Review

Estimation of quantitative genetic parameters

Robin Thompson[1,2,3,*]

[1]*School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK*
[2]*Centre for Mathematical and Computational Biology, and* [3]*Department of Biomathematics and Bioinformatics, Rothamsted Research, Harpenden AL5 2JQ, UK*

This paper gives a short review of the development of genetic parameter estimation over the last 40 years. The need to analyse genetic processes in both animal selection experiments and animal breeding improvement programmes motivated the majority of this work. The usage of animal model in conjunction with residual maximum likelihood (REML) techniques for mixed models has revolutionized the methods. These methods to estimate quantitative genetic parameters have recently been advocated for use in evolutionary studies of natural populations. Therefore, it is perhaps timely to discuss the development of REML methods and their application to the analysis of artificial selection experiments and breeding programmes in animals. This should give extra insight into the methods and hopefully lead to synergy between both the areas.

Keywords: genetic parameters; variance components; heritability; residual maximum likelihood; selection

1. INTRODUCTION

Given the recent enthusiasm for the usage of residual maximum likelihood (REML) techniques for the analysis of mixed models in conjunction with the animal model for the genetic analysis of natural populations (Kruuk 2004), it is perhaps timely to review the development of these techniques in which the author has been actively involved. In writing this paper, the author had two aims.

The first aim is to explain some background of the development of the REML method. It has been explained in §2 by discussing the animal breeding problems that motivated the author's interest and why existing techniques needed improvement. How the technique best linear unbiased prediction (BLUP) helps both in computational terms and at a conceptual level is also shown. It is also explained how when taking account of the structure of the model, resulting prediction equations and likelihood can make it easier to estimate parameters. Finally, it is shown how the computational methods have evolved and how the variety of genetic models that can be estimated has increased.

The second aim is to review the animal quantitative genetics work on selection and relate it to evolutionary studies with the hope of contrasting the strengths of the two approaches. In §3, it is shown how the REML techniques were extended to deal with animal selection experiments and improvement programmes with emphasis on the estimation of variance parameters taking account of selection and prediction of breeding values. In §4, these approaches are compared with those used in evolutionary studies with emphasis on measuring selection responses. It is hoped that this might lead to synergy between the analysis of artificial and natural populations.

*robin.thompson@bbsrc.ac.uk

2. ESTIMATION IN MIXED MODELS

(a) *Motivation*

In the early 1950s, a system of dairy cattle genetic improvement programmes in the UK was introduced which followed the paradigm of choosing the best of tested young bulls to father more daughters and the next generation of young bulls. The genetic merit of young bulls for milk production was evaluated by comparing the milk production of their daughters with that of daughters of other bulls. Owing to the possible environmental effects of herd, year and season, daughters were grouped into contemporary groups, i.e. cows in the same herd, year and season. The subsequent design was very unbalanced with a small contemporary group size and although each bull was used in many herds, all bulls were not used in each contemporary group. When first introduced, the computational facilities were very limited and the evaluation procedure was two-staged. In the first stage, the daughter means were corrected for the environmental effects of contemporary groups. In the second stage, the breeding values of the sires were predicted by regressing these adjusted progeny means to the overall mean, with the regression coefficient depending on the additive genetic variance and a measure of residual variance in the progeny mean. In effect, this was a mixed model with contemporary groups as fixed effects, random sire effects and the variance of the sire effects depending on the additive variance. The emphasis was not only mainly on evaluating the bulls with the contemporary effects, just as a nuisance factor, but there was also interest in estimating genetic variances and covariances between traits. In the 1970s, the computing facilities were improving. It became computationally feasible to make better adjustments for the contemporary group effects taking account of the genetic differences between contemporary groups arising owing to unbalance (Thompson 1976).

(b) *Early analysis*

Mixed models had been used implicitly in agricultural experiments since the 1930s. The emphasis was on estimating experimental treatment effects (fixed effects) as accurately as possible by taking account of random effects. The random effects depend on the context. In animal experiments on pigs, one might standardize the litter size and apply different treatments to the individual pigs in each litter and treat litter effects as random effects. In crop experiments, the random effects might be blocks or groups of similar plots. If the number of treatments is larger than the size of the litter, then one would like to use a design allocating treatments to pigs within litters which allowed the variances of all treatment comparisons corrected for litter effects to be the same. Yates introduced designs to do this, called balanced incomplete block (BIB) designs. Further there are two possible types of comparisons, one within litters, which have no litter component, and one from comparisons of litter totals which include different sets of treatments, which do have a litter component. One would like to combine this information and take account of the different variances in the different comparisons. Yates (1940) suggested constructing an efficient weighted average of these two estimates and called this the recovery of inter-block information. Yates used an analysis for BIB designs, based on an analysis of variance, by first separating variation into two parts, or strata, within and between blocks (the latter component is also termed 'interblock'). The weights depend on the variation in the two strata and the variance of the treatment effect in each stratum. There were three sources of information to estimate the variation: the two residual sums of squares in the two strata and the comparison of treatment effects in the two strata. Later Nelder (1968) extended the method to generally balanced designs which allowed more than two sources of variation, or strata, and allowed the treatment effects to be partitioned into sets with different efficiencies. In these examples, estimation of variance components was essentially used to provide appropriate weights for the fixed effects and give appropriate standard errors for treatment comparisons.

