

Efficient Sampling Methods for Truncated Multivariate Normal and Student-t Distributions Subject to Linear Inequality Constraints

Yifang Li ^{*} and Sujit K. Ghosh [†]

Department of Statistics, North Carolina State University
NC State Department of Statistics Technical Report #2649

Abstract

Sampling from a truncated multivariate distribution subject to multiple linear inequality constraints is a recurring problem in many areas in statistics and econometrics, such as the order restricted regressions, censored data models, and shape-restricted nonparametric regressions. However, the sampling problem remains non-trivial due to the analytically intractable normalizing constant of the truncated multivariate distribution. We first develop an efficient rejection sampling method for the truncated univariate normal distribution, and analytically establish its superiority in terms of acceptance rates compared to some of the popular existing methods. We then extend our methodology to obtain samples from a truncated multivariate normal distribution subject to convex polytope restriction regions. Finally, we generalize the sampling method to truncated multivariate Student-t distributions. Empirical results are presented to illustrate the superior performance of our proposed Gibbs sampler in terms of various criteria (e.g., accuracy, mixing and convergence rate).

Key words: truncated multivariate distribution, rejection sampling, Gibbs sampler.

^{*}email: yli40@ncsu.edu

[†]email: sujit.ghosh@ncsu.edu

1 Introduction

The necessity of sampling from a truncated multivariate normal (TMVN) distribution subject to multiple linear inequality constraints arises in many research areas in statistics and econometrics. Robert (1995) gives several examples in the context of order restricted (or isotonic) regressions and censored data models. Several other applications are discussed by Gelfand et al. (1992), Liechty and Lu (2010) and Yu and Tian (2011), including the truncated multivariate probit models in market research, and modeling co-skewness of the skew-normal distributions, shape-restricted nonparametric regressions, and so on. The Bayesian normal linear regression model subject to linear inequality restrictions is also a common application of the TMVN distributions, which is investigated in great details by Geweke (1996) and Rodrigues-Yam et al. (2004).

Despite the wide range of applications, an efficient method to generate samples from a TMVN distribution is not straightforward. One of the main barriers arises from the complex normalizing constant involved in the probability density function (pdf) of TMVN. Although a naive rejection sampling method can be used, it is hard to identify a well working envelope function for the target distribution, and thus very low acceptance rates can occur for some constrained regions. Methods for sampling from TMVN distributions also need to cope with the fact that its marginal distributions are not TMVN and have very complicated pdf's. Much research is done in seeking efficient sampling methods avoiding the evaluation of the normalizing constant.

The Gibbs sampler (Geman and Geman (1984) and Gelfand and Smith (1990)) has been a popular technique for sampling from TMVN distributions as all full conditional distributions of a TMVN are truncated univariate normal (TUVN) distributions. An early survey is provided by Hajivassiliou and Mcfadden (1990). Thus an efficient sampling method for TUVN is key in developing a sampling method for TMVN. Methods such as the classical inversion technique (Gelfand et al. (1992)) and naive rejection sampling using normal or uniform envelope functions (Breslaw (1994)) are recognized as not accurate or efficient in many cases. Geweke (1991) and

Robert (1995) develop methods based on various combinations of standard rejection sampling methods for TUVN within the Gibbs sampler for TMVN. Both of these methods have low acceptance rates for certain types of restrictions. Geweke (1991)'s Gibbs sampler for TMVN suffers from slow convergence, and our empirical studies illustrate that this is especially the case for unbounded regions. Rodrigues-Yam et al. (2004) propose a type of transformation focusing on simplifying the covariance by uncorrelating the random vectors, which somewhat fix the poor mixing property of Geweke (1991)'s method. However, most of these previous work does not provide theoretical properties (e.g., acceptance rates of rejection sampling etc.).

In this paper, we develop an efficient mixed rejection sampling method for TUVN distributions and a Gibbs sampler for TMVN distributions. The layout of the paper is as follows. In Section 2, we describe our sampling methods for both of the TUVN and TMVN distributions. The analytical acceptance rates of the proposed mixed rejection sampling method for TUVN distributions are derived analytically and compared to Geweke (1991)'s and Robert (1995)'s methods. The results demonstrate that our method has uniformly larger analytical acceptance rates than both of the existing methods. This sampling method is generalized to truncated multivariate Student-t distribution (TMVT) in Section 3. In Section 4, we present the empirical acceptance rates for TUVN distributions. Further numerical examples with our Gibbs sampler for TMVN and TMVT distributions with various restriction regions are presented, to demonstrate the accuracy and the fast convergence of the proposed method. In Section 5, we provide conclusions and further discussions in this area. Proofs of the lemmas and additional tables and figures are provided in the appendices. An R package (TMVN) has been developed to carry out all empirical work which is available from the first author.

2 Efficient Truncated Normal Sampling Method

The main goal of this paper is to develop an efficient sampling method for the TMVN distribution. A p -dimensional random variable \mathbf{W} is said to follow a truncated multivariate (p -variate) normal distribution subject to linear inequality constraints, if its pdf is

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{\exp \left\{ -\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu}) \right\}}{\int_{\mathbf{c} \leq \tilde{\mathbf{R}}\mathbf{w} \leq \mathbf{d}} \exp \left\{ -\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu}) \right\} d\mathbf{w}} I(\mathbf{c} \leq \tilde{\mathbf{R}}\mathbf{w} \leq \mathbf{d}), \quad (1)$$

where $\boldsymbol{\Sigma}$ is a non-negative definite matrix, I is the indicator function, and $\tilde{\mathbf{R}}$ is a $m \times p$ matrix with rank $m \leq p$. The inequality notation of $\mathbf{c} \leq \tilde{\mathbf{R}}\mathbf{w} \leq \mathbf{d}$ means that the inequality holds element-wisely, i.e., $c_i \leq [\tilde{\mathbf{R}}\mathbf{w}]_i \leq d_i$ for each $i = 1, 2, \dots, m$, with c_i 's and d_i 's allowed to be $-\infty$ and $+\infty$, respectively. We denote this TMVN distribution as $\mathbf{W} \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d})$. A good collection of statistical properties of TMVN distributions can be found in Horrace (2005).

We state two key properties of TMVN distributions, which enable the development of the proposed sampling methods.

Proposition 1. *Let $\mathbf{Y} = \mathbf{A}\mathbf{W}$, where \mathbf{A} is a $q \times p$ matrix with rank $q \leq p$. Then,*

$$\mathbf{Y} \sim TN_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T; \mathcal{T}), \quad (2)$$

where $\mathcal{T} = \{\mathbf{A}\mathbf{w} : \mathbf{c} \leq \tilde{\mathbf{R}}\mathbf{w} \leq \mathbf{d}\}$.

Proposition 2. *Partition \mathbf{X} , $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as*

$$X = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where \mathbf{X}_1 is a p_1 -dimensional random vector and \mathbf{X}_2 is a p_2 -dimensional random vector, and

$p_1 + p_2 = p$. Then the conditional distribution of \mathbf{X}_1 given $\mathbf{X}_2 = \mathbf{x}_2$ is

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim TN_{p_1}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}; \mathcal{R}_1(\mathbf{x}_2)), \quad (3)$$

where

$$\mathcal{R}_1(\mathbf{x}_2) = \left\{ \mathbf{x}_1 \in \mathbb{R}^{p_1} : \mathbf{a} \leq \mathbf{R} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \leq \mathbf{b} \right\}.$$

A similar result holds for the conditional distribution of \mathbf{X}_2 given $\mathbf{X}_1 = \mathbf{x}_1$.

It is worth mentioning that in both (2) and (3), the restriction regions can not be explicitly written in the form of $\mathbf{c}^* \leq \tilde{\mathbf{R}}^* \mathbf{x}^* \leq \mathbf{d}^*$ for some $\tilde{\mathbf{R}}^*$, \mathbf{c}^* and \mathbf{d}^* , however, the constraints are still linear inequalities.

With these properties, we are able to use transformations to simplify general TMVN distributions and apply Gibbs sampler as the full conditionals are TUVN distributed. Therefore, an efficient and accurate TUVN distribution sampler is key to the problem.

2.1 Algorithms for Truncated Univariate Normal Distributions

In this section, we introduce our method for generating samples from a TUVN distribution. A random variable W is said to follow a TUVN distribution, if its pdf is

$$f_W(w) = \frac{\frac{1}{\sigma} \phi\left(\frac{w - \mu}{\sigma}\right)}{\Phi\left(\frac{d - \mu}{\sigma}\right) - \Phi\left(\frac{c - \mu}{\sigma}\right)} I(c \leq w \leq d) \propto \phi\left(\frac{w - \mu}{\sigma}\right) I(c \leq w \leq d), \quad (4)$$

where ϕ denotes the pdf of $N(0, 1)$, and Φ denotes its cumulative distribution function (cdf). In the restriction, the bounds c and d are allowed to be $-\infty$ and $+\infty$, respectively. We denote

the distribution as $W \sim TN(\mu, \sigma^2; c, d)$. It easily follows that

$$X = \frac{W - \mu}{\sigma} \sim TN(0, 1; a, b), \quad (5)$$

where $a = (c - \mu)/\sigma$, and $b = (d - \mu)/\sigma$, and the transformed random variable X has the pdf given by $f_X(x) = \frac{\phi(x)}{\Phi(b) - \Phi(a)} I(a \leq x \leq b)$. Notice that a sampled value x generated from $TN(0, 1; a, b)$ can be used to obtain a sample from $TN(\mu, \sigma^2; c, d)$ by the transformation $w = \sigma x + \mu$. Therefore, without loss of generality, we establish an efficient sampling method for the standard TUVN random variable X .

2.1.1 Mixed Rejection Algorithm

In this section, we develop a new mixed accept-reject algorithm, by combining four standard rejection sampling methods with the goal of optimizing the acceptance rate uniformly.

Following the work of Geweke (1991) and Robert (1995), the proposed sampling algorithm is based on basic accept-reject algorithm, which is a common method for generating samples from a distribution. More details can be seen in Chapter 2 of Robert and Casella (2004). This algorithm is stated as a lemma below.

