Density-Based Skewness and Kurtosis Functions

Frank Critchley and M.C. Jones

New, functional, concepts of skewness and kurtosis are introduced for large classes of continuous univariate distributions. They are the first skewness and kurtosis measures to be defined directly in terms of the probability density function and its derivative, and are directly interpretable in terms of them. Unimodality of the density is a basic prerequisite. The mode defines the centre of such densities, separating their left and right parts. Skewness is then simply defined by suitably invariant comparison of the distances to the right and left of the mode at which the density is the same, positive function values arising when the former distance is larger. Our skewness functions are, thus, directly interpretable right-left comparisons which characterise asymmetry, vanishing only in the symmetric case. Kurtosis is conceived separately for the left and right parts of a unimodal density, these concepts coinciding in the symmetric case. By reflection in the mode, it suffices to consider right kurtosis. This, in turn, is directly and straightforwardly defined as skewness of an appropriate unimodal function of the right density derivative, two alternative functions being of particular interest. Dividing the right density into its peak and tail parts at the mode of such a function, (right) kurtosis is seen as a corresponding tail-peak comparison. A number of properties and illustrations of both skewness and kurtosis functions are presented and a concept of relative kurtosis addressed. Estimation of skewness and kurtosis functions, via kernel density estimation, is briefly considered and illustrated. Scalar summary skewness and kurtosis measures based on suitable averages of their functional counterparts are also considered and a link made to a popular existing scalar skewness measure. Further developments are briefly

indicated.

KEY WORDS: Density derivative; Density inverse; Kernel estimation; Khintchine's theorem; Mode; Unimodal distribution.

Frank Critchley and M.C. Jones are Professors, Department of Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK (E-mail: *f.critchley@open.ac.uk*, *m.c.jones@open.ac.uk*).

1. INTRODUCTION

Beginning statistics students are taught to initially assess the skewness (departure from symmetry) of a continuous univariate distribution by looking at its probability density function or an estimate thereof, typically a histogram. They are then introduced to one or more scalar measures of skewness, these being both suitably invariant (unaffected by location and scale changes, changing sign under reflection) and signed (positive values corresponding to skewness to the right, in some sense). However, they find that these skewness measures are only indirectly derived from the density function *per se* through quantities such as moments or the distribution function or its inverse, the quantile function. They will also find that these scalar measures do not fully capture the notion of skewness, vanishing for many asymmetric distributions. This last is unsurprising since, just as symmetry is an essentially *functional* concept, so too is its opposite, skewness. In this paper, we remedy this state of affairs by defining suitably invariant signed skewness functions by direct and immediately interpretable reference to the density, these skewness functions vanishing *only* in the symmetric case. Scalar skewness summary measures can, if required, be defined as certain averages of the skewness functions. The basic ideas and theory behind our proposals pertaining to skewness functions are given in Section 2. They are defined for large classes of unimodal densities identified there.

Compared to skewness (asymmetry), the concept of kurtosis is much harder to make precise. Kurtosis has been variously defined as a measure of peakedness or of heavy tails or of some kind of combination of the two, sometimes involving bimodality (e.g. Darlington, 1970, Ruppert 1987, Balanda and MacGillivray, 1988, and references therein). By separately considering left and right hand parts of unimodal density functions, we link kurtosis to skewness by defining left and right kurtosis to be skewness of appropriate unimodal functions of the derivative of left and right hand parts of the density. Two such functions — yielding similar qualitative results — will be our particular focus, left and right kurtosis being defined for corresponding large subclasses of unimodal densities identified in Section 3. We summarise this as: *kurtosis is gradient skewness*. Overall, our approach: (i) naturally defines kurtosis to be a functional concept (with averages of kurtosis functions providing scalar kurtosis measures); (ii) yields specific and natural definitions of peak and tail and of the contrast between them; (iii) is as immediately applicable to skew distributions as it is to symmetric; and (iv) makes for immediate and general transfer of results obtained for skewness to kurtosis.

Accordingly, Section 2 on skewness and Section 3 on kurtosis have structure in common: Sections 2.1 and 3.1 clarify the simple ideas alluded to above, Sections 2.2 and 3.2 give some theoretical support and properties, while Sections 2.3 and 3.3 show examples of skewness and kurtosis functions, respectively, for a number of familiar distributions. Our general kurtosis formulation in Section 3.1 is framed in terms of one of our two favoured specific kurtosis functions; the other is described and discussed in Section 3.4. The notion of relative kurtosis is, additionally, considered in Section 3.5. An illustrative example of skewness and kurtosis functions estimated from data is offered in Section 4; kernel density estimation plays a central role here. The possibilities for, and links with, scalar skewness and kurtosis measures are explored in Section 5. Of particular interest is a link with Arnold and Groeneveld's (1995) skewness measure. Our conclusions are given in Section 6, where further developments are briefly indicated.

The provision of skewness and kurtosis functions and measures which are clearly interpretable by direct reference to the density and/or its derivative is our primary goal. As such, our work differs from virtually all the existing work on skewness and/or kurtosis. A partial list of the more important of these works — beyond the classical moment-based measures — includes van Zwet (1964), Oja (1981), Groeneveld and Meeden (1984), Ruppert (1987), Balanda and MacGillivray (1988, 1990), Hosking (1992), Arnold and Groeneveld (1995), Benjamini and Krieger (1996), Groeneveld (1998), Serfling (2004) and Brys, Hubert and Struyf (2005). Aspects of our approach are touched on by Averous, Fougères & Meste (1996).

We end this introduction by establishing some further basic terminology and notation, used throughout. We use 'unimodal' to refer to a continuous function on an open interval that is strictly increasing up to a finite maximum and strictly decreasing thereafter, such a function being called 'rooted' if its limiting value at both endpoints is zero. For any $a < b$, the class of all rooted unimodal densities with support (a, b) is denoted $\mathcal{F}(a, b)$, or simply $\mathcal{F}$ when no confusion is possible. By invariance, there is no loss of generality in restricting attention to three possibilities — $(a, b) = (0, 1), (0, \infty)$ or $\mathcal{R}$ — depending on whether a and/or b is finite, (rootedness being automatic in this last case). We use 'k-smooth' to abbreviate 'k-times continuously differentiable'. Denoting by $\mathcal{F}_k$ the subclass of all k-smooth members of $\mathcal{F}$, we have the inclusions $\mathcal{F} = \mathcal{F}_0 \supset \mathcal{F}_1 \supset \ldots \supset \mathcal{F}_\infty$. Note that, whereas $f \in \mathcal{F}$ may not be differentiable, in particular at its mode m, $f'(m)$ necessarily vanishes for each $f \in \mathcal{F}_1$.

2. SKEWNESS

2.1 Skewness Functions

Our skewness functions are defined for any rooted unimodal density supported on an open interval, that is, for any $f \in \mathcal{F}(a, b)$, $a < b$. The simple ideas underlying our proposed functions are easily understood by reference to Figure 1, which shows such a density with mode at m. For any $0 < p < 1$, and hence for $0 < pf(m) < f(m)$ as indicated by the horizontal dashed line in Figure 1, there are two points $x_L(p)$ and $x_R(p)$, one each side of m, satisfying $f(x_L(p)) = f(x_R(p)) = pf(m)$. Their distances from the mode are, respectively, $\tau_L(p) = m - x_L(p)$ and $\tau_R(p) = x_R(p) - m$. Only if f is symmetric do we have $\tau_L(p) = \tau_R(p)$ for all $0 < p < 1$. In general, a comparison of $\tau_L(p)$ with $\tau_R(p)$ directly reflects the skewness of the density at each level $0 < p < 1$.

* * * FIGURE 1 ABOUT HERE * * *

Our prototype skewness function is, therefore,

$$\rho(p) = \tau_R(p)/\tau_L(p), \qquad 0 < p < 1.$$

It is location and scale invariant, a basic requirement of any skewness measure. However, it takes values on $(0, \infty)$ with $\rho(p) = 1$, $0 < p < 1$, corresponding to symmetry, $\rho(p) > 1$ indicating skewness to the right at level p and reflection of the density leading to $\rho(p) \to 1/\rho(p)$. It is more appealing, and in line with existing scalar skewness measures, to transform ρ to an entirely equivalent function γ, say, for which symmetry corresponds to $\gamma(p) = 0$, $0 < p < 1$, positive values to $\tau_R(p) > \tau_L(p)$ and reflection entails $\gamma(p) \to -\gamma(p)$. Thus, each skewness function value $\gamma(p)$ is both suitably invariant and signed. This can be achieved by a variety of strictly increasing transformations. The unique Box-Cox transformation with these properties is $\gamma^0 = \log \rho$ which takes values in $\mathcal{R}$. Again, for any $\lambda \geq 1$, we may use the transformation $\rho \to (\rho^\lambda - 1)/(\rho^\lambda + 1)$ which takes values in $(-1, 1)$. Our preferred choice is its $\lambda = 1$ version:

$$\gamma^*(p) = \frac{\tau_R(p) - \tau_L(p)}{\tau_R(p) + \tau_L(p)} = \frac{x_R(p) - 2m + x_L(p)}{x_R(p) - x_L(p)}, \quad 0 < p < 1, \tag{1}$$

this quantity being directly interpretable from the density as the signed proportionate difference between τ_R and τ_L at level p.

Note that skewness functions can be defined without the rootedness condition, the price for this extra generality being restriction of their domains to $\max\{f(a+)/f(m), f(b-)/f(m)\} < p < 1$.

2.2 Theoretical Support

2.2.1 The Class of Densities for Which Skewness is Defined. Each class of densities $\mathcal{F}(a,b)$, $a < b$, for which skewness is defined is location and scale equivariant. That is, under any transformation $x \to t(x) = c(x - x_0)$ ($x_0 \in \mathcal{R}$, $c > 0$),

$$t(f) \in \mathcal{F}(t(a), t(b)) \Leftrightarrow f \in \mathcal{F}(a,b)$$

where $t(f)$ denotes the density of $t(X)$ induced by $X \sim f$, with $\sim$ denoting 'is distributed as'. Also, $\mathcal{F}(a,b)$ is equivariant under reflection in the mode. That is,

$$f_m \in \mathcal{F}(b_m, a_m) \Leftrightarrow f \in \mathcal{F}(a,b)$$

where $x \to x_m = 2m - x$ induces $f \to f_m$; of course, $a_m = 2m - a$, $b_m = 2m - b$. Thus, overall, $\mathcal{F}(a,b)$ is affine equivariant.

Now, it is clear that f_m and f share the same mode, that $(f_m)_m = f$ and that symmetry is precisely the functional equation $f_m = f$. These basic facts reflect a simple left/right duality that both gives insight and cuts some work in half, entirely equivalent 'twin' results occurring in left/right pairs. To this end, we decompose f into its left and right parts, f_L and f_R say, defined as its restrictions to (a,m) and to (m,b) respectively. Under reflection in m, $f_L \to (f_m)_R$ while $f_R \to (f_m)_L$, a second reflection getting us back where we started. Skewness arises precisely when f_L, say, does not reflect onto f_R, that is, under any departures from $(f_m)_R = f_R$. As $a_m = b$ is necessary for this, exact symmetry is impossible on semi-finite intervals (a,b), represented here by $(0,\infty)$.

