BAYESIAN ANALYSIS OF CONSTRAINED PARAMETER AND TRUNCATED DATA PROBLEMS

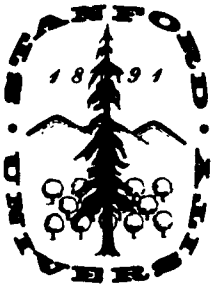
BY

A. E. GELFAND, A. F. M. SMITH and T-M. LEE

TECHNICAL REPORT NO. 439
JANUARY 4, 1991

PREPARED UNDER CONTRACT
N00014-89-J-1627 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



BAYESIAN ANALYSIS OF CONSTRAINED PARAMETER AND TRUNCATED DATA PROBLEMS

BY

A. E. GELFAND, A. F. M. SMITH and T-M. LEE

TECHNICAL REPORT NO. 439
JANUARY 4, 1991

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction

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, parameter constraints and data truncation are both present.

The statistical literature on such problems is very extensive, reflecting both their widespread occurrence in applications and the methodological challenges which they pose. However, it is striking that hardly any of this applied and theoretical literature involves a parametric Bayesian perspective (although some exceptions will be noted in the context of the examples given later, in Section 4). From a technical perspective, this is perhaps not difficult to understand. The fundamental tool for Bayesian calculations in typical realistic models is (multi–dimensional) numerical integration. This is often problematic in unconstrained contexts; it can be well–nigh impossible for the kinds of constrained problems we shall be considering.

Our goal in this paper is to show that Bayesian calculations <u>can</u> be routinely implemented for constrained parameter and truncated data problems by means of the Gibbs sampler (introduced by Geman and Geman, 1984, in the context of image processing and subsequently proposed as a general method for Bayesian calculations by Gelfand and Smith, 1990).

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

In Section 2, we briefly review the Gibbs sampler and comment on experience with its use for other classes of statistical problems. In Section 3, 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 4, we develop detailed analyses for a variety of examples. These are chosen with a view to giving the reader an appreciation 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 5, we provide illustrative analyses of two artificial data sets, generated from models chosen

to present extremely awkward inference problems. Finally, in Section 6 we provide a summary discussion.

2. The Gibbs Sampler

Our subsequent development will use the following notational conventions. Densities will be developed generically by square braclets, so that joint, conditional and marginal forms for random variables U,V, appear, respectively, 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 $[U_1, U_2,...,U_k]$ the full conditional densities can thus be denoted by $[U_s|U_r, r \neq s]$, $s = 1,2,...,k$, and the marginal densities by $[U_s]$, $s = 1,2,...,k$.

Suppose we now consider the following problem. 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, $r\neq s$, $s = 1,2,...,k$, can we find an iterative scheme which enables us to make sample–based estimates, $[\hat{U}_s]$, say, of the marginal densities, $[U_s]$, $s = 1,2,...,k$? We shall make the connection with Bayesian posterior calculations later; for the moment, we note that the general question is answered affirmatively by the following procedure.

Gibbs sampling is a Markovian updating scheme which proceeds as follows. Given an arbitrary starting set of values $U_1^{(0)},...,U_k^{(0)}$, we draw $U_1^{(1)}$ from $[U_1|U_2^{(0)},...,U_k^{(0)}]$, then $U_2^{(1)}$ from $[U_2|U_1^{(1)}, U_3^{(0)},...,U_k^{(0)}]\cdots$ and so on up to $U_k^{(1)}$ from $[U_k|U_1^{(1)}...,U_{k-1}^{(1)}]$ to complete one iteration of the scheme. After t such iterations we would arrive at joint a sample $(U_1^{(t)},...,U_k^{(t)})$. Geman and Geman show under mild conditions that $(U_1^{(t)},...,U_k^{(t)} \xrightarrow{d} (U_1,...,U_k) \sim [U_1,U_2,...,U_k]$ as $t \to \infty$. Hence for t large enough, $U_s^{(t)}$ for example can be regarded as a sample variate from $[U_s]$. Parallel replication of this process m times yields m iid k–tuples $(U_{1j}^{(t)},...,U_{kj}^{(t)})$ $j = 1,2,...,m$. Note that sample size at, say, the w–th iteration may be increased from m to any specified size by randomly reusing the $U_{sj}^{(w-1)}$ with replacement.

A kernel density estimate for $[U_s]$ based upon the $U_{sj}^{(t)}$ can be readily obtained (see e.g., Silverman, 1986) and should be adequate if, a the last iteration the number of replications, m, is large enough. However, using a Rao Blackwell argument (see Gelfand

and Smith, 1990) a density estimate of the form

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

is better under a wide range of loss functions. This is not surprising since (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 and is 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 upon the $U_{sj}^{(t)}$ or possibly as a "Rao–Blackwellized" version analogous to (1) based upon $E(h(U_s)|U_r, r \neq s)$. Now consider a function of the U_i, say $W(U_1,...,U_k)$. Each k–tuple, $(U_{1j}^{(t)},...,U_{kj}^{(t)})$, provides an observed $W_j^{(t)} \equiv W(U_{1j}^{(t)},...,U_{kj}^{(t)})$ whose marginal distribution, as $t \to \infty$, is approximately $[W]$, whence a kernel density estimator for $[W]$ using these $W_j^{(t)}$ can be developed. If, say, U_s actually appears as an argument of W 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) can also 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) which 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 full 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 employ tailored versions of general random variate generating procedures such as the ratio–of–uniforms and rejection methods (see, for example, Devroye, 1986, or Ripley, 1986).

Finally, we note that complete implementation of the Gibbs sampler requires that a determination of t be made and that across iterations, choice(s) of m be specified. In a

challenging application some experimentation with different settings for t and m will likely be necessary. We do not view this as a deterrent since random generation is generally inexpensive and since in many cases there may be no feasible alternative. In the examples of Section 5, convergence was evaluated in a univariate manner by plotting marginal posterior density estimates of the form (1) five iterations apart in order to judge stability. Typically, a somewhat small m is used until convergence is concluded, at which point, for a final iteration, m is increased by an order of magnitude to develop the presented density estimate. We make no claims for the optimality of this procedure. Assessment of convergence is a complex issue which is currently very much in the empirical domain.

However, extensive computational experience in a wide variety of parametric models has given us considerable confidence in the pragmatic procedure described above. See, for example: Gelfand and Smith (1989, 1990); Gelfand et al (1989); Carlin et al (1989); Racine–Poon et al (1990).

Since the cited papers contain extensive discussion of very detailed specification of the Gibbs sampler in a number of situations, we avoid unnecessary detail in what follows, concentrating instead on the structural insights which underlie the specification of the required full conditionals. Having followed the general discussion, the reader can easily supply the missing detail in any specific example.

3. 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. As indicated in Section 2, the implementation problem reduces to identification of the appropriate full conditional distributions and methods for drawing samples from them.

