

Bayesian Estimation for Skew Normal Distributions Using Data Augmentation

Hea-Jung Kim¹⁾

Abstract

In this paper, we develop a MCMC method for estimating the skew normal distributions. The method utilizing the data augmentation technique gives a simple way of inferring the distribution where fully parametric frequentist approaches are not available for small to moderate sample cases. Necessary theories involved in the method and computation are provided. Two numerical examples are given to demonstrate the performance of the method.

Keywords : Skew normal distribution, Bayesian estimation, MCMC, data augmentation

1. Introduction

Azzalini(1985) and Henze(1986) worked on the so-called skew normal distribution, a family of distributions including the standard normal, but with an extra parameter to regulate skewness. A random variable X is said to be skew normal with parameters μ , σ , and λ , written $SN(\lambda; \mu, \sigma)$, if its density function is

$$f(x|\mu, \sigma, \lambda) = (2/\sigma)\Phi(\lambda z)\phi(z), \quad -\infty < x < \infty, \quad (1.1)$$

where $-\infty < \mu, \lambda < \infty$, $\sigma > 0$, $z = (x - \mu)/\sigma$, and $\phi(z)$ and $\Phi(z)$ are the standard normal density and distribution functions, respectively. The skewness is regulated by the parameter λ ; $\lambda = 0$ yields the normal model. We refer to Azzalini(1986), Arnold et al.(1993), Chen, Dey and Shao(1999), Kim(2002) for the application of the distribution. Multivariate extensions of this model have been developed by Azzalini and Dalla Valle(1996) and Azzalini and Capitanio(1999). Other models handling skewness and heavy tails simultaneously are also developed by Azzalini(1986), Branco(2001), and Kim(2002).

A number of investigators studied inferential aspects of the $SN(\lambda; \mu, \sigma)$ distribution. The information matrix of the MLEs in the distribution has been derived by Azzalini(1985). Other

1) Professor, Department of Statistics, Dongguk University, Seoul 100-715, Korea
E-mail : kim3hj@dongguk.edu

estimation issues of the distribution have been considered by Azzalini and Dalla Valle(1996) and Azzalini and Capitanio(1999). As pointed out by Azzalini and Capitanio(1999), however, the likelihood shape and the MLE are problematic in the inference of the $SN(\lambda; \mu, \sigma)$ distribution. In addition to difficulties with numerical maximization, anomalies of the ML estimate exist due to the intrinsic properties of the likelihood function, which are not removable by a change of parameterization. Specifically, the estimate of λ , obtainable from profile likelihood of λ , tends to diverge when sample size is small to moderate and the source of this sort of anomaly has not been fully explained. Furthermore, exact distribution of the MLEs is not known so that inferences about μ , σ , and λ are not possible for small to moderate sample size.

In this paper, we suggest an alternative estimation method that resolves the problems involved in the ML estimation of $SN(\lambda; \mu, \sigma)$ distribution. This is done by Bayesian estimation via Markov chain Monte Carlo(MCMC) that allows us to sample from the distribution of parameters of $SN(\lambda; \mu, \sigma)$ distribution, by exploiting a particular mathematical representation involving the skew normal density. The interest in studying the estimation comes from both theoretical and applied directions. On the theoretical side, it provides small to moderate sample inference for $SN(\lambda; \mu, \sigma)$ distribution. In applied view point, the suggested estimation method provides yet another example of data augmentation technique for MCMC method.

2. Bayesian Inference

2.1 The Joint Posterior Density

The Bayesian paradigm is to update parameter knowledge, in the form of a density function $\pi(\theta)$, using observed data \mathbf{x} , through the parametric model density function $p(\mathbf{x}|\theta)$. The result is a posterior density given by $\pi(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)\pi(\theta)$. Suppose the model density for \mathbf{x} alone is a complex form, but a simple form of the model density is available if we express it jointly with some extra random variables \mathbf{y} in the form of $p(\mathbf{x}, \mathbf{y}|\theta)$. Then the posterior is formally obtainable through Bayes' theorem by integrating out the unwanted variables, so that $\pi(\theta|\mathbf{x}) \propto \int p(\mathbf{x}, \mathbf{y}|\theta)\pi(\theta)d\mathbf{y}$. Such a representation is available for \mathbf{x} modeled by the skew normal distributions, and we begin with the following theorem.

