Austral. J. Statist., 30(3), 1988, 319–337

THE RELATIONSHIPS BETWEEN SKEWNESS AND KURTOSIS

H.L. MacGillivray and K.P. Balanda

Department of Mathematics, University of Queensland, St. Lucia, QLD, 4067, Australia.

Summary

Theoretical considerations of kurtosis, whether of partial orderings of distributions with respect to kurtosis or of measures of kurtosis, have tended to focus only on symmetric distributions. With reference to historical points and recent work on skewness and kurtosis, this paper defines anti-skewness and uses it as a tool to discuss the concept of kurtosis in asymmetric univariate distributions. The discussion indicates that while kurtosis is best considered as a property of symmetrised versions of distributions, symmetrisation does not simply remove skewness. Skewness, anti-skewness and kurtosis are all inter-related aspects of shape. The Tukey g and h family and the Johnson S_U family are considered as examples.

Key words: Anti-skewness; asymmetry; kurtosis; measures; moments; orderings; quantiles; shape; skewness; spread.

1. Introduction

The concepts of skewness and kurtosis were both introduced through parameters that were felt to be representative measures of certain shape properties of distributions (Pearson, 1895; Bowley, 1901). Since then, other skewness and kurtosis parameters have been introduced for a variety of reasons, although for many statisticians the word kurtosis is still attached to the original parameter, the standardised fourth central moment. All are location and scale-free in the sense that they are invariant under a linear transformation $aX + b$ of the random variable X except for a change of sign for skewness measures for $a < 0$. Almost all the kurtosis measures are ratios of measures of scale, as suggested by Bickel & Lehmann (1976), and skewness measures are zero for symmetric distributions. Interpretation of these measures has received considerable attention and, particularly for kurtosis, is still doing so (for example, Ruppert (1987)). Such interpretations usually consider conditions on two distributions which imply an ordering on a parameter; this includes considering changes to a distribution that increase a parameter.

Unfortunately, these implications are often erroneously used in reverse. For example, it is incorrect to assume that an increase in the standardised third central moment implies "longer right tails and shorter left tails" of the distribution, or that an increase in a single-valued kurtosis parameter implies "heavier tails" and/or "more peakedness". Reverse implications hold only if the distributions are ordered by the appropriate conditions. Dissatisfaction with such use of single-valued parameters contributed to the partial ordering approach to skewness and kurtosis (van Zwet, 1964; Oja, 1981). In such an approach only parameters that preserve a skewness or kurtosis ordering are used as measures of the property for ordered distributions.

Such partial orderings have been used to some extent in areas such as assessment of inferential procedures (for example, Lawrence (1974) and Loh (1984)). There is a need for understanding and using a range of distributional shapes (Barnard, 1987). For example, in detailed studies of the performance of location estimators in symmetric distributions (Crow & Siddiqui, 1967; Andrews *et al.*, 1972; Rosenberger & Gasko, 1983), comparisons are made for a small group of distributions such as normal, double exponential, Cauchy, slash, logistic and one or two specific contaminated normals, in the "hope that these distributions span the range of reasonable situations" (Rosenberger & Gasko, 1983). Yet the shape comparison of the double exponential and the Cauchy is not simple (Balanda, 1987).

Formulisation of the concept of shape must identify the roles of different orderings and measures of skewness and kurtosis that have been used in different contexts, remembering that the weaker the conditions imposed by the ordering, the larger the distributional classes that can be ordered. MacGillivray (1986) considers skewness and asymmetry; Balanda (1986) and Balanda & MacGillivray (1987) consider kurtosis. However questions that arise in applications require also joint consideration of skewness and kurtosis. Questions include: how do we interpret kurtosis parameters in asymmetric distributions? can we formulate skewness and/or kurtosis alternatives as we can location or scale alternatives? can we separate shape into skewness and kurtosis components or is it some indefinable mixture of the two?

Parameters purporting to measure kurtosis are used for both symmetric and asymmetric distributions, but almost all theoretical considerations of kurtosis have been restricted to the symmetric case. Orderings with respect to kurtosis, and hence peakedness and tailedness, for asymmetric distributions are needed not only for theoretical applications but also to justify and interpret kurtosis parameters when used for asymmetric distributions.

Balanda & MacGillivray (1987) extend work on kurtosis in symmetric distributions to asymmetric distributions by exploiting the role of spread in kurtosis. This approach identifies the orderings behind all measures of kurtosis that are ratios of measures of spread. However this definition essentially considers kurtosis as a property of symmetrised versions of distributions and raises questions relating to distributional equivalence and interpretation of kurtosis parameters in terms of the original asymmetric distributions. In examining these questions, this paper defines and uses antiskewness as a tool in understanding the roles of skewness and kurtosis in describing shape; antiskewness is a possible alternative generalisation of kurtosis to the asymmetric case. We conclude that the definition of kurtosis in terms of spread functions is more consistent with current usage, but that the roles of skewness and kurtosis in describing shape are interrelated. These inter-relationships are examined here to enhance understanding not only of the properties themselves but also of the parametrs that measure them.

Asymmetric distributions require a choice of location reference point; both the median and the mean are considered here. The median corresponds to quantile-based orderings and measures which are potentially richer and appear to have found more use in applications than moment-based orderings using the mean as location parameter. However because the third and fourth standardised central moments are the traditional parameters of skewness and kurtosis, it is important that they be considered; problems associated with their interpretation are mentioned where appropriate. For simplicity of discussion we consider here only continuous random variables whose densities have interval support.

2. Kurtosis in Asymmetric Distributions

2.1 Kurtosis Orderings and Spread Functions

Van Zwet's (1964) skewness ordering on continuous univariate distributions defines the distribution $F(x)$ as being less skew to the right than the distribution $G(x)$, $F \leq_c G$, iff $G^{-1}[F(x)]$ is convex on $I_F = \{x;\ 0 < F(x) < 1\}$. For symmetric distributions $F(x)$ and $G(x)$, van Zwet's kurtosis ordering defines F as having less kurtosis than G, $F \leq_s G$, iff $G^{-1}[F(x)]$ is convex for $x > \mu_F$ ($\equiv m_F$) in I_F; the symmetry of F and G then dictates that $G^{-1}[F(x)]$ is concave for $x < \mu_F$ ($\equiv m_F$), where μ_F, m_F denote the mean, median of F. A distribution F is skew to the right according to $\leq_c$ if $\bar{F} \leq_c F$, where $\bar{F}(x) = 1 - F(-x)$ is the distribution of $-X$ when $X \sim F$. $F \leq_c G$ and $G \leq_c F$ if and only if there exist constants a, b such that $F(x) = G(ax + b)$; similarly for $F \leq_s G$ and $G \leq_s F$ in the class of symmetric F and G, and hence we give the following definition which is consistent with statistical usage.

