GIBBS SAMPLING APPROACH FOR GENERATION OF TRUNCATED MULTIVARIATE GAUSSIAN RANDOM VARIABLES

Jayesh H. Kotecha and Petar M. Djurić

Department of Electrical and Computer Engineering
State University of New York at Stony Brook, Stony Brook, NY 11790
jkotecha@ee.sunysb.edu, djuric@ee.sunysb.edu

ABSTRACT

In many Monte Carlo simulations, it is important to generate samples from given densities. Recently, researchers in statistical signal processing and related disciplines have shown increased interest for a generator of random vectors with truncated multivariate normal probability density functions (pdf's). A straightforward method for their generation is to draw samples from the multivariate normal density and reject the ones that are outside the acceptance region. This method, which is known as rejection sampling, can be very inefficient, especially for high dimensions and/or relatively small supports of the random vectors. In this paper we propose an approach for generation of vectors with truncated Gaussian densities based on Gibbs sampling, which is simple to use and does not reject any of the generated vectors.

1. INTRODUCTION

In signal processing the generation of samples from various distributions is often needed [5]. Some existing methods include rejection sampling [3] and various Markov chain Monte Carlo (MCMC) methods [2], [4]. In Gaussian noise models, we often come across situations where samples from truncated Gaussian densities are required. For example, in autoregressive (AR) parameter estimation [6], [8], if Gaussian noise and uniform priors (over the stability region) for the AR parameters are assumed, the posterior is approximately a truncated Gaussian. In such cases if rejection sampling is used, whereby samples not in the acceptable region are rejected, it is very inefficient and the most time consuming part of the simulation.

We propose here a Gibbs sampling approach, which is a special MCMC sampling method, for drawing samples from multivariate truncated Gaussian densities. For implementation of the Gibbs sampler, we need a transformation that maps a univariate Gaussian random variable to a truncated Gaussian one.

In Section 2 we discuss such transformation, which we then exploit in the multivariate case, described in Section 3. There we propose a Gibbs sampling scheme for generation of n-dimensional truncated multivariate Gaussian variables whose support is an n-dimensional cuboid. The support, however, can be any convex region. In Section 4 we show simulation results that illustrate our approach. We also perform nonparametric tests on the drawn samples to further substantiate that they are indeed distributed according to the desired truncated Gaussian densities.

2. SAMPLING A TRUNCATED GAUSSIAN: UNIVARIATE CASE

Suppose we want to generate samples from a univariate truncated Gaussian random variable $\tilde{y}$ with a support set $\mathcal{R}_s \equiv (a,b]$.[1] That is, we need a sample from

$$f_{\tilde{y}}(y) = \begin{cases} c^{-1}(2\pi\sigma^2)^{-1/2}\exp(-(y-\mu)^2/2\sigma^2), & a < y \leq b \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where

$$c = \int_a^b (2\pi\sigma^2)^{-1/2}\exp(-(x-\mu)^2/2\sigma^2)dx. \tag{2}$$

We denote the above density as $TN(\mu, \sigma^2, a, b)$.

A simple method to do this is rejection sampling. We generate samples x from the required Gaussian density and discard the ones whose generated values are outside $\mathcal{R}_s$. But in cases where the probability mass above the support region is small, the rejection rate becomes unacceptably high, leading to a very inefficient sampling scheme.

Here we present a procedure which does not discard the samples x from the Gaussian density. It is based on the idea of mapping x to y, where $y \in \mathcal{R}_s$, in such a way that the mapped samples have the required density $f_{\tilde{y}}(y)$. Denote the mapping by $y = g(x)$, where $\tilde{x}$ is distributed as $N(\mu, \sigma^2)$. First we find the transformation $g(\cdot)$.

If $F_{\tilde{x}}(x)$ and $F_{\tilde{y}}(y)$ denote the cumulative distribution functions (cdf's) of $\tilde{x}$ and $\tilde{y}$ respectively, then

$$F_{\tilde{y}}(y) = \begin{cases} 0, & y \leq a \\ c^{-1}(F_{\tilde{x}}(y) - F_{\tilde{x}}(a)), & a < y \leq b \\ 1, & y > b \end{cases} \tag{3}$$

and $F_{\tilde{x}}(x) = \Phi((x-\mu)/\sigma)$, where $\Phi(\cdot)$ is the cdf of the standard Gaussian random variable. If $g(x)$ is one-to-one, we can write

$$F_{\tilde{y}}(y) = F_{\tilde{x}}(x). \tag{4}$$

This work was supported by the National Science Foundation under Award No. MIP-9506743.

