On Distributions Generated by Transformation of Scale

M.C. Jones

Abstract

This paper concerns the surprisingly wide 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 transformation of scale function. Note the simplicity of the normalising constant and its lack of dependence on the transformation function. It turns out that t should be the inverse of a first iterated symmetric distribution function (or of certain extensions thereof). Transformation of scale distributions have a close link with densities of the form $2\pi(x)g(x)$ where π is a skewing function, through a certain transformation of random variables. A particular case of this construction is the Cauchy-Schlömilch transformation recently introduced into statistics by R. Baker. Transformation of scale distributions have a number of attractive tractabilities: modality properties, explicit density-based asymmetry functions and a beautiful Khintchine-type theorem being chief amongst them. Other theoretical properties and one aspect of likelihood inference are also explored.

KEY WORDS: Asymmetry function, Cauchy-Schlömilch transformation, iterated distribution function, Khintchine theorem, normalising constant, self-inverse function, skew distributions, two-piece distributions.

1. INTRODUCTION

Let g be a continuous univariate density function on $\mathbb{R}$ that is symmetric about zero and t a one-to-one onto transformation function with range the support of g. This paper is concerned with families of distributions derived from these two components by what I call 'transformation of scale' (as opposed to the more familiar transformation of random variable). These have density of the form

$$f(x) = 2g(t(x)). \tag{1}$$

Typically, g and t will each involve a shape parameter, that in g controlling some aspect(s) of kurtosis and that in t introducing and controlling skewness. Location and scale parameters should be introduced for practical work in the usual way but can be set equal to 0 and 1, respectively, for theoretical clarity.

The claim that the exceptionally simple form (1) is a density for symmetric g and a reasonably wide 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. Moreover, the normalising constant provided in (1) is just twice that of g! In particular, any parameter(s) involved in t does not enter the normalising constant. (By rescaling g, the constant of proportionality could have been taken to be 1, but the choice of 2 makes the following development fit in better with existing related work. Alternatively, one can think of $2g$ as a density on $\mathbb{R}^+$.)

A trivial example of (1) is the location-scale shift $t(x) = 2x + b$. Some tractable non-trivial examples of other transformations of scale that also work in this way are:

$$t_1(x) = x - \frac{b}{x}, \qquad b > 0, \quad x \in \mathbb{R}^+ \tag{2}$$

(Baker, 2008);

$$t_2(x) = \frac{1}{a}\log(e^{ax} - 1), \qquad a > 0, \quad x \in \mathbb{R}^+ \tag{3}$$

(Jones, 2009a);

$$t_3(x) = c\left(2\sqrt{\frac{x}{c}} - 1\right) I\left(0 < x < c\right) + xI\left(x \geq c\right), \quad c > 0,\ x \in \mathbb{R}^+; \tag{4}$$

$$t_4(x) = d\left(1 - \frac{x^2}{d^2}\right) I\left(x < -d\right) + d\left(\frac{x}{d} + \frac{1}{2} + \sqrt{\frac{x}{d} + \frac{5}{4}}\right) I\left(x \geq -d\right), \tag{5}$$

$d > 0,\ x \in \mathbb{R}$;

$$t_5(x) = \frac{2x}{1+\alpha} I\,(x < 0) + \frac{2x}{1-\alpha} I\,(x \ge 0)\,, \qquad -1 < \alpha < 1, \quad x \in \mathbb{R}, \quad (6)$$

the celebrated two-piece density (Fechner, 1897, Fernández and Steel, 1998) in essentially the parameterisation preferred, for example, by Mudholkar and Hutson (2000); and a novel variation thereon,

$$\begin{aligned} t_6(x) &= \frac{2x}{1+\alpha} I\left(x < -\tfrac{1}{2}(1+\alpha)\right) \\ &+ \frac{1}{\alpha}\left(1 - \sqrt{1-\alpha(4x+\alpha)}\right) I\left(-\tfrac{1}{2}(1+\alpha) \le x < \tfrac{1}{2}(1-\alpha)\right) \\ &+ \frac{2x}{1-\alpha} I\left(x \ge \tfrac{1}{2}(1-\alpha)\right), \qquad -1 < \alpha < 1, \quad x \in \mathbb{R}. \qquad (7) \end{aligned}$$

