Parameter estimation for univariate Skew-Normal distribution based on the modified empirical characteristic function

Gege Hou\textsuperscript{a}, Ancha Xu\textsuperscript{b}, Fengjing Cai\textsuperscript{c}, and You-Gan Wang\textsuperscript{d}

\textsuperscript{a}School of Mechanical Engineering, Northwestern Polytechnical University, China; \textsuperscript{b}Department of Statistics, Zhejiang Gongshang University, China; \textsuperscript{c}College of Mathematics and Information Science, Wenzhou University, China; \textsuperscript{d}School of Mathematical Sciences, Queensland University of Technology, Australia

\textbf{ABSTRACT}

Parameter estimation for the skew-normal distribution is challenging, since the profile likelihood function of shape parameter has a stationary point at zero, which hampers the use of traditional methods, such as maximum likelihood method. We present a modified empirical characteristic function method to perform parameter estimation for the skew-normal distribution. The proposed approach is flexible and easy to implement. We show that the estimators converge to the true values in probability. The simulation study and data analysis suggest that the proposed method performs well, even for the case of small sample size.

\section{1. Introduction}

The skew-normal (SN) distribution, which extends the standard one via adding an extra shape parameter, first appeared in Roberts (1966) and was named by Azzalini Azzalini (1985). Azzalini and Capitanio (1999) demonstrated that it is rather practical to model real data sets with skewness, since the SN distribution possesses prominent properties in term of mathematical tractability. The probability density function of a SN random variable $X$ is

$$f_X(x; \alpha) = 2\phi(x)\Phi(\alpha x), x, \alpha \in \mathbb{R},$$

where $\alpha$ is a shape parameter that controls the shape of the density function. Here $\phi(x)$ and $\Phi(x)$ denote the probability density function and the cumulative distribution function of the standard normal distribution, respectively. If $\alpha = 0$, the distribution becomes the standard normal distribution; The distribution is left skewed if $\alpha > 0$, and is right skewed if $\alpha < 0$. Apart from that, the larger the absolute value of $\alpha$ is, the larger the skewness is. When $\alpha$ goes to infinity, the SN distribution will converge to the half-normal (or folded normal) distribution. It is general to consider a linear transformation, $Y = \omega X + \xi$. Then the transformed variable $Y$ follows a skew-normal distribution with
a parameter set \( \theta = (\xi, \omega, \alpha) \), one may write \( Y \sim SN(\theta) \), and the probability density function of \( Y \) is

\[
f(y; \xi, \omega, \alpha) = \frac{2}{\omega} \phi\left( \frac{y - \xi}{\omega} \right) \Phi\left( \frac{\alpha}{\omega} \frac{y - \xi}{\omega} \right), y, \xi \in \mathbb{R}, \omega > 0,
\]

where \( \xi \) and \( \omega \) are the location and scale parameters, respectively.

However, in many cases, maximum likelihood (ML) inference on the skew-normal parameters poses some difficulties. Firstly, two of the elements of the gradient of the log-likelihood may become linearly dependent as \( \alpha \) goes to zero. Secondly, the profile likelihood function of \( \alpha \) has a stationary point at zero, which is independent of the observed sample. The two problems result in singular Fisher information matrix and inconsistent estimators of \( \alpha \). Researchers have developed many methodologies to solve the two problems. For example, Azzalini (1985) proposed a parametrization method to avoid singular Fisher information matrix, which can also be seen in Chiogna (1998) and Pewsey (2000). Azzalini and Capitanio (1999) stated that the second obstacle is challenging. Azzalini and Arellano (2013) proposed the maximum penalized likelihood estimation (MPLE) by introducing an appropriate penalty function

\[
Q(\alpha) = c_1 \log (1 + c_2 \alpha^2),
\]

where \( c_1 = 0.875913 \) and \( c_2 = 0.856250 \). The modified log-likelihood function is

\[
l_p(\theta) = l(\theta) - Q(\alpha),
\]

where \( l(\theta) \) is the log-likelihood function. This method works well only for large sample size. According to Azzalini and Arellano (2013), this method results in infinite estimators for shape parameter \( \alpha \) for small to moderate sample sizes. Besides, the moment method (MM) can also be applied to estimate \( \alpha \) by inverting the skewness equation based on sample skewness. The estimator of the transformation of \( \alpha \) is

\[
\hat{\alpha} = \frac{\hat{\delta}}{\sqrt{1 - \hat{\delta}^2}}.
\]

