Statistical aspects in the extended skew-normal model

(Article begins on next page)

METRON - *International Journal of Statistics*
2011, vol. LXIX, n. 3, pp. 279-295

ANTONIO CANALE

Statistical aspects of the scalar extended skew-normal distribution

Summary - This paper presents some inferential results about the *extended skew-normal* family in the scalar case. For this family many inferential aspects are still unexplored. The expected information matrix is obtained and some of its properties are discussed. Some simulation experiments and an application to real data are presented pointing out not infrequent estimation problems such as different estimates in function of the starting values of the algorithm which leads to substantially equivalents densities. All these issues underline a problem of *near unidentifiability*.

Key Words - Extended skew-normal distribution; Information matrix; Skew-normal distribution.

1. Introduction

The skew-normal family (Azzalini, 1985) is a family of probability distributions which includes the normal one as a special case. It retains much of the mathematical tractability and some nice formal properties of the normal family. In addition in the last years a strong stream of statistical literature has led to several inferential results for this family of distributions both in the univariate and in the multivariate case. In the scalar case, if a random variable X is distributed as a skew-normal with location ξ, scale ω and shape α, written $X \sim \mathrm{SN}(\xi, \omega, \alpha)$, the density function of X is

$$f_X(x) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and the distribution function of a $N(0, 1)$ respectively, $\xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}$.

By generalizing this model adding a fourth parameter τ, one can obtain another family of probability distributions known as *extended skew-normal*. In the univariate case this family of distributions is already presented in the seminal

Received January 2009 and revised August 2011.

paper of Azzalini (1985) and it is studied subsequently in Arnold *et al.* (1993). Extensions to the multivariate context are studied in Arnold and Beaver (2000) and in Capitanio *et al.* (2003). A random variable Y is distributed as an extended skew-normal with position ξ, scale ω, shape α and truncation τ, written $Y \sim \text{ESN}(\xi, \omega, \alpha, \tau)$ if its density function is

$$f_Y(y) = \frac{1}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\alpha_0(\tau) + \alpha\frac{y-\xi}{\omega}\right)/\Phi(\tau), \tag{1}$$

where $\alpha_0(\tau) = \tau\sqrt{\alpha^2+1}$. The reason of the name for the fourth parameter will be explained later. If $\tau = 0$, we are back to the $\text{SN}(\xi, \omega, \alpha)$ distribution.

In the next section, some properties of this distribution will be discussed. In Section 3, the expression of the expected information matrix is presented and some of its characteristics and similarities with those of the skew-normal and of the normal are discussed. In Section 4, some of the characteristics of the likelihood function are discussed, and the maximum likelihood estimates for some simulated samples are obtained showing some numerical difficulties. Finally, in Section 5, an application to the well known AIS dataset is discussed.

2. Properties of the density function

Variations of the first two parameters lead to position and scale changes while the increase in absolute value of the α parameter increases the skewness of the density. Opposite values of α lead to density function mirrored on the x axis as in Property C of the work of Azzalini (19685). As τ increases on the positive semi-axis, the skewness reduces and the density function tends to be normal while as τ increases in absolute value, but on the negative semi-axis, the density function still seems to lose its skewness showing a shift (to positive values if α is positive, to negative values otherwise) proportional to the value of τ. This kind of behaviour can be seen in Figure 1.

The reason of this and of the name *truncation* for the parameter τ rises from one of the possible generation mechanisms of the density already presented in the work of Arnold *et al.* (1993) as *hidden truncation model*. Consider the random vector $(V, Y)^\top$ distributed as a bivariate normal distribution with zero mean vector and variance matrix with unit diagonal and correlation δ. Said $\tilde{f}_{V,Y}(v, y)$ its density function, let us define

$$f_{V,Y}(v, y) = \begin{cases} \dfrac{\tilde{f}_{V,Y}(v, y)}{1-\Phi(-\tau)} & v \geq -\tau \\ 0 & \text{otherwise} \end{cases}$$

