The Canadian Journal of Statistics
Vol. 19, No. 4, 1991, Pages 399–405
La Revue Canadienne de Statistique

Inference for nonconjugate Bayesian Models using the Gibbs sampler

Bradley P. CARLIN and Nicholas G. POLSON

University of Minnesota and University of Chicago

Key words and phrases: Bayesian model choice, Gibbs sampler, nonlinear models, nonnormal errors.
AMS 1985 subject classifications: 62F15, 62F99.

ABSTRACT

A Bayesian approach to modeling a rich class of nonconjugate problems is presented. An adaptive Monte Carlo integration technique known as the Gibbs sampler is proposed as a mechanism for implementing a conceptually and computationally simple solution in such a framework. The result is a general strategy for obtaining marginal posterior densities under changing specification of the model error densities and related prior densities. We illustrate the approach in a nonlinear regression setting, comparing the merits of three candidate error distributions.

RÉSUMÉ

Une approche bayesienne est présentée afin d'aborder une classe étendue de problèmes du type non conjugué. Une technique adaptative d'intégration Monte Carlo, l'échantillonneur de Gibbs, est proposée; elle apporte une solution simple tant du point de vue conceptuel que du point de vue calculatoire. Le résultat procure une stratégie générale dans le but d'obtenir des densités a posteriori marginales lorsque les lois des erreurs, et les lois a priori correspondantes, sont altérées. L'approche est illustrée dans un contexte de régression non linéaire où les mérites de trois lois pour les erreurs sont comparés.

1. INTRODUCTION

Consider the common statistical model $y_j = \mu_j + \epsilon_j$, $j = 1, \ldots, n$, where the μ_j are unknown mean structures (perhaps depending on the value of a covariate x_j or a parameter vector $\boldsymbol{\theta}$, common to all n observations), and the ϵ_j are independent random errors having density f with mean 0. Let $\mathbf{y} = (y_1, \ldots, y_n)$ denote the collection of data. A Bayesian analysis might assume an i.i.d. prior specification for the μ_j conditional on some parameter η, say $\mu_j = g(\eta) + \tau_j$, where $g(\eta)$ is a prior mean structure and the τ_j are random errors distributed independently of the ϵ_j according to a density π having mean 0. A common difficulty faced by practicing Bayesians is the appropriate specification of the model and prior error densities, f and π. For mathematical convenience, one might proceed by selecting a prior π which is *conjugate* with f, i.e., one which leads to posterior distributions $p(\mu_j|\mathbf{y})$ whose normalized forms are analytically available. Unfortunately, this computationally simple approach is often not possible, since the existence of a conjugate prior is not guaranteed. Further, even when a conjugate prior is available, it is often unrealistic (for example, imposing a normal prior on an already tenuous normal-likelihood assumption). At the very least, we would be interested in assessing the effect on the posterior distribution of departing from such an artificial assumption. On the

other hand, realistic choices for f and π typically lead to arduous, high-dimensional, and occasionally intractable numerical integration in order to obtain $p(\mu_j|\mathbf{y})$ for any j.

In practice, we would like a formal method for assessing the relative merits of two or more competing specifications for f (model choice) or π (prior robustness). The Bayesian decision paradigm has long offered an attractive theoretical setting for such a comparison (see for example Box and Tiao 1973, Smith 1983, and Berger 1985), but the aforementioned computational concerns have traditionally hampered its applicability. Recently, however, an adaptive Monte Carlo integration scheme known as the Gibbs sampler (Gelfand and Smith 1990) has proved to be a simple yet powerful tool in analyzing conjugate Bayesian hierarchical models (see for example Gelfand *et al.* 1990). In this paper we extend this applicability to nonconjugate scenarios via a highly parametrized model which exploits the simplicity of the Gibbs sampler in the conjugate case.

Section 2 discusses our Bayesian model specification, and provides a general paradigm for the Bayesian modeling of nonnormal errors. The methodology is developed with our computational tool, the Gibbs sampler, in mind. In particular, we focus on the wide range of error densities available as scale mixtures of normal distributions. Finally, Section 3 illustrates our approach by considering a nonlinear response surface, and offers a numerical example where we compare normal, t, and double-exponential errors.

