Thailand Statistician
July 2012; 10(2): 225-246
http://statassoc.or.th
Contributed paper

Inferences on The Standard Skew-Normal Distribution

Nabendu Pal [a]*, Wooi K. Lim [b], Ampai Thongteeraparp [c]

[a] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana, USA 70504.

[b] Department of Mathematics, William Paterson University, Wayne, New Jersey, USA 07470.

[c] Department of Statistics, Kasetsart University, Bangkok, Thailand 10900.

*Author for correspondence; e-mail: nxp3695@louisiana.edu

Received: 9 August 2012
Accepted: 9 October 2012

Abstract

It is a common practice among applied researchers to assume normal distribution for naturally occurring data over the real line. But often one is not sure about the assumption of normality for various reasons, including the fact that the standard goodness of fit tests are not effective enough always, especially for small sample sizes. In such a scenario one would be better off by starting with a more versatile skew-normal distribution which is defined over the whole real line, and is a natural generalization of the usual normal distribution. This paper deals with the standard skew-normal distribution which can reduce to the standard normal distribution if the skew parameter takes the value zero. Depending on the value of the skew parameter, the standard skew-normal distribution can be either positively skewed, symmetric (standard normal), or negatively skewed. This paper is devoted to various estimation and hypothesis testing methods for the skew parameter which, to the best of our knowledge, is the first comprehensive work in this direction.

Keywords: Skew parameter, asymptotic distribution, penalized likelihood estimation, parametric bootstrap.

This paper was presented at the International Conference in Mathematics and Applications (ICMA - MU 2011), organized by Mahidol University, Thailand. The authors would like to express their gratitude to the organizers for their invitation and hospitality.

1. Introduction

There has been a growing interest lately in the skew-normal distribution(‘SND’ hereafter) due to its flexibility in modeling real-life data sets. The SND density, which has an extra skew (or shape) parameter to regulate skewness, includes the usual normal density as a special case. As a result, the SND class of densities incorporates negatively skewed, positively skewed as well as the symmetric normal densities which makes it quite versatile in fitting data over the real-line.

A random variable X is said to have a SND with location parameter μ, scale parameter σ and skew (or shape) parameter λ, henceforth denoted by $SND(\mu,\sigma,\lambda)$, provided the *pdf* of X is given by

$$f(x\,|\,\mu,\sigma,\lambda) = (2/\sigma)\phi((x-\mu)/\sigma)\ \Phi(\lambda(x-\mu)/\sigma),\ \ x\in\mathbb{R},\ \mu\in\mathbb{R},\ \sigma\in\mathbb{R}^{+},\ \lambda\in\mathbb{R}; \qquad (1.1)$$

where ϕ and Φ are the standard normal pdf and cdf respectively. A positive (negative) value of λ indicates positive (negative) skewness of the distribution. Also note that $SND(\mu,\sigma,\lambda=0)\equiv N(\mu,\sigma^2)$ distribution.

In many applied problems it is customary to assume normality for the data set(s) mainly due to convenience. Usual goodness of fit tests, like Anderson-Darling test, Shapiro-Wilk test, etc., are often applied to justify the use of normality. However, it should be kept in mind that such tests are asymptotic in nature, which means,—they are not very effective for small sample sizes, and do not possess good power unless the alternative model is highly skewed. Further, there aren’t too many skewed distributions defined on the real-line which are widely used. Therefore, it would make sense if one starts with a SND for the given data, and then proceeds with the next step of inferences. After assuming a SND for a given dataset, the first logical step would be to estimate the skew parameter λ, followed by a test to see if λ is actually zero or not. If the null hypothetical value of zero is accepted for λ then the model reduces to a regular normal distribution; otherwise, one ought to use the SND model for further inferences.

The SND in the present form (see (1.1)) has been made popular mainly by Azzalini [1] who coined the name of the distribution. Initially, SND was used in a few applications by Roberts [2], O'Hagan and Leonard [3], and Aigner and Lovell [4]. Subsequently many interesting properties of the distribution have been studied by other researchers, notably – Azzalini [5]; Azzalini and Dalla Valle [6]; Arnold and Lin [7]; Gupta, Nyuyen and Sanqui [8]. The last two papers are particularly important as they discuss many characterization properties of the distribution. These characterizations are often helpful in simulation studies to generate SND data.

More recently, Pal, Chang and Lin [9] extended the famous Stein's normal identity (see Stein [10]) for the SND. A further generalization of the univariate SND has been introduced by Gomez, Salinas and Bolfarine [11], whereas Gupta, Chen and Tang [12] provided a multivariate generalization.

The focus of this paper is a special case of the SND, that is the standard skew-normal distribution (' $SSND(\lambda)$ ' hereafter) which is obtained from $SND(\mu,\sigma,\lambda)$ with $\mu=0$ and $\sigma=1$. Our interest lies in inferences on λ for $SSND$. The tools and ideas developed here will be used for general SND which will be discussed in a follow-up future paper.

As we will see in the rest of the paper, inferences with SND (or $SSND$) is not easy. First of all, the sampling distributions are intractable; and secondly, there are many computational challenges one has to overcome.

Before we go into the structure of this paper and further details about inferences on λ, let us revisit some basic and interesting results of SND.

If $X \sim SND(\mu,\sigma,\lambda)$, then

$Z=(X-\mu)/\sigma \sim SND(0,1,\lambda) \equiv SSND(\lambda).$ Further,

(i) $E(Z)=\sqrt{(2/\pi)}(\lambda/\sqrt{1+\lambda^2})$;

(ii) $Var(Z)=1-(2/\pi)\lambda^2/(1+\lambda^2)$;

(iii) $(-Z) \sim SND(0,1,-\lambda) \equiv SSND(-\lambda)$;

(iv) $Z^2 \sim \chi_1^2$ (Chi-square with 1 df);

(v) $M_Z(t)=E[\exp(tZ)]=2\exp(t^2/2)\Phi(\lambda t/\sqrt{1+\lambda^2})$.

(vi) The SND identity (Pal, Chang and Lin [9]) says that if g is a real valued differentiable function such that $g(u)\phi(u) \to 0$ as $u \to \pm\infty$ then

$$E[(X-\mu)g(X)] = \sigma^2 E[g'(X)] + \sigma\sqrt{(2/\pi)}(\lambda/\sqrt{1+\lambda^2}E[g(X_*)]$$

where X_* follows $N(\mu, \sigma^2/(1+\lambda^2))$, and provided all expectations exist.

(vii) If V_1 and V_2 are iid $N(0,1)$ random variables, then

$$V = \{(\lambda/\sqrt{1+\lambda^2})|V_1| + (1/\sqrt{1+\lambda^2})V_2\} \sim SSND(\lambda).$$

The above identity (vi) comes handy to evaluate moments of SND in a very convenient way. Also, the property (vii) can be used to generate $SSND(\lambda)$ observations.

Since this paper deals with the special case of $SSND(\lambda)$ (i.e., $SND(\mu=0, \sigma=1, \lambda)),$ the rest of the paper is organized in a straightforward way. In Section 2 we consider point estimation of the skew parameter λ, whereas Section 3 deals with hypothesis testing of λ. Each section provides comprehensive simulation results to justify our suggested methods.

