Statistica Sinica **24** (2014), 749-771
doi:http://dx.doi.org/10.5705/ss.2011.304

GENERATING DISTRIBUTIONS BY TRANSFORMATION OF SCALE

M. C. Jones

The Open University

Abstract: This paper investigates the surprisingly wide and practicable class of continuous distributions that have densities of the form $2g\{t(x)\}$ where g is the density of a symmetric distribution and t is a suitable invertible transformation of scale function which introduces skewness. Note the simplicity of the normalising constant and its lack of dependence on the transformation function. It turns out that the key requirement is that $\Pi = t^{-1}$ satisfies $\Pi(y)-\Pi(-y) = y$ for all y; Π thus belongs to a class of functions that includes first iterated symmetric distribution functions but is also much wider than that. Transformation of scale distributions have a link with 'skew-g' densities of the form $2\pi(x)g(x)$, where $\pi = \Pi'$ is a skewing function, by using Π to transform random variables. A particular case of the general construction is the Cauchy-Schlömilch transformation recently introduced into statistics by Baker (2008); another is the long extant family of 'two-piece' distributions. Transformation of scale distributions have a number of further attractive tractabilities, modality properties, explicit density-based asymmetry functions, a beautiful Khintchine-type theorem and invariant entropy being chief amongst them. Inferential questions are considered briefly.

Key words and phrases: Asymmetry function, Cauchy-Schlömilch transformation, invariant entropy, iterated distribution function, Khintchine theorem, normalising constant, self-inverse function, skew distributions, skewing function, two-piece distributions.

1. Introduction

The setting for this paper is that of parametric modelling using continuous distributions with several, but not too many, parameters. Practical reasons for interest in such distributions are the introduction of flexibility tempered by parsimony, sometimes for understanding specific aspects of distributional shape such as skewness or tail weight, but mostly to achieve robustness to these aspects of distributional shape in estimating location and perhaps scale via statistical modelling; see, for example, Azzalini and Genton (2008). I will concentrate in this paper on univariate continuous distributions. This is nothing like as limiting as it sounds since they often form the marginal and/or conditional random components of more complex models acting, for simple example, as response distributions in regression, or having major roles in graphical models, hierarchical models, etc. Bayesian priors in multiparameter situations also often comprise well chosen combinations of univariate marginals and conditionals. Some comments on the multivariate case can nonetheless be found in Section 6.3.

A natural approach to providing wide families of univariate distributions with familiar special cases yet asymmetric and possibly otherwise altered shape characteristics is as follows. Start from a relatively simple symmetric 'base' distribution, such as the normal, and generate further distributions from it by, for example, judicious use of an appropriate transformation function. To this end, let g be a continuous univariate density function on support $\mathcal{S}_g \ni 0$ that is symmetric about zero, and let $t : \mathcal{S}_f \to \mathcal{D}$, where $\mathcal{D} \supseteq \mathcal{S}_g$, be a monotonically increasing transformation function. This paper is concerned with particular families of distributions derived from these components by what I call 'transformation of scale'. The distributions of interest have densities of the form

$$f(x) = 2g\{t(x)\}, \qquad x \in \mathcal{S}_f, \tag{1.1}$$

for appropriate t, chosen such that their inverses satisfy (2.1) below. The 'transformation of scale' terminology is intended to convey the notion of 'pulling g about' by transforming its point of evaluation x, as opposed to the more familiar technique of transforming the random variable associated with g. The latter results in distributions with densities of the form $t'(x)g\{t(x)\}$ which are clearly not equal to $2g\{t(x)\}$ in general.

Typically, in (1.1), g and t will each involve a shape parameter, that in g controlling some aspect of kurtosis and that in t introducing and controlling skewness. Location and scale parameters, $\mu \in \mathbb{R}$ and $\sigma > 0$ when $\mathcal{S}_f = \mathbb{R}$ or just $\sigma > 0$ when $\mathcal{S}_f = \mathbb{R}^+$, should be introduced for practical work in the usual way via $\sigma^{-1}f(\sigma^{-1}(x-\mu))$ or $\sigma^{-1}f(\sigma^{-1}x)$, but can be set equal to 0 and 1, respectively, for theoretical clarity. It is the location and/or scale parameters on which much practical modelling work is focussed, in order to understand their dependence on covariates.

The claim that the exceptionally simple form (1.1) is a density for symmetric g and a certain range of choices of t is not entirely trivial. In general, a putative density core of the form $g\{t(x)\}$ might not be integrable and, when it is, will typically require calculation of a new and likely complicated normalising constant. However, the normalising constant in (1.1) is just twice that of g! In particular, any parameters involved in t do not enter the normalising constant, simplifying their estimation. Of course, by rescaling g to $g_2(x) = 2g(2x)$ and t to $t_{1/2}(x) = t(x)/2$, I could have written $f(x) = g_2(t_{1/2}(x))$, so that the coefficient would have been 1. However, the choice of 2 makes the following development fit better with related work.

A trivial example of (1.1) is the specific scale shift $t(x) = 2x$, which is the identity function scaled to match the choice of normalising constant. (This is also the only example common to transformation of scale and transformation of random variable, since $t'(x) = 2$ in this case.) Some tractable non-trivial examples of other transformations of scale that also work in this way and influence shape by introducing skewness are:

$$t_1(x) = x - (\frac{b}{x}), \qquad b > 0, \quad x \in \mathbb{R}^+ \tag{1.2}$$

(Baker (2008));

$$t_2(x) = \frac{1}{a}\log(e^{ax} - 1), \qquad a > 0, \quad x \in \mathbb{R}^+ \tag{1.3}$$

(Jones (2010));

$$t_3(x) = \left(2\sqrt{cx} - c\right) I\left(0 < x < c\right) + xI\left(x \geq c\right), \quad c > 0,\ x \in \mathbb{R}^+; \tag{1.4}$$

$$t_4(x) = x\left(2-dx\right) I\left(dx<0\right) + \left\{x - \frac{1}{2d}(1-\sqrt{1+4dx})\right\} I\left(dx \geq 0\right),\ d \in \mathbb{R}\backslash\{0\}, \tag{1.5}$$

with $\lim_{d\to 0} t_4(x) = 2x$ for all $x \in \mathbb{R}$;

$$t_5(x) = 2x\left\{\frac{1}{1-\alpha} I\left(x < 0\right) + \frac{1}{1+\alpha} I\left(x \geq 0\right)\right\}, \quad -1 < \alpha < 1, \quad x \in \mathbb{R}, \tag{1.6}$$

leading to the celebrated two-piece density (Fechner (1897), Fernández and Steel (1998)) in essentially the parameterisation preferred, for example, in Mudholkar and Hutson (2000), which the reader might also see for further references in the case of normal g; and

$$t_6(x) = \frac{1}{1-\alpha^2}\left\{2x - \alpha\ \sqrt{4x^2 + 1 - \alpha^2}\right\}, \quad -1 < \alpha < 1, \quad x \in \mathbb{R}. \tag{1.7}$$

That all these transformations of scale should work with no change to the normalising constant seems, at first glance, little short of astonishing. In particular, that $2g\{t_1(x)\}$, $x \in \mathbb{R}^+$, is a valid density is a consequence, as observed by Baker (2008), of the Schlömilch, or Cauchy-Schlömilch, transformation used in the evaluation of certain definite integrals (Boros and Moll (2004, Sec. 13.2)). My primary purpose here is to identify a wide class of functions t, through their inverses $\Pi = t^{-1}$, that afford densities of form (1.1); see Proposition 1 of Section 2. This yields a general approach to generating appropriate transformations of scale, all of which are readily shown necessarily to introduce asymmetry.

There turns out to be a close link between the proposed construction and an alternative, popular, methodology for generating families of distributions from g, namely the 'skew-g' distributions arising from the seminal work of Azzalini (1985). This is spelt out in Section 3.1. Simple and tractable versions of Π and hence t are explored in Sections 3.2 to 3.5. Transformation of scale distributions have several other attractive and tractable properties which are explored in Section 4. These characteristics include their modality properties, tailweight control, skewness properties as understood through density-based asymmetry functions, simple random variate generation, an especially attractive extended Khintchine theorem, and invariance of entropy. Inferential questions are considered briefly in Section 5. A discussion of the wider place of transformation of scale distributions completes the paper in Section 6. This concentrates principally on the univariate case but also briefly visits the cases of circular and multivariate distributions.

