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

is the conditional posterior mode of θ given b, τ, and ρ. Also,

$$M(b|\theta,\tau,\rho) = [\tau u'u + A(\rho)]^{-1}\tau u'(y - X\theta), \tag{4.21}$$

$$M(\tau^{-1}|\theta,b,\rho) = \frac{2\beta + (y - X\theta - ub)'(y - X\theta - ub)}{n + 2\alpha + 2} \tag{4.22}$$

and

$$M(\tau_i^{-1}|\theta,b,\tau) = \frac{2\beta_i + b_i'b_i}{m_i + 2\alpha_i + 2} \tag{4.23}$$

are the conditional modes of b, τ^{-1}, and τ_i^{-1}, i = 1, 2, …,c, given the other parameters. The last two modes refer to the error variance τ^{-1} and the variance components τ_i^{-1} and the four equations for the conditional modes will be used to find a mode of the joint posterior density, (4.5), of all the parameters θ, b, τ, and ρ.

Suppose a mode of the joint density is given by a solution to the equations

$$\frac{\partial}{\partial\theta}p(\theta,b,\tau,\rho|y) = 0, \tag{4.24}$$

$$\frac{\partial}{\partial b}p(\theta,b,\tau,\rho|y) = 0, \tag{4.25}$$

$$\frac{\partial}{\partial\tau}p(\theta,b,\tau,\rho|y) = 0, \tag{4.26}$$

and

$$\frac{\partial}{\partial\tau_i}p(\theta,b,\tau,\rho|y) = 0, \tag{4.27}$$

where i = 1, 2, …,c.

Now consider (4.24), then it is known that the joint density factors as

$$p(\theta,b,\tau,\rho|y) = p_1(\theta|b,\tau,\rho,y)p_2(b,\tau,\rho,|y),$$

where p_1 is the conditional density of θ given b, τ, and ρ and p_2 the marginal density of b, τ, and ρ. If $p_2 > 0$ for $b \in R^m$, $\tau > 0$, and all $\tau_i > 0$, the first equation (4.24) is equivalent to

$$\frac{\partial}{\partial\theta}p_1(\theta|b,\tau,\rho,y) = 0, \tag{4.28}$$

and in a similar fashion

$$\frac{\partial}{\partial b}p_3(b|\theta,\tau,\rho,y) = 0, \tag{4.29}$$

$$\frac{\partial}{\partial\tau}p_5(\tau|b,\theta,\rho,y) = 0, \tag{4.30}$$

and

$$\frac{\partial}{\partial\tau_i}p_7(\tau_i|\rho,\theta,b,y) = 0, \quad i = 1,2,\ldots,c \tag{4.31}$$

where p_3, p_5, and p_7 are the conditional densities of b, τ, and τ_i given the other parameters, and this system of four equations is equivalent to the system (4.24), (4.25), (4.26), (4.27).

Since each conditional density is unimodal for all values of the conditioning variables, this system is equivalent to the four modal equations (4.20), (4.21), (4.22), (4.23). In order to find the joint mode, this system must be solved simultaneously for θ, b, τ, and τ_i, i = 1,2, …, c, over the space $b \in R^m$, $\theta \in R^p$, $\tau > 0$, and $\tau_i > 0$, i = 1, 2, … c.

Starting with some initial value for b, one may find a θ value from the first modal equation (4.20), then a τ^{-1} value from (4.22), and finally c τ_i^{-1} values from (4.23), then the cycle is repeated, and if the process converges, the solution is a solution to the system (4.24), (4.25), (4.26), (4.27), and a mode of the joint posterior density is obtained.

Let us see how this method works when applied to some simple mixed models. First, consider the one-way random balanced model of the previous section, and let the initial value of b be the least squares estimator of b, namely ${}_0b = (\bar{y}_1 - \bar{y}, \bar{y}_2 - \bar{y}, \ldots, \bar{y}_m - \bar{y})'$ then the first modal equation is ${}_0\mu = \bar{y}$ where θ = μ. Further substitution gives where θ = μ. Further substitution gives

$$ {}_0\tau^{-1} = \frac{2\beta+\sum_1^m\sum_1^t(y_{ij}-\bar{y}_i)^2}{mt+2\alpha+2} $$

and

$$ {}_0\tau_1^{-1} = \frac{2\beta_1+\sum_1^m(\bar{y}_i-\bar{y})^2}{m+2\alpha_1+2} $$

as the first stage estimates of the error variance and between variance component, respectively. Gharraf (1979) shows that ${}_0b$, ${}_0\mu$, ${}_0\tau$, and ${}_0\tau_1^{-1}$ are conditional Bayes (conditional posterior modes) estimators of b, μ, τ^{-1}, and τ_1^{-1}, respectively.

The cycle is repeated using ${}_1b = m(b \mid {}_0\mu, {}_0\tau^{-1}, {}_0\tau_1^{-1})$ and continued until the value of the estimates stabilizes. Note the first stage estimators depend on the familiar analysis of variance sum of squares.

