Journal of the Korean Statistical Society (2005), **34**: *3, pp 245–261*

ON BAYESIAN ESTIMATION AND PROPERTIES OF THE MARGINAL DISTRIBUTION OF A TRUNCATED BIVARIATE t-DISTRIBUTION

HEA-JUNG KIM[1] AND JU SUNG KIM[1]

ABSTRACT

The marginal distribution of X is considered when (X, Y) has a truncated bivariate t-distribution. This paper mainly focuses on the marginal nontruncated distribution of X where Y is truncated below at its mean and its observations are not available. Several properties and applications of this distribution, including relationship with Azzalini's skew-normal distribution, are obtained. To circumvent inferential problem arises from adopting the frequentist's approach, a Bayesian method utilizing a data augmentation method is suggested. Illustrative examples demonstrate the performance of the method.

AMS 2000 subject classifications. Primary 62F15; Secondary 62E15.
Keywords. Truncated bivariate t, Skew-t distribution, Bayesian method, Markov chain Monte Carlo method.

1. INTRODUCTION

A random variable is said to be skew-normal with parameter θ, written $Z \sim SN(\theta)$, if its density function is

$$2\phi(z)\Phi(\theta z), \quad -\infty < z < \infty, \tag{1.1}$$

where $\phi(z)$ and $\Phi(z)$ denote the standard normal density and distribution function, respectively. The parameter $\theta \in (-\infty, \infty)$ regulates the skewness, and $\theta = 0$ corresponds to the standard normal density. A systematic treatment of the distribution has been independently given by Azzalini (1985) and Henze (1986) and extensions of the distribution are considered by Azzalini and Dalla Valle (1996),

Received June 2005; accepted Septamber 2005.
[1]Department of Statistics, Dongguk University, Seoul 100-715, Korea
(e-mail: kim3hj@dongguk.edu)

Branco and Dey (2001), Azzalini and Capitanio (2003), and Ma and Genton (2004). The distribution is suitable for the analysis of data exibiting a unimodal empirical distribution but with some skewness present, a situation often occurring in practical problems. See Azzalini (1985) and Arnold et al. (1993) for some applications of the distribution.

Kim (2002) modified the distribution to obtain a skew-t distribution, $St(\theta, \nu)$, whose density function is

$$2f_\nu(t)F_{\nu+1}\left(\frac{\theta t\sqrt{\nu+1}}{\sqrt{\nu+t^2}}\right), \quad -\infty < t < \infty, \tag{1.2}$$

where $f_\nu(\cdot)$ and $F_{\nu+1}(\cdot)$ are the standard t_ν density and $t_{\nu+1}$ distribution function, respectively. It is seen that $St(\theta, \nu)$ distribution leads to a parametric class of distributions that have the properties; (i) strict inclusion of t_ν distribution, (ii) mathematical tractability, (iii) wide range of the indices of skewness. An application of $St(\theta, \nu)$ distribution in fitting the binary regression model is given by Kim(2002).

The purpose of the present paper is to introduce further properties of $St(\theta, \nu)$ distribution, some relations with a bivariate t-distribution. Such properties are potentially relevant for practical applications, since in data analysis there are a few parametric distributions available to dealing with both symmetric and skewed data, especially for the problem of fitting data from a screening process. See Arnold et al.(1993, 2002) and Cohen(1991) for reviews of the literature in the screening process. Another is to develop a Bayesian estimation of the generalized $St(\theta, \nu)$ distribution, utilizing a data augmentation method, in order to circumvent inferential problem arises from adopting the frequentist's approach. Furthermore, this paper gives illustrative examples to demonstrate the utility of the suggested properties of $St(\theta, \nu)$ distribution.

2. The Properties

Suppose that $h_\nu(x, y)$ is the density of $t_\nu(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ distribution, i.e. a bivariate t distribution with degrees of freedom ν, mean vector (μ_1, μ_2), scale vector (σ_1^2, σ_2^2), and correlation ρ, and suppose the joint density of (X, Y) is

$$h_{X,Y}(x, y) = 2h_\nu(x, y), \quad -\infty < x < \infty, \; y > \mu_2. \tag{2.1}$$

Clearly, (X, Y) has a truncated bivariate t_ν distribution so that Y is truncated below at its expectation. In this paper we are concerned with the marginal distribution of X, the untruncated variable. By direct integration one obtains

$$h_X(x) = 2/\sigma_1 f_\nu(t) F_{\nu+1}\left(\frac{\theta t\sqrt{\nu+1}}{\sqrt{\nu+t^2}}\right), \quad -\infty < x < \infty, \tag{2.2}$$

where $t = (x - \mu_1)/\sigma_1$ and $\theta = \rho/\sqrt{1-\rho^2}$, respectively. From now on, we will denote the marginal distribution of X with the density (2.2) as $St(\theta, \nu, \mu_1, \sigma_1)$.

LEMMA 2.1. *Let (W_1, W_2) be a $t_\nu(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ random variable, and let $T_i = (W_i - \mu_1)/\sigma_i$ for $i = 1, 2$ so that (T_1, T_2) is a $t_\nu(0, 0, 1, 1, \rho)$ random variable. Then the conditional distribution of T_2 given that $T_1 = t_1$ is $t_{\nu+1}(\alpha, \beta)$ distribution, where $\alpha = \rho t_1$, $\beta = \sqrt{(\nu + t_1^2)(1-\rho^2)/(\nu+1)}$ and $t_{\nu+1}(\alpha, \beta)$ denotes a univariate $t_{\nu+1}$ distribution with the mean α and the scale parameter β^2.*

PROOF. Since the marginal distribution of T_1 is $t_\nu \equiv t_\nu(0, 1)$, complicated but straightforward derivation of the conditional density of T_2 leads to

$$f(t_2|t_1) = \frac{(\nu+1)^{(\nu+1)/2}}{\beta B(1/2, (\nu+1)/2]}\left\{(\nu+1) + \left(\frac{t_2-\alpha}{\beta}\right)^2\right\}^{-(\nu+2)/2}, \quad -\infty < t_2 < \infty \tag{2.3}$$

for $-\infty < t_1 < \infty$, where $B[\cdot, \cdot]$ is the beta function. (2.3) is the density of $t_{\nu+1}(\alpha, \beta)$ distribution, and hence the result. □

Applying the relations $T_i = (W_i - \mu_i)/\sigma_i$ to (2.3), we can immediately obtain the following conditional distribution.

