4

THE MIXED MODEL

INTRODUCTION

The mixed model is more general than the fixed because in addition to containing fixed effects the model contains so called random effects. The random factors of the model are unobservable random variables and these have variances called variance components which are the primary parameters of interest. In the traditional analysis, the fixed effects are parameters but are not regarded as random variables.

Fixed, mixed, and random are terms which came from the sampling theory framework and are not applicable to the Bayesian approach; nevertheless we will continue to use them, but within a Bayesian context.

The fixed model is the general linear model $Y = X\theta + e$ of Chapter 1, where one's main interest lies with the posterior distribution of θ and not with the parameters per se of the posterior distribution of θ.

With a mixed model

$$y = X_\theta + \sum_{1}^{c} u_i b_i + e, \tag{4.1}$$

the random factors $b_1, b_2, \ldots, b_c$ are independent random vectors with zero means and dispersion matrices $\sigma_i^2 I_{m_i}$, and one's main interest is with the variance components σ_i^2, $i = 1, 2, \ldots, c$, not the posterior distribution of the random factors. If θ is a vector of fixed effects, one is also interested in the posterior distribution of θ, not with the parameter of the posterior distribution, although these are indispensable for describing the posterior of 0. The usual assumption about the random factors is that they are independent and normally distributed, namely $N[0, \tau_i^{-1} I_{m_i}]$, where $\tau_i = \sigma_1^{-2}$, which is equivalent to a prior assumption about the random factor parameters. Now, since the variance components are parameters of the random factors, one must introduce a prior distribution for their analysis.

Note there are two levels of parameters. The primary level consists of θ, the vector of fixed effects, the random factors $b_1, \ldots, b_c$, and the error precision $\tau = \sigma^{-2}$, where the random error term $e \sim N[0, \tau^{-1} I_n]$, and is independent of the random factors. The secondary level consists of the vector of precision components, $\rho = (\tau_1, \tau_2, \ldots, \tau_c)'$. Thus, the precision components (or variance components) are parameters of some of the primary level parameters of the model.

For a complete description of the model we let X be a full rank $n \times p$ matrix, and let the u_i be $n \times m_i$ known design matrices.

There are $p + m + c + 1$ parameters, with $p + m + 1$ primary level parameters, and c secondary level parameters. The random factors are usually regarded as nuisance parameters, thus there are $p + c + 1$ parameters of interest.

Most of the work concerning random ($p = 1$) and mixed models has centered on estimating the variance components τ_i^{-1} and until 1967 the methodology was to equate the analysis of variance sum of squares to their expectations and solve the resulting system of linear equations for estimates of the variance components. For example, in a one-way random model, the between and within mean squares are equated to their expectations, giving analysis of variance estimates of the between and within components. See Searle (1971), pp. 385–389 for an example. The original methodology was introduced by Daniels (1939) and Winsor and Clarke (1940), and the sampling properties of the estimators studied by Graybill (1954), Graybill and Wortham (1956), and Graybill and Hultquist (1961).

For unbalanced data, Henderson (1953) introduced three analysis of variance methods of estimating variance components, and the properties of these estimators were examined by Searle (1971). The next new development was maximum likelihood estimation given by Hartley and Rao in 1967 and since then there have appeared many new methodologies including restricted maximum likelihood, minimum norm quadratic unbiased estimation or MINQUE, iterative MINIQUE, MIVQUE, or minimum variance quadratic unbiased estimation. These and other methods of estimating variance components are described by Searle (1978).

Box and Tiao (1973) is the only Bayesian book dealing with the variance components of random and mixed models. They give a very thorough treatment of the subject and the methodology is based on a numerical determination of the one- and two-dimensional marginal posterior distribution of the variance components. Thus with the one-way random model, which has two variance components, the one two-dimensional and two one-dimensional marginal posterior distributions are given for two data sets of Chapter 4. One interesting feature of their analysis is the choice of prior distribution. They show a Jeffreys' type vague non-informative prior on the variance components is equivalent to a vague prior on the expected mean squares of the analysis of variance.

They continue the study of variance components by examining the two-fold nested random model, the two-way random, and the two-way mixed models. Other studies from a Bayesian approach have been taken by Hill (1965, 1967), who studied the one-way model, and Stone and Springer (1965) who criticize the Box and Tiao choice of prior distribution.

