ESTIMATING THE CLOSED SKEW-NORMAL DISTRIBUTIONS PARAMETERS USING WEIGHTED MOMENTS

BY

Cédric Flecher
Philippe Naveau
Denis Allard

Research Report No. 40
Septembre 2009

Unité Biostatistique et Processus Spatiaux
Institut National de la Recherche Agronomique
Avignon, France
http://www.biosp.org

Estimating the closed skew-normal distributions parameters using weighted moments

by C. Flecher[1,2], P. Naveau[1] and D. Allard[2]

[1] Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France

[2] Biostatistics and Spatial Processes, INRA, Agroparc 84914 Avignon, France

September 7, 2008

Abstract

Skewness is often present in a wide range of applied problems. One possible approach to model this skewness is based on the class of skew normal distributions. Fitting such distributions remains an inference challenge in various cases. In this paper, we propose and study novel estimators using weighted moments for the closed multivariate skew-normal distribution.

1 Introduction

Multivariate skew-normal distributions are extensions of the normal distribution which admit skewness whilst retaining most of the interesting properties of the Gaussian distribution. The book edited by Genton (2004) provides an overview of theoretical and applied developments related to skewed distributions. More recently, a special issue in the journal *Communications in statistics* edited by Pewsey and González-Farías (2007) was centered on the different aspects of skew distributions. In the large class of skew distributions, we focus on the multivariate closed skew-normal distribution proposed by González-Farías *et al.* (2004). Such a distribution has many interesting properties inherited from the Gaussian distribution. Concerning its definition, an k-dimensional random vector $\mathbf{Y}$ is said to have a multivariate closed skew-normal distribution, denoted by $\mathrm{CSN}_{k,l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$, if its density function is of the form

$$f_{k,l}(y) = c_l \; \phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \; \Phi_l(\mathbf{D}^t(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta}), \text{ with } c_l^{-1} = \Phi_l(\mathbf{0}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D^t \Sigma D}), \quad (1)$$

where $\boldsymbol{\mu} \in \mathbb{R}^k$, $\boldsymbol{\nu} \in \mathbb{R}^l$, $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{l \times l}$ are both covariance matrices, $\mathbf{D} \in \mathbb{R}^{k \times l}$, $\phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the probability distribution function (pdf) and cumulative distribution function (cdf), respectively, of the k-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{D}^t$ is the transpose of the matrix $\mathbf{D}$. If $\mathbf{D} = \mathbf{0}$, the density (1) reduces to the multivariate normal one. When $l = 1$, the density of Azzalini (2005) is obtained, i.e. the variable $\mathbf{Y}$ follows a $\mathrm{CSN}_{k,1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, 0, 1)$ distribution, where $\boldsymbol{\alpha}$ is a vector of length k.

Concerning the inference of the Azzalini skew normal distribution parameters, Azzalini and Capitanio (1999) studied the classical maximum likelihood estimation (mle) approach. In sections 5 and 6 of their paper, these authors noticed that they are numerous statistical issues with the mle procedure, even for the univariate case. They wrote that *an alternative estimation method is called for.* This is one motivation for this work.

One of the advantages of parametrization (1) resides in the additive stability of the distribution. González-Farías *et al.* (2004) proved the closure under linear transformation of independent CSN random vectors. In particular, the sum of independent CSN vectors is still a CSN vector. They also derived the moments of $\mathbf{Y}$. The first moment equals to

$$\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} + \frac{\Phi_l^*(\mathbf{D}^t\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{D}^t\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}, \quad (2)$$

where

$$\Phi_l^*(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D}) = \sum_{i=1}^{k}\sum_{j=1}^{l}(\mathbf{D}^t\boldsymbol{\Sigma})_{ij}\Phi_l^{\{j\}}(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})\mathbf{e}_i, \tag{3}$$

and $\mathbf{e}_i$ is a $(k \times 1)$ vector with one in the i^{th} position and zero elsewhere, and

$$\Phi_l^{\{j\}}(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D}) = \phi([\mathbf{D}^t\boldsymbol{\mu}]_j;\boldsymbol{\nu}_j,[\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D}]_{jj})\Phi_{l-1}([\mathbf{D}^t\boldsymbol{\mu}]_{-j};\boldsymbol{\nu}_{-j},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D}|[\mathbf{D}^t\boldsymbol{\mu}]_j).$$

where $[\mathbf{D}^t\boldsymbol{\mu}]_j$ is the j^{th} element of $\mathbf{D}^t\boldsymbol{\mu}$ and $[\mathbf{D}^t\boldsymbol{\mu}]_{-j}$ is the vector $\mathbf{D}^t\boldsymbol{\mu}$ without the j^{th} element. The variance can be expressed as

$$\mathbb{V}ar(\mathbf{Y}) = \boldsymbol{\Sigma} + \frac{\Phi_l^{**}(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})} - \mathbb{E}(\mathbf{Y}-\boldsymbol{\mu})\mathbb{E}(\mathbf{Y}-\boldsymbol{\mu})^t, \tag{4}$$

where

$$\Phi_l^{**}(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D}) = \frac{\partial}{\partial\mathbf{t}\partial\mathbf{t}^t}\Phi_l(\mathbf{D}^t(\boldsymbol{\mu}+\boldsymbol{\Sigma}\mathbf{t});\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})|_{\mathbf{t}=\mathbf{0}}. \tag{5}$$

