# ALPHA-SKEW-NORMAL DISTRIBUTION [*]

*DAVID ELAL-OLIVERO*
*UNIVERSIDAD DE ATACAMA, CHILE*

## Abstract

*The main object of this paper is to introduce an alternative form of generate asymmetry in the normal distribution that allows to fit unimodal and bimodal data sets. Basic properties of this new distribution, such as stochastic representation, moments, maximum likelihood and the singularity of the Fisher information matrix are studied. The methodology developed is illustrated with a real application.*

*Key Words :* *Asymmetry; Bimodality; Skew-normal distribution; Stochastic representation; Maximum likelihood estimation; Singular information matrix*

---

## 1. Introduction

The univariate skew-normal (SN) distribution has been studied by Azzalini (1985, 1986), Henze (1986), Pewsey (2000), and others, and synthetized in the book edited by Genton (2004). In the univariate case, this type of distributions are denominated skew-symmetric, and they have been used in multiple applications to study the asymmetric behavior of empirical data sets coming from different research areas. Thus, in the last years, different families of skew-symmetric distributions have been generated, some of which are related with the SN model introduced by Azzalini (1985). Examples of these families are those considered by Arellano-Valle *et al.* (2004) and Gómez *et al.* (2006). Most of those classes include the normal distribution as a particular case and satisfy similar properties as the normal family. Mudholkar and Hutson (2000) proposed an asymmetric normal family of distributions with a different structure of the SN class considered by Azzalini (1985), which is called epsilon-skew-normal (ESN) and is denoted $\{ESN(\epsilon) : |\epsilon| < 1\}$ where $\epsilon$ represents the asymmetry parameter, so that $ESN(0)$ corresponds to the normal distribution. On the other hand, there has been a number of works exploring bimodality arising from skew distributions. See Ma and Genton (2004), Azzalini and Capitonio (2003), Arellano-Valle *et al.* (2005) and Elal-Olivero *et al.* (2009).

The aim of this article is to introduce a new family of distributions that is flexible enough to support both unimodal and bimodal shape. Many data sets arising in practice can be adequately modeled in this way and so the proposal plays a unifying role in this context.
This new family is called alpha-skew-normal (ASN) and is denoted by $\{ASN(\alpha) : \alpha \in R\}$ where $\alpha$ represents the asymmetric parameter with effect of uni-bimodality, so that $ASN(0)$ corresponds to the normal distribution.

The rest of this article is organized as follow. Section 2 presents the new family, and develops its main property. In particular, we show how uni-bimodality shape are obtained. Section 3, shows important probabilistic properties of the new family distribution, including the stochastic representation. Section 4 considers maximum likelihood estimation, with emphasis on the derivation of the Fisher information matrix. Real data applications are reported in section 5. In what follows, density and accumulative function of the standard normal distribution will be expressed as $\phi(\cdot)$ and $\Phi(\cdot)$ respectively.

## 2. Alpha-Skew-Normal Distribution

**Definition 2.1.** *If a random variable $Y$ has density function,*

(2.1)                         $$f(y) = y^2 \phi(y), \quad y \in R,$$

*then we say that $Y$ is a* bimodal-normal *random variable. We denote this as $Y \sim BN$.*

**Definition 2.2.** *If a random variable $X$ has density function,*

(2.2)                     $$f(x) = \frac{(1 - \alpha x)^2 + 1}{2 + \alpha^2} \phi(x), \quad x \in R,$$

*where $\alpha \in R$, then we say that $X$ is a* alpha-skew-normal *random variable with parameter $\alpha$. We denote this as $X \sim ASN(\alpha)$.*

If $X \sim ASN(\alpha)$, the following properties are deduced immediately from the definition:

(i) If $\alpha = 0$, then $X \sim N(0, 1)$.

(ii) If $\alpha \to \pm\infty$, then $X \xrightarrow{d} BN$.

(iii) $-X \sim ASN(-\alpha)$.

