Statistical Inference for a General Class of Asymmetric Distributions

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

Fernando A. Quintana‡

November, 2003

Abstract

We consider a general class of asymmetric univariate distributions depending on a real-valued parameter α, which includes the entire family of univariate symmetric distributions as a special case. We discuss the connections between our proposal and other families of skew distributions that have been studied in the statistical literature. A key element in the construction of such families of distributions is that they can be stochastically represented as the product of two independent random variables. From this representation we can readily derive theoretical properties, easy-to-implement simulation schemes as well as extensions to the multivariate case. We also study statistical inference for this

*Departamento de Estadística, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, CHILE. e-mail: reivalle@mat.puc.cl. Partially supported by grant FONDECYT 8000004.

†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, Pontificia Universidad Católica de Chile. e-mail: quintana@mat.puc.cl. Partially supported by grants FONDECYT 1990430 and 1020712. Corresponding author.

class based on the method of moments and maximum likelihood. We give special attention to the skew-power exponential distribution, but other cases like the skew-t distribution are also considered. Finally, the statistical methods are illustrated with three examples based on real datasets.

KEY WORDS: Kurtosis; Skewness; Stochastic Representation; Symmetric Distributions.

1 Introduction

There has been an increasing interest over the last decade in the construction of flexible parametric families of distributions that exhibit skewness and kurtosis differing from those of the normal distribution. Many of these efforts have been motivated by the need to improve on modelling capabilities in connection with linear and nonlinear regression models. In this regard, a very popular choice is given by the family of elliptic distributions introduced by Kelker (1970) and systematically studied in Cambanis, Huang and Simons (1981). This class contains a vast set of known distributions such as normal, compound normal, t-Student, power exponential and logistic. A comprehensive review of elliptic distributions can be found in Fang, Kotz and Ng (1990) and in Arellano-Valle (1994). However, the main disadvantage of the whole class of elliptical distributions is that they include only symmetric distributions, so their practical application is restricted to situations where the symmetry assumption seems reasonable.

Consequently, a major research effort has been focused in the study of parametric families of asymmetric distributions which are analytically tractable, accommodate practical values of skewness and kurtosis, and have the normal distribution as a particular case. Such distributions are particularly useful for data modelling, statistical analysis and the study of robustness of methods based on normal theory. The works by Azzalini(1985, 1986) introduced a class of asymmetric univariate normal distributions, called skew-normal (SN), which includes the normal distribution as a special case. See also Henze (1986). The construction inherent to the SN distribution has motivated several extensions. Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) introduced multivariate versions of the SN distributions. See also Genton, He and Liu (2001). The idea was further extended in Branco and Dey (2001) who introduced the class of multivariate skew-elliptical (SE) distributions, which contains the SN and coincides with Azzalini's construction in the univariate case. Indeed, the univariate SE is obtained as

$$g(x|\lambda) = 2f(x)F(\lambda x), \tag{1}$$

where F is the cdf of a symmetric density f on $\mathbb{R}$, i.e. $f(x) = f(-x)$, and $\lambda \in \mathbb{R}$ is the asymmetry parameter. The SN corresponds to the case $f(x) = \phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$. Recent extensions of theses ideas are discussed in Genton and Loperfido (2002), Arellano-Valle, del Pino and San Martín (2002), and Arellano-Valle and Genton (2003).

The work by Fernández and Steel (1998) presents an alternative procedure for generating skew distributions from symmetric kernels, based on the introduction of a positive asymmetry parameter (see also Fernández et al. 1995). They define a family of distributions with density given by

$$h(x|\gamma) = \frac{2}{\gamma + \gamma^{-1}} \left(f(\gamma x) I\{x < 0\} + f(x/\gamma) I\{x \geq 0\} \right), \tag{2}$$

where f is a symmetric density (around the origin) and $\gamma > 0$. Bayesian inference for a regression analysis under the skew-t distribution obtained from this class is also considered there. More recently, Mudholkar and Hutson (2000) study the epsilon-skew-normal (ESN) distribution, a form of which can be tracked back to Fechner (1897). They study parameter estimation using the moment and maximum likelihood methods. We note here that the ESN distribution is a special case of a more general family with density defined as

$$h(x|\epsilon) = f(x/(1+\epsilon)) I\{x < 0\} + f(x/(1-\epsilon)) I\{x \geq 0\}, \tag{3}$$

where $|\epsilon| < 1$. The ESN follows by choosing $f(x) = \phi(x)$.

This work introduces a general family of skew distributions that includes (2) and (3) as special cases. A key element in our construction is that the members of the new class can be stochastically represented as the product $U_\alpha \cdot V$, where V is the absolute value of a random variable with symmetric distribution, and U_α is discrete and independent of V, with support $\{-b(\alpha), a(\alpha)\}$, where $a(\alpha)$ and $b(\alpha)$ are known positive asymmetry functions and the asymmetry parameter α is defined on some subset of $\mathbb{R}$. The asymmetry functions are such that $a(\alpha) = b(\alpha)$ reduces down to the symmetric case. Thus, the whole class of symmetric continuous distributions is obtained as a particular case. Furthermore, the construction can be equivalently formulated in terms of densities supported on $\mathbb{R}^+$.

The rest of the article is organized as follows. Section 2 gives the definition of the new family, showing how this can be expressed as a class of location-scale distributions. At the same time, we discuss connections between the sub-families (2) and (3). We also obtain the stochastic representation, which turns out to be quite useful for the study of properties of the proposed family, such as moments, skewness and kurtosis. In Section 3 we briefly describe the results of applying the construction to some relevant special cases, such as the normal, t and logistic distributions. Statistical inference for the new family is discussed in Section 4. In particular, we study the properties of moments estimates for the general case, and maximum likelihood estimation for the important case of the skew-power exponential distribution. Section 5 shows how the methods apply in three examples, and final discussions and multivariate extensions are given in Section 6.

2 A General Family of Skew Distributions

The basic idea in Fechner (1897) is to form a distribution by joining at $x = 0$ two half normals with different scale parameters. This is crucial for the developments presented in Mudholkar and Hutson (2000). The same principle applies to the construction given in Fernández and Steel (1998). We extend both families of distributions as follows:

Definition 1 Let f be a symmetric density. We say that X is a skew random variable if its density function is given by

$$h(x|\alpha) = \frac{2}{a(\alpha)+b(\alpha)} \left\{ f\left(\frac{x}{a(\alpha)}\right) I\{x \geq 0\} + f\left(\frac{x}{b(\alpha)}\right) I\{x < 0\} \right\}, \tag{4}$$

where α is an asymmetry parameter and $a(\alpha)$ and $b(\alpha)$ are known and positive asymmetry functions. We denote this by $X \sim S(f, \alpha)$, or alternatively, by $X \sim Sf(\alpha)$.

It follows that if the asymmetry functions are chosen to be equal, then (4) reduces down to a scale family based on f, which does not modify the underlying symmetry. Hence, the entire class of symmetric densities is a special case of (4). Since any symmetric density on $\mathbb{R}$ can be uniquely determined from a density on $\mathbb{R}^+$, the entire $Sf(\alpha)$ family can be defined in terms of densities supported on $\mathbb{R}^+$ only.

From (4), and letting F denote the cdf of f it is immediately seen that the cdf of a skew distribution is

$$H(x|\alpha) = \begin{cases} \frac{2b(\alpha)}{a(\alpha)+b(\alpha)} F\left(\frac{x}{b(\alpha)}\right) & \text{if } x < 0 \\ \frac{b(\alpha)-a(\alpha)}{a(\alpha)+b(\alpha)} + \frac{2a(\alpha)}{a(\alpha)+b(\alpha)} F\left(\frac{x}{a(\alpha)}\right) & \text{if } x \geq 0 \end{cases} \tag{5}$$

Some basic properties of the $S(f, \alpha)$ family are stated next. The proofs are straightforward and therefore, omitted.

