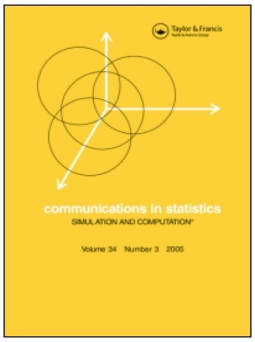

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Bias-corrected estimators of scalar skew normal

Guoyi Zhang & Rong Liu

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2017, VOL. 46, NO. 2, 831–839
http://dx.doi.org/10.1080/03610918.2014.980512

Bias-corrected estimators of scalar skew normal

Guoyi Zhang[a] and Rong Liu[b]

[a]Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA;
[b]Department of Mathematics and Statistics, University of Toledo, Toledo, Ohio, USA

ABSTRACT
One problem of skew normal model is the difficulty in estimating the shape parameter, for which the maximum likelihood estimate may be infinite when sample size is moderate. The existing estimators suffer from large bias even for moderate size samples. In this article, we proposed five estimators of the shape parameter for a scalar skew normal model, either by bias correction method or by solving a modified score equation. Simulation studies show that except bootstrap estimator, the proposed estimators have smaller bias compared to those estimators in literature for small and moderate samples.

ARTICLE HISTORY
Received 27 February 2014
Accepted 21 October 2014

KEYWORDS
Bias-corrected estimators; Bias prevention; Scalar skew normal; Score function; Simulations

MATHEMATICS SUBJECT CLASSIFICATION
62-Statistics; 62F12-Asymptotic properties of estimators

1. Introduction

The skew normal $Y \sim SN(\mu, \sigma, \lambda)$ is a class of distributions that includes the normal distribution ($\lambda = 0$) as a special case. Its density function is as follows:

$$f(y; \lambda, \mu, \sigma) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda \cdot \frac{y-\mu}{\sigma}\right),$$

where ϕ and Φ are the $N(0, 1)$ density and distribution function, parameters μ, σ, and λ regulate location, scale, and shape, respectively. The distribution is positively or negatively asymmetric, in agreement with the sign of λ.

Azzalini (1985, 1986) introduced scalar skew normal problem and derived properties of the scalar skew normal density function. Generalization to the multivariate case are given by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), and Azzalini (2005, 2011). The skew t family has been investigated by Branco and Dey (2001), Azzalini and Capitanio (2003), Gupta (2003) and Lagos-Álvarez and Jiménez-Gamero (2012). Based on the method introduced by Firth (1993), Sartori (2006) investigated bias prevention of the maximum likelihood estimate (MLE) for scalar skew normal and t distribution. If the MLE is subject to a positive bias $b(\lambda)$ (true for skew normal), Firth (1993) suggested shifting the score function $U(\lambda)$ downward by an amount of $U'(\lambda)b(\lambda)$ at each point of λ (illustrated in Fig. 1) to derive a modified score function $U(\lambda) + U'(\lambda)b(\lambda)$. It is proved by Firth (1993) that bias of the MLE could be reduced by modifying the score function.

CONTACT Guoyi Zhang gzhang123@gmail.com Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131-0001.

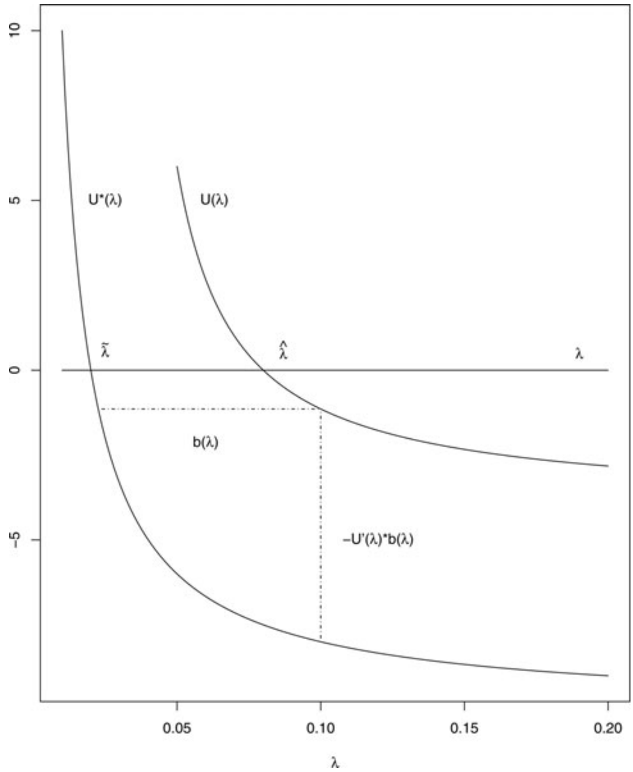

Figure 1. Modifications of the unbiased score function.