2. BAYESIAN MODEL SPECIFICATION AND THE GIBBS SAMPLER

Consider the specification of the error density f (specification of the prior π follows analogously). Assume we have a series of independent errors $\epsilon_j|\sigma,\lambda_j \sim \mathsf{N}(0,\lambda_j\sigma^2)$, $j = 1,\ldots,n$. Then placing a prior on λ_j enables a wide variety of model error densities $f(\epsilon_j|\sigma)$ to emerge as scale mixtures of normal distributions (Andrews and Mallows 1974). That is, $f(\epsilon_j|\sigma) = \int_\Lambda p(\epsilon_j|\sigma,\lambda_j)p(\lambda_j)\,d\lambda_j$, $j = 1,\ldots,n$. In particular, the following list identifies the necessary functional forms for $p(\lambda_j)$ to obtain a wide range of densities which represent departures from normality:

t-family errors: If $\nu/\lambda_j \sim \chi^2_\nu$ then $\epsilon_j|\sigma \sim t_\nu(0,\sigma)$.

Double-exponential errors: If $\lambda_j \sim \mathrm{E}(2)$, the exponential distribution having mean 2, then $\epsilon_j|\sigma \sim \mathrm{DE}(0,\sigma)$, where DE denotes the double-exponential distribution.

Exponential-power-family errors: If $p(\lambda_j)$ is stable with parameter $\alpha/2$, then $\epsilon_j|\sigma \sim \mathrm{EP}(\alpha,\sigma)$, where EP denotes the exponential power distribution.

Logistic errors: If $1/\sqrt{\lambda_j}$ has the asymptotic Kolmogorov distance distribution, then $\epsilon_j|\sigma$ is logistic (see Andrews and Mallows 1974).

Several authors, including Box and Tiao (1973, p. 157) and Spiegelhalter (1977), have attempted to use this scale-mixtures approach to modelling in order to investigate Bayesian robustness issues. However, analytical computations using these nonnormal distributional assumptions, even with improper priors, are notoriously difficult. This is especially true when μ_j is a nonlinear function of $\boldsymbol{\theta}$. Common examples of such models include the exponential growth model $\mu_j = \theta_1 e^{\theta_2 x_j}$, the inverse polynomial model (Nelder 1966) $\mu_j = (x_j - \theta_1)/\{\theta_2 + \theta_3(x_j + \theta_1)\}$, and more application-specific models (see the example in Section 3 below) such as the Michaelis-Menten model $\mu_j = \theta_1 + \theta_2 x_j/(\theta_3 + x_j)$. Furthermore, it is in precisely these nonlinear settings where departures from assumptions can have the greatest impact on results obtained from standard distributional forms.

Fortunately, the Gibbs sampler provides a computational mechanism tailor-made for this situation. The approach is a Monte Carlo integration method which proceeds by a particular Markov chain, and as such is closely related to the "data augmentation" approach to missing-data problems introduced by Tanner and Wong (1987). The reader is referred

to Gelfand and Smith (1990) for a discussion of this method and its properties. Put simply, given a collection of k (possibly vector-valued) random variables $U_1, \ldots, U_k$, all that we require to generate a random sample from their full joint distribution $p(U_1, \ldots, U_k)$ is the ability to successively sample from each of the complete conditional distributions $p(U_s|U_r, r \neq s)$, $s = 1, \ldots, k$. Provided these complete conditional distributions uniquely determine the full joint distribution, Geman and Geman (1984) show that the k-tuple produced at the tth iteration of the sampling scheme, $(U_1^{(t)}, \ldots, U_k^{(t)})$, converges in distribution to a random variate from $p(U_1, \ldots, U_k)$. Replicating this process G times (a total of ktG random generations), we obtain k-tuples $\{U_{1g}^{(t)}, \ldots, U_{kg}^{(t)}), g = 1, \ldots, G\}$ for estimation of any desired marginal density. In particular, if $p(U_s|U_r, r \neq s)$ is available in closed form, then

$$\hat{p}(U_s) = \frac{1}{G} \sum_{g=1}^{G} p(U_s|U_{rg}^{(t)}, r \neq s). \tag{1}$$

