediction of breeding values of animals. Assuming each dam reared her progeny and full-sib families were kept under similar environmental conditions, the estimates of common environmental effects indicate that dam 4 provided the best environment for her progeny compared with dams 2 and 5. Also, dam 4 has the highest breeding value among the dams and would therefore be the first dam of choice, whether selection is for dams of the next generation on the basis of breeding value only or selection is for future performance of the dams in the same herd, which will be based on some combination of breeding value and estimate of common environmental effect.

The environmental covariance among full-sibs or maternal half-sibs might be due to influences from the dam (mothering ability or maternal effect); therefore, differences in mothering ability among dams would cause environmental variance between families. For instance, resemblance among progeny of the same dam in body weight could be due to the fact they share the same milk supply and variation in milk yield among dams would result in differences between families in body weight. This variation in mothering ability of dams has a genetic basis and, to some degree, is due to genetic variation in some character of the dams. In Chapter 7, the genetic component of maternal effect is examined and the appropriate model that accounts for the genetic component in genetic evaluation is presented.

Downloaded From Vet-eBooks.com

5 Best Linear Unbiased Prediction of Breeding Value: Multivariate Animal Models

5.1 Introduction

Selection of livestock is usually based on a combination of several traits of economic importance that may be phenotypically and genetically related. Such traits may be combined into an index on which animals are ranked. A multiple trait evaluation is the optimum methodology to evaluate animals on these traits because it accounts for the relationship between them. A multiple trait analysis involves the simultaneous evaluation of animals for two or more traits and makes use of the phenotypic and genetic correlations between the traits. The first application of best linear unbiased prediction (BLUP) for multiple trait evaluation was by Henderson and Quaas (1976).

One of the main advantages of multivariate best linear unbiased prediction (MBLUP) is that it increases the accuracy of evaluations. The gain in accuracy is dependent on the absolute difference between the genetic and residual correlations between the traits. The larger the differences in these correlations, the greater the gain in accuracy of evaluations (Schaeffer, 1984; Thompson and Meyer, 1986). When, for instance, the heritability, genetic and environmental correlations for two traits are equal, multivariate predictions are equivalent essentially to those from univariate analysis for each trait. Moreover, traits with lower heritabilities benefit more when analysed with traits with higher heritabilities in a multivariate analysis. Also, there is an additional increase in accuracy with multivariate analysis resulting from better connections in the data due to residual covariance between traits (Thompson and Meyer, 1986).

In some cases, one trait is used to decide whether animals should remain in the herd and be recorded for other traits. For instance, only calves with good weaning weight may be allowed the chance to be measured for yearling weight. A single trait analysis of yearling weight will be biased since it does not include information on weaning weight on which the selection was based. This is often called culling bias. However, a multi-trait analysis on weaning and yearling weight can eliminate this bias. Thus MBLUP accounts for culling selection bias.

One of the disadvantages of a multiple trait analysis is the high computing cost. The cost of multiple analysis of n traits is more than the cost of n single analyses. Second, a multiple trait analysis requires reliable estimates of genetic and phenotypic correlations among traits and these may not be readily available.

In this chapter, MBLUP involving traits affected by the same effects (equal design matrices) and situations in which different traits are affected by different factors (non-identical design matrices) are discussed. In the next chapter, approximations of MBLUP when design matrices are equal with or without missing records are also examined.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

5.2 Equal Design Matrices and No Missing Records

Equal design matrices for all traits imply that all effects in the model affect all traits in the multivariate analysis and there are no missing records for any trait.

5.2.1 Defining the model

The model for a multivariate analysis resembles a stack of the univariate models for each of the traits. For instance, consider a multivariate analysis for two traits, with the model for each trait of the form given in Eqn 3.1, that is, for trait 1:

$$\mathbf{y}_1 = \mathbf{X}_1\mathbf{b}_1 + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{e}_1$$

and for trait 2:

$$\mathbf{y}_2 = \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_2\mathbf{a}_2 + \mathbf{e}_2$$

If animals are ordered within traits, the model for the multivariate analysis for the two traits could be written as:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & 0\\ 0 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & 0\\ 0 & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} \tag{5.1}$$

where $\mathbf{y}_i$ = vector of observations for the ith trait, $\mathbf{b}_i$ = vector of fixed effects for the ith trait, $\mathbf{a}_i$ = vector of random animal effects for the ith trait, $\mathbf{e}_i$ = vector of random residual effects for the ith trait, and $\mathbf{X}_i$ and $\mathbf{Z}_i$ are incidence matrices relating records of the ith trait to fixed and random animal effects, respectively.

It is assumed that:

$$\text{var}\begin{bmatrix}a_1\\ a_2\\ e_1\\ e_2\end{bmatrix} = \begin{bmatrix}g_{11}\mathbf{A} & g_{12}\mathbf{A} & 0 & 0\\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & 0 & 0\\ 0 & 0 & r_{11} & r_{12}\\ 0 & 0 & r_{21} & r_{22}\end{bmatrix}$$

where $\mathbf{G}$ = additive genetic variance and covariance matrix for animal effect with each element defined as: g_{11} = additive genetic variance for direct effects for trait 1; $g_{12} = g_{21}$ = additive genetic covariance between both traits; g_{22} = additive genetic variance for direct effects for trait 2; $\mathbf{A}$ is the relationship matrix among animals; and $\mathbf{R}$ = variance and covariance matrix for residual effects.

The MME are of the same form as in Section 3.2 and these are:

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}'\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1}\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{a}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix} \tag{5.2}$$

where:

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & 0\\ 0 & \mathbf{X}_2\end{bmatrix},\quad \mathbf{Z} = \begin{bmatrix}\mathbf{Z}_1 & 0\\ 0 & \mathbf{Z}_2\end{bmatrix},\quad \hat{\mathbf{b}} = \begin{bmatrix}\hat{\mathbf{b}}_1\\ \hat{\mathbf{b}}_2\end{bmatrix},\quad \hat{\mathbf{a}} = \begin{bmatrix}\hat{\mathbf{a}}_1\\ \hat{\mathbf{a}}_2\end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}$$

Downloaded From Vet-eBooks.com

Writing out the equations for each trait in the model separately, the MME become:

$$\begin{bmatrix}\hat{\mathbf{b}}_1\\ \hat{\mathbf{b}}_2\\ \hat{\mathbf{a}}_1\\ \hat{\mathbf{a}}_2\end{bmatrix}=\begin{bmatrix}\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2\\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2\\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1+\mathbf{A}^{-1}g^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2+\mathbf{A}^{-1}g^{12}\\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1+\mathbf{A}^{-1}g^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2+\mathbf{A}^{-1}g^{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1'\mathbf{R}^{11}\mathbf{y}_1+\mathbf{X}_1'\mathbf{R}^{12}\mathbf{y}_2\\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{y}_1+\mathbf{X}_2'\mathbf{R}^{22}\mathbf{y}_2\\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{y}_1+\mathbf{Z}_1'\mathbf{R}_2^{12}\\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{y}_1+\mathbf{Z}_2'\mathbf{R}_2^{22}\end{bmatrix} \tag{5.3}$$

where g^{ij} are r^{ij} elements of $\mathbf{G}^{-1}$ and $\mathbf{R}^{-1}$, respectively. It should be noted that if $\mathbf{R}^{12}$, $\mathbf{R}^{21}$, g^{12} and g^{21} were set to zero, the matrices in the equations above reduce to the usual ones computed when carrying out two single trait analyses since the two traits become uncorrelated and there is no flow of information from one trait to the other.

5.2.2 An illustration

Example 5.1

Assume the data in Table 5.1 to be the pre-weaning gain (WWG) and post-weaning gain (PWG) for five beef calves. The objective is to estimate sex effects for both traits and to predict breeding values for all animals using a MBLUP analysis. Assume that the additive genetic covariance (**G**) matrix is:

$$\mathbf{G}=\begin{matrix}\\ \text{WWG}\\ \text{PWG}\end{matrix}\begin{matrix}\begin{matrix}\text{WWG} & \text{PWG}\end{matrix}\\ \begin{bmatrix}20 & 18\\ 18 & 40\end{bmatrix}\end{matrix}\text{ and the residual covariance matrix }\mathbf{R}=\begin{matrix}\\ \text{WWG}\\ \text{PWG}\end{matrix}\begin{matrix}\begin{matrix}\text{WWG} & \text{PWG}\end{matrix}\\ \begin{bmatrix}40 & 11\\ 11 & 30\end{bmatrix}\end{matrix}$$

The inverses of **G** and **R** are:

$$\mathbf{G}^{-1}=\begin{bmatrix}0.084 & -0.038\\ -0.038 & 0.042\end{bmatrix}\quad\text{and}\quad\mathbf{R}^{-1}=\begin{bmatrix}0.028 & -0.010\\ -0.010 & 0.037\end{bmatrix}$$

Setting up the design matrices

The matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ relate records for WWG and PWG, respectively, to sex effects. Both matrices are exactly the same as **X** in Section 3.3.1. Considering only animals with records, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ relate records for WWG and PWG to animals, respectively. Both matrices are identity matrices since animals have only one record each for WWG and PWG. The matrix **y** is a vector of observations for WWG ($\mathbf{y}_1$) and PWG ($\mathbf{y}_2$). Thus its transpose is:

$$\mathbf{y}'=\begin{bmatrix}\mathbf{y}_1' & \mathbf{y}_2'\end{bmatrix}=\begin{bmatrix}4.5 & 2.9 & 3.9 & 3.5 & 5.0 & 6.8 & 5.0 & 6.8 & 6.0 & 7.5\end{bmatrix}$$

The other matrices in the MME can then easily be calculated from the design matrices and vector of observations through matrix multiplication. Examples of some blocks of equations are given below.

Downloaded From Vet-eBooks.com

Table 5.1. Pre-weaning gain (kg) and post-weaning gain (kg) for five beef calves.

Calves	Sex	Sire	Dam	WWG	PWG
4	Male	1	–	4.5	6.8
5	Female	3	2	2.9	5.0
6	Female	1	2	3.9	6.8
7	Male	4	5	3.5	6.0
8	Male	3	6	5.0	7.5

From Eqns 5.2 and 5.3, the fixed effects by fixed effects block of equations for both traits in the coefficient matrix of the MME is:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} 0.084 & 0.000 & -0.03 & 0.00 \\ 0.00 & 0.056 & 0.00 & -0.02 \\ -0.03 & 0.00 & 0.101 & 0.00 \\ 0.00 & -0.02 & 0.00 & 0.074 \end{bmatrix}$$

The right-hand side for the levels of sex effects for both traits is:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{22}\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 0.364 + (-0.203) \\ 0.190 + (-0.118) \\ -0.130 + 0.751 \\ -0.068 + 0.437 \end{bmatrix}$$

The inverse of the relationship matrix for the example data is the same as that given in Example 3.1. The matrices $\mathbf{A}^{-1}g^{11}$, $\mathbf{A}^{-1}g^{12}$ and $\mathbf{A}^{-1}g^{22}$ are added to $\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1$, $\mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2$ and $\mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2$, respectively, to obtain the MME. For example, the matrix $\mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{12}$ is:

$$\mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{12} = \begin{bmatrix} -0.069 & -0.019 & 0.000 & 0.025 & 0.000 & 0.038 & 0.000 & 0.000 \\ -0.019 & -0.076 & -0.019 & 0.000 & 0.038 & 0.038 & 0.000 & 0.000 \\ 0.000 & -0.019 & -0.076 & 0.000 & 0.038 & -0.019 & 0.000 & 0.038 \\ 0.025 & 0.000 & 0.000 & -0.080 & -0.019 & 0.000 & 0.038 & 0.000 \\ 0.000 & 0.038 & 0.038 & -0.019 & -0.105 & 0.000 & 0.038 & 0.000 \\ 0.038 & 0.038 & -0.019 & 0.000 & 0.000 & -0.105 & 0.000 & 0.038 \\ 0.000 & 0.000 & 0.000 & 0.038 & 0.038 & 0.000 & -0.086 & 0.000 \\ 0.000 & 0.000 & 0.038 & 0.000 & 0.000 & 0.038 & 0.000 & -0.086 \end{bmatrix}$$

The MME have not been presented because they are too large, but solving the MME by direct inversion of the coefficient matrix gives the solutions shown below. See also the solutions from a univariate analysis of each trait.

Downloaded From Vet-eBooks.com

	Multivariate analysis traits		Univariate analysis traits	
Effects	WWG	PWG	WWG	PWG
Sex				
Male	4.361	6.800	4.358	6.798
Female	3.397	5.880	3.404	5.879
Animals				
1	0.151	0.280	0.098	0.277
2	–0.015	–0.008	–0.019	–0.005
3	–0.078	–0.170	–0.041	–0.171
4	–0.010	–0.013	–0.009	–0.013
5	–0.270	–0.478	–0.186	–0.471
6	0.276	0.517	0.177	0.514
7	–0.316	–0.479	–0.249	–0.464
8	0.244	0.392	0.183	0.384

The differences between the solutions for males and females for WWG and PWG in the multivariate analysis are more or less the same as those obtained in the univariate analyses of both traits. The solutions for fixed effects in the multivariate analysis from the MME can be calculated as:

$$\begin{bmatrix}\hat{b}_{1j}\\ \hat{b}_{2j}\end{bmatrix}=\begin{bmatrix}n_j r^{11} & n_j r^{12}\\ n_j r^{21} & n_j r^{22}\end{bmatrix}^{-1}\left[\mathbf{R}^{-1}\begin{bmatrix}y_{1j}-\hat{a}_{1j}-g^{12}\hat{a}_{2j}\\ y_{2j}-g^{21}\hat{a}_{1j}-\hat{a}_{2j}\end{bmatrix}\right] \tag{5.4}$$

where y_{ij} and $\hat{a}_{ij}$ are the sums of observations and EBVs, respectively, for calves for trait i in sex subclass j, $\hat{b}_{ij}$ is the solution for trait i in sex subclass j and n_j is the number of observations for sex subclass j. Using the above equation, the solutions for sex effects for males for WWG and PWG are:

$$\begin{bmatrix}\hat{b}_{11}\\ \hat{b}_{21}\end{bmatrix}=\begin{bmatrix}3r^{11} & 3r^{12}\\ 3r^{21} & 3r^{22}\end{bmatrix}^{-1}\left[\begin{bmatrix}r^{11} & r^{12}\\ r^{21} & r^{22}\end{bmatrix}\begin{bmatrix}13.0-(0.082)-g^{12}(-0.10)\\ 20.3-g^{21}(-0.082)-(-0.10)\end{bmatrix}\right]=\begin{bmatrix}4.361\\ 6.800\end{bmatrix}$$

5.2.3 Partitioning animal evaluations from multivariate analysis

An equation similar to Eqn 3.8 for the partitioning of evaluations from multivariate model was presented by Mrode and Swanson (2004) in the context of a random regression model (see Chapter 9). Since the yield records of animals contribute to the breeding values through the vector of yield deviations (**YD**), equations for calculating **YD** are initially presented. From Eqn 5.1, the equations for the breeding values of animals are:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1}\mathbf{G}^{-1})\hat{\mathbf{a}}=\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})$$

Therefore:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1}\mathbf{G}^{-1})\hat{\mathbf{a}}=(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\,\mathbf{YD} \tag{5.5}$$

with:

$$\mathbf{YD}=(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})^{-1}\,(\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})) \tag{5.6}$$

Downloaded From Vet-eBooks.com

Just as in the univariate model, **YD** is a vector of the weighted average of a cow's yield records corrected for all fixed effects in the model.

Transferring the left non-diagonal terms of $\mathbf{A}^{-1}$ in Eqn 5.5 to the right side of the equation (VanRaden and Wiggans, 1991) gives:

$$(\mathbf{Z'R^{-1}Z} + \mathbf{G}^{-1}\alpha_{anim})\hat{\mathbf{a}}_{anim} = \mathbf{G}^{-1}\alpha_{par}(\hat{\mathbf{a}}_{sire} + \hat{\mathbf{a}}_{dam}) + (\mathbf{Z'R}^{-1}\mathbf{Z})\mathbf{YD} + \mathbf{G}^{-1}\sum \alpha_{prog}(\mathbf{a}_{prog} - 0.5\hat{\mathbf{a}}_{mate})$$

where α_{par} = 1, $\frac{2}{3}$ or $\frac{1}{2}$ if both, one or neither parents are known, respectively, and α_{prog} = 1 if the animal's mate is known and $\frac{2}{3}$ if unknown. Note that $\alpha_{anim} = 2\alpha_{par} + 0.5\alpha_{prog}$.

The above equation can be expressed as:

$$(\mathbf{Z'}\,\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})\hat{\mathbf{a}}_{anim} = 2\mathbf{G}^{-1}\alpha_{par}\,(\mathbf{PA}) + (\mathbf{Z'R}^{-1}\mathbf{Z})\mathbf{YD} + 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - \hat{\mathbf{a}}_{mate}) \quad (5.7)$$

where **PA** = parent average.

Pre-multiplying both sides of the equation by $(\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})^{-1}$ gives:

$$\hat{\mathbf{a}}_{anim} = \mathbf{W}_1\mathbf{PA} + \mathbf{W}_2\,\mathbf{YD} + \mathbf{W}_3\mathbf{PC} \quad (5.8)$$

with:

$$\mathbf{PC} = \sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - \hat{\mathbf{a}}_{mate}) \Big/ \sum \alpha_{prog}$$

The weights $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ = **I**, with $\mathbf{W}_1$ = $(\mathbf{DIAG})^{-1}2\mathbf{G}^{-1}\alpha_{par}$, $\mathbf{W}_2$ = $(\mathbf{DIAG})^{-1}(\mathbf{Z'R}^{-1}\mathbf{Z})$ and $\mathbf{W}_3$ = $(\mathbf{DIAG})^{-1}0.5\mathbf{G}^{-1}\Sigma\alpha_{prog}$, where (**DIAG**) = $(\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})$. Equation 5.8 is similar to Eqn 3.8 but the weights are matrices of the order of traits in the multivariate analysis. Equation 5.8 is illustrated below using calf 8 in Example 5.1.

Since **Z** = **I** for calf 8, then Eqn 5.6 becomes $\mathbf{YD} = \mathbf{RR}^{-1}(\mathbf{y} - \mathbf{Xb}) = \mathbf{y} - \mathbf{Xb}$. Thus:

$$\begin{pmatrix} YD_{81} \\ YD_{82} \end{pmatrix} = \begin{pmatrix} y_{81} - \hat{b}_1 \\ y_{82} - \hat{b}_2 \end{pmatrix} = \begin{pmatrix} 5.0 - 4.361 \\ 7.5 - 6.800 \end{pmatrix} = \begin{pmatrix} 0.639 \\ 0.700 \end{pmatrix}$$

Both parents of calf 8 are known, therefore:

$$\mathbf{DIAG}_8 = \mathbf{R}^{-1} + 2\mathbf{G}^{-1} = \begin{pmatrix} 0.1958 & -0.0858 \\ -0.0858 & 0.1211 \end{pmatrix}$$

and:

$$\mathbf{W}_1 = (\mathbf{DIAG})^{-1}2\mathbf{G}^{-1} = \begin{pmatrix} 0.8476 & -0.1191 \\ -0.0237 & 0.6092 \end{pmatrix} \text{ and}$$

$$\mathbf{W}_2 = \mathbf{I} - \mathbf{W}_1 = \begin{pmatrix} 0.1524 & 0.1191 \\ 0.0237 & 0.3908 \end{pmatrix}$$

Then, from Eqn 5.8:

$$\begin{pmatrix} \hat{a}_{81} \\ \hat{a}_{82} \end{pmatrix} = \mathbf{W}_1\begin{pmatrix} PA_{81} \\ PA_{82} \end{pmatrix} + \mathbf{W}_2\begin{pmatrix} YD_{81} \\ YD_{82} \end{pmatrix} = \mathbf{W}_1\begin{pmatrix} 0.099 \\ 0.1735 \end{pmatrix} + \mathbf{W}_2\begin{pmatrix} 0.639 \\ 0.700 \end{pmatrix}$$

$$= \begin{pmatrix} 0.06325 \\ 0.10335 \end{pmatrix} + \begin{pmatrix} 0.18075 \\ 0.28870 \end{pmatrix} = \begin{pmatrix} 0.244 \\ 0.392 \end{pmatrix}$$

Downloaded From Vet-eBooks.com

In both traits, the contributions from **PA** accounted for about 26% of the breeding value of the calf.

In general, the estimates of breeding value for PWG from the multivariate analysis above are similar to those from the univariate analysis. The maximum difference between the multivariate and univariate breeding values is 0.008 kg (calf 8). The similarity of the evaluations for PWG from both models is due to the fact that genetic regression of WWG on PWG (0.45) is almost equal to the phenotypic regression (0.41) (Thompson and Meyer, 1986). However, the breeding values for WWG from the multivariate analysis are higher than those from the univariate analysis, with a maximum difference of 0.10 kg (calf 8) in favour of the multivariate analysis. Thus much of the gain from the multivariate analysis is in WWG and this is due to its lower heritability, as mentioned earlier. However, there was only a slight re-ranking of animals for both traits in the multivariate analysis.

5.2.4 Accuracy of multivariate evaluations

One of the main advantages of MBLUP is the increase in the accuracy of evaluations. Presented below are estimates of reliabilities for the proofs for WWG and PWG from the multivariate analysis and the univariate analysis of each trait.

	Multivariate analysis					
	Diagonals[a]		Reliability		Univariate analysis reliability	
Animal	WWG	PWG	WWG	PWG	WWG	PWG
1	18.606	35.904	0.070	0.102	0.058	0.102
2	19.596	38.768	0.020	0.031	0.016	0.031
3	17.893	33.799	0.105	0.155	0.088	0.155
4	16.506	29.727	0.175	0.257	0.144	0.256
5	16.541	29.865	0.173	0.253	0.144	0.253
6	17.152	31.504	0.142	0.212	0.116	0.212
7	17.115	31.364	0.144	0.216	0.116	0.216
8	16.285	29.160	0.186	0.271	0.156	0.270

[a]Diagonal elements of the inverse of the coefficient matrix from multivariate analysis.

The reliability for the proof of animal i and trait j (r^2_{ij}) in the multivariate analysis was calculated as $r^2_{ij} = (g_{jj} - PEV_{ij})/g_{jj}$, where PEV_{ij} is the diagonal element of the coefficient matrix pertaining to animal i and trait j. This formula is obtained by rearranging the equation given for reliability in Section 3.3.3. For instance, the reliabilities for the proofs for WWG and PWG for animal 1, respectively, are:

$$r^2_{11} = (20 - 18.606)/20 = 0.070$$

and:

$$r^2_{21} = (40 - 35.904)/40 = 0.102$$

Similar to the estimates of breeding values, the reliabilities for animals for PWG from the multivariate analysis were essentially the same from the univariate analysis as $G_{ij} = r_p G_{jj}$ (Thompson and Meyer, 1986), where the jth trait is PWG and r_p is the phenotypic

Downloaded From Vet-eBooks.com

correlation. However, there was an increase of about 20% in reliability for WWG for each animal under the multivariate analysis compared with the univariate analysis. Again much of the gain in accuracy from the multivariate analysis is observed in WWG.

5.2.5 Calculating daughter yield deviations in multivariate models

The equations for calculating daughter yield deviations (DYDs) with a multivariate model are similar to Eqn 3.12 for the univariate model except that the weights are matrices of order equal to the order of traits. The equations can briefly be derived (Mrode and Swanson, 2004) as follows.

Given the daughter (prog) of a bull, with no progeny of her own, Eqn 5.8 becomes:

$$\hat{\mathbf{a}}_{prog} = \mathbf{W}_{1\mathbf{prog}}\mathbf{PA} + \mathbf{W}_{2\mathbf{prog}}(\mathbf{YD}) \tag{5.9}$$

Let **PC** be expressed as in Eqn 5.7:

$$\mathbf{PC} = 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - \hat{\mathbf{a}}_{mate}) \tag{5.10}$$

Substituting Eqn 5.9 into Eqn 5.10 gives:

$$\mathbf{PC} = 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(\mathbf{W}_{1_{\text{prog}}}\hat{\mathbf{a}}_{anim} + \mathbf{W}_{1_{\text{prog}}}\hat{\mathbf{a}}_{mate} + \mathbf{W}_{2_{\text{prog}}}2\mathbf{YD} - \hat{\mathbf{a}}_{mate})$$

Since the daughter has no offspring of her own, $\mathbf{W}_3 = 0$, therefore $\mathbf{W}_{1\mathbf{prog}} = \mathbf{I} - \mathbf{W}_{2\mathbf{prog}}$. Then:

$$\mathbf{PC} = 0.5\mathbf{G}^{-1}\sum \alpha_{prog}((\mathbf{I} - \mathbf{W}_{2_{\text{prog}}})\hat{\mathbf{a}}_{anim} + \mathbf{W}_{2\mathbf{prog}}(2\mathbf{YD} - \hat{\mathbf{a}}_{mate})) \tag{5.11}$$

Substituting Eqn 5.11 into Eq 5.7 and moving all terms involving $\hat{\mathbf{a}}_{anim}$ to the left-hand side gives:

$$\begin{aligned}&(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + 2\mathbf{G}^{-1}\alpha_{par} + 0.5\mathbf{G}^{-1}\sum \mathbf{W}_{2_{\text{prog}}}\alpha_{prog})\hat{\mathbf{a}}_{anim}\\ &\quad = 2\mathbf{G}^{-1}\alpha_{par}\mathbf{PA} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{YD} + 0.5\mathbf{G}^{-1}\sum \mathbf{W}_{2_{\text{prog}}}\alpha_{prog}(2\mathbf{YD} - \hat{\mathbf{a}}_{mate})\end{aligned}$$

Pre-multiplying both sides of the equation by the inverse coefficient matrix gives:

$$\hat{\mathbf{a}}_{anim} = \mathbf{M}_1(\mathbf{PA}) + \mathbf{M}_2(\mathbf{YD}) + \mathbf{M}_3(\mathbf{DYD}) \tag{5.12}$$

where:

$$\mathbf{DYD} = \sum \mathbf{W}_{2_{\text{prog}}}\alpha_{prog}(2\mathbf{YD} - \hat{\mathbf{u}}_{mate}) \Big/ \sum \mathbf{W}_{2_{\text{prog}}}\alpha_{prog} \tag{5.13}$$

and $\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 = \mathbf{I}$, with $\mathbf{M}_1 = (\mathbf{DIAG})^{-1}2\mathbf{G}^{-1}\alpha_{par}$, $\mathbf{M}_2 = (\mathbf{DIAG})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})$ and $\mathbf{M}_3 = (\mathbf{DIAG})^{-1}0.5\mathbf{G}^{-1}\Sigma\mathbf{W}_{2\mathbf{prog}}\alpha_{prog}$ where $(\mathbf{DIAG}) = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + 2\mathbf{G}^{-1}\alpha_{par} + 0.5\mathbf{G}^{-1}\Sigma\mathbf{W}_{2\mathbf{prog}}\alpha_{prog})$. The matrix $\mathbf{W}_{2\mathbf{prog}}$ in the equation for DYD is not symmetrical and is of the order of traits and the full matrix has to be stored. This could make the computation of DYD cumbersome, especially with a large multivariate analysis or when a random regression model is implemented (see Chapter 9). For instance, in the Canadian test day model, which involves analysing milk, fat and protein yields and somatic cell count (SCC) in the first three lactations, it is a matrix of order 36 (Jamrozik *et al.*, 1997).

Downloaded From Vet-Books.com

Thus for computational ease, pre-multiply $\mathbf{W}_{2\text{prog}}$ with $\mathbf{G}^{-1}$, and the equation for **DYD** becomes:

$$\mathbf{DYD} = \sum \mathbf{G}^{-1}\mathbf{W}_{2\text{prog}}\alpha_{prog}(2\mathbf{YD} - \hat{\mathbf{u}}_{mate}) \Big/ \mathbf{G}^{-1}\mathbf{W}_{2\text{prog}}\alpha_{prog}$$

The product of $\mathbf{G}^{-1}\mathbf{W}_{2\text{prog}}$ is symmetric and only upper or lower triangular elements need to be stored. The computation of DYD is illustrated in Section 5.4.2, using the example dairy data.

5.3 Equal Design Matrices with Missing Records

When all traits in a multivariate analysis are not observed in all animals, the same methodology described in Section 5.2 can also be employed to evaluate animals, except that different residual covariance matrices have to be set up corresponding to a different combination of traits present. If the loss of traits is sequential, that is, the presence of the *i*th record implies the presence of 1 to $(i - 1)$ records, then the number of residual covariance matrices is equal to the number of traits. In general, if there are n traits, there are $(2^n - 1)$ possible combinations of observed traits and therefore residual covariance matrices (Quaas, 1984).

5.3.1 An illustration

Example 5.2
For illustrative purposes, consider the data set below, obtained by modifying the data in Table 5.1.

Calf	Sex	Sire	Dam	WWG (kg)	PWG (kg)
4	Male	1	–	4.5	–
5	Female	3	2	2.9	5.0
6	Female	1	2	3.9	6.8
7	Male	4	5	3.5	6.0
8	Male	3	6	5.0	7.5
9	Female	7	–	4.0	–

The model for the analysis is the same as in Section 5.2.1 and the same genetic parameters applied in Example 5.1 are assumed. The loss of records is sequential; there are therefore two residual covariance matrices. For animals with missing records for PWG, the residual covariance matrix ($\mathbf{R}_m$) and its inverse ($\mathbf{R}_m^{-1}$) are $\mathbf{R}_m = r_{m11} = 40$ and $\mathbf{R}_m^{-1} = r_m^{11} = \frac{1}{40} = 0.025$. For animals with records for both WWG and PWG, the residual covariance matrix ($\mathbf{R}_o$) and its inverse ($\mathbf{R}_o^{-1}$) are:

$$\mathbf{R}_o = \begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix} \quad \text{and} \quad \mathbf{R}_o^{-1} = \begin{bmatrix} 0.028 & -0.010 \\ -0.010 & 0.037 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

The $\mathbf{X}_1'$ and $\mathbf{X}_2'$ matrices, which relate sex effects for WWG and PWG, respectively, are:

$$\mathbf{X}_1' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2' = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{X}_1'\mathbf{X}_1 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2'\mathbf{X}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

In setting up $\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1$, it is necessary to account for the fact that animals (one male and one female) have missing records for PWD. Thus:

$$\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 = r_m^{11}\mathbf{W}'\mathbf{W} + r_o^{11}\mathbf{B}'\mathbf{B} = 0.025\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 0.028\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0.081 & 0.000 \\ 0.000 & 0.081 \end{bmatrix}$$

where the matrix $\mathbf{W}$ relates WWG records for animals 4 and 9 with missing records for PWD to sex effects and $\mathbf{B}$ relates WWG records for calves 5, 6, 7 and 8 to sex effects. The matrices $\mathbf{W}'$ and $\mathbf{B}'$ are:

$$\mathbf{W}' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B}' = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

However, all animals recorded for PWG also had records for WWG, therefore:

$$\mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 = r_o^{22}\mathbf{X}_2'\mathbf{X}_2 = 0.037\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0.074 & 0 \\ 0 & 0.074 \end{bmatrix}$$

and:

$$\mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 = r_o^{12}\mathbf{X}_1'\mathbf{X}_2 = -0.010\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -0.02 & 0.00 \\ 0.00 & -0.02 \end{bmatrix}$$

Excluding ancestors, the matrix $\mathbf{Z}_1$ is an identity matrix because every animal has a record for WWG. Therefore, $\mathbf{Z}_1'\mathbf{Z}_1 = \mathbf{I}$ and:

$$\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 = \text{diag}(0.025, 0.028, 0.028, 0.028, 0.028, 0.025)$$

However:

$$\mathbf{Z}_2 = \text{diag}(0, 1, 1, 1, 1, 0)$$

indicating that calves 4 and 9 have no records for PWG, and:

$$\mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 = \text{diag}(0.0, 0.037, 0.037, 0.037, 0.037, 0.0)$$

To account for ancestors (animals 1 to 3), $\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1$ and $\mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2$ given above augmented with three rows and columns of zeros.

The other matrices in the MME can be calculated through matrix multiplication. The matrix $\mathbf{A}^{-1}$ can be set up and $\mathbf{A}^{-1}{}^{*}\mathbf{G}^{-1}$ (where $*$ means the Kronecker product) added to the appropriate matrices, as described in Section 5.2.2, to obtain the MME. The MME are too large to be presented but solutions from solving the equations are shown below, together with solutions from the univariate analyses of WWG and PWG.

Downloaded From Vet-eBooks.com

	Multivariate analysis		Univariate analysis	
Effects	WWG	PWG	WWG	PWG
Sex				
Male	4.367	6.834	4.364	6.784
Female	3.657	6.007	3.648	5.873
Animal				
1	0.130	0.266	0.077	0.273
2	–0.084	–0.075	–0.081	0.000
3	–0.098	–0.194	–0.058	–0.165
4	0.007	0.016	0.003	–0.025
5	–0.343	–0.555	–0.250	–0.463
6	0.192	0.440	0.098	0.517
7	–0.308	–0.483	–0.237	–0.460
8	0.201	0.349	0.143	0.392
9	–0.018	–0.119	0.010	–0.230

The differences for sex solutions for WWG from the multivariate and univariate analyses are very similar to those in Section 5.2 since there are no missing records in WWG. However, sex differences in the two analyses are different for PWG due to the missing records. Again, most of the benefit in terms of breeding values from the multivariate analysis was observed in WWG, as explained in Section 5.2. However, for the calves with missing records for PWG, there was a substantial change in their proofs compared with the estimates from the univariate analysis. The proofs for these calves for PWG are based on pedigree information only in the univariate analysis but include information from the records for WWG in the multivariate analysis due to the genetic and residual correlations between the two traits. Thus the inclusion of a correlated trait in a multivariate analysis is of much benefit to animals with missing records for the other trait.

5.4 Unequal Design Matrices

Unequal design matrices for different traits arise when traits in the multivariate analysis are affected by different fixed or random effects – for instance, the multivariate analysis of yields in different lactations as different traits. Due to the fact that calving in different parities occur in different years, herd–year–season (HYS) effects associated with each lactation are different, and an appropriate model should include different HYS for yield in each parity. An example where random effects might be different for different traits is the joint analysis for weaning weight and lean per cent in beef cattle. It might be considered that random maternal effect (see Chapter 7) is only important for weaning weight and the model for the analysis will include maternal effects only for weaning weight.

5.4.1 Numerical example

Example 5.3

Using the fat yield data in Chapter 4 analysed with a repeatability model, the principles of a multivariate analysis with unequal design are illustrated below, considering

Downloaded From Vet-eBooks.com

yield in each parity as different traits and fitting a different HYS effect for each trait. The data with each lactational yield treated as different traits and HYS recoded for each trait is:

Cow	Sire	Dam	HYS1	HYS2	FAT1	FAT2
4	1	2	1	1	201	280
5	3	2	1	2	150	200
6	1	5	2	1	160	190
7	3	4	1	1	180	250
8	1	7	2	2	285	300

HYS1, HYS2, herd–year–season for parity 1 and 2, respectively; FAT1, FAT2, fat yield in parity 1 and 2.

The aim is to carry out a multivariate estimate of breeding values for fat yield in lactation 1 (FAT1) and 2 (FAT2) as different traits. Assume the genetic parameters are:

$$\mathbf{R} = \begin{bmatrix} 65 & 27 \\ 27 & 70 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 35 & 28 \\ 28 & 30 \end{bmatrix} \quad \text{with} \quad \mathbf{R}^{-1} = \begin{bmatrix} 0.018 & -0.007 \\ -0.007 & 0.017 \end{bmatrix} \quad \text{and}$$

$$\mathbf{G}^{-1} = \begin{bmatrix} 0.113 & -0.105 \\ -0.105 & 0.132 \end{bmatrix}$$

The model for the analysis is the same as in Section 5.2 but the MME are different from those in Section 5.2 because HYS effects are peculiar to each trait. The MME with the equations written out separately for each trait are:

$$\begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 + \mathbf{A}^{-1}g^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{12} \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1 + \mathbf{A}^{-1}g^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{22} \end{bmatrix}^{-1}$$
$$\begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{X}_1'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{22}\mathbf{y}_2 \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{y}_2 \end{bmatrix}$$

Setting up the design matrices and MME

The matrix $\mathbf{X}_1$ now relates HYS effects to FAT1 while $\mathbf{X}_2$ relates HYS effects to FAT2. The transposes of these matrices are:

$$\mathbf{X}_1' = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2' = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are equal and they are identity matrices of order 5 by 5 considering only animals with records. The matrix $\mathbf{A}^{-1}$ has been presented in Section 4.2.2. The remaining matrices in the MME can be obtained as described in

Downloaded From Vet-eBooks.com

previous sections. The MME have not been presented because they are too large. The solutions to the MME are:

	Solutions			
	Multivariate analysis		Univariate analysis	
Effects	FAT1	FAT2	FAT1	FAT2
HYS				
1	175.7	243.2	175.8	237.1
2	219.6	240.6	220.4	250.0
Animal				
1	8.969	8.840	6.933	8.665
2	–2.999	–2.777	–2.59	–2.244
3	–5.970	–6.063	–4.341	–6.422
4	11.754	11.658	9.103	12.197
5	–16.253	–15.824	–12.992	–15.563
6	–17.314	–15.719	–15.197	–11.149
7	8.690	8.138	7.566	7.696
8	22.702	20.931	19.417	15.560

Similar to the results in Section 5.2.2, the largest increase in breeding value under the multivariate analysis compared with the univariate was in FAT2. This may be due to the lower heritability of FAT2 compared with FAT1, as explained earlier.

Compared with the results from the repeatability model (Section 4.2.2) on the same data with corresponding estimates of genetic parameters, the mean breeding values for FAT1 and FAT2 for animals in the multivariate analysis are similar to the breeding value estimates from the former. The ranking of animals is the same under both models. Also, the differences between solutions for corresponding levels of HYS are very similar. In general, the repeatability model on successive records of animals is very efficient compared with the multivariate model, especially when the genetic correlation among records is high. The genetic correlation used for the multivariate analysis was 0.86. Visscher (1991) reported a loss of 0 to 5% in efficiency in genetic gain with a repeatability model on first and second fat yield compared with the multivariate model using a selection index. Mrode and Swanson (1995) reported a rank correlation of 0.98 between breeding value estimates for milk yield in first and second lactations, from a repeatability model and multivariate analysis for bulls with 60 or more daughters. The benefit of the repeatability model compared with the multivariate is that it is less computationally demanding and fewer estimates of genetic parameters are required.

If there are missing records in addition to unequal design matrices for traits in a multivariate analysis, the analysis can be carried out using the same principles outlined in Section 5.3, defining different residual covariance matrices for each pattern of missing traits.

5.4.2 Illustrating the computation of DYD from a multivariate model

The computation of DYD from a multivariate model is illustrated using sire 1 with three daughters (cows 4, 6 and 8) in Example 5.3. As shown in Section 5.2, since

Downloaded From Vet-eBooks.com

each daughter has one record per each trait, $\mathbf{YD}_{ij}$ for the daughter i and trait j equals $(y_{ij} - x_{ij}\hat{b})$. Thus:

$$\begin{pmatrix} YD_{41} \\ YD_{42} \end{pmatrix} = \begin{pmatrix} 201 - 175.5 \\ 280 - 243.2 \end{pmatrix} = \begin{pmatrix} 25.7 \\ 36.8 \end{pmatrix}; \quad \begin{pmatrix} YD_{61} \\ YD_{62} \end{pmatrix} = \begin{pmatrix} 160 - 219.6 \\ 190 - 243.2 \end{pmatrix} = \begin{pmatrix} -59.6 \\ -53.2 \end{pmatrix}$$

and:

$$\begin{pmatrix} YD_{81} \\ YD_{82} \end{pmatrix} = \begin{pmatrix} 285 - 219.6 \\ 300 - 240.6 \end{pmatrix} = \begin{pmatrix} 65.4 \\ 59.4 \end{pmatrix}$$

For all three daughters, the dams are known, therefore $\mathbf{W}_{2\text{prog}}$ in Eqn 5.13 is the same for all daughters and is:

$$\mathbf{W}_{2\text{prog}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + 2\mathbf{G}^{-1})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})$$
$$= \begin{pmatrix} 0.2439 & -0.2176 \\ -0.2176 & 0.2802 \end{pmatrix}^{-1} \begin{pmatrix} 0.0183 & -0.0071 \\ -0.0071 & 0.0170 \end{pmatrix} = \begin{pmatrix} 0.1713 & 0.0821 \\ 0.1078 & 0.1244 \end{pmatrix}$$

The correction of the daughters' YD for the breeding values of the mates of the sire is follows:

$$\begin{pmatrix} 2YD_{41} - \hat{a}_{21} \\ 2YD_{42} - \hat{a}_{22} \end{pmatrix} = \begin{pmatrix} 51.4 - (-2.999) \\ 73.6 - (-2.777) \end{pmatrix} = \begin{pmatrix} 54.399 \\ 76.377 \end{pmatrix}$$
$$\begin{pmatrix} 2YD_{61} - \hat{a}_{51} \\ 2YD_{62} - \hat{a}_{52} \end{pmatrix} = \begin{pmatrix} -119.2 - (-16.253) \\ -106.4 - (-15.824 \end{pmatrix} = \begin{pmatrix} -102.947 \\ -90.576 \end{pmatrix}$$
$$\begin{pmatrix} 2YD_{81} - \hat{a}_{71} \\ 2YD_{82} - \hat{a}_{72} \end{pmatrix} = \begin{pmatrix} 130.8 - 8.690 \\ 118.8 - 8.138 \end{pmatrix} = \begin{pmatrix} 122.110 \\ 110.662 \end{pmatrix}$$

Since α_{prog} equals 1 for all daughters of the bull, **DYD** for sire 1, using Eqn 5.13, is:

$$\mathbf{DYD} = (3\mathbf{W}_{2_{\text{prog}}})^{-1} \left[\mathbf{W}_{2_{\text{prog}}} \begin{pmatrix} 53.399 \\ 76.377 \end{pmatrix} + \mathbf{W}_{2_{\text{prog}}} \begin{pmatrix} -102.947 \\ -90.576 \end{pmatrix} + \mathbf{W}_{2_{\text{prog}}} \begin{pmatrix} 122.110 \\ 110.662 \end{pmatrix} \right]$$
$$= \begin{pmatrix} 24.5207 \\ 32.1543 \end{pmatrix}$$

Using Eqn 5.12, the breeding value of sire 1 can be calculated as:

$$\begin{pmatrix} \hat{a}_{11} \\ \hat{a}_{12} \end{pmatrix} = \mathbf{M}_3 \begin{pmatrix} 24.5207 \\ 32.1543 \end{pmatrix} = \begin{pmatrix} 7.439 \\ 7.387 \end{pmatrix}$$

where:

$$\mathbf{M}_3 = \left(2G^{-1}0.5 + 0.5G^{-1} \sum_3 \mathbf{W}_{2_{\text{prog}}} \alpha_{prog} \right)^{-1} \left(0.5G^{-1} \sum_3 \mathbf{W}_{2_{\text{prog}}} \alpha_{prog} \right)$$
$$\mathbf{M}_3 = \begin{pmatrix} 0.1247 & -0.1110 \\ -0.1110 & 0.1432 \end{pmatrix}^{-1} \begin{pmatrix} 0.0120 & -0.0058 \\ -0.0058 & 0.0116 \end{pmatrix} = \begin{pmatrix} 0.1937 & 0.0836 \\ 0.1099 & 0.1459 \end{pmatrix}$$

The vector of breeding value calculated for sire 1 using Eqn 5.12 is slightly lower than that shown earlier in the table of results as contributions from the grand-progeny of the sire are not included.

Downloaded From Vet-eBooks.com

5.5 Multivariate Models with No Environmental Covariance

In some cases, a multivariate analysis may be necessary when individual animals have records for one trait (or subset of traits) but relatives have records on a different trait (or subset of traits). For instance, in beef cattle, if selection is for dual-purpose sires, male and female calves might be reared in different environments (different feedlots) and body weight recorded in male calves and milk yield in female calves. The evaluation of the sires will be based on multivariate analysis of these two traits. A special feature of such a multivariate analysis is that there is no environmental covariance between the traits as the two traits are not observed in the same individual. In Section 5.5.1, the details of such a model are discussed and its application to example data is illustrated.

Also, when the same trait is measured on relatives in different environments such that the genetic correlation between performances in the two environments is not one, a multivariate analysis might be the optimum means to evaluate sires. For example, milk yield may be recorded on the daughters of a bull in two different environments, say, in a tropical environment and a temperate environment. Such a multivariate analysis will treat milk yield in the various environments as different traits. However, as the number of environments increases, the data might be associated with a heterogeneous fixed effects structure that might be difficult to model correctly in multivariate analysis, such that it might be useful, for practical purposes of implementation, to analyse not the original data but summaries of the data. A very good illustration of such a multivariate analysis is the multi-trait sire model used by the international bull evaluation service Interbull (Uppsala, Sweden), for the across-country evaluation of dairy sires. This multi-trait sire model, commonly referred to as MACE (multi-trait across-country evaluations), analyses deregressed breeding values (DRB) of sires in different countries as different traits. The use of DRB could be regarded as utilizing a variable that summarizes the daughter performances of bulls in different countries. This avoids the need to model at the Interbull centre the heterogeneous fixed effects structure, such as different herd management systems and complex national climatic conditions associated with the daughters' milk performance records in the different countries. MACE plays a very important role in the international trade of dairy cattle and in Section 5.5.2 the model for MACE is discussed and illustrated.

5.5.1 Different traits recorded on relatives

Defining the model

In this situation, with different traits recorded on relatives in different environments, the different traits are not observed on the same individual, and so there is not environmental covariance between the traits. Therefore, the residual covariance matrix $\mathbf{R}$ is diagonal. Thus for n traits:

$$\mathbf{R} = \text{diag}(\sigma^2_{e1}, \sigma^2_{e2},\ldots,\sigma^2_{en}) = \text{diag}(r_{11}, r_{22},\ldots,r_{nn})$$

and:

$$\mathbf{R}^{-1} = \text{diag}(r^{11}, r^{22},\ldots,r^{nn})$$

Downloaded From Vet-eBooks.com

However, var(**a**), where $\mathbf{a}' = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$, the vector of breeding values, is:

$$\text{var}(\mathbf{a}) = \mathbf{A}^{*}\mathbf{G}$$

where * refers to the direct product, **A** is the relationship matrix and **G** the covariance matrix for additive genetic effects. Schaeffer *et al.* (1978) discussed this model in detail but from the standpoint of variance component estimation.

Assuming there are two traits, the model for the analysis is as given in Eqn 5.1 but with **R** and **G** defined as above. The MME for the BLUP of **a** and estimable functions of **b** are:

$$\begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} \begin{bmatrix} r^{11}\mathbf{X}_1'\mathbf{X}_1 & 0 & r^{11}\mathbf{X}_1'\mathbf{Z}_1 & 0 \\ 0 & r^{22}\mathbf{X}_2'\mathbf{X}_2 & 0 & r^{22}\mathbf{X}_2'\mathbf{Z}_2 \\ r^{11}\mathbf{Z}_1'\mathbf{X}_1 & 0 & r^{11}\mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}g^{11} & \mathbf{A}^{-1}g^{12} \\ 0 & r^{22}\mathbf{Z}_2'\mathbf{X}_2 & \mathbf{A}^{-1}g^{21} & r^{22}\mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}^{-1}g^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' r^{11}\mathbf{y}_1 \\ \mathbf{X}_2' r^{22}\mathbf{y}_2 \\ \mathbf{Z}_1' r^{11}\mathbf{y}_1 \\ \mathbf{Z}_2' r^{22}\mathbf{y}_2 \end{bmatrix}$$

An illustration

Example 5.4

Consider the following data on the progeny of three sires born in the same herd; assuming that selection is for dual-purpose sires, such that the male and female calves are raised on different feeding regimes, with males recorded for yearling weight and females for fat yield:

Calf	Sex	Sire	Dam	HYS	Yearling weight (kg)	Fat yield (kg)
4	Female	1	Unknown	–	–	–
9	Male	1	4	1	375.0	–
10	Male	2	5	2	250.0	–
11	Male	1	6	2	300.0	–
12	Male	3	Unknown	1	450.0	–
13	Female	1	7	1	–	200.0
14	Female	3	8	2	–	160.0
15	Female	2	Unknown	3	–	150.0
16	Female	2	13	2	–	250.0
17	Female	3	15	3	–	175.0

HYS, herd–year–season.

The aim is to estimate HYS effects for both traits and predict breeding values for yearling weight and fat yield for all animals, carrying out a multivariate analysis. Note that animal 4 is just an ancestor and has no yield record for either trait. Assume that the additive genetic covariance matrix (**G**) is:

$$\mathbf{G} = \begin{bmatrix} 43 & 18 \\ 18 & 30 \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \text{diag}(77, 70)$$

Downloaded From Vet-eBooks.com

Then $\mathbf{R}^{-1}$ = diag(1/77, 1/70) and:

$$\mathbf{G}^{-1} = \begin{bmatrix} 0.0311 & -0.0186 \\ -0.0186 & 0.0445 \end{bmatrix}$$

The MME given earlier can easily be set up using the principles discussed so far in this chapter. Solving the MME by the direct inverse of the coefficient matrix gave the following solutions:

	Solutions	
Effects	Yearling weight (kg)	Fat (kg)
HYS		
1	411.833	193.299
2	275.955	205.344
3	–	163.315
Animal		
1	–0.472	2.519
2	–3.350	0.381
3	0.856	–3.208
4	–5.142	–3.936
5	–4.778	–2.000
6	4.778	2.000
7	2.177	3.628
8	–4.940	–5.251
9	–10.234	–3.817
10	–8.842	–2.810
11	6.932	4.260
12	11.568	3.060
13	3.029	6.701
14	–6.395	–11.485
15	–2.797	–1.680
16	4.193	10.797
17	0.526	0.050

Selection of dual-purpose sires will be based on some combination of breeding value estimates for yearling weight and fat yield. If equal weights were given to yearling weight and fat yield, sire 1 would be the best of the three sires, followed by sire 3.

5.5.2 The multi-trait across-country evaluations (MACE)

The sire model for MACE was originally proposed by Schaeffer (1994) and involved the analysis of the DYD of bulls in different countries as different traits, with the number of daughters of a bull used as a weighting factor. The genetic correlations among DYDs of bulls in different countries were incorporated. The genetic correlations accounted for genotype by environment (G × E) interactions and differences in national models for genetic evaluations among the countries. The genetic correlations among several countries used by Interbull are usually of medium to high value.

Downloaded From Vet-eBooks.com

However, due to the inability of some countries to compute DYDs for bulls, the deregressed proofs (DRP) of bulls became the variable of choice (Sigurdsson and Banos, 1995) and the weighting factor became the effective daughter contributions (EDC) of bulls (Fiske and Banos, 2001). The model in matrix notation is:

$$\mathbf{y}_i = 1\boldsymbol{\mu}_i + \mathbf{Z}_i\mathbf{Q}\mathbf{w}_i + \mathbf{Z}_i\mathbf{a}_i + \mathbf{e}_i \tag{5.14}$$

where $\mathbf{y}_i$ is the vector of DRP from country i for one trait such as milk yield, $\boldsymbol{\mu}_i$ is a mean effect for country i, which reflects the definition of the genetic base for that country, $\mathbf{w}_i$ is the vector of genetic group effects of phantom parents, $\mathbf{a}_i$ is the vector random sire proof for country i and $\mathbf{e}_i$ is the vector of random mean residuals.The matrix $\mathbf{Q}_i$ relates sires to phantom groups (see Section 3.6) and $\mathbf{Z}_i$ relates DRP to sires. Given two countries, the variance–covariance matrix for $\mathbf{w}$, $\mathbf{s}$ and $\mathbf{e}$ is:

$$\text{var}\begin{pmatrix}\mathbf{w}_1\\ \mathbf{w}_2\\ \mathbf{s}_1\\ \mathbf{s}_2\\ \mathbf{e}_1\\ \mathbf{e}_2\end{pmatrix} = \begin{pmatrix}\mathbf{A}_{pp}g_{11} & \mathbf{A}_{pp}g_{12} & \mathbf{A}_{pn}g_{11} & \mathbf{A}_{pn}g_{12} & 0 & 0\\ \mathbf{A}_{pp}g_{21} & \mathbf{A}_{pp}g_{22} & \mathbf{A}_{pn}g_{21} & \mathbf{A}_{pn}g_{22} & 0 & 0\\ \mathbf{A}_{np}g_{11} & \mathbf{A}_{np}g_{12} & \mathbf{A}_{nn}g_{11} & \mathbf{A}_{nn}g_{12} & 0 & 0\\ \mathbf{A}_{np}g_{21} & \mathbf{A}_{np}g_{22} & \mathbf{A}_{nn}g_{21} & \mathbf{A}_{nn}g_{22} & 0 & 0\\ 0 & 0 & 0 & 0 & \mathbf{D}_1\sigma^2_{e1} & 0\\ 0 & 0 & 0 & 0 & 0 & \mathbf{D}_2\sigma^2_{e2}\end{pmatrix}$$

where n and p are the number of bulls and groups, respectively, g_{ij} is the sire genetic (co)variance between countries i and j, and $\mathbf{A}$ is the additive genetic relationship for all bulls and phantom parent groups based on the maternal grandsire (MGS) model (see Section 3.6), σ^2_{ei} is the residual variance for country i, and $\mathbf{D}_i$ is the reciprocal of the effective daughter contribution of the bull in the ith country.

The variable DRP, analysed in Eqn 5.14, are obtained by deregressing the national breeding values of bulls such that they are independent of all country group effects and additive genetic relationships among bulls, their sires and paternal grandsires, which are included in the MACE analysis (Sigurdsson and Banos, 1995). DRP may therefore contain additive genetic contributions from the maternal pedigree, which are included at the national level but not in MACE. The deregression procedure involves solving the MME associated with Eqn 5.14 for the right-hand side details. The details of the procedure are outlined in Appendix F. The computation of the EDC of bulls used as the weighting factor for the analysis of DRP in Eqn 5.14 is dealt with in a subsequent section.

The MME for the above model, which are modified such that sire solutions have group solutions incorporated (see Section 3.6) are:

$$\begin{pmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & 0\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1}\otimes\mathbf{G}^{-1} & -\mathbf{A}^{-1}\mathbf{Q}\otimes\mathbf{G}^{-1}\\ 0 & -\mathbf{Q}'\mathbf{A}^{-1}\otimes\mathbf{G}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}\otimes\mathbf{G}^{-1}\end{pmatrix}\begin{pmatrix}\hat{\mathbf{c}}\\ \mathbf{Q}\hat{\mathbf{w}}+\hat{\mathbf{a}}\\ \hat{\mathbf{w}}\end{pmatrix} = \begin{pmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\\ 0\end{pmatrix} \tag{5.15}$$

Genetic groups are defined for unknown sires and MGS on the basis of country of origin and year of birth of their progeny. Also, maternal granddams (MGDs) are always assumed unknown and assigned to phantom groups on the same basis.

Downloaded From Vet-eBooks.com

Then $\mathbf{A}^{-1}$ can be obtained by the rules outlined in Section 2.4, which can be briefly summarized in the table below, taking into account the contribution to the groups for MGDs. Given a list of pedigrees with the *i*th line consisting of a bull, its sire or group, its MGS or group and a group for its MGD, then contributions to $\mathbf{A}^{-1}$ are as follows:

	Bull	Sire	MGS	MGD
Bull	d	$-0.5d$	$-0.25d$	$-0.25d$
Sire	$-0.5d$	$0.25d$	$0.125d$	$0.125d$
MGS	$-0.25d$	$0.125d$	$0.0625d$	$0.0625d$
MGD	$-0.25d$	$0.125d$	$0.0626d$	$0.0625d$

where $d = 16/(11 + m)$ and $m = 0$ if both sire and MGS are known, $m = 1$ if the sire is known but MGS is unknown, $m = 4$ if the sire is unknown and the MGS is known, and $m = 5$ if both sire and MGS are unknown.

Usually there are dependencies among group effect equations and 1 is added to the diagonals of the phantom group effects in the inverse of the relationship matrix to overcome these dependencies. Then the group solutions sum to zero, and so the solutions for bulls are relative to the same genetic base within each country. The addition of 1 to the diagonals of the phantom groups implies that group effects are random, with expected values of zero. Since group effects represent differences in the effects of previous selection, which should not have expected values of zero, Schaeffer (1994) indicated that this approach could also be regarded as a biased estimation of the fixed effects of phantom groups. That is, a small amount of bias in the estimates of the phantom groups is accepted in exchange for the hope of getting estimates with smaller mean square errors.

Computing effective daughter contribution

The use of EDC instead of the number of daughters as a weighting factor was proposed by Fiske and Banos (2001) from a simulation study in which they demonstrated that using the numbers of daughters resulted in biased estimates of sire variances used in MACE and international reliabilities. The computation of EDC for a bull accounts for such factors as contemporary group (CG) structure for the bull's daughters, the correlation between observations on the same daughter and the reliability of the performance of the daughters' dams. Thus the EDC provides a measure of the precision of the daughter information used to compute the DRP of the bull. The formula for the computation of EDC (Fiske and Banos, 2001), which included the performance of the dam of the daughter k of bull i is:

$$\mathrm{EDC}_i = \sum_k \frac{\lambda rel_{k(o)}}{4 - rel_{k(o)} \cdot (1 + rel_{dam(o)})}$$

where the summation is over all the k daughters of the bull, $\lambda = (4 - h^2)/h^2$, $rel_{dam(o)}$ is the reliability of the dam's own performance, $rel_{k(o)}$ is the reliability of the animal k's own performance computed as:

$$rel_{k(o)} = \frac{n_k h^2}{1 + (n_k - 1)r}$$

with r being the reliability of the animal's records, n_k the number of lactations of the daughter k of the sire adjusted for the CG size computed as:

Downloaded From Vet-eBooks.com

$$n_k = \sum_i 1 - 1/n_{jkl}$$

where n_{jkl} is the size of the CG_j in which the daughter k of sire i made her lth lactation.

An example of MACE for two countries

Example 5.5

The data set below consists of bull breeding values (kg) and DRP for fat yield for six bulls from two countries. Two of the bulls have evaluations in both countries and in addition each country had two other bulls, which were the only progeny tested in that country. A MACE is implemented using the data set. Assume residual variances of 206.5 kg^2 and 148.5 kg^2 for countries 1 and 2, respectively, with corresponding sire additive genetic variances of 20.5 kg^2 and 9.5 kg^2. The sire genetic covariance between fat yield in both countries was assumed to be 12.839 kg, giving a genetic correlation of 0.92.

	Country 1			Country 2		
Sire	EDC	BV	DRP	EDC	BV	DRP
1	58	9.0	9.7229	90	13.5	14.5088
2	150	10.1	9.9717	65	7.6	7.7594
3	20	15.8	19.2651	–	–	–
4	25	–4.7	–8.5711	–	–	–
5	–	–	–	30	19.6	23.9672
6	–	–	–	55	–5.3	–9.6226

EDC = effective daughter contribution; BV = breeding value; DRP = deregressed proof.

Assume that the sires in the data set have the following pedigree structure, with unknown sires, MGS and MGD assigned to group G_i, with $i = 1, \ldots 5$.

Bull	Sire	MGS	MGD
1	7	G3	G5
2	8	9	G5
3	7	2	G5
4	1	G2	G5
5	8	G3	G4
6	1	9	G4
7	G1	G2	G4
8	G1	G2	G4
9	G1	G3	G4

Computing sire breeding values

The matrix $\mathbf{G}^{-1}$ for Example 5.5 is:

$$\mathbf{G}^{-1} = \begin{pmatrix} 0.31762 & -0.42925 \\ -0.42925 & 0.68539 \end{pmatrix}$$

Downloaded From Vet-eBooks.com

The inverses of the matrix of residual variances for countries 1 and 2 are:

$$\mathbf{R}_1^{-1} = \text{diag}(0.2809, 0.7264, 0.0969, 0.1211, 0, 0)$$

and:

$$\mathbf{R}_2^{-1} = \text{diag}(0.6061, 0.4377, 0, 0, 0.2020, 0.3704)$$

The design matrix **X** is:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

and:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{pmatrix} 1.2252 & 0 \\ 0 & 1.6162 \end{pmatrix}$$

The matrix **Z** is an identity matrix of order 12, considering only bulls with evaluations. The matrix $\mathbf{A}^{-1}$ is set up using the rules outlined earlier. The remaining matrices in Eqn 5.15 could be obtained through matrix multiplication and addition. The MME are of the order of 30 by 30 and have not been shown. Solutions to the MME by direct inversion gave the following results:

	Solutions			
Effects	Country 1		Country 2	
Country effect	7.268		9.036	
Animal/group	A	B	A	B
1	2.604	9.871	2.661	11.697
2	2.176	9.444	0.403	9.439
3	8.059	15.327	5.001	14.037
4	–9.865	–2.597	–5.605	3.431
5	13.634	20.902	9.728	18.764
6	–18.086	–10.818	–13.203	–4.167
7	4.310	11.578	3.071	12.106
8	7.015	14.283	4.489	13.525
9	–6.299	0.969	–5.059	3.977
G1	0.174	7.442	–0.092	8.944
G2	–0.124	7.144	0.126	9.162
G3	–0.071	7.197	0.264	9.300
G4	0.087	7.355	–0.288	8.748
G5	–0.067	7.201	–0.010	9.026

A = solutions for animals and groups from the MME; B = solutions for animals and groups expressed in each country scale.

The solutions for animals and groups were expressed in each country scale by adding the solution for country effects for country *i* to the animal and group solutions of the *i*th country. As indicated earlier, the sum of the group solutions is zero.

Downloaded From Vet-eBooks.com

In the next section, some of the bull solutions are partitioned to contributions from various sources to gain a better understanding of MACE.

Equations for partitioning bull evaluations from MACE

The equations for sire proofs from Eqn 5.15 are:

$$(\mathbf{Z'R^{-1}Z} + \mathbf{A^{-1}} \otimes \mathbf{G^{-1}})\hat{\mathbf{a}} = (\mathbf{A^{-1}Q} \otimes \mathbf{G^{-1}})\hat{\mathbf{g}} + \mathbf{Z'R^{-1}}\,(\mathbf{y} - \mathbf{X}\hat{\mathbf{c}}) \tag{5.16}$$

where:

$$\hat{\mathbf{a}} = \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{s}}$$

Thus Eqn 5.16 can be expressed as:

$$(\mathbf{Z'R^{-1}Z} + \mathbf{A^{-1}} \otimes \mathbf{G^{-1}})\hat{\mathbf{a}} = (\mathbf{A^{-1}Q} \otimes \mathbf{G^{-1}})\hat{\mathbf{g}} + \mathbf{Z'R^{-1}\,Z(CD)} \tag{5.17}$$

where:

$$\mathrm{CD} = (\mathbf{Z'R^{-1}Z})^{-1}\,(\mathbf{Z'R^{-1}}(\mathbf{y} - \mathbf{X}\hat{\mathbf{c}}))$$

CD (country deviation) is simply a vector of weighted average of corrected DRP in all countries where the bull has a daughter, the weighting factor being the reciprocal of EDC multiplied by the residual variance in each country. Since $\mathbf{R^{-1}}$ is diagonal, **CD** is equal to the vector $(\mathbf{y} - \mathbf{X}\hat{\mathbf{c}})$.

For a particular bull with a direct progeny (e.g. son), Eqn 5.17 can be written as:

$$\begin{aligned}(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})\hat{\mathbf{a}}_{bull} = {} & \mathbf{G^{-1}}\alpha_{par}\,(\hat{\mathbf{a}}_{sire} + 0.5(\hat{\mathbf{a}}_{mgs} + \hat{\mathbf{g}})) + \mathbf{Z'R^{-1}Z(CD)} \\ & + \mathbf{G^{-1}}\sum \alpha_{prog}\,(\hat{\mathbf{a}}_{prog} - 0.25\hat{\mathbf{a}}_{mate})\end{aligned} \tag{5.18}$$

where $\alpha_{par} = \frac{8}{11}, \frac{8}{15}, \frac{2}{3}$ or $\frac{1}{2}$ if both sire and MGS (maternal grandsire), only MGS, only sire or no parents are known, respectively; and $\alpha_{prog} = \frac{8}{11}$ if bull's mate (MGS of the progeny) is known or $\frac{2}{3}$ if unknown. The above values for α_{par} and α_{prog} are based on the assumption that $\mathbf{A^{-1}}$ has been calculated without accounting for inbreeding. Note that in Eqn 5.18:

$$\alpha_{bull} = 2\alpha_{par} + 0.5\alpha_{prog}$$

Equation 5.18 can be expressed as:

$$\begin{aligned}(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})\hat{\mathbf{a}}_{bull} = {} & 2\mathbf{G^{-1}}\alpha_{par}\,(\mathbf{PA}) + (\mathbf{Z'R^{-1}Z})\mathbf{CD} \\ & + 0.5\mathbf{G^{-1}}\sum \alpha_{prog}\,(2\hat{\mathbf{a}}_{prog} - 0.5\hat{\mathbf{a}}_{mate})\end{aligned}$$

where:

$$\mathbf{PA} = 0.5\hat{\mathbf{a}}_{sire} + 0.25(\hat{\mathbf{a}}_{mgs} + \hat{\mathbf{g}})$$

Pre-multiplying both sides of the equation by $(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})^{-1}$ gives:

$$\hat{\mathbf{a}}_{bull} = \mathbf{W}_1\mathbf{PA} + \mathbf{W}_2\mathbf{YD} + \mathbf{W}_3\mathbf{PC} \tag{5.19}$$

where:

$$\mathbf{PC} = \sum \alpha_{prog}\,(2\hat{\mathbf{a}}_{prog} - 0.5\hat{\mathbf{a}}_{mate}) / \sum \alpha_{prog} \quad \text{and} \quad \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3 = \mathbf{I}$$

Downloaded From Vet-eBooks.com

The matrices $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ are the product of $(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})^{-1}$ and $2\mathbf{G^{-1}}\alpha_{par}$, $\mathbf{Z'R^{-1}Z}$, and $0.5\mathbf{G^{-1}}\sum\alpha_{prog}$, respectively. Using Eqn 5.19, the contributions from different sources of information from different countries to the MACE of a bull can be computed.

If the progeny in Eqn 5.19 is not a direct progeny of the bull but a maternal grandson of the bull then α_{prog} equals $\frac{4}{11}$ if mate (sire) is known or $\frac{4}{15}$ if unknown. Then Eqn 5.19 becomes:

$$(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})\hat{\mathbf{a}}_{bull} = \mathbf{G^{-1}}\alpha_{par}(\hat{\mathbf{a}}_{sire} + 0.5(\hat{\mathbf{a}}_{mgs} + \hat{\mathbf{g}})) + \mathbf{Z'R^{-1}Z(CD)} + \mathbf{G^{-1}}\sum\alpha_{prog}(\hat{\mathbf{a}}_{prog} - 0.5\hat{\mathbf{a}}_{mate})$$

and α_{bull} now equals $2\alpha_{par} + 0.25\alpha_{prog}$ and $0.5\hat{\mathbf{a}}_{mate} = 0.5\hat{\mathbf{a}}_s$, the sire of the progeny. The above can be expressed as:

$$(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})\hat{\mathbf{a}}_{bull} = 2\mathbf{G}^{-1}_{\alpha_{par}}(\mathbf{PA}) + (\mathbf{Z'R^{-1}Z})\mathbf{CD} + 0.25\mathbf{G^{-1}}\sum\alpha_{prog}(4\hat{\mathbf{a}}_{prog} - 2\hat{\mathbf{a}}_{mate})$$

Pre-multiplying both sides by $(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})^{-1}$ gives the same equation as Eqn 5.19 but with:

$$\mathbf{PC} = \sum\alpha_{prog}\left(4\hat{\mathbf{a}}_{prog} - 2\hat{\mathbf{a}}_{mate}\right)\Big/\sum\alpha_{prog} \qquad (5.20)$$

and $\mathbf{W}_3$ now equals $(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull})^{-1}(0.25\,\mathbf{G^{-1}}\sum\alpha_{prog})$

The use of Eqn 5.19 to partition proofs from MACE is illustrated for two bulls, one with no progeny and another with a maternal grandson. First, consider bull 3 in Example 5.5 that has DRPs only in country 1 and has no progeny. Therefore, $\mathbf{CD}_{3i}$ for bull 3 in country i is:

$$CD_{31} = y_{31} - \mu_1 = 19.2651 - 7.268 = 11.997 \quad \text{and} \quad CD_{32} = 0$$

Parent average for bull 3 ($\mathbf{PA}_{3i}$) in country i is:

$$PA_{31} = 0.5(\hat{a}_{71}) + 0.25(\hat{a}_{21} + \hat{g}_{G51}) = 0.5(4.310) + 0.25(2.176 + (-0.067)) = 2.68225$$

and:

$$PA_{32} = 0.5(\hat{a}_{72}) + 0.25(\hat{a}_{22} + \hat{g}_{G52}) = 0.5(3.071) + 0.25(0.403 + (-0.010)) = 1.63375$$

where $\hat{a}_{ji}$ is the breeding value of animal j in country i and $\hat{g}_{Gji}$ is the solution for group j and in the ith country.

The residual variance for bull 3 in country 1 $(\mathbf{r}_{31}) = (\frac{1}{20})206.5 = 10.325$ and its inverse equals 0.09685. Both sire and MGS of bull 3 are known, therefore $\alpha_{bull} = \frac{16}{11}$. Then:

$$\left(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}\alpha_{bull}\right) = \begin{pmatrix} 0.09685 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0.4620 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix} = \begin{pmatrix} 0.55884 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}$$

Downloaded From Vet-eBooks.com

The matrices of weights ($\mathbf{W}_i$) using Eqn 5.19 are:

$$\mathbf{W}_1 = \begin{pmatrix} 0.55884 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}^{-1} \begin{pmatrix} 0.4620 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix} = \begin{pmatrix} 0.4229 & -0.3614 \\ -0.3614 & 1.0000 \end{pmatrix}$$

and:

$$\mathbf{W}_2 = \begin{pmatrix} 0.55884 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}^{-1} \begin{pmatrix} 0.09685 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.5771 & 0 \\ 0.3614 & 0 \end{pmatrix}$$

Therefore the vector proofs of bull 3 are:

$$\begin{pmatrix} \hat{a}_{31} \\ \hat{a}_{32} \end{pmatrix} = \mathbf{W}_1 \begin{pmatrix} 2.68225 \\ 1.63375 \end{pmatrix} + \mathbf{W}_2 \begin{pmatrix} 11.9971 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.1343 \\ 0.6644 \end{pmatrix} + \begin{pmatrix} 6.9235 \\ 4.3358 \end{pmatrix} = \begin{pmatrix} 8.058 \\ 5.000 \end{pmatrix}$$

The contribution from the DRP of bull 3 in country 1 accounts for over 85% of the MACE proof in both countries, although the bull has no DRP in country 2. Thus, with only 20 daughters, parental contribution was not very large, although in general, parental contributions will be influenced by the heritability of the traits in both countries and the genetic correlation between them.

When a bull has a proof only in country *i* and not in *j*, its proof in country *j* can be obtained (Mrode and Swanson, 1999) as:

$$\hat{\mathbf{a}}_j = \mathbf{PA}_j - (g_{ij}/g_{ii})(\hat{\mathbf{a}}_i - \mathbf{PA}_i) \tag{5.21}$$

where g_{ii} is the genetic variance in country *i* and g_{ij} the genetic covariance between countries *i* and *j*. Therefore, if interest was only in calculating the proof of bull 3 in country 2, it can be obtained from the above equation as:

$$\hat{a}_{32} = 1.63375 - (12.839/20.5)(8.059 - 2.68225) = 5.001$$

Equation 5.21 can be derived from Eqn 5.18 as follows. The equation for $\hat{a}_{32}$ from Eqn 5.18 is:

$$(g^{22}\alpha_{bull})\hat{a}_{32} = g^{22}\alpha_{par}(\hat{a}_{sire2} + 0.5(\hat{a}_{mgs2} + \hat{g}_{mgd2}) + g^{21}\alpha_{par}(\hat{a}_{sire1} + 0.5(\hat{a}_{mgs1} + \hat{g}_{mgd1}) + (g^{21}\alpha_{bull})\hat{a}_{31}$$

where $\hat{a}_{sirej}$, $\hat{a}_{mgsj}$ and $\hat{g}_{mgdj}$ are the proofs for the sire, MGS and solution for the MGD in country *j*, respectively, and g^{ii} are the inverse elements of $\mathbf{G}^{-1}$. Since $\alpha_{bull} = 2\alpha_{par}$ for bull 3, multiplying the above equation by $(2\alpha_{par})^{-1}$ gives:

$$g^{22}\hat{a}_{32} = g^{22}(PA_2) + g^{21}(PA_1) - g^{21}\hat{a}_{31}$$

$$g^{22}\hat{a}_{32} = g^{22}(PA_2) - g^{21}(\hat{a}_{31} - PA_1)$$

$$\hat{a}_{32} = PA_2 - g^{21}/g^{22}(\hat{a}_{31} - PA_1)$$

$$\hat{a}_{32} = PA_2 - g_{21}/g_{22}(\hat{a}_{31} - PA_1)$$

Thus the proof of a bull in country *j* is dependent on the parent average of the bull in country *j* and the Mendelian sampling of the bull in the *i*th country.

Partitioning the proof of bull 2 with records in both countries and a maternal grandson (bull 3) is as follows. The country deviations for bull 2 in both countries are:

$$CD_{21} = y_{21} - \mu_1 = 9.9717 - 7.268 = 2.7037$$

Downloaded From Vet-Books.com

and:

$$CD_{22} = y_{22} - \mu_2 = 7.7594 - 9.036 = -1.2766$$

Parent average for sire 2 (PA_{2i}) for country i is:

$$PA_{21} = 0.5(\hat{a}_{81}) + 0.25(\hat{a}_{91} + \hat{g}_{G51}) = 0.5(7.015) + 0.25(-6.299 + (-0.067)) = 1.916$$

$$PA_{22} = 0.5(\hat{a}_{82}) + 0.25(\hat{a}_{92} + \hat{g}_{G52}) = 0.5(4.489) + 0.25(-5.059 + (-0.010)) = 0.97725$$

Progeny contributions (PC) from bull 3 to sire 2 (PC_{32i}) in country i are:

$$PC_{21} = 4(\hat{a}_{31}) - 2(\hat{a}_{71}) = 4(8.059) - 2(4.310) = 23.616$$

$$PC_{22} = 4(\hat{a}_{32}) - 2(\hat{a}_{72}) = 4(5.001) - 2(3.071)) = 13.862$$

The residual variance for bull 2 in country 1, $(r_{21}) = (\frac{1}{150})206.5$ and country 2, $(r_{22}) = (\frac{1}{65})148.5$. Corresponding inverses were 0.72639 and 0.43771, respectively. Since both sire and MGS of bull 2 are known and he has a maternal grandson, $\alpha_{bull} = 2\alpha_{par} + 0.25\alpha_{prog} = 2(\frac{8}{11}) + 0.25(\frac{4}{11}) = 1.54545$. Therefore:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull}) = \begin{pmatrix} 0.72630 & 0 \\ 0 & 0.43771 \end{pmatrix} + \begin{pmatrix} 0.49087 & -0.66338 \\ -0.66338 & 1.05924 \end{pmatrix}$$

$$= \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}$$

From Eqn 5.19, the matrices of weights ($\mathbf{W}_i$) are:

$$\mathbf{W}_1 = \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}^{-1} \begin{pmatrix} 0.4620 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix} = \begin{pmatrix} 0.2007 & -0.1977 \\ -0.3281 & 0.5783 \end{pmatrix}$$

$$\mathbf{W}_2 = \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}^{-1} \begin{pmatrix} 0.72639 & 0 \\ 0 & 0.43771 \end{pmatrix} = \begin{pmatrix} 0.7867 & 0.2101 \\ 0.3487 & 0.3855 \end{pmatrix}$$

$$\mathbf{W}_3 = \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}^{-1} \begin{pmatrix} 0.02887 & -0.03902 \\ -0.03902 & 0.06231 \end{pmatrix} = \begin{pmatrix} 0.0125 & -0.0124 \\ -0.02051 & 0.0361 \end{pmatrix}$$

The vector of proof for bull 2 is:

$$\begin{pmatrix} \hat{a}_{21} \\ \hat{a}_{22} \end{pmatrix} = \mathbf{W}_1 \begin{pmatrix} 1.9160 \\ 0.9773 \end{pmatrix} + \mathbf{W}_2 \begin{pmatrix} 2.7037 \\ -1.2766 \end{pmatrix} + \mathbf{W}_3 \begin{pmatrix} 23.616 \\ 13.862 \end{pmatrix} = \begin{pmatrix} 2.176 \\ 0.403 \end{pmatrix}$$

Again, similar to bull 3 above, the contributions from the DRPs in both countries accounted for much of the MACE proofs of the bull 2 in countries 1 and 2.

Recently, Interbull has modified the MACE systems to use sire and dam pedigree instead of sire and maternal sire pedigree. Partitioning of bull proofs can be done as in Section 5.2.3.

Downloaded From Vet-eBooks.com

6 Methods to Reduce the Dimension of Multivariate Models

6.1 Introduction

One of the limitations of multivariate analysis is the large computational requirements of such high-dimensional analyses. The number of effects in multivariate analyses tend to increase linearly with the number of traits considered. In some cases, the available or small number of records for some of the traits can hamper the reliable estimation of a large number of covariance components simultaneously. However, developments in methodologies to model higher-dimensional data more parsimoniously (Kirkpatrick and Meyer, 2004) implies that such multivariate analysis is more feasible in terms of parameter estimation and therefore genetic evaluation.

Reducing the dimension of multivariate analysis includes methods such as canonical transformation and Cholesky decomposition, which involve the transformation of the vector of observations in addition to residual and genetic covariance matrices. Other approaches, such as principal component analysis and factor analysis, only involve reducing the rank of the genetic covariance matrix. Initially, methods that include the transformation of the vector of observations are discussed.

6.2 Canonical Transformation

In the example discussed in Section 5.2.2 both traits were affected by the same fixed effect and all animals were measured for both traits. Thus the design matrices $\mathbf{X}$ and $\mathbf{Z}$ were the same for both traits or, in other words, the traits are said to have equal design matrices. In addition there was only one random effect (animal effect) for each trait apart from the residual effect. Under these circumstances, the multivariate analysis can be simplified into n (number of traits) single trait analyses through what is called a canonical transformation (Thompson, 1977b). Canonical transformation involves using special matrices to transform the observations on several correlated traits into new variables that are uncorrelated with each other. These new variables are analysed by the usual methods for single trait evaluation, but the results (predictions) are transformed back to the original scale of the observations. Ducrocq and Besbes (1993) have presented a methodology for applying canonical transformation when design matrices are equal for all traits but with some animals having missing traits; details of the methodology, together with an illustration, are given in Appendix E, Section E.2.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

Let **y** be vectors of observations:

$$\text{var}(\mathbf{y}) = \mathbf{G} + \mathbf{R} \tag{6.1}$$

where **G** and **R** are variance and covariance matrices for the additive genetic and residual effects, respectively. Assuming **G** and **R** are positive definite matrices, then there exists a matrix Q, such that:

$$\mathbf{QRQ}' = \mathbf{I} \quad \text{and} \quad \mathbf{QGQ}' = \mathbf{W}$$

where **I** is an identity matrix and **W** is a diagonal matrix (Anderson, 1958). This implies that pre- and post-multiplication of **R** by the transformation matrix (**Q**) reduces it to an identity matrix and **G** to a diagonal matrix. The multiplication of **y** by **Q** yields a new vector of observations $\mathbf{y}^*$ that are uncorrelated:

$\mathbf{y}^* = \mathbf{Qy}$
$\text{var}(\mathbf{y}^*) = \mathbf{W} + \mathbf{I}$; which is a diagonal matrix.

Since there are no covariances between the transformed traits, they can be independently evaluated. The procedure for calculating the transformation matrix **Q** is given in Appendix E, Section E.1.

6.2.1 The model

A single trait analysis is usually carried out on each of the transformed variables. The model for the *i*th transformed variable can be written as:

$$\mathbf{y}_i^* = \mathbf{Xb}_i^* + \mathbf{Za}_i^* + \mathbf{e}_i^* \tag{6.2}$$

where $\mathbf{y}_i^*$ = vector of transformed variables for the *i*th transformed trait; $\mathbf{b}_i^*$ = vector of fixed effects for the *i*th transformed variable *i*; $\mathbf{a}_i^*$ = vector of random animal effects for transformed trait *i*; $\mathbf{e}_i^*$ = vector of random residual errors for the *i*th transformed trait; and **X** and **Z** are incidence matrices relating records to fixed and random effects, respectively.

The MME to be solved to obtain the BLUE of $\mathbf{b}_i^*$ and the BLUP of $\mathbf{a}_i^*$ are the same as those presented in Section 3.2 for the univariate model. These equations are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha_i \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_i^* \\ \hat{\mathbf{a}}_i^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}_i^* \\ \mathbf{Z}'\mathbf{y}_i^* \end{bmatrix}$$

As explained earlier, it is assumed for the *i*th trait that:

$$\text{var}(\mathbf{a}_i^*) = \mathbf{A}w_{ii}; \quad \text{var}(\mathbf{e}_i^*) = \mathbf{I} \quad \text{and} \quad \text{var}(\mathbf{y}_i^*) = \mathbf{ZAZ}'w_{ii} + \mathbf{I}$$

where w_{ii} refers to the *i*th element of the diagonal matrix **W**.

The MME are solved for $\mathbf{b}_i^*$ and $\mathbf{a}_i^*$ and the transformation back to the original scale is achieved as:

$$\mathbf{b}_i = \mathbf{Q}^{-1}\mathbf{b}_i^* \tag{6.3}$$

$$\mathbf{a}_i = \mathbf{Q}^{-1}\mathbf{a}_i^* \tag{6.4}$$

Thus the multivariate analysis is simplified to *i* single trait evaluations.

Downloaded From Vet-eBooks.com

6.2.2 An illustration

Example 6.1

The multivariate analysis for WWG and PWG in Section 5.2.2 is repeated below, carrying out a canonical transformation assuming the same genetic parameters.

The calculation of the transformation **Q** and the diagonal matrix **W** are given in Appendix E, Section E.1. Presented in Table 6.1 are the data for all calves in the original scale and as transformed variables (VAR1 and VAR2). The observations are transformed into new uncorrelated variables using the matrix **Q**. Thus for animal 4, the record would be transformed as:

$$\mathbf{Q}y_4 = \begin{bmatrix} 0.1659 & -0.0792 \\ 0.0168 & 0.1755 \end{bmatrix} \begin{bmatrix} 4.5 \\ 6.8 \end{bmatrix} = \begin{bmatrix} 0.208 \\ 1.269 \end{bmatrix}$$

The residual variance for each of the transformed variables is 1, thus heritability for the ith transformed variable = $w_{ii}/(1 + w_{ii})$ and $\alpha_i = 1/w_{ii}$.

Therefore $h_1^2 = 0.247$, $h_2^2 = 0.573$, $\alpha_1 = 1/0.3283 = 3.046$ and $\alpha_2 = 1/1.3436 = 0.744$. A single trait analysis is carried out on the transformed variates for WWG and PWG using the model and the MME in Section 5.3.1 and solutions are transformed back to the original scale.

Setting up the design matrices

The matrix **X**, which relates records for either VAR1 or VAR2 to sex effects, is exactly as the matrix $\mathbf{X}_1$ in Section 5.2.2. Similarly, **Z** is the same as $\mathbf{Z}_1$ in Section 5.2.2. For animals with records, the vector of observations $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ are equal to the column of transformed variates for WWG and PWG gains, respectively, in Table 6.1. The matrices in the MME are easily obtained through matrix multiplication and the addition to the animal equations of $\mathbf{A}^{-1}\alpha_1$ for VAR1 and $\mathbf{A}^{-1}\alpha_2$ for VAR2. $\mathbf{A}^{-1}$ has been given earlier in Section 5.2.2. For instance, the MME for VAR1 only are:

$$\begin{bmatrix} \hat{b}_1^* \\ \hat{b}_2^* \\ \hat{a}_1^* \\ \hat{a}_2^* \\ \hat{a}_3^* \\ \hat{a}_4^* \\ \hat{a}_5^* \\ \hat{a}_6^* \\ \hat{a}_7^* \\ \hat{a}_8^* \end{bmatrix} = \begin{bmatrix} 3.0 & 0.0 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.0 & 2.0 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.0 & 0.0 & 5.584 & 1.523 & 0.000 & -2.031 & 0.000 & -3.046 & 0.000 & 0.000 \\ 0.0 & 0.0 & 1.523 & 6.092 & 1.523 & 0.000 & -3.046 & -3.046 & 0.000 & 0.000 \\ 0.0 & 0.0 & 0.000 & 1.523 & 6.092 & 0.000 & -3.046 & 1.523 & 0.000 & -3.046 \\ 1.0 & 0.0 & -2.031 & 0.000 & 0.000 & 6.584 & 1.523 & 0.000 & -3.046 & 0.000 \\ 0.0 & 1.0 & 0.000 & -3.046 & -3.046 & 1.523 & 8.615 & 0.000 & -3.046 & 0.000 \\ 0.0 & 1.0 & -3.046 & -3.046 & 1.523 & 0.000 & 0.000 & 8.615 & 0.000 & -3.046 \\ 1.0 & 0.0 & 0.000 & 0.000 & 0.000 & -3.046 & -3.046 & 0.000 & 7.092 & 0.000 \\ 1.0 & 0.0 & 0.000 & 0.000 & -3.046 & 0.000 & 0.000 & -3.046 & 0.000 & 7.092 \end{bmatrix}^{-1} \begin{bmatrix} 0.549 \\ 0.194 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.208 \\ 0.085 \\ 0.108 \\ 0.105 \\ 0.235 \end{bmatrix}$$

Solving the MME for each transformed trait by direct inversion of the coefficient matrix gives the following solutions on the canonical scales. Given also are solutions for WWG and PWG after transforming the solutions for the transformed variates to the original scale.

Downloaded From Vet-eBooks.com

Table 6.1. Weaning gain and post-weaning gain for beef calves on the original and transformed scales.

				Original scale		Transformed scale	
Calves	Sex	Sire	Dam	WWG	PWG	VAR1	VAR2
4	Male	1	–	4.5	6.8	0.208	1.269
5	Female	3	2	2.9	5.0	0.085	0.926
6	Female	1	2	3.9	6.8	0.109	1.259
7	Male	4	5	3.5	6.0	0.106	1.112
8	Male	3	6	5.0	7.5	0.236	1.400

	Canonical scale		Original scale	
Effects	VAR1	VAR2	WWG	PWG
Sex				
Male	0.185	1.266	4.361	6.800
Female	0.098	1.089	3.397	5.880
Animals				
1	0.003	0.052	0.151	0.280
2	–0.002	–0.002	–0.015	–0.008
3	0.000	–0.031	–0.078	–0.170
4	–0.001	–0.002	–0.010	–0.013
5	–0.007	–0.088	–0.270	–0.478
6	0.005	0.095	0.276	0.517
7	–0.015	–0.089	–0.316	–0.479
8	0.009	0.073	0.244	0.392

The solutions are exactly the same as those obtained from the multivariate analysis in Section 5.2. The solutions are transformed to the original scale using Eqns 6.3 and 6.4. For instance, the solutions for animal 1 for both traits on the original scale are:

$$\begin{bmatrix} \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} = \begin{bmatrix} 5.7651 & 2.6006 \\ -0.5503 & 5.4495 \end{bmatrix} \begin{bmatrix} 0.0029 \\ 0.0516 \end{bmatrix} = \begin{bmatrix} 0.151 \\ 0.280 \end{bmatrix}$$

6.3 Cholesky Transformation

When all records are measured in all animals, MBLUP may be simplified by a canonical transformation as described in Section 6.2. However, if animals have some records missing and the loss of records is sequential then a Cholesky transformation can be applied (Quaas, 1984). Such situations can arise, for example, in dairy cattle due to sequential culling and different lactations being regarded as different traits.

6.3.1 Calculating the transformation matrix and defining the model

Cholesky transformation involves forming transformed variables (traits) that are environmentally independent of each other; that is, there is no residual covariance among them, therefore the residual covariance matrix for the transformed traits is an

Downloaded From Vet-eBooks.com

identity matrix. The transformation matrix $\mathbf{T}^{-1}$ is obtained by carrying out a Cholesky decomposition of **R**, the residual covariance matrix for the traits, such that:

$$\mathbf{R} = \mathbf{T}\mathbf{T}'$$

where **T** is a lower triangular matrix. The transformation matrix $\mathbf{T}^{-1}$ is the inverse of **T**. The formula for calculating **T** is given in Appendix E, Section E.3.

The vector of observations $\mathbf{y}_{ki}$ for the ith animal is transformed as:

$$\mathbf{y}^{*}_{ki} = \mathbf{T}^{-1}\mathbf{y}_{ki}$$

where k is the number of traits recorded and $\mathbf{y}^{*}_{ki}$ is the transformed vector.

If traits are missing in $\mathbf{y}_{ki}$, then the corresponding rows of $\mathbf{T}^{-1}$ are set to zero when transforming the vector of observation. Thus if $\mathbf{y}_{ki}$ is a vector of observations of n traits for the ith animal, the transformation of **y** can be illustrated as:

$$\begin{aligned}
y^{*}_{11} &= t^{11}y_{11} \\
y^{*}_{21} &= t^{21}y_{11} + t^{22}y_{21} \\
&\cdots \\
&\cdots \\
&\cdots \\
y^{*}_{n1} &= t^{n1}y_{11} + t^{n2}y_{21} + t^{nn}y_{n1}
\end{aligned}$$

where the t^{ij} above are the elements of $\mathbf{T}^{-1}$.

Given that the variance of $\mathbf{y}_{ki}$ is:

$$\text{var}(\mathbf{y}) = \mathbf{G} + \mathbf{R}$$

and the variance of the transformed variables becomes:

$$\text{var}(\mathbf{y}^{*}) = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})' + \mathbf{I} = \mathbf{G}^{*} + \mathbf{I} = \mathbf{M} + \mathbf{I} \tag{6.5}$$

where **G** is the covariance matrix for additive genetic effects and $\mathbf{G}^{*}$ is the transformed additive genetic covariance matrix. Note that $\mathbf{G}^{*}$ is not diagonal. Vectors of solutions ($\mathbf{b}^{*}_{i}$ and $\mathbf{a}^{*}_{i}$) are transformed back to the original scale ($\mathbf{b}_{i}$ and $\mathbf{a}_{i}$) as:

$$\mathbf{b}_{i} = \mathbf{T}\mathbf{b}^{*}_{i} \tag{6.6}$$

$$\mathbf{a}_{i} = \mathbf{T}\mathbf{a}^{*}_{i} \tag{6.7}$$

6.3.2 An illustration

Example 6.2

The methodology is illustrated using the growth data on beef calves in Section 5.4.1. The residual and additive genetic covariance matrices were:

$$\mathbf{R} = \begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 20 & 18 \\ 18 & 40 \end{bmatrix}$$

Now carry out a Cholesky decomposition of **R** such that $\mathbf{R} = \mathbf{T}\mathbf{T}'$. For the **R** above:

$$\mathbf{T} = \begin{bmatrix} 6.324555 & 0.000 \\ 1.739253 & 5.193746 \end{bmatrix} \quad \text{with} \quad \mathbf{T}^{-1} = \begin{bmatrix} 0.1581139 & 0.000 \\ -0.052948 & 0.1925393 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

The transformed additive genetic covariance matrix (**M**) is:

$$\mathbf{M} = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})' = \begin{bmatrix} 0.5000 & 0.380539 \\ 0.380539 & 1.171972 \end{bmatrix} \quad \text{and} \quad \mathbf{M}^{-1} = \begin{bmatrix} 2.654723 & -0.862556 \\ -0.862556 & 1.133334 \end{bmatrix}$$

The transformed variables are calculated using the transforming matrix $\mathbf{T}^{-1}$. For the first two animals the transformation is as follows:
Animal 1:

$$y^*_{11} = t^{11}y_{11} = 0.1581139(4.5) = 0.712$$

Animal 2:

$$y^*_{11} = t^{11}y_{11} = 0.1581139(2.9) = 0.459$$
$$y^*_{22} = t^{21}y_{11} + t^{22}y_{22} = -0.052948(2.9) + 0.1925393(5.0) = 0.809$$

where y_{ij} and y_{ij}^* are the original and transformed observations, respectively, for the ith trait and jth animal. The transformed variables for all calves are shown in the table below.

				Original traits		Transformed traits	
Calves	Sex	Sire	Dam	WWG	PWG	y_1^*	y_2^*
4	Male	1	–	4.5	–	0.712	–
5	Female	3	2	2.9	5.0	0.459	0.809
6	Female	1	2	3.9	6.8	0.617	1.103
7	Male	4	5	3.5	6.0	0.553	0.970
8	Male	3	6	5.0	7.5	0.791	1.179
9	Female	7	–	4.0	–	0.632	–

The model for analysis is the same as in Section 5.4.1 except that the variance of $\mathbf{y}^*$ now is:

$$\text{var}(\mathbf{y}^*) = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})' + \mathbf{I} = \mathbf{M} + \mathbf{I}$$

The MME for the transformed variables are:

$$\begin{bmatrix} \hat{\mathbf{b}}_1^* \\ \hat{\mathbf{b}}_2^* \\ \hat{\mathbf{a}}_1^* \\ \hat{\mathbf{a}}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & 0 & \mathbf{X}_1'\mathbf{Z}_1 & 0 \\ 0 & \mathbf{X}_2'\mathbf{X}_2 & 0 & \mathbf{X}_2'\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{X}_1 & 0 & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}m^{11} & \mathbf{A}^{-1}m^{12} \\ 0 & \mathbf{Z}_2'\mathbf{X}_2 & \mathbf{A}^{-1}m^{21} & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}^{-1}m^{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1^* \\ \mathbf{X}_2'\mathbf{y}_2^* \\ \mathbf{Z}_1'\mathbf{y}_1^* \\ \mathbf{Z}_2'\mathbf{y}_2^* \end{bmatrix}$$

The design matrices $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and the inverse of the relationship matrix are exactly as in Section 5.4.1. The vector observations $\mathbf{y}^*$ now contain the transformed variables shown in the above table. All other matrices in the MME above can be derived from the design matrices and vector of observations through matrix multiplication and the addition of the $\mathbf{A}^{-1}m^{11}$ and $\mathbf{A}^{-1}m^{22}$ to the animal equations for trait one and two, respectively, and $\mathbf{A}^{-1}m^{12}$ to animal equations for trait one by trait two and $\mathbf{A}^{-1}m^{21}$ to equations for trait one by trait two that pertains to animals. The MME have not been shown because they are too large. However, solving the MME gives

Downloaded From Vet-eBooks.com

the following solutions on the transformed scale. The solutions transformed to the original scale are also shown.

	Transformed scale		Original scale	
Effects	WWG	PWG	WWG	PWG
Sex				
Male	0.691	1.085	4.367	6.834
Female	0.578	0.963	3.657	6.007
Animals				
1	0.021	0.044	0.130	0.266
2	–0.013	–0.010	–0.084	–0.075
3	–0.015	–0.032	–0.098	–0.194
4	0.001	0.003	0.007	0.016
5	–0.054	–0.089	–0.343	–0.555
6	0.030	0.075	0.192	0.440
7	–0.049	–0.077	–0.308	–0.483
8	0.032	0.056	0.201	0.349
9	–0.003	–0.022	–0.018	–0.119

These are exactly the same solutions as those obtained in Section 5.3 without any transformation. The number of non-zero elements was 188 in the analysis on the transformed variables, compared with 208 when no transformation is carried out. This difference could be substantial with large data sets and reduces storage requirements when data is transformed. The solutions were transformed to the original scale using Eqns 6.6 and 6.7. Thus the solutions for male calves on the original scale are:

$$\begin{bmatrix}\hat{b}_{11}\\ \hat{b}_{12}\end{bmatrix}=\begin{bmatrix}6.324555 & 0.000\\ 1.739253 & 5.193746\end{bmatrix}\begin{bmatrix}0.69063\\ 1.0846\end{bmatrix}=\begin{bmatrix}4.367\\ 6.834\end{bmatrix}$$

6.4 Factor and Principal Component Analysis

In Sections 6.2 and 6.3, the simplification of multivariate analysis using canonical transformation and Cholesky decomposition were discussed. Both approaches involved the transformation of the vector of observations as well as the residual and genetic covariance matrices. However, for multivariate analysis with a large number of traits and with high genetic correlations among the traits, a factorial or principal component analysis might be more appropriate in reducing the dimension of such analysis. Neither of these methods involve the transformation of the vector of observations. The principal component and factor analysis (FA) methods provide efficient means for reducing the rank of the genetic covariance matrix in multivariate analysis, resulting in the substantial sparsity of the MME for genetic evaluation and estimation of genetic parameters (Meyer, 2009). Therefore, both methodologies have attracted considerable attention in multivariate analysis involving many traits for parameter estimation and genetic evaluation (Kirkpatrick and Meyer, 2004; Meyer, 2005, 2007; Tyriseva *et al*., 2011a, 2011b).

FA is mainly concerned with identifying the common factors that give rise to correlations between variables. It assumes that the traits studied are linear combinations of few latent variables, referred to as common factors. Then any variance not explained by these common factors is modelled separately as trait specific, by fitting

Downloaded From Vet-eBooks.com

corresponding specific factors. Since the factors are assumed to be uncorrelated, substantial sparsity of the MME is achieved.

On the other hand, PC aims to identify factors that explain the maximum amount of variation and does not imply any underlying model. The first PC explains the maximum amount of genetic variation in the data and each successive PC explains the maximum amount of the remaining variation. Thus for highly correlated traits, only the leading PC have a practical influence on genetic variation and those with negligible effect can be omitted without reducing the accuracy of estimation. For example, with t traits, k independent principal components ($k \leq t$) can be derived that explain a maximum proportion of the total multivariate system. Similar to the FA, the PC approach requires decomposing the genetic covariance matrix into pertaining matrices of eigenvalues and eigenvectors. The eigenvector or PC can be regarded as a linear combination of the traits and they are independent, while the corresponding eigenvalues gives the variance explained.

6.4.1 Factor analysis

Assume that $\mathbf{w}$ is a vector of n variables with covariance matrix equal to $\mathbf{G}$ and that $\mathbf{w}$ can be modelled as:

$$\mathbf{w} = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{c} + \mathbf{s}$$

where $\boldsymbol{\mu}$ is the vector of means, $\mathbf{c}$ is a vector of common factors of length m, $\mathbf{s}$ is the vector of residuals or specific effects of length n and $\boldsymbol{\Phi}$ is the matrix of order $n \times m$ of the so-called factor loadings. In the most common form of FA, the columns of $\boldsymbol{\Phi}$ are orthogonal, i.e. $\varphi_i\varphi_j = 0$, for $i \neq j$ and thus the elements of $\mathbf{c}$ are uncorrelated and assumed to have unit variance, var($\mathbf{c}$) = $\mathbf{I}$. The columns φ_i are determined as corresponding eigenvectors of $\mathbf{G}$, scaled by the square root of the respective eigenvalues (Meyer, 2009).

