On Families of Distributions With Shape Parameters

M.C. Jones

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

E-mail: m.c.jones@open.ac.uk

Summary

Univariate continuous distributions are one of the fundamental components on which statistical modelling, ancient and modern, frequentist and Bayesian, multidimensional and complex, is based. In this article, I review and compare some of the main general techniques for providing families of typically unimodal distributions on $\mathbb{R}$ with one or two, or possibly even three, shape parameters, controlling skewness and/or tailweight, in addition to their all-important location and scale parameters. One important and useful family is comprised of the 'skew-symmetric' distributions brought to prominence by Azzalini. As these are covered in considerable detail elsewhere in the literature, I focus more on their complements and competitors. Principal amongst these are distributions formed by transforming random variables, by what I call 'transformation of scale' — including two-piece distributions — and by probability integral transformation of non-uniform random variables. I also treat briefly the issues of multivariate extension, of distributions on subsets of $\mathbb{R}$, and of distributions on the circle. The review and comparison is not comprehensive, necessarily being selective and therefore somewhat personal.

Key words: circular distributions; interpretable parameters; kurtosis; multivariate; probability integral transform; skew-symmetric; skewness; tailweight; transformation of random variable; transformation of scale; two-piece; unimodality; univariate continuous.

1 Introduction

Univariate continuous distributions are one of the fundamental components on which statistical modelling, ancient and modern, frequentist and Bayesian, multidimensional and complex, is based. On the whole of $\mathbb{R}$, the distributions mostly used are simple ones, usually symmetric, typically with two parameters, one, $\mu \in \mathbb{R}$, controlling the location of the distribution, the second, $\sigma > 0$, controlling its scale; the normal distribution is, of course, the premier such example. Here, I am interested in families of distributions on $\mathbb{R}$ with additional shape parameters, typically two in number, between them controlling skewness (or asymmetry) and tailweight (or kurtosis or elongation). Of course, this is because data often display such traits in ways not accommodated by the basic distributions. The importance of empirically modelling such traits is usually for the purpose of model-based robust estimation of the location and perhaps the scale, which remain the parameters of main interest. This is especially so in regression contexts when the interest parameters in turn depend on covariates, from a simple binary treatment effect to a hugely complicated multi-faceted regression model. Such families of distributions also have roles in complex statistical models as link functions, random effect distributions and as families of informative priors for Bayesian analysis. Understanding the skewness and/or tailweight is only more occasionally of direct interest in itself. The case for interest in such distributions in parametric modelling is, I think, clear (see also Azzalini & Genton, 2008); the purpose of this article is to explore and compare some of the many possible ways of providing them.

One particular such family of distributions has gained a high profile in the literature in recent years. These are the skew-symmetric distributions of Azzalini (1985); for book-length treatments, see Genton (2004) and Azzalini and Capitanio (2014). In this article, I especially wish to review and compare some of the alternative families of distributions that have also been proposed. It seems to me that each of these families of distributions has its pros and its cons — and there are plenty of equivalences too — and that no one of them should be unequivocally pre-eminent. Rather, choices between them should mostly be made on the basis of important problem-specific requirements, though are partly a matter of taste.

Some personal preferences for aspects of such distributions will be to the fore in this article, including (i) unimodality, (ii) tractability, (iii) interpretability and (iv) parsimony. Let me consider each briefly in turn.

(i) In my view, bi- and multi-modal data are best modelled *interpretably* by finite mixtures of unimodal component distributions. What might be called 'collateral bimodality', arising as an often unintended consequence of particular choices of parameter values in a mathematical formula for a density, is usually unattractive on interpretability grounds, constraining in the type of bimodality available and more likely to force a bimodal model on data showing no such trait.

(ii) It can certainly be argued that in an age of fast computers, mathematical tractability is not an issue of overwhelming importance. However, straightforward mathematical formulae describing features of distributions remain a springboard to insight, interpretation and clarity of exposition, as well as improving computational speed and convenience. All other things being equal — which they never are! — tractability is still to be preferred to non-tractability.

(iii) Interpretability, by which I mean interpretability of the roles of the parameters, has already featured in points (i) and (ii). Understanding particular features of distributions is greatly enhanced if they correspond to specific parameters. Perhaps more importantly but also perhaps less concretely, interpretable parameters implies parameters describing different features of distributions which should imply parameters being better estimated, as different aspects of the data affect different aspects of the distribution (this might be manifest, for example, in parameter orthogonality).

(iv) The role of parsimony in statistical modelling, to aid interpretation (that word again!), to facilitate estimation and particularly prediction, affording generalisability of results by avoiding overfitting, is clear. In developing families of distributions, however, the watchword has usually been flexibility, and parsimony is little mentioned. I am all for flexibility at the level of the (three-, four-, perhaps five-parameter) unimodal distributions that are the topic of this paper, but it is hard to justify more (except in multimodal situations). Remember too that complex models might involve many univariate distributions.

The common basic ingredients underlying the families of distributions considered in this paper will be described in Section 2. In Section 3, I will review four main families, one with a prominent sub-family. A comparison of the merits and demerits of these families, bearing in mind the discussion above, will take place in Section 4.

Despite the main focus of the paper being on univariate continuous distributions with support the whole of $\mathbb{R}$, I will briefly address some of the many related issues to do with distributions on different supports in Section 5. Perhaps the most obvious is the multivariate case, but I argue in Section 5.1 that it is not necessarily the most interesting in various senses, one of which being that univariate distributions can be generalised to the multivariate case using existing *general* techniques (e.g. copulas). Distributions on subsets of $\mathbb{R}$ and simple manifolds, specifically the circle, are considered in Sections 5.2 and 5.3. Discrete distributions will not be considered in this article. Sections 2 to 4 will be focussed on three- and four-parameter distributions on $\mathbb{R}$ but the position of the five-parameter (unimodal) case will be briefly addressed in Section 6. The paper finishes with further remarks and discussion in Section 7. Throughout, use the notation '$\sim$' to denote 'follows the distribution with density'.

It should be stressed that this review and comparison, although wide-ranging, is not intended to be comprehensive. The literature is too vast: this is a topic which lends itself to industrial-scale paper production based on writing down formulae for special cases. The review is therefore selective and the comparison is necessarily personal.

2 Ingredients

All families of distributions considered in this article will be based on extensions of a simple, *symmetric*, *unimodal*, distribution on $\mathbb{R}$ which has density function g and distribution function G. (If this distribution is the normal distribution, I shall follow the usual convention of writing $g = \phi$ and $G = \Phi$.) In practice, this distribution would be used in the form of a location-scale family, $g_{\mu,\sigma}(x) = \sigma^{-1}g(\sigma^{-1}(x-\mu))$. However, for most purposes in this paper, those all-important location and scale parameters can be ignored by setting $\mu = 0$, $\sigma = 1$. Non-zero location and non-unit scale parameters should be reintroduced for practical work in the same manner for extensions of g. (And μ and perhaps other parameters might depend on covariates.) It is often assumed in what follows that g has no other, shape, parameter, and is therefore a two-parameter distribution, though sometimes it is useful to employ a (symmetric) g with a third, tailweight-controlling, parameter.

