A NEW CLASS OF SKEW-NORMAL DISTRIBUTIONS

Reinaldo B. Arellano-Valle* Héctor W. Gómez †

Fernando A. Quintana‡

November, 2003

Abstract

We introduce a new family of asymmetric normal distributions that contains Azzalini's skew-normal (SN) distribution as a special case. We study the main properties of this new family, showing in particular that it may be generated via mixtures on the SN asymmetry parameter when the mixing distribution is normal. This property provides a Bayesian interpretation of the new family.

Key Words: Asymmetry, Kurtosis, Skew-Normal, Skew-Curved Normal and Skew-Generalized Normal Distributions.

1 Introduction

The class of distributions we study in this article is closely related to the skew-normal (SN) distribution introduced by Azzalini (1) and defined as follows: a random variable X has

*Departamento de Estadística, Facultad de Matemática, Pontificia Universidad Católica de Chile. E-mail: reivalle@mat.puc.cl

†Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, CHILE. E-mail: hgomez@matematica.uda.cl. Partially supported by grant MECESUP PUC-0103.

‡Departamento de Estadística, Facultad de Matemática, Pontificia Universidad Católica de Chile. E-mail: quintana@mat.puc.cl. Partially supported by grant FONDECYT 1020712. Corresponding author.

skew-normal distribution with asymmetry parameter $\lambda \in \mathbb{R}$ if

$$f(x\,|\lambda) = 2\phi(x)\Phi(\lambda x) \qquad x \in \mathbb{R}, \tag{1.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the $N(0,1)$ density and CDF respectively. We denote this by $X \sim SN(\lambda)$. When $\lambda = 0$ (1.1) becomes the $N(0,1)$ distribution. Henze (2) gives a stochastic representation of this distribution, obtaining explicitly its odd moments. Azzalini (3) also discusses stochastic representations in a more general context. Multivariate extensions of the univariate SN distribution have been studied by several authors. See, for example, Azzalini and Dalla Valle (4), Branco and Dey (5) and Arellano-Valle and Genton (6).

A limitation of the $SN(\lambda)$ model is that for moderate values of λ nearly all the mass accumulates either on the positive or negative numbers, as determined by the sign of λ. In such cases, (1.1) closely resembles the half-normal density, with a nearly linear shape in the side with smaller mass.

To partially mitigate such limitation we introduce a family of distributions that exhibits a better behavior, particularly at the side with smaller mass. This class will be referred to as *skew-generalized normal* (SGN) because it contains the SN distribution (1.1) as a special case.

The rest of this article is organized as follows. Section 2 gives the definition of the SGN class and presents its basic properties. We also consider an important subclass, called the skew-curved normal family, which includes the normal distribution as a special case, but is disjoint of Azzalini's subclass. Section 3 gives important probabilistic properties of the new family and Section 4 is related to the moments of SGN random variables. Section 5 deals with a location-scale extension of the SGN distribution and establishes an interesting invariance property of such distributions under normal sampling. Finally, Section 6 gives an application based on real data.

2 The Skew-Generalized Normal Distribution

In this section we introduce a new class of skew normal distributions, which generalizes (1.1) while preserving most of its properties.

Definition 1 *We say that a random variable X has the skew-generalized normal distribution if its density is given by*

$$f(x\,|\lambda_1,\lambda_2) = 2\phi(x)\Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}\right), \qquad x \in \mathbb{R}, \tag{2.1}$$

where $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \geq 0$. We denote this by $X \sim SGN(\lambda_1,\lambda_2)$. The resulting distribution for the special case $\lambda_2 = \lambda_1^2$ is called skew-curved normal (SCN) and is denoted by $X \sim SCN(\lambda_1)$.

We note that (2.1) is indeed a density, because as shown by Ellison (7), if $Z \sim N(\mu,\sigma^2)$ then

$$E\left[\Phi\left(Z\right)\right] = \Phi\left(\frac{\mu}{\sqrt{1+\sigma^2}}\right). \tag{2.2}$$

Figure 1 illustrates several of the possible shapes obtained from (2.1) under various choices of (λ_1,λ_2).

