$\mu+bi+\varepsilon ij$; $i=1,2,\ldots,m$; $j=1,2,\ldots,t$,

where b_i is normal with mean 0 and variance σ_1^2, and the ε_{ij} follow a normal distribution with zero mean and error variance σ^2 for all i and j. This is the mixed model (4.1) with c = 1, m = m, p = 1, and n = mt. Also

$$
\begin{aligned}
&y = (Y_{11},\ldots,y_{1t}; y_{21},\ldots,y_{2t};\ldots;y_{m1},\ldots,y_{mt})',\\
&x = j_n, \text{a } n\times 1 \text{ vector of ones},\\
&\theta = \mu, \text{a real scalar},\\
&U = \text{Diagonal } (j_t, j_t, \ldots, j_t), \text{of order } n\times m, \text{and}\\
&b = (b_1, b_2, \ldots, b_m)'.
\end{aligned}
$$

For this particular case, it can be shown that

$$
\begin{aligned}
&R = I_n - (1/n)J_n, J_n \text{ a } n\times n \text{ matrix of ones},\\
&u'Ru = t[I_m - (1/m)J_m],\\
&(u'Ru)^- = (1/t)[I_m - (1/m)J_m],\\
&\hat{b} = (y_1 - y, y_2 - y, \ldots, y_m - y)',\\
&\hat{b}' u'R\hat{b} = s_1 \text{ the between group sum of squares},\\
&A_1 = (n - m + 2\alpha - 3)(2\beta + s_2)^{-1}[tI_m - (t/m)J_m],
\end{aligned}
$$

where s_2 is the within group sum of squares,

$$
\begin{aligned}
&A_2 = [(\alpha_1 - 1)/\beta_1]I_m,\\
&A^* = A_1 + A_2 = (at + g)I_m - (at/m)J_m, \text{ where}\\
&a = (n - m + 2\alpha - 3)(2\beta + s_2)^{-1}, \text{ and}\\
&g = (\alpha_1 - 1)/\beta_1.
\end{aligned}
$$

Using the above results, one may show that

$$
\begin{aligned}
&(A^*)^{-1} = (at + g)^{-1}I_m + at(gm)^{-1}(at + g)^{-1}J_m,\\
&b^* = at(at + g)^{-1}\hat{b},\\
&u'Ru(A^*)^{-1} = t(at+g)^{-1}[I_m - (1/m)J_m],\\
&b^{*\prime}b^* = a^2t(at+g)^{-2}s_1,\\
&(b^* - \hat{b})'u'Ru(b^* - \hat{b}) = g^2(at + g)^{-2}s_1.
\end{aligned}
$$

The conditional Bayes estimators conditional on b* are

$$
\begin{aligned}
&\tilde{\alpha}^2 = (n + 2\alpha - 3)^{-1}[2\beta + s_2 + g^2 s_1(at + g)^{-2}],\\
&\tilde{\alpha}_1^2 = (m + 2\alpha_1 - 2)^{-1}[2\beta_1 + a^2ts_1(at + g)^{-2}].
\end{aligned}
\tag{4.18}
$$

Approximate Bayes estimators, based on the approximations to the marginal posterior distribution of the parameters are given by (4.14) and (4.16) and reduce to the estimators

$$
\begin{aligned}
&\tilde{\alpha}^2 = (n + 2\alpha - 3)^{-1}[2\beta + s_2 + t(m - 1)(at + g)^{-1} + g^2s_1(at + g)^{-2}],\\
&\tilde{\alpha}_1^2 = (m + 2\alpha_1 - 2)^{-1}[2\beta_1 + (at + mg)g^{-1}(at + g)^{-1} + a^2ts_1(at + g)^{-2}].
\end{aligned}
\tag{4.19}
$$

The approximate Bayes estimators (means of posterior densities) and the conditional Bayes estimators (means of the conditional posterior distribution) are very similar. For example, we see they each have two terms in common, only differing in the third term $t(m - 1).(at + g)^{-1}$ for the error variance and $(at + mg)g^{-1}(at + g)^{-1}$ for the between variance component.