Notwithstanding it seems a reasonable way to represent the shape parameter, the method of moments shows even worse results. Because it could not find a solution for Eq. (4) when the sample skewness is large and the transformation \( |\hat{\delta}| \geq 1 \). Alternatively, by setting \( \hat{\delta} = 1 \) in Eq. (3), the maximum theoretical skewness is \( \hat{\gamma} \approx 0.9952717 \). Thus, let

\[
|\hat{\gamma}| = \min\left\{ 0.99, \frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_i - \bar{y}}{s} \right)^3 \right\}
\]
be the sample skewness, where $\bar{y}$ and $s$ are the mean and standard deviation, respectively. The estimators of the location parameter and scale parameter are

$$\hat{\xi} = \bar{y} - \sqrt{\frac{2}{\pi}} \cdot \hat{\delta}, \hat{\omega} = \sqrt{\frac{s^2}{1 - \frac{2}{\pi} \cdot \hat{\delta}^2}}.$$

From Eq. (3), we notice that the efficiency of the MM is affected by the sample skewness. When a sample is extremely skew, the estimator of $\alpha$ will deviate far from its true value, which results in lowering estimation efficiency of the MM. Furthermore, the estimation accuracy of the MM is not high for a small sample as well.

The multivariate skew-normal distribution was generalized by Azzalini and Dalla (1996) and Azzalini and Capitanio Azzalini and Capitanio (1999). Unlike multivariate normal, where the parameters estimates can be obtained using the sample mean and sample variance-covariance matrix, there are some obstacles we will meet when we estimate multivariate skew-normal parameters. Similar to the univariate version, the estimators of the MPLE for a multivariate skew-normal distribution is not robust as well. Although the probability density function is asymmetric, Azzalini, Genton, and Scarpa (2010) introduced estimating equations for a multivariate skew-normal case based on its invariance. However, it is quite problematic to estimate parameters when we actually implement this method for the multivariate skew-normal distribution because of high multiplicity of its solutions.

In this paper, we approach the problem of parameter estimation in a manner that is different from the aforementioned methods. We introduce the modified empirical characteristic function (MECF) method to estimate the model parameters. A major motivation for this work is that the empirical characteristic function (ECF) method is a conceptually simple, flexible, and effective parameter estimation technique for which the only major drawback was the estimator may be far from optimal, as its error terms $\varepsilon(t, \theta)$ cannot be justified to follow a normal distribution. Hence, we will modify the distance function by the idea of M-estimator, which in turn improves estimation efficiency in small sample cases, and minimize the modified distance function. As we will demonstrate, it is more robust to the initial values than the original ECF method. The proposed new estimators converge to the true values in probability as well. For multivariate data, the MECF gives a way to test symmetry under a multivariate skew-symmetry model as well as provides a new direction to estimate its parameters.

The rest of this paper is arranged as follows. Section 2 reviews the method of ECF and introduce the proposed method concretely. In Section 3, the simulation will appraise the performance of the proposed method under different sample sizes. We analyze a data set for illustration in Section 4. Finally, we present conclusions of the paper and show a brief exploration of the proposed method for estimate parameters for multivariate skew-normal distribution.

2. Proposed algorithm

2.1. Review of the ECF method

The ECF method shares the same mechanism with the empirical moment generating function method, as it replaces the moment generating function by the characteristic
function. It was previously developed by Feuerverger and Mureika (1977) and Heathcote (1977), and applied in parameters estimation by Tran (1998) because of two essential advantages of characteristic functions. First, characteristic functions are uniformly bounded, thus the resulting solution is numerically stable. Second, when moment generating functions of distributions does not exist, it is appropriate to adopt the characteristic function for heavy tail distribution.

