Proc. Assoc. Advmt. Anim. Breed. Genet. ***17****:407–410*

SCOPE FOR ESTIMATION OF VARIANCES DUE TO SEX-LINKED, MATERNAL AND DOMINANCE EFFECTS IN MIXED MODEL ANALYSES

Karin Meyer

Animal Genetics and Breeding Unit[1], University of New England, Armidale, NSW 2351

SUMMARY

Simulation and calculations based on the likelihood function are used to examine the sampling properties of maximum likelihood estimates of variance components due to sex-linked, maternal and dominance effects. Data are assumed to have been collected in an experiment designed specifically to provide a sufficient set of covariances between relatives for this task. Results show that variances due to sex-linked and dominance effects are inherently difficult to estimate. Moreover, the design suggested appears to be unsuitable to separate maternal, permanent environmental and dominance variances. The investigations shown are recommended as part of the design stage of quantitative genetic experiments.

INTRODUCTION

Mixed model analyses fitting the animal model utilise all types of covariances between relatives in complex pedigrees simultaneously. This allows variances due to multiple genetic or environmental effects to be separated, provided records for appropriate relatives are available. Fairbairn and Roff (2006) considered a model which distinguished between autosomal, sex-linked and maternal additive genetic effects, as well as autosomal and sex-linked dominance effects, and maternal, permanent environmental effects. In addition, the authors described an experimental design to generate the types of relationships between animals needed to estimate the associated variance components. This paper examines the expected sampling properties of estimates for their design, using simulation and likelihood calculations.

MATERIAL AND METHODS

Experimental design. The design of Fairbairn and Roff (2006) involves 3 generations. considering a number of unrelated families. Generation 1 comprises 8 unrelated individuals, forming 4 pairs of grandparents. Each pair is assumed to have 4 off-spring, resulting in 16 animals in generation 2; 2 male and 2 female offspring for pairs 1 and 2 and 4 female offspring for pairs 3 and 4. Each of the 2 full-sib males (from pairs 1 and 2) is then mated to the unrelated females in the other 3 families (12 matings in total), generating sets of full- and half-sibs and single- and double-first cousins in generation 3.

Model. Let $\mathbf{y}$ denote the vector of observations for a trait of interest, and $\mathbf{a}$, $\mathbf{s}$ and $\mathbf{m}$ be the corresponding vectors of autosomal, sex-linked and maternal, additive genetic effects. Further, let $\mathbf{da}$ and $\mathbf{ds}$ be the vectors of autosomal and sex-linked dominance effects, $\mathbf{c}$ represent maternal, permanent environmental effects, and $\mathbf{e}$ the vector of residuals. This gives model of analysis

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}\left(\mathbf{a} + \mathbf{s} + \mathbf{da} + \mathbf{ds}\right) + \mathbf{W}\left(\mathbf{m} + \mathbf{c}\right) + \mathbf{e}$$

with $\boldsymbol{\beta}$ a vector of fixed effects, and $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ the design matrices for fixed, animal and maternal effects. Allowing for a direct-maternal, additive genetic covariance, σ_{AM}, gives variance of $\mathbf{y}$

$$\mathbf{V} = \mathbf{Z}\left(\sigma^2_A\mathbf{A} + \sigma^2_S\mathbf{S} + \sigma^2_{DA}\mathbf{D_A} + \sigma^2_{DS}\mathbf{D_S}\right)\mathbf{Z}' + \sigma_{AM}\left(\mathbf{ZAW}' + \mathbf{WAZ}'\right) + \mathbf{W}\left(\sigma^2_M\mathbf{A} + \sigma^2_C\mathbf{I}\right)\mathbf{W}' + \sigma^2_E\mathbf{I}$$

[1] AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

with **A**, **S**, $\mathbf{D_A}$ and $\mathbf{D_S}$ the matrices of autosomal and sex-linked genetic relationships for additive and dominance effects, respectively, **I** an identity matrix, and σ^2_A, σ^2_S, σ^2_M, σ^2_C, σ^2_{DA}, σ^2_{DS} and σ^2_E the variance components due to **a**, **s**, **m**, **c**, **da**, **ds** and **e**.