Regardless of one's approach, it is safe to say that to estimate variance components is indeed a difficult task. There are so many sampling methods, it is difficult to declare any one as superior to the others, while with the existing Bayesian techniques one must rely on multi-dimensional numerical integrations. I suspect, variance components, except for the error variance, are difficult to estimate because they are secondary level parameters (at least more difficult to estimate than primary level parameters). This suspicion is motivated by Goel and DeGroot (1980), who assess the amount of information of hyperparameters in hierarchical models.

Inference about parameters of a mixed model will be accomplished by first determining the one-dimensional marginal posterior distribution of the variance components and the marginal posterior distribution of the fixed effects θ, then finding the joint posterior mode of all the parameters in the model.

THE PRIOR ANALYSIS

THE PRIOR ANALYSIS

To recapitulate, consider the model (4.1), where y is a $n \times 1$ vector of observations, X an $n \times p$ known full-rank design matrix, u_i a $n \times m_i$ known design matrix, θ a $p \times 1$ real unknown parameter vector, b_i a $m_i \times 1$ random real parameter vector, and e a $n \times 1$ normal vector with mean θ and precision τI_n. Furthermore assume $b_1, b_2, \ldots, b_c$, and e are independent.

Thus given θ, the b_i, and $\tau(> 0)$, y is normal with mean $X\theta + ub$ and precision τI_n.

The prior information is introduced in two stages: First, the conditional prior distribution of b_i given τ_i. is $b_i \sim N(0,\tau_i{}^{-1}I_{m_i}), i = 1, 2, \ldots, c$, which is the usual assumption about the random factors, θ has a marginal prior density which is constant over R^P, and τ is gamma with positive parameters a and β. Also assume θ, b, and τ are independent, a priori. Secondly, assume the secondary parameters $\tau_1, \tau_2, \ldots, \tau_c$ are independent and that τ_i is gamma with positive parameters α_i and β_i.

Why this choice of prior distribution? As will be seen, this form of the prior density, namely,

$$p(\theta,b,\tau|\rho) \propto \prod_{i=1}^{c} \tau_i^{m_i/2} e^{-\tau_i b_i' b_i/2} \tau^{\alpha-1} e^{-\tau\beta}, \tag{4.2}$$

$$p(\rho) \propto \prod_{i=1}^{c} \tau_i^{\alpha_i-1} e^{-\tau_i\beta_i}$$

for $\theta \in R^P$, $\tau > 0$, $b_i \in R^{m_i}$, and $\tau_i > 0$, produces an analytically tractable posterior distribution for the parameters. If one believes this prior is not flexible enough to express one's prior opinion about θ, ρ, and τ, one may use mixtures of these distributions which will allow more flexibility. A normal prior distribution could have been used with θ, but this yields messy posterior distributions. Or we could give θ and τ a normal-gamma distribution, b and ρ a normal-gamma, and let θ and b be independent. This would give messy but tractable posterior distributions. Note that (4.2) is an improper prior density.

The likelihood function or probability density of y given θ, b, and τ is

$$L(\theta, b, \tau) \propto \tau^{n/2} \exp -\frac{\tau}{2}(y - x_\theta - ub)\prime(y - x_\theta - ub) \tag{4.3}$$

for $\theta \in R^P$, $b \in R^m$, and $\tau > 0$, where $m = \sum_1^c m_i$.

This will be combined with (4.2), by Bayes' theorem, to give the posterior distribution of all the parameters.

THE POSTERIOR ANALYSIS

In this section, joint and marginal posterior distributions for the parameters will be derived, thereby giving us the foundation for making inferences about the parameters of the model.

First, the joint distribution of all the parameters is found by combining the likelihood function with the two-stage prior density, then the conditional posterior distributions of θ, b, τ, and ρ, given the other parameters is found. Second, the conditional distributions of θ given b, τ given b, and ρ given b are derived and lastly, the marginal posterior density of ρ and τ is found.

Combining the likelihood function (4.3) with the prior density (4.2), we have

$$\begin{aligned} p(\theta,b,\tau,\rho|y) \propto\ & \tau^{(n+2\alpha)/2-1} \exp -\frac{\tau}{2}\Big\{2\beta + y'Ry - \hat{b}'u'Ru\hat{b} \\ & + (b-\hat{b})\prime u\prime Ru(b-\hat{b}) \\ & + (\theta-\hat{\theta})\prime X\prime X(\theta-\hat{\theta})\Big\} \prod_1^c \tau_i^{((m_i+2\alpha_i)/2)-1} \\ & \times \exp -\frac{\tau_i}{2}(2\beta_i + b'_i b_i), \end{aligned} \tag{4.4}$$

where $\theta \in R^P$, $b \in R^m$, $\tau > 0$, and $\tau_i > 0$, as the joint posterior density of all the parameters. The various quantities in this expression are $\hat{b} = (u'Ru)^{-}u'Ry$, $R = I - X(X'X)^{-1}$, and $\hat{\theta} = (X'X)^{-1}X'(y - ub)$, where A is the unique Moore-Penrose generalized inverse of the matrix A.