The extensions of g will all be based on the application of a function W say, or its derivative $w = W'$, to the distribution G, in some appropriate sense. This function will depend on one or two parameters; a two-parameter function W or w applied to a two-parameter g will result in a four-parameter family of distributions, while a similar effect can be produced by applying a one-parameter W or w to a three-parameter g, in which case W or w might be called a skewing function; three-parameter subfamilies of these four-parameter families are often of interest also. For the purposes of unification, I will take $W : \mathbb{R} \to \mathbb{R}$ to be a monotone transformation-type function and note that, if W is convex (concave) increasing, w might be a distribution (survival) function $w : \mathbb{R} \to [0, 1]$. (And later on I will need the derivative of w, denoted w', which in the concave case would then be a density function.) This unification will be stretched at times by the addition/removal of some important further requirements on W and/or w.

3 The Menu: Four (or Five) Main Families

3.1 Family 1: Azzalini–Type Skew–Symmetric Distributions

In its greatest (univariate) generality (Wang, Boyer & Genton, 2004), the density of a member of this type of distribution has the beautifully simple form

$$f_A(x) = 2w(x)g(x) \tag{1}$$

where $w > 0$ additionally satisfies the key oddness property

$$w(x) + w(-x) = 1. \tag{2}$$

It is the evenness of g and the oddness of $2w-1$ that immediately lead to the normalising constant being 2. The archetypal, but not only, examples of w are distribution functions of symmetric distributions themselves, $w(x) = D(\lambda x)$ say, where D may or may be the same as G and the scale parameter λ of D acts as a skewing parameter for f_A (Azzalini, 1985). The not-especially-attractive name skew-symmetric is most often used for these distributions (though they are sometimes called 'perturbed' distributions and Azzalini has recently proposed 'modulated' instead), reflecting use of w principally as a skewing mechanism; in this context, unimodal four-parameter distributions arise most usually by one-parameter w being applied to three-parameter g.

These distributions and their multivariate generalisations have become extremely popular in the literature (e.g. Azzalini & Capitanio, 2003, 2014, Genton, 2004, Azzalini, 2005, Azzalini & Genton, 2008 — Azzalini's works are mentioned so much because the more recent provide excellent overviews of developments, including those by others). This seems to be partly because of modelling mechanisms which lead to them (see below), and partly because of their apparent simplicity (and partly, I might add, because they are not quite as simple or well-behaved as might be expected and hence there are papers to be written on improving matters!). Among the attractive properties of f_A, arising from (2), is that, if $X_A \sim f_A$, then $|X_A| \sim g$; for example, if $g = \phi$, those aspects of normal theory based on the (chi-squared) distribution of Y^2 where $Y \sim g$ remain valid for $X_A \sim f_A$. (I think I agree that this is attractive though some have worried that this shows that f_A does not deviate sufficiently far from g.)

But what are the distribution and quantile functions of f_A? The distribution function, even for $w = G = \Phi$, is not in closed form; it's a non-standard but well-behaved, one-dimensional, integral (so does this lack of tractability really matter?). Is f_A unimodal? Not necessarily, although it is in many important special cases; see Azzalini & Regoli (2012) for a general investigation. Even if unimodal, the mode is typically not explicit. Several papers point to real and perceived difficulties with maximum likelihood estimation for these distributions but I am unsure how important these are; see Section 4.

Densities f_A arise from a variety of simple modelling mechanisms. Principally, they are the marginal distribution of the untruncated variable when another, hidden, variable in a bivariate (e.g. normal) distribution is truncated. This results in what, more generally, is a weighted distribution (one with density proportional to $w(x)g(x)$

for a general $w > 0$) which is a distribution arising from a biased sampling mechanism (e.g. Patil & Rao, 1977). Weighted distributions themselves form families of distributions with shape parameters and have been considered so a little. But here, the particular form of w as given by (2) is so special that not only does it lead to the normalising constant 2 (in weighted distribution terms, $E_g(w(Y)) = 1/2$) but it also allows the weighting mechanism to translate to a random sign mechanism: $X_A \sim f_A$ is the result of attaching a random sign — positive with probability $w(y)$, given $Y = y$ — to a random variable $Y \sim g$. (Hence the invariant distributions of $|X_A|$ and X_A^2 above.) In some cases, skew-symmetric distributions can also arise as the distributions of weighted sums of random variables following g and 'half-g' distributions, or as the distributions of maxima and minima from suitable correlated multivariate distributions. I will return to the importance of these characterisations in Section 4 below.

The above should nonetheless have established that, however attractive these skew-symmetric distributions might be — and they *are* attractive — it is worth considering alternatives which may, or may not, prove equally or more attractive.

3.2 Family 2: Transformation of Random Variable

Family 2 is based on transformations of random variables as usually understood. If $Y \sim g$, then consider the distribution of $X_R = W(Y)$ where W is a monotone transformation function, depending on one or two parameters, that I shall take to be increasing with domain and range $\mathbb{R}$. The identity transformation should be a special case of W, but otherwise I impose no further requirement on W in most of this subsection. The density of X_R is then, from undergraduate statistics,

$$f_R(x) = g(W^{-1}(x))/w(W^{-1}(x)). \tag{3}$$

If W is explicitly invertible, then f_R is tractable, with distribution function $G(W^{-1}(x))$ and quantile function $W(G^{-1}(u))$. Unfortunately, unimodality is not, in general, invariant to transformation of random variables.

A particular favourite of mine is a flexible and tractable two-parameter transformation called the sinh-arcsinh transformation:

$$W_{sas}(x) = \sinh(a + b \sinh^{-1}(x)) \tag{4}$$

(Jones & Pewsey, 2009, who parametrise by $\delta = 1/b$, $\epsilon = a/b$). Here, $a \in \mathbb{R}$ controls skewness, its sign corresponding to the sign of skewness produced, and $b > 0$ controls tailweight, $b > 1$ leading to heavier tails (than those of g), $b < 1$ to lighter tails. In this family, the parameters are true skewness and kurtosis parameters in the quantile-based sense of van Zwet (1964). It is especially attractive that g lies in the 'centre' of the family of distributions, with lighter tails allowed as well as heavier; g corresponds to $a = 0$, $b = 1$. Note also that W_{sas} is explicitly invertible with W_{sas}^{-1} taking the same form as W_{sas}. When $g = \phi$, $f_{R;sas}$ is unimodal for all $a \in \mathbb{R}$, $b > 0$ (Jones & Pewsey, 2009).

