

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

Yifang Li

Department of Statistics, North Carolina State University

2311 Stinson Dr., Raleigh, NC, 27695

yli40@ncsu.edu

Sujit K. Ghosh

Department of Statistics, North Carolina State University

2311 Stinson Dr., Raleigh, NC, 27695

sujit.ghosh@ncsu.edu

NC State Department of Statistics Technical Report # 2649

Abstract

Sampling from a truncated multivariate normal 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 still appears non-trivial due to the existence of the analytically intractable normalizing constant of the truncated multivariate normal distribution. In this paper, to start with, we develop an efficient mixed 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. As the full conditional distributions of a truncated multivariate normal distribution are truncated univariate normals, we employ the proposed superior univariate sampling method and implement the Gibbs sampler for sampling from a truncated multivariate normal distribution with convex polytope restriction regions. We also 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 normal distribution, truncated multivariate Student-t distribution, rejection sampling, Gibbs sampler, MCMC.

1 Introduction

The necessity of sampling from a truncated multivariate normal (TMVN) distribution subject to multiple linear inequality constraints arises in many applied areas of research in statistics and econometrics. Robert (1995) discusses several examples in order restricted (or isotonic) regressions and censored data models. Several other applications are illustrated 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 mentioned above, an efficient method to generate samples from a TMVN distribution is not so straightforward. One of the main barriers arises from the complex normalizing constant involved in the probability density function (pdf) of TMVN. Although the 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. Moreover, even if we have an efficient sampling method for the truncated univariate normal (TUVN) distribution, we cannot use direct sampling for each element either. This is because of the fact that the marginal and the forward conditional distributions of a TMVN distribution are not TUVN. Therefore, researchers have been seeking

more efficient methods to avoid the evaluation of the normalizing constant, and also to utilize the properties of the TMVN distribution itself.

The Gibbs sampler (Geman and Geman (1984) and Gelfand and Smith (1990)) has been a popular technique in sampling from TMVN distributions. The Gibbs sampler is well suited for the problem of sampling from TMVN distributions because all full conditional distributions of a TMVN distribution are TUVN distributions. Majorities of the sampling methods for generating samples from TMVN are based on the Gibbs sampler. An early survey is provided by Hajivassiliou and Mcfadden (1990). Since the Gibbs sampler generates samples iteratively from the full conditionals, the efficiency of this method is effectively determined by the efficiency of sampling from the TUVN distributions. Breslaw (1994) suggests to use the uniform rejection sampling method for the univariate full conditionals, which is recognized as not accurate or efficient in many cases. Geweke (1991) and Robert (1995) both develop methods based on various combinations of standard rejection sampling methods within the Gibbs sampler. These methods are now widely used in statistics, and a few R packages are based on these methods. Geweke (1991) also proposes a transformation on the restrictions of the TMVN distribution before the Gibbs sampler is carried out. However, both of their methods for TUVN distributions have low acceptance rates for certain types of interval restriction. Moreover, Geweke (1991)'s Gibbs sampler suffers from slow convergence, and our empirical studies illustrate that this is especially the case for unbounded regions. Rodrigues-Yam et al. (2004) propose another type of transformation focusing on simplifying the covariance by uncorrelating the random vectors using the square root of the covariance matrix instead of simplifying the restrictions, which somewhat fix the poor mixing property of Geweke (1991)'s method.

The slice sampler (Neal (2003)) is also an appropriate technique for the sampling problem of the TMVN distributions. Using this technique, Damien and Walker (2001) proposed to introduce a single auxiliary variable and employ the univariate slice sampler to sample the TMVN random vector one component at a time. In a recent paper, Liechty and Lu (2010) develop two slice

samplers using multiple auxiliary variables and updating the entire random vector at a time. This avoids the potential slow convergence of the Markov chain which can be caused by high correlation between the components of the TMVN random vector. In their paper, they also compare their method with the naive rejection method, and the one described in Neal (1997), where a single auxiliary variable is used and the entire random vector is updated at a time, and claim that their method is more efficient and accurate than the existing slice sampling methods. However, their method only works for rectangular restriction regions, which limits the applications of this method. Other methods can be seen in Yu and Tian (2011), where the authors propose two sampling methods for TMVN distributions. One is a data augmentation algorithm based on the EM algorithm, and the other one is a non-iterative inverse Bayes formulae sampling procedure. These methods also update the random vector components together at once.

In this paper, first we develop an efficient improved mixed rejection sampling method for TUVN distributions that is shown to have uniformly larger acceptance rates than the existing and widely used Geweke (1991)'s and Robert (1995)'s methods. Next we develop a Gibbs sampler using our proposed TUVN sampling method to sample from the full conditionals of the TMVN distributions subject to general linear inequality restrictions. Empirical results show that our proposed Gibbs sampler has good mixing property and fast convergence regardless of the type of the restriction regions. Moreover, as the multivariate Student-t distribution is a scale mixture of the multivariate normal distribution, the sampling method of TMVN distributions can be easily generalized to the truncated multivariate Student-t (TMVT) distributions. In fact, any scale mixture of the multivariate normal distributions can be derived based on this efficient sampling method. In this paper, we describe the algorithm for sampling from TMVT distributions, and empirical study shows that this sampling method inherits the good mixing property and fast convergence from the Gibbs sampler for TMVN.

In Section 2, we describe our sampling methods for both of the TUVN and TMVN distri-

butions. The analytical acceptance rates of our mixed rejection sampling method for TUVN distributions are calculated 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 in Section 3. In Section 4, we present the empirical acceptance rates and display the histograms of the samples along with the true densities for TUVN distributions overlayed. Further numerical examples with our Gibbs sampler for different 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 other related additional results are included in the appendices.

2 Efficient Truncated Normal Sampling Method

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

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

where $\boldsymbol{\Sigma}$ is a non-negative definite matrix, I is the indicator function. 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 $i = 1, 2, \dots, p$, with c_i 's and d_i 's allowed to be $-\infty$ and $+\infty$, respectively. We denote this TMVN distribution as

$$\mathbf{W} \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d}), \quad (2)$$

A good collection of statistical properties of TMVN distributions can be found in Horrace (2005).

In this paper, the proposed sampling method requires $\tilde{\mathbf{R}}$ to be a $m \times p$ matrix with rank $m \leq p$.

We state two key propositions of truncated multivariate normal distributions, which enable the development of the proposed sampling methods. In the following propositions, we assume that \mathbf{W} has the distribution defined in (2).

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

$$\mathbf{Y} \sim TN_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T; \mathcal{T}), \quad (3)$$

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

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

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

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 given by

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim TN_{p_1}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}; \mathcal{R}_1(\mathbf{x}_2)), \quad (4)$$

where

$$\mathcal{R}_1(\mathbf{x}_2) = \{\mathbf{x}_1 \in \mathbb{R}^{p_1} : \mathbf{a} \leq \mathbf{R}(\mathbf{x}_1, \mathbf{x}_2)^T \leq \mathbf{b}\}.$$

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 of (3) and (4), 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. Hence, according to Proposition 1, for any general random vector \mathbf{W}

distributed as in (2), we have

$$\mathbf{X} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{W} - \boldsymbol{\mu}) \sim TN_p(\mathbf{0}, \mathbf{I}; \mathbf{R}, \mathbf{a}, \mathbf{b}), \quad (5)$$

where

$$\mathbf{R} = \tilde{\mathbf{R}}\boldsymbol{\Sigma}^{1/2}, \quad \mathbf{a} = \mathbf{c} - \tilde{\mathbf{R}}\boldsymbol{\mu}, \quad \text{and} \quad \mathbf{b} = \mathbf{d} - \tilde{\mathbf{R}}\boldsymbol{\mu} \quad (6)$$

In (5), matrix $\boldsymbol{\Sigma}^{-1/2}$ denotes the inverse of $\boldsymbol{\Sigma}^{1/2}$, which is the lower triangular Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$. Thus, without loss of generality, we focus on the sampling method for the distribution in (5), which has a simpler form, instead of the general TMVN random vector \mathbf{W} .