[1] $\tilde{y}$ represents a random variable, and y a value of $\tilde{y}$.

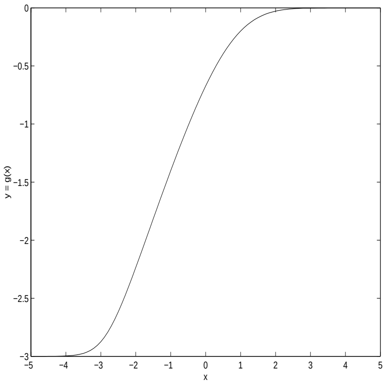

Figure 1: Plot of the univariate transformation equation for window (-3,0) for a standard normal random variable.

From (3) and (4) it follows that

$$c^{-1}(F_{\tilde{x}}(y) - F_{\tilde{x}}(a)) = F_{\tilde{x}}(x). \tag{5}$$

For the range $a < y \leq b$, we have

$$\begin{aligned} y &= F_{\tilde{x}}^{-1}(cF_{\tilde{x}}(x) + F_{\tilde{x}}(a)) \\ &= F_{\tilde{x}}^{-1}\left(F_{\tilde{x}}(x)(F_{\tilde{x}}(b) - F_{\tilde{x}}(a)) + F_{\tilde{x}}(a)\right). \end{aligned} \tag{6}$$

From (4) we deduce that $F_{\tilde{y}}(y) = 0$ when $x = -\infty$, and $F_{\tilde{y}}(y) = 1$ when $x = \infty$. If $x = -\infty$, we obtain from (6) $y = a$, and if $x = \infty$, $y = b$. Since (3) is the desired cdf of $\tilde{y}$, we conclude that (6) is the transformation $g(x)$ which produces $\tilde{y}$ from the Gaussian random variable $\tilde{x}$. It is important to notice that $g(x)$ is a composition of two strictly monotonically increasing functions, $g(x) = F_{\tilde{x}}^{-1}(F_{\tilde{x}}(x)(F_{\tilde{x}}(b) - F_{\tilde{x}}(a)) + F_{\tilde{x}}(a))$, and therefore is also strictly monotonically increasing function and hence one-to-one.

In summary, to get a sample from $TN(\mu, \sigma^2, a, b)$, we simply draw x from $N(\mu, \sigma^2)$ and map it to y by $g(x)$. Figure 1 shows an example plot of (6) for $a = -3$ and $b = 0$, where $\tilde{x} \sim N(0, 1)$.

3. MULTIVARIATE CASE – GIBBS SAMPLING SCHEME:

Gibbs sampling is a Markov chain Monte Carlo (MCMC) method for generation of samples from high dimensional densities by drawing the samples from full conditional densities [4]. In other words, to implement the Gibbs sampling for an n-dimensional vector $\tilde{\mathbf{y}}$, we need to find full conditionals of its subvectors which are easy to sample from.

Let $\tilde{\mathbf{x}} \sim N(\mu, \mathbf{C})$ and $\tilde{\mathbf{y}}^T = [\tilde{y}_1 \tilde{y}_2 \dots \tilde{y}_n] \sim TN(\mu, \mathbf{C}, \mathcal{R}_s)$,[2] where $\mathcal{R}_s = \prod_{i=1}^{n} \mathcal{R}_i$ is the support of $\tilde{\mathbf{y}}$ with $\mathcal{R}_i = (a_i, b_i]$. Denote by $\mathbf{y}^T = [y_1 y_2 \dots y_n]$ a sample of $\tilde{\mathbf{y}}$.

Note that,

$$f_{\tilde{\mathbf{y}}}(\mathbf{y}) = \frac{f_{\tilde{\mathbf{x}}}(\mathbf{y})}{\int_{\mathcal{R}_s} f_{\tilde{\mathbf{x}}}(\mathbf{x}) d\mathbf{x}}, \quad \mathbf{y} \epsilon \mathcal{R}_s. \tag{7}$$