Proposition 1 With the above notation, the following properties hold:

1. The median of the skew family is obtained as

$$H^{-1}(1/2|\alpha) = \begin{cases} b(\alpha)F^{-1}\left(\frac{a(\alpha)+b(\alpha)}{4b(\alpha)}\right) & \text{if } a(\alpha) < b(\alpha) \\ a(\alpha)F^{-1}\left(\frac{3a(\alpha)-b(\alpha)}{4a(\alpha)}\right) & \text{if } a(\alpha) \geq b(\alpha) \end{cases} \tag{6}$$

2. $H(0|\alpha) = \frac{b(\alpha)}{a(\alpha)+b(\alpha)}$.

3. If $X \sim S(f, \alpha)$ then

$$\frac{P(X \geq 0|\alpha)}{P(X < 0|\alpha)} = \frac{a(\alpha)}{b(\alpha)}. \tag{7}$$

Remark 1 The skew class (2) of Fernández and Steel (1998) is a special case of (4), obtained by choosing $a(\gamma) = \gamma$ and $b(\gamma) = \frac{1}{\gamma}$, for $\gamma > 0$. In turn, (3) follows by choosing $a(\epsilon) = 1 - \epsilon$ and $b(\epsilon) = 1 + \epsilon$, with $|\epsilon| < 1$. It is interesting to compare the behavior exhibited by these two special cases. First note that both sub-families retain the mode at $x = 0$. Under (2), we have symmetry if and only if $\gamma = 1$. Besides, $h(-x|\gamma) = h(x|\gamma^{-1})$, i.e. if $X \sim h(x|\gamma)$ then $-X \sim h(x|\gamma^{-1})$, and (7) becomes γ^2. In contrast, under (3), symmetry is obtained if and only if $\epsilon = 0$, and $h(-x|\epsilon) = h(x| - \epsilon)$ so that $X \sim Sf(\epsilon)$ implies $-X \sim Sf(-\epsilon)$. Also, (7) becomes $(1 - \epsilon)/(1 + \epsilon)$. One may be tempted to think of (3) as a reparametrization of (2), by simply matching the expressions for (7), which yields $\epsilon = (1 - \gamma^2)/(1 + \gamma^2)$, a one-to-one transformation between $\mathbb{R}^+$ and $(-1, 1)$. This is of course not true. For instance, taking f as the $U(-1, 1)$ density, (2) yields the $U(-1/\gamma, \gamma)$ distribution, while (3) gives, using the above transformation, the $U(-2/(1 + \gamma^2), 2\gamma^2/(1 + \gamma^2))$ distribution. These are clearly different. Besides, under (2),

$$\lim_{\gamma \to \infty} h(x|\gamma) = 0 \qquad \text{and} \qquad \lim_{\gamma \to 0^+} h(x|\gamma) = 0,$$

while under (3)

$$\lim_{\epsilon \to -1^+} h(x|\epsilon) = f\left(\frac{x}{2}\right) I\{x \geq 0\} \qquad \text{and} \qquad \lim_{\epsilon \to 1^-} h(x|\epsilon) = f\left(\frac{x}{2}\right) I\{x < 0\},$$

so that the asymptotic behavior when the corresponding asymmetry parameter approaches the extremes is radically different.

We next show that any random variable X with density function given by (4) can be represented as the product of two independent random variables.

Proposition 2 Let f be a symmetric density and consider known and positive asymmetry functions $a(\alpha)$ and $b(\alpha)$. Then $X \sim Sf(\alpha)$ if and only if there are two independent random variables V and U_α with $V \sim 2f(x)I\{x \geq 0\}$ and $P(U_\alpha = a(\alpha)) = a(\alpha)/(a(\alpha) + b(\alpha))$, $P(U_\alpha = -b(\alpha)) = b(\alpha)/(a(\alpha) + b(\alpha))$ such that $X = U_\alpha V$.

Proof: The sufficiency of the condition can be obtained by direct computation of the density function of $U_\alpha V$. To show necessity, define

$$U_\alpha = \begin{cases} a(\alpha) & \text{if } X \geq 0 \\ -b(\alpha) & \text{if } X < 0 \end{cases} \qquad V = \begin{cases} \frac{X}{a(\alpha)} & \text{if } X \geq 0 \\ \frac{-X}{b(\alpha)} & \text{if } X < 0 \end{cases}$$

Then it follows that $X = U_\alpha V$, U_α and V are independent and have the appropriate marginal distributions. □

Remark 2 The stochastic representation depicted in Proposition 2 and its proof show an alternative way of constructing random variables $X \sim S(f, \alpha)$, which is particularly useful for simulation purposes: simply generate $Y \sim f$, and U_α with $P(U_\alpha = a(\alpha)) = a(\alpha)/(a(\alpha) + b(\alpha)) = 1 - P(U_\alpha = -b(\alpha))$ independent of Y, and set $X = U_\alpha|Y|$. Moreover, we may think of the family (4) as the result of generating first a random variable V with the same distribution as $|Y|$, followed by a random assignment of the sign and a rescaling, as indicated by the U_α factor. Furthermore, we can rewrite $U = U_\alpha/|U_\alpha| = X/|X|$, where $P(U = 1) = \frac{a(\alpha)}{a(\alpha)+b(\alpha)} = 1-P(U = -1)$, and letting $R = |U_\alpha|V = |X|$ we get $X = RU$, with R and U being independent. Note that the sign of X is, in general, chosen with unequal probabilities. When $a(\alpha) = b(\alpha) = 1$ the rescaling vanishes, and (4) is generated by choosing the magnitude first, followed by a random reflection about the origin. This is just the stochastic characterization of any symmetric distribution about 0, as discussed in Fang et al. (1990), and later generalized to a $\mathcal{C}$-class of symmetric distributions by Arellano-Valle et al. (2002).

We next extend the family of skew distributions to include location and scale parameters.

Definition 2 Let $X \sim S(f, \alpha)$. The family of location-scale skew distributions, is defined as the distribution of $Z = \mu + \sigma X$ for $\mu \in \mathbb{R}$ and $\sigma > 0$. The corresponding density is given by

$$h(z|\boldsymbol{\theta}) = \frac{2}{\sigma(a(\alpha) + b(\alpha))} \left\{ f\left(\frac{z-\mu}{\sigma a(\alpha)}\right) I\{z \geq \mu\} + f\left(\frac{z-\mu}{\sigma b(\alpha)}\right) I\{z < \mu\} \right\}, \qquad (8)$$

where $\boldsymbol{\theta} = (\mu, \sigma, \alpha)$ and we denote this by $Z \sim S(f, \mu, \sigma, \alpha)$, or $Z \sim Sf(\mu, \sigma, \alpha)$ or $Z \sim Sf(\boldsymbol{\theta})$.