3.1. Constrained Parameter Models

Consider a parametric model for data $\underset{\sim}{Y}$ involving a k–dimensional parameter vector $\underset{\sim}{\theta}$, constrained to lie in a subset $S_{\underset{\sim}{Y}}^{k}$ of R^k. Often the constraint set $S_{\underset{\sim}{Y}}^{k}$ is determined by order or other inequality relationships among the components θ_i, i =

1,2,...,k, of $\underset{\sim}{\theta}$, in which case, $S^k = S^k_{\underset{\sim}{Y}}$ does not depend on $\underset{\sim}{Y}$. In other cases, constraints occur because the region of positive support for the distribution of $\underset{\sim}{Y}$ depends on $\underset{\sim}{\theta}$, so that $\underset{\sim}{Y}$ occurs explicitly in $S^k_{\underset{\sim}{Y}}$: see, for example, Section 4.5, where $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_k)$ and $\theta_i \leq Y_i$, i = 1,2,...,k. In the former case, it is natural to think of the constraint as built into the specification of the prior distribution, $[\underset{\sim}{\theta} | \underset{\sim}{\lambda}]$, where $\underset{\sim}{\lambda}$ is some hyperparameter; in the latter case, it is natural to think of the constraint as built into the likelihood, $[\underset{\sim}{Y} | \underset{\sim}{\theta}]$. In either case, it suffices to note that the constrained Bayesian model (likelihood x prior) is given by

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

where $S = \{(y,\theta): \theta \epsilon S^k_y\}$. In general, $[\underset{\sim}{Y} | \underset{\sim}{\theta}]$, $[\underset{\sim}{\theta} | \underset{\sim}{\lambda}]$, as functions of θ have the functional forms they would have had if constraints were ignored. It follows immediately (generalizing slightly a remark in Box and Tiao, 1973, p. 67) that the posterior distribution for $\underset{\sim}{\theta}$, given the constraints, is simply the unconstrained posterior appropriated normalized so that

$$[\underset{\sim}{\theta} | \underset{\sim}{Y}] = \frac{[\underset{\sim}{Y} | \underset{\sim}{\theta}] \cdot [\underset{\sim}{\theta} | \underset{\sim}{\lambda}]}{\int_{S^k_{\underset{\sim}{Y}}} [\underset{\sim}{Y} | \underset{\sim}{\theta}] \cdot [\underset{\sim}{\theta} | \lambda]} \qquad \underset{\sim}{\theta} \epsilon S^k_{\underset{\sim}{Y}} \tag{2}$$

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

$$[\theta_i | \underset{\sim}{Y}, \underset{\sim}{\lambda}, \theta_j, j\neq i] \propto [\underset{\sim}{Y} | \underset{\sim}{\theta}] \cdot [\underset{\sim}{\theta} | \underset{\sim}{\lambda}], \quad \theta_i \epsilon S_i^k(\theta_j, j\neq i), \tag{3}$$

where the right–hand side is regarded as a function of θ_i for specified θ_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) will then simply 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 $S_{\underset{\sim}{Y}}^k$ is, to implement the Gibbs sampler we need only consider $S_{\underset{\sim}{Y}}^k$ 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, [a,b], say. Then, if U is a uniform (0,1) variate, it is noted by Devroye (1986, p.38) that $\theta_i = F_i^{-1}[F_i(a) + U(F_i(b) - F_i(a))]$ is a drawing from the constrained full conditional. Thus we sample "one–for–one" from the constrained full conditional. This is easily extended to the case where the cross–section is a collection of disjoint intervals, $\bigcup_{j=1}^{r}[a_j,b_j]$, say. 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 striking ease of implementation of the Gibbs sampler, enabling one to carry out full Bayesian calculations for complex constrained parameter problems which were previously 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 the above to a hierarchical model by adding a prior $[\underline{\lambda}]$ for the hyperparameter $\underline{\lambda}$, thus far assumed to be known. The full conditionals for the θ_i are unchanged and are given by (3). However, the full conditional for $\underline{\lambda}$ doesn't depend upon $\underline{Y}$ and takes the form

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

where $c(\lambda) = (\int_S [\underline{Y}\,|\,\underline{\theta}][\underline{\theta}\,|\,\underline{\lambda}])^{-1}$. If $\underline{\theta}$ is not constrained by $\underline{Y}$, $c(\lambda)$ simplifies to $(\int_k [\theta\,|\,\lambda])^{-1}$ but regardless $c(\lambda)$ will typically not be available explicitly (see, e.g. Sections 4.1, 4.2) making sampling from (4) almost impossible.

3.2 Censored Data Models

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

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

in terms of parameters $\underline{\theta}$ and $\underline{\eta}$, and define component–wise a further random n–vector $\underline{Z}$ by

$$Z_j = \begin{cases} V_j & Y_j \le V_j \\ Y_j & \text{if} \quad V_j < Y_j < W_j \\ W_j & Y_j \ge W_j \end{cases} \qquad j=1,2,\ldots,n \tag{5}$$

We shall consider $\underline{Z}$ to be observed data arising as a censored form of $\underline{Y}$ through the censoring process $[\underline{V},\underline{W}\,|\,\underline{\eta}]$, with $\underline{V}$ and $\underline{W}$ also observed. In this very general formulation, $\underline{V}$ and $\underline{W}$ 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 include as special cases.

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

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

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

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

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

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

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

$$[\underline{Z}|\underline{Y},\underline{V},\underline{W}][\underline{Y}|\underline{V},\underline{W},\underline{\theta}][\underline{V},\underline{W}|\eta][\underline{\theta}|\underline{\lambda}][\underline{z}][\underline{\lambda}] \tag{7}$$

Here $[\underline{Z}|\underline{Y},\underline{V},\underline{W}]$ is, of course, a degenerate distribution and in typical applications we

shall have $[\underline{Y}|\underline{V},\underline{W},\underline{\theta}] = [\underline{Y}|\underline{\theta}]$. Note that now

$$[\theta_i|\underline{Z},\underline{Y},\underline{V},\underline{W},\theta_j, j\neq i, \underline{\eta},\underline{\lambda}] \propto [\underline{Y}|\underline{V},\underline{W},\underline{\theta}]\,[\underline{\theta}|\underline{\lambda}] \tag{8}$$

The right hand side of (8) is now an explicit function of θ_i and sampling no longer presents a problem. As we remarked in the previous section, under conjugacy, sampling from θ_i will simply involve sampling from a standard distribution. Without conjugacy 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 $[\underline{\eta}|\underline{\lambda},\underline{Y},\underline{V},\underline{W},\underline{\theta},\underline{\lambda}] \propto [\underline{V},\underline{W}|\underline{\eta}][\underline{\eta}]$, and $[\underline{\lambda}|\underline{Z},\underline{Y},\underline{V},\underline{W},\underline{\theta},\underline{\eta}] \propto [\underline{\theta}|\underline{\lambda}][\underline{\lambda}]$ to which similar remarks apply, and finally $[\underline{Y}|\underline{Z},\underline{V},\underline{W},\underline{\theta},\underline{\eta},\underline{\lambda}] \propto [\underline{Z}|\underline{Y},\underline{V},\underline{W}] \cdot [\underline{Y}|\underline{V},\underline{W},\underline{\theta}]$. Again, for illustration, consider the typical case where $[\underline{Y}|\underline{V},\underline{W},\underline{\theta}] = [\underline{Y}|\underline{\theta}] = \prod_{j=1}^{n} f_j(Y_j|\underline{\theta})$. Then the Y_j are conditionally independent and the full conditional distribution for Y_j has the following form: it is degenerate at Z_j if $V_j < Z_j < W_j$; it has the distribution $f_j(\cdot|\underline{\theta})$ restricted to $(-\infty, V_j]$ if $Z_j = V_j$, and has the distribution $f_j(\cdot|\underline{\theta})$ 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" methods described in Section 3.1.