Bayes and Branco (2007) developed a simple closed form for the bias correction factor suggested by Sartori (2006) through a rescaled logistic distribution. Azzalini and Arellano-Valle (2013) formulated a general frame work for penalization of the log-likelihood function and proposed maximum penalized likelihood estimate (MPLE) to correct some undesirable behavior of the MLE. Genton (2004) gives a general overview of the skew distributions and their applications.

The existing work of skew normal and t distribution mainly include the bias prevention estimators: Sartori (2006)'s estimator (call $\tilde{\lambda}_1$), Bayes and Branco (2007)'s estimator (call $\tilde{\lambda}_2$), and Azzalini and Arellano-Valle (2013)'s estimator (call $\tilde{\lambda}_3$). With a moderate sample $n = 20$, and shape parameter $\lambda = 10$, the probability that all observations are nonnegative reaches 52.5%, for which MLE $= \infty$ and bias is ∞ as well. For such situations, $\tilde{\lambda}_1$, $\tilde{\lambda}_2$, and $\tilde{\lambda}_3$ provided finite solutions for the shape parameter λ, but with large bias. For example, simulations from Sartori (2006) show that under the setting with $\lambda = 10$, $n = 20$, bias of $\tilde{\lambda}_1$ reached -5.897. Similar results can be found from $\tilde{\lambda}_2$ and $\tilde{\lambda}_3$. The bias prevention estimators work well only for large samples.

In this article, we proposed five estimators for the shape parameter λ from different perspectives: bias correction approach and score function modification approach. This article is organized as follows. In Section 2, we give a background review of Sartori (2006)'s bias prevention estimator, Bayes and Branco (2007)'s approximation estimator and Azzalini and Arellano-Valle (2013)'s MPLE. In Section 3, we propose five estimators. In Section 4, we perform simulation studies and compare the proposed estimators with those reviewed in Section 2. Section 5 gives conclusions.

2. Background

Let $Z_1, Z_2, \ldots, Z_n$ be a random sample from $SN(0, 1, \lambda)$ and let $l(\lambda)$ be the log-likelihood function denoted as

$$l(\lambda) = \text{constant} + \sum_{i=1}^{n} \log\{2\Phi(\lambda Z_i)\}.$$

Let $U(\lambda)$ be the score function of $l(\lambda)$,

$$U(\lambda) = \sum_{i=1}^{n} \frac{\phi(\lambda Z_i)}{\Phi(\lambda Z_i)} Z_i.$$

$U'(\lambda)$ can be derived as follows,

$$U'(\lambda) = -\lambda \sum_{i=1}^{n} \frac{\phi(\lambda Z_i)}{\Phi(\lambda Z_i)} Z_i^3 - \sum_{i=1}^{n} \left(\frac{\phi(\lambda Z_i)}{\Phi(\lambda Z_i)}\right)^2 Z_i^2.$$

Based on Firth (1993), Sartori (2006) modified the usual score equation $U(\lambda) = 0$ by adding an order O(1) term $M(\lambda) = E\{U'(\lambda)b(\lambda)\}$ (the expected value is used to remove the first-order bias of $\hat{\lambda}$), so that the modified score equation is

$$U(\lambda) + M(\lambda) = 0. \tag{1}$$

Sartori's estimator $\tilde{\lambda}_1$ is the solution of Eq. (1) after replacing $M(\lambda)$ by $M_1(\lambda)$ as follows,

$$M_1(\lambda) = -\frac{\lambda}{2} \cdot \frac{a_{42}(\lambda)}{a_{22}(\lambda)},$$

where $a_{kh}(\lambda) = E\left\{Z^k \left(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)}\right)^h\right\}$, and the expected values need to be numerically computed.

