Nonparametric Multivariate Kurtosis and Tailweight Measures

Jin Wang[1]

Northern Arizona University

and

Robert Serfling[2]

University of Texas at Dallas

November 2004 – final preprint version, to appear in *Journal of Nonparametric Statistics*, 2005

[1]Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, Arizona 86011-5717, USA. Email: jin.wang@nau.edu.

[2]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/~serfling. Support by NSF Grant DMS-0103698 is gratefully acknowledged.

Abstract

For nonparametric exploration or description of a distribution, the treatment of location, spread, symmetry and skewness is followed by characterization of *kurtosis*. Classical moment-based kurtosis measures the dispersion of a distribution about its "shoulders". Here we consider quantile-based kurtosis measures. These are robust, are defined more widely, and discriminate better among shapes. A univariate quantile-based kurtosis measure of Groeneveld and Meeden (1984) is extended to the multivariate case by representing it as a transform of a dispersion functional. A family of such kurtosis measures defined for a given distribution and taken together comprises a real-valued "kurtosis functional", which has intuitive appeal as a convenient two-dimensional curve for description of the kurtosis of the distribution. Several multivariate distributions in any dimension may thus be compared with respect to their kurtosis in a single two-dimensional plot. Important properties of the new multivariate kurtosis measures are established. For example, for elliptically symmetric distributions, this measure determines the distribution within affine equivalence. Related tailweight measures, influence curves, and asymptotic behavior of sample versions are also discussed.

AMS 2000 Subject Classification: Primary 62G05 Secondary 62H05.

Key words and phrases: Kurtosis; Tailweight; Depth functions; Influence curves.

1 Introduction

In developing a nonparametric description of a distribution, the natural step after treating location, spread, symmetry, and skewness is to characterize *kurtosis*. This feature, however, has proved more difficult to characterize and interpret, due to rather sophisticated linkage with spread, peakedness and tailweight, and with asymmetry if present. Here we treat kurtosis measures based on *quantile functions*, developing a general extension of existing univariate results to the *multivariate case*.

The classical notion of kurtosis is *moment-based*, given in the univariate case by the standardized fourth central moment $\kappa = E\{(X-\mu)^4\}/\sigma^4$. Although sometimes construed as simply a discriminator between heavy peakedness and heavy tails, it has become better understood as a measure concerning the structure of the distribution in the region that falls between, and links, the center and the tails. The "middle" of this region represents, in picturesque language, the "shoulders" of the distribution. More precisely, writing

$$\kappa = \mathrm{Var}\left\{\left(\frac{X-\mu}{\sigma}\right)^2\right\} + \left(E\left\{\left(\frac{X-\mu}{\sigma}\right)^2\right\}\right)^2 = \mathrm{Var}\left\{\left(\frac{X-\mu}{\sigma}\right)^2\right\} + 1, \tag{1}$$

it is seen that κ measures the dispersion of $(\frac{X-\mu}{\sigma})^2$ about its mean 1, or equivalently the dispersion of X about the points $\mu \pm \sigma$, which are viewed as the "shoulders". Thus classical univariate kurtosis measures in a location- and scale-free sense the *dispersion of probability mass away from the shoulders*, toward either the center or the tails or both. Rather than treat kurtosis simply as tailweight, it is more illuminating to treat *peakedness*, *kurtosis* and *tailweight* as distinct, although very much interrelated, descriptive features of a distribution. As probability mass diminishes in the region of the shoulders and increases in either the center or the tails or both, producing higher peakedness or heavier tailweight or both, the dispersion of X about the shoulders increases. As this dispersion increases indefinitely, heavy tailweight becomes a necessary component. See Finecan (1964), Chissom (1970), Darlington (1970), Hildebrand (1971), Horn (1983), Moors (1986), and Balanda and MacGillivray (1988) for discussion and illustration.

For the *multivariate* case, given a distribution in $\mathbb{R}^d$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the classical univariate kurtosis is generalized (Mardia, 1970) to

$$\kappa = E\{[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^2\},$$

i.e., the fourth moment of the Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}$. The multivariate extension of (1) shows that κ measures the dispersion of the squared Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}$ about its mean d, or equivalently the dispersion of $\boldsymbol{X}$ about the points on the ellipsoid $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d$, which surface thus comprises the "shoulders" of the distribution. Higher kurtosis thus arises when probability mass is diminished near the shoulders and greater either near $\boldsymbol{\mu}$ (greater peakedness) or in the tails (greater tailweight) or both.

Our purposes here concern *quantile-based* kurtosis measures, which in comparison with moment-based types are *robust*, are defined more widely, and discriminate better among distributional shapes. For symmetric univariate distributions, a quantile-based notion of kurtosis was formulated by Groeneveld and Meeden (1984), whose definition we broaden to allow asymmetry and then extend to the *multivariate* setting. Taken together over the range of quantile levels, these kurtosis measures comprise a real-valued *kurtosis functional*, which in turn may be represented as a transform of a *dispersion functional* based on the given quantile function. With an appropriately modified notion of "shoulders", each such quantile-based kurtosis measure in the family compares the relative sizes of two regions, taken just within and just without the shoulders, whose boundaries deviate from the shoulders by equal shifts of an outlyingness parameter. In this fashion the trade-off between peakedness and tailweight becomes characterized. Consequently, the moment- and quantile-based kurtosis measures tend to extract complementary pieces of information regarding the nature of the distribution in the region of "shoulders". The kurtosis functional has special intuitive appeal as a convenient two-dimensional curve for describing or comparing a number of multivariate distributions in any dimension with respect to their kurtosis. For elliptically symmetric distributions, and under suitable conditions on the chosen multivariate quantile function, the kurtosis functional possesses the very useful property of actually determining the form of the distribution up to affine transformation.

The above general formulation and key properties are developed in Section 2, along with pictorial illustration. Several complementary topics are treated briefly in Section 3: related tailweight measures, influence curves, asymptotic behavior of sample versions, and a new test of multivariate normality. Further related topics such as the ordering of distributions with respect to kurtosis, and the interrelations between kurtosis and skewness, are beyond the focus and scope of the present paper and left for subsequent investigation.

2 Quantile-Based Kurtosis Measures

As seen above, classical moment-based kurtosis quantifies the *dispersion* of probability mass in the region of the "shoulders" but does not characterize *shape*. Here we consider *quantile-based* kurtosis measures which indeed provide shape information, via a kurtosis *functional* that tends to serve as a discriminator between high peakedness and heavy tailweight, and which for elliptical distributions characterizes the shape within affine equivalence.

2.1 The univariate case

In describing the shape of an asymmetric distribution, kurtosis becomes entangled with skewness, a complication many authors avoid by restricting to symmetric distributions, or to symmetrized versions of asymmetric distributions. See MacGillivray and Balanda (1988) and Balanda and MacGillivray (1990) for details. In a review of quantile-based *skewness* measures, which have a long history, Groeneveld and Meeden (1984) show that reasonable *kurtosis* measures may be generated in the case of a *symmetric* distribution F by applying skewness measures to the distribution of the "folded" random variable $|X - M_F|$, where $M_F = F^{-1}(\frac{1}{2})$ denotes the median of F. In particular, application of a skewness functional of Oja (1981),

$$b_2(\alpha) = \frac{F^{-1}(\alpha) + F^{-1}(1-\alpha) - 2\,M_F}{F^{-1}(1-\alpha) - F^{-1}(\alpha)},\ 0 < \alpha < \tfrac{1}{2},$$

to the distribution $F_{|X-M_F|}(t) = 2F_X(M_F + t) - 1$, via the corresponding quantile function $F^{-1}_{|X-M_F|}(p) = F_X^{-1}(\frac{1}{2} + \frac{p}{2}) - M_F$, yields the *kurtosis functional*

$$k_F(p) = \frac{F^{-1}(\frac{3}{4} - \frac{p}{4}) + F^{-1}(\frac{3}{4} + \frac{p}{4}) - 2\,F^{-1}(\frac{3}{4})}{F^{-1}(\frac{3}{4} + \frac{p}{4}) - F^{-1}(\frac{3}{4} - \frac{p}{4})},\ 0 < p < 1. \tag{2}$$

See also Balanda and MacGillivray (1988, 1990), Groeneveld (1998) and Gilchrist (2000) for relevant discussion.