Usually $\boldsymbol{\Phi}$ is not unique but is often orthogonally transformed to obtain factor loadings that are more interpretable than those derived from the eigenvectors. The specific effects ($\mathbf{s}$) are assumed to be independently distributed and therefore the variance of $\mathbf{s}$ is a diagonal matrix $\mathbf{S}$ of order n. Therefore:

$$\text{var}(\mathbf{w}) = \mathbf{G}_{FA} = \boldsymbol{\Phi}\boldsymbol{\Phi}' + \mathbf{S} \tag{6.8}$$

The above indicates that all the covariances between the levels of $\mathbf{w}$ are modelled through the common factors while the specific factors account for the additional individual variances of the elements of $\mathbf{w}$. Thus the $n(n + 1)/2$ elements of $\mathbf{G}$ are modelled through the n elements of the specific variances and $m(2n - m + 1)/2$ elements of $\boldsymbol{\Phi}$ and additional $m(m - 1)/2$ of $\boldsymbol{\Phi}$ which is determined by the orthogonal constraints. For example if n is 4 and m = 1, then the 10 elements of $\mathbf{G}$ are modelled by the four elements of $\mathbf{S}$ and the four elements of $\boldsymbol{\Phi}$. FA with a small m thus provides a parsimonious way to model the covariances among a large number of variables. When all the elements of $\mathbf{S}$ are non-zero, then four traits is the minimum number of variables for which imposing an FA structure results in a reduction of the parameters (Meyer, 2009).

Mixed model equations

Assume the following multi-trait linear mixed model in Eqn 5.1 is presented as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \tag{6.9}$$

Downloaded From Vet-eBooks.com

with terms defined as in Eqn 5.1 and MME as in Eqn 5.2. If **G** is represented by an FA structure (Eqn 6.8), then an equivalent model to Eqn 6.9 is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}(\mathbf{I}q \times \mathbf{\Phi})\mathbf{c} + \mathbf{Ws} + \mathbf{e} = \mathbf{Xb} + \mathbf{Z}^*\mathbf{c} + \mathbf{Ws} + \mathbf{e} \tag{6.10}$$

with q being the number of individuals, **c** is a vector of common factor effects of order m, $\mathbf{Z}^* = \mathbf{Z}(\mathbf{I}q \times \mathbf{\Phi})$, and **s** is the vector for the specific factor effects. In some contexts, application of Eqn 6.10, i.e. with elements of $\mathbf{S} \neq 0$, is referred to as the extended factor analysis (XFA) compared with models with no specific effects ($\mathbf{S} = 0$), which is simply referred to as factor analysis (FA). The MME for XFA then are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^* & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}^{*\prime}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{*\prime}\mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{I}_m \otimes \mathbf{A}^{-1} & \mathbf{Z}^{*\prime}\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z}^* & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{S}^{-1} \otimes \mathbf{A}^{-1} \end{bmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{c}} \\ \hat{\mathbf{s}} \end{pmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^{*\prime}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{6.11}$$

The vector $\hat{\mathbf{a}}_i$ of solutions for animal i can be obtained as:

$$\hat{\mathbf{a}}_i = \mathbf{\Phi}\hat{\mathbf{c}}_i + \hat{\mathbf{s}}_i \tag{6.12}$$

The number of equations in the MME (Eqn 5.2) for the usual multivariate model are equal to the number of equations for **b** and **s** in Eqn 6.11. However, there are an additional mq equations for the common effects and $\mathbf{Z}^*$, which is a vector of order m with elements φ_{ij}, is denser than **Z** in Eqn 5.2, which contains a single element of unity in a row or column. However, the section of the coefficient matrix for random effects is much sparser as effects are genetically uncorrelated and $\mathbf{A}^{-1}$ contributes only $(m + n)$ non-zero elements compared to n^2 for Eqn 5.1. For the estimation of covariance estimates using REML, Thompson *et al.* (2003) showed that the sparsity of the MME with an XFA structure imposed dramatically reduced computational requirements compared to the standard multivariate model. Note that fitting an FA structure to **G** with no specific effects, the MME are similar to Eqn 6.11 but with the row of equations for **s** omitted, and the $\mathbf{Z}^*$ will be a vector of order n.

An illustration

Example 6.3

The data on pre-weaning weight gain (WWG) and post-weaning gain (PWG) in Example 5.1 is extended to include two additional traits of muscle score (MS) and backfat thickness (BFAT), and data is presented below. The objective is to undertake multi-trait analysis imposing an XFA on **G** and the results obtained compared to those from full MBLUP or FA structure on **G** with no specific factors.

Calf	Sex	WWG	PWG	MS	BFAT
4	Male	4.5	6.8	5.0	0.226
5	Female	2.9	5.0	3.0	0.573
6	Female	3.9	6.8	12.0	0.386
7	Male	3.5	6.0	8.0	0.290
8	Male	5.0	7.5	15.0	0.175

Downloaded From Vet-eBooks.com

Assuming that **G** and **R**, respectively, are:

$$\mathbf{G} = \begin{pmatrix} 20 & 18 & 4 & 9 \\ 18 & 40 & 9 & 20 \\ 4 & 9 & 25 & 4.5 \\ 9 & 20 & 4.5 & 32 \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} 40 & 11 & 16 & 9 \\ 11 & 30 & 12 & 14 \\ 16 & 12 & 70 & 10 \\ 9 & 14 & 10 & 55 \end{pmatrix}$$

Applying Eqn 6.8 to **G** using the function factanal in the **R** package (The R Development Core Team, 2010) gives:

$\mathbf{\Phi}'$ = (2.8532 6.3056 1.4250 3.1678) and $\mathbf{S}'$ = (11.860 0.200 22.975 21.952)

This implies that the number of common factors, m, is equal to 1 for the example **G** above. Thus the column vector $\mathbf{z}_i^*$ for animal i in the matrix $\mathbf{Z}^*$ in Eqn 6.11 equals $\mathbf{\Phi}$. Therefore, for animal i with a record, $\mathbf{z}_i^{*\prime}\mathrm{r}^{-1}\mathbf{z}_i^*$ is 1.361. However, for animal i, $\mathbf{W}_i$ is a diagonal matrix and therefore $\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_i$ is computed as described for the MBLUP model in Section 5.2. Thus for animal i, $\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_i$ is:

$$\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_i = \begin{pmatrix} 0.0297 & & & \\ -0.0079 & 0.0419 & \text{symmetric} & \\ -0.0052 & -0.0041 & 0.0163 & \\ -0.0019 & -0.0086 & -0.0011 & 0.0209 \end{pmatrix}$$

Although there were 48 equations in the MME defined in Eqn 6.11 for this example compared with 40 in the usual MBLUP, there were only 502 non-zero elements in the XFA compared with 620 in MBLUP, illustrating the increased sparsity of the MME with the XFA model. Solving the MME gave the following solutions. The results from the usual MBLUP gave exactly the same solutions and these have not been presented.

Solutions for sex of calf effects

	WWG	PWG	MSC	BFAT
M	4.352	6.795	9.412	0.231
F	3.487	5.959	7.095	0.535

Animal and specific solutions

		Specific effects solutions				Transformed solutions[b]			
	COM[a]	WWG	PWG	BFAT	MSC	WWG	PWG	MSC	BFAT
1	0.036	–0.008	0.095	0.000	0.005	0.095	0.227	0.340	0.010
2	–0.012	–0.001	–0.073	0.021	0.000	–0.089	–0.073	0.313	–0.050
3	–0.027	0.068	0.031	0.208	0.000	–0.086	–0.169	0.031	–0.032
4	0.021	0.046	0.113	–0.021	0.000	0.168	0.136	–0.855	0.113
5	–0.064	–0.191	0.000	0.005	0.000	–0.191	–0.407	–0.539	–0.029
6	0.046	0.290	0.021	0.000	0.000	0.017	0.290	1.350	–0.082
7	–0.063	–0.813	0.208	0.000	0.029	–0.208	–0.399	–0.813	–0.015
8	0.028	–0.101	–0.021	0.000	0.000	–0.017	0.178	1.431	–0.101

[a]COM, solutions for common factor. [b]Transformed solutions from Eqn 6.12.

Downloaded From Vet-eBooks.com

Analysis with FA model

The main differences with fitting an FA model with no specific effects is that $\mathbf{Z}^*$ for an animal in Eqn 6.11 is of order n by n and the last row of Eqn 6.11 is omitted. Note that $\mathbf{Z}^*$ is now a product of the eigenvectors of **G** and the square root of a diagonal matrix of eigenvalues (see Section 6.4.2). Thus $\mathbf{Z}^*$ is:

$$\mathbf{Z}_i^* = \begin{pmatrix} 2.2974 & 3.2100 & -0.1259 & -2.0983 \\ -1.7761 & 5.8683 & -0.1348 & -1.5333 \\ 0.2453 & 2.0018 & 4.4348 & 1.1272 \\ 0.5878 & 4.3506 & -1.7658 & 3.0977 \end{pmatrix} \quad \text{and for animal } i\text{:}$$

$$\mathbf{Z}_i^{*\prime}\mathbf{R}^{-1}\mathbf{Z}_i^* = \begin{pmatrix} 0.372 & -0.226 & -0.042 & 0.054 \\ -0.226 & 1.236 & -0.141 & -0.246 \\ -0.042 & -0.141 & 0.410 & 0.019 \\ 0.054 & -0.246 & 0.019 & 0.537 \end{pmatrix}$$

Setting the MME follows the usual rules and the MME has 40 equations for the example but with only 388 non-zero elements. The low number of non-zero elements is due to the fact that only n elements of A^{-1} are contributed compared with n^2 for the MBLUP. Solving the equations gives the following solutions.

Solutions for calf sex effects

	WWG	PWG	MSC	BFAT
M	4.352	6.795	9.412	0.231
F	3.487	5.959	7.095	0.535

Animal solutions

	Untransformed solutions				Transformed solutions[a]			
	WWG	PWG	BFAT	MSC	WWG	PWG	MSC	BFAT
1	–0.011	0.035	0.063	–0.008	0.095	0.227	0.340	0.010
2	–0.003	–0.005	0.066	0.028	–0.089	–0.073	0.313	–0.050
3	0.010	–0.020	0.010	0.021	–0.086	–0.169	0.031	–0.032
4	0.000	0.002	–0.177	–0.067	0.168	0.136	–0.855	0.113
5	0.015	–0.062	–0.099	0.019	–0.191	–0.407	–0.539	–0.029
6	–0.022	0.061	0.267	0.045	0.017	0.290	1.350	–0.082
7	0.003	–0.069	–0.153	0.005	–0.208	–0.399	–0.813	–0.015
8	–0.007	0.050	0.285	0.060	–0.017	0.178	1.431	–0.101

[a]Transformed solutions = vectors of solutions multiplied by $\mathbf{Z}^*$

6.4.2 Principal component analysis

Analysis with full PC model

The application of a full PC model with no rank reduction is similar to the FA analysis except that $\mathbf{Z}^*$ is now a matrix of eigenvectors of order n by n and $\mathbf{Z}^{*\prime}\mathbf{R}^{-1}\mathbf{Z}^* + (\mathbf{I}_m \otimes \mathbf{A}^{-1})$

Downloaded From Vet-eBooks.com

in the second row of Eqn 6.11 is replaced by $\mathbf{Z}^{*\prime}\mathbf{R}^{-1}\mathbf{Z}^{*} + (\mathbf{D}_n \otimes \mathbf{A}^{-1})$, where $\mathbf{D}_n$ is a diagonal matrix of eigenvalues. Again, the last row of Eqn 6.11 is omitted. It therefore involves decomposing **G** to a matrix of eigenvectors ($\mathbf{Z}^{*}$) and corresponding eigenvalues (**D**). Thus $\mathbf{Z}^{*}$ and **D**, respectively, are:

$$\mathbf{Z}^{*} = \begin{pmatrix} 0.7710 & 0.3896 & -0.02940 & -0.5029 \\ -0.5983 & 0.7139 & -0.0268 & -0.3628 \\ 0.0865 & 0.2427 & 0.9288 & 0.2664 \\ 0.2000 & 0.5288 & -0.3685 & 0.7379 \end{pmatrix} \quad \text{and}$$

$$\mathbf{D} = \text{diag}(8.8159\ 67.6963\ 22.8286\ 17.6592)$$

Thus $\mathbf{Z}_i^{*\prime}\mathbf{R}^{-1}\mathbf{Z}_i^{*}$ for animal *i* is:

$$\mathbf{Z}_i^{*\prime}\mathbf{R}^{-1}\mathbf{Z}_i^{*} = \begin{pmatrix} 0.042 & -0.009 & -0.003 & 0.004 \\ -0.009 & 0.018 & -0.004 & -0.007 \\ -0.003 & -0.004 & 0.018 & 0.001 \\ 0.004 & -0.007 & 0.001 & 0.030 \end{pmatrix}$$

The MME are set up as usual. Similar again to the FA model, the PC has 40 equations and 388 non-zero elements. The solutions for the various effects from solving the MME are:

Solutions for sex of calf effects

	WWG	PWG	MSC	BFAT
M	4.352	6.795	9.412	0.231
F	3.488	5.959	7.095	0.535

Animal solutions

	Untransformed solutions				Transformed solutions			
	WWG	PWG	BFAT	MSC	WWG	PWG	MSC	BFAT
1	–0.032	0.287	0.303	–0.032	0.094	0.227	0.340	0.010
2	–0.009	–0.038	0.314	0.118	–0.090	–0.073	0.313	–0.050
3	0.031	–0.163	0.047	0.089	–0.086	–0.169	0.030	–0.032
4	–0.002	0.015	–0.844	–0.279	0.170	0.136	–0.855	0.113
5	0.045	–0.511	–0.473	0.078	–0.190	–0.407	–0.539	–0.029
6	–0.062	0.496	1.276	0.186	0.014	0.290	1.350	–0.083
7	0.007	–0.571	–0.732	0.022	–0.207	–0.400	–0.812	–0.015
8	–0.018	0.413	1.362	0.252	–0.019	0.178	1.431	–0.101

6.4.3 Analysis with reduced rank PC model

The diagonal matrix **D** with the full PC model in Section 6.4.2 indicates that the first principal component accounts for about 8.82% of the total genetic variance. Deleting the first

Downloaded From Vet-eBooks.com

eigenvalue gives a diagonal $\mathbf{D}^*$ of order 3 as $\mathbf{D}^*$= diag(67.6963 22.8286 17.6592). Then $\mathbf{G}^*$, a new genetic covariance matrix, can be computed as $\mathbf{M}'\mathbf{D}^*\mathbf{M}$, where $\mathbf{M}$ is equivalent to $\mathbf{Z}^*$ in Section 6.4.2 with a full PC model fitted but with the first column deleted. Thus:

$$\mathbf{G}^* = \mathbf{M}'\mathbf{D}^*\mathbf{M} = \begin{pmatrix} 14.759 & 22.067 & 3.412 & 7.640 \\ 22.067 & 36.844 & 9.456 & 21.055 \\ 3.412 & 9.456 & 24.934 & 4.347 \\ 7.640 & 21.055 & 4.347 & 31.647 \end{pmatrix} \quad \text{with}$$

$$\mathbf{M} = \begin{pmatrix} 0.3896 & -0.02940 & -0.5029 \\ 0.7139 & -0.0268 & -0.3628 \\ 0.2427 & 0.9288 & 0.2664 \\ 0.5288 & -0.3685 & 0.7379 \end{pmatrix}$$

The application of reduced rank PC is similar to the full PC analysis with $\mathbf{Z}^*$ replaced by $\mathbf{M}$ and $\mathbf{D}$ by $\mathbf{D}^*$. Thus for animal i, $\mathbf{M}_i'\mathbf{R}^{-1}\mathbf{M}_i$ is:

$$\mathbf{M}_i'\mathbf{R}^{-1}\mathbf{M}_i = \begin{pmatrix} 0.018 & -0.004 & -0.007 \\ -0.004 & 0.018 & 0.001 \\ -0.007 & 0.001 & 0.030 \end{pmatrix}$$

The MME for the reduced PC has 32 equations and 284 non-zero elements. The solutions for the various effects from solving the MME are:

Solutions for sex of calf effects

	WWG	PWG	MSC	BFAT
M	4.349	6.798	9.412	0.230
F	3.480	5.963	7.093	0.533

Solutions for animal effects

	Untransformed solutions			Transformed solutions[a]			
				WWG	PWG	MSC	BFAT
1	0.295	0.305	–0.033	0.123	0.214	0.346	0.019
2	–0.037	0.314	0.118	–0.083	–0.078	0.314	–0.048
3	–0.170	0.046	0.090	–0.113	–0.156	0.025	–0.041
4	0.017	–0.844	–0.279	0.171	0.136	–0.854	0.115
5	–0.523	–0.476	0.080	–0.230	–0.390	–0.548	–0.042
6	0.511	1.279	0.185	0.069	0.263	1.362	–0.066
7	–0.576	–0.734	0.022	–0.215	–0.400	–0.816	–0.018
8	0.419	1.364	0.251	–0.003	0.171	1.435	–0.096

[a]Transformed solutions = vector of solutions multiplied by **M**

The deletion of the first eigenvalue in the reduced PC analysis had very little effect in terms of the EBVs of animals for traits 3 and 4. Thus there was no ranking for MSC

Downloaded From Vet-eBooks.com

and only two animals swapped places for BFAT compared with the result from the full PC analysis. However, only the top four and six animals were the same for WWG and PWG, respectively, compared with the full PC analysis, indicating more re-ranking was observed in WWG due to the reduction in variance. In practice, models with reduced ranks are usually applied in the analysis of many traits as in Meyer (2007), resulting in no re-ranking in the top animals, which are mainly of interest.

Downloaded From Vet-eBooks.com

7 Maternal Trait Models: Animal and Reduced Animal Models

7.1 Introduction

The phenotypic expression of some traits in the progeny, such as weaning weight in beef cattle, is influenced by the ability of the dam to provide a suitable environment in the form of better nourishment. Thus the dam contributes to the performance of the progeny in two ways: first, through her direct genetic effects passed to the progeny and second, through her ability to provide a suitable environment, for instance in producing milk. Traits such as birth and weaning weights in beef cattle fall into this category and are termed maternally influenced traits. The ability of the dam to provide a suitable environment for the expression of such traits in her progeny is partly genetic and partly environmental. Similar to the genetic component of an individual, the maternal genetic component can be partitioned into additive, dominance and epistatic effects (Willham, 1963). The environmental part may be partitioned into permanent and temporary environmental components. It is the maternal additive genetic component of the dam that is passed on to all her offspring, but it is expressed only when the female offspring have progeny of their own.

In the usual mixed linear model for maternally influenced traits (Eqn 7.1) the phenotype is partitioned into:

1. Additive genetic effects from the sire and the dam, usually termed the direct genetic effect.
2. Additive genetic ability of the dam to provide a suitable environment, usually termed the indirect or maternal genetic effect.
3. Permanent environmental effects, which include permanent environmental influences on the dam's mothering ability and the maternal non-additive genetic effects of the dam.
4. Other random environmental effects, termed residual effects.

In this chapter, the mixed model methodology for genetic evaluation in models with maternal effects is discussed, considering a univariate situation, and the extension to multivariate analysis is also briefly presented. The application of BLUP to models with maternal effects was first presented by Quaas and Pollak (1980).

When repeated measurements for maternally influenced traits are available over a range of ages (for instance, body weight from birth to 630 days), a random regression model (see Chapter 9) might be more appropriate to analyse such a trait. A random regression model for maternally influenced traits is briefly defined in Section 9.3.6.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

7.2 Animal Model for a Maternal Trait

The model for maternally influenced traits in matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Spe} + \mathbf{e} \tag{7.1}$$

where **y** = vector of observations, **b** = vector of fixed effects, **u** = vector of random animal effects, **m** = vector of random maternal (indirect) genetic effects, **pe** = vector of permanent environmental effects as explained in item **3** in Section 7.1, **e** = vector of random residual effects, and **X**, **Z**, **W** and **S** are incidence matrices relating records to fixed, animal, maternal genetic and permanent environmental effects, respectively. It is assumed that:

$$\text{var}\begin{bmatrix}\mathbf{u}\\ \mathbf{m}\\ \mathbf{pe}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & 0 & 0\\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & 0 & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{pe} & 0\\ 0 & 0 & 0 & \mathbf{I}\sigma^2_{e}\end{bmatrix}$$

where g_{11} = additive genetic variance for direct effects, g_{22} = additive genetic variance for maternal effects, g_{12} = additive genetic covariance between direct and maternal effects, σ^2_{pe} = variance due to permanent environmental effects and σ^2_e = residual error variance.

The variance of **y**, using the same arguments as in Section 3.2, is:

$$\text{var}(\mathbf{y}) = \begin{bmatrix}\mathbf{Z} & \mathbf{W}\end{bmatrix}\begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A}\\ g_{21}\mathbf{A} & g_{22}\mathbf{A}\end{bmatrix}\begin{bmatrix}\mathbf{Z}'\\ \mathbf{W}'\end{bmatrix} + \mathbf{SI}\sigma^2_{pe}\,\mathbf{S}' + \mathbf{I}\sigma^2_e$$

The BLUE of estimable functions of **b** and the BLUP of **u**, **m** and **pe** in Eqn 7.1 are obtained by solving the following MME:

$$\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{u}}\\ \hat{\mathbf{m}}\\ \hat{\mathbf{pe}}\end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} & \mathbf{X}'\mathbf{S}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\mathbf{A}^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W}+\mathbf{A}^{-1}\alpha_2 & \mathbf{Z}'\mathbf{S}\\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z}+\mathbf{A}^{-1}\alpha_2 & \mathbf{W}'\mathbf{W}+\mathbf{A}^{-1}\alpha_3 & \mathbf{W}'\mathbf{S}\\ \mathbf{S}'\mathbf{X} & \mathbf{S}'\mathbf{Z} & \mathbf{S}'\mathbf{W} & \mathbf{S}'\mathbf{S}+\mathbf{I}\alpha_4\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\\ \mathbf{W}'\mathbf{y}\\ \mathbf{S}'\mathbf{y}\end{bmatrix} \tag{7.2}$$

$$\text{with } \mathbf{G} = \begin{bmatrix} g_{11} & g_{12}\\ g_{21} & g_{22}\end{bmatrix};\ \mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12}\\ g^{21} & g^{22}\end{bmatrix} \text{ and } \begin{bmatrix}\alpha_1 & \alpha_2\\ \alpha_2 & \alpha_3\end{bmatrix} = \sigma^2_e\begin{bmatrix} g^{11} & g^{12}\\ g^{21} & g^{22}\end{bmatrix} \text{ and } \alpha_4 = \sigma^2_e/\sigma^2_{pe}$$

7.2.1 An illustration

Example 7.1

Assume the data in Table 7.1 to be the birth weight for a group of beef calves. The aim is to estimate solutions for herd and pen effects and predict solutions for direct and maternal effects for all animals and permanent environmental effects for dams of

Downloaded From Vet-eBooks.com

Table 7.1. Birth weight for group of beef calves.

Calf	Sire	Dam	Herds	Pen	Birth weight (kg)
5	1	2	1	1	35.0
6	3	2	1	2	20.0
7	4	6	1	2	25.0
8	3	5	1	1	40.0
9	1	6	2	1	42.0
10	3	2	2	2	22.0
11	3	7	2	2	35.0
12	8	7	3	2	34.0
13	9	2	3	1	20.0
14	3	6	3	2	40.0

progeny with records. Suppose that the genetic parameters are $g_{11} = 150$, $g_{12} = -40$, $g_{22} = 90$, $\sigma^2_{pe} = 40$ and $\sigma^2_e = 350$. Then:

$$\mathbf{G}^{-1} = \begin{bmatrix} 0.00756 & 0.00336 \\ 0.00336 & 0.0126 \end{bmatrix} \text{ and } \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} = \begin{bmatrix} 2.647 & 1.176 \\ 1.176 & 4.412 \end{bmatrix}$$

and $\alpha_4 = 350/40 = 8.75$.

The model for the analysis is as presented in Eqn 7.1.

Setting up the design matrices

Considering only animals with records, the first three rows of matrix **X** relate records to herd effects and the last two rows to pen effects. The transpose of **X** is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Excluding ancestors, each animal has one record; therefore **Z** is an identity matrix. However, **Z** is augmented with columns of zeros equal to the number of ancestors to take account of ancestors in the pedigree. The matrices **W** and **S** relate records through the dam to their effects, i.e. maternal genetic effect and permanent environmental effect, respectively. However, since maternal effect is genetic and is passed from parent to offspring, estimates of maternal effect are for all animals in the analysis while estimates of permanent environmental effects are only for dams of progeny with records. Thus, in setting up **W**, all animals are considered, while only four dams with progeny having records are taken into account for **S**.

Downloaded From Vet-eBooks.com

For the example data set, **W** (with rows and columns numbered by the relevant animal they relate to) is:

$$\mathbf{W} = \begin{array}{c} \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \mathbf{0} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

and:

$$\mathbf{S}' = \begin{array}{c} \\ 2 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \mathbf{0} & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

The matrix **S** above implies, for instance, that animals 5, 6, 10 and 13 have the same dam (animal 2), while animals 11 and 12 are from another dam (animal 7).

The transpose of the vector of observations is:

$\mathbf{y}'$ = [35 20 25 40 42 22 35 34 20 40]

The other matrices in the MME can be calculated through matrix multiplication. The inverse of the relationship matrix is calculated applying the rules in Section 2.4.1. The matrix $\mathbf{A}^{-1}\alpha_1$ is added to animal equations, $\mathbf{A}^{-1}\alpha_2$ to the equations for maternal genetic effects, $\mathbf{A}^{-1}\alpha_3$ to the animal by maternal genetic equations and α_4 to the diagonals of the equations for permanent environmental effects to obtain the MME. The MME are not presented because they are too large. There is dependency between the equations for herds and pen; thus the row for the first herd was set to zero in solving the MME by direct inversion. Solutions to the MME are:

Effects	Solutions
Herd–year–season	
1	0.000
2	3.386
3	1.434
Pen	
1	34.540
2	27.691

Continued

Downloaded From Vet-eBooks.com

(*Continued*)

Effects	Solutions	
Animals		
	Direct effects	Maternal effects
1	0.564	0.262
2	–1.244	–1.583
3	1.165	0.736
4	–0.484	0.586
5	0.630	–0.507
6	–0.859	0.841
7	–1.156	1.299
8	1.917	–0.158
9	–0.553	0.660
10	–1.055	–0.153
11	0.385	0.916
12	0.863	0.442
13	–2.980	0.093
14	1.751	0.362
Permanent environment		
2	–1.701	
5	0.415	
6	0.825	
7	0.461	

The solutions show little difference between the herds, but calves in pen 1 were heavier than those in pen 2 by about 6.85 kg at birth. The solution for level *i* of the fixed effect *n* can be calculated using Eqn 4.3 except that the sum of yields for the level of fixed effect is corrected in addition for maternal effects. That is:

$$\hat{b}_{in} = \frac{\sum_{f=1}^{\text{diag}_{in}} y_{inf} - \sum_{j} \hat{b}_{inj} - \sum_{k} \hat{a}_{ink} - \sum_{l} \hat{m}_{inl} - \sum_{t} \hat{p}e_{int}}{\text{diag}_{in}} \tag{7.3}$$

where m_{inl} is the solution for level *l* of genetic maternal effects within level *i* of the *n*th fixed effect and all other terms are as defined in Eqn 4.3. Thus the solution for level 1 of pen effect is:

$$\begin{aligned}\hat{b}_{11} &= [137 - (2\hat{h}d_1 + \hat{h}d_2 + \hat{h}d_3) - (\hat{a}_5 + \hat{a}_8 + \hat{a}_9 + \hat{a}_{13}) \\ &\quad - (2\hat{m}_2 + \hat{m}_5 + \hat{m}_6) - (2\hat{p}e_2 + \hat{p}e_5 + \hat{p}e_6)]/4 \\ &= [137 - 4.82 - (-0.986) - (-2.832) - (-2.162)]/4 \\ &= 34.540\end{aligned}$$

where $\hat{h}d_j$ is the solution for level *j* of herd effect.

From the MME, the solutions for direct and maternal effects for animal *i* with progeny *o* are:

$$\begin{bmatrix}\hat{u}_i \\ \hat{m}_i\end{bmatrix} = \begin{bmatrix}n_1 + (d + k_1)\alpha_1 & (d + k_1)\alpha_2 \\ (d + k_1)\alpha_2 & n_2 + (d + k_1)\alpha_3\end{bmatrix}^{-1} \mathbf{H}k_2 \begin{bmatrix}\hat{u}_s + \hat{u}_d \\ \hat{m}_s + \hat{m}_d\end{bmatrix}$$
$$+ \begin{bmatrix}y_i - \hat{b}_i - \hat{m}_{dam} - \hat{p}_{dam} \\ y_o - \hat{b}_o - \hat{u}_o - \hat{p}_i\end{bmatrix} + \mathbf{H}k_3 \begin{bmatrix}\hat{a}_o - 0.5(\hat{a}_{mate}) \\ \hat{m}_o - 0.5(\hat{m}_{mate})\end{bmatrix} \tag{7.4}$$

Downloaded From Vet-eBooks.com

where n_1 is the number of records for animal i; n_2 is the number of progeny records with animal i as the dam; d = 2, $\frac{4}{3}$ or 1 when both, one or no parents of animal i are known; k_2 = 1 or $\frac{2}{3}$ when both or one parent of animal i are known; $k_1 = \frac{1}{2}$ and k_3 = 1 when the mate of animal i is known or $k_1 = \frac{1}{3}$ and $k_3 = \frac{2}{3}$ with the mate unknown and:

$$\mathbf{H} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix}$$

For instance, the solutions for direct and genetic maternal effects for animal 5 are:

$$\begin{bmatrix} \hat{u}_5 \\ \hat{m}_5 \end{bmatrix} = \begin{bmatrix} 1+(2+0.5)2.647 & (2+0.5)1.176 \\ (2+0.5)1.176 & 1+(2+0.5)4.412 \end{bmatrix}^{-1} \mathbf{H}k_2 \begin{bmatrix} u_1+u_2 \\ m_2+m_2 \end{bmatrix} + \begin{bmatrix} y_5-\hat{b}_1-\hat{m}_2-\hat{p}_2 \\ y_8-\hat{b}_1-\hat{u}_8-\hat{p}_5 \end{bmatrix}$$
$$+ \mathbf{H}k_3 \begin{bmatrix} \hat{a}_8 - 0.5(\hat{a}_3) \\ \hat{m}_8 - 0.5(\hat{m}_3) \end{bmatrix}$$
$$\begin{bmatrix} \hat{u}_5 \\ \hat{m}_5 \end{bmatrix} = \begin{bmatrix} 1+(2+0.5)2.647 & (2+0.5)1.176 \\ (2+0.5)1.176 & 1+(2+0.5)4.412 \end{bmatrix}^{-1} \mathbf{H}(1) \begin{bmatrix} 0.564+(-1.244) \\ 0.262+-(1.583) \end{bmatrix}$$
$$+ \begin{bmatrix} 35-0-34.54-(-1.583)-(-1.701) \\ 40-0-34.54-1.917-0.415 \end{bmatrix}$$
$$+ \mathbf{H}(1) \begin{bmatrix} 1.917-0.5(1.165) \\ -0.158-0.5(0.736) \end{bmatrix} = \begin{bmatrix} 0.630 \\ -0.507 \end{bmatrix}$$

The solution for the permanent environmental effect for dam j from the MME is:

$$\hat{p}e_j = (y_o - \hat{b}_o - \hat{u}_o - \hat{m}_j)/(n_2 + \alpha_3) \quad (7.5)$$

where all terms are as defined in Eqn 7.4. For animal 5, the solution for the permanent environmental effect is:

$$\hat{p}e_5 = 40 - 0 - 34.54 - 1.917 - (-0.507)/(1 + 8.75) = 0.415$$

Additive genetic maternal effects represent good mothering ability, which is passed on from dams to progeny, while permanent environment effects refer to permanent environmental and maternal non-additive genetic influences on the mothering ability of the dam. Thus selection of dams for the next generation in a maternal line would place emphasis on good genetic maternal effects in addition to a good estimate of breeding value. If equal emphasis is placed on both effects, dams 7 and 5 would be the top two dams in the example while dam 2 ranks lowest. However, if the main interest is the performance of the future dams in the same herd, then selection of dams would be based on some combination of the solutions for direct, maternal genetic and permanent environmental effects for the dams. Again, in the example data, dam 2 ranks lowest while the best two dams are dams 6 and 7 if equal emphasis is placed on the three components.

In the case of males, the selection of sires for a maternal line, for instance, would be based on a combination of solutions for direct and maternal genetic effects. Obviously, sires 3 and 1 would be the top two bulls for such a purpose. However, if the emphasis is only on direct genetic effects, probably to breed a bull, then sire 8 in the example would be the bull of choice.

Downloaded From Vet-eBooks.com

7.3 Reduced Animal Model with Maternal Effects

In Section 3.5, the use of the reduced animal model (RAM), with only one random effect apart from residual error in the model, was considered. The records of non-parents in the MME were expressed as the average of parental breeding values plus Mendelian sampling. This has the advantage of reducing the number of random animal equations in the MME. The application of RAM with multiple random effects in the model is illustrated in this section using the example data used for the full animal model in Section 7.2. The model for the analysis is the same but design matrices and the variance of non-parental animals are different. From the arguments in Section 3.5, the model for the RAM can be expressed as:

$$\begin{bmatrix} \mathbf{y}_p \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p \\ \mathbf{X}_n \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{Z}_p \\ \mathbf{Z}_n \end{bmatrix} \mathbf{u}_p + \mathbf{Z}_2 \mathbf{m} + \mathbf{Z}_3 \mathbf{pe} + \begin{bmatrix} \mathbf{e}_p \\ \mathbf{e}_n \end{bmatrix} \tag{7.6}$$

where $\mathbf{y}_p$, $\mathbf{y}_n$ = vector of observations for parent and non-parents, respectively, $\mathbf{b}$ = vector of fixed effects, $\mathbf{u}_p$ = vector of random animal effect for parents, $\mathbf{m}$ = vector of maternal genetic effects for parents, $\mathbf{pe}$ = vector of permanent environmental effects and $\mathbf{e}_p$, $\mathbf{e}_n$ = vector of residual error for parents and non-parents, respectively.

The incidence matrices $\mathbf{Z}_2$ and $\mathbf{Z}_3$ relate records to maternal genetic and permanent environmental effect, respectively. The matrices $\mathbf{Z}_p$ and $\mathbf{X}_p$ relate records of parents to animal and fixed effects, respectively, while $\mathbf{Z}_n$ and $\mathbf{X}_n$ relate records of non-parents to parents (animal effect) and fixed effects, respectively.

It is assumed that:

$$\text{var}\begin{bmatrix} \mathbf{u}_p \\ \mathbf{m} \\ \mathbf{pe} \\ \mathbf{e}_p \\ \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & 0 & 0 & 0 \\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I}\sigma^2_{pe} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\sigma^2_{ep} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I}\sigma^2_{en} \end{bmatrix}$$

where σ^2_{ep} is the residual variance for parents, which is equal to σ^2_e in Section 7.2, σ^2_{en} is the residual variance for non-parents and is equal to $\mathbf{I} + \mathbf{D}g_{11}$, with $\mathbf{D}$ being a diagonal matrix containing elements d_{jj}, which are equal to $\frac{3}{4}$ or $\frac{1}{2}$ depending on whether one or both parents are known. The matrix $\mathbf{G}$ and σ^2_{pe} are defined as in Section 7.2. Let:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_p \\ \mathbf{X}_n \end{bmatrix}, \quad \mathbf{Z}_1 = \begin{bmatrix} \mathbf{Z}_p \\ \mathbf{Z}_n \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \mathbf{I}\sigma^2_{ep} & 0 \\ 0 & \mathbf{I}\sigma^2_{en} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_n & 0 \\ 0 & \mathbf{R}_p \end{bmatrix} \quad \text{and} \quad \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{R}_p^{-1} & 0 \\ 0 & \mathbf{R}_n^{-1} \end{bmatrix}$$

Again, the MME provide the basis of the BLUE of estimable functions of $\mathbf{b}$ and BLUP of $\mathbf{a}$, $\mathbf{m}$ and $\mathbf{pe}$ in Eqn 7.6. The relevant MME are:

Downloaded From Vet-eBooks.com

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{m}} \\ \hat{\mathbf{pe}} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_3 \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1' + \mathbf{A}_p^{-1}g^{11} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{A}_p^{-1}g^{12} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_3 \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1' + \mathbf{A}_p^{-1}g^{21} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{A}_p^{-1}g^{22} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_3 \\ \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_2 & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_3' + \mathbf{I}\,1/\sigma_{pe}^2 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{7.7}$$

where g^{ii} are the elements of the inverse of **G**.

As shown in Section 3.5, each block of equations in the MME above can be expressed as the sum of the contributions from parents' records and non-parents' records. Thus:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \mathbf{X}_p'\ \mathbf{R}_p^{-1}\mathbf{X}_p + \mathbf{X}_n'\ \mathbf{R}_n^{-1}\mathbf{X}_n$$

Expressing Eqn 7.7 as shown for the equations for the block of fixed effects above and multiplying by R_p gives:

$$\begin{bmatrix} \mathbf{X}_p'\ \mathbf{X}_p + \mathbf{X}_n'\ \mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{X}_p'\ \mathbf{Z}_p + \mathbf{X}_n'\ \mathbf{R}_n^{-1}\mathbf{Z}_n & \mathbf{X}_p'\ \mathbf{Z}_2 + \mathbf{X}_n'\ \mathbf{R}_n^{-1}\mathbf{Z}_2 & \mathbf{X}_p'\ \mathbf{Z}_3 + \mathbf{X}_n'\ \mathbf{R}_n^{-1}\mathbf{Z}_3 \\ \mathbf{Z}_p'\ \mathbf{X}_p + \mathbf{Z}_n'\ \mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}_p'\ \mathbf{Z}_p + \mathbf{Z}_n'\ \mathbf{R}_n^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z}_p'\ \mathbf{Z}_2 + \mathbf{Z}_n'\ \mathbf{R}_n^{-1}\mathbf{Z}_2 + \mathbf{A}^{-1}\alpha_2 & \mathbf{Z}_p'\ \mathbf{Z}_3 + \mathbf{Z}_n'\ \mathbf{R}_n^{-1}\mathbf{Z}_3 \\ \mathbf{Z}_2'\ \mathbf{X}_p + \mathbf{Z}_2'\ \mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}_2'\ \mathbf{Z}_p + \mathbf{Z}_2'\ \mathbf{R}_n^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\alpha_2 & \mathbf{Z}_2'\ \mathbf{Z}_2 + \mathbf{Z}_2'\ \mathbf{R}_n^{-1}\mathbf{Z}_2 + \mathbf{A}^{-1}\alpha_3 & \mathbf{Z}_2'\ \mathbf{Z}_3 + \mathbf{Z}_2'\ \mathbf{R}_n^{-1}\mathbf{Z}_3 \\ \mathbf{Z}_3'\ \mathbf{X}_p + \mathbf{Z}_3'\ \mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}_3'\ \mathbf{Z}_p + \mathbf{Z}_3'\ \mathbf{R}_n^{-1}\mathbf{Z}_n & \mathbf{Z}_3'\ \mathbf{Z}_3 + \mathbf{Z}_3'\ \mathbf{R}_n^{-1}\mathbf{Z}_3 & \mathbf{Z}_3'\ \mathbf{Z}_3 + \mathbf{Z}_3'\ \mathbf{R}_n^{-1}\mathbf{Z}_3 + \mathbf{I}\alpha_4 \end{bmatrix}$$
$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{m}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p'\ \mathbf{y}_p + \mathbf{X}_n'\ \mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}_p'\ \mathbf{y}_p + \mathbf{Z}_n'\ \mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}_2'\ \mathbf{y}_p + \mathbf{Z}_2'\ \mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}_3'\ \mathbf{y}_p + \mathbf{Z}_3'\ \mathbf{R}_n^{-1}\mathbf{y}_n \end{bmatrix}$$

The α terms are as defined in Eqn 7.2 and $\mathbf{R}_n^{-1}$ now equals $1/(1 + \mathbf{D}\alpha^{-1})$. The MME for the solutions of **b**, **u**, **m** and **pe** can therefore be set up as shown above or as in Eqn 7.7.

7.3.1 An illustration

Example 7.2

The same data set and genetic parameters as in Section 7.2 are used below to demonstrate the principles for setting up a RAM with maternal effects in the model using Eqn 7.5. Recollect that:

$$\mathbf{G} = \begin{bmatrix} 40 & 10 \\ 10 & 20 \end{bmatrix} \text{ and } \mathbf{G}^{-1} = \begin{bmatrix} 0.029 & -0.014 \\ -0.014 & 0.057 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

The residual variance for parents σ^2_{ep} = 350, and because both parents of non-parents in the data are known:

$$\sigma^2_{en} = \sigma^2_e + \tfrac{1}{2}(g_{11}) = 350 + \tfrac{1}{2}(150) = 425$$

with:

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}\sigma^2_{ep} & 0 \\ 0 & \mathbf{I}\sigma^2_{en} \end{bmatrix}$$

Then:

$$\mathbf{R} = \text{diag}(350, 350, 350, 350, 350, 425, 425, 425, 425, 425)$$

and:

$$\mathbf{R}^{-1} = \text{diag}(0.00286, 0.00286, 0.00286, 0.00286, 0.00286, 0.00235, 0.00235, 0.00235, 0.00235, 0.00235)$$

$$1/\sigma^2_{pe} = \tfrac{1}{40} = 0.025$$

Setting up the design matrices

The matrix **X**, which relates records to fixed effects, is the same as in Section 7.2.1, considering only animals with records. The matrix $\mathbf{X'R^{-1}X}$ in the MME can be calculated through matrix multiplication from **X** and $\mathbf{R}^{-1}$ already set up. For illustrative purposes, the matrix $\mathbf{X'R^{-1}X}$, when expressed as the sum of the contributions from parents' and non-parents' records, is:

$$\mathbf{X'R^{-1}X} = \mathbf{r}^{11}_p \mathbf{X}'_p \mathbf{X}_p + \mathbf{r}^{11}_n \mathbf{X}'_n \mathbf{X}_n$$

$$= \begin{bmatrix} 0.0114 & 0.0 & 0.0 & 0.0057 & 0.0057 \\ 0.0 & 0.0029 & 0.0 & 0.0029 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0057 & 0.0029 & 0.0 & 0.0086 & 0.0 \\ 0.0057 & 0.0 & 0.0 & 0.0 & 0.0057 \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0047 & 0.0 & 0.0 & 0.0047 \\ 0.0 & 0.0 & 0.0071 & 0.0024 & 0.0047 \\ 0.0 & 0.0 & 0.0024 & 0.0024 & 0.0 \\ 0.0 & 0.0047 & 0.0047 & 0.0 & 0.0094 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0114 & 0.0 & 0.0 & 0.0057 & 0.0057 \\ 0.0 & 0.0076 & 0.0 & 0.0029 & 0.0047 \\ 0.0 & 0.0 & 0.0071 & 0.0024 & 0.0047 \\ 0.0057 & 0.0029 & 0.0024 & 0.0109 & 0.0 \\ 0.0057 & 0.0047 & 0.0047 & 0.0 & 0.0151 \end{bmatrix}$$

where $\mathbf{X}_p$ and $\mathbf{X}_n$ are matrices relating parents and non-parents to fixed effects, respectively, and are:

$$\mathbf{X}'_p = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}'_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

The matrix $\mathbf{Z}_1$, which relates records to animal effect is:

$$\mathbf{Z}_1 = \begin{array}{c} \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 \end{bmatrix} \end{array}$$

The first five rows correspond to animals 5 to 9, which are parents, and each has one record. The last five rows correspond to the records for animals 10 to 14 (non-parents), which are related to their parents. The matrices $\mathbf{Z}_2$ and $\mathbf{Z}_3$ are exactly the same as **W** and **S** in Section 7.2.1, respectively, and the vector of observation, **y**, is the same as in Section 7.2.1. Apart from the relationship matrix, all the matrices in the MME can easily be calculated through matrix multiplication from the design matrices and vector of observation set up above. The inverse of the relationship matrix is set up only for parents ($\mathbf{A}_p^{-1}$), i.e. for animals 1 to 9, using the procedure outlined in Chapter 2. The matrix $\mathbf{A}_p^{-1}g^{11}$ is added to animal equations, $\mathbf{A}_p^{-1}g^{22}$ to the equations for maternal genetic effects, $\mathbf{A}_p^{-1}g^{12}$ to the animal by maternal genetic equations, $\mathbf{A}_p^{-1}g^{21}$ to the maternal genetic by animal equations and $1/\sigma^2_{pe}$ to the diagonals of the equations for permanent environmental effects to obtain the MME. The MME are not presented because they are too large. Solving the MME by direct inversion with the equation for the first herd set to zero gives the same solutions as from the animal model (Example 7.1). However, the number of non-zero elements in the coefficient matrix was 329 compared with 429 in the animal model, due to the reduced number of equations, indicating the advantages of the RAM.

Back-solving for non-parents

The solutions for direct animal and maternal effects for non-parents are back-solved after the MME have been solved.

Back-solving for direct effects

Solutions for direct animal effect for the non-parents are obtained from parent average and an estimate of Mendelian sampling using Eqn 3.27. Thus the solution for the non-parent *i* is:

$$\hat{u}_i = 0.5(\hat{u}_s + \hat{u}_d) + k_i(y_i - \hat{b}_j - \hat{m}_d - \hat{p}e_d - 0.5(\hat{u}_s + \hat{u}_d)) \qquad (7.8)$$

with:

$$k_i = r^{-1}/(r^{-1} + d^{-1}g^{-1}) = 1/(1 + d^{-1}\alpha); \quad \alpha = \sigma^2_e/\sigma^2_a$$

Downloaded From Vet-eBooks.com

where d is either $\frac{1}{2}$ if both parents are known or $\frac{3}{4}$ if only one parent is known. For the example data, both parents of the non-parent individuals are known, therefore:

$$k_i = 1/(1 + 2(2.333)) = 0.17647$$

For animal 10, for instance, the breeding value is:

$$\begin{aligned}\hat{u}_{10} &= 0.5(\hat{u}_3 + \hat{u}_2) + k(y_{10} - \hat{b}_2 - \hat{b}_5 - \hat{m}_2 - \hat{p}_2 - 0.5(\hat{u}_3 + \hat{u}_2)) \\ &= 0.5(1.165 + -1.244) + 0.17647(22 - 3.386 - 27.691 - (-1.583) \\ &\quad - (-1.701) - 0.5(1.165 + -1.244)) \\ &= -1.055\end{aligned}$$

Back-solving for maternal effects

The equation for obtaining genetic maternal effects for non-parents can be derived as follows. From the MME, the equation for direct and genetic maternal effects for non-parent i is:

$$\begin{bmatrix} r^{-1} + n^{-1}g^{11} & n^{-1}g^{12} \\ n^{-1}g^{21} & n^{-1}g^{22} \end{bmatrix} \begin{bmatrix} \hat{u}_i \\ \hat{m}_i \end{bmatrix} = G^{-1}k_2 \begin{bmatrix} \hat{u}_s + \hat{u}_d \\ \hat{m}_s + \hat{m}_d \end{bmatrix} + r^{-1} \begin{bmatrix} y_i - \hat{b}_i - \hat{m}_{dam} - \hat{p}e_{dam} \\ 0 \end{bmatrix} \qquad (7.9)$$

where n is defined in Eqn 7.8 and other terms are as defined in Eqn 7.4.

From the above equations:

$$\begin{aligned}\hat{m}_i &= [g^{22}(\hat{m}_s + \hat{m}_d) + g^{21}(\hat{u}_s + \hat{u}_d) - n^{-1}g^{21}(\hat{u}_i)]/n^{-1}g^{22} \\ \hat{m}_i &= n(\hat{m}_s + \hat{m}_d) + [(g^{21}n(\hat{u}_s + \hat{u}_d) - g^{21}\hat{u}_i)/g^{22}] \\ \hat{m}_i &= n(\hat{m}_s + \hat{m}_d) + g^{21}/g^{22}(n(\hat{u}_s + \hat{u}_d) - \hat{u}_i)\end{aligned}$$

Note that:

$$\begin{aligned}g^{21}/g^{22} &= \{-g_{12}/(g_{11}g_{22} - g_{12}g_{21})\}\{(g_{11}g_{22} - g_{12}g_{21})/g_{11}\} \\ &= -g_{12}/g_{11}\end{aligned}$$

Therefore:

$$\hat{m}_i = n(\hat{m}_s + \hat{m}_d) + g_{12}/g_{11}(\hat{u}_i - d(\hat{u}_s + \hat{u}_d)) \qquad (7.10)$$

When both parents are known:

$$\hat{m}_i = 0.5(\hat{m}_s + \hat{m}_d) + (g_{12}/g_{11})(\hat{u}_i - 0.5(\hat{u}_s + \hat{u}_d))$$

For instance, for animal 10:

$$\begin{aligned}\hat{m}_{10} &= 0.5(\hat{m}_3 + \hat{m}_2) + (g_{12}/g_{11})(\hat{u}_5 - 0.5(\hat{u}_3 + \hat{u}_2)) \\ &= 0.5(0.736 + (-1.583)) + (-40/150)(-1.055 - 0.5(1.165 + -1.244)) \\ &= -0.153\end{aligned}$$

The solutions for direct and maternal effects of all non-parents in the example data (animals 10 to 14) applying Eqns 7.7 and 7.9 are exactly the same as obtained for these animals in the animal model.

7.4 Sire and Maternal Grandsire Model

In some cases, due to the structure of the available data, a sire and maternal grandsire model may be fitted for traits affected by direct and maternal genetic effects.

Downloaded From Vet-eBooks.com

This tends to be more common for calving traits such as calving ease or stillbirth (Wiggans *et al*., 2003). The calving event is regarded as a direct effect of the service sire (direct effect). This predicts how easily his progeny are born and is computed by fitting the service sire. The maternal effect, which predicts how easily the bull daughters calve, is computed by fitting the MGS, hence the name sire-maternal grandsire (S-MGS) model.

The model then is similar to Eqn 7.1 and can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{Wmgs} + \mathbf{Spe} + \mathbf{e} \qquad (7.11)$$

where **y** = vector of observations, **s** = vector of random service sire (direct) effects, **mgs** = vector of random MGS (indirect) genetic effects and other terms defined as in Eqn 7.1, but **Z** and **W** are now incidence matrices relating records to service sire and MGS genetic effects, respectively. Note that if only first lactation data is being analysed, then the **pe** can be omitted from the model.

It is assumed that:

$\text{var}(\mathbf{s}) = \mathbf{A}\sigma_s^2$, with $\sigma_s^2 = 0.25\sigma_u^2$; $\text{var}(\mathbf{mgs}) = \mathbf{A}\sigma_{mgs}^2$ with $\sigma_{mgs}^2 = (\frac{1}{16}\sigma_u^2 + \frac{1}{4}\sigma_m^2 + \frac{1}{16}\sigma_{u,m})$,

where σ_u^2 and σ_m^2 are the additive genetic variance and maternal genetic variance, respectively.

$\text{cov}(\mathbf{s}, \mathbf{mgs}) = \mathbf{A}\sigma_{s,mgs}$, with $\sigma_{s,mgs} = \frac{1}{8}\sigma_u^2 + \frac{1}{4}\sigma_{u,m}$

$\text{var}(\mathbf{pe}) = \sigma_{pe}^2 = (\frac{3}{16}\sigma_u^2 + \frac{3}{4}\sigma_m^2 + \frac{3}{4}\sigma_{u,m} + \sigma_{pe}^2)$ and

$\text{var}(\mathbf{e}) = \sigma_e^2 = (\frac{1}{2}\sigma_u^2 + \frac{1}{4}\sigma_{te}^2)$

The same principles described in Section 7.2 can be used in the application of Eqn 7.11 to estimate breeding values and solutions for fixed effects. Note, however, that MME from such an analysis will produce predicted transmitting abilities (PTAs) (which is half of the EBV) for the service sire (direct effect) and PTAs for maternal effect are computed as:

PTA maternal effect = PTA for MGS from MME – 0.25*(PTA for direct effect)

The variance components for a S-MGS model can be converted to variances for an animal model direct and maternal effects from the details of the components of the variances defined above. Thus the direct genetic variance component $(\sigma_u^2) = 4\sigma_s^2$, the covariance between direct and maternal component $(\sigma_{u,m}) = 4^*(\sigma_{s,mgs}) - 0.5\sigma_u^2$ and the maternal genetic variance component $(\sigma_m^2) = 4\sigma_{mgs}^2 - 0.25\sigma_u^2 - \sigma_{u,m}$. The computation of maternal genetic component (σ_m^2) can be illustrated as:

$$\sigma_m^2 = (4\sigma_{mgs}^2 - 0.25\sigma_u^2 - \sigma_{u,m}) = 4(\tfrac{1}{16}\sigma_u^2 + \tfrac{1}{4}\sigma_m^2 + \tfrac{1}{4}\sigma_{u,m}) - 0.25\sigma_u^2 - \sigma_{u,m} = \sigma_m^2$$

Downloaded From Vet-eBooks.com

8 Social Interaction Models

8.1 Introduction

Social interaction among animals, such as competition and cooperation, can have a profound effect on the expressions of performance and welfare traits in domestic livestock populations (Muir, 2005; Bijma *et al*., 2007a). When a group of animals rely on a limiting resource (e.g. feed) to achieve an outcome (e.g. growth) the observed phenotype of an individual (e.g. growth rate) can be influenced by both the phenotype (e.g. ability to fight for food) and the genotype (which confers this ability) of the competitors in the group. So the growth rate of piglets, for instance, can be reduced due to competition for food. In laying hen production systems, social interactions can result in mortality due to cannibalism when hens are housed in groups, and this poses both economic and welfare problems.

Although a major component of the social interaction among group members may appear to be environmental, there is a genetic component (Wolf *et al*., 1998) attributable to the genes carried by others in the group which affects how they compete; generally referred to as indirect genetic effects (IGE) (Cheverud and Moore, 1994; Moore *et al*., 1997). A selection experiment to reduce mortality due to cannibalism in domestic chickens (Muir, 1996) has shown that heritable interactions (or IGE) can contribute substantially to response to selection. Selection schemes that ignore this social effect of an individual on the phenotypes of its group members could result in less optimum response or even response in the opposite direction (Griffing, 1967). This social effect or indirect genetic effect (Cheverud and Moore, 1994) is often referred to as an associative effect (Griffing, 1967). In addition, Bijma *et al*. (2007b) indicated that the existence of social interaction among individuals may increase the total heritable variance in a trait. They found that heritable variance in survival days expressed as a proportion of phenotypic variance increased from 7 to 20% due to social interactions, indicating that about two-thirds of heritable variation is due to interactions among individuals. One possible solution for improving traits affected by social interaction is to undertake group selection (Griffing, 1967). However, an optimum individual selection scheme to improve traits affected by interactions among individuals will involve the use of models that account for:

1. direct effects due to the direct effects of the genes of the individual; and
2. indirect effects due to the associative effect of the individual on its group members.

The phenotype (P_i) of an animal i for a trait influenced by social interaction belonging to a group with n members where interaction occurs may be modelled as:

$$P_i = A_{D,i} + Q_{D,i} + E_{D,i} + \sum_{j \neq i}^{n-1} A_{S,j} + \sum_{j \neq i}^{n-1} Q_{S,j} + E_{S,j}$$

where j is one of the n – 1 group mates, $A_{D,i}$ and $A_{S,j}$ are the additive direct effect and sum of the additive indirect effects of each of the n – 1 group mates, with corresponding

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

non-additive components $Q_{D,i}$ and $Q_{S,i}$ and environmental components $E_{D,i}$ and $E_{S,j}$. The non-additive components may be combined with the environmental components such that the equation may be expressed:

$$P_i = A_{D,i} + E_{D,i} + \sum_{j \neq i}^{n-1} A_{S,j} + E_{S,j}$$

Therefore the phenotypic variance can be derived as:

$$\text{var}(P) = \text{var}[A_{D,i} + E_{D,i} + \sum_{j \neq i}^{n-1} A_{S,j} + E_{S,j}]$$

Given that $\text{cov}(E_{D,i}, E_{S,j}) = 0$ when $i \neq j$, and $\text{cov}(A, E) = 0$ for all i, j, then:

$$\text{var}(P) = \sigma^2_{A_D} + \sigma^2_{E_D} + \text{var}(\sum_{j \neq i}^{n-1} A_{S,j}) + \text{var}(\sum_{j \neq i}^{n-1} E_{S,j}) + 2\text{cov}(A_{D,i}, \sum_{j \neq i}^{n-1} A_{S,j})$$

with $\text{cov}((E_{S,j}, E_{S,j'}) = 0$
when:

$$j \neq j', \text{var}\left(\sum_{j \neq i}^{n-1} E_{s,j}\right) = (n-1)\sigma^2_{E_S}$$

Also given that $\text{cov}((A_{S,j}, A_{S,j'}) = r_{jj'}\sigma^2_{A_S}$ where $r_{jj'}$ is the relatedness between animals j and j', then:

$$\text{var}\left(\sum_{j \neq i}^{n-1} A_{s,j}\right) = (n-1)\sigma^2_{A_S} + (n-1)(n-2)r\sigma^2_{A_S}$$

with r equal to the mean relatedness within the groups. Finally:

$$\text{cov}(A_{D,i}, \left(\sum_{j \neq i}^{n-1} A_{s,j}\right) = (n-1)r\sigma_{A_{DS}}$$

Collecting all the terms together gives the phenotypic variance as:

$$\sigma^2_p = \sigma^2_{A_D} + \sigma^2_{E_D} + (n-1)(\sigma^2_{A_S} + \sigma^2_{E_S}) + (n-1)r[2\sigma_{A_{DS}} + (n-2)\sigma^2_{A_S}]$$

However, the total breed value (TBV; Bijma *et al*., 2007a) for individual i is:

$$\text{TBV}_i = A_{D,i} + (n-1)\, A_{s,j} \tag{8.1}$$

Note that TBV is what the progeny of the individual i will inherit and is the relevant breeding value in computing response for selection for traits affected by associative effects. Therefore, the total heritable variance of the trait equals the variance of the TBVs (σ^2_{TBV}) among individuals and is:

$$\sigma^2_{TBV} = \sigma^2_{A_D} + 2(n-1)\sigma_{A_{DS}} + (n-1)^2\sigma^2_{A_S}$$

where $\sigma^2_{A_D}$, $\sigma^2_{A_S}$ and $\sigma_{A_{DS}}$ are the variance of direct breeding value (DBV), associative breeding value (SBV) and the covariance between DBV and SBV, respectively. The sign of this covariance provides a measure of the competition versus cooperation among group members. Negative values may be interpreted as 'heritable competition' in the sense that animals' positive DBV on the basis of their phenotype has a negative heritable impact on the phenotypes of their associates. On the other hand, a positive covariance may be interpreted as 'heritable cooperation' (Bijma *et al*., 2007b).

Thus the ratio of total heritable variance to the phenotypic variance (τ^2) for traits with associative effects (Bergsma *et al*., 2008) can be expressed as $\tau^2 = \sigma^2_{TBV}/\sigma^2_p$. A comparison of τ^2 to the classical heritability ($h^2 = \sigma^2_A/\sigma^2_p$) indicates the proportional contribution of indirect additive effects to the total heritable variance for traits with associative effects.

Downloaded From Vet-eBooks.com

Bijma *et al*. (2007a) presented this general formula for total genetic response per generation ($\Delta\bar{G}$) to selection for traits with associative effects.

$$\Delta\bar{G} = \{wt(n-1)(r+1)\sigma^2_{TBV} + (1-wt)\sigma_{p,TBV}\} \kappa / \sigma_I$$

where $\sigma_{p,TBV}$ is the covariance between the phenotype of the individual and TBV, r measures the degree of genetic relatedness, which is twice the coefficient of coancestry, κ is the selection intensity, σ_I is the standard deviation of the index (**I**) that combines individual phenotypes and phenotypes of group members, and wt defines the weights on individuals versus phenotypes of group members, such that:

$$\mathbf{I} = wtP_i + (1-wt)\sum_{j\neq i}^{n} p_j$$

Thus for a given r, n, wt and selection intensity, response is dependent on the σ^2_{TBV} and the covariance between the phenotype of the individual and TBV. Therefore, response to selection may not necessarily follow the same direction as the selection pressure as in classical quantitative theory. The interactions among individuals affect both the direction and magnitude of selection response. Strong competition, for instance a negative $\sigma_{p,TBV}$ due to a large and negative $\sigma_{A_{DS}}$, will result in a response opposite in direction to the direction of selection.

8.2 Animal Model with Social Interaction Effects

Usually, data with associative effects tend to include animals that are full-sibs and therefore there is the need to account for the common environmental effects in the model. Thus the MME for a trait with social interaction effects could be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{u}_D + \mathbf{Z}_S\mathbf{u}_S + \mathbf{Wc} + \mathbf{e} \tag{8.2}$$

where **b** is the vector of fixed effects, $\mathbf{u}_D$ and $\mathbf{u}_S$ are the vectors for direct and associative genetic effects, respectively, **c** is the vector for common environmental effects and **e** is the vector for residual error.
It is also assumed that:

$$\text{var}\begin{bmatrix}\mathbf{u}_D \\ \mathbf{u}_S\end{bmatrix} = \begin{bmatrix}g_{11}\mathbf{A} & g_{12}\mathbf{A} \\ g_{21}\mathbf{A} & g_{22}\mathbf{A}\end{bmatrix}$$

and if there are n animals in a group, then for the ith animal:

$$\text{var}(e_i) = (\text{var}(E_{D,i} + E_{S,j}), j = 1, n-1 \text{ and } i \neq j) = \sigma^2_{E_D} + (n-1)\sigma^2_{E_S} \tag{8.3}$$

Assuming that n = 3, with animals i, j and k in the group, then the residual covariance between animal i and j in the same group or pen is:

$$\begin{aligned}\text{cov}_{penmates} &= \text{cov}(e_i, e_j) = \text{cov}(E_{D,i} + E_{S,j} + E_{S,k};\ E_{D,j} + E_{S,i} + E_{S,k}) \\ &= \text{cov}(E_{D,i}, E_{S,i}) + \text{cov}(E_{S,j}, E_{D,j}) + \text{cov}(E_{S,k}, E_{S,k}) = 2\sigma_{E_{DS}} + (n-2)\sigma^2_{E_S}\end{aligned} \tag{8.4}$$

Therefore, the correlation among animals in the same group (ρ) can be defined as:

$$\rho = \text{cov}(e_i, e_j) / \text{var}(e) = [2\sigma_{E_{DS}} + (n-2)\sigma^2_{E_S}] / [\sigma^2_{E_D} + (n-1)\sigma^2_{E_S}]$$

Assuming that residual covariance among different groups is zero, the residual variance structure can then be defined as var(**e**) = **R**, with $r_{ii} = \sigma^2_e$, $r_{ij} = \rho(\sigma^2_e)$ for animals i and j in the same group and r_{ij} = 0 for animals i and j in different groups. Thus **R** is block diagonal and with n = 3, the block diagonal for one group is:

Downloaded From Vet-eBooks.com

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \sigma_e^2$$

All elements between the various block diagonals are zero. However, Bergsma *et al.* (2008) indicated that the residual covariance within groups ($\text{cov}_{penmates}$) equals the variance among group means (σ_g^2). Thus when $\text{cov}_{penmates}$ or ρ is > 0, instead of fitting the correlated residual structure described above, a random group effect can be fitted as an equivalent model, with:

$$\sigma_g^2 = 2\sigma_{E_{DS}} + (n-2)\sigma_{E_S}^2$$

and residual variance now defined as:

$$\sigma_{e^*}^2 = \sigma_e^2 - \sigma_g^2$$

Therefore, the equivalent model to Eqn 8.2 is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{u}_D + \mathbf{Z}_S\mathbf{u}_S + \mathbf{Vg} + \mathbf{Wc} + \mathbf{e} \tag{8.5}$$

where $\mathbf{g}$ is the vector of random group effects with $\mathbf{g} \sim N(0, \mathbf{I}_g\sigma_g^2)$. The MME to be solved then are:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z_D} & \mathbf{X'Z_S} & \mathbf{X'V} & \mathbf{X'W} \\ \mathbf{Z'_DX} & \mathbf{Z'_DZ_D} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z'_DZ_S} + \mathbf{A}^{-1}\alpha_2 & \mathbf{Z'_DV} & \mathbf{Z'_DW} \\ \mathbf{Z'_SX} & \mathbf{Z'_SZ_D} + \mathbf{A}^{-1}\alpha_2 & \mathbf{Z'_S\,Z_S} + \mathbf{A}^{-1}\alpha_3 & \mathbf{Z'_SV} & \mathbf{Z'_SW} \\ \mathbf{V'X} & \mathbf{V'Z_D} & \mathbf{V'Z_S} & \mathbf{V'V} + \mathbf{I}\alpha_4 & \mathbf{V'W} \\ \mathbf{W'X} & \mathbf{W'Z_D} & \mathbf{W'Z_S} & \mathbf{W'V} & \mathbf{W'W} + \mathbf{I}\alpha_5 \end{pmatrix}$$

$$\begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_D \\ \hat{\mathbf{u}}_S \\ \hat{\mathbf{g}} \\ \hat{\mathbf{c}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Z'_Dy} \\ \mathbf{Z'_Sy} \\ \mathbf{V'y} \\ \mathbf{W'y} \end{pmatrix} \tag{8.6}$$

If $\mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}$ then $\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} = \sigma_{e^*}^2 \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}$, $\alpha_4 = \sigma_{e^*}^2/\sigma_g^2$ and $\alpha_5 = \sigma_{e^*}^2/\sigma_c^2$

However, when $\text{cov}_{penmates}$ is ≤ 0, then the MME to be solved are:

$$\begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{X} & \mathbf{X'R}^{-1}\mathbf{Z_D} & \mathbf{X'R}^{-1}\mathbf{Z_S} & \mathbf{X'R}^{-1}\mathbf{W} \\ \mathbf{Z'_DR}^{-1}\mathbf{X} & \mathbf{Z'_DR}^{-1}\mathbf{Z_D} + \mathbf{A}^{-1}g^{11} & \mathbf{Z'_DR}^{-1}\mathbf{Z_S} + \mathbf{A}^{-1}g^{12} & \mathbf{Z'_DR}^{-1}\mathbf{W} \\ \mathbf{Z'_SR}^{-1}\mathbf{X} & \mathbf{Z'_SR}^{-1}\mathbf{Z_D} + \mathbf{A}^{-1}g^{21} & \mathbf{Z'_SR}^{-1}\mathbf{Z_S} + \mathbf{A}^{-1}g^{22} & \mathbf{Z'_SR}^{-1}\mathbf{W} \\ \mathbf{W'R}^{-1}\mathbf{X} & \mathbf{W'R}^{-1}\mathbf{Z_D} & \mathbf{W'R}^{-1}\mathbf{Z_S} & \mathbf{W'R}^{-1}\mathbf{W} + \mathbf{I}\sigma_c^2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_D \\ \hat{\mathbf{u}}_S \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{y} \\ \mathbf{Z'_DR}^{-1}\mathbf{y} \\ \mathbf{Z'_SR}^{-1}\mathbf{y} \\ \mathbf{W'R}^{-1}\mathbf{y} \end{bmatrix} \tag{8.7}$$

Downloaded From Vet-eBooks.com

Although the number of equations to be fitted for Eqn 8.6 is usually more than for Eqn 8.7, the systems of equations for Eqn 8.7 are denser and more difficult to set up.

8.2.1 Illustration of a model with social interaction

Example 8.1

Table 8.1 contains the growth rate data of nine pigs housed in three pens during the finishing period in groups of three. The pigs are from three different litters and the aim is to estimate the direct and associative breeding values for all pigs, estimate sex effect and common environment effect as some of the pigs are full-sibs. It is assumed that genetic variances for direct and associative effects are 25.70 g^2 and 3.60 g^2, respectively, with a covariance of 2.25 g between them. Also, it is assumed that the variance for common environmental variance (σ_c^2) is 12.5 g^2 and residual variances for direct ($\sigma_{E_D}^2$) and associative ($\sigma_{E_S}^2$) effects are 40.6 g^2 and 10.0 g^2, respectively, and the correlation among pigs in the same pen (ρ) is 0.2.

The MME in Eqn 8.6 are initially used to analyse the data. Based on the given genetic parameters:

$$\text{var}(\mathbf{e}) = \sigma_{E_D}^2 + (n-1)\sigma_{E_S}^2 = 40.6 + (3-1)10 = 60.6$$

Since $\rho = \text{cov}(e_i, e_j)/\text{var}(\mathbf{e}) = 0.2$, and $\text{cov}(e_i, e_j) = \sigma_g^2$ in Eqn 8.6, then $\sigma_g^2 = \rho\ \text{var}(\mathbf{e}) = 0.2^*60.6 = 12.12$.

Therefore, the residual variance relevant to the analysis using Eqn 8.6 with groups fitted is $\text{var}(e^*) = \text{var}(e) - \sigma_g^2 = 60.6 - 12.12 = 48.48$ and:

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} \text{in}[8.6] = \mathbf{G}^{-1}\,\sigma_{e^*}^2 = \begin{bmatrix} 1.9956 & -1.2472 \\ -1.2472 & 14.2462 \end{bmatrix}$$

Setting up the incidence matrices $\mathbf{X}$, $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{Z}_D$ in Eqn 8.6 follows the pattern already described for other models in previous chapters, with $\mathbf{Z}_D$ being a diagonal matrix for animals with records and:

$$\mathbf{V} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

relating records to pen (groups).

Table 8.1. The growth rate of a set of finishing pigs.

Animal	Sire	Dam	Pen	Sex	Growth rate (g/day)*10
7	1	4	1	Male	5.50
8	1	4	1	Female	9.80
9	2	5	1	Female	4.90
10	1	4	2	Male	8.23
11	2	5	2	Female	7.50
12	3	6	2	Female	10.00
13	2	5	3	Male	4.50
14	3	6	3	Female	8.40
15	3	6	3	Male	6.40

Downloaded From Vet-eBooks.com

The matrix $\mathbf{Z}_S$ that relates an individual to other members of the same group is:

$$\mathbf{Z}_S = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Setting up the MME therefore follows a similar pattern to that described in previous chapters. Solving the MME (Eqn 8.6) for this example gives the following set of solutions. The results from an analysis that ignored associative effects but fitted random animal and common environmental effects and fixed effects of sex of pigs and pen effects are also presented.

Model with associative effects				Model with no associative effects[a]
Sex of pig effects				
Male	6.006			0.000
Female	8.241			2.169
Animal effects				
	DBV	SBV	TBV	
1	0.298	–0.044	0.210	0.336
2	–0.487	0.028	–0.431	–0.478
3	0.189	0.016	0.221	0.142
4	0.298	–0.044	0.210	0.336
5	–0.487	0.028	–0.431	–0.478
6	0.189	0.016	0.221	0.142
7	0.129	–0.075	–0.022	0.279
8	0.527	–0.098	0.330	0.652
9	–0.878	0.010	–0.858	–0.738
10	0.538	–0.003	0.531	0.412
11	–0.494	0.083	–0.328	–0.628
12	0.400	0.059	0.517	0.216
13	–0.578	0.019	–0.539	–0.547
14	0.156	0.004	0.164	0.162
15	0.201	0.002	0.204	0.192
Common environment effects				
1	0.325			0.327
2	–0.504			–0.465
3	0.178			0.139
Group effects				
1	–0.275			
2	0.367			
3	–0.092			

DBV, direct EBV; SBV, associative EBV; TBV, total EBV = (DBV + (n – 1)SBV).

[a]Model also fitted pen effects, and solutions were 5.160, 7.131 and 5.838 for pens 1, 2 and 3, respectively.

Downloaded From Vet-eBooks.com

Although solutions for sex of pig and common environmental effects were generally in the same direction in models with or without associative effects, there was a major re-ranking of animals based on the EBVs. Griffing (1967) has indicated that selection schemes that ignore this social effect of an individual on the phenotypes of its group members could result in less optimum response, while Bijma *et al.* (2007a) observed that the presence of social interaction among individuals may increase the total heritable variance in a trait.

8.3 Partitioning Evaluations from Associative Models

The equations for DBV and SBV for animal i can be written as:

$$\begin{bmatrix} \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{D}} + \mathbf{A}^{-1}_{\alpha_1} & \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{S}} + \mathbf{A}^{-1}_{\alpha_2} \\ \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{D}} + \mathbf{A}^{-1}_{\alpha_2} & \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{S}} + \mathbf{A}^{-1}_{\alpha_3} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{\mathrm{D}} \\ \hat{\mathbf{u}}_{\mathrm{S}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_{i\mathrm{D}}(\mathbf{y}_i - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}_{j\mathrm{S}}\hat{\mathbf{u}}_{j\mathrm{S}}) \\ \mathbf{Z}'_{i\mathrm{S}}(\mathbf{y}_j - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}_{j\mathrm{S}}\hat{\mathbf{u}}_{j\mathrm{S}} - \mathbf{Z}_{j\mathrm{D}}\hat{\mathbf{u}}_{j\mathrm{D}}) \end{bmatrix}$$

$$= \begin{pmatrix} \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{D}} & 0 \\ 0 & \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{S}} \end{pmatrix} \begin{pmatrix} \mathbf{yd}_1 \\ \mathbf{yd}_2 \end{pmatrix} \tag{8.8}$$

with $i \neq j$ and $j = (1, n-1)$, where n is the number of animals in the same group and:

$$\mathbf{yd}_1 = (\mathbf{Z}'_{i\mathrm{D}}\,\mathbf{Z}_{i\mathrm{D}})^{-1}\mathbf{Z}'_{i\mathrm{D}}\left(y_i - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}_{j\mathbf{S}}\hat{\mathbf{u}}_{j\mathbf{S}}\right) \text{ and}$$

$$\mathbf{yd}_2 = (\mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{S}})^{-1}\mathbf{Z}'_{i\mathrm{S}}(\mathbf{y}_i - \mathbf{X}\mathbf{b} - \mathbf{Z}_{j\mathrm{S}}\hat{\mathbf{u}}_{j\mathrm{S}} - \mathbf{Z}_{j\mathrm{D}}\hat{\mathbf{u}}_{j\mathrm{D}})$$

Thus $\mathbf{yd}_1$ is the yield record of animal i corrected for all fixed effects and the SBVs of all other members in the same group, and $\mathbf{yd}_2$ is the average of the yield records of all animals in the same group apart from animal i corrected for all fixed effects, the DBVs and SBVs of the members of the group. Transferring the left non-diagonal terms of $\mathbf{A}^{-1}$ in Eqn 8.8 to the right side of the equation gives:

$$\begin{bmatrix} \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{D}} + a^{ii}\alpha_1 & \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{S}} + a^{ii}\alpha_2 \\ \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{D}} + a^{ii}\alpha_2 & \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{S}} + a^{ii}\alpha_3 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{\mathrm{D}} \\ \hat{\mathbf{u}}_{\mathrm{S}} \end{bmatrix} = 2a_{par} \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix} \mathbf{PA}_1 \\ \mathbf{PA}_2 \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{D}} & 0 \\ 0 & \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{S}} \end{pmatrix} \begin{pmatrix} \mathbf{yd}_1 \\ \mathbf{yd}_2 \end{pmatrix} + 0.5a_{prog} \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix} 2\hat{\mathbf{u}}_{\mathrm{D}prog} - \hat{\mathbf{u}}_{\mathrm{D}mate} \\ 2\hat{\mathbf{u}}_{\mathrm{S}prog} - \hat{\mathbf{u}}_{\mathrm{S}mate} \end{pmatrix}$$

where $\mathbf{PA}_1$ and $\mathbf{PA}_2$ are the parent averages for DBV and SBV for animal i; $a_{par} = 1$, $\frac{2}{3}$ or $\frac{1}{2}$ if both, one or neither parents are known, respectively; and $a_{prog} = 1$ if the animal's mate is known and $\frac{2}{3}$ if unknown. Note that $a^{ii} = 2a_{par} + 0.5a_{prog}$, therefore pre-multiplying both sides of the above equation by the inverse of **DIAG**, with:

$$\mathbf{DIAG} = \begin{bmatrix} \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{D}} + a^{ii}\alpha_1 & \mathbf{Z}'_{i\mathrm{D}}\mathbf{Z}_{i\mathrm{S}} + a^{ii}\alpha_2 \\ \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{D}} + a^{ii}\alpha_2 & \mathbf{Z}'_{i\mathrm{S}}\mathbf{Z}_{i\mathrm{S}} + a^{ii}\alpha_3 \end{bmatrix}$$

gives:

$$\begin{bmatrix} \hat{\mathbf{u}}_{\mathrm{D}} \\ \hat{\mathbf{u}}_{\mathrm{S}} \end{bmatrix} = \mathbf{WT}_1 \begin{pmatrix} \mathbf{PA}_1 \\ \mathbf{PA}_2 \end{pmatrix} + \mathbf{WT}_2 \begin{pmatrix} \mathbf{yd}_1 \\ \mathbf{yd}_2 \end{pmatrix} + \mathbf{WT}_3 \begin{pmatrix} \mathbf{PC1} \\ \mathbf{PC2} \end{pmatrix} \tag{8.9}$$

Downloaded From Vet-eBooks.com

where:

$$\begin{pmatrix} \mathbf{PC1} \\ \mathbf{PC2} \end{pmatrix} = \sum a_{prog} \begin{pmatrix} 2\hat{\mathbf{u}}_{\mathbf{D}prog} - \hat{\mathbf{u}}_{\mathbf{D}mate} \\ 2\hat{\mathbf{u}}_{\mathbf{S}prog} - \hat{\mathbf{u}}_{\mathbf{S}mate} \end{pmatrix} \Big/ \sum a_{prog}$$

The weights $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3 = \mathbf{I}$, with:

$$\mathbf{W}_1 = (\mathbf{DIAG})^{-1} \left(2a_{par} \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix} \right), \mathbf{W}_2 = (\mathbf{DIAG})^{-1} \begin{pmatrix} \mathbf{Z}'_{i\mathbf{D}} \mathbf{Z}_{i\mathbf{D}} & 0 \\ 0 & \mathbf{Z}'_{i\mathbf{S}} \mathbf{Z}_{i\mathbf{S}} \end{pmatrix}$$

and $\mathbf{W}_3 = (\mathbf{DIAG})^{-1} 0.5 \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix} \Sigma a_{prog}$

Equation 8.9 is illustrated below using pig 7 in Example 8.1. For pig 7:

$$\mathbf{yd}_1 = (y_7 - \hat{b}_1 - \hat{u}_{s8} - \hat{u}_{s9} - \hat{c}_1 - \hat{g}_1) = (5.50 - 6.006 - (-0.098) - 0.010 - 0.325 - (-0.274) = -0.469 \quad \text{and}$$

$$\mathbf{yd}_2 = 1/(n-1)((y_8 + y_9) - 2\hat{b}_1 - \hat{u}_{D8} - \hat{u}_{D9} - \hat{u}_{s8} - \hat{u}_{s9} - \hat{c}_1 - \hat{c}_2 - 2\hat{g}_1)$$

$$= \frac{1}{2}\big((9.8 + 4.9) - 2(8.241) - 0.527 - (-0.878) - (-0.098) - 0.010 - 0.325 - (-0.504) - 2(-0.274)\big) = -0.308$$

Since both parents are known:

$$\mathbf{DIAG} = 2 \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4.991 & -2.494 \\ -2.494 & 30.492 \end{pmatrix}$$

Therefore:

$$\mathbf{WT}_1 = (\mathbf{DIAG})^{-1} 2 \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix} = \begin{pmatrix} 0.791 & -0.034 \\ -0.017 & 0.932 \end{pmatrix} \quad \text{and}$$

$$\mathbf{WT}_2 = (\mathbf{DIAG})^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0.209 & 0.034 \\ 0.017 & 0.068 \end{pmatrix}$$

From Eqn 8.9:

$$\begin{pmatrix} \hat{\mathbf{u}}_{\mathbf{D}7} \\ \hat{\mathbf{u}}_{\mathbf{S}7} \end{pmatrix} = \mathbf{WT}_1 \begin{bmatrix} \mathbf{PA}_1 \\ \mathbf{PA}_2 \end{bmatrix} + \mathbf{WT}_2 \begin{pmatrix} \mathbf{yd}_1 \\ \mathbf{yd}_2 \end{pmatrix} = \mathbf{WT}_1 \begin{bmatrix} 0.298 \\ -0.044 \end{bmatrix} + \mathbf{WT}_2 \begin{pmatrix} -0.469 \\ -0.308 \end{pmatrix} = \begin{pmatrix} 0.129 \\ -0.075 \end{pmatrix}$$

The weights indicate that the relative emphasis on parent contribution was higher for the SBV compared to the DBV. This might be due to the lower genetic variance for associative effects in the model.

8.4 Analysis Using Correlated Error Structure

The analysis of the same data using Eqn 8.7 gave the same solutions obtained from Eqn 8.6. Since the major difference is the structure of the residual covariance, **R**, this section has only focused on illustrating the structure of **R**, for this example. Although the number of equations using Eqn 8.7 were three less compared to Eqn 8.6, the number of non-zero elements was higher (481 compared with 462 for Eqn 8.6). This is due

Downloaded From Vet-eBooks.com

to the correlated residual variance structure in Eqn 8.7. As mentioned earlier, residual error structure is block diagonal with all elements between the various block diagonals being zero. Thus for the example data in Table 8.1, with $n = 3$, the **R** block diagonal structure for one group is:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \sigma_e^2 = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{pmatrix} 60.6 \quad \text{and}$$

$$\mathbf{R}^{-1} = \begin{pmatrix} 0.01768 & -0.00295 & -0.00295 \\ -0.00295 & 0.01768 & -0.00295 \\ -0.00295 & -0.00295 & 0.01768 \end{pmatrix}$$

The MME can then easily be set following the usual principles.

Downloaded From Vet-eBooks.com

9 Analysis of Longitudinal Data

9.1 Introduction

In Chapter 4, the use of a repeatability model to analyse repeated measurements on individuals was discussed and illustrated. The basic assumption of the model was that repeated measurements were regarded as expression of the same trait over time. In other words, a genetic correlation of unity was assumed between repeated measurements. The model has been employed mostly in the genetic evaluation of milk production traits of dairy cattle in most countries up to 1999 (Interbull, 2000). The main advantages of this model are its simplicity, fewer computation requirements and fewer parameters compared to a multivariate model (see Chapter 5). However, the model has some drawbacks. First, test day records within lactation are assumed to measure the same trait during the whole lactation length and are used to compute 305-day yields. These test day records are actually repeated observations measured along a trajectory (days in milk), and the mean and covariance between measurements change gradually along the trajectory. Several studies have reported that heritability of daily milk yields varied with days in milk. In addition, genetic correlations between repeated measurements usually tended to decrease as the time between them increases (Meyer, 1989; Pander *et al*., 1992). The extension of test records to compute 305-day yields is unable to account for these changes in the covariance structure. Second, the assumption that 305-day yields across parities measure the same trait suffers from the same limitations.

However, in beef cattle, repeated measurements of growth have been analysed somewhat differently, with the assumption that measurements are genetically different but correlated traits. Usually, a multivariate model has been employed in the genetic evaluation of these traits. While the multivariate model is an improvement on the repeatability model by accounting for the genetic correlations among different records, it would be highly over-parameterized if records were available at many ages or time periods. For instance, a multivariate model for daily body weight up to yearly weight in beef cattle as different traits will not only be over-parameterized but it will be difficult to obtain accurate estimates of the necessary genetic parameters.

An appropriate model for the analysis of repeated measurements over time or age (also termed longitudinal data) should account for the mean and covariance structure that changes with time or age and should be feasible in terms of estimating the required genetic parameters. In 1994, Schaeffer and Dekkers introduced the concept of the random regression (RR) model for the analysis of test day records in dairy cattle as a means of accounting for the covariance structure of repeated records over time or age. Almost at the same time, Kirkpatrick *et al*. (1990, 1994) introduced

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

covariance functions (CFs) to handle the analysis of longitudinal data, illustrating their methodology with growth data. The application of RR models in animal breeding for the analysis of various types of data has been comprehensively reviewed by Schaeffer (2004). Prior to the development of the RR model for genetic evaluation, milk yield test day records were analysed by Ptak and Schaefer (1993) using a fixed regression model. The details of this model are discussed and illustrated in the next section, followed by its extension to an RR model. This is then followed by a brief presentation of CF, and the equivalence of the RR model and CF is demonstrated.

9.2 Fixed Regression Model

The theoretical framework for the fixed regression model and its application for the analysis of longitudinal data such as test day milk production traits were presented by Ptak and Schaefer in 1993. On a national scale, a fixed regression model was implemented for the genetic evaluation of test day records of milk production traits and somatic cell counts in Germany from 1995 until 2002. The model involved the use of individual test day records, thereby avoiding the problem of explicitly extending test day yields into 305-day yield, and accounted for the effects peculiar to all cows on the same test day within herds (herd–test–day (HTD) effect). Therefore, corrections for temporary environmental effects on the day of test are more precise compared to evaluations based on 305-day yields. The model also accounted for the general shape of the lactation curve of groups of similar age, and calving in the same season and region. The latter was accomplished by regressing lactation curve parameters on days in milk (hence the name of the model) within the groupings for cows. Inclusion of the curve therefore allows for correction of the means of test day yields at different stages of lactation. Fitting residual variances relevant to the appropriate stage of lactation could also account for the variation of test day yields with days in milk. The only major disadvantage is that the volume of data to be analysed is much larger, especially in the dairy situation, as ten or more test day observations are stored relative to a single 305-day yield.

Similar to the repeatability model, at the genetic level, the fixed regression model assumes that test day records within a lactation are repeated measurements of the same trait, i.e. a genetic correlation of unity among test day observations. Usually, the permanent environmental effect is included in the model to account for environmental factors with permanent effects on all test day yields within lactation.

The fixed regression model is of the form:

$$y_{tij} = htd_i + \sum_{k=0}^{nf} \phi_{tijk}\beta_k + u_j + pe_j + e_{tij}$$

where y_{tij} is the test day record of cow j made on day t within HTD subclass i; β_k are fixed regression coefficients; u_j and pe_j are vectors of animal additive genetic and permanent environmental effects, respectively, for animal j; ϕ_{tijk} is the vector of the kth Legendre polynomials or any other curve parameter, for the test day record of cow j made on day t; nf is the order of fit for Legendre polynomials used to model the fixed regressions (fixed lactation curves) and e_{tij} is the random residual. In matrix notation, the model may be written as:

$$\mathbf{y = Xb + Qu + Zpe + e} \qquad (9.1)$$

Downloaded From Vet-eBooks.com

where **y** is the vector of TD yields, **b** is a vector of solutions for HTD and fixed regressions, and **u** and **pe** are vectors of animal additive genetic and permanent environmental effects, respectively. The variances of **u** and **pe** are as defined in Eqn 4.1. The matrices **X**, **Q** and **Z** are incidence matrices and are described in detail in the next section, which illustrates the application of the model. It is assumed that var(**u**) = $\mathbf{A}\sigma_u^2$, and var(**pe**) = $\mathbf{I}\sigma_p^2$, and var(**e**) = $\mathbf{I}\sigma_e^2 = \mathbf{R}$. The MME for Eqn 9.1 are:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Q} & \mathbf{X'Z} \\ \mathbf{Q'X} & \mathbf{Q'Q} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Q'Z} \\ \mathbf{Z'X} & \mathbf{Z'Q} & \mathbf{Z'Z} + \alpha_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{p}}\mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Q'y} \\ \mathbf{Z'y} \end{pmatrix}$$

with $\alpha_1 = \sigma_e^2/\sigma_u^2$ and $\alpha_2 = \sigma_e^2/\sigma_{p.}^2$

9.2.1 An illustration

Example 9.1

Given in Table 9.1 are the test day fat yields of five cows in a herd with details of HTD and days in milk (DIM). The aim is to estimate solutions for HTD effects, regression coefficients for a fixed lactation curve fitting Legendre polynomials of order 4, solutions for permanent environmental effects and breeding values for animal effects using Eqn 9.1. Assume that the estimated variances for additive genetic effects, permanent environmental effects and residual variances were 5.521 kg^2, 8.470 kg^2 and 3.710 kg^2, respectively. Then:

$$\alpha_1 = \sigma_e^2/\sigma_u^2 = 3.710/5.521 = 0.672$$

and:

$$\alpha_2 = \sigma_e^2/\sigma_p^2 = 3.710/8.470 = 0.438$$

Table 9.1. Test day fat yields (TDY) for some cows in a herd.

	Animals									
	4		5		6		7		8	
DIM	HTD	TDY	HTD	TDY	HTD	TDY	HTD	TDY	HTD	TDY
4	1	17.0	1	23.0	6	10.4	4	22.8	1	22.2
38	2	18.6	2	21.0	7	12.3	5	22.4	2	20.0
72	3	24.0	3	18.0	8	13.2	6	21.4	3	21.0
106	4	20.0	4	17.0	9	11.6	7	18.8	4	23.0
140	5	20.0	5	16.2	10	8.4	8	18.3	5	16.8
174	6	15.6	6	14.0			9	16.2	6	11.0
208	7	16.0	7	14.2			10	15.0	7	13.0
242	8	13.0	8	13.4					8	17.0
276	9	8.2	9	11.8					9	13.0
310	10	8.0	10	11.4					10	12.6

DIM, days in milk; HTD, herd–test–day.

Downloaded From Vet-eBooks.com

The modelling of the fixed lactation curve by means of Legendre polynomials implies the need to compute $\mathbf{\Phi}$, which is the matrix of Legendre polynomials evaluated at the different DIM. The matrix $\mathbf{\Phi}$ is of order t (the number of DIM) by k (where k is the order of fit) with element $\phi_{ij} = \phi_j(a_t)$, which is the jth Legendre polynomial evaluated at the standardised DIM t (a_t). Therefore $\mathbf{\Phi} = \mathbf{M\Lambda}$, where $\mathbf{M}$ is the matrix containing the polynomials of the standardized DIM values and $\mathbf{\Lambda}$ is a matrix of order k containing the coefficients of Legendre polynomials. The calculation of $\mathbf{\Phi}$ is outlined in Appendix G and matrix $\mathbf{\Phi}$ for Example 9.1 is shown in Eqn g.1.

Setting up the incidence matrices for the MME

In Eqn 9.1, let $\mathbf{Xb} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2$, then in Example 9.1, the matrix $\mathbf{X}_1$, which relates records to HTD effects, is of order n_{td} (number of TD records) and is too large to be presented. However, $\mathbf{X}_1'\mathbf{X}_1$ is diagonal and is:

$\mathbf{X}_1'\mathbf{X}_1$ = diagonal [3, 3, 3, 4, 4, 5, 5, 5, 5, 5]

The matrix $\mathbf{X}_2$ of order n_{td} by nf contains Legendre polynomials (covariables) corresponding to the DIM of the ith TD yield. Thus the ith row of $\mathbf{X}_2$ contains elements of the row of $\mathbf{\Phi}$ corresponding to the DIM for the ith record. The matrix $\mathbf{X}_2$, with rows for the first three TD records of cow 4 and the last three TD records of cow 8 is:

$$\begin{bmatrix} 0.7071 & -1.2247 & 1.5811 & -1.8704 & 2.1213 \\ 0.7071 & -0.9525 & 0.6441 & -0.0176 & -0.6205 \\ 0.7071 & -0.6804 & -0.0586 & 0.7573 & -0.7757 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.7071 & 0.6804 & -0.0586 & -0.7573 & -0.7757 \\ 0.7071 & 0.9525 & 0.6441 & -0.0176 & -0.6205 \\ 0.7071 & 1.2247 & 1.5811 & 1.8704 & 2.1213 \end{bmatrix}$$

and $\mathbf{X}_2'\mathbf{X}_2$ is:

$$\mathbf{X}_2'\mathbf{X}_2 = \begin{bmatrix} 20.9996 & -4.4261 & 4.0568 & -0.8441 & 8.7149 \\ -4.4261 & 24.6271 & -4.7012 & 11.1628 & -3.0641 \\ 4.0568 & -4.7012 & 31.0621 & -6.6603 & 19.0867 \\ -0.8441 & 11.1628 & -6.6603 & 38.6470 & -8.8550 \\ 8.7149 & -3.0641 & 19.0867 & -8.8550 & 48.2930 \end{bmatrix}$$

Considering only animals with records, $\mathbf{Q} = \mathbf{Z}$ and is a matrix of order 5 (number of animals) by n_{td}. The matrix $\mathbf{Q}'$ could be represented as:

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{q}_4' & 0 & 0 & 0 & 0 \\ 0 & \mathbf{q}_5' & 0 & 0 & 0 \\ 0 & 0 & \mathbf{q}_6' & 0 & 0 \\ 0 & 0 & 0 & \mathbf{q}_7' & 0 \\ 0 & 0 & 0 & 0 & \mathbf{q}_8' \end{bmatrix}$$

Downloaded From Vet-eBooks.com

where $\mathbf{q}_i'$ is a vector of ones with size equal to the number of TD records for the *i*th cow. The matrices $\mathbf{Q}'\mathbf{Q}$ and $\mathbf{Z}'\mathbf{Z}$ are both diagonal and equal. Thus:

$$\mathbf{Q}'\mathbf{Q} = \mathbf{Z}'\mathbf{Z} = \text{diag}[10, 10, 5, 7, 10]$$

The matrix $\mathbf{A}^{-1}$ has been given in Example 4.1. The remaining matrices in the MME could be obtained as outlined in earlier chapters. Solving the MME, with the solution for the 10th level of HTD effects constrained to zero, give the following results:

Effects	Solutions	
HTD		
1	10.9783	
2	7.9951	
3	8.7031	
4	8.2806	
5	6.3813	
6	3.1893	
7	3.3099	
8	3.3897	
9	0.6751	
10	0.0000	
Fixed regression coefficients		
1	16.3082	
2	–0.5227	
3	–0.1245	
4	0.5355	
5	–0.4195	
Animal effect		
	EBV for daily yield	EBV for 305-day yield
1	–0.3300	–100.6476
2	–0.1604	–48.9242
3	0.4904	149.5718
4	0.0043	1.3203
5	–0.2449	–74.7065
6	–0.8367	–255.2063
7	1.1477	350.0481
8	0.3786	115.4757
Permanent environmental effects		
Cow	Solutions for daily yield	Solutions for daily yield
4	–0.6156	–187.7634
5	–0.4151	–126.6150
6	–1.6853	–514.0274
7	2.8089	856.7092
8	–0.0928	–28.3035

EBV, estimated breeding value.

The solutions for the fixed regressions are regression coefficients from which plots of lactation curves can be obtained. In practice, the fixed regressions are usually fitted within group of cows calving in the same season in the same parity and of similar age. Thus the curves obtained for various groups of cows are useful for examining the influence of different environmental factors on lactation curves. In Example 9.1,

Downloaded From Vet-eBooks.com

one fixed lactation curve was fitted for all cows and a vector (**v**) of actual daily fat yield (kg) from days 4 to 310 can be obtained as:

$$\mathbf{v} = \mathbf{\Phi}\hat{\mathbf{b}} = \sum_{i=4}^{310}\sum_{j=1}^{nf} \phi_{ij}\, \hat{b}_{2j}$$

where $\mathbf{\Phi}$ is a matrix of Legendre polynomials evaluated from 4 to 310 DIM, as described in Appendix G. From the above equation, $\mathbf{v}_{38}$, for instance, is:

$$\mathbf{v}_{38} = \begin{bmatrix} 0.7071 & -0.9525 & 0.6441 & -0.0176 & -0.6205 \end{bmatrix} \hat{\mathbf{b}}_2 = 12.2001$$

For the DIM in the example data set, **v** is:

(DIM)	4	38	72	106	140	174	208	242	276	310
v =	[10.0835	12.2001	12.6254	12.2077	11.5679	11.0407	10.9156	11.1111	11.2500	10.8297]

A graph of the fixed lactation curve can be obtained by plotting the elements of **v** against DIM.

The EBV for animals and solutions for permanent environmental effect obtained by solving the MME are those for daily fat yield. To obtain EBV or solutions for **pe** effects on the *n*th DIM, these solutions are multiplied by *n*. This is implicit from the assumptions stated earlier of genetic correlations of unity among TD records. Thus EBVs for 305 days, shown in the table of results above, were obtained by multiplying the solutions for daily fat yield by 305.

Partitioning breeding values and solutions for permanent environmental effects

Similar to the repeatability model, EBVs of animals can be partitioned in terms of contributions from various sources, using Eqn 3.8. The YD for an animal is now calculated as the average of corrected TD records. The correction is for effects of HTD, fixed regressions and **pe**. Thus for cow 6 with five TD records, YD_6 is:

$$YD_6 = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\mathbf{y}_6 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \hat{\mathbf{p}}\mathbf{e})$$

$$\text{with } \mathbf{y}_6 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \hat{\mathbf{p}}\mathbf{e} = \mathbf{y}_c$$

$$= \begin{bmatrix} 10.4 \\ 12.3 \\ 13.2 \\ 11.6 \\ 8.4 \end{bmatrix} - \begin{bmatrix} 3.1893 \\ 3.3099 \\ 3.3897 \\ 0.6751 \\ 0.0000 \end{bmatrix} - \begin{bmatrix} 10.0835 \\ 12.2001 \\ 12.6254 \\ 12.2077 \\ 11.5679 \end{bmatrix} - \begin{bmatrix} -1.6853 \\ -1.6853 \\ -1.6853 \\ -1.6853 \\ -1.6853 \end{bmatrix} = \begin{bmatrix} -1.1875 \\ -1.5247 \\ -1.1298 \\ 0.4025 \\ -1.4826 \end{bmatrix}$$

and:

$$YD_6 = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\mathbf{y}_c) = \frac{1}{5}\left(\mathbf{q}'_6 \begin{bmatrix} -1.1875 \\ -1.5247 \\ -1.1298 \\ 0.4025 \\ -1.4826 \end{bmatrix} \right) = -4.9221/5 = -0.9844$$

Downloaded From Vet-eBooks.com

Then the solution for additive genetic effect for animals 6 using Eqn 3.8 is:

$$\hat{u}_6 = w_1((\hat{u}_1 + \hat{u}_5)/2) + w_2(YD_6)$$
$$= w_1((-0.3300 + -0.2449)/2) + w_2(-0.9844) = -0.8367$$

with $w_1 = 2(0.672)/6.344$, $w_2 = 5/6.344$ and 6.344 = the sum of the numerators of w_1 and w_2.

For animal 8 with ten TD records, the solution for additive genetic effect is:

$$\hat{u}_8 = w_1((\hat{u}_1 + \hat{u}_7)/2) + w_2(YD_8)$$
$$= w_1((-0.3300 + 1.1477)/2) + w_2(0.3746) = 0.3786$$

with $w_1 = 2(0.672)/11.344$, $w_2 = 10/11.344$ and 11.344 = the sum of the numerators of w_1 and w_2. The weights on YDs were 0.7882 and 0.8815 for animals 6 and 8, respectively. This illustrates the fact that as the number of TD increases, more emphasis is placed on performance records of the animal. Considering animal 4 with ten TD records and a progeny, her breeding value can be calculated as:

$$\hat{u}_4 = w_1((\hat{u}_1 + \hat{u}_2)/2) + w_2(YD_4) + w_3(\hat{u}_7 - 0.5\hat{u}_3)$$
$$= w_1((-0.3300 + -0.1604)/2) + w_2(-0.0226)$$
$$+ w_3(2(1.1477) - 0.4934) = 0.0043$$

where $w_1 = 2(0.672)/11.68$, $w_2 = 10/11.68$ and $w_3 = 0.5(0.672)/11.68$ and 11.68 is the sum of the numerators of w_1, w_2 and w_3. There was a slight reduction to the weight given to parent average from 0.1185 (animal 8) to 0.1151 (animal 4) due to the additional information from progeny.