Lemma 1. *To draw a sample from $X \sim f(x)$, if for all x , there exists a constant $M \geq 1$, and a known density function $g(x)$ defined on a common support with $f(x)$, such that $f(x) \leq Mg(x)$, then it is sufficient to generate $x \sim g$ and $u \sim Unif[0, 1]$, and take x as a sample from $f(x)$, until $u \leq f(x)/(Mg(x))$. The resulting acceptance rate is then $1/M$, and this acceptance rate is maximized at $M_{opt} \doteq \sup_{x: g(x) > 0} \frac{f(x)}{g(x)}$.*

The proposed algorithm is called mixed rejection algorithm because, depending on the type of restriction, the envelope is one of four standard distributions, and we define the corresponding basic rejection sampling methods as: the normal rejection sampling (NR), the half-normal

rejection sampling (HR), the uniform rejection sampling (UR), and the translated-exponential rejection sampling (ER). With the goal of maximizing the acceptance rate globally, we propose the following algorithm:

Proposed Algorithm for TUVN:

Case 1 For the truncated interval $[a, \infty)$; if

- (i) $a \leq 0$: use the normal rejection sampling,
- (ii) $0 < a < a_0$: use the half-normal rejection sampling,
- (iii) $a \geq a_0$: use the one-sided translated-exponential.

Case 2 For the truncated interval $0 \in [a, b]$; if

- (i) $b \leq a + \sqrt{2\pi}$: use the uniform rejection sampling,
- (ii) $b > a + \sqrt{2\pi}$: use the normal rejection sampling.

Case 3 For the truncated interval $[a, b]$, $a \geq 0$; if

- (i) $0 \leq a < a_0$, and if
 - (a) $b \leq b_1(a)$: use the uniform rejection sampling,
 - (b) $b > b_1(a)$: use the half-normal rejection sampling;
- (ii) $a \geq a_0$, and if
 - (a) $b \leq b_2(a)$: use the uniform rejection sampling,
 - (b) $b > b_2(a)$: use the two-sided translated-exponential rejection sampling.

Case 4 For the truncated interval $(-\infty, b]$: use the symmetric algorithm to Case 1.

Case 5 For the truncated interval $[a, b]$, $b \leq 0$: use the symmetric algorithm to Case 3.

In the above algorithm, the constants are given by $a_0 = 0.2570$, $b_1(a) = a + \sqrt{\frac{\pi}{2}} \exp\left\{\frac{a^2}{2}\right\}$, and $b_2(a) = a + \frac{2}{a + \sqrt{a^2 + 4}} \exp\left\{\frac{a^2 - a\sqrt{a^2 + 4}}{4} + \frac{1}{2}\right\}$. We then show that this mixed accept-reject algorithm has globally optimal acceptance rates and is thus more efficient than those proposed by Geweke (1991) and Robert (1995).

We first describe each of the four basic parts of the proposed algorithm, and provide the optimized acceptance rate for each part. From here on, a and b are finite numbers, unless mentioned differently.

Normal rejection sampling. Normal rejection sampling is a natural yet naive sampling method for TUVN, in which we set the envelope function $g(x) = \phi(x)$. Hence we draw a candidate x from $N(0, 1)$, accept x as a sample of the TUVN if it is in the range of $[a, b]$. The resulting acceptance rate is

$$p_N \doteq p_N(a, b) = \Phi(b) - \Phi(a).$$

Here a and b are allowed to take the values as $\pm\infty$ respectively.

Half-normal rejection sampling. When $a \geq 0$, we can consider the half-normal rejection sampling. In this method, a candidate x is drawn from $N(0, 1)$, retain x if $|x| \in [a, b]$. It is easy to show that the acceptance rate is

$$p_H \doteq p_H(a, b) = 2(\Phi(b) - \Phi(a)).$$

Here b is allowed to be $+\infty$. Notice that this is more efficient than normal rejection sampling when $a \geq 0$.

Uniform rejection sampling. When the interval is bounded, uniform distribution can be considered as the envelope distribution for sampling from TUVN distribution. We set $g(x) =$

$1/(b-a)$ as the pdf of $Unif[a, b]$. Since $\phi(x)$ is maximized at $x = 0$ if x is unrestricted, the constant that maximizes the acceptance rate is given by $M_{opt} = \frac{\Phi(b) - \Phi(a)}{(b-a)\phi\{bI(b \leq 0) + aI(a \geq 0)\}}$. Hence, the corresponding acceptance rate is

$$\begin{aligned} p_{U,1} &\doteq p_{U,1}(a, b) = \frac{\sqrt{2\pi}}{b-a}(\Phi(b) - \Phi(a)), \text{ if } a \leq 0 \leq b; \\ p_{U,2} &\doteq p_{U,2}(a, b) = \frac{\sqrt{2\pi}}{b-a}e^{\frac{a^2}{2}}(\Phi(b) - \Phi(a)), \text{ if } a \geq 0; \\ p_{U,3} &\doteq p_{U,3}(a, b) = \frac{\sqrt{2\pi}}{b-a}e^{\frac{b^2}{2}}(\Phi(b) - \Phi(a)), \text{ if } b \leq 0. \end{aligned}$$

Translated-exponential rejection sampling. The translated-exponential rejection sampling method is initially considered for the type of the one-side restriction intervals with the form of $[a, \infty]$ for $a \geq 0$. The envelope density of this method is a translated-exponential distribution with pdf

$$g(x) = \lambda \exp\{-\lambda(x-a)\}I(x \geq a), \quad (6)$$

which is the standard exponential distribution with parameter λ shifted to the right by a unit. As the lower bound a gets large, this translated-exponential distribution resembles the TUVN distribution. We state the optimized acceptance rate of the one-sided translated-exponential rejection sampling in the following lemma.

Lemma 2. *The maximized acceptance rate of the one-sided translated-exponential rejection sampling is*

$$p_{E,1} \doteq p_{E,1}(a, b) = \sqrt{2\pi}\lambda^* \exp\left\{-\frac{\lambda^{*2}}{2} + \lambda^*a\right\} \Phi(-a),$$

where

$$\lambda^* \doteq \lambda^*(a) = \frac{a + \sqrt{a^2 + 4}}{2}. \quad (7)$$

It is worth pointing out that, as a approaches ∞ , the acceptance rate $p_{E,1}$ goes to 1, which

illustrates that as $a \rightarrow \infty$, the translated-exponential distribution starts to resemble the TUVN distribution. Using λ^* does not increase the computation time as for any given value of a , if the translated-exponential rejection sampling is chosen for that region, λ^* only needs to be evaluated once.

The translated-exponential rejection sampling method can also be used for two-sided restriction intervals as $[a, b]$, provided that the upper bound b is relatively large. In the two-sided translated-exponential rejection sampling, the candidate x is generated from the translated-exponential distribution with pdf defined in (6), until $x \leq b$. Following a similar procedure of maximizing the acceptance rate of the one-sided translated-exponential distribution (as in Lemma 2), we get the maximized acceptance rate for the two-sided case as

$$p_{E,2} \doteq p_{E,2}(a, b) = \sqrt{2\pi} \lambda^* \exp \left\{ \lambda^* a - \frac{1}{2} \lambda^{*2} \right\} (\Phi(b) - \Phi(a)).$$

We then compare the four basic rejection sampling methods using a case-by-case basis by maximizing the acceptance probability for all restriction intervals. Based on the different natures and suitabilities of the basic rejection sampling methods, the restriction intervals are divided into five categories as listed in the algorithm. Note that the sampling method of TUVN with restriction region $[-b, -a]$ can be directly derived from the method for the restriction region $[a, b]$. To sample $x \sim TN(0, 1; -b, -a)$, one can generate $y \sim TN(0, 1; a, b)$, and take $x = -y$. In this sense, we call the regions $[a, b]$ and $[-b, -a]$ symmetric to each other. The above argument also holds when $a = -\infty$, or $b = \infty$. Hence, Case 4 is symmetric to Case 1, and Case 5 to Case 3. Therefore, we focus on optimizing the sampling methods for the first three cases.

It is well known that naively using only NR is impractical for TUVN sampling method. NR may work well when the mean 0 is contained in the region $[a, b]$. When a is several standard deviations away to the right of 0, NR is very inefficient. For the one-sided region, NR is only valid when $a \leq 0$. Note that p_H is uniformly larger than p_N , hence HR is favored over NR as

long as it is suited for the type of regions. UR generally produces satisfactory acceptance rate for narrow finite restriction regions. When the whole region is above the mean 0, for either one-sided region, or the two-sided region with a relatively large upper bound, ER is likely to be preferred, as the shape mimics the TUVN distribution better. We now explore the four basic rejection sampling methods and obtain the best acceptance rates.

Case 1. When $a \leq 0$, NR is the only available method for the one-sided restriction regions.

When $a > 0$, both HR and ER are also candidates, but with higher acceptance rates than NR. Therefore, we only consider these two methods when $a > 0$. The comparison between them is stated in the following lemma, and it results the choice of envelope functions stated in the algorithm.

Lemma 3 (HR vs. ER). *For the one-sided region with the form $[a, \infty)$ with $a \geq 0$, there exists $a_0 > 0$, such that when $0 < a < a_0$, $p_H > p_E$. When $a \geq a_0$, $p_E \geq p_H$. In above, a_0 is the solution of the equation*

$$\lambda^* e^{\frac{\lambda^{*2}}{2} - 1} = \sqrt{\frac{2}{\pi}},$$

which is approximately 0.2570 (up to 4 decimal places), and λ^ is given as in (7).*

Case 2. For this type of two-sided finite restriction regions, only NR and UR are suitable. The results of comparing the acceptance rates of the two methods are given in the lemma below and the choice of envelope functions for this case follows.

Lemma 4 (NR vs. UR). *For the two-sided region with the form $[a, b]$, where $a < 0 < b$, when $b - a \leq \sqrt{2\pi}$, $p_{U,1} \geq p_N$; when $b - a > \sqrt{2\pi}$, $p_N > p_{U,1}$.*

Case 3. This type of two-sided regions require the most detailed comparative analysis since all four rejection sampling methods are available. However, as p_H is uniformly larger than p_N for any a and b , we do not consider NR for this case. We then focus on the comparisons among the other three methods and the results are stated as follows.