The author was interested in designs where there was little balance or orthogonality in the designs, but it was not clear how to extend the existing methods. The extension that was developed was built on Henderson's work on best linear unbiased prediction (BLUP).

(c) *Best linear unbiased prediction*

Some of the main points arise when considering a linear model

$$y = Xb + Zu + e, \tag{2.1}$$

with

$$\operatorname{var}(y) = ZGZ' + R = V, \quad \operatorname{var}(u) = G \quad \text{and}$$
$$\operatorname{var}(e) = R, \tag{2.2}$$

where y represents an $n \times 1$ vector of observations, X and Z are $n \times t$ and $n \times c$ matrices representing design matrices for the fixed and random effects in the $t \times 1$ and $c \times 1$ vectors b and u. The random effects vector, u, and the residual vector, e, are assumed to be multivariate normally distributed. The matrices V, G and R represent the variance matrices of y, u and e, respectively. Since this

linear model includes fixed effects, a, and random effects, b, it is called a mixed model. It has many applications. In the analysis of experiments, interest is in estimation of treatment effects, b, taking account of the variance matrix V. In some genetic applications, there is interest in the estimation of the genetic variances and covariances in G, adjusting the data for the fixed effects a. In other applications, which particularly motivated Henderson, there is interest in predicting the random effects, u.

If the variance parameters are known, then the weighted least squares estimate, $\hat{b}$, of b is given by

$$X'V^{-1}X\hat{b} = X'V^{-1}y. \tag{2.3}$$

Henderson in Henderson *et al.* (1959) showed that the inversion of V, the $n \times n$ variance matrix, can be avoided and that $\hat{b}$ satisfies

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \tag{2.4}$$

This set of equations includes terms for both fixed and random effects. Owing to the similarity of these equations to least squares equations, which are the same as (2.4) but without the G^{-1} term, these are now called Henderson's mixed model equations. Henderson pointed out that the effects $\hat{u}$ in (2.4) are not just computational artefacts arising out of the solution and he related $\hat{u}$ to Lush's (1949) 'most probable producing ability' or predicted breeding value. Henderson (1950) suggested estimating u based on maximizing the log-joint density of the data and random effects and later he considered the more general problem of predicting linear functions of the unknown b and u (Henderson 1973). The case when we wish to predict the individual random effects is of most interest. Then we wish to predict the vector u with elements u_j. Henderson suggested using a BLUP $\hat{u}$ for u where: (i) the mean square error $E(\hat{u}_j - u_j)^2$ is minimized (best), (ii) $\hat{u}$ is a linear function of y (linear), and (iii) $\hat{u}_j$ is an unbiased estimator of u_j, i.e. $E(\hat{u}_j - u_j) = 0$ (unbiased). The expectations correspond to values in the hypothetical repetitions of sampling and predicting and involve averaging over u and e. Henderson showed that this best linear unbiased predictor of u satisfies $\hat{u} = GZ'V^{-1}(y - X\hat{b})$ with b satisfying the weighted least squares equations (2.3). We note that GZ' is the covariance of u with y and V is the variance of y, so that $GZ'V^{-1}$ is informally $\operatorname{cov}(u, y)$ *divided by* $\operatorname{var}(y)$ and can be thought of as regression coefficients of y (corrected for the fixed effects) on u. Henderson (1973) also noted that if b was known, one could combine all the information available on relatives in a selection index in order to predict u. This selection index is an optimal way of selecting individuals as parents based on a linear index combining information on relatives and traits when there are no fixed effects associated with the observations. A selection index predictor of w is then $\bar{w} = GZ'V^{-1}y$. Hence the BLUP predictor is of the same form as the selection index predictor with the intuitively sensible correction of the data using the weighted least squares estimate $\hat{b}$.

(d) *Estimation of variance parameters*

The estimation of variance components in the mid-1970s mainly followed Henderson's (1953) methods that essentially constructed an analysis of variance by

sequentially fitting a sequence of linear models. Variance components were found by equating sums of squares involving random effects to their expectation. It is a confusing procedure in that a fixed effect model motivates the sums of squares and then a random effect model is used to find their expectation. Cunningham & Henderson (1968) had suggested sums of squares analogous to those of Henderson (1953), though using terms based on the mixed model equations rather than the fixed effect equations. These methods were not necessarily efficient, depending on the degree of unbalance and the relative size of the ratio of variance components. It was suspected whether that sums of squares would be more efficient.

