Flexible Class of Skew-Symmetric Distributions

YANYUAN MA and MARC G. GENTON

North Carolina State University

ABSTRACT. We propose a flexible class of skew-symmetric distributions for which the probability density function has the form of a product of a symmetric density and a skewing function. By constructing an enumerable dense subset of skewing functions on a compact set, we are able to consider a family of distributions, which can capture skewness, heavy tails and multimodality systematically. We present three illustrative examples for the fibreglass data, the simulated data from a mixture of two normal distributions and the Swiss bills data.

Key words: dense subset, generalized skew-elliptical, multimodality, skewness, skew-normal

1. Introduction

A popular approach to achieve departures from normality consists of modifying the probability density function (pdf) of a random vector in a multiplicative fashion. Wang *et al.* (2004) showed that any p-dimensional multivariate pdf $g(\mathbf{x})$ admits, for any fixed location parameter $\xi \in \mathbb{R}^p$, a unique skew-symmetric (SS) representation:

$$g(\mathbf{x}) = 2f(\mathbf{x} - \xi)\pi(\mathbf{x} - \xi), \tag{1}$$

where $f : \mathbb{R}^p \to \mathbb{R}_+$ is a symmetric pdf and $\pi : \mathbb{R}^p \to [0, 1]$ is a skewing function satisfying $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$. Conversely, any function g of the type defined by (1) is a valid pdf. By symmetric, we mean $f(\mathbf{x}) = f(-\mathbf{x})$ and we will use 'symmetric pdf' and the property $f(\mathbf{x}) = f(-\mathbf{x})$ interchangeably in the sequel. Throughout this paper, we restrict our interest on functions $f \in C_0(\mathbb{R}^p)$ and continuous skewing functions $\pi(\mathbf{x})$, where $C_0(\mathbb{R}^p)$ denotes continuous functions on $\mathbb{R}^p$ with the property $f(\mathbf{x}) \to 0$ when $\|\mathbf{x}\|_2 \to \infty$, and $\|\cdot\|_2$ denotes the L^2 norm. Genton & Loperfido (2002) considered the subfamily of generalized skew-elliptical (GSE) distributions for which the pdf f in (1) is elliptically contoured rather than only symmetric. Many definitions of skewed distributions found in the literature can be written in the form of an SS distribution (1). For instance, Azzalini & Dalla Valle's (1996) multivariate skew-normal distribution corresponds to $f(\mathbf{x}) = \phi_p(\mathbf{x}; \mathbf{0}, \Omega)$ and $\pi(\mathbf{x}) = \Phi(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, where $\phi_p(\mathbf{x}; \boldsymbol{\mu}, \Omega)$ is the p-dimensional multivariate normal pdf with mean vector $\boldsymbol{\mu}$ and correlation matrix Ω, Φ is the standard normal cumulative distribution function (cdf), and $\boldsymbol{\alpha}$ is a shape parameter controlling skewness. Similarly, multivariate distributions such as skew-t (Branco & Dey, 2001; Azzalini & Capitanio, 2003; Jones & Faddy, 2003; Sahu *et al.*, 2003), skew-Cauchy (Arnold & Beaver, 2000) and other skew-elliptical ones (Azzalini & Capitanio, 1999; Branco & Dey, 2001; Sahu *et al.*, 2003) can be represented by the SS distribution (1) with appropriate choices of f and π.

In this article, we propose a flexible class of distributions (1) by constructing an enumerable dense subset of the skewing functions π on a compact set. The result is a family of distributions which can systematically capture skewness, heavy tails and multimodality. The construction of the subset is through polynomials, which has a similar flavour as the semi-non-parametric (SNP) representation proposed by Gallant & Nychka (1987). The latter is defined as the product of the standard normal pdf and the square of a polynomial. The SNP distribution requires the coefficients in the polynomial to be constrained in order to yield a valid density. It

also relies on rejection sampling schemes to simulate random samples. These difficulties do not occur with our construction.

The paper is organized as follows. In section 2, we describe a subset of skewing functions based on odd polynomials and prove that it results in a dense subset of the SS distributions. In particular, we define flexible skew-normal and skew-t distributions that can have more than one mode. This is an essential property for some situations and provides an alternative to modelling with mixtures of distributions. The flexibility and possible multimodality of the new class of distributions is illustrated in section 3. We present three illustrative examples in section 4, and a discussion in section 5.

2. A dense subset of skew-symmetric distributions

In this section, we construct a dense subset of SS distributions through approximating the skewing function π on a compact set. Any continuous skewing function π can be written as:

$$\pi(\mathbf{x}) = H(w(\mathbf{x})), \tag{2}$$

where $H : \mathbb{R} \to [0, 1]$ is the cdf of a continuous random variable symmetric around 0, and $w : \mathbb{R}^p \to \mathbb{R}$ is an odd continuous function, that is $w(-\mathbf{x}) = -w(\mathbf{x})$. In fact, for a chosen H such that H^{-1} exists, $w(\mathbf{x}) = H^{-1}(\pi(\mathbf{x}))$ is a continuous odd function. This representation has been used by Azzalini & Capitanio (2003) to define certain distributions by perturbation of symmetry. Note, however, that the representation (2) is not unique due to the many possible choices of H.

