On a Class of Distributions
with Simple Exponential Tails

By M.C. JONES

Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, U.K.

m.c.jones@open.ac.uk

Summary

A simple general construction is put forward which covers many unimodal univariate distributions with simple exponentially decaying tails (e.g. asymmetric Laplace, log F and hyperbolic distributions as well as many new models). The proposed family is a special subset of a regular exponential family, and many properties flow therefrom. Two main practical points are made in the context of maximum likelihood fitting of these distributions to data. The first of these is that three, rather than an apparent four, parameters of the distributions suffice. The second is that maximum likelihood estimation of location in the new distributions is precisely equivalent to a standard form of kernel quantile estimation, choice of kernel being equivalent to specific choice of model within the class. This leads to a maximum likelihood method for bandwidth selection in kernel quantile estimation, but its practical performance is shown to be somewhat mixed. Further distribution theoretical aspects are also pursued, particularly distributions related to the main construction as special cases, limiting cases or by simple transformation.

Some key words: Asymmetric Laplace distribution; Bandwidth selection; Exponential family; Hyperbolic distribution; Kernel quantile estimation; Log F distribution; Maximum likelihood.

1. Introduction

A continuous univariate distribution on $\mathcal{R}$ has simple exponential tails if its density f has the properties

$$f(x) \sim e^{\alpha x} \text{ as } x \to -\infty, \qquad f(x) \sim e^{-\beta x} \text{ as } x \to \infty, \tag{1}$$

for some $\alpha, \beta > 0$. The archetypal example of such a distribution is the asymmetric Laplace distribution given by

$$f_{AL}(x) = \frac{\alpha\beta}{\alpha + \beta} \exp\left\{\alpha x I(x < 0) - \beta x I(x \geq 0)\right\}. \tag{2}$$

For an excellent treatment of this distribution, see Kotz, Kozubowski and Podgórski (2001, Chapter 3). It is particularly simple, its one drawback, to some, being its 'pointed' nature at $x = 0$ and its non-differentiability there.

Which other distributions share the property of simple exponential tails? I could think of two before starting this work — the log F and hyperbolic distributions — and they will feature below. (Their properties include much smoother behaviour than the asymmetric Laplace around $x = 0$.) The purpose of this paper is to present a simple general construction involving the two parameters $\alpha, \beta > 0$ which affords a wide variety of distributions with tail behaviour (1) (and of which the asymmetric Laplace, log F and hyperbolic distributions remain probably the most important). The family proposed will be a special subset of a regular exponential family.

Input to this construction will simply be one's favourite (simple, symmetric) distribution which has random variable X_G, density g, distribution function G and first iterated (left-tail) distribution function $G^{[2]}(x) = \int_{-\infty}^{x} G(t) dt$ which is $G(x)$ times the mean residual life function (e.g. Bassan, Denuit and Scarsini, 1999, and references therein). The latter does not exist if $g(x)$ goes as $|x|^{-(\gamma+1)}$ for $0 < \gamma \leq 1$ as $x \to -\infty$, so any such (very heavy tailed) distributions — 'Cauchy and heavier' — are disqualified from consideration. Then, $G^{[2]}(x) = E\{(x - X_G)I(X_G < x)\}$. Taking g to be symmetric (about zero) is a convenience that affords particularly elegant simplifications without losing importantly in generality and which will be followed virtually throughout this paper.

The main construction and numerous basic properties are given in Section 2. A variety of special cases are considered in Section 3. Distributions linked to the main construction as limiting cases are derived in Section 4. In

2

Sections 5 and 6, two major practical points are made in the context of maximum likelihood fitting of the new distributions to data. The first of these (Section 5) explores whether the new construction really needs all four of its parameters in practice. The answer is negative: three parameters suffice. The second of these sections (Section 6) observes that maximum likelihood estimation of location in the new distributions is precisely equivalent to a standard form of kernel quantile estimation. (This was a major motivation for the current work: specific choice of kernel is equivalent to specific choice of model within the class.) This leads to a maximum likelihood method for bandwidth selection in kernel quantile estimation, but its practical performance is shown to be somewhat mixed. Finally, a number of distributions related to our main construction through simple transformations are explored in Section 7.

2 GENERAL CONSTRUCTION AND PROPERTIES

The proposed general family of distributions with simple exponential tails has density

$$f_G(x) = \mathcal{K}_G^{-1}(\alpha, \beta) \exp\{\alpha x - (\alpha + \beta)G^{[2]}(x)\}. \tag{3}$$

It is clear that as $x \to -\infty$, $G^{[2]}(x) \to 0$ and that — the real key to the construction — as $x \to \infty$, $G^{[2]}(x) \sim x$. That f_G has simple exponential tails as at (1) is thus clear. Note that this holds regardless of the weight of the tails of allowed G.

The exponential tails also ensure integrability of f_G so that the claim that it defines a density is confirmed, albeit one for which the normalisation constant $\mathcal{K}_G(\alpha, \beta)$ is not available in closed form in general. Likewise, the exponential tails imply the existence of all moments of the distribution, but their explicit formulae are also available only on a case-by-case basis. These comments are reflected in the moment generating function associated with (3) which, for $-\alpha < t < \beta$, is immediately seen to take the form $\mathcal{K}_G(\alpha + t, \beta - t)/\mathcal{K}_G(\alpha, \beta)$. Similarly, the characteristic function is $\mathcal{K}_G(\alpha + it, \beta - it)/\mathcal{K}_G(\alpha, \beta)$. Define $\mathcal{K}_G^{ij}(\alpha, \beta) = \partial^{i+j}\mathcal{K}_G(\alpha, \beta)/\partial\alpha^i\partial\beta^j$. Then, inter alia, the mean of distribution (3) is $\{\mathcal{K}_G^{10}(\alpha, \beta) - \mathcal{K}_G^{01}(\alpha, \beta)\}/\mathcal{K}_G(\alpha, \beta)$.