That all these transformations of scale should work with no change to the normalising constant seems, at first glance, little short of astonishing. However, they can all be proved to give valid densities by integration by substitution and, perhaps, a little legerdemain. 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 (e.g. Boros and Moll, 2004, Amdeberhan, Glasser, Jones, Moll, Posey, and Varela, 2009). For this reason, I will refer to the whole methodology as involving Cauchy-Schlömilch-type transformations. My first purpose here is to dispel the 'magic' of such transformations by offering an alternative, more intuitive, proof of the veracity of such distributions. More importantly, this yields a general approach to *generating* appropriate transformations of scale. There turns out to be a close link to a much better established and popular methodology for generating families of distributions from g, namely the 'skew-g' distributions arising from the seminal work of Azzalini (1985). See Section 2 for this main development.

As well as their simple normalising constants, transformation of scale distributions have certain other attractive and tractable properties which are explored in Section 3. These include modality properties and tailweight control (Section 3.1). Transformation of scale distributions are especially amenable to tractable analysis of their skewness properties through density-based asymmetry functions (Section 3.2). The link with skew-g distributions developed in Section 2.1 affords (Section 3.3) simple random variate generation. An extended Khintchine theorem associated with unimodal transformation of scale distributions is especially attractive, and is provided in

Section 3.4. I also investigate properties of maximum likelihood estimation of parameters in Section 4, with a view towards maximising the amount of parameter orthogonality. It turns out that the only members of the wide class of distributions explored in this paper that display a high degree of parameter orthogonality are the two-piece distributions associated with (6) above. Finally, in Section 5, I assess the pros and cons of transformation of scale distributions relative to their direct competitors/complements such as skew-g distributions and the distributions of transformed random variables.

2. CAUCHY-SCHLÖMILCH-TYPE TRANSFORMATION OF SCALE

2.1 The Essence of Cauchy-Schlömilch-Type Transformation of Scale

I will start this section with a demonstration of a simple mathematical result which underlies, explicitly or implicitly, much of the appeal of the 'skew-g-type' distributions mentioned in the introduction. Let κ and π denote functions on $\mathbb{R}$ such that κ is an even function and $2\pi - 1$ is an odd function (so that $\pi(x) = 1 - \pi(-x)$). Then, assuming integrability of κ and $\pi\kappa$ but not necessarily of π itself,

$$\int_{-\infty}^{\infty} \pi(x)\kappa(x)dx = \int_{-\infty}^{\infty} \pi(-x)\kappa(-x)dx = \int_{-\infty}^{\infty} (1 - \pi(x))\kappa(x)dx$$

so that, rearranging,

$$2\int_{-\infty}^{\infty} \pi(x)\kappa(x)dx = \int_{-\infty}^{\infty} \kappa(x)dx.$$

Now, let $\kappa = g$ be a density function symmetric about zero. Then, with the added requirement that π be a nonnegative function, $2\pi g$ is also a density, but no longer a symmetric one. This is the general formulation pursued by Wang, Boyer, and Genton (2004). Specialising further, if π is taken to be monotone, and hence the distribution or survival function of a distribution symmetric about zero, P say, then $2Pg$ is also a (skew) density function. And specialising once more, a one-parameter extension of g which allows skewness is afforded by taking $P(x) = Q(\lambda x)$, say, where $\lambda \in \mathbb{R}$ and Q is the distribution function of a symmetric distribution with no further shape parameters. This, the density $2Q(\lambda x)g(x)$, is the 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, Genton (2004) and Azzalini (2005). The big point of this paragraph as far as the current article is concerned is that if one proposes to modify a symmetric density g by multiplication by a skewing function π with the appropriate oddness property then the normalising constant is trivial, it's 2, and independent of any parameters that might be present in π.

Now define W to be such that $W'(x) = \pi(x)$. Since W' is nonnegative, W will be an invertible function with domain $\mathbb{R}$ and range $\mathbb{S}$ to be discussed below. Then, making the transformation of random variable $X = W(Y)$ where $Y \sim 2\pi g$, X on $\mathbb{S}$ follows the transformation of scale distribution of form (1) with density

$$2g(W^{-1}(x)). \tag{8}$$

Therefore, Cauchy-Schlömilch-type transformations of scale work — in the sense of simplicity of normalising constant — because they arise from appropriate transformation of the random variable associated with skew-g distributions — which have the same simple normalising constant.