It may appear that there has been some overlapping of our work with some recent works, especially with Liseo and Loperfido [13], Sartori [14], Pewsey [15] and Dalla Valle [16], but most of the results presented here are new to the best of our knowledge.

2. Estimation of the Skew Parameter

Assume that iid observations $X_1, X_2, \cdots, X_n$ are available from $SSND(\lambda)$ with the following pdf (from (1.1))

$$f(x|\lambda) = 2\phi(x)\Phi(\lambda x), \lambda > 0, x \in \mathbb{R} \quad (2.1)$$

Maximum likelihood estimation of λ based on $\underset{\sim}{X} = (X_1, X_2, \cdots, X_n)$ remains a contentious issue. The log-likelihood function

$$L_n = L_n(\lambda \mid \underset{\sim}{X}) = n\,ln\,2 + \sum_{i=1}^{n} ln\,\varphi(X_i) + \sum_{i=1}^{n} ln\,\Phi(\lambda X_i) \quad (2.2)$$

can not be maximized always to find the MLE of λ. Note that if all observations are nonnegative, then L is nondecreasing in λ, and hence the MLE of λ becomes

($+\infty$). Similarly, on the other hand, if all observations are nonpositive, then the MLE of λ becomes ($-\infty$).

Let $p_\lambda = P(X_i > 0)$. Then, the probability of all observations becoming nonnegative (nonpositive) is $p_\lambda^n((1-p_\lambda)^n)$. The following table (Table 2.1) gives the values of p_λ for different values of λ. Because $\hat{\lambda}_{ML}$, the MLE of λ, can take $\pm\infty$ with positive probability, it is not meaningful to evaluate its bias and/or MSE (mean squared error) in the usual sense.

Table 2.1 : Values of p_λ for various $\lambda \in \mathbb{R}$

λ	-7	-6	-5	-4	-3	-2	-1	0
p_λ	0.0452	0.0526	0.0628	0.0780	0.1024	0.1476	0.2500	0.500

λ	1	2	3	4	5	6	7	8
p_λ	0.7500	0.8524	0.8976	0.9220	0.9372	0.9474	0.9548	0.9604

Further, since $\hat{\lambda}_{ML}$ takes the boundary values ($\pm\infty$) of the parameter space of λ (which is $\mathbb{R}$), the standard asymptotic theory doesn't apply to study the asymptotic distribution of $\hat{\lambda}_{ML}$. For the model (2.1) Fisher information per observation is

$$i(\lambda) = -E(l''(\lambda \mid X)) = 2\int_{-\infty}^{\infty} u^2\, \varphi(u)(\varphi(\lambda u))^2 / \Phi(\lambda u)\, du, \tag{2.3}$$

where $l(\lambda \mid X) = ln f(X \mid \lambda)$, and $l''(\lambda \mid X) = \partial^2 l(\lambda \mid X)/\partial\lambda^2$. Liseo and Loperfido [17] explored the properties of $i(\lambda)$ and observed that (i) $i(\lambda)$ is symmetric about $\lambda = 0,$ and is decreasing in $|\lambda|$; (ii) the tails of $i(\lambda)$ are of order $O(\lambda^{-3})$.

Though a lot has been studied about the properties of the SND, relatively little attention has been paid to inferences on λ. As an alternative to the MLE, Liseo and Loperfido [17] proposed a Bayesian estimation technique which we will discuss later. But first we look into the possible frequentist methods.

When the MLE is not readily available the common sense dictates that we look at the method of moment(s) estimator (MME). By equating the first sample moment with the population one we have

$$E(X_i) = \sqrt{(2/\pi)}(\lambda/\sqrt{1+\lambda^2}) \approx \bar{X}, \tag{2.4}$$

which, upon solving for λ, yields the MME of λ as

$$\hat{\lambda}_{MM} = (sign\ of\ \ \bar{X})\left\{(2/\pi)(1/\bar{X})^2 - 1\right\}^{-1/2}. \tag{2.5}$$

(From (2.4) it is obvious that the MME of λ must carry the sign of $\bar{X}$.) The square-root of the term inside the $\{\}$ in (2.5) is not meaningful unless the term is nonnegative. Hence $\{\}$ needs to be truncated at 0 if it ever becomes negative. In other words, the MME of λ is modified to look as

$$\hat{\lambda}_{MM+} = (sign\ of\ \ \bar{X})\left\{(2/\pi)(1/\bar{X})^2 - 1\right\}_+^{-1/2}, \tag{2.6}$$

where for any real value c, $c_+ = \max(0, c).$

Since the MLE of λ can assume $\pm\infty$ with positive probability, we propose an alternative approach through 'Penalized Maximum Likelihood Estimation' (PMLE) where a suitable penalty function is attached to the log-likelihood function as

$$\tilde{L}_n = \tilde{L}_n(\lambda \mid \underset{\sim}{X}) = L_n(\lambda \mid \underset{\sim}{X}) - h_n(\lambda) \tag{2.7}$$

where $h_n(\lambda)$ is a suitable penalty function, may be dependent on n, goes to ∞ as λ approaches $\pm\infty$. For notational and computational convenience we drop the first two terms of (2.2) which are free from λ. Therefore, we consider the reduced expression of $\tilde{L}_n$ as

$$\tilde{L}_n(\lambda \mid \underset{\sim}{X}) = \sum_{i=1}^{n} ln\ \Phi(\lambda X_i) - h_n(\lambda). \tag{2.8}$$

The following result gives a sufficient condition for obtaining $\hat{\lambda}_{PML}$, a penalized MLE of λ.

Proposition 2.1 The penalized likelihood function $\tilde{L}_n$ in (2.8) attains a maximum at a finite λ provided the penalty function $h_n(\lambda)$ is strictly convex, i.e., $h_n''(\lambda) > 0.$

Proof. From (2.8),

$$\partial\tilde{L}_n / \partial\lambda^2 = \sum_{i=1}^{n}\left\{\partial\, ln\, \Phi(\lambda X_i)/\partial\lambda^2\right\} - h_n''(\lambda).$$

If $h_n''(\lambda) > 0$ then it is enough to show that each $ln\, \Phi(\lambda X_i)$ is concave with respect to λ so that $\tilde{L}_n$ is concave to attain a maximum at finite λ . Note that

$$\partial ln\, \Phi(\lambda X_i)/\partial\lambda^2 = [-\lambda^{-2}u_i^2\{u_i\phi(u_i)\Phi(u_i) + (\phi(u_i))^2\}]/(\Phi(u_i))^2. \tag{2.9}$$

where $u_i = (\lambda X_i)$. So, $ln\Phi(\lambda X_i)$ is concave w.r.t. λ for each $i = 1, 2, \cdots, n,$ provided $u_i\phi(u_i)\Phi(u_i) + (\phi(u_i))^2 > 0 \quad \forall u_i \in \mathbb{R}$;

i.e., $\kappa(u_i) = u_i\Phi(u_i) + \phi(u_i) > 0 \quad \forall u_i \in \mathbb{R}$. Note that $\kappa'(u_i) = \Phi(u_i) > 0$ $\forall u_i \in \mathbb{R}$; i.e., $\kappa(u_i)$ is increasing. Therefore, it is enough to show that $\kappa(-\infty) = 0$ such that $\kappa(u_i) \geq 0$ always. By L'Hospital's rule, the limit of $(u_i\Phi(u_i)),$ as $u_i \to -\infty$, is $(-u_i^2\phi(u_i)),$ which is zero.