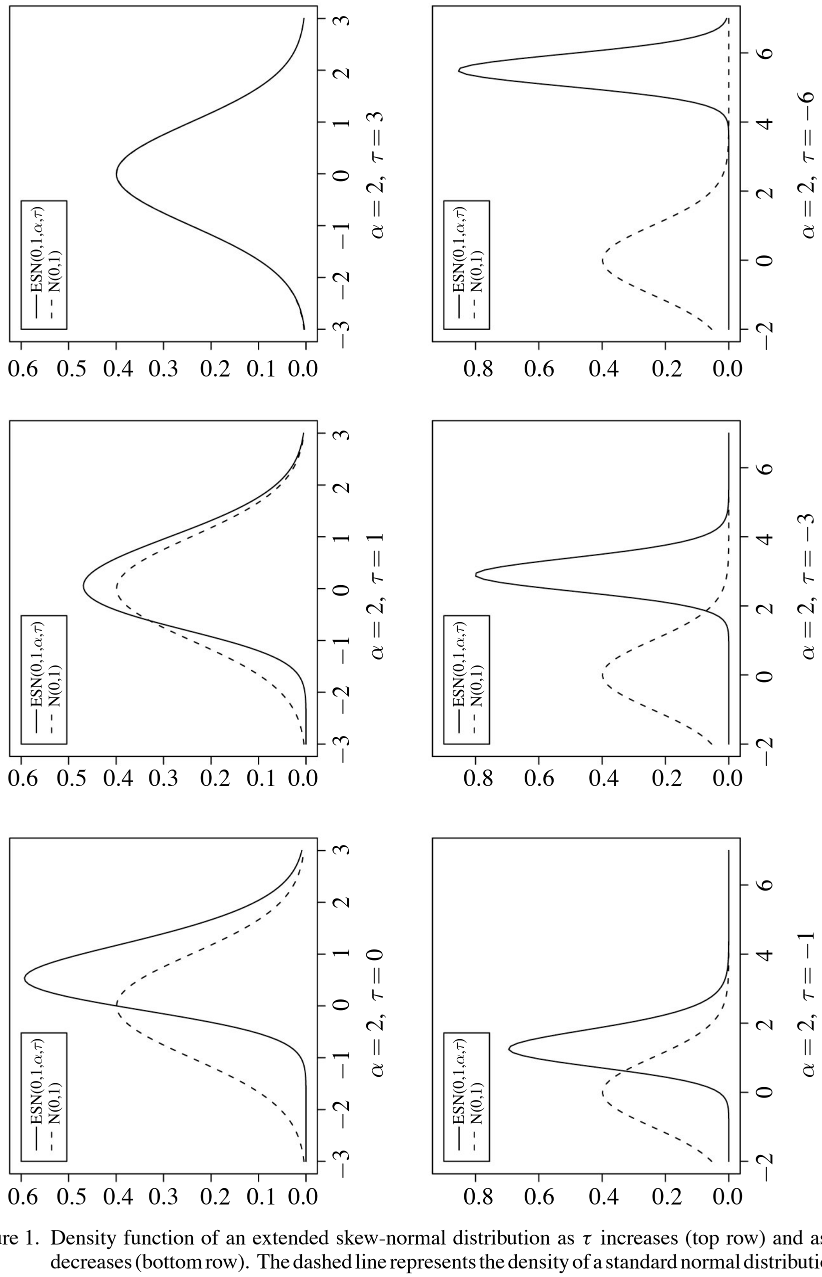


Figure 1. Density function of an extended skew-normal distribution as τ increases (top row) and as τ decreases (bottom row). The dashed line represents the density of a standard normal distribution.

the density function of the vector $(V, Y)^\top$ with V truncated below $-\tau$. Marginalizing out with respect to the truncated component, we obtain $Y \sim \text{ESN}(0,1,\alpha,\tau)$, where $\alpha = \delta/\sqrt{1-\delta^2}$. This argument explains the behaviour of the density for changes in τ. In fact if $\tau \to +\infty$, the truncation tends to $-\infty$ and hence it is negligible, with the marginal of Y being normal. Without loss of generality, let $\xi = 0$ and $\omega = 1$ to simplify notation, then

$$\lim_{\tau \to +\infty} \frac{1}{\Phi(\tau)} \phi\,(y)\,\Phi\,(\alpha_0(\tau) + \alpha\,(y)) = \phi(y). \tag{2}$$

On the other side, if $\tau \to -\infty$, the truncation is very strong ameaning that we are observing the tail of the marginal.

Analytically, using the de l'Hospital theorem, we obtain

$$\lim_{\tau \to -\infty} \frac{1}{\Phi(\tau)} \phi(y)\Phi(\alpha_0(\tau) + \alpha y) = \lim_{\tau \to -\infty} \frac{1}{\sqrt{1-\delta^2}} \phi\left(\frac{y + \delta\tau}{\sqrt{1-\delta^2}}\right), \tag{3}$$

where

$$\delta = \frac{\alpha}{\sqrt{\alpha^2 + 1}}. \tag{4}$$

As $\tau \to -\infty$ the density tends to be $N(-\delta\tau, 1 - \delta^2)$ and hence the limit distribution when $|\tau| \to \infty$ is normal. Practically we can achieve very non-skewed densities also for finite values of τ.

Define the $\zeta_0(\cdot)$ function and its derivatives as

$$\begin{aligned} \zeta_0(x) &= \log(\Phi(x)), \\ \zeta_m(x) &= \frac{\mathrm{d}^m}{\mathrm{d}x^m}\zeta_0(x) \quad (m = 1, 2, \dots) \end{aligned} \tag{5}$$

and consider the third standardized cumulant, given by

$$\gamma_1 = \frac{\delta^3 \zeta_3(\tau)}{\left[1 + \delta^2 \zeta_2(\tau)\right]^{3/2}}.$$

The third standardized cumulant was already introduced in Capitanio *et al.* (2003) for the multivariate extended skew-normal. In the multivariate case it coincides with the index of multivariate skewness of Mardia (1970). Fixed τ the γ_1 index is monotone with respect to α, as in the skew-normal distribution, while it is not the case if we let τ vary and we fixed α. A plot of the γ_1 index in function of τ for some α is given in Figure 2. The skewness is close to zero for $\tau > 3$ ($\gamma_1 < 0.036$ for $\alpha = 15$ and $\tau = 3$) and it reduces also for decreasing τ. The value of τ for which γ_1 is approximately zero heavily depends on α. The maximum value of γ_1 is about 1.995.

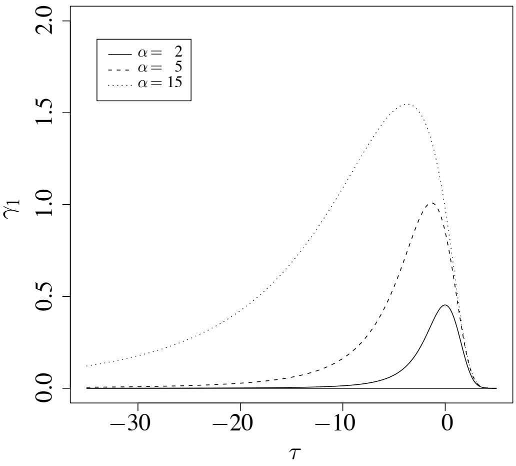


Figure 2. Skewness index γ_1 in function of τ for some α.

Some peculiar shapes of the density function can be obtained for some combinations of the parameters α and τ. In particular there are cases in which the shape of the density shows an evident sharp drop on one tail as those of Figure 3. This behaviour can be seen for $|\alpha| > 7$ and τ greater than 0 and less than 3. Clearly these indications are merely qualitative, because there is not a pointwise threshold for which the sharp drop shape is evident or not. We can add that the greater is the absolute value of α, the more evident is the sharp drop of the tail.

