Exact Sampling from the Student's t Model

Dobrin Marchev
Department of Statistics and Computer Information Systems
City University of New York, Baruch College

James P. Hobert
Department of Statistics
University of Florida

October 7, 2008

Abstract

We consider the Bayesian version of the location-scale Student's t model with non-informative prior on the location vector $\boldsymbol{\mu}$ and the scale matrix $\boldsymbol{\Sigma}$. Direct analysis of the resulting posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is challenging and MCMC methods have been developed in the literature to make draws from this intractable density. In this paper we propose a method for obtaining an independent and identically distributed (iid) sample from the target posterior density based on the accept-reject method.

Corresponding author: Dobrin Marchev,

One Bernard Baruch Way, Box B 11-220, New York, NY 10010-5585.

Some key words and phrases: Accept-reject, Student's t model.

1 Introduction

The multivariate Student's t model is an alternative to the normal model whenever the data, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$[1], are believed to come from a heavy tailed distribution. We assume that given $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are conditionally independent with

$$\begin{aligned} \boldsymbol{y}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma} &\overset{\text{iid}}{\sim} t_d(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}),\ i = 1, \ldots, n; \\ \boldsymbol{\mu}, \boldsymbol{\Sigma} &\sim \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \tag{1}$$

where $t_d(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the d-dimensional Student's t distribution with ν degrees of freedom and density

$$f_{t_d}(\boldsymbol{y}; \nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}) := \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \Gamma\left(\frac{d+\nu}{2}\right)}{\pi^{\frac{d}{2}} \Gamma\left(\frac{\nu}{2}\right) \nu^{\frac{d}{2}}} \left[1 + \frac{(\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})}{\nu}\right]^{-\frac{d+\nu}{2}},$$

and $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{d+1}{2}}$ is the standard non-informative prior for a multivariate location-scale problem. The goal is to analyze the posterior given by

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{y}) = \frac{f(\boldsymbol{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\int_S \int_{\mathbb{R}^n} f(\boldsymbol{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\mu} d\boldsymbol{\Sigma}}, \tag{2}$$

where

$$S = \left\{ \boldsymbol{z} \in \mathbb{R}^{\frac{d(d+1)}{2}} : \text{ unvech}(\boldsymbol{z}) \text{ is pos. definite} \right\},$$

unvech$(\boldsymbol{z})$ is the symmetric matrix corresponding to the elements of $\boldsymbol{z}$.

While the Student's t model may be more realistic than the normal model in some cases, there are practical limitations to its use: making inferences about $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is not as straightforward as in the normal case. For example, within the frequentist context, the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are not available in closed form and can only be approximated with the EM algorithm. Similarly, only approximate methods, like the Gibbs sampler, are available to sample from the posterior distribution resulting from Equation (1). We concentrate on the Bayesian

[1]Throughout the paper, we use bold for variables that are vectors or matrices

paradigm and provide an exact method to sample from the posterior distribution associated with Equation (1).

There has been considerable interest in the t model. For example, Geweke (1993) was among the first to study a Bayesian t model with independent errors. He was also among the first to suggest a Gibbs sampler to get draws from the posterior and to actually apply his results to a real data set involving U.S. macroeconomic time series, studied by Nelson and Plosser (1982). More recently, van Dyk and Meng (2001) and Meng and van Dyk (1999) showed how to improve on the standard Gibbs method for sampling from the t posterior, and Marchev and Hobert (2004) proved their algorithm and the standard Gibbs sampler are both geometrically ergodic under mild regularity conditions. However, none of these articles deals with the problem of exact sampling. In this paper we show how the rejection sampler can be applied to obtain exact draws from the target posterior. This is done by utilizing an inequality, which we prove in a separate section, because it is an interesting topic on its own.

2 Making Exact Draws from the Target Posterior

The t model is usually analyzed indirectly using the augmented likelihood, which is based on the so called "missing data" $\boldsymbol{q} := (q_1, \ldots, q_n)^\top \in \mathcal{Q}$, where $\mathcal{Q}$ is usually a subset of some Euclidean space.