2. The Main Result

Proposition 1. *Let $\Pi : \mathcal{D} \to \mathcal{S}_f$ be a piecewise differentiable monotone increasing function with inverse t, where $\mathcal{D} \supseteq \mathcal{S}_g \ni 0$. Suppose that*

$$\Pi(y) - \Pi(-y) = y, \quad \text{for all } y \in \mathcal{D}. \tag{2.1}$$

Then if $g(x)$ is a density on $\mathcal{S}_g$ that is symmetric about zero, $f(x) = 2g\{t(x)\} = 2g\{\Pi^{-1}(x)\}$ is a density on $\mathcal{S}_f$.

Proof. Nonnegativity of f follows from that of g. It remains to show that $\int_{\mathcal{S}_f} f(x)dx = 1$. This is done here for differentiable Π for simplicity, with obvious extension to the piecewise differentiable case. Making the substitution $y = t(x)$ and writing $\pi(y) = \Pi'(y)$, which is positive for all $y \in \mathcal{D}$,

$$\int_{\mathcal{S}_f} f(x)dx = 2\int_{\mathcal{S}_f} g\{t(x)\}dx = 2\int_{\mathcal{D}} g(y)\pi(y)dy = 2\int_{\mathcal{S}_g} g(y)\pi(y)dy.$$

That $2\pi(y)g(y)$ is a density and hence integrates to one is a standard result readily shown via the evenness of g and the oddness of $\pi - 1/2$ as at (3.1) to follow (Azzalini (1985), Wang, Boyer, and Genton (2004)).

Corollary 1. *The only transformation $t = \Pi^{-1}$ with Π satisfying (2.1) such that $f(x) = 2g\{t(x)\}$ is a symmetric density is $t(x) = 2x$.*

Proof. Symmetry requires $\Pi(y) = -\Pi(-y)$ for all y. Combining the latter with (2.1) yields $\Pi(y) = y/2$.

3. Relationships With Earlier Work and Special Cases

3.1. Relationship with skew-g distributions

The substitution involved in the proof of Proposition 1 is equivalent to making the transformation of random variable $X = \Pi(Y)$ where X follows the distribution with density (1.1), Y follows the distribution with density $2\pi(y)g(y)$

and $\pi(y) = \Pi'(y)$. Note that (2.1) implies that

$$\pi(y) + \pi(-y) = 1, \quad \text{for all } y \in \mathcal{D}. \tag{3.1}$$

The distribution of Y, with constraint (3.1), is precisely the general formulation for skewing symmetric distributions pursued in Wang, Boyer, and Genton (2004). Specialising from the most general case, if π is taken to be the distribution or survival function of a distribution symmetric about zero, P say, then $2P(y)g(y)$ is also a skew density function. And specialising a little more, a one-parameter extension of g which allows skewness is afforded by taking $P(y) = Q(\lambda y)$, say, where $\lambda \in \mathbb{R}$ and Q is the distribution function of a symmetric distribution with no further shape parameters. This, the density $2Q(\lambda y)g(y)$, is the celebrated formulation of Azzalini (1985) in which Q is often taken to be the distribution function G corresponding to g. An enormous literature has arisen on this kind of skew-g model; see, for example, Azzalini (2005) and Genton (2004).

It can be argued, therefore, that transformations of scale work in the sense of simplicity of normalising constant because they arise by appropriate transformation of the random variable associated with skew-g distributions, which have the same simple normalising constant. The families of distributions with densities of the forms $2g\{\Pi^{-1}(x)\}$ and $2\pi(x)g(x)$ are, however, quite different and competing/complementary; some comparison of the two is made in Section 6.1.

3.2. Special cases I: $\mathcal{S}_f = \mathbb{R}^+$

When π is a distribution function, $\Pi(y) = \int_{-\infty}^{y} \pi(w)dw$ when the definite integral exists and it is then called the first iterated distribution function (Bassan, Denuit, and Scarsini (1999)). Integration by parts shows this $\Pi(y)$ also to equal $\int_{-\infty}^{y}(y - w)\pi'(w)dw$ which, inter alia, is the numerator of the mean residual life function associated with density π'. These distribution functions yield Π with range $\mathcal{S}_f = \mathbb{R}^+$. In particular, when $\mathcal{D} = \mathbb{R}$, a sufficient condition both for existence of the definite integral and $\lim_{y\to-\infty} \Pi(y) = 0$ is that $\pi(y) = o(|y|^{-1})$ as $y \to -\infty$; this retains all distributions with finite mean, that is, all distributions except those with Cauchy-like tails and heavier. Precisely these Πs were utilised in a quite different context in Jones (2008a).

What are good choices of distribution function π in this context? Few choices of π are fully tractable in the sense of possessing explicit formulae both for its integral Π and for the inverse of that integrated function, $t = \Pi^{-1}$. For example, the natural, and popular in terms of skew-g distributions, choice of $\pi_N(y) = \Phi(\lambda y)$, the normal, $N(0, 1/\lambda^2)$, distribution function, yields $\Pi_N(y) = y\Phi(\lambda y) + \lambda^{-1}\phi(\lambda y)$, where ϕ is the standard normal density, but this is not explicitly invertible. This need not, however, be a total bar to employing $t = \Pi_N^{-1}$ in (1.1).

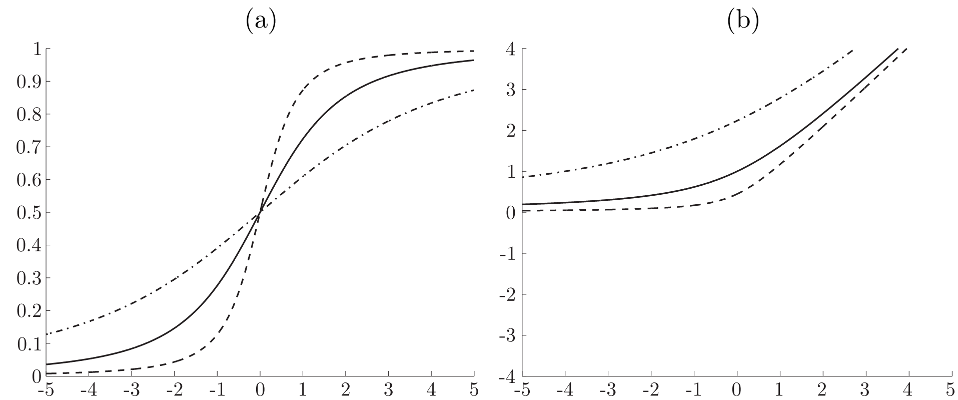

Figure 1. Examples of the functions (a) π_T and (b) Π_T belonging to the $t(2)$ distribution: $b = 0.2$ (dashed lines), $b = 1$ (solid lines), $b = 5$ (dot-dashed lines).

On the whole real line, two symmetric distributions stand out as being especially tractable in this kind of respect. These are the t distribution on two degrees of freedom, $t(2)$, and the logistic distribution; they have distribution functions, suitably scaled, given by

$$\pi_T(y) = \frac{1}{2}\left\{1 + \frac{y}{\sqrt{4b+y^2}}\right\} \quad \text{and} \quad \pi_L(y) = \frac{e^{ay}}{1+e^{ay}},$$

respectively, $a, b > 0$. I first found the tractable nature of these distribution functions to be advantageous in Jones (2004), and then even more so because of their explicit first iterated distribution functions in Jones (2008a). They star again here because those first iterated distribution functions are also explicitly invertible:

- for the $t(2)$ distribution, $\Pi_T(y) = (1/2)(y + \sqrt{4b+y^2})$ so that $\Pi_T^{-1}(x) = x - (b/x)$, that is, $\Pi_T^{-1}(x) = t_1(x)$ given at (1.2). The remarkable $t(2)$ distribution (Jones (2002a)) strikes again: the $t(2)$ distribution turns out to be at the heart of why the Cauchy-Schlömilch transformation works! π_T and Π_T are shown in Figure 1;
- for the logistic distribution, $\Pi_L(y) = a^{-1}\log(1+e^{ay})$ which leads to $\Pi_L^{-1}(x) = a^{-1}\log(e^{ax}-1)$: the logistic distribution therefore leads precisely to the alternative transformation of scale of Jones (2010), $\Pi_L^{-1}(x) = t_2(x)$ given at (1.3).

