Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling

ALAN E. GELFAND, ADRIAN F. M. SMITH, and TAI-MING LEE*

Constrained parameter problems arise in a wide variety of applications, including bioassay, actuarial graduation, ordinal categorical data, response surfaces, reliability development testing, and variance component models. Truncated data problems arise naturally in survival and failure time studies, ordinal data models, and categorical data studies aimed at uncovering underlying continuous distributions. In many applications both parameter constraints and data truncation are present. The statistical literature on such problems is very extensive, reflecting both the problems' widespread occurrence in applications and the methodological challenges that they pose. However, it is striking that so little of this applied and theoretical literature involves a parametric Bayesian perspective. From a technical viewpoint, this perhaps is not difficult to understand. The fundamental tool for Bayesian calculations in typical realistic models is (multidimensional) numerical integration, which often is problematic in unconstrained contexts and can be well-nigh impossible for the kinds of constrained problems we consider. In this article we show that Bayesian calculations *can* be implemented routinely for constrained parameter and truncated data problems by means of the Gibbs sampler. Specific models discussed include constrained multinormal parameters, constrained linear model parameters, ordered parameters in experimental family models, data and order restricted parameters from exponential distributions, straight line regression with censoring and bivariate grouped data models. Analysis of data sets illustrating the first two of these settings is provided.

KEY WORDS: Bayesian inference; Constrained parameters; Gibbs sampler; Truncated data.

Constrained parameter problems arise in a wide variety of applications, including bioassay, actuarial graduation, ordinal categorical data, response surfaces, reliability development testing, and variance component models. Truncated data problems—to be understood as encompassing both censoring and scoring or grouping mechanisms—arise naturally in survival and failure time studies, ordinal data models, and categorical data studies aimed at uncovering underlying continuous distributions. In many applications both parameter constraints and data truncation occur.

The parametric Bayes perspective is attractive for examining such models. For example, consider ordered parameter (slippage) models, which in a classical setting might employ isotonic regression of maximum likelihood estimates to obtain point estimates. A more satisfying analysis would develop and compare posterior distributions arising from priors (possibly vague) that reflect the order restrictions. However, analytic approaches (exact or approximate) for carrying out required multi dimensional integrations in this case (and in fact for all the aforementioned problems) will be well-nigh impossible.

This article aims to show that Bayesian calculations *can* be implemented routinely for constrained parameter and truncated data problems by means of the Gibbs sampler. The Gibbs sampler was introduced by Geman and Geman (1984) in the context of image processing, see also Hastings (1970) for an early recognition. Later, it was proposed as a general method for Bayesian calculations by Gelfand and Smith (1990).

In general we shall assume that the desired outcome of a Bayesian analysis is the calculation and display of marginal posterior (predictive) densities of parameters (unobserved data) of interest, although summaries (for example, via modes, moments, quantiles) often will suffice. As we shall see, the Gibbs sampler will provide the basis for whatever form of final inference summary we require.

In Section 1, we briefly review the Gibbs sampler and comment on experience with its use for other classes of statistical problems. In Section 2, we present a general overview formulation of the structure of constrained parameter and truncated data problems and the resulting form of the Gibbs sampler. In Section 3 we develop detailed analyses for various examples chosen to give an overview of the power and scope of the Gibbs sampler in reducing seemingly impossibly complex computational tasks to simple, easily implemented, iterative sampling schemes. In Section 4 we provide illustrative analyses of two artificial data sets generated from models chosen to present extremely awkward inference problems. Finally, in Section 5 we provide a summary discussion.

1. THE GIBBS SAMPLER

Our subsequent development uses the following notational conventions. Densities are developed generically by square brackets so that joint, conditional, and marginal forms for random variables U, V, appear as $[U, V]$, $[U|V]$, and $[V]$. The usual marginalization by integration is denoted by forms such as $[U] = \int [U|V]\cdot[V]$. For a collection of random variables $[U_1, U_2, \ldots, U_k]$, the full conditional densities thus can be denoted by $[U_s|U_r, r \neq s]$, $s = 1, 2, \ldots, k$, and the marginal densities can be denoted by $[U_s]$, $s = 1, 2, \ldots, k$.

Consider the following problem. If we are given the ability to draw random variate samples of U_s from $[U_s|U_r, r \neq s]$ for specified values of U_r, $r \neq s$, $s = 1, 2, \ldots, k$, can we find an iterative scheme that enables us to make sample-based estimates, $[\hat{U}_s]$, say of the marginal densities $[U_s]$, $s = 1, 2, \ldots, k$? We shall make the connection with Bayesian cal-

* Alan E. Gelfand is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269-3120. Adrian F. M. Smith is Professor, Department of Mathematics, Imperial College of Science Technology and Medicine, London SW7 2BZ, U.K. Tai-Ming Lee is a graduate student, Department of Statistics, University of Connecticut, Storrs, CT 06269-3120. The first author's research was supported in part by a grant from the U.K. Science and Engineering Research Council.

culations later; for the moment, we note that the general question is answered affirmatively by the following procedure.

Gibbs sampling is a Markovian updating scheme that proceeds as follows. Given an arbitrary starting set of values $U_1^{(0)}, \ldots, U_k^{(0)}$, we draw $U_1^{(1)}$ from $[U_1 | U_2^{(0)}, \ldots, U_k^{(0)}]$, then $U_2^{(1)}$ from $[U_2 | U_1^{(1)}, U_3^{(0)}, \ldots, U_k^{(0)}]$, and so on up to $U_k^{(1)}$ from $[U_k | U_1^{(1)}, \ldots, U_{k-1}^{(1)}]$ to complete one iteration of the scheme. After t such iterations we would arrive at joint a sample $(U_1^{(t)}, \ldots, U_k^{(t)})$. Geman and Geman (1984) showed that under mild conditions $(U_1^{(t)}, \ldots, U_k^{(t)}) \xrightarrow{d} (U_1, \ldots, U_k) \sim [U_1, U_2, \ldots, U_k]$ as $t \to \infty$. Hence for t large enough, $U_s^{(t)}$ for example can be regarded as a sample variate from $[U_s]$. If this process is replicated in parallel m times iid k-tuples $(U_{lj}^{(t)}, \ldots, U_{kj}^{(t)})$ $j = 1, 2, \ldots, m$ result.

A kernel density estimate for $[U_s]$ based on the $U_{sj}^{(t)}$ can be obtained readily (Silverman 1986) and should be adequate if at the last iteration the number of replications, m, is large enough. Using a Rao–Blackwell argument (Gelfand and Smith 1990), however, a density estimate of the form

$$[\hat{U}_s] = \sum_{j=1}^{m} [U_s | U_{rj}^{(t)}, r \neq s] \Big/ m \qquad (1)$$

is better under a wide range of loss functions. This is not surprising, because (1) takes advantage of the known structure in the model whereas the kernel density estimate does not. The form (1) is a discrete mixture distribution, essentially a Monte Carlo integration to accomplish the desired marginalization. Similarly, the expectation $E(h(U_s))$ can be obtained either as a sample estimate based on the $U_{sj}^{(t)}$ or possibly as a "Rao–Blackwellized" version analogous to (1) based on $E(h(U_s) | U_r, r \neq s)$. Now consider a function of the U_i, say $W(U_1, \ldots, U_k)$. Each k-tuple, $(U_{1j}^{(t)}, \ldots, U_{kj}^{(t)})$, provides an observed $W_j^{(t)} \equiv W(U_{1j}^{(t)}, \ldots, U_{kj}^{(t)})$ with marginal distribution, as $t \to \infty$, approximately $[W]$, whence a kernel density esitmator for $[W]$ using these $W_j^{(t)}$ can be developed. If, say, U_s appears as an argument of W, then the full conditional density $[W | U_r, r \neq s]$ can be obtained by univariate transformation from $[U_s | U_r, r \neq s]$. Thus a "Rao–Blackwellized" density estimate for $[W]$ analogous to (1) also can be obtained.

