THE t FAMILY AND THEIR CLOSE AND DISTANT RELATIONS

M.C. Jones[1]

Abstract

Student's t distribution arose, of course, and has huge application, as a normal-based sampling distribution. In modern times, the t distribution has also found considerable use as a symmetric heavy-tailed distribution for empirical data modelling. In this the centenary year of Student's introduction of the t distribution, this paper constitutes a personal — and sometimes somewhat tangential — tribute to the Student-t family of distributions, by way of an exploration of some of its close and not-so-close relations. I start with a brief reminder of the (symmetric) Student-t family, with particular focus on the t distribution on two degrees of freedom which turns out to play an important role in what follows. The "close relations" of the t distribution include a number of 'skew-t' distributions, of which I shall briefly introduce and comment on three. The "distant relations" of the t distribution are a further trio of four-parameter families of distributions allowing control of skewness and tailweights, with particular emphasis on members of the families with (t-like) 'power' density tails.

AMS 2000 subject classification. Primary 60E05: Secondary 62-02.

Keywords. Exponential tails, heavy-tailed distribution, order statistics, power tails, sinh-arcsinh, skew-t distribution, Student, t_2.

Running head. Extended t-Type Distributions

[1]Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, U.K. (e-mail: m.c.jones@open.ac.uk)

1. Introduction

This paper appears in the centenary year of Student's t distribution. In 1908, one of the seminal papers of modern statistics, Student (1908), was published. In that paper was the first presentation and use of the t distribution. The author, of course, was really W.S. Gosset who published under the pseudonym 'Student' by arrangement with his employers, the Guinness brewing company in Dublin, and it is this pseudonym that has remained attached to the name of the distribution ever since. Gosset actually studied the distribution of the scaled statistic $z = t/\sqrt{n-1}$ (where n is the sample size); in correspondence with Gosset and in his early papers, R.A. Fisher shifted the focus from z to t, "but the choice of the letter "t" to denote the new form was due to "Student"" (Eisenhart, 1979). I am not a historian of statistics, so I will say little more on the history of the t distribution. Because of the centenary, other journals have published historical accounts, two excellent ones already published at the time of writing being Hanley *et al.* (2008) and Zabell (2008).

The t distribution was introduced, of course, as a normal-based sampling distribution (for the standardised sample mean) and it is in that, or closely related, guises that it has, necessarily, seen a very large part of its use and application over the years. However, beside its ubiquitous role as a sampling distribution, in recent years the t distribution — along with many members of its extended family — has also been much used as a purely empirical, heavy-tailed, model for data. Recall that the density of the t distribution on $\nu > 0$ degrees of freedom is proportional to $(\nu + x^2)^{-(\nu+1)/2}$ and that this decreases to zero like $|x|^{-(\nu+1)}$ as $|x| \to \infty$. It is these heavy, 'power', density tails, that are the key to the burgeoning popularity of the t distribution in this sense.

My tribute to the Student-t distribution on its centenary is this exploration of the t family and, in particular, of some of their close and distant relations which are (potentially) useful for empirical statistical modelling. I apologise if, in so doing, the paper later develops too much into a review of some of my own recent work.

N.B. I will remain with the univariate case — which is wider than it sounds because it covers e.g. response distributions in regression — throughout this paper although, of course, there is also much to be said about multivariate t distributions (e.g. Kotz and Nadarajah, 2004). In fact, the world that the t family and virtually all their relations in this paper inhabit is the whole real line, $\mathbb{R}$.

2. The t Family

Mr Normal and Miss Chi-Square were married in 1908 (by Mr Gosset) and, later, changed their family name to "Student-t".

"Marriage" here means division (of independent random variables) or scale-mixing in mathematical/statistical terms! Also, we have already seen that "Student-t" was not their first choice of family name.

They had many children, so many that they are mostly referred to by numbers, t_1, t_2, ..., t_ν, etc.

I'm not really sure whether Normal and Chi-square were very good parents. They referred to their numbered children as degrees of freedom which I assume referred to the amount of freedom they were losing by having such a large family!

In fact, the first of these children, t_1, was much older than the others, an adopted child who more usually went by his earlier name, Cauchy.

Actually Cauchy was already aged 55 when Normal and Chi-Square got married. By this I mean that A.L. Cauchy had written much about the t_1 distribution in 1853. There again, it seems that t_1 might have been called Poisson who wrote about the "Cauchy distribution" in 1824 (some 84 years before the marriage). See Stigler (1974).

