SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1093/BIOMET/ASP053

Sinh-arcsinh distributions — Source link

M. C. Jones, Arthur Pewsey

Institutions: Open University, University of Extremadura

Published on: 01 Dec 2009 - Biometrika (Oxford University Press)

Topics: Stability (probability), Heavy-tailed distribution, Normal distribution, Skew normal distribution and Location parameter

Related papers:

- A class of distributions which includes the normal ones
- On Bayesian modeling of fat tails and skewness
- Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution
- Systems of frequency curves generated by methods of translation.
- An Analysis of Transformations

Share this paper:

View more about this paper here: https://typeset.io/papers/sinh-arcsinh-distributions-hkch4o5ar3

Open Research Online

The Open University's repository of research publications
and other research outputs

Sinh-arcsinh distributions

Journal Item

How to cite:

Jones, M. C. and Pewsey, Arthur (2009). Sinh-arcsinh distributions. Biometrika, 96(4) pp. 761–780.

For guidance on citations see FAQs.

© 2009 Biometrika Trust

Version: [not recorded]

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.1093/biomet/asp053

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

Sinh-arcsinh distributions: a broad family giving rise to powerful tests of normality and symmetry

M.C. Jones

The Open University, UK

and Arthur Pewsey

University of Extremadura, Spain

Summary. We introduce the 'sinh-arcsinh transformation' and thence, by applying it to random variables from some 'generating' distribution with no further parameters beyond location and scale (which we take for most of the paper to be the normal), a new family of 'sinh-arcsinh distributions'. This four parameter family has both 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 appear to be superior to the state-of-the-art because of the range of alternatives against which they are very powerful. Likelihood ratio tests of symmetry are also available and very successful. Three-parameter symmetric and asymmetric subfamilies of the full family are of interest too. Heavy-tailed symmetric sinh-arcsinh distributions behave like Johnson S_U distributions while light-tailed symmetric sinh-arcsinh distributions behave like Rieck and Nedelman's 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 reparametrisation. A multivariate version is considered. Options and extensions are discussed.

Keywords: Heavy tails; Johnson's S_U distribution; Light tails; Sinh-normal distribution; Skew-normal distribution; Skewness; Transformation.

Address for correspondence: M.C. Jones, Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
`E-mail: m.c.jones@open.ac.uk`

1. Introduction

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

Broadly speaking, most of these families of distributions have the normal distribution as a special, often a limiting, case with other members of the families having heavier tails than the normal. In this paper, we propose a novel relatively simple and tractable four-parameter family of distributions on $\mathbb{R}$ with the normal distribution 'situated centrally' and other members having *both lighter and heavier tails*. This has practical benefits especially in affording excellent tests of the appropriateness of the normal distribution.

To describe the new distributions, consider their canonical case in which location $\mu \in \mathbb{R}$ and scale $\sigma > 0$ are removed; they can be reinstated for practical work in the usual way by utilising $\sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu))$ where $f_{\epsilon,\delta}(x)$ 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 tailweight. Associate random variables Z and $X_{\epsilon,\delta}$ with the standard normal density ϕ and $f_{\epsilon,\delta}$, respectively. Then, we propose to define $f_{\epsilon,\delta}$ by what we shorthandedly call the 'sinh-arcsinh transformation'

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

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

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

where $C_{\epsilon,\delta}(x) = \cosh(\epsilon + \delta \sinh^{-1}(x)) = \sqrt{1 + S^2_{\epsilon,\delta}(x)}$. Of course, $f_{0,1}(x) = \phi(x)$. Examples of densities (2) can be seen in Fig. 1 to follow. We note in passing that, unlike some other families of distributions, no special functions appear in the definition of the density of the sinh-arcsinh distribution above.

Properties of the full family (2) are considered in Section 2 and further properties of the three-parameter symmetric subfamily thereof (corresponding to $\epsilon = 0$) in Section 3. A considerable degree of tractability is evident in the provision of distribution and quantile functions, unimodality and moments. Tailweights are also considered. It is shown that ϵ (Section 2.2) and δ (in the symmetric case; Section 3.1) are skewness and kurtosis parameters in the sense of van Zwet (1964). A three-parameter subfamily of 'skew-normal' distributions is briefly described in Section 2.5. In Section 3.3 it is shown how, in the symmetric case, the small δ (heavy-tailed) members of family (2) behave like Johnson's (1949) S_U distributions while the large δ (light-tailed) members behave like Rieck and Nedelman's (1991) sinh-normal distributions. In this sense, the symmetric sinh-arcsinh distributions form a seamless combination of the two, the single-parameter δ controlling the transition from one to the other via the normal distribution ($\delta = 1$).

Likelihood fitting of the sinh-arcsinh distribution in the form of (2) with location and scale parameters introduced is considered in Section 4. Asymptotic properties are considered in Section 4.1, leading to a useful reparametrisation in Section 4.2. Although these subsections concentrate on the three-parameter symmetric subfamily of sinh-arcsinh distributions, we employ (and recommend) the same reparametrisation for use in fitting the full four-parameter family (Section 4.3). An example illustrating the modelling flexibility of the full sinh-arcsinh family is presented in Section 4.4.

Likelihood ratio tests (LRTs) of normality are immediately available within the sinh-arcsinh family: $H_0 : \epsilon = 0, \delta = 1$. The performance of these tests is investigated in a substantial simulation study reported in Section 5. We actually consider testing for normality against either symmetric or asymmetric alternatives and against alternatives both within and beyond the sinh-arcsinh family. We compare performance with that of seven of the best performing omnibus tests of normality and conclude that our LRTs appear to provide the best tests of normality.

A similar large simulation study of the sinh-arcsinh LRT for testing symmetry ($H_0 : \epsilon = 0$) was undertaken and is reported in Section 6. Again, we observe excellent performance and show that it outperforms two competing

omnibus tests chosen as representing the 'state-of-the-art'.

There is an immediate and straightforward extension of the univariate distributions above to the multivariate case by marginal transformation of a multivariate normal distribution. The resulting multivariate distributions are considered relatively briefly in Section 7 with some emphasis on their dependence properties.

In Section 8, we consider three ways in which the sinh-arcsinh distribution (2) might be/have been formulated differently. In Section 8.1, we discuss the choice of transformation function within the class of transformations of the form $H(\epsilon + \delta H^{-1}(X))$. In Section 8.2, we investigate alternative options to the normal for the role of the 'central' symmetric distribution in the family. And in Section 8.3, we explore a different approach to skewing the (same) symmetric members of the family. While there prove to be a number of interesting considerations and alternatives, the end result is a justification — for most general use — of the choices made in (2).

We close with discussion in Section 9.

2. Properties of family (2)

2.1. Basic properties

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

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

Also, $\sinh^{-1}(Z) = \epsilon + \delta\sinh^{-1}(X_{\epsilon,\delta})$ or $X_{\epsilon,\delta} = \sinh[\delta^{-1}\{\sinh^{-1}(Z) - \epsilon\}]$. Random variate generation is immediate using the latter formula.

Second, the distribution function associated with density (2) is readily written as

$$F_{\epsilon,\delta}(x) = \Phi(S_{\epsilon,\delta}(x)),$$

where Φ is the standard normal distribution function.

Third, since $S^{-1}_{\epsilon,\delta}(z) = S_{-\epsilon/\delta,1/\delta}(z)$, the quantile function associated with density (2) is

$$Q_{\epsilon,\delta}(u) = S_{-\epsilon/\delta,1/\delta}(\Phi^{-1}(u)), \qquad 0 < u < 1. \qquad (5)$$

In particular, the median of the distribution is $-\sinh(\epsilon/\delta)$.

Fourth, density (2) is always unimodal. To see this, the first derivative of $\log f_{\epsilon,\delta}(x)$ is of the form

$$-\frac{x}{1+x^2} - \frac{\delta S^3_{\epsilon,\delta}(x)}{\sqrt{1+x^2}\; C_{\epsilon,\delta}(x)}.$$

Any point x_0 for which this derivative is zero satisfies

$$\frac{\delta S^3_{\epsilon,\delta}(x_0)}{\sqrt{1+S^2_{\epsilon,\delta}(x_0)}} = -\frac{x_0}{\sqrt{1+x_0^2}}.$$

But the left-hand side of this equation is a monotonically increasing function of x_0 taking all real values while the right-hand side is a monotonically decreasing function of x_0 taking values from 1 to -1. It follows that there can only be one crossing point and so the density is unimodal. Of course, when $\epsilon = 0$, $x_0 = 0$, else $x_0 \neq 0$.

2.2. Skewness

First, in this subsection, let us note that $f_{-\epsilon,\delta}(x) = f_{\epsilon,\delta}(-x)$.

We can show that, for fixed δ, ϵ acts as a skewness parameter in the sense of van Zwet's (1964) skewness ordering. This ordering defines $G_1 \leq_2 G_2$ if $G_2^{-1}(G_1)$ is convex for all x. So now let $G_1 = F_{\epsilon_1,\delta}$ and $G_2 = F_{\epsilon_2,\delta}$ for $\epsilon_1 > \epsilon_2$. Then $F^{-1}_{\epsilon_2,\delta}(F_{\epsilon_1,\delta}(x)) = S_{c,1}(x)$, where $c = (\epsilon_1 - \epsilon_2)/\delta > 0$, and

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

which is positive because $S_{c,1}(x) > x$ for $c > 0$. Note that distribution (2) is parametrised in such a way that, while the absolute value of skewness increases with increasing $|\epsilon|$, positive skewness corresponds to negative ϵ.