Theorem 2.1. *Let $\lambda = \rho/(1 - \rho^2)^{1/2}$, and define a density function of $X = \sigma Z + \mu$ and Y , conditional on λ as*

$$f(x, y | \lambda) = (2/\sigma)g(z, y), \quad 0 < y < \infty, \quad -\infty < x < \infty, \quad (2.1)$$

where $g(z, y)$ denotes the joint pdf of a standard bivariate normal random variable with correlation coefficient $\rho(|\rho| < 1)$ so that $-\infty < \lambda < \infty$, and $z = (x - \mu)/\sigma$. Then

$f(x, y|\lambda)$ is a proper bivariate probability density for the distribution of (X, Y) , and the marginal distribution of X is $SN(\lambda; \mu, \sigma)$.

Proof. Note that, for all λ , $f(x, y|\lambda) > 0$ holds for the space of (X, Y) and

$$\int_{-\infty}^{\infty} \int_0^{\infty} f(x, y|\lambda) dy dx = (2/\sigma)(1+\lambda^2)^{1/2} \int_{-\infty}^{\infty} \phi(z) \int_0^{\infty} \phi((1+\lambda^2)^{1/2}y - \lambda z) dy dx$$

$$= (2/\sigma) \int_{-\infty}^{\infty} \phi(z) \Phi(\lambda z) dz = 1.$$

Thus our function (2.1) is indeed a proper bivariate density function. Moreover, the marginal distribution of X has the density function of $SN(\lambda; \mu, \sigma)$ distribution.

Corollary 2.1. The density for Y conditional on X , $f(y|\lambda, x)$, is given by

$$f(y|\lambda, x) \propto \phi((1+\lambda^2)^{1/2}y - \lambda x), \quad 0 < y < \infty \tag{2.2}$$

for $-\infty < x < \infty$.

Proof. The proof is trivial.

From Theorem 2.1, we can have an alternative expression for the joint distribution of $n \times 1$ vector of observations \mathbf{x} from $SN(\lambda; \mu, \sigma)$, $\lambda = \rho/(1 - \rho^2)^{1/2}$, distribution. More specifically, let (Z_i, Y_i) be a standard bivariate normal random variable with correlation coefficient ρ ($|\rho| < 1$) and let $Z_i = (X_i - \mu)/\sigma$, then the joint pdf in (1.1) can be expressed as

$$f(\mathbf{x}|\mu, \sigma, \lambda) = 2^n \int_0^{\infty} \dots \int_0^{\infty} \prod_{i=1}^n \sigma^{-1} g(z_i, y_i) d\mathbf{y},$$

where $z_i = (x_i - \mu)/\sigma$ for $i = 1, 2, \dots, n$ and \mathbf{y} is a $n \times 1$ vector of auxiliary independent standard normal variables y_i . Therefore, under the reparameterization $\rho = \lambda/(1 + \lambda^2)^{1/2}$ and some prior density $\pi(\mu, \sigma, \rho)$, the posterior density of the skew normal parameters is given by

$$\pi(\mu, \sigma, \rho|\mathbf{x}) \propto \int_0^{\infty} \dots \int_0^{\infty} \left(\prod_{i=1}^n \sigma^{-1} g(z_i, y_i) \right) \pi(\mu, \sigma, \rho) d\mathbf{y}. \tag{2.3}$$

As we shall show, the practical exploitation of this representation requires the use of Markov chain Monte Carlo (MCMC) simulation. This combination of MCMC and the forgoing density representation will enable us to produce samples from the joint posterior density of the skew normal parameters denoted by $\pi(\mu, \sigma, \lambda|\mathbf{x})$.

2.2 The Gibbs Sampler Using Data Augmentation

The Gibbs sampler is a Markovian updating scheme originally developed by Geman and Geman(1984) for use in image processing 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.

Exploiting Theorem 2.1 and the idea of running the Gibbs sampler on an augmented vector of unknowns, we generate from $\pi(\mu, \sigma, \rho | \mathbf{x})$ as follows. For each observation x_i , we generate a y_i from $f(y_i | \mu, \sigma, \rho, x_i)$, given up to proportionality by Corollary 2.1. Once we have generated the entire $n \times 1$ vector \mathbf{y} , we generate μ , σ , and ρ from respective full conditional posterior distributions:

$$\pi(\mu | \sigma, \rho, \mathbf{x}, \mathbf{y}), \pi(\sigma | \mu, \rho, \mathbf{x}, \mathbf{y}), \text{ and } \pi(\rho | \mu, \sigma, \mathbf{x}, \mathbf{y}).$$

A long run, iterating this procedure, enables us to estimate and summarize features of $\pi(\mu, \sigma, \rho | \mathbf{x})$, and hence $\pi(\mu, \sigma, \lambda | \mathbf{x})$ 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 skew normal distribution in closed form.