Let $P_K(\mathbf{x})$ be an odd polynomial of order K. A polynomial of order K in $\mathbb{R}^p$ is defined as a linear combination of terms of the form $\prod_{i=1}^{p} x_i^{r_i}$, where $k = \sum_{i=1}^{p} r_i \leq K$. If each term has an odd order (all ks are odd), then the polynomial is called an odd polynomial, whereas if each term has an even order (all ks are even), it is called an even polynomial. We define flexible skew-symmetric (FSS) distributions by restricting (1) to:

$$2f(\mathbf{x} - \boldsymbol{\xi})\pi_K(\mathbf{x} - \boldsymbol{\xi}), \tag{3}$$

where $\pi_K(\mathbf{x}) = H(P_K(\mathbf{x}))$ and H is any cdf of a continuous random variable symmetric around 0. Note that there are no constraints on the coefficients of the polynomial P_K in order to make (3) a valid pdf. In particular, (3) defines flexible generalized skew-elliptical (FGSE) distributions when the pdf f is elliptically contoured. For instance, flexible generalized skew-normal (FGSN) distributions are defined by:

$$2\phi_p(\mathbf{x}; \boldsymbol{\xi}, \Omega)\Phi(P_K(A(\mathbf{x} - \boldsymbol{\xi}))), \tag{4}$$

and flexible generalized skew-t (FGST) distributions are defined by:

$$2t_p(\mathbf{x}; \boldsymbol{\xi}, \Omega, \boldsymbol{v})T(P_K(A(\mathbf{x} - \boldsymbol{\xi})); \boldsymbol{v}), \tag{5}$$

where we use the Choleski decomposition $\Omega^{-1} = A^{\mathrm{T}}A$, t_p denotes a p-dimensional multivariate t pdf, and T denotes a univariate t cdf, both with degrees of freedom $\boldsymbol{v}$. Note that we could use Φ, or any other symmetric cdf, instead of T for the skewing function in (5). In practice, a popular choice for the cdf H would be Φ or the univariate cdf corresponding to the symmetric pdf f. Effectively, the following proposition shows that FSS distributions can approximate SS distributions arbitrarily well.

Proposition 1
Let the class of FSS distributions consist of distributions with pdf given in (3) and the class of SS distributions of distributions with pdf given in (1), where $f \in C_0(\mathbb{R}^p)$ in both classes and π is continuous. Then the class of FSS distributions is dense in the class of SS distributions under the L^∞ norm.

Proof. An arbitrary distribution in the SS class can be written as $2f(\mathbf{x} - \xi)H(w(\mathbf{x} - \xi))$, where f and H are continuous, H^{-1} exists, and w is a continuous odd function. Because $f \in C_0(\mathbb{R}^p)$, for any arbitrary $\epsilon > 0$, we can find a compact set D which is symmetric around ξ (if $\mathbf{x} - \xi \in D$ then $\xi - \mathbf{x} \in D$), such that for any $\mathbf{x} - \xi \notin D, f(\mathbf{x} - \xi) < \epsilon/4$. Thus, for any $\mathbf{x} - \xi \notin D, |2f(\mathbf{x} - \xi)\pi(\mathbf{x} - \xi) - 2f(\mathbf{x} - \xi)H(P(\mathbf{x} - \xi))| < \epsilon$ for any odd polynomial P.