Remark 2.1. For the penalty function, $h_n(\lambda) = 0$ is not a choice since this leads to the MLE which can be $\pm\infty$ with positive probability. We must have $h_n(\lambda)$ such that $h_n''(\lambda)$ is strictly positive so that it has a dampening effect on $\{\sum_{i=1}^{n} ln\, \Phi(\lambda X_i)\}.$

The PMLE (penalized MLE) of λ , i.e., $\hat{\lambda}_{PML}$ is obtained by solving the equation

$$\sum_{i=1}^{n}\left\{X_i\; \varphi(\lambda X_i)/\Phi(\lambda X_i)\right\} - h_n'(\lambda) = 0. \tag{2.10}$$

Remark 2.2. The choice of the penalty function $h_n(\lambda)$ is critical in estimating λ by $\hat{\lambda}_{PML}$. Following the standard asymptotic theory we need to choose $h_n(\lambda)$ such that

$h_n(\lambda) = \lambda^2 c_n$, where $c_n/\sqrt{n} \to c_0$ for some $c_0 \geq 0$ .The following general result gives us an idea about the asymptotic behavior of $\hat{\lambda}_{PML}$.

Proposition 2.2 Let $h_n(\lambda) = c_n\lambda^2$ where $c_n/\sqrt{n} \to c_0$ as $n \to \infty$. Then, under standard regularity conditions (the ones used for Cramér-Rao inequality), asymptotically

$$\sqrt{n}(\hat{\lambda}_{PML} - \lambda) \xrightarrow{D} N(-2c_0\lambda/i(\lambda), 1/i(\lambda))$$

where $i(\lambda)$ is given in (2.3).

Proof. See Appendix A.1

Using the penalty function as given in Proposition 2.2, obtaining $\hat{\lambda}_{PML}$ now boils down to solving the equation

$$\sum_{i=1}^{n} \{X_i \phi(\lambda X_i)/\Phi(\lambda X_i)\} - 2c_n\lambda = 0, \tag{2.11}$$

where $c_n/\sqrt{n} \to c_0$.

For a demonstration purpose we generate $n = 5$ observations from $SSND(\lambda)$
(i.e., $SND(0,1,\lambda)$) with $\lambda = 1$ as $\{0.5501, -0.4498, 1.5790, 1.9250, 0.2601\}$

(2.12)

(using the template available at *http://azzalini.stat.unipd.it/SN/index.html*). The following Figure 2.1 plots the LHS of (2.11) as a function of λ , given the above dataset (2.12), with $c_n = 0.5$. Here $\hat{\lambda}_{PML} = 0.723281$. Using $c_n = 0.5, c_0 = 0,$ and hence $\hat{\lambda}_{PML}$ is asymptotically unbiased.

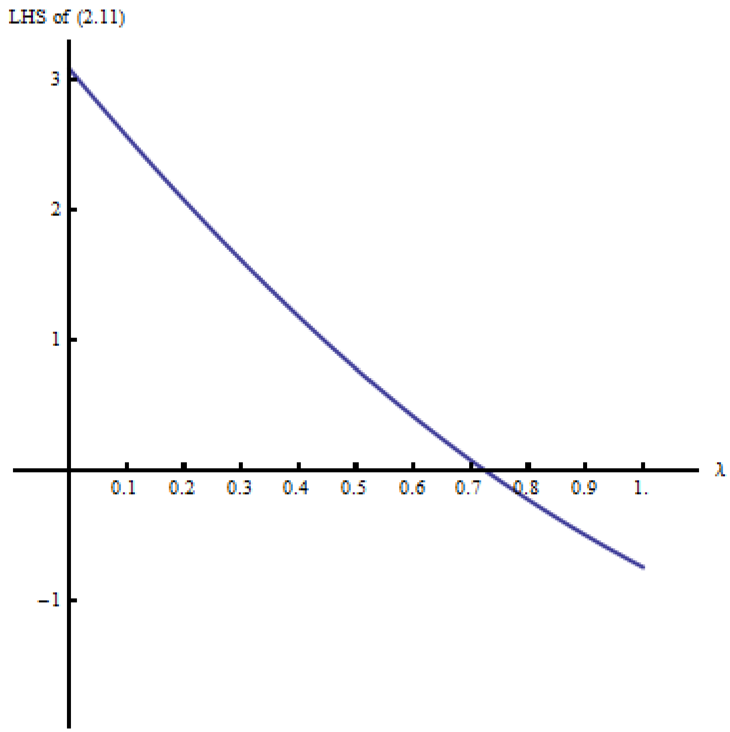

Figure 2.1. Plot of LHS of (2.11) with $n=5$, $c_n=0.5$ for the dataset in (2.12).

Apart from the frequentist estimators discussed above one can also consider a Bayes estimator under a suitable prior $\pi(\lambda)$ of λ. Based on iid observations from (2.1), the posterior distribution of λ given $X=(X_1, X_2, \cdots, X_n)$ has the form

$$\pi(\lambda|\underset{\sim}{X}) = \pi(\lambda)\{\prod_{i=1}^{n}\Phi(\lambda X_i)\} / \int_{-\infty}^{\infty}\pi(\lambda)\{\prod_{i=1}^{n}\Phi(\lambda X_i)\}\,d\lambda. \qquad (2.13)$$

There doesn't appear to be any natural conjugate prior for the SND skew parameter. The easiest prior one can think of is

$$\pi(\lambda|\alpha,\lambda_0) = |\lambda-\lambda_0|^{-\alpha}, \qquad (2.14)$$

where λ_0 is any real value that λ is thought to be close to. The posterior distribution with the above prior (2.14) is

$$\pi_\alpha(\lambda|\underset{\sim}{X}) = |\lambda-\lambda_0|^{-\alpha}\{\prod_{i=1}^{n}\Phi(\lambda X_i)\} / \int_{-\infty}^{\infty}|\lambda-\lambda_0|^{-\alpha}\{\prod_{i=1}^{n}\Phi(\lambda X_i)\}\,d\lambda, \quad (2.15)$$

Provided $\pi_\alpha(\lambda|\underset{\sim}{X})$ is integrable, i.e.,

$$\kappa(\alpha,\lambda_0) = \int_{-\infty}^{\infty}|\lambda-\lambda_0|^{-\alpha}\{\prod_{i=1}^{n}\Phi(\lambda X_i)\}\,d\lambda < \infty. \quad (2.16)$$

The following Figures 2.2a, 2.2b show the plots of the posterior distribution in (2.15) for various values of α. However, the above condition (2.16) does not hold for all $\alpha \geq 0.$ If $\alpha = 0$, then

$$\kappa(0,\lambda_0) = \int_{-\infty}^{\infty}\{\prod_{i=1}^{n}\Phi(\lambda X_i)\}\,d\lambda = \int_{-\infty}^{\infty}\{\prod_{i=1}^{n}\Phi(\lambda X_{(i)})\}\,d\lambda$$

$$= \int_{0}^{\infty}\{\prod_{i=1}^{n}\Phi(\lambda X_{(i)})\}\,d\lambda + \int_{0}^{\infty}\{\prod_{i=1}^{n}\Phi(\lambda(-X_{(i)}))\}\,d\lambda. \quad (2.17)$$