Since $f_L : (a,m) \to (0, f(m))$ and $f_R : (m,b) \to (0, f(m))$ are both strictly monotone (increasing and decreasing, respectively) and onto, they are invertible, so that $x_L(p) = f_L^{-1}(pf(m))$ and $x_R(p) = f_R^{-1}(pf(m))$. Our approach is therefore distinctive in being based on inversion of (part) density, rather than distribution or survival, functions.

The density function f determines a *proportion function* denoted $p_f(\cdot)$, or simply $p(\cdot)$, via $p(x) = f(x)/f(m)$, which is also rooted and unimodal at m with support (a, b), its modal value being unity. Its left part p_L is a distribution function on (a, m) and its right part p_R a survival function on (m, b) while $x_L = p_L^{-1}$ and $x_R = p_R^{-1}$. It follows at once that τ_L and τ_R – and, hence, any function of them, such as ρ or γ – depend on f only via its proportion function p_f. And, hence, that our skewness functions can be defined for any positive multiple cf, $c > 0$, of a density $f \in \mathcal{F}(a, b)$, integration to unity not being required. This fact is useful in developing our kurtosis functions (Section 3).

The functions τ_L and τ_R can be thought of as left and right scale functions and their sum, $\sigma(p) = x_R(p) - x_L(p)$, is the overall scale function suggested by Averous et al. (1996). We make only a couple of observations about this scale function here. First, its value at $p = 1/2$ is nothing other than the 'full width at half maximum' beloved of the physics community. And second, changing from vertical to horizontal the direction of integration of the area $\int_a^b p(x)dx$ under the graph of $p(\cdot)$, we find that

$$\int_0^1 \sigma(p)dp = \int_0^1 \{\tau_L(p) + \tau_R(p)\}dp = 1/f(m). \tag{2}$$

Thus, the scale function $\sigma(p)$ determines the scalar summary scale parameter $1/f(m)$ (this being, for example, $\sigma\sqrt{2\pi}$ in the normal density case).

We end this section by noting a number of decompositions of f into two or more components, each of which is directly interpretable from the graph of f (Figure 1) and from which f itself can be completely recovered. We write, for example, $f \overset{1-1}{\leftrightarrow} (f_L, f_R)$. Similar decompositions apply to related functions of interest. In particular, we may further decompose f_L into the scalar location and scale parameters, m and $1/f(m)$, and the left scale function τ_L. Combining this with its twin $f_R \overset{1-1}{\leftrightarrow} (m, 1/f(m), \tau_R)$ and using (2), we find the alternative decompositions

$$f \overset{1-1}{\leftrightarrow} (f_L, f_R) \overset{1-1}{\leftrightarrow} (m, \tau_L, \tau_R) \overset{1-1}{\leftrightarrow} (m, \tau_L, \gamma) \overset{1-1}{\leftrightarrow} (f_L, \gamma)$$

which hold for any skewness function γ, obvious twin decompositions applying.

2.2.2 Properties of Skewness Functions. By definition, ρ is location and scale invariant while, additionally, its strictly increasing transformations γ^0 and γ^* change sign under reflection. Each of these three functions completely characterises the skewness of f level-by-level. For any $0 < p < 1$,

$$\tau_R(p) = \tau_L(p) \Leftrightarrow \rho(p) = 1 \Leftrightarrow \gamma^0(p) = 0 \Leftrightarrow \gamma^*(p) = 0$$

in which case f is called symmetric at level p, overall symmetry occurring if and only if this condition holds for every level p. By strict monotonicity,

$$\tau_R(p) - \tau_L(p) > 0 \Leftrightarrow \rho(p) > 1 \Leftrightarrow \gamma^0(p) > 0 \Leftrightarrow \gamma^*(p) > 0$$

in which case f is called right or positive skew at level p, f being called *totally* right or positive skew if this condition holds for every $0 < p < 1$. The reverse inequalities characterise (total) left or negative skew (at level p) in the obvious way. In particular, both skewness functions are directly interpretable right-left comparisons.

Visibly (see Figure 1), it is clear that ρ – equivalently, γ^0 or γ^* – contains precisely the information required to move, level-by-level, between f_L and f_R. Algebraically, f_L and ρ together determine f_R via, for each $a < x < m$:

$$f_R\left\{m + \rho\left(f_L(x)/f(m)\right)(m - x)\right\} = f_L(x),$$

or equivalently, for each $m < x < b$:

$$f_L\left\{m - \frac{(x - m)}{\rho\left(f_R(x)/f(m)\right)}\right\} = f_R(x).$$

Now, given a symmetric rooted unimodal density h, say, on $\mathcal{R}$ or finite (a, b), 'two-piece' or 'split' densities are a fairly popular method of 'skewing' h; these define the skew density to be proportional to

$$h(x-m)I(x \leq m) + h((x-m)/\lambda)I(x > m), \qquad \lambda > 0, \tag{3}$$

(or equivalent parametrisations; see Fernandez and Steel, 1998, Mudholkar and Hutson, 2000, Jones, 2005). It follows that

$$\text{skewness function constant} \Leftrightarrow f \text{ is a two} - \text{piece distribution.}$$

In particular, under (3), $\rho(p) = \lambda$ for all $0 < p < 1$.

Whenever $f \in \mathcal{F}_2(a, b)$ and $f''(m) < 0$, Taylor expansion about $x = m$ gives

$$(x_R(p), x_L(p)) \to m \pm \sqrt{\frac{2(1-p)f(m)}{(-f''(m))}} \text{ as } p \to 1-.$$

Accordingly, in the same limit, $\rho(p) \to 1$ and hence $\gamma^0(p) \to 0$ and $\gamma^*(p) \to 0$. Whereas what happens on $\mathcal{R}$ as $p \to 0+$ depends on the specific tail-behaviours of f, in the semi-finite and finite support cases we always have $\rho(0+) = \infty$ and $\rho(0+) = (b-m)/(m-a)$, respectively. In particular, for every $f \in \mathcal{F}(0, \infty)$, $\gamma^0(p) \to \infty$ as $p \to 0+$ while $\gamma^*(p) \to 1$ in the same limit. Note that the final result means that neither γ skewness function for f on $(0, \infty)$ can ever be entirely negative for all p, but see Section 2.3.2.

2.3 Examples of Skewness Functions

Both our examples in this section have support $(0, \infty)$.

2.3.1 The Gamma Distribution. Setting the usual scale parameter, without loss of generality, to 1, the gamma density is proportional to $x^{\alpha-1}e^{-x}$, $\alpha > 0$. Our skewness functions are defined provided that $\alpha > 1$. If $\alpha \leq 1$, the gamma density (including the exponential for $\alpha = 1$) is a monotone decreasing function and hence is 'all right part' and has undefined skewness in our sense. The gamma skewness function $\gamma^*(p)$ is calculated numerically via the two solutions, $x_L(p)$ and $x_R(p)$, of $x^{\alpha-1}e^{-x} = pk_\alpha$ where $k_\alpha = (\alpha-1)^{\alpha-1}e^{-(\alpha-1)}$. It is shown for $\alpha = 2, 3, 5, 10, 20$ and 100 in Figure 2. We observe that all the skewness functions are strictly decreasing functions of p, the gamma densities are ordered in terms of decreasing skewness in the sense that each skewness function lies completely above the next and, for large α, the skewness is tending to zero as the gamma distribution tends to normality.

* * * FIGURE 2 ABOUT HERE * * *

2.3.2 The Weibull Distribution. The Weibull density is proportional to $x^{\beta-1}e^{-x^\beta}$, $\beta > 0$. Again, our skewness functions are defined for $\beta > 1$ and calculated numerically. $\gamma^*(p)$ is shown for $\beta = 2, 3, 3.6, 5, 10$ and 100 in Figure 3. We observe (again) that these skewness functions are ordered, the Weibull distribution skewness passing from being positive for smaller β to being essentially negative for larger β. The skewness function for $\beta = 3.6$ was chosen for display because it is approximately the value at which the classical third-moment skewness summary measure changes from being positive to being negative (Dubey, 1967, Johnson, Kotz and Balakrishnan, 1994, Section 21.2). As far as is visible on Figure 3, our skewness function describes a small negative skewness for all p when $\beta = 3.6$. The words 'essentially' and 'visible' were used here for the negativity of the Weibull skewness function because, as noted at the end of Section 2.2.2, $\gamma^*(0+) = 1$. However, we note from Figures 2 and, especially, 3 that this effect can be 'very asymptotic' and not of practical importance. This is because the density for large x in the (semi-)infinite support case can be too small for its effect to be noticed in numerical practice.

* * * FIGURE 3 ABOUT HERE * * *

3. KURTOSIS

3.1 Kurtosis Functions

Kurtosis is here conceived separately for the left and right parts of (appropriate subclasses of) unimodal densities f in $\mathcal{F}(a, b)$, these concepts coinciding in the symmetric case to form a single kurtosis function. By reflection in the mode, it suffices to consider right kurtosis. This, in turn, is directly and straightforwardly defined as skewness of an appropriate unimodal function of the right density derivative f_R' or, for short, kurtosis is gradient skewness. Dividing the right density into its peak and tail parts at the mode of such a function, (right) kurtosis is immediately seen as a corresponding tail-peak comparison. Two alternative functions of f_R' are of particular interest, but introduction of the second of these is delayed until Section 3.4.

The first of our kurtosis functions is motivated by considering Figure 4 which plots f_R for the density of Figure 1. Over this region (m, b), the gradient f_R' is continuous and negative, while its magnitude $-f_R'$ is strictly increasing up to a finite maximum, achieved at an intermediate point π_R, and strictly decreasing thereafter, its limiting values at both endpoints being zero. In other words, $-f_R'$ is a positive multiple of a density in $\mathcal{F}(m, b)$ (since p_R is a survival function, this multiple is $f(m)$). Decomposing f_R into its right peak f_{RP} and right tail f_{RT}, defined as its restrictions to (m, π_R) and to (π_R, b) respectively, it is natural to compare, for any $0 < p < 1$, the distances from π_R of the unique points, $x_{RP}(p)$ and $x_{RT}(p)$ say, in these sub-regions at which the negative slope is the same fraction, p, of its maximum value $-f_R'(\pi_R)$. Recalling that skewness is unaffected by positive multiplication of a density (Section 2.2.1), this first, natural idea is to define the right kurtosis of such a density f as the skewness of $-f_R'$.

