Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities

John Geweke
Department of Economics
University of Minnesota
Minneapolis, MN 55455

First draft: April, 1991

Phone: (612)625-7563 Fax: (612)624-0209
E-mail: geweke@atlas.socsci.umn.edu

Abstract

The construction and implementation of a Gibbs sampler for efficient simulation from the truncated multivariate normal and Student-t distributions is described. It is shown how the accuracy and convergence of integrals based on the Gibbs sample may be constructed, and how an estimate of the probability of the constraint set under the unrestricted distribution may be produced.

Keywords: Bayesian inference; Gibbs sampler; Monte Carlo; multiple integration; truncated normal

This paper was prepared for a presentation at the meeting Computing Science and Statistics: the Twenty-Third Symposium on the Interface, Seattle, April 22-24, 1991. Research assistance from Zhenyu Wang and financial support from National Science Foundation Grant SES-8908365 are gratefully acknowledged. The software for the examples may be requested by electronic mail, and will be returned by that medium.

1. Introduction

The generation of random samples from a truncated multivariate normal distribution, that is, a multivariate normal distribution subject to multiple linear inequality restrictions, is a recurring problem in the evaluation of integrals by Monte Carlo methods in econometrics and statistics. Sampling from a truncated multivariate Student-t distribution is a closely related problem. The problem is central to Bayesian inference, where a leading example is the normal linear regression model subject to linear inequality restrictions on the coefficients (Geweke, 1986). But it also arises in classical inference, when integrals enter the likelihood function; McFadden (1989) has proposed the use of Monte Carlo integration in one such instance.

Recently several promising solutions of this problem have been investigated. A survey of methods along with several contributions is provided in Hajivassiliou and McFadden (1990). One of these methods, the Gibbs sampler, is especially well suited to the problem. It uses a simple algorithm that generates samples with great computational efficiency, but at the cost of introducing three complications. First, the drawings are not independent, which complicates the evaluation of the accuracy of the approximation using standard methods like those proposed in Geweke (1989). Second, the distribution from which the drawings are made converges to the truncated multivariate normal distribution, but is not identical with it at any stage. Third, the method does not produce an estimate of the probability of the constraint set under the unrestricted multivariate normal distribution. Such an estimate is often desirable and is provided by other methods, which sets the Gibbs sampler at a relative disadvantage.

In this paper we contribute to the resolution of these three problems. The solution of the first two problems is an application of methods developed elsewhere (Geweke, 1991), while the solution of the third is wholly developed here. The ability to generate variates from a truncated univariate normal distribution is a central building block in the solution of the more general problem. Section 2 describes an algorithm for the generation of variates from a truncated univariate normal distribution that is substantially more efficient and flexible than the method that has been favored in the past. Drawing from the truncated multivariate normal distribution with the Gibbs sampler is fully developed in Section 3, including the evaluation of the accuracy of the numerical approximation and construction of diagnostics for convergence. These methods are extended to the multivariate Student-t distribution in Section 4. The estimation of the probability of the constraint set under the unrestricted distribution is taken up in Section 5.

Throughout the paper some standard but not universal notation is employed. The univariate normal probability density function is $\phi(\cdot)$, the corresponding cumulative distribution function is $\Phi(\cdot)$, and the inverse cumulative distribution function is $\Phi^{-1}(\cdot)$. The uniform distribution on the interval [a, b] is denoted U[a, b]. The univariate truncated normal distribution TN(a, b) is the univariate normal restricted to (a, b): its density is $[\Phi^{-1}(b) - \Phi^{-1}(a)]^{-1}\phi(\cdot)$ on (a, b) and 0 elsewhere; $a = -\infty$ and $b = +\infty$ are permitted special cases.

2. The mixed rejection algorithm for truncated univariate normal sampling

All of the methods described in this paper assume the ability to draw i.i.d. samples from a truncated univariate normal distribution. It is well recognized that rejection sampling from a univariate normal distribution is impractical. Inverse c.d.f. sampling (Devroye, 1986) is a feasible alternative. If $x \sim TN(a, b)$, then $x = \Phi^{-1}(u)$, $u \sim U[\Phi(a), \Phi(b)]$. This method requires the evaluation of one integral for each draw, and if the values of a and b change with the draws, then three evaluations are required. The computation of $\Phi^{-1}(w)$ requires more time as $w \to 0$ or $w \to 1$, and the double precision implementation in the IMSL/STAT library is unable to compute $w = \Phi^{-1}(p)$ if $|w| > 8$. Here, we shall suggest a different algorithm, whose execution times are substantially smaller than inverse c.d.f. sampling, and can draw $x \sim TN(a, b)$ for any $a < b$ so long as $|a| \leq 35$ and $|b| \leq 35$, when programmed in double precision (64-bit) floating point arithmetic.