Next consider a two-fold nested random balanced model

$$ y_{ijk} = \theta + a_i + b_{ij} + \varepsilon_{ijk}, \tag{4.32} $$

where i = 1, 2, …, a; j = 1, 2, …, g; k = 1, 2, …, d. In terms of the general mixed model, n = agd, c = 2 (two variance components), m_1 = a and m_2 = g. The observations are $\{y_{ijk}\}$, θ is a real unknown scalar, the sets $\{a_i\}$, $\{b_{ij}\}$, and $(\varepsilon_{ijk}\}$ are independent and the a_i are independent and $n(0, \sigma_i^2)$, the b_{ij} are independent and $n(0, \sigma_2^2)$, and the ε_{ijk} are independent and $n(0, \sigma^2)$. If the starting value of b is ${}_0b = (j'Ru)^- u'Ry$, the modal equations reduce to

$$ {}_0\theta = \bar{y}, $$

$$ {}_0\sigma^2 = \frac{2\beta+\sum_1^a\sum_1^g\sum_1^d(y_{ijk}-\bar{y})2}{agd+2\alpha+2} $$

$$ {}_0\sigma_1^2 \frac{2\beta_1+g^2\sum_1^a(\bar{y}_{ijk}-\bar{y})^2/(g+1)^2}{a+2\alpha_1+2} $$

and

$$ {}_0\sigma_2^2 \frac{2\beta_2+\sum_1^a\sum_1^g[\bar{y}_{ijk}-g\bar{y}_i/(g+1)^2-\bar{y}/(g+1)]^2}{g+2\alpha_2+2} $$

for the first stage estimates of the parameters. The cycle is repeated beginning with a new value of b, namely

$$ {}_1b = [{}_0\tau u'y + A({}_0\rho)]^{-1}{}_0\tau u'(y - x_0\theta], $$

where ${}_0\rho = ({}_0\tau_1, {}_0\tau_2)$ and $\tau_i^{-1} = \sigma_i^{-2}$, i = 1, 2. This model is given as an example by Box and Tiao (1973) to illustrate the Bayesian approach, by Hartley and Rao (1968), who use maximum likelihood estimation, and by Gharraf (1979), who showed the first stage estimators were conditional Bayes estimators.

The starting value of b was the least squares estimator of b, however other reasonable estimators of b will suffice.

There are some problems with this procedure. First, when does it converge and if it converges does it converge to the joint posterior mode? Also if it converges, how does the rate of convergence depend on the starting value? There are also some other computational problems with inverting large matrices. The generalized Moore-Penrose inverse of u′Ru and the inverse of τu′u + A(ρ) are of order m × m, the total number of random effects of the model. This problem also occurs with maximum likelihood estimation, thus perhaps the W transform introduced by Hemmerle and Lorens (1976) is a way to calculate inverses of patterned matrices such as u'Ru.

This section is concluded with numerical examples for the oneway and twofold random models. Again, the data set generated by Box and Tiao (1973) is used as in the previous section. With regard to the first mode the parameters were set at τ^{-1} = 16, τ_1^{-1} = 4, and θ = 5, and there are six groups, m_1 = 6, and five observations per group. The analysis here uses gamma prior densities for σ^2 and σ_1^2, which differs from Box and Tiao (1973), who place an improper vague prior on the analysis of variance expected mean squares. See pp. 246–276 of Box and Tiao (1973) for a detailed account. The analysis of variance estimators is $\widehat{\sigma^2}$ = 14.495 and $\widehat{\sigma_1^2}$ = 1.3219, and the joint posterior mode of (σ^2, σ_1^2) is (13.796,0).

Table 4.2 gives a solution to the modal equations (4.20), (4.21), (4.22), (4.23) corresponding to a one-way model using different values for the hyperparameters α, α_1, β, and β_1, and in all cases the starting value of b was the least squares estimator. The random effect estimates are not shown since b is a 6 × 1 vector and the other parameters are our primary concern. In all cases $\hat{\theta}$ is 5.666 because the conditional mode of θ is $\bar{y}$ and is not affected by prior information.

Table 4.2 demonstrates the effect of the prior distribution on the modal estimates. For example, the first row uses the “true” value of the variance components as the prior mean of these parameters, and the estimates of the within and between components are 11.4861 and .713785, respectively. On the other hand, when the prior means are set at 10, the modal estimates are 11.7734 and .0200043.

Table 4.2
Modal Estimates of Parameters: The One-Way Model

α	α_1	β	β_1	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\sigma}_1^2$

2	2	4	4	5.6656	11.4861	.713785
2	2	2	2	5.6656	10.9041	.349768
2	2	10	10	5.6656	10.8208	1.8334
2	2	20	20	5.6656	12.9161	3.59343
5	5	4	4	5.6656	9.93803	.462915
5	5	2	2	5.6656	9.41301	.228348
5	5	50	20	5.6656	11.1399	2.35272
5	5	10	10	5.6656	9.3689	1.1868
10	10	16	4	5.6656	8.08161	.293158
10	10	2	2	5.6656	7.63936	.145248
10	10	15	20	5.6656	9.05468	1.49166
10	10	100	20	5.6656	11.0234	1.47858
10	10	10	10	5.6656	7.63557	.748002
1.1	1.1	1	1	5.6656	11.5485	.20321
1.01	1.01	.1	.1	5.6656	11.7507	.0200421
1.001	1.001	.01	.01	5.6656	11.7734	.0200043

The modal estimate of θ is not sensitive to prior information, but the modal estimates of the within and between components are sensitive; however, the within estimate is less sensitive than the between estimate. For instance, the ratio of the largest to smallest of the within estimates is only 1.5419 while the ratio is 179.6328 for the between estimates.

