

BIBLIOTECA DI SCIENZE STATISTICHE

SERVIZIO BIBLIOTECARIO NAZIONALE

BID P00891405 BID

ACQ. 37 / '03 INV. 82856

COLL. _____ CLASS. 5-COLL. WP. 2003/1

**Bias prevention of
maximum likelihood
estimates: skew normal and
skew t distributions**

N. Sartori

2003.1

**Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova**

Gennaio 2003

RECEIVED
JAN 10 1964
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C. 20250

This document is
intended to provide
information about
the use of the
equipment.

Page 1

Department of Agriculture
Washington, D.C. 20250
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C. 20250

Bias prevention of maximum likelihood estimates: skew normal and skew t distributions

Nicola Sartori*

Department of Statistics
University of Padova

January 7, 2003

Summary

The skew normal model is a class of distributions that extends the normal one by including a shape parameter. Despite its nice properties, this model presents some problems with the estimation of the shape parameter. In particular, for moderate sample sizes, the maximum likelihood estimator is infinite with positive probability. In this paper, we use the bias preventive method proposed by Firth (1993) to estimate the shape parameter. It is proved that this modified maximum likelihood estimator always exists finite. When regression and scale parameters are present, the method is combined with maximum likelihood estimators for these parameters. Finally, also the skew t distribution is considered, which may be viewed as an extension of the skew normal. The same method is applied to this model, considering the degrees of freedom as known.

Some key words: Bias reduction; Maximum likelihood; Monotone likelihood; Skewness; Skew normal; Skew t .

1 Introduction

The skew normal is a class of distribution that includes the normal one as a special case. In particular, a random variable Z is said to be skew normal with shape parameter λ , written $Z \sim SN(\lambda)$, if its density function at z is

$$f(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad (1)$$

where ϕ and Φ denote the $N(0, 1)$ density and distribution functions, respectively, and $z, \lambda \in \mathbb{R}$. We note that, when $\lambda = 0$, Z is a standard normal variable. More generally, positive (negative) values of λ correspond to positive (negative) skewness in the density (1). As a limit case, when λ tends to $+\infty$ ($-\infty$), Z tends to a positive (negative) half normal random variable. The skew

*For correspondence: Nicola Sartori, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, via C. Battisti, 241, 35121, Padova, Italy. E-mail: sartori@stat.unipd.it

normal shares with the normal distribution some nice theoretical properties and has the advantage of being suitable for the analysis of data with unimodal and skewed empirical distributions.

In practice, it is common to work with a location and scale transformation $Y = \mu + \sigma Z$, with $\mu \in \mathbb{R}$ and $\sigma > 0$. Hence, the density for the random variable Y , written $Y \sim SN(\mu, \sigma, \lambda)$, is

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

A systematic treatment of the skew normal distribution has been given in Azzalini (1985, 1986); generalizations to the multivariate case are given in Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999), while Azzalini & Capitanio (2003) propose a further extension with a skew t distribution, which extends the skew normal as the Student t extends the normal.

Despite its nice properties, the skew normal family has some inferential problems. Indeed, the estimation of the shape parameter may be problematic. For instance, consider the simple case with fixed location and scale parameters ($\mu = 0$ and $\sigma = 1$). If the observations are all positive (negative), the likelihood is monotone increasing (decreasing), giving an infinite maximum likelihood estimate (Liseo, 1990). Moreover, this occurs with positive probability, although this probability decreases rapidly with increasing sample sizes.

Also in the case with $Y \sim SN(\mu, \sigma, \lambda)$, this problem may arise. In some sense, things are even more complicated, because it is not so clear how to identify samples which lead to an infinite estimate of λ . The problem is however confined to the shape parameter. Indeed, empirical evidence shows that maximum likelihood estimates of location and scale parameters always exist.

Maximum likelihood is not the only frequentist method which encounters problems with the skew normal distribution. For instance, even the methods of moments can give very bad results. The same is true for maximum product of spacings (Cheng & Traylor, 1995), which sometimes gives an infinite estimate, even though not as frequently as maximum likelihood.

The behavior of the likelihood function suggests that some additional information from the model may be needed. Liseo & Loperfido (2002) perform a default Bayesian analysis for the skew normal distribution, and they prove that the Jeffreys' prior for the shape parameter is proper. In this paper, the additional information from the model is used in a frequentist method for bias prevention of the maximum likelihood estimator (Firth, 1993).

In the one parameter case, the score function is modified in such a way that the corresponding estimator has lower bias than the maximum likelihood estimator. The modification involves some expected quantities of log-likelihood derivatives, which need to be computed numerically. It is proved that this modified maximum likelihood estimator always exist finite, although simulation results show that it tends to underestimate the true parameter value for small sample sizes. In the general case, this method is combined with maximum likelihood estimates for location and scale parameters.

The skew t distribution extends the skew normal distribution. Unfortunately, the problem with the estimation of the skewness parameter extends as well. In particular, for fixed degrees of freedom, the problem occurs with exactly the same probability. This probability is lower when we consider the degrees of freedom as unknown, although other problems arise with the estimation of this parameter. Indeed, the maximum likelihood estimator for the degrees of freedom is very often infinite. Unknown degrees of freedom are not considered in this paper. Firth's method for bias prevention is applied also to the skew t model, leading to result similar to those of the skew

normal model. In this case, a formal proof of the existence of the estimate is not given, although in simulations the estimate always existed finite.

The structure of the paper is as follows. Section 2 describes the method for bias prevention due to Firth (1993). In Section 3, this method is applied to the skew normal distribution. In particular, in the scalar parameter case, it is proved the the modified maximum likelihood estimate always exists. Then, in the regression, scale and shape case, the method is combined with maximum likelihood estimates for location and scale parameters. Section 4 extends the method to the skew t distribution, with fixed degrees of freedom. Finally, a brief discussion is given.

2 Bias prevention for maximum likelihood estimator

In this section we briefly summarize the method introduced by Firth (1993). Let us consider a regular parametric model with log-likelihood $l(\theta)$. We assume for simplicity that the parameter θ is scalar, although the result holds for vector parameter as well. We denote by $U(\theta) = l'(\theta)$ and $j(\theta) = -l''(\theta)$, the score function and the observed information respectively, where l' and l'' denote first and second derivative of l . Some expected values of likelihood quantities will be needed. In particular, we will use the expected information, $i(\theta) = E_{\theta}\{j(\theta)\}$, and

$$\nu_{\theta,\theta,\theta}(\theta) = E_{\theta}\{l'(\theta)^3\}, \quad \nu_{\theta,\theta\theta}(\theta) = -E_{\theta}\{l'(\theta)l''(\theta)\}.$$

Typically, in samples with n independent observations, $j(\theta)$ is of order $O_p(n)$ and $i(\theta)$, $\nu_{\theta,\theta,\theta}(\theta)$ and $\nu_{\theta,\theta\theta}(\theta)$ are quantities of order $O(n)$.