The usual AOV estimators, see Box and Tiao (1973), are $\hat{\sigma}^2 = s_2t^{-1}(m - 1)^{-1}$ and $\hat{\sigma}_1^2 = [s_2t^{-1}(m - 1)^{-1} - s_1(m - 1)^{-1}](t - 1)$ and are quite different than the conditional Bayes and approximate Bayes estimators, which are not linear functions of the AOV sum of squares (as are the AOV estimators), but instead are ratios of integer powers of the AOV sum of squares. For this model, the AOV estimators are minimum variance quadratic unbiased estimators (but are negative with positive probability), see Graybill and Wortham (1956), thus it would be interesting to compare the sampling mean square errors of Bayes and AOV estimators.

We now continue with our example of the one-way random model by plotting the posterior density of the variance components and calculating the posterior mean and variance of these parameters for a set of data employed by Box and Tiao (1973). For the generated data the true value of σ^2 is 16 and is 4 for the between variance component σ_1^2.

Many different sets of values of the hyperparameters (α, β, α_1, β_1) were taken so as to examine the effects of the prior parameter values on the two marginal posterior distributions and to see the effect on the closeness of the approximation.

For each set of hyperparameters, the posterior means and variances of the variance components were calculated in two ways. First from (4.14)–(4.15)

For each set of hyperparameters, the posterior means and variances of the variance components were calculated in two ways. First from (4.14), (4.15), (4.16), (4.17), the approximate moments were computed, then the true moments, mean and variance, were calculated from the exact posterior densities via numerical integration. The eigenvalues and other constants needed to evaluate the approximate densities (4.12) and (4.13) and their moments, (4.14), (4.15), (4.16), (4.17), were computed with the MATRIX procedure in SAS (Statistical Analysis System). The true posterior distributions are given by (4.8), and the relevant marginal densities and their moments were evaluated using numerical integrations. This was done with a FORTRAN program (WATFIV). To evaluate the approximate densities (4.12) and (4.13) only 20 terms in the mixture were taken, because the contribution from the remaining terms was negligible. The constant c was chosen to be the smallest eigenvalue in order to make (4.12) and (4.13) mixtures. The calculations were done on an IBM 370/168 at Oklahoma State University by M. Rajagopalan (1980).

Table 4.1 provides one with the approximate and true values of the posterior means and variances of the variance components for various values of the hyperparameters.

The exact and approximate marginal posterior densities of the two parameters are plotted in Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6 for six sets of hyperparameters.

Results of the numerical study indicate the approximations are close, generally, to the true means and variance.

The closeness of the approximations increases with α, even for values as low as 8, and is independent of the β value. That this must be so follows from the fact that as the α parameter increases, the degrees of freedom of each factor of the poly-t density increase, thus each factor is well approximated by the normal density, and the posterior density of b is well approximated by a normal.

This insures the accuracy of the approximation, however a large a parameter also insures a small prior standard deviation for the parameter. For example, the prior standard deviation of σ^2 is $\beta(\alpha-1)^{-1}(\alpha-2)^{-1/2}$, thus the approximate posterior distribution of σ^2 is accurate if one has a strong informative prior for the variance component. When precise prior information is available the approximations are efficient. The accuracy of the approximations is confounded with one's prior beliefs.

Also, when the α-parameter is large and the prior means of the components are close to their true values, the posterior means are close to the true means. Thus in such cases, the two posterior distributions are centered very close to the true value of the variance components and the posterior variances are small.

Furthermore, when α is large, changes in the α parameter do not significantly affect the posterior distribution of the within component as compared to the between component. Informative priors influence the between component relatively more than the within. Hence, unless one is quite sure of what one is doing, then the value of the α parameter for the between variance component should not be selected as large even though it does not matter much for the within variance component. The reasons are explained by referring to the posterior density of b, (4.7), where in the first factor the degrees of freedom, which introduces prior information about a (or τ), is $(n - p + 2\alpha + m)$ which is likely to be large for most data, regardless of the value of α. Variations in the α parameter do not affect this factor, but on the other hand, the second part, consisting of c factors, of the density of b, and the degrees of freedom of the i-th factor, is only m_i, and variations in the value of α_i significantly affect the posterior distribution of the between component.