It is well known that all full conditional distributions of TMVN are TUVN distributions. This fact motivates us to use the Gibbs sampler (Geman and Geman (1984)) for generating samples from the TMVN distribution, provided we have an efficient and accurate sampling method to generate samples from the truncated univariate normal (TUVN) distribution.

The Gibbs sampler is well suited for the case where the joint distribution is non-standard and is difficult to sample from, however, all of its full conditional distributions are some standard distributions which are relatively easier to sample from. The Gibbs sampler iteratively samples from the full conditional distribution producing an ergodic Markov chain which converges to the joint distribution under a set of mild regularity conditions (Liu et al. (1995)).

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 satisfies

$$f_W(w) = \frac{\frac{1}{\sigma} \phi\left(\frac{w-\mu}{\sigma}\right)}{\Phi\left(\frac{d-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right)} I(c \leq w \leq d) \propto \phi\left(\frac{w-\mu}{\sigma}\right) I(c \leq w \leq d), \quad (7)$$

where ϕ denotes the pdf of the standard normal distribution with mean 0 and variance 1, and Φ denotes its cumulative distribution function (cdf). These notations will be used throughout this paper. In the restriction, the bounds c and d are allowed to be $-\infty$ and $+\infty$, respectively. We denote the distribution as

$$W \sim TN(\mu, \sigma^2; c, d). \quad (8)$$

It easily follows that

$$X = \frac{W - \mu}{\sigma} \sim TN(0, 1; a, b), \quad (9)$$

where $a = (c - \mu)/\sigma$, and $b = (d - \mu)/\sigma$, and the transformed random variable X has the pdf given by

$$f_X(x) = \frac{\phi(x)}{\Phi(b) - \Phi(a)} I(a \leq x \leq b). \quad (10)$$

Suppose the sample generated from (9) is x , samples from W can be obtained by the transformation $w = \sigma x + \mu$. Therefore, without loss of generality, we will establish an efficient sampling method for the TUVN random variable distributed as in (9).

Much of the research has been done on sampling methods for TUVN distributions. Gelfand et al. (1992) propose to use the classical cdf inversion technique for sampling from $f_X(x)$. The procedure is to sample u from $Unif[\Phi(a), \Phi(b)]$, then the sample x can be obtained by setting $x = \Phi^{-1}(u)$. This method requires numerical approximations for computing the normal cdf Φ

and its inverse Φ^{-1} for each draw. It is especially inefficient when the bounds are extreme as the numerical evaluations of $\Phi(a)$ and $\Phi(b)$ are only reasonably accurate when $\max(|a|, |b|) \leq 4$, otherwise Φ is essentially 0 or 1, which may cause computational issues for the bounds of the uniform distribution. Furthermore, as u gets closer to 0 or 1, the computational time increases and the precision gets poorer for the evaluation of $\Phi^{-1}(u)$.

Later work on seeking efficient sampling methods mainly focuses on avoiding the numerical limitations of evaluating the cdf of normal distribution and its inverse. Geweke (1991) and Robert (1995) develop the sampling methods based on accept-reject algorithm. The two algorithms involve combinations of several rejection sampling methods. These mixed rejection sampling methods are substantially more efficient than the cdf inversion technique in terms of computational time and accuracy as no numerical integrations are required to sample random variables. However, for a selected types of constraint intervals, they both suffer from low acceptance rates.

In this section, we will develop a new mixed accept-reject algorithm, by combining four standard rejection sampling methods. We seek the optimum acceptance rate among the four rejection sampling methods for every type of restriction boundaries, hence improve the acceptance rate substantially compared to these two widely used sampling methods.

2.1.1 Basic Rejection Sampling Methods

The accept-reject algorithm 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

$$x \sim g \text{ and } u \sim \text{Unif}[0, 1],$$

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

$$M = M_{opt} = \sup_{x:g(x)>0} \frac{f(x)}{g(x)}.$$

Four basic rejection sampling methods are used in our mixed accept-reject algorithm for generating samples from TUVN distributions, which are defined by choosing the envelope function $g(x)$. The methods are the normal rejection sampling, the half-normal rejection sampling, the uniform rejection sampling, and the translated-exponential rejection sampling. In this section, we will describe each of the rejection sampling methods, and provide the optimized acceptance rate for each method. 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

$$\Phi(b) - \Phi(a). \tag{11}$$

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

$$2(\Phi(b) - \Phi(a)). \quad (12)$$

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 $\text{Unif}[a, b]$. 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

$$\frac{\sqrt{2\pi}}{b-a}(\Phi(b) - \Phi(a)), \quad \text{if } a \leq 0 \leq b; \quad (13)$$

$$\frac{\sqrt{2\pi}}{b-a}e^{\frac{a^2}{2}}(\Phi(b) - \Phi(a)), \quad \text{if } a \geq 0; \quad (14)$$

$$\frac{\sqrt{2\pi}}{b-a}e^{\frac{b^2}{2}}(\Phi(b) - \Phi(a)), \quad \text{if } b \leq 0.$$

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 a translated-exponential distribution defined as

$$g(x) = \lambda \exp\{-\lambda(x-a)\}I(x \geq a), \quad (15)$$

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*

$$\sqrt{2\pi} \lambda^* \exp \left\{ -\frac{\lambda^{*2}}{2} + \lambda^* a \right\} \Phi(-a), \quad (16)$$

where

$$\lambda^* \doteq \lambda^*(a) = \frac{a + \sqrt{a^2 + 4}}{2}. \quad (17)$$

It is worth pointing out that, as a approaches ∞ , the acceptance rate described in (16) goes to 1, which illustrates that as $a \rightarrow \infty$, the translated-exponential distribution starts to resemble the TUVN distribution.

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 (15), 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

$$\sqrt{2\pi} \lambda^* \exp \left\{ \lambda^* a - \frac{1}{2} \lambda^{*2} \right\} (\Phi(b) - \Phi(a)), \quad (18)$$

where λ^* is defined as in (17).

It is important to point out that choosing the value of λ as λ^* does not increase the computation time as stated in Geweke (1991). This is because for any given value of a , if the translated-exponential rejection sampling is chosen for that region, λ^* only needs to be evaluated once.

2.1.2 Mixed Rejection Algorithm

We now compare the four basic rejection sampling methods described in Section 2.1.1 using a case-by-case basis by maximizing the acceptance probability for all restriction intervals. According to the different natures and suitabilities of the basic rejection sampling methods, the restriction intervals are divided into five cases. The five cases considered are:

- Case 1: one-sided region $[a, \infty)$;
- Case 2: two-sided region $[a, b]$, where $a < 0 < b$;
- Case 3: two-sided region $[a, b]$, where $a \geq 0$;
- Case 4: one-sided region $(-\infty, b]$;
- Case 5: two-sided region $[a, b]$, where $b \leq 0$.

Note that the sampling method of TUVN with restriction region $[-b, -a]$ can be directly derived from the method for the restriction region $[a, b]$, provided the method for the latter region is given. 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 is symmetric to Case 3. Therefore, we will focus on optimizing the sampling methods for the first three cases, and the method for the last two cases can be obtained from their symmetric versions respectively.

Generally speaking, it is well known that naively using only the normal rejection sampling is impractical for TUVN sampling method. The normal rejection sampling may work well when the mean 0 is contained in the region $[a, b]$. However, when a is several standard deviations away to the right of 0, the normal rejection method should be very inefficient. For the one-sided region, the normal rejection method is only valid when $a \leq 0$. Note that the acceptance rate

of the half-normal rejection sampling method described in (12) is uniformly larger than the acceptance rate of the normal rejection sampling as in (11), as long as half-normal rejection sampling is suited for the type of regions. Hence, whenever half-normal rejection sampling can be used, we favor it over the normal rejection sampling. The uniform rejection sampling method generally produces satisfactory acceptance rate for narrow finite restriction regions. When the whole region is above the mean 0, and for either one-sided region, or the two-sided region with a relatively large upper bound, the translated-exponential rejection sampling is likely to be preferred, as the shape mimics the TUVN distribution better and thus the method will result in larger acceptance rates. We now explore the four basic rejection sampling methods and obtain the best acceptance rates.