This attractive result about monotonicity of skewness allows us to calculate the limits to the achievable range of skewness values in family (2). Consider the Bowley skewness (e.g. Bowley, 1937) defined by

$$B_{\epsilon,\delta} \equiv \frac{Q_{\epsilon,\delta}(3/4) - 2Q_{\epsilon,\delta}(1/2) + Q_{\epsilon,\delta}(1/4)}{Q_{\epsilon,\delta}(3/4) - Q_{\epsilon,\delta}(1/4)}.$$

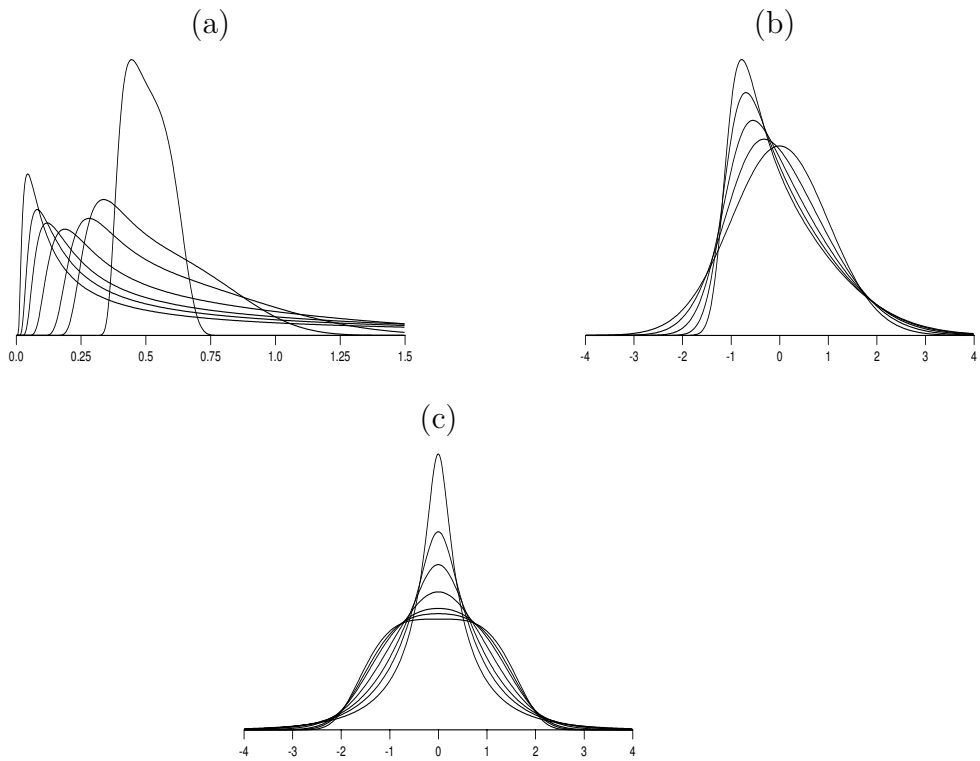

Figure 1: (a) densities $f_{-\infty,\delta}$ for, reading from left to right, $\delta = 0.5, 0.625, 0.75, 1, 1.5, 2, 5$; (b) normalised densities $\sigma_{\epsilon,1} f_{\epsilon,1}(\sigma_{\epsilon,1} x + \mu_{\epsilon,1})$ for, in increasing degree of skewness, $\epsilon = 0, -0.25, -0.5, -0.75, -1$; (c) scaled densities $\sigma_{0,\delta} f_{0,\delta}(\sigma_{0,\delta} x)$ for, in decreasing value of $\sigma_{0,\delta} f_{0,\delta}(0)$, $\delta = 0.5, 0.625, 0.75, 1, 1.5, 2, 5$.

This measure is monotone in ϵ because the distribution follows van Zwet's skewness ordering (Groeneveld and Meeden, 1984) and, in general, can take any values between -1 and 1. It is easy to show that, as $\epsilon \to \pm\infty$, $B_{\epsilon,\delta} \to \mp(k_\delta - 1)/(k_\delta + 1)$ where $k_\delta \equiv \exp(\sinh^{-1}(\Phi^{-1}(3/4))/\delta) \approx \exp(0.6316/\delta)$.

It is possible to identify the limiting densities $f_{\epsilon,\delta}$ as $\epsilon \to \pm\infty$. For concreteness, let us work with negative ϵ (positive skewness) and call the limiting densities $f_{-\infty,\delta}$. Employing suitable normalisation of mean and location, the limiting densities turn out to be

$$f_{-\infty,\delta}(y) = \frac{1}{\sqrt{2\pi}} \frac{\delta \cosh(\delta \log 2y)}{y} \exp\left\{ -\frac{1}{2} \sinh^2(\delta \log 2y) \right\},$$

with support $y > 0$. 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 Fig. 1(c).) Note that all the densities in Fig. 1(a) have median 1/2! Density $f_{5,-\infty}$ — which we shall shortly confirm is associated with very light tails — is not very skew, but most of the others are. Any limitations associated with the range of available skewness values determined above seem mild.

Similar consideration of the kurtosis role of δ is delayed until consideration of the symmetric subfamily in Section 3.1.

2.3. Tailweight

As $|x| \to \infty$, $S_{\epsilon,\delta}(x) \sim 2^{\delta-1}\mathrm{sgn}(x)\exp(\mathrm{sgn}(x)\epsilon)|x|^{\delta}$ and $C_{\epsilon,\delta}(x) \sim 2^{\delta-1}\exp(\mathrm{sgn}(x)\epsilon)|x|^{\delta}$. It follows that, retaining the position of ϵ (but not other constants) in asymptotic formulae even though it does not affect rates,

$$f_{\epsilon,\delta}(|x|) \sim \exp(\mathrm{sgn}(x)\epsilon)|x|^{\delta-1}\exp(-e^{\mathrm{sgn}(x)2\epsilon}|x|^{2\delta}). \tag{6}$$

Such tails 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 also see the effect of ϵ, through $\exp(\pm\epsilon)$, on the relative scales of the tails of the distribution. This is a major contributory factor to the way in which ϵ controls skewness.

2.4. Moments

The moments — which necessarily all exist as a consequence of the tail behaviour given by (6) — are available for family (2). Using the version of (3) associated with the inverse sinh-arcsinh transformation, we have

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

where

$$P_q = E\left\{\left(Z+\sqrt{Z^2+1}\right)^q\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(x+\sqrt{x^2+1}\right)^q e^{-x^2/2}dx$$

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

using (3.471.12) of Gradshteyn and Ryzhik (1994). A property of the modified Bessel function is that $K_{-\nu}(z) = K_\nu(z)$. It follows that $P_{-q} = P_q$, which confirms that odd moments of X are, indeed, zero in the symmetric case where $\epsilon = 0$.

In particular, we have for the mean

$$
\begin{aligned}
\mu_{\epsilon,\delta} &\equiv \mathrm{E}(X_{\epsilon,\delta}) = -\sinh(\epsilon/\delta) P_{1/\delta} \\
&= -\sinh(\epsilon/\delta) \frac{e^{1/4}}{\sqrt{8\pi}} \left(K_{(1+\delta)/(2\delta)}(1/4) + K_{(1-\delta)/(2\delta)}(1/4)\right)
\end{aligned}
$$

and for the variance,

$$
\begin{aligned}
\sigma^2_{\epsilon,\delta} &\equiv \mathrm{Var}(X_{\epsilon,\delta}) = \frac{1}{2}\left(\cosh(2\epsilon/\delta) P_{2/\delta} - 1\right) - \mu^2_{\epsilon,\delta} \\
&= \cosh(2\epsilon/\delta) \frac{e^{1/4}}{\sqrt{32\pi}} \left(K_{(2+\delta)/(2\delta)}(1/4) + K_{(2-\delta)/(2\delta)}(1/4)\right) - \frac{1}{2} - \mu^2_{\delta,\epsilon}.
\end{aligned}
$$

When $\epsilon = 0, \delta = 1$, it can be readily checked that $\mathrm{Var}(X_{0,1}) = 1$.

2.5. An asymmetric subfamily

There may also be some specific interest in the particular three-parameter subfamily of (2) in which $\delta = 1$. In this case transformation (1) has, through (4), the attractively simple form

$$
S_{\epsilon,1}(X) = \sinh(\epsilon)\ \sqrt{1 + X^2} + \cosh(\epsilon)\ X.
$$

These densities, some of which are displayed in Fig. 1(b), are true 'skew-normal' distributions in the sense of admitting normality as well as asymmetry and, by (6), retaining two normal-like tails. They share this property with 'two-piece' normal distributions (Fechner, 1897, Fernandez and Steel, 1998, Mudholkar and 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 popular 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

When $\epsilon = 0$ in transformation (1), density (2) is symmetric about 0. The properties discussed in Section 2 translate to the current special case in a straightforward way. (Inter alia, the mean, median and mode, of course, all reduce to 0 in this case.) In addition, computations strongly suggest that the tails of $f_{0,\delta}$ are sufficiently light for $f_{0,\delta}$ to be log-concave for all $\delta \geq 1$.

3.1. Kurtosis

We can show that, for $\epsilon = 0$, δ acts as a kurtosis parameter in the sense of van Zwet's (1964) ordering, which defines $G_1 \leq_S G_2$ for distributions G_1 and G_2 symmetric about zero if the function $G_2^{-1}(G_1)$ 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 $F_{0,\delta_2}^{-1}(F_{0,\delta_1}(x)) = S_{0,\delta}(x)$ where $\delta = \delta_1/\delta_2 > 1$. Then

$$\frac{d^2 F_{0,\delta_2}^{-1}(F_{0,\delta_1}(x))}{d^2 x} = \frac{\delta\sqrt{1+S_{0,\delta}^2(x)}}{1+x^2}\left(\frac{\delta S_{0,\delta}(x)}{\sqrt{1+S_{0,\delta}^2(x)}} - \frac{x}{\sqrt{1+x^2}}\right),$$

two d's have been changed to δ's here which is positive because $\delta > 1$ and, correspondingly, $S_{0,\delta}(x) > x$ for $x > 0$.

From Section 2.4, we find

$$\begin{aligned} E(X_{0,\delta}^4) &= \frac{e^{1/4}}{\sqrt{512\pi}}\left\{K_{(4+\delta)/(2\delta)}(1/4) + K_{(4-\delta)/(2\delta)}(1/4)\right. \\ &\qquad \left. - 4\left(K_{(2+\delta)/(2\delta)}(1/4) + K_{(2-\delta)/(2\delta)}(1/4)\right)\right\} + \frac{3}{8}\,. \end{aligned}$$

It can be checked that $E(X_{0,1}^4) = 3$. Given that $f_{0,\delta}$ obeys van Zwet's ordering, the classical kurtosis measure $\beta_2 = E(X_{0,\delta}^4)/\sigma_{0,\delta}^4$ must be monotone decreasing in δ.

3.2. Graphs of density

A range of symmetric members of family (2) is plotted in Fig. 1(c). The densities have been scaled to unit variance by use of the formula for the variance in Section 2.4 (with $\epsilon = 0$). Because of this, the densities are in the reverse order at 0 to what they would have been unscaled, for then $f_{0,\delta}(0) = \delta/\sqrt{2\pi}$. Unimodality, tailweight and kurtosis properties from above are well illustrated by this picture. Notice how the densities vary from the heavy tailed when δ is small, through the normal when $\delta = 1$, to 'wide-bodied'/light tailed densities when δ is large.

3.3. Links to Johnson S_U and sinh-normal distributions

Consider again 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 $\mathbb{R}$. Again, δ controls tailweight. This paper, of course, concerns the transformation $T_\delta(X) = S_{0,\delta}(X) = \sinh(\delta \sinh^{-1}(X))$. The two 'component parts' of transformation $S_{0,\delta}(X)$, the sinh and arcsinh transformations, have each previously been employed separately in the same manner. First, when $T_\delta(X) = \delta \sinh^{-1}(X)$, we have the symmetric members of Johnson's (1949) S_U distributions, part of the famous family of transformation-based distributions which also have members on $\mathbb{R}^+$ and $[0, 1]$. See Johnson et al. (1994, Section 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 and 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 and Nedelman (using different notation) the sinh-normal distribution is log-concave for $\delta' \geq 1$, but there is a problem for $\delta' < 1$: the distribution is then bimodal. This is unattractive both because of the form of the bimodality which 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

$$f_{0,\delta}(x) \simeq \frac{1}{\sqrt{2\pi}} \frac{\delta}{\sqrt{1+x^2}} \exp\left[-\frac{1}{2}\{\delta \sinh^{-1}(x)\}^2\right].$$

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

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

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 which 'centre on' the normal distribution, behave very much like Johnson's S_U distributions for tailweights heavier than normal, and like Rieck and 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 and 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. For small δ'', $T_{\delta''}(X) \simeq \delta'' \sinh(X)$ which, again, affords Rieck and Nedelman's (1991) sinh-normal distributions. However, these correspond to small $\delta' = \delta''$ cases of the sinh-normal distribution and hence to bimodality.