The characteristic function of an arbitrary distribution with the probability density function \( f(y; \theta) \) is written as

\[
c(t, \theta) = \int e^{ity} f(y; \theta) dy,
\]

where \( i \) is an imaginary unit and \( \theta \) represents the unknown parameter vector of the distribution. If \( \{y_1, y_2, ..., y_n\} \) is a random set of sample from the distribution, the ECF is defined as

\[
C_n(t) = \frac{1}{n} \sum_{i=1}^{n} \exp \{ity_i\} = \frac{1}{n} \sum_{i=1}^{n} (\cos ty_i + i \sin ty_i).
\]

We can separate the characteristic function into two parts and unite them as a vector

\[
F(t, \theta) = (\text{Re}[c(t_1, \theta)], \ldots, \text{Re}[c(t_m, \theta)], \text{Im}[c(t_1, \theta)], \ldots, \text{Im}[c(t_m, \theta)])',
\]

where \( t_1, \ldots, t_m \) are \( m \) fixed grid points, \( \text{Re}[c(t, \theta)] \) and \( \text{Im}[c(t, \theta)] \) denote the real part and imaginary part of \( c(t, \theta) \), respectively. The counterpart of the ECF \( C_n(t) \) is

\[
Z_n(t) = (\text{Re}[C_n(t_1)], \ldots \text{Re}[C_n(t_m)], \text{Im}[C_n(t_1)], \ldots, \text{Im}[C_n(t_m)])',
\]

where \( \text{Re}[C_n(t_k)] = \frac{1}{n} \sum_{i=1}^{n} \cos t_k y_i \) and \( \text{Im}[C_n(t_k)] = \frac{1}{n} \sum_{i=1}^{n} \sin t_k y_i, k = 1, \ldots, m, \) are real and imaginary parts, respectively.

Referring to the results of Feuerverger and McDunnough (1981), we have

\[
n^{1/2} [Z_n(t) - F(t, \theta)] \xrightarrow{d} N(0, \Omega_\theta),
\]

where

\[
\Omega_\theta = \begin{pmatrix}
\Omega_{RR} & \Omega_{RI} \\
\Omega_{IR} & \Omega_{II}
\end{pmatrix}
\]

is the variance-covariance matrix with the \((p, q)\)th elements in each partition

\[
(\Omega_{RR})_{pq} = \frac{1}{2} \{\text{Re}[c(t_p + t_q, \theta)] + \text{Re}[c(t_p - t_q, \theta)]\} - \text{Re}[c(t_p, \theta)] \cdot \text{Re}[c(t_q, \theta)],
\]

\[
(\Omega_{RI})_{pq} = (\Omega_{IR})_{qp} = \frac{1}{2} \{\text{Im}[c(t_p + t_q, \theta)] - \text{Im}[c(t_p - t_q, \theta)]\} - \text{Re}[c(t_p, \theta)] \cdot \text{Im}[c(t_q, \theta)],
\]

\[
(\Omega_{II})_{pq} = \frac{1}{2} \{\text{Re}[c(t_p - t_q, \theta)] - \text{Re}[c(t_p + t_q, \theta)]\} - \text{Im}[c(t_p, \theta)] \cdot \text{Im}[c(t_q, \theta)].
\]

If we set \( e(t, \theta) = Z_n(t) - F(t, \theta) \), this problem can be regarded as the weighted least-square, where \( Z_n(t) \) is assumed as a dependent variable and \( F(t, \theta) \) is assumed as a right-side explanatory function. Thus, for a set of grid points \( t_1, \ldots, t_m \), the optimal estimator of \( \theta, \hat{\theta} \), will be finalized simply by minimizing
Note that $\Omega_\theta^{-1}$ would dramatically increase computational complexity while we optimize Eq.(7) directly. Therefore, we consider to replace $X/C0h$ by its consistent estimator, $\hat{X}_n$, whose form is

$$\hat{X}_n = \left( \begin{array}{ccc} X_{RR} \quad X_{RI} \quad X_{II} \\ X_{IR} \quad X_{II} \quad X_{II} \\ X_{II} \quad X_{II} \quad X_{II} \end{array} \right)$$

with the $(p, q)$th elements in each partition being

$$(\Omega_{RR})_{pq} = \frac{1}{2} \{ \text{Re}[C_n(tp + tq)] + \text{Re}[C_n(tp - tq)] \} - \text{Re}[C_n(tp)] \cdot \text{Re}[C_n(tq)],$$

$$(\Omega_{RI})_{pq} = \frac{1}{2} \{ \text{Im}[C_n(tp + tq)] - \text{Im}[C_n(tp - tq)] \} - \text{Re}[C_n(tp)] \cdot \text{Im}[C_n(tq)],$$

$$(\Omega_{II})_{pq} = \frac{1}{2} \{ \text{Re}[C_n(tp - tq)] - \text{Re}[C_n(tp + tq)] \} - \text{Im}[C_n(tp)] \cdot \text{Im}[C_n(tq)].$$