Likelihood. For N unrelated families of equal structure, $\mathbf{V} = \mathbf{I}_N \otimes \mathbf{V}_0$. In the absence of fixed effects, the log likelihood ($\log \mathcal{L}$) and its derivatives can be obtained manipulating only matrices of size proportional to the number of observations per family (e.g. Thompson 1976)

$$\log \mathcal{L} = const. - \frac{d}{2}\left(\log|\mathbf{V}_0| + \mathrm{tr}\left(\mathbf{V}_0^{-1}\mathbf{M}\right)\right) \qquad \mathbf{H} = -E\left[\frac{\partial^2 \log \mathcal{L}}{\partial\theta_k \partial\theta_m}\right] = \frac{d}{2}\,\mathrm{tr}\left(\mathbf{V}_0^{-1}\frac{\partial \mathbf{V}_0}{\partial\theta_k}\mathbf{V}_0^{-1}\frac{\partial \mathbf{V}_0}{\partial\theta_m}\right)$$

with $\mathbf{M} = \sum_{i=1}^{N} \mathbf{y}_i\mathbf{y}_i'/d$, the matrix of mean squares and cross-products, accumulated across families (where $\mathbf{y}_i$ is the sub-vector of $\mathbf{y}$ for family i), d the associated degrees of freedom, and $\boldsymbol{\theta} = \{\theta_k\}$ the vector of variance components to be estimated. $\mathbf{H}$ is the 'expected' information matrix. $\mathbf{H}^{-1}$, with elements h_{km}, gives the asymptotic, lower bound sampling covariances among the elements of $\hat{\boldsymbol{\theta}}$. Using the normal approximation, the large sample, 95% confidence interval (CI) for the $k-$th parameter is given by the interval $\left[\theta_k - 1.96\sqrt{h_{kk}}, \theta_k + 1.96\sqrt{h_{kk}}\right]$.

Profile likelihood. Partition $\boldsymbol{\theta}$ into a subset of interest, $\boldsymbol{\theta}_1$ of length p, and the remaining parameters, $\boldsymbol{\theta}_2$. The log profile likelihood $\log \mathcal{P}(\boldsymbol{\theta}_1)$ at $\boldsymbol{\theta}_1^*$ is the value of $\log \mathcal{L}$ obtained by fixing $\boldsymbol{\theta}_1$ at $\boldsymbol{\theta}_1^*$ and maximising with respect to $\boldsymbol{\theta}_2$, deviated from the maximum of $\log \mathcal{L}$. This is the quantity computed to carry out a likelihood ratio test of the hypothesis that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^*$. Asymptotically, $-2\log \mathcal{P}(\boldsymbol{\theta}_1)$ has a χ^2 distribution with p degrees of freedom. Hence, 95% confidence limits for a single parameter θ_k are given by the values for which $\log \mathcal{P}(\theta_k)$ is equal to -1.92 (Meyer and Hill 1992). These are readily determined numerically.

Simulation. The design of Fairbairn and Roff (2006) was simulated assuming that each mating in generation 2 resulted in 4 offspring, two of either sex. This gave 72 animals in each family. To include maternal effects, 8 dams of animals in generation 1, without records themselves, needed to be added to the pedigree. The simulation assumed a reasonably large experiment comprising 200 families of identical structure, i.e. 14 400 animals recorded, sampling **M** from a Wishart distribution to represent the 'data part' of $\log \mathcal{L}$. Population values assumed were $\sigma^2_A = 400$, $\sigma^2_S = 100$, $\sigma^2_M = 120$, $\sigma_{AM} = -30$, $\sigma^2_C = 150$, $\sigma^2_{DA} = 60$, $\sigma^2_{DS} = 20$ and $\sigma^2_E = 600$. Matrices **A** and **S** were obtained from the pedigree, using the tabular method (Emik and Terrill 1949; Fernando and Grossman 1990). Matrices $\mathbf{D_A}$ and $\mathbf{D_S}$ were constructed from the coefficients in the expectations of covariances between relatives given by Fairbairn and Roff (2006). Estimates were obtained using a Method of Scoring type algorithm

