KRESGE
BUS. ADM.
LIBRARY

RESEARCH SUPPORT
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

JULY 1996

SAMPLING PROBABILITY DENSITIES VIA UNIFORM RANDOM VARIABLES AND A GIBBS

WORKING PAPER #9612-05

PAUL DAMIEN
AND
STEPHEN WALKER

UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

Sampling Probability Densities via Uniform Random Variables and a Gibbs Sampler

Paul Damien[1]
and
Stephen Walker[2]

[1] Department of Statistics and Management Science, School of Business, University of Michigan, Ann Arbor, 48109-1234, USA.

[2] Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ

SUMMARY

Continuous densities on the real line are sampled given only an iid sequence of uniform random variables on the interval $(0,1)$ and using a Gibbs sampling scheme. The method is exemplified via several examples. A feature of the paper is to provide an alternative sampling algorithm to rejection based methods and other sampling approaches such as the Metropolis-Hastings algorithms, especially from a Bayesian perspective.

Keywords: Gibbs sampler, Latent variables, Uniform density.

1 Introduction

Let f be a continuous density function defined on the real line. We address the problem of generating a random variate X from f. Starting with the assumption that it is possible to sample uniform random variables from the interval $(0,1)$ to provide an iid sequence $U_1, U_2, \ldots$ of such variables.

The basic idea is to introduce a latent variable Y, construct the joint density of Y and X with marginal density for X given by f, and to use the Gibbs sampler (see, for example, Smith and Roberts, 1993) to generate random variates from f. This is done by simulating a Markov chain $\{X_n\}$ where given $X_n = x$, Y is taken from $f(y|x)$ and then X_{n+1} is taken from $f(x|Y = y)$. Under mild regularity conditions $X_n \to_d X \sim f$. Additionally we are looking for the conditional distributions which can be sampled using the appropriate uniform random variables. For a historical overview of Markov chain methods and the use of latent (auxilliary) variables the reader is referred to Besag and Green (1993).

With the widespread use of the Gibbs sampler this paper, especially from a Bayesian perspective, is relevant. In particular, the new algorithm, after some preliminary analysis (details later), provides a Gibbs sampler in which all the full conditional densities are of known type. This then may lead to a more efficient method than the Metropolis-Hastings algorithm (Tierney, 1994), the adaptive rejection method for log-concave densities (Gilks and Wild, 1992) and other rejection algorithms, in many contexts.

Preliminaries

Firstly being able to sample a uniform random variable from the interval $(0,1)$ allows us to sample a uniform random variable from any interval (a,b) and we write such a density as $\mathcal{U}(a,b)$.

Let $f(x) \propto \exp(-x\beta)I(a < x < b)$, where $\beta > 0$, $0 \le a < b \le \infty$ and I represents the indicator function. We write such a density as $\mathcal{E}(\beta,a,b)$. Sampling X from $\mathcal{E}(\beta,a,b)$ can be done by taking U from $\mathcal{U}(1-\exp(-a\beta), 1-\exp(-b\beta))$ and taking $X = -1/\beta \log(1-U)$.

Let $f(x) \propto (1-x)^{\beta-1}I(a < x < b)$, where $\beta > 0$ and $0 \le a < b \le 1$. We write such a density as $\mathcal{B}(\beta,a,b)$ and it is easy to see that if $f(y) = \mathcal{E}(\beta, -\log(1-a), -\log(1-b))$ and $X = 1-\exp(-Y)$ then $f(x) = \mathcal{B}(\beta,a,b)$. Thus sampling an X from $\mathcal{B}(\beta,a,b)$ follows from above. Alternatively, one could use the inverse transform method to sample from $\mathcal{B}(\beta,a,b)$.

In subsequent sections, we will show that $\mathcal{E}(\beta,a,b)$, $\mathcal{B}(\beta,a,b)$ and $\mathcal{U}(a,b)$, along with an appropriately defined Gibbs sampler, are sufficient to generate random variates from most of the standard densities, and many nonstandard

AUG 1 2 1996 /

and complicated densities as well.