A popular alternative two-parameter transformation is that associated with the g-and-h distributions (Hoaglin, 1985):

$$W_{gh}(x) = g^{-1}(e^{gx} - 1)e^{hx^2/2},$$

$g \in \mathbb{R}$, $h \geq 0$. Here, I use the eponymous parametrization that hopefully causes no confusion with my use of g as base density; g-and-h distributions themselves take my $g = \phi$ but other g could obviously be accommodated. A full comparison of g-and-h with sinh-arcsinh distributions remains to be performed. Certainly, g-and-h distributions share some of sinh-arcsinh's nice properties: van Zwet parameter interpretation, unimodality when $g = \phi$. However, they lack an explicit inverse for W_{gh}, have ϕ only as a parameter boundary case ($g \to 0$, $h = 0$) and, when $g \neq 0$, $h = 0$, W_{gh} has a one-sided range with boundary value $-1/g$ which is a lower boundary if $g > 0$, or upper boundary if $g < 0$. Consequentially, when $h = 0$, inference becomes irregular. On the other hand, in symmetric cases, g-and-h and sinh-arcsinh distributions have interestingly different tail behaviour. For variations on the g-and-h approach, see Rayner & MacGillivray (2002), Fischer, Horn & Klein (2007) and the many references therein. For other transformation forms applied to $g = \phi$, see Shore (2014).

In the different situation of the next subsection, the transformation function W is subject to the extra restriction (6). If (6) is applied to transformation of random variables, as here, quantile-based measures of kurtosis acquire the attractive property of being invariant to skewness (Jones, Rosco & Pewsey, 2011); this applies to sinh-arcsinh transformations, among others.

3.3 Family 3: Transformation of Scale, including Family 3A: Two-Piece

Densities of this family of distributions take the form

$$f_S(x) = 2g(W^{-1}(x)) \tag{5}$$

... like f_R, but without the Jacobian! This third simple form is not, of course, a valid density in general but is a valid density if W satisfies the relationship

$$W(x) - W(-x) = x \tag{6}$$

(Jones, 2014a). Moreover, differentiating each side of (6) with respect to x leads to $w(x)+w(-x) = 1$, that is, to precisely the same requirement (2) as is associated with w in skew-symmetric distributions. (This is no accident; see below.)

The particular beauty of f_A is that it is immediately unimodal if g is; moreover, the mode of f_A is explicitly at $W(0)$. This is a consequence of f_S taking the same values as g in the same order as x increases; W acts as a 'transformation of scale' (only) moving through the values of $g(x)$ at a varying rate. Skewness properties of f_S can be readily interpreted in terms of density-based skewness (Avérous, Fougères & Meste, 1996, Boshnakov, 2007, Critchley & Jones, 2008), as opposed to the quantile-based skewness of van Zwet (1964). See also Fujisawa & Abe (2013) for more on skewness of f_S. In transformation of scale distributions, entropy is invariant to the transformation used. Also, there is a beautiful Khintchine theorem: $X_S \sim f_S$ is unimodal if and only if X_S has the distribution of $UZ + W(Z) - Z$ where U is distributed as $U(0,1)$ and, independently, $Z \sim -zg'(z)$. Other tractability questions concerning f_A (e.g. explicit distribution function?) do not have such positive answers.

Families 1 to 3 are linked in the following way. As $X_R \sim f_R$ arises from $Y \sim g$ through $X_R = W(Y)$, so $X_S \sim f_S$ arises from $X_A \sim f_A$ by the same transformation $X_S = W(X_A)$. That is, transformation of scale distributions arise by transforming the random variable associated with skew-symmetric distributions, provided the W and $w = W'$ functions involved are the same and satisfy (2) and (6).

It may seem presumptious to include transformation of scale distributions in a list of 'major' families of distributions with shape parameters, given that the main paper describing them (Jones, 2014a) is yet to appear (Jones, 2010, contains some preliminary results). A celebrated, and much older, special case, however, justifies their inclusion because it would otherwise have formed, and in practice continues to form, an important family of distributions with shape parameters in its own right ('Family 3A'). These are the two-piece (scale) distributions which are versions of f_S with

$$W_{tp}^{-1}(x) = 2x \left\{ \frac{1}{1-\alpha} I(x < 0) + \frac{1}{1+\alpha} I(x \geq 0) \right\}, \tag{7}$$

$-1 < \alpha < 1$. As well as the unimodality and density skewness properties of more general f_S — indeed, two-piece distributions have constant density skewness — $f_{S;tp}(x) = 2g(W_{tp}^{-1}(x))$ has excellent tractabilities thanks to the simple two-piece nature of W_{tp}^{-1} (and W_{tp} which is of the same form). Less immediately, it has van Zwet skewness interpretability too (Klein & Fischer, 2006). In this case, interpretability translates — uniquely among four-parameter distributions on $\mathbb{R}$, as far as I know — to considerable parameter orthogonality in which the pair of parameters representing location and skewness can be estimated asymptotically independently from the pair of parameters representing scale and tailweight (Jones & Anaya-Izquierdo, 2011). Alternatively, further reparametrization leads to the practically appealing property that location can be estimated asymptotically independently of all the other parameters. When $g = \phi$, two-piece distributions date back to Fechner (1897); see Wallis (2014). Other prominent references include Fernández & Steel (1998), Mudholkar & Hutson

(2000) and Arellano-Valle, Gómez & Quintana (2005).

3.4 Family 4: Probability Integral Transformation of $(0,1)$ *Random Variable*

Let U be a $U(0,1)$ random variable and set $Y = G^{-1}(U)$. Then, $Y \sim g$. This is, of course, the (inverse) probability integral transformation (PIT) which forms the basis of random variate generation. If g is symmetric, shape parameters can be introduced into the distribution of Y by replacing U by a random variable V with a different distribution on $(0,1)$. If the distribution function of V is w then the density of the resulting family of distributions is

$$f_V(x) = g(x)w'(G(x)); \tag{8}$$

the distribution function is $w(G(x))$ and the quantile function is $G^{-1}(w^{-1}(u))$.

A general treatment of this approach which provides bespoke construction of a suitable w so that f_V has certain desired properties is given by Ferreira & Steel (2006). Their specific construction deals brilliantly with the case where skewness is introduced by W while the weights of the tails of f_V are unaffected.

A huge tranche of literature is based on picking a standard distribution for w. The beta distribution is a natural two-parameter choice; this was first published by Eugene, Lee & Famoye (2002) when $g = \phi$ and applied to general g in Jones (2004) — where, admittedly, this view of the family was secondary to one based on order statistics which has since proved less popular. Another popular choice for w is the Kumaraswamy distribution (e.g. Cordeiro & de Castro, 2011). The power law one-parameter special cases of both beta and Kumaraswamy distributions have also been considered, corresponding to Lehmann alternatives (Lehmann, 1953). Alzaatreh, Lee & Famoye (2012) choose w by transforming distributions on $\mathbb{R}^+$. Unfortunately, in general this is an area which has suffered particularly badly from an over-emphasis on writing out formulae for special cases (of both w and g together).