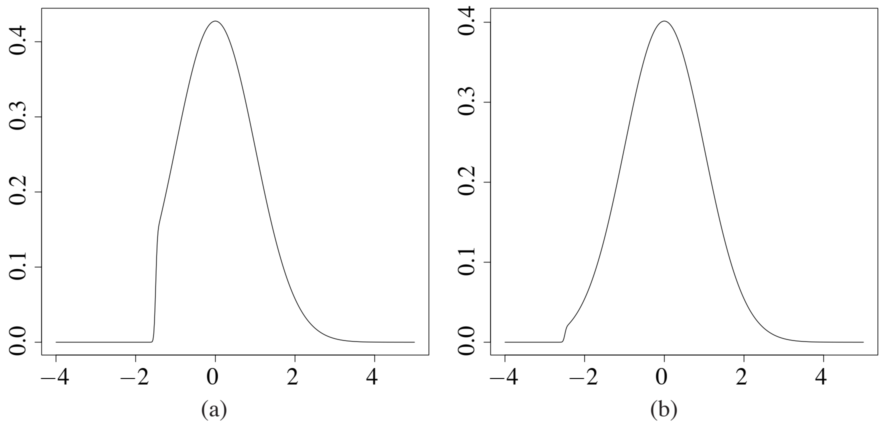


Figure 3. Sharp drop on the shape of the density function of an extended skew-normal. The densities are ESN(0, 1, 30, 1.5) (a) and ESN(0, 1, 30, 2.5) (b).

We want to conclude this section pointing out that there is not an analytic expression for the cumulative distribution function of the scalar extended skew-normal. However an expression for $P(Y > 0)$ is obtained in the following lemma.

Lemma 1. *Let* $Y \sim \mathrm{ESN}(0, 1, \alpha, \tau)$ *and* $T(h, a)$ *the function studied by Owen* (1956). *Then*

$$P(Y > 0) = \frac{1}{2} + \frac{T(\tau, \alpha)}{\Phi(\tau)}. \tag{6}$$

Proof. Let us obtain first the result $P(Y < 0)$.

$$\begin{aligned} P(Y < 0) &= \int_{-\infty}^{0} \Phi(\tau)^{-1}\phi(y)\Phi(\tau\sqrt{\alpha^2+1} + \alpha y)\,\mathrm{d}y \\ &= \Phi(\tau)^{-1}\int_{-\infty}^{0}\int_{-\tau}^{+\infty} \frac{1}{2\pi\sqrt{1-\delta^2}} \exp\left\{-\frac{(t^2 - 2\delta y t + y^2)}{2(1-\delta^2)}\right\} \mathrm{d}t\,\mathrm{d}y, \end{aligned}$$

using well known results about the transformation of correlated bivariate normal variates to uncorrelated ones, we can arrive at

$$\begin{aligned} P(Y < 0) &= \Phi(\tau)^{-1}\int_{-\tau}^{+\infty}\int_{-\infty}^{-\alpha u} \phi(u)\phi(v)\,\mathrm{d}v\,\mathrm{d}u, \\ &= \frac{1}{2} - \frac{T(\tau, \alpha)}{\Phi(\tau)} \end{aligned}$$

that gives equation (6). □

In Gupta and Pillai (1965) an equivalent expression for Lemma 1 is given. The results in their Lemma 2 gives an expression for the integral on the positive real line of the normal cdf of a linear combination of a normal random variable multiplied for its density. With suitable notation changes and normalizing constant this can be seen as the probability of obtaining a positive draw from an extended skew normal distribution.

3. Likelihood function and expected information matrix

3.1. *The expected information matrix*

Let $Y \sim \mathrm{ESN}(\xi, \omega, \alpha, \tau)$. Denoting with $\theta = (\xi, \omega, \alpha, \tau)$ the vector of the parameters, given an observation y of Y, the log-likelihood function is

$$\ell(\theta) = -\zeta_0(\tau) - \frac{1}{2}\log(\omega^2) - \frac{1}{2}\left(\frac{y-\xi}{\omega}\right)^2 + \zeta_0\left(\alpha_0 + \alpha\frac{y-\xi}{\omega}\right). \tag{7}$$

Let $z = (y - \xi)/\omega$, the score function is

$$\frac{\partial \ell}{\partial \theta} = \begin{pmatrix} \dfrac{1}{\omega}\left[z - \alpha\zeta_1(\alpha_0 + \alpha z)\right] \\ \dfrac{1}{\omega}\left[z^2 - \alpha z \zeta_1(\alpha_0 + \alpha z) - 1\right] \\ \zeta_1(\alpha_0 + \alpha z)\left(\dfrac{\tau\alpha}{\sqrt{\alpha^2+1}} + z\right) \\ \zeta_1(\alpha_0 + \alpha z)\sqrt{\alpha^2+1} - \zeta_1(\tau) \end{pmatrix},$$

while the second partial derivatives of the log-likelihood (7) are

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \xi^2} &= \frac{1}{\omega^2}\left(\alpha^2\zeta_2(\alpha_0 + \alpha z) - 1\right), \\
\frac{\partial^2 \ell}{\partial \xi \partial \omega} &= \frac{1}{\omega^2}\left(-2z + \alpha\zeta_1(\alpha_0 + \alpha z) + \alpha^2 z \zeta_2(\alpha_0 + \alpha z)\right), \\
\frac{\partial^2 \ell}{\partial \xi \partial \alpha} &= -\frac{1}{\omega}\left(\frac{\tau\alpha^2 + \alpha z\sqrt{1+\alpha^2}}{\sqrt{\alpha^2+1}}\zeta_2(\alpha_0 + \alpha z) + \zeta_1(\alpha_0 + \alpha z)\right), \\
\frac{\partial^2 \ell}{\partial \xi \partial \tau} &= -\frac{\alpha\sqrt{\alpha^2+1}}{\omega}\zeta_2(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \omega^2} &= \frac{1}{\omega^2}\left(-3z^2 + 2\alpha z \zeta_1(\alpha_0 + \alpha z) + (\alpha z)^2\zeta_2(\alpha_0 + \alpha z) + 1\right), \\
\frac{\partial^2 \ell}{\partial \omega \partial \alpha} &= -\frac{z}{\omega}\left(\left(\frac{\tau\alpha^2}{\sqrt{\alpha^2+1}} + \alpha z\right)\zeta_2(\alpha_0 + \alpha z) + \zeta_1(\alpha_0 + \alpha z)\right), \\
\frac{\partial^2 \ell}{\partial \omega \partial \tau} &= -\frac{\alpha z\sqrt{\alpha^2+1}}{\omega}\zeta_2(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \alpha^2} &= \left(\frac{\tau\alpha}{\sqrt{\alpha^2+1}} + z\right)^2\zeta_2(\alpha_0 + \alpha z) + \frac{\tau}{(\alpha^2+1)^{3/2}}\zeta_1(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \alpha \partial \tau} &= (\alpha\tau + z\sqrt{\alpha^2+1})\zeta_2(\alpha_0 + \alpha z) + \frac{\alpha}{\sqrt{\alpha^2+1}}\zeta_1(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \tau^2} &= (\alpha^2+1)\zeta_2(\alpha_0 + \alpha z) - \zeta_2(\tau).
\end{aligned}$$