As f is continuous, f is bounded on D. We denote the bound by C, i.e. $f(\mathbf{x} - \xi) \leq C$ for any $\mathbf{x} - \xi \in D$. We use D_1 to denote the image space of w, i.e. $D_1 = \{w(\mathbf{x}) | \mathbf{x} \in D\}$. Because of the continuity of w, which is a result of the continuity of both H and π, D_1 is also compact. The continuous function H is uniformly continuous on the compact set D_1. Hence there exists $\eta > 0$ such that for any $y_1, y_2 \in D_1$ and $|y_1 - y_2| < \eta$, we get $|H(y_1) - H(y_2)| < \epsilon/(2C)$. From the Stone–Weierstrass theorem (see e.g. Rudin, 1973, p. 115), there exists a polynomial P such that $|w(\mathbf{x} - \xi) - P(\mathbf{x} - \xi)| < \eta$ for any $\mathbf{x} - \xi \in D$. We decompose P into an even term P_e and an odd term P_o, i.e. $P = P_e + P_o$. Then $|w(\mathbf{x} - \xi) - P_e(\mathbf{x} - \xi) - P_o(\mathbf{x} - \xi)| < \eta$ and $|w(\xi - \mathbf{x}) - P_e(\xi - \mathbf{x}) - P_o(\xi - \mathbf{x})| < \eta$. Because w and P_o are odd, and P_e is even, we get $|-w(\mathbf{x} - \xi) - P_e(\mathbf{x} - \xi) + P_o(\mathbf{x} - \xi)| < \eta$. Notice that $2|w(\mathbf{x} - \xi) - P_o(\mathbf{x} - \xi)| \leq |w(\mathbf{x} - \xi) - P_e(\mathbf{x} - \xi) - P_o(\mathbf{x} - \xi)| + |-w(\mathbf{x} - \xi) - P_e(\mathbf{x} - \xi) + P_o(\mathbf{x} - \xi)| < 2\eta$, so $|w(\mathbf{x} - \xi) - P_o(\mathbf{x} - \xi)| < \eta$. Combining these results, we know that for an arbitrary member $2f(\mathbf{x} - \xi)H(w(\mathbf{x} - \xi))$ in SS and an arbitrary $\epsilon > 0$, we can find a member $2f(\mathbf{x} - \xi)H(P_o(\mathbf{x} - \xi))$ in FSS such that $|2f(\mathbf{x} - \xi)H(w(\mathbf{x} - \xi)) - 2f(\mathbf{x} - \xi)H(P_o(\mathbf{x} - \xi))| < \epsilon$ for any $\mathbf{x} - \xi \in D$.

Hence FSS is dense in SS with respect to the L^∞ norm.

Remark 1. The requirement $f \in C_0(\mathbb{R}^p)$ in Proposition 1 can be relaxed to allow that f has a finite number, m say, of poles. In this case, FSS is dense in SS with respect to almost uniform convergence (uniform in a set whose complement is of measure arbitrarily small). Indeed, let $\mathbb{R}^p(r)$ denote $\mathbb{R}^p$ minus the union of m open balls of radius r centred at the m poles. Then FSS is dense in SS on $\mathbb{R}^p(r)$ under the L^∞ norm. Letting $r \to 0$, the result follows.

Proposition 1 shows in particular that the class of GSE, skew-t, and skew-normal distributions can be approximated arbitrarily well by their flexible versions.

3. Flexibility and multimodality

In Fig. 1, we illustrate the shape flexibility of the FGSN distribution in the univariate case. Its pdf for $K = 3$ is defined by:

$$2\phi_1(x; \xi, \sigma^2)\Phi(\alpha(x - \xi)/\sigma + \beta(x - \xi)^3/\sigma^3). \tag{6}$$

Figure 1(a) depicts the pdf of the FGSN model for $\xi = 0$, $\sigma^2 = 1$, $\alpha = 4$, and $\beta = 0$, i.e. it reduces to Azzalini's (1985) univariate skew-normal distribution. However, when $\beta \neq 0$, the pdf (6) can exhibit bimodality as shown in Fig. 1(b) with $\alpha = 1$, and $\beta = -1$. In general, as the degree K of the odd polynomial in the skewing function becomes large, the number of modes allowed in the pdf increases, thus inducing a greater flexibility in the available shapes. Unfortunately, the number of modes depends on the degree K of the odd polynomial, on the symmetric pdf f, and on the cdf H of the skewing function π_K in a complex fashion. Indeed, even for the univariate situation given by $p = 1$, the modes are determined by zeros of the first derivative of the FSS distribution (3) given by:
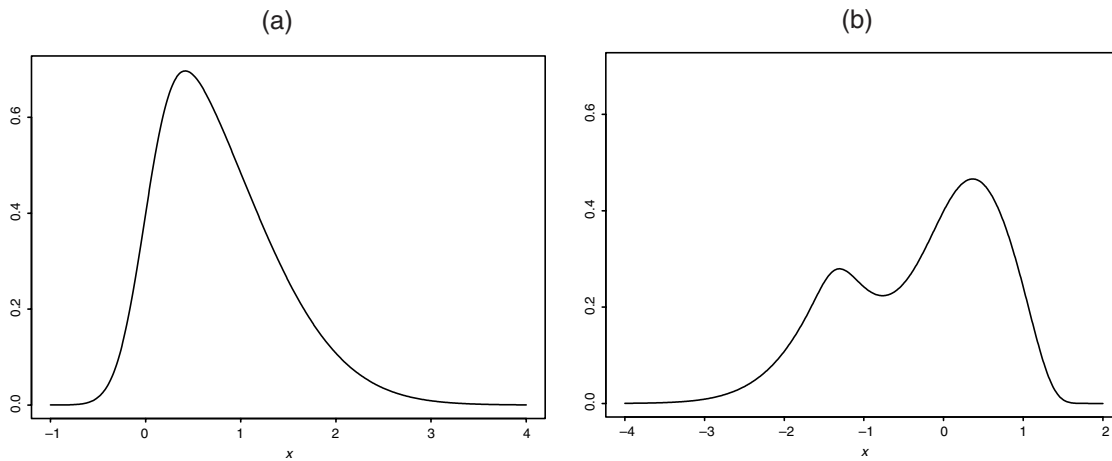
Fig. 1. Two members of the univariate FGSN family of distributions with $K = 3$, $\xi = 0$, $\sigma^2 = 1$: (a) $\alpha = 4$, $\beta = 0$ (skew-normal); (b) $\alpha = 1$, $\beta = -1$.