THEOREM 2.1. *If (W_1, W_2) is a $t_\nu(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ random variable, and X is set to equal to W_1 conditionally on $W_2 > \mu_2$. Then $X \sim St(\theta, \nu, \mu_1, \sigma_1)$.*

PROOF. The conditional density of W_1 given that $W_2 > \mu_2$ is obtained from that of T_1 given that $T_2 > 0$, where the variables T_1 and T_2 are the same ones defined in Lemma 2.1. Using (2.3), one can obtain the conditional density of T_1 given that $T_2 > 0$ from $2\int_0^\infty f(t_2|t_1) f_\nu(t_1) dt_2$, where $f_\nu(t_1)$ is the t_ν density. Straightforward integration with respect to t_2, and then transforming t_1 to w_1 and setting w_1 to x gives (2.2). □

COROLLARY 2.1. *For $T \sim St(\theta, \nu)$, the distribution of $\sigma_1 T + \mu_1$ is $St(\theta, \nu, \mu_1, \sigma_1)$.*

PROOF. Considering the transformation $X = \sigma_1 T + \mu_1$ with density of T in (1.2) gives the result. □

This representation of $St(\theta, \nu)$ distribution is interesting since it links the distribution to a censoring operation on t variates, a situation naturally occurring in a large number of practical cases. A similar conditioning mechanism can be used to obtain the distribution as a prior distribution for the mean of a t variable, in a Bayesian framework. We see, from Theorem 2.1 and Corollary 2.1, that a random variable X with distribution $St(\theta, \nu, \mu_1, \sigma_1)$ can be generated by the following acceptance-rejection method. Sample a pair (T_1, T_2) from $t_\nu(0, 0, 1, 1, \rho)$ distribution. If $T_2 > 0$, then put $X = \sigma_1 T_1 + \mu_1$, otherwise restart sampling a new pair of variables (T_1, T_2), until the condition $T_2 > 0$ is satisfied. On average two pairs (T_1, T_2) are necessary to produce X. Theorem 2.1 and Corollary 2.1 also lead to following properties:

PROPERTY 1. For $\mu_1 = 0$ and $\sigma_1 = 1$, $St(\theta, \nu, \mu_1, \sigma_1) \equiv St(\theta, \nu)$.

PROPERTY 2. For $\mu_1 = 0$ and $\sigma_1 = 1$; (i) as $\nu \to \infty$, $St(\theta, \nu, \mu_1, \sigma_1)$ distribution tends to $SN(\theta)$ distribution; (ii) as $\nu \to \infty$ and $\theta \to 0$, the distribution tends to $N(0, 1)$ distribution; (iii) as $\theta \to \infty$, the distribution tends to half-t_ν distribution.

PROPERTY 3. If $X \sim St(\theta, \nu, \mu_1, \sigma_1)$, then $-X$ is a $St(-\theta, \nu, \mu_1, -\sigma_1)$ random variable.

PROPERTY 4. If $Z \sim St(\theta, \nu)$, then $Z^2 \sim F_{1,\nu}$, a F distribution with degrees of freedom 1 and ν.

PROPERTY 5. For $\rho = 0$, $St(\theta, \nu, \mu_1, \sigma_1) \equiv t_\nu(\mu_1, \sigma_1)$.

Above properties yield useful distributional results. Especially, Property 4 leads to a counter example for showing that the reverse of a well-known property of t_ν distribution is not true. That is, for $Z \sim t_\nu$, then $Z^2 \sim F_{1,\nu}$, but the reverse is not true by Property 4. Furthermore, Property 5 can be used to test $H_0 : \rho = 0$ against $H_1 \neq 0$ for a $St(\theta, \nu, \mu_1, \sigma_1)$ distribution. The following further representation of $St(\theta, \nu)$ distribution relates the skew-normal distribution, $SN(\theta)$.

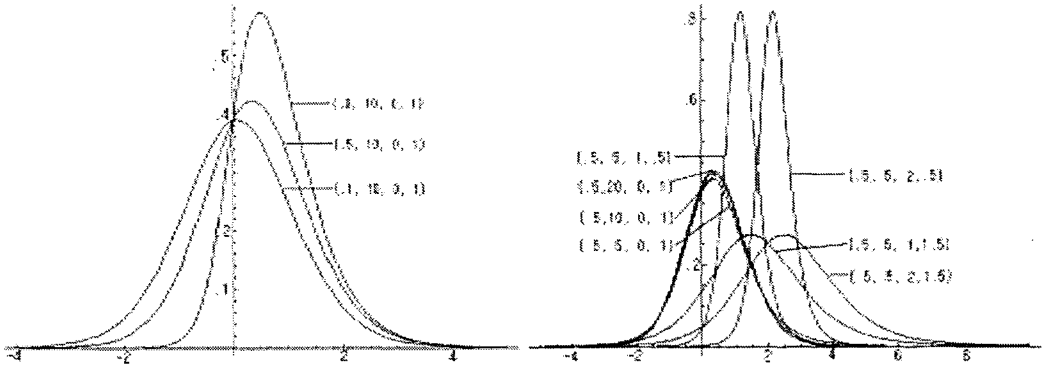

FIGURE 2.1 *Shape of Density of $St(\theta, \nu, \mu_1, \sigma_1)$ Distribution for Various Set of Parameter Values* $(\theta, \nu, \mu_1, \sigma_1)$.

THEOREM 2.2. *Let $Z \sim SN(\theta)$, then a scale mixture distribution of $\lambda^{-1/2}Z$ is $St(\theta, \nu)$, where the scale mixture distribution is $\lambda \sim Gamma(\nu/2, 2/\nu)$, a gamma distribution with the mean 1 and the variance $2/\nu$.*

PROOF. See Kim (2002).