These expressions can quickly become cumbersome in practice. The main reason for this difficulty is the absence of analytical representations for $\Phi_l(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})$ and its derivatives, even when $\boldsymbol{\mu} = \mathbf{0}$. To apply the CSN to real case studies, the practitioner needs to be able to easily implement an estimation method that provides accurate estimators of the CSN parameters. In its most general form, it is unlikely that such an objective can be reached for the CSN because of its large number of parameters. This aim can be achieved in some specific cases, basically when (2) and (4) can be simplified. Such cases still can offer an added flexibility and skewness with respect to the classical Gaussian distribution. To reach this goal in such cases our strategy is to develop a novel Method of Moments (MOM) approach that will be described in Section 2.

Before closing this section, we derive a lemma that will help us to compute the mean and variance of a CSN. All our proofs can be found in the Appendix.

Lemma 1. *Let* $\Phi_l(\cdot;\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D^t\Sigma D})$ *be a* l*-dimensional Gaussian cdf with mean* $\boldsymbol{\nu}\in\mathbb{R}^l$ *and covariance matrix* $\boldsymbol{\Delta}+\mathbf{D^t\Sigma D}$*. Then we can write*

$$\frac{\Phi_l^*(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}=\frac{\Phi_l^*(\mathbf{0};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{0};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}. \tag{6}$$

and

$$\frac{\Phi_l^{**}(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}=\frac{\Phi_l^{**}(\mathbf{0};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{0};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}, \tag{7}$$

where $\Phi_l^*(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})$ *and* $\Phi_l^{**}(\mathbf{D}^t\boldsymbol{\mu};\boldsymbol{\nu},\boldsymbol{\Delta}+\mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})$ *are defined by (3) and (5), respectively.*

In practice, this lemma indicates that, although $\boldsymbol{\mu}$ has still to be estimated, $\boldsymbol{\mu}$ can be set equal to zero in the ratios (6) and (7).

2 Our inference method

The key concept for any MOM approach is to be able to explicitly express moments in terms of the unknown parameters. From a CSN vector, there are at least three quantities to estimate: the covariance matrix, the mean vector and the skewness parameters. Hence, working with the first two moments does not provide enough equations. A possible path is to look at higher moments but this strategy is associated with a least two drawbacks. The mathematical expression of the third and fourth CSN moments are too complex to be of any practical use. In addition, the estimation of the third and fourth CSN moments are classically tainted by large variances. To avoid these hurdles, we modify a MOM approach used in Extreme Value Theory.

Probability Weighted Moments are a class of MOM estimators introduced by Hosking *et al.* (1985) and their collaborators. Such approaches have been extended by Diebolt *et al.* (2008). The basic idea is to compute and estimate moments of the type $\mathbb{E}(Z^sF^r(Z))$ where F corresponds to the cumulative distribution function of the univariate random variable Z and r,s are integers, usually equal to one or two. For the extreme value distributions, such quantities have explicit analytical expressions (e.g.

Hosking *et al.*, 1985; Diebolt *et al.*, 2008). In the case of the CSN, the computation of these so-called Probability Weighted Moments seems very difficult to derive because the "weight" $F^r(Z)$ is very complex for a CSN. However, replacing $F^r(.)$ by the cumulative Gaussian distribution $\Phi^r(.;\mu,\sigma^2)$ greatly simplifies the problem. This is even true for the multivariate case. The following proposition summarizes our findings on this topic.