Densities (3) are, immediately, unimodal for all $\alpha, \beta > 0$ with mode x_0 satisfying $G(x_0) = \alpha/(\alpha+\beta)$ i.e. the mode of f_G is at the $\alpha/(\alpha+\beta)$'th quantile

3

of G. Moreover, densities (3) are all log-concave, i.e. strongly unimodal, in x. Let X_{F_G} follow the distribution with density f_G. It is also the case that $E(G(X_{F_G})) = \alpha/(\alpha + \beta)$.

For symmetric g, two apparent alternative formulations turn out to be essentially the same as (3). Let $\bar{G}^{[2]}(x) = \int_x^\infty \{1 - G(t)\}dt = E\{(X_G - x)I(X_G > x)\}$ be the first iterated right-tail distribution function; it is easy to see that for symmetric g, $\bar{G}^{[2]}(x) = G^{[2]}(-x)$. First, one might consider the density proportional to $\exp\{-\beta x - (\alpha + \beta)\bar{G}^{[2]}(x)\}$ but this is just the distribution of $-X_{F_G}$ with the roles of α and β swopped. Second, one might consider the more symmetric formulation in which the density is proportional to

$$\exp\{-\alpha\bar{G}^{[2]}(x) - \beta G^{[2]}(x)\}, \tag{4}$$

but this turns out to be nothing other than density (3) again. This is because, for symmetric g,

$$G^{[2]}(x) - \bar{G}^{[2]}(x) = E(x - X_G) = x.$$

Formulation (4), in particular, makes it immediately clear that f_G is symmetric (about zero) if and only if $\alpha = \beta$ (for symmetric g). Indeed, in that case, symmetric densities are proportional to the α'th power of density (3) with $\alpha = \beta = 1$.

3. Special cases

3·1. The asymmetric Laplace distribution

The asymmetric Laplace density (2) is the very special case of density (3) when G corresponds to a point mass at zero: $G(x) = I(x \geq 0)$, $G^{[2]}(x) = xI(x \geq 0)$.

3·2. The log F distribution

Now let G be the logistic distribution so that $G(x) = e^x/(1 + e^x)$ and $G^{[2]}(x) = \log(1 + e^x)$. It follows that the resulting density

$$f_{LF}(x) \propto \frac{e^{\alpha x}}{(1 + e^x)^{\alpha + \beta}}$$

4

and it can readily be calculated that $\mathcal{K}_{LF}(\alpha, \beta) = B(\alpha, \beta)$ where $B(\cdot, \cdot)$ is the beta function. This is none other than the log F distribution which dates back to R.A. Fisher (as the z distribution) and which has appeared from time to time and in a variety of guises in the literature since then. For a partial review and references, see Jones (2006a).

The logistic distribution also 'generates' the log F distribution in the following way. The ith order statistic of an i.i.d. sample of size n from the logistic distribution follows the log F distribution with $\alpha = i$, $\beta = n + 1 - i$. Moreover, in Jones (2004), I argue that replacing the integers i and n by a pair of real parameters provides a general method for generating distributions with two extra shape parameters from a simple initial distribution.

3·3. The hyperbolic distribution

Now let G be the (scaled) t_2 distribution (the Student t distribution on two degrees of freedom) such that $G(x) = (1/2)(1 + x/\sqrt{1 + x^2})$ and $G^{[2]}(x) = (1/2)(x + \sqrt{1 + x^2})$. The resulting density is that of the hyperbolic distribution of Barndorff-Nielsen (1977), see also Barndorff-Nielsen and Blaesild (1983):

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

It turns out that $\mathcal{K}_H(\alpha, \beta) = (\alpha + \beta) K_1(\sqrt{\alpha\beta})/\sqrt{\alpha\beta}$ where $K_1(\cdot)$ is a Bessel function. This parametrisation is not, perhaps, the most usual one which takes as parameters $\pi = (\alpha - \beta)/2\sqrt{\alpha\beta}$ and $\xi = \sqrt{\alpha\beta}$ (Barndorff-Nielsen and Blaesild, 1983), but it is one of the alternative forms listed by those authors. Of course, consideration of $\log f_H$ and its hyperbolic form makes the hyperbolic distribution an especially natural member of the class of distributions with simple exponential tails from the viewpoint of linear asymptotes for the log density.

In Jones (2004), I argued that the two most tractable and useful order statistic distributions were the log F distribution, generated by the logistic, and the skew t distribution of Jones and Faddy (2003), generated via the order statistics of the t_2 distribution. In this paper, I find myself suggesting that the two most obviously tractable and useful (smooth) alternatives (with exponential tails) to the asymmetric Laplace distribution are, again, the log F distribution, generated in an alternative fashion by the logistic,

together with a rather different distribution, the hyperbolic distribution, but one which turns out also to be generated by the t_2 distribution. I find the place of the t_2 distribution at the heart of this kind of distribution theory intriguing, even more so now than when I wrote extolling the simple virtues of the t_2 distribution in Jones (2002).