Bayes and Branco (2007)'s estimator is the solution of Eq. (1) after replacing $M(\lambda)$ by

$$M_2(\lambda) = -\frac{3\lambda}{2}\left(1 + \frac{8\lambda^2}{\pi^2}\right)^{-1},$$

where $M_2(\lambda)$ is a simple closed form approximation of $M_1(\lambda)$ using a rescaled logistic distribution.

Azzalini and Arellano-Valle (2013) proposed MPLE $\tilde{\lambda}_3$. They replace $M(\lambda)$ in Eq. (1) by

$$M_3(\lambda) = -2C_1C_2\frac{\lambda}{1 + C_2\lambda^2}, \tag{2}$$

where $C_1 = 0.875913, C_2 = 0.856250$. It is easy to see that $M_1(\lambda) \approx M_2(\lambda) \approx M_3(\lambda) = O(\lambda^{-1})$. Hence, the finite solution of λ exists for all of the three methods. It can be shown that for $\tilde{\lambda}_1, \tilde{\lambda}_2$, and $\tilde{\lambda}_3$, $E(\tilde{\lambda}_i - \lambda) = O(n^{-2})$.

3. Bias reduction techniques for scalar skew normal

All the three estimators $\tilde{\lambda}_1, \tilde{\lambda}_2$, and $\tilde{\lambda}_3$ suffer from large bias when sample size is small or moderate. One intuitive way is to estimate the bias and subtract the bias from the estimator. Also notice the systematic negative bias of the three estimators from simulation studies, we propose adjusting the score function to offset the systematic trend. We also examined jackknife and bootstrap bias correction methods for comparison purpose.

3.1. Bias correction for MLE and $\tilde{\lambda}_3$

For a general MLE $\hat{\lambda}$, it is well known that $\hat{\lambda}$ is consistent with asymptotic distribution

$$\sqrt{n}(\hat{\lambda}-\lambda) \xrightarrow{d} N(0, i(\lambda)^{-1}), n \to \infty,$$

where $i(\lambda)$ is the expected Fisher information for a single observation. Consider the second order expression for the mean of the limiting distribution of $\hat{\lambda}$,

$$0 = U(\hat{\lambda}) = U(\lambda) + (\hat{\lambda}-\lambda)U'(\lambda) + \frac{1}{2}(\hat{\lambda}-\lambda)^2 U''(\lambda) + O_p\left(n^{-\frac{1}{2}}\right). \tag{3}$$

Taking expectations through (3), we obtain

$$\begin{aligned} & E(\hat{\lambda}-\lambda)E\left\{U'(\lambda)\right\} + \mathrm{cov}(\hat{\lambda}, U'(\lambda)) \\ & + \frac{1}{2}E(\hat{\lambda}-\lambda)^2 E\{U''(\lambda)\} + \frac{1}{2}\mathrm{cov}\{(\hat{\lambda}-\lambda)^2, U''(\lambda)\} \\ = {} & O\left(n^{-\frac{1}{2}}\right). \end{aligned}$$

Let l_2 be the log-likelihood for one single observation. For convenience, define

$$K_{rs}(\lambda) = E[\{l_2'(\lambda)\}^r \{l_2''(\lambda) + i(\lambda)\}^s].$$

We can show that

$$\begin{aligned} E\{l_2'''(\lambda)\} &= -3K_{11}(\lambda) - K_{30}(\lambda), \\ \mathrm{cov}\{\hat{\lambda}, U'(\lambda)\} &= o(n^{-1}), \end{aligned}$$

and

$$\mathrm{cov}\{(\hat{\lambda}-\lambda)^2, U''(\lambda)\} = o(n^{-1}).$$

For detailed derivation of the above equations in this section, please refer to (Cox and Hinkley, 1974, page 309) and Cox and Snell (1968). Some manipulation, then, gives

$$\begin{aligned} b(\lambda) = E(\hat{\lambda}-\lambda) &= -\frac{K_{11}(\lambda) + K_{30}(\lambda)}{2ni^2(\lambda)} + o(n^{-1}) \\ &= \frac{1}{2} \cdot \frac{\lambda a_{42}(\lambda)}{n a_{22}^2(\lambda)} + o(n^{-1}). \end{aligned}$$

The proposed bias-corrected MLE takes the form of

$$\hat{\lambda}_{\mathrm{bc}} = \hat{\lambda} - b(\hat{\lambda}), \tag{4}$$

with $b(\lambda) = \lambda a_{42}(\lambda)/2n a_{22}^2(\lambda)$. If the MLE doesn't exist, the bias prevention estimator $\tilde{\lambda}_1$ will be used instead.