Using Theorem 2.2 and the moment generating function of $Z \sim SN(\theta)$ in Azzalini (1985), we have at once that the moment generating function of $T \sim St(\theta, \nu)$ is

$$M_T(\eta) = 2E_\lambda[\exp(\eta^2/2\lambda)\Phi(\rho\lambda^{-1/2}\eta)], \; \infty < \eta < \infty. \tag{2.4}$$

□

Corollary 2.1 notes that it suffices to obtain the moments of T to get those of $X \sim St(\theta, \nu, \mu_1, \sigma_1)$. From (2.3), we obtain

$$E[T] = \rho(\nu/\pi)^{1/2}\Gamma[(\nu-1)/2]/\Gamma[\nu/2], \tag{2.5}$$

$$E[T^2] = \nu/(\nu-2) \quad \text{for } \nu > 2, \tag{2.6}$$

$$E[T^3] = \frac{3\rho\nu^{3/2}(1+2\theta^2/3)\Gamma[(\nu-3)/2]}{2\pi^{1/2}(1+\theta^2)\Gamma[\nu/2]} \quad \text{for } \nu > 3, \tag{2.7}$$

$$E[T^4] = 15\left(\frac{\nu}{2}\right)^3 \frac{\Gamma[(\nu-6)/2]}{\Gamma[\nu/2]} \quad \text{for } \nu > 6, \tag{2.8}$$

and skewness β_3 is $A/var(T)^{3/2}$, where

$$A = \theta\left[\left(\frac{\nu}{\pi}\right)^{1/2} \frac{3\nu}{(1+\theta^2)^{3/2}} \frac{\Gamma[(\nu-1)/2]}{(\nu-2)\Gamma[\nu/2]} \left\{\frac{1}{\nu-3} + \theta^2 B\right\}\right]$$

and

$$B = \left(\frac{2(\nu-2)}{3(\nu-3)} + \frac{2(\nu-2)}{3\pi} \frac{\Gamma[(\nu-1)/2]^2}{\Gamma[\nu/2]^2} - 1 \right). \tag{2.9}$$

A numerical evaluation by *Mathematika* showed that $B > 0$ for all integer values of ν for which $\nu > 3$, and hence the skewness β_3 depends only on the sign of ρ value in $\theta = \rho/\sqrt{1-\rho^2}$. This implies that $St(\theta, \nu, \mu_1, \sigma_1)$ distributions are skewed to the right (the left) when $\rho > 0$ ($\rho < 0$). Figure 2.1 depicts various shapes of density of the $St(\theta, \nu, \mu_1, \sigma_1)$distribution.

3. BAYESIAN ESTIMATION

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n obtained from $St(\theta, \nu, \mu_1, \sigma_1)$ Distribution. From (2.2), the log-likelihood can be written as

$$\ln L(\theta, \nu, \mu_1, \sigma_1) = n \ln(2/\sigma_1) + \sum_{i=1}^{n} \ln f_\nu(t_i) + \sum_{i=1}^{n} \ln F_{\nu+1} \left(\frac{\nu+1)^{1/2}\theta t_{1i}}{(\nu+t_i^2)^{1/2}} \right), \tag{3.1}$$

where $t_i = (x_i - \mu_1)/\sigma_1$. Unfortunately, general statistical analysis based on t_ν distributions has been hindered by the nonexistence of a simple estimator of the degrees of freedom ν (see Johnson *et al.* 1995, p.399). This implies that the likelihood function (3.1), a function of f_ν and $F_{\nu+1}$, is too complicate to obtain the maximum likelihood estimators of the four parameters in the $St(\theta, \nu, \mu_1, \sigma_1)$ distribution. Moreover, the moment estimators of the parameters obtained from using (2.5) through (2.8) would not be represented in a closed form. These, in turn, lead to absence of exact distributions of the estimators, and hence inference for the parameters is not possible, especially for small and intermediate sample size n. We shall show, by exploiting a particular mathematical representation involving the $St(\theta, \nu, \mu_1, \sigma_1)$ density, that Bayesian inference for the parameters can be implemented by using MCMC (Markov chain Monte Carlo) scheme, particularly the Gibbs sampler.

3.1. Joint Posterior Density

From Theorem 2.1, we can have an alternative expression for the joint distribution of $n \times 1$ vector of observations $\mathbf{x}$ from the $St(\theta, \nu, \mu_1, \sigma_1)$ distribution. More specifically, let (T_1, T_2) be a $t_\nu(0, 0, 1, 1, \rho)$ random variable, and let $T_{1i} = (X_i - \mu_1)/\sigma_1$, then the joint pdf of $X_1, X_2, \ldots, X_n$ can be expressed as

$$f(\mathbf{x} \,|\theta, \nu, \mu_1, \sigma_1) = (2/\sigma_1)^n \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n h_\nu(t_{1i}, t_{2i}) d\mathbf{t_2},$$

where $t_{1i} = (x_i - \mu_1)/\sigma_1$ for $i = 1, 2, \ldots, n$ and $\mathbf{t_2}$ is a $n \times 1$ vector of auxiliary independent t_ν variables t_{2i}. Therefore, under the reparametrization $\rho = \theta/\sqrt{1+\theta^2}$ and some prior density $\pi(\rho, \nu, \mu_1, \sigma_1)$, the posterior density of parameters is given by

$$\pi(\rho, \nu, \mu_1, \sigma_1 \,|\mathbf{x}) \propto \sigma_1^{-n} \int_0^\infty \cdots \int_0^\infty \pi(\rho, \nu, \mu_1, \sigma_1) \prod_{i=1}^n h_\nu(t_{1i}, t_{2i}) d\mathbf{t_2}. \tag{3.2}$$

This representation requires the use of MCMC schemes. The combination of MCMC and the forgoing posterior density representation enables us to produce samples from the joint posterior density of $St(\theta, \nu, \mu_1, \sigma_1)$parameters denoted by $\pi(\rho, \nu, \mu_1, \sigma_1 \,|\mathbf{x})$.

3.2. Gibbs Sampler with Data Augmentation

The Gibbs sampler is a Markovian updating scheme developed by Geman and Geman (1984) and introduced as a powerful tool in general Bayesian statistics by Gelfand and Smith (1990). The Gibbs sampler needs not to be restricted just to parameters. When a model includes auxiliary variables (missing data, for example), Gelfand *et al.* (1992) showed that such unobservable variables can simply be added to the parameter vector and the Gibbs sampler can be constructed for the augmented vector.