It is difficult to state any definitive conclusions about these estimates and a more thorough sensitivity analysis needs to be done. The estimates are closest to the true value of the parameters when the prior means are set at 15 and 20 and in all cases the estimates are smaller than the true values of the parameters.

The next example is the twofold nested model, (4.32), and is discussed by Box and Tiao (1973) beginning on page 282. This model has four parameters, the fixed effect θ, the error variance σ^2, and two variance components, σ_1^2 and σ_2^2. The authors generated the data and set the parameter values at θ = 0, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 4.0$, and $\sigma^2 = 2.25$. Also a = 10 and g = d = 2 and the analysis of variance estimators are $\hat{\sigma}_1^2 = 1.04$, $\hat{\sigma}_2^2 = 5.62$, and $\hat{\sigma}^2 = 3.60$. Assuming a vague prior distribution on the expected mean squares of the analysis of variance, Box and Tiao numerically determine the marginal distribution of the variance components.

Suppose one uses independent gamma densities for the precision components $\tau_1 (= \sigma_1^{-2})$ τ_2, and the error precision τ $(= \sigma^{-2})$ with hyperparameters α_1, β_1, α_2, β_2, α, and β and a constant prior density for the fixed effect. For various values of the hyperparameters the model equations were solved and the solution listed in Table 4.3, where in all cases the starting value of b is the least squares estimator (u'Ru) u'Ry.

The conditional mode of θ is $\bar{y}$ and is not affected by prior information. The estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ of the two variance components are more sensitive to prior information than the error variance estimate $\hat{\sigma}^2$. For example, the ratios of the largest to smallest estimates are 5.86, 5703.23, and 14.22 for $\hat{\sigma}^2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ respectively. Definitive conclusions are difficult to state on the basis of such a limited study.

In this section, a mode of the posterior marginal distribution of all the parameters is found for a general mixed model. The estimates are solutions to the modal equations based on the conditional mode of each set of parameters, conditional on the remaining.

In all cases the iterative scheme rapidly converged and was illustrated with two random models. The estimates of the variance components were more sensitive to prior information than the error variance, and estimates of the fixed effects were not sensitive to prior information.

The program of the iterative procedure was written in SAS.

Table 4.3
Modal Estimates for Twofold Nested Model

$\hat{\sigma}_2^2$							
	9019	6223	9854	6469	9156	5159	4959

α_1	α_2	β_1	β_2	α	β	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\sigma}_1^2$	$\hat{c}$
2	2	10	10	2	10	.488375	.94567	2.6596	3.09
2	2	4	2.25	2	1	.488375	.510797	.681372	4.86
2	2	.1	.1	2	.1	.488375	.466851	.0125578	.49
1	1	2	2	1	2	.488375	.578268	.32538	6.26
1.001	1.001	.01	.01	1.001	.01	.488375	.483015	.00142911	7.09
2	2	50	50	2	50	.488375	2.73923	8.15055	5.65
1	1	10	10	1	10	.488375	.984649	3.24959	3.24

SUMMARY AND CONCLUSIONS

This chapter has introduced a Bayesian analysis of the mixed model. Generally two methodologies were given. First, the one-dimensional marginal distributions of all the variance components and of the error variance are derived and are based on a normal approximation to the marginal posterior density of the random factors. This method applies to any mixed model, be it unbalanced or otherwise, and general formulas for the posterior means and variances are derived.

variances are derived.

A numerical study of the one-way random model reveals the accuracy of the approximation and the sensitivity of the marginal posterior densities to the parameters of the prior density.

Second, modal estimates of all the parameters may be found by solving a set of modal equations based on the conditional posterior distributions given in Theorem 4.1. The sensitivity of the modal estimates to prior information was investigated for the one-way and twofold nested models.

The advantage of the first method is that one may plot each marginal posterior density, thus, giving us at a glance (see Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6) inferences about the parameters based on all the information in the sample. But, a major disadvantage is that this method is based on an approximation to the marginal posterior densities.

With regard to the second method, estimates of all the parameters, including the random and fixed effects, are easily found, however the posterior precision of the parameters is not available.

These two methodologies should both be done in a Bayesian analysis and are not competing methods, but should augment each other. The second is easier to program than the first.

As we have seen, making inferences for the parameters of a mixed model presents many challenges. With regard to the variance components, further research, perhaps, will give us an accurate approximation to the marginal distribution of b, in which case, the Bayesian solution will be complete.