Case 1. As stated above, when $a \leq 0$, the normal rejection sampling is the only available method for the one-sided restriction regions. Thus it will be used when $a \leq 0$. When $a > 0$, both half-normal and translated-exponential rejection sampling methods can be considered. Clearly, both of their acceptance rates are higher than the normal's in (11). Therefore, we only consider these two methods when $a > 0$. The comparison between them is stated in the following lemma.

Lemma 3 (Half-norm vs. Translated-exponential). *For the one-sided region with the form $[a, \infty)$ with $a \geq 0$, there exists $a_0 > 0$, such that when $0 < a < a_0$, the half-normal rejection sampling method has larger acceptance rate, i.e., (12) is larger than (16). When $a \geq a_0$, the translated-exponential rejection sampling method has larger acceptance rate, i.e., (16) is larger than (12). 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 (17).

According to the result in Lemma 3, when $0 < a < a_0$, we choose the half-normal rejection sampling over the translated-exponential rejection sampling. When a is larger than a_0 , the translated-exponential rejection sampling will be used.

Case 2. For this type of two-sided finite restriction regions, only the normal and the uniform rejection sampling methods are suited. The results of comparing the acceptance rates of the two methods are stated as follows.

Lemma 4 (Norm vs. Uniform). *For the two-sided region with the form $[a, b]$, where $a < 0 < b$, when $b - a \leq \sqrt{2\pi}$, the uniform rejection sampling method has larger acceptance rate, i.e., (13) is larger than (11); when $b - a > \sqrt{2\pi}$, the normal rejection sampling method has larger acceptance rate, i.e., (11) is larger than (13).*

Thus, when $b - a \leq \sqrt{2\pi}$, the uniform rejection sampling will be employed. Otherwise, normal rejection sampling method will be used. This result complies with the fact that the uniform rejection sampling performs better for shorter restriction regions.

Case 3. This type of two-sided regions require the most detail comparative analysis since all four rejection sampling methods are available. However, as the acceptance rate of the half-normal rejection sampling method described in (12) is uniformly larger than the normal rejection sampling method for any a and b , we do not consider the normal rejection sampling for this case. We then focus on the comparisons among the other three methods and the results are stated as follows.

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

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

the uniform rejection sampling method has larger acceptance rate, i.e., (14) is larger than

(12). Otherwise, the half-normal rejection sampling has larger acceptance rate, i.e., (12) is larger than (14).

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

$$b \leq a + \frac{2}{a + \sqrt{a^2 + 4}} \exp \left\{ \frac{a^2 - a\sqrt{a^2 + 4}}{4} + \frac{1}{2} \right\} \doteq b_2(a), \quad (19)$$

the uniform rejection sampling has larger acceptance rate, i.e., (14) is larger than (18). Otherwise, the two-sided translated-exponential rejection sampling has larger acceptance rate, i.e., (18) is larger than (14).

Taking considerations of the results stated in Lemma 3, Lemma 5, and Lemma 6, for Case 3, we first discriminate the situations according to the comparison between the half-normal rejection sampling and the translated-exponential rejection sampling, as this comparison only involves the value of a . Hence, we first consider the case when $0 \leq a < a_0$. Under this circumstance, when $b \leq b_1(a)$, we will use the uniform rejection sampling method; otherwise, we use the half-normal rejection sampling method. We then consider the case when $a \geq a_0$. In this latter case, when $b \leq b_2(a)$, the uniform rejection sampling is implemented for the TUVN sampling; otherwise the two-sided translated-exponential sampling will be used.

Combining all the comparisons among the basic rejection sampling methods for all cases, a globally optimal algorithm of our proposed mixed rejection sampling method is described as follows:

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

2.1.3 Comparisons of Analytical Acceptance Rates

Our numerical studies show that generally, Robert (1995)'s outperforms Geweke (1991)'s for Case 1 and Case 3, while the dominance reverses for Case 2, in terms of acceptance rates. As a result, it is of interest to compare our proposed algorithm with that of Robert (1995)'s for Cases 1 and 3, and with that of Geweke (1991)'s for Case 2. The analytical acceptance rates for some of the representative regions listed in Tables 1 through 3 demonstrate that our new

Table 1: Analytical Acceptance Rates for Case 1: $[a, \infty)$

a	New	Geweke	Robert	a	New	Geweke	Robert
-2	0.977	0.977	0.977	0.2	0.841	0.417	0.790
-1	0.841	0.841	0.841	0.45	0.822	0.326	0.822
-0.5	0.691	0.691	0.691	1	0.876	0.656	0.876
0	1.000	0.500	0.760	5	0.983	0.964	0.983

 Table 2: Analytical Acceptance Rates for Case 2: $0 \in [a, b]$

	New	Geweke	Robert	New	Geweke	Robert	
b	$a = -2$			$a = -1$			
.5	0.670	0.669	0.670	0.890	0.890	0.890	
1	0.819	0.819	0.684	0.856	0.856	0.856	
2	0.954	0.954	0.598	0.819	0.819	0.684	
b	$a = -0.5$			$a = -0.1$			
2	0.670	0.669	0.670	0.617	0.517	0.617	

mixed rejection sampling method uniformly outperforms both of the two methods for all types of regions.

These tables clearly show the sub-optimal performance in acceptance rates of the two widely used existing methods, since none of them were created by maximizing the acceptance rates. In Geweke (1991)'s algorithm, there are 4 constants serving as critical cut-off values of deciding the separations of the utility of the basic rejection sampling methods, which are claimed to find via computational empirical experiments only. However, theoretical analysis shows that this may cause very low acceptance rates, such as in Case 1 when $a = 0.45$, which is approximately the constant Geweke (1991) uses to divide the use of the normal and the translated-exponential

 Table 3: Analytical Acceptance Rates for Case 3: $[a, b], a \geq 0$

	New	Geweke	Robert		New	Geweke	Robert		New	Geweke	Robert
b	$a = 0$			b	$a = 1$			b	$a = 2$		
2	0.955	0.955	0.726	3	0.869	0.650	0.869	4	0.932	0.842	0.932
1	0.856	0.856	0.856	2	0.751	0.562	0.751	3	0.878	0.793	0.878
0.5	0.960	0.960	0.960	1.5	0.759	0.759	0.759	2.5	0.679	0.613	0.679
0.1	0.998	0.998	0.998	1.1	0.950	0.950	0.950	2.1	0.905	0.905	0.905

rejection sampling methods. The acceptance rate is as low as 0.326, which leads to a highly inefficient sampling method, especially when used within a Gibbs sampler. As for Robert (1995)'s method, the half-normal rejection sampling method is not considered. Hence, the algorithm does not take advantage of the similar nature of the half-normal distributions to the TUVN distribution when the lower bound is close to 0 for one-sided regions over the translated-exponential distribution, or for loose bounded regions over the uniform distribution. For an example of the low acceptance rate caused by Robert (1995)'s method, consider Case 1, when $a = 0.2$, and Case 3, when $a = 0$ and $b = 2$.