Proposition 2. *Let* $\mathbf{Y}$ *be a* $CSN_{k,l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \mathbf{0}, \boldsymbol{\Delta})$ *defined by (1) and* $h(\mathbf{y}) = h(y_1, \ldots, y_k)$ *be any real valued function such that* $\mathbb{E}(h(\mathbf{Y}))$ *is finite, then*

$$\mathbb{E}(h(\mathbf{Y})\Phi_k^r(\mathbf{Y};\mathbf{0},\mathbf{I}_k)) = \frac{\Phi_{rk+l}(\mathbf{0};\boldsymbol{\nu}_+, \boldsymbol{\Delta}_+ + \mathbf{D}_+^{\mathbf{t}}\boldsymbol{\Sigma}\mathbf{D}_+)}{\Phi_l(\mathbf{0};\mathbf{0}, \boldsymbol{\Delta} + \mathbf{D}^{\mathbf{t}}\boldsymbol{\Sigma}\mathbf{D})}\mathbb{E}(h(\mathbf{Y}_+)), \tag{8}$$

where $\mathbf{Y}_+$ *follows a* $CSN_{k,rk+l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}_+, \nu_+, \boldsymbol{\Delta}_+)$ *with* $\boldsymbol{\Delta}_+ = \begin{bmatrix} \mathbf{I}_{rk} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta} \end{bmatrix}$, $\mathbf{D}_+ = \begin{bmatrix} \mathbf{D}_*^t \\ \mathbf{D}^t \end{bmatrix}^t$, $\mathbf{D}_*$ *a* $k \times rk$ *matrix defined by* $\mathbf{D}_* = \begin{bmatrix} \mathbf{I}_k \\ \ldots \\ \mathbf{I}_k \end{bmatrix}^t$ *with* $\mathbf{I}_k$ *the identity matrix of size* k*, and* $\boldsymbol{\nu}_+$ *a* $rk + l$ *vector defined by* $\boldsymbol{\nu}_+^t = (-\boldsymbol{\mu}, \ldots, -\boldsymbol{\mu}, \mathbf{0}_l)$.

To understand Equation (8) we consider two special cases. First, suppose that $r = 1$ and $h(y_1, \ldots, y_k) = y_i$ for some $i = 1, \ldots, k$. Then, using Lemma 1 and Equation (2) within (8), we can write

$$\mathbb{E}(Y_i\Phi_k(\mathbf{Y};\mathbf{0},\mathbf{I}_k)) = \mu_i\frac{\Phi_{k+l}(\mathbf{0};\boldsymbol{\nu}_+, \boldsymbol{\Delta}_+ + \mathbf{D}_+^{\mathbf{t}}\boldsymbol{\Sigma}\mathbf{D}_+)}{\Phi_l(\mathbf{0};\mathbf{0}, \boldsymbol{\Delta} + \mathbf{D}^{\mathbf{t}}\boldsymbol{\Sigma}\mathbf{D})} + \frac{\left[\Phi_{k+l}^*(\mathbf{0};\boldsymbol{\nu}_+, \boldsymbol{\Delta}_+ + \mathbf{D}_+^{\mathbf{t}}\boldsymbol{\Sigma}\mathbf{D}_+)\right]_i}{\Phi_l(\mathbf{0};\mathbf{0}, \boldsymbol{\Delta} + \mathbf{D}^{\mathbf{t}}\boldsymbol{\Sigma}\mathbf{D})}. \tag{9}$$

The left hand side of (9) can be viewed as a type of weighted moments. Concerning the right hand side, it basically means that, if $\mathbf{D}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ are chosen such as the first moment (2) and second moment (4) have explicit expressions or can be reasonably approximated, then the weight moment defined by (9) can also be derived. In other words, if one can compute or approximate the first and second moments expressions, then a weighted moment can also be obtained and a MOM approach with a third equation can be implemented.

Our second case concerns the bivariate sub vector $(Y_i, Y_j)^t$ for some $i, j = 1, \dots, k$ that follows a $\mathrm{CSN}_{2,l}(\mathbf{0}, \mathbf{\Sigma}_{ij}, \mathbf{D}_{ij}, \mathbf{0}, \mathbf{\Delta}_{ij})$ with obvious notations for the matrices. To apply Proposition 2 we choose $h(y_i, y_j) = y_i y_j$ and $r = 2$. If $\Phi(x)$ denote a shortcut notation for the standardized cumulative Gaussian distribution $\Phi_1(x; 0, 1)$, we can then write

$$\mathbb{E}(Y_i Y_j \Phi(Y_i)\Phi(Y_j)) = \frac{\Phi_{4+l}(\mathbf{0}; \mathbf{0}, \mathbf{\Delta}_{+ij} + \mathbf{D}^t_{+ij}\mathbf{\Sigma}\mathbf{D}_{+ij})}{\Phi_l(\mathbf{0}; \mathbf{0}, \mathbf{\Delta}_{ij} + \mathbf{D}^t_{ij}\mathbf{\Sigma}_{ij}\mathbf{D}_{ij})} \mathbb{E}(Y_{+i} Y_{+j}), \qquad (10)$$

where $(Y_{+i}, Y_{+j})^t$ follows a $CSN_{2,2+l}(\mathbf{0}, \mathbf{\Sigma}_{ij}, \mathbf{D}_{+ij}, \mathbf{0}, \mathbf{\Delta}_{+\mathbf{ij}})$ and these elements can be obtained from (4). Equation (10) will be useful to derive the variance properties of our proposed estimators. In Equation (10) we took $(\mu_i, \mu_j) = (0, 0)$. When this assumption does not held, we can take advantage of Lemma 1.