To compute the elements of the expected information matrix, it is useful to introduce the following lemma.

Lemma 2. *Let* $Z \sim \mathrm{ESN}(0, 1, \alpha, \tau)$*,* $\zeta_1(x)$ *the first derivative of the function defined in expression* (5)*,* $\alpha_0 = \tau\sqrt{\alpha^2+1}$ *and* $g(x) = \alpha\tau + \sqrt{\alpha^2+1}x$*, then*:

$$\mathbb{E}[f(Z)\zeta_1(\alpha_0 + \alpha Z)] = \frac{\zeta_1(\tau)}{\sqrt{\alpha^2+1}}\,\mathbb{E}[f(g^{-1}(U))]$$

where $U \sim N(0,1)$ and $f(\cdot)$ is a function from $\mathbb{R}$ in $\mathbb{R}$ such that the involved integrals exist.

Proof.

$$\begin{aligned}
\mathbb{E}[f(Z)\zeta_1(\alpha_0+\alpha Z)] &= \int_{\mathbb{R}} f(z)\frac{\phi(\alpha_0+\alpha z)}{\Phi(\alpha_0+\alpha z)}\frac{1}{\Phi(\tau)}\phi(z)\Phi\left(\alpha_0+\alpha z\right)\mathrm{d}z \\
&= \frac{1}{\Phi(\tau)}\int_{\mathbb{R}} f(z)\phi(z)\phi(\alpha_0+\alpha z)\mathrm{d}z \\
&= \frac{1}{\Phi(\tau)}\int_{\mathbb{R}} f(z)\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(z^2+(\alpha_0+\alpha z)^2\right)\right\}\mathrm{d}z \\
&= \frac{1}{\Phi(\tau)}\int_{\mathbb{R}} f(z)\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(\alpha\tau+\sqrt{\alpha^2+1}z\right)^2\right\} \\
&\quad\times\exp\left\{-\frac{\tau^2}{2}\right\}\mathrm{d}z \\
&= \frac{\phi(\tau)}{\Phi(\tau)}\int_{\mathbb{R}} f(z)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\alpha\tau+\sqrt{\alpha^2+1}z\right)^2\right\}\mathrm{d}z \\
&= \frac{\zeta_1(\tau)}{\sqrt{\alpha^2+1}}\ \mathbb{E}[f(g^{-1}(U))].
\end{aligned}$$

□

Some useful results obtained using Lemma 2 are

$$\begin{aligned}
\mathbb{E}[\zeta_1(\alpha_0+\alpha Z)] &= \frac{\zeta_1(\tau)}{\sqrt{\alpha^2+1}}, \\
\mathbb{E}[Z\zeta_1(\alpha_0+\alpha Z)] &= -\alpha\tau\frac{\zeta_1(\tau)}{\alpha^2+1}, \\
\mathbb{E}[Z^2\zeta_1(\alpha_0+\alpha Z)] &= (1+\alpha^2\tau^2)\frac{\zeta_1(\tau)}{(\alpha^2+1)^{3/2}}, \\
\mathbb{E}[Z^3\zeta_1(\alpha_0+\alpha Z)] &= -(3\alpha\tau+\alpha^3\tau^3)\frac{\zeta_1(\tau)}{(\alpha^2+1)^2}, \\
\mathbb{E}[(\alpha_0+\alpha Z)\zeta_1(\alpha_0+\alpha Z)] &= \frac{\tau\zeta_1(\tau)}{\alpha^2+1}, \\
\mathbb{E}[\zeta_2(\alpha_0+\alpha Z)] &= -\frac{\tau\zeta_1(\tau)}{\alpha^2+1} - \mathbb{E}[\zeta_1(\alpha_0+\alpha Z)^2],
\end{aligned}$$

where the last expression is obtained exploits the formula

$$\zeta_2(x) = -\zeta_1(x)\left[x+\zeta_1(x)\right].$$

Lengthy algebra leads to the expressions for each of the elements I_{ij} of the expected information matrix. They are

$$
\begin{aligned}
I_{11} &= \frac{1}{\omega^2}\left(1+\delta^2\tau\zeta_1(\tau)+\alpha^2 a_0\right),\\
I_{12} &= \frac{1}{\omega^2}\left[\zeta_1(\tau)\delta\left(\frac{1+2\alpha^2-(\alpha^2+1)\delta^2\tau^2}{\alpha^2+1}\right)+\alpha^2 a_1\right],\\
I_{13} &= \frac{1}{\omega}\left(\frac{\zeta_1(\tau)}{(1+\alpha^2)^{3/2}}-\tau\alpha\delta a_0-\alpha a_1\right),\\
I_{14} &= -\frac{1}{\omega}\left(\delta\tau\zeta_1(\tau)+\alpha\sqrt{\alpha^2+1}a_0\right),\\
I_{22} &= \frac{1}{\omega^2}\left(2+\alpha^2 a_2+\delta^4\tau(\tau^2-3)\zeta_1(\tau)\right),\\
I_{23} &= \frac{1}{\omega}\left(\frac{-2\alpha\tau\zeta_1(\tau)}{(1+\alpha^2)^2}-\tau\alpha\delta a_1-\alpha a_2\right),\\
I_{24} &= \frac{\alpha\sqrt{\alpha^2+1}}{\omega}\left(\frac{\alpha\zeta_1(\tau)(\tau^2-1)}{(1+\alpha^2)^{3/2}}-a_1\right),\\
I_{33} &= (\tau\delta)^2 a_0+2\tau\delta a_1+a_2,\\
I_{34} &= \alpha\tau a_0+\sqrt{\alpha^2+1}\ a_1,\\
I_{44} &= (\alpha^2+1)a_0-\zeta_1(\tau)^2,
\end{aligned}
$$

where δ is given by expression (4) and the quantity a_k, with $k = 0, 1, 2$, are redefined similarly to the analogous for the skew-normal distribution as