An alternative method of estimation could be based on the method of maximum likelihood (ML), but one drawback is that ML estimation of variance parameters can lead to very biased estimates. In the simple case of a linear regression with t regressors, an unbiased estimate of residual variance is the residual sum of squares, S, divided by $n-t$ with n as the number of observations. However, the ML estimate is S/n, i.e. independent of the number of regressors, and ML ignores the estimation of fixed effects in the estimation of variance parameters. When considering the equations of Hartley & Rao (1967) for the first differentials of the log-likelihood equations, two sums of squares, one a weighted residual sum of squares and the other a weighted sum of squares of predicted values, seemed more natural terms to use. The author carried out simulations on a non-orthogonal design, suggested for rotation experiments by Patterson (1964), and expected that his estimation scheme would produce unbiased estimates. Because ML produces biased estimates, it was surprising that the efficiency of the author's estimation scheme was exactly the same as those quoted by Patterson (1964) for a ML scheme. H. D. Patterson (1969, personal communication) explained to the author that in fact he had not used the full likelihood of the data but that based on a set of error contrasts, i.e. comparisons that did not provide information on fixed effects. This method agreed with the analysis of variance methods in balanced cases and effectively eliminated bias in variance estimation due to estimation of the fixed effects. We were able to show using a spectral decomposition that there were $n-t-c$ comparisons with expectation equal to the residual variance and $c-1$ components with expectation equal to a function of the residual and block components. A weighted least squares scheme was developed essentially fitting a linear model to the independent mean squares with parameters given by the variance components with the slight complication that the mean squares had a variance that depended on their expectation. In fact this was a special case of generalized linear models (McCullagh & Nelder 1989). These models extend linear models with normally distributed data to allow both the expectation of the data to be a function of a linear predictor, and distributions that are members of the exponential family to be used. In Patterson & Thompson (1971) the connections with Henderson's mixed model equations were noted and the method extended to more general mixed models.

We called the method a modified ML method and later, in a review paper, Harville (1977) called the method a restricted maximum likelihood method. T. P. Speed (1982, personal communication) commented that a more appropriate name might be a residual maximum likelihood method and this was the name we preferred. There have been other justifications of the likelihood: Harville (1974) suggested it arose in a Bayesian analysis with vague knowledge of fixed effects; and Speed (1991) pointed out that REML was equivalent to a technique called generalized ML introduced by Wahba (1990) for the estimation of spline parameters. A convenient form of the residual log-likelihood (Harville 1974) is

$$L = -(1/2)[(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}) + \log \det(\boldsymbol{V})$$
$$+ \log \det(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})].$$

This is of the same form as the full likelihood with the addition of the last term that is sometimes thought of as a penalty for estimating the fixed effects. More recently, Lee & Nelder (1996) used Henderson's joint density in a more general setting (called the hierarchical or h-likelihood) and suggested that REML includes two penalties, effectively one for estimation of fixed effects and one for the prediction of random effects. In some ways these alternative justifications add extra strength to the method.

(e) *Computation of variance parameters*
Thompson *et al.* (2005) reviewed more recent computing developments in deriving ML estimates. Some are related to the maximization of a complex nonlinear function.

It is often useful to express relevant quantities in terms of the projection matrix

$$\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$$

so that

$$L = -(1/2)[\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y} + \log \det(\boldsymbol{V}) + \log \det(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})].$$

If the variance matrix is a function of parameters $\boldsymbol{\theta}$, then the estimation of a variance parameter θ_i involves setting to zero the first derivatives

$$\frac{\partial L}{\partial \theta_i} = \boldsymbol{y}'\boldsymbol{P}\left(\frac{\partial \boldsymbol{V}}{\partial \theta_i}\right)\boldsymbol{P}\boldsymbol{y} - \text{tr}\left[\boldsymbol{P}\left(\frac{\partial \boldsymbol{V}}{\partial \theta_i}\right)\right].$$

This could be thought of as equating a function of the data to its expectation. The finding of a maximum of the likelihood usually requires an iterative scheme. One suggested by Patterson & Thompson (1971) is based on the expected value of the second differentials, which can be constructed from

$$E\left[\frac{\partial^2 L}{\partial \theta_i \, \partial \theta_j}\right] = -(1/2)\text{tr}\left[\boldsymbol{P}\left(\frac{\partial \boldsymbol{V}}{\partial \theta_i}\right)\boldsymbol{P}\left(\frac{\partial \boldsymbol{V}}{\partial \theta_j}\right)\boldsymbol{P}\right].$$

If these second differentials of $\boldsymbol{\theta}$ are collected in a matrix $\boldsymbol{E}$, it is called the expected information matrix of $\boldsymbol{\theta}$. Then, we can update $\hat{\boldsymbol{\theta}}$ using

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \boldsymbol{E}^{-1}\left(\frac{\partial L}{\partial \boldsymbol{\theta}}\right),$$

and this expression can be computed from the solution of (2.4). The expected information matrix $\boldsymbol{E}$ gives an asymptotic variance matrix of $\hat{\boldsymbol{\theta}}$. Smith & Graser (1986) developed a convenient sequential way of calculating the data and determinant terms in the log-likelihood; and Misztal & Perez-Enciso (1993) found a convenient way of

calculating terms in the first differentials of the likelihood, taking full account of the fact that a large proportion of the coefficients in the matrices involved in the mixed model equations are zero. Efron & Hinkley (1978) suggested, however, that an observed information matrix, based on the actual second differentials, could be more appropriate than the expected information matrix for giving asymptotic variances. Gilmour *et al.* (1995) showed that the average of the observed and expected information matrices is much easier to compute than that of either the expected or observed information matrices. The coefficients of this average information matrix, F, are given by

$$F_{ij} = -(1/2)y'P\left(\frac{\partial V}{\partial \theta_i}\right)P\left(\frac{\partial V}{\partial \theta_j}\right)Py.$$

These terms can be calculated in the same way as the data term, $y'Py$ in the log-likelihood and L is calculated by replacing y with the 'working variable' $(\partial V/\partial \theta_j)Py$ in algorithms to calculate the weighted sum of squares, $y'Py$.

