SORT 37 (1) January-June 2013, 19-28

A note on the Fisher information matrix for the skew-generalized-normal model

Reinaldo B. Arellano-Valle[1], Héctor W. Gómez[2] and Hugo S. Salinas[3]

Abstract

In this paper, the exact form of the Fisher information matrix for the skew-generalized normal (SGN) distribution is determined. The existence of singularity problems of this matrix for the skew-normal and normal particular cases is investigated. Special attention is given to the asymptotic properties of the MLEs under the skew-normality hypothesis.

MSC: 62E20

Keywords: Asymptotic distribution, Kurtosis, maximum likelihood estimation, singular information matrix, Skewness.

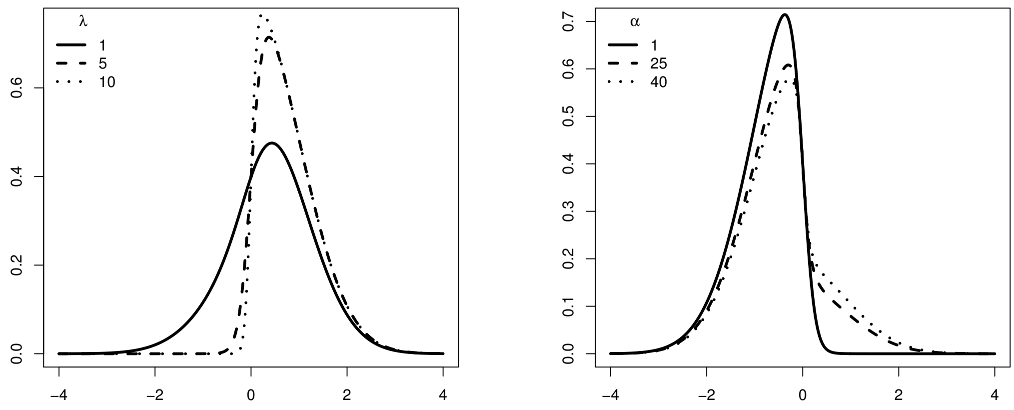
1. Introduction

Arellano-Valle, Gómez and Quintana (2004) introduced the skew-generalized-normal (SGN) distribution with density

$$f(z;\lambda,\alpha) = 2\phi(z)\Phi\left(\frac{\lambda z}{\sqrt{1+\alpha z^2}}\right),\ \ z\in\mathbb{R},\ \lambda\in\mathbb{R},\ \alpha\geq 0, \tag{1}$$

and denoted by $\mathrm{SGN}(\lambda,\alpha)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density function and cumulative distribution function of the standardized normal distribution, respectively. The skewness of the SGN distribution (1) is regulated by the parameters λ and α, so that it reduces

[1] Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Chile.
E-mail: reivalle@mat.puc.cl
[2] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile.
E-mail: hgomez@uantof.cl.
[3] Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Chile.
E-mail: hsalinas@mat.uda.cl.
Received: March 2011
Accepted: November 2011



(a) SGN$(\lambda, \alpha = 1)$ (b) SGN$(\lambda = -5, \alpha)$

Figure 1: *Examples of the skew-generalized normal density.*

to the skew-normal (SN) distribution when $\alpha = 0$ and to the normal (N) one when $\lambda = 0$. Note, however, that the value of α is irrelevant when $\lambda = 0$. The same occurs with λ for the limiting case when $\alpha \to \infty$. In both of these situations, the normality is attained from the SGN model, producing there a local identifiability problem. Further special models can also be obtained by reparametrizing α in terms of λ or viceversa. For example, by making $\alpha = \lambda^2$ we obtain in (1) the so-called skew-curved normal (SCN) distribution in Arellano-Valle *et al.* (2004). This flexibility of the SGN distribution allows to incorporate a wide range of models in a neighbourhood of the normal distribution. Figure 1 shows the behaviour of the SGN density for different values of the parameters λ and α. Only positive values of λ are considered in the plots of Figure 1(a); when the sign of λ is reversed, the density is reflected about the origin, as in Figure 1(b).

