Comment

Duane Steffey and Robert E. Kass

Dr. Robinson's well-written article provides a variety of perspectives on the general linear model and its applications. Particularly welcome are the examples of Section 6 illustrating the widespread utility of these models. We limit discussion here to three main points. First, further details are provided on approximate Bayesian methods for inference about unit-specific parameters ("random" effects). Next, we amplify Dr. Robinson's comment on the often close agreement between Bayesian and frequentist inferences. Specifically, we give an approximation to the variance of the marginal posterior distribution of a unit-specific parameter and conjecture that the expression may also be justified on frequentist grounds as an approximation to the sampling variance of the BLUP estimator. Finally, we discuss the desirability of using all relevant information and mention some possible mechanisms for incorporating prior knowledge about animal breeding into the general linear model.

1. APPROXIMATE BAYESIAN INFERENCE

We here consider the marginal posterior distribution of a unit-specific parameter and provide a rather general variance approximation that satisfies the often-identified need (e.g., noted by Robinson in Section 5.6) to account for the uncertainty in estimating the common dispersion parameters. We begin by rewriting Robinson's model (1.1). Switching to a formulation similar to that of Laird and Ware (1982), we consider the general linear model in which there are k experimental units and, for the ith unit,

$$Y_i = X_i\beta + Z_i u_i + e_i.$$

Here, Y_i is an $n_i \times 1$ vector of responses, β is a $p \times 1$ vector of unknown population parameters and X_i is a known $n_i \times p$ matrix linking β to Y_i. In addition, u_i is a $q \times 1$ vector of unknown individual effects, Z_i is a known $n_i \times q$ matrix linking u_i to Y_i and e_i is an $n_i \times 1$ vector of random errors. The vectors e_i, $i = 1, \ldots, k$ are assumed to be independent and normally distributed with $E(e_i) = 0$ and $VAR(e_i) = R_i(\theta)$. The vectors u_i are taken to be independent (of each other and of the e_i) and normally distributed with $E(u_i) = 0$ and $VAR(u_i) = G(\theta)$. (We suppress the vector θ in the subsequent discussion.) That is, the model has the structure

$$\text{(1)}\qquad \begin{aligned} Y_i \mid \beta, u_i, \theta &\sim \text{Normal}(X_i\beta + Z_i u_i, R_i) \\ u_i \mid \theta &\sim \text{Normal}(0, G), \end{aligned}$$

so that given β and θ the vector pairs (Y_i, u_i) are conditionally independent for $i = 1, \ldots, k$. Kass and Steffey (1989) refer to models characterized by this structure as *conditionally independent hierarchical models* (CIHMs).

For example, in the context of Robinson's animal breeding problem (Section 1), Y_i is the vector of first lactation yields for dairy cows from the ith sire. In this case, $k = 4$, $p = 3$, and $q = 1$. Letting $i = 4$ identify Sire D, we have $n_4 = 5$ and

$$y_4 = \begin{pmatrix} 100 \\ 100 \\ 100 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} u_4 + \begin{pmatrix} e_{4(1)} \\ e_{4(2)} \\ e_{4(3)} \\ e_{4(4)} \\ e_{4(5)} \end{pmatrix}.$$

If an improper uniform prior is specified for β and integration with respect to β is performed, the posterior distribution of u_i given y_i and θ is Normal with

$$E(u_i \mid y_i, \theta) = GZ_i^T P_i y_i$$

$$VAR(u_i \mid y_i, \theta) = (G^{-1} + Z_i^T S_i Z_i)^{-1},$$

where

$$P_i = V_i^{-1} - V_i^{-1} X_i \left(\sum_{j=1}^{k} X_j^T V_j X_j \right)^{-1} X_i^T V_i^{-1}$$

$$S_i = R_i^{-1} - R_i^{-1} X_i \left(\sum_{j=1}^{k} X_j^T R_j^{-1} X_j \right)^{-1} X_i^T R_i^{-1}$$

with $V_i = R_i + Z_i G Z_i^T$. Approximations to the posterior mean and variance of u_i given $y =$