Taking the results of Feuerverger and Mcdunnough (1981) and the Delta-method into account, we can obtain that the ECF estimators of $\theta, \hat{\theta}$, is consistent and asymptotically normal with asymptotic covariance

$$\frac{1}{n} \left( A' \Omega_\theta^{-1} A \right)^{-1},$$

where $A = \frac{\partial E(t, \theta)}{\partial \theta}$. However, the efficiency of the ECF is momentously affected by the choice of $\{t_1, ..., t_m\}$ as $m$ approaches the number of unknown parameters. Schmidt (1982) has proved that the asymptotic efficiency of the EFC will generally improve as $m$ increases, but will converge to a constant while $m$ is large enough. In fact, Eq. (10) can be close to the Cramer-Rao lower bound arbitrarily when $m$ is sufficiently large. Furthermore, based on the analysis of Feuerverger and Mcdunnough (1981) and Schmidt (1982), grid points $t_1, ..., t_m$ must distinguish from each other. Given a fixed $m$, we adopt the equal interval, $\tau$, for the grid points, $t_i = i\tau, i = 1, 2, ..., m$. Specifically, $\tau$ is an arbitrary real constant and is obtained by minimizing the size of the asymptotic covariance matrix Eq. (10). Thus, one way is to initialize $\tau$ at the preliminary values of the unknown parameters and iterate them together, since the optimal $\tau$ only bonds with the model parameters.

### 2.2. The modified empirical characteristic function

Assume that $y_1, y_2, ..., y_n$ are sampled from the $SN(\theta)$ with the characteristic function

$$c(t, \theta) = \int e^{iyf(y; \theta)}dy = \exp \left\{ it\xi - \frac{(t\delta)^2}{2} \right\} \cdot \left[ 1 + i\text{Efri} \left( \frac{\delta t\xi}{\sqrt{2}} \right) \right],$$

where $\delta = \frac{x}{\sqrt{1+x^2}}$ and $\text{Efri}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is imaginary error function (N.J.A 2007). The partial derivatives of the characteristic function (Feuerverger and Mcdunnough 1981) with respect to the three parameters are
residuals. Rey (1983) has proposed many different M-estimator, derived by Peter and Huber (2011), is regarded as a desirable method in achieve a tradeoff between efficiency and stability in the presence of infinite estimates. justified to follow a normal distribution. Many robust estimators have been proposed to

\[
\frac{\partial c(t, \theta)}{\partial \hat{c}} = \exp \left\{ \frac{it \hat{c} - (t \cdot \omega)^2}{2} \right\} \cdot \left[ \frac{it - t \cdot \text{Efri} \left( \frac{\delta \theta t}{\sqrt{2}} \right)}{2} \right],
\]

\[
= t \cdot \exp \left\{ -\frac{(t \cdot \omega)^2}{2} \right\} \cdot \left[ \left[ -\sin t \hat{c} + \cos t \hat{c} \cdot \text{Efri} \left( \frac{\delta \theta t}{\sqrt{2}} \right) \right] - i \left[ \cos t \hat{c} + \sin t \hat{c} \cdot \text{Efri} \left( \frac{\delta \theta t}{\sqrt{2}} \right) \right] \right\},
\]

\[
\frac{\partial c(t, \theta)}{\partial \omega} = \exp \left\{ \frac{it \hat{c} - (t \omega)^2}{2} \right\} \cdot t^2 \cdot \left[ -1 - i \text{Efri} \left( \frac{\delta \theta t}{\sqrt{2}} \right) \right] + \exp \left\{ \frac{it \hat{c} - (t \omega)^2}{2} \right\} \left[ \frac{it - t \sqrt{2}}{\sqrt{\pi}} \exp \left\{ \frac{\delta \theta t^2}{2} \right\} \right].
\]

\[
\frac{\partial c(t, \theta)}{\partial x} = \frac{\sqrt{2}}{\sqrt{\pi} \cdot \text{Efri} \left( \frac{\omega}{\sqrt{2}} \right)} \cdot \left\{ (\cos t \hat{c} - (t \omega)^2) \right\} (1 + x^2)^{-\frac{1}{2}} (-\sin t \hat{c} + i \cos t \hat{c}).
\]

The rest of the estimation process is the same as the previous subsection 2.1.