A third set of proposals fits into Family 4 implicitly. This is where authors have taken the distribution function associated with f_V, F_V, to be a function of G. It is easy to see that that function must itself be a distribution function on $(0,1)$. Proportional hazard and proportional reverse hazard families (Marshall & Olkin, 2007, Section 7.E) correspond precisely to the power law w mentioned above. The proportional odds family of Marshall & Olkin (1997) corresponds to $w(v) = w_{PO}(v) = v/\{\gamma + (1-\gamma)v\}$, $0 < v < 1$, $\gamma > 0$. I must stress that, as their names suggest, these families are largely intended to be used in a lifetime distribution context — and work on them by Marshall & Olkin is well motivated and, of course, first class — but they have also been suggested for use as skewing mechanisms on $\mathbb{R}$; Rubio & Steel (2012) show the choice w_{PO} to be poor in this regard in general.

Family 4 is related to Family 2 by reversal of roles. Each takes a transformation function, T, say, and an initial random variable I, say, and is the distribution of $T(I)$. In Family 2, $I \sim g$ and T is 'general'; in Family 4, $T = G^{-1}$ and I is 'general'.

4 Comparison

In this section, I am going to make some general comparisons between Families 1 to 4. This is necessarily a potentially contentious undertaking because many things are not clear-cut and there is inevitably room for disagreement with the topics I've chosen to address (and those I have not) and the 'marks' I've chosen to give them. The relative importance given to each consideration should to some extent be problem-specific, but is also necessarily partly personal.

An overview of the comparisons is given as a ten-point checklist in Table 1. Cells of the table include ticks for 'yes' and crosses for 'no', but also numerous pale ticks for 'partial yes'. Explanatory words and phrases are also often included. The interested reader is recommended to look over the table at this point. Brief further discussions associated with each row of Table 1 are given below.

`Rows 1 and 2.` As discussed under (i) in the Introduction, unimodality is a particular preference of mine. While the main special cases of all families tend to be unimodal (if g is), for Families 1, 2 and 4 they have to be checked on a case-by-case basis and typically don't afford explicit expressions for their modes. Transformation of scale (Family 3) and its two-piece special case (Family 3A) are greatly advantageous in their universal unimodality (if g is) and explicit modes. In general, transformation of random variables and modality properties of densities do not interact well!

`Row 3.` Despite the conceptual and empirical observation that the shape parameter in three-parameter skew-symmetric distributions acts as a skewness parameter, I am not aware of any mathematical sense in which this can be proved. Families 2, 3 and 3A have skewness parameters in the classical quantile-based sense of van Zwet (1964) or, and the reader might think this slightly contrived, in the density asymmetry sense of Critchley & Jones (2008) or, in the case of two-piece distributions, both.

`Row 4.` Families 2 and 4 share full control over tailweights by choice of W/w'. Families 3 and 3A force the weights of tails to be the same (selected through choice of g). If w in Family 1 is a distribution/survival function on the whole of $\mathbb{R}$, tailweights are different; however, weights of tails are the same in, for example, the popular skew-t distribution of Azzalini & Capitanio (2003) where scale mixing leads to w being a distribution function on a finite interval.

Table 1. *Comparison of families 1 to 4.*

	Family 1 *Skew-Symmetric*	**Family 2** *Transformation of Random Variable*	**Family 3** *Transformation of Scale*	**Family 3A** *Two-Piece*	**Family 4** *PIT of* $(0,1)$ *Random Variable*
1. `Unimodal?`	usually	often	✓	✓	often
2. `When unimodal, explicit mode?`	x	x	✓	✓	x
3. `Skewness ordering?`	seems well-behaved	✓ (van Zwet)	✓ (density asymmetry)	✓ ✓ (both)	x
4. `Full control over tails?`	✓	✓	x (must be the same)	x (must be the same)	✓
5. `Straightforward c.d.f?`	x	✓	x	✓	✓
6. `Tractable quantiles?`	x	✓	x	✓	✓
7. `Easy r.v. generation?`	✓	✓	✓	✓	✓
8. `Easy maximum likelihood estimation?`	✓ (“problem” overblown?)	✓	✓	✓	✓
9. `Nice Fisher information matrix?`	x (singularity in one case)	✓ (full FI)	✓ (full FI)	✓ (much parameter orthogonality)	✓ (full FI)
10. `‘‘Physical’’ mechanism(s)?`	✓	perhaps?	x	perhaps?	sometimes

`Rows 5,6 and 7.` If you don't care about tractability, as discussed under (ii) in the Introduction, you won't care about the comparisons made in Rows 5 and 6. The full set of ticks in Row 7 arises from the tractable distribution, quantile and/or transformation functions plus the simple probabilistic mechanisms underlying Family 1, as discussed in Section 3.1.

`Row 8.` At last, to inference! I would like to put a full set of ticks in Row 8 also, in the following sense. These are typically distributions with four parameters plus perhaps one or two more in simple regression situations. Brute force maximum likelihood (ML) using general optimisation methods (such as those in `R`) with multiple starts, possibly combined with careful numerical computation of the log-likelihood itself, will find the global ML estimates almost all the time. This partly complacent view might require modification when several-parameter distributions are embedded in much more complicated models. Wherever ML estimation is feasible, so is Bayes; and the more complex models will probably be given a Bayesian treatment anyway. The pale tick under Family 1 reflects a literature that sees as a problem the tendency of three-parameter skew-symmetric distributions to sometimes have the ML estimate of their skewing parameter $\hat{\lambda}$, say, at $\hat{\lambda} = \pm\infty$. In Table 1 I have asked "problem overblown?" but I really think "problem non-existent". Infinite $\hat{\lambda}$ is a perfectly good ML solution corresponding to a half-g distribution; such arises either because the half-g is a good model for the data or, probably more often, because the skew-symmetric model based on g doesn't fit at all, and the half-g is the best of a bad bunch. Effort put into forcing a finite value of $\hat{\lambda}$ which is so large that the model fitted is almost indistinguishable from half-g anyway seems misplaced.

`Row 9.` For all families, given the ML estimates, observed Fisher information is numerically available, either as a by-product of the optimisation or from (possibly tortuous) explicit differentiation of the log-likelihood. Row 9 concerns the expected Fisher information matrix (FI). "Full FI" means it has no, or almost no, zeroes; as discussed in Section 3.3, the two-piece distribution wins clearly on parameter orthogonality grounds (but that evokes reactions from statisticians ranging from "great" to "I don't care"!). There is a problem with singularity of FI at certain important parameter values for Family 1 that does not appear to occur in other families. But even in Family 1, the problem only occurs for some choices of g and w, depending on their relationship: see Hallin & Ley (2012) for a complete analysis of this (and for earlier references; the univariate case appears to be fully solved in their Section 2).