We shall prove that the full conditionals $f_{\tilde{y}_i}(y_i | \mathbf{y}_{-i})$, $i = 1, 2, \dots, n$, where $\mathbf{y}_{-i}^T = [y_1 \;\; y_2 \;\; \dots \;\; y_{i-1} \;\; y_{i+1} \;\; \dots \;\; y_n]$, are themselves univariate truncated Gaussian densities. We have

$$\begin{aligned} f_{\tilde{y}_i}(y_i | \mathbf{y}_{-i}) &= \frac{f_{\tilde{\mathbf{y}}}(\mathbf{y})}{f_{\tilde{\mathbf{y}}_{-i}}(\mathbf{y}_{-i})} \\ &= \frac{f_{\tilde{\mathbf{y}}}(\mathbf{y})}{\int_{a_i}^{b_i} f_{\tilde{\mathbf{y}}}(\mathbf{y}) dy_i}, \quad \mathbf{y} \in \mathcal{R}_s \end{aligned} \tag{8}$$

and using (7), we can write

$$f_{\tilde{y}_i}(y_i | \mathbf{y}_{-i}) = \frac{f_{\tilde{\mathbf{x}}}(\mathbf{y})}{\int_{a_i}^{b_i} f_{\tilde{\mathbf{x}}}(\mathbf{y}) dy_i}, \quad \mathbf{y} \in \mathcal{R}_s. \tag{9}$$

Finally, from (9) the conditional density $f_{\tilde{y}_i}(y_i | \mathbf{y}_{-i})$ may be rewritten as

$$f_{\tilde{y}_i}(y_i | \mathbf{y}_{-i}) = \frac{f_{\tilde{\mathbf{x}}}(y_i | \mathbf{y}_{-i})}{\int_{a_i}^{b_i} f_{\tilde{\mathbf{x}}}(y_i | \mathbf{y}_{-i}) dy_i}, \quad \mathbf{y} \in \mathcal{R}_s \tag{10}$$

which is nothing but a truncated Gaussian pdf whose samples can be generated as shown in the previous section.

Now we present the Gibbs sampling scheme used to produce samples from $f_{\tilde{\mathbf{y}}}(\mathbf{y})$. First, we get initial values $\{y_i^{(0)}\}_{i=2}^{n}$ from the support set $\mathcal{R}_s$ and then for $j = 1, \dots, M$ generate samples according to

$$\begin{array}{lcl} f(y_1 \mid y_2^{(j-1)}, y_3^{(j-1)}, \dots, y_n^{(j-1)}) & \longrightarrow & y_1^{(j)} \\ f(y_2 \mid y_1^{(j)}, y_3^{(j-1)}, \dots, y_n^{(j-1)}) & \longrightarrow & y_2^{(j)} \\ \vdots & & \\ f(y_{n-1} \mid y_1^{(j)}, \dots, y_{n-2}^{(j)}, y_n^{(j-1)}) & \longrightarrow & y_{n-1}^{(j)} \\ f(y_n \mid y_1^{(j)}, y_2^{(j)}, \dots, y_{n-1}^{(j)}) & \longrightarrow & y_n^{(j)} \end{array} \tag{11}$$

where the arrows represent sampling from the corresponding density. Note that each y_i is sampled according to the univariate scheme in Section 2.

A good set of initial values for the Markov chain is a sample that corresponds to the maximum of $f_{\tilde{\mathbf{y}}}(\mathbf{y})$, where $\mathbf{y} \in \mathcal{R}_s$. It is easy to find this starting point because we know the shape of the distribution. With this starting value there is no need for burn-in period (in general required until convergence to the equilibrium distribution is achieved) except for the first few samples, after which the subsequent draws become uncorrelated with the initial point.

4. SIMULATION RESULTS

We performed a goodness-of-fit test on the samples obtained in the univariate case [1]. Assume we have n samples of a univariate random variable $\tilde{y}$ with a cdf $F_{\tilde{y}}(y)$. We need to test the hypothesis

$$H : F_{\tilde{y}}(y) = F_T(y) \tag{12}$$

where $F_T(y)$ is the cdf of a univariate truncated Gaussian random variable. The alternative hypothesis is K, where

$$K : F_{\tilde{y}}(y) \neq F_T(y). \tag{13}$$

---

[2] Note that μ and $\mathbf{C}$ are not the mean and the covariance matrix of $\tilde{\mathbf{y}}$.

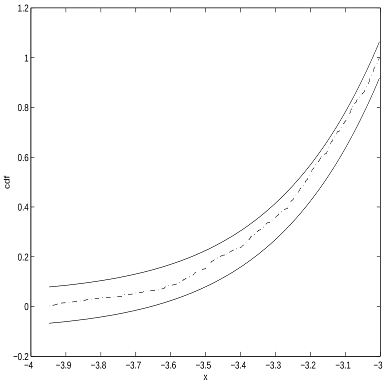

Figure 2: Plot of $F_T(y) \pm \epsilon$ (— curve) and $F_n(y)$ (-.- curve) for $TN(0, 1, -4, -3)$, and $\alpha = 0.01$.