2 Densities from Johnson and Kotz

While there is not an obvious theorem underlying the proposed method there *is* a common thread which will be exemplified. Here we state this as a 'rule-of-thumb':

(1) Given a target density f, defined up to a constant of proportionality, identify factors in f which can be substituted easily via latent variables that induces full conditional densities, and which can be sampled using a uniform random variable.

(2) Set up the full conditional distributions for the Gibbs Sampler based on (1).

Example 1. Normal

First we consider the normal(0, 1) density with density function $f(x) \propto \exp(-0.5x^2)$. We introduce the latent variable Y, defined on $(0, \infty)$, which has joint density function with X given, up to a constant of proportionality, by

$$f(x, y) \propto \exp(-0.5y) I(y > x^2).$$

Clearly the marginal density for X is the normal(0, 1) density. The conditional densities are given by

$$f(y|x) = \mathcal{E}\Big(0.5, x^2, \infty\Big)$$

and

$$f(x|y) = \mathcal{U}\Big(-\sqrt{y}, +\sqrt{y}\Big).$$

For a normal(μ, σ) density we simply generate a $X_{(0,1)}$ variable and take $X_{(\mu,\sigma)} = \sigma X_{(0,1)} + \mu$.

Note that the sampling of the normal(0, 1) density could also be done by defining the joint density by

$$f(x, y) \propto I\Big(y < \exp(-0.5x^2)\Big).$$

The conditional densities are now given by

$$f(y|x) = \mathcal{U}\Big(0, \exp(-0.5x^2)\Big)$$

and

$$f(x|y) = \mathcal{U}\Big(-\sqrt{-2\log(y)}, +\sqrt{-2\log(y)}\Big).$$

It is no more difficult to sample a truncated normal$(0,1)$ variate with density $f(x) \propto \exp(-0.5x^2)I(a < x < b)$. The joint density of X and Y is given by

$$f(x,y) \propto \exp(-0.5y)I\Big(y > x^2, a < x < b\Big).$$

The conditional density $f(y|x)$ remains unchanged but the conditional density $f(x|y)$ becomes $\mathcal{U}(\max\{-\sqrt{y}, a\}, \min\{+\sqrt{y}, b\})$.

Example 2. Gamma

We first consider the gamma$(\alpha, 1)$ density with density function given up to a constant of proportionality by $f(x) \propto x^{\alpha-1}\exp(-x)I(x > 0)$ for $\alpha > 0$. We introduce the latent variable Y, defined on $(0,\infty)$, which has joint density function with X given, up to a constant of proportionality, by

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

The conditional densities are given by

$$f(y|x) = \mathcal{U}\Big(0, x^{\alpha-1}\Big)$$

and

$$f(x|y) = \begin{cases} \mathcal{E}(1, y^{1/(\alpha-1)}, \infty) & \text{if } \alpha > 1 \\ \mathcal{E}(1, 0, (1/y)^{1/(1-\alpha)}) & \text{if } \alpha < 1. \end{cases}$$

Trivially $\alpha = 1$ implies $f(x) = \mathcal{E}(1, 0, \infty)$. Such a Gibbs algorithm may be appropriate when $\alpha = \epsilon < 1$ is very small. For a gamma(α, β) density we sample a $X_{(\alpha,1)}$ variable and take $X_{(\alpha,\beta)} = 1/\beta X_{(\alpha,1)}$. Again, as in example 1, sampling a truncated gamma density will pose no extra problem.

Example 3. Beta

Here we have $f(x) \propto x^{\alpha-1}(1-x)^{\beta-1}I(0 < x < 1)$ for $\alpha, \beta > 0$. We introduce the latent variable Y, defined on $(0,\infty)$, which has joint density function with X given, up to a constant of proportionality, by

$$f(x,y) \propto I(y < x^{\alpha-1}, 0 < x < 1)(1-x)^{\beta-1}.$$

The conditional densities are given by

$$f(y|x) = \mathcal{U}\Big(0, x^{\alpha-1}\Big)$$

and

$$f(x|y) = \begin{cases} \mathcal{B}(\beta, y^{1/(\alpha-1)}, 1) & \text{if } \alpha > 1 \\ \mathcal{B}(\beta, 0, (1/y)^{1/(1-\alpha)}) & \text{if } \alpha < 1. \end{cases}$$

Trivially $\alpha = 1$ implies $f(x) = \mathcal{B}(\beta, 0, 1)$. Again, as in examples 1 and 2, sampling a truncated beta density will pose no extra problem. We note that the above method overcomes numerical problems when one wishes to sample from a beta distribution with parameters that are less than one *and* very small.

