

7A

Mixed-Model and Bayesian Analysis of Short-Term Selection Experiments

Draft version 14 Nov 2005

Up to this point, we have been analyzing selection experiments by least-squares (LS) analysis, which requires only the phenotypic means and the selection differentials. However, in some case (such as animal breeding experiments) we have much more extensive information, such as the measures (or **records**) of all individuals throughout the course of the selection experiment and the pedigree of all these individuals as well. Even when such data is available, LS analysis only concerns itself with the covariances between the mean values from different generations (e.g., Equation 6.15), estimating response from differences in phenotypic means and estimating the realized heritability from a suitable (LS) regression of response on selection differential. As a result, LS analysis ignores much of the structure of the data, using only the records from a particular generation to compute the response and weighting all selected individuals equally. **Mixed-model (MM) analysis** (introduced in LW Chapters 26 and 27), on the other hand, consider *all* covariances between observations (both those between individuals within a generation and those between individuals from different generations). A MM analysis incorporates the fact that some families provide more information on response and weights individuals from these families accordingly, using records from the current and all previous generations to estimate the response. By virtue of using this additional information, a MM analysis is potentially far more powerful than a LS analysis (Sorensen and Kennedy 1983, 1984a, 1984b; Kennedy and Sorensen 1988). For example, under the infinitesimal model, incorporation of the full covariance structure by a mixed-model analysis completely accounts for any within- and between-line changes under genetic drift, as well as changes in σ_A^2 from gametic-phase disequilibrium due to selection and/or assortative mating.

This chapter concerns itself with the application of mixed model to the analysis of selection experiments. We remind the reader that the difficulty of topics reviewed here ranges from intermediate to rather advanced, and the material covered in Chapter 6 may be sufficient for many analyses, if for no other reason than many experiments do not record all the information required for a MM analysis

(i.e., records for all individuals over the course of the experiment and their pedigree structure). While the details (and extensions) of mixed model analysis takes up the bulk of this chapter, we conclude by introducing **Bayesian approaches** based on mixed-models. As opposed to the point estimators (means, variances) used by **classical statistics**, **Bayesian statistics** is concerned with generating the posterior distribution of the unknown parameters given both the data and some prior density for these parameters. As such, Bayesian statistics provides a much more complete picture of the uncertainty in the estimation of the unknown parameters, especially after the confounding effects of nuisance parameters are removed.

MIXED MODEL vs. LEAST SQUARES ANALYSIS

Figure 7.1 illustrates the result of a mixed-model analysis of selection response. Note that instead of measuring response from the observed phenotypic means (the LS approach), response is measured from the estimated mean breeding values. Further, instead of estimating a realized heritability, a mixed-model analysis estimates the additive genetic variance in the base population.

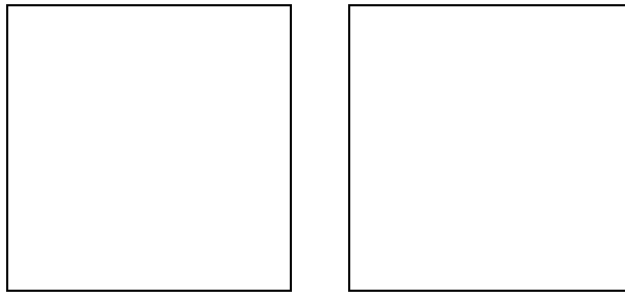


Figure 7.1. Results from high and low selection on 6-week weight in mice. **Left:** Observed mean phenotypic values in the up and down-selected lines. <M/H Fig 1.e>. **Right:** Estimated mean breeding values for both selected populations and the control (middle curve). <M/H Fig 2.3>. See Examples 2 and 4 for more detail on this experiment. After Meyer and Hill (1991).

Mixed-models readily allow the records to be adjusted for any number of **fixed effects**. For example, one might correct for character differences between sexes, between individuals from different size litters, age effects, known environmental factors, etc. Such adjustments for fixed-factors result in more accurate prediction of an individual's genetic value (and a more accurate estimate of the population's genetic response). One common fixed-factor correction in a LS analysis is the use of a control population to adjust the means from common envi-

ronmental trends that are unrelated to genetic changes (Chapter 6). While any number of fixed factors are readily incorporated into a mixed-model analysis, it can be more problematic to adjust for fixed-factors in a LS analysis. For example, a LS analysis cannot separate genetic from environmental trends when only a single line is considered. By contrast, under a properly-formatted MM analysis, one can separate phenotypic changes into genetic and environmental components without using a control population. As well will see, this is made possible by using the full covariance structure associated with the complete pedigree of all measured individuals in the experiment.

A final advantage of mixed model analysis is its great flexibility in handling any number of selection designs within the same framework. For example, a MM analysis allows for overlapping generations (e.g., when a parent contributes offspring over several different years of selection), while LS analysis of response in overlapping generations can be difficult to formulate correctly. Likewise, the mixed-model framework applies to mass (individual) selection, within-family selection, or any other index of selection based on additional information from (measured) relatives. On the other hand, a LS analysis requires that different analysis be formulated for each of these different selection schemes.

For all the power of mixed-model analysis, there are tradeoffs relative to the far simpler LS analysis. First, a MM analysis requires far greater record keeping than a LS analysis and is much more computationally demanding. Second, a MM analysis can be rather model-sensitive, in particular the infinitesimal model assumption is critical. If selection-induced changes in allele frequencies result in significant changes in the genetic variance during the course of the experiment, the assumptions of a MM analysis are violated. A LS analysis, on the other hand, make no such assumptions about the underlying genetics and is hence far more robust than a MM analysis.

BASICS OF MIXED-MODEL ANALYSIS

While we provide a brief overview of the analysis of mixed models, we strongly encourage the reader to review LW Chapters 26 and 27 before proceeding. These chapters provide many worked examples to give the reader a feel for mixed models, as well as considering advanced topics in MM analysis in far great detail than we do here.

Mixed models are so named because they consider both fixed and random effects. Recall that fixed effects are unknown constants while random effects are drawn from some underlying distribution (LW Chapters 8, 26). Hence, any particular value for a random effect represents just one possible realization from this underlying distribution. Typically, statisticians speak of *estimating* fixed effects and *predicting* the realized values of random effects. Both LS and MM analyses estimate the fixed effects in a model, while MM analysis also predicts the values of

the random effects, using the covariances between observations (after adjusting for fixed effects).

The standard mixed model for a vector \mathbf{y} of n observations is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (7.1)$$

where $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed effects, \mathbf{a} is a $p \times 1$ vector of random effects (in our case, these are the breeding values of the individuals in our experiment), \mathbf{e} the vector of residuals (which are also random effects), and \mathbf{X} and \mathbf{Z} are $n \times k$ and $n \times p$ **incidental matrices** associated with the fixed and random effects. If each individual is measured exactly once, then $\mathbf{Z} = \mathbf{I}_{n \times n}$, an identity matrix of dimension n . (Examples of the structure for \mathbf{X} and \mathbf{Z} for different models are given in LW Chapter 26.) In the absence of the vector of random effects \mathbf{a} , Equation 7.1 reduces to a least squares model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with ordinary (unweighted) least squares (OLS) used to estimate $\boldsymbol{\beta}$ if the residuals are uncorrelated and homoscedastic (having the same variances, giving the covariance matrix for the vector of residuals as $\text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}$). More generally, if the covariance structure of the residuals is more complex, $\text{Var}(\mathbf{e}) = \mathbf{V}$ where the only constraints on \mathbf{V} is that it is a symmetric and positive definite (Chapter 15), then general (weighted) least squares (GLS) is used (see LW Chapter 8).

In order to solve Equation 7.1, we need to specify the covariance structure for the vectors of random effects. It is generally assumed that the residuals are uncorrelated and homoscedastic, $\text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}$. The covariance of \mathbf{a} (in our case, the vector of breeding values) has a more complicated covariance structure governed by the structure of the pedigree, $\text{Var}(\mathbf{a}) = \sigma_A^2 \mathbf{A}$. Here \mathbf{A} is a matrix of known constants (the **numerator relationship matrix** whose elements are given by the pedigree structure). The resulting the $n \times n$ covariance matrix $\text{Var}(\mathbf{y}) = \mathbf{V}$ for the vector of observations \mathbf{y} becomes

$$\mathbf{V} = \sigma_A^2 \mathbf{Z}\mathbf{A}\mathbf{Z}^T + \sigma_e^2 \mathbf{I} \quad (7.2a)$$

The covariance matrix \mathbf{V} is a function of the variance components σ_A^2 and σ_e^2 . Since we will assume throughout that \mathbf{a} is a vector of breeding values, so that σ_A^2 is the additive genetic variance, we can alternatively express \mathbf{V} as a function of the heritability (h^2) and phenotypic variance (σ_z^2) of the trait of interest,

$$\mathbf{V} = \sigma_z^2 \left(h^2 \mathbf{Z}\mathbf{A}\mathbf{Z}^T + (1 - h^2) \mathbf{I} \right) \quad (7.2b)$$

as $\sigma_e^2 = \sigma_z^2 - \sigma_A^2 = \sigma_z^2(1 - h^2)$ when only additive variance is present. Estimation of the vector of fixed-effects $\boldsymbol{\beta}$ follows from GLS using the covariance matrix \mathbf{V} (LW Chapter 8),

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \quad (7.3a)$$

Equation 7.3a is called the **best linear unbiased estimator (BLUE)** of the vector of fixed effects. **Estimability** of the fixed effects can be an issue, as the structure of the data (indicated the column rank of \mathbf{X}) may not allow for unique estimates of all fixed effects. In such cases, generalized inverses can be used to obtain unique estimates of certain linear combinations of the fixed effects (LW Appendix 2). If \mathbf{X} has column rank $\ell \leq q$, then exactly ℓ combinations of fixed effects can be estimated, see LW Chapter 26 and LW Appendix 2 for further details. Finally, note that although the BLUEs are a function of \mathbf{V} (and hence σ_A^2 and σ_e^2), applying Equation 7.2b shows that the phenotypic variance σ_z^2 in \mathbf{V} cancels out in Equation 7.3a (\mathbf{V}^{-1} scales as $1/\sigma_z^2$, while the inverse of \mathbf{V}^{-1} scales as σ_z^2), leaving the BLUE estimate as a function of just the heritability.

The **best linear unbiased predictor (BLUP)** of the vector of random effects is given by

$$\hat{\mathbf{a}} = \text{Var}(\mathbf{a})\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sigma_A^2\mathbf{A}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (7.3b)$$

this is the regression of \mathbf{a} on \mathbf{y} , the vector of observations adjusted for fixed effects ($\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$) and then suitably scaled. Even if the number of random effects exceeds the number of actual observations (i.e., $p > n$), Equation 7.3b still provides unique estimates of each. This occurs because the covariance structure for the \mathbf{a} is incorporated in the model. As with the BLUEs, BLUPs are just functions of h^2 as $\sigma_A^2 = h^2\sigma_z^2$, while \mathbf{V}^{-1} scales as $1/\sigma_z^2$, leaving $\hat{\mathbf{a}}$ as only a function of the heritability (Equation 7.2b).

In practice, Equations 7.3a and 7.3b are often not used, as they require inversion of the potentially very large matrix \mathbf{V} . Instead, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{a}}$ are obtained without computing an inverse by numerically solving (for example, by Gaussian elimination) **Henderson's mixed model equations**

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix} \quad (7.4)$$

where $\lambda = \sigma_e^2/\sigma_A^2 = (1 - h^2)/h^2$. See LW Chapter 26 for more details.