It is well known that the maximum likelihood estimator has, in general, a bias of order $O(n^{-1})$ (see, for instance, §9.2 of Cox & Hinkley, 1974). In particular, through a simple Taylor expansion of the likelihood equation, $U(\hat{\theta}) = 0$, it is straightforward to see that

$$E_{\theta}\{\hat{\theta} - \theta\} = b(\theta) + O(n^{-2}),$$

where

$$b(\theta) = -\frac{1}{2}i(\theta)^{-2}\{\nu_{\theta,\theta,\theta}(\theta) + \nu_{\theta,\theta\theta}(\theta)\} = O(n^{-1}).$$

Hence, the bias corrected estimator

$$\hat{\theta}_{BC} = \hat{\theta} - b(\hat{\theta}),$$

is such that $E_{\theta}\{\hat{\theta}_{BC} - \theta\} = O(n^{-2})$. The drawback of this method is that it requires $\hat{\theta}$ to exist finite. The same is true for other bias correcting methods, such as jackknife or bootstrap.

The idea of Firth is to prevent the bias $b(\theta)$ before computing $\hat{\theta}$, by adding a modification $M(\theta)$, of order $O(1)$, to the score function. This leads to the modified likelihood equation

$$U^*(\theta) = U(\theta) + M(\theta) = 0. \quad (3)$$

The solution of (3) gives the estimate $\hat{\theta}^*$. The modification $M(\theta)$ is chosen in such a way that

$$E_{\theta}\{\hat{\theta}^* - \theta\} = O(n^{-2}). \quad (4)$$

Using a Taylor expansion for $U^*(\hat{\theta}^*) = 0$ around θ , we can find an expansion for $\hat{\theta}^* - \theta$. Then, imposing condition (4), we find

$$M(\theta) = -i(\theta)b(\theta). \quad (5)$$

An asymptotically equivalent form of the modification replaces the expected information with the observed information in (5), giving $M^{obs}(\theta) = -j(\theta) b(\theta)$. The two forms of the modification are equal in full exponential models. In general models, Firth mistakenly claimed that $M^{obs}(\theta)$ is preferable in terms of second order efficiency. This is not true: there is no gain in efficiency using $M^{obs}(\theta)$ in place of $M(\theta)$ (Firth, 1995).

The estimator $\hat{\theta}^*$ is second order equivalent to $\hat{\theta}_{BC}$ and it has the advantage that it does not depend on the finiteness of $\hat{\theta}$. Hence, there could be instances in which $\hat{\theta}$ is not finite, while this is true for $\hat{\theta}^*$. Although this property has not been proved in general, one example is given in binomial logistic regression (see Example 3.3 in Firth, 1993). Another case is given in this paper.

We note that there is a connection between $M(\theta)$ and Bayesian model-based priors. Indeed, in full exponential models, $\hat{\theta}^*$ coincides with the mode of the posterior distribution obtained using Jeffreys prior $|i(\theta)|^{1/2}$, where in this case $i(\theta)$ is the expected information for a single observation (see section 3.1 in Firth, 1993). While this equivalence is not true in general models, we can still think the modification to the score function as the derivative of a model based penalization of the log-likelihood. This interpretation is definitely correct for the expected version $M(\theta)$, which does not depend on the observed sample, while is less adequate when we consider $M^{obs}(\theta)$.

3 Application to the skew normal distribution

In this section we apply Firth's method for bias prevention to the scalar skew normal distribution. The density function of a single observation y from a $SN(\mu, \sigma, \lambda)$ is given by (2). First, in section 3.1, we consider the standard case, with $\mu = 0$ and $\sigma = 1$, and show that $U^*(\lambda)$ has always a finite solution. Then, in section 3.2, we propose an extension to the case with unknown μ and σ .

3.1 Shape parameter

Let us consider a sample $z_1, \dots, z_n$ of independent observations from $Z \sim SN(\lambda)$. The log-likelihood for λ is given by

$$l(\lambda) = \sum_{i=1}^n \zeta_0(\lambda z_i),$$

where $\zeta_0(x) = \log\{2\Phi(x)\}$. We also define the derivative

$$\zeta_m(x) = \frac{d^m}{dx^m} \zeta_0(x)$$

and note that $\zeta_1(x) = \phi(x)/\Phi(x)$. All subsequent derivatives can be expressed as functions of $\zeta_1(x)$; in particular $\zeta_2(x) = -x\zeta_1(x) - \zeta_1(x)^2$ (Azzalini & Capitanio, 1999).

The maximum likelihood estimate may be infinite with positive probability. Indeed, when all the observations are positive, then $l(\lambda)$ is an increasing function that tends to a constant value as λ tends to $+\infty$. The problem is specular when all observations are negative, since if $Z \sim SN(\lambda)$ then $-Z \sim SN(-\lambda)$. For this reason, in the following, we will restrict our attention to positive values of λ . Of course, the probability of having a monotone log-likelihood depends on n and on the true value of λ (Liseo, 1990). For instance, if $\lambda = 5$ and $n = 20$, we have probability 0.273 of having all positive observations. This probability decreases to 0.002 if $n = 100$.

The behavior of the likelihood function suggests the need for some additional information from the model. One way to incorporate this information is through default Bayesian analysis. This is what is done in Liseo & Loperfido (2002). In the following, we consider an alternative, although related, frequentist approach: we try to reduce the bias of the maximum likelihood estimate by means of a penalization to the log-likelihood function. This penalization makes use of some expected values of log-likelihood derivatives and is implicitly defined by the modified score function (3).

The score function and the observed information are

$$U(\lambda) = \sum_{i=1}^n \zeta_1(\lambda z_i) z_i, \quad j(\lambda) = \lambda \sum_{i=1}^n \zeta_1(\lambda z_i) z_i^3 + \sum_{i=1}^n \zeta_1(\lambda z_i)^2 z_i^2.$$

The expected quantities that are needed to compute $M(\lambda)$ are

$$\begin{aligned} i(\lambda) &= n a_{22}(\lambda), \\ \nu_{\lambda, \lambda, \lambda}(\lambda) &= n a_{33}(\lambda), \\ \nu_{\lambda, \lambda \lambda}(\lambda) &= -n \{ \lambda a_{42}(\lambda) + a_{33}(\lambda) \}, \end{aligned}$$

where the quantities

$$a_{kh}(\lambda) = E_{\lambda} \{ Z^k \zeta_1(\lambda Z)^h \}$$

need to be computed numerically. Note that $a_{k1}(\lambda) = 0$ when k is odd, and that $a_{kh}(\lambda) \geq 0$ when both k and h are even. Hence, modification (5) has the following form

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

and it can be easily obtained using numerical integration. The shape of $M(\lambda)$ is shown in Figure 1. The function is odd and negative for all positive values of λ . It has minimum value $M(1.07) = -0.83$ and it is increasing to 0 for values greater than 1.07.

In the following, we prove that the modified likelihood equation (3) always gives a solution when the maximum likelihood estimate does not exist. Indeed, suppose that all observations are positive. Then the likelihood equation, $U(\lambda) = 0$, has no solution, because the score function is always positive. Moreover, $U(\lambda) \rightarrow 0^+$, as λ tends to $+\infty$. In particular, it is straightforward to see that $U(\lambda) = O_p(e^{-\lambda^2})$. The modification $M(\lambda)$ is negative for positive values of λ . In the Appendix, it is proved that $M(\lambda)$ tends to 0^- , as λ tends to infinity, and that it increases as $-\lambda^{-1}$. This guarantees that $U^*(\lambda) = U(\lambda) + M(\lambda) \rightarrow 0^-$. Finally, noting that $U^*(0) = U(0) > 0$, we have that there must exist at least a point λ^* such that $U^*(\lambda^*) = 0$. We have not proved that this solution is unique, although numerically we could not find cases with multiple solutions.