Duane Steffey is Assistant Professor, Department of Mathematical Sciences, San Diego State University, San Diego, California 92182. Robert E. Kass is Associate Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

$(y_1, \ldots, y_k)$ are obtained by applying the results of Kass and Steffey (1989):

$$(2) \qquad E(u_i \mid y) \doteq E(u_i \mid y_i, \tilde{\theta})$$

$$(3) \quad VAR(u_i \mid y) \doteq VAR(u_i \mid y_i, \tilde{\theta}) + \sum_{j,k} \tilde{\sigma}_{jk} \tilde{\delta}_j \tilde{\delta}_k,$$

where $\tilde{\sigma}_{jk}$ is the (j, k)-component of

$$\tilde{\Sigma} = \left[-D^2 \log L(\theta) \pi(\theta) \mid_{\theta = \tilde{\theta}} \right]^{-1}$$

and

$$\tilde{\delta}_j = \frac{\partial}{\partial \theta_j} E(u_i \mid y_i, \theta) \mid_{\theta = \tilde{\theta}}.$$

Here, $L(\theta)$ denotes the likelihood function, $\pi(\theta)$ denotes the prior for θ, and D^2 denotes second-order partial differentiation with respect to the elements of θ. (For specific formulas, see Kass and Steffey, 1989; Harville, 1977.) Note the convenient interpretation of (3): the first term is the conditional posterior variance with θ set equal to $\tilde{\theta}$, and the second term accounts for the additional uncertainty in estimating θ.

2. APPROXIMATE VERSUS EMPIRICAL BAYES

The approximations (2) and (3) have an alternative frequentist interpretation. The approximate posterior mean (2) is commonly recognized as the parametric empirical Bayes (PEB) estimator. In addition, we believe that under appropriate regularity conditions, the approximate posterior variance (3) may also be justified fairly generally as a frequentist approximation to the sampling variance of the PEB estimator. In the context of the general linear model (1), when $\pi(\theta) = 1$ the expression (3) is very similar to Kackar and Harville's (1984) approximation to the mean squared error of the PEB estimator, differing only in that Kackar and Harville use estimated expected information rather than observed information. Specifically, with $G(y_i, \theta) = E(u_i \mid y_i, \theta)$, we conjecture that under suitable conditions

$$(4) \quad \begin{aligned} &VAR[G(Y_i, \hat{\theta})] \\ &\quad = \hat{V}\{1 + O_p(n_i^{-1/2})\}\{1 + O_p(k^{-1/2})\}, \end{aligned}$$

where $\hat{\theta}$ is the MLE of θ, $\hat{V}$ is the approximate posterior variance given in (3) and our O_p statements refer to the two-stage model—i.e., the joint distribution of $(Y_1, u_1), \ldots, (Y_k, u_k)$ for some fixed $\theta \in \Theta$.

To indicate heuristically why we make this conjecture, we write $V1 = VAR[E(G(Y_i, \hat{\theta}) \mid Y_i = y_i)]$ and $V2 = E[VAR(G(Y_i, \hat{\theta}) \mid Y_i = y_i)]$, and use the conditional variance formula

$$VAR[G(Y_i, \hat{\theta})] = V1 + V2.$$

That the second term of (3) estimates $V2$ can be readily seen by carrying out a Taylor series expansion and standard delta method arguments. The conditional posterior variance $VAR(u_i \mid y_i, \hat{\theta})$ can then be shown to estimate $V1$ by linking three steps: (i) $VAR(u_i \mid y_i, \theta)$ is approximately equal to the variance of $\hat{u}_i$ conditional on u_i, while (ii) the latter is approximately equal to the variance of the conditional posterior expectation $VAR[G(Y_i, \theta)]$; finally, (iii) this quantity $VAR[G(Y_i, \theta)]$ can be shown to estimate $V1$. Substituting $\hat{\theta}$ for θ then yields the desired result. These arguments are most easily made by taking advantage of the asymptotic independence of Y_i and $\hat{\theta}$ as $k \to \infty$. In words, the influence of any single observation on the estimation of θ becomes negligible as the number of experimental units increases.