When π is monotone increasing and hence a distribution function, $W(x) = \int_{-\infty}^{x} \pi(w)dw$ has interpretation as the first iterated distribution function (e.g. Bassan, Denuit and Scarsini, 1999). This, in turn, is a certain truncated mean and hence the numerator of the mean residual life function. When π is monotone decreasing and hence a survival function, $W(x) = \int_{-x}^{\infty}(1-\pi(w))dw$ plays the same role.

2.2 Special Cases I: $\mathbb{S} = \mathbb{R}^+$

Any distribution function π of a distribution symmetric about zero yields an extended Schlömilch-type transformation of scale density of form (8). Most such distribution functions yield W with domain $\mathbb{S} = \mathbb{R}^+$. In particular, a sufficient condition for $\lim_{x\to-\infty} W(x) = 0$ is that $\pi(|x|) = o(|x|^{-1})$ as $x \to -\infty$ which retains all distribution functions except those with Cauchy-like tails and heavier. (Precisely these Ws were utilised in a quite different context by Jones, 2008a.)

Few choices of π are fully tractable in the sense of explicit formulae for W and for W^{-1}. For example, the natural (and popular in terms of skew-g distributions) choice of $\pi_N(x) = \Phi(\lambda x)$, the $N(0, 1/\lambda^2)$ distribution function, yields, for $\lambda > 0$, $W_N(x) = x\Phi(\lambda x) + \lambda^{-1}\phi(\lambda x)$ but this is not explicitly invertible. This need not, however, be a total bar to employing W_N^{-1} in (8).

On the whole real line, in my experience, two symmetric distribution functions stand out as being especially tractable. These are the t distribution on two degrees of freedom ($t(2)$) and the logistic distribution with distribution functions, suitably scaled,

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

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, $W_T(x) = \frac{1}{2}(x + \sqrt{4b+x^2})$ so that $W_T^{-1}(x) = x - (b/x)$, that is, (2). The remarkable $t(2)$ distribution (Jones, 2002a) strikes again: the $t(2)$ distribution is at the heart of why the original Cauchy-Schlömilch transformation works!

• for the logistic distribution, $W_L(x) = \log(1+e^{ax})/a$ which leads to $W_L^{-1}(x) = \log(e^{ax}-1)/a$: the logistic distribution therefore leads precisely to the alternative transformation of scale of Jones (2009a), given at (3).

Symmetric distributions on finite support also lead to distributions on $\mathbb{R}^+$. In fact, they are of the form of two-piece Cauchy-Schlömilch-type transformations of scale. A tractable example corresponds to the uniform distribution on $(-c, c)$ ... which leads to transformation of scale t_3 given at (4).

But there is no real need for π to be monotonic. As just one toy illustration of this, take π to be the piecewise linear 'up-and-down' form

$$\pi(x) = \begin{cases} 0 & \text{if } x < -3, \\ \frac{1}{2}(x+3) & \text{if } -3 \le x < -1, \\ \frac{1}{2}(1-x) & \text{if } -1 \le x < 1, \\ \frac{1}{2}(x-1) & \text{if } 1 \le x < 3, \\ 1 & \text{if } x \ge 3. \end{cases}$$

Then

$$t(x) = \begin{cases} 2\sqrt{x} - 3 & \text{if } 0 < x < 1, \\ 1 - 2\sqrt{2-x} & \text{if } 1 \le x < 2, \\ 1 + 2\sqrt{x-2} & \text{if } 2 \le x < 3, \\ x & \text{if } x \ge 3. \end{cases}$$

2.2.1 Inverse Iterated Distribution Functions and Self-Inverse Functions

The special cases of Cauchy-Schlömilch-type transformation that result in distributions on $\mathbb{R}^+$ were also the subject — from a less well understood standpoint — of Jones (2009a). In this section, I have claimed that $t(x)$ in (1) can then be the inverse of a function W where W' is a nonnegative function with values at $-\infty$ falling to zero sufficiently quickly, and with $2W'(x) - 1$ an odd function. (If W' is also monotone increasing, then t is the inverse of the first iterated distribution function of a symmetric distribution.) But in Jones (2009a), 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 $s^{-1}(x) = s(x)$. Are these two formulations equivalent? The answer is yes, except for cases where $W : \mathbb{R} \to \mathbb{R}^+$ corresponds to π functions on finite support. Here is a demonstration of this equivalence.