Definition 1. The distributions F and G have the same shape if there exist constants, a, b such that $F(x) = G(ax + b)$ for all x.

The convexity of $G^{-1}[F(x)]$ on an interval I may be re-expressed by either of two characterisations.

(i) $Q(x) \equiv [G^{-1}\{F(x)\} - G^{-1}\{F(x_0)\}]/(x - x_0)$ is non-decreasing in I

$$\text{for any } x_0 \in I \tag{2.1}$$

or, using Karlin's S^- notation (1968, p.20) to denote number of sign changes,

(ii) $$S^-[G^{-1}\{F(x)\} - ax - b] \leq 2 \quad \text{in } I \text{ for any } a, b\,, \tag{2.2}$$

with the sign changes from positive to negative to positive if equality holds. Condition (2.1) is the starting point for quantile-based orderings and measures while condition (2.2) is more appropriate to moment-based measures.

Lawrence (1974), Oja (1981) and Loh (1984) have defined orderings that weaken the convexity condition, while van Zwet (1979) and Doksum (1975) have considered conditions that order $\bar{F}$ and F. MacGillivray (1986) and Balanda & MacGillivray (1987) absorb these orderings into hierarchies of skewness and kurtosis orderings respectively, identifying the orderings behind all measures of skewness and kurtosis that have been used.

Van Zwet's orderings have no reference to any particular measures of location or scale; these are introduced as the convexity condition is weakened. Starting with (2.1) there are three main steps in the weakening procedure with different branches of the hierarchies corresponding to different choices of reference points. The first step chooses a particular x_0 in (2.1) giving the star-shaped orderings of Lawrence (1974) and Oja (1981); x_0 determines the location reference point. The next step generalises Doksum's (1975) concept of strong skewness and involves a second choice of reference point which is particularly important in kurtosis as it determines the scaling measure. The third step considers crossings of the resultant scaled distributions. Previous work corresponds to choosing the median as both reference points.

For kurtosis, by expressing the orderings in terms of the spread function

$$S_F(u) = F^{-1}(u) - F^{-1}(1-u)\,, \quad u \in (\tfrac{1}{2}, 1)\,,$$

instead of $F^{-1}(u)$, they may be generalised immediately to asymmetric distributions. For example, $F \leq_s G$ becomes $S_G(S_F^{-1}(x))$ convex for $x > 0$.

The branch of the kurtosis hierarchy with the median as both reference point choices is given by (Balanda & MacGillivray, 1987)

$$F \leq_s G \Rightarrow F \underset{\text{star}}{<^0} G \Rightarrow F \underset{\text{D}}{\leq^0_0} G \Rightarrow F \leq^0_0 G \begin{array}{l} \Longrightarrow F \leq^0_0 G(\text{tail}) \\ \Longrightarrow F \leq^0_0 G(\text{peak})\,. \end{array} \tag{2.3}$$

The D-level is so-named because it generalises Doksum's (1975) concept of strong skewness.

For the orderings $\leq_s$, $\underset{\text{star}}{<^0}$, $\underset{\text{D}}{\leq^0_0}$ and $\leq^0_0$, we have $F \leq G$ and $G \leq F$, if and only if the spread functions of F and G differ only by a constant multiple; symmetric F and G will thus have the same shape.

The part of the skewness hierarchy (MacGillivray, 1986) that takes the median as both reference point choices is, using Oja's (1981) $\leq_2$ instead of $\leq_c$,

$$F \leq_2 G \Rightarrow F \underset{\text{star}}{\leq^m_2} G \Rightarrow F \underset{\text{D}}{\leq^m_2} G \Rightarrow F \leq^m_2 G \begin{array}{l} \Longrightarrow F \leq^{m,t}_2 G \\ \Longrightarrow F \leq^{m,c}_2 G\,. \end{array} \tag{2.4}$$

Distributions are equivalent with respect to any of the orderings $\leq_2$, $\underset{\text{star}}{\leq^m_2}$, $\underset{\text{D}}{\leq^m_2}$ and $\leq^m_2$, if and only if they have the same shape.

The spread function generalisation defines kurtosis as a property of the class of distributions whose spread functions are equal up to a constant multiple, and hence for an asymmetric distribution as the kurtosis of its symmetrised version F_0 defined by $F_0^{-1}(u) = \frac{1}{2}S_F(u) + m_F$. F_0 is the unique symmetric distribution with median m_F having the same spread function, $S_F(u)$ as F. Amongst symmetric distributions with median m_F, F_0 also most closely approximates F in the sense of minimising $\sup_u |F^{-1}(u) - F_0^{-1}(u)|$ (Doksum, 1975).

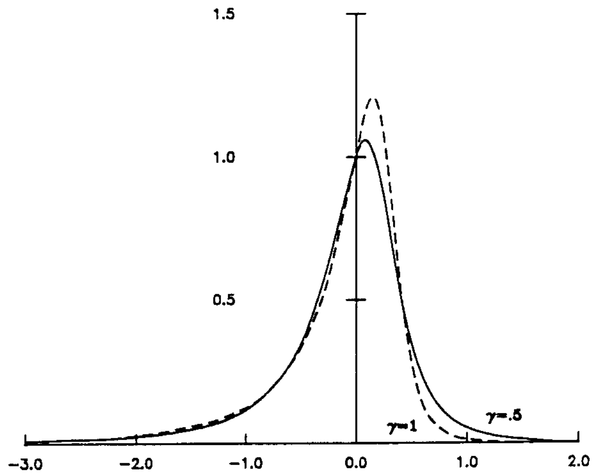

Fig. 1.—Densities of Johnson S_U distributions $F_{\gamma,\delta}$; $F_{1,1} \underset{D}{<_2^m} F_{.5,1}$.

Figure 1 shows the densities of two distributions ordered by $\underset{D}{\leq_2^m}$ but whose spread functions differ only by a constant multiple. They are members of the S_U Johnson family with distribution function $F_{\gamma,\delta}(x) = \Phi(\delta \sinh^{-1} x+\gamma)$, where Φ is the standard normal. Here $\delta = 1$ and $\gamma = .5, 1$, the densities are standardised to have median $= 0$ and density at the median $= 1$, and it can be shown that $F_{1,1} \underset{D}{<_2^m} F_{.5,1}$.