Now, we consider bias correction of the estimator $\tilde{\lambda}_3$. Recall that $\tilde{\lambda}_3$ is the MPLE proposed by Azzalini and Arellano-Valle (2013). Let

$$U^*(\lambda) = U(\lambda) + M_3(\lambda), \tag{5}$$

where $M_3(\lambda)$ is defined as in Eq. (2). Take derivative of $U^*(\lambda)$, we have

$$U^{*'}(\lambda) = U'(\lambda) + M_3'(\lambda),$$

where $M_3'(\lambda) = -2C_1C_2(1-C_2\lambda^2)/(1+C_2\lambda^2)^2$. It is easy to show that $M_3(\lambda) = O(\lambda^{-1})$ and $M_3'(\lambda) = O(\lambda^{-2})$. Apply Taylor theorem for $U^*(\tilde{\lambda}_3)$ at the neighborhood of λ, we have

$$0 = U^*(\tilde{\lambda}_3) = U^*(\lambda) + U^{*'}(\lambda)(\tilde{\lambda}_3 - \lambda). \tag{6}$$

Replace $U^{*'}(\lambda)$ by $E\{U^{*'}(\lambda)\}$ and use the fact that $ni(\lambda) = -E\{U'(\lambda)\}$, $\tilde{\lambda}_3 - \lambda$ can be expressed as follows,

$$\begin{aligned}\tilde{\lambda}_3 - \lambda &= -\frac{U^*(\lambda)}{E\{U^{*'}(\lambda)\}} \\ &= \frac{U(\lambda)+M_3(\lambda)}{ni(\lambda)-M_3'(\lambda)}.\end{aligned} \tag{7}$$

Use the result in Eq. (7) and take expectation through Eq. (6), we have

$$\begin{aligned}0 &= E\{U^*(\lambda)\} + E\{U^{*'}(\lambda)\}E(\tilde{\lambda}_3-\lambda) + \text{cov}\{U^{*'}(\lambda), \tilde{\lambda}_3 - \lambda\} \\ &= M_3(\lambda) + \{M_3'(\lambda) - na_{22}(\lambda)\}E(\tilde{\lambda}_3 - \lambda) \\ &\quad + \frac{1}{na_{22}(\lambda) - M_3'(\lambda)}\{-n(\lambda a_{42}(\lambda) + a_{33}(\lambda))\}.\end{aligned}$$

Therefore, the bias of $\tilde{\lambda}_3$ is

$$\begin{aligned}E(\tilde{\lambda}_3 - \lambda) &= \frac{\dfrac{\lambda na_{42}(\lambda)+na_{33}(\lambda)}{na_{22}(\lambda)-M_3'} - M_3}{M_3' - na_{22}(\lambda)} \\ &= -\frac{\lambda na_{42}(\lambda) + na_{33}(\lambda) + M_3'M_3 - na_{22}(\lambda)M_3}{(M_3' - na_{22}(\lambda))^2}.\end{aligned}$$

The proposed bias-corrected $\tilde{\lambda}_3$ takes the form of

$$\tilde{\lambda}_{SC} = \tilde{\lambda}_3 - b(\tilde{\lambda}_3), \tag{8}$$

with $b(\tilde{\lambda}_3) = -\left\{\lambda na_{42}(\lambda) + na_{33}(\lambda) + M_3'M_3 - na_{22}(\lambda)M_3\right\}/(M_3' - na_{22}(\lambda))^2$.

3.2. *Adjusted estimator*

Consider Fig. 1, $U(\lambda)$ cross the x-axis when Z_is are with opposite sign numbers ($\hat{\lambda}$ exists); and $U(\lambda)$ approaches x-axis without crossing it when Z_is are all positive or all negative ($\hat{\lambda} = \pm\infty$). For $\hat{\lambda} = \pm\infty$ cases, the bias prevention idea is to shift the score function by an amount of $\{-U'(\lambda)b(\lambda)\}$ to force it cross the x-axis to obtain a finite MLE. From simulation studies, we have noticed systematic negative biases of the three estimators $\tilde{\lambda}_1$, $\tilde{\lambda}_2$, and $\tilde{\lambda}_3$. This means that the amount of shift $\{-U'(\lambda)b(\lambda)\}$ is too large for the three estimators. Therefore, it should be reduced by a certain amount to allow the score function $U(\lambda)$ cross the x-axis but produce less bias. We propose adding $M_4(\lambda)$ to the score function $U(\lambda)$, so that

$$U(\lambda) + M_4(\lambda) = 0, \tag{9}$$

where

$$M_4(\lambda) = -\frac{n}{n+d\lambda}\cdot\frac{\lambda a_{42}(\lambda)}{2a_{22}(\lambda)}.$$

