Sampling Truncated Normal, Beta, and Gamma Densities

Paul DAMIEN and Stephen G. WALKER

We consider the Bayesian analysis of constrained parameter and truncated data problems within a Gibbs sampling framework and concentrate on sampling truncated densities that arise as full conditional densities within the context of the Gibbs sampler. In particular, we restrict attention to the normal, beta, and gamma densities. We demonstrate that, in many instances, it is possible to introduce a latent variable which facilitates an easy solution to the problem. We also discuss a novel approach to sampling truncated densities via a "black-box" algorithm, based on the latent variable idea, valid outside of the context of a Gibbs sampler.

Key Words: Gibbs sampler; Latent variables; Uniform random variables.

1. INTRODUCTION

Gelfand, Smith, and Lee (1992) described problems for which, in the context of a Gibbs sampler, there is a need to sample from truncated densities. Sampling from truncated densities can be nontrivial, even when the density is of a standard type, such as the normal, beta, or gamma. This article introduces a general methodology for sampling truncated densities, which arise as full conditionals within a Gibbs sampling framework (Smith and Roberts 1993; Tierney 1994). We also introduce a "black-box" algorithm for sampling truncated densities, applicable ouside of the context of a Gibbs sampler.

The methods rely on the introduction of latent variable(s). We show that for truncated normals, betas, and gammas, the introduction of a single latent variable reduces the problem of sampling such truncated densities to the sampling of a couple of uniform random variables. This is, from a coding and practical perspective, very appealing.

The idea of introducing a latent variable to improve or facilitate an "easy-to-sample" Gibbs sampler is not new. The method dates back at least to Swendsen and Wang (1987) who introduced latent variables for the Ising model. Recently, Higdon (1998) and Damien, Wakefield, and Walker (1999) demonstrated the use of latent variables for a large class of

Paul Damien is Assistant Professor, University of Michigan Business School, 701 Tappan, D3229, Ann Arbor, MI (E-mail: pdamien@umich.edu). Stephen G. Walker is Professor, Department of Mathematical Sciences, University of Bath, UK (E-mail: massgw@bath.ac.uk).

Journal of Computational and Graphical Statistics, Volume 10, Number 2, Pages 206–215

statistical models, the latter authors concentrating on Bayesian nonconjugate, nonlinear, and generalized linear mixed models.

Recall that the latent variable is being introduced into an existing Gibbs sampler and therefore it is the convergence of the overall chain that is important. The effect of introducting a latent variable into a Gibbs sampler is difficult to ascertain, but there is no reason why a dramatic loss in convergence should be the consequence. In fact, latent variables have been introduced in many instances for simplifying Gibbs samplers: Wakefield, Smith, Racine-Poon, and Gelfand (1994), Laud, Smith, and Damien (1996), Polson (1996) to cite a few. In particular, Polson notes that if latent variables induce densities which are log-concave then convergence can be improved. Damien et al. (1999) noted that ease of coding should also be taken into consideration when constructing Gibbs samplers, particularly in one-off applications.

Section 2 discusses sampling from truncated normal densities and Section 3 explores truncated beta and gamma densities, when they need to be sampled as a full conditional within an existing Gibbs sampler (Gelfand et al. 1992). We show that our approach relies on the introduction of a single latent variable, extending the loop by one more full conditional, which turns out to be a uniform density. Section 4 introduces a novel approach to sampling truncated densities, applicable in any context. Finally, Section 5 concludes with a discussion.