The observed version of the score modification is given by

$$M^{obs}(\lambda) = -\frac{j(\lambda)}{n} \frac{\lambda a_{42}(\lambda)}{2 a_{22}(\lambda)^2}.$$

In this case the proof of the existence of the associated estimate is not given. It can be seen that $M^{obs}(\lambda) = O_p(\lambda^3 e^{-\lambda^2})$. Anyway, this is only an upper bound and it does not ensure that

$U(\lambda) + M^{obs}(\lambda)$ is negative, when $\lambda \rightarrow \infty$. Moreover, there are two additional reasons to prefer the expected modification. First, $M^{obs}(\lambda)$ is not very stable numerically. Second, in simulations, $M(\lambda)$ gave better results in terms of bias. For all these reasons, in the following we concentrate only on the expected version.

Example 1. Consider the following data $z=(0.195, 0.847, 0.726, -0.139, 1.788, 0.570, 2.069, 0.452, 0.868, 1.199, 0.894, 0.887, 1.258, 0.918, 0.496, 0.183, 0.119, 1.207, 0.446, 2.579)$, that have been generated from a $SN(10)$. In this case, $\hat{\lambda} = 5.39$ and $\hat{\lambda}^* = 2.84$. The right panel of Figure 2 shows the score function, $U(\lambda)$, and the modified score function, $U^*(\lambda)$, for the data z , together with the modification $M(\lambda)$. Let us consider now the absolute value of z . We note that only the fourth observation changes. In this case, $\hat{\lambda}$ does not exist, while $\hat{\lambda}^* = 3.92$. The right hand panel of Figure 2 shows $U(\lambda)$ and $U^*(\lambda)$, for the data $|z|$, together with the modification $M(\lambda)$.

As noted in the previous section, in full exponential models $\hat{\lambda}^*$ coincides with the mode of the posterior density, obtained using Jeffreys' prior, here denoted by $\hat{\lambda}^J$. In this case, the two estimates are not equal. In particular, $\hat{\lambda}^J$ is the solution of the equation $U(\lambda) + M^J(\lambda) = 0$, where

$$M^J(\lambda) = -\frac{1}{2}i(\theta)^{-2}\{\nu_{\theta,\theta,\theta}(\theta) + 2\nu_{\theta,\theta\theta}(\theta)\} = -\frac{1}{2}\frac{2\lambda a_{42}(\lambda) + a_{33}(\lambda)}{a_{22}(\lambda)}$$

is the derivative of the logarithm of Jeffreys' prior. Although $M^J(\lambda)$ is not equal to $M(\lambda)$, their shapes are very similar, as shown in Figure 3. This suggests that $\hat{\lambda}^J$ might be close to $\hat{\lambda}^*$ in practice. This relation is not further explored in the present context.

Tables 1 reports some simulation results, for $\lambda = 5$ and $\lambda = 10$, in which we compared the bias of $\hat{\lambda}$ and $\hat{\lambda}^*$. As expected, $\hat{\lambda}^*$ always exists. Moreover, in all simulated samples the solution to (3) was unique. On the other hand, for small samples and large values of λ , $\hat{\lambda}^*$ has a negative bias which can be substantial. Anyway, this bias should be compared with the corresponding proportion of non existence of $\hat{\lambda}$. In other words, this is the price to pay to get a finite estimate when $\hat{\lambda}$ does not exist. For instance, when $n = 20$ and $\lambda = 10$, $\hat{\lambda}^*$ has a bias equal to -5.933 , but $\hat{\lambda}$ is infinite in more than a half of the samples.

3.2 Shape parameter with location and scale

In this section we consider inference on the general case, with unknown location, scale and shape parameters. If we consider a sample $y_1, \dots, y_n$ of independent and identically distributed observations from $Y \sim SN(\mu, \sigma, \lambda)$, with density given by (2), the log-likelihood for the parameter $\theta = (\mu, \sigma, \lambda)$ has the form

$$l(\theta) = l(\mu, \sigma, \lambda) = -n \log \sigma - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma} \right)^2 + \sum_{i=1}^n \zeta_0 \left\{ \lambda \left(\frac{y_i - \mu}{\sigma} \right) \right\}.$$

Even in this setting $\hat{\lambda}$ might be infinite, although less frequently than in the case with only the shape parameter. In particular, this happens when the maximum likelihood estimate of the location parameter, $\hat{\mu}$, tends to the smallest (largest) observation in the sample. In such cases, the log-likelihood is monotone increasing (decreasing) and tends to a constant value when λ tends

to plus (minus) infinity. We note that only the shape parameter has the problem of existence of the estimate. Indeed, empirical evidence indicates that, in cases with $\hat{\lambda} = \infty$, we have

$$\lim_{\lambda \rightarrow \infty} l(\hat{\mu}_\lambda, \hat{\sigma}_\lambda, \lambda) = l(\hat{\mu}, \hat{\sigma}, \infty) = k,$$

where $\hat{\mu}_\lambda$ and $\hat{\sigma}_\lambda$ denote the maximum likelihood estimate for fixed λ of the location and shape parameters, respectively, and k is a constant value. Hence, maximum likelihood estimates of μ and σ always exist (finite).

In principle, it would be possible to apply Firth's method simultaneously to all components of the parameter θ . This approach is not proposed here for several reasons. First of all, it requires the computation of a larger number of expected values $a_{kh}(\lambda)$. Indeed, the quantities $\nu_{\theta, \theta, \theta}(\theta)$ and $\nu_{\theta, \theta \theta}(\theta)$ are $3 \times 3 \times 3$ arrays. The dimension is even larger if we consider a regression setting, with $\mu_i = x_i^T \beta$, where x_i is a vector of explanatory variables and β is a vector of unknown parameters, both of dimension p . Hence, the numerical computation of the modification $M(\theta)$ may not be as fast as in the one parameter case.

A second, and more important, reason is that the method becomes numerically unstable. In fact, very often, algorithms for solving the modified likelihood equation (3) do not reach convergence. Moreover, even when they do converge, the behavior of $\hat{\mu}^*$ and $\hat{\sigma}^*$ is quite suspicious: in some cases $\hat{\mu}^*$ is far from the observed sample and $\hat{\sigma}^*$ gets much larger than $\hat{\sigma}$. This behavior looks strange in particular when compared with the behavior of $\hat{\mu}$ and $\hat{\sigma}$. Indeed, simulation results showed that maximum likelihood estimates of μ and σ have quite good properties in terms of bias and mean square error, as also shown at the end of this section.

The nice behavior of $\hat{\mu}$ and $\hat{\sigma}$ suggests a two-steps method of estimation in the location, scale and shape case. In the first step, we compute maximum likelihood estimates of μ and σ . Then, we consider $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$ as fixed, and we apply Firth's method for the shape parameter only, as in the previous section. In practice, we keep the part of the maximum likelihood estimate that works fine and we modify only the other component.