2.3 Random Variate Generation of \mathbf{y}

We begin with Gibbs sampling of \mathbf{y} . Efficient generation from

$$f(y_i | \mu, \sigma, \rho, x_i) \propto \exp \left\{ -\frac{[y_i - \rho z_i]^2}{2(1 - \rho^2)} \right\}, \quad 0 < y_i < \infty; i = 1, \dots, n, \quad (2.4)$$

is vital to the successful implementation of this method of skew normal analysis, because a y_i value is required for every x_i value, at every iteration of the Gibbs sampler, where $z_i = (x_i - \mu)/\sigma$. The conditional posterior distribution of Y_i is a truncated normal, $N(\rho z_i, 1 - \rho^2)$, with lower truncation point at 0. The efficient one-for-one method by Devroye(1986) is available for generating variate Y_i from the truncated distribution (2.4): Let u_i be an observation generated from Uniform(0, 1) distribution, then

$$y_i = \rho z_i + (1 - \rho^2)^{1/2} \Phi^{-1}[p(u_i; \rho, z_i)], \quad i = 1, \dots, n \quad (2.5)$$

is a drawing from the distribution (2.4), where

$$p(u_i; \rho, z_i) = \Phi(-\rho z_i / (1 - \rho^2)^{1/2}) + u_i (1 - \Phi(-\rho z_i / (1 - \rho^2)^{1/2})) .$$

2.4 Random Variate Generation of μ

From (2.3), we see that the conditional posterior density for μ , with a prior density $\pi(\mu)$,

is given as

$$\pi(\mu|\sigma, \rho, \mathbf{x}, \mathbf{y}) \propto \exp \left\{ -\frac{n[\mu - (\bar{x} - \sigma\rho\bar{y})]^2}{2\sigma^2(1-\rho^2)} \right\} \pi(\mu) \tag{2.6}$$

for $-\infty < \mu < \infty$. Assume that μ, σ and ρ are priori independent and we have a conjugate prior, $\pi(\mu) \propto \exp \{ -(\mu - \delta)^2 / (2\tau^2) \}$, then the full conditional posterior distribution of μ is

$$\mu|\sigma, \rho, \mathbf{x}, \mathbf{y} \sim N \left(\frac{n\tau^2(\bar{x} - \sigma\rho\bar{y}) + \sigma^2(1-\rho^2)\delta}{n\tau^2 + \sigma^2(1-\rho^2)}, \frac{\tau^2\sigma^2(1-\rho^2)}{n\tau^2 + \sigma^2(1-\rho^2)} \right). \tag{2.7}$$

An assessment of the hyperparameters δ and τ^2 leads to Gibbs sampling of μ from the normal distribution (2.7).

2.5 Random Variate Generation of σ

If we assume $\pi(\sigma) \propto \sigma^{-m-1} \exp \{ -S / (2\sigma^2) \}$, a generalized inverse Chi-density on m degrees of freedom as given in Lee(1997). The full conditional posterior density for σ with the prior density is

$$\pi(\sigma|\mu, \rho, \mathbf{x}, \mathbf{y}) \propto \sigma^{-(n+m+1)} \exp \left\{ -\frac{(1-\rho^2)S + \sum_{i=1}^n (x_i - \mu)^2 - 2\sigma\rho \sum_{i=1}^n (x_i - \mu)y_i}{2\sigma^2(1-\rho^2)} \right\} \tag{2.8}$$

for $\sigma > 0$, where m and S are the hyperparameters to be assessed. Generating σ from the distribution of (2.8) is not trivial since no real information on the shape of (2.8) is available, and we must resort to the Metropolis algorithm. Therefore, we consider the following random walk(RW) Metropolis algorithm with a de-constraint transformation to sample σ . Since $\sigma > 0$, we let $\xi = \ln \sigma$, $-\infty < \xi < \infty$. Then

$$\pi(\xi|\mu, \rho, \mathbf{x}, \mathbf{y}) = \pi(\sigma|\mu, \rho, \mathbf{x}, \mathbf{y}) e^\xi. \tag{2.9}$$