4. On maximum likelihood estimation

For fitting to one-sample data, family (2) is expanded to a four-parameter family by the addition of location, μ, and scale, σ, parameters in the usual way i.e. by fitting $\sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x - \mu))$. The theoretical work to follow in Sections 4.1 and 4.2 concentrates specifically on the symmetric, $\epsilon = 0$, case. However, this work informs our fitting of the full model also, as described in Section 4.3. Note also that one-sample considerations generalise readily to the important wide class of regression situations in which the sinh-arcsinh distribution can be used to provide a general family of response conditional distributions and location (and possibly one or more other parameters) is modelled as a simple parametric, e.g. linear, function of covariates.

4.1. Maximum likelihood asymptotics in the symmetric case

Manipulations to derive the score equations and elements of the observed information matrix are standard if tedious, and are not given here. We move straight to consideration of the expected information matrix which is n times the matrix made up of values of $\iota_{\eta\xi} = E\left\{-(\partial^2 \ell / \partial\eta\partial\xi)(Y)\right\}$, $\eta, \xi = \{\mu, \sigma, \delta\}$. We find we have the special structure

$$\iota_{\mu\mu} = f_m(\delta)/\sigma^2, \qquad \iota_{\mu\sigma} = 0, \qquad \iota_{\mu\delta} = 0,$$

$$\iota_{\sigma\sigma} = f_s(\delta)/\sigma^2, \qquad \iota_{\sigma\delta} = f_c(\delta)/\sigma, \qquad \iota_{\delta\delta} = f_d(\delta),$$

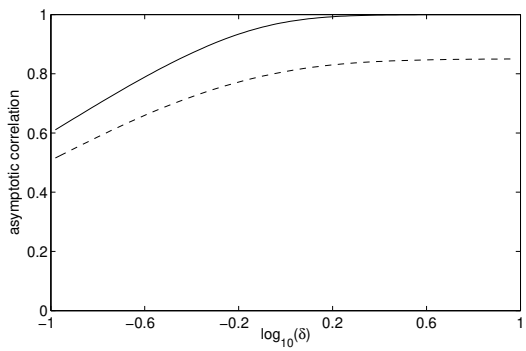

Figure 2: The asymptotic correlations between $\hat{\sigma}$ and $\hat{\delta}$ (solid line) and between $\hat{\sigma}_\delta$ and $\hat{\delta}$ (dashed line) in the symmetric case, plotted as a function of $\log_{10}\delta$.

say, where the f functions are all independent of μ and σ. This structure is a consequence of the symmetry of the fitted model. In fact, we have

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

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

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

and

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

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

It is immediately clear that the location and scale parameters are asymptotically independent as are the location and shape (δ) parameters. However, because $\iota_{\sigma\delta} \neq 0$, the scale and shape parameters are not asymptotically independent. In fact, $\mathrm{Corr}(\hat{\sigma}, \hat{\delta})$, which does not depend on (μ or) σ asymptotically, equals

$$-\frac{\iota_{\sigma\delta}}{\sqrt{\iota_{\sigma\sigma}\iota_{\delta\delta}}} = -\frac{f_c(\delta)}{\sqrt{f_s(\delta)f_d(\delta)}}.$$

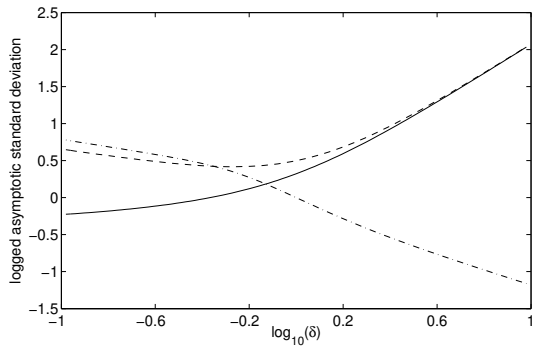

Figure 3: Log_{10} of the asymptotic standard deviation plus $\frac{1}{2}\log_{10} n$ for $\hat{\delta}$ (solid line), $\hat{\mu}$ (dot-dashed line) and $\hat{\sigma}$ (dashed line) in the symmetric case plotted as a function of $\log_{10}\delta$. The solid curve is actually the log standard deviation minus $\log_{10}\delta$ while the other two curves depict the log standard deviation minus $\log_{10}\sigma$.

It is clear that $\iota_{\sigma\delta} < 0$ and hence that the asymptotic correlation between $\hat{\sigma}$ and $\hat{\delta}$ is positive. This correlation can be plotted as a function of δ (solid line in Fig. 2). The correlation is very high for almost all δ. At first, this is disappointing, but it proves to be a standard property of scale/tailweight families of symmetric distributions and reflects the fact that one cannot really tell the difference between changing scale and changing tailweight at all easily in practice.

It is also the case that the asymptotic variance of $\hat{\delta}$ does not depend on σ; it is given by n^{-1} times

$$\frac{\iota_{\sigma\sigma}}{\iota_{\sigma\sigma}\iota_{\delta\delta} - \iota_{\sigma\delta}^2} = \frac{f_s(\delta)}{f_s(\delta)f_d(\delta) - f_c^2(\delta)}.$$

The logged *relative* asymptotic standard deviation (plus $\frac{1}{2}\log_{10} n$) is plotted as the solid curve in Fig. 3; it is necessarily rather large. (See Section 4.3 for comments on the practical effect of this.) While the location parameter μ is in the happy position of being estimated asymptotically independently of σ and δ, the asymptotic variances of the estimates of each are of the form $n^{-1}\sigma^2 h_i(\delta)$ where $i = \mu, \sigma$. So, reasonably enough, both standard deviations increase in direct proportion to the value of σ. We have that

$$h_\mu(\delta) = \frac{1}{f_m(\delta)} \quad \text{and} \quad h_\sigma(\delta) = \frac{f_d(\delta)}{f_s(\delta)f_d(\delta) - f_c^2(\delta)}.$$

These two functions are also shown, square rooted and logged, in Fig. 3 as dotted and dashed lines, respectively.

4.2. Reparametrisation

In principle, at least, it is possible to provide an orthogonal parametrisation of the form $(\mu, \sigma F(\delta), \delta)$. Since the correlation between $\hat{\sigma} F(\hat{\delta})$ and $\hat{\delta}$ is proportional to $(\log F)'(\delta) f_s(\delta) - f_c(\delta)$, this would be achieved by setting $(\log F)'(\delta) = f_c(\delta)/f_s(\delta)$. Unfortunately, this is insufficiently tractable to provide a workable formula. However, as shown in Fig. 2, the asymptotic correlation between $\hat{\sigma}$ and $\hat{\delta}$, which we are trying to alleviate via reparametrisation, is highest for large δ. This suggests seeking a large δ approximation to the above.

To this end, we find that, for large δ, $f_c(\delta) \simeq -(C+S)/\delta$ and $f_s(\delta) \simeq 1+S$ where

$$C = E\left\{\frac{Z^3 \sinh^{-1}(Z)}{\sqrt{1+Z^2}}\right\} \quad \text{and} \quad S = E\left[\frac{Z^2\{\sinh^{-1}(Z)\}^2(3+2Z^2)}{1+Z^2}\right].$$

Numerically, we find that $C \approx 1$, at least correct to 7 decimal places (we have been unable to prove exact equality to unity). We then find that $F(\delta) \simeq \delta^{-1}$, so suggesting a simple reparametrisation in which σ is replaced by $\sigma_\delta \equiv \sigma/\delta$. The asymptotic correlation between $\hat{\sigma}_\delta$ and $\hat{\delta}$ is

$$-\frac{\delta f_c(\delta) + f_s(\delta)}{\sqrt{f_s(\delta)\{\delta^2 f_d(\delta) + 2\delta f_c(\delta) + f_s(\delta)\}}}.$$

This is plotted as the dashed line in Fig. 2. It is clear that we have achieved a general lowering of the asymptotic correlation to less extreme values. We have not achieved the very small correlation for large δ that might have been expected because the variance of $\hat{\sigma}_\delta$ tends to zero alongside the covariance for large δ. However, the reduction in correlation that we have achieved proves to make a considerable difference in practice.

4.3. Practical implementation in the general case

We employ (and recommend) the reparametrisation just derived in fitting the full four-parameter sinh-arcsinh distribution (as well as its symmetric subfamily) to data, i.e. utilising $\{\mu, \sigma_\delta, \epsilon, \delta\}$ and then setting $\hat{\sigma} = \hat{\sigma}_\delta \hat{\delta}$. This solved severe numerical problems encountered in the original parametrisation when $\delta > 1$. We made use of the Nelder and Mead (1965) simplex

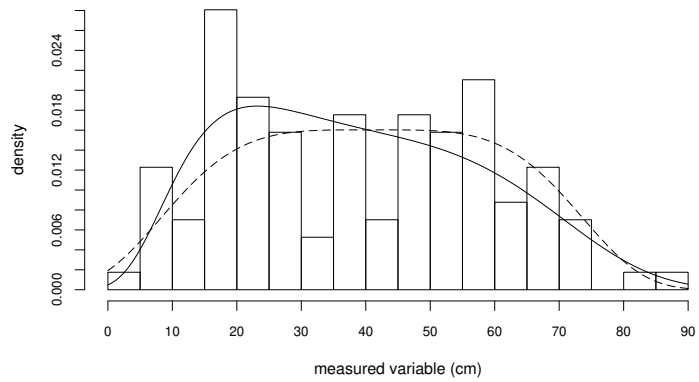

Figure 4: Histogram of the snow depth data together with the fitted densities for family (2) (solid line) and its symmetric subfamily (dashed line).

algorithm to perform maximisation of the log-likelihood. Using this direct search approach, it proves helpful to optimise over $\mu/\sqrt{1+\mu^2} \in (-1,1)$, $\sigma_\delta/(1+\sigma_\delta) \in (0,1)$, $\epsilon/\sqrt{1+\epsilon^2} \in (-1,1)$ 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 log-likelihood surface. However, as is generally the case when using numerical optimisation techniques, it is advisable to try a range of different starting values in an attempt to ensure that the global maximum is identified. We find that each of μ, σ_δ and ϵ is estimated well but large δ-values are not estimated so precisely. The log-likelihood surface remains flat when δ is large, corresponding to the large asymptotic variance of $\hat{\delta}$ shown in Fig. 3.

4.4. Example

In order to briefly illustrate the modelling flexibility of family (2), we present an analysis of $n = 114$ measurements of the depth of snow (in cm) taken on an ice floe in the eastern Asmundsen Sea, Antarctica, in March 2003. See Banks (2006, Chapter 6) 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. 4. 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

Table 1: Parameter estimates for the fits to the 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 maximised log–likelihood (l_{max}), AIC and BIC values, and p-value for the chi–squared goodness-of-fit test, are included as fit diagnostics.