* * * Figures 4 and 5 about here * * *

Again, Figure 4, in terms of f_R, translates directly to Figure 5, in terms of $-f'_R$. And Figure 5 is a direct analogue of Figure 1, reflecting the fact that kurtosis is gradient skewness. The role of m in the skewness case is now taken by π_R, which is the right point of inflection of f; the roles of a and b are taken by m and b, respectively; the roles of $x_L(p)$ and $x_R(p)$ are transferred to $x_{RP}(p)$ and $x_{RT}(p)$, respectively; and $\tau_L(p)$ and $\tau_R(p)$ are replaced by $\tau_{RP}(p) = \pi_R - x_{RP}(p)$ and $\tau_{RT}(p) = x_{RT}(p) - \pi_R$, respectively. The prototype right kurtosis function is therefore

$$\kappa_R(p) = \tau_{RT}(p)/\tau_{RP}(p), \qquad 0 < p < 1,$$

and its transformed versions include $\delta_R^0(p) = \log \kappa_R(p)$ and our preferred

$$\delta_R^*(p) = \frac{\kappa_R(p) - 1}{\kappa_R(p) + 1} = \frac{\tau_{RT}(p) - \tau_{RP}(p)}{\tau_{RT}(p) + \tau_{RP}(p)} = \frac{x_{RT}(p) - 2\pi_R + x_{RP}(p)}{x_{RT}(p) - x_{RP}(p)}. \qquad (4)$$

Note that there is no concept of 'flank' or 'shoulder' of f here as there is in some discussions of kurtosis. Nor is there any implication that tails have small probabilities. Note also that kurtosis functions (with restricted domains) could be defined for more general cases without rootedness of f'.

3.2 Theoretical Support

3.2.1 The Classes of Densities for Which Kurtosis is Defined. We denote by

$$\mathcal{F}'_R(a,b) = \{f \in \mathcal{F}(a,b) : f_R \text{ is 1−smooth and } -f'_R/f(m) \in \mathcal{F}(m,b)\}$$

the subclass of $\mathcal{F}(a,b)$ for which our definition of right kurtosis is possible. We call such densities right inflected since each such f is strictly concave on its right peak and strictly convex on its right tail. In the 2-smooth case, $f''(\pi_R) = 0$. Exploiting duality, define the left kurtosis function δ_L^* for left inflected densities in

$$\mathcal{F}'_L(a,b) := \{f \in \mathcal{F}(a,b) : f_L \text{ is 1−smooth and } f'_L/f(m) \in \mathcal{F}(a,m)\}$$

in an exactly analogous way. Both δ_L^* and δ_R^* are defined if and only if $f \in \mathcal{F}'(a,b) = \mathcal{F}'_L(a,b) \cap \mathcal{F}'_R(a,b)$. We say that such f are *inflected* since then f is strictly convex on its left and right tails, defined as (a, π_L) and (π_R, b), respectively, and is strictly concave on its peak region, defined as (π_L, π_R), π_L and π_R being the points of inflection of f. Note that for inflected f, $f'(m)$ is defined and zero (as are $f'(a+)$ and $f'(b-)$). A density $f \in \mathcal{F}(a,b)$ may admit either, neither, or both left and right kurtosis functions, δ_L^* and δ_R^*. For example, if f is the beta density with parameters η_L and η_R, say, then the left kurtosis function is defined for $\eta_L > 2$ and the right kurtosis function for $\eta_R > 2$. When f is symmetric inflected, $\delta_L^* = \delta_R^* = \delta^*$, say. Clearly, $\mathcal{F}'_R(a,b)$ is location and scale equivariant while under reflection in m,

$$f \in \mathcal{F}'_R(a,b) \Leftrightarrow f_m \in \mathcal{F}'_L(b_m, a_m),$$

a natural twin statement applying. Thus, $\mathcal{F}'(a,b)$ is affine equivariant.

For each $f \in \mathcal{F}'_R(a,b)$, let $\widetilde{f}_R = -p'_R = -f'_R/f(m)$ denote the density function corresponding to the survival function p_R. Then δ_R^* contains precisely the information required to move, level by level, between $\widetilde{f}_{RP}$ and $\widetilde{f}_{RT}$ (in an obvious notation), equivalently, given $f(m)$, between f_{RP} and f_{RT}. Defining $\widetilde{f}_L = p'_L = f'_L/f(m)$ for each $f \in \mathcal{F}'_L(a,b)$, exactly similar statements apply to δ_L^*.

Finally in this subsection, we note some further decompositions of $f \in \mathcal{F}'_R(a,b)$, exactly similar ones applying in the twin case. Using, first, rootedness (at b) and, then, continuity (at m) of f, we have successively:

$$f_R \overset{1-1}{\leftrightarrow} -f'_R \overset{1-1}{\leftrightarrow} (f(m), \widetilde{f}_R),$$

giving $f_{RP} \overset{1-1}{\leftrightarrow} (f(m), \widetilde{f}_{RP})$ and its tail analogue. Recall also (Section 2.2.1) that $f \overset{1-1}{\leftrightarrow} (f_L, \gamma) \overset{1-1}{\leftrightarrow} (\gamma, f_R)$ so that, as $\widetilde{f}_R \in \mathcal{F}(m,b)$, $\widetilde{f}_R \overset{1-1}{\leftrightarrow} (\widetilde{f}_{RP}, \delta_R^*) \overset{1-1}{\leftrightarrow} (\delta_R^*, \widetilde{f}_{RT})$. Combining these, we have

$$f \overset{1-1}{\leftrightarrow} (\gamma, \delta_R^*, f_{RP}) \overset{1-1}{\leftrightarrow} (\gamma, \delta_R^*, f_{RT}).$$

For inflected f, given γ, $f_L \overset{1-1}{\leftrightarrow} f_R$ gives $f_{LP} \overset{1-1}{\leftrightarrow} f_{RP}$ and its tail analogue. Thus, for such f, we have overall:

$$f \overset{1-1}{\leftrightarrow} (\gamma, \delta_R^*, f_\bullet) \overset{1-1}{\leftrightarrow} (\gamma, \delta_L^*, f_\bullet)$$

where $f_\bullet$ denotes any of the four part densities f_{LT}, f_{LP}, f_{RP} and f_{RT} of f. In particular, given γ and any $f_\bullet$, δ_L^* and δ_R^* are equivalent.

3.2.2 Properties of Kurtosis Functions. Return now to explicit consideration of the right kurtosis function only. Immediately, each of κ_R, δ_R^0 and δ_R^* are location and scale invariant; δ_R^0 and δ_R^* change sign under reflection in π_R since then tail and peak are interchanged. Since

$$\delta_R^0(p) > (<)\ 0 \Leftrightarrow \delta_R^*(p) > (<)\ 0 \Leftrightarrow \kappa_R(p) > (<)\ 1,$$

we can define right inflected f to have positive (resp. negative) right kurtosis at level p if these inequalities hold for a particular value of p, and to have totally positive (resp. negative) right kurtosis if the inequalities hold for all $0 < p < 1$. Notice that positive (resp. negative) right kurtosis corresponds to the right tail being 'heavy' (resp. 'light') relative to the right peak.

Totally zero total right kurtosis ($\delta_R^0(p) = \delta_R^*(p)$ for all $0 < p < 1$) can be characterised in a number of entirely equivalent ways: (a) $-f_R'$ is symmetric about π_R; (b) p_R is the survival function of a symmetric distribution on (m, b); (c) f_R is an odd function about π_R; (d) f_R has the form of symmetry expressed by $f_R(\pi_R + y) + f_R(\pi_R - y) = 2f(\pi_R) = f(m)$. As m is finite, this can only occur if b is also finite, in which case $m + b = 2\pi_R$. A symmetric f with totally zero kurtosis is therefore a density on finite (a, b) of the form

$$f(x) \propto G(x)I(x \leq m) + G(x_m)I(x > m)$$

where G is the distribution function of a symmetric distribution on (a, m), $m = (a + b)/2$ so that $x_m = a + b - x$, and $G(x_m) = 1 - G(a + x - m)$.

Further analogues of properties of skewness functions given in Section 2.2.2 also arise immediately for right kurtosis functions by replacing f by $-f'_R$, but full details do not warrant repetition here. Suffice it to say that constant non-zero right kurtosis right parts of densities are themselves of two-piece form on (m, b) and, therefore, that densities which have constant skewness and equal, constant, right and left kurtosis are four-piece distributions with joins at π_L, m and π_R. (Three different constants for skewness, left and right kurtosis are not possible because of properties given in the previous subsection.) In addition, whenever f_R is 3-smooth and $f'''_R(\pi_R) < 0$, $\kappa_R(1-) = 1$, $\delta^0_R(1-) = 0$ and $\delta^*_R(1-) = 0$. It is also the case that $\kappa_R(0+) = \infty$, $\delta^0_R(0+) = \infty$ and $\delta^*_R(0+) = 1$ when the support of f is either $\mathcal{R}$ or $(0, \infty)$. When b is finite, $\kappa_R(0+) = (b - \pi_R)/(\pi_R - m)$.

3.3 Examples of Kurtosis Functions

3.3.1 The t and Symmetric Beta Distributions. The Student's t distribution on $\mathcal{R}$ with $\nu > 0$ degrees of freedom has (scaled) density proportional to $(1+x^2)^{-(\nu+1)/2}$. Its limit as $\nu \to \infty$ is, of course, the normal distribution and, as ν becomes small, the t distribution acquires very heavy tails (it includes the Cauchy distribution when $\nu = 1$). The symmetric beta distribution on $(-1, 1)$ with parameter $\eta > 0$ has density proportional to $(1 - x^2)^{\eta-1}$. The distribution is uniantimodal for $\eta < 1$, is the uniform density for $\eta = 1$ and is unimodal for $\eta > 1$, also tending to the normal distribution as $\eta \to \infty$. The symmetric beta, normal and t distributions are the symmetric members of the Pearson family of distributions. Notice that we can speak of kurtosis here rather than just of right kurtosis because the symmetry of the distributions means that left kurtosis = right kurtosis = kurtosis.

* * * FIGURE 6 ABOUT HERE * * *

The kurtosis functions of the t (solid lines), normal (thicker solid line) and symmetric beta distributions (dotted lines) are shown in Figure 6. They were calculated numerically. The main impression given by Figure 6 is that these kurtosis functions are ordered: the t distributions with small ν are most kurtotic, with kurtosis decreasing as ν increases; then, starting from the kurtosis of the normal distribution, kurtosis continues to decrease as η decreases in the symmetric beta distributions. Notice that these comparisons appear to hold totally for all values of p. The kurtosis function does not exist for $\eta \leq 2$ beyond which, in our sense, the symmetric beta distribution has no tail (it is 'all peak').

3.3.2 The Exponential Power Distribution. The exponential power distribution has density proportional to $\exp(-|x|^\beta)$, $x \in \mathcal{R}$, $\beta > 0$. It reduces to the Laplace and normal distributions for $\beta = 1$ and $\beta = 2$, respectively, and is very short tailed for large β. Its kurtosis function has already been shown in Figure 3. This is because the (right) kurtosis is the skewness of the negative derivative of (the right part of) f, and the negative derivative of (the right part of) f is proportional to the Weibull density whose skewness function is plotted in Figure 3. Note, therefore, that the kurtosis function of the exponential power distribution exists only for $\beta > 1$ and is a decreasing function of β for all p, eventually becoming negative.