Symmetric distributions on finite support also lead to distributions on $\mathbb{R}^+$. In fact, they result in transformations of scale that are defined by separate formulae above and below some threshold value. A tractable example corresponds to the uniform distribution on $(-c, c)$ which leads to transformation of scale t_3 given at (1.4). One has to be careful with finite support π to ensure that Π is then applied only with g on support $\mathcal{S}_g \subseteq \mathcal{D}$.

Yet other examples, not shown here, can be constructed by taking π to be non-monotonic. But such πs typically introduce multimodality in a way that, to this author at least, is far less meaningful and interpretable than employing finite mixtures of unimodal components.

3.3. Relationship with self-inverse functions

The special cases of transformation of scale that result in distributions on $\mathbb{R}^+$ were also the subject, from a much less well understood standpoint, of Jones (2010). Above, I claimed that $t(x)$ in (1.1) equals $\Pi^{-1}(x)$ where $\Pi(y)-\Pi(-y) = y$ for all $y \in \mathcal{D}$ and here I constrain $\mathcal{D} = \mathbb{R}$ and Π such that $\lim_{y\to-\infty} \Pi(y) = 0$. But in Jones (2010), I claimed that $t(x)$ should have the form $x-s(x)$ where $s : \mathbb{R}^+ \to \mathbb{R}^+$ is an onto monotone decreasing function that is self-inverse i.e. $s\{s(x)\} = x$ or $s^{-1}(x) = s(x)$. Here is a demonstration that these two formulations are equivalent. Write $s(x) = x - t(x)$ for $x = \Pi(y) > 0$ and $y \in \mathcal{D}$. Then,

$$
\begin{aligned}
& s\{s(x)\} = x \\
\Leftrightarrow \quad & s(x) - t\{s(x)\} = s(x) + t(x) \\
\Leftrightarrow \quad & t(x) = -t\{x - t(x)\} \\
\Leftrightarrow \quad & t\{\Pi(y)\} = -t[\Pi(y) - t\{\Pi(y)\}] \\
\Leftrightarrow \quad & -y = t\{\Pi(y) - y\} \\
\Leftrightarrow \quad & \Pi(-y) = \Pi(y) - y.
\end{aligned}
$$

The essential equivalence between inverse first iterated distribution functions of most symmetric distributions, and some extensions thereof, and self-inverse functions would appear to be new. It means that examples of one can, of course, be used to generate examples of the other. For example, from the special cases in Section 3.2, the inverse iterated $t(2)$ distribution function leads to the self-inverse function b/x, while the inverse iterated logistic distribution function leads to the self-inverse function $-a^{-1}\log(1 - e^{-ax})$. Likewise, self-inverse functions (e.g. Jones (2010), Kucerovsky, Marchand, and Small (2005)) can be used to generate symmetric distributions. However, it seems to be easier to first think of symmetric distribution functions and to then work out their equivalent self-inverse functions than to first think of self-inverse functions and to then work back to their equivalent symmetric distribution functions. Moreover, the Π^{-1} formulation is more general than the self-inverse formulation, as the following section bears witness.

3.4. Special cases II: $\mathcal{S}_f = \mathbb{R}$

Suppose that $\mathcal{D} = \mathbb{R}$ and, for sufficiently large y, $\pi(y) > \pi(-y)$ where $\pi = \Pi'$. If for $y \to -\infty$, $\pi(y)$ tends to zero as, or more slowly than, $|y|^{-1}$ then $f(x) = 2g\{\Pi^{-1}(x)\}$ will be a valid density on the whole of $\mathbb{R}$. The precise manner in which π — which we hereafter take to be monotone — behaves for minus large y influences the weight of the left-hand tail of f which, in turn, induces positive skewness; see Sections 4.2 and 4.3. The densities $f(-x)$, $x \in \mathbb{R}$, are all available too by changing the transformation from the inverse of $\Pi(y)$ to the inverse of $-\Pi(-y) = y - \Pi(y)$. They have negative skewness and can be thought of as emanating from π being the survival function counterpart of a distribution function.

A natural distribution to consider with Cauchy-like tails is, of course, the Cauchy distribution itself. It can readily be seen that for the Cauchy distribution with signed scale parameter $d \in \mathbb{R}\backslash\{0\}$,

$$\Pi_C(y) = \frac{1}{2}\left[y + \frac{1}{\pi}\left\{2y\tan^{-1}(dy) - \frac{1}{d}\log(1 + d^2y^2)\right\}\right], \qquad y \in \mathbb{R},$$

and $\lim_{d\to 0}\Pi_C(y) = y/2$. Following Azzalini (1985), the device of allowing d to take positive and negative values within π is used to accommodate both distribution and survival function versions at once. However, Π_C^{-1} is not explicitly available.

A more tractable alternative that corresponds to a heavy-tailed density on $\mathbb{R}$ results in transformation t_4 given at (1.5). This corresponds to the very heavy-tailed distribution with density

$$\frac{d}{4\,(1 + d|y|)^{3/2}}, \qquad y \in \mathbb{R}, \tag{3.2}$$

for $d > 0$. Both distribution and survival functions for π are accommodated if $d \in \mathbb{R}\backslash\{0\}$, in which case

$$\pi_D(y) = I(dy > 0) - \frac{\mathrm{sgn}(dy)}{2\sqrt{1 + |dy|}}, \qquad y \in \mathbb{R},$$

and the corresponding Π function, which has the explicit inverse t_4, is given by

$$\Pi_D(y) = yI(dy > 0) + \frac{1}{d}\left(1 - \sqrt{1 + |dy|}\right), \qquad y \in \mathbb{R}, \tag{3.3}$$

and $\lim_{d\to 0}\Pi_D(y) = y/2$ by appropriate choice of constant of integration. π_D and Π_d are shown for some values of d in Figure 2.

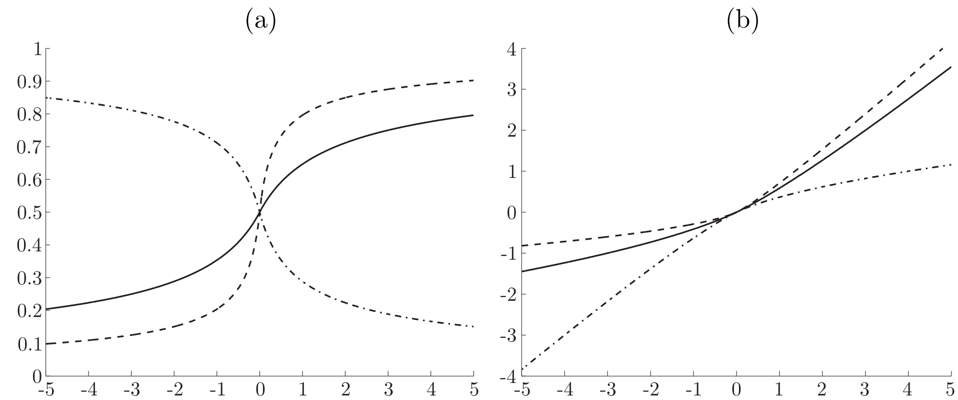

Figure 2. Examples of the functions (a) π and (b) Π_D belonging to the distribution with density (3.2): $d = 5$ (dashed lines), $d = 1$ (solid lines), $d = -2$ (dot-dashed lines).

3.5. An interesting class

The need for very heavy-tailed distributions to act as π is obviated, and all other symmetric distribution functions, R say, brought back into the reckoning, if distribution functions are replaced by monotone functions tending to nonzero constants between zero and one. The latter are conveniently parametrised as $(1-\alpha)/2$ and $(1+\alpha)/2$ for some $-1 < \alpha < 1$ as $y \to -\infty$ and ∞, respectively. That is, take

$$\pi(y) = \frac{1}{2}(1-\alpha) + \alpha R(y), \qquad y \in \mathbb{R},$$

and hence

$$\Pi_P(y) = \frac{1}{2}(1-\alpha)y + \alpha \Pi_R(y), \qquad y \in \mathbb{R}. \tag{3.4}$$

Here, any scale parameter in R is set to a fixed value and its previous role of controlling skewness is transferred to α. Notice that $\alpha = 0$ corresponds to $\Pi_P(y) = 1/2$ for all $y \in \mathbb{R}$ while $\alpha = 1$ and $\alpha = -1$, if allowed, would correspond to $\pi = R$ and $1 - R$, respectively. The sense in which α is a skewness parameter will be explored in Section 4.3.