It is apparent that this method can be applied to other well known densities as well. For example, the Student's t, chi-squared and Weibull densities can be sampled using transformations or mixtures of the above.

Example 4. Cauchy

The Cauchy$(0,1)$ has density given by $f(x) \propto 1/(1+x^2)$. We define the joint density of X and Y, a random variable defined on $(0,1)$, by $f(x,y) \propto I(y < 1/(1+x^2))$. The full conditional densities are given by $f(y|x) = \mathcal{U}(0, 1/(1+x^2))$ and $f(x|y) = \mathcal{U}(-\sqrt{1/y - 1}, +\sqrt{1/y-1})$.

Example 5. Pareto

The Pareto(a, α) density is given, up to a constant of proportionality, by $f(x) \propto 1/x^{a+1} I(x > \alpha)$, where $\alpha, a > 0$. We define the joint density of X and Y, a random variable defined on $(0, \infty)$, by $f(x,y) \propto I(y < 1/x^{a+1}, x > \alpha)$. The full conditional densities are given by $f(y|x) = \mathcal{U}(0, 1/x^{a+1})$ and $f(x|y) = \mathcal{U}(\alpha, 1/y^{1/(a+1)})$.

Example 6. Inverse Gaussian

We consider the Inverse Gaussian with density given, up to a constant of proportionality, by

$$f(x) \propto \sqrt{1/x^3} \exp\Big(-x - 1/x\Big) I(x > 0).$$

Here we introduce two latent variables Y and Z, defined on $(0, \infty)$ and $(0,1)$, respectively, such that their joint density with X is given, up to a constant of proportionality, by

$$f(x,y,z) \propto \exp(-x) I\Big(y < \sqrt{1/x^3}\Big) I\Big(z < \exp(-1/x)\Big) I(x > 0).$$

The full conditional densities are given by

$$f(y|x,z) = \mathcal{U}\Big(0, \sqrt{1/x^3}\Big),$$

$$f(z|y,x) = \mathcal{U}\Big(0, \exp(-1/x)\Big)$$

and

$$f(x|y,z) = \mathcal{E}\Big(1, -1/\log(z), (1/y)^{2/3}\Big).$$

3 Nonstandard Densities

This method and other related concepts in the context of *neutral to the right* processes (Doksum, 1974) are studied in detail in Walker and Damien (1996). In this section we exemplify our method to densities encountered within the context of Bayesian nonparametrics, and which have appeared in recent literature: see, for example, Damien (1994), Damien et al. (1995,1996), Laud et al. (1993,1996) and Walker (1995,1996). In addition, we illustrate the method for some Bayesian non-conjugate models.

Example 7. D-Distributions

The class of D-distributions was introduced by Laud (1977). A random variable X on $(0,\infty)$ is said to have a D-distribution with parameters $\alpha, \beta > 0$, $\gamma \geq 0$, if its density function is given, up to a constant of proportionality, by

$$f(x) \propto x^{\alpha-1} \exp(-\beta x)\{1 - \exp(-x)\}^{\gamma}.$$

Here we introduce the latent variable Y, defined on the interval $(0,1)$, such that the joint density of X and Y is given, up to a constant of proportionality, by

$$f(x,y) \propto x^{\alpha-1} \exp(-\beta x)(1-y)^{\gamma-1} I\Big(y > \exp(-x)\Big).$$

The conditional densities are given by

$$f(y|x) = \mathcal{B}\Big(\gamma, \exp(-x), 1\Big)$$

and

$$f(x|y) \propto x^{\alpha-1}\exp(-\beta x)I\Big(x > -\log(y)\Big),$$

a truncated gamma(α,β) density (see, example 2). The D-distributions are special cases from the class of SD-distributions (Damien et al., 1995) and the method of this paper can be used to sample from such distributions (Walker, 1995). We omit details here.