3·4. The doubly double exponential distribution

One can actually take g to be a Laplace distribution in which case the following interesting new distribution arises:

$$f_{DDE}(x) = \mathcal{K}_{DDE}^{-1}(\alpha, \beta) \begin{cases} \exp(\alpha x - ce^x) & \text{if } x < 0, \\ \exp(-\beta x - ce^{-x}) & \text{if } x \geq 0, \end{cases} \tag{5}$$

where $c = (\alpha + \beta)/2$, $\mathcal{K}_{DDE}(\alpha, \beta) = c^{-\alpha}\Gamma_c(\alpha) + c^{-\beta}\Gamma_c(\beta)$ and $\Gamma_c(d) = \int_0^c z^{d-1}e^{-z}dz$ is the incomplete gamma function.

In the case where α is an integer, the distribution with density of the form $\exp(\alpha x - ce^x)$, $x \in \mathcal{R}$, is the asymptotic distribution of the α'th largest order statistic of an i.i.d. sample from a distribution with exponential tails (Gumbel, 1958), shifted in location by an amount depending on c. Density (5) consists, therefore, of splicing together a Gumbel extreme value distribution with parameter α and a negative Gumbel extreme value distribution with parameter β (appropriately located). Density (5) is differentiable everywhere, non-differentiability of g translating to lack of a *second* continuous derivative of $f_{DDE}(x)$ at $x = 0$. Note that the mode of (5) is at $x_0 = \log(\alpha/c)I(\alpha < \beta) - \log(\beta/c)I(\alpha \geq \beta)$, not 0. Both Laplace and Gumbel distributions are sometimes known as double exponential distributions, so with what can be conceived to be dual use of both such distributions, the doubly double exponential distribution seems a good name!

3·5. Other smooth distributions

Further smooth f's arise from further smooth distributions G with support the whole of $\mathcal{R}$, but none seems more attractive than, or as tractable as, those already considered. One example that it is natural to consider is the normal-based distribution with

$$f_N(x) \propto \exp\left[\alpha x - (\alpha + \beta)\{x\Phi(x) + \phi(x)\}\right]$$

where ϕ and Φ are the standard normal density and distribution functions.

Differences between (smooth) densities are, in any case, not very marked. See Fig. 1 for three sets of log F, hyperbolic and normal-based densities (the latter computed numerically), all normalised to have unit variance and means strategically placed along the line to allow clarity of viewing. What differences there are between densities show up mainly in the centres of the distributions.

$$* * *\quad \text{Fig. 1 about here}\quad * * *$$

My colleague Karim Anaya made the excellent observation that $H^{[2]}(x) \equiv xH(x) + h(x)$ has the properties of a first iterated left-tail distribution function, provided again that the otherwise arbitrary density function h (and distribution function H) are such that h is not 'very heavy left-tailed' in the sense described in Section 1. The distribution function associated with $H^{[2]}$ is, of course, not H but $H(x) + xh(x) + h'(x)$, which differs from H except in the case $H = \Phi$. However, this method of construction of $G^{[2]}$'s tends to add complication and so no examples will be pursued.

3·6. Three-piece distributions

The asymmetric Laplace distribution is a two-piece distribution in the sense that its density can be thought of as being made up of two smooth parts joined together continuously but, in this case, not differentiably. If I drop the requirement that G be symmetric and employ instead a distribution on $\mathcal{R}^+$, further two-piece distributions ensue, but they will not be considered here.

Instead, consider G to be a symmetric distribution on finite support (which, without loss of generality, I shall take to be $(-1, 1)$). These result in three-piece distributions. The simplest case is that G be uniform so that $G(x) = (1/2)(1 + x)I(-1 < x < 1) + I(x \geq 1)$. It follows that $G^{[2]}(x) = (1/4)(1 + x)^2 I(-1 < x < 1) + xI(x \geq 1)$ and thence that

$$f_U(x) \propto \begin{cases} \exp(\alpha x) & \text{if } x < -1, \\ \exp\left(-\frac{\alpha\beta}{\alpha+\beta}\right) \exp\left\{-(1/4)(\alpha + \beta)\left(x - \frac{\alpha-\beta}{\alpha+\beta}\right)^2\right\} & \text{if } -1 \leq x < 1, \\ \exp(-\beta x) & \text{if } x \geq 1. \end{cases} \quad (6)$$

It can readily be evaluated that

$$\mathcal{K}_U(\alpha, \beta) = \frac{1}{\alpha}e^{-\alpha} + \frac{1}{\beta}e^{-\beta}$$

$$+ \ 2\sqrt{\frac{\pi}{\alpha + \beta}} \exp\left\{-\frac{\alpha\beta}{\alpha + \beta}\right\} \left\{\Phi\left(\beta\sqrt{\frac{2}{\alpha + \beta}}\right) - \Phi\left(-\alpha\sqrt{\frac{2}{\alpha + \beta}}\right)\right\}.$$

Density (6) is the result of the very simple piecewise method of continuously — but not differentiably — joining two lines and a quadratic centre on the log density scale. Equivalently, it consists of a normal centre on to which exponential, rather than normal, tails have been grafted. In the symmetric case with $\alpha = \beta$, (6) is the density associated with the 'most robust' M-estimator of Huber (1964, p.75; rescale (6) by factor k and take $\alpha = k^2$ to match Huber's parameterisation).

Higher order contact between pieces can be achieved by replacing the quadratic by a higher order polynomial by, for example, replacing uniform g by other symmetric beta $g(x) \propto (1 - x^2)^{m-1}I(-1 < x < 1)$ for integer $m > 1$. See also Section 6.