The solution for pe of an animal can be calculated as in Section 4.2.2, using Eqn 4.4. Here, the correction of the TD records is for the estimates for HTD effects and fixed regressions and animal effect. Thus for cow 6, $\hat{\mathbf{pe}}_6$ can be calculated as:

$$\hat{\mathbf{pe}}_6 = \left(\mathbf{t}' \begin{bmatrix} 10.4 \\ 12.3 \\ 13.2 \\ 11.6 \\ 8.4 \end{bmatrix} - \begin{bmatrix} 3.1893 \\ 3.3099 \\ 3.3897 \\ 0.6751 \\ 0.0000 \end{bmatrix} - \begin{bmatrix} 10.0835 \\ 12.2001 \\ 12.6254 \\ 12.2077 \\ 11.5679 \end{bmatrix} - \begin{bmatrix} -0.8367 \\ -0.8367 \\ -0.8367 \\ -0.8367 \\ -0.8367 \end{bmatrix} \right) \Big/ 5.4380$$
$$= -9.1650/5.4380 = -1.6853$$

where $\mathbf{t}$ is a column vector of order 5 (number of TD records for the animal), with all elements equal to one. However, in contrast to pe estimates in Example 4.1, these pe estimates represent permanent environmental factors affecting TD records within lactation.

9.3 Random Regression Model

In Section 9.2, the advantage of including fixed regressions on days in milk in the model was to account for the shape of the lactation curve for different groups of cows. However, the breeding values estimated represented genetic differences between animals at the height of the curves. Although different residual variances

Downloaded From Vet-eBooks.com

associated with different stages of lactation could be fitted with the fixed regression model, the model did not account for the covariance structure at the genetic level. Schaeffer and Dekkers (1994) extended the fixed regression model for genetic evaluation by considering the regression coefficients on the same covariables as random, therefore allowing for between-animal variation in the shape of the curve. Thus the genetic differences among animals could be modelled as deviations from the fixed lactation curves by means of random parametric curves (see Guo and Swalve, 1997) or orthogonal polynomials such as Legendre polynomials (Brotherstone *et al*., 2000), or even non-parametric curves such as natural cubic splines (White *et al*., 1999). Most studies have used Legendre polynomials as they make no assumption about the shape of the curve and are easy to apply. The RR model has also been employed for the analysis of growth data in pigs (Andersen and Pedersen, 1996) and beef cattle (Meyer, 1999). An additional benefit of the RR model in dairy cattle is that it provides the possibility of genetic evaluation for persistence of the lactation. A typical random regression model (RRM) especially for the analysis of dairy cattle test day records is of the form:

$$y_{tijk} = htd_i + \sum_{k=0}^{nf} \phi_{jtk}\,\boldsymbol{\beta}_k + \sum_{k=0}^{nr} \phi_{jtk}\,\mathbf{u}_{jk} + \sum_{k=0}^{nr} \phi_{jtk}\,\mathbf{pe}_{jk} + e_{tijk}$$

where y_{tijk} is the test day record of cow j made on day t within HTD subclass i; $\boldsymbol{\beta}_k$ are fixed regression coefficients; $\mathbf{u}_{jk}$ and $\mathbf{pe}_{jk}$ are vectors of the kth random regression for animal and permanent environmental effects, respectively, for animal j; ϕ_{jtk} is the vector of the kth Legendre polynomials for the test day record of cow j made on day t; nf is the order of polynomials fitted as fixed regressions; nr is the order of polynomials for animal and **pe** effects; and e_{tljk} is the random residual. The model in matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Qu} + \mathbf{Zpe} + \mathbf{e}$$

The vectors **y**, **b** and the matrix **X** are as described in Example 9.1. However, **u** and **pe** are now vectors of random regressions for animal additive genetic and pe effects. The matrices **Q** and **Z** are covariable matrices and, if only animals with records are considered, the ith row of these matrices contains the orthogonal polynomials (covariables) corresponding to the DIM of the ith TD yield. If the order of fit is the same for animal and pe effects, **Q** = **Z**, considering only animals with records. This would not be the case if the order of fit is different for animal and pe effects. In general, considering animals with records, the order of either **Q** or **Z** is n_{td} (number of TD records) by nk, where nk equals nr times the number of animals with records. It is assumed that var(**u**) = **A*****G**, var(**pe**) = **I*****P** and var(**e**) = $\mathbf{I}\sigma_e^2$ = **R**, where **A** is the numerator relationship matrix, * is the Kronecker product and **G** and **P** are of the order of polynomial fitted for animal and pe effects. The MME are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Q} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Q}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} + \mathbf{A}^{-1} \otimes \mathbf{G} & \mathbf{Q}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Q} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{P} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{pe}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Q}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

Downloaded From Vet-eBooks.com

9.3.1 Numerical application

Example 9.2

Analysis of the data in Table 9.1 is undertaken fitting an RR model with Legendre polynomials of order 4 fitted for the fixed lactation curve and Legendre polynomials of order 2 fitted for both random animal and **pe** effects. The covariance matrices for the random regression coefficients for animal effect and **pe** effects are:

$$\mathbf{G} = \begin{bmatrix} 3.297 & 0.594 & -1.381 \\ 0.594 & 0.921 & -0.289 \\ -1.381 & -0.289 & 1.005 \end{bmatrix}; \quad \mathbf{P} = \begin{bmatrix} 6.872 & -0.254 & -1.101 \\ -0.254 & 3.171 & 0.167 \\ -1.101 & 0.167 & 2.457 \end{bmatrix}$$

and the residual variance equals 3.710 for all stages of lactation.

As indicated earlier, the above **G** or **P** matrix models the genetic or permanent environment covariance structure of fat yields over the whole lactation length. Thus the genetic covariance between DIM i and j along the trajectory can be calculated from **G**. For instance, the genetic variance for DIM i, (v_{ii}) can be calculated as:

$$v_{ii} = \mathbf{t}_i \mathbf{G} \mathbf{t}_i'$$

where $\mathbf{t}_i = \phi_{ik}$, the ith row vector of $\mathbf{\Phi}$, for day i, and k is the order of fit. The genetic covariance between DIM i and j (v_{ij}) therefore is:

$$v_{ij} = \mathbf{t}_i \mathbf{G} \mathbf{t}_j'$$

Using the **G** matrix in Example 9.1, the genetic variance for DIM 106 equals 2.6433 kg^2, with $\mathbf{t}_{106}$ = [0.7071 –0.4082 –0.5271], and the genetic covariance between DIM 106 and 140 equals 3.0219 kg, with $\mathbf{t}_{140}$ = [0.7071 –0.1361 –0.7613]. The plots of daily genetic and permanent environmental variances against DIM are shown in Fig. 9.1, indicating how these variances change through the lactation length.

Setting up the matrices for the MME

The setting of the matrix **X** has been described in Example 9.1. The matrix $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$ can easily be obtained by matrix multiplication. Considering only animals with records, **Q**′ can be represented as:

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{Q}_4' & 0 & 0 & 0 & 0 \\ 0 & \mathbf{Q}_5' & 0 & 0 & 0 \\ 0 & 0 & \mathbf{Q}_6' & 0 & 0 \\ 0 & 0 & 0 & \mathbf{Q}_7' & 0 \\ 0 & 0 & 0 & 0 & \mathbf{Q}_8' \end{bmatrix}$$

where $\mathbf{Q}_i'$ is the matrix of order nr by k (number of TD records for animal i). Thus for animal 6, $\mathbf{Q}_6'$ is:

$$\mathbf{Q}_6' = \begin{bmatrix} 0.7071 & 0.7071 & 0.7071 & 0.7071 & 0.7071 \\ -1.2247 & -0.9525 & -0.6804 & -0.4082 & -0.1361 \\ 1.5811 & 0.6441 & -0.0586 & -0.5271 & -0.7613 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

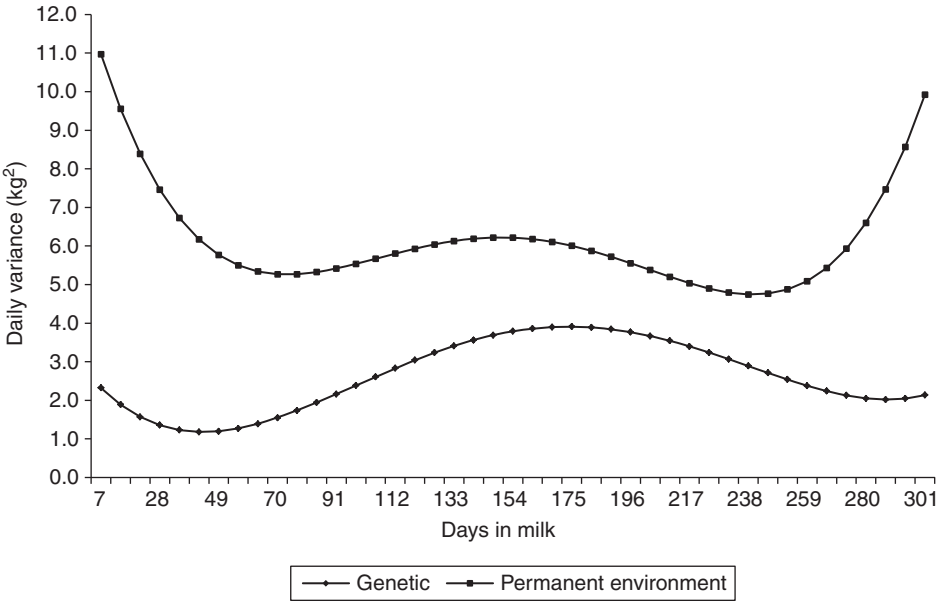


Fig. 9.1. The estimates of daily genetic and permanent environmental variances by days in milk.

For all animals with records, $\mathbf{Q'R^{-1}Q} = \mathbf{Z'R^{-1}Z}$ and are block diagonal. For instance, $\mathbf{Q'R^{-1}Q}$ for the first three cows (cows 4, 5 and 6) with records is:

$$\begin{bmatrix}
1.348 & 0.000 & 0.335 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.000 & 1.647 & 0.000 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.335 & 0.000 & 2.035 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1.348 & 0.000 & 0.335 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.000 & 1.647 & 0.000 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.335 & 0.000 & 2.035 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.674 & -0.648 & 0.167 \\
0 & 0 & 0 & 0 & 0 & 0 & -0.648 & 0.824 & -0.591 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.167 & -0.591 & 1.018
\end{bmatrix}$$

When all animals are considered, $\mathbf{Q'R^{-1}Q}$ is augmented by *nr* columns and rows per ancestor without records (i.e. animals 1–3). The matrix $\mathbf{G^{-1}}$ is then added to $\mathbf{Q'R^{-1}Q}$ and $\mathbf{P^{-1}}$ added to $\mathbf{Z'R^{-1}Z}$ to obtain the MME. Solving the MME by direct inversion with the solution for level 10 of HTD effects constrained to zero gave the following results:

Effects	Solutions
HTD	
1	10.0862
2	7.5908
3	8.5601
4	8.2430

Continued

Downloaded From Vet-eBooks.com

(*Continued*)

Effects	Solutions			
5	6.3161			
6	3.0101			
7	3.1085			
8	3.1718			
9	0.5044			
10	0.0000			
Fixed regression				
1	16.6384			
2	–0.6253			
3	–0.1346			
4	0.3479			
5	–0.4218			
Animal		Regression coeffients		305-day breeding value
1	–0.0583	0.0552	–0.0442	–12.3731
2	–0.0728	–0.0305	–0.0244	–15.7347
3	0.1311	–0.0247	0.0686	28.1078
4	0.3445	0.0063	–0.3164	74.8132
5	–0.4537	–0.0520	0.2798	–98.4153
6	–0.5485	0.0730	0.1946	–118.4265
7	0.8518	–0.0095	–0.3131	184.1701
8	0.2209	0.0127	–0.0174	47.6907
Permanent environmental effects				
Cow		Regression coefficients		305-day solutions
4	–0.6487	–0.3601	–1.4718	–138.4887
5	–0.7761	0.1370	0.9688	–168.5531
6	–1.9927	0.9851	–0.0693	–427.2378
7	3.5188	–1.0510	–0.4048	756.9415
8	–0.1013	0.2889	0.9771	–22.6619

The solutions for HTD and fixed regression for the RRM are similar to those from the fixed regression model. Lactation curves can be constructed from the fixed regression, as described in Section 9.2.1, and influences of different environmental factors on the curves can be evaluated. Each animal has *nr* regression coefficients as solutions for animal and permanent environmental effects. These are not useful for ranking animals and need to be converted to breeding values for any particular day of interest. Usually, in dairy cattle, values are calculated for 305-day yields and these have been shown above in the table of results. The EBV from days 6 to *m* for animal *k* (EBV_{km}) is calculated as:

$$EBV_{km} = \mathbf{t}\hat{\mathbf{u}}_k; \quad with \quad \mathbf{t} = t_j = \sum_{i=6}^{m}\sum_{j=0}^{nr}\phi_{ij} \tag{9.2}$$

where **t** is a row vector of order *nr*, with the *j*th elements equal to the sum of the *j*th orthogonal polynomial from days 6 to *m* and $\hat{\mathbf{u}}_k$ is vector for the regression coefficient of animal *k*. For Example 9.2, the matrix **Φ** for days 4 to 310 has not been shown because of the size but can be generated as described in Appendix G. Assuming 305-day breeding values are computed from days 6 to 310, then the vector **t** for Example 9.2 calculated from days 6 to 310 is:

t = [215.6655 2.4414 –1.5561]

Downloaded From Vet-eBooks.com

The breeding value for 305-day yield for animal 4, for instance, can be calculated as:

$$\mathbf{t}\hat{\mathbf{u}}_4 = [215.6655 \quad 2.4414 \quad -1.5561]\begin{pmatrix} 0.3445 \\ 0.0063 \\ -0.3164 \end{pmatrix} \approx 74.81$$

Over the lactation length, daily breeding values can be computed for each animal from the random regression coefficients. Genetic lactation curves can be obtained for each animal by plotting these daily breeding values against DIM and differences between curves for different animals can then be studied. Let **v** be a vector containing daily breeding values for days 6 to 310, then **v** can be calculated as:

$$\mathbf{v} = \mathbf{T}\hat{\mathbf{u}}_k; \quad \text{with} \quad \mathbf{T} = t_{ij} = \sum_{i=6}^{310}\sum_{j=0}^{nr}\phi_{ij}$$

The plots of the daily breeding values for animals 2, 3 and 8 are shown in Fig. 9.2. The plots indicate that the animal with the highest 305-day breeding value for fat yield also had the highest daily breeding values along the lactation length.

If the trait being analysed is milk yield, persistence breeding values can be calculated from the daily breeding values. For instance, persistence predicted transmitting ability (PS_{PTA}) for milk yield can be calculated (Schaeffer *et al.*, 2000) as:

$$PS_{PTA} = \frac{PTA_{280} - PTA_{60} + \bar{y}_{280}}{\bar{y}_{60}}(100)$$

where PTA_{60} and PTA_{280} are predicted transmitting abilities for day milk yield for an animal at days 60 and 280, respectively, and y_{60} and y_{280} are the average milk yields of cows in the genetic base at days 60 and 280, respectively.

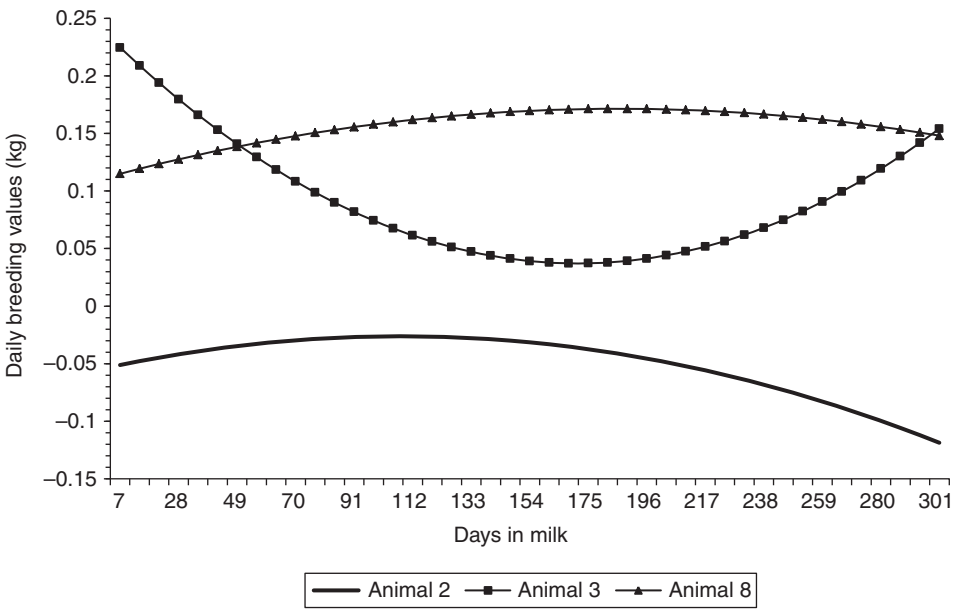


Fig. 9.2. The estimates of daily breeding values for some animals by days in milk.

Downloaded From Vet-eBooks.com

9.3.2 Partitioning animal solutions from random regression model

Equations for calculating the contribution of information from various sources to the solutions (random regression coefficients) of an animal from an RRM were presented by Mrode and Swanson (2002). These equations are the same as those presented in Section 5.2.3 for the multivariate model. Test day records of cows contribute to random regressions for the animal effect through the yield deviations. The calculation of the vector of yield deviations (**YD**) is first examined. Using the same argument for deriving Eqn 5.7, the equation for **YD** for an RRM is:

$$\mathbf{YD} = (\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q})^{-1}(\mathbf{Q}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\mathbf{p}\hat{\mathbf{e}})) \tag{9.3}$$

While this equation is similar to Eqn 5.6 for yield deviation under a multivariate model, here **YD** is a vector of weighted regressions of the animal's TD yields adjusted for all effects other than additive genetic effect, on orthogonal polynomials for DIM. Since **YD** is a vector of regressions, it can be used to generate actual yield deviations for any DIM using Eqn 9.2. Thus actual yield deviation (yd*) for day *m*, for instance, equals $\mathbf{v}'\mathbf{YD}$, where **v** is a vector of order *nr* with $v_m = \phi_{mj}$ and $j = 1,nr$. The actual yield deviation for 305-day yield can be calculated using Eqn 9.2 but with $\hat{\mathbf{u}}$ replaced with **YD**.

The calculation of **YD** for cow 6 in Example 9.2 is illustrated below. First, the vector of TD records for cow 6 corrected for all effects ($\mathbf{y}_c$) other than the additive genetic effect is:

$$\mathbf{y}_c = \mathbf{y}_6 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \hat{\mathbf{p}}\mathbf{e}$$

$$\mathbf{y}_c = \begin{bmatrix}10.4\\12.3\\13.2\\11.6\\8.4\end{bmatrix} - \begin{bmatrix}3.0101\\3.1085\\3.1718\\0.5044\\0.0000\end{bmatrix} - \begin{bmatrix}10.7725\\12.5295\\12.7890\\12.3454\\11.7641\end{bmatrix} - \begin{bmatrix}-2.7251\\-2.3920\\-2.0752\\-1.7746\\-1.4904\end{bmatrix} = \begin{bmatrix}-0.6576\\-0.9460\\-0.6856\\0.5249\\-1.8738\end{bmatrix}$$

where $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$ are vectors of solutions for HTD and fixed regression coefficients. The matrices $\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q}$ and $\mathbf{Q}'\mathbf{R}^{-1}\mathbf{y}_c$ are:

$$\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} = \begin{bmatrix}0.6738 & -0.6484 & 0.1674\\-0.6484 & 0.8235 & -0.5906\\0.1674 & -0.5906 & 1.0177\end{bmatrix} \quad \text{and} \quad \mathbf{Q}'\mathbf{R}^{-1}\mathbf{y}_c = \begin{bmatrix}-0.6934\\0.5967\\-0.1237\end{bmatrix}$$

Using Eqn 9.3, yield deviation for cow 6 ($\mathbf{YD}_6$) is:

$$\mathbf{YD}_6 = (\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{R}^{-1}\mathbf{y}_c = \begin{bmatrix}-5.0004\\-4.6419\\-1.9931\end{bmatrix}$$

Downloaded From Vet-eBooks.com

The actual yield deviation at 305 DIM for cow 6 using Eqn 9.2 with $\hat{\mathbf{u}}$ replaced with $\mathbf{YD}_6$ is –1086.6450.

The equation for the partitioning of random regression coefficients for animals to contributions for parent average, yield deviations and progeny is:

$$\hat{\mathbf{u}}_{anim} = \mathbf{W}_1\mathbf{PA} + \mathbf{W}_2(\mathbf{YD}) + \mathbf{W}_3\mathbf{PC} \tag{9.4}$$

with:

$$\mathbf{PC} = \sum \alpha_{prog} \left(2\,\hat{\mathbf{u}}_{prog} - \hat{\mathbf{u}}_{mate}\right) / \sum \alpha_{prog} \quad \text{and} \quad \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3 = \mathbf{I}$$

This is the same equation as Eqn 5.8, which partitioned breeding values under the multivariate model. The weights $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ are as defined in Eqn 5.8, but here $\mathbf{W}_i$ is of the order of orthogonal polynomials for animal effects. Illustrating with cow 6, the weights on parent average ($\mathbf{W}_1$) and yield deviation ($\mathbf{W}_2$) can be calculated as:

$$\mathbf{W}_1 = \underbrace{\begin{bmatrix} 2.1520 & -0.9957 & 2.0986 \\ -0.9957 & 3.2921 & -0.3580 \\ 2.0986 & -0.3580 & 5.7284 \end{bmatrix}^{-1}}_{(\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} + \mathbf{G}^{-1}\alpha_6)^{-1}} \underbrace{\begin{bmatrix} 1.4781 & -0.3473 & 1.9313 \\ -0.3473 & 2.4685 & 0.2326 \\ 1.9313 & 0.2326 & 4.7107 \end{bmatrix}}_{(2\mathbf{G}^{-1}\alpha_{par})}$$

$$= \underbrace{\begin{bmatrix} 0.6156 & 0.1940 & 0.2987 \\ 0.0935 & 0.8107 & 0.2402 \\ 0.1175 & 0.0202 & 0.7279 \end{bmatrix}}_{\mathbf{W}_1}$$

and:

$$\mathbf{W}_2 = \underbrace{\begin{bmatrix} 2.1520 & -0.9957 & 2.0986 \\ -0.9957 & 3.2921 & -0.3580 \\ 2.0986 & -0.3580 & 5.7284 \end{bmatrix}^{-1}}_{(\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} + \mathbf{G}^{-1}\alpha_6)^{-1}} \underbrace{\begin{bmatrix} 0.6738 & -0.6484 & 0.1674 \\ -0.6484 & 0.8235 & -0.5906 \\ 0.1674 & -0.5906 & 1.0177 \end{bmatrix}}_{(\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q})}$$

$$= \underbrace{\begin{bmatrix} 0.3844 & -0.1940 & -0.2987 \\ -0.0935 & 0.1893 & -0.2402 \\ -0.1175 & -0.0202 & 0.2721 \end{bmatrix}}_{\mathbf{W}_2}$$

The contributions from **PA** and **YD** to the random regression coefficients for cow 6 are:

$$\begin{bmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \mathbf{W}_1 \begin{bmatrix} -0.2560 \\ 0.0016 \\ 0.1178 \end{bmatrix} + \mathbf{W}_2 \begin{bmatrix} -5.0004 \\ -4.6419 \\ -1.9931 \end{bmatrix} = \begin{bmatrix} -0.1221 \\ 0.0057 \\ 0.0557 \end{bmatrix} + \begin{bmatrix} -0.4265 \\ 0.0674 \\ 0.1389 \end{bmatrix} = \begin{bmatrix} -0.5485 \\ 0.0730 \\ 0.1946 \end{bmatrix}$$

Downloaded From Vet-eBooks.com

For cow 8 with ten TD records and no progeny, Eqn 9.4 is:

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \underbrace{\begin{bmatrix}0.3893 & -0.0844 & 0.1763\\ -0.0604 & 0.5903 & 0.0353\\ 0.1576 & 0.0425 & 0.6379\end{bmatrix}}_{(\mathbf{W}_1)}\underbrace{\begin{bmatrix}0.3967\\ 0.0228\\ -0.1787\end{bmatrix}}_{(\mathbf{PA})} + \underbrace{\begin{bmatrix}0.6107 & 0.0844 & -0.1763\\ 0.0604 & 0.4097 & -0.0353\\ -0.1576 & -0.0425 & 0.3621\end{bmatrix}}_{(\mathbf{W}_2)}\underbrace{\begin{bmatrix}0.2102\\ 0.0574\\ 0.1893\end{bmatrix}}_{(\mathbf{YD}_8)}$$

and:

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \begin{bmatrix}0.1210\\ -0.0168\\ -0.0505\end{bmatrix} + \begin{bmatrix}0.0998\\ 0.0295\\ 0.0330\end{bmatrix} = \begin{bmatrix}0.2208\\ 0.0127\\ -0.0175\end{bmatrix}$$

Considering cow 4, with ten TD records and a progeny:

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \underbrace{\begin{bmatrix}0.3488 & -0.0684 & 0.1393\\ -0.0490 & 0.5132 & 0.0284\\ 0.1245 & 0.0343 & 0.5451\end{bmatrix}}_{(\mathbf{W}_1)}\underbrace{\begin{bmatrix}-0.0655\\ 0.0123\\ -0.0343\end{bmatrix}}_{(\mathbf{PA})} + \underbrace{\begin{bmatrix}0.5640 & 0.0856 & -0.1741\\ 0.0613 & 0.3585 & -0.0355\\ -0.1557 & -0.0428 & 0.3186\end{bmatrix}}_{(\mathbf{W}_2)}$$
$$\begin{bmatrix}0.2711\\ -0.0508\\ -0.6412\end{bmatrix}_{(\mathbf{YD})} + \underbrace{\begin{bmatrix}0.0872 & -0.0171 & 0.0348\\ -0.0123 & 0.1283 & 0.0071\\ 0.0311 & 0.0086 & 0.1363\end{bmatrix}}_{(\mathbf{W}_3)}\underbrace{\begin{bmatrix}1.5725\\ 0.0057\\ -0.6948\end{bmatrix}}_{(\mathbf{PC})}$$

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \begin{bmatrix}-0.0285\\ 0.0086\\ -0.0264\end{bmatrix} + \begin{bmatrix}0.2602\\ 0.0212\\ -0.2443\end{bmatrix} + \begin{bmatrix}0.1128\\ -0.0235\\ -0.0457\end{bmatrix} = \begin{bmatrix}0.3445\\ 0.0063\\ -0.3164\end{bmatrix}$$

Equation 9.4 is useful in explaining the evaluations for animals in terms of contributions from different sources of information, and how these contributions vary with different DIM could also be examined. However, Eqn 9.4 relates to random regression coefficients. Usually, the EBV at a particular stage of the longitudinal scale, such as 305 days for milk yield or body weight at 1 year of age, is published. Therefore, the interest might be in calculating the contributions from the various sources of information to the published EBV. Using milk yield as an example, the contribution to 305-day estimated BV from various sources of information can be calculated as:

$$\begin{aligned}\hat{\mathbf{u}}_{(305)anim} &= \mathbf{V}_1\mathbf{PA} + \mathbf{V}_2\mathbf{YD} + \mathbf{V}_3\mathbf{PC}\\ \hat{\mathbf{u}}_{(305)anim} &= \mathbf{PA}^* + \mathbf{YD}^* + \mathbf{PC}^*\end{aligned} \qquad (9.5)$$

Downloaded From Vet-eBooks.com

where $\mathbf{V}_i = \mathbf{DW}_i$, with **D** being a diagonal matrix such that $d_{ii} = t_i$, with t_i being the element of the row vector **t** in Eqn 9.2, $\mathbf{PA}^* = \mathbf{V}_1\mathbf{PA}$, $\mathbf{YD}^* = \mathbf{V}_2\mathbf{YD}$ and $\mathbf{PC}^* = \mathbf{V}_3\mathbf{PC}$. However, $\mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3 \neq \mathbf{I}$. Thus the estimated BV at 305 days ($BV_{(305)anim}$) from Eqn 9.5 is:

$$BV_{(305)anim} = \sum_{i=1}^{nr} \hat{u}_{(305)anim} = \sum_{i=1}^{nr} PA_i^* + \sum_{i=1}^{nr} YD_i^* + \sum_{i=1}^{nr} PC_i^*$$

where the contributions to the EBV at 305 days from PA, YD and PC are:

$$\sum_{i=1}^{nr} PA_i^*, \sum_{i=1}^{nr} YD_i^*, \text{ and } \sum_{i=1}^{nr} PC_i^*, \text{ respectively.}$$

Using Eqn 9.5, the contributions from various sources of information can be calculated for EBV at days or ages *j* to *n* along the longitudinal scale, and this could be plotted to examine how the contributions vary with days or age.

Using cow 6 in Example 9.2, the matrix **D** used in calculating the **V** terms in Eqn 9.5 is:

$$\mathbf{D} = \text{diag}(215.6655, 2.4414, -1.5561)$$

Using the $\mathbf{W}_1$ and $\mathbf{W}_2$ calculated earlier for cow 6:

$$\mathbf{V}_1 = \mathbf{DW}_1 = \begin{pmatrix} 132.7637 & 41.8391 & 64.4193 \\ 0.2283 & 1.9792 & 0.5864 \\ -0.1828 & -0.0314 & -1.1327 \end{pmatrix}$$

$$\mathbf{V}_2 = \mathbf{DW}_2 = \begin{pmatrix} 82.9018 & -41.8391 & -64.4193 \\ -0.2283 & 0.4622 & -0.5864 \\ 0.1828 & 0.0314 & -0.4234 \end{pmatrix} \quad \text{and}$$

$$\hat{\mathbf{u}}_{(305)6} = \mathbf{V}_1\mathbf{PA} + \mathbf{V}_2\mathbf{YD} = \begin{pmatrix} -26.3320 \\ 0.0138 \\ 0.1178 \end{pmatrix} + \begin{pmatrix} -91.9351 \\ 0.1649 \\ -0.2160 \end{pmatrix}$$

Therefore, contributions from **PA** and **YD** are –26.4049 and –91.9862, respectively, and:

$$BV_{(305)6} = -26.4049 + 91.9862 = -118.3911$$

Thus contribution from parent average is about 22% of the EBV at 305 days. The EBV at 305 days calculated above is slightly different from the value of –118.4265 shown earlier, due to rounding.

9.3.3 Calculating daughter yield deviations

The equation for calculating daughter yield deviation under an RRM is the same as Eqn 5.12 presented for the multivariate models. However, with the RRM, **DYD** in Eqn 5.12 is a vector of random regression coefficients and the weights $\mathbf{M}_1$, $\mathbf{M}_2$ and

Downloaded From Vet-eBooks.com

$\mathbf{M}_3$ are of the order *nr*. Actual daughter yield deviation for any DIM can be generated using Eqn 9.2.

As indicated in Section 5.2, for ease of computation, $\mathbf{W}_{2prog}$ in Eqn 5.12 is pre-multiplied with $\mathbf{G}^{-1}$, such that the equation for **DYD** becomes:

$$\mathbf{DYD} = \sum \mathbf{G}^{-1}\mathbf{W}_{2prog}\alpha_{prog}\left(2\mathbf{YD} - \hat{\mathbf{u}}_{mate}\right) / \mathbf{G}^{-1}\mathbf{W}_{2prog}\alpha_{prog}$$

9.3.4 Reliability of breeding values

The reliability of an EBV depends on its prediction error variance (PEV) relative to the genetic variance. It can therefore be regarded as a statistic summarizing the value of information available in calculating the EBV. The published EBV from an RR model is usually a linear function of the random regression coefficients obtained by solving the MME. The principles for calculating PEV and reliability under this situation are presented using the diagonal elements of the inverse of the coefficient matrix of the MME for Example 9.2.

Let $\mathbf{k}'\mathbf{u}$ define the EBV for the trait of interest for animal *i* from the RR model. The vector $\mathbf{k} = \mathbf{w}_i\mathbf{t}$, where $\mathbf{w}_i$ might be the weighting factor for the *i*th age or lactation if the study was on body weight at several ages or fat yield in different lactations analysed as different traits. For instance, if fat yields in lactations 1 and 2 were analysed as different traits, $\mathbf{w}_i'$ might be [0.70 0.3], indicating a weight of 0.7 and 0.3, respectively, for first and second lactation EBV. The vector **t** defines how within lactation EBV was calculated and is the same as in Eqn 9.2. For Example 9.2, *k* is a scalar with a value of 1. Given that **G** is the additive genetic covariance matrix for random regression effect for animal effects and **P** is the covariance matrix for pe effects, then the additive genetic variance of $\mathbf{k}'\mathbf{u} = g = \mathbf{k}'\mathbf{G}\mathbf{k}$ and the variance for the pe effect for the trait of interest = $p = \mathbf{k}'\mathbf{P}\mathbf{k}$. The heritability of $\mathbf{k}'\mathbf{u}$ can therefore be calculated as $(g/(g + p + e)$ and $\alpha = (4 - h^2)/h^2$.

Let $\mathbf{C}^{ii}$ be the subset of the inverse of MME corresponding to the genetic effect for the *i*th animal. Then for animal *i*, prediction error variance (PEV_i) = $\mathbf{k}'\mathbf{C}^{ii}\mathbf{k}$. The reliability of $\mathbf{k}'\mathbf{u}$ can therefore be calculated as $1 - PEV_i/g$. As an illustration, in Example 9.2, $\mathbf{k}' = \mathbf{w}\mathbf{T}$ = [215.6655 2.4414 −1.5561], $g = \mathbf{k}'\mathbf{G}\mathbf{k}$ = 154896.766 kg^2, $p = \mathbf{k}'\mathbf{P}\mathbf{k}$ = 323462.969 kg^2 and h^2 = 0.32. For animal 1, the matrix $\mathbf{C}^{11}$ is:

$$\mathbf{C}^{11} = \begin{bmatrix} 2.9911 & 0.5159 & -1.2295 \\ 0.5159 & 0.8683 & -0.2480 \\ -0.2295 & -0.2480 & 0.9183 \end{bmatrix}$$

and:

$$PEV_1 = \mathbf{k}'\mathbf{C}^{11}\mathbf{k} = 140499.97$$

Therefore, reliability for animal 1 equals 1 – 140499.97/154896.766 = 0.09. The reliabilities for the animals in Example 9.2 are:

Downloaded From Vet-eBooks.com

Animal	Reliability
1	0.09
2	0.04
3	0.07
4	0.12
5	0.15
6	0.06
7	0.10
8	0.05

In practice, calculating the inverse of the MME is not feasible for large populations and PEV has to be approximated. As indicated earlier, EBV from RR models are linear functions of the random regressions; therefore, methods to approximate reliabilities should simultaneously approximate PEV and the prediction error covariance (PEC) among the individual random regressions (Liu *et al*., 2002; Meyer and Tier, 2003). Such an approximation method presented by Meyer and Tier (2003) is outlined in Appendix D, Section D.2.

9.3.5 Random regression models with spline function

Random regression models with Legendre polynomials have been considered to have better convergence properties as the regressions are orthogonal. However, some studies have reported high genetic variances at the extremes of the lactation and negative correlations between the most distant test days. In order to overcome this limitation, some workers have fitted RRM using splines (Misztal, 2006; Bohmanova *et al*., 2008). Splines are piece-wise functions consisting of independent segments that are connected in knots. The segments are described by lower-order polynomials. Linear splines are the simplest spline function where the segments are fitted by linear polynomials between two knots adjacent to the record and zero between all other knots. Thus the system of equations is sparse as only two coefficients are non-zero for a given record. The use of cubic splines for the modelling of the lactation curve has also been presented by White *et al*. (1999). However, the linear spline is considered in this section.

Let **T** be a vector of n knots, then the covariables of the linear spline for DIM t ($\Phi_i(t)$) located between knots T_i and T_{i+1} can be calculated as:

$$\Phi_i(t) = (t - T_i)/(T_{i+1} - T_i)$$
$$\Phi_{i+1}(t) = (T_{i+1} - t)/(T_{i+1} - T_i)$$
$$= 1 - \Phi_i(t) \text{ and } \Phi_{1\ldots i-1,\, i+2\ldots n} = 0$$

If $t = T_i$, $\Phi_i(t) = 1$ and $\Phi_{1\ldots i-1,\, i+1\ldots n} = 0$.

Thus the vector $\mathbf{\Phi}$ for DIM t has at most two non-zero elements, which sum up to one. The above formula assumes that $T_i \leq t < T_n$. If, however, $t < T_i$ or $t > T_n$, the following can be used and the sum of the elements of the vector will not sum up to one:

if $t < T_1$, $\Phi_1(t) = t/T_1$ and $\Phi_{1+\,i\ldots n} = 0$
if $t > T_n$, $\Phi_n(t) = T_S/t$ and $\Phi_{1\ldots n-1} = 0$

Using the data in Example 9.1, assume that the four knots are fitted for the fixed lactation curve and knots are placed at days 4, 106, 208 and 310, the covariables for the spine function for particular DIM are as follows:

Downloaded From Vet-eBooks.com

DIM	Φ_0	Φ_1	Φ_2	Φ_3
4	1.000	0.000	0.000	0.000
38	0.333	0.667	0.000	0.000
72	0.667	0.333	0.000	0.000
106	0.000	1.000	0.000	0.000
140	0.000	0.333	0.667	0.000
174	0.000	0.667	0.333	0.000
208	0.000	0.000	1.000	0.000
242	0.000	0.000	0.333	0.667
276	0.000	0.000	0.667	0.333
310	0.000	0.000	0.000	1.000

As an illustration, the covariates for DIM 38 can be computed as:

$\Phi_1(38) = (38 - 4)/106 - 4) = 0.333$ and $\Phi_2(38) = 1 - 0.333 = 0.667$

Thus $\mathbf{\Phi}(38) = [0.333, 0.667, 0\ 0]$

A random regression model can therefore be fitted as:

$$y_{tijk} = htd_i + \sum_{k=0}^{nf} \phi_{jtk}\,\boldsymbol{\beta}_k + \sum_{k=0}^{nr} \phi_{jtk}\,\mathbf{u}_{jk} + \sum_{k=0}^{nr} \phi_{jtk}\,\mathbf{pe}_{jk} + e_{tij}$$

where all terms are as defined in Section 9.3 but the ϕ_{jtk} is the vector of the kth spline function for the test day record of cow j made on day t. The same procedure described in Section 9.3.1 can be used in the application of the model for the analysis of data and interpretation of results.

9.3.6 Random regression model for maternal traits

Maternal genetic effects are important in growth traits in beef cattle, and models that account for these effects have been discussed in Chapter 7. However, the RR model could also be augmented to include random regressions for maternal genetic and maternal permanent environmental effects. Albuquerque and Meyer (2001) examined different orders of fit for the random regressions for both effects. One of the favoured models was the one in which the order of Legendre polynomials for direct genetic, maternal genetic, animal pe and maternal pe effects were 5, 5, 5 and 3, respectively.

Such a model, excluding all fixed effects, could be written as:

$$y_{ijktd} = \sum_{i=0}^{k_1-1} \phi_{jti}\,\mathbf{u}_{ji} + \sum_{i=0}^{k_2-1} \phi_{jti}\,\mathbf{m}_{ji} + \sum_{i=0}^{k_3-1} \phi_{jti}\,\mathbf{pe}_{ji} + \sum_{i=0}^{k_4-1} \phi_{dti}\,\mathbf{pp}_{di} + e_{ijktd}$$

where y_{ijktd} is the body weight of cow j taken at age t that has a dam d; $\mathbf{u}_{ji}$, $\mathbf{m}_{ji}$ and $\mathbf{pe}_{ji}$ are the random regressions for direct, maternal genetic and animal pe effects for animal j, respectively; $\mathbf{pp}_{di}$ is the random regression for dam pe effects and e_{ijktd} is random error; ϕ_{jti} and ϕ_{dti} are the vector of the ith Legendre polynomial for body weight at age t for cow j and dam d, respectively. They assumed a zero covariance between direct and maternal genetic effects to simplify the computation. The variance for direct effects increased from birth to 365 days while maternal genetic variance increased from birth to about 115 days and decreased thereafter.

Downloaded From Vet-eBooks.com

9.4 Covariance Functions

Kirkpatrick *et al*. (1990, 1994) introduced the concept of analysing repeated records taken along a trajectory such as time or age by means of covariance functions. In view of the fact that such a trait can take on a value at each of an infinite number of ages and its value at each age can be regarded as a distinct trait, the trajectory for such a trait could be regarded as an infinite-dimensional trait. Thus the growth trajectory or milk yield trajectory of an individual could be represented by a continuous function. Covariance function describes the covariance structure of an infinite-dimensional character as a function of time. Therefore, the covariance function is the infinite-dimensional equivalent of a covariance matrix for a given number of records taken over time at different ages. The value of the phenotypic covariance function, $þ(t_i, t_j)$, gives the phenotypic covariance between the value of the trait at ages t_i and t_j. Similarly, the value of the additive genetic covariance function, $f(t_i, t_j)$, gives the additive genetic covariance between the value of the trait at ages t_i and t_j. In mathematical terms, given t ages, the covariance between breeding values u_l and u_m on an animal at ages a_l and a_m could be written as:

$$\text{cov}(u_l, u_m) = f(a_l, a_m) = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \phi_i(a_l)\phi_j(a_m)C_{ij} \tag{9.6}$$

$$= \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \tau_{ij} a_l^i a_m^j \tag{9.7}$$

where f with factors τ_{ij} is the covariance function (CF), **C** is the coefficient matrix associated with the CF with elements C_{ij}, a_l is the lth age standardized to the intervals for which the polynomials are defined and k is the order of fit. Kirkpatrick *et al*. (1990, 1994) used Legendre polynomials that span the interval –1 to + 1. The ages can be standardized as described in Appendix G.

Given that **G** is the observed genetic covariance matrix of order t, and assuming a full order polynomial fit ($k = t$), Eqn 9.6 can be written in matrix notation as:

$$\hat{\mathbf{G}} = \mathbf{\Phi}\,\hat{\mathbf{C}}\,\mathbf{\Phi}' \tag{9.8}$$

and $\hat{\mathbf{C}}$ can be estimated as:

$$\hat{\mathbf{C}} = \mathbf{\Phi}^{-1}\hat{\mathbf{G}}\,(\mathbf{\Phi}^{-1}) \tag{9.9}$$

where $\mathbf{\Phi}$ is the matrix of Legendre polynomials of order t by k with element $\phi_{ij} = \phi_j(a_t)$ = the jth polynomial evaluated at standardized age t.

As an illustration, assume body weight measurements in beef cattle have been taken at three different ages, 90, 160 and 240 months old, and that the genetic covariance matrix ($\hat{\mathbf{G}}$) estimated was:

$$\hat{\mathbf{G}} = \begin{bmatrix} 132.3 & 127.0 & 136.6 \\ 127.0 & 172.8 & 200.8 \\ 136.6 & 200.8 & 288.0 \end{bmatrix}$$

Using the method described in Appendix G, the vector of standardized ages is:

$\mathbf{a}'$ = [–1.0 –0.0667 1.000]

Downloaded From Vet-eBooks.com

and **M** becomes:

$$\mathbf{M} = \begin{bmatrix} 1.0000 & -1.0000 & 1.0000 \\ 1.0000 & -0.0667 & 0.0044 \\ 1.0000 & 1.0000 & 1.0000 \end{bmatrix}$$

Thus for $t = 3$, $\mathbf{\Lambda}$ is:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.7071 & 0.0000 & -0.7906 \\ 0.0000 & 1.2247 & 0.0000 \\ 0.0000 & 0.0000 & 2.3717 \end{bmatrix}$$

and $\mathbf{\Phi}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} 0.7071 & -1.2247 & 1.5811 \\ 0.7071 & -0.0816 & -0.7801 \\ 0.7071 & 1.2247 & 1.5811 \end{bmatrix}$$

and from Eqn 9.9, the coefficient matrix $\hat{\mathbf{C}}$ is:

$$\hat{\mathbf{C}} = \begin{bmatrix} 344.7117 & 45.2787 & -3.2062 \\ 45.2787 & 24.5185 & -0.1475 \\ -3.2062 & -0.1475 & 3.2768 \end{bmatrix}$$

The covariance between two different ages can be calculated using Eqn 9.8. For instance, the variances at days 90 and 200 of body weight and the covariance between body weight on both days are $\mathbf{\Phi}_{90}\hat{\mathbf{C}}\mathbf{\Phi}'_{90} = 132.30$, $\mathbf{\Phi}_{200}\hat{\mathbf{C}}\mathbf{\Phi}'_{200} = 218.50$, $\mathbf{\Phi}_{90}\hat{\mathbf{C}}\mathbf{\Phi}'_{200} = 129.71$, respectively, with:

$$\mathbf{\Phi}_{90} = \mathbf{m}_{90}\mathbf{\Lambda} = [0.7071 \ -1.2247 \ 1.5811]$$

and:

$$\mathbf{\Phi}_{200} = \mathbf{m}_{200}\mathbf{\Lambda} = [0.7071 \ 0.5716 \ -0.2740]$$

where $\mathbf{m}_i$ are the appropriate row vectors of the matrix **M**.

Also, from Eqn 9.8 and Appendix G, $\hat{\mathbf{G}}$ can be written as:

$$\hat{\mathbf{G}} = \mathbf{M}\mathbf{\Lambda}\hat{\mathbf{C}}\mathbf{\Lambda}'\mathbf{M}'$$

Therefore, $\hat{\mathbf{G}} = \mathbf{MTM}'$ with $\mathbf{T} = \mathbf{\Lambda}\hat{\mathbf{C}}\mathbf{\Lambda}$ or calculated as $\mathbf{T} = \mathbf{M}^{-1}\hat{\mathbf{G}}(\mathbf{M}^{-1})'$, where **T** is the matrix with elements τ_{ij} in Eqn 9.7. Substituting **T** in Eqn 9.7 the full estimate of the CF, $f(a_l, a_m)$, can be obtained. Using the example data:

$$\mathbf{T} = \begin{pmatrix} 177.99 & 39.35 & -11.52 \\ 39.35 & 36.78 & -0.43 \\ -11.52 & -0.43 & 18.43 \end{pmatrix}$$

Therefore, the full estimate of the covariance function, $f(a_l, a_m)$, is:

$$f(a_l, a_m) = 177.99 + 39.35(a_l + a_m) + 36.78a_l a_m - 11.52(a_l^2 + a_m^2) - 0.43(a_l^2 a_m + a_l a_m^2) + 18.43a_l^2 a_m^2$$

Downloaded From Vet-eBooks.com

The application of CF in genetic evaluation involves defining an equivalent model using Eqn 9.8. For instance, using the example of the body weight of beef cattle, assume that the multivariate model for observations measured on one animal is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{a} + \mathbf{e}$$

where **y**, **X**, **b**, **a** and **e** are vectors defined as in Eqn 5.1 with $i = t$, with var(**a**) = $\breve{\mathbf{G}}$ and var(**e**) = **R**. Assuming a CF has also been fitted for the covariance matrix for environmental effects with a term included to account for measurement error, then:

$$\mathbf{R} = \mathbf{\Phi C}_p\mathbf{\Phi}' + \mathbf{I}\sigma^2\varepsilon$$

where $\mathbf{C}_p$ contains the coefficient matrix associated with the CF for pe and variance $\boldsymbol{\varepsilon}$ is $\mathbf{I}\sigma^2\varepsilon$. Using this equation and Eqn 9.8, an equivalent model to the multivariate model can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{\Phi u} + \mathbf{\Phi pe} + \boldsymbol{\varepsilon}$$

where **u** and **pe** are now vectors of random regression coefficients for random animal and pe effects. Then var(**u**) = $\mathbf{\Phi C\Phi}'$ and var(**pe**) = $\mathbf{\Phi C}_p\mathbf{\Phi}'$. The application of the above model in genetic evaluation is illustrated in Example 9.2. Thus the breeding value a_n for any time n can be calculated as:

$$a_n = \sum_{i=0}^{k-1} \phi_i(t_n)\mathbf{u}_i$$

where $\phi(t_n)$ is the vector of Legendre polynomial coefficients evaluated at age t_n. Thus with a full order fit, the covariance function model is exactly equivalent to the multivariate model. However, in practice, the order of fit is chosen such that the estimated covariance matrix can be appropriately fitted with as few parameters as possible. In the next section, the fitting of a reduced-order CF is discussed.

9.4.1 Fitting a reduced order covariance function

Equation 9.8 and the illustration given in the above section assumed a full-order polynomial fit of **G** ($k = t$). Therefore, it was possible to get an inverse of $\mathbf{\Phi}$ and hence estimate **C**. However, for a reduced-order ($k < t$) fit, $\mathbf{\Phi}$ has only k columns and a direct inverse may not be possible. With the reduced fit, the number of coefficients to be estimated are reduced to $k(k + 1)/2$. This is particularly important for large $\mathbf{\Lambda}$, such as test day milk yield within a lactation with t equal to 10 or 305 assuming monthly or daily sampling and requiring $t(t + 1)/2$ coefficients to be estimated. Thus a reduced order fit with k substantially lower than t could be very beneficial.

Kirkpartrick *et al.* (1990) proposed weighted least squares as an efficient method of obtaining an estimate of the reduced coefficient matrix ($\breve{\mathbf{C}}$) from the linear function of the elements of $\breve{\mathbf{G}}$. They outlined the following steps for the weighted least-square procedure. The procedure is illustrated using the example $\breve{\mathbf{G}}$ for the body weight in beef cattle given earlier, fitting polynomials of order one, i.e. only the first two Legendre polynomials are fitted, thus $k = 2$. Initially, a vector $\breve{\mathbf{g}}$ of order t^2 is formed by stacking the successive columns of $\breve{\mathbf{G}}$. Thus:

$$\breve{\mathbf{g}}' = [\breve{G}_{11},\ldots,\breve{G}_{n1}, \breve{G}_{12},\ldots,\breve{G}_{n2}, \breve{G}_{1n},\ldots,\breve{G}_{nn}]$$

Downloaded From Vet-eBooks.com

Thus for the example $\breve{\mathbf{G}}$:

$\breve{\mathbf{g}}' = [132.3 \quad 127.0 \quad 136.6 \quad 172.8 \quad 200.8 \quad 288.0]$

Define $\mathbf{\Phi}_r$ of order t by k, obtained by deleting $(t - k)$ columns of $\mathbf{\Phi}$ corresponding to those ϕ_j not in the reduced-order fit. The relationship between the observed covariance matrix, $\breve{\mathbf{g}}$, and the coefficient matrix of the reduced fit to be estimated is given by the following regression equation:

$$\breve{\mathbf{g}} = \mathbf{X}_s \breve{\mathbf{c}} + \mathbf{e} \tag{9.10}$$

where $\mathbf{e}$ is the vector of the difference between observed covariances and those predicted by the covariance function, $\breve{\mathbf{c}}$ is a vector of dimension k^2, containing the elements of the coefficient matrix of the reduced fit ($\breve{\mathbf{C}}$). The order of elements of $\breve{\mathbf{C}}$ in $\breve{\mathbf{c}}$ is the same as in $\breve{\mathbf{g}}$: that is, $\breve{\mathbf{c}} = (\breve{C}_{00}, \ldots, \breve{C}_{k0}, \ldots, \breve{C}_{k1}, \ldots, \breve{C}_{kk})$. $\mathbf{X}_s$ is the Kronecker product of $\mathbf{\Phi}_r$ with itself ($\mathbf{X}_s = \mathbf{\Phi}_r \times \mathbf{\Phi}_r$) and is of the order t^2 by k^2. Since only the first two polynomials are fitted, the matrix $\mathbf{\Phi}_r$ can be derived by deleting from $\mathbf{\Phi}$ the third column, corresponding to the missing second-degree polynomial. Thus for the beef cattle example:

$$\mathbf{\Phi}_r = \begin{bmatrix} 0.7071 & -1.2247 \\ 0.7071 & -0.0816 \\ 0.7071 & 1.2247 \end{bmatrix}$$

and $\mathbf{X}_s$ is:

$$\mathbf{X}_s = \begin{bmatrix} 0.5000 & -0.8660 & -0.8660 & 1.4999 \\ 0.5000 & -0.0577 & -0.8660 & 0.0999 \\ 0.5000 & 0.8660 & -0.8660 & -1.4999 \\ 0.5000 & -0.8660 & -0.0577 & 0.0999 \\ 0.5000 & -0.0577 & -0.0577 & 0.0067 \\ 0.5000 & 0.8660 & -0.0577 & -0.0999 \\ 0.5000 & -0.8660 & 0.8660 & -1.4999 \\ 0.5000 & -0.0577 & 0.8660 & -0.0999 \\ 0.5000 & 0.8660 & 0.8660 & 1.4999 \end{bmatrix}$$

The application of weighted least squares to obtain solutions for $\breve{\mathbf{c}}$ in Eqn 9.10 requires the covariance matrix ($\mathbf{V}$) of sampling errors of $\breve{\mathbf{g}}$. Kirkpatrick *et al.* (1990) presented several methods for estimating $\mathbf{V}$, examining three different experimental designs. However, in animal breeding, most estimates of $\breve{\mathbf{G}}$ are from field data and may not fit strictly to the designs they described, but estimates of sampling variances from REML analysis could be used. For the example $\breve{\mathbf{G}}$ for the beef cattle data, $\mathbf{V}$ has been estimated using the formula given by Kirkpatrick *et al.* (1990) for a half-sib design, assuming that 60 sires were each mated to 20 dams. The mean cross-product for the residual effect ($\hat{W}_e$) was estimated as $\hat{W}_{e,ij} = P_{ij} - 0.25\breve{G}_{ij}$ and that among sires ($\hat{W}_a$) as $\hat{W}_{a,ij} = (n - 1/4)\breve{G}_{ij} + P_{ij}$, where P_{ij} is the phenotypic variance and n is the number of dams. Sampling variance for $\breve{\mathbf{g}}$ was then calculated as: $\mathbf{V} = (16/n^2)[\text{cov}(\hat{W}_{a,ij}, \hat{W}_{a,kl}) + \text{cov}(\hat{W}_{e,ij}, \hat{W}_{e,kl})]$, where $\text{cov}(\hat{W}_{ij}, \hat{W}_{kl}) = (\hat{W}_{ik}\hat{W}_{jl} + \hat{W}_{il}\hat{W}_{jk})/\text{df}$, with df = number of degrees of freedom

Downloaded From Vet-eBooks.com

plus 2. In estimating cov($\hat{W}_{a,ij}$, $\hat{W}_{a,kl}$) and cov ($\hat{W}_{e,ij}$, $\hat{W}_{e,kl}$), df = (s – 1) + 2 and $s(n$ – 1) + 2, respectively. The estimated **V** therefore is:

$$\hat{\mathbf{V}} = \begin{bmatrix} 3450.0 & 2256.4 & 2184.6 & 2256.4 & 1480.3 & 1434.7 & 2184.6 & 1434.7 & 1390.9 \\ 2256.4 & 2959.6 & 2430.9 & 2959.6 & 2903.5 & 2530.2 & 2430.9 & 2530.2 & 2180.1 \\ 2184.6 & 2430.9 & 3889.7 & 2430.9 & 2249.1 & 3181.6 & 3889.7 & 3181.6 & 4051.5 \\ 2256.4 & 2959.6 & 2430.9 & 2959.6 & 2903.5 & 2530.2 & 2430.9 & 2530.2 & 2180.1 \\ 1480.3 & 2903.5 & 2249.1 & 2903.5 & 5711.4 & 4410.0 & 2249.1 & 4410.0 & 3417.5 \\ 1434.7 & 2530.2 & 3181.6 & 2530.2 & 4410.0 & 5818.8 & 3181.6 & 5818.8 & 6354.3 \\ 2184.6 & 2430.9 & 3889.7 & 2430.9 & 2249.1 & 3181.6 & 3889.7 & 3181.6 & 4051.5 \\ 1434.7 & 2530.2 & 3181.6 & 2530.2 & 4410.0 & 5818.8 & 3181.6 & 5818.8 & 6354.3 \\ 1390.9 & 2180.1 & 4051.5 & 2180.1 & 3417.5 & 6354.3 & 4051.5 & 6354.3 & 11835.0 \end{bmatrix}$$

However, the symmetry of $\breve{\mathbf{G}}$ resulted in redundancies in the vector $\breve{\mathbf{g}}$ such that **V** is singular. The vector $\breve{\mathbf{g}}$ can be redefined to be of the order s by 1, which contains only the elements in the lower half of $\breve{\mathbf{G}}$, where $s = t(t + 1)/2$. Therefore, delete from $\breve{\mathbf{g}}$ the elements $\breve{\mathbf{G}}_{ij}$ for which $i < j$. Thus for the example $\breve{\mathbf{G}}$, the vector $\breve{\mathbf{g}}$ becomes:

$$\breve{\mathbf{g}} = [132.3 \quad 127.0 \quad 136.6 \quad 172.8 \quad 200.8 \quad 288.0]$$

Then delete from **V** those columns and rows corresponding to elements $\breve{\mathbf{G}}_{ij}$ with $i < j$. This involves deleting rows and columns 4, 7 and 8 from the matrix **V** given above. The **V** of reduced order (s by s) is:

$$\hat{\mathbf{V}} = \begin{bmatrix} 3450.0 & 2256.4 & 2184.6 & 1480.3 & 1434.7 & 1390.9 \\ 2256.4 & 2959.6 & 2430.9 & 2903.5 & 2530.2 & 2180.1 \\ 2184.6 & 2430.9 & 3889.7 & 2249.1 & 3181.6 & 4051.5 \\ 1480.3 & 2903.5 & 2249.1 & 5711.4 & 4410.0 & 3417.5 \\ 1434.7 & 2530.2 & 3181.6 & 4410.0 & 5818.8 & 6354.3 \\ 1390.9 & 2180.1 & 4051.5 & 3417.5 & 6354.3 & 11835.0 \end{bmatrix}$$

Similarly, the rows corresponding to those elements of $\breve{\mathbf{g}}$ for which $\breve{\mathbf{G}}_{ij}$ has $i < j$ are deleted from $\mathbf{X}_s$. In the example $\mathbf{X}_s$, rows 4, 7 and 8 are deleted. Thus $\mathbf{X}_s$ becomes:

$$\mathbf{X}_s = \begin{bmatrix} 0.5000 & -0.8660 & -0.8660 & 1.4999 \\ 0.5000 & -0.0577 & -0.8660 & 0.0999 \\ 0.5000 & 0.8660 & -0.8660 & -1.4999 \\ 0.5000 & -0.0577 & -0.0577 & 0.0067 \\ 0.5000 & 0.8660 & -0.0577 & -0.0999 \\ 0.5000 & 0.8660 & 0.8660 & 1.4999 \end{bmatrix}$$

Also, for each element of $\breve{\mathbf{c}}$ for which $\breve{\mathbf{C}}_{ij}$ has $i < j$, add the corresponding column of $\mathbf{X}_s$ to the column corresponding to $\breve{\mathbf{C}}_{ji}$, then delete the former column. For the beef cattle example, the vector of coefficients, $\breve{\mathbf{c}}' = [\breve{C}_{00} \ \breve{C}_{10} \ \breve{C}_{01} \ \breve{C}_{11}]$. Therefore, the third column of $\mathbf{X}_s$ corresponding to $\breve{C}_{01}$ is added to the second column and the third column is deleted. The matrix $\mathbf{X}_s$ then becomes:

Downloaded From Vet-eBooks.com

$$\mathbf{X}_s = \begin{bmatrix} 0.5000 & -1.7320 & 1.4999 \\ 0.5000 & -0.9237 & 0.0999 \\ 0.5000 & 0.0000 & -1.4999 \\ 0.5000 & -0.1154 & 0.0067 \\ 0.5000 & 0.8083 & -0.0999 \\ 0.5000 & 1.7320 & 1.4999 \end{bmatrix}$$

Finally, delete from $\check{\mathbf{c}}$ the elements for which $\check{\mathbf{C}}$ has elements $i < j$. The matrix $\check{\mathbf{c}}$ is now of the order $k(k + 1)/2$ by 1. For the example data, $\check{\mathbf{c}}' = [\check{C}_{00}\ \check{C}_{10}\ \check{C}_{11}]$. The vector $\check{\mathbf{c}}$ can now be calculated by a weighted least-square procedure as:

$$\check{\mathbf{c}} = (\mathbf{X}_s' \hat{\mathbf{V}}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s' \hat{\mathbf{V}}^{-1} \check{\mathbf{g}}$$

For the example data, $\check{\mathbf{c}}$ calculated using the above equation is:

$$\check{\mathbf{c}} = [341.8512 \quad 45.0421 \quad 24.5405]$$

The reduced coefficient matrix $\check{\mathbf{C}}$ is then constructed from the calculated $\check{\mathbf{c}}$. Then a row and column of zeros are inserted in positions corresponding to those polynomials not included to obtain $\check{\mathbf{C}}$. For the example data, $\check{\mathbf{C}}$ is now:

$$\check{\mathbf{C}} = \begin{bmatrix} 341.8512 & 45.0421 & 0.0 \\ 45.0421 & 24.5405 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$$

Kirkpatrick *et al*. (1990) presented the following chi-square statistic to test the goodness of fit of the reduced covariance function to $\check{\mathbf{G}}$:

$$\chi^2_{(m-p)} = (\check{\mathbf{g}} - \mathbf{X}_s \hat{\mathbf{c}})' \hat{\mathbf{V}}^{-1} (\check{\mathbf{g}} - \mathbf{X}_s \hat{\mathbf{c}})$$

where $m = t(t + 1)/2$ is the number of degrees of freedom in $\check{\mathbf{G}}$ and $p = (k(k + 1)/2$ is the number of parameters being fitted. A significant result indicates that the model is inconsistent with the data, and a higher order of fit may be needed. For the beef cattle example, the value of χ^2 was 0.2231 with $m = 6$ and $p = 3$. This value of χ^2 was not significant with three degrees of freedom and thus the reduced covariance function was not significantly different from $\check{\mathbf{G}}$.

Another method of fitting a reduced-order CF, proposed by Mantysaari (1999), involved eigenvalue decomposition of the coefficient matrix. The largest k eigenvalues of $\hat{\mathbf{C}}$ in Eqn 9.9, for instance, are kept in a diagonal matrix ($\mathbf{D}_a$) and the matrix $\mathbf{\Phi}$ replaced by the k corresponding eigenfunctions. Thus $\hat{\mathbf{G}}$ in Eqn 9.7 can be approximated as:

$$\hat{\mathbf{G}} \approx \mathbf{\Phi}[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_k]\mathbf{D}_a[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_k]'\mathbf{\Phi}' = \mathbf{T}\mathbf{D}_a\mathbf{T}'$$

where the $\mathbf{v}_i$ are the eigenvectors of $\hat{\mathbf{C}}$ corresponding to eigenvalues in $\mathbf{D}_a$.

Similarly, if CF has been fitted to the environmental covariance matrix, a similar reduction can be carried as follows:

$$\begin{aligned} \mathbf{R} &= \mathbf{\Phi}\mathbf{C}_p\mathbf{\Phi} + \mathbf{I}\sigma^2\varepsilon \\ &= \mathbf{\Phi}[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_k]\mathbf{D}_p[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_k]'\mathbf{\Phi}' + \mathbf{I}\sigma^2\varepsilon = \mathbf{Q}\mathbf{D}_p\mathbf{Q}' + \mathbf{I}\sigma^2\varepsilon \end{aligned} \tag{9.11}$$

Downloaded From Vet-eBooks.com

where $\mathbf{D}_p$ contains the k largest eigenvalues of $\mathbf{C}_p$. However, Mantysaari (1999) indicated that with several biological traits, Eqn 9.11 could easily lead to a non-positive definite $\mathbf{C}_p$ and the decomposition may not be possible. He used an expectation maximization (EM) algorithm to fit the CF to the environmental covariance matrix. However, if $\mathbf{C}_p$ has been estimated directly using REML (Meyer and Hill, 1997), the EM algorithm would not be necessary and the covariance matrix for pe can be approximated as $\mathbf{QD}_p\mathbf{Q}'$. In addition to reducing the number of equations to k per animal in the MME with this method, the system of equations is very sparse since $\mathbf{D}_a$ or $\mathbf{D}_p$ are diagonal.

9.5 Equivalence of the Random Regression Model to the Covariance Function

Meyer and Hill (1997) indicated that the RR model is equivalent to a covariance function model. The equivalence of the RR model fitting either a parametric curve or Legendre polynomials to the CF model is presented below. Similar to the model in Section 9.3, the RR model with a parametric curve can be represented as:

$$y_{jt} = F_{jt} + \sum_{m=0}^{f-1} z_m(t)\beta_m + \sum_{m=0}^{k-1} z_m(t)\alpha_{jm} + \sum_{m=0}^{k-1} z_m(t)\lambda_{jm} + e_{jt} \tag{9.12}$$

where y_{jt} is the test day record of cow j made on day t; β_m are fixed regressions coefficients; α_{jm} and λ_{jm} are the additive genetic and permanent environmental random regressions for cow j; F_{jt} represents the remaining fixed effects in the model; $z_m(t)$ is the mth parameter of a parametric function of days in milk; and e_{jt} is the random error term. For example, in the model of Jamrozik *et al.* (1997), z was a function of days in milk with five parameters: $z = (1\ c\ c^2\ d\ d^2)$, where $c = t/305$ and $d = \ln(1/c)$, with ln being the natural logarithm. Then the covariance between breeding values u_i and u_l on an animal recorded at DIM t_i and t_l is:

$$\text{cov}(u_i, u_l) = f(t_i, t_l) = \sum_{m=0}^{k-1}\sum_{r=0}^{k-1} z_m(t_i) z_r(t_l)\ \text{cov}(\alpha_m, \alpha_r) \tag{9.13}$$

However, instead of a parametric curve, assume that orthogonal polynomials such as Legendre polynomials were fitted in an RR model as described in Section 9.3. Let a_i and a_l represent TD records on days t_i and t_l of animal j standardized to the interval –1 to 1 as outlined in Appendix G. Furthermore, assume that the mth Legendre polynomial of a_i be $\phi_m(a_l)$, for $m = 0, \ldots, k-1$. The covariance between breeding values u_i and u_l on an animal recorded at DIM a_i and a_l could then be represented as:

$$\text{cov}(u_i, u_l) = f(a_i, a_l) = \sum_{m=0}^{k-1}\sum_{r=0}^{k-1} \phi_m(a_i)\phi_r(a_l)\ \text{cov}(\alpha_m, \alpha_r) \tag{9.14}$$

The right-hand sides of Eqns 9.13 and 9.14 are clearly equivalent to the right-hand side of Eqn 9.6, with $\text{cov}(\alpha_m, \alpha_r)$ equal to $\mathbf{C}_{ij}$, the ijth element of the coefficient matrix of the covariance function. This equivalence of the RR model with the covariance function is useful when analysing data observed at many ages or time periods, as only k regression coefficients and their $k(k+1)/2$ covariances need to be estimated for each source of variation in a univariate model.

Downloaded From Vet-eBooks.com

10 Use of Genetic Markers in Breeding Value Prediction

10.1 Introduction

A genetic marker is a fragment of DNA that is associated with a certain location (chromosome) within the genome. In the 1990s, most genetic markers used in livestock studies were microsatellites. DNA microsatellites, also referred to as simple sequence repeats (SSR), consist of a specific sequence of DNA bases or nucleotides which contain mono, di, tri, or tetra tandem repeats; for example, AAAAAAAAAAA or CTGCTGCTGCTG, which may be referred to as $(A)_{11}$ or $(CTG)_4$, respectively. Alleles at a specific location (locus) can differ in the number of repeats (polymorphic) and hence they are used as genetic markers. Microsatellites are inherited in a Mendelian fashion and are typically co-dominant, that is, the heterozygote genotype could be distinguished from either homozygote.

Genetic markers are useful in identifying portions of the chromosomes that are associated with particular quantitative traits. The incorporation of information on marker loci that are linked to quantitative trait loci (QTL), together with phenotypic information in a genetic evaluation procedure, would increase the accuracy of evaluations and therefore of selection. The use of breeding values with marker information incorporated in the selection of animals in a breeding programme is termed marker-assisted selection (MAS). The gains from MAS depend on the amount of genetic variation explained by the marker information and are larger for traits with low heritabilities, and therefore EBV from phenotype are of low accuracy (Goddard and Hayes, 2002). Similarly, MAS should result in larger increases in accuracies for traits that are sex-limited, such as milk yield, or measured in only in culled animals, for instance, carcass traits.

Fernando and Grossman (1989) presented a methodology that incorporated marker information into the BLUP procedure for the genetic evaluation of animals. This method is discussed and illustrated in this chapter. The extension of the method of Fernando and Grossman (1989) by Goddard (1992) to handle information on QTL bracketed by two markers is examined. This chapter deals with linkage analysis, i.e. the use of microsatellites as markers for the purpose of MAS.

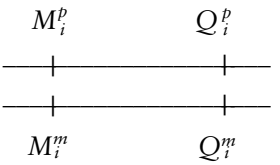
10.2 Defining a Model with Marker Information

Consider a single polymorphic marker locus (ML), which is closely linked to a quantitative trait locus (MQTL). Assume individual i inherited M_i^p and M_i^m at the ML from

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

its paternal (p) and its maternal (m) parents. Also, let Q_i^p and Q_i^m denote alleles at the quantitative trait loci linked to M_i^p and M_i^m as illustrated below:



Let v_i^p and v_i^m be the genetic additive effects of Q_i^p and Q_i^m, respectively, and u_i the genetic additive effects of the remaining quantitative trait loci not linked to the ML. Then the additive genetic value (a_i) of individual i is:

$$a_i = v_i^p + v_i^m + u_i \tag{10.1}$$

Given only phenotypic information, the usual BLUP equation for additive genetic effects (Section 3.2) is:

$$y_i = x_i\beta + a_i + e_i \tag{10.2}$$

Replacing a_i above by Eqn 10.1 gives:

$$y_i = x_i\beta + v_i^p + v_i^m + u_i + e_i \tag{10.3}$$

From Section 2.2, the covariance matrix for u_i, **A**, is the usual relationship matrix (Henderson, 1976) but the covariance for v_i, $\mathbf{G}_v$, depends on both the relationship matrix and marker information. Thus given **A** and $\mathbf{G}_v$, the BLUP of v_i and u_i can be obtained using the usual MME. The calculation of **A** and its inverse has been covered in Chapter 2. The calculation of $\mathbf{G}_v$ and its inverse are covered in the next section.

10.3 Calculating the Covariance Matrix ($\mathbf{G}_v$) for MQTL Effects

The matrix $\mathbf{G}_v\sigma_v^2$ represents the covariance between the additive effects of the MQTL alleles. For simplicity, consider only maternal MQTL. Assume two arbitrary individuals b and b' inherit MQTL alleles Q_b^m and $Q_{b'}^m$ with additive effects v_b^m and $v_{b'}^m$ from dams d and d', respectively. The covariance between the additive effects v_b^m and $v_{b'}^m$ for the maternal MQTL in b and b' is:

$$\begin{aligned}\mathrm{cov}(v_b^m, v_{b'}^m) &= \mathrm{cov}(v_b^m, v_{b'}^m \mid Q_b^m \equiv Q_{b'}^m) \cdot P(Q_b^m \equiv Q_{b'}^m) \\ &= \mathrm{var}(v_b^m) \cdot P(Q_b^m \equiv Q_{b'}^m) \\ &= \sigma_v^2 \mathbf{G}_{v(b,b')}\end{aligned} \tag{10.4}$$

where $\mathrm{var}(v_b^m) = \sigma_v^2$ is the variance of the MQTL allele, $P(Q_b^m \equiv Q_{b'}^m)$ is the probability that Q_b^m is identical by descent (IBD) to $Q_{b'}^m$ and the matrix $\mathbf{G}_{v(b,b')}$ is the covariance matrix for the MQTL between b and b'. Given that b is not a direct descendant of b', Q_b^m can only be identical by descent to $Q_{b'}^m$ in two mutually exclusive manners: (i) if Q_b^m is IBD to $Q_{d'}^p$, the paternal MQTL allele of the dam of b', and b' has inherited $Q_{d'}^p$; or (ii) Q_b^m is IBD to $Q_{d'}^m$, the maternal MQTL allele of the dam of b', and b' has inherited $Q_{d'}^m$. This is akin to calculating **A** where the relationship, say, between b and b' is evaluated through the relationship of b with the parents of b'.

Downloaded From Vet-eBooks.com

With marker information available, the conditional probability that b' inherits $Q^m_{d'}$, given that it has inherited $M^m_{d'}$, is $(1 - r)$, with r being the recombination rate between the ML and the MQTL. Thus if b' inherits $M^m_{d'}$, the probability in Eqn 10.4 can be calculated recursively as:

$$P(Q^m_b \equiv Q^m_{b'}) = P(Q^m_b \equiv Q^p_{d'}) \cdot r + P(Q^m_b \equiv Q^m_{d'}) \cdot (1 - r) \tag{10.5}$$

Similarly, given that b' inherits $M^p_{d'}$, then:

$$P(Q^m_b \equiv Q^m_{b'}) = P(Q^m_b \equiv Q^p_{d'}) \cdot (1 - r) + P(Q^m_b \equiv Q^m_{d'}) \cdot r \tag{10.6}$$

If it is not known whether b' inherits $M^m_{d'}$ or $M^p_{d'}$ due to lack of marker information, then $Q^m_{d'}$ and $Q^p_{d'}$ have equal probability of being transmitted to b'. Therefore, r is replaced by 0.5 in Eqns 10.5 and 10.6.

Using the above information, Fernando and Grossman (1989) developed a tabular method for constructing $\mathbf{G}_v$, which is similar to that for calculating $\mathbf{A}$. The rows and columns of $\mathbf{G}_v$ should be such that those for parents precede those for progeny. It should be noted that there are two rows for an individual in $\mathbf{G}_v$: one each for the paternal and maternal MQTL alleles. Let g_{ij} be the ij element of $\mathbf{G}_v$ and i^p_o, i^m_o be the rows of $\mathbf{G}_v$ corresponding to the additive effects of MQTL alleles (v^p_o, v^m_o) of the oth individual. Similarly, let i^p_s, i^m_s be the rows for the additive effects of the MQTL alleles (v^p_s, v^m_s) of its sire (s) additive effects and i^p_d, i^m_d be the rows for the effects of the MQTL alleles (v^p_d, v^m_d) of its dam (d).Then the elements of the row i^p_o below the diagonal, using Eqns 10.4 to 10.6, can be calculated as:

$$g_{i^p_o,j} = (1-\rho^p_o)g_{i^p_s,j} + \rho^p_o g_{i^m_s,j}; \quad \text{for } j = 1,...,i^p_o - 1 \tag{10.7}$$

with $\rho^p_o = r$ if b inherits M^p_s or $\rho^p_o = (1 - r)$ if o inherits M^m_s. Similarly, elements of row i^m_o below the diagonal are:

$$g_{i^m_o,j} = (1-\rho^m_o)g_{i^p_d,j} + \rho^m_o g_{i^m_d,j}; \quad \text{for } j = 1,...,i^m_o - 1 \tag{10.8}$$

where $\rho^m_o = r$ if o inherits M^p_d or $\rho^m_o = (1 - r)$ if o inherits M^m_d. Since $\mathbf{G}_v$ is symmetric then:

$$g_{j,i^p_o} = g_{i^p_o,j} \quad \text{and} \quad g_{j,i^m_o} = g_{i^m_o,j}$$

It is obvious from Eqn 10.4 that, if $o = o'$, that is, the same individual, then cov(v^m_o, $v^m_{o'}$) = var(v^m_o) as $P(Q^m_o \equiv Q^m_{o'}) = 1$. Therefore, the diagonal elements of $\mathbf{G}_v$ equal unity. If it is not possible to determine which of the two marker alleles o inherited from its sire or dam, then ρ^p_o in Eqn 10.7 and ρ^m_o in Eqn 10.8 are replaced by 0.5.

10.3.1 Numerical application

Example 10.1

Given in the table below are the post-weaning gain data of five calves with the genotype at the marker locus given. The aim at this stage is to construct the covariance matrix $\mathbf{G}_v$ for the MQTL among the five calves.

Downloaded From Vet-eBooks.com

Calf	Sex of calf	Marker inheritance				PWG (kg)
		Sire	Dam	Sire	Dam	
1	Male	–	–	–	–	6.8
2	Female	–	–	–	–	4.5
3	Male	1	2	M_1^p	M_2^m	8.5
4	Female	1	3	M_1^m	M_3^p	6.0
5	Female	4	3	M_4^p	M_3^p	7.0

For ease of illustration, let rows i_o^p and i_o^m for animal o in $\mathbf{G}_v$ be coded as ip and im, respectively. Thus, for example, for animal 1, i_1^p and i_1^m will be coded as $1p$ and $1m$, respectively, for animal 2, i_2^p and i_2^m will be coded as $2p$ and $2m$ and for animal 5, i_5^p and i_5^m will be coded as $5p$ and $5m$, respectively. The $\mathbf{G}_v$ for the example, therefore, is:

	1p	1m	2p	2m	3p	3m	4p	4m	5p	5m
1p	1.000	0.000	0.000	0.000	0.900	0.000	0.100	0.810	0.171	0.810
1m	0.000	1.000	0.000	0.000	0.100	0.000	0.900	0.090	0.819	0.090
2p	0.000	0.000	1.000	0.000	0.000	0.100	0.000	0.010	0.001	0.010
2m	0.000	0.000	0.000	1.000	0.000	0.900	0.000	0.090	0.009	0.090
3p	0.900	0.100	0.000	0.000	1.000	0.000	0.180	0.900	0.252	0.900
3m	0.000	0.000	0.100	0.900	0.000	1.000	0.000	0.100	0.010	0.100
4p	0.100	0.900	0.000	0.000	0.180	0.000	1.000	0.162	0.916	0.162
4m	0.810	0.090	0.010	0.090	0.900	0.100	0.162	1.000	0.246	0.820
5p	0.171	0.819	0.001	0.009	0.252	0.010	0.916	0.246	1.000	0.228
5m	0.810	0.090	0.010	0.090	0.900	0.100	0.162	0.820	0.228	1.000

The calculation of $\mathbf{G}_v$ for the first three animals is illustrated as below. For the first two animals, the parents are unknown, therefore:

$$g_{1p,1p} = g_{1m,1m} = g_{2p,2p} = g_{2m,2m} = 0$$

At the ML, animal 3 inherited M_s^p from his father; therefore, for row $3p$ in $\mathbf{G}_v$, corresponding to the effects of the paternal alleles of the MTQL for animal 3, $r = 0.1$. Hence, from Eqn 10.7:

$$\begin{aligned}
g_{3p,1p} &= (1 - 0.1)g_{1p,1p} + (0.1)g_{1m,1p} = (0.9)1 + (0.1)0 = 0.9\\
g_{3p,1m} &= (1 - 0.1)g_{1p,1m} + (0.1)g_{1m,1m} = (0.9)0 + (0.1)1 = 0.1\\
g_{3p,2p} &= (1 - 0.1)g_{1p,2p} + (0.1)g_{1m,2p} = (0.9)0 + (0.1)0 = 0\\
g_{3p,2m} &= (1 - 0.1)g_{1p,2m} + (0.1)g_{1m,2m} = (0.9)0 + (0.1)0 = 0\\
g_{3p,3p} &= 1.0
\end{aligned}$$

At the ML, animal 3 inherited M_d^m from his mother; therefore, for row $3m$ in $\mathbf{G}_v$, corresponding to the effects of the maternal alleles of the MTQL for animal 3, $r = 0.9$. Hence, from Eqn 10.8:

$$\begin{aligned}
g_{3m,1p} &= (1 - 0.9)g_{2p,1p} + (0.9)g_{2m,1p} = (0.1)0 + (0.9)0 = 0\\
g_{3m,1m} &= (1 - 0.9)g_{2p,1m} + (0.9)g_{2m,1m} = (0.1)0 + (0.9)0 = 0\\
g_{3m,2p} &= (1 - 0.9)g_{2p,2p} + (0.9)g_{2m,2p} = (0.1)1 + (0.9)0 = 0.1
\end{aligned}$$

Downloaded From Vet-eBooks.com

$g_{3m,2m} = (1 - 0.9)g_{2p,2m} + (0.9)g_{2m,2m} = (0.1)0 + (0.9)1 = 0.9$
$g_{3m,3p} = (1 - 0.9)g_{2p,3p} + (0.9)g_{2m,3p} = (0.1)0 + (0.9)0 = 0$
$g_{3m,3m} = 1.0$

10.4 An Alternative Approach for Calculating $\mathbf{G}_v$

An alternative recursive method for the calculation of $\mathbf{G}_v$ and its inverse was presented by Van Arendonk *et al*. (1994) using matrix notation. Their method accounts for inbreeding and can be used to calculate a combined numerator relationship matrix ($\mathbf{A}_a$) and its inverse. The matrix $\mathbf{A}_a = \mathbf{A}_u + \mathbf{A}_v$, where $\mathbf{A}_u$ is the numerator relationship matrix for animals for QTL not linked to the marker and $\mathbf{A}_v$ is the relationship matrix for animals for MQTL linked to the marker. The inverse of $\mathbf{A}_a$ is useful for the direct prediction of total additive genetic merit, i.e. additive genetic merit with information from markers directly included.

The principles of their methodology are initially illustrated briefly using the calculation of the relationship matrix ($\mathbf{A}$) among animals in the absence of marker information. The representation of the rules for building $\mathbf{A}_i$ for animals 1 to i in matrix form is:

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{i-1} & \mathbf{A}_{i-1}\mathbf{s}_i \\ \mathbf{s}_i' \mathbf{A}_{i-1} & a_{ii} \end{bmatrix} \tag{10.9}$$

where $\mathbf{s}_i$ is the column vector of $i - 1$ elements containing two elements, $\frac{1}{2}$, corresponding to the sire or dam (if known) and zero elsewhere. $\mathbf{A}_{i-1}$ is the numerator relationship matrix for animals 1 to $(i - 1)$ and a_{ii} is the diagonal element of $\mathbf{A}$ for animal i and is equal to $1 + F_i$, where F_i is the inbreeding coefficient of the ith animal. Using the data in Example 10.1, the $\mathbf{A}$ matrix ignoring marker information is:

$$\mathbf{A} = \begin{bmatrix} 1.000 & 0.000 & 0.500 & 0.750 & 0.625 \\ 0.000 & 1.000 & 0.500 & 0.250 & 0.375 \\ 0.500 & 0.500 & 1.000 & 0.750 & 0.875 \\ 0.750 & 0.250 & 0.750 & 1.250 & 1.000 \\ 0.625 & 0.375 & 0.875 & 1.000 & 1.375 \end{bmatrix}$$

For animal 5, $\mathbf{s}_5'$ = [0 0 0.5 0.5]; therefore, the column vector above the diagonal for animal 5 ($\mathbf{q}_5$) in $\mathbf{A}$ using Eqn 10.9 can be calculated as $\mathbf{q}_5 = \mathbf{A}_4\mathbf{s}_5$.Thus the row vector $\mathbf{q}_5' = \mathbf{s}_5'\mathbf{A}_4$= [0.625 0.375 0.875 1.00] and the diagonal element for animal 5, $a_{55} = 1 + 0.5(a_{34}) = 1.375$. Note also that given $\mathbf{q}_i$, $\mathbf{s}_i$ can be computed as:

$$\mathbf{s}_i = \mathbf{A}_{i-1}^{-1}\, \mathbf{q}_i \tag{10.10}$$

This relationship will be used in subsequent sections when it is not possible to calculate s_i directly.

Given $\mathbf{A}_{i-1}^{-1}$, for animal $i - 1$, Tier and Solkner (1993) demonstrated that the effect of adding an additional row to $\mathbf{A}$ on the elements of $\mathbf{A}^{-1}$ as:

$$\mathbf{A}_i^{-1} = \begin{bmatrix} \mathbf{A}_{i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \left(a_{ii} - \mathbf{s}_i' \mathbf{A}_{i-1}\mathbf{s}_i\right)^{-1} \begin{bmatrix} \mathbf{s}_i\mathbf{s}_i' & -\mathbf{s}_i \\ -\mathbf{s}_i' & 1 \end{bmatrix} \tag{10.11}$$

Downloaded From Vet-eBooks.com

When both sire (f) and dam (d) of i are known, $\mathbf{s}_i'\mathbf{A}_{i-1}\mathbf{s} = \frac{1}{4}(a_{ff} + a_{fd} + a_{df} + a_{dd})$ where a_{ij} are the elements of $\mathbf{A}_{i-1}$ for f and d. Since $a_{ii} = (1 + a_{ff})$, then $(a_{ii} - \mathbf{s}_i'\mathbf{A}_{i-1}\mathbf{s})^{-1}$ can be written as $(1 - \frac{1}{4}(a_{ff} + a_{dd}))^{-1}$. The application of Eqn 10.11 to calculate $\mathbf{A}^{-1}$ for the pedigree in Example 10.1 is straightforward. For instance, for the first two animals with parents unknown, $\mathbf{A}_2^{-1}$ is an identity matrix of order 2. Then $\mathbf{A}_3^{-1}$ can then be calculated using Eqn 10.11. Given that $\mathbf{A}_4^{-1}$ has been calculated, the inverse of $\mathbf{A}$ for all five animals can be illustrated as follows:

For animal 5, $(a_{55} - \mathbf{s}_5'\mathbf{A}_4\mathbf{s}_5)^{-1} = (1 - \frac{1}{4}(a_{33} + a_{44}))^{-1} = (1 - \frac{1}{4}(1 + 1.25))^{-1} = 2.286$. Then Eqn 10.11 is:

$$\mathbf{A}_5^{-1} = \begin{bmatrix} 2.000 & 0.500 & -0.500 & -1.000 & 0.000 \\ 0.500 & 1.500 & -1.000 & 0.000 & 0.000 \\ -0.500 & -1.000 & 2.500 & -1.000 & 0.000 \\ -1.000 & 0.000 & -1.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix} + (2.286) \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.25 & 0.25 & -0.5 \\ 0.0 & 0.0 & 0.25 & 0.25 & -0.5 \\ 0.0 & 0.0 & -0.5 & -0.5 & 1.0 \end{pmatrix}$$

$$= \begin{bmatrix} 2.000 & 0.500 & -0.500 & -1.000 & 0.000 \\ 0.500 & 1.500 & -1.000 & 0.000 & 0.000 \\ -0.500 & -1.000 & 3.071 & -0.429 & -1.143 \\ -1.000 & 0.000 & -0.429 & 2.571 & -1.143 \\ 0.000 & 0.000 & -1.143 & -1.143 & 2.28 \end{bmatrix}$$

where $\mathbf{s}_5' = (0 \quad 0 \quad 0.5 \quad 0.5)$.

Applying Eqns 10.9, Van Arendonk *et al.* (1994) showed that when alleles are ordered chronologically, $\mathbf{G}_{v,i}$ can be calculated as:

$$\mathbf{G}_{v,i} = \begin{bmatrix} \mathbf{G}_{v,i-1} & \mathbf{G}_{v,i-1}\mathbf{s}_i \\ \mathbf{s}_i'\mathbf{G}_{v,i-1} & g_{ii} \end{bmatrix} \tag{10.12}$$

where $\mathbf{s}_i$ is the column vector of $i - 1$ elements containing non-zero elements relating allele i to paternal and maternal alleles of parent (if known) and zeros elsewhere; $\mathbf{G}_{v,i-1}$ is the covariance matrix for MQTL for alleles 1 to $(i - 1)$; and g_{ii} is the diagonal element of $\mathbf{G}_v$ for the k allele, which is equal to 1. Using the same notation for the rows in $\mathbf{G}_v$ shown in Section 10.3.1, $\mathbf{s}_i$ for animals 3, 4 and 5 are: $\mathbf{s}_{3p}' = [1 - r \; r \; 0 \; 0]$, $\mathbf{s}_{3m}' = [0 \; 0 \; r \; (1 - r) \; 0]$, $\mathbf{s}_{4p}' = [r \; 1 - r \; 0 \; 0 \; 0 \; 0]$, $\mathbf{s}_{4m}' = [0 \; 0 \; 0 \; 0 \; (1 - r) \; r \; 0]$, $\mathbf{s}_{5p}' = [0 \; 0 \; 0 \; 0 \; 0 \; 0 \; (1 - r) \; r]$ and $\mathbf{s}_{5m}' = [0 \; 0 \; 0 \; 0 \; (1 - r) \; r \; 0 \; 0 \; 0]$. Thus $\mathbf{G}_v$ can easily be constructed using Eqn 10.12.

10.5 Calculating the Inverse of $\mathbf{G}_v$

Fernando and Grossman (1989) used an approach similar to that for setting up $\mathbf{A}^{-1}$ in calculating the inverse of $\mathbf{G}_v$. They showed that $\mathbf{G}_v$ could be expressed as:

$$\mathbf{G}_v = (\mathbf{Q}^{-1})'\mathbf{H}\mathbf{Q}^{-1}$$

Therefore, $\mathbf{G}_v^{-1}$ can be written as:

$$\mathbf{G}_v^{-1} = \mathbf{Q}\mathbf{H}^{-1}\mathbf{Q}' \tag{10.13}$$

Downloaded From Vet-eBooks.com

where $\mathbf{Q} = (\mathbf{I} - \mathbf{P}')$ and $\mathbf{P}$ is a matrix that relates the effect of the MQTL allele of an individual to the paternal and maternal MQTL alleles of its parent. Each row of $\mathbf{P}$ contains only two non-zero elements if the parent is known, otherwise only zeros if the parent is unknown. For instance, for individual i with sire (s) known, row i_o^p will have $(1 - \rho_o^p)$ in the column corresponding to i_s^p, and ρ_o^p in the column corresponding to column i_s^m. Similarly, if dam (d) is known, row i_o^m will contain $(1 - \rho_o^m)$ in the column corresponding to i_d^p and ρ_o^m in the column corresponding to i_d^m. The row of $\mathbf{P}$ for allele i is equal to s_i in Eqn 10.12. The matrix $\mathbf{P}$ for the pedigree in Example 10.1 is:

	1p	1m	2p	2m	3p	3m	4p	4m	5p	5m
1p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1m	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2m	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3p	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3m	0.0	0.0	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0
4p	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4m	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0
5p	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0
5m	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0

The matrix $\mathbf{H}$ is a diagonal matrix for the covariance of residual effects after adjusting the effect of the MQTL allele of an individual for the effects of the parent's paternal and maternal MQTL alleles. For example, the residual effect (ε_o^p) for a paternal MQTL allele of an individual with sire s known is:

$$\varepsilon_o^p = v_o^p - (1 - \rho_o^p)v_s^p + \rho_o^p v_s^m$$

and the variance of ε_o^p is:

$$\text{var}(\varepsilon_o^p) = \text{var}(v_o^p) - (1 - \rho_o^p)^2 \cdot \text{var}(v_s^p) - (\rho_o^p)^2 \cdot \text{var}(v_s^m) - 2(1 - \rho_o^p)\rho_o^p \cdot \text{cov}(v_s^p, (v_s^m))$$

Since $\text{var}(v_o^p) = \text{var}(v_s^p) = \text{var}(v_s^m) = \sigma_v^2$ and $\text{cov}(v_s^p, v_s^m) = \text{var}(v_s^p) \cdot P(Q_s^p \equiv Q_s^m) = \text{var}(v_s^p) \cdot F_s = \sigma_v^2 F_s$, the above equation can be written as:

$$\begin{aligned}\text{var}(\varepsilon_o^p) &= 2\sigma_v^2(\rho_o^p) - 2\sigma_v^2(\rho_o^p)^2 - 2\sigma_v^2(1 - \rho_o^p)\rho_o^p F_s \\ &= 2\sigma_v^2((1 - \rho_o^p)\rho_o^p - (1 - \rho_o^p)\rho_o^p F_s) \\ &= 2\sigma_v^2(1 - \rho_o^p)\rho_o^p(1 - F_s) \\ \text{var}(\varepsilon_o^p)/\sigma_v^2 &= h_o^p = 2(1 - \rho_o^p)\rho_o^p(1 - F_s) \qquad (10.14)\end{aligned}$$

where $(1 - \rho_o^p)\rho_o^p = (1 - r)r$ for $\rho_o^p = r$ or $(1 - r)$, F_s is the inbreeding coefficient at the MQTL of the sire and h_o^p is the diagonal element of $\mathbf{H}$ for the paternal MQTL of individual o. Therefore, if the sire is not inbred, $h_o^p = 2(1 - r)r$ with marker information or $h_o^p = 0.5$ with no marker information and $h_o^p = 1$ if the sire is unknown. Similarly, for the maternal MQTL of o:

$$\text{var}(\varepsilon_o^m)/\sigma_v^2 = h_o^m = 2(1 - \rho_o^m)\rho_o^m(1 - F_d) \qquad (10.15)$$

Downloaded From Vet-eBooks.com

where $(1 - \rho_o^m)\rho_o^m = (1 - r)r$ for $\rho_o^m = r$ or $(1 - r)$, F_d is the inbreeding coefficient at the MQTL of the dam and h_o^m is the diagonal element of **D** for the paternal MQTL of individual *o*. Therefore, if the dam is not inbred, $h_o^m = 2(1 - r)r$ with marker information or $h_o^m = 0.5$ with no marker information and $h_o^m = \sigma_v^2$ if the dam is unknown.

Equation 10.13 may be written as:

$$\mathbf{G}_v^{-1} = \sum_{j=1}^{n} q_j q_j h_j^{-1}$$

where *n* is the number of individuals in the pedigree, q_j is the column of **Q** and h_j is the *j*th diagonal element of **H**. Since $\mathbf{Q} = (1 - \mathbf{P}')$, the *j*th element of q_j (i.e. the diagonal element) is unity and q_j has at most only two other non-zero elements. If the sire of *o* is known, $j = i_o^p$, element $i_s^p = -(1 - \rho_o^p)$ and element $i_s^m = -\rho_o^p$. Similarly, if the dam is known, then for $j = i_o^m$, element $i_d^p = -(1 - \rho_o^m)$ and element $i_d^m = -\rho_o^m$. Therefore, the contribution corresponding to the paternal and maternal MQTL alleles of an individual to $\mathbf{G}_v^{-1}$ can easily be calculated from parent and marker information.

Fernando and Grossman (1989) gave the following rules for obtaining $\mathbf{G}_v^{-1}$. First, calculate the diagonals of **H** using Eqns 10.14 and 10.15 and its inverse. Second, set $\mathbf{G}_v^{-1}$ to zero and for each offspring *o*, with sire *s* and dam *d*, add the following to the indicated elements of $\mathbf{G}_v^{-1}$:

If the sire is known, add:

$(1 - \rho_o^p)^2 h_{i_o^p}^{-1}$ to diagonal element $i_s^p i_s^p$;

$-(1 - \rho_o^p)\, h_{i_o^p}^{-1}$ to elements $i_s^p i_o^p$ and $i_o^p i_s^p$;

$(1 - \rho_o^p)\rho_o^p\, h_{i_o^p}^{-1}$ to elements $i_s^p i_s^m$ and $i_s^m i_s^p$;

$(\rho_o^p)^2 h_{i_o^p}^{-1}$ to diagonal element $i_s^m i_s^m$;

$-\rho_o^p\, h_{i_o^p}^{-1}$ to elements $i_s^m i_o^p$ and $i_o^p i_s^m$

If the dam is known, add:

$(1 - \rho_o^m)^2 h_{i_o^m}^{-1}$ to diagonal element $i_d^p i_d^p$;

$-(1 - \rho_o^m)\, h_{i_o^m}^{-1}$ to elements $i_d^p i_o^m$ and $i_o^m i_d^p$;

$(1 - \rho_o^m)\rho_o^m\, h_{i_o^m}^{-1}$ to elements $i_d^p i_d^m$ and $i_d^m i_d^p$;

$(\rho_o^m)^2 h_{i_o^m}^{-1}$ to diagonal element $i_d^m i_d^m$;

$-\rho_o^m\, h_{i_o^m}^{-1}$ to elements $i_d^m i_o^m$ and $i_o^m i_d^m$

And always add:

$h_{i_o^p}^{-1}$ to element i_o^p, i_o^p and $h_{i_o^m}^{-1}$ to element i_o^m, i_o^m

Downloaded From Vet-eBooks.com

Applying these rules, the calculation of the inverse of $\mathbf{G}_v^{-1}$ for the pedigree in Example 10.1 is illustrated. For this pedigree, the matrix **H** and its inverse are:

$\mathbf{H}$ = diag(1 1 1 1 0.18 0.18 0.18 0.18 0.1508 0.18) and
$\mathbf{H}^{-1}$ = diag(1 1 1 1 5.556 5.556 5.556 5.556 6.630 5.556)

Note that in calculating the diagonal element for the paternal MQTL of animal 4 ($d_{4p,4p}$), an inbreeding coefficient of 0.162 (covariance between the maternal and paternal MQTL alleles of the sire and dam, respectively) has been accounted for. Set $\mathbf{G}_v^{-1}$ with elements represented as $g^{ii,jj}$ to zero and the contribution from the first three animals can be calculated as follows.

For animals 1 and 2, parents are unknown; the diagonal elements are equal to 1 for the MQTL alleles of these animals. Therefore, add 1 to $g^{1p,1p}$, $g^{1m,1m}$, $g^{2p,2p}$ and $g^{2m,2m}$, using the same coding as for the rows of $\mathbf{G}_v$ as in Section 10.3. For paternal MQTL allele of animal 3, $\rho_o^p = 0.1$ and $d_{3p,3p}$ equals 5.556. Add $(1 - 0.1)^2 h^{-1}_{3p,3p} = 4.50$ to g^{1p1p}, $(1 - 0.1)0.1(h^{-1}_{3p,3p}) = 0.5$ to $g^{1p,1m}$, $-(1 - 0.1)h^{-1}_{3p3p} = -5.00$ to $g^{1p,3p}$, $(0.1)^2 h^{-1}_{3p3p} = 0.056$ to $g^{1m,1m}$, $(-0.1)h^{-1}_{3p3p} = 0.556$ to $g^{1m,3p}$ and h^{-1}_{3p3p} to $g^{3p,3p}$. For the maternal allele of animal 3, $\rho_o^p = 0.9$ and $h^{-1}_{3m,3m} = 5.556$. Add $(1 - 0.9)^2 h^{-1}_{3m,3m} = 0.056$ to $g^{2p,2p}$, $(1 - 0.9)0.9(h^{-1}_{3m,3m}) = 0.5$ to $g^{2p,2m}$, $-(1 - 0.9)h^{-1}_{3m3m} = -0.556$ to $g^{2p,3m}$, $(0.9)^2 h^{-1}_{3m3m} = 4.50$ to $g^{2m,2m}$, $(-0.9)h^{-1}_{3m3m} = -0.500$ to $g^{2m,3m}$ and h^{-1}_{3m3m} to $g_{3m,3m}$. Applying the rules to all animals in the pedigree gives $\mathbf{G}_v^{-1}$as:

	1p	1m	2p	2m	3p	3m	4p	4m	5p	5m
1p	5.556	1.000	0.000	0.000	–5.000	0.000	–0.556	0.000	0.000	0.000
1m	1.000	5.556	0.000	0.000	–0.556	0.000	–5.000	0.000	0.000	0.000
2p	0.000	0.000	1.056	0.500	0.000	–0.556	0.000	0.000	0.000	0.000
2m	0.000	0.000	0.500	5.500	0.000	–5.000	0.000	0.000	0.000	0.000
3p	–5.000	–0.556	0.000	0.000	14.556	1.000	0.000	–5.000	0.000	–5.000
3m	0.000	0.000	–0.556	–5.000	1.000	5.667	0.000	–0.556	0.000	–0.556
4p	–0.556	–5.000	0.000	0.000	0.000	0.000	10.925	0.597	–5.967	0.000
4m	0.000	0.000	0.000	0.000	–5.000	–0.556	0.597	5.622	–0.663	0.000
5p	0.000	0.000	0.000	0.000	0.000	0.000	–5.967	–0.663	6.630	0.000
5m	0.000	0.000	0.000	0.000	–5.000	–0.556	0.000	0.000	0.000	5.556

Similarly, the inverse of $\mathbf{G}_{v,i}^{-1}$ can be obtained using Eqn 10.11 (Van Arendonk *et al*., 1994) as:

$$\mathbf{G}_{v,i}^{-1} = \begin{bmatrix} \mathbf{G}_{v,i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \left(g_{ii} - \mathbf{s}_i' \mathbf{G}_{v,i-1} \mathbf{s}_i\right)^{-1} \begin{bmatrix} \mathbf{s}_i \mathbf{s}_i' & -\mathbf{s}_i \\ -\mathbf{s}_i' & \mathbf{1} \end{bmatrix} \tag{10.16}$$

The application of Eqn 10.16 for the calculation of $\mathbf{G}_v^{-1}$ is briefly illustrated. It has been shown earlier that $\mathbf{G}_v^{-1}$ for the MQTL alleles of the first two animals is an

Downloaded From Vet-eBooks.com

identity matrix of order 4. The matrix $\mathbf{G}_v^{-1}$ with the paternal MQTL allele of animal 3 added can be computed as:

$$\mathbf{G}_{v,3p}^{-1} = \begin{bmatrix} \mathbf{G}_{v,2m}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + (1 - 0.82)^{-1} \begin{bmatrix} 0.81 & 0.09 & 0 & 0 & -0.9 \\ 0.09 & 0.01 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.9 & -0.1 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 5.500 & 0.500 & 0.0 & 0.0 & -5.000 \\ 0.500 & 1.056 & 0.0 & 0.0 & -0.556 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.500 & -0.556 & 0.0 & 0.0 & 5.556 \end{bmatrix}$$

10.6 Prediction of Breeding Values with Marker Information

The model in Eqn 10.2 for breeding value prediction with marker information can be written in matrix notation as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Wv} + \mathbf{e} \tag{10.17}$$

where $\mathbf{y}$ is the vector of observation, $\boldsymbol{\beta}$ is the vector of fixed effects, $\mathbf{u}$ is the random vector for additive genetic effects due to loci not linked to ML, $\mathbf{v}$ is the random vector with allelic effects at the MQTL and $\mathbf{e}$ is random residual effects. The matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ are incidence matrices. $\text{Var}(u) = \mathbf{A}_u\sigma_u^2$, $\text{var}(v) = \mathbf{G}_v\sigma_v^2$, $\text{var}(e) = \mathbf{I}\sigma_e^2$ and $\text{cov}(u, v) = \text{cov}(u, e) = \text{cov}(v, e) = 0$.