Notice that in Table 1, when $a = 0$, our acceptance rate is exactly 1, because $TN(0, 1; 0, \infty)$ is exactly the half-normal distribution. On the contrary, for this case, Robert (1995)'s uses the translated-exponential rejection sampling, and Geweke (1991) uses the naive normal rejection sampling, which are inefficient compared to our proposed method. In Table 2 for Case 2, when $a + \sqrt{2\pi} = 0.5066$ with $a = 2$, which is very close to the value of $b = 0.5$, hence the uniform rejection sampling method is used and results in a slightly larger acceptance rate than Geweke (1991)'s. Since in this case, Robert (1995)'s also utilizes the uniform rejection sampling, the acceptance rate is exactly the same as ours. Also in Table 2, when $a = -0.1$ and $a = -0.5$, the lower bound to use the normal rejection sampling is $b = 2.01$ and $b = 2.4$ respectively, hence in both of the cases when $b = 2$, the uniform rejection sampling is used, which again results in a larger acceptance rate than that of Geweke (1991)'s method which uses the normal rejection sampling method. Moreover, Table 2 gives a clear illustration of the fact that the uniform rejection sampling only performs well for shorter restriction intervals, as for Case 2, which is the only sampling method that is used by Robert (1995). It can be seen that as the interval becomes wider, the acceptance rate decreases significantly for the Robert (1995) method, yielding a low acceptance rate of only 0.598 when $a = -2$ and $b = 2$, compared to our normal rejection which has an acceptance rate of 0.954 in this case. Finally, consider the Case 3 shown in Table 3. In this case when $a = 0$, although the numerical acceptance rates are the same of our proposed

method and that of Geweke (1991) method, the empirical methods used in fact can be different. As in our method, the lower bound of using the half-normal rejection sampling is $b = 1.253$, while in Geweke (1991)'s, that value is 1.249, which is very close to our cut-off value, hence it is hard to make a difference between the two methods. Other than this value of a , Geweke (1991)'s method yields much smaller acceptance rates. Furthermore, Geweke (1991)'s method requires much more complicated computation involving two constants obtained empirically, and the evaluations of the ratio of $\phi(a)$ to $\phi(b)$. Moreover, in the case of $a = 0$ and $b = 2$, our method uses the half-normal rejection sampling, while Robert (1995)'s method uses the uniform rejection sampling, which produces smaller acceptance rate.

Therefore, in any case of the restriction regions, our newly proposed method outperforms both of the widely used existing methods with uniformly larger acceptance rates. This increases the efficiency of sampling from the TUVN distributions, which serves as the building blocks for the sampling method from the TMVN distributions within the Gibbs sampler. Moreover, as our method does not involve much complicated computations, the running time of our sampling program is similar to that of the other existing methods. We illustrate the empirical performance of the proposed method in Section 4.

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 from the TMVN distribution, which is the central interest of this paper. Although naively applying the multivariate normal rejection method can be used, it is well known that such a method would be highly impractical and inefficient. Direct rejection sampling methods are also difficult to construct as it is hard to find an all purpose envelope function for TMVN distributions. Methods for sampling from TMVN distributions also need to cope with the fact that its marginal distributions are not TMVN and are very complicated functions. As such,

many researchers have chosen the Gibbs sampler to sample from TMVN. In this paper, we will use this technique and focus on developing an efficient Gibbs sampler for generating samples from TMVN distributions.

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 $\mathbf{W} \sim TN_p(\boldsymbol{\mu}, \Sigma; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d})$ as in (2), Geweke (1991) proposes to use the transformation as

$$\mathbf{Z} = \tilde{\mathbf{R}}(\mathbf{W} - \boldsymbol{\mu}). \quad (20)$$

The motivation behind (20) 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, according to Proposition 1, the distribution of \mathbf{Z} is $TN_p(\mathbf{0}, \tilde{\mathbf{R}}\Sigma\tilde{\mathbf{R}}^T; \mathbf{I}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, hence the covariance matrix is not an identity matrix. By Proposition 2, the resulting full conditional distributions required within the Gibbs sampler are also 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 nonidentity 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.

In this paper, as argued at the beginning of this section, we introduce the transformation described in (5) as

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

The distribution of the obtained random vector is denoted as $\mathbf{X} \sim TN_p(\mathbf{0}, \mathbf{I}; \mathbf{R}, \mathbf{a}, \mathbf{b})$ in (9), where $\tilde{\mathbf{R}}$ and \mathbf{R} are both $m \times p$ matrices. This transformation focuses on transforming the mean and the variance of the multivariate normal distribution, instead of the constraints of the TMVN distribution. As a result, we do not require $\tilde{\mathbf{R}}$ to be a square matrix, and hence we allow that $m \leq p$ and require that the rows are linearly independent. The motivation of the transformation (5) is that it greatly simplifies the pdf of \mathbf{W} given in (1), as the transformed random vector \mathbf{X} has a truncated standard multivariate normal distribution with mean $\mathbf{0}$, and covariance matrix \mathbf{I} . By Proposition 2, the resulting full univariate conditional distribution is

$$x_i | \mathbf{x}_{-i} \sim TN(0, 1; a_i^*(\mathbf{x}_{-i}), b_i^*(\mathbf{x}_{-i})), \quad i = 1, \dots, p, \quad (21)$$

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 $\mathbf{a} \leq \mathbf{Rx} \leq \mathbf{b}$. The full conditional distribution described in (21) shows the advantage of making such a transformation, that in each updating step of the Gibbs sampler, we will 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 trans-

formations, and one matrix inversion on 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 $\mathbf{a} \leq \mathbf{Rx} \leq \mathbf{d}$ 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

$$a_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i} \leq r_{ji}x_i \leq b_j - \mathbf{r}_{j,-i}\mathbf{x}_{-i}, \quad (22)$$

for all $j = 1, \dots, m$, 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 $r_{ji} > 0$, (22) 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

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

When $r_{ji} < 0$. For all j 's such that $r_{ji} < 0$, (22) 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

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

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 $\{\mathbf{x} : \mathbf{a} \leq \mathbf{Rx} \leq \mathbf{b}\}$ 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

$$x_i^{(t)} | x_1^{(t)}, \dots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \dots, x_p^{(t-1)} \sim TN(0, 1; a_i^{*(t)}, b_i^{*(t)}), \quad (23)$$

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 very fast to the true joint 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-square 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

$$\mathbf{Y} \sim TT_p(\mu, \Sigma, \nu; \tilde{\mathbf{R}}, \mathbf{c}, \mathbf{d}). \quad (24)$$

Following the property of a multivariate Student-t distribution, a similar transformation as in (5) can be done such that

$$\mathbf{T} = \Sigma^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) \sim TT_p(\mathbf{0}, \mathbf{I}, \nu; \mathbf{R}, \mathbf{a}, \mathbf{b}), \quad (25)$$

where $\Sigma^{-1/2}$, $\tilde{\mathbf{R}}$, \mathbf{a} , and \mathbf{b} are defined as in (6). Therefore, without loss of generality, we develop the sampling algorithm for \mathbf{T} . To obtain a sample \mathbf{y} of the distribution given in (24) from a sample \mathbf{t} of that in (25), we only require a simple transformation as

$$\mathbf{y} = \Sigma^{1/2}\mathbf{t} + \boldsymbol{\mu}.$$

We 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-square 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

$$\mathbf{a} \leq \mathbf{R}\mathbf{T} = \frac{\mathbf{R}\mathbf{X}}{\sqrt{u/\nu}} \leq \mathbf{b},$$

which implies that

$$\mathbf{a}\sqrt{u/\nu} \leq \mathbf{R}\mathbf{X} \leq \mathbf{b}\sqrt{u/\nu}.$$

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 the distribution given in (25) 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, simulation 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 Simulation Studies

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

Table 4: Simulated Acceptance Rates for Case 1: $[a, \infty)$

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 5: Simulated Acceptance Rates for Case 2: $0 \in [a, b]$, $a < 0 < b$

	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	

4.1 Sampling from TUVN

In this sub-section, we present the empirical results for our proposed mixed rejection sampling algorithm for TUVN distributions.

Tables 4 to 6 show the empirical acceptance rates along with the theoretical analytical acceptance rates given in Table 1 to 3. For each of the restriction intervals, 10,000 samples are generated using three methods: our 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. Similarly, the acceptance rates are calculated for Cases 1 to 3 only, as Case 4 and Case 5 are symmetric to Case 1 and Case 3 respectively.

Figures 1 and 2 display the histograms of the samples obtained by our algorithm with the exact density curves overlayed. In these examples, a $TN(\mu, \sigma^2; c, d)$ with parameters $\mu = 2$ and $\sigma = 1$ are used. The bounds are provided in the captions of the figures. The histograms presented in Figures 1 and 2 clearly indicate that sampling distribution capture the true shape

Table 6: Simulated Acceptance Rates for Case 3: $[a, b]$, $a \geq 0$

	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

of the density of the TUVN. It is worth noticing that the figures displayed in Figure 2 for Case 4 and 5 and in Figure 1 for Cases 1 and 3 are symmetrical to the line $x = \mu$, respectively, which further re-establishes the fact that we only require symmetric algorithms for Cases 4 and 5. In addition to depicting histograms, we also carried out statistical tests of goodness of fit (such as Anderson-Darling test, Kolmogorov-Smirnov test, etc.), all of which further confirms the correct generating mechanism of the proposed method.