**Proposition 2.1.** *The density function (2.2) has at most two modes.*

**Proof 2.1.** *Differentiating (2.2), we obtain*

$$
\begin{aligned}
f'(x) &= \frac{1}{2 + \alpha^2} \left[ (-2\alpha + 2\alpha^2 x)\phi(x) + (2 - 2\alpha x + \alpha^2 x^2)(-x\phi(x)) \right] \\
&= \frac{\phi(x)}{2 + \alpha^2} \left[ (-\alpha^2 x^3 + 2\alpha x^2 - (2 - 2\alpha^2)x - 2\alpha \right],
\end{aligned}
$$

*thus, $f'(x)$ has at most three zeros, hence $f(x)$ has at most two modes.*

**Remark 2.1.** *Considering the cubic term inside the square brackets in $f'(x)$, in proof above, and applying a numerical method, we conclude that the transition from bimodality to unimodality of $f(x)$ occurs around $\alpha = \pm 1.34$.*

Figure 1 : Plots of the skewed density $ASN(\alpha)$ for different choice of $\alpha$.

**Proposition 2.2.** *Let* $X \sim ASN(\alpha)$, *then*

$$(2.3) \qquad E[X^{2k}] = \frac{2 + \alpha^2(2k+1)}{2 + \alpha^2} \prod_{j=1}^{k}(2j-1), \ \text{for } k = 1, 2, 3, \ldots,$$

$$(2.4) \qquad E[X^{2k-1}] = -\frac{2\alpha}{2 + \alpha^2} \prod_{j=1}^{k}(2j-1), \ \text{for } k = 1, 2, 3, \ldots.$$

**Proof 2.2.** *knowing that*

$$E[Z^{2k}] = \prod_{j=1}^{k}(2j-1) \text{ and } E[Z^{2k-1}] = 0, \text{ for } k = 1,2,3,\ldots, \text{ when } Z \sim N(0,1),$$

*we obtain*

$$
\begin{aligned}
E[X^{2k}] &= \frac{1}{2+\alpha^2}\left(2E\left[Z^{2k}\right] - 2\alpha E\left[Z^{2k+1}\right] + \alpha^2 E\left[Z^{2(k+1)}\right]\right) \\
&= \frac{1}{2+\alpha^2}\left(2\prod_{j=1}^{k}(2j-1) + \alpha^2\prod_{j=1}^{k+1}(2j-1)]\right) \\
&= \frac{2+\alpha^2(2k+1)}{2+\alpha^2}\prod_{j=1}^{k}(2j-1),
\end{aligned}
$$

*on the other hand*

$$
\begin{aligned}
E[X^{2k-1}] &= \frac{1}{2+\alpha^2}\left(2E\left[Z^{2k-1}\right] - 2\alpha E\left[Z^{2k}\right] + \alpha^2 E\left[Z^{2k+1}\right]\right) \\
&= -\frac{2\alpha}{2+\alpha^2}\prod_{j=1}^{k}(2j-1).
\end{aligned}
$$

**Proposition 2.3.** *Let $\gamma_1$ and $\gamma_2$ the indices of skewness and kurtosis respectively of model $ASN(\alpha)$, then*

(2.5)          $-0.811 < \gamma_1 < 0.811 \quad and \quad -1.\overline{3} < \gamma_2 < 0.7489.$

**Proof 2.3.** *Let $X \sim ASN(\alpha)$. Using (2.3) and (2.4), we obtain*

$$E[X] = \mu_1 = -\frac{2\alpha}{2+\alpha^2}, \qquad E[X^2] = \mu_2 = \frac{2+3\alpha^2}{2+\alpha^2},$$

$$E[X^3] = \mu_3 = -\frac{6\alpha}{2+\alpha^2} \quad and \quad E[X^4] = \mu_4 = \frac{6+15\alpha^2}{2+\alpha^2}.$$

*The definition of $\gamma_1$ and $\gamma_2$, leads to*

$$\gamma_1 = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{\frac{3}{2}}} = \frac{12\alpha^5 + 8\alpha^3}{(3\alpha^4 + 4\alpha^2 + 4)^{\frac{3}{2}}},$$

*and*

$$\gamma_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} - 3 = \frac{15\alpha^8 + 120\alpha^6 + 168\alpha^4 + 96\alpha^2 + 48}{(3\alpha^4 + 4\alpha^2 + 4)^2} - 3,$$