The MME for the above linear model are:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}_a^{-1}\alpha_1 & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{G}_v^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \tag{10.18}$$

where:

$$\alpha_1 = \sigma_e^2/\sigma_u^2 \quad \text{and} \quad \alpha_2 = \sigma_e^2/\sigma_v^2$$

10.6.1 An illustration

Example 10.2

Using the data for Example 10.1, the breeding value of animals for QTL not linked to ML (simply referred to subsequently as breeding values), additive MQTL effects are predicted for the beef calves and sex effects are estimated. It is assumed that $\sigma_u^2 = 0.3$, $\sigma_v^2 = 0.05$ and $\sigma_e^2 = 0.6$. Therefore, $\alpha_1 = 0.6/0.3 = 2$ and $\alpha_2 = 0.6/0.05 = 12$. The parameters are expressed as a proportion of the phenotypic variance. Note that the total genetic variance $\sigma_a^2 = (\sigma_u^2 + 2\sigma_v^2) = 0.3 + 2(0.05) = 0.40$. Thus 40% of the phenotypic variance is due to additive genetic variance, of which 25% can be explained by the MQTL.

Downloaded From Vet-eBooks.com

The matrix **X** is formed as discussed in Example 3.1, **Z** is an identity matrix and the matrix **W** is:

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The matrices $\mathbf{A}_u^{-1}$ and $\mathbf{G}_v^{-1}$ have been calculated for the example data. The remaining matrices in the MME are calculated through matrix multiplication and addition. The MME are too large to be shown, but solving the equations by direct inversion gives the following results:

Effects	Solutions
Sex	
Male	7.357
Female	5.529
Animals	Breeding values
1	0.092
2	–0.091
3	0.341
4	0.329
5	0.515
MQTL alleles of animals	Additive effects
1p	0.064
1m	0.011
2p	–0.065
2m	–0.011
3p	0.083
3m	–0.004
4p	0.028
4m	0.076
5p	0.043
5m	0.086

The additive genetic effects of the MQTL accounted for about 45% of the total genetic merit of animals 1 and 2 but only about 20% for animals 3 and 5.

In Germany, with Holstein dairy cattle, the method used in Example 10.2 has been used to incorporate QTL information into routine estimation of breeding values (Szyda *et al*., 2003). In the study, 13 markers were used for routine genotyping of animals, and regions representing QTL for milk, protein, fat yields and somatic cell counts were identified on several chromosomes. The QTL information has been incorporated into BLUP, analysing DYD as the dependent variable. As a percentage of the polygenic variance, the variances of the MQTL in their study varied from 3 to 5% for milk, fat and protein yields in the first lactation.

Downloaded From Vet-eBooks.com

10.7 Directly Predicting the Additive Genetic Merit at the MQTL

Another approach to reduce the number of equations in the MME is to directly predict the combined additive genetic effects for the paternal and maternal alleles at the MQTL of an individual. The number of equations per animal would therefore be two: one for the additive genetic effects not linked to the MQTL and the other for MQTL. This implies predicting the additive genetic effects at the MQTL at the animal level; therefore, a covariance matrix ($\mathbf{A}_v$) for the MQTL at the animal level is needed. The covariance matrix $\mathbf{A}_v$ can be obtained from $\mathbf{G}_v$ as $\mathbf{A}_v = \frac{1}{2}\mathbf{B}\mathbf{G}_v\mathbf{B}'$; where $\mathbf{B} = \mathbf{I}_n \otimes [1\ 1]$, with n being the number of animals, and $\otimes$ denotes the Kronecker product. For Example 10.1, the matrix $\mathbf{B} = \mathbf{W}$ in Section 10.5 and $\mathbf{A}_v$ is:

$$\mathbf{A}_v = \begin{bmatrix} 1.000 & 0.000 & 0.500 & 0.950 & 0.945 \\ 0.000 & 1.000 & 0.500 & 0.050 & 0.055 \\ 0.500 & 0.500 & 1.000 & 0.590 & 0.631 \\ 0.950 & 0.050 & 0.590 & 1.162 & 1.072 \\ 0.945 & 0.055 & 0.631 & 1.072 & 1.228 \end{bmatrix}$$

Equation 10.11 can be used to obtain the inverse of $\mathbf{A}_v$. However, the vector $\mathbf{s}_i$ containing the contributions from ancestors is needed and this can be computed using Eqn 10.10. The vector $\mathbf{s}_i$ for the ith animal needed to calculate $\mathbf{A}_v^{-1}$ is shown in Table 10.1.

The inverse of $\mathbf{A}_v$ is:

$$\mathbf{A}_v^{-1} = \begin{bmatrix} 4.966 & 0.286 & -0.148 & -2.723 & -1.382 \\ 0.286 & 1.519 & -1.068 & 0.013 & 0.249 \\ -0.148 & -1.068 & 2.245 & -0.298 & -0.732 \\ -2.723 & 0.013 & -0.298 & 5.978 & -2.971 \\ -1.382 & 0.249 & -0.732 & -2.971 & 4.836 \end{bmatrix}$$

The model for the prediction now becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{q} + \mathbf{e} \qquad (10.19)$$

where all terms are as defined in Eqn 10.17 except that $\mathbf{W}$ is now identical to $\mathbf{Z}$ and relates additive genetic effects at the MQTL to animals. Both matrices $\mathbf{Z}$ and $\mathbf{W}$ are identity matrices and are of the order of animals. The vector $\mathbf{q}$ is the vector of additive genetic effects at the MQTL and is equal to the sum of the additive genetic effects of the paternal and maternal alleles for the animal. The variance–covariance matrix of $\mathbf{q} = 2\mathbf{A}_v\sigma_v^2 = \mathbf{A}_v\sigma_q^2$, since $\sigma_q^2 = 2\sigma_v^2$. The MME for the above model are:

Table 10.1. Vector ($\mathbf{s}_i$) with contributions at the MQTL from ancestors (animals 1 to 4) to animals 2 to 5 using the pedigree in Example 10.1.

	Elements in $\mathbf{s}_i$ relating to animal			
Animal	1	2	3	4
2	0.0000			
3	0.5000	0.5000		
4	0.8600	–0.0400	0.1800	
5	0.2857	–0.0514	0.1514	0.6143

Downloaded From Vet-eBooks.com

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}_a^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \mathbf{A}_v^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{q}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \tag{10.20}$$

with:

$$\alpha_1 = \sigma_e^2/\sigma_u^2 \quad \text{and} \quad \alpha_2 = \sigma_e^2/\sigma_q^2$$

10.7.1 An illustration

Example 10.3

Using the same data set as in Example 10.1 and the same genetic parameters, the prediction of additive genetic effects breeding values at the QTL not linked to the MQTL, and combined additive genetic effect of the MQTL at the animal level, is illustrated.

From the parameters, $\alpha_1 = 0.6/0.3 = 2$ and $\alpha_2 = 0.6/0.10 = 6$. The design matrices $\mathbf{X}$ and $\mathbf{Z}$ are as defined in Example 10.2 and $\mathbf{W}$ is now equal to $\mathbf{Z}$. The MME is too large to show but the matrix $\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{A}_v^{-1}\alpha_2$ is:

$$\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{A}_v^{-1}\alpha_2 = \begin{bmatrix} 30.796 & 1.716 & -0.888 & -16.338 & -8.292 \\ 1.716 & 10.114 & -6.408 & 0.078 & 1.494 \\ -0.888 & -6.408 & 14.470 & -1.788 & -4.392 \\ -16.338 & 0.078 & -1.788 & 36.868 & -17.826 \\ -8.292 & 1.494 & -4.392 & -17.826 & 30.016 \end{bmatrix}$$

Solving the MME gave the following solutions:

Effects	Solutions
Sex	
Male	7.356
Female	5.529
Animal	Additive genetic effects not linked to MQTL
1	0.091
2	–0.091
3	0.341
4	0.329
5	0.515
Animal	Combined additive genetic effects at the MQTL
1	0.076
2	–0.076
3	0.079
4	0.104
5	0.130

The solutions for the additive effect at the MQTL are the same as the sum of estimated effects in Examples 10.1 and 10.2. The application of this model may be

Downloaded From Vet-eBooks.com

limited to populations of small size as the tabular method of calculating $\mathbf{A}_v$ and its inverse may not be computationally feasible in large populations.

10.8 Predicting Total Additive Genetic Merit

Van Arendonk (1994) showed that total additive genetic merit (**a**) for animals that includes marker information could be predicted directly. This implies that only a single equation is needed for an animal in the MME to predict breeding values with marker information included. Let Eqn 10.17 be written as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e} \tag{10.21}$$

where $\mathbf{a} = \mathbf{u} + \mathbf{Kv}$ with **u** and **v** as defined in Eqn 10.17. The matrix **K**, which relates allelic effects to animals, is identical to **W** in Eqn 10.17 when all animals have observations. The variance–covariance matrix of **a** ($\mathbf{V}_a$) is:

$$\begin{aligned}\mathbf{V}_a &= \text{var}(\mathbf{u} + \mathbf{Kv}) \\ &= \text{var}(\mathbf{u}) + \mathbf{K}\text{var}(\mathbf{v})\mathbf{K}' \\ &= \mathbf{A}_u\sigma_u^2 + \mathbf{K}\mathbf{G}_v\mathbf{K}'\sigma_v^2 \\ &= \mathbf{A}_u\sigma_u^2 + 2\mathbf{A}_v\sigma_v^2 \\ &= \mathbf{A}_u\sigma_u^2 + \mathbf{A}_v\sigma_q^2\end{aligned}$$

The combined numerator relationship matrix among animals with marker information included ($\mathbf{A}_a$) is:

$$\mathbf{A}_a = \mathbf{A}_u\sigma_u^2/\sigma_a^2 + \mathbf{A}_v\sigma_q^2/\sigma_a^2 \tag{10.22}$$

with:

$$\sigma_a^2 = \sigma_u^2 + \sigma_q^2$$

The MME for Eqn 10.21 are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}_a^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \tag{10.23}$$

where:

$$\alpha = \sigma_e^2/\sigma_a^2$$

The use of Eqn 10.23 would require the inverse of $\mathbf{A}_a$ to be calculated. Initially, $\mathbf{A}_a$ is computed using Eqn 10.22, then $\mathbf{A}_a^{-1}$ can be calculated using Eqn 10.11, with the vector $\mathbf{s}_i$ containing the contributions from ancestors computed using Eqn 10.10. The calculation of both matrices is illustrated in the following example.

10.8.1 Numerical application

Example 10.4

Using the same data set as in Example 10.1 and the same genetic parameters, the total additive genetic effects of animals, which included marker information, are directly

Downloaded From Vet-eBooks.com

predicted. From the genetic parameters in Example 10.1, $\sigma_a^2 = \sigma_u^2 + \sigma_q^2 = 0.3 + 0.1 = 0.4$ and $\sigma_e^2 = 0.6$; therefore, $\alpha = \sigma_e^2/\sigma_a^2 = 0.6/0.4 = 1.5$. The **Z** matrix in Eqn 10.21 is now an identity matrix considering animals with records.

The matrix $\mathbf{A}_a$ below was calculated as the sum of $\mathbf{A}_{u(0.3/0.4)}$ and $\mathbf{A}_{v(0.1/0.4)}$. The matrices $\mathbf{A}_u$ and $\mathbf{A}_v$ have been calculated in Examples 10.2 and 10.4.

$$\mathbf{A}_a = \begin{bmatrix} 1.000 & 0.000 & 0.500 & 0.800 & 0.705 \\ 0.000 & 1.000 & 0.500 & 0.200 & 0.295 \\ 0.500 & 0.500 & 1.000 & 0.710 & 0.814 \\ 0.800 & 0.200 & 0.710 & 1.228 & 1.018 \\ 0.705 & 0.295 & 0.814 & 1.018 & 1.338 \end{bmatrix}$$

The vector $\mathbf{s}_i$ for the ith animal needed to calculate $\mathbf{A}_a^{-1}$ is shown in Table 10.2.

The matrix $\mathbf{A}_a^{-1}$ calculated using Eqn 10.11 is:

$$\mathbf{A}_a^{-1} = \begin{bmatrix} 2.2641 & 0.4854 & -0.4101 & -1.2080 & -0.1314 \\ 0.4854 & 1.5007 & -1.0218 & 0.0030 & 0.0327 \\ -0.4101 & -1.0218 & 2.7536 & -0.3673 & -0.9544 \\ -1.2080 & 0.0030 & -0.3673 & 2.9811 & -1.4088 \\ -0.1314 & 0.0327 & -0.9544 & -1.4088 & 2.4619 \end{bmatrix}$$

The MME (Eqn 10.23) for the example data is as follows:

$$\begin{bmatrix} 2.000 & 0.000 & 1.000 & 0.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 3.000 & 0.000 & 1.000 & 0.000 & 1.000 & 1.000 \\ 1.000 & 0.000 & 4.396 & 0.728 & -0.615 & -1.812 & -0.197 \\ 0.000 & 1.000 & 0.728 & 3.251 & -1.533 & 0.005 & 0.049 \\ 1.000 & 0.000 & -0.615 & -1.533 & 5.130 & -0.551 & -1.432 \\ 0.000 & 1.000 & -1.812 & 0.005 & -0.551 & 5.472 & -2.113 \\ 0.000 & 1.000 & -0.197 & 0.049 & -1.432 & -2.113 & 4.693 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{bmatrix} = \begin{bmatrix} 15.3 \\ 17.5 \\ 6.8 \\ 4.5 \\ 8.5 \\ 6.0 \\ 7.0 \end{bmatrix}$$

Table 10.2. Vector ($\mathbf{s}_i$) with contributions from ancestors (animals 1 to 4) to animals 2 to 5, using the pedigree in Example 10.1.

	Elements in $\mathbf{s}_i$ relating to animal			
Animal	1	2	3	4
2	0.0000			
3	0.5000	0.5000		
4	0.5900	–0.0100	0.4200	
5	0.0534	–0.0133	0.3877	0.5722

Downloaded From Vet-eBooks.com

Solving the MME equations gave these results:

Effects	Solutions
Sex	
Male	7.356
Female	5.529
Animal	Total additive genetic merit including marker information
3	0.167
4	–0.167
5	0.419
6	0.432
7	0.645

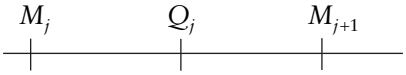
The application of Eqn 10.21 is valuable as only one equation is fitted per animal, but its application to a large data set may be limited because of the tabular method of calculating the relationship matrix needed and its inverse.

10.9 Analysis of Data with QTL Bracketed by Two Markers

This section deals with the extension of the model of Fernando and Grossman (1989) by Goddard (1992) to handle situations in which MQTL is bracketed between two markers. The use of marker information when MQTL is bracketed between two markers should enhance the accuracy of EBVs compared with information with a single marker.

10.9.1 Basic model

Consider a chromosome with a series of marker loci with at most one QTL located between each pair of markers:



Each animal inherits two alleles at the Q_j locus: one from its sire and the other from its dam. A marker haplotype consisting of the marker alleles at M_j and M_{j+1} would be associated with each of the MQTL alleles. Let the jth chromosome segment that animal i inherited from its sire be of the marker haplotype (kl) and the value of the MQTL allele be $v_{ij(kl)}$ or simply $v_{ij(p)}$. Similarly, let the value of the MQTL allele from its dam be $v_{ij(m)}$. Summed over all chromosome segments, the breeding value of animal i (a_i) is:

$$a_i = u_i + \sum_j v_{ij(p)} + \sum_j v_{ij(m)}$$

Similar to Eqn 10.3 the model for the phenotypic record of animal i is:

$$y_i = x_i\beta + u_i + \sum_j v_{ij(p)} + \sum_j v_{ij(m)} + e$$

Downloaded From Vet-eBooks.com

or in matrix notation the model is:

$$\mathbf{y} = \mathbf{x\beta} + \mathbf{Zu} + \sum_j \mathbf{W}_j \mathbf{v}_j + \mathbf{e}$$

The terms are as defined in Eqn 10.17. The vector $\mathbf{v}_j$ contains the effects of the paternal and maternal MQTL alleles at each locus. The summation is over chromosome segments bounded by markers. The variance of $\mathbf{u}$ and $\mathbf{v}_j$ are as defined in Eqn 10.19, such that:

$$\text{var}(\mathbf{v}_j) = \mathbf{G}_{vj} J_{vj}^2$$

Assuming $j = 2$, the BLUP equations for the above model are:

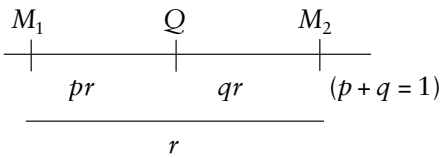
$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W}_1 & \mathbf{X'W}_2 \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z'W}_1 & \mathbf{Z'W}_2 \\ \mathbf{W}_1'\mathbf{X} & \mathbf{W}_1'\mathbf{Z} & \mathbf{W}_1'\mathbf{W}_1' + \mathbf{G}_{v1}^{-1}\alpha_2 & \mathbf{W}_1'\mathbf{W}_2 \\ \mathbf{W}_2'\mathbf{X} & \mathbf{W}_2'\mathbf{Z} & \mathbf{W}_2'\mathbf{W}_1 & \mathbf{W}_2'\mathbf{W}_2 + \mathbf{G}_{v2}^{-1}\alpha_3 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W}_1'\mathbf{y} \\ \mathbf{W}_2'\mathbf{y} \end{pmatrix} \quad (10.24)$$

where:

$$\alpha_1 = \sigma_e^2/\sigma_u^2, \quad \alpha_2 = \sigma_e^2/\sigma_{v1}^2 \quad \text{and} \quad \alpha_3 = \sigma_e^2/\sigma_{v2}^2$$

10.9.2 Calculating the covariance matrix, G

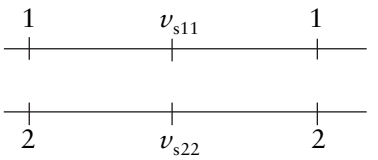
Consider a single MQTL bounded by two marker loci with marker distances as follows:



With the assumption of no crossover, the recombination rates are (Haldane, 1919) between:

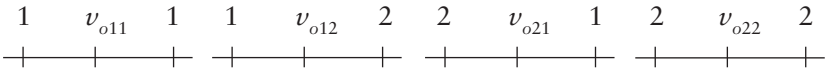
M_1 and $M_2 = a = 0.5(1 - e^{-2r})$
M_1 and $Q = b = 0.5(1 - e^{-2pr})$
Q and $M_2 = c = 0.5(1 - e^{-2qr})$

Similar to the situation with a single marker, the variance of $\mathbf{v}$ depends on the relationship among the $\mathbf{v}$ terms. The MQTL alleles in the progeny can be expressed in terms of parental MQTL. Thus given, for instance, that the genotype of the sire is:



Downloaded From Vet-eBooks.com

The sire will produce the following four types of gametes on the basis of marker haplotypes:



Assuming no double recombination between markers, the frequency, means and approximate means for the four gametes (Goddard, 1992) are:

Haplotype		Frequency	Mean	Approximate mean
1	1	$\frac{1}{2}(1-a)$	$[(1-b)(1-c)/1-a]v_{s11} + [bc/1-a]v_{s22}$	v_{s11}
1	2	$\frac{1}{2}a$	$[(1-b)c/a]v_{s11} + [b(1-c)/a]v_{s22}$	$qv_{s11} + pv_{s22}$
2	1	$\frac{1}{2}a$	$[b(1-c)/a]v_{s11} + [(1-b)c/a]v_{s22}$	$pv_{s11} + qv_{s22}$
2	2	$\frac{1}{2}(1-a)$	$[bc/1-a]v_{s11} + [(1-b)(1-c)/1-a]v_{s22}$	v_{s22}

Given, for instance, that $r = 0.2$, $p = 0.8$ and $q = 0.2$, then a, b and c are 0.1649, 0.1370 and 0.0385, respectively. The means for the haplotypes are $0.99v_{s11}$ and $0.01v_{s11}$ for (1 1), $0.2v_{s11}$ and $0.8v_{s11}$ for (1 2), $0.8v_{s11}$ and $0.2v_{s11}$ for (2 1) and $0.01(v_{s11})$ and $0.99(v_{s11})$ for (2 2). The approximate means are very similar to these estimates. The maximum errors associated with the above approximate means are when $p = q = 0.5$ for haplotypes (1 1) and (2 2) (Goddard, 1992). Using the approximate means, the value of the MQTL in each gamete can be written in terms of the parental MQTL as:

$$\begin{pmatrix} v_{o11} \\ v_{o12} \\ v_{o21} \\ v_{o22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ q & p \\ p & q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_{s11} \\ v_{s22} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix} \tag{10.25}$$

where ε_{ij} is the deviation of each gamete from the mean of the haplotype. Since v_{o11} is identical to v_{s11} and v_{o22} to v_{s22} with the approximate means, then $\varepsilon_{11} = \varepsilon_{22} = 0$. Eqn 10.25 may be expressed as:

$$\mathbf{v} = \mathbf{P}\mathbf{v} + \boldsymbol{\varepsilon}$$

where $\mathbf{P}$ is as defined in Section 10.5 and has at most two non-zero elements, which sum to unity. Thus:

$$\mathbf{v} = (\mathbf{I} - \mathbf{P})^{-1}\boldsymbol{\varepsilon}$$

Therefore:

$$\mathbf{G} = \text{var}(\mathbf{v}) = (\mathbf{I} - \mathbf{P})^{-1}\text{var}(\boldsymbol{\varepsilon})((\mathbf{I} - \mathbf{P})^{-1})'$$

and:

$$\mathbf{G}^{-1} = (\mathbf{I} - \mathbf{P})'\mathbf{H}^{-1}(\mathbf{I} - \mathbf{P}) \tag{10.26}$$

where $\mathbf{H}\sigma_v^2 = \text{var}(\boldsymbol{\varepsilon})$ and $\mathbf{H}$ is a diagonal matrix. Since $\varepsilon_{11} = \varepsilon_{22} = 0$, $\text{var}(\varepsilon_{11}) = \text{var}(\varepsilon_{22}) = 0$. The main interest therefore is in calculating $\text{var}(\varepsilon_{12})$ and $\text{var}(\varepsilon_{21})$. The calculation of

Downloaded From Vet-eBooks.com

either var(ε_{12}) or var(ε_{21}) is similar to that for var(ε) in Section 10.5. For instance, for the oth progeny:

$$\begin{aligned}
\varepsilon_{o12} &= v_{o12} - q(v_{s11}) - p(v_{s22}) \\
\text{var}(\varepsilon_{o12}) &= \text{var}(v_{o12}) - q^2\text{var}(v_{s11}) - p^2\text{var}(v_{s22}) - 2qp\text{cov}(v_{s11}, v_{s22}) \\
&= \sigma_v^2 - (1-p)^2\sigma_v^2 + p^2\sigma_v^2 - 2(1-p)p\sigma_v^2 F_s \\
&= 2\sigma_v^2((1-p)p - (1-p)pF_s) \\
&= 2\sigma_v^2(1-p)p(1-F_s) = 2\sigma_v^2\, pq(1-F_s) \\
\text{var}(\varepsilon_{o12})/\sigma_v^2 &= \mathbf{H} = 2pq(1-F_s)
\end{aligned}$$

Therefore, if the sire is not inbred, the diagonal element of **H** for progeny o(h_{oo}) with the allele v_{o12} equals $2pq$. If the sire is unknown, $h_{oo} = 1$. Similarly, for a progeny o with allele v_{o21}, $h_{oo} = 2qp$ if the sire is known, otherwise 0 if the sire is unknown.

The matrix **G** can be calculated using rules similar to those defined in Section 10.3. The relationship of the MQTL paternal allele of a progeny o with MQTL alleles of individuals 1 to (o – 1) can be calculated using Eqn 10.7, with $\rho_o^p = p$ when o inherits marker haplotype v_{s12} or $\rho_o^p = (1-p)$ when o inherits marker haplotype v_{s21}. Similarly, for the maternal MQTL allele, Eqn 10.8 can be used with $\rho_o^m = p$ when o inherits marker haplotype v_{m12} or $\rho_o^m = (1-p)$ when o inherits marker haplotype v_{m21}.

Using Eqn 10.26, Goddard (1992) derived the following rules for calculating $\mathbf{G}_v^{-1}$.

1. Replace v_{o11} with v_{s11} in all equations and then delete the row and column for v_{o11} in $\mathbf{G}^{-1}$. Similarly, replace v_{o22} with v_{s22}. Set $\mathbf{G}^{-1}$ to zero.
2. For progeny allele v_{o12}, add:
   $q/2p$ to the element corresponding to (v_{s11}, v_{s11})
   $p/2q$ to the element corresponding to (v_{s22}, v_{s22})
   $1/2pq$ to the element corresponding to (v_{o12}, v_{o12})
   $-1/2p$ to the element corresponding to (v_{s11}, v_{o12}) and (v_{o12}, v_{s11})
   $-1/2q$ to the element corresponding to (v_{s22}, v_{o12}) and (v_{o12}, v_{s22})
   1/2 to the element corresponding to (v_{s11}, v_{s22}) and (v_{s22}, v_{s11})
3. For a progeny allele v_{o21}, replace p with q and v_{o12} with v_{o21} in the rules above.
4. For an allele v_{s11} without known parents, add 1 to element corresponding to (v_{s11}, v_{s11})

Goddard (1992) indicated that the use of the approximate means to calculate **P** implies that v_{s11} and v_{o11} are forced to be identical even if double crossover occurs. Therefore, it might be desirable to use a correlation (r) slightly less than unity between v_{s11} and v_{o11}. This is achieved by using:

$$v_{o11} = (1 - r^2/4)v_{s11} + r^2/4v_{s22} + \varepsilon_{11}$$

Then the row and column for v_{o11} are retained in $\mathbf{G}_v^{-1}$, and, in the above rules, v_{o12} is replaced by v_{o11} and p by $r^2/4$.

10.9.3 An illustration

Example 10.5

Consider that the four calves in the following data set have the following genotype at two linked loci.

Downloaded From Vet-eBooks.com

			Genotype at the two linked markers	
Animal	Sire	Dam	Marker 1	Marker 2
1	–	–	11	22
2	–	–	33	44
3	1	3	12	44
4	4	3	21	14

Assuming no double crossing over, the genetic parameters as in Example 10.2 and letting p and q be equal to 0.8 and 0.2, respectively, predict the effects of the sex of the calf, additive genetic effects (breeding values) not linked to the MQTL for animals and additive genetic effects for the MQTL alleles of animals.

The alleles at the MQTL can be defined from the genotypes at the two linked marker loci. Thus the paternal and maternal MQTL alleles for animal 1 will be v_{s11} and v_{s22}, respectively. Correspondingly, those for animal 4 will be v_{o21} and v_{o14}, respectively. As in Example 10.3, $\alpha_1 = 0.6./0.3 = 2$ and $\alpha_2 = 0.6/0.05 = 12$. With the assumption of no double crossing over, for calf 3, $v_{o44} = v_{m44}$ (calf 2); therefore, the row and column for v_{o44} are deleted from $\mathbf{G}_v$ and the MME.

The design matrix $\mathbf{Z}$ is an identity matrix of order four and $\mathbf{W}$ is:

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{W'W} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The covariance matrix $\mathbf{G}_v$ is:

$$\mathbf{G}_v = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.200 & 0.800 & 0.040 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.800 & 0.200 & 0.160 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.800 \\ 0.200 & 0.800 & 0.000 & 0.000 & 1.000 & 0.320 & 0.200 \\ 0.800 & 0.200 & 0.000 & 0.000 & 0.320 & 1.000 & 0.064 \\ 0.040 & 0.160 & 0.000 & 0.800 & 0.200 & 0.064 & 1.000 \end{bmatrix}$$

The calculation of $\mathbf{G}_v$ with elements $g(i,j)$ for the first few animals is as follows. For the first two animals, both parents are unknown; therefore, the diagonal element of $\mathbf{G}$ for either the paternal or maternal allele is 1 for these animals. Calf 3 inherited marker haplotype v_{s12} from its sire; therefore, $\rho_o^p = p$ in Eqn 10.7. Thus:

$$g(_{3p3p,1p1p}) = (1 - p)g(_{1p1p,1p1p}) + pg(_{1m1m,1p1p}) = q(1) + p(0) = q = 0.2$$
$$g(_{3p3p,1m1m}) = (1 - p)g(_{1p1p,1m1m}) + pg(_{1m1m,1m1m}) = q(0) + p(1) = p = 0.8$$
$$g(_{3p3p,2p2p}) = (1 - p)g(_{1p1p,2p2p}) + pg(_{1m1m,2p2p}) = q(0) + p(0) = 0$$
$$g(_{3p3p,2m2m}) = (1 - p)g(_{1p1p,2m2m}) + pg(_{1m1m,2m2m}) = q(0) + p(0) = 0$$

Downloaded From Vet-eBooks.com

The marker haplotype inherited by calf 4 from its sire is v_{s21}; therefore, $\rho_o^p = q$ in Eqn 10.7. Thus:

$$g(_{4p4p,1p1p}) = (1-q)g(_{1p1p,1p1p}) + qg(_{1m1m,1p1p}) = p(1) + q(0) = p = 0.8$$
$$g(_{4p4p,1m1m}) = (1-q)g(_{1p1p,1m1m}) + qg(_{1m1m,1m1m}) = p(0) + q(1) = q = 0.2$$
$$g(_{4p4p,2p2p}) = (1-q)g(_{1p1p,2p2p}) + qg(_{1m1m,2p2p}) = p(0) + q(0) = 0$$
$$g(_{4p4p,2m2m}) = (1-q)g(_{1p1p,2m2m}) + qg(_{1m1m,2m2m}) = p(0) + q(0) = 0$$
$$g(_{4p4p,3m3m}) = (1-q)g(_{1p1p,3m3m}) + qg(_{1m1m,3m3m}) = p(q) + q(p) = 2pq = 0.32$$

The inverse of $\mathbf{G}_v$ is:

$$\mathbf{G}_v^{-1} = \begin{bmatrix} 3.125 & 1.000 & 0.000 & 0.000 & -0.625 & -2.500 & 0.000 \\ 1.000 & 3.125 & 0.000 & 0.000 & -2.500 & -0.625 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 3.000 & 0.500 & 0.000 & -2.500 \\ -0.625 & -2.500 & 0.000 & 0.500 & 3.250 & 0.000 & -0.625 \\ -2.500 & -0.625 & 0.000 & 0.000 & 0.000 & 3.125 & 0.000 \\ 0.000 & 0.000 & 0.000 & -2.500 & -0.625 & 0.000 & 3.125 \end{bmatrix}$$

The matrix $\mathbf{G}^{-1}$ was computed using the rules outlined earlier. Thus for the first two animals (first four alleles), add 1 to the diagonal elements since parents of both calves are unknown. For paternal allele of calf 3, add $1/2pq$ to the diagonal element $(3p3p,3p3p)$ of G^{-1}, $q/2p$ to element $(1p1p,1p1p)$, $p/2q$ to element $(1m1m,1m1m)$, $-1/2p$ to elements $(1p1p,3p3p)$ and $(3p3p,1p1p)$, $-1/2q$ to elements $(1m1m,3p3p)$ and $(3p3p,1m1m)$ and 0.5 to the elements $(1p1p,1m1m)$ and $(1m1m,1p1p)$.

The matrix $\mathbf{A}^{-1}$ for the example data can be calculated using the usual rules; therefore, the MME can easily be set up from the design matrices and inverse of the covariance matrices given. Solving the MME by direct inversion gave the following results:

Effects	Solutions
Sex of calf	
Male	7.475
Female	5.091
Breeding values for animals	
1	0.034
2	–0.034
3	0.246
4	0.280
Additive effects for animals at the MQTL	
1p	–0.008
1m	0.005
2p	–0.047
2m	0.049
3p	0.024
4p	0.010
4m	0.059

A similar model to that in Example 10.5 has been used by Boichard *et al.* (2002) for incorporating MQTL information into genetic evaluation for milk production traits in young bulls.

Downloaded From Vet-eBooks.com

11 Computation of Genomic Breeding Values and Genomic Selection

11.1 Introduction

In outbreeding populations, the incorporation of molecular information in breeding programmes on the basis of the linkage analysis, as discussed in Chapter 10, is limited, as the marker maps are rather sparse and linkage between the markers and QTL may not be sufficiently close enough to persist across the population. Thus the linkage phase between marker and QTL must be established for every family in which the marker is intended to be used for selection.

However, a huge amount of variation has been discovered in the genome at the DNA level as a result of sequencing the genomes of most livestock species. The most abundant form of variation is the single nucleotide polymorphisms (SNPs). An SNP is a DNA sequence variation occurring when a single nucleotide (A, T, C or G) in the genome differs between paired chromosomes in an individual. For example, two sequenced DNA fragments from different individuals, AAGC*C*TA to AAGC*T*TA, contain a difference in a single nucleotide. In this case we say that there are two alleles: C and T. Generally, SNPs are diallelic. In view of the high frequency of SNPs in the genome, and developments in genotyping technology that mean many thousands of SNPs can be genotyped very cheaply, they have been proposed as markers for use in QTL analysis and in association studies in place of microsatellites.

The main emphasis of this chapter is the use of SNPs to directly compute EBVs of animals, which are often called direct genomic breeding values (DGV). This is usually combined with some measure of the traditional EBV, say parent index, from an animal model to produce what is termed genomic breeding values (GEBV), which are officially published and used for the selection of animals.

The use of GEBV in the selection of animals has been referred to as genomic selection. Genomic selection requires that markers (SNPs) are in linkage disequilibrium (LD) with the QTLs across the whole population. LD can be defined as the non-random association between the alleles of two loci (e.g. between alleles of a marker and a QTL). Given a marker locus, A (with alleles A_1, A_2), and a QTL locus, B (with alleles B_1 and B_2), on the same chromosome, LD can be measured as the squared correlation (r^2) between the marker and the QTL as:

$$D = \text{freq}(A_1B_1)*\text{freq}(A_2B_2) - \text{freq}(A_1B_2)*\text{freq}(A_2B_1)$$
$$r^2 = D^2/[\text{freq}(A_1)*\text{freq}(A_2)*\text{freq}(B_1)*\text{freq}(B_2)]$$

The r^2 between the marker and the QTL indicates the proportion of the variance for the QTL that can be explained at the marker.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Downloaded From Vet-eBooks.com

The basic assumption is that the use of SNPs as markers enables all QTL in the genome to be traced through the tracing of chromosome segments defined by adjacent SNPs. It is assumed that the effects of the chromosome segments will be the same across the population as a result of the LD between the SNPs and QTL. Thus it is important that marker density is high enough to ensure that all QTL are in LD with at least a marker.

The main advantages of genomic selection are similar to those outlined in Chapter 10 with MAS. Briefly, it results in a reduction of the generation interval, as young animals can be genotyped early in life and their GEBV computed for the purposes of selection. In the dairy cattle situation, GEBV computed early in life can be used to select young bulls, thereby reducing the cost of progeny testing, provided the GEBV are accurate enough. In addition, higher accuracy of GEBV, about 20–30% above that from a parent average, has been reported for young bulls. The computation of GEBV for an individual on the basis of the SNPs it has inherited means that the differences in the genomic merit of full-sibs can be captured.

The implementation of genomic selection involves estimating the SNP effects in a reference population that consists of individuals with phenotypic records and genotypes. This is then followed by prediction of GEBV for selection candidates that do not yet have phenotypes of their own.

11.2 General Linear Model

The general linear model underlying genomic evaluation is of the form:

$$\mathbf{y} = \mathbf{Xb} + \sum_{i}^{m} \mathbf{M}_i \mathbf{g}_i + \mathbf{e} \tag{11.1}$$

where m is the number of SNPs or markers across the genome, $\mathbf{y}$ is the data vector, $\mathbf{b}$ the vector for mean or fixed effects, $\mathbf{g}_i$ the genetic effect of the ith SNP genotype and $\mathbf{e}$ is the error. The matrices $\mathbf{X}$ and $\mathbf{M}_i$ are design matrices for the mean or fixed effects and the ith SNPs, respectively. The matrix $\mathbf{M}$ is of dimension n (number of animals) and m. The assumption is that all the additive genetic variance is explained by all the marker's effects such that the estimate of an animal's total genetic merit or breeding value ($\mathbf{a}$) is: $\mathbf{a} = \sum_{i}^{m} \mathbf{M}_i \mathbf{g}_i$. However, if it is assumed that a certain proportion of the additive genetic variance is not explained by markers, then the model can be extended to include a residual polygenic effect ($\mathbf{u}$), which is the proportion of the additive genetic variance not captured by markers. The model can then be written as:

$$\mathbf{y} = \mathbf{Xb} + \sum_{i}^{m} \mathbf{M}_i \mathbf{g}_i + \mathbf{Wu} + \mathbf{e} \tag{11.2}$$

where $\mathbf{W}$ is the design matrix linking records to random animal or sire effects if an animal or model has been fitted.

11.3 Coding and Scaling Genotypes

As explained in Eqn 11.1, $\mathbf{M}$ is the genotypic matrix that contains which marker alleles each individual inherited. The genotypes of animals are commonly coded as

Downloaded From Vet-eBooks.com

2 and 0 for the two homozygotes (AA and BB) and 1 for the heterozygotes (AB or BA). If alleles are expressed in terms of nucleotides, and the reference allele at a locus is G and the alternative allele is C, then the code is 0 = GG, 1 = GC and 2 = CC. The diagonal elements of MM′ then indicate the individual relationship with itself (inbreeding) and the off-diagonals indicate the number of alleles shared by relatives (VanRaden, 2007).

Commonly, in genomic evaluations (VanRaden, 2008), the elements of **M** are scaled to set the mean values of the allele effects to zero and account for differences in allele frequencies of the various SNPs. Let the frequency of the second or alternative allele at locus j be p_j and then elements of **M** can be scaled by subtracting $2p_j$. Let the element for column j of a matrix **P** equal $2p_j$, then the matrix **Z**, which contained the scaled elements of **M**, can be computed as **Z = M – P**. Note that the sum of the elements of each column of **Z** equals zero. Furthermore, the elements of **Z** can be normalized by dividing the column for marker j by its standard deviation, which is assumed to be $\sqrt{2p_j(1-p_j)}$. This is assuming that the locus is at Hardy Weinberg equilibrium. However, in this chapter **Z** computed as **M – P** has been used.

11.4 Fixed Effect Model for SNP Effects

Several methods for genomic selection were presented by Meuwissen *et al.* (2001), and one such method includes the least squares approach with chromosome segments or SNPs considered as fixed. There is no assumption made about the distribution of the SNP effects and it usually involves two steps.

1. Analysis of each SNP using the simple model in Eqn 11.1, with $\mathbf{g}_i$ defined as the vector of fixed ith SNP effect.
2. Select the k most significant SNPs and estimate their effects simultaneously (in the same data) using a multiple regression with the term for SNP effects in Eqn 11.1 equal to:

$$\sum_{i}^{k} \mathbf{M}_i \mathbf{g}_i$$

This approach suffers from two major limitations. First, the estimation of effects based on an SNP selected by single SNP analysis will result in overestimation of the SNP effects, as the large amount of multiple testing ensures the selected SNPs are those with positive error terms. Second, determining the level of significance for the choice of SNPs to include in the final analysis is far from straightforward.

In an animal breeding context, assuming the few SNPs that have significant effects on a trait have been identified, then these SNPs can fitted as fixed effects in a model that includes the polygenic effect as a random effect. Thus the genomic breeding value for animal i ($GEBV_i$) can be computed as a sum of the direct genomic breeding value (DGV_i) calculated from the marker (SNP) effects as $\mathbf{M}_i\hat{\mathbf{g}}_i$ and the polygenic effects ($\hat{\mathbf{u}}_i$).

Such a linear model could be written as:

$$\mathbf{y = Xb + Zg + Wu + e} \tag{11.3}$$

where **g** represents the fixed marker or SNP effects, **Z** is the scaled matrix of genotypes defined in Section 11.2, which relates SNPs to phenotypes, and other terms are defined as in Eqn 11.2.

Downloaded From Vet-eBooks.com

The equations for obtaining the solutions for SNP and polygenic effects are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{X} & \mathbf{W}\mathbf{W} + \mathbf{A}^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{pmatrix} \tag{11.4}$$

where:

$\alpha = \sigma_e^2/\sigma_u^2$

If the vector of observations, **y** in Eqn 11.3, are de-regressed breeding values of bulls (see Section 5.5.2), then each observation may be associated with differing reliabilities. Thus a weighted analysis may be required to account for these differences in bull reliabilities. The weight (wt_i) for each observation could be the reciprocal of the effective daughter contribution (see Section 5.5.2) or $wt_i = (1/rel_{dtr}) - 1$, where rel_{dtr} is the bull's reliability from daughters with parent information excluded (VanRaden, 2008). Then the MME are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}\mathbf{R}^{-1}\mathbf{W} + \mathbf{A}^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{11.5}$$

where **R** = **D** and **D** is a diagonal matrix with diagonal element $i = wt_i$.

Example 11.1

Given below is the real genotype for the first ten SNPs of a popular dairy bull and those of his sons and some other unrelated bulls genotyped using the 50K Illumin chip. The genotypes of animals are coded as described in Section 11.3. The observations are the DYDs for fat yield, and the effective daughter contribution (EDC) for each bull is also given. The EDC can be used as weights in the analysis. It is assumed the genetic variance for fat yield is 35.241 kg^2 and residual variance of 245 kg^2, and animals 13 to 20 are assumed as the reference population and 21 to 26 as selection candidates. Assuming that the first three SNPs have been identified as having the most significant effect, the aim is to fit Eqn 11.3 with and without weights using these three SNPs:

Animal	Sire	Dam	Mean	EDC	Fat DYD	SNP Genotype									
13	0	0	1	558	9.0	2	0	1	1	0	0	0	2	1	2
14	0	0	1	722	13.4	1	0	0	0	0	2	0	2	1	0
15	13	4	1	300	12.7	1	1	2	1	1	0	0	2	1	2
16	15	2	1	73	15.4	0	0	2	1	0	1	0	2	2	1
17	15	5	1	52	5.9	0	1	1	2	0	0	0	2	1	2
18	14	6	1	87	7.7	1	1	0	1	0	2	0	2	2	1
19	14	9	1	64	10.2	0	0	1	1	0	2	0	2	2	0
20	14	9	1	103	4.8	0	1	1	0	0	1	0	2	2	0
21	1	3	1	13	7.6	2	0	0	0	0	1	2	2	1	2
22	14	8	1	125	8.8	0	0	0	1	1	2	0	2	0	0
23	14	11	1	93	9.8	0	1	1	0	0	1	0	2	2	1

Continued

Downloaded From Vet-eBooks.com

(*Continued*)

Animal	Sire	Dam	Mean	EDC	Fat DYD	SNP Genotype									
24	14	10	1	66	9.2	1	0	0	0	1	1	0	2	0	0
25	14	7	1	75	11.5	0	0	0	1	1	2	0	2	1	0
26	14	12	1	33	13.3	1	0	1	1	0	2	0	1	0	0

EDC, effective daughter contribution; DYD, daughter yield deviation.

The prediction of marker effects and polygenic effects for the reference population and selection candidates can be done simultaneously by including $\mathbf{A}^{-1}$ for all animals but using only the fat yield records for the reference animals. Thus $\mathbf{y}'$ = (9.0 13.4 12.7 15.4 5.9 7.7 10.2 4.8). The incidence matrix $\mathbf{X} = \mathbf{I}_q$, with q = 8 (the number of animals in the reference population).

Computing the matrix Z

The computation of $\mathbf{Z}$ requires calculating the allele frequency for each SNP. The allele frequency for the ith SNP was computed as:

$$\frac{\sum_{j}^{n} m_{ij}}{2*n}$$

where n = 14, the number of animals with genotypes, and m_{ij} are elements of $\mathbf{M}$. The allele frequencies for the ten SNPs were 0.312, 0.179, 0.357, 0.357, 0.143, 0.607, 0.071, 0.964, 0.571 and 0.393, respectively. However, only the first three SNPs are needed for this example, therefore $\mathbf{Z}$ is of order 8 by 3 with elements $z_{i,j} = m_{i,j} - p_{i,j}$, with j =1, 3. Thus:

$$\mathbf{Z} = \begin{pmatrix} 1.357 & -0.357 & 0.286 \\ 0.357 & -0.357 & -0.714 \\ 0.357 & 0.643 & 1.286 \\ -0.643 & -0.357 & 1.286 \\ -0.643 & 0.643 & 0.286 \\ 0.357 & 0.643 & -0.714 \\ -0.643 & -0.357 & 0.286 \\ -0.643 & 0.643 & 0.286 \end{pmatrix}$$

The $\mathbf{W}$ matrix is a diagonal matrix for the eight reference animals with records. This is augmented with 12 columns of zeros to account for ancestors 1 to 12. For the weighted analysis, the $\mathbf{R}$ was a diagonal matrix with the diagonal elements equal to the EDC of the first eight animals in the data set. The matrix $\mathbf{A}^{-1}$ is computed using the usual rules for all 26 animals and α = 245/35.241 = 6.952. Solving the system of equations gives the following results:

	Unweighted analysis	Weighted analysis
Mean effect		
	9.895	9.178
SNP effect		
1	0.607	2.655
2	–4.080	–4.640
3	1.934	2.951

Continued

Downloaded From Vet-eBooks.com

(*Continued*)

	Unweighted analysis		Weighted analysis	
Reference animals				
	DGV	Polygenic	DGV	Polygenic
13	2.834	–0.299	7.070	–0.001
14	0.293	0.256	1.464	0.000
15	0.081	0.142	2.726	0.000
16	3.554	0.254	4.711	0.002
17	–2.460	–0.085	–2.880	–0.001
18	–3.787	0.271	–3.176	0.002
19	1.620	–0.092	1.760	–0.002
20	–2.460	–0.181	–2.880	–0.002
Selection animals				
	DGV	Polygenic	DGV	Polygenic
25	0.900	0.000	4.119	0.000
26	–0.314	0.128	–1.191	0.000
27	–2.460	0.128	–2.880	0.000
28	0.293	0.128	1.464	0.000
29	–0.314	0.128	–1.191	0.000
30	2.227	0.128	4.415	0.000

With this small amount of data, it seems that when records are properly weighted, polygenic effects were very close to zero. The GEBV for reference and selection animals equals $\mathbf{Z}\hat{\mathbf{g}} + \hat{\mathbf{u}}$. This would be equal to 2.535 for animal 13 for instance. The $\mathbf{Z}$ has been given for reference animals and for the selection candidates the corresponding matrix $\mathbf{Z}_2$ is:

$$\mathbf{Z}_2 = \begin{pmatrix} 1.357 & -0.357 & -0.714 \\ -0.643 & -0.357 & -0.714 \\ -0.643 & 0.643 & 0.286 \\ 0.357 & -0.357 & -0.714 \\ -0.643 & -0.357 & -0.714 \\ 0.357 & -0.357 & 0.286 \end{pmatrix}$$

11.5 Mixed Linear Model for Computing SNP Effects

Several methods that fit SNP effects as random have been presented by various researchers (Meuwissen *et al*., 2001; VanRaden *et al*., 2008; Habier *et al*., 2011). The most common random model used in the national evaluation centres for genomic evaluation, especially of dairy animals, assumes the effect of the SNP are normally distributed and all SNP are from a common normal distribution (e.g. the same genetic variance for all SNPs). There are two equivalent models with these assumptions:

1. A model fitting individual SNP effects simultaneously. In this model (SNP-BLUP), DGVs for selection candidates are calculated as DGV = $\mathbf{Z}\hat{\mathbf{g}}$, where $\hat{\mathbf{g}}$ are the estimates of random SNP effects. This method involves knowing σ^2_g, but this may not be the case in practice, and σ^2_g may have to be approximated from σ^2_a, the additive genetic variance. In such situations, this method is also referred as ridge regression.

Downloaded From Vet-eBooks.com

2. A model estimating breeding values directly, with the (co)variance among breeding values $\mathbf{G}\sigma_a^2$ fitted, where **G** is the genomic relationship matrix. The matrix **G** represents the realized proportion of the genome that animals share in common and is estimated from the SNPs.

These models will now be described in more detail.

11.5.1 SNP-BLUP model

In matrix form, the mixed linear model for estimating SNP effects can be written as (Meuwissen *et al*., 2001; VanRaden, 2008):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e} \tag{11.6}$$

where **g** is a vector of additive genetic effects corresponding to allele substitution effects for each SNP and all other terms defined as in Eqn 11.3. The matrix **Z** relates SNP effects to the phenotypes. The sum of **g** over all marker loci is assumed equal to the vector of breeding values (**a**), i.e. DGV = **a** = **Zg**. The MME for Eqn 11.6 are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{I}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{11.7}$$

where $\alpha = \sigma_e^2/\sigma_g^2$ and **R** is a diagonal matrix of weights (see Eqn 11.5). The MME in Eqn 11.7 can easily be set up and solutions obtained for each SNP and the fixed effects. However, in practice, the value of σ_g^2 may not be known and σ_g^2 could be obtained either as $\sigma_g^2 = \sigma_a^2/m$, with m = the number of markers, or as $\sigma_g^2 = \sigma_a^2/2\Sigma p_j(1-p_j)$. The latter is preferred as it takes into account the differences in allele frequencies. With the latter, $\alpha = 2\Sigma p_j(1-p_j)^*[\sigma_e^2/\sigma_a^2]$, with σ_a^2 being the additive genetic variance for the trait and p_j is as defined in Section 11.3. Hayes and Daetwyler (2013) indicated that there is a potential problem with this estimate as it assumes the LD between SNP and QTL is perfect and all genetic variance is captured by the SNP. This may not be the case in practice and they recommended the method described by Moser *et al.* (2010) for estimating α through cross-validation. The method involves estimating SNP effects with different values of α and predicting DGV in validation data sets that have not contributed to the estimation of SNP effects. The value of α that minimizes the mean square error between the DGV and **y** is taken as the appropriate estimate. This process can be repeated, dropping out different subsets of the data and obtaining an estimate of α by averaging across data sets.

Example 11.2

Using the data and genetic parameters given in Example 11.1, SNP effects are predicted using Eqn 11.6 and all ten SNPs. Then DGVs are computed for the reference and validation animals. Initially, analyses are carried out without weights, thus $\mathbf{R} = \mathbf{I}\sigma_e^2$. Then the data were re-analysed using EDCs as weights, with **R** in Eqn 11.7 being a diagonal matrix containing EDCs for reference bulls.

COMPUTING THE REQUIRED MATRICES AND α

The allele frequencies for the ten SNPs have been calculated in Example 11.1. Using those frequencies, $2\Sigma p_j(1-p_j) = 3.5383$. Thus $\alpha = 3.5383^*(245/35.242) = 24.598$.

Downloaded From Vet-eBooks.com

The matrix **X** in Eqn 11.7 is the same as **X** in Example 11.1 and **Z** computed as **Z** = **M** – **P** is:

$$\mathbf{Z} = \begin{pmatrix} 1.357 & -0.357 & 0.286 & 0.286 & -0.286 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ 0.357 & -0.357 & -0.714 & -0.714 & -0.286 & 0.786 & -0.143 & 0.071 & -0.143 & -0.786 \\ 0.357 & 0.643 & 1.286 & 0.286 & 0.714 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ -0.643 & -0.357 & 1.286 & 0.286 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & 0.214 \\ -0.643 & 0.643 & 0.286 & 1.286 & -0.286 & -1.214 & -0.143 & 0.071 & -0.143 & 1.214 \\ 0.357 & 0.643 & -0.714 & 0.286 & -0.286 & 0.786 & -0.143 & 0.071 & 0.857 & 0.214 \\ -0.643 & -0.357 & 0.286 & 0.286 & -0.286 & 0.786 & -0.143 & 0.071 & 0.857 & -0.786 \\ -0.643 & 0.643 & 0.286 & -0.714 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & -0.786 \end{pmatrix}$$

The MME in Eqn 11.7 can then be easily set up. The solutions for the mean and SNP effects from solving the MME, either using weights or no weights, are shown in Table 11.1. The DGV for the reference animals is then computed as $\mathbf{Z}\hat{\mathbf{g}}$. The results are shown in Table 11.2.

Similarly, the DGV of the validation animals are computed as $\mathbf{Z}_2\hat{\mathbf{g}}$, where $\mathbf{Z}_2$ contains the centralized genotypes for the selection candidates. Thus for the unweighted analysis:

$$\begin{bmatrix} \hat{a}_{21} \\ \hat{a}_{22} \\ \hat{a}_{23} \\ \hat{a}_{24} \\ \hat{a}_{25} \\ \hat{a}_{26} \end{bmatrix} = \begin{pmatrix} 1.357 & -0.357 & -0.714 & -0.714 & -0.286 & -0.214 & 1.857 & 0.071 & -0.143 & 1.214 \\ -0.643 & -0.357 & -0.714 & 0.286 & 0.714 & 0.786 & -0.143 & 0.071 & -1.143 & -0.786 \\ -0.643 & 0.643 & 0.286 & -0.714 & -0.286 & -0.214 & -0.143 & 0.071 & 0.857 & 0.214 \\ 0.357 & -0.357 & -0.714 & -0.714 & 0.714 & -0.214 & -0.143 & 0.071 & -1.143 & -0.786 \\ -0.643 & -0.357 & -0.714 & 0.286 & 0.714 & 0.786 & -0.143 & 0.071 & -0.143 & -0.786 \\ 0.357 & -0.357 & 0.286 & 0.286 & -0.286 & 0.786 & -0.143 & -0.929 & -1.143 & -0.786 \end{pmatrix} \begin{pmatrix} 0.087 \\ -0.311 \\ 0.262 \\ -0.080 \\ 0.110 \\ 0.139 \\ 0.000 \\ 0.000 \\ -0.061 \\ -0.016 \end{pmatrix}$$

$$= \begin{pmatrix} 0.027 \\ 0.114 \\ -0.240 \\ 0.143 \\ 0.054 \\ 0.354 \end{pmatrix}$$

11.5.2 Equivalent models: GBLUP

An equivalent model to Eqn 11.6 is the application of the usual BLUP MME but with the inverse of the numerator relationship matrix ($\mathbf{A}^{-1}$) replaced by the inverse of the genomic relationship matrix ($\mathbf{G}^{-1}$) (Habier *et al*., 2007; Hayes *et al*., 2009). This tends to be referred to generally as GBLUP. The DGVs are computed directly from the MME as the sum of the SNP effects (**a** = **Zg**), with the assumption that SNP effects are normally distributed. Assume the following mixed linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wa} + \mathbf{e} \tag{11.8}$$

where **y** is the vector of observations, **a** is the vector of DGVs and **W** is the design matrix linking records to breeding value (random animal or sire effect if an animal or sire model has been fitted) and **e** is random residual effect. Given that **a** = **Zg**, then:

$$\text{var}(\mathbf{a}) = \mathbf{ZZ}'\sigma_g^2$$

Downloaded From Vet-eBooks.com

Table 11.1. Solutions for mean and SNP effects from various models.

	Unweighted	Weighted
Mean effect		
	9.944	11.876
SNP effects solutions		
1	0.087	–0.633
2	–0.311	–3.041
3	0.262	3.069
4	–0.080	–1.267
5	0.110	2.600
6	0.139	4.447
7	0.000	0.000
8	0.000	0.000
9	–0.061	–3.240
10	–0.016	1.883

Table 11.2. Direct genomic breeding (DGV) values from various models.

	SNP-BLUP	GBLUP	Selection index	SNP-BLUP (weighted)
Reference animals				
13	0.070	0.069	0.070	–2.651
14	0.111	0.116	0.111	1.307
15	0.045	0.049	0.045	0.611
16	0.253	0.260	0.253	1.007
17	0.495	–0.500	–0.495	–5.693
18	–0.357	–0.359	–0.357	–4.358
19	0.145	0.146	0.146	0.502
20	–0.224	–0.231	–0.225	–5.718
Selection candidates				
21	0.027	0.028	0.028	–0.006
22	0.114	0.115	0.115	6.513
23	–0.240	–0.240	–0.240	–3.835
24	0.143	0.143	0.143	2.701
25	0.054	0.054	0.054	3.273
26	0.354	0.353	0.353	6.350

Noting that:

$$\sigma_a^2 = \frac{\sigma_g^2}{2\Sigma p_j(1-p_j)}$$

then the matrix $\mathbf{ZZ'}$ can be scaled such that:

$$\mathbf{G} = \frac{\mathbf{ZZ'}}{2\Sigma p_j(1-p_j)}$$

and $\text{var}(\mathbf{a}) = \mathbf{G}\sigma_a^2$. The above division scales **G** to be analogous to the numerator relationship matrix (**A**). The genomic inbreeding coefficient for individual i is $G_{ii} - 1$, and the genomic relationship between individuals i and k, which are analogous to the relationship coefficients (Wright, 1922), can be obtained by dividing the elements G_{ij} by the square roots of the diagonals of G_{ii} and G_{jj}. The matrix **G** is generally positive

Downloaded From Vet-eBooks.com

semi-definite but can be singular if two individuals have identical genotypes or the number of markers (m) is less than genotyped individuals (n). If number of markers are limited ($m < n$), an improved non-singular matrix $\mathbf{G}_{wt}$ can be obtained as wt$\mathbf{G}$ + (1 – wt)$\mathbf{A}$. VanRaden (2008) indicated that wt = 0.90, 0.95 and 0.98 gave good results.

Another method for computing $\mathbf{G}$ involves scaling $\mathbf{ZZ'}$ by the reciprocals of the expected variance of marker loci (VanRaden, 2008). Thus $\mathbf{G} = \mathbf{ZDZ'}$, where $\mathbf{D}$ is diagonal with:

$$d_{ii} = \frac{1}{m[2p_j(1-p_j)]}$$

The MME for Eqn 11.8 are:

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}W + G^{-1}}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{pmatrix} \tag{11.9}$$

where α now equals σ_e^2/σ_a^2.

This approach for genomic evaluation has the advantage that existing software for genetic evaluation can be used by replacing $\mathbf{A}$ with $\mathbf{G}$ and the systems of equations are of the size of animals, which tend to be fewer than the number of SNPs. In pedigree populations, $\mathbf{G}$ discriminates among sibs, and other relatives, allowing us to say whether these sibs are more or less alike than expected, so we can capture information on Mendelian sampling. Also, the method is attractive for populations without good pedigree, as $\mathbf{G}$ will capture this information among the genotyped individuals (Hayes and Daetwyler, 2013).

Note that Eqn 11.9 assumes all the additive genetic variance (σ_a^2) is captured by the SNP, but this may not be the case if the linkage disequilibrium between SNP and QTL is not perfect. Later, in Section 11.6, a model is discussed that might capture any residual polygenic variance not captured by the SNPs. Another possible limitation is that there are no direct rules for computing $\mathbf{G}^{-1}$ and in large populations the computation may not be feasible.

Example 11.3

The data in Example 11.1 is analysed using Eqns 11.8 and 11.9 and the same genetic parameters to compute DGVs for both the reference and validation animals without using weights.

The matrix $\mathbf{X}$ in Eqn 11.9 is the same $\mathbf{X}$ as in Example 11.1, $\mathbf{W}$ is a diagonal matrix for the eight reference animals with records and α = 245/35.25 = 6.950.

The $\mathbf{G}$ matrix constructed from $\mathbf{Z}$ for the ten SNPs as:

$$\frac{\mathbf{ZZ'}}{2\Sigma p_j(1-p_j)}$$

with $2\Sigma p_j(1 - p_j)$ = 3.5383 is:

13	1.472								
14	–0.446	0.746							
15	0.988	–0.930	1.634						
16	0.059	–0.446	0.422	0.907					
17	0.685	–0.950	1.048	0.402	1.593				
18	–0.163	0.180	–0.365	–0.163	–0.102	0.746			
19	–0.708	0.201	–0.627	0.423	–0.365	0.201	0.786		
20	–0.547	0.079	–0.183	0.301	–0.203	0.079	0.382	0.826	

Downloaded From Vet-eBooks.com

```
21   0.887   0.100   0.120  –0.526  –0.183   0.100  –0.728  –0.567   2.280
22  –0.789   0.402  –0.708  –0.506  –0.446  –0.163   0.140  –0.264  –0.526   1.190
23  –0.203  –0.143   0.160   0.362   0.140   0.140   0.160   0.604  –0.224  –0.486   0.665
24  –0.143   0.483  –0.345  –0.708  –0.648  –0.365  –0.345  –0.183   0.120   0.705  –0.405  1.068
25  –0.829   0.362  –0.748  –0.264  –0.486   0.079   0.382  –0.022  –0.567   0.867  –0.244  0.382  0.826
26  –0.264   0.362  –0.466  –0.264  –0.486  –0.203   0.100  –0.304  –0.284   0.584  –0.526  0.382  0.261  1.109
```

For the purposes of comparison, the **G** matrix computed from 41866 SNPs ($\mathbf{G}_{all}$) with $2\Sigma p_j(1 - p_j)$ = 15555.80 and the **A** computed from a five-generation pedigree are shown below:

$\mathbf{G}_{all}$ =

```
13   0.957
14  –0.108   0.973
15   0.452  –0.116   1.182
16   0.209  –0.058   0.424   1.025
17   0.234  –0.083   0.425   0.312   1.037
18  –0.040   0.438   0.097  –0.047  –0.043   1.151          symmetric
19  –0.089   0.458   0.039  –0.067  –0.070   0.426   1.175
20  –0.093   0.460   0.053  –0.058  –0.063   0.432   0.707   1.183
21   0.077  –0.082   0.064   0.104   0.082  –0.071  –0.069  –0.069   1.031
22  –0.056   0.418   0.093  –0.046  –0.038   0.408   0.355   0.342  –0.044  1.139
23  –0.005   0.464  –0.038  –0.035  –0.038   0.206   0.223   0.215   0.011  0.280  0.993
24  –0.070   0.468   0.075  –0.027  –0.053   0.403   0.521   0.550  –0.079  0.424  0.260  1.198
25  –0.052   0.416   0.098  –0.009  –0.031   0.386   0.363   0.342  –0.038  0.370  0.219  0.419  1.125
26  –0.070   0.493  –0.084  –0.039  –0.044   0.258   0.241   0.270  –0.072  0.253  0.178  0.259  0.214  1.009
```

A =

```
13  1.008
14  0.033  1.037
15  0.545  0.021  1.041
16  0.288  0.021  0.536  1.016
17  0.285  0.031  0.541  0.293  1.020
18  0.047  0.580  0.036  0.028  0.032  1.062
19  0.033  0.613  0.021  0.021  0.031  0.365  1.095          symmetric
20  0.033  0.613  0.021  0.021  0.031  0.365  0.613  1.095
21  0.099  0.031  0.082  0.118  0.074  0.028  0.031  0.031  1.021
22  0.046  0.586  0.032  0.031  0.039  0.351  0.373  0.373  0.044  1.068
23  0.096  0.569  0.067  0.043  0.047  0.329  0.357  0.357  0.042  0.338  1.050
24  0.041  0.574  0.027  0.019  0.026  0.331  0.406  0.406  0.028  0.335  0.335  1.056
25  0.033  0.548  0.035  0.039  0.039  0.315  0.336  0.336  0.037  0.321  0.310  0.310  1.029
26  0.035  0.588  0.023  0.024  0.039  0.337  0.376  0.376  0.036  0.347  0.341  0.348  0.325  1.070
```

The matrix **A** is more similar to $\mathbf{G}_{all}$ than to **G**, thus with more SNPs, the genomic relationship matrix captures more relationships.

The matrices required for Eqn 11.9 have been described. Solving Eqn 11.9 gives the DGVs directly for both the reference and selection animals and these are shown in Table 11.2. The solution for the mean effects was 9.944. Thus the model gave the same results as the SNP model.

11.5.3 Equivalent models: selection index approach

VanRaden (2008) presented a selection index approach which is equivalent to Eqn 11.9. The method is of limited use in practice as it is assumed that the solutions of the vector of fixed ($\hat{\mathbf{b}}$) effects are known. It does, however, demonstrate the equivalence of the selection index approach to GBLUP.

Downloaded From Vet-eBooks.com

Selection index equations to predict DGV ($\hat{\mathbf{a}}$) are constructed as the covariance between **y** and **a** multiplied by the inverse of the variance of **y** and the deviation of **y** from fixed effects solutions. Thus:

$$\hat{\mathbf{a}} = \mathbf{G}\left(\mathbf{G} + \mathbf{R}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right)^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{11.10}$$

The vector of estimates of SNP effects ($\hat{\mathbf{g}}$) can be obtained from Eqn 11.10 as:

$$\hat{\mathbf{g}} = \left(\frac{1}{2\sum p_j(1-p_j)}\right) \quad \mathbf{Z}'\left(\mathbf{G} + \mathbf{R}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right)^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{11.11}$$

The DGV of validation candidates without records can then be computed with the selection index approach as:

$$\hat{\mathbf{a}} = \mathbf{C}\left(\mathbf{G} + \mathbf{R}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right)^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{11.12}$$

where **C** is the genomic covariance between animals with and without records computed as:

$$\frac{\mathbf{Z}_2\mathbf{Z}'}{2\sum p_j(1-p_j)}$$

with $\mathbf{Z}_2$ being the matrix of centralized genotypes for the validation animals (see Example 11.3).

Example 11.4

The data in Example 11.1 is again analysed using Eqn 11.10 and the same genetic parameters to compute DGVs for the reference animals without using weights. The solution of 9.994 has been assumed for the mean.

The **X** matrix in Eqn 11.10 equals **X** in Example 11.1, the **G** matrix is of order 8 for the reference animals only and corresponds to the first eight rows and columns of **G** computed in Example 11.3 and $\mathbf{R} = \mathrm{I}\sigma_e^2$, assuming no weights are used in the analysis.

Solutions from solving Eqn 11.10 are shown in Table 11.2. Similarly, the DGV of the selection candidates were obtained by Eqn 11.12 and these are also shown in Table 11.2. The same solutions were obtained for both reference and validation animals as obtained from the SNP or GBLUP models.

11.6 Mixed Linear Models with Polygenic Effects

The genomic BLUP model used to estimate SNP effects in most livestock populations is based on chips with densities of about 60K, and it is usually assumed that these SNPs explain all the genetic variation for the traits analysed. However, fitting a residual polygenic effect (RP) may account for the fact that SNPs may not explain all the genetic variance and it has also been found to render SNP effects less biased

Downloaded From Vet-eBooks.com

(Solberg *et al*., 2009). Liu *et al*. (2011) have demonstrated that the optimum level of RP may differ for traits of different heritabilities but tends to vary between 10 and 20% of the genetic variance.

A mixed linear model with polygenic effects included is of this form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wu} + \mathbf{Zg} + \mathbf{e} \tag{11.13}$$

where **u** is the vector of random residual polygenic effects, **W** is the design matrix that relates records to animals and other terms are defined as in Eqn 11.6. If a SNP-BLUP model is fitted, the MME to be solved are:

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}W} & \mathbf{X'R^{-1}Z} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}W} + \mathbf{A^{-1}}\alpha_1 & \mathbf{W'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}W} & \mathbf{Z'R^{-1}Z} + \mathbf{I}\alpha_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{W'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{pmatrix} \tag{11.14}$$

where $\alpha_1 = \sigma_e^2/\sigma_u^2$, with σ_u^2 equal to the chosen percentage of the additive genetic variance fitted as polygenic effect and $\alpha_2 = \sigma_e^2/\sigma_g^2$, with σ_g^2 calculated to account for the percentage of additive genetic variance attributed to the polygenic effect. Thus $\alpha_2 = (\sigma_a^2 - \sigma_u^2)/m$ with m = number of markers or $2\Sigma p_j(1 - p_j)^*[\sigma_e^2/(\sigma_a^2 - \sigma_u^2)]$.

However, if a GBLUP model is to be fitted, then the mixed linear model is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wu} + \mathbf{Wa} + \mathbf{e} \tag{11.15}$$

where **a** is the vector of DGVs and all other terms are as defined in Eqn 11.8. The MME to be solved are:

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}W} & \mathbf{X'R^{-1}Z}^* \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}W} + \mathbf{A^{-1}}\alpha_1 & \mathbf{W'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}W} & \mathbf{W'R^{-1}W} + \mathbf{G^{-1}}\alpha_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{W'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{pmatrix} \tag{11.16}$$

where:

$\alpha_1 = \sigma_e^2/\sigma_u^2$ and $\alpha_2 = \sigma_e^2/(\sigma_a^2 - \sigma_u^2)$

Example 11.5

The data in Example 11.1 is analysed assuming the same genetic parameters to compute DGVs for the reference animals without using weights. It is also assumed that 10% of the additive genetic variance is due to residual polygenic effect in the model. The analysis has been carried out using both Eqns 11.14 and 11.16 without any weights.

Given that $\sigma_a^2 = 35.241$, then $\sigma_u^2 = 0.1*35.241 = 3.5241$. Therefore, for both Eqns 11.14 and 11.16, $\alpha_1 = \sigma_e^2/\sigma_u^2 = 245/3.5241 = 69.521$. However, for Eqn 11.14, $\alpha_2 = \sigma_e^2/\sigma_g^2$ and now equals $2\Sigma p_j(1 - p_j)^*[\sigma_e^2/(\sigma_a^2 - \sigma_u^2)] = 3.5383*(245/(35.241 - 3.5241) = 27.332$, while in Eqn 11.16, $\alpha_2 = \sigma_e^2/(\sigma_a^2 - \sigma_u^2) = 7.725$.

The matrix **Z** in Eqn 11.14 is as defined in Example 11.2, while **W** in Eqns 11.14 and 11.16 have been set up in Example 11.3. The matrix $\mathbf{A^{-1}}$ is for the eight reference animals. All matrices for Eqns 11.14 and 11.16 have therefore been defined. The mean and SNP solutions from solving the MME in Eqn 11.14 are given in Table 11.3.

Downloaded From Vet-eBooks.com

Table 11.3. Mean and SNP effects from SNP-BLUP model with polygenic effects.

Mean effects	
	9.940
SNP effects	
1	0.078
2	–0.280
3	0.234
4	–0.075
5	0.098
6	0.128
7	0.000
8	0.000
9	–0.054
10	–0.018

Table 11.4. Direct genomic breeding values from models with polygenic effects.

	SNP-BLUP model		GBLUP	
	Polygenic	DGV	Polygenic	DGV
Reference animals				
13	0.011	0.066	0.011	0.064
14	–0.007	0.102	–0.007	0.106
15	0.043	0.071	0.043	0.074
16	0.076	0.299	0.076	0.305
17	–0.015	–0.473	–0.015	–0.477
18	–0.025	–0.343	–0.025	–0.345
19	–0.021	0.115	–0.021	0.115
Selection candidates				
20	–0.056	–0.254	–0.056	–0.260
21	0.005	0.028	0.005	0.029
22	–0.006	0.102	–0.006	0.102
23	–0.004	–0.220	–0.004	–0.220
24	–0.008	0.125	–0.008	0.125
25	–0.003	0.051	–0.003	0.051
26	–0.006	0.316	–0.006	0.315

The mean solution from solving Eqn 11.16 was 9.940. The DGVs for the reference and validation populations from both sets of MME are given in Table 11.4. As expected, Eqns 11.14 and 11.16 gave similar results, but for this example, the inclusion of 10% polygenic effects decreased the range of SNP solutions slightly but increased the range for DGVs.

11.7 Single-step Approach

Since the genomic predictions are usually based on a subset of data used for national evaluation, the DGV are usually combined with some measure of conventional breeding values to incorporate additional information in the conventional evaluations.

Downloaded From Vet-eBooks.com

The combined evaluations are called genotypic breeding values (GEBV) and these are usually the published values for selection. The combination of DGVs and the conventional evaluations is based on some sort of selection index approach. The selection index presented by VanRaden *et al.* (2009) was:

$$\text{GEBV} = wt_1\text{DGV} + wt_2 PTA_1 + wt_3 PTA_2$$

for animals in the reference population. Similarly, for selection candidates with no daughter information:

$$\text{GEBV} = wt_1\text{DGV} + wt_2 PA_1 + wt_3 PA_2$$

where PTA_1 and PTA_2 are predicted transmitting abilities from the official evaluations based on all records and the evaluations of only the bulls in the reference population using the **A** matrix, respectively. Correspondingly, PA_1 and PA_2 are parent averages from the respective evaluations. The weights (wt_i) were computed as $\mathbf{c}'\mathbf{V}^{-1}$. The matrix **V** is of order 3 × 3 with diagonal elements equal to the reliabilities for DGV, PTA_1 (PA_1) and PTA_2 (PA_2), respectively. The off-diagonal elements were calculated as $v_{12} = v_{22}$, $v_{23} = v_{22}$ and $v_{13} = v_{22} + (v_{11} - v_{22})(v_{33} - v_{22})/(1 - v_{22})$. The vector **c** has elements v_{11}, v_{22} and v_{33}.

Misztal *et al.* (2010) presented a method called the single-step approach that combines conventional and DGVs in one step, resulting in the direct prediction of EBVs for non-genotyped and GEBV for genotyped animals.

Assume the following mixed linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wa} + \mathbf{e} \tag{11.17}$$

where **y** = vector of phenotypes or de-regressed breeding values, **a** = vector breeding values and **W** is a design matrix that relates records to all animals including genotyped and ungenotyped animals. Suppose **a** is portioned as $\mathbf{a}_1$ for ungenotyped animals and $\mathbf{a}_2$ for genotyped animals, then:

$$\text{var}\begin{pmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{pmatrix} = \begin{pmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{G}\end{pmatrix}\sigma_a^2 = \mathbf{A} + \begin{pmatrix}0 & 0\\ 0 & \mathbf{G} - \mathbf{A}_{22}\end{pmatrix}\sigma_a^2 \tag{11.18}$$

where $\mathbf{A}_{22}$ is the relationship matrix for only the genotyped animals.
It has already been shown in Section 11.5.2 that $\mathbf{a}_2 = \mathbf{Zg}$ and $\text{var}(\mathbf{a}_2) = \mathbf{G}\sigma_a^2$.

Based on selection index theory, $\mathbf{a}_1$ can be predicted from the genotyped animals (Legarra *et al.*, 2009) as:

$$\mathbf{a}_1 = \mathbf{A}_{12}\mathbf{A}^{-1}{}_{22}\mathbf{Zg} + \omega$$

where ω is the residual term, such that:

$$\text{var}(\mathbf{a}_1) = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\,\mathbf{A}_{22}^{-1}\,\mathbf{A}_{21} + \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\,\mathbf{A}_{21}$$

and this reduces to:

$$\text{var}(\mathbf{a}_1) = \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\,\mathbf{A}_{22}^{-1}\,\mathbf{A}_{21}$$

Finally, $\text{cov}(\mathbf{a}_1, \mathbf{a}_2) = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}$.

Putting all terms together into a matrix **H**, a covariance matrix of breeding values including genomics information (Legarra *et al.*, 2009; Christensen and Lund, 2010) is:

Downloaded From Vet-eBooks.com

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{pmatrix}$$

The matrix **H** could be regarded as a matrix that combines pedigree and genomic relationships.

The single-step methodology involves the use of matrix **H**, and Aguilar *et al*. (2010) and Christensen and Lund (2010) found the inverse of **H** has the following simple form:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}$$

where $\mathbf{A}_{22}^{-1}$ is inverse of the relationship matrix for genotyped animals.
This implies that by replacing $\mathbf{A}^{-1}$ with $\mathbf{H}^{-1}$ in the usual MME, direct prediction of EBVs and genomic evaluations can be obtained for ungenotyped and genotyped animals. Therefore, the MME for the single-step procedure (Eqn 11.17) are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{H}^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{11.19}$$

where:

$$\alpha = \sigma_e^2/\sigma_a^2$$

The main advantage of the single-step approach is that existing software for genetic predictions can easily be modified to implement this method. However, the computation of $\mathbf{H}^{-1}$ requires efficient computation of $\mathbf{G}^{-1}$. Thus this could be a major limitation, with large numbers of animals genotyped, since there are no simple rules for computing the inverse of **G**. Another complication is that **G** must be on exactly the same scale (e.g. scaled to the same base animals) as **A**, otherwise animals with genotypes will have biased GEBV.

Example 11.6

The data in Example 11.1 is analysed using Eqn 11.17 assuming the same genetic parameters, but the data is modified as follows. The first five animals (13 to 17) are treated as ungenotyped animals with records, the next five animals (18 to 22) are regarded as genotyped animals with records, while the remaining four animals (23 to 26) are regarded as genotyped animals with no records. A weighted analysis was carried out using the EDCs.

Therefore, the $\mathbf{A}_{22}$ matrix for the nine genotyped animals was extracted from the last nine rows of **A** given in Example 11.3.

The **G** matrix is computed as:

$$\frac{\mathbf{Z}\mathbf{Z}'}{2\sum p_j(1-p_j)}$$

for the nine genotyped animals. However, due to the small size of the data, the **G** used in the analysis was computed as $\mathbf{G} = 0.95\mathbf{G} + 0.05\mathbf{A}$ (Misztal *et al*., 2010), to enable inversion of the matrix (see Section 11.5.2). The matrix **G** then is:

Downloaded From Vet-eBooks.com

$$\mathbf{G} = \begin{bmatrix} 0.762 & 0.209 & 0.093 & 0.096 & -0.137 & 0.149 & -0.330 & 0.091 & -0.176 \\ 0.209 & 0.801 & 0.394 & -0.690 & 0.152 & 0.170 & -0.307 & 0.380 & 0.114 \\ 0.093 & 0.394 & 0.839 & -0.537 & -0.232 & 0.592 & -0.154 & -0.004 & -0.270 \\ 0.096 & -0.690 & -0.537 & 2.217 & -0.497 & -0.211 & 0.115 & -0.537 & -0.268 \\ -0.137 & 0.152 & -0.232 & -0.497 & 1.184 & -0.445 & 0.686 & 0.840 & 0.572 \\ 0.149 & 0.170 & 0.592 & -0.211 & -0.445 & 0.684 & -0.368 & -0.216 & -0.483 \\ -0.330 & -0.307 & -0.154 & 0.115 & 0.686 & -0.368 & 1.067 & 0.378 & 0.380 \\ 0.091 & 0.380 & -0.004 & -0.537 & 0.840 & -0.216 & 0.378 & 0.836 & 0.264 \\ -0.176 & 0.114 & -0.270 & -0.268 & 0.572 & -0.483 & 0.380 & 0.264 & 1.107 \end{bmatrix}$$

The $\mathbf{H}^{-1}$ for this example was constructed from the inverses of **A** in Example 11.1, of **G** and $\mathbf{A}_{22}$ shown above. The matrices in Eqn 11.19 have all been defined and solving these equations with $\alpha = 245/35.241 = 6.952$ gives the following solutions:

Mean effects	
	6.895
EBVs for animals with records	
13	3.114
14	1.697
15	4.200
16	3.842
17	2.861
GEBV for genotyped animals	
18	1.477
19	1.410
20	0.572
21	0.691
22	1.526
23	0.036
24	0.564
25	1.765
26	0.527

It is not possible to compare these results with the other models considered so far in this chapter as the data structure was modified.

11.8 Bayesian Methods for Computing SNP Effects

The assumption of equal variance explained by all loci in the SNP-BLUP or GBLUP model has the advantage that only one variance has to be estimated. However, this may be unrealistic across all traits, which may have different genetic architecture. Also, one of the problems with GBLUP is that it does not allow for moderate to large QTL effects; if these are actually present they will be severly reduced. The other problem is that with GBLUP, SNP effects cannot be zero, they always have (often very small) effects. Meuwissen *et al.* (2001) presented a Bayesian method that assumes *t* distributions at the level of the SNP effect, modelled using different genetic variances for each SNP (the so-called BayesA method) and another method in which some SNPs are assumed to have effects following a *t*-distribution, and others have zero effects

Downloaded From Vet-eBooks.com

(BayesB). Other variations of the Bayesian methods such as BayesC and BayesCπ (where some SNPs are assumed to have zero effects, and others are assumed to follow a normal distribution) have been published by Habier *et al.* (2011). This section presents some of these methods.

11.8.1 BayesA

Instead of the assumption of a normal distribution for SNP effects as in the SNP-BLUP model, another possible assumption is that the distribution follows a Student's *t*-distribution. This allows for a higher probability of moderate to large SNP effects than a normal distribution. However, the *t*-distribution is not easy to incorporate into prediction of marker effects, so a mathematically tractable way of achieving this is to assume that each SNP effect comes from a normal distribution but σ^2_g can be varied among the SNPs. Thus if σ^2_g is large then $\hat{g}$ will be large and if σ^2_g is small, then $\hat{g}$ will likely be small as it will regress towards zero (Hayes and Daetwyler, 2013). This leads to modelling the data at two levels: first at the level of the data that is similar to SNP-BLUP to estimate the SNP effects and second at the variances of the chromosome segments or SNPs, which are assumed to be different at every segment or locus. The procedure uses a Gibbs sampling approach, which involves sampling from the posterior distributions conditioned on other effects. If the reader is not familiar with Gibbs sampling, they may want to read Chapter 16, where application of the Gibbs sampling for the estimation of genetic parameters is discussed.

Thus given the linear model in Eqn 11.6, the conditional distribution that generates the data, **y**, is:

$$\mathbf{y} \mid \mathbf{b}, \mathbf{g}, \sigma^2_e \sim \mathrm{N}(\mathbf{Xb} + \mathbf{Zg} + \mathbf{R}\sigma^2_e)$$

Prior distributions

Specification of the Bayesian model involves defining the prior distributions. Usually, an improper or 'flat' prior distribution is assigned to **b**. Thus $P(\mathbf{b}) \sim$ constant.

The overall mean effect (**b**) is then sampled from the following conditional distribution as:

$$\mathbf{X'Xb} \mid \mathbf{g}, \sigma^2_{gi}, \sigma^2_e, \mathbf{y} \sim \mathrm{N}(\mathbf{X'}(\mathbf{y} - \mathbf{Zg}), \mathbf{X'X}\sigma^2_e)$$

Therefore:

$$\mathbf{b} \mid \mathbf{g}, \sigma^2_{gi}, \sigma^2_e, \mathbf{y} \sim \mathrm{N}(\hat{\mathbf{b}}, (\mathbf{X'X})^{-1}\sigma^2_e) \tag{11.20}$$

where $\hat{\mathbf{b}} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{y} - \mathbf{Zg})$

A scaled inverted chi distribution, $\chi^{-2}(\nu, S)$ is usually used as priors for the variance components, with ν being the degrees of freedom and S the scaled parameter (Wang *et al.*, 1993). Thus for the residual variance, prior uniform distribution ($\chi^{-2}(-2, 0)$) or flat prior can be assumed. Sampling is then from the following conditional posterior distribution:

$$\sigma^2_e \mid e_i \sim \chi^{-2}(n-2, e'_i e_i) \tag{11.21}$$

Downloaded From Vet-eBooks.com

where $e_i = (y_i - \mathbf{x}_i\mathbf{b} - \mathbf{z}_i\mathbf{g})$; $i = 1, n$ with n equal to the number of records or animals.

Similarly, σ^2_{gi} is sampled from the following conditional posterior distribution:

$$\sigma^2_{gi}|\mathbf{g}_i \sim \chi^{-2}(\upsilon + k_i, S + \mathbf{g}'_i\mathbf{g}_i) \tag{11.22}$$

with $\upsilon = 4.012$ and S derived as:

$$\frac{\tilde{\sigma}^2(\upsilon - 2)}{\upsilon}$$

where $\tilde{\sigma}^2$ is the a prior value of σ^2_{gi} and k_i equals 1 for the ith SNP.

Other researchers (Xu, 2003; Ter Braak *et al*., 2005) have published similar approaches with different priors for estimating σ^2_{gi}.

Finally, $\hat{\mathbf{g}}_i$ for the ith SNP is sampled from the following distribution as:

$$\mathbf{g}_i \mid \mathbf{b}, \mathbf{g}_j, \sigma^2_{gi}, \sigma^2_e, \mathbf{y} \sim N(\hat{\mathbf{g}}_i, (\mathbf{z}'_i\mathbf{z}_i + \alpha)^{-1}\sigma^2_e);\ i \neq j \tag{11.23}$$

with:

$$\hat{\mathbf{g}}_i = (\mathbf{z}'_i\mathbf{z}_i + \alpha)^{-1}\mathbf{z}'_i(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{z}_j\mathbf{g}_j) \text{ and } \alpha = \sigma^2_e/\sigma^2_{gi}$$

The Gibbs sampling procedure then consists of setting initial values for $\mathbf{b}$, $\mathbf{g}$, σ^2_e and σ^2_g, and iteratively sampling successively from Eqns 11.20 to 11.23, using updated values of the parameters from the i round in the $i + 1$ round. Assuming that p rounds of iteration were performed, then p is called the length of the chain. The first j samples are usually discarded as the burn-in period. This is to ensure that samples saved are not influenced by the priors but are drawn from the posterior distribution. Posterior means are then computed from the saved samples.

Example 11.7

Using the data in Example 11.1, the application of BayesA is illustrated using residual updating (Legarra and Misztal, 2008). The data for the reference animals is analysed by fitting the model in Eqn 11.6. Thus n, the number of records, is 8 and a flat prior has been assumed for $\mathbf{b}$. It is also assumed that $\upsilon = 4.012$ and S is derived as:

$$\frac{\tilde{\sigma}^2(\upsilon - 2)}{\upsilon} = 0.352$$

where $\tilde{\sigma}^2 = 0.702$. Note that the matrix of genotypes $\mathbf{Z}$ used in the computation below has not been centralized and there $\mathbf{Z}$ equals $\mathbf{M}$ in Section 11.2.

The starting value for $\hat{\mathbf{b}}$ was computed as $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = 79.1/8 = 9.888$ and those for $\hat{\mathbf{g}}$ and σ^2_{gi} were 0.05 and 0.702, respectively, for all SNPs. The starting value for σ^2_a was set as 2.484, thus $\sigma^2_{gi} = \sigma^2_a/2\Sigma p_j(1 - p_j) = 0.702$. The starting values for DGV for animals in the reference population were computed as $\mathbf{a} = \mathbf{Z}\hat{\mathbf{g}}$. Thus:

$\mathbf{a}'$ = (0.45 0.30 0.55 0.45 0.45 0.50 0.40 0.35)

Initially, a vector of residuals $\hat{\mathbf{e}}$ was computed as $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{g}}$. Thus:

Downloaded From Vet-eBooks.com

$$\begin{pmatrix}\hat{e}_1\\ \hat{e}_2\\ \hat{e}_3\\ \hat{e}_4\\ \hat{e}_5\\ \hat{e}_6\\ \hat{e}_7\\ \hat{e}_8\end{pmatrix} = \begin{pmatrix}9.0\\ 13.4\\ 12.7\\ 15.4\\ 5.9\\ 7.7\\ 10.2\\ 4.8\end{pmatrix} - \begin{pmatrix}9.888\\ 9.888\\ 9.888\\ 9.888\\ 9.888\\ 9.888\\ 9.888\\ 9.888\end{pmatrix} - \begin{pmatrix}0.45\\ 0.30\\ 0.55\\ 0.45\\ 0.45\\ 0.50\\ 0.40\\ 0.35\end{pmatrix} = \begin{pmatrix}-1.388\\ 3.213\\ 2.263\\ 5.063\\ -4.438\\ -2.688\\ -0.088\\ -5.437\end{pmatrix}$$

From the above, $\hat{\mathbf{e}}'\hat{\mathbf{e}} = 99.345$, and thus given the value of 8.131 sampled from the inverted χ^2 distribution with $n - 2$ degrees of freedom, $\sigma_e^{2[1]} = 99.345/8.131 = 12.218$, using Eqn 11.21. The superscript in brackets denotes the iteration number.

Then sample $\mathbf{b}^{[1]}$ using Eqn 11.20, with $\hat{\mathbf{b}}$ calculated as $(\mathbf{x}_j'\mathbf{x})^{-1}\mathbf{1}'\hat{\mathbf{e}} = 9.456$ after initially updating $\hat{\mathbf{e}}$, the vector of residuals to include information on $\mathbf{b}$ as:

$$\hat{\mathbf{e}}_i = \hat{\mathbf{e}}_i + \mathbf{X}\hat{\mathbf{b}} \text{ with } i = 1, n$$

Assuming the random number generated from a normal distribution is 0.873 and $(\mathbf{x}_j'\mathbf{x})^{-1}\sigma_e^2 = 12.218/8 = 1.527$, then $\mathbf{b}_1^{[1]} = (9.456 + 0.873\sqrt{(1.527)} = 10.535$.

After sampling for $\mathbf{b}$, the $\hat{\mathbf{e}}$ is updated to exclude the information on $\mathbf{b}$ as:

$$\hat{\mathbf{e}}_i = \hat{\mathbf{e}}_i - \mathbf{X}\hat{\mathbf{b}} \text{ with } i = 1, n$$

Using Eqn 11.22, σ_{gi}^2 for the ith SNP effect is sampled from the inverted χ^2 distribution with degrees of freedom 5.012 and $S = 0.352$ computed earlier. For the first SNP, $\hat{\mathbf{g}}_1^2 = 0.003$, thus given the value of 11.422 sampled from the inverted χ^2 distribution $\sigma_{g1}^{2[1]} = (S + \hat{\mathbf{g}}^2)/11.422 = 0.031$. The variance estimates for other SNPs in the first iteration are shown in Table 11.5.

Finally, estimates of $\hat{\mathbf{g}}$ are sampled from the normal distribution using Eqn 11.23. First update the vector of residuals to include information on the jth SNP. Thus for the jth SNP effect:

Table 11.5. SNP solutions and variances from BayesA and BayesB.

	BayesA				BayesB			
	First iteration		Posterior means		First iteration		Posterior means	
SNP	Effects	Var	Effects	Var	Effects	Var	Effects	Var
1	0.289	0.031	0.018	0.170	2.187	1.105	0.038	0.316
2	0.279	0.049	–0.064	0.179	–1.565	0.516	–0.107	0.319
3	–0.010	0.070	0.058	0.179	–0.156	0.124	0.067	0.293
4	0.023	0.097	–0.023	0.176	–0.309	0.118	–0.034	0.300
5	0.045	0.052	0.022	0.167	0.413	0.363	0.047	0.328
6	–0.321	0.050	0.025	0.171	–0.521	0.161	0.031	0.283
7	0.411	0.256	–0.006	0.186	0.000	0.000	0.009	0.335
8	0.408	0.056	–0.008	0.168	–0.010	0.431	0.008	0.261
9	0.115	0.034	–0.003	0.162	0.000	0.000	–0.006	0.294
10	–0.578	0.152	–0.008	0.165	0.000	0.000	–0.017	0.286

Var, SNP variances.

Downloaded From Vet-eBooks.com

$\hat{\mathbf{e}}_i = \hat{\mathbf{e}}_i + \mathbf{z}_{ij}\,\hat{\mathbf{g}}_j$ with $i = 1, n$

Thus for the first SNP effect, $\hat{\mathbf{g}}_1 = (\mathbf{z}'_{i1}\mathbf{z}_{i1} + \alpha)^{-1}\mathbf{z}'_{i1}\hat{\mathbf{e}}_i = (7 + 393.201)^{-1}(-2.775) = -0.007$. Assuming the random number generated from a normal distribution is 1.692 then $\mathbf{g}_j^{[1]}$ can be sampled as $\mathbf{g}_1^{[1]} = -0.007 + 1.692\sqrt{(12.218/400.201)} = 0.289$. After computing $\mathbf{g}_1^{[1]}$, the residual vector is updated as ($\hat{\mathbf{e}}_i = \hat{\mathbf{e}}_i - \mathbf{z}_{i1}\mathbf{g}_1^{[1]}$, $i = 1, n$) before computing the next SNP effect. The estimates of $\mathbf{g}_2^{[1]}$ to $\mathbf{g}_8^{[1]}$ are given in Table 11.5. The next cycle of sampling then begins again with sampling residual variance without setting up of the vector of residuals.

For this example, the Gibbs sampling chain was ran 10,000 times, with the first 3000 considered as burn-in period. The posterior means computed from the remaining 7000 samples for $\hat{\mathbf{b}}$ and σ_e^2 were 9.890 kg and 33.119 kg^2, respectively. The estimates for $\hat{\mathbf{g}}$ and σ_{gi}^2 are given in Table 11.4.

The DGV of animals in the validation set can then be predicted using the solutions for the SNP effects in Table 11.5 as $\mathbf{Z}_2\hat{\mathbf{g}}$, where $\mathbf{Z}_2$ is a matrix of genotypes for the validation of test animals given in Example 11.2.

11.8.2 BayesB

The basic assumption in BayesA is that there is genetic variance at every loci or chromosome segment. It is possible that some SNPs will have zero effects as they are in genomic regions with no QTL. The prior density of BayesA does not account for such SNPs with zero effects as BayesA density peak at $\sigma_{gi}^2 = 0$; in fact its probability of $\sigma_{gi}^2 = 0$ is infinitesimal (Meuwissen *et al.*, 2001). It is possible that genetic variance may be observed in relatively few marker loci containing QTL. Meuwissen *et al.* (2001) introduced BayesB to address this situation. Thus the prior distribution of BayesB is a mixture distribution with some SNPs with zero effects and the rest with a *t*-distribution (Hayes and Daetwyler, 2013). BayesB, therefore, uses a prior that has a high density, π, at $\sigma_{gi}^2 = 0$ and has an inverted chi-squared distribution for $\sigma_{gi}^2 > 0$. Thus the prior distribution for BayesB is:

$$\begin{aligned}&\sigma_{gi}^2 = 0 \text{ with probability } \pi\\ &\sigma_{gi}^2 \sim \chi^{-2}(\upsilon, S) \text{ with probability } (1 - \pi)\end{aligned} \tag{11.24}$$

where S is the scaling parameter, υ the degrees of freedom and π is assumed known. They set S to be to 0.0429 and computed it as in Eqn 11.22 while υ was set to 4.234.

While the Gibbs sampling algorithm used for BayesA can also be used for BayesB, it will not, however, move through the entire sampling space, as the sampling of $\sigma_{gi}^2 = 0$ is not possible if $(\mathbf{g}_i'\mathbf{g}_i)$ is greater than zero. Also, if $\sigma_{gi}^2 = 0$, the sampling of $\mathbf{g}_i$ has an infinitesimal probability. This problem is overcome by sampling σ_{gi}^2 and $\mathbf{g}_i$ simultaneously from the distribution:

$$p(\sigma_{gi}^2, \mathbf{g}_i|\mathbf{y}^*) = p(\sigma_{gi}^2|\mathbf{y}^*) \times p(\mathbf{g}_i|\sigma_{gi}^2, \mathbf{y}^*) \tag{11.25}$$

where $\mathbf{y}^*$ is the data vector $\mathbf{y}$ corrected for the mean and all genetic effects apart from $\mathbf{g}_i$. The first term in Eqn 11.25 implies sampling σ_{gi}^2 without conditioning on $\mathbf{g}_i$ and then sampling from the second term of Eqn 11.25 for $\mathbf{g}_i$ conditional on σ_{gi}^2 and $\mathbf{y}^*$ as in BayesA. The distribution $p(\sigma_{gi}^2|\mathbf{y}^*)$ cannot be expressed in the form of a known distribution, therefore Meuwissen *et al.* (2001) used the Metropolis–Hastings (MH)

Downloaded From Vet-eBooks.com

algorithm to sample from $p(\sigma^2_{gi}|\mathbf{y}^*)$ using the prior distribution, $p(\sigma^2_{gi})$, as the driver distribution to suggest updates for the MH chain as follows:

1. Sample $\sigma^2_{gi(\text{new})}$ from the prior distribution $p(\sigma^2_{gi})$.
2. Replace the current σ^2_{gi} by $\sigma^2_{gi(\text{new})}$ with a probability of k:

$$k = \text{minimize}\{p(\mathbf{y}^*|\sigma^2_{gi(\text{new})})/p(\mathbf{y}^*|\sigma^2_{gi}); 1\}$$

and then go to step 1:

where $p(\mathbf{y}^*|\sigma^2_{gi})$ is the likelihood of the data given σ^2_{gi}. The likelihood can be calculated as:

$$L(\mathbf{y}^*|\sigma^2_{gi}) = \frac{1}{2\pi^{1/2n}\sqrt{|\mathbf{V}|}} e^{-1/2(\mathbf{y}^{*\prime}\mathbf{V}^{-1}\mathbf{y})} \tag{11.26}$$

where $\mathbf{V} = \mathbf{z}_i(\mathbf{I}\sigma^2_{gi})\mathbf{z}_i' + \mathbf{I}\sigma^2_e$ and $|\mathbf{V}|$ is the determinant of $\mathbf{V}$. Note that if σ^2_{gi} is zero, as will happen in the course of the MH sampling, then $\mathbf{V} = \mathbf{I}\sigma^2_e$.