If $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are both assumed to be unknown, then they cannot be identified in (1) because the expression $\Phi_l(\mathbf{D}^t(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \mathbf{\Delta})$ is over-parametrized in this case. Following González-Farías *et al.* (2004) we decide to set $\boldsymbol{\nu} = \mathbf{0}$ in the rest of the paper. In our applications, we also suppose that the matrix $\mathbf{\Delta}$ is a function of the other three unknown parameters $\boldsymbol{\mu}, \mathbf{\Sigma}$ and $\mathbf{D}$. Conceptually, it may be possible to deal with the estimation of $\mathbf{\Delta}$ by working with a moment of the type $\mathbb{E}(Y_i \Phi_k^2(\mathbf{Y}; \mathbf{0}, \mathbf{I}_k))$. But it becomes burdensome in terms of notations and complicated in terms of optimization.

3 Examples and estimation procedures

3.1 The univariate case

Assume that $X_1, \dots, X_n$ are n independent and identically distributed (iid) $CSN_{1,1}(\mu, \sigma^2, \delta, 0, \Delta)$ random variables with μ, σ^2 and δ unknown. Define $\lambda = \frac{\sigma\delta}{\sqrt{\Delta + \sigma^2\delta^2}}$, the first and second moment expressions are easily derived from (2) and (4) with $k = l = 1$

$$\mathbb{E}(X_i) = \mu + 2\Phi_1^*(0; 0, \Delta + \delta^2\sigma^2) = \mu + \frac{2}{\sqrt{2\pi}}\sigma\lambda,$$

and

$$\mathbb{V}ar(X_i) = \sigma^2 + 2\Phi_1^{**}(0; 0, \Delta + \delta^2\sigma^2) - (2\Phi_1^*(0; 0, \Delta + \delta^2\sigma^2))^2 = \sigma^2\left(1 - \frac{2}{\pi}\lambda^2\right).$$

A direct application of Equation (9) gives

$$\mathbb{E}(X_i\Phi(X_i)) = 2\mu\ \Phi_2\left(\mathbf{0}; \boldsymbol{\nu}_{2+}, \mathbf{A}_2\right) + 2\Phi_2^*\left(\mathbf{0}; \boldsymbol{\nu}_{2+}, \mathbf{A}_2\right) \tag{11}$$

where $\Phi_2^*\left(\mathbf{0}; \boldsymbol{\nu}_{2+}, \mathbf{A}_2\right) = \sigma^2\left\{\Phi_2^{\{1\}}\left(\mathbf{0}; \boldsymbol{\nu}_{2+}, \mathbf{A}_2\right) + \delta\ \Phi_2^{\{2\}}\left(\mathbf{0}; \boldsymbol{\nu}_{2+}, \mathbf{A}_2\right)\right\}$ and

$$\boldsymbol{\nu}_{2+} = \begin{pmatrix} -\frac{\mu}{\sqrt{1+\sigma^2}} \\ 0 \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \mathbf{A}_2 = \begin{bmatrix} 1 & \frac{\sigma}{\sqrt{1+\sigma^2}}\lambda \\ \frac{\sigma}{\sqrt{1+\sigma^2}}\lambda & 1 \end{bmatrix}. \tag{12}$$

In terms of estimation, the sample mean denoted $\overline{X}_n = (X_1 + \cdots + X_n)/n$ and the sample variance $s_n^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ can be plugged in the above moment expressions. The weighted moment $m = \mathbb{E}(X_i\Phi(X_i))$ can be estimated by the unbiased statistic $m_n = \frac{1}{n}\sum_{i=1}^n X_i\Phi(X_i)$. The following system of three equations with unknowns $\hat{\mu}$, $\hat{\delta}$ and $\hat{\sigma}$ follows

$$\begin{cases} \overline{x}_n = \hat{\mu} + \sqrt{\frac{2}{\pi}}\hat{\sigma}\hat{\lambda} \\ s_n^2 = \hat{\sigma}^2\left(1 - \frac{2}{\pi}\hat{\lambda}^2\right) \\ m_n = 2\hat{\mu}\ \Phi_2\left(\mathbf{0}; \hat{\boldsymbol{\nu}}_{2+}, \hat{\mathbf{A}}_2\right) + 2\Phi_2^*\left(\mathbf{0}; \hat{\boldsymbol{\nu}}_{2+}, \hat{\mathbf{A}}_2\right) \end{cases}, \tag{13}$$

where $\hat{\boldsymbol{\nu}}_{2+}$ and $\hat{\mathbf{A}}_2$ are defined by (12). Concerning the asymptotic properties of our moments, they can described by the following proposition.