This method has the advantage of being computationally simple and fast. The result of the previous section also guarantees that an estimate of the shape parameter always exists. Moreover, extensions to the regression setting are straightforward. On the other hand, we note that this procedure uses plug-in estimates of unknown parameters. Indeed, $z_i = (y_i - \mu)/\sigma$ are independent observations from a $SN(\lambda)$. What we do is to standardize the observations using the maximum likelihood estimates of location and scale, $\hat{z}_i = (y_i - \hat{\mu})/\hat{\sigma}$, and then assume that $\hat{z}_1, \dots, \hat{z}_n$ are still independent observations from a $SN(\lambda)$. Anyway, the simulation results showed that $\hat{\lambda}^*$ is not strongly affected by this assumption for reasonable sample sizes.

Example 2. Consider the dataset discussed in Section 6.3 of Azzalini & Capitanio (1999), also available at <http://azzalini.stat.unipd.it/SN/index.html>, which is a random sample of size 50 from a $SN(0, 1, 5)$. This dataset, reproduced in Figure 4, has some interest because it leads to infinite maximum likelihood estimate of λ . Similar problems arise with other frequentist methods. Instead, the procedure described above gives $\hat{\lambda}^* = 9.14$. In Figure 4, the solid curve represents the density corresponding to $\hat{\lambda}$, while the dashed curve represents the one corresponding to $\hat{\lambda}^*$.

Example 3. Consider the `ais` data, included in the library `sn` for R (available for download at <http://azzalini.stat.unipd.it/SN/index.html>). We define the model

$$\text{Bfat}_i = \beta_0 + \beta_1 \text{Wt}_i + \sigma Z_i, \quad i = 1, \dots, 202,$$

where Bfat_i is the body fat percentage of the i -th subject, Wt_i is the weight and $Z_1, \dots, Z_{202}$ are assumed independent realizations of a $SN(\lambda)$. The maximum likelihood estimates of the parameters are $\hat{\beta}_0 = 5.288$, $\hat{\beta}_1 = 0.308$, $\hat{\sigma} = 47.585$ and $\hat{\lambda} = +\infty$. The modified estimate of the shape parameter is $\hat{\lambda}^* = 23.65$. Figure 5 shows the histogram of the residuals $\hat{z}_i = (\text{Bfat}_i - \hat{\beta}_0 - \hat{\beta}_1 \text{Wt}_i) / \hat{\sigma}$. Solid and dashed lines represents the skew normal density with shape parameter $\hat{\lambda}$ and $\hat{\lambda}^*$, respectively.

Table 2 reports bias of maximum likelihood and modified maximum likelihood estimators in some simulation results, for $\mu = 0$, $\sigma = 1$ and $\lambda = 5, 10$. We see that the behaviour of the maximum likelihood estimator for location and scale parameters is quite good and it does not seem to be affected by the value of the shape parameter. The percentage of cases in which $\hat{\lambda} = +\infty$ is not negligible and is larger for small sample sizes and $\lambda = 10$. Even in this case, $\hat{\lambda}^*$ is always finite, although its bias has a strange non-monotone behaviour as n increases. This is due to large estimates that, in the mean, balance the tendency to underestimate the true λ , in particular for small sample sizes. The median does not have this problem. Indeed, in the simulation with $\lambda = 5$, the median of the simulated empirical distribution of $\hat{\lambda}^*$ is equal to 4.469 when $n = 50$, 4.688 when $n = 100$ and 4.828 when $n = 200$. This fact is confirmed in Figure 6, where nonparametric density estimates of the empirical distribution of $\hat{\lambda}^*$ are reported. Similar considerations apply to the case with $\lambda = 10$.

4 Application to the skew t distribution

In the following we apply the method of the previous section to the skew t distribution, with known degrees of freedom. The skew t distribution is a natural generalization of the skew normal distribution, which allows to model the tails of the distribution as well as its skewness. Below, we give the definition of Azzalini & Capitanio (2003), although this coincides with the proposal of Branco & Dey (2001). Both papers deal with the general multivariate case. Here, we restrict our attention to the scalar case.

Let us consider a random variable $Z \sim SN(\lambda)$, with density given by (1), and another random variable V , independent of Z , with distribution $V \sim \chi_\nu^2 / \nu$. Then

$$Y = \mu + \sigma V^{-1/2} Z \tag{6}$$

has the skew t distribution, written $Y \sim ST(\mu, \sigma, \lambda, \nu)$. The density of Y is given by

$$f(y; \mu, \sigma, \lambda, \nu) = 2 t(y; \nu) T \left\{ \lambda \frac{y - \mu}{\sigma} \left(\frac{\nu + 1}{Q_y + \nu} \right)^{1/2}; \nu + 1 \right\}, \tag{7}$$

where

$$Q_y = \left(\frac{y - \mu}{\sigma} \right)^2,$$

$$t(y; \nu) = \frac{1}{\sigma} \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma\{(\nu+1)/2\}}{\Gamma(\nu/2)} \left(1 + \frac{Q_y}{\nu} \right)^{-(\nu+1)/2}$$

and $T(x; \nu_0)$ is the t distribution function with ν_0 degrees of freedom. We note that, as ν goes to infinity, the skew t distribution tends to the skew normal distribution. In the present context, we consider the degrees of freedom as fixed and we apply Firth's method. As with the skew normal distribution, we deal first with the shape parameter only ($\mu = 0$ and $\sigma = 1$) and then we combine this solution with the maximum likelihood estimates for μ and σ .

Let us consider a sample $y_1, \dots, y_n$ of independent and identically distributed observations from $Y \sim ST(0, 1, \lambda, \nu)$, with known $\nu > 0$. The log-likelihood for λ is given by

$$l(\lambda) = \sum_{i=1}^n \log T(\lambda u_i; \nu + 1),$$

where $u_i = y_i \sqrt{(\nu+1)/(y_i^2 + \nu)}$.

As with the skew normal distribution, the maximum likelihood estimate may be infinite with positive probability. In particular, also in this case, when all the observations are positive, then $l(\lambda)$ is an increasing function that tends to a constant value as λ tends to $+\infty$. We note that, for given n and λ , the probability of having $\hat{\lambda} = \infty$ is the same as the one of the corresponding skew normal log-likelihood, irrespectively of the (fixed) degrees of freedom. This can be easily seen in (6). Here, we apply the same method of the previous section.

The score function and observed information are

$$U(\lambda) = \sum_{i=1}^n \frac{t(\lambda u_i; \nu + 1) u_i}{T(\lambda u_i; \nu + 1)},$$

$$j(\lambda) = \sum_{i=1}^n \frac{1}{T(\lambda u_i; \nu + 1)} \left\{ \frac{\nu+1}{\nu+2} \lambda u^3 \left(1 + \frac{\lambda^2 u_i^2}{\nu+1} \right)^{-1} t(\lambda u_i; \nu + 1) \right\} +$$

$$\sum_{i=1}^n \frac{t(\lambda u_i; \nu + 1)^2 u_i^2}{T(\lambda u_i; \nu + 1)^2}.$$

Using these quantities we can compute the needed ingredients for the modification $M(\lambda)$, as defined by (5), through numerical integration. Figure 7 shows $M(\lambda)$ for different values of ν , together with $M(\lambda)$ for the skew normal distribution. The modifications for the skew t distribution have shape similar to the skew normal one, in particular as ν increases. For small values of ν the penalization of the score function for small values of $|\lambda|$ is lower than what occurs in the skew normal case.