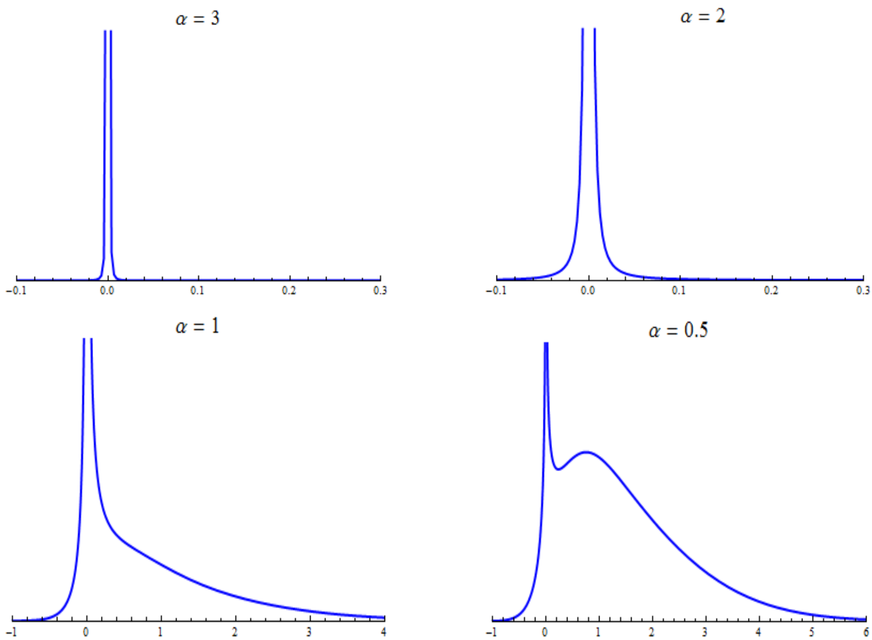

Figure 2.2a. Plots of the posterior *pdf* for the dataset in (2.12) under prior (2.14).

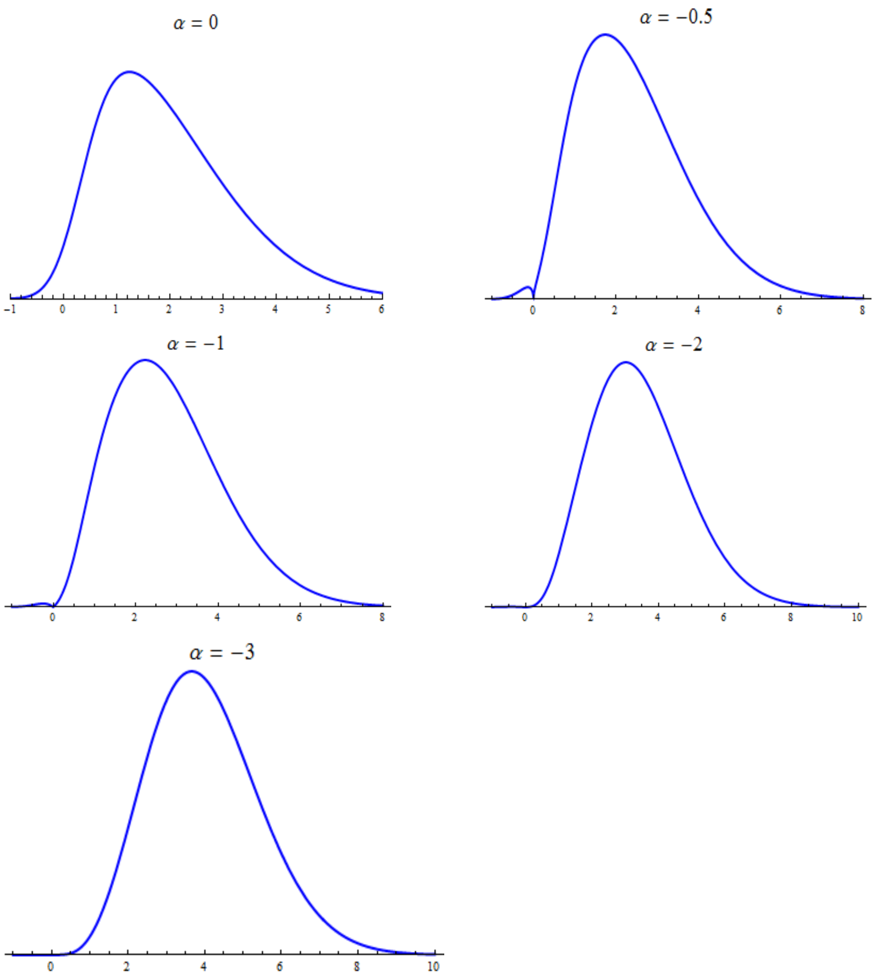

Figure 2.2b. Plots of the posterior *pdf* for the dataset in (2.12) under prior (2.14).

Without loss of generality, assume that $X_{(1)} > 0.$ Then the first term of (2.17) is bounded below by $\int_0^\infty (\Phi(\lambda(X_{(1)})))^n \, d\lambda$. Since this integrand goes to 1 as $\lambda \to \infty$, the whole integral is ∞. However, a negative value of $\alpha(<0)$ can ensure that the posterior $\pi_\alpha(\lambda | \underset{\sim}{X})$ is proper. In our subsequent derivations we will be using $\alpha < 0$ and $\lambda_0 = 0$ (without loss of generality).

One can also use Jeffrey's noninformative prior which seems to be a natural candidate for estimating λ , and it is given as

$$\pi_J(\lambda) = (i(\lambda))^{1/2} \quad (see\ ((2.3)). \tag{2.18}$$

The posterior distribution of λ under $\pi_J(\lambda)$ is

$$\pi_J(\lambda|\underset{\sim}{X}) = \kappa\{\prod_{i=1}^{n}\Phi(\lambda X_i)\}\{\int_{-\infty}^{\infty} u^2\varphi(u)(\varphi(\lambda u))^2 / \Phi(\lambda u)\, du\}^{1/2} \tag{2.19}$$

where κ is the normalizing constant. Liseo and Loperfido [13] has discussed details of the mean of the posterior (2.19) as an estimator of the skew parameter.

As a demonstration, we plot the posterior pdf of λ under Jeffery's prior for the dataset (2.12) in Figure 2.3.

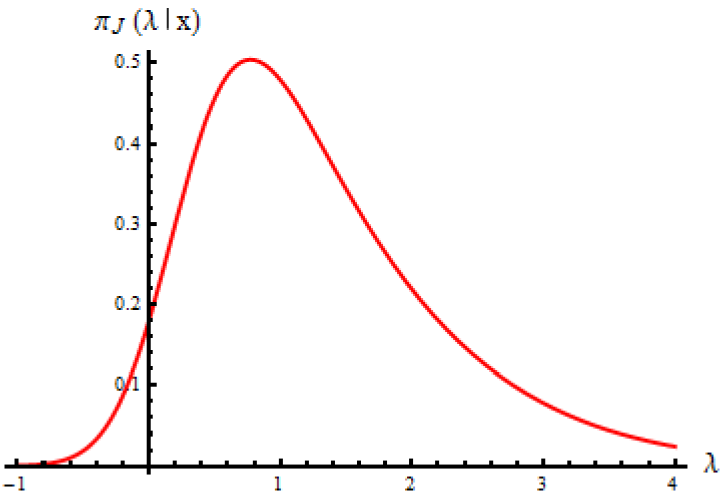

Figure 2.3. Plot of the posterior pdf for the dataset in (12) under Jeffery's prior.