Lemma 5 (HR vs. UR). *For the two-sided regions with the form $[a, b]$, where $a \geq 0$, when*

$$b < a + \sqrt{\frac{\pi}{2}} \exp \left\{ \frac{a^2}{2} \right\} \doteq b_1(a),$$

$p_{U,2} > p_H$. Otherwise, $p_H \geq p_{U,2}$.

Lemma 6 (UR vs. ER). *For the two-sided regions with the form $[a, b]$, where $a \geq 0$, if*

$$b \leq a + \frac{2}{a + \sqrt{a^2 + 4}} \exp \left\{ \frac{a^2 - a\sqrt{a^2 + 4}}{4} + \frac{1}{2} \right\} \doteq b_2(a), \quad (8)$$

$p_{U,2} \geq p_{E,2}$. Otherwise, $p_{E,2} > p_{U,2}$.

Combining Lemma 3, Lemma 5, and Lemma 6, for Case 3, we first discriminate the situations according to the comparison between HR and ER, as this comparison only involves the value of a , and then consider the comparison between UR and ER. As we always favor the method with larger acceptance rate, the choice stated in the algorithm results.

2.1.2 Comparisons of Analytical Acceptance Rates

Our studies show that generally, Robert (1995)'s method outperforms Geweke (1991)'s for Case 1 and Case 3, while the dominance reverses for Case 2, in terms of acceptance rates. To compare the acceptance rates of the proposed method with the two existing ones, we plot them for Case 1 to 3, shown in Figures 1 to 3. As the switching criteria for Cases 2 and 3 involves both a and b , we present some representative values of a for illustration. These figures clearly show the sub-optimal performance in acceptance rates of the two widely used existing methods.

Generally speaking, in Geweke (1991)'s algorithm, there are four ad-hoc constants serving as critical values of deciding which of the four basic rejection sampling methods to be used. However, theoretical analysis shows that this ad-hoc choice may cause very low acceptance rates

in some cases. Moreover, in Case 2 and 3, Geweke (1991)'s switching criteria involve evaluations of $\phi(a)$ and $\phi(b)$, which may also cause low efficiency when using in a Gibbs sampler. As for Robert (1995)'s method, HR is not considered, hence the algorithm does not take advantage of the similar nature of the half-normal distributions to the TUVN when the lower bound is close to 0 for one-sided regions, or for relatively wide bounded regions.

In Figure 1 for Case 1, the two blue lines indicate the switching values for the proposed method: 0 and $a_0 = 0.2570$. Between these two values, the choice of HR outperforms both of the other two methods. Note that when $a = 0$, our acceptance rate is exactly 1, because $TN(0, 1; 0, \infty)$ is exactly the half-normal distribution. For this case, Robert (1995) uses ER with acceptance rate 0.760, and Geweke (1991) uses the naive NR with acceptance rate 0.500. In Geweke (1991)'s method, the value separating NR and ER (non-optimal) is 0.45, and the resulted acceptance rate is as low as 0.326, compared to ours at 0.822, which illustrates that Geweke (1991)'s choice of the four constants is not optimal.

In Figure 2 for Case 2, acceptance rates are plotted for four values of a . The blue lines indicate the switching value $a + \sqrt{2\pi}$, depending on a . This figure gives a clear illustration that UR only performs well for narrow restriction intervals. However, it is the only sampling method used by Robert (1995) for this case. Although Geweke (1991) uses the similar idea as ours, since the switching criteria are not optimized, it has smaller acceptance rates for relatively narrow regions.

In Figure 3 for Case 3, acceptance rates of four values of a are plotted, two of which are less than $a_0 = 0.2570$, two of which are larger. The switching value indicated by the blue line is $b_1(a)$ for a less than a_0 , and $b_2(a)$ otherwise. When $0 \leq a < a_0$, Robert (1995)'s acceptance rates are smaller for wide restriction regions because it uses ER instead of HR. Geweke (1991) uses either HR or ER (non-optimal) for wide regions. Due to the choice of envelope functions, the acceptance rates are low for certain regions.

These comparisons give an illustration of how the proposed method outperforms the two

existing ones. Moreover, as it does not involve much complicated computations, the running time is similar to that of the other existing methods when sampling TUVN alone.

2.2 Algorithm for Truncated Multivariate Normal Distributions

Now that we have an efficient sampling method for TUVN distributions, we focus on the sampling method for the TMVN distribution, which is the central interest of this paper.

To simplify the procedure of the Gibbs sampler, usually some transformations are done before the Gibbs sampler is implemented. For any general TMVN distributed random vector $\mathbf{W} \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d})$, Geweke (1991) proposes to use the transformation as $\mathbf{Z} = \tilde{\mathbf{R}}(\mathbf{W} - \boldsymbol{\mu})$. The motivation behind this is that the resulting restrictions of the random vector \mathbf{Z} becomes $\boldsymbol{\alpha} \leq \mathbf{Z} \leq \boldsymbol{\beta}$, where $\boldsymbol{\alpha} = \mathbf{a} - \tilde{\mathbf{R}}\boldsymbol{\mu}$ and $\boldsymbol{\beta} = \mathbf{b} - \tilde{\mathbf{R}}\boldsymbol{\mu}$, hence the restrictions are explicitly defined for every component without any linear summation involved. Although this transformation simplifies the restrictions, the distribution of \mathbf{Z} is $TN_p(\mathbf{0}, \tilde{\mathbf{R}}\boldsymbol{\Sigma}\tilde{\mathbf{R}}^T; \mathbf{I}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, hence the covariance matrix is complicated. The resulting full conditional distributions required within the Gibbs sampler are not truncated standard univariate normal. As the sampling method is defined for $TN(0, 1; a, b)$ only, in each of the Gibbs updating step, two transformations are required, one of which is to transform the nonstandard TUVN to the standard TUVN, and another one is to transform the generated sample back to the original full conditional distribution for later updating steps, and in each of the step, one matrix inversion is needed. This causes inefficiency in Geweke (1991)'s sampling method for TMVN especially when a large number of samples are of interest from a high dimensional TMVN distribution. The non-diagonal covariance matrix is also a source for the poor mixing in the Gibbs sampler. Moreover, since the backward transformation is needed, the matrix $\tilde{\mathbf{R}}$ is required to be squared and invertible, which limits the application of the method.

To improve the mixing, we introduce the transformation as

$$\mathbf{X} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{W} - \boldsymbol{\mu}), \quad (9)$$

where $\boldsymbol{\Sigma}^{-1/2}$ denotes the inverse of $\boldsymbol{\Sigma}^{1/2}$, which is the lower triangular Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$. According to Proposition 1, $\mathbf{X} \sim TN_p(\mathbf{0}, \mathbf{I}; \mathbf{R}, \mathbf{a}, \mathbf{b})$, with $\mathbf{R} = \tilde{\mathbf{R}}\boldsymbol{\Sigma}^{1/2}$, $\mathbf{a} = \mathbf{c} - \tilde{\mathbf{R}}\boldsymbol{\mu}$, and $\mathbf{b} = \mathbf{d} - \tilde{\mathbf{R}}\boldsymbol{\mu}$. This transformation greatly simplifies the pdf given in (1), instead of only the constraints of the TMVN distribution. As a result, we do not require $\tilde{\mathbf{R}}$ to be an invertible matrix, and hence allow that $m \leq p$ with linearly independent rows. By Proposition 2, the resulted full univariate conditional distribution is

$$x_i | \mathbf{x}_{-i} \sim TN(0, 1; a_i^*(\mathbf{x}_{-i}), b_i^*(\mathbf{x}_{-i})), \quad i = 1, \dots, p, \quad (10)$$

where $\mathbf{x}_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$, and $a_i^*(\mathbf{x}_{-i})$ and $b_i^*(\mathbf{x}_{-i})$ are determined suitably such that $\mathbf{a} \leq \mathbf{R}\mathbf{x} \leq \mathbf{b}$. The full conditional distribution described in (10) shows the advantage of making such a transformation, that in each updating step of the Gibbs sampler, we deal with a truncated standard univariate normal distribution. This procedure improves the poor mixing of the Gibbs sampler proposed by Geweke (1991), especially for wider restriction regions. Without loss of generality, we develop the sampling method for the random vector \mathbf{X} . To obtain the samples for the original TMVN vector \mathbf{W} , we simply use the following transformation $\mathbf{w} = \boldsymbol{\Sigma}^{1/2}\mathbf{x} + \boldsymbol{\mu}$, where \mathbf{x} is obtained by the Gibbs sampling. Our method only requires two linear matrix transformations, and one matrix inversion in the Cholesky decomposition of the covariance matrix.

We next explain how to evaluate the lower and the upper bounds a_i^* and b_i^* in every updating step. Let \mathbf{R}_i denote the i th column of \mathbf{R} , \mathbf{R}_{-i} denote the $m \times (p-1)$ matrix by removing the i th column of \mathbf{R} . Hence, the restrictions $\mathbf{a} \leq \mathbf{R}\mathbf{x} \leq \mathbf{b}$ is equivalent to $\mathbf{a} - \mathbf{R}_{-i}\mathbf{x}_{-i} \leq \mathbf{R}_i x_i \leq$

$\mathbf{b} - \mathbf{R}_{-i}\mathbf{x}_{-i}$, which is equivalent to the following coordinate-wise inequalities

$$a_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i} \leq r_{ji}x_i \leq b_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}, \quad j = 1, \dots, m, \quad (11)$$

where $\mathbf{r}_{j,-i}$ denotes the j th row of the matrix \mathbf{R}_{-i} , and r_{ji} denotes the j th entry of the vector \mathbf{R}_i . Depending on the sign of r_{ji} , we have three different scenarios:

When $r_{ji} > 0$. For all j 's such that $r_{ji} > 0$, (11) is equivalent to $\frac{a_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}} \leq x_i \leq \frac{b_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}}$.

Hence we define $l_i^+ = \max_{\{j: r_{ji} > 0\}} \frac{a_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}}$ and $u_i^+ = \min_{\{j: r_{ji} > 0\}} \frac{b_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}}$.