The orderings of (2.3) and (2.4) have arisen from a variety of uses. Many measures preserve the weakest level of orderings and hence the preceding ones in the hierarchy, so that only the weakest ordering is required for interpretation of the measures. An important point is that $S_F(v)/S_F(u), u < v$, used by Crow & Siddiqui (1967) and Andrews *et al.* (1972), for particular values of u and v, does not preserve $\leq_0^0$ but does preserve $\underset{star}{\leq^0}$. A different branch of the kurtosis hierarchy involving a different second reference point is required for this measure; Section 2.3 discusses this.

Using $a = \sigma_G/\sigma_F$ and $b = \mu_G - \sigma_G\mu_F/\sigma_F$ in (2.2) considers the mean and standard deviation as measures of location and scale and leads to moment measures of shape. Van Zwet (1964) proved that the standardised odd moments μ_{2k+1}/σ^{2k+1} preserve $\leq_2$ and the standardised even moments μ_{2k}/σ^{2k} preserve $\leq_s$ in the symmetric case, these results following from (2.2) with the above values of a, b. The general results may be stated as follows, denoting the rth central moment of a distribution F by $\mu_{F,r}$ if it exists.

(a) If

$$S^-[F(\mu_F + \sigma_F x) - G(\mu_G + \sigma_G x)] = 2 \quad \text{for } -\infty < x < \infty \,, \quad \text{from } \geq 0 \text{ to } \leq 0 \text{ to } \geq 0 \tag{2.5}$$

then

$$\mu_{F,2k+1}/\sigma_F^{2k+1} < \mu_{G,2k+1}/\sigma_G^{2k+1} \,, \quad k = 1, \ldots \;.$$

(b) If F and G are symmetric distributions and

$$S^-[F(\mu_F+\sigma_F x)-G(\mu_G+\sigma_G x)] = 1 \quad \text{for } x > 0, \text{ from } \leq 0 \text{ to } \geq 0 \,, \tag{2.6}$$

then

$$\mu_{F,2k}/\sigma_F^{2k} < \mu_{G,2k}/\sigma_G^{2k} \,, \quad k = 2, \ldots \;.$$

In (b) symmetry dictates the sign changes for $x < 0$ analogous to (2.5).

Oja's (1981) orderings $<_2^*$ and $<_s^*$ are the special cases of (2.5) and (2.6) that are implied by $<_2$ and $<_2$ respectively. MacGillivray's (1986) $<_2^\mu$ is (2.5). For proofs of (a) and (b) see Oja (1981) and MacGillivray (1985).

Whereas all the quantile-based orderings are transitive, none of $<_2^*$, $<_2^\mu$ and $<_s^*$ are, and hence it is debatable whether they can be called orderings. However they are useful because distribution crossings are basic to interpreting the standardised moments, and $<_2^*$ is important in considering the sign of μ_{2k+1} in skewed distributions (see, for example, Runnenburg 1978; MacGillivray, 1981 and 1982). It should be stressed that if the standardised distributions cross more than the minimum number of times, any relationship between the standardised third or fourth central moments is possible.

Symmetrising to F_0 can be thought of as symmetrising with respect to the median. F can also be symmetrised with respect to the mean to give the distribution

$$F^0(x) = \left[F(x) - F(2\mu_F - x) + 1\right] \,, \quad -\infty < x < \infty \,. \tag{2.7}$$

$F^0(x)$ is the unique distribution symmetric around μ_F for which $\sup_x |F(x) - F^0(x)|$ is minimised. If, for a distribution F, we define

$$\varphi_F(x) = F(\mu_F + x) - F(\mu_F - x) \,, \quad x > 0$$

then $F^0(x)$ is also the unique distribution symmetric around μ_F with the same φ function as F. $\varphi_F \leq \varphi_G$ is Bickel & Lehmann's (1976) dispersion ordering $\leq_{\text{disp}}$ generalised directly to asymmetric distributions. It is preserved by all the even central moments and φ_F can be thought of as an alternative spread function to $S_F(u)$; thus F^0 has the same alternative spread function as F.

If $F^0 <_s^* G^0$, then $\mu_{F,2k}/\sigma_F^{2k} < \mu_{G,2k}/\sigma_G^{2k}$, $k \geq 2$. Therefore, if kurtosis with respect to the mean is defined for asymmetric distributions by the orderings $<_s$ and $<_s^*$ on the symmetrised versions F^0 and G^0, the standardised even moments preserve these generalised orderings.

2.2 Anti-Skewness

In defining partial orderings such as $\leq_s$ on symmetric distributions, the condition needs to be stated only for one side of the centre since symmetry dictates the behaviour of the other side. Hence another way of generalising the ordering to asymmetric distributions is to specify the behaviour that is imposed by symmetry. This requires firstly a choice of location reference point. The method is then equivalent to imposing a skewness condition on one side of this point and the reverse condition on the other side. Taking the median as location point and using skewness and reverse skewness conditions with respect to the median gives the following structure, where f and g denote the densities of F and G respectively.

Definition 2.

$$F \leq_a G \text{ iff } G^{-1}(F(x)) \text{ is } \begin{pmatrix}\text{concave}\\ \text{convex}\end{pmatrix} \text{ for } x \binom{\leq}{\geq} m_F\,. \tag{2.8}$$

$$\begin{aligned} F \underset{\text{star}}{\leq_a^m} G &\text{ iff } [G^{-1}(u) - m_G]/[F^{-1}(u) - m_F] \\ &\text{is } \begin{pmatrix}\text{nonincreasing}\\ \text{nondecreasing}\end{pmatrix} \text{ for } u \binom{\leq}{\geq} \tfrac{1}{2} \text{ in } [0,1]\,. \end{aligned} \tag{2.9}$$

$$\begin{aligned} F \underset{\text{D}}{\leq_a^m} G &\text{ iff } G^{-1}(u)g(m_G) - F^{-1}(u)f(m_F) \\ &\text{is nondecreasing for } u \in [0,1]\,. \end{aligned} \tag{2.10}$$

$$\begin{aligned} F \leq_a^m G &\text{ iff } [G^{-1}(u) - m_G]g(m_G) \binom{\leq}{\geq} [F^{-1}(u) - m_F]f(m_F) \\ &\text{for } u \binom{\leq}{\geq} \tfrac{1}{2} \text{ in } [0,1]\,. \end{aligned} \tag{2.11}$$

$$\begin{aligned} F \leq_a^{m,c} G &\text{ iff (2.11) holds for } \binom{\frac{1}{2}-\alpha}{\frac{1}{2}} \leq u \leq \binom{\frac{1}{2}}{\frac{1}{2}+\alpha} \\ &\text{for some } \alpha \in (0, \tfrac{1}{2})\,. \end{aligned} \tag{2.12}$$

$$\begin{aligned} F \leq_a^{m,t} G &\text{ iff (2.11) holds for } \binom{0}{1-\alpha} \leq u \leq \binom{\alpha}{1} \\ &\text{for some } \alpha \in (0, \tfrac{1}{2})\,. \end{aligned} \tag{2.13}$$

Theorem 1.

$$F \le_a G \Rightarrow F \underset{\text{star}}{\le_a^m} G \Rightarrow F \underset{\text{D}}{\le_a^m} G \Rightarrow F \le_a^m G \begin{array}{l} \Longrightarrow F \le_a^{m,c} G \\ \Longrightarrow F \le_a^{m,t} G\,. \end{array} \tag{2.14}$$

Proof follows immediately from Theorem 2.2 of MacGillivray (1986).