2. TRUNCATED MULTIVARIATE NORMAL DENSITY

This section describes a method for sampling truncated multivariate normal variables, within the context of a Gibbs sampler. To introduce our ideas we consider the standard univariate normal distribution. Let $X \sim \mathrm{N}_1(0,1)$; that is,

$$f_X(x) \propto \exp\left(-x^2/2\right).$$

Introduce the latent variable Y which has joint density with X given by

$$f_{X,Y}(x,y) \propto I_{(0,\exp(-x^2/2))}(y).$$

We then have the following full conditional densities:

$$\begin{aligned} Y|(X=x) &\sim \mathrm{U}\big(0, \exp(-x^2/2)\big), \\ X|(Y=y) &\sim \mathrm{U}\left(-\sqrt{-2\log y}, \sqrt{-2\log y}\right). \end{aligned}$$

Implementing this idea for a truncated standard normal variable is no more complicated. Suppose we wish to sample from the density given by

$$f_X(x) \propto \exp\left(-x^2/2\right) I\big(x \in (a,b)\big).$$

Again, we introduce the latent variable Y which has joint density with X given by

$$f_{X,Y}(x,y) \propto I_{(0,\exp(-x^2/2))}(y) I\big(x \in (a,b)\big),$$

leading to the full conditionals:

$$\begin{aligned} Y|(X=x) &\sim \mathrm{U}\big(0, \exp(-x^2/2)\big) \\ X|(Y=y) &\sim \mathrm{U}\left(\max\left\{a, -\sqrt{-2\log y}\right\}, \min\left\{b, \sqrt{-2\log y}\right\}\right). \end{aligned}$$

The algorithm extends the Gibbs loop by one more full conditional, which is a uniform distribution. The new full conditional for X is also uniform. This is, from a coding perspective, more appealing than a rejection-based algorithm (see, e.g., Devroye 1986; Robert 1995).

Robert (1995) proposed an extension of his rejection algorithm for the multivariate normal density. We can also simplify this method via the introduction of a latent variable. Consider

$$f_{X_1,\ldots,X_p}(x_1,\ldots,x_p) \propto \exp\left(-1/2(x-\mu)'\Sigma^{-1}(x-\mu)\right) I(x \in A),$$

where we assume, as did Robert, that the bounds for x_i given x_{-i} are available and given by (a_i, b_i). Therefore,

$$f_{X_i|X_{-i}}(x_i|x_{-i}) \propto \exp\left(-1/2(x_i-\nu_i)^2/\sigma_i^2\right) I\big(x_i \in (a_i,b_i)\big),$$

are the full conditionals, $\nu_i = \mu_i - \sum_{j\neq i}(x_j - \mu_i)e_{ij}/e_{ii}$ and $\sigma_i^2 = 1/e_{ii}$, where e_{ij} is the ijth element of Σ^{-1}. Robert (1995) used his rejection algorithm for sampling these truncated univariate normal densities. However, since we are already in a Gibbs sampler, it seems appropriate to implement the latent variable idea. We do not need to introduce p latent variables, one is sufficient.

We define the joint density of $(X_1,\ldots,X_p,Y)$ by

$$f_{X_1,\ldots,X_p,Y}(x_1,\ldots,x_p,y) \propto \exp(-y/2) I\left(y > (x-\mu)'\Sigma^{-1}(x-\mu)\right) I(x \in A).$$

The full conditional distributions are given by

$$f_{X_i|X_{-i},Y}(x_i|x_{-i},y) \propto I(x_i \in A_i),$$

where

$$A_i = (a_i,b_i) \cap B_i,$$

and B_i is the set $\{x_i|x_{-i} : (x-\mu)'\Sigma^{-1}(x-\mu) < y\}$ and so the bounds for B_i are obtained by solving a quadratic equation. The full conditional for $Y|X$ is a truncated exponential distribution which can be sampled using the cdf inversion technique.

Therefore, we have a Gibbs sampler that runs on $p+1$ full conditionals which can all be sampled directly using uniform variates, replacing the p full conditionals of Robert (1995) which are sampled via rejection algorithms.