In our case, letting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, we require the complete conditional distributions $p(\lambda_j|(\boldsymbol{\lambda} \setminus \lambda_j), \boldsymbol{\mu}, \sigma, \mathbf{y})$, $j = 1, \ldots, n$. Under the independence prior on the components of $\boldsymbol{\lambda}$ we have $\lambda_j|(\boldsymbol{\lambda} \setminus \lambda_j), \boldsymbol{\mu}, \sigma, \mathbf{y} \sim \lambda_j|\mu_j, \sigma, y_j$. But by Bayes's theorem, $p(\lambda_j|\mu_j, \sigma, y_j) \propto p(y_j|\mu_j, \sigma, \lambda_j)p(\lambda_j)$, where the appropriate normalization constant, $p(y_j|\mu_j, \sigma)$, is known by construction. Hence the complete conditional for λ_j will always be of known functional form. Generation of the required samples may be done directly if this form is a standard density; otherwise, a carefully selected rejection method may be employed.

Since the remaining complete conditionals (for σ^2 and the μ_j) are determined given $\boldsymbol{\lambda}$, convenient prior specifications may often be employed with the normal likelihood, again leading to direct sampling. Failing this, we may again employ rejection sampling; the example in the next section offers an illustration.

3. SELECTION OF AN ERROR DISTRIBUTION FOR A NONLINEAR MODEL

Consider the model choice problem where $\mathcal{M}_i$ denotes the nonlinear model $y_j = \mu_j + \epsilon_{ij}$, in which $\mu_j = f(x_j, \theta)\boldsymbol{\beta}$, i indexes the model error distribution, and j indexes the observation, as before. Here $\boldsymbol{\beta}$ is a k-dimensional vector of linear nuisance parameters, and $\mathbf{f}(\mathbf{x}_j, \theta) = (f_1(\mathbf{x}_j, \theta), \ldots, f_k(\mathbf{x}_j, \theta))$ is a collection of known functions, nonlinear in θ. Creating the $n \times k$ matrix $\mathbf{F}_\theta = (\mathbf{f}(\mathbf{x}_j, \theta))$, the log likelihood is

$$\log p(\mathbf{y}|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2, \mathcal{M}_i) = -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{F}_\theta\boldsymbol{\beta})^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{y} - \mathbf{F}_\theta\boldsymbol{\beta}) - n \log \sigma - \frac{1}{2}\sum_{j=1}^{n} \log \lambda_j, \tag{2}$$

where again $\mathbf{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{\Sigma}_i = \text{Diag}(\lambda_1, \ldots, \lambda_n)$. Assume that $\boldsymbol{\beta} \sim \mathsf{N}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_0$ known. In addition, let $\sigma^2 \sim \text{IG}(a_0, b_0)$, where IG denotes the inverse gamma distribution, and let θ have prior distribution $p(\theta)$.

A conceptually simple solution to estimating the $p(\mathcal{M}_i|\mathbf{y})$, and hence a Bayes factor (Jeffreys 1961) between any two of the models, is to include a model indicator as a further parameter in the model, as the notation in (2) above suggests. Specifically, think of $\mathcal{M}$ as a discrete parameter taking values in the range $\{1, \ldots, m\}$. Then the complete conditional for $\mathcal{M}$ is given by

$$p(\mathcal{M}_i|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2, \mathbf{y}) = \frac{p(\mathbf{y}|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2, \mathcal{M}_i)p(\mathcal{M}_i|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)}{\sum_{k=1}^{m} p(\mathbf{y}|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2, \mathcal{M}_k)p(\mathcal{M}_k|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2)}, \quad i = 1, \ldots, m. \tag{3}$$