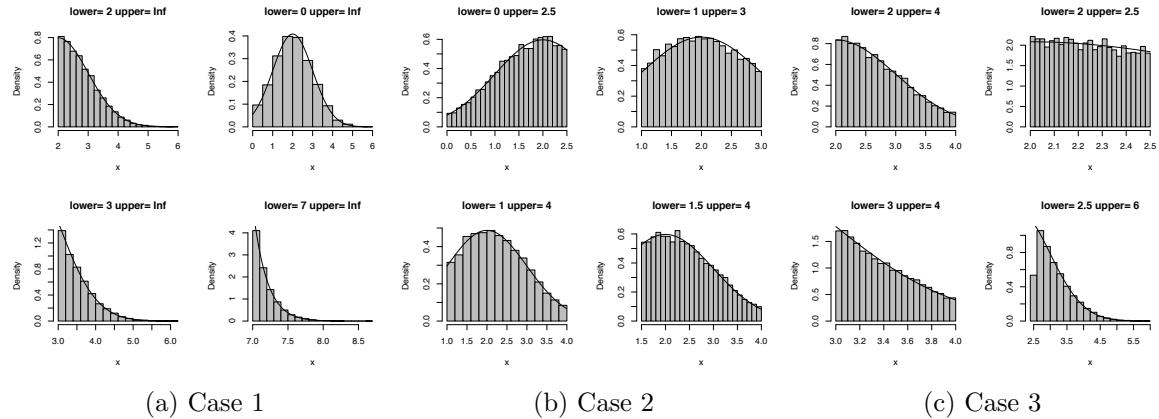
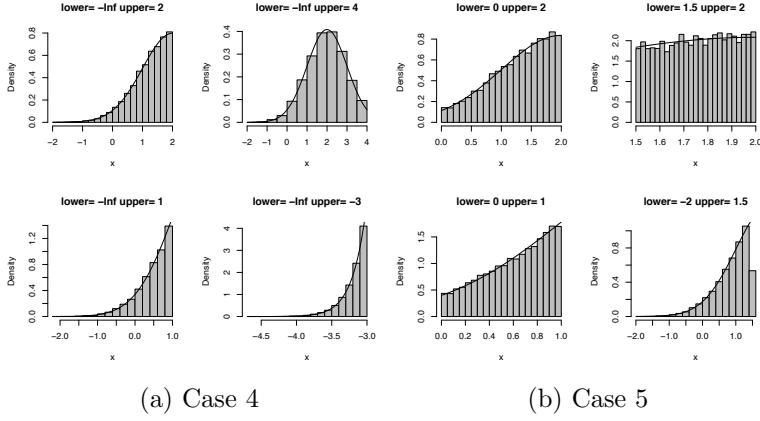


Figure 1: Case 1, 2 and 3 ($\mu = 2, \sigma = 1$)

4.2 Sampling from TMVN

In this sub-section, we illustrate our sampling method for several examples to explore how well our method works for different type of TMVN distributions.



(a) Case 4

(b) Case 5

Figure 2: Case 4 and 5 ($\mu = 2, \sigma = 1$)

Example 1. In this example, we employ our method to truncated bivariate normal (TBVN) distributions. We choose these TBVN distributions only for the sake of conveniently displaying the accuracy of our 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.

In this example, our method is implemented for a random vector $\mathbf{X} = (X_1, X_2)^T$ which is a bivariate normal distribution with $\boldsymbol{\mu} = (0, 0)^T$, $Var(X_1) = 10$, $Var(X_2) = 0.1$, and the correlation matrix

$$\mathbf{V} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

truncated by the linear constraints given as

$$\tilde{\mathbf{R}} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

We investigate 6 different restriction regions. We define

$$\begin{cases} \sigma_1 = \sqrt{\text{Var}(X_1 + X_2)} = \sqrt{10.1 + 2\rho} \\ \sigma_2 = \sqrt{\text{Var}(X_1 - X_2)} = \sqrt{10.1 - 2\rho} \end{cases} \quad \text{and } \mathbf{sd} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}.$$

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 final one is an extreme case where the lower and the upper bounds are set to be $\pm\infty$ respectively, hence it is in fact an ordinary bivariate normal distribution. For each of the restriction regions, we also test a moderate correlation between X_1 and X_2 with $\rho = 0.5$, and a very high correlation with $\rho = 0.98$. The initial values for the fifth case is $(1, 0)^T$, while those of all the others are $(0, 0)^T$.

We generate 10,000 samples for every combination. Figures 3 and 4 display the contour plots of the density function overlayed with the last 5,000 generated samples of the bounded restriction region represented in case 1, and the one-sided open restriction region represented in case 4, with $\rho = 0.5$ and 0.98 , respectively. 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 very fast to target distribution. Figures 3 and 4 also show that with higher correlation between the two components, the samples become more concentrated on a line, while with lower correlation, the samples are more spread over the plane.

The mixing of a MCMC chain shows how fast the MCMC chain converges to the stationary distribution. Trace plots and auto-correlation function(ACF) plots are good visual indicators of the mixing property. These plots are shown in Figures 5 and 6 with 10,000 sample generations and 1,000 burn-ins for $\rho = 0.98$. It is very satisfactory to see that even with this extremely high

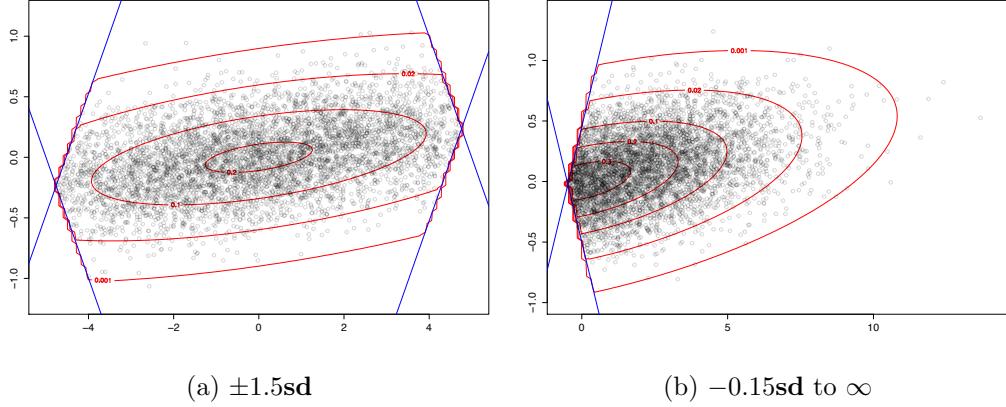


Figure 3: Contour plots $\rho = 0.5$

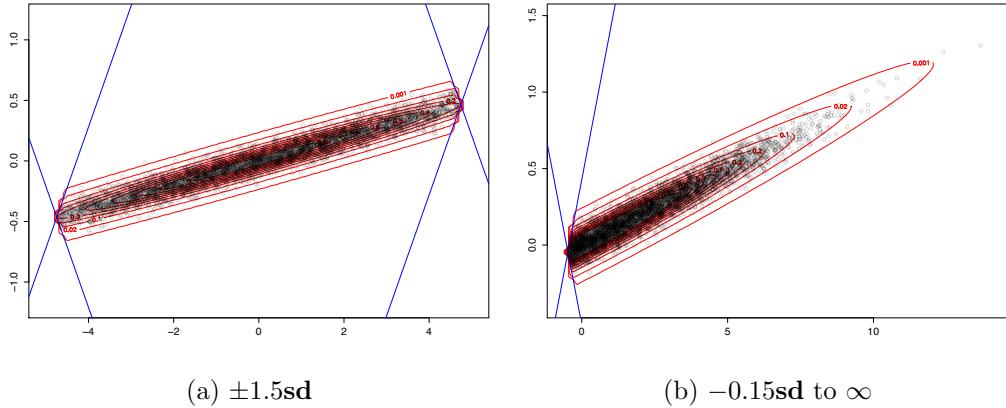


Figure 4: Contour plots $\rho = 0.98$

correlation between the two components, the samples randomly cover the the marginal restriction regions, and the auto-correlations decay to 0 very fast for both bounded and unbounded regions. These illustrate that our Gibbs sampler has very good mixing property and the chains converge fast to the true correct densities. For all cases, we also use the R function `rtmvnrm` in package `tmvtnorm` which is based on the Gibbs sampler proposed by Geweke (1991). However, for most of the cases, especially unbounded restriction regions and wider bounded restriction regions, even when the truncation is very moderate such as $\pm 1.5\text{sd}$, the trace plots and the ACF plots

show that this Gibbs sampler has very poor mixing property and the chains converge very slowly. These plots are displayed in Figures 7 and 8.