More specifically, conditional on $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $(\boldsymbol{y}_i, q_i)$, $i = 1, \ldots, n$, be iid pairs satisfying the relations

$$\begin{aligned} \boldsymbol{y}_i | q_i, \boldsymbol{\mu}, \boldsymbol{\Sigma} &\sim \mathrm{N}_d\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{q_i}\right) \\ q_i | \boldsymbol{\mu}, \boldsymbol{\Sigma} &\sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right) . \end{aligned} \tag{3}$$

Then the implied joint density of $\boldsymbol{y}$ and $\boldsymbol{q}$, given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, is

$$f(\boldsymbol{y}, \boldsymbol{q} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} f(\boldsymbol{y}_i | q_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(q_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

and it is easy to see that

$$\int_{\mathbb{R}_+^n} f(\boldsymbol{y}, \boldsymbol{q}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{q} = f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the same as in Equation (2).

First notice that

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{y}) = \int_{\mathbb{R}_+^n} f(\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{y}) d\boldsymbol{q},$$

where

$$f(\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(\boldsymbol{q}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\int_{\mathcal{Q}} \int_S \int_{\mathbb{R}^n} f(\boldsymbol{y}|\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) f(\boldsymbol{q}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\mu} d\boldsymbol{\Sigma} d\boldsymbol{q}}.$$

Thus, our problem would be solved if we could make draws from $f(\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{y})$ since we could then just ignore the $\boldsymbol{q}$ component. Now, since

$$f(\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{y}) = f(\boldsymbol{q}|\boldsymbol{y}) f(\boldsymbol{\Sigma}|\boldsymbol{q}, \boldsymbol{y}) f(\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{q}, \boldsymbol{y}),$$

we can make draws from $f(\boldsymbol{q}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{y})$ sequentially. Actually the draws from $f(\boldsymbol{\Sigma}|\boldsymbol{q}, \boldsymbol{y})$ and $f(\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{q}, \boldsymbol{y})$ are trivial, since it can be shown that

$$\boldsymbol{\Sigma}|\boldsymbol{q}, \boldsymbol{y} \sim \mathrm{IW}_d \left[n-1, \left(\sum_{j=1}^n q_j (\boldsymbol{y}_j - \hat{\boldsymbol{\mu}})(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}})^\top \right)^{-1} \right],$$

and

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{q}, \boldsymbol{y} \sim \mathrm{N}_d \left(\hat{\boldsymbol{\mu}}, \frac{\boldsymbol{\Sigma}}{q.} \right),$$

where $q. := \sum_{j=1}^n q_j$ and $\hat{\boldsymbol{\mu}} = q.^{-1} \sum_{j=1}^n q_j \boldsymbol{y}_j$, and IW_d is the Inverse Wishart distribution (see; e.g. Mardia, Kent and Bibby (1979)) with density

$$f_{\mathrm{IW}_p}(\boldsymbol{X}; k, \boldsymbol{\Lambda}) = \frac{|\boldsymbol{X}|^{-\frac{k+p+1}{2}} e^{-\frac{\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{X}^{-1})}{2}}}{2^{\frac{kp}{2}} \pi^{\frac{p(p-1)}{4}} |\Lambda|^{\frac{k}{2}} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(k+1-i)\right)}.$$

Unfortunately, drawing from $f(\boldsymbol{q}|\boldsymbol{y})$ is not straightforward. However, as we will show, it is possible to construct rejection samplers for $f(\boldsymbol{q}|\boldsymbol{y})$ that are reasonably efficient when d is small.

Straightforward calculations show that

$$f(\boldsymbol{q}|\boldsymbol{y}) \propto \left(\prod_{j=1}^{n} q_j\right)^{\frac{\nu+d-2}{2}} (q.)^{-\frac{d}{2}}\, e^{-\frac{\nu}{2}q.} \left|\sum_{j=1}^{n} q_j \boldsymbol{y}_j \boldsymbol{y}_j^\top - q.\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\right|^{-\frac{n-1}{2}}. \tag{4}$$