While this distribution must be restandardized at each stage of the Gibbs algorithm, generation of the corresponding $\mathcal{M}^{(g)}$ samples is a straightforward programming exercise. Assuming convergence of the Markov chain to its stationary distribution, the quantity $\sum_{g=1}^{G}(\text{no. of } \mathcal{M}^{(g)\prime}\text{s equal to } i)/G$ provides a simple point estimate of $p(\mathcal{M}_i|\mathbf{y})$ having maximum standard error $\sqrt{1/4G}$. Next, standard hierarchical Bayes calculations (see Lindley and Smith 1972) yield the following complete conditional distributions:

$$\theta|\boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma, \mathbf{y}, \mathcal{M}_i \propto h(\theta) = \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{F}_\theta\boldsymbol{\beta})^\mathsf{T}\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}-\mathbf{F}_\theta\boldsymbol{\beta})\right\} p(\theta);$$

$$\boldsymbol{\beta}|\theta, \boldsymbol{\lambda}, \sigma, \mathbf{y}, \mathcal{M}_i \sim \mathsf{N}(\mathbf{Bb}, \mathbf{B}),$$

where $\mathbf{B}^{-1} = (1/\sigma^2)\mathbf{F}_\theta^\mathsf{T}\boldsymbol{\Sigma}_i^{-1}\mathbf{F}_\theta + \boldsymbol{\Sigma}_0^{-1}$ and $\mathbf{b} = (1/\sigma^2)\mathbf{F}_\theta^\mathsf{T}\boldsymbol{\Sigma}_i^{-1}\mathbf{y} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0$; and

$$\sigma^2|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{y}, \mathcal{M}_i \sim \text{IG}\left(a_0 + \frac{n}{2}, \left\{b_0^{-1} + \frac{1}{2}\sum_{j=1}^{n}\frac{\{y_j - \mathbf{f}(\mathbf{x}_j, \theta)\boldsymbol{\beta}\}^2}{\lambda_j}\right\}^{-1}\right).$$

Finally, we must determine the λ_j complete conditionals. Suppose that, due to uncertainty about the error density and the impact of possible outliers, we wish to compare the $m = 2$ models $\mathcal{M}_1 : \epsilon_j \sim t(0, \sigma^2, \nu = 2)$ and $\mathcal{M}_2 : \epsilon_j \sim \text{DE}(0, \sigma)$. For $\mathcal{M}_1$, we have the conditional distribution

$$\lambda_j|\theta, \boldsymbol{\beta}, \sigma, \mathbf{y}, \mathcal{M}_1 \sim \text{IG}\left(\frac{\nu+1}{2}, \left\{\frac{1}{2}\left[\frac{\{y_j - \mathbf{f}(\mathbf{x}_j, \theta)\boldsymbol{\beta}\}^2}{\sigma^2} + \nu\right]\right\}^{-1}\right), \quad j = 1, \ldots, n,$$

in a manner very similar to that of Andrews and Mallows (1974). For $\mathcal{M}_2$ we have the complete conditional

$$\lambda_j|\theta, \boldsymbol{\beta}, \sigma, \mathbf{y}, \mathcal{M}_2 \propto \lambda_j^{-1/2}\exp\left\{-\frac{1}{2}\left(\lambda_j + \frac{\{y_j - \mathbf{f}(\mathbf{x}_j, \theta)\boldsymbol{\beta}\}^2}{\lambda_j\sigma^2}\right)\right\},$$

that is, $\lambda_j|\theta, \boldsymbol{\beta}, \sigma, \mathbf{y}, \mathcal{M}_2 \sim \text{GIG}(\frac{1}{2}, 1, \{\mathbf{y}_j - \mathbf{f}(\mathbf{x}_j, \theta)\boldsymbol{\beta}\}^2/\sigma^2)$, $j = 1, \ldots, n$, where GIG denotes the generalized inverse Gaussian distribution (see Devroye 1986, p. 478). In order to sample from this density, we note that it is the reciprocal of an inverse Gaussian $\left(|\sigma/\{\mathbf{y}_j - \mathbf{f}(\mathbf{x}_j, \theta)\boldsymbol{\beta}\}|, 1\right)$, a density from which we may easily sample. In fact, all of the preceeding distributions are available in closed form with the exception of that of θ, for which we would use a rejection sampling method in order to obtain the necessary Gibbs samples.