The idea of running the Gibbs sampler on an augmented vector of unknowns, we generate from $\pi(\rho, \nu, \mu_1, \sigma_1 \,|\mathbf{x})$ as follows. For each observation x_i, we generate a t_{2i} from $f(t_{2i} \,|\mu, \sigma, \rho, x_i)$, given by Lemma 2.1. Once we have generated the entire $n \times 1$ vector $\mathbf{t_2} = (t_{21}, t_{22}, \ldots, t_{2n})'$, we generate μ_1, σ_1, ρ and ν from respective full conditional posterior distributions:

$$\pi(\mu_1 \,|\rho, \nu, \sigma_1, \mathbf{x}, \mathbf{t_2}),\ \pi(\sigma_1 \,|\rho, \nu, \mu_1, \mathbf{x}, \mathbf{t_2}), \pi(\rho \,|\nu, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2}),\ \pi(\nu \,|\rho, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2}).$$

A long run, iterating this procedure, enables us to estimate and summarize features of $\pi(\rho, \nu, \mu_1, \sigma_1 \,|\mathbf{x})$, and hence $\pi(\theta, \nu, \mu_1, \sigma_1 \,|\mathbf{x})$ as required. The crucial feature of this method of analysis is that, by introducing auxiliary variables, we have circumvented the problem of being unable to specify the Gibbs sampler for the joint posterior (3.2) in a closed form.

3.3. Random Variate Generation of $\mathbf{t_2}$

We begin with the data augmentation, *i.e.* Gibbs sampling of $\mathbf{t_2} = (t_{21}, t_{22}, \ldots, t_{2n})'$. Efficient generation from

$$T_{2i} \,|(\rho, \nu, \mu_1, \sigma_1, x_i) \sim t_{\nu+1}(\alpha_i, \beta_i) I(T_{2i} \geq 0), \quad i = 1, \ldots, n, \tag{3.3}$$

where $t_{\nu+1}(\alpha_i, \beta_i) I(T_{2i} \geq 0)$, denotes a truncated $t_{\nu+1}(\alpha_i, \beta_i)$ distribution with lower truncation point at 0. Here $\alpha_i = \rho t_{1i}$, $\beta_i = \sqrt{(\nu + t_{1i}^2)(1 - \rho^2)/(\nu + 1)}$, $t_{1i} = (x_i - \mu_1)/\sigma_1$. The data augmentation is vital to the successful implementation of the $St(\theta, \nu, \mu_1, \sigma_1)$ analysis, because a t_{2i} value is required for every x_i value, at every iteration of the Gibbs sampler. The efficient one-for-one method by Devroye(1986) is available for generating variate T_{2i} from the truncated distribution (3.3): Let u_i be an observation generated from $Uniform(0, 1)$ distribution, then

$$t_{2i} = \alpha_i + \beta_i \; F_{\nu+1}^{-1}[P(u_i \,; \alpha_i, \beta_i)], \quad i = 1, \ldots, n \tag{3.4}$$

is a drawing from the distribution (3.3), where $F_{\nu+1}$ is the distribution function of $t_{\nu+1}$-distribution, and $P(u_i \,; \alpha_i, \beta_i) = F_{\nu+1}(-\alpha_i/\beta_i) + u_i(1 - F_{\nu+1}(-\beta_i/\alpha_i))$.

3.4. Random Variate Generation of μ_1

From Lemma 2.1 and (3.2), we see that the full conditional posterior density for μ_1, with a prior density $\pi(\mu_1)$, is given as

$$\pi(\mu_1 \,|\rho, \nu, \sigma_1, \mathbf{x}, \mathbf{t_2}) \propto \pi(\mu_1) \prod_{i=1}^{n} \left\{ 1 + \frac{[\mu_1 - (x_i - \sigma_1 \rho t_{2i})]^2}{\sigma_1^2 (1 - \rho^2)(\nu + t_{2i}^2)} \right\}^{-(\nu+2)/2} \tag{3.5}$$

for $-\infty < \mu_1 < \infty$. The fact that we have little knowledge of the shape of the density suggests using the Metropolis-Hastings sampling algorithm (see, Gustafson 1998 and references therein). Since the distribution of the full conditional posterior is continuous distribution, we adopt the RW (random walk) Metropolis algorithm that works as follows: Assume that we currently performing the kth iteration of the sampler then updating procedure from $\mu_1^{(k)}$ to $\mu_1^{(k+1)}$ is

1. Generate μ_1^* from $\mu_1^{(k)} + Z$, where $Z \sim N(0, \eta^2)$.

2. Generate u from a $Uniform(0, 1)$.

3. If $u < \pi(\mu_1^* \,|\rho, \nu, \sigma_1, \mathbf{x}, \mathbf{t_2}) / \pi(\mu_1^{(k)} \,|\rho, \nu, \sigma_1, \mathbf{x}, \mathbf{t_2})$, then $\mu_1^{(k+1)} = \mu_1^*$; otherwise, $\mu_1^{(k+1)} = \mu_1^{(k)}$.

3.5. Random Variate Generation of σ_1

The full conditional posterior density for σ_1 with a prior density $\pi(\sigma_1)$ is

$$\pi(\sigma_1 \,|\rho,\nu,\mu_1,\mathbf{x},\mathbf{t_2}) \propto \pi(\sigma_1)\sigma_1^{-n} \prod_{i=1}^{n} \left\{1 + \frac{[\mu_1 - (x_i - \sigma_1 \rho t_{2i})]^2}{\sigma_1^2(1-\rho^2)(\nu + t_{2i}^2)}\right\}^{-(\nu+2)/2} \tag{3.6}$$

for $\sigma_1 > 0$. Generating σ_1 from the distribution of (3.6) is not trivial since no real information on the shape of (3.6) is available, so once again, we use to the Metropolis- Hastings algorithm. We consider the following RW Metropolis algorithm with a de-constraint transformation to sample σ_1. Since $\sigma_1 > 0$, we let $\xi = \ln \sigma_1$, $-\infty < \xi < \infty$. Then

$$\pi(\xi \,|\rho,\nu,\mu_1,\mathbf{x},\mathbf{t_2}) = \pi(\sigma_1 \,|\rho,\nu,\mu_1,\mathbf{x},\mathbf{t_2})\, e^{\xi}. \tag{3.7}$$

Instead of directly sampling σ_1, we generate ξ by choosing a proposal transition density that adds noise to the current state. The algorithm to generate ξ operates as follows: Assume that we are currently performing the kth iteration of the sampler, then updating procedure from $\xi^{(k)}$ to $\xi^{(k+1)}$ is to use the RW Metropolis step as in Section 3.4. After we obtain $\xi^{(k+1)}$, we compute $\sigma^{(k+1)}$ by using the relation $\xi = \ln \sigma_1$.