	Model			
Parameter	Normal	Normal tails	Symmetric	Family (2)
μ	39.24	24.66	40.49	-52.91
σ	20.20	17.88	349139.3	34.27
δ	(1)	(1)	14028.5	3.99
ϵ	(0)	-0.52	(0)	-6.75
l_{max}	-504.39	-502.50	-497.72	-494.98
AIC	1012.78	1011.00	1001.44	997.96
BIC	1018.25	1019.21	1009.65	1008.90
p-value	0.016	0.004	0.228	0.213

1. All three likelihood-based diagnostics in Table 1 indicate that the fit for the full family, with its lighter than normal tails ($\delta > 1$) and positive asymmetry ($\epsilon < 0$), is best, followed by that for its symmetric subfamily. The p-values of likelihood-ratio tests for normality, normal tails and symmetry, calculated using the usual asymptotic chi-squared approximation, are 0.000, 0.000 and 0.019, respectively. So, the fit for the full family appears to offer a significantly better fit than any of its three submodels. Table 1 also contains the p-values for the chi-squared goodness-of-fit test performed using the class intervals of the histogram shown, some of which were combined to obtain expected values of at least 5. The p-values support the adequacy of the two best fits and rule out the normal and normal-tailed submodels. The densities of the two best fits are superimposed on the histogram in Fig. 4. Comparing them with the histogram, there is perhaps some indication of multimodality in the data. However, this could be an artifact of the binning used and the rounding of the data to the nearest whole cm during measurement. It would certainly be difficult to conceive of a better unimodal fit to the data.

In the next two sections we present a substantial practical investigation of the use of sinh-arcsinh distributions in testing normality and symmetry.

5. Testing normality

The central position of the normal distribution within family (2) allows testing of normality within the family via standard likelihood ratio tests (LRTs). However, since family (2) is sufficiently broad to, at some level, provide an approximation to any unimodal distribution, we propose that sinh-arcsinh-based LRTs of normality also be used as general purpose tests of normality. To that end, in this section, we explore the size and power of sinh-arcsinh-based LRTs of normality both within and beyond the sinh-arcsinh family of distributions.

5.1. Testing normality against symmetric alternatives I: size

It is probably most usual to test for normality in a situation where one is willing to assume symmetry of the distribution of interest. In that case, the appropriate LRT is a statistic of the form $L = -2\log(\ell_0/\ell_1)$ where ℓ_0 represents the maximum of the log-likelihood function for an assumed normal distribution and ℓ_1 the maximum of the log-likelihood function assuming that the sample was drawn from a *symmetric* sinh-arcsinh distribution i.e. $\sigma^{-1}f_{0,\delta}(\sigma^{-1}(x-\mu))$ where $f_{\epsilon,\delta}(x)$ is given by (2). ℓ_1 has to be calculated numerically. This being a regular problem, the asymptotic distribution of L is, of course, χ^2_1 (the single degree of freedom being associated with setting $\delta = 1$ in the symmetric subfamily to achieve normality). For testing normality against asymmetric alternatives, see Section 5.3.

We investigated the distribution of 10,000 values of L based on samples generated from the standard normal distribution. The χ^2 approximation to the sampling distribution of L is, as expected, poor for small sample sizes, rapidly improving with increasing n such that for $n \geq 50$ it would appear to provide a very good approximation to the true sampling distribution. However, the tails of the χ^2 approximation to the simulated sampling distribution are in reasonably close agreement even for small n. Indeed, in addition to being adequate for large n, the asymptotic critical values for the test are closest to the simulated critical values in certain cases associated with small n! Overall, we consider it reasonable, as well as simplest, to use the critical values of the asymptotic χ^2_1 distribution when analysing samples of any size. This is further vindicated in the studies that follow in the next section.

5.2. Testing normality against symmetric alternatives II: power

There exists a well-established literature addressing the problem of testing univariate data for normality. Renewed recent interest in this inferential problem can be found in the papers of Zhang and Wu (2005) and Thadewald and Büning (2007), amongst others. In the light of the findings presented in those two papers, we conducted a simulation study designed to compare the performance of the LRT of normality with those of the following seven competitive tests for a nominal significance level of 5% (the description of each test is preceded by the abbreviation we will use when referring to it):

JB. The (one-sided) test of Jarque and Bera (1980), the test statistic of which is a function of the coefficients of skewness and kurtosis. We used the corrected critical values for this test presented in Table 2 of Thadewald and Büning (2007).

D. The (two-sided) test of D'Agostino (1971, 1972). Up to a constant, the test statistic is the ratio of Downton's (1966) linear estimator of the standard deviation to the sample standard deviation. The critical values for this test are given in D'Agostino's papers.

AD. The (one-sided) empirical distribution function (EDF)-based test of Anderson and Darling (1952). We used the corrected critical values for this test presented under the name CMW in Table 2 of Thadewald and Büning (2007). (Those authors do not seem to realise that their CMW statistic is in fact the A^2 statistic of Anderson and Darling.)

CM. The (one-sided) Cramér-von Mises EDF-based test with statistic identified as CM in Thadewald and Büning (2007). We used the corrected critical values given in their Table 2.

SW. The (one-sided) test of Shapiro and Wilk (1965). We used Algorithm AS R94 of Royston (1995) to compute the test statistic and its p-value.

ZA and ZC. The (one-sided) nonparametric likelihood-ratio-based tests with test statistics Z_A and Z_C of Zhang and Wu (2005). We used the corrected critical values for these tests given in Tables 1 and 2, respectively, of that paper.

Our power simulations concern two sets of alternative distributions; here is the first. For each combination of $n = 10, 20, 50, 100, 200$ and $\delta = 0.2, 0.4, 0.6, 1, 2, 10$, $10,000$ random samples of size n were simulated from the symmetric subfamily of (2) with $\epsilon = 0$. All eight tests were then applied to

each simulated sample. (Setting $\mu = 0$ and $\sigma = 1$ throughout is appropriate since all tests are location/scale invariant.) In Fig. 5, the proportion of the 10,000 samples for which the null hypothesis of normality was rejected in a nominally 5% test is plotted against $\lambda = \delta/(1+\delta)$. Remember that $\delta = 1$, i.e. $\lambda = 0.5$, corresponds to a normal distribution, so this figure provides information concerning both the true size and power of the different tests.

Considering the content of Fig. 5, we can draw the following conclusions. Firstly, all seven rival tests maintain the nominal significance level of 5% very closely. So does the LRT, in general, although it is slightly conservative for $n = 10$ (size $\simeq 0.03$) and slightly liberal for $n = 50$ (size $\simeq 0.08$). Secondly, the LRT and D tests have the best overall power characteristics; the LRT and CM tests are the most powerful against alternatives with $0 < \lambda < 0.45$ (i.e. distributions with tails that are far heavier than normal), and, for $n \geq 20$, the LRT is the most powerful against alternatives with lighter than normal tails ($\lambda > 0.5$). D has second best power for these latter alternatives; it also has high power for λ-values in the region of 0.6–0.8, as does JB. However, for alternatives with lighter than normal tails, JB has the worst power signature. Indeed, even for samples as large as 100, its power generally lies below the nominal significance level. As is to be expected, the power of all eight tests generally increases with n for fixed δ. The increase in power with n is particularly noteworthy against the alternatives with lighter than normal tails ($\lambda > 0.5$). Note that these results provide interesting further information concerning the relative performance of the competing test to complement the findings of Zhang and Wu (2005) and Thadewald and Büning (2007), particularly concerning the ZA and ZC tests in the former and the Jarque-Bera test in the latter.

Of course, testing *within* the symmetric sinh-arscinh family, as just considered, is the situation for which our LRT was designed and for which it must be expected to be particularly strong as, gratifyingly, it proved. The second set (of three) alternative distributions are not members of class (2). These are: (i) the very heavy-tailed t distribution on two degrees of freedom (t_2); (ii) the fairly-heavy-tailed logistic distribution; and (iii) a light-tailed distribution due to M.L. Tiku with density $16(1+x^2/4)^2\phi(x)/27$ (e.g. Tiku, Islam and Selcuk, 2001). The results of our power simulations against these alternative distributions are given in Fig. 6.

From Fig. 6(a), corresponding to the t_2 alternative distribution, it can be seen that all eight tests are relatively powerful. For larger values of n there is little difference in their powers; for $n = 20, 50$, JB and D tend to dominate.

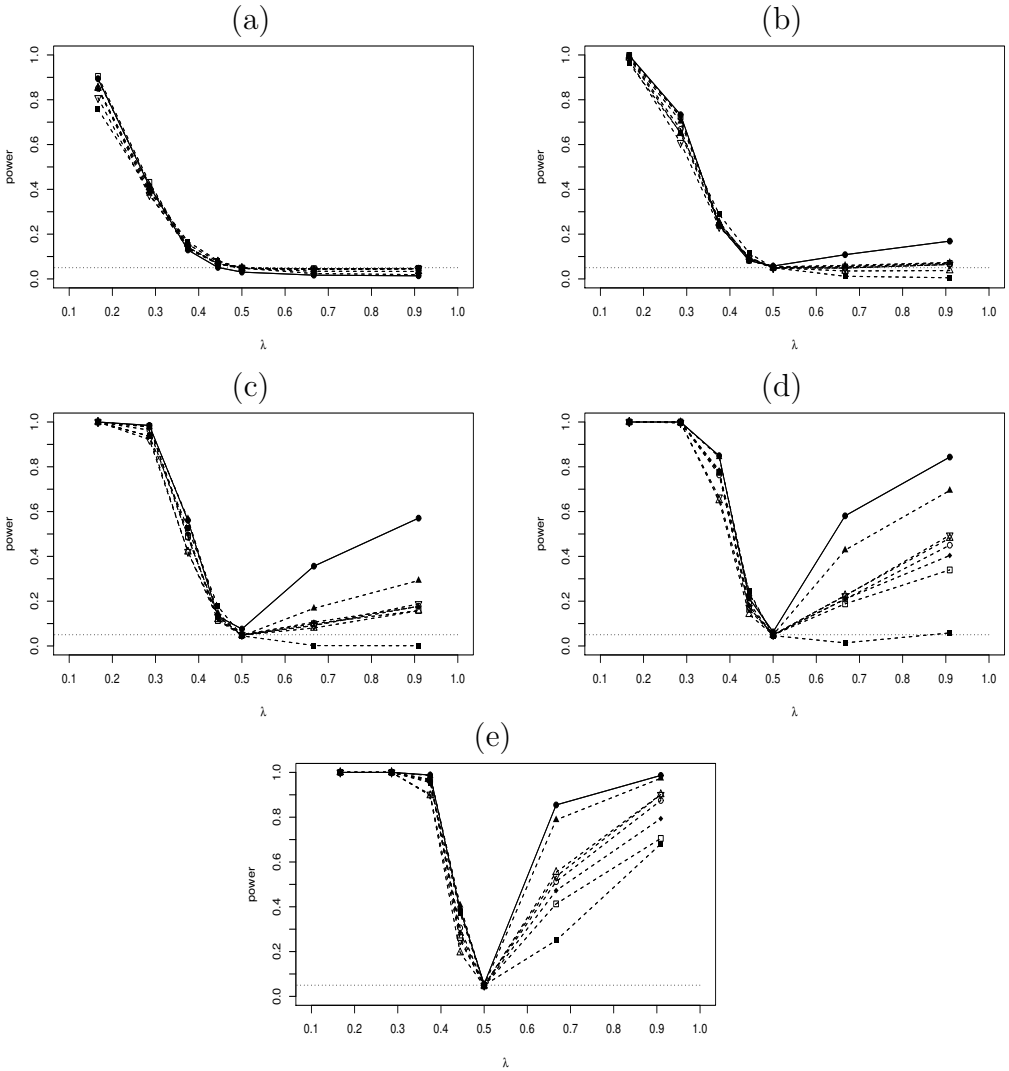

Figure 5: The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from alternative sinh-arcsinh distributions with $\epsilon = 0$, $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other seven tests: JB (solid square); D (solid triangle); AD (solid diamond); CM (open square); SW (open circle); ZA (open triangle); ZC (open inverted triangle). The dotted line is at the nominal level of 0.05.