4. Related distributions I: limiting cases

For the purposes of this section, consider

$$\frac{1}{\sigma}f_G\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\mathcal{K}_G(\alpha, \beta)} \exp\left\{\alpha\frac{(x - \mu)}{\sigma} - (\alpha + \beta)G^{[2]}\left(\frac{x - \mu}{\sigma}\right)\right\} \quad (7)$$

with (symmetric) G not being a degenerate distribution. It turns out that one can take $\mu = 0$ in Sections 4.1 and 4.2.

4·1. $\alpha, \beta \to 0$

Immediately, the asymmetric Laplace is the limiting form of (7) obtained by letting σ tend to zero. This normalisation is, clearly, appropriate for the situation where $\alpha, \beta \to 0$. In particular, in the symmetric case of $\alpha = \beta$ with limiting (symmetric) Laplace distribution, one can take $\sigma = \alpha$.

From (7) with $\alpha = \beta$,

$$
\begin{aligned}
\mathcal{K}_G(\alpha, \alpha) &= \int_{-\infty}^{\infty} \exp\{\alpha(x - 2G^{[2]}(x))\}dx \\
&= \exp(-2\alpha G^{[2]}(0)) \int_{-\infty}^{\infty} \exp[\alpha\{x - 2(G^{[2]}(x) - G^{[2]}(0))\}]dx \\
&\simeq \exp(-2\alpha G^{[2]}(0)) \int_{-\infty}^{\infty} \exp\{-\alpha x^2 g(0)\}dx \\
&= \exp(-2\alpha G^{[2]}(0))\sqrt{\frac{\pi}{\alpha g(0)}},
\end{aligned}
$$

the Taylor approximation being justified by the integrand in the second line being 1 for $x = 0$ and vanishingly small otherwise. (Note that $2G^{[2]}(0) = E(|X_G|)$ which has already implicitly been assumed to exist.) Now take $\sigma = \sqrt{2\alpha g(0)}$. Then,

$$
\begin{aligned}
-\log \sigma - \log f_G\left(\frac{x}{\sigma}\right) &\simeq -\frac{1}{2}\log(2\pi) + \sqrt{\frac{\alpha}{2g(0)}}x \\
&\qquad - 2\alpha\left\{G^{[2]}\left(\frac{x}{\sqrt{2\alpha g(0)}}\right) - G^{[2]}(0)\right\} \\
&\simeq -\frac{1}{2}\log(2\pi) - \frac{1}{2}x^2
\end{aligned}
$$

and the standard normal distribution ensues.

Define $\ell(x) > 0$ to be the limiting expression for $G^{[2]}(x) - x$ as $x \to \infty$. Take μ and σ large such that $\alpha\ell((x-\mu)/\sigma) \sim \ell_1(x) > 0$ for large x and some function ℓ_1. Then, the exponential part of (7) is

$$
\exp\left[-\beta\frac{(x-\mu)}{\sigma} - (\alpha + \beta)\left\{G^{[2]}\left(\frac{(x-\mu)}{\sigma}\right) - \frac{(x-\mu)}{\sigma}\right\}\right]
$$

and this affords a limiting density of the form

$$
\mathcal{K}(\beta)^{-1}\exp\{-\beta x - \ell_1(x)\} \tag{8}
$$

9

on appropriate support.

Formula (8) works for the asymmetric Laplace distribution because then $\ell_1(x) = 0$ and the limiting case as the left-hand tail parameter $\alpha \to \infty$ is, of course, the exponential distribution on $\mathcal{R}^+$. (Ditto for all distributions generated by g's on finite support.) For the log F distribution, the appropriate normalisation is $\mu = -\log \alpha$, $\sigma = 1$, so that $\ell_1(x) = e^{-x}$ and the limiting density is proportional to $e^{-\beta x} \exp(-e^{-x})$, $x \in \mathcal{R}$. This Gumbel extreme value limiting distribution corresponds to the log F's interpretation as an order statistic distribution (Jones, 2004, Section 4.6). It is a consequence of the logistic's exponential tails and the same limiting distribution applies to e.g. the doubly double exponential distribution. For the hyperbolic distribution, take $\mu = 0$, $\sigma = 4/\alpha$, so that the limiting density is proportional to $\exp\{-(\beta x + (1/x))\}$, $x \in \mathcal{R}^+$. This is the positive hyperbolic distribution (e.g. Barndorff-Nielsen and Blaesild, 1983, whose formula (7) incorporates a scale parameter).

5. Maximum likelihood estimation I: too many scale parameters

Let $X_1, ..., X_n$ be an i.i.d. sample from the location-scale version (7) of density f_G and assume that G is twice continuously differentiable. The asymmetric Laplace distribution is therefore disqualified from consideration on two counts, the second being the lack of a role for σ which cannot be separated from α and β in that case. (See Section 3.5 of Kotz, Kozubowski and Podgórski, 2001, for a full account of maximum likelihood estimation for the asymmetric Laplace distribution.) The (exact) unidentifiability of α, β and σ in the asymmetric Laplace case suggests that there might be what might be called a practical unidentifiability of α, β and σ in other cases of f_G. This proves to be so in the sense that the asymptotic correlation between the maximum likelihood estimators of at least one pair of these parameters is necessarily extremely high and therefore that there is no hope of estimating all these parameters well from data, nor indeed is there any need to: in practice, one parameter can be dropped. This is because α, β and σ all act as scale parameters, yet there are clear roles for only two scale parameters, one associated with the left-tail of the distribution, the other with the right (or perhaps one overall scale parameter and one parameter controlling

10

the left-right difference). Relatedly, the tails of $\sigma^{-1}f_G(\sigma^{-1}(x - \mu))$ go like $e^{(\alpha/\sigma)x}$ as $x \to -\infty$ and $e^{-(\beta/\sigma)x}$ as $x \to \infty$.