Interpretation of (2). As with the classical measure κ, there is an orientation toward "shoulders", which now, however, are given not by $\mu \pm \sigma$, but by the *1st and 3rd quartiles.* It suffices to consider simply the right-hand side of the symmetric distribution, for which the "shoulder" $F^{-1}(\frac{3}{4})$ partitions a "central part" from a complementary "tail part". The numerator of (2) expresses the difference in the lengths $\ell_1(p) = F^{-1}(\frac{3}{4} + \frac{p}{4}) - F^{-1}(\frac{3}{4})$ and $\ell_2(p) = F^{-1}(\frac{3}{4}) - F^{-1}(\frac{3}{4} - \frac{p}{4})$ of regions of equal probability $\frac{p}{4}$ taken just within and without the "shoulder", while the denominator is the sum of these two lengths. That is,

$$k_F(p) = \frac{\ell_1(p) - \ell_2(p)}{\ell_1(p) + \ell_2(p)},\ 0 < p < 1. \tag{3}$$

Clearly, $|k_F(p)| \le 1$, with values near $+1$ suggesting a pronounced shift of probability mass away from the tails and toward the center, values near -1 suggesting a U-shaped distribution, and values near 0 suggesting a rather uniform distribution. ∎

Below we extend this functional to the multivariate case. As a prelude, we now carry out two steps which both clarify the univariate case and set the stage for appropriate generalization. The first step extends to the univariate *asymmetric* case, by combining right- and left-hand versions of (2), leading to

$$\frac{[F^{-1}(\frac{3}{4} + \frac{p}{4}) + F^{-1}(\frac{3}{4} - \frac{p}{4}) - 2\,F^{-1}(\frac{3}{4})] - [F^{-1}(\frac{1}{4} + \frac{p}{4}) + F^{-1}(\frac{1}{4} - \frac{p}{4}) - 2\,F^{-1}(\frac{1}{4})]}{[F^{-1}(\frac{3}{4} + \frac{p}{4}) - F^{-1}(\frac{3}{4} - \frac{p}{4})] + [F^{-1}(\frac{1}{4} + \frac{p}{4}) - F^{-1}(\frac{1}{4} - \frac{p}{4})]},\ 0 < p < 1. \tag{4}$$

Interpretation of (4). The shoulders $F^{-1}(\frac{1}{4})$ and $F^{-1}(\frac{3}{4})$ bound a "central region" of probability $\frac{1}{2}$ which is complemented by a two-sided "tail region". With $\ell_1(p)$ now the total length of the (two-sided) region of probability $p/2$ just *without* the central region and $\ell_2(p)$ now the total length of the (two-sided) region of probability $p/2$ just *within* the central region, the quantity $k_F(p)$ given by (4) may also be expressed in the form (3). ■

Next we express $k_F(p)$ in terms of a well-known *dispersion* functional (e.g., Balanda and MacGillivray, 1990),

$$d_F(p) = F^{-1}(\tfrac{1+p}{2}) - F^{-1}(\tfrac{1-p}{2}),\ 0 \le p < 1, \tag{5}$$

which for each p gives the width of the interquantile region of probability p with tails of equal probability $(1-p)/2$. By rearrangement of terms, (4) becomes

$$\frac{[F^{-1}(\frac{3}{4}+\frac{p}{4}) - F^{-1}(\frac{1}{4}-\frac{p}{4})] + [F^{-1}(\frac{3}{4}-\frac{p}{4}) - F^{-1}(\frac{1}{4}+\frac{p}{4})] - 2\,[F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4})]}{[F^{-1}(\frac{3}{4}+\frac{p}{4}) - F^{-1}(\frac{1}{4}-\frac{p}{4})] - [F^{-1}(\frac{3}{4}-\frac{p}{4}) - F^{-1}(\frac{1}{4}+\frac{p}{4})]},\ 0 < p < 1,$$

which corresponds to

$$k_F(p) = \frac{d_F(\frac{1}{2}-\frac{p}{2}) + d_F(\frac{1}{2}+\frac{p}{2}) - 2\,d_F(\frac{1}{2})}{d_F(\frac{1}{2}+\frac{p}{2}) - d_F(\frac{1}{2}-\frac{p}{2})},\ 0 < p < 1. \tag{6}$$

Thus the kurtosis functional (4) may be represented succinctly in terms of dispersion rather than in terms of quantiles (as already given in the symmetric case by Groeneveld, 1998). The representation (6) in terms of dispersion provides a further way to understand and interpret $k_F(p)$, just as we saw the moment-based κ to be usefully represented in terms of variance. Further, this form for $k_F(p)$ is the appropriate one for generalization to the multivariate case, as seen in the next subsection. Figure 1 shows (6) for the uniform distribution and several common symmetric unimodal distributions.

Why look at plots of $k_F(p)$ instead of the underlying quantile function $F^{-1}(p)$ which already contains all the information about F? when the kurtosis functional $k_F(p)$ is a function of the quantile function and so inherently contains no additional information? Plots directly of the quantile function itself, even when standardized for location and scale, are difficult to interpret, just as are plots of cumulative distribution functions. For the latter, plots of densities are helpful. But even then it is very informative to exhibit specific key features of a model by employing suitable *descriptive measures*. For such purposes, kurtosis provides an important additional tool along with location, spread, and skewness. Of course, one might use location centered quantile-quantile plots to compare a pair of distributions in an overall way, but interpretation in terms of descriptive features is difficult to extract from such a plot. Further, only two models can be compared at a time in such a plot. On the other hand, a location, spread, skewness or kurtosis functional permits several distributions to be compared with respect to the selected descriptive feature and within a single two-dimensional plot.