Basic properties of the location-scale skew distributions can be obtained exactly as before. For instance, the median is given by

$$H^{-1}(1/2|\alpha) = \begin{cases} \mu + \sigma b(\alpha) F^{-1}\left(\frac{a(\alpha)+b(\alpha)}{4b(\alpha)}\right) & \text{if } a(\alpha) < b(\alpha) \\ \mu + \sigma a(\alpha) F^{-1}\left(\frac{3a(\alpha)-b(\alpha)}{4a(\alpha)}\right) & \text{if } a(\alpha) \geq b(\alpha). \end{cases} \tag{9}$$

The stochastic representation in Proposition 2 allows to easily compute moments of the skew distributions. For $r \in \{1, 2, \ldots\}$ denote

$$d_r = \int_{-\infty}^{\infty} |x|^r f(x)\, dx,$$

i.e., $d_r = E(V^r)$ in the notation of Proposition 2. It is clear that $d_r < \infty$ is a necessary and sufficient condition for the existence of moments up to order r of (4) or (8). Thus, assuming existence, expressions for these moments are given next.

Proposition 3 Let $X \sim Sf(0, 1, \alpha)$ and $Z \sim Sf(\mu, \sigma, \alpha)$, and denote $\mu_r = E(X^r)$ and $\mu_r' = E(Z^r)$. Then we have, for $r = 1, 2, \ldots$

$$\mu_r = \left\{ \frac{a(\alpha)^{r+1} + (-1)^r b(\alpha)^{r+1}}{a(\alpha) + b(\alpha)} \right\} d_r, \quad \text{and} \quad \mu_r' = \sum_{k=0}^{r} \binom{r}{k} \sigma^k \mu^{r-k} \mu_k.$$

Proof: The result follows by using the independence of U_α and V in Proposition 2, and the Binomial Theorem. $\square$.

In what follows we will specialize our discussion to the sub-family (3), i.e. we assume $a(\alpha) = 1 - \alpha$ and $b(\alpha) = 1 + \alpha$. This is motivated mainly by two reasons. First, (4) is quite general, and any practical application of such distributions require adopting specific forms for the asymmetry functions. And secondly, (3) is an interesting class by itself, which has more stable behavior than, e.g. (2). See Remark 1 above. It is then an immediate consequence of Proposition 3, that expectation and variance of the corresponding location-scale skew distribution are given by

$$E(Z) = \mu - 2\alpha\sigma d_1 \qquad \text{and} \qquad \mathrm{Var}(Z) = \sigma^2 \left\{ (1 + 3\alpha^2) d_2 - 4(\alpha d_1)^2 \right\}. \tag{10}$$

Likewise, the (standardized) skewness and kurtosis coefficients are given by

$$\alpha_3 = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{\{(1+3\alpha^2)d_2 - 4(\alpha d_1)^2\}^{3/2}} \qquad \text{and} \qquad \alpha_4 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{\{(1+3\alpha^2)d_2 - 4(\alpha d_1)^2\}^2}, \quad (11)$$

where μ_r, $r = 1, 2\ldots$ is defined in Proposition 3.

3 Some Special Cases

Table 1 gives specific results for the following particular cases of density f in (4), assuming $a(\alpha) = 1 - \alpha$ and $b(\alpha) = 1 + \alpha$: (i) $f(x) = c_1 \exp\{-\frac{1}{2}|x|^{2/(1+\beta)}\}$ with $c_1^{-1} = 2^{(3+\beta)/2}\Gamma((3+\beta)/2)$ (Power Exponential); (ii) $f(x) = c_2(1 + x^2/\nu)^{-(1+\nu)/2}$ with $c_2 = (\nu\pi)^{-1/2}\Gamma(\nu/2)^{-1}\Gamma((\nu+1)/2)$ (Student's t); (iii) $f(x) = e^{-x}/(1+e^{-x})^2$ (logistic); and (iv) $f(x) = (2\Gamma(\nu))^{-1}\beta^\nu|x|^{\nu-1}e^{-\beta|x|}$ (symmetrized gamma).

$f(x)$	d_r	$E(Z)$	$\mathrm{Var}(Z)$
Power Exponential	$\frac{2^{r(1+\beta)/2}\Gamma((1+\beta)(1+r)/2)}{\Gamma((1+\beta)/2)}$	$\mu - \frac{2^{(3+\beta)/2}\alpha\sigma\Gamma(1+\beta)}{\Gamma((1+\beta)/2)}$	$\frac{\sigma^2 2^{1+\beta}}{\Gamma((1+\beta)/2)}\left\{(1+3\alpha^2)\Gamma\left(\frac{3(1+\beta)}{2}\right) - \frac{4\alpha^2\Gamma(1+\beta)^2}{\Gamma((1+\beta)/2)}\right\}$
t (*)	$2c_2\nu^{\frac{r+1}{2}}\sum_{j=0}^{\frac{r-1}{2}}\frac{(-1)^{\frac{r-1-2j}{2}}}{(\nu-2j-1)}\binom{\frac{r-1}{2}}{j}$, if r is odd $\frac{\nu^{r/2}\cdot 1\cdot 3\cdots(r-1)}{(\nu-r)(\nu-r+2)\cdots(\nu-2)}$, if r is even	$\mu - \frac{4c_2\nu\alpha\sigma}{\nu-1}$	$\sigma^2\left\{\frac{\nu(1+3\alpha^2)}{\nu-2} - \frac{16(c_2\alpha\nu)^2}{(\nu-1)^2}\right\}$
Logistic	$d_1 = 2\log(2)$ $d_r = \frac{2\Gamma(r+1)(2^r-2)\zeta(r)}{2^r}$, $r = 2, 3, \ldots$	$\mu - 4\alpha\sigma\log(2)$	$\sigma^2\left\{\frac{(1+3\alpha^2)\pi^2}{3} - 16(\alpha\log(2))^2\right\}$
Symmetrized Gamma	$\frac{\Gamma(r+\nu)}{\beta^r\Gamma(\nu)}$	$\mu - \frac{2\alpha\sigma\nu}{\beta}$	$\sigma^2\left\{\frac{(1+3\alpha^2)\nu(\nu+1)-4(\alpha\nu)^2}{\beta^2}\right\}$

Table 1: *Summaries for some particular cases of interest. Here $Z \sim S(f, \mu, \sigma, \alpha)$ for $f(x)$ indicated in the first column. (*) In this case, $d_r < \infty$ if and only if $\nu > r$.*

Remark 3 We note here that the power exponential distribution has a "strict inclusion" property. Indeed, the normal and double exponential distributions are obtained by taking $\beta = 0$ and $\beta = 1$ respectively. In addition, great flexibility in the specification of skewness and kurtosis coefficients is obtained. For instance, setting $\beta > 0$ yields heavier tails than the normal distribution, and the contrary is true for $\beta < 0$. We also note that the moments shown in Table 1 for the logistic distribution (iii) involve the Riemann Zeta function $\zeta(r) = \sum_{j=1}^{\infty} \frac{1}{j^r}$, which exists for integers $r = 2, 3, \ldots$

Table 2 illustrates the wide ranges of skewness and kurtosis computed as (11) that can be obtained for different choices of the generating density $f(x)$. We remark here that the ranges for the $S(\phi, \alpha)$ distribution coincide with those for Azzalini's $SN(\lambda)$ (Mudholkar and Hutson 2000). From (4) we can also see that the mode of the $S(\phi, \alpha)$ is always located at $x = 0$, while for the $SN(\lambda)$ distribution this varies with λ. In addition, if $X \sim SN(\lambda)$ then $X^2 \sim \chi^2(1)$, the chi-square distribution with 1 degree of freedom. Therefore, the even moments of X do not depend on the asymmetry parameter λ, which is not the case for the $S(f, \alpha)$ model.