Lastly, when there is little prior information, it seems reasonable to let the value of the α parameter be close to 2 so the prior variance is large. Even when α is close to 2, the approximations are close, but whether the posterior distributions of the variance components are close to their true values depends on the β parameter. From Table 4.1, one can see that when α = 2 and β varies from 2 to 20, the posterior mean of σ^2 goes from 13.06 to 14.82, a small change. Note the true value of $\sigma^2 = 16$. We conclude, therefore, that the posterior distribution of the within variance component is not very sensitive to changes in β, for values of α close to 2. This is not the case with the between component since when α is close to 2, the posterior mean and variance heavily depend on the β parameter. Thus a fairly good estimate of the true value is needed for a proper choice of the β parameter.

Table 4.1
Mean and Variance of the True and the Approximate Marginal Posterior Distributions of the Within and Between Variance Components for Box's Data

Prior parameters				Within variance component				Between variance component			
				mean		variance		mean		variance	
α	β	α_1	β_1	true	app	true	app	true	app	true	app
2	2	2	2	12.96	13.06	11.72	12.24	1.22	1.62	0.84	1.07
2	5	2	5	13.14	13.56	11.15	13.67	2.40	3.01	1.97	2.66
2	8	2	8	13.49	13.81	12.84	14.22	3.29	3.69	2.63	4.02
2	20	2	20	14.59	14.82	16.16	17.02	7.44	8.97	16.87	24.40
2	8	5	5	13.35	13.39	12.30	12.68	1.11	1.17	0.30	0.30
3	3	3	3	12.21	12.28	9.92	10.13	1.17	1.33	0.56	0.60
3	5	3	5	12.32	12.50	9.63	10.73	1.79	2.05	1.07	1.23
3	10	3	10	12.75	12.97	10.96	11.67	3.07	3.36	2.12	2.61
3	20	3	20	13.56	13.86	12.92	13.58	5.77	6.50	9.37	11.90
3	50	3	50	15.80	16.18	17.86	19.08	12.07	12.23	24.22	26.73
3	3	5	5	12.21	12.26	9.91	10.07	1.12	1.16	0.30	0.30

4	5	4	5	11.59	11.69	8.24	8.74	1.38	1.50	0.56	0.56
4	10	4	10	11.96	12.08	9.23	9.40	2.54	2.77	1.72	2.00
4	20	4	20	12.68	12.85	10.52	10.77	4.64	5.03	5.15	6.08
4	50	4	50	14.76	14.85	14.53	14.81	10.25	10.70	18.29	20.76
5	5	5	5	10.95	11.00	7.23	7.24	1.11	1.17	0.30	0.30
5	10	5	10	11.26	11.34	7.67	7.79	2.07	2.20	0.95	1.03
5	20	5	20	11.91	12.04	8.70	8.89	3.85	4.08	2.98	3.35
5	50	5	50	13.85	13.98	11.98	12.15	8.76	9.13	13.86	14.37
8	8	8	8	9.48	9.53	4.98	4.65	1.08	1.09	0.16	0.16
10	10	20	10	8.79	8.81	3.55	3.57	0.52	0.52	0.01	0.01
10	40	20	50	10.10	10.14	4.77	4.82	2.51	2.52	0.32	0.32
10	100	20	100	12.86	12.94	7.55	7.87	4.90	4.89	1.10	1.25
10	200	20	200	17.56	17.31	14.62	14.90	9.70	9.74	4.65	4.71
20	100	50	100	8.82	8.84	2.49	2.50	2.01	2.02	0.08	0.06
20	400	50	200	18.14	18.02	10.58	11.26	4.00	4.00	0.43	0.33
32	500	20	80	15.49	15.51	5.52	5.55	3.98	3.96	0.80	0.79