`Row 10.` If the situation at hand submits to a justifiable modelling process from which emerges a particular family of distributions, use that family, period. As described in Section 3.1, Family 1 scores very well indeed on generating mechanisms, and so has greatest potential for being the 'right' choice on occasions; in particular, underlying truncation and selection mechanisms are cited to justify some uses of skew-symmetric distributions (e.g. Arnold, Beaver, Groeneveld & Meeker, 1993, Copas & Li, 1997, Arnold & Beaver, 2002, Marchenko & Genton, 2012). Some choices of w in Family 4 have order statistic or systems justifications when their parameters take integer values, but their relevance when the distributions are employed with real parameter values is less clear. Generating mechanisms for other families might be contrived e.g. by meaningful choice of transformation? However, the vast majority of situations in which the families of distributions considered in this article are used are purely empirical (and then all the other considerations come back into play). Even when they exist, I find it rare that, with the occasional exception of Family 1, underlying generating mechanisms are even mentioned when it comes to most examples of the application of these families of distributions.

5 On Other Supports

5.1 The Multivariate Case

Natural bi- and multi-variate generalisations exist for each of the univariate families of distributions under consideration. The selection and truncation mechanisms underlying Family 1, skew-symmetric distributions, apply also to the multivariate case. Indeed, many publications on that topic (e.g. Branco & Dey, 2001, Arnold & Beaver, 2002, Azzalini & Capitanio, 2003, Arellano-Valle & Genton, 2005, Arellano-Valle, Branco & Genton, 2006, Kim, 2008) work in the multivariate case because it adds little more than notational complexity. Family 2 can be generalised to the multivariate case by transforming the marginals of suitable multivariate symmetric distributions (e.g. Jones & Pewsey, 2009, Section 7, Field & Genton, 2006). Ditto Family 3 if one starts from a suitable multivariate skew-symmetric distribution (Jones, 2014b). Family 4 requires appropriate distributions on $(0, 1)^d$ (where d denotes dimensionality) which will also be transformed marginally; for example, w might be the extended uniform order statistic distribution in Jones & Larsen (2004) or the bivariate beta distribution of Arnold & Ng (2011), as recently suggested in a talk by J.M. Sarabia.

But, as I have protested elsewhere (Jones, 2014a,b), does 'natural' necessarily translate to 'best' or even 'appropriate'? If the mechanistic derivations of, in particular, Family 1 apply, then, yes, they do. Otherwise, it all depends on what you mean by a multivariate extension of such-and-such a distribution, and I would argue that general methods for constructing multivariate distributions are almost always more appropriate than, and alleviate the need for, bespoke multivariate extensions. For example, I suspect that the desire to model skewness often arises from the shapes of marginal distributions (on the original scale) and so distributions with specified univariate marginals are then what is wanted. Copulas (e.g. Joe, 1997, Nelsen, 2010) provide the obvious (but not only) means for generating such multivariate extensions. Indeed, the natural multivariate extension of Family 2 given above is precisely to choose a copula, the same one as is perhaps implicitly chosen for the underlying symmetric multivariate g. Similarly, their marginal transformation relationship means that multivariate Families 3 and 1 share the same copula. And the scheme above for multivariate Family 4 amounts to choosing a copula too. Or perhaps one wishes to use several-parameter univariate distributions as conditional distributions, either wholly, raising the difficulties of consistent conditional specification (Arnold, Castillo & Sarabia, 1999), or in conjunction with one or more marginals using some of the basic rules of probability (e.g. Arnold, Castillo & Sarabia, 2006).

But then much of multivariate analysis doesn't stick with the original variables at all but works with linear combinations of them. The basic multivariate construction of Family 1 (Branco & Dey, 2001, Azzalini & Capitanio, 2003) works in this way, skewness effectively being introduced along a well-chosen direction in d dimensions. Marginal and conditional distributions on the original scale are secondary. The waters are muddied here a little by the special properties of the multivariate normal distribution implying that univariate marginals of multivariate skew-normal distributions are skew-normal too, but that is not true for other choices of g. Another general construction in the same spirit is to model rotated data by *independent* univariate distributions as in Ferreira & Steel (2007); this is suggested for use with transformation of scale distributions (Family 3) by Fujisawa & Abe (2014).

Multivariate normal and elliptical distributions, useful as they are, give the wrong impression: for most purposes, 'one-lump' multivariate distributions with identical marginals and/or conditionals are not required. (If nothing else, high-dimensional datasets will encompass variables of different types as well as different distributions.) One needs to understand how some variables depend on others and/or how low-dimensional subsets of variables vary jointly (requiring bivariate and perhaps other low-dimensional joint distributions). That is, most multivariate models should be made up of meaningful combinations of low-dimensional marginals and conditionals (many of them univariate). This is well reflected in the copula literature where high-dimensional direct copulas are becoming largely eschewed in favour of pair-copula constructions (e.g. Bedford & Cooke, 2002, Kurowicka & Cooke, 2006, Aas, Czado, Frigessi & Bakken, 2009, Czado, 2010).

5.2 Distributions on Finite and Semi-Finite Intervals

The univariate action in this article has so far wholly occurred on the whole of $\mathbb{R}$.

Families of continuous distributions are, however, just as useful on finite and semi-finite intervals of $\mathbb{R}$, which I will take without loss of generality to be $(-1, 1)$ and $\mathbb{R}^+$, respectively. These are huge topics worthy of review papers and books themselves. Here, I shall just look briefly at their relationship with some of the ideas involved in the families of distributions on $\mathbb{R}$ in this article.

Distributions on $(-1, 1)$ can be dealt with quite quickly because they retain much in common with distributions on $\mathbb{R}$, when symmetry, and lack of it, is a focus. (Like most people, I usually take the canonical finite interval to be $(0, 1)$ but taking $(-1, 1)$ allows me to avoid shifting the centre of symmetry to 1/2.) In fact, Families 1 to 4 can be applied directly provided all the functions involved are redefined to have the appropriate support (and location and scale parameters are not just invisible but nonexistent!). Of these, two-piece distributions (Family 3A) seem to have been used most widely in this context (e.g. Kotz & van Dorp, 2004) .

Symmetry is not possible on $\mathbb{R}^+$, so the main initial raison-d'être for Families 1 to 4 doesn't apply: all distributions on $\mathbb{R}^+$ are skew! Families of distributions on $\mathbb{R}^+$ can, however, still have shape parameters controlling the right-hand tail and the behaviour of the density near 0 (and possibly another in the spirit of Section 6 to follow) in addition to their scale parameter. On $\mathbb{R}^+$ there also already exist numerous effective competitors. This is partly because distributions on $\mathbb{R}^+$ are most often 'lifetime' or 'survival' distributions arising when the axis represents progression in time. Underlying physical mechanisms — stochastic processes — are therefore more directly relevant and common. Even then, extra parameters can be added meaningfully. A masterclass in so doing is given by Marshall & Olkin (2007, Part III). There, extra parameters are interpreted in terms of proportionality of hazard functions or of odds functions, etc. Most, but not all, of these are cases of Family 4 (though not based on symmetric g).