The use of this average information matrix has removed a computational straightjacket in the estimation of mixed models. Before its use, the estimation was usually constrained to estimating a small number of variance parameters on relatively small sets of data with a restricted set of variance structures. Now it is possible to fit realistically up to a 100 parameters on much larger sets of data with a much wider class of variance structures including multivariate models, autoregressive models, random regression models, spatial models and interactions between these models (Gilmour *et al.* 2002).

(f) Genetic models

The application of mixed models to animal breeding data has motivated several of the developments in mixed model estimation. The other key result that allowed a much wider class of models to be fitted was to allow estimation of correlated random effects. Henderson (1976) was interested in BLUP using information from all relatives. He introduced an animal model of the form of (2.1) with the element u_i representing the additive genetic effect for the ith individual. Then the variance matrix for the animal effects is $G = A\sigma_a^2$, where A represents the additive relationship matrix and σ_a^2 is the additive genetic variance. Henderson showed that it is much easier to form A^{-1} (the term used in the mixed model equations) than A. This follows from the fact that, if the individuals are listed with ancestors above descendants, A can be written as TMT' where M is a diagonal matrix and T is a lower triangular matrix with non-zero diagonal elements and i, jth elements are non-zero if the jth individual is an ancestor of the ith (Thompson 1979). The matrix T has a simple inverse with both the diagonal elements and i, jth elements being non-zero if the jth individual is a parent of the ith individual. Hence A has a simple inverse. It is interesting to note that an animal effect can be written as an accumulation of independent terms from its ancestors and itself, i.e. $u_j = \sum_i T_{ij}u_i^*$, where the independent effects u_i^*, which can be thought of as Mendelian sampling terms, have variance matrix M. Since if, for example, we have a non-inbred animal with parents and grandparents, then its genetic effect is made up of one Mendelian sampling term from itself, two Mendelian sampling terms with coefficient one-half from its parents and four Mendelian sampling terms with coefficient one-quarter from its four grandparents. In this case, the estimation procedure can be thought of as equating $\hat{u}'A^{-1}\hat{u}$ or equivalently $\hat{u}^{*\prime}M^{-1}\hat{u}^*$ to its expectation.

The advantages of using the animal model are that one can take account of all additive genetic relationships and unequal family sizes. On the other hand, one might be using a composite heritability that might be biased by non-additive variances, for example, dominance or environmental covariances associated with full sibs if these terms are not included in the model.

For computational reasons the REML methods were used initially for sire models, i.e. relating a sire effect with each observation and ignoring other relationships, particularly for dairy cattle populations where there are relatively few sires each with many progeny. As computational facilities improved it became possible to fit animal models, relating an animal effect with each individual, both for univariate and multivariate data taking account of additive relationships (Meyer 1985, 1997; Jensen *et al.* 1997). The basic framework was extended to include more biologically appropriate models with genetic components, including maternal models that introduce terms into the model for an individual that depends on the mother of the individual or her phenotype (Wilham 1963; Falconer 1965; Koerhuis & Thompson 1997). Misztal (1997) showed how dominance terms could be included in the model. More recently, there has been interest in introducing extra heritable components to model interactions, including competition effects, between individuals (Bijma *et al.* 2007). The usage of the animal model with REML was developed and followed initially for livestock populations only, but subsequently much more widely. Kruuk (2004) has shown the usage of these models in the analysis of natural populations.

Before the introduction of the animal model, parameter estimation was a three-stage procedure: adjusting data for environmental effects, estimating covariances between relatives, and interpreting them in terms of genetic parameters. Now, it consists of fitting a mixed linear model with specification of expectation and variance submodels.

3. EFFECTS OF SELECTION

Sometimes, either through design or accident, parents are chosen on their phenotypic performance. Then some of the usual methods of estimation are biased; for example, heritability if estimated by sib covariances, or genetic correlations if estimated by parent–offspring regression. Some of these difficulties are removed if ML methods are used.

