Biometrika (2009), **96**, 4, *pp.* 761–780
doi: 10.1093/biomet/asp053

Advance Access publication 25 October 2009
Printed in Great Britain

Sinh-arcsinh distributions

BY M. C. JONES

Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, U.K.

m.c.jones@open.ac.uk

AND ARTHUR PEWSEY

Department of Mathematics, Escuela Politécnica, University of Extremadura, Avenida de la Universidad s/n, 10071 Cáceres, Spain

apewsey@unex.es

SUMMARY

We introduce the sinh-arcsinh transformation and hence, by applying it to a generating distribution with no parameters other than location and scale, usually the normal, a new family of sinh-arcsinh distributions. This four-parameter family has symmetric and skewed members and allows for tailweights that are both heavier and lighter than those of the generating distribution. The central place of the normal distribution in this family affords likelihood ratio tests of normality that are superior to the state-of-the-art in normality testing because of the range of alternatives against which they are very powerful. Likelihood ratio tests of symmetry are also available and are very successful. Three-parameter symmetric and asymmetric subfamilies of the full family are also of interest. Heavy-tailed symmetric sinh-arcsinh distributions behave like Johnson S_U distributions, while their light-tailed counterparts behave like sinh-normal distributions, the sinh-arcsinh family allowing a seamless transition between the two, via the normal, controlled by a single parameter. The sinh-arcsinh family is very tractable and many properties are explored. Likelihood inference is pursued, including an attractive reparameterization. Illustrative examples are given. A multivariate version is considered. Options and extensions are discussed.

Some key words: Heavy tail; Johnson's S_U distribution; Light tail; Sinh-normal distribution; Skew-normal distribution; Skewness; Transformation.

1. INTRODUCTION

Families of distributions with four parameters, accounting for location, scale and, in some senses, skewness and tailweight, cover many of the most important aspects of any unimodal distribution on $\mathbb{R}$. They can be used as the random parts of regression-type models where, typically, they allow potentially complex modelling of location, and perhaps scale, parameters while acting robustly with respect to asymmetry and weight of tails. The Pearson and Johnson families of distributions contain famous examples (Johnson et al., 1994, Chapter 12); stable laws (Samorodnitsky & Taqqu, 1994), generalized hyperbolic distributions (Barndorff-Nielsen, 1978), Tukey's g-and-h distributions (Hoaglin, 1986), two-piece distributions (Fernández & Steel, 1998), generalized distributions of order statistics (Jones, 2004) and a popular class of skew distributions in which a symmetric density is multiplied by a rescaled symmetric distribution function (Azzalini, 1985; Genton, 2004) are other examples. Many more families live on finite or semi-infinite support.

Broadly, most such families of distributions have the normal distribution as a special, often limiting, case with other family members having heavier tails than the normal. Exceptions are the exponential power and skew-exponential power distributions (Box & Tiao, 1973; DiCiccio & Monti, 2004). Here, we propose a novel relatively simple and tractable four-parameter family of distributions on $\mathbb{R}$, which also has the normal distribution situated centrally and contains other members with both lighter and heavier tails. Practical benefits include excellent tests of the appropriateness of the normal distribution.

Consider the canonical case of the new distributions in which location $\xi \in \mathbb{R}$ and scale $\eta > 0$ are removed; they can be reinstated for practical work in the usual way through $\eta^{-1} f_{\epsilon,\delta}\{\eta^{-1}(x-\xi)\}$, where $f_{\epsilon,\delta}(\cdot)$ is the density of a member of the new family. Here, $\epsilon \in \mathbb{R}$ will turn out to be a skewness parameter and $\delta > 0$ will control the tailweight. Associate random variables Z and $X_{\epsilon,\delta}$ with the standard normal density ϕ and $f_{\epsilon,\delta}$, respectively. Then, we define $f_{\epsilon,\delta}$ by the sinh-arcsinh transformation

$$Z = S_{\epsilon,\delta}(X_{\epsilon,\delta}) \equiv \sinh\{\delta \sinh^{-1}(X_{\epsilon,\delta}) - \epsilon\}. \tag{1}$$

It follows that the density of the sinh-arcsinh distribution is given by

$$f_{\epsilon,\delta}(x) = \{2\pi(1+x^2)\}^{-1/2} \delta C_{\epsilon,\delta}(x) \exp\{-S^2_{\epsilon,\delta}(x)/2\}, \tag{2}$$

where $C_{\epsilon,\delta}(x) = \cosh\{\delta \sinh^{-1}(x) - \epsilon\} = \{1 + S^2_{\epsilon,\delta}(x)\}^{1/2}$. It turns out that skewness increases with increasing ϵ, positive skewness corresponding to $\epsilon > 0$. Tailweight decreases with increasing δ, $\delta < 1$ yielding heavier tails than the normal distribution, $\delta > 1$ yielding lighter tails. Of course, $f_{0,1}(x) = \phi(x)$ and $f_{0,\delta}(x)$ is a new family of symmetric distributions. Examples of densities (2) are given in Fig. 1. No special functions appear in the density of the sinh-arcsinh distribution.

Further details of what follows, especially of the simulations of §§ 6 and 7, can be found in Open University Technical Report #08/06, available at http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm, which we refer to as our technical report hereafter. Inter alia, our technical report works in terms of $-\epsilon$ rather than ϵ.

2. PROPERTIES OF FAMILY (2)

2·1. *Basic properties*

We begin by noting several equivalent formulations of transformation (1):

$$\begin{aligned} S_{\epsilon,\delta}(X) &= \frac{1}{2}\left[e^{-\epsilon} \exp\{\delta \sinh^{-1}(X)\} - e^{\epsilon} \exp\{-\delta \sinh^{-1}(X)\}\right], \\ &= \frac{1}{2}\left[e^{-\epsilon}\{(X^2+1)^{1/2} + X\}^{\delta} - e^{\epsilon}\{(X^2+1)^{1/2} + X\}^{-\delta}\right], \qquad (3) \\ &= \frac{1}{2}\left[e^{-\epsilon}\{(X^2+1)^{1/2} + X\}^{\delta} - e^{\epsilon}\{(X^2+1)^{1/2} - X\}^{\delta}\right]. \qquad (4) \end{aligned}$$

Also, $\sinh^{-1}(Z) = \delta \sinh^{-1}(X_{\epsilon,\delta}) - \epsilon$ and $X_{\epsilon,\delta} = S_{-\epsilon/\delta,1/\delta}(Z) = \sinh[\delta^{-1}\{\sinh^{-1}(Z) + \epsilon\}]$. Random variate generation is immediate using the latter formula. The distribution function associated with density (2) is $F_{\epsilon,\delta}(x) = \Phi\{S_{\epsilon,\delta}(x)\}$, where Φ is the standard normal distribution function. The quantile function associated with (2) is $Q_{\epsilon,\delta}(u) = S_{-\epsilon/\delta,1/\delta}\{\Phi^{-1}(u)\}$, $0 < u < 1$. In particular, the median is $\sinh(\epsilon/\delta)$. Finally, density (2) is always unimodal. This is

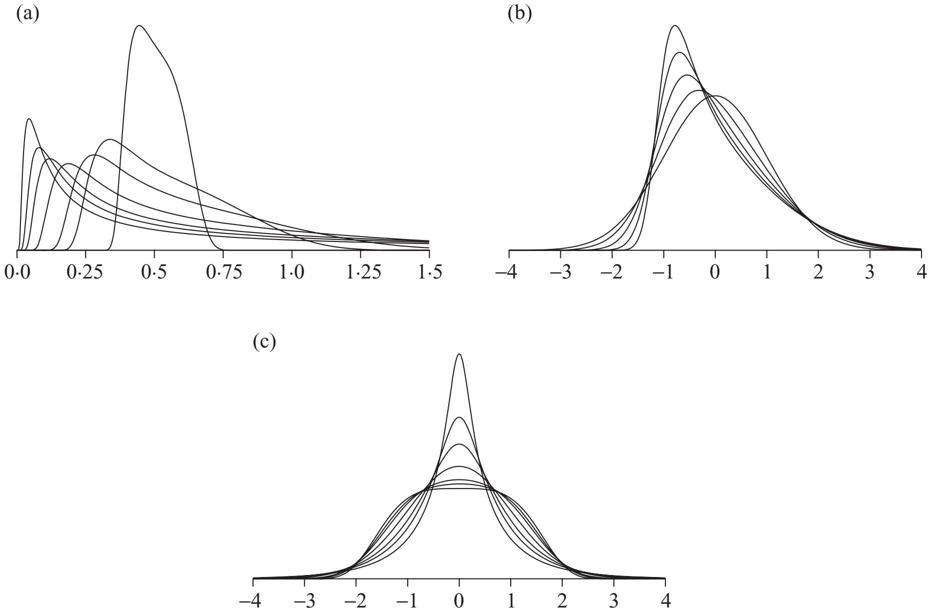

Fig. 1. (a) Densities $f_{\infty,\delta}$ for, reading from left to right, $\delta = 0{\cdot}5, 0{\cdot}625, 0{\cdot}75, 1, 1{\cdot}5, 2, 5$; (b) standardized densities $\sigma_{\epsilon,1} f_{\epsilon,1}(\sigma_{\epsilon,1}x + \mu_{\epsilon,1})$ for, in increasing degree of skewness, $\epsilon = 0, 0{\cdot}25, 0{\cdot}5, 0{\cdot}75, 1$; (c) scaled densities $\sigma_{0,\delta} f_{0,\delta}(\sigma_{0,\delta}x)$ for, in decreasing value of $\sigma_{0,\delta} f_{0,\delta}(0)$, $\delta = 0{\cdot}5, 0{\cdot}625, 0{\cdot}75, 1, 1{\cdot}5, 2, 5$.

demonstrated in our technical report. When $\epsilon = 0$, the mode $x_0 = 0$, else x_0 has the sign of ϵ and $0 < |x_0| < \sinh(|\epsilon|/\delta)$.

