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

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

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.