Example 8. Diaconis/Kemperman

Here we consider the unusual density uncovered by Diaconis and Kemperman (1996) which is given, up to a constant of proportionality, by

$$f(x) \propto (1-x)^{-1+x}x^{-x}\sin(x\pi)I(0<x<1).$$

Now we introduce five latent variables U, V, W, all defined on $(0,1)$, Y and Z, both defined on $(0,\infty)$, such that their joint density with X is given, up to a constant of proportionality, by

$$f(u,v,x,w,y,z) \propto I(w<\sin(x\pi))$$

$$\times I\Big(u<\exp(-xy), v<\exp\{-(1-x)z\}, y>-\log(1-x), z>-\log(x)\Big).$$

The conditional densities are given by

$$f(u|v,w,x,y,z) = \mathcal{U}\Big(0,\exp(-xy)\Big),$$

$$f(v|w,x,y,z,u) = \mathcal{U}\Big(0,\exp\{-(1-x)z\}\Big),$$

$$f(w|x,y,z,u,v) = \mathcal{U}\Big(0,\sin(x\pi)\Big),$$

$$f(y|z,u,v,w,x) = \mathcal{U}\Big(-\log(1-x), -x^{-1}\log(u)\Big),$$

$$f(z|u,v,w,x,y) = \mathcal{U}\Big(-\log(x), -(1-x)^{-1}\log(v)\Big)$$

and

$$f(x|y,z,u,v,w) =$$

$$\mathcal{U}\Big(\max\{a_w,\exp(-z),1+z^{-1}\log(v)\}, \min\{b_w,-y^{-1}\log(u),1-\exp(-y)\}\Big),$$

where $(a_w,b_w) = A_w = \{x : \sin(x\pi) > w)$. It is not clear that this density could be sampled in any other way.

Next we consider some common Bayesian non-conjugate models.

Bayesian Non-conjugate Models

If a posterior is given by $f(x) \propto l(x)\pi(x)$, where $l(.)$ represents the likelihood and $\pi(.)$ the prior, then the general idea is to introduce the latent variable Y, defined on the interval $(0, \infty)$ or more strictly the interval $(0, l(\hat{\theta}))$, where $\hat{\theta}$ maximises $l(.)$, and define the joint density with X by

$$f(x, y) \propto I\Big(y < l(x)\Big)\pi(x).$$

The full conditional for Y is $\mathcal{U}(0, l(x))$ and the full conditional for X is π restricted to the set $A_y = \{x : l(x) > y\}$.

Example 9. Poisson/log-normal model

A Poisson likelihood with log-normal prior produces the posterior

$$f(x) \propto \exp\{nx - \exp(x)\}\exp(-0.5x^2),$$

where we assume without loss of generality that the prior is normal$(0, 1)$. We introduce the latent variable Y, defined on the interval $(0, \infty)$, such that the joint density with X is given by

$$f(x, y) \propto \exp(-y)I\Big(y > \exp(x)\Big)\exp\{-0.5(x^2 - 2nx)\},$$

which leads to the conditional densities given by

$$f(y|x) = \mathcal{E}\Big(1, \exp(x), \infty\Big)$$

and

$$f(x|y) \propto \exp\{-0.5(x - n)^2\}I\Big(-\infty < x < \log(y)\Big),$$

a truncated normal$(n, 1)$ density (see, example 1).