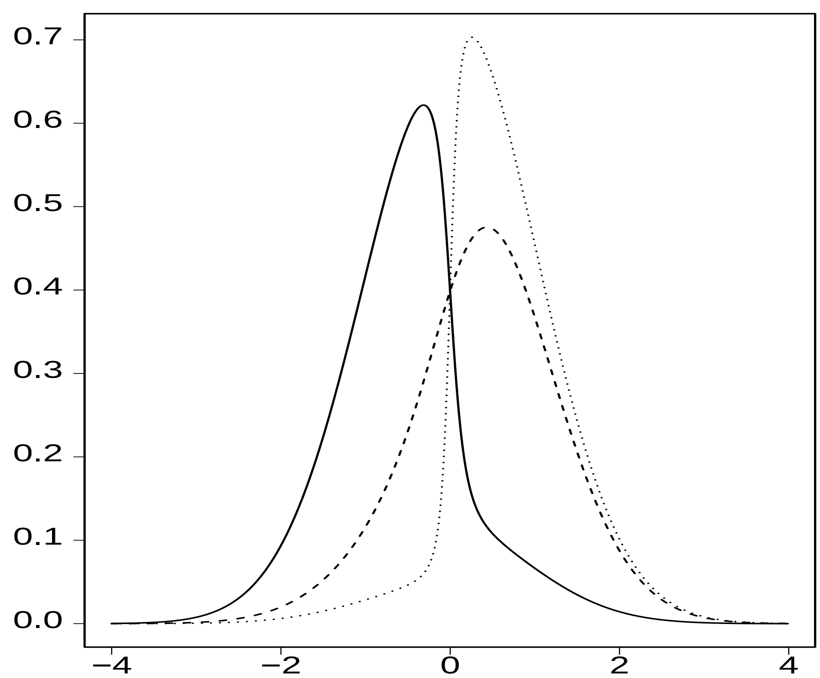

Figure 1: *Examples of the $SGN(\lambda_1,\lambda_2)$ density for $(\lambda_1,\lambda_2) = (-5,20)$ (solid line), $(\lambda_1,\lambda_2) = (1,1)$ (dashed line) and $(\lambda_1,\lambda_2) = (10,40)$ (dotted line).*

Figure 2 compares the $SCN(\lambda_1)$ and the $SN(\lambda_1)$ densities for various choices of λ_1.

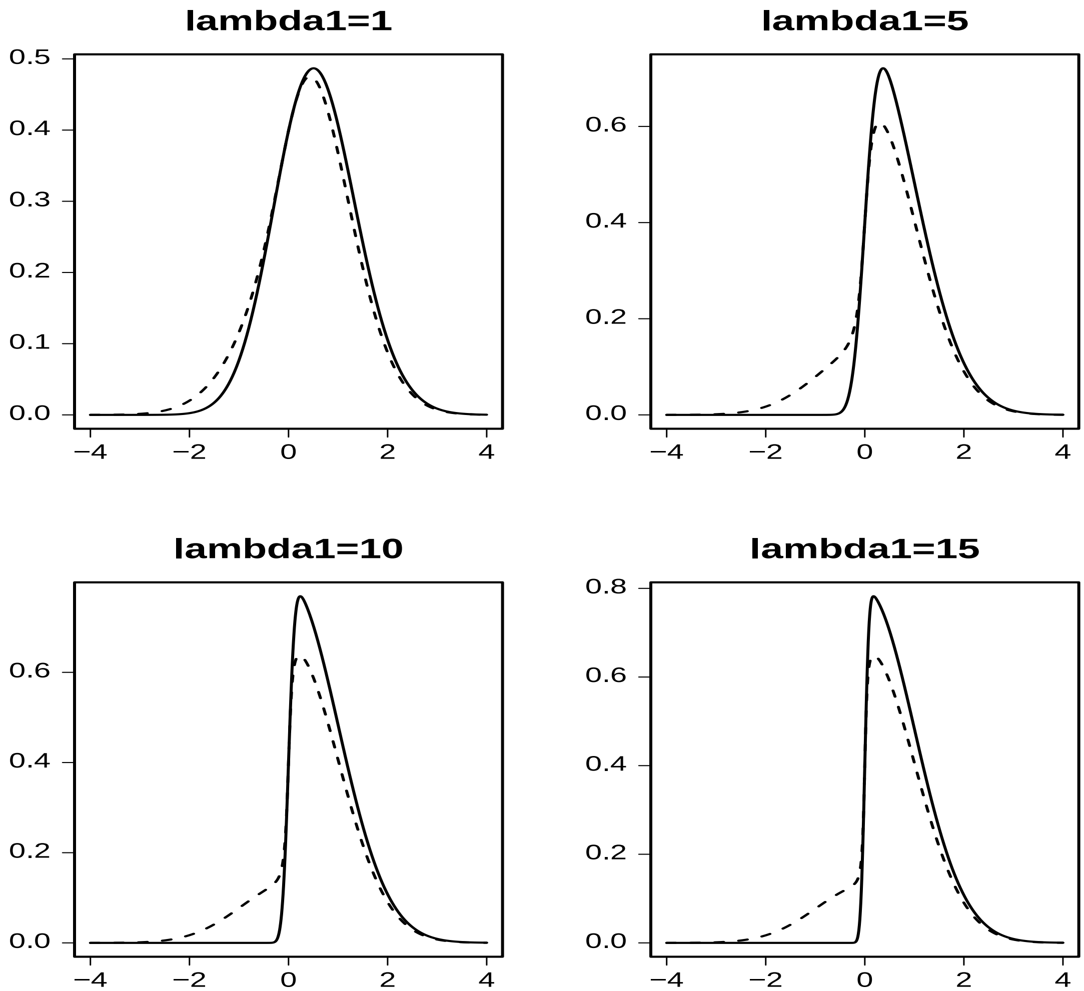

Figure 2: *The $SCN(\lambda_1)$ (dotted line) and $SN(\lambda_1)$ (solid line) densities for different values of λ_1.*

Some basic properties of the $\{SGN(\lambda_1, \lambda_2) : \lambda_1 \in \mathbb{R}; \lambda_2 \geq 0\}$ family arise directly from (2.1).