Although the ECF method could estimate parameters of the SN distribution, the estimator may be far from optimal, as the distribution of its error terms \(\varepsilon(t, \theta)\) cannot be justified to follow a normal distribution. Many robust estimators have been proposed to achieve a tradeoff between efficiency and stability in the presence of infinite estimates. M-estimator, derived by Peter and Huber (2011), is regarded as a desirable method in regression modeling and is robust to departure from the normality assumption. The fundamental idea of M-estimator is to replace the squared residuals by a function \(\rho()\) of residuals. Rey (1983) has proposed many different \(\rho\) functions, and all \(\rho\) functions have the following properties.

**Assumption 1.** Let \(\rho()\) be a real function satisfying the following properties:

1. \(\rho(0) = 0\) and \(\rho(x) = \rho(-x)\);
2. if \(0 < x_1 < x_2\), then \(\rho(x_1) \leq \rho(x_2)\);
3. \(\rho\) is continuously differentiable.

Let

\[
\varepsilon^\ast(t, \theta) = \Omega_n^{-1/2} \varepsilon(t, \theta) = (\varepsilon_1^\ast(\theta), ..., \varepsilon_{2m}^\ast(\theta))^\prime.
\]

We consider using function \(\rho()\) to eliminate the effects of large absolute values in \(\varepsilon^\ast(t, \theta)\). Consequently, the estimation of \(\theta\) can be obtained by minimizing

\[
\frac{1}{2m} \sum_{i=1}^{2m} \rho(\varepsilon_i^\ast(\theta)).
\]

In practice, the Tukey biweight function, proposed by Yohai and Zamar (1988), is the most ubiquitous \(\rho\) function, defined as

\[
\rho(u) = \begin{cases} 
\frac{u^2}{2} \left( 1 - \frac{u^2}{c^2} + \frac{u^4}{3c^4} \right), & \text{if } |u| \leq c, \\
\frac{c^2}{6}, & \text{if } |u| > c,
\end{cases}
\]

where \(c\) is the tuning parameter. Because the M-estimator with the Turkey biweight function has an efficiency of 95% under independent Gaussian error by setting \(c = 4.6851\). The following theorem gives rise to the asymptotic properties of \(\hat{\theta}\) obtained by minimizing Eq. (12). The proof of theorem 1 is similar to the proof of theorem 1 in Chen, Ye, and Zhao (2017).
Theorem 1. Let $t_1, t_2, \ldots, t_m$ be discrete fixed grid points. The estimator of $\theta$ based on Equation (12), $\hat{\theta}$, converges to the true value $\theta_0$ in probability.

As for confidence intervals of parameters, we would meet obstacles if we directly construct them based on Eq. (12). Alternatively, resampling algorithms, including the bootstrap resampling method, are frequently adopted for intervals estimates. DiCiccio et al. (1996) proved that bootstrap is asymptotically more accurate than the standard intervals obtained using sample variance and assumptions of normality. Efron and Tibshirani (1986) exhibited that bootstrap percentile confidence intervals have invariant properties. Thus, we perform interval estimation using the bootstrap percentile confidence interval for the parameters of the SN distribution. The procedures are as follows.

Step 1 For the given sample size $n$, draw $B$ samples with replacement from the observed data $y_1, y_2, \ldots, y_n$.

Step 2 For each bootstrap sample, estimate the parameters by minimising Expression (Feuerverger and Mureika 1977).

Step 3 Let $\hat{\theta}^*_{\beta/2}$ and $\hat{\theta}^*_{1-\beta/2}$ be the $\beta/2$ and $1 - \beta/2$ quantiles among the $B$ bootstrap estimates, where $\beta$ is the given significance level. Then the $1 - \beta$ interval estimate of $\theta$ is