When $r_{ji} < 0$. For all j 's such that $r_{ji} < 0$, (11) is equivalent to $\frac{b_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}} \leq x_i \leq \frac{a_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}}$.

Hence we define $l_i^- = \max_{j: r_{ji} < 0} \frac{b_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}}$ and $u_i^- = \min_{j: r_{ji} < 0} \frac{a_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}}{r_{ji}}$.

When $r_{ji} = 0$. There is no restriction on x_i for this j .

Combining all three scenarios discussed above, we have

$$a_i^* = \max\{l_i^+, l_i^-\} \text{ and } b_i^* = \max\{u_i^+, u_i^-\}.$$

If $a_i^* \leq b_i^*$, these bounds can be used for sampling from the full conditional distribution in each step of the Gibbs sampler. In fact, the region defined by $\{\mathbf{x} : \mathbf{a} \leq \mathbf{R}\mathbf{x} \leq \mathbf{b}\}$ is a convex polytope. Hence as long as the starting value is within the region, our method remains feasible in every updating step. Therefore, in the i th step of the t th pass in the Gibbs sampler, we sample

$$x_i^{(t)} | x_1^{(t)}, \dots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \dots, x_p^{(t-1)} \sim TN(0, 1; a_i^{*(t)}, b_i^{*(t)}), \quad (12)$$

where $a_i^{*(t)} = a_i^*(x_1^{(t)}, \dots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \dots, x_p^{(t-1)})$, and $b_i^{*(t)} = b_i^*(x_1^{(t)}, \dots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \dots, x_p^{(t-1)})$, for $i = 1, \dots, p$.

Empirical results in Section 4 demonstrate that the proposed Gibbs sampler has very good mixing property and converges fast to the true truncated multivariate normal distribution.

3 Efficient Truncated Student-T Sampling Method

The sampling method for TMVN distributions can be easily generalized to a method for sampling from truncated multivariate Student-t (TMVT) distributions, as a TMVT can be obtained as the ratio of a TMVN to the square root of an independent chi-squared random variable divided by its degree of freedom. In this section, we describe the algorithm of sampling from TMVT distributions based on the efficient sampling method for TMVN distributions.

A random vector \mathbf{Y} following a truncated multivariate (p -variate) Student-t distribution with degree of freedom ν subject to linear inequality constraints $\mathbf{c} \leq \tilde{\mathbf{R}}\mathbf{y} \leq \mathbf{d}$ is denoted as

$$\mathbf{Y} \sim TT_p(\mu, \Sigma, \nu; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d}).$$

Following the property of a multivariate Student-t distribution, a similar transformation as in (9) can be done such that $\mathbf{T} = \Sigma^{-1/2}(\mathbf{Y} - \mu) \sim TT_p(\mathbf{0}, \mathbf{I}, \nu; \tilde{\mathbf{R}}, \mathbf{a}, \mathbf{b})$, where $\Sigma^{-1/2}$, $\tilde{\mathbf{R}}$, \mathbf{a} , and \mathbf{b} are defined as before. Therefore, without loss of generality, we develop the sampling algorithm for \mathbf{T} . To obtain a sample \mathbf{y} of the distribution of \mathbf{Y} from a sample \mathbf{t} of \mathbf{T} , we only require a simple transformation as $\mathbf{y} = \Sigma^{1/2}\mathbf{t} + \mu$.

Let $\mathbf{T}^* \sim T_p(\mathbf{0}, \mathbf{I}, \nu)$ denote an untruncated p -variate Student-t distributed random variable with degree of freedom ν , then by definition, $\mathbf{T}^* = \frac{\mathbf{X}^*}{\sqrt{U/\nu}}$, where $\mathbf{X}^* \sim N_p(\mathbf{0}, \mathbf{I})$ is a standard p -variate normal distributed random variable, and $U \sim \chi^2(\nu)$ is a chi-squared distributed random variable with degree of freedom ν , independent of \mathbf{X}^* . Hence, for a given sample u from the distribution of U , the linear constraints of \mathbf{T} can be expressed by $\mathbf{a} \leq \mathbf{RT} = \frac{\mathbf{RX}}{\sqrt{u/\nu}} \leq \mathbf{b}$, which implies that $\mathbf{a}\sqrt{u/\nu} \leq \mathbf{RX} \leq \mathbf{b}\sqrt{u/\nu}$. These facts indicate that a Gibbs sampler can be used to sample from TMVT distributions based on the existing TMVN sampling method.

The algorithm of the Gibbs sampler to obtain a sample \mathbf{t} from $TT_p(\mu, \Sigma, \nu; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d})$ is described as follows:

1. *Sample* $u \sim \chi^2(\nu)$.
2. *Sample* $\mathbf{x} \sim TN_p(\mathbf{0}, \mathbf{I}; \mathbf{R}, \mathbf{a}\sqrt{u/\nu}, \mathbf{b}\sqrt{u/\nu})$.
3. *Set* $\mathbf{t} = \frac{\mathbf{x}}{\sqrt{u/\nu}}$.

In the second step, our proposed Gibbs sampler for TMVN sampling will be implemented. As the efficiency and accuracy of the TMVT distributions are highly influenced by that of the TMVN distributions, empirical studies given in Section 4 illustrate that this Gibbs sampler inherits the good mixing property and fast convergence from the sampling method for TMVN distributions.

4 Empirical performnaces

In this section, we present several numerical illustrations for the proposed sampling methods of TUVN, TMVN, and TMNT distributions, and comparisons with other popular methods.

4.1 Sampling from TUVN

We first present the empirical results for the proposed mixed rejection sampling algorithm for TUVN distributions.

As mentioned previously, in terms of acceptance rates, Robert (1995)'s algorithm outperforms Geweke (1991)'s for Case 1 and Case 3, and vice versa for Case 2. As a result, it is of interest to compare our proposed algorithm with Robert (1995)'s for Cases 1 and 3, and with Geweke (1991)'s for Case 2. The empirical acceptance rates for some representative regions are listed in Tables 1 through 3. For each of the restriction intervals, 10,000 samples are generated using three methods: the newly proposed method, Geweke (1991)'s, and Robert (1995)'s. For each method, the empirical acceptance rates are computed as $10,000/N_{try}$, where N_{try} denotes the total number of trials until 10,000 samples are accepted.

Figure 4 is the histograms of the samples obtained by the proposed algorithm with the exact density curves overlayed for $TN(2, 1; c, d)$. The bounds are provided in the captions of the figures. It clearly indicates that sampling distribution captures the true shape of the density of the TUVN. In addition to depicting histograms, we also carried out a few standard goodness of fit tests (Anderson-Darling, Kolmogorov-Smirnov, etc), all of which further confirms the correct generating mechanism of the proposed method.

4.2 Sampling from TMVN and TMVT

We illustrate our sampling method for several examples to explore how well our method works for different types of TMVN and TMVT distributions.

Example 1. In this example, we apply our method to truncated bivariate normal (TBVN) distributions. For TBVN, we are able to display the accuracy of the sampling method on a contour plot of the distribution. This is difficult for higher dimensional TMVN distributions due to the property that the marginal distributions are not TMVN and have very complex pdf's.

Consider a random vector $\mathbf{X} = (X_1, X_2)^T$ with $\boldsymbol{\mu} = (0, 0)^T$, $Var(X_1) = 10$, $Var(X_2) = 0.1$, and correlation ρ between two components. We investigate 6 different restriction regions. Define $\sigma_1 = \sqrt{Var(X_1 + X_2)}$, $\sigma_2 = \sqrt{Var(X_1 - X_2)}$, and $\mathbf{sd} = (\sigma_1, \sigma_2)^T$. In the first 3 cases, the restriction regions are bounded in which \mathbf{X} is truncated by $\pm 1.5\mathbf{sd}$, $\pm 0.15\mathbf{sd}$, and $\pm 0.05\mathbf{sd}$, respectively. The fourth and the fifth restriction regions are one-sided open regions, with the lower bounds set to be $-0.15\mathbf{sd}$ and $0.15\mathbf{sd}$, respectively, and the upper bounds both set to be $+\infty$. The sixth is an extreme case where the lower and the upper bounds are set to be $\pm\infty$, hence it is in fact an ordinary bivariate normal distribution. For each of the restriction regions, we test a moderate correlation $\rho = 0.5$, and a very high correlation $\rho = 0.98$. The initial values for the fifth case is $(1, 0)^T$, while those of all the others are $(0, 0)^T$. Results are based on 10,000 generated samples and 1,000 burn-ins.

Figures 5 and 6 display the contour plots of the density function overlaid with the last 5,000 generated samples of the bounded restriction region between $\pm 1.5\mathbf{sd}$, and the one-sided open restriction region $(-0.15\mathbf{sd}, \infty)$, with both $\rho = 0.5$ and 0.98 . It can be seen that the generated samples stay within the restricted regions, and sample points distribute according to the steps of the contour lines, which illustrates that the Gibbs sampler converges to target distribution.

Trace plots and auto-correlation function (ACF) plots are good visual indicators of the mixing property, which indicates how fast the MCMC chain converges to the stationary distribution. High correlations between two components may cause poor mixing. Note that even when $\rho = 0.98$, Figures 7 and 8 illustrate that our Gibbs sampler has very good mixing property and the chains converge fast to the true densities. This is the case for all scenarios using the proposed Gibbs sampler. We also apply the Gibbs sampler proposed by Geweke (1991) using the R function `rtmvnorm` in package `tmvtnorm`. For most of the scenarios, especially unbounded restriction regions and wider bounded restriction regions, even when the truncation is very moderate such as $\pm 1.5\mathbf{sd}$, this Gibbs sampler has poor mixing property and the chains converge very slowly, as seen in Figures 9 and 10. Changing to a lower correlation $\rho = 0.5$ improves the mixing property and accelerate the convergence for Geweke (1991)'s method. But the mixing is still not very satisfactory even for moderate truncated restriction regions as the trace plots are sticky and the ACF's decay to 0 fairly slowly, as seen in Figure 11. These results strongly illustrate that our Gibbs sampler fixes the poor mixing property and slow convergence caused by Geweke (1991)'s method.