Instead of directly sampling σ , 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 i th iteration of the sampler, then updating procedure from ξ_i to ξ_{i+1} is

1. Generate ξ from $\xi_i + Z$, where $Z \sim N(0, \eta^2)$.
2. Generate u from a Uniform(0, 1).
3. If $u < \pi(\xi|\mu_{i+1}, \rho_i, \mathbf{x}, \mathbf{y}_{i+1}) / \pi(\xi_i|\mu_{i+1}, \rho_i, \mathbf{x}, \mathbf{y}_{i+1})$ then $\xi_{i+1} = \xi$;
otherwise, $\xi_{i+1} = \xi_i$.

After we obtain ξ_{i+1} , we compute σ_{i+1} by using the relation $\xi = \ln \sigma$. The variance, η^2 of the normal proposal transition density can be tuned so that the optimal acceptance rate 0.23 of Robert et al.(1997) is achieved.

2.6 Random Variate Generation of ρ

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

$$\pi(\rho|\mu, \sigma, \mathbf{x}, \mathbf{y}) \propto (1 - \rho^2)^{-n/2} \exp \left\{ - \frac{\sum_{i=1}^n [y_i^2 - 2\rho y_i z_i + z_i^2]}{2(1 - \rho^2)} \right\} \quad (2.10)$$

for $-1 < \rho < 1$. Generating ρ from (2.10) is not trivial since (2.10) is not log-concave. Therefore, we consider the following random walk Metropolis algorithm with a de-constraint transformation to sample ρ . Since $-1 < \rho < 1$, we let

$$\rho = (-1 + e^\zeta)/(1 + e^\zeta), \quad -\infty < \zeta < \infty. \quad (2.11)$$

Then

$$\pi(\zeta|\mu, \sigma, \mathbf{x}, \mathbf{y}) = \pi(\rho|\mu, \sigma, \mathbf{x}, \mathbf{y}) \frac{2e^\zeta}{(1 + e^\zeta)^2}.$$

Instead of directly sampling ρ , we generate ζ by using the RW Metropolis algorithm. We can use the algorithm as is used to generate from (2.10) as follows: Assume that we are currently performing the i th iteration of the sampler, then:

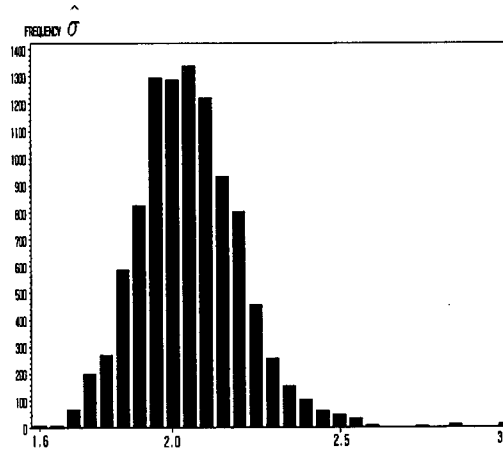
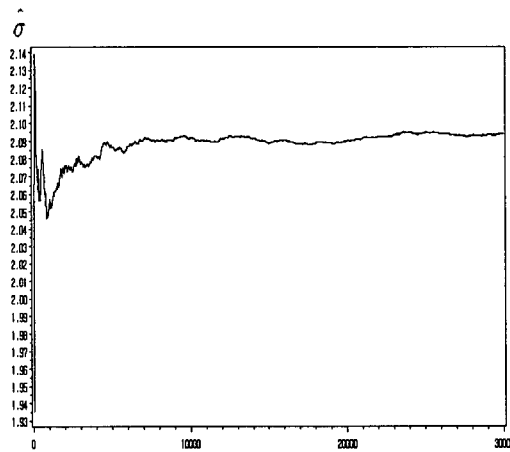
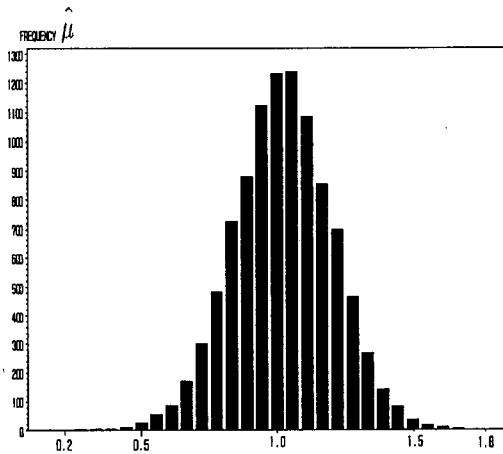
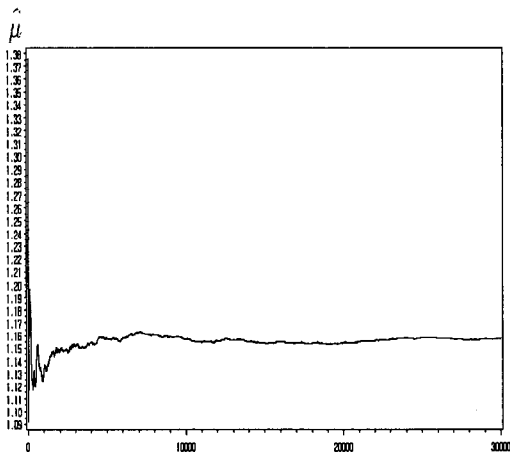
1. Generate ζ from $\zeta_i + Z^*$, where $Z^* \sim N(0, \nu^2)$.
2. Generate u from a Uniform(0, 1).
3. If $u < \pi(\zeta_{i+1}|\mu_{i+1}, \sigma_{i+1}, \mathbf{x}, \mathbf{y}_{i+1})/\pi(\zeta_i|\mu_{i+1}, \sigma_{i+1}, \mathbf{x}, \mathbf{y}_{i+1})$ then $\zeta_{i+1} = \zeta$;
otherwise, $\zeta_{i+1} = \zeta_i$.