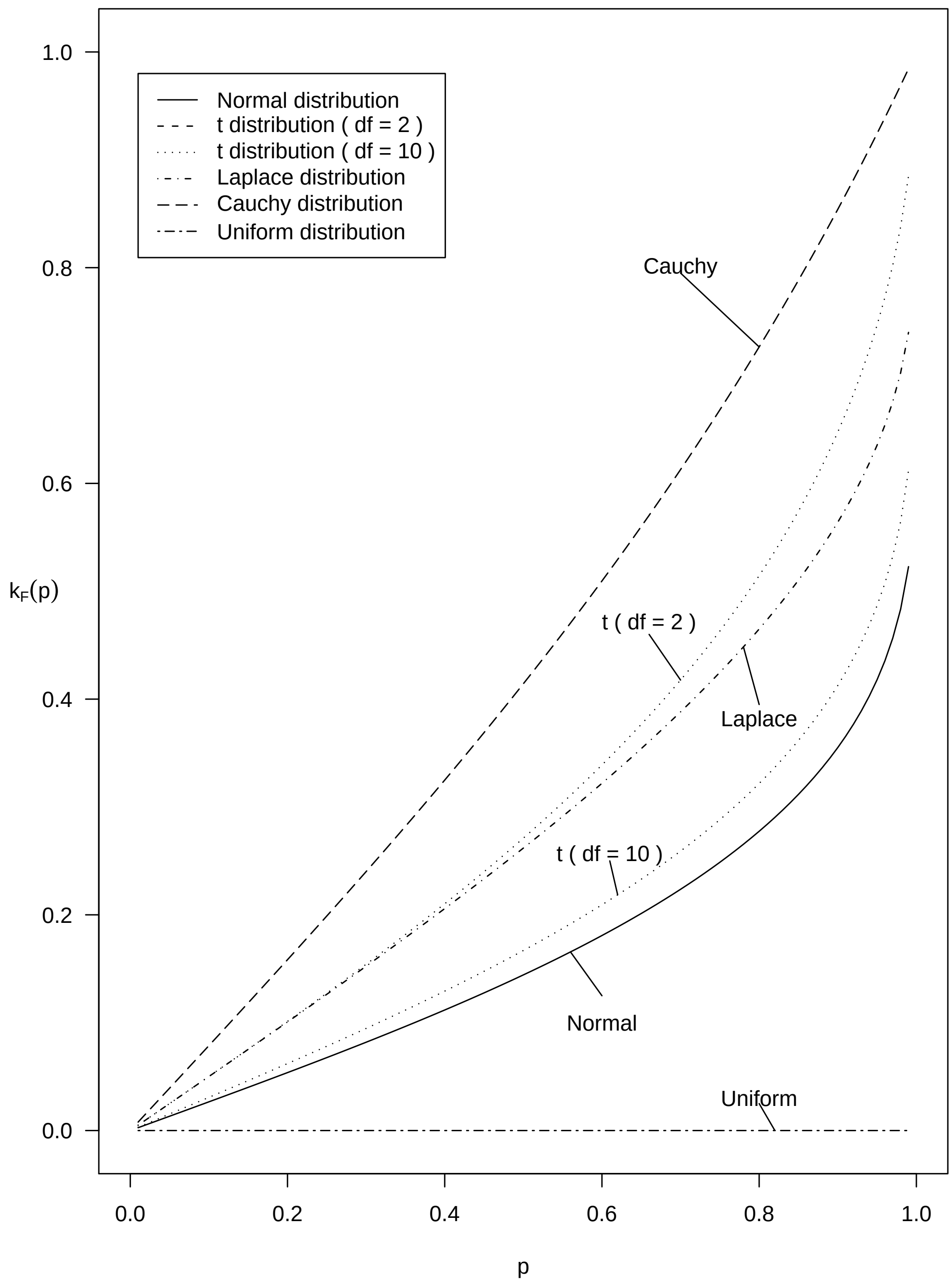

Figure 1: Kurtosis curves of some symmetric univariate distributions.

2.2 The multivariate case

Analogues of (5) and (6) are developed as follows, utilizing notions of multivariate quantile functions and associated central regions. First we select as the "center" $\boldsymbol{M}_F$ any notion of multidimensional median. Various possibilities are discussed in Small (1990), Liu, Parelius and Singh (1999), and Zuo and Serfling (2000a). Next we consider any family of "central regions" $\mathcal{C} = \{C_F(r) : 0 \le r < 1\}$ which are nested about $\boldsymbol{M}_F$ and reduce to $\boldsymbol{M}_F$ as $r \to 0$. For $\boldsymbol{x} \in \mathbb{R}^d$, let r index the corresponding central region with $\boldsymbol{x}$ on its boundary, let $\boldsymbol{v}$ be the unit vector toward $\boldsymbol{x}$ from $\boldsymbol{M}_F$, and put $\boldsymbol{u} = r\boldsymbol{v}$. Setting $Q_F(\boldsymbol{u}) = \boldsymbol{x}$, with $Q_F(\boldsymbol{0}) = \boldsymbol{M}_F$, the points $\boldsymbol{x} \in \mathbb{R}^d$ thus generate a *quantile function* $Q_F(\boldsymbol{u})$ for $\boldsymbol{u}$ taking values in the unit ball $\mathbb{B}^d$ in $\mathbb{R}^d$. Here $\|\boldsymbol{u}\|$ denotes the *outlyingness* of the point $\boldsymbol{x} = Q_F(\boldsymbol{u})$, and the index r may be interpreted as an *outlyingness parameter* describing the extent of the region $C_F(r)$. As in Liu, Parelius and Singh (1999), quantile functions and central regions might arise in connection with the contours of equal depth of a statistical depth function, with $C_F(r)$ the central region having probability weight r. An alternative approach is given by the spatial quantile of Chaudhuri (1996), also treated in Serfling (2003), where $\boldsymbol{u}$ is not the direction between $Q_F(\boldsymbol{u})$ and $\boldsymbol{M}_F$, but rather the expected direction from $Q_F(\boldsymbol{u})$ to $\boldsymbol{X}$ having cdf F. Other appropriate multivariate quantile functions may be considered as well.

A judicious choice of family of central regions is one whose contours follow the distribution F in the sense of agreeing with outlyingness contours with respect to some meaningful notion of outlyingness. A natural approach is provided by depth functions, as seen below.

For a chosen family $\mathcal{C} = \{C_F(r) : 0 \le r < 1\}$, a corresponding *volume functional* is defined as

$$V_{F,\mathcal{C}}(r) = \mathrm{volume}(C_F(r)),\ 0 \le r < 1.$$

An increasing function of the variable r, $V_{F,\mathcal{C}}(r)$ characterizes the dispersion of F in terms of the expansion of the central regions $C_F(r)$. The volume functional plays in higher dimensions the role of the univariate dispersion functional (5) based on *interquantile central regions with equiprobable tails.*

As in the univariate case with the dispersion functional, the volume functional yields other descriptive shape information besides measuring scale: namely, a *kurtosis functional*, as a natural analogue of (6):

$$k_{F,\mathcal{C}}(r) = \frac{V_{F,\mathcal{C}}(r)(\frac{1}{2}-\frac{r}{2}) + V_{F,\mathcal{C}}(r)(\frac{1}{2}+\frac{r}{2}) - 2\,V_{F,\mathcal{C}}(r)(\frac{1}{2})}{V_{F,\mathcal{C}}(r)(\frac{1}{2}+\frac{r}{2}) - V_{F,\mathcal{C}}(r)(\frac{1}{2}-\frac{r}{2})},\ 0 < r < 1. \tag{7}$$

For convenience, in the sequel we will simply write $k_F(\cdot)$, leaving $\mathcal{C}$ implicit.

The nature of $k_F(r)$ is easily understood via Figure 2, which exhibits regions A and B for which $k_F(r)$ is represented as the difference of their volumes divided by the sum of their volumes:

$$\frac{\mathrm{volume}(B) - \mathrm{volume}(A)}{\mathrm{volume}(B) + \mathrm{volume}(A)}.$$

The boundary of the central region $C_F(\frac{1}{2})$ represents the "shoulders" of the distribution and separates a designated "central part" from a corresponding "tail part". The quantity $k_F(r)$ thus measures the relative volumetric difference between a region A just within the shoulders and a region B just without, in the central and tail parts, respectively, which are defined by modifying equally the "outlyingness" parameter $r = 1/2$ of the shoulders by adding and subtracting a given amount $r/2$.

Within certain typical classes of distribution we can attach intuitive interpretations to ranges of values for $k_F(\cdot)$. For example, if we confine attention to a class of distributions for which either F is unimodal, or F is uniform, or $1-F$ is unimodal, then, for any fixed r, a value of $k_F(r)$ near $+1$ suggests *peakedness*, a value near -1 suggests a *bowl-shaped* distribution, and a value near 0 suggests *uniformity*. Thus higher $k_F(r)$ arises when probability mass is greater in the "central part", or greater in the "tail part", or both, which is consistent with the interpretation of the moment-based kurtosis measure.