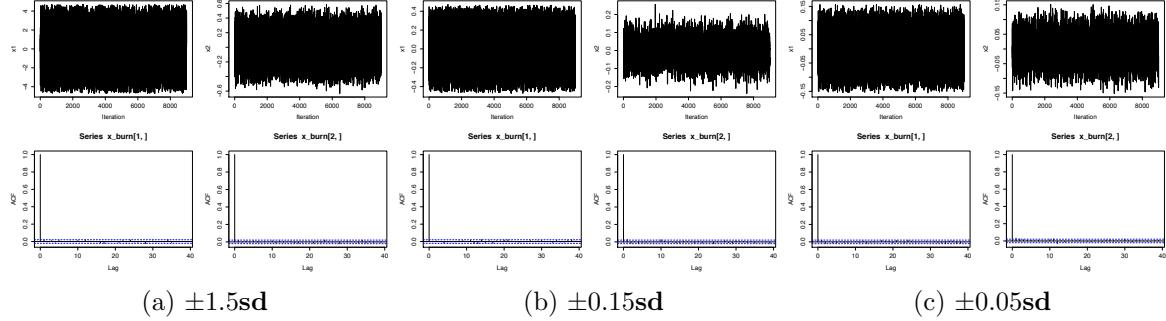


Figure 5: Trace and ACF plots for 1-3 when $\rho = 0.98$

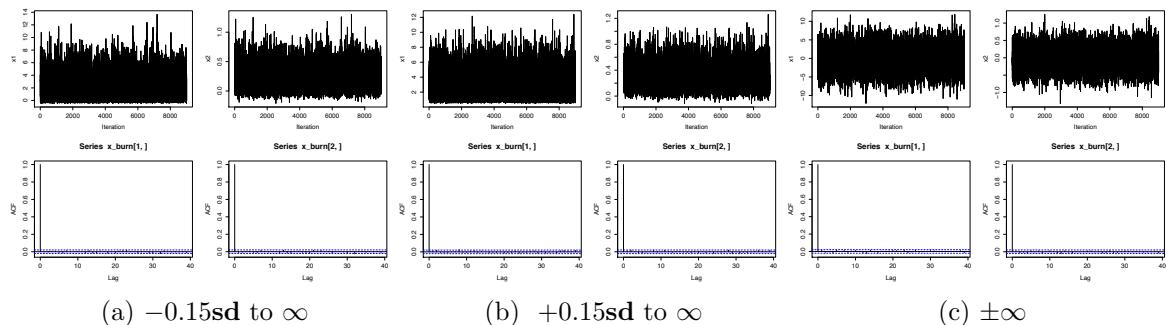


Figure 6: Trace and ACF plots for 4-6 when $\rho = 0.98$

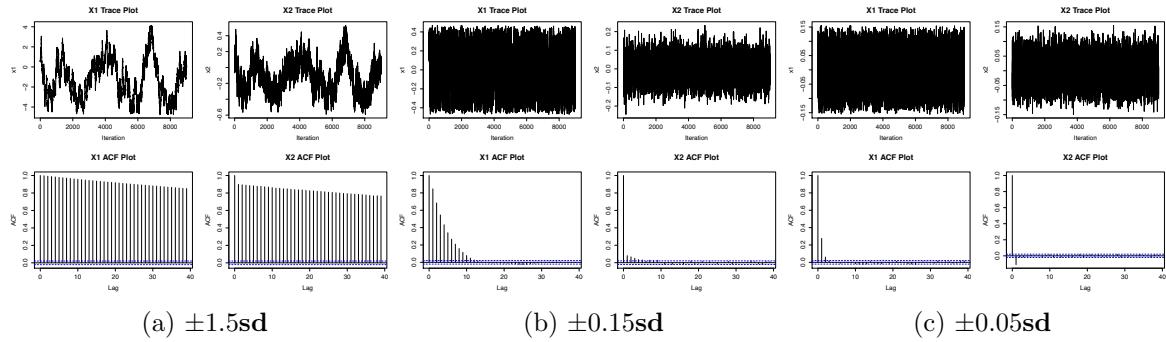


Figure 7: Trace and ACF plots for 1-3 when $\rho = 0.98$ by Geweke (1991)'s

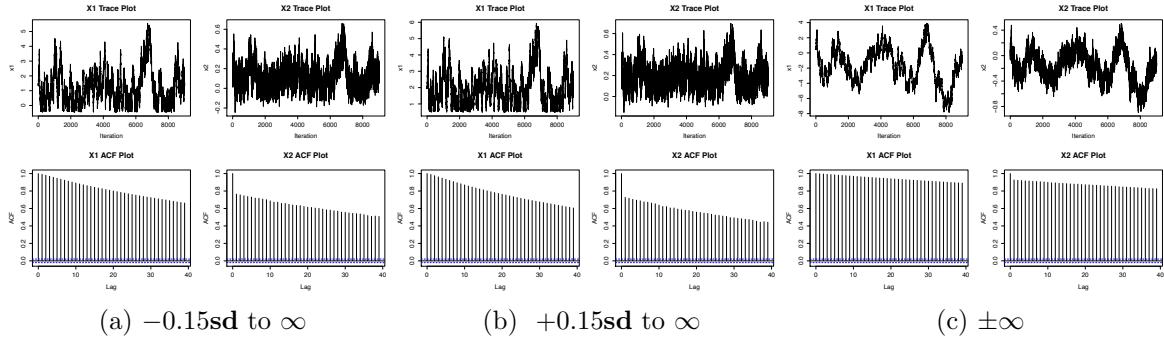


Figure 8: Trace and ACF plots for 4-6 when $\rho = 0.98$ by Geweke (1991)'s

Changing to a lower correlation between the two variables should improve the mixing property and accelerate the convergence even more. However, since our Gibbs sampler already performs extremely well, the improvement is not explicit when $\rho = 0.5$. The trace plots and the ACF plots for some representative cases are given in Figure 9. This improvement is clearer when using Geweke (1991)'s method, whose results are shown in Figure 10. However, although the correlation is lowered to 0.5, the mixing property and the convergence for moderate truncated restriction regions and the one-sided regions are still not very satisfactory as the trace plots are sticky and the ACF's decay to 0 fairly slowly.