The elements of the observed and expected information matrices associated with maximum likelihood estimation in the four-parameter distribution (7) are given in the Appendix. The main point concerning the unnecessary nature of one of the scale parameters can, however, be demonstrated clearly in the symmetric three-parameter case with $\alpha = \beta$, as follows. The symmetry of the distribution means that the location estimate $\hat{\mu}$ is asymptotically independent of $\hat{\sigma}$ and $\hat{\alpha}$. Using manipulations similar to those underlying the Appendix, the elements of the submatrix of the expected information matrix associated with $\hat{\sigma}$ and $\hat{\alpha}$ are n times

$$J_{\sigma\sigma} = \frac{1}{\sigma^2}(1 + 2\alpha E(X_{F_G}^2 g(X_{F_G}))), \qquad J_{\sigma\alpha} = -\frac{1}{\sigma\alpha} \quad \text{and} \quad J_{\alpha\alpha} = \mathcal{M}_G''(\alpha)$$

where $\mathcal{M}_G(\alpha) = \log(\mathcal{K}_G(\alpha, \alpha))$. The asymptotic correlation, r say, of $\hat{\sigma}$ and $\hat{\alpha}$ is therefore the following function of α only:

$$r(\alpha) = \frac{1}{\alpha\{\mathcal{M}_G''(\alpha)(1 + 2\alpha E(X_{F_G}^2 g(X_{F_G})))\}^{1/2}}. \tag{9}$$

When $\alpha \to \infty$, the manipulations at the start of Section 4.2 can be extended to show that $\mathcal{M}_G''(\alpha) \sim 1/(2\alpha^2)$ and $E(X_{F_G}^2 g(X_{F_G})) \sim 1/(2\alpha)$ so that $\lim_{\alpha\to\infty} r(\alpha) = 1$. An asymptotic approximation of 1 for $\lim_{\alpha\to 0} r(\alpha)$ also seems to arise from other calculations. Indeed, an extraordinary closeness of $r(\alpha)$ to unity for all α is obtained in numerical calculations. For the log F and hyperbolic distributions, the *minimum* correlations that I obtained numerically were 0.992 and 0.994, respectively! I did a similar analysis for the four-parameter log F distribution in Jones (2006a) and obtained a (now rather less impressive!) minimum correlation between $\hat{\sigma}$ and each of $\hat{\alpha}$, $\hat{\beta}$ and $2/(\hat{\alpha} + \hat{\beta})$ of "almost 0.9".

Treating the log F distribution as a three parameter distribution must alleviate the computational problems noted with fitting the four-parameter distribution by Brown et al. (1996) and Dupuis (2001). For more on the theory of maximum likelihood estimation for the log F distribution see Prentice (1975) and for the hyperbolic distribution see Barndorff-Nielsen and Blaesild (1981).

6 MAXIMUM LIKELIHOOD ESTIMATION II: KERNEL QUANTILE ESTIMATION

The first likelihood equation reads $n^{-1} \sum_{i=1}^{n} G\left((X_i - \mu)/\sigma\right) = \alpha/(\alpha + \beta)$ or equivalently

$$n^{-1} \sum_{i=1}^{n} G\left(\frac{\mu - X_i}{\sigma}\right) = \frac{\beta}{\alpha + \beta} \equiv p. \tag{10}$$

The left-hand side of (10) is nothing other than the kernel estimator of the distribution function at the point μ with bandwidth σ and kernel distribution function G. Solving (10) for μ, the resulting $\hat{\mu}(p)$ *is precisely the inversion kernel quantile estimator at p* (Nadaraya, 1964, Azzalini, 1981). It is well known that maximum likelihood location estimation in the asymmetric Laplace distribution is equivalent to sample quantile estimation (e.g. Koenker and Machado, 1999); here, for the first time, is a simple generalisation to the case of kernel smoothed quantile estimation. It is somewhat intriguing to note that the more tractable choices of G from a distribution theory perspective and the usual preferred choices of G from a kernel estimation perspective (e.g. normal and Epanechnikov and other symmetric beta kernels; Sections 3.5 and 3.6) differ. However, the relative indifference to precise choice of kernel, bar perhaps smoothness considerations, matches with the relative similarity of members of the class f_G as in Fig. 1.

Define $\alpha + \beta = \delta$ and fix p by choice of quantile. In this parametrisation, the tails of the underlying density go like $e^{(1-p)(\delta/\sigma)x}$ as $x \to -\infty$ and $e^{-p(\delta/\sigma)x}$ as $x \to \infty$. This makes it clear (again) that σ and δ are both acting as scale parameters, but for current purposes it is appropriate to set $\delta = 1$ (and hence completely fix α and β as $1 - p$ and p, respectively) and retain σ, formula (10) still holding. Interestingly, the special case of the log F distribution with $\delta = 1$ that corresponds to use of the logistic kernel in (10) is precisely the NEF-GHS (natural exponential family generalized hyperbolic secant) distribution of Morris (1982). In addition, when $p = 1/2$, Huber's 'most robust' location M-estimator mentioned in Section 3.6 can now be newly interpreted as an inversion kernel median estimator using a uniform kernel.