3.3 Grouped Data Models

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

$$[\underline{S}|\underline{\theta}]\,[\underline{\theta}|\underline{\lambda}]\,[\underline{\lambda}],$$

where $[\underline{S}|\underline{\theta}]$ is induced by $[\underline{Y}|\underline{\theta}]$.

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

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

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

$$[\underline{S}|\underline{Y}]\,[\underline{Y}|\underline{\theta}]\,[\underline{\theta}|\underline{\lambda}]\,[\underline{\lambda}]$$

and the full conditionals are given by

$$[\underline{\lambda}|\underline{S},\underline{Y},\underline{\theta}] \propto [\underline{\theta}|\underline{\lambda}]\,[\underline{\lambda}]$$

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

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

$$[\underline{Y}|\underline{S},\underline{\theta},\underline{\lambda}] \propto [\underline{S}|\underline{Y}]\,[\underline{Y}|\underline{\theta}].$$

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

4. Models: Specific Examples

In this section, we make explicit the forms of the Gibbs sampler arising from a variety of examples of the general structures discussed in Section 3. Our development is designed to illuminate for the reader the astonishing simplicity with which the appropriately defined Gibbs sampler solves the problem of Bayesian computation in constrained parameter and truncated data contexts.

As we remarked in Section 1, there is remarkably little literature on Bayesian approaches to these problems and that which exists typically does not solve the problem of calculating marginal densities, but only attempts limited inference summaries in the form of modes or means. Ordered restricted inference is discussed at length from a frequentist perspective in the books by Barlow et al (1972) and Robertson et. al. (1988). The former has some discussion of Bayesian inference for ordered exponential family parameters, but this is largely limited to a discussion of the joint posterior mode as an isotonic regression. The latter provides 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.

4.1. Ordered Binomial Parameters

Suppose that we have conditionally independent observations Y_i - Binomial (n_i,p_i), $i = 1,2,...,k$, where it is known that $p_1 \le p_2 \le ... \le p_k$ and we seek to make inferences about the p_i (or functions thereof, such as $p_{i+1} - p_i$ or $(p_{i+1} - p_i)/p_i$). This problem is discussed in Broffitt (1987) and in Sharples (1988). It arises naturally in reliability development testing, where interest often focuses on p_k (see, for example, Smith, 1977, and Fard and Dietrich, 1987) and in bioassay, where $p_i = p(x_i)$ may be thought of in terms of a monotonic increasing relationship with dose or stimulus. The function $p(x)$ is referred to as a potency curve. See, for example, Kuo, 1988, for recent discussion and related references.

A flexible, convenient form of prior distribution over the simplex $S^k = \{(p_1,p_2,...,p_k) : 0 \le p_1 \le p_2 \le ... \le p_k \le 1\}$ takes the form

$$c_k(\alpha_1,\alpha_2,...,\alpha_k ; \beta_1,\beta_2,...,\beta_k) \prod_{i=1}^{k} p_i^{\alpha_i-1} (1-p_i)^{\beta_i-1}, \qquad (9)$$

where c_k is the normalizing constant and α_i,β_i are chosen to reflect prior beliefs. Note that this is equivalent to unconstrained p_i having been drawn independently from Beta(α_i,β_i) distributions, and that if $\alpha_i = \alpha$, $\beta_i = \beta$, it is equivalent to the p_i being order statistics from a sample of k iid Beta(α,β) variables. For integer α_i,β_i, the constant c_k can be obtained as a finite multidimensional summation (see Sharples, 1988).

By conjugacy, the joint posterior $[\underline{p}|\underline{Y}]$ has the same form as (9), but with the α_i,β_i replaced by $\alpha_i + Y_i$, $\beta_i + n_i - Y_i$, respectively. (Again, in the case of integer α_i,β_i the exact marginal posterior densities for the p_i can be identified as very complicated weighted averages of Beta distributions; see Sharples, 1988.) Implementation of the Gibbs sampler is extremely simple. From (9), we see immediately that, for $i = 1,2,...,k$, $[p_i|\underline{Y}, p_j, j\neq i] = \text{Beta}(\alpha_i + Y_i, \beta_i + n_i - Y_i)$ restricted to $p_{i-1} \le p_i \le p_{i+1}$, with

$p_0 = 0, p_{k+1} = 1$, so that sampling reduces to interval restricted sampling from a standard distribution, as discussed in Section 3.1. We need never concern ourselves with calculation of c_k.

4.2 Ordered Exponential Family Parameters

For conditionally independent observations from one–parameter exponential family models, with increasing parameters and a constrained form of conjugate prior, the analysis given in Section 4.1 generalizes in an obvious and straightforward way. The resulting full conditionals again reduce to interval restricted sampling from the standard posterior forms arising from the conjugate analysis.

Detailed developments are given by Broffit (1984), motivated by graduation problems in actuarial science. He considers ordered parameters from a family of models of the form

$$f(Y \mid \theta) = a(Y)\theta^{b(Y)} e^{-\theta c(Y)}, \quad \theta > 0 \tag{10}$$

which includes models such as gamma with known shape parameters , normal with known mean and Poisson.

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

$$d_k(\delta_1,...,\delta_k; \gamma_1,...,\gamma_k) \prod_{i=1}^{k} \frac{\theta_i^{\delta_i - 1}}{\gamma_i^{\delta_i}} \frac{t^{-\theta_i/\gamma_i}}{\Gamma(\delta_i)} \tag{11}$$

where d_k is the normalizing constant and δ_i, γ_i are chosen to reflect prior beliefs. Note that (11) arises from independent Gamma priors were the θ_i unconstrained. In the case where the δ_i are integers Broffitt obtains c_k as a finite multidimensional sum. The joint

posterior $[\underline{\theta}|\underline{Y}]$ has the same form as (11) with δ_i replaced by $\delta_i^* = \delta_i + \sum_{j=1}^{n_i} b(Y_{ij})$ and γ_i replaced by $\gamma_i^* = (\frac{1}{\gamma_i} + \sum_{j=1}^{n_i} c(Y_{ij}))^{-1}$. The posterior mean for θ_i under the unrestricted problem is $\theta_i^* = \delta_i^* \gamma_i^*$. Using isotonic regression Broffitt obtains the order restricted Bayes estimate for θ_i under squared error loss as a function of θ_i^* and two d_k's.

To implement the Gibbs sampler we only require the full conditional distribution $[\theta_i|\underline{Y},\theta_j, j\neq i]$, $i = 1,2,...,k$. Under (11), this is merely a Gamma (δ_i^*, γ_i^*) restricted to the interval $[\theta_{i-1}, \theta_{i+1}]$ where $\theta_0 \equiv 0$, $\theta_{k+1} \equiv \infty$. As in the previous example, we need never concern ourselves with calculation of the normalizing constant d_k.

4.3 Ordered Multinomial Parameters

Sedransk, et al (1985) discuss the problem of Bayes estimation of finite population parameters when a random variable X assumes one of a finite set of values $\{b_1,...,b_k\}$ with probabilities $p_1,p_2,...,p_k$, respectively. A particular application is the case of household income, where b_j might denote a central value for the j^{th} income category.