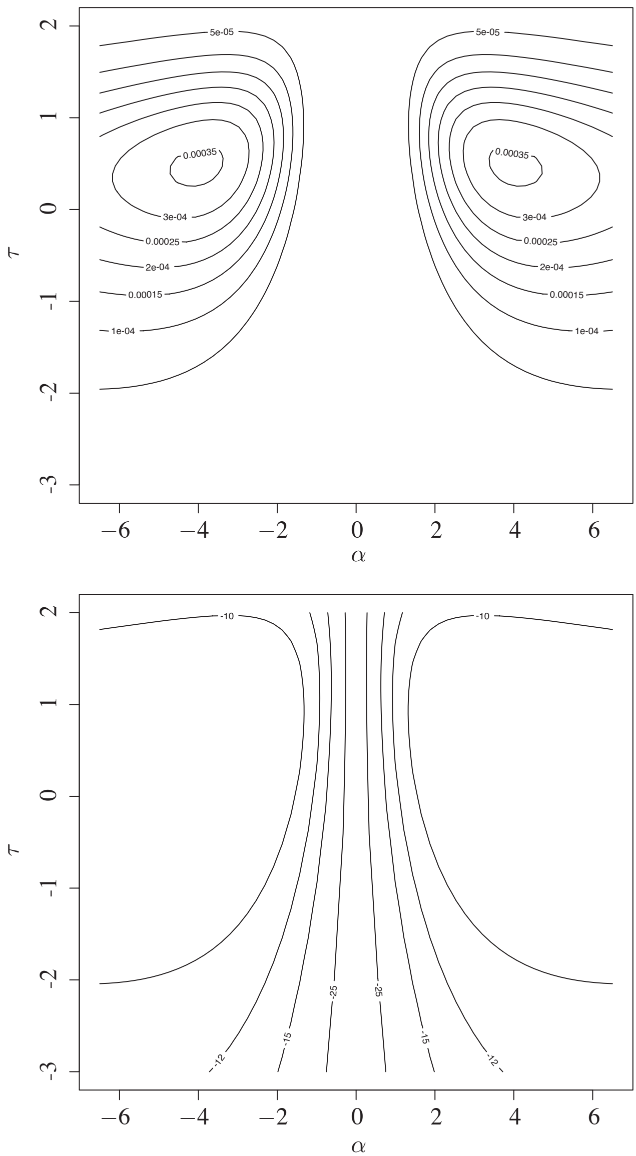
$$a_k(\alpha,\tau) = \mathbb{E}\left[Z^k\left(\zeta_1(\tau\sqrt{\alpha^2+1}+\alpha Z)\right)^2\right]. \tag{8}$$

The a_ks depend on both parameters α and τ, and need to be evaluated numerically.

3.2. Some properties of the expected information matrix

It is easy to check that the information matrix in Subsection 3.1 when $\tau = 0$ is exactly that of Azzalini (1985) for the parameter (ξ, ω, α). Moreover the analogies with the skew-normal information matrix do not refer only to the case $\tau = 0$.

Another common characteristic is that the information matrix tends to be singular as $\alpha \to 0$ for all values of τ. Moreover the information matrix tends to be singular also for $\tau \to \pm\infty$ with the determinant being very small even for finite values of τ (*e.g.* $\tau > 2$ or $\tau < -3$). In Figure 4 the contour level



(a)fl

(b)

Figure 4. Contour levels of the value of the determinant of the expected information matrix (a) and its log transformation (b) in function of (α, τ).

plot of the determinant of the expected information matrix is reported as a function of α and τ.

Another peculiar feature of the expected information matrix is that the elements of the upper left 2×2 block of the matrix are, for $\tau \to \pm\infty$, exactly those of the expected information matrix of a normal distribution.

4. Maximum likelihood estimation

4.1. Behaviour of the likelihood function

As for the skew-normal distribution, the profile log-likelihood for the α parameter of the extended skew-normal distribution presents a stationary point in $\alpha = 0$. This characteristic is clearly visible in all samples and for any sample size as illustrated in Figure 5, where the relative profile log-likelihood