Now, while most statisticians think of Cauchy as the 'nicest' of the t family children — in the sense of mathematical tractability — I'd like to put in a word for his little sister, t_2. She has the most delightfully simple density function,

$$g_2(x) = (2 + x^2)^{-3/2},$$

distribution function

$$G_2(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{2 + x^2}}\right)$$

and quantile function

$$Q_2(u) = G_2^{-1}(u) = \frac{2u - 1}{\sqrt{2u(1 - u)}}.$$

The shape of her density is very attractive, inheriting symmetry from her father Normal but displaying tailweights more like those of her brother Cauchy. See Jones (2002) for further details on the t_2 distribution, including her portrait!

For much more on the distribution theory of the Student-t family, see, for example, Chapter 28 of Johnson *et al.* (1994). There is also a non-trivial story to be told regarding inferential issues if the degrees of freedom parameter is to be estimated from data (in addition to location and scale parameters). However, here I will proceed to looking at other members of the extended t family and then at yet others who seem to have pretensions to joining the extended family.

3. Some Skew-t Families

Now, the Student-t family are a pretty old-fashioned, straight-laced, bunch who all follow the path of *symmetry*. But they have many cousins who have renounced symmetry and, daringly, have embraced the religion of *skewness*.

They call themselves the skew-t's.

I shan't list every skew-t distribution that's in the literature here, but concentrate briefly on three of them, each associated with a general approach to skewing symmetric distributions.

3.1. The 'Azzalini' skew-t distributions

The Italian branch of the skew-t family has become the most well known as they run a whole global industry devoted to their approach. This has its genesis in Azzalini (1985). These skew-t distributions are part of an attractive method of generating skewed distributions from symmetric: set

$$f(x; \lambda) = 2g(x)G(\lambda x), \qquad \lambda \in \mathbb{R}, \tag{3.1}$$

where g and G are the density and distribution functions, respectively, of a symmetric distribution. This most famously provides a skew-normal distribution when g is the standard normal density ϕ. The parameter λ controls skewness, positive (negative) λ leading to positive (negative) skewness, $f(x; -\lambda) = f(-x; \lambda)$ and $\lambda = 0$ corresponding to symmetric density g. Numerous extensions exist, replacing $G(\lambda x)$ in (3.1) by a host of other skewing functions.

In the skew-t case, probably the nicest version of this — which arises from the marriage of a Mr Skew-Normal to another of the Chi-Square sisters (i.e. scale-mixing the skew-normal) — has density

$$f(x;\nu,\lambda) = 2g_\nu(x)G_{\nu+1}\left(\lambda_\nu \frac{x}{\sqrt{\nu+x^2}}\right), \tag{3.2}$$

where g_ν and G_ν are the density and distribution functions, respectively, of the t_ν distribution and $\lambda_\nu = \lambda\sqrt{\nu+1}$ (Branco and Dey, 2001, Azzalini and Capitanio, 2003). Of course, this is itself a variation on the theme of (3.1).

Here, I will make just a couple of brief notes of important properties for comparison with other skew-t distributions in this section:

- the most (positive) skew limit of (3.2) (as $\lambda \to \infty$) is the half-t distribution; (3.1.i)
- the left- and right-hand density tails of (3.2) both go as $|x|^{-(\nu+1)}$. (3.1.ii)

Re (3.1.i), I guess the half-t family arose from divorce, remarriage and further offspring of a member of the t family! More seriously, concerning (3.1.ii), observe that the tailweights of this skew-t distribution are the same as those of the symmetric Student-t distribution. This seems to me to be an extra, unheralded, advantage of formulation (3.2) over a more naive skew-t distribution based directly on (3.1) (or, indeed, the skew-normal distribution); in (3.1), the skewness parameter conflates its skewing role with alteration of the weight of one of the two tails in a way that is not well understood.

For entry into the vast literature in this area, see, for example, Genton (2004), Arellano-Valle and Azzalini (2006), Azzalini and Genton (2008), etc., etc., etc. However, if we can escape the clutches of the skew-t 'Statia', we will find that there are plenty of alternative skew-t families out there as well.