The algorithm produces i.i.d. samples from TN(a, b), including the cases $a = -\infty$ and $b = +\infty$. It employs four different kinds of rejection sampling, depending on the values of a and b. In *normal rejection sampling*, x is drawn from N(0, 1) and accepted if $x \in [a, b]$. In *half-normal rejection sampling*, x is drawn from N(0, 1) and |x| is accepted if $x \in [a, b]$ (where $a \geq 0$). In *uniform rejection sampling*, x is drawn from U[a, b], u is drawn independently from U(0,1), and x is accepted if $u \leq \phi(x)/\phi(x^*)$, $x^* = \text{argmax}_{[a,b]}[\phi(x)]$.

Exponential rejection sampling is key to the algorithm, and requires description in some detail. The motivating example is $TN(a, \infty)$, where $a > 0$, and possibly $\Phi(a)$ is close to 1. As $a \to \infty$, the $TN(a, \infty)$ distribution comes to resemble the exponential distribution as detailed in Geweke (1986, Appendix A). Suppose z is drawn from an exponential distribution on $[a, \infty)$ with kernel $\exp(-\lambda z)$ for $z \geq a$. Consider fixing λ so as to minimize the probability of rejection. The acceptance probability must be proportional to $\exp(-\frac{1}{2}z^2)/\exp(-\lambda z)$, for $z \in [a, \infty)$. Computing the constants of proportionality, we find acceptance probabilities

$\exp[-\frac{1}{2}(z^2 + a^2)] \exp(-\lambda z)$ if $\lambda \le a$, $\exp[-\frac{1}{2}(z - \lambda)^2)]$ if $\lambda \ge a$.

The first expression is maximized at $\lambda = a$ for all z. Integrating the second expression with respect to the exponential density $\lambda\exp[-\lambda(z-a)]$ on $[a, \infty)$, we find that the acceptance probability is $\lambda \exp(\lambda a - \frac{1}{2}\lambda^2)(2\pi)^{1/2}[1 - \Phi(a)]$. This is maximized when $\lambda = \frac{1}{2}[a + (a^2 + 4)^{1/2}]$. As $a \to \infty$, $\lambda/a \to 1$, and the acceptance probability converges to unity. Experimentation within the context of the algorithm presently described has shown that the increase in computing time from using the suboptimal but simpler choice $\lambda = a$, is less than the time required to compute $\frac{1}{2}[a + (a^2 + 4)^{1/2}]$. Hence we use $\lambda = a$ in this algorithm.

The algorithm employs four constants (t_i, i = 1, .. , 4) whose values have been set through experimentation with computation time. The selected value is indicated when the constant is introduced. The sampling procedure depends on the relative configuration of a and b, as follows. Except in case (1), a and b are finite.

(1) On (a, ∞): normal rejection sampling if $a \le t_4$ (= .45); exponential rejection sampling if $a > t_4$.

(2) On (a, b) if $0 \in [a, b]$:
   (a) If $\phi(a) \le t_1$ (= .150) or $\phi(b) \le t_1$: normal rejection sampling;
   (b) If $\phi(a) > t_1$ or $\phi(b) \ge t_1$: uniform rejection sampling.

(3) On (a, b) if $a > 0$:
   (a) If $\phi(a)/\phi(b) \le t_2$ (= 2.18): uniform rejection sampling;
   (b) If $\phi(a)/\phi(b) > t_1$ and $a < t_3$ (= .725): half-normal rejection sampling;
   (c) If $\phi(a)/\phi(b) > t_1$ and $a \ge t_3$: exponential rejection sampling.

The omitted cases $(-\infty, b)$, and (a, b) with $b < 0$, are symmetric to the cases (1) and (3), respectively, and are treated in the same way. Software for the mixed rejection algorithm was tested by comparing the distributions of sampled variates produced, with those produced by inverse c.d.f. sampling. Each was programmed in double precision Fortran-77 using the IMSL/STAT library, on a Sun Sparcstation 4/40 (IPC). Computation times for 10,000 sampled variates are shown in Table 1. Times for the inverse c.d.f. algorithm range from 2.24 to 4.51 seconds, those for the mixed rejection algorithm from 0.67 to 1.28 seconds. On a case-by-case basis the mixed rejection algorithm is from 2.47 to 6.24 times faster than the inverse c.d.f. algorithm.

