Statistica Sinica **18**(2008), 1101-1110

ON A CLASS OF DISTRIBUTIONS WITH SIMPLE EXPONENTIAL TAILS

M. C. Jones

The Open University

Abstract: A general construction is put forward that covers many unimodal univariate distributions with simple, exponentially decaying tails (e.g., asymmetric Laplace, log F, and hyperbolic distributions, plus many new models). The proposed family is a 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 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 equivalent to a standard form of kernel quantile estimation, choice of kernel corresponding to choice of model within the class. This leads to a maximum likelihood method for bandwidth selection in kernel quantile estimation, but with somewhat mixed practical performance.

Key words and phrases: 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, for some $\alpha, \beta > 0$, its density f has the properties

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

The archetypal example of such a distribution has density

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

this is the asymmetric Laplace distribution (Kotz, Kozubowski and Podgórski (2001, Chap. 13)). It is particularly tractable and straightforward, its one drawback, to some, being its 'pointed', non-differentiable nature at $x = 0$.

Which other distributions share the property of simple exponential tails? I initially knew of two − the log F and hyperbolic distributions − and they surface below (their properties include smoother behaviour around $x = 0$). The first 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). Input to this construction is a simple, symmetric, 'generating' distribution with random variable X_G, density g, distribution function G, and first iterated (left-tail) distribution function $G^{[2]}(x) = \int_{-\infty}^{x} G(t)dt = E\{(x - X_G)I(X_G < x)\}$ which, inter alia, is $G(x)$ times the mean residual life function (Bassan, Denuit and Scarsini (1999), and references therein). $G^{[2]}(x)$ does not exist for any x if $g(x) \sim |x|^{-(\gamma+1)}$ as $x \to -\infty$ for some $0 < \gamma \leq 1$, so such very heavy tailed generating distributions — 'Cauchy and heavier' — are disqualified from consideration. Taking g to be symmetric (about zero) is a convenience that affords particularly elegant simplifications without losing importantly in generality, and is followed virtually throughout this paper. The main construction and basic properties are given in Section 2. Special cases are considered in Section 3. The proposed family is a special subset of a regular exponential family.

The second purpose of this paper is to make two major practical points in the context of maximum likelihood fitting of the proposed family of distributions to data. The first of these (Section 4) explores whether the new construction really needs all four of its parameters in practice. The answer is negative: three parameters suffice. The second (Section 5) observes that maximum likelihood estimation of location in the new family is precisely equivalent to a standard form of kernel quantile estimation (Azzalini (1981)); specific choice of kernel is equivalent to specific choice of model within the family. 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, in Section 6, two closely related topics are briefly described. Some details missing from this paper can be found in its technical report version, Jones (2006).

2. General Construction and Properties

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

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

The key to this construction is that as $x \to -\infty$, $G^{[2]}(x) \to 0$, and 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 tails of allowed G. Of course, the location-scale extension $\sigma^{-1} f_G\{\sigma^{-1}(x - \mu)\}$ would appear to be available for practical work, but see Section 4.

The exponential tails ensure integrability of f_G, confirming that it is a density, albeit one for which the normalisation constant $\mathcal{K}_G(\alpha, \beta)$ is not generally available in closed form. Likewise, the exponential tails imply the existence of all moments of the distribution, but their explicit formulae are 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 seen to be $\mathcal{K}_G(\alpha+t,\beta-t)/\mathcal{K}_G(\alpha,\beta)$ (and 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 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 of G. Moreover, densities (3) are log-concave in x, i.e., strongly unimodal. Let X_{F_G} follow the distribution with density f_G. Then, also, $E\{G(X_{F_G})\} = \alpha/(\alpha+\beta)$.

For symmetric g, two 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; 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 swapped. 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 density (3) again. This is because, for symmetric g, $G^{[2]}(x) - \bar{G}^{[2]}(x) = E(x - X_G) = x$. Another apparent generalisation of (3) is also fruitless: introduction of a scale parameter into G results only in a rescaled and reparametrised version of (3).

Formulation (4), in particular, makes it clear that f_G is symmetric (about zero) if and only if $\alpha = \beta$ (for symmetric g). Indeed, 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 (3) when G corresponds to a point mass at zero: $G(x) = I(x \ge 0)$, $G^{[2]}(x) = xI(x \ge 0)$.