Suppose we have observations on $n_1 + n_2$ female animals, y_i, $(i=1, n_1+n_2)$, of which the first n_1 animals are selected at random and used as parents, and we have n_1 offspring observations, z_i, $(i=1, n_1)$. Suppose both y_i and z_i are normally distributed with means μ_1 and μ_2, variances V_{11} and V_{22} and covariance V_{12} between y_i and z_i. Let $\bar{y}$, $\bar{y}^*$ and $\bar{z}$ be the mean values for all observations in the parental generation, parents and offspring, respectively. Then ML estimates of μ_1 and μ_2 satisfy (e.g. Curnow 1961)

$$\bar{y} = \mu_1 \tag{3.1}$$

$$\bar{z} = \mu_2 + V_{12}V_{11}^{-1}(\bar{y}^* - \mu_1). \tag{3.2}$$

An instructive form for the likelihood follows if we partition the likelihood into two independent parts. The first part is the likelihood of the parental data, denoted L_1, and the second part the likelihood of $z_i - V_{21}V_{11}^{-1}y_i$, denoted $L_{2.1}$, which can be thought of as the offspring record conditional on the parental record. So ML makes use of three pieces of information: the parental data give information on V_1, regression of z_i on y_i gives information on $V_{21}V_{11}^{-1}$ and the variance of $z_i - V_{21}V_{11}^{-1}y_i$ gives information on the conditional variance $V_{22c} = V_{22} - V_{21}V_{11}^{-1}V_{21}$. Suppose now that parents are chosen on the basis of their own phenotypic values or a function of these. Following Kempthorne and Von Krosigk (in Henderson *et al.* 1959) we can again write the log-likelihood as the sum of the log-likelihood of the parental values and the log-likelihood of offspring values given as parental values. The estimates of the fixed effects again follow (3.1) and (3.2) and have the natural interpretation whose response $(\bar{z} - \mu_2)$ equals regression coefficient $(V_{12} - V_{11}^{-1})$ times selection differential $(\bar{y}^* - \mu_1)$, and the residual log-likelihood is again of the same form. Curnow (1961) notes that the information matrix used in the calculation of the asymptotic variance matrix of the estimates needs modification. He suggests that expected values for covariances $\text{cov}(z, y)$ and $\text{var}(z)$, terms depending on the parental values, should be calculated using the observed parental values. If the effect of selection is to reduce the variance of the parental observations to a proportion k then

$$\text{cov}(z, y) = kV_{21} \quad \text{and}$$

$$\text{var}(z) = V_{22} - (1-k)V_{21}V_{11}^{-1} - V_{12}. \tag{3.3}$$

Note that if $k < 1$ the variance in the offspring generation is reduced, but REML still estimates the variance in the absence of selection, V_{22}.

An alternative and more general way of thinking about selection processes was introduced by Rubin (1976) who suggests thinking about the 'complete data', comprising observed data and missing data and introducing a missing value indicator. In our example, the observed data are y_i, $(i = 1, n_1 + n_2)$ and z_i, $(i = 1, n_1)$; the missing data are z_i, $(i = n_1 + 1, n_1 + n_2)$; and the missing value indicators are R_{y_i}, with $R_{y_i} = 1$ $(i = 1, n_1 + n_2)$, and R_{z_i}, with $R_{z_i} = 1$ $(i = 1, n_1)$ and $R_{z_i} = 0$ $(i = n_2 + 1, n_1 + n_2)$. He shows that if the missing value indicator depends only on observed data and not the missing data, then the inference can be based on the observed data. In this case, the data are said to be missing at random. This gives extra motivation to our sequential formation of likelihood and the estimates of $\boldsymbol{u}$ from the mixed model equations. Rubin (1976) also suggests methods for other cases when the missing data are 'missing completely at random' and 'not missing at random'. Hadfield (2008) has discussed the application of these ideas to the analysis of natural populations.

A logical extension of this missing-at-random paradigm is that, when we measure and make selection decisions based on several variates, a multivariate approach estimating covariances between variates will be more appropriate than a sequence of univariate analyses. More motivation for this approach is given by the comment by Robertson (1966) that in predicting response on one trait, one cannot merely multiply

heritability by selection differential to give response if other traits influence selection decisions. The corollary is that one cannot either predict unbiased breeding values or obtain unbiased estimates of variance parameters if the relevant selection traits are not included in the analysis.

Henderson (1950, 1973) had used both likelihood based and sampling arguments to motivate the development of BLUP. With random selection of parents, both arguments give the same predictors. When there is a selection of parents, then the methods do not necessarily agree. Henderson (1975), using sampling arguments with a model based on (2.1) and (2.2), suggests conditioning on contrasts $\boldsymbol{L}'\boldsymbol{y}$ with the coefficients of $\boldsymbol{L}$ depending on how selection is carried out. The author found this paradigm hard to understand (Thompson 1979), it can lead to difficulty in generating simulation samples (Schaeffer 1987) and to loss of efficiency (Im *et al.* 1989). Rubin (1976) suggests that, based on the consideration of studies with active intervention, inferences based on likelihoods are less sensitive than sampling distribution inferences to the process that cause missing data, in that likelihood methods are more likely to take account of the fact that the follow-up data might not be missing at random.

Essentially, analyses of many sets of animal breeding data were carried out to predict breeding values for animals, but it was soon recognized that use of the predicted breeding values also enabled formation of the estimates of genetic trend (Van Vleck 1977). However, it is not always obvious how to interpret these estimators. The estimates of mean genetic values in different generations certainly have an interesting structure due to the accumulation of genetic drift in each generation (Hill 1972; Thompson & Atkins 1994). In some cases, including the earlier two generations example, the predicted genetic mean value in the second generation is a function of the heritability and the selection differential and does not depend on the individual values in the second generation. Hadfield (2008) has given other instances when the use of predicted breeding values for estimating selection differentials is naive.