Notice that

$$\sum_{j=1}^{n} q_j \boldsymbol{y}_j \boldsymbol{y}_j^\top - q.\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top = q. \sum_{j=1}^{n} \frac{q_j}{q.}(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}})(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}})^\top =: q.C(\boldsymbol{q}, \boldsymbol{y}),$$

and therefore Equation (4) becomes

$$\begin{aligned} f(\boldsymbol{q}|\boldsymbol{y}) &\propto \left(\prod_{i=1}^{n} q_i\right)^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}\sum_{i=1}^{n} q_i} \frac{\left(\prod_{i=1}^{n} q_i\right)^{\frac{d}{2}}}{q.^{\frac{d}{2}} |q.C(\boldsymbol{q},\boldsymbol{y})|^{\frac{n-1}{2}}} \\ &\propto f_{\Gamma(\frac{\nu}{2},\frac{\nu}{2})}(q_1)\dots f_{\Gamma(\frac{\nu}{2},\frac{\nu}{2})}(q_n) \sqrt{\frac{\left(\prod_{i=1}^{n} \frac{q_i}{q.}\right)^d}{|C(\boldsymbol{q},\boldsymbol{y})|^{n-1}}}. \end{aligned} \tag{5}$$

It is interesting to note here, that when $n = d + 1$ (throughout the paper we assume that $n \geq d + 1$, which is a necessary and sufficient condition for the posterior to be proper), the part of Equation (5) that is under the square root sign is equal to 1, which means that $f(\boldsymbol{q}|\boldsymbol{y})$ is simply a product of iid $\Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ densities and, therefore, can be sampled directly. In the general case, we consider a rejection sampler in which the candidate is exactly this product of Gamma densities. To do this, we need an upper bound B such that

$$\frac{\left(\prod_{i=1}^{n} \frac{q_i}{q.}\right)^d}{|C(\boldsymbol{q},\boldsymbol{y})|^{n-1}} \leq B. \tag{6}$$

Once B is established, we can apply the rejection sampler as follows:

1. Simulate n iid $q_i \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$.
2. Simulate $u \sim \mathrm{U}(0, 1)$.
3. If $\frac{\left(\prod_{i=1}^{n} \frac{q_i}{q.}\right)^d}{|C(\boldsymbol{q},\boldsymbol{y})|^{n-1}} < u^2B^2$, then accept $\boldsymbol{q}$ as a draw from $f(\boldsymbol{q}|\boldsymbol{y})$. Otherwise, go to step 1.

In the next section we show how an upper bound B can be found.

3 An Inequality

We first provide two lemmas that are needed to establish our bound.

Lemma 1. *Suppose that $\boldsymbol{Y}$ is a discrete random vector putting mass p_i at $\boldsymbol{y}_i$, $i = 1, 2, \ldots, n$. Then*

$$cov(\boldsymbol{Y}) = \sum_{1 \leq i < j \leq n} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top.$$

Proof.

This follows straightforwardly from the fact that the covariance matrix of Y can be represented as $(1/2)E(X - Y)(X - Y)^T$ where X is an iid copy of Y.

■

Before we proceed any further, we need to develop some notation. For $m \leq n$, let C_m^n be the set of all possible subsets of size m from $\mathbb{N}_n := \{1, 2, \ldots, n\}$; that is,

$$C_m^n := \{(k_1, \ldots, k_m) : k_1 < \cdots < k_m \text{ and } k_i \in \mathbb{N}_n, i = 1, \ldots, m\}.$$

For example, $C_2^3 := \{(1, 2), (1, 3), (2, 3)\}$. Also, for a given matrix $\boldsymbol{A}$ of dimension $m \times n$, let $\boldsymbol{A}_{r(T)}$ denote the submatrix obtained from $\boldsymbol{A}$ using only the rows that are in the index set T (of course, here $T \subset \mathbb{N}_m$). Similarly, let $\boldsymbol{A}_{c(T)}$ be the submatrix obtained from the columns of $\boldsymbol{A}$ that are in the set T. We now state the Binet-Cauchy formula for the determinant of a product of two matrices. It can be found, for example, in Karlin (1968).

Lemma 2 (Binet-Cauchy Formula)**.** *Let $\boldsymbol{A}$ be an $m \times n$ matrix, and let $\boldsymbol{B}$ be an $n \times m$ matrix. Then*

$$|\boldsymbol{AB}| = \sum_{T \in C_m^n} \left|\boldsymbol{A}_{c(T)}\right| \left|\boldsymbol{B}_{r(T)}\right|.$$

Let's now go back to the question of how to bound the expression in Equation (6).