$$2f'(x)H(P_K(x)) + 2f(x)H'(P_K(x))P_K'(x), \tag{7}$$

for which the number of zeros cannot be easily computed. Even with restrictions to some specific f and H functions, a general statement on the relation between the number of modes and the order of the polynomial seems unavailable. However, in the univariate case, if we consider a normal pdf $f = \phi_1$ and a standard normal cdf $H = \Phi$ with an odd polynomial of order $K = 3$, we have the following proposition.

Proposition 2

The class of FGSN distributions with pdf $2\phi_1(x;\xi,\sigma^2)\Phi(\alpha(x - \xi)/\sigma + \beta(x - \xi)^3/\sigma^3)$ has at most two modes.

Proof. Without loss of generality, we can set $\xi = 0$, $\sigma = 1$, assume $\beta > 0$, and only need to prove that $\psi(x) = 2\phi(x)\Phi(\alpha x + \beta x^3)$ has at most two modes. We prove this by contradiction. If $\psi(x)$ has more than two modes, then $\psi'(x)$ has at least five zeros. In the following proof, we show that this cannot be the case. We have $\psi'(x) = 2\phi(x)((\alpha + 3\beta x^2)\phi(\alpha x + \beta x^3) - x\Phi(\alpha x + \beta x^3))$ and need to consider three cases:

Case 1 ($\alpha = 0$). We write $\psi'(x) = 2x\phi(x)\eta(x)$, where $\eta(x) = 3\beta x\phi(\beta x^3) - \Phi(\beta x^3)$. We can verify that $\eta'(x) = 3\beta\phi(\beta x^3)\eta_1(y)$, where $y = x^2$ and $\eta_1(y) = 1 - y - 3\beta^2 y^3$. As $\eta_1(y)$ is a decreasing function on $y \geq 0$, $\eta'(x)$ has at most two zeros. Thus, $\eta(x)$ has at most three zeros, hence $\psi'(x)$ has at most four zeros.

Case 2 ($\alpha > 0$). Notice that $\psi'(x) > 0$ for $x \leq 0$. For $\gamma_1(x) = \psi'(x)/(2x\phi(x)) = \phi(\alpha x + \beta x^3)(\alpha + 3\beta x^2)/x - \Phi(\alpha x + \beta x^3)$, we get $\gamma_1'(x) = \phi(\alpha x + \beta x^3)/(-9\beta x^2)\gamma_2(y)$, where $y = \alpha + 3\beta x^2 > 0$ and $\gamma_2(y) = y^4 + \alpha y^3 + (3 - 2\alpha^2)y^2 - (3\alpha + 9\beta)y + 18\alpha\beta$. As $\gamma_2''(y) = 12y^2 + 6\alpha y + (6 - 4\alpha^2)$ has at most one positive zero, and $\gamma_2'(y) = 4y^3 + 3\alpha y^2 + (6 - 4\alpha^2)y - (3\alpha + 9\beta) < 0$ at $y = 0$, we know that $\gamma_2'(y)$ has at most one positive zero. Thus $\gamma_2(y)$ has at most two positive zeros. This means $\gamma_1'(x)$ has at most two positive zeros, so $\psi'(x)$ has at most three positive zeros.

Case 3 ($\alpha < 0$). Notice that $\psi'(x) < 0$ for $x \in [0, \sqrt{-\alpha/(3\beta)}]$ and $\psi'(x) > 0$ for $x \in (-\infty, -\sqrt{-\alpha/(3\beta)}]$. So we look only for solutions $x \in (\sqrt{-\alpha/(3\beta)}, \infty)$ and $x \in (-\sqrt{-\alpha/(3\beta)}, 0)$. Let $y = \alpha + 3\beta x^2$, then there is a one-to-one mapping between the x in the above range and $y \in (\alpha, \infty)$. Let $\gamma_1(x)$ and $\gamma_2(y)$ have the same expressions as in case 2. We

have that $\gamma_2(y)$ has at most four zeros as it is a fourth-order polynomial. Notice that $\gamma_2(\alpha) < 0$, $\gamma_2(-\infty) > 0$, so $\gamma_2(y)$ has at most three zeros in (α, ∞). This means $\gamma_1'(x)$ has at most three zeros, hence $\psi'(x)$ has at most four zeros.

Figure 1 illustrates the result of proposition 2 by depicting a unimodal and a bimodal pdf from the univariate FGSN with $K = 3$. For $K = 1$, the pdf is always unimodal as was already noted by Azzalini (1985) for the univariate skew-normal distribution.