Further properties of the SGN model are investigated by Arellano-Valle *et al.* (2004). In particular, they gave formulas for the moments of a SGN random variable, $Z \sim$ SGN(λ, α). They showed that the even moments of Z are equal to the corresponding even moments of a standardized normal random variable. For the odd moments of Z, they obtained expressions involving an implicit formula,

$$E(Z^{2k+1}) = 2c_k - 2^k\Gamma(1+k)(2/\pi)^{1/2}, \quad k = 0, 1, 2, \ldots,$$

where $c_k := c_k(\lambda, \alpha) = \int_0^\infty u^k \phi(\sqrt{u})\Phi\left(\frac{\lambda\sqrt{u}}{\sqrt{1+\alpha u}}\right) du$. The location-scale extension of the SGN distribution (1) was also obtained by Arellano-Valle *et al.* (2004) by letting $X = \mu + \sigma Z$, where $Z \sim$ SGN(λ, α), and where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters, respectively. In this case, the notation $X \sim$ SGN$(\mu, \sigma, \lambda, \alpha)$ will be used. Finally, for the mean and variance of $X \sim$ SGN$(\mu, \sigma, \lambda, \alpha)$ we can note that $E(X) = \mu + \sigma\mu_z$ and $Var(X) = (1 - \mu_z^2)\sigma^2$, where $\mu_z = 2c_0 - (2/\pi)^{1/2}$ is the mean of Z.

The Fisher information matrix has an important role in statistical analysis (classical and Bayesian) as well as in information theory. In the location-scale skew-normal distribution, however, the Fisher information matrix is singular (Azzalini, 1985) when the skewness/shape parameter is zero, i.e., under the normality hypothesis. This fact violates the standard regularity conditions leading to the asymptotic normal distribution of the MLEs. A situation of this type falls under nonstandard asymptotic theory studied by Rotnitzky et al. (2000), who showed that in these circumstances the rate of convergence estimate is slower than the usual one. Motivated by this fact, we consider it important to obtain and analyse the behaviour of the Fisher information matrix in an generalization of the skew-normal distribution.

In this note, we determine the exact form of the Fisher information matrix for the skew-generalized-normal (SGN) distribution. Next, we examine the existence of singularity problems of this matrix for the skew-normal and normal special cases, giving a special attention to the asymptotic properties of the MLEs under the skew-normality hypothesis ($\lambda = 0$).

This paper is organized as follows. The elements of the expected information matrix for the full location-scale SGN model are derived in Section 2. Solutions for the singularity problems in the full information matrix for the normal particular cases are also discussed there. The technical details are given in an Appendix.

2. Maximum likelihood estimation

This section is related to the asymptotic properties of the MLEs of the location-scale SGN model. Specifically, the ingredients to compute the expected information matrix for the full location-scale SGN model are given. Hence, the study is focused on the asymptotic behaviour of the MLEs for the particular skew-normal and normal models.

2.1. Likelihood score functions

Let $X_1,\ldots,X_n$ be a random sample drawn from the SGN$(\mu,\sigma,\lambda,\alpha)$ distribution. The log-likelihood function for $\theta = (\mu,\sigma,\lambda,\alpha)^\top$ is $\sum_{i=1}^n l(\theta,X_i)$, where $l(\theta,X)$ is the log-likelihood for θ based on a single observation X, that is,

$$l(\theta,X) := \log f(X;\theta) = \frac{1}{2}\log\left(\frac{2}{\pi}\right) - \log(\sigma) - \frac{Z^2}{2} + \log\Phi(W), \qquad (2)$$

where $Z = (X-\mu)/\sigma$ and $W = W(Z) = \lambda Z/(1+\alpha Z^2)^{1/2}$. The score function is $\sum_{i=1}^n S_\theta(\theta,X_i)$, where $S_\theta(\theta,X) = \partial l(\theta,X)/\partial\theta$ is the vector $(S_\mu,S_\sigma,S_\lambda,S_\alpha)^\top$ with elements

$$S_\mu = \frac{Z}{\sigma} - \frac{1}{\sigma}\frac{\phi(W)}{\Phi(W)}\frac{\partial W}{\partial Z}, \quad S_\sigma = -\frac{1}{\sigma} + \frac{Z^2}{\sigma} - \frac{1}{\sigma}\frac{\phi(W)}{\Phi(W)}\frac{\partial W}{\partial Z}Z,$$

