BAYESIAN ESTIMATION OF VARIANCE COMPONENTS IN MIXED LINEAR MODELS

Alicia L. Carriquiry
Department of Statistics, Iowa State University
Ames, Iowa 50011, USA

SUMMARY

A Bayesian method for estimating variance components in a mixed linear model with two random vectors is presented. The procedure allows for the estimation of the ratio of variances and of any (monotone, differentiable) function of the ratio, such as heritability h^2. Prior information is incorporated into the process of inference in a general manner. The methodology can therefore be used for a wide range of prior beliefs, with no modifications. Point estimators of the variance ratio (or of h^2) can be obtained by numerical evaluation of one (or in some cases two) one–dimensional integral. The solution of the eigensystem of a certain positive semi–definite matrix simplifyes computations.

INTRODUCTION

Animal breeding data are usually assumed to follow a mixed linear model. Estimates of genetic parameters such as heritability (h^2), and predictors of breeding values, depend on the – frequently unknown – variances of the random effects in the model.

A wide array of methods in available for estimating variance components. Likelihood–based methods, and in particular, Restricted Maximum Likelihood (REML; Patterson and Thompson, 1971) have interesting frequentist properties, such as asymptotic normality and efficiency, and consistency.

An alternative are Bayesian methods for estimation of variance components. In the Bayesian framework, a (subjective) probability distribution is assigned to all the parameters in the model. From this prior distribution, and from the information provided by the data through the likelihood function, the posterior (conditional on the data) distribution of the parameters in the model can be derived. Inferences about each parameter are based on the corresponding marginal posterior distributions (Berger, 1985). The Bayesian approach allows for the formal incorporation of prior knowledge into the process of inference; in this sense, REML can be viewed as a Bayesian estimator. It corresponds to the case where "flat" (no information) prior distributions are assigned to all parameters (Harville, 1974).

Within the Bayesian framework, it is possible to obtain (via analytical or numerical integration) the marginal posterior distributions of subsets of parameters of interest. In animal breeding, for example, one is often interested in the ratio of the random effects to the residual variance. This, in general, is not possible in the classical approach.

In what follows, we present a Bayesian procedure for estimating the variance components in an unbalanced mixed linear model with two random vectors. We extend results presented by Gianola et al. (1986), Macedo and Gianola (1988), Carriquiry (1989), and Harville (1989). Following Harville (1977), the methodology takes advantage of the intimate relation that exists between the estimation of variance components and that of the fixed and random effects in the model.

THE MIXED LINEAR MODEL

Let

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{s} + \mathbf{e}, \qquad [1]$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown fixed effects, $\mathbf{X}$ and $\mathbf{Z}$ are fixed, known matrices with $\text{rank}(\mathbf{X}) = p$ (w.l.g.), $\mathbf{s}$ is a $q \times 1$ vector of unobservable random effects such that $\mathbf{s} \sim (0, \sigma^2_s\mathbf{A})$, with $\mathbf{A}$ a $q \times q$ fixed, positive definite

known matrix, and $\mathbf{e}$ is an $n \times 1$ vector of unobservable random residuals such that $\mathbf{e} \sim (0, \sigma^2_e\mathbf{I})$. Assume that σ^2_s and σ^2_e are unknown, scalar–valued parameters.

Define $\gamma = \sigma^2_s/\sigma^2_e$, $\theta = (\gamma, \sigma^2_e)'$, and assume that $\theta \in \Omega = \{(\gamma, \sigma^2_e): \sigma^2_e > 0, 0 \leq \gamma \leq u\}$, $u > 0$. Further, consider any linear combination of the elements of $\mathbf{s}$, $\varphi'\mathbf{s}$, where φ is a $q \times 1$ vector of known constants.

Under [1], $\mathbf{y} \sim (\mathbf{X}\beta, \mathbf{V})$, where $\mathbf{V} = \sigma^2_e(\mathbf{I} + \gamma\mathbf{ZAZ}')$.

The objective is to obtain an estimator for γ or for functions of γ.