Proposition 1

1. *$f(x|\lambda_1 = 0, \lambda_2) = \phi(x)$ for all $x \in \mathbb{R}$ and $\lambda_2 \geq 0$, i.e. $SGN(\lambda_1 = 0, \lambda_2) = N(0, 1)$.*

2. *$f(x|\lambda_1, \lambda_2 = 0) = 2\phi(x)\Phi(\lambda_1 x)$ for all $x \in \mathbb{R}$, i.e. $SGN(\lambda_1, \lambda_2 = 0) = SN(\lambda_1)$.*

3. *If* $X \sim SGN(\lambda_1, \lambda_2)$ *then* $-X \sim SGN(-\lambda_1, \lambda_2)$.

4. $f(x|-\lambda_1, \lambda_2) + f(x\,|\lambda_1, \lambda_2) = 2\phi(x)$ *for all* $x \in \mathbb{R}$.

5. $\lim_{\lambda_1 \to +\infty} f(x|\lambda_1, \lambda_2) = 2\phi(x) I\{x \geq 0\}$ *and* $\lim_{\lambda_1 \to -\infty} f(x|\lambda_1, \lambda_2) = 2\phi(x) I\{x \leq 0\}$ *for all* λ_2.

6. $\lim_{\lambda_1 \to \pm\infty} f(x|\lambda_1, \lambda_1^2) = 2\phi(x)\Phi(\pm\,\mathrm{sgn}(x))$.

Remark 1 *Item 6 of Proposition 1 gives a hint at the basic motivation for considering the special case* $SCN(\lambda_1)$*. When* λ_1 *grows, the resulting SCN density has heavier tail to the left than the SN distribution. In fact, the asymptotic distribution obtained from the* $SCN(\lambda_1)$ *when* $\lambda_1 \to \infty$ *is not half-normal. Indeed, this corresponds to a distribution concentrating most of its mass at one side of* 0*, but still leaving positive probability at the other side.*

From (2.1) the CDF of the $SGN(\lambda_1, \lambda_2)$ distribution can be expressed as

$$F\left(x\,|\lambda_1, \lambda_2\right) = \int_{-\infty}^{x} 2\phi(t)\Phi\left(\frac{\lambda_1 t}{\sqrt{1+\lambda_2 t^2}}\right) dt. \tag{2.3}$$

This immediately leads to the following properties:

Proposition 2

1. $F\left(x|\lambda_1 = 0, \lambda_2\right) = \Phi(x)$ *for all* $x \in \mathbb{R}$ *and* $\lambda_2 \geq 0$.

2. $F\left(-x|\lambda_1, \lambda_2\right) = 1 - F\left(x|-\lambda_1, \lambda_2\right)$.

3. $$\lim_{\lambda_1 \to +\infty} F\left(x|\lambda_1, \lambda_1^2\right) = \begin{cases} 2\left(1-\Phi\left(1\right)\right)\Phi\left(x\right) & \textit{if } x < 0 \\ 1 - 2\Phi\left(1\right)\left(1-\Phi\left(x\right)\right) & \textit{otherwise.} \end{cases}$$

4. $$\lim_{\lambda_1 \to -\infty} F\left(x|\lambda_1, \lambda_1^2\right) = \begin{cases} 2\Phi\left(1\right)\Phi\left(x\right) & \textit{if } x < 0 \\ 2\left[\Phi\left(1\right) + \left(1-\Phi\left(1\right)\right)\Phi\left(x\right) - 1/2\right] & \textit{otherwise.} \end{cases}$$

3 Some Important Properties

We now derive the main properties of the SGN distribution. The first result shows that, just as Azzalini's SN, the distribution of the absolute value of a SGN random variable is half-normal. In particular, this implies that all the moments of the SGN distribution are finite, and its even moments coincide with those of the standard normal distribution.

Proposition 3 *Let* $X \sim SGN(\lambda_1, \lambda_2)$ *and* $Y \sim N(0,1)$. *Then* $|X|$ *and* $|Y|$ *are identically distributed, i.e.* $|X| \stackrel{d}{=} |Y| \sim HN(0,1)$, *where* $HN(0,1)$ *denotes the standard half-normal distribution.*

Proof: It is well known that $Z = |Y|$ has density $2\phi(z)I\{z > 0\}$. On the other hand, by (2.1) the density of $W = |X|$ is

$$\begin{aligned} f_W(w) &= f_X(w) + f_X(-w) \\ &= 2\phi(w)\Phi\left(\frac{\lambda_1 w}{\sqrt{1+\lambda_2 w^2}}\right) + 2\phi(-w)\Phi\left(-\frac{\lambda_1 w}{\sqrt{1+\lambda_2 w^2}}\right) \\ &= 2\phi(w) \qquad \text{for } w > 0, \end{aligned}$$

and since $f_W(w) = 0$ for $w \leq 0$, we have $f_W(w) = f_Z(w)$ for all $w \in \mathbb{R}$. $\square$

The following result is a direct consequence of Proposition 3:

Proposition 4 *Let* $X \sim N(0,1)$, $Y \sim SGN(\lambda_1, \lambda_2)$ *and let* h *be an even function. Then* $h(X)$ *and* $h(Y)$ *are identically distributed. In particular,* $Y^2 \sim \chi^2(1)$.