4. INFERRING SELECTION

In selection experiments, one has at least some control of the selection processes and it is then easier to take account of selection in the resulting analyses. In other cases, we may wish to infer something about the unknown selection processes from the observed data. A useful case to consider is if we measure q traits with genetic variance $\boldsymbol{G}$ and phenotypic variance $\boldsymbol{P}$. Then if the selection differential, the difference between the mean phenotype before and after an episode of selection, is $\boldsymbol{s}$ then the application of (3.2) (or extension of (3.2)) shows that the response to selection, $\boldsymbol{r}$, is given by $\boldsymbol{r} = \boldsymbol{G}\boldsymbol{P}^{-1}\boldsymbol{s}$ (e.g. Lande 1979), assuming that the individuals are unrelated. Animal breeding literature tends to emphasize the estimation of the genetic variances and covariances rather than the selection differentials.

An early attempt to infer culling rules was to incorporate a culling variate into modelling dairy cattle culling procedures (Robertson 1966). Meyer & Thompson (1984) extended the two-trait example to

include sire families and investigated also a threshold trait, with values 0 and 1, correlated with the first trait y. They showed that it was possible to obtain information on the relationships of the underlying trait with y but not on the relationships between the culling variable and the second trait, essentially because the culling variable was 1 whenever a value was available on the second trait. This analysis essentially treated culling as a discrete variable.

Studies in evolution biology have an emphasis different from those in animal breeding. In terms of the equation to predict response, $r = GP^{-1}s$, there is more interest in measuring the phenotypic selection gradient, $\beta = P^{-1}s$, than the response r. For example, Kingsolver *et al.* (2001) give a comprehensive review of recent estimates of selection gradients. Lande & Arnold (1983) give the underlying theory. They make extensive use of an identity introduced by Robertson (1966) and developed by Price (1970) and called by Lynch & Walsh (1998) the Robertson–Price identity. This theory shows that the vector of selection differentials of traits can be expressed in terms of the covariances between the traits and relative fitness. Then if there is a linear relationship between fitness and the traits, $P^{-1}s$ can be thought of as a vector of the (phenotypic) regression coefficients of relative fitness on the q traits which are recorded. Arguments similar to those in the formation of (3.3) can give the change in variances the following selection.

Alternatively, if we add relative fitness to the set of q traits and use the Robertson–Price identity to note that the selection differential for relative fitness is equal to the phenotypic variance of relative fitness, we find that $r = g_f$, where g_f represents the genetic covariance between the traits and relative fitness. Presumably, if fitness is linearly related to the other traits in the analysis, a mixed model analysis of the traits and fitness could give alternative estimates. Lande & Arnold (1983) also extend their methods to include a quadratic regression of traits on fitness and if the traits are normally distributed there is a natural interpretation of this extension. They also suggest canonical transformations, which essentially form linear combinations of the variates, so that the fitness surface and the changes in variance due to selection can be expressed in a simpler form. This use of canonical transformation is different from that used in animal breeding studies to simultaneously diagonalize the G and P matrices and simplify the calculation of predictions and estimation of genetic parameters (Thompson 1977; Meyer 1985).

There is an interesting contrast between the animal breeding approach with its emphasis on estimating genetic and phenotypic variances and the evolutionary biology approach with its emphasis on fitness and selection gradients. It could be useful to attempt a synthesis. It does not seem appropriate always to treat fitness as just another trait in a standard mixed model analysis of normally distributed traits.

There are several reasons. (i) The consequence of treating fitness as a trait that is totally dependent, either linearly or nonlinearly (Lande & Arnold 1983), at the phenotypic or genetic level on other measured traits has the consequence that the usual assumption of multivariate mixed models of positive definite variance matrices cannot hold, nor does it seem simple to take

account of nonlinear relationships between traits; (ii) in equilibrium there might be no additive genetic variance in fitness and again the usual assumption of a positive defined additive genetic matrix will not hold; and (iii) fitness might not be normally distributed (or not be easily transformed to be so). One suggestion for a trait that follows a distribution in the exponential family (e.g. Poisson, gamma or binomial distribution) is to use a generalized linear mixed model. There are rather obvious extensions to the generalized linear model weighted least squares equations (McCullagh & Nelder 1989) and Henderson's mixed model equations to give approximate estimators in this case (Thompson 1979; Breslow & Clayton 1993; Lee & Nelder 1996). These estimators use approximations and in the animal breeding context are thought to behave better for larger family sizes, i.e. for sire models rather than animal models (Ducrocq & Casella 1996; Moreno *et al.* 1997). The methods Lee & Nelder (1996) introduce are thought to be slightly different and they argue that these behave better for essentially smaller family sizes, but no numerical results have yet been given for animal models.