$f(x)$	Skewness	Kurtosis
Power Exponential(0)	$-0.995 < \alpha_3 < 0.995$	$3.000 < \alpha_4 < 3.869$
Power Exponential(0.5)	$-1.504 < \alpha_3 < 1.504$	$4.200 < \alpha_4 < 6.008$
Power Exponential(-0.5)	$-0.443 < \alpha_3 < 0.443$	$2.180 < \alpha_4 < 2.440$
$t_{(5)}$	$-2.550 < \alpha_3 < 2.550$	$9.000 < \alpha_4 < 23.109$
Logistic	$-1.540 < \alpha_3 < 1.540$	$4.200 < \alpha_4 < 6.584$
Symmetrized Gamma(6,3)	$-0.816 < \alpha_3 < 0.816$	$1.714 < \alpha_4 < 4.000$

Table 2: *Possible values for the standardized coefficients of skewness and kurtosis for some particular cases when the asymmetry parameter α ranges over* $(-1, 1)$.

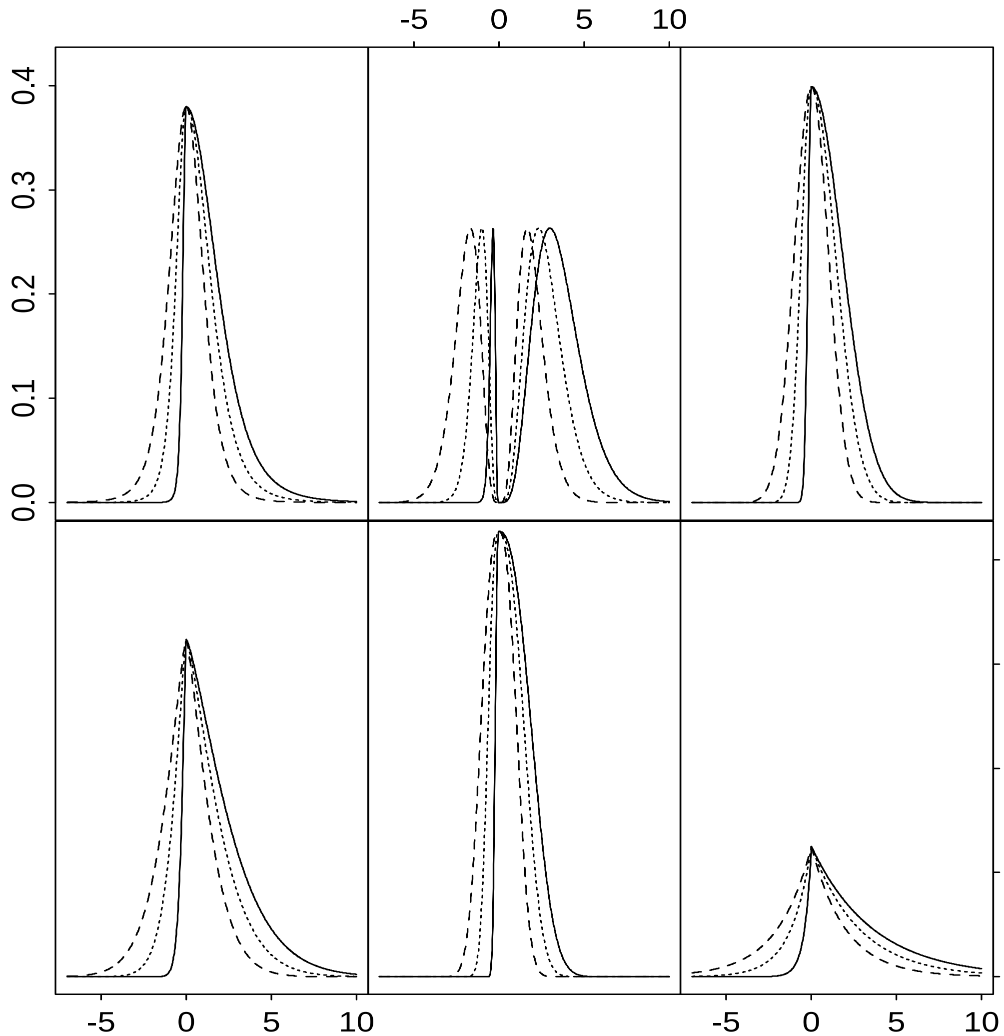

Figure 1: *Examples of $S(f, \alpha)$ density for $\alpha = 0$ (dashed line), $\alpha = -0.4$ (dotted line) and $\alpha = -0.8$ (solid line). From left to right, upper panel shows the cases when $f(x)$ is: (i) the $t_{(5)}$ density; (ii) the symmetrized $\Gamma(6, 3)$ density; and (iii) the $SPE(0, 1, \alpha, 0)$ density. From left to right, the lower panel shows the cases when $f(x)$ is the $SPE(0, 1, \alpha, \beta)$ density with (i) $\beta = 0.5$; (ii) $\beta = -0.2$; and (iii) $\beta = 1$.*

Figure 1 shows plots of the skew densities $S(f, \alpha)$ for different choices of α and $f(x)$. The left plot in the upper panel corresponds to the $t_{(5)}$ distribution, which is a heavy-tailed version of the case $f(x) = \phi(x)$, studied by Mudholkar and Hutson (2000). The middle plot in the upper panel depicts the case when $f(x)$ is the symmetrized $\Gamma(6, 3)$ density, which results in bimodal shapes. We note that these modes are not symmetrically located with respect to the origin, and the locations depend on the value of α. The remaining plots represent the various shapes that the resulting distributions may adopt when $f(x)$ is the skew-power exponential density for different choices of β.

4 Inferential Aspects

We discuss now issues concerning statistical inference in the location-scale skew family (8), under the assumption $a(\alpha) = 1 - \alpha$ and $b(\alpha) = 1 + \alpha$. In the general case, we discuss estimation via the method of moments as well as some asymptotic results. For the particular case of the power exponential distribution, we study maximum likelihood estimation. Our developments follow the methodology discussed in Mudholkar and Hutson (2000), thus extending some of their results, particularly those concerning asymptotic theory, to a larger class of distributions.

4.1 Estimation by the Method of Moments

Let $Z_1, \ldots, Z_n$ be a sample from the $Sf(\mu, \sigma, \alpha)$ density (8). Then

$$\mu = E(Z) + 2\alpha\sigma d_1 \qquad \text{and} \qquad \sigma^2 = \frac{Var(Z)}{(1 + 3\alpha^2)d_2 - 4\alpha^2 d_1^2}. \tag{12}$$

When f does not depend on additional parameters, the d_i coefficients are completely specified, so we use the sample skewness coefficient γ_1 and solve

$$\gamma_1 = \frac{-4\hat{\alpha}(1 + \hat{\alpha}^2)d_3 + 6\hat{\alpha}(1 + 3\hat{\alpha}^2)d_1 d_2 - 16\hat{\alpha}^3 d_1^3}{\{(1 + 3\hat{\alpha}^2)d_2 - 4\hat{\alpha}^2 d_1^2\}^{3/2}} \tag{13}$$

for $\hat{\alpha}$, using an appropriate numerical method. Let $\hat{\alpha}_M$ be the solution. Next, we substitute in (12) α by $\hat{\alpha}_M$, $E(Z)$ by the sample mean and $Var(Z)$ by the sample variance, from which the moments estimates $(\hat{\mu}_M, \hat{\sigma}^2_M, \hat{\alpha}_M)$ are readily obtained.