2·2. *Skewness, tailweight and moments*

First, $f_{-\epsilon,\delta}(x) = f_{\epsilon,\delta}(-x)$. We can also show that, for fixed δ, ϵ acts as a skewness parameter in the sense of van Zwet's (1964) skewness ordering. This defines $G_1 \leqslant_2 G_2$, that is, G_2 is more positively skew than G_1, if $G_2^{-1}(G_1)$ is convex for all x. So let $G_1 = F_{\epsilon_1,\delta}$ and $G_2 = F_{\epsilon_2,\delta}$ for $\epsilon_2 > \epsilon_1$. Then $F^{-1}_{\epsilon_2,\delta}\{F_{\epsilon_1,\delta}(x)\} = S_{c,1}(x)$, where $c = (\epsilon_1 - \epsilon_2)/\delta < 0$, and

$$\frac{d^2 F^{-1}_{\epsilon_2,\delta}\{F_{\epsilon_1,\delta}(x)\}}{d^2x} = \frac{\{1 + S^2_{c,1}(x)\}^{1/2}}{1 + x^2}\left[\frac{S_{c,1}(x)}{\{1 + S^2_{c,1}(x)\}^{1/2}} - \frac{x}{(1 + x^2)^{1/2}}\right],$$

which is positive because $S_{c,1}(x) > x$ for $c < 0$.

It is possible to identify the limiting densities $f_{\epsilon,\delta}$ as $\epsilon \to \pm\infty$. For concreteness, let us work with $\epsilon > 0$, that is, positive skewness, and call the limiting densities $f_{\infty,\delta}$. Employing suitable standardization of location and scale, we find that

$$f_{\infty,\delta}(y) = (2\pi)^{-1/2} y^{-1} \delta \cosh(\delta \log 2y) \exp\{-\sinh^2(\delta \log 2y)/2\} \quad (y > 0).$$

These are the densities of $Y = \exp\{\sinh^{-1}(Z)/\delta\}/2$, where Z is standard normal, and are plotted in Fig. 1(a) for a range of values of δ. The reader might prefer to look first at the less extreme members of family (2) shown in Fig. 1(b) and (c). The density $f_{\infty,5}$, associated with very light tails, is not very skew, but most others are. Any limitations associated with the range of available skewness seem mild.

With regard to tailweight, as $|x| \to \infty$, $S_{\epsilon,\delta}(x) \approx 2^{\delta-1}\mathrm{sgn}(x)\exp\{-\mathrm{sgn}(x)\epsilon\}|x|^{\delta}$ and $C_{\epsilon,\delta}(x) \approx 2^{\delta-1}\exp\{-\mathrm{sgn}(x)\epsilon\}|x|^{\delta}$. It follows that

$$f_{\epsilon,\delta}(|x|) \approx \exp\{-\mathrm{sgn}(x)\epsilon\}|x|^{\delta-1}\exp\{-e^{-\mathrm{sgn}(x)2\epsilon}|x|^{2\delta}\}. \tag{5}$$

These are closely related to Weibull and semi-heavy tails for small δ, being heavier than exponentially decaying tails and lighter than tails decreasing as a power of $|x|$. We see the effect of ϵ, through $e^{\pm\epsilon}$, on the relative scales of the tails of the density.

The moments, which all exist because of the tail behaviour given by (5), are available for family (2). Using the version of (3) associated with $S^{-1}_{\epsilon,\delta}$, we have

$$\begin{aligned} E(X^r_{\epsilon,\delta}) &= \frac{1}{2^r}E\left([e^{\epsilon/\delta}\{Z+(Z^2+1)^{1/2}\}^{1/\delta} - e^{-\epsilon/\delta}\{Z+(Z^2+1)^{1/2}\}^{-1/\delta}]^r\right) \\ &= \frac{1}{2^r}\sum_{i=0}^{r}\binom{r}{i}(-1)^i\exp\left\{(r-2i)\frac{\epsilon}{\delta}\right\}P_{(r-2i)/\delta} \end{aligned}$$

for integer r, where

$$\begin{aligned} P_q = E\left[\{Z+(Z^2+1)^{1/2}\}^q\right] &= \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty}\{x+(x^2+1)^{1/2}\}^q e^{-x^2/2}dx \\ &= \frac{1}{(8\pi)^{1/2}}\int_0^{\infty} w^q\left(1+\frac{1}{w^2}\right)\exp\left\{-\frac{1}{8}\left(w-\frac{1}{w}\right)^2\right\}dw \\ &= \frac{e^{1/4}8^{(q+1)/2}}{(32\pi)^{1/2}}\int_0^{\infty} z^{(q-1)/2}\left(1+\frac{1}{8z}\right)\exp\left\{-\left(z+\frac{1}{64z}\right)\right\}dz \\ &= \frac{e^{1/4}}{(8\pi)^{1/2}}\left\{K_{(q+1)/2}(1/4)+K_{(q-1)/2}(1/4)\right\}, \end{aligned}$$

using (3.471.12) of Gradshteyn & Ryzhik (1994). Here, K is the modified Bessel function of the second kind: $K_{-\nu}(z) = K_{\nu}(z)$ so that $P_{-q} = P_q$, confirming $E(X^r_{0,\delta}) = 0$ for odd r. The first four moments are therefore

$$\begin{aligned} &E(X_{\epsilon,\delta}) = \sinh(\epsilon/\delta)P_{1/\delta}, \quad E(X^2_{\epsilon,\delta}) = \tfrac{1}{2}\{\cosh(2\epsilon/\delta)P_{2/\delta}-1\}, \\ &E(X^3_{\epsilon,\delta}) = \tfrac{1}{4}\{\sinh(3\epsilon/\delta)P_{3/\delta} - 3\sinh(\epsilon/\delta)P_{1/\delta}\}, \\ &E(X^4_{\epsilon,\delta}) = \tfrac{1}{8}\{\cosh(4\epsilon/\delta)P_{4/\delta} - 4\cosh(2\epsilon/\delta)P_{2/\delta}+3\}. \end{aligned}$$

The above expressions yield formulae for the classical skewness, γ_1, and the excess kurtosis $\gamma_2 = \beta_2 - 3$. These facilitate, via a substantial computational effort, Fig. 2, which shows the attainable region of skewness and kurtosis values for the sinh-arcsinh distribution. We show the half of the plot associated with nonnegative skewness, utilize a nonstandard but helpful rescaling of the measures and include the lower bound of kurtosis values as a function of skewness for all distributions. We also plot the corresponding regions for the skew-exponential power (DiCiccio & Monti, 2004) and skew-t (Branco & Dey, 2001; Azzalini & Capitanio, 2003) distributions. The former has a rather surprising shape for light tails corresponding, it seems, to its skewness parameter not necessarily having the same sign as the skewness of the distribution nor behaving in a monotone fashion with respect to kurtosis; see also Figs. 3 and 4 of Azzalini (1986). Figure 2 shows that the sinh-arcsinh region is a superset of the skew-t region, the latter, of course, not covering negative excess kurtosis. The skew-exponential power region, on the other hand, is

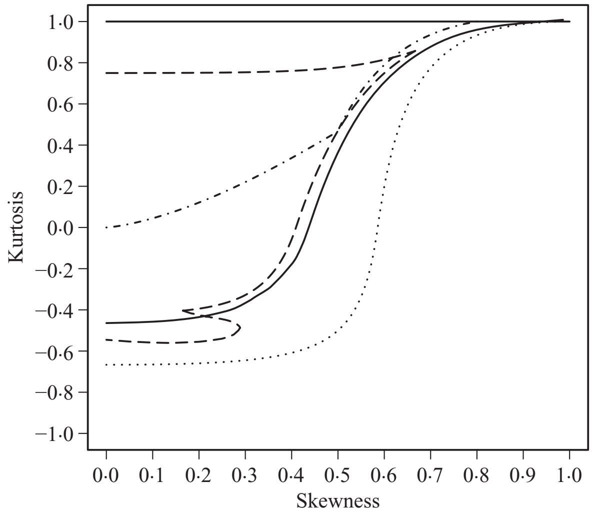

Fig. 2. Attainable regions of nonnegative skewness and excess kurtosis, represented in terms of the measures $\gamma_2/(1+|\gamma_2|) \in (-1, 1)$ versus $\gamma_1/(1+\gamma_1) \in [0, 1)$, for the sinh-arcsinh (solid), skew-exponential power (long dashed) and skew-*t* (dot-dashed) families. Also included is the limiting lower bound of kurtosis as a function of skewness (dotted).

slightly larger in the low skewness and kurtosis area, but fails to encompass the largest values of kurtosis. The sinh-arcsinh distributions allow any value of γ_1 and $\gamma_2 > -0{\cdot}863$. Fig. 2 is very supportive of a wide role for the sinh-arcsinh distribution.

2·3. *An asymmetric subfamily*

There may be some interest in the three-parameter subfamily of (2) in which $\delta = 1$. In this case transformation (1) reduces, through (4), to $S_{\epsilon,1}(X) = \cosh(\epsilon)\ X - \sinh(\epsilon)\ (1 + X^2)^{1/2}$. These densities, some of which are displayed in Fig. 1(b), are true skew-normal densities in the sense of retaining two normal-like tails; see (5). They share this property with two-piece normal distributions (Fechner, 1897; Fernández & Steel, 1998; Mudholkar & Hutson, 2000) in which two differentially scaled halves of a normal distribution are joined together. However, unlike the two-piece normal, density (2) is infinitely differentiable at all $x \in \mathbb{R}$. The current and two-piece densities differ from the skew-normal distribution with density $2\phi(x)\Phi(\lambda x)$ (Azzalini, 1985; Genton, 2004) for which a side-effect of introducing the skewness parameter λ is a change to the weight in one of the tails.