For further diagnostics, we also compute the integrated auto-correlation time (IACT) for each individual chain. The closer the IACT is to 1, the more effective the sampler is. The average of the IACT for all of our combinations is 1.013, which is fairly close to 1. The CPU time for generating 10,000 samples using our Gibbs sampler for each combination is also recorded, and the average is about 4.4 seconds on a DELL Dual Processor Xeon Twelve Core 3.6 GHz machines with 80GB RAM running 64Bit CentOS Linux 5.0. The IACT and CPU time do not

vary much when the restriction regions and the correlations are changed. We also conduct the Geweke’s convergence diagnosis test and every chain passes the test. It is also worth pointing out that, in our other numerical experiments, we discover that the proposed Gibbs sampler has a better mixing property and faster convergence as the restriction regions become wider. However, the above results show that our method still work well for extremely tight truncation restriction regions, and we have not found a case where our method has trouble converging. The Gibbs sampler proposed by Geweke (1991) has a reverse trend, that is, it works better for tighter restriction regions, but this method becomes problematic for even moderately truncated regions and one-sided restriction regions, which are very commonly encountered in practice.

Example 2. In this example, we consider a 3-dimentional TMVN with the linear constraint matrix $\tilde{\mathbf{R}}$ having smaller rank than the dimension, which is given as

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}, \quad \text{and} \quad \tilde{\mathbf{R}} = \begin{pmatrix} 1 & -2 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

The constraints with two types of restriction regions, bounded and one-sided open, are specified as $\mathbf{c} = (0, 0)^T$, and $\mathbf{d} = (1, 2)^T$ or $(\infty, \infty)^T$. Results are based on 10,000 generations with 1,000 burn-ins. In examining the trace plots and the ACF plots, the proposed method has good mixing and converges fast to the true density. As it is hard to display samples with the contour plots of the true joint densities, or even the 2-dimensional marginal densities, to show the accuracy of our sampling method, we numerically compute the 1-dimensional marginal density for each of the component for the two types of the restriction regions. The numerical computation of the marginal densities involves evaluating the probability of the restriction region and the density of a multivariate normal distribution, which is done by employing the R package `mvtnorm`. The histograms of the generated samples with the marginal densities are given by Figures 12 and

13. These figures not only show that our Gibbs sampler gives a correct and accurate sampling method, but also reveals the interesting effects resulted by truncation. In both of the cases, the marginal distributions of X_1 and X_2 are clearly skewed, while those of X_3 are symmetric and appear to be normally distributed, which is expected because no constraints are imposed on X_3 .

Example 3. In this example, we apply the proposed Gibbs sample for a TMVT distribution that has the same parameter settings as in Example 1, with the degrees of freedom $\nu = 5$. Trace and ACF plots show good mixing property and fast convergence for all scenarios. Results for two selected scenarios can be seen in Figure 14, and the contour plots overlayed with the last 5,000 samples are displayed in Figure 15 for $\rho = 0.5$. Compared with Figure 5, there are more points generated at very low density area for TMVT. These demonstrate that our sampling method successfully captures the heavy-tailed nature of the multivariate Student-t distributions. Diagnostics shows that every chain in both of the cases has IACT approximately 1 and passes the Geweke diagnosis test.

5 Conclusions and Discussions

In this paper, we develop the sampling methods for both TUVN and TMVN distributions, and generalize the method to TMVT distributions. For TUVN distributions, we establish the mixed rejection sampling method, utilizing four basic rejection sampling methods. We prove that the proposed method has the optimized acceptance rate over the set of all intervals, thus is more efficient than the existing methods proposed by Geweke (1991) and Robert (1995). We then develop a Gibbs sampler for TMVN in which the TUVN full conditionals are updated using the proposed TUVN sampling method. When applying this sampler to a general TMVN distribution, a linear transformation is done aiming to uncorrelate the components and thus contributes to fix the potential poor mixing property suffered by Geweke (1991)’s method. Empirical studies

show that our method has great accuracy, and good mixing property regardless of the type of the restriction regions. This method is also computationally efficient as it only requires one matrix inversion in the Cholesky decomposition of the covariance matrix. We generalize the method to sampling from TMVT distributions as a multivariate Student-t distribution is a scale mixture density of a multivariate normal density. In fact, the sampling method for any scale mixture of the multivariate normal distributions can be derived based on this efficient method. Empirical study shows that the Gibbs sampler for TMVT inherits the good mixing property and fast convergence from the that for TMVN.

In addition to the sampling methods based on rejection sampling and Gibbs sampler, there are other methods including techniques based on the slice sampler (Neal (2003)), see Damien and Walker (2001) and Liechty and Lu (2010). In a recent paper, Yu and Tian (2011) propose methods using EM algorithm and a non-iterative inverse Bayes formulae sampling procedure. However, as theoretical properties of these methods are not currently available, it is not clear how to compare our proposed algorithm with these non-Gibbs type sampling methods. This remains an area of future research.

A Proof of Lemmas

In this appendix, we use the notation given in Lemma 1, and we further denote that $\rho(x) = \frac{f(x)}{Mg(x)}$.

Proof of Lemma 2. The feasible region for the one-sided translated-exponential rejection sampling is $[a, \infty)$, where $a \geq 0$. As the envelope distribution function is the pdf of a translated exponential distribution $Exp(\lambda; a)$, then

$$\frac{f(x)}{g(x)} = \frac{\frac{1}{1-\Phi(a)}\phi(x)I(x \geq a)}{\lambda e^{-\lambda(x-a)}I(x \geq a)} = \frac{1}{\sqrt{2\pi}\lambda(1-\Phi(a))} \exp\{-\frac{1}{2}x^2 + \lambda(x-a)\}.$$

To find M_{opt} , since e^x is increasing, we only need to maximize

$$Q \doteq -\frac{1}{2}x^2 + \lambda(x - a) = -\frac{1}{2}(x - \lambda)^2 + \frac{\lambda^2}{2} + \lambda a.$$

For known and fixed a , according to the relations between λ and a , we discuss the maximization in two separate scenarios.

Scenario 1: If $\lambda \geq a$, Q is maximized at $x = \lambda$, and $Q_{max} = \frac{\lambda^2}{2} - \lambda a$. Accordingly, $M_{opt} = \frac{1}{\sqrt{2\pi\lambda(1-\Phi(a))}} \exp\{\lambda^2 - \lambda a\}$, and $\rho(x) = \exp\{-\frac{1}{2}(x - \lambda)^2\}I(x \geq a)$. Hence the resulting acceptance rate is $E[\rho(X)] = \sqrt{2\pi}\lambda \exp\{-\frac{\lambda^2}{2} + \lambda a\}\Phi(-a)$. To maximize this acceptance rate, it is equivalent to maximize $E \doteq \lambda \exp\{-\frac{\lambda^2}{2} + \lambda a\}$, which is achieved by solving the equation

$$\frac{\partial E}{\partial \lambda} = (-\lambda^2 + a\lambda + 1)e^{-\frac{\lambda^2}{2} + \lambda a} = 0 \iff -\lambda^2 + a\lambda + 1 = 0.$$

Hence the probability $E[\rho(X)]$ is maximized at λ^* defined in (7). Then the optimized acceptance rate is $p_{E,1}$. Notice that $a = \lambda^* - \frac{1}{\lambda^*}$, hence we also have $p_{E,1} = \sqrt{2\pi}\lambda^* \exp\{\frac{\lambda^* a}{2} - \frac{1}{2}\}\Phi(-a)$.

Scenario 2: If $\lambda < a$, Q is maximized at $x = a$, and $Q_{max} = -\frac{a^2}{2}$. Accordingly, $M_{opt} = \frac{1}{\sqrt{2\pi\lambda(1-\Phi(a))}} \exp\{-\frac{a^2}{2}\}$, and $\rho(x) = \exp\{-\frac{1}{2}(x - \lambda)^2 + \frac{1}{2}(\lambda - a)^2\}I(x \geq a)$. Hence, the acceptance rate is $E[\rho(X)] = \sqrt{2\pi}\lambda \exp\{\frac{a^2}{2}\}\Phi(-a)$. Clearly, the acceptance rate is maximized at $\lambda = a$. Then the optimized acceptance rate is

$$P_2 = \sqrt{2\pi}a \exp\{\frac{a^2}{2}\}\Phi(-a).$$

To pick a λ for the rejection sampling, we then compare the two acceptance rates $p_{E,1}$ and P_2 resulted by two scenarios. By comparing $D = \lambda^* \exp\{-\frac{\lambda^{*2}}{2} + \lambda^* a\} - a \exp\{\frac{a^2}{2}\}$ with 0, we find that for any value of λ , $D > 0$. Hence $p_{E,1}$ is always greater than P_2 . Therefore, we choose

$\lambda = \lambda^*$ and the resulting maximized acceptance rate of the one-sided translate-exponential rejection sampling method is given as $p_{E,1}$. \square

Proof of Lemma 3. The type of region we compare for the two methods are $[a, \infty)$ with $a \geq 0$. We define D to be the kernel of the difference of the two acceptance rates p_H evaluated at $b = \infty$ and $p_{E,1}$, i.e., $D = 2 - \sqrt{2\pi}\lambda^*e^{\frac{\lambda^*a}{2}-\frac{1}{2}}$. We then examine the relation between D and 0. This can be done by solving the equation $D = 0$. Since λ^* is obtained by solving $-\lambda^2 + a\lambda + 1 = 0$, we have $a\lambda^* = \lambda^{*2} - 1$, which is equivalent to $\lambda^*e^{\frac{\lambda^{*2}}{2}-1} = \sqrt{\frac{2}{\pi}}$. Numerically solving the equation shows that when $0 < a < 0.2570$, $D > 0$, which means that the half-normal rejection sampling has a larger acceptance rate; when $a \geq 0.2570$, $D \leq 0$, which means that the translated-exponential rejection sampling has a larger acceptance rate. \square