Define a constant c such as

$$c = \sup\{d | U(\lambda^*) + M_4(\lambda^*) = 0, \text{ where } \quad \lambda^* \quad \text{has negative bias}\}. \tag{10}$$

We can see that for any fixed d and λ, $|M_4(\lambda)| < |M_1(\lambda)|$, i.e., the shifted amount $M_4(\lambda)$ of the score function is smaller than that of $\tilde{\lambda}_1$. As $n \to \infty$, $n/(n + d\lambda) \to 1$, hence, $M_4(\lambda) \longrightarrow M_1(\lambda)$. The Eq. (10) indicates that $d \in [0, c]$, and that we are looking for a constant c such that λ^* has smallest negative bias (close to the true value). The proposed adjusted estimator $\tilde{\lambda}_{ad}$ is naturally follows as the solution of Eq. (11),

$$U(\lambda) + M_5(\lambda) = 0, \tag{11}$$

with $M_5(\lambda) = -\dfrac{n}{n + c\lambda} \cdot \dfrac{\lambda a_{42}(\lambda)}{2a_{22}(\lambda)}$. The following theorem can be derived.

Theorem 3.1. *The adjusted estimator $\tilde{\lambda}_{ad}$ has the following properties: (1) $\tilde{\lambda}_{ad}$ has finite solution; (2) Bias$(\tilde{\lambda}_{ad}) = O(n^{-2})$; and (3) $\tilde{\lambda}_{ad}$ converges in probability to Sartori (2006)'s estimator $\tilde{\lambda}_1$ as $n \to \infty$, i.e., $\tilde{\lambda}_{ad} \xrightarrow{p} \tilde{\lambda}_1$.*

Proof. Proof follows from Sartori (2006). □

3.3. Jackknife and bootstrap bias correction

Follow Lagos-Álvarez et al. (2011) for bias correction in the type I generalized logistic distribution, we consider jackknife and bootstrap bias correction. The jackknife was introduced by Quenouille (1949, 1956) to reduce bias of estimators. Shao and Tu (1995) discussed several forms of the jackknife. The bootstrap was introduced by Efron (1990) for estimating the sampling distribution of a statistic and its characteristics. Both jackknife and bootstrap are popularly used since then. In the following, we will consider delete-1 jackknife and bootstrap bias correction of the estimator $\tilde{\lambda}_3$.

Recall that $Z_1, Z_2, \ldots, Z_n$ is a random sample from $SN(0, 1, \lambda)$. Let $\tilde{\lambda}_{3(i)}$ be the solution of the equation

$$U(\lambda) + M_3(\lambda) = 0, \tag{12}$$

with observation Z_i deleted. Define $\bar{\tilde{\lambda}}_3 = \sum_{i=1}^{n} \tilde{\lambda}_{3(i)}/n$. The jackknife bias is defined as $\widehat{bias}_{jack} = (n-1)(\bar{\tilde{\lambda}}_3 - \tilde{\lambda}_3)$ and the jackknife bias-corrected estimator of λ is

$$\tilde{\lambda}_{jack} = \tilde{\lambda}_3 - \widehat{bias}_{jack} = n\tilde{\lambda}_3 - (n-1)\bar{\tilde{\lambda}}_3. \tag{13}$$

For bootstrap bias correction, we use nonparametric bootstrap to approximate the bias of $\tilde{\lambda}_3$. First, we draw B independent bootstrap samples from $Z_1, Z_2, \ldots, Z_n$ with replacement. Let $Z_1^{(i)}, Z_2^{(i)}, \ldots, Z_n^{(i)}$, $i = 1, \ldots, B$, be the ith bootstrap sample, and $\tilde{\lambda}_3^{(i)}$ be the solution of Eq. (12) with the ith bootstrap samples. The bias can be estimated as follows:

$$\widehat{bias}_{boot} = \frac{\sum_{b=1}^{B} \tilde{\lambda}_3^{(i)}}{B} - \tilde{\lambda}_3.$$