The variance-covariance matrix for $\hat{\mathbf{a}}$ and $\hat{\boldsymbol{\beta}}$ also follow from the mixed-model equations. Denote the inverse of the matrix in Equation 7.4 by

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^T & \mathbf{C}_{22} \end{pmatrix} \quad (7.5a)$$

where \mathbf{C}_{11} , \mathbf{C}_{12} , and \mathbf{C}_{22} are, respectively, $k \times k$, $k \times p$, and $p \times p$ submatrices. Using this notation, Henderson (1975) showed that the covariance matrix for $\boldsymbol{\beta}$ is given by

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma_e^2\mathbf{C}_{11} \quad (7.5b)$$

while the covariance matrix of the prediction errors ($\hat{\mathbf{a}} - \mathbf{a}$) is

$$\mathbf{Var}(\hat{\mathbf{a}} - \mathbf{a}) = \sigma_e^2 \mathbf{C}_{22} \quad (7.5c)$$

and finally the covariances between estimated effects and prediction errors is

$$\boldsymbol{\sigma}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{a}} - \mathbf{a}) = \sigma_e^2 \mathbf{C}_{12} \quad (7.5d)$$

One concern is that the above BLUPs and BLUEs may be biased by selection. Henderson (1975) showed that these are unbiased if (i) selection decisions are based on linear combinations of data (such as truncation selection based on individual phenotypes or a linear index based on the phenotypes of an individual and its relatives) and (ii) that if selection is based on records that are adjusted for fixed effects, the model used estimates these fixed effects in an unbiased fashion when selection is absent. Hence, Henderson's conditions for BLUPs and BLUEs being unbiased by selection hold under many reasonable forms of artificial selection.

REML Estimation of Unknown Variance Components

The variance components (σ_A^2 and σ_e^2), or at a minimum the heritability $h^2 = \sigma_A^2 / (\sigma_A^2 + \sigma_e^2)$ must be specified to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{a}}$. These variances are generally unknown, but can be estimated using **restricted maximum likelihood (REML)**. REML is closely related to BLUP, with (roughly speaking) REML estimates obtained from iterating and updating BLUP estimates until suitable convergence. One advantage of REML estimates (over other variance estimation procedures) is that they are unbiased by the estimates of fixed effects, as the restricted likelihood refers to that part of the likelihood function unaffected by fixed effects (Patterson and Thompson 1971). We refer the reader to the extensive discussion of REML variance estimation in LW Chapter 27 for further details. For the remainder of this chapter, we assume we have already obtained REML estimates of the variances before proceeding the BLUP analysis.

The covariance matrix for the REML estimates can be approximated by using the best quadratic fit of the restricted likelihood surface, centered at the REML estimates (Smith and Graser 1986, Graser et al. 1987). If $\boldsymbol{\sigma} = (\sigma_A^2, \sigma_e^2)^T$ is a vector of assumed variances, one computes the restricted likelihood $L(\boldsymbol{\sigma})$ for a grid of values close to the REML solution and then fits the best quadratic surface to the data,

$$L(\boldsymbol{\sigma}) = \mathbf{b}_0 + \boldsymbol{\sigma}^T \mathbf{b}_1 + \boldsymbol{\sigma}^T \mathbf{S} \boldsymbol{\sigma} \quad (7.6a)$$

where the vectors \mathbf{b}_i and the symmetric matrix \mathbf{S} are fitted using the data. (Chapter 16 discusses fitting the best quadratic surfaces in the context of fitness surface estimation.) The approximate covariance matrix for the vector of REMLs, $\hat{\boldsymbol{\sigma}}$, is given by

$$\mathbf{Var}(\hat{\boldsymbol{\sigma}}) \simeq (-2\mathbf{S})^{-1} \quad (7.6b)$$

The rationale for this approach is that the inverse of the matrix of second-order partial derivatives of the likelihood surface at the likelihood estimate approaches (the large-sample) covariance matrix of these estimates (LW Appendix 4). Equation 7.6a is a (second-order) multidimensional Taylor series, with $2S$ corresponding to the matrix of second-order partial derivatives of the likelihood function (Chapter 15).

As with BLUPs and BLUEs, REML estimates of variance are often unbiased by selection. In particular, if the base population consists of unselected and noninbred individuals and phenotypic data are available for all selected and unselected individuals, then under the infinitesimal model REML yields essentially unbiased estimates of the additive genetic variance in the base population (Henderson 1949, Henderson et al. 1959, Curnow 1961, Thompson 1973, Rothschild et al. 1979, Sorensen and Kennedy 1984b, Gianola and Fernando 1986, Gianola et al. 1988, Fernando and Gianola 1990). On the other hand, van der Werf and colleagues (van der Werf 1990, van der Werf and de Boer 1990, van der Werf and Thompson 1992) show that when the base population consists of previously selected individuals, REML provides no protection from biased estimates of the additive genetic variance in the population prior to selection, even if the entire pedigree of individuals back to the base population is included. Likewise, if selection acts on a suite of unmeasured characters that are correlated with characters included in the model, REML can generate biased estimates of the variances and covariances of the measured characters (Schaeffer and Song 1978).

ANIMAL-MODEL ANALYSIS OF SELECTION EXPERIMENTS

The basic building block of mixed-model analysis of selection experiments is the **animal model**, which estimates the breeding (or additive genetic) values of all individuals measuring during the course of experiment (LW Chapter 26). We examine the simplest version of the animal model first, considering various elaborations in later sections. While the animal model has its origin in the animal breeding literature, it has very widespread applicability. We trust that plant scientists will not be not greatly offended, as the “animal” model can be used to analyze plant selection experiments as well.

To apply the animal model to selection experiments, first vectorize the observations from the entire experiment by letting y_{ij} denote the j measured individual from generation i , where $0 \leq i \leq t$ (generation 0 representing the unselected base population) and $1 \leq j \leq n_i$. Let the vector y denote all measured individuals

from the entire experiment,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_t \end{pmatrix}, \quad \text{where} \quad \mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \vdots \\ \mathbf{y}_{in_i} \end{pmatrix}$$

The vector \mathbf{y}_i includes the values for all measured individuals from generation i , including those culled as well as those allowed to reproduce. The simplest animal model for these data is

$$y_{ij} = \mu + a_{ij} + e_{ij} \quad (7.7a)$$

where μ is an overall mean, a_{ij} the breeding value of the j th measured individual from generation i , and e_{ij} the deviation between breeding and phenotypic values. With exactly one record per individual, $\mathbf{Z} = \mathbf{I}$. In this simple model the only fixed effect is the mean, giving $\boldsymbol{\beta} = (\mu)$ and $\mathbf{X} = \mathbf{1}$ (a vector of ones), reducing Equation 7.1 to

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{a} + \mathbf{e} \quad (7.7b)$$

Here \mathbf{a} is the vector of breeding values for all individuals measured during the course of the experiment, with $\text{Var}(\mathbf{a}) = \sigma_A^2 \mathbf{A}$. The relationship matrix \mathbf{A} is the key to mixed-model analysis, as it includes all the pedigree information. The diagonal elements of \mathbf{A} describe the amount of inbreeding, with $A_{ii} = (1 + f_i)$, while the off-diagonal elements $A_{ij} = 2\Theta_{ij}$ (twice the coefficient of coancestry, see LW Chapters 7, 26) describe the relatedness of individuals i and j . Recursive methods for obtain the elements of \mathbf{A} are discussed in LW Chapter 26. The simple animal model assumes that all genetic variance is additive, so that there is no (genetic) covariance between residuals. In this case, it is generally assumed that $\text{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$.

For Equation 7.7b, the mixed-model equations (7.4) simplify to

$$\begin{pmatrix} n & \mathbf{1}^T \\ \mathbf{1} & \mathbf{I} + \lambda \mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} n \bar{y} \\ \mathbf{y} \end{pmatrix} \quad (7.7c)$$

where n is the total number of individuals in the experiment, $\lambda = \sigma_e^2 / \sigma_A^2 = (1 - h^2) / h^2$, $\hat{\mathbf{a}}$ is an n -dimensional vector of the predicted breeding values of all measured individuals, and $\mathbf{1}$ is a vector of ones.

Under a mixed-model analysis, response is measured by the change in the mean breeding value of a selected population over time. The estimated mean breeding value in generation k is simply given by

$$\hat{\mathbf{a}}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \hat{a}_{kj} \quad (7.8a)$$

Total response at generation t is estimated by $\bar{a}_k - \bar{a}_0 = \bar{a}_k$, as the predicted mean breeding value from generation 0 (the unselected base population) is zero by construction. In matrix notation, the vector $\bar{\mathbf{a}}$ of mean breeding values is estimated by

$$\hat{\bar{\mathbf{a}}} = \mathbf{K}^T \hat{\mathbf{a}} \tag{7.8b}$$

where the i th row of the matrix \mathbf{K} consists of $1/n_j$ when the column corresponds to an individual from generation j , otherwise all elements in that row are zero. Thus, for t generations of data (corresponding to $t - 1$ generations of selection, as the analysis includes the unselected base population, generation 0), \mathbf{K} is $n \times t$, with $\mathbf{K}^T \mathbf{1}_n = \mathbf{1}_t$ (a $t \times 1$ vector of ones).

From Equation 7.5c, and recalling that $\text{Var}(\mathbf{Bx}) = \mathbf{B} \text{Var}(\mathbf{x}) \mathbf{B}^T$ (LW Equation 8.21b), the covariance matrix for the vector of genotypic means becomes

$$\text{Var}(\hat{\bar{\mathbf{a}}}) = \sigma_e^2 \mathbf{K}^T \mathbf{C}_{22} \mathbf{K} \tag{7.8c}$$

where the $n \times n$ matrix \mathbf{C}_{22} is the solution to Equation 7.5a.

Example 1. Suppose in the base population (unrelated and non-inbred) individuals 1-4 have trait values 3, 6, 5, and 2, respectively. The two largest individuals are allowed to reproduce and their resulting offspring (individuals 5-8) have values 4, 5, 6, 5. Assuming the only fixed effect is the mean, the resulting animal model is $\mathbf{y} = \mathbf{1}\beta + \mathbf{a} + \mathbf{e}$, where

$$\mathbf{y} = \begin{pmatrix} 3 \\ 6 \\ 5 \\ 2 \\ 4 \\ 5 \\ 6 \\ 5 \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix}, \quad \beta = (\mu)$$

What is the relationship matrix \mathbf{A} ? Since individuals 2 and 3 are the parents, and all offspring are full-sibs, all related individuals have values of $1/2$ as $2\theta_{ij} = 1/2$ for both parent-offspring and full-sibs. The resulting numerator relationship matrix becomes

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 0 & 1 & 0 & 1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 1 & 1/2 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 1/2 & 1 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 1/2 & 1/2 & 1 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 1/2 & 1/2 & 1/2 & 1 \end{pmatrix}$$

Suppose the heritability of the trait is $h^2 = 0.3$. Applying 7.3a gives

$$\hat{\mu} = (\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{V}^{-1} \mathbf{y} = 4.22$$

where we have computed \mathbf{V} using Equation 7.2b scaled to remove the phenotypic variance σ_z^2 . Substituting into 7.3b gives the BLUP estimate of the genetic values as

$$\hat{\mathbf{a}} = \begin{pmatrix} -0.366 \\ 0.666 \\ 0.366 \\ -0.666 \\ 0.386 \\ 0.562 \\ 0.739 \\ 0.562 \end{pmatrix}. \quad \text{Here } \mathbf{K} = \frac{1}{4} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{implying } \mathbf{K}^T \hat{\mathbf{a}} = \begin{pmatrix} 0 \\ 0.562 \end{pmatrix}$$

Hence, the estimated response (for $h^2 = 0.3$) is 0.562. Under a least squares analysis, $\bar{y}_0 = 4$ and $\bar{y}_1 = 5$ for an estimated response of 1. The estimated response for different assumed heritabilities are as follows:

h^2	Estimated response	h^2	Estimated response
0.0	0	0.6	0.940
0.1	0.211	0.7	1.026
0.2	0.398	0.8	1.083
0.4	0.707	0.9	1.095
0.5	0.833	1.0	1

Note that a REML/BLUP analysis of a selection experiment has a very different character from a LS analysis. In the latter, one estimates the realized heritability from a suitable regression of phenotypic means on selection differentials. With a strictly BLUP analysis, one starts with an assumed base-population heritability and then computes a genetic trend by plotting the mean estimated breeding values (Figure 7.1). Response is underestimated if the assumed heritability is less than the true value, while it is overestimated if the heritability is overestimated (Sorensen and Kennedy 1984a). As Figure 7.1 shows, the BLUP response curves of mean breeding values are much smoother than the curves for phenotypic means. This is because BLUP compares an individual's estimated breeding value with an index based on information from its relatives. Individual breeding values are regressed towards the value predicted by the index, smoothing out excessive fluctuations (Sorensen and Kennedy 1986). In a REML/BLUP analysis, one first estimates the additive genetic variance in the base population (via REML), using this value in the subsequent BLUP analysis. Kackar and Harville (1981) and Gianola et al. (1986) show that using the REML estimates does not result in biased

values for BLUPs, but that the resulting predictors may not be “best” (there may be other linear predictors with smaller errors).