3.3.3 The Gamma Distribution. As an example of left and right kurtosis functions for an asymmetric density, we return to the gamma distribution with parameter α. We find it convenient to draw $\kappa_L(p)$ as a function of what is labelled $-p$ leftwards along the negative axis. Left kurtosis exists for $\alpha > 2$, while right kurtosis exists for $\alpha > 1$. The gamma kurtosis functions are drawn for the skew case of $\alpha = 3$ (solid lines) and the fairly symmetric case of $\alpha = 100$ (dotted lines) in Figure 7. The $\alpha = 100$ case yields a right kurtosis function that is essentially that of the normal distribution but, interestingly, the left kurtosis function is a little lower indicating that any discrepancies that remain from the gamma's limiting case are towards the left of the mode (no doubt because of the finite lower limit of the gamma's support). When $\alpha = 3$, however, the left kurtosis is quite different and indeed negative, indicating a very light lower tail relative to the peak.

* * * FIGURE 7 ABOUT HERE * * *

3.4 The Alternative Kurtosis Function

3.4.1 Definition, Rationale and Properties. Denote by d the unimodal function of the density derivative on which kurtosis is based, and which up to now has been $d(x) = d_1(x) = |f'(x)|$. In this section, we introduce the alternative d function $d_2(x) = -(x-m)f'(x)$. Notationally, whenever we need to refer specifically to elements of kurtosis functions based on d_1 or d_2 we will incorporate a '1' or '2' into the appropriate subscript e.g. $d_{1,R}$, $x_{2,RT}(p)$, $\delta^*_{1,R}(p)$. An entirely analogous figure to Figure 5 arises when $d_{2,R}$ replaces $d_{1,R}$. By the way, it is easy to see that $\pi_{2,R} \geq \pi_{1,R}$.

Why do we consider basing kurtosis on d_2? Well, any unimodal density is 'singly Khintchine' in the sense that if $X \sim f$, then $X =^d m + U_{(1)}W$ where $U_{(1)} \sim U(0,1)$ and $W \sim g$ where $g(w) = -wf'(w+m)$ are independent; here $=^d$ denotes 'has the same distribution as'. This is Khintchine's theorem (Khintchine, 1938, Shepp, 1962, Feller, 1971, Jones, 2002) which gives the natural link between unimodality of f and consideration of the density derivative-based function $d_2(x) = g(x-m)$. When $d_{2,L}$ and $d_{2,R}$ are themselves unimodal and rooted, we say that f is doubly Khintchine; this is the d_2 analogue of f being inflected and, of course, $f \in \mathcal{F}(a,b)$ can otherwise admit either, neither, or both left and right kurtosis functions. For doubly Khintchine densities, if $W_R \sim g_R(w) = -wf'(w+m)/(1-F(m)), w > 0$, ($F$ being the distribution function of f) is 1-smooth (equivalently, f_R is 2-smooth), then, treating g_R as the unimodal density of interest, $W_R =^d U_{(2)}Z$ where $U_{(2)} \sim U(0,1)$ independently of $Z \sim -zg'_R(z), z > 0$. Defining $X_R = (X-m)I(X>m)$, combination of the Khintchine relationships for f and g_R

results in $X_R =^d VZ$ where $V \sim -\log(v),\ 0 < v < 1$, and $Z \sim z\{f'(z+m)$ $+zf''(z+m)\}/(1-F(m))$.

Returning to W_R, the random variable associated with g_R, note that it has the distribution of $\tau_R(P) = x_R(P) - m$ where

$$P = \frac{Y}{f(m)} \quad \text{and} \quad Y \sim \frac{f_R^{-1}(y) - m}{1 - F(m)}\, I(0 < y < f(m)), \tag{5}$$

the latter being none other than the Y-marginal distribution when (X, Y) are uniformly chosen from the region bounded by f_R and the axes. The X-marginal of this bivariate uniform distribution is proportional to f_R and the bivariate distribution is fundamental to random variate generation, as argued by Jones (2002). From this viewpoint, this natural choice of distribution for Y contrasts favourably with that leading to $-f_R'(x)/f(m),\ x > m$, which is the distribution of $x_R(P)$ when $P \sim U(0,1)$.

The whole range of properties given in Section 3.2 specifically for $d = d_1$ have immediate analogues for $d = d_2$ but these will not be explicitly given here.

3.4.2 Examples. Replacing d_2 by d_1 in the examples of Section 3.3 makes for relatively minor changes, at least qualitatively, in general. For example, a very similar figure (not shown) to Figure 6 is obtained for t, normal and symmetric beta distributions: shapes of kurtosis functions are much as they are for d_1 with perhaps a wider spread over t's (low degrees of freedom more kurtotic, normal less kurtotic) while betas are somewhat squashed down for larger a. The range of existence is the same in both cases.

In the exponential power case, swapping d_2 for d_1 increases the range of existence of the kurtosis function all the way to $\beta > 0$. In particular, the case $\beta = 1$ which for d_1 was disallowed now affords a kurtosis function for the Laplace distribution (equivalent to a right kurtosis function for the exponential distribution). While for d_1, the right kurtosis of the exponential power distribution is the skewness of the Weibull distribution, for d_2 the right kurtosis is the skewness of the distribution with density proportional to $x^\beta e^{-x^\beta}$, sometimes called the pseudo-Weibull distribution (Murthy, Xie and Jiang, 2004, Section 7.2). Figure 8 displays the d_2 kurtosis for a range of exponential power distributions for two further reasons: (i) to show the high kurtosis attributed to the exponential distribution and (ii) to demonstrate the qualitative similarity between d_1 and d_2 kurtosis functions in other cases.

* * * FIGURE 8 ABOUT HERE * * *

For the gamma distribution, conditions on existence of left kurtosis are the same for d_1 and d_2, but right kurtosis exists for d_2 if $\alpha \geq 1$, thereby also accommodating the exponential distribution.

3.5 Relative Kurtosis

Absolute skewness and relative skewness using either $\gamma(p)$ measure are the same thing because skewness is naturally measured relative to symmetry for which $\gamma(p) \equiv 0$. It is not so clear that totally zero kurtosis necessarily has the same role as zero skewness. One might prefer to measure kurtosis relative to some other distribution's kurtosis. This is often done for the classical moment-based measure where the value 3, associated with the normal distribution, is subtracted off to give a measure of kurtosis relative to that of the normal distribution.

Let $\delta_R \equiv \delta_R^f$ be a given (absolute) right kurtosis function and, for some suitable $h(\cdot|\cdot)$, let $\delta_R^{f|f_1}$ defined by

$$\delta_R^{f|f_1}(p) = h(\delta_R^f(p)|\delta_R^{f_1}(p)), \qquad 0 < p < 1,$$

denote a corresponding relative right kurtosis function for f with respect to f_1. Two requirements will fix the function $h(\cdot|\cdot)$. First, it is reasonable to require that working relative to a density whose kurtosis vanishes at level

p should leave absolute kurtosis unchanged at that level, i.e. $h(u|0) = u$. Second, there should be an explicit way to transform, level-by-level, kurtosis measured relative to different reference densities which would most simply be $h(u|u_2) = h(u|u_1) + h(u_1|u_2)$. The unique function satisfying these requirements is $h(u|u_1) = u - u_1$ so that

$$\delta_R^{f|f_1}(p) = \delta_R^f(p) - \delta_R^{f_1}(p), \qquad 0 < p < 1.$$

Further intuitive and interpretable properties of the relative kurtosis follow immediately:

(i) reciprocity: $h(u_2|u_1) = -h(u_1|u_2)$;
(ii) self-nullity: $h(u|u) = 0$;
(iii) inflection-equivariance: $h(-u|-u_1) = -h(u|u_1)$;
(iv) monotonicity: $u > u_1 \Rightarrow h(u|u_1) > 0$.

So if, at level p, the tail of f is 'heavier' relative to the peak of f than is the tail of f_1 relative to its peak, then f has positive kurtosis relative to f_1 at that level. It is sometimes useful to compare left and right part behaviour vis-à-vis kurtosis via $\delta_R^f(p) - \delta_L^f(p)$, the right kurtosis of f relative to its left kurtosis.

4. ESTIMATING SKEWNESS AND KURTOSIS FUNCTIONS: AN EXAMPLE

In this section, we will indicate the kinds of issues involved with estimating our skewness and kurtosis functions and illustrate the results of an initial implementation with pragmatic choice of details. There is scope for much theoretical and practical work on estimation issues that we barely touch on. The principal tool is kernel density estimation (Silverman, 1986, Wand and Jones, 1995) where the density f is estimated by

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

Here, $X_1, ..., X_n$ is a random sample from f, K will be taken to be the standard normal density function and h is the smoothing parameter, also called the bandwidth, which will be estimated below. Estimated skewness and kurtosis functions will be obtained for 'sample 1' of Table 1 of Smith and Naylor (1987); here, $n = 63$ and the data are breaking strengths of 1.5cm long glass fibres, originally obtained at the UK National Physical Laboratory.

4.1 Estimating the Skewness Function

The ingredients of $\gamma^*(p)$, see (1), are $x_R(p)$, $x_L(p)$ and m. The first two of these are values of the inverse of f_R and f_L at a value dependent on $f(m)$. Most of these ingredients are directly dependent on f and therefore a value of h appropriate to estimating f itself rather than any other functional of f is suggested. (The unusual step of inverting f makes no difference in this regard; see Jones, 2000, for a relatively non-technical introduction to the interplay between bandwidth choice and functionals of f.) But the mode also plays an important role in γ^* and, optimally, estimation of the mode requires an order of magnitude larger value for h because of its close link with estimation of f' through $f'(m) = 0$ (Müller, 1984, Jones, 2000). Yet we need to use the same bandwidth for each element of $\gamma^*(p)$ if a coherent skewness function is to be obtained. Our compromise between these requirements is to use the simple rule-of-thumb $h = h_\gamma = s\{4/(3n)\}^{1/5}$, where s is the sample standard deviation (Silverman, 1986). This arises from the formula for the bandwidth that minimises asymptotic integrated mean squared error for density estimation by assuming f to be a normal distribution. Typically, this rule-of-thumb oversmooths a little in terms of estimating the density per se.

There is, however, a further requirement: skewness is defined only for unimodal distributions. If the underlying f is unimodal, $\hat{f}$ using h_γ will typically be unimodal too and this is the case for the glass fibre data; see Figure

9(a). If $\hat{f}_{h_\gamma}$ is not unimodal it can be made so by increasing h (monotonically reducing the number of modes in the case of the normal kernel, Silverman, 1981). A general strategy might be to utilise Silverman's (1981) test of unimodality which depends on the size of the smallest $h = h_c$ necessary to obtain a single mode (see also Fisher, Mammen and Marron, 1994, Fraiman and Meloche, 1999, Hall and York, 2001). If the test accepts unimodality, then use $h = h_\gamma$ if $\hat{f}_{h_\gamma}$ is unimodal and $h = h_c$ otherwise; if Silverman's test rejects unimodality, do not proceed. (For more sophisticated methods of forcing smooth unimodality see, e.g., Bickel and Fan, 1996, Eggermont and La Riccia, 2000, Hall and Huang, 2002, Hall and Kang, 2005).