*applying a numerical method, we obtain (2.5)*

**Proposition 2.4.** *If $F(t)$ is the distribution function of $X \sim ASN(\alpha)$, then*

$$F(t) = \Phi(t) + \alpha \left( \frac{2 - \alpha t}{2 + \alpha^2} \right) \phi(t).$$

**Proof 2.4.**

$$
\begin{aligned}
F(t) &= \frac{1}{2 + \alpha^2} \int_{-\infty}^{t} (2 - 2\alpha x + \alpha^2 x^2)\phi(x)dx \\
&= \frac{1}{2 + \alpha^2} \left[ 2\int_{-\infty}^{t} \phi(x)dx - 2\alpha \int_{-\infty}^{t} x\phi(x)dx + \alpha^2 \int_{-\infty}^{t} x^2\phi(x)dx \right] \\
&= \frac{1}{2 + \alpha^2} \left[ 2\Phi(t) - 2\alpha(-\phi(t)) + \alpha^2(-t\phi(t) + \Phi(t)) \right] \\
&= \Phi(t) + \alpha \left( \frac{2 - \alpha t}{2 + \alpha^2} \right) \phi(t).
\end{aligned}
$$

**Proposition 2.5.** *If $M_X(t)$ is the moment generating function of $X \sim ASN(\alpha)$, then*

$$M_X(t) = \left[ 1 - \alpha t \left( \frac{2 - \alpha t}{2 + \alpha^2} \right) \right] \exp\left( \frac{t^2}{2} \right).$$

**Proof 2.5 :**

$$
\begin{aligned}
M_x(t) &= \frac{1}{2+\alpha^2} \int_{-\infty}^{\infty} \exp(tx)(2 - 2\alpha x + \alpha^2 x^2)\phi(x)dx \\
&= \frac{1}{2+\alpha^2} \left[ 2\int_{-\infty}^{\infty} \exp(tx)\phi(x)dx - 2\alpha \int_{-\infty}^{\infty} \exp(tx)x\phi(x)dx \right. \\
&\quad \left. + \alpha^2 \int_{-\infty}^{\infty} \exp(tx)x^2\phi(x)dx \right] \\
&= \frac{1}{2+\alpha^2} \left[ 2\exp\left(\frac{t^2}{2}\right) - 2\alpha \left( t \exp\left(\frac{t^2}{2}\right) \right) + \alpha^2 \left( t^2 \exp\left(\frac{t^2}{2}\right) + \exp\left(\frac{t^2}{2}\right) \right) \right] \\
&= \left[ 1 - \alpha t \left( \frac{2 - \alpha t}{2 + \alpha^2} \right) \right] \exp\left( \frac{t^2}{2} \right).
\end{aligned}
$$

## 3. Stochastic Representation

The next proposition shows a stochastic representation for the model $BN$.

**Proposition 3.1.** *Let $T$, $V$ be two independent random variables, where $T \sim \chi^2_{(3)}$ and $V$ is such that $P[V = 1] = P[V = -1] = \frac{1}{2}$. If $Y = \sqrt{T}V$, then*

$$Y \sim BN.$$

**Proof 3.1.** *If $h(y)$ is the density function of $H = \sqrt{T}$, then*

$$h(y) = 2y^2 \phi(y), \qquad y \geq 0,$$

*on the other hand, if $Y = HV$ is easy to show that the density function of $Y$ is*

$$f(y) = y^2 \phi(y), \quad y \in R,$$

*which proves the required result.*

**Remark 3.1.** *From the above proof we can deduce that:*

$$Y \sim NB \Leftrightarrow Y \sim \pm \chi_3$$

*where $\pm \chi_3$ is the square root of the $\chi_3^2$ random variable, with a random sign attached.*

**Proposition 3.2.** *If $M_Y(t)$ is the moment generating function of $Y \sim BN$, then*

$$M_Y(t) = E[\exp(tY)] = (1 + t^2) \exp\left(\frac{t^2}{2}\right).$$

**Proof 3.2.**

$$
\begin{aligned}
M_Y(t) &= \int_{-\infty}^{\infty} \exp(ty)\, y^2 \phi(y) dy, \\
&= t^2 \exp\left(\frac{t^2}{2}\right) + \exp\left(\frac{t^2}{2}\right) \\
&= (1 + t^2) \exp\left(\frac{t^2}{2}\right).
\end{aligned}
$$