Example 4. We generated a sample of size 30 from a $ST(0, 1, 5, 3)$:
 $y = (0.4493, 0.9244, 2.3975, 1.6091, 1.1315, 1.7909, 0.2530, 1.0535, 0.8879, 0.0635, 0.6291, 2.7096,$

0.2444, 2.5884, 0.6907, 0.0996, 0.3889, 0.0651, 0.7204, 0.7187, 1.0224, 0.3116, 1.1170, 0.8715, -0.0491, 0.6654, 1.7649, 0.3744, 0.5656, 0.5284).

This sample has only one negative observation, equal to -0.0491. The maximum likelihood estimate is $\hat{\lambda} = 13.496$ while the modified estimate is $\hat{\lambda}^* = 6.586$. If we consider the absolute value of the negative observation, we have $\hat{\lambda} = +\infty$, while $\hat{\lambda}^* = 12.042$.

In what done so far, we assumed ν as known. Hence, it may be interesting to consider what happens to $\hat{\lambda}$ and $\hat{\lambda}^*$ when ν changes. Figure 8 shows the two estimates for values of ν that range from 1 to 50, for the data of Example 4. In this case, $\hat{\lambda}^*$ is more affected by the choice of ν than $\hat{\lambda}$. Anyway, in both cases, the behavior of the estimates is smooth and it tends to the horizontal line, which corresponds to the estimate in the skew normal case.

As for the skew normal distribution, the observed modification $M^{obs}(\lambda)$ gave numerically unstable results. Therefore, we restricted our attention to the expected version. On the other hand, in this case we have not proved the existence of $\hat{\lambda}^*$, when $\hat{\lambda} = \infty$. Anyway, it seems reasonable to us that a proof along the lines of the skew normal one could probably be done, since, in all simulated samples, $\hat{\lambda}^*$ was found to exist and to be unique. Table 3 reports some of these simulations results. Similarly to the skew normal case, $\hat{\lambda}^*$ always exists but it tends to have a negative bias, particularly noticeable for small n and large λ . Anyway, this bias is lower than the corresponding one for the skew normal case, and it is lower for small values of ν .

Finally, let us consider the general case, where $y_1, \dots, y_n$ are independent and identically distributed observations from a $ST(\mu, \sigma, \lambda, \nu)$, with ν known. The log-likelihood is

$$l(\mu, \sigma, \lambda) = \sum_{i=1}^n \log f(y_i; \mu, \sigma, \lambda, \nu),$$

where $f(y_i; \mu, \sigma, \lambda, \nu)$ is given by (7). As with the skew normal distribution, also in the more general case, $\hat{\lambda}$ might be infinite. However, this problem affects only the skewness parameter. Indeed, $\hat{\mu}$ and $\hat{\sigma}$ always exist finite and they have a nice asymptotic behavior.

Hence, we follow the same procedure of Section 3.2: we first compute $\hat{\mu}$ and $\hat{\sigma}$. Then, we standardize the observation, giving $\hat{z}_i = (y_i - \hat{\mu})/\hat{\sigma}$, and we apply Firth's method for the shape parameter only, using $(\hat{z}_1, \dots, \hat{z}_n)$ as observations. Some simulation results are reported in Table 4, for various values of ν and λ . These results show the same behavior of those in Table 2. Therefore, for some comments, see the final paragraph of Section 3.2.

5 Discussion

We proposed a modified maximum likelihood estimator for the shape parameter of the skew normal and skew t distributions. By construction, this estimator has lower bias than the maximum likelihood estimator and, in the present context, has the advantage of being always finite. To our knowledge, this is the first example in which it is proved that Firth's method always yields a finite estimate when the maximum likelihood estimate is infinite.

The proposed estimator is not invariant with respect to reparameterizations. Hence, a careful choice of the parameterization is crucial. Working with the somehow natural shape parameter λ has the advantage that $\hat{\lambda}^*$ has a distribution close to normality. An alternative parameterization

could be $\delta = \lambda/(1 + \lambda^2)^{1/2}$, which gives all values between -1 and 1 as the parameter space. In this case, $\hat{\delta}$ has values on the boundary of the parameter space with positive probability, while $\hat{\delta}^*$ does not have this problem. However, the distribution of $\hat{\delta}^*$ is far from being normal (except for large n). Therefore, it seems preferable to work with the parameter λ . Of course, with large sample sizes, in practice absence of invariance tends to vanish.

Nothing has been said in this paper about standard errors and confidence intervals based on $\hat{\lambda}^*$. Since the modification $M(\lambda)$ is of order $O(1)$, the first order asymptotic variance of $\hat{\lambda}^*$ is the same as that of $\hat{\lambda}$, namely $i(\lambda)^{-1}$, and this can be used in the usual way for providing standard errors. In the location, scale and shape case, for the same reason, we can use the element corresponding to λ of the matrix $i(\theta)^{-1}$.

Appendix

In the following we prove that $M(\lambda) = \Omega(\lambda^{-1})$, where we use the order notation $f = \Omega(g) \iff (f = O(g) \text{ and } g = O(f))$. Intuitively, this means that f/g is bounded above and below. This condition guarantees that $U^*(\lambda) = 0$ has a solution when $\hat{\lambda}$ does not exist, as shown in Section 3.1.

The order of the modification $M(\lambda) = -\lambda a_{42}(\lambda)/\{2 a_{22}(\lambda)\}$ depends on the order of $a_{22}(\lambda)$ and $a_{42}(\lambda)$. In particular, the result $M(\lambda) = \Omega(\lambda^{-1})$ follows from the fact that $a_{42}(\lambda) = \Omega(\lambda^{-5})$ and $a_{22}(\lambda) = \Omega(\lambda^{-3})$. Both $a_{42}(\lambda)$ and $a_{22}(\lambda)$ are positive even functions, and they are decreasing for $\lambda > 0$. In Appendix A of Liseo & Loperfido (2002) this is proved for $a_{22}(\lambda)$. It is also proved that $a_{22}(\lambda) = O(\lambda^{-3})$. With a slight modification of their argument, it is also possible to show that $a_{22}(\lambda) = \Omega(\lambda^{-3})$. Since the proof for $a_{42}(\lambda) = \Omega(\lambda^{-5})$ uses exactly the same steps, only this one is given in what follows.

The symmetry of $a_{42}(\lambda)$ is proved by splitting the integral in two parts

$$\begin{aligned} a_{42}(\lambda) &= 2 \int_0^\infty z^4 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz + 2 \int_0^\infty z^4 \phi(-z) \frac{\phi^2(-\lambda z)}{1 - \Phi(\lambda z)} dz \\ &= 2 \int_0^\infty z^4 \phi(z) \phi^2(-\lambda z) \frac{1}{\Phi(\lambda z) \{1 - \Phi(\lambda z)\}} dz = a_{42}(-\lambda). \end{aligned}$$

The monotonicity of $a_{42}(\lambda)$, for $\lambda > 0$, is proved noting that $\partial a_{42}(\lambda)/\partial \lambda$ is negative. Indeed,

$$\frac{\partial a_{42}(\lambda)}{\partial \lambda} = -2 \int_0^\infty z^5 \frac{\phi(z) \phi^2(\lambda z)}{\Phi^2(\lambda z) \{1 - \Phi(\lambda z)\}^2} H(\lambda z) dz,$$

where $H(s) = 2 s \Phi(s) \{1 - \Phi(s)\} + \phi(s) \{1 - 2\Phi(s)\}$. Since, for all $s > 0$, $s\Phi(-s) < \phi(s)$, then $H(s) > 0$ and the result follows.