It is clear from Figure 2 that this kurtosis functional can be defined without regard to whether F is symmetric or not. Here, for generality, the central regions are "generically" hand-drawn, and these particular contours only coincidentally reflect approximate spherical symmetry. Of course, any actual asymmetry becomes reflected in the contours, and thus asymmetry and kurtosis are somewhat confounded, as is well-known. Further, this figure provides a clarification for the univariate case: (4) and (6) not only extend (2) to include asymmetric distributions but actually represent in their own right the natural and most intuitive way to define a quantile-based notion of kurtosis.

Example *Kurtosis curves based on halfspace depth.* In correspondence with the well-known *halfspace depth function*,

$$D(\boldsymbol{x}, F) = \inf\{F(H) : \boldsymbol{x} \in H \in \boldsymbol{\mathcal{H}}\},$$

where $\boldsymbol{\mathcal{H}}$ is the class of closed halfspaces of $\mathbb{R}^d$, we may take M_F as the point of maximal halfspace depth and the central region $C_{F,D}(p)$ as the set of form $\{\boldsymbol{x} : D(\boldsymbol{x}, F) \geq \alpha\}$ having probability weight p, for $0 \leq p < 1$. Thus here the probability weight p is interpreted as an "outlyingness parameter". Figure 3 exhibits the corresponding kurtosis curves k_F with central regions based on halfspace depth, for F given by d-variate normal distributions with $d = 1, 5, 10, 15, 20$. Likewise, Figures 4 and 5 show k_F for F given by various choices of 8-variate symmetric Pearson Type II and Kotz type distributions, respectively. We note that these F are elliptically symmetric (see Fang, Kotz and Ng, 1990) and, consequently, other depth functions which like the halfspace depth are affine invariant yield the same (ellipsoidal) central regions and hence the same kurtosis curves. In fact, for such F, the contours of equal affine invariant depth agree with the contours of equal probability density. ■

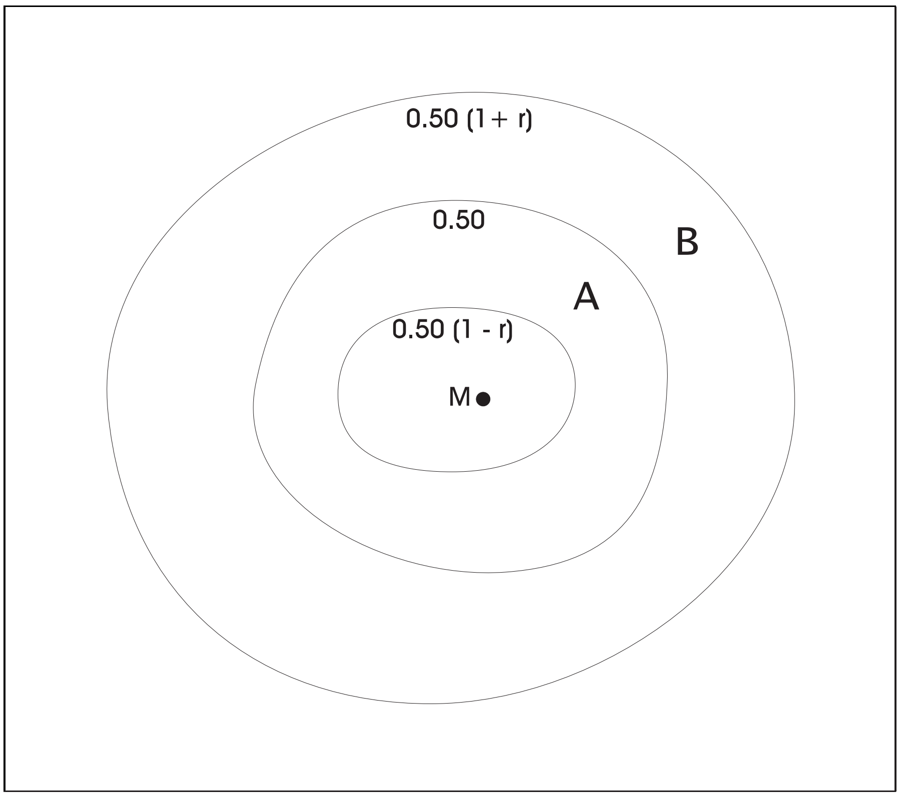

Figure 2: Median $M = M_F$ and central regions $C_F(\frac{1}{2} - \frac{r}{2})$, $C_F(\frac{1}{2})$, and $C_F(\frac{1}{2} + \frac{r}{2})$, with $A = C_F(\frac{1}{2}) - C_F(\frac{1}{2} - \frac{r}{2})$ and $B = C_F(\frac{1}{2} + \frac{r}{2}) - C_F(\frac{1}{2})$.