THE BAYESIAN APPROACH

Prior Distribution: The first step in the Bayesian analysis is to determine the form of the prior distribution of the parameters in the model. In the present context, a prior distribution should be assigned to β and to the distributions of $\mathbf{s}$ and $\mathbf{e}$. This turns the problem into an infinite dimensional one. To avoid this, we restrict attention to the class of prior distributions for $\mathbf{s}$ and $\mathbf{e}$ that are defined by their first two moments, we fix the first moment at $\mathbf{0}$, and assign a prior to the second moments. We assume that:

1. The prior distribution of β is MVN(α, $\epsilon\mathbf{I}$), α, ϵ known, and has a p.d.f. $\tau(\beta)$.
2. β and θ are independent a priori.
3. The distribution of θ has a p.d.f. $\pi(\theta)$, and is a member of the general class of distributions with p.d.f.'s of the form

$$\pi(\theta) = G_1(\gamma)(\sigma^2_e)^{G_2(\gamma)}\exp\{-(2\sigma^2_e)^{-1}G_3(\gamma)\}, \qquad [2]$$

where $G_1(\gamma)$, $G_2(\gamma)$, and $G_3(\gamma)$ are (arbitrary) functions of γ such that $G_1(\gamma) > 0$, $G_2(\gamma) < (n–p–2)/2$, and $G_3(\gamma) \geq 0$. We assume, w.l.g., that $G_2(\gamma) = 1$. Note that the family of distributions with p.d.f.'s of the form [2] includes all the commonly used prior distributions for variance components, among them the Jeffreys, the inverted gamma, and the conjugate families.

Posterior Distribution: Inferences about θ are based on the posterior distribution of θ, $h(\theta|\mathbf{y})$. By Bayes' theorem,

$$h(\theta|\mathbf{y}) = p_1(\theta,\mathbf{y})\,[p_2(\mathbf{y})]^{-1},$$

where $p_1(\theta,\mathbf{y}) = \int\ldots\int p_3(\mathbf{y},\theta,\beta)\,d\beta$, $p_2 = \int\int p_1(\theta,\mathbf{y})\,d\theta$, and $p_3(\mathbf{y},\theta,\beta)$ is the p.d.f. of the joint distribution of $\mathbf{y}$ and all the parameters in [1]. Obtaining $h(\theta|\mathbf{y})$ in this manner requires integration in $(p + 2)$ dimensions, which is unfeasible in most animal breeding situations.

We present an alternative derivation, which takes advantage of the following assumption on the joint conditional distribution of $\mathbf{y}$ and $\varphi'\mathbf{s}$ given β and θ, and of a limit theorem regarding the prior distribution of β. Assume that, given β and θ, $\mathbf{y}$ and $\varphi'\mathbf{s}$ have a joint MVN distribution. It follows that, given θ, the joint conditional distribution of $\mathbf{y}$ and $\varphi'\mathbf{s}$ (which has a p.d.f. denoted by $g(\mathbf{y},\varphi'\mathbf{s}|\,\theta)$) is also MVN.

Further,

$$g(\mathbf{y},\varphi'\mathbf{s}|\,\theta) = g_1(\varphi'\mathbf{s}|\,\mathbf{y},\theta)\,g_2(\mathbf{y}|\,\theta),$$

where g_1 and g_2 are the p.d.f.'s of a N and a MVN distribution, respectively. It can be shown (Sallas and Harville, 1981) that

$$\lim_{\epsilon \to \infty} g_2(\mathbf{y}|\,\theta)(2\pi\epsilon)^a \propto g_3(\mathbf{z}|\,\theta),$$

where $\mathbf{z}$ is an $(n–p) \times 1$ vector of linearly independent error contrasts (as those that arise in REML), and, for $a = p / 2$, $g_3(\mathbf{z}|\,\theta)$ is the p.d.f. of a MVN distribution with mean vector 0. Harville (1989) gives a convenient expression, from a computational standpoint, of the function $g_3(\mathbf{z}|\,\theta)$. If the vector $\mathbf{z}$ of linearly independent error contrasts is taken to be the

vector of observations for estimating the variance components, then

$$h(\theta|\mathbf{z}) = g_3(\mathbf{z}|\theta)\,\pi(\theta)\,[\int\int g_3(\mathbf{z}|\theta)\pi(\theta)d\theta]^{-1}. \qquad [3]$$

For the special case where $\pi(\theta) = 1$, the value of θ which maximizes $h(\theta|\mathbf{z})$ is the REML estimate of θ.

Marginal posterior distribution of γ or of functions of γ: Inferences about γ should be based on the marginal posterior distribution of γ, where

$$h_1(\gamma|\mathbf{z}) = \int h(\theta|\mathbf{z})\, d\sigma^2_e = c_1^{-1}G_1(\gamma)\, K_1(\gamma,\Delta)\, [K_2(\gamma,\mathbf{z}) + G_3(\gamma)]^{-d},$$

and where Δ is a vector of $r < q$ vector of positive eigenvalues of a certain $q \times q$ matrix. This marginal posterior distribution depends trivially on the prior distribution of θ; different prior information can be entertained by substituting $G_1(\gamma)$ and $G_3(\gamma)$ by appropriate functional forms.