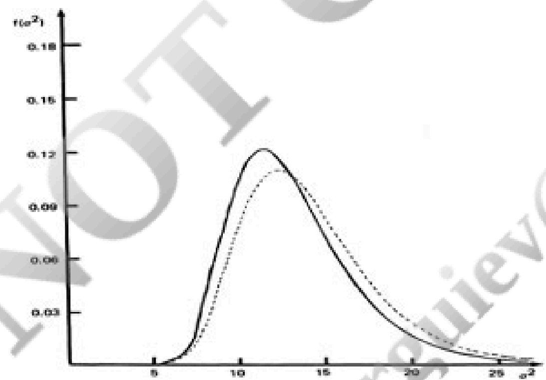

Figure 4.1. Marginal posterior density of the within-variance component; $\alpha = 2$, $\beta = 5$, $\alpha_1 = 2$, $\beta_1 = 5$; ---- = approximation, — = true distribution.

These conclusions are based on a limited study of one data set. More extensive numerical studies need to be done before one can formulate general rules on the closeness of the approximations and the choice of the hyperparameters.

Overall, one may say that the results of the numerical study indicate the approximations are good. Generally, the accuracy of the approximations increases with the a parameter (α or α_1) of the prior distribution of the variance components and the approximations are close for values of α as small as 8. The posterior distribution of σ^2 is less sensitive to variations in the β parameters of the gamma priors as compared to that of the between variance component. When precise prior information is available, it doesn't seem unreasonable to set the α parameter close to 2. Generally, in order to fix the β parameters, a good prior estimate of the true value of the variance components is needed. But, more detailed numerical studies are needed before one can put forward these conclusions as general rules.

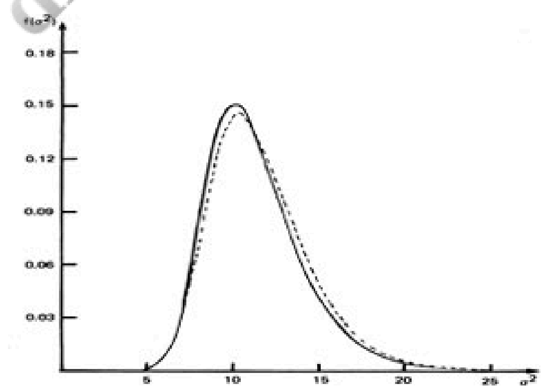

Figure 4.2. Marginal posterior density of the within-variance component; $\alpha = 5$, $\beta = 10$, $\alpha_1 = 5$, $\beta_1 = 10$; ---- = approximation, — = true distribution.

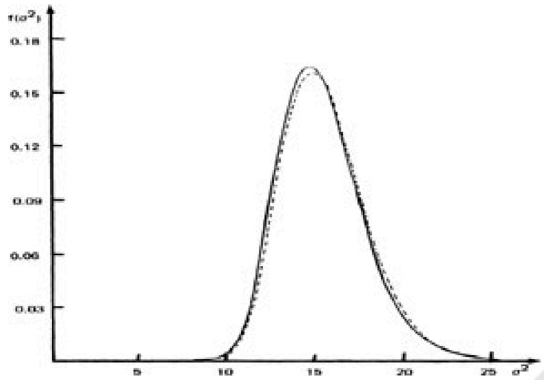

Figure 4.3. Marginal posterior density of the within-variance component; $\alpha = 32$, $\beta = 500$, $\alpha_1 = 20$, $\beta_1 = 80$; ---- = approximation, — = true distribution.

Joint Modal Estimators of Parameters of Mixed Models

Another way to estimate the parameters of mixed linear models is with the mode of the joint posterior distribution of all the parameters. This approach is somewhat similar to the method of maximum likelihood estimation, which was first developed by Hartley and Rao (1968). The mode of the joint posterior distribution is found in the same way as done by Lindley and Smith (1972), who used iterative procedures to find modal estimates of the parameters. These modal estimators were viewed as large-sample approximations to posterior means.

The view taken here is to regard joint modal estimators as “natural” Bayesian estimators, since they correspond to points of highest posterior density, and to use the posterior mean if a quadratic loss function is appropriate.