The most extreme case of this construction in terms of choice of R is to set

$$\pi_{TP}(y) = \frac{(1-\alpha)}{2} I(y < 0) + \frac{(1+\alpha)}{2} I(y \geq 0), \qquad -1 < \alpha < 1,$$

which leads to the two-piece density, which is based on transformation t_5 given at (1.6). As implied by the references given in the Introduction, this is an especially important and well-established special case of the methodology of this paper.

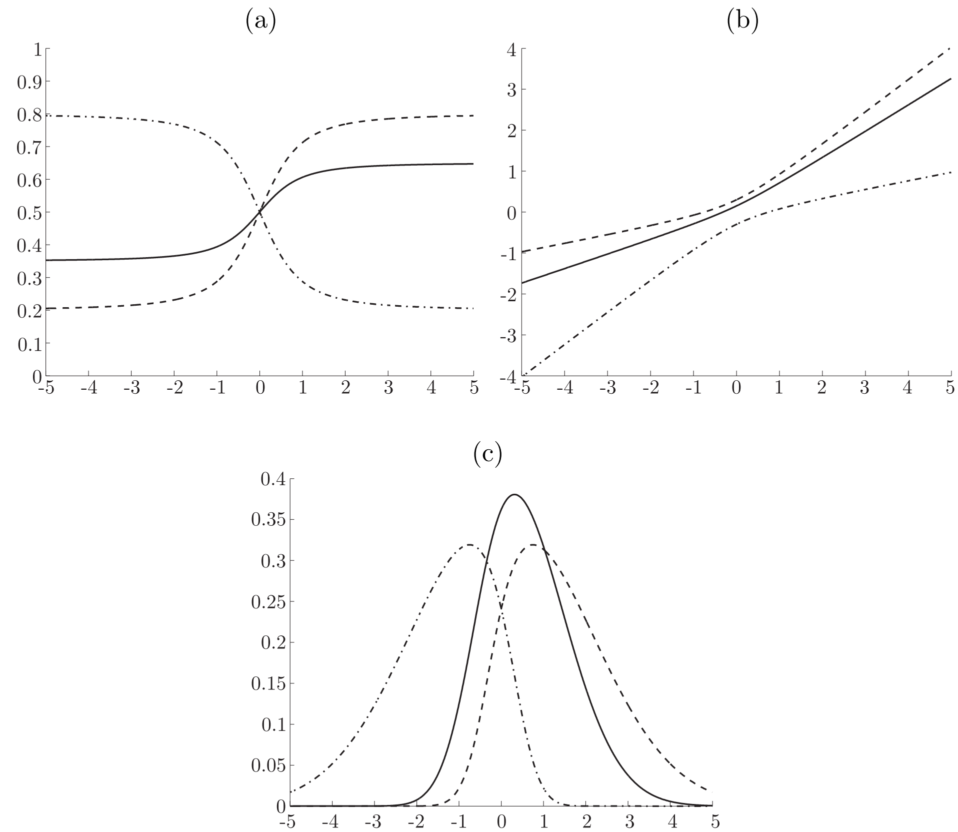

Figure 3. Examples of the functions (a) π_S, (b) Π_S (both given by (3.5)) and (c) density (3.6) when $g = \phi$: $\alpha = 0.6$ (dashed lines), $\alpha = 0.3$ (solid lines), $\alpha = -0.6$ (dot-dashed lines) .

It is the abrupt step in π_{TP} at $y = 0$ that causes the transformation of scale density derived from it to exhibit its two-piece nature; in fact, the two-piece density consists of differently scaled versions of the left- and right-hand halves of g joined in a continuous fashion at zero. Amelioration of the step in π_{TP} by utilising a smooth, non-degenerate, R leads to other transformations of scale of which a tractable one is given by t_6 at (1.7): this corresponds to setting $R(y) = \pi_T(y)$ with $b = 1/4$, that is, once more, the t distribution on 2 degrees of freedom. The corresponding π and Π functions, called π_S and Π_S, respectively, are graphed in Figure 3(a),(b) for certain choices of α. Their formulae are simply

$$\pi_S(y) = \frac{1}{2}\Big(1 + \frac{\alpha y}{\sqrt{1+y^2}}\Big), \quad \Pi_S(y) = \frac{1}{2}\left(y + \alpha\,\sqrt{1+y^2}\right), \quad y \in \mathbb{R}. \qquad (3.5)$$

The function Π_S and its inverse given at (1.7) are closely related to the "sinh-arcsinh transformation" introduced in the context of transforming random variables in Jones and Pewsey (2009). Explicitly, in the transformation of scale case, if one rescales the distribution $f_S(x) = 2g\{t_6(x)\}$ by the factor $2/\sqrt{1-\alpha^2}$, the rescaled density is

$$\sqrt{1-\alpha^2}\; g\left\{\frac{1}{\sqrt{1-\alpha^2}}(x-\alpha\sqrt{1+x^2})\right\} = \frac{1}{\cosh\epsilon}\; g\left[\sinh\{\sinh^{-1}(x)-\epsilon\}\right], \quad (3.6)$$

where $\epsilon = \tanh^{-1}\alpha$.

Alternatively, use of general b in π_T used as R within Π_P leads, for suitably small $b > 0$, to explicit smoothed approximations to the two-piece density.

The distributions of this subsection potentially have the greatest practical relevance on $\mathbb{R}$: densities (3.6) are graphed when g is the standard normal density, ϕ, in Figure 3(c).

4. Attractive Properties of Transformation of Scale

I have already stressed the simplicity of the normalising constant in (1.1) relative to the difficulty of obtaining the normalising constant in models of the form $g\{t(x)\}$ in general, and the fact that the normalising constant does not depend on any parameters involved in $t = \Pi^{-1}$. Model (1.1) has other attractive properties, the most striking of which are described here; see Jones (2010) for more when $\mathcal{S}_f = \mathbb{R}^+$.

4.1. Modality

As a 'pulling about' of the density $g_2(x) = 2g(2x)$, the density at (1.1) takes precisely the same values as g_2 in precisely the same order, just occurring at a different 'rate' (determined by $t = \Pi^{-1}$). Immediately, then, the density at (1.1) has the same modality as g_2 and hence g and, in particular, unimodal g leads to unimodal f. Moreover, since the mode of unimodal g is at 0, the mode of unimodal f is explicitly given by $x_0 = \Pi(0)$. For example, the modes of unimodal distributions using transformations of scale $t_1, \ldots, t_6$ are $\sqrt{b}$, $(\log 2)/a$, $c/4$, $1/d$, 0 and $\alpha/2$, respectively. Few families of distributions outside the transformation of scale class have such immediate and explicit modality properties.

4.2. Tails

For clarity in this subsection, consider tail behaviour in the case that $\pi(y) > \pi(-y)$ for sufficiently large y, that is, for transformations related to distribution functions rather than survival functions. Of course, tails are switched in the other case when $t(-x)$ replaces $t(x)$.

First, consider the cases corresponding to $\lim_{y\to-\infty}\pi(y) = 0$. Since then $\Pi^{-1}(x) \sim x$ as $x \to \infty$, $f(x) = 2g(\Pi^{-1}(x))$ retains the same right-hand tail behaviour as $g(x)$. These transformations of scale otherwise work by lightening the left-hand tail of f relative to that of g, to the point of restricting the range of f to positive values of x, and then to controlling its behaviour at and near zero, as described below.

When $\mathcal{S}_f = \mathbb{R}^+$, $f(0) = 0$. The precise order of contact of f to the horizontal axis as $x \to 0$ is controlled by the left-hand tail behaviour of π: the heavier its tail, up to but not including the $O(|y|^{-1})$ restriction mentioned in Section 3.2, the lighter the contact of f. For example, for π functions with $O(|y|^{-\tau})$ left-hand tail, $\tau > 1$, $f(x) \sim 2g(x^{-1/(\tau-1)})$ as $x \to 0$.