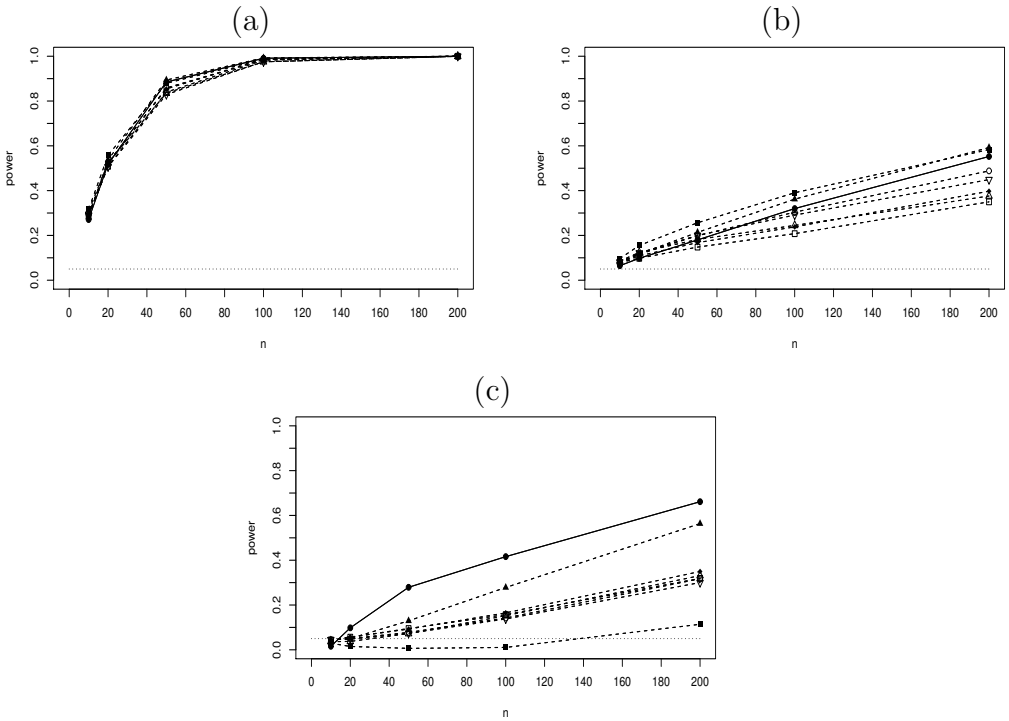

Figure 6: The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test plotted against n. The proportions were calculated using 10,000 random samples of size $n = 10, 20, 50, 100, 200$ from the: (a) t_2; (b) logistic; and (c) Tiku short-tailed distributions. The solid lines connect the results of the LRT, and the dashed lines those for the other seven tests: JB (solid square); D (solid triangle); AD (solid diamond); CM (open square); SW (open circle); ZA (open triangle); ZC (open inverted triangle). The dotted line is at the nominal level of 0.05.

Fig. 6(b) portrays the equivalent results for the logistic distribution. Clearly, none of the tests is very powerful against this alternative. The JB test has the best overall performance. D also performs relatively well, particularly for larger values of n. The LRT performs relatively poorly for samples of size $n \leq 50$ but its relative performance improves with increasing n. The performance of CM is worst overall. Finally, the results for Tiku's short-tailed distribution are displayed in Fig. 6(c). Again, none of the tests is particularly powerful. Indeed, for samples of size 10 and 20 the power lies below, and bobs around, respectively, the nominal level of the tests. LRT is clearly the most powerful, followed by D. The powers of five of the other six tests are very similar, with the JB test being very poor.

Overall, the LRT seems very competitive in most symmetric situations with the best of existing tests which would appear to be D'Agostino's test D.

5.3. Testing normality against asymmetric alternatives

If one is not willing to assume symmetry, testing for normality can still be accomplished from within the full four-parameter sinh-arcsinh family. The appropriate LRT now compares the maximised log-likelihood function for an assumed normal distribution with the maximum of the log-likelihood assuming the sample was drawn from $\sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu))$, the asymptotic distribution of the LRT statistic now being χ^2_2. Simulations from the normal distribution yielded results in keeping with the test's ability, as in the symmetric case, to maintain its nominal significance level using its asymptotic sampling distribution.

Because symmetry is no longer being assumed in constructing the test statistic, the power of the 'asymmetric LRT' is necessarily a little lower than that of the previous 'symmetric LRT' when normality is tested within a truly symmetric situation. The effect is quite small and the overall performance of the asymmetric LRT remains excellent. For example, if the powers of the symmetric LRT in Fig. 5 were replaced by those of the asymmetric LRT: (a) for large δ (light tails), the previous superiority of the symmetric LRT is reduced to a performance essentially on a par with the second-based method, namely D; (b) for very small δ (the heaviest tails), the LRT remains almost as good as the (otherwise best) CM test; (c) for other $\delta < 1$ (tails heavier than normal), the LRT continues to have a quality of performance which is in the middle of the pack of tests considered. These observations are also reflected

for the non-sinh-arcsinh symmetric alternatives of Fig. 6 in accordance with their relative tail weights.

Fig. 7 shows simulated powers for the asymmetric LRT and the same set of seven competing tests for normality within a set of asymmetric distributions, namely sinh-arcsinh densities with $\epsilon = 1$. Overall, the LRT is best. Indeed, the ordering of the power performances of the tests is by and large the same as in Fig. 7's symmetric counterpart (Fig. 5) with two notable exceptions: (i) the D test, which was previously competitive with the LRT, is very badly affected by the presence of asymmetry; and (ii) the LRT maintains its 'first place' even for alternatives with slightly heavier tails than those of the normal. The performance of D is particularly poor for a middle range of values of δ including fairly heavy tails when n is small, normal tails when n is small and moderate, and fairly light (but not the lightest) tails even when $n = 200$. We note also that, for small n, the combination of non-light tails ($\delta \leq 1$) and skewness makes for a greater disparity in power performance between the best and the poorest tests. In further experiments with a number of asymmetric alternatives outside the sinh-arcsinh class, the relative performances of tests described here — including the mostly leading performance of the LRT and the many poor performances of the D test — were upheld, again in accordance with their levels of tail weight.

5.4. Conclusion

Taking both symmetric and asymmetric alternatives into account, the LRT seems to be the best of the options considered here (and its competitors have been chosen because of claims of leading performance elsewhere).

6. Testing symmetry

We can also test for symmetry (about an unknown centre) by employing an LRT within the full sinh-arcsinh family of the null hypothesis that $\epsilon = 0$. The asymptotic null distribution of the test, which we shall use, is, again, χ^2_1.

We will compare the size and power performance of our LRT of symmetry (again, for a nominal significance level of 0.05) with those of two other general tests of symmetry. These particular tests were chosen because they were found to perform well in extensive simulation comparisons reported in Cabilio and Masaro (1996). They are:

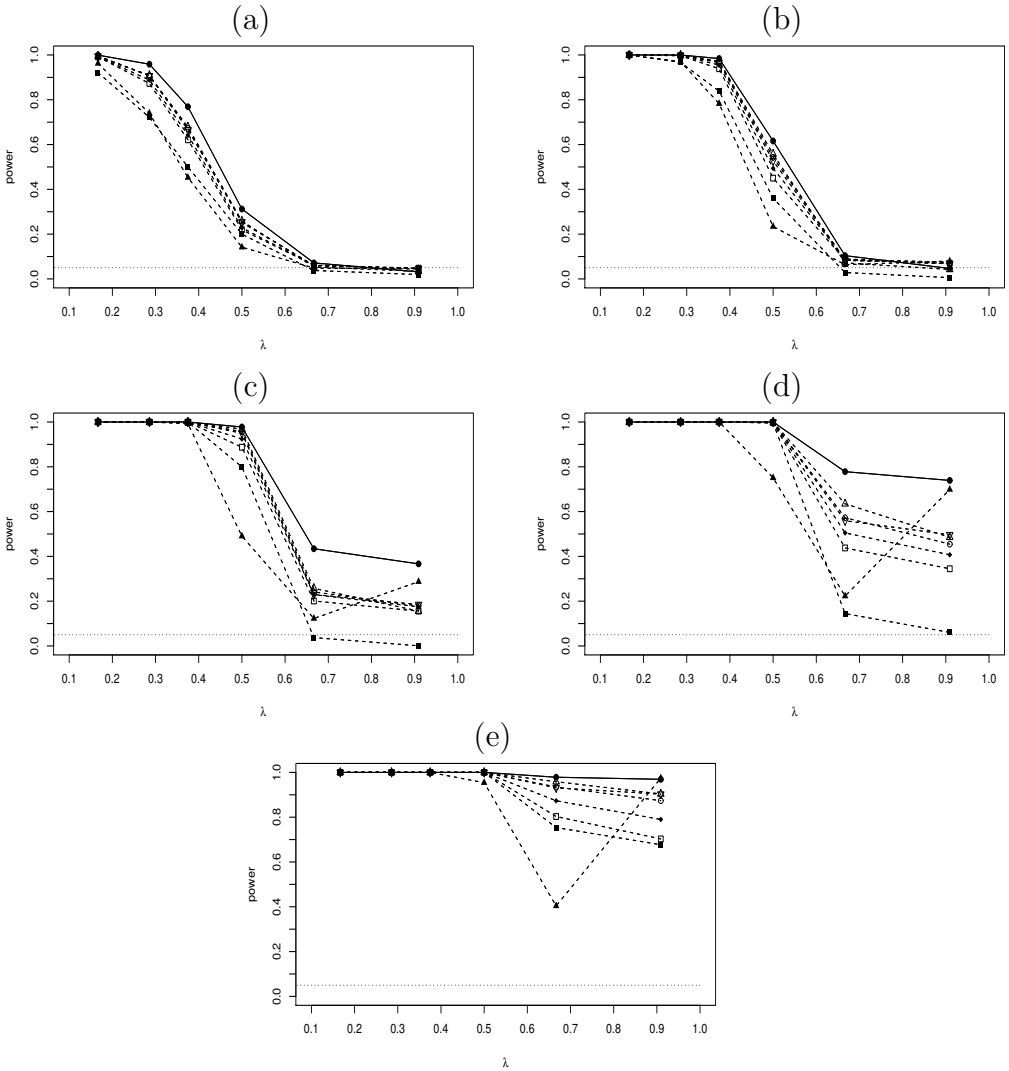

Figure 7: The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from asymmetric sinh-arcsinh distributions with $\epsilon = 1$, $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other seven tests: JB (solid square); D (solid triangle); AD (solid diamond); CM (open square); SW (open circle); ZA (open triangle); ZC (open inverted triangle). The dotted line is at the nominal level of 0.05.

