SAMPLING TRUNCATED POISSON AND MULTIVARIATE NORMAL DENSITIES VIA THE GIBBS SAMPLER

WORKING PAPER #9612-40

CRAIG CUMBUS
KIMBERLY-CLARK CORPORATION
AND
PAUL DAMIEN
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL
AND
STEPHEN WALKER
IMPERIAL COLLEGE, LONDON, UK

Sampling Truncated Poisson and Multivariate Normal Densities via the Gibbs Sampler

Craig Cumbus

Kimberly-Clark Corporation. Atlanta. GA. USA

Paul Damien

School of Business. University of Michigan, Ann Arbor, MI, USA

Stephen Walker

Department of Mathematics, Imperial College, London. UK.

ABSTRACT

Gelfand et al. (1992) consider the Bayesian analysis of constrained parameter and truncated data problems within a Gibbs sampling framework. In this paper we consider the truncated data problem and concentrate on sampling truncated Poisson and multivariate normal densities within a Gibbs sampler. The proposed method bypasses the need for rejection sampling and/or algorithms such as the Metropolis-Hastings and sampling-resampling. In the truncated multivariate normal case we introduce a Gibbs sampler in which all the conditional distributions can be sampled via uniform variates. whereas the method of Robert (1995), which also runs on a Gibbs sampler, has conditional distributions requiring rejection algorithms to obtain random variates.

Key Words: Latent variables, Gibbs sampler, uniform random variables.

I. Introduction

Let f be a continuous density function defined on the real line. The solution 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 of such variables, was developed in Damien and Walker (1996 a). To motivate their central result stated as a Theorem later, let $f(x) \propto \exp(-x\beta)I(a < x < b)$, where $\beta > 0$, $-\infty \leq a < b \leq \infty$ and I represents the indicator function; write such a density as $\mathcal{E}(\beta, a, b)$. Let $f(x) \propto (1-x)^{\beta-1}$, where $\beta > 0$ and $0 \leq a < b \leq 1$; write such a density as $\mathcal{B}(\beta, a, b)$. Finally, let $U(a, b)$ denote the uniform distribution on (a, b). It is obvious that both $\mathcal{E}(\beta, a, b)$ and $\mathcal{B}(\beta, a, b)$ can be sampled via the inverse transform method. Now, 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 (Smith and Roberts, 1993) to generate random variates from f. In particular, Damien and Walker (1996 a) show that all the conditional distributions in such a Gibbs sampler will be one of the three distributions, $U(a, b)$, $\mathcal{E}(\beta, a, b)$ or $\mathcal{B}(\beta, a, b)$. To this end they prove:

Theorem (Damien and Walker 1996 a).

If

$$f(x) \propto \prod_{l=1}^{L} g_l(x),$$

where the g_l are nonnegative invertible functions (not necessarily) densities; that is, if $g_l(x) > y$ then it is possible to obtain the set $A_l(y) = \{x : g_l(x) > y\}$, then it is possible to implement a Gibbs sampler for generating random variates from f in which all the full conditionals are uniform distributions.

Damien and Walker (1996 a) exemplify the use of this Theorem by sampling all univariate continuous densities that appear in Johnson and Kotz. Also they illustrate the method for several Bayesian nonconjugate models and nonparametric models. Damien and Walker (1996 b), Walker and Damien (1996 a b) apply the above Theorem to provide a full Bayesian analysis of circular data, neutral to the right processes, and mixtures of Dirichlet processes, respectively.

Since an indicator function satisfies the conditions of a g_l it is no more difficult to sample a truncated or constrained random variate using the approach of Damien and Walker than in the nontruncated or unconstrained case. Therefore. if the sampling from f. or a truncated version of f. is required within the context of a Gibbs sampler the Theorem provides a means of doing this via the introduction of latent variables which has the effect of extending the Gibbs loop (i.e. full conditionals) by the number of latent variables introduced.

In this paper, the Theorem is used to generate random variates from truncated Poisson (Section II) and multivariate normal densities (Section III). A brief discussion in Section IV concludes the paper.

II. Truncated Poisson density

We consider the random variate generation of a left truncated Poisson variable required within the context of a Gibbs sampler which is 'running' over a sequence of full conditionals. one of which turns out to be a truncated Poisson, for example, in a truncated data problem involving a Poisson model (Gelfand et al.. 1992). Devroye (pg. 489) identifies evaluating factorials as a problem with the sampling of a Poisson variate. Our approach by-passes the need to evaluate factorials.