**Remark 3.2.** *The density function (2.2), of model $ASN(\alpha)$, can be represented as sum of two functions, as shown below*

$$
\begin{aligned}
f(x) &= \left(\frac{(1 - \alpha x)^2 + 1}{2 + \alpha^2}\right) \phi(x) \\
&= \left(\frac{2 + \alpha^2 x^2}{2 + \alpha^2}\right) \phi(x) + \left(\frac{-2\alpha}{2 + \alpha^2}\right) x \phi(x),
\end{aligned}
$$

*where the first summand is a symmetric density function, which is considered in the following definition.*

**Definition 3.1.** *If a random variable $S$ has density function,*

$$(3.1) \qquad f_1(x) = \left( \frac{2 + \alpha^2 x^2}{2 + \alpha^2} \right) \phi(x), \quad x \in R,$$

*where $\alpha \in R$, then we say that $S$ is a* symmetric-component *random variable of the model $ASN(\alpha)$. We denote this as $S \sim SCASN(\alpha)$.*

If $S \sim SCASN(\alpha)$, the following properties are deduced immediately from the definition:

(i) If $\alpha = 0$, then $S \sim N(0,1)$.

(ii) If $\alpha \to \pm\infty$, then $S \xrightarrow{d} BN$.

**Remark 3.3.** *Note that the density function (3.1) is a mixture between a normal density and a bimodal-normal density, as shown below*

$$
\begin{aligned}
f_1(x) &= \left( \frac{2 + \alpha^2 x^2}{2 + \alpha^2} \right) \phi(x) \\
&= \left( \frac{2}{2 + \alpha^2} \right) \phi(x) + \left( \frac{\alpha^2}{2 + \alpha^2} \right) x^2 \phi(x).
\end{aligned}
$$

**Proposition 3.3.** *If $M_S(t)$ is the moment generating function of $S \sim SCASN(\alpha)$, then*

$$M_S(t) = E[\exp(tY)] = \left( 1 + \frac{\alpha^2 t^2}{2 + \alpha^2} \right) \exp\left( \frac{t^2}{2} \right).$$

**Proof 3.3.** *Let $M_Z(t)$ and $M_Y(t)$ moment generating functions of $Z \sim N(0,1)$ and $Y \sim BN$ respectively. Considering the previous remark, we can deduced that*

$$
\begin{aligned}
M_S(t) &= \frac{2}{2 + \alpha^2} M_Z(t) + \frac{\alpha^2}{2 + \alpha^2} M_Y(t) \\
&= \frac{2}{2 + \alpha^2} \exp\left( \frac{t^2}{2} \right) + \frac{\alpha^2}{2 + \alpha^2} (1 + t^2) \exp\left( \frac{t^2}{2} \right) \\
&= \left( 1 + \frac{\alpha^2 t^2}{2 + \alpha^2} \right) \exp\left( \frac{t^2}{2} \right).
\end{aligned}
$$

The next proposition shows a stochastic representation for the model $SCASN(\alpha)$.

**Proposition 3.4.** *Let $Y \sim BN$ and $Z \sim N(0,1)$ independent random variables. If*

$$S = \sqrt{\frac{\alpha^2}{2 + \alpha^2}}\, Y + \sqrt{\frac{2}{2 + \alpha^2}}\, Z,$$

*then*

$$S \sim SCASN(\alpha).$$

**Proof 3.4.** *Let $a = \sqrt{\frac{\alpha^2}{2+\alpha^2}}$ and $b = \sqrt{\frac{2}{2+\alpha^2}}$, then*

$$
\begin{aligned}
M_S(t) = E[\exp(tS)] &= E[\exp[t(aY + bZ)]]\\
&= E[\exp(atY)]E[\exp(btZ)]\\
&= \left(1 + a^2t^2\right)\exp\left(\frac{a^2t^2}{2}\right)\exp\left(\frac{b^2t^2}{2}\right)\\
&= \left(1 + \frac{\alpha^2t^2}{2 + \alpha^2}\right)\exp\left(\frac{(a^2 + b^2)t^2}{2}\right)\\
&= \left(1 + \frac{\alpha^2t^2}{2 + \alpha^2}\right)\exp\left(\frac{t^2}{2}\right),
\end{aligned}
$$