SK. The test of Cabilio and Masaro (1996), the test statistic of which is the simple function $S_K = \sqrt{n}(\bar{X} - m)/s$ where $\bar{X}$, m and s denote the sample mean, median and standard deviation (with divisor n), respectively. We used the critical values for this test presented in Table 1 of Cabilio and Masaro (1996) which were calibrated against the normal distribution.

TN. The second test statistic was the more involved one of Boos (1982) which is based on the Hodges-Lehmann estimator. We used the critical values for this test presented in Table 1 of Boos (1982) which were calibrated against the logistic distribution.

6.1. Testing symmetry I: size

In Fig. 8, simulated values of the size of each of the LRT, SK and TN tests are presented for a variety of symmetric members of the sinh-arcsinh family. It can be seen that the LRT is by far the best test in terms of its overall ability to maintain the nominal significance level. SK tends to be very liberal when the distribution is either heavy- or light-tailed. TN is extremely liberal when the distribution is heavy-tailed and marginally liberal when it is light-tailed. When the underlying distribution is normal ($\lambda = 0.5$), all three tests maintain the nominal level increasingly well with increasing n.

We also computed the size of the tests of symmetry for data simulated from the t_2, logistic and Tiku distributions used above as symmetric alternatives to the normal distribution in Fig. 6. Summarising our results: (a) for the heavy-tailed t_2 distribution, SK holds the nominal level best, whilst the LRT and especially TN are markedly liberal; (b) for the logistic distribution, TN holds the nominal level well, SK is marginally conservative and the LRT marginally liberal; and (c) for Tiku's light-tailed distribution, all three tests hold the nominal level pretty well, with the LRT and TN holding it best, SK being rather liberal. These results collectively chime with earlier observations: Cabilio and Masaro (1996) observed that the size of their test, SK, is inflated when the underlying distribution is uniform or Cauchy, while both Boos (1982) and Cabilio and Masaro (1996) noted that TN can be extremely sensitive to heavy-tailed distributions, tending to mistakenly confuse such tails with asymmetry.

6.2. Testing symmetry II: power

We start this section by investigating the power of the LRT, SK and TN tests against alternative, asymmetric, distributions taken from the sinh-arcsinh

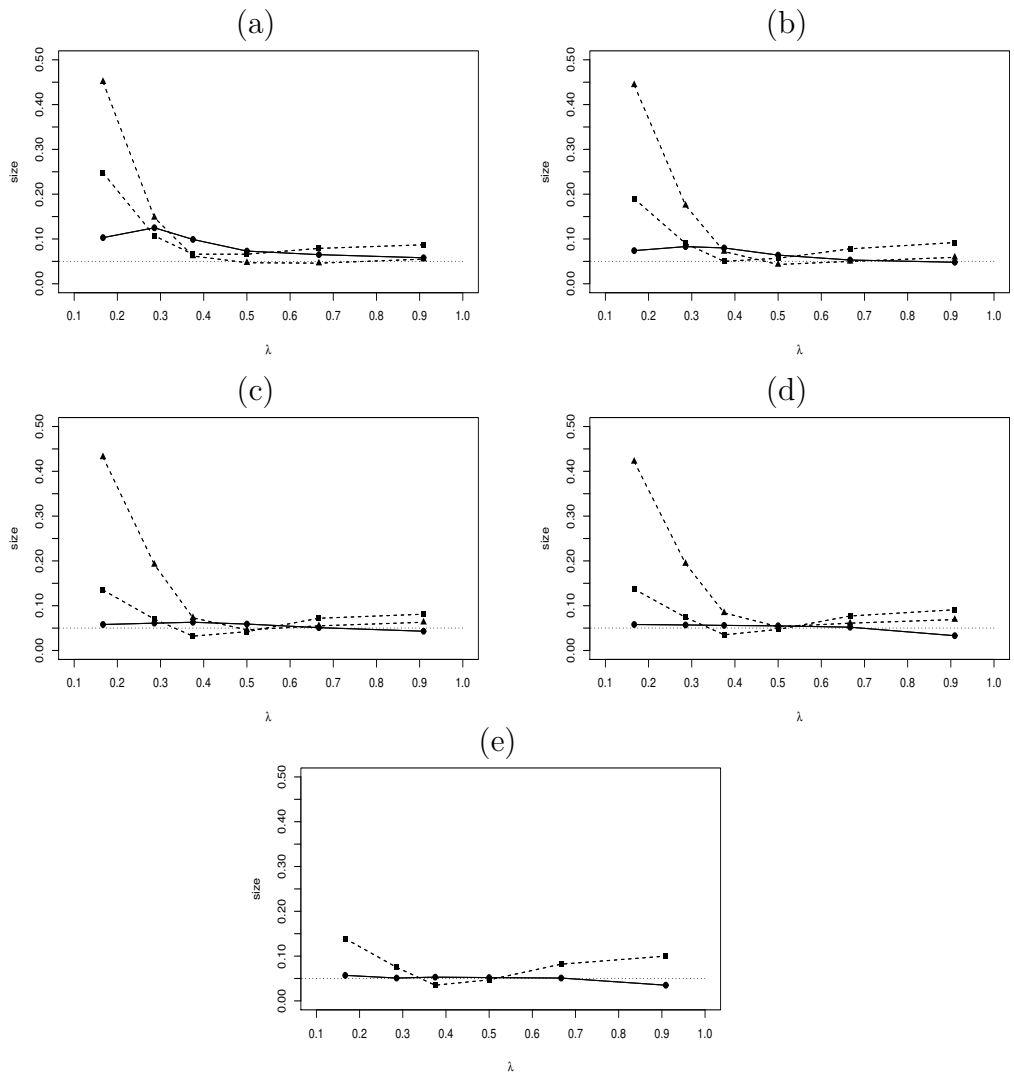

Figure 8: The proportion of samples for which the null hypothesis of symmetry was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from symmetric sinh-arcsinh distributions ($\epsilon = 0$), with $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other two tests: SK (square); TN (triangle). The dotted line is at the nominal level of 0.05. The results for TN are missing from panel (e) as the computational burden (in terms of storage) proved too much for our programs to handle.

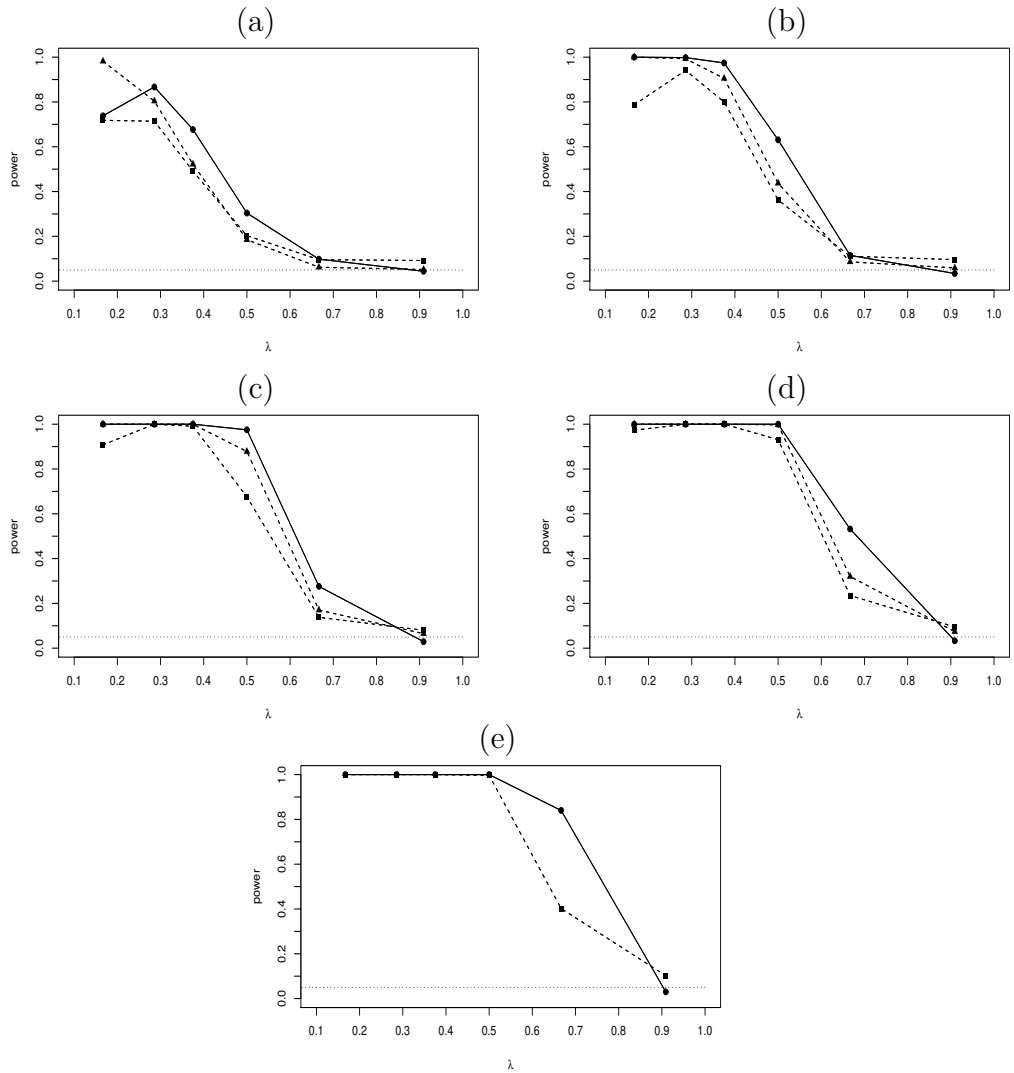

Figure 9: The proportion of samples for which the null hypothesis of symmetry was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from asymmetric sinh-arcsinh distributions with $\epsilon = 1$, $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other two tests: SK (square); TN (triangle). The dotted line is at the nominal level of 0.05. The results for TN are again missing from panel (e).