Write $s(x) = x - W^{-1}(x)$ for $x = W(y) > 0$ and $y \in \mathbb{R}$. Then,

$$\begin{aligned}
& s(s(x)) = x \\
\text{iff} \quad & W^{-1}(x) = -W^{-1}(x - W^{-1}(x)) \\
\text{iff} \quad & W^{-1}(W(y)) = -W^{-1}(W(y) - W^{-1}(W(y))) \\
\text{iff} \quad & -y = W^{-1}(W(y) - y) \\
\text{iff} \quad & W(-y) = W(y) - y \\
\text{iff} \quad & W'(y) + W'(-y) = 1.
\end{aligned}$$

The last statement is sufficient for the previous one (as well as being obviously necessary for it) because

$$\begin{aligned}
W(x) - W(-x) &= \int_{-x}^{x} W'(y)dy = \int_{0}^{x} W'(y)dy + \int_{-x}^{0} W'(y)dy \\
&= \int_{0}^{x} W'(y)dy + \int_{0}^{x} (1 - W'(y))dy = \int_{0}^{x} dy = x.
\end{aligned}$$

The essential equivalence between inverse first iterated distribution functions of symmetric distributions on the whole of $\mathbb{R}$ (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 2.2, 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 $-\log(1-e^{-ax})/a$. Likewise, self-inverse functions (e.g. Kucerovsky, Marchand and Small, 2005, Jones, 2009a) can correspond to inverse iterated symmetric distribution functions and hence correspond to symmetric distribution functions themselves. Other explicitly tractable examples of this correspondence are, however, hard to come by.

That said, the W^{-1} formulation is more general than the self-inverse formulation because W might also be based on π on finite support or on non-monotone π.

2.3 Special Cases II: $\mathbb{S} = \mathbb{R}$

Suppose that, for sufficiently large x, $\pi(x) > \pi(-x)$. If for $x \to -\infty$, $\pi(x)$ tends to zero sufficiently slowly or if $\lim_{x\to-\infty} \pi(x)$ equals a constant conveniently written as $\frac{1}{2}(1+\alpha)$ for some $-1 < \alpha < 0$, then $f(x) = 2g(W^{-1}(x))$ will be a valid density on the whole of $\mathbb{R}$. The manner in which π behaves for minus large x influences the weight of the left-hand tail of f; see Section 3.

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 scale parameter $d > 0$,

$$W_C(x) = \frac{1}{2}\left[x + \frac{1}{\pi}\left\{2x\tan^{-1}(dx) - \frac{1}{d}\log(1+d^2x^2)\right\}\right]$$

but the inverse of this is not explicitly available.

A more tractable alternative that corresponds to a heavy-tailed density on $\mathbb{R}$ is transformation t_4 given at (5). This corresponds to the very heavy-tailed distribution with density

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

Now allow non-zero limiting values for π (like atoms of probability at $\pm\infty$). The most extreme case of this is to set

$$\pi(x) = \tfrac{1}{2}(1+\alpha)I(x<0) + \tfrac{1}{2}(1-\alpha)I(x \geq 0),$$

$-1 < \alpha \leq 0$, which leads to the two-piece density with transformation t_5 given at (6).

Amelioration of the abrupt step at $x = 0$ leads to other Cauchy-Schlömilch-type transformations of which a tractable one is given by t_6 at (7): this corresponds to a linear portion inserted in π between -1 and 1 continuously joining $\frac{1}{2}(1+\alpha)$ and $\frac{1}{2}(1-\alpha)$. (More generally, but at the cost of further parameters, the linear portion could be on a different interval and need not join up continuously.)

In the case of $\mathbb{S} = \mathbb{R}$, decreasing π produces transformations of the form $-t(-x)$ and sometimes these are equivalent to expanding the range of the parameter involved in the transformation of scale since $-t(-x;\alpha) = t(x;-\alpha)$. In particular, this accounts for the wider range of α in (6) and (7) than is directly allowed in the increasing π case above.