Parameter–free transformations to change supports provide, of course, another approach, but one that has to be used with care: interpretability and particularly modality properties are readily compromised. The mantra 'take logs' (of random variables on $\mathbb{R}^+$) allows one particularly useful family of distributions on $\mathbb{R}^+$, namely the log-location-scale distributions (e.g. Lawless, 2003). (These distributions can be 'log-symmetric', a meaningful alternative to ordinary symmetry, along with 'R-symmetry', Mudholkar & Wang, 2007, on $\mathbb{R}^+$; see Jones, 2008.)

The logarithmic transformation also changes tailweights (that, indeed, is often its very purpose in practice). In particular, distributions on $\mathbb{R}^+$ produced by the exponential transformation of random variables on $\mathbb{R}$ have considerably increased tailweight, so that moderate to heavy-tailed distributions on $\mathbb{R}^+$ arise from light to moderate-tailed distributions (which are not so much our usual focus) on $\mathbb{R}$. An alternative transformation from $Y \in \mathbb{R}^+$ to $X \in \mathbb{R}$ is $X = \frac{1}{2}(Y - 1/Y) = \sinh(\log(Y))$ (Jones, 2007), in which the sinh function 'corrects' the effects on tails of the log function. If $Z \in (0,1)$ (a probability), its basic transformation to $\mathbb{R}^+$ is $Y = Z/(1-Z)$ (odds), so that the X given above is (half) the difference between odds for and against. Interestingly, the odds transformation, which does not affect tails (interpreted as behaviour at a boundary as well as at ∞, Jones, 2007) is often combined with the log transformation, which does (giving the logistic transformation), in an inconsistent manner viewed from this perspective! Odds/log/sinh are the triumvirate at the heart of the classical Johnson (1949) distributions. (And their choice can be given an interpretation in terms of the logistic distribution function, Jones, 2007.) Notice that (non-log cases of) the Box-Cox transformation (from $\mathbb{R}^+$) are inappropriate here because they do not have the correct range.

5.3 Distributions on the Circle

Families of distributions with shape parameters are also appropriate for use on other manifolds. By way of example, one on which a similar program of investigation has recently been carried out, and which is described briefly here, is the circle, $\mathbb{C}$, say. Again, unimodality will be very much to the fore, although on $\mathbb{C}$ "multiplicative mixtures" such as the generalized von Mises distribution (e.g. Gatto & Jammalamadaka, 2007) compete with ordinary mixtures for modelling multimodality.

The ingredients are much the same on $\mathbb{C}$ as they are on $\mathbb{R}$: a symmetric unimodal distribution on $\mathbb{C}$ with density g (e.g. the von Mises or wrapped Cauchy distributions); location and concentration parameters which will often be hidden; and, say, two shape parameters accounting for skewness and, in this case, a different notion of 'kurtosis' (e.g. just peakedness) in the absence of tails per se. Again, those shape parameters will be part of a monotone transformation function $W : \mathbb{C} \to \mathbb{C}$, and its derivative $w = W'$. Interestingly, just two particular cases of W have currency on $\mathbb{C}$. One, first used in different form by Batschelet (1981), is

$$W_B(\theta) = \theta - \nu - \nu \cos\theta; \tag{9}$$

the parameter $-1 \le \nu \le 1$ is a skewness parameter, and the transformation might be used with three-parameter symmetric g (e.g. the symmetric family of Jones & Pewsey, 2005). The second is the famous Möbius transformation which has numerous manifestations but here will be written as a two-parameter 'arctan-tan' transformation,

$$W_M^{-1}(\theta) = \nu + 2\tan^{-1}[\omega \tan\{\tfrac{1}{2}(\theta - \nu)\}], \tag{10}$$

$\nu \in \mathbb{C}$, $\omega \ge 1$.

Family 1 on $\mathbb{C}$ is usually represented by the sine-skewed distributions (Umbach & Jammalamadaka, 2009, Abe & Pewsey, 2011) which correspond to $w(\theta) = \frac{1}{2} w_B(\theta) = \frac{1}{2}(1+\nu \sin\theta)$; this w satisfies (2) but is not monotone. For many choices of g, however, $f_{SS}(x) = 2w_B(x)g(x)$ is not unimodal and there is also, for some g, a problem with the amount and even the sign of the skewness induced by ν: see Web Figure 1b in the supplementary materials associated with Jones & Pewsey (2012). Family 2 on $\mathbb{C}$ is well represented by the Möbius-transformed distributions (Kato & Jones, 2010) which use $W_M(\theta)$ as transformation function. This has a number of nice properties, especially with regard to circular-circular regression, but it too is far from guaranteeing unimodality.

Transformation of scale still yields unimodal distributions on $\mathbb{C}$ provided g is. Batschelet suggested using, essentially, the inverse of W_B in that context, resulting in densities $f_{S1} \propto g(\theta - \nu - \nu \cos\theta)$ (see also Abe, Pewsey & Shimizu, 2013); note the proportionality sign. Jones & Pewsey (2012) suggested "inverse Batschelet distributions" with densities $f_{S2} = g(W_B^{-1}(\theta))$, yes, with an equality sign. The lack of an explicit inverse of (9) is a slight disadvantage but the lack of a change to the normalisation constant is a big plus. The latter comes about here, as on $\mathbb{R}$, because of the link between Family 3 and Family 1: if $\Theta \sim f_{S2}$, then $\Phi = W_B^{-1}(\Theta) \sim f_{SS}$.

The best current family of four-parameter unimodal distributions on $\mathbb{C}$ arises, however, by compromising some desiderata. The family of Kato & Jones (2014) has as a subset a version of f_{S1} using W_M (not f_{S2} because W_M' does not satisfy (2)) and is built on the specific choice of wrapped Cauchy g. Despite these apparent limitations, it has tractability, interpretability and practical advantages even over Batschelet and inverse Batschelet distributions, for which see Kato & Jones (2014).

6 More Parameters?

The basic question here, on $\mathbb{R}$, is whether it can be appropriate to add a third shape parameter to a four-parameter distribution, while retaining unimodality. (Of course, more shape parameters are need in bi- and multi-modal situations.) Recall that the two shape parameters often individually control skewness and kurtosis/tailweight (as in most incarnations of Families 1 to 3) but sometimes control the weights of the two tails, differences between them leading to skewness (e.g. Family 4 using beta w). The next step might be to combine the two: three shape parameters could comprise two parameters controlling the weights of the tails, their difference implying 'tail-skewness', and a third controlling density shape close to its mode, a 'main-body-skewness' parameter, perhaps. Good examples include the extended two-piece t distribution of Zhu & Galbraith (2010) and the use of generalized beta w in Family 4 (Alexander, Cordeiro, Ortega & Sarabia, 2012).