When $\mathcal{S}_f = \mathbb{R}$:

- π functions with $O(|y|^{-1})$ left-hand tail make $\Pi(|y|) \sim -\log|y|$ as $y \to -\infty$, so that the left-hand tail of $f(x) = 2g(\Pi^{-1}(x))$ decreases as $g(e^{-x})$, which is very much lighter than the original left-hand tail of g;
- π functions with $O(|y|^{-\tau})$ left-hand tail, $0 < \tau < 1$, make the left-hand tail of f decrease as $g\{|x|^{1/(1-\tau)}\}$ which is also lighter than the original left-hand tail of g but not so light as $g(e^{-x})$. The closer is τ to 1, the lighter the tail of f.
- heavier left-hand tails of f arise when $\lim_{y\to-\infty} \pi(y) = (1-\alpha)/2$ is a nonzero constant as in Section 3.5. In fact, π functions with constant limits retain the orders of magnitude of both of g's tails. The tails of f still differ from each other when $\alpha \neq 0$, but by being rescaled by different scaling factors in the ratio $(1-\alpha) : (1+\alpha)$. In this respect, these transformation of scale distributions all behave like their two-piece special case.

The left-hand tail behaviour described above is illustrated in the case that g has simple exponential tails, $g(x) \sim e^{-|x|}$ as $x \to \pm\infty$, as, for example, in the Laplace and logistic distributions, in Figure 4. There, the cases of a non-zero limit of π and of $\tau = 1/2, 2$ and 5 are illustrated, corresponding to tails of f going as 2 times $\exp(-|x|)$, $\exp(-x^2)$, $\exp(-1/x)$ and $\exp(-1/x^{1/4})$, respectively.

The differences between the left- and right-hand tails of f are responsible for its skewness which is investigated further in the next subsection.

4.3. Skewness

Unimodal transformation of scale distributions lend themselves admirably to investigation via the density-based asymmetry function of Avérous, Fougères, and Meste (1996), Boshnakov (2007), Critchley and Jones (2008), and O'Hagan (1994, Sec. 2.6). This asymmetry function arises by comparing, in a suitably scaled way, the distances from the mode of a unimodal distribution to the points to its left and to its right at which the density function has the same level. See Figure 1 of Critchley and Jones (2008) for graphical illustration. Write $x_L(p)$ and

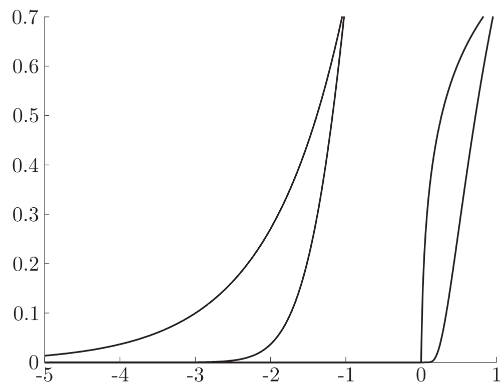

Figure 4. Left-hand tails of transformation of scale densities based on g with simple exponential tails and, from the left, a non-zero limit of π and of $\tau = 1/2, 5$ and 2, respectively.

$x_R(p)$ for the left- and right-hand solutions of $f(x) = pf(x_0)$, where x_0 is the mode of unimodal f. Then, in general, the density-based asymmetry function is given by

$$\gamma(p) \equiv \frac{x_R(p) - 2x_0 + x_L(p)}{x_R(p) - x_L(p)}, \qquad 0 < p < 1.$$

Now write $g_R(y) = g(y)I(y > 0)$ for the 'right-hand half' of g, and take

$$c_g(p) = g_R^{-1}\{pg(0)\} > 0.$$

This function is typically explicitly available for simple symmetric g; for example, when $g = \phi$, the standard normal density, $c_g(p) = \phi_R^{-1}(p/\sqrt{2\pi}) = \sqrt{2(-\log p)}$. Then, it can easily be seen that, as well as $x_0 = \Pi(0)$, we have $x_R(p) = \Pi\{c_g(p)\}$ and $x_L(p) = \Pi\{-c_g(p)\} = \Pi\{c_g(p)\} - c_g(p)$ so that, for transformation of scale distributions,

$$\gamma(p) = \frac{2}{c_g(p)}[\Pi\{c_g(p)\} - \Pi(0)] - 1, \qquad 0 < p < 1. \tag{4.1}$$

Since π and $c_g(p)$ are both typically explicitly available, transformation of scale distributions afford explicit mathematical formulae for $\gamma(p)$ through (4.1). Such tractability of density-based asymmetry functions has previously been singularly lacking in the literature. Indeed, it was the quest for distributions with tractable density-based asymmetry functions that initially inspired the current work.

Now concentrate on increasing π. Then, $\gamma(p) > 0$ for all $0 < p < 1$ because $\Pi(y) - 2\Pi(0) + \Pi(-y)$ is an increasing function of y for all $y > 0$, and hence is positive for all $y > 0$. That is, transformation of scale based on increasing π induces positive skew in f in this quite strong sense.

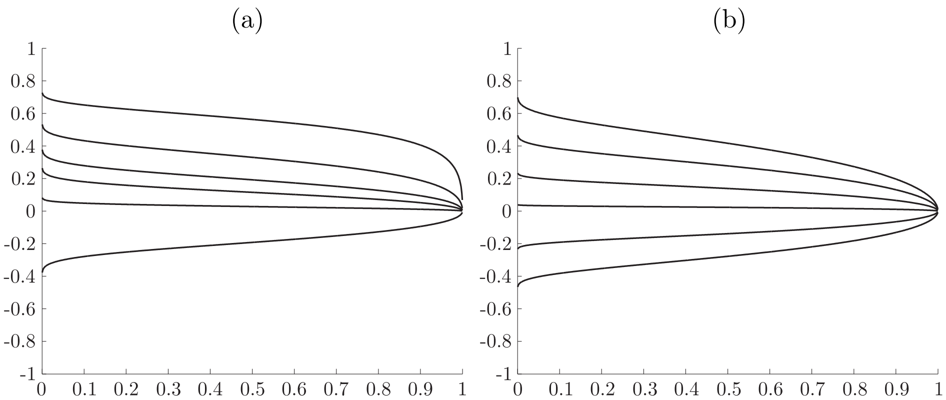

Figure 5. Asymmetry functions $\gamma(p)$ when (a) $\Pi = \Pi_D$ given by (3.3) and, in order of increasing amount of skewness, $d = -1, 0.1, 0.5, 1, 2.5, 10$, and (b) $\Pi = \Pi_S$ given by (3.5) and, in order of increasing amount of skewness, $\alpha = -0.6, -0.3, 0.05, 0.3, 0.6, 0.9$.

Next, let π involve a scale parameter $\lambda > 0$ so that π is written $\pi(\lambda x)$ as in Azzalini (1985) and Π becomes $\lambda^{-1}\Pi(\lambda x)$. Then,

$$\frac{\partial\gamma(p)}{\partial\lambda} = \frac{2}{\lambda^2 c_g(p)} \left[\lambda c_g(p)\pi\{\lambda c_g(p)\} + \Pi(0) - \Pi\{\lambda c_g(p)\}\right].$$

The quantity in square brackets is an increasing function of $\lambda c_g(p)$ and, since its value at 0 is 0, is consequently positive. Therefore, density-based asymmetry increases with increasing $\lambda > 0$. This applies to transformations $t_1, \ldots, t_4$ when $\lambda = 1/2b^{1/2}$, a, $1/c$, and d, respectively. See Jones (2010, Figures 4 and 5) for graphs of asymmetry functions when Π corresponds to transformations t_1 and t_2, and Figure 5(a) in this paper when $\Pi = \Pi_D$ given at (3.3) so that $\Pi_D^{-1} = t_4$.