From the symmetry of $a_{42}(\lambda)$, it is enough to study the right tail only. Let

$$\begin{aligned} a_{42}(\lambda) &= A(\lambda) + B(\lambda) \\ &= 2 \int_0^\infty z^4 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz + 2 \int_0^\infty z^4 \phi(z) \frac{\phi^2(\lambda z)}{1 - \Phi(\lambda z)} dz. \end{aligned}$$

For all $\lambda > 0$ and $z > 0$, $1 < 1/\Phi(\lambda z) < 2$. Then $A^*(\lambda) < A(\lambda) < 2A^*(\lambda)$, with

$$A^*(\lambda) = 2 \int_0^\infty z^4 \phi(z) \phi^2(\lambda z) dz = \frac{3}{2\pi(1+2\lambda^2)^{5/2}}.$$

Hence, $A(\lambda) = \Omega(\lambda^{-5})$. We also note that $B(\lambda) > A^*(\lambda) = O(\lambda^{-5})$, since $1 - \Phi(\lambda z) < 1$. Therefore, it only remains to prove that $B(\lambda) = O(\lambda^{-5})$. Let us write

$$\begin{aligned} B(\lambda) &= B_1(\lambda) + B_2(\lambda) \\ &= 2 \int_0^{1/\lambda} z^4 \phi(z) \frac{\phi^2(\lambda z)}{1 - \Phi(\lambda z)} dz + 2 \int_{1/\lambda}^\infty z^4 \phi(z) \frac{\phi^2(\lambda z)}{1 - \Phi(\lambda z)} dz. \end{aligned}$$

When $0 < z < \lambda^{-1}$, $\Phi(\lambda z) < \Phi(1)$; then $\{1 - \Phi(\lambda z)\}^{-1} < 6.30 = \{1 - \Phi(1)\}^{-1} = c$. It follows that $B_1(\lambda) < c/2B_1^*(\lambda)$, with

$$B_1^*(\lambda) = 4 \int_0^{1/\lambda} z^4 \phi(z) \phi^2(\lambda z) dz = \frac{4}{\pi^{3/2}} \left(\frac{1}{1+2\lambda^2} \right)^{5/2} \tilde{\Gamma} \left(\frac{5}{2}, 1 + \frac{1}{2\lambda^2} \right),$$

and where $\tilde{\Gamma}(a, x)$ denotes the incomplete Gamma function

$$\tilde{\Gamma}(a, x) = \int_0^x t^{a-1} \exp -t dt.$$

It follows that $B_1(\lambda) = O(\lambda^{-5})$. Finally, for fixed positive λ and $z > \lambda^{-1}$ we use the following inequality (Feller, 1971)

$$\frac{1}{1 - \Phi(\lambda z)} \leq \frac{\lambda^3 z^3}{(\lambda^2 z^2 - 1)\phi(\lambda z)}.$$

Hence

$$\begin{aligned} B_2(\lambda) &\leq 2\lambda^3 \int_{1/\lambda}^\infty \frac{z^7}{\lambda^2 z^2 - 1} \phi(z) \phi(\lambda z) dz \\ &\leq 2\lambda^3 \int_{1/\lambda}^\infty z^7 \phi(z) \phi(\lambda z) dz < \frac{2}{2\pi} \frac{\lambda^3}{(1 + \lambda^2)^4} \int_0^\infty t^7 \phi(t) dt = \frac{38.298}{2\pi} \frac{\lambda^3}{(1 + \lambda^2)^4}, \end{aligned}$$

giving $B(\lambda) = O(\lambda^{-5})$. This also implies that $B(\lambda) = \Omega(\lambda^{-5})$ and completes the proof.

Acknowledgments

I am grateful to Adelchi Azzalini for helpful discussions and encouragement. This research has been partially supported by a grant (PRIN 2000) from Ministero dell'Istruzione, dell'Università e della Ricerca, Italy.

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171-178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **46**, 199-208.
- Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Statist. Soc.*, B **61**, 579-602.
- Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc.*, B **65**, to appear.
- Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew normal distribution. *Biometrika*, **83**, 715-726.
- Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew elliptical distributions. *J. Multivariate Anal.*, **79**, 99-113.
- Cheng, C.H. & Traylor, L. (1995). Non-regular maximum likelihood problems (with discussion). *J. Roy. Statist. Soc.*, B **57**, 3-44.
- Cox, D.R. & Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- Feller, W. (1971). *An Introduction to Probability Theory, Vol. II*. Wiley, New York.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27-38.
- Firth, D. (1995). *Amendments and corrections: Bias reduction of maximum likelihood estimates. Biometrika*, **82**, 667-668.
- Liseo, B. (1990). The skew-normal class of densities: Inferential aspects from a Bayesian viewpoint (Italian). *Statistica*, **50**, 59-70.
- Liseo, B. & Loperfido, N. (2002). Default Bayesian analysis of the skew normal distribution. Submitted. Available at:
<http://geostasto.eco.uniroma1.it/utenti/liseo/liseosent.pdf>.

λ	n	$E_{\lambda}\{\hat{\lambda} - \lambda\}^{(*)}$	$E_{\lambda}\{\hat{\lambda}^* - \lambda\}$	$\% (\hat{\lambda} < +\infty)$
5	20	2.030	-1.687	72.6 (72.7)
	50	1.815	-0.254	95.8 (96.1)
	100	0.825	0.006	99.8 (99.8)
10	20	1.926	-5.983	48.3 (47.5)
	50	4.195	-2.635	80.2 (80.1)
	100	4.181	-0.580	96.1 (96.0)

(*) computed when $\hat{\lambda} < +\infty$

Table 1: Skew normal: simulation results with 5000 iterations. Bias of $\hat{\lambda}$ (computed only when $\hat{\lambda} < +\infty$), bias of $\hat{\lambda}^*$ and empirical (theoretical) percentage of cases when $\hat{\lambda}$ exists.

λ	n	$E_{\theta}\{\hat{\mu} - \mu\}$	$E_{\theta}\{\hat{\sigma} - \sigma\}$	$E_{\theta}\{\hat{\lambda} - \lambda\}^{(*)}$	$E_{\theta}\{\hat{\lambda}^* - \lambda\}$	$\% (\hat{\lambda} < +\infty)$
5	50	0.026	-0.023	1.560	0.098	87.2
	100	0.010	-0.009	1.384	0.437	98.0
	200	0.004	-0.004	0.556	0.186	99.9
10	50	0.023	-0.024	0.889	-2.248	66.3
	100	0.008	-0.009	3.257	0.302	87.7
	200	0.003	-0.002	2.720	0.971	98.2

(*) computed when $\hat{\lambda} < +\infty$

Table 2: Skew normal: simulation results when $\mu = 0$, $\sigma = 1$ and $\lambda = 5, 10$ and with 5000 iterations. Bias of $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ (computed only when $\hat{\lambda} < +\infty$) and $\hat{\lambda}^*$. Last column reports the empirical percentage of cases when $\hat{\lambda}$ exists.