Note that the conjecture (4) applies to all CIHMs, of which the linear model (1) is a special case. Hence, the approximation (3) with a vague prior such as $\pi(\theta) = 1$ may find usage among frequentist statisticians as a means of overcoming the mathematical complications typically encountered in estimating the precision of BLUP estimators and, more generally, of PEB estimators in hierarchical models. For related recent work on PEB inference, see Laird and Louis (1987), Carlin and Gelfand (1990), and Hill (1990).

3. PRIOR INFORMATION AND ANIMAL BREEDING

We have undertaken only a cursory review of the animal breeding literature, but we suspect that, after generations of animal breeding experiments, considerable knowledge has been accumulated about the distributions of herd and sire effects. Does the author concur in this view?

Statistical models that incorporate available prior information will typically yield inferences about quantities of interest that are more accurate than those obtained from models that ignore relevant information. Gianola and Fernando (1986) discuss Bayesian methods for estimating breeding value (the u vector) and genetic parameters (the θ vector). They note that prior information is often available and that its use can preclude anomalies such as nonpositive definite estimated covariance matrices and "ridiculous estimates of heritability" (page 219). Also, they explain how the Bayesian approach provides a logical framework for handling problems such as those involving sequential experiments and those in which only indirect data (from relatives) are available in predicting an individual's breeding value.

Robinson notes (Section 4.2) that the BLUP estimates may be viewed as approximate Bayes estimates with improper uniform priors on β and θ. However, the results using proper prior distributions (even only mildly informative ones) that reflect pre-experiment knowledge can be substantially different from those obtained using flat priors; for instance, see Example 2 in Kass and Steffey (1989). For further details of Bayesian analyses with informative priors for β and θ in the linear model (1), see Gianola and Fernando (1986), Broemeling (1985) and the references contained therein.

The difficulty in finding ways to incorporate prior information has led many applied statisticians to question the practical value of Bayesian methods. Establishing sensible methodologies has been a goal of continuing research by statistical scientists, cognitive psychologists and econometricians. While translating uncertainty into probability distributions can be challenging, the potential scientific rewards for doing so can be substantial. Some authors have advocated generating probability distributions from statements made by substantive experts as a mechanism for incorporating prior information. For example, Kadane, Dickey, Winkler, Smith and Peters (1980) present a method for specifying a conjugate prior for (β, σ^2) in the normal linear regression model. That procedure is based on collecting responses from substantive experts (nonstatisticians) to questions about the predictive distribution of the response vector given values of the predictor variables. Such a procedure may be adaptable to the mixed effects models considered here.

Along with Kadane (1990) we would emphasize the need for more work on elicitation and would add that the need is especially great when elicitation is taken, in its broadest sense, to refer to the general process of constructing probability distributions from available background information.

We look forward to future modeling efforts that tap all sources of relevant information in order to improve inferences in the statistical problems encountered in animal breeding and many other fields of science.

Comment

Robin Thompson

Dr. Robinson's paper is valuable as it shows the links of BLUP, suggested for animal breeding applications, with methods used in other areas. An alternative way of thinking about the models used is in terms of an expectation and a variance for y that leads to a natural interpretation for prediction in terms of regressing future observations on present observations. I wonder why Dr. Robinson did not use such a formulation. With regard to making inferences on random estimates, can Dr. Robinson say if it is sensible to use the suggestion of most likely unobservables to construct confidence intervals for random estimates? I would also like to know which likelihood to use when testing fixed effects.

Robin Thompson is Head of the Biometrical Genetics Department, Edinburgh Research Station, Institute of Animal Physiology and Genetics Research, Roslin, Midlothian, EH25 9PS, Scotland.