After we obtain ζ_{i+1} , we compute ρ_{i+1} by using the relation (2.11), and hence we generate λ_{i+1} by using the relation $\lambda = \rho/(1 - \rho^2)^{1/2}$. The variance, ν^2 , of the normal proposal transition density should be tuned so that the acceptance rate is approximately 0.23.

3. Numerical Examples

Our examples are illustration of extensive studies we have undertaken to validate the MCMC method. We generated n observations from a $SN(\lambda; \mu, \sigma)$ distribution using the algorithm by Henze(1986) and then ran the Gibbs sampler for 30,000 iterations. According to the moments of the distribution(Azzalini 1985), it would appear that the sample mean minus b/k and sample standard deviation over $(1 - (ck)^2)^{1/2}$ are reasonable starting points for μ and σ , where $c = (2/\pi)^{1/2}$, $k = \lambda/(1 + \lambda^2)^{1/2}$, and b is equal to c times the starting point of σ . In an attempt to test the robustness of the sampler, we started λ well away from their true values, i.e. true value of $\lambda+2$.

For the analysis of the MCMC method given in Section 2, the hyperparameter prior specification was defined by $\delta = 0$, $\tau^2 = 100$, $m = 3$, $S = 100$, reflecting rather vague initial vague information relative to that to be provided by the data. The iterative process was monitored by observing trace of the Gibbs samples, where the raw output for the three conditionals (μ, σ, λ) is displayed in the form of time series plots across iterations. Figure 1 shows the ergodic averages of the trace of three parameters, leading us to believe that convergence has been attained before 20,000 iterations. Thus the first 20,000 samples were removed from the output. Figure 2 contains histograms of the marginals of the three parameters, which are all centered about their true values and apparently unimodal. The sample means are not perfect as parameter estimates, because of the amount of skew in the Gibbs samples as shown in Figure 2; however, they produced accurate estimates for the $SN(\lambda; \mu, \sigma)$ distribution.



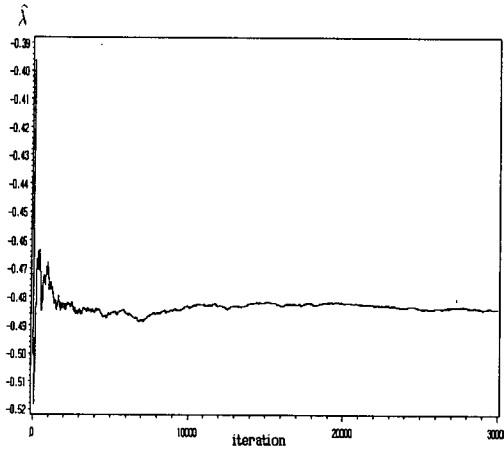


Figure 1. Ergodic averages of $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$.