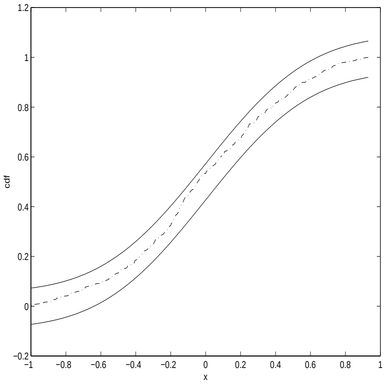

Figure 3: Plot of $F_T(y_5|\mathbf{y}_{-5}) \pm \epsilon$ (— curve) and $F_n(y_5|\mathbf{y}_{-5})$ (-.- curve) for the $TN(\mathbf{0}, \mathbf{C}, \prod_{i=1}^{10}(-1, 1])$, where $c_{ii} = 1$, $c_{ij} = 0.8$, $i \neq j$, and $\alpha = 0.01$.

The Kolmogorov statistic for testing the hypothesis H is given by [1],

$$D_n = \sup_y |F_n(y) - F_T(y)| \tag{14}$$

where $F_n(y) = [\text{number of } y_i \leq y]/n$ and n is the total number of samples. We performed the Kolmogorov test for $n = 500$ samples for different acceptance regions as shown in the Table 1, and significance level $\alpha = 0.01$. We conclude that our hypothesis is not rejected in any of the cases. Figure 2 shows a plot of $F_T(y) \pm \epsilon$ (where ϵ corresponds to $\alpha = 0.01$) and $F_n(y)$ for $TN(0, 1, -4, -3)$.

We repeated the test for the density $f_{\tilde{y}_5}(y_5|\mathbf{y}_{-5})$, where $\mathbf{y}_{-5}^T = [y_1 y_2 \dots y_4 y_6 \dots y_{10}]$. The truncated Gaussian was defined by $TN(\mathbf{0}, \mathbf{C}, \mathcal{R}_s)$ where the diagonal elements of $\mathbf{C}$, $c_{ii} = 1$, and all the off-diagonal elements $c_{ij} = 0.8$. The support region was $\mathcal{R}_s = \prod_{i=1}^{10}(-1, 1]$, and the significance level was $\alpha = 0.01$. Thus, we wanted to verify if indeed the conditional densities are truncated Gaussians as claimed in the previous section. A result of the test is shown in Figure 3, and again it shows agreement with our claim.

To compare the efficiencies of the rejection and Gibbs sampling schemes, in Figures (4) and (5) we show scat-

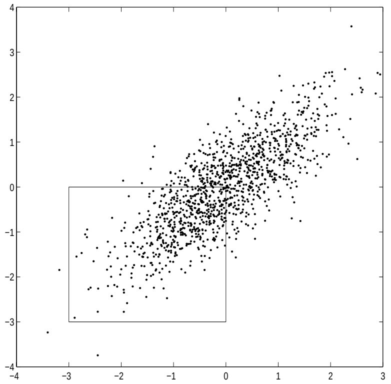

Figure 4: Rejection sampling for generating samples of truncated Gaussian random variables. The acceptance region is indicated by the square.

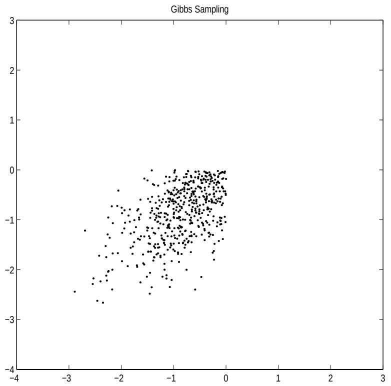

Figure 5: Gibbs sampling for the same random variable as in Figure 4.

No.	Acceptance Region	D_n	Acceptable Value
1.	(-3,0)	0.0295	0.0729
2.	(-3,-2)	0.0421	0.0729
3.	(-4,-3)	0.0422	0.0729

Table 1: Kolmogorov test for different acceptance regions for the univariate case $TN(0, 1, -4, -3)$ and significance level $\alpha = 0.01$.

Total Samples	Acceptance region			
	(-1,1;...;-1,1)		(-4,-3;...;-4,-3)	
	RS	GS	RS	GS
30000	7993	30000	1	30000
100000	26583	100000	2	100000

Table 2: Number of generated samples of a 10-dimensional $TN(\mathbf{0}, \mathbf{C}, \prod_{i=1}^{10}(-1, 1])$, where $c_{ii} = 1$, $c_{ij} = 0.8$ ($i \neq j$).