Proof of Lemma 4. For the type of regions defined by $[a, b]$ with $a < 0 < b$, we compare the two acceptance rates given in p_N and $p_{U,2}$. It is equivalent to compare $D \doteq \frac{\sqrt{2\pi}}{b-a} - 1$ with 0. It is easy to see that when $b - a > \sqrt{2\pi}$, $D < 0$, which means that the normal rejection sampling has a larger acceptance rate, and when $b - a < \sqrt{2\pi}$, $D > 0$, which means that the uniform rejection has a larger acceptance rate. \square

Proof of Lemma 5. When the type of regions are a bounded region defined as $[a, b]$ with $a \geq 0$, we compare the acceptance rates of p_H and $p_{U,1}$. This is equivalent to compare 2 vs. $\frac{\sqrt{2\pi}}{b-a}e^{\frac{a^2}{2}}$. Since $\frac{\sqrt{2\pi}}{b-a}e^{\frac{a^2}{2}} - 2 > 0$ is equivalent to $b < \sqrt{\frac{\pi}{2}}e^{\frac{a^2}{2}} + a$, we conclude that when that holds, the uniform rejection sampling has a larger acceptance rate, otherwise, the half-normal rejection sampling does. \square

Proof of Lemma 6. For the bounded region defined as $[a, b]$ with $a \geq 0$, we compare the acceptance rates $p_{U,1}$ and $p_{E,2}$. This is equivalent to compare $\frac{\sqrt{2\pi}}{b-a}e^{\frac{a^2}{2}}$ and $\sqrt{2\pi}\lambda^* \exp\left\{-\frac{\lambda^{*2}}{2} + \lambda^*a\right\}$.

Since λ^* is defined by (7), $\frac{\sqrt{2\pi}}{b-a}e^{\frac{a^2}{2}} < \sqrt{2\pi}\lambda^* \exp\left\{-\frac{\lambda^{*2}}{2} + \lambda^*a\right\}$ is equivalent to

$$b > a + \frac{2}{a + \sqrt{a^2 + 4}} \exp\left\{\frac{a^2 - a\sqrt{a^2 + 4}}{4} + \frac{1}{2}\right\},$$

and this proves the lemma. □

B Tables and Figures

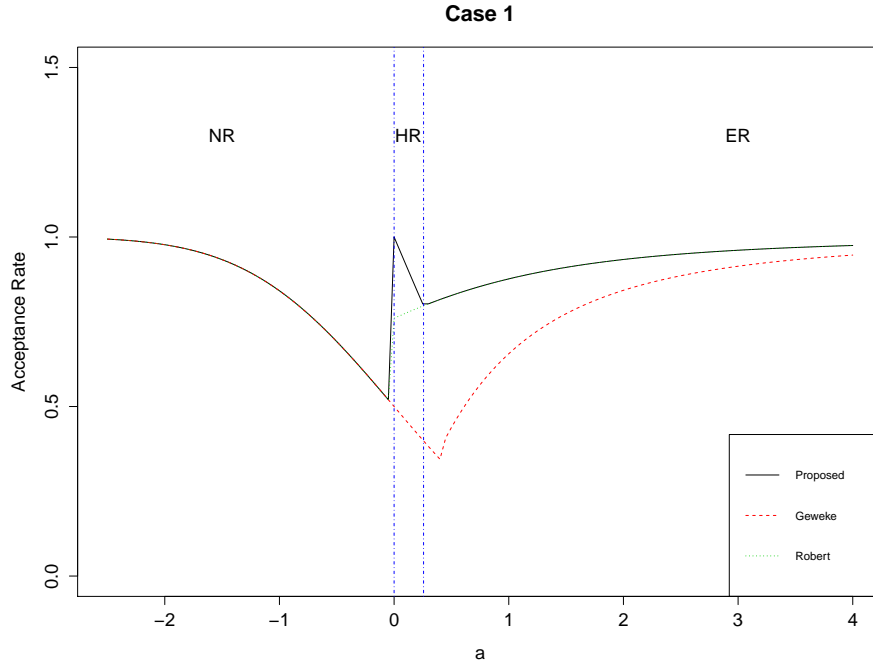


Figure 1: Comparing analytical acceptance rates for Case 1

Table 1: Comparing empirical acceptance rates for Case 1

a	New	Geweke	Robert	a	New	Geweke	Robert
-2	0.977	0.977	0.977	0.2	0.840	0.421	0.790
-1	0.840	0.840	0.840	0.45	0.823	0.324	0.823
-0.5	0.695	0.695	0.695	1	0.880	0.656	0.880
0	1.000	0.495	0.763	5	0.984	0.969	0.984

Table 2: Comparing empirical acceptance rates for Case 2

	New	Geweke	Robert		New	Geweke	Robert
b	$a = -2$				$a = -1$		
.5	0.672	0.660	0.672	0.897	0.897	0.897	
1	0.819	0.819	0.687	0.860	0.860	0.860	
2	0.954	0.954	0.606	0.819	0.819	0.690	
b	$a = -0.5$				$a = -0.1$		
2	0.680	0.673	0.680	0.624	0.515	0.624	

Table 3: Comparing empirical acceptance rates for Case 3

	New	Geweke	Robert		New	Geweke	Robert		New	Geweke	Robert
b	$a = 0$			b	$a = 1$			b	$a = 2$		
2	0.954	0.954	0.728	3	0.872	0.652	0.872	4	0.937	0.843	0.937
1	0.861	0.861	0.861	2	0.756	0.562	0.756	3	0.882	0.793	0.882
0.5	0.964	0.964	0.964	1.5	0.766	0.766	0.766	2.5	0.684	0.616	0.684
0.1	0.999	0.999	0.999	1.1	0.954	0.954	0.954	2.1	0.911	0.911	0.911