3. The Gibbs algorithm for truncated multivariate normal sampling

The central problem addressed in this paper is the construction of samples from an n-variate normal distribution subject to linear inequality restrictions,

$$x \sim N(\mu, \Sigma), \qquad a \leq Dx \leq b. \tag{3.1}$$

The matrix D is $n \times n$ of rank n, individual elements of a may be $-\infty$, and individual elements of b may be $+\infty$. This accommodates fewer than n linearly independent restrictions. It does not allow more than n linearly independent restrictions, and the method set forth here cannot be extended to these cases, at least in a tidy way. In the applications described in the introduction the truncated multivariate normal distribution arises in the form (3.1). The problem is equivalent to the construction of samples from the n-variate normal distribution subject to linear restrictions,

$$z \sim N(0, T), \qquad \alpha \leq z \leq \beta, \tag{3.2}$$

where

$$T = D\Sigma D', \quad \alpha = a - D\mu, \quad \beta = b - D\mu,$$

and we then take $x = \mu + D^{-1}z$.

Several approaches to the solution are possible; see Hajivassiliou and McFadden (1990, Appendix B) for a brief survey of these methods, and Hajivassiliou, McFadden, and Ruud (1990) for an application of importance sampling to the special case of orthant restrictions. Naive rejection sampling from $N(\mu, \Sigma)$ can be employed directly in (3.1), but is impractical in general since the ratio of rejected to accepted variates is astronomical for many commonly arising problems. More sophisticated procedures must cope with the fact that the marginal distributions of the elements of z, and of x, are not univariate truncated normal. The method set forth here exploits the fact that the distribution of each element of z, *conditional* on all of the other elements of z, is truncated normal. This method has also been described by Hajivassiliou and McFadden (1990), but as outlined in the introduction we pursue several extensions here.

The algorithm employed is the Gibbs sampler, whose systematic application to problems of this form dates from Geman and Geman (1984); see also Gelfand and Smith (1990). The general problem is to sample from a multivariate density $f(x)$ for an n-dimensional random vector x, when no practical algorithm is available for doing so directly. But suppose that the conditional distributions,

$x_i \mid \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\} \sim f_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \qquad (i = 1, \ldots, n)$

are known, and are of a form that synthetic i.i.d. random variables can be generated readily and efficiently from each of the $f_i(\cdot)$. Let $x^{0\prime} = (x_1^0, \ldots, x_n^0)$ be an arbitrary point in the support of $f(x)$. Generate successive synthetic random variables,

$$x_i^1 \mid \{x_1^1, \ldots, x_{i-1}^1, x_{i+1}^0, \ldots, x_n^0\} \sim f_i(x_1^1, \ldots, x_{i-1}^1, x_{i+1}^0, \ldots, x_n^0) \quad (i = 1, \ldots, n) \qquad (3.3)$$

These n *steps* constitute the first *pass* of the Gibbs sampler. The second and successive passes are performed similarly. At the i'th step of the j'th pass,

$$x_i^j \mid \{x_1^j, \ldots, x_{i-1}^j, x_{i+1}^{j-1}, \ldots, x_n^{j-1}\} \sim f_i(x_1^j, \ldots, x_{i-1}^j, x_{i+1}^{j-1}, \ldots, x_n^{j-1}),$$

and the composition of the vector becomes

$$x^{(j,i)\prime} = (x_1^j, \ldots, x_i^j, x_{i+1}^{j-1}, \ldots, x_n^{j-1})'$$

at the end of this step. At the end of the j'th pass the composition of the vector is

$$x^{(j)\prime} = (x_1^j, \ldots, x_n^j)'.$$

Gelfand and Smith (1990) have outlined weak conditions under which $x^{(j)}$ converges in distribution and has limiting distribution give by the density $f(x)$, and the rate of convergence is geometric in the L_1 norm. These conditions pertain to the truncated multivariate normal density in (3.2). The conditional densities $f_i(\cdot)$ for this problem are truncated univariate normal, and the algorithm described in the previous section may be used to generate the required successive synthetic random variables. Suppose that in the non-truncated distribution $N(0, T)$,

$$E[z_i \mid z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n] = \sum_{j \neq i} c_{ij} z_j.$$