Assuming the categories to be 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 $\underline{p}$ could also 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 $[\underline{Y}|\underline{p}] = \text{Mult}(n; p_1, p_2,...,p_k)$. Following Sedransk et al, given t we specify a prior $[\underline{p}|t]$ the form

$$c(\beta_1,...,\beta_k; t) \sum_{j=1}^{k} p_j^{\beta_j - 1} \qquad (12)$$

over $S^k = \{(p_1,...,p_k) : p_1 \le p_2 \le ... \le p_t \ge p_{t+1} \ge ... p_k, 0 \le p_j \le 1, \sum_{j=1}^{k} p_j = 1\}$ where $c(\beta_1,...,\beta_k;t)$ is the normalizing constant. Note that (12) arises from a Dirichlet prior over $\underline{p}$ were the p_i unconstrained.

Sedransk et al assume t is known and only compute desired posterior expections using Monte Carlo itegration, employing importance sampling to avoid calculation of c. To implement the Gibbs sampler requires the full conditional distribution, for p_i, $i = 1,...,k-1$ (p_k is a function of these p_i), $[p_i | \underline{Y}, p_j, j = 1,...,k-1, j \neq i, t]$. 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 the Gibbs sampler also avoids calculation of c. Moreover, empirical work (as in 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.

Suppose t is unknown whence we take it to be random with discrete prior $Pr(t = j) = \tau_j$, $j = 1,2,...,k$. We note that $[t | \underline{Y}, \underline{p}]$ is a degenerate distribution. Therefore the Gibbs sampler can not be directly employed since 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). There 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

$$Pr(t=j|\underline{Y}) = \frac{c(\beta_1,...,\beta_k;j)\tau_j / c(\beta_1+Y_1, ...,\beta_k+Y_k; j)}{\sum_{t=1}^{k} c(\beta_1,...,\beta_k;t)\tau_t / c(\beta_1+Y_1,..., B_k+Y_k;t)} \qquad (13)$$

Evaluation of the $2k$ c's in (13) can be done straightforwardly using Monte Carlo integration with importance sampling as in Sedransk et al. Thus we can estimate the marginal posteriors for the p_i by using the relationship

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

For each given t we can use the Gibbs sampler in the customary manner to obtain samples approximately from $[p_i|\underline{Y},t]$. We can then resample from these samples according to $[t|\underline{Y}]$ to obtain observations approximately from $[p_i|\underline{Y}]$. Full details of all the required sampling in the context of an illustrative example are given in section 5.1.

Note that in a different context the sequence p_i might, for instance, be assumed to have a bimodal form, e.g. 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 and 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.

4.4 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 appears in Gelfand et al (1989); application to ordered slopes in a change–point regression model is given in Carlin et al (1989). Extensions to other problems will be obvious from the following development.

Consider, then, a model of the form

$$Y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}, \; i = 1,2,...,I, \; j = 1,2,...,J, \tag{14}$$

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, etc. Other applications occur in consumer preference studies (Green 1974, Green & Srinivasan, 1978): here, Y_{ij} might to 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 context level.

The discussion of 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, and that for t is obtained using the technique described in the previous section. Full detail on the required sampling in the context of an illustrative example is given in section 5.2.

4.5 Ordered and Data Constrained Parameters

As an illustration of a situation where the constraint set $S_{\underline{Y}}^{k}$ discussed in Section 3.1 depends on $\underline{Y}$, consider the following model, which has applications to reliability development studies and survival analysis. We suppose that Y_{ij}, $i = 1,2,...,k$, $j = 1,2,...,n_i$ 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 amongst the parameters there is recent decision–theoretic discussion of simultaneous point estimation of the location parameters in such models, assuming known scale parameters and vice versa. See for example Ebrahimi and Hosmane (1988), Das Gupta et al (1988).

Here, we shall complete a Bayesian model specification by assuming, for purposes of illustration, that $0 < \theta_1 \leq \theta_2 \leq ... \leq \theta_k$ 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 $[b^a/\Gamma(a)]\ [\exp\{-b|\sigma_i\}/\sigma_i^{a+1}]$, 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 which we note is extremely unpleasant using standard Monte Carlo integration

due to the awkward nature of $S_{\underline{Y}}^{k}$, defined by $Y_{ij} \geq \theta_i > 0$ and $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$.

However, approached via the Gibbs sampler, 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

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

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

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

restricted to the interval $\theta_{i-1} \leq \theta_i \leq (\min_j Y_{ij}) \wedge \theta_{i+1}$, where $\theta_0 \equiv 0$, $\theta_{k+1} \equiv \infty$. These full conditional distribution are therefore easily sampled and the Bayesian analysis straightforwardly implemented.

4.6 Straight–line Regression With Censoring

As a first illustration of a truncated data problem, consider the special case of the structure introduced in Section 3.2 where $[\underline{Y} | \underline{\theta}]$ corresponds to conditionally independent straight–line observations generated from $Y_{ij} = \alpha + \beta X_i + \epsilon_{ij} \sim N(0,\sigma^2)$, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,n_i$ and $\underline{Z}$ is defined by

$$Z_{ij} = \begin{cases} Y_{ij} & Y_{ij} \leq d_i \\ \text{if} & \\ d_i & Y_{ij} > d_i \end{cases}.$$

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

In contrast, the implementation of the Gibbs sampler using the approach set out in

Section 3.2 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 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.

4.7 Truncated Bivariate Normal Data

Consider a bivariate normal process where X_i is observed and then Y_i is observed only if $Y_i \leq X_i$, $i = 1,...,n$. Such data arises, for example, in paired survival time studies (perhaps using logarithms of survival time) where observation is terminated when the first patient dies. The more general version where observation is terminated when either patient dies can be handled in a similar fashion to the development given below.

More precisely, assume iid pairs (X_i, Y_i) such that for $i = 1,2,...,n$

$$[\begin{pmatrix} X_i \\ Y_i \end{pmatrix}] = N(\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}) \qquad (15)$$

We observe the pairs (X_i, Z_i) with $Z_i = Y_i$ if $Y_i \leq X_i$; otherwise we observe $(X_i, *)$ where $Z_i = *$ indicates that $Y_i > X_i$. Inference for θ_x, θ_y is of interest as well as, perhaps, for Σ, the covariance matrix in (14). A convenient prior form for $\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}$ is $N(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, V)$ and for Σ an inverse Wishart, so that $[\Sigma^{-1}] = W((\rho R)^{-1}, \rho)$ with μ_x, μ_y, V, ρ and R assumed known. Interest focuses on the marginal posterior distributins of θ_x, θ_y and Σ. Following the development of Section 3.2, these may be routinely obtained using the Gibbs sampler. As in section 3.2, treating Y as an unobservable, we will need the full conditional distributions of θ_x, θ_y, Σ and Y_i.

Letting $\underset{\sim}{T}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix}$, $\underset{\sim}{\theta} = \begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}$, $\underset{\sim}{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ and using standard distribution theory it is straightforward to verify that:

$$[\underset{\sim}{\theta} | \underset{\sim}{T}_i, i=1,\ldots,n, \underset{\sim}{Z}, \Sigma] = N(V(n^{-1}\Sigma+V)^{-1}\overline{\underset{\sim}{T}} + n^{-1}\Sigma(n^{-1}\Sigma+V)\underset{\sim}{\mu}, (n\Sigma^{-1}+V^{-1})^{-1}) \qquad (16)$$

where $\overline{\underset{\sim}{T}} = n^{-1} \sum_{i=1}^{n} \underset{\sim}{T}_i$ and from (16) the full conditional distributions for θ_x and θ_y are routinely obtained,

$$[\Sigma^{-1} | T_i, i=1,\ldots,n, \underset{\sim}{Z}, \underset{\sim}{\theta}] = W((\Sigma(\underset{\sim}{T}_i - \underset{\sim}{\theta})(\underset{\sim}{T}_i - \underset{\sim}{\theta})^T + \rho R)^{-1}, n+\rho) \qquad (17)$$

from which sample generation is easily achieved as detailed in Gelfand et al (1989). Routine analysis also shows that the Y_i are conditionally independent with

$$[Y_i | X_i, Z_i, \underset{\sim}{\theta}, \Sigma] = N(\theta_y + \sigma_{xy}(X_i - \theta_x)/\sigma_x^2, \sigma_y^2 - \sigma_{xy}/\sigma_x^2)$$

if $Z_i = *$, and Y_i degenerate at Z_i otherwise.