Because of their construction we call these orderings anti-skewness orderings. When F and G are symmetric, the orderings in (2.14) coincide with the kurtosis orderings in (2.3). The anti-skewness orderings are reflexive and transitive, and as with the skewness orderings, distributions are equivalent with respect to any of the orderings $\le_a$, $\underset{\text{star}}{\le_a^m}$, $\underset{\text{D}}{\le_a^m}$ and $\le_a^m$, if and only if they have the same shape.

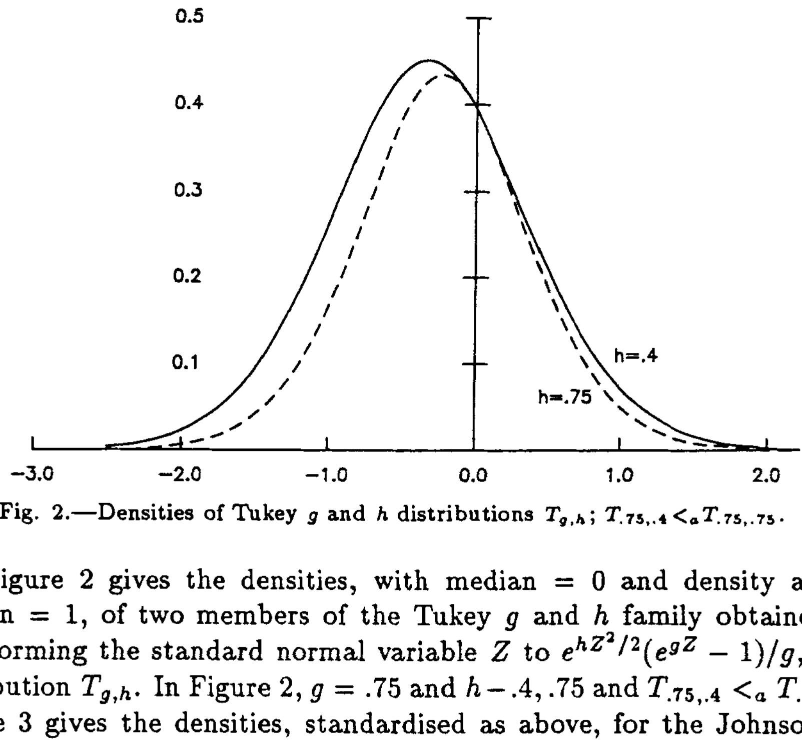

Fig. 2.—Densities of Tukey g and h distributions $T_{g,h}$; $T_{.75,.4} <_a T_{.75,.75}$.

Figure 2 gives the densities, with median $= 0$ and density at the median $= 1$, of two members of the Tukey g and h family obtained by transforming the standard normal variable Z to $e^{hZ^2/2}(e^{gZ}-1)/g$, with distribution $T_{g,h}$. In Figure 2, $g = .75$ and $h - .4, .75$ and $T_{.75,.4} <_a T_{.75,.75}$. Figure 3 gives the densities, standardised as above, for the Johnson S_U distributions with $\gamma = .5$ and $\delta = .8, 1.5$; it can be shown that $F_{.5,.8} <_2^{m,c} F_{.5,1.5}$ and $F_{.5,1.5} <_a^{m,t} F_{.5,.8}$.

The following corollary to Theorem 1 gives two measures that are invariant under linear transformations of random variables and preserve the orderings stated, and hence the orderings preceding in the hierarchy (2.14).

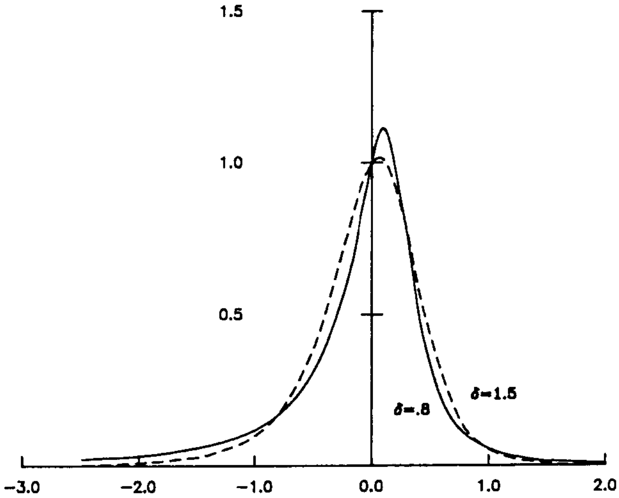

Fig. 3.—Densities of Johnson S_U distributions $F_{\gamma,\delta}$; $F_{.5,.8} <_2^{m,c} F_{.5,1.5}$ and $F_{.5,1.5} <_a^{m,t} F_{.5,.8}$.

Corollary 1

(i) $S_F(u)f(m)$ *preserves* $\leq_a^m$, $\frac{1}{2} \leq u < 1$

(ii) $[S_F(v) - S_F(u)]f(m)$ *preserves* $\underset{D}{\leq_a^m}$, $\frac{1}{2} \leq u < v < 1$.

Proof: follows from (2.11) and (2.10).

$S_F(v)/S_F(u)$ does not preserve any anti-skewness ordering; see Section 2.3.