The computation of the required likelihood is easier to implement in a log-likelihood form. Fernando (2010) presented the following algorithm for the log-likelihood:

$$\text{logLH} = -0.5(\log(\mathbf{V})) + (((\mathbf{z}_i'\mathbf{y}^*)'\mathbf{V}^{-1})\,\mathbf{z}_i'\mathbf{y}^*) \tag{11.27}$$

with:

$$\mathbf{V} = (\mathbf{z}_i'\mathbf{z}_i\mathbf{I}\sigma^2_{gi},\mathbf{z}_i'\mathbf{z}_i) + \mathbf{z}_i'\mathbf{z}_i{}^*\sigma^2_e \text{ or } \mathbf{V} = \mathbf{z}_i'\mathbf{z}_i{}^*\sigma^2_e \text{ when } \sigma^2_{gi} \text{ is zero}$$

In practice, a required number of MH cycles are implemented per cycle of Gibbs sampling. The implementation of each MH cycle involves:

1. Using Eqn 11.26 or 11.27 compute an initial likelihood (LH1) using the current σ^2_{gi}. Note that the current σ^2_{gi} could be zero and LH1 is also computed but with $\mathbf{V}$ appropriately defined.
2. Then commence the MH cycle, by drawing r from a uniform distribution. Set $\sigma^2_{gi(\text{new})}$ to be zero. If $r < (1 - \pi)$, sample a $\sigma^2_{gi(\text{new})}$ from the driver distribution using Eqn 11.25. Compute likelihood (LH2) using $\sigma^2_{gi(\text{new})}$ and calculate k as k = minimize (LH2/LH1; 1). Note that if log-likelihood Eqn 11.27 is used, then $k = \exp(\text{LH2} - \text{LH1})$. The value of k is compared with a number s drawn from a uniform distribution. If s is less than k, then accept $\sigma^2_{gi(\text{new})}$ and then set LH1 = LH2. Go to step 1 and begin another MH cycle until required MH cycles are complete.

After the required number of MH cycles, if $\sigma^2_{gi(\text{new})}$ is > 0 then $\mathbf{g}_i$ is sampled as in BayesA, otherwise $\mathbf{g}_i = 0$. Similarly, the sampling of $\mathbf{b}$ and σ^2_e is implemented as described in BayesA.

Example 11.8

The application of BayesB is illustrated using the data in Example 11.1 with residual updating. The data for the reference animals is analysed with the model in Eqn 11.6. The initial parameters are the same as outlined for BayesA in Example 11.7 and the starting value of π was set at 0.30.

The starting values for $\hat{\mathbf{b}}$, $\hat{\mathbf{g}}$, σ^2_{gi} and $\hat{\mathbf{a}}$ were the same as for BayesA. The sampling procedure for parameters is the same as for BayesA apart from sampling for σ^2_{gi}.

Initially, the vector of residuals, $\hat{\mathbf{e}}$, is set up and this has been given in Example 11.7. Therefore, in the first iteration $\sigma^{2[1]}_e = 99.345/8.131 = 12.218$.

Downloaded From Vet-eBooks.com

Similarly, $\mathbf{b}_1^{[1]} = (9.456 + 0.873\sqrt{(1.527)} = 10.535$, as in Example 11.7.

Using the steps outlined for the MH cycle for BayesB, σ^2_{gi} for the *i*th SNP effect is then sampled, which could result in either $\sigma^2_{gi} = 0$ or $\sigma^2_{gi} > 0$. In this example, 20 MH samples were evaluated per each round of Gibbs sampling, and for the first SNP, the estimate of $\sigma^2_{g1} = 1.105$. Therefore, $\hat{\mathbf{g}}_1$ was sampled from the normal distribution using Eqn 11.23 as described in Example 11.7 but with $\alpha_1 = 12.218/1.105 = 11.057$. In this example, σ^2_{gi} and $\hat{\mathbf{g}}_i$ for SNP_i, with $i = 7$, 9 and 10 were zero in the first round of iteration. The solutions for σ^2_{gi} and $\hat{\mathbf{g}}_i$ for the first round of iteration are presented in Table 11.5.

The Gibbs sampling was run for 10,000 cycles, with the first 3000 regarded as the burn-in period. The posterior means computed from the remaining 7000 samples for $\hat{\mathbf{b}}$ and σ^2_e were 9.792 kg and 34.930 kg^2, respectively. The estimates for $\hat{\mathbf{g}}$ and σ^2_{gi} are given in Table 11.5. The DGV of animals in the validation set can then be predicted using the solutions for the SNP effects in Table 11.2 as $\mathbf{Z}_2\hat{\mathbf{g}}$, where $\mathbf{Z}_2$ is defined as in Example 11.7.

11.8.3 BayesC

Habier *et al.* (2011) indicated the estimation of individual SNP variances in BayesA and BayesB has only one additional degree of freedom compared with its prior, and so the shrinkage of SNP effects is largely dependent on the scale parameter, *S*. To overcome this limitation, they proposed BayesC, which involves estimating a single variance that is common to all SNPs, thereby reducing the influence of the scale parameter. Similar to BayesB, BayesC allows for some SNPs to have zero effects with probability π while the remaining SNPs have non-zero effect with probability $(1 - \pi)$. Habier *et al.* (2011) indicated that since the priors of all SNP effects have a common variance, the effect of an SNP fitted with probability $(1 - \pi)$ comes from a mixture of multivariate Student's *t*-distributions.

In BayesC, it is assumed that π is known and the decision to include SNP_i depends on the full conditional posterior of an indicator variable δ_i. This indicator variable equals 1 if SNP_i is fitted, otherwise it is zero. Thus the decision to include the *i*th SNP involves computing the probability *k* of $\delta_i = 1$ as $k = 1/\{1 + (p(\mathbf{y}^* \mid \delta_i = 0, \ominus)/p(\mathbf{y}^* \mid \delta_i = 1, \sigma^2_g, \ominus))\}$, where $(p(\mathbf{y}^* \mid \delta_i = 1, \ominus))$ denotes the likelihood of the data given that SNP_i is fitted with common variance σ^2_g, $\ominus$ refers to accepted values for all other parameters, $(p(\mathbf{y}^* \mid \delta_i = 0, \ominus))$ denotes the likelihood of the data model without the *i*th SNP and where $\mathbf{y}^*$ is the data vector $\mathbf{y}$ corrected for the mean and all genetic effects apart from $\mathbf{g}_i$.

The computation of the required likelihood is easier to implement in a log-likelihood form. Fernando (2010) presented such an algorithm based on the log-likelihood.

Given current estimates of σ^2_g, and σ^2_e, logLH1 with $\delta_i = 1$ is computed as:

$$\text{logLH1} = -0.5(\log(\mathbf{V})) + (\mathbf{z}_i'\mathbf{y}^*)'\mathbf{V}^{-1}\mathbf{z}_i'\mathbf{y}^* + \log(1 - \pi) \text{ with}$$
$$\mathbf{V} = (\mathbf{z}_i'\mathbf{z}_i\mathbf{I}\sigma^2_g, \mathbf{z}_i'\mathbf{z}_i) + \mathbf{z}_i'\mathbf{z}_i{}^*\sigma^2_e$$

Similarly, the log-likelihood when $\delta_i = 0$ is computed as $\text{logLH0} = -0.5(\log(\mathbf{V})) + (\mathbf{z}_i'\mathbf{y}^*)'\mathbf{V}^{-1}\mathbf{z}_i'\mathbf{y}^* + \log(\pi)$ but with $\mathbf{V} = \mathbf{z}_i'\mathbf{z}_i{}^*\sigma^2_e$.
Then compute probability *k* of $\delta_i = 1$ as $k = 1/(1 + \exp(\text{logLH0} - \text{logLH1}))$.

If *k* is greater than *r*, where *r* is a random drawn from a uniform distribution, then SNP_i is fitted and $\mathbf{g}_i^{[j]}$ is sampled from the normal distribution using Eqn 11.23, otherwise $\mathbf{g}_i^{[j]} = 0$.

Downloaded From Vet-eBooks.com

After sampling the vector $\mathbf{g}$, σ_g^2 is sampled from the following conditional posterior distribution as:

$$\sigma_g^2 \mid \mathbf{g}_i \sim \chi^{-2}(\upsilon + k^{[j]}, S + \mathbf{g}_i' \mathbf{g}_i) \tag{11.28}$$

with terms defined as in Eqn 11.22 but with degrees of freedom equal to $\upsilon + k^{[j]}$, where $k^{[j]}$ is the number of SNPs with non-zero effects fitted in the jth iteration.

Example 11.9

The data in Example 11.1 is used to illustrate BayesC by applying the model in Eqn 11.6. The assumptions and the starting values for $\hat{\mathbf{b}}$, $\hat{\mathbf{g}}$ and $\mathbf{a}$ were the same as outlined for BayesA in Example 11.7. The starting value of π was assumed at 0.30 while the starting value of σ_g^2 was set at 0.702.

The sampling procedure for σ_e^2 and $\mathbf{b}$ were as outlined in BayesA and therefore with the same solutions in the first iteration. Then for the ith SNP, the probability of $\hat{\mathbf{g}}_i$ having a zero effect or otherwise was computed as described earlier in this section. In the first iteration, the first SNP has a non-zero effect; therefore, $\hat{\mathbf{g}}_1 = (\mathbf{z}_{i1}'\mathbf{z}_{i1} + \alpha)^{-1}$ $\mathbf{z}_{i1}'\hat{\mathbf{e}}_i = (7 + 17.045)^{-1}$ $(-2.775) = -0.115$, with $\alpha = 12.218/0.702$. Assuming the random number generated from a normal distribution is 0.748, $\mathbf{g}_j^{[1]}$ was sampled using Eqn 11.23 as $\mathbf{g}_1^{[1]} = -0.115 + 0.748\sqrt{(12.218/24.045)} = 0.418$. In the first round of iteration, two SNPs (5 and 10) had zero effects. The solutions for $\hat{\mathbf{g}}_i$ in the first iteration are presented in Table 11.6.

The sampling of common variance was done using Eqn 11.28. For this example, eight SNPs had non-zero effects in the first iteration; therefore, σ_g^2 in the first iteration was sampled from the inverted χ^2 distribution with degrees of freedom now equal to $8 + 4.012 = 12.012$, $S = 0.352$ and $\sum_i \hat{\mathbf{g}}_i^2 = 1.435$. Thus given the value of 16.294 sampled from the inverted χ^2 distribution, then in the first iteration $\sigma_g^{2\,[1]} = (S + \sum \hat{\mathbf{g}}_i^2) / 16.294 = 0.110$.

The Gibbs sampling was run for 10,000 cycles, with the first 3000 regarded as the burn-in period. The posterior means computed from the remaining 7000 samples for $\hat{\mathbf{b}}$, σ_e^2 and σ_g^2 were 9.828 kg, 32.377 kg^2 and 0.184 kg^2, respectively. The estimates for $\hat{\mathbf{g}}$ are given in Table 11.6.

Table 11.6. Solutions for SNP effects from BayesC and BayesCπ.

	BayesC		BayesCπ
SNP	First iteration	Posterior means	Posterior means
1	0.416	0.015	0.010
2	–0.360	–0.045	–0.029
3	–0.590	0.044	0.028
4	0.465	–0.014	–0.018
5	0.000	0.014	0.013
6	0.360	0.025	0.010
7	–0.586	–0.002	0.004
8	–0.307	0.009	0.003
9	–0.041	–0.013	–0.011
10	0.000	–0.002	–0.006

Downloaded From Vet-eBooks.com

11.8.4 BayesCπ

In BayesC there is the implicit assumption that the probability, $\pi > 0$, i.e. a SNP has zero effect, is regarded as known. Habier *et al.* (2011) argued that the shrinkage of SNP effects is affected by π and should be estimated from the data and proposed BayesCπ, which incorporates this estimation step. Thus compared to BayesC, the additional feature of BayesCπ is estimating π from the data. The sampling procedure for parameters in BayesCπ is therefore the same as BayesC apart from the additional step of sampling for π. Thus only the procedure for sampling π is described.

The parameter π is sampled from a beta distribution, with shape parameters $(m - k^{[j]} + 1)$ and $(k^{[j]} + 1)$, with m equal to the total number of SNPs in the analysis and $k^{[j]}$ is the number of SNPs with non-zero effects fitted in the jth iteration.

Example 11.10

The application of BayesCπ is illustrated using the data in Example 11.1. The reference animals are analysed by applying the model in Eqn 11.6 using residual updating. The initial parameters are the same as outlined for BayesA in Example 11.7. The starting values of π and σ_g^2 were set at 0.30 and 0.702, respectively.

The sampling procedure for σ_e^2 and $\mathbf{b}$ were as outlined in BayesA and therefore with the same solutions in the first iteration. Then for the ith SNP, the probability of $\hat{\mathbf{g}}_i$ having a zero effect or otherwise was computed as described earlier in this section. In the first iteration, the first SNP has a non-zero effect; therefore, $\hat{\mathbf{g}}_1 = (\mathbf{z}'_{i1}\mathbf{z}_{i1} + \alpha)^{-1}$ $\mathbf{z}'_{i1}\hat{\mathbf{e}}_i = (7 + 17.045)^{-1}$ $(-2.775) = -0.115$, with $\alpha = 12.218/0.702$. Assuming the random number generated from a normal distribution is 0.748, $\mathbf{g}_j^{[1]}$ was sampled using Eqn 11.23 as $\mathbf{g}_1^{[1]} = -0.115 + 0.748\sqrt{(12.218/24.045)} = 0.418$. In the first round of iteration, two SNPs (5 and 10) had zero effects and the solutions for $\mathbf{g}$ were the same as obtained for BayesC (Table 11.6).

The sampling of common variance follows the same procedure for BayesC, again with the degrees of freedom equal to the number of SNPs with non-zero effects. For this example, eight SNPs had non-zero effects in the first iteration; therefore, σ_g^2 in the first iteration was sampled from the inverted χ^2 distribution with degrees of freedom now equal to 8 + 4.012 = 12.012, $S = 0.352$ and $\sum_i \hat{\mathbf{g}}_i^2 = 1.435$. Thus given the value of 16.294 sampled from the inverted χ^2 distribution, then in the first iteration $\sigma_1^{2[1]} = (S + \sum_i \hat{\mathbf{g}}_i^2)/16.294 = 0.110$.

Then $\pi^{[1]}$ was sampled from the beta distribution with shape parameters $((m - k^{[1]} + 1) = 3)$ and $((k^{[1]} + 1) = 9)$, given eight SNPs had non-zero effects. A value of 0.339 was sampled for π.

A total of 10,000 cycles was implemented for the Gibbs sampling and the first 3000 were discarded as the burn-in period. The posterior means computed from the remaining 7000 samples for $\hat{\mathbf{b}}$, σ_e^2, σ_g^2 and π were 9.898 kg, 32.343 kg^2, 0.162 kg^2 and 0.51, respectively. The estimates for $\hat{\mathbf{g}}$ were given in Table 11.6.

The estimates for $\hat{\mathbf{b}}$ and σ_e^2 were very consistent for the Bayesian models considered. Similarly, BayesC and BayesCπ gave very similar estimates of σ_g^2, which were consistent with estimates for BayesA but SNP solutions were different from the different models. The estimates of σ_{gi}^2 for BayesB were almost double those from the other models.

Downloaded From Vet-eBooks.com

11.9 Cross-validation and Genomic Reliabilities

As described in previous sections, the computation of SNP effects is usually in a reference population using animals with observations. In the case of the dairy industry, the estimation of SNP effects has been carried out using mostly bulls with high reliability as the reference population with deregressed breeding values (DRB) used as observations. Recently, some countries have started including cows in the reference populations, which require weighting the cow records appropriately. Ideally, it is necessary that the estimates of SNP effects are validated in another data set, which has not contributed any information to the reference population to assess accuracy of prediction. In practice, the cross-validation should be evaluated in differently randomly sampled validation data sets to avoid any bias.

The DGV computed for the validation data sets are compared with their DRP. An estimate of the correlation between the DGV and the DRP in the validation animals provides an estimate of the accuracy of genomic predictions, although this does not take into account the accuracy of the DRP themselves. For the purposes of illustration, the correlation between the DGVs from the SNP or GBLUP models with the DRPs for the validation animals in the data for Example 11.1 is 0.49, which gives a reliability of 0.24. The accuracies or reliabilities from the cross-validation studies are usually referred to as realized reliabilities.

Theoretical reliabilities, as calculated in traditional BLUP, can also be computed from the inverse of equations similar to those used to compute DGVs. For individuals with observations, reliabilities for the DGV can be computed (VanRaden, 2008) by first computing **B** as follows:

$$\mathbf{B} = \mathbf{G}\left(\mathbf{G} + \mathbf{R}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right)^{-1}\mathbf{G}$$

Then reliability for animal i = $\text{rel}_i = 1 - (b_{ii}^* \sigma_e^2/\sigma_a^2)$, where b_{ii} is the diagonal element of **B** for the animal. Similarly, for validation candidates with no records, **B** is:

$$\mathbf{B} = \mathbf{C}\left(\mathbf{G} + \mathbf{R}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right)^{-1}\mathbf{C}'$$

Then reliability is computed from the diagonal elements of **B** as described for the reference animals.

However, these theoretical reliability estimates tend to be too high. These can be scaled by the realized reliabilities from the cross-validation study. In addition, with a large data set, the inversion required for the computation of the reliabilities could be a source of limitation to the use of the methodology.

11.10 Understanding SNP Solutions from the Various Models

The vector **g** can be computed from the second row of the MME in Eqn 11.7. Thus:

$$\hat{\mathbf{g}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{I}\alpha)^{-1}(\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}))$$

Downloaded From Vet-eBooks.com

For the *i*th SNP, this can be expressed (Mrode *et al*., 2010) as:

$$\hat{g}_i = (z_{ji}'r^{-1}z_{ji} + \alpha)^{-1}z_{ji}'r^{-1}z_{ji}(yd_i),\ j = 1, n \text{ (the number of animals)}$$

$$\hat{g}_i = wt_i(yd_i) \tag{11.29}$$

where yd_i is the SNP deviation for the *i*th SNP, i.e. data information for that SNP corrected for all effects apart from the SNP and the SNP deviation can be defined as $yd_i = (z'_{ji}r^{-1}z_{ji})^{-1}z'_{ji}r^{-1}(y_j - z_{jk}\hat{g}_k - x_j\hat{b}),\ k \neq i$ and $wt_i = (z'_{ji}r^{-1}z_{ji} + \alpha)^{-1}z_{ji}'r^{-1}z_{ji}$. The DGV_j of animal *j* therefore is:

$$DGV_j = \sum_i z_{ji}wt_i(yd_i)$$

For illustration purposes, the SNP solution for SNP 1, $\hat{g}_1$ in Example 11.2, can be computed using Eqn 11.29 as follows:

The **Z** in Example 11.2, $(z'_{j1}z_{j1}) = 3.878$, and $(z'_{j1}z_{j1} + \alpha) = 28.476$.
The SNP deviation, $yd_1 = 0.638$; therefore, $wt_1 = 3.878/28.476 = 0.136$ and $\hat{g}_1 = 0.136\ (0.638) = 0.087$. Similar calculations indicated that for SNP 7, $yd_7 = -0.001$, $wt_7 = 0.007$, and $\hat{g}_7 = 0.00$.

In the case of Bayesian methods, there is an additional component as a result of sampling from the conditional posterior distribution of **g**, such that:

$$\hat{g}_i = wt_i(yd_i) + N(\hat{g}_i, (z_i'r^{-1}z_i + \alpha_i)^{-1}\sigma^2 e) \tag{11.30}$$

The second term on the right-hand side of Eqn 11.30 tends towards zero averaged over all samples after the burn-in period.

Equation 11.29 indicates that with the SNP-BLUP model, the SNP solutions are a function of the SNP deviations, which could be regarded as the unregressed SNP allele substitution effects and the weight. Given that α is constant for the SNP-BLUP model, the weight is therefore very dependent on the allele frequencies. Thus alleles of lower frequencies will have a lower weight on their SNP deviations. In the calculations above, the weight for SNP1 with an allele frequency of 0.312 was much higher than that for SNP 7. Mrode *et al*. (2010) obtained a correlation between the weights and allele frequencies of 0.99 from the SNP-BLUP model. However, for BayesA and BayesB, the estimation of individual variances meant that α, and therefore weights, were different for each SNP. Thus SNP deviations were differentially weighted not only on the basis of their allele frequencies but also on the basis of their genetic variance, i.e. by the amount of available information.

Downloaded From Vet-Books.com

12 Non-additive Animal Models

12.1 Introduction

The models considered in the previous chapters have dealt with only additive genetic effects. Henderson (1985) provided a statistical framework for modelling additive and non-additive genetic effects when there is no inbreeding. This chapter covers some of these models. The ability to separate non-additive genetic effects implies removal of some of the confounding that would otherwise bias the results from the analysis. Moreover, the availability of estimates of non-additive genetic effects for individuals could be used in mate selection, which would maximize the use of both additive and non-additive genetic variance. In this chapter, the prediction of dominance and epistatic effects using mixed model methodology is discussed. In practice, the application of non-additive models in genetic evaluation has been limited due to lack of genetic parameters and due to the fact that these effects tend to be highly confounded with others, such as common maternal environment.

12.2 Dominance Relationship Matrix

Dominance genetic effects result from the action of pairs of alleles at a locus on a trait. If two animals have the same set of parents or grandparents, it is possible that they possess the pair of alleles in common. The dominance relationship between two such animals represents the probability that they have the same pair of alleles in common. Thus for a group of animals, the dominance genetic relationship matrix (**D**) among them can be set up. The dominance relationship between an individual x with parents s and d and y with parents f and m in a non-inbred population can be calculated (Cockerham, 1954) as:

$$d_{xy} = 0.25(u_{sf}u_{dm} + u_{sd}u_{fm}) \tag{12.1}$$

where u_{ij} represents the additive genetic relationship between i and j. For instance, for two full-sibs with both parents unrelated to each other:

$$d = 0.25[(1)(1) + (0)(0)] = 0.25$$

with the assumption that there is no common environmental variance.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

Thus **D** can be generated from the additive genetic relationship. However, the prediction of dominance effects requires the inverse of **D**. This could be obtained by calculating **D** by Eqn 12.1 and inverting it: this is not computationally feasible with large data sets. Hoeschele and VanRaden (1991) developed a methodology for obtaining a rapid inversion of **D** and this is presented in Section 12.4. Initially, the principles involved in using $\mathbf{D}^{-1}$ from Eqn 12.1 for the prediction of dominance effects are discussed.

12.3 Animal Model with Dominance Effect

The model with dominance included is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wd} + \mathbf{e} \tag{12.2}$$

where **y** = vector of observations, **b** = vector of fixed effects, **a** = vector for random animal additive genetic effects, **d** = vector of random dominance effects and **e** = random residual error.

It is assumed that:

$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2, \quad \text{var}(\mathbf{d}) = \mathbf{D}\sigma_d^2 \quad \text{and} \quad \text{var}(\mathbf{e}) = \sigma_e^2$$

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}' + \mathbf{WDW}' + \mathbf{I}\sigma_e^2$$

The MME to be solved for the BLUP of **a** and **d** and the BLUE of **b** are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \mathbf{A}^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \tag{12.3}$$

with $\alpha_1 = \sigma_e^2/\sigma_a^2$ and $\alpha_2 = \sigma_e^2/\sigma_d^2$. However, we are interested in the total genetic merit (**g**) of the animal, which is $\mathbf{g} = a + \mathbf{d}$. The MME could be modified such that the total genetic merit is solved for directly. Since $\mathbf{g} = a + \mathbf{d}$, then:

$$\text{var}(\mathbf{g}) = \mathbf{G} = \mathbf{A}\sigma_a^2 + \mathbf{D}\sigma_d^2$$

The MME become:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \tag{12.4}$$

The individual components of **g** can be obtained as:

$$\hat{\mathbf{a}} = \sigma_a^2\mathbf{A}\mathbf{G}^{-1}\hat{g} \quad \text{and}$$

$$\hat{\mathbf{d}} = \sigma_d^2\mathbf{D}\mathbf{G}^{-1}\hat{g}$$

12.3.1 Solving for animal and dominance genetic effects separately

Example 12.1
Suppose the data below are the weaning weights for some piglets in a herd.

Pig	Sire	Dam	Sex	Weaning weight (kg)
5	1	2	Female	17.0
6	3	4	Female	20.0
7	6	5	Female	18.0
8	0	5	Female	13.5
9	3	8	Male	20.0
10	3	8	Male	15.0
11	6	8	Male	25.0
12	6	8	Male	19.5

The aim is to estimate sex effects and predict solutions for animal and dominance genetic effects, assuming that $\sigma_e^2 = 120$, $\sigma_a^2 = 90$ and $\sigma_d^2 = 80$. This has been illustrated below, solving for animal and dominance effects separately (Eqn 12.3). From the above parameters, $\alpha_1 = 1.333$ and $\alpha_2 = 1.5$.

Setting up the MME

The matrix **X** relates records to sex effects. Its transpose, considering only animals with records, is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The matrices **Z** and **W** are both identity matrices since each animal has one record. The transpose of the vector of observations $\mathbf{y}'$ = [17 20 18 13.5 20 15 25 19.5].

The other matrices in the MME, apart from $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$, can be obtained through matrix multiplication from the matrices already calculated. The inverse of the additive relationship matrix is set up using rules outlined in Section 2.4.1. Using Eqn 12.1, the dominance relationship matrix is:

$$\mathbf{D} = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.062 & 0.062 & 0.125 & 0.125 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.062 & 0.000 & 1.000 & 0.250 & 0.125 & 0.125 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.062 & 0.000 & 0.250 & 1.000 & 0.125 & 0.125 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.125 & 0.000 & 0.125 & 0.125 & 1.000 & 0.250 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.125 & 0.000 & 0.125 & 0.125 & 0.250 & 1.000 \end{bmatrix}$$

and its inverse is:

$$\mathbf{D}^{-1} = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.028 & 0.000 & -0.032 & -0.032 & -0.096 & -0.096 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.032 & 0.000 & 1.084 & -0.249 & -0.080 & -0.080 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.032 & 0.000 & -0.249 & 1.084 & -0.080 & -0.080 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.096 & 0.000 & -0.080 & -0.080 & 1.092 & -0.241 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.096 & 0.000 & -0.080 & -0.080 & -0.241 & 1.092
\end{bmatrix}$$

The matrices $\mathbf{A}^{-1}\alpha_1$ and $\mathbf{D}^{-1}\alpha_2$ are added to $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{W}'\mathbf{W}$ in the MME. The MME are of the order 26 by 26 and are too large to be presented. However, the solutions to the MME by direct inversion of the coefficient matrix are:

Effects	Solutions	
Sex		
Female	16.980	
Male	20.030	
Animal	BV[a]	DV[a]
1	–0.160	0.000
2	–0.160	0.000
3	0.059	0.000
4	0.819	0.000
5	–0.320	0.136
6	1.259	0.705
7	0.555	0.237
8	–0.998	–0.993
9	–0.350	0.000
10	–1.350	–1.333
11	1.061	1.428
12	–0.039	–0.038

[a]BV, DV, solutions for random animal and dominance effects, respectively.

The results indicate that males were heavier than females by about 3.05 kg at weaning. The breeding value for animal i, $\hat{a}_i$, from the MME can be calculated using Eqn 3.8, except that yield deviation is corrected not only for fixed effects but also for dominance effect. Thus the solution for animal 6 can be calculated as:

$$\begin{aligned}
\hat{a}_6 &= n_1((\hat{a}_3 + \hat{a}_4)/2) + n_2(y_6 - \hat{b}_1 - \hat{d}_6) + n_3(2\hat{a}_{12} - \hat{a}_8) + n_3(2\hat{a}_{11} - \hat{a}_8) + n_3(\hat{a}_7 - \hat{a}_5) \\
&= n_1(0.059 + 0.819)/2 + n_2(20 - 16.980 - 0.705) + n_3(2(-0.039) - (-0.998)) \\
&\quad + n_3(2(1.061) - (-0.998)) + n_3(2(0.555) - (-0.320)) \\
&= 1.259
\end{aligned}$$

where $n_1 = 2\alpha_1/wt$, $n_2 = 1/wt$, $n_3 = 0.5\alpha_1/wt$, with wt equal to the sum of the numerator of n_1, n_2 and $3(n_3)$.

The solution for the dominance effect of animal i from the MME is:

$$\hat{d}_i = \left[-\alpha_2 \left(\sum_j c_{ij} \hat{d}_j \right) + \left(y_i - \hat{b}_k - \hat{a}_i \right) \right] \Big/ \left(n + c_{ii} \alpha_2 \right)$$

where c_{ij} is the inverse element of **D** between animal i and j, and n is the number of records. For instance, the dominance effect of animal 6 is:

$\hat{d}_6 = (0 + (20 - 16.980 - 1.259))/(1 + 1.5) = 0.705$

The dominance effect for an individual represents interactions of pairs of genes from both parents and Mendelian sampling; it therefore gives an indication of how well the genes from two parents combine. This could be used in the selection of mates.

12.3.2 Solving for total genetic merit directly

Example 12.2

Using the same data and genetic parameters as in Example 12.1, solving directly for total genetic merit (**â** + **d**) applying Eqn 12.4 is illustrated.

Setting up the MME

The design matrices **X** and **Z** are exactly the same as in Eqn 12.3. However, in Eqn 12.4, $\mathbf{G} = \mathbf{A}\sigma_a^2 + \mathbf{D}\sigma_d^2$. The matrix **D** has been given earlier and **A** can be calculated as outlined in Section 2.2. Then $\mathbf{G}^{-1}\sigma_e^2$ is added to $\mathbf{Z'Z}$ to obtain the MME (Eqn 12.4). Solving the MME by direct inversion of the coefficient matrix gives the following solutions:

Effects	Solutions		
Sex			
Female	16.980		
Male	20.030		
Animal + dominance		*Animal + dominance*	
1	–0.160	7	0.792
2	–0.160	8	–1.991
3	0.059	9	–0.349
4	0.819	10	–2.683
5	–0.184	11	2.489
6	1.963	12	–0.078

The vector of solutions for additive genetic effects can then be calculated as $\hat{\mathbf{a}} = \sigma_a^2 \mathbf{A}\mathbf{G}^{-1}\mathbf{g}$ and as $\mathbf{d} = \sigma_d^2 \mathbf{D}\mathbf{G}^{-1}\mathbf{g}$ for dominance effects, as mentioned earlier. It should be noted that the sum of $\hat{a}_i$ and d_i for animal i in Example 12.1 equals the solution for animal i above, indicating that the two sets of results are equivalent. The advantage of using Eqn 12.4 is the reduction in the number of equations to be solved.

12.4 Method for Rapid Inversion of the Dominance Matrix

Hoeschele and VanRaden (1991) developed a method for computing directly the inverse of the dominance relationship matrix for populations that are not inbred, by including sire and dam or sire and maternal grandsire subclass effects in the model. However, only the inclusion of sire and dam subclasses is considered in this text. Dominance effects result from interaction of pairs of genes and are not inherited through individuals. Since animals receive half of their genes from the sire and half from the dam, the dominance effect of an individual could be expressed as:

$$d = f_{S,D} + \varepsilon \tag{12.5}$$

where f represents the average dominance effect of many hypothetical full-sibs produced by sire (S) and dam (D) and ε is the Mendelian sampling deviation of the individual from the S by D subclass effect. Variance of S by D subclass effects, σ_f^2, is equal to the covariance among full-sibs due to dominance, i.e. $\sigma_f^2 = 0.25\sigma_d^2$; therefore, $\text{var}(\varepsilon) = 0.75\sigma_d^2$. On the basis of Eqn 12.5, Hoeschele and VanRaden developed simple recurrence formulae for dominance effects using pairs of animals (sire and dam) and interaction between their parents.

For a particular sire and dam subclass (f_{SD}), the combination effect results from the interactions between the sire and the parents of D, interactions of the dam with the parents of S and interactions of the parents of S with the parents of D. Thus:

$$\begin{aligned} f_{SD} = {} & 0.5(f_{S,SD} + f_{S,DD} + f_{SS,D} + f_{DS,D}) \\ & - 0.25(f_{SS,SD} + f_{SS,DD} + f_{DS,SD} + f_{DS,DD}) + e \end{aligned} \tag{12.6}$$

where SS and DS denote sire and dam of sire, respectively, and SD and DD corresponding parents for the dam. Equation 12.6 can also be obtained by regressing f_{SD} on its parent subclasses effects as:

$$f_{SD} = \mathbf{b}'\mathbf{f}_{par} + e$$

where $\mathbf{f}_{par}$ is a vector of eight parent subclasses in Eqn 12.6 and $\mathbf{b}$ is a vector of corresponding partial regression coefficients with:

$$\mathbf{b}' = \text{cov}(f_{SD}, \mathbf{f}_{par})/\text{var}(\mathbf{f}_{par}) \tag{12.7}$$

and:

$$\text{var}(e) = \sigma_f^2 - \mathbf{b}'\text{var}(\mathbf{f}_{par})\mathbf{b} \tag{12.8}$$

The covariance between subclasses in Eqn 12.7, for instance between f_{SD} and f_{PM}, is:

$$\text{cov}(f_{SD}, f_{PM}) = (a_{SP}a_{DM} + a_{SM}a_{DP})\sigma_f^2 \tag{12.9}$$

with a_{ij} being the additive relationship between i and j. Thus:

$$\text{cov}(f_{SD}, f_{SS,DD}) = (a_{S,SS}a_{D,DD} + a_{S,DD}a_{D,SS})\sigma_f^2 = (0.5(0.5)) + (0(0)) = 0.25\sigma_f^2$$

and:

$$\text{cov}(f_{SD}, f_{S,SD}) = (a_{SS}a_{D,SD} + a_{S,SD}a_{D,S})\sigma_f^2 = (1(0.5)) + (0(0)) = 0.5\sigma_f^2$$

If the nine subclasses in Eqn 12.6 are identified by 1, 2, 3, 4, 5, 6, 7, 8 and 9 (i.e. f_{SD} = 1, $f_{S,SD}$ = 2, etc.), the covariances between f_{SD} and its parent subclasses ($\text{cov}(f_{SD}, \mathbf{f}_{par})/\sigma_f^2$) using Eqn 12.9 are:

$$\begin{array}{cccccccccc} & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \\ 1 & [0.5 & 0.5 & 0.5 & 0.5 & 0.25 & 0.25 & 0.25 & 0.25] & \end{array} \quad (12.10)$$

and the relationship matrix among parent subclasses ($\text{var}(\mathbf{f}_{par})/\sigma_f^2$) using Eqn 12.9 is:

$$\begin{bmatrix} 1.0 & 0.0 & 0.25 & 0.25 & 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 1.0 & 0.25 & 0.25 & 0.0 & 0.5 & 0.0 & 0.5 \\ 0.25 & 0.25 & 1.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.25 & 0.25 & 0.0 & 1.0 & 0.0 & 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (12.11)$$

From the two matrices above (Eqns 12.10 and 12.11) the regression coefficients (Eqn 12.7) are:

$$\mathbf{b}' = [0.5\ 0.5\ 0.5\ 0.5\ -0.25\ -0.25\ -0.25\ -0.25] \quad (12.12)$$

which are identical to the coefficients in Eqn 12.6. It should be noted that there is no need to add more remote ancestors of *S* and *D* as the partial regression of these are zero.

12.4.1 Inverse of the relationship matrix of subclass effects

The recurrences in Eqn 12.6 could be represented as:

$$\mathbf{f} = \mathbf{Q}\mathbf{f} + \boldsymbol{\varepsilon} \quad (12.13)$$

where **f** is the vector of sire by dam subclasses and the row *i* of **Q** contains the elements of **b** from Eqn 12.7 in columns pertaining to identified parent subclasses of subclass *i*. The relationship matrix for subclasses in **f** is $\mathbf{F} = \text{var}(\mathbf{f})/\sigma_f^2$. From Eqn (12.13):

$$\mathbf{f} = (\mathbf{I} - \mathbf{Q})^{-1}\boldsymbol{\varepsilon}$$

The variance–covariance of **f** is:

$$\text{var}(\mathbf{f}) = \mathbf{F}\sigma_f^2 = (\mathbf{I} - \mathbf{Q}')^{-1}\mathbf{R}(\mathbf{I} - \mathbf{Q})^{-1}\sigma_f^2$$

with:

$$\mathbf{R}\sigma_f^2 = \text{var}(\boldsymbol{\varepsilon})$$

Therefore:

$$\mathbf{F}^{-1} = (\mathbf{I} - \mathbf{Q}')\ \mathbf{R}^{-1}(\mathbf{I} - \mathbf{Q}) \quad (12.14)$$

The diagonal elements **R** can be obtained from Eqn 12.8. The off-diagonals are zeros if all ancestor subclasses providing relationship ties are included in **f**. To ensure a diagonal **R**, Hoeschele and VanRaden specified two conditions to be used in deciding which subclasses should be included in **f** as known. These are:

1. A subclass should remain in **f** if any of its parent subclasses remain in **f**.
2. A subclass should remain in **f** if **f** contains two or more of its immediate progeny subclasses.

Equation 12.14 implies that $\mathbf{F}^{-1}$ can be calculated from a list of subclasses and their parent subclass effects by computing for the ith subclass, r^{ii} (the diagonal element i of $\mathbf{R}^{-1}$) and c_i (the ith row of $(\mathbf{I} - \mathbf{Q})$). Then the contribution of the ith subclass to $\mathbf{F}^{-1}$ is calculated as $c_i c_i' r^{ii}$. In summary, the following procedure could therefore be used to calculate $\mathbf{F}^{-1}$:

1. List animals and their sires and dams. Parents not in the list of animals with more than one progeny should be added to the list while those with one progeny may be treated as unknown.
2. Form a list of all filled (S and D known) subclasses and add ancestor subclasses that provide ties. Ancestors are identified by listing subclasses for the sire with parents of the dam and for the dam with parents of the sire for each filled subclass and then repeating this process for the subclasses just added until no further ancestors are known. The same sex subclasses of animal i with animal j and of animal j with animal i should be treated as identical when listing ancestor subclasses. The list of subclasses is sorted such that progeny subclass precedes its parent subclasses. Commencing with the oldest ancestor subclass, subclasses could be regarded as unknown if they are not filled, have no known parents and provide no ties for at least two filled descendant subclasses.

The number of connections provided by an ancestor subclass may be approximately determined from counts formed when ancestor subclasses are being identified originally. Progeny subclass (f_{SD}) would contribute 1 to parent subclasses of type $f_{S,SD}$ and $f_{SS,D}$ but –1 to parent subclasses of type $f_{SS,SD}$. The substraction of 1 is due to the fact that $f_{S,SD}$ and $f_{SS,D}$ are regarded as progeny subclasses of $f_{SS,SD}$ and both may have come from one f_{SD}. It should be noted, however, that some subclasses which should be deleted for having a count of less than 2 may be needed in order to achieve a diagonal **R**. Thus if both $f_{S,SD}$ and $f_{SS,D}$ are known, for instance, it may be necessary to add back subclasses of type $f_{SS,SD}$ if they have been deleted for a count of less than 2.
3. Go through the list of all subclasses and calculate contributions (coefficients) of each subclass i to $\mathbf{F}^{-1}$ as $r^{11} c_i c_i'$. The vector $\mathbf{c}_i$ contains non-zero coefficients, which is equal to 1 in subclass i and equal to **–b** for parent subclasses, with **b** computed as in Eqn 12.7.
4. Sort the coefficients by columns within rows and sum those with identical columns and rows to obtain $\mathbf{F}^{-1}$.

12.4.2 Prediction of dominance effects

So far, the discussion has been on the inverse of the relationship matrix for subclass effects but the major interest is the prediction of dominance effects.

Since the inheritance of dominance effects is from subclass effects, dominance effects can be predicted by the inclusion of the inverse of the relationship matrix ($\mathbf{D}_*$) among dominance effects and subclass effects in the MME. From Eqns 12.5 and 12.13, the dominance ($\mathbf{d}$) and subclass effect ($\mathbf{f}$) may be predicted as:

$$\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix}=\begin{bmatrix}0 & \mathbf{S}\\ 0 & \mathbf{Q}\end{bmatrix}\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{e}\end{bmatrix}$$

with:

$$\text{var}\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix}=\mathbf{D}_*\sigma_d^2 \quad \text{and} \quad \text{var}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}0.75\mathbf{I} & 0\\ 0 & 0.25\mathbf{R}\end{bmatrix}\sigma_d^2$$

where $\mathbf{S}$ is the incidence matrix relating $\mathbf{d}$ to $\mathbf{f}$, and $\boldsymbol{\beta}$ equals $\mathbf{d}$ minus $\mathbf{Sf}$. Therefore:

$$\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix}=\begin{bmatrix}\mathbf{I} & -\mathbf{S}\\ 0 & \mathbf{I}-\mathbf{Q}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{e}\end{bmatrix}$$

and the inverse of $\mathbf{D}_*$ can be computed as:

$$\mathbf{D}_*^{-1}=\begin{bmatrix}\mathbf{I} & 0\\ -\mathbf{S}' & \mathbf{I}-\mathbf{Q}'\end{bmatrix}\begin{bmatrix}\left(\frac{4}{3}\right)\mathbf{I} & 0\\ 0 & 4\mathbf{R}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{S}\\ 0 & \mathbf{I}-\mathbf{Q}\end{bmatrix} \tag{12.15}$$

From the above, the inverse of $\mathbf{D}_*$ is similar to $\mathbf{F}^{-1}$ with coefficients of $\frac{4}{3}$ on the diagonals of dominance effects, $-\frac{4}{3}$ of off-diagonals linking dominance to subclass effects, and the coefficients contributed by the subclass effects are multiplied by 4. The matrix $\mathbf{D}_*^{-1}$ can then be included in the MME, resulting in the prediction of both dominance and subclass effects. The only disadvantage is that the inclusion of subclass effects in the MME will increase the order of equations, but the method can easily be applied to large data sets.

12.4.3 Calculating the inverse of the relationship matrix among dominance and subclass effects for example data

Example 12.3
Using the pedigree information in Example 12.1, the calculations of $\mathbf{F}^{-1}$ and $\mathbf{D}_*^{-1}$ are illustrated.

Setting up $\mathbf{F}^{-1}$

Application of rules 1 to 2 in Section 12.4.1 for calculating $\mathbf{F}^{-1}$ generated Table 12.1. Creating a list of filled subclasses in the first pass (pass 1) through the pedigree in reverse order generated subclasses A to E (sorted by sire) in Table 12.1. Passes 2 and 3 through this list identified all ancestor subclasses (subclasses F to N). Counts to

Table 12.1. List of filled sire × dam subclasses and ancestor subclasses.

Φ	Sire × dam subclass S	D	Pass subclass added	Counts from progeny subclasses	Status	φ	Known parent subclasses		
A	6	8	1		KN	1	2	3	6
B	6	5	1	1	KN	2	3	6	
C	3	8	1	1	KN	3	6		
D	3	4	1		KN	4			
E	1	2	1		KN	5			
F	4	8	2	1	UK				
G	3	5	2	1 + 1 – 1 = 1	KN	6			
H	6	1	2	1	UK				
I	6	2	2	1	UK				
J	4	5	2	1 + 1 – 1 = 1	UK				
K	3	1	3	1 + 1 – 1 = 1	UK				
L	3	2	3	1 + 1 – 1 = 1	UK				
M	4	1	3	1 + 1 – 1 = 1	UK				
N	4	2	3	1 + 1 – 1 = 1	UK				

Φ, consecutive label for subclasses.
S, sire; D, dam; KN, known; UK, regarded as unknown.
φ, consecutive number for known subclasses.

determine whether ancestor subclasses are treated as known or unknown were calculated as specified earlier. Subclasses of the types $f_{S,SD}$ and $f_{SS,SD}$ received a count of 1 and –1, respectively, from progeny subclass f_{SD}. Thus subclass $f_{3,5}$ received a count of 1 from each of its progeny subclasses, $f_{3,8}$ and $f_{6,5}$, and a count of –1 from $f_{6,8}$. Again, $f_{4,1}$ received 1 each from $f_{6,1}$ and $f_{4,5}$ and –1 from $f_{6,5}$. Proceeding through the ancestor subclasses (F to N), those with a count of 1 and with at least two progeny subclasses known are regarded as unknown. Only the ancestor subclass $f_{3,5}$ was regarded as known because two of its progeny subclasses ($f_{3,8}$ and $f_{6,5}$) were known although it had a count of 1.

Using rule 3, the contribution of subclass i regarded as known (subclasses 1 to 6 (see Table 12.1)) to $\mathbf{F}^{-1}$ is then calculated as $\mathbf{c}_i\mathbf{c}_i'r^{ii}$. For example, for the subclass $f_{6,8}$ (subclass 1), three parent subclasses are known: 2, 3 and 6, which are of the subclass type $f_{S,DD}$, $f_{SS,D}$ and $f_{SS,DD}$, respectively. Therefore, $\mathbf{b}_1' = [0.5\ 0.5\ -0.25]$, $\mathbf{c}_2' = [1 - \mathbf{b}_1'] = [1\ -0.5\ -0.5\ 0.25]$. The matrix, $\mathbf{F}_1$, the relationship among parent subclasses 2, 3 and 6 (see 12.14) is:

$$\mathbf{F}_1 = \begin{array}{c} \begin{matrix} 2 & 3 & 6 \end{matrix} \\ \begin{bmatrix} 1.00 & 0.25 & 0.50 \\ 0.25 & 1.00 & 0.50 \\ 0.50 & 0.50 & 1.00 \end{bmatrix} \end{array}$$

The contribution of $f_{6,8}$ to $\mathbf{F}^{-1}$ therefore is:

$$\mathbf{c}_1\mathbf{c}_1' r^{11} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 6 \end{array} \\ \begin{bmatrix} 1.00 & -0.50 & -0.50 & 0.25 \\ -0.50 & 0.25 & 0.25 & -0.125 \\ -0.50 & 0.25 & 0.25 & -0.125 \\ 0.25 & -0.125 & 0.125 & 0.0625 \end{bmatrix} \end{array} 1.778$$

where $r^{11} = 1/(1 - (\mathbf{b}_1'\mathbf{F}_1\mathbf{b}_1)) = 1/(1 - 0.4375) = 1.778$ (see Eqn 12.8).

Processing of all subclasses gives $\mathbf{F}^{-1}$ as:

$$\mathbf{F}^{-1} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1.778 & -0.889 & -0.889 & 0.000 & 0.000 & 0.445 \\ -0.889 & 1.778 & 0.445 & 0.000 & 0.000 & -0.889 \\ -0.889 & 0.445 & 1.778 & 0.000 & 0.000 & -0.889 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 \\ 0.445 & -0.889 & -0.889 & 0.000 & 0.000 & 1.778 \end{bmatrix} \end{array}$$

The methodology can be verified by calculating the dominance relationship matrix among animals as $\mathbf{D} = (0.25)\mathbf{SFS}' + \mathbf{I}(0.75)$, which should give the same **D** as that calculated using Eqn 12.1. **S**, as defined earlier, relates dominance effects to subclass effects. For the example pedigree:

$$\mathbf{S}' = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} 5 & 6 & 7 & 9 & 10 & 11 & 12 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

and:

$$\mathbf{D} = (0.25)\mathbf{SFS}' + \mathbf{I}(0.75)$$

$$= \begin{array}{c} \begin{array}{ccccccc} 5 & 6 & 7 & 9 & 10 & 11 & 12 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.0625 & 0.0625 & 0.125 & 0.125 \\ 0 & 0 & 0.0625 & 1 & 0.25 & 0.125 & 0.125 \\ 0 & 0 & 0.0625 & 0.25 & 1 & 0.125 & 0.125 \\ 0 & 0 & 0.125 & 0.125 & 0.125 & 1 & 0.25 \\ 0 & 0 & 0.125 & 0.125 & 0.125 & 0.25 & 1 \end{bmatrix} \end{array}$$

which is the same as the **D** (Section 12.3.1) calculated from the pedigree using Eqn 12.1.

Let $\mathbf{D}_*^{-1}$ be partitioned as:

$$\mathbf{D}_*^{-1} = \begin{bmatrix} \mathbf{D}_{*11}^{-1} & \mathbf{D}_{*12}^{-1} \\ \mathbf{D}_{*21}^{-1} & \mathbf{D}_{*22}^{-1} \end{bmatrix}$$

where $\mathbf{D}_{*11}^{-1}$ is the top 12 by 12 block for dominance effects for animals, $\mathbf{D}_{*22}^{-1}$ is the bottom 6 by 6 block for subclass effects and $\mathbf{D}_{*12}^{-1}$ is the block for dominance by subclass effects. For the example data using Eqn 12.15, the submatrices of $\mathbf{D}_*^{-1}$ are:

$$\mathbf{D}_{*11}^{-1} = \text{diag}\left(4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3,\ 4/3\right)$$

$$\mathbf{D}_{*12}^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.333 & -1.333 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1.333 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.333 & -1.333 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1.333 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\mathbf{D}_{*12}^{-1}$ is the transpose of $\mathbf{D}_{*21}^{-1}$, and:

$$\mathbf{D}_{*22}^{-1} = \begin{bmatrix} 7.111 & -3.556 & -3.556 & 0 & 0 & 1.778 \\ -3.556 & 7.111 & 1.778 & 0 & 0 & -3.556 \\ -3.556 & 1.778 & 7.111 & 0 & 0 & -3.556 \\ 0 & 0 & 0 & 4.0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.0 & 0 \\ 1.778 & -3.556 & -3.556 & 0 & 0 & 7.111 \end{bmatrix}$$

The matrix $\mathbf{D}_*^{-1}$ can be included in the usual MME for the prediction of dominance and subclass effects.

12.5 Epistasis

Epistasis refers to the interaction among additive and dominance genetic effects; for instance, additive by additive, additive by dominance, additive by additive by dominance, etc. The epistasis relationship matrix can be derived from **A** and **D** as:

A#A	for additive by additive
D#D	for dominance by dominance
AA#D	for additive by additive by dominance

where # represents the Hadamard product of the two matrices. The *ij* element of the Hadamard product of the two matrices is the product of the *ij* elements of the two matrices. Thus if **M** = **A#B**, then $m_{ij} = (a_{ij})(b_{ij})$ where the matrices **A** and **B** should be of the same order.

The model in Eqn 12.2 can be expanded to include epistatic effects as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wd} + \mathbf{Sep} + \mathbf{e}$$

where **ep** is the vector of interaction (epistatic) effects. The evaluation can be carried out as described in Section 12.3 but the major limitation is obtaining the inverse of the epistatic relationship matrix for large data sets. However, VanRaden and Hoeschele (1991) presented a rapid method for obtaining the inverse of the epistatic relationship matrix when epistasis results from interactions between additive by additive (**A** × **A**) genetic effects when the population is inbred or not. The approach is similar to the method described for obtaining the inverse of the dominance relationship matrix and it involves including sire × dam subclasses; consequently, the details of the method have not been covered in this section. The method involves calculating the inverse of **U**, the relationship matrix among epistatic and subclass effects, and $\mathbf{U}^{-1}$ is then included in the usual MME for the prediction of epistatic and sire × dam subclass effects.

The rules for obtaining $\mathbf{U}^{-1}$ for a population that is not inbred are given in the next section, with an illustration.

12.5.1 Rules for the inverse of the relationship matrix for epistatic and subclass effects

The inverse of **U** can be computed by going through a list of individuals and their parents and sire × dam subclasses. See rules 1 and 2 in Section 12.4.1 on how such a list should be set up. The contribution of individual *i* in the list to $\mathbf{U}^{-1}$ is computed by the following rules:

1. For an individual *i* with both parents and subclass effects known, the contribution to $\mathbf{U}^{-1}$ is:

$$\begin{array}{c} c \quad s \quad d \quad (s,d) \\ \begin{bmatrix} 16 & -4 & -4 & -16 \\ -4 & 1 & 1 & 4 \\ -4 & 1 & 1 & 4 \\ -16 & 4 & 4 & 16 \end{bmatrix} \end{array} (1/12) \qquad (12.16)$$

2. For an individual with both parents known but subclass effects treated as unknown, the contribution to $\mathbf{U}^{-1}$ is:

$$\begin{array}{c} c \quad s \quad d \\ \begin{bmatrix} 16 & -4 & -4 \\ -4 & 1 & 1 \\ -4 & 1 & 1 \end{bmatrix} \end{array} (1/14) \qquad (12.17)$$

3. If only one parent, say *s*, is known, then the contribution is:

$$\begin{array}{c} c \quad s \\ \begin{bmatrix} 16 & -4 \\ -4 & 1 \end{bmatrix} \end{array} (1/15) \qquad (12.18)$$

4. If both parents and subclass are unknown, add 1 to the individual diagonal.
5. For sire × dam subclasses, the contribution of the *i*th subclass to $\mathbf{U}^{-1}$ is the same as for the inverse of the dominance matrix (see rule 3 in Section 12.4.1) except that the coefficients are multiplied by 8.
6. Sort coefficients by row and by columns within a row, and sum coefficients with identical row and columns to obtain $\mathbf{U}^{-1}$.

The method can be verified by inverting $\mathbf{U}^{-1}$ to form **U**. The animal by animal submatrix of **U** should be equal to the epistatic relationship matrix calculated as **A#A**.

12.5.2 Calculating the inverse relationship matrix for epistasis and the subclass matrix for an example pedigree

Example 12.4
The calculation of $\mathbf{U}^{-1}$ is illustrated below using the pedigree information in Example 12.1.

The identification of sire and dam subclasses and their ancestors subclasses treated as known has been discussed in Section 12.4.3. Thus the list of animals and known subclasses is:

Animal	Sire	Dam
1	0	0
2	0	0
3	0	0
4	0	0
5	1	2
6	3	4
7	6	5
8	0	5
9	3	8
10	3	8
11	6	8
12	6	8

Subclasses	Parent subclasses
6,8	6,5, 3,8, 3,5
6,5	3,8, 3,5
3,8	3,5
3,4	
1,2	
3,5	

In setting up $\mathbf{U}^{-1}$, animals 1 to 12 have been regarded as rows 1 to 12 while subclasses have been assigned rows 13 (subclass (6,8)) to 18 (subclass (3,5)). The first four animals have both parents and sire–dam subclasses unknown and therefore each contributes 1 to their respective diagonals. The parents of animals 5, 6, 7, 10, 11 and 12 and their sire × dam subclass effects are known; therefore, the contributions of

each of these animals to $\mathbf{U}^{-1}$ are computed using rule 1 in Section 12.5.1. For animals 8 and 9, only one of their parents is known and rule 3 is applicable when processing these animals. The calculation of the contributions of subclass effects has been given in Section 12.4.3 (Example 12.3); these are multiplied by 8, as mentioned earlier. After processing all animals and subclass effects, the top 12 by 12 submatrix of $\mathbf{U}^{-1}$ (block for animals only) is:

$$\begin{bmatrix}
1.083 & 0.083 & 0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.083 & 1.083 & 0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.250 & 0.083 & 0.000 & -0.333 & 0.000 & 0.167 & -0.333 & -0.333 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.083 & 1.083 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
-0.333 & -0.333 & 0.000 & 0.000 & 1.483 & 0.083 & -0.333 & -0.267 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & -0.333 & -0.333 & 0.083 & 1.583 & -0.333 & 0.167 & 0.000 & 0.000 & -0.333 & -0.333 \\
0.000 & 0.000 & 0.000 & 0.000 & -0.333 & -0.333 & 1.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.167 & 0.000 & -0.267 & 0.167 & 0.000 & 1.400 & -0.333 & -0.333 & -0.333 & -0.333 \\
0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 1.333 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 0.000 & 1.333 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 0.000 & -0.333 & 0.000 & 0.000 & 1.333 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 1.333
\end{bmatrix}$$

The top 12 by 12 submatrix of the inverse $\mathbf{U}^{-1}$ is the epistatic relationship matrix for the animals and is:

$$\begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.250 & 0.000 & 0.063 & 0.063 & 0.016 & 0.016 & 0.016 & 0.016 \\
0.000 & 1.000 & 0.000 & 0.000 & 0.250 & 0.000 & 0.063 & 0.063 & 0.016 & 0.016 & 0.016 & 0.016 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.250 & 0.063 & 0.000 & 0.250 & 0.250 & 0.063 & 0.063 \\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.250 & 0.063 & 0.000 & 0.000 & 0.000 & 0.063 & 0.063 \\
0.250 & 0.250 & 0.000 & 0.000 & 1.000 & 0.000 & 0.250 & 0.250 & 0.063 & 0.063 & 0.063 & 0.063 \\
0.000 & 0.000 & 0.250 & 0.250 & 0.000 & 1.000 & 0.250 & 0.000 & 0.063 & 0.063 & 0.250 & 0.250 \\
0.063 & 0.063 & 0.063 & 0.063 & 0.250 & 0.250 & 1.000 & 0.063 & 0.063 & 0.063 & 0.141 & 0.141 \\
0.063 & 0.063 & 0.000 & 0.000 & 0.250 & 0.000 & 0.063 & 1.000 & 0.250 & 0.250 & 0.250 & 0.250 \\
0.016 & 0.016 & 0.250 & 0.000 & 0.063 & 0.063 & 0.063 & 0.250 & 1.000 & 0.250 & 0.141 & 0.141 \\
0.016 & 0.016 & 0.250 & 0.000 & 0.063 & 0.063 & 0.063 & 0.250 & 0.250 & 1.000 & 0.141 & 0.141 \\
0.016 & 0.016 & 0.063 & 0.063 & 0.063 & 0.250 & 0.141 & 0.250 & 0.141 & 0.141 & 1.000 & 0.250 \\
0.016 & 0.016 & 0.063 & 0.063 & 0.063 & 0.250 & 0.141 & 0.250 & 0.141 & 0.141 & 0.250 & 1.000
\end{bmatrix}$$

It is equal to the epistatic relationship matrix calculated as **A#A**. The matrix $\mathbf{U}^{-1}$ can then be incorporated into the usual MMEs for the prediction of epistatic and subclass effects.

13 Analysis of Ordered Categorical Traits

13.1 Introduction

Some traits of economic importance in animal breeding, such as calving ease or litter size, are expressed and recorded in a categorical fashion. For instance, in the case of calving ease, births may be assigned to one of several distinct classes, such as difficult, assisted and easy calving, or litter size in pigs might be scored 1, 2, 3 or more piglets born per sow. Usually, these categories are ordered along a gradient. In the case of calving ease, for example, the responses are ordered along a continuum measuring the ease with which birth occurred. These traits are therefore termed ordered categorical traits. Such traits are not normally distributed, and animal breeders have usually attributed the phenotypic expression of categorical traits to an underlying continuous unobservable trait that is normally distributed, referred to as the liability (Falconer and Mckay, 1996). The observed categorical responses are therefore due to animals exceeding particular threshold levels (t_i) of the underlying trait. Thus with m categories of responses, there are $m - 1$ thresholds such that $t_1 < t_2 < t_3, \ldots, t_{m-1}$. For traits such as survival to a particular age or stage, the variate to be analysed is coded 1 (survived) or 0 (not survived) and there is basically only one threshold.

Linear and non-linear models have been applied for the genetic analysis of categorical traits with the assumption of an underlying normally distributed liability. Usually, the non-linear (threshold) models are more complex and have higher computing requirements. The advantage of the linear model is the ease of implementation, as programs used for analysis of quantitative traits could be utilized without any modifications. However, Fernando *et al*. (1983) indicated that some of the properties of BLUP do not hold with categorical traits. Such properties include the invariance of BLUP to certain types of culling (selection) and the ability of BLUP to maximize the probability of correct pairwise ranking. Also, Gianola (1982) indicated that the variance of a categorical trait is a function of its expectation and the application of a linear model that has fixed effects in addition to an effect common to all observations results in heterogeneity of variance.

In a simulation study, Meijering and Gianola (1985) demonstrated that with no fixed effects and constant or variable number of offspring per sire, an analysis of a binary trait with either a linear or non-linear model gave similar sire rankings. This was independent of the heritability of the liability or incidence of the binary trait. However, with the inclusion of fixed effects and a variable number of progeny per sire, the non-linear model gave breeding values that were more similar to the true breeding values compared with the linear model. The advantage of the threshold model increased as the incidence of the binary trait and its heritability decreased. Thus for traits with low heritability and low incidence, a threshold model might be the method of choice.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

The principles required to apply a linear model for the analysis of categorical traits are the same as discussed in the previous chapters; therefore, the main focus of this chapter is on threshold models, assuming a normal distribution for the liability. Cameron (1997) illustrated the analysis of a binary trait with a threshold model using a logit function. In this chapter, sample data used for the illustration with the threshold model have also been analysed with a linear model for the purposes of comparison.

13.2 The Threshold Model

13.2.1 Defining some functions of the normal distribution

The use of the threshold model involves the use of some functions of the normal distribution and these are briefly defined. Assume the number of lambs born alive to ewes in the breeding season is scored using four categories. The distribution of liability for the number of lambs born alive with three thresholds (t_j) can be illustrated as in Fig. 13.1, where N_j is the number of ewes with the jth number of lambs and are those exceeding the threshold point t_{j-1}, when $j > 1$ and $j \leq m - 1$.

With the assumption that the liability (l) is normally distributed ($l \sim N(0,1)$), the height of the normal curve at t_j ($\phi(t_j)$) is:

$$\phi(t_j) = \exp(-0.5t_j^2) / \sqrt{2\pi} \tag{13.1}$$

For instance, given that $t_j = 0.779$, then $\phi(0.779) = 0.2945$.

The function $\Phi()$ is the standard cumulative distribution function of the normal distribution. Thus $\Phi(k)$ or Φ_k gives the areas under the normal curve up to and including the kth category. Given that there are m categories, then $\Phi_k = 1$ when the kth category equals m. For a variable x, for instance, drawn from a normal distribution, the value Φ_x can be computed, using a subroutine from the IMSL (1980) library. Thus if $x = 0.560$, then $\Phi(0.560) = 0.7123$.

$P(k)$ defines the probability of a response being observed in category k assuming a normal distribution. This is also the same probability that a response is between the thresholds defined by category k. Thus $P(k)$ or P_k may be calculated as

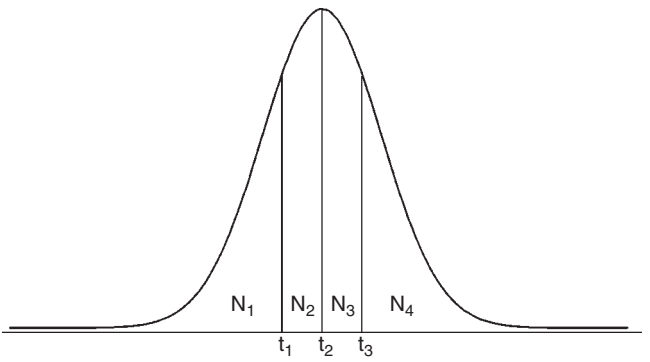


Fig. 13.1. The distribution of liability for number of lambs born alive with four categories and three thresholds.

$P(k) = \Phi(k) - \Phi(k-1)$ with $\Phi(k-1) = 0$, when $k = 1$; or expressed in terms of thresholds defining the category k, $P_k = \Phi t_k - \Phi t_{(k-1)}$. For instance in Fig. 13.1, the probability of response in the k category (P_k) can be computed as:

$$P_1 = \Phi(t_1) \tag{13.2}$$
$$P_2 = \Phi(t_2) - \Phi(t_1) \tag{13.3}$$
$P_3 = \Phi(t_3) - \Phi(t_2)$ and
$P_4 = 1 - \Phi(t_3)$

13.2.2 Data organization and the threshold model

Usually, the data are organized into an s by m contingency table (Table 13.1), where the s rows represent individuals or herd–year subclasses of effects, such as herd, and the m columns indicate ordered categories of response. If the rows represent individuals, then all n_{jk} will be zero except one and the $n_{j.} = 1$, for $j = 1,\ldots, s$.

The linear model for the analysis of the liability is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of liability on a normal scale, $\mathbf{b}$ and $\mathbf{u}$ are vectors of fixed and random (sire or animal) effects, respectively, and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating data to fixed effects and responses effects, respectively. Since $\mathbf{y}$ is not observed, it is not possible to solve for $\mathbf{u}$ using the usual MME.

Given that $\mathbf{H}' = [\mathbf{t}', \mathbf{b}', \mathbf{u}']$, where $\mathbf{t}$ is the vector for the threshold effects, Gianola and Foulley (1983) proceeded to find the estimator $\hat{\mathbf{H}}$ that maximizes the log of the posterior density L($\mathbf{H}$). The resulting set of equations involved in the differentiation were not linear with respect to $\mathbf{H}$. They therefore provided the following non-linear iterative system of equations based on the first and second derivatives, assuming a normal distribution to obtain solutions for $\Delta\mathbf{t}$, $\Delta\mathbf{b}$ and $\Delta\mathbf{u}$:

$$\begin{bmatrix} \mathbf{Q} & \mathbf{L'X} & \mathbf{L'Z} \\ \mathbf{X'L} & \mathbf{X'WX} & \mathbf{X'WZ} \\ \mathbf{Z'L} & \mathbf{Z'WX} & \mathbf{Z'WZ} + \mathbf{A}^{-1}\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{b} \\ \Delta\mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \mathbf{X'v} \\ \mathbf{Z'v} - \mathbf{A}^{-1}\mathbf{G}^{-1}\mathbf{u} \end{bmatrix} \tag{13.4}$$

Table 13.1. Ordered categorical data arranged as an s by m contingency table.

	Categories[a]						
Subclasses	1	2	…	k	…	m	Totals[b]
1	n_{11}	n_{12}	…	n_{1k}	…	n_{1m}	$n_{1.}$
2	n_{21}	n_{22}	…	n_{2k}	…	n_{2m}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
j	n_{j1}	n_{j2}	…	n_{jk}	…	n_{jm}	$n_{j.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
s	n_{s1}	n_{s2}	…	n_{sk}	…	n_{sm}	$n_{s.}$

[a] n_{jk} = number of counts in category k of response in row j.
[b] $n_{j.} = \sum_{k=1}^{m} n_{jk}$

with $\mathbf{G} = \mathbf{I}\sigma_s^2$ or $\mathbf{I}\sigma_u^2$ if a sire or an animal model is being fitted in a univariate situation. They presented equations for the calculation of the matrices in Eqn 13.4, which are outlined below. The calculation of most of these matrices involves P_{jk} (see Eqn 13.2) and it is initially described. P_{jk}, the response in the kth category under the conditions of the jth row, is:

$$P_{jk} = \Phi(t_k - a_j) - \Phi(t_{k-1} - a_j);\ k = 1,\ m-1;\ j = 1,\ldots,\ s \tag{13.5}$$

where $a_j = (\mathbf{x}_j\mathbf{b} + \mathbf{z}_j\mathbf{u})$, with x_j and z_j being the jth row of $\mathbf{X}$ and $\mathbf{Z}$, respectively. This equation is no different from that in Section 13.2.1, but it shows that the distribution of response probabilities by category is a function of the distance between a_j and the threshold. Similarly, the height of the normal curve at t_k (Eqn 13.1) under the conditions of the jth row becomes:

$$\phi_{jk} = \phi(t_k - a_j) \tag{13.6}$$

The formulae for computing the various matrices and vectors in Eqn 13.4 are outlined below.

The jth element of vector $\mathbf{v}$ can be calculated as:

$$v_j = \sum_{k=1}^{m} n_{jk}\left(\frac{\phi_{j(k-1)} - \phi_{jk}}{p_{jk}}\right) \tag{13.7}$$

The elements of the matrix $\mathbf{W}$, which is a weighting factor, is computed as:

$$w_{jj} = n_{j.}\sum_{k=1}^{m}\frac{\left(\phi_{j(k-1)} - \phi_{jk}\right)^2}{p_{jk}} \tag{13.8}$$

The matrix $\mathbf{Q}$ is an $(m-1)$ by $(m-1)$ banded matrix and the diagonal elements are calculated as:

$$q_{kk} = \sum_{j=1}^{s} n_{j.}\frac{P_{jk} + P_{j(k+1)}}{P_{jk}\,P_{j(k+1)}}\,\phi_{jk}^2, \qquad \text{for } k = 1 \text{ to } (m-1) \tag{13.9}$$

and the off-diagonal elements are:

$$q_{(k+1)k} = -\sum_{j=1}^{s} n_{j.}\frac{\phi_{j(k+1)}\,\phi_{jk}}{P_{j(k+1)}}, \quad \text{for } k = 1 \text{ to } (m-2) \tag{13.10}$$

with the element $q_{k(k+1)} = q_{(k+1)k}$.

The matrix $\mathbf{L}$ is of order s by $(m-1)$ and its jkth element is calculated as:

$$l_{jk} = -n_{j.}\phi_{jk}\left(\frac{\phi_{jk} - \phi_{j(k-1)}}{P_{jk}} - \frac{\phi_{j(k+1)} - \phi_{jk}}{P_{j(k+1)}}\right) \tag{13.11}$$

The vector $\mathbf{p}$ is accumulated over all subclasses and its elements are:

$$p_k = \left\{\sum_{j=1}^{s}\left[\frac{n_{jk}}{p_{jk}} - \frac{n_{j(k+1)}}{p_{j(k+1)}}\right]\phi_{jk}\right\};\ \ k = 1,\ m-1 \tag{13.12}$$

The remaining matrices in Eqn 13.4 can be computed by matrix multiplication.

13.2.3 Numerical example

Example 13.1

The analysis of categorical traits is illustrated below, using the calving ease data described by Gianola and Foulley (1983) but with a relationship matrix included for the sires and the age of dam effect omitted from the model. The data consisted of calving ease scores from 28 male and female calves born in two herd–years from cows mated to four sires. Cows were scored for calving ease using three ordered categories: 1 = normal birth, 2 = slight difficulty and 3 = extreme difficulty. The data set is presented in Table 13.2.

The following pedigree was assumed for the four sires:

Animal	Sire	Dam
1	0	0
2	0	0
3	1	0
4	3	0

The sire variance used in the analysis was assumed to be $\frac{1}{19}$. In the underlying scale, residual variance equals one; therefore, $\sigma_e^2/\sigma_s^2 = 4 - h^2/h^2 = 19$. Thus the σ_s^2 assumed corresponded to a heritability of 0.20 on the underlying scale.

Table 13.2. Distribution of calving ease score by herd–year and sex of calf subclasses.

Herd	Sex of calf	Sire of calf	Category of response[a]			Total
			1	2	3	
1	Male	1	1	0	0	1
1	Female	1	1	0	0	1
1	Male	1	1	0	0	1
1	Female	2	0	1	0	1
1	Male	2	1	0	1	2
1	Female	2	3	0	0	3
1	Male	3	1	1	0	2
1	Female	3	0	1	0	1
1	Male	3	1	0	0	1
2	Female	1	2	0	0	2
2	Male	1	1	0	0	1
2	Male	1	0	0	1	1
2	Female	2	1	0	1	2
2	Male	2	1	0	0	1
2	Female	3	0	1	0	1
2	Male	3	0	0	1	1
2	Male	4	0	1	0	1
2	Female	4	1	0	0	1
2	Female	4	2	0	0	2
2	Male	4	2	0	0	2

[a]1, normal birth; 2, slight difficulty; 3, extreme difficulty.

The vectors of solutions in Eqn 13.4 for the example data are:

$\mathbf{t}' = (t_1\ t_2)$, since there are two thresholds
$\mathbf{b}' = (h_1\ h_2\ \eta_1\ \eta_2)$
$\mathbf{u}' = (u_1\ u_2\ u_3\ u_4)$

where h_i and η_i represent solutions for level i of herd–year and the sex of calf effects, respectively; and **u** is the vector of solutions for sires.

The inverse of the relationship for the assumed pedigree is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.3333 & 0.0000 & -0.6667 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ -0.6667 & 0.0000 & 1.6667 & -0.6667 \\ 0.0000 & 0.0000 & -0.6667 & 1.3333 \end{bmatrix}$$

For the example data, the transpose of matrix **X**, which relates subclasses to herd–year and sex of calf effects, and that of matrix **Z**, which relates subclasses to sires, are:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & \mathbf{1} & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

and:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & \mathbf{0} & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Starting values for **t**, **b** and **u** are needed to commence the iterative process. Let **b** = **u** = 0, but starting values for t_i can be computed from the proportion of records in all categories of response preceding t_i. In this example, there is only one category before t_1 and 0.679 of the records are in this category. The first two categories precede t_2 and 0.857 of the records are observed in both categories. Using these proportions, the values of t can be obtained from the usual table of standardized normal deviates of the normal distribution. From these proportions, $t_1 = 0.468$ and $t_2 = 1.080$ and these were used as starting values. However, using various starting values of t, Gianola and Foulley (1983) demonstrated that the system of equations converged rapidly. It seems, therefore, that the system of equations is not very sensitive to starting values for t. The calculations of the various matrices in the equations have been illustrated below using solutions obtained after the first iteration. The solutions obtained at the end of the first iteration and the updated estimates for the effects (which are now the starting values for the second iteration) are:

Solutions at the end of iteration one	Updated[a] estimates after iteration one
$\Delta t_1 = -0.026992$	$t_1 = 0.441008$
$\Delta t_2 = -0.035208$	$t_2 = 1.044792$
$\Delta \hat{h}_1 = 0.000000$	$\hat{h}_1 = 0.000000$

Continued

(*Continued*)

Solutions at the end of iteration one	Updated[a] estimates after iteration one
$\Delta\hat{h}_2 = 0.286869$	$\hat{h}_2 = 0.286869$
$\Delta\eta_1 = 0.000000$	$\eta_1 = 0.000000$
$\Delta\eta_2 = -0.358323$	$\eta_2 = -0.358323$
$\Delta u_1 = -0.041528$	$u_1 = -0.041528$
$\Delta u_2 = 0.057853$	$u_2 = 0.057853$
$\Delta u_3 = 0.039850$	$u_3 = 0.039850$
$\Delta u_4 = -0.065178$	$u_4 = -0.065178$

[a]The updated estimates were obtained as the sum of the starting values and the solutions at the end of the first iteration.

The following steps are involved in calculating P_{jk}, which is required to calculate subsequent matrices in Eqn 13.4 for the example data. In each round of iteration and for each subclass, i.e. for $j = 1,\ldots s$:

1. Initially calculate $(t_k - a_j)$ in Eqn 13.5 for $k = 1,\ldots m-1$. Therefore:

$$d_{jk} = (t_k - a_j) = t_k - x_j - z_j \text{ for } k = 1,\ldots m-1$$

where x_j and z_j are the jth rows of **X** and **Z**.

For the example data in the second iteration:

$$d_{11} = t_1 - \hat{h}_1 - \hat{\eta}_1 - \hat{u}_1$$

$$d_{11} = 0.441008 - 0 - 0 - (-0.041528) = 0.482536$$

$$d_{12} = t_2 - \hat{h}_1 - \hat{\eta}_1 - \hat{u}_1$$

$$d_{12} = 1.044792 - 0 - 0 - (-0.041528) = 1.086320$$

$$d_{21} = t_1 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1$$

$$d_{21} = 0.441008 - 0 - (-0.358323) - (-0.041528) = 0.840859$$

$$d_{22} = t_2 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1$$

$$d_{22} = 1.044792 - 0 - (-0.358323) - (-0.041528) = 1.444643$$

$$\vdots$$

$$d_{201} = t_1 - \hat{h}_2 - \hat{\eta}_1 - \hat{u}_4$$

$$d_{201} = 0.441008 - 0.286869 - 0 - (-0.065178) = 0.219317$$

$$d_{202} = t_2 - \hat{h}_2 - \hat{\eta}_1 - \hat{u}_4$$

$$d_{202} = 1.044792 - 0.286869 - 0 - (-0.065178) = 0.823101$$

2. Using the values of d_{jk} computed above, calculate ϕ_{jk} (see Eqn 13.6) and Φ_{jk}, for $k = 0,\ldots, m$. Note that in all cases, when $k = 0$, $\phi_{jk} = \Phi_{jk} = 0$ and when $k = m$, $\phi_{jk} = 0$ and $\Phi_{jk} = 1$.

In the second round of iteration for the example data:

$$\phi_{11} = \phi(0.482536) = 0.355099 \text{ and } \Phi_{11} = \Phi(0.482536) = 0.685288$$

$$\phi_{12} = \phi(1.086320) = 0.221135 \text{ and } \Phi_{12} = \Phi(1.086320) = 0.861331$$

$\phi_{21} = \phi(0.840859) = 0.280142$ and $\Phi_{21} = \Phi(0.840859) = 0.799787$

$\phi_{22} = \phi(1.444643) = 0.140516$ and $\Phi_{22} = \Phi(1.444643) = 0.925721$

$\vdots$

$\phi_{201} = \phi(0.219317) = 0.389462$ and $\Phi_{201} = \Phi(0.219317) = 0.586799$

$\phi_{202} = \phi(0.823101) = 0.284311$ and $\Phi_{202} = \Phi(0.823101) = 0.794775$

3. Then calculate P_{jk} as $\Phi_{jk} - \Phi_{j(k-1)}$ for $k = 1, \ldots, m$
In the second round of iteration, for Example 13.1:

$P_{11} = \Phi_{11} - \Phi_{10} = 0.685288 - 0 = 0.685288$

$P_{12} = \Phi_{12} - \Phi_{11} = 0.861331 - 0.685288 = 0.176044$

$P_{13} = \Phi_{13} - \Phi_{12} = 1.0 - 0.861331 = 0.138669$

$P_{21} = \Phi_{21} - \Phi_{20} = 0.799787 - 0 = 0.799787$

$P_{22} = \Phi_{22} - \Phi_{21} = 0.925721 - 0.799787 = 0.125934$

$P_{23} = \Phi_{23} - \Phi_{22} = 1.0 - 0.925721 = 0.074279$

$\vdots$

$P_{201} = \Phi_{201} - \Phi_{200} = 0.586799 - 0 = 0.586799$

$P_{202} = \Phi_{202} - \Phi_{201} = 0.794775 - 0.586799 = 0.207976$

$P_{203} = \Phi_{203} - \Phi_{202} = 1.0 - 0.794775 = 0.205225$

The calculation of the remaining matrices in the MME can now be illustrated for the example data. The first elements of **W** using Eqn 13.8 for the example data are:

$$w_{11} = 1\left[\frac{(0-0.355099)^2}{0.685288} + \frac{(0.355099-0.221135)^2}{0.176044} + \frac{(0.221135-0)^2}{0.138669}\right] = 0.638589$$

and:

$\mathbf{W}$ = diag[0.638589 0.518748 0.638589 0.554385 1.332860 1.663156
1.323206 0.548036 0.661603 1.233768 0.710404 0.710404 1.293402
0.728641 0.641496 0.725614 0.705526 0.609417 1.218834 1.411052]

For the vector **v**, the first element can be calculated from Eqn 13.7 as:

$$v_1 = \frac{1(0-0.355099)}{0.685288} + \frac{0(0.355099-0.221135)}{0.176044} + \frac{0(0.221135-0)}{0.138669} = -0.518175$$

and the transpose of **v** is:

$\mathbf{v}'$ = [–0.518175 –0.350270 –0.518175 1.012257 0.943660 –1.179520 0.120754
1.029729 –0.561257 –0.963633 –0.677635 1.366976 1.039337 –0.737615
0.751341 1.304294 0.505592 –0.470090 –0.940181 –1.327414]

The matrix **L** is order 20 by 2 for the example data. The elements in the first row of **L** from Eqn 13.11 can be calculated as:

$$l_{11} = (-1)(0.355099)\left[\frac{(0.355099 - 0)}{0.685288} - \frac{(0.221135 - 0.355099)}{0.176044}\right] = -0.454223$$

$$l_{12} = (-1)(0.221135)\left[\frac{(0.221135 - 0.355099)}{0.176044}\frac{(0 - 0.221135)}{0.138669}\right] = -0.184365$$

The matrix **L** has not been shown because it is too large but the elements of the last row, l_{201} and l_{202}, are –0.910795 and –0.500257, respectively.

The elements of **Q** calculated using Eqns 13.9 and 13.10 are:

$$q_{11} = \frac{1(0.355099)^2(0.685287 + 0.176044)}{(0.685286 * 0.176044)} + \frac{1(0.280142)^2(0.799787 + 0.125934)}{(0.799787 * 0.125934)} + \ldots$$
$$+ \frac{2(0.389462)^2(0.586799 + 0.207976)}{(0.586799 * 0.207976)} = 25.072830$$

$$q_{12} = \frac{-[1(0.355099)(0.221135)}{0.176044} + \frac{1(0.280142)(0.140516)}{0.125934} + \ldots$$
$$+ \frac{2(0.389462)(0.284311)}{0.207976]} = -12.566598$$

$$q_{22} = \frac{1(0.221135)^2(0.176044 + 0.138669)}{(0.176044 * 0.138669)} + \frac{1(0.140516)^2(0.125934 + 0.074279)}{(0.125934 * 0.074279)} + \ldots$$
$$+ \frac{2(0.284311)^2(0.207976 + 0.205225)}{(0.207976 * 0.205225)} = 17.928093$$

Since **Q** is symmetric, $q_{21} = q_{12}$.

Lastly, the elements of **p** can be calculated using Eqn 13.12 as:

$$p_1 = 0.355099\left(\frac{1}{0.685287} - \frac{0}{0.176044}\right) + 0.280142\left(\frac{1}{0.799787} - \frac{0}{0.125934}\right) + \ldots$$
$$+ 0.389462\left(\frac{2}{0.586799} - \frac{0}{0.207976}\right) = -0.288960$$

and:

$$p_2 = 0.221135\left(\frac{0}{0.176044} - \frac{0}{0.138669}\right) + 0.140516\left(\frac{0}{0.125934} - \frac{0}{0.074279}\right) + \ldots$$
$$+ 0.284311\left(\frac{0}{0.207976} - \frac{0}{0.205225}\right) = 0.458984$$

The matrices in Eqn 13.4 can now be obtained by matrix multiplication and $\mathbf{A}^{-1}$ is added to $\mathbf{Z'WZ}$. The matrix $\mathbf{Z'WZ} + \mathbf{A}^{-1}\mathbf{G}^{-1}$ is illustrated below:

$$\mathbf{Z'WZ}+\mathbf{A}^{-1}\mathbf{G}^{-1} = \begin{bmatrix} 29.783773 & 0.000000 & -12.666731 & 0.000000 \\ 0.000000 & 24.572445 & 0.000000 & 0.000000 \\ -12.666731 & 0.000000 & 35.566685 & -12.666731 \\ 0.000000 & 0.000000 & -12.666731 & 29.278162 \end{bmatrix}$$

Then Eqn 13.4 is:

$$\begin{bmatrix} 25.073 & -12.567 & -5.733 & -6.773 & -6.366 & -6.140 & -3.123 & -3.977 & -2.699 & -2.707 \\ -12.567 & 17.928 & -2.146 & -3.215 & -3.220 & -2.141 & -1.327 & -1.595 & -1.201 & -1.238 \\ -5.733 & -2.146 & 7.879 & 0.000 & 4.595 & 3.284 & 1.796 & 3.550 & 2.533 & 0.000 \\ -6.773 & -3.215 & 0.000 & 9.989 & 4.992 & 4.997 & 2.655 & 2.022 & 1.367 & 3.945 \\ -6.366 & -3.220 & 4.595 & 4.992 & 9.586 & 0.000 & 2.698 & 2.062 & 2.710 & 2.117 \\ -6.140 & -2.141 & 3.284 & 4.997 & 0.000 & 8.281 & 1.753 & 3.511 & 1.190 & 1.828 \\ -3.123 & -1.327 & 1.796 & 2.655 & 2.698 & 1.753 & 29.784 & 0.000 & -12.667 & 0.000 \\ -3.977 & -1.595 & 3.550 & 2.022 & 2.062 & 3.511 & 0.000 & 24.572 & 0.000 & 0.000 \\ -2.699 & -1.201 & 2.533 & 1.367 & 2.710 & 1.190 & -12.667 & 0.000 & 35.567 & -12.667 \\ -2.707 & -1.238 & 0.000 & 3.945 & 2.117 & 1.828 & 0.000 & 0.000 & -12.667 & 29.278 \end{bmatrix}$$

$$\begin{bmatrix} \Delta\hat{t}_1 \\ \Delta\hat{t}_2 \\ \Delta\hat{b}_1 \\ \Delta\hat{b}_2 \\ \Delta\hat{\eta}_1 \\ \Delta\hat{\eta}_2 \\ \Delta\hat{u}_1 \\ \Delta\hat{u}_2 \\ \Delta\hat{u}_3 \\ \Delta\hat{u}_4 \end{bmatrix} = \begin{bmatrix} -0.289 \\ 0.459 \\ -0.021 \\ -0.149 \\ -0.099 \\ -0.071 \\ -0.104 \\ -0.021 \\ 0.031 \\ -0.076 \end{bmatrix}$$

The equations were solved with the solutions for $\Delta\hat{b}_1$ and $\Delta\eta_1$ set to zero. The equations converged rapidly, and solutions at various different iteration numbers and the final solutions are given below. Solution from an analysis using a linear model with an α value of 19 are also shown:

	Iteration number				
Effects	1	2	3	7	Solutions from linear models
Threshold					
1	0.4410	0.4375	0.4378	0.4378 ± 0.44[a]	–
2	1.0448	1.0661	1.0675	1.0675 ± 0.47	–
Herd–year					
1	0.0000	0.0000	0.0000	0.0000 ± 0.00	0.0
2	0.2869	0.2763	0.2774	0.2774 ± 0.49	1.0604

Continued

(*Continued*)

Effects	Iteration number 1	2	3	7	Solutions from linear models
Sex of calf					
Male	0.0000	0.0000	0.0000	0.0000 ± 0.00	0.0
Female	–0.3583	–0.3577	–0.3589	–0.3590 ± 0.48	0.5193
Sires					
1	–0.0415	–0.0431	–0.0434	–0.0434 ± 0.22	0.2229
2	0.0579	0.0586	0.0592	0.0592 ± 0.21	0.2751
3	0.0399	0.0410	0.0412	0.0412 ± 0.22	0.3162
4	–0.0652	–0.0653	–0.0660	–0.0660 ± 0.22	0.0985

[a]Standard errors.

The standard errors associated with the results from the last iteration were computed from the square root of the diagonals of the generalized inverse. Sire rankings from the linear model were similar to those from the threshold model except for sires 2 and 3, which ranked differently.

Usually of interest is calculating the probability of response in a given category under specific conditions. For instance, the proportion of calving in the *j*th category of response, considering only female calves in HYS subclass 1 for sire 1 can be estimated as:

$$P_{11} = \Phi(t_1 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) = \Phi(0.4378 - 0 - (-0.3590) - (-0.0434)) = \Phi(0.8402) = 0.800$$

$$P_{12} = \Phi(t_2 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) - \Phi(t_1 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) = \Phi(1.0675 - 0 - (0.3590) - (-0.0434)) - \Phi(0.800) = \Phi(1.4699) - \Phi(0.800) = 0.129$$

$$P_{13} = 1 - \Phi(t_2 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) = 1 - \Phi(1.4699) = 0.071$$

Calculating this probability distribution by category of response for all sires gives the following:

	Probability in category of response 1	2	3
Sire 1	0.800	0.129	0.071
Sire 2	0.770	0.145	0.086
Sire 3	0.775	0.142	0.083
Sire 4	0.803	0.129	0.068

The results indicate that the majority of heifers calving in HYS subclass 1 for all four sires were normal, with a very low proportion of extreme difficulties.

Since sires are used across herds, the interest might be the probability distribution of heifer calvings for each sire across all herds and sexes. Such a probability for each sire in category 1 of response per herd–year–sex subclass (Z_{1kji}) can be calculated as follows:

$$Z_{1kji} = \Phi(t_1 - (\hat{h}_k + \hat{\eta}_j + \hat{u}_i)); \qquad k = 1, 2; \quad j = 1, 2, \; i = 1,\ldots,4$$

Since there are four herd–year–sex subclasses, the probability for sire i in category 1 (S_{1i}) can be obtained by weighting Z_{1kji} by factors that sum up to one. Thus:

$$S_{1i} = \sum_{i=1}^{4}\sum_{k=1}^{2}\sum_{m=1}^{2} a_{km} Z_{1ikm}$$

where $a_{km} = a_{11} + a_{12} + a_{21} + a_{22} = 1$. In the example data, $a_{11} = a_{12} = a_{21} = a_{22} = 0.25$.

Similarly, the probability for each sire in category 2 of response per herd–year–sex subclass (Z_{2kji^*}) can be calculated as:

$$Z_{2kji} = Z_{2kji^*} - Z_{1kji}$$

where:

$$Z_{1kji^*} = \Phi(t_2 - (\hat{h}_k + \hat{\eta}_j + \hat{u}_i)); \; k = 1, 2; \; j = 1, 2; \; i = 1, \ldots, 4$$

Finally, the probability for each sire in category 3 of response per herd–year–sex subclass (Z_{3kji}) can be calculated as:

$$Z_{3kji} = 1 - Z_{2kji^*}$$

For Example 13.1, the probability distribution of heifer calvings for each sire across all herds and sexes in all categories are as follows:

	Probability in category of response		
	1	2	3
Sire 1	0.695	0.175	0.131
Sire 2	0.659	0.188	0.153
Sire 3	0.665	0.186	0.149
Sire 4	0.702	0.172	0.126

13.3 Joint Analysis of Quantitative and Binary Traits

Genetic improvement may be based on selecting animals on an index that combines both quantitative and categorical traits. Optimally, a joint analysis of the quantitative and categorical traits is required in the prediction of breeding values in such a selection scheme to adequately account for selection. A linear multivariate model might be used for such analysis. However, such an analysis suffers from the limitations associated with the use of a linear model for the analysis of discrete traits mentioned in Section 13.2. In addition, such a multivariate linear model will not properly account for the correlated effects of the quantitative traits on the discrete trait.

Foulley *et al.* (1983) presented a method of analysis to handle the joint analysis of quantitative and binary traits using a Bayesian approach. It involves fitting a linear model for the quantitative traits and a non-linear model for the binary trait. This section presents this methodology and illustrates its application to an example data set.

13.3.1 Data and model definition

Assume that a quantitative trait, such as birth weight, and a binary trait, such as calving difficulty (easy versus difficult calving), is being analysed. As in

Section 13.2.2, the data for calving difficulty could be represented in an s by 2 contingency table:

	Response category	
Row	Easy calving	Difficult calving
1	n_{11}	$n_{1.} - n_{11}$
2	n_{21}	$n_{2.} - n_{21}$
⋮	⋮	⋮
j	n_{j1}	$n_{j.} - n_{j1}$
⋮	⋮	⋮
s	n_{s1}	$n_{s.} - n_{s1}$

where the s rows refer to conditions affecting an individual or grouped records. Note that n_{i1} or $n_{i.} - n_{i1}$ in the above table can be null, as responses in the two categories are mutually exclusive, but $n_i \neq 0$.

Assume that a normal function has been used to describe the probability of response for calving ease. Let $\mathbf{y}_1$ be the vector for observations for the quantitative trait, such as birth weight, and $\mathbf{y}_2$ be the vector of the underlying variable for calving difficulty. The model for trait 1 would be:

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{e}_1 \tag{13.13}$$

and for the underlying variable for trait 2:

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}_2 \tag{13.14}$$

where $\boldsymbol{\beta}_1$ and $\mathbf{u}_1$ are vectors of fixed effect and sire solutions for trait 1, and $\mathbf{X}_1$ and $\mathbf{Z}_1$ are the usual incidence matrices. The matrices $\mathbf{X}_2$ and $\mathbf{Z}_2$ are incidence matrices for the liability. The matrix $\mathbf{Z}_2 = \mathbf{Z}_1$ and $\mathbf{X}_2 = \mathbf{X}_1\mathbf{H}$, where $\mathbf{H}$ is an identity matrix if all factors affecting the quantitative traits also affect the liability. However, if certain fixed effects affecting the quantitative trait have no effect on the liability, $\mathbf{H}$ is obtained by deleting the columns of an identity matrix of appropriate order corresponding to such effects. It is assumed that:

$$\text{var}\begin{pmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{pmatrix} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{pmatrix}$$
$$\text{var}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} = \mathbf{A} \otimes \mathbf{G} \tag{13.15}$$

where $\mathbf{G}$ is the genetic covariance matrix for both traits and $\mathbf{A}$ is the numerator relationship matrix.

Let $\boldsymbol{\theta}' = [\boldsymbol{\beta}_1, \boldsymbol{\tau}, \mathbf{u}_1, \boldsymbol{\nu}]$, the vector of location parameters in Eqns 13.13 and 13.14 to be estimated, where $\boldsymbol{\tau} = \boldsymbol{\beta}_2 - b\mathbf{H}\boldsymbol{\beta}_1$ and $\boldsymbol{\nu} = \mathbf{u}_2 - b\mathbf{u}_1$, where b is the residual regression coefficient of the underlying variate on the quantitative trait. The calculation of b is illustrated in the next section. Since the residual variance of liability is unity, the use of b is necessary to properly adjust the underlying variate for the effect of the residual covariance between both

traits. The use of b can be thought of as correcting calving difficulty for other 'risk' factors affecting calving and, in this example, the birth weight of the calf. Thus Eqn 13.15 may be written as:

$$\text{var}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2 - b\mathbf{u}_1\end{pmatrix} = \begin{pmatrix}\mathbf{u}_1\\ \nu\end{pmatrix} = \mathbf{A}\otimes\mathbf{G}_c$$

where:

$$\mathbf{G}c = \begin{pmatrix}\mathbf{I} & 0\\ -b\mathbf{I} & \mathbf{I}\end{pmatrix}\begin{pmatrix}g_{11} & gg_{12}\\ g_{21} & gg_{22}\end{pmatrix}\begin{pmatrix}\mathbf{I} & -b\mathbf{I}\\ 0 & \mathbf{I}\end{pmatrix} \tag{13.16}$$

with g_{ij} being the elements of **G.**

Using a Bayesian approach, Foulley *et al.* (1983) calculated the mode of the posterior density of $\boldsymbol{\theta}$ by equating the derivatives of the log-posterior density of $\boldsymbol{\theta}$ to zero. The resulting system of equations were not linear in $\boldsymbol{\theta}$. They set up the following iterative system of equations for $\boldsymbol{\theta}$ to be estimated:

$$\begin{pmatrix}\mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1 & 0 & 0\\ \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}\mathbf{g}_c^{11} & 0 & \mathbf{A}^{-1}\mathbf{g}_c^{12}\\ 0 & 0 & \mathbf{X}_2'\mathbf{W}^{[i-1]}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{W}^{[i-1]}\mathbf{Z}_2\\ 0 & \mathbf{A}^{-1}\mathbf{g}_c^{21} & \mathbf{Z}_2'\mathbf{W}^{[i-1]}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{W}^{[i-1]}\mathbf{Z}_2 + \mathbf{A}^{-1}\mathbf{g}_c^{22}\end{pmatrix}$$

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}^{[i]}\\ \hat{\mathbf{u}}^{[i]}\\ \boldsymbol{\Delta\tau}^{[i]}\\ \boldsymbol{\Delta\nu}^{[i]}\end{pmatrix} = \begin{pmatrix}\mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{y}_1\\ \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{y}_1\\ \mathbf{X}_2'\mathbf{q}^{[i-1]}\\ \mathbf{Z}_2'\mathbf{q}^{[i-1]}\end{pmatrix} - \begin{pmatrix}0\\ \mathbf{A}^{-1}\mathbf{g}_c^{12}\boldsymbol{\nu}^{[i-1]}\\ 0\\ \mathbf{A}^{-1}\mathbf{g}_c^{22}\boldsymbol{\nu}^{[i-1]}\end{pmatrix} \tag{13.17}$$

The matrices and vectors in Eqn 13.17 have been defined earlier, apart from $\mathbf{q}$ and $\mathbf{W}$. Initially, P_{jk}, the probability of response in category k, given the conditions in the jth row, is defined for the category trait. With only two categories of response for calving difficulty, then from Eqn 13.5:

$$P_{j1} = \Phi(t - a_j) \quad \text{and} \quad P_{j2} = 1 - P_{j1}$$

with a_j regarded as the mean of the liability in the jth row or as defined in Eqn 13.5.

However, with only one threshold, the value of t by itself is of no interest; the probability of response in the first category for the jth row can then be written as:

$$P_{j1} = \Phi(t - a_j) = \Phi(\mu_j)$$

where μ_j can be defined as the expectation of $\mathbf{y}_{2j}$ given $\boldsymbol{\beta}$, $\mathbf{u}$ and $\mathbf{y}_{1j}$, and this is worked out in the next section.

The vector $\mathbf{q}$ is of order s by 1 with elements:

$$\mathbf{q}_j = -\{n_{j1}d_{j1} + (n_{j.} - n_{j1})d_{j2}\}, \quad j = 1,\ldots, s \tag{13.18}$$

where $d_{j1} = -\phi(\mu_j)/P_{j1}$ and $d_{j2} = \phi\mu_j/(1 - P_{j1})$, with P_{j1} calculated as $\Phi(\mu_j)$.

$\mathbf{W}$ is an s by s diagonal matrix with the following elements:

$$w_{jj} = \mu_j q_j + n_{j1}d^2_{j1} + (n_{j.} - n_{j1})d^2_{j2}, \quad j = 1,\ldots, s \tag{13.19}$$

Calculating μ and the residual regression coefficient

From Eqn 13.14, the model for the *j*th row of the contingency table may be written as:

$$\mathbf{y}_{2j} = \mathbf{x}'_{2j}\boldsymbol{\beta}_2 + \mathbf{z}'_{2j}\mathbf{u}_2 + \mathbf{e}_{2j}$$

where $\mathbf{x}'_{2j}$ and $\mathbf{z}'_{2j}$ are vectors *j* of the $\mathbf{X}_2$ and $\mathbf{Z}_2$, respectively. Similarly, observations for trait 1, corresponding to the *j*th row of the contingency table, may be modelled as:

$$\mathbf{y}_{1j} = \mathbf{x}'_{1j}\boldsymbol{\beta}_1 + \mathbf{z}'_{1j}\mathbf{u}_1 + \mathbf{e}_{1j}$$

Let μ_j be the expectation of $\mathbf{y}_{2j}$ given $\boldsymbol{\beta}$, $\mathbf{u}$ and $\mathbf{y}_{1j}$. Thus:

$$\mu_j = \mathrm{E}(\mathbf{y}_{2j}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_{1j}) = \mathbf{x}'_{2j}\boldsymbol{\beta}_2 + \mathbf{z}'_{2j}\mathbf{u}_2 + \mathrm{E}(e_{2j}|e_{1j}) \tag{13.20}$$

given that e_{2j} is only correlated with e_{1j}. Assuming e_{2j} and e_{1j} are bivariately normally distributed:

$$\begin{aligned}\mathrm{E}(e_{2j} \mid e_{1j}) &= \frac{\sigma_{e(2,1)}}{\sigma^2_{e1}}(e_{1j}) \\ &= r_{12}\left(\frac{\sigma_{e2}}{\sigma_{e1}}\right)e_{1j}\end{aligned} \tag{13.21}$$

where σ^2_{ei} is the residual variance of trait *i*, $\sigma_{ei,k}$ and r_{ik} are the residual covariance and correlation between traits *i* and *k*, and σ_{ei} is the residual standard deviation of the *i*th trait. Similarly:

$$\mathrm{var}(\mathbf{y}_{2j}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_{1j}) = \mathrm{var}(e_{2j}|e_{1j}) = \sigma^2_{e2}(1 - r^2_{12})$$

Since the unit of the conditional distribution of the underlying trait, given $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{y}_{1j,}$ is the standard deviation, then from the above equation:

$$\sigma_{e2} = \frac{1}{\sqrt{(1-r^2_{12})}}$$

Therefore, Eqn 13.21 can be written as:

$$\mathrm{E}(e_{2j} \mid e_{1j}) = r_{12}\left(\frac{1}{\sigma_{e1}}\right)\frac{1}{\sqrt{1-r^2_{12}}}\, e_{1j} = be_{1j} \tag{13.22}$$

In general, Eqn 13.20 can be expressed as:

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + b\mathbf{e}_1 \\ &= \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + b(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{Z}_1\mathbf{u}_1)\end{aligned} \tag{13.23}$$

The above equation may be written as:

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{X}_2(\boldsymbol{\beta}_2 - b\mathbf{H}\boldsymbol{\beta}_1) + \mathbf{Z}_2(\mathbf{u}_2 - b\mathbf{u}_1) + b\mathbf{y}^*_1 \\ \boldsymbol{\mu} &= \mathbf{X}_2\boldsymbol{\tau} + \mathbf{Z}_2\boldsymbol{\nu} + b\mathbf{y}^*_1\end{aligned} \tag{13.24}$$

with the solutions of factors affecting calving difficulty corrected for the residual relationship between the two traits and $\mathbf{y}^*_1 = (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{Z}_1\mathbf{u}_1)$ or $\mathbf{y}^*_1$ may be calculated as:

$\mathbf{y}^*_1 = (\mathbf{y}_1 - \bar{y}_1)$, where $\bar{y}_1$ is the mean of $\mathbf{y}_1$.