3.2. The 'two-piece' skew-t distributions

An alternative branch of the skew-t family comprises 'two-piece' distributions made up of differently scaled halves of a symmetric distribution. In general, one might write

$$f(x;\gamma) = \frac{2\gamma}{1+\gamma^2}\left\{g(\gamma x)I(x<0) + g(x/\gamma)I(x \geq 0)\right\}, \qquad \gamma > 0, \tag{3.3}$$

although various equivalent parametrisations exist (see Jones, 2006a, and references therein). The parameter γ controls skewness, $\gamma >$ ($<$) 1 leading to positive (negative) skewness, $f(x;1/\gamma) = f(-x;\gamma)$ and $\gamma = 1$ corresponding to symmetric density g. The skew-t version of (3.3), corresponding to $g = t_\nu$ and denoted $f(x;\nu,\gamma)$ below, has been nicely treated by Fernández and Steel (1998).

For comparison with (3.1.i) and (3.1.ii) we have:

• the most (positive) skew limit of (3.3) with $g = t_\nu$ (as $\gamma \to \infty$) is the half-t distribution; (3.2.i)

• the left- and right-hand density tails of (3.3) with $g = t_\nu$ both go as $|x|^{-(\nu+1)}$. (3.2.ii)

These are, of course, essentially identical with (3.1.i) and (3.1.ii).

Indeed, the considerable correspondence between (3.2) and (3.3) with $g = t_\nu$, which I shall now write as $k_\nu(\nu + x^2)^{-(\nu+1)/2}$, does not seem to have been noted before (partly because there is no such similar correspondence between (3.1) and (3.3) in general). In particular, since the skewness parameter, γ, in the two-piece skew-t distribution can be interpreted in terms of the ratio of what can be thought of as separate left- and right-hand scale parameters, a similar interpretation can be provided (for the first time) for skew-t distribution (3.2): comparing the ratio of the values of

$$\lim_{x\to-\infty} f(x;\nu,\lambda) = \frac{2k_\nu(1-T_{\nu,\lambda})}{|x|^{(\nu+1)}} \quad \text{and} \quad \lim_{x\to\infty} f(x;\nu,\lambda) = \frac{2k_\nu T_{\nu,\lambda}}{x^{(\nu+1)}},$$

where $T_{\nu,\lambda} = T_{\nu+1}(\lambda_\nu)$, with the ratio of the values of

$$\lim_{x\to-\infty} f(x;\nu,\gamma) = \frac{2k_\nu}{(1+\gamma^2)\gamma^\nu|x|^{(\nu+1)}} \quad \text{and} \quad \lim_{x\to\infty} f(x;\nu,\gamma) = \frac{2k_\nu\gamma^{\nu+2}}{(1+\gamma^2)x^{-(\nu+1)}}$$

yields

$$\gamma \simeq \left(\frac{T_{\nu,\lambda}}{1-T_{\nu,\lambda}}\right)^{1/\{2(\nu+1)\}} \qquad \text{or} \qquad \lambda \simeq \frac{1}{\sqrt{\nu+1}} T_{\nu+1}^{-1}\left(\frac{\gamma^{2(\nu+1)}}{1+\gamma^{2(\nu+1)}}\right).$$

This match-up yields *identical* tail behaviour for the two skew-t distributions.

The skew-t families in Sections 3.1 and 3.2 seem to have more genetic material in common than they realised!

3.3. The Jones and Faddy skew-t distribution

There is also a small Welsh branch of the skew-t family (Jones, not Faddy) currently living over the border in England. (Unfortunately, I'm not a native Welsh speaker, but it seems the appropriate Welsh name for this family *might* be "lletbai-t". Pronunciation of the Welsh "ll" is, according to all the websites, a difficult one to describe. Here's one effort: "The tongue is held in the l position and breath is forced out between the side of the tongue and the upper teeth". Good luck!) Anyway, Jones and Faddy (2003) introduced the skew-t family with density

$$\begin{aligned} f(x;a,b) &= \frac{1}{B(a,b)2^{a+b-1}\sqrt{a+b}} \\ &\times \left(1+\frac{x}{\sqrt{a+b+x^2}}\right)^{a+1/2}\left(1-\frac{x}{\sqrt{a+b+x^2}}\right)^{b+1/2}, \end{aligned} \tag{3.4}$$

$a,b>0$. Here, $B(\cdot,\cdot)$ is the beta function. Note that the Student-t distributions are retrieved whenever $a=b$ (although the parametrisation here is such that the degrees of freedom are then $2a$). Otherwise, $a>\ (<)\ b$ corresponds to positive (negative) skewness. This, too, arises as part of a general methodology for skewing symmetric distributions but I'll leave that until Section 4.1.