Then in the truncated normal distribution of (3.2), the distribution of z_i conditional on $\{z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n\}$ has the construction,

$$z_i = \sum_{j \neq i} c_{ij} z_j + h_i \varepsilon_i, \qquad \varepsilon_i \sim TN[(\alpha_i - \sum_{j \neq i} c_{ij} z_j)/h_i, \ (\beta_i - \sum_{j \neq i} c_{ij} z_j)/h_i].$$

Denote the vectors of coefficients in the conditional means,

$$c^i = (c_{i1}, \ldots, c_{i,i-1}, c_{i,i+1}, \ldots, c_{in})' \qquad (i = 1, \ldots, n).$$

From the conventional theory for the conditional multivariate normal distribution (Rao, 1965, p. 441) and expressions for the inverse of a partitioned symmetric matrix (Rao, 1965, p. 29),

$$c^i = -(T^{ii})^{-1} T^{i,<i}, \quad h_i^2 = (T^{ii})^{-1},$$

where T^{ii} is the element in row i and column i of T^{-1}, and $T^{i,<i}$ is row i of T^{-1} with T^{ii} deleted. These computations need only be performed once, before sampling begins. An initial value $z^{(0)}$ may be selected by setting $z = 0$ and then successively applying (3.3) for $i = 1, \ldots, n$. At the end of each pass we compute $x^{(j)} = \mu + D^{-1}z^{(j)}$.

Samples from the truncated multivariate normal distribution are typically used to estimate the expected value of a function $g(\cdot)$ of the random vector x,

$$\bar{g} = \int_X g(x)f(x)dx.$$

An assessment of the reliability of this estimate must take into account the facts that in general $\{x^{(j)}\}$ is a serially correlated process, whose unconditional distribution converges to $f(\cdot)$ rather than being identical with $f(\cdot)$. These problems are taken up for the general case in Geweke (1991), where standard spectral analytic techniques are used to produce diagnostics for the convergence of the sampled distributions to $f(\cdot)$ and to provide a numerical standard error for the reliability of the estimated expected value. Using this approach, five statistics from the sample $\{x^{(j)}\}_{j=1}^{p}$ provide information about the expected value of the function in question.

(1)The simple arithmetic *mean* $\bar{g}_p = p^{-1} \sum_{j=1}^{p} g(x^{(j)})$ is the most efficient estimate of $\bar{g}$ from a Gibbs sample of size p passes, assuming that departures from convergence in the sample are negligible. The function $g(x)$ is computed at the end of each pass, following the transformation from z to x.

(2) The sampling variance of $\bar{g}_p$ is $S_g(0)/p$, where $S_g(\omega)$ denotes the spectral density of the Gibbs-sampled $g(x)$ process at frequency ω. The *numerical standard error (NSE)* of $\bar{g}_p$. is $[p^{-1}\hat{S}_g(0)]^{1/2}$, where $\hat{S}_g(\omega)$ is a consistent (in p) estimator of $S_g(\omega)$.

(3) The variance of $g(\cdot)$ is estimated in the same way as the mean of $g(\cdot)$. The ratio of this variance to $S_g(0)$ indicates the ratio of the number of i.i.d. draws that would have been required, were such an algorithm available, to the number of passes required with the Gibbs sampler, to produce an estimate of $\bar{g}$ of equivalent reliability. Following Geweke (1989), this ratio is called the *relative numerical efficiency (RNE)* of the Gibbs sampling procedure.

(4) A *convergence diagnostic (CD)* is computed based on subsamples of the sampled g(x); see Geweke (1991, Section 3.2) for details. Under a stationary distribution for $\{x^{(j)}\}_{j=1}^{p}$ this statistic has a standard normal distribution.

(5) The spectral density provides further details on the characteristics of the process $\{g(x^{(j)})\}_{j=1}^{p}$. If the spectral density is nearly flat, or is lower near $\omega = 0$ than at other frequencies, then the Gibbs sampling process is efficient relative to i.i.d. sampling. But if the spectral density is much higher near $\omega = 0$ than elsewhere, the process is inefficient. Thus, there is a correspondence between the shape of the spectral density and the RNE of the Gibbs sampler.