3. ATTRACTIVE PROPERTIES OF CAUCHY-SCHLÖMILCH-TYPE TRANSFORMATION OF SCALE

3.1 Some Basic Properties

I have already stressed the simplicity of the normalising constant in (1) relative to the difficulty of obtaining the normalising constant in models of the form $g(t(x))$ for t not of the form W^{-1} in (8), and the fact that the normalising constant does not depend on any parameters involved in t. Model (8) has other attractive properties, only the most striking of which are described here; see Jones (2009a) for more in the case that $\mathbb{S} = \mathbb{R}$.

Immediately, density (8) has the same modality as g and, in particular, unimodal g leads to unimodal f. Moreover, the mode of unimodal f is explicitly given by $x_0 = W(0)$.

Since $W^{-1}(x) \sim x$ as $x \to \infty$, f retains precisely the same right-hand tail behaviour as g. 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 and then to controlling its behaviour at and near zero), the heaviest left-hand tail of f being the same as that of g when $\lim_{x\to-\infty} \pi(x)$ is a non-zero constant. When $\mathbb{S} = \mathbb{R}^+$, finite support π allows $0 < f(0) < \max_x g(x)$, else $f(0) = 0$. The precise order of contact of f at zero is controlled by the left-hand tail behaviour of π: the heavier its tail, up to $O(|x|^{-1})$, the lighter the contact of f. When $\mathbb{S} = \mathbb{R}$, π functions with $O(|x|^{-\alpha})$ left-hand tail, $0 < \alpha \leq 1$, make the left-hand tail of f decrease as $g(|x|^{1/(1-\alpha)})$ which is lighter than the original tail of g, the larger α, the lighter the tail; π functions with constant limits retain the order of magnitude of g's tails but introduce a relative scaling factor of the form $(1-\alpha)/(1+\alpha)$. (Of course, tails are switched when $t(-x)$ replaces $t(x)$.)

3.2 Skewness

Transformation of scale distributions lend themselves admirably to investigation via the density-based asymmetry measure of Avérous, Fougères and Meste (1996), Boshnakov (2007) and Critchley and Jones (2008) (here studied in the version given by the last reference). Moreover, transformation of scale distributions afford explicit mathematical formulae for the asymmetry function $\gamma(p)$, $0 < p < 1$; this tractability of asymmetry functions had previously been singularly lacking. Assume that g is unimodal and write $c_g(p) = g_R^{-1}(pg(0)) > 0$ which is typically explicitly available. Here, $g_R(x) = g(x)I(x > 0)$. Then, it can easily be seen that

$$\gamma(p) = \frac{W(c_g(p)) - 2W(0) + W(-c_g(p))}{W(c_g(p)) - W(-c_g(p))}$$

which, by virtue of $W(x) - W(-x) = x$, reduces to

$$\gamma(p) = 2c_g(p)^{-1}\{W(c_g(p)) - W(0)\} - 1. \tag{9}$$

Now concentrate on increasing π. Then, $W(c_g(p)) - W(0) = \int_0^{c_g(p)} \pi(y)dy > c_g(p)\pi(0) = \frac{1}{2}c_g(p)$. It follows that $\gamma(p) > 0$ for all $0 < p < 1$, that is, transformation of scale based on increasing π induces positive skew in f in this quite strong sense. Moreover, let π involve a scale parameter λ so that π is written $\pi(\lambda x)$ (as in Azzalini, 1985) and W becomes $\lambda^{-1}W(\lambda x)$. Then,

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

because $W(\lambda c_g(p)) - W(0) = \int_0^{\lambda c_g(p)} \pi(y)dy < \lambda c_g(p)\pi(c_g(p))$. That is, density-based skewness increases with λ. This applies to transformations $t_1, ..., t_4$ when $\lambda = 1/b, a, 1/c, 1/d$, respectively. See Jones (2009a) for further investigation of γ when $\mathbb{S} = \mathbb{R}^+$ and graphs of asymmetry functions when

W corresponds to transformations t_1 and t_2. The two-piece distribution associated with t_5 has the special property of constant density-based skewness (Boshnakov, 2007, Critchley and Jones, 2008); in fact, $\gamma(p) = -\alpha$ for all p.

The skewness functions associated with transformation of scale can be seen to arise from the way such transformations (in the case of increasing π) reduce the weight of the left-hand tail of g relative to that of its right-hand tail.