Theorem 1. *For $d \in \mathbb{N}$, $n \geq d+1$, $\boldsymbol{y} \in \mathbb{R}^{dn}$ and $\boldsymbol{q} \in \mathbb{R}^n_+$,*

$$\frac{\left(\prod_{i=1}^n \frac{q_i}{q.}\right)^d}{|C(\boldsymbol{q}, \boldsymbol{y})|^{n-1}} \leq \frac{1}{\binom{n-1}{d}^{n-1} d(\boldsymbol{y})}, \tag{7}$$

where

$$d(\boldsymbol{y}) = \begin{cases} \frac{1}{(y_1-y_2)^2} & \text{if } n = 2 \\ \left(\frac{1}{c_1(\boldsymbol{y})^{\frac{1}{d(n-2)}}} + \cdots + \frac{1}{c_n(\boldsymbol{y})^{\frac{1}{d(n-2)}}}\right)^{-d(n-2)} & \text{if } n > 2, \end{cases}$$

$$c_i(\boldsymbol{y}) = \prod_{T \in C_d^{n-1}} \left|\boldsymbol{M}^{(i)}_{r(T)}\right|^{\frac{2d!(n-d-1)!}{(n-2)!}},$$

and $\boldsymbol{M}^{(i)}$ is an $(n-1) \times d$ matrix with rows $(\boldsymbol{y}_i - \boldsymbol{y}_1)^\top, \ldots, (\boldsymbol{y}_i - \boldsymbol{y}_{i-1})^\top, (\boldsymbol{y}_i - \boldsymbol{y}_{i+1})^\top, \ldots, (\boldsymbol{y}_i - \boldsymbol{y}_n)^\top$, for $i = 1, \ldots, n$.

Proof. First notice that when $n = 2$, d must be 1, and in this case both sides of (7) are equal to $\frac{1}{(y_1-y_2)^2}$. For the nontrivial case when $n > 2$, denote $p_i = \frac{q_i}{q.}, i = 1, \ldots, n$ and notice that because of Lemma 1

$$\sup_{\boldsymbol{q} \in \mathbb{R}^n_+} \frac{\left(\prod_{i=1}^n \frac{q_i}{q.}\right)^d}{|C(\boldsymbol{q}, \boldsymbol{y})|^{n-1}} = \sup_{\boldsymbol{p} \in \mathcal{P}} \frac{\left(\prod_{i=1}^n p_i\right)^d}{\left|\sum_{i<j} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top\right|^{n-1}},$$

where $\mathcal{P} = \{(p_1, \ldots, p_n) \in (0,1)^n : \sum_{i=1}^n p_i = 1\}$.

What follows is true for any $p_k, k = 1, \ldots, n$; but for ease of exposition, we'll show the details only for p_1. First, notice that in the sense of matrix ordering[2],

$$\begin{aligned} \sum_{i<j} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top &\geq \sum_{i=2}^n p_1 p_i (\boldsymbol{y}_1 - \boldsymbol{y}_i)(\boldsymbol{y}_1 - \boldsymbol{y}_i)^\top \\ &= p_1 \boldsymbol{M}^{(1)^\top} \boldsymbol{P}_* \boldsymbol{M}^{(1)}, \end{aligned} \tag{8}$$

where $\boldsymbol{p}_* = (p_2, p_3, \vdots, p_n)^\top$ and $\boldsymbol{P}_*$ is a diagonal matrix with $\boldsymbol{p}_*$ on the diagonal. Now applying the Binet-Cauchy formula with $\boldsymbol{A} = \boldsymbol{M}^{(1)^\top}$ and $\boldsymbol{B} = \boldsymbol{P}_* \boldsymbol{M}^{(1)}$, we obtain

$$\left|\boldsymbol{M}^{(1)^\top} \boldsymbol{P}_* \boldsymbol{M}^{(1)}\right| = \sum_{T \in C_d^{n-1}} \left|\left(\boldsymbol{M}^{(1)^\top}\right)_{c(T)}\right| \left|\left(\boldsymbol{P}_* \boldsymbol{M}^{(1)}\right)_{r(T)}\right|. \tag{9}$$