3.6. Random Variate Generation of ρ

Assuming a uniform prior $Uniform(-1,1)$ for ρ, the full conditional posterior density for ρ is given by

$$\pi(\rho \,|\nu,\mu_1,\sigma_1,\mathbf{x},\mathbf{t_2}) \propto (1-\rho^2)^{-n/2} \prod_{i=1}^{n} \left\{1 + \frac{[\mu_1 - (x_i - \sigma_1 \rho t_{2i})]^2}{\sigma_1^2(1-\rho^2)(\nu + t_{2i}^2)}\right\}^{-(\nu+2)/2} \tag{3.8}$$

for $-1 < \rho < 1$. Generating ρ from (3.8) is not trivial since (3.8) is not log-concave. Therefore, we consider the following RW Metropolis algorithm with a de-constraint transformation to sample ρ. Since $-1 < \rho < 1$, we let

$$\rho = (-1 + e^{\zeta})/(1 + e^{\zeta}), \;\; -\infty < \zeta < \infty. \tag{3.9}$$

Then

$$\pi(\zeta \,|\mu_1,\sigma_1,\mathbf{x},\mathbf{t_2}) = \pi(\rho \,|\mu_1,\sigma_1,\mathbf{x},\mathbf{t_2}) \frac{2e^{\zeta}}{(1+e^{\zeta})^2}.$$

Instead of directly sampling ρ, we generate ζ by using the RW Metropolis step in Section 3.4. After we generate $\zeta^{(k+1)}$, we compute $\rho^{(k+1)}$ in each iteration by using the relation (3.9), and hence we generate $\theta^{(k+1)}$ by using the relation $\theta = \rho/\sqrt{1-\rho^2}$.

3.7. Random Variate Generation of ν

Finally, we sample ν from its full conditional posterior density that is

$$\pi(\nu \mid \rho, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2}) \propto \pi(\nu) h(\nu \mid \rho, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2}), \tag{3.10}$$

where

$$h(\nu \mid \rho, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2}) = \left(\frac{\nu^{\nu/2} \Gamma[(\nu+2)/2]}{\Gamma[\nu/2]} \right)^n \prod_{i=1}^n \left\{ \nu + \frac{[t_{1i}^2 - 2\rho t_{1i} t_{2i} + t_{2i}^2]}{1-\rho^2} \right\}^{-(\nu+2)/2},$$

and $t_{1i} = (x_i - \mu_1)/\sigma_1$. A way of generating ν is to use a Metropolis-Hastings algorithm (practiced by Chib and Greenberg, 1995) using the *Uniform* prior on ψ $(0 < \psi \leq 1)$, where $\psi = 1/\nu$. Thus we set a proposal density $q(\psi, \psi^*) = \pi(\psi^*)$ which supplies candidate values ψ^* given the current value of ψ. In this case, the probability of move requires only the computation of h function. Thus the kth iteration of the Metropolis step is given by

1. Generate ψ^* from a $Uniform(0,1)$.
2. Generate u from a $Uniform(0,1)$.
3. If $u < h(\psi^* \mid \rho, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2}) / h(\psi^{(k)} \mid \rho, \mu_1, \sigma_1, \mathbf{x}, \mathbf{t_2})$ then $\psi^{(k+1)} = \psi^*$; otherwise, $\psi^{(k+1)} = \psi^{(k)}$.

After we obtain $\psi^{(k+1)}$, we compute $\nu^{(k+1)}$ by using the relation $\nu = 1/\psi$. Note that the proposal density need not to enforce the interval constraint, because it is uniform distribution on $0 < \psi \leq 1$. Thus we use a Metropolis step to draw μ_1, σ_1, ρ, and ν and the Gibbs sample is obtained by drawing $\mathbf{t_2}$, μ_1, σ_1, ρ, and ν in turn, after convergence.

4. NUMERICAL EXAMPLES

4.1. Simulation Study

Our examples are illustration of extensive studies we have undertaken to validate the MCMC method. We generated n observations from a $X \sim St(\theta, \nu, \mu_1, \sigma_1)$ distribution using the algorithm in Section 2 and then ran the Gibbs sampler for 60,000 iterations. For the starting points of the sampler, it appears that the sample mean and the sample standard deviation are reasonable starting points for μ_1 and σ_1. In an attempt to test the robustness of the sampler, we started ρ

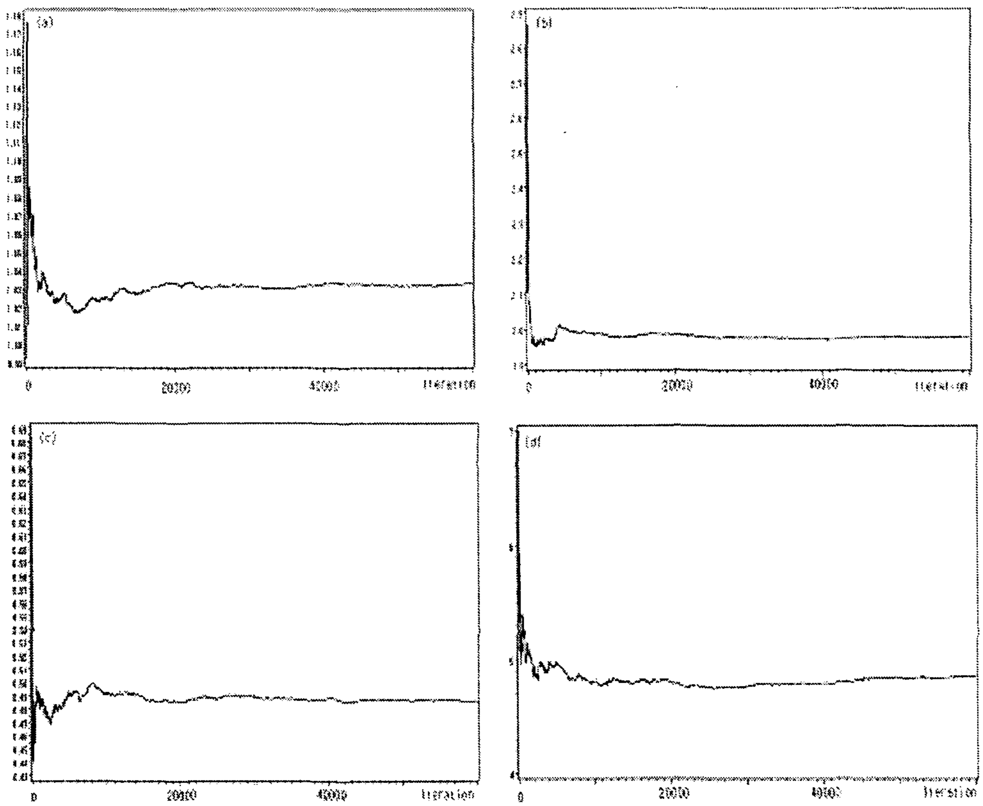

FIGURE 4.1 *The Ergodic Average of the Trace of the Parameters; (a) Average of Gibbs sample of μ_1 up to kth iteration; (b) Average of Gibbs sample of σ_1 up to kth iteration; (c) Average of ρ; (d) Average of ν.*

and ν well away from their true values, *i.e.* true value of ρ plus 0.2 and that of ν plus 2.