*which is the moment generating function of $SCASN(\alpha)$ model.*

The stochastic representations, of the models $BN$ and $SCASN(\alpha)$, presented in propositions 3.1 and 3.4, make possible the application of "acceptance-rejection" method to generate random numbers for model $ASN(\alpha)$, algorithm that is described in the next proposition

**Proposition 3.5.** *(The "acceptance-rejection" algorithm )*
*Let $f(x)$ density function of $X \sim ASN(\alpha)$ and $f_1(x)$ density function of $S \sim SCASN(\alpha)$, with*

$$M = \sup_x \frac{f(x)}{f_1(x)} = \frac{2 + \sqrt{2}}{2}.$$

To generate a random variable $X \sim ASN(\alpha)$:

a. *Generate $U \sim uniform(0,1)$, $S \sim SCASN(\alpha)$, independent.*

b. If $U < \frac{1}{M}\frac{f(S)}{f_1(S)} = \frac{2[(1+\alpha S)^2+1]}{(2+\sqrt{2})(2+\alpha^2 S^2)}$, set $X = S$; otherwise, return to step *a..*

**Remark 3.4.** *Given that*

$$f(x) = \frac{(1-\alpha x)^2 + 1}{2 + \alpha^2}\phi(x) \quad and \quad f_1(x) = \left(\frac{2 + \alpha^2 x^2}{2 + \alpha^2}\right)\phi(x),$$

*is easy to prove that*

$$M = \sup_x \frac{f(x)}{f_1(x)} = \frac{2 + \sqrt{2}}{2},$$

*and therefore*

$$\frac{1}{M}\frac{f(S)}{f_1(S)} = \frac{2[(1 + \alpha S)^2 + 1]}{(2 + \sqrt{2})(2 + \alpha^2 S^2)},$$

*on the other hand*

$$P\left[U < \frac{1}{M}\frac{f(S)}{f_1(S)}\right] = \frac{1}{M} = \frac{2}{2 + \sqrt{2}},$$

*so the number of trials needed to generate one $X$ is a geometric($\frac{1}{M}$) random variable, and $M = \frac{2+\sqrt{2}}{2} = 1.707$ is the expected number of trials.*

### 3.0.1. Location and Scale

**Definition 3.2.** *Let $X \sim ASN(\alpha)$. The alpha-skew-normal density of location and scale is defined as the distribution of $W = \mu + \sigma X$ for $\mu \in R$ and $\sigma > 0$. The corresponding density function is given by*

(3.2) $$f(x|\theta) = \left(\frac{[1 - \alpha(\frac{x-\mu}{\sigma})]^2 + 1}{\sigma(2 + \alpha^2)}\right)\phi\left(\frac{x - \mu}{\sigma}\right),$$

*where $\theta = (\mu, \sigma, \alpha)$. We denote this as $W \sim ASN(\theta)$.*

The distribution function of (3.2) is given by

$$F(t) = \Phi\left(\frac{t-\mu}{\sigma}\right) + \alpha\left(\frac{2\sigma - \alpha(x-\mu)}{\sigma(2+\alpha^2)}\right)\phi\left(\frac{x-\mu}{\sigma}\right),$$

and the moment of order n by

$$E[W^n] = \sum_{j=0}^{n} \binom{j}{n}\mu^{n-j}\sigma^j E[X^j] \quad where \quad X \sim ASN(\alpha).$$

## 4. Maximum likelihood estimation

This section concerns likelihood inference about the parameter $\theta = (\mu, \sigma, \alpha)$ of the location-scale family defined in (3.2), In particular, the Fisher information matrix of the maximum likelihood estimators (MLEs) of these parameter is obtained.

Let $Z_1, Z_2, ..., Z_n$ be a random sample from the $ASN(\theta)$. Thus, the likelihood function is given by

(4.1)                    $$L(\theta|Z_1, Z_2, ..., Z_n) = \Pi_{i=1}^{n} f(Z_i|\theta),$$

where $f(Z_i|\theta)$ is given by (3.2). The MLE of $\theta$ is obtained maximizing (4.1), for which a numerical algorithm is necessary. In this work, the subroutine *nlminb* of the software S-PLUS is used.