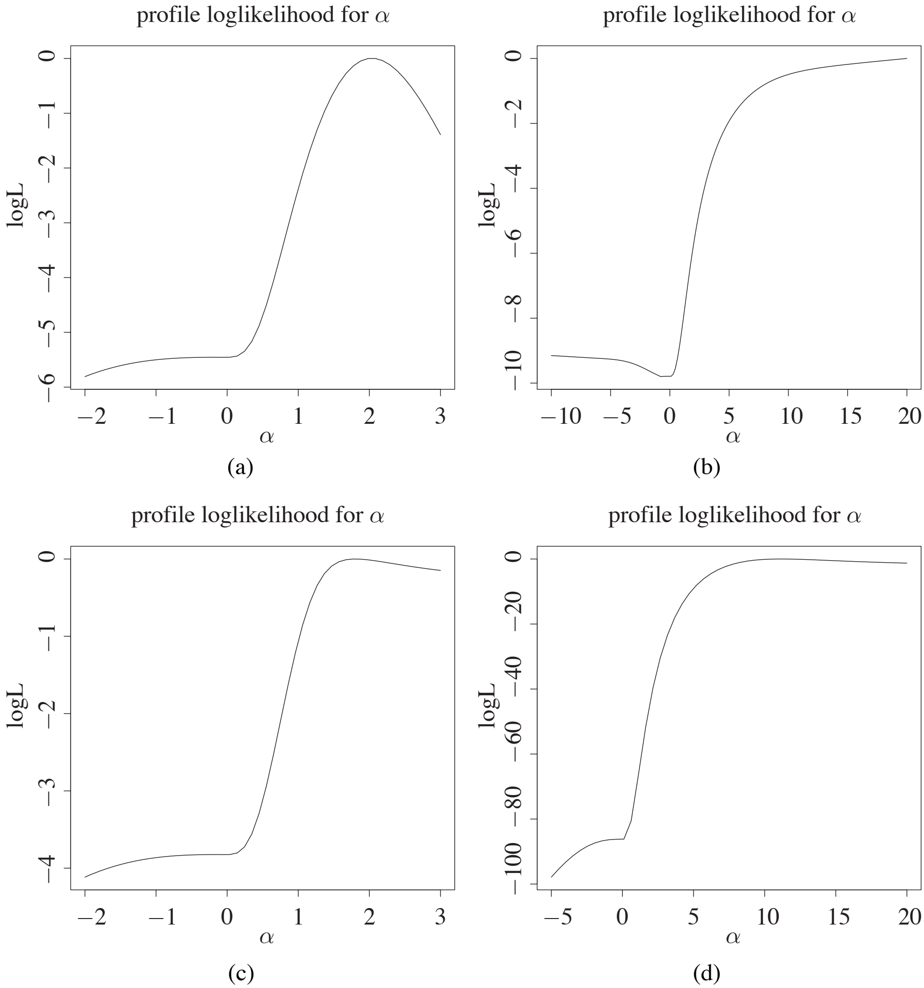


Figure 5. Relative profile log-likelihood for α in simulated samples with $n = 500$ from ESN(0, 1, 2, 1) (a), ESN(0, 1, 15, 2) (b), ESN(0, 1, 2, −1) (c), ESN(0, 1, 15, −2) (d).

for four samples of size 500 is plotted. The presence of a stationary point at $\alpha = 0$ can be easily proved solving the likelihood equations

$$\frac{\partial \ell}{\partial \theta} = 0,$$

because for $\alpha = 0$, the fourth likelihood equation is

$$\frac{\partial \ell}{\partial \tau} = \zeta_1(\tau) - \zeta_1(\tau) = 0$$

that is verified for all τ.

Let us analyze now the profile log-likelihood for the fourth parameter τ. We calculated the profile log-likelihood only for the range $-6 < \tau < 3$ because values outside this range can lead to numerical instability of the optimization routines. In addition $\tau > 3$ leads the skewness of the distribution to become tiny (*i.e.* $\gamma_1 < 0.037$) and, on the other side, $\tau < -6$ leads to a location change of the curve showed in Section 2. The role of τ outside this range is confused with the one of α (regulate skewness) and of ξ (change location).

The plots in Figure 6 show that when the real value of τ is small in absolute value and positive (*i.e.* between 0 and 3), determining particular extended skew normal shapes of the density such as the sharp drop tail, the likelihood function is able to detect the right region of the parameter space, as in plots (a) and (b). On the other side, when the τ parameter is high and positive (*i.e.* > 3) determining moderate skewed and symmetrical densities or when it is negative but with a high values of α determining highly skewed densities without the particular sharp drop shape, the profile log-likelihood behaves bad. See plot (c) and (d) for moderate skewness ($\gamma_1 = 0.39$) and for non-skewness ($\gamma_1 = 0.002$) and plot (e) for high skewness ($\gamma_1 = 1.45$) in the densities. Moreover, from the values along the vertical axis of plots (c) and (d) it is clear that the profile log-likelihood is quite flat for a wide region of the parameter space, despite the graphical impression while in plot (e) the likelihood has its maximum around 0.5 with the real value of τ being in a lower likelihood value region.

4.2. Numerical evaluation of the maximum likelihood estimates

We tried to estimate the parameters of the distribution via numerical maximization of the likelihood function from some simulated samples. We used both quasi-Newton and simulating annealing algorithms with the full log-likelihood or the profile log-likelihood for α or τ as objective functions. The study shows that both the maximization algorithms hardly reach the global maximum and, changing the starting values, they converge to local maxima. This can be seen in Table 1-3, in which we report the results for the quasi-Newton max-