$S_F(u)f(m)$ also preserves $\leq_a^{m,c}$ for $\frac{1}{2} \leq u \leq \frac{1}{2}+\alpha$ and $\leq_a^{m,t}$ for $1-\alpha \leq u \leq 1$. Horn (1983) uses a parameter that increases with $f(m)[F^{-1}(u)-m]$ for $u > \frac{1}{2}$ as a measure of peakedness for symmetric distributions. As it stands $f(m)[F^{-1}(u) - m]$ preserves both $\leq_2^m$ and $\leq_a^m$ for asymmetric distributions; generalising it through the spread function gives $S_F(u)f(m)$. Groeneveld & Meeden (1984) suggest using $[F^{-1}(.75+p)+F^{-1}(.75-p)-2\pi_F]/[F^{-1}(.75+p) - F^{-1}(.75-p)]$, $0 < p < \frac{1}{4}$ as a measure of kurtosis for symmetric distributions where $\pi_F \equiv F^{-1}(.75)$. There is an immediate similarity between this measure and the $\gamma_F(u)$ skewness measures, where $\gamma_F(u) = [F^{-1}(u)+F^{-1}(1-u)-2m_F]/[F^{-1}(u)-F^{-1}(1-u)]$, $\frac{1}{2} < u < 1$. Groeneveld & Meeden's measure preserves both $\leq_2$ and $\leq_a$ since it refers only to behaviour on the right of the median. It arises through choosing $x_0 = \pi_F$ in (2.1), and again choosing π_F as the second reference point. As it stands, although it preserves $\leq_s$ for symmetric distributions, it does not preserve $\leq_s$ in general, nor any of the orderings derived from $\leq_s$ in the general case. However substituting $\frac{1}{2}S_F(u) + m_F$ for $F^{-1}(u)$ gives $[S_F(.75+p) + S_F(.75-p) - 2S_F(.75)]/[S_F(.75+p) - S_F(.75-p)]$ which does preserve $\le_s$ for all distributions. Horn's and Groeneveld & Meeden's measures illustrate the problems that can arise if kurtosis measures are introduced for symmetric distributions but then also applied to asymmetric distributions. It should be noted that the numerator in both the original version and our generalisation of Groeneveld & Meeden's measure does not preserve Bickel & Lehmann's (1976) spread-ordering and so it is not a ratio of scale measures.

The measures in Corollary 1 also preserve $\le_0^0$ and $\underset{\mathrm{D}}{\le_o^o}$ respectively and a question that arises immediately is in what way are the kurtosis and anti-skewness orderings related. The following theorem provides some answers. The skewness mesure $\nu_F(u)$ is given by $[F^{-1}(u)+F^{-1}(1-u)-2m_F]f(m_F)$.

Theorem 2.

(i) $F \le_a^m G \Rightarrow F \le_0^0 G$, *and if* $\gamma_F(u) = \gamma_G(u)$ *for all* $u \in (\frac{1}{2}, 1)$, *then* $F \le_a^m G \Leftrightarrow F \le_0^0 G$.

(ii) $F \underset{\mathrm{D}}{\le_a^m} G \Rightarrow F \underset{\mathrm{D}}{\le_o^o} G$, *and if* $\nu_F(u) = \nu_G(u)$ *for all* $u \in (\frac{1}{2}, 1)$, *then* $F \underset{\mathrm{D}}{\le_a^m} G \Leftrightarrow F \underset{\mathrm{D}}{\le_o^o} G$.

(iii) If $\gamma_F(u) = \gamma_G(u)$ *for all* $u \in (\frac{1}{2}, 1)$, *then* $F \underset{\mathrm{star}}{\le_a^m} G \Leftrightarrow F \underset{\mathrm{star}}{\le^o} G$.

(iv) If $f[F^{-1}(u)]/g[G^{-1}(u)] = f[F^{-1}(1-u)]/g[G^{-1}(1-u)]$ *then* $F \le_a G \Leftrightarrow F \le_s G$.

Proof: (i) $F \le_a^m G \Rightarrow S_G(u)g(m_G) \ge S_F(u)f(m_F) \Leftrightarrow F \le_0^0 G$.

$$\gamma_F(u) = \gamma_G(u) \Leftrightarrow \frac{G^{-1}(1-u) - m_G}{F^{-1}(1-u) - m_F} = \frac{G^{-1}(u) - m_G}{F^{-1}(u) - m_F} \tag{2.15}$$

$F \le_0^0 G$ and (2.15) $\Rightarrow F \le_a^m G$.

(ii) Similarly, $F \underset{\mathrm{D}}{\le_a^m} G$ implies that $S_G(u)g(m_G) - S_F(u)f(m_F)$ is non-decreasing for $u \in (\frac{1}{2}, 1)$, that is, $F \underset{\mathrm{D}}{\le_o^o} G$. Denote $f[F^{-1}(u)]$ by $q_F(u)$ and similarly for $q_G(u)$. If $\nu_F(u) = \nu_G(u)$, the definition of $F \underset{\mathrm{D}}{\le_o^o} G$ becomes $2[g(m_G)/q_G(u) - f(m_F)/q_F(u)] \ge 0$, which is the definition of $F \underset{\mathrm{D}}{\le_a^m} G$.

(iii) $F \underset{\mathrm{star}}{\le_a^m} G$ and (2.15) give, with some inequality manipulation, $S_G(v)S_F(u) \ge S_G(u)S_F(v)$, for $\frac{1}{2} < u < v < 1$, which is $F \underset{\mathrm{star}}{\le^o} G$. Conversely, $F \underset{\mathrm{star}}{\le^o} G$ and (2.15) give $[G^{-1}(v) - m_G][F^{-1}(u) - m_F] \ge [G^{-1}(u) - m_G][F^{-1}(v) - m_F]$ for $\frac{1}{2} < u < v < 1$, and the reverse inequality for $0 < u < v < \frac{1}{2}$, that is $F \underset{\mathrm{star}}{\le_a^m} G$.

(iv)

$$F \leq_a G \Leftrightarrow q_F(u)/q_G(u) \begin{pmatrix} \text{nonincreasing} \\ \text{nondecreasing} \end{pmatrix}$$

for $u \left(\substack{<\\>}\right) \frac{1}{2}$. If $q_F(u)/q_G(u) = q_F(1-u)/q_G(1-u)$, then $q_{F_0}(u)/q_{G_0}(u) = q_F(u)/q_G(u)$ and $F \leq_a G \Leftrightarrow F \leq_s G$.

We now consider the mean as location reference point. We could define another generalisation of van Zwet's $\leq_s$ to concavity-convexity about the mean but this is not transitive and is useful only in implying three crossings of the standardised distributions; we consider the case of three crossings.

Definition 3.

$$F <^*_\mu \text{ iff } S^-\{F(\mu_F + \sigma_F x) - G(\mu_G + \sigma_G x)\} = 3 \text{ for } -\infty < x < \infty$$
$$\text{from } \leq 0 \text{ to} \geq 0 \text{ to} \leq 0 \text{ to} \geq 0\,. \tag{2.16}$$

For F and G symmetric, (2.16) $\Leftrightarrow$ (2.6). The 'ordering' $<^*_\mu$ is a natural generalisation of $<^*_s$ in that the two distributions, after standardisation to have zero means and unit variances, cross each other exactly three times (they must cross at least three times), with G having the heavier tails. The behaviour near the centre of the distributions is not as readily interpretable for $<^*_\mu$ as for $<^*_s$, but then for asymmetric distributions the centre itself is not well defined. Again $<^*_\mu$ is not transitive. There is also no theorem corresponding to Theorem 2.