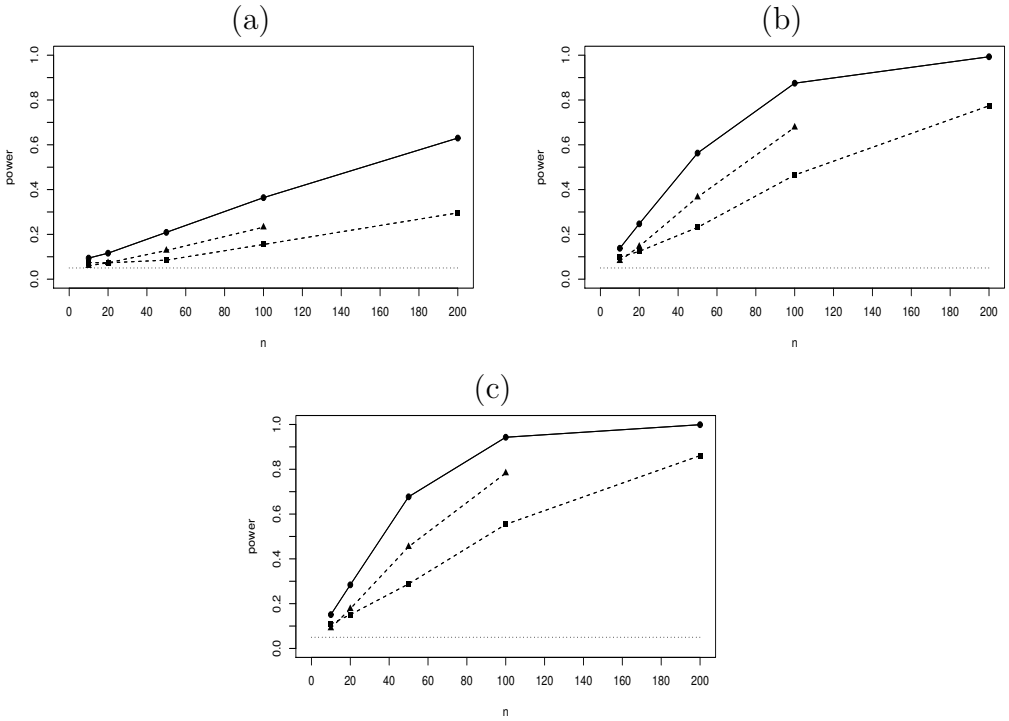

Figure 10: The proportion of samples for which the null hypothesis of symmetry was rejected in a nominally 5% test plotted against n. The proportions were calculated using 10,000 random samples of size $n = 10, 20, 50, 100, 200$ from the log F distribution with: (a) 4 and 2; (b) 16 and 2; and (c) 64 and 2 degrees of freedom. The solid lines connect the results of the LRT, and the dashed lines those for the other two tests: SK (square); TN (triangle). The dotted line is at the nominal level of 0.05. The results for TN are again missing for $n = 200$.

family with $\epsilon = 1$. See Fig. 9 for results. It can be seen that, overall, the LRT is the most powerful of the tests. For the smallest-sized sample from the heaviest-tailed distribution, TN has higher power but it should be recalled that this test is unable to maintain the nominal level for heavy-tailed symmetric distributions within family (2). The LRT also performs quite poorly against extremely light-tailed alternatives, but all three tests have very low power in such cases.

We also made extensive investigations of the power of these three tests to detect asymmetry outside the sinh-arcsinh family. To this end, we investigated — in order of increasing tail weight — the extreme value distribution and a range of skew-normal (Azzalini, 1985), log F (e.g. Baghdachi and Balakrishnan, 2008) and skew t (Jones and Faddy, 2003) distributions. Results for each distribution were broadly similar: the LRT was most powerful, followed by the TN test and then the SK test. Fig. 10 shows these results for a range of log F distributions, those for extreme value and skew-normal alternatives looking very similar. Only in the case of the heavy-tailed skew t distributions was the power performance of the LRT closely matched by that of TN.

6.3. Conclusion

The sinh-arcsinh-based LRT clearly outperforms two omnibus tests for symmetry that we chose for comparison as being 'state-of-the-art', both (unsurprisingly) within the sinh-arcsinh family but also (much less necessarily) in a wide range of situations outside the sinh-arcsinh family too.

7. The multivariate case

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

$$f_{\epsilon,\delta}(x) = \frac{1}{\sqrt{(2\pi)^d |R|}} \prod_{i=1}^{d} \left\{ \frac{\delta_i C_{\epsilon_i,\delta_i}(x_i)}{\sqrt{1+x_i^2}} \right\} \exp\left(-\frac{1}{2} S_{\epsilon,\delta}(x)' R^{-1} S_{\epsilon,\delta}(x) \right). \quad (7)$$

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

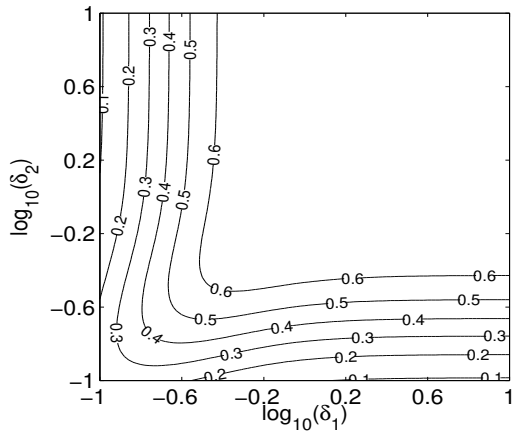

Figure 11: The correlation between X and Y in the symmetric marginals case, plotted as a function of $\log_{10}\delta_1$ and $\log_{10}\delta_2$: here, $\rho = 0.7$.

The univariate marginals of this distribution are sinh-arcsinh distributions by construction. If z is partitioned into (z_1, z_2) and x, X and R are partitioned conformably, $X_1|x_2$ is the distribution of $S^{-1}_{\epsilon_1,\delta_1}(Z_1)|z_2 = S_{\epsilon_2,\delta_2}(x_2)$ where $Z_1|z_2 \sim N(R_{12}(R_{22})^{-1}z_2, R_{11} - R_{12}^T(R_{22})^{-1}R_{12})$. Notice 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 further. The maintenance of unimodality in 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. 11 as a function of δ_1, the parameter in the x-direction, and δ_2, the parameter in the y-direction, for $\rho = 0.7$. 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 because of the positive (negative) quadrant dependence of the bivariate normal distribution with $\rho >$ ($<$) 0 and the strictly increasing nature of the marginal transformations (see, for example, results in Joe, 1997). Second, $|\rho_{12}| \leq |\rho|$. This inequality can be found in literature stemming from Gebelein (1941), see, for example, Koyak (1987) and references therein.

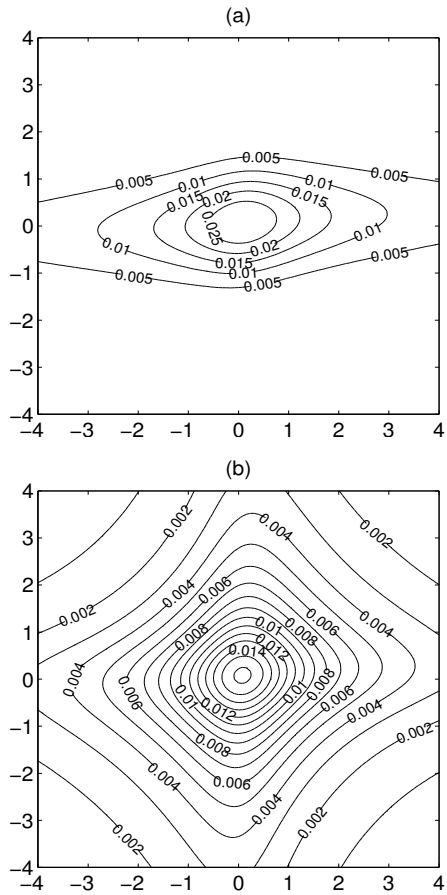

Figure 12: The bivariate sinh-arcsinh density with $\rho = 0.7$ and (a) $\delta_1 = 0.135$, $\delta_2 = 1$, (b) $\delta_1 = \delta_2 = 0.27$.

Concentrating on the symmetric marginals case as in Fig. 11, we note that: (i) the value of the correlation is $\rho = 0.7$ only at the point $\delta_1 = \delta_2 = 1$ and is lower elsewhere; (ii) the value of the correlation remains close to $\rho = 0.7$ for all $\delta_1, \delta_2 \geq 1$ i.e. lighter tails; and (iii) 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. 12(a); (iii) this last effect is reduced somewhat if both tails get heavier. The density for $\delta_1 = \delta_2 = 0.27$ is plotted in Fig. 12(b).

It may also be of interest to consider the local dependence function de-

fined 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 localised correlation coefficient by Jones (1996). Either directly, or by noting that $\gamma(x,y) = \rho/(1-\rho^2)$ for the normal distribution and that γ transforms in the same way as density functions, in our case we have

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

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 and is otherwise divorced from the influence of the other parameters is a nice feature of this transformation approach. Also, x- and y-dependence are separated out, so we consider, say, $L_{\epsilon,\delta}(x) \equiv \delta C_{\epsilon,\delta}(x)/\sqrt{1+x^2}$ only. In the symmetric case, $L_{0,\delta}(0) = \delta$ and it can readily be shown that $L_{0,\delta}$ symmetrically decreases (increases) towards zero (infinity) if $\delta <$ $(>)$ 1. In the general case, $L_{\epsilon,\delta}(0) = \delta \cosh \epsilon$, while both 'tails' of $L_{\epsilon,\delta}$ still go to zero (infinity) if $\delta <$ $(>)$ 1.

8. Options and extensions

Readers may be discomfited by some of the specific choices that have been made in this paper so far. In particular, a question that we have been asked more than once is: "why is the sinh function at the heart of this methodology rather than some other monotone function?" Second, it is clear that the normal distribution is only one of a number of possible choices for the 'central distribution' in this approach. And there is a third, perhaps less obvious, question that concerns the way in which skewness has been introduced into our model. In this section, we address each of these issues in turn.

8.1. Which transformation function?

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

$$Z = T_{\epsilon,\delta}(X_{\epsilon,\delta}) \equiv H\{\epsilon + \delta H^{-1}(X_{\epsilon,\delta})\}. \tag{8}$$

This formulation, involving both H and H^{-1}, is key to setting the normal distribution at the centre of the transformed family and allowing both heavier and lighter tails. This is most easily seen when $\epsilon = 0$: for small δ, $T(X) \sim$

$\delta h(0)H^{-1}(x)$, and for large δ, $T(X/\delta) \sim H(x/h'(0))$, division of X by δ being the 'suitable scaling' employed in Section 3.3.

Anticipating the main consideration of Section 8.2, replace the normal density as the object of transformation by a generic simple symmetric distribution with distribution function G. Apply the 'H-arcH' transformation to obtain the transformed family of distributions with distribution function

$$\mathcal{G}(x) \equiv G(H(\epsilon + \delta H^{-1}(x))). \tag{9}$$

The following result concerns the conditions required on H so that ϵ and δ act as skewness and kurtosis parameters in the sense of van Zwet (1964).