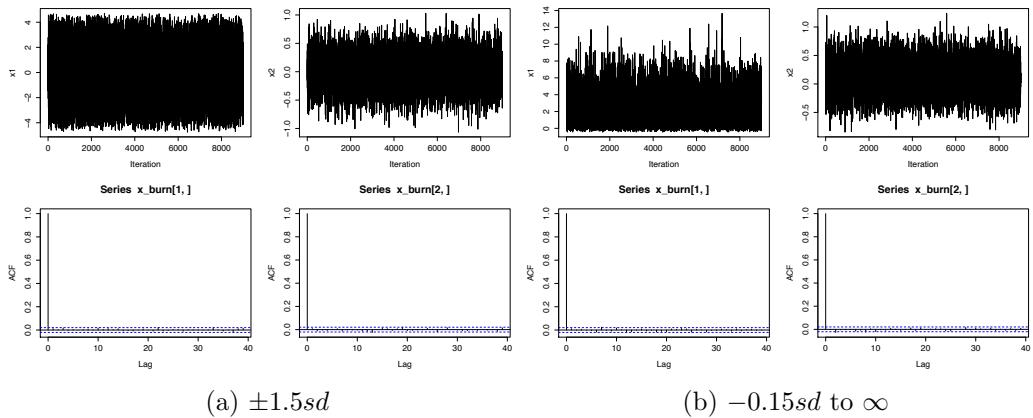


Figure 9: Trace and ACF plots when $\rho = 0.5$

These results strongly illustrate that our Gibbs sampler fixes the poor mixing property and

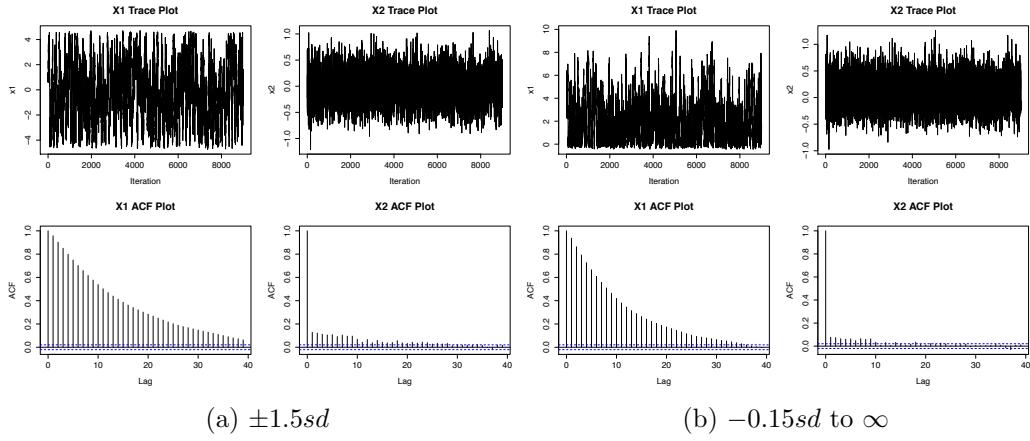


Figure 10: Trace and ACF plots when $\rho = 0.5$ by Geweke (1991)'s

slow convergence caused by Geweke (1991)'s method. For every combination in this example, 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 all of the individual chains pass the tests. It is also worth pointing out that, in our other simulation experiments, we discover that the Gibbs sampler proposed by us has a better mixing property and faster convergence as the restriction regions become wider. However, the above results show that our method still work satisfactorily 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, however, this method becomes problematic for even moderately truncated regions and one-sided restriction regions, which are very commonly encountered in practice.

Example 2. In the description of our method, we claim that the linear constraint matrix $\tilde{\mathbf{R}}$ can have smaller rank than the dimension of the TMVN random vector. In this example, we consider an 3-dimensional TMVN random vector \mathbf{X} with mean and covariance matrix given by

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}.$$

The constraints with two types of restriction regions, bounded and one-sided open, are specified as

$$\tilde{\mathbf{R}} = \begin{pmatrix} 1 & -2 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ and } \mathbf{d} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ or } \begin{pmatrix} \infty \\ \infty \end{pmatrix}.$$

In examining the trace plots and the ACF plots with 10,000 generations and 1,000 burn-ins, for both of the cases, our method has very good mixing property and converges fast to the true density. However, for a 3-dimensional TMVN random vector, it is hard to display samples with the contour plots of the true joint densities, or even the 2-dimensional marginal densities, as the marginal distribution of a TMVN distribution is not TMVN, and the pdf has a very complex expression. Therefore, 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 our generated samples with the marginal densities overlayed are given by Figures 11 and 12 for the bounded region and the one-sided open region, respectively. These figures not only show that our Gibbs sampler gives a correct and accurate sampling generation 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 . [provide the algm in Appendix?]

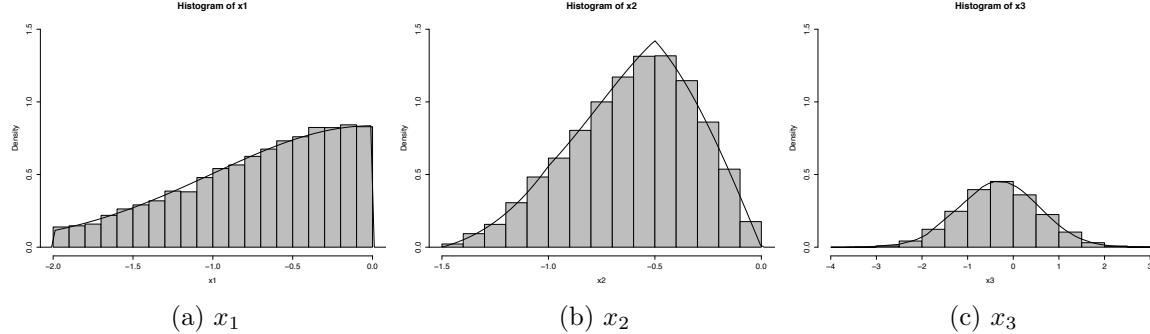


Figure 11: The bounded region

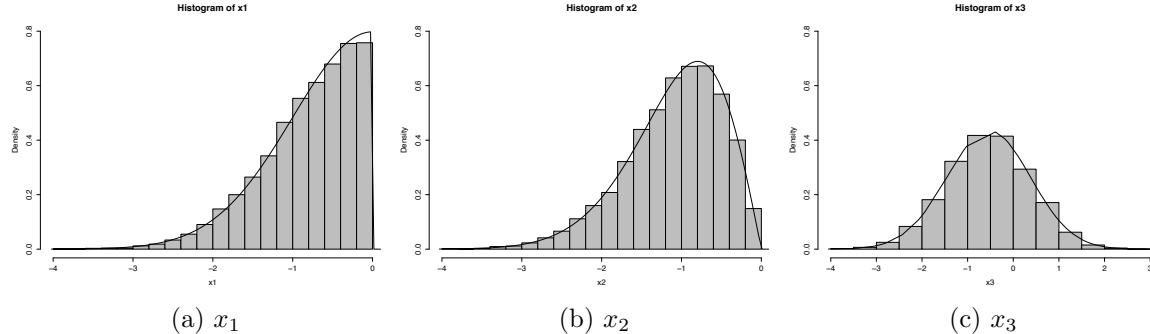


Figure 12: The unbounded region

4.3 Sampling from TMVT

In this sub-section, we illustrate the empirical results of using our proposed Gibbs sampler in sampling from TMVN distributions.

Example 3. We use the same μ , Σ and $\tilde{\mathbf{R}}$ as given in Example 1 for the TMVN distributed random vector. In this example, we only display the results for two types of restriction regions, which are represented by first case in Example 1 with truncation by $\pm 1.5\mathbf{s}\mathbf{d}$, and the fourth case

with the lower bound -0.15sd and the upper bound $+\infty$. We also set the correlation at $\rho = 0.5$. The degrees of freedom are $\nu = 5$. Figure 13 shows the trace and the ACF plots of 10,000 samples with 1,000 burn-ins. These illustrate that our proposed Gibbs sampler for TMVT distributions has good mixing property and fast convergence as the ACF's decay to 0 very fast. The contour plots overlayed with the last 5,000 samples are displayed in Figure 14. The plot for the one-sided open region in Figure 14b is actually truncated as there are a few points generated at very low density area. These clearly demonstrate that our sampling method successfully captures the heavy-tailed nature of the multivariate Student-t distributions. Numerical examinations shows that every individual chain in both of the cases has IACT approximately 1 and passes the Geweke diagnosis tests.