Proposition 3. *Assume that $X_1, \ldots, X_n$ are n iid $CSN_{1,1}(\mu, \sigma^2, \delta, 0, \Delta)$ variables where μ, σ^2 and δ are unknown and define $\lambda = \frac{\sigma\delta}{\sqrt{\Delta+\sigma^2\delta^2}}$. The unbiased estimator vector $\sqrt{n}\left(\overline{x}_n - \mathbb{E}(X), s_n^2 - \mathbb{V}ar(X), m_n - \mathbb{E}(X\Phi(X))\right)^t$ converges in distribution, as* *n goes to infinity, to a zero-mean Gaussian vector with a covariance matrix whose elements c_{ij}/n equal*

$$
\begin{aligned}
c_{1,1} &= \sigma^2\left(1-\frac{2}{\pi}\lambda\right), \quad c_{2,2}=\sigma^4\left(4(\frac{2}{\pi}-\frac{3}{\pi^2})\lambda^4+6\frac{2}{\pi}\lambda^2+3\right), \\
c_{3,3} &= \mathbb{E}(X^2\Phi^2(X))-\mathbb{E}^2(X\Phi(X)), \\
c_{1,2} &= c_{2,1}=\frac{2}{\sqrt{2\pi}}\sigma^3\lambda^3\left(\frac{2}{\pi}-1\right), \\
c_{1,3} &= c_{3,1}=\mathbb{E}(X^2\Phi(X))-\mathbb{E}(X)\mathbb{E}(X\Phi(X)), \\
c_{2,3} &= c_{3,2}=\mathbb{E}(X^3\Phi(X))-\mathbb{E}(X)\mathbb{E}(X^2\Phi(X))-\mathbb{V}ar(X)\mathbb{E}(X\Phi(X)),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}(X^2\Phi^2(X)) &= 2(\sigma^2+\mu^2)\Phi_3(\mathbf{0};\boldsymbol{\nu}_{3+},\mathbf{A}_3)+4\mu\Phi_3^*(\mathbf{0};\boldsymbol{\nu}_{3+},\mathbf{A}_3)+2\Phi_3^{**}(\mathbf{0};\boldsymbol{\nu}_{3+},\mathbf{A}_3), \\
\mathbb{E}(X^2\Phi(X)) &= 2\mu\Phi_2(\mathbf{0};\boldsymbol{\nu}_{2+},\mathbf{A}_2)+2\Phi_2^*(\mathbf{0};\boldsymbol{\nu}_{2+},\mathbf{A}_2), \\
\mathbb{E}(X^3\Phi(X)) &= 2\Phi_2(\mathbf{0};\boldsymbol{\nu}_{2+},\mathbf{A}_2)\mathbb{E}(X_+^3),
\end{aligned}
$$

with $\boldsymbol{\nu}_{2+}$ *and* $\mathbf{A}_2$ *defined by (12),*

$$
\boldsymbol{\nu}_{3+} = \begin{pmatrix} -\frac{\mu}{\sqrt{1+\sigma^2}} \\ -\frac{\mu}{\sqrt{1+\sigma^2}} \\ 0 \end{pmatrix}, \text{ and } \mathbf{A}_3 = \begin{bmatrix} 1 & \frac{\sigma^2}{1+\sigma^2} & \frac{\sigma}{\sqrt{1+\sigma^2}}\lambda \\ \frac{\sigma^2}{1+\sigma^2} & 1 & \frac{\sigma}{\sqrt{1+\sigma^2}}\lambda \\ \frac{\sigma}{\sqrt{1+\sigma^2}}\lambda & \frac{\sigma}{\sqrt{1+\sigma^2}}\lambda & 1 \end{bmatrix}.
$$

To assess our MOM estimation scheme with simulations, we focus on the model described by Allard and Naveau (2007) and defined as a $CSN_{1,1}(\mu,\sigma^2,\lambda\sigma^{-1},0,1-\lambda^2)$ for some $0\leq|\lambda|<1$. In this case, the system (13) can be written as

$$
\begin{cases}
\overline{x}_n = \hat{\mu}+\sqrt{\frac{2}{\pi}}\hat{\sigma}\hat{\lambda} \\
s_n^2 = \hat{\sigma}^2\left(1-\frac{2}{\pi}\hat{\lambda}^2\right) \\
m_n = 2\hat{\mu}\ \Phi_2\left(\mathbf{0};\hat{\boldsymbol{\nu}}_+,\hat{\mathbf{A}}_2\right)+2\Phi_2^*\left(\mathbf{0};\hat{\boldsymbol{\nu}}_+,\hat{\mathbf{A}}_2\right)
\end{cases}, \tag{14}
$$