3. The symmetric subfamily

3·1. *Basic properties*

When $\epsilon = 0$ in transformation (1), density (2) is symmetric about zero. The properties discussed in § 2 translate to this special case straightforwardly. Inter alia, the mean, the median and the mode, of course, all reduce to zero. Also, computations strongly suggest that $f_{0,\delta}$ is log-concave for all $\delta \geqslant 1$.

We can show that, for $\epsilon = 0$, δ acts as a kurtosis parameter in the sense of van Zwet's (1964) ordering, which defines $G_1 \leqslant_S G_2$ for distributions G_1 and G_2 symmetric about zero if $G_2^{-1}(G_1)$ is convex for $x > 0$. In our case, let $G_1 = F_{0,\delta_1}$ and $G_2 = F_{0,\delta_2}$ for $\delta_1 > \delta_2$. Then

$F^{-1}_{0,\delta_2}\{F_{0,\delta_1}(x)\} = S_{0,\delta}(x)$, where $\delta = \delta_1/\delta_2 > 1$ and

$$\frac{d^2 F^{-1}_{0,\delta_2}\{F_{0,\delta_1}(x)\}}{d^2x} = \frac{\delta\{1+S^2_{0,\delta}(x)\}^{1/2}}{1+x^2}\left[\frac{\delta S_{0,\delta}(x)}{\{1+S^2_{0,\delta}(x)\}^{1/2}} - \frac{x}{(1+x^2)^{1/2}}\right].$$

This is positive because $\delta > 1$ and, correspondingly, $S_{0,\delta}(x) > x$ for $x > 0$.

A range of symmetric members of family (2) are plotted in Fig. 1(c). These densities have been scaled to unit variance. Unimodality, tailweight and kurtosis properties from above are well illustrated by this figure. Note how the densities vary from the heavy tailed when δ is small, through the normal when $\delta = 1$, to wide bodied, and light tailed, densities when δ is large.

3·2. *Links to Johnson S_U and sinh-normal distributions*

Consider again the transformations of a standard normal random variable Z of the form $Z = T_\delta(X)$ for some odd function T_δ generating symmetric distributions for X also on $\mathbb{R}$. The two-component parts of transformation $S_{0,\delta}(X)$, the sinh and arcsinh transformations, have each previously been employed separately in this manner. First, when $T_\delta(X) = \delta \sinh^{-1}(X)$, we have the symmetric subset of Johnson's (1949) S_U distributions, part of the famous family of transformation-based distributions that also have members on $\mathbb{R}^+$ and [0, 1]; see Johnson et al. (1994, § 12.4.3). These distributions all have tails that are heavier than those of the normal. Second, when Z is normal and $T_{\delta'}(X) = \delta' \sinh(X)$, we have Rieck & Nedelman's (1991) sinh-normal distributions. These symmetric distributions all have tails that are lighter than those of the normal. Indeed, as noted by Rieck & Nedelman using different notation, the sinh-normal distribution is log-concave for $\delta' \geqslant 1$, but there is a problem for $\delta' < 1$: the distribution is then bimodal. This is unattractive both because of the form of the bimodality that seems unlikely to be of practical interest and because we feel it better to model bi- and multi-modality through interpretable mixtures of unimodal components.

Now, when δ is small, it is immediate from (2) that

$$f_{0,\delta}(x) \approx \{2\pi(1+x^2)\}^{-1/2}\delta \exp[-\{\delta \sinh^{-1}(x)\}^2/2].$$

This is precisely the symmetric Johnson S_U density. It can also be shown that, suitably scaled, the limiting form of $f_{0,\delta}$ when $\delta \to \infty$ is

$$f_{0,\infty}(x) = (2\pi)^{-1/2}\cosh(x)\exp\{-\sinh^2(x)/2\}.$$

This is the unimodal special case of the sinh-normal distribution with $\delta' = 1$. These results are very gratifying. They show that by the use of transformation $\sinh\{\delta \sinh^{-1}(X)\}$, we have achieved a seamless family of distributions that centre on the normal distribution, behave very much like Johnson's S_U distributions for tailweights heavier than normal, and like Rieck & Nedelman's sinh-normal distributions for tailweights lighter than normal. Furthermore, recalling that the normal distribution corresponds to $\delta' = \infty$, the correspondence with the sinh-normal distribution only goes down as far as Rieck & Nedelman's $\delta' = 1$, i.e. automatically stopping just before bimodality kicks in. Similar reasoning shows why the dual transformation $T_{\delta''}(X) = \sinh^{-1}\{\delta'' \sinh(X)\}$ is not to be recommended for further investigation: it allows small $\delta' = \delta''$ cases of the sinh-normal distribution and hence bimodality.

4. On maximum likelihood estimation

4·1. *Preamble*

Let $X_1, \ldots, X_n$ denote a random sample from family (2) expanded to a four-parameter family of the form $\eta^{-1} f_{\epsilon,\delta}\{\eta^{-1}(x-\xi)\}$ and ℓ the corresponding loglikelihood. Manipulations to derive the score equations and the information matrices are standard if tedious; the observed information matrix is given in the Appendix and expectations of elements of the matrix are available through numerical integration. The usual regularity conditions for likelihood inference hold in this case. Moreover, we have not come across singular information matrices for particular parameter values, as occur for many Azzalini-type skew distributions (Azzalini & Genton, 2008), for example. The further theoretical work in §§ 4·2 and 4·3 concentrates on the symmetric subfamily. However, this work informs our fitting of the full model as described in § 4·4. One-sample considerations generalize readily to regression models in which the sinh-arcsinh distribution provides a family of response conditional distributions and location, and possibly other parameters, and is modelled as a parametric, e.g. linear, function of covariates.

4·2. *Maximum likelihood asymptotics in the symmetric case*

Consider the expected information matrix when $\epsilon = 0$, which is n times the matrix of values of $\iota_{\psi\theta} = E[-(\partial^2\ell/\partial\psi\partial\theta)\{\eta^{-1}(X_{0,\delta}-\xi)\}]$, $\psi, \theta = \{\xi, \eta, \delta\}$. The formulae to follow can be confirmed to correspond to the formulae given in the Appendix when $\epsilon = 0$ and those that are no longer relevant are ignored. We find that

$$\iota_{\xi\xi} = f_m(\delta)/\eta^2, \quad \iota_{\xi\eta} = 0, \quad \iota_{\xi\delta} = 0,$$
$$\iota_{\eta\eta} = f_s(\delta)/\eta^2, \quad \iota_{\eta\delta} = f_c(\delta)/\eta, \quad \iota_{\delta\delta} = f_d(\delta),$$

say, where the fs are all independent of ξ and η. This structure is a consequence of the symmetry of this location-scale model. In fact,

$$f_m(\delta) = E\left[\frac{\delta^2 Z^2(3+2Z^2)}{C^2_{0,1/\delta}(Z)(1+Z^2)} - \frac{\delta S_{0,1/\delta}(Z)Z^3}{C^3_{0,1/\delta}(Z)(1+Z^2)^{1/2}} + \frac{\{1-S^2_{0,1/\delta}(Z)\}}{C^4_{0,1/\delta}(Z)}\right],$$

$$f_s(\delta) = E\left(S^2_{0,1/\delta}(Z)\left[\frac{\delta^2 Z^2(3+2Z^2)}{C^2_{0,1/\delta}(Z)(1+Z^2)} - \frac{\delta S_{0,1/\delta}(Z)Z^3}{C^3_{0,1/\delta}(Z)(1+Z^2)^{1/2}} + \frac{\{1-S^2_{0,1/\delta}(Z)\}}{C^4_{0,1/\delta}(Z)}\right]\right) + 1,$$

$$f_c(\delta) = -E\left[\frac{S_{0,1/\delta}(Z)Z^2}{C_{0,1/\delta}(Z)(1+Z^2)}\left\{Z(1+Z^2)^{1/2} + (3+2Z^2)\sinh^{-1}(Z)\right\}\right],$$

and

$$f_d(\delta) = \delta^{-2}(1 + E[(1+Z^2)^{-1}Z^2(3+2Z^2)\{\sinh^{-1}(Z)\}^2]),$$

where $Z \sim N(0, 1)$.

The estimated location parameter, $\hat{\xi}$, is asymptotically independent of the estimated scale and the shape parameters, $\hat{\eta}$ and $\hat{\delta}$, respectively. However, $\hat{\eta}$ and $\hat{\delta}$ are not asymptotically independent: as $n \to \infty$,

$$\text{corr}(\hat{\eta}, \hat{\delta}) \approx -(\iota_{\eta\eta}\iota_{\delta\delta})^{-1/2}\iota_{\eta\delta} = -\{f_s(\delta)f_d(\delta)\}^{-1/2}f_c(\delta),$$

a function of δ only. Clearly, $\iota_{\eta\delta} < 0$ and hence $\text{corr}(\hat{\eta}, \hat{\delta})$ is asymptotically positive. This correlation is high for almost all δ, rising from around 0·6 when $\delta \to 0$, through 0·975 when $\delta = 1$ to 1 as $\delta \to \infty$. This seems disappointing, but proves to be a standard property of scale/tailweight families of symmetric distributions, reflecting the fact that one cannot easily tell the difference between changing scale and changing tailweight in practice.

Also, the asymptotic variance of $\hat{\delta}$ depends only on δ, while the asymptotic variances of $\hat{\xi}$ and $\hat{\eta}$ are each of the form $n^{-1}\eta^2 h(\delta)$. Formulae for, and a pictorial representation of, these variances are given in our technical report.