$$S_\lambda = \frac{\phi(W)}{\Phi(W)}\frac{\partial W}{\partial \lambda} \quad \text{and} \quad S_\alpha = \frac{\phi(W)}{\Phi(W)}\frac{\partial W}{\partial \alpha},$$

where $\partial W/\partial Z = \lambda/(1+\alpha Z^2)^{3/2}$, $\partial W/\partial \lambda = Z/(1+\alpha Z^2)^{1/2}$ and $\partial W/\partial \alpha = -\lambda Z^3/2(1+\alpha Z^2)^{3/2}$.

2.2. Fisher information matrix

By definition, the SGN-expected information matrix for θ can be computed as $\mathbf{I}_\theta = E[\mathbf{S}_\theta \mathbf{S}_\theta^\top]$, where $\mathbf{S}_\theta$ is the SGN-score vector above. Thus, the elements $I_{\theta_i\theta_j} = E[\mathbf{S}_{\theta_i}\mathbf{S}_{\theta_j}^\top]$ of this matrix are shown in the Appendix to be

$$I_{\mu\mu} = \frac{1}{\sigma^2} + \frac{\lambda^2}{\sigma^2}\eta_{03}, \quad I_{\mu\sigma} = \frac{2}{\sigma^2}(c_1 - c_0) - \frac{(2/\pi)^{1/2}}{\sigma^2} - \frac{2\lambda}{\sigma^2}\rho_{23} + \frac{\lambda}{\sigma^2}\rho_{03} + \frac{\lambda^2}{\sigma^2}\eta_{13},$$

$$I_{\mu\lambda} = \frac{1}{\sigma}\rho_{21} - \frac{\lambda}{\sigma}\eta_{12}, \quad I_{\mu\alpha} = -\frac{\lambda}{2\sigma}\rho_{43} + \frac{\lambda^2}{2\sigma}\eta_{33}, \quad I_{\sigma\sigma} = \frac{2}{\sigma^2} + \frac{\lambda^2}{\sigma^2}\eta_{23},$$

$$I_{\sigma\lambda} = -\frac{\lambda}{\sigma}\eta_{22}, \quad I_{\sigma\alpha} = \frac{\lambda^2}{2\sigma}\eta_{43}, \quad I_{\lambda\lambda} = \eta_{21}, \quad I_{\lambda\alpha} = -\frac{\lambda}{2}\eta_{42} \text{ and } I_{\alpha\alpha} = \frac{\lambda^2}{4}\eta_{63},$$

where the coefficients ρ_{nm} and η_{nm} are defined in Proposition 1 given in the Appendix. These coefficients must be computed numerically.

For the nonnormal cases with $\lambda \neq 0$ and $0 \leq \alpha < \infty$, the above information matrix is always nonsingular, so that the usual $\sqrt{n}$-asymptotic behaviour holds for the MLEs. In particular, estimation of the standard errors of the parameter estimates can be taken from the diagonal elements of the inverse Fisher information matrix. Moreover, the submatrix of the full information matrix corresponding to the vector of parameters $(\mu, \sigma, \lambda)^\top$ coincides with the SN-information matrix obtained by Azzalini (1985). In addition, for the skew-normal special case with $\alpha = 0$, the full associated information matrix is also nonsingular. See Section 2.3 below.

For the normal case that follows when $\lambda = 0$, the information matrix of $\theta = (\mu, \sigma, \lambda, \alpha)^\top$ is

$$\begin{pmatrix} \frac{1}{\sigma^2} & 0 & \frac{2(2/\pi)^{1/2}}{\sigma} d_1(\alpha) & 0 \\ & \frac{2}{\sigma^2} & 0 & 0 \\ & & \frac{2}{\pi} d_2(\alpha) & 0 \\ & & & 0 \end{pmatrix},$$

where $d_1(\alpha) = \int_0^\infty \frac{z^2\phi(z)}{(1+\alpha z^2)^{1/2}} dz$ and $d_2(\alpha) = \frac{1}{\alpha}\left(1 - (2\pi/\alpha)^{1/2} e^{\frac{1}{2\alpha}} \Phi(-\alpha^{-1/2})\right)$ (see Corollary 1 in the Appendix). Although the first three columns of this matrix are linearly independent, it leads to a singular information matrix because of a final column (corresponding to the parameter α) of 0s. This fact is obvious from (1), since α is non-identifiable when $\lambda = 0$. Properties of the MLEs when the SGN model reduces to the normal case are considered in Section 2.4.