where $\Phi_2\left(\mathbf{0};\hat{\boldsymbol{\nu}}_+,\hat{\mathbf{A}}_2\right)=\Phi_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix};\begin{pmatrix} -\frac{\hat{\mu}}{\sqrt{1+\hat{\sigma}^2}} \\ 0 \end{pmatrix},\begin{pmatrix} 1 & \hat{\lambda}\frac{\hat{\sigma}}{\sqrt{1+\hat{\sigma}^2}} \\ \hat{\lambda}\frac{\hat{\sigma}}{\sqrt{1+\hat{\sigma}^2}} & 1 \end{pmatrix}\right]$ and the probability $\Phi_2^*\left(\mathbf{0};\hat{\boldsymbol{\nu}}_+,\hat{\mathbf{A}}_2\right)$ equals

$$\hat{\sigma}^2\,\phi\left(\hat{\mu};0,1+\hat{\sigma}^2\right)\Phi\left(-\frac{\hat{\sigma}}{1+\hat{\sigma}^2}\hat{\lambda}\hat{\mu};0,\frac{1+\hat{\sigma}^2(1-\hat{\lambda}^2)}{1+\hat{\sigma}^2}\right)+\frac{\hat{\sigma}\hat{\lambda}}{\sqrt{2\pi}}\Phi\left(\hat{\mu};0,1+\hat{\sigma}^2(1-\hat{\lambda}^2)\right).$$

For our simulations, 1000 samples of sizes $n=50$ and $n=100$ are generated from a $CSN_{1,1}(0,1,\lambda,0,1-\lambda^2)$ with λ in $\{0.71,0.89,0.97\}$, these values correspond to δ in $\{1,2,4\}$. Table 1 compares the Mean Squared Errors for each parameter obtained from our weighted moment approach with the ones derived with the mle method studied by Azzalini and Capitanio (1999). The column of percentages corresponds to the number of cases for which both methods give the correct sign for λ. The lowest MSE value for the two methods is represented in bold letters. This table indicates that our approach performs better than the mle for the small sample size $n=50$, especially for the parameter σ. As the sample size increases, both methods improve and become comparable (although the mle has still difficulties for σ).

Table 1: Mean Squared Error computed from 1000 samples of sizes $n=50$ and $n=100$ generated from a $CSN_{1,1}(0,1,\lambda,0,1-\lambda^2)$ with λ in $\{0.71,0.89,0.97\}$

		Weighted Moment Method			Maximum Likelihood Method		
		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$
$n=50$							
$\lambda=0.71$	57.9%	**0.086**	**0.043**	**0.031**	0.091	0.24	0.034
$\lambda=0.89$	86.6%	**0.032**	**0.024**	**0.010**	0.049	0.14	0.024
$\lambda=0.97$	99.4%	**0.024**	**0.022**	**0.0061**	0.028	0.088	0.012
$n=100$							
$\lambda=0.71$	64.4%	0.060	**0.027**	0.025	**0.052**	0.12	**0.023**
$\lambda=0.89$	94.8%	**0.025**	**0.016**	**0.0092**	0.030	0.071	0.015
$\lambda=0.97$	99.9%	0.013	**0.012**	0.0021	**0.012**	0.054	**0.0012**

3.2 A multivariate example

Following our univariate example, we study the following multivariate distribution $CSN_{k,k}\left(\boldsymbol{\mu},\boldsymbol{\Sigma},\lambda\boldsymbol{\Sigma}^{-1/2},0,(1-\lambda^2)\mathbf{I}_k\right)$, where $\boldsymbol{\mu}$ is a location vector of size k, $\boldsymbol{\Sigma}$ a $k\times k$

variance covariance matrix and λ a scalar in $]-1,1[$. Such a parametrization corresponds to $\mathbf{\Delta}+\mathbf{D^t}\mathbf{\Sigma}\mathbf{D}=\mathbf{I}_k$ and the moments are

$$\mathbb{E}(\mathbf{Y})=\boldsymbol{\mu}+\frac{2}{\sqrt{2\pi}}\lambda\mathbf{\Sigma}^{1/2}\mathbf{1}_k,\ \mathbb{V}ar(\mathbf{Y})=\mathbf{\Sigma}\left(1-\frac{2}{\pi}\lambda^2\right)\ \text{and}$$

$$\mathbb{E}(\Phi_k(\mathbf{Y},\mathbf{0},\mathbf{Id}_n))=2^k\Phi_{2k}\left(\mathbf{0};\begin{bmatrix}-\boldsymbol{\mu}\\ \mathbf{0}\end{bmatrix},\begin{bmatrix}\mathbf{\Sigma}+\mathbf{I}_k & \lambda\mathbf{\Sigma}^{1/2}\\ \lambda\mathbf{\Sigma}^{1/2} & \mathbf{I}_k\end{bmatrix}\right)$$