Example 10. Binomial/logit model

Here we have $m|x \sim$ binomial(x, n) with a normal$(0, 1)$ prior for $\log\{x/(1-x)\}$ which produces the posterior

$$f(x) \propto \exp(mx)\{1 + \exp(x)\}^{-n}\exp(-0.5x^2).$$

We introduce the latent variable Y, defined on the interval $(0, 1)$, such that the joint density with X is given by

$$f(x, y) \propto I\Big(y < \{1 + \exp(x)\}^{-n}\Big)\exp\{-0.5(x^2 - 2mx)\},$$

which leads to the conditional densities given by

$$f(y|x) = \mathcal{U}\Big(0, \{1+\exp(x)\}^{-n}\Big)$$

and

$$f(x|y) \propto \exp\{-0.5(x-m)^2\} I\Big(-\infty < x < \log\{1/y^{1/n} - 1\}\Big),$$

a truncated normal$(m, 1)$ density (see, example 1).

Example 11. Bernoulli/logistic regression model

Consider the following model in which

$$y_i | X = x \sim \text{Bernoulli}\Big(1/\{1+\exp(-\mu - xz_i)\}\Big),\ i = 1, ..., n,$$

where $X \sim$ normal$(0, 1)$ is the prior (we assume μ is known). The posterior density for X is given, up to a constant of proportionality, by

$$f(x) \propto \exp(-0.5x^2)\Pi_{i=1}^n \Big\{1+\exp(-\mu - xz_i)\Big\}^{-y_i} \Pi_{i=1}^n \Big\{1+\exp(\mu + xz_i)\Big\}^{y_i - 1}.$$

Here we introduce the latent variables $U = (U_1, ..., U_n)$ and $V = (V_1, ..., V_n)$, where both U and V are defined on $(0,1)^n$, such that their joint density with X is given, up to a constant of proportionality, by

$$f(x, u, v) \propto \exp(-0.5x^2)$$

$$\times \Pi_{i=1}^n \Big(u_i < \Big\{1+\exp(-\mu - xz_i)\Big\}^{-y_i}\Big) \Pi_{i=1}^n \Big(v_i < \Big\{1+\exp(\mu + xz_i)\Big\}^{y_i - 1}\Big).$$

The full conditional densities $f(u_i | u_{-i}, v, x)$ and $f(v_i | v_{-i}, u, x)$ are all uniform. For example,

$$f(u_i | u_{-i}, v, x) = \mathcal{U}\Big(0, \Big\{1+\exp(-\mu - xz_i)\Big\}^{-y_i}\Big)$$

and

$$f(v_i | v_{-i}, u, x) = \mathcal{U}\Big(0, \Big\{1+\exp(\mu + xz_i)\Big\}^{y_i - 1}\Big).$$

Let $\mathcal{S} = \{i : y_i = 1\} \cap \{i : z_i \neq 0\}$ and $\mathcal{R} = \{i : y_i = 0\} \cap \{i : z_i \neq 0\}$. Then

$$f(x | u, v) \propto \exp(-0.5x^2) I(x \in A_{uv}),$$

where $A_{uv} = (\max_{i\in\mathcal{S}}\{a_i\}, \min_{i\in\mathcal{R}}\{b_i\})$, $a_i = \{\log(1/u_i - 1) - \mu\}/z_i$ and $b_i = \{\log(1/v_i - 1) - \mu\}/z_i$. This is then a truncated normal density. Note that if $\mathcal{S} = \emptyset$ then replace $\max_{i\in\mathcal{S}}\{a_i\}$ by $-\infty$ and if $\mathcal{R} = \emptyset$ then replace $\min_{i\in\mathcal{R}}\{b_i\}$ by $+\infty$.

Example 12. Probit model

Here we have the posterior density given, up to a constant of proportionality, by

$$f(\beta) \propto \Pi_{i=1}^n \Big\{\Phi(\beta_0 + \beta_1 z_i)\Big\}^{y_i} \Pi_{i=1}^n \Big\{1 - \Phi(\beta_0 + \beta_1 z_i)\Big\}^{n_i - y_i} \pi(\beta),$$

where we assume a multivariate normal(μ, Σ) prior for β and Φ is the standard normal distribution function. We introduce the latent variables $U = (U_1, ..., U_n)$ and $V = (V_1, ..., V_n)$, where both U and V are defined on $(0,1)^n$, such that their joint density with β is given, up to a constant of proportionality, by

$$f(\beta, u, v) \propto \Pi_{i=1}^n \Big(u_i < \Big\{\Phi(\beta_0 + \beta_1 z_i)\Big\}^{y_i}\Big) \Pi_{i=1}^n \Big(v_i < \Big\{1 - \Phi(\beta_0 + \beta_1 z_i)\Big\}^{n_i - y_i}\Big) \pi(\beta).$$