ν	λ	n	$E_{\lambda}\{\hat{\lambda} - \lambda\}^{(*)}$	$E_{\lambda}\{\hat{\lambda}^* - \lambda\}$	% ($\hat{\lambda} < +\infty$)
3	5	20	2.174	-0.935	72.1 (72.7)
		50	1.760	0.185	95.8 (96.1)
		100	0.753	0.117	99.8 (99.8)
	10	20	0.745	-4.668	47.8 (47.5)
		50	3.592	-1.119	79.9 (80.1)
		100	3.113	0.235	95.7 (96.0)
10	5	20	2.048	-1.477	73.3 (72.7)
		50	1.742	-0.138	96.1 (96.1)
		100	0.756	0.042	99.8 (99.8)
	10	20	1.894	-5.483	48.6 (47.5)
		50	4.482	-2.058	80.0 (80.1)
		100	3.470	-0.225	95.5 (96.0)

(*) computed when $\hat{\lambda} < +\infty$

Table 3: Skew t : simulation results when $\nu = 3, 10$, $\lambda = 5, 10$ and with 5000 iterations. Bias of $\hat{\lambda}$ (computed only when $\hat{\lambda} < +\infty$), bias of $\hat{\lambda}^*$ and empirical (theoretical) percentage of cases when $\hat{\lambda}$ exists.

ν	λ	n	$E_{\theta}\{\hat{\mu} - \mu\}$	$E_{\theta}\{\hat{\sigma} - \sigma\}$	$E_{\theta}\{\hat{\lambda} - \lambda\}^{(*)}$	$E_{\theta}\{\hat{\lambda}^* - \lambda\}$	% ($\hat{\lambda} < +\infty$)	
3	5	50	0.016	-0.006	1.156	1.084	85.5	
		100	0.009	-0.003	1.592	0.866	98.2	
		200	0.006	-0.003	0.586	0.273	99.9	
	10	50	0.015	-0.010	-0.081	0.336	60.6	
		100	0.007	-0.004	3.766	1.983	89.5	
		200	0.002	-0.001	3.029	1.721	98.5	
	5	5	50	0.017	-0.009	1.439	0.890	84.8
			100	0.009	-0.005	1.607	0.755	98.3
			200	0.006	-0.004	0.568	0.226	100.0
10		50	0.016	-0.013	1.003	-0.389	62.7	
		100	0.005	-0.002	4.327	1.569	90.0	
		200	0.003	-0.002	3.118	1.280	99.1	
10	5	50	0.017	-0.014	1.984	0.504	87.5	
		100	0.004	-0.004	1.801	0.769	97.9	
		200	0.007	-0.006	0.516	0.184	100.0	
	10	50	0.017	-0.018	1.908	-1.312	67.2	
		100	0.006	-0.006	4.595	0.972	89.5	
		200	0.003	-0.002	3.009	1.076	98.8	

(*) computed when $\hat{\lambda} < +\infty$

Table 4: Skew t : simulation results when $\mu = 0$, $\sigma = 1$, $\nu = 3, 5, 10$ and $\lambda = 5, 10$ and with 5000 iterations. Bias of $\hat{\mu}$, $\hat{\sigma}$, $\hat{\lambda}$ (computed only when $\hat{\lambda} < +\infty$) and $\hat{\lambda}^*$. Last column reports the empirical percentage of cases when $\hat{\lambda}$ exists.

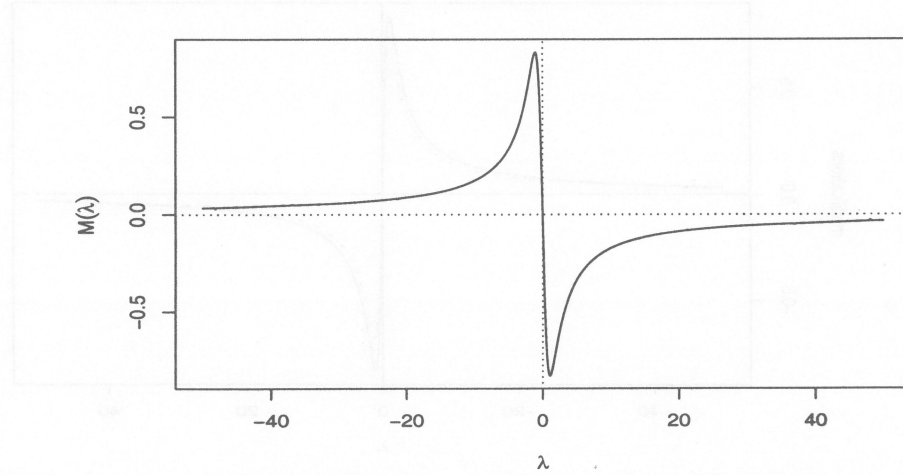


Figure 1: Modification to the score function.

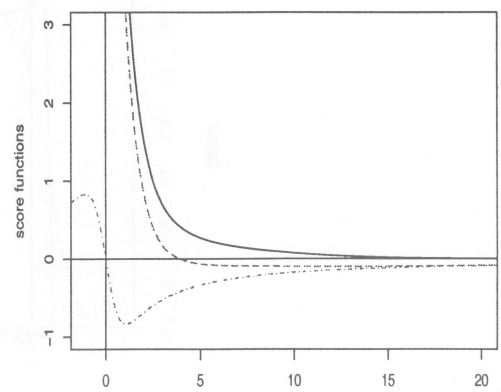
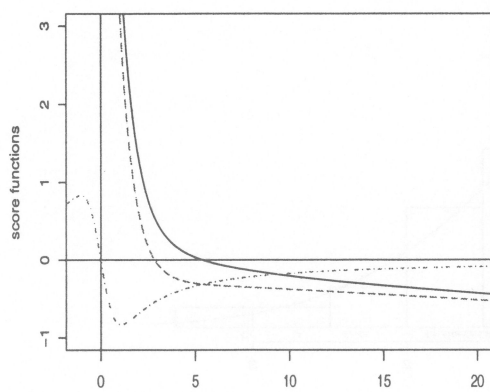


Figure 2: Score functions: $U(\lambda)$ (solid), $U^*(\lambda)$ (dashed) and the modification $M(\lambda)$ (dot-dashed). Data: z (left panel), $|z|$ (right panel).

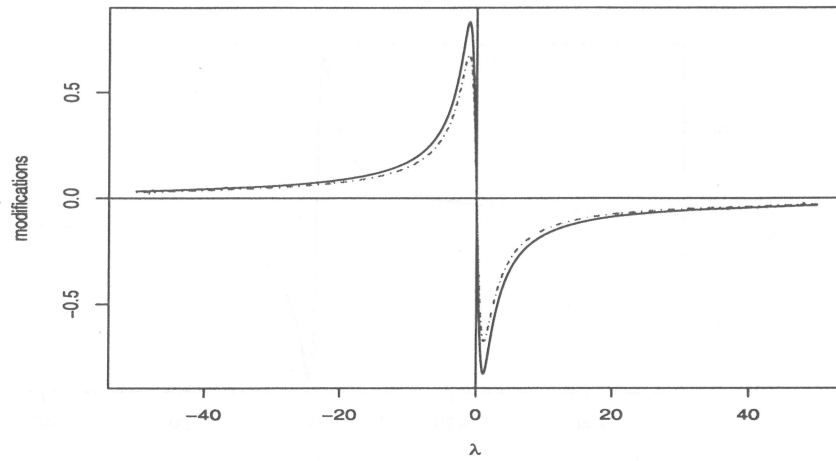


Figure 3: Modifications to the score function: $M(\lambda)$ (solid) and $M^J(\lambda)$ (dot-dashed).