The bootstrap bias-corrected estimator of λ is

$$\tilde{\lambda}_{boot} = \tilde{\lambda}_3 - \widehat{bias}_{boot} = 2\tilde{\lambda}_3 - \sum_{b=1}^{B} \tilde{\lambda}_3^{(i)}/B. \tag{14}$$

Table 1. Bias comparison among eight estimators: $\tilde{\lambda}_1$ (Startori, 2006), $\tilde{\lambda}_2$ (Bayes and Branco, 2007), $\tilde{\lambda}_3$ (Azzalini and Arellano-Valle, 2013), $\hat{\lambda}_{bc}$ (bias-corrected MLE), $\tilde{\lambda}_{sc}$ (bias-corrected $\tilde{\lambda}_3$), $\tilde{\lambda}_{ad}$ (adjusted estimator), $\tilde{\lambda}_{jack}$ (jackknife estimator), and $\tilde{\lambda}_{boot}$ (bootstrap estimator).

		Bias Comparison									
λ	n	$\tilde{\lambda}_1$	$\tilde{\lambda}_2$	$\tilde{\lambda}_3$	$\hat{\lambda}_{bc}$	$\tilde{\lambda}_{sc}$	$\tilde{\lambda}_{ad}$	$\tilde{\lambda}_{jack}$	$\tilde{\lambda}_{boot}$	$(\hat{\lambda} < +\infty)$ %	Theoretical %
5	5	− 3.8367	− 3.8378	− 3.8169	1.5487	− 2.8557	− 0.6169	− 3.0074	− 3.8370	29.45	27.70
	10	− 2.9321	− 2.9799	− 2.9317	2.2052	− 1.9755	0.1329	− 1.4557	− 3.0246	49.30	47.73
	20	− 1.7206	− 1.7886	− 1.6813	1.5455	− 0.5224	− 0.0725	0.2866	− 1.9125	74.45	72.68
	50	− 0.2506	− 0.4367	− 0.3167	0.8166	0.7034	0.3035	0.6606	− 0.8950	95.20	96.10
	100	0.0130	− 0.0455	− 0.0139	0.2286	0.6055	0.1164	0.1411	− 0.5138	99.85	99.84
10	5	− 8.7893	− 8.8078	− 8.7728	− 2.0213	− 7.7877	− 4.2517	− 7.8898	− 8.7821	14.70	14.88
	10	− 7.7206	− 7.8064	− 7.6863	− 0.4815	− 6.6637	− 3.4285	− 5.8375	− 7.8352	27.85	27.55
	20	− 5.9499	− 6.0674	− 5.8859	0.7907	− 4.3533	− 2.6139	− 2.8286	− 6.1351	47.65	47.52
	50	− 2.5310	− 2.8728	− 2.5830	0.6848	− 0.3144	− 0.5205	1.3768	− 3.0968	81.40	80.05
	100	− 0.5412	− 0.7596	− 0.5230	0.7716	1.3282	0.2560	0.6022	− 1.3407	95.90	96.02

The last two columns are the estimated percentage of $\hat{\lambda} < \infty$ samples and the theoretical percentage, respectively.

4. Simulation studies

In this section, a small simulation study was conducted to evaluate the five proposed estimators. We consider the shape parameter $\lambda = 5$ and $\lambda = 10$, and generate 2,000 skew normal $SN(\lambda)$ samples with sizes $n = 5, 10, 20, 50,$ and 100. For each generated sample, the following estimators and their bias were computed: $\tilde{\lambda}_1$ (Sartori 2006), $\tilde{\lambda}_2$ (Bayes and Branco 2007), $\tilde{\lambda}_3$ (Azzalini and Arellano-Valle 2013), $\hat{\lambda}_{bc}$ (bias-corrected MLE), $\tilde{\lambda}_{sc}$ (bias-corrected $\tilde{\lambda}_3$), $\tilde{\lambda}_{ad}$ (adjusted estimator), $\tilde{\lambda}_{jack}$ (jackknife bias-corrected estimator), and $\tilde{\lambda}_{boot}$ (bootstrap bias-corrected estimator). The adjusted estimator $\tilde{\lambda}_{ad}$ is calculated as the solution of (9) with $d = 2$, which is found by a comparison of several numbers of d in reducing the bias and was used to approximate the constant c in (10). Empirical mean bias, mean variance, and mean mean square error (MSE) are reported by Tables 1, 2, and 3, respectively. Notice that the three estimators $\tilde{\lambda}_1$, $\tilde{\lambda}_2$, and $\tilde{\lambda}_3$ perform similarly without any noticeable difference in bias and variance.