In the Bayesian context the U_i are the unknown parameters (or possibly unobserved data) in the model; W would be any function of the parameters (or unobserved data) that is of interest. All distributions are viewed as conditional on the observed data, whence the marginal densities, $[U_s]$, become the desired marginal posterior distributions of the parameters (or unobserved data).

So far as ease of drawing samples from the complete conditional distributions is concerned, in many cases the likelihood and prior forms specified in the Bayesian model lead to familiar standard full conditional forms, such as normals and gammas, and implementation is immediate. In other cases we simply have, up to proportionality, a mathematical form for the full conditional and must use tailored versions of general random variate generating procedures, such as the ratio of uniforms and rejection methods (see, e.g., Devroye 1986 or Ripley 1987).

Finally, we note that complete implementation of the Gibbs sampler requires determination of t and m. These settings will vary with the application, and some experimentation with different settings likely will be necessary. We do not view this as a deterrent, because random generation is generally inexpensive and there may be no feasible computational alternative. In the examples of Section 4, convergence is evaluated in a univariate manner by plotting marginal posterior density estimates of the form (1) five iterations apart to judge stability. Extensive computational experience with this assessment procedure in a wide variety of parametric models was reported in Gelfand and Smith (1990, 1991), Gelfand, Hills, Racine-Poon, and Smith (1990), and Carlin, Gelfand, and Smith (1992).

Because these cited papers contain discussion of very detailed specification of the Gibbs sampler in various situations, we avoid unnecessary detail in this article and concentrate instead on the structural insights that underlie the specification of the required full conditionals. Having followed the general discussion, the reader can supply the missing detail in any specific example.

2. MODELS: GENERAL STRUCTURE

In this section we provide a discussion of the Gibbs sampler structures arising from rather general formulations of Bayesian parametric versions of constrained parameter and truncated data problems. The implementation problem reduces to identification of the appropriate full conditional distributions and methods for drawing samples from them.

2.1 Constrained Parameter Models

Consider a parametric model for data $\mathbf{Y}$ involving a k-dimensional parameter vector $\boldsymbol{\theta}$, constrained to lie in a subset $S_{\mathbf{Y}}^k$ of R^k. Often the constraint set $S_{\mathbf{Y}}^k$ is determined by order or other inequality relationships among the components θ_i, $i = 1, 2, \ldots, k$, of $\boldsymbol{\theta}$, in which case $S^k = S_{\mathbf{Y}}^k$ does not depend on $\mathbf{Y}$. In other cases constraints occur because the region of positive support for the distribution of $\mathbf{Y}$ depends on $\boldsymbol{\theta}$, so that $\mathbf{Y}$ occurs explicitly in $S_{\mathbf{Y}}^k$; see, for example, Section 3.5, in which $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ and $\theta_i \leq Y_i$, $i = 1, 2, \ldots, k$. In the former case, it it natural to think of the constraint as built into the specification of the prior distribution, $[\boldsymbol{\theta} | \boldsymbol{\lambda}]$, where $\boldsymbol{\lambda}$ is some hyperparameter; whereas in the latter case it is natural to think of the constraint as built into the likelihood, $[\mathbf{Y} | \boldsymbol{\theta}]$. In either case it suffices to note that the constrained Bayesian model (likelihood $\times$ prior) is given by

$$\begin{cases} [\mathbf{y} | \boldsymbol{\theta}] \cdot [\boldsymbol{\theta} | \boldsymbol{\lambda}], & (\mathbf{y}, \boldsymbol{\theta}) \in S \\ 0, & (\mathbf{y}, \boldsymbol{\theta}) \notin S, \end{cases}$$

where $S = \{(\mathbf{y}, \boldsymbol{\theta}): \boldsymbol{\theta} \in S_y^k\}$. In general $[\mathbf{Y} | \boldsymbol{\theta}]$ and $[\boldsymbol{\theta} | \boldsymbol{\lambda}]$, as functions of $\boldsymbol{\theta}$, have the functional forms they would have had if constraints had been ignored. It follows immediately (generalizing slightly a remark in Box and Tiao [1973, p. 67]) that the posterior distribution for $\boldsymbol{\theta}$, given the con-

straints, is simply the unconstrained posterior appropriately normalized so that

$$[\theta|\mathbf{Y}] = \frac{[\mathbf{Y}|\theta]\cdot[\theta|\lambda]}{\int_{S_{\mathbf{Y}}^k}[\mathbf{Y}|\theta]\cdot[\theta|\lambda]}, \qquad \theta \in S_{\mathbf{Y}}^k. \tag{2}$$

Now let $S_i^k(\theta_j, j \neq i)$ denote the cross-section of $S_{\mathbf{Y}}^k$ defined by the constraints on component θ_i at a specified set of values $\theta_j, j \neq i$ (where for the cross-section we have suppressed possible dependence on $\mathbf{Y}$ for notational convenience). In the case of scalar components $S_i^k(\theta_j, j \neq i)$ is a subset of R^1, typically an interval or a collection of intervals. It then follows immediately from (2) that the complete posterior conditional distribution for θ_i is defined by

$$[\theta_i|\mathbf{Y}, \lambda, \theta_j, j \neq i] \propto [\mathbf{Y}|\theta]\cdot[\theta|\lambda], \qquad \theta_i \in S_i^k(\theta_j, j \neq i), \tag{3}$$

where the right side is regarded as a function of θ_i for specified $\theta_j, j \neq i$. When for θ_i the likelihood and prior combine to give a conjugate Bayesian form, the unconstrained version of the full conditional for θ_i will be a familiar standard distribution defined by the conjugate prior-to-posterior updating. The constrained form (3) then simply will be the standard distribution restricted to $S_i^k(\theta_j, j \neq i)$.

This latter point is critical. Regardless of how complicated the overall constraint set $S_{\mathbf{Y}}^k$ is, to implement the Gibbs sampler we need only consider $S_{\mathbf{Y}}^k$ in univariate cross-sections. Moreover, to carry out the actual sampling we need only identify the full conditionals under the unconstrained model and then make the restriction to the cross-sections.

One way of doing this is simply to generate from the unconstrained full conditional and retain the variate value only if it falls in the cross-section constraint region. Alternately, suppose the form of the distribution function, F_i, say of the full conditional for θ_i is available and the cross-section is an interval, say $[a, b]$. Then if U is a uniform (0, 1) variate, as was noted by Devroye (1986, p. 38) $\theta_i = F_i^{-1}[F_i(a) + U(F_i(b) - F_i(a))]$ is a drawing from the constrained full conditional. Thus we sample "one for one" from the constrained full conditional. This easily is extended to the case where the cross-section is a collection of disjoint intervals, say $\cup_{j=1}^{r}[a_j, b_j]$. In this case we choose $J = j$ with probability $[\sum_{j=1}^{r}(F_i(b_j) - F_i(a_j))]^{-1}[F_i(b_j) - F_i(a_j)]$ and, given j, set $\theta_i = F_i^{-1}[F_i(a_j) + U(F_i(b_j) - F_i(a_j))]$, where U again is a uniform (0, 1) variate.

In general, sampling from constrained full conditionals will not be particularly efficient, especially in the case of nonstandard unnormalized distributions. But this is more than compensated for by the ease of implementation of the Gibbs sampler, enabling one to carry out full Bayesian calculations for complex constrained parameter problems that previously were unanalyzable by standard numerical integration techniques.