On the other hand, the two-piece distribution associated with t_5 has the special property of constant density-based asymmetry (Boshnakov (2007), Critchley and Jones (2008)); in fact, $\gamma(p) = \alpha$ for all p. Moreover, the asymmetry function associated with all transformations of scale of the form (3.4) is

$$\gamma(p) = \alpha\left(\frac{2}{c_g(p)}\left[\Pi_R\{c_g(p)\} - \Pi_R(0)\right] - 1\right),$$

which is simply α times the asymmetry function associated with Π_R. This is a very explicit sense in which α acts as a skewness parameter in densities (1.1) based on transformations (3.4). The example of the asymmetry function associated with Π_S given by (3.5) is shown in Figure 5(b).

Both frames of Figure 5 show asymmetry functions that are fairly 'flat' as p varies, implying skewness that arises from a consistent 'widening' of the densities to the right at all heights, as opposed to skewness induced, for example, by a heavier tail to the right than to the left.

The scalar skewness measure most naturally associated with density-based asymmetry (Critchley and Jones (2008)) is that of Arnold and Groeneveld (1995): $\gamma = 1 - 2F(x_0)$ where F is the distribution function associated with unimodal f and x_0 is again the mode. In fact, because

$$\gamma = f(x_0) \int_0^1 \gamma(p)\{x_R(p) - x_L(p)\}dp \propto \int_0^1 \gamma(p) c_g(p) dp$$

and, in the case of transformation of scale (3.4), $\gamma(p)$ is directly proportional to α, then so is γ. We also readily find that, in general,

$$F(x_0) = 1 - 2\int_0^\infty \pi(y)g(y)dy \quad \text{so that} \quad \gamma = 4\int_0^\infty \pi(y)g(y)dy - 1.$$

Note that this is not the same as the Arnold–Groeneveld measure associated with skew-g density $2\pi(y)g(y)$.

It is also the case that, in the classical van Zwet (1964) sense, density (1.1) is more positively skewed than the density $2\pi(y)g(y)$ whenever π is increasing because the transformation between the two, $X = \Pi(Y)$, is convex.

On $\mathbb{R}$, the most positively skewed version of f when π is a scaled distribution function is the half-g density $2g(y)I(y > 0)$, but when Π is of the form (3.4) it is $2g\{\Pi_R^{-1}(y)\}$.

More on the skewness properties of transformation of scale distributions may be found in Fujisawa and Abe (2012).

4.4. Random variate generation

If a random variable A from the underlying symmetric distribution with density g is available, then a random variable $X \sim 2g\{\Pi^{-1}(x)\}$ can be obtained from A with the help of an independent uniform (0,1) random variable, U. In fact, simply take

$$X = \Pi(A) - A\; I\{U > \pi(A)\}. \tag{4.2}$$

This is because $Y \sim 2\pi(y)g(y)$ can be obtained as $A\; I\{U \le \pi(A)\} - A\; I\{U > \pi(A)\}$ (Wang, Boyer, and Genton (2004, p.1262)). $X = \Pi(Y)$ and $\Pi(-y) = \Pi(y) - y$.

4.5. An extended Khintchine theorem

On $\mathbb{R}$, $X \sim g$ for unimodal g, with mode at zero, if and only if X has a representation as a uniform scale mixture of the following form: $X \sim UZ$ where $U \sim U(0,1)$ and $Z \sim f_Z$ where $f_Z(z) = -zg'(z)$, $z \in \mathbb{R}$, and U and Z

are independent. This is Khintchine's theorem (Dharmadhikari and Joag-Dev (1988, Sec. 1.2), Feller (1971), Jones (2002b), Khintchine (1938), Shepp (1962)). Mudholkar and Wang (2007) proved a Khintchine-type theorem for unimodal R-symmetric densities; R-symmetric densities are densities, f_R, of R, say, on support $\mathbb{R}^+$ such that, for some $\theta > 0$, $f_R(r) = f_R(\theta/r)$ for all $r > 0$. See also Chaubey, Mudholkar, and Jones (2010), where R-symmetric distributions were identified with Cauchy-Schlömilch transformation of scale distributions, that is, using (1.2). Their development can be extended, as is done in the web-appendix, to prove the following simple and elegant extension of the Khintchine theorem for general unimodal distributions of type (1.1).

Proposition 2. *$X \sim f$ where f is of form* (1.1) *and is unimodal if and only if*

$$X \sim UZ + \Pi(Z) - Z \tag{4.3}$$

where $U \sim U(0,1)$, $Z \sim -zg'(z)$, $z \in \mathbb{R}$, and U and Z are independent.

Note that this theorem reduces immediately to the usual Khintchine theorem for symmetric unimodal g when $\Pi(Z) = Z$ and, after a little more work, to the result in Chaubey, Mudholkar, and Jones (2010) when $\Pi(Z) = \Pi_T(Z)$.

Formula (4.3) provides an alternative means of random variate generation from f to that in (4.2) when f is unimodal.

4.6. Entropy

Transformation of scale densities stay close to their symmetric density roots in the sense that certain properties of f are inherited directly from g. One such is modality (Section 4.1). Another is entropy, a recent statistical overview of which is provided in Ebrahimi, Soofi, and Soyer (2010). The Shannon entropy, $S(f)$, of a continuous density, f, is given by $-\int_{S_f} f(x)\log\{f(x)\}dx$; the Rényi entropy is $R_\rho(f) = (1-\rho)^{-1}\log\int_{S_f} f^\rho(x)dx$, $\rho > 0$, $\rho \neq 1$.

Proposition 3. *Let g_2 be the rescaled version $g_2(x) = 2g(2x)$ of $g(x)$, and let E denote either the Shannon entropy S or the Rényi entropy R_ρ. Then*

$$E(f) = E(g) - \log 2 = E(g_2).$$

The proof of Proposition 3 for the Shannon entropy can be found in the web-appendix, while that for the Rényi entropy is completely omitted.

5. On Maximum Likelihood Estimation

Let $X_1, \ldots, X_n \in \mathbb{R}$ be a random sample assumed to come from the location-scale version of (1.1) with support $\mathbb{R}$, namely the distribution with density

$$2\sigma^{-1}g\left(\Pi_\lambda^{-1}(\sigma^{-1}(x-\mu));\nu\right) \tag{5.1}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and, for concreteness, ν is a shape parameter associated with g and λ is a skewness parameter that forms part of $\Pi = \Pi_\lambda$. (When the data are nonnegative, μ can usually be left out.)

5.1. Expected information matrix

Generically, write $\iota_{\theta_1\theta_2}$ for the element of the expected information matrix associated with θ_1 and θ_2, divided by n. In this section, a prime denotes differentiation with respect to x, and superscripts ν and λ denote differentiation with respect to each of those parameters. For example, $(\log g)'^{\lambda}(x)$ denotes $\partial^2(\log g(x))/\partial x \partial\lambda$. By way of shorthand, the arguments of integrands will be omitted. Also, hats over parameters will denote their maximum likelihood estimators.

Then, it turns out that, after a certain amount of manipulation,

$$
\begin{aligned}
\iota_{\mu\mu} &= \frac{2}{\sigma^2}\left\{\int \frac{\Pi''}{(\Pi')^2} g' - \int \frac{1}{\Pi'}(\log g)'' g\right\},\\
\iota_{\mu\sigma} &= \frac{2}{\sigma^2}\left\{\int \frac{\Pi''\Pi}{(\Pi')^2} g' - \int \frac{\Pi}{\Pi'}(\log g)'' g\right\},\\
\iota_{\mu\lambda} &= -\frac{2}{\sigma}\left\{\int \left(\frac{\Pi^\lambda}{\Pi'}\right)' g' + \int \frac{\Pi^\lambda}{\Pi'}(\log g)'' g\right\},\\
\iota_{\mu\nu} &= 0,\\
\iota_{\sigma\sigma} &= \frac{1}{\sigma^2} + \frac{2}{\sigma^2}\left\{\int \frac{\Pi''\Pi^2}{(\Pi')^2} g' - \int \frac{\Pi^2}{\Pi'}(\log g)'' g\right\},\\
\iota_{\sigma\lambda} &= -\frac{2}{\sigma}\left\{\int \left(\frac{\Pi^\lambda}{\Pi'}\right)' \Pi g' + \int \frac{\Pi^\lambda \Pi}{\Pi'}(\log g)'' g\right\},\\
\iota_{\sigma\nu} &= \frac{2}{\sigma}\int \Pi(\log g)'^{\nu} g = \frac{1}{\sigma}\int x(\log g)'^{\nu} g,\\
\iota_{\lambda\lambda} &= -2\left\{\int \left(\frac{\Pi^\lambda}{\Pi'}\right)^{\lambda} \Pi^\lambda g' + \int \frac{(\Pi^\lambda)^2}{\Pi'}(\log g)'' g\right\},\\
\iota_{\lambda\nu} &= 0,\\
\iota_{\nu\nu} &= -2\int \Pi'(\log g)^{\nu\nu} g = -\int (\log g)^{\nu\nu} g.
\end{aligned}
$$