There are a number of obvious extensions and generalizations of this kind of structure. Examples include: further modeling of θ_x and θ_y as parametric functions of exploratory variables; trivariate or multivariate data and bivariate exponential family models with conjugate priors. The reader can doubtless think of others. The point we wish to stress is that systematic use of the ideas of Section 3.2 will lead in each case to a relatively straightforward form of Gibbs sampler, no matter how seemingly complex the initial model structure is.

4.8 Bivariate Grouped Data

Suppose that data from an underlying continuous bivariate distribution has been grouped into an IxJ 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 (15) given above in Section 4.7. 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 $[b_{j-1}, b_j]$, $j = 1,\ldots,J$ where $a_0 = b_0 = 0$

(technically these should be $-\infty$), $a_I = b_J = \infty$. The data consist of counts n_{ij} $i = 1,...,I$, $j = 1,...,J$ 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 $\text{Mult}(n; p_{11}, p_{12},...,P_{IJ})$ 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 (15). For illustration, we adopt the same prior structure for $\underset{\sim}{\theta}$ and Σ as given in the previous section, i.e. $[\underset{\sim}{\theta}] = N(\underset{\sim}{\mu}, V)$ and $[\Sigma^{-1}] = W((\rho R)^{-1}, \rho)$, and we seek the marginal posterior distributions $[\theta_x|\underset{\sim}{n}]$, $[\theta_y|\underset{\sim}{n}]$ and $[\Sigma|\underset{\sim}{n}]$ where $\underset{\sim}{n} = (n_{11}, n_{12},...,n_{IJ})$.

The Gibbs sampler is most easily implemented if we include the $\underset{\sim}{T}_s = \begin{pmatrix} X_s \\ Y_s \end{pmatrix}$, $s = 1,...,n$ in the model as unobservables. We then require the full conditional distributions for θ_x, θ_y, Σ and $\underset{\sim}{T} = (\underset{\sim}{T}_1,...,\underset{\sim}{T}_n)$. But $[\underset{\sim}{\theta}|\underset{\sim}{T},\underset{\sim}{n},\Sigma]$ is exactly (16) from which the full conditional distributions for θ_x and θ_y can be readily obtained. Similarly, $[\Sigma|\underset{\sim}{T},\underset{\sim}{n},\underset{\sim}{\theta}]$ is exactly (17). Finally, we need to generate $\underset{\sim}{T}_s$, $s = 1,...,n$ given $\underset{\sim}{n},\underset{\sim}{\theta}$ and Σ. But this merely requires, for each pair i,j, that we generate n_{ij} independent observations from (15) 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, i.e., product multinomial sampling (see Bishop, Fienberg and Holland, 1975). To be concrete, suppose this third variable is age and the bivariate groups do actually correspond to height and weight. That is, grouped height and weight data is supplied (using the same groups) for a sample of say 5 year old children, a sample of 6 year old children, etc.. Under (15) 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 not be surprised to find that the Gibbs sampler is very straightforward, despite the seeming awkwardness of the model and parameter constraints.

5. Illustrative Analysis

In this section, we analyse two artificial data sets, derived from models based on those discussed in Sections 4.3 and 4.4. It will be clear from Section 4 that 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.

5.1 Multinomial With Ordered Parameters

As an example of the problem discussed in Section 4.3, Table 1 shows a $k = 8$ cell multinomial model and the results of 40 random draws from this model

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

Table 1: Multinomial Population and Generated Data

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

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

Table 2: Marginal Posterior Distribution of t

As remarked in Section 4.3, to obtain the marginal posteriors, $[p_i|\underline{Y}]$, we implement the Gibbs sampler in a slightly different way. We use general k and $\underline{\alpha} = (\alpha_1,...,\alpha_k)$ in the ensuing details. Since

$$[p_i|\underset{\sim}{Y}] = \sum_{l=1}^{k} [p_i|\underset{\sim}{Y},t]\,[t|\underset{\sim}{Y}] \tag{18}$$

and since $[t|\underset{\sim}{Y}]$ has already been obtained we propose to sample from $[p_i|\underset{\sim}{Y}]$ by randomly selecting t according to $[t|\underset{\sim}{Y}]$ and then sampling p_i from $[p_i|\underset{\sim}{Y},t]$. The densities $[p_i|\underset{\sim}{Y},t]$ can be obtained using the Gibbs sampler in the customary fashion, as indicated in Section 4.3. More precisely, we only require the full conditional distributions for p_i, $i=1,...,k-1$ since $p_k = 1 = \sum_{i=1}^{k-1} p_i$. But $[p_i|\underset{\sim}{Y},t,p_j,\ j=1,2,...,k-1,\ j\neq i] = \text{Beta}(\alpha_i + Y_i,\ \alpha_k + Y_k)$ scaled to the interval $[0, a_i]$ where $a_i = 1 - \sum_{j=1}^{k-1} p_j, j\neq i$, and then restricted to an interval determined by the other p_j's and t, i.e., with $p_0 \equiv 0$

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

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

$$\text{it } i = t,\ \max(p_{t-1},\ p_{t+1},\ a_t - p_{k-1}) \leq p_t \leq a_t.$$

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

Suppose we run the Gibbs sampler in this manner for each t, $t = 1,...,k$. Then in theory we could obtain a kernel density estimate for each $[p_i|\underset{\sim}{Y},t]$, $t = 1,...,k$ and thus via (18) a density estimate which 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, $j = 1,...,m$. This resampling procedure results in an observation approximately distributed as $[p_i|\underset{\sim}{Y}]$. Repeating this process a large number of times (1000 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 2 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.

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

Table 3: Marginal Posterior Modes of the $[p_i \mid \underline{Y}]$

5.2 Two Way Layout With Ordered Parameters

Consider the problem discussed in Section 4.4 Table 4 presents a set of data, $\underline{Y}$, generated from (14) of Section 4.4 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 while for each row they increase to the middle column and then decrease. The data is rather noisy often at odds with these expectations. Ordinary least squares analysis ignoring known order restrictions is terribly misleading; $\hat{\alpha}_1 = 1.064$, $\hat{\alpha}_2 = -1.163$, $\hat{\alpha}_3 = .536$, $\hat{\alpha}_4 = -5.203$, $\hat{\beta}_1 = -1.737$, $\hat{\beta}_2 = .758$, $\hat{\beta}_3 = .283$, $\hat{\beta}_4 = -3.344$, $\hat{\beta}_5 = -1.917$, $\hat{\sigma}^2 = 3.590$. The analyst obtains estimates for the α_i and β_j which fail to meet the restrictions and are often far from the true values. Some sort of constrained least squares solution (an isotonic regression) would be a better frequentist approach.