Figures 1 and 2 summarize, with boxplots and histograms, our MOM estimations obtained with 1000 samples of size 500 generated from a bivariate random vector $CSN_{2,2}\left(\boldsymbol{\mu},\mathbf{\Sigma},\lambda\mathbf{\Sigma}^{-1/2},0,(1-\lambda^2)\mathbf{I}_2\right)$ with $\mu_1=\mu_2=0$, $\sigma_1=\sigma_2=1$, $\lambda=0.89$ and $\rho=0.8$. The boxplots correspond to the parameters μ and σ and the histograms to ρ and λ. These graphs indicate that our approach provides reasonable results in this bivariate case.

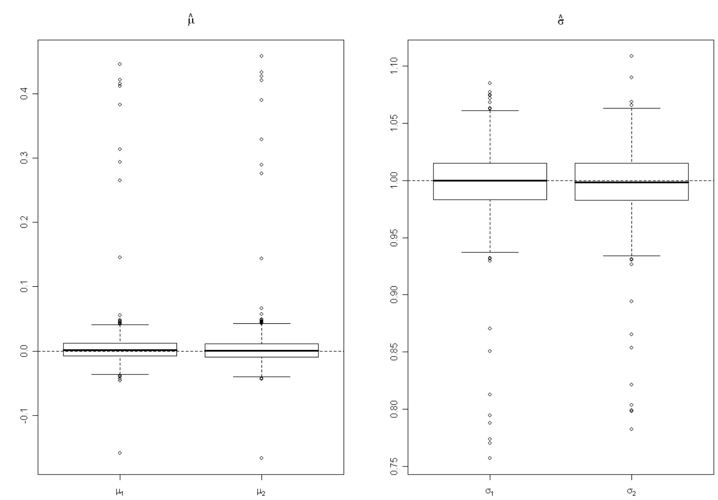

Figure 1: Estimated values of $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1$ and $\hat{\sigma}_2$ obtained from 1000 samples of size 500 generated from a bivariate $CSN_{2,2}\left(\boldsymbol{\mu},\mathbf{\Sigma},\lambda\mathbf{\Sigma}^{-1/2},0,(1-\lambda^2)\mathbf{I}_2\right)$ with $\mu_1=\mu_2=0$, $\sigma_1=\sigma_2=1$, $\lambda=0.89$ and $\rho=0.8$. Solid lines represent true values.

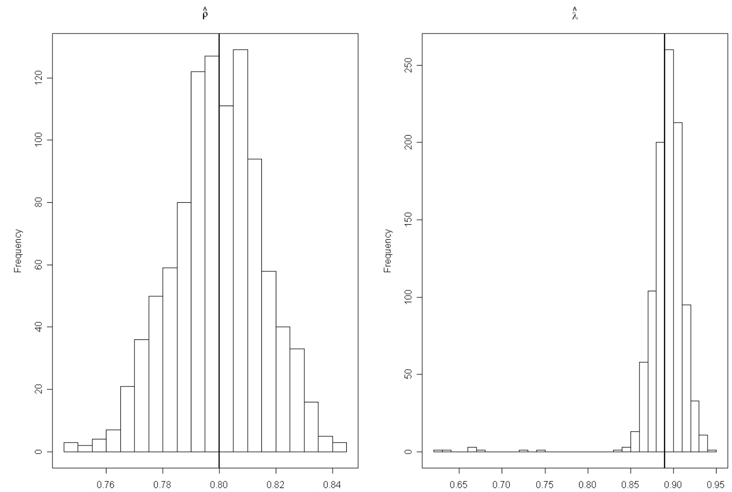

Figure 2: Histograms of estimated values for $\hat{\rho}$ (left) and $\hat{\lambda}$ (right) from the same simulations described in Figure 1

4 Concluding remarks

In this paper, we have introduced a new inference approach for closed skew normal distribution by modifying a MOM approach used in Extreme Value Theory. In the simulation study, we compare our estimators with mle ones in the univariate case. Our approach seems to outperform the mle for small sample sizes. As illustrated by our bivariate example, the basic principles of our method can be applied to some multivariate cases. In conclusion, our goal was not to show the superiority of our approach over the mle. Instead and as advocated by Azzalini and Capitanio (1999), we aimed at providing an alternative inference scheme for skew normal distributions. This method could be used to initialize mle algorithms or whenever the mle provides unreasonable values.

Acknowledgments. This work was supported by the CLIMATOR project, the NCAR Weather and Climate Impact Assessment Science Initiative and the ANR-AssimilEx project. The authors would also like to credit the contributors of the R project.