Next we investigate the flexibility of the FGSN distribution in the bivariate case. Its pdf for $K = 3$, $\xi = 0$, and $\Omega = I_2$ is given by:

$$2\phi_2(x_1, x_2; \mathbf{0}, I_2)\Phi(\alpha_1 x_1 + \alpha_2 x_2 + \beta_1 x_1^3 + \beta_2 x_2^3 + \beta_3 x_1^2 x_2 + \beta_4 x_1 x_2^2). \tag{8}$$

Figure 2 depicts the contours of four different pdfs (8) for various combinations of values of the skewness parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$ and β_4. In particular, for $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, the pdf is exactly the bivariate skew-normal proposed by Azzalini & Dalla Valle (1996), and known to be unimodal (see Fig. 2a). However, Figs 2(b–d) show that many different distributional shapes can be obtained with the parameters $\beta_1, \ldots, \beta_4$, in particular bimodal and trimodal distributions. Additional flexibility can be imposed on the tail behaviour by choosing pdfs other than the normal for the symmetric pdf f, for example, a t distribution. This yields FGST distributions and will prove useful for applications as they can allow for both fat tails and skewness (see section 4, example 1).

Finally, note that the stochastic representation of FSS distributions follows from the stochastic representation of SS distributions described by Wang *et al.* (2004), see also Azzalini & Capitanio (2003). It provides a quick way to generate pseudo-realizations from the FSS distribution (3).

4. Model fitting and examples

In this section, we present three applications of FGSE distributions. We carry out the estimation and model fitting by maximizing the likelihood function corresponding to (3) for a given order K. Unlike SNP distributions, there are no constraints on the parameters of the skewing function π_K and standard optimization techniques are used. To avoid local maxima, we carry out the optimization routine with several different starting values that are widely scattered in the feasible region. The order K is chosen adaptively via model selection strategies. Because for a given symmetric pdf f and skewing function π_K the models induced by (3) are nested when K decreases, likelihood ratio tests (LRT) can be used to identify an appropriate value of K. Model selection criteria such as AIC (twice the log-likelihood minus twice the number of parameters) and BIC (twice the log-likelihood minus the number of parameters times the logarithm of the sample size) can be used as well. In practice, $K = 3$ seems to provide enough flexibility to model unimodal and bimodal pdfs.

4.1. Example 1: fibreglass data

This example is concerned with a unidimensional data set of breaking strengths values of 1.5 cm long glass fibres. Jones & Faddy (2003) and Azzalini & Capitanio (2003) fit two forms of skew-t distributions to these data. They both noted skewness on the left as well as heavy tail behaviour.

We fit an FGST distribution (5) with pdf:

$$2t_1(x; \xi, \sigma^2, v)T(P_K((x - \xi)/\sigma); v), \tag{9}$$

for $K = 1$ and $K = 3$. The fitted parameters, obtained by maximizing the corresponding likelihood function, are listed in Table 1.
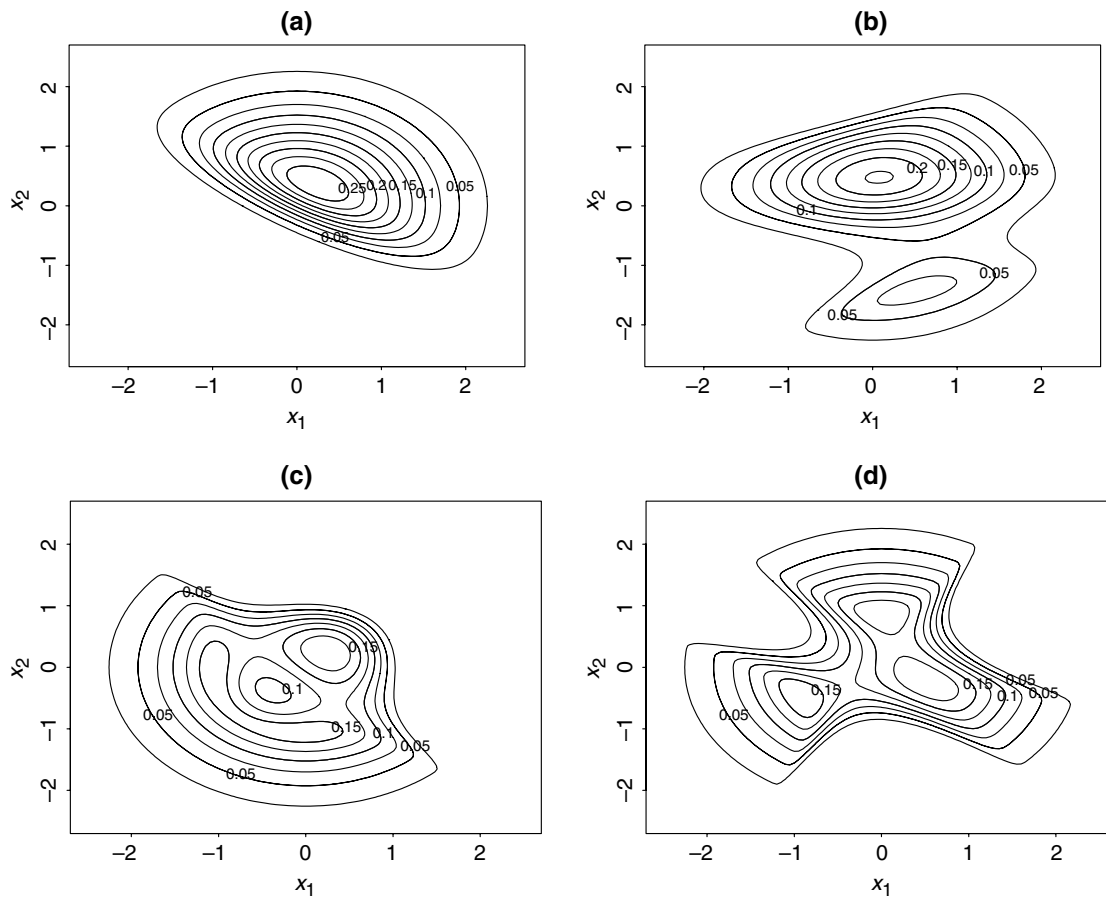
Fig. 2. Four members of the bivariate FGSN family of distributions with $K = 3$, $\boldsymbol{\xi} = \mathbf{0}$ and $\Omega = I_2$: (a) $\alpha_1 = 2$, $\alpha_2 = 3$, $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$, $\beta_4 = 0$ (skew-normal; unimodal); (b) $\alpha_1 = 0$, $\alpha_2 = 2$, $\beta_1 = 0$, $\beta_2 = -1$, $\beta_3 = 0$, $\beta_4 = 1$ (bimodal); (c) $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = -2$, $\beta_2 = -2$, $\beta_3 = -1$, $\beta_4 = -1$ (bimodal); (d) $\alpha_1 = 1$, $\alpha_2 = 0$, $\beta_1 = -1$, $\beta_2 = 2$, $\beta_3 = -4$, $\beta_4 = -1$ (trimodal).

Note the small values for $\hat{\nu}$, indicating tails much heavier than the normal distribution. The fitted pdfs are depicted in Fig. 3 for $K = 1$ (solid line) and $K = 3$ (dotted line), along with a histogram of the fibreglass data.

There appear not to be too much difference between the two models. We use an LRT for the null hypothesis $H_0 : \beta = 0$ with the approximate asymptotic distribution χ_1^2 and use the AIC, BIC criteria for model selection. The results are tabulated in Table 1. All three methods favour the FGST model with $K = 1$, which does not allow for bimodality.

4.2. Example 2: mixture of normals

This example illustrates that FSS distributions can be used as an alternative to mixtures of distributions in situations where multimodality is desirable. We consider a simulated data set of size 100 from a mixture of two bivariate normal distributions:

Table 1. *Fitted values of the univariate FGST model for $K = 1$ and $K = 3$ for the fibreglass data, as well as model selection criteria*

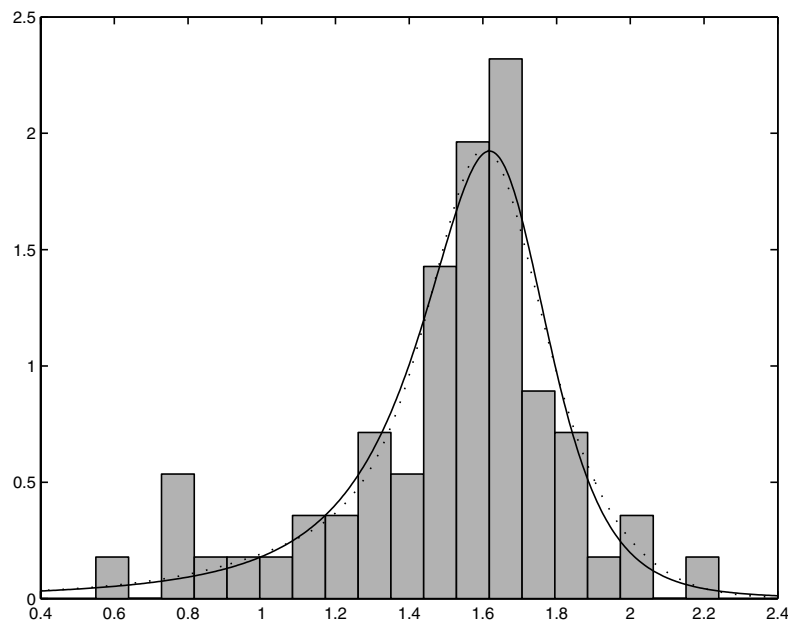|	$\hat{\xi}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$	LRT (*p*-value)	AIC	BIC
$K = 1$	1.67	5.14	−0.60	–	2.05	–	−31.9	−40.4
$K = 3$	1.60	5.43	0.07	−0.04	2.11	0.42	−33.2	−43.9