For better or for worse, this skew-t distribution behaves rather *differently* from those in Sections 3.1 and 3.2. In particular, this is clear in the following observations:

• the most (positive) skew limit of (3.4) (as $a \to \infty$) is an extreme-value-type distribution (actually the distribution of the reciprocal of the square-root of a gamma random variable with shape parameter b); (3.3.i)

• the left- and right-hand density tails of (3.4) go as $|x|^{-(2a+1)}$ and $x^{-(2b+1)}$, respectively. (3.3.ii)

What I'd like to stress here is the different way in which skewness is controlled in (3.4) compared with (3.2) or (3.3). All three skew-t families necessarily have two shape parameters and at least one of them controls tailweight in the sense of being the power in a power tail. In (3.2) and (3.3) there is just one such parameter yielding the same power in each tail; skewness is introduced/controlled by different *scales* in each tail. In contrast, in (3.4) there is no differential scale parameter, but separate left- and right- *power tail parameters*, skewness arising through differences between such tailweight parameters. The same consideration, of 'pure tail' versus 'relative scale' parameters, permeates what follows in Section 4.

4. Some General Families

In this section, I will leave the (already broad) t/skew-t family temporarily to one side and take a look at some other aspirants to joining the extended t family. In particular, I will look at *three* alternative families of distributions for continuous univariate data on $\mathbb{R}$, power-tailed versions of which will be prime candidates for joining the extended t family. I will (for most of this section) continue to be concerned with models involving four parameters: location, μ, scale, σ, and two shape parameters, a and b, say, accounting for skewness and tailweight *in some way*. The models, therefore, like the skew-t's in empirical work, are all of the form $\sigma^{-1}f(\sigma^{-1}(x-\mu);a,b)$ where the invisible ingredient, along with the visible a and b, is again a simple symmetric 'generating' distribution G with density g, here with no further unspecified parameters. (Here and earlier, for distribution theory work concerning shape parameters, we can set $\mu=0$, $\sigma=1$.)

4.1. Generalised distribution of order statistics

The two frames of Figure 1 of Jones (2004) show a logistic density in the role of g and firstly $n=5$ and secondly $n=35$ other densities generated from g. These densities are the (marginal) densities associated with the n order statistics of g. They exhibit varying amounts of skewness, the greatest amounts being in the distributions of the most extreme order statistics, and, to some extent, variation in tailweight; other symmetric distributions arise as the distributions of sample medians from g.

However, why choose $n=5$ or $n=35$? Why choose n integer? Why stick with integer values $1,2,...,n$ for the order statistics; why not 1.5, or 27.3, or ...? Indeed, why choose one value of n, why not include all n together? The answer is clear ("because you have to"!) if you are considering genuine order statistics of a sample, in which case the density of the ith order statistic from a sample of size n from G is

$$\frac{g(x)G(x)^{i-1}(1-G(x))^{n-i}}{B(i,n+1-i)}.$$

However, if thinking just in terms of providing a family of models for empirical work, the answer is "you don't have to"! Replace i and, most conveniently, $n+1-i$ by positive real parameters a and b and you have

$$f_{OS}(x) = \frac{g(x)G(x)^{a-1}(1-G(x))^{b-1}}{B(a,b)}. \tag{4.1}$$

An alternative way of looking at this first general family of distributions is as the probability integral transformation (PIT) applied to the beta distribution with parameters a and b; that is, as the distribution of $G^{-1}(B)$ where $B \sim \text{Beta}(a,b)$. (The usual PIT involves the Beta(1,1), or uniform, distribution.) Papers continue to be produced on individual "beta-G" distributions despite the general treatment given in Jones (2004).

It is worth reminding ourselves of the roles of a and b:

• a controls the left-hand tailweight (as it happens, parametrised here such that smaller a yields a heavier tail);

• b controls the right-hand tailweight;

• skewness is a spin-off when $a \neq b$;

• $a = b$ yields a symmetric subfamily (with a controlling tailweight).

In this case, we also have that $a = b = 1$ corresponds to $f = g$.