In the above, there is no dependence on μ whilst the submatrix associated with μ and σ has elements proportional to σ^{-2}, that associated with λ and ν has elements independent of σ, and the remainder are proportional to σ^{-1}. A relatively unusual feature of four-parameter models that we have here is the existence of two zero elements, namely the elements associating $\hat{\nu}$ with each of $\hat{\mu}$

and $\hat{\lambda}$. These zeroes arise because of the evenness of g and Π^{λ} and the oddness of $(\log g)'^{\nu}$. The second formulae for $\iota_{\sigma\nu}$ and $\iota_{\nu\nu}$ arise from the properties of Π; the information matrix elements associated with $\hat{\nu}$ are, therefore, all independent of Π.

What requirements are there on Π and its derivatives for further zeroes to appear in the information matrix? Since g and $(\log g)''$ are even and $(\log g)'$ is odd, and given that Π should not depend on g, $\iota_{\mu\sigma} = 0$ requires Π/Π' to be an odd function and $\Pi''\Pi/(\Pi')^2$ to be an even function. Since, in addition, Π^{λ} is even and Π'^{λ} is odd, these requirements also result in $\iota_{\sigma\lambda} = 0$. Now from $\Pi(x) - \Pi(-x) = x$, $\Pi'(x) + \Pi'(-x) = 1$ and hence $\Pi''(x) = \Pi''(-x)$, the first requirement translates, via $-(\Pi/\Pi')(x) = (\Pi/\Pi')(-x) = (\Pi(x) - x)/(1 - \Pi'(x))$, to

$$x\Pi'(x) = \Pi(x). \tag{5.2}$$

Since this leads to $x\Pi''(x) = 0$, the second requirement is also satisfied. Equation (5.2) is satisfied only if Π is piecewise constant. The only one parameter transformation of this type corresponds precisely to transformation t_5 and the two-piece distribution. The achievement of these four zeroes in the expected information matrix implies an attractive asymptotic independence between the location-skewness pair of parameters $\{\mu, \lambda\}$ and the scale-tail pair of parameters $\{\sigma, \nu\}$. This is the only known four-parameter family of distributions on $\mathbb{R}$ with this property (Jones and Anaya-Izquierdo (2011)). Further four-parameter distributions on $\mathbb{R}$ with a high level of parameter orthogonality, within the transformation of scale class as well as without, therefore remain elusive.

The specific property of some skew-g distributions that their expected information matrices are singular at parameter configurations corresponding to symmetry (e.g., Ley and Paindaveine (2011)) does not seem to arise for transformation of scale distributions (or other skew-symmetric distributions).

5.2. A small simulation study

In order to check that maximum likelihood estimation is practicable, a small simulation study was run. Using the random variate generation algorithm given in Section 4.4, $10,000$ samples of sizes $n = 100$ and $n = 500$ were generated from transformation of scale distributions with density (3.6) with added location and scale parameters as at (5.1); and with $g = \phi$, $\mu = 0$, $\sigma = 1$ and $\alpha = 0.25, 0.5, 0.75$. Table 1 shows the bias, mean squared error (MSE) and 95% confidence interval (CI) coverage of each of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\alpha}$. The confidence intervals are those obtained by inverting the observed information matrix. Standard errors of estimated biases and MSEs are given in brackets.

Briefly, biases and mean squared errors are generally small, with the expected improvement as n increases. Where biases are significant, they are consistently

Table 1. Simulation Results for Maximum Likelihood Estimation of Parameters of Location-Scale Form of Distribution With Density (3.6).

	bias (s.e.)	MSE (s.e.)	95% CI coverage
$n = 100, \alpha = 0.25$			
$\hat{\mu}$	-0.0121 (0.0040)	0.1617 (0.0026)	0.927
$\hat{\sigma}$	-0.0388 (0.0011)	0.0132 (0.0005)	0.941
$\hat{\alpha}$	0.0093 (0.0020)	0.0395 (0.0007)	0.925
$n = 100, \alpha = 0.5$			
$\hat{\mu}$	-0.0413 (0.0043)	0.1900 (0.0029)	0.923
$\hat{\sigma}$	-0.0704 (0.0020)	0.0442 (0.0014)	0.954
$\hat{\alpha}$	0.0237 (0.0019)	0.0361 (0.0006)	0.911
$n = 100, \alpha = 0.75$			
$\hat{\mu}$	-0.0759 (0.0048)	0.2403 (0.0032)	0.855
$\hat{\sigma}$	-0.2066 (0.0041)	0.2097 (0.0034)	0.852
$\hat{\alpha}$	0.0349 (0.0016)	0.0261 (0.0003)	0.818
$n = 500, \alpha = 0.25$			
$\hat{\mu}$	-0.0006 (0.0016)	0.0262 (0.0004)	0.947
$\hat{\sigma}$	-0.0063 (0.0004)	0.0015 (0.0000)	0.953
$\hat{\alpha}$	0.0010 (0.0008)	0.0061 (0.0001)	0.950
$n = 500, \alpha = 0.5$			
$\hat{\mu}$	-0.0049 (0.0017)	0.0305 (0.0004)	0.944
$\hat{\sigma}$	-0.0094 (0.0006)	0.0035 (0.0001)	0.950
$\hat{\alpha}$	0.0032 (0.0007)	0.0056 (0.0001)	0.940
$n = 500, \alpha = 0.75$			
$\hat{\mu}$	-0.0161 (0.0020)	0.0422 (0.0006)	0.942
$\hat{\sigma}$	-0.0270 (0.0012)	0.0158 (0.0004)	0.956
$\hat{\alpha}$	0.0066 (0.0006)	0.0042 (0.0001)	0.936

negative for estimation of μ and σ and positive for estimation of α. Estimation quality decreases as (positive) α increases. Confidence interval coverage is good, if slightly low, when $n = 100$ and $\alpha = 0.25, 0.5$ but deteriorates somewhat when $n = 100$ and $\alpha = 0.75$; by the time $n = 500$, CI coverage is very good for all tested values of α.

Many further inferential issues such as possible alternative parameter estimation methods for small n, methodology for specification of elements of the model, especially Π, and robustness to misspecification thereof, are important topics for future work. Two examples of the application of transformation of scale models to data can be found in Fujisawa and Abe (2012).

6. Discussion

6.1. Comparison

So, where do distributions with densities given by (1.1) potentially fit relative to competing existing general strategies for generating general families of distributions based on symmetric density g and skewing functions t or equivalently Π? Here, I am thinking principally of the Azzalini-type skew-g model of the form utilised in Sections 2 and 3.1, and transformation of random variable models based on densities of the form $2t'(x)g\{t(x)\}$. Also relevant is the very special transformation of scale distribution that is the two-piece distribution. Yet other important competitors, such as that espoused by Ferreira and Steel (2006), are not considered here.

On $\mathbb{R}$, transformation of scale distributions seem to have some tractability advantages, listed here. Given unimodal g, they are immediately unimodal with explicit mode, whereas skew-g and transformation of random variable models are not necessarily so; both the latter often are unimodal, but this has to be checked on a case-by-case basis. See Section 3.2 of Azzalini and Regoli (2012) for what can be said in general in the skew-g case. Relatedly, only the transformation of scale approach yields such a beautiful Khintchine-type theorem. Also, only the transformation of scale approach is amenable to explicit density-based asymmetry analysis. And only the transformation of scale approach provides a way to alter skewness aspects of the density g without affecting the entropy of the distribution.