But now we also have a (semi-)principled method of bandwidth selection by choosing σ and μ simultaneously by maximum likelihood (in the model with $\delta = 1$). The second likelihood equation that should be solved in con-

junction with (10) is

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\left\{p - G\left(\frac{\mu - X_i}{\sigma}\right)\right\} = \sigma. \tag{11}$$

Uniqueness of the estimators of μ and σ is assured. In fact, it can be shown that the left-hand sides of (10) and (11) are monotone decreasing in μ for fixed σ and in σ for fixed μ, respectively, over appropriate ranges of values and hence that simple (e.g. bisection) methods can be used successfully to compute $\hat{\mu}$ and $\hat{\sigma}$.

Simulation results using this methodology are, however, mixed. As an example, Table 1 gives results for $n = 50$ and the standard normal distribution; results for $n = 100$ and other distributions are qualitatively similar. The four methods compared in Table 1 are the sample quantile, the Harrell and Davis (1982) estimator and two estimators based on (10) with logistic G: the first takes σ to be the 'rule-of-thumb' bandwidth associated with minimisation of asymptotic mean squared error (Azzalini, 1981) assuming normality — which is in fact the right assumption here; the second utilises (11). Taking $50,000$ replications resulted in standard errors such that the simulated mean squared errors are (approximately) correct to the number of decimal places shown.

* * *   Table 1 about here   * * *

The kernel method with σ chosen by (11) performs particularly well at the median. This is because we are fitting a smooth symmetric log F distribution rather than the sample quantile's implicit Laplace distribution. This, of course, can also be considered to be good robust estimation of location via a particular M-estimator. Performance is rather worse for other quantiles. The new estimator proves to be of roughly comparable quality to the Harrell-Davis estimator (which is well thought of in the study of Sheather and Marron, 1990) but not as good as the rule-of-thumb kernel estimator (whose good performance persists for non-normal distributions). It struggles particularly when $p = 0.75$ but seems to improve again for higher p. The somewhat disappointing overall performance of the new estimator away from the median must be associated with the fitting of particular skew log F distributions that bear little relation to the symmetric distribution underlying the data (although the same is true of the asymmetric Laplace distribution underlying the sample quantile). Hence the words "a (semi-)principled method of bandwidth selection" above!

13

It is intended to explore the consequences of the above for kernel quantile regression elsewhere.

7. RELATED DISTRIBUTIONS II: EXPONENTIAL TAILS AND POWER TAILS

In this section, I will briefly explore distributions related to f_G by simple transformation.

7.1. Distributions with power tails on $\mathcal{R}^+$

Probably the most obvious transformation link to make is that associated with 'taking logs'. Let $Y = e^X$, $X = \log Y$. Then the density $f_{G;+,p}(y)$, $y > 0$, of Y has the form

$$f_{G;+,p}(y) = \frac{y^{\alpha-1}}{\mathcal{K}_G(\alpha, \beta)} \exp\{-(\alpha + \beta)G^{[2]}(\log(y))\}. \tag{12}$$

In this way, the simple exponential tails of density f_G translate to simple power tails for $f_{G;+,p(y)}$:

$$f_{G;+,p}(y) \sim y^{\alpha-1} \text{ as } y \to 0, \qquad f_{G;+,p}(y) \sim y^{-(\beta+1)} \text{ as } y \to \infty.$$

Elsewhere (Jones, 2006b) I have argued that this behaviour at 0 — that of the reciprocal of a random variable with a $y^{-(\alpha+1)}$ right-hand density tail — is the natural analogue of the power tail at infinity. Formula (12) might be seen as directly generating densities with power tails on $\mathcal{R}^+$ starting from a simple symmetric distribution on $\mathcal{R}$.

Immediately and unsurprisingly, the power-tailed distribution associated with the log F distribution on $\mathcal{R}$ is the F distribution on $\mathcal{R}^+$. The distribution associated with f_H is known as the log hyperbolic distribution and it is in that guise that it is most often used as a model for (positive) data (e.g. Barndorff-Nielsen, 1977). The distribution associated with f_{AL} has the simple two-piece density given by

$$f_{AL;+,p}(y) = \frac{\alpha\beta}{\alpha + \beta} \left\{ y^{\alpha-1} I(0 < y < 1) + y^{-(\beta+1)} I(y \geq 1) \right\};$$

Fieller, Flenley and Olbricht (1992) put this 'log-skew-Laplace' distribution forward as a more tractable alternative to the log hyperbolic distribution. Further distributions with power tails on $\mathcal{R}^+$ can, of course, be derived from other examples of f_G.

14

In Jones (2006b), I argued that the following simple transformation from $Y \in \mathcal{R}^+$ to $Z \in \mathcal{R}$ has the useful property of maintaining the power tails of $f_{G;+,p}(y)$ in the density $f_{G;p}(z)$, say:

$$Z = \frac{1}{2}\left(Y - \frac{1}{Y}\right) = \sinh(\log(Y)). \tag{13}$$

Combining this transformation with the exponential transformation to Y from X leads to densities with power tails on $\mathcal{R}$ in the sense that

$$f_{G;p}(z) \sim z^{-(\alpha+1)} \text{ as } z \to -\infty, \qquad f_{G;p}(z) \sim z^{-(\beta+1)} \text{ as } z \to \infty.$$

But the combined transformation is nothing other than $Z = \sinh(X)$. And the associated density is

$$f_{G;p}(x) = \mathcal{K}_G^{-1}(\alpha, \beta)\frac{(z + \sqrt{1+z^2})^\alpha}{\sqrt{1+z^2}} \exp\{-(\alpha + \beta)G^{[2]}(\sinh^{-1}(z))\}. \tag{14}$$