* * * FIGURE 9 ABOUT HERE * * *

A quick glance at Figure 9(a) suggests left or negative skewness. But what is meant by that in this case? The skewness function $\gamma^*(p)$ corresponding to $\hat{f}_{h_\gamma}$ is shown in Figure 9(b). The skewness function is only very slightly negative over most of the range of (larger) p. This corresponds to an almost-symmetry of the main body of the density estimate. (The 'glitch' in Figure 9(b) near $p = 1$ appears to be a numerical problem.) Negative skewness is stronger for (approximately) $0 < p < 0.16$. This reflects the strong 'bump' in the left hand tail of $\hat{f}_{h_\gamma}$: it is the presence of this bump that causes an overall impression of negative skewness.

Note that kernel estimation is not reliable in the far tails of a distribution and both it and Silverman's test of unimodality would be strongly affected by isolated points in the tails, so estimation of $\gamma(p)$ is generally not to be trusted for very small p.

4.2 Estimating the Kurtosis Functions

Once again, because our kurtosis functions are just skewness functions applied to $|f'|$ or $-(x - m)f'$ rather than f, the technology for estimating skewness functions transfers pretty much directly to estimating kurtosis functions. In particular, our starting point is kernel density derivative estimation given by

$$\hat{f}'_h(x) = \frac{1}{nh^2}\sum_{i=1}^{n} K'\left(\frac{x - X_i}{h}\right)$$

for K the standard normal kernel. The rule-of-thumb bandwidth for estimation of the first derivative of the density, for use when $d = d_1$, is $h = h_{1,\delta} = s\{4/(5n)\}^{1/7}$. A similar calculation for estimation of $-(x-m)f'$ with m taken to be known, for use when $d = d_2$, yields the slightly different value $h = h_{2,\delta} = s\{8/(11n)\}^{1/7}$. For the case of the glass fibre data, our figures pertain to taking $d = d_1 = |f'|$. Then, $h_{1,\delta} = 0.1737$ but $\hat{f}'$ to the left of the mode is not itself unimodal. This reflects the fact that $\hat{f}$ in Figure 9(a), admittedly based on a smaller bandwidth, is not inflected. Multiplying $h_{1,\delta}$ by 1.35 turns out to (approximately) yield h_c for f', and the corresponding $|f'|$ is plotted in Figure 10(a).

* * * FIGURE 10 ABOUT HERE * * *

The left and right kurtosis functions based on $|\hat{f}'_{h_c}|$ are plotted in Figure 10(b). It should be noted that, in general, left and right kurtosis functions deserve separate left and right bandwidths. In this example, one could use h_c for the left kurtosis function and $h_{1,\delta}$ for the right kurtosis function, but the pictorial difference would be negligible in this case. Moreover, extension of testing procedures to test for unimodality of left and right parts of $|f'|$ — and to declare left or right kurtosis undefined if the corresponding test is failed — is warranted, but not yet pursued. (Kernel estimation of left and right kurtosis functions separately by dividing the dataset depending on position relative to the estimated mode, $\hat{m}$, seems to offer no advantages because of the consequent need to allow for the boundary introduced at $\hat{m}$.) The right kurtosis function is very reminiscent of those for t distributions with moderate degrees of freedom (Figure 6). The left kurtosis function is similar for large p but increases much more rapidly at around $p = 0.24$, again reflecting the 'bump' in the left hand tail of this distribution. When $d = d_2$, similar plots (not shown) accentuate the left hand bump rather more and give a larger left kurtosis function for more (small) values of p.

The glass fibre data have recently been used to illustrate the fitting of various skew t distributions (Jones and Faddy, 2003, Ferreira and Steel, 2004). These four-parameter distributions fit much of the dataset well but treat the bump simply as a heavy tail. This is defensible given the relatively small size of the dataset and is sufficient for most purposes. But more data would be needed to shed further light on whether there is really a small 'second group' or just a more widespread heavy tail to the left.

5. SKEWNESS AND KURTOSIS SCALARS

A theme of this paper is the functional nature of skewness and kurtosis. Nonetheless, there is still some role for scalar measures of skewness and kurtosis and it is natural to provide them by some appropriate averaging of the skewness, generically $\gamma(p)$, and left and right kurtosis, generically $\delta_L(p)$ and $\delta_R(p)$, functions. So, define the skewness measure

$$\gamma = \int_0^1 \gamma(p) w_f(p) dp$$

and the right kurtosis measure

$$\delta_R = \int_0^1 \delta_R(p) w_d(p) dp$$

(likewise the left kurtosis measure) for some density functions w_f and w_d on $(0, 1)$ (which might be the same in which case we write them as w). Immediately, γ and δ_R have the same range as the functions $\gamma(p)$ and $\delta(p)$. If f is symmetric, then $\gamma(p) = 0$ for all p and so $\gamma = 0$. However, γ can also be zero in cases where positive and negative parts of $\gamma(p)$ cancel out, a disadvantage of insisting on scalar measures.

One obvious choice for w is the Dirac delta function at, say, p_0; that is, take $\gamma(p_0)$ and $\delta_R(p_0)$ as scalar summaries of the whole γ or δ_R functions. A particularly natural choice might be the median-type choice $p_0 = 1/2$. So, for example, skewness might be measured by

$$\gamma^*(1/2) = \frac{x_R(1/2) - 2m + x_L(1/2)}{x_R(1/2) - x_L(1/2)}.$$

The denominator is the full width at half maximum scale measure mentioned in Section 2.2.1.

Another obvious choice is to take a uniform average, $w^{11}(p) = 1$, $0 < p < 1$. The complementary choices of Beta$(2,1)$ and Beta$(1,2)$ densities, $w^{21}(p) = 2p$ and $w^{12}(p) = 2(1-p)$, $0 < p < 1$, put more (resp. less) weight where the density or density derivative is larger. Adopting an obvious abbreviation, the three corresponding scalar skewness measures are related through $\gamma^{11} = (\gamma^{21} + \gamma^{12})/2$ (similarly for scalar kurtosis measures). Weight functions that put more weight where the density is larger will be better estimated from data. One might therefore consider pursuing this by taking w as the density of $f(X)$ (Troutt, Pang and Hou, 2004). However, being an integrated quantity, even the natural uniform choice, e.g. $\gamma = \int_0^1 \gamma(p)dp$, yields a quantity which is much more robust to specific choice of bandwidth (Jones, 2000) than is any unaveraged skewness or kurtosis function.

For the remainder of this section, let ℓ stand for densities f or normalised $d_{1,R}$ or $d_{2,R}$ and L for the associated distribution function. That is, $\ell = f$, $-f'(x)/f(m)$ or $-(x-m)f'(x)/\{1-F(m)\}$ and $L = F$, $1-\{f(x)/f(m)\}$ and $\{F(x) - F(m) - (x-m)f(x)\}/\{1 - F(m)\}$, respectively. Let π denote the corresponding mode $m, \pi_{1,R}$ or $\pi_{2,R}$ in each of these cases. Inspired by Section 3.4.1, there are further natural choices for w, specifically that associated with $P = Y/\ell(\pi)$ when (X, Y) are uniformly chosen from the region bounded by ℓ and the horizontal axis. Taking into account both left and right parts of the unimodal density ℓ, it is readily seen that

$$Y \sim \{\ell_R^{-1}(y) - \ell_L^{-1}(y)\}\ I(0 < y < \ell(\pi)),$$

that is, Y's density is the scale function associated with ℓ, and hence that

$$P \sim w_\ell(p) = \ell(\pi)\{\ell_R^{-1}(p\ell(\pi)) - \ell_L^{-1}(p\ell(\pi))\}\ I(0 < p < 1).$$

This makes for attractive simplifications as follows.

Let $x_R(p)$, $x_L(p)$ and $\sigma(p)$ also refer to any version of ℓ and let $\psi(p)$ and ψ denote either $\gamma(p)$ and γ or $\delta_R(p)$ and δ_R. Provided they exist for all $0 < p < 1$, it is easy to see that

$$\int_0^1 x_R(p)dp = \pi + \frac{(1 - L(\pi))}{\ell(\pi)} \quad \text{and} \quad \int_0^1 x_L(p)dp = \pi - \frac{L(\pi)}{\ell(\pi)}.$$

It follows that

$$\int_0^1 \sigma(p)dp = \frac{1}{\ell(\pi)}$$

(implicit above) which reduces to (2) when $\ell = f$. In addition,

$$\psi = \int_0^1 \psi(p)w_\ell(p)dp = \ell(\pi)\int_0^1 \{x_R(p) - 2\pi + x_L(p)\}dp = 1 - 2L(\pi).$$

In all cases, $\psi \to 1$ (resp. -1) as the mode π of ℓ tends to the lower (resp. upper) end of its support.

The scalar skewness measure that arises from these considerations is nothing other than

$$\gamma = 1 - 2F(m),$$

the Arnold and Groeneveld (1995) measure, for unimodal distributions with $f \to 0$ at both support endpoints, and is undefined otherwise (including, for example, for the exponential distribution, cf. Section 2.3.1).

The corresponding scalar kurtosis measures are novel. In the d_1 case,

$$\delta_{1,R} = 2\,\frac{f(\pi_R)}{f(m)} - 1.$$

This is an intriguing simple scalar kurtosis measure. This kurtosis is zero if the point of inflection has density one-half the density at the mode. Otherwise, it makes a very simple tail/peak comparison by being more and more positive (resp. negative) the larger (resp. smaller) the density is at the point of inflection relative to the density at the mode. Alternatively,

$$\delta_{2,R} = 1 - 2\ \frac{\{F(\pi_R) - F(m) - (\pi_R - m) f(\pi_R)\}}{1 - F(m)}.$$

6. CONCLUSIONS AND FURTHER DEVELOPMENTS

The main claims of this paper are that:

(i) Skewness and kurtosis, in their fullest senses, are functional concepts not scalars. This is not entirely new. For example, the final version of (1) is reminiscent of a quantile-based measure in which $x_R(p)$ is replaced by $F^{-1}(1-p)$, $x_L(p)$ by $F^{-1}(p)$ and m by the median (Hinkley, 1975, Groeneveld and Meeden, 1984). However, such measures are typically reduced to scalar measures by specific choice of p or by some kind of averaging and are rarely treated as functions of p per se; Benjamini and Krieger (1996) is one exception.

(ii) Our skewness function is the first to be defined directly — and hence immediately interpretably — in terms of the probability density function. Our kurtosis functions are defined simply in terms of the density derivative (which in one case, at least, translates readily back to interpretation in terms of the density function itself).

(iii) Skewness and kurtosis are well defined concepts only for unimodal distributions.

(iv) Left and right kurtosis are defined separately, left and right parts of a unimodal density being defined by the position of the mode. For symmetric densities, left and right kurtosis functions coincide to form a single kurtosis function.