Truncated $N_2(0,\Sigma)$ density. We consider the example presented by Robert (1995) that involves sampling a truncated $N_2(0,\Sigma)$ distribution, where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

so

$$\Sigma^{-1} = \begin{pmatrix} (1-\rho^2)^{-1} & -\rho(1-\rho^2)^{-1} \\ -\rho(1-\rho^2)^{-1} & (1-\rho^2)^{-1} \end{pmatrix}$$

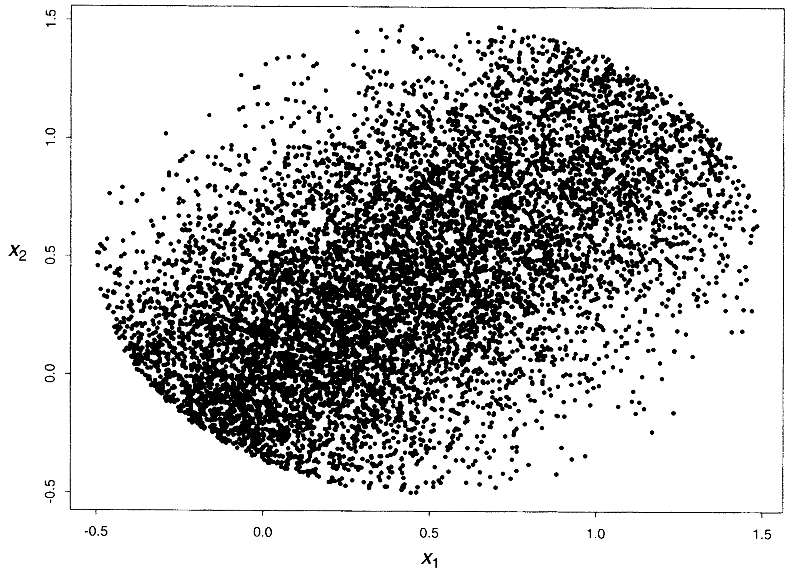

Figure 1. Plot of 10,000 samples from truncated $N_2(0, \Sigma)$ distribution obtained from Gibbs sampler.

and A is the circle centered on (γ_1, γ_2) with radius r. Therefore,

$$a_i = \gamma_i - \sqrt{r^2 - (\gamma_j - x_j)^2},$$

and

$$b_i = \gamma_i + \sqrt{r^2 - (\gamma_j - x_j)^2},$$

and

$$B_i = \left(x_j\rho - \sqrt{(y - x_j^2)(1 - \rho^2)}, x_j\rho + \sqrt{(y - x_j^2)(1 - \rho^2)}\right),$$

for $i(j) = 1(2), 2(1)$. Note that $y > x_i^2$ since $Y|X_1, X_2$ is an exponential distribution restricted to the set $([x_1^2 - 2\rho x_1 x_2 + x_2^2]/(1 - \rho^2), \infty)$ and $x_i^2(1 - \rho^2) < x_1^2 - 2\rho x_1 x_2 + x_2^2$, for $i = 1, 2$.

We performed the Gibbs sampler for the truncated bivariate normal distribution given and took $\rho = 0.9$, $(\gamma_1, \gamma_2) = (1/2, 1/2)$, and $r = 1$. We took 10,000 samples (with computing time of 5 seconds on a SPARC workstation). A scatterplot of the samples appears in Figure 1.

3. TRUNCATED BETA AND GAMMA DENSITIES

This section considers the problem of sampling from truncated beta and gamma distributions which arise as full conditionals within a Gibbs sampling framework.

Truncated beta density. We consider the density given up to proportionality by

$$f_X(x) \propto x^{\alpha-1}(1 - x)^{\beta-1} I\big(x \in (a, b)\big),$$

where $\alpha, \beta > 0$ and $0 \leq a < b \leq 1$. We introduce the latent variable Y which has joint density with X given by

$$f_{X,Y}(x,y) \propto x^{\alpha-1} I\big(y < (1-x)^{\beta-1}, x \in (a,b)\big).$$