The Gibbs algorithm was programmed in double precision Fortran-77 using the IMSL/STAT library, on a Sun Sparcstation 4/40 (IPC). The routine was tested by comparing the distribution of truncated normal samples with those generated by a naive accept/reject procedure. To provide some indication of the efficiency of the procedure, we present two examples here.

The first example is a truncated bivariate normal, with parameters chosen so that convergence ought to be especially slow. Both variables have mean zero. The variance of x_1 is 10, while the variance of x_2 is 0.1, and the restrictions are of the form $a_1 \le x_1 + x_2 \le b_1$, $a_2 \le x_1 - x_2 \le b_2$. Consequently the transformed variables z_1 and z_2 have correlation .98. As elaborated in Geweke (1991), this implies that the process $z^{(j)}$, and hence $x^{(j)}$, will exhibit strong positive serial correlation: e.g., if $a_i = -\infty$ and $b_i = +\infty$, then each element of $z^{(j)}$ will follow a first order autoregressive process with parameter .96. Results are presented in Table 2, which shows the five statistics for five different configurations of truncation points (a_i, b_i), and for three choices of the number of passes, p = 400, 2000, or 10,000. In each case p preliminary passes were performed before the functions of interest $g_i(x)$ (i = 1, ... , 4) were computed and averaged over the next p passes. Computation times varied about 20% depending on the (a_i, b_i) configurations, averaging about .35 seconds for p = 400 and 7.1 seconds for p = 10,000.

The results, presented in Table 2, confirm that convergence is slow for the untruncated normal distribution, panel A. (This is presented as a limiting case; obviously Gibbs sampling is not the method of choice for this problem.) Even when p = 10,000, results are unreliable, as indicated by the convergence diagnostics. The problem arises from the strong serial correlation in the processes $\{g(x^{(j)})\}$, which is not fully evident in the estimated spectral densities for the smaller values of p; correspondingly, computed RNE falls as p increases. These results persist in the second case, in which z_2 is truncated at about 1.5 standard deviations above and below (Panel B), but are not so strong. In both cases the convergence diagnostic is an imperfect indicator of unreliable estimates of $\bar{g}$, for there are

several cases in which $\bar{g}_p$ is more than three times NSE from 0 (the known true value of $\bar{g}$ in all cases except D) and yet CD is less than 1.5 in absolute value. In the third case z_2 is truncated at about .15 standard deviations above and below (Panel C), and performance is satisfactory for all values of p. The same is true in the fourth case, in which the bivariate normal distribution is truncated to an extreme tail in both dimensions (Panel D), and in the fifth case, in which the the truncation produces a distribution closer to uniform than to bivariate normal (Panel E). Severe truncation diminishes the potential for strong serial correlation in the $x^{(j)}$, and thereby increases the efficiency of the Gibbs sampler.

The second example is constructed to resemble the truncated multivariate normal distribution that might be encountered in Bayesian inference with a multivariate probit model with panel data and serial correlation in equation disturbances for the same sampling unit and different years. Assuming three choices, five years, and a first-order autoregressive process for the disturbance leads to a variance matrix $\Sigma = R \otimes I_3$, $r_{ij} = \rho^{|i-j|}$ in a 15-variate normal with truncation restrictions that require one of x_{3j+1}, x_{3j+2}, and x_{3j+3} to be greater than the other two, for $j = 1, \ldots, 5$. Results are presented in Table 3 for four different values of ρ, ranging from $\rho = .00$ to $\rho = .95$. The number of passes and preliminary passes are the same as those in the previous example, and computation times range from 2.5 seconds for $p = 400$ to about 60 seconds for $p = 10{,}000$. As ρ increases, serial correlation in $z^{(j)}$ and hence $x^{(j)}$ increases, diminishing the efficiency and reliability of the Gibbs sampling algorithm. The convergence diagnostic proves to be a reliable indicator of the reliability of the estimates $\bar{g}_p$. For $\rho = .00$, 400 passes are reliable, despite some modest serial correlation; for $\rho = .50$, 2000 passes are required, and for $\rho = .80$, 10,000 passes are required. For $\rho = .95$, even 10,000 passes do not produce reliable results.