Theorem. The parameters ϵ (for fixed δ) and δ (for $\epsilon = 0$) in (9) 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 $\epsilon_i, \delta_i, i = 1, 2$. Then

$$\mathcal{G}_2^{-1}(\mathcal{G}_1(x)) = H(c + dH^{-1}(x))$$

(independently of G) where $c = (\epsilon_1 - \epsilon_2)/\delta_2$ and $d = \delta_1/\delta_2$. Then,

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

where $p(x) = dh(c + dH^{-1}(x))/h^2(H^{-1}(x)) > 0\ \forall x$. For fixed δ, i.e. $d = 1$, consider the case $c > 0$ i.e. $\epsilon_1 > \epsilon_2$; then $t_{c,1}(x) > 0$, the requirement for ϵ to act as a skewness parameter, corresponds precisely 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 certainly also sufficient that $\log h$ is increasing if $d > 1$ or that $\log h$ is decreasing if $d < 1$. ■

Another requirement that potentially further narrows the field of potential H's is unimodality of all members of the resulting family of distributions. We are keen on this since we believe that we are in the business of providing 'component' unimodal distributions which can be combined, interpretably, by mixture modelling if multimodality is present in one's data. Unfortunately, unimodality 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

Section 3.4). That said, it reinforces the requirement that $h(x) > 0\ \forall x$ else, if $h(x_0) = 0$ for some x_0, the density associated with distribution (9) 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 any viable alternative to $H(x) = \sinh(x)$.

8.2. Which central density?

The normal distribution is, of course, but one particular choice for the 'central' simple symmetric distribution mentioned in Section 8.1; let this distribution have density g. Then the transformed family of distributins has densities of the form

$$g_{\epsilon,\delta}(x) = \frac{\delta C_{\epsilon,\delta}(x)}{\sqrt{1+x^2}} g\left\{S_{\epsilon,\delta}(x)\right\}. \tag{10}$$

Several of the properties developed for the normal distribution hold immediately for other g too: examples include its distribution and quantile function (in terms of G and G^{-1}), skewness and kurtosis ordering properties, etc; some properties need to be investigated on a case-by-case basis. A sufficient condition for unimodality is that

$$1 + x(1+x^2)(\log g)'(x) + (1+x^2)^2(\log g)''(x) < 0\ \forall x$$

which has been satisfied for all the g we have considered.

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 Section 5.

A second consideration might be the tailweight properties of g-based families. For small δ, and ignoring all constants,

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

for example, simple exponential tails like those of the logistic lead to 'Weibull-type' tails, $|x|^{\delta-1}\exp(-|x|^\delta)$, while power tails, of the form $g(|x|) \sim |x|^{-(\alpha+1)}$, $\alpha > 0$, lead to 't-type' power tails for $g_{\epsilon,\delta}$ of the limiting form $|x|^{-(\nu+1)}$

where $\nu = \alpha\delta$. The Cauchy distribution as g leads to the particularly simple expression

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

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, $\epsilon = 0$, the Cauchy-based family tends to the hyperbolic secant density $\{\pi \cosh(x)\}^{-1}$, which is both 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.

8.3. Which method of introducing skewness?

Formulae (3) and (4) suggest an alternative method of introducing skewness into the symmetric sinh-arcsinh transformation. Instead of $S_{\epsilon,\delta}(X)$ as defined there, consider

$$S_{\delta,\gamma}(X) \equiv \frac{1}{2}\left\{\exp(\delta \sinh^{-1}(X)) - \exp(-\gamma \sinh^{-1}(X))\right\}, \tag{11}$$

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. Note that (the same) symmetric cases now arise from setting $\gamma = \delta$. In fact, δ now 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. This can be contrasted with the way in which skewness in (2) is introduced and controlled by differential scaling of tails. It is clear that $f_{\gamma,\delta}(x) = f_{\delta,\gamma}(-x)$.

Many properties of these skew sinh-arcsinh distributions can also be determined although the family is a little less tractable than is that based on $S_{\epsilon,\delta}$. Briefly, the distribution function and quantile functions associated with (11) — in the normal case — are $F_{\delta,\gamma}(x) = \Phi(S_{\delta,\gamma}(x))$ and $Q_{\delta,\gamma}(u) = S^{-1}_{\delta,\gamma}(\Phi^{-1}(u))$; the latter is not explicitly invertible in general although its inverse is easy to compute. A nice property of the family based on (11) is that its median is always zero. We have much numerical evidence that $f_{\gamma,\delta}$ remains unimodal for all values of $\delta, \gamma > 0$, but have been unable to prove it. Plots of the Bowley skewness (not shown) indicate that the entire range of Bowley skewness values, from -1 to $+1$, is achieved within this family.

Tests of normality and symmetry can, of course, be based on fitting $f_{\gamma,\delta}$ in the same way as they were in Sections 5 and 6 for $f_{\epsilon,\delta}$. We repeated all the simulations reported there for the alternative skewness family too. Much the most striking feature of the results is their extreme similarity; in almost all cases, results like those shown in Figs 5 to 10 provide excellent approximations to the equivalent results for the alternative family. In terms of testing normality, and from the viewpoint of the alternative family, we might claim a slightly better holding of size for $n \geq 100$ but slightly worse for $n \leq 50$, with slightly lower power for heavy-tailed distributions and slightly higher power for lighter-tailed distributions. But emphasis here is on the word 'slightly'. Something similar was observed in the context of testing symmetry save for one exceptional case: the test based on $f_{\gamma,\delta}$ was rather less able to maintain size than was the test based on $f_{\epsilon,\delta}$ for the (heavy-tailed) t_2 distribution. (A nominal 5% test exhibited significance levels around 10% for the LRT based on $f_{\epsilon,\delta}$, rising to 20% and more for the test based on $f_{\gamma,\delta}$.)

All told, however, 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 would like to 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 currently very sparsely populated. On the one hand, many if not most families of distributions on $\mathbb{R}$ concentrate on providing 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; see, for example, Jones and Faddy (2003) and Azzalini and Genton (2008). On the other hand, few families of distributions on $\mathbb{R}$ have much in the way of light-tailed membership. An exception is the exponential power distributions (Box and Tiao, 1973, Tadikamalla, 1980) and their natural two-piece skew counterparts. The new distributions fill something of a gap between these two sorts of distributions. Like skew-t distributions, they allow tails considerably heavier than the normal, although their tails are not quite as heavy as the t's power tails can be, but unlike skew-t distributions they allow lighter than normal tails also. Like exponen-

tial power distributions, the new distributions allow much lighter tails than normal (though not as light as the uniform limit of the exponential power) and heavier tails than the normal, but in the latter case escape the purely exponential nature of the exponential power tails. We reiterate that the sinh-arcsinh distributions achieve these properties in a manner something like an amalgamation of Johnson S_U and sinh-normal distributions. Indeed, the sinh-arcsinh distribution can be seen as a generalised 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 also especially appealing, in our view, to have such a family of distributions 'centred' on the normal distribution in order, as exemplified in Section 5, to allow standard likelihood ratio testing for normality against skew and light- and heavy-tailed distributions within the sinh-arcsinh family. This is in contrast to 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 LRTs within this class.

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 and Rigby, 2007).

Acknowledgements

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, and to various colleagues for the enthusiasm with which they have greeted versions of these ideas. The work of the second author was partially supported by project MTM2007–61470 of the Spanish Ministry of Science and Education.

References

Anderson, T.W. and 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**, 171–178.

Azzalini, A. and Genton, M.G. (2008) Robust likelihood methods based on the skew-t and related distributions. *Internat. Statist. Rev.*, **76**, 106–129.

Baghdachi, J. and Balakrishnan, N., eds. (2008) *Handbook of the Logistic Distribution*, Second Edition, to appear.

Banks, C.J. (2006) Sea ice thickness and iceberg distribution in the Southern Ocean. Ph.D. thesis, Department of Earth Sciences, The Open University, Milton Keynes.

Barndorff-Nielsen, O.E. (1978) Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.*, **5**, 151–157.

Boos, D.D. (1982) A test for asymmetry associated with the Hodges-Lehmann estimator. *J. Am. Statist. Assoc.*, **77**, 647–651.

Bowley, A.L. (1937) *Elements of Statistics*, Sixth Edition. New York: Scribner.

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

Cabilio, P. and Masaro, J. (1996) A simple test of symmetry about an unknown median. *Can. J. Statist.*, **24**, 349–361.

D'Agostino, R.B. (1971) An omnibus test of normality for moderate and large sample size. *Biometrika*, **58**, 341–348.

D'Agostino, R.B. (1972) Small sample probability points for the D test of normality. *Biometrika*, **59**, 219–221.

Downton, F. (1966) Linear estimates with polynomial coefficients. *Biometrika*, **53**, 129–141.

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

Fernandez, C. and Steel, M.F.J. (1998) On Bayesian modeling of fat tails and skewness. *J. Amer. Statist. Assoc.*, **93**, 359–371.

Gebelein, H. (1941) Das statistische Problem der Korrelation als Variations – und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.*, **21**, 364–379.

Genton, M.G., ed. (2004) *Skew-Elliptical Distributions and Their Applications; a Journey Beyond Normality.* Boca Raton: Chapman and Hall/CRC.

Gradshteyn, I.S. and Ryzhik, I.M. (1994) *Table of Integrals, Series, and Products*, Fifth Edition. San Diego: Academic Press.

Groeneveld, R.A. and Meeden, G.L. (1984) Measuring skewness and kurtosis. *Statistician*, **33**, 391–399.

Holland, P.W. and Wang, Y.J. (1987) Dependence function for continuous bivariate densities. *Commun. Statist. Theory Meth.*, **16**, 863–876.

Jarque, C. and Bera, A. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economet. Lett.*, **6**, 255–259.

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. and Balakrishnan, N. (1994) *Continuous Univariate Distributions*, Volume 1, Second Edition. 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**, 546–547.

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

Koyak, R.A. (1987) On measuring internal dependence in a set of random variables. *Ann. Statist.*, **15**, 1215–1228.

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

Nelder, J.A. and Mead, R. (1965) A simplex-method for function minimization. *Comput. J.*, **7**, 308–313.

Rieck, J.R. and Nedelman, J.R. (1991) A log-linear model for the Birnbaum–Saunders distribution. *Technometrics*, **33**, 51–60.

Royston, P. (1965) AS R94 A remark on algorithm AS 181: the W-test for normality. *Appl. Statist.*, **44**, 547–551.

Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes; Stochastic Models with Infinite Variance.* Boca Raton: Chapman and Hall/CRC.

Shapiro, S. and Wilk, M. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.

Stasinopoulos, D.M. and Rigby, R.A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *J. Statist. Software*, **23**, Issue 7.

Tadikamalla, P.R. (1980) Random sampling from the exponential power distribution. *J. Am. Statist. Assoc.*, **75**, 683–686.

Thadewald, T. and 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. and Selcuk, A.S. (2001) Non-normal regression II: symmetric distributions. *Commun. Statist. Theory Meth.*, **30**, 1021–1045.

Zhang, J. and Wu, Y. (2005) Likelihood-ratio tests for normality. *Comput. Statist. Data Anal.*, **49**, 709–721.

van Zwet, W.R. (1964) *Convex Transformations of Random Variables.* Amsterdam: Mathematisch Centrum.