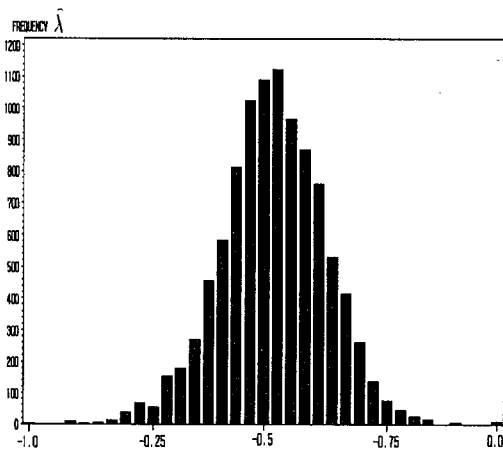


Figure 2. Histograms of $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$.

Table 1. Results of the Posterior Estimates and Standard Deviations

True value (μ, σ, λ)	MCMC result	$n = 20$ $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$	$n = 50$ $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$	$n = 100$ $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$
(1, 2, -.2)	Estimates	(1.1316, 1.8906, -.1958)	(0.9172, 2.0903, -.1994)	(1.0024, 2.0439, -.1979)
	S.D.	(0.3592, 0.2078, 0.1696)	(0.3005, 0.1737, 0.1445)	(0.1186, 0.2023, 0.1118)
(1, 2, .2)	Estimates	(0.8557, 2.0367, 0.1856)	(0.9479, 2.2776, 0.1922)	(1.1304, 2.0569, 0.1936)
	S.D.	(0.4002, 0.2300, 0.1678)	(0.3196, 0.1747, 0.1408)	(0.2035, 0.1198, 0.1129)
(1, 2, .5)	Estimates	(1.3718, 1.8871, 0.4830)	(0.8307, 2.1934, 0.4813)	(1.2600, 1.9094, 0.5097)
	S.D.	(0.3412, 0.2508, 0.2144)	(0.2736, 0.1937, 0.1463)	(0.1775, 0.1384, 0.1168)
(1, 2, -.5)	Estimates	(1.1566, 2.2050, -.4766)	(0.8989, 2.1347, -.4942)	(1.1639, 2.1024, -.4909)
	S.D.	(0.3505, 0.2585, 0.2056)	(0.1942, 0.1498, 0.1116)	(0.1968, 0.1401, 0.1144)
(1, 2, 1)	Estimates	(0.9658, 2.1481, 1.0104)	(1.2515, 2.0635, 0.9591)	(1.0347, 1.9047, 0.9604)
	S.D.	(0.3219, 0.3431, 0.3107)	(0.2245, 0.2024, 0.2053)	(0.1430, 0.1371, 0.1374)
(1, 2, -1)	Estimates	(1.1794, 1.9094, -.9924)	(0.9679, 2.1376, -.9982)	(1.0235, 2.0413, -.9450)
	S.D.	(0.4772, 0.4260, 0.2998)	(0.2292, 0.2339, 0.2142)	(0.1538, 0.1518, 0.1365)
(1, 2, 2)	Estimates	(1.0276, 2.0431, 1.6126)	(1.0236, 1.8520, 1.8145)	(0.9377, 2.0040, 1.9264)
	S.D.	(0.2667, 0.2866, 0.4067)	(0.1630, 0.2199, 0.3541)	(0.1241, 0.1456, 0.2275)
(1, 2, -2)	Estimates	(0.8299, 2.014, -1.8290)	(1.0482, 2.1869, -1.7581)	(1.0654, 1.9293, -1.9358)
	S.D.	(0.3106, 0.3464, 0.8341)	(0.1920, 0.2309, 0.3350)	(0.1186, 0.1893, 0.2156)

Table 1 shows estimation results for various true values of μ, σ, λ and sample size n .