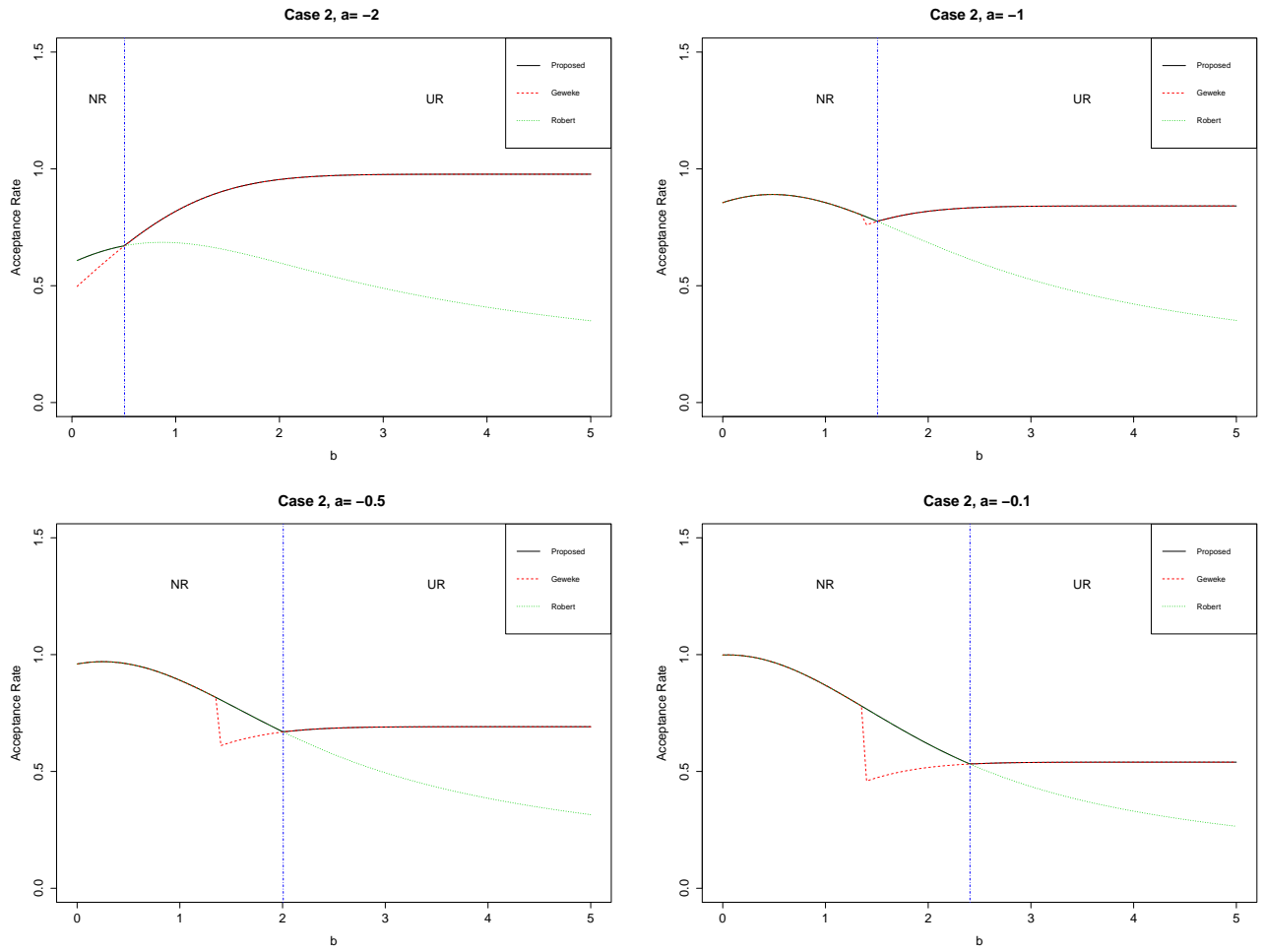


Figure 2: Comparing analytical acceptance rates for Case 2

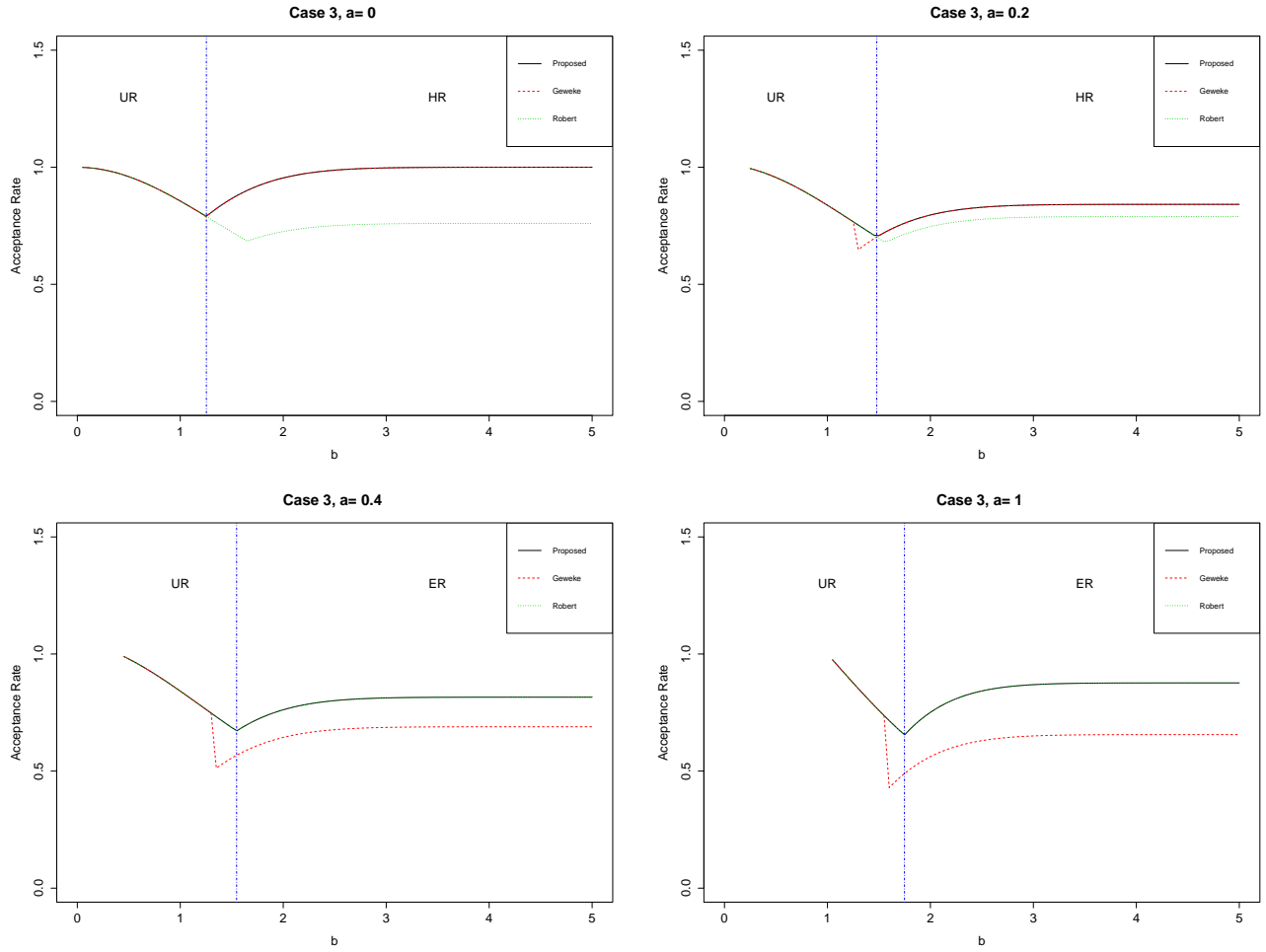


Figure 3: Comparing analytical acceptance rates for Case 3

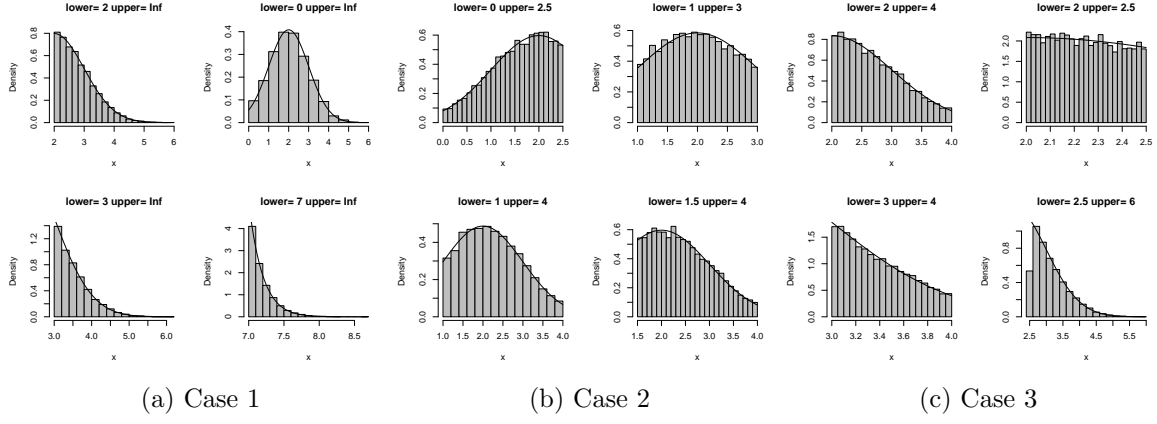


Figure 4: Histograms of samples obtained for Case 1, 2 and 3 ($\mu = 2, \sigma = 1$) in Section 4.1