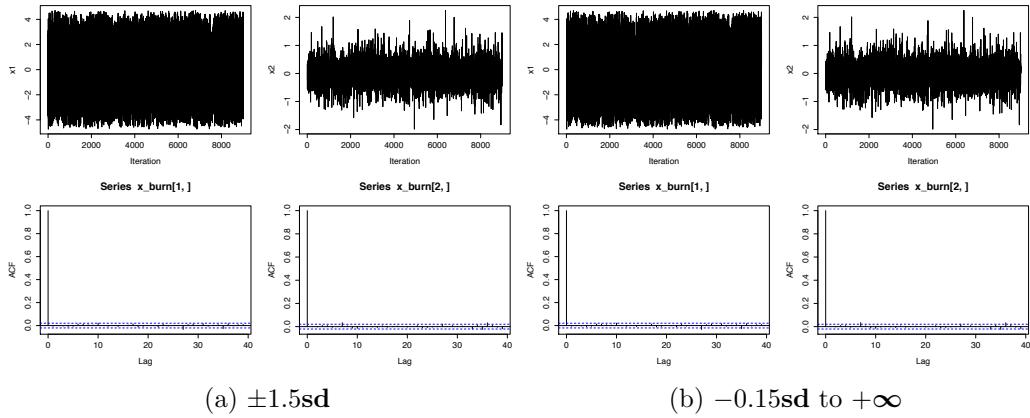


Figure 13: Trace and ACF plots for TMVT $\rho = 0.5 \nu = 5$

5 Conclusions

In this paper, we develop the sampling methods for both TUVN and TMVN distributions, and generalize the methods to TMVT distributions. For TUVN distributions, we establish the mixed rejection sampling method, utilizing four basic rejection sampling methods. For each type of restriction intervals, we compute the analytical acceptance rates for all four rejection

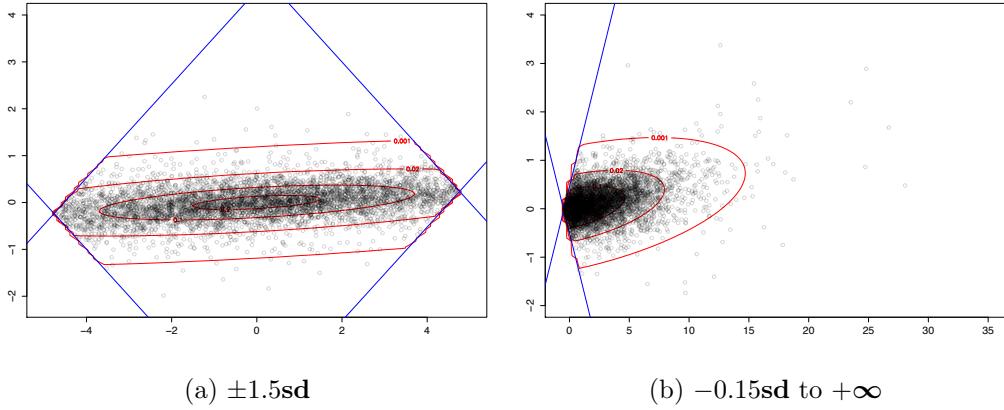


Figure 14: Contour plots for TMVT $\rho = 0.5 \nu = 5$

sampling methods and select the one with the largest acceptance rate. Hence our method has the optimized acceptance rate over the set of all restriction regions. Therefore our method is more efficient than the existing widely used mixed rejection sampling methods by Geweke (1991) and Robert (1995). For the TMVN distributions, we employ the Gibbs sampler taking advantage of the distribution's property that the full conditional distributions are still TMVN. Before implementing the Gibbs sampler, a linear transformation is operated aiming to simplify the functional form to a standard truncated multivariate normal distribution. This transformation also contributes to fix the potential poor mixing property suffered by Geweke (1991)'s method. Simulation studies show that our method has great accuracy, and good mixing property regardless of the type of the restriction regions. To obtain the samples from a general target TMVN distribution, we only need a simple linear transformation after the Gibbs sampling is done. This method is also computationally efficient as it only requires one matrix inversion on the Cholesky decomposition of the covariance matrix. We generalize the method to sampling from TMVN distributions as a multivariate Student-t distribution is a scale mixture density of a multivariate normal density. The sampling method for the TMVT is also a Gibbs sampler with the Gibbs sampler for the TMVN as the central piece of the algorithm, and thus inherits the efficiency

and the accuracy from the sampling method for TMVN distributions. The codes (written in R language) for generating samples from TUVN, TMVN AND TMVT distributions are available from the first author, upon request.

Future work includes generalizing the proposed sampling method for a wider range of scale mixture densities of the multivariate normal distributions. Generalizations to the TMVN distributions with non-linear constraints are also interesting topics for further exploration.

A Appendix

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 $Exp(\lambda; a)$, which is defined as

$$g(x) = \lambda \exp(-\lambda(x - a))I(x \geq a),$$

then

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

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,

$$M_{opt} = \frac{1}{\sqrt{2\pi}\lambda(1 - \Phi(a))} \exp\{\lambda^2 - \lambda a\}, \text{ and } \rho(x) = \exp\{-\frac{1}{2}(x - \lambda)^2\}I(x \geq a).$$

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 the root which is

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

Then the optimized acceptance rate is

$$P_1 \doteq \sqrt{2\pi}\lambda^* \exp\{-\frac{\lambda^{*2}}{2} + \lambda^* a\}\Phi(-a) \tag{26}$$

Notice that $a = \lambda^* - \frac{1}{\lambda^*}$, hence we also have

$$P_1 = \sqrt{2\pi}\lambda^* \exp\left\{-\frac{\lambda^* a}{2} - \frac{1}{2}\right\} \Phi(-a)$$

Scenario 2: If $\lambda < a$, Q is maximized at $x = a$, and

$$Q_{max} = -\frac{a^2}{2}.$$

Accordingly,

$$M_{opt} = \frac{1}{\sqrt{2\pi}\lambda(1 - \Phi(a))} \exp\left\{-\frac{a^2}{2}\right\}, \text{ and } \rho(x) = \exp\left\{-\frac{1}{2}(x - \lambda)^2 + \frac{1}{2}(\lambda - a)^2\right\} I(x \geq a).$$

Hence, the acceptance rate is

$$E[\rho(X)] = \sqrt{2\pi}\lambda \exp\left\{-\frac{a^2}{2}\right\} \Phi(-a)$$

Clearly, the acceptance rate is maximized at $\lambda = a$. Then the optimized acceptance rate is

$$P_2 = \sqrt{2\pi}a \exp\left\{-\frac{a^2}{2}\right\} \Phi(-a). \quad (27)$$

To pick a λ for the rejection sampling, we then compare the two acceptance rates, (26) and (27) resulted by two scenarios. By comparing

$$D = \lambda^* \exp\left\{-\frac{\lambda^{*2}}{2} + \lambda^* a\right\} - a \exp\left\{-\frac{a^2}{2}\right\}$$

with 0, we find that for any value of λ , $D > 0$. Hence (26) is always greater than (27). Therefore, we choose $\lambda = \lambda^*$ and the resulting maximized acceptance rate of the one-sided

translate-exponential rejection sampling method is given as in (26). \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 given in (12) evaluated at $b = \infty$ and (16) respectively, i.e.,

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

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 $0 < a < 0.25696$, $D > 0$, which means that the half-normal rejection sampling has a larger acceptance rate; when $a \geq 0.25696$, $D \leq 0$, 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 (11) and (14). 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 (12) and (13). 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

$$b < \sqrt{\frac{\pi}{2}} e^{\frac{a^2}{2}} + a, \quad (28)$$

we conclude that when (28) 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 given by (13) and (18). This is equivalent to compare

$$\frac{\sqrt{2\pi}}{b-a} e^{\frac{a^2}{2}} \text{ and } \sqrt{2\pi} \lambda^* \exp \left\{ -\frac{\lambda^{*2}}{2} + \lambda^* a \right\}.$$

Since λ^* is defined by (17),

$$\frac{\sqrt{2\pi}}{b-a} e^{\frac{a^2}{2}} < \sqrt{2\pi} \lambda^* \exp \left\{ -\frac{\lambda^{*2}}{2} + \lambda^* a \right\}$$

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. \square

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.