2.3. Properties of the MLEs in the skew-normal case

Suppose that the parameter vector is $\theta^* = (\mu^*, \sigma^*, \lambda^*, 0)^\top$, that is, the data are drawn from the SN$(\mu^*, \sigma^*, \lambda^*)$ distribution. At $\theta = \theta^*$, the components of the score vector S_θ are

$$S^*_\mu = \frac{1}{\sigma^*}\left[Z^* - \lambda^* \frac{\phi(\lambda^* Z^*)}{\Phi(\lambda^* Z^*)}\right], S^*_\sigma = \frac{1}{\sigma^*}\left[Z^{*2} - 1 - \lambda^* \frac{\phi(\lambda^* Z^*)}{\Phi(\lambda^* Z^*)} Z^*\right],$$
$$S^*_\lambda = \frac{\phi(\lambda^* Z^*)}{\Phi(\lambda^* Z^*)} Z^* \quad \text{and} \quad S^*_\alpha = -\frac{\lambda^*}{2} \frac{\phi(\lambda^* Z^*)}{\Phi(\lambda^* Z^*)} Z^{*^3},$$

where $Z^* = (X - \mu^*)/\sigma^*$. Linear dependence does not exist between the elements of the score function when $\lambda^* \neq 0$. Consequently, the information matrix is not singular in this case. In the full parameter case, the vector $n^{1/2}(\hat{\mu} - \mu^*, \hat{\sigma} - \sigma^*, \hat{\lambda} - \lambda^*, \hat{\alpha})$, where $(\hat{\mu}, \hat{\sigma}, \hat{\lambda}, \hat{\alpha})$ is the MLE of $(\mu, \sigma, \lambda, \alpha)$, converges in distribution to (Y_1, Y_2, Y_3, Y_4), where $(Y_1, Y_2, Y_3, Y_4)^\top$ is a multivariate normal random vector with mean vector $(0,0,0,0)^\top$ and covariance matrix

$$\begin{pmatrix} \frac{1}{\sigma^2}(1+\lambda^2 a_0) & \frac{1}{\sigma^2}\left(\frac{\lambda(2/\pi)^{1/2}(1+2\lambda^2)}{(1+\lambda^2)^{3/2}} + \lambda^2 a_1\right) & \frac{1}{\sigma}\left(\frac{(2/\pi)^{1/2}}{(1+\lambda^2)^{3/2}} - \lambda a_1\right) & \frac{1}{2\sigma}\left(-\frac{3\lambda(2/\pi)^{1/2}}{(1+\lambda^2)^{5/2}} + \lambda^2 a_3\right) \\ & \frac{1}{\sigma^2}(2+\lambda^2 a_2) & -\frac{\lambda}{\sigma} a_2 & \frac{\lambda^2}{2\sigma} a_4 \\ & & a_2 & -\frac{\lambda}{2} a_4 \\ & & & \frac{\lambda^2}{4} a_6 \end{pmatrix}^{-1}$$

where $a_k := a_k(\lambda) = \frac{1}{\pi}\int_0^\infty z^k \phi(\sqrt{1+2\lambda^2}\,z)\left[\frac{(-1)^k}{\Phi(-\lambda z)} + \frac{1}{\Phi(\lambda z)}\right] dz$ for $k = 0,1,3,4,6$, which have to be evaluated numerically (see Proposition 2 in the Appendix).