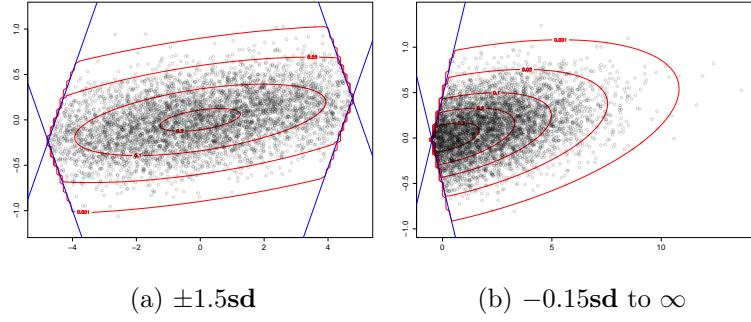


Figure 5: Selected Contour plots when $\rho = 0.5$ in Example 1

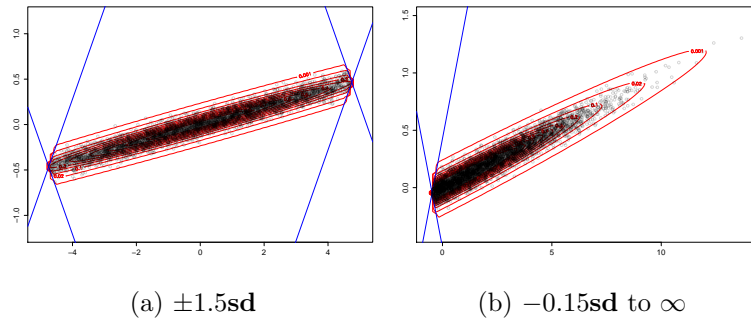


Figure 6: Selected Contour plots when $\rho = 0.98$ in Example 1

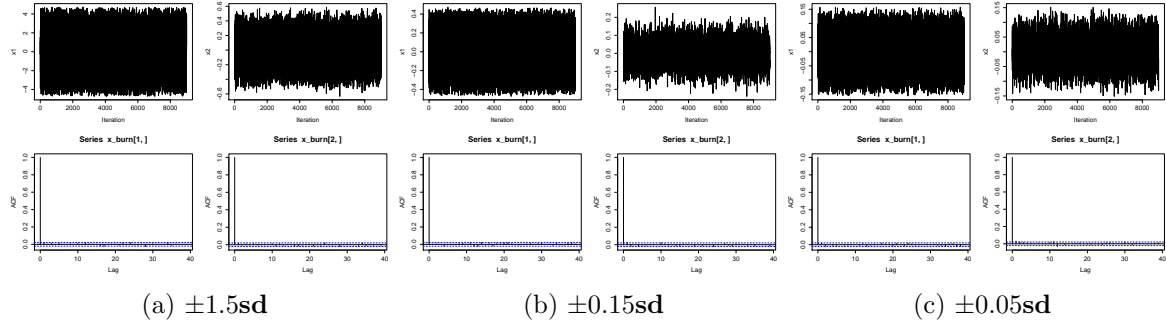


Figure 7: Trace and ACF plots when $\rho = 0.98$ in Example 1

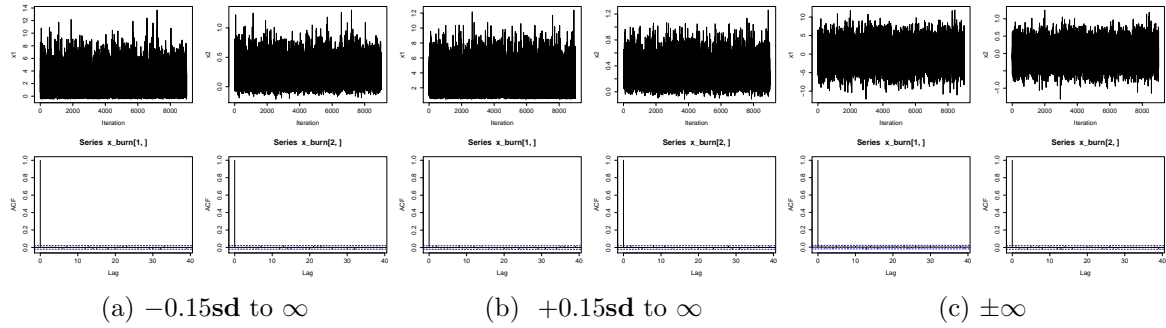


Figure 8: Trace and ACF plots when $\rho = 0.98$ in Example 1

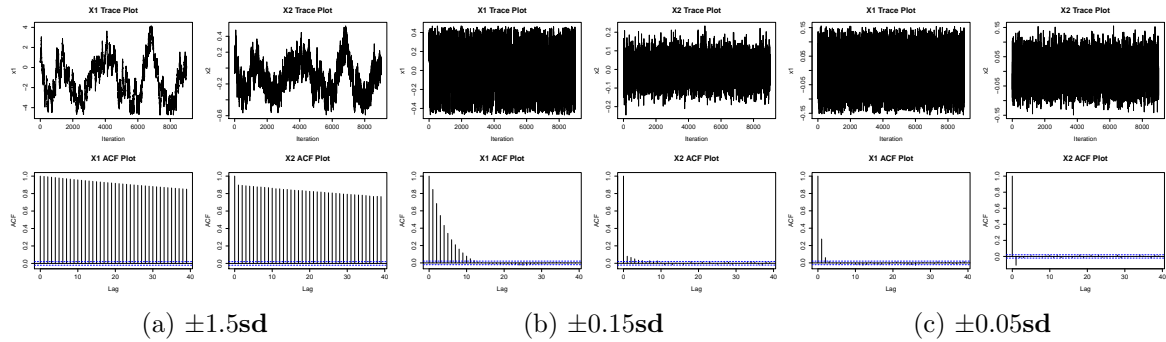


Figure 9: Trace and ACF plots when $\rho = 0.98$ by Geweke (1991)'s in Example 1

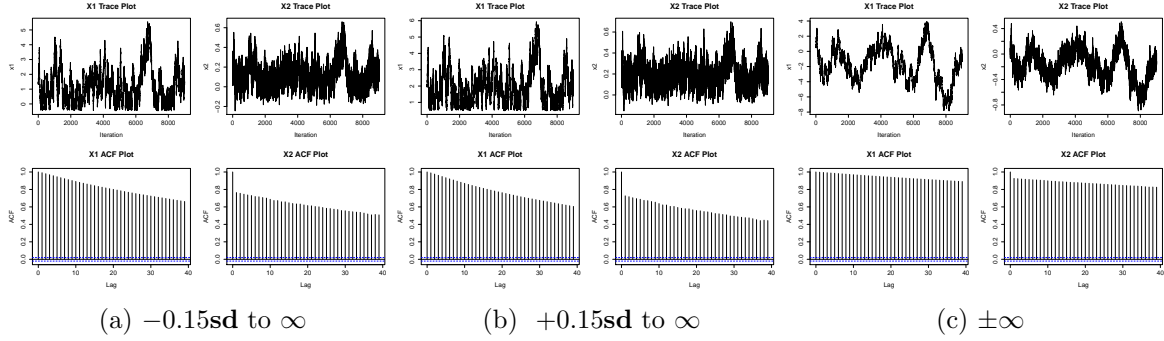


Figure 10: Trace and ACF plots when $\rho = 0.98$ by Geweke (1991)'s in Example 1

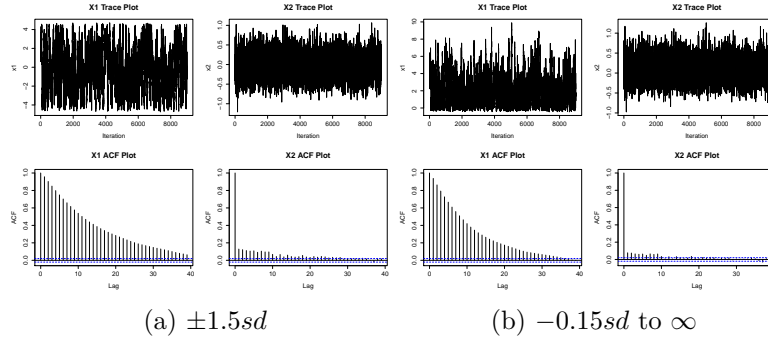


Figure 11: Selected Trace and ACF plots when $\rho = 0.5$ by Geweke (1991)'s in Example 1

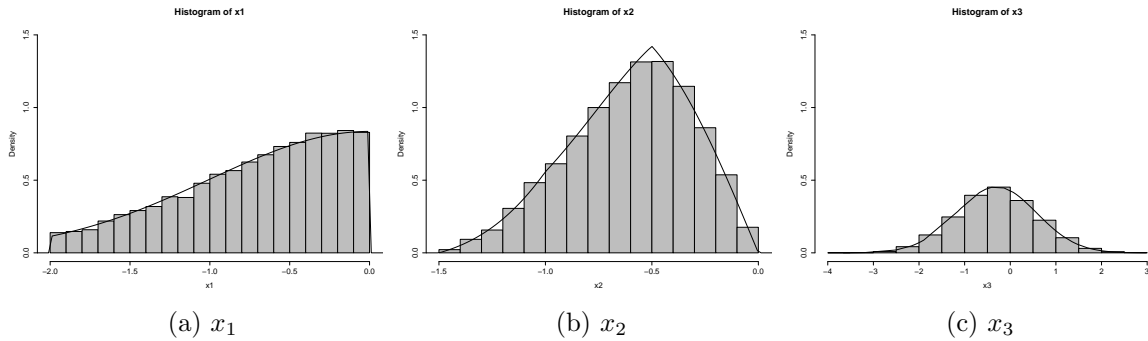


Figure 12: Histograms of generated samples with analytical marginal densities for the bounded region in Example 2

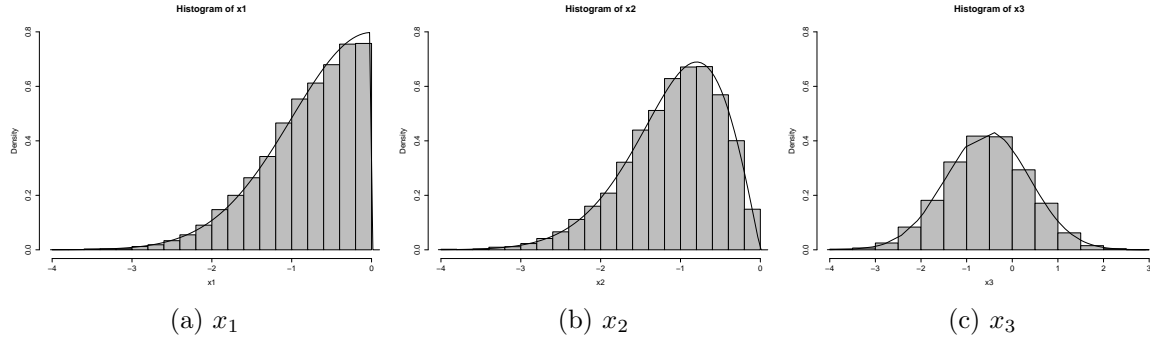


Figure 13: Histograms of generated samples with analytical marginal densities for the unbounded region in Example 2

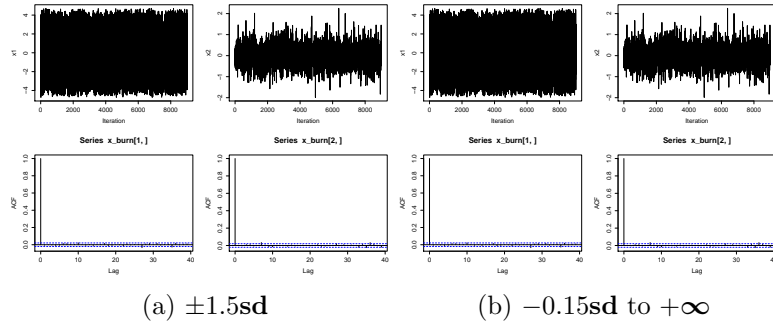


Figure 14: Selected Trace and ACF plots for TMVT when $\rho = 0.5$, $\nu = 5$ in Example 3

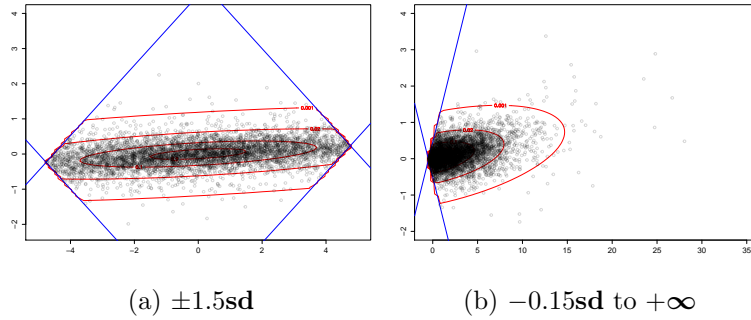


Figure 15: Selected Contour plots for TMVT when $\rho = 0.5$, $\nu = 5$ in Example 3

References

- Breslaw, J. (1994). Random sampling from a truncated multivariate normal distribution. *Applied Mathematics Letters*, 7(1):1 – 6.
- Damien, P. and Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2):206–215.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85(410):398–409.
- Gelfand, A. E., Smith, A. F. M., and Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *J. Amer. Statist. Assoc.*, 87(418):523–532.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741.
- Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In Keramidas, E. M., editor, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 571–578, Fairfax Station, VA. Interface Foundation of North America, Inc.
- Geweke, J. (1996). Bayesian inference for linear models subject to linear inequality constraints. In Lee, J. C., Johnson, W. O., and Zellner, A., editors, *Modelling and Prediction: Honoring Seymour Geisser*, pages 248–263. Springer, New York.
- Hajivassiliou, V. A. and Mcfadden, D. L. (1990). The method of simulated scores for the estimation of ldv models: with an application to the problem of external debt crises.cowles foundation discussion paper no. 967.

- Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. *J. Multivariate Anal.*, 94(1):209–221.
- Liechty, M. W. and Lu, J. (2010). Multivariate normal slice sampling. *J. Comput. Graph. Statist.*, 19(2):281–294. With supplementary material available online.
- Liu, J. S., Wong, W. H., and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B*, 57(1):157–169.
- Neal, R. (1997). Markov chain Monte Carlo methods based on ‘slicing’ the density function. Technical report, University of Toronto, Department of Statistics.
- Neal, R. M. (2003). Slice sampling. *Ann. Statist.*, 31(3):705–767. With discussions and a rejoinder by the author.
- Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5:121–125.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.
- Rodrigues-Yam, G., Davis, R. A., and Scharf, L. L. (2004). Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression. Technical report, Colorado State University.
- Yu, J. and Tian, G. (2011). Efficient algorithms for generating truncated multivariate normal distributions. *Acta Mathematicae Applicatae Sinica (English Series)*, 27:601–612.