(v) Left and right kurtosis are defined directly as skewness of very simple functions of the left and right parts of the density derivative (of what we have called either inflected or doubly Khintchine densities). This is vaguely reminiscent of skewness quantile measures which are applied to halves of the distribution, where half is defined by the median (Groeneveld, 1998, Brys et al., 2005).

(vi) We have a straightforward definition of left and right kurtosis in terms of a tail-peak comparison where, for example in the case of right kurtosis, the right tail and right peak are simply and explicitly defined. In particular, for right inflected densities, the right tail is the region between the right hand point of inflection and the right hand end of the support, and the right peak is the region between the mode and the right hand point of inflection.

(vii) A scalar skewness measure that arises as a natural average of our skewness function is $1 - 2F(m)$, the popular measure of Arnold & Groeneveld (1995). The analogous scalar kurtosis measures are novel.

(viii) Our skewness and kurtosis functions come complete with a natural location measure, the mode m, and a natural (overall) measure of scale, $1/f(m)$. It is then also natural to think of a collection of four items (two scalars, two functions) such as $\{m, 1/f(m), \gamma^*(p), \delta_R^*(p)\}$ as a useful set of summary descriptors analogous to familiar sets of scalar summaries based, say, on the first four moments. Indeed, we can extend further to sets of measures from which the density function f itself can be reconstructed, examples including:

$$f \overset{1-1}{\leftrightarrow} (m, \tau_L, \tau_R) \overset{1-1}{\leftrightarrow} (\gamma, f_R) \overset{1-1}{\leftrightarrow} (\gamma, \delta_R, f_\bullet)$$

where $f_\bullet$ is as defined in Section 3.2.1.

The ideas and methodology presented here can be developed in a variety of directions, including:

(a) practical implementation, involving refinement of smoothing procedures;

(b) inferential uses such as testing symmetry;

(c) uniantimodal densities;

(d) multivariate densities

and

(e) higher order derivatives.

This last extends the maxim with which we finish: *kurtosis is gradient skewness*.

REFERENCES

Arnold, B.C., and Groeneveld, R.A. (1995), "Measuring Skewness With Respect to the Mode," *The American Statistician*, 49, 34–38.

Averous, J., Fougères, A.L., and Meste, M. (1996), "Tailweight With Respect to the Mode for Unimodal Distributions," *Statistics and Probability Letters*, 28, 367–373.

Balanda, K.P., and MacGillivray, H.L. (1988), "Kurtosis: A Critical Review," *The American Statistician*, 42, 111–119.

Balanda, K.P., and MacGillivray, H.L. (1990), "Kurtosis and Spread," *Canadian Journal of Statistics*, 18, 17–30.

Benjamini, Y., and Krieger, A.M. (1996), "Concepts and Measures for Skewness with Data-Analytic Implications," *Canadian Journal of Statistics*, 24, 131–140.

Bickel, P.J., and Fan, J.Q. (1996), "Some Problems on the Estimation of Unimodal Densities," *Statistica Sinica*, 6, 23–45.

Brys, G., Hubert, M., and Struyf, A. (2005), "Robust Measures of Tail Weight," *Computational Statistics and Data Analysis*, to appear.

Darlington, R.B. (1970), "Is Kurtosis Really "Peakedness?"," *The American Statistician*, 24, 2, 19–22.

Dubey, S.D. (1967), "Normal and Weibull Distributions," *Naval Research Logistics Quarterly*, 14, 69–79.

Eggermont, P.P.B., and La Riccia, V.N. (2000), "Maximum Likelihood Estimation of Smooth Monotone and Unimodal Densities," *Annals of Statistics*, 28, 922–947.

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

Fernández, C., and Steel, M.J.F. (1998), "On Bayesian Modelling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

Ferreira, J.T.A.S., and Steel, M.J.F. (2004), "A Constructive Representation of Univariate Skewed Distributions," Technical Report, Department of Statistics, University of Warwick, U.K.

Fisher, N.I., Mammen, E., and Marron, J.S. (1994), "Testing for Multimodality," *Computational Statistics and Data Analysis*, 18, 499–512.

Fraiman, R., and Meloche, R. (1999), "Counting Bumps," *Annals of the Institute of Statistical Mathematics*, 51, 541–569.

Groeneveld, R.A. (1998), "A Class of Quantile Measures for Kurtosis," *The American Statistician*, 51, 325–329.

Groeneveld, R.A., and Meeden, G. (1984), "Measuring Skewness and Kurtosis," *The Statistician*, 33, 391–399.

Hall, P., and Huang, L.S. (2002), "Unimodal Density Estimation Using Kernel Methods," *Statistica Sinica*, 12, 965–990.

Hall, P., and Kang, K.H. (2005), "Unimodal Kernel Density Estimation by Data Sharpening," *Statistica Sinica*, 15, 73–98.

Hall, P., and York, M. (2001), "On the Calibration of Silverman's Test for Multimodality," *Statistica Sinica*, 11, 515–536.

Hinkley, D.V. (1975), "On Power Transformations to Symmetry," *Biometrika*, 62, 101–111.

Hosking, J.R.M. (1992), "Moments or L-Moments? An Example Comparing Two Measures of Distributional Shape," *The American Statistician*, 46, 186–189.

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

Jones, M.C. (2000), "Rough-and-Ready Assessment of the Degree and Importance of Smoothing in Functional Estimation," *Statistica Neerlandica*, 54, 37–46.

Jones, M.C. (2002), "On Khintchine's Theorem and Its Place in Random Variate Generation," *The American Statistician*, 56, 304–307.

Jones, M.C. (2005), "A Note on Rescalings, Reparametrizations and Classes of Distributions," *Journal of Statistical Planning and Inference*, to appear.

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, 159–74.

Khintchine, A.Y. (1938), "On Unimodal Distributions," *Izvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki*, 2, 1–7.

Mudholkar, G.S., and Hutson, A. (2000), "The Epsilon-Skew-Normal Distribution for Analyzing Near-Normal Data," *Journal of Statistical Planning and Inference*, 83, 291–309.

Müller, H.G. (1984), "Smooth Optimal Kernel Estimators of Densities, Regression Curves and Modes," *Annals of Statistics*, 12, 766–774.

Murthy, D.N.P., Xie, M., and Jiang, R. (2004), *Weibull Models*, Hoboken, NJ: Wiley.

Oja, H. (1981), "On Location, Scale, Skewness and Kurtosis of Univariate Distributions," *Scandinavian Journal of Statistics*, 8, 154–168.

Ruppert, D. (1987), "What is Kurtosis? An Influence Function Approach," *The American Statistician*, 41, 1–5.

Serfling, R. (2004), "Nonparametric Multivariate Descriptive Measures Based on Spatial Quantiles," *Journal of Statistical Planning and Inference*, 123, 259–278.

Shepp, L.A. (1962), "Symmetric Random Walk," *Transactions of the American Mathematical Society*, 104, 144–153.

Silverman, B.W. (1981), "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society, Series B*, 43, 97–99.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Smith, R.L., and Naylor, J.C. (1987), "A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull Distribution," *Applied Statistics*, 36, 358–369.

Troutt, M.D., Pang, W.K., and Hou, S.H. (2004), *Vertical Density Representation and Its Applications*, New Jersey: World Scientific.

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

Wand, M.P., and Jones, M.C. (1995), *Kernel Smoothing*, London: Chapman and Hall.

Figure 1. A Unimodal Density f on (a, b) With Mode m. The horizontal dashed line is at height $pf(m)$ and this defines the points $x_L(p)$, $x_R(p)$ and the distances $\tau_L(p)$ and $\tau_R(p)$ as shown.

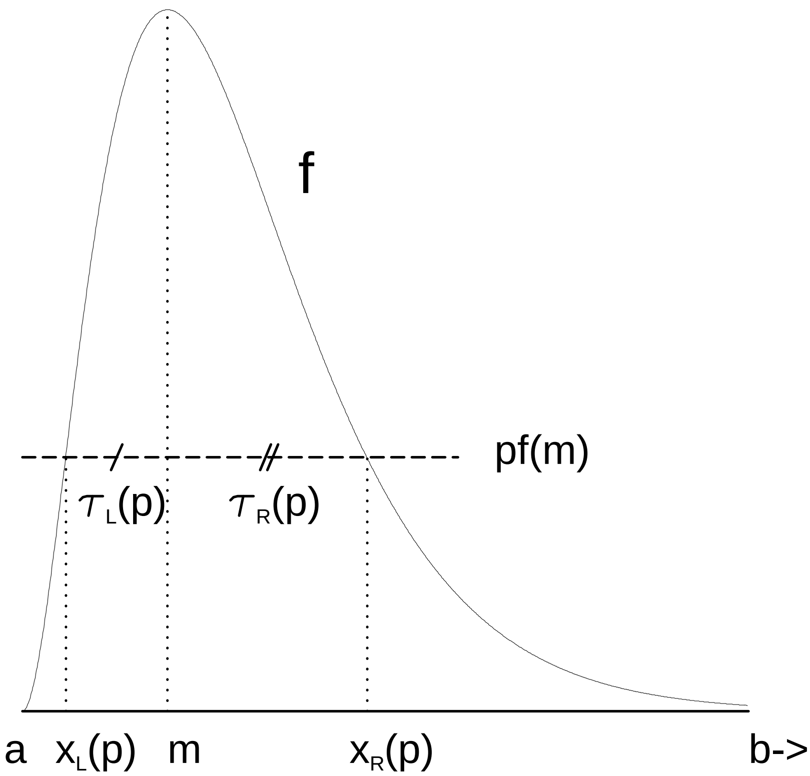

Figure 2. The Skewness Function $\gamma^*(p)$, $0 < p < 1$, *for Gamma Densities With Parameter, in Order of Decreasing Value of Skewness,* $\alpha = 2, 3, 5, 10, 20$ *and* 100.