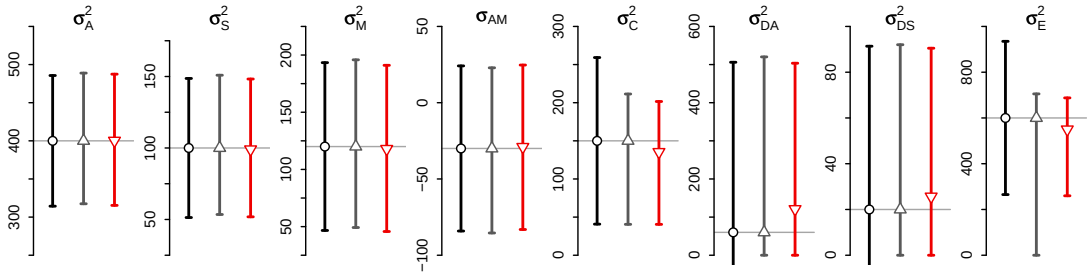

Figure 1. Means and confidence intervals (○ information, △ profile likelihood, ▽ simulation).

(Thompson 1976), together with a simple derivative-free search (Nelder and Mead 1965) to ensure convergence. Variance components were constrained to be non-negative, and the direct-maternal correlation was forced to be within the interval $[-1, 1]$. In addition to the full model with 8 covariance components, data were simulated and analysed considering subsets of effects. For each scenario considered, 50 000 replicates were carried out. Empirical 95% confidence limits were obtained as the values truncating the top and bottom 2.5% of estimates for each parameter.

RESULTS AND DISCUSSION

Figure 1 summarises likelihood derived and empirical CIs for analyses fitting all 8 covariance components. For σ^2_A, σ^2_S, σ^2_M and σ_{AM}, CIs derived from the information matrix and the profile likelihood show good agreement, with only slight asymmetry in the latter noticeable. This indicates that the shape of the profile likelihood is well approximated by the quadratic form assumed under large sample theory. Furthermore, mean estimates over replicates are close to the population values and empirical CIs deviate little from their likelihood based counterparts. While CIs are sizable, this suggests that the experimental design provides sufficient contrasts to estimate these components reliably. For the remaining 4 components, however, there are large discrepancies between the 3 measures of confidence. As shown in Figure 2, profile likelihoods for these components are asymmetrical and, especially for σ^2_{DA}, rather flat (dashed horizontal line marks -1.92). A substantial proportion (>30%) of replicates yielded estimates of σ^2_{DA} or σ^2_{DS} which were effectively zero (≤ 0.01). Consequently, mean estimates for these components are biased upwards. In turn, mean estimates for σ^2_C and σ^2_E are biased downwards. Table 1 shows sampling correlations between parameters, giving expected values from the information matrix below and observed values across replicates above the diagonal. On the whole there is good agreement between observed and expected values, with some differences attributable to constraining estimates to the parameter space.

Table 1. Sampling correlations (see text)

	σ^2_A	σ^2_S	σ^2_M	σ_{AM}	σ^2_C	σ^2_{DA}	σ^2_{DS}	σ^2_E
σ^2_A		-29	28	-63	-10	-3	6	-15
σ^2_S	-30		-9	-5	6	2	-21	1
σ^2_M	28	-9		-65	-33	-16	1	10
σ_{AM}	-63	-5	-66		16	8	2	4
σ^2_C	-6	3	-14	7		-80	-5	80
σ^2_{DA}	-4	4	-22	11	-90		-5	-96
σ^2_{DS}	8	-28	3	2	1	-12		-14
σ^2_E	-9	0	18	-3	90	-97	-6	

Clearly, there are strong associations between σ^2_C, σ^2_{DA} and σ^2_E. A substantial negative correlation between σ^2_C and σ^2_{DA} implies that we are able to estimate the sum of these components with reasonable accuracy, but that we have little information to partition them according to their causal effects. This is not surprising : only litter mates are subject to maternal, permanent environmental effects. For the design considered, all litter mates are full sibs and information to estimate dominance variances comes only from the covariances among full sibs and double first cousins. If feasible, strategies such as embryo transfer or cross-fostering may thus reduce this inherent sampling correlation. This would also reduce sampling correlations between σ^2_A, σ^2_M and σ_{AM} (Meyer 1992).