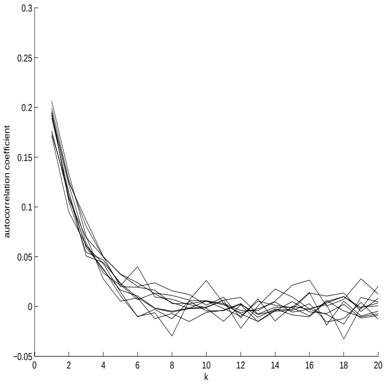

Figure 6: Plot of the autocorrelations $\hat{r}_{y_i y_i}(k)$, $i = 1, 2, \ldots, 10$, as functions of the lag k.

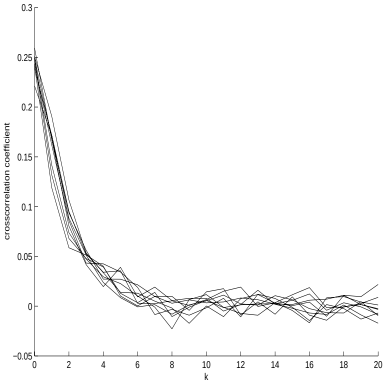

Figure 7: Plot of the crosscorrelations $\hat{r}_{y_i y_1}(k)$, $i = 2, \ldots, 10$, as functions of the lag k.

ter plots of two dimensional truncated Gaussians generated by rejection sampling and our method, where $\mathcal{R}_s = \prod_{i=1}^{2}(-3, 0]$. Figure (4) shows visually the amount of rejection, where the square window is the acceptance region.

In continuing the comparison, with the two schemes, we generated samples from a 10-dimensional truncated Gaussian. The results shown in Table 2 are for two different acceptance regions. The column "Total Samples" indicates the total number of samplings done, and the columns RS and GS indicate the number of samples in the support set of the truncated Gaussian density. From the table we see that as the acceptance region slides into the low probability region of the multivariate Gaussian random variable, the rejections of the rejection sampling scheme increase, resulting in lesser and lesser efficiency. On the other hand, the Gibbs sampling scheme keeps all the generated samples. The discrepancy in efficiency of the two schemes increases with the dimension of the generated random variables.

In many Monte Carlo simulations, we would like to use samples which are independent of each other. The rejection sampling scheme generates independent samples, but most MCMC schemes like ours, generates samples which are correlated. To examine the correlation, we generated samples from the 10-dimensional density defined above, and calculated the estimates of the autocorrelation and crosscorrelation coefficients of the individual random variables. Some results are shown in Figures 6 and 7. In Figure 6 we see the plot of the autocorrelations $\hat{r}_{y_i y_i}(k), i = 1, \ldots, 10$ and in Figure 7 the crosscorrelations $\hat{r}_{y_i y_1}(k), i = 2, \ldots, 10$, as a function of the lag k. We see that the correlations decrease rapidly and become negligible for $k \geq 5$.

5. CONCLUSION

A procedure for generating samples from multivariate truncated Gaussian random variables using Gibbs sampling has been proposed. We see from the results that our scheme is far more efficient than the rejection sampling approach, especially when the dimension of the multivariate density is high and/or the acceptance region lies in a low probability region. This is because a large number of rejections occur in rejection sampling scheme, whereas in the Gibbs sampling scheme all the generated samples are in the acceptance region. The scheme is easy to implement and can be generalized to other types of convex regions.

6. REFERENCES

[1] P. J. Bickel and K. A. Doksum, *Mathematical Statistics*. Oakland, CA: Holden Day, Inc., 1977.

[2] G. Casella, E. I. George, "Explaining the Gibbs sampler", *The American Statistician,* vol. 46, pp. 167-174, 1992.

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

[4] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice.* New York: Chapman & Hall, 1996.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing - Detection Theory.* Upper Saddle River, NJ: Prentice Hall PTR, 1998.

[6] J. J. Rajan, P. J. W. Rayner, S. J. Godsill, "A Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler," *IEE Proc. Vision, Image & Signal Processing*, vol. 144, no. 4, pp. 249-256, 1997.

[7] B. D. Ripley, *Stochastic Simulation.* New York: John Wiley & Sons, 1987.

[8] P. T. Troughton, S. J. Godsill, "Bayesian model selection for time series using Markov chain Monte Carlo," *IEEE Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3233-3736, 1997.