2.4. Properties of the MLEs in the normal case

Suppose now that the parameter vector is $\theta^* = (\mu^*, \sigma^*, 0, \alpha^*)^\top$, that is, the data are obtained from a N$(\mu^*, {\sigma^*}^2)$ distribution. At $\theta = \theta^*$, the components of S_θ are

$$S^*_\mu = \frac{Z^*}{\sigma^*}, \quad S^*_\sigma = \frac{Z^{*2} - 1}{\sigma^*}, \quad S^*_\lambda = \frac{(2/\pi)^{1/2} Z^*}{(1 + \alpha^* Z^{*2})^{1/2}} \quad \text{and} \quad S^*_\alpha = 0,$$

where $Z^* = (X - \mu^*)/\sigma^*$.

In this case, the components of $(S_\mu^*, S_\sigma^*, S_\lambda^*)$ are linearly independent at least that $\alpha^* = 0$, and so the singularity of the information matrix of θ^* is due to the fact that $S_\alpha^* = 0$. Moreover, the score component of interest S_λ^* cannot be expressed as a linear combination of the components of $(S_\mu^*, S_\sigma^*, S_\alpha^*)$, that is, there is not a vector $\mathbf{c} \neq 0$ of constants such that $S_\lambda^* = \mathbf{c}^\top (S_\mu^*, S_\sigma^*, S_\alpha^*)^\top$, and so the condition (28) considered by Rotnitzky *et al.* (2000) is not satisfied. Consequently, the methodology proposed by these authors cannot be applied to study the asymptotic properties of the MLEs in the normal case ($\lambda = 0$), since there is no vector $\mathbf{c} \neq (0,0,0,0)^\top$ to initialize the iterative process in order to obtain an appropriate reparametrization for which the information matrix is of full rank. As was mentioned above, this conclusion derives from the fact that α is non-identifiable when $\lambda = 0$.

If, in addition, $\alpha^* = 0$, i.e., $\theta^* = (\mu^*, \sigma^*, 0, 0)^\top$, we then find in the above score functions the relation $S_\lambda^* = (2/\pi)^{1/2}\sigma^* S_\mu^*$. Hence, at θ^* the full information matrix has rank 2, which violates the condition (27) of Rotnitzky *et al.* (2000).

A similar fact occurs when $\alpha \to \infty$, which is another form to obtain the normal model. That is, for $\theta^* = (\mu^*, \sigma^*, \lambda^*, \infty)^\top$, we have whatever the value of λ^* that S_μ^* and S_σ^* are as before, but $S_\lambda^* = S_\alpha^* = 0$. Therefore, again the the methodology proposed by Rotnitzky *et al.* (2000) is not appropriated to study the asymptotic properties of the MLEs in the normal case.

However, if the objective is to study the normality hypothesis only, then a natural and convenient strategy is the following:

a) Use the SGN model to test the skew-normality hypothesis $\alpha = 0$ (see Section 2.3).
b) If the skew-normal model is not rejected, then use this model to test the normality hypothesis $\lambda = 0$. In this case, the Rotnitzky *et al.* (2000) methodology (see Chiogna, 2005) as well as the centred parametrization (see Azzalini, 1985) can be used.

Acknowledgments

The research of R. B. Arellano-Valle was supported by Grant FONDECYT (Chile) 1120121|1090411. The work of H. W. Gómez was supported by Grant FONDECYT (Chile) 1090411 and the work of H. S. Salinas was supported by Grant DIUDA (Chile) 221229. The authors thank the editor and two referees whose constructive comments led to a far improved presentation.

Appendix

This appendix provides preliminary calculations needed to derive the elements of the SGN expected information matrix. To simplify the notation, let $W := W(Z) = \frac{\lambda Z}{\sqrt{1+\alpha Z^2}}$ and $R = R(W) = \frac{\phi(W)}{\Phi(W)}$, where $Z \sim \text{SGN}(\lambda, \alpha)$.

Proposition 1 *Let* $\rho_{nm} = E_Z\left(\frac{Z^n R}{(1+\alpha Z^2)^{m/2}}\right)$ *and* $\eta_{nm} = E_Z\left(\frac{Z^n R^2}{(1+\alpha Z^2)^{m}}\right)$, $n,m = 0,1,\ldots$, *where* $Z \sim SGN(\lambda,\alpha)$. *Then,*

$$\rho_{nm} = \begin{cases} 0, & for\, n = 2k+1\ (odd), \\ E_Y\left(\frac{2Y^{2k}\phi(W(Y))}{(1+\alpha Y^2)^{m/2}}\right), & for\, n = 2k\ (even), \end{cases}$$

and

$$\eta_{nm} = E_Y\left(\left[\frac{(-1)^n}{\Phi(-W(Y))} + \frac{1}{\Phi(W(Y))}\right]\frac{Y^n\phi^2(W(Y))}{(1+\alpha Y^2)^m}\right),$$

where $Y \sim 2\phi(y)I(y \geq 0)$.