Finally we note a further feature that arises in implementing the Gibbs sampler were we to seek to extend these calculations to a hierarchical model by adding a prior $[\lambda]$ for the hyperparameter λ, thus far assumed to be known. The full conditionals for the θ_i are unchanged and are given by (3), but the full conditional for λ does not depend on $\mathbf{Y}$ and takes the form

$$[\lambda|\mathbf{Y}, \theta] \propto [\theta|\lambda][\lambda]c(\lambda), \tag{4}$$

where $c(\lambda) = (\int_S [\mathbf{Y}|\theta][\theta|\lambda])^{-1}$. If θ is not constrained by $\mathbf{Y}$, $c(\lambda)$ simplifies to $(\int_{S^k}[\theta|\lambda])^{-1}$, but regardless $c(\lambda)$ typically will not be available explicitly (see, for example, Section 3.1), making sampling from (4) almost impossible.

2.2 Censored Data Models

To develop a general framework for censored data models, consider random n vectors $\mathbf{Y}, \mathbf{V}, \mathbf{W}$ with joint density defined by

$$[\mathbf{Y}, \mathbf{V}, \mathbf{W}|\theta, \eta] = [\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta]\cdot[\mathbf{V}, \mathbf{W}|\eta]$$

in terms of parameters θ and η and define componentwise a further random n vector $\mathbf{Z}$ by

$$\begin{aligned} Z_j &= V_j \quad \text{if} \quad Y_j \leq V_j \\ &= Y_j \quad \text{if} \quad V_j < Y_j < W_j, \qquad j = 1, 2, \ldots, n. \\ &= W_j \quad \text{if} \quad Y_j \geq W_j. \end{aligned} \tag{5}$$

We shall consider $\mathbf{Z}$ to be observed data arising as a censored form of $\mathbf{Y}$ through the censoring process $[\mathbf{V}, \mathbf{W}|\eta]$, with $\mathbf{V}$ and $\mathbf{W}$ also observed. In this very general formulation $\mathbf{V}$ and $\mathbf{W}$ are random, but the process could of course be degenerate for either or both. In particular, right or left censoring only (corresponding to $W_j = -\infty$, $V_j = +\infty$, respectively) are included as special cases.

To complete the Bayesian model let us assume that prior distributions are specified in the form $[\theta|\lambda][\eta][\lambda]$, so that the full model becomes

$$[\mathbf{Z}|\mathbf{V}, \mathbf{W}, \theta]\cdot[\mathbf{V}, \mathbf{W}|\eta]\cdot[\theta|\lambda]\cdot[\eta]\cdot[\lambda], \tag{6}$$

where the form of $[\mathbf{Z}|\mathbf{V}, \mathbf{W}, \theta]$ is defined by $[\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta]$ and (5). Other forms of prior specification could of course be considered, but the form given here, involving a hyperparameter λ in the construction of the prior for θ, will suffice for our later illustrative examples. We assume that interest focuses on the marginal posterior distributions for the components of θ, $[\theta_i|\mathbf{Z}, \mathbf{V}, \mathbf{W}]$, $i = 1, 2, \ldots, k$, as well as perhaps $[\eta|\mathbf{Z}, \mathbf{V}, \mathbf{W}]$.

At first sight it appears natural to try to implement the Gibbs sampler using the full conditional distributions for $\theta_1, \theta_2, \ldots, \theta_k, \theta_k, \eta$ and λ. We note, however, that

$$[\theta_i|\mathbf{Z}, \mathbf{V}, \mathbf{W}, \theta_j, j \neq i, \eta, \lambda] \propto [\mathbf{Z}|\mathbf{V}, \mathbf{W}, \theta]\cdot[\theta|\lambda],$$

with the right side considered a function of θ_i for specified $\theta_j, j \neq i$. This leads to difficulties, because the density $[\mathbf{Z}|\mathbf{V}, \mathbf{W}, \theta]$ generally will be awkward to deal with. Suppose, for example, that $[\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta] = [\mathbf{Y}|\theta] = \prod_{j=1}^{n} f_j(Y_j|\theta)$. Then $[Z_j|\mathbf{V}, \mathbf{W}, \theta] = f_j(Z_j|\theta)$ if $V_j < Z_j < W_j$, but has point masses $\underline{\delta}_j(V_j, \theta) = \int_{-\infty}^{V_j} f_j(Z|\theta)dZ$ at $Z_j = V_j$ and $\bar{\delta}_j(W_j, \theta) = \int_{W_j}^{\infty} f_j(Z|\theta)dZ$ at $Z_j = W_j$. Generally, $\underline{\delta}_j$, and $\bar{\delta}_j$ will not be available in explicit form, which means that this also will be the case for $[\mathbf{Z}|\mathbf{V}, \mathbf{W}, \theta]$ whenever any Z_j equals either V_j or W_j (i.e., whenever censoring occurs).

To avoid this difficulty, suppose instead that we treat $\mathbf{Y}$ as an unobservable and include it in the Gibbs sampler. The model (6) now becomes, in its most general form,

$$[\mathbf{Z}|\mathbf{Y}, \mathbf{V}, \mathbf{W}][\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta][\mathbf{V}, \mathbf{W}|\eta][\theta|\lambda][\mathbf{z}][\lambda]. \quad (7)$$

Here $\{\mathbf{Z}|\mathbf{Y}, \mathbf{V}, \mathbf{W}]$ is of course a degenerate distribution, and in typical applications we shall have $\{\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta] = \{\mathbf{y}|\theta)$. Note that now

$$[\theta_i|\mathbf{Z}, \mathbf{Y}, \mathbf{V}, \mathbf{W}, \theta_j, j \neq i, \eta, \lambda] \propto [\mathbf{Y}|\mathbf{V} \cdot \mathbf{W} \cdot \theta][\theta|\lambda]. \quad (8)$$

The right side of (8) is now an explicit function of θ_i, and sampling no longer presents a problem. As noted in Section 2.1, under conjugacy sampling from θ_i will simply involve sampling from a standard distribution. Without conjugacy, however, we will need to sample from a nonstandardized density using, for example, ratio of uniforms or rejection techniques. The remaining full conditionals required for the Gibbs sampler are given by $[\eta|\mathbf{Z}, \mathbf{Y}, \mathbf{V}, \mathbf{W}, \theta, \lambda] \propto [\mathbf{V}, \mathbf{W}|\eta][\eta]$, and $[\lambda|\mathbf{Z}, \mathbf{Y}, \mathbf{V}, \mathbf{W}, \theta, \eta] \propto [\theta|\lambda][\lambda]$, for which similar remarks apply, and finally $[\mathbf{Y}|\mathbf{Z}, \mathbf{V}, \mathbf{W}, \theta, \eta, \lambda] \propto [\mathbf{Z}|\mathbf{Y}, \mathbf{V}, \mathbf{W}\} \cdot [\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta]$. Again for illustration consider the typical case where $[\mathbf{Y}|\mathbf{V}, \mathbf{W}, \theta] = [\mathbf{Y}|\theta] = \prod_{j=1}^{n} f_j(Y_j|\theta)$. Here the Y_j are conditionally independent and the full conditional distribution for Y_j is degenerate at Z_j if $V_j < Z_j < W_j$ and has the distribution $f_j(\cdot|\theta)$ restricted to $(-\infty, V_j]$ if $Z_j = V_j$ and the distribution $f_j(\cdot|\theta)$ restricted to $[W_j, \infty)$ if $Z_j = W_j$. Sampling the Y_j is therefore routine, the latter two cases being handled perhaps by the "one for one" method described in Section 2.1.