Remark 4 A similar procedure can be applied when the d_i coefficients are unknown, i.e., f depends on an additional real-valued parameter β. In this case, we resort to the sample skewness (γ_1) and kurtosis (γ_2), adding to (13) equation

$$\gamma_2 = \frac{(1+10\hat{\alpha}^2+5\hat{\alpha}^4)\hat{b}_4 - 32\hat{\alpha}^2(1+\hat{\alpha}^2)\hat{b}_1\hat{b}_3 + 24\hat{\alpha}^2(1+3\hat{\alpha}^2)\hat{b}_1^2\hat{b}_2 - 48\hat{\alpha}^4\hat{b}_1^4}{\{(1+3\hat{\alpha}^2)d_2 - 4\hat{\alpha}^2 d_1^2\}^2}, \tag{14}$$

where $\hat{b}_r = d_r(\hat{\beta})$. Solving (13) and (14) for α and β, using some numerical method, we obtain $\hat{\alpha}_M$ and $\hat{\beta}_M$, which are later used in (12) to produce $\hat{\mu}_M$ and $\hat{\sigma}^2_M$. Note in this case that d_r in (13) needs to be replaced by $\hat{b}_r$, $r = 1, 2, 3$.

Asymptotic normality of these estimates is established next.

Proposition 4 Let $Z_1, \ldots, Z_n$ be a sample from the $Sf(\mu, \sigma, \alpha)$ distribution with known f. Let $\boldsymbol{\theta} = (\mu, \sigma^2, \alpha)$ and for $k = 1, 2, 3$, write $\mu'_k = \mu'_k(\boldsymbol{\theta}) = E(Z^k)$. If $\mu'_6(\boldsymbol{\theta}) < \infty$ and $\hat{\boldsymbol{\theta}}_M(n)$ is the corresponding moments estimate, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_M(n) - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_3(\mathbf{0}, \Gamma(\boldsymbol{\theta}))$$

as $n \to \infty$, where $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = H^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}[H^{-1}(\boldsymbol{\theta})]^T$, $\boldsymbol{\Sigma} = \{(\mu'_{i+j} - \mu'_i\mu'_j)_{i,j}\}$ and $H(\boldsymbol{\theta})$ is given by

$$\begin{pmatrix} 1 & \frac{-\alpha d_1}{\sigma} & -2\sigma d_1 \\ 2\mu - 4\sigma\alpha d_1 & \tau_3 d_2 - \frac{2\mu\alpha d_1}{\sigma} & -4\mu\sigma d_1 + 6\sigma^2\alpha d_2 \\ 3\mu^2 - 12\mu\sigma\alpha d_1 + 3\sigma^2\tau_3 d_2 & 3\mu\tau_3 d_2 - \frac{3\mu^2\alpha d_1}{\sigma} - 6\sigma\alpha\tau_1 d_3 & 18\mu\sigma^2\alpha d_2 - 6\mu^2\sigma d_1 - 4\sigma^3\tau_3 d_3 \end{pmatrix},$$

with $\tau_k = (1 + k\alpha^2)$.

Sketch of the Proof: Define $A(\boldsymbol{\theta}) = (\mu'_1(\boldsymbol{\theta}), \mu'_2(\boldsymbol{\theta}), \mu'_3(\boldsymbol{\theta}))$ and $H(\boldsymbol{\theta}) = \partial A(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. It follows that $H(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ and has complete rank. Then the result is a consequence of a standard Taylor series expansion of $A(\boldsymbol{\theta})$:

$$\sqrt{n}\left(\boldsymbol{M}_n - A(\boldsymbol{\theta})\right) = \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) H(\boldsymbol{\theta})^T + \mathcal{O}_p(n),$$

where $\boldsymbol{M}_n$ is the vector of the first three sample moments. $\square$

4.2 Maximum Likelihood Estimation for the Skew-Power Exponential Distribution

Maximum likelihood estimation (MLE) for the $Sf(\mu, \sigma, \alpha)$ family is, conceptually, straightforward. In practice, however, computations require some care. To illustrate these calculations, we analyze here the skew-power exponential (SPE) distribution with known β. The methodology extends the results presented in Mudholkar and Hutson (2000), who considered the particular case of the skew-normal distribution (i.e., $\beta = 0$).

Let $z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(n)}$ be the order statistics associated to a random sample $Z_1, \ldots, Z_n$ from the $SPE(\mu, \sigma, \alpha, \beta)$ distribution. Denote $z_{(0)} = -\infty$ and $z_{(n+1)} = \infty$. Let $k \equiv k(z_{(1)}, \ldots, z_{(n)}, \mu)$ be the random integer such that $z_{(k)} < \mu < z_{(k+1)}$. By the continuity of the SPE distribution, k is well defined with probability 1, and ranges over $\{0, 1, \ldots, n\}$. The log-likelihood function can then be expressed as

$$l(\mu, \sigma^2, \alpha) = n \log(c_1) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^{\frac{2}{(1+\beta)}}} g_k(\mu, \alpha), \tag{15}$$

where

$$g_k(\mu, \alpha) = \left\{(1+\alpha)^{-\frac{2}{(1+\beta)}} \sum_{i=1}^{k} |z_{(i)} - \mu|^{\frac{2}{(1+\beta)}} + (1-\alpha)^{-\frac{2}{(1+\beta)}} \sum_{i=k+1}^{n} |z_{(i)} - \mu|^{\frac{2}{(1+\beta)}}\right\}. \tag{16}$$

To compute the MLE of $\boldsymbol{\theta}$ we need to maximize (15), which we can do by fixing k first, and then finding the corresponding optimal values $\hat{\boldsymbol{\theta}}_k$. The MLE, is $\hat{\boldsymbol{\theta}}_k^* = \arg\max\, l(\hat{\boldsymbol{\theta}}_k)$ over the set $k = 0, \ldots, n$.

The cases $k = 0$ and $k = n$ can be easily solved. When $k = 0$ then (15) reduces to

$$-\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^{\frac{2}{(1+\beta)}}}\left\{\sum_{i=1}^{n}\frac{|z_{(i)}-\mu|^{\frac{2}{(1+\beta)}}}{(1-\alpha)^{\frac{2}{(1+\beta)}}}\right\},$$

which, in terms of α, is maximized at $\hat{\alpha}_0 = -1$. In addition, $\mu < z_{(1)}$ which implies that $\hat{\mu}_0 = z_{(1)}$, and from this it follows that

$$\hat{\sigma}_0^2 = \left[\frac{1}{2^{\frac{2}{1+\beta}}n(1+\beta)}\sum_{i=2}^{n}|z_{(i)}-z_{(1)}|^{\frac{2}{1+\beta}}\right]^{1+\beta}.$$

Similarly, when $k = n$ we obtain $\hat{\alpha}_n = 1$, $\hat{\mu}_n = z_{(n)}$ and

$$\hat{\sigma}_n^2 = \left[\frac{1}{2^{\frac{2}{1+\beta}}n(1+\beta)}\sum_{i=1}^{n-1}|z_{(i)}-z_{(n)}|^{\frac{2}{1+\beta}}\right]^{1+\beta}.$$

When $1 \le k < n$, we need to solve the likelihood equations

$$\frac{\partial}{\partial\mu}g_k(\mu,\alpha) = 0, \quad \frac{\partial}{\partial\alpha}g_k(\mu,\alpha) = 0, \quad \frac{\partial}{\partial\sigma^2}l(\mu,\sigma^2,\alpha) = 0.$$