[2]For nonnegative definite $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \geq \boldsymbol{B}$ iff $\boldsymbol{A} - \boldsymbol{B}$ is nonnegative definite

Since

$$\boldsymbol{P}_*\boldsymbol{M}^{(1)} = \begin{pmatrix} p_2(\boldsymbol{y}_1 - \boldsymbol{y}_2)^\top \\ \cdots \\ p_n(\boldsymbol{y}_1 - \boldsymbol{y}_n)^\top \end{pmatrix},$$

for any $T \in C_d^{n-1}$, we can write

$$\left|\left(\boldsymbol{P}_*\boldsymbol{M}^{(1)}\right)_{r(T)}\right| = \left(\prod_{i\in T} p_{i+1}\right)\left|\boldsymbol{M}^{(1)}_{r(T)}\right|.$$

Thus,

$$\left|\boldsymbol{M}^{(1)\top}\boldsymbol{P}_*\boldsymbol{M}^{(1)}\right| = \sum_{T\in C_d^{n-1}}\left(\prod_{i\in T} p_{i+1}\right)\left|\boldsymbol{M}^{(1)}_{r(T)}\right|^2,$$

and consequently from Equations (8), (9) and the fact that if $\boldsymbol{A} \geq \boldsymbol{B}$ then $|\boldsymbol{A}| \geq |\boldsymbol{B}|$, it follows that

$$\left|\sum_{i<j} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top\right| \geq p_1^d \sum_{T\in C_d^{n-1}}\left(\prod_{i\in T} p_{i+1}\right)\left|\boldsymbol{M}^{(1)}_{r(T)}\right|^2. \tag{10}$$

Now by the inequality between the arithmetic and geometric means, we obtain

$$\begin{aligned} &\sum_{T\in C_d^{n-1}}\left(\prod_{i\in T} p_{i+1}\right)\left|\boldsymbol{M}^{(1)}_{r(T)}\right|^2 \\ &\qquad \geq \binom{n-1}{d}\left[\prod_{T\in C_d^{n-1}}\left(\prod_{i\in T} p_{i+1}\right)\left|\boldsymbol{M}^{(1)}_{r(T)}\right|^2\right]^{\binom{n-1}{d}^{-1}}. \end{aligned} \tag{11}$$

Notice that each index $i \in \{1, \ldots, n\}$ is contained in exactly $\binom{n-2}{d-1}$ elements of the set C_d^{n-1}. Therefore,

$$\prod_{T\in C_d^{n-1}}\prod_{i\in T} p_{i+1} = (p_2 \ldots p_n)^{\binom{n-2}{d-1}},$$

implying that

$$\begin{aligned} &\left[\prod_{T\in C_d^{n-1}}\left(\prod_{i\in T} p_{i+1}\right)\left|\boldsymbol{M}^{(1)}_{r(T)}\right|^2\right]^{\binom{n-1}{d}^{-1}} \\ &\qquad = (p_2 \ldots p_n)^{\frac{d}{n-1}}\left(\prod_{T\in C_d^{n-1}}\left|\boldsymbol{M}^{(1)}_{r(T)}\right|^2\right)^{\binom{n-1}{d}^{-1}}. \end{aligned} \tag{12}$$

From Equations (10), (11) and (12) we obtain that

$$\left|\sum_{i<j} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top\right|^{n-1} \geq \binom{n-1}{d}^{n-1} p_1^{d(n-1)} (p_2 \dots p_n)^d c_1(\boldsymbol{y}),$$

and consequently

$$\frac{\left(\prod_{i=1}^n p_i\right)^d}{\left|\sum_{i<j} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top\right|^{n-1}} \leq \frac{1}{\binom{n-1}{d}^{n-1} p_1^{d(n-2)} c_1(\boldsymbol{y})}.$$