On the other hand, it is only the two-piece and transformation of random variable approaches that yield straightforward distribution and quantile functions in terms of those of g. The transformation of random variable approach, by its very nature, is especially well suited to skewness analysis by the classical transformation-based approach of van Zwet (1964). In what might also be considered a slender advantage, both transformation of scale and transformation of random variable distributions afford skew distributions, including the two-piece, whose tails have the same asymptotic order as those of g. This is not always true of skew-g distributions, for which typically only one tail remains the same as the corresponding tail of g. That all said, transformation of scale is probably the weakest of the three main competitors considered here in terms of plausible underlying physical generation processes.

In a different direction, as discussed in Section 5, transformation of scale, like many competing approaches, appears not to have any of the difficulties with maximum likelihood inference associated with the skew-g distribution. It has been argued elsewhere (Jones (2008b), rejoinder) that two-piece distributions are at least broadly comparable to skew-g distributions in general, and that the likelihood fitting considerations of Jones and Anaya-Izquierdo (2011) give it a slight edge.

On $\mathbb{R}^+$, the transformation of scale distributions compete only, amongst the above classes, with transformation of random variable models. When $t(x) = \log(x)$, the latter are the important class of log-symmetric distributions (Lawless (2003), Seshadri (1965)). They are natural competitors/complements to the R-symmetric distributions (Mudholkar and Wang (2007)) as discussed in Chaubey, Mudholkar, and Jones (2010) and Jones (2008c), which are none other than transformation of scale distributions using $t_1(x)$ (Baker (2008), Chaubey, Mudholkar, and Jones (2010).

6.2. Other domains

The general approach also works on other domains and, at least in one case, with real practical promise. In Jones and Pewsey (2012), a similar approach affords a transformation of scale family of distributions which appears to be amongst the best four-parameter unimodal families of distributions on the circle. Retaining unimodality whilst 'skewing', which is trivial in this approach, seems especially difficult for many other approaches on the circle.

6.3. Multivariate considerations

Finally, some brief words on the multivariate case. A natural approach is to apply marginal transformations $X_i = \Pi_i(Y_i)$ to an Azzalini-type multivariate skew-g distribution (e.g., Azzalini and Capitanio (2003), Azzalini (2005)) with marginals $2\pi_i(y)g(y)$, $i = 1, \ldots, d$. This is explored in Jones (2014). However, 'most natural' does not necessarily correspond to 'most useful'. If all marginals being of transformation of scale type remains the focus, a general purpose methodology, most obviously copulas (e.g., Nelsen (2010)), can be employed. In higher dimensions, this has the added advantage of being able to cope with a variety of marginals of different type both within and without the transformation of scale class. If marginals of original variables are secondary, a multivariate distribution might be based on a general linear transformation of independent transformation of scale components (Ferreira and Steel (2007)). Expanding $t(x)$ in (1.1) to $t(x_1, \ldots, x_d)$ results in distributions defined by contour shape (Fernández, Osiewalski, and Steel (1995), Arnold, Castillo, and Sarabia (2008)), but choosing $t(x_1, \ldots, x_d)$ to retain a simple normalising constant is a challenge. See Section 1 of Jones (2014) for a slightly fuller discussion.

Acknowledgement

I am very grateful to the Editor, Associate Editor and two referees for their encouraging comments, and for prompting a beneficial expansion of the article.

References

Arnold, B. C., Castillo, E. and Sarabia, J. M. (2008). Multivariate distributions defined in terms of contours. *J. Statist. Plann. Inference* **138**, 4158-4171.

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *Amer. Statist.* **49**, 34-38.

Avérous, J., Fougères, A. L. and Meste, M. (1996). Tailweight with respect to the mode for unimodal distributions. *Statist. Probab. Lett.* **28**, 367-373.

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

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.* **32**, 159-200.

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

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

Azzalini, A. and Regoli, G. (2012). Some properties of skew-symmetric distributions. *Ann. Inst. Statist. Math.* **64**, 857-879.

Baker, R. (2008). Probabilistic applications of the Schlömilch transformation. *Commun. Statist. Theory Meth.* **37**, 2162-2176.

Bassan, B., Denuit, M. and Scarsini, M. (1999). Variability orders and mean differences. *Statist. Probab. Lett.* **45**, 121-130.

Boros, G. and Moll, V. H. (2004). *Irresistible Integrals; Symbolics, Analysis and Experiments in the Evaluation of Integrals.* Cambridge University Press, Cambridge.

Boshnakov, G. N. (2007). Some measures for asymmetry of distributions. *Statist. Probab. Lett.* **77**, 1111-1116.

Chaubey, Y. P., Mudholkar, G. S. and Jones, M. C. (2010). Reciprocal symmetry, unimodality and Khintchine's theorem. *Proc. Roy. Soc. A* **466**, 2079-2096.

Critchley, F. and Jones, M. C. (2008). Asymmetry and gradient asymmetry functions: density-based skewness and kurtosis. *Scand. J. Statist.* **35**, 415-437.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, Convexity, and Applications.* Academic Press, Boston.

Ebrahimi, N., Soofi, E. S. and Soyer, R. (2010). Information measures in perspective. *Internat. Statist. Rev.* **78**, 383-412.

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

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York.

Fernández, C., Osiewalski, J., and Steel, M. F. J. (1995). Modeling and inference with ν-spherical distributions. *J. Amer. Statist. Assoc.* **90**, 1331-1340.

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

Ferreira, J. T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *J. Amer. Statist. Assoc.* **101**, 823-829.

Ferreira, J. T. A. S. and Steel, M. F. J. (2007). A new class of skewed multivariate distributions with applications to regression analysis. *Statist. Sinica* **17**, 505-529.

Fujisawa, H. and Abe, T. (2012). A family of unimodal skew-symmetric distributions with mode-invariance. Manuscript.

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

Jones, M. C. (2002a). Student's simplest distribution. *J. Roy. Statist. Soc. Ser. D* **51**, 41-49.

Jones, M. C. (2002b). On Khintchine's theorem and its place in random variate generation. *Amer. Statist.* **56**, 304-307.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion). *Test* **13**, 1-43.

Jones, M. C. (2008a). On a class of distributions with simple exponential tails. *Statist. Sinica* **18**, 1101-1110.

Jones, M. C. (2008b). The t family and their close and distant relations (with discussion). *J. Kor. Statist. Assoc.* **37**, 293-311.

Jones, M. C. (2008c). On reciprocal symmetry. *J. Statist. Plann. Inference* **138**, 3039-3043.

Jones, M. C. (2010). Distributions generated by transformation of scale using an extended Schlömilch transformation. *Sankhyā A* **72**, 359-375.

Jones, M. C. (2014). On bivariate transformation of scale distributions. *Commun. Statist. Theory Meth.* to appear.

Jones, M. C. and Anaya-Izquierdo, K. (2011). On parameter orthogonality in symmetric and skew models. *J. Statist. Plann. Inference* **141**, 758-770.

Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika* **96**, 761-780.

Jones, M. C. and Pewsey, A. (2012). Inverse Batschelet distributions. *Biometrics* **68**, 183-193.

Khintchine, A. Y. (1938). On unimodal distributions. *Izvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki* **2**, 1-7.

Kucerovsky, D., Marchand, E. and Small, R. D. (2005). On the equality in distribution of the random variables X and $g(X)$. *Internat. J. Pure Appl. Math.* **23**, 93-114.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* Second edition. Wiley, Hoboken, NJ.

Ley, C. and Paindaveine, D. (2011). On the singularity of multivariate skew-symmetric models. *J. Multivariate Anal.* **101**, 1434-1444.

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

Mudholkar, G. S. and Wang, H. (2007). IG-symmetry and R-symmetry: inter-relations and applications to the inverse Gaussian theory. *J. Statist. Plann. Inference* **137**, 3655-3671.

Nelsen, R. B. (2010). *An Introduction to Copulas.* Second edition. Springer, New York.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference.* Edward Arnold, London.

Seshadri, V. (1965). On random variables which have the same distribution as their reciprocals. *Canad. Math. Bull.* **8**, 819-824.

Shepp, L. A. (1962). Symmetric random walk. *Trans. Amer. Math. Soc.* **104**, 144-153.

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

Wang, J. Z., Boyer, J. and Genton, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica* **14**, 1259-1270.

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

E-mail: chris.jones@open.ac.uk

(Received December 2011; accepted March 2013)