To introduce our approach. let X be a Poisson random variable with $f_x \propto \lambda^x / x! \cdot I(x \in \Omega)$ where $f_x = P(X = x)$. $\lambda > 0$. and $\Omega = \{0. 1. 2, ...\}$. Following the Theorem we introduce two latent variables Y and Z such that the joint density with X is given by

$$f_{X,Y,Z}(x,y,z) \propto I\left(y \leq \lambda^x, z \leq x!^{-1}\right).$$

The resulting full conditionals are given by

$$Y|(X = x) \qquad \sim U\left(0, \lambda^x\right)$$

$$Z|(X = x) \qquad \sim U\left(0, x!^{-1}\right)$$

$$X|(Y = y, Z = z) \quad \sim U\left(\{n(y), ..., m(y, z)\}\right).$$

where, if $\lambda > 1$ then $n(y)$ is the smallest integer greater than $\log y / \log \lambda$ if $y > 1$, and $n(y) = 0$ if $y \leq 1$, else if $\lambda \leq 1$ then $n(y) = 0$. If $\lambda \geq 1$

3

then $m(y,z)$ is the largest integer $m(z)$ such that $m(z)! \leq 1/z$, else if $\lambda < 1$ then $m(y,z) = \min\{m(z), m(y)\}$, where $m(y)$ is the largest integer less than $|\log y / \log \lambda|$. Note we can take $z = \tau / x!$ where $\tau \sim U(0,1)$ leading to

$$m(z) = \mathrm{int}\left\{\Gamma^{-1}\left(\tau^{-1}\Gamma(x+1)\right) - 1\right\}.$$

where Γ denotes the gamma function. Perhaps a less computational approach to obtaining $m(z)$ is that $m(z)$ is the largest integer for which

$$\sum_{k=x+1}^{m(z)} \log k \leq -\log \tau.$$

and so there is no need for us to evaluate a factorial.

Suppose then we need to sample from $f_x \propto \lambda^x / x! \cdot I(x \in D)$ where $D = \{a, a+1, ...\}$. The algorithm is now given by the Gibbs sampler with full conditionals

$$Y|(X = x) \qquad \sim U(0, \lambda^x)$$

$$Z|(X = x) \qquad \sim U(0, x!^{-1})$$

$$X|(Y = y, Z = z) \quad \sim U\left(\{\max\{a, n(y)\}, ..., m(y, z)\}\right).$$

and so is essentially no more difficult than the non-truncated case.

III. Truncated multivariate normal density

In this Section we describe a method for sampling truncated multivariate normal variables using a Gibbs sampler. To introduce our ideas we consider the standard univariate normal distribution. Let $X \sim N_1(0, 1)$, 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:

$$Y|(X = x) \sim U(0, \exp(-x^2/2))$$

$$X|(Y = y) \sim U\left(-\sqrt{-2\log y}, \sqrt{-2\log y}\right).$$

4

Since we are only going to be interested in the X samples we can define $Y = \tau \exp(-X^2/2)$ where $\tau \sim U(0,1)$ and therefore we can construct the sequence $\{X_n\}$ via

$$X_{n+1}|(X_n = x, \tau_n) \sim U\left(-\sqrt{x^2 - 2\log \tau_n}, \sqrt{x^2 - 2\log \tau_n}\right).$$

where the τ_n are iid $U(0,1)$ and also independent of the X_n. Obviously, as $n \to \infty$ then $X_n \to_d N_1(0,1)$.

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

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

Our required algorithm is now easily seen to be given by the sequence

$$X_{n+1}|(X_n = x, \tau_n) \sim U\left(\max\{a, -\sqrt{x^2 - 2\log \tau_n}\}, \min\{b, \sqrt{x^2 - 2\log \tau_n}\}\right).$$

Robert (1995) proposes a rejection based algorithm for sampling a truncated univariate normal distribution. We anticipate that the vast majority of cases involving sampling truncated normal distributions would arise in the context of a Gibbs sampler, so our algorithm is effectively extending each Gibbs loop by only one more full conditional, which happens to be a uniform distribution.

Robert (1995) then proposes a Gibbs sampler for sampling from a truncated multivariate normal distribution. We can greatly simplify his algorithm using our latent variable idea. Therefore, consider

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

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

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

for $i = 1, \dots, p$, are the full conditionals, and $\nu_i = \mu_i - \sum_{j \neq i}(x_j - \mu_i)e_{ij}/e_{ii}$ and $\sigma_i^2 = 1/e_{ii}$, where e_{ij} is the ijth element of Σ^{-1}. Robert uses his rejection algorithm for sampling these truncated univariate normal densities. However, since we are already in a Gibbs sampler it seems appropriate to implement

5

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 clearly a truncated exponential distribution which can be sampled using the cdf inversion technique.

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

Example Truncated $N_2(0, \Sigma)$.