An interesting interpretation of the SGN distribution is given next.

Proposition 5 *If* $X|Y = y \sim SN(y)$ *and* $Y \sim N(\lambda_1, \lambda_2)$ *then* $X \sim SGN(\lambda_1, \lambda_2)$.

Proof:

$$\begin{aligned} f(x|\lambda_1, \lambda_2) &= \frac{2}{\sqrt{\lambda_2}} \int_{-\infty}^{\infty} \phi(x)\Phi(xy)\phi\left(\frac{y-\lambda_1}{\sqrt{\lambda_2}}\right) dy = \int_{-\infty}^{\infty} 2\phi(x)\Phi(\sqrt{\lambda_2}xz + \lambda_1 x)\phi(z)\, dz \\ &= 2\phi(x)E\left[\Phi(\sqrt{\lambda_2}\, xZ + \lambda_1 x)\right]. \end{aligned}$$

The result now follows by applying (2.2) with $Z \sim N(0,1)$, and the definition of conditional density. $\square$

Proposition 5 states that the skew-generalized (curved) normal distribution can be represented as a mixture of the skew-normal on the asymmetry parameter, using a (curved) normal distribution as the mixing distribution. This stochastic representation is also quite useful for drawing samples from the $SGN(\lambda_1, \lambda_2)$ distribution via the method of composition (see, e.g. Tanner (8)) and the *Sn* package of Azzalini, available for downloaded at *http://azzalini.stat.unipd.it*. Alternatively, (2.1) is easily shown to be a log-concave density, which can be also used to draw samples via the Gilks and Wild (9) algorithm.

4 Moments

Taking $h(x) = x^k$ for k even in Proposition 4 we conclude that the even moments of the skew-generalized and standard normal distributions are identical. This also implies the existence of the odd moments, but there is no explicit expression for them. Nevertheless, an implicit formula can be derived as follows. Let

$$a_k(\lambda_1, \lambda_2) = \int_0^\infty \frac{u^k}{\sqrt{2\pi}} e^{-u/2}\Phi\left(\frac{\lambda_1\sqrt{u}}{\sqrt{1+\lambda_2 u}}\right) du. \tag{4.1}$$

Note that for all $\lambda_2 \geq 0$,

$$a_k(0, \lambda_2) = a_k(0, 0) = 2^k\Gamma(k+1)/\sqrt{2\pi}. \tag{4.2}$$

Proposition 6 *If $X \sim SGN(\lambda_1, \lambda_2)$ then for $k = 0, 1, 2, ...$ we have*

$$E\left(X^{2k+1}\right) = 2\{a_k(\lambda_1, \lambda_2) - a_k(0, 0)\}. \tag{4.3}$$

Proof:

$$\begin{aligned} E\left(X^{2k+1}\right) &= 2\int_0^\infty 2x^{2k+1}\phi(x)\Phi\left(\frac{\lambda_1 x}{\sqrt{1+\lambda_2 x^2}}\right) dx - 2\int_0^\infty x^{2k+1}\phi(x)\,dx \\ &= 2a_k(\lambda_1, \lambda_2) - 2^{k+1}\Gamma(1+k)/\sqrt{2\pi}. \qquad \square \end{aligned}$$

In deriving further properties of moments of the SGN distribution the next result turns out to be useful:

Lemma 1 *For any $\lambda_1 > 0$, $\lambda_2 > 0$ and $k = 0, 1, 2, \ldots$,*

$$a_k(0, 0) \leq a_k(\lambda_1, \lambda_2) \leq 2a_k(0, 0)\Phi\left(\frac{\lambda_1}{\sqrt{\lambda_2}}\right) \tag{4.4}$$

and

$$a_k(0, 0) \leq a_k(\lambda_1, \lambda_2) \leq a_k(\lambda_1, 0) \leq 2a_k(0, 0). \tag{4.5}$$

Proof: Since Φ is an increasing function, we have for any $\lambda_1 > 0$ and $u > 0$ that

$$\frac{1}{2} \leq \Phi\left(\frac{\lambda_1\sqrt{u}}{\sqrt{1+\lambda_2 u}}\right) = \Phi\left(\frac{\lambda_1}{\sqrt{\lambda_2}}\frac{\sqrt{\lambda_2 u}}{\sqrt{1+\lambda_2 u}}\right) \leq \Phi\left(\frac{\lambda_1}{\sqrt{\lambda_2}}\right)$$

and the proof of (4.4) follows by combining this last set of inequalities with (4.1) and using the fact that

$$\int_0^\infty \frac{u^{(k+1)-1}}{2^{k+1}\Gamma(k+1)} e^{-u/2}\, du = 1.$$

The argument to show (4.5) is very similar, but now using

$$\frac{1}{2} \le \Phi\left(\frac{\lambda_1\sqrt{u}}{\sqrt{1+\lambda_2 u}}\right) \le \Phi(\lambda_1\sqrt{u}). \qquad \square$$

It can be easily seen that the inequalities in (4.4) and (4.5) are reversed when $\lambda_1 < 0$. Therefore, (4.4) and (4.3) imply that $E(X) > 0$ if and only if $\lambda_1 > 0$. This property is generalized next.