One caveat that must be stressed is that the estimated mean breeding values should not themselves be used to estimate heritabilities. For example, Blair and Pollak (1984) regressed the BLUP-estimated mean breeding values on cumulative selection differentials to obtain a realized heritability estimate. The problem with this approach is that the heritability (or additive variance) used in the BLUP analysis very strongly influences the results. Hence, realized heritability estimates obtained from regressions based on BLUP estimates of mean breeding values depend on the assumed heritability, not the actual population heritability (Thompson 1986). The correct estimate of heritability in a mixed-model analysis should be based on REML (or other) estimates of the base-population variance components.

The Relationship Matrix Accounts for Drift and Disequilibrium

As was reviewed in Chapter 5, selection changes the additive variance by generating gametic-phase disequilibrium even in the absence of allele frequency changes. Additionally, with a finite number of loci selection also changes allele frequencies, further changing the genetic variances. While gametic-phase disequilibrium changes in the genetic variances are generally restricted to the first few generations of selection (Chapter 5), allele frequency changes become increasingly more important as selection proceeds (Chapters 10, 11). Thus, the additive variance in a particular generation after selection is likely different from the variance in the unselected base population. A least-squares analysis does not account for these changes, but rather assumes that the realized heritability is the same in each generation. Given that the reduction in h^2 from disequilibrium reaches its equilibrium value in only a few generations of directional selection, the LS assumption of a constant h^2 may not induce too serious of an error, *provided* the reduction is corrected for.

Under the infinitesimal model (in particular, assuming no significant selection-induced changes in allele frequencies), a mixed-model analysis fully accounts for the effects of gametic-phase disequilibrium as well as genetic drift (Sorensen and Kennedy 1983, 1984a). This occurs because under the infinitesimal model, even in the face of selection and drift, the variance-covariance matrix of the vector of breeding values remains the product of the base-population additive genetic variance and the numerator relationship matrix, $\text{Var}(\mathbf{a}) = \sigma_A^2 \mathbf{A}$.

This independence of the covariance relationships of \mathbf{a} from selection and drift (given \mathbf{A}) follows as a consequence of the behavior of the residual in regression of the breeding value of an individual A_i on the breeding values of its sire (A_{m_i}) and dam (A_{f_i}). Recall from Equation 5.8a that

$$A_i = \frac{1}{2}A_{f_i} + \frac{1}{2}A_{m_i} + s_i \tag{7.9}$$

where the **segregation residual** s (also referred to as **Mendelian sampling**) re-

sults from segregation of alleles at heterozygous loci in the parents. Under the infinitesimal model, s is independent of parental breeding values and has mean zero and variance $(1 - \bar{f}_i) \sigma_A^2/2$. Here \bar{f}_i is the mean inbreeding of the parents of individual i and σ_A^2 the base-population (before selection) additive variance. More generally, provided the vector \mathbf{s} of Mendelian sampling residuals remains multivariate normal, then $\mathbf{s} \sim \text{MVN}(\mathbf{0}, (\sigma_A^2/2) \mathbf{F})$. The matrix \mathbf{F} a diagonal with i th element $(1 - \bar{f}_i)$, one minus the average inbreeding in the parents. If k and j are the parents of offspring i , then

$$F_{ii} = (1 - \bar{f}_i) = \left(1 - \frac{f_k + f_j}{2}\right) = \left(2 - \frac{A_{kk} + A_{jj}}{2}\right) \quad (7.10)$$

where $A_{kk} = (1 + f_k)$ denotes the k th diagonal element of the relationship matrix \mathbf{A} . The effects of drift on the additive variance enter through the inbreeding coefficients f . Thus, the distribution of the Mendelian sampling terms \mathbf{s} is unaffected by the breeding values of the parents (and hence by selection and assortative mating). When we have the complete pedigree of all individuals in the selection experiment, along with all their records, we can express any breeding value as a linear function of the base population breeding values (the coefficients following from \mathbf{A}) and Mendelian sampling terms (\mathbf{s}) not affected by selection. In particular, we can express \mathbf{a} as a linear function of the Mendelian sampling terms, $\mathbf{a} = \mathbf{T}\mathbf{s}$. The resulting covariance matrix is $\text{Var}(\mathbf{a}) = \mathbf{T} \text{Var}(\mathbf{s}) \mathbf{T}^T$, where $\text{Var}(\mathbf{s}) = (\sigma_A^2/2) \mathbf{F}$ is independent of selection and assortative mating, while \mathbf{F} accounts for the reduction in additive variance from genetic drift.

To show this, we follow Sorensen and Kennedy (1984a). Ordering individuals so that parents proceed their offspring, Henderson (1976) and Thompson (1977) show that \mathbf{A} can be written as a function of a diagonal matrix, \mathbf{D} , and an upper triangular matrix \mathbf{T} ,

$$\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}^T \quad (7.11)$$

\mathbf{T} traces the passage of genes from one generation to the next, while \mathbf{D} is the variance in offspring breeding value, conditioned on the parental breeding values (the Mendelian sampling variance). To see this last point, consider the transformation $\mathbf{g} = \mathbf{T}^{-1}\mathbf{a}$ of the vector of breeding values. From Equation 7.11, the covariance matrix for \mathbf{g} becomes

$$\text{Var}(\mathbf{g}) = \mathbf{T}^{-1} \text{Var}(\mathbf{a}) (\mathbf{T}^T)^{-1} = \sigma_A^2 \mathbf{T}^{-1} \mathbf{T} \mathbf{D} \mathbf{T}^T (\mathbf{T}^T)^{-1} = \sigma_A^2 \mathbf{D} \quad (7.12)$$

which follows using the general matrix relationship $(\mathbf{B}^T)^{-1} = (\mathbf{B}^{-1})^T$ (the inverse of transpose equals the transpose of the inverse). Sorensen and Kennedy show that the i th element of \mathbf{g} equals

$$g_i = A_i - \frac{1}{2}A_{f_i} - \frac{1}{2}A_{m_i} = s_i \quad (7.13)$$

which is simply the segregation residual. Hence, $\mathbf{g} = \mathbf{s}$, and since $\text{Var}(\mathbf{s}) = (\sigma_A^2/2)\mathbf{F}$, implies $\mathbf{D} = \mathbf{F}/2$. Writing $\mathbf{a} = \mathbf{T}\mathbf{T}^{-1}\mathbf{a} = \mathbf{T}\mathbf{g} \equiv \mathbf{T}\mathbf{s}$, shows that the vector of breeding values \mathbf{a} is a simple vector transformation of the vector of Mendelian sampling residuals \mathbf{s} .

Thus, the variance of the vector of breeding values is a linear transformation of the variance of the Mendelian sampling residuals. Under the infinitesimal model, the distribution of these residuals is unaffected by selection. However, if the infinitesimal model does not hold, then residual values may indeed vary with parental breeding values, in which case selection can certainly influence the distribution of residuals. If, however, the change in allele frequencies is small over the course of the experiment, the bias may not be too serious. Likewise, another key element is that the distribution of residuals does not significantly deviate from normality. Chapter 11 examines this very technical issue in some detail.

Model Validation

Given the sensitivity of a mixed-model analysis to the validity of the assumptions (in particular, the infinitesimal model), some form of model validation is required to apply MM methods with confidence. One approach is to test the infinitesimal model prediction that estimates of the base population σ_A^2 should remain stable as additional generations of selection are considered. If the infinitesimal model holds, \mathbf{A} completely accounts for changes in the additive variance in these later generations of selection. On the other hand, if σ_A^2 is changing in ways not predictable from the infinitesimal model, using data from additional generations of selection may result in dramatically different estimates of the base-population additive variance. Likewise, if the same base population is used to form both the control and selected lines, the estimated the base population additive variance has the same expected value in both lines. Departures from either of these two predictions indicates failure of the model.

Example 2. One of the first REML/BLUP analyses of a selection experiment was by Meyer and Hill (1991), who examined the response to selection for adjusted food intake (AFI) in mice. AFI is defined as food intake between 4 and 6 weeks corrected for 4-week weight. Meyer and Hill had three replicates, each consisting of a high, low, and control lines, for a total of almost 11,000 mice over the course of the experiment. Within-family selection on AFI was followed for 23 generations. Meyer and Hill included a number of fixed effects in their model, as well as adding a random effect to control for common litter effects (see Example 4 for details).

As a check of the validity of the assumptions (in particular, the infinitesimal model), Meyer and Hill compared variance estimates based on data from generations 5-7 with estimates based on generations 14-23. In both cases, the full pedigree structure was incorporated into the numerator relationship matrix (i.e., the relationships among individuals in generation 5 were used to generate the

submatrix of \mathbf{A} associated with this generation, and similarly for future generations). While incorporation of the complete pedigree information reduces the bias in estimates of the base-population additive variance, it does not completely reduce the bias if the records for all individuals from previous generations (back to the base population) are ignored (van der Werf and de Boer 1990). Even with this caveat in mind, Meyer and Hill observed a dramatic decline in the estimated additive variances (from 7.2 based on generations 5-7 to 2.5 based on generations 14-23). Under the infinitesimal model, both estimates should be for the base population variance. This large decrease suggested that the infinitesimal model may not be appropriate for this trait. It is interesting to note that this decrease occurred even as the total variance increased dramatically (from 23.88 to 33.93). This increase resulted mainly from an increase in the environmental variance (from 12.9 to 25.5), although there was also a slight increase in the litter-effects variance (from 4.78 to 5.96).

Several other REML/BLUP analyses of selection experiments in mice also found differences in estimates of base population additive variance when comparing data from early generations versus data from later generations. Beniwal et al. (1992a,b) observed decreases in the additive variance (in body weight, litter size, and lean mass), while Heath et al. (1995) observed an increase in the additive variance in body weight.

MODIFICATIONS OF THE BASIC ANIMAL MODEL

In many cases, it is prudent to modify the simple animal model by considering additional fixed and random effects. For example, genetic and environmental effects can be separated without a control population by adding fixed effects to account for environmental trends. Likewise, it is often reasonable to include additional random effects, such as material/litter effects. Another modification of the basic model occurs when the phenotypic scores (records) of parents are unknown, in which case it may be reasonable to treat the unknown parental breeding values as fixed effects. We examine all these modifications below. We conclude this section by examining how to estimate the additive genetic variance present in any particular generation and how to extend the analysis to models with significant non-additive genetic variance.

Separating Genetic and Environmental Trends

The mean phenotype can be altered by both genetic and environmental changes. In a least-squares analysis, any underlying environmental trend is assumed to be removed by contrasting selected and control populations (or contrasting populations selected in opposite directions). The rationale is that the k th individual from

population j in generation t can be written as

$$y_{tjk} = \mu + d_t + a_{tjk} + e_{tjk} \tag{7.14}$$

where d_t is the environmental trend. If the common environmental value is the same in both selected and control populations, then the difference in phenotypic means in generation t is

$$\bar{z}_{s,t} - \bar{z}_{c,t} = (\bar{a}_{s,t} - \bar{a}_{c,t}) + (e_{s,t} - e_{c,t}) \tag{7.15}$$

The residuals e have expected value zero, hence the contrast provides an estimate of $\bar{a}_{s,t}$, provided there is no significant drift in the mean breeding value of the control population (so that $\bar{a}_{c,t} \simeq 0$). However, if genotype-environment interactions are present, so that the environment value for generation t is different in both populations, then Equation 7.14 has an additional term $(d_{s,t} - d_{c,t})$. Hence, even when a control population is used, a least-squares analysis can still give biased results if there is significant drift in the mean of the control population and/or significant $G \times E$.