The density generated by logistic G is particularly interesting:

$$f_{LF;p}(z) = \frac{1}{B(\alpha, \beta)} \frac{(z + \sqrt{1+z^2})^\alpha}{\sqrt{1+z^2}(1 + z + \sqrt{1+z^2})^{\alpha+\beta}}.$$

This is the $k = 1$ special case of distribution (6.2) of Jones (2004) and, as such, is, when α and β are integers, the distribution of an order statistic of a random sample from the distribution with density

$$\frac{z + \sqrt{1+z^2}}{\sqrt{1+z^2}(1 + z + \sqrt{1+z^2})^2}.$$

(It is also interesting to note that the (scaled) t_2 distribution is nothing other than the distribution of $\sinh(L/2)$ where L follows the logistic distribution, a simple relationship buried in Jones, 2004, but missed by Jones, 2002.) The two-piece distributions associated with the asymmetric Laplace distribution have density

$$f_{AL;p}(z) = \frac{\alpha\beta}{(\alpha+\beta)\sqrt{1+z^2}}\left\{(z+\sqrt{1+z^2})^\alpha I(z < 0) + (\sqrt{1+z^2} - z)^\beta I(z \geq 0)\right\}.$$

15

Taking logs and sinhs of X's with particular distributions, different from those considered here, is at the heart of the Johnson system of distributions (Johnson, 1949, Johnson, Kotz and Balakrishnan, 1994a, Section 12.4.3). See Jones (2006b) for material on the interplay between Johnson distributions and transformation (13).

7·3. Distributions with exponential tails on $\mathcal{R}^+$

The inverse of transformation (13), $Y = X + \sqrt{1 + X^2} = \exp(\sinh^{-1}(X))$, can also be applied to densities with exponential tails on $\mathcal{R}$ to 'maintain' exponential tails on $\mathcal{R}^+$ in the sense that

$$f_{G;+,e}(y) \sim y^{-2} e^{-\alpha/(2y)} \text{ as } y \to 0, \qquad f_{G;+,e}(y) \sim e^{-(\beta/2)y} \text{ as } y \to \infty.$$

(The extra scaling factor of $1/2$ is inconsequential.) The 'exponential tail behaviour' at zero is actually that of the reciprocal of a random variable with exponential tail behaviour at infinity and is also rather similar to that of the inverse Gaussian distribution, for which the power -2 is replaced by $-3/2$.

To cut a longer story short, probably the most attractive distribution in this family turns out to be that associated with the hyperbolic distribution:

$$f_{H;+,e}(y) = \frac{1}{2\mathcal{K}_H(\alpha, \beta)} \left(1 + \frac{1}{y^2} \right) \exp\left\{ \frac{1}{2} \left(-\frac{\alpha}{y} - \beta y \right) \right\}.$$

This is a mixture of the positive hyperbolic distribution and its version weighted by $1/y^2$, with mixture probabilities $\alpha/(\alpha+\beta)$ and $\beta/(\alpha+\beta)$, respectively. However, this is in competition with the log hyperbolic distribution itself which arises from the limiting process of Section 7.3 and behaves as $e^{-1/y}$ as $y \to 0$ (as well as $e^{-\beta y}$ as $y \to \infty$.). Note, however, that the limiting process approach is less general than the transformation approach in that not all limiting densities (8) have support $\mathcal{R}^+$. A class of cases that have the required support arises from g having power upper tail $x^{-(\gamma+1)}$, $\gamma > 1$, for then $\ell_1(x) \propto x^{-(\gamma-1)}$ with density behaviour $e^{-1/y^{(\gamma-1)}}$ as $y \to 0$.

16

APPENDIX

Observed and expected information for the four-parameter case

Write $\mathcal{M}_G(\alpha, \beta) = \log \mathcal{K}_G(\alpha, \beta)$. Based on the log-likelihood

$$-n \log \sigma - n\mathcal{M}_G(\alpha, \beta) + \alpha \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} - (\alpha + \beta) \sum_{i=1}^{n} G^{[2]} \left(\frac{X_i - \mu}{\sigma} \right),$$

it can readily be shown that the elements of the observed information matrix (minus the second derivative of the log-likelihood) in the four-parameter case are as follows:

$$\iota_{\mu\mu} = \frac{(\alpha + \beta)}{\sigma^2} \sum_{i=1}^{n} g \left(\frac{X_i - \mu}{\sigma} \right);$$

$$\iota_{\mu\sigma} = \frac{(\alpha + \beta)}{\sigma^2} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} g \left(\frac{X_i - \mu}{\sigma} \right);$$

$$\iota_{\sigma\sigma} = \frac{1}{\sigma^2} \left\{ n + (\alpha + \beta) \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} g \left(\frac{X_i - \mu}{\sigma} \right) \right\};$$

$$\iota_{\mu\alpha} = \frac{1}{\sigma} \left\{ n - \sum_{i=1}^{n} G \left(\frac{X_i - \mu}{\sigma} \right) \right\} = \frac{\beta n}{(\alpha + \beta)\sigma};$$

$$\iota_{\mu\beta} = -\frac{1}{\sigma} \sum_{i=1}^{n} G \left(\frac{X_i - \mu}{\sigma} \right) = -\frac{\alpha n}{(\alpha + \beta)\sigma};$$

$$\iota_{\sigma\alpha} = \frac{1}{\sigma} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} \left\{ 1 - G \left(\frac{X_i - \mu}{\sigma} \right) \right\} = \frac{n}{(\alpha + \beta)\sigma} \left(\beta \frac{(\bar{X} - \mu)}{\sigma} - 1 \right);$$

$$\iota_{\sigma\beta} = -\frac{1}{\sigma} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} G \left(\frac{X_i - \mu}{\sigma} \right) = -\frac{n}{(\alpha + \beta)\sigma} \left(1 + \alpha \frac{(\bar{X} - \mu)}{\sigma} \right);$$

$$\iota_{\alpha\alpha} = n\mathcal{M}_G^{20}(\alpha, \beta); \qquad \iota_{\alpha\beta} = n\mathcal{M}_G^{11}(\alpha, \beta) \qquad \iota_{\beta\beta} = n\mathcal{M}_G^{02}(\alpha, \beta).$$

Here, $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ as usual.