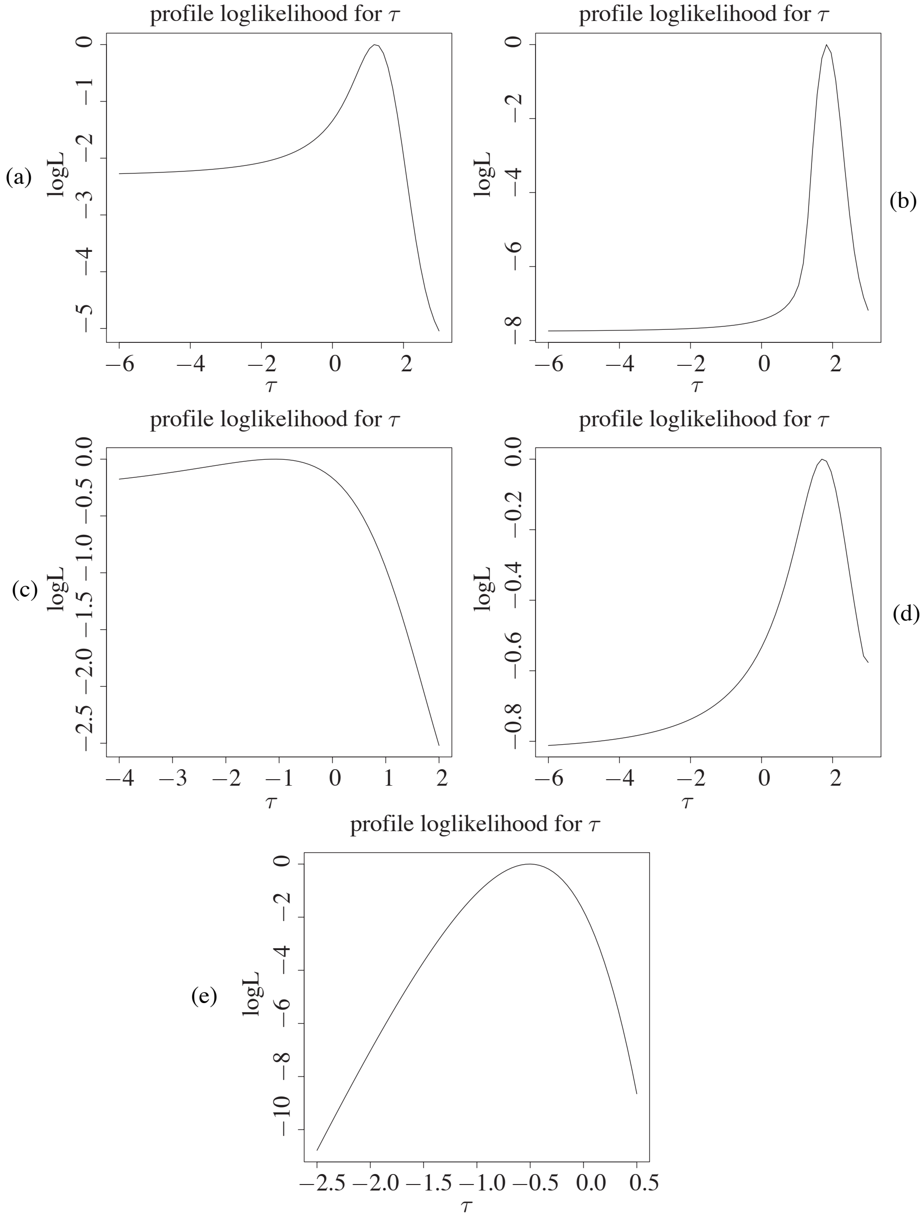


Figure 6. Relative profile log-likelihood for τ in simulated samples with $n = 500$ from (a) ESN(0, 1, 2, 1), (b) ESN(0, 1, 15, 2), (c) ESN(0, 1, 2, −1), (d) ESN(0, 1, 5, 4) and (e) ESN(0, 1, 15, −2).

imization of the log-likelihood from three simulated samples of size 500 from an ESN(0, 1, 15, 2), ESN(0, 1, 1, 0) and ESN(0, 1, 15, −2). These three choices represent a typical sharp drop shape of the extended skew-normal, a genuine skew-normal and a highly skewed distribution ($\gamma_1 = 1.45$) respectively.

Consider for example row 3 of Table 1 in which the maximization algorithm is initialized with the real value of the parameters. The algorithm does not move the final estimates from the starting values even though the obtained values are not the maximum likelihood estimates, in particular for the α parameter as we can see comparing line 3 with line 1, 2 or 5 of the same table which report similar parameter values with the likelihood function being equal up to the third decimal and higher than that of line 3. The estimated density functions are plotted in Figure 7 (a) where it is evident that they are essentially equivalent and nearly overlappable. Consider now Table 2. The estimates are very different among each other and also very far from the real values of the parameters. In line 5 for example we can find a very high estimate for the location and a negative, high in absolute value, estimate for τ. As shown in Section 2, indeed, high in absolute value and negative τ shifts the density. Here the likelihood adjusts the estimates of the other parameters leading a density equivalent to those obtained using the parameters of the other rows, plotted in Figure 7 (b). Rows 1 and 2 of Table 3 were initialized at the same way but for the τ parameter with the algorithm converging to different parameter estimates. The difference among the estimated densities reported in Figure 7 (c) is now more perceptible but always small.

TABLE 1: *Maximum likelihood estimation for some starting values and objective functions. (Sample of size 500 from* ESN(0, 1, 15, 2)).

	Obj. fun.	Initialization	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\tau}$	$\log L$
1	$\ell(\theta)$	moments ($\tau = 0$)	0.012	0.994	8.800	1.914	−205.244
2	$\ell(\theta)$	moments ($\tau = 2$)	0.012	0.995	8.800	1.914	−205.244
3	$\ell(\theta)$	(0,1,15,2)	0.022	0.987	15.000	1.995	−205.406
4	$\ell_{P(\alpha)}$	moments	0.007	0.995	5.000	1.870	−205.820
5	$\ell_{P(\tau)}$	moments	0.012	0.995	8.696	1.913	−205.244

TABLE 2: *Maximum likelihood estimation for some starting values and objective functions. (Sample of size 500 from* ESN(0, 1, 1, 0)).

	Obj. fun.	Initialization	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\tau}$	$\log L$
1	$\ell(\theta)$	moments ($\tau = 0$)	4.075	1.278	−1.414	−3.087	−129.727
2	$\ell(\theta)$	(0,1,1,0)	0.558	0.787	0.008	−0.304	−129.913
3	$\ell(\theta)$	(0,1,0,0)	0.558	0.787	0.010	−0.086	−129.913
4	$\ell_{P(\alpha)}$	moments	8.972	1.712	−2.088	−5.265	−129.712
5	$\ell_{P(\tau)}$	moments	23.743	2.604	−3.367	−9.180	−129.706

TABLE 3: *Maximum likelihood estimation for some starting values and objective functions. (Sample of size 500 from* ESN(0, 1, 15, −2)).