If (2.16) holds with the sign changes occurring at a_1, a_2, a_3, then

$$\left[\frac{\mu_{F,4}}{4\sigma_F^4} - \frac{\mu_{G,4}}{4\sigma_G^4}\right] - (a_1 + a_2 + a_3)\left[\frac{\mu_{F,3}}{3\sigma_F^3} - \frac{\mu_{G,3}}{3\sigma_G^3}\right] < 0\,. \tag{2.17}$$

This generalises Dyson (1943).

Although (2.17) indicates to some extent the role of μ_4/σ^4 and the influence of the skewness through μ_3/σ^3, the interpretation of Σa_i is difficult. Results relating distribution crossings to moment crossings follow from the variation diminishing properties of the strictly totally positive kernel x^r for $x \geq 0$ and all real r (Karlin, 1968, p.15). Thus the essential results are for the moments of the positive and negative parts of the standardised distributions, with results for odd and even moments as corollaries. If $F <^*_\mu G$, similar techniques to those of MacGillivray (1985) give results which can probably best be summarised by saying that either $\mu_{F,4}/\sigma_F^4 \leq \mu_{G,4}/\sigma_G^4$, or the third and fourth moments of the positive and negative parts of the (standardised) random variables behave as they would if the distributions were comparable with respect to the skewness ordering $<_2^*$ (or $<_2$). Even if $a_2 = 0$ we can say only that at least one of $\mu_{F,4}/\sigma_F^4$ and $\mu_{F,3}/\sigma_F^3$ is less than the corresponding parameter for G.

2.3 Anti-Skewness and Kurtosis

Anti-skewness is a possible generalisation of kurtosis to the asymmetric case but there are two major flaws, described in (a) and (b) below.

(a) The above comments show that $<_\mu^*$ can not be used to interpret μ_4/σ^4 whereas this parameter arises naturally in comparisons of distributions symmetrised with respect to the mean as in (2.7). The anti-skewness orderings do not generally give rise to measures that are ratios of measures of scale, and in particular $S_F(v)/S_F(u)$ does not preserve $\leq_a$. A couple of examples illustrate this point.

As described in Section 2.1, the successive levels in the hierarchies (2.3), (2.4) and (2.14) require choices of reference points which determine location reference point and scaling method. To show that $S_F(v)/S_F(.75)$ preserves $\underset{\text{star}}{\leq^{\circ}}$, it is necessary to choose the 75% quantile of the symmetrised versions as the second reference point, thus choosing the interquartile range as the measure of scale instead of $1/f(m_F)$. A similar procedure may also be used in successively weakening $\leq_a$ but the interquartile range is not obtained and hence measures such as $S_F(v)/S_F(.75)$ do not arise.

In weakening $\leq_a$, if the median is chosen at the first step to the star-shaped level, we get $\underset{\text{star}}{\leq_a^m}$. If we then consider the quartiles at the next step to the Doksum level we obtain after a little algebra, denoting 75% quartiles by π and 25% quartiles by η, that

$$F \underset{\text{star}}{\leq_a^m} G \Rightarrow \begin{array}{l} [G^{-1}(u) - m_G]/(\pi_G - m_G) + [G^{-1}(1-u) - m_G]/(\eta_G - m_G) \\ -[F^{-1}(u) - m_F]/(\pi_F - m_F) - [F^{-1}(1-u) - m_F]/(\eta_F - m_F) \\ \left\{ \begin{array}{c} \geq \\ \leq \end{array} \right\} 0 \text{ for } \left\{ \begin{array}{c} .75 < u < 1 \\ .5 < u < .75 \end{array} \right\} . \end{array} \tag{2.18}$$

Similarly, if we choose the quartiles at the first step to the star-shaped level and then again at the next step to the Doksum level, we obtain

$$\begin{array}{l} g(\pi_G)(G^{-1}(u) - \pi_G) - g(\eta_G)(G^{-1}(1-u) - \eta_G) \\ -f(\pi_F)(F^{-1}(u) - \pi_F) - f(\eta_F)(F^{-1}(1-u) - \eta_F) \\ \left\{ \begin{array}{c} \geq \\ \leq \end{array} \right\} 0 \text{ for } \left\{ \begin{array}{c} .75 < u < 1 \\ .5 < u < .75 \end{array} \right\} . \end{array} \tag{2.19}$$

In comparison the kurtosis orderings on asymmetric distributions that are obtained by symmetrising to F_0 give rise to the measures

$$[(F^{-1}(u) - F^{-1}(1-u)]/(\pi_F - \eta_F) = S_F(u)/S_F(.75)\,,$$

and

$$[F^{-1}(u) - F^{-1}(1-u)]/[1/f(\pi_F) + 1/f(\eta_F)]\,.$$

These measures are ratios of measures of scale, whereas the measures of (2.18) and (2.19) are sums of ratios of measures of scale; in symmetric distributions the two types coincide. To obtain ratios of measures of scale as Bickel & Lehmann (1976) recommend, including those that have been used quite extensively as measures of kurtosis or of tailweight only, it is necessary to consider kurtosis as a property of the spread functions of the distributions, and thus essentially as a property of the symmetrised versions.

(b) From the definitions it follows that two distributions cannot be compared simultaneously with respect to skewness and anti-skewness within the same probability range unless they have the same shape, that is, are equal with respect to these properties, within that range. Hence if kurtosis is defined as the anti-skewness property, it becomes an alternative to skewness; two distributions of different shape would be comparable (within a certain probability range) *either* with respect to skewness *or* with respect to kurtosis but not both. This occurs to some extent when kurtosis considerations are limited to symmetric distributions, since symmetric distributions are not comparable according to a skewness ordering unless they are actually (shape) equivalent. However, kurtosis as an alternative to skewness is unlikely to appeal to statisticians, especially as it is the custom to consider simultaneously parameters of skewness and kurtosis for distributions.

Hence although it is possible to discuss shape properties of distributions entirely in terms of skewness and anti-skewness, (a) and (b) above show that parameters and usage of the terms skewness and kurtosis could not be fully interpreted. Defining kurtosis through spread functions provides the best formal basis to what has become accepted usage statistically, and we therefore adopt this definition. Kurtosis thus defined relates to the behaviour of a spread function in which the possible differences of behaviour between the left and right half-spreads (a term due to Hoaglin, 1985) are averaged out.

3. Shape

Although kurtosis for asymmetric distributions is best defined through spread functions the consequence is that skewness and kurtosis are then not distinct properties but are highly interrelated aspects of shape.