In the following we compare the estimators developed above in terms of standardized bias (SB) and standardized MSE (SMSE). For an estimator $\hat{\lambda}$ of λ , SB and SMSE are defined as

$$SB(\hat{\lambda} \mid \lambda, n) = E(\hat{\lambda} - \lambda)/\lambda; \quad SMSE(\hat{\lambda} \mid \lambda, n) = E(\hat{\lambda} - \lambda)^2/\lambda^2. \tag{2.20}$$

The estimators of primary interest are :

$\hat{\lambda}_{mm+}$ = the (modified) MME as given in (2.6)

$\hat{\lambda}_{\alpha(mean)}$ = posterior mean of $\pi_\alpha(\lambda | \underset{\sim}{X})$ in (2.15)

$\hat{\lambda}_{\alpha(median)}$ = posterior median of $\pi_\alpha(\lambda | \underset{\sim}{X})$ in (2.15)

$\hat{\lambda}_{\alpha(mode)}$ = posterior mode of $\pi_\alpha(\lambda | \underset{\sim}{X})$ in (2.15)

$\hat{\lambda}_{J(mean)}$ = posterior mean of $\pi_J(\lambda | \underset{\sim}{X})$ in (2.19)

$\hat{\lambda}_{J(median)}$ = posterior median of $\pi_J(\lambda | \underset{\sim}{X})$ in (2.19)

$\hat{\lambda}_{J(mode)}$ = posterior mode of $\pi_J(\lambda | \underset{\sim}{X})$ in (2.19)

$\hat{\lambda}_{PMLE}$ = by solving equation (2.11)

These eight estimators are compared in terms of SB and SMSE by simulation. In this simulation we use $\alpha = -1$ for $\hat{\lambda}_{\alpha(mean)}$, $\hat{\lambda}_{\alpha(median)}$ and $\hat{\lambda}_{\alpha(mode)}$. We choose $c_n = 1$ to obtain the estimator $\hat{\lambda}_{PMLE}$ by solving the equation (2.11). For a fixed n and λ, iid observations $X_1, \cdots, X_n$ are generated from $SSND(\lambda)$ (using the property (vii) in Section 1) in a particular replication. In the k^{th} replication, $1 \le k \le K,$ the estimator $\hat{\lambda}$ (which is any one of the eight estimators) is denoted by $\hat{\lambda}^{(k)}$. Then the SB and SMSE of $\hat{\lambda}$ are approximated as

$$SB(\hat{\lambda} | \lambda, n) \approx \{\sum_{k=1}^{K} (\hat{\lambda}^{(k)} - \lambda)/\lambda\}/K \quad and$$

$$SMSE(\hat{\lambda} | \lambda, n) \approx \{\sum_{k=1}^{K} (\hat{\lambda}^{(k)} - \lambda)^2/\lambda^2\}/K.$$

The simulated SB and SMSE values are λ presented in the Tables 3.1 (a,b,c) and 3.2 (a,b,c), respectively. The estimated SB and SMSE of each estimator are computed based on $K = 5000$ replications for each $\lambda = 0.1, 0.5, 1, 2, 3, \cdots, 8$. If we look at SB and SMSE criteria then it is obvious that there is no clear-cut winner. Some estimators perform well when λ is close to zero, and others perform better asdeviates greatly from zero. However, taking overall performance into consideration, it appears that out of 8 estimators $\hat{\lambda}_{J(median)}$, $\hat{\lambda}_{J(mode)}$ and $\hat{\lambda}_{PMLE}$ are the top three estimators compared to other estimators.

Table 3.1a. Table of SB for $n=5$.

		λ									
n	Estimator	0.1	0.5	1	2	3	4	5	6	7	8
5	$\hat{\lambda}_{MM+}$	0.7386	0.1062	-0.2541	-0.5698	-0.7446	-0.7579	-0.8175	-0.8423	-0.872	-0.888
	$\hat{\lambda}_{\alpha(mean)}$	1.0105	1.9039	1.2671	0.5732	0.1254	-0.1254	-0.292	-0.4059	-0.4901	-0.5514
	$\hat{\lambda}_{\alpha(median)}$	15.4213	3.1729	1.4988	00.65	0.178	-0.0822	-0.255	-0.3745	-0.4631	-0.5271
	$\hat{\lambda}_{\alpha(mode)}$	10.5722	2.8513	0.8203	0.6609	0.1878	-0.0903	-0.2599	-0.3817	-0.4703	-0.0712
	$\hat{\lambda}_{J(mean)}$	0.1119	0.5449	0.047	-0.2653	-0.4763	-0.5937	-0.6747	-0.7275	-0.7659	-0.7952
	$\hat{\lambda}_{J(median)}$	0.0638	0.397	0.4024	0.1002	-0.1755	-0.3172	-0.4281	-0.5087	-0.5254	-0.3214
	$\hat{\lambda}_{J(mode)}$	1.1424	0.105	-0.2177	-0.4322	-0.5868	-0.6638	-0.7245	-0.7653	-0.8005	-0.8219
	$\hat{\lambda}_{pmle}$	-0.5292	-0.3979	0.5523	-0.6923	-0.7834	-0.8289	-0.863	-0.8848	-0.9005	-0.9127

Table 3.1b. Table of SB for $n=10$.

		λ									
n	Estimator	0.1	0.5	1	2	3	4	5	6	7	8
10	$\hat{\lambda}_{MM+}$	0.802	0.1565	-0.1482	-0.4466	-0.6338	-0.6813	-0.7783	-0.7724	-0.8337	-0.8624
	$\hat{\lambda}_{\alpha(mean)}$	1.5705	1.2102	0.983	0.5096	0.1576	-0.1027	-0.2679	-0.3814	-0.4675	-0.5305
	$\hat{\lambda}_{\alpha(median)}$	5.7053	1.4594	0.9973	0.5406	0.1985	-0.0667	-0.2364	-0.3529	-0.4426	-0.5081
	$\hat{\lambda}_{\alpha(mode)}$	6.179	1.4038	0.6891	0.4244	0.4261	0.4442	0.4523	0.2571	0.1003	-0.0127
	$\hat{\lambda}_{J(mean)}$	0.5772	0.3474	0.143	-0.1736	-0.4011	-0.5479	-0.6414	-0.7056	-0.7495	-0.7848
	$\hat{\lambda}_{J(median)}$	0.5125	0.2677	0.3197	0.2263	0.0761	-0.1308	-0.262	-0.3496	-0.424	-0.4392
	$\hat{\lambda}_{J(mode)}$	0.4098	0.0876	-0.0088	-0.1813	-0.3048	-0.4449	-0.5365	-0.5895	-0.644	-0.6752
	$\hat{\lambda}_{pmle}$	-0.0892	-0.233	-0.3633	-0.5705	-0.6835	-0.7585	-0.8035	-0.835	-0.8575	-0.8743

Table 3.1c. Table of SB for $n = 25$.