For the analysis of the MCMC method given in Section 3, we need to specify the priors of μ_1 and σ_1. Because μ_1 and σ_1 correspond to location and scale, it may be relatively straightforward for a particular application to assign informative prior distributions to these parameters (because it is usually reasonable to assume independent between the two). The many guideline for prior selection of the location and scale of a normal distribution may be followed as a good benchmark (see, for example, Berger 1985). Thus we assume $\pi(\mu_1) \propto \exp\left\{-(\mu_1 - \delta)^2/(2\tau^2)\right\}$, a normal prior, and $\pi(\sigma_1) \propto \sigma_1^{-m-1} \exp\left\{-S/(2\sigma_1^2)\right\}$, a generalized inverse Chi-density on m degrees of freedom as given in Lee (1997). The hyperparameter

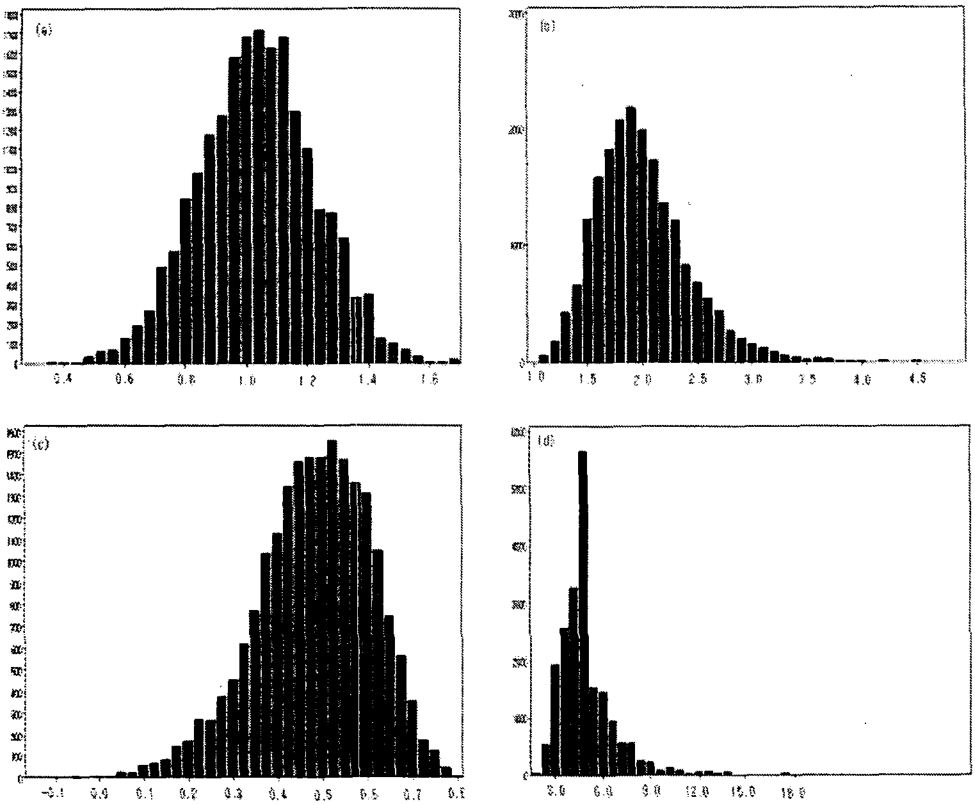

FIGURE 4.2 *Frequency Histogram of* $L = 20,000$ *Gibbs Sample of* $(\rho, \nu, \mu_1, \sigma_1)$; *(a) Histogram for Gibbs sample of* μ_1; *(b) That of* σ_1; *(c) That of* ρ; *(d) That of* ν.

specification was defined by $\delta = 0$, $\tau = 10$, $m = 3$, and $S = 100$, reflecting rather vague initial information relative to that to be provided by the data. A simulation study with various sample sizes and set of parameter values is conducted and estimation results by the MCMC method are obtained as listed in Table 4.1.

For calculating the estimates, the iterative process was monitored by observing trace of the Gibbs samples. The diagnostics we used are described in Cowles and Carlin (1996). For each data set, we used 40,000 iterations to "burn in" the sampler; the decision is based on the trace plots in Figure 4.1. Figure 4.1 shows the ergodic averages of the trace of the parameters, $(\rho, \nu, \mu_1, \sigma_1)$, leading us to believe that convergence has been attained before 40,000 iterations. By adjusting the turning constant (standard deviation of the transition density in the RW Metropolis algorithm), we were able to keep the jumping probabilities between .23

TABLE 4.1 *Posterior Summaries for $St(\theta, \nu, \mu_1, \sigma_1)$ distribution Parameters*