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

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

so

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

and A is the ball 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

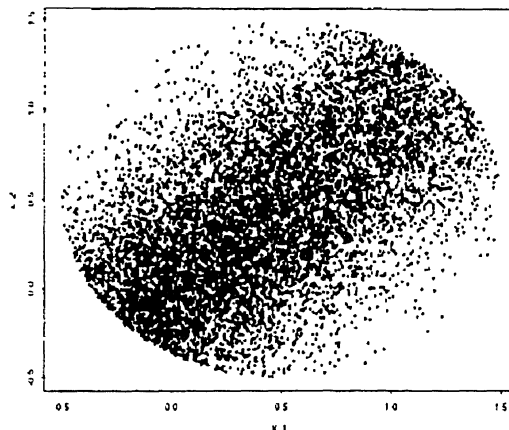$$B_i = \left(x_j\rho - \sqrt{(y - x_j^2)(1 - \rho^2)}, x_j\rho + \sqrt{(y - x_j^2)(1 - \rho^2)}\right).$$

Figure 1: Plot of 10.000 samples from truncated $N_2(0, \Sigma)$ distribution obtained from Gibbs sampler

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 and a scatter plot of the samples appears in Figure 1.

IV. Extensions and conclusions

In this paper we have proposed and illustrated a method for generating random variates from truncated Poisson and multivariate normal densities. The main advantage of the method is that it obviates the difficulties associated with alternative methods such as rejection algorithms (Devroye. 1986). Metropolis-Hastings (Tierney. 1994; Chib and Greenberg, 1996) and sampling-resampling (Smith and Gelfand, 1992). These algorithms require identifying dominating densities, calculating supremums, and acceptance rates: these may be difficult to obtain in many contexts. Required within the context of a Gibbs sampler our algorithms are particularly appropriate since the method leads to the addition of one or two uniform full conditionals

7

into the original Gibbs loop.

We restricted ourselves to the Poisson density since simulating from other discrete densities. using our method. is straightforward. Likewise. it is no more difficult to simulate other multivariate densities such as the Student-t. using our approach. As examples, we give here the full conditionals to sample the geometric and negative binomial distributions.

1. Geometric. Here we have $f_x \propto (1-p)^x$ where $p \in (0.1)$ and $x \in \{0.1....\}$. The joint density of interest is given by

$$f_{X,Y}(x.y) \propto I\left(y \le (1-p)^x\right).$$

leading to the full conditionals

$$f_{Y|X}(y|x) = U\left(0,(1-p)^x\right).$$

and

$$f_{X|Y}(x|y) = U\left(\{0,1,...,n(y)\}\right),$$

where $n(y) = \text{int}[\log y/\log(1-p)]$.

2. Negative binomial. Here we have $f_x \propto (x-1)!(1-p)^x/(x-k)!$ for $p \in (0.1)$ and $x \in \{k, k+1....\}$. The joint density of interest is given by

$$f_{X,Y,Z,W}(x.y.z.w) \propto I\left(y \le (x-1)!, z(x-k)! \le 1, w \le (1-p)^x\right),$$

leading to the full conditonals

$$f_{W|X}(w|x) = U\left(0,(1-p)^x\right),$$

$$f_{Y|X}(y|x) = U\left(0,(x-1)!\right),$$

$$f_{Z|X}(z|x) = U\left(0,(x-k)!^{-1}\right),$$

and

$$f_{X|W,Y,Z}(x|w,y,z) = U\left(\{n(y),...,\min\{n(w),n(z)\}\}\right),$$

where $n(y)$ is the smallest integer such that $(n(y)!-1) \ge y$, $n(z)$ is the largest integer $\ge k$ such that $(n(z)-k)! \le 1/z$, and $n(w) = \text{int}[\log w/\log(1-p)]$. As with the Poisson case we can circumvent the need to evaluate factorials.

8

References

Chib. S. and Greenberg. E. (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician.* 49, 327-335.

Damien P. and Walker S.G. (1996 a) Sampling probability densities via uniform random variables and a Gibbs sampler. *Submitted for publication.*

Damien P.and Walker S.G. (1996 b) A full Bayesian analysis of circular data using von Mises distribution. *Submitted for publication.*

Devroye. L. (1986) *Non-uniform random variate generation.* Springer. New York.

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.

Robert. C.P. (1995). Simulation of truncated normal variables. *Statistics and Computing* 5, 121-125.

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 computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B.* 55. 3-24.

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

Walker. S.G. and Damien. P. (1996 a) A full Bayesian nonparametric analysis involving a neutral to the right process. *Submitted for publication.*

Walker. S.G. and Damien. P. (1996 b) Sampling a Dirichlet process mixture model. *Submitted for publication.*

10