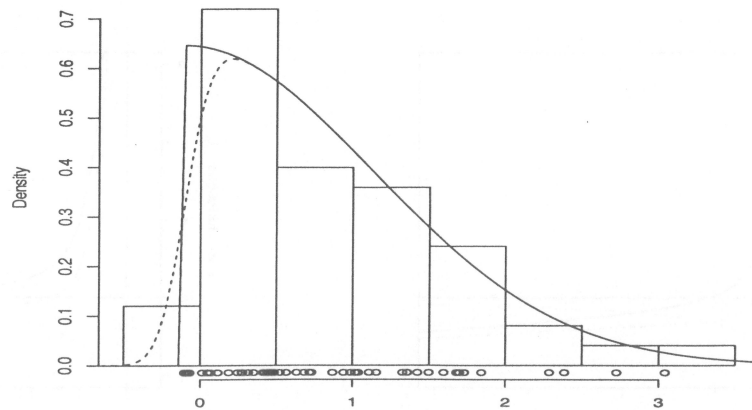


Figure 4: Histogram of simulated data points (small circles), leading to $\hat{\lambda} = +\infty$ (solid curve). The modified estimate is $\hat{\lambda}^* = 9.14$ (dashed curve).

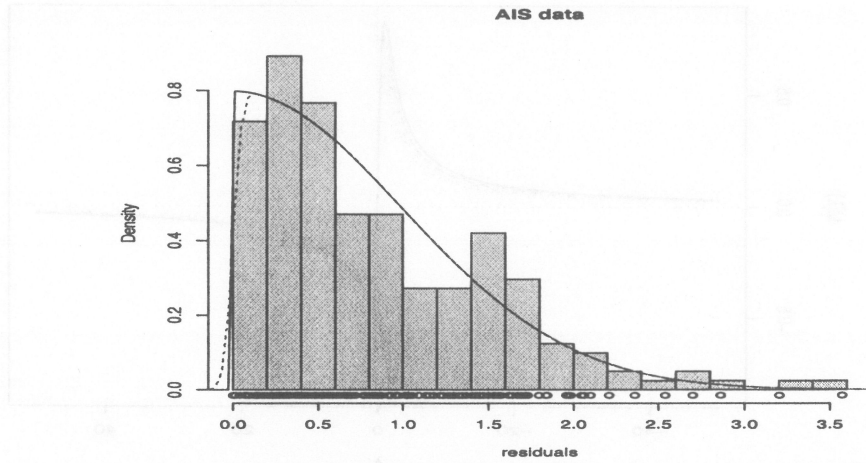


Figure 5: Ais data. Histogram of the residuals (small circles), leading to $\hat{\lambda} = +\infty$ (solid curve). The modified estimate is $\hat{\lambda}^* = 23.65$ (dashed curve).

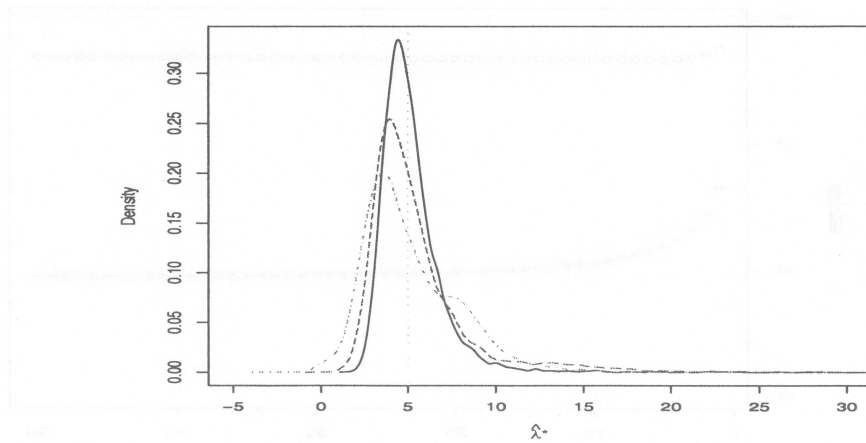


Figure 6: Simulation with $\xi = 0$, $\sigma = 1$ and $\lambda = 5$ and with 5000 iterations. Nonparametric density estimates of the distribution of $\hat{\lambda}^*$: $n = 50$ (dot-dashed), $n = 100$ (dashed) and $n = 200$ (solid).

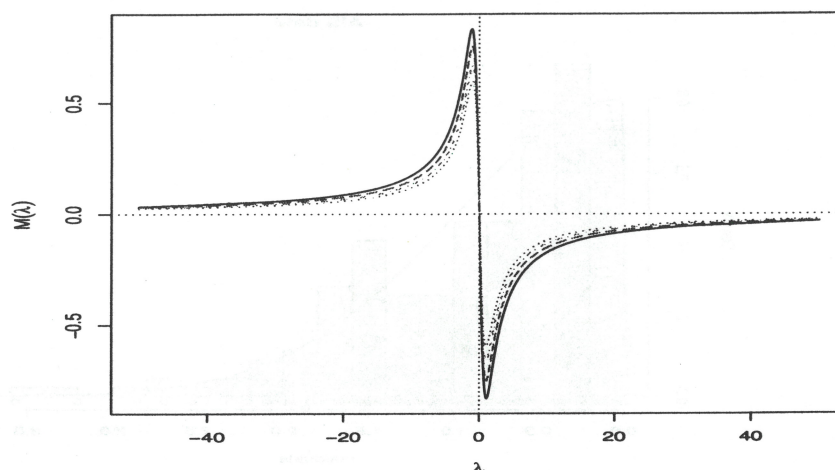


Figure 7: Modifications to the score function for the skew t distribution: $\nu = 3$ (dotted), 5 (dot-dashed), 10 (dashed). The solid line corresponds to the modification for the skew normal distribution.

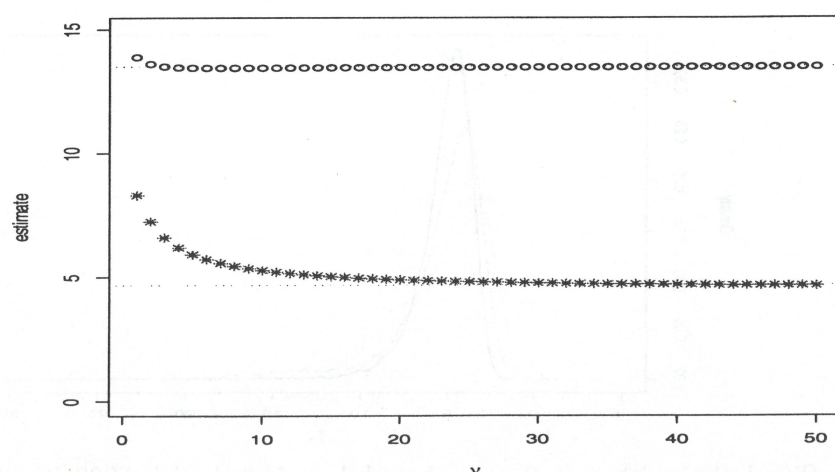


Figure 8: Example 4. Maximum likelihood estimate and its modification for the skew t distribution and for various values of ν : $\hat{\lambda}$ (o) and $\hat{\lambda}^*$ (*). The horizontal dotted lines correspond to the estimates for the skew normal distribution.