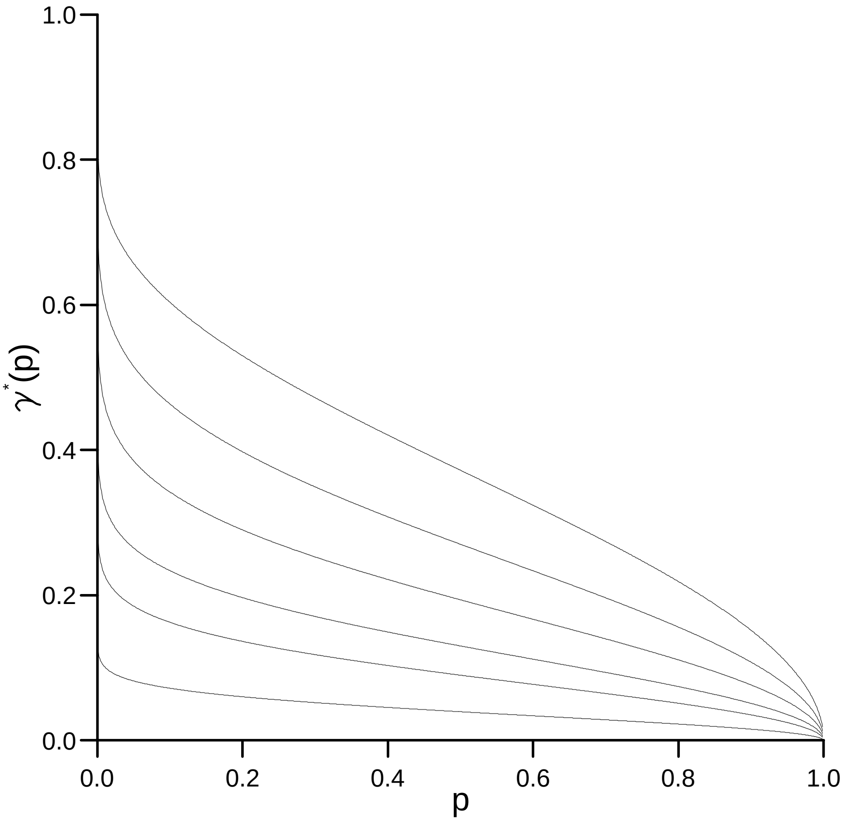

Figure 3. The Skewness Function $\gamma^*(p)$, $0 < p < 1$, *for Weibull Densities With Parameter, in Order of Decreasing Value of Skewness,* $\beta = 2, 3, 3.6, 5, 10$ *and* 100.

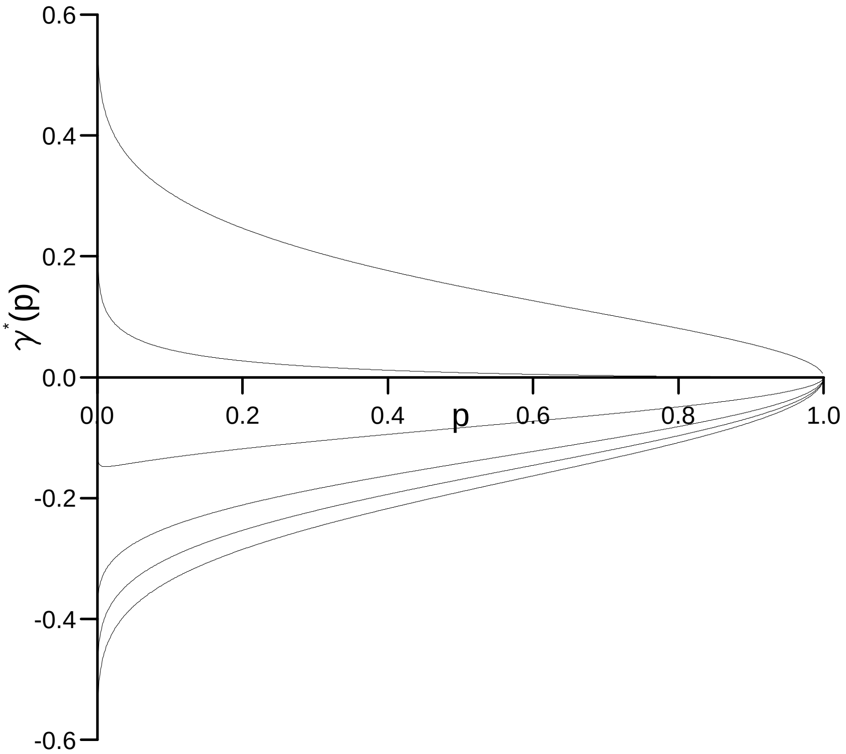

Figure 4. The Right Part Density f_R From Figure 1, Plotted on (m, b). The parallel slanting lines show the equality of f'_R at $x_{RP}(p)$ and $x_{RT}(p)$. The horizontal dashed line is now at an arbitrary height.

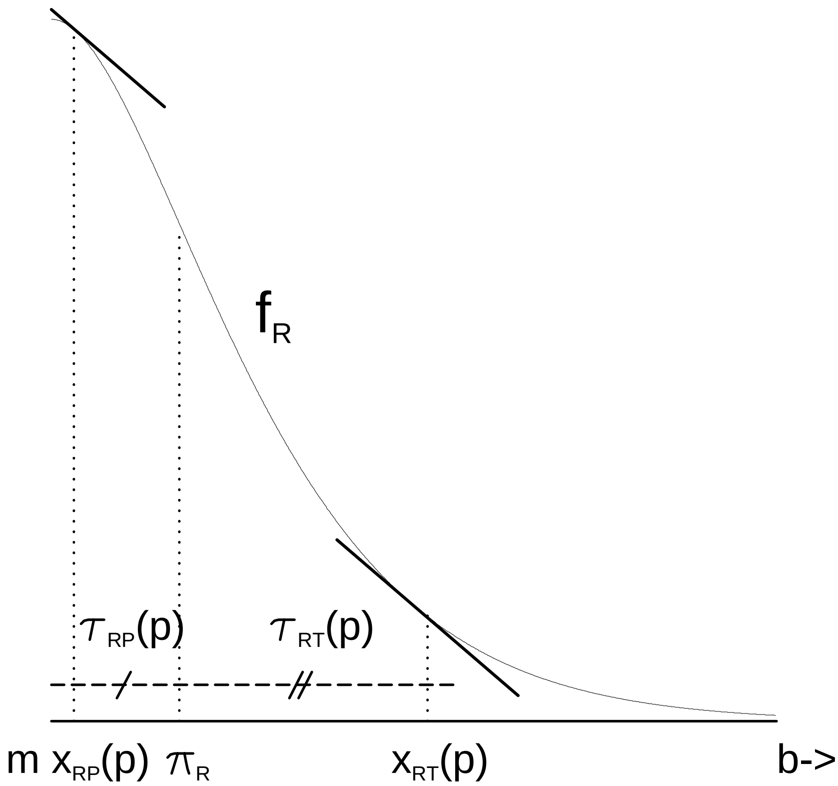

Figure 5. The Negative Derivative $-f_R'$ on (m, b) for the Density f in Figure 1, With Maximum at π_R. The horizontal dashed line is at height $p(-f_R'(\pi_R)) = p(-f_R'(\pi_R))$ and this defines the points $x_{RP}(p)$, $x_{RT}(p)$ and the distances $\tau_{RP}(p)$ and $\tau_{RT}(p)$ as shown.

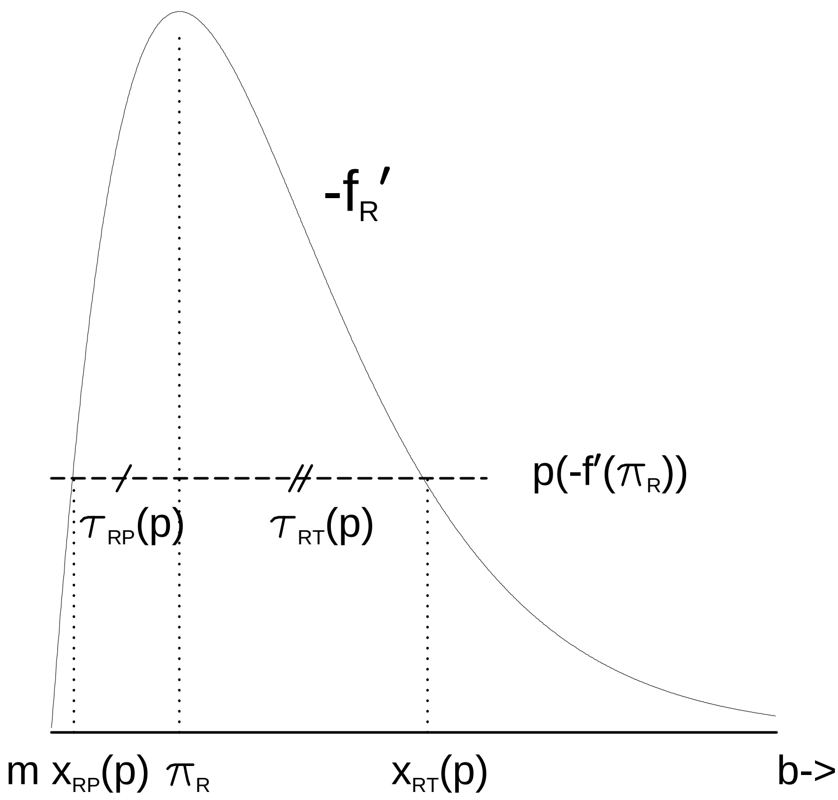

Figure 6. The Kurtosis Function $\delta^*(p)$, $0 < p < 1$, for: t densities with parameter, in order of decreasing value of kurtosis, $\nu = 0.5, 1, 2, 5, 10$ and 20 (thin solid lines); the normal density (thicker solid line); and symmetric beta densities with parameter, in order of decreasing value of kurtosis, $\eta = 10, 5, 4, 3, 2.5$ and 2.1 (dotted lines).

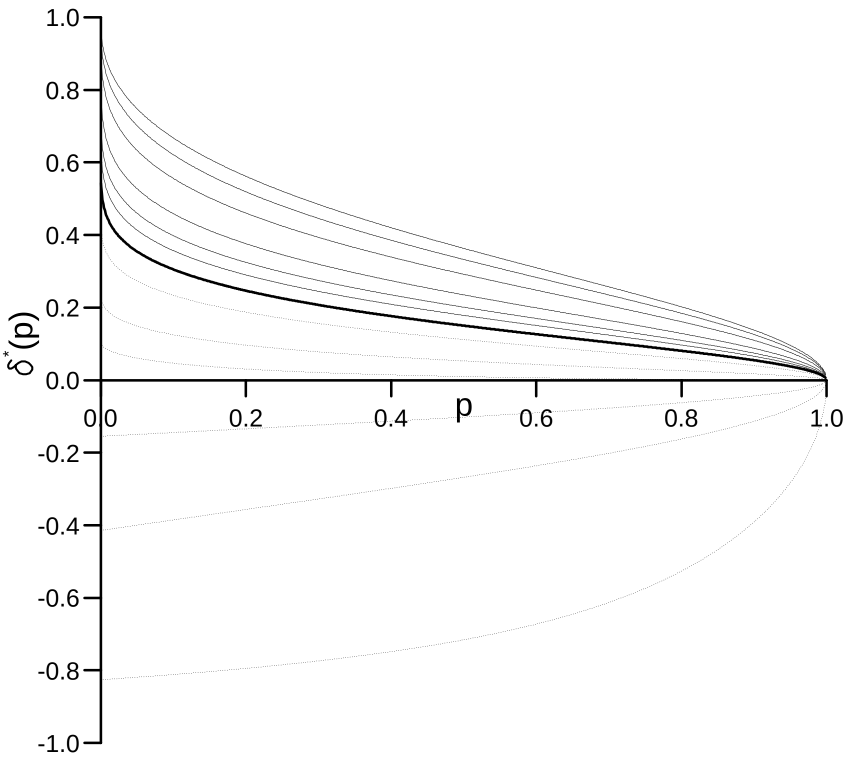

Figure 7. Left and Right Kurtosis Functions for Gamma Densities With Parameter $\alpha = 3$ (Solid Lines) and $\alpha = 100$ (Dotted Lines).