The scalar skewness measure most naturally associated with density-based skewness (Critchley and Jones, 2008) is that of Arnold & Groeneveld (1995): $\gamma = 1 - 2F(x_0)$ where F is the distribution function associated with unimodal f and x_0 is the mode. In our case, we find, simply and attractively, that

$$\gamma = 4\int_0^\infty \pi(x)g(x)dx - 1.$$

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

It is also the case that, in the classical van Zwet (1964) sense, density (8) is more positively (resp. negatively) skewed than the density $2\pi g$ whenever π is increasing (resp. decreasing) because the transformation between the two, $X = W(Y)$, is convex (resp. concave).

In all the explicit cases in the paper, the most skew version of f is the half-g density $2g(x)I(x > 0)$.

3.3 Random Variate Generation

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

$$X = W(A) - A\ I(U > \pi(A)). \tag{10}$$

This is because $Y \sim 2\pi g$ can be obtained as $A\ I(U \leq \pi(A)) - A\ I(U > \pi(A))$, $X = W(Y)$ and $W(-x) = W(x) - x$.

3.4 An Extended Khintchine Theorem

First, some background. On $\mathbb{R}$, $X \sim g$ for unimodal (but not necessarily symmetric) g has a representation as a uniform scale mixture: setting the mode to zero (as is appropriate to symmetric g), $X \sim UZ$ 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 (Khintchine, 1938, Shepp, 1962, Feller, 1971, p.158, Dharmadhikari and Joag-Dev, 1988, Section 1.2, Jones, 2002b). Mudholkar and Wang (2007) proved a Khintchine-type theorem for unimodal R-symmetric densities; see also Chaubey, Mudholkar and Jones (2009) who identify R-symmetric distributions with Cauchy-Schlömilch-type distributions with transformation of scale (2). Their development can be extended to prove an elegant extension of the Khintchine theorem for general unimodal distributions of type (8) as follows.

Theorem. *If $X \sim f$ where f is of form (8) and is unimodal, then*

$$X \sim UZ + W(Z) - Z \tag{11}$$

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

Proof. Let $f_\ell(x) = f(x)I(x < x_0), g_\ell(x) = g(x)I(x < 0)$ and $f_r(x) = f(x)I(x > x_0), g_r(x) = g(x)I(x > 0)$; these functions are invertible and $f_r^{-1}(y) = W(g_r^{-1}(y/2)), f_\ell^{-1}(y) = W(-g_r^{-1}(y/2))$. If (X,Y) is uniformly distributed over the region between the x-axis and f, then, unconditionally, $X \sim f$, but conditionally

$$X|Y = y \sim U(f_\ell^{-1}(y), f_r^{-1}(y)).$$

It follows that $X|Z = z_1 \sim U(W(z_1)-W(-z_1))+W(-z_1) = Uz_1+W(z_1)-z_1$ where $z_1 = g_r^{-1}(y/2) \in (0,\infty)$ or $X|Z = z_1 \sim (1-U)(W(z_1) - W(-z_1)) + W(-z_1) = W(z_1)-Uz_1 = Uz_2+W(z_2)-z_2$ where $z_2 = -z_1$. Unconditionally, the form of (11) follows. It remains to deduce the distribution of $Z_1 = g_r^{-1}(Y/2)$ (and hence of $Z = \pm Z_1$). A minor modification of a calculation in Chaubey, Mudholkar and Jones (2009) shows that $Y \in (0, 2g(0))$ has distribution function

$$F_Y(y) = F(f_\ell^{-1}(y)) + y\{f_r^{-1}(y) - f_\ell^{-1}(y)\} + 1 - F(f_r^{-1}(y))$$

and hence $Z \in \mathbb{R}^+$ has distribution function

$$F_Z(z) = 1 - \{F(W(-z)) + 2zg_r(z) + 1 - F(W(z))\}.$$

Differentiating,

$$f_Z(z) = -2\{-\pi(-z)g(z) + g(z) + zg'(z) - \pi(z)g(z)\} = -2zg'(z)$$

and the result follows. $\square$

Note that this theorem reduces immediately to the usual Khintchine theorem when $W(Z) = Z$. It also reduces to the result in Chaubey, Mudholkar and Jones (2009) when W is the inverse of transformation t_1. Formula (11) provides an alternative means of random variate generation from f to that in (10) when f is unimodal.