EXAMPLE (Michaelis-Menten model). As a concrete example, we will entertain a model commonly used in biochemical kinetics, the Michaelis-Menten model. Here we suppose that $\mu_j = \gamma + \alpha x_j/(\theta + x_j)$, where $\alpha, \gamma \in \mathbb{R}$ and $\theta \in \mathbb{R}^+$, so that we have design matrix

$$\mathbf{F}_\theta^\mathsf{T} = \begin{pmatrix} 1 & \cdots & 1 \\ \dfrac{x_1}{\theta + x_1} & \cdots & \dfrac{x_n}{\theta + x_n} \end{pmatrix}$$

and $\boldsymbol{\beta}^\mathsf{T} = (\gamma, \alpha)$. This model has been previously analyzed in a Bayesian framework by Eaves (1983), and in a classical framework by Currie (1982), among many others [see the books of Ratkowsky (1983) and Bates and Watts (1988) for additional references].

TABLE 1: Puromycin experiment data.

Case j	x_j	y_j	Case j	x_j	y_j
1	0.02	76	7	0.22	159
2	0.02	47	8	0.22	152
3	0.06	97	9	0.56	191
4	0.06	107	10	0.56	201
5	0.11	123	11	1.10	207
6	0.11	139	12	1.10	200

The dataset we shall use was collected by Treloar (1974), is reproduced in Bates and Watts (1988), and is displayed in Table 1. These data record the "velocity" y_j of an enzymatic reaction (in counts/min^2) as a function of substrate concentration x_j (in ppm), where the enzyme has been treated with Puromycin. Our goal is to obtain estimates of the marginal posterior density of the parameter α, and also the marginal posterior density of the mean velocity at $X = 0.5$, a substrate concentration not represented in the original data.

We implement the Gibbs sampler using the complete conditional distributions given above, taking vague priors on $\boldsymbol{\beta}$ and σ^2 by letting $\boldsymbol{\Sigma}_0 = \mathbf{0}$, $a_0 = 3.44$, and $b_0 = 0.00341$ (σ^2 has prior mean 120 and prior variance 100^2), and choose a lognormal prior for θ, $\log \theta \sim \mathsf{N}(-5, 2^2)$. Since the complete conditional distribution of θ is available only up to proportionality to $h(\theta)$, $\theta_g^{(t)}$-samples must be generated using a rejection algorithm (see for example Devroye 1986, p. 40). We chose to transform to $\eta = \log \theta$ and use rejection from a multiple of its asymptotic density, a normal with mean $\mu_r = \hat{\eta}$, the generalized MLE of η [i.e., the value of η which maximizes $g(\eta) = e^{\eta} h(e^{\eta})$], and variance σ_r^2 equal to the reciprocal of the Fisher information for η. To ensure that our rejection density would "blanket" $g(\eta)$ in the tails, we actually used a heavier-tailed t_5 distribution in place of a normal. Note that, while we have suppressed this in the notation, μ_r and σ_r^2 are both functions of the current Gibbs samples $(\boldsymbol{\beta}_g^{(t)}, \sigma_g^{(t)}, \boldsymbol{\lambda}_g^{(t)})$, and hence must be recomputed at each iteration of the algorithm.

Running the Gibbs sampler for each of the $\mathcal{M}_i$ separately with $G = 2500$, we monitored the collection $\{\alpha_g^{(t)}, g = 1, \ldots, G\}$ over 5-unit t-increments ($t = 5, 10, \ldots$). Stabilization of the empirical $\alpha_g^{(t)}$ quantiles and successive density estimates obtained using Equation (1) indicated convergence of the algorithm within $t = 50$ iterations. At this point a final estimate $\hat{p}(\alpha|\mathbf{y}, \mathcal{M}_i)$ was computed for model i. The two resulting density estimates are shown as broken lines in Figure 1(a). (As a comparison, a density estimate for the standard model which assumes normal errors is also shown.) We then ran the algorithm again, this time including the model parameter $\mathcal{M}$ in the sampling order and generating it according to the distribution given in (3), where we simply let $p(\mathcal{M}_i|\theta, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) = \frac{1}{2}$, $i = 1, 2$. The resulting estimated posterior probabilities (observed $\mathcal{M}^{(g)}$ frequencies) are (to two decimal places) $\hat{p}(\mathcal{M}_1|\mathbf{y}) = 0.73$ and $\hat{p}(\mathcal{M}_2|\mathbf{y}) = 0.27$, where again we used $G = 2500$ to ensure that the standard error associated with each of these point estimates would be less than 0.01 (assuming convergence of the Markov chain). Finally, we used these estimated model probabilities to mix the t_2 and double exponential estimated posteriors, obtaining the final overall estimate $\hat{p}(\alpha|\mathbf{y})$ shown as the solid line in Figure 1(a). Note that the estimated posterior seems overly concentrated under the standard-normal-errors model. Of the two more reasonable candidates, the t_2 seems slightly preferable (posterior odds of nearly 3 : 1). The overall mixture estimate offers a reasonable compromise between