In terms of α and σ^2 we can easily find that

$$\begin{aligned}
\hat{\alpha}_k &= \frac{\left(\sum_{i=1}^{k}|z_{(i)}-\hat{\mu}_k|^{\frac{2}{1+\beta}}\right)^{\frac{1+\beta}{3+\beta}} - \left(\sum_{i=k+1}^{n}|z_{(i)}-\hat{\mu}_k|^{\frac{2}{1+\beta}}\right)^{\frac{1+\beta}{3+\beta}}}{\left(\sum_{i=1}^{k}|z_{(i)}-\hat{\mu}_k|^{\frac{2}{1+\beta}}\right)^{\frac{1+\beta}{3+\beta}} + \left(\sum_{i=k+1}^{n}|z_{(i)}-\hat{\mu}_k|^{\frac{2}{1+\beta}}\right)^{\frac{1+\beta}{3+\beta}}}\\
\hat{\sigma}_k^2 &= \frac{1}{n(1+\beta)}\left\{\sum_{i=1}^{k}\frac{|z_{(i)}-\hat{\mu}_k|^{\frac{2}{1+\beta}}}{(1+\hat{\alpha}_k)^{\frac{2}{1+\beta}}} + \sum_{i=k+1}^{n}\frac{|z_{(i)}-\hat{\mu}_k|^{\frac{2}{1+\beta}}}{(1-\hat{\alpha}_k)^{\frac{2}{1+\beta}}}\right\}.
\end{aligned}$$

Additionally, it follows that $g_k(\cdot)$ is differentiable with respect to μ provided that $-1 < \beta < 1$. However, this does not imply that $\frac{\partial}{\partial\mu}g_k(\mu,\alpha) = 0$ can be solved in the interval $(z_{(k)}, z_{(k+1)})$. If this is the case then $\hat{\mu}_k$ will be either $z_{(k)}$ or $z_{(k+1)}$ according to wether $\frac{\partial}{\partial\mu}g_k(\mu,\alpha)$ is negative or positive over $(z_{(k)}, z_{(k+1)})$. On the other hand, if $g_k(\mu,\alpha)$ does have a local minimum at an interior point of the interval $(z_{(k)}, z_{(k+1)})$ then this point is a solution of

$$\frac{\sum_{i=1}^{k}|z_{(i)}-\mu|^{\frac{1-\beta}{1+\beta}}}{\sum_{i=1}^{k}|z_{(i)}-\mu|^{\frac{1}{3+\beta}}} = \frac{\sum_{i=k+1}^{n}|z_{(i)}-\mu|^{\frac{1-\beta}{1+\beta}}}{\sum_{i=k+1}^{n}|z_{(i)}-\mu|^{\frac{1}{3+\beta}}}.$$

Putting the pieces together, we have shown the following:

Proposition 5 The MLE of $\boldsymbol{\theta}$ is $(\hat{\mu}_k, \hat{\sigma}_k^2, \hat{\alpha}_k)$, where k is such that $l(\hat{\mu}_k, \hat{\sigma}_k^2, \hat{\alpha}_k) \geq l(\hat{\mu}_j, \hat{\sigma}_j^2, \hat{\alpha}_j)$ for all $j \in \{0, \ldots, n\}$.

Asymptotic properties of the MLE are established next.

Proposition 6 Let $Z \sim Sf(\mu, \sigma, \alpha)$. Assume that f has two continuous derivatives and that the following quantities exist for $i = 0, 1, 2$ and $j = 0, 1$:

$$\begin{aligned} \kappa_i \equiv \kappa_i(f) &= \int_0^\infty \left(\frac{(f'(t))^2}{f(t)} - f''(t) \right) t^i \, dt \\ s_j \equiv s_j(f) &= \int_0^\infty t^j f'(t) \, dt. \end{aligned}$$

Then the MLE $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}^2, \hat{\alpha})$ is consistent and asymptotically normal: $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_3(\mathbf{0}, \boldsymbol{V})$ as $n \to \infty$, with

$$\boldsymbol{V} = \begin{pmatrix} \frac{\sigma^2(1-\alpha^2)(\kappa_2 - 2s_1)}{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]} & 0 & \frac{\sigma(1-\alpha^2)(s_0 - \kappa_1)}{2[(s_0 - \kappa_1)^2 - \kappa_0(\kappa_2 - 2s_1)]} \\ 0 & \frac{2\sigma^4}{\kappa_2 - 3s_1 - 1} & 0 \\ \frac{\sigma(1-\alpha^2)(s_0 - \kappa_1)}{2[(s_0 - \kappa_1)^2 - \kappa_0(\kappa_2 - 2s_1)]} & 0 & \frac{\kappa_0(1-\alpha^2)}{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]} \end{pmatrix}.$$

Proof: Straightforward (e.g. Lehmann 1999) and therefore omitted. □

Table 3 shows the values of κ_i and s_j for the examples considered in Section 3.

$f(x)$	s_0	s_1	κ_0	κ_1	κ_2
Power Exponential	$-c_1$	$-\frac{1}{2}$	$\frac{c_1(1-\beta)\Gamma((1-\beta)/2)}{2^{(1+\beta)/2}(1+\beta)}$, $\beta < 1$	$\frac{c_1(1-\beta)}{1+\beta}$	$\frac{1-\beta}{2(1+\beta)}$
$t_{(\nu)}$	$-c_2$	$-\frac{1}{2}$	$\frac{\nu+1}{2(\nu+3)}$	$\frac{c_2(\nu-1)}{3+\nu}$	$\frac{\nu-3}{2(\nu+3)}$
Logistic	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{1}{12} + \frac{\log(2)}{2}$	$-\frac{1}{3} + \frac{\pi^2}{18}$
Symmetrized Gamma	$-\frac{1}{2}$	0	$\frac{\beta^2}{2(\nu-2)}$, $\nu > 2$	$\frac{\beta}{2}$, $\nu > 1$	$\frac{\nu-1}{2}$

Table 3: *Values of κ_i and s_j for the examples considered in Section 3.*

Corollary 1 If $Z \sim SPE(\mu, \sigma, \alpha, \beta)$ with known β then the asymptotic covariance matrix $\boldsymbol{V}$ from Proposition 6 is given by

$$\boldsymbol{V} = \begin{pmatrix} \frac{2^{1+\beta}\sigma^2(1+\beta)(3+\beta)(1-\alpha^2)\Gamma\left(\frac{3+\beta}{2}\right)^2}{(1-\beta)(3+\beta)\Gamma\left(\frac{1-\beta}{2}\right)\Gamma\left(\frac{3+\beta}{2}\right)-4} & 0 & \frac{2^{\frac{3+\beta}{2}}\sigma(1+\beta)(1-\alpha^2)\Gamma\left(\frac{3+\beta}{2}\right)}{(1-\beta)(3+\beta)\Gamma\left(\frac{1-\beta}{2}\right)\Gamma\left(\frac{3+\beta}{2}\right)-4} \\ 0 & 2(1+\beta)\sigma^4 & 0 \\ \frac{2^{\frac{3+\beta}{2}}\sigma(1+\beta)(1-\alpha^2)\Gamma\left(\frac{3+\beta}{2}\right)}{(1-\beta)(3+\beta)\Gamma\left(\frac{1-\beta}{2}\right)\Gamma\left(\frac{3+\beta}{2}\right)-4} & 0 & \frac{(1-\beta^2)(1-\alpha^2)\Gamma\left(\frac{1-\beta}{2}\right)\Gamma\left(\frac{3+\beta}{2}\right)}{(1-\beta)(3+\beta)\Gamma\left(\frac{1-\beta}{2}\right)\Gamma\left(\frac{3+\beta}{2}\right)-4} \end{pmatrix}.$$