The Fisher information for the parameter $\theta = (\mu, \sigma, \alpha)$ is easily computed, obtaining

$$F = \begin{pmatrix} \frac{4\alpha^2 b_0 - \alpha^2 + 2}{\sigma^2(2+\alpha^2)} & \frac{4\alpha^2 b_1 - 2\alpha}{\sigma^2(2+\alpha^2)} & \frac{2+4\alpha b_1}{\sigma(2+\alpha^2)} \\ \frac{4\alpha^2 b_1 - 2\alpha}{\sigma^2(2+\alpha^2)} & \frac{4\alpha^2 b_2 + 2\alpha^2 + 4}{\sigma^2(2+\alpha^2)} & \frac{4\alpha(b_2-1)}{\sigma(2+\alpha^2)} \\ \frac{2+4\alpha b_1}{\sigma(2+\alpha^2)} & \frac{4\alpha(b_2-1)}{\sigma(2+\alpha^2)} & \frac{4(\alpha^2 b_2 + 2b_2 - \alpha^2)}{(2+\alpha^2)^2} \end{pmatrix},$$

where

$$Z \sim N(0,1) \quad and \quad b_k = b(k) = E\left[Z^k \frac{(1-\alpha Z)^2}{(1-\alpha Z)^2 + 1}\right] \quad k = 0,1,2,$$

which has to be evaluated numerically.

On the other hand, in the normal model when $\alpha = 0$, the Fisher information matrix is

$$F = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & \frac{1}{\sigma} \\ 0 & \frac{2}{\sigma^2} & 0 \\ \frac{1}{\sigma} & 0 & 1 \end{pmatrix}.$$

Note from this matrix that the column corresponding to the parameters $\mu$ and $\alpha$ are linearly dependent, implying that it is a singular matrix. This irregularity is discussed by Azzalini (1985) in the context of the SN model, and afterwards it is studied systematically by Chiogna (2005) in some other context. Di Ciccio and Monti (2004) studied this singularity problem in the context of the skew-exponential power distribution and Salinas *et al.* (2007) generalized it in the context of the extended skew-exponential power distribution. In this latter context, the authors make use of the methodology considered in Rotnitzky *et al.* (2000), so that an adequate reparametrization can be found for which the asymptotic properties of the maximum likelihood estimators remain valid.

Using the same procedure as studied in Chiogna (2005), Di Ciccio and Monti (2004) and Salinas *et al.* (2007), start by supposing that the parameter value is $\theta^* = (\mu^*, \sigma^*, 0)$, that is, data set has been generated by a normal distribution with mean value $\mu^*$ and variance $\sigma^{*2}$. To $\theta = \theta^*$, leading to the score vector $S_\theta(\theta^*, W) = \partial \log L(\theta^*)/\partial \theta = (S_\mu^*, S_\sigma^*, S_\alpha^*)$ is given by

$$S_\theta(\theta^*, W) = \left[ \frac{Z^*}{\sigma^*}, \frac{Z^{*2} - 1}{\sigma^*}, -Z^* \right]$$

where $Z^* = (W - \mu^*)/\sigma^*$. The linear dependence between the components $S_\mu^*$ and $S_\alpha^*$ causes the singularity problem for the information matrix.

The general result in Rotnitzky *et al.* (2000) concerning the maximum likelihood estimator for the parameter vector $\theta = (\theta^1, \theta^2, ..., \theta^q)$ can be applied in this context to establish consistency and asymptotic normality for the MLEs. Rotnitzky *et al.* (2000) derive the asymptotic distributions for the MLEs $\theta = (\theta^1, \theta^2, ..., \theta^q)$ under two conditions: one, the components of the score vector, say $S_{\theta^1}$, is zero for $\theta = \theta^*$ and, secondly, higher order partial derivatives of $S_{\theta^1}$ with respect to $\theta^1$ is possibly zero at this point; however, the first derivative is not a linear combination of the other components of $S_{\theta^2}, S_{\theta^3}, ..., S_{\theta^q}$. Even though the above conditions are not valid for the ASN normal model, it is possible to use the iterative procedure proposed in Rotnitzky *et al.* (2000) to find a reparametrization so that such conditions are satisfied. This procedure has been detailed in Chiogna (2005) and by Di Ciccio and Monti (2004). After extensive algebraic manipulations, it can be shown that such reparametrization is given by $\widetilde{\theta} = \widetilde{\theta}(\theta) = (\widetilde{\mu}, \widetilde{\sigma}, \alpha)$, where