Alternatively there has been the recent development of methods based on Markov Chain Monte Carlo (MCMC) sampling. This paradigm often makes it simpler to think about general models without being constrained by the linearity of expectation or normality of distribution. They can be used to give ML estimates (Guo & Thompson 1994), but most of their usage has been in the development of Bayesian methods. Sorensen & Gianola (2002) review their application to quantitative genetics. More recent relevant papers consider multivariate analysis when the underlying traits are multivariate normal, but some measured traits may be grouped or censored (Korsgaard *et al.* 2003) and the joint analysis of survival and an underlying normally distributed trait (Damgaard & Korsgaard 2006*a*,*b*). The Bayesian approach allows the natural incorporation of 'prior' information on parameters. This prior evidence is not often well explained, however. Thompson *et al.* (2005) note that at a recent conference five-sixths of the papers using a Bayesian analysis did not quantify the prior knowledge or used vague or flat priors. Some papers at least paid lip service to the Bayesian paradigm.

5. CONCLUSIONS

One of the aims in writing this paper was to outline the development of REML and the animal model and the applications which provided the stimulation for this. Another aim was that the author had hoped to tie up what he perceived as some loose ends, which include treating fitness as another trait in a standard multivariate mixed model analysis. In fact some of them seem to have been tied up before the paper is published. In a paper in this issue, Hadfield (2008) has built on Rubin's missing data theory to extend survival analysis to show how it can be used to provide a basis for a better analysis of viability selection in natural populations.

REFERENCES

Bijma, P., Muir, W. M., Ellen, E. D., Wolf, J. B. & Van Arendonk, J. A. M. 2007 Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**, 289–299. (doi:10. 1534/genetics.106.062729)

Breslow, N. & Clayton, D. 1993 Approximate inference in generalized linear models. *J. Am. Stat. Assoc.* **88**, 9–25. (doi:10.2307/2290687)

Cunningham, E. P. & Henderson, C. R. 1968 An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics* **24**, 13–25. (doi:10.2307/2528457)

Curnow, R. N. 1961 The estimation of repeatability and heritability from records subject to culling. *Biometrics* **17**, 553–566. (doi:10.2307/2527855)

Damgaard, L. H. & Korsgaard, I. R. 2006a A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait. *Genet. Select. Evol.* **38**, 45–64. (doi:10.1051/gse:200502)

Damgaard, L. H. & Korsgaard, I. R. 2006b A bivariate quantitative genetic model for a threshold trait and a survival trait. *Genet. Select. Evol.* **38**, 565–581. (doi:10. 1051/gse:2006022)

Ducrocq, V. & Casella, G. 1996 A Bayesian analysis of mixed survival models. *Genet. Select. Evol.* **28**, 505–529. (doi:10. 1051/gse:19960604)

Efron, B. & Hinkley, D. 1978 Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65**, 457–482. (doi:10. 1093/biomet/65.3.457)

Falconer, D. S. 1965 Maternal effects and selection response. In *Proc. XI Int. Congress on Genetics*, vol. 3, pp. 763–774. Oxford, UK: Pergamon Press.

Gilmour, A. R., Thompson, R. & Cullis, B. R. 1995 Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440–1450. (doi:10.2307/2533274)

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. 2002 *ASREML user guide, release 1.0*. Hemel Hempstead, UK: VSN International.

Guo, S. W. & Thompson, E. A. 1994 Monte-Carlo estimation of mixed models for large complex pedigrees. *Biometrics* **50**, 417–432. (doi:10.2307/2533385)

Hadfield, J. D. 2008 Estimating evolutionary parameters when viability selection is operating. *Proc. R. Soc. B* **275**, 723–734. (doi:10.1098.rspb.2007.1013)

Hartley, H. O. & Rao, J. N. K. 1967 Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

Harville, D. A. 1974 Bayesian inference for variance components using only error contrasts. *Biometrika* **62**, 383–385. (doi:10.1093/biomet/61.2.383)

Harville, D. A. 1977 Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* **72**, 320–338. (doi:10.2307/2286796)

Henderson, C. R. 1950 Estimation of genetic parameters (abstract). *Ann. Math. Stat.* **21**, 309–310.

Henderson, C. R. 1953 Estimation of variance and covariance components. *Biometrics* **9**, 226–252. (doi:10. 2307/3001853)

Henderson, C. R. 1973 Sire evaluation and genetic trends. In *Proc. Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush*, pp. 10–41. Champaign, IL: American Society of Animal Science and American Dairy Science Association.

Henderson, C. R. 1975 Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447. (doi:10.2307/2529430)

Henderson, C. R. 1976 A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**, 69–83. (doi:10.2307/2529339)

Henderson, C. R., Kempthorne, O., Searle, S. R. & von Krosigk, C. M. 1959 The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218. (doi:10.2307/2527669)

Hill, W. G. 1972 Estimation of genetic change. I. General theory and design of control populations. *Anim. Breed. Abstr.* **32**, 1–15.