References

Allard, D. and Naveau, P., 2007. A new spatial skew-normal random field model. *Communication in Statistics*, **36**, 1821–1834.

Azzalini, A. and Capitanio, A., 1999. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, B.*, **61, part 3**, 579–602.

Azzalini, A., 2005. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, **32**, 159–188.

Diebolt, J., A. Guillou, P. Naveau and P. Ribereau, 2008. Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT - Statistical Journal*, **6–1**, 3350.

Genton, M.G. (Ed.), 2004. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality.* Boca Raton, FL: Chapman & Hall/CRC.

González-Farías, G., J. Domínguez-Molina and A. Gupta, 2004. Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, **126**, 521–534.

Hosking, J.R.M., J.R. Wallis and E.F. Wood, 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, **27**, 251-261.

Pewsey A. and G. González-Farías, 2007. Special issue preface. *Communications in Statistics - Theory and Methods*, **36–9**.

5 Appendix

Proof of Lemma 1. According to González-Farías *et al.* (2004), if $\mathbf{Y}$ follows a $\mathrm{CSN}_{k,l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$, then $\mathbf{Y} + \mathbf{c}$ follows a

$$\mathrm{CSN}_{k,l}(\boldsymbol{\mu} + \mathbf{c}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta}) \tag{15}$$

for any real vector $\mathbf{c}$ of length k. This implies the following equalities

$$\begin{aligned}
\frac{\Phi_l^*(\mathbf{0}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{0}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})} &= \mathbb{E}(\mathrm{CSN}_{k,l}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})), \text{ from (2)}, \\
&= \mathbb{E}(\mathrm{CSN}_{k,l}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta}) + \boldsymbol{\mu}) - \boldsymbol{\mu}, \\
&= \mathbb{E}(\mathrm{CSN}_{k,l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})) - \boldsymbol{\mu}, \text{, because of (15)}, \\
&= \left[\frac{\Phi_l^*(\mathbf{D}^t\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{D}^t\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\nu}\boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})} + \boldsymbol{\mu}\right] - \boldsymbol{\mu}, \text{ from (2)}, \\
&= \frac{\Phi_l^*(\mathbf{D}^t\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}{\Phi_l(\mathbf{D}^t\boldsymbol{\mu}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}^t\boldsymbol{\Sigma}\mathbf{D})}.
\end{aligned}$$

With a similar argument, this time using (4), Equation (7) can be derived.

Proof of Proposition 2 . We write that

$$\begin{aligned}
\mathbb{E}(h(\mathbf{Y})\Phi_k^r(\mathbf{Y}); \mathbf{0}, \mathbf{I}_k)) &= c_l \int \dots \int h(\mathbf{y})\ \Phi_k^r(\mathbf{y}, \mathbf{0}, \mathbf{I}_k)\ \phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\ \Phi_l(\mathbf{D}^t(\mathbf{y} - \boldsymbol{\mu}); \mathbf{0}, \boldsymbol{\Delta}) d\mathbf{y}, \\
&= c_l \int \dots \int h(\mathbf{y})\ \phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\ \left[\Phi_{rk}(\mathbf{D}_*^t(\mathbf{y} - \boldsymbol{\mu}); -\boldsymbol{\mu}, \mathbf{I}_{rk})\Phi_l(\mathbf{D}^t(\mathbf{y} - \boldsymbol{\mu}); \mathbf{0}, \boldsymbol{\Delta})\right] d\mathbf{y}, \\
&= c_l \int \dots \int h(\mathbf{y})\ \phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\ \left[\Phi_{rk+l}(\mathbf{D}_+^t(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}_+, \boldsymbol{\Delta}_+)\right] d\mathbf{y}.
\end{aligned}$$

To complete the computation of $\mathbb{E}(h(\mathbf{Y})\Phi_k(\mathbf{Y} - \boldsymbol{\mu}; \mathbf{0}, \mathbf{I}_k))$, we remark that the density of $\mathbf{Y}_+$ follows a $CSN_{k,rk+l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}_+, \nu_+, \boldsymbol{\Delta}_+)$ defined in Proposition 2 is

$$f_{k,rk+l}(\mathbf{y}) = \frac{1}{\Phi_{rk+l}(\mathbf{0}; \mathbf{0}, \boldsymbol{\Delta}_+ + \mathbf{D}_+^t\boldsymbol{\Sigma}\mathbf{D}_+)}\phi_k(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\ \Phi_{rk+l}(\mathbf{D}_+^t(\mathbf{y} - \boldsymbol{\mu}); \mathbf{0}, \boldsymbol{\Delta}_+).$$

The expected result follows.