4·3. *Reparameterization*

In principle, it is possible to provide an orthogonal parameterization of the form $\{\xi, \eta F(\delta), \delta\}$. Since, asymptotically, $\text{corr}\{\hat{\eta}F(\hat{\delta}), \hat{\delta}\} \propto (\log F)'(\delta) f_s(\delta) - f_c(\delta)$, this would be achieved by setting $(\log F)'(\delta) = f_c(\delta)/f_s(\delta)$, but this specification is insufficiently tractable to provide a workable formula. However, the asymptotic correlation between $\hat{\eta}$ and $\hat{\delta}$ is highest for large δ. This suggests seeking a large δ approximation to the above. To this end, we find that, for large δ, $f_c(\delta) \approx -(C + S)/\delta$ and $f_s(\delta) \approx 1 + S$, where

$$\begin{aligned} C = E\{(1 + Z^2)^{-1/2} Z^3 \sinh^{-1}(Z)\} &= \int_{-\infty}^{\infty} x \sinh^3 x\, \phi(\sinh x) dx \\ &= [\Phi(\sinh x) - \phi(\sinh x) x \cosh x]_{-\infty}^{\infty} = 1 \end{aligned}$$

and

$$S = E[(1 + Z^2)^{-1} Z^2 \{\sinh^{-1}(Z)\}^2 (3 + 2Z^2)] \simeq 3{\cdot}498 > 0.$$

Thus, $F(\delta) \approx \delta^{-1}$, suggesting a simple reparameterization in which η is replaced by $\eta_\delta \equiv \eta/\delta$. The asymptotic correlation between $\hat{\eta}_\delta$ and $\hat{\delta}$ is

$$-[f_s(\delta)\{\delta^2 f_d(\delta) + 2\delta f_c(\delta) + f_s(\delta)\}]^{-1/2}\{\delta f_c(\delta) + f_s(\delta)\}.$$

A general lowering of the asymptotic correlation is achieved; it increases monotonically from around 0·5 as $\delta \to 0$ to a little over 0·85 as $\delta \to \infty$. We have not achieved the very small correlation for large δ that might have been expected because the variance of $\hat{\eta}_\delta$ tends to zero as well as the covariance for large δ. However, the reduction that we have achieved proves to make a considerable difference in practice.

4·4. *Practical implementation in the general case*

We employ and recommend the reparameterization just derived in fitting the four-parameter sinh-arcsinh distribution to data, as well as its symmetric subfamily, i.e. utilizing $\{\xi, \eta_\delta, \epsilon, \delta\}$ and then setting $\hat{\eta} = \hat{\eta}_\delta \hat{\delta}$. This solved numerical problems encountered in the original parameterization when $\delta > 1$. We used the Nelder & Mead (1965) simplex algorithm to perform the maximization of the loglikelihood. Using this direct search approach, it proves helpful to optimize over $\xi/(1 + \xi^2)^{1/2} \in (-1, 1)$, $\eta_\delta/(1 + \eta_\delta) \in (0, 1)$, $\epsilon/(1 + \epsilon^2)^{1/2} \in (-1, 1)$ and $\delta/(1 + \delta) \in (0, 1)$ and then back-transform. In practice, we have not come across examples of multiple maxima occurring on the loglikelihood surface. However, as is generally the case when using numerical optimization, it is advisable to try a range of starting values to try to ensure that the global maximum is identified. We find that each of ξ, η_δ and ϵ is estimated well but large δ-values are not estimated so precisely. It is clear from Fig. 1(c) that the shape of density (2) does not vary much with large δ; in consequence and inconsequentially the loglikelihood surface is quite flat when δ is large.

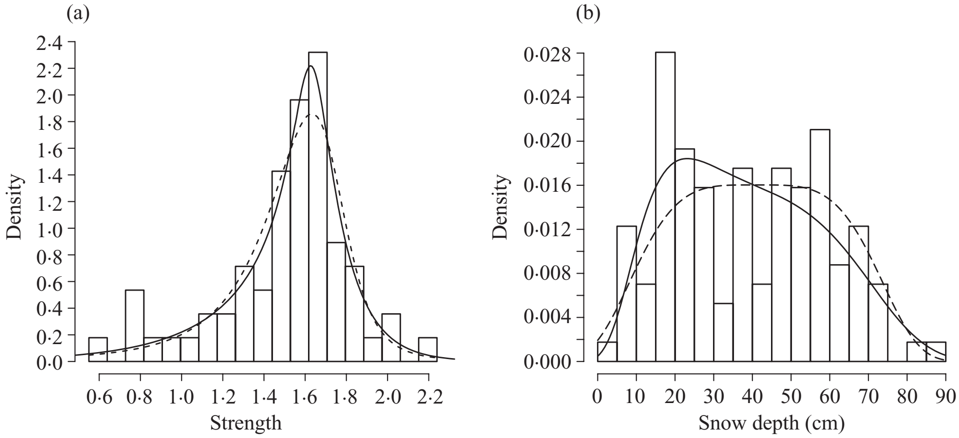

Fig. 3. Histogram and fitted densities of: (a) the strengths of glass fibres with densities of family (2) (solid) and the skew-t model of Azzalini & Capitanio (2003) (dashed); (b) the ice floe snow depth data together with the fitted densities for family (2) (solid) and its symmetric subfamily (dashed).

5. Examples

5·1. *Strengths of glass fibres*

As the first illustrative example, we present an analysis of the breaking strengths of $n = 63$ glass fibres of length 1·5 cm. These data were originally obtained by workers at the U.K. National Physical Laboratory and first published by Smith and Naylor (1987). They have previously been modelled using three different four-parameter skew-t distributions by Jones & Faddy (2003), in the University of Padua technical report version of Azzalini & Capitanio (2003) and by Ma & Genton (2004). The fits for the skew-t models proposed in these three papers are, like the data, negatively skewed and heavy-tailed; they have maximized values of their loglikelihoods of $-11{\cdot}93$, $-11{\cdot}70$ and $-11{\cdot}93$, respectively. Figure 3 of Azzalini & Capitanio's technical report shows that the fitted densities for the first two models are very similar. Figure 3(a) presents a histogram of the data with the same class intervals as those used to produce the histogram in Fig. 3 of Ma & Genton (2004). Superimposed on the histogram is the density of the best fitting skew-t model and that of the full family considered here, which corresponds to the higher maximized loglikelihood value of $-10{\cdot}00$. The sinh-arcsinh model does a much better job of capturing the shape around the peak of the density and has correspondingly slimmer flanks, yet its tails are very similar to those of all three fitted skew-t densities. These last have tails decreasing like those of a Student's t distribution with 2 or 3 degrees of freedom. This example therefore illustrates the ability of the sinh-arcsinh distribution to mimic the tails of distributions with what would usually be regarded as heavy power tails.

Detailed results for the maximum likelihood fits of family (2) and its normal, $\delta = 1$, $\epsilon = 0$, normal-tailed, $\delta = 1$, and symmetric, $\epsilon = 0$ submodels are given in Table 1. The likelihood-based criteria identify the full sinh-arcsinh family as providing a superior fit to any of the three submodels considered in Table 1. These findings are supported by p-values of 0·009, 0·007 and 0·000, respectively, for likelihood ratio tests of symmetry, normal tails and normality. The chi-squared goodness-of-fit test is less able to discriminate between the quality of the fits of the various models, the full model and its symmetric submodel being judged to provide adequate fits to the data, the normal distribution not doing so and the fit of the normal-tailed submodel being borderline. Nominally 95% confidence intervals for ξ, η_δ, δ and ϵ, obtained using the observed information matrix corresponding to this parameterization, which in turn is readily obtained from

Table 1. *Parameter estimates for the fits to the glass fibre strengths of, reading from right to left, family* (2) *and its symmetric* ($\epsilon = 0$), *normal-tailed* ($\delta = 1$) *and normal* ($\delta = 1, \epsilon = 0$) *submodels. The maximized loglikelihood,* $l_{\max}$*; Akaike information criterion,* AIC*; Bayesian information criterion,* BIC*; and p-value for the chi-squared goodness-of-fit test are included as fit diagnostics.*

	Model			
Parameter	Normal	Normal tails	Symmetric	Family (2)
ξ	1·51	1·66	1·59	1·63
η_δ	0·32	0·29	0·18	0·18
δ	1	1	0·52	0·57
ϵ	0	−0·42	0	−0·27
Diagnostic				
$l_{\max}$	−17·92	−13·63	−13·46	−10·00
AIC	39·84	33·26	32·92	28·00
BIC	44·13	39·69	39·35	36·57
p-value	0·01	0·07	0·23	0·57

that in the Appendix, are (1·57, 1·69), (0·12, 0·24), (0·44, 0·70) and (−0·47, −0·07), respectively. Confidence intervals obtained using the profile likelihood are very similar to these. The latter two intervals confirm the heavier than normal tails and negative skewness apparent in the histogram.

5·2. *Ice floe snow depths*

In this subsection, we show how the sinh-arcsinh distribution can also model light-tailed datasets. To this end, we present an analysis of $n = 114$ measurements of the depth of snow, in centimetres, taken on an ice floe in the eastern Amundsen Sea, Antarctica, in March 2003; see Chapter 6 of C. J. Banks's 2006 Open University Department of Earth Sciences Ph.D. thesis for details, noting that these data pertain to Floe 2 and Banks's analysis included preliminary use of a symmetric sinh-arcsinh distribution. A histogram of the data appears in Fig. 3(b). The results for the maximum likelihood fits of family (2) and its normal, normal-tailed and symmetric submodels are given in Table 2. All three likelihood-based diagnostics in Table 2 indicate that the fit for the full family is best, followed by that for its symmetric subfamily; the chi-squared goodness-of-fit test suggests that the two best-fitting models provide adequate fits to the data. The fit for the full family has lighter than normal tails, $\hat{\delta} = 4{\cdot}24 > 1$, and positive asymmetry, $\hat{\epsilon} = 17{\cdot}09 > 0$. The densities for the two best fits are superimposed upon the histogram in Fig. 3(b). Comparing them with the histogram, there is perhaps some indication of multimodality in the data. However, this could be an artefact of the binning used and the rounding of the data to the nearest whole centimetre during measurement. It would be difficult to conceive of a better unimodal fit to the data.