Proof: For $n = 2k+1$, we have after a simple algebra that

$$\rho_{nm} = 2\int_{-\infty}^{\infty}\frac{z^{2k+1}\phi(z)}{(1+\alpha z^2)^{m/2}}\phi\left(\frac{\lambda z}{\sqrt{1+\alpha z^2}}\right)dz = 2\int_{-\infty}^{\infty} z h_0(z)dz = 0,$$

since for all $k,m = 0,1,\ldots$, the function $h_0(z) = \frac{z^{2k}\phi(z)}{(1+\alpha z^2)^{m/2}}\phi\left(\frac{\lambda z}{\sqrt{1+\alpha z^2}}\right)$ is even. Similarly, for $n = 2k$, we have

$$\begin{aligned}\rho_{nm} &= 2\int_{-\infty}^{\infty}\frac{z^{2k}\phi(z)}{(1+\alpha z^2)^{m/2}}\phi\left(\frac{\lambda z}{\sqrt{1+\alpha z^2}}\right)dz \\ &= 2\int_0^{\infty}\frac{2y^{2k}\phi(y)}{(1+\alpha y^2)^{m/2}}\phi\left(\frac{\lambda y}{\sqrt{1+\alpha y^2}}\right)dy = 2E_Y\left(\frac{Y^{2k}\phi(W(Y))}{(1+\alpha Y^2)^{m/2}}\right).\end{aligned}$$

Finally, for η_{nm} we have

$$\begin{aligned}\eta_{nm} &= 2\int_{-\infty}^{\infty}\frac{z^n\phi(z)}{(1+\alpha z^2)^m}\frac{\phi^2\left(\frac{\lambda z}{\sqrt{1+\alpha z^2}}\right)}{\Phi\left(\frac{\lambda z}{\sqrt{1+\alpha z^2}}\right)}dz \\ &= 2\int_0^{\infty}\frac{(-y)^n}{\Phi\left(\frac{-\lambda y}{\sqrt{1+\alpha y^2}}\right)}h_1(y)dy + 2\int_0^{\infty}\frac{y^n}{\Phi\left(\frac{\lambda y}{\sqrt{1+\alpha y^2}}\right)}h_1(y)dy \\ &= 2\int_0^{\infty}\left[\frac{(-1)^n}{\Phi(-W(y))} + \frac{1}{\Phi(W(y))}\right]\frac{y^n\phi^2(W(y))}{(1+\alpha y^2)^m}\phi(y)dy \\ &= E_Y\left(\left[\frac{(-1)^n}{\Phi(-W(Y))} + \frac{1}{\Phi(W(Y))}\right]\frac{Y^n\phi^2(W(Y))}{(1+\alpha Y^2)^m}\right),\end{aligned}$$

where it is used that the function $h_1(t) = \frac{\phi^2(W(t))\phi(t)}{(1+\alpha t^2)^m}$ is even for all $m = 0, 1, \ldots$; concluding thus the proof. ■

From Proposition 1 we have, after some straightforward algebra, that the entries $I_{\theta_i\theta_j} = E(S_{\theta_i}S_{\theta_j})$ of the information matrix $\mathbf{I}_\theta$ are as follows:

$$I_{\mu\mu} = E\left(\frac{Z^2}{\sigma^2} - \frac{2\lambda ZR}{\sigma^2(1+\alpha Z^2)^{3/2}} + \frac{\lambda^2 R^2}{\sigma^2(1+\alpha Z^2)^3}\right) = \frac{1}{\sigma^2} + \frac{\lambda^2}{\sigma^2}\eta_{03},$$