The full conditional densities $f(u_i|u_{-i}, v, \beta)$ and $f(v_i|v_{-i}, u, \beta)$ are uniform (see example 11 to see how to obtain them) and interest is in $f(\beta_k|\beta_{-k}, u, v)$. Let $a_i = \Phi^{-1}(\tau_i) - \beta_1 z_i$, $b_i = (\Phi^{-1}(\tau_i) - \beta_0)/z_i$, $c_i = \Phi^{-1}(\lambda_i) - \beta_1 z_i$ and $d_i = (\Phi^{-1}(\lambda_i) - \beta_0)/z_i$, where $\tau_i = u_i^{1/y_i}$ and $\lambda_i = 1 - v_i^{1/(n_i - y_i)}$. As in example 11 b_i and d_i are only defined for those $z_i \neq 0$, a_i and b_i are only defined when $y_i > 0$ and c_i and d_i are only defined when $n_i > y_i$. Then

$$f(\beta_0|\beta_1, u, v) \propto \pi(\beta_0|\beta_1) I\Big(\max_i\{a_i\} < \beta_0 < \min_i\{c_i\}\Big)$$

and

$$f(\beta_1|\beta_0, u, v) \propto \pi(\beta_1|\beta_0) I\Big(\max_i\{b_i\} < \beta_1 < \min_i\{d_i\}\Big).$$

Numerical Examples

Our first example is the Cauchy(0,1) density (example 4). Using the Gibbs sampler algorithm 1000 random variates were generated and the results are shown in Figure 1.

Our second example is the binomial GLM with a logit link function and a quadratic logistic model given by

$$y_i|\pi_i \sim \text{binomial}(n_i, \pi_i)$$

and

$$\log\Big(\pi_i/(1-\pi_i)\Big) = \beta_1 + Z_i\beta_2 + Z_i^2\beta_3 = X_i\beta, \ \ i = 1, ..., n.$$

Further details are provided in Dellaportas and Smith (1993). With a multivariate normal prior for β, say $N(\mu, \Sigma)$, the posterior distribution is given by

$$f(\beta) \propto \Big\{\Pi_{i=1}^n e^{y_i X_i \beta}/(1+e^{X_i\beta})^{n_i}\Big\}\exp\Big(-0.5(\beta-\mu)\Sigma^{-1}(\beta-\mu)\Big).$$

Here we introduce the latent variable $U = (U_1, ..., U_n)$ such that the joint density with β is given, up to a constant of proportionality, by

$$f(\beta, u) \propto \Big\{\Pi_{i=1}^n I\Big(u_i < \{1+\exp(X_i\beta)\}^{-n_i}\Big)\Big\}\exp\Big(-0.5(\beta-\mu)\Sigma^{-1}(\beta-\mu)+\nu\beta\Big),$$

where $\nu = \sum_{i=1}^n y_i X_i$. The full conditional distributions for each of the U_i are uniform given by

$$f(u_i|u_{-i}, \beta) = \mathcal{U}\Big(0, \{1+\exp(X_i\beta)\}^{-n_i}\Big).$$

However, the real interest is in sampling from $f(\beta|u)$.

First the condition $u_i < (1+e^{X_i\beta})^{-n_i}$ implies $\exp(X_i\beta) < 1/u_i^{1/n_i} - 1$. Therefore define the sets

$$A_{ku} = \Big\{\beta_k : \beta_k < \min_i\{\log(1/u_i^{1/n_i} - 1)/X_{ki} - X_{li}\beta_l/X_{ki} - X_{mi}\beta_m/X_{ki}\}\Big\},$$

where $\{k, l, m\}$ are, in some order, the elements $\{1, 2, 3\}$. Sampling from $f(\beta|u)$ can now be done by sampling successively from $f(\beta_k|\beta_{-k}, u)$ which involves sampling from a univariate normal distribution restricted to the set A_{ku}. This univariate normal distribution is given by $\pi(\beta_k|\beta_{-k})$ where $\pi(\beta)$ is the multivariate normal distribution with mean $\mu + \Sigma\nu$ and covariance matrix Σ.