Fig. 3. Histogram of the fibreglass data and fitted pdfs of the FGST model with $K = 1$ (solid line) and $K = 3$ (dotted line) by maximum likelihood.

$$(1 - \varepsilon)N_2(\boldsymbol{\mu}_1, \Sigma_1) + \varepsilon N_2(\boldsymbol{\mu}_2, \Sigma_2), \tag{10}$$

with $\varepsilon = 0.4$, $\boldsymbol{\mu}_1 = (0,0)^{\mathrm{T}}$, $\boldsymbol{\mu}_2 = (5,4)^{\mathrm{T}}$ and

$$\Sigma_1 = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}. \tag{11}$$

Figure 4(a) depicts the 100 simulated data along with the bivariate contours of the pdf corresponding to (10), which shows bimodality.

We fit a bivariate FGSN distribution (4) with $K = 1$ and $K = 3$. The parameters, estimated by maximizing the corresponding likelihood function, are listed in Table 2.

The contours of the fitted bivariate pdfs are depicted in Fig. 4(b) for $K = 1$ and in Fig. 4(c) for $K = 3$. The case $K = 1$ corresponds to Azzalini & Dalla Valle's (1996) bivariate skew-normal distribution, which cannot capture the bimodality. The fit with $K = 3$ captures the bimodality and adapts closely to the shape of the simulated data. We test the model with the LRT and AIC, BIC criteria, and find that all three select the model with $K = 3$. We further fit a model with $K = 5$ and find that the LRT and AIC score indicate that $K = 5$ is a better fit for the data, while BIC suggests that $K = 5$ imposes too much model complexity for the gain. We decide to adopt a more complex model only when all three methods indicate so, hence we keep $K = 3$ as our final model. The results of the model selection are tabulated in Table 3.

4.3. Example 3: old Swiss 1000 franc bills data

This example consists of measurements on 100 genuine and 100 forged old Swiss 1000 franc bills analysed by Flury & Riedwyl (1988). We consider two variables: the distance X_1 from the inner frame to the lower border, and the length X_2 of the diagonal of the bills.

Figure 5 depicts scatter plots of (X_1, X_2) along with fitted FGSN pdfs for various orders of the polynomial P_K in the skewing function. Panel (a) represents a fit with $K = 1$ which corresponds to Azzalini & Dalla Valle's (1996) bivariate skew-normal pdf and implies
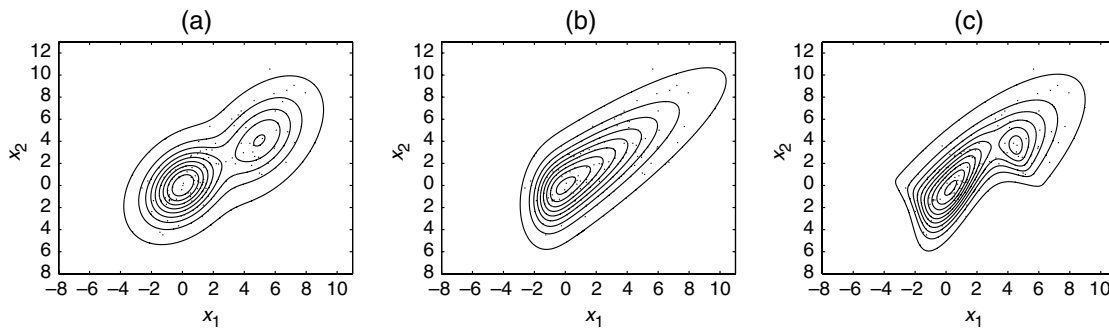
Fig. 4. Simulated data set of size 100 from a mixture of two normal distributions, with contours of the corresponding bivariate pdf in panel (a). Contours of the bivariate fitted pdf from the FGSN model with $K = 1$ in panel (b) and with $K = 3$ in panel (c).

Table 2. *Fitted values of the bivariate FGSN model for K = 1 and K = 3 for the simulated mixture data*

	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{22}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$K = 1$	−1.64	−1.93	0.36	−0.29	0.28	2.42	6.84	–	–	–	–
$K = 3$	1.72	2.08	0.42	−0.26	0.27	−0.78	−1.60	0.21	0.40	0.31	0.56

Table 3. *Model selection criteria for K = 1, 3, 5 for the simulated mixture data*