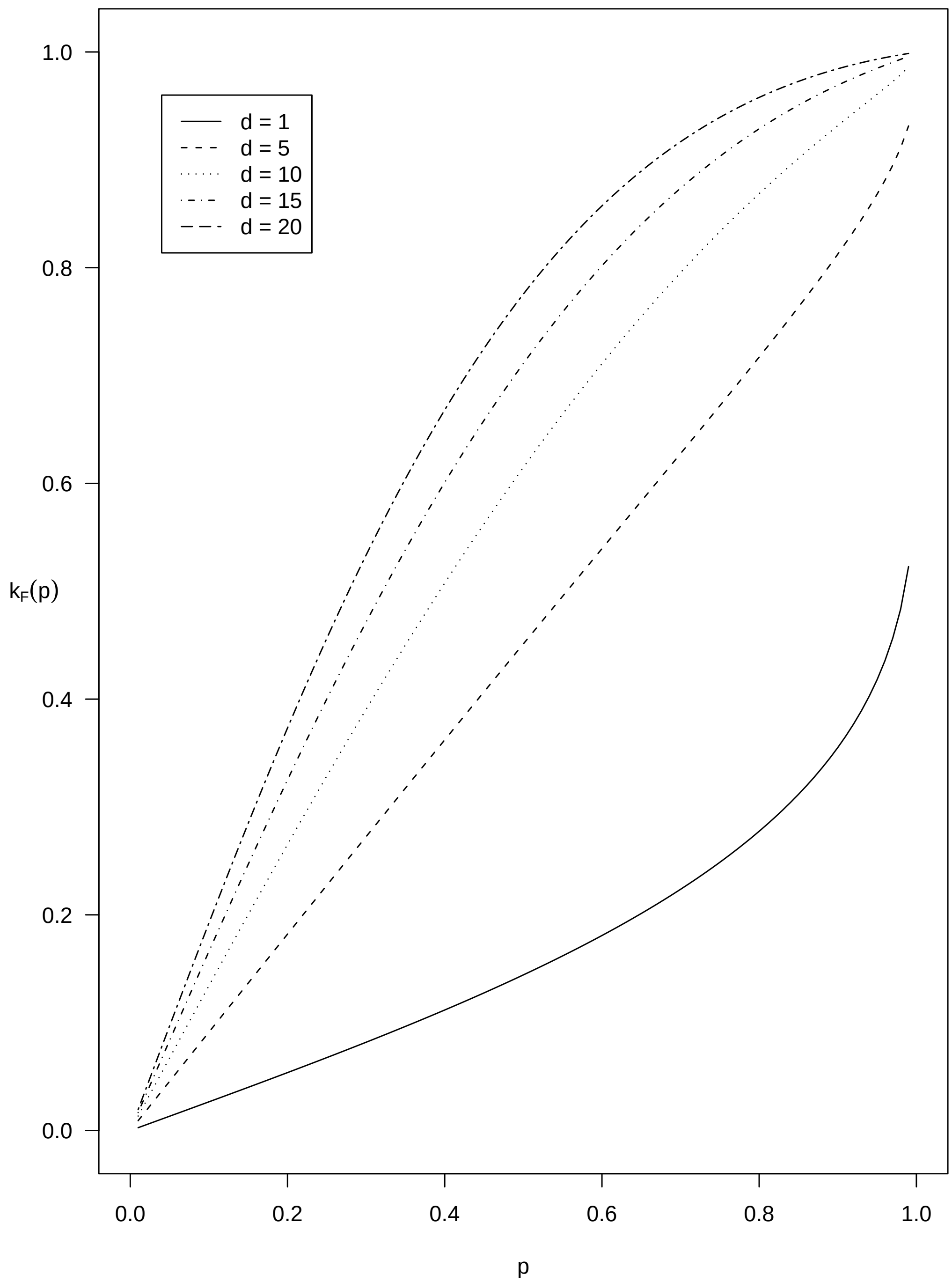

Figure 3: Kurtosis curves for normal distributions for $d = 1, 5, 10, 15, 20$, based on halfspace or other affine invariant depth, or equivalently equidensity contours.

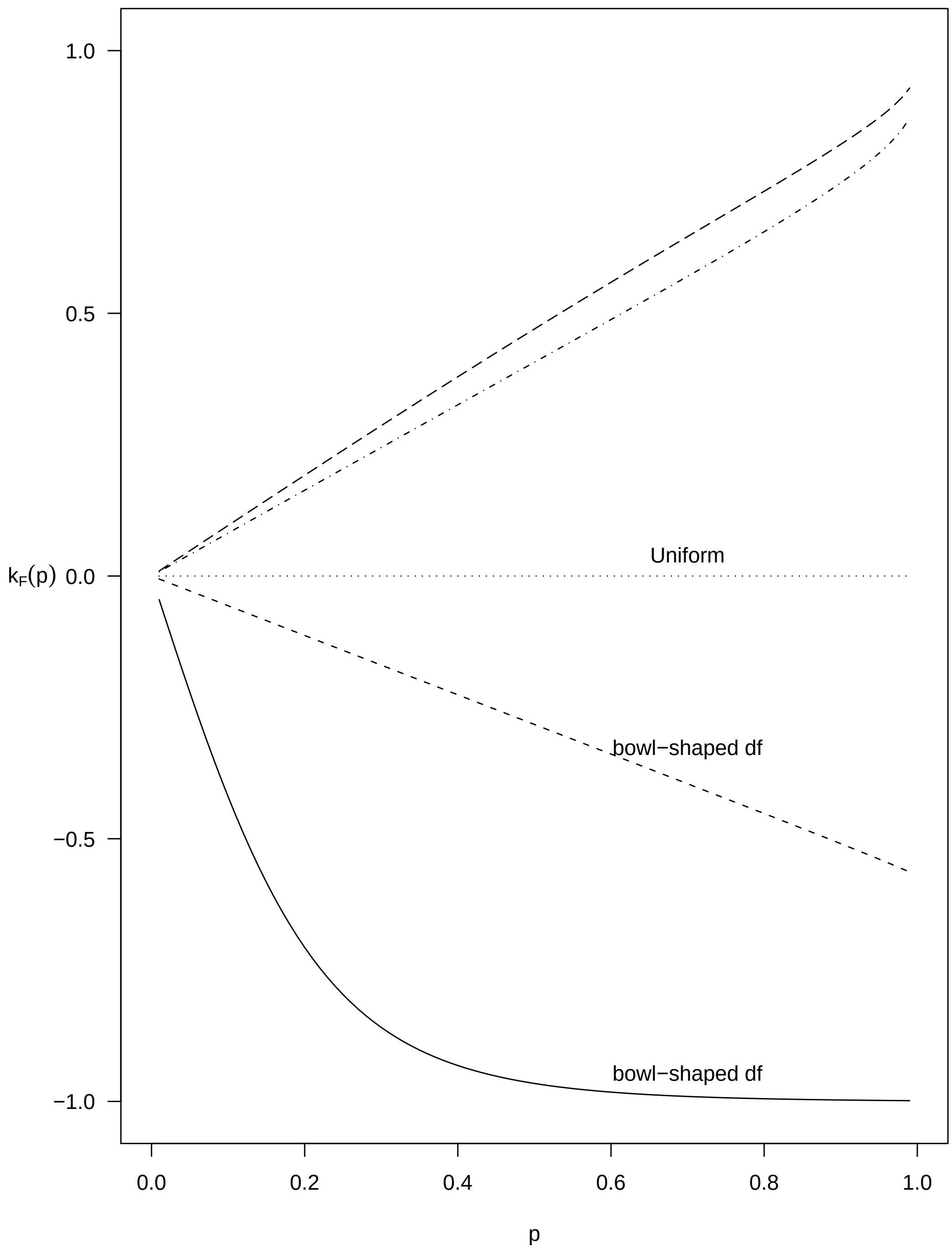

Figure 4: Kurtosis curves for symmetric Pearson Type II distributions, $d = 8$, based on halfspace or other affine invariant depth, or equivalently equidensity contours.