$$\left(\hat{\theta}^*_{\beta/2}, \hat{\theta}^*_{1-\beta/2}\right).$$

3. Simulation

This section gives a simulation to investigate the performance of different methods for the SN parameters. Specifically, the proposed method in this article is compared with the MPLE and the MM. In the proposed method, we choose the Tukey biweight function defined in Expression (Heathcote 1977) as the $\rho$ function, where $c = 4.6851$. Besides, we use the MM estimators as the initial value for performing the MEFC method. Due to the randomness, not all eigenvalues of $\hat{\Omega}_n$ are positive. If such a case is encountered, we just let the negative eigenvalues as 0, then perform the optimal computation of (Feuerverger and Mureika 1977) to obtain $\hat{\theta}$. Assume that $y_1, y_2, \ldots, y_n$ from $\text{SN}(\xi, \omega, \alpha)$. Without loss of generality, we assume $\xi = 0$ and $\omega = 1$. The shape parameter $\alpha$ is chosen as $-3, -2, -1, 1, 2$ and 3, respectively. And the sample size is selected as $n = 50, 100$ and 200, respectively. We generate 2000 times with different sample sizes $n$. The comparisons are based on the absolute bias and the mean squared error (MSE). We set the eigenvalue as 0, if it is less than 0.

First, we need to determine the optimal number of grid points, $m$, under the condition of different values of $\alpha$. Figures 1 and 2 show that when $n = 50$, the size of grid points has no significant effect on the estimation of the location parameter $\xi$ and scale parameter $\omega$. As for the shape parameter $\alpha$, the lines of biases and MSE fluctuate slightly when $m \leq 15$, and keep steady when $m \geq 20$. When $m = 10$, the biases and MSEs become stable. The case of $n = 100$ and 200 show the similar phenomena, and the figures are shown in the Supplemental file. Thus, for balancing the efficiency of estimation and computation, we choose the number of grid points $m = 10$.

Then, with optimal $m$, we compare the proposed method with the MPLE and the MM. Azzalini Azzalini (1985) stated that compared with the shape parameter $\alpha$, the
location parameter $\xi$ and the scale parameter $\omega$ are easier to be estimated. This conclusion can also be verified in the Figures 3 and 4 because of small biases and MSEs in all the scenarios. Thus, we just utilize the results of the shape parameter $\alpha$ to assess location parameter $\xi$ and the scale parameter $\omega$ are easier to be estimated. This conclusion can also be verified in the Figures 3 and 4 because of small biases and MSEs in all the scenarios. Thus, we just utilize the results of the shape parameter $\alpha$ to assess
Figure 2. When the sample size $n = 50$ and shape parameter $\alpha = 1, 2$ and $3$, the number of grid points $m$ versus the estimation efficiency in terms of abstract biases (left panel) and MSEs (right panel). The different colors and shapes of lines represent estimated results of the parameters, that is, location parameter $\xi$ by a green dot line, scale parameter $\omega$ by a red dashed line, and shape parameter $\alpha$ by a solid black line.
Figure 3. Comparison of estimation efficiency among the MECF, the MPLE and the MM according to sample size $n = 50$ and the shape parameter $\alpha$ with different values. The different colors and shapes of lines represent the results of different methods, that is, the MM by green dot line, the MPLE by red dashed line, and the MECF method by solid black line.
Figure 4. Comparison of estimation efficiency among the MECF, the MPLE and the MM according to sample size $n = 100$ and the shape parameter $\alpha$ with different values. The different colors and shapes of lines represent the results of different methods, that is, the MM by green dot line, the MPLE by red dashed line, and the MECF method by solid black line.
estimation efficiency. As can be seen from Figures 3 and 4, the MSEs of the MM are much larger in all the case comparing with the other two methods, although it has small biases, especially for the case of the small sample size. Thus, the estimation efficiency of the MM is the lowest one. MPLE estimates being infinite is increasing as the sample size becomes small, which can be seen in the Table 1. Thus, the results obtained by MPLE may be non-convergent. Moreover, even if we take out these infinite values, the MPLE still works worse than the MEFC in terms of biases and MSEs, but better than the MM. However, the proposed method, the MEFC, performs the best in overall, since its results are closer to the true value and show the smallest MSEs, even for the small sample size. When the sample size \( n = 200 \), the MPLE also performs well, which proves the conclusion of Azzalini and Arellano (2013), and the results are listed in the Supplemental file. Therefore, the MEFC has a better performance comparing with the other two methods.