Remark 5 We can also extend the result of Theorem 4.11 in Mudholkar and Hutson (2000). Indeed, taking the transformation

$$g(\hat{\mu}, \hat{\sigma}^2, \hat{\alpha}) = \left(\hat{\mu}, \sqrt{\frac{\kappa_2 - 3s_1 - 1}{2}}\log(\hat{\sigma}^2), \sqrt{\frac{2[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}{\kappa_0}}\arcsin(\hat{\alpha})\right),$$

then as $n \to \infty$

$$\sqrt{n}(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} \mathcal{N}_3\left[\mathbf{0}, \begin{pmatrix} \frac{\sigma^2(1-\alpha^2)(\kappa_2-2s_1)}{2[\kappa_0(\kappa_2-2s_1)-(s_0-\kappa_1)^2]} & 0 & \frac{\sigma(\kappa_1-s_0)\sqrt{1-\alpha^2}}{\sqrt{2[\kappa_0(\kappa_2-2s_1)-(s_0-\kappa_1)^2]}} \\ 0 & 1 & 0 \\ \frac{\sigma(\kappa_1-s_0)\sqrt{1-\alpha^2}}{\sqrt{2[\kappa_0(\kappa_2-2s_1)-(s_0-\kappa_1)^2]}} & 0 & 1 \end{pmatrix}\right].$$

This result can be used to construct approximate confidence intervals for individual parameters. For instance,

$$\left(\sin\left\{\arcsin(\hat{\alpha}) - \Phi^{-1}\left(1 - \frac{\gamma}{2}\right)\sqrt{\frac{\kappa_0}{2n[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}}\right\},\right.$$
$$\left.\sin\left\{\arcsin(\hat{\alpha}) + \Phi^{-1}\left(1 - \frac{\gamma}{2}\right)\sqrt{\frac{\kappa_0}{2n[\kappa_0(\kappa_2 - 2s_1) - (s_0 - \kappa_1)^2]}}\right\}\right)$$

is an approximate $100(1-\gamma)\%$ confidence interval for α.

Finally, the case where β is unknown can be conceptually handled with no extra difficulty. This simply requires solving an enlarged set of likelihood equations that includes the extra variable and the additional equation given by $\frac{\partial}{\partial\beta}l(\mu, \sigma^2, \alpha, \beta) = 0$. Asymptotic theory for this new scenario requires deriving the corresponding and also enlarged Information matrix, which can be easily done using standard calculations.

5 Examples

We now apply the methods discussed earlier to three examples. The first one is the dataset of 219 heights of volcanos used in Mudholkar and Hutson (2000). The second example is taken from a dataset consisting of several variables recorded on 202 Australian athletes and reported in Cook and Weisberg (1994). Concretely, we analyze here measurements of the plasma ferritin concentration. The third example considers the dataset provided by the Chilean National Institute of Statistics (INE), consisting of births that resulted in death of the child before reaching the age of 1 year. Data on an equal size sample of births that survived 1 year are also available. Specifically, we analyze here a random subsample of 250 of the mother ages.

Table 4 presents some summary statistics for the three datasets. These summaries suggest that the first two examples correspond to leptokurtic densities, while the third one is platykurtic. Additionally, in all the three cases positive skewness is present.

Example	n	$\bar{X}$	S^2	γ_1	γ_2
1	219	70.246	1850.56	0.840	3.482
2	202	76.876	2256.37	1.290	4.486
3	250	26.408	49.27	0.283	2.398

Table 4: *Summary descriptive statistics for each of the examples. Here,* γ_1 *and* γ_2 *represent the sample skewness and kurtosis.*

We fitted the skew-power exponential (SPE) distribution to the examples, using the method of moments (MM) and maximum likelihood (ML) as described in Section 4. The results are summarized in Table 5, also including asymptotic 95% confidence intervals for all estimates.

		Estimation Method	
Example	Parameter	MM	ML
1	μ	27.773 (-8.929,64.476)	26.774 (16.251,37.298)
	σ^2	1588.71 (133.74,2983.42)	1768.43 (872.38,2664.48)
$(n = 219)$	α	-0.6789 (-1.000,-0.084)	-0.684 (-0.835,-0.534)
	β	-0.009 (-0.605,0.586)	-0.076 (-0.356,0.204)
2	μ	(*)	35.000 (28.814,41.186)
	σ^2	(*)	615.98 (122.11,1109.85)
$(n = 202)$	α	(*)	-0.695 (-0.806,-0.584)
	β	(*)	0.503 (0.107,0.899)
3	μ	23.333 (20.445,26.222)	21.523 (18.299,24.747)
	σ^2	70.63 (39.23,102.03)	89.62 (58.96,120.28)
$(n = 250)$	α	-0.271 (-0.495,-0.046)	-0.426 (-0.673,-0.179)
	β	-0.237 (-0.499,0.026)	-0.422 (-0.627,-0.216)

Table 5: *Parameter estimates and asymptotic confidence intervals for each of the three examples described in Section 5, using the skew-power exponential model. Estimation is done via the method of moments (MM), and maximum likelihood (ML). (*) In this case there is no solution to the equation defining the MM estimators.*

We remark here that for the second example there is no solution to the set of equations determined by the MM, despite the fact that the sampling skewness and kurtosis are well within the ranges that correspond to the SPE distribution with $\beta = 0.5$ (see Table 2). In contrast, computing the MLE poses no practical difficulties, and in fact, $\hat{\beta}$ is close to the value 0.5 quoted above. This suggests that the MM needs to be used with caution, particularly in the case of highly skewed leptokurtic datasets.

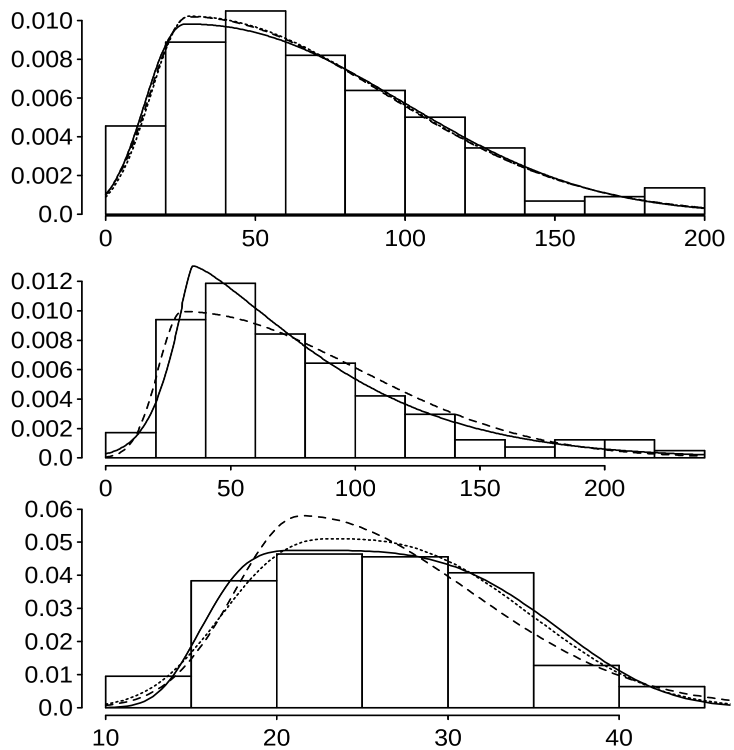

Figure 2: *Fitted SPE densities for each of the three examples considered. Estimation procedures are: maximum likelihood (solid line), method of moments (dotted line) and maximum likelihood restricted to $\beta = 0$ (dashed line). Top plot represents the volcano heights dataset, middle plot corresponds to the plasma ferritin concentration dataset and bottom plot is the age of Chilean women who gave birth in 1996. Note that for middle plot no estimated density could be obtained using the method of moments.*

Graphical display of the data and the estimated densities can be found in Figure 2. For the sake of comparison we also included the MLE of the skew-normal distribution obtained by imposing $\beta = 0$ in the general SPE distribution. For the first example, the three densities are nearly identical. The corresponding (asymptotic) confidence interval for β suggests the adequacy of the skew-normal distribution, for which estimation via MM works generally well. In contrast, for the plasma ferritin data, the MLE for the general SPE distribution is seen to yield the most satisfactory results, with β being significantly different from 0. In the third example we have again a value of β significantly negative, which is convenient for fitting the model to a platykurtic dataset.