	Obj. fun.	Initialization	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\tau}$	$\log L$
1	$\ell(\theta)$	moments ($\tau = -1.5$)	−1.128	1.141	15.461	−2.758	409.544
2	$\ell(\theta)$	moments ($\tau = -2$)	0.788	0.806	11.164	−1.515	410.167
3	$\ell(\theta)$	(0,1,15,-2)	0.793	0.805	11.150	−1.511	410.167
4	$\ell_{P(\alpha)}$	moments	1.738	0.543	5.000	−0.522	401.141
5	$\ell_{P(\tau)}$	moments	0.750	0.814	11.304	−1.546	410.165

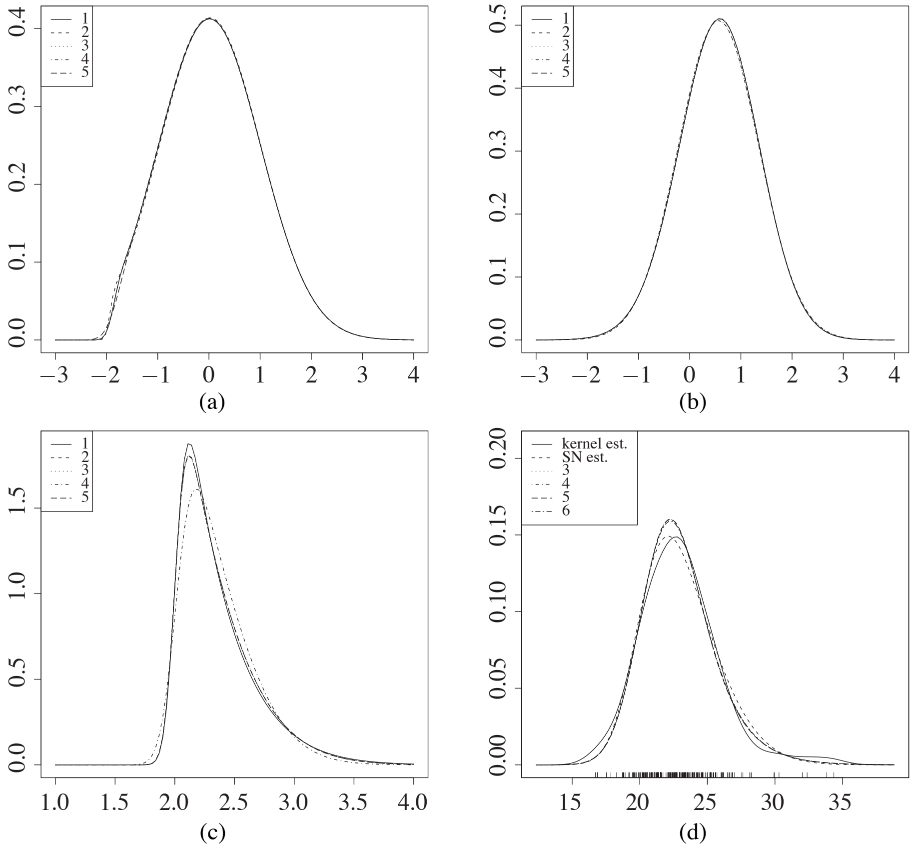


Figure 7. Density functions with the parameters of Table 1, Table 2, Table 3 and Table 4 obtained fitting the ESNto the AIS dataset.

5. APPLICATION TO REAL DATA

We fitted the extended skew-normal model to the well known AIS (Australian Institute of Sport) dataset. The dataset refers to a number of biomedical measurements on 202 Australian athletes, of which we used only the body mass index (bmi).

As in the simulated samples, different starting values of the maximization algorithm lead to different parameter estimates. Table 4 reports the maximum likelihood estimates assuming a skew-normal model (row 1) and the results of the maximization of the log-likelihood and profile log-likelihood for τ assuming an extended skew normal model (rows 2-5). The obtained estimates are extremely different from those obtained assuming a skew-normal model and among each other even if the correspond to an extremely similar value of the log-likelihood. Figure 7 (d) shows the non parametric density estimation using the kernel method and the six densities obtained with the parameters of Table 4. All the five vector of parameters obtained maximizing the extended skew-normal model yield to equivalent densities.

TABLE 4: *Maximum likelihood estimation for some starting values and objective functions. (AIS bmi data).*

	Obj. fun.	Initialization	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\tau}$	$\log L$
1	$\ell_{SN}(\theta)$	moments	19.97	4.13	2.31	-	-
2	$\ell(\theta)$	moments ($\tau = 0$)	-40.04	12.00	6.36	-5.13	-301.54
3	$\ell(\theta)$	moments ($\tau = -1$)	-47.05	12.57	6.67	-5.46	-301.51
4	$\ell(\theta)$	moments ($\tau = -5$)	-166.78	20.21	10.71	-9.32	-301.40
5	$\ell_{P(\tau)}$	moments	-15.64	9.57	5.00	-3.88	-301.67

6. DISCUSSION

An expression for the Fisher information matrix of the extended skew-normal in the scalar case has been obtained and studied. In particular we have seen that the matrix can be singular when τ grows in absolute value and for $\alpha = 0$.

In the simulation study and in the application on AIS data, some characteristic of this distribution have been analyzed stressing a problem of *near unidentifiability* of the model in the scalar case. Numerical algorithms, in fact, hardly reach the global maximum and different values of the parameters lead to almost equivalent densities.

ACKNOWLEDGMENTS

We would like to thank professor Adelchi Azzalini for the comments he kindly offered in early versions of the paper and the referees for the useful suggestions.

REFERENCES

ARNOLD, B. C. and BEAVER, R. J. (2000) Hidden truncation model, *Sankhyā, series A*, 62, 22–35.

ARNOLD, B. C., BEAVER, R. J., GROENEVELD, R. A. and MEEKER, W. Q. (1993) The non truncated marginal of a truncated bivariate normal distribution, *Psychometrika*, 58, 471–488.

AZZALINI, A. (1985) A class of distributions which includes the normal ones, *Scand. J. Statist.*, 12, 171–178.

CAPITANIO, A., AZZALINI, A. and STANGHELLINI, E. (2003) Graphical models for skew-normal variates, *Scand. J. Statist.*, 30, 129–144.

GUPTA, S. S. and PILLAI, K. C. S. (1965) On linear functions of ordered correlated normal random variables, *Biometrika*, 53, 367–379.

MARDIA, K. V. (1970) Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 57, 519–530.

OWEN, D. B. (1956) Tables for computing bivariate normal probabilities, *Ann. Math. Statist.*, 27, 1075–1090.

ANTONIO CANALE
Dip. Sc. Statistiche
Università di Padova,
via C. Battisti, 241
35121 Padova (Italia)
Tel.: +39-049-8274111
Fax: +39-049-8274170
canale@stat.unipd.it