I have two favourite specific examples of this construction. The first is the skew-t distribution of Jones and Faddy (2003) described in Section 3.3. It must be, therefore, that when a and b take integer values, these skew-t distributions arise as the distributions of order statistics from some distribution ... and that distribution, remarkably, is the t distribution on 2 degrees of freedom! Inter alia, this means that Student-t distributions when ν is an even integer are the distributions of the medians of samples of size $\nu-1$ from the t_2 distribution. (The 'beta-t_2' viewpoint gives rise to various interesting transformation relationships for this skew-t distribution, for which see Jones and Faddy, 2003, p.160.)

My other favourite example is a much older one, the superbly simple family of distributions that I will call the log F distribution; it has density

$$f_L(x) = \frac{1}{B(a,b)} \frac{e^{ax}}{(1+e^x)^{a+b}}. \tag{4.2}$$

Apart from a location shift, this is the distribution of the log of an F random variable on $\{2a, 2b\}$ degrees of freedom. It is, of course, a symmetric distribution when $a = b$. This distribution goes back to R.A. Fisher in the 1920s and has a somewhat intermittent history since then; see Jones (2004, 2008a) for information and references. The $a = b = 1$, generating-by-order-statistics, distribution is the logistic.

The tails of the log F distribution decay exponentially: $f_L(x) \sim e^{ax}$ as $x \to -\infty$, $f_L(x) \sim e^{-bx}$ as $x \to \infty$. Thus, a still controls the left-hand tail and b the right-hand tail, and $a \neq b$ is still responsible for asymmetry. But in this case a and b are left- and right-scale parameters, not power parameters. This means that the log F distribution is rather different from any skew-t distribution! (Indeed, that is why I try to promote both the Jones and Faddy skew-t and log F distributions in Jones, 2004.) Please bear with me for some time (until Section 5.2) before drawing this family of distributions closer to the skew-t fold.

4.2. Distributions with simple exponential tails

The reader's indulgence is begged yet further in this section, as I plough on, for now, with distributions which have quite different tail behaviour from that of members of the t family. Indeed, I am going to pursue further the notion of a general family of distributions, each of which has the same 'simple exponential tail' behaviour as the log F distribution:

$$f(x) \sim e^{ax} \text{ as } x \to -\infty, \qquad f(x) \sim e^{-bx} \text{ as } x \to \infty.$$

I call these tails 'simple' because there are no powers or other non-identity functions of x involved either within or without the exponential term.

This simple exponential tail property is shared by at least three 'well known' distributions. One is the log F. Another is the asymmetric Laplace distribution (e.g. Kotz, Kozubowski & Podgórski, 2001); it has density

$$f_{AL}(x) = \frac{ab}{a+b} \exp\left\{axI(x<0) - bxI(x \geq 0)\right\}.$$

I think of this as consisting of *just* the two exponential tails which meet with a discontinuity in derivative at zero. A third distribution which, like the log F, incorporates a smooth joining of the exponential tails is the hyperbolic distribution of Barndorff-Nielsen (1977) with density

$$f_H(x) \propto \exp\left\{\left(\frac{a-b}{2}\right)x - \left(\frac{a+b}{2}\right)\sqrt{1+x^2}\right\}.$$

Here is a general form for a wide family of distributions each with simple exponential tails. Different members of the family are generated, once more, by a simple symmetric distribution G where, this time, we also need its iterated version

$$G^{[2]}(x) = \int_{-\infty}^{x} G(t)dt.$$

(This quantity crops up in various places in the literature such as, for example, in the mean residual life function of reliability theory.) The general form for this density is

$$f_{SET}(x) \propto \exp\left\{ax - (a+b)G^{[2]}(x)\right\} \tag{4.3}$$

(Jones, 2008b). The key to this construction (and to its exponential tails) is that $G^{[2]}(x) \sim 0$ as $x \to -\infty$, $G^{[2]}(x) \sim x$ as $x \to \infty$. It is perhaps less clear that $a = b$, as usual, results in symmetry (although (4.3) is also available in an alternative, 'more symmetric looking', version).

The asymmetric Laplace, log F and hyperbolic distributions correspond to (4.3) with G taken to be the degenerate distribution with point mass at zero, the logistic distribution ... and the t_2 distribution, respectively. So, as in Section 4.1, the logistic and the t_2 generate the most interesting tractable families! Numerous other families exist (Jones, 2008b) including the one based on uniform G which has close links with Huber M-estimation.