3.2. The log F distribution

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 $f_{LF}(x) \propto (1+e^x)^{-(\alpha+\beta)}e^{\alpha x}$ and, in fact, $\mathcal{K}_{LF}(\alpha,\beta) = B(\alpha,\beta)$ where $B(\cdot,\cdot)$ is the beta function. This is the density of the log F distribution which dates back to R.A. Fisher and appears from time to time, and in a variety of guises, in the literature. For a partial review and references, see Jones (2008).

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, Jones (2004) argues 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 $G^{[2]}(x) = (1/2)(x+\sqrt{1+x^2})$. This results in the hyperbolic distribution of Barndorff-Nielsen (1977) (see also Barndorff-Nielsen and Blaesild (1983)):

$$f_H(x) \propto \exp\left[\left\{(\alpha-\beta)\,x - (\alpha+\beta)\,\sqrt{1+x^2}\right\} \times \frac{1}{2}\right];$$

$\mathcal{K}_H(\alpha,\beta) = (\alpha+\beta)K_1(\sqrt{\alpha\beta})/\sqrt{\alpha\beta}$, where $K_1(\cdot)$ is a Bessel function. This parametrisation is not the most usual one that 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. Consideration of the hyperbolic form of $\log f_H$ makes this distribution an especially natural one with simple exponential tails from the viewpoint of linear asymptotes for the log density.

Jones (2004) argues that the two most tractable and useful order statistic distributions are the log F distribution, generated by the logistic, and the skew t distribution of Jones and Faddy (2003), generated by the t_2. Here, I suggest that the two most tractable and useful (smooth) alternatives to the asymmetric Laplace distribution with exponential tails are, again, the log F distribution, generated in an alternative fashion by the logistic, together with a different distribution, the hyperbolic, which turns out also to be generated by the t_2 distribution. I find the central place of the t_2 distribution in this kind of distribution theory intriguing, even more so now than seen in Jones (2002).

3.4. Other members of family (3)

One can take g to be a Laplace distribution, in which case the following new ('doubly double exponential') distribution arises:

$$f_{DDE}(x) \propto \begin{cases} \exp(\alpha x - ce^{x}) & \text{if } x < 0, \\ \exp(-\beta x - ce^{-x}) & \text{if } x \geq 0, \end{cases}$$

where $c = (\alpha+\beta)/2$, see Jones (2006, Sec. 3.4). This density is differentiable everywhere but lacks a *second* continuous derivative at $x = 0$.

Further smooth f's arise from further smooth G with support $\mathcal{R}$, but none seems more attractive than those already considered. An obvious example is the normal-based distribution with $f_N(x) \propto \exp[\alpha x - (\alpha+\beta)\{x\Phi(x) + \phi(x)\}]$, where ϕ and Φ are the standard normal density and distribution functions. Differences between such smooth densities are, in any case, not very marked: see Figure 1 of Jones (2006) for a graphical comparison.

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

$$f_U(x) \propto \begin{cases} \exp(\alpha x) & \text{if } x < -1, \\ \exp\left(-\frac{\alpha\beta}{\alpha+\beta}\right) \exp\left\{-\frac{1}{4}(\alpha+\beta)\left(x - \frac{\alpha-\beta}{\alpha+\beta}\right)^2\right\} & \text{if } -1 \leq x < 1, \\ \exp(-\beta x) & \text{if } x \geq 1. \end{cases} \tag{5}$$

($\mathcal{K}_U(\alpha, \beta)$ is given in Jones (2006)). Density (5) is the result of continuously – but not differentiably – joining two lines and a quadratic center on the log density scale. Equivalently, it consists of a normal center onto which exponential tails have been grafted. In the symmetric case, (5) is the density associated with the 'most robust' M-estimator of Huber (1964, p.75; rescale (5) by factor k and take $\alpha = \beta = k^2$ to match Huber's parameterisation). Higher-order contact between pieces can be achieved by replacing the quadratic by a higher-order polynomial, such as another symmetric beta density. See also Section 5.