4. MAXIMUM LIKELIHOOD CONSIDERATIONS

Let $X_1, ..., X_n \in \mathbb{R}$ be a random sample assumed to come from the location-scale version of (8), namely the distribution with density

$$2\sigma^{-1}g(W_\lambda^{-1}(\sigma^{-1}(x-\mu)); \nu) \tag{12}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and, for concreteness, ν is the shape parameter associated with g and λ the skewness parameter that forms part of W. (When the data are nonnegative, μ can be dispensed with.) I shall use a prime to denote differentiation with respect to x and superscripts ν and λ to denote differentiation with respect to each of those parameters; for example, $(\log g)'^\lambda(x)$ denotes $\partial^2(\log g)(x)/\partial x \partial\lambda$. Also, hats over parameters will denote their maximum likelihood estimators.

Elements of the expected information matrix, denoted by ι with appropriate subscripts, are given in the Appendix. 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 $\hat{\lambda}$.

What requirements are there on W and its derivatives for further zeroes to appear in this matrix? Since g and $(\log g)''$ are even and $(\log g)'$ is odd, and given that W should not depend on g, $\iota_{\mu\sigma} = 0$ requires W/W' to be an odd function and $W''W/(W')^2$ to be an even function. Since, in addition, W^λ is even and W'^λ is odd, these requirements also result in $\iota_{\sigma\lambda} = 0$. Now from

$W(x) - W(-x) = x$, $W'(x) + W'(-x) = 1$ and hence $W''(x) = W''(-x)$, the first requirement translates, via $-(W/W')(x) = (W(x) - x)/(1 - W'(x))$, to

$$xW'(x) = W(x). \tag{13}$$

Since this leads to $xW''(x) = 0$, the second requirement is also satisfied. (13) is satisfied only if W 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 $\{\sigma, \nu\}$. I know of only this very one extant four-parameter family of distributions with this property (Jones, 2009b). Other piecewise constant transformation of scale functions, such as

$$t_7(x) = \frac{2x}{1+\alpha} I\left(x < -\beta\right) + 2xI\left(-\beta \le x < \beta\right) + \frac{2x}{1-\alpha} I\left(x \ge \beta\right),$$

$-1 < \alpha < 1$, $\beta > 0$, work only if their other parameter(s), here β, remain fixed, which seems inappropriate for data analysis. Further four-parameter distributions with a high level of parameter orthogonality therefore remain elusive.

5. DISCUSSION

So, where do Cauchy-Schlömilch-type distributions with densities given by (1) or (8) potentially fit relative to competing existing general strategies for generating general families of distributions based on symmetric g and t or equivalently W? Here, I am thinking principally of the Azzalini-type skew-g model of the form utilised in Section 2.1 and transformation-of-random-variable (henceforth, just 'transformation') models based on densities of the form $2t'(x)g(t(x))$. Also relevant is the very special Cauchy-Schlömilch-type distribution that is the two-piece distribution.

Aside from two-piece distributions, Cauchy-Schlömilch-type distributions with support $\mathbb{R}$ are, perhaps, rather thinner on the ground than their opposite numbers. When they do exist, they have some tractability advantages. Given unimodal g, they are unimodal when skew-g and transformation models are not necessarily so (both the latter often are unimodal, but this has to be checked on a case-by-case basis). Relatedly, only the Cauchy-Schlömilch approach yields such a beautiful Khintchine-type theorem. Also, only the Cauchy-Schlömilch approach is amenable to explicit density-based skewness analysis. On the other hand, it is only the two-piece and transformation approaches that yield straightforward distribution and quantile functions in terms of those of g. The transformation approach, by its very nature, is especially well suited to skewness analysis by the classical transformation-based approach of van Zwet (1964). I have argued elsewhere (rejoinder to Jones, 2008b) that two-piece distributions are at least broadly comparable to skew-g distributions, and the likelihood fitting considerations of Jones (2009b) and Section 4 give it a slight edge at present, at least in my mind.