By symmetry, for any $k \in \mathbb{N}_n$,

$$\frac{\left(\prod_{i=1}^n p_i\right)^d}{\left[\sum_{i<j} p_i p_j (\boldsymbol{y}_i - \boldsymbol{y}_j)(\boldsymbol{y}_i - \boldsymbol{y}_j)^\top\right]^{n-1}} \leq \min_{1 \leq k \leq n} \frac{1}{\binom{n-1}{d}^{n-1} p_k^{d(n-2)} c_k(\boldsymbol{y})},$$

and consequently,

$$\sup_{\boldsymbol{q} \in \mathbb{R}_+^n} \frac{\left(\prod_{i=1}^n \frac{q_i}{q.}\right)^d}{|C(\boldsymbol{q}, \boldsymbol{y})|^{n-1}} \leq \sup_{\boldsymbol{p} \in \mathcal{P}} \min_{1 \leq k \leq n} \frac{1}{\binom{n-1}{d}^{n-1} p_k^{d(n-2)} c_k(\boldsymbol{y})}.$$

Finally, we will show that

$$\sup_{\boldsymbol{p} \in \mathcal{P}} \min_{1 \leq i \leq n} \frac{1}{p_i^{d(n-2)} c_i(\boldsymbol{y})} = \frac{1}{d(\boldsymbol{y})}. \tag{13}$$

Let $b_i = c_i(\boldsymbol{y})^{-\frac{1}{d(n-2)}}$; $\quad b. = \sum_{i=1}^n b_i$; $\quad p_i^* = \frac{b_i}{b.}$. Then

$$\begin{aligned}
\sup_{\boldsymbol{p} \in \mathcal{P}} \min_{1 \leq i \leq n} \frac{1}{p_i^{d(n-2)} c_i(\boldsymbol{y})} &= \left[\inf_{\boldsymbol{p} \in \mathcal{P}} \max_{1 \leq i \leq n} p_i c_i(\boldsymbol{y})^{\frac{1}{d(n-2)}}\right]^{-d(n-2)} \\
&= b.^{d(n-2)} \left(\inf_{\boldsymbol{p} \in \mathcal{P}} \max_{1 \leq i \leq n} \frac{p_i}{p_i^*}\right)^{-d(n-2)}.
\end{aligned}$$

Assuming $p_i^* > 0 \ \forall \ i \in \mathbb{N}_n$, it is clear that if $\boldsymbol{p} = \boldsymbol{p}^*$, then $\max_{1 \leq i \leq n} \frac{p_i}{p_i^*} = 1$, whereas for any other choice of $\boldsymbol{p}$, $\max_{1 \leq i \leq n} \frac{p_i}{p_i^*} > 1$, since if $p_i \neq p_i^*$ for some i, then there exists a $j \in \mathbb{N}_n$ such that $p_j > p_j^*$. ■

In the univariate case ($d = 1$), the constants $c_i(\boldsymbol{y})$ can be simplified considerably. We state this simplification of Theorem 1 as a corollary.

Corollary 1. *For $n \geq 2$, $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{q} \in \mathbb{R}^n_+$,*

$$\frac{\prod_{i=1}^{n} \frac{q_i}{q.}}{[v(\boldsymbol{q}, \boldsymbol{y})]^{n-1}} \leq \left[(n-1)^{n-1} d_1(\boldsymbol{y})\right]^{-1},$$

where

$$v(\boldsymbol{q}, \boldsymbol{y}) = \frac{1}{q.} \sum_{i=1}^{n} q_i (y_i - \hat{\mu})^2, \quad d_1(\boldsymbol{y}) = \left[\frac{1}{e_1(\boldsymbol{y})^{\frac{1}{n-2}}} + \cdots + \frac{1}{e_n(\boldsymbol{y})^{\frac{1}{n-2}}}\right]^{-(n-2)},$$

and $e_i(\boldsymbol{y}) = \prod_{j \neq i} (y_i - y_j)^2, \ i = 1, \ldots, n.$

Proof. First notice that $C_1^{n-1} := \{1, \ldots, n-1\}$ and therefore

$$\prod_{j \in C_1^{n-1}} \left|\boldsymbol{M}_{r(j)}^{(i)}\right| = \prod_{j \neq i} (y_i - y_j)^2.$$

The rest of the proof is just substitution. ■

The bound proved in Theorem 1 is not sharp, which reduces the acceptance rate of the resulting rejection sampler. Unfortunately, we were unable to get a better bound in the general case, but we can state a conjecture about the exact supremum of (6) when $d = 1$, which is presented next.