4. The Gibbs algorithm for truncated multivariate Student-t sampling

A closely related problem arising in Bayesian inference is the generation of samples from the multivariate Student-t distribution subject to linear restrictions,

$$x \sim T(\mu, \Sigma; m),\ a \le Dx \le b.$$

We continue to make the same assumptions about a, b, and D. The genesis of the multivariate Student-t as the ratio of a multidimensional normal to an independent $[\chi^2(m)/m]^{1/2}$ leads immediately to a Gibbs sampling algorithm for $(w, z_1, \ldots, z_n)$ followed by the construction $x = \mu + D^{-1}zw^{-1}$.

At the start of pass j, $w^{(j-1)}$ and $z^{(j-1)}$ are available from the previous pass. In the first step draw $w^{(j)} \sim [\chi^2(m)/m]^{1/2}$ subject to the restrictions

$$\alpha_i w^{(j)} \le z_i^{(j-1)} \le \beta_i w^{(j)} \qquad (i = 1, \ldots, n),$$

using an accept/reject procedure. In steps 2, ... , n+1, draw z from a multivariate normal distribution conditional on $w^{(j)}$, the pertinent z's, and the restrictions $\alpha_i w^{(j)} \le z_i^{(j)} \le \beta_i w^{(j)}$:

$$z_i^{(j)} = \sum_{j=1}^{i-1} c_{ij} z_i^{(j)} + \sum_{j=i+1}^{n} c_{ij} z_i^{(j-1)} + h_i \varepsilon_i,$$

$$\varepsilon_i \sim TN[(\alpha_i w^{(j)} - \sum_{j=1}^{i-1} c_{ij} z_i^{(j)} - \sum_{j=i+1}^{n} c_{ij} z_i^{(j-1)})/h_i, \ (\beta_i w^{(j)} - \sum_{j=1}^{i-1} c_{ij} z_i^{(j)} - \sum_{j=i+1}^{n} c_{ij} z_i^{(j-1)})/h_i)].$$

At the end of the pass, $x^{(j)} = \mu + D^{-1} z^{(j)} w^{(j)}$.

This algorithm was programmed in double precision Fortran-77 using the IMSL/STAT library. The routine was tested by comparing the distribution of truncated Student-t samples with those generated by a naive accept/reject procedure. No appreciable increases in computation time over corresponding problems with the truncated multivariate normal distribution were noted. In particular, the accept/reject procedure for $w^{(j)}$ appears quite efficient, even for $m = 2$. No considerations with respect to the efficiency of the Gibbs sampling algorithm, beyond those for the multivariate normal, have been noted.

5. Approximating the truncated multivariate normal integral

In some circumstances it is desirable to evaluate the the integral of the multivariate normal, or multivariate Student-t, over the region $a \le Dx \le b$. For example, in Bayesian inference one may wish to evaluate the posterior probability of a set of restrictions under a diffuse prior, while simultaneously generating variates from the restricted distribution. The estimation of this probability is a ready byproduct of the naive accept/reject procedure, although this method of estimation becomes impractical for very small probabilities as does the the sampling procedure itself. In the Gibbs sampling algorithm set forth here, the evaluation of this probability is not an immediate byproduct. Here, we describe an augmentation of the Gibbs algorithm that provides estimates of these probabilities. It entails a substantial increase in computing time, and thus it is competitive with the naive accept/reject method for small values of p.

A heuristic overview of the procedure will help the reader to follow the details. Consider a small neighborhood about any point in the region $a \le Dx \le b$, of sufficiently regular shape that its Euclidean volume is known. The probability of this neighborhood under the untruncated multivariate normal distribution may is approximately the product of the normal density at the point and the Euclidean volume of the neighborhood. The probability of this neighborhood under the truncated multivariate normal distribution may

be estimated from the Gibbs sample by the ratio of the number of $x^{(j)}$ in the neighborhood to the number of passes, p. Taking the limit as the size of the small neighborhood diminishes and p increases, the ratio of the first probability to the second converges to the probability of the region $a \le Dx \le b$ under the unrestricted multivariate normal distribution.

The following algorithm embodies this idea and has been incorporated in the software described previously. It assumes several constants, whose values are given parenthetically as they are introduced. These constants have been chosen to provide efficient and reliable estimates of the truncated multivariate normal integral based on experience with several problems, but there is no well-defined sense in which they are optimal. How the algorithm is extended to the multivariate Student-t distribution should be clear; this extension has been incorporated in the software.