$$I_{\mu\sigma} = E\left(-\frac{Z}{\sigma^2} + \frac{Z^3}{\sigma^2} - \frac{2\lambda Z^2 R}{\sigma^2(1+\alpha Z^2)^{3/2}} + \frac{\lambda R}{\sigma^2(1+\alpha Z^2)^{3/2}} + \frac{\lambda^2 Z R^2}{\sigma^2(1+\alpha Z^2)^3}\right)$$

$$= \frac{2}{\sigma^2}(c_1 - c_0) - \frac{(2/\pi)^{1/2}}{\sigma^2} - \frac{2\lambda}{\sigma^2}\rho_{23} + \frac{\lambda}{\sigma^2}\rho_{03} + \frac{\lambda^2}{\sigma^2}\eta_{13},$$

$$I_{\mu\lambda} = E\left(\frac{Z^2 R}{\sigma(1+\alpha Z^2)^{1/2}} - \frac{\lambda Z R^2}{\sigma(1+\alpha Z^2)^2}\right) = \frac{1}{\sigma}\rho_{21} - \frac{\lambda}{\sigma}\eta_{12},$$

$$I_{\mu\alpha} = E\left(-\frac{\lambda Z^4 R}{2\sigma(1+\alpha Z^2)^{3/2}} + \frac{\lambda^2 Z^3 R^2}{2\sigma(1+\alpha Z^2)^3}\right) = -\frac{\lambda}{2\sigma}\rho_{43} + \frac{\lambda^2}{2\sigma}\eta_{33},$$

$$I_{\sigma\sigma} = E\left(\frac{1}{\sigma^2} - \frac{2Z^2}{\sigma^2} + \frac{2\lambda ZR}{\sigma^2(1+\alpha Z^2)^{3/2}} + \frac{Z^4}{\sigma^2} - \frac{2\lambda Z^3 R}{\sigma^2(1+\alpha Z^2)^{3/2}} + \frac{\lambda^2 Z^2 R^2}{\sigma^2(1+\alpha Z^2)^3}\right)$$

$$= \frac{1}{\sigma^2} - \frac{2}{\sigma^2} + \frac{3}{\sigma^2} + \frac{\lambda^2}{\sigma^2}\eta_{23} = \frac{2}{\sigma^2} + \frac{\lambda^2}{\sigma^2}\eta_{23},$$

$$I_{\sigma\lambda} = E\left(-\frac{ZR}{\sigma(1+\alpha Z^2)^{1/2}} + \frac{Z^3 R}{\sigma(1+\alpha Z^2)^{1/2}} - \frac{\lambda Z^2 R^2}{\sigma(1+\alpha Z^2)^2}\right) = -\frac{\lambda}{\sigma}\eta_{22},$$

$$I_{\sigma\alpha} = E\left(\frac{\lambda Z^3 R}{2\sigma(1+\alpha Z^2)^{3/2}} - \frac{\lambda Z^5 R}{2\sigma(1+\alpha Z^2)^{3/2}} + \frac{\lambda^2 Z^4 R^2}{2\sigma(1+\alpha Z^2)^3}\right) = \frac{\lambda^2}{2\sigma}\eta_{43},$$

$$I_{\lambda\lambda} = E\left(\frac{Z^2 R^2}{(1+\alpha Z^2)}\right) = \eta_{21},$$

$$I_{\lambda\alpha} = E\left(-\frac{\lambda Z^4 R^2}{2(1+\alpha Z^2)^2}\right) = -\frac{\lambda}{2}\eta_{42},$$

$$I_{\alpha\alpha} = E\left(\frac{\lambda^2 Z^6 R^2}{4(1+\alpha Z^2)^3}\right) = \frac{\lambda^2}{4}\eta_{63}.$$

Corollary 1 *If $\lambda = 0$, then the entries $I_{\mu\lambda}$ and $I_{\lambda\lambda}$ of the information matrix $\mathbf{I}_\theta$ reduce to*

$$I_{\mu\lambda} = \frac{2(2/\pi)^{1/2}}{\sigma} d_1(\alpha) \quad and \quad I_{\lambda\lambda} = \frac{2}{\pi} d_2(\alpha), \tag{3}$$

where $d_1(\alpha) = \int_0^\infty \frac{z^2\phi(z)}{(1+\alpha z^2)^{1/2}} dz$ *and* $d_2(\alpha) = \frac{1}{\alpha}\left(1 - (2\pi/\alpha)^{1/2} e^{\frac{1}{2\alpha}} \Phi(-\alpha^{-1/2})\right)$.