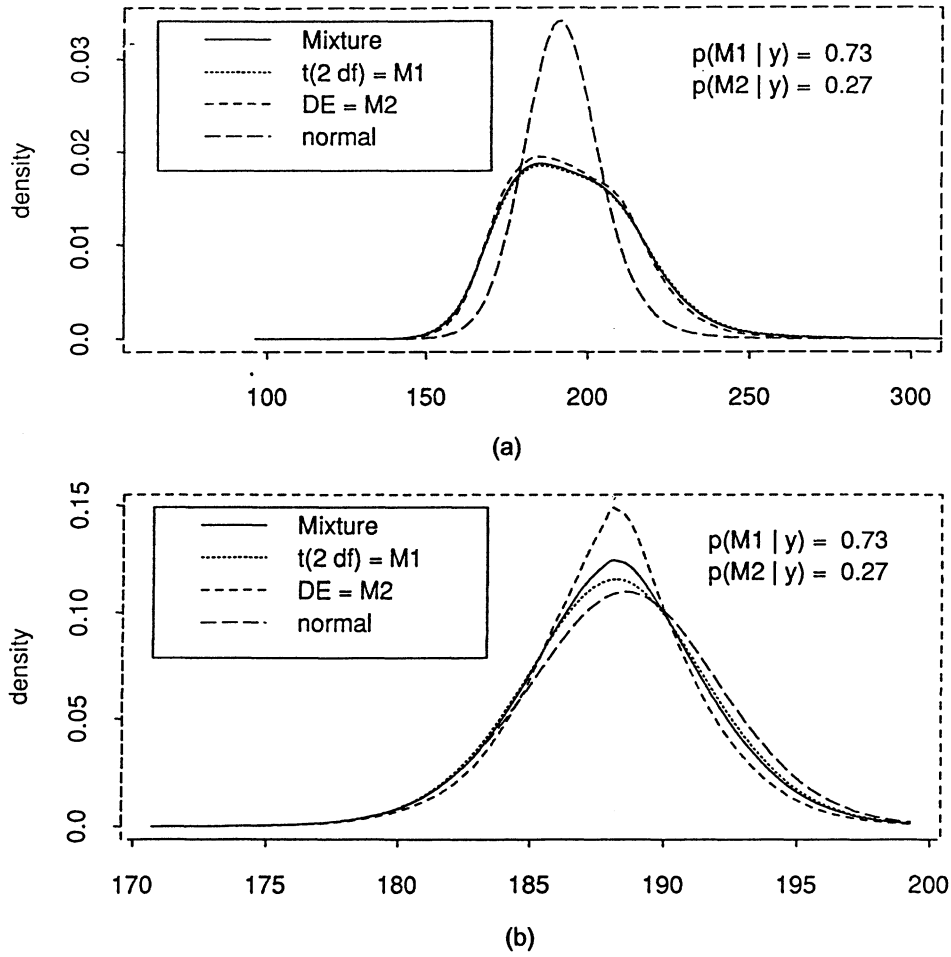

FIGURE 1: Estimated posteriors, Puromycin data. $t = 50$, $G = 2500$. (a) Posteriors for α; mode of mixture = 185. (b) Posteriors for mean response at $x = 0.5$; mode of mixture = 188.1.

the heavier-tailed t_2-errors assumption and the sharply peaked double-exponential-errors model.