Conjecture 1. *For $n \geq 2, \boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{q} \in \mathbb{R}^n_+$,*

$$\sup_{\boldsymbol{q} \in \mathbb{R}^n_+} \frac{\prod_{i=1}^{n} \left(\frac{q_i}{q.}\right)}{[v(\boldsymbol{q}, \boldsymbol{y})]^{n-1}} = \left[(n-1)^{n-1} c(\boldsymbol{y})\right]^{-1}, \tag{14}$$

where $c(\boldsymbol{y}) = \min_{1 \leq i \leq n} \prod_{j \neq i} (y_i - y_j)^2.$

Proof. We were able to prove the conjecture only when $n = 2$ or $n = 3$ and the proof is given in the appendix. ■

4 Some Simulation Results

Clearly, the bounds established in the previous section depend on $\boldsymbol{y}, n$ and d. Furthermore, the candidate density depends on ν. It is, therefore, interesting to investigate the behavior of the algorithm when changing the values of these parameters. Generally speaking, the results of our computer simulations show that the acceptance rate of the algorithm varies greatly depending on the specific data set, $\boldsymbol{y}$, and suggest that it works better if either n is small and/or ν is big. Also, the acceptance rate tends to be higher when the y_is are equally spaced. Here are some specific examples.

Example 2.1: Suppose a particular set of y_is is given by[3]

$$\boldsymbol{y} = (-1.449605, -0.996631, 0.228872, 0.068414, -0.126978, -0.563358, 0.766889)^\top$$

and is assumed to have t_1 distribution with $\nu = 5$. A computer simulation using the bound in Corollary 1 resulted in an Accept - Reject algorithm with acceptance rate of 0.00131 estimated with a standard error of 0.00011 (using the well-known formula for the standard deviation of one observation from the Binomial distribution). On the other hand, using the bound from the conjecture on the same data, results in acceptance rate of 0.0384 estimated with a standard error of 0.00061. Assuming that $\nu = 10$, changes the acceptance rates for the proved and unproved bound to 0.00139 (0.00012) and 0.04141 (0.00063), respectively (the numbers in parenthesis are the standard errors).

Increasing the dimensionality of the data quickly reduces the acceptance rate of the algorithm, which is true for any rejection sampler. For example, even with $d = 3$, the acceptance rate was about 7-8 out of 10,000,000. However, keep in mind that simulating 10,000,000 iterations is a matter of a few minutes of computational time, and therefore obtaining an exact sample from the t model is doable for reasonbly small d.

[3]The actual values were obtained by simulation

5 Concluding Remarks

Exact sampling from the Student's t model is a viable alternative to the approximate draws via Gibbs samplers. In this paper, we show that obtaining an iid sample from the multivariate t model is possible by applying a rejection sampler. This is done by proving an interesting inequality. The efficiency of our sampler can be approved siginificantly if an alternative bound is used.

Appendix

Here we prove our conjecture for the cases $n = 2$ and $n = 3$.

Notice that when $n = 2, v(\boldsymbol{q}, \boldsymbol{y}) = p_1 p_2 (y_1 - y_2)^2, c(\boldsymbol{y}) = (y_1 - y_2)^2$ and consequently (14) is obvious. The non-trivial case is when $n = 3$.

We will first prove that

$$\frac{\prod_{i=1}^{n} \left(\frac{q_i}{q_\cdot}\right)}{[v(\boldsymbol{q}, \boldsymbol{y})]^{n-1}} \leq \left[(n-1)^{n-1} c(\boldsymbol{y})\right]^{-1} \tag{15}$$

and then we will establish the attainment of the sup. Suppose that

$$\min_{1 \leq i \leq 3} \prod_{j \neq i} (y_i - y_j)^2 = \prod_{j \neq 1} (y_1 - y_j)^2;$$

i.e., the minimum is achieved at $i = 1$. This simply means that y_1 is between y_2 and y_3. Without loss of generality, we will assume that the ordering of y_is is $y_2 \leq y_1 \leq y_3$. Notice that (15) is now equivalent to

$$\sqrt{\frac{p_1 p_2}{p_3}} (y_1 - y_2)^2 + \sqrt{\frac{p_1 p_3}{p_2}} (y_1 - y_3)^2 + \sqrt{\frac{p_2 p_3}{p_1}} (y_2 - y_3)^2 \geq 2(y_1 - y_2)(y_3 - y_1). \tag{16}$$