A mixed-model analysis estimates the mean *breeding value*, rather than the *phenotypic mean*, of the population. Hence a MM analysis allows for the separation of the genetic change from any environmental change (Blair and Pollak 1984). This occurs because \mathbf{A} tracks the flow of genes through the population, allowing for estimates of breeding values independent of environmental effects. Of course, this is strictly dependent on the model assumptions holding, but if they do, a mixed-model analysis does not require a control population. Common environment effects are incorporated into the basic animal model by simply adding a fixed effect d_i for common environmental effect in generation i ,

$$y_{ij} = \mu + d_i + a_{ij} + e_{ij} \tag{7.16}$$

The vector of fixed effects now becomes

$$\boldsymbol{\beta} = (\mu, d_1, d_2, \dots, d_T)^T$$

and the corresponding incident (or design) matrix \mathbf{X} has ones in the columns corresponding to the generation in which the individual was scored, viz.,

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix}$$

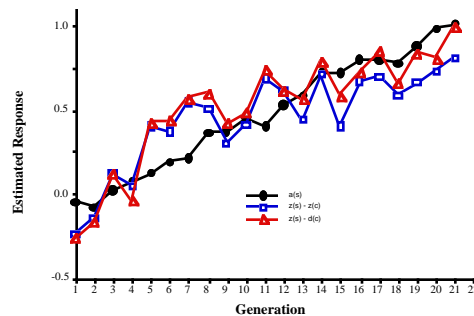
Example 3. To examine the potential bias from $G \times E$ and drift in the control population, Blair and Pollak (1984) examined a seven-generation selection experiment on 14-month greasy fleece weight in sheep. The model they assumed was that the m th individual in generation t with fixed sex effect (male/female) i , fixed age of dam effect (mature/immature) j , and fixed rearing rank effect (single/twin) k had a phenotypic value y of

$$y_{tmijk} = sx_i + b_j + r_k + d_t + a_{tm} + e_{tmijk}$$

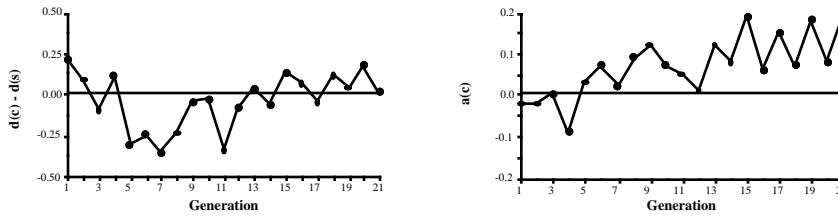
In matrix form, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$, where the vector of fixed effects $\boldsymbol{\beta}$ contains the fixed-effects for sex (sx), dam age (b), and rearing rank (r) in addition to along with the fixed effects for years (d). Both the selected and control line were subjected to a BLUP analysis using this model of fixed effects, and three different estimates of selection response were considered.

- (i) $\widehat{z}_{ts} - \widehat{z}_{tc}$, the estimated phenotypic means following adjustment for the fixed-effects (sx , b , and r), obtained by $\widehat{z}_t = \widehat{a}_t + \widehat{d}_t$. This is an unbiased estimate of the response if there is no significant drift in the control population ($\bar{a}_{ct} \simeq 0$) and no $G \times E$, so that $d_{ct} = d_{st}$.
- (ii) $\widehat{z}_{st} - \widehat{d}_{ct}$, the (fixed-effects adjusted) phenotypic mean of the selected population minus the common environmental effect, as estimated from the control population.
- (iii) \widehat{a}_{st} , the BLUP estimate of the mean breeding value in the selected population.

Estimate (i) mimics that used in a least-squares analysis, and Blair and Pollak show it is independent of the assumed heritability. Estimates (ii) and (iii) are highly dependent on the assumed (or estimated) heritabilities in the control and selected populations. As the figure below shows, all three estimates show a positive genetic trend (following a reversed response over the first few generations). The estimated response using only the predicted mean breeding value is smoother (as expected) than the other two estimates.



The potential biases in a least-squares analysis of this data using the contrast between the control and selected phenotypic means are seen in the two graphs below. The left graph plots the difference in the estimated common environmental effects ($\hat{d}_{ct} - \hat{d}_{st}$). Ignoring the inherent variance in estimating the \hat{d} 's, the difference (via a paired t -test) is not significantly different from zero. The right graph plots the predicted mean breeding value of the control population, which is assumed to be zero under the least-squares analysis. As the figure shows, there is a slight, but positive, trend. When the control mean is subtracted off, the net result is that the LS analysis slightly underestimates the true response. Thus, there is no evidence of error being introduced by $G \times E$ differences between the control and selected lines, but error is introduced by the mean breeding value of the control population departing significantly from zero.



As illustrated in Example 1, the estimated environmental (d) and additive-genetic (a) effects are highly-dependent on the assumed (or estimated) base-population heritability h^2 . Hence, using BLUP to separate genetic from environmental values is highly dependent on the heritability used being close to the true value. Further, the complete relationship matrix and corresponding measures for all individuals must be known. Finally, any other departures from mixed-model assumptions (infinitesimal model, and that the BLUP and REML estimates are not affected by selection) can also result an incorrect assignment of the relative importance of environmental versus genetic values. The infinitesimal model assumption, namely that there are no significant selection-induced changes in the allele frequencies, it likely to break down as the selection experiment continues.

Thus, one must be wary of relying solely on a BLUP analysis to separate genetic from environmental effects, and it is best to have some sort of control population(s) where possible. However, a BLUP analysis on the control population can estimate the amount of drift in the mean breeding value from its expected value of zero. Separate BLUP analyses of selected and control populations can provide initial estimates of potential environmental values. If these are reasonably consistent, then a joint analysis (assuming the same environmental values in both populations) may yield more precise estimates of the generational environmental values. Likewise, if the estimates are significantly different, the possibility of either $G \times E$ and/or different environmental values in selected versus controlled lines needs to be seriously considered. Thus, even with a control population, there

is still much to be gained by subjecting each to a BLUP analysis.

It is straightforward (although potentially quite computationally demanding) to jointly analyze multiple lines simultaneously. For k lines, write the total vector of observations as $\mathbf{y}^T = (\mathbf{y}_{i1}^T, \mathbf{y}_{i2}^T, \dots, \mathbf{y}_{ik}^T)$ where \mathbf{y}_{li} is the vector of total observations from line i . If the generational environmental effects are assumed to be the same in each line, the model for the i th individual in generation t from line k is

$$y_{kti} = \mu + d_t + a_{kti} + e_{kti} \quad (7.17a)$$

Alternatively, if the environmental effects are potentially different in each population, then

$$y_{kti} = \mu + d_{kt} + a_{kti} + e_{kti} \quad (7.17b)$$

The power of combining multiple lines arises when the effects are assumed to be the same in multiple lines, in which case the effective sample size for estimating each effect is increased, and (presumably) the resulting sampling variance decreased, improving the precision of the estimates. The assumed covariance matrix for the vector of joint breeding values can also take several forms. If the founding members for each line are drawn from the same base population (but otherwise unrelated), then the covariance matrix for \mathbf{a} has block-diagonal form, with the i th block corresponding to $\sigma_A^2 \mathbf{A}(i)$, the numerator relationship matrix for line i times the base-population additive variance. If the founding members of at least some different lines are related, then \mathbf{A} is more complex, reflecting these relationships. Further modifications for joint analysis have been proposed by Visscher and Thompson (1990), and extended by Beniwal et al. (1992a,b) and Heath et al. (1995) by allowing the additive variance to change. For example, one might assume that the additive variance remains constant in the control population and for the first few generations of the selected populations, after which it assumes a different value in the selected populations. This is a logical, but still ad-hoc, approach towards dealing with potential departures from the infinitesimal model.

Models with Additional Random Effects

We have assumed that the residuals are uncorrelated and homoscedastic, giving their covariance matrix as $\text{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$. When additional random effects are present, but ignored by the model, they are subsumed into the residuals, potentially introducing correlations and heteroscedasticity. For example, if sibs share a common maternal environment, this introduces correlations between sibs beyond those accounted for by \mathbf{A} . If the model only includes \mathbf{a} and \mathbf{e} , this additional covariance appears between the residuals and the covariance matrix for \mathbf{e} is no longer diagonal, leading to biased estimates of the BLUEs and BLUPs. By suitably incorporating additional random effects, we recover a model where the residuals again have the simple covariance structure, $\text{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$.

Suppose there is a second vector \mathbf{u} of m random effects in addition to the

vector \mathbf{a} of p breeding values and vector of residuals \mathbf{e} . Equation 7.1 becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{u} + \mathbf{e} \quad (7.18a)$$

where \mathbf{X} , \mathbf{Z} , and \mathbf{W} are $n \times q$, $n \times p$ and $n \times m$ incident matrices. The covariance structure assumed is $\text{Var}(\mathbf{a}) = \sigma_A^2 \mathbf{A}$, $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}$, and $\text{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$, giving the covariance matrix for \mathbf{y} as

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{A}\mathbf{Z}^T \sigma_A^2 + \mathbf{W}\mathbf{W}^T \sigma_u^2 + \mathbf{I} \sigma_e^2 \quad (7.18b)$$

If we had incorrectly assumed the true model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$, the covariance matrix for the residuals becomes $\text{Var}(\mathbf{e}) = \mathbf{W}\mathbf{W}^T \sigma_u^2 + \mathbf{I} \sigma_e^2$, showing how the additional random effects alters the covariance matrix. The resulting mixed-model equations for 7.18a become

$$\begin{pmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} & \mathbf{X}^T \mathbf{W} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \lambda_A \mathbf{A}^{-1} & \mathbf{Z}^T \mathbf{W} \\ \mathbf{W}^T \mathbf{X} & \mathbf{W}^T \mathbf{Z} & \mathbf{W}^T \mathbf{W} + \lambda_u \mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{c}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \\ \mathbf{W}^T \mathbf{y} \end{pmatrix} \quad (7.19a)$$

where

$$\lambda_A = \frac{\sigma_e^2}{\sigma_A^2} \quad \text{and} \quad \lambda_u = \frac{\sigma_e^2}{\sigma_u^2} \quad (7.19b)$$

Additional vectors of random effects can be incorporated in a similar manner, see LW Chapters 26 and 27 for details. The mixed-model equation again form the basis for iterative REML estimates of the unknown variance components (σ_A^2 , σ_u^2 , and σ_e^2), as discussed in detail in LW Chapter 27.

The following examples illustrate two common models with additional random effects — maternal effects and repeated measures. A third example, considered in a later section, is the addition of random effects to account for nonadditive genetic effects such as dominance.