Tables 1, 2, and 3 show that except bootstrap method, all the four proposals work very well for small and medium samples ($n \leq 20$) in bias reduction. For large samples, the existing methods work better. We also notice that bias correction is more needed for samples with

Table 2. Variance comparison among eight estimators: $\tilde{\lambda}_1$ (Startori, 2006), $\tilde{\lambda}_2$ (Bayes and Branco, 2007), $\tilde{\lambda}_3$ (Azzalini and Arellano-Valle, 2013), $\hat{\lambda}_{bc}$ (bias-corrected MLE), $\tilde{\lambda}_{sc}$ (bias-corrected $\tilde{\lambda}_3$), $\tilde{\lambda}_{ad}$ (adjusted estimator), $\tilde{\lambda}_{jack}$ (jackknife estimator) and $\tilde{\lambda}_{boot}$ (bootstrap estimator).

		Variance Comparison							
λ	n	$\tilde{\lambda}_1$	$\tilde{\lambda}_2$	$\tilde{\lambda}_3$	$\hat{\lambda}_{bc}$	$\tilde{\lambda}_{sc}$	$\tilde{\lambda}_{ad}$	$\tilde{\lambda}_{jack}$	$\tilde{\lambda}_{boot}$
5	5	0.0733	0.0675	0.0781	49.9059	0.1746	32.115	0.4369	0.0778
	10	0.3078	0.2815	0.3002	52.8567	0.7884	39.1468	2.5469	0.3582
	20	1.1035	0.9790	1.1616	42.6702	2.7217	16.1576	8.3396	1.2477
	50	3.3111	2.4887	2.7447	15.4587	5.6216	8.5271	16.6452	2.7521
	100	2.5556	2.2694	2.5770	3.9826	3.7198	3.0563	8.9424	2.4476
10	5	0.0567	0.0569	0.0570	60.5340	0.1472	45.0309	0.3815	0.0625
	10	0.3970	0.2752	0.3695	69.2983	0.9511	47.9203	2.9509	0.4101
	20	1.3551	1.2935	1.5519	65.8202	4.2194	34.7103	10.7657	1.7632
	50	6.6020	5.0796	6.1174	41.4082	14.0655	27.7209	40.6141	7.2765
	100	11.4856	10.5395	10.8033	27.2896	23.5940	20.3222	37.6914	13.2780

Table 3. MSE comparison among eight estimators: $\tilde{\lambda}_1$ (Startori, 2006), $\tilde{\lambda}_2$ (Bayes and Branco, 2007), $\tilde{\lambda}_3$ (Azzalini and Arellano-Valle, 2013), $\hat{\lambda}_{\mathrm{bc}}$ (bias-corrected MLE), $\tilde{\lambda}_{\mathrm{sc}}$ (bias-corrected $\tilde{\lambda}_3$),$\tilde{\lambda}_{\mathrm{ad}}$ (adjusted estimator), $\tilde{\lambda}_{jack}$ (jackknife estimator) and $\tilde{\lambda}_{boot}$ (bootstrap estimator).