There remain unanswered questions of how best to implement this, and how important it is to do so. On the first point, addition of the fifth parameter (in a natural way) to two-piece distributions destroys parameter orthogonality; is there scope for more interpretable and better estimable alternative extensions/parametrisations? On the second, large samples would seem to be needed, robustness of location estimates would probably be affected little, and would one often really need to model a distribution so carefully? That said, a role for a five-parameter distribution as an overarching model within which one can better choose between competing (tail-skewness and body-skewness) four-parameter models can be envisaged, examples in Rubio & Steel (2013) suggesting that this might be doable effectively.

7 Further Discussion

It might be surprising that an article principally on univariate distributions has no figures, either density plots or fitted models! These abound, of course, in most of the papers under review, to which the reader is referred. Comparative density plots are not produced here because they can only be shown for special cases and, with careful choice of parameter values, densities from different families can be made to look much the same, at least at a superficial level.

No data modelling exercise is provided partly because in the simplest cases such is straightforward (this is a big advantage of these models!). In the same way that details of simple mathematical arguments are omitted from statistical papers, so it seems to me that often details of simple practical applications can be omitted too, where these lead only via standard methods to the expected answers. Most published applications are simple dataset fitting exercises, period. Few instances of more complex applications are yet available except for some uses of Family 1. That said, the potential uses of these models are voluminous. And, as I said in Sections 1 and 4, choices between families should be made on the basis of particular requirements (as outlined in Sections 3 and 4), but remain partly a matter of taste.

The average reader is probably also by this time dismayed not to have found any mention of her/his favourite family of distributions! I will mention just a few more here. The most obvious, perhaps, is the classical Pearson system (e.g. Johnson, Kotz & Balakrishnan, 1994, Section 12.4.1) based on a ratio of polynomials form for the logarithmic derivative of the density and including several well known distributions. Another is Tukey's lambda distributions and generalisations thereof (e.g. Karian & Dudewicz, 2000), which form perhaps the leading family based on the interesting alternative approach of assuming a simple form for the quantile function (e.g. Gilchrist, 2000). Both suffer, however, from their generality in including distributions on different supports depending on the values that happen to be chosen/estimated for their parameters. Better in this respect, to my mind, are methods like that of Johnson (1949; Section 5.2 above) which take different supports into account explicitly.

The approach which came closest to forming Family 5 in this paper is that based on location/scale mixtures. (It was omitted because it seems to me that there has not been sufficient development of the approach in general as yet, but such may well be warranted.) Allowing (symmetric) g to have a scale parameter $s > 0$ following some distribution on $\mathbb{R}^+$ retains symmetry, but typically tailweight is varied, usually becoming heavier. Normal scale mixtures (such as the t distribution) are especially well known and used, and are particularly convenient to extend to the multivariate case; so-called 'slash' distributions and their variants, which specify particular distributions for the scale, are members of this class too. On the other hand, skewness can be introduced by location mixing; such distributions are those of convolutions of g and a mixing distribution on $\mathbb{R}$. An interesting special case is the Azzalini skew-normal distribution: if $g = \phi$ and the mixing distribution is (scaled) half-normal, a scaled version of normal-based f_A arises. Continuity of the mixing distribution is important in the current context; simple discrete distributions for M lead to standard mixture distributions which are typically models for multi-modality.

Natural extensions of these two sets of, typically, three-parameter distributions would seem to be to apply both devices together. This has been done with independent location and scale mixing distributions, but perhaps more prominently, forms of location and scale mixing have been considered based on a *single* mixing variable following a one- or two-parameter distribution. For dimensional reasons, the most natural version of this would seem to me to take, for example in terms of normal g, mean and standard deviation to be proportional to the same random variable, but this seems intractable. Instead, focus in the literature is on taking mean and variance to be proportional to the same random variable. When g is normal and the mixing distribution is the generalized inverse Gaussian distribution this yields the popular generalised hyperbolic distributions. See, for example, Barndorff-Nielsen, Kent & Sørensen (1982), Barndorff-Nielsen & Shephard (2001), Aas & Haff (2006) and Paolella (2007).

As well as the papers already cited that cover Family 1 and extensions thereof, another more general review of families of distributions has recently been published by Lee, Famoye & Alzaatreh (2013).

It is one of the easiest things in Statistics to invent new univariate distributions; after all, any nonnegative integrable function is the core of a density function. The ongoing challenge is to extract from the overwhelming plethora of possibilities those relatively few with the best and most appropriate properties that are of real potential value in practical applications.

References

Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insur. Math. Econ.*, **44**, 182–198.

Aas, K. & Haff, I.H. (2006). The generalised hyperbolic skew Student's t-distribution. *J. Fin. Economet.*, **4**, 275–309.

Abe, T. & Pewsey, A. (2011). Sine-skewed circular distributions. *Statist. Pap.*, **52**, 683–707.

Abe, T., Pewsey, A. & Shimizu, K. (2013). Extending circular distributions through transformation of argument. *Ann. Inst. Statist. Math.*, **65**, 833–858.

Alexander, C., Cordeiro, G.M., Ortega, E.M.M. & Sarabia, J.M. (2012). Generalized beta-generated distributions. *Comput. Statist. Data Anal.*, **56**, 1880–1897.

Alzaatreh, A., Lee, C. & Famoye, F. (2012). A new method for generating families of continuous distributions. *Metron*, **71**, 63–79.

Arellano-Valle, R.B., Branco, M.D. & Genton, M.G. (2006). A unified view on skewed distributions arising from selections. *Canad. J. Statist.*, **34**, 581–601.

Arellano-Valle, R.B. & Genton, M.G. (2005). On fundamental skew distributions. *J. Multivar. Anal.*, **96**, 93–116.

Arellano-Valle, R.B., Gómez, H.W. & Quintana, F.A. (2005). Statistical inference for a general class of asymmetric distributions. *J. Statist. Planning Inference*, **128**, 427–443.

Arnold, B.C. & Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7–54.

Arnold, B.C., Beaver, R.J., Groeneveld, R.A. & Meeker, W.Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471–488.

Arnold, B.C., Castillo, E. & Sarabia, J.M. (1999). *Conditional Specification of Statistical Models*. New York: Springer.

Arnold, B.C., Castillo, E. & Sarabia, J.M. (2006). Families of multivariate distributions involving the Rosenblatt construction. *J. Amer. Statist. Assoc.*, **101**, 1652–1662.