Prior to the computation of the Gibbs sample $\{z^{(j)}\}_{j=1}^{p}$, at most n^* (= 1,000) points $s^{(1)}, s^{(2)}, \ldots$ are selected using the Gibbs sampling algorithm. (The fact that these points are not independent will turn out to be irrelevant. They could be selected by many other procedures, but the Gibbs sampler is quick and convenient.) About each point a neighborhood of the form $N(s^{(i)}) = \{z: (z-s^{(i)})'T^{-1}(z-s^{(i)}) \le r_i^2\}$ is designated. The shape is chosen because of its known Euclidean volume, $\pi^{n/2} r_i^n |T|^{1/2} / \Gamma(\frac{n}{2}+1)$ (which the reader may verify), and because the orientation of its axes parallel to those of the $N(0, T)$ distribution then minimizes the variation in the value of the corresponding density over the neighborhood. The radius r_i is chosen so that two criteria are satisfied. First, the ratio of the highest to the lowest value of the density should not exceed c^* (= 1.2). This implies $r_i = q_i^{1/2}(1 - a_i)$, with $a_i = [1 + (c/q_i)]^{1/2}$, $q_i = s^{(i)\prime}T^{-1}s^{(i)}$, and $c = 2\log(c^*)$. Second, the neighborhood should lie entirely within the set $\alpha \le z \le \beta$, equivalent to the $2n$ constraints $(\alpha_j - s_j^{(i)})T^{jj} \le r_i^2$, $(\beta_j - s_j^{(i)})T^{jj} \le r_i^2$. If necessary the value of r_i is reduced to satisfy these constraints. After r_i is chosen the approximate probability of the neighborhood under the multivariate normal distribution, $v_i = 2^{-n/2} r_i^n \exp(-\frac{1}{2} z'T^{-1}z) / \Gamma(\frac{n}{2}+1)$, is recorded. The construction of points and neighborhoods stops when $i = n^*$ or the sum of the v_i exceeds v^* (= 10.0), whichever comes first. Let v^+ denote the final sum of the v_i.

As the Gibbs sample proper, $\{z^{(j)}\}_{j=1}^{p}$, is constructed, each point $z^{(j)}$ is checked to see whether or not it is within each neighborhood. A running total of the number of times a point lies in a neighborhood is maintained; for each point $z^{(j)}$, the total is increased by the number of neighborhoods in which that point lies. Let $\tilde{n}$ denote the value of the running total at the end of the p passes. Then the estimated value of the multivariate normal integral $N(0, T)$ over the region $\alpha \le z \le \beta$ is $pv^+/\tilde{n}$.

Experimentation with this algorithm has produced probability estimates within .01 of the true value, where the true value has been computed analytically in simple cases or estimated by naive accept/reject methods when analytical computations are not possible, and the true probability is between .2 and .8. For smaller values of the true probability accuracy appears to be much better. A numerical standard error that measures the variability in pv+/ñ due to ñ may be constructed, but this would not reflect the approximation error due to $c^* > 1$ and the consequent bias in v as computed. These questions are left for future research.

References

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

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

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

Geweke, J., 1986: "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics* **1**, 127-142.

Geweke, J., 1989: "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* **57**, 1317-1339.

Geweke, J., 1991: "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," University of Minnesota manuscript. Presented at the Fourth Valencia International Meeting on Bayesian Statistics.

Hajivassiliou, V.A., and D.L. McFadden, 1990: "The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises," Colwes Foundation working paper, Yale University.

Hajivassiliou, V.A., D.L. McFadden, and P.A. Ruud, 1990: "Simulation of Multivariate Normal Orthan Probabilities: Methods and Programs," M.I.T. mimeo.

McFadden, D., 1989: "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica* **57**.

Rao, C.R., 1965: *Linear Statistical Inference and Its Applications*. New York: Wiley.

References

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

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

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

Geweke, J., 1986: "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics* **1**, 127-142.

Geweke, J., 1989: "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* **57**, 1317-1339.

Geweke, J., 1991: "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," University of Minnesota manuscript. Presented at the Fourth Valencia International Meeting on Bayesian Statistics.

Hajivassiliou, V.A., and D.L. McFadden, 1990: "The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises," Colwes Foundation working paper, Yale University.

Hajivassiliou, V.A., D.L. McFadden, and P.A. Ruud, 1990: "Simulation of Multivariate Normal Orthan Probabilities: Methods and Programs," M.I.T. mimeo.

McFadden, D., 1989: "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica* **57**.

Rao, C.R., 1965: *Linear Statistical Inference and Its Applications*. New York: Wiley.