	LRT (*p*-value)	AIC	BIC
$K = 1$	–	−898.7	−916.9
$K = 3$	0.0009	−888.0	−916.7
$K = 5$	0.0085	−882.8	−927.0

unimodality. Therefore, this fit cannot capture the difference between the genuine and forged bills. Panel (b) represents a fit with $K = 3$ which takes into account the bimodality arising from the two types of bills. A fit with $K = 5$ is presented in panel (c). It shows a complex distributional pattern in the data and detects a third mode. Although this additional mode appears to be located over a hole of the scatter plot, it suggests a possible third group near that region. This pattern has also been discovered by various non-parametric kernel density estimators, see Simonoff (1996, pp. 111–114). LRT and AIC favour the model with $K = 5$ while BIC favours $K = 3$. We keep the model with $K = 3$ in panel (b), that is the bimodal pdf which detects the two groups of genuine and forged Swiss bills.
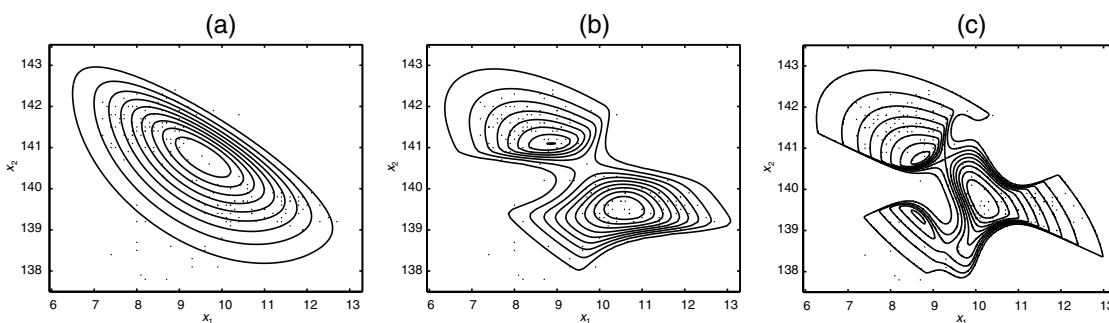

Fig. 5. Contours of the bivariate FGSN pdf fitted by maximum likelihood for the Swiss bills data. (a) $K = 1$; (b) $K = 3$; (c) $K = 5$.

5. Discussion

We have investigated FSS distributions, a flexible class that can take skewness, heavy tails and multimodality, into account. It is based on an approximation of the skewing function by a univariate symmetric cdf evaluated at odd polynomials. Using the Stone–Weierstrass theorem, we have proved that this approximation can be made arbitrarily accurate by increasing the order of the odd polynomial. However, the number of coefficients of the polynomial increases quickly with its order K, especially when the dimension p of the distribution is also large. In this case, Monte Carlo Markov chain methods might be more appropriate than direct likelihood maximization. In light of the examples in section 4, it seems that $K = 3$ is sufficient for unimodal and bimodal practical applications.

The choice of the symmetric pdf f is of practical importance. Although it has been shown that any multivariate pdf g can be represented by an SS distribution, a parametric form for the pdf f needs to be specified for our applications. It turns out that the normal and the t pdfs are the most natural ones, yielding the flexible models FGSN and FGST. The latter model is particularly well suited to capture heavy tails, possibly Cauchy-like, in addition to skewness and multimodality. More sophisticated symmetric pdfs f could be used as well, e.g. such as the slash distribution. The choice of the cdf H has theoretically no impact, but in practice it can influence the order K of the polynomial P_K. For applications, H should be chosen to facilitate computations. A natural choice is $H = \Phi$ or the cdf corresponding to the symmetric pdf f.

References

Arnold, B. C. & Beaver, R. J. (2000). The skew-Cauchy distribution. *Statist. Probab. Lett.* **49**, 285–290.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.

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

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Statist. Soc. Ser. B* **61**, 579–602.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc. Ser. B* **65**, 367–389.

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* **79**, 99–113.

Flury, B. & Riedwyl, H. (1988). *Multivariate statistics, a practical approach*. Cambridge University Press, Cambridge.

Gallant, A. R. & Nychka, D. W. (1987). Seminonparametric maximum likelihood estimation. *Econometrica* **55**, 363–390.

Genton, M. G. & Loperfido, N. (2002). Generalized skew-elliptical distributions and their quadratic forms. *Institute of Statistics Mimeo Series No. 2539*, under review (http://www.stat.ncsu.edu/~mggenton/publications.html).

Jones, M. C. & Faddy, M. J. (2003). A skew extension of the t distribution, with applications. *J. Roy. Statist. Soc. Ser. B* **65**, 159–174.

Rudin, W. (1973). *Functional analysis*. McGraw-Hill, Inc. New York.

Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Statist.* **31**, 129–150.

Simonoff, J. S. (1996). *Smoothing methods in statistics*. Springer, New York.

Wang, J., Boyer, J. & Genton, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica*, in press.

Received January 2003, in final form January 2004

Yanyuan Ma, CRSC, North Carolina State University, Box 8205, Raleigh, NC 27695-8205, USA.
 E-mail: yma@unity.ncsu.edu