We analyse a data set relevant to the above example. The data set and prior distribution used are given in Dellaportas and Smith (1993). We start the chain by taking β as the location of the prior distribution and then proceed to sample U and then back to β. We ran the chain for 20000 iterations (taking only several seconds) and collected the last 2000 for parameter estimation. We can report, as was to be expected, that our parameter estimates ($\hat{\beta}_1 = -2.36$, $\hat{\beta}_2 = 0.21$ and $\hat{\beta}_3 = -0.004$) coincide with those obtained by Dellaportas and Smith. These authors used the adaptive rejection sampling scheme (Gilks and Wild, 1992) which depends on the posterior density being log-concave. (We need no such condition.) Additionally, in Figure 2 we give kernel density estimates of the marginals for β obtained from the output of the Gibbs sampler.

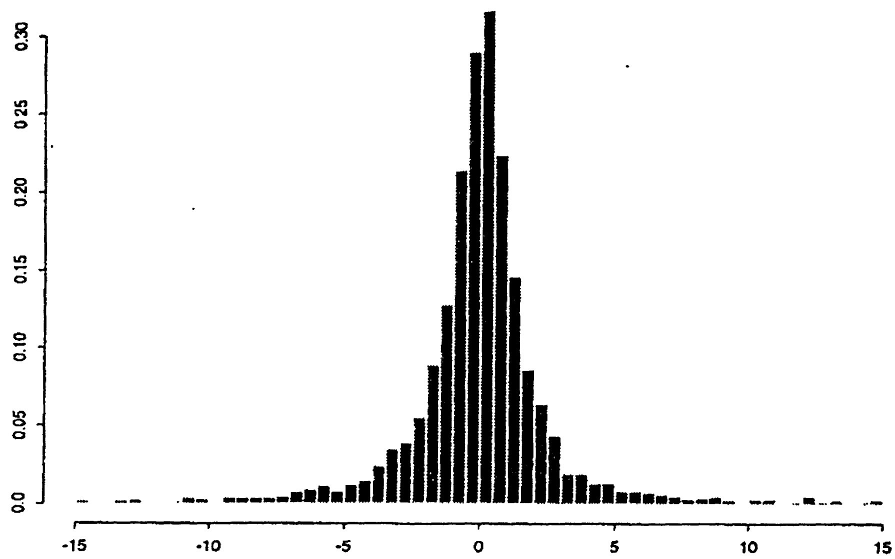

Figure 1: Histogram representation from output obtained using the Gibbs Sampler for the Cauchy(0,1) example.

4 Discussion and Conclusions

With the increasing use of the Gibbs sampler in Bayesian analysis, faster, efficient, and *simpler* methods for generating random variates are required. In this paper, we have proposed and illustrated a method that appears to