4. Maximum Likelihood Estimation I: Too Many Scale Parameters

Let $X_1, \ldots, X_n$ be an i.i.d. sample from the location-scale version

$$\frac{1}{\sigma} f_G\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\mathcal{K}_G(\alpha,\beta)} \exp\left\{\alpha\frac{(x-\mu)}{\sigma} - (\alpha+\beta)G^{[2]}\left(\frac{x-\mu}{\sigma}\right)\right\} \tag{6}$$

of density f_G, and assume that G is twice continuously differentiable. The asymmetric Laplace distribution is thus disqualified from consideration on two counts, the second being the lack of 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 an account of maximum likelihood estimation for the asymmetric Laplace distribution.) The (exact) non-identifiability of α, β and σ in the asymmetric Laplace case suggests a practical non-identifiability 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 extremely high, and therefore there is no hope of estimating all these parameters well from data: 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. 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 (6) are given in Jones (2006). The main point concerning the unnecessary nature of one scale parameter can, however, be demonstrated in the symmetric, three-parameter case ($\alpha = \beta$), as follows. The symmetry of the distribution means that the location estimate $\hat{\mu}$ is asymptotically independent of $\hat{\sigma}$ and $\hat{\alpha}$. The elements of the submatrix of the expected information matrix associated with $\hat{\sigma}$ and $\hat{\alpha}$ turn out to be n times

$$J_{\sigma\sigma} = \frac{1}{\sigma^2}[1 + 2\alpha E\{X_{F_G}^2 g(X_{F_G})\}],$$

$$J_{\sigma\alpha} = -\frac{1}{\sigma\alpha} \quad \text{and} \quad J_{\alpha\alpha} = \mathcal{M}_G''(\alpha),$$

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

$$r(\alpha) = \frac{1}{\alpha\left(\mathcal{M}_G''(\alpha)[1 + 2\alpha E\{X_{F_G}^2 g(X_{F_G})\}]\right)^{\frac{1}{2}}}.$$

When $\alpha \to \infty$, manipulations in Section 4.2 of Jones (2006) 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 value of 1 for $\lim_{\alpha\to 0} r(\alpha)$ arises from other calculations. Indeed, an extraordinary closeness of $r(\alpha)$ to unity for all α is obtained in numerical computations: for the log F and hyperbolic distributions, the *minimum* correlations obtained were 0.992 and 0.994, respectively! A similar analysis was done for the four-parameter log F distribution in Jones (2008) where a (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" was found.

Treating the log F distribution as a three parameter distribution alleviates the computational problems with fitting the four-parameter distribution noted by Brown, Spears, Levy, Lovato and Russell (1996) and Dupuis (2001). The user is left with a choice between, essentially, $(\mu, \sigma, 1-p, p)$, $0 < p < 1$, and $(\mu, 1, \alpha, \beta)$ that one expects to be unimportant in practical performance terms. Within these parametrisations, further reparametrisations such as $(\mu, \sigma, (1-p)\sigma, p\sigma)$ or $(\mu, 1, p_0, q_0)$, where $p_0 > 0, q_0$ are new parameters suggested by Prentice (1975), are possible.

5. Maximum Likelihood Estimation II: Kernel Quantile Estimation

The likelihood equation associated with differentiation with respect to μ in (6) reads $n^{-1}\sum_{i=1}^n G\{(X_i - \mu)/\sigma\} = \alpha/(\alpha+\beta)$, or equivalently

$$n^{-1}\sum_{i=1}^n G\left(\frac{\mu - X_i}{\sigma}\right) = \frac{\beta}{\alpha+\beta} \equiv p. \tag{7}$$

The left-hand side of (7) is the kernel estimator of the distribution function at the point μ, with bandwidth σ and kernel distribution function G. Solving (7) for μ, the resulting $\hat{\mu}(p)$ *is precisely the inversion kernel quantile estimator at* p (Nadaraya (1964) and 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. Intriguingly, the more tractable choices of G from a distributional perspective and the usual choices of G from a kernel perspective (e.g., normal and Epanechnikov and other symmetric beta kernels) differ. However, the relative indifference to precise choice of kernel, bar perhaps smoothness considerations, matches the relative similarity of members of the class f_G.