Example 4. A common source of variation is shared family effects between sibs from the same litter (or, more generally, different litters from the same mother). Meyer and Hill (1991), examining the response to selection on adjusted food intake (AFI) in mice (Example 2), formulate a model incorporating shared family values c as random effects. In addition, their model accounts for fixed-effects due to generations (d , 22 levels), lines (ln , 3 levels), sex (sx , male/female), and litter size (lt , 7 levels for litters of size 6 to 12 individuals). Under their model, the observed value for AFI from the k th individual from generation t , line ℓ and full-sib family i

$$y_{tlik} = d_t + ln_\ell + sx_j + lt_m + a_{tlik} + c_{tli} + e_{tlik}$$

where this individual has sex j and litter size m . In matrix form, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{c} + \mathbf{e}$. The vector of fixed effects $\boldsymbol{\beta}$ contains the d , ln , sx , and lt values, while the random effects are the vector of common family effects \mathbf{c} , the vector of additive genetic values \mathbf{a} , and the vector of residuals \mathbf{e} . The incident matrix \mathbf{W} has as its ij -th element 1 if individual i ($1 \leq i \leq n$) is from family j ($1 \leq j \leq n_c$), else the element is zero. Note that Meyer and Hill have two model variables to account for litter effects — a fixed effect lt common to all litters of the same size and a random effect $c_{t\ell i}$ that varies between families but is the same for all individuals from a particular family. The resulting REML estimate for heritability was 0.15, while the fraction c^2 of the total variation (after removal of the fixed effects) accounted for by random family effects was estimated to be $c^2 = \sigma_c^2 / \sigma_z^2 = 0.22$ ($\sigma_z^2 = \sigma_A^2 + \sigma_c^2 + \sigma_e^2$ is the trait variance following the removal of fixed effects). Hence, a larger fraction of the resemblance between sibs (the intraclass correlation, $t = h^2/2 + c^2$) is due to shared family environments, rather than shared genes. One caveat is that the model assumes no additive genetic variation. If present, sibs share dominance variance ($\sigma_D^2/4$), and under this model this is incorporated into σ_c^2 .

Example 5. Often the same trait is measured multiple times in the same individual, for example, the sizes of different litters from a female. When multiple records are present for at least some individuals, a **repeatability model** should be used (LW Chapter 26). Repeated measures from the same individual have three components: a genetic value a_k , a common (permanent) environmental value p_k that is the same in each measurement, and the residual environmental value e varying between each measurement, giving the i th measurement of k th individual as $a_k + p_k + e_{ki}$. The **repeatability** r of the trait is given by $r = (\sigma_A^2 + \sigma_p^2) / \sigma_z^2$, giving the variance of the residuals as $\sigma_e^2 = (1 - r)\sigma_z^2$. Likewise, it follows that the variance of permanent environmental effects is $\sigma_p^2 = (r - h^2)\sigma_z^2$.

The repeatability model was used by Estany et al. (1989) to examine the selection response on litter size in rabbits. Their model assumed two groups of fixed effects, d_t the year-season (environmental) effect which had 22 levels in this experiment and the reproductive state l_i of the doe (l has three levels: l_1 for primiparous does, l_2 for lactating does, and l_3 for non-primiparous and non-lactating does). Since only two of these factors are estimable, l_1 was assigned a value zero. Their model had three random effects, a_k and p_k for the additive genetic and permanent non-genetic effect of the k th doe, and the residual e , giving the overall model as

$$y_{tk\ell i} = l_i + d_t + a_k + p_k + e_{tk\ell i}$$

where $y_{tk\ell i}$ denotes the litter size for the ℓ th litter of doe k in reproductive state i in season-year t . In matrix form, the mixed-model becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{Zp} + \mathbf{e}$$

where \mathbf{a} and \mathbf{p} are $n \times 1$ vectors corresponding to the n does, $\mathbf{Var}(\mathbf{a}) = \sigma_A^2 \mathbf{A}$, $\mathbf{Var}(\mathbf{p}) = \sigma_p^2 \mathbf{I}$, and $\mathbf{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$. \mathbf{X} and \mathbf{Z} are incident matrices, and the vector of fixed effects is

$$\boldsymbol{\beta} = \begin{pmatrix} l_1 \\ l_2 \\ d_1 \\ d_2 \\ \vdots \\ d_{22} \end{pmatrix}$$

The mixed-model equations are given by Equation 7.19a with

$$\lambda_A = \frac{\sigma_e^2}{\sigma_A^2} = \frac{1-r}{h^2} \quad \text{and} \quad \lambda_u = \frac{\sigma_e^2}{\sigma_p^2} = \frac{1-r}{r-h^2}$$

Treating Certain Breeding Values as Fixed effects

How should one proceed if the base-population has itself been under selection? If this is known or suspected to be the case, Graser et al. (1987) suggest that the base population breeding values be treated as fixed, rather than random, effects. The motivation for this suggestion is that if parents are selected, they are not a random sample from the base population. Since REML estimates are unbiased by fixed-effects, any bias in the variance of the initial sample is ignored by treating the original parental breeding values as fixed. Simulation studies, however, show that even bias is reduced by treating the parents as fixed, selection on the resulting offspring (or future generations) introduces additional bias (van der Werf 1992). Despite this reservation, we briefly review the approach here as parents whose records are missing are also often treated as fixed, and treating these breeding values as fixed requires some modifications of the mixed-model equations. Let \mathbf{a}_b be the vector of breeding values for the base population and \mathbf{a}_r breeding values of the remaining individuals that descent from the base population. Following Graser et al. (1987), we can express the dependence of \mathbf{a}_r on the base population breeding values \mathbf{a}_b as follows

$$\begin{pmatrix} \mathbf{a}_b \\ \mathbf{a}_r \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P}_1 & \mathbf{P}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{s} \end{pmatrix} \tag{7.20}$$

where \mathbf{s} is the vector of segregational residuals (Equation 7.9) and \mathbf{P}_1 and \mathbf{P}_2 are matrices with values of 1/2 in the parents' column in each row. Here, \mathbf{a}_r is a random effect because it is a function of a fixed effect (\mathbf{a}_b) and a random effect (\mathbf{s})

The resulting mixed-model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_b + \mathbf{Z}_2\mathbf{a}_r + \mathbf{e} \tag{7.21}$$

Graser et al. show that the resulting mixed-model equations can be written as

$$\begin{pmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z}_1 & \mathbf{X}^T \mathbf{Z}_2 \\ \mathbf{Z}_1^T \mathbf{X} & \mathbf{Z}_1^T \mathbf{Z}_1 + \lambda \mathbf{Q}^T \mathbf{G}^{-1} \mathbf{Q} & -\lambda \mathbf{Q}^T \mathbf{G}^{-1} \\ \mathbf{Z}_2^T \mathbf{X} & -\lambda \mathbf{G}^{-1} \mathbf{Q} & \mathbf{Z}_2^T \mathbf{Z}_2 + \lambda \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}}_b \\ \hat{\mathbf{a}}_r \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}_1^T \mathbf{y} \\ \mathbf{Z}_2^T \mathbf{y} \end{pmatrix} \quad (7.22a)$$

where $\lambda = (2\sigma_e^2/\sigma_A^2)$ and

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P}_2)^{-1} \mathbf{P}_1 \quad \text{and} \quad \mathbf{G} = (\mathbf{I} - \mathbf{P}_2)^{-1} \mathbf{F} [(\mathbf{I} - \mathbf{P}_2)^{-1}]^T \quad (7.22b)$$

and the elements of the diagonal matrix \mathbf{F} are given by Equation 7.10.

Estimating the Additive Variance at Generation t

Even under the infinitesimal model, the additive variance changes over time. While REML provides an estimate of the base-population additive variance (which is unbiased provided the model assumptions hold), it does not immediately provide estimates of the actual additive variance in any particular generation of selection. The most straightforward approach is to use the parent-offspring regression for each generation of selection to estimate the additive genetic variance in the parents. With parents from generation t and offspring in generation $t + 1$, the regression estimates the heritability of the parents, $h_A^2(t)$. The drawback to this approach is the typically small sample size associated with each generation (resulting in large standard errors for each heritability/variance estimate). Ideally, one would like to be able to combine information across generations in such a way as to improve the variance estimates.

Sorenson and Kennedy (1984a) suggest one approach for combining information to estimate the variance in generation t is to use a mixed-model analysis treating generation t as the base population. In particular, one considers only the data from generation t onward (say to generation T) and the relationship matrix is adjusted to assume that generation t is the base population. The resulting covariance matrix for the breeding values becomes

$$\text{Var} \begin{pmatrix} \mathbf{a}_t \\ \mathbf{a}_{t+1} \\ \vdots \\ \mathbf{a}_T \end{pmatrix} = \sigma_A^2(t) \begin{pmatrix} \mathbf{I} & \mathbf{A}_{t,t+1} & \cdots & \mathbf{A}_{t,T} \\ \mathbf{A}_{t,t+1} & \mathbf{A}_{t+1,t+1} & \cdots & \mathbf{A}_{t+1,T} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{A}_{t,T} & \mathbf{A}_{t+1,T} & \cdots & \mathbf{A}_{T,T} \end{pmatrix} \quad (7.23)$$

where \mathbf{a}_k is the vector of breeding values in generation k . \mathbf{A}_{jk} is the relationship matrix of associations between individuals in generations j and k . By assuming $\text{Var}(\mathbf{a}_t) = \mathbf{I}$, we are assuming that all individuals in generation t are unrelated and noninbred, as this is now our assumed base population. All measured individuals

from generation t (including those not leaving offspring) are included in the base population. While this approach seems logical, it is still somewhat ad-hoc and is not exact. Simulation studies by van der Werf and de Boer (1990) show that Sorenson and Kennedy's approach tends to overestimate the true variance.

Another potential candidate would be to use the variance among the predicted breeding values within a generation, viz.,

$$\text{Var}(A_t) = \frac{1}{n-1} \sum_{i=1}^{n_t} (\hat{a}_{ti} - \hat{a}_t)^2 \tag{7.24}$$

While again this is a reasonable suggestion, there are complications. First, there is an additional level of uncertainty in that the breeding values are all estimated. Second is that the assumed genetic variance used to obtain the BLUP estimates has a strong influence on the values of the estimated \hat{a}_{ti} . Hence, there is circularity in this approach and it is not recommended.

Clearly, there is not presently a satisfactory solution to the problem of combining information across generations to estimate the variance in a particular generation. This is certainly an important area for future research.

Incorporating Nonadditive Genetic Variance

Up to this point, we have been assuming all genetic variation is additive, so that we need only consider the vector \mathbf{a} of breeding values and the numerator relationship matrix \mathbf{A} . When nonadditive genetic variance is present, it creates additional genetic correlations between relatives beyond those accounted for \mathbf{A} . The simplest setting is when dominance is present. Letting the vector \mathbf{d} denote the dominance effects, the mixed model becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{d} + \mathbf{e} \tag{7.25a}$$

The overall genetic merit of an individual is estimated by $\hat{\mathbf{g}} = \hat{\mathbf{a}} + \hat{\mathbf{d}}$. Turning to the covariance structure of this model, as before $\text{Var}(\mathbf{a}) = \sigma_A^2 \mathbf{A}$ and $\text{Var}(\mathbf{e}) = \sigma_E^2 \mathbf{I}$, while the covariance matrix for dominance effects is $\text{Var}(\mathbf{d}) = \sigma_D^2 \mathbf{D}$, giving

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{A}\mathbf{Z}^T \sigma_A^2 + \mathbf{Z}\mathbf{D}\mathbf{Z}^T \sigma_D^2 + \mathbf{I} \sigma_e^2 \tag{7.25b}$$

The elements of the **dominance genetic relationship matrix** \mathbf{D} are obtained as follows. The covariance between dominance effects for (non-inbred) individuals i and j is the product of the dominance genetic variance and the coefficient of fraternity, $\sigma_D^2 \Delta_{ij}$. From LW Equation 7.7, the latter is given by

$$\Delta_{ij} = \Theta_{gk} \Theta_{hl} + \Theta_{gl} \Theta_{hk} \tag{7.26a}$$

where i 's parents are indexed by g and h and j 's by k and l , and (as above), Θ is the coefficient of coancestry. Recalling that the elements of the numerator

relationship matrix \mathbf{A} are $2\Theta_{ij}$, the off-diagonal elements of \mathbf{D} can be computed from the elements of \mathbf{A} by

$$D_{ij} = \frac{A_{gk} A_{hl} + A_{gl} A_{hk}}{4} \quad (7.26b)$$

whereas the diagonal elements are all $D_{ii} = 1$.

The mixed-model equations for Equation 7.25a become

$$\begin{pmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \lambda_A \mathbf{A}^{-1} & \mathbf{Z}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} & \mathbf{Z}^T \mathbf{Z} + \lambda_D \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{d}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \end{pmatrix} \quad (7.27)$$

where $\lambda_A = \sigma_e^2 / \sigma_A^2$, and $\lambda_D = \sigma_e^2 / \sigma_D^2$.