3.1 Kurtosis for Skew-Comparable Distributions

Theorem 2 examines how anti-skewness affects the kurtosis behaviour of the (symmetrised) distributions. To consider the effect of skewness properties, we consider the symmetrisation of skew-comparable distributions.

Suppose $F \leq_2^m G$, and consider the distribution F_α defined by $F_\alpha^{-1}(u) \equiv (\frac{1}{2}+\alpha)F^{-1}(u)-(\frac{1}{2}-\alpha)F^{-1}(1-u)$; and similarly for G_α. $F_{\frac{1}{2}} \equiv F$ and as α decreases from $\frac{1}{2}$ to 0, F_α becomes less asymmetric until the symmetric distribution F_0 is obtained at $\alpha = 0$. To compare the skewness or anti-skewness of F_α and G_α at the $\leq^m$ level we consider the expression

$$(\tfrac{1}{2}+\alpha)[h_G(u)-h_F(u)]-(\tfrac{1}{2}-\alpha)[h_G(1-u)-h_F(1-u)]\,, \quad u \in (\tfrac{1}{2},1) \quad (3.1)$$

where $h_F(u) = (F^{-1}(u) - m_F)f(m_F)$.

For $F \leq_2^m G$, as α decreases from $\frac{1}{2}$ to 0, (3.1) decreases for each u from some value > 0. If it is still ≥ 0 for all $u > \frac{1}{2}$ at $\alpha = 0$, then $F \leq_0^0 G$; the skewness difference between F and G at all the right-hand quantiles must dominate the skewness difference at all the corresponding left-hand quantiles. If (3.1) is ≤ 0 for all u at $\alpha = 0$, then $G \leq_0^0 F$. In either of these cases, the comparison of F and G changes from a skewness to a kurtosis comparison as α changes, with the direction of the kurtosis comparison depending on which side of the skewness comparison dominates for all u. If the domination is not for all u, there will not be a kurtosis comparison for the whole range of u.

Similarly, at the D-level, if one side of the skewness comparison dominates the other side, the skewness comparison changes to a kurtosis comparison.

Although it is possible to give similar expressions for comparisons of F_α and G_α at the star-shaped level and van Zwet's level, the contributions of the skewness properties of F and G are not as simple as in (3.1). But again for $F \underset{star}{\leq_2^m} G$ or $F \leq_2 G$, a kurtosis comparison of F and G arises when one side of the skewness comparison dominates the other.

Figures 1 and 3 give examples of distributions for which a skewness comparison becomes, on symmetrisation, a kurtosis comparison. In Figure 1, $F_{1,1}$ and $F_{.5,1}$ have the same kurtosis; in Figure 3, on symmetrisation,

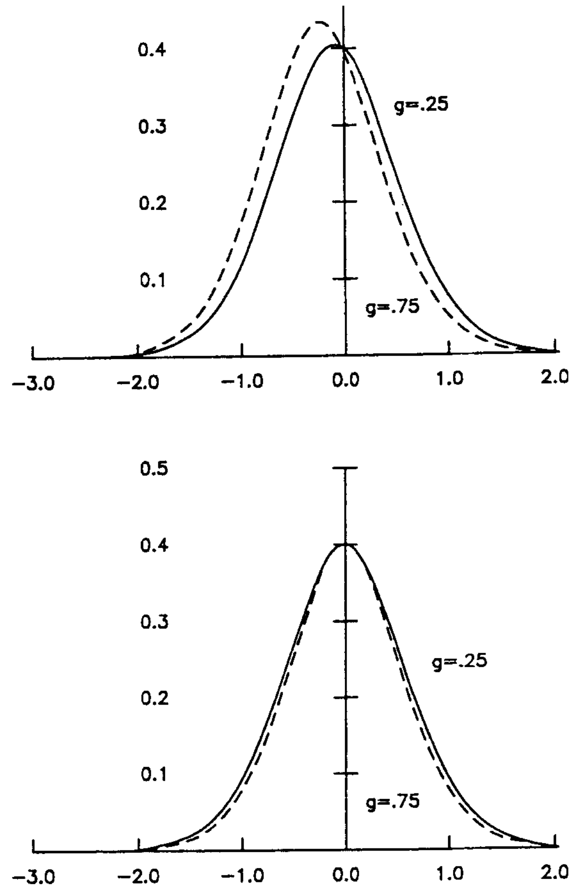

Fig. 4.—Densities and symmetrised densities of Tukey g and h distributions $T_{g,h}$; $T_{.25,.75} <_2 T_{.75,.75}$ and $T_{.25,.75} \underset{star}{<^0} T_{.75,.75}$.

the central skewness comparison becomes a kurtosis comparison with the ordering of the distributions reversed, and joins with the tail anti-skewness to give $F_{.5,1.5} <_0^0 F_{.5,8}$. Figure 4 shows the densities of $T_{.25,.75}$ and $T_{.75,.75}$ and their symmetrised versions, standardised as in the previous figures, illustrating the result that $T_{.25,.75} <_2 T_{.75,.75}$ and $T_{.25,.75} \underset{star}{<^0} T_{.775,.75}$.

As in Section 2, the moment-based properties taking the mean as location reference are not as amenable to consideration. The third and fourth order moments of the positive and negative parts of the standardised distributions are of assistance in considering the effect of symmetrisation on $F <_2^* G$, because the comparability of F^0 and G^0 depends on the relative weight of the left and right hand tails of F and G. But although skewness and kurtosis comparisons with respect to the mean of asymmetric distributions are interrelated, the relationships are not as directly interpretable as in orderings and measures with respect to the median.

3.2 Shape Contributions

Theorem 2 and the above section illustrate that the roles of skewness and anti-skewness relationships are not eliminated by symmetrisation; symmetrisation focusses on an averaged aspect of the shape relationship. An important point that requires examination is the somewhat unequal roles of skewness and kurtosis in describing shape, in that skewness (or anti-skewness) equivalence is shape equivalence, whereas kurtosis equivalence is not. However skewness and kurtosis tend to be regarded as being different components of shape, so leaving kurtosis defined through spread functions, we ask if we can define skewness, consistent with statistical usage, so that skewness and kurtosis are both needed to determine shape. With reference to Theorem 2(iv), although it has not been identified or previously used as a measure of skewness, $\varphi_F(u) \equiv \log\{f[F^{-1}(1-u)]/f[F^{-1}(u)]\}$, $u \geq \frac{1}{2}$, preserves $\underset{D}{\leq_2^m}$ and changes only by sign (a) under a linear transformation $aX+b$. It is of interest to note that the limit of $\varphi_F(u)$ as $u \to 1$, is proportional to Parzen's (1979) (right tail index — left tail index). Thus Theorem 2 shows that, at each level of strength of the properties, anti-skewness and kurtosis are equivalent for distributions having common values of a certain skewness function that is a skewness measure for skew-comparable distributions. From this we obtain Theorem 3 below, giving conditions for shape equivalence.