Im, S., Fernando, R. L. & Gianola, D. 1989 Likelihood inferences in animal breeding under selection: a missing-data theory view point. *Genet. Select. Evol.* **21**, 399–414. (doi:10.1051/gse:19890401)

Jensen, J., Mantysaari, E. A., Madsen, P. & Thompson, R. 1997 Residual maximum likelihood estimation of (Co) variance components in multivariate mixed linear models using average information. *J. Ind. Soc. Agric. Stat.* **49**, 215–236.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, N., Hill, C. E., Hoang, A., Gibert, P. & Beerli, P. 2001 The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261. (doi:10. 1086/319193)

Koerhuis, A. N. M. & Thompson, R. 1997 Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genet. Select. Evol.* **29**, 225–299. (doi:10. 1051/gse:19970208)

Korsgaard, I. R., Lund, M. S., Sorensen, D., Gianola, D., Madsen, P. & Jensen, J. 2003 Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Select. Evol.* **35**, 159–183. (doi:10.1051/gse:2003002)

Kruuk, L. E. B. 2004 Estimating genetic parameters in natural populations using the 'animal model'. *Phil. Trans. R. Soc. B* **359**, 873–890. (doi:10.1098/rstb.2003.1437)

Lande, R. 1979 Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–416. (doi:10.2307/2407630)

Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated traits. *Evolution* **37**, 1210–1226. (doi:10. 2307/2408842)

Lee, Y. & Nelder, J. A. 1996 Hierarchical generalized linear models (with discussion). *J. R. Stat. Soc. B* **58**, 619–678.

Lush, J. L. 1949 *Animal breeding plans*, 3rd edn. Ames, IA: Iowa State College Press.

Lynch, M. & Walsh, B. 1998 *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

McCullagh, P. & Nelder, J. A. 1989 *Generalized linear models*, 2nd edn. New York, NY: Chapman and Hall.

Meyer, K. 1985 Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* **41**, 153–165. (doi:10.2307/2530651)

Meyer, K. 1997 An 'average information' restricted maximum likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices. *Genet. Select. Evol.* **29**, 97–116. (doi:10.1051/gse:19970201)

Meyer, K. & Thompson, R. 1984 Bias in variance and covariance component estimation due to selection on a correlated trait. *J. Anim. Breed. Genet.* **101**, 33–50.

Misztal, I. 1997 Estimation of variance components with large scale dominance models. *J. Dairy Sci.* **80**, 965–974.

Misztal, I. & Perez-Enciso, M. 1993 Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation–maximization. *J. Dairy Sci.* **76**, 1479–1483.

Moreno, C., Sorensen, D., García-Cortés, L. A., Varona, L. & Altarriba, J. 1997 On biased inferences about variance components in the binary threshold model. *Genet. Select. Evol.* **29**, 145–160. (doi:10.1051/gse:19970203)

Nelder, J. A. 1968 The combination of information in generally balanced designs. *J. R. Stat. Soc. B* **30**, 303–311.

Patterson, H. D. 1964 Theory of cyclic rotation experiments (with discussion). *J. R. Stat. Soc. B* **26**, 1–45.

Patterson, H. D. & Thompson, R. 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554. (doi:10.1093/biomet/58.3.545)

Price, G. R. 1970 Selection and covariance. *Nature* **227**, 520–521. (doi:10.1038/227520a0)

Robertson, A. 1966 A mathematical model of culling processes in dairy cattle. *Anim. Prod.* **8**, 95–108.

Rubin, D. B. 1976 Inference and missing data. *Biometrika* **63**, 581–592. (doi:10.1093/biomet/63.3.581)

Schaeffer, L. R. 1987 Estimation of variance components under a selection model. *J. Dairy Sci.* **70**, 661–671.

Smith, S. P. & Graser, H. U. 1986 Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* **69**, 1156–1165.

Sorensen, D. A. & Gianola, D. 2002 *Likelihood, Bayesian and MCMC methods in quantitative genetics*. New York, NY: Springer.

Speed, T. P. 1991 That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* **6**, 42–44.

Thompson, R. 1976 Relationship between the cumulative difference and best linear unbiased predictor methods of evaluating bulls. *Anim. Prod.* **23**, 15–24.

Thompson, R. 1977 Estimation of quantitative genetic parameters. In *Proc. Int. Conf. on Quantitative Genetics* (eds O. Kempthorne, E. Pollock & T. Bailey), pp. 639–657. Ames, IA: Iowa State University Press.

Thompson, R. 1979 Sire evaluation. *Biometrics* **35**, 339–353. (doi:10.2307/2529955)

Thompson, R. & Atkins, K. D. 1994 Sources of information for estimating heritability from selection experiments. *Genet. Res. Camb.* **63**, 49–55.

Thompson, R., Brotherstone, S. & White, I. M. S. 2005 Estimation of quantitative genetic parameters. *Phil. Trans. R. Soc. B* **360**, 1469–1477. (doi:10.1098/rstb.2005.1676)

Van Vleck, L. D. 1977 Theoretical and actual genetic progress in dairy cattle. In *Proc. Int. Conf. on Quantitative Genetics* (eds O. Kempthorne, E. Pollock & T. Bailey), pp. 543–568. Ames, IA: Iowa State University Press.

Wahba, G. 1990 *Spline models for observational data*. Philadelphia, PA: SIAM.

Wilham, R. L. 1963 The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**, 18–27. (doi:10.2307/2527570)

Yates, F. 1940 The recovery of inter-block information in balanced incomplete block designs. *Ann. Eug.* **10**, 317–325.