### 4. An application

In this section, we compare the performance of the bootstrap percentile confidence interval based on the MEFC method, the MPLE method and the MM. We consider the dataset downloaded from [http://azzalini.stat.unipd.it/SN/frontier.dat](http://azzalini.stat.unipd.it/SN/frontier.dat). It consists of simulated data from \( \text{SN}(0, 1, 5) \) with \( n = 50 \). The ML estimation is not proper since the profile likelihood of \( \alpha \) doesn’t have a close-form equation. Some researchers have approached this problem in different methods. Sartori (2006) obtained \( \hat{\alpha} = 9.14 \) by using the corrected ML method. Azzalini and Capitanio (1999) used the graphical approach to fit the data by setting \( \alpha = 8.1 \). Liseo and Loperfido (2006) utilized the reference priors to estimate \( \alpha \). But its posterior does not have finite mean and its median, which is about 15.9, is regarded as the reasonable point estimate, with the 95% highest posterior density region given by (4.2, 52.5).

| True value of \( \alpha \) | \( n = 50 \) | \( n = 100 \) | \( n = 200 \) |
|--------------------------|-------------|-------------|-------------|
| \( -3 \)                | 0.194       | 0.074       | 0.004       |
| \( -2 \)                | 0.058       | 0.03        | 0           |
| \( -1 \)                | 0.01        | 0           | 0           |
| \( 1 \)                 | 0.008       | 0           | 0           |
| \( 2 \)                 | 0.072       | 0.012       | 0           |
| \( 3 \)                 | 0.138       | 0.074       | 0.002       |

### Table 1. Proportions of infinity estimators given by the MPLE under different sample sizes.

| Method | Parameters | Mean  | SD    | 95% confidence interval |
|--------|------------|-------|-------|-------------------------|
| MEFC   | \( \zeta \) | -0.054| 0.005 | (-0.492, 1.004)         |
|        | \( \omega \) | 1.200 | 0.005 | (0.994, 1.738)          |
|        | \( \alpha \) | 5.530 | 2.239 | (2.366, 9.684)          |
| MPLE   | \( \zeta \) | -0.096| 7.6e-04| (-0.103, 0.019)        |
|        | \( \omega \) | 1.520 | 0.028 | (1.194, 1.852)          |
|        | \( \alpha \) | 1.504  | 8.988e+14| (6.929, 1.024e+08)   |
| MM     | \( \zeta \) | -0.059| 0.138 | (-0.166, 0.114)        |
|        | \( \omega \) | 1.385  | 0.041 | (1.057, 1.331)          |
|        | \( \alpha \) | 5.557  | 2.218 | (2.891, 7.018)          |

### Table 2. Comparison of confidence intervals constructed by the MEFC, the MPLE and the MM for parameters of the SN distribution.
We combine the MEFC with the bootstrap method to construct the percentile confidence intervals of the parameters, by setting number of bootstrap replications $B = 2000$. The results are shown in Table 2. From Table 2, we can see that the bootstrap percentile confidence intervals of the parameters based on the MECF cover all the true values, and the lengths of intervals are short. However, the MM and the MPLE fail to achieve this results. Therefore, the MEFC method outperforms the MM and the MPLE in this example.

5. Conclusion

We have presented the MEFC method to estimate parameters of the SN distribution and construct their confidence intervals based on bootstrap. By leveraging $\rho()$ function, the modified ECF is extrapolated to more complex situations, where the distribution of its error terms $\varepsilon(t, \theta)$ cannot be justified to follow a normal distribution, thereby improving estimation accuracy even for small to mediate sample sizes. By comparison with the MPLE and the MM, the MEFC works well no matter how small sample size is. Data analysis also shows that the MEFC performs well.

As for multivariate versions, the MEFC method could be a way to test symmetry under multivariate skew-symmetry models and explore new direction to estimate their parameters. It is well-known that the characteristic function of a symmetric distribution has only the real part. Then the multivariate data could be projected onto the direction of maximal skewness (Loperfido 2018) and the characteristic function of this univariate projection could be used to test its skewness. If the corresponding p-value were high enough, then the sampled multivariate distribution could be taken as symmetric. Combined with the method in Jiménez-Gamero et al. (2016), the parameters of multivariate skew-normal distribution could be obtained. Any skewed and multivariate probability density function can be expressed as the product of a symmetric probability density function and another function, the proposed method provides a possible new way to test symmetry under a multivariate skew-symmetry model.