6 Discussion and Extensions

We have introduced a new class of asymmetric distributions that is generated from a symmetric density f acting as a kernel function. When the asymmetry parameter α is such that the asymmetry functions are both equal to 1, the skew density reduces to symmetric kernel f. The stochastic representation given in Proposition 2 yields both, a simple interpretation, and a method to generate arbitrary members of the class.

The class of distributions (4) can be further extended as follows. Consider U_α as defined in Proposition 2, independent of

$$V \sim \frac{f(x+\beta)}{1-F(\beta)} I\{x \geq 0\},$$

and let $X = U_\alpha V$. This induces a new type of asymmetric densities, which we denote by $X \sim Sf2(\alpha, \beta)$, a two-parameter distribution with $\beta \in \mathbb{R}$. From this notation we immediately have that $Sf(\alpha) \equiv Sf2(\alpha, 0)$. It can be shown that when $\beta > 0$, the $Sf2(\alpha, \beta)$ distribution has more flexible asymmetries than (4), both to the left and the right. When $\beta < 0$ the shapes of the $Sf2(\alpha, \beta)$ densities are very similar to those obtained using (4).

Another relevant extension concerns the multivariate case. The basic idea consists of applying construction (4) to each coordinate of a density f with the property of being symmetric in each coordinate when holding the remaining ones fixed. We make this explicit for the case $a(\alpha) = 1 - \alpha$ and $b(\alpha) = 1 + \alpha$. Note first that (4) can be rewritten as $h(x|\alpha) = f(x/(1-\text{sgn}(x)\alpha))$, where $\text{sgn}(x) = x/|x|$ if $x \neq 0$ and $\text{sgn}(0) =$

0. Let now f be a k-variate density with the property that $f(x_i, \boldsymbol{x}_{-i}) = f(-x_i, \boldsymbol{x}_{-i})$ for all $i = 1, \ldots, k$, $x_i \in \mathbb{R}$, and $\boldsymbol{x}_{-i} \in \mathbb{R}^{k-1}$, where $\boldsymbol{x}_{-i}$ denotes the k-dimensional vector $\boldsymbol{x}$ with the ith coordinate removed. The class of k-variate skew distributions with asymmetry parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k) \in (-1, 1)^k$ is defined as

$$h(\boldsymbol{x}|\alpha) = f\left(\frac{x_1}{1 - \mathrm{sgn}(x_1)\alpha_1}, \ldots, \frac{x_k}{1 - \mathrm{sgn}(x_k)\alpha_k}\right), \qquad \boldsymbol{x} \in \mathbb{R}^k. \tag{17}$$

We denote a vector $\boldsymbol{X} = (X_1, \ldots, X_k)$ drawn from (17) as $\boldsymbol{X} \sim SM(f, k, \boldsymbol{\alpha})$. In the bivariate case (17) spells out as

$$\begin{aligned} h(x_1, x_2|\boldsymbol{\alpha}) = f\left(\frac{x_1}{1+\alpha_1}, \frac{x_2}{1+\alpha_2}\right) I\{x_1 < 0, x_2 < 0\} \\ + f\left(\frac{x_1}{1+\alpha_1}, \frac{x_2}{1-\alpha_2}\right) I\{x_1 < 0, x_2 \geq 0\} + f\left(\frac{x_1}{1-\alpha_1}, \frac{x_2}{1+\alpha_2}\right) I\{x_1 \geq 0, x_2 < 0\} \\ + f\left(\frac{x_1}{1-\alpha_1}, \frac{x_2}{1-\alpha_2}\right) I\{x_1 \geq 0, x_2 \geq 0\}. \end{aligned}$$

The stochastic representation from Proposition 2 can be also naturally extended to the multivariate $SM(f, k, \boldsymbol{\alpha})$ family. Indeed, using essentially the same proof, we can show that a random vector $\boldsymbol{X} = (X_1, \ldots, X_k) \sim SM(f, k, \boldsymbol{\alpha})$ can be expressed as $(U_1 \cdot V_1, \ldots, U_k \cdot V_k)$ where $\boldsymbol{U} = (U_1, \ldots, U_k)$ is independent of $\boldsymbol{V} = (V_1, \ldots, V_k)$. Moreover, the components of $\boldsymbol{U}$ are independent with $P(U_i = 1 - \alpha_i) = 1 - P(U_i = -1 - \alpha_i) = (1 - \alpha_i)/2$ for $i = 1, \ldots, k$ and $\boldsymbol{V} \stackrel{d}{=} (|Y_1|, \ldots, |Y_k|)$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_k) \sim f$. Based on the work by Arellano-Valle et al. (2002), we say that $\boldsymbol{X}$ is a $\mathcal{C}_\alpha$-vector, which extends the symmetric class of $\mathcal{C}$-random vectors to include asymmetry. Many properties of the $\mathcal{C}$ class can be easily extended to the $\mathcal{C}_\alpha$ class. For instance, the marginal distributions of a $\mathcal{C}_\alpha$-random vector belong to the $\mathcal{C}_\alpha$ class. Because the $\mathcal{C}$ class includes as special cases independence and spherical symmetry, it follows that the $\mathcal{C}_\alpha$ class contains many interesting cases and can be used for modelling purposes as a flexible and general class of distributions.

Acknowledgments

We thank Professor Guillermo Marshall, Pontificia Universidad Católica de Chile for facilitating us the dataset on births in Chile. We also thank a referee for the valuable comments and suggestions that helped to improve this manuscript.

References

Arellano-Valle, R. (1994). *Distribuçoes elípticas: propiedades, inferencia e aplicaçoes a modelos de regressao*, PhD thesis, IME-USP, BRASIL.

Arellano-Valle, R. B. and Genton, M. G. (2003). On fundamental skew distributions, *Technical Report 2551*, North Carolina State University, Institute of Statistics. Under review.

Arellano-Valle, R. B., del Pino, G. and San Martín, E. (2002). Definition and probabilistic properties of skew-distributions, *Statistics and Probability Letters* **58**: 111–121.

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

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

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society, Series B* **61**: 579–602.

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

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

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions, *Journal of Multivariate Analysis* **11**: 368–385.

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

Fang, K.-T., Kotz, S. and Ng, K.-W. (1990). *Symmetric multivariate and related distributions*, London: Chapman-Hall.

Fechner, G. T. (1897). *Kollectivmasslehre*, Leipzig: Engleman.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness, *Journal of the American Statistical Association* **93**: 359–371.

Fernández, C., Osiewalski, J. and Steel, M. F. J. (1995). Modeling and inference with v-spherical distributions, *Journal of the American Statistical Association* **90**: 1331–1340.

Genton, M. G. and Loperfido, N. (2002). Generalized skew-elliptical distributions and their quadratic forms, *Technical Report 2539*, North Carolina State University, Institute of Statistics. Under review.

Genton, M. G., He, L. and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms, *Statistics and Probability Letters* **51**: 319–325.

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

Kelker, D. (1970). Distribution theory of special distributions and location-scale parameters, *Sankhya Ser. A* **32**: 319–340.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*, New York: Springer.

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