2.3 Grouped Data Models

To illustrate scored or grouped ordinal data models, suppose that instead of observing the actual coordinates of a random n vector $\mathbf{Y}$ we only observe a score, $R_j = b_t$ if $a_{t-1} \leq Y_j \leq a_t$, $j = 1, 2, \ldots, n$, and $t = 1, 2, \ldots, T$, where a_t, b_t are known fixed constants (often with $a_0 = -\infty$, $a_T = +\infty$). Assuming $\mathbf{Y}$ as having a parametric distribution $[\mathbf{Y}|\theta]$ and θ as having a prior defined by $[\theta|\lambda]$, $[\lambda]$, the Bayesian model is given by

$$[\mathbf{R}|\theta][\theta|\lambda][\lambda],$$

where $[\mathbf{R}|\theta]$ is induced by $[\mathbf{Y}|\theta]$.

As in the previous section, the natural Gibbs sampler defined directly in terms of the full conditionals $[\lambda|\mathbf{R}, \theta]$ and

$$[\theta_i|\mathbf{R}, \theta_j, j \neq i, \lambda] \propto [\mathbf{R}|\theta][\theta|\lambda]$$

runs into trouble due to the presence of $[\mathbf{R}|\theta]$, which generally is not an explicit expression in terms of $\theta_1, \theta_2, \ldots, \theta_k$. A solution again is to include the unknown $\mathbf{Y}$ as part of the Gibbs sampler. The Bayesian model then becomes

$$[\mathbf{R}|\mathbf{Y}][\mathbf{Y}|\theta][\theta|\lambda][\lambda]$$

and the full conditionals are given by

$$[\lambda|\mathbf{R}, \mathbf{Y}, \theta] \propto [\theta|\lambda][\lambda]$$

$$[\theta_i|\mathbf{R}, \mathbf{Y}, \theta_j, j \neq i, \lambda] \propto [\mathbf{Y}|\theta][\theta|\lambda]$$

together with the conditionals for Y_j given Y_i, $i \neq j$, derived from

$$[\mathbf{Y}|\mathbf{R}, \theta, \lambda] \propto [\mathbf{R}|\mathbf{Y}][\mathbf{Y}|\theta].$$

Sampling is now straightforward. In particular if $[\mathbf{Y}|\theta] = \prod_{j=1}^{n} f_j(Y_j|\theta)$ and $R_j = b_t$, then the full conditional for Y_j is simply $f_j(\cdot|\theta)$ restricted to $[a_{t-1}, a_t]$.

3. MODELS: SPECIFIC EXAMPLES

In this section we make explicit the forms of the Gibbs sampler arising from various examples of the general structures discussed in Section 2. Our development is designed to illuminate the astonishing simplicity with which the appropriately defined Gibbs sampler solves the problem of Bayesian computation in constrained parameter and truncated data contexts.

As noted earlier, there is remarkably little literature on Bayesian approaches to such problems and that which does exist typically does not solve the problem of calculating marginal densities but rather attempts only limited inference summaries in the form of modes or means. Ordered restricted inference was discussed at length from a frequentist perspective in Barlow et al. (1972) and Robertson, Wright, and Dykstra (1988). The former presented some discussion of Bayesian inference for ordered exponential family parameters, but this was largely limited to a discussion of the joint posterior mode as an isotonic regression. The latter provided a convenient review of the brief Bayesian literature on ordered parameters. We know of no systematic discussion of truncated data problems from a Bayesian perspective.

3.1 Ordered Exponential Family Parameters

Motivated by graduation problems in actuarial science, Broffit (1984) considered ordered parameters from a family of models of the form

$$f(Y|\theta) = a(Y)\theta^{b(Y)}e^{-\theta c(Y)}, \qquad \theta > 0. \quad (9)$$

(This family includes models such as Gamma with known shape parameter, normal with known mean, and Poisson.)

Suppose then that conditionally independent observations Y_{ij}, $i = 1, 2, \ldots, k$, and $j = 1, 2, \ldots, n_i$ are available from $f(\cdot|\theta_j)$, where it is assumed that $\theta \in S_k = \{\theta = (\theta_1, \ldots, \theta_k): 0 < \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k\}$. Broffitt (1984) suggested a convenient and flexible prior family for θ over S_k of the form

$$d_k(\delta_1, \ldots, \delta_k; \gamma_1, \ldots, \gamma_k) \prod_{i=1}^{k} \frac{\theta_i^{\delta_i - 1}}{\gamma_i^{\delta_i}} \frac{e^{-\theta_i/\gamma_i}}{\Gamma(\delta_i)}, \quad (10)$$

where d_k is the normalizing constant and δ_i, γ_i are chosen to reflect prior beliefs. Note that if the θ_i were unconstrained (10) becomes a product of independent Gamma priors. In the case where the δ_i are integers Broffitt obtained d_k as a finite multidimensional sum. The joint posterior $[\theta|\mathbf{Y}]$ has the same form as (10), but with δ_i replaced by $\delta_i^* = \delta_i + \sum_{j=1}^{n_i} b(Y_{ij})$ and γ_i replaced by

$$\gamma_i^* = \left(\frac{1}{\gamma_i} + \sum_{j=1}^{n_i} c(Y_{ij})\right)^{-1}.$$

The posterior mean for θ_i under the unrestricted problem is $\hat{\theta}_i^* = \delta_i^* \gamma_i^*$. Using isotonic regression Broffitt obtained the order-restricted Bayes estimate for θ_i under squared error loss as a function of $\hat{\theta}_i^*$ and two d_k's.

To implement the Gibbs sampler we require only the full conditional distribution $[\theta_i | \mathbf{Y}, \theta_j, j \neq i]$, $i = 1, 2, \ldots, k$. Under (10) this is merely a Gamma (δ_i^*, γ_i^*) restricted to $[\theta_{i-1}, \theta_{i+1}]$, where $\theta_0 \equiv 0$, $\theta_{k+1} \equiv \infty$. Thus sampling reduces to interval-restricted sampling from a standard distribution, as discussed in Section 2.1. We need not concern ourselves with calculation of the normalizing constant d_k.

Extension from one-parameter exponential family models to conditionally independent observations with increasing parameters and a constrained form of conjugate prior follows in an obvious and straightforward way.

3.2 Ordered Multinomial Parameters

Sedransk, Monahan, and Chiu (1985) discussed the problem of Bayes estimation of finite population parameters when a random variable X assumes one of a finite set of values $\{b_1, \ldots, b_k\}$ with probabilities $p_1, p_2, \ldots, p_k$. A particular application is the case of household income, where b_j might denote a central value for the jth income category.

Assuming that the categories are arranged in increasing order, we would expect that the p_j would increase up to some category, t, say ($1 \leq t \leq k$), and then decrease thereafter. Typically, t would be unknown. The quantity of primary interest in such a situation might be the finite population mean, $\sum_{j=1}^k b_j p_j$, although other functions of $\mathbf{p}$ also could be of interest. A possible Bayesian model for such problems is given by defining Y_j = # of observations in category j, with $\sum_{j=1}^k Y_j = n$, so that $[\mathbf{Y} | \mathbf{p}] = \text{Mult}(n; p_1, p_2, \ldots, p_k)$. Following Sedransk et al. (1985), given t we specify a prior $[\mathbf{p} | t]$ of the form

$$c(\beta_1, \ldots, \beta_k; t) \sum_{j=1}^{k} p_j^{\beta_j - 1} \tag{11}$$

over $S^k = \{(p_1, \ldots, p_k): p_1 \leq p_2 \leq \cdots \leq p_t \geq p_{t+1} \geq \cdots p_k, 0 \leq p_j \leq 1, \sum_{j=1}^k p_j = 1\}$, where $c(\beta_1, \ldots, \beta_k; t)$ is the normalizing constant. Note that if the p_i were unconstrained, (11) becomes a Dirichlet prior over the k dimensional simplex.