Funding

The research is supported by National Natural Science Foundation of China (11671303) and Foundation of Zhejiang Province(LY19A010014).

ORCID

You-Gan Wang http://orcid.org/0000-0003-0901-4671

References

Azzalini, A. 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12 (2):171–78.
Azzalini, A., and A. Capitanio. 1999. Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (3): 579–602. doi:10.1111/1467-9868.00194.
Azzalini, A., and A. Dalla. 1996. Valle. The multivariate skew-normal distribution. *Biometrika* 83 (4):715–26. doi:10.1093/biomet/83.4.715.

Azzalini, A., and R. B. Arellano. 2013. Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference* 143 (2):419–33. doi:10.1016/j.jspi.2012.06.022.

Azzalini, A., M. Genton, and B. Scarpa. 2010. Invariance-based estimating equations for skew-symmetric distributions. *Metron* - *METRON* 68 (3):275–98. doi:10.1007/BF03263540.

Chen, P., Z. Ye, and X. Zhao. 2017. Minimum distance estimation for the generalized pareto distribution. *Technometrics* 59 (4):528–41. doi:10.1080/00401706.2016.1270857.

Chiojna, M. 1998. Some results on the scalar skew-normal distribution. *Journal of the Italian Statistical Society* 7 (1):1–13. doi:10.1007/BF03178918.

DiCiccio, T. J., B. Efron, P. Hall, M. A. Martin, A. J. Canty, A. C. Davison, D. V. Hinkley, L. J. Gleser, S. M. S. Lee, G. A. Young, et al. 1996. Bootstrap confidence intervals. *Statistical Science* 11 (3):189–212. doi:10.1214/ss/1032280214.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1 (1):54–75. doi:10.1214/ss/1177013815.

Feuerverger, A., and P. Mcdunnough. 1981. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society: Series B (Methodological)* 43 (1):20–7. 01 doi:10.1111/j.2517-6161.1981.tb01143.x.

Feuerverger, A., and R. Mureika. 1977. The empirical characteristic function and its applications. *Annals of Statistics* 5 (1):87–97.

Heathcote, C. R. 1977. Integrated mean square error estimation of parameters. *Biometrika* 64 (2):255–64. doi:10.1093/biomet/64.2.255.

Jiménez-Gamero, M. D., M. V. Alba-Fernández, P. Jodra, and Y. Chalco Cano. 2016. Testing for the symmetric component in skew distributions. *Mathematical Methods in the Applied Sciences* 39 (16):4713–22. doi:10.1002/mma.3410.

Liseo, B., and N. Loperfido. 2006. A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* 136 (2):373–89. doi:10.1016/j.jspi.2004.06.062.

Loperfido, N. 2018. Skewness-based projection pursuit: A computational approach. *Computational Statistics & Data Analysis* 120:42–57. doi:10.1016/j.csda.2017.11.001.

N.J.A, S. 2007. The on line encyclopedia of integer sequences. *Notices of the American Mathematical Society* 41 (8):219–34.

Peter, J. Huber. 2011. *Robust Statistics*, 1248–51. Berlin, Heidelberg: Springer.

Pewsey, A. 2000. Problems of inference for azzalini’s skewnormal distribution. *Journal of Applied Statistics* 27 (7):859–70. doi:10.1080/02664760050120542.

Rey, W. J. J. 1983. Type M estimators. In *Introduction to Robust and Quasi-Robust Statistical Methods*, 90–116. Berlin, Heidelberg: Springer Berlin Heidelberg.

Roberts, C. 1966. A correlation model useful in the study of twins. *Journal of the American Statistical Association* 61 (316):1184–90. doi:10.1080/01621459.1966.10482202.

Sartori, N. 2006. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference* 136 (12):4259–75. doi:10.1016/j.jspi.2005.08.043.

Schmidt, P. 1982. An improved version of the quandt-ramsey mgf estimator for mixtures of normal distributions and switching regressions. *Econometrica* 50 (2):501–16. doi:10.2307/1912640.

Tran, K. 1998. Estimating mixtures of normal distributions via empirical characteristic function. *Econometric Reviews* 17 (2):167–83. doi:10.1080/07474939808800410.

Yohai, V. J., and R. H. Zamar. 1988. High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association* 83 (402):406–13. doi:10.1080/01621459.1988.10478611.