		λ									
n	Estimator	0.1	0.5	1	2	3	4	5	6	7	8
25	$\hat{\lambda}_{MM+}$	0.0092	0.1697	0.1181	-0.2743	-0.4608	-0.6338	-0.7003	-0.5803	-0.7747	-0.803
	$\hat{\lambda}_{\alpha(mean)}$	0.4511	0.4712	0.4342	0.3659	0.1494	-0.0621	-0.2189	-0.338	-0.4281	-0.4966
	$\hat{\lambda}_{\alpha(median)}$	-1.0286	0.3506	0.4049	0.3647	0.178	-0.0347	-0.1913	-0.313	-0.4061	-0.4768
	$\hat{\lambda}_{\alpha(mode)}$	2.3163	0.4413	0.3213	0.3423	0.3048	0.3758	0.3661	0.2377	0.0863	-0.0169
	$\hat{\lambda}_{J(mean)}$	-0.0373	0.0838	0.129	0.0981	-0.0535	-0.2066	-0.3286	-0.4257	-0.5011	-0.5593
	$\hat{\lambda}_{J(median)}$	0.0324	0.0596	0.0965	0.0724	-0.0621	-0.2017	-0.3182	-0.4134	-0.4891	-0.5354
	$\hat{\lambda}_{J(mode)}$	-0.1026	0.0048	0.0263	0.0352	-0.0416	-0.1024	-0.1783	-0.2475	-0.3164	-0.3704
	$\hat{\lambda}_{pmle}$	-0.1579	-0.0979	-0.181	-0.3973	-0.5493	-0.6451	-0.7096	-0.7557	-0.7885	-0.8147

Table 3.2a. Table of $SMSE$ for $n = 5$.

		λ									
n	Estimator	0.1	0.5	1	2	3	4	5	6	7	8
5	$\hat{\lambda}_{MM+}$	166.0259	7.7135	2.5598	1.0613	0.7268	0.7579	0.8835	0.7697	0.8007	0.8293
	$\hat{\lambda}_{\alpha(mean)}$	325.608	15.2418	3.304	0.4601	0.0361	0.0195	0.087	0.1655	0.2406	0.3042
	$\hat{\lambda}_{\alpha(median)}$	378.3882	15.5778	3.5538	0.5597	0.0576	0.0122	0.0671	0.1411	0.215	0.2781
	$\hat{\lambda}_{\alpha(mode)}$	292.1193	22.9046	2.5853	0.692	0.1035	0.0333	0.0817	0.1549	0.2282	0.0405
	$\hat{\lambda}_{J(mean)}$	90.2509	3.9596	0.3939	0.1034	0.233	0.3536	0.4559	0.5296	0.5868	0.6324
	$\hat{\lambda}_{J(median)}$	72.2579	3.2004	1.4229	0.2524	0.1109	0.1318	0.2007	0.269	0.3009	0.187
	$\hat{\lambda}_{J(mode)}$	32.6716	1.1489	0.3067	0.231	0.3606	0.4488	0.5298	0.5891	0.6428	0.6774
	$\hat{\lambda}_{pmle}$	12.8321	0.584	0.3795	0.4862	0.6157	0.6877	0.7452	0.7831	0.8112	0.8332

Table 3.2 b. Table of $SMSE$ for $n=10$.

		λ									
n	Estimator	0.1	0.5	1	2	3	4	5	6	7	8
10	$\hat{\lambda}_{MM+}$	113.08	2.7959	1.0502	1.1919	0.7658	1.223	0.8952	2.0565	0.7983	0.8186
	$\hat{\lambda}_{\alpha(mean)}$	107.6475	6.0347	2.0517	0.4112	0.0482	0.0168	0.0738	0.1462	0.2189	0.2817
	$\hat{\lambda}_{\alpha(median)}$	123.0445	6.008	2.1295	0.4845	0.072	0.0135	0.0591	0.1256	0.1965	0.2584
	$\hat{\lambda}_{\alpha(mode)}$	96.8018	6.9892	1.6921	0.7478	0.5607	0.4417	0.3527	0.1477	0.0601	0.0294
	$\hat{\lambda}_{J(mean)}$	38.1697	2.0836	0.3501	0.0642	0.1664	0.3026	0.4126	0.4988	0.5624	0.6164
	$\hat{\lambda}_{J(median)}$	33.5793	1.8212	1.1005	0.4405	0.1518	0.0849	0.1026	0.1418	0.1929	0.2118
	$\hat{\lambda}_{J(mode)}$	22.3294	1.1428	0.3916	0.1557	0.1548	0.23	0.309	0.3632	0.4254	0.4657
	$\hat{\lambda}_{pmle}$	11.8783	0.4526	0.1956	0.3324	0.4686	0.576	0.6459	0.6974	0.7354	0.7645

Table 3.2c. Table of $SMSE$ for $n=25$.

		λ									
n	Estimator	0.1	0.5	1	2	3	4	5	6	7	8
25	$\hat{\lambda}_{MM+}$	8.0383	1.1141	1.5411	1.1236	1.0464	0.7564	0.7714	0.8064	0.9098	0.816
	$\hat{\lambda}_{\alpha(mean)}$	18.7428	0.8948	0.6041	0.3014	0.0556	0.0127	0.0499	0.115	0.1837	0.2468
	$\hat{\lambda}_{\alpha(median)}$	74.4606	1.3795	0.5735	0.3262	0.0724	0.0134	0.0393	0.099	0.1654	0.2276
	$\hat{\lambda}_{\alpha(mode)}$	24.7823	0.6695	0.4656	0.4857	0.3493	0.3795	0.2737	0.1302	0.0499	0.0249
	$\hat{\lambda}_{J(mean)}$	8.3374	0.4616	0.2889	0.1383	0.0381	0.055	0.1118	0.183	0.2521	0.3134
	$\hat{\lambda}_{J(median)}$	8.3155	0.4283	0.2648	0.1423	0.0474	0.0575	0.1069	0.1736	0.2408	0.2929
	$\hat{\lambda}_{J(mode)}$	7.3245	0.3652	0.2293	0.2651	0.1657	0.1303	0.1141	0.1243	0.1499	0.1766
	$\hat{\lambda}_{pmle}$	6.1883	0.2438	0.0905	0.1665	0.3036	0.4168	0.5038	0.5712	0.6218	0.6638

3. Testing on the Skew Parameter

Based on the iid observations from (2.1) we want to test

$$H_0 : \lambda = \lambda_0 \quad vs \quad H_A : \lambda \neq \lambda_0 \tag{3.1}$$

for any $\lambda_0 \in \mathbb{R}$. However, our interest lies primary in $\lambda_0 = 0$ since this leads to the standard normal distribution, and hence all our computational results are provided accordingly. However, here we present the theory for general λ_0.

As mentioned in the previous sections the main challenge of SND (or $SSND$) is the sampling distributions of the estimators of λ. As a result we will focus on deriving tests based on the asymptotic theory as well as parametric bootstrap.