Proposition 7 *If $X \sim SGN(\lambda_1, \lambda_2)$ and $Y \sim SN(\lambda_1)$ then*

$$|E(X^r)| \le |E(Y^r)| \qquad \textit{for all } \lambda_1 \in \mathbb{R},\ \lambda_2 > 0 \textit{ and } r = 1, 2, \ldots$$

Proof: The equality is trivially true for all r when $\lambda_1 = 0$. When r is even it is easily seen from Proposition 4 that $E(X^r) = E(Y^r)$. When r is odd, let us first consider the case where $\lambda_1 > 0$. By (4.3)

$$E(X^r) = 2\{a_{2(r-1)}(\lambda_1, \lambda_2) - a_{2(r-1)}(0,0)\} \quad \text{and} \quad E(Y^r) = 2\{a_{2(r-1)}(\lambda_1, 0) - a_{2(r-1)}(0,0)\}$$

and the result follows from (4.5). Finally, the case $\lambda_1 < 0$ is handled in a similar way. $\square$

Remark 2 *As an immediate consequence of Proposition 7 we have that if $X \sim SGN(\lambda_1, \lambda_2)$ and $Y \sim SN(\lambda_1)$ then $Var(Y) \le Var(X) \le 1$ for all $\lambda_1 \in \mathbb{R}$ and $\lambda_2 > 0$. For the special case $X \sim SCN(\lambda_1)$ the skewness and kurtosis coefficients $\sqrt{\beta_1}$ and β_2 verify*

$$\sqrt{\beta_1} = 0.5(4-\pi)\left[\frac{E^2(X)}{Var(X)}\right]^{3/2}, \qquad \beta_2 = 2(\pi - 3)\left[\frac{E^2(X)}{Var(X)}\right]^2. \tag{4.6}$$

Considering all the possible values for λ_1 we get

$$-0.376 \le \sqrt{\beta_1} \le 0.376 \qquad \textit{and} \qquad 3.0000 \le \beta_2 \le 4.332. \tag{4.7}$$

If $X \sim SN(\lambda)$ the corresponding intervals are shown to be

$$-0.9953 \le \sqrt{\beta_1} \le 0.9953 \qquad \textit{and} \qquad 3.0000 \le \beta_2 \le 3.8692. \tag{4.8}$$

Thus, the skew-curved normal model provides a wider range of kurtosis but less skewness than the skew-normal model. On the other hand, we find that the SGN family provides identical skewness range as the SN case, while allowing for kurtosis ranging over a slightly wider interval than the SCN distribution (data not shown). This is attained when λ_2 takes on values somewhat larger than λ_1^2. However, for large values of λ_1 the skewness and kurtosis of the $SGN(\lambda_1, \lambda_2)$ and $SN(\lambda_1)$ distributions exhibit nearly identical behavior.

The moment generating function (MGF) of $X \sim SGN(\lambda_1, \lambda_2)$ can be easily obtained as

$$M_X(t) = 2e^{t^2/2} E\left[\Phi\left(\frac{\lambda_1 (Y+t)}{\sqrt{1+\lambda_2 (Y+t)^2}}\right)\right], \tag{4.9}$$

where $Y \sim N(0,1)$. An alternative representation is derived next.

Proposition 8 *Let $W \sim N(a, b^2)$ be independent of $Z \sim N(c, d^2)$. Then*

$$E\left[\Phi\left(\frac{cW}{\sqrt{1+d^2W^2}}\right)\right] = E\left[\Phi\left(WZ\right)\right].$$

In particular, if U and V are i.i.d. $N(0,1)$ then $E\left[\Phi\left(UV\right)\right] = 1/2$.

Proof:

$$\begin{aligned} E\left[\Phi\left(WZ\right)\right] &= \frac{1}{bd}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{w} z\phi(zu)\phi\left(\frac{z-c}{d}\right)\phi\left(\frac{w-a}{b}\right) du\, dz\, dw \\ &= \frac{1}{b}\int_{-\infty}^{\infty}\int_{-\infty}^{\frac{cw}{\sqrt{1+d^2w^2}}} \phi(x)\phi\left(\frac{w-a}{b}\right) dx\, dw \\ &= E\left[\Phi\left(\frac{cW}{\sqrt{1+d^2W^2}}\right)\right]. \qquad \square \end{aligned}$$

Proposition 9 *Let $X \sim SGN(\lambda_1, \lambda_2)$. Then*

$$M_X(t) = 2e^{t^2/2} E\left[\Phi\left(WZ\right)\right],$$

where $W \sim N(t, 1)$ is independent of $Z \sim N\left(\lambda_1, \lambda_2\right)$.