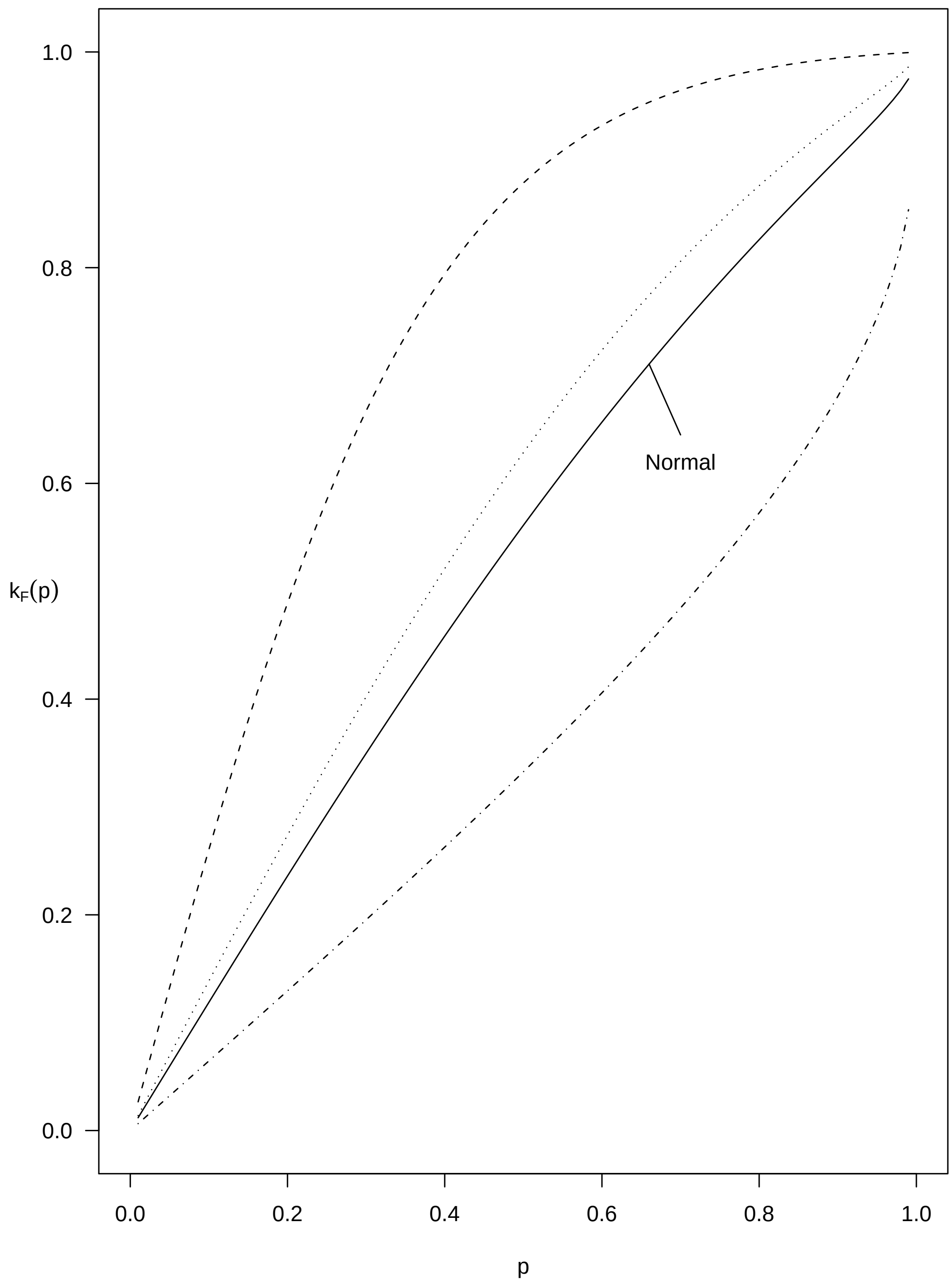

Figure 5: Kurtosis curves for symmetric Kotz type distributions, $d = 8$, based on halfspace or other affine invariant depth, or equivalently equidensity contours.

2.3 Properties of the kurtosis functional

The following basic properties for $k_F(r)$ are straightforward (see Wang, 2003, for details).

1. $-1 \leq k_F(r) \leq 1$.
2. $k_F(r)$ is defined without moment assumptions.
3. For uniform distributions, $k_F(r) \equiv 0$.
4. For unimodal distributions, $k_F(r) > 0$.
5. For bowl-shaped distributions (i.e., $1 - F$ unimodal), $k_F(r) < 0$.
6. If $V_{F,\mathcal{C}}(r)(r)$ is a continuous function of r, then

$$\lim_{r \to 1} k_F(r) = \begin{cases} \dfrac{\text{volume(support}(F)) \; - \; 2V_{F,\mathcal{C}}(r)(\frac{1}{2})}{\text{volume(support}(F))}, & \text{volume(support}(F)) \; < \infty, \\ 1, & \text{otherwise.} \end{cases}$$

7. If $V_{F,\mathcal{C}}(r)(r)$ is differentiable with nonzero derivative at $r = 0.5$, then

$$\lim_{r \to 0} k_F(r) = 0.$$

Two very important further properties are established in the following results. The first states that if the underlying quantile function $Q_F(\cdot)$ is affine equivariant, then the corresponding kurtosis functional is affine invariant. For this we discuss

Equivariance of multivariate quantile functions. The condition for Q_F to be equivariant with respect to the transformation $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ is that

$$Q_G\left(\frac{\boldsymbol{A}\boldsymbol{u}}{\|\boldsymbol{A}\boldsymbol{u}\|}\|\boldsymbol{u}\|\right) = \boldsymbol{A}Q_F(\boldsymbol{u}) + \boldsymbol{b}, \;\; \boldsymbol{u} \in \mathbb{B}^d, \tag{8}$$

where G denotes the cdf of $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ when $\boldsymbol{X}$ has cdf F. That is, the point $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ has a quantile representation given by that of $\boldsymbol{x}$ similarly transformed, subject to a re-indexing that keeps the indices in the unit ball and keeps the outlyingness of $\boldsymbol{x}$ unchanged under transformation. In particular, setting $\boldsymbol{u} = \boldsymbol{0}$, the medians satisfy $\boldsymbol{M}_G = \boldsymbol{A}\boldsymbol{M}_F + \boldsymbol{b}$. Further, it is easily seen that the corresponding rth central regions satisfy

$$C_G(r) = \boldsymbol{A}C_F(r) + \boldsymbol{b}. \tag{9}$$

If (8) holds for all nonsingular $d \times d$ $\boldsymbol{A}$ and all $\boldsymbol{b} \in \mathbb{R}^d$, then Q_F is an *affine equivariant* functional of F. ■

Theorem 2.1 *If $k_F(\cdot)$ is based on an affine equivariant quantile function $Q_F(\cdot)$, then, for each r, $k_F(r)$ is an affine invariant functional of F: for all nonsingular $d \times d$ $\boldsymbol{A}$ and all $\boldsymbol{b} \in \mathbb{R}^d$,*

$$k_G(r) = k_F(r), \tag{10}$$

where G denotes the cdf of $\boldsymbol{AX} + \boldsymbol{b}$ when $\boldsymbol{X}$ has cdf F.

Proof. Using (9), we have

$$\begin{aligned} V_{G,\mathcal{C}}(r) &= \int \mathbf{1}\{\boldsymbol{y} \in C_G(r)\}\, d\boldsymbol{y} \\ &= \int \mathbf{1}\{\boldsymbol{y} \in \boldsymbol{A}Q_F(r) + \boldsymbol{b}\}\, d\boldsymbol{y} \\ &= \int \mathbf{1}\{\boldsymbol{z} \in C_F(r)\}\, |\det(\boldsymbol{A})|\, d\boldsymbol{z} \\ &= |\det(\boldsymbol{A})|\, V_{F,\mathcal{C}}(r)(r). \end{aligned}$$

From this with the definition of $k_F(\cdot)$, we readily obtain (10). ∎

Our next result establishes a partial converse to Theorem 2.1. The distribution F is called *elliptically symmetric* if it has a density of form

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} h((\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})),\ \boldsymbol{x} \in \mathbb{R}^d,$$

for a nonnegative function $h(\cdot)$ with $\int_0^\infty t^{d/2-1} h(t)dt < \infty$ and a positive definite matrix $\boldsymbol{\Sigma}$.

Theorem 2.2 *Let $Q_F(\cdot)$ be affine equivariant. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ have elliptically symmetric distributions F and G, respectively, and suppose that*

$$k_G(r) = k_F(r),\ 0 \le r < 1.$$

Then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are affinely equivalent in distribution: $\boldsymbol{Y} \overset{d}{=} \boldsymbol{AX} + \boldsymbol{b}$ for some nonsingular matrix $\boldsymbol{A}$ and some vector $\boldsymbol{b}$.

We will need the following lemma.

Lemma 2.1 *Let F be elliptically symmetric and $Q_F(\cdot)$ affine equivariant. Then, for some nondecreasing function $\gamma(\cdot)$,*

$$(Q_F(\boldsymbol{u}) - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(Q_F(\boldsymbol{u}) - \boldsymbol{\mu}) = \gamma(\|\boldsymbol{u}\|) \tag{11}$$

and hence the corresponding central regions are ellipsoidal,

$$C_F(r) = \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \le \gamma(r)\}.$$

Proof of Lemma 2.1. For $\boldsymbol{X}$ elliptically symmetric, $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$ is spherically symmetric and thus, for any orthogonal $d \times d$ matrix $\boldsymbol{T}$,

$$\boldsymbol{T}\boldsymbol{Z} \overset{d}{=} \boldsymbol{Z}.$$

Fix a unit vector $\boldsymbol{u}_0$, for example $\boldsymbol{u}_0 = (1, 0, \ldots, 0)'$. Then, for any $\boldsymbol{u} \in \mathbb{B}^d(\boldsymbol{0})$, there exists an orthogonal matrix $\boldsymbol{U}$ such that $\boldsymbol{u} = \boldsymbol{U}\boldsymbol{u}_0\|\boldsymbol{u}\|$. Let G denote the cdf of $\boldsymbol{Z}$ and H that of $\boldsymbol{U}\boldsymbol{Z}$. Then affine equivariance of Q_F yields

$$Q_G(\boldsymbol{u}) = Q_G(\boldsymbol{U}\boldsymbol{u}_0\|\boldsymbol{u}\|) = Q_H(\boldsymbol{U}\boldsymbol{u}_0\|\boldsymbol{u}\|) = \boldsymbol{U}Q_G(\boldsymbol{u}_0\|\boldsymbol{u}\|).$$