Epistatic terms can be included in the mixed model equations in a similar fashion, see LW Chapter 26 for details and LW Chapter 27 for modifications of the REML equations to estimate nonadditive variances.

While the above treatment suggests that dominance is easily incorporated, this is misleading, as we have made the tacit assumption of no selection and (much more importantly) no inbreeding. Assuming the infinitesimal model, selection does not significantly change the dominance variance (Chapter 5, Bulmer 1971b) and hence the estimates of \mathbf{a} and \mathbf{d} are not likely to be biased by selection. Inbreeding (which occurs in all selection experiments), on the other hand, introduces major complications. First, there may be inbreeding depression. In some situations, this can be dealt with by including inbreeding depression as a covariate, for example by using a model such as

$$y_{ti} = \mu + I f(t) + a_{ti} + d_{ti} + e_{ti} \quad (7.28)$$

where $f(t)$ is the inbreeding in generation t and I is the inbreeding depression under complete inbreeding. Recall (LW Chapter 10) that with only dominance, inbreeding depression is a linear function of the inbreeding f , while with epistasis it is a nonlinear function of f . Thus, if there is significant epistasis, Equation 7.28 may not properly correct for inbreeding depression, especially at high values of f (those approaching one). One can also include a control population with known levels of inbreeding to provide an independent estimation of I . In a LS square analysis, inbreeding depression is typically assumed to be corrected for by estimating response by subtracting off the mean of such a control population. This assumes that the control and selected lines have the same level of inbreeding, when in fact the selected population is typically far more inbred (Chapter 12). If the levels of inbreeding are very similar in both populations, the use of a control can account for nonlinear inbreeding depression.

A second, and far more serious, complication is that the covariance between inbred relatives with dominance is no longer a function of just σ_A^2 and σ_D^2 . As discussed in Chapter 3, these covariances depend upon four other quadratic components as well ($\sigma_{DI}^2, \sigma_{ADI}, i^*, i^2 - i^*$), see Equations 3.14-3.16. Recall that these additional components fully account for inbreeding depression. While one could formulate a mixed model incorporating all six quadratic components (using the covariances given by Equation 3.14-3.16), the resulting model is extremely complex and numerically very demanding. A start in this direction has been developed by Smith and Mäki-Tanila (1990), who should be consulted for more details. A second approach is to use Equation 7.35, with a **D** matrix that approximates the elements under inbreeding, as suggested by Smith and Allaire (1985), who again should be consulted for details. One could also combine the use of a modified **D** with a covariate for inbreeding depression (Equation 7.28), but this is still a largely ad-hoc approach to a complex problem. The situation with epistasis is even more complex than with simple dominance. Clearly, estimating response with nonadditive variance is an important area for future research.

BAYESIAN ANALYSIS OF SELECTION EXPERIMENTS

An important breakthrough in the estimate of selection response is offered by Bayesian approaches. While these approaches are often touted for their ability to incorporate prior information, we view their main utility as providing a complete description of the uncertainty of an estimate (say breeding values **a**) that itself relies on other parameters being estimated (e.g., σ_A^2 , **b**). We start by first introducing some of the basic ideas of Bayesian statistics and the powerful Gibbs sampler method used to approximate the posterior distributions used in Bayesian analysis. We then discuss Sorensen et al.'s (1994) application of the Gibbs sampler for a Bayesian analysis of selection response and conclude by applying their approach to estimate the response in little size components in pigs.

Introduction to Bayesian Statistics

Our treatment here is intentionally quite brief and we refer the reader to Lee (1997) for a complete introduction to Bayesian analysis. While very deep (and very subtle) differences in philosophy separate hard-core **Bayesians** from hard-core **frequentists** (Efron 1986, Glymour 1981), our treatment here of Bayesian methods is motivated simply by their use as a powerful statistical tool. Its widespread introduction to quantitative genetics can be credited (or blamed) on the influential paper of Gianola and Fredando (1986), which reviews Bayesian methods in animal breeding.

The foundation of Bayesian Statistics is **Bayes' theorem**. Suppose there are n possible outcomes (b_1, b_2, \dots, b_n) of a random variable that we cannot observe. Given the observed outcome of an observed variable y correlated with b , what is

the probability of b_j ? From the definition of conditional probability, $\Pr(b_j | y) = \Pr(b_j, y) / \Pr(y)$. Bayes' theorem follows by noting that $\Pr(b_j, y) = \Pr(b_j) \Pr(y | b_j)$ and $\Pr(y) = \sum_i^n \Pr(b_i) \Pr(y | b_i)$, giving

$$\Pr(b_j | y) = \frac{\Pr(b_j) \Pr(y | b_j)}{\Pr(y)} = \frac{\Pr(b_j) \Pr(y | b_j)}{\sum_{i=1}^n \Pr(b_i) \Pr(y | b_i)}$$

$\Pr(b_j)$ is the **prior distribution** of the possible b values, while $\Pr(b_j | y)$ is the **posterior distribution** of b given the observed data y . The origin of Bayes' theorem has a fascinating history (Stigler 1983). It is named after the rev. Thomas Bayes, a priest who never published a mathematical paper in his life. The paper in which the theorem appears was posthumously read before the Royal Society by his friend Richard Price in 1764. Stigler (1983) suggests it was first discovered by Nicholas Saunderson, a blind mathematician/optician who, at age 29, became Lucasian Professor of Mathematics at Cambridge (the position held earlier by Issac Newton)

Example 6. Suppose a major gene (with alleles **Q** and **q**) underlies the character of interest. The distribution of phenotypic values given major gene genotype follow a normal distribution with variance one and means 2.1, 3.5, and 1.3 for the genotypes **QQ**, **Qq**, and **qq** (respectively). Suppose the frequencies of these genotypes for a random individual drawn from the population are 0.3, 0.2, and 0.5 (for **QQ**, **Qq**, and **qq** respectively). If an individual from this population has a phenotypic value of 3, what is the probability of it being **QQ**? **Qq**? **qq**?

Let $\varphi(z | \mu, 1) = (2\pi)^{-1/2} e^{-(z-\mu)^2/2}$ denote the density function for a normal with mean μ and variance one. To apply Bayes' theorem, the values for the priors and the conditionals are as follows:

Genotype, G	Pr(G)	Pr(Y G)	Pr(G)·Pr(Y G)
QQ	0.3	$\varphi(3 2.1, 1) = 0.177$	0.053
Qq	0.2	$\varphi(3 3.5, 1) = 0.311$	0.062
qq	0.5	$\varphi(3 1.3, 1) = 0.022$	0.011

Since $\sum_G \Pr(G) \cdot \Pr(Y | G) = 0.126$, Bayes' theorem gives the posterior probabilities for the genotypes given the observed value of 3 as:

$$\Pr(\mathbf{QQ} | y = 3) = (0.177 \cdot 0.3) / 0.126 = 0.053 / 0.126 = 0.421$$

$$\Pr(\mathbf{Qq} | y = 3) = (0.311 \cdot 0.2) / 0.126 = 0.062 / 0.126 = 0.491$$

$$\Pr(\mathbf{qq} | y = 3) = (0.022 \cdot 0.5) / 0.126 = 0.011 / 0.126 = 0.088$$

Thus, there is a 42 percent chance this individual has genotype **QQ**, a 49 percent chance it is **Qq**, and only an 8.8 percent chance it is **qq**.

More generally, the continuous, vector-valued version of Bayes' theorem is given by:

$$p(\Theta | \mathbf{y}) = \frac{p(\mathbf{y} | \Theta) p(\Theta)}{\int p(\mathbf{y}, \Theta) d\Theta} = \frac{p(\mathbf{y} | \Theta) p(\Theta)}{p(\mathbf{y})} \quad (7.29)$$

where $\Theta^T = (\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(k)})$ is a vector of k (potentially) continuous variables. As with the discrete case, $p(\Theta)$ is the assumed prior distribution of the unknown parameters, while $p(\Theta | \mathbf{y})$ is the posterior distribution given the prior $p(\Theta)$ and the data \mathbf{y} .

The dependence of the posterior on the prior (which can easily be assessed by trying different priors) provides an indication of how much information on the unknown parameter values is contained in the data. If the posterior is highly dependent on the prior, then the data likely has little signal, while if the posterior is largely unaffected by the shape of the assumed prior, the data are highly informative.

Often, only a subset of the unknown parameters is really of concern the us, the rest being **nuisance parameters** that we wish to remove. Write the vector of unknown parameters as $\Theta^T = (\Theta_1^T, \Theta_n^T)$, where Θ_n is the vector of nuisance parameters. Thus, we would like to condition out the dependence of the parameters of interest from the effects of the nuisance parameters, generating a **marginal posterior distribution** that does not depend on the nuisance parameters. Integrating over Θ_n gives the desired marginal as

$$p(\Theta_1 | \mathbf{y}) = \int p(\Theta | \mathbf{y}) p(\Theta_n | \mathbf{y}) d\Theta_n = E_{\Theta_n} [p(\Theta_1 | \Theta_n, \mathbf{y})] \quad (7.30)$$

The conditional probabilities in Equation 7.30 can be evaluated using Equation 7.29.

Example 7. In the context of analyzing a selection response experiment, the vector of breeding values \mathbf{a} is of interest, while the fixed effects (\mathbf{b}) and variances (σ_A^2, σ_e^2) are generally regarded as nuisance parameters. In this case, Equation 7.30 gives the marginal distribution of the breeding values \mathbf{a} given the data as

$$p(\mathbf{a} | \mathbf{y}) = \int p(\mathbf{a}, \mathbf{b}, \sigma_A^2, \sigma_e^2 | \mathbf{y}) p(\mathbf{b}, \sigma_A^2, \sigma_e^2 | \mathbf{y}) d\mathbf{b} d\sigma_A^2 d\sigma_e^2$$

This conditioning removes any dependencies of estimates of the response on estimates of the variance components (and fixed effects). Uncertainties introduced

by estimating these nuisance parameters are automatically accommodated when considering the marginal distribution. While this multidimensional integral is complex, the Gibbs sampler (below) can be used to approximate drawings from this marginal distribution.

With the marginal density $p(\mathbf{a} | \mathbf{y})$ in hand, one can obtain estimates of the response to selection $\mathbf{K}^T \mathbf{a}$ that are *independent* of the assumed (or estimated) additive variance σ_A^2 . The error due to estimation of the additive variance from the data is directly incorporated when the marginal is computed as we integrate over possible values of σ_A^2 and their support given the data. This independence of the estimate of response from the estimate of additive variance and the subsequent incorporation of the error in estimating σ_A^2 in the estimate of the response are two very compelling reasons for a Bayesian analysis of response.

The Gibbs Sampler

One practical limitation towards more widespread implementation of Bayesian approaches is that obtaining the posterior distribution often requires the integration of high dimensional functions (e.g., Equations 7.29 and 7.30). This can be computationally very difficult, but several approaches short of direct integration have been proposed (Smith 1991, Evans and Swartz 1995). Perhaps the most powerful is the **Gibbs sampler** method, one of a family of **Markov Chain Monte Carlo (MCMC)** approaches. The ideal behind Monte Carlo methods is that instead of having to compute an extremely difficult multidimensional integral, we can draw samples from the multidimensional distribution associated with the integral and use these to compute the desired quantity of interest.