Proof: Let $Y \sim N(0,1)$ and $W = Y + t$. Using (4.9) we obtain

$$M_X(t) = 2e^{t^2/2} E\left[\Phi\left(\frac{\lambda_1 W}{\sqrt{1+\lambda_2 W^2}}\right)\right],$$

and the result follows from Proposition 8. $\square$

5 Location-Scale Extension

The same basic principle used for the definition of (2.1) can be extended to a location-scale family.

Definition 2 *The location-scale skew-generalized normal distribution is defined as that of* $Z = \mu + \sigma X$*, where* $X \sim SGN(\lambda_1, \lambda_2)$*,* $\mu \in \mathbb{R}$ *and* $\sigma > 0$*. Its density is given by*

$$f(z|\boldsymbol{\theta}) = \frac{2}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\frac{\lambda_1 (z-\mu)}{\sqrt{\sigma^2 + \lambda_2 (z-\mu)^2}}\right), \tag{5.1}$$

where $\boldsymbol{\theta} = (\mu, \sigma, \lambda_1, \lambda_2)$*. We denote this by* $Z \sim SGN(\mu, \sigma; \lambda_1, \lambda_2)$*. In the special case* $\lambda_2 = \lambda_1^2$ *we say that* X *follows a location-scale skew-curved normal distribution and denote this by* $X \sim SCN(\mu, \sigma; \lambda_1)$*.*

When $\lambda_1 = 0$, the family of densities in (5.1) becomes the $N(\mu, \sigma^2)$ distribution. Additionally, when $\lambda_2 = 0$, (5.1) coincides with the $SN(\mu, \sigma^2, \lambda_1)$ density, i.e., the distribution of $\mu + \sigma A$ with $A \sim SN(\lambda_1)$. A straightforward stochastic representation of the location-scale SGN distribution is obtained as an extension of Proposition 5.

Proposition 10 *If* $Z|Y = y \sim SN(\mu, \sigma; y)$ *and* $Y \sim N(\lambda_1, \lambda_2)$ *then* $Z \sim SGN(\mu, \sigma; \lambda_1, \lambda_2)$*, where* $SN(\mu, \sigma; y)$ *is the location-scale skew-normal distribution.*

The stochastic representation stated by Proposition 10 results in an interesting invariance property of the skew-normal distribution under normal sampling models. The same type of invariance property also holds for the skew-curved normal distribution, as we show next.

Proposition 11 *Let* $Z|Y = y \sim SCN(\mu, \sigma, y)$ *and* $Y \sim N(0, \tau^2)$*. Then* $Z \sim N(\mu, \sigma^2)$ *and* $Y|Z = z \sim SCN\left(0, \tau^2, \left(\frac{\tau}{\sigma}(z-\mu)\right)^2\right)$*.*

Proof: Note first that

$$\begin{aligned}
f_Z(z) &= \frac{2}{\sigma\tau}\int_{-\infty}^{\infty} \phi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\frac{y(z-\mu)}{\sqrt{\sigma^2 + y^2(z-\mu)^2}}\right)\phi\left(\frac{y}{\tau}\right) dy \\
&= \frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)\int_{-\infty}^{\infty} 2\phi(t)\Phi\left(\frac{\tau(z-\mu)t}{\sqrt{\sigma^2 + \tau^2(z-\mu)^2 t^2}}\right) dt \\
&= \frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right).
\end{aligned}$$

To prove the other part observe that

$$
\begin{aligned}
f_{Y\,|\,Z}(y) &= \frac{\frac{2}{\sigma\tau}\phi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\frac{y(z-\mu)}{\sqrt{\sigma^2+y^2(z-\mu)^2}}\right)\phi\left(\frac{y}{\tau}\right)}{\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)} \\
&= \frac{2}{\tau}\phi\left(\frac{y}{\tau}\right)\Phi\left(\frac{\frac{\tau}{\sigma}\left[(z-\mu)\right]y}{\sqrt{\tau^2+\frac{\tau^2}{\sigma^2}\left[(z-\mu)\right]^2y^2}}\right). \qquad \square
\end{aligned}
$$

The result of Proposition 11 can be particularly useful in the context of Bayesian inference. Unlike the usual conjugate models, we have shown that for a skew-curved normal likelihood, a normal prior yields a normal marginal and a skew-curved normal posterior. Finally, the next result shows that a similar invariance property is also true for the skew-generalized normal distribution, but no learning from the data is possible for the second parameter under this particular choice of prior.