13.3.2 Numerical application

Example 13.2

The bivariate analysis of a quantitative trait and a binary trait is illustrated using the data presented by Foulley *et al*. (1983) but with a sire–maternal grandsire relationship matrix included for the sires and pelvic opening omitted from the analysis. The data consisted of birth weight (BW) and calving difficulty (CD) on 47 Blonde d'Aquitaine heifers, with information on region of origin, sire of the heifer, calving season and sex of the calf included. Calving difficulty was summarized into two categories: easy or difficult calving. The data set is presented below:

Heifer origin	Sire	Season	Sex of calf	BW	CD[a]	Heifer origin	Sire	Season	Sex of calf	BWT	CD[a]
1	1	1	M	41.0	E	1	4	2	M	47.0	D
1	1	1	M	37.5	E	1	4	2	F	51.0	D
1	1	1	F	41.5	E	1	4	2	F	39.0	E
1	1	2	F	40.0	E	2	4	1	M	44.5	E
1	1	2	F	43.0	E	1	5	1	M	40.5	E
1	1	2	F	42.0	E	1	5	1	F	43.5	E
1	1	2	F	35.0	E	1	5	2	M	42.5	E
2	1	1	F	46.0	E	1	5	2	M	48.8	D
2	1	1	F	40.5	E	1	5	2	M	38.5	E
2	1	2	F	39.0	E	1	5	2	M	52.0	E
1	2	1	M	41.4	E	1	5	2	F	48.0	E
1	2	1	M	43.0	D	2	5	1	F	41.0	E
1	2	2	F	34.0	E	2	5	1	M	50.5	D
1	2	2	M	47.0	D	2	5	2	M	43.7	D
1	2	2	M	42.0	E	2	5	2	M	51.0	D
2	2	2	M	44.5	E	1	6	1	F	51.6	D
2	2	2	M	49.0	E	1	6	1	M	45.3	D
1	3	1	M	41.6	E	1	6	1	F	36.5	E
2	3	1	M	36.0	E	1	6	2	M	50.5	E
2	3	1	F	42.7	E	1	6	2	M	46.0	D
2	3	2	F	32.5	E	1	6	2	M	45.0	E
2	3	2	F	44.4	E	1	6	2	F	36.0	E
2	3	2	M	46.0	E	2	6	1	F	43.5	E
						2	6	1	F	36.5	E

[a]CD, calving difficulty; D, difficulty, E, easy.

A summary of the data, in terms of marginal means of calving variables by level of factors considered, is shown in the following table:

Factor		Number	Birth weight (kg)	Frequency CD[a]
Heifer origin	1	30	43.02	0.267
	2	17	43.02	0.176
Calving season	1	20	42.23	0.200
	2	27	43.61	0.259

Continued

(*Continued*)

Factor		Number	Birth weight (kg)	Frequency CD[a]
Sex of calf	M	25	–	0.360
	F	22	–	0.091
Sire of heifer	1	10	40.55	0.000
	2	7	42.99	0.286
	3	6	40.53	0.000
	4	4	45.38	0.500
	5	11	45.46	0.364
	6	9	43.43	0.333

[a]Frequency of calving difficulty.

The following sire–maternal grandsire relationship matrix was assumed among the sires:

Bull	Sire	Maternal grandsire
1	0	0
2	0	0
3	1	0
4	2	1
5	3	2
6	2	3

The inverse of the sire–maternal grandsire relationship matrix obtained for the above pedigree using the rules in Section 2.5 is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.424 & 0.182 & -0.667 & -0.364 & 0.000 & 0.000 \\ 0.182 & 1.818 & 0.364 & -0.727 & -0.364 & -0.727 \\ -0.667 & 0.364 & 1.788 & 0.000 & -0.727 & -0.364 \\ -0.364 & -0.727 & 0.000 & 1.455 & 0.000 & 0.000 \\ 0.000 & -0.364 & -0.727 & 0.000 & 1.455 & 0.000 \\ 0.000 & -0.727 & -0.364 & 0.000 & 0.000 & 1.455 \end{bmatrix}$$

The residual variance (σ^2_{e1}) for BW was assumed to be 20 kg^2 and the residual correlation (r_{12}) between BW and CD was assumed to be 0.459. Therefore, from Eqn 13.20, b equals 0.1155. The matrix **G** assumed was:

$$\mathbf{G} = \begin{pmatrix} 0.7178 & 0.1131 \\ 0.1131 & 0.0466 \end{pmatrix}$$

Therefore, from Eqn 13.16:

$$\mathbf{G}_C = \begin{pmatrix} 1 & 0 \\ -0.1155 & 1 \end{pmatrix} \begin{pmatrix} 0.7178 & 0.1131 \\ 0.1131 & 0.0466 \end{pmatrix} \begin{pmatrix} 1 & -0.1155 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.7178 & 0.0302 \\ 0.0302 & 0.0300 \end{pmatrix}$$

Thus the heritabilities for BW and CD are 0.14 and 0.18, respectively, with a genetic correlation of 0.62 between the two traits.

The model in Eqn 13.13 was used for the analysis of BW, thus $\boldsymbol{\beta}_1$ is the vector of solutions for origin of heifer, calving season and sex of calf and $\mathbf{u}_1$ is the vector of solutions for sire effects. The same effects were fitted for CD, with $\boldsymbol{\tau}$ being the vector of solutions for the fixed effects and $\boldsymbol{\nu}$ for the sire effects. Let $\boldsymbol{\theta}$ be as follows:

$$\boldsymbol{\beta}_1' = (d_1, d_2, s_1, s_2, f_1, f_2)$$

$$\mathbf{u}_1' = (\hat{u}_{11}, \hat{u}_{12}, \hat{u}_{13}, \hat{u}_{14}, \hat{u}_{15}, \hat{u}_{16})$$

$$\boldsymbol{\tau}' = (d_1', d_2', s_1', s_2', f_1', f_2')$$

$$\boldsymbol{\nu}' = (\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6)$$

where $d_i(d_i')$, $s_i(s_i')$ and $f_i(f_i')$ are level i of the effects of heifer origin, calving season and sex of calf, respectively; for BW (CD), $\hat{u}_{1j}$ and ν_j are the solutions for the sire j for BW and CD, respectively.

The matrix $\mathbf{X}_1$, which relates records for BW to the effects of heifer origin, calving season and sex of calf, can be set by principles already outlined in previous chapters. For the example data, all fixed effects affecting BW also affect CD; therefore, $\mathbf{H}$ is an identity matrix and $\mathbf{X}_2 = \mathbf{X}_1$. Similarly, the matrix $\mathbf{Z}_1 = \mathbf{Z}_2$. The remaining matrix in Eqn 13.17 can be obtained through matrix multiplication and addition.

Equation 13.17 needs starting values for $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$ to commence the iterative process. The starting values used were solutions ($\boldsymbol{\tau}^{(0)}$ and $\boldsymbol{\nu}^{(0)}$) from Eqn 13.17 with $\mathbf{W}^{[i-1]} = \mathbf{I}$, $\mathbf{q}^{[i-1]}$ = a vector of (0,1) variables (1, difficulty; 0, otherwise) and $\boldsymbol{\nu}^{[i-1]} = 0$. The solutions to Eqn 13.17 using these starting values are shown in Table 13.3, with equations for the second levels of calving season and sex of calf effects set to zero because of dependency in the systems of equations. Using these solutions, the calculation of $\mathbf{q}^{(0)}$ and $\mathbf{W}^{(0)}$ in the next round of iteration are illustrated for the first and last two animals in the example data.

First, $\boldsymbol{\mu}$ in Eqn 13.18 is calculated for these animals using Eqn 13.24.

For animals 1 and 2:

$$\mathbf{X}_2\boldsymbol{\tau} + \mathbf{Z}_2\boldsymbol{\nu} = (d_1' + \hat{s}_1' + f_1' + \hat{v}_1) = 0.1873 + -0.0874 + 0.2756 + (-0.1180) = 0.2575$$

Therefore, from Eqn 13.24, using the mean of birth weight, μ_1 is:

$$\mu_1 = 0.2575 + 0.1155(41 - 43.02) = 0.0242$$

and:

$$\mu_2 = 0.2575 + 0.1155(37.5 - 43.02) = -0.3800$$

For animals 46 and 47:

$$\mathbf{X}_2\boldsymbol{\tau} + \mathbf{Z}_2\boldsymbol{\nu} = (d_2' + \hat{s}_1' + f_2' + \hat{v}_6) = 0.1484 + -0.0874 + 0.0 + 0.0079 = 0.0690$$

Therefore, from Eqn 13.22:

$$\mu_{46} = 0.0690 + 0.1155(43.5 - 43.02) = 0.1244$$

and:

$$\mu_{47} = 0.0690 + 0.1155(36.5 - 43.02) = -0.6841$$

Table 13.3. Solutions to Example 13.2 using Eqn 13.17.

Trait[a]	Factor	Iteration number 0	1	4	8	13	Linear model
BW	*Heifer origin*						
	1	41.6633	41.5471	41.6262	41.6182	41.6195	41.6175
	2	42.2530	42.1409	42.2178	42.2099	42.2112	42.2022
	Calving season						
	1	–1.2350	–1.2345	–1.2346	–1.2343	–1.2344	–1.2387
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sex of calf						
	Male	3.1589	3.1890	3.1687	3.1690	3.1690	3.1845
	Female	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sire						
	1	–0.4155	–0.2633	–0.3671	–0.3580	–0.3595	–0.3268
	2	0.1048	0.1687	0.1246	0.1311	0.1300	0.1171
	3	–0.3315	–0.2280	–0.3007	–0.2939	–0.2950	–0.2641
	4	0.1364	0.3365	0.2035	0.2139	0.2122	0.1886
	5	0.2730	0.3261	0.2893	0.2979	0.2965	0.2688
	6	0.1545	0.2270	0.1770	0.1821	0.1813	0.1690
CD	*Heifer origin*						
	1	0.1873	–1.0189	–1.4072	–1.3915	–1.3943	0.1349
	2	0.1484	–1.2813	–1.7342	–1.7472	–1.7452	0.0876
	Calving season						
	1	–0.0874	0.1871	0.1327	0.1415	0.1401	–0.0311
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sex of calf						
	Male	0.2756	0.3218	0.8621	0.8369	0.8411	0.2410
	Female	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sire						
	1	–0.1180	0.0471	–0.0656	–0.0561	–0.0577	–0.0527
	2	0.0144	0.0705	0.0319	0.0379	0.0369	0.0285
	3	–0.0850	0.0185	–0.0546	–0.0477	–0.0488	–0.0427
	4	–0.0380	0.1698	0.0319	0.0424	0.0407	0.0350
	5	–0.0048	0.0362	0.0075	0.0163	0.0148	0.0195
	6	0.0079	0.0702	0.0270	0.0315	0.0308	0.0323

[a]BW, birth weight; CD, calving difficulty

Using Eqn 13.18, the elements of **q** for animals 1, 2, 46 and 47 are:

$$\mathbf{q}(1) = -\{0(-1)\phi(0.0242)/\Phi(0.0242) + (1-0)\phi(0.0242)/(1-\Phi(0.0242))\}$$
$$= -\{0(-1)0.3988/0.5097 + 1(0.3988/0.4903)\} = -0.8134$$

$$\mathbf{q}(2) = -\{0(-1)\phi(-0.3800)/\Phi(-0.3800) + (1-0)\phi(-0.3800)/(1-\Phi(-0.3800))\}$$
$$= -\{0(-1)0.3712/0.3520 + 1(0.3712/0.6480)\} = -0.5727$$

$$\mathbf{q}(46) = -\{0(-1)\phi(0.1244)/\Phi(0.1244) + (1-0)\phi(0.1244)/(1-\Phi(0.1244))\}$$
$$= -\{0(-1)0.3959/0.5495 + 1(0.3959/0.4505)\} = -0.8787$$

$$\mathbf{q}(47) = -\{0(-1)\phi(-0.6841)/\Phi(-0.6841) + (1-0)\phi(-0.6841)/(1-\Phi(-0.6841))\}$$
$$= -\{0(-1)0.3157/0.2470 + 1(0.3157/0.7530)\} = -0.4193$$

The diagonal elements of $\mathbf{W}$ for each of the four animals above can be calculated using Eqn 13.19 as:

$$w(1,1) = 0.0242(-0.8134)\{0(-1)[\phi(0.0242)/\Phi(0.0242)]^2 + (1-0)[\phi(0.0242)/(1-\Phi(0.0242))]^2\} = 0.6419$$

$$w(2,2) = -0.3800(-0.5727)\{0(-1)[\phi(-0.3800)/\Phi(-0.3800)]^2 + (1-0)[\phi(-0.3800)/(1-\Phi(-0.3800))]^2\} = 0.5458$$

$$w(46,46) = 0.1244(-0.8787)\{0(-1)[\phi(0.1244)/\Phi(0.1244)]^2 + (1-0)[\phi(0.1244)/(1-\Phi(0.1244))]^2\} = 0.6629$$

$$w(47,47) = -0.6841(-0.4193)\{0(-1)[\phi(-0.6841)/\Phi(-0.6841)]^2 + (1-0)[\phi(-0.6841)/(1-\Phi(-0.6841))]^2\} = 0.4626$$

The equations were solved iteratively and were said to have converged at the 15th round of iteration when $\Delta'\Delta/20 \leq 10^{-7}$, where $\Delta = \boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}$. Solutions at convergence at the 13th round of iteration and at some intermediate rounds are shown in Table 13.3. Results from an analysis using a linear model fitting the same effects with the **G** matrix and residual variances of 20 kg^2 for BW, 1.036 for CD and residual covariance of 2.089 between the two traits are also presented.

The results indicate that the probability of a difficult calving is higher for a male calf than for a female calf. Similarly, there is a slightly higher probability for calving difficulty for calving in the first season.

In general, sire rankings from the threshold and linear models were similar, except for sires 2 and 6 slightly changing rankings in the two models. The ranking of sires for calving difficulty based on the results from the threshold model could be based on $\hat{\mathbf{u}}_2 = \boldsymbol{\nu} + b_1\hat{\mathbf{u}}_1$ using the information provided by BW. However, the interest might be on ranking sires in terms of probability of calving difficulty, under a given set of conditions. For instance, what is the probability that a heifer sired by the *j*th bull born in region 2, calving a male calf in season 1, will experience a calving difficulty? This probability (V_{211j}) can be calculated as:

$$V_{211j} = \Phi[\hat{d}'_2 + \hat{s}'_1 + \hat{f}'_1 + \hat{v}_j + b_1(\hat{d}_2 + \hat{s}_1 + \hat{f}_1 - 43.02)] \quad (13.25)$$

Using the above equation, this probability for sire 1 is:

$$V_{211j} = \Phi[-1.7452 + 0.1401 + 0.8411 + (-0.0577) + 0.1155(42.2112 + (-1.2344) + 3.1690 - 43.02)] = 0.245$$

Similar calculations gave probabilities of 0.275, 0.247, 0.276, 0.268 and 0.273 for sires 2, 3, 4, 5 and 6, respectively. In general, there might be interest in the probability of difficult calving associated with using the *j*th sire across all regions of origin by season of calving and sex of calf subclasses. Such a probability can be calculated as:

$$V_{\ldots j} = \Sigma_{ikl}\lambda_{ikl}V_{iklj} \quad (13.26)$$

with V_{iklj} estimated as Eqn 13.25 and λ_{ikl} is an arbitrary weight such that $\Sigma_{ikl}\lambda_{ikl} = 1$. For the example data, λ can be set to be equal to $\frac{1}{8}$, as there are eight region–season–sex of calf subclasses. The probabilities obtained using Eqn 13.26 with $\lambda = \frac{1}{8}$ were 0.167, 0.188, 0.169, 0.189, 0.183 and 0.187 for sires 1, 2, 3, 4, 5 and 6, respectively.

The analysis of a binary trait with a quantitative trait has been discussed and illustrated in this section. However, if the category trait has several thresholds, then the method discussed in Section 13.2 would be used for the analysis of the categorical trait.

14 Survival Analysis

14.1 Introduction

Survival is one of the most important functional economic traits in livestock production, affecting profitability through the rate of replacement and farm production levels. In dairy cattle, the average herd life or survival of dairy cows has an economic value approximately half that of protein yield on a genetic standard deviation basis (Visscher *et al*., 1999). Consequently, most of the earlier research work on survival in terms of genetic evaluation and inclusion in breeding programmes has been in dairy cattle.

Various traits have been defined as the basis of evaluating survival in the dairy cow. These usually include some measure of survival for a period or length of life such as stayability until certain months of life defined as a binary trait (Everett *et al*., 1976), or in terms of the length of life or length of productive life (VanRaden and Klaaskate, 1993), or number of lactations (Brotherstone *et al*., 1997) or survival per lactation as a binary trait. Linear models are generally used – either a repeatability model (Madgwick and Goddard, 1989) or a multivariate model (Jairath *et al*., 1998). Similar definitions of survival have been applied to other livestock species. The length of productive life between first farrowing and culling has been analysed in pigs (Tarrés *et al*., 2006; Mészáros *et al*., 2010). In rabbits, survival has been defined as the length of productive life, referring to the days between date of the first positive pregnancy diagnosis and date of culling or death (Piles *et al*., 2006).

14.2 Functional Survival

Another important element of evaluating survival is the concept of functional survival or longevity. Functional longevity refers to survival that is independent of production such as milk yield for dairy cattle or litter size in pigs. The reasoning is that voluntary culling is based mostly on production, thus adjusting for production (usually at the phenotypic level) in the analysis of survival produces EBVs for animals that defines their ability to avoid involuntary culling.

14.3 Censoring

The traits used in survival analysis involve measuring the length of time between two events, usually a start and end point (also called 'failure'). However, at the time of analysis, some animals might still be alive, not having had the opportunity to reach the end point. Their measure of survival is based on their current status and does not

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

therefore reflect their true measure of survival. This phenomenon is referred to as censoring and such records are regarded as censored. There are several types of censoring. When records are based on current values that are less than the unknown end point, this is called right censoring. Left censoring can occur when, for instance, an animal has been alive for a certain time before entering the study or the start of data collection. Interval censoring can occur when there is a break in data collection and the cow fails somewhere in that interval. However, the most common is right censoring and this is the only type of censoring considered in this chapter.

14.4 Models for Analysis of Survival

14.4.1 Linear models

The linear models described in Chapters 3 or 5 have been used by various researchers for the analysis of survival traits, including those defined as a binary trait (Everett *et al*., 1976; Madgwick and Goddard, 1989; Jairath *et al*., 1998).

One of the major limitations with analysis of survival traits using a linear model is the inability or the difficulty of accounting for censoring. Various authors have attempted to address this problem. Brotherstone *et al*. (1997) introduced the concept of lifespan, which is the number of lactations a cow has survived or is expected to survive. Thus if p_n is the probability of survival to lactation $n + 1$ of an animal that has survived to complete lactation n, the expected lifespan (LS) of a cow that has completed n lactations but has not had time to complete $n + 1$ is:

$$LS = n + p_n + p_n^* p_{n+1} + p_n^* p_{n+1}^* p_{n+2} +$$

Thus if all p values above are constant and cows have completed their first lactation and have had no time restriction in the opportunity to express LS, then:

$$\text{Prob}(Ls = x) = (1 - p)P^{x-1} \quad \text{with} \quad x = (1, 2, 3 \ldots)$$

indicating that LS has a geometric distribution with mean = $1 + p/(1 - p)$ and variance = $p/(1 - p)^2$.

Similarly, VanRaden and Klaaskate (1993) evaluated survival using length of productive life, and censored records were predicted using phenotypic multiple regression. Madgwick and Goddard (1989) proposed a multi-trait model for the analysis of survival in each lactation, with observations in individual lactations treated as a different trait. Information on the current lactation of living cows can then be included as observed while their later (future) lactations are treated as missing records, hence accounting for all information.

While some of these linear models have included methods to predict expected survival for censored animals, these models are generally inadequate to handle time-dependent effects. Thus HYS effects, for instance, might be based on information from first calving, even for cows that have survived several lactations.

14.4.2 Random regression models for survival

Veerkamp *et al*. (1999) introduced the concept of fitting a random regression model (RRM) for the analysis of survival defined in terms of survival to the fourth lactation

as another approach to handle censored records in a linear model. In addition, time-dependent variables could be fitted with an RRM. The records in lactations 1 to 4 were coded as 1 if next lactation was present or 0 otherwise. For censored animals, current lactations were coded as described but later (future) lactations were regarded as missing. Thus for uncensored animals, there would be four observations and censored animals would have a number of observations equal to the current lactation at which they were censored. In addition to the fixed effects of HYS of calving, quadratic regressions for milk yield and age within herd, and a linear regression for Holstein percentage, they modelled the survival records of cows fitting a fixed cubic polynomial for lactation number and orthogonal polynomial of order 3 for additive animal genetic effects. It is not clear why a permanent environmental effect was not included in their model. They concluded that RRM could be considered as an alternative to a proportional hazard model in terms of handling time-dependent variables, but that the RRM was not very efficient at handling culling towards the end of lactation 4. This was attributed to lack of adequate data in the last lactation in the study. The same approach could be used to model survival defined in terms of days or months of productive life. The details of the methodology of fitting an RRM have been covered in Chapter 9, therefore only an outline is presented here.

Considering the data in Table 14.1 and assuming 60 months as the maximum length of productive life, the data can be analysed using an RRM considering herd and year–season–parity (YSP) as the only fixed (FIX) effects with the following model:

$$y_{tijk} = FIX_i + \sum_{k=0}^{nf} \phi_{jtk}\boldsymbol{\beta}_k + \sum_{k=0}^{nr} \phi_{jtk}\mathbf{u}_{jk} + \sum_{k=0}^{nr} \phi_{jtk}\mathbf{p}_{jk} + e_{tijk} \quad (14.1)$$

where y_{tijk} is the record for cow j, which is either 1 (alive) or 0 (dead) at time t (tth month of productive life) associated with the ith level of fixed effects (FIX_i); $\boldsymbol{\beta}_k$ are fixed regression coefficients; $\mathbf{u}_{jk}$ and $\mathbf{p}_{jk}$ are vectors of the kth random regression for animal and permanent environmental (pe) effects, respectively, for animal j; ϕ_{jtk} is the vector of the kth Legendre polynomial for the cow j at time t; nf is the order of polynomials

Table 14.1. Length of productive life (LPL) in months for some cows reared in two herds.

Cow	Sire	Dam	Herd	Parity	YSP	Code	LPL
8	1	2	1	2	3	0	40
9	1	3	1	2	4	1	47
10	4	2	1	1	1	0	22
11	4	9	1	1	2	1	28
12	5	3	1	2	3	1	50
13	5	8	1	1	1	1	33
14	1	6	2	2	4	1	49
15	1	7	2	1	1	1	29
16	5	14	2	1	2	0	23
17	5	6	2	2	3	1	37
18	4	7	2	2	4	0	35
19	4	3	2	1	2	1	30

YSP, year–season–parity. Code: 1, uncensored; 0, censored.

fitted as fixed regressions; *nr* is the order of polynomials for animal and pe effects; and e_{tijk} is the random residual. Note that for cows 8, 10, 16 and 18, which are censored, their records consist of the number of observations equal to their last month alive when they were censored. Thus cows 8 and 10 have 40 and 22 observations consisting of ones, respectively. However, for uncensored cows, each has 60 observations consisting of ones and zeros. Thus cow 9 has a record consisting of 47 ones and 13 zeros. The model in Eqn 14.1 can be fitted as described in Section 9.3.

14.4.3 Proportional hazard models

In view of the peculiarities associated with survival traits in terms of censoring of records and the presence of time-dependent covariates (i.e. whose values change with time), the proportional hazard model has been considered a more appropriate method of handling survival data. Its wide usage in the analysis of animal breeding data has been facilitated by the 'Survival kit' software by Ducrocq and Solkner (1998). A new version of the 'Survival kit', written in Fortran 90 with an R interface to make it user friendly, has recently been released (Mészáros *et al*., 2013). The new version offers the opportunity to account for the correlated nature of two random effects, either by specifying a known correlation coefficient or estimating it from the data. In addition to the computational complexities of the proportional hazard model, the other disadvantage of the method is the difficulty of applying it in a multi-trait situation with more than two traits. This is important as most direct measures of survival traits are obtained late in life; therefore, various traits, mostly linear or composite type traits such as fore-udder attachment, udder depth, mammary composite and legs and feet composite, have been used as indirect predictors of survival.

Subsequently in this section, it is assumed that censoring is random, such that the end time or censoring is independent for all individuals.

Defining some distributions

The basic idea is that survival time follows a distribution (for example, Fig. 14.1) and the goal is to use data to estimate the parameters of this distribution. Let T be the random continuous variable denoting the failure time (death) of an animal, then the survival function $S(t)$, which is the probability that the animal survives at least until time t, is:

$$S(t) = \Pr(T \geq t) = 1 - \Pr(T < t) = 1 - F(t)$$

where $F(t)$ is the cumulative distribution of T and $S(t)$ can be regarded as the proportion of animals still alive at time t.

One of the approaches for modelling the survival function is through the hazard function $h(t)$, which measures the risk of failure of an individual at time t. It specifies the instantaneous rate of failure at time t, given that the individual has survived up to time t. The usefulness of the $h(t)$ stems from the fact it can provide the failure rate over time even when the exact nature of the survival curve is not known. It can be denoted as:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

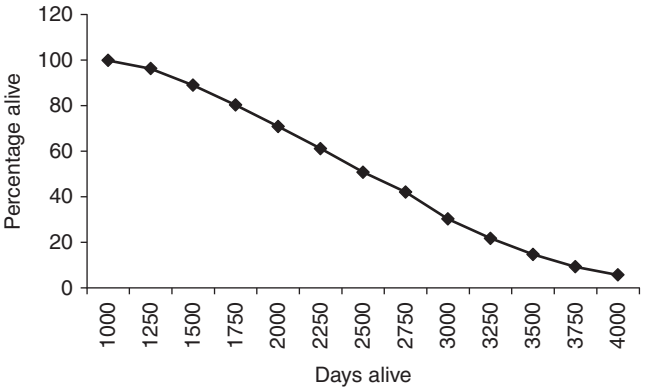


Fig. 14.1. Distribution of length of productive life for a group of Holstein dairy cows in the United Kingdom.

where $f(t)$ is the density function that equals $h(t)S(t)$. Another way of looking at the $h(t)$ is that for short periods of time (Δt), the probability that an animal fails is approximately equal to ($h(t)\Delta t$) (Kachman, 1999).

Exponential distribution

Several distributions can be used to define $h(t)$. If $h(t)$ is assumed to be constant over time then this is an exponential distribution. This implies that the chance of an animal surviving, for instance, an additional 2 years, is the same independent of how old the animal is. Assuming the exponential distribution, then $h(t) = \lambda$ and $S(t) = \exp(-\lambda t)$, where λ is the parameter of the exponential distribution.

Weibull distribution

The Weibull distribution, which is a two-parameter generalization of the exponential distribution, has also been used to model the hazard function to account for increasing or decreasing hazard function. With the Weibull distribution, $h(t)$ and $S(t)$ are:

$$h(t) = \rho\lambda(\lambda t)^{\rho-1} \quad \text{and} \quad S(t) = \exp(-(\lambda t)\rho)$$

with $\rho > 0$ and $\lambda > 0$. When $\rho = 1$, the Weibull distribution reduces to the exponential distribution. The Weibull distribution has a decreasing hazard function when $\rho < 1$ and an increasing hazard function when $\rho > 1$ (Fig. 14.2). Kachman (1999) showed that at a given λ, survival functions based on a Weibull model will all intersect at $t = 1/\lambda$, and that at $t = 1/\lambda$, the percentage survival is equal to $\exp(-1) \approx 37\%$. The role of the λ is to adjust the intercept.

Other possible distributions to model the hazard function include the gamma distribution, log-logistics and the log-normal distribution (Ducrocq, 1997). A summary of the commonly used distributions and parameters are given in Table 14.2.

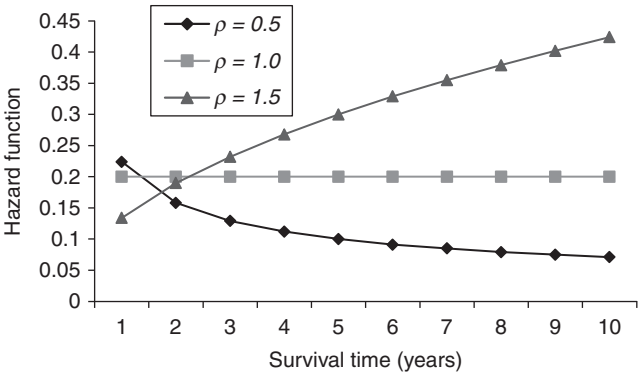


Fig. 14.2. The Weibull hazard function with a $\lambda = 0.20$ and with various ρ values.

Table 14.2. Some commonly used survival distributions and their parameters.

Distributions	$h(t)$	$S(t)$	$f(t)$
Exponential	λ	$\exp-(\lambda t)$	$\lambda\exp-(\lambda t)\rho$
Weibull	$\rho\lambda(\lambda t)^{\rho-1}$	$\exp-(\lambda t)\rho$	$\rho\lambda(\lambda t)^{\rho-1}\exp-(\lambda t)\rho$
Log-logistic	$\frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}$	$\frac{1}{1+\lambda t^{\rho}}$	$\frac{\lambda\rho t^{\rho-1}}{(1+\lambda t^{\rho})^2}$

14.4.4 Non-parametric estimation of the survival function

The survival function, $S(t)$, can be estimated from the parametric functions mentioned above. A non-parametric estimation of the survival function can be obtained using the Kaplan–Meier estimator (Kaplan and Meier, 1958). Let T_i represent failure times ordered from the first occurrence to the last. At T_i, let the number of animals that could have died (at risk) be denoted by n_i and the number that actually died as d_i. The Kaplan–Meier estimator then is:

$$\hat{S}(t) = \prod_{i|T_{i<t}} \left(\frac{n_i - d_i}{n_i} \right)$$

The usefulness of the Kaplan–Meier estimate of the survival function is that it could be used to check if the survival trait follows a particular parametric distribution. For instance, the appropriateness of a Weibull model can be evaluated by plotting $\log(-\log(\hat{S}(t)))$ versus $\log(t)$, where $\hat{S}(t)$ is the Kaplan–Meier estimate. This should result in a straight line with intercept $\rho\log(t)$ and slope ρ, given that:

$$S(t) = \exp(-(\lambda t)\rho) \rightarrow -\log(S(t)) = \lambda t^{\rho} \rightarrow \log(-\log(S(t))) = \log(\lambda) + \rho\log(t)$$

Similarly for the exponential distribution:

$$S(t) = \exp(-\lambda t) \rightarrow -\log(S(t)) = \lambda t$$

Therefore, the test for an exponential model will involve the plot of $-\log(\hat{S}(t))$ versus t, which should give straight line passing through the origin with slope λ.

14.4.5 Regression survival models

Initially, a fixed effects survival model is considered to introduce the concept. Assume that **x** is a vector of risk fixed effect factors or variables that influence failure time and **b** is the vector of corresponding solutions. One of the most popular procedures used to associate the hazard function $h(t)$ and **x** is the proportional hazard model (Cox, 1972; Ducrocq, 1997). The hazard function with vector of risk factors can be written as:

$$h(t; \mathbf{x}) = h_o(t)\exp(\mathbf{x}'\mathbf{b}) \tag{14.2}$$

where $h_o(t)$ is the baseline hazard function, representing the ageing process of the whole population. Thus the hazard function has been factored into two parts. First, the baseline hazard function ($h_o(t)$), which is independent of the risk factors, and hence the ratio of the hazard functions of two animals, is equal to a constant at any time, i.e. their hazard functions are proportional (Ducrocq, 1997). Second, the remaining part of the equation, $\exp(\mathbf{x}'\mathbf{b})$, can be regarded as the scalar that does not depend on time and denotes the specific risk associated with animals with the factors **x** and acts multiplicatively on the baseline hazard function.

When $h_o(t) = \lambda$ = a constant, then the baseline hazard is exponential. When the baseline hazard function is left completely arbitrary, then the proportion model is termed a Cox model (Cox, 1972).

With the Weibull model, the baseline hazard function can be derived as:

$$\begin{aligned} h(t; \mathbf{x}) &= \rho\lambda(\lambda t)^{\rho-1}\exp(\mathbf{x}'\mathbf{b}) \\ &= \rho t^{\rho-1}\exp(\rho\log(\lambda) + \mathbf{x}'\mathbf{b}) \\ &= h_o(t)\exp(\mathbf{x}'\mathbf{b}) \end{aligned} \tag{14.3}$$

where $h_o(t) = \rho t^{\rho-1}$ models the baseline hazard function and $\exp(\mathbf{x}'\mathbf{b})$, the scalar, models the relative risk above or below the baseline risk. Note that the $\mathbf{x}'\mathbf{b}$ in Eqn 14.3 includes the intercept term such that $\mathbf{x} = (\mathbf{1}, \mathbf{x}')$ and $\mathbf{b} = (\rho\log(\lambda), \mathbf{b})$.

The corresponding survival function (Kachman, 1999) is:

$$S(t; \mathbf{x}) = \exp\{-t^{\rho}\exp(\mathbf{x}'\mathbf{b})\}$$

Stratified proportional hazard model

At times, the assumption of a single baseline hazard function for the whole population in proportional hazard models may be inappropriate. Therefore, data may be divided into subclasses on the basis of factors such as year or season of birth, treatment or region. Then for individuals in a subclass c, a baseline hazard function can be fitted as:

$$h(t; \mathbf{x}, c) = h_{o,c}(t)\exp(\mathbf{x}'\mathbf{b})$$

Therefore, the hazards of two animals A1 and A2 in the same subclass with covariates $\mathbf{x}_{A1}$ and $\mathbf{x}_{A2}$, respectively, are proportional:

$$\frac{h(t; \mathbf{x}_{A1,}c)}{h(t; \mathbf{x}_{A2,}c)} = \exp[(\mathbf{x}'_{A1} - \mathbf{x}'_{A2})\mathbf{b}] = \text{constant}$$

and the baseline can have a known parametric form or be left arbitrary.

Accelerated failure time model

The accelerated failure time is another procedure to associate the hazard functions and the risk factors. Here it is assumed that the risk factors not only act multiplicatively but also accelerate or decelerate the failure time. Let $S_c(t)$ and $S_b(t)$ denote the survival functions of cows housed on concrete floors and floors with straw bedding, respectively. If it is assumed that the survival of cows housed on concrete floors is lower by a factor γ than those housed on floors with bedding, then the accelerated failure time model assumes that $S_b(t) = S_c(\gamma t)$ and $\gamma > 0$ is the so-called accelerating factor.

Thus Eqn 14.2 can be written as:

$$h(\mathrm{t};\, \mathbf{x}) = h_o(\exp[\mathbf{x}'\mathbf{b}]t)\exp(\mathbf{x}'\mathbf{b})$$
$$= h_o(t^*)\exp(\mathbf{x}'\mathbf{b})$$

where this change in timescale from t to t^* denotes an acceleration or a deceleration depending on whether $\exp(\mathbf{x}'\mathbf{b})$ is smaller or greater than unity (Ducrocq, 1997).

Time-dependent risk factors

In the analysis of survival data that span a good length of time, it is possible that some of the risk factors may change with time. In livestock situations, the effect of such factors such as year–season of calving or herd management effects are likely to change over time. Such factors are termed time-dependent variables or risk factors. The proportional hazard model can be extended to incorporate time-dependent variables and Eqn 14.2 can then be written as:

$$h(t;\, \mathbf{x}(t)) = h_o(t)\exp(\mathbf{x}(t)'\mathbf{b})$$

where as usual $\mathbf{x}(t)$ represents a vector of risk factors, but some of them will be time-dependent variables. Ducrocq (2000) showed that it is possible to define time-dependent variables such as HYS as a sequence of indicator variables $\mathbf{x}(t) = (0 \ldots 1 \ldots 0)$ with $x_i(t) = 1$ if the observation is affected by the ith HYS at time t, or $x_i(t) = 0$ otherwise.

14.4.6 Mixed survival models

Mixed survival models, usually called frailty models, refer to the extension of the proportional hazard function to include random effects such as genetic effects. The random or frailty term $\mathbf{u}_m$ is defined as an unobserved random quantity that acts multiplicatively on the hazard of individuals or a group of animals (Ducrocq, 1997). The random vector $\mathbf{u}_m$ can be defined for individual animals or daughters of a sire m as in a sire model. With the simple transformation $\mathbf{a}_m = \log(\mathbf{u}_m)$, the frailty term can be included in the exponential part of the proportional hazards. Thus the mixed survival model can be written as:

$$h(t;\, \mathbf{x},\mathbf{z}) = h_o(t)\exp(\mathbf{x}'\mathbf{b} + \mathbf{z}'\mathbf{a}) \quad \text{and}$$
$$S(t;\, \mathbf{x},\mathbf{z}) = \exp\{-t^{\rho}\exp(\mathbf{x}'\mathbf{b} + \mathbf{z}'\mathbf{a})\} \qquad (14.4)$$

where **z** is an incidence matrix for random effects and the baseline hazard function can assume a parametric or arbitrary form. Ducrocq *et al*. (1988a, 1988b) and Ducrocq (1997) discussed the various distributions (gamma or log-gammas or inverse Gaussian) that have been assumed for the frailty term and various estimation procedures for the parameters of the frailty model. In the following section, the parametric model presented by Kachman (1999) is used to illustrate the prediction of **a** in the frailty model.

The parameters of interest in a survival model with or without the frailty term can be estimated using non-parametric, semi-parametric or parametric approaches. In this section, a brief outline of the parametric approach is presented. The basic parametric approach involves obtaining the joint likelihood of the survival time and the random effects, getting the marginal likelihood of survival time by integrating over the random effects or taking a second-order Taylor's series expansion of the joint log-likelihood. The joint log-likelihood for the Weibull function can be written (Kachman, 1999) as:

$$L(\mathbf{b},\mathbf{u},\boldsymbol{\rho}) = \sum_i \{\log h_o(t_i) + (\mathbf{x}_i\mathbf{b} + z_i\mathbf{a}) - H_o(t_i)\exp(\mathbf{x}_i\mathbf{b} + z_i\mathbf{a})\} - 1/2\log|\mathbf{G}| - 1/2\,\mathbf{a}'\mathbf{G}\mathbf{a} \qquad (14.5)$$

The posterior mode estimates of the fixed and random effects can then be obtained by taking the first and second partial derivatives of Eqn 14.5. The resulting equations for the estimation are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}\mathbf{X} & \mathbf{X}'\mathbf{R}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}\mathbf{X} & \mathbf{Z}'\mathbf{R}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}^* \\ \mathbf{Z}'\mathbf{y}^* \end{bmatrix} \qquad [14.6]$$

where **R** is a diagonal matrix with elements $r_{ii} = w_i \times \exp(\mathbf{x}_i\mathbf{b} + \mathbf{z}_i\mathbf{a})$, with $w_i = \exp(\rho^* \log(t_i))$ and t_i is the survival record for animal i, and $y_i^* = q_i - r_{ii}\{1 - (\mathbf{x}_i\mathbf{b} + \mathbf{z}_i\mathbf{a})\}$, with $q_i = 1$ for uncensored records or 0 if records are censored.

The use of Eqn 14.6 involves an iterative procedure with $d_i = (\mathbf{x}_i\mathbf{b} + \mathbf{z}_i\mathbf{a})$ being initially computed for record or individual i, then r_{ii} and y_i^* are calculated assuming that the estimate ρ is known for the data. Then Eqn 14.6 can be set up. Once all records have been processed, estimates of $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ are obtained by solving Eqn 14.6. The new estimates of $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ are then fed into the iterative procedure again until convergence is achieved.

Example 14.1

Presented in Table 14.1 is the length of productive life in months for a group of cows in two herds. The aim is to undertake a survival analysis using Eqn 14.6, fitting herd and year–season–parity as fixed risk factors and random animal effects. It is assumed that ρ is 1 and the genetic variance is 20. The full pedigree is incorporated into the analysis.

Considering the fixed effects, the design matrix X is:

$$\mathbf{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Thus given **M** = [**X, Z**], where **Z** is a diagonal matrix considering the cows with records and:

$$\mathbf{u} = \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{pmatrix}$$

then vector **d** = **Mu.** For instance, if starting values for **u** in the first iteration were set to 0.1, then for the first animal, $d_1 = \mathbf{m}_1\mathbf{u} = 0.30$ with $\mathbf{m}_1$ = (1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0).

Then for animal *i*, compute r_{ii} and y_i^* and set up the row of equations for *i*th animal. For animal 1, r_{11} = exp(1*log(40))*exp(0.3) = 53.994 and y_1^* = 0 – 53.994 (1 – 0.3) = –37.760.

Equation 14.6 is built up and solved after all animals are processed and the iteration continued until convergence. Due to dependencies in the system of equations, the first levels of herd and year–season–parity effects have been constrained to zero. The solutions obtained at convergence and the risk ratios (RRS) are:

Herd					
	Solution	RRS			
1	0.000	1.000			
2	–1.631	0.196			
Year–season–parity					
	Solution	RRS			
1	0.000	1.000			
2	–2.346	0.094			
3	–3.149	0.043			
4	–2.982	0.051			
Animal	Solution	RRS	Animal	Solution	RRS
1	–0.779	0.459	11	–0.706	0.494
2	–1.233	0.291	12	–0.476	0.621
3	–0.062	0.940	13	–1.902	0.149
4	–0.750	0.472	14	–0.178	0.837
5	–0.758	0.469	15	–0.842	0.431
6	0.238	1.269	16	–0.519	0.595
7	–0.328	0.720	17	–0.115	0.891
8	–1.477	0.228	18	–0.578	0.561
9	–0.533	0.587	19	–0.290	0.748
10	–1.753	0.173			

The estimates b_i can be expressed in relative risk (hazard) ratio by the transformation RRS(b_i) = exp(b_i). This expression gives the RRS of culling due to that effect and it follows from the assumption of the proportional hazard model:

$$h_o(t; \mathbf{x}_a)/h_o(t; \mathbf{x}_c) = \exp((\mathbf{x}'_a - \mathbf{x}'_c)b)$$

implying that the relative hazard for two animals with covariates described by $\mathbf{x}_a$ and $\mathbf{x}_c$, respectively, is independent of time and of other covariates. Thus the RRS denotes the relative risk of a cow being culled in a certain fixed effect class compared to a cow in a reference class with risk set to unity. The estimates of RRS in Table 14.1 indicate, for instance, that cows in herd 2 are 20% more likely to be culled compared to herd 1. Also, cows in YSP subclasses 2 and 3 are 10% and 4% more likely to be culled compared to cows in YSP subclass 1. For the random animal effect, the RRS estimates

indicate the relative risk of daughters of these animals being culled. Usually these estimates are transformed to relative breeding values, say, with mean 100 and standard deviation of 12, so they are comparable with breeding values of other traits.

The results can also be presented in several other forms. The interest may be to predict the percentage of live daughters for sires at 40 months of productive life; for instance, (i) in herd 1 and in the 4th YSP or (ii) across all herds and YSPs. If (i), then for sire 1:

$$d_1 = \hat{b}_1 + \hat{b}_6 + \hat{a}_1 = 0.00 + -2.982 + -0.779 = -3.761$$

Then using Eqn 14.4, $S(40, \mathbf{x},\mathbf{z}) = \exp\{-40^{\rho}\exp(d_1)\} = 0.394$.
For (ii), a weighted mean for fixed effect solutions might be computed based on the number of daughters the sire has in each fixed effect subclass. Thus for sire 1:

$$d_1 = (2\hat{b}_1 + 2\hat{b}_2)/4 + (1\hat{b}_3 + 0\hat{b}_4 + 1\hat{b}_5 + 2\hat{b}_6)/4 + \hat{a}_1$$

$$= (2*0.0 + 2*0.196)/4 + (0.0 + 0 + -3.149 + 2*-2.982)/4 + -0.779 = -2.959$$

and $S(40, \mathbf{x},\mathbf{z}) = \exp\{-40^{\rho}\exp(d_1)\} = 0.126$

Equation 14.6 and its application in Example 14.1 was mainly to illustrate the basic principles of survival analysis using proportional hazard models with a frailty term. The parameter ρ has been assumed known and in practice this has to be estimated simultaneously, and usually more terms including time-dependent variables are included in the models. The 'Survival kit' (Ducrocq and Solkner, 1998; Mészáros *et al*., 2013) is currently used for the genetic evaluation of survival traits at the national level by a number of countries. A summary of methods utilized for the evaluation of survival at the national level for the Holstein breed on the Interbull website (http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm) indicates that eight countries (France, Germany, Italy, the Netherlands, Hungary, Slovenia, Spain and Switzerland) use proportional hazard models in their genetic evaluation systems. Similarly, nine countries (Canada, Denmark, Finland, Japan, New Zealand, Sweden and the UK) currently use a multi-trait animal model, while the USA, Israel and Australia employ a single-trait animal model. The only country that uses a random regression animal model is Belgium (Walloon region).

14.4.7 Group data survival model

When survival is defined as a discrete trait such as number of lactations completed or number of years completed, the Cox and Weibull models may not be suitable for the analysis of such traits. This is because these models assume continuity of the baseline hazard distribution and/or absence of ties between ordered failure times. Thus, with discrete survival traits, the grouped data version of the proportional hazards model introduced by Prentice and Gloeckler (1978) can be used. The group data proportional hazard model involves grouping failure time into intervals $Q_i = (q_{i-1}, q_i)$, $i = 1, \ldots, r$ with $q_0 = 0$, $q_r = +\text{infinite}$ and failure times in Q_i are recorded as t_i. Thus the regression vector is assumed to be time-dependent but fixed within each time interval. Grouped data models have been used in beef cattle (Phocas and Ducrocq, 2006) and rabbits (Piles *et al*., 2006). Mészáros *et al*. (2010) demonstrated this grouped data model was more appropriate in length of productive life in pigs.

15 Estimation of Genetic Parameters

ROBIN THOMPSON

Rothamsted Research, Harpenden, UK

15.1 Introduction

In order to carry out prediction of breeding values, estimates of variance components are usually needed. In this chapter the estimation of variance parameters is considered using univariate sire and animal models.

15.2 Univariate Sire Model

To motivate this work, the mixed effect sire model introduced in Chapter 3 is used. This model (Eqn 3.15) has:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e}$$

and:

$$\text{var}(\mathbf{s}) = \mathbf{A}\sigma_s^2$$

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_s^2 + \mathbf{R}$$

where $\mathbf{A}$ is the numerator relationship matrix for sires, $\sigma_s^2 = 0.25\sigma_a^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$. The aim is to estimate σ_s^2 and σ_e^2. The simplest case with this sire model is when $\mathbf{X}$ is a $n \times 1$ matrix with elements 1, $\mathbf{b}$ having one element representing an overall effect and the q sires being unrelated, so that $\mathbf{A} = \mathbf{I}$.

An analysis of variance can be constructed by fitting: (i) a model with the overall effect $\mathbf{b}$; and (ii) a model with sire effects, these models giving residual sums of squares that can be put into an analysis of variance of the form:

Source	Degrees of freedom	Sums of squares
Overall	Rank ($\mathbf{X}$) = 1	$\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ = F
Sires	Rank ($\mathbf{Z}$) – rank ($\mathbf{X}$) = $q - 1$	$\mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ = S
Residual	n – rank ($\mathbf{Z}$) = $n - q$	$\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ = R

Essentially, the effects $\mathbf{b}$ and $\mathbf{s}$ are thought of as fixed effects to construct an unweighted analysis. If estimates of σ_s^2 and σ_e^2 are required, then the sums of squares S and R can be equated to their expectation E(R) = $(n - q)\sigma_e^2$ and E(S) = $(q - 1)\sigma_e^2$ + trace($\mathbf{Z}'\mathbf{SZ}$)σ_s^2 where $\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

15.3 Numerical Example of Sire Model

Consider the data in Table 15.1 for the pre-weaning gain (WWG) of beef calves. The objective is to illustrate the estimation of variance components on a very small example so that the calculations can be expressed concisely.

The model to describe the observations is:

$$y_{ij} = o + s_j + e_i$$

where y_{ij} = the WWG of the ith calf, o = the overall effect, s_j = random effect of the jth sire (j = 1, 2, 3) and e_i = random error effect (i = 1, 2, 3, 4).

In matrix notation, the model is the same as described in Eqn 3.1, with n = 4, p = 1 and q = 3.

The matrix **X** in the MME relates records to the overall effects. For the example data set, its transpose is:

$$\mathbf{X}' = [1\ 1\ 1\ 1]$$

The matrix **Z** then relates records to sires. In this case it is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

An analysis of variance can be constructed as:

Source	Degrees of freedom	Sums of squares (kg^2)
Overall	1	F = 48.3085
Sire	2	S = 0.6075
Residual	1	R = 0.1800

with:

$$\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = F = (2.9 + 4 + 3.5 + 3.5)^2/4 = 48.3025$$

$$\mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = S = (4)^2/1 + (2.9 + 3.5)^2/2 + (3.5)^2/1 - 48.3085 = 0.6075$$

$$\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = R = (2.9)^2 + (4)^2 + (3.5)^2 + (3.5)^2 - F - S = 0.18$$

Table 15.1. Pre-weaning gain (kg) for four beef calves.

Calf	Sire	WWG (kg)
4	2	2.9
5	1	4.0
6	3	3.5
7	2	3.5

In this case:

$$\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix} \quad \text{and} \quad \mathbf{Z'SZ} = \begin{bmatrix} 0.75 & -0.50 & -0.25 \\ -0.50 & 1.00 & -0.25 \\ -0.25 & -0.50 & 0.75 \end{bmatrix}$$

so that:

$$E(R) = \sigma_e^2 = 0.18 \quad \text{and} \quad E(S) = 2\sigma_e^2 + 2.5\sigma_s^2 = 0.6075$$

Then estimates of σ_e^2 and σ_s^2 are:

$$\sigma_e^2 = 0.18 \text{ (kg}^2\text{)} \quad \text{and} \quad \sigma_s^2 = 0.027 \text{ (kg}^2\text{)}$$

15.4 Extended Model

The model and analysis hold if the model is extended to allow **X** to represent an environmental effect with p levels. If sires are nested within levels of the environmental factor so that daughters of each sire are only associated with one level of the environmental factor, then the above analysis could be used. If, however, as usually happens, daughters of a sire are associated with more than one level, a slightly more complicated analysis is appropriate.

Source	Degrees of freedom	Sums of squares
Fixed effects	Rank (**X**) = p	$\mathbf{y'X(X'X)^{-1}X'y} = F$
Sires corrected for fixed effects	Rank (**Z′SZ**) = df_S	$\mathbf{y'SZ(Z'SZ)^{-1}Z'Sy} = S$
Residual	n − rank (**X**) − rank (**Z′SZ**) = $n - p - df_S = df_R$	$\mathbf{y'y} - F - S = R$

Now R and S have expectation:

$$E(R) = df_R\sigma_e^2 \quad \text{and} \quad E(S) = df_S\sigma_e^2 + \text{trace}(\mathbf{Z'SZ})\sigma_s^2$$

The term involved in the trace (**Z′SZ**) can sometimes have a simple interpretation. If **X** represents a fixed effect matrix with p levels, then the ith diagonal element of **Z′SZ** is $n_{i.} - \Sigma n_{ij}^2/n_{.j}$ (summation is from $j = 1$ to p) where n_{ij} is the number of daughters of sire i in fixed effect level j and $n_{.j} = \Sigma n_{ij}$ (summation is from $j = 1$ to p) and $n_{i.} = \Sigma n_{ij}$ (summation is from $i = 1$ to s). This number was called the effective number of daughters of sire i by Robertson and Rendel (1954) and measures the loss of information on a sire because his daughters are measured in different environmental classes. This method

of analysis is called Henderson's method 3 (Henderson, 1953). These methods of analysis were very popular in that they related to sequential fitting of models and were relatively easy to compute. One problem is that the terms are generated under a fixed effect model with $\mathbf{V} = \mathbf{I}\sigma_e^2$ and then sums of squares are equated to their expectation under a different variance model. Only in special balanced cases will estimation based on R and S lead to efficient estimates of σ_s^2 and σ_e^2. In general, B is based on $\mathbf{Z}'\mathbf{S}\mathbf{y}$ with variance matrix $\mathbf{Z}'\mathbf{S}\mathbf{Z}\sigma_e^2 + \mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{S}\mathbf{Z}\sigma_s^2$ and these can be transformed to df_S independent values $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ by using arguments similar to those used in Section 6.2 on the canonical transformation, where $\mathbf{Q}$ is a df_S n matrix and $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{Q} = \mathbf{I}$ and $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{Q} = \mathbf{W}$, where $\mathbf{W}$ is a diagonal matrix of size df_S with ith diagonal element w_i. The variance matrix of $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ is then $\mathbf{I}\sigma_e^2 + \mathbf{W}\sigma_s^2$. Then an analysis of variance can be constructed from squaring each of the df_S elements of $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ with ith sum of squares u_i with expectation $\sigma_e^2 + w_i\sigma_s^2$ and R is the residual sum of squares with expectation $E(R) = df_R\sigma_e^2$. The individual u_i are distributed as chi-squared variables with variance $E(u_i)^2$. A natural scheme is to fit a linear model in σ_s^2 and σ_e^2 to u_i and R. One can also use an iterative scheme with the weight dependent on the estimated parameters.

15.5 Numerical Example

For the example with data in Table 15.1, it was shown that:

$$\mathbf{Z}'\mathbf{S}\mathbf{Z} = \begin{bmatrix} 0.75 & -0.50 & -0.25 \\ -0.50 & 1.00 & -0.25 \\ -0.25 & -0.50 & 0.75 \end{bmatrix}$$

so that the sires have 0.75, 1.0 and 0.75 effective daughters, respectively.

It can be found that with $\mathbf{A} = \mathbf{I}$:

$$\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{S}\mathbf{Z} = \begin{bmatrix} 0.875 & -0.750 & -0.125 \\ -0.750 & 1.500 & -0.750 \\ -0.125 & -0.750 & 0.875 \end{bmatrix}$$

The algorithm in Appendix E, Section E.1, can be used to calculate the eigenvalues $\mathbf{Q}$ so that:

$$\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{Q} = \mathbf{I} \quad \text{and} \quad \mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{Q} = \mathbf{W}$$

In this case:

$$\mathbf{Q} = \begin{bmatrix} -0.3333 & 0.6667 & -0.3333 \\ 0.7071 & 0.0000 & -0.7071 \end{bmatrix}$$

So $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{Q} = \mathbf{I}$ and $\mathbf{Q}'\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{S}\mathbf{Z}\mathbf{Q} = \mathbf{W}$ with:

$$\mathbf{W} = \begin{bmatrix} 1.5 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$$

The contrasts $\mathbf{Q'Z'Sy}$ are now:

$$\mathbf{Q'Z'S} = \begin{bmatrix} 0.5000 & -0.5000 & -0.5000 & 0.5000 \\ 0.0000 & 0.7071 & -0.7071 & 0.0000 \end{bmatrix} \mathbf{y}$$

So the first contrast, $(y_1 - y_2 - y_3 + y_4)/2 = (2.9 - 4.0 - 3.5 + 3.5)/2 = -1.1/2 = -0.55$ is a scaled contrast comparing sire 2 with sire 1 and sire 3 and the second contrast $(y_2 - y_3)/\sqrt{2} = (-4.0 = 3.5)/\sqrt{2} = (-0.5)/\sqrt{2}$ is a scaled contrast between sire 1 and sire 3.

An analysis of variance can be constructed:

Source	Degrees of freedom	Sums of squares (kg^2)	Expected mean squares (kg^2)
Overall	1	F = 48.3085	
Sire 2 compared with sires 1 and 3	1	$(-0.55)^2 = 0.3025$	$\sigma_e^2 + 1.5\sigma_s^2$
Sire 1 compared with sire 3	1	$(-0.5)^2/2 = 0.1250$	$\sigma_e^2 + \sigma_s^2$
Residual	1	R = 0.1800	σ_e^2

Fitting a linear model in σ_e^2 and σ_s^2 to the three sums of squares 0.3025, 0.1250 and 0.1800, gives estimates of $\sigma_e^2 = 0.143$ (kg^2) and $\sigma_s^2 = 0.079$ (kg^2). If a generalized linear model is fitted iteratively to the sum of squares with weights proportional to the variance of the sum of squares when the procedure converges, the estimate of σ_e^2 is 0.163 (kg^2) and of σ_s^2 is 0.047 (kg^2). The estimated variances of these estimates (from the inverse of the generalized least squares coefficient matrix) are 0.216 (kg^2) and 0.234 (kg^2).

15.6 Animal Model

It has been shown that estimates can be obtained from analysis of variance for some models. Now consider a more general model – the animal model introduced in Chapter 3. This linear model (Eqn 3.1) is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

and the variance structure is defined, with:

$$\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}; \quad \text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G} \quad \text{and} \quad \text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$$

where $\mathbf{A}$ is the numerator relationship matrix, and there is interest in estimating σ_a^2 and σ_e^2. A popular method of estimation is by restricted (or residual) maximum likelihood (REML) (Patterson and Thompson, 1971). This is based on a log-likelihood of the form:

$$L \propto (\tfrac{1}{2})\{-(\mathbf{y} - \mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb}) - \text{logdet}(\mathbf{V}) - \text{logdet}(\mathbf{X'V}^{-1}\mathbf{X})\}$$

where $\mathbf{b}$ is the generalized least squares (GLS) solution and satisfies:

$$\mathbf{X'V}^{-1}\mathbf{Xb} = \mathbf{X'V}^{-1}\mathbf{y}$$

There are three terms in L: the first is a weighted sum of squares of residuals; the second, a term that depends on the variance matrix; and a third that depends on the variance matrix of the fixed effects and can be thought of as a penalty because fixed effects are estimated. MME (Chapter 3) play an important part in the analysis process.

For the particular model these can be written as (Eqn 3.4):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

with $\alpha = \sigma_e^2/\sigma_a^2$ or $(1 - h^2)/h^2$.

Extensive use is made of the prediction error matrix of **a**. In this case the prediction error matrix is PEV = var(**a** − **â**) = $\mathbf{C}^{22}\sigma_e^2$ (Eqn 3.14), where $\mathbf{C}^{22}$ is associated with the coefficient matrix of the MME.

Estimates of σ_a^2 and σ_e^2 are chosen to maximize L. It is useful to express relevant terms in this estimation process in terms of the projection matrix **P**:

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$$

Then:

$$\mathrm{L}\ \alpha(\tfrac{1}{2})\{-\mathbf{y}'\mathbf{P}\mathbf{y} - \mathrm{logdet}(\mathbf{V}) - \mathrm{logdet}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\} \qquad (15.1)$$

Estimation of a variance parameter θ_i ($\theta_1 = \sigma_e^2$, $\theta_2 = \sigma_a^2$) involves setting to zero the first derivatives:

$$\partial \mathrm{L}/\partial\theta_i = (\tfrac{1}{2})\{\mathbf{y}'\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}\mathbf{y} - \mathrm{trace}[\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)]\}$$

These equations could be thought of as equating a function of data (the first term in the expression) to its expectation.

Normally, finding a maximum requires an iterative scheme. One suggested by Patterson and Thompson (1971) was based on using the expected value of the second differential matrix. In this case these are:

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j) = -(\tfrac{1}{2})\mathrm{trace}[\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}(\partial\mathbf{V}/\partial\theta_j)]$$

Using the first and expected second differentials one can update $\boldsymbol{\theta}$ using terms that depend on the solution of the MME and PEVs. For the particular animal model that is being considered, then:

$$\partial \mathrm{L}/\partial\sigma_e^2 = (\tfrac{1}{2})\{(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathrm{b} - \mathbf{Z}\mathbf{a})/\sigma_e^4 - (n - p - q)/\sigma_e^2 - \mathrm{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]/\sigma_a^2\} \qquad (15.2)$$

$$\partial \mathrm{L}/\partial\sigma_a^2 = (\tfrac{1}{2})\{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/\sigma_a^4 + q/\sigma_a^2 - \mathrm{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2/\sigma_a^4\} \qquad (15.3)$$

and:

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\sigma_e^4) = -(\tfrac{1}{2})\{(n - p - q)/\sigma_e^4 + \mathrm{trace}[(\mathbf{C}^{22}\mathbf{A}^{-1})^2]/\sigma_a^4\}$$

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\sigma_a^4) = -(\tfrac{1}{2})\{\mathrm{trace}[\{\mathbf{I} - \mathbf{C}^{22}\mathbf{A}^{-1}(\sigma_e^2/\sigma_a^2)\}^2]/\sigma_a^4\}$$

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\sigma_a^2\partial\sigma_e^2) = -(\tfrac{1}{2})\{\mathrm{trace}[\{\mathbf{I} - \mathbf{C}^{22}\mathbf{A}^{-1}(\sigma_e^2/\sigma_a^2)\}\{\mathbf{C}^{22}\mathbf{A}^{-1}\}]\sigma_a^4\}$$

Thinking of the variance parameters and the first differentials as vectors $\boldsymbol{\theta}$ and $\partial\mathbf{L}/\partial\boldsymbol{\theta}$ with ith (i = 1, 2) element θ_i and $\partial \mathrm{L}/\partial\theta_i$, respectively, and **Einf**, the expected information matrix, a matrix with i,jth element $-\mathrm{E}(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j)$, suggests an iterative scheme with the new estimate $\boldsymbol{\theta}_n$ satisfying:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta} + \mathbf{Einf}^{-1}(\partial \mathbf{L}/\partial \boldsymbol{\theta}) \tag{15.4}$$

There are two problems with this approach. First, the parameters might go negative and one would want estimates of variances to stay essentially positive. One popular way of avoiding this property is to note that at a maximum of the likelihood the first differentials are zero and to manipulate Eqns 15.1 and 15.2 in the form:

$$(n - p)\sigma_e^2 = (\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{y}) \tag{15.5}$$

$$q\sigma_a^2 = \mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2 \tag{15.6}$$

so that it can be seen that σ_e^2 is estimated from a sum of squares of residuals and σ_a^2 is estimated from a weighted sum of squares of predicted values and their PEV. This algorithm is an expectation maximization (EM) algorithm (Dempster *et al*., 1977) and successive iterates are positive. The algorithm can be written in the form of the updating formula if **Einf** is replaced by a matrix that depends on the information derived, as if one could directly observe the residuals and breeding values rather than predicting them. This algorithm can be slow to converge in animal breeding applications.

A second problem is that the expected second differentials are difficult to calculate. Sometimes it is recommended to use observed second differentials. These are of the form:

$$(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j) = -\mathbf{y}'\mathbf{P}(\partial \mathbf{V}/\partial\theta_i)\mathbf{P}(\partial \mathbf{V}/\partial\theta_j)\mathbf{Py} + (\tfrac{1}{2})\text{trace}[\mathbf{P}(\partial \mathbf{V}/\partial\theta_i)\mathbf{P}(\partial \mathbf{V}/\partial\theta_j)]$$

but again, these terms involve the complicated trace terms. One suggestion (Gilmour *et al*., 1995) is to use the average of the expected and observed information terms. These are of the form:

$$\mathrm{A}(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j) = -(\tfrac{1}{2})\{\mathbf{y}'\mathbf{P}(\partial \mathbf{V}/\partial\theta_i)\mathbf{P}(\partial \mathbf{V}/\partial\theta_j)\mathbf{Py}\}$$

These terms are similar to $\mathbf{y}'\mathbf{Py}$ in that they could be thought of as a weighted sum of squares matrix with **y** replaced by two columns $(\partial \mathbf{V}/\partial\theta_i)\mathbf{Py}$ $(i = 1, 2)$. In this particular case:

$$(\partial \mathbf{V}/\partial\sigma_e^2)\mathbf{Py} = (\mathbf{y} - \mathbf{Xb} - \mathbf{Za})/\sigma_e^2$$

and:

$$(\partial \mathbf{V}/\partial\sigma_a^2)\mathbf{Py} = \mathbf{Za}/\sigma_a^2$$

As in the formation of **Einf,** we can construct and base an iterative scheme on Eqn 15.3 and on **Ainf**, a matrix with elements $-\mathrm{A}(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j)$. Once the iterative scheme has converged, then the asymptotic variance matrix of $\boldsymbol{\theta}$ can be estimated from $\mathbf{Ainf}^{-1}$ or $\mathbf{Einf}^{-1}$. The animal model and estimation procedure introduced can easily be extended to deal with other models, just as prediction procedures can be developed for a variety of models. Software for estimating variance parameters using this average information algorithm is described by Jensen and Madsen (1997) and Gilmour *et al*. (2003).

15.7 Numerical Example

Consider the data in Table 15.2 for the pre-weaning gain (WWG) of beef calves. This is very similar to the data of Table 3.1, with the data changed to give positive variance estimates.

The model to describe the observations is:

$$y_{ijk} = p_i + a_j + e_{ijk}$$

where y_{ij} = the WWG of the jth calf of the ith sex, p_i = the effect of the ith sex, a_j = random effect of the jth calf and e_{ijk} = random error effect.

In matrix notation, the model is the same as described in Eqn 3.1.

Again, the objective is to illustrate the estimation of variance components σ_e^2 and σ_a^2 on a very small example so that the calculations can be expressed concisely.

In matrix notation, the model is the same as described in Eqn 3.1 with $n = 5$, $p = 2$ and $q = 8$, with the design matrices as given in Section 3.3. Now $\mathbf{y}' = [2.6, 0.1, 1.0, 3.0, 1.0]$ and, using initial estimates of $\sigma_e^2 = 0.4$ and $\sigma_a^2 = 0.2$, solutions to MME (Eqn 3.15) are:

Sex effects	
Male	2.144
Female	0.602
Animals	
1	0.117
2	–0.025
3	–0.222
4	–0.254
5	–0.135
6	0.032
7	0.219
8	–0.305

Then:

$$(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})' = [0.2022\ -0.3661\ 0.3661\ 0.6374\ -0.8395]$$

$$\mathbf{C}^{22}\sigma_e^2 = \begin{bmatrix} 0.1884 & 0.0028 & 0.0131 & 0.0878 & 0.0180 & 0.0883 & 0.0554 & 0.0537 \\ 0.0028 & 0.1968 & -0.0041 & 0.0082 & 0.0949 & 0.0981 & 0.0479 & 0.0443 \\ 0.0131 & -0.0041 & 0.1826 & 0.0193 & 0.0805 & 0.0090 & 0.0504 & 0.0871 \\ 0.0878 & 0.0082 & 0.0193 & 0.1711 & 0.0188 & 0.0510 & 0.0971 & 0.0493 \\ 0.0180 & 0.0949 & 0.0805 & 0.0188 & 0.1712 & 0.0679 & 0.0879 & 0.0712 \\ 0.0883 & 0.0981 & 0.0090 & 0.0510 & 0.0679 & 0.1769 & 0.0609 & 0.0877 \\ 0.0554 & 0.0479 & 0.0504 & 0.0971 & 0.0879 & 0.0609 & 0.1767 & 0.0672 \\ 0.0537 & 0.0443 & 0.0871 & 0.0493 & 0.0712 & 0.0877 & 0.0672 & 0.1689 \end{bmatrix}$$

$\mathbf{y}'\mathbf{Py} = 4.8193$, $\log\det(\mathbf{V}) = -2.6729$ and $\log\det(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = 2.6241$ so $L = -2.3852$ from Eqn 15.1.

Then Eqns 15.2 and 15.3 give:

$$\partial L/\partial\sigma_e^2 = (0.5)\{(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})/\sigma_e^4 - (n - p - q)/\sigma_e^2 - \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]/\sigma_a^2\}$$

Table 15.2. Pre-weaning gain (kg) for five beef calves.

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	–	2.6
5	Female	3	2	0.1
6	Female	1	2	1.0
7	Male	4	5	3.0
8	Male	3	6	1.0

$\partial L/\partial\sigma_e^2 = (0.5)\{8.8753 - (-12.5000) - 18.0733\} = 1.6510$

$\partial L/\partial\sigma_a^2 = (0.5)\{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/\sigma_a^4 - q/\sigma_a^2 + \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2/\sigma_a^4\}$

$\partial L/\partial\sigma_a^2 = (0.5)\{6.3461 - 40.0000 + 36.1466\} = 1.2464$

and:

$A(\partial L^2/\partial\sigma_e^4) = -(0.5)\{(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'\mathbf{P}(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})/\sigma_e^4\}$

$A(\partial L^2/\partial\sigma_e^4) = -(0.5)16.5346 = -8.2673$

$A(\partial L^2/\partial\sigma_a^4) = -(0.5)\{\mathbf{a}'\mathbf{Z}'\mathbf{PZa}\}/\sigma_a^4$

$A(\partial L^2/\partial\sigma_a^4) = -(0.5)9.1163 = -4.5582$

$A(\partial L^2/\partial\sigma_a^2\partial\sigma_e^2) = -(0.5)\{\mathbf{a}'\mathbf{Z}'\mathbf{P}(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})\}/(\sigma_e^2\sigma_a^4)$

$A(\partial L^2/\partial\sigma_a^2\partial\sigma_e^2) = -(0.5)11.3070 = -5.6535$

and:

$$\mathbf{Ainf} = \begin{bmatrix} 8.2673 & 5.6535 \\ 5.6535 & 4.5582 \end{bmatrix} \quad \text{so } \mathbf{Ainf}^{-1} = \begin{bmatrix} 0.7967 & -0.9882 \\ -0.9882 & 1.4450 \end{bmatrix}$$

Using Eqn 15.4 and replacing **Einf** by **Ainf**:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta} + \mathbf{Ainf}^{-1}(\partial L/\partial\boldsymbol{\theta}) = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.7967 & -0.9882 \\ -0.9882 & 1.4450 \end{bmatrix}\begin{bmatrix} 1.6510 \\ 1.2464 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.0838 \\ 0.1695 \end{bmatrix}$$

so that new estimates of σ_e^2 and σ_a^2 are 0.4838 (kg^2) and 0.3695 (kg^2), respectively.

Table 15.3 gives six successive iterates and log-likelihood for this data.

In the last iteration:

$$\mathbf{Ainf}^{-1} = \begin{bmatrix} 2.4436 & -3.2532 \\ -3.2532 & 5.3481 \end{bmatrix}$$

so that the estimate of σ_e^2 is 0.4835 with standard error $\sqrt{2}.4436 = 1.563$ and the estimate of σ_a^2 is 0.5514 with standard error $\sqrt{5}.3481 = 2.313$.

By contrast, if estimates of $\sigma_e^2 = 0.4$ and $\sigma_a^2 = 0.2$ are used in conjunction with Eqns 15.5 and 15.6 then: $(n - p)\sigma_e^2 = (\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{y})$ so $3\sigma_e^2 = 1.9277$ so $\sigma_e^2 = 0.6426$ (kg^2) and $q\sigma_a^2 = \mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2$ so $8\sigma_a^2 = 0.2538 + 1.4458$ so $\sigma_a^2 = 0.2125$ (kg^2) with L = –2.3852. After 1000 iterations, the algorithm gives $\sigma_e^2 = 0.4842$ (kg^2) and $\sigma_a^2 = 0.5504$ (kg^2) with L = –2.1817, showing that this algorithm is slower to converge.

Table 15.3. Estimates of σ_e^2 and σ_a^2 and **L**.

Iterate	σ_e^2 (kg^2)	σ_a^2 (kg^2)	**L**
1	0.4000	0.2000	–2.3852
2	0.4838	0.3695	–2.2021
3	0.4910	0.5126	–2.1821
4	0.4839	0.5500	–2.1817
5	0.4835	0.5514	–2.1817
6	0.4835	0.5514	–2.1817

16 Use of Gibbs Sampling in Variance Component Estimation and Breeding Value Prediction

16.1 Introduction

Gibbs sampling is a numerical integration method and is one of several Markov chain Monte Carlo (MCMC) methods. They involve drawing samples from specified distributions; hence they are called Monte Carlo and are referred to as Markov chain because each sample depends on the previous sample. Specifically, Gibbs sampling involves generating random drawings from marginal posterior distributions through iteratively sampling from the conditional posterior distributions. For instance, given that $\mathbf{Q}' = (\mathbf{Q}_1, \mathbf{Q}_2)$ and $P(\mathbf{Q}_1, \mathbf{Q}_2)$ is the joint distribution of $\mathbf{Q}_1$ and $\mathbf{Q}_2$, Gibbs sampling involves sampling from the full conditional posterior distributions of $\mathbf{Q}_1$, $P(\mathbf{Q}_1|\mathbf{Q}_2)$ and $\mathbf{Q}_2$, $P(\mathbf{Q}_2|\mathbf{Q}_1)$.

Thus given that the joint posterior distribution is known to proportionality, the conditional distributions can be generated. However, defining the joint density involves the use of Bayes' thereom. In general, given that the probability of two events occurring together, $P(B, Y)$, is:

$$P(B,Y) = P(B)P(Y|B) = P(Y)P(B|Y)$$

then:

$$P(B|Y) = P(B)P(Y|B)/P(Y) \tag{16.1}$$

Equation 16.1 implies that inference about the variable B depends on the prior probability of its occurrence, $P(B)$. Given that observations on Y are available, this prior probability is then updated to obtain the posterior probability or density of B, $(P(B|Y)$. Equation 16.1 is commonly expressed as:

$$P(B|Y) \propto P(B)P(Y|B) \tag{16.2}$$

as the denominator is not a function of B. Therefore, the posterior density of B is proportional to the prior probability of B times the conditional distribution of Y given B. Assuming that B in Eqn 16.2 is replaced by $\mathbf{W}$, a vector of parameters, such that $\mathbf{W}' = (W_1, W_2, W_3)$, and that the joint posterior distribution is known to proportionality (Eqn 16.2), the full conditional probabilities needed for the Gibbs sampler can be generated for each parameter as $P(W_1|W_2, W_3, Y)$, $P(W_2|W_1, W_3, Y)$ and $P(W_3|W_1, W_2, Y)$. Assuming starting values $W_1^{[0]}$, $W_2^{[0]}$ and $W_3^{[0]}$, the implementation of the Gibbs sampler involves iterating the following loop:

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

1. Sample $W_1^{[i+1]}$ from $P(W_1|W_2^{[i]}, W_3^{[i]}, Y)$
2. Sample $W_2^{[i+1]}$ from $P(W_2|W_1^{[i+1]}, W_3^{[i]}, Y)$
3. Sample $W_3^{[i+1]}$ from $P(W_3|W_2^{[i+1]}, W_3^{[i+1]}, Y)$

Usually, the initial samples are discarded (the so-called burn-in period). In summary, the application of the Gibbs sampler involves defining the prior distributions and the joint posterior density and generating the full conditional posterior distributions and sampling from the latter.

The Gibbs sampler was first implemented by Geman and Geman (1984). In animal breeding, Wang *et al*. (1993, 1994) used Gibbs sampling for variance component estimation in sire and animal models. It has been implemented for the study of covariance components in models with maternal effects (Jensen *et al*., 1994), in threshold models (Sorensen *et al*., 1995) and in random regression models (Jamrozik and Schaeffer, 1997). It has recently been employed for the purposes of variance component estimation and breeding value prediction in linear threshold models (Heringstad *et al*., 2002; Wang *et al*., 2002). Detailed presentations of the Gibbs sampling within the general framework of Bayesian inference and its application for variance components estimation under several models have been published by Sorensen and Gianola (2002). In this chapter, the application of the Gibbs sampler for variance component estimation and prediction of breeding values with univariate and multivariate animal models are presented and illustrated.

16.2 Univariate Animal Model

Consider the following univariate linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where terms are as defined in Eqn 3.1 but with $\mathbf{u} = \mathbf{a}$ in Eqn 3.1. The conditional distribution that generates the data, **y**, is:

$$\mathbf{y}|\mathbf{b}, \mathbf{u}, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Zu} + \mathbf{R}\sigma_e^2) \quad (16.3)$$

16.2.1 Prior distributions

Prior distributions of **b**, **u**, σ_u^2 and σ_e^2 are needed to complete the Bayesian specification of the model (Wang *et al*., 1993). Usually, a flat prior distribution is assigned to **b**. Thus:

$$P(\mathbf{b}) \sim \text{constant} \quad (16.4)$$

This represents an improper or 'flat' prior distribution, denoting lack of prior knowledge about this vector. However, if there is information a priori about value of **b** in terms of upper or lower limits, this can be incorporated in defining the posterior distribution of **b**. Such a prior distribution will be called a proper prior distribution.

Assuming an infinitesimal model, the distribution of **u** is multivariate normal and is:

$$\mathbf{u}|\mathbf{A}, \sigma_u^2 \sim N(O, \mathbf{A}\sigma_u^2) \quad (16.5)$$

A scaled inverted chi-squared distribution (χ^2) is usually used as priors for the variance components (Wang *et al*., 1993). Thus for the residual variance:

$$P(\sigma_e^2 \mid \upsilon_e, s_e^2) \propto (\sigma_e^2)^{-\left(\frac{\upsilon_e}{2}+1\right)} \exp\left(-\frac{\upsilon_e s_e^2}{2\sigma_e^2}\right) \tag{16.6}$$

and the additive genetic variance:

$$P(\sigma_u^2 \mid \upsilon_u, s_u^2) \propto (\sigma_u^2)^{-\left(\frac{\upsilon_u}{2}+1\right)} \exp\left(-\frac{\upsilon_u s_u^2}{2\sigma_u^2}\right) \tag{16.7}$$

where υ_e (υ_u) is a 'degree of belief' parameter and s_e^2 (s_u^2) can be interpreted as a prior value of the appropriate variance component. Alternatively, prior uniform distribution could be assigned to the variance components such that:

$$P(\sigma_j^2) \propto \text{constant} \tag{16.8}$$

where $\sigma_j^2 = \sigma_e^2$ or σ_u^2 and an upper limit might be assigned for σ_j^2 based on prior knowledge. Setting υ_e or υ_u to –2 and s_e^2 or s_a^2 to 0 in Eqns 16.6 or 16.7 gives Eqn 16.8.

16.2.2 Joint and full conditional distributions

The joint posterior distribution of the parameters (**b**, **u**, σ_e^2 or σ_u^2) is proportional to the product of the likelihood function and the joint prior distribution. Using Eqns 16.3 to 16.7, the joint posterior distribution can be written as:

$$P(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2 \mid \mathbf{y}) \propto (\sigma_e^2)^{-\left(\frac{n+\upsilon_e}{2}+1\right)} \exp\left[-\frac{(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu}) + \upsilon_e s_e^2}{2\sigma_e^2}\right]$$
$$(\sigma_u^2)^{-\left(\frac{m+\upsilon_u}{2}+1\right)} \exp\left(\frac{(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + \upsilon_u s_u^2}{2\sigma_u^2}\right) \tag{16.9}$$

assuming n observations and m animals. Setting υ_e or υ_a and s_e^2 or s_a^2 to zero gives the joint posterior distributions for the uniform distribution in Eqn 16.8.

The full conditional posterior distribution of each parameter is obtained by regarding all other parameters in Eqn 16.9 as known. Thus for **b**:

$$P(\mathbf{b} \mid \mathbf{u}, \sigma_u^2, \sigma_e^2, \mathbf{y}) \propto \exp\left(-\frac{(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})}{2\sigma_e^2}\right) \tag{16.10}$$

A corresponding distribution to the above is:

$$\mathbf{Xb}|\mathbf{u},\ \sigma_u^2,\ \sigma_e^2,\ \mathbf{y} \sim N(\mathbf{y} - \mathbf{Zu},\ \mathbf{I}\sigma_e^2)$$

or:

$$\mathbf{X'Xb}|\mathbf{u},\ \sigma_u^2,\ \sigma_e^2,\ \mathbf{y} \sim N(\mathbf{X'}(\mathbf{y} - \mathbf{Zu}),\ \mathbf{X'X}\sigma_e^2)$$

Therefore:

$$\mathbf{b} \mid \mathbf{u},\ \sigma_u^2,\ \sigma_e^2,\ \mathbf{y} \sim N(\hat{\mathbf{b}},\ (\mathbf{X'X})^{-1}\sigma_e^2)$$

where:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\mathbf{u})$$

Thus for the *j*th level of **b**:

$$\hat{b}_j | \mathbf{b}_{-j}, \mathbf{u}, \sigma_u^2, \sigma_e^2, \mathbf{y} \sim \mathbf{N}(\hat{b}_j, (\mathbf{x}_j'\mathbf{x}_j)^{-1}\sigma_e^2) \tag{16.11}$$

with $\hat{b}_j = (\mathbf{x}_j'\mathbf{x}_j)^{-1}\mathbf{x}_j'(\mathbf{y}_j - \mathbf{X}_{-j}\mathbf{b} - \mathbf{Z}\mathbf{u})$, which is equivalent to Eqn 3.5, $\mathbf{x}_j$ is the jth row of **X** and $\mathbf{b}_{-j}$ is the vector **b** with level *j* deleted.

Similarly, the distribution for the *j*th random effect is:

$$\mathbf{u}_j | \mathbf{b}, \mathbf{u}_{-j}, \sigma_u^2, \sigma_e^2, \mathbf{y} \sim \mathrm{N}(\hat{\mathbf{u}}_j, (\mathbf{z}_j'\mathbf{z}_j + \mathbf{A}^{-1}_{j,j}\alpha)^{-1}\sigma_e^2) \tag{16.12}$$

with:

$$\hat{\mathbf{u}}_j = (\mathbf{z}_j'\mathbf{z}_j + \mathbf{A}^{-1}_{j,j}\alpha)^{-1}\mathbf{z}_j'(y - \mathbf{X}\mathbf{b} - \mathbf{A}^{-1}_{j,-j}, \alpha\mathbf{u}_{-j})$$

which is equivalent to Eqn 3.8.

The full conditional of distribution of the residual variance is derived from Eqn 16.9 by considering only terms that involve σ_e^2 and is in the scaled inverted χ^2 form (Wang *et al*., 1993). Thus for the residual variance:

$$P(\sigma_e^2 \mid \mathbf{b}, \mathbf{u}, \sigma_u^2, \mathbf{y}) \propto (\sigma_e^2)^{-\left(\frac{n+\nu_e}{2}+1\right)} \exp\left(-\frac{\ddot{\nu}_e \ddot{s}_e^2}{2\sigma_e^2}\right)$$

where $\ddot{\nu}_e = n + \nu_e$ and $\ddot{s}_e^2 = ((\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u})) + \nu_e s_e^2) / \ddot{\nu}_e$
Hence:

$$\sigma_e^2 \mid \mathbf{b}, \mathbf{u}, \sigma_u^2, \mathbf{y} \sim \ddot{\nu}_e \ddot{s}_e^2 \chi^{-2}_{\ddot{\nu}_e} \tag{16.13}$$

which involve sampling from an inverted χ^2 distribution with scale parameter, $(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u}) + \nu_e s_e^2$ and $\ddot{\mathbf{v}}_e$ degrees of freedom.

Similarly, the full conditional distribution of σ_u^2 is also in the form of an inverted chi-square. Thus:

$$P(\sigma_u^2 \mid \mathbf{b}, \mathbf{u}, \sigma_e^2, \mathbf{y}) \propto (\sigma_u^2)^{-\left(\frac{m+\nu_u}{2}+1\right)} \exp\left(-\frac{\ddot{\nu}_u \ddot{s}_u^2}{2\sigma_u^2}\right)$$

where $\ddot{\nu}_u = m + \nu_u$ and $\ddot{s}_u^2 = ((\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}) + \nu_u s_u^2) / \ddot{\nu}_u$
Thus:

$$\sigma_u^2 \mid \mathbf{b}, \mathbf{u}, \sigma_u^2, \mathbf{y} \sim \ddot{\nu}_u \ddot{s}_u^2 \chi^{-2}_{\ddot{\nu}_u} \tag{16.14}$$

which involves sampling from an inverted χ^2 distribution with scale parameter $(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}) + \nu_u s_u^2$ and $\ddot{\mathbf{v}}_u$ degrees of freedom.

The Gibbs sampling then consists of setting initial values for **b**, **u**, σ_u^2 and σ_e^2 and iteratively sampling successively from Eqns 16.11 to 16.14, using updated values of the parameters from the *i* round in the *i* + 1 round. Assuming that *k* rounds of iteration were performed, then *k* is called the length of the chain. As mentioned earlier, the first *j* samples are usually discarded as the burn-in period. This is to ensure that samples saved are not influenced by the priors but are drawn from the posterior

distribution. The size of j is determined rather arbitrarily, but a graphical illustration could help.

Several strategies can be implemented in using the Gibbs sampler and these have an effect on the degree of correlation between the values sampled. Details of various strategies are discussed in detail by Sorensen and Gianola (2002) and therefore not presented here. One approach is to run a single long chain. A sample ($\mathbf{b}$, $\mathbf{u}$, σ_u^2, σ_e^2) is saved at every dth iterate until a total t samples are saved and analysed. The larger d is, the lower the degree of autocorrelation between the samples. Another strategy, known as the multiple chain or short chain approach, involves carrying out several parallel t runs and saving the last nth sample from each run. Thus this approach produces $m = nt$ samples. The different chains will produce different realizations, even if the same starting values are used. However, if the parameters in the model are highly correlated, it might be useful to utilize different starting values in the different chains.

Determining convergence with the Gibbs sampler is not very straightforward, but it is advisable, depending on the size of the problem, to run several chains and check convergence graphically.

16.2.3 Inferences from the Gibbs sampling output

The samples saved are usually analysed to estimate posterior means or variances of the posterior distribution. Detailed discussion of various estimation methods is given in Sorensen and Gianola (2002) and not presented here. Given that $\mathbf{w}$ is a vector of size k, containing the saved samples, then the posterior mean and variance can be computed, respectively, as:

$$\mu_f = \frac{\sum_{i=1}^{k} f(w_i)}{k} \tag{16.15}$$

and:

$$\text{var}(\mu_f) = \frac{\sum_{i=1}^{k} (f(w_i) - \mu_f)^2}{k}$$

where $f(\mathbf{w})$ is a function of interest of the variables in $\mathbf{w}$. For instance, in the linear animal model in Section 6.2, the function of interest would be the variance components (σ_u^2 and σ_e^2) and the vectors $\mathbf{b}$ and $\mathbf{u}$.

The above estimates from the posterior distribution are associated with sampling variance (Monte Carlo variance). The larger the number of samples analysed, the smaller the sampling variance. It is usually useful to get an estimate of the sampling variance associated with the estimates from the posterior distributions. An empirical estimate could be obtained by running several independent runs and then computing the between-chain variance of the estimates obtained for each run. This is not computationally feasible in most practical situations and various methods are used to estimate this variance. A number of such estimators are fully discussed by Sorensen and Gianola (2002). A simple method that could be used involves calculating the

batch effective chain size. Given a chain of size k, successive samples are grouped into b batches, each of size t. The average of the jth batch can be computed as:

$$\bar{u}_j = \frac{\sum_{i=1}^{t} f(w_i)}{t}$$

The batch estimator of the variance of μ in Eqn 16.14 is:

$$\text{var}_b(\mu) = \frac{\sum_{j=1}^{b} (\bar{u}_j - \mu)^2}{b(b-1)}$$

The batch effective chain size can be obtained as:

$$\psi_b = \frac{\sum_{i=1}^{k} [f(w_i) - \mu]^2}{(k-1)\,\text{var}_b(u)}$$

If samples are uncorrelated, then $\psi = k$. The difference between ψ and k gives an idea of the degree of the autocorrelation among the samples in the chain.

16.2.4 Numerical application

Example 16.1

Using the data in Example 3.1 and the variance components, the application of Gibbs sampling for estimation of variance components and the prediction of breeding values is illustrated. Uniform priors are assumed for the variance components such that $\nu_e = \nu_a = -2$ and $s_e^2 = s_u^2 = 0$. A flat prior is assumed for $\mathbf{b}$, and $\mathbf{u}$ is assumed to be normally distributed.

First, sample $\mathbf{b}_1^{[1]}$, where the superscript in brackets denotes iteration number, using Eqn 16.11, with $\hat{\mathbf{b}}_1$ calculated using Eqn 3.5 and $(\mathbf{x}_j'\mathbf{x})^{-1}\sigma_e^2 = (3)^{-1}40 = 13.333$. From Eqn 3.5:

$$\hat{\mathbf{b}}_1 = [(4.5 + 3.5 + 5.0) - (0 + 0 + 0)]/3 = 4.333$$

Assuming the random number (RN) generated from a normal distribution, N(0,1), is 0.1704, then $\mathbf{b}_1$ from Eqn 16.11 is:

$$\mathbf{b}_1^{[1]} = 4.333 + 0.1704\sqrt{13.333} = 4.955$$

Then sample $\mathbf{b}_2$ using Eqn 16.11 with $(\mathbf{x}_j'\mathbf{x})^{-1}\sigma_e^2 = (2)^{-1}40 = 20$ and $\hat{\mathbf{b}}_2$ is:

$$\hat{\mathbf{b}}_2 = [(2.9 + 3.9) - (0 + 0)]/2 = 3.40$$

Assuming the RN from N(0,1) is –0.1294, then:

$$\mathbf{b}_2^{[1]} = 3.40 - 0.1294\sqrt{20} = 2.821$$

The vector of solution $\mathbf{u}_j$ for animal j is sampled using Eqn 16.12, with $\hat{\mathbf{u}}_j$ calculated using Eqn 3.8. Thus for animal 1:

$\hat{\mathbf{u}}_1 = 0$ and $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}^{-1}_{1,1}\alpha)^{-1}\sigma_e^2 = (3.667)^{-1}40 = 10.908$

The value of $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}^{-1}_{1,1}\alpha)^{-1}$ is taken from the diagonal element of the coefficient matrix of the MME for Example 3.1. Assuming the RN from N(0,1) is 0.2067:

$$\mathbf{u}_1^{[1]} = 0 + 0.2067\sqrt{10.98} = 0.683$$

For animal 2, $\hat{\mathbf{u}}_2$ from Eqn 3.8 = –0.171, and $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}^{-1}_{1,1}\alpha)^{-1}\sigma_e^2 = (4)^{-1}40 = 10$. Then from Eqn 16.12, assuming RN from N(0,1) is –1.8025:

$$\mathbf{u}_2^{[1]} = -0.171 + -1.8025\sqrt{10} = -5.871$$

Similarly, given that $\hat{\mathbf{u}}_3$ from Eqn 3.8 = 1.468, $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}^{-1}_{1,1}\alpha)^{-1}\sigma_e^2 = (4)^{-1}40 = 10$ and RN = –0.5558, then:

$$\mathbf{u}_3^{[1]} = 1.468 - 0.5558\sqrt{10} = -0.290$$

For animal 4, $\hat{\mathbf{u}}_4 = 0.0976$ from Eqn 3.8, $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}^{-1}_{1,1}\alpha)^{-1}\sigma_e^2 = (4.667)^{-1}40 = 8.571$ and RN = –1.8654, then:

$$\mathbf{u}_4^{[1]} = 0.0976 - 1.8654\sqrt{8.571} = -5.364$$

Similar calculations using Eqn 16.12 gave estimates of $\mathbf{u}_5^{[1]}$, $\mathbf{u}_6^{[1]}$, $\mathbf{u}_7^{[1]}$ and $\mathbf{u}_8^{[1]}$ to be –3.097, –2.577, –1.621 and 0.697, respectively.

The vector of residuals, $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u}$, is:

$$\begin{pmatrix} \hat{e}_4 \\ \hat{e}_5 \\ \hat{e}_6 \\ \hat{e}_7 \\ \hat{e}_8 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{pmatrix} - \begin{pmatrix} 4.955 \\ 2.821 \\ 2.821 \\ 4.955 \\ 4.955 \end{pmatrix} - \begin{pmatrix} -5.364 \\ -3.097 \\ -2.577 \\ -1.621 \\ -0.697 \end{pmatrix} = \begin{pmatrix} 4.908 \\ 3.176 \\ 3.656 \\ 0.165 \\ 0.742 \end{pmatrix}$$

and $\hat{\mathbf{e}}'\hat{\mathbf{e}} = 48.118$. Sampling from the inverted χ^2 distribution with three degrees of freedom (Eqn 16.13) gave an estimate of 39.870 for the residual variance.

Using Eqn 16.14, sampling for σ_u^2 is again from the inverted χ^2 distribution, with $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} = 93.11$ and degrees of freedom being 6. An estimate of 23.913 was obtained for σ_u^2. Note that it is easier to compute $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$ using Eqn 2.3. Thus $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} = \mathbf{u}'(\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1}\mathbf{u} = \mathbf{m}'\mathbf{D}\mathbf{m}$ where $\mathbf{m} = \mathbf{T}^{-1}\mathbf{u}$, with $\mathbf{m}$ being a vector of Mendelian sampling for animals calculated using Eqn 2.2.

The next round of iteration is then commenced using the updated values computed for the parameters.

16.3 Multivariate Animal Model

In this section, the Gibbs sampling algorithm developed by Jensen *et al.* (1994) for models with maternal genetic effects is generalized for a multivariate situation. Given that animals are ordered within traits, the multivariate model for two traits could be written as:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix}$$

where terms are as defined in Eqn 5.1, with $\mathbf{u} = \mathbf{a}$. The conditional distribution of the complete data, given that animals are ordered within traits, is:

$$\left\langle \begin{matrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{matrix} \middle| \mathbf{b}_1, \mathbf{b}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{R} \right\rangle \sim \mathbf{N}\left[\begin{matrix} \mathbf{X}_1\mathbf{b}_1 + \mathbf{Z}_1\mathbf{u}_1 \\ \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_2\mathbf{u}_2 \end{matrix}, \mathbf{R} \otimes \mathbf{I} \right] \tag{16.16}$$

It is assumed that:

$$\left\langle \begin{matrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{matrix} \middle| \mathbf{G}, \mathbf{A} \right\rangle \sim \mathbf{N}\left[\begin{matrix} 0 \\ 0 \end{matrix}, \mathbf{G} \otimes \mathbf{A} \right] \tag{16.17}$$

where **G** is the genetic covariance matrix and **A** is the numerator relationship matrix.

16.3.1 Prior distributions

Assume that proper uniform distributions are defined for the fixed effects:

$$P(\mathbf{b}_1) \propto \text{constant}; \quad P(\mathbf{b}_2) \propto \text{constant}$$

with:

$$\mathbf{b}_i(\min) \leq \mathbf{b}_i \leq \mathbf{b}_i(\max)$$

An inverted Wishart distribution (Jensen *et al*., 1994) is used as prior distribution for the genetic and residual covariances. Thus the prior distribution for the residual covariance is:

$$P(\mathbf{R} \mid \mathbf{V}_e, \nu_e) \propto |\mathbf{R}|^{-\frac{1}{2}(\nu_e + p + 1)} \exp[-\tfrac{1}{2}\operatorname{tr}(\mathbf{R}^{-1}\mathbf{V}_e^{-1})] \tag{16.18}$$

The above is a p-dimensional inverse Wishart distribution (IW_2), where p is the order of **R**, $\mathbf{V}_e$ is a parameter of the prior distribution and ν_e is the degrees of freedom. If $\mathbf{V}_e = 0$ and $\nu_e = -(p + 1)$, the above reduces to a uniform distribution. Similarly, for the genetic covariance, the following prior distribution is assumed:

$$P(\mathbf{G} \mid \mathbf{V}_u, \nu_u) \propto |\mathbf{G}|^{-\frac{1}{2}(\nu_u + p + 1)} \exp[-\tfrac{1}{2}\operatorname{tr}(\mathbf{G}^{-1}\mathbf{V}_u^{-1})] \tag{16.19}$$

with terms $\mathbf{V}_u$ and ν_u equivalent to $\mathbf{V}_e$ and ν_e, respectively, in Eqn 16.18.

The joint posterior distribution assuming n traits and using Eqns 16.16 to 16.19, is:

$$\begin{aligned} &P(\mathbf{b}_1,\ldots,\mathbf{b}_n, \mathbf{u}_1,\ldots,\mathbf{u}_n, \mathbf{R}, \mathbf{G}) \\ &\quad \propto p(\mathbf{y}_1,\ldots,\mathbf{y}_n|\mathbf{b}_1,\ldots,\mathbf{b}_n, \mathbf{u}_1,\ldots\mathbf{u}_n, \mathbf{R})p(\mathbf{u}_1,\ldots\mathbf{u}_n,|\mathbf{G})p(\mathbf{G})p(\mathbf{R}) \end{aligned} \tag{16.20}$$

16.3.2 Conditional probabilities

Using the same principles as those for obtaining Eqns 16.11 and 16.12, the conditional distribution for the level k of the ith trait is:

$$\mathbf{b}_{i,k} \mid \mathbf{b}_{i,-k}, \mathbf{b}_j, \mathbf{u}, \mathbf{R}_e, \mathbf{G}, \mathbf{y} \sim \mathbf{N}(\mathbf{b}_{i,k}, (\mathbf{x'}_{i,k}\mathbf{r}^{ii}\mathbf{x}_{i,k})^{-1}); \quad j = 1, n \text{ and } j \neq i \tag{16.21}$$

with:

$$\hat{\mathbf{b}}_{i,k} = (\mathbf{x}'_{i,k}\mathbf{r}^{ii}\mathbf{x}_{i,k})^{-1}\mathbf{x}'_{i,k}(\mathbf{r}^{ii}\mathbf{y}_i + \mathbf{r}^{ij}\mathbf{y}_j) - \mathbf{r}^{ii}(\mathbf{x}'_{i,-k}\mathbf{b}_{i,-k} + \mathbf{z}_i\mathbf{u}_i) - \mathbf{r}^{ij}(\mathbf{x}_j\mathbf{b}_j + \mathbf{z}_j\mathbf{u}_j); j = 1, n \text{ and } j \neq i$$

Similarly, for the random animal effect, the conditional distribution for animal *k* of the *i*th trait is:

$$\mathbf{u}_{i,k} \mid \mathbf{u}_{i,-k}, \mathbf{u}_j, \mathbf{b}, \mathbf{R}_e, \mathbf{G}, \mathbf{y} \sim \mathbf{N}(\hat{\mathbf{u}}_{i,k}, (\mathbf{r}^{ii}\mathbf{z}'_{i,k}\mathbf{z}_{i,k} + \mathbf{g}^{ii}\mathbf{A}^{-1}_{k,k})^{-1}), \; j = 1, n \text{ and } j \neq i \tag{16.22}$$

with:

$$\hat{\mathbf{u}}_{i,k} = \left(\mathbf{z}'_{i,k}\mathbf{r}^{ii}\mathbf{z}_{i,k} + \mathbf{A}^{-1}_{k,k}\mathbf{g}^{ii}\right)^{-1}\left\{\mathbf{z}'_{i,k}\left(\mathbf{r}^{ii}\mathbf{y}_i + \mathbf{r}^{ij}\mathbf{y}_j - \mathbf{r}^{ii}\mathbf{x}_i\mathbf{b}_i - \mathbf{r}^{ij}\mathbf{x}_j\mathbf{b}_j\right) - \left(\mathbf{z}'_{i,k}r^{ij}\mathbf{z}_{j,k} + \mathbf{A}^{-1}_{k,k}\mathbf{g}^{ij}\mathbf{u}_{j,k}\right) - \mathbf{A}^{-1}_{k,s}\left(\mathbf{g}^{ii}\mathbf{u}_{i,s} + \mathbf{g}^{ij}\mathbf{u}_{j,s}\right)\right\}$$

where *s* represents the known parents of the *k*th animal.