Liseo and Loperfido [17] proposed a Bayesian test for (3.1) based on the prior (2.18). The procedure is computationally intensive, and requires the MCMC algorithm. Dalla Valle [16] proposed a general goodness of fit test for the SND based on a modified version of the Anderson-Darling test (usually used for a normal goodness of fit test). Though critical values have been provided for large sample sizes, how this test performs for small sample sizes is yet to be seen. The three test procedures developed in this section are not only easy to implement, but also two of the three tests attain the nominal level closely even for small sample sizes.

3.1. A Simple Asymptotic Test

Possibly the simplest asymptotic test one can think of is based on the sample average $\bar{X}$ when iid observations are available from $SSND(\lambda)$. Using the moment properties discussed in Section 1, and using the Central Limit Theorem (CLT), it is easy to see that as $n \to \infty$

$$\bar{X} \xrightarrow{D} N\left(\sqrt{(2\pi)}(\lambda/\sqrt{1+\lambda^2}), (1-(2/\pi)(\lambda^2/(1+\lambda^2)))/n\right) \tag{3.2}$$

Therefore, under $H_0 : \lambda = \lambda_0$,

$$Q_1 = \sqrt{n}\{\bar{X} - \sqrt{(2\pi)}(\lambda_0 / \sqrt{1+\lambda_0^2} \ )\} / \left\{1-(2\pi)\lambda_0^2 / (1+\lambda_0^2)\right\}^{1/2} \sim N(0,1) \tag{3.3}$$

asymptotically as $n \to \infty$. Therefore, the test based on Q_1 suggests that

$$reject\ H_0\ if\ |Q_1| > z_{(\alpha/2)} \tag{3.4}$$

where $z_{(\alpha/2)}$ is the upper $(\alpha/2)$ tail probability cut off point of $N(0,1)$.

3.2. Asymptotic Test Based on PMLE

Under $H_0 : \lambda = \lambda_0$, and using the Proposition 2.2, one has

$$\sqrt{n}(\hat{\lambda}_{PML} - \lambda_0) \xrightarrow{D} N(-2c_0\lambda_0/i(\lambda_0), 1/i(\lambda_0)) \tag{3.5}$$

i.e., $Q_2 = \sqrt{n}\sqrt{i(\lambda_0)}(\hat{\lambda}_{PML} - \lambda_0) + 2c_0\lambda_0/\sqrt{i(\lambda_0)} \xrightarrow{D} N(0,1)$ as $n \to \infty$. Note that, as stated in Section 2, by our choice $c_0 = 0$. Hence, the test based on Q_2 suggests that

$$reject\ H_0\ if\ |Q_2| > z_{(\alpha/2)}. \tag{3.6}$$

3.3. A Parametric Bootstrap Test

In the following we present the parametric bootstrap test (PBT) as given in Pal, Lim and Ling [18]. The PBT is implemented through the following steps:

Step-1: Obtain $\hat{\lambda}_{PML}$ (as shown in Section 2 by solving (2.11) with $h_n'(\lambda) = 2c_n\lambda$).

Step-2: Assume that $\lambda = \lambda_0$. Generate $X_1^*, X_2^*, \cdots, X_n^*$ *iid* from $SSND(\lambda_0)$, and recalculate the PMLE using this bootstrap data, and call the PBT PMLE as $\hat{\lambda}_{PML}^*$.

Step-3: Repeat the above Step-2 a large number of times (say, M times), and the resultant PBT PMLEs are denoted by $\hat{\lambda}_{PML}^{*1}, \cdots, \hat{\lambda}_{PML}^{*M}$. Further, order them as $\hat{\lambda}_{PML}^{*(1)} \le \cdots \le \hat{\lambda}_{PML}^{*(M)}$.

Step-4: The upper and lower cut-off points of the PBT are found as $\hat{\lambda}_L = \hat{\lambda}_{PML}^{*((\alpha/2)M)}$ and $\hat{\lambda}_U = \hat{\lambda}_{PML}^{*((1-\alpha/2)M)}$. Retain H_0 if $\hat{\lambda}_L \le \hat{\lambda}_{PML}$ (from Step-1) $\le \hat{\lambda}_U$; reject H_0 otherwise.

3.4. Numerical Comparison of the Three Tests

In the following we provide the simulated size/power of the above three tests through a large number of replications. For fixed n and $\lambda \in \mathbb{R}$ we test (3.1) with $\lambda_0 = 0$.

From a practical point of view one should be interested to know if the $SSND(\lambda)$ can be reduced to the $N(0,1)$ distribution or not. We generate iid observations from $SSND(\lambda)$ a large number of times, say N times, and apply each test ($Q_1, Q_2, \& \ PBT$). We had chosen several values for $c_n = 1, 0.5, 0.25$ and $1/\sqrt{n}$; however, we only report the results for $c_n = 0.5$. The simulated size/power is computed by the proportion of times (out of N) the test rejects H_0.The following table (Table 3.7) shows the simulation results with $N = 10^4$ (and for PBT we have used $M = 10^4$).

Table 3.3. Simulated size/power for testing $H_0 : \lambda = 0 \ \ versus \ H_A : \lambda \neq 0$ at $\alpha = 0.05$.

		λ								
n	Test	0	0.5	1	1.5	2	2.5	3	3.5	4
5	Q_1	0.0488	0.1125	0.1983	0.2626	0.2953	0.3209	0.3292	0.3335	0.3343
	Q_2 (c_n = 0.5)	0.0093	0.0183	0.0425	0.0714	0.0985	0.1126	0.1317	0.1346	0.1453
	PBT (c_n =0.5)	0.0475	0.0974	0.2017	0.3059	0.4004	0.4561	0.5088	0.5385	0.5688
10	Q_1	0.0519	0.1888	0.4104	0.5572	0.6589	0.7057	0.7333	0.7632	0.7756
	Q_2 (c_n = 0.5)	0.0642	0.2240	0.5353	0.7888	0.9239	0.9685	0.9891	0.9947	0.9981
	PBT (c_n = 0.5)	0.0511	0.1812	0.4700	0.7347	0.8855	0.9495	0.9786	0.9895	0.9956
15	Q_1	0.054	0.2677	0.6084	0.7953	0.8733	0.9166	0.9333	0.9491	0.9557
	Q_2 (c_n = 0.5)	0.075	0.3204	0.7614	0.9497	0.9929	0.9992	1.0000	1.0000	1.0000
	PBT (c_n = 0.5)	0.0506	0.2399	0.6695	0.9109	0.9843	0.9969	0.9998	0.9998	1.0000
20	Q_1	0.0514	0.3469	0.7431	0.9176	0.9608	0.9819	0.9910	0.9923	0.9934
	Q_2 (c_n = 0.5)	0.0687	0.4003	0.8796	0.9906	0.9996	1.0000	1.0000	1.0000	1.0000
	PBT (c_n =0.5)	0.0479	0.3182	0.8195	0.9811	0.9989	1.0000	1.0000	1.0000	1.0000
25	Q_1	0.0489	0.4258	0.8504	0.9680	0.9928	0.9968	0.9985	0.9992	0.9994
	Q_2 (c_n = 0.5)	0.0663	0.4825	0.9382	0.9977	0.9999	1.0000	1.0000	1.0000	1.0000
	PBT (c_n =0.5)	0.0534	0.438	0.922	0.9964	0.9999	1.0000	1.0000	1.0000	1.0000
50	Q_1	0.0471	0.7237	0.9932	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
	Q_2 (c_n = 0.5)	0.0597	0.7744	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	PBT (c_n =0.5)	0.0503	0.7680	0.9992	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Remark 3.1: From the computational results it is clear that the test based on Q_2 is not very good in maintaining the nominal level α. For $n = 5$ the test is very conservative, while for other moderate sample sizes it is quite liberal. Only for large sample size ($n = 50$) does it have a proper size; on the other hand, both the tests based on Q_1 and PBT are quite good in the following nominal level $\alpha = 0.05$. Yet, between these two tests, PBT exhibits higher power as λ moves moderately away from the null hypothesis value.