Sedransk et al. (1985) assumed that t is known and computed only desired posterior expectations through Monte Carlo integration, using importance sampling to avoid calculation of c. To implement the Gibbs sampler requires the full conditional distribution for p_i, $i = 1, \ldots, k-1$ (with p_k as a function of these p_i), $[p_i | \mathbf{Y}, p_j, j = 1, \ldots, k-1, j \neq i, t]$. This is clearly a Beta distribution scaled to $[0, 1 - \sum_{\substack{j=1 \\ j \neq i}}^{k-1} p_j]$ and then suitably restricted according to the constraints determined by t. Thus if t is known, then the Gibbs sampler also avoids calculation of c. Moreover, empirical work (Gelfand and Smith 1990) suggests that iterative Monte Carlo integration using the Gibbs sampler will be more efficient than noniterative Monte Carlo integration such as that used by Sedransk et al. (1985).

Suppose t is unknown and assigned a discrete prior $\Pr(t = j) = \tau_j, j = 1, 2, \ldots, k$. We note that $[t | \mathbf{Y}, \mathbf{p}]$ is a degenerate distribution. Thus the Gibbs sampler cannot be employed directly, because a condition for its convergence is that transitions from one t to any other are possible. This hierarchical situation differs from that in expression (4), where λ is a hyperparameter having nothing to do with the order restrictions. Here t determines the restrictions.

Fortunately the marginal posterior for t can be calculated directly, taking the form

$$\Pr(t = j | \mathbf{Y}) = \frac{c(\beta_1, \ldots, \beta_k; j)\tau_j / c(\beta_1 + Y_1, \ldots, \beta_k + Y_k; j)}{\sum_{t=1}^k c(\beta_1, \ldots, \beta_k; t)\tau_t / c(\beta_1 + Y_1, \ldots, \beta_k + Y_k; t)}. \tag{12}$$

Evaluation of the $2k$ constants in (12) can be done straightforwardly using Monte Carlo integration with importance sampling, as in Sedransk et al. (1985). Thus we can estimate the marginal posteriors for the p_i by using the relationship

$$[p_i | \mathbf{Y}] = \sum_{t=1}^{k} [p_i | \mathbf{Y}, t][t | \mathbf{Y}].$$

For each given t we can use the Gibbs sampler in the customary manner to obtain samples approximately from $[p_i | \mathbf{Y}, t]$. We then can *resample* from these samples according to $[t | \mathbf{Y}]$ to obtain observations approximately from $[p_i | \mathbf{Y}]$. Full details of the required sampling in the context of an illustrative example are given in Section 4.1.

Note that in a different context the sequence p_i might for instance be assumed to have a bimodal form; for example, for grouped data arising from samples of exam scores or from samples of heights or of weights. It is clear that our formulation of the present example can be extended to handle such cases. Also note that this example is a nonparametric version of the grouped ordinal data problem. We are concerned only with the probabilities for the income categories, not with an underlying parametric model for the incomes themselves.

Finally, extension to models involving collections of independent multinomials with order restrictions perhaps both across and within populations is straightforward.

4.3 Ordered Linear Model Parameters

We demonstrate the potential of the Gibbs sampler for the Bayesian analysis of constrained parameters in general normal linear models by considering an illustrative analysis of a simple two-way layout. Application to normal means without linear structure appeared in Gelfand et al. (1991); application to ordered slopes in a change-point regression model was given in Carlin et al. (1992). Extensions to other problems will be obvious from the following development.

Consider then a model of the form

$$Y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}, \qquad i = 1, 2, \ldots, I; \quad j = 1, 2, \ldots, J, \tag{13}$$

where the ϵ_{ij} are independent $N(0, \sigma^2)$ and prior knowledge about the linear parameters constrains the α_i to be decreasing in i and the β_j to be increasing up to some unknown level t and then decreasing. Such a model generalizes the "response surface" priors discussed in Smith (1973) and finds application in many contexts where factor levels correspond to increasing (decreasing) levels of a treatment, fertilizer, and so forth. Other applications occur in consumer preference studies (Green 1974; Green and Srinivasan 1978): here Y_{ij} might be a scoring or rating of a product, such as a candy bar, with factor α_i corresponding to price level and β_j to sugar content.

The discussion in the previous sections indicates the obvious way to proceed. We place a multivariate normal prior on the set of α_i, β_j, independent of the ϵ_{ij}, *ignoring the order restrictions.* To complete the Bayesian specification we place, say, an inverse Gamma prior on σ^2 and a discrete distribution on t. Simple conjugate analysis (using, for example, the algebraic forms given in Lindley and Smith [1972]) straightforwardly reveals the full conditionals for the α_i and β_j to be univariate normals suitably constrained (the constraints for β_j being dependent on t). The full conditional for σ^2 is the conjugately updated inverse Gamma, whereas the full conditional for t is obtained using the technique described in the previous section. Full details on the required sampling in the context of an illustrative example are given in Section 4.2.

3.4 Ordered and Data Constrained Parameters

To illustrate a situation in which the constraint set $S_{\mathbf{Y}}^k$ discussed in Section 2.1 depends on $\mathbf{Y}$, consider the following model, which has applications to reliability development studies and survival analysis. We suppose that Y_{ij}, $i = 1, 2, \ldots, k$, and $j = 1, 2, \ldots, n_i$ are conditionally independent observations from location and scale exponential models, so that Y_{ij} has density

$$f(Y_{ij}|\theta_i, \sigma_i) = \frac{1}{\sigma_i} \exp\{-(Y_{ij} - \theta_i)/\sigma_i\},$$

$$Y_{ij} \geq \theta_i > 0, \ \sigma_i > 0.$$

In the absence of order restriction among the parameters, there has been recent decision-theoretic discussion of simultaneous point estimation of the location parameters in such models assuming known scale parameters and estimation of scale parameters assuming known location parameters. See, for example, Ebrahimi and Hosmane (1988) and Das Gupta, Dey, and Gelfand (1988).

Here we shall complete a Bayesian model specification by assuming for purposes of illustration that $0 < \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$ are the k order statistics from the exponential density $\lambda^{-1}\exp\{-\theta/\lambda\}$, with λ known, and that the σ_i are iid from IG(a, b), the inverse Gamma density $[b^a/\Gamma(a)] \times [\exp\{-b|\sigma_i\}/\sigma_i^{a+1}]$, with a, b known. We are interested in obtaining the marginal posterior densities for the θ_i and σ_i (or functions thereof), a problem that is extremely difficult using standard Monte Carlo integration due to the awkward nature of $S_{\mathbf{Y}}^k$, defined by $Y_{ij} \geq \theta_i > 0$ and $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Approached via the Gibbs sampler, however, the analysis is very straightforward. In particular, consider the full conditional distributions for the θ_i and σ_i. The σ_i are conditionally independent with

$$[\sigma_i|\mathbf{Y}, \theta] = \text{IG}(a + n_{i/2}, b + n_i(\bar{Y}_i - \theta_i)),$$

where $\bar{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$. For θ_i we have

$$[\theta_i|\mathbf{Y}, \sigma, \theta_j, j \neq i] \propto e^{-\theta_i(1/\lambda - 1/\sigma_i)}$$

restricted to the interval $\theta_{i-1} \leq \theta_i \leq \min(\min_j Y_{ij}, \theta_{i+1})$, where $\theta_0 \equiv 0$, $\theta_{k+1} \equiv \infty$. Thus these full conditional distributions are easily sampled, and the Bayesian analysis can be implemented straightforwardly.