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

Table 4: Generated Two Way Layout Data

Bayesian analysis using the Gibbs sampler is easily implemented in this case yielding marginal posterior distributions for the α_i, the 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 respectively. That is, ignoring

restrictions let α_i i.i.d. $N(0,\sigma_\alpha^2)$, β_j i.i.d. $N(0,\sigma_\beta^2)$. (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 σ^2-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, b = 1. The full conditional distributions using general σ_α^2, σ_β^2, a, b and incorporating the order restrictions are:

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

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

$$[\beta_j | \underline{Y}, \beta_s, s\neq j, \alpha_i, \sigma^2] = N\left(\frac{4\sigma_\beta^2(\overline{Y}_{\cdot j}-\alpha_\cdot)}{4\sigma_\beta^2+\sigma^2}, \frac{\sigma_\beta^2\sigma^2}{4\sigma_\beta^2+\sigma^2}\right), \quad j=1,2,\ldots,5$$

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

$$[\sigma^2 | \underline{Y}, \alpha_i, i=1,\ldots,4, \beta_j, j=1,\ldots,5] = IG(a+10, b+\textstyle\sum_i \sum_j (Y_{ij}-\alpha_i-\beta_j)^2/2).$$

As output from running the Gibbs sampler Figures 2, 3, and 4 show the marginal posterior distributions for the α_i, the β_j and σ^2 respectively. Figures 2 and 3 show that these marginal posteriors respond to the order restrictions. In particular marginal posterior modes are $\alpha_1^*= 1.480$, $\alpha_2^*= .197$, $\alpha_3^*= -.507$, $\alpha_4^*= -3.684$, $\beta_1^*= -1.039$, $\beta_2^*= .635$, $\beta_3^*= 1.261$, $\beta_4^*= -1.149$, $\beta_5^*= -1.790$, $\sigma^{2*} = 3.975$. These are in accord with the restrictions and generally much closer to the true values than the ordinary least squares estimates. While a constrained least squares solution would no doubt produce comparably good point estimates, because the Gibbs sampler enables marginal posterior distributions for the α_i,

β_j for e.g. $\alpha_i+\beta_j$ for e.g. $\alpha_r-\alpha_s$, $\beta_r-\beta_s$, etc. we can in addition easily obtain Bayesian interval estimates for any functions of the model parameters which may be of interest.

In the above 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 the data contained some outliers, we might robustify the Bayesian analysis by assuming that the distribution of the errors in (14) is $[\epsilon_{ij}|\lambda_{ij}] = N(0, \lambda_{ij}^{-1}\sigma^2)$ where $\nu\lambda_{ij}$~Gamma$(\nu/2, 1/2)$ so that marginally the ϵ_{ij}~$t_\nu(0,\sigma^2)$. To implement the Gibbs sampler we would include the λ_{ij} as unobserved variables. All full conditional distributions are straightforward to obtain. We omit details.

6. Summary

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

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*. J. Wiley & Sons, London.

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

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

Broffitt, J.D. (1984). A Bayes estimator for ordered parameters and isotonic Bayesian graduation. *Scand. Actuarial J.*, 1984, 231–247.

Broffit, J.D. (1987). Restricted Bayes estimators for binomial parameters. *Probability and Bayesian Statistics* ed. R. Viertl, New York, Plenum Press, 61–72.

Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1989). Heirarchical Bayesian analysis of change point problems. *University of Connecticut Tech. Rpt. 89–21.*

DasGupta, 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. *Sankhya A*, 50(2), 269–281.

Devroye, L. (1986). Non–uniform random variate generation. Springer–Verlag, New York.

Ebrahimi, N. and Hosmane, B. (1988). On shrinkage estimation of the exponential location parameter. *Commun. in Statist. A*, 16, 2623–2637.

Fard, N.S. and Dietrich, D.L. (1987). A Bayesian note on reliability growth during a development testing program. *I.E.E.E. Trans. Rel.*, R–36, 568–572.

Gelfand, A.E. and Smith, A.F.M. (1989). Gibbs sampling for marginal posterior expectations. *University of Connecticut Tech Rpt.89–28.*

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal density. *J. Amer. Statist. Assoc.* (to appear).

Gelfand, A.E., Hills, S.E., Racine–Poon, A. and Smith, A.F.M. (1989). Illustration of Bayesian inference in normal data models using Gibbs sampling. *University of Nottingham Tech Rpt.*

Geman, D., and Geman, S. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I.E.E.E. 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.

Kuo, L. (1988). Linear Bayes estimators of the potency curve in bioassay. *Biometrika*, 25, 91–96.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with

discussion). *J. Royal Statist. Soc.*, 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.

Racine–Poon, A., Smith, A.F.M., and Gelfand, A.E. (1990). Gibbs sampling implementation of the Bayesian hierarchical model. *University of Nottingham Tech Rpt.*

Ripley, B. (1986). Stochastic Simulation. J. Wiley & Sons, New York.

Robertson, J., Wright, F.T. and Dykstra, R.L. (1988). Order restricted Statistical Inference. J. Wiley & Sons, Chichester.

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. *J.R. Statist. Soc.*, B, 47, 519–527.

Sharples, L.D. (1988). Aspects of robustness and approximation in hierarchical models. Unpublished Ph.D. thesis, University of Nottingham.

Silverman, B. (1986). Density estimation for statistics and data analysis. Chapman and Hall, London.

Smith, A.F.M. (1973). Bayes estimates in one–way and two–way models. *Biometrika*, 60, 319–329.

Smith, A.F.M. (1977). A Bayesian note on reliability growth during a development testing program. *I.E.E.E. Trans. Rel.* R–26, 346–347.