The three-parameter version of (6) that yields (7) directly is parametrised by $(\mu, \sigma, 1-p, p)$, but now take p to be fixed by choice of quantile. Interestingly, the special case of the log F distribution that corresponds to use of the logistic kernel in (7) is the 'NEF-GHS' distribution of Morris (1982). In addition, when $p = 1/2$, Huber's 'most robust' location M-estimator (Section 3.4) can now be interpreted as an inversion kernel median estimator using a uniform kernel. Quite generally, inversion kernel median estimators are seen to be equivalent to M-estimators of location. On the other hand, inversion kernel quantile estimators differ from kernel-smoothed order statistic estimators (for several of which, see Sheather and Marron (1990)) but exhibit broadly comparable behaviour (see below for some evidence). Note that (7) is readily solved numerically, and that there is no need to worry about boundary effects near $p = 0$ or 1.

But now we also have a (semi-)principled method of bandwidth selection by choosing σ and μ simultaneously by maximum likelihood. The second likelihood equation that should be solved in conjunction with (7) is

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\left\{p - G\left(\frac{\mu - X_i}{\sigma}\right)\right\} = \sigma. \qquad (8)$$

Uniqueness of the estimators of μ and σ is assured. In fact, the left-hand sides of (7) and (8) are monotone decreasing in μ for fixed σ and in σ for fixed μ, respectively, over appropriate ranges of values and hence, attractively, simple (e.g., bisection) methods can be used successfully to compute both $\hat{\mu}$ and $\hat{\sigma}$.

Simulation results using this methodology are, however, mixed. As examples, Table 1 gives results for $n = 25, 50, 100$ and the standard normal and Laplace distributions. The four methods compared in Table 1 are the sample quantile, the Harrell and Davis (1982) estimator, and two estimators based on (7) with 'biweight' (i.e. beta(3, 3)) G: the first takes σ to be the 'rule-of-thumb' bandwidth associated with minimisation of asymptotic mean squared error (Azzalini (1981)) assuming normality; the second utilises (8). Taking $50,000$ replications resulted

Table 1. Mean squared errors (and biases in brackets beneath them) associated with the estimation of normal and Laplace quantiles from samples of size $n = 25, 50, 100$ for specified p and four estimation methods. The biweight kernel was used in the kernel methods. $50,000$ replications.

	Normal distribution				Laplace distribution			
p	Sample quantile	Harrell-Davis	Kernel; r-o-t b'width	Kernel; b'width via (8)	Sample quantile	Harrell-Davis	Kernel; r-o-t b'width	Kernel; b'width via (8)
	$n = 25$							
0.50	0.062	0.052	0.062	0.056	0.054	0.050	0.054	0.050
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
0.75	0.073	0.061	0.064	0.067	0.113	0.110	0.161	0.109
	(−0.037)	(0.018)	(0.086)	(−0.003)	(−0.021)	(0.078)	(0.247)	(0.030)
0.9	0.111	0.096	0.098	0.108	0.354	0.373	0.314	0.347
	(−0.017)	(0.054)	(0.095)	(−0.014)	(0.010)	(0.195)	(0.186)	(0.015)
	$n = 50$							
0.50	0.032	0.027	0.032	0.028	0.025	0.023	0.025	0.023
	(0.025)	(0.000)	(0.025)	(0.000)	(0.023)	(0.000)	(0.023)	(0.000)
0.75	0.037	0.032	0.032	0.033	0.060	0.055	0.077	0.054
	(−0.003)	(0.009)	(0.057)	(0.002)	(0.010)	(0.040)	(0.172)	(0.022)
0.9	0.063	0.049	0.050	0.054	0.219	0.175	0.161	0.172
	(0.049)	(0.028)	(0.063)	(−0.005)	(0.116)	(0.103)	(0.123)	(0.017)
	$n = 100$							
0.50	0.016	0.014	0.016	0.014	0.012	0.011	0.012	0.011
	(0.012)	(−0.001)	(0.012)	(−0.001)	(0.011)	(0.000)	(0.011)	(0.000)
0.75	0.019	0.016	0.016	0.017	0.032	0.028	0.036	0.027
	(0.014)	(0.004)	(0.036)	(0.003)	(0.026)	(0.020)	(0.115)	(0.018)
0.9	0.030	0.026	0.025	0.027	0.098	0.082	0.079	0.085
	(0.024)	(0.014)	(0.041)	(−0.002)	(0.055)	(0.050)	(0.077)	(0.009)

in standard errors such that the simulated mean squared errors are (approximately) correct to the number of decimal places shown.