Simulating random vectors drawn from some complex **target distribution** can be a very difficult task. The idea behind MCMC approaches is to successively draw samples from far simpler distributions in such a way that the distribution of the samples converges to the target distribution. MCMC approaches are so-named because one uses the previous sample values to randomly generate the next sample value, generating a **Markov chain** (the transition probabilities between sample values are only functions of the most recent sample value). The roots of MCMC methods go back to attempts by mathematical physicists to integrate very complex functions by random sampling (Metropolis and Ulam 1949, Metropolis et al. 1953, Hastings 1970). Under the **Metropolis-Hastings algorithm**, one simulates draws from a complex target distribution by first drawing a random variable from a specified distribution and then using a second probability distribution to decide whether to keep that realization or reject it. The Gibbs sampler (introduced in the context of image processing by Geman and Geman 1984), is a special case of Metropolis-Hastings sampling wherein the random value is always accepted. The task remains to specify how to construct a Markov Chain whose values converge to the target distribution. The key to the Gibbs sampler is that one only consid-

ers conditional distributions — the distribution when all of the random variables but one are assigned fixed values. Such conditional distributions are far easier to simulate than complex joint distributions and usually have simple forms (often being normals). Thus, one simulates n random variables sequentially from the n univariate conditionals rather than generating a single n -dimensional vector in a single pass using the full joint distribution.

To introduce the Gibbs sampler, consider a bivariate random variable (x, y) , and suppose we wish to compute one or both marginals, $p(x)$ and $p(y)$. The idea behind the sampler is that it is far easier to consider a sequence of conditional distributions, $p(x | y)$ and $p(y | x)$, than it is to obtain the marginal by integration of the joint density $p(x, y)$, e.g., $p(x) = \int p(x, y)dy$. The sampler starts with some initial value y_0 for y and obtains x_0 by generating a random variable from the conditional distribution $p(x | y = y_0)$. The sampler then uses x_0 to generate a new value of y_1 , drawing from the conditional distribution based on the value x_0 , $p(y | x = x_0)$. The sampler proceeds as follows

$$x_i \sim p(x | y = y_{i-1}) \tag{7.31a}$$

$$y_i \sim p(y | x = x_i) \tag{7.31b}$$

Repeating this process k times, generates a **Gibbs sequence** of length k , where a subset of points (x_j, y_j) for $1 \leq j \leq m < k$ are taken as our simulated draws from the full joint distribution. To obtain the desired total of m sample points, one samples the chain (i) after a sufficient **burn-in** to removal the effects of the initial sampling values and (ii) at set time points (say every n samples) following the burn-in. The Gibbs sequence converges to a stationary (equilibrium) distribution that is independent of the starting values, and by construction this stationary distribution is the target distribution we are trying to simulate (Tierney 1991, 1994).

Example 8. The following distribution is from Casella and George (1992). Suppose the joint distribution of $x = 0, 1, \dots, n$ and $0 \leq y \leq 1$ is given by

$$p(x, y) = \frac{n!}{(n-x)!x!} y^{x+\alpha-1} (1-y)^{n-x+\beta-1}$$

Note that x is discrete while y is continuous. While the joint density is complex, the conditional densities are simple distributions. To see this, first recall that a binomial random variable z has a density proportional to

$$p(z | q, n) \propto \frac{q^z(1-q)^{n-z}}{z!(n-z)!} \quad \text{for } 0 \leq z \leq n$$

where $0 \leq q \leq 1$ is the success parameter and n the number of traits, and we denote $z \sim B(n, p)$. A distribution that may be less familiar to the reader is the **beta distribution** with parameters a and b , whose density is proportional to

$$p(z | a, b) \propto z^{a-1}(1-z)^{b-1} \quad \text{for } 0 \leq z \leq 1$$

Denote a random variable z following a beta distribution with shape parameters a and b by $z \sim \text{Beta}(a, b)$. With these probability distributions in hand, note that $p(x | y) \sim B(n, y)$, while $p(y | x) \sim \text{Beta}(x + \alpha, n - x + \beta)$.

The power of the Gibbs sampler is that by computing a sequence of these univariate conditional random variables (a binomial and then a beta) we can compute any feature of either marginal distribution. Suppose $n = 10$, and $\alpha = 1, \beta = 2$. Start the sampler with (say) $y_0 = 1/2$, and we will take the sampler through three full iterations.

- (i) x_0 is obtained by generating a random $B(n, y_0) = B(10, 1/2)$ random variable, giving $x_0 = 5$ in our simulation.
- (ii) y_1 is obtained from a $\text{Beta}(x_0 + \alpha, n - x_0 + \beta) = \text{Beta}(5 + 1, 10 - 5 + 2)$ random variable, giving $y_1 = 0.33$.
- (iii) x_1 is a realization of a $B(n, y_1) = B(10, 0.33)$ random variable, giving $x_1 = 3$.
- (iv) y_2 is obtained from a $\text{Beta}(x_1 + \alpha, n - x_1 + \beta) = \text{Beta}(3 + 1, 10 - 3 + 2)$ random variable, giving $y_2 = 0.56$.
- (v) x_2 is obtained from a $B(n, y_2) = B(10, 0.56)$ random variable, giving $x_2 = 0.7$.

Our particular realization of the Gibbs sequence after three iterations is thus (5, 0.5), (3, 0.33), (7, 0.56). We can continue this process to generate a chain of the desired length. Obviously, the initial values in the chain are highly dependent upon the y_0 value chosen to start the chain. This dependence decays as the sequence length increases and so we typically start recording the sequence after a sufficient number of burn-in iterations have occurred to remove any effects of the starting conditions.

When more than two variables are involved, the sampler is extended in the obvious fashion. In particular, the value of the k th variable is drawn from the distribution $p[\theta^{(k)} | \Theta^{(-k)}]$ where $\Theta^{(-k)}$ denotes a vector containing all off the variables but k . Thus, during the i th iteration of the sample, to obtain the value of $\theta_i^{(k)}$ we draw from the distribution

$$\theta_i^{(k)} \sim p[\theta^{(k)} | \theta^{(1)} = \theta_i^{(1)}, \dots, \theta^{(k-1)} = \theta_i^{(k-1)}, \theta^{(k+1)} = \theta_{i-1}^{(k+1)}, \dots, \theta^{(n)} = \theta_{i-1}^{(n)}]$$

For example, if there are four variables, (w, x, y, z) , the sampler becomes

$$\begin{aligned} w_i &\sim p(w \mid x = x_{i-1}, y = y_{i-1}, z = z_{i-1}) \\ x_i &\sim p(x \mid w = w_i, y = y_{i-1}, z = z_{i-1}) \\ y_i &\sim p(y \mid w = w_i, x = x_i, z = z_{i-1}) \\ z_i &\sim p(z \mid w = w_i, x = x_i, y = y_i) \end{aligned}$$

One can either use a single long chain (Geyer 1992, Raftery and Lewis 1992b) or multiple chains each starting from different initial values (Gelman and Rubin 1992). Note that with parallel processing machines, using multiple chains may be computationally more efficient than a single long chain. Diagnostics to check whether the chain has indeed approached the stationary distribution (and hence samples from it are indeed samples from the target distribution) are discussed by Geyer (1992), Gelman and Rubin (1992), Raftery and Lewis (1992a, 1992b), Geweke (1992), Robert (1995), . The simplest approach is to look at the autocovariance structure between samples with different time lags, setting the sampling interval as the point where the autocovariance approaches zero. Gelfand and Smith (1990) illustrated the power of the Gibbs sampler to address a wide variety of statistical issues, while Smith and Roberts (1993) showed the natural marriage of the Gibbs sampler with Bayesian statistics (in obtaining posterior distributions). A nice introduction to the sampler is given by Casella and George (1992), while further details can be found in Tanner (1991), Besag et al. (1995), and Lee (1997). Finally, note that the Gibbs sampler can be thought of as a stochastic analog to the EM (Expectation-Maximization) approaches used to obtain likelihood functions when missing data are present (LW Chapter 27, LW Appendix 4). In the sampler, random sampling replaces the expectation and maximization steps.

Using the Gibbs Sampler to Approximate Marginal Distributions

Any feature of interest for the marginals can be computed from the m realizations of the Gibbs sequence. For example, the expectation of any function f of the random variable x is approximated by

$$E[f(x)]_m = \frac{1}{m} \sum_{i=1}^m f(x_i) \tag{7.32a}$$

This is the **Monte-Carlo (MC) estimate** of $f(x)$, as $E[f(x)]_m \rightarrow E[f(x)]$ as $m \rightarrow \infty$. Likewise, the MC estimate for any function of n variables $(\theta^{(1)}, \dots, \theta^{(n)})$ is given by

$$E[f(\theta^{(1)}, \dots, \theta^{(n)})]_m = \frac{1}{m} \sum_{i=1}^m f(\theta_i^{(1)}, \dots, \theta_i^{(n)}) \tag{7.32b}$$

Example 9. Although the sequence of length 3 computed in Example 8 is too short (and too dependent on the starting value) to be a proper Gibbs sequence, for illustrative purposes we can use it to compute Monte-Carlo estimates. The MC estimate of the means of x and y are

$$\bar{x}_3 = \frac{5 + 3 + 7}{3} = 5, \quad \bar{y}_3 = \frac{0.5 + 0.33 + 0.56}{3} = 0.46$$

Similarly, $\left(\overline{x^2}\right)_3 = 27.67$ and $\left(\overline{y^2}\right)_3 = 0.22$, giving the MC estimates of the variances of x and y as

$$\text{Var}(x)_3 = \left(\overline{x^2}\right)_3 - (\bar{x}_3)^2 = 2.67$$

and

$$\text{Var}(y)_3 = \left(\overline{y^2}\right)_3 - (\bar{y}_3)^2 = 0.25$$

While computing the MC estimate of any moment using the sampler is straightforward, computing the actual shape of the marginal density is slightly more involved. While one might use the Gibbs sequence of (say) x_i values to give a rough approximation of the marginal distribution of x , this turns out to be inefficient, especially for obtaining the tails of the distribution. A better approach is to use the average of the conditional densities $p(x | y = y_i)$, as the function form of the conditional density contains more information about the shape of the entire distribution than the sequence of individual realizations x_i (Gelfand and Smith 1990, Liu et al. 1991). Since

$$p(x) = \int p(x | y) p(y) dy = E_y [p(x | y)] \quad (7.33a)$$

one can approximate the marginal density using

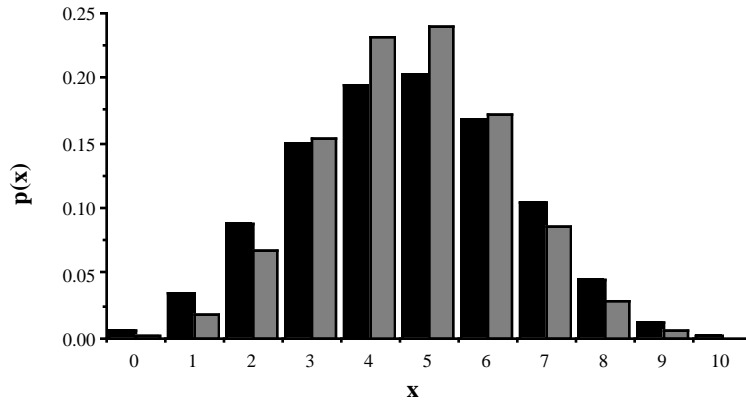
$$\hat{p}_m(x) = \frac{1}{m} \sum_{i=1}^m p(x | y = y_i) \quad (7.33b)$$

Example 10. Returning to the Gibbs sequence generated in Example 8, recall that the distribution of x given y is binomial, with $x | y \sim B(n, y)$. Applying Equation 7.33b the estimate (based on this sequence) of the marginal distribution

of x is the weighted sum of three binomials with success parameters 0.5, 0.33, and 0.56, giving

$$p_3(x) = 10! \left[\frac{0.5^x(1 - 0.5)^{10-x} + 0.33^x(1 - 0.33)^{10-x} + 0.56^x(1 - 0.56)^{10-x}}{3 x!(10 - x)!} \right]$$

As the figure below shows, the resulting distribution (solid bars), although a weighted sum of binomials, departs substantially from the best-fitting binomial (success parameter 0.46333, stripped bars)



The Method of Sorensen, Wang, Jensen, and Gianola

Sorensen, Wang, Jensen, and Gianola (1994) applied the above machinery to suggest a develop Gibbs-sampler based Bayesian analysis of selection response based on the animal model,

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \tag{7.34a}$$

As before, the conditionals are

$$\mathbf{y} | \mathbf{b}, \mathbf{a}, \sigma_e^2 \sim \text{MVN}(\mathbf{Xb} + \mathbf{Za}, \mathbf{I}\sigma_e^2) \tag{7.34b}$$

while the infinitesimal model is assumed, with

$$\mathbf{a} | \mathbf{A}, \sigma_A^2 \sim \text{MVN}(\mathbf{0}, \mathbf{A}\sigma_A^2) \tag{7.34c}$$

Sorensen et al. provide expressions for the $p + q + 2$ conditional distributions $p(\theta^{(i)}, | \Theta^{(-i)})$ for the $p + q + 2$ variables in the model (p breeding values a_i , q fixed effects b_j , and the two variances σ_A^2 and σ_e^2). They assume a uniform prior for \mathbf{b} , a normal prior for \mathbf{a} , and both uniform and inverted χ^2 priors for the

variances. For example, the prior for \mathbf{a} is given by $p(\mathbf{a} | \sigma_A^2, \mathbf{A}) \cdot p(\sigma_A^2)$, where the first distribution is a multivariate normal (Equation 7.34c) and the second either a uniform or an inverted χ^2 . Using these expressions, the Gibbs sampler can be run, from which Monte Carlo estimates of the mean and variance for the additive variance and mean breeding values can be obtained. Likewise, posterior marginals for the additive variance and mean breeding values can be obtained (Example 7).