The full conditional for Y is given by the uniform density on the interval $(0,(1-x)^{\beta-1})$. The full conditional for X depends on whether $\beta < 1$ or $\beta > 1$: if $\beta = 1$, then we can sample X directly using the inverse cdf technique and there would not be any need to introduce a latent variable. Let us assume that $\beta > 1$, leading to the full conditional for X given by

$$f_{X|Y}(x|y) \propto x^{\alpha-1} I\left(x \in \left(a, \min\left\{b, 1 - y^{1/(\beta-1)}\right\}\right)\right).$$

This is easily sampled using the inverse cdf technique and so involves sampling a uniform random variable. If $\beta < 1$, then the full conditional for X is the same but the indicator function for the full conditional for Y becomes

$$I\left(x \in \left(\max\left\{a, 1 - y^{1/(\beta-1)}\right\}, b\right)\right).$$

Truncated gamma density. Here we consider the density given by

$$f_X(x) \propto x^{\alpha-1} \exp(-x) I\big(x \in (a,b)\big),$$

where $0 \leq a < b \leq \infty$. We introduce the latent variable Y with joint density with X given by

$$f_{X,Y}(x,y) \propto x^{\alpha-1} I\big(y < \exp(-x), x \in (a,b)\big)$$

leading to the full conditional for Y being the uniform density on the interval $(0, \exp(-x))$ and the full conditional for X being given by the density

$$f_{X|Y}(x|y) \propto x^{\alpha-1} I\big(x \in \big(a, \min\{b, -\log y\}\big)\big)$$

which can also be sampled using the inverse cdf technique.

4. ADAPTIVE UNIFORM REJECTION SAMPLING

The material in this section arise as a direct consequence of the work in Sections 2 and 3. Based on the earlier development, note that the marginal density function of the latent variables always satisfy a monotone decreasing property on a bounded interval. This makes it possible to sample using an adaptive rejection routine described in this section. The type of density function which can be sampled using this method is of the type $f(x) \propto l(x)\pi(x)$, where

1. $\pi(\cdot)$ is a proper density for which it is possible to sample from $\pi(\cdot)$, and truncated versions of $\pi(\cdot)$, and $\Pi(A) = \int_A \pi(dx)$.

2. $l(\cdot) \leq c < \infty$ and $A_u = \{x : l(x) > u\}$ where $\Pi(A_0) = 1$, $\Pi(A_c) = 0$ and $A_u \subset A_v$ iff $v < u$.

If $f(x, u) \propto I(u < l(x))\pi(x)$, then the marginal density for u is given by

$$f(u) \propto \Pi(A_u) I(0 < u < c),$$

and, obviously, the marginal density for x is given by $f(x) \propto l(x)\pi(x)$. If we let $h(u) = \Pi(A_u)I(0 < u < c)$ then $h(\cdot)$ is a monotone decreasing function on $(0, c)$, $h(0) = 1$ and $h(c) = 0$. Thus, instead of using a Gibbs sampler on $f(x, u)$ we obtain $f(u)$, sample from $f(u)$, and then sample from $f(x|u)$.

Next we show how to sample from the density proportional to $h(\cdot)$ using an adaptive rejection sampling algorithm, similar in spirit to that of Gilks and Wild (1992).

The AURS algorithm. Let $h(\cdot)$ be a nonincreasing continuous function on $(0, c)$ such that $0 < h(0) < \infty$ and $h(c) = 0$ for some $0 < c < \infty$. We sample from $f(u) \propto h(u)$ using an iterative adaptive rejection sampling algorithm. At the ith iteration of the algorithm we have evaluated $h(u_j)$ at u_j for $j = 1, \ldots, i+1$, where $0 = u_1 < u_2 < \cdots < u_i < u_{i+1} = c$. Let

$$g_i(u) \propto \sum_{j=1}^{i} h(u_j) I(u_j < u < u_{j+1}).$$

Take u^* from $g_i(\cdot)$ and w from the uniform distribution on the interval $(0, 1)$. If

$$w < \frac{h(u^*)}{h(u_j)},$$

where $u_j \leq u^* < u_{j+1}$, then we accept u^* as a random variate from $f(u)$, else we proceed to the $(i+1)$th iteration with $g_{i+1}(\cdot)$, which includes the $(u^*, h(u^*))$. Note then the algorithm is straightforward to implement, not requiring maximizations or differentiation.