An equivalent representation of this density is

$$\begin{aligned} p(\theta, b, \tau,\rho|y) \propto\ & \tau^{(n+2\alpha)/2-1} \exp -\frac{\tau}{2}\{2\beta + (y - X_\theta - ub)\prime \\ & \times (y - X_\theta - ub)] \prod_1^c \tau_i{}^{(m_i+2\alpha_i)/2-1} \\ & \times \exp -\frac{\tau_i}{2}(b_i' b_i + 2\beta_i) \end{aligned} \tag{4.5}$$

where $\theta \in R^P$, $\tau > 0$, $\tau_i > 0$, and $b \in R^m$. This form of the density allows one to readily deduce the many conditional posterior distributions of the parameters. Thus we have

Theorem 4.1. The conditional posterior distribution of θ given b, τ, and ρ is normal with mean $\hat{\theta}$ and precision matrix $\tau X'X$. The conditional posterior distribution of b given θ, τ, and ρ is normal with mean $[\tau u'u + A(\rho)]^{-1} \tau u'(y - X\theta)$ and precision matrix $\tau u'u + A(\rho)$, where $A(\rho)$ is the $m \times m$ block diagonal matrix with i-th diagonal matrix $\tau_i I_{m_i}$, $i = 1, 2, \ldots, c$. The conditional posterior distribution of τ given θ, b, and ρ is gamma with parameters (n +

diagonal matrix with i th diagonal matrix $\tau_i I_{m_i}$, i = 1, 2, …,c. The conditional posterior distribution of τ given θ, b, and ρ is gamma with parameters (n + 2α)/2 and [2β + (y − Xθ − ub)′(y − Xθ − ub)]/2.

Finally, the conditional distribution of ρ given θ, b, and τ is such that $\tau_1, \tau_2, \ldots, \tau_c$ are independent and τ_i is gamma with parameters $(m_i + 2\alpha_i)/2$ and $(2\beta_i + b'_i b_i)/2$, i = 1, 2, …, c.

Thus the conditional distribution of each parameter given the other is a well-known density and, as will be shown, provides a convenient way to estimate the mode of the joint posterior density.

Of course, our goal is to determine the marginal posterior distribution of θ, τ, and ρ, regarding b as a nuisance parameter, but first consider the conditional distributions of θ given b and (τ, ρ) given b.

Theorem 4.2. The conditional posterior distribution of θ conditional on b is a multivariate t with mean vector $\hat{\theta}$ and precision matrix

$$T = \frac{(n-p+2\alpha)X'X}{2\beta + y'Ry - \hat{b}'u'Ru\hat{b} + (b-\hat{b})'u'Ru(b-\hat{b})}. \tag{4.6}$$

The conditional distribution of τ conditional on b is gamma with parameters (n − p + 2α)/2 and $[(2\beta + y'Ry - \hat{b}'u'uRu\hat{b} + (b-\hat{b})'u'Ru(b-\hat{b})]/2$. Also, conditional on b; $\tau, \tau_1, \tau_2 \ldots, \tau_c$ are independent and τ_i is gamma with parameters $(m_i + 2\alpha_i)/2$ and $(2\beta_i + b'_i b_i)/2$, i = 1, 2, …, c.

We will use the conditional distributions of the previous theorem in order to find the marginal posterior distributions of θ, τ, and ρ, and to do this, we need, also, the marginal posterior distribution of b, which is given by

Theorem 4.3. The marginal posterior density of b, the vector of random effects, is

$$\begin{aligned} p(b|y) &\propto \left[1 + \left(b - \hat{b}\right)' A \left(b - \hat{b}\right)\right]^{-(n-p+2\alpha)/2} \\ &\quad \times \prod_1^c \left(1 + b'_i A_i b_i\right)^{-(m_i+2\alpha_i)/2}, \end{aligned} \tag{4.7}$$

where $b \in R^m$,

$$A = \frac{u'Ru}{2\beta + y'Ru - \hat{b}' u'Ru\hat{b}},$$

and

$$A_i = (2\beta_i)^{-1} I_{m_i}, \quad i = 1, 2, \ldots, c.$$