But $(y_2 - y_3)^2 = (y_1 - y_2)^2 + (y_3 - y_1)^2 + 2(y_3 - y_1)(y_1 - y_2)$, and because of the assumed ordering, (16) is equivalent to

$$\frac{1}{2} \frac{y_1 - y_2}{y_3 - y_1} \left(\sqrt{\frac{p_1 p_2}{p_3}} + \sqrt{\frac{p_2 p_3}{p_1}} \right) + \frac{1}{2} \frac{y_3 - y_1}{y_1 - y_2} \left(\sqrt{\frac{p_1 p_3}{p_2}} + \sqrt{\frac{p_2 p_3}{p_1}} \right) + \sqrt{\frac{p_2 p_3}{p_1}} \geq 1.$$

Now an application of the inequality between the arithmetic and geometric mean gives us

$$\frac{1}{2}\frac{y_1-y_2}{y_3-y_1}\left(\sqrt{\frac{p_1p_2}{p_3}}+\sqrt{\frac{p_2p_3}{p_1}}\right)+\frac{1}{2}\frac{y_3-y_1}{y_1-y_2}\left(\sqrt{\frac{p_1p_3}{p_2}}+\sqrt{\frac{p_2p_3}{p_1}}\right)$$
$$\geq\sqrt{\left(\sqrt{\frac{p_1p_2}{p_3}}+\sqrt{\frac{p_2p_3}{p_1}}\right)\left(\sqrt{\frac{p_1p_3}{p_2}}+\sqrt{\frac{p_2p_3}{p_1}}\right)}=\sqrt{1+\frac{p_2p_3}{p_1}}.$$

Consequently, (15) will be established, if we can show that

$$\sqrt{1+\frac{p_2p_3}{p_1}}+\sqrt{\frac{p_2p_3}{p_1}}\geq 1,$$

which is obviously true.

Notice that the last inequality becomes equality when $p_1 \to 1$ and $p_{2,3} \to 0$ with the rest of the inequalities suggesting that $\frac{p_2}{p_3}=\frac{(y_3-y_1)^2}{(y_1-y_2)^2}$. It is easy to verify that in terms of q_is this is satisfied when, for example, we take

$$q_1=t,\ q_2=\frac{1}{t}\text{ and }q_3=\frac{(y_1-y_2)^2}{t(y_3-y_1)^2}$$

and let $t\to\infty$. Then

$$\frac{q_1}{q.}=\frac{t}{t+\frac{1}{t}+\frac{(y_1-y_2)^2}{t(y_3-y_1)^2}}\to 1;\quad \frac{q_2q_3}{q_1q.}=\frac{\frac{(y_1-y_2)^2}{t^2(y_3-y_1)^2}}{t\left[t+\frac{1}{t}+\frac{(y_1-y_2)^2}{t(y_3-y_1)^2}\right]}\to 0,$$

as $t\to\infty$, implying that

$$\lim_{t\to\infty}\left[\frac{1}{2}\frac{y_1-y_2}{y_3-y_1}\left(\sqrt{\frac{p_1p_2}{p_3}}+\sqrt{\frac{p_2p_3}{p_1}}\right)+\frac{1}{2}\frac{y_3-y_1}{y_1-y_2}\left(\sqrt{\frac{p_1p_3}{p_2}}+\sqrt{\frac{p_2p_3}{p_1}}\right)+\sqrt{\frac{p_2p_3}{p_1}}\right]=1.$$

Acknowledgement: The authors are gratefull to Brett Presnell for his helpful comments.

References

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model, *Journal of Applied Econometrics* **8**: S19–S40.

Karlin, S. (1968). *Total Positivity*, Vol. 1, Stanford University Press, Stanford, California.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van dyk and meng's algorithm for the multivariate student's t model, *Journal of the American Statistical Association* **99**: 228–238.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press, London.

Meng, X.-L. and van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation, *Biometrika* **86**: 301–320.

Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics* **10**: 139–162.

van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation (with discussion), *Journal of Computational and Graphical Statistics* **10**: 1–50.