3.5 Straight-line Regression With Censoring

As a first illustration of a truncated data problem, consider the special case of the structure introduced in Section 2.2 where $[\mathbf{Y}|\theta]$ corresponds to conditionally independent straight-line observations generated from $Y_{ij} = \alpha + \beta X_i + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2)$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n_i$, and $\mathbf{Z}$ is defined by

$$Z_{ij} = Y_{ij} \quad \text{if} \quad Y_{ij} \leq d_i$$
$$= d_i \quad \text{if} \quad Y_{ij} > d_i.$$

Thus at each setting X_i of the regression variable there is a cutoff d_i above which the value of Y_{ij} cannot be observed. An application of this model was given by Schmee and Hahn (1979), and a Bayesian analysis using adaptive Gauss–Hermite quadrature was given by Naylor and Smith (1982), who noted various subtleties required in performing the numerical integration.

In contrast implementing the Gibbs sampler using the approach set out in Section 2.2 is straightforward. We include the unobserved Y_{ij} (i.e., those where $Y_{ij} > d_i$) as further unknowns in the model. Then, given conjugate normal priors for α, β and an inverse Gamma prior for σ^2, it is clear that the full conditionals for α, β and σ^2 are the updated conjugate forms obtained by standard Bayesian analysis *assuming all the Y_{ij} to be observed.* The full conditional for any unobserved Y_{ij} is simply $N(\alpha + \beta X_i, \sigma^2)$, restricted to the range $Y_{ij} > d_i$. The required sampling from all full conditionals is therefore immediate.

3.6 Bivariate Grouped Data

Suppose that data from an underlying continuous bivariate distribution have been grouped into an $I \times J$ table, and we wish to make inferences about the parameters of an assumed bivariate parametric form for the unobserved continuous data. In what follows we shall assume, for illustration, an underlying bivariate normal population of the form

$$\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = N\left(\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right). \tag{14}$$

For convenience of nomenclature, we shall refer to the two component variables as "height" and "weight," with height groups $[a_{i-1}, a_i]$, $i = 1, \ldots, I$, and weight groups $[b_{j-1}, b_j]$,

Table 1. Multinomial Population and Generated Data

	P_i							
	.03	.07	.10	.25	.30	.12	.08	.05
Y_i	1	4	1	12	13	4	4	1

$j = 1, \ldots, J$, where $a_0 = b_0 = 0$ (which technically should be $-\infty$), $a_I = b_J = \infty$. The data consist of counts n_{ij}, $i = 1, \ldots, I, j = 1, \ldots, J$, where n_{ij} denotes the observed number of individuals in height group $[a_{i-1}, a_i]$ and in weight group $[b_{j-1}, b_j]$. If $\sum_i \sum_j n_{ij} = n$ the n_{ij} are distributed as Mult$(n; p_{11}, p_{12}, \ldots, P_{IJ})$, where $p_{ij} = p_{ij}(\theta, \Sigma) = \Pr(a_{i-1} \le X \le a_i, b_{j-1} \le Y \le b_j)$ under (14). Let θ and Σ denote the mean and covariance in (14). For illustration we adopt a normal-inverse Wishart prior structure for θ and Σ; that is, $[\theta] = N(\mu, V)$ and $[\Sigma^{-1}] = W((\rho R)^{-1}, \rho)$. We seek the marginal posterior distributions $[\theta_x|\mathbf{n}]$, $[\theta_y|\mathbf{n}]$, and $[\Sigma|\mathbf{n}]$, where $\mathbf{n} = (n_{11}, n_{12}, \ldots, n_{IJ})$.

The Gibbs sampler is implemented most easily if we include the $\mathbf{T}_s = \binom{X_s}{Y_s}$, $s = 1, \ldots, n$ in the model as unobservables. We then require the full conditional distributions for θ_x, θ_y, Σ and $\mathbf{T} = (\mathbf{T}_1, \ldots, \mathbf{T}_n)$. But $[\theta|\mathbf{T}, \mathbf{n}, \Sigma]$ is an updated bivariate normal that readily yields the full conditional distributions for θ_x and θ_y; similarly, $[\Sigma|\mathbf{T}, \mathbf{n}, \theta]$ is an updated inverse Wishart. Finally, we need to generate $\mathbf{T}_s$, $s = 1, \ldots, n$ given $\mathbf{n}$, θ and Σ. But this merely requires that for each pair i, j, we generate n_{ij} independent observations from (14) restricted to $[a_{i-1}, a_i] \times [b_{j-1}, b_j]$. Each such generation can be implemented by drawing X from $N(\theta_x, \sigma_x^2)$ restricted to $[a_{i-1}, a_i]$ and then Y given $X = x$ from $N(\theta_y + \sigma_{xy}(x - \theta_x)/\sigma_x^2, \sigma_y^2 - \sigma_{xy}^2/\sigma_x^2)$ restricted to $[b_{j-1}, b_j]$.

We note the obvious extension to higher dimensional tables arising from an underlying multivariate normal model. Another interesting extension arises if we have a collection of independent two-way tables arising from a third classification variable; that is, product multinomial sampling (Bishop, Fienberg, and Holland 1975). To be concrete, suppose this third variable is age and that the bivariate groups actually correspond to height and weight. That is, grouped height and weight data is supplied (using the same groups) for a sample of say five-year-old children, a sample of six-year-old children, and so forth. Under (14) it seems reasonable that both θ_x and θ_y should increase with age. Thus we have both grouped data and ordered parameters within one model. We leave details of this extension to the reader, who by now will be surprised to find that the Gibbs sampler is very straightforward despite the seeming awkwardness of the model and parameter constraints.

4. ILLUSTRATIVE ANALYSIS

In this section we analyze two artificial data sets derived from models based on those discussed in Sections 3.2 and 3.3. Real applications of these and the other models discussed exist in abundance. Our purpose in analyzing artificial data sets generated from known models is to provide insight into and calibration of the performance of the methodology we have presented.

4.1 Multinomial With Ordered Parameters

As an example of the problem discussed in Section 4.2, Table 1 shows a $k = 8$ cell multinomial model and the results of 40 random draws from this model. We assume that the p_i increase $i = 1, 2, \ldots, t$ and then decrease thereafter but otherwise p_i and t are unknown. We take the generalized uniform Dirichlet, $\alpha_i = 1$, $i = 1, 2, \ldots, 8$, and calculate the constants $c(1, \ldots, 1; t)$, $c(Y_1 + 1, \ldots, Y_8 + 1; t)$, for $t = 1, 2, \ldots, 8$ as described in Sedransk et al. (1985). Using (13) we obtain the marginal posterior, $[t|\mathbf{Y}]$, which is shown in Table 2. Note that despite only 40 draws from an eight-cell table and a flat prior, $[t|\mathbf{Y}]$ places nearly all its mass on $t = 4$ and 5.

As remarked in Section 3.2, to obtain the marginal posteriors, $[p_i|\mathbf{Y}]$, we implement the Gibbs sampler in a slightly different way. We use general k and $\alpha = (\alpha_1, \ldots, \alpha_k)$ in the ensuing details. Because

$$[p_i|\mathbf{Y}] = \sum_{l=1}^{k} [p_i|\mathbf{Y}, t][t|\mathbf{Y}] \tag{15}$$

and because $[t|\mathbf{Y}]$ already has been obtained, we propose to sample from $[p_i|\mathbf{Y}]$ by randomly selecting t according to $[t|\mathbf{Y}]$ and then sampling p_i from $[p_i|\mathbf{Y}, t]$. The densities $[p_i|\mathbf{Y}, t]$ can be obtained using the Gibbs sampler in the customary fashion, as indicated in Section 3.3. More precisely, we require only the full conditional distributions for p_i, $i = 1, \ldots, k-1$, because $p_k = 1 - \sum_{i=1}^{k-1} p_i$. But $[p_i|\mathbf{Y}, t, p_j, j = 1, 2, \ldots, k-1, j \ne i] = \text{Beta}(\alpha_i + Y_i, \alpha_k + Y_k)$ scaled to the interval $[0, a_i]$, where $a_i = 1 - \sum_{j=1}^{k-1} p_j, j \ne i$, and then restricted to an interval determined by the other p_j's and t; that is, with $p_o \equiv 0$

if $i < t$, $\max(p_{i-1}, a_i - p_{k-1}) \le p_i \le \min(p_{i+1}, a_i)$;

if $i > t$, $\max(p_{i+1}, a_i - p_{k-1}) \le p_i \le \min(p_{i-1}, a_i)$;

if $i = t$, $\max(p_{t-1}, p_{t+1}, a_t - p_{k-1}) \le p_t \le a_t$.