Therefore, to sample from $f(x)$, we sample from $f(u) \propto h(u)$ using the AURS scheme, then sample from $f(x|u) \propto \pi(x)I(x \in A_u)$. The algorithm splits up into two parts:

(a) Use the AURS algorithm on $h(\cdot)$ to sample u.

(b) Sample x from $\pi(x)I(x \in A_u)$.

We demonstrate the efficiency of the AURS algorithm; that is, for part (a) only, on particular beta densities $f(u) \propto (1-u)^p I(0 < u < 1)$. For large p this will be the "worst" type of density that could be put through the AURS algorithm. This is because it is far from uniform, which is the starting proposal density. In the following, we simulate 10,000 random variables from $f(u) \propto (1-u)^p I(0 < u < 1)$ for a range of values of p and compute the mean number of iterations of the algorithm per sample and the length of time for the task to be completed. The results are presented in Table 1. For the last case, when $p = 1{,}000$, we provide the histogram of the samples in Figure 2. The mean of the sample is 0.000997.

Table 1. Beta Density AUR Sampling

p	*Mean number of iterations per sample*	*Time for samples (secs)*
1	1.79	1
5	3.18	2
10	3.88	2
50	5.59	2.5
100	6.31	3.0
1,000	8.69	3.5

Next we demonstrate how to use the AURS algorithm to sample truncated normal, beta, and gamma densities. This makes it possible to sample outside of the full conditional context of a Gibbs sampler using the latent variable idea.

Normal density. Let $f(x) \propto \exp(-x^2/2)I(a < x < b)$. We will consider three cases: (i) a, b finite, (ii) $a = -\infty$, b finite, and (iii) a finite, $b = +\infty$.

(i) If we take $f(x, u) \propto I(u < \exp(-x^2/2))I(a < x < b)$, then we have $h(u) = |(a, b) \cap A_u| I(0 < u < 1)$, where $|A| = \int_A dx$, $\pi(x)$ is the uniform distribution on the interval (a, b) and $A_u = (-\sqrt{-\log u}, +\sqrt{-\log u})$. Therefore, $f(x|u)$ is the uniform distribution on the interval $(a, b) \cap A_u$.

(ii) We introduce the joint density

$$f(x, u) \propto I\left(u < \exp(-0.5x^2)\right) \exp(-0.5x^2) I(x < b),$$

leading to

$$h(u) = \left[\Phi\left(\min\left\{b, \sqrt{-2\log u}\right\}\right) - \Phi\left(-\sqrt{-2\log u}\right)\right] I(0 < u < 1),$$

$\pi(x)$ is the standard normal distribution, restricted to the set $(-\infty, b)$, and

$$A_u = \left(-\sqrt{-2\log u}, +\sqrt{-2\log u}\right).$$

Here $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. We can sample from $f(x|u) \propto \pi(x)I(x \in A_u)$ using (i).

(iii) This follows from an obvious modification of (ii).

Beta density. We let $f(x) \propto x^{\alpha-1}(1-x)^{\beta}I(a < x < b)$, $\alpha > 0, \beta > -1$, with $0 < a < b \leq 1$ or $0 \leq a < b < 1$.

At first we assume that $\beta > 0$ and introduce the joint density

$$f(x, u) \propto I\left(u < (1-x)^{\beta}\right) x^{\alpha-1} I(a < x < b),$$

leading to

$$h(u) = \left[\max\left\{b^{\alpha}, \left(1 - u^{1/\beta}\right)^{\alpha}\right\} - a^{\alpha}\right] I\left(0 < u < (1-a)^{\beta}\right),$$

$\pi(x)$ is the beta$(\alpha, 1)$ distribution on the interval (a, b) and $A_u = (0, 1 - u^{1/\beta})$. We can sample from $f(x|u)$ using the inverse cdf technique.