$$\widetilde{\mu} = \mu - \sigma^* \alpha, \qquad \widetilde{\sigma} = \sigma$$

Finally, making use of theorem 3 in Rotnitzky *et al.* (2000), with the above reparametrization, we can conclude that:

1. The MLE of $\theta$ is unique with probability tending to 1, and it is consistent.

2. The likelihood ratio statistic for testing the simple null hypothesis $H_0 : \theta = \theta^*$ converges in distribution to the $\chi_3^2$.

3. The random vector

$$[n^{\frac{1}{2}}(\widehat{\mu} - \mu^* - \sigma^*\widehat{\alpha}), n^{\frac{1}{2}}(\widehat{\sigma} - \sigma^*), n^{\frac{1}{6}}\widehat{\alpha}],$$

   converges under $\theta = \theta^*$ to $(Z_1, Z_2, Z_3^{\frac{1}{3}})$ where $(Z_1, Z_2, Z_3)$ is a normal random vector with mean zero and covariance matrix equals to the inverse of the covariance matrix of the vector

$$\left[\frac{Z}{\sigma^*}, \frac{Z^2 - 1}{\sigma^*}, -\frac{Z^3}{6}\right]^t$$

   which is given by

$$F = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & -\frac{1}{2\sigma} \\ 0 & \frac{2}{\sigma^2} & 0 \\ \frac{1}{-2\sigma} & 0 & \frac{5}{12} \end{pmatrix}^{-1}.$$

## 5. An illustrative application

In this section, we illustrate the use of the estimation procedures described in the previous section. The variable to be considered is the average length of stay for patients who are in hospital for acute care because of problems, hepatobiliary system and pancreas, and die for this cause. The sample, under study, corresponds to 1082 hospitals in 10 states of the United States. For more information see columns 4 in http://lib.stat.cmu.edu/data-expo/ 1997/ascii/p07.dat. Table 1 shows descriptive statistics for the data set.

Table 5.1: Summary statistics of the average length of stay of patients in 1082 hospitals, where $g_1$ and $g_2$ represent the coefficients of asymmetry and kurtosis respectively.

| Data | $n$ | $\bar{x}$ | $s$ | $g_1$ | $g_2$ |
|------|-----|-----------|-----|-------|-------|
|      | 1082 | 5.7593 | 1.614 | 0.6723 | 1.4205 |

The models $N(\mu, \sigma)$, $ASN(\mu, \sigma, \alpha)$ and $SN(\mu, \sigma, \alpha)$ are fitted to the data set using the maximum likelihood approach, where $SN(\mu, \sigma, \alpha)$ is the skew-normal model introduced by Azzalini (1985). Results are reported in table 2 with the estimators standard errors being estimated by using the observed information matrix.

Table 5.2: Estimated parameters and maximized log-likelihood function for models $N(\mu, \sigma)$, $ASN(\mu, \sigma, \alpha)$ and $SN(\mu, \sigma, \alpha)$. Standard errors are reported in parenthesis.

| Parameter estimates | $N$ | $ASN$ | $SN$ |
|---|---|---|---|
| $\mu$ | 5.7593(0.0961) | 6.7578(0.1261) | 4.1192(0.0602) |
| $\sigma$ | 1.614(0.068) | 1.4873(0.0711) | 2.3006(0.0206) |
| $\alpha$ | - | 0.9559(0.1444) | 2.0608(0.1501) |
| log-likelihood | $-2052.824$ | $-2018.41$ | $-2023.653$ |
| AIC | - | 4042.82 | 4053.306 |