Table 2 contains a comparison of estimates obtained from the suggested MCMC method with those obtained from the method of moments(MM) and the maximum likelihood (ML) method described in Arnold et al.(1993). Estimates of the standard deviations of the ML estimators, based on the estimated inverse of the information matrix(given in Azzalini 1985) are also included in Table 2. But, due to complex distribution of the estimators, the standard deviations of MM estimators are not available. The comparison is based on two simulated data sets of size $n = 100$. Both data sets come from $SN(\lambda; \mu, \sigma)$ distribution are the same data sets used for obtaining Table 1. Data Set 1 is the simulated one for the case of $(\mu, \sigma, \lambda)=(1, 2, .5)$ and Data Set 2 is that for the case of $(\mu, \sigma, \lambda)=(1, 2, -.5)$. An optimization procedure due to Dennis, Gay, and Welsch(1981) implemented in S-PLUS, was used to maximize the likelihood function of (1.1) and to compute profile likelihood functions. To summarize, there is good agreement between the estimates of μ and σ for the three methods(MCMC method, MM and ML method). The values of the estimate of λ , however, disagree with each other. In fact, the values of λ disagree in sign. Thus this study highlights the performance of the suggested MCMC method and

Table 2. Values of Estimates: Quantities in Parentheses are the Standard Deviations

True Parameter	Data Set 1			Data Set 2		
	MCMC	MLE	MME	MCMC	MLE	MME
$\mu=1$	1.2600 (0.1775)	1.2613 (0.1841)	1.2713	1.1639 (0.1968)	1.1264 (0.2418)	1.2793
$\sigma=2$	1.9094 (0.1384)	1.9123 (0.2135)	1.8437	2.1024 (0.1401)	2.1312 (0.2159)	2.1852
$\lambda=0.5$	0.5097 (0.1168)	-0.2546 (0.8421)	-0.1983	-.4909 (0.1144)	0.2341 (0.3261)	-0.1746

reveals the same problem of the frequentist estimation methods(MM and ML method) detected by the simulation study in Arnold et al.(1993).

4. Conclusions

Very little work on the skew normal distribution parameter estimation and inference appears in the literature due to the complex likelihood function. But we have shown that Bayesian computation via MCMC allows us to sample from the distribution of the parameters of the skew normal distribution, by exploiting a particular mathematical representation involving the skew normal density. We have shown how MCMC method can be used to generate Bayesian posterior samples from the parameters of a skew normal 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 example in section 3 demonstrated good performance of the MCMC method. As a minor numerical digression, it should be noted that, for small λ value ($\lambda = .2, -.2$), the standard deviation of the posterior sample is large compared to the estimate. This yields inclusion of zero in a Bayesian confidence interval of λ . The problem can be easily eliminated by introducing constraint $\rho > 0$ (or $\rho < 0$) in (2.10) for the MCMC method. The constraint can be attainable from the shape of the sample distribution.

References

- [1] Arnold, B. C., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q.(1993). The nontruncated marginal of a truncated bivariate normal distribution, *Psychometrika*, 58, 471-478.
- [2] Azzalini, A.(1985). A class of distributions which includes the normal one *Scandinavian Journal of Statistics*, 12, 171-178.
- [3] Azzalini, A.(1986). Further results on a class of distributions which includes the normal ones, *Statistica*, 46, 199-208.
- [4] Azzalini, A. and Capitanio, A.(1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society*, B, 61, 579-602.
- [5] Azzalini, A. and Dalla Valle, A.(1996). The multivariate skew-normal distribution, *Biometrika*, 83, 715-726.
- [6] Branco, M. D.(2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis*, 79, 99-113.
- [7] 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.
- [8] Dennis, J. E., Gay, D. M., and Welsch, R. E. (1981). An adoptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7, 348-384.
- [9] Devroye, L.(1986). *Non-Uniform Random Variate Generation*, New York: Springer Verlag.
- [10] 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. parameter
- [11] Gelfand, A. E., Smith, A. F. M., and Lee, T. M. (1992). Bayesian analysis of constrained and truncated data problems using Gibbs sampling, *Journal of the American Statistical Association*, 87, 523-532.
- [12] 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

- [13] Henze, N.(1986). A probabilistic representation of the skew-normal distribution, *Scandinavian Journal of Statistics*, 13, 271-275.
- [14] Kim, H. J.(2002). Binary regression with a class of skewed t link models, *Communications in Statistics-Theory and Methods*, 1863-1886.
- [15] Lee. P. M. (1997). *Bayesian Statistics*, 2nd ed. New York: John Wiley.
- [16] 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.

[Received March 2005, Accepted April 2005]