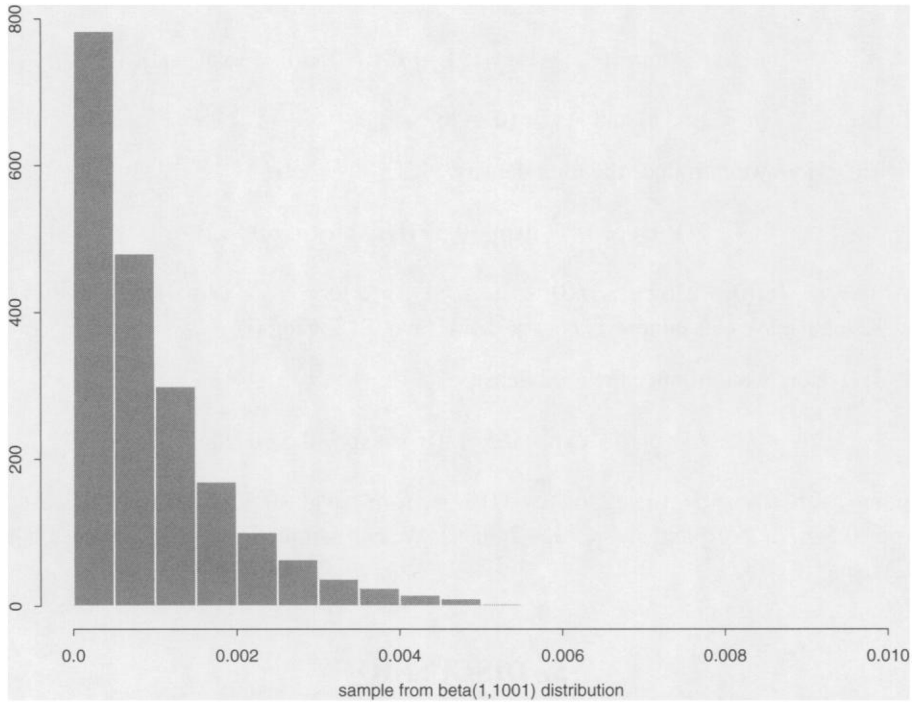

Figure 2. Plot of 10,000 samples from beta(1,1,001) distribution obtained via AURS algorithm.

Our method here depends on $\beta > 0$. If $-1 < \beta < 0$ and $\alpha > 1$, then we simply use

$$f(x,u) \propto I\left(u < x^{\alpha-1}\right)(1-x)^{\beta} I(a < x < b)$$

instead. If both $\alpha < 1$ and $\beta < 0$ then the problem is that both $x^{\alpha-1}$ and $(1-x)^{\beta}$ are unbounded on $(0,1)$, though we can obviously proceed if either $a \neq 0$ or $b \neq 1$, which is assumed.

Gamma density. Here we consider the density given by

$$f(x) \propto x^{\alpha-1}\exp(-x) I(a < x < b),$$

and look at four cases: (i) $a = 0$, b finite, (ii) $a > 0$, b finite, (iii) $a = 0$, $b = \infty$, (iv) $a > 0$, $b = \infty$. Throughout, we will use $IG(a,b;\alpha)$ to denote the integral $\int_{a<x<b} x^{\alpha-1}$ $\exp(-0.5x)dx$.

(i) Here we introduce the joint density

$$f(x,u) \propto x^{\alpha-1} I(u < \exp(-x)) I(0 < x < b).$$

Then we have $h(u) = \min\{b^{\alpha}, (-\log u)^{\alpha}\} I(0 < u < 1)$, $\pi(x) \propto x^{\alpha-1} I(0 < x < b)$, and $A_u = (0, -\log u)$.