The kernel method with σ chosen by (8) performs particularly well at the median. This is because we are utilising the variance of a well-fitting model as bandwidth. This, of course, can also be considered to be good robust estimation of location via a particular M-estimator. Performance tends to be rather worse for other quantiles, although the new method is best for the 0.75 quantile of the Laplace distribution. Overall, 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 is often beaten by the rule-of-thumb kernel estimator away from the median. The latter observation must be associated with the fitting of particular skew distributions that bear little relation to the symmetric distributions 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.

6. Related Methodology

Formula (3) also provides distributions for use on support $\mathcal{R}^+$, most obviously via log transformation. The simple exponential tails of f_G translate to simple power tails for the transformed density $f_{G;+}(y)$: $f_{G;+}(y) \sim y^{\alpha-1}$ as $y \to 0, f_{G;+}(y) \sim y^{-(\beta+1)}$ as $y \to \infty$. Elsewhere (Jones (2007)) 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 a natural analogue of the power tail at infinity. The F, log hyperbolic, and 'log-skew-Laplace' distributions arise as special cases: it is in the guise of the log hyperbolic that the hyperbolic distribution is most often used as a model for data (e.g., Barndorff-Nielsen (1977)), while Fieller, Flenley and Olbricht (1992) put the log-skew-Laplace distribution forward as a more tractable alternative to the log hyperbolic.

The work of Section 5 on kernel quantile estimation can be extended to the regression context. To do so, take the version of (6) used in Section 5 as distribution for the response variable, and pursue the usual approach of introducing a further kernel function 'in the x-direction' to fit lines (or other polynomials) locally (e.g., Loader (1999)). This yields a more principled version of the "double kernel local linear quantile regression" approach of Yu and Jones (1998) which turns out to consistently, if not always substantially, outperform the original version. See Jones and Yu (2007) for much more on this improved double kernel local linear quantile regression.

Acknowledgement

I am grateful to Karim Anaya, Frank Critchley, and the referees for their helpful suggestions and questions regarding this work.

References

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

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London A* **353**, 401-419.

Barndorff-Nielsen, O. and Blaesild, P. (1983). Hyperbolic distributions. *Encyclopedia of Statistical Sciences*, (Edited by N. L. Johnson, S. Kotz and C. B. Read), 700-707. Wiley, New York.

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

Brown, B. W., Spears, F. M., Levy, L. B., Lovato, J. and Russell, K. (1996). Algorithm 762: LL-DRLF, log-likelihood and some derivatives for log-F models. *ACM Trans. Math. Software* **22**, 372-382.

Dupuis, D. J. (2001). Fitting log-F models robustly, with an application to the analysis of extreme values. *Comput. Statist. Data Anal.* **35**, 321-333.

Fieller, N. R. J., Flenley, E. C. and Olbricht, W. (1992). Statistics of particle size data. *Appl. Statist.* **41**, 127-146.

Harrell, F. E. and Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika* **69**, 635-640.

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

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. (2006). On a class of distributions with simple exponential tails. Open University Department of Statistics Technical Report 06/02; see http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm

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). The logistic and the log F distribution. *Handbook of the Logistic Distribution*, Second Edition, (Edited by J. Baghdachi and N. Balakrishnan), Dekker, New York, to appear.

Jones, M. C. and Faddy, M. J. (2003). A skew extension of the t distribution, with applications. *J. Roy. Statist. Soc. Ser. B* **65**, 159-174.

Jones, M. C. and Yu, K. (2007). Improved double kernel local linear quantile regression. *Statist. Modelling* **7**, 377-389.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.* **94**, 1296-1310.

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.

Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York.

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**, 607-614.

Sheather, S. J. and Marron, J. S. (1990). Kernel quantile estimators. *J. Amer. Statist. Assoc.* **85**, 410-416.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228-237.

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

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

(Received July 2006; accepted March 2007)