This data-fitting exercise demonstrates some difficulties associated with the similarity of distributions with a wide range of values of $\delta > 1$, mentioned at the end of § 4·4. First, we did many runs of our optimization programmes from a wide range of starting values. Many rather different parameter combinations resulted in loglikelihood values very close to the quoted −494·93, but their densities are effectively indistinguishable from the one portrayed in Fig. 3(b). Second, this resulted, when using the original parameterization, in ill-conditioning of the observed information matrix; this effect is mitigated somewhat by our alternative parameterization. Third, the asymptotic variability measures associated with the observed information matrix remain, in this case, mostly very large and, we think, unreliable. We have more faith in confidence intervals obtained

Table 2. *Parameter estimates for the fits to the ice floe snow depth data of, reading from right to left, family* (2) *and its symmetric* ($\epsilon = 0$)*, normal-tailed* ($\delta = 1$) *and normal* ($\delta = 1, \epsilon = 0$) *submodels. The maximized loglikelihood,* $l_{\max}$*; Akaike information criterion,* AIC*; Bayesian information criterion,* BIC*; and p-value for the chi-squared goodness-of-fit test are included as fit diagnostics.*

	Model			
Parameter	Normal	Normal tails	Symmetric	Family (2)
ξ	39·24	24·66	40·94	−65·26
η_δ	20·20	17·88	24·89	0·85
δ	1	1	19263·4	4·24
ϵ	0	0·52	0	17·09
Diagnostics				
$l_{\max}$	−504·39	−502·50	−497·72	−494·93
AIC	1012·78	1011·00	1001·44	997·86
BIC	1018·25	1019·21	1009·65	1008·80
p-value	0·031	0·006	0·311	0·253

from profile likelihoods. Nominally 95% confidence intervals for the shape parameters turn out to be (1·72, 25·50) for δ and (0·54, ∞) for ϵ. Of course, these are very wide, but they reflect well the facts that the tails are certainly light and the asymmetry is certainly positive.

6. Testing normality and symmetry

6·1. *Testing normality: the competition*

The central position of the normal distribution allows testing of normality within family (2) via standard likelihood ratio tests. However, we propose that sinh-arcsinh-based likelihood ratio tests of normality also be used as general purpose tests of normality. There exists a well-established literature addressing this problem; renewed recent interest can be found in Zhang & Wu (2005) and Thadewald & Büning (2007), among others. We conducted a simulation study to compare the performance of the likelihood ratio tests with those of the following seven competing tests for a nominal significance level of 5%:

(i) the test of Bowman & Shenton (1975) based on skewness and kurtosis. This test was rediscovered by Jarque & Bera (1980). We used corrected critical values given in Table 2 of Thadewald & Büning (2007);
(ii) the test of D'Agostino (1971, 1972) comparing Downton's (1966) linear estimator of the standard deviation with the sample standard deviation;
(iii) the empirical distribution function-based test of Anderson & Darling (1952). We used corrected critical values from Table 2 of Thadewald & Büning (2007), where the test is renamed as the weighted Cramér–von Mises test;
(iv) the Cramér–von Mises empirical distribution function-based test with corrected critical values from Table 2 of Thadewald & Büning (2007);
(v) the test of Shapiro & Wilk (1965) as implemented by Royston (1995);
(vi/vii) the nonparametric likelihood-ratio-based tests Z_A and Z_C of Zhang & Wu (2005).

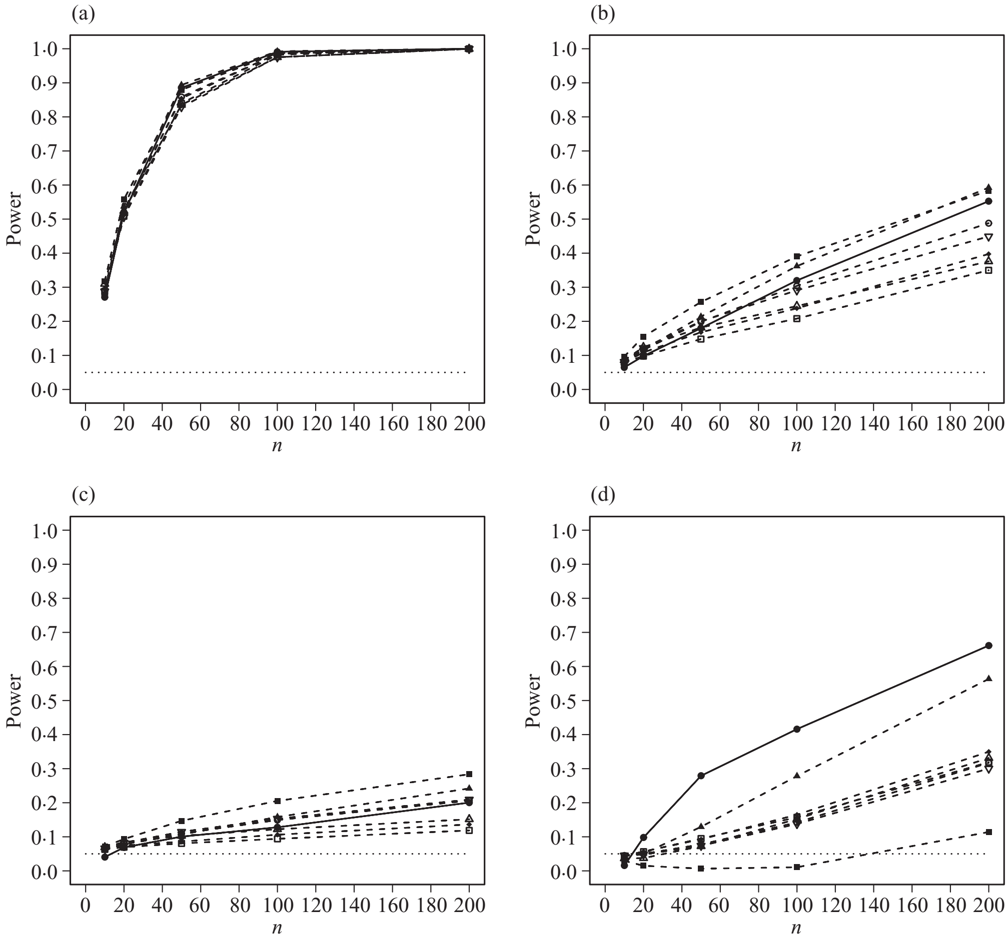

Fig. 4. The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test plotted against n. The proportions were calculated using 10 000 random samples of size $n = 10, 20, 50, 100, 200$ from the (a) t_2; (b) logistic; (c) t_{15}; and (d) Tiku short-tailed distributions. The solid lines connect the results of the likelihood ratio test, and the dashed lines those for the other seven tests: Bowman–Shenton (solid square); D'Agostino (solid triangle); Anderson–Darling (solid diamond); Cramér–von Mises (open square); Shapiro–Wilk (open circle); Z_A (open triangle); and Z_C (open inverted triangle). The dotted line is at the nominal level of 0·05.

6·2. *Testing normality against symmetric alternatives*

One scenario in which one might wish to test for normality is that in which the distribution of interest is assumed to be symmetric. The appropriate likelihood ratio test statistic is $L = -2(\ell_0 - \ell_1)$, where ℓ_0 is the maximum of the loglikelihood function for an assumed normal distribution and ℓ_1 is the maximized loglikelihood assuming that the sample was drawn from a symmetric sinh-arcsinh distribution, ℓ_1 being calculated numerically. The asymptotic distribution of L is χ^2_1. We investigated the sampling distribution of L under normality via simulation. The χ^2 approximation, initially poor, rapidly improves with increasing n such that for $n \geqslant 50$ it gives a very good approximation to the true sampling distribution. Moreover, the tails and critical values of the χ^2 approximation are good even for small n. Our results vindicate using the critical values of the asymptotic χ^2_1 distribution for any reasonable n. We also checked that all seven rival tests maintain the nominal significance level of 5% very closely.

Here are the results of some power simulations against four alternative symmetric distributions, namely: (a) the very heavy-tailed t distribution on two degrees of freedom, t_2; (b) the fairly heavy-tailed logistic distribution; (c) the rather normal-like t distribution on 15 degrees of freedom; and (d) a light-tailed distribution according to M. L. Tiku with density $16(1 + x^2/4)^2\phi(x)/27$ (e.g. Tiku et al., 2001). In Fig. 4, the proportion of 10 000 simulated samples for which the null hypothesis of normality was rejected in a nominally 5% test is plotted against n. All eight tests are relatively powerful against the t_2 alternative; see Fig. 4(a). For larger values of n, there is little difference in their powers; for $n = 20, 50$, the Bowman–Shenton and D'Agostino tests tend to dominate. None of the tests is very powerful against the logistic alternative; see Fig. 4(b). The Bowman–Shenton test has the best overall performance. The D'Agostino test also performs relatively well, particularly for larger n. The likelihood ratio test performs relatively poorly for $n \leqslant 50$ but its performance improves with increasing n. The performance of the Cramér–von Mises test is worst. A similar pattern, but at lower power levels, is observed in Fig. 4(c) for the t_{15} alternative. Finally, none of the tests is particularly powerful against Tiku's short-tailed distribution; see Fig. 4(d). The likelihood ratio test is clearly the most powerful, followed by the D'Agostino test. The powers of five of the other six tests are very similar, with the Bowman–Shenton test being very poor. Overall, the sinh-arcsinh likelihood ratio test seems very competitive in most symmetric situations with the best of existing tests which appears to be D'Agostino's test. Parallel results using alternatives from within the sinh-arcsinh family of distributions can be found in our technical report.