The marginal posterior distribution of h^2 (or of any other monotone, differentiable function of γ) can be obtain via an inverse transformation. Letting $h^2 = t(\gamma)$, and $\gamma = v(h^2)$,

$$f(h^2|\mathbf{z}) = h_1(v(h^2)|\mathbf{z})\ |v'(h^2)|, \qquad t(0) = 0 \le h^2 \le t(u) = 1,$$

(e.g., Hogg and Tannis, 1977).

Point Estimation of γ, or of h^2: A Bayesian estimator is optimal from a (decision theoretic) Bayesian viewpoint, if it minimizes the expected loss. It is well known that the posterior mean $E(\gamma|\mathbf{z})$ and the posterior mode minimize squared error loss and "all or none" loss, respectively. The first moment of the distribution with p.d.f. $h_1(\gamma|\mathbf{z})$ is obtained by numerically evaluating

$$E(\gamma|\mathbf{z}) = \int_0^u \gamma h_1(\gamma|\mathbf{z})d\gamma.$$

It is apparent, however, that no numerical integration is required to obtain a modal estimator.

The second moment of the posterior distribution reflects our degree of belief on the estimate obtained for the parameter. A posterior distribution with a large second moment assigns a large probability to values of the parameter that deviate greatly from the point estimate.

Since inferences are conditioned on the data, it is not necessary to invoque the concept of repeated sampling in order to assess the "goodness" of an estimator.

DISCUSSION

We have presented a Bayesian method for estimating the ratio of variances γ, or functions of γ. The estimators proposed are optimal, in the sense that they minimize the expected error loss.

The estimators presented are general, since they are given for a broad class of prior assumptions about the distributions of σ^2_e and γ. Further, the methodology allows for the existence of an upper bound on the value of γ.

The marginal posterior distribution of γ (or of h^2) can be obtained in closed form (except for the normalizing constant). The methodology can be easily extended to the case of more than two variance components, but in general, numerical integrations in several dimensions will be required to obtain the marginal posterior distributions for each ratio of variances. It is possible, under certain simplifying assumptions (Macedo and Gianola, 1988), to obtain these marginal posterior distributions analytically.

The marginal posterior distribution of γ (or of h^2) contains all of the information about the parameter available to the practitioner. Point and interval estimators, probability statements, and measures of our degree of belief in the estimates can be obtained from the marginal posterior distribution. Moreover, distributions of (monotone, differentiable) functions of γ can be derived from the posterior in a straight forward manner.

Following Harville and Fenech (1985) and Harville (1989), the eigenvalues and eigenvectors of a certain $q \times q$ positive semi–definite matrix can be used to simplify computations. In fact, once these eigenvalues and eigenvectors have been obtained, the rest of the computations turn out to be trivial. There are, for example, no matrices to be inverted. The evaluation of $h_1(\gamma|\mathbf{z})$ for values of γ may be somewhat difficult if the data is highly unbalanced, in which case the range in the elements of Δ is very large.

Bayes estimators depend not only on the information about the parameters contained in the data, but also on prior knowledge. Therefore, it is to be expected that the Bayesian methodology will do better than the classical procedures when the data contain little information about the parameters of interest.

The methodology presented was developed, at first, for the estimation of linear combinations of the elements of β and $\mathbf{s}$, and later extended to the variance component case (Carriquiry, 1989).

REFERENCES

Berger, J. O. 1985. Statistical Decision Theory and Bayesian Analysis. Springer–Verlag.

Carriquiry, A. L. 1989. PhD Dissertation. Departments of Statistics and Animal Science, Iowa State University, Ames.

Gianola, D. J. L. Foulley, and R. L. Fernando. 1986. Genet. Sel. Evol., 18:485–498.

Harville, D. A. 1974. Biometrika 61:383–385.

Harville, D. A. 1977. J Am. Stat. Assoc. 72:320–340.

Harville, D. A. 1989. in: D. Gianola and K. Hammond, eds. Advances in Statistical Methods for Genetic Improvement of Livestock. Springer–Verlag.

Harville, D. A., and A. P. Fenech. 1985. Biometrics 41:137–152.

Hogg, R. V., and E. A. Tannis. 1977. Probability and Statistical Inference. Macmillan.

Macedo, F. W., and D. Gianola. 1988. Proc. XXXVIII An. Meet. Europ. Assoc. An. Prod. 113–129.

Patterson, H. D., and R. Thompson. 1971. Biometrika 58:545–554.

Sallas, W. M., and D. A. Harville. 1981. J. Amer. Stat. Assoc. 76:860–869.