The output from m replications of the Gibbs sampler will be vectors $\mathbf{p}_j^t = (p_{ij}^t, \ldots, p_{kj}^t)$ and $j = 1, 2, \ldots, m$, such that the $\mathbf{p}_j^t$ are approximately distributed as $[\mathbf{p}|\mathbf{Y}, t]$; thus the p_{ij}^t, $j = 1, \ldots, m$, are approximately distributed as $[p_i|\mathbf{y}, t]$.

Table 2. Marginal Posterior Distribution of t

	t							
	1	2	3	4	5	6	7	8
$p(t|\mathbf{Y})$	.0000	.0001	.0013	.3527	.6350	.0104	.0005	.0000

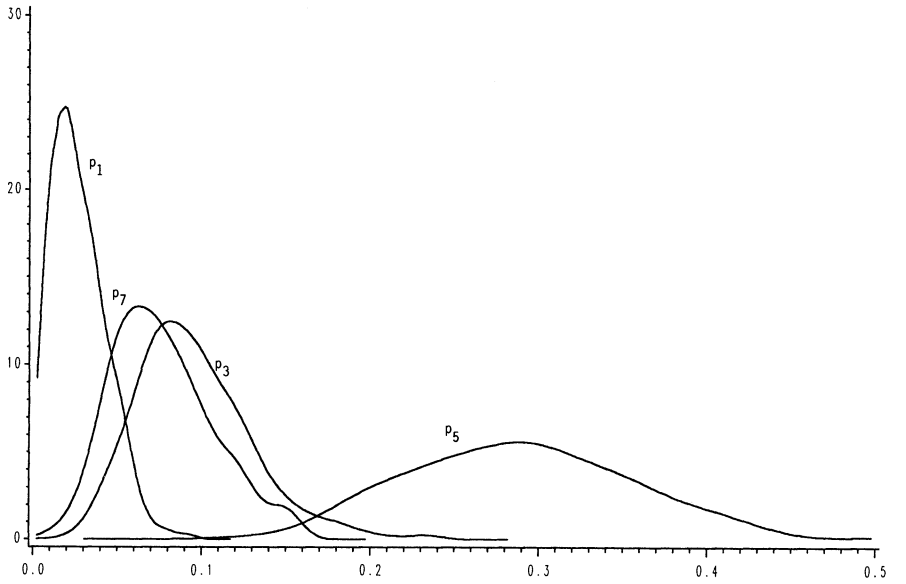

Figure 1. Multinomial Model With Ordered Parameters (Section 4.1). Marginal posterior distributions for selected cell probabilities.

Suppose we run the Gibbs sampler in this manner for each $t, t = 1, \ldots, k$. Then in theory we could obtain a kernel density estimate for each $[p_i|\mathbf{Y}, t]$, $t = 1, \ldots, k$ and thus via (15) a density estimate that is a finite mixture of these. In practice we would merely randomly select t according to $[t|Y]$ and then make an equally likely choice from the set of p_{ij}^t, $j = 1, \ldots, m$. This resampling procedure results in an observation approximately distributed as $[p_i|\mathbf{Y}]$. Repeating this process many times (1,000 times to create the plots in Table 2), we can compute a kernel density estimate for $[p_i|Y]$.

Returning to the analysis at the beginning of this section, in Figure 1 we plot such kernel density estimates for the illustrative set p_1, p_3, p_5, p_7. We note that these posteriors reflect the order restrictions and have modes close to the respective true values. The complete set of posterior modes is given in Table 3.

4.2 Two-Way Layout With Ordered Parameters

Consider the problem discussed in Section 3.3. Table 4 presents a set of data, $\mathbf{Y}$, generated from (13) with $\alpha_1 = 2$, $\alpha_2 = 1$, $\alpha_3 = 0$, $\alpha_4 = -2$, $\beta_1 = -1$, $\beta_2 = 0$, $\beta_3 = 2$, $\beta_4 = -1$, $\beta_5 = -2$, and $\sigma^2 = 3$. Thus for each column cell expectations decrease, whereas for each row expectations increase to the middle column and then decrease. The data is rather noisy and often at odds with these expectations. Ordinary least squares (OLS) analysis ignoring known order restrictions is terribly misleading: $\hat{\alpha}_1 = 1.064$, $\hat{\alpha}_2 = -1.163$, $\hat{\alpha}_3 = .536$, $\hat{\alpha}_4 = -5.203$, $\hat{\beta}_1 = -1.737$, $\hat{\beta}_2 = .758$, $\hat{\beta}_3 = .283$, $\hat{\beta}_4 = -3.344$, $\hat{\beta}_5 = -1.917$, and $\hat{\sigma}^2 = 3.590$. This analysis yields estimates for the α_i and β_j that fail to meet the restrictions and often are far from the true values. Some sort of constrained least squares solution (an isotonic regression) would be a better frequentist approach.

Bayesian analysis using the Gibbs sampler is implemented easily in this case, yielding marginal posterior distributions for the α_i, the β_j, and σ^2. In the process, using say posterior modes, the isotonic regression problem is solved.

Specific details are as follows. Suppose for simplicity we assume conjugate normal and inverse Gamma forms for the α_i, β_j and for σ^2. That is, ignoring restrictions, let α_i be iid $N(0, \sigma_\alpha^2)$ and β_j be iid $N(0, \sigma_\beta^2)$. (For convenience we have centered these priors at 0 and have chosen the above α_i, β_j to be approximately centered at 0 as well.) Let $\sigma^2 \sim \mathrm{IG}(a, b)$ independent of the α_i and β_j. We make these priors rather vague by setting $\sigma_\alpha^2 = 5$, $\sigma_\beta^2 = 5$, $a = 0$, and $b = 1$. The full conditional distributions using general σ_α^2, σ_β^2, a, b and incorporating the order restrictions are

$$[\alpha_i|\mathbf{Y}, \alpha_r, r \neq i, \beta_j, \sigma^2] = N\left(\frac{5\sigma_\alpha^2(\bar{Y}_{i\cdot} - \beta_\cdot)}{5\sigma_\alpha^2 + \sigma^2}, \frac{\sigma_\alpha^2\sigma^2}{5\sigma_\alpha^2 + \sigma^2}\right), \qquad i = 1, 2, 3, 4$$

restricted to $(\alpha_{i-1}, \alpha_{i+1})$, where $\alpha_0 \equiv -\infty$, $\alpha_5 \equiv +\infty$, $\bar{Y}_{i\cdot} = \sum_{j=1}^5 Y_{ij}/5$, and $\beta_\cdot = \sum_{j=1}^5 \beta_j/5$;

Table 3. Marginal Posterior Modes of the $[p_i|\mathbf{Y}]$

	i							
	1	2	3	4	5	6	7	8
Mode of $[p_i\|\mathbf{Y}]$	.019	.061	.082	.246	.289	.096	.063	.019

Table 4. Generated Two-Way Layout Data

	i/j				
	1	2	3	4	5
1	.982	1.902	3.797	−1.531	.570
2	−1.417	1.356	1.287	−3.629	−3.413
3	−1.601	4.713	.814	.834	−2.082
4	−4.912	−4.541	−4.768	−9.051	−2.744