Acknowledgment: The first author wishes to thank his colleague Dr. Mihai Giurcanu for his helpful suggestions regarding the proof of Proposition 2.2. Also, the authors appreciate the constructive comments made by an anonymous referee which led to this improved version of the paper.

References

[1] Azzalini, A. A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 1985; 12: 171–178.

[2] Roberts, C. A correlation model useful in the study of twins, *Journal of the American Statistics Association,* 1966; 61: 1184–1190.

[3] O' Hagan, A. and Leonard, T. Bayes estimation subject to uncertainty about parameter constraints,*Biometrika*, 1976; 63: 201–203

[4] Aigner, D. J. and Lovell, C. A. K. Formulation and estimation of stochastic frontier production function model, *Econometrics*, 1977; 12: 21–37.

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

[6] Azzalini, A. and Dalla Valle, A. The multivariate skew-normal distribution, *Biometrika*, 1996; 83: 715–726.

[7] Arnold, B. C. and Lin, G. D. Characterizations of the skew-normal and generalized chi distributions, *Sankhyā,* 2004; 66: 593–606.

[8] Gupta, A. K., Nguyen, T. T. and Sanqui, J. A. Characterization of the skew-normal distribution, *The Annals of the Institute of Statistical Mathematics*, 2004; 56: 351–360.

[9] Pal, N. Chang, C. H. and Lin, J. J. An identity for the skew-normal distribution, *Journal of Applied Probability and Statistics*, 2008; 3: 305–309.

[10] Stein, C. Estimation of the mean of a multivariate normal distribution, *Annals of Statistics*, 1981; 9: 1135–1151.

[11] Gomez, H. W., Salinas, H. S. and Bolfarine, H. Generalized skew-normal models: properties and inference, *Statistics*, 2006; 40: 495–505.

[12] Gupta, A. K., Chen, J. T. and Tang, J. A multivariate two-factor skew model, *Statistics*, 2007; 41: 301–309.

[13] Liseo, B. and Loperfido, N. A note on reference priors for the scalar skew-normal distribution, *Journal of Statistical Planning and Inference*, 2006; 136: 373–389.

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

[15] Pewsey, A. Problems of inference for Azzalini's skew-normal distribution, *Journal of Applied Statistics*, 2000; 27(7): 859–870.

[16] Dalla Valle, A. A test for the hypothesis of skew-normality in a population, *Journal of Statistical Computation and Simulation*, 2007; 77(1): 63–77.

[17] Liseo, B. and Loperfido, N. Default Bayesian analysis of the skew-normal distribution. 2006. Available at: http://geostasto.eco.uniroma1.it/utenti/liseo/liseosent.pdf.

[18] Pal, N., Lim, W. K. and Ling, C. H. A Computational Approach to Statistical Inferences, *Journal of Applied Probability & Statistics*, 2007; Vol 2, No. 1: 13–35.

Appendix

A.1. Proof of Proposition 2.2

The proof is quite general, and hence it is given in a quite general form.

Let $\underset{\sim}{X} = (X_1, \cdots, X_n)$ be a random sample where each $X_i \sim f(x \mid \theta) \in \mathscr{F} = \{f(x \mid \theta) : \theta \in \Theta\}$. We assume that the regularity conditions (those of Cramér-Rao inequality) hold. Using the penalty function $h_n(\theta) = c_n\theta^2$, the PMLE of θ is given by

$$\hat{\theta}_{PML} = argmax\{\sum_{i=1}^{n} l(\theta \mid X_i) - c_n\theta^2\}, \tag{a.1}$$

where $l(\theta \mid x) = ln\ f(x \mid \theta)$ is the score function and $c_n > 0$ is the regularization parameter. Let $\tilde{L}_n(\theta) = \sum_{i=1}^{n} l(\theta \mid X_i) - c_n\theta^2$ be the $l_2 -$penalized likelihood function and let $T_n = \sqrt{n}(\hat{\theta}_{PML} - \theta)$ be the scaled and centered l_2 -penalized MLE, respectively. Note that $\hat{\theta}_{PML} = \theta + n^{-1/2}T_n,$ and since $\hat{\theta}_{PML}$ maximizes $\tilde{L}_n(\theta)$ w.r.t. θ, we have

$$\tilde{L}_n'(\theta + n^{-1/2}T_n) = 0$$

i.e., $\tilde{L}_n'(\theta + n^{-1/2}T_n)n^{-1/2} = 0$

i.e., T_n is the solution of $\tilde{L}_n'(\theta + n^{-1/2}t)n^{-1/2} = 0$;

i.e., T_n is the value of t which also maximizes

$$\Delta = \tilde{L}_n(\theta + n^{-1/2}t) - \tilde{L}_n(\theta). \qquad \text{(a.2)}$$

Therefore,

$$T_n = argmax_t\{\tilde{L}_n(\theta + n^{-1/2}t) - \tilde{L}_n(\theta)\}$$

$$= argmax_t\{\sum_{i=1}^{n}(l(\theta + n^{-1/2}t \mid x_i) - l(\theta \mid x_i)) - c_n((\theta + n^{-1/2}t)^2 - \theta^2)\}$$

$$= argmax_t\{tS_n(\theta) - (1/2)\,t^2 i_n(\theta) - 2n^{-1/2}c_n\theta t + O_P(n^{-1/2})\}, \qquad \text{(a.3)}$$

Where $S_n(\theta) = n^{-1/2}\sum_{i=1}^{n} l'(\theta \mid x_i)$ is the sample score function, $i_n(\theta) = -n^{-1}\sum_{i=1}^{n} l''(\theta \mid x_i)$ is the sample Fisher information (per observation), and the remainder term is of order $O_p(n^{-1/2})$ uniformly on compact set in t (i.e., $\sqrt{n}$ (remainder term) goes to 0 in probability as n increases to ∞). Since $S_n(\theta)$ converges to $N(0, i(\theta))$ in distribution, $i_n(\theta)$ converges to $i(\theta)$ in probability, and by assumption $n^{-1/2}c_n \to c_0,$ it thus follows that $T_n = \sqrt{n}(\hat{\theta}_{PML} - \theta)$ converges in distribution to $N(-2c_0\theta/i(\theta), 1/i(\theta)),$ where $i(\theta) = E(-l''(\theta \mid x))$ is the Fisher information per observation. [The above asymptotic distribution of T_n comes from noting that T_n, which maximizes the term inside $\{\ \}$ in (3), except the $O_P(n^{-1/2})$ term, is in fact $S_n(\theta) - 2n^{-1/2}c_n\theta)/i(\theta).$]