True Value		Posterior Mean				Posterior s.d.			
$(\rho, \nu, \mu_1, \sigma_1)$	n	$\hat{\rho}$	$\hat{\nu}$	$\hat{\mu_1}$	$\hat{\sigma_1}$	s.d$(\hat{\rho})$	s.d$(\hat{\nu})$	s.d$(\hat{\mu_1})$	s.d$(\hat{\sigma_1})$
(.7, 5, -1, 2)	20	.627	6.126	-.993	2.383	.132	1.892	.198	.241
	50	.645	5.452	-1.019	1.872	.126	1.564	.188	.194
	100	.689	4.865	-.941	2.010	.086	1.221	.176	.154
(.7, 10, -1, 2)	20	.751	9.123	-.961	2.126	.089	1.762	.187	.281
	50	.750	11.324	-.968	1.985	.099	1.634	.185	.201
	100	.664	9.628	-.971	2.172	.089	1.469	.182	.153
(.5, 5, 1, 2)	20	.562	6.917	1.033	2.175	.143	1.813	.189	.287
	50	.485	4.982	1.029	2.092	.117	1.794	.191	.216
	100	.487	5.164	.983	2.036	.097	1.263	.181	.163
(.5, 10, 1, 2)	20	.404	8.952	.955	2.231	.185	1.712	.195	.210
	50	.462	9.417	.954	2.071	.150	1.599	.198	.175
	100	.483	10.535	.980	2.023	.127	1.486	.190	.138
(.5, 20, 1, 2)	20	.524	18.164	1.015	1.885	.170	1.936	.195	.204
	50	.541	19.361	.946	1.983	.136	1.827	.191	.171
	100	.537	19.621	.934	2.258	.105	1.612	.183	.151
(.3, 5, 3, 2)	20	.359	6.198	2.853	2.279	.183	1.912	.204	.315
	50	.312	5.976	2.825	2.156	.158	1.654	.197	.208
	100	.304	5.297	2.783	2.118	.111	1.015	.185	.157
(.3, 10, 3, 2)	20	.293	11.786	2.888	2.344	.198	1.753	.193	.277
	50	.283	11.019	2.868	2.019	.163	1.534	.191	.168
	100	.286	10.592	2.824	2.062	.108	1.212	.180	.139
(.3, 20, 3, 2)	20	.320	22.549	2.857	2.321	.181	2.387	.197	.293
	50	.316	21.154	2.893	1.944	.163	2.146	.200	.164
	100	.303	19.259	2.889	2.150	.118	1.789	.191	.138

and .5 (see, Gelman *et al.* 1996; Robert *et al.* 1997). The frequency histogram of the Gibbs sample of each parameter was also considered. As given in Figure 4.2, all the histograms were centered about their true values and seemingly unimodal. The sample means (estimates of posterior means) are not perfect as parameter estimates, because of the amount of skew in the Gibbs samples as depicted by the histograms; however, as given in Table 4.1, they produced accurate estimates for the parameters of $St(\theta, \nu, \mu_1, \sigma_1)$ distribution.

4.2. An Example with a Screened Data

The screening process is frequently based upon an individual's score on one or more screening variables. A $St(\rho, \nu, \mu_1, \sigma_1)$ model is useful for describing the

TABLE 4.2 *Summary of Statistics*

Variable	Mean	s.d.	Correlation
Environment Score (X)	$\mu_1 = 55.462$	$\sigma_1\sqrt{\nu}/\sqrt{\nu-2} = 15.974$	$\rho = .844$
Index Score(Y)	$\mu_2 = 52.683$	$\sigma_2\sqrt{\nu}/\sqrt{\nu-2} = 18.863$	

screening process. As an example for this, we consider a data set from KEBIX 2003 (Korea e-Business Index 2003) data base. The data set consists of 509 pairs of environment score (X) and e-business index score (Y) obtained from companies (or public organizations) that are operating in Korea in 2003. The environment score (X) measures external environment that effect e-business activities and the e-business index score (Y) measures e-business activities of a company (or public organization). Summary of statistics of the 509 pairs of scores is listed in Table 4.2.

For an illustrative purpose, we set up a screening process that drops companies whose index scores achieve $Y < \mu_2$. So that we consider here the case in which Y represents the screened variable and X represents the variable that is measured following initial screening. The process dropped 266 companies so that 243 companies have passed the screening process. We assume that the only observations available are environmental scores (X) of 243 companies that passed the screening process and that corresponding index scores (Y) are not available. Figure 4.3 depicts a Box-plot of the 243 scores of the unscreened variable of X, indicating moderate right skewness of the score distribution. The mean and standard deviation of the 243 observed scores are $\bar{x} = 69.617$ and $s_X = 8.5682$, respectively. Given the 243 scores, applying the MCMC method in Section 3 (with 10,000 iterations to burn in the sampler and Gibbs sample of size =1,000), we estimate the location and scale parameter of the unscreened population environmental scores as $\hat{\mu_1} = 56.915(1.495)$ and $\hat{\sigma_1} = 14.596(.694)$. The quantities in parentheses are the estimated standard deviations of the estimates. The MCMC method also gives estimates of ρ and ν, the degrees of freedom of the $St(\rho, \nu, \mu_1, \sigma_1)$ distribution. They are $\hat{\rho} = .899(.075)$ and $\nu = 18.826(2.410)$. Comparing the estimates with the true values listed in Table 4.2, we see that $\hat{\mu_1} = 56.915$ and $\hat{\sigma_1}\sqrt{\hat{\nu}}/\sqrt{\hat{\nu}-2} = 15.439$ give fairly good estimates of the theoretical mean (55.462) and standard deviation (15.974) of in the unscreened population, while the values, $\bar{x} = 69.617$ and $s_X = 8.5682$, are population mean and standard deviation of the marginal distribution of X, subject to $Y > \mu_2$.

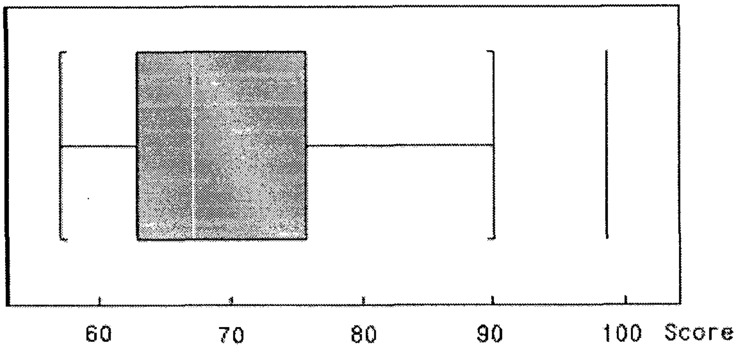

FIGURE 4.3 *Box-plot of the Screened Data*

To evaluate the fit of the posterior distribution of the $St(\rho, \nu, \mu_1, \sigma_1)$ model, we compare the observed data to the posterior predictive distribution using the posterior predictive p-value suggested by Gelman *et al.* (2000, p.169). The p-value is defined as the probability that the predicted data could be more extreme than the observed data, as measured by the test quantity:

$$\text{Bayes } p\text{-value} = Pr(T(\mathbf{x}^{pre}, \Psi) \geq T(\mathbf{x}, \Psi) \,|\mathbf{x}), \tag{4.1}$$

where the probability is taken over the posterior distribution of $\Psi = (\nu, \rho, \mu_1, \sigma_1)'$. Here $\mathbf{x}^{pre}$ and $\mathbf{x}$ denote the vectors of predicted and realized observations, respectively. As the test quantity, $T(\mathbf{x}, \Psi)$, we choose a general goodness-of-fit measure that is the discrepancy quantity, written here in terms of univariate observation x_i:

$$T(\mathbf{x}, \Psi) = \sum_{i=1}^{n} \frac{(x_i - E(x_i \,|\Psi))^2}{Var(x_i \,|\Psi)}. \tag{4.2}$$