6·3. *Testing normality against asymmetric alternatives*

If one is not willing to assume symmetry, testing for normality can be accomplished from within the full sinh-arcsinh family in the obvious way. The asymptotic sampling distribution of the likelihood ratio test statistic is now χ^2_2; simulations confirmed the test's ability to maintain its nominal significance level using this distribution. The power of the asymmetric likelihood ratio test is necessarily a little lower than that of the previous symmetric likelihood ratio test when normality is tested within an assumed symmetric situation. The effect is quite small and the overall performance of the asymmetric test remains excellent; indeed, it proves to be the best. Indeed, the ordering of the power performances of the tests is by and large the same as in Fig. 4 with two notable exceptions: the D'Agostino test, which was previously competitive with the likelihood ratio test, is very badly affected by the presence of asymmetry; and the likelihood ratio test maintains its first place even for alternatives with slightly heavier tails than those of the normal. Details are given in our technical report. In further unreported simulations, we have also shown that the sinh-arcsinh-based likelihood ratio test performs essentially as well as the skew-exponential power-based likelihood ratio test for the range of skew-exponential power distributions considered in Table 7 of DiCiccio & Monti (2004). Taking both symmetric and asymmetric alternatives into account, the likelihood ratio test seems to be the best of the options considered here. Note that its competitors were chosen because of claims of leading performance elsewhere.

6·4. *Testing symmetry*

We can also test for symmetry about an unknown centre by employing a likelihood ratio test within the full sinh-arcsinh family of the null hypothesis that $\epsilon = 0$. The asymptotic null distribution of the test is, again, χ^2_1. We show in our technical report that the sinh-arcsinh-based likelihood ratio test clearly outperforms two omnibus tests for symmetry chosen as being state of the art, namely those of Boos (1982) and Cabilio & Masaro (1996).

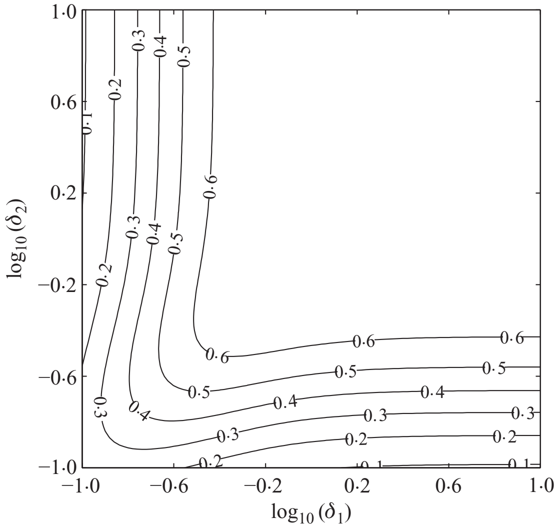

Fig. 5. The correlation between X and Y in the symmetric marginals case, plotted as a function of $\log_{10}\delta_1$ and $\log_{10}\delta_2$: here, $\rho = 0{\cdot}7$.

7. THE MULTIVARIATE CASE

Multivariate extensions of the univariate distributions arise naturally by transforming the univariate marginals of a standardized, but correlated, multivariate normal distribution. We thereby model skewness and/or tailweight variations on the original scales of the variables. In d dimensions, let R be a correlation matrix and define the vector X by $Z_i = S_{\epsilon_i,\delta_i}(X_i)$, $i = 1, \ldots, d$, where $Z \sim N_d(0, R)$, so that

$$f_{\epsilon,\delta}(x) = \{(2\pi)^d |R|\}^{-1/2} \prod_{i=1}^{d} \left\{ \left(1 + x_i^2\right)^{-1/2} \delta_i C_{\epsilon_i,\delta_i}(x_i) \right\} \exp\{-S_{\epsilon,\delta}(x)' R^{-1} S_{\epsilon,\delta}(x)/2\}.$$

In an abuse of notation, the vector z has been written $S_{\epsilon,\delta}(x)$.

The univariate marginals are sinh-arcsinh distributions by construction. If z is partitioned into (z_1, z_2) and x, X and R are partitioned conformably, $X_1 \mid x_2$ is the distribution of $S^{-1}_{\epsilon_1,\delta_1}(Z_1) \mid z_2 = S_{\epsilon_2,\delta_2}(x_2)$, where $Z_1 \mid z_2 \sim N\{R_{12}(R_{22})^{-1} z_2, R_{11} - R_{12}^T (R_{22})^{-1} R_{12}\}$. Note that now the transformation is applied to an unstandardized normal distribution, which means that conditional distributions are members of a wider, and not very tractable, family of distributions that will not be pursued. The unimodality of univariate distributions augurs well for the unimodality of the multivariate case, and we have no counterexamples from our limited experience with these distributions. All moments of the distribution, of course, exist.

The covariance between any two elements of X is not generally tractable. It is, however, plotted in the symmetric marginals, $\epsilon_1 = \epsilon_2 = 0$, case in Fig. 5 as a function of δ_1, the parameter in the x-direction, and δ_2, the parameter in the y-direction, for $\rho = 0{\cdot}7$. A number of properties of the multivariate distribution are illustrated by this plot. First, the sign of $\rho_{12} = \mathrm{corr}\{S^{-1}_{\epsilon_1,\delta_1}(Z_1), S^{-1}_{\epsilon_2,\delta_2}(Z_2)\}$ is the same as the sign of ρ for all $\epsilon_1, \delta_1, \epsilon_2, \delta_2$. This follows from the positive, respectively negative, quadrant dependence of the bivariate normal distribution with $\rho > 0$, respectively < 0, and the strictly increasing nature of the marginal transformations; see, for example, Joe (1997). Second, $|\rho_{12}| \leqslant |\rho|$. This inequality can be found in literature stemming from Gebelein (1941); see, for example, Koyak (1987) and references therein. Concentrating on the symmetric marginals

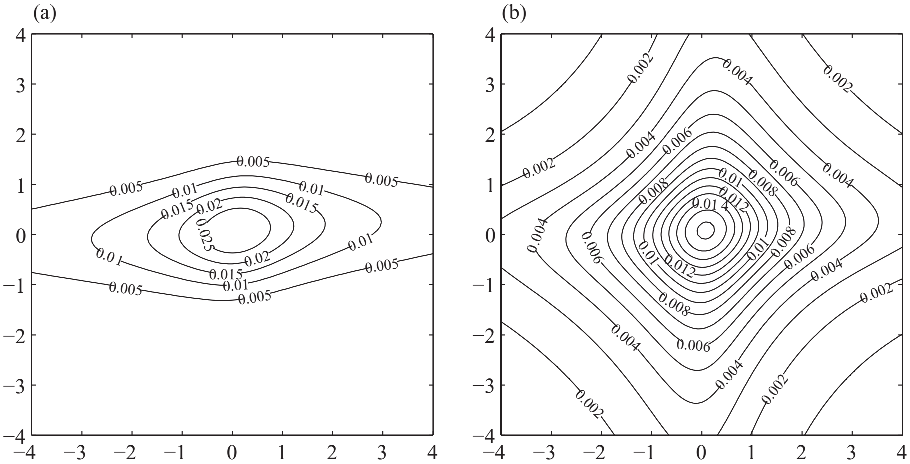

Fig. 6. The bivariate sinh-arcsinh density with $\rho = 0{\cdot}7$ and (a) $\delta_1 = 0{\cdot}135$, $\delta_2 = 1$; (b) $\delta_1 = \delta_2 = 0{\cdot}27$.

case in Fig. 5, we note that: the correlation is $\rho = 0{\cdot}7$ only at the point $\delta_1 = \delta_2 = 1$ and is lower elsewhere; the correlation remains close to $\rho = 0{\cdot}7$ for all $\delta_1, \delta_2 \geqslant 1$, i.e. lighter tails; and the absolute value of the correlation decreases as one or both tails get heavier. In particular, this makes sense in the case $\delta_1 < 1$, $\delta_2 = 1$, where the density is spread much more in the x-direction than in the y-direction. For an illustration of this, see the density plotted in Fig. 6(a); this effect is reduced somewhat if both tails get heavier. The density for $\delta_1 = \delta_2 = 0{\cdot}27$ is plotted in Fig. 6(b).

It is also of interest to consider the local dependence function defined as $\gamma(x, y) = \partial^2 \log f_{\epsilon,\delta}(x, y)/\partial x \partial y$. This was introduced as a continuous analogue of the local log odds ratio by Holland and Wang (1987) and alternatively justified as a localized correlation coefficient by Jones (1996). In our case, we have

$$\gamma(x, y) = \frac{\rho}{1-\rho^2} \frac{\delta_1 C_{\epsilon_1,\delta_1}(x)}{(1+x^2)^{1/2}} \frac{\delta_2 C_{\epsilon_2,\delta_2}(y)}{(1+y^2)^{1/2}}.$$

Note that $\gamma(x, y)$ has the same sign as ρ for all x, y. The way that ρ affects only the overall size of local dependence is a nice feature of this transformation approach. Also, x- and y-dependences are separated out, so we consider, say, $L_{\epsilon,\delta}(x) \equiv \delta C_{\epsilon,\delta}(x)/(1+x^2)^{1/2}$ only. In the symmetric case, $L_{0,\delta}(0) = \delta$ and it can readily be shown that $L_{0,\delta}$ symmetrically decreases, respectively increases, towards zero, respectively infinity, if $\delta < 1$, respectively > 1. In the general case, $L_{\epsilon,\delta}(0) = \delta \cosh \epsilon$, while both tails of $L_{\epsilon,\delta}$ still go to zero, respectively infinity, if $\delta < 1$, respectively > 1.

8. Options and extensions

8·1. *Questions*

Readers may be discomfited by some of the choices made in this paper. First, a question that has been asked is: why is the sinh function at the heart of this methodology rather than some other monotone function? Second, the normal distribution is but one of a number of possibilities for the central distribution in this approach. And there is a third, less obvious, question concerning the way in which skewness is introduced into our model. Here, we address each of these issues.