It is clear that, on taking expectations, the elements of the expected information matrix are of the forms n times $j_{\mu\mu}/\sigma^2$, $j_{\mu\sigma}/\sigma^2$, $j_{\sigma\sigma}/\sigma^2$, $j_{\mu\alpha}/\sigma$, $j_{\mu\beta}/\sigma$, $j_{\sigma\alpha}/\sigma$, $j_{\sigma\beta}/\sigma$, $j_{\alpha\alpha}$, $j_{\alpha\beta}$ and $j_{\beta\beta}$, respectively, where the j's are functions of α and β only. The j's look much like the ι's above except that the first three depend on $E(X_{F_G}^r g(X_{F_G}))$, $r = 0, 1, 2$, and $E((\bar{X} - \mu)/\sigma) = \mathcal{M}_G^{10}(\alpha, \beta) - \mathcal{M}_G^{01}(\alpha, \beta)$. Therefore, the expected information matrix does not depend on μ at all and asymptotic *correlations* are independent of σ too.

References

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68**, 326–8.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London, Ser. A* **353**, 401–19.

Barndorff-Nielsen, O. & Blaesild, P. (1981). Hyperbolic distributions and ramifications: contributions to theory and application. In *Statistical Distributions in Scientific Work*, Vol. 4, Eds. C. Taillie et al., pp. 19–44. Reidel.

Barndorff-Nielsen, O. & Blaesild, P. (1983). Hyperbolic distributions. In *Encyclopedia of Statistical Sciences*, Eds. N.L. Johnson, S. Kotz and C.B. Read, pp. 700–7. New York: Wiley.

Bassan, B., Denuit, M. & Scarsini, M. (1999). Variability orders and mean differences. *Statist. Probab. Lett.* **45**, 121–30.

Brown, B.W., Spears, F.M., Levy, L.B., Lovato, J. & Russell, K. (1996). Algorithm 762: LLDRLF, log-likelihood and some derivatives for log-F models. *ACM Trans. Math. Software* **22**, 372–82.

Dupuis, D.J. (2001). Fitting log-F models robustly, with an application to the analysis of extreme values. *Comput. Statist. Data Anal.* **35**, 321–33.

Fieller, N.R.J., Flenley, E.C. & Olbricht, W. (1992). Statistics of particle size data. *Appl. Statist.* **41**, 127–46.

Gumbel, E.J. (1958). *Statistics of Extremes.* New York: Columbia University Press.

Harrell, F.E. & Davis, C.E. (1982). A new distribution-free quantile estimator. *Biometrika* **69**, 635–40.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73–101.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, Second Edition. New York: Wiley.

Jones, M.C. (2002). Student's simplest distribution. *Statistician* **51**, 41–9.

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

Jones, M.C. (2006a). The logistic and the log F distribution. In *Handbook of the Logistic Distribution*, Second Edition, Ed. N. Balakrishnan, New York: Dekker, to appear. Open University Department of Statistics Technical Report 06/01; see http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm

Jones, M.C. (2006b). Connecting distributions with power tails on the real line, the half line and the interval. Under consideration. Open University Department of Statistics Technical Report 05/12; see http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm

Jones, M.C. & Faddy, M.J. (2003). A skew extension of the t distribution, with applications. *J. Roy. Statist. Soc. Ser. B* **65**, 159–74.

Koenker, R. & Machado, J.A.F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.* **94**, 1296–310.

Kotz, S., Kozubowski, T.J. & Podgórski, K. (2001). *The Laplace Distribution and Generalizations; A Revisit With Applications to Communications, Economics, Engineering, and Finance.* Boston: Birkhauser.

Morris, C.N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10**, 65–80.

Nadaraya, E.A. (1964). Some new estimates for distribution functions. *Theor. Probab. Applic.* **15**, 497–500.

Prentice, R.L. (1975). Discrimination among some parametric models. *Biometrika* **62**, 607–14.

Sheather, S.J. & Marron, J.S. (1990). Kernel quantile estimators. *J. Amer. Statist. Assoc.* **85**, 410–16.

Table 1: *Mean squared errors associated with the estimation of normal quantiles from samples of size $n = 50$ for specified p and four estimation methods. The logistic kernel was used in the kernel methods. $50,000$ replications*

p	Sample quantile	Harrell-Davis	Kernel; rule-of-thumb bandwidth	Kernel; bandwidth via (11)
0.50	0.032	0.027	0.032	0.022
0.75	0.037	0.032	0.031	0.035
0.9	0.063	0.049	0.047	0.049
0.95	0.086	0.076	0.068	0.075

Fig. 1: Log F (solid), hyperbolic (dashed) and normal-based (dotted) distributions with variance unity and $\beta = 2$ with, from left, means $-8, 0$ and 8 and $\alpha = 8$, 2 and 0.25, respectively.