The expected value and the variance in (4.2) are readily evaluated by (2.5) and (2.6). For evaluating $T(\mathbf{x}^{pre}, \Psi)$, we use the following simulation: If we already have Gibbs sample $\{\Psi_\ell \;; \ell = 1, \ldots, L\}$ of size $L = 1,000$ from the posterior density of Ψ, we just draw one observation vector $\mathbf{x}_\ell^{pre}$ of size $n = 243$ from the predictive distribution obtained from ℓth simulated value of Ψ, Ψ_ℓ. Thus we have L draws of $\mathbf{x}_\ell^{pre}$. The posterior predictive check is the comparison between the realized test quantities, $T(\mathbf{x}, \Psi)$, and predictive test quantities, $T(\mathbf{x}_\ell^{pre}, \Psi)$. The estimated p-value is just the proportion of these simulations for which the test quantity equals or exceeds its realized value; that is, for which

$$T(\mathbf{x}_\ell{}^{rep}, \Psi_\ell) \geq T(\mathbf{x}, \Psi_\ell), \quad \ell = 1, 2, \ldots, L.$$

In this example, the estimated p-value based on $L = 1,000$ draws is 0.342. This implies that the unscreened environmental scores (X) artificially obtained from KEBIX 2003 data seem to well described by the $St(\rho, \nu, \mu_1, \sigma_1)$ model.

5. Concluding Remarks

We have considered a distribution of the truncated bivariate t random variable, the $St(\rho, \nu, \mu_1, \sigma_1)$ distribution. Little work on the estimation appears in the literature due to the complex likelihood function. Estimation of the four parameters of the distribution, corresponding to degrees of freedom, skewness, location, and scale has been carried out using a MCMC method exploiting the result of Theorem 2.1, specifically the Gibbs sampler consisting of a data augmentation technique derived by Theorem 2.1 and Metropolis-Hastings algorithm with each Metropolis step obtained using an independence chain. We have shown how the MCMC method can be used to generate posterior samples from the parameters of the distribution. With these samples, we are in a position not only to estimate parameter values, but also to make more general inference. We could, for example, readily compute parameter quantiles or estimate arbitrary functions of the parameters.

The validation examples in Section 4 demonstrated good performance of the MCMC method. Moreover, it is seen that the $St(\rho, \nu, \mu_1, \sigma_1)$ model is suitable for fitting data from a screening process where Y variable is screened but attention is concentrated on the distribution of the unscreened variable X. The estimation of the model provide both estimates of the expectation and the standard deviation of X, as well as an estimate of the correlation between the two variables in the original unscreened population, even though no observations of Y are available.

References

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q. (1993). "The nontruncated marginal of a truncated bivariate normal distribution", *Psychometrica*, **58**, 471-478.

Arnold, B. C., Castillo, A. and Sarabia, J. M. (2002). "Conditionally specified multivariate skew distributions", *Sankhyā*, **A64**, 206-226.

Azzalini, A. (1985). "A class of distributions which includes the normal ones", *Scandinavian Journal of Statistics*, **12**, 171-178.

AZZALINI, A. AND CAPITANIO, A. (2003). "Distributions generated by perturbation of symmetry with emphasis on a multivariate distribution", *Journal of the Royal Statistical Society*, **B35**, 367-389.

AZZALINI, A. AND DALLA VALLE, A. (1996). "The multivariate skew-normal distribution", *Biometrika*, **83**, 715-726.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, 2nd ed.*, New York: Springer-Verlag.

BRANCO, M. D. AND DEY, D. K. (2001). "A general class of multivariate skew-elliptical distributions", *Journal of Multivariate Analysis*, **79**, 99-113.

CHEN, M. H., DEY, D. K., AND SHAO, Q. M. (1999). "A new skewed link model for dichotomous quantal response data", *Journal of the American Statistical Association*, **94**, 1172-1185.

CHIB, S. AND GREENBERG, E. (1995). "Understanding the Metropolis-Hastings Algorithm", *The American Statistician*, **49**, 327-335.

COHEN, A. C. (1991). *Truncated and Censored Samples: Theory and Applications*, New York: Marcel Dekker.

COWLES, M. AND CARLIN, B. (1996). "Markov chain Monte Carlo diagnostics: A comparative review", *Journal of the American Statistical Association*, **91**, 883-904.

DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*, New York: Springer Verlag.

GELFAND, A. E. AND SMITH, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities", *Journal of the American Statistical Association*, **85**, 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", *Journal of the American Statistical Association*, **87**, 523-532.

GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. (2000). *Bayesian Data Analysis*, New York: Chapman and Hall.

GELMAN, A. ROBERT, G. O., AND GILKS, W. R. (1996). "Efficient Metropolis jumping rules", in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford UK: Oxford University Press, 599-607.

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.

GUSTAFSON, P.(1998). "A guided walk Metropolis algorithm", *Statistics and Computing*, **8**, 357-364.

HENZE, N. (1986). "A probabilistic representation of the skew-normal distribution", *Scandinavian Journal of Statistics*, **13**, 271-275.

JOHNSON, N. L., KOTZ, S., AND BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, **Vol. 2**, New York: John Wiley.

KIM, H. J. (2002). "Binary regression with a class of skewed t link models", *Communications in Statistics-Theory and Methods*, **31**, 1863-1886.

LEE. P. M. (1997). *Bayesian Statistics*, 2nd ed., New York: John Wiley.

MA, T. AND GENTON, M. G. (2004). "A flexible class of skew-symmetric distributions. To appear in *Scandinavian Journal of Statistics.* http:www4.stat.ncsu.edu/ mggenton/publications.html.

ROBERT, G. O., GELMAN, A., AND GILKS, W. R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithm, *Annals of Applied Probability*, **7**, 110-120.