Now, the asymmetric Laplace is a three parameter distribution, inclusion of each of σ, a and b leading to non-identifiability. This is not (exactly) the case for other members of family (4.3). However, the fourth parameter in (4.3) is indeed redundant in practice (we have a form of practical near-non-identifiability). In particular, asymptotic correlations between maximum likelihood estimates of σ and either of a and b are very near 1; and I mean "very near", correlations typically being greater than 0.99 (Jones, 2008a,b). The reason is clear: σ, a and b are, in the case of simple exponential tails, all scale parameters, yet only two such parameters are needed to describe the main scale-related aspects of the distribution. These can either be: (i) a left- and a right-scale (a and b, $\sigma = 1$); or (ii) an overall scale and a left-right comparison parameter (e.g. σ and p where $a = 1 - p$, $b = p$).

Let us continue with the second three-parameterisation just suggested. Then, the score equation for the third, unmentioned parameter, μ, satisfies

$$p = \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{\mu - X_i}{\sigma}\right). \tag{4.4}$$

The right-hand side of (4.4) is nothing other than a kernel estimator of the distribution function underlying the data $X_1, ..., X_n$, using kernel G and bandwidth σ and hence:

Formula (4.4), as an equation defining $\hat{\mu}$, is precisely the (inversion) kernel estimator of the pth quantile!

What we have here, then, is a parametric model underlying a kernel quantile estimator; previously it was known only that maximum likelihood estimation of location in the asymmetric Laplace distribution gives rise to the (unsmoothed) sample quantile. And, interestingly, an early reference for kernel quantile estimator (4.4) is none other than Azzalini (1981).

The above has practical consequences, though they are not quite as exciting as one might first hope. First, there is, of course, a second score equation, for the scale parameter σ. This affords a (readily computable) method for the thorny issue of bandwidth selection for the kernel quantile estimator. Unfortunately (Jones, 2008b), this works very well only for the median (and somewhat erratically for other quantiles) when tested in symmetric distributions. There is a good reason for this: (4.4) is something of a 'vehicle model' rather than a true model for the data. For instance, consider estimating the 0.95 quantile. Then, distribution (4.4) will be strongly negatively skewed (decided by the value of p) even if the data exhibit strong positive skewness!

Second, utilisation of (4.4) gives rise to a (slightly) more principled and less ad hoc version of the double kernel local linear method for nonparametric quantile regression (DKLLQR) of Yu and Jones (1998). It turns out to afford a consistently (if not always greatly) improved version of DKLLQR that we recommend as a replacement for the original (Jones and Yu, 2007). Again, this approach helps only a little with bandwidth selection.

4.3. Sinh-arcsinh distributions

These distributions are based on a specific two-parameter transformation of a random variable W, say, which has distribution G:

$$W = S_{a,b}(X) = \sinh(a + b\sinh^{-1}(X)), \tag{4.5}$$

$a \in \mathbb{R}$, $b > 0$. I call this the *sinh-arcsinh transformation*. It gives rise to the sinh-arcsinh family of distributions for X with density

$$f_{SAS}(x) = b\sqrt{\frac{1 + S_{a,b}^2(x)}{1 + x^2}}\, g(S_{a,b}(x)). \tag{4.6}$$

(Note that $\sqrt{1+S_{a,b}^2(x)} = \cosh(a + b\sinh^{-1}(x))$.) In particular, in our work we have focussed on the normal-based special case, $g = \phi$, in (4.6). This family exhibits considerable tractability and practicability; see Jones and Pewsey (2008) for the full and original description of this work.

The big 'selling point' here is that the sinh-arcsinh family accommodates both heavier and lighter tails than the normal (or, more generally, g). This allows likelihood ratio testing of normality against alternatives with (possible asymmetry and) both types of tail behaviour (see Jones and Pewsey, 2008, where testing for symmetry is also studied). This normality test is very successful as a general tool, proving to be more powerful in most cases than competing omnibus tests of normality.

Let us consider the symmetric subfamily of (4.6) with $g = \phi$ briefly; this corresponds to $a = 0$. Then, $b = 1$ gives the normal distribution; small $b > 0$ gives heavier tailed distributions with something akin to what are sometimes called 'semi-heavy' tails ($f_{SAS}(|x|) \sim |x|^{b-1}\exp(-|x|^{2b})$ as $|x| \to \infty$); and large b gives lighter tailed distributions dominated by their somewhat 'uniform-like' main bodies.