Defining

$$\gamma(r) = Q_G(\boldsymbol{u}_0\, r)'Q_G(\boldsymbol{u}_0\, r),\ \ 0 \le r \le 1,$$

which we note is nondecreasing due to nestedness of central regions about the median, we thus have

$$Q_G(\boldsymbol{u})'Q_G(\boldsymbol{u}) = Q_G(\boldsymbol{u}_0\|\boldsymbol{u}\|)'Q_G(\boldsymbol{u}_0\|\boldsymbol{u}\|) = \gamma(\|\boldsymbol{u}\|).$$

On the other hand,

$$\boldsymbol{\Sigma}^{-1/2}(Q_F(\boldsymbol{u}) - \boldsymbol{\mu}) = Q_G\left(\frac{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}}{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}\|}\|\boldsymbol{u}\|\right),$$

and it follows that

$$\begin{aligned}(Q_F(\boldsymbol{u}) - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(Q_F(\boldsymbol{u}) - \boldsymbol{\mu}) &= Q_G\left(\frac{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}}{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}\|}\|\boldsymbol{u}\|\right)' Q_G\left(\frac{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}}{\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}\|}\|\boldsymbol{u}\|\right) \\ &= \gamma(\|\boldsymbol{u}\|),\end{aligned}$$

completing the proof. ■

Proof of Theorem 2.2. For F elliptically symmetric as assumed, denote by F_R and f_R the cdf and density of the squared Mahalanobis distance

$$R = (\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}).$$

Then we have

$$f_R(r) = \frac{\pi^{d/2} r^{d/2-1} h(r)}{\Gamma(d/2)}$$

and

$$P(\boldsymbol{X} \in C_F(r)) = P(R \le \gamma(r)) = F_R(\gamma(r)).$$

It readily follows that

$$V_{F,\mathcal{C}}(r)(r) = \text{volume}(C_F(r)) = \frac{\pi^{d/2}[F_R^{-1}(r)]^{d/2}|\mathbf{\Sigma}|^{1/2}}{\Gamma(d/2+1)},$$

from which we obtain

$$k_F(r) = \frac{[F_R^{-1}(\frac{1}{2}+\frac{r}{2})]^{d/2} + [F_R^{-1}(\frac{1}{2}-\frac{r}{2})]^{d/2} - 2[F_R^{-1}(\frac{1}{2})]^{d/2}}{[F_R^{-1}(\frac{1}{2}+\frac{r}{2})]^{d/2} - [F_R^{-1}(\frac{1}{2}-\frac{r}{2})]^{d/2}}. \tag{12}$$

Likewise, for $\boldsymbol{Y}$ elliptically symmetric with parameters $\boldsymbol{\mu}_1$, $\mathbf{\Sigma}_1$ and $h_1(\cdot)$, we obtain an expression for $k_G(r)$ similar to (12) with F_R replaced by F_{R^*}, the cdf of

$$R^* = (\boldsymbol{Y} - \boldsymbol{\mu}_1)'\mathbf{\Sigma}_1^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_1).$$

Equating the two expressions for $k_F(r)$ and $k_G(r)$, we obtain after some reduction

$$\frac{[F_R^{-1}(\frac{1}{2}+\frac{r}{2})]^{d/2}}{[F_{R^*}^{-1}(\frac{1}{2}+\frac{r}{2})]^{d/2}} = \frac{[F_R^{-1}(\frac{1}{2}-\frac{r}{2})]^{d/2}}{[F_{R^*}^{-1}(\frac{1}{2}-\frac{r}{2})]^{d/2}} = \frac{[F_R^{-1}(\frac{1}{2})]^{d/2}}{[F_{R^*}^{-1}(\frac{1}{2})]^{d/2}} \ := q^{d/2}, \text{ say.}$$

This yields

$$F_R^{-1}(r) = q\, F_{R^*}^{-1}(r),\ 0 \le r \le 1.$$

Equivalently,

$$F_{R^*}(x) = F_{R/q}(x),\ 0 \le x < \infty,$$

which leads to

$$q^{1/2}\mathbf{\Sigma}_1^{-1/2}(\boldsymbol{Y} - \boldsymbol{\mu}_1) \stackrel{d}{=} \mathbf{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}),$$

i.e.,

$$\boldsymbol{Y} \stackrel{d}{=} q^{-1/2}\mathbf{\Sigma}_1^{1/2}\mathbf{\Sigma}^{-1/2}\boldsymbol{X} + (\boldsymbol{\mu}_1 - q^{-1/2}\mathbf{\Sigma}_1^{1/2}\mathbf{\Sigma}^{-1/2}\boldsymbol{\mu}).$$

This completes the proof. ■

In the univariate case, Theorem 2.2 reduces to the following simple result.

Corollary 2.1 *Let X and Y have univariate symmetric distributions F and G, respectively, and suppose that*

$$k_G(r) = k_F(r),\ 0 \le r < 1,$$

with $k_F(\cdot)$ defined by (2). Then $Y \stackrel{d}{=} aX + b$ for some $a \neq 0$ and b.

Theorem 2.2 establishes, for elliptically symmetric distributions, that the kurtosis functional determines the distribution up to affine equivalence. An important potential practical application is discussed below.

3 Complements

In practice one must rely on sample versions of the kurtosis functional $k_F(\cdot)$ and apply related asymptotics. For example, one desires results such as uniform strong convergence of sample versions to their population versions, and uniform weak convergence of corresponding empirical processes to limiting Gaussian processes. Under some restrictions on F, such results are available in Wang (2003).

Classical multivariate analysis relies heavily on the assumption of multivariate normality, for which a variety of hypothesis tests has been developed. A number of these use the classical moment-based notion of kurtosis (Mardia, 1970, Malkovich and Afifi, 1973, and Isogai, 1983, for example). The kurtosis functional $k_F(\cdot)$ introduced here lends itself to an alternative approach based on Theorem 2.2. Details are provided in Wang (2003).

Quantile-based *peakedness* and *tailweight* measures are also of interest, and these are discussed in Section 3.1 below. An intuitive way to understand the functional $k_F(\cdot)$ and related peakedness and tailweight measures is through their *influence curves*, which are discussed in Section 3.2.

3.1 Peakedness and Tailweight Measures

As emphasized above, we interpret kurtosis as interrelated with peakedness and tailweight but not equated with either. Here we discuss these related measures.

A family of *tailweight* measures based on a quantile function through its volume functional is given by

$$t_F(r,s) = \frac{V_{F,\mathcal{C}}(r)(r)}{V_{F,\mathcal{C}}(r)(s)},\ 0 < r < s < 1, \tag{13}$$

which reduce in the univariate case to ratios of evaluations of the dispersion functional (5) at different points (see Balanda and MacGillivray, 1990, for discussion). Such an extension using depth-based central regions is proposed by Liu, Parelius and Singh (1999), who use the term "kurtosis" to mean tailweight and introduce a "fan plot" exhibiting in a single plot the two-dimensional curves $t_F(r,s)$ for a fixed choice of r and selected choices of s. Thus several multivariate distributions or data sets can be compared with respect to tailweight on the basis of their fan plots. These authors also introduce other forms of depth-based tailweight diagnostics, i.e, a Lorenz curve and a "shrinkage plot".