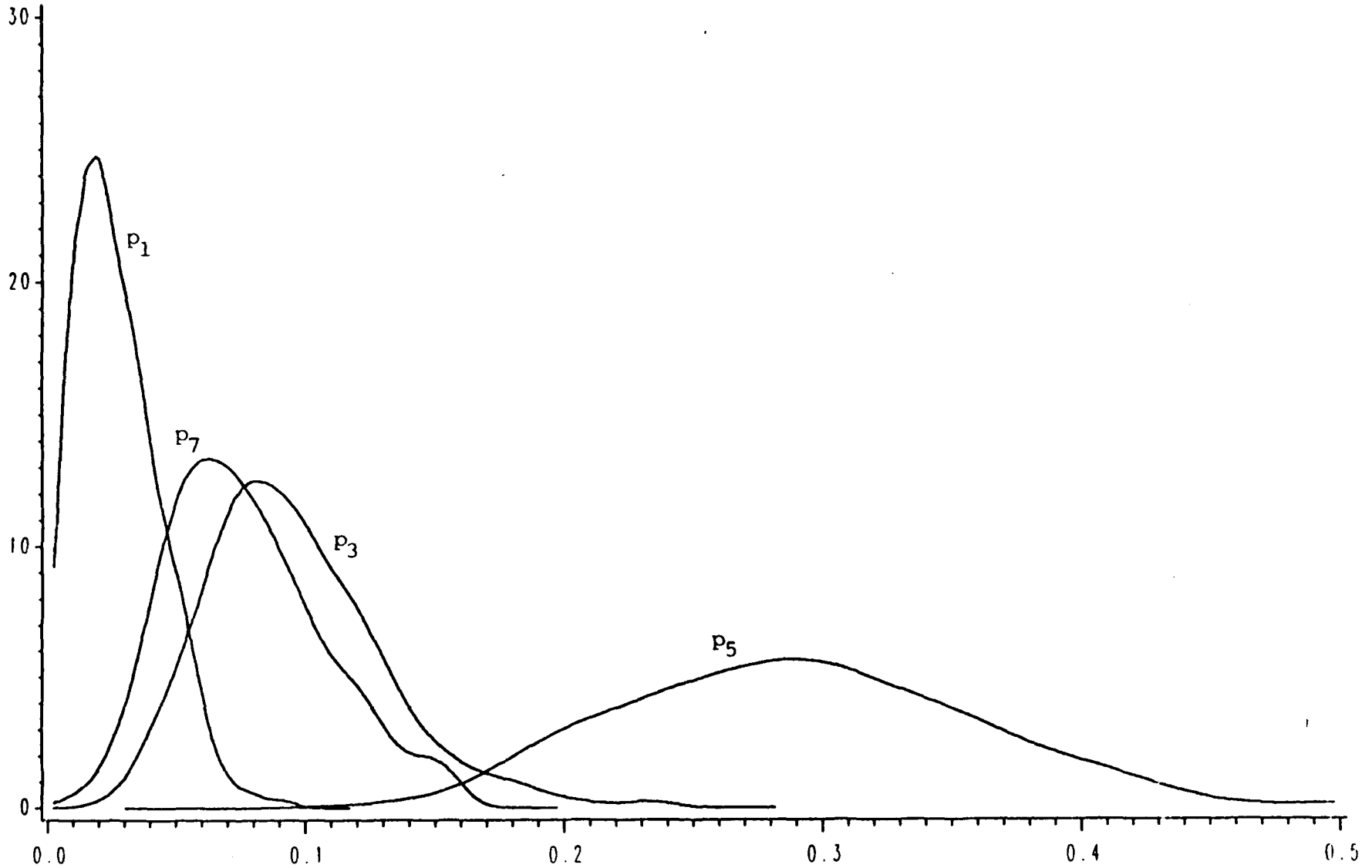


Figure 1: Marginal Posterior Distributions for Selected p_i's, Sec. 5.1

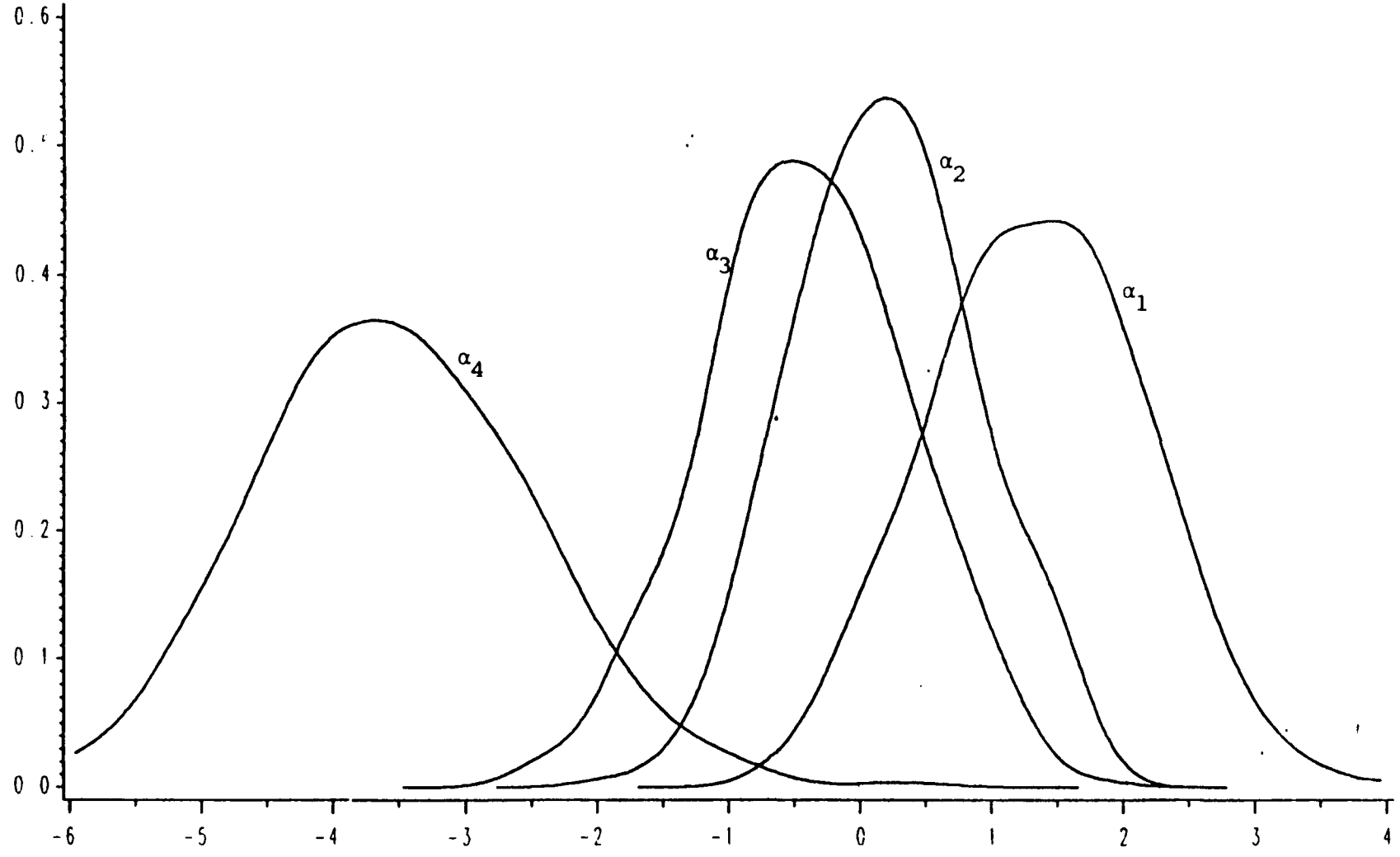


Figure 2: Marginal Posterior Distributions for α_i's, Sec. 5.2

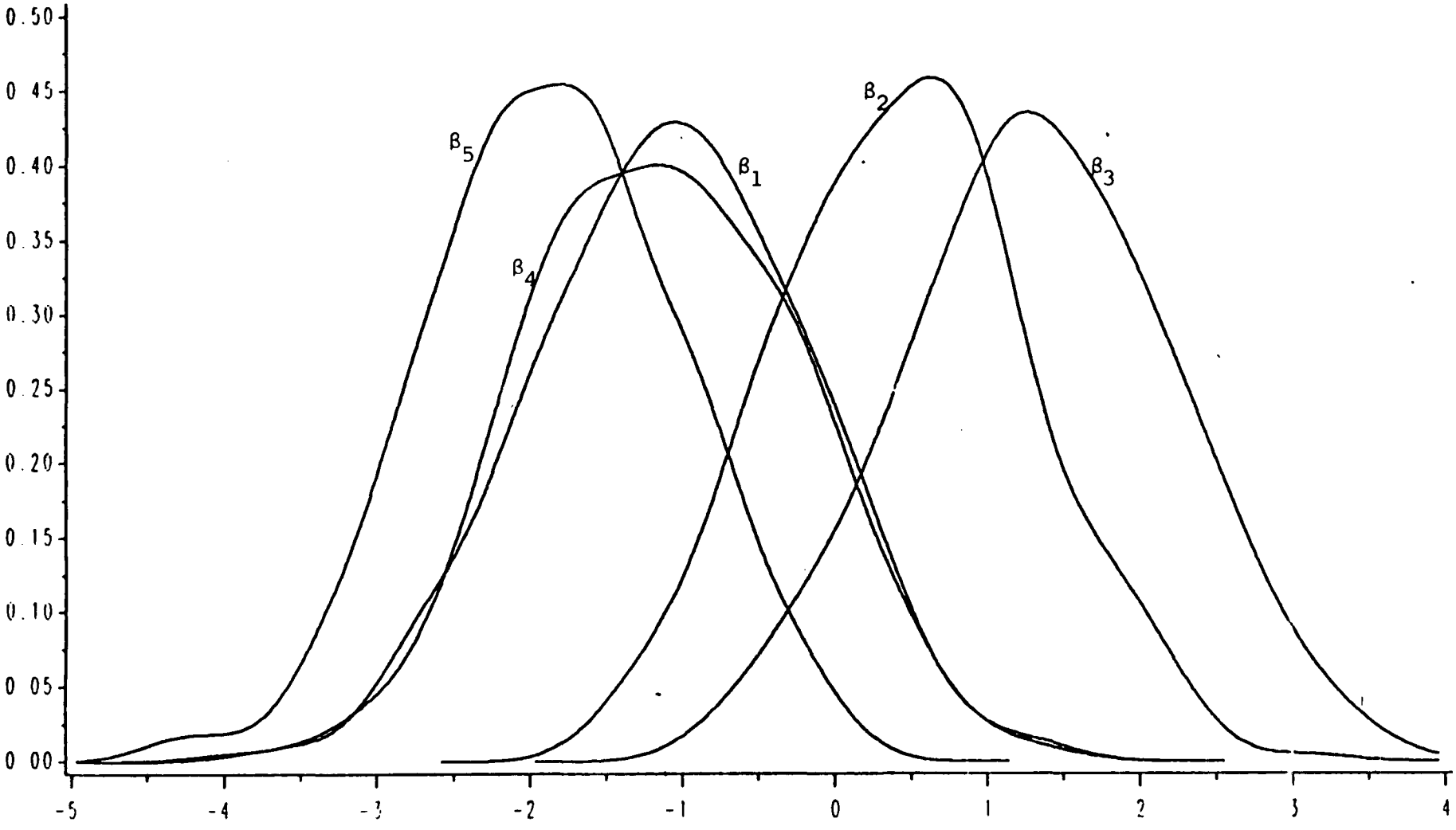


Figure 3: Marginal Posterior Distributions for β_i's, Sec. 5.2

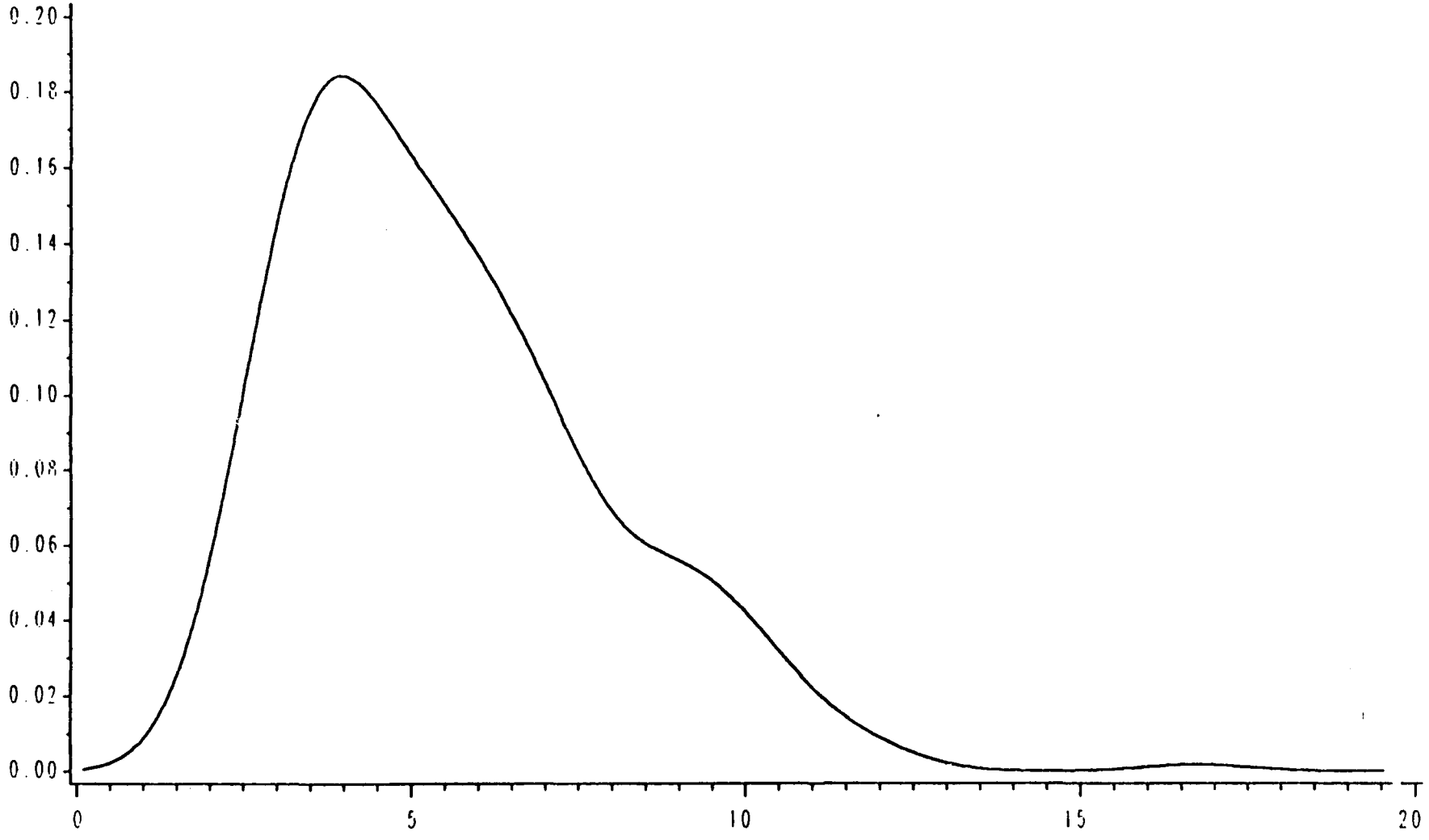


Figure 4: Marginal Posterior Distribution for σ^2, Sec. 5.2

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 439	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Bayesian Analysis Of Constrained Parameter And Truncated Data Problems		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A.E. Gelfand, A.F.M. Smith and T-M. Lee		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE January 4, 1991 13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayesian inference; Gibbs sampler; Constrained parameters; Truncated data.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Bayesian analysis of constrained parameter and truncated data problems is complicated by the seeming need for, typically multidimensional, numerical integrations over awkwardly defined regions. This paper illustrates how the Gibbs sampler approach to Bayesian calculation (Gelfand and Smith, 1990) avoids these difficulties and leads to straightforwardly implemented procedures, even for apparently very complicated model forms.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)