Proposition 12 *Let $Z|(Y_1 = y_1, Y_2 = y_2) \sim SGN(\mu, \sigma; y_1, y_2)$, where the joint distribution of (Y_1, Y_2) is defined as $Y_1|Y_2 = y_2 \sim N\left(0, \tau^2(y_2)\right)$ and $Y_2 \sim f_{Y_2}$. Then Z and Y_2 are independent, $Z \sim N(\mu, \sigma^2)$, and*

$$
f_{Y_1|Z}(y_1) = \int_0^\infty \frac{2}{\tau(y_2)}\phi\left(\frac{y_1}{\tau(y_2)}\right)\Phi\left(\frac{\tau(y_2)\,(z-\mu)}{\sqrt{\sigma^2+y_2\,(z-\mu)^2}}\frac{y_1}{\tau(y_2)}\right)f_{Y_2}(y_2)\,dy_2.
$$

Proof: Note first that

$$
\begin{aligned}
f_{Z|Y_2}(z) &= \int_{-\infty}^{\infty}\frac{2}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\frac{y_1(z-\mu)}{\sqrt{\sigma^2+y_2(z-\mu)^2}}\right)\frac{1}{\tau(y_2)}\phi\left(\frac{y_1}{\tau(y_2)}\right)dy_1 \\
&= \frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)\int_{-\infty}^{\infty}2\phi(t)\Phi\left(\frac{\tau(y_2)(z-\mu)t}{\sqrt{\sigma^2+y_2(z-\mu)^2}}\right)dt \\
&= \frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right).
\end{aligned}
$$

Since this conditional density does not depend on y_2, we have that Z and Y_2 are independent, with $Z \sim N(\mu, \sigma^2)$. It is now easy to see that

$$
f_{Y_1\,|\,Z,Y_2}(y_1) = \frac{2}{\tau(y_2)}\phi\left(\frac{y_1}{\tau(y_2)}\right)\Phi\left(\frac{\tau(y_2)\,(z-\mu)}{\sqrt{\sigma^2+y_2\,(z-\mu)^2}}\frac{y_1}{\tau(y_2)}\right),
$$

that is,

$$Y_1|(Y_2 = y_2, Z = z) \sim SN\left(0, \tau^2(y_2); \frac{\tau(y_2)\,(z-\mu)}{\sqrt{\sigma^2 + y_2\,(z-\mu)^2}}\right),$$

and the rest of the proof follows by noting that the independence of Z and Y_2 implies

$$f_{Y_1|Z}(y_1) = \int f_{Y_1|Y_2,Z}(y_1) f_{Y_2}(y_2)\, dy_2. \qquad \square$$

The above proof makes it clear that the inability of learning about Y_2 from Z is due precisely to their independence. In fact, the above calculations imply that

$$f_{Y_1,Y_2|Z}(y_1, y_2) = f_{Y_1|Y_2,Z}(y_1) f_{Y_2}(y_2).$$

Of course, this is only true under the zero-mean normality assumption for the conditional prior $f_{Y_1|Y_2}(y_1)$ and the fact that this depends on the y_2 values only through the scale parameter.

6 Data Illustration

We consider now data concerning the heights (in centimeters) of 100 Australian athletes. The data have been previously analyzed in Cook and Weisberg (10) and are available for download at *http://azzalini.stat.unipd.it/SN/index.html*. Table 1 shows summary statistics for these data. Note that the sample skewness is not included in the range supported by the $SCN(\lambda_1)$ family and the same holds for the sample kurtosis with respect to the $SN(\lambda_1)$ distribution.

n	$\bar{Z}$	S^2	$\sqrt{b_1}$	b_2
100	174.594	67.934	−0.568	4.321

Table 1: *Summary statistics for the heights of* 100 *Australian athletes.*

We estimate parameters by numerically maximizing the likelihood function

$$\log L(\boldsymbol{\theta}) = -\frac{n}{2}\log\frac{\pi\sigma^2}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-\mu)^2 + \sum_{i=1}^{n}\log\Phi\left[\frac{\lambda_1(z_i-\mu)}{\sqrt{\sigma^2+\lambda_2\,(z_i-\mu)^2}}\right]$$

with respect to the components of $\boldsymbol{\theta} = (\mu, \sigma, \lambda_1, \lambda_2)$. The results are summarized in Table 2, considering the full $SGN(\lambda_1, \lambda_2)$ model and three special cases: normal, skew-normal and skew-curved normal.

Parameter estimates	N	SN	SCN	SGN
μ	174.594	174.5834	180.751	170.320
σ^2	67.255	67.255	105.160	85.518
λ_1	–	0.0016	-4.628	4.381
λ_2	–	–	–	24.184
log-likelihood	-352.318	-352.318	-351.066	-347.239

Table 2: *Maximum likelihood parameter estimates for the heights of Australian athletes data under the SGN model and three particular sub-models: Normal* $(\lambda_1 = \lambda_2 = 0)$*, skew-normal* $(\lambda_2 = 0)$ *and skew-curved normal* $(\lambda_2 = \lambda_1^2)$*.*

These summaries illustrate a key point of the SGN model. The flexibility and increased skewness and kurtosis ranges were able to capture features of the data that the other models missed. For instance, it is clear that fitting a $SN(\lambda_1)$ model to these data would be inadequate due to the sample kurtosis value. And this is reflected in that the fitted skew-normal is almost indistinguishable from a Normal model. These points are further illustrated in Figure 3, where a histogram of the data is plotted together with the fitted densities.