Theorem 3. *F and G have the same shape if $S_G(u)/S_F(u)$ is a constant and either (i) $\gamma_F(u) = \gamma_G(u)$ or (ii) $\nu_F(u) = \nu_G(u)$ or (iii) $\varphi_F(u) = \varphi_G(u), u \epsilon (\frac{1}{2}, 1)$.*

Proof. The proof follows from Theorem 2 and the results that antiskewness equivalence $\Leftrightarrow$ shape equivalence, but can also be obtained directly using for example, (2.15) in (1).

Therefore if skewness were defined using one of the functions $\gamma(u)$, $\nu(u)$ or $\varphi(u)$, instead of in terms of half-spreads, shape equivalence would require both skewness and kurtosis equivalence. However it can be shown that the remaining two functions and other measures would not preserve the ordering. Also the different levels of strength of the property would not be available as a hierarchy of transitive orderings except under restrictions on the individual skewness properties of the distributions being compared; essentially the function being used for the definition would need to be monotonic.

Thus if it desired to consider skewness and kurtosis as jointly determining shape, skewness can be defined only weakly by inequalities on say, $\gamma(u)$, and measured only by this function, while kurtosis is defined through spread functions. This may be sufficient and appropriate for some applications. Van Zwet's orderings arose from considering transformations of random variables that increase skewness, or, for symmetric distributions only, that increase kurtosis. The skewness and anti-skewness orderings continue the transformation approach and this may be more appropriate in some theoretical considerations. Different uses of the theory may be more appropriate to different areas of application.

There is not a theorem corresponding to Theorem 3 for the mean as location reference and the spread functions φ_F and φ_G. Even if $\varphi_F = \varphi_G$ and $\mu_{F,2k+1} = \mu_{G,2k+1}$ for all $k = 1, 2, \ldots$, it does not follow that $F(\mu_F + x) = G(\mu_G + x)$. This again illustrates the limitations of moment measures of shape in the general context.

4. Conclusion

If kurtosis is not defined for asymmetric distributions, then kurtosis parameters cannot be justifiably used in the asymmetric case. In considering asymmetry, it is easy to see qualitatively that skewness and kurtosis are linked. This paper has examined as quantitatively as possible the roles of parameters and orderings that have been used to describe skewness and kurtosis, to provide insight into how they interrelate in representing shape properties. Such insight provides a fuller background for both applications and theoretical considerations.

Acknowledgement

We thank the referee for helpful comments.

References

ANDREWS, D.F., BICKEL, P.J., HAMPEL, F.R., HUBER, P.J., ROGERS, W.H. & TUKEY, J.W. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton, N.J.: Princeton University Press.

BALANDA, K.P. (1986). *Kurtosis, Peakedness and Tailweight.* Unpublished M.Sc. thesis, University of Queensland.

BALANDA, K.P. (1987). Kurtosis comparison of the Cauchy and double exponential distributions. *Comm. Statist. A—Theory Methods* **16**(2), 579–592.

BALANDA, K.P. (1987). Kurtosis and spread. Submitted for publication.

BICKEL, P.J. & LEHMANN, E.L. (1976). Descriptive statistics for non-parametric models. III. Dispersion. *Ann. Statist.* **4**, 1139–1158.

BOWLEY, A.L. (1901). *Elements of Statistics.* London: Stapler Press Ltd.

CROW, E.L. & SIDDIQUI, M.M. (1967). Robust estimation of location. *J. Amer. Statist. Assoc.* **52**, 353–389.

DOKSUM, K.A. (1969). Starshaped transformations and the power of rank tests. *Ann. Math. Statist.* **40**, 1167–1176.

DOKSUM, K.A. (1975). Measures of location and asymmetry. *Scand. J. Statist.* **2**, 11–22.

DYSON, F.J. (1943). A note on kurtosis. *J. Roy. Statist. Soc. Ser. B* **106**, 360–361.

GOENEVELD, R.A. & MEEDEN, G. (1984). Measuring skewness and kurtosis. *The Statistician* **33**, 391–399.

HOAGLIN, D.C. (1985). Summarizing shape numerically: the g-and-h distributions. In *Data Analysis for Tables, Trends and Shapes: Robust and Exploratory Techniques*, eds. D.C. Hoaglin, F. Mosteller and J.W. Tukey. New York: Wiley.

HORN, P. (1983). A measure for peakedness. *Amer. Statist.* **37**, 55–56.

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

KARLIN, S. (1968). *Total Positivity.* Vol.1. Stanford: Stanford University Press.

LAWRENCE, M.J. (1974). Inequalities of s-ordered distributions. *Ann. Statist.* **3**, 413–428.

LOH, W.J. (1984). Bounds on AREs for restricted classes of distributions defined via tail orderings. *Ann. Statist.* **12**, 685–701.

MACGILLIVRAY, H.L. (1981). The mean, median, mode inequality and skewness for a class of densities. *Austral. J. Statist.* **23**, 247–250.

MACGILLIVRAY, H.L. (1982). Skewness properties of asymmetric forms of Tukey lambda distributions. *Comm. Statist. A—Theory Methods* **11**, 2239–2248.

MACGILLIVRAY, H.L. (1985). A crossing theorem for distributions and their moments. *Bull. Austral. Math. Soc.* **31**, 413–419.

MACGILLIVRAY, H.L. (1986). Skewness and asymmetry: measures and orderings. *Ann. Statist.* **14**, 994–1011.

OJA, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.* **8**, 154–168.

PARZEN, E. (1979). Nonparametric statistical data modelling (with discussion). *J. Amer. Statist. Assoc.* **74**, 105–121.

PEARSON, K. (1895). Contributions to the mathematical theory of evolution II. Skew variation in homogeneous material. *Philos. Trans.* **186**, 343.

RUNNENBERG, J. Th. (1978). Mean, median, mode. *Statist. Neerlandica* **32**, 73–79.

RUPPERT, D. (1987). What is kurtosis? An influence function approach. *Amer. Statist.* **41**, 1–6.

VAN ZWET, W.R. (1964). *Convex transformations of random variables.* Amsterdam: Math. Centrum.

VAN ZWET, W.R. (1979). Mean, median, mode II. *Statist. Neerlandica* **33**, 1–5.

Received June 1987; revised March 1988