8·2. *Which transformation function?*

Introduce a one-to-one onto function $H : \mathbb{R} \to \mathbb{R}$ with $H(0) = 0$ and write $h(x) = H'(x) > 0$ for all x. Consider transformations of the form

$$Z = T_{\epsilon,\delta}(X_{\epsilon,\delta}) \equiv H\{\delta H^{-1}(X_{\epsilon,\delta}) - \epsilon\}.$$

This formulation, involving both H and H^{-1}, is key to allowing both heavier and lighter tails. This is easily seen when $\epsilon = 0$: for small δ, $T(X) \approx \delta h(0) H^{-1}(x)$, and for large δ, $T(X/\delta) \approx H\{x/h'(0)\}$, division of X by δ being the suitable scaling employed in § 3·4. Anticipating the main consideration of § 8·3, replace the normal density by a generic simple symmetric distribution with distribution function G. Apply the transformation to obtain the family of distribution functions

$$\mathcal{G}(x) \equiv G[H\{\delta H^{-1}(x) - \epsilon\}]. \tag{6}$$

THEOREM 1. *The parameters ϵ, for fixed δ, and δ, for $\epsilon = 0$, in* (6) *act as a pair of skewness and kurtosis parameters in the sense of van Zwet* (1964) *if and only if* $\log h$ *is either a convex or a concave function of x.*

Proof. Let $\mathcal{G}_i$ denote $\mathcal{G}$ when the parameters are ϵ_i, δ_i $(i = 1, 2)$. Then $\mathcal{G}_2^{-1}\{\mathcal{G}_1(x)\} = H\{dH^{-1}(x) - c\}$, independently of G, where $c = (\epsilon_1 - \epsilon_2)/\delta_2$ and $d = \delta_1/\delta_2$. Then,

$$t_{c,d}(x) \equiv \frac{d^2\mathcal{G}_2^{-1}\{\mathcal{G}_1(x)\}}{dx^2} = p(x)[d(\log h)'\{dH^{-1}(x) - c\} - (\log h)'\{H^{-1}(x)\}],$$

where $p(x) = dh\{dH^{-1}(x) - c\}/h^2\{H^{-1}(x)\} > 0$ for all x. For fixed δ, i.e. $d = 1$, consider the case $c < 0$, i.e. $\epsilon_2 > \epsilon_1$; then $t_{c,1}(x) > 0$, the requirement for ϵ to act as a skewness parameter, corresponds to $(\log h)''(x) > 0$ for all x. Likewise, $c > 0$ requires $(\log h)''(x) < 0$ for all x. Now fix $c = 0$ for the symmetric case $\epsilon_1 = \epsilon_2 = 0$. For δ to be a kurtosis parameter, we need $t_{0,d}(x) > 0$ for $x > 0$ and for this it is also sufficient that $\log h$ is increasing if $d > 1$ or that $\log h$ is decreasing if $d < 1$. □

Another requirement that narrows the field of Hs is unimodality of all members of the resulting family of distributions. Unfortunately, this seems to require verification on a case-by-case basis, though it was used to disqualify $H(x) = \sinh^{-1}(x)$ for normal G in § 3·4. That said, it reinforces the requirement that $h(x) > 0$ for all x: otherwise, if $h(x_0) = 0$ for some x_0, the density associated with distribution (6) will be zero at $x = H\{(x_0 + \epsilon)/\delta\}$ and nonzero to either side; this removes candidates of the form $H(x) = |x|^\gamma$, $\gamma > 0$. Other considerations include explicit invertibility of H, differentiability perhaps, and the type and breadth of effect on tails. We have not been able to come up with a viable alternative to $H(x) = \sinh(x)$.

8·3. *Which central density?*

Let the central simple symmetric distribution mentioned in § 8·2 have density g. Then the transformed family of distributions has densities of the form

$$g_{\epsilon,\delta}(x) = (1 + x^2)^{-1/2} \delta C_{\epsilon,\delta}(x) g\{S_{\epsilon,\delta}(x)\}.$$

Several properties developed for normal g hold immediately for other g too: its distribution and quantile function in terms of G and G^{-1}, skewness and kurtosis ordering properties, etc; some properties need investigation on a case-by-case basis. A sufficient condition for unimodality, usually satisfied, is that $1 + x(1 + x^2)(\log g)'(x) + (1 + x^2)^2(\log g)''(x) < 0$ for all x.

A major reason for choosing a different g would be if testing for some other simple symmetric distribution, such as the logistic, were of interest. We would expect likelihood ratio testing within a g-based family to perform as well as it does for the normality case in § 6. A further consideration might be the tailweight properties of g-based families. For small δ, and ignoring all constants,

$$g_{\epsilon,\delta}(|x|) \approx |x|^{\delta-1} g(|x|^{\delta}) \text{ as } |x| \to \infty;$$

for example, simple exponential tails like those of the logistic lead to Weibull-type tails, $|x|^{\delta-1}\exp(-|x|^{\delta})$, while t-type power tails, of the form $g(|x|) \approx |x|^{-(\alpha+1)}$, $\alpha > 0$, lead to power tails for $g_{\epsilon,\delta}$ of the form $|x|^{-(\nu+1)}$, where $\nu = \alpha\delta$. The Cauchy distribution as g leads to the particularly simple expression

$$g_{\epsilon,\delta}(x) = \frac{\delta}{\pi} \frac{1}{(1+x^2)^{1/2}\, C_{\epsilon,\delta}(x)}.$$

But centring the family of distributions at such a heavy-tailed case has consequences for the lightness of tails as $\delta \to \infty$. In the symmetric case, the Cauchy-based family tends to the hyperbolic secant density $\{\pi \cosh(x)\}^{-1}$, which is intriguing and indicative of relatively heavy light tails; they are of simple exponential form. Again, aside from distributional testing requirements, it is difficult to see beyond the normal-based family as the most useful general tool.

8·4. *Which method of introducing skewness?*

Formulae (3) and (4) suggest an alternative method of introducing skewness into the symmetric sinh-arcsinh transformation. Consider

$$S_{\delta,\gamma}(X) \equiv \tfrac{1}{2}[\exp\{\delta \sinh^{-1}(X)\} - \exp\{-\gamma \sinh^{-1}(X)\}],$$

where $\delta, \gamma > 0$. Then define $X_{\delta,\gamma}$ by $Z = S_{\delta,\gamma}(X_{\delta,\gamma})$, $X_{\delta,\gamma}$ having density $f_{\delta,\gamma}$, not shown to save space. The same symmetric cases now arise from setting $\gamma = \delta$. In fact, δ controls the weight of the right-hand tail of the distribution, while γ controls the left-hand tail in the same way. Skewness arises implicitly from the imbalance between the tails when $\delta \neq \gamma$: if $\delta < \gamma$, the left-hand tail is lighter than the right and the resulting skewness is positive, if $\delta > \gamma$, negative skewness ensues. It is clear that $f_{\gamma,\delta}(x) = f_{\delta,\gamma}(-x)$.

Many properties of these skew sinh-arcsinh distributions can be determined although the family is a little less tractable than the one based on $S_{\epsilon,\delta}$; see our technical report for details. Tests of normality and symmetry can, of course, be based on fitting $f_{\gamma,\delta}$ in the same way as they were in §§ 5 and 6 for $f_{\epsilon,\delta}$. We repeated all the simulations reported there for the alternative skewness family too. The most striking feature of the results is their extreme similarity; again, the details are in our technical report. All told, there is relatively little to choose between $f_{\epsilon,\delta}$ and $f_{\gamma,\delta}$ in many respects. We have focussed on the former in this paper primarily because of its greater tractability and secondarily because of minor practical advantages.

9. Discussion

We argue that, far from being just another four-parameter family of distributions on the real line with rather similar properties, the distributions of this paper fill a niche that is sparsely populated. Most families of distributions on $\mathbb{R}$ provide only tailweights heavier than those of the normal, often with the normal distribution as their lightest tailed limit. Examples include stable laws and various skew-t distributions which include Student's t distributions as their symmetric special cases, for example, Fernández & Steel (1998), Branco & Dey (2001), Azzalini & Capitanio (2003), Jones & Faddy (2003) and Azzalini & Genton (2008). Few families of distributions on $\mathbb{R}$ have much in the way of light-tailed membership. An exception is the exponential power distribution (Box & Tiao, 1973; Tadikamalla, 1980) and its skew counterparts, of which a popular version is studied by Azzalini (1986) and DiCiccio & Monti (2004). The new distributions fill a gap between these two sorts of distributions. Like skew-t distributions, they allow tails considerably heavier than the normal, although not quite as heavy as the t's power tails can be, but unlike skew-t distributions they also allow lighter than normal tails. Like exponential power distributions, the new distributions allow lighter tails than normal, though not as light as the uniform limit of the exponential power, and heavier tails than the normal, but escaping the purely exponential nature of the exponential power's tails, indeed to what in practice is a quite high degree of tailweight. We reiterate that the sinh-arcsinh distributions achieve all this like an amalgamation of Johnson S_U and sinh-normal distributions. In fact, the sinh-arcsinh distribution can be seen as a generalized Johnson distribution where the sinh transformation, as in Johnson (1949), is applied not to the normal distribution but to the sinh-normal distribution.

It is especially appealing, in our view, to have such a family of distributions centred on the normal distribution in order, as exemplified in § 5, to allow standard likelihood ratio testing for normality against skew and light- and heavy-tailed distributions. This contrasts with families in which the normal distribution is a limiting case. Moreover, the resulting tests are widely applicable: they turn out to compete with, and essentially outperform, existing omnibus tests of normality against alternatives not in the sinh-arcsinh family. Essentially, of course, the tests work by approximating the distribution of the data by a member of the sinh-arcsinh family, which proves to be an adequate approximation at least for most unimodal densities. Similar remarks apply to testing for symmetry via likelihood ratio tests within this class. Of course, the general approach to testing normality and symmetry through wide parametric families is not new; see e.g. Kappenman (1988) and Cassart et al. (2008). It is, however, beyond the scope of this paper to compare the results using our family with those using another, such as the skew-exponential power, although we believe that much of the foregoing points to the sinh-arcsinh family as being a particularly good choice.