On $\mathbb{R}^+$, the Cauchy-Schlömilch-type distributions compete only, amongst the above classes, with transformation models. When $t(x) = \log(x)$, the latter are the important class of log-symmetric distributions (Seshadri, 1965, Lawless, 2003). They are natural competitors/complements to the R-symmetric distributions (Mudholkar & Wang, 2007) as discussed in Jones (2008c) and Chaubey, Mudholkar & Jones (2009) ... which are none other than Cauchy-Schlömilch-type distributions using $t_1(x)$ (Baker, 2008, Chaubey, Mudholkar & Jones, 2009). Other complements, with what, more generally, I called "S-symmetric" distributions in Jones (2009a), have not yet been explored. t_1 has, however, been explored in a transformation of random variable context in Jones (2007).

Finally, it can be argued that some of the distributions generated, especially on $\mathbb{R}^+$, might not be of immense immediate practical value (except where they coincide with existing distributions such as the lognormal). The general approach works on other domains, however. A particularly exciting development is being pursued for distributions on the circle in joint work with Arthur Pewsey. A Cauchy-Schlömilch-type approach taken there affords a transformation-of-scale family of distributions which appears to offer advantages over currently existing distributions closely related to distributions described in Batschelet (1981, Section 15.6); the new distributions appear to be amongst the most practically promising four-parameter *unimodal* families of distributions on the circle.

APPENDIX: ELEMENTS OF THE 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. By way of shorthand the arguments of integrands will be omitted. Then, it turns out that

$$\iota_{\mu\mu} = \frac{2}{\sigma^2}\left\{\int \frac{W''}{(W')^2}g' - \int \frac{1}{W'}(\log g)''g\right\},$$

$$\iota_{\mu\sigma} = \frac{2}{\sigma^2}\left\{\int \frac{W''W}{(W')^2}g' - \int \frac{W}{W'}(\log g)''g\right\},$$

$$\iota_{\mu\lambda} = -\frac{2}{\sigma}\left\{\int \left(\frac{W^\lambda}{W'}\right)' g' + \int \frac{W^\lambda}{W'}(\log g)''g\right\},$$

$$\iota_{\mu\nu} = 0,$$

$$\iota_{\sigma\sigma} = \frac{1}{\sigma^2} + \frac{2}{\sigma^2}\left\{\int \frac{W''W^2}{(W')^2}g' - \int \frac{W^2}{W'}(\log g)''g\right\},$$

$$\iota_{\sigma\lambda} = -\frac{2}{\sigma}\left\{\int \left(\frac{W^\lambda}{W'}\right)' Wg' + \int \frac{W^\lambda W}{W'}(\log g)''g\right\},$$

$$\iota_{\sigma\nu} = \frac{2}{\sigma}\int W(\log g)'^{\nu}g,$$

$$\iota_{\lambda\lambda} = -2\left\{\int \left(\frac{W^\lambda}{W'}\right)^\lambda W^\lambda g' + \int \frac{(W^\lambda)^2}{W'}(\log g)''g\right\},$$

$$\iota_{\lambda\nu} = 0,$$

$$\iota_{\nu\nu} = -2\int W'(\log g)^{\nu\nu}g.$$

The zeroes here arise because of the evenness of g and W^λ and the oddness of $(\log g)'^{\nu}$.

REFERENCES

Amdeberhan, T., Glasser, M.L., Jones, M.C., Moll, V.H., Posey, R. and Varela, D. (2009) The Cauchy–Schlömilch transformation. Preprint.

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.

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.

Batschelet, E. (1981) *Circular Statistics in Biology*. Academic Press, London.

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

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. (2009) Reciprocal symmetry, unimodality and Khintchine's theorem. Under consideration.

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.* Boston: Academic Press.

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

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

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

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

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. (2007). Connecting distributions with power tails on the real line, the half line and the interval. *Internat. Statist. Rev.*, **75**, 58–69.

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. Planning Inference*, **138**, 3039–3043.

Jones, M.C. (2009a) Distributions generated by transformation of scale using an extended Schlömilch transformation. Open University Statistics Group Technical Report 09/01, `http://stats-www.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`

Jones, M.C. (2009b) On parameter orthogonality in symmetric and skew models. Open University Statistics Group Technical Report 09/02, `http://stats-www.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`

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*, 2nd Edition. Wiley, Hoboken, NJ.

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

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

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.* Amsterdam: Mathematisch Centrum.

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