However, instead of sampling for each level of fixed or random effects for one trait at a time, it is more efficient to implement block sampling for each level of fixed or random effect across all traits at once. The conditional distribution for level *k* of a fixed effect required for block sampling, assuming *n* = 2, is:

$$\left\langle \begin{matrix}\mathbf{b}_{1,k}\\ \mathbf{b}_{2,k}\end{matrix} \middle| \mathbf{b}_{-k}, \mathbf{u}, \mathbf{R}, \mathbf{G}, \mathbf{y} \right\rangle \sim \mathrm{N}\left[\begin{matrix}\hat{\mathbf{b}}_{1,k}\\ \hat{\mathbf{b}}_{2,k}\end{matrix}, (\mathbf{X}'_k\,\mathbf{R}^{-1}\mathbf{X}_k)^{-1}\right] \tag{16.23}$$

where:

$$\begin{pmatrix}\hat{\mathbf{b}}_{1,k}\\ \hat{\mathbf{b}}_{2,k}\end{pmatrix} = (\mathbf{X}'_k\,\mathbf{R}^{-1}\mathbf{X}_k)^{-1}(\mathbf{X}'_k\,\mathbf{R}^{-1}(\mathbf{y}_k - \mathbf{X}_{-k}\mathbf{b}_{-k} - \mathbf{Z}\hat{\mathbf{u}}))$$

which is equivalent to Eqn 5.4.

For the random animal effect, block sampling for animal *k*, assuming *n* = 2, the conditional distribution is:

$$\left\langle \begin{matrix}\mathbf{u}_{1,k}\\ \mathbf{u}_{2,k}\end{matrix} \middle| \mathbf{b}, \mathbf{u}_{j,-k}, \mathbf{R}, \mathbf{G}, \mathbf{y} \right\rangle \sim \mathrm{N}\left[\begin{matrix}\hat{\mathbf{u}}_{1,k}\\ \hat{\mathbf{u}}_{2,k}\end{matrix}, (\mathbf{Z}'_k\,\mathbf{R}^{-1}\mathbf{Z}_k + \mathbf{A}^{-1}_{k,k} \otimes \mathbf{G}^{-1})^{-1}\right] \tag{16.24}$$

where:

$$\begin{pmatrix}\hat{\mathbf{u}}_{1,k}\\ \hat{\mathbf{u}}_{2,k}\end{pmatrix} = (\mathbf{Z}'_k\,\mathbf{R}^{-1}\mathbf{Z}_k + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1})^{-1}\{(\mathbf{Z}'_k\,\mathbf{R}^{-1}(\mathbf{y}_k - \mathbf{X}\mathbf{b}) - \mathbf{A}^{-1} \otimes \mathbf{G}^{-1}(\hat{\mathbf{u}}_s + \hat{\mathbf{u}}_d)\}$$

where *s* and *d* are the sire and dam of the *k*th animal.

From Eqn 16.20, the full conditional distribution of the residual variance is:

$P(\mathbf{R}|\mathbf{b}, \mathbf{u}, \mathbf{y}) \propto P(\mathbf{R})P(\mathbf{y}|\mathbf{b}, \mathbf{u}, \mathbf{R})$

Including the prior distribution, the above can be expressed (Jensen *et al.*, 1994) as:

$$P(\mathbf{R} \mid \mathbf{b}, \mathbf{u}, \mathbf{y}) \propto |\mathbf{R}|^{-\frac{1}{2}(v_e+p+1+m)} \exp\left[-\frac{1}{2}\mathrm{tr}\{\mathbf{R}^{-1}(\mathbf{S}_e^2 + \mathbf{V}_e^{-1})\}\right]$$

where m is the number of records and $\mathbf{S}_e^2$ is:

$$\mathbf{S}_e^2 = \begin{pmatrix} \hat{\mathbf{e}}_1' \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_1' \hat{\mathbf{e}}_2 \\ \hat{\mathbf{e}}_2' \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2' \hat{\mathbf{e}}_2 \end{pmatrix}$$

assuming that $n = 2$ and $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\mathbf{b}_i - \mathbf{Z}_i\mathbf{u}_i$, $i = 1, n$.

Thus:

$$\mathbf{R} \mid \mathbf{b}, \mathbf{u}, \mathbf{y} \sim \mathrm{IW}_2((\mathbf{S}_e^2 + \mathbf{V}_e^{-1})^{-1}, \mathbf{v}_e + m) \tag{16.25}$$

which is in the form of a p-dimensional inverted Wishart distribution with $\mathbf{v}_e + m$ degrees of freedom and scale parameter $(\mathbf{S}_e^2 + \mathbf{V}_e^{-1})$.

Similarly, the conditional distribution for the additive genetic variance is:

$$P(\mathbf{G} \mid \mathbf{b}, \mathbf{u}, \mathbf{y}) \propto P(\mathbf{G})P(\mathbf{u} \mid \mathbf{G})$$

Including the prior distribution, the above can be expressed (Jensen *et al.*, 1994) as:

$$P(\mathbf{G} \mid \mathbf{b}, \mathbf{u}, \mathbf{y}) \propto |\mathbf{G}|^{-\frac{1}{2}(v_u+p+1+q)} \exp\left[-\frac{1}{2}\mathrm{tr}\{\mathbf{G}^{-1}(\mathbf{S}_u^2 + \mathbf{V}_u^{-1})\}\right]$$

where q is the number of animals and, assuming $n = 2$, $\mathbf{S}_u^2$ is:

$$\mathbf{S}_u^2 = \begin{pmatrix} \mathbf{u}_1'\mathbf{A}^{-1}\mathbf{u}_1 & \mathbf{u}_1'\mathbf{A}^{-1}\mathbf{u}_2 \\ \mathbf{u}_2'\mathbf{A}^{-1}\mathbf{u}_1 & \mathbf{u}_2'\mathbf{A}^{-1}\mathbf{u}_2 \end{pmatrix}$$

Thus:

$$\mathbf{G} \mid \mathbf{b}, \mathbf{u}, \mathbf{y} \sim \mathrm{IW}_2((\mathbf{S}_u^2 + \mathbf{V}_u^{-1})^{-1}, \mathbf{v}_u + q) \tag{16.26}$$

which again is in the form of a p-dimensional inverted Wishart distribution with $\mathbf{v}_u + q$ degrees of freedom and scale parameter $(\mathbf{S}_u^2 + \mathbf{V}_u^{-1})$.

16.3.3 Numerical illustration

Example 16.2

Using the data in Example 5.1 and the variance components, the application of Gibbs sampling to estimating variance components and predicting breeding values is illustrated. Uniform priors are assumed for the variance components such that $\mathbf{v}_e = \mathbf{v}_u = -3$ and $\mathbf{V}_e = \mathbf{V}_u = 0$. A flat prior is assumed for **b**, and **u** is assumed to be normally distributed.

Processing data and accumulating right-hand side (rhs) and diagonals (Diag) for level j of sex of calf effects as:

$$\text{rhs}_{1j} = \text{rhs}_{1j} + \mathbf{R}^{11}(\mathbf{y}_1 - \mathbf{u}_{1i}) + \mathbf{R}^{12}(\mathbf{y}_2 - \mathbf{u}_{2i})$$
$$\text{rhs}_{2j} = \text{rhs}_{2j} + \mathbf{R}^{21}(\mathbf{y}_2 - \mathbf{u}_{1i}) + \mathbf{R}^{22}(\mathbf{y}_2 - \mathbf{u}_{2i})$$
$$\mathbf{Diag}_j = \mathbf{Diag}_j + \mathbf{R}$$

When all data have been read, calculate solutions for level *j* of sex effect as:

$$\begin{pmatrix} \hat{b}_{1j} \\ \hat{b}_{2j} \end{pmatrix} = \mathbf{Diag}_j^{-1} \begin{pmatrix} rhs_{1j} \\ rhs_{2j} \end{pmatrix}$$

Sample $\mathbf{b}_j$ in Eqn 16.23 as:

$$\mathbf{b}_j = \begin{pmatrix} \hat{b}_{1j} \\ \hat{b}_{2j} \end{pmatrix} + \{\text{CHOL}(\mathbf{Diag}_j^{-1})\}\mathbf{h}$$

where **h** is the vector of normal deviates from a population of mean zero and variance 1 and CHOL is the Cholesky decomposition of the inverse of the matrix **Diag**.

Next, process data and accumulate right-hand side (rhs) and diagonals (Diag) for animal *i* as:

$$\text{rhs}_{1i} = \text{rhs}_{1i} + \mathbf{R}^{11}(\mathbf{y}_1 - \mathbf{b}_{1j}) + \mathbf{R}^{12}(\mathbf{y}_2 - \mathbf{b}_{2j})$$
$$\text{rhs}_{2i} = \text{rhs}_{2i} + \mathbf{R}^{21}(\mathbf{y}_2 - \mathbf{b}_{1j}) + \mathbf{R}^{22}(\mathbf{y}_2 - \mathbf{b}_{2j})$$
$$\mathbf{Diag}_i = \mathbf{Diag}_i + \mathbf{R}$$

When all data have been read, calculate solutions for animal *i* as:

$$\begin{pmatrix} \hat{u}_{1i} \\ \hat{u}_{2i} \end{pmatrix} = \mathbf{Diag}_i^{-1} \begin{pmatrix} rhs_{1i} \\ rhs_{2i} \end{pmatrix}$$

Sample $\mathbf{u}_i$ in Eqn 16.24 as:

$$\mathbf{u}_i = \begin{pmatrix} \hat{u}_{1i} \\ \hat{u}_{2i} \end{pmatrix} + \{\text{CHOL}(\mathbf{Diag}_i^{-1})\}\mathbf{h}$$

All data is then processed to obtain residual effects as:

$$\hat{\mathbf{e}} = \begin{pmatrix} \hat{\mathbf{e}}_1 \\ \hat{\mathbf{e}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{Z}_1\hat{\mathbf{u}}_1 \\ \mathbf{y}_2 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \mathbf{Z}_2\hat{\mathbf{u}}_2 \end{pmatrix}$$

and calculate residual sums of squares, $\mathbf{S}_e^2 = \hat{\mathbf{e}}\hat{\mathbf{e}}'$. Then compute $\mathbf{T} = (\mathbf{S}_e^2 + \mathbf{V}_e^{-1})^{-1}$. Cholesky decomposition of **T** is carried out to obtain **LL′**, where **L** is a lower triangular matrix. Sampling from a Wishart distribution with **L** as the input matrix and $\mathbf{v}_e + m$ degrees of freedom (Eqn 16.25) generates a new sample value of **R**.

Similarly, to compute a new sample value of **G** using Eqn 16.26, first compute $\mathbf{T}^{-1} = (\mathbf{S}_u^2 + \mathbf{V}_u^{-1})^{-1}$. Decompose **T** to obtain **LL′** and sample from a Wishart distribution with **L** as the input matrix and $\mathbf{v}_u + q$ degrees of freedom. Another cycle of sampling is then initiated until the desired length of chain is achieved. Post-processing of results can be carried out, as discussed in Section 16.2.3.

17 Solving Linear Equations

17.1 Introduction

Different methods can be used to solve the MME covered in the previous chapters. These various methods could broadly be divided into three main categories:

1. Direct inversion (Section 17.2).
2. Iteration on the MME (Section 17.3).
3. Iteration on the data (Section 17.4).

The manner in which the MME are set up depends on the method to be used in solving these equations. As shown in Section 17.4, the third method, for instance, does not involve setting up the MME directly.

17.2 Direct Inversion

The solutions to the MME in the various examples given so far in this book have been based on this method. It involves setting up the MME and inverting the coefficient matrix. Solutions are obtained by multiplying the right-hand side (RHS) by the inverse of the coefficient matrix. Thus **b**, the vector of solution, is calculated as:

$$\hat{\mathbf{b}} = \mathbf{C}^{-1}\mathbf{y}$$

where **C** is the coefficient matrix and **y** is the RHS. Since the coefficient matrix is symmetric, only the upper triangular portion is usually set up and inverted. The major limitation of this approach is that it can only be applied to small data sets in view of the memory requirements and computational difficulties of inverting large matrices.

17.3 Iteration on the Mixed Model Equations

This involves setting up the MME and iterating on these equations until convergence is achieved at a predetermined criterion. The iterative procedures are based on the general theory for solving simultaneous equations. For instance, given two simultaneous equations with unknown parameters, b_1 and b_2, the first equation can be solved for b_1 in terms of b_2. This value of b_1 can then be substituted in the

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

second equation to solve for b_2. The value of b_2 is then substituted in the first equation to calculate b_1. This is the principle upon which the iterative procedures are based. In the iterative procedure, the above process is continued until the solutions for the b's are more or less the same in each round of iteration and the equations are said to have converged. There are various iterative procedures that can be used, and some are described below.

17.3.1 Jacobi iteration

One of the simplest methods is Jacobi iteration or total step iteration.

Consider the following set of simultaneous equations:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

These equations can also be written as:

$$c_{11}b_1 + c_{12}b_2 + c_{13}b_3 = y_1$$
$$c_{21}b_1 + c_{22}b_2 + c_{23}b_3 = y_2$$
$$c_{31}b_1 + c_{32}b_2 + c_{33}b_3 = y_3$$

or as:

$$\mathbf{Cb} = \mathbf{y} \tag{17.1}$$

The system of equations is rearranged so that the first is solved for b_1, the second for b_2 and the third for b_3. Thus:

$$b_1^{r+1} = (1/c_{11})(y_1 - c_{12}b_2^r - c_{13}b_3^r)$$
$$b_2^{r+1} = (1/c_{22})(y_2 - c_{21}b_1^r - c_{23}b_3^r) \tag{17.2}$$
$$b_3^{r+1} = (1/c_{33})(y_3 - c_{31}b_1^r - c_{32}b_2^r)$$

The superscript r refers to the number of the round of iteration. In the first round of iteration, r equals 1 and b_1 to b_3 could be set to zero or an assumed set of values that are used to solve the equations to obtain a new set of solutions (**b** terms). The process is continued until two successive sets of solutions are within previously defined allowable deviations and the equations are said to converge. One commonly used convergence criterion is the sum of squares of differences between the current and previous solutions divided by the sum of squares of the current solution. Once this is lower than a predetermined value, for instance 10^{-9}, the equations are considered to have converged.

From the set of equations above, the solution for b_i was obtained by dividing the adjusted RHS by the diagonal (a_{ii}). It is therefore mandatory that the diagonal element, often called the pivot element, is not zero. If a zero pivot element is encountered during the iterative process, the row containing the zero should be exchanged with a row below it in which the element in that column is not zero. To avoid the problem of encountering a zero pivot element and

generally improving the efficiency of the iterative process, it is sometimes recommended that the system of equations should be ordered such that the coefficient of b_1 of the greatest magnitude occurs in the first equation, the coefficient of b_2 of the greatest magnitude in the remaining equations occurs in the second equation, etc.

The iterative procedure described above is usually called Jacobi iteration as all new solutions in the current (r) round of iteration are obtained using solutions only from the previous ($r - 1$) round of iteration. The Jacobi iterative procedure is inefficient in handling systems of equations that are not constrained (i.e. with no restrictions placed on the solutions for the levels of an effect) and convergence is not guaranteed (Maron, 1987; Misztal and Gianola, 1988). When a random animal effect is involved in the system of equations with relationships included, it is usually necessary to use a relaxation factor of below 1.0, otherwise equations may not converge (Groeneveld, 1990). The relaxation factor refers to a constant estimated on the basis of the linear changes in the solutions during the iteration process and applied to speed up the solutions towards convergence. When iterating on the data (Section 17.4), the Jacobi iterative procedure involves reading only one data file, even with several effects in the model. With large data sets this has the advantage of reducing memory requirement and processing time compared with the Gauss–Seidel iterative procedure (see Section 17.3.2).

The Jacobi iterative procedure can be briefly summarized as follows.

Following Ducrocq (1992), Eqn 17.1 can be written as:

$$[\mathbf{M} + (\mathbf{C} - \mathbf{M})]\mathbf{b} = \mathbf{y}$$

if **M** is the diagonal matrix containing the diagonal elements of **C**; then the algorithm for Jacobi iteration is:

$$\mathbf{b}^{(r+1)} = \mathbf{M}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b}^{(r)}) + \mathbf{b}^{(r)} \qquad (17.3)$$

When a relaxation factor (w) is applied, the above equation becomes:

$$\mathbf{b}^{(r+1)} = w[\mathbf{M}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b}^{(r)})] + \mathbf{b}^{(r)}$$

Another variation of the Jacobi iteration, called second-order Jacobi, is usually employed in the analysis of large data sets and it can increase the rate of convergence. The iterative procedure for second-order Jacobi is:

$$\mathbf{b}^{(r+1)} = \mathbf{M}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b}^{(r)} + \mathbf{b}^{(r)} + w(\mathbf{b}^{(r)} - \mathbf{b}^{(r-1)}))$$

Example 17.1

Using the coefficient matrix and the RHS for Example 3.1, Jacobi iteration (Eqn 17.2) is carried out using only the non-zero element of the coefficient matrix. Solutions for sex effect (**b** vector) and random animal effect (**u** vector) are shown below with the round of iteration. The convergence criterion (CONV) was the sum of squares of differences between the current and previous solutions divided by the sum of squares of the current solution.

	Rounds of iteration									
Effects	0[a]	1	2	3	4	16	17	18	19	20
b_1	4.333	4.333	4.381	4.370	4.368	4.358	4.358	4.358	4.358	4.358
b_2	3.400	3.400	3.433	3.365	3.414	3.404	3.404	3.404	3.404	3.404
$\hat{u}_1$	0.000	0.267	0.164	0.185	0.131	0.099	0.099	0.099	0.099	0.099
$\hat{u}_2$	0.000	0.000	–0.073	–0.003	–0.039	–0.018	–0.018	–0.018	–0.018	–0.018
$\hat{u}_3$	0.000	–0.033	–0.080	–0.049	–0.070	–0.041	–0.041	–0.041	–0.041	–0.041
$\hat{u}_4$	0.167	–0.138	–0.007	–0.035	0.000	–0.008	–0.008	–0.008	–0.008	–0.008
$\hat{u}_5$	–0.500	–0.411	–0.248	–0.265	–0.204	–0.185	–0.185	–0.185	–0.185	–0.185
$\hat{u}_6$	0.500	0.345	0.318	0.237	0.236	0.178	0.178	0.178	0.177	0.177
$\hat{u}_7$	–0.833	–0.406	–0.390	–0.301	–0.295	–0.249	–0.249	–0.249	–0.249	–0.249
$\hat{u}_8$	0.667	0.400	0.286	0.232	0.207	0.183	0.183	0.183	0.183	0.183
CONV	1.000	2.3^{-2}	3.9^{-3}	1.4^{-3}	5.9^{-4}	4.2^{-8}	1.6^{-8}	1.0^{-8}	4.1^{-9}	3.0^{-9}

[a]Starting values.

The starting solutions for sex effect were the mean yield for each sex subclass and, for animals with records, starting solutions were the deviation of their yields from the mean yield of their respective sex subclass and zero for ancestors. The final solutions obtained after the 20th round of iteration were exactly the same as obtained in Section 3.2 by direct inversion of the coefficient matrix. The solutions for sex effect were obtained using Eqn 17.2. Thus in the first round of iteration the solution for males was:

$$b_1 = \frac{1}{c_{11}}\left(\sum_{k=1}^{m} y_k - (1)\hat{u}_4 - (1)\hat{u}_7 - (1)\hat{u}_8\right)$$

where c_{ii} is the diagonal element of the coefficient matrix for level i of sex effect and m is the number of records for males.

$$b_1 = 1/3(13.0 - 0.167 - (-0.833) - 0.667) = 4.333$$

However, using Eqn 17.2 to obtain animal solutions caused the system of equations to diverge. A relaxation factor (w) of 0.8 was therefore employed and solutions for animal j were computed as:

$$\hat{u}_j^r = w\left[\left(\frac{1}{c_{ll}}\right)\left(y_j - c_{li}\hat{b}_i^{r-1} - \sum_k c_{lt}\hat{u}_k^{r-1}\right) - \hat{u}_j^{r-1}\right] + \hat{u}_j^{r-1}$$

where $l = j + n$, $t = k + n$, with $n = 2$; the total number of levels of fixed effect, c_{lt} and c_{li}, for instance, are the elements of the coefficient matrix between animals j and k, and animal j and level i of sex effect, respectively. Thus in the first round of iteration, solutions for animals 1 and 8 are calculated as:

$$\begin{aligned}\hat{u}_1^1 &= w[\{1/c_{33}(y_1 - (1)\hat{u}_2 - (-1.333)\hat{u}_4 - (-2)\hat{u}_6)\} - \hat{u}_1^0] + \hat{u}_1^0\\ &= w[\{1/3.667(0 - 0 - (-0.223) - (-1))\} - 0] + 0\\ &= 0.8(0.334 - 0) + 0 = 0.267\end{aligned}$$

and:

$$\begin{aligned}\hat{u}_8^1 &= w[\{1/c_{1010}(y_8 - (1)b_1 - (-2)\hat{u}_3 - (-2)\hat{u}_6)\} - \hat{u}_8^0] + \hat{u}_8^0\\ &= w[\{1/5(5 - 4.333 - 0 - (-1))\} - 0.667] + 0.667\\ &= 0.8(0.333 - 0.667) + 0.667 = 0.400\end{aligned}$$

17.3.2 Gauss–Seidel iteration

Another iterative procedure commonly used is Gauss–Seidel iteration. This is similar to Jacobi iteration except that most current solutions are calculated from the most recent available solution rather than the solution from the previous round of iteration. Using the same set of simultaneous equations as in Eqn 17.1, solutions for b_1, b_2 and b_3 in the first round of iteration become:

$$\begin{aligned} b_1^{r+1} &= (1/c_{11})(y_1 - c_{12}b_2^r - c_{13}b_3^r) \\ b_2^{r+1} &= (1/c_{22})(y_2 - c_{21}b_1^{r+1} - c_{23}b_3^r) \\ b_3^{r+1} &= (1/c_{33})(y_3 - c_{31}b_1^{r+1} - c_{32}b_2^{r+1}) \end{aligned} \tag{17.4}$$

Thus the solution for b_2 in the $r + 1$ round of iteration is calculated using the most recent solution for b_1 (b_1^{r+1}) instead of the previous solution (b_1^r), and the current solution for b_3 is calculated from the current solutions for b_1 (b_1^{r+1}) and b_2 (b_2^{r+1}). If, in Eqn 17.3, **L** is strictly the lower triangular of **C** and **D** the diagonal of **C**, then Eqn 17.3 becomes the Gauss–Seidel iteration when **M** = **L** + **D**. The convergence criteria could equally be defined as discussed in Section 17.3.1. Generally, equations are guaranteed to converge with the Gauss–Seidel iterative procedure. However, when iterating on the data, this iterative procedure involves reading one data file for each effect in the model. With large data sets, the setting up of data files for each effect could result in large memory requirement and the reading of several files in each round of iteration could increase processing time.

Example 17.2

Using the same coefficient matrix, RHS and starting values as in Example 17.1 above, the Gauss–Seidel iteration (Eqn 17.4) is carried out for the same number of iterations as in Jacobi's method and the results are shown below. The convergence criterion is as defined in Example 17.1.

	Rounds of iteration									
Effects	0	1	2	3	4	16	17	18	19	20
b_1	4.333	4.333	4.400	4.372	4.364	4.359	4.359	4.359	4.359	4.359
b_2	3.400	3.400	3.392	3.403	3.407	3.405	3.405	3.405	3.405	3.405
$\hat{u}_1$	0.000	0.333	0.194	0.149	0.115	0.098	0.098	0.098	0.098	0.098
$\hat{u}_2$	0.000	−0.083	−0.035	−0.006	−0.008	−0.019	−0.019	−0.019	−0.019	−0.019
$\hat{u}_3$	0.000	−0.021	−0.136	−0.109	−0.076	−0.041	−0.041	−0.041	−0.041	−0.041
$\hat{u}_4$	0.167	−0.119	0.001	0.004	−0.003	−0.009	−0.009	−0.009	−0.009	−0.009
$\hat{u}_5$	−0.500	−0.376	−0.261	−0.218	−0.199	−0.186	−0.186	−0.186	−0.186	−0.186
$\hat{u}_6$	0.500	0.392	0.254	0.204	0.185	0.177	0.177	0.177	0.177	0.177
$\hat{u}_7$	−0.833	−0.364	−0.284	−0.260	−0.253	−0.250	−0.250	−0.250	−0.250	−0.250
$\hat{u}_8$	0.667	0.282	0.167	0.164	0.171	0.182	0.183	0.183	0.183	0.183
CONV	1.000	1.9^{-2}	3.4^{-3}	3.1^{-4}	1.0^{-4}	7^{-10}	4^{-10}	2^{-10}	1^{-10}	8^{-11}

CONV, convergence criterion.

The solutions obtained are the same as those obtained from Jacobi iteration and by direct inversion of the coefficient matrix in Example 3.1. In addition, the equations converged faster than when using Jacobi iteration and no relaxation factor was applied.

Iterating on the MME could be carried out as described above, once the equations have been set up, using only the stored non-zero elements of the coefficient matrix. In practice, it may be necessary to store the non-zero elements and their rows and columns on disk for large data sets because of the memory requirement, and these are read in each round of iteration.

17.4 Iterating on the Data

This is the most commonly used methodology in national genetic evaluations, which usually involve millions of records. Schaeffer and Kennedy first presented this method in 1986. It does not involve setting up the coefficient matrix directly, but it involves setting up equations for each level of effects in the model as the data and pedigree files are read and solved using either Gauss–Seidel or Jacobi iteration or a combination of both or a variation of any of the iterative procedures such as second-order Jacobi. Presented below are the basic equations for the solutions of various effects under several models and these form the basis of the iterative process for each of the models.

The equation for the solution of level i for a fixed effect in the model in a univariate animal situation is Eqn 3.5, which is derived from the MME and can be generalized as:

$$\hat{b}_i = \frac{\sum_{k=1}^{n_i} y_{ki} - \sum_{j=1}^{m} \hat{w}_{ij}}{n_i} \tag{17.5}$$

where y_{ki} is the kth record in level i, m is the total number of levels of other effects within subclass i of the fixed effect and $\hat{w}_{ij}$ is the solution for the jth level, and n_i is the number of records in fixed effect subclass i. However, when there are many fixed effects in the model, the above formula may be used to obtain solutions for the major fixed effect with many levels such as HYS, while the vector of solutions ($\mathbf{f}$) for other minor fixed effects with few levels may be calculated as:

$$\mathbf{f} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \hat{\mathbf{w}} - \hat{\mathbf{b}}) \tag{17.6}$$

where $\mathbf{y}$ is the vector of observations, $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of the coefficient matrix for the minor fixed effects, and $\hat{\mathbf{w}}$ and $\hat{\mathbf{b}}$ are vectors of solutions for effects as defined in Eqn 17.4. The matrix $\mathbf{X}'\mathbf{X}$ could be set up in the first round of iteration and stored in the memory for use in subsequent rounds of iterations.

The solution ($\hat{u}$) for the level j (animal j) of the random animal effect in the univariate animal model is calculated using Eqn 3.8, which can be rewritten (replacing n_3 by k) as:

$$\hat{u}_j = [n_1\alpha(\hat{u}_s + \hat{u}_d) + n_2 yd + \sum_o \{k_o\alpha(\hat{u}_o - 0.5(\hat{u}_{mo}))\}]/\text{diag}_j \tag{17.7}$$

with:

$$\text{diag}_j = 2(n_1)\alpha + n_2 + \sum_o \{(k_o/2)\alpha\}$$

where $\hat{u}_s$, $\hat{u}_d$ and $\hat{u}_o$ are solutions or EBVs for the sire, dam and oth progeny of animal j, respectively; $\hat{u}_{mo}$ is the solution of the mate of animal j with respect to progeny o; yd is yield deviation, i.e. yield of animal j corrected for all other effects in the model; $n_1 = 1$ or $\frac{2}{3}$ if both or one parent of animal j is known; n_2 is the number of records; k_o is 1 or $\frac{2}{3}$ if the other parent of progeny o (mate of animal j) is known or not known; and $\alpha = \sigma_e^2/\sigma_a^2$.

In the multivariate animal model situation with equal design and random animal effect as the only random effect in addition to residual effects, the solutions for the levels of fixed effect and animal effects are obtained using Eqns 5.4 and 5.8, respectively, which are derived from the MME (Eqn 5.3).

For maternal animal model equations, the solutions for fixed effects could be calculated using Eqn 7.3. The equations for animal and genetic maternal effects are based on Eqn 7.4, given earlier. From Eqn 7.4, the solution ($\hat{u}$) for direct effect for animal i is:

$$\begin{aligned}\hat{u}_i = [&n_1\alpha_1(\hat{u}_s + \hat{u}_d) + n_1\alpha_2(\hat{m}_s + \hat{m}_d) - n_4\alpha_2(\hat{m}_i) - (k_o/2)\alpha_2(\hat{m}_i) \\ &+ n_2(y_i - b_j - \hat{m}_d - \hat{pe}_d) + \sum_o\{k_o\alpha_1(\hat{u}_o - 0.5(\hat{u}_{mo}))\} \\ &+ \sum_o\{k_o\alpha_2(\hat{m}_o - 0.5(\hat{m}_{mo}))\}]/\text{diag}_i \end{aligned} \quad (17.8)$$

with:

$$\text{diag}_i = 2(n_1)\alpha_1 + n_2 + \sum_o\{(k_o/2)\alpha_1\}$$

where $\hat{m}_i$, $\hat{m}_s$, $\hat{m}_d$, $\hat{m}_o$ and $\hat{m}_{mo}$ are solutions for genetic maternal effects for animal i, sire, dam, oth progeny of animal i and mate of animal i, respectively; y_i is the yield for animal i; b_j is the solution for fixed effect j; $\hat{pe}_d$ is the permanent environmental effect for the dam of animal i; n_1, n_2 and k_o are as defined above and $n_4 = 2(n_1)$; and α terms are as defined in Eqn 7.4.

The solution (m) for genetic maternal effect for animal i from Eqn 7.4 is:

$$\begin{aligned}m_i = [&n_1\alpha_2(\hat{u}_s + \hat{u}_d) + n_1\alpha_3(\hat{m}_s + \hat{m}_d) - n_4\alpha_2(\hat{u}_i) - (k_o/2)\alpha_2(\hat{u}_i) \\ &+ n_2(y_i - b_j - \hat{u}_i - \hat{m}_d - \hat{pe}_d) + \sum_o\{k_o\alpha_2(\hat{u}_o - 0.5(\hat{u}_{mo}))\} \\ &+ \sum_o\{k_o\alpha_3(\hat{m}_o - 0.5(\hat{m}_{mate}))\}]/\text{diag}_i \end{aligned} \quad (17.9)$$

with:

$$\text{diag}_i = 2(n_1)\alpha_3 + n_2 + \sum_o\{(k_o/2)\alpha_3\}$$

Solutions for permanent environmental effect are obtained using Eqn 7.5.

The computational procedure for a reduced animal model was presented by Schaeffer and Wilton (1987) using a bivariate analysis. The procedure is similar to the animal model described above except that records for non-parents are written twice, one record for each parent. Consequently, the residual variance of non-parental records (r_2) is multiplied by 2, that is:

$$r_2 = 2(\sigma_e^2 + d(\sigma_a^2)) = 2(1 + d\alpha^{-1})\sigma_e^2$$

where $d = \frac{1}{2}$ or $\frac{3}{4}$ if both or one parent is known and the contribution of non-parents' records to the diagonal of their parents is 0.5 instead of 0.25 (see Example 7.2).

The equations for solutions for levels of fixed and random effects are similar to those defined earlier. From Eqn 7.3, if the residual variance for parental records is defined as r_1, the contribution of parental records to the RHS for level i of a major fixed effect is:

$$\text{RHS}_i = \sum_{k=1}^{n_i}(r_1^{-1}(y_{ik} - \hat{w}_{kj})) \quad (17.10)$$

where n_i is the number of parental records in level i of fixed effect and $\hat{w}_{kj}$ is the solution for the jth level of other effects in the model affecting record k. The contribution of non-parental records to the RHS is included as:

$$\text{RHS}_i = \text{RHS}_i + \sum_{k=1}^{m_i}(r_{2k}^{-1}(y_{ki} - 0.5(\hat{u}_s + \hat{u}_d) - \hat{w}_{kj})) \tag{17.11}$$

where m_i is the number of non-parental records in level i of fixed effect, $\hat{u}_s$ and $\hat{u}_d$ are solutions for the sire and dam of the non-parent with record k, r_{2k}^{-1} is the inverse of the residual variance for the non-parental record k and $\hat{w}_{kj}$ is the solution for level j of other effects in the model apart from random animal effects affecting record k. Then:

$$b_i = \frac{\text{RHS}_i}{\sum_{k=1}^{n_i} r_1^{-1} + \sum_{j=1}^{m_i} r_2^{-1}}$$

The equation for the breeding value of the jth animal, which is a parent with its own yield record, a non-parental record from progeny i and information from another progeny (o), who is itself a parent, is:

$$\hat{u}_j = [n_1\alpha(\hat{u}_s + \hat{u}_d) + n_2 r_1^{-1}(yd_j) + n_3 r_2^{-1}(yd_i - (0.5)\hat{u}_{mi}) + \sum_o\{k_o\alpha(\hat{u}_o - 0.5(\hat{u}_{mo}))\}]/\text{diag}_j \tag{17.12}$$

with:

$$\text{diag}_j = 2(n_1)\alpha + n_2 r_1^{-1} + (0.5)n_3 r_2^{-1} + \sum_o\{(k_o/2)\alpha\}$$

where yd_j and yd_i are yield deviations for animal j and progeny i, which is a non-parent, $\hat{u}_{mi}$ is the breeding value for the mate of animal j with respect to the ith progeny (non-parent), n_2 is the number of observations (records) on animal j, n_3 is the number of non-parental records, r_1^{-1} and r_2^{-1} are as defined earlier and all other terms are as defined in Eqn 17.7. Note that contributions from the oth progeny in the above equation refer to those progeny of animal j who are themselves parents and that non-parental records are adjusted for half the breeding value of the mate of animal j. If animal j has no non-parental records from its progeny, Eqn 17.12 is the same as Eqn 17.7.

The principles of evaluation based on iterating on the data are illustrated below using a univariate animal model and a reduced animal model with maternal effects.

17.4.1 Animal model without groups

Example 17.3

Using the same data as in Example 3.1 (Table 3.1) on the weaning weight of beef calves, parameters and model, the principles of predicting breeding values and estimating solutions for fixed effects iterating on the data are illustrated using Gauss–Seidel iteration.

DATA ARRANGEMENT

Gauss–Seidel iteration requires the data files to be sorted by the effect to be solved for. The pedigree file is needed when solving for animal solutions.

The pedigree file is created and ordered in such a manner that contributions to the diagonal and RHS of an animal from the pedigree due to the number of parents known (see type 1 record below) and from progeny accounting for whether mate is known (type 2 record), can be accumulated while processing the animal. Thus, initially, a pedigree file is created consisting of two types of records:

1. Type 1 record for all animals in the data comprising the animal identity, record type, and sire and dam identities.
2. Type 2 record for each parent in the data comprising the parent identity, record type, identities for progeny and other parent (mate) if known. The type 2 records are used to adjust the contribution of the progeny to each parent for the mate's breeding value when solving for animal solutions.

The pedigree file is sorted by animal and record type. The sorted pedigree file for the example data is given below.

Animal	Code	Sire or progeny	Dam or mate
1	1	0	0
1	2	4	0
1	2	6	2
2	1	0	0
2	2	5	3
2	2	6	1
3	1	0	0
3	2	5	2
3	2	8	6
4	1	1	0
4	2	7	5
5	1	3	2
5	2	7	4
6	1	1	2
6	2	8	3
7	1	4	5
8	1	3	6

Second, a data file is set up consisting of animal identity, fixed effects, covariates and traits. If there is a major fixed effect with many levels, two data files need to be set up: one sorted by the major fixed effects such as herd or HYS (file A), to be used when solving for the major fixed effect; and the other sorted by animal identity (file B), to be used to solve for animal solutions. Assuming sex effect to be the major fixed effect in the example data, the data sorted by sex are as follows:

Calf	Sex	Weaning weight gain (kg)
4	Male	4.5
7	Male	3.5
8	Male	5.0
5	Female	2.9
6	Female	3.9

Iteration stage

Let $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ be vectors of solutions for sex and animal effects. Starting values for sex and animal effect are assumed to be the same as in Example 17.1.

Solving for fixed effects. In each round of iteration, file A is read one level of sex effect at a time with adjusted right-hand sides (ARHS) and diagonals (DIAG) accumulated for the *i*th level as:

$$\text{ARHS}_i = \text{ARHS}_i + y_{ik} - \hat{u}_k$$

$$\text{DIAG}_i = \text{DIAG}_i + 1$$

At the end of the *i*th level, the solution for the level is computed as:

$$\hat{b}_i = \text{ARHS}_i/\text{DIAG}_i$$

The above step essentially involves adjusting the yields for animal effects using previous solutions and calculating solutions for each level of sex effect. For example, the solution for level one of sex effect in the first round of iteration is:

$$\hat{b}_1 = [(4.5 - 0.167) + (3.5 - (-0.833)) + (5.0 - 0.667)]/3 = 4.333$$

After calculating solutions for fixed effect in the current round of iteration, file B and the pedigree file are processed to compute animal solutions.

Solving for animal solutions. DIAG and ARHS are accumulated as data for each animal and read from the pedigree file or from both the pedigree file and file B for animals with records. When processing type 1 records in the pedigree file for the *k*th animal, the contribution to the DIAG and ARHS according to the number of parents known is as follows:

Number of parents known		
None	One (sire (*s*))	Both
$\text{ARHS}_k = 0$ $\text{DIAG}_k = \alpha$	$\text{ARHS}_k = (\frac{2}{3})\alpha(\hat{u}_s)$ $\text{DIAG}_k = (\frac{4}{3})\alpha$	$\text{ARHS}_k = \alpha(\hat{u}_s + \hat{u}_d)$ $\text{DIAG}_k = 2\alpha$

where $\hat{u}_s$ and $\hat{u}_d$ are current solutions for the sire and dam, respectively.

When processing type 2 records in the pedigree file for the *k*th animal, the contribution to the DIAG and ARHS according to whether the mate of animal *k* is known or not is as follows:

Mate unknown	Mate known
$\text{ARHS}_k = \text{ARHS}_k + (\frac{2}{3})\alpha(\hat{u}_o)$ $\text{DIAG}_k = \text{DIAG}_k + (\frac{1}{3})\alpha$	$\text{ARHS}_k = \text{ARHS}_k + \alpha(\hat{u}_o - 0.5\hat{u}_m)$ $\text{DIAG}_k = \text{DIAG}_k + (\frac{1}{2})\alpha$

where $\hat{u}_o$ and $\hat{u}_m$ are current solutions for the progeny and mate, respectively, of the *k*th animal. If the *k*th animal has a yield record:

$$\text{ARHS}_k = \text{ARHS}_k + y_{ik} - \hat{b}_i$$

$$\text{DIAG}_k = \text{DIAG}_k + 1$$

where $\hat{b}_i$ are current solutions for level i of sex effect.

When all pedigree and yield records for the kth animal have been processed, the solution for the animal is computed as:

$$\hat{u}_k = \text{ARHS}_k/\text{DIAG}_k$$

For the example data, the solutions for animal 5 in the first round of iteration is computed as follows.

Contribution to diagonal from pedigree is:

$$\text{DIAG}_5 = (2 + 0.5)\alpha = 5.00$$

Accounting for yield record, diagonal becomes:

$$\text{DIAG}_5 = 5.00 + 1 = 6.00$$

Contribution to RHS from yield is:

$$\text{ARHS}_5 = 2.9 - 3.40 = -0.5$$

Contribution to RHS from parents and progeny (pedigree) is:

$$\begin{aligned}\text{ARHS}_5 &= \text{ARHS}_5 + \alpha(\hat{u}_2 + \hat{u}_3) + \alpha(\hat{u}_7 - 0.5(\hat{u}_4)) \\ &= -0.5 + 2(-0.083 + (-0.021)) + 2(-0.833 - 0.5(-0.119)) \\ &= -2.255\end{aligned}$$

and:

$$\hat{u}_5 = -2.255/6.00 = -0.376$$

When all animals have been processed, the current round of iteration is completed. However, the iteration process is continued for sex and animal effects until convergence is achieved. The convergence criterion can be defined as in Section 17.3.1. In this example, solutions were said to have converged when the sum of squares of differences between the current and previous solutions divided by the sum of squares of the current solution was less than 10^{-7}. The solutions for all effects in the first round of iteration and at convergence at the 20th iteration are as follows:

	Solutions	
Effects	At round 1	At convergence
Sex		
Male	4.333	4.359
Female	3.400	3.404
Animal		
1	0.333	0.098
2	–0.083	–0.019
3	–0.021	–0.041
4	–0.119	–0.009
5	–0.376	–0.186
6	0.392	0.177
7	–0.364	–0.249
8	0.282	0.183

These solutions are the same as those obtained by direct inversion of the coefficient matrix in Section 3.2 or iterating on the coefficient matrix in Section 17.2. However, as stated earlier, the advantage of this method is that the MME are not set up and therefore memory requirement is minimal and can be applied to large data sets.

17.4.2 Animal model with groups

Example 17.4

With unknown parents assigned to phantom groups, the procedure is very similar to that described in Section 17.4.1, with no groups in the model except in the way the pedigree file is set up and animal solutions are computed. Using the same data, parameters and model as in Example 3.4, the methodology is illustrated below.

DATA PREPARATION

The pedigree file is set up as described in Section 17.4.1 with ancestors with unknown parentage assigned to groups. The assignment of unknown parents for the example pedigree has been described in Section 3.6. However, there is also an additional column for each animal indicating the number of unknown parents for each animal.

The pedigree with unknown parents assigned to groups and the additional column indicating the number of unknown parents is as follows:

Calf	Sire	Dam	Number of unknown parents
1	9	10	2
2	9	10	2
3	9	10	2
4	1	10	1
5	3	2	0
6	1	2	0
7	4	5	0
8	3	6	0

and the ordered pedigree for the analysis is:

Animal	Code	Sire or progeny	Dam or mate	Number of unknown parents
1	1	9	10	2
1	2	4	10	1
1	2	6	2	0
2	1	9	10	2
2	2	5	3	0
2	2	6	1	0
3	1	9	10	2
3	2	5	2	0
3	2	8	6	0

Continued

(*Continued*)

Animal	Code	Sire or progeny	Dam or mate	Number of unknown parents
4	1	1	10	1
4	2	7	5	0
5	1	3	2	0
5	2	7	4	0
6	1	1	2	0
6	2	8	3	0
7	1	4	5	0
8	1	3	6	0
9	2	1	10	2
9	2	2	10	2
9	2	3	10	2
10	2	1	9	2
10	2	2	9	2
10	2	3	9	2
10	2	4	1	1

The arrangement of yield data is the same as in Section 17.4.1 in the animal model analysis without groups.

Iterative stage

SOLVING FOR FIXED EFFECTS. This is exactly as described for the animal model without groups in Section 17.4.1, with yield records adjusted for other effects in the model and solutions for fixed effects computed.

SOLVING FOR ANIMAL SOLUTIONS. Solutions for animals are computed one at a time as both pedigree and data file sorted by animals are read, as described for the animal model without groups. Therefore, only the differences in terms of the way diagonals and ARHSs are accumulated are outlined.

For the kth animal in the pedigree file, calculate:

$$w_k = \alpha(4/(2 + \text{no. of unknown parents}))$$

For the type 1 record in the pedigree file for the kth animal:

$$\text{ARHS}_k = \text{ARHS}_k + (\hat{u}_s + \hat{u}_d)0.5w_k$$

$$\text{DIAG}_k = \text{DIAG}_k + w_k$$

For the type 2 record in the pedigree file for the kth animal:

$$\text{ARHS}_k = \text{ARHS}_k + (\hat{u}_o - 0.5\hat{u}_m)0.5w_k$$

Accumulation of ARHSs from the data file is as specified in Section 17.4.1 in the model without groups.

The solution for the kth animal is computed as $\text{ARHS}_k/\text{DIAG}_k$ when all records for the animal in the pedigree and data file have been read. The solutions in the first

round of iteration and at convergence without and with constraint on group solutions, as in Example 3.4, are as follows:

	Solutions		
Effects	At round 1	At convergence	At convergence[a]
Sex			
Male	4.333	4.509	5.474
Female	3.400	3.364	4.327
Animal			
1	0.333	0.182	–0.780
2	–0.083	0.026	–0.937
3	–0.021	–0.014	–0.977
4	–0.119	–0.319	–1.287
5	–0.376	–0.150	–1.113
6	0.392	0.221	–0.741
7	–0.364	–0.389	–1.355
8	0.282	0.181	–0.782
9	0.153	0.949	0.000
10	–0.176	–0.820	–1.795

[a]With solutions for groups constrained to those in Example 3.4.

When the solutions for groups are constrained as those in Example 3.4, this method gives the same solutions. However, when there is no constraint on group solutions, the ranking of animals is the same and linear differences between levels of effects are more or less the same as when there is a contraint on group solutions.

17.4.3 Reduced animal model with maternal effects

Example 17.5

The principles of genetic evaluation iterating on the data with a reduced animal model with maternal effects are illustrated using the same data set, parameters and model as in Example 6.2. The genetic parameters were:

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}g_{11} & g_{12} & 0 & 0\\ g_{21} & g_{22} & 0 & 0\\ 0 & 0 & \sigma^2_{pe} & 0\\ 0 & 0 & 0 & \sigma^2_e\end{bmatrix}=\begin{bmatrix}150 & -40 & 0 & 0\\ -40 & 90 & 0 & 0\\ 0 & 0 & 40 & 0\\ 0 & 0 & 0 & 350\end{bmatrix}$$

and:

$$\mathbf{G}^{-1}=\begin{bmatrix}g^{11} & g^{12}\\ g^{21} & g^{22}\end{bmatrix}=\begin{bmatrix}0.00756 & 0.00336\\ 0.00336 & 0.01261\end{bmatrix}$$

The inverse of the residual variance for parental records is $1/\sigma^2_e = r^{-1}_{pa} = 0.002857$ and for non-parental records is $1/(\sigma^2_e + dg_{11}) = r^{-1}_{np}$, where $d = 3/4$ or $1/2$ when one or both parents are known and the inverse of the variance due to permanent environmental effect = $1/\sigma^2_{pe} = 0.025$.

DATA ARRANGEMENT

The pedigree file is set up as described in Section 17.4.1 but only for animals that are parents. The pedigree file for the example data is:

Animal	Code	Sire or progeny	Dam or mate
1	1	0	0
1	2	5	2
1	2	9	6
2	1	0	0
2	2	5	1
2	2	6	3
3	1	0	0
3	2	6	2
3	2	8	5
4	1	0	0
4	2	7	6
5	1	1	2
5	2	8	3
6	1	3	2
6	2	7	4
6	2	9	1
7	1	4	6
8	1	3	5
9	1	1	6

A data file is set up consisting of a code to identify parents and non-parents. For non-parents, one record is set up for each parent, comprising the parent, a code indicating it is a non-parent, the animal that has the yield record, the other parent (mate), the sire and dam of the animal with the yield record, fixed effects, covariates (if any) and traits. A single record is set up for parents, comprising the animal, a code indicating it is a parent, the animal again, a field set to zero corresponding to the column for the other parent in non-parents' records, the sire and dam of the animal, fixed effects, covariates (if any) and traits. The data file may be sorted in three sequences if there is a major fixed effect in the model: sorted by major fixed effect, such as HYS (file A); sorted by animal (file B); and third sorted by dam code (file C). For the example, file A is:

Parent/ animal	Code[a]	Animal	Mate	Sire	Dam	Herd	Sex	Birth weight (kg)
5	0	5	0	1	2	1	Male	35.0
6	0	6	0	3	2	1	Female	20.0
7	0	7	0	4	6	1	Female	25.0
8	0	8	0	3	5	1	Male	40.0
9	0	9	0	1	6	2	Male	42.0
3	1	10	2	3	2	2	Female	22.0
2	1	10	3	3	2	2	Female	22.0
3	1	11	7	3	7	2	Female	35.0
7	1	11	3	3	7	2	Female	35.0
8	1	12	7	8	7	3	Female	34.0

Continued

(*Continued*)

Parent/ animal	Code[a]	Animal	Mate	Sire	Dam	Herd	Sex	Birth weight (kg)
7	1	12	8	8	7	3	Female	34.0
9	1	13	2	9	2	3	Male	20.0
2	1	13	9	9	2	3	Male	20.0
3	1	14	6	3	6	3	Female	40.0
6	1	14	3	3	6	3	Female	40.0

[a]0, parental record; 1, non-parental record.

Iteration stage

The solution vectors for herd ($\hat{\mathbf{hd}}$), sex ($\hat{\mathbf{b}}$), direct animal effect ($\hat{\mathbf{u}}$), genetic maternal effect ($\hat{\mathbf{m}}$) and permanent environmental effect ($\hat{\mathbf{pe}}$) are initially set to zero.

SOLVING FOR FIXED EFFECTS. Data file A is read at each round of iteration one herd at a time with ARHS and DIAG accumulated for the *i*th herd as:

$$\text{ARHS}_i = \text{ARHS}_i + r_{pa}^{-1}(y_{ijklt} - \hat{b}_j - \hat{u}_k - \hat{m}_l - \hat{pe}_t)$$

for parental records (Eqn 17.10):

$$\text{ARHS}_i = \text{ARHS}_i + r_{np}^{-1}(y_{ijklt} - \hat{b}_j - 0.5(\hat{u}_s + \hat{u}_d) - \hat{m}_l - \hat{pe}_t)$$

for non-parent records (Eqn 17.11):

$$\text{DIAG}_i = \text{DIAG}_i + r_n^{-1}$$

where r_n^{-1} is the inverse of the residual variance of the *n*th record being read.

At the end of records for the *i*th herd, the solution is computed as:

$$\hat{hd}_i = \text{ARHS}_i/\text{DIAG}_i$$

In the first round of iteration, the solution for the first herd is:

$$\begin{aligned}\hat{hd}_1 &= [r_{pa}^{-1}(y_1 - \hat{b}_1 - \hat{u}_5 - \hat{m}_2 - \hat{pe}_2) + (y_2 - \hat{b}_2 - \hat{u}_6 - \hat{m}_2 - \hat{pe}_2) \\ &\quad + (y_3 - \hat{b}_2 - \hat{u}_7 - \hat{m}_6 - \hat{pe}_6) + (y_4 - \hat{b}_1 - \hat{u}_8 - \hat{m}_5 - \hat{pe}_5)]/4(r_{pa}^{-1}) \\ &= [r_{pa}^{-1}((35 - 0 - 0 - 0 - 0) + (20 - 0 - 0 - 0 - 0) + (25 - 0 - 0 - 0 - 0) \\ &\quad + (40 - 0 - 0 - 0 - 0)]/4(r_{pa}^{-1}) \\ &= 0.3432/0.01144 = 30.00\end{aligned}$$

While reading data file A, ARHSs consisting of yield adjusted for previous animal, maternal and permanent environmental solutions are accumulated for each level of sex effect. Thus for the *j*th level of sex effect:

$$\text{ARHS}_j = \text{ARHS}_j + r_{pa}^{-1}(y_{ijklt} - \hat{u}_k - \hat{m}_l - \hat{pe}_t)$$

for parent records:

$$\text{ARHS}_j = \text{ARHS}_j + r_{np}^{-1}(y_{ijklt} - 0.5(\hat{u}_s + \hat{u}_d) - \hat{m}_l - \hat{pe}_t)$$

and for non-parent records:

$$\text{DIAG}_j = \text{DIAG}_j + r_n^{-1}$$

Afer reading file A, the solution for the j sex class is computed as:

$$\text{ARHS}_j = \text{ARHS}_j - nr_{ij}^{-1}\hat{h}d_i$$

$$\hat{b}_j = \text{ARHS}_j/\text{DIAG}_j$$

where $\hat{h}d_i$ is the current solution of herd i and nr_{ij}^{-1} is the sum of the inverse of the residual variance for records of the jth level of sex effect in herd i. The latter is accumulated while reading file A. For the example data, solutions for sex effect in the first round of iteration are:

$$\hat{b}_1 = \text{ARHS}_1 - 2r_{pa}^{-1}(\hat{h}d_1) - r_{pa}^{-1}(\hat{h}d_2) - 2r_{np}^{-1}(\hat{h}d_3)/[3r_{pa}^{-1} + 2r_{np}^{-1}]$$

$$= (0.38134 - 2r_{pa}^{-1}(30.0) - r_{pa}^{-1}(33.638) - 2r_{np}^{-1}(31.333)/0.01092$$

$$= 3.679$$

After obtaining solutions for fixed effects in the current round of iteration, the solutions for animals are solved for.

SOLVING FOR ANIMAL SOLUTIONS. As described in Section 17.4.1, animal solutions are computed one at a time as the pedigree file and file B are read. Briefly, for a type 1 record in the pedigree file for the kth animal, contributions to DIAG and ARHS according to the number of parents known (Eqn 17.8) are:

Number of parents known		
None	One (sire (s))	Both
$\text{ARHS}_k = 0$ $\text{DIAG}_k = g^{11}$	$\text{ARHS}_k = \frac{2}{3}g^{11}(\hat{u}_s)$ $\text{DIAG}_k = \frac{4}{3}g^{11}$	$\text{ARHS}_k = g^{11}(\hat{u}_s + \hat{u}_d)$ $\text{DIAG}_k = 2g^{11}$

where $\hat{u}_s$ and $\hat{u}_d$ are current solutions for direct effects for the sire and dam of the animal k.

The ARHS is augmented by contributions from the maternal effect as a result of the genetic correlation between animal and maternal effects. These contributions are from the sire, dam and the kth animal (see Eqn 17.9) and these are:

Number of parents known		
None	One (sire (s))	Both
–	$\text{ARHS}_k = \text{ARHS}_k + (\hat{m}_s)\frac{2}{3}g^{12}$	$\text{ARHS}_k = \text{ARHS}_k + (\hat{m}_s + \hat{m}_d)g^{12}$
$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)g^{12}$	$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)\frac{4}{3}g^{12}$	$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)2g^{12}$

where $\hat{m}_s$, $\hat{m}_d$ and $\hat{m}_k$ are current maternal solutions for the sire and dam of animal k respectively.

In processing a type 2 record in the pedigree file for the kth animal, contributions to DIAG and ARHS according to whether the mate of k is known are:

Mate is unknown	Mate is known
$ARHS_k = ARHS_k + \frac{2}{3} g^{11}(\hat{u}_o)$ $DIAG_k = DIAG_k + \frac{1}{3} g^{11}$	$ARHS_k = ARHS_k + (\hat{u}_o - 0.5\hat{u}_{ma})g^{11}$ $DIAG_k = DIAG_k + \frac{1}{2} g^{11}$

where $\hat{u}_o$ and $\hat{u}_{ma}$ are current solutions for direct effects for the progeny and mate of the animal k.

Accounting for contributions from the maternal effect to ARHS:

Mate is unknown	Mate is known
$ARHS_k = ARHS_k + \frac{2}{3} g^{12}(\hat{m}_o)$ $ARHS_k = ARHS_k - (\hat{m}_k)1/3g^{12}$	$ARHS_k = ARHS_k + (\hat{m}_o - 0.5\hat{m}_{ma})g^{12}$ $ARHS_k = ARHS_k - (\hat{m}_k)½g^{12}$

where $\hat{m}_o$ and $\hat{m}_{ma}$ are current maternal solutions for the progeny and mate of the animal k.

If the animal has a yield record:

$$DIAG_k = DIAG_k + r_n^{-1} \text{if it is a parent}$$

or:

$$DIAG_k = DIAG_k + (r_n^{-1})0.5 \text{ if it is a non-parent}$$

The diagonals of non-parents are multiplied by 0.5 instead of 0.25 because records of non-parents have been written twice (see Section 17.4).

Contributions to the RHS are accumulated as:

$$ARHS_k = ARHS_k + r_{pa}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - \hat{m}_l - \hat{p}e_t)$$

for parent records and:

$$ARHS_k = ARHS_k + r_{np}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - 0.5(\hat{u}_{ma}) - \hat{m}_l - \hat{p}e_t$$

for non-parent records. In the equations above, $\hat{h}d_i$, $\hat{b}_j$, $\hat{m}_l$, $\hat{p}e_t$ and $\hat{u}_{ma}$ are current solutions for herd i, jth level for sex effect, lth maternal effect level, tth level of permanent environment effect and animal solution for the other parent (mate), respectively. The solution for animal k is computed as usual when all records for the animal in the pedigree and data file have been read as:

$$\hat{u}_k = ARHS_k/DIAG_k$$

The solution for animal 2 in the example data in the first round of iteration is as follows.

The contribution to the diagonal from pedigree is:

$$DIAG_2 = (1 + \tfrac{1}{2} + \tfrac{1}{2})0.00756 = 0.01512$$

The contribution to the diagonal from yield is:

$$\text{DIAG}_2 = \text{DIAG}_2 + 2(0.00059) = 0.01512 + 0.00118 = 0.0163$$

The contribution to the ARHS from the pedigree is zero since both parents are unknown and solutions for progeny are zero in the first round of iteration. The contribution to ARHS from yield record is:

$$\begin{aligned}\text{ARHS}_2 &= r_{np}^{-1}(y_{10} - \hat{h}d_2 - \hat{b}_2 - \hat{u}_3 - \hat{m}_3 - \hat{p}e_3) \\ &\quad + r_{np}^{-1}(y_{13} - \hat{h}d_3 - \hat{b}_1 - \hat{u}_9 - \hat{m}_2 - \hat{p}e_2)\end{aligned}$$

$$\begin{aligned}\text{ARHS}_2 &= r_{np}^{-1}(22 - (-2.567) - 33.600 - 0 - 0 - 0\ ) \\ &\quad + r_{np}^{-1}(20 - 3.679 - 31.333) = -0.02818\end{aligned}$$

Therefore:

$$\hat{u}_2 = -0.02818/0.0163 = -1.729$$

After processing all animals in the pedigree and data file in the current round of iteration, equations for maternal effects are set and solved as described below.

SOLUTIONS FOR MATERNAL EFFECT. Solutions for maternal effects are computed using both the pedigree file and the data file sorted by dam. Records for the *l*th animal are read in from the pedigree file and from file C if it is a dam that has progeny with a yield record, while accumulating DIAG and ARHS. For the type 1 record in the pedigree file for animal *l*, contributions to ARHS and DIAG according to the number of parents known are as follows:

Number of parents known		
None	One (dam(d))	Both
$\text{ARHS}_l = 0$ $\text{DIAG}_l = g^{22}$	$\text{ARHS}_l = \frac{2}{3}\, g^{22}(\hat{m}_d)$ $\text{DIAG}_l = \frac{4}{3}\, g^{22}$	$\text{ARHS}_l = g^{22}(\hat{m}_s + \hat{m}_d)$ $\text{DIAG}_l = 2g^{22}$

Taking into account contributions from animal effects to the ARHS due to genetic correlation gives:

Number of parents known		
None	One (dam(d))	Both
– $\text{ARHS}_l = \text{ARHS}_l - (\hat{u}_l)g^{12}$	$\text{ARHS}_l = \text{ARHS}_l + (\hat{u}_d)\ \frac{2}{3}\ g^{12}$ $\text{ARHS}_l = \text{ARHS}_l - (\hat{u}_l)\ \frac{4}{3}\ g^{12}$	$\text{ARHS}_l = \text{ARHS}_l + (\hat{u}_s + \hat{u}_d)g^{12}$ $\text{ARHS}_l = \text{ARHS}_l - (\hat{u}_l)2g^{12}$

For the type 2 record in the pedigree file for animal *l*, contributions to the ARHS and DIAG according to whether the mate of animal *l* is known or not are:

Mate is unknown	Mate is known
$\text{ARHS}_l = \text{ARHS}_l + (\frac{2}{3})g^{22}(\hat{m}_o)$ $\text{DIAG}_l = \text{DIAG}_l + (\frac{1}{3})g^{22}$	$\text{ARHS}_l = \text{ARHS}_l + g^{22}(\hat{m}_o - 0.5\hat{m}_{ma})$ $\text{DIAG}_l = \text{DIAG}_l + (1/2)g^{22}$

Taking into account contributions from animal effect (see Eqn 17.6) gives:

Mate is unknown	Mate is known
$ARHS_l = ARHS_l + (\frac{2}{3})g^{12}(\hat{u}_o)$	$ARHS_l = ARHS_l + (\hat{u}_o - 0.5\hat{u}_{ma})g^{12}$
$ARHS_l = ARHS_l - (\hat{u}_l)(\frac{1}{3})g^{12}$	$ARHS_l = ARHS_l - (\hat{u}_l)(\frac{1}{2})g^{12}$

For the animal *l*, which is a dam with progeny having yield records, DIAG and ARHS from the pedigree is augmented with information from yield as:

$$DIAG = DIAG + r_n^{-1}$$

and:

$$ARHS_l = ARHS_l + r_{pa}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - \hat{u}_k - \hat{p}e_t)$$

for parent records and:

$$ARHS_l = ARHS_l + r_{np}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - 0.5(\hat{u}_s + \hat{u}_d) - \hat{p}e_t)$$

for non-parent records.

After processing all records from pedigree and yield records for the *l*th animal, the solution for the maternal effect is computed as:

$$\hat{m}_l = ARHS_l/DIAG_l$$

The calculation of the solution for animal 5 in the first round of iteration is as follows.

The contribution from a type 1 record in the pedigree is:

$$\begin{aligned} ARHS_5 &= (\hat{m}_1 + \hat{m}_2)g^{22} + (\hat{u}_1 + \hat{u}_2)g^{12} - (\hat{u}_5 2g^{12}) \\ &= (0.0217 + -1.7027)0.01261 + (0 + (-1.7294))0.00336 - ((-0.5831)(2)0.0336) \\ &= -0.02309 \end{aligned}$$

$$DIAG_5 = (2)0.01261 = 0.02522$$

The contribution from a type 2 record in the pedigree is:

$$\begin{aligned} ARHS_5 &= ARHS_5 + (\hat{m}_8 - \tfrac{1}{2}\hat{m}_3)g^{22} + (\hat{u}_8 - \tfrac{1}{2}\hat{u}_3)g^{12} - (\hat{u}_5 \tfrac{1}{2} g^{12}) \\ &= -0.02309 + (0 - \tfrac{1}{2}(0.4587))0.01261 \\ &\quad + (1.4382 - \tfrac{1}{2}(0.8960))0.00336 - ((-0.5831)(\tfrac{1}{2})0.00336) \\ &= -0.021675 \end{aligned}$$

$$DIAG_5 = DIAG_5 + \tfrac{1}{2} g^{22} = 0.02522 + 0.0063 = 0.03153$$

The contribution from yield of progeny (animal 8) for dam 5 is:

$$\begin{aligned} ARHS_5 &= ARHS_5 + r_{pa}^{-1}(y_8 - \hat{h}d_1 - \hat{b}_1 - \hat{u}_8 - \hat{p}e_5) \\ &= -0.021675 + r_{pa}^{-1}(40 - 30.00 - 3.679 - (1.4382) - 0) \\ &= -0.007724 \end{aligned}$$

$$DIAG_5 = DIAG_5 + r_{pa}^{-1} = 0.03153 + 0.002857 = 0.034387$$

and the solution is:

$$\hat{m}_5 = -0.007724/0.034387 = -0.225$$

Solutions for permanent environmental effects are solved for after processing all animals for maternal effects in the current round of iteration.

SOLVING FOR PERMANENT ENVIRONMENTAL *(pe)* EFFECTS. Only the data file sorted by dams is required to obtain solutions for pe effects.

The records for the tth dam are read from file C while ARHS and DIAG are accumulated as:

$$\text{ARHS}_t = \text{ARHS}_t + r_{pa}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - \hat{u}_k - \hat{m}_l)$$

for parent records and:

$$\text{ARHS}_t = \text{ARHS}_t + r_{np}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - 0.5(\hat{u}_s + \hat{u}_d) - \hat{m}_l)$$

for non-parent records.

$$\text{DIAG}_t = \text{DIAG}_t + r_n^{-1}$$

At the end of records for the tth dam, solutions are computed as:

$$\hat{p}e_t = \text{ARHS}_t/(\text{DIAG}_t + 1/\sigma_p^2)$$

The solution for permanent environmental effect for animal 5 in the first round of iteration is:

$$\begin{aligned}\text{ARHS}_5 &= r_{pa}^{-1}(y_8 - \hat{h}d_1 - \hat{b}_1 - \hat{u}_8 - \hat{m}_5)\\ &= r_{pa}^{-1}(40 - 3.679 - 30.0 - 1.4822 - (-0.2246))\\ &= 0.01459\end{aligned}$$

$$\text{DIAG}_5 = r_{pa}^{-1} + 0.025 = 0.02786$$

and:

$$\hat{p}e_5 = 0.01459/0.02786 = 0.524$$

Further iterations are carried out until convergence is achieved. The convergence criteria defined in Section 17.3.1 could also be used. The solutions for the first round of iteration and at convergence are shown below.

	Solutions	
Effects	At round 1	At convergence
Herd		
1	30.000	30.563
2	33.600	33.950
3	31.333	31.997
Sex of calf		
Male	3.679	3.977
Female	–2.657	–2.872
Animal		
1	0.000	0.564
2	–1.729	–1.246
3	0.896	1.166
4	0.000	–0.484
5	–0.583	0.630
6	–0.554	–0.859

Continued

(*Continued*)

Effects	Solutions	
	At round 1	At convergence
7	–0.020	–1.156
8	1.438	1.918
9	–0.396	–0.553
Maternal		
1	0.022	0.261
2	–1.703	–1.582
3	0.459	0.735
4	0.046	0.586
5	–0.225	–0.507
6	0.425	0.841
7	0.788	1.299
8	–0.224	–0.158
9	0.255	0.659
Permanent environment		
2	–1.386	–1.701
5	0.524	0.415
6	0.931	0.825
7	0.527	0.461

These solutions are exactly the same obtained as those obtained in Section 7.3 by directly inverting the coefficient matrix.

Back-solving for non-parents

The solutions for direct animal and maternal effects for non-parents are calculated after convergence has been achieved, as described in Section 7.3. The solutions for non-parents for this example have been calculated in Section 7.3.

17.5 Preconditioned Conjugate Gradient Algorithm

Berger *et al*. (1989) investigated the use of the plain or Jacobi conjugate gradient iterative scheme for solving MME for the prediction of sire breeding values. They indicated that plain conjugate gradient was superior to a number of other iterative schemes, including Gauss–Seidel. Strandén and Lidauer (1999) implemented the use of the preconditioned conjugate gradient (PCG) in genetic evaluation models for the routine evaluation of dairy cattle with very large data. In the PCG method, the linear systems of equations (Eqn 17.1, for instance) is made simpler by solving an equivalent system of equations:

$$\mathbf{M}^{-1}\mathbf{Cb} = \mathbf{M}^{-1}\mathbf{r}$$

where **M** is a symmetric, positive definite, preconditioner matrix that approximates **C** and **r** is the right-hand side. In the plain conjugate gradient method, the preconditioner **M** is an identity matrix.

The implementation of the PCG method requires storing four vectors of size equal to the number of unknowns in the MME: a vector of residuals (**e**), a search-direction vector (**d**), a solution vector (**b**) and a work vector (**v**). The PCG method can be implemented with less memory by storing the solution vector on disk and reading it in during the iteration. The pseudo-code for the PCG method (Lidauer *et al*., 1999) is outlined below, assuming that starting values are:

$$\mathbf{b}^{(0)} = 0, \quad \mathbf{e}^{(0)} = \mathbf{r} - \mathbf{C}\mathbf{b} = \mathbf{r}, \quad \mathbf{d}^{(0)} = \mathbf{M}^{-1}\mathbf{e}^{(0)} = \mathbf{M}^{-1}\mathbf{r}^{(0)}$$

For $k = 1,2,\ldots, n$

$$\mathbf{v} = \mathbf{C}\mathbf{d}^{(k-1)}$$

$$\omega = \mathbf{e}'^{(k-1)}\mathbf{M}^{-1}\mathbf{e}^{(k-1)}/(\mathbf{d}'^{(k-1)}\mathbf{v})$$

$$\mathbf{b}^{(k)} = \mathbf{b}^{(k-1)} + \omega\mathbf{d}^{(k-1)}$$

$$\mathbf{e}^{(k)} = \mathbf{e}^{(k-1)} - \omega\mathbf{v}$$

$$\mathbf{v} = \mathbf{M}^{-1}\mathbf{e}^{(k)}$$

$$\beta = \mathbf{e}'^{(k)}\mathbf{v}/(\mathbf{e}'^{(k-1)}\mathbf{M}^{-1}\mathbf{e}^{(k-1)})$$

$$\mathbf{d}^{(k)} = \mathbf{v} + \beta\mathbf{d}^{(k-1)}$$

If not converged, continue iteration until converged, and ω and β are step sizes in the PCG method.

17.5.1 Computation strategy

The major task in the PCG algorithm above is calculating **Cd**, where **C** is the coefficient matrix of the MME. The vector **d** is the search direction vector and every iteration of the PCG minimizes the distance between the current and the true solutions in the search direction. Strandén and Lidauer (1999) presented an efficient computation strategy for computing **Cd** for a multivariate model. Assuming, for instance, that data are ordered by animals, the MME for the multivariate model (Eqn 5.2) can be written as:

$$\begin{pmatrix} \sum_{i=1}^{N}\mathbf{x}_i\mathbf{R}_i^{-1}\mathbf{x}_i' & \sum_{i=1}^{N}\mathbf{x}_i\mathbf{R}_i^{-1}\mathbf{z}_i' \\ \sum_{i=1}^{N}\mathbf{z}_i\mathbf{R}_i^{-1}\mathbf{x}_i' & \sum_{i=1}^{N}\mathbf{z}_i\mathbf{R}_i^{-1}\mathbf{z}'+\mathbf{A}^{-1}\otimes\mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N}\mathbf{x}_i\mathbf{R}_i^{-1}\mathbf{y}_i \\ \sum_{i=1}^{N}\mathbf{z}_i\mathbf{R}_i^{-1}\mathbf{y}_i \end{pmatrix}$$

where N is the number of animals with records, $\mathbf{x}_i'$ and $\mathbf{z}'_i$ are matrices having rows with l_i equal to the number of traits observed on animal i. Denote $\mathbf{w}' = [\mathbf{x}_i' \; \mathbf{z}_i']$ and **V** as:

$$\mathbf{V} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{A}\otimes\mathbf{G} \end{pmatrix}$$

Computing **Cd** then implies calculating:

$$\sum_{i=1}^{N} \mathbf{w}_i \mathbf{R}_i^{-1} \mathbf{w}_i' \mathbf{d} + \mathbf{V}^{-1}\mathbf{d} = \sum_{i=1}^{N} \mathbf{v}_i + \mathbf{v}_d \tag{17.13}$$

If solving the MME with iteration on the data for a univariate model without any regression effects, this calculation can be achieved by accumulating for each individual i, the product $\mathbf{v}_i = \mathbf{T}_i\mathbf{d}$, where the coefficients in $\mathbf{T}_i = \mathbf{w}_i\mathbf{R}_i^{-1}\mathbf{w}_i'$ can be deduced without performing any of the products, as $\mathbf{w}_i$ contains zeros and ones only and $\mathbf{R}_i^{-1}$ is a scalar or $\mathbf{R}_i^{-1}$ is factored out (Eqn 3.4). For a multivariate model, the principles for computing $\mathbf{T}_i$ are essentially the same but with scalar contributions replaced by matrix $\mathbf{R}_i$. Strandén and Lidauer (1999) suggested the following three-step method for calculating the product $\mathbf{w}_i\mathbf{R}_i^{-1}\mathbf{d}$:

$$\mathbf{s}_i \leftarrow \mathbf{w}_i'\mathbf{d}; \quad \mathbf{s}_i^* \leftarrow \mathbf{R}_i^{-1}\mathbf{s}_i; \quad \mathbf{v}_i \leftarrow \mathbf{w}_i\mathbf{s}_i^*$$

where vectors $\mathbf{s}_i$ and $\mathbf{s}_i^*$ are of size equal to the number of traits observed on individual i (l_i). They demonstrated that this three-step approach reduced substantially the number of floating point operations (multiplications) compared with a multivariate accumulation technique as used by Groeneveld and Kovac (1990). For instance, given that q_i is the number of effects over traits observed for individual i, the number of floating point operations were 720 with $l_i = 3$ and $q_i = 15$ using the multivariate accumulation technique compared with 78 with the three-point approach. They also suggested that $\mathbf{v}_d = \mathbf{V}^{-1}\mathbf{d}$ in Eqn 17.13 can be evaluated in a two-step approach:

$$\mathbf{x} \leftarrow (\mathbf{I} \otimes \mathbf{A}^{-1})\mathbf{d}; \quad \mathbf{v}_d \leftarrow (\mathbf{G}^{-1} \otimes \mathbf{I})\mathbf{x}$$

17.5.2 Numerical application

Example 17.6

The application of PCG to solve MME is illustrated using data for Example 3.1 for a univariate model and iterating on the data.

COMPUTING STARTING VALUES

Initially, the pedigree is read and diagonal elements of $\mathbf{A}^{-1}$ multiplied by α are accumulated for animals, where the variance ratio α is 2, as in Example 3.1. This is straightforward and has not been illustrated, but elements for animals 1 to 8 stored in a vector **h** are:

$$\mathbf{h}' = [3.667 \quad 4.0 \quad 4.0 \quad 3.667 \quad 5.0 \quad 5.0 \quad 4.0 \quad 4.0]$$

Second, read through the data as shown in Table 3.1 and accumulate right-hand side (**r**) for all effects, diagonals for the levels of sex of calf effect and add contribution of information from data to diagonals from $\mathbf{A}^{-1}\alpha$ for animals. Assuming that diagonals for all effects are stored as diagonal elements of **M**, such that the first two elements are for the two levels of sex of calf effect and the remaining elements for animals 1 to 8, then **r** and **M** are:

$$\mathbf{r}' = [13.0 \quad 6.8 \quad 0.0 \quad 0.0 \quad 0.0 \quad 4.5 \quad 2.9 \quad 3.9 \quad 3.5 \quad 5.0]$$

and:

$$\mathbf{M} = \text{diag}[3.0 \quad 2.0 \quad 3.667 \quad 4.0 \quad 4.0 \quad 4.667 \quad 6.0 \quad 6.0 \quad 5.0 \quad 5.0]$$

The starting values for PCG can now be calculated. Thus:

$$\mathbf{b}^{(0)} = 0, \quad \mathbf{e}^{(0)} = \mathbf{r} - \mathbf{C}\mathbf{b}^{(0)} = \mathbf{y} \quad \text{and} \quad \mathbf{d}^{(0)} = \mathbf{M}^{-1}\mathbf{r}$$

Thus:

$$\mathbf{d}^{(0)\prime} = [4.333 \quad 3.4 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.964 \quad 0.483 \quad 0.650 \quad 0.70 \quad 1.0]$$

Iterative stage

Reading through the data and performing the following calculations in each round of iteration, start the PCG iterative process. Calculations are shown for the first round of iteration.

The vector $\mathbf{v} = \mathbf{C}\mathbf{d}$ is accumulated as data are read. For the ith level of fixed effect:

$$\mathbf{v}(i) = \mathbf{v}(i) + 1(\mathbf{d}(i)) + 1(\mathbf{d}(\text{anim}_k))$$

where anim_k refers to the animal k associated with the record. Thus for the level 1 of sex of calf effect:

$$\mathbf{v}(1) = 3(4.333) + \mathbf{d}(\text{anim}_4) + \mathbf{d}(\text{anim}_7) + \mathbf{d}(\text{anim}_8) = 15.663$$

As each record is read, calculate:

$z = 4/(2 + \text{number of unknown parents for animal with record})$

$xx = -0.5(z)\alpha$ if either parent is known, otherwise $xx = 0$

$xm = 0.25(z)\alpha$ if both parents are known, otherwise $xm = 0$

If only one parent, p, of animal k is known, then accumulate:

$$\mathbf{v}(\text{anim}_k) = \mathbf{v}(\text{anim}_k) + 1(\mathbf{d}(i)) + \mathbf{M}_{k,k}(\mathbf{d}(\text{anim}_k)) + xx(\mathbf{d}(\text{anim}_p)) \tag{17.14}$$

where $\mathbf{d}(i)$ refers to the ith level of the fixed effect and $\mathbf{M}_{k,k}$ the diagonal element of $\mathbf{M}$ for animal k.

Accumulate the contribution to the known parent, p, of k at the same time:

$$\mathbf{v}(\text{anim}_p) = \mathbf{v}(\text{anim}_p) + xx(\mathbf{d}(\text{anim}_k))$$

If both parents p and j of animal k are known, then accumulate for animal k as:

$$\mathbf{v}(\text{anim}_k) = \mathbf{v}(\text{anim}_k) + 1(\mathbf{d}(i)) + \mathbf{M}_{k,k}(\mathbf{d}(\text{anim}_k)) + xx(\mathbf{d}(\text{anim}_p) + \mathbf{d}(\text{anim}_j)) \tag{17.15}$$

Accumulate for both parents as:

$$\mathbf{v}(\text{anim}_p) = \mathbf{v}(\text{anim}_p) + xx(\mathbf{d}(\text{anim}_k))$$

$$\mathbf{v}(\text{anim}_p) = \mathbf{v}(\text{anim}_p) + xm(\mathbf{d}(\text{anim}_j))$$

$$\mathbf{v}(\text{anim}_j) = \mathbf{v}(\text{anim}_j) + xx(\mathbf{d}(\text{anim}_k))$$

$$\mathbf{v}(\text{anim}_j) = \mathbf{v}(\text{anim}_j) + xm(\mathbf{d}(\text{anim}_p))$$

After processing all animals with records, the contribution for animals in the pedigree without records is accumulated. The equations for accumulating contributions for these animals is the same as shown above except that the coefficient for $\mathbf{d}(i)$ in Eqns 17.13 and 17.14 is zero instead of one, indicating no contribution from records.

For example, for animal 4 with only the sire known:

$$\mathbf{v}(4) = \mathbf{v}(4) + \mathbf{d}(1) + \mathbf{M}_{4,4}(\mathbf{d}(\text{anim}_4)) + (-2/3)\alpha(\mathbf{d}(\text{anim}_1)) = 8.833$$

Add contribution from progeny when processing the record for animal 7:

$$\mathbf{v}(4) = 8.833 + -1.0\alpha(\mathbf{d}(\text{anim}_7)) + 0.25\alpha(\mathbf{d}(\text{anim}_5)) = 7.917$$

The vector $\mathbf{v}$ for all effects is:

$$\mathbf{v}' = [15.664\ 7.933\ -2.586\ -2.267\ -2.317\ 7.917\ 5.864\ 5.300\ 4.938\ 8.033]$$

Next ω is computed using matrix multiplication and scalar division as:

$$\omega = 95.1793/120.255 = 0.7915$$

The solution vector is then computed as $\mathbf{b}^{(1)} = \mathbf{b}^{(0)} + \omega\mathbf{d}^{(0)}$. The vector $\mathbf{b}^{(1)}$ is:

$$\mathbf{b}'^{(1)} = [3.430\quad 2.691\quad 0.0\quad 0.0\quad 0.0\quad 0.763\quad 0.383\quad 0.514\quad 0.554\quad 0.791]$$

The updated vector of residuals $\mathbf{e}^{(1)}$ is computed as $\mathbf{e}^{(0)} - \omega\mathbf{v}$. For the example data $\mathbf{e}^{(1)}$ is:

$$\mathbf{e}'^{(1)} = [0.602\ 0.521\ 2.047\ 1.794\ 1.834\ -1.766\ -1.741\ -0.295\ -0.408\ -1.358]$$

The vector $\mathbf{v}$ is then computed as $\mathbf{M}^{-1}\mathbf{e}^{(1)}$. For the example data, $\mathbf{v}$ is:

$$\mathbf{v}' = [0.201\ 0.260\ 0.558\ 0.449\ 0.458\ -0.378\ -0.290\ -0.049\ -0.082\ -0.272]$$

Next, compute the scalar β. The denominator of β is equal to the numerator of ω and this has already been computed. Using the example data:

$$\beta = 4.634/95.179 = 0.0487$$

Finally, $\mathbf{d}^{(1)}$, the search-direction vector for the next iteration is computed as $\mathbf{v} + \beta\mathbf{d}^{(0)}$. This vector for the example data is:

$$\mathbf{d}'^{(1)} = [0.412\ 0.426\ 0.558\ 0.449\ 0.458\ -0.331\ -0.267\ -0.017\ -0.048\ -0.223]$$

The next cycle of iteration is continued until the system of equations converges. Convergence can either be monitored using the criteria defined in Example 17.1 or the relative difference between the right-hand and left-hand sides:

$$c_d^{(r)} = \frac{\left\|\mathbf{y} - \mathbf{C}\mathbf{b}^{(r+1)}\right\|}{\|\mathbf{y}\|}$$

where:

$$\|\mathbf{x}\| = \left(\sum_i x_i^2\right)^{1/2}$$

Using the convergence criteria used in Example 17.1, the iteration was stopped at the 10th iteration when equations converged to 8.3^{-07}. Some intermediary and final solutions are shown in the following table.

Effects	Iteration number				
	1	3	5	7	10
Sex of calf					
Male	3.430	3.835	4.280	4.367	4.359
Female	2.691	3.122	3.154	3.377	3.404
Animals					
1	0.000	0.475	0.170	0.092	0.098
2	0.000	0.224	0.116	0.012	–0.019
3	0.000	0.272	0.058	–0.056	–0.041
4	0.763	0.390	0.032	–0.029	–0.009
5	0.383	0.249	–0.072	–0.155	–0.186
6	0.514	0.547	0.435	0.194	0.177
7	0.554	0.193	–0.178	–0.231	–0.249
8	0.791	0.537	0.334	0.171	0.183

The equations converged at the 10th round of iteration compared with 20 iterations on the data in Example 17.3.

This page intentionally left blank

Appendix A: Introduction to Matrix Algebra

The basic elements of matrix algebra necessary to understand the principles involved in the prediction of breeding values are briefly covered in this appendix. Little or no previous knowledge of matrix algebra is assumed. For a detailed study of matrix algebra, see Searle (1982).

A.1 Matrix: A Definition

A matrix is a rectangular array of numbers set in rows and columns. These elements are called the elements of a matrix. The matrix **B**, for instance, consisting of two rows and three columns, may represented as:

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

or:

$$\mathbf{B} = \begin{bmatrix} 2 & 4 & 5 \\ 6 & 8 & 9 \end{bmatrix}$$

The element b_{ij} is called the ij element of the matrix, the first subscript referring to the row the element is in and the second to the column. The order of a matrix is the number of rows and columns. Thus a matrix of r rows and c columns has order $r \times c$ (read as r by c). The matrix **B** above is of the order 2×3 and can be written as $\mathbf{B}_{2\times3}$.

A matrix consisting of a single row of elements is called a row vector. A row vector consisting of three elements may be represented as:

$$\mathbf{c} = \begin{bmatrix} 2 & 6 & -4 \end{bmatrix}$$

Only one subscript is needed to specify the position of an element in a row vector. Thus the ith element in the row vector **c** above refers to the element in the ith column. For instance, $\mathbf{c}_3 = -4$.

Similarly, a matrix consisting of a single column is called a column vector. Again, only one supscript is needed to specify the position of an element, which refers to the row the element is in, since there is only one column. A column vector **d** with four elements can be shown as below:

$$\mathbf{d} = \begin{bmatrix} -20 \\ 60 \\ 8 \\ 2 \end{bmatrix}$$

A scalar is a matrix with one row and one column.

A.2 Special Matrices

A.2.1 Square matrix

A matrix with an equal number of rows and columns is referred to as a square matrix. Shown below is a square matrix **G** of order 3 × 3.

$$\mathbf{G} = \begin{bmatrix} 2 & 1 & 6 \\ 4 & 2 & 7 \\ 0 & 4 & 8 \end{bmatrix}$$

The *ij* elements in a square matrix with *i* equal to *j* are called the diagonal elements. Other elements of the square matrix are called off-diagonal or non-diagonal elements. Thus the diagonal elements in the **G** matrix above are 2, 2 and 8.

A.2.2 Diagonal matrix

A square matrix having zero for all of its off-diagonal elements is referred to as a diagonal matrix. For example, a diagonal matrix **B** can be shown as below:

$$\mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 18 \end{bmatrix}$$

When all the diagonal elements of a diagonal matrix are one, it is referred to as an identity matrix. Given below is an identity matrix, **I**:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A.2.3 Triangular matrix

A square matrix with all elements above the diagonal being zero is called a lower triangular matrix. When all the elements below the diagonal are zeros, it is referred to as an upper triangular matrix. For instance, the matrices **D**, a lower triangular matrix and **E**, an upper triangular matrix, can be illustrated as:

$$\mathbf{D} = \begin{bmatrix} 4 & 0 & 0 \\ 1 & 3 & 0 \\ -2 & 7 & 9 \end{bmatrix}; \quad \mathbf{E} = \begin{bmatrix} 3 & 9 & 1 \\ 0 & 4 & 8 \\ 0 & 0 & 6 \end{bmatrix}$$

The transpose (see A.3.1) of an upper triangular matrix is a lower triangular matrix and vice versa.

A.2.4 Symmetric matrix

A symmetric matrix is a square matrix with the elements above the diagonal equal to the corresponding elements below the diagonal, i.e. element *ij* is equal to element *ji*. The matrix **A** below is an example of a symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 2 & -4 & 0 \\ -4 & 6 & 3 \\ 0 & 3 & 7 \end{bmatrix}$$

A.3 Basic Matrix Operations

A.3.1 Transpose of a matrix

The transpose of a matrix **A** is usually written as $\mathbf{A}'$ or $\mathbf{A}^{\mathrm{T}}$ and is the matrix whose *ji* elements are the *ij* elements of the original matrix, i.e. $a'_{ji} = a_{ij}$. In other words, the columns of $\mathbf{A}'$ are the rows of **A** and the rows of $\mathbf{A}'$ the columns of **A**. For instance, the matrix **A** and its transpose $\mathbf{A}'$ are illustrated below:

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \\ 4 & 0 \end{bmatrix}; \quad \mathbf{A}' = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & 0 \end{bmatrix}$$

Note that **A** is not equal to $\mathbf{A}'$ but the transpose of a symmetric matrix is equal to the symmetric matrix. Also $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, where **AB** refers to the product (see A.3.3) of **A** and **B**.

A.3.2 Matrix addition and subtraction

Two matrices can be added together *only* if they have the same number of rows and columns, i.e. they are of the same order and they are said to be conformable for addition. Given that **W** is the sum of the matrices **X** and **Y**, then $w_{ij} = x_{ij} + y_{ij}$. For example, if **X** and **Y**, both of order 2 × 2, are as illustrated below:

$$\mathbf{X} = \begin{bmatrix} 40 & 10 \\ 39 & -25 \end{bmatrix}; \quad \mathbf{Y} = \begin{bmatrix} -2 & 20 \\ 4 & 40 \end{bmatrix}$$

Then the matrix **W**, the sum of **X** and **Y**, is:

$$\mathbf{W} = \begin{bmatrix} 40 + (-2) & 10 + 20 \\ 39 + 4 & -25 + 40 \end{bmatrix} = \begin{bmatrix} 38 & 30 \\ 43 & 15 \end{bmatrix}$$

Matrix subtraction follows the same principles used for matrix addition. If **B** = **X** – **Y**, then $b_{ij} = x_{ij} - y_{ij}$. Thus the matrix **B** obtained by subtracting **Y** from **X** above is:

$$\mathbf{B} = \mathbf{X} - \mathbf{Y} = \begin{bmatrix} 40 - (-2) & 10 - 20 \\ 39 - 4 & -25 - 40 \end{bmatrix} = \begin{bmatrix} 42 & -10 \\ 35 & -65 \end{bmatrix}$$

A.3.3 Matrix multiplication

Two matrices can be multiplied only if the number of columns in the first matrix equals the number of rows in the second. The order of the product matrix is equal to the number of rows of the first matrix by the number of columns in the second. Given that **C** = **AB**, then:

$$\mathbf{C} = c_{ij} = \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{z} a_{ik} b_{kj}$$

where m = number of columns in **B**, n = number of rows in **A** and z = number of rows in **B**. Let:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & -1 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 2 & 5 \\ 4 & 3 \\ 6 & 1 \end{bmatrix}$$

Then **C** can be obtained as:

$c_{11} = 1(2) + 4(4) + -1(6) = 12$ (row 1 of **A** multiplied by column 1 of **B**)
$c_{21} = 2(2) + 5(4) + 0(6) = 24$ (row 2 of **A** multiplied by column 1 of **B**)
$c_{31} = 3(2) + 6(4) + 1(6) = 36$ (row 3 of **A** multiplied by column 1 of **B**)
$c_{12} = 1(5) + 4(3) + -1(1) = 16$ (row 1 of **A** multiplied by column 2 of **B**)
$c_{22} = 2(5) + 5(3) + 1(1) = 26$ (row 2 of **A** multiplied by column 2 of **B**)
$c_{23} = 3(5) + 6(3) + 1(1) = 34$ (row 3 of **A** multiplied by column 2 of **B**)

$$\mathbf{C} = \begin{bmatrix} 12 & 16 \\ 24 & 26 \\ 36 & 34 \end{bmatrix}$$

Note that **C** has order 3 × 2, where 3 equals the number of rows of **A** and 2 the number of columns in **B**. Also, note that **AB** is not equal to **BA**, but **IA** = **AI** = **A**, where **I** is an identity matrix. If **M** is the product of a scalar g and a matrix **B**, then **M** = $b_{ij}g$, i.e. each element of **M** equals the corresponding element in **B** multiplied by g.

A.3.4 Direct product of matrices

Given a matrix **G** of order n by m and **A** of order t by s, the direct product is:

$$\mathbf{G} \otimes \mathrm{A} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} \\ g_{21}\mathbf{A} & g_{22}\mathbf{A} \end{bmatrix}$$

The direct product is also known as the Kronecker product and is of the order *nt* by *ms*. For instance, assuming that:

$$\mathbf{G} = \begin{bmatrix} 10 & 5 \\ 5 & 20 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 4 \\ 2 & 4 & 1 \end{bmatrix}$$

the Kronecker product is:

$$\mathbf{G} \otimes \mathbf{A} = \begin{bmatrix} 10 & 0 & 20 & 5 & 0 & 10 \\ 0 & 10 & 40 & 0 & 5 & 20 \\ 20 & 40 & 10 & 10 & 20 & 5 \\ 5 & 0 & 10 & 20 & 0 & 40 \\ 0 & 5 & 20 & 0 & 20 & 80 \\ 10 & 20 & 5 & 20 & 80 & 20 \end{bmatrix}$$

The Kronecker product is useful in multiple trait evaluations.

A.3.5 Matrix inversion

An inverse matrix is one which when multiplied by the original matrix gives an identity matrix as the product. The inverse of a matrix **A** is usually denoted as $\mathbf{A}^{-1}$ and, from the above definition, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, where **I** is an identity matrix. Only square matrices can be inverted and for a diagonal matrix the inverse is calculated simply as the reciprocal of the diagonal elements. For instance, the diagonal matrix **B** and its inverse are:

$$\mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 18 \end{bmatrix} \quad \text{and} \quad \mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{18} \end{bmatrix}$$

For a 2 × 2 matrix, the inverse is easy to calculate and is illustrated below. Let:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

First, calculate the determinant, which is the difference between the product of the two diagonal elements and the two off-diagonal elements ($a_{11}a_{22} - a_{12}a_{21}$). Second, the inverse is obtained by reversing the diagonal elements, multiplying the off-diagonal elements by –1 and dividing all elements by the determinant. Thus:

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

For instance, given that:

$$\mathbf{A} = \begin{bmatrix} 8 & 4 \\ 6 & 4 \end{bmatrix} \text{ then } \mathbf{A}^{-1} = \frac{1}{(8)(4)-(6)(4)} \begin{bmatrix} 4 & -4 \\ -6 & 8 \end{bmatrix} = \begin{bmatrix} 0.50 & -0.50 \\ -0.75 & 1.00 \end{bmatrix}$$

Note that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$, as stated earlier. Calculating the inverse of a matrix becomes more cumbersome as the order increases, and inverses are usually obtained using computer programs. The methodology has not been covered in this text. It is obvious from the above that an inverse of a non-diagonal matrix cannot be calculated if the determinant is equal to zero. A square matrix with a determinant equal to zero is said to be singular and does not have an inverse. A matrix with a non-zero determinant is said to be non-singular.

Note that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. The inverses of matrices may be required when solving linear equations. Thus given the following linear equation:

$$\mathbf{Ab} = \mathbf{y}$$

pre-multiplying both sides by $\mathbf{A}^{-1}$ gives the vector of solutions **b** as:

$$\mathbf{b} = \mathbf{A}^{-1}\mathbf{y}$$

A.3.6 Rank of a matrix

The rank of a matrix is the number of linearly independent rows or columns. A square matrix with the rank equal to the number of rows or columns is said to be of full rank. In some matrices, some of the rows or columns are linear combinations of other rows or columns; therefore, the rank is less than the number of rows or columns. Such a matrix is not of full rank. Consider the following set of equations:

$$\begin{aligned} 3x_1 + 2x_2 + 1x_3 &= y_1 \\ 4x_1 + 3x_2 + 0x_3 &= y_2 \\ 7x_1 + 5x_2 + 1x_3 &= y_3 \end{aligned}$$

The third equation is the sum of the first and second equations; therefore, the vector of solutions, $\mathbf{x}(\mathbf{x}' = [x_1\ x_2\ x_3])$, cannot be estimated due to the lack of information. In other words, if the system of equations were expressed in matrix notation as:

$$\begin{bmatrix} 3 & 2 & 1 \\ 4 & 3 & 0 \\ 7 & 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ ys_3 \end{bmatrix}$$

that is, as:

$$\mathbf{Dx} = \mathbf{y}$$

a unique inverse does not exist for **D** because of the dependency in the rows. Only two rows are linearly independent in **D** and it is said of to be of rank 2, usually written as $r(\mathbf{D}) = 2$. When a square matrix is not of full rank, the determinant is zero and hence a unique inverse does not exist.

A.3.7 Generalized inverses

While an inverse does not exist for a singular matrix, a generalized inverse can, however, be calculated. A generalized inverse for a matrix **D** is usually denoted as $\mathbf{D}^-$ and satisfies the expression:

$$\mathbf{D}\mathbf{D}^-\mathbf{D} = \mathbf{D}$$

Generalized inverses are not unique and may be obtained in several ways. One of the simplest ways to calculate a generalized inverse of a matrix, say **D** in Section A.3.6, is to initially obtain a matrix **B** of full rank as a subset of **D**. Set all elements of **D** to zero. Calculate the inverse of **B** and replace the elements of **D** with corresponding elements of **B** and the result is $\mathbf{D}^-$. For instance, for the matrix **D** above, the matrix **B**, a full rank subset of **D**, is:

$$\mathbf{B} = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B}^{-1} = \begin{bmatrix} 3 & -2 \\ -4 & 3 \end{bmatrix}$$

Replacing elements of **D** with the corresponding elements of **B** after all elements of **D** have been set to zero gives $\mathbf{D}^-$ as:

$$\mathbf{D}^- = \begin{bmatrix} 3 & -2 & 0 \\ -4 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A.3.8 Eigenvalues and eigenvectors

Eigenvalues are also referred to as characteristic or latent roots and are useful in simplifying multivariate evaluations when transforming data. The sum of the eigenvalues of a square matrix equals its trace (sum of the diagonal elements of a square matrix) and their product equals its determinant (Searle, 1982). For symmetric matrices, the rank equals the number of non-zero eigenvalues.

For a square matrix **B**, the eigenvalues are obtained by solving:

$$|\mathbf{B} - d\mathbf{I}| = 0$$

where the vertical lines denote finding the determinant.

With the condition specified in the above equation, **B** can be represented as:

$$\mathbf{B}\mathbf{L} = \mathbf{L}\mathbf{D}$$
$$\mathbf{B} = \mathbf{L}\mathbf{D}\mathbf{L}^{-1} \qquad \text{(a.1)}$$

where **D** is a diagonal matrix containing the eigenvalues of **B**, and **L** is a matrix of corresponding eigenvectors. The eigenvector (k) is found by solving:

$$(\mathbf{B} - d_k\mathbf{I})l_k = 0$$

where d_k is the corresponding eigenvalue.

For symmetric matrices **L** is orthogonal (that is, $\mathbf{L}^{-1} = \mathbf{L}'$; $\mathbf{L}\mathbf{L}' = \mathbf{I} = \mathbf{L}'\mathbf{L}$); therefore, given that **B** is symmetric, (Eqn a.1) can be expressed as:

$$\mathbf{B} = \mathbf{L}\mathbf{D}\mathbf{L}'$$

Usually, eigenvalues and eigenvectors are calculated by means of computer programs.

Appendix B: Fast Algorithms for Calculating Inbreeding Based on the L Matrix

In this appendix, two algorithms based on the **L** matrix for calculating inbreeding are discussed.

B.1 Meuwissen and Luo Algorithm

The algorithm given by Quaas (1976) invoves the calculation of one column of **L** at a time. The algorithm requires $n(n + 1)/2$ operations and computational time is proportional to n^2, where n is the size of the data set. It suffers from the disadvantage of not being readily adapted for updating when a new batch of animals is available without restoring a previously stored **L**. Meuwissen and Luo (1992) presented a faster algorithm, which involves computing the elements of **L** row by row.

The fact that each row of **L** is calculated independently of other rows makes it suitable for updating. The row i of **L** for animal i gives the fraction of genes the animal derives from its ancestors. If s_i and d_i are the sire and dam of animal i, then $l_{is_i} = l_{id_i} = 0.5$. The ith row of **L** can be calculated by proceeding through a list of i's ancestors from the youngest to the oldest and updating continually as $l_{is_j} = l_{is_j} + 0.5l_{ij}$ and $l_{id_j} = l_{id_j} + 0.5l_{ij}$, where j is an ancestor of i. The fraction of genes derived from an ancestor is:

$$l_{ij} = \sum_{k \varepsilon \mathbf{P}_j} 0.5 l_{ik}$$

where $\mathbf{P}_j$ is a set of identities of the progeny of j. However, $l_{ij} = 0$ only when k is not an ancestor of i or k is not equal to i. Thus if **AN** is the set of identities of the number of ancestors of i, then:

$$l_{ij} = \sum_{k \varepsilon \mathbf{AN} \cap \mathbf{p}_j} 0.5 l_{ik}$$

that is, the summation of $0.5l_{ik}$ is over those k animals that are both ancestors of i and progeny of j. This forms the basis of the algorithm given below for the calculation of the row i of **L**, one row at a time. As each row of **L** is calculated, its contribution to the diagonal elements of the relationship matrix (a_{ii}) is accumulated. Initially, set row i of **L** and a_{ii} to zero. The list of ancestors whose contributions to a_{ii} are yet to be included are added to the vector **AN** (if not already there) as each row of **L** is being processed.

The algorithm is:

$F_0 = -1$

For $i = 1, N$ (all rows of **L**):

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

$\mathbf{AN}_i = i$

$l_{ii} = 1$

$\mathbf{D}_{ii} = [0.5 - 0.25(F_{s_i} + F_{d_i})]$, if both parents are known; otherwise use appropriate the formula (see Chapter 2)

Do while $\mathbf{AN}_i$ is not empty.

$j = \max(\mathbf{AN}_i)$, (j = youngest animal in $\mathbf{AN}_i$)

If s_j is known, add s_j to $\mathbf{AN}_i$:

$l_{is_j} = l_{is_j} + 0.5l_{ij}$

If d_j is known, add d_j to $\mathbf{AN}_i$:

$l_{id_j} = l_{id_j} + 0.5l_{ij}$

$a_{ii} = a_{ii} + l_{ij}^2 \boldsymbol{D}_{jj}$

Delete j from $\mathbf{AN}_i$

End while:

$F_i = a_{ii} - 1$

B.1.1 Illustration of the algorithm

Using the pedigree in Table 2.1, the algorithm is illustrated for animals 1 and 5.