		Mean Square Errors Comparison							
λ	n	$\tilde{\lambda}_1$	$\tilde{\lambda}_2$	$\tilde{\lambda}_3$	$\hat{\lambda}_{\mathrm{bc}}$	$\tilde{\lambda}_{\mathrm{sc}}$	$\tilde{\lambda}_{\mathrm{ad}}$	$\tilde{\lambda}_{jack}$	$\tilde{\lambda}_{boot}$
5	5	14.7938	14.7969	14.6470	52.2797	8.3296	32.4795	9.4817	14.801
	10	8.9050	9.1615	8.8955	57.6934	4.6910	39.1449	4.6647	9.5066
	20	4.0637	4.1777	3.9881	46.4340	2.9933	16.1548	8.4176	4.9049
	50	3.3722	2.6783	2.8437	16.1178	6.1136	8.6150	17.0733	3.5518
	100	2.5545	2.2703	2.5759	4.0328	4.0847	3.0684	8.9534	2.7092
10	5	77.3094	77.6356	77.0203	64.5895	60.7964	63.0859	62.6313	77.1880
	10	60.0059	61.2160	59.4488	69.4955	45.3560	59.6515	37.0267	61.8014
	20	36.7557	38.1067	36.1955	66.4125	23.1692	41.5257	18.7614	39.4021
	50	13.0050	13.3305	12.7865	41.8565	14.1574	27.9780	42.4897	16.8631
	100	11.7729	11.1113	11.0714	27.8714	25.3464	20.3776	38.0165	15.0624

large shape parameter. From MSE perspective, only $\tilde{\lambda}_{sc}$ is admissible for small and moderate samples. We think that there is still room to improve $\tilde{\lambda}_{\mathrm{ad}}$. In simulation study, we used $d = 2$ to approximate the constant c defined in (10). Future research may consider looking for a better approximation of the constant c.

5. Conclusions

The difficulty of the shape parameter estimation in a scalar skew normal model lies in the fact that there is a considerable percentage of samples in which MLE goes to infinity. The bias prevention estimators in literature are based on large sample properties, therefore, they don't work well for small and moderate samples. In this research, we have studied this problem from different perspectives, such as bias correction approach and score function modification approach. Simulation studies show that $\hat{\lambda}_{\mathrm{bc}}$ (bias-corrected MLE), $\tilde{\lambda}_{\mathrm{sc}}$ (bias-corrected $\tilde{\lambda}_3$), $\tilde{\lambda}_{\mathrm{ad}}$ (adjusted estimator), and $\tilde{\lambda}_{jack}$ (jackknife bias-corrected estimator) are all effective in reducing bias for small and moderate samples. However, the price paid for reduced bias is the relatively large variance. For scalar skew normal shape parameter estimation, if sample size is large, the existing estimators $\tilde{\lambda}_1$, $\tilde{\lambda}_2$, $\tilde{\lambda}_3$ all work well, there is no need to perform bias correction; if sample size is small or moderate, we suggest using the proposed estimators $\tilde{\lambda}_{\mathrm{sc}}$ since it has smaller bias and MSE.

Acknowledgments

The authors thank the referees for their constructive and insightful comments and suggestions to improve the manuscript.

References

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* 46:199–208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scandinavian Journal of Statistics* 32:159–188.

Azzalini, A. (2011). Skew-normal distribution. *International Encyclopedia of Statistical Sciences* 19:1342–1344.

Azzalini, A., Arellano-Valle, R. B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference* 143:419–433.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B* 61:579–602.

Azzalini, A., Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society: Series B* 65:367–389.

Azzalini, A., Dalla Valle, A. (1996). The multivariate skew normal distribution. *Biometrika* 83:715–726.

Bayes, C. L., Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21:141–163.

Branco, M. D., Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* 79:99–113.

Cox, D. R., Hinkley, D. V. (1974). *Theoretical Statistics*. Boca Raton, *1st ed*. FL: Chapman & Hall/CRC.

Cox, D. R., Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B* 30:248–275.

Efron, B. (1990). More efficient bootstrap computations. *Journal of the American Statistical Association* 85:79–89.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80:27–38.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. *1st ed.* Boca Raton, FL: Chapman & Hall/CRC.

Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics* 37:359–363.

Lagos-Álvarez, B., Jiménez-Gamero, M. D. (2012). A note on bias reduction of maximum likelihood estimates for the scalar skew t distribution. *Journal of Statistical Planning and Inference* 142:608–612.

Lagos-Álvarez, B., Jiménez-Gamero, M. D., Alba Fernández, M. (2011). Bias correction in the type I generalized logistic distribution. *Communication in Statistics-Simulation and Computation* 40:511–531.

Quenouille, M. H. (1949). Problems in plane sampling. *Annals of Mathematical Statistics* 20:355–375.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* 43:353–360.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference* 136:4259–4275.

Shao, J., Tu, D. (1995). *The Jackknife and Bootstrap*. *1st ed.* New York: Springer-Verlag.