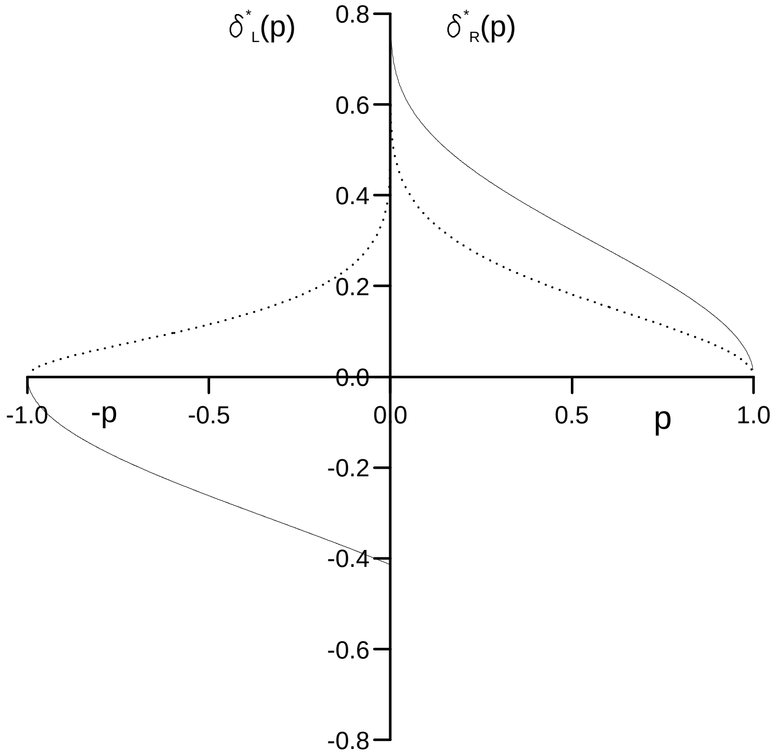

Figure 8. The d_2-Based Kurtosis Function $\delta_2^*(p)$, $0 < p < 1$, for Exponential Power Densities With Parameter, in Order of Decreasing Value of Kurtosis, $\beta = 1, 2, 3, 3.6, 5, 10$ *and* 100.

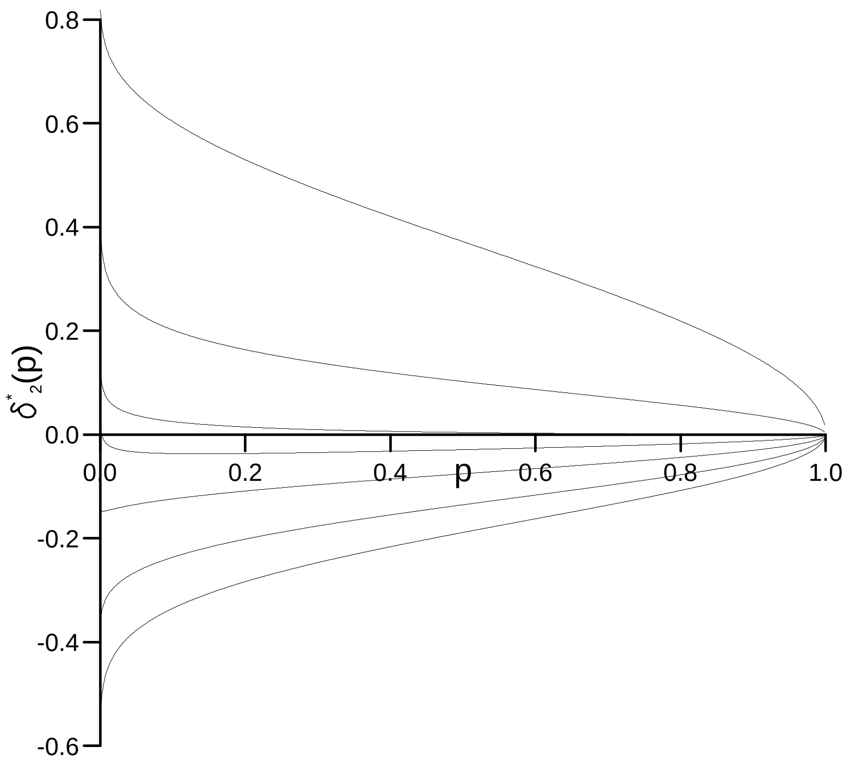

Figure 9. (a) Kernel Density Estimate and (b) Skewness Function Estimate for Glass Fibre Data; $n = 63$, $h = h_\gamma = 0.1499$.

FIGURE 9(a)

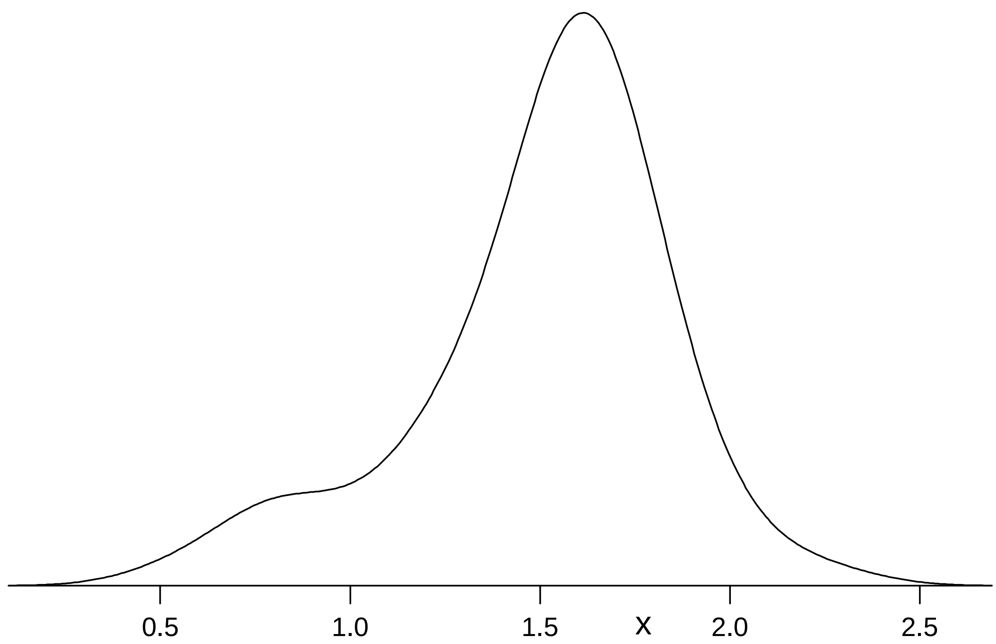

FIGURE 9(b)

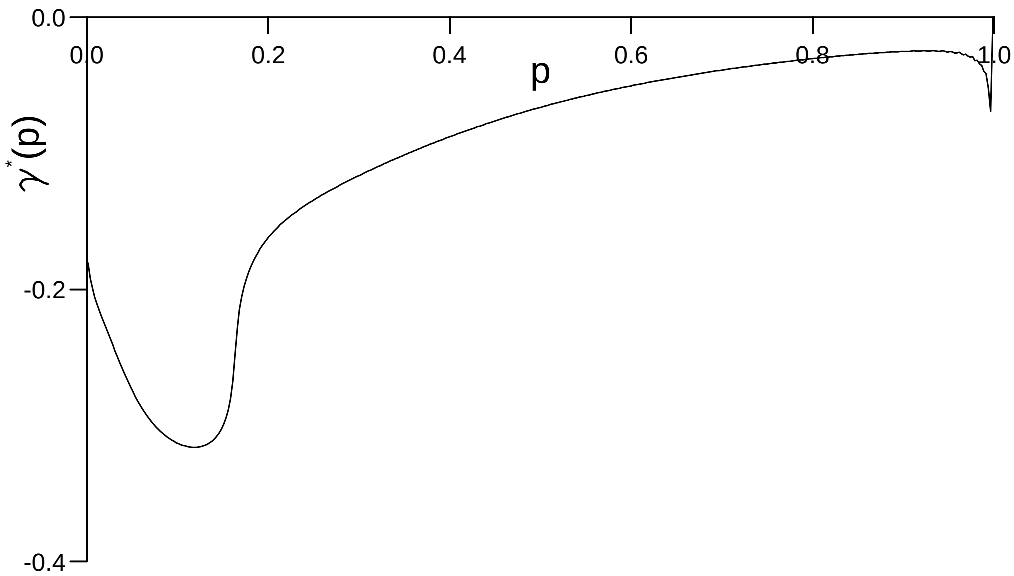

Figure 10. (a) Absolute Value of Kernel Density Derivative Estimate and (b) Left and Right Kurtosis Function Estimates for Glass Fibre Data; $n = 63$, $h = 1.35h_{1,\delta} = 0.2345$.

FIGURE 10(a)

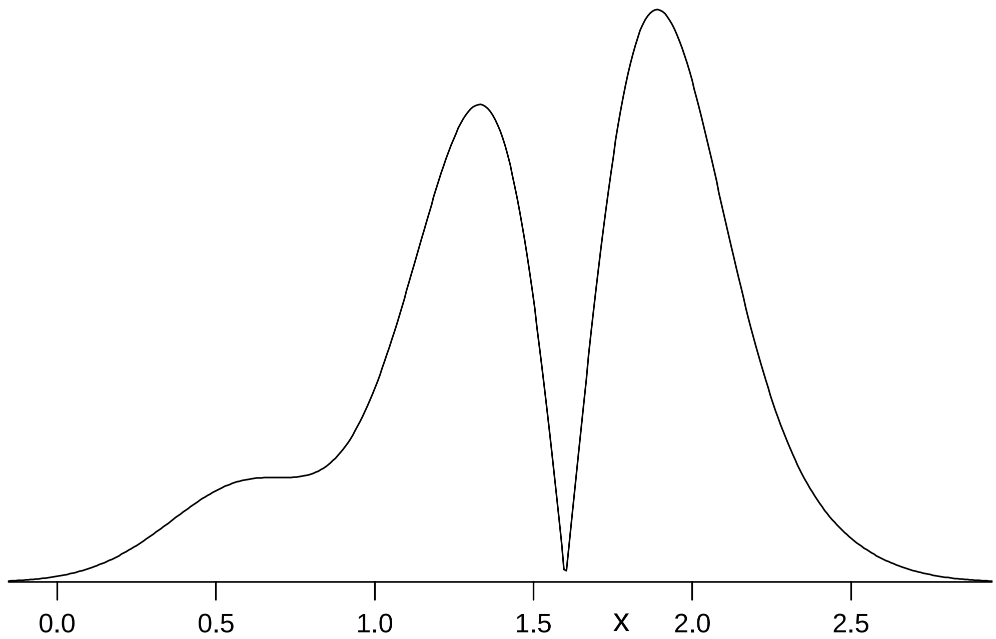

FIGURE 10(b)