However, as shown in Figure 2, even for a 'minimum' model which does not include anything other than the respective additive genetic effects, estimates of the dominance variances are likely to have wide CIs, i.e. substantial sampling variances, unless large data sets are available to estimate them. Similarly, σ^2_S appears to be difficult to estimate accurately. For a model comprising σ^2_A, σ^2_S and σ^2_E only, the profile likelihood derived CI for σ^2_S is still as wide as $[61.6, 141.9]$, compared to $[53.5, 150.9]$ for the full model with maternal and dominance effects. For a model omitting $\sigma^2_{DA} = \sigma^2_{DS} = 0$, CIs

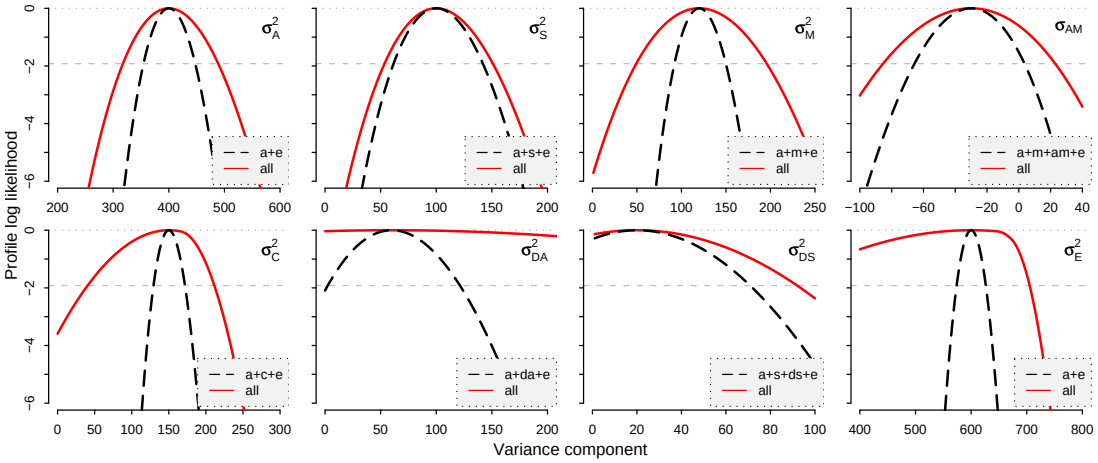

Figure 2. Profile likelihood for 'minimum' and complete models.

for σ^2_A, σ^2_S, σ^2_M and σ_{AM} were slightly reduced over those for the full model, while CIs for σ^2_E and σ^2_C were narrowed to $[555, 644]$ and $[106, 195]$, compared to $[0, 705]$ and $[41, 211]$ as shown in Figure 1. In addition, mean estimates from simulation were equal to population values, indicating that few estimates required constraining to the parameter space.

CONCLUSIONS

Separating variances between animals into causal components due to different modes of gene action is inherently difficult. The design suggested comprised sufficient types of covariances between relatives so that contrasts could be constructed to estimate all the components of interest. However, as profile likelihoods and sampling correlations derived from the inverse of the information matrix show, scope for distinguishing between variances due to maternal, permanent environmental and dominance effects is poor. The investigations shown proved illuminating, and are straightforward and computationally undemanding. They can be recommended as part of the design stage of quantitative genetic experiments.

REFERENCES

Emik, L. O. and Terrill, C. E. (1949) *J. Hered.* **40**:51.
Fairbairn, D. J. and Roff, D. A. (2006) *Heredity* **97**:319.
Fernando, R. L. and Grossman, M. (1990) *Theor. Appl. Genet.* **80**:75.
Meyer, K. (1992) *Genet. Select. Evol.* **24**:487.
Meyer, K. and Hill, W. G. (1992) *J. Anim. Breed. Genet.* **109**:264.
Nelder, J. A. and Mead, R. (1965) *Computer J.* **7**:308.
Thompson, R. (1976) *Biometrics* **32**:903.

ACKNOWLEDGEMENTS

This work was supported by Meat and Livestock Australia under grant BFGEN.100B.