Arnold, B.C. & Ng, H.K.T. (2011). Flexible bivariate beta distributions. *J. Multivar. Anal.*, **102**, 1194–1202.

Avérous, J., Fougères, A.L. & 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. & 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. & Capitanio, A. (2014). *The Skew-Normal and Related Families.* Cambridge: IMS Monographs, Cambridge University Press.

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

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

Barndorff-Nielsen, O., Kent, J. & Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *Internat. Statist. Rev.*, **50**, 145–159.

Barndorff-Nielsen, O.E. & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlen- beck-based models and some of their uses in financial econometrics (with discussion). *J. Roy. Statist. Soc. Ser. B*, **63**, 167–207.

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

Bedford, T. & Cooke, R.M. (2002). Vines — a new graphical model for dependent random variables. *Ann. Statist.*, **30**, 1031–1068.

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

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

Copas, J.B. & Li, H.G. (1997). Inference for non-random samples (with discussion). *J. Roy. Statist. Soc. Ser. B*, **59**, 55–95.

Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *J. Statist. Comput. Simul.*, **81**, 883–898.

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

Czado, C. (2010). Pair-copula constructions of multivariate copulas. *Copula Theory and Its Applications*, Eds. P. Jaworski, F. Durante, W. Härdle & W. Rychlik, pp. 93–109. Dordrecht: Springer.

Eugene, N., Lee, C. & Famoye, F. (2002). Beta-normal distribution and its application. *Commun. Statist. Theory Meth.*, **31**, 497–512.

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

Fernández, C. & 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. & Steel, M.F.J. (2006). A constructive representation of univariate skewed distributions. *J. Amer. Statist. Assoc.*, **101**, 823–829.

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

Field, C. & Genton, M.G. (2006). The multivariate g-and-h distribution. *Technometrics*, **48**, 104–111.

Fischer, M., Horn, A. & Klein, I. (2007). Tukey-type distributions in the context of financial data. *Commun. Statist. Theory Meth.*, **36**, 23–35.

Fujisawa, H. & Abe, T. (2013). A family of skew-unimodal distributions with mode-invariance through transformation of scale. Manuscript.

Fujisawa, H. & Abe, T. (2014). A family of multivariate skew distributions with monotonicity of skewness. Manuscript.

Gatto, R. & Jammalamadaka, S.R. (2007). The generalized von Mises distribution. *Statist. Meth.*, **4**, 341–353.

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

Gilchrist, W.G. (2000). *Statistical Modelling with Quantile Functions*. Boca Raton, FL: Chapman & Hall/CRC.

Hallin, M. & Ley, C. (2012). Skew-symmetric distributions and Fisher information — a tale of two densities. *Bernoulli*, **18**, 747–763.

Hoaglin, D.C. (1985). Summarizing shape numerically; the g-and-h distributions. *Exploring Data, Tables, Trends, and Shapes*, Eds: D.C. Hoaglin, F. Mosteller & J.W. Tukey, pp. 461–511. New York: Wiley.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–176.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, **1**, 2nd ed. New York: Wiley.

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. (2008). On reciprocal symmetry. *J. Statist. Planning 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. (2014a). Generating distributions by transformation of scale. *Statist. Sinica*, to appear.

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

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

Jones, M.C. & Larsen, P.V. (2004). Multivariate distributions with support above the diagonal. *Biometrika*, **91**, 975–986.

Jones, M.C. & Pewsey, A. (2005). A family of symmetric distributions on the circle. *J. Amer. Statist. Assoc.*, **100**, 1422–1428.

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

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

Jones, M.C., Rosco, J.F. & Pewsey, A. (2011). Skewness-invariant measures of kurtosis. *Amer. Statist.*, **65**, 89–95.

Karian, Z.A. & Dudewicz, E.J. (2000). *Fitting Statistical Distributions: the Generalized Lambda Distribution and Generalized Bootstrap Methods.* Boca Raton, FL: CRC Press.

Kato, S. & Jones, M.C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Amer. Statist. Assoc.*, **105**, 249–262.

Kato, S. & Jones, M.C. (2014). A tractable and interpretable four-parameter family of unimodal distributions on the circle. Manuscript.

Kim, H.J. (2008). A class of weighted multivariate normal distributions and its properties. *J. Multivar. Anal.*, **99**, 1758–1771.

Klein, I. & Fischer, M. (2006). Skewness by splitting the scale parameter. *Commun. Statist. Theory Meth.*, **35**, 1159–1171.

Kotz, S. & van Dorp, J.R. (2004). *Beyond Beta; Other Continuous Families of Distributions with Bounded Support and Applications*. New Jersey: World Scientific.

Kurowicka, D. & Cooke, R. (2006). *Uncertainty Analysis With High Dimensional Dependence Modelling*. Chichester: Wiley.

Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd ed. Hoboken, NJ: Wiley.

Lee, C., Famoye, F. & Alzaatreh, A. (2013). Methods for generating families of continuous distributions in the recent decades. *WIREs Comput. Statist.*, **5**, 219–238.

Lehmann, E.L. (1953). The power of rank tests. *Ann. Statist.*, **24**, 23–43.

Marchenko, J.V. & Genton, M.G. (2012). A Heckman selection-t model. *J. Amer. Statist. Assoc.*, **107**, 304–317.

Marshall, A.W. & Olkin, I. (1997). A new method of adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrika*, **84**, 641–652.

Marshall, A.W. & Olkin, I. (2007). *Life Distributions; Structure of Nonparametric, Semiparametric, and Parametric Families*. New York: Springer.

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

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

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

Paolella, M. (2007). *Intermediate Probability: a Computational Approach.* Chichester: Wiley.

Patil, G.P. & Rao, C.R. (1977). Weighted distributions: a survey of their applications. *Applications of Statistics*, Ed: P.R. Krishnaiah, pp. 383–405. Amsterdam: North-Holland.

Rayner, G.D. & MacGillivray, H.L. (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statist. Comput.*, **12**, 57–75.

Rubio, F.J. & Steel, M.F.J. (2012). On the Marshall-Olkin transformation as a skewing mechanism. *Comput. Statist. Data Anal.*, **56**, 2251–2257.

Rubio, F.J. & Steel, M.F.J. (2013). Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations. Research Paper 13–10, Centre for Research in Statistical Methodology, University of Warwick.

Shore, H. (2014). A general model of random variation. *Commun. Statist. Theory Meth.*, to appear.

Umbach, D. & Jammalamadaka, S.R. (2009). Building asymmetry into circular distributions. *Statist. Probab. Lett.*, **79**, 659–663.

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

Wallis, K.F. (2014). The two-piece normal, binormal, or double Gaussian distribution: its origin and rediscoveries. *Statist. Sci.*, to appear.

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

Zhu, D. & Galbraith, J.W. (2010). A generalized asymmetric Student-t distribution with applications to financial econometrics. *J. Economet.*, **157**, 297–305.