To obtain the estimated marginal posterior distribution for the mean response level at $X = 0.5$ (a value not present in the original data), we employ a transformation of $\boldsymbol{\beta}$, whose complete conditional distribution lends itself nicely to this purpose. Putting $\delta = \mathcal{E}[Y|X = 0.5] = \mathbf{l}^T\boldsymbol{\beta}$, where $\mathbf{l}^T = (1, 0.5/(\theta + 0.5))$, we easily obtain $p(\delta|\boldsymbol{\lambda}, \sigma, \mathbf{y}, \mathcal{M}_i) = \mathbf{N}(\mathbf{l}^T\mathbf{B}\mathbf{b}, \mathbf{l}^T\mathbf{B}\mathbf{l})$. Mixing this distribution over $\{(\theta_g^{(t)}, \sigma_g^{(t)}, \boldsymbol{\lambda}_g^{(t)}), g = 1, \ldots, G\}$, we obtain the density estimates $\hat{p}(\delta|\mathbf{y}, \mathcal{M}_i)$, $i = 1, 2$, shown again as dashed lines in Figure 1(b). We again include the corresponding normal-model density estimate for comparison. Using the same model weights $\hat{p}(\mathcal{M}_i|\mathbf{y})$ as above, we again may obtain an overall density estimate $\hat{p}(\delta|\mathbf{y})$, again shown as a solid line in the figure. This time the double-exponential density estimate departs more markedly from the t_2-errors estimate; again the mixture distribution provides a compromise.

Figure 1(b) illustrates the method's ability to derive estimated posteriors for functions of the model parameters without rerunning the algorithm. As a final remark, it is perhaps worth noting that despite the rather large value of G used and the repeated maximization

and conservative asymptotic variance inherent in the $\theta^{(g)}$ generation, the algorithm is surprisingly fast, completing the Gibbs sampling and all density estimation in roughly 10 minutes on a DECStation 3100.

REFERENCES

Andrews, D.F., and Mallows, C.L. (1974). Scale mixtures of normality. *J. Roy. Statist. Soc. Ser. B*, 36, 99–102.

Bates, D.M., and Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*. Wiley, New York.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Edition. Springer-Verlag, New York.

Box, G.E.P., and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

Currie, D.J. (1982). On estimating Michaelis-Menten parameters: Bias, variance, and experimental design. *Biometrics*, 38, 907–919.

Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer-Verlag, New York.

Eaves, D.M. (1983). On Bayesian nonlinear regression with an enzyme example. *Biometrika*, 70, 373–379.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Amer. Statist. Assoc.*, 85, 972–985.

Gelfand, A.E., and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85, 398–409.

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. and Machine Intelligence*, 6, 721–741.

Jeffreys, H. (1961). *Theory of Probability*. Third Edition. Clarendon Press, Oxford.

Lindley, D.V., and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B*, 34, 1–41.

Nelder, J.A. (1966). Inverse polynomials, a useful group of multifactor response functions. *Biometrics*, 22, 128–140.

Ratkowsky, D. (1983). *Nonlinear Regression Modeling*. Marcel Dekker, New York.

Smith, A.F.M. (1983). Bayesian approaches to outliers and robustness. *Specifying Statistical Models* (J. Florens *et al., eds.*) Springer-Verlag, New York, 13–35.

Spiegelhalter, D.J. (1977). A test for normality against symmetric alternatives. *Biometrika*, 64, 415–418.

Tanner, M.A., and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.*, 82, 528–550.

Treloar, M.A. (1974). Effects of Puromycin on galactosyltransferase of golgi membranes. Master's Thesis, University of Toronto.

Received 22 August 1990
Revised 5 March 1991
Accepted 2 April 1991

Division of Biostatistics
Box 197, Mayo Memorial Bldg.
School of Public Health
University of Minnesota 55455
Minneapolis, Minnesota
U.S.A.

Graduate School of Business
University of Chicago
Chicago, Illinois
U.S.A.