Finally, it is possible to carry out asymptotic likelihood ratio tests to compare the SGN model against the sub-models mentioned above. Doing so we conclude that the SGN model is the most appropriate for the particular example analyzed here (data not shown).

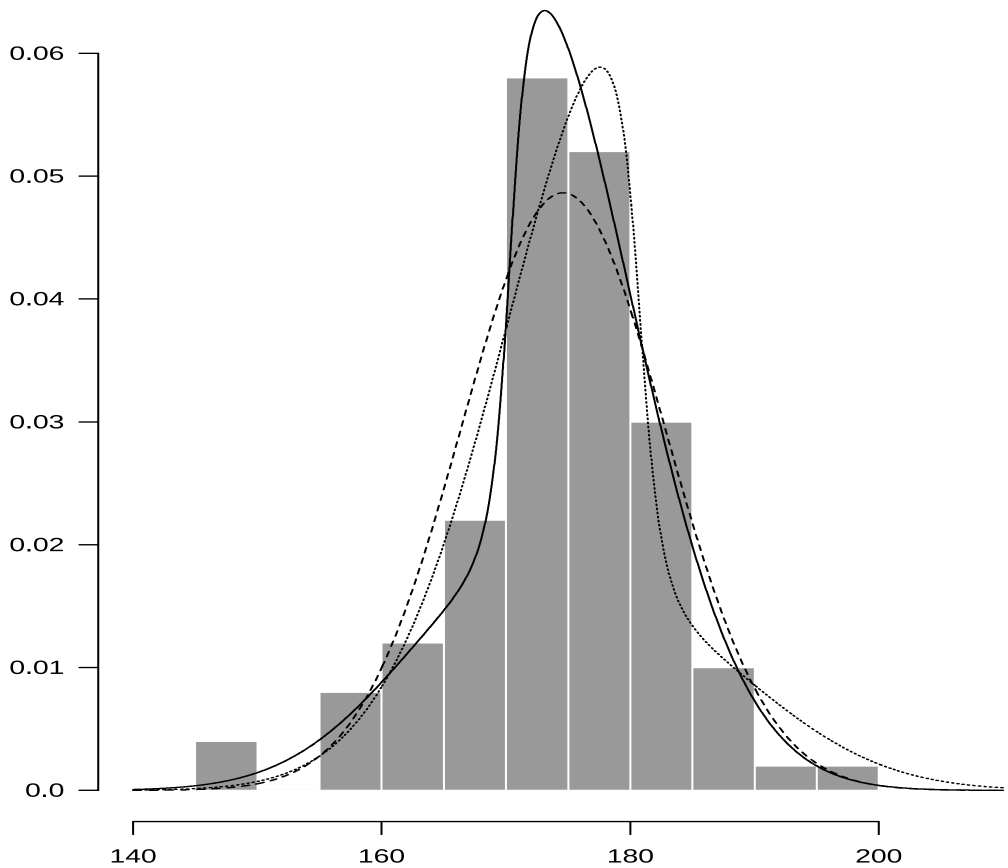

Figure 3: *Histogram of heights of* 100 *Australian athletes. The lines represent distributions fitted using maximum likelihood estimation:* $SGN(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_1, \hat{\lambda}_2)$ *(solid line);* $SCN(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_1, \hat{\lambda}_1^2)$ *(dotted line); and* $SN(\hat{\mu}, \hat{\sigma}, \hat{\lambda}_1)$ *(dashed line), which in this case is almost indistinguishable from the* $N(\hat{\mu}, \hat{\sigma})$ *density.*

BIBLIOGRAPHY

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

(2) Henze, N. A probabilistic representation of the skew-normal distribution. Scandinavian

Journal of Statistics, **1986**, *13*, 271-275.

(3) Azzalini, A. Further results on a class of distributions which includes the normal ones. Statistica. **1986**, *46*, 199-208.

(4) Azzalini, A. and Dalla Valle, A. The multivariate skew-normal distribution. Biometrika. **1996**, *83*, 715-726.

(5) Branco, M. and Dey, D. A general class of multivariate elliptical distributions. Journal of Multivariate Analysis. **2001**, *79*, 99-113.

(6) Arellano-Valle, R. B. and Genton, M. G. On fundamental skew distributions. Technical Report, North Carolina State University, Institute of Statistics, **2003**.

(7) Ellison, B. Two theorems for inferences about the normal distribution with applications in acceptance sampling. Journal of the American Statistical Association, **1964**, *59*, 89-95.

(8) Tanner, M.A. *Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions. Third Edition.* New York: Springer-Verlag. **1996**.

(9) Gilks, W.R. and Wild, P. Adaptive rejection sampling for Gibbs sampling. Journal of Applied Statistics, **1992**, *41*, 337-348.

(10) Cook, R. D. and Weisberg, S. *An introduction to regression graphics.* New York: Wiley. **1994**.