Bickel and Lehmann (1975) suggest that a measure of "kurtosis" (meaning tailweight) is given by any suitable ratio of two scale measures. Such a restriction is too restrictive for the more refined notion of kurtosis emphasized in the present paper to be interpreted as a tailweight measure. While typical tailweight measures are indeed of this form, the numerator of (2), for example, is not a scale measure (see MacGillivray and Balanda, 1988, and Balanda and MacGillivray, 1990, for discussion). Thus we properly distinguish our kurtosis measure as *not* a tailweight measure.

The term "peakedness" is traditionally considered synonymous with "concentration" and equivalent (inversely) to "dispersion" or "scatter". By this equivalence, peakedness too is distinct from kurtosis. For key definitions and developments regarding peakedness and dispersion, see in the univariate case Brown and Tukey (1946), Birnbaum (1948), and Bickel and Lehmann (1976), and in the multivariate case Sherman (1955), Eaton (1982), Oja(1983), Olkin and Tong (1988), and Zuo and Serfling (2000b).

3.2 Influence Curves

For quantile functions based on Type D depth (Zuo and Serfling, 2000a), which includes the halfspace depth, it is shown in Wang and Serfling (2003) that the influence function (IF) of the corresponding depth-based kurtosis $k_F(r)$ is a step function with jumps at the boundaries of the $(\frac{1}{2}-\frac{r}{2})$th central region (upward), the $\frac{1}{2}$th central region (downward), and the $(\frac{1}{2}+\frac{p}{2})$th central region (upward). One of these contours defines the interquartile region or "shoulders", and the other two demark the annuli of equal probability $r/2$ within and without the shoulders. This IF is bounded and thus has finite gross error sensitivity, in contrast with the unbounded IF's of moment-based kurtosis measures. In particular, for F elliptically symmetric with $h(\cdot)$ continuous and D the halfspace depth, we have

$$\begin{aligned}
&\mathrm{IF}(\boldsymbol{y}, k_{\cdot,D}(r), F) \\
&= \frac{2\Gamma(d/2+1)}{\pi^{d/2}[(F_R^{-1}(\frac{1}{2}+\frac{r}{2}))^{d/2}-(F_R^{-1}(\frac{1}{2}-\frac{r}{2}))^{d/2}]^2} \times \\
&\quad \Big\{[(F_R^{-1}(\tfrac{1}{2}))^{d/2}-(F_R^{-1}(\tfrac{1}{2}-\tfrac{r}{2}))^{d/2}] \cdot \frac{(\frac{1}{2}+\frac{r}{2}) - \mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \le F_R^{-1}(\frac{1}{2}+\frac{r}{2})\}}{h(F_R^{-1}(\frac{1}{2}+\frac{r}{2}))} \\
&\quad +[(F_R^{-1}(\tfrac{1}{2}+\tfrac{r}{2}))^{d/2}-(F_R^{-1}(\tfrac{1}{2}))^{d/2}] \cdot \frac{(\frac{1}{2}-\frac{r}{2}) - \mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \le F_R^{-1}(\frac{1}{2}-\frac{r}{2})\}}{h(F_R^{-1}(\frac{1}{2}-\frac{p}{2}))} \\
&\quad -[(F_R^{-1}(\tfrac{1}{2}+\tfrac{r}{2}))^{d/2}-(F_R^{-1}(\tfrac{1}{2}-\tfrac{r}{2}))^{d/2}] \cdot \frac{\frac{1}{2} - \mathbf{1}\{(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \le F_R^{-1}(\frac{1}{2})\}}{h(F_R^{-1}(\frac{1}{2}))}\Big\}.
\end{aligned}$$

As noted earlier, other affine invariant depth functions also yield ellipsoidal central regions in this case and hence the same kurtosis curves.

Influence functions are also available for the tailweight measures (13) and have similar features (Wang, 2003). For discussion of IF's of some kurtosis measures in the univariate case, see Ruppert (1987) and Groeneveld (1998).

Acknowledgment

Very constructive suggestions and insightful comments by an Associate Editor are greatly appreciated and have been utilized to improve the paper considerably. Also, support by NSF Grant DMS-0103698 is gratefully acknowledged.

References

[1] Balanda, K. P. and MacGillivray, H. L. (1988). Kurtosis: a critical review. *The American Statistician* **42** 111–119.

[2] Balanda, K. P. and MacGillivray, H. L. (1990). Kurtosis and spread. *The Canadian Journal of Statistics* **18** 17–30.

[3] Bickel, P. J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. I. Introduction. *Annals of Statistics* **3** 1038–1044.

[4] Bickel, P. J., and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Annals of Statistics* **4** 1139–1158.

[5] Brown, G., and Tukey, J. W. (1946). Some distributions of sample means. *Annals of Mathematical Statistics* **7** 1–12.

[6] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872.

[7] Chissom, B. S. (1970). Interpretation of the kurtosis statistic. *The American Statistician* **24**, (6) 19–22.

[8] Darlington, R. B. (1970). Is kurtosis really "peakedness"? *The American Statistician* **24**, (2) 19–22.

[9] Eaton, M. L. (1983). *Multivariate Statistics*. John Wiley & Sons, New York.

[10] Fang, K.-T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall.

[11] Finecan, H. M. (1964). A note on kurtosis. Journal of the Royal Statistical Society, *Series B* **26** 111–112.

[12] Gilchrist, W. G. (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall.

[13] Groeneveld, R. A. (1998). A class of quantile measures for kurtosis. *The American Statistician* **51** 325–329.

[14] Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician* **33** 391–399.

[15] Hildebrand, D. K. (1971). Kurtosis measures bimodality? *The American Statistician* **25**, (1) 42–43.

[16] Horn, P. S. (1983). A measure for peakedness. *The American Statistician* **37**, (1) 55–56.

[17] Isogai, T. (1983). On measures of multivariate skewness and kurtosis. *Mathematica Japonika* **28** 251–261.

[18] Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Annals of Statistics* **27** 783–858.

[19] MacGillivray, H. L. and Balanda, K. P. (1988). The relationships between skewness and kurtosis. *Australian Journal of Statistics* **30** 319–337.

[20] Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association* **68** 176–179.

[21] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57** 519–530.

[22] Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician* **40** 283–284.

[23] Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics* **8** 154–168.

[24] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters* **1** 327–333.

[25] Olkin, I. and Tong, Y. L. (1988). Peakedness in multivariate distributions. In *Statistical Decision Theory and Related Topics IV*, Vol. 2 (S.S. Gupta and J. O. Berger, eds.), pp. 373–384, Springer.

[26] Ruppert, D. (1987). What is kurtosis? An influence function approach. *The American Statistician* **41** 1–5.

[27] Serfling, R. (2003). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference*, in press.

[28] Sherman, S. (1955). A theorem on convex sets with applications. *Annals of Mathematical Statistics* **26** 763–767.

[29] Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Institute Review* **58** 263–277.

[30] Wang, J. (2003). *On Nonparametric Multivariate Scale, Kurtosis, and Tailweight Measures.* Ph.D. dissertation, Department of Mathematical Sciences, University of Texas at Dallas, 2003.

[31] Wang, J. and Serfling, R. (2003). Influence functions for a general class of depth-based generalized quantile functions. Submitted.

[32] Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *Annals of Statistics* **28** 461–482.

[33] Zuo, Y. and Serfling, R. (2000b). Nonparametric notions of multivariate "scatter measure" and "more scattered" based on statistical depth functions. *Journal of Multivariate Analysis* **75** 62–78.