The outline for the sampler is as follows:

1. Set initial values for \mathbf{a} , \mathbf{b} , σ_A^2 , σ_e^2 .
2. Using the current values of \mathbf{a} , \mathbf{b} , σ_A^2 , σ_e^2 and the conditional distributions (see Sorensen et al. for exact expressions):
 - (i) update the fixed effects by sequentially drawing (for $i = 1, \dots, q$) from the conditionals (which are univariate normals)

$$b_{j,i} \sim p(b_j | b_{1,i}, \dots, b_{j-1,i}, b_{j+1,i-1}, \dots, b_{q,i-1}, \mathbf{a}_{i-1}, \sigma_{A,i-1}^2, \sigma_{e,i-1}^2)$$

- (ii) Update the breeding values by sequentially drawing (for $i = 1, \dots, p$) from the conditionals (again univariate normals)

$$a_{j,i} \sim p(a_j | \mathbf{b}_i, a_{1,i}, \dots, a_{j-1,i}, a_{j+1,i-1}, \dots, a_{p,i-1}, \sigma_{A,i-1}^2, \sigma_{e,i-1}^2)$$

- (iii) Update the additive variance, drawing from the conditional (a univariate inverted χ^2 distribution)

$$\sigma_{A,i}^2 \sim p(\sigma_A^2 | \mathbf{b}_i, \mathbf{a}_i, \sigma_{e,i-1}^2)$$

- (iv) Update the error variance, drawing from the conditional (again, a univariate inverted χ^2 distribution)

$$\sigma_{e,i}^2 \sim p(\sigma_e^2 | \mathbf{b}_i, \mathbf{a}_i, \sigma_{A,i}^2)$$

3. Using the updated values, repeat (2) until k samples are obtained, from which m are extracted for the Gibbs sampler chain.

Note that the Bayesian analysis of the animal model makes most of the standard animal-model assumptions, in particular that the infinitesimal model holds (so that covariance matrix for breeding values is $\sigma_A^2 \mathbf{A}$). Thus, both MM and Bayesian analysis are potentially biased by changes in allele frequencies. The advantage of a Bayesian analysis is that it has all the advantages of a MM analysis (over a LS analysis) and, in addition, the posterior marginals correctly give the distribution of any parameter of interest independent of the values assumed for other parameters. The uncertainty introduced by estimating these additional parameters is reflected in the posterior marginal. Thus, the Bayesian approach gives

the correct distribution (assuming the model assumptions hold) for the estimated response independent of the additive genetic variance. By contrast, a MM analysis is highly dependent on the assumed (or estimated) additive variance, and the standard error of a REML/BLUP estimate for the response (Equation 7.8c) does not account for the variance in REML estimation of σ_A^2 .

Application: Estimating Response in Pig Litter Size Components

Blasco, Sorensen, and Bidanel (1998) used the method of Sorensen et al. (1994) to estimate the response to selection on ovulation rate and prenatal survival in French Large White pigs. Three lines were followed, one selecting on each trait and a control line. The relevant selection and control lines were jointly analyzed to estimate response. Ovulation rate was examined using the standard animal model,

$$\mathbf{y} | \mathbf{b}, \mathbf{a}, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Za}, \mathbf{I}\sigma_e^2), \quad \mathbf{a} | \mathbf{A}, \sigma_A^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_A^2)$$

Prenatal survival (as a function of the mother) was examined using the repeatability model,

$$\mathbf{y} | \mathbf{b}, \mathbf{a}, \mathbf{c}, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Za} + \mathbf{Wc}, \mathbf{I}\sigma_e^2)$$

$$\mathbf{a} | \mathbf{A}, \sigma_A^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_A^2), \quad \mathbf{c} | \sigma_c^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_c^2)$$

Among the fixed effects in \mathbf{b} are terms for the **parity** of the mother (1st parity = 1st litter, 2nd parity = second litter and so on). The marginal posterior distribution for breeding values (and hence for the response via $\mathbf{K}^T \mathbf{a}$) was obtained by using the Gibbs sampler approach of Sorensen et al. (1994). For each trait, two independent chains of length 100,000 were computed, with the first 10,000 samples discarded (to remove burn-in effects) and sampling at every 30 iterations thereafter, generating a sampler of length 3000. The authors obtained these burn-in and resampling values after several initial runs and using the diagnostics suggested by Raftery and Lewis (1992a) for level of precision and Geyer (1992) for autocorrelation between samples. A uniform prior was taken for the fixed effects, while different priors used for the variances (discussed below).

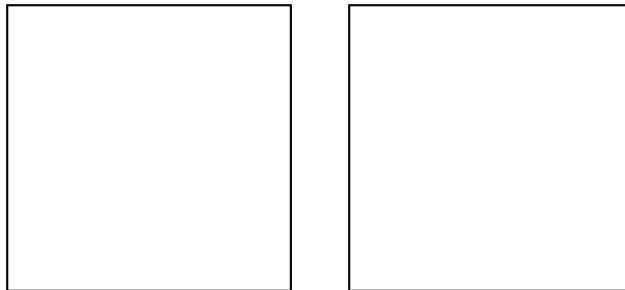


Figure 7.2. Analysis of ovulation rate at puberty in French Large White pigs. **Left:** Assumed priors for σ_A^2 (see text for details). **Right:** The Bayesian estimate of

response is given by the posterior density for the mean breeding value in ovulation rate in the last generation of selection. The distribution is approximately normal (solid curve). After Blasco et al. (1998).

Consider the results for ovulation rate at puberty first. Figure 7.2 shows the three priors assumed for the additive variance in this trait based on prior information. The phenotypic variance of this trait is 6.25, setting an upper limit on σ_A^2 . Prior one is a uniform distribution (an **uninformative prior**) that weights all values in the parameter space equally. Priors two and three (based on scaled inverted χ^2 distributions, chosen for analytic tractability) reflect additional information. Published heritabilities for this trait in pigs and rabbits ranges from 0.1 to 0.6, and prior two assumes a broad distribution around the approximate medium value ($\sigma_A^2 = 0.4 \cdot 6.25 = 2.5$). A study specifically in French Large Whites gave an estimate of $h^2 = 0.11 \pm 0.02$ and the tight distribution around this value is reflected in prior three. Using the approach for Sorensen et al. (1994), Blasco et al. obtain Monte-Carlo estimates of the (base population) heritability under these three priors of $h^2 = 0.39 \pm 0.07$, 0.39 ± 0.06 , and 0.32 ± 0.06 . Table 7.1 shows the estimated response during each of the four generations of selection, comparing these with the least squares (differences between generation means) and mixed model (REML/BLUP) estimates. Note that the three different priors give very consistent estimates of response, implying that the data contain sufficient information to overpower most of the signal coming from the assumed prior. The Bayesian and MM analysis give very similar results, while the LS results give a slightly different estimates of response.

Table 7.1 Estimated response to selection for ovulation rate at puberty and prenatal survival in French Large White pigs. Bayesian analysis with three different priors (Figure 7.2 for ovulation, Figure 7.3 for prenatal survival) were used to obtain Monte-Carlo estimates of the mean response and their associated standard errors (the later incorporating the additional error from estimating the additive variance and other parameters). For comparison, least squares (LS) estimates ($\bar{z}_{i+1} - \bar{z}_i$) and mixed-model (REML/BLUP) estimates are also included. After Blasco et al. (1998).

Ovulation Rate at Puberty				
Method	Gen 1	Gen 2	Gen 3	Gen 4
Bayesian, Prior 1	0.30 ± 0.31	0.51 ± 0.35	1.03 ± 0.39	1.58 ± 0.43
Bayesian, Prior 2	0.31 ± 0.30	0.51 ± 0.34	1.05 ± 0.38	1.55 ± 0.42
Bayesian, Prior 3	0.31 ± 0.31	0.51 ± 0.35	1.01 ± 0.35	1.53 ± 0.38
LS	-0.09	0.35	1.98	1.87
REML/BLUP	0.27	0.45	1.00	1.54

Prenatal Survival

Method	Gen 1	Gen 2	Gen 3	Gen 4
Bayesian, Prior 1	-0.53 ± 1.44	1.23 ± 1.61	2.83 ± 1.94	2.89 ± 2.12
Bayesian, Prior 2	-0.64 ± 1.70	1.50 ± 1.87	3.46 ± 2.05	3.49 ± 2.30
Bayesian, Prior 3	-0.46 ± 1.45	1.22 ± 1.60	2.84 ± 1.82	2.90 ± 2.01
LS	-5.71	2.11	4.13	-2.82
REML/BLUP	-0.54	1.49	3.26	3.42

While the results for ovulation rate are consistent across the three priors and with the MM analysis, the results are more problematic for prenatal survival (Table 7.1). Figure 7.2 shows the assumed different priors. As with ovulation rate, prior one is the uninformative prior, weighting all potential additive variances equally. Prior two (as with prior 2 for ovulation rate) assumes a broad distribution around the mean heritability ($h^2 \simeq 0.2$) for a number of studies, while prior three uses the estimate of $h^2 = 0.03 \pm 0.03$ found using French Large Whites. The three priors give Monte Carlo estimates of heritability (and its standard error) of $h^2 = 0.12 \pm 0.06$, 0.16 ± 0.04 , and 0.11 ± 0.04 . Likewise, these priors give Monte Carlo estimates of the repeatability of 0.23 ± 0.05 , 0.23 ± 0.04 , and 0.19 ± 0.04 . As Table 7.1 shows, the standard errors for the Monte Carlo estimates of mean response are very large, but that the three priors and the MM analysis give consistent results, while the LS results are quite different. Clearly, the information in the experiment is sufficiently small that the posterior is strongly influenced by the prior.

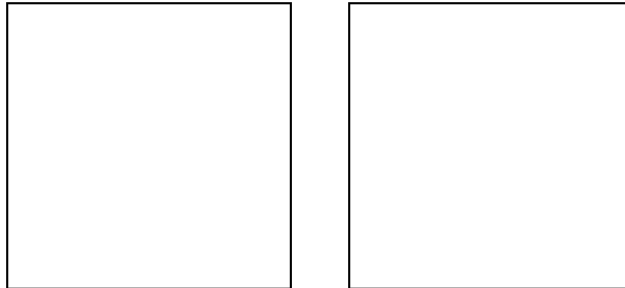


Figure 7.3. Analysis of prenatal survival in French Large White pigs. **Left:** Assumed priors for the additive variance (see text for details). **Right:** Posterior distribution of mean breeding values (at generation four) in prenatal survival. After Blasco et al. (1998).