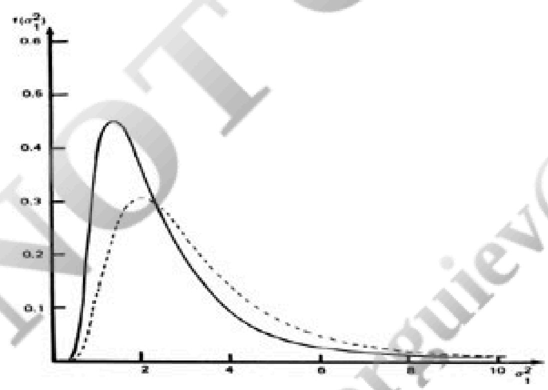

Figure 4.4. Marginal posterior density of the between-variance component; $\alpha = 2$, $\beta = 5$, $\alpha_1 = 2$, $\beta_1 = 5$; ---- = approximation, — = true distribution.

Of course, a complete Bayesian analysis dictates that one give the marginal posterior distributions of all the parameters, thus for a model with k parameters, there are $2^k - 1$ marginal distributions to determine. If one considers a one-way model, it has three parameters, excluding the random factor, one should specify seven marginal distributions, something which has not been done.

Instead, one is forced to use summary characteristics of the posterior distributions such as posterior means, modes, medians, variances and so on of the one- and two-dimensional marginal distributions of the two parameters σ^2 and σ_1^2. This approach usually involves numerical integration of functions of several variables.

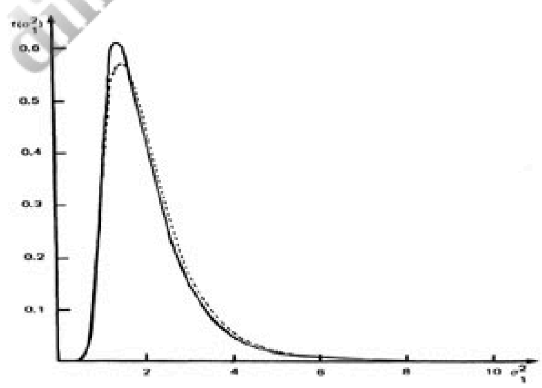

Figure 4.5. Marginal posterior density of the between-variance component; $\alpha=5$, $\beta = 10$, $\alpha_1 = 5$, $\beta_1 = 10$; ---- = approximation, — = true distribution.

In a related study, Rajagopalan (1980) presents a way to compute all one-dimensional marginal densities of the parameters of mixed linear models and this method was presented in the previous sections of the chapter. One advantage of his approach is that it avoids numerical integration of functions of several variables.

Lindley and Smith (1972) and later Smith (1973) propose using the mode of the joint posterior density of the parameters in the model. They employed their method with a two-factor model with two variance components, the error variance and the row and column effects. They were primarily interested in estimating the main effect factors, but in principle, their method will work with any subset of parameters.

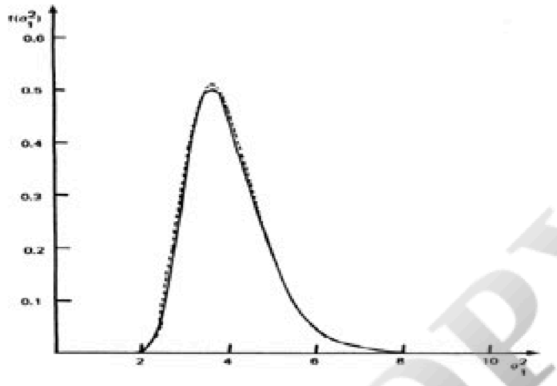

Figure 4.6. Marginal posterior density of the between-variance component; $\alpha = 32$, $\beta = 500$, $\alpha_1 = 20$, $\beta_1 = 80$; ---- = approximation, — = true distribution.

Consider the joint posterior density (4.5) of all the parameters of a mixed linear model. Theorem 4.1 gives the conditional posterior distributions of θ, b, τ, and ρ, given the other parameters of the model, thus all the conditional modes are known, namely

$$M(\theta|b, \tau, \rho) = (X'X)^{-1}X'(y - ub) \tag{4.20}$$