The likelihood radio test (LRT) for the hypothesis $H_0 : \alpha = 0$ ($N(\mu, \sigma)$ model) versus $H_1 : \alpha \neq 0$ ($ASN(\mu, \sigma, \alpha)$ model) is such that $-2\log(LRT) = -2(-2023.653 + 2018.41) = 10.486$. Hence, comparing this quantity with the 95% critical value, namely $\chi^2_{(1)} = 3.84$, there is sufficient evidence to reject the null hypothesis, that is, parameter $\alpha$ is significantly different from zero, concluding that $ASN$ gives a better fit to the data than the normal model. Since the models $ASN(\mu, \sigma, \alpha)$ and $SN(\mu, \sigma, \alpha)$ are not nested the Akaike Information Criterion (AIC) has been used for comparison. According to this criterion, the ASN model provides a better fit than the SN model ($4042.82 < 4053.306$). These conclusions are also corroborated by the plots of the fitted densities (using the maximum likelihood estimators), presented in Figure 2.

Figure 2 : The histograms correspond to the average length of stay of patients in 1082 hospitals. The lines represent fitted distributions using maximum likelihood: $ASN(\widehat{\mu}, \widehat{\sigma}, \widehat{\alpha})$ (solid line), $N(\widehat{\mu}, \widehat{\sigma})$ (dotted line in (a)) and $SN(\widehat{\mu}, \widehat{\sigma}, \widehat{\alpha})$ (dotted line in (b)).

# References

[1] Arellano-Valle, R. B., Gómez, H. W. and Quintana, F.A. A New Class of Skew-Normal Distributions. Communications in Statistics : Theory and Methods: **33** (7), pp. 1465-1480, (2004).

[2] Arellano-Valle, R. B., Gómez, H.W. and Quintana, F. A. Statistical inference for a general class of asymmetric distributions. Journal of Statistical Planning and Inference: **128**(2), pp. 427 - 443, (2005).

[3] Azzalini, A. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics: **12**, pp. 171-178, (1985).

[4] Azzalini, A. Further results on a class of distributions which includes the normal ones. Statistica: **46**(2), pp. 199-208, (1986).

[5] Azzalini, A. and Capitonio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society series B: **65**(2), pp. 367 - 389, (2003).

[6] Chiogna, M. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. Statistical Methods and Applications: **14**, pp. 331-341, (2005).

[7] DiCiccio,T. J. and Monti, A. C. Inferencial Aspects of the Skew Exponcial Power Distribution. Journal of the American Statistical Association: **99**(466), pp. 439-450, (2004).

[8] Elal - Olivero, D., Gómez, H. W. and Quintana, F. A. Bayesian modeling using a class of bimodal skew-elliptical distributions. Journal of Statistical Planning and Inference: **139**, pp. 1484-1492, (2009).

[9] Genton, M. G., Ed. Skew-elliptical distributions and their applications : a journey beyond normality. Chapman & Hall/CRC, (2004).

[10] Gómez H. ., Salinas, H. S. and Bolfarine, H. Generalized skew-normal models: Properties and Inference. Statistic: **40**(6), pp. 495-505, (2006).

[11] Henze, N. A probabilistic representation of the skew-normal distribution. Scandinavian Journal of Statistics: **13**, pp. 271-275, (1986).

[12] Ma Y. and Genton M, G. Flexible Class of Skew-Symmetric Distribution. Scandinavian Journal of Statistics: **31**, pp. 459 - 468, (2004).

[13] Mudholkar G.S. and Hutson, A. D. The epsilon-skew-normal distribution for analysing near-normal data. Journal of Statistical Planning and Inference: **83**, pp. 291-309, (2000).

[14] Pewsey, A. Problems of inference for Azzalini's skew-normal distribution. Journal of Applied Statistics,Vol. 27, N$^{o}$:**7**, pp. 859-870, (2000).

[15] Rotnitzky, A. Cox, D. R.,Bottai, M. and Robins, J. Likelihood - based inference with singular information matrix. Bernoulli: **6**(2), pp. 243-284, (2000).

[16] Salinas, H. S., Arellano-Valle, R. B. and Gómez H. W. The extended skew-exponencial power distribution and its derivations. Communications in Statistics : Theory and Methods: **36**(9), pp. 1673-1689, (2007).

**David Elal-Olivero**
Departamento de Matemáticas,
Facultad de Ingeniería,
Universidad de Atacama,
Copiapó
CHILE
E-mail : delal@mat.uda.cl