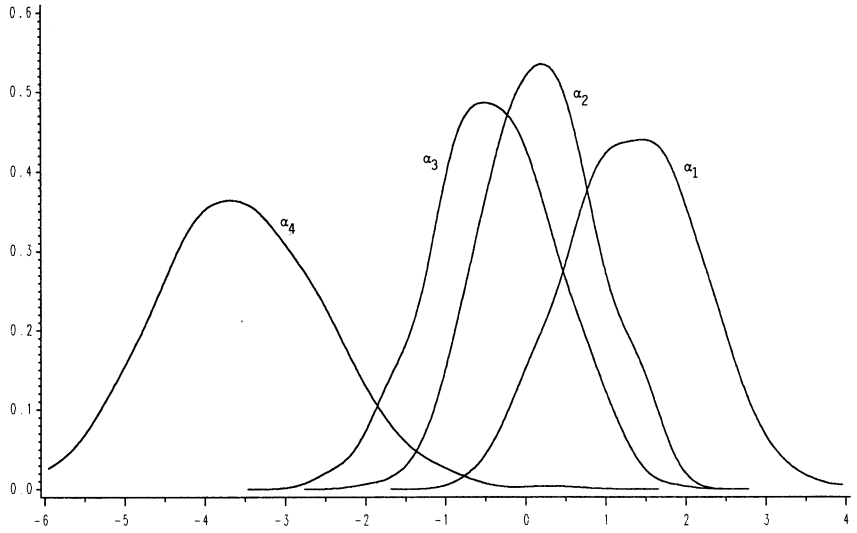

Figure 2. Two-Way Layout With Ordered Parameters (Section 4.2). Marginal posterior distributions for the first factor effects.

$$[\beta_j|\mathbf{Y}, \beta_s, s \neq j, \alpha_i, \sigma^2]$$

$$= N\left(\frac{4\sigma_\beta^2(\bar{Y}_{\cdot j} - \alpha_\cdot)}{4\sigma_\beta^2 + \sigma^2}, \frac{\sigma_\beta^2\sigma^2}{4\sigma_\beta^2 + \sigma^2}\right), \qquad j = 1, 2, 3, 4, 5$$

restricted to $(\beta_{i-1}, \beta_{i+1})$ $j = 1, 2$, $(\beta_{j+1}, \beta_{j-1})$, $j = 4, 5$, $[\max(\beta_2, \beta_4), \infty)$, and $j = 3$, where $\beta_0 \equiv -\infty$, $\beta_6 \equiv +\infty$, $\bar{Y}_{\cdot j} = \sum_{i=1}^4 Y_{ij/4}$, and $\alpha_\cdot = \sum_{i=1}^4 (\alpha_{i/4})$;

$$[\sigma^2|\mathbf{Y}, \alpha_i, i = 1, \ldots, 4, \beta_j, j = 1, \ldots, 5]$$

$$= \mathrm{IG}\left(a + 10, b + \sum_i \sum_j (Y_{ij} - \alpha_i - \beta_j)^2 \Big/ 2\right).$$

Figure 2 shows the marginal posterior distribution for the α_i as output from running the Gibbs sampler, revealing that they respond to the order restrictions. Similar figures may be developed for the β_j and for σ^2. The marginal posterior modes $\alpha_1^* = 1.480$, $\alpha_2^* = .197$, $\alpha_3^* = -.507$, $\alpha_4^* = -3.684$, $\beta_1^* = -1.039$, $\beta_2^* = .635$, $\beta_3^* = 1.261$, $\beta_4^* = -1.149$, $\beta_5^* = -1.790$, and $\sigma^{2^*} = 3.975$ are in accordance with the restrictions and generally much closer to the true values than are the OLS estimates. Although some version of a constrained least squares algorithm might produce comparably good point estimates, because the Gibbs sampler enables marginal posterior distributions for the α_i, β_j as well as for $\alpha_i + \beta_j$, $\alpha_r - \alpha_s$, $\beta_r - \beta_s$, and so forth, we also can easily obtain Bayesian interval estimates for any functions of the model parameters that may be of interest.

In these calculations we have assumed that β_3 was known to be the largest β. If we did not know which subscript denoted the largest β, we could use an approach analogous to that described in the previous example. If we felt that the data contained some outliers, we might robustify the Bayesian analysis by assuming that the distribution of the errors in (14) is $[\varepsilon_{ij}|\lambda_{ij}] = N(0, \lambda_{ij}^{-1}\sigma^2)$, where $\nu\lambda_{ij} \sim \mathrm{Gamma}(\nu/2, 1/2)$, so that marginally the $\varepsilon_{ij} \sim t_\nu(0, \sigma^2)$. To implement the Gibbs sampler we would include the λ_{ij} as unobserved variables. Obtaining all full conditional distributions would be straightforward; we omit details.

5. SUMMARY

Our intent has been to describe how Bayesian analysis of a broad range of ordered parameter and truncated data problems can be implemented straightforwardly using the Gibbs sampler. This approach avoids well-nigh impossible numerical integrations over high dimensional sets defined by complex restrictions. Rather, it requires only sampling from univariate full conditional distributions, restricted to easily described subsets of R^1. With conjugacy, the needed full conditional distributions are standard probability distributions; without conjugacy, tailored versions of general random variate generation procedures are needed. Although sampling may be inefficient, this is more than compensated for by the ability to carry out full Bayesian calculations for many problems that were previously inaccessible. Two illustrative examples show how much stronger inference can be when restrictions are taken into account in the modeling process.

[*Received March 1990. Revised December 1990, May 1991.*]

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference Under Order Restrictions,* New York: John Wiley.

Bishop, Y., Fienberg, S., and Holland, P. (1975), *Discrete Multivariate Analysis: Theory and Practice,* Cambridge, MA: MIT Press.

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

Broffitt, J. D. (1984), "A Bayes Estimator for Ordered Parameters and Isotonic Bayesian Graduation," *Scandinavian Actuarial Journal,* 4, 231–247.

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian Analysis of Change Point Problems," *Applied Statistics,* 41, 389–405.

Das Gupta, A., Dey, D. K., and Gelfand, A. E. (1988), "A New Admissibility Theorem With Applications To Estimation of Survival and Hazard Rates and Means in the Scale Parameter Family," *Sankhyā A,* 50, 269–281.

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

Ebrahimi, N., and Hosmane, B. (1988), "On Shrinkage Estimation of the Exponential Location Parameter," *Communications in Statistics A,* 16, 2623–2637.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* 85, 398–409.

——— (1991), "Gibbs Sampling for Marginal Posterior Expectations," *Communications in Statistics B,* 20, 1747–1766.

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," *Journal of the American Statistical Association,* 85, 972–985.

Geman, D., and Geman, S. (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6, 721–741.

Green, P. E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," *Journal of Consumer Research,* 1, 61–68.

Green, P. E., and Srinivasan, V. (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research,* 5, 103–123.

Hastings, W. K. (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika,* 57, 97–109.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 34, 1–41.

Naylor, J. C., and Smith, A. F. M. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics,* 31, 214–225.

Ripley, B. (1987), *Stochastic Simulation,* New York; John Wiley.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order-Restricted Statistical Inference,* New York: John Wiley.

Schmee, J., and Hahn, G. J. (1979), "A Simple Method for Regression Analysis With Censored Data," *Technometrics,* 21, 417–432.

Sedransk, J., Monahan, J., and Chiu, H. Y. (1985), "Bayesian Estimation of Finite Population Parameters in Categorical Data Models Incorporating Order Restrictions," *Journal of the Royal Statistical Society,* Ser. B, 47, 519–527.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis,* London: Chapman & Hall.

Smith, A. F. M. (1973), "Bayes Estimates in One-Way and Two-Way Models," *Biometrika,* 60, 319–329.