Finally, this paper has been rather long in gestation and the first author has talked on the topic a number of times, including Jones (2005). It is therefore the case that the sinh-arcsinh distribution has already been implemented, under the acronym shash, in the GAMLSS software package (Stasinopoulos & Rigby, 2007).

ACKNOWLEDGEMENT

We are very grateful to Professor Mohsen Pourahmadi for drawing our attention to the Gebelein inequality, to Dr Chris Banks and Professor Paul Garthwaite for the snow depth data, to various colleagues for the enthusiasm with which they greeted versions of these ideas and to the referees for a positive reception and many useful suggestions that have greatly influenced the final article. The work of the second author was partially supported by the Spanish Ministry of Science and Education.

APPENDIX

Elements of the observed information matrix

Let $Y_i = (X_i - \xi)/\eta \;\; (i = 1, \ldots, n)$. The elements of the observed information matrix are:

$$-\frac{\partial^2 \ell}{\partial \xi^2} = \frac{1}{\eta^2}\left[\sum_{i=1}^{n} \frac{1 - Y_i^2}{\left(1 + Y_i^2\right)^2} + \delta \sum_{i=1}^{n} S_{\epsilon,\delta}^2(Y_i) \frac{\delta\left(1 + Y_i^2\right)^{1/2}\left\{3 + 2S_{\epsilon,\delta}^2(Y_i)\right\} - Y_i S_{\epsilon,\delta}(Y_i) C_{\epsilon,\delta}(Y_i)}{C_{\epsilon,\delta}^2(Y_i)\left(1 + Y_i^2\right)^{3/2}}\right],$$

$$-\frac{\partial^2\ell}{\partial\xi\partial\eta} = \frac{1}{\eta^2}\left[\sum_{i=1}^{n}\frac{Y_i\left(1-Y_i^2\right)}{\left(1+Y_i^2\right)^2} + \delta\sum_{i=1}^{n} Y_i S_{\epsilon,\delta}^2(Y_i)\frac{\delta\left(1+Y_i^2\right)^{1/2}\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\} - Y_i S_{\epsilon,\delta}(Y_i)C_{\epsilon,\delta}(Y_i)}{C_{\epsilon,\delta}^2(Y_i)\left(1+Y_i^2\right)^{3/2}}\right],$$

$$-\frac{\partial^2\ell}{\partial\xi\partial\delta} = -\frac{1}{\eta}\sum_{i=1}^{n} S_{\epsilon,\delta}^2(Y_i)\frac{S_{\epsilon,\delta}(Y_i)C_{\epsilon,\delta}(Y_i) + \delta\sinh^{-1}(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)\left(1+Y_i^2\right)^{1/2}},$$

$$-\frac{\partial^2\ell}{\partial\xi\partial\epsilon} = \frac{\delta}{\eta}\sum_{i=1}^{n}\frac{S_{\epsilon,\delta}^2(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)\left(1+Y_i^2\right)^{1/2}},$$

$$-\frac{\partial^2\ell}{\partial\eta^2} = \frac{1}{\eta^2}\left[\sum_{i=1}^{n}\frac{2Y_i^2}{\left(1+Y_i^2\right)^2} + \delta\sum_{i=1}^{n} Y_i S_{\epsilon,\delta}^2(Y_i)\frac{\delta Y_i\left(1+Y_i^2\right)^{1/2}\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\} + S_{\epsilon,\delta}(Y_i)C_{\epsilon,\delta}(Y_i)}{C_{\epsilon,\delta}^2(Y_i)\left(1+Y_i^2\right)^{3/2}}\right],$$

$$-\frac{\partial^2\ell}{\partial\eta\partial\delta} = -\frac{1}{\eta}\sum_{i=1}^{n} Y_i S_{\epsilon,\delta}^2(Y_i)\frac{S_{\epsilon,\delta}(Y_i)C_{\epsilon,\delta}(Y_i) + \delta\sinh^{-1}(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)\left(1+Y_i^2\right)^{1/2}},$$

$$-\frac{\partial^2\ell}{\partial\eta\partial\epsilon} = \frac{\delta}{\eta}\sum_{i=1}^{n}\frac{Y_i S_{\epsilon,\delta}^2(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)\left(1+Y_i^2\right)^{1/2}},$$

$$-\frac{\partial^2\ell}{\partial\delta^2} = \frac{n}{\delta^2} + \sum_{i=1}^{n}\frac{\{\sinh^{-1}(Y_i)\}^2 S_{\epsilon,\delta}^2(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)},$$

$$-\frac{\partial^2\ell}{\partial\delta\partial\epsilon} = -\sum_{i=1}^{n}\frac{\sinh^{-1}(Y_i)S_{\epsilon,\delta}^2(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)},$$

$$-\frac{\partial^2\ell}{\partial\epsilon^2} = \sum_{i=1}^{n}\frac{S_{\epsilon,\delta}^2(Y_i)\left\{3+2S_{\epsilon,\delta}^2(Y_i)\right\}}{C_{\epsilon,\delta}^2(Y_i)}.$$

References

Anderson, T. W. & Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193–212.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–8.

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

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. R. Statist. Soc.* B **65**, 367–89.

Azzalini, A. & Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *Int. Statist. Rev.* **76**, 106–29.

Barndorff-Nielsen, O. E. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.* **5**, 151–7.

Boos, D. D. (1982). A test for asymmetry associated with the Hodges–Lehmann estimator. *J. Am. Statist. Assoc.* **77**, 647–51.

Bowman, K. O. & Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika* **62**, 243–50.

Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

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

Cabilio, P. & Masaro, J. (1996). A simple test of symmetry about an unknown median. *Can. J. Statist.* **24**, 349–61.

Cassart, D., Hallin, M. & Paindaveine, D. (2008). Optimal detection of Fechner-asymmetry. *J. Statist. Plan. Infer.* **138**, 2499–525.

D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large sample size. *Biometrika* **58**, 341–8.

D'Agostino, R. B. (1972). Small sample probability points for the D test of normality. *Biometrika* **59**, 219–21.

DiCiccio, T. J. & Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *J. Am. Statist. Assoc.* **99**, 439–50.

DOWNTON, F. (1966). Linear estimates with polynomial coefficients. *Biometrika* **53**, 129–41.
FECHNER, G. T. (1897). *Kollectivmasslehre*. Leipzig: Engleman.
FERNÁNDEZ, C. & STEEL, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *J. Am. Statist. Assoc.* **93**, 359–71.
GEBELEIN, H. (1941). Das statistische Problem der Korrelation als Variations und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.* **21**, 364–79.
GENTON, M. G. Ed. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC, 2004.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1994). *Table of Integrals, Series, and Products*, 5th ed. San Diego, CA: Academic Press.
HOAGLIN, D. C. (1986). Summarizing shape numerically: the g-and-h distributions. In *Exploring Data Tables, Trends and Shapes*, Ed. D. C. Hoaglin, F. Mosteller and J. W. Tukey, pp. 461–513. New York: Wiley.
HOLLAND, P. W. & WANG, Y. J. (1987). Dependence function for continuous bivariate densities. *Commun. Statist.* A **16**, 863–76.
JARQUE, C. & BERA, A. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economet. Lett.* **6**, 255–9.
JOE, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.
JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76.
JOHNSON, N. L., KOTZ, S. & BALAKRISHNAN, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd ed. New York: Wiley.
JONES, M. C. (1996). The local dependence function. *Biometrika* **83**, 899–904.
JONES, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion). *Test* **13**, 1–43.
JONES, M. C. (2005). Contribution to the discussion of 'Generalized additive models for location, scale and shape', by R. A. Rigby and D. M. Stasinopoulos. *Appl. Statist.* **54**, 546–7.
JONES, M. C. & FADDY, M. J. (2003). A skew extension of the t distribution, with applications. *J. R. Statist. Soc.* B **65**, 159–74.
KAPPENMAN, R. F. (1988). Detection of symmetry or lack of it and applications. *Commun. Statist.* A **17**, 4163–77.
KOYAK, R. A. (1987). On measuring internal dependence in a set of random variables. *Ann. Statist.* **15**, 1215–28.
MA, Y. & GENTON, M. G. (2004). Flexible class of skew-symmetric distributions. *Scand. J. Statist.* **31**, 459–68.
MUDHOLKAR, G. S. & HUTSON, A. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *J. Statist. Plan. Infer.* **83**, 291–309.
NELDER, J. A. & MEAD, R. (1965). A simplex-method for function minimization. *Comp. J.* **7**, 308–13.
RIECK, J. R. & NEDELMAN, J. R. (1991). A log-linear model for the Birnbaum–Saunders distribution. *Technometrics* **33**, 51–60.
ROYSTON, P. (1995). AS R94: a remark on algorithm AS 181: the W-test for normality. *Appl. Statist.* **44**, 547–51.
SAMORODNITSKY, G. & TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Boca Raton, FL: Chapman and Hall/CRC.
SHAPIRO, S. & WILK, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.
SMITH, R. L. & NAYLOR, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Appl. Statist.* **36**, 358–69.
STASINOPOULOS, D. M. & RIGBY, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS). *J. Statist. Software* **23**, Issue 7.
TADIKAMALLA, P. R. (1980). Random sampling from the exponential power distribution. *J. Am. Statist. Assoc.* **75**, 683–6.
THADEWALD, T. & BÜNING, H. (2007). Jarque–Bera test and its competitors for testing normality: a power comparison. *J. Appl. Statist.* **34**, 87–105.
TIKU, M. L., ISLAM, M. Q. & SELCUK, A. S. (2001). Non-normal regression II – symmetric distributions. *Commun. Statist.* A **30**, 1021–45.
VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Amsterdam: Mathematisch Centrum.
ZHANG, J. & WU, Y. (2005). Likelihood-ratio tests for normality. *Comp. Statist. Data Anal.* **49**, 709–21.

[*Received August* 2008. *Revised April* 2009]