Especially appealing is the following behaviour when $a = 0$ (Jones and Pewsey, 2008, Section 3.3). Asymptotically, when b is small, f_{SAS} behaves like Johnson's (1949) S_U density, which is the distribution of $X = \sinh(Z/b)$ where Z is a standard normal random variable; as $b \to \infty$, f_{SAS} behaves like a special case of Rieck and Nedelman's (1991) sinh-normal density, the distribution of $X = \sinh^{-1}(Z)$. The sinh-arcsinh transformation allows a 'seamless' transition from heavy-tailed, essentially S_U, distributions to light-tailed, essentially sinh-normal, distributions via the normal, all controlled by the single parameter b. And the sinh-arcsinh distributions are all unimodal, automatically excluding the very lightest-tailed, bimodal, sinh-normal distributions.

Another interesting subclass of (4.6) with $g = \phi$ corresponds to $b = 1$, $a \neq 0$. These are truly skew-normal distributions in the sense of being skew and having normal tails, as is also the case with two-piece normal distributions but not Azzalini's (1985) skew-normal distribution. The alert reader will have noticed that a and b are not both tail parameters here, but there is an alternative (similarly behaving but slightly less tractable) version of the sinh-arcsinh transformation in which they are.

5. Who Can Join the Extended t Family?

So, how many members of the other wide families of distributions that I've described in Section 4 are really members of the extended (skew-)t family, in the sense of allowing skewness and having power tails to their densities?

5.1. Which generalised distributions of order statistics?

Any family of the form (4.1) will have power tails if the generating distribution G itself has power tails. Of course, the Jones and Faddy (2003) skew-t distribution with density (3.4) is (already!) a member of the extended t family but there are other aspirants like the family based on Cauchy g. But the link with t_2 means that the Jones and Faddy family is much the nicest (i.e. most tractable).

5.2. Which distributions with simple exponential tails?

None, of course!

However, one can transform distributions of the form (4.3) to have power tails. Let E follow (4.3). Then any transformation of E whose absolute value increases exponentially as $x \to \infty$ will do the trick. Probably the neatest, but not the only, example is the following version of the sinh transformation: $X = \sinh(E/2)$. (For brief consideration of $X = \sinh(E)$ — which is a particular sinh-arcsinh transformation of the above! — see Section 7.2 of Jones, 2006b.) This results in a family of densities of the form

$$f_{TSET}(x) \propto \frac{\left(x+\sqrt{1+x^2}\right)^{2a}}{\sqrt{1+x^2}} \exp\left\{-(a+b)G^{[2]}(2\sinh^{-1}(x))\right\}.$$

The tails of these densities go as $|x|^{-(2a+1)}$ as $x \to -\infty$ and $x^{-(2b+1)}$ as $x \to \infty$.

Again, the most immediate special case of this construction is the Jones and Faddy skew-t distribution! (But there are many others.) This arises because of the fundamental link between the logistic and t_2 distributions (buried in Section 6.2 of Jones, 2004) that drives the conspicuous presence of t_2 in so much of the above:

If $L \sim$ logistic and $T_2 \sim t_2$, then the two are connected by the sinh transformation

$$T_2 = \sqrt{2}\sinh(L/2).$$

5.3. Which sinh-arcsinh distributions?

The answer is essentially the same as in Section 5.1: any family of the form (4.6) will have power tails if the generating distribution G itself has power tails. There is a nice example based on g being the t_2 distribution. But there's also one based on the Cauchy which, for once, is possibly nicer; its density is given by

$$f_C(x) = \frac{b}{\pi\sqrt{1+x^2}\sqrt{1+S^2_{a,b}(x)}}.$$

6. Conclusions

I hope you have enjoyed this rapid tour of the extended (and growing!) family of t-related distributions, some important points for the present and future being:

• the Student-t distribution can itself be, and is, used as a symmetric empirical model for heavy-tailed data;

• there are many ways of producing four-parameter distributions of skew-t type, particularly if one allows further broadening to include power tails in general. Production of new families is the easy bit; much more attention needs to be paid to the comparative advantages and disadvantages of each, and to determining the *most useful* such families;

• there remain a lot of inferential challenges (which have mostly only been taken up at all as yet in the context of the 'Azzalini' skew-t family);

and last, but not least,

• *four-parameter families of distributions like these should have a much more central role in statistics*. This is because they provide automatic robustness to both heavy tails and skewness while retaining the benefits of parsimony; in the memorable phrase of my colleague Frank Critchley, they are "as flexible as is estimable". (One caveat: I've been considering unimodal distributions throughout, multimodality generally requiring mixtures of appropriate unimodal components.)