(ii) Here we introduce the joint density

$$f(x,u) \propto x^{\alpha-1} I(u < \exp(-x)) I(a < x < b),$$

leading to

$$h(u) = \left[\min\{b^{\alpha}, (-\log u)^{\alpha}\} - a^{\alpha}\right] I(0 < u < \exp(-a)),$$

$\pi(x) \propto x^{\alpha-1} I(a < x < b)$ and $A_u = (0, -\log u)$.

(iii) Here we introduce the joint density

$$f(x, u) \propto x^{\alpha-1} \exp(-0.5x) I(u < \exp(-0.5x)),$$

so $h(u) = IG(0, -2\log u; \alpha) I(0 < u < 1)$, $\pi(x) \propto x^{\alpha-1}\exp(-0.5x)$ and $A_u = (0, -2\log u)$. We can sample $f(x|u) \propto \pi(x) I(x \in A_u)$ using (i).

(iv) Here we introduce the joint density

$$f(x, u) \propto x^{\alpha-1} \exp(-0.5x) I(u < \exp(-0.5x)) I(x > a),$$

leading to $h(u) = IG(a, -2\log u; \alpha) I(0 < u < \exp(-0.5a))$, with $\pi(x) \propto x^{\alpha-1} \exp(-0.5x) I(x > a)$ and $A_u = (0, -2\log u)$. We can sample from $f(x|u) \propto \pi(x) I(x \in A_u)$ using (ii).

5. DISCUSSION

Our approach to sampling truncated densities is based on the introduction of strategic latent variables. Having done this, one can proceed in one of two ways, which will depend on the context in which the truncated sample is required. The first is applicable in the context of a Gibbs sampler. One incorporates the sampling of the full conditional density of the latent variable into the original Gibbs loop. The idea of latent variables being introduced into Gibbs samplers is not new in general, however it appears to be new with respect to truncated densities. The second relies on obtaining the marginal density of the latent variable, which takes on a special form, allowing it to be sampled directly via an adaptive algorithm. With this, the original truncated density can be sampled directly via the conditional density, conditional on the latent variable. This idea is new. Moreover, from the work presented in Section 4, it is clear that the algorithm here can be used to sample a large class of densities, not just truncated versions of well known densities.

[Received August 1999. Revised April 2000.]

ACKNOWLEDGMENTS

The authors are grateful to a referee for constructive comments on an earlier draft of the article.

REFERENCES

Damien, P., Wakefield, J. C., and Walker, S. G. (1999), "Gibbs Sampling for Bayesian Non-conjugate and Hierarchical Models Using Auxiliary Variables," revised for *Journal of the Royal Statistical Society*, Series B, 61, 331–344.

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

Gelfand, A. E., Smith, A. F. M., and Lee, T. M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling," *Journal of the American Statistical Association*, 87, 523–532.

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Higdon, D. (1998), "Auxiliary Variable Methods for Markov Chain Monte Carlo Methods With Applications," *Journal of the American Statistical Association*, 93, 585–595.

Laud, P. W., Smith, A. F. M., and Damien, P. (1996), "Monte Carlo Methods for Approximating a Posterior Hazard Rate Process," *Statistics and Computing*, 6, 77–84.

Polson, N. G. (1996), "Convergence of Markov Chain Monte Carlo Algorithms," in *Bayesian Statistics 5*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, Oxford: Oxford University Press, pp. 297–321.

Robert, C. P. (1995), "Simulation of Truncated Normal Variables," *Statistics and Computing*, 5, 121–125.

Smith, A. F. M., and Roberts, G. O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Series B, 55, 3–24.

Swendsen, R. H., and Wang, J. S. (1987), "Nonuniversal Critical Dynamics in Monte Carlo Simulations," *Physical Review Letters*, 58, 86–88.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E. (1994), "Bayesian Analysis of Linear and Nonlinear Population Models Using the Gibbs Sampler," *Applied Statistics*, 43, 201–221.