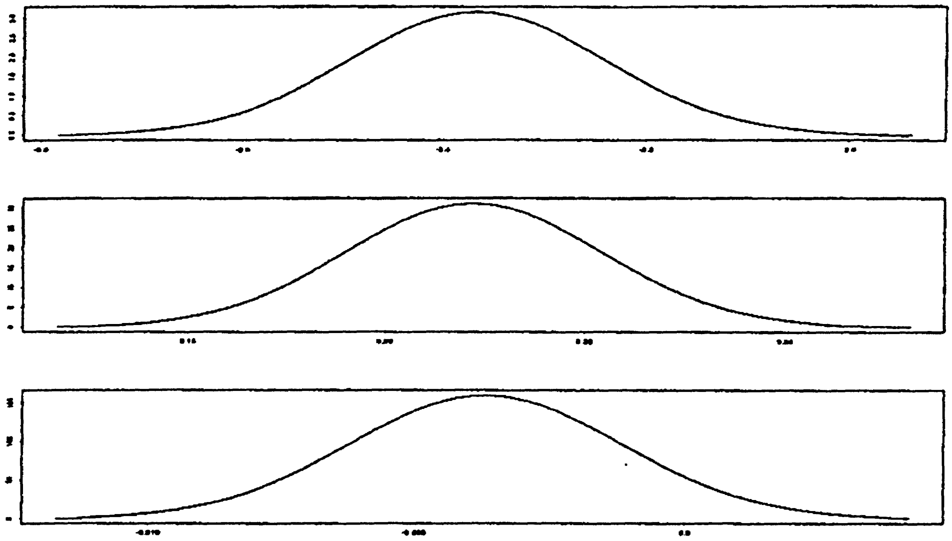

Figure 2: Marginal posterior distibutions for β (β_1 top, β_3 bottom)

be, in some sense, ubiquitous when sampling from univariate continuous densities, from a Bayesian perspective. Here we simply point out that in all the examples we have presented, the use of popular algorithms such as the Metropolis-Hastings, sampling-resampling (Smith and Gelfand, 1992), adaptive-rejection, ratio-of-uniforms, or any rejection method, whether these methods are used within a Gibbs loop or not, is bypassed. (For an excellent discussion of these algorithms, see, for example, Chib and Greenberg, 1995, and Mueller, 1995). The striking feature of our approach is that it obviates the difficulties associated with these alternative approaches, namely identifying dominating densities, calculating supremums, and acceptance rates.

Acknowledgements

The authors are grateful to Adrian Smith for helpful discussions and critical comments on earlier drafts of the paper.

References

Besag,J. and Green,P.J. (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society Series B* **55**, 25-37.

Chib, S. and Greenberg, E. (1995). *The American Statistician.* **49**, 327-335.

Damien, P. (1994). *Some contributions to Bayesian nonparametric inference.* Ph.D. Thesis. Imperial College, London.

Damien, P., Laud, P.W. and Smith, A.F.M. (1995). Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference. *Journal of the Royal Statistical Society Series B* **57**, 547-564.

Damien, P., Laud, P.W. and Smith, A.F.M. (1996). Implementation of Bayesian nonparametric inference based on beta processes. *Scandinavian Journal of Statistics.* **23**, 27-36.

Dellaportas,P. and Smith,A.F.M. (1993). Bayesian inference for generalised linear and proportional hazards models via Gibbs sampling. *Applied Statistics* **42**, 443-459.

Diaconis,P. and Kemperman,J. (1996). Some new tools for Dirichlet priors. *Bayesian Statistics 5.* pp.97-106. Eds, J.M.Bernado,J.O.Berger,A.P.Dawid and A.F.M.Smith. Oxford University Press.

Doksum,K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability* **2**, 183-201.

Gilks,W.R. and Wild,P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.

Laud,P.W. (1977). Bayesian nonparametric inference in reliability. Ph.D. Dissertation, University of Missouri, Columbia, MO.

Laud, P.W., Damien, P. and Smith, A.F.M. (1993). Random variate generation from D-distributions. *Statistics and Computing* **3**, 109-112.

Laud, P.W., Smith, A.F.M., and Damien, P. (1996). Monte Carlo methods approximating a posterior hazard rate process. *Statistics and Computing* **3**, 109-112.

Mueller, P. (1995). A generic approach to posterior integration and Gibbs sampling. *Manuscript.*

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian statistics without tears: a sampling resampling perspective. *The American Statistician* **46**, 84-88.

Smith,A.F.M. and Roberts,G.O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B* **55**, 3-23.

Tierney,L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**, 1701-1762.

Walker, S.G. (1995) Generating random variates from D-distributions via substitution sampling. *Statistics and Computing* **5**, 311-315.

Walker, S.G. (1996) Random variate generation from an infinitely divisible distribution via Gibbs sampling. *Technical Report*, Imperial College.

Walker, S.G. and Damien, P. (1996) A full Bayesian nonparametric analysis involving a neutral to the right process. *Technical Report*, Imperial College.