I wonder what Gosset would have made of the use of the t distribution as an empirical model some way from its Gaussian roots, and of the globalisation of the extended (skew-)t family at least by theorists if not as yet quite as widely by practitioners.

References

Arellano-Valle, R. B. and Azzalini, A. (2006). "On the unification of families of skew-normal distributions", *Scandinavian Journal of Statistics*, **33**, 561–574.

Azzalini, A. (1981). "A note on the estimation of a distribution function and quantiles by a kernel method", *Biometrika*, **68**, 326–328.

Azzalini, A. (1985). "A class of distributions which includes the normal ones", *Scandinavian Journal of Statistics*, **12**, 171–178.

Azzalini, A. and Capitanio, A. (2003). "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution", *Journal of the Royal Statistical Society, Series B*, **65**, 367–389.

Azzalini, A. and Genton, M. G. (2008). "Robust likelihood methods based on the skew-t and related distributions", *International Statistical Review*, **76**, 106–129.

Barndorff-Nielsen, O. (1977). "Exponentially decreasing distributions for the logarithm of particle size", *Proceedings of the Royal Society of London, Series A*, **353**, 401–419.

Branco, M. D. and Dey, D. K. (2001). "A general class of multivariate skew-elliptical distributions", *Journal of Multivariate Analysis*, **79**, 99–113.

Eisenhart, C. (1979). "On the transition from Student's z to Student's t", *The American Statistician*, **33**, 6–10.

Fernández, C. and Steel, M.F.J. (1998). "On Bayesian modeling of fat tails and skewness", *Journal of the American Statistical Association*, **93**, 359–371.

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

Hanley, J. A., Julien, M. and Moodie, E. E. M. (2008). "Student's z, t, and s: what if Gosset had R?", *American Statistician*, **62**, 64–69.

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

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 2, Second Edition, Wiley, New York.

JONES, M. C. (2002). "Student's simplest distribution", *Statistician*, **51**, 41–49.

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

JONES, M. C. (2006a). "A note on rescalings, reparametrizations and classes of distributions", *Journal of Statistical Planning and Inference*, **136**, 3730–3733.

JONES, M. C. (2006b). "On a class of distributions with simple exponential tails", Open University Department of Statistics Technical Report 06/02 available from `http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`

JONES, M. C. (2008a). "The logistic and the log F distribution", Chapter 10 in *Handbook of the Logistic Distribution*, Second Edition (Baghdachi, J. and Balakrishnan N., eds.), to appear. Version available as Open University Department of Statistics Technical Report 06/01 from `http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`

JONES, M. C. (2008b). "On a class of distributions with simple exponential tails", *Statistica Sinica,* to appear. Version with additional material available as Jones (2006b).

JONES, M. C. AND FADDY, M. J. (2003). "A skew extension of the t-distribution, with applications", *Journal of the Royal Statistical Society, Series B*, **65**, 159–174.

JONES, M. C. AND PEWSEY, A. (2008). "Sinh-arcsinh distributions: a broad family giving rise to powerful tests of normality and symmetry", in preparation. Version will be available very soon as an Open University Technical Report.

JONES, M. C. AND YU, K. (2007). "Improved double kernel local linear regression", *Statistical Modelling*, **7**, 377–389.

KOTZ, S., KOZUBOWSKI, T. J. AND PODGÓRSKI, K. (2001). *The Laplace Distribution and Generalizations; A Revisit With Applications to Communications, Economics, Engineering, and Finance*, Birkhauser, Boston.

KOTZ, S. AND NADARAJAH, S. (2004). *Multivariate t-Distributions and Their Applications*, Cambridge University Press, Cambridge.

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

Stigler, S. (1974). “Cauchy and the witch of Agnesi: an historical note on the Cauchy distribution”, *Biometrika*, **61**, 375–380.

Student (1908). “The probable error of a mean”, *Biometrika*, **6**, 1–25.

Yu, K. and Jones, M. C. (1998). “Local linear quantile regression”, *Journal of the American Statistical Association*, **93**, 228–237.

Zabell, S. A. (2008). “On Student’s 1908 article “The probable error of a mean”” (with discussion), *Journal of the American Statistical Association*, **103**, 1–20.