Proof: In fact, if $\lambda = 0$, then $W \equiv 0$, and so $R \equiv (2/\pi)^{1/2}$. Thus, $I_{\mu\lambda} = \frac{1}{\sigma} E\left(\frac{Z^2}{(1+\alpha Z^2)^{1/2}} R\right)$ $= \frac{(2/\pi)^{1/2}}{\sigma} \int_{-\infty}^\infty \frac{z^2\phi(z)}{(1+\alpha z^2)^{1/2}} dz = \frac{2(2/\pi)^{1/2}}{\sigma} \int_0^\infty \frac{z^2\phi(z)}{(1+\alpha z^2)^{1/2}} dz$ since the function $\frac{z^2\phi(z)}{(1+\alpha z^2)^{1/2}}$ is even. Note that this integral has be computed numerically when $\alpha > 0$. For $I_{\lambda\lambda}$ we have $I_{\lambda\lambda} = E\left(\frac{Z^2}{1+\alpha Z^2} R^2\right) = \frac{2}{\pi} \int_{-\infty}^\infty \frac{z^2\phi(z)}{1+\alpha z^2} dz = \frac{4}{\pi} \int_0^\infty \frac{z^2\phi(z)}{1+\alpha z^2} dz$ since that function $\frac{z^2\phi(z)}{1+\alpha z^2}$ is even. Hence, the result follows by noting from Mathematica (Wolfram Research, 2008) that

$$d_2(\alpha) := 2\int_0^\infty \frac{z^2\phi(z)}{1+\alpha z^2} dz = \frac{1}{\alpha} - \frac{(\pi/2)^{1/2} e^{\frac{1}{2\alpha}} (1 - \operatorname{erf}(\frac{\sqrt{2}}{2\sqrt{\alpha}}))}{\alpha^{3/2}},$$

for $\alpha > 0$, where $\operatorname{erf}\left(\frac{\sqrt{2}}{2} t\right) = 2\Phi(t) - 1$. ∎

Proposition 2 *Let $Z \sim SGN(\lambda, 0)$. Then*

$$a_k(\lambda) := E_Z\left(Z^k \left\{\frac{\phi(\lambda Z)}{\Phi(\lambda Z)}\right\}^2\right) = \frac{1}{\pi}\int_0^\infty z^k \phi(\sqrt{1+2\lambda^2} z)\left[\frac{(-1)^k}{\Phi(-\lambda z)} + \frac{1}{\Phi(\lambda z)}\right] dz.$$

Proof: Since $\phi^2(\lambda z)\phi(z) = \frac{1}{2\pi}\phi(\sqrt{1+2\lambda^2} z)$ we have after a simple algebra that

$$\begin{aligned} E_Z\left(Z^k \left\{\frac{\phi(\lambda Z)}{\Phi(\lambda Z)}\right\}^2\right) &= 2\int_{-\infty}^\infty \frac{z^k \phi^2(\lambda z)\phi(z)}{\Phi(\lambda z)} dz \\ &= \frac{1}{\pi}\int_{-\infty}^\infty \frac{z^k \phi(\sqrt{1+2\lambda^2} z)}{\Phi(\lambda z)} dz \\ &= \frac{1}{\pi}\int_0^\infty z^k \phi(\sqrt{1+2\lambda^2} z)\left[\frac{(-1)^k}{\Phi(-\lambda z)} + \frac{1}{\Phi(\lambda z)}\right] dz. \end{aligned}$$

∎

References

Arellano-Valle, R. B., Gómez, H. W. and Quintana, F. A. (2004). A new class of skew-normal distributions. *Communications in Statistics: Theory and Methods*, 33(7), 1465–1480.

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

Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods & applications*, 14, 331–341.

Rotnitzky, A., Cox, D. R., Bottai, M. and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6, 243–284.

Wolfram Research, Inc., Mathematica, Version 7.0, Champaign, IL (2008).