For animal 1:

$a_{11} = 0$

$\mathbf{AN}_1 = 1$, $l_{11} = 1$

Since both parents are unknown:

$\mathbf{D}_{11} = 1$

Processing animals in $\mathbf{AN}_1$:

$j = \max(\mathbf{AN}_1) = 1$

Both parents of j are unknown:

$a_{11} = a_{11} + l_{11}^2\mathbf{D}_{11} = (1^2)1 = 1$

Delete animal 1 from $\mathbf{AN}_1$; $\mathbf{AN}_1$ is now empty.

$\mathbf{F}_1 = 1 - 1 = 0$

For animal 5:

$a_{55} = 0$

$\mathbf{AN}_5 = 5$, $l_{55} = 1$

$\mathbf{D}_{55} = 0.5$, since neither parent is inbred.

Processing animals in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 5$

Add sire and dam of 5 (animals 4 and 3) to AN_5:

$l_{54} = l_{54} + 0.5l_{55} = 0.5$

$l_{53} = l_{53} + 0.5l_{55} = 0.5$

$a_{55} = a_{55} + l_{55}^2\mathbf{D}_{55} = 1^2(0.5) = 0.5$

Delete animal 5 from $\mathbf{AN}_5$; animals 4 and 3 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 4$

Add sire of 4 (animal 1) to $\mathbf{AN}_5$:

$l_{51} = l_{51} + 0.5l_{54} = 0.25$

$a_{55} = a_{55} + l_{54}^2\mathbf{D}_{44} = 0.5 + (0.5)^2(0.75) = 0.6875$

Delete animal 4 from $\mathbf{AN}_5$; animals 3 and 1 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 3$

Since animal 1, the sire of j, is already in $\mathbf{AN}_5$, add only the dam of 3 (animal 2) to $\mathbf{AN}_5$:

$l_{51} = l_{51} + 0.5l_{53} = 0.25 + (0.5)0.5 = 0.5$

$l_{52} = l_{52} + 0.5l_{53} = 0 + (0.5)0.5 = 0.25$

$a_{55} = a_{55} + l_{53}^2\mathbf{D}_{33} = 0.6875 + (0.5)^20.5 = 0.8125$

Delete animal 3 from $\mathbf{AN}_5$; animals 1 and 2 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 2$

Both parents are unknown:

$a_{55} = a_{55} + l_{52}^2\mathbf{D}_{22} = 0.8125 + (0.25)^21 = 0.875$

Delete animal 2 from $\mathbf{AN}_5$; animal 1 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 1$

Both parents are unknown:

$a_{55} = a_{55} + l_{51}^2\mathbf{D}_{11} = 0.875 + (0.5)^21 = 1.125$

Delete 1 from $\mathbf{AN}_5$; $\mathbf{AN}_5$ is empty.

$\mathbf{F}_5 = 1.125 - 1 = 0.125$

which is the same inbreeding coefficient as that obtained for animal 5 in Section 2.2.

B.2 Modified Meuwissen and Luo Algorithm

The approach of Meuwissen and Luo given above was modified by Quaas (1995) to improve its efficiency. The disadvantage of the above method is that, while calculating a row of **L** at a time (Henderson, 1976), it is accumulating diagonal elements of **A**, as in Quaas (1976), and this necessitates tracing the entire pedigree for i, but what is really needed is only the common ancestors. Thus a more efficient approach is to accumulate $a_{s_id_i}$ as $\mathbf{\Sigma}_k l_{s_ik}l_{d_ik}\mathbf{D}_{kk}$ (Henderson (1976) and calculate F_i as $0.5a_{s_id_i} = \mathbf{\Sigma}_k l_{s_ik}l_{d_ik}(0.5\mathbf{D}_{kk})$. Instead of computing the ith row of **L**, only the non-zero elements in the rows for the sire and dam of i are calculated. Quaas (1995) suggested setting up a separate ancestor list ($\mathbf{AS}_{s_i}$) for s_i and another ($\mathbf{AD}_{d_i}$) for d_i; then $\mathbf{F}_i = 0.5a_{s_id_i} = \mathbf{\Sigma}_k \varepsilon_{s_i} U_{d_i} l_{s_ik}l_{d_ik}(0.5\mathbf{D}_{kk})$.

Similar to the approach of Meuwissen and Luo (1992), the two lists can be set up simultaneously while processing the ith animal by continually adding the parents of the next youngest animal in either list to the appropriate list. If the next youngest in each list is the same animal, say k, then it is a common ancestor and $\mathbf{F}_i$ is updated as $\mathbf{F}_i = \mathbf{F}_i + l_{s_ik}l_{d_ik}(0.5\mathbf{D}_{kk})$. When ancestors of one of the parents have been processed, the procedure can be stopped, and it is not necessary to search both lists completely. The algorithm for this methodology is:

$\mathbf{F}_0 = -1$

For i =1, N:

$\mathbf{F}_i = 0$

If s_i is known, add s_i to $\mathbf{AS}_{s_i}$, $l_{s_is_i} = 1$.
If d_i is known, add d_i to $\mathbf{AD}_{d_i}$, $l_{d_id_i} = 1$.
Do while $\mathbf{AS}_{\hat{s}_i}$ not empty and $\mathbf{AD}_{d_i}$ not empty.

$j = \max(\mathbf{AS}_{s_i})$, $k = \max(\mathbf{AD}_{d_i})$

If $j > k$ then (next youngest j is in $\mathbf{AS}_{s_i}$):
If s_j is known, add s_j to $\mathbf{AS}_{s_i}$; $l_{s_is_j} = l_{s_is_j} + 0.5l_{s_ij}$
If d_j is known, add d_j to $\mathbf{AS}_{s_i}$; $l_{s_id_j} = l_{s_id_j} + 0.5l_{s_ij}$
Delete j from $\mathbf{AS}_{s_i}$
Else if $k > j$ then (next youngest k is in $\mathbf{AD}_{d_i}$):
If s_k is known, add s_k to $\mathbf{AD}_{d_i}$; $l_{d_is_k} = l_{d_is_k} + 0.5l_{d_ik}$
If d_k is known, add d_k to $\mathbf{AD}_{d_i}$; $l_{d_id_k} = l_{d_id_k} + 0.5l_{d_ik}$
Delete k from $\mathbf{AD}_{di}$
Else (next youngest ancestor $j = k$ is a common ancestor):
If s_j is known, add s_j to $\mathbf{AS}_{s_i}$; $l_{s_is_j} = l_{s_is_j} + 0.5l_{s_ij}$
add s_j to $\mathbf{AD}_{d_i}$; $l_{d_is_j} = l_{d_is_j} + 0.5l_{d_ij}$
If d_j is known, add d_j to $\mathbf{AS}_{s_i}$; $l_{s_id_j} = l_{s_id_j} + 0.5l_{s_ij}$
add d_j to $\mathbf{AD}_{d_i}$; $l_{d_id_j} = l_{d_id_j} + 0.5l_{d_ij}$

$F_i = F_i + l_{s_ij}l_{d_ij}0.5(\mathbf{D}_{jj})$

Delete j from $\mathbf{AN}_{\hat{s}_i}$ and $\mathbf{AD}_{d_i}$
End if
End while
End do

B.2.1 Illustration of the algorithm

Using the pedigree in Table 2.1, the algorithm is illustrated for animal 5, which is inbred.

For animal 5:

$\mathbf{F}_5 = 0$

Both parents known, s = 4 and d = 3.

Add 4 to $\mathbf{AD}_4$; $l_{44} = 0.5$
Add 3 to $\mathbf{AD}_3$; $l_{33} = 0.5$

Processing animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 4$; $k = 3$

$j > k$ therefore.

Add sire of 4, animal 1 to $\mathbf{AS}_4$; $l_{41} = l_{41} + 0.5l_{44} = 0.5$
Delete animal 4 from $\mathbf{AS}_4$.

Next animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 1$, $k = 3$

$k > j$ therefore.

Add sire of 3, animal 1 to $\mathbf{AD}_3$; $l_{31} = l_{31} + 0.5l_{33} = 0.5$
Add dam of 3, animal 2 to $\mathbf{AD}_3$; $l_{32} = l_{32} + 0.5l_{33} = 0.5$
Delete 3 from $\mathbf{AD}_3$.

Next animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 1$, $k = 2$

$k > j$

Both parents of 2 are unknown.
Delete 2 from $\mathbf{AD}_3$.

Next animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 1$, $k = 1$

$j = k$

Both parents are unknown.

$$F_5 = F_5 + l_{41}l_{31}0.5(\mathbf{D}_{11}) = 0.5(0.5)(0.5)(1) = 0.125$$

which is the same inbreeding coefficient as that obtained from the algorithm in Section B.1.

Appendix C

C.1 Outline of the Derivation of the Best Linear Unbiased Prediction (BLUP)

Consider the following linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \tag{c.1}$$

where the expectations are:

$$E(\mathbf{y}) = \mathbf{Xb}; E(\mathbf{a}) = E(\mathbf{e}) = 0$$

and:

$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G}, \text{var}(\mathbf{e}) = \mathbf{R} \text{ and } \text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$$

Then, as shown in Section 3.2:

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}, \text{cov}(\mathbf{y}, \mathbf{a}) = \mathbf{ZG} \text{ and } \text{cov}(\mathbf{y}, \mathbf{e}) = \mathbf{R}$$

The prediction problem involves both **b** and **a**. Suppose we want to predict a linear function of **b** and **a**, say $\mathbf{k}'\mathbf{b} + \mathbf{a}$, using a linear function of **y**, say $\mathbf{L}'\mathbf{y}$, and $\mathbf{k}'\mathbf{b}$ is estimable. The predictor $\mathbf{L}'\mathbf{y}$ is chosen such that:

$$E(\mathbf{L}'\mathbf{y}) = E(\mathbf{k}'\mathbf{b} + \mathbf{a})$$

that is, it is unbiased and the prediction error variance (PEV) is minimized (Henderson, 1973). Now PEV (Henderson, 1984) is:

$$\begin{aligned}\text{PEV} &= \text{var}(\mathbf{L}'\mathbf{y} - \mathbf{k}'\mathbf{b} + \mathbf{a}) \\ &= \text{var}(\mathbf{L}'\mathbf{y} - \mathbf{a}) \\ &= \mathbf{L}'\text{var}(\mathbf{y})\mathbf{L} + \text{var}(\mathbf{a}) - \mathbf{L}'\text{cov}(\mathbf{y}, \mathbf{a}) - \text{cov}(\mathbf{a}, \mathbf{y})\mathbf{L} \\ &= \mathbf{L}'\mathbf{V}\mathbf{L} + \mathbf{G} - \mathbf{L}'\mathbf{Z}\mathbf{G} - \mathbf{Z}\mathbf{G}'\mathbf{L}\end{aligned} \tag{c.2}$$

Minimizing PEV subject to $E(\mathbf{L}'\mathbf{y}) = E(\mathbf{k}'\mathbf{b} + \mathbf{a})$ and solving (see Henderson, 1973, 1984 for details of derivation) gives:

$$\mathbf{L}'\mathbf{y} = \mathbf{k}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$$

Let $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}\mathbf{V}^{-1}\mathbf{y}$, the generalized least square solution for **b**, then the predictor can be written as:

$$\mathbf{L}'\mathbf{y} = \mathbf{k}'\hat{\mathbf{b}} + \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{c.3}$$

which is the BLUP of $\mathbf{k}'\mathbf{b} + \mathbf{a}$.

Note that if $\mathbf{k}'\mathbf{b} = 0$, then:

$$\mathbf{L}'\mathbf{y} = \text{BLUP}(\mathbf{a}) = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{c.4}$$

which is equivalent to the selection index. Thus BLUP is the selection index with the GLS solution of **b** substituted for **b**.

C.2 Proof that $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ from MME are the GLS of b and BLUP of a, Respectively

In computation terms, the use of Eqn c.3 to obtain the BLUP of $\mathbf{k}'\mathbf{b} + \mathbf{a}$ is not feasible because the inverse of **V** is required. Henderson (1950) formulated the MME that are suitable for calculating solutions for **b** and **a**, and showed later that $\mathbf{k}'\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$, where $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ are solutions from the MME, are the best linear unbiased estimator (BLUE) of $\mathbf{k}'\mathbf{b}$ and BLUP of **a**, respectively.

The usual MME for Eqn c.1 are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}' \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{c.5}$$

The proof that $\hat{\mathbf{b}}$ from the MME is the GLS of **b** and therefore $\mathbf{k}'\hat{\mathbf{b}}$ is the BLUE of $\mathbf{k}'\mathbf{b}$ was given by Henderson *et al.* (1959). From the second row of Eqn c.5:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{a}} = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{c.6}$$

From the first row of Eqn c.5:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{a}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$$

Substituting the solution for $\hat{\mathbf{a}}$ into the above equation gives:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$$

where $\mathbf{W} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}$:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} - (\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z})(\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$$

$$\mathbf{X}'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{X}\mathbf{b} = \mathbf{X}'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{y}$$

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

with $\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}$:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{c.7}$$

It can be shown that:

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}$$

by pre-multiplying the right-hand side by **V** and obtaining an identity matrix (Henderson *et al.*, 1959):

$$\begin{aligned} \mathbf{V}[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}] &= (\mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}')(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}) \\ &= \mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1} \\ &= \mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}(\mathbf{I} + \mathbf{G}\mathbf{Z}'\mathbf{R}\mathbf{Z})\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1} \\ &= \mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}\mathbf{G}(\mathbf{G}^{-1} + \mathbf{Z}'\mathbf{R}\mathbf{Z})\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1} \\ &= \mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}\mathbf{G}(\mathbf{W}^{-1})\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1} \\ &= \mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} \\ &= \mathbf{I} \end{aligned}$$

Thus the solution for $\mathbf{b}$ from the MME is equal to the GLS solution for $\mathbf{b}$ in Eqn c.3.

The proof that $\hat{\mathbf{a}}$ from the MME is equal to $\mathbf{GZ'V^{-1}(y - X\hat{b})}$ in Eqn c.3 was given by Henderson (1963). Replace $\mathbf{V^{-1}}$ in $\mathbf{GZ'V^{-1}(y - X\hat{b})}$ by $\mathbf{R^{-1} - R^{-1}ZWZ'R^{-1}}$, thus:

$$\begin{aligned}
\mathbf{GZ'V^{-1}(y - X\hat{b})} &= \mathbf{GZ'(R^{-1} - R^{-1}ZWZ'R^{-1})(y - X\hat{b})} \\
&= \mathbf{G(Z'R^{-1} - Z'R^{-1}ZWZ'R^{-1})(y - X\hat{b})} \\
&= \mathbf{G(I - Z'R^{-1}ZW)Z'R^{-1}(y - X\hat{b})} \\
&= \mathbf{G(W^{-1} - Z'R^{-1}Z)WZ'R^{-1}(y - X\hat{b})} \\
&= \mathbf{G((Z'RZ + G^{-1}) - Z'R^{-1}Z)WZ'R^{-1}(y - X\hat{b})} \\
&= \mathbf{GZ'R^{-1}Z + I - GZ'R^{-1}Z)WZ'R^{-1}(y - X\hat{b})} \\
&= \mathbf{(I)WZ'R^{-1}(y - X\hat{b})} \\
&= \mathbf{WZ'R^{-1}(y - X\hat{b}) = \hat{a}} \qquad \text{(See Eqn c.6)}
\end{aligned}$$

Thus the BLUP of $\mathbf{k'b + a = k'\hat{b} + \hat{a}}$, where $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ are solutions to the MME.

C.3 Deriving the Equation for Progeny Contribution (PC)

Considering an individual *i* that has one record with both sire (*s*) and dam (*d*) known, the MME for the three animals can be written (assuming the sire and dam are ancestors with unknown parents) as:

$$\begin{bmatrix} u_{ss}\,\alpha & u_{sd}\,\alpha & u_{si}\,\alpha \\ u_{ds}\,\alpha & u_{dd}\,\alpha & u_{di}\,\alpha \\ u_{is}\,\alpha & u_{id}\,\alpha & (1+u_{ii})\,\alpha \end{bmatrix} \begin{bmatrix} \hat{a}_s \\ \hat{a}_d \\ \hat{a}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1'y \end{bmatrix} \qquad \text{(c.8)}$$

where the *u* terms are elements of $\mathbf{A^{-1}}$.
From Eqn c.8, the equation for solution of the sire is:

$$u_{ss}\alpha\hat{a}_s = 0 - u_{sd}\alpha\hat{a}_d - u_{si}\alpha\hat{a}_i$$
$$u_{ss}\alpha\hat{a}_s = PC$$

with:

$$PC = 0 - u_{sd}\,\alpha\hat{a}_d - u_{si}\alpha\hat{a}_i$$

When the mate is known:

$$PC = 0 - \tfrac{1}{2}\alpha\hat{a}_d + (1)\alpha\hat{a}_i$$
$$PC = \alpha(\hat{a}_i - \tfrac{1}{2}\,\hat{a}_d) = 0.5\alpha(2\hat{a}_i - \hat{a}_d)$$

In general, assuming sire *s* has *k* progeny:

$$PC_s = 0.5\alpha\Sigma_k u_{prog}(2\hat{a}_i - \hat{a}_m)/\Sigma_k u_{prog}$$

where u_{prog} is 1 when the mate of *s* is known or $\frac{2}{3}$ when the mate is not known.

Appendix D: Methods for Obtaining Approximate Reliability for Genetic Evaluations

D.1 Computing Approximate Reliabilities for an Animal Model

Presented below is a method published by Meyer (1989) for obtaining approximate values of repeatability or reliability for genetic evaluations from an animal model and has been used to estimate reliabilities in the national dairy evaluation system in Canada. The reliability for each animal is derived from the corresponding diagonal element in the MME, adjusting for selected off-diagonal coefficients. For instance, the section of the coefficient matrix (**C**) pertaining to an animal i with parents f and m and with a record in a subclass h of a major fixed effect as HYS could be represented as:

$$\begin{bmatrix} c_{ii} & -\alpha & -\alpha & 1 \\ -\alpha & c_{ff} & 0.5\alpha & 0 \\ -\alpha & 0.5\alpha & c_{mm} & 0 \\ 1 & 0 & 0 & n_h \end{bmatrix}$$

where n_h is the number of records in subclass h of the major fixed effect and $\alpha = \sigma_e^2/\sigma_a^2$. If this were the complete coefficient matrix for this animal, $\mathbf{C}^{-1}$ and hence true reliability could be obtained using partition matrix results. Thus the coefficient c^{ii} can be calculated as the reciprocal of the ith diagonal element of **C** after absorbing all other rows and columns. For animal i:

$$c^{ii} = (c_{ii} - 1/n_h - \alpha^2(c_{ff} + c_{mm} - \alpha)/(c_{ff}\, c_{mm} - \tfrac{1}{4}\,\alpha^2))^{-1}$$

and for parent f:

$$c^{ff} = (c_{ff} - Q - (\tfrac{1}{2}\,\alpha - Q)^2/(c_{mm} - Q))^{-1}$$

with:

$$Q = \alpha^2(c_{ii} - 1/n_h)^{-1}$$

Exchange m for f for parent m.

However, if there are other off-diagonals for animal i, the above equations will yield approximations of the diagonal elements of **C** and hence reliability. Based on the above principle of forming and inverting the submatrix of the MME for each animal, Meyer outlined three steps for calculating approximate r^2, which were similar to the true r^2 from her simulation study. These steps are:

1. Diagonal elements (D) of animals with records are adjusted for the effect of the major fixed effects such as HYS. Thus:

$$D_{1i} = D_{0i} - 1/n_h$$

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

and for animals without records:

$$D_{1i} = D_{0i}$$

where D_{0i} is the diagonal element for animal i in the MME and, in general, its composition, depending on the amount of information available on the animal, is:

$$D_{0i} = x_i + n_i\alpha + n_{1i}\alpha/3 + n_{2i}\alpha/2$$

where $x_i = 1$ if the animal has a record and otherwise it is zero, n_i equals 1 or $\frac{4}{3}$ or 2 if none or one or both parents are known, n_{1i} and n_{2i} are number of progeny with one or both parents known, respectively.

2. Diagonal elements for parents (f and m) are adjusted for the fact the information on their progeny is limited.

For each progeny i with only one parent known, adjust the diagonal element of the parent as:

$$D_{2f} = D_{1f} - \alpha^2(\tfrac{4}{9}\, D_{1i}^{-1})$$

and if both parents are known, adjust the diagonal of parent f as:

$$D_{2f} = D_{1f} - \alpha^2 D_{1i}^{-1}$$

Replace subscript f with m for the other parent. For animals that are not parents:

$$D_{2i} = D_{1i}$$

3. Adjustment of progeny diagonals for information on parents.

This involves initially unadjusting the diagonals of the parents for the contribution of the ith progeny in question by reversing step 2 before adjusting progeny diagonals for parental information. If only one parent f is known, the diagonal is unadjusted initially as:

$$D_{2f}^{*} = \mathrm{D}_{2f} + \alpha^2(\tfrac{4}{9} D_{2i}^{-1})$$

and if both parents are known as:

$$D_{2f}^{*} = D_{2f} - \alpha^2\, D^{-1}_{2i}$$

for parent f. Exchange m for f in the above equation to calculate for parent m.

Adjustment of progeny i diagonal then is:

$$D_{3i} = D_{2i} - \alpha^2 \tfrac{4}{9} D_{2f}^{*-1}$$

if only parent f is known and:

$$D_{3i} = D_{2i} - \alpha^2((D_{2f}^{*} + D_{2m}^{*} - \alpha)/(D_{2f}^{*}\, D_{2m}^{*} - \tfrac{1}{4}\alpha^2))$$

when both parents f and m are known.

For animals with unknown parents:

$$D_{3i} = D_{2i}$$

Reliability for progeny i is calculated as:

$$r^2 = \text{const.}(1 - \alpha D_{3i}^{-1})$$

where const. is a constant of between 0.90 and 0.95 from Meyer simulation studies which gave the best estimate of r^2.

D.2 Computing Approximate Reliabilities for Random Regression Models

Meyer and Tier (2003) extended the method in Appendix D.1 to estimate reliabilities for multivariate and random regression models. They outlined several steps.

D.2.1 Determine value of observation for an animal

Compute the diagonal block ($\mathbf{D}_i$) for animal i in the MME, based on the information from the data, as:

$$\mathbf{D}_i = \mathbf{Z}_i' \mathbf{R}_i^{-1} \mathbf{Z}_i$$

However, to account for the limited subclass sizes of contemporary group effect, such as HTD in dairy cattle, $\mathbf{D}_i$ can better be calculated as:

$$\mathbf{D}_i = \mathbf{Z}_i' (\mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} (\mathbf{S}_i^{-1}) \mathbf{R}_i^{-1}) \mathbf{Z}_i$$

where $\mathbf{Z}_i$ and $\mathbf{R}_i^{-1}$ are submatrices of $\mathbf{Z}$ and $\mathbf{R}^{-1}$ for animal i, and $\mathbf{S}_i$ is the block of coefficient matrix pertaining to the contemporary groups of which animal i is a member. Then the permanent environmental (pe) effects are also absorbed into the block corresponding to animal genetic effects:

$$\mathbf{D}_i^* = \mathbf{D}_i - \mathbf{Z}_i' \mathbf{R}_i^{-1} \mathbf{Q}_i (\mathbf{Q}_i' \mathbf{R}_i^{-1} \mathbf{Q}_i + \mathbf{P}^{-1}) \mathbf{Q}_i' \mathbf{R}_i^{-1} \mathbf{Z}_i$$

where $\mathbf{Q}_i$ is a submatrix of the matrix $\mathbf{Q}$ defined in Section 9.3. Limited subclass effects of pe can be accounted for by using weights $w_m = (n_m - k)/n_m \leq 1$, for the mth record, with n_m the size of the subclass to which the record belongs and k the number of 'repeated' records it has in that subclass. Then $\mathbf{R}_i$ in the above equation is replaced with $\mathbf{R}_i^* = \text{Diag}(w_m \sigma_e^2)$.

D.2.2 Value of records on descendants

In this second step, the contributions from progeny and other descendants are accumulated for each animal, processing the pedigree from youngest to the oldest. Let $\mathbf{E}_i$ be the block of contributions for animal i that has n_i progeny. Then:

$$\mathbf{E}_i = \tfrac{1}{3}\mathbf{G}^{-1} - \tfrac{4}{9}\mathbf{G}^{-1} \left(\mathbf{D}_i^* + \sum_{k=1}^{n_i} \mathbf{E}_k + \tfrac{4}{3}\mathbf{G}^{-1} \right)^{-1} \mathbf{G}^{-1}$$

This block is accumulated for both sire and dam of the ith animal. This equation can be derived by assuming each progeny has only one parent known and that the parent has no other information; then the MME are set up for the animal and the parent and the equations for the animal are absorbed into those of the parent. The above equation will give an overestimate of the individual's contribution to its parents if it were in a contemporary group with many of its half-sibs. This can be discounted by weighting contributions with a factor dependent on the proportion of sibs in a subclass. Let $\mathbf{H}_i$ be a diagonal matrix of weights $w_m < 1$, with $w_m = \sqrt{(n_m - s_m)/n_m}$, where n_m is the total number of records in the subclass for trait m and s_m the total number of sibs of animal i in the subclass. Calculate $\mathbf{D}_i^{**} = \mathbf{H}_i \mathbf{D}_i^* \mathbf{H}_i$ and then replace $\mathbf{D}_i^*$ with $\mathbf{D}_i^{**}$.

D.2.3 Value of records on ancestors

In the third step, contributions from parents, ancestors and collateral relatives are accumulated for each animal, processing the pedigree from oldest to youngest. However, in step two, contributions from descendants were accumulated for all animals, hence $\mathbf{E}_j$ for parent j of animal i includes the contribution from animal i. The contributions of animal i have to be removed from $\mathbf{E}_j$ to avoid double counting. The corrected block is:

$$\mathbf{E}_j^* = \tfrac{1}{3}\mathbf{G}^{-1} - \tfrac{4}{9}\mathbf{G}^{-1}(-\mathbf{E}_i + \mathbf{F}_j + \tfrac{4}{3}\mathbf{G}^{-1})^{-1}\mathbf{G}^{-1}$$

where $\mathbf{F}_j$ is the sum of contributions from all sources of information for parent j. As parents are processed in the pedigree before progeny, $\mathbf{F}_j$ is always computed before the contribution of parent j to animal i is required. For animal i, $\mathbf{F}_j$ is:

$$\mathbf{F}_i = \sum_{j=1}^{t_i} \mathbf{E}_j^* + \mathbf{D}_i^* + \sum_{k=1}^{n_i} \mathbf{E}_k$$

with t_i = 0, 1 or 2 denoting the number of parents of animal i that are known.

The matrix $\mathbf{T}_i$ of the approximate PEV and PEC for the genetic effects for animal i is:

$$\mathbf{T}_i = (\mathbf{F}_i + \mathbf{G})^{-1}$$

The approximate reliability for a linear function of EBVs for animal i then is:

$$r_i^2 = 1 - \mathbf{k}'\mathbf{T}_i\mathbf{k}/\mathbf{k}'\mathbf{G}\mathbf{k}$$

with $\mathbf{k}$ calculated as described in Section 9.3.4.

Appendix E

E.1 Canonical Transformation: Procedure to Calculate the Transformation Matrix and its Inverse

The simplification of a multivariate analysis into n single trait analyses using canonical transformation involves transforming the observations of several correlated traits into new uncorrected traits (Section 6.2). The transformation matrix **Q** can be calculated by the following procedure, which has been illustrated by the **G** and **R** matrices for Example 6.1 in Section 6.2.2.

The **G** and **R** matrices are, respectively:

$$\begin{matrix} \text{WWG} \\ \text{PWG} \end{matrix} \begin{bmatrix} 20 & 18 \\ 18 & 40 \end{bmatrix} \text{ and } \begin{matrix} \text{WWG} \\ \text{PWG} \end{matrix} \begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix}$$

where WWG is the pre-weanng gain and PWG is the post-weaning gain.

1. Calculate the eigenvalues (**B**) and eigenvectors (**U**) of **R**:

$$\mathbf{R} = \mathbf{UBU}'$$

For the above **R**:

$$\mathbf{B} = \text{diag}(47.083, 22.917)$$

and:

$$\mathbf{U} = \begin{bmatrix} 0.841 & -0.541 \\ 0.541 & 0.841 \end{bmatrix}$$

2. Calculate **P** and **PGP′**:

$$\mathbf{P} = \mathbf{U}\sqrt{\mathbf{B}^{-1}}\,\mathbf{U}'$$

$$\mathbf{P} = \begin{bmatrix} 0.1642 & -0.0288 \\ -0.0288 & 0.1904 \end{bmatrix} \text{ and } \mathbf{PGP}' = \begin{bmatrix} 0.403 & 0.264 \\ 0.264 & 1.269 \end{bmatrix}$$

3. Calculate the eigenvalues (**W**) and eigenvectors (**L**) of **PGP′**:

$$\mathbf{PGP}' = \mathbf{LWL}'$$

$$\mathbf{W} = \text{diag}(0.3283, 1.3436)$$

and:

$$\mathbf{L} = \begin{bmatrix} 0.963 & 0.271 \\ -0.271 & 0.963 \end{bmatrix}$$

© R.A. Mrode 2014. *Linear Models for the Prediction of Animal Breeding Values, 3rd Edition* (R.A. Mrode)

4. The transformation matrix **Q** can be obtained as:

$$\mathbf{Q} = \mathbf{L}'\mathbf{P}$$

$$\mathbf{Q} = \begin{bmatrix} 0.1659 & -0.0792 \\ 0.0168 & 0.1755 \end{bmatrix} \text{ and } \mathbf{Q}^{-1} = \begin{bmatrix} 5.7651 & 2.6006 \\ -0.5503 & 5.4495 \end{bmatrix}$$

E.2 Canonical Transformation with Missing Records and Same Incidence Matrices

Ducrocq and Besbes (1993) presented a methodology for applying canonical transformation when all effects in the model affect all traits but there are missing traits for some animals. The principles of the methodology are briefly discussed and illustrated by an example.

Let **y**, the vector of observations, be partitioned as $\mathbf{y}' = [\mathbf{y}_v, \mathbf{y}_m]$ and $\mathbf{u} = [\mathbf{b}', \mathbf{a}']$, where $\mathbf{y}_v$ and $\mathbf{y}_m$ are vectors of observed and missing records, respectively, **b** is the vector of fixed effects and **a** is the vector of random effects. Assuming that the distribution of **y** given **u** is multivariate normal, Ducrocq and Besbes (1993) showed that the following expectation maximization (EM) algorithm gives the same solutions for **a** and **b** as when the usual multivariate MME are solved:

E step: at iteration k, calculate $\hat{\mathbf{y}}^{[k]} = \mathrm{E}[\mathbf{y}|\mathbf{y}_v, \hat{\mathbf{u}}^{[k]}]$
M step: calculate $\hat{\mathbf{u}}^{[k+1]}$ = BLUE and BLUP solutions of **b** and **a**, respectively, given $\hat{\mathbf{y}}^{[k]}$

The E step implies doing nothing to observed records but replacing the missing observations by their expectation given the current solutions for **b** and **a**, and the observed records. The equation for the missing records for animal i is:

$$\hat{\mathbf{y}}_{im}^{[k]} = \mathbf{x}'_{im}\mathbf{b}^{[k]} + \hat{\mathbf{a}}_{im}^{[k]} + \hat{\mathbf{e}}_{im}^{[k]} \qquad \text{(e.1)}$$

If **X** is the matrix that relates fixed effects to animals, $\mathbf{x}'_{im}$ denotes the row of **X** corresponding to missing records for animal i and $\hat{\mathbf{e}}_{im}^{[k]}$ is the regression of the residuals of missing records on the current estimates of the residuals for observed traits. Thus:

$$\hat{\mathbf{e}}_{im}^{[k]} = \mathrm{E}[\mathbf{e}_{im}|\mathbf{y}_{iv}, \mathbf{u} = \hat{\mathbf{u}}^{[k]}] = \mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}[\mathbf{y}_{iv} - \mathbf{x}'_{iv}\mathbf{b}^{[k]} - \hat{\mathbf{a}}_{iv}^{[k]}]$$

where $\mathbf{R}_{mv}$ and $\mathbf{R}_{vv}$ are submatrices obtained through partitioning of **R**, the residual covariance matrix. $\mathbf{R}_{vv}$ represents the residual variance of observed traits and $\mathbf{R}_{mv}$ is the covariance between missing traits and observed traits. If three traits are considered, for example, and trait 2 is missing for animal i, then $\mathbf{R}_{vv}$ is the submatrix obtained by selecting in **R** the elements at intersection of rows 1 and 3 and columns 1 and 3. The submatrix $\mathbf{R}_{mv}$ is the element at the intersection of row 2 and columns 1 and 3. Once the missing observations have been estimated, records are now available on all animals and the analysis can be carried out as usual, applying canonical transformation as when all records are observed.

The application of the method in genetic evaluation involves the following steps at each iteration k, assuming **Q** is the transformation matrix to canonical scale and $\mathbf{Q}^{-1}$ the back-transforming matrix:

1. For each animal i with missing observations:
   (1a) calculate $\hat{\mathbf{y}}_{im}^{[k]}$, given $\hat{\mathbf{b}}^{[k]}$ and $\hat{\mathbf{a}}^{[k]}$ using Eqn e.1;
   (1b) transform $\hat{\mathbf{y}}_i$ to the canonical scale: $\hat{\mathbf{y}}_i^* = \mathbf{Q}\hat{\mathbf{y}}_i$.
2. Solve the MME to obtain solutions in the canonical scale: $\hat{\mathbf{b}}^{*[k+1]}$ and $\hat{\mathbf{a}}^{*[k+1]}$.
3. Back-transform using $\mathbf{Q}^{-1}$ to obtain $\hat{\mathbf{b}}^{[k+1]}$ and $\hat{\mathbf{a}}^{[k+1]}$.
4. If convergence is not achieved, go to 1.

Ducrocq and Besbes (1993) showed that it is possible to update $\mathbf{y}$ (step 1) without back-transforming to the original scale (step 3) in each round of iteration. Suppose that the vector of observations for animal i with missing records, $\mathbf{y}_i$, is ordered such that observed records precede missing values: $\mathbf{y}'_i = [\mathbf{y}'_{iv}, \mathbf{y}'_{im}]$, and rows and columns of $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ are ordered accordingly. Partition $\mathbf{Q}$ as $(\mathbf{Q}_v \mid \mathbf{Q}_m)$ and $\mathbf{Q}^{-1}$ as:

$$\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{Q}^{\mathrm{v}} \\ \mathbf{Q}^{\mathrm{m}} \end{bmatrix}$$

then from Eqn e.1, the equation for $\mathbf{Q}\hat{\mathbf{y}}_i$ or $\hat{\mathbf{y}}_i^*$ (see 1b) is:

$$\hat{\mathbf{y}}_i^* = \mathbf{Q}_v\mathbf{y}_{iv} + \mathbf{Q}_m[\mathbf{x}'_{im}\hat{\mathbf{b}}^{[k]} + \hat{\mathbf{a}}_{im}^{[k]} + \mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}(\mathbf{y}_{iv} - \mathbf{x}'_{iv}\hat{\mathbf{b}}^{[k]} - \hat{\mathbf{a}}_{iv}^{[k]})] \qquad \text{(e.2)}$$

However:

$$\begin{bmatrix} \hat{\mathbf{b}}_{\mathrm{iv}} \\ \hat{\mathbf{b}}_{\mathrm{im}} \end{bmatrix} = \mathbf{Q}^{-1}\hat{\mathbf{b}}^* = \begin{bmatrix} \mathbf{Q}^{\mathrm{v}}\hat{\mathbf{b}}^* \\ \mathbf{Q}^{\mathrm{m}}\hat{\mathbf{b}}^* \end{bmatrix}$$

and a similar expression exists for $\hat{\mathbf{a}}$. Substituting these values for $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ in Eqn e.2:

$$\begin{aligned} \hat{\mathbf{y}}_i^* &= (\mathbf{Q}_v + \mathbf{Q}_m\mathbf{R}_{mv}\mathbf{R}_{vv}^{-1})\mathbf{y}_{iv} + (\mathbf{Q}_m\mathbf{Q}^m - \mathbf{Q}_m\mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}\mathbf{Q}^v)(\mathbf{x}'_i\hat{\mathbf{b}}^{*[k]} + \hat{\mathbf{a}}^{*[k]}) \\ &= \mathbf{Q}_1\mathbf{y}_{iv} + \mathbf{Q}_2(\mathbf{x}_i'\hat{\mathbf{b}}^{*[k]} + \hat{\mathbf{a}}^{*[k]}) \qquad \text{(e.3)} \end{aligned}$$

with $\mathbf{Q}_1 = \mathbf{Q}_v + \mathbf{Q}_m\mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}$ and $\mathbf{Q}_2 = (\mathbf{Q}_m\mathbf{Q}^m - \mathbf{Q}_m\mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}\mathbf{Q}^v)$

Thus for an animal with missing records, $\hat{\mathbf{y}}_i^*$ in Eqn e.3 is the updated vector of observation transformed to canonical scale (steps 1a and 1b above) and this is calculated directly without back-transformation to the original scale (step 3). The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ in Eqn e.3 depend on the missing pattern and if there are n missing patterns, n such matrices of each type must be set up initially and stored for use at each iteration.

E.2.1 Illustration

Using the same genetic parameters and data as for Example 5.3, the above methodology is employed to estimate sex effects and predict breeding values for pre-weaning weight and post-weaning gain iterating on the data (see Section 17.4).
From Section E.1, $\mathbf{Q}$ is:

$$\mathbf{Q} = \begin{bmatrix} 0.1659 & -0.0792 \\ 0.0168 & 0.1755 \end{bmatrix} \text{ and } \mathbf{Q}^{-1} = \begin{bmatrix} 5.7651 & 2.6006 \\ -0.5503 & 5.4495 \end{bmatrix}$$

Partitioning $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ as specified above gives the following matrices:

$$\mathbf{Q}_v = \begin{bmatrix} 0.1659 \\ 0.0168 \end{bmatrix}, \quad \mathbf{Q}_m = \begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix},$$

$$\mathbf{Q}^v = \begin{bmatrix} 5.7651 & 2.6006 \end{bmatrix} \text{ and } \mathbf{Q}^m = \begin{bmatrix} -0.5503 & 5.4495 \end{bmatrix}$$

From the residual covariance matrix in Section E.1:

$$\mathbf{R}_{mv}\mathbf{R}_{vv}^{-1} = 11/40 = 0.275$$

The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$, respectively, are:

$$\mathbf{Q}_1 = \begin{bmatrix} 0.1659 \\ 0.0168 \end{bmatrix} + \begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix} 0.275 = \begin{bmatrix} 0.1441 \\ 0.0654 \end{bmatrix}$$

and:

$$\mathbf{Q}_2 = \left[\begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix} \begin{bmatrix} -0.5503 & 5.4495 \end{bmatrix} - \begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix} 0.275 \begin{bmatrix} 5.7651 & 2.6006 \end{bmatrix} \right]$$

$$= \begin{bmatrix} 0.1691 & -0.3750 \\ -0.3748 & 0.8309 \end{bmatrix}$$

Employing steps 1 to 4 given earlier to the data in Example 5.2, using the various transformation matrices given above and solving for sex and animal solutions by iterating on the data (see Section 17.4), gave the following solutions on the canonical scale at convergence. The solutions on the original scale are also presented.

	Canonical scale		Original scale	
Effects	VAR1	VAR2	WWG	PWG
Sex				
Male	0.180	1.265	4.326	6.794
Female	0.124	1.108	3.598	5.968
Animal				
1	0.003	0.053	0.154	0.288
2	–0.006	–0.010	–0.059	–0.054
3	0.003	–0.030	–0.062	–0.163
4	0.002	0.007	0.027	0.037
5	–0.010	–0.097	–0.307	–0.521
6	0.001	0.088	0.235	0.477
7	–0.011	–0.084	–0.280	–0.452
8	0.013	0.076	0.272	0.407
9	0.009	0.010	0.077	0.051

VAR1, $\mathbf{Q}y_1$, VAR2, $\mathbf{Q}y_2$ with WWG = y_1 and PWG = y_2.

These are similar to the solutions obtained from the multivariate analysis in Section 5.3 or the application of the Cholesky transformation in Section 6.3. The advantage of this methodology is that the usual univariate programs can easily be modified to incorporate missing records.

The prediction of the missing record (PWG) for animal 4 using solutions on canonical and original scales at convergence is illustrated below.

Using Eqn e.1:

$$\begin{aligned} y_{42} &= \hat{b}_{12} + \hat{a}_{42} + \mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}(y_{41} - \hat{b}_{11} - \hat{a}_{41}) \\ &= 6.836 + 0.016 + 0.275(4.5 - 4.366 - 0.007) \\ &= 6.9 \end{aligned}$$

where y_{ij} and $\hat{a}_{ij}$ are the record and EBV, respectively, for animal i and trait j, and $\hat{b}_{kj}$ is the fixed effect solution for level k for trait j.

Using Eqn e.2:

$$\begin{bmatrix} \hat{y}^*_{41} \\ \hat{y}^*_{42} \end{bmatrix} = \mathbf{Q}_1 y_{41} + \mathbf{Q}_2 (\mathbf{x}'\hat{\mathbf{b}}^* + \hat{\mathbf{a}}^*_4)$$

$$\begin{bmatrix} \hat{y}^*_{41} \\ \hat{y}^*_{42} \end{bmatrix} = \begin{bmatrix} 0.648 \\ 0.294 \end{bmatrix} + \mathbf{Q}_2 \begin{bmatrix} 0.183 \\ 1.273 \end{bmatrix} + \mathbf{Q}_2 \begin{bmatrix} 0.000 \\ 0.003 \end{bmatrix}$$

$$= \begin{bmatrix} 0.648 \\ 0.294 \end{bmatrix} + \begin{bmatrix} -0.446 \\ 0.989 \end{bmatrix} = \begin{bmatrix} 0.202 \\ 1.283 \end{bmatrix}$$

These predicted records for animal 4 are on the canonical scale and they are used for the next round of iteration if convergence has not been achieved. These predicted records can be transformed to the original scale as:

$$\begin{bmatrix} \hat{y}_{41} \\ \hat{y}_{42} \end{bmatrix} = \mathbf{Q}^{-1} \begin{bmatrix} 0.202 \\ 1.283 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 6.9 \end{bmatrix}$$

The record for WWG is as observed and predicted missing record for PWG is the same as when using Eqn e.1.

E.3 Cholesky Decomposition

Any positive semi-definite symmetric matrix **R** can be expressed in the form **TT′**, where **T** is a lower triangular matrix. The matrix **T** can be calculated using the following formulae. The ith diagonal element of **T** is calculated as:

$$t_{ii} = r_{ii} - \sqrt{\sum_{j=1}^{i-1} t_{ij}^2}$$

and the lower off-diagonal element of the ith row and the kth column of **T** as:

$$t_{ik} = \frac{1}{t_{kk}} \left(r_{ik} - \sum_{j=1}^{k-1} t_{ij} t_{kj} \right)$$

Appendix F: Procedure for Computing Deregressed Breeding Values

The deregressed breeding values (DRB) of bulls used in multi-trait across-country evaluations (MACE) are obtained by solving Eqn 5.15 for **y** considering data from only one country at a time. Jairath *et al*. (1998) presented an algorithm for calculating DRP. For instance, Eqn 5.15 for country *i* can be written as:

$$\begin{pmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{1} & \mathbf{1}'\mathbf{R}_i^{-1} & 0 & 0 \\ \mathbf{R}_i^{-1}\mathbf{1} & \mathbf{R}_i^{-1}+\mathbf{A}_{nn}^{-1}\alpha_i & \mathbf{A}_{np}^{-1}\alpha_i & \mathbf{A}_{ng}^{-1}\alpha_i \\ 0 & \mathbf{A}_{pn}^{-1}\alpha_i & \mathbf{A}_{pp}^{-1}\alpha_i & \mathbf{A}_{pg}^{-1}\alpha_i \\ 0 & \mathbf{A}_{gn}^{-1}\alpha_i & \mathbf{A}_{gp}^{-1}\alpha_i & \mathbf{A}_{gg}^{-1}\alpha_i \end{pmatrix} \begin{pmatrix} \hat{\mu}_i \\ \mathbf{Q}\mathbf{g}_i+\hat{\mathbf{s}}_i \\ \hat{\mathbf{p}}_i \\ \hat{\mathbf{g}}_i \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i \\ \mathbf{R}_i^{-1}\mathbf{y}_i \\ 0 \\ 0 \end{pmatrix} \tag{f.1}$$

where $\mathbf{p}_i$ is the vector of identified parents without EBV and $\mathbf{A}_{jj}^{-1}$ are blocks of the inverse of the relationship (see Chapter 3, Section 3.6) with $j = n, p$ and g for animals with records, ancestors and genetic groups, respectively, and $\alpha_i = (4 - h_i^2)/h_i^2$, the ratio of residual variance to sire variance for the *i*th country. The deregression of EBV involves solving Eqn f.1 for $\mathbf{y}_i$. The constant $\boldsymbol{\mu}_i$ and vectors $\mathbf{s}_i$, $\mathbf{p}_i$, $\mathbf{g}_i$ and $\mathbf{y}_i$ are unknown but $\mathbf{a}_i$, the vector of genetic evaluations for sires, is known, as well as matrices $\mathbf{Q}$, $\mathbf{R}_i^{-1}$ and $\mathbf{A}_{jj}^{-1}$. Let $\mathbf{a}_i = \mathbf{1}\boldsymbol{\mu}_i + \mathbf{Q}\mathbf{g}_i + \mathbf{s}_i$. The following iterative procedure can be used to compute the vector of DRB, $\mathbf{y}_i$:

1. Set $\mathbf{1}\boldsymbol{\mu}_i$, $\mathbf{p}_i$, $\mathbf{s}_i$ and $\mathbf{g}_i$ to 0.
2. Calculate $\mathbf{Q}\mathbf{g}_i + \mathbf{s}_i = \mathbf{a}_i - \mathbf{1}\boldsymbol{\mu}_i$.
3. Compute:

$$\begin{pmatrix} \hat{\mathbf{p}}_i \\ \hat{\mathbf{g}}_i \end{pmatrix} = -\begin{pmatrix} \mathbf{A}_{pp}^{-1} & \mathbf{A}_{pg}^{-1} \\ \mathbf{A}_{gp}^{-1} & \mathbf{A}_{gg}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}_{pn}^{-1} \\ \mathbf{A}_{gn}^{-1} \end{pmatrix} \left(\mathbf{Q}\hat{\mathbf{g}}_i + \hat{\mathbf{s}}_i\right)$$

4. Generate:

$$\mathbf{R}_i^{-1}\mathbf{y}_i = \mathbf{R}_i^{-1}1\mu_i + (\mathbf{R}_i^{-1} + \mathbf{A}_{nn}^{-1})(\mathbf{Q}\hat{\mathbf{g}}_i + s_i) + \mathbf{A}_{pn}^{-1}\mathbf{p}_i\alpha_i + \mathbf{A}_{gn}^{-1}\hat{\mathbf{g}}_i\alpha_i$$

and $1'\mathbf{R}_i^{-1}\mathbf{y}_i$.

5. Now calculate:

$$\boldsymbol{\mu}_i = (\mathbf{1}'\mathbf{R}_i^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i$$

6. Continue at step 2 until convergence is achieved.
7. Then compute DRB as $\mathbf{y}_i = \mathbf{R}_i\,(\mathbf{R}_i^{-1}\mathbf{y}_i)$.

Using the data for country 1 in Example 5.5, the deregression steps above are illustrated in the first iteration. For country 1, $\alpha_1 = 206.50/20.5 = 10.0732$ and, considering only the bulls with evaluations, $\mathbf{R}_1 = \text{diag}(0.0172, 0.0067, 0.0500, 0.0400)$.

The pedigree structure (see Example 5.5) used for the deregression of breeding values in country 1 is:

Bull	Sire	MGS	MGD
1	5	G2	G3
2	6	7	G4
3	5	2	G4
4	1	G2	G4
5	G1	G2	G3
6	G1	G2	G3
7	G1	G2	G3

The matrix $\mathbf{A}_1^{-1}$ was calculated according to the rules in Section 5.5.2.
In the first round of iteration, the transpose of the vector $\mathbf{Qg}_1 + \mathbf{s}_1$ in step 2 above is:

$$(\mathbf{Qg}_1 + \mathbf{s}_1)' = (9.0\ 10.1\ 15.8\ -4.7)$$

The vector of solutions for $\mathbf{p}_1$ and $\mathbf{g}_1$ in step 3 is computed as:

$$\begin{bmatrix}\hat{\mathbf{p}}_1\\ \hat{\mathbf{g}}_1\end{bmatrix} = \begin{bmatrix} 17.094 & 0.000 & 0.000 & -5.037 & -0.839 & -0.839 & 1.831\\ 0.000 & 13.736 & 1.831 & -5.037 & -2.518 & -2.518 & 1.831\\ 0.000 & 1.831 & 10.989 & -5.037 & -2.518 & -2.518 & 0.916\\ -5.037 & -5.037 & -5.037 & 8.555 & 3.777 & 3.777 & 0.000\\ -0.839 & -2.518 & -2.518 & 3.777 & 4.568 & 2.728 & 0.839\\ -0.839 & -2.518 & -2.518 & 3.777 & 2.728 & 3.728 & 0.000\\ 1.831 & 1.831 & 0.916 & 0.000 & 0.839 & 0.000 & 3.671 \end{bmatrix}^{-1}$$

$$\begin{bmatrix} 6.716 & -1.831 & 7.326 & 0.000\\ 0.000 & 7.326 & 0.000 & 0.000\\ 0.000 & 3.663 & 0.000 & 0.000\\ 0.000 & 0.000 & 0.000 & 0.000\\ 1.679 & 0.000 & 0.000 & 3.357\\ 3.357 & 0.000 & 0.000 & 0.000\\ -1.679 & 2.747 & 3.663 & 3.357 \end{bmatrix} \begin{bmatrix} 9.0\\ 10.1\\ 15.8\\ -4.7 \end{bmatrix} = \begin{bmatrix} 16.330\\ 12.861\\ 12.622\\ 23.481\\ -9.801\\ 12.375\\ -0.564 \end{bmatrix}$$

The transpose of the vector $(\mathbf{R}_1^{-1}\mathbf{y}_1)$ in step 4 is: (30.2 9.0 10.1 15.8) and:

$$1'\mathbf{R}_1^{-1}\mathbf{y}_1 = 2235.50$$

Therefore, in the first round of iteration (step 4):

$$\boldsymbol{\mu}_1 = 2235.50/253 = 8.835$$

Convergence was achieved after about six iterations. The transpose of the vector $(\mathbf{R}_1^{-1}\mathbf{y}_1)$ after convergence is:

$$(\mathbf{R}_1^{-1}\mathbf{y}_1)' = (563.928\ 1495.751\ 385.302\ -214.278)$$

with $\mathbf{R}_1^{-1}$ = diag(0.0172, 0.0067, 0.050, 0.04), the transpose of the vector of DRB calculated in step 7 is:

$$\mathbf{y}_1' = (9.7229\ 9.9717\ 9.2651\ -8.5711)$$

Appendix G: Calculating Φ, a Matrix of Legendre Polynomials Evaluated at Different Ages or Time Periods

The matrix $\mathbf{\Phi}$ is of order t (the number of days in milk or ages) by k (where k is the order of fit) with element $\phi_{ij} = \phi_j(a_t)$ equals the jth Legendre polynomial evaluated at the tth standardized age or days in milk (DIM). Thus a_t is the tth DIM or age standardized to the interval for which the polynomials are defined. Kirkpatrick *et al*. (1990, 1994) used Legendre polynomials that span the interval –1 to +1. Defining d_{min} and d_{max} as the first and latest DIM on the trajectory, DIM d_t can be standardized to a_t as:

$$a_t = -1 + 2(d_t - d_{min})/(d_{max} - d_{min})$$

In matrix notation, $\mathbf{\Phi} = \mathbf{M\Lambda}$, where $\mathbf{M}$ is the matrix containing the polynomials of the standardized DIM values and $\mathbf{\Lambda}$ is a matrix of order k containing the coefficients of Legendre polynomials. The elements of $\mathbf{M}$ can be calculated as $m_{ij} = (a_i^{(j-1)}$, $i = 1,\ldots t$; $j = 1,\ldots k$). For instance, given that $k = 5$ and that $t = 3$ (three standardized DIM), $\mathbf{M}$ is:

$$\mathbf{M} = \begin{bmatrix} 1 & a_1 & a_1^2 & a_1^3 & a_1^4 \\ 1 & a_2 & a_2^2 & a_2^3 & a_2^4 \\ 1 & a_3 & a_3^2 & a_3^3 & a_3^4 \end{bmatrix}$$

Using the fat yield data in Table 9.1 as an illustration, with ten DIM, the vector of standardized DIM is:

$$\mathbf{a}' = [-1.0\ -0.7778\ -0.5556\ -0.3333\ -0.1111\ 0.1111\ 0.3333\ 0.5556\ 0.7778\ 1.0]$$

and $\mathbf{M}$ is:

$$\mathbf{M} = \begin{bmatrix} 1.0000 & -1.0000 & 1.0000 & -1.0000 & 1.0000 \\ 1.0000 & -0.7778 & 0.6049 & -0.4705 & 0.3660 \\ 1.0000 & -0.5556 & 0.3086 & -0.1715 & 0.0953 \\ 1.0000 & -0.3333 & 0.1111 & -0.0370 & 0.0123 \\ 1.0000 & -0.1111 & 0.0123 & -0.0014 & 0.0002 \\ 1.0000 & 0.1111 & 0.0123 & 0.0014 & 0.0002 \\ 1.0000 & 0.3333 & 0.1111 & 0.0370 & 0.0123 \\ 1.0000 & 0.5556 & 0.3086 & 0.1715 & 0.0953 \\ 1.0000 & 0.7778 & 0.6049 & 0.4705 & 0.3660 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \end{bmatrix}$$

Next, the matrix $\mathbf{\Lambda}$ of Legendre polynomials needs to be computed. The jth Legendre polynomial evaluated at age $t(P_j(t))$ can in general be evaluated by the formula given by Abramowitz and Stegun (1965). In general, for the j integral:

$$P_j(t) = \frac{1}{2^j} \sum_{r=0}^{j/2} \frac{(-1)^r (2j-2r)!}{r!(j-r)!(j-2r)!} t^{j-2r}$$

where $j/2 = (j-1)/2$ if j is odd. The first five Legendre polynomials therefore are:

$P_0(t) = 1$; $P_1(t) = t$; $P_2(t) = \frac{1}{2}(3t^2 - 1)$

$P_3(t) = \frac{1}{2}(5t^3 - 3t)$; and $P_4(t) = \frac{1}{8}(35t^4 - 30t^2 + 3)$

The normalized value of the jth Legendre polynomial evaluated at age t ($\phi_j(t)$) can be obtained as:

$$\phi_j(t) = \sqrt{\frac{2n+1}{2} P_j(t)}$$

Thus:

$$\phi_0(t) = \sqrt{\tfrac{1}{2}}\, P_0(t) = 0.7071; \quad \phi_1(t) = \sqrt{\tfrac{3}{2}}\, P_1(t) = 1.2247(t)$$

$$\phi_2(t) = \sqrt{\tfrac{5}{2}}\, P_2(t) = 2.3717(t^2) - 0.7906; \quad \phi_3(t) = \sqrt{\tfrac{7}{2}}\, P_3(t) = 4.6771(t^3) - 2.8067(t)$$

$$\text{and} \quad \phi_4(t) = \sqrt{\tfrac{9}{2}} P_4(t) = 9.2808(t^4) - 7.9550(t^2) + 0.7955$$

Therefore, for $t = 5$ in Example 9.1, $\mathbf{\Lambda}$ is:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.7071 & 0.0000 & -0.7906 & 0.0000 & 0.7955 \\ 0.0000 & 1.2247 & 0.0000 & -2.8067 & 0.0000 \\ 0.0000 & 0.0000 & 2.3717 & 0.0000 & -7.9550 \\ 0.0000 & 0.0000 & 0.0000 & 4.6771 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 9.2808 \end{bmatrix}$$

and $\mathbf{\Phi} = \mathbf{M\Lambda}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} 0.7071 & -1.2247 & 1.5811 & -1.8704 & 2.1213 \\ 0.7071 & -0.9525 & 0.6441 & -0.0176 & -0.6205 \\ 0.7071 & -0.6804 & -0.0586 & 0.7573 & -0.7757 \\ 0.7071 & -0.4082 & -0.5271 & 0.7623 & 0.0262 \\ 0.7071 & -0.1361 & -0.7613 & 0.3054 & 0.6987 \\ 0.7071 & 0.1361 & -0.7613 & -0.3054 & 0.6987 \\ 0.7071 & 0.4082 & -0.5271 & -0.7623 & 0.0262 \\ 0.7071 & 0.6804 & -0.0586 & -0.7573 & -0.7757 \\ 0.7071 & 0.9525 & 0.6441 & 0.0176 & -0.6205 \\ 0.7071 & 1.2247 & 1.5811 & 1.8704 & 2.1213 \end{bmatrix} \qquad \text{(g.1)}$$

References

Abramowitz, M. and Stegun, I.A. (1965) *Handbook of Mathematical Functions*. Dover, New York.

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. and Lawlor, T.J. (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93, 743–752.

Albuquerque, L.G. and Meyer, K. (2001) Estimates of covariance functions for growth from birth to 630 days in Nelore cattle. *Journal of Animal Science* 79, 2776–2789.

Andersen, S. and Pedersen, B. (1996) Growth and food intake curves for group-housed gilts and castrated male pigs. *Animal Science* 63, 457–464.

Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Berger, P.J., Luecke, G.R. and Hoekstra, J.A. (1989) Iterative algorithms for solving mixed model equations. *Journal of Dairy Science* 72, 514–522.

Bergsma, R., Kanis, E., Knol, E.F. and Bijma, P. (2008) The contribution of social effects to heritable variation in finishing traits of domesticated pigs (*Sus scrofa*). *Genetics* 178, 1559–1570.

Bijma, P., Muir, W.M. and Van Arendonk, J.A.M. (2007a) Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics* 175, 277–288.

Bijma, P., Muir, W.M., Ellen, E.D., Wolf, J.B. and Van Arendonk, J.A.M. (2007b) Multilevel selection 2: Estimating the genetic parameters determining inheritance and response to selection. *Genetics* 175, 289–299.

Bohmanova, J., Miglior, F., Jamrozik, J., Misztal, I. and Sullivan, P.G. (2008) Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. *Journal of Dairy Science* 91, 3627–3638.

Boichard, D., Bonaiti, B., Barbat, A. and Mattalia, S. (1995) Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of Dairy Science* 78, 431–437.

Boichard, D., Fritz, S., Rossignol, M.N., Boscher, M.Y., Malafosse, A. and Colleau, J.J. (2002) Implementation of marker-assisted selection in French dairy cattle. In: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, August 2002, Montpellier, France, Communication no. 22-03.

Brotherstone, S. and Hill, W.G. (1991) Dairy herd life in relation to linear type traits and production. 2 Genetic analyses for pedigree and non-pedigree cows. *Animal Production* 53, 289–297.

Brotherstone, S., Veerkamp, R.F. and Hill, W.G. (1997) Genetic parameters for a simple predictor of the lifespan of Holstein-Friesian dairy cattle and its relationship to production. *Animal Science* 65, 31.

Brotherstone, S., White, I.M.S. and Meyer, K. (2000) Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Animal Science* 70, 407–415.

Bulmer, M.G. (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

Cameron, N.D. (1997) *Selection Indices and Prediction of Genetic Merit in Animal Breeding*. CAB International, Wallingford.

Cheverud, J.M. and Moore, A.J. (1994) Quantitative genetics and the role of the environment provided by relatives in the evolution of behaviour. In: Boake, C.R.B. (ed.) *Quantitative Genetics of Behavioral Evolution*. University of Chicago Press, Chicago, pp. 60–100.

Christensen, O. and Lund, M. (2010) Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42, 2.

Cockerham, C.C.(1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39, 859–882.

Cox, D.R. (1972) Regression models and life tables. *Journal of the Royal Statistical Society B* 34, 187–220.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society B* 39, 1–38.

Ducrocq, V. (1992) Solving animal model equations through an approximate incomplete Cholesky decomposition. *Genetics Selection Evolution* 24, 193–209.

Ducrocq, V. (1997) Survival analysis, a statistical tool for longevity data. *48th Annual Meeting of the European Association for Animal Production*, 25–28 August 1997, Vienna, Austria.

Ducrocq, V. (2000) *Survival Analysis Applied to Animal Breeding*. Station de Génétique Quantitative et Appliqueé. Institut National de la Recherche Agronomique. F-&8352 Jouy-en-Josas Cedex, France, p. 89.

Ducrocq, V. and Besbes, B. (1993) Solutions of multiple trait animal models with missing data on some traits. *Journal of Animal Breeding and Genetics* 110, 81–89.

Ducrocq, V. and Solkner, J. (1998) Implementation of a routine breeding value evaluation for longevity of dairy cows using survival analysis techniques. In: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, January 1998, Armidale, Australia, pp. 359–362.

Ducrocq, V., Quaas, R.L., Pollak, E.J. and Casella, G. (1988a) Length of productive life of dairy cows. 1. Justification of a Weibull model. *Journal of Dairy Science*, 71, 3061–3070.

Ducrocq, V., Quaas, R.L., Pollak, E.J. and Casella, G. (1988b) Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation. *Journal of Dairy Science* 71, 3071–3079.

Everett, R., Keown, J.R. and Clapp, E.E. (1976) Relationships among type, production and stayability in Holstein sire evaluation. *Journal of Dairy Science* 59, 1277–1285.

Falconer, D.S. and Mckay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, Harlow.

Fernando, R.L. (2010) Genomic selection in livestock. Animal breeding and genetics short courses. Summer 2010. Iowa State University, Ames, IA, USA.

Fernando, R.L. and Grossman, M. (1989) Marker-assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* 21, 467–477.
Fernando, R.L., Billingsley, R.D. and Gianola, D. (1983) Effects of method of scaling on heritability estimates and sire evaluations for frame size at weaning in Angus cattle. *Journal of Animal Science* 56, 1047–1056.
Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society, Edinburgh* 52, 399–433.
Fiske, W.F. and Banos, G. (2001) Weighting factors of sire daughter information in international genetic evaluations. *Journal of Dairy Science* 84, 1759–1767.
Foulley, J.L., Gianola, D. and Thompson, R. (1983) Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genetics Selection Evolution* 15, 401–424.
Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.
Gianola, D. (1982) Theory and analysis of threshold characters. *Journal of Animal Science* 54, 1079–1096.
Gianola, D. and Foulley, J.L. (1983) Sire evaluation for ordered categorical data with a threshold model. *Genetics Selection Evolution* 15, 201–224.
Gilmour, A.R., Thompson, R. and Cullis, B. (1995) Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440–1450.
Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. (2003) *ASREML UserGuide Release 1.0*. VSN International, Hemel Hempstead, 310 pp.
Goddard, M.E. (1985) A method of comparing sires evaluated in different countries. *Livestock Production Science* 13, 321–331.
Goddard, M.E. (1992) A mixed model for the analyses of data on multiple genetic markers. *Theoretical and Applied Genetics* 83, 878–886.
Goddard, M.E. (1999) New technology to enhance genetic improvement of pigs. *Manipulating Pig Production* 7, 44–52.
Goddard, M.E. and Hayes, B.J. (2002) Optimisation of response using molecular data. In: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, August 2002, Montpellier, France, Communication no. 22-01.
Griffing, B. (1967) Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Australian Journal of Biological Science* 20, 127–142.
Groeneveld, E.(1990) *PEST User's Manual*. Institute of Animal Husbandry and Animal Behaviour, Federal Agricultural Research Centre, Neustadt, Germany.
Groeneveld, E. and Kovac, M. (1990) A generalised computing procedure for setting up and solving mixed models. *Journal of Dairy Science* 73, 513–531.
Groeneveld, E., Kovac, M. and Wang, T. (1990) Pest, a general purpose BLUP package for multivariate prediction and estimation. In: *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production* 13, 488–491.
Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetical Research Cambridge* 72, 159–168.

Guo, Z. and Swalve, H.H. (1997) Comparison of different lactation curve sub-models in test day models. *Proceedings of the Annual Interbull Meeting, Vienna. Interbull Bulletin* 16, 75–79.
Habier, D., Fernando, R.L. and Dekkers, J.C.M. (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397.
Habier, D., Fernando, R.L, Kizikaya, K. and Garrick, D.J. (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186–197.
Haldane, J.B.S. (1919) The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* 8, 299.
Hayes, B. and Daetwyler, H. (2013) Genomic selection in the era of genome sequencing. Course notes, March 2013, Piacenza, Italy.
Hayes, B.J., Visscher, P.M. and Goddard, M.E. (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* 91, 47–60.
Henderson, C.R. (1949) Estimation of changes in herd environment. *Journal of Dairy Science* 32, 709 (abstract).
Henderson, C.R. (1950) Estimation of genetic parameters. *Annals of Mathematical Statistics* 21, 309.
Henderson, C.R. (1953) Estimation of variance and covariance components. *Biometrics* 9, 226–252.
Henderson, C.R. (1963) Selection index and expected genetic advance. In: Hanson, W.D. and Robinson, H.F. (eds), *Statistical Genetics and Plant Breeding*. Publication 982, National Academy of Sciences, National Research Council, Washington DC, pp. 141–163.
Henderson, C.R. (1973) Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium in Honour of J.L. Lush*. American Society for Animal Science, Blackburgh, Champaign, IL, pp. 10–41.
Henderson, C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447.
Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69–83.
Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph Press, Guelph, Canada.
Henderson, C.R. (1985) Best linear unbiased prediction of non-additive genetic merits in non-inbred populations. *Journal of Animal Science* 60, 111–117.
Henderson, C.R. and Quaas, R.L. (1976) Multiple trait evaluation using relatives' records. *Journal of Animal Science* 43, 1188–1197.
Henderson, C.R., Kempthorne, O., Searle, S.R. and von Krosigk, C.M. (1959) The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.
Heringstad, B., Chang, Y.M., Gianola, D. and Klemetsdal, G. (2002) Analysis of clinical mastitis in Norwegian cattle with a longitudinal threshold model. In: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, August 2002, Montpellier, France, Communication no. 20-03.
Hoeschele, I. and VanRaden, P.M. (1991) Rapid inverse of dominance relationship matrices for non-inbred populations by including sire by dam subclass effects. *Journal of Dairy Science* 74, 557–569.
IMSL (1980) *Library Reference Manual*. IMSL, Houston, Texas.

Interbull (2000) *National Genetic Evaluation Programmes for Dairy Production Traits Practised in Interbull Member Countries 1999–2000*. Department of Animal Breeding and Genetics, Uppsala, Sweden, Bulletin 24.

Jairath, L., Dekkers, J.C.M., Schaeffer, L.R., Liu, Z., Burnside, E.B. and Kolstad, B. (1998) Genetic evaluation for herd life in Canada. *Journal of Dairy Science* 81, 550–652.

Jamrozik, J. and Schaeffer, L.R. (1997) Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. *Journal of Dairy Science* 80, 762–770.

Jamrozik, J., Schaeffer, L.R. and Dekkers, J.C.M. (1997) Genetic evaluation of dairy cattle using test day yields and random regression model. *Journal of Dairy Science* 80, 1217–1226.

Jensen, J. and Madsen, P. (1997) *A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models*. National Institute of Animal Science, Research Center, Foulum, Denmark.

Jensen, J., Wang, C.S., Sorensen, D.A. and Gianola, D. (1994) Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agriculturæ Scandinavica* 44, 193–201.

Kachman, S.D. (1999) Applications in survival analysis. *Journal of Animal Science* 77, 147–153.

Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–481.

Kemp, R.A. (1985) The effects of positive assortative mating and preferential treatment of progeny on the estimation of breeding values. PhD thesis (unpublished), University of Guelph, Guelph, Canada.

Kempthorne, O. and Nordskog, A.W. (1959) Restricted selection indexes. *Biometrics* 15, 10–19.

Kennedy, B.W. (1989) *Animal Model BLUP*. Erasmus Intensive Graduate Course, Trinity College, Dublin.

Kirkpatrick, M. and Meyer, K. (2004) Direct estimation of genetic principal components: Simplified analysis of complex phenotypes. *Genetics* 168, 2295–2306.

Kirkpatrick, M., Lofsvold, D. and Bulmer, M. (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979–993.

Kirkpatrick, M., Hill, W.G. and Thompson, R. (1994) Estimating the covariance of traits during growth and ageing, illustrated with lactation in dairy cattle. *Genetics Research Cambridge* 64, 57–69.

Legarra, A. and Misztal, I. (2008) Technical note: Computing strategies in genome-wide selection. *Journal of Dairy Science* 91, 360–366.

Legarra, A., Aguilar, I. and Misztal, I. (2009) A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92, 4656–4663.

Lidauer, M., Stranden, I., Mäntysaari, E.A, Pösö, J. and Kettunen, A. (1999) Solving large test-day models by iteration on the data and preconditioned conjugate gradient. *Journal of Dairy Science* 82, 2788–2796.

Lidauer, M., Matilainen, K., Mäntysaari, E.A. and Stradén, I. (2011) Mixed model equations Solver, MiX99. Biometrical Genetics, MTT Agrifood Research, Finland. http://www.mtt.fi/BGE/Software/MiX99

Liu, Z., Reinhardt, F. and Reents, R. (2002) The multiple trait effective daughter contribution method applied to approximate reliabilities of estimated breeding

values from a random regression test day model for genetic evaluation in dairy cattle. In: *Proceedings of the 7th World Congress Applied to Livestock Production*, August 2002, Montpellier, France, Communication no. 20-15.
Liu, Z., Seefried, F.R., Reinhardt, F., Rensing, S., Thaller, G. and Reents, R. (2011) Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genetics Selection Evolution* 43, 19.
Madgwick, P.A. and Goddard, M.E. (1989) Genetic and phenotypic parameters of longevity in Australian dairy cattle. *Journal of Dairy Science* 72, 2624–2632.
Mantysaari, E.A. (1999) Derivation of multiple trait reduced random regression model for the first lactation test day records of milk, protein and fat. In: *Proceedings 50th Annual Meeting of EAAP*, August 1999, Zurich, Switzerland.
Maron, M.J. (1987) *Numerical Analysis*. Macmillan Publishing Company, New York.
Meijering, A. and Gianola, D. (1985) Linear versus nonlinear methods of sire evaluations for categorical traits: A simulation study. *Genetics Selection Evolution* 17, 115–132.
Mészáros, G., Pálos, J., Ducrocq, V. and Sölkner, J. (2010) Heritability of longevity in Large White and Landrace sows using continuous time and grouped data models. *Genetics Selection Evolution* 42, 13.
Mészáros, G., Sölkner, J. and Ducrocq, V. (2013) The Survival Kit: Software to analyse possible survival data including correlated random effects. *Computer Methods and Programs in Biomedicine* 110, 503–510.
Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* 75, 934–940.
Meuwissen, T.H.E. and Luo, Z. (1992) Computing inbreeding coefficients in large populations. *Genetics Selection Evolution* 24, 305–313.
Meuwissen, T.H., Hayes, B. and Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
Meyer, K. (1989) Approximate accuracy of genetic evaluation under an animal model. *Livestock Production Science* 21, 87–100.
Meyer, K. (1999) Estimates of genetic and phenotypic covariance functions for post-weaning growth and mature weight of beef cows. *Journal of Animal Breeding and Genetics* 116, 181–205.
Meyer, K. (2005) Genetic principal components for live ultrasound scan traits of Angus cattle. *Animal Science* 81, 337–345.
Meyer, K. (2007) Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. *Journal of Animal Breeding and Genetics* 124, 50–64.
Meyer, K. (2009) Factor-analytic models for genotype x environment type problems and structured covariance matrices. *Genetics Selection Evolution* 41, 21–31.
Meyer, K. and Hill, W.G. (1997) Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. *Livestock Production Science* 47, 185–200.
Meyer, K. and Tier, B. (2003) Approximate prediction covariances among multiple estimated breeding values for individuals. In: *Proceedings of the Annual Interbull Meeting, Rome. Interbull Bulletin* 31, 133–136.
Misztal, I. (2006) Properties of random regression models using linear splines. *Journal of Animal Breeding and Genetics* 123, 74–80.

Misztal, I. and Gianola, D. (1988) Indirect solution of mixed model eqautions. *Journal of Dairy Science* 71(Suppl 2), 99–106.
Misztal, I., Aguilar, I., Legarra, A., Tsuruta, S., Johnson, D.L. and Lawlor, T.J. (2010) A unified approach to utilize phenotypic, full pedigree and genomic information for genetic evaluation. In: *Proceedings of the 9th World Congress Applied to Livestock Production*, Leipzig, Germany, Communication no. 0050.
Moore, A.J., Brodie Jr, E.D. and Wolf, J.B. (1997) Interacting phenotypes and the evolutionary process. I. Direct and indirect genetic effects of social interactions. *Evolution* 51, 1352–1362.
Moser, G., Khatkar, M.S., Hayes, B.J. and Raadsma, H.W. (2010) Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genetics Selection Evolution* 42, 37.
Mrode, R.A. and Swanson, G.J.T. (1995) Comparison of the efficiency of the repeatability model on observed or transformed yields to a multi-variate analysis. *Summaries of Paper, British Society of Animal Science Winter Meeting*, March 1995, paper no. 129.
Mrode, R.A. and Swanson, G.J.T. (1999) Simplified equations for evaluations of bulls in the Interbull international evaluation system. *Livestock Production Science* 62, 43–52.
Mrode, R.A. and Swanson, G.J.T. (2004) Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. *Livestock Production Science* 86, 253–260.
Mrode, R., Coffey, M. and Berry, D.P. (2010) Understanding genomic evaluations from various evaluation methods and GMACE. *Interbull Bulletin* 42, 52–55.
Muir, W.M. (1996) Group selection for adaptation to multiple-hen cages: Selection program and direct responses. *Poultry Science* 75, 447–458.
Muir, W.M. (2005) Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* 170, 1247–1259.
Pander, B.L., Hill, W.G. and Thompson, R. (1992) Genetic parameters of test day records of British Holstein heifers. *Animal Production* 55, 11–21.
Patterson, H.D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
Phocas, F. and Ducrocq, V. (2006) Discrete vs continuous time survival analysis of productive life of Charolais cows. In: *Proceedings of the 8th World Congress Applied to Livestock Production*, Belo Horizonte, Communication no. 03-13.
Piles, M., Garreau, H., Rafel, O., Larzul, C., Ramon, J. and Ducrocq, V. (2006) Survival analysis in two lines of rabbits selected for reproductive traits. *Journal of Animal Science* 84, 1658–1665.
Prentice, R.L. and Gloeckler, L.A. (1978) Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 34, 57–67.
Ptak, E. and Schaeffer, L.R. (1993) Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science* 34, 23–34.
Quaas, R.L. (1976) Computing the diagonal elements of a large numerator relationship matrix. *Biometrics* 32, 949–953.
Quaas, R.L. (1984) *Linear Prediction in BLUP School Handbook: Use of Mixed Models for Prediction and Estimation of (Co)variance Components*. Animal Genetics and Breeding Unit, University of New England, NSW, Australia, pp. 1–76.
Quaas, R.L. (1995) Fx Algorithms. An unpublished note.

Quaas, R.L. and Pollak, E.J. (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *Journal of Animal Science* 51, 1277–1287.
Quaas, R.L. and Pollak, E.J. (1981) Modified equations for sire models with groups. *Journal of Dairy Science* 64, 1868–1872.
Robertson, A. and Rendel, J.M. (1954) The performance of heifers got by artificial insemination. *Journal of Agricultural Science, Cambridge* 44, 184–192.
Ronningen, K. and Van Vleck, L.D. (1985) Selection index theory with practical applications. In: Chapman, A.B. (ed.) *General and Quantitative Genetics*. World Animal Science, A4, Elsevier Science Publishers, Oxford.
Schaeffer, L.R. (1984) Sire and cow evaluation under multiple trait models. *Journal of Dairy Science* 67, 1567–1580.
Schaeffer, L.R. (1994) Multiple-country comparison of dairy sires. *Journal of Dairy Science* 77, 2671–2678.
Schaeffer, L.R. (2004) Application of random regression models in animal breeding. *Livestock Production Science* 86, 35–45.
Schaeffer, L.R. and Dekkers, J.C.M. (1994) Random regression in animal models for test-day production in dairy cattle. In: *Proceedings of the 5th World Congress Applied to Livestock Production*, Guelph, Canada, pp. 443–446.
Schaeffer, L.R. and Kennedy, B.W. (1986) Computing solutions to mixed model equations. In: *Proceedings of the 3rd World Congress on Genetics Applied to Livestock Production*, Lincoln, NE, p. 392.
Schaeffer, L.R. and Wilton, J.W. (1987) RAM computing strategies and multiple traits. Prediction of genetic value for beef cattle. In: *Proceedings Workshop II*, Winrock, Kansas City, MO.
Schaeffer, L.R., Wilton, J.W. and Thompson, R. (1978) Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* 34, 199–208.
Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. and van Doormaal, B.J. (2000) Experience with a test day model. *Journal of Dairy Science* 83, 1135–1144.
Searle, S.R. (1982) *Matrix Algebra Useful for Statistics*. John Wiley & Sons, New York.
Sigurdsson, A. and Banos, G. (1995) Dependent variables in international sire evaluations. *Acta Agriculturæ Scandinavica* 45, 209–217.
Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Ødegard, J. and Meuwissen, T.H.E. (2009) Persistence of accuracy of genomic-wide breeding values over generations when including a polygenic effect. *Genetics Selection Evolution* 41, 53.
Sorensen, D. and Gianola, D. (2002) *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer-Verlag, New York.
Sorensen, D.A. and Kennedy, B.W. (1983) The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theoretical and Applied Genetics* 66, 217–220.
Sorensen, D.A., Andersen, S. and Gianola, D. (1995) Bayesian inference in threshold models using Gibbs sampling. *Genetics Selection Evolution* 27, 229–249.
Stranden, I. and Lidauer, M. (1999) Solving large mixed linear models using preconditioned conjugate gradient iteration. *Journal of Dairy Science* 82, 2779–2787.
Szyda, J., Liu, Z., Reinhardt, F. and Reents, R. (2003) Incorporation of QTL information into routine estimation of breeding values for German Holstein dairy cattle. In: *Proceedings of the 54th Annual Meeting of the European Association for Animal Production*, Rome, 2003, p. 11.

Tarrés, J., Bidanel, J.P., Hofer, A. and Ducrocq, V. (2006) Analysis of longevity and exterior traits on Large White sows in Switzerland. *Journal of Animal Science* 84, 2914–2924.

Ter Braak, C.J.F., Boer, M.P. and Bink, M.C.A.M. (2005) Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. *Genetics* 170, 1435–1443.

The R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Thompson, R. (1977a) The estimation of heritability with unbalanced data. II. Data available on more than two generations. *Biometrics* 33, 497–504.

Thompson, R. (1977b) Estimation of quantitative genetic parameters. In: Pollak, E., Kempthorne, O. and Bailey, T.B. (eds) *Proceedings of the International Conference on Quantitative Genetics*, Iowa State University Press, Ames, IO, pp. 639–657.

Thompson, R. (1979) Sire evaluation. *Biometrics* 35, 339–353.

Thompson, R. and Meyer, K. (1986) A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science* 15, 299–313.

Thompson, R., Cullis, B.R., Smith, A.B. and Gilmour, A.R. (2003) A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian and New Zealand Journal of Statistics* 45, 445–459.

Tier, B. and Solkner, J. (1993) Analysing gametic variation with an animal model. *Theoretical and Applied Genetics* 85, 868–872.

Tyriseva, A.M., Meyer, K., Freddy, F., Ducrocq, V., Jakobsen, J., Lidauer, M.H. and Mantysaari, E. (2011a) Principal component approach in variance component estimation in international sire evaluation. *Genetics Selection Evolution* 43, 21.

Tyriseva, A.M., Meyer, K., Freddy, F., Ducrocq, V., Jakobsen, J., Lidauer, M.H. and Mantysaari, E. (2011b) Principal component and factor analytic models in international sire evaluation. *Genetics Selection Evolution* 43, 33.

Van Arendonk, J.A.M., Tier, B. and Kinghorn, B.P. (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319–329.

VanRaden, P.M. (2007) Genomic measures of relationship and inbreeding. *Interbull Bulletin* 37, 33–36.

VanRaden, P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414–4423.

VanRaden, P.M. and Hoeschele, I. (1991) Rapid inverse of additive by additive relationship matrices by including sire-dam combination effects. *Journal of Dairy Science* 74, 570–579.

VanRaden, P.M. and Klaaskate, E.J.H. (1993) Genetic evaluation of length of productive life including predicted longevity of live cows. *Journal of Dairy Science* 76, 2758–2764.

VanRaden, P.M. and Wiggans, G.R. (1991) Derivation, calculation and use of national animal model information. *Journal of Dairy Science* 74, 2737–2746.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegart, T.S., Schnabel, R.D., Taylor, J.F. and Schenkel, F.S. (2009) Invited Review. Reliability of genomic predictions for North American Holsteins bulls. *Journal of Dairy Science* 92, 16–24.

Veerkamp, R., Brotherstone, S. and Meuwissen, T.H.E. (1999) Survival analysis using random regression models. *Interbull Bulletin* 21, 36–40.

Veerkamp, R.F., Hill, W.G., Stott, A.W., Brotherstone, S. and Simm, G. (1995) Selection for longevity and yield in dairy cows using transmitting abilities for type and yield. *Animal Science* 61, 189–197.
Visscher, P., Thompson, R., Yazdi, H., Hill, W.G. and Brotherstone, S. (1999) Genetic analysis of longevity data in the UK: Present practice and considerations for the future. *Interbull Bulletin* 21, 16–23.
Visscher, P.M. (1991) Estimation of genetic parameters in dairy cattle using an animal model and implications for genetic evaluation. PhD thesis (unpublished), University of Edinburgh, Edinburgh.
Wang, C.S., Rutledge, J.J. and Gianola, D. (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genetics Selection Evolution* 25, 41–62.
Wang, C.S., Rutledge, J.J. and Gianola, D. (1994) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics Selection Evolution* 26, 91–115.
Wang, Y., Miller, S.P., Schenkel, F.S., Wilton, J.W. and Boettcher, P.J. (2002) Performance of a linear-threshold model to evaluate calving interval and birth weight in a multibreed beef population. In: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production,* August 2002, Montpellier, France, Communication no. 17-10.
Weller, J.I. (2001) *Quantitative Trait Loci Analysis in Animals*. CAB International, Wallingford.
Westell, R.A. and VanVleck, L.D. (1987) Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *Journal of Dairy Science* 70, 1006–1017.
Westell, R.A., Quaas, R.L. and VanVleck, L.D. (1988) Genetic groups in an animal model. *Journal of Dairy Science* 71, 1310–1318.
White, I.M.S., Thompson, R. and Brotherstone, S. (1999) Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science* 82, 632–638.
Wiggans, G.R., VanRaden, P.M. and Powell, R.L. (1992) A method for combining United States and Canadian bull proofs. *Journal of Dairy Science* 75, 2834–2839.
Wiggans, G.R., Misztal, I. and Van Tassell, C.P. (2003) Calving ease (co)variance componenets for a sire-maternal grandsire threshold model. *Journal of Dairy Science* 86, 1845–1848.
Willham, R.L. (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* 19, 18–27.
Wolf, J.B., Brodie, E.D., Cheverud, J.M., Moore, A.J. and Wade, M.J. (1998) Evolutionary consequences of indirect genetic effects. *Trends in Ecological Evolution* 13, 64–69.
Wright, S. (1922) Coefficients of inbreeding and relationship. *American Naturalist* 56, 330–338.
Xu, S.Z (2003) Estimating polygenic effects using markers of the entire genome. *Genetics* 163, 789–801.

Index

adjusted right-hand side (ARHS) 280
animal evaluation model
  accuracy, evaluations 44–46
  construction, mixed model equations
    breeding values, progeny 42
    EBV 42
    least squares equations (LSE) 38–39
    male calves 40
    Mendelian sampling 41
    MME 39–40
    non-zero off-diagonals 40
    predicted breeding values 42
    progeny contribution 40–41
  description 37
  in groups
    assigned parents 56
    data set and genetic parameters 56
    design matrices and MME 57–60
    G1 recoding 57
    matrix notation 55
    Mendelian sampling 56
    pedigree file 57
    relationship matrix 54
    sires 55
    strategy 55
  pre-weaning gain 37
  PYD *see* progeny yield deviation (PYD)
approximate reliability, genetic evaluations
  animal model 314–315
  random regression models
    ancestors 317
    observation, animal 316
ARHS *see* adjusted right-hand side (ARHS)
associative breeding value (SBV) 122–123, 127–128

Bayesian methods, SNP
  BayesA
    chi distribution 194
    conditional distribution 194
    Gibbs sampling chain 197
    random number 196
    reference population 195
    solutions and variances 196–197
    t-distribution 194
  BayesB 197–199
  BayesC 199–200
  BayesCπ 201
  description 193
  genetic variances 193
best linear unbiased prediction (BLUP)
  animal evaluation model *see* animal evaluation model
  animal models and groups 54–60
  breeding value 36
  data span 37
  derivation 311–312
  description 34
  genetic evaluation 34
  identity matrix 36
  minimization 35
  mixed linear model 35
  MME 36, 312–313
  numerator relationship matrix 35
  progeny contribution (PC) 313
  properties 34
  reduced animal model 49–54
  selection index 36
  sire model *see* sire model
birth weight (BW) 235–238
BLUP *see* best linear unbiased prediction (BLUP)
breeding value prediction
  correlated response (CR) 11
  genetic markers *see* genetic markers
  pedigree 9–10
  progeny records *see* progeny records
  repeated records
    accuracy, EBV 5
    environmental correlations 4
    heritability and repeatability 4
    mean yield, milk 6
    multiple measurements 3
    percentage 5
    ratio, accuracy 5
  single record 2–3
BW *see* birth weight (BW)

calving difficulty (CD) 235–238
canonical transformation
  calculation, variables 97
  calculation procedure 318–319
  data, pre-weaning weight and gain 320–321
  design matrices 95, 97
  identity matrix 96
  missing records and same incidence matrices 319–320
  single trait analysis 95, 96
  solutions 321–322
  traits observations 318
  variance and covariance matrices 96
  WWG and PWG, beef calves 97–98
CD *see* calving difficulty (CD)
CFs *see* covariance functions (CFs)
Cholesky decomposition 322
Cholesky transformation
  defining 98–99
  loss of records 98
  residual and additive genetic covariance, beef calves 99–100
  solutions 101
common environmental effects, BLUP
  definition 67
  resemblance 66–67
  variation 67
  weaning weight, piglets 67–69
covariance functions (CFs)
  body weight measurements 149–151
  description 149
  reduced order fitting 151–155
  standardized ages 149
covariance matrix calculation
  alternative recursive method 160–161
  inverse calculation
    description 161–162
    identity matrix 164–165
    inbreeding coefficient 162
    maternal and paternal MQTL 164
    residual effect 162
    sire and dam 163
  maternal MQTL 157
  numerical application 158–160
  recombination rate 158
  tabular method 158

daughter yield deviations (DYDs) 77–78, 82–83, 86–87
  see also progeny yield deviation (PYD)
DBV *see* direct breeding value (DBV)
deregressed breeding values (DRB)
  iterative procedure 323
  and MACE 323
  pedigree structure 324
deregressed proofs (DRP) 87, 89
diagonal matrix 300
direct breeding value (DBV) 122–123, 127–128
dominance matrix
  inverse, relationship matrix $\mathbf{F}^{-1}$ 212–215
  subclass effects 210–211
  pairs of genes 209
  partial regression coefficient 209
  prediction 211–212
DRB *see* deregressed breeding values (DRB)
DRP *see* deregressed proofs (DRP)
DYDs *see* daughter yield deviations (DYDs)

EDC *see* effective daughter contributions (EDC)
effective daughter contributions (EDC) 88–89
epistasis
  Hadamard product 215
  inverse, relationship matrix
    calculation 217–218
    rules 216–217

fat yield in lactation 1, FAT1 and FAT2 81, 82
fixed regression model
  incidence matrices, MME 133–135
  lactation curve 133
  permanent environmental effects 135–136
  TDY, cows 132

Gauss–Seidel iteration 275–276
GEBV *see* genomic breeding values (GEBV)
genetic covariance
  BLUP 22
  decomposing, relationship matrix 23–25
  description 22
  numerator relationship matrix *see* numerator relationship matrix
  sire and MGS model 30–33

genetic evaluation
additive effects 1
breeding value prediction *see* breeding value prediction
description 1
dominance and epistasis 1–2
pedigree 9–10
phenotypic observation 1
prediction, breeding value 2
progeny records, breeding value prediction 6–8
selection index *see* selection index
genetic markers
additive genetic merit, MQTL 167–169
BLUP procedure 156
chromosomes identification 156
covariance matrix calculation 157–160
definition 156–157
DNA microsatellites 156
prediction, breeding values 165–166
QTL bracketed, markers *see* quantitative trait loci (QTL)
total additive genetic merit prediction 169–171
genetic parameters
eigenvalues 254
environmental effect 253
estimation, animal model 255–257
measures, daughters of sire 253–254
prediction, breeding values 251
scaled contrast 255
univariate sire model 251
WWG, beef calves 252–253, 257–259
genomic breeding values (GEBV)
Bayesian methods 193–201
chromosome segments 178
coding and scaling genotypes 178–179
cross-validation and genomic reliabilities 202
DGV and GEBV 177
fixed effect model, SNP 179–182
genomic selection 178
linear model 178
mixed linear model
polygenic effects 188–190
SNP effects 182–188
MME 202
QTL 177
reference and selection 182
selection of animals 177
single-step approach 190–193
SNP-BLUP model 203
SNPs *see* single nucleotide polymorphisms (SNPs)
young animals 178
Gibbs sampling
Gibbs sampler 261
joint posterior distribution 260
MCMC 260
multivariate animal model *see* multivariate animal model
univariate animal model *see* univariate animal model

herd–test–day (HTD) effect 132
herd–year–season (HYS) 80

IGE *see* indirect genetic effects (IGE)
inbreeding calculation, L matrix
calculation 306–307
modification 308–310
pedigree 307–308
indirect genetic effects (IGE) 121
iteration, mixed model equations
animal model with groups
data preparation 282–283
description 282
iterative stage 283–284
animal model without groups
ARHS 280
convergence criterion 281
pedigree file 278–279
predicting breeding 278
description 271–272
Gauss–Seidel iteration 275–276
genetic maternal effects 277
Jacobi iteration *see* Jacobi iteration
progeny 278
random animal effect 276
reduced animal model, maternal effects *see* reduced animal model (RAM), maternal effects

Jacobi iteration
convergence criterion 272, 273
diagonal element 272–273
random animal effect 273
sex effect 274
simultaneous equations 272
joint and full conditional distributions
Gibbs sampling 263

joint and full conditional distributions (*continued*)
multiple chain/short chain approach 264
parameters 262

Legendre polynomials
evaluation 325–326
linear equations
description 271
direct inversion 271
iteration, mixed model equations *see* iteration, mixed model equations
PCG *see* preconditioned conjugate gradient (PCG)
longitudinal data
beef cattle 130
CFs *see* covariance functions (CFs)
covariance function and RRM equivalence 155
fixed regression model 131–136
repeated measurements 130
RRM *see* random regression model (RRM)
test day records 130

MACE *see* multi-trait across-country evaluations (MACE)
Markov chain Monte Carlo (MCMC) methods 260
maternal grandsire (MGS) model
description 30
inbreeding coefficients 31
pedigree 32
pertains to males 30–31
recoding sires 32
$\mathbf{T}^{-1}$ and $\mathbf{D}^{-1}$ matrix, pedigree 32–33
without inbreeding 32
maternal trait models
animal
birth weight, beef calves 110–111
BLUE 110
design matrices 111–114
genetic and permanent environmental effects 110
RAM 115–119
components 109
measurements 109
phenotypic expression 109
progeny performance 109
sire and grandsire 119–120
matrix algebra
addition and subtraction 302
column vector 299
definition 299
diagonal matrix 300
direct product 302–303
eigenvalues and eigenvectors 305
inverse matrix 303–304
multiplication 302
rank of matrix 304
singular matrix 305
square matrix 300
symmetric matrix 301
transpose, matrix 301
triangular matrix 300–301
MBLUP *see* multivariate best linear unbiased prediction (MBLUP)
MCMC *see* Markov chain Monte Carlo (MCMC) methods
MGS model *see* maternal grandsire (MGS) model
mixed model equations (MME)
animal and dominance genetic effects 206–208
total genetic merit 208
MME *see* mixed model equations (MME)
multi-trait across-country evaluations (MACE)
analysis, DYD 86–87
computing
EDC 88–89
sire breeding values 89–91
dependencies 88
and DRP 87, 89
maternal grandsire model 87–88
MME 87
partitioning, bull evaluations 91–94
multivariate analysis
effects 95
factor analysis
covariances 102
FA model 105
loadings 102
multi-trait linear mixed model 102–103
WWG and PWG 103–104
limitations 95
parameter estimation and genetic evaluation 95
principal component analysis

eigenvalues 105–106
MME 106
reduced rank 106–108
transformation
canonical *see* canonical transformation
Cholesky *see* Cholesky transformation
multivariate animal model
conditional probabilities 267–268
description 266–267
numerical illustration 269–270
prior distributions 267
multivariate best linear unbiased prediction (MBLUP)
advantages and disadvantages 70
environments
different traits recorded, relatives 84–86
DRB 84, 89
genetic correlation 84
MACE *see* multi-trait across-country evaluations (MACE)
milk yield 84
equal design matrices with missing records
differences, sex solutions 80
genetic parameters 78
loss of traits 78
MME 79
WWG and PWG 78–79
equal design matrices with no missing records
accuracy 76–77
calculation, DYDs 77–78
defining 71–72
MME calculation 73–74
partitioning animal evaluations 74–76
WWG and PWG 72–73
trait evaluation 70
unequal design matrices
computation, DYD 82–83
FAT1 and FAT2 81, 82
HYS effects 80, 81
MME 81–82

non-additive animal models
and dominance effect
genetic effects 206–208
MME and BLUP 205
total genetic merit 208
dominance relationship matrix 204–205
epistasis *see* epistasis
inversion, dominance matrix *see* dominance matrix
statistical framework 204
numerator relationship matrix
decomposing 23–25
inbreeding coefficient 23
inverse
$\mathbf{A}^{-1}$ to zero 26
accounting, inbreeding 28–30
animals, pedigree 27
inbreeding ignorance 27
pedigree 27–28
pedigree, six animals 23
probability, identical genes 22

ordered categorical traits
animal breeding 219
binary trait 219–220
joint analysis, quantitative and binary traits
Bayesian approach 230
data and model definition 230–233
numerical application 234–239
linear and non-linear models 219
threshold model *see* threshold model

PCG *see* preconditioned conjugate gradient (PCG)
post-weaning gain (PWG) 72–73, 78–79
pre-weaning gain (WWG) 72–73, 78–79, 252–253, 257–259
preconditioned conjugate gradient (PCG)
computation strategy 293
computing starting values 294–295
convergence criteria 296–297
genetic evaluation models 292
iterative stage 295–296
multivariate accumulation technique 294
pseudo-code 293
progeny records
breeding value 7
calculations 7
description 6
EBV 8
half-sib progeny 8

progeny records (*continued*)
intra-class correlation 7
paternal half-sibs 6
predicted breeding value 8
progeny yield deviation (PYD)
accuracy, predictions 44–46
breeding value 43
dairy cattle evaluation 42
description 42
and DYD 42
univariate animal model 43
PWG *see* post-weaning gain (PWG)
PYD *see* progeny yield deviation (PYD)

QTL *see* quantitative trait loci (QTL)
quantitative trait loci (QTL)
accuracy, EBVs 171
BLUP equations 172
chromosome, marked loci 171
covariance matrix calculation
design matrices 176
genetic parameters 175
genotype, sire 172
inherited marker haplotype 175
loci, marker 172
marker haplotype 174
milk production traits 176
parental MQTL 173
types of gametes 173

RAM *see* reduced animal model (RAM)
random environmental effects, BLUP
breeding 61
common *see* common environmental effects, BLUP
component 61
repeatability model
breeding values 61–62
data analysis 61
definition 62
DYD 66
HYS effects, dairy 62–66
random regression model (RRM)
animal pe and maternal pe effects 137
animal solutions partition 142–145
breeding values 146–147
and covariance function equivalence 155
dairy cattle test day records 137
DIM calculation 138
DYD calculation 145–146
genetic differences 136
Legendre polynomials 138
maternal traits 148
matrices, MME 138–141
spline function 147–148
reduced animal model (RAM)
ancestors 54
back-solving
direct effects 118–119
maternal effects 119
non-parents 118
data set and genetic parameters 116–117
definition 49–51
design matrices 116–118
error 115
genetic and permanent environmental effect 115
genetic parameters 51
maternal effects
ARHS 289
data file 284–285
fixed effects 286
genetic parameters 284
parental records 284
pedigree file 284
permanent environmental effects 291–292
solving, animal solutions 286–288
MME 51–53, 115–116
non-parents solutions 53–54
regression survival models
accelerated failure time 247
group data 250
mixed 247–250
stratified proportional hazard 246
time-dependent risk factors 247
residual regression coefficient 223
RRM *see* random regression model (RRM)

S-MGS *see* sire-maternal grandsire (S-MGS)
SBV *see* associative breeding value (SBV)
selection index
accuracy 12–13
aggregate genotype prediction 16–18
animal and relatives records 15–16
data available, correlated traits 14
description 11
economic indices 18–19
matrix form 12

multiple linear regression 12
phenotype and genetic marker information 21
properties 11
restricted 19–20
single records, individual and relatives 14–15
single nucleotide polymorphisms (SNPs)
chromosome segments 178
DNA sequence variation 177
fixed effect model
animal breeding context 179
and EDC 180
GEBV 182
marker effects prediction 181
MME 180
and polygenic effects 180
reference animals 181
steps 179
mixed linear model
GBLUP 184–187
selection index approach 187–188
SNP-BLUP 183–184
sire model
description 46
design matrices and MME 47–48
in matrix notation 46
sire-maternal grandsire (S-MGS) 119–120
SNPs *see* single nucleotide polymorphisms (SNPs)
social interaction
animal model
associative effects 123
MME 123
pigs 125–127
residual covariance 124–125
correlated error structure 128–129
DBV and SBV 122–123, 127–128
direct effects 121
economic and welfare problems 121
food competition 121
IGE 121
non-additive components 122
schemes 121–122
square matrix 300
survival analysis
censoring 240–241
definition 240
function 240
linear models 241
livestock production 240
non-parametric estimation 245
proportional hazard models
definition 243–244
exponential distribution 244
Weibull distribution 244–245
random regression models 241–243
regression models *see* regression survival models
symmetric matrix 301

TDY *see* test day fat yields (TDY)
test day fat yields (TDY) 132
threshold model
calving ease score distribution 223
categorical traits 223
and data organization 221–222
functions, normal distribution 220–221
inverse, pedigree 224
iteration, subclass 225–230
pedigree 223
probability, category 229–230
system of equations 224–225
triangular matrix 300–301

univariate animal model
Gibbs sampling output 264–265
joint and full conditional distributions 262–264
numerical application 265–266
prior distributions 261–262

WWG *see* pre-weaning gain (WWG)