Kurtosis: A Critical Review

KEVIN P. BALANDA and H. L. MACGILLIVRAY*

We critically review the development of the concept of kurtosis. We conclude that it is best to define kurtosis vaguely as the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails and to recognize that it can be formalized in many ways. These formalizations are best expressed in terms of location- and scale-free partial orderings on distributions and the measures that preserve them. The role of scale-matching techniques and placement of shoulders in the formalizations that have appeared in the literature are emphasized.

KEY WORDS: Measure; Ordering; Peakedness; Shape; Tail weight.

1. INTRODUCTION

The terms kurtosis, peakedness, and tail weight are often used in descriptive statistics and statistical inference. There has been a variety of uses and interpretations of these terms, however; and in this article we discuss the development of these concepts as components of distributional shape.

In Section 2 we describe the various attempts that have appeared in the literature to interpret the value of the standardized fourth central moment β_2. It is apparent that although moments play an important role in statistical inference they are very poor indicators of distributional shape. Kurtosis, peakedness, and tail weight are nevertheless important distributional concepts and several authors have proposed alternative measures, some of which are discussed in Section 3.

These studies have identified a shape characteristic that we call kurtosis and can be vaguely defined as the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails. Like location, scale, and skewness, kurtosis should be viewed as a "vague concept" (Mosteller and Tukey 1977) that can be formalized in many ways. We argue that the various interpretations of β_2 and alternative measures can be obtained from this definition by taking a specific scale-matching technique and a particular placement of shoulders. In Section 4 we outline an approach to descriptive problems based on partial orderings on distributions and measures that preserve them. Only a few orderings and measures have appeared in the literature, defined only on symmetric distributions, and these are discussed in Section 5. The ordering-based approach has been more successful, but some areas require further attention. In Section 6 we outline some ongoing work.

*Kevin P. Balanda is Lecturer and H. L. MacGillivray is Senior Lecturer, Department of Mathematics, University of Queensland, St. Lucia, Queensland 4067, Australia. The authors thank the referees for helpful comments.

2. INTERPRETATION OF THE STANDARDIZED FOURTH CENTRAL MOMENT

Kurtosis is traditionally defined operationally. The oldest and most commonly used definition is that the kurtosis of a distribution H is that characteristic measured by its standardized fourth central moment $\beta_2(H)$ (provided it exists) defined by $\beta_2(H) = \mu_4(H)/(\mu_2(H))^2$. The normal distribution, with a value of β_2 equal to 3, is often used as a standard, and the quantity $\gamma_2(H)$ defined by $\gamma_2(H) = \beta_2(H) - 3$ is sometimes called the *kurtosis* (or *coefficient of excess*) of the distribution H. The terms *platykurtic*, *leptokurtic*, and *mesokurtic* appear to have been first used by Pearson (1905). Dyson (1943) gave two amusing mnemonics attributed to Student for these names: platykurtic curves, like platypuses, are squat with short tails whereas leptokurtic curves are high with long tails, like kangaroos—noted for "lepping"! The terms supposedly refer to the general shape of a distribution, with platykurtic distributions ($\beta_2 < 3$) being flat-topped compared with the normal, leptokurtic distributions ($\beta_2 > 3$) being more sharply peaked than the normal and mesokurtic distributions ($\beta_2 = 3$) having shape comparable to that of the normal.

Because of the "averaging" nature of moments, however, the relationship of β_2 to shape is far from clear; in this section we discuss some of the attempts that have appeared in the literature to describe the distributional shapes corresponding to large values of β_2. These investigations concentrate on symmetric distributions and fall into two major areas:

1. Many form part of a more general inquiry into the relationship between moment crossings and density crossings. Typically, conditions on the crossings of two standardized (usually symmetric) densities f and g that ensure $\mu_4(f) \leq \mu_4(g)$ are described. Earlier results deal only with kurtosis, whereas more recent works discuss more general crossings and give results about kurtosis as corollaries (see Sec. 2.1).
2. In the second type of investigation, a (usually discrete) distribution is modified in some way and the effect on the value of β_2 is noted. Although some of these studies are not very sophisticated, they have provided some interesting interpretations of β_2 such as the one in terms of bimodality discussed in Section 2.2

2.1 Moment Crossings and Density Crossings

Dyson (1943) proved the following result. If f and g are standardized to have mean 0 and equal variances, and there exist constants a_1, a_2, a_3, and a_4 with $a_1 < a_2 < a_3 < a_4$ such that

$$\text{(a)}\quad \left.\begin{array}{c} -\infty < x < a_1 \\ a_2 < x < a_3 \\ a_4 < x < \infty \end{array}\right\} \Rightarrow \quad f(x) \leq g(x),$$

$$\text{(b)} \quad \left.\begin{array}{l} a_1 < x < a_2 \\ a_3 < x < a_4 \end{array}\right\} \Rightarrow f(x) \geq g(x),$$

and (c) $[a_1 + a_2 + a_3 + a_4]$ and $[\mu_3(f) - \mu_3(g)]$ are not both strictly positive or both strictly negative, then $\mu_4(f) \leq \mu_4(g)$. An example of two standardized symmetric densities f and g is given, showing the condition

$$f(x) < g(x) \text{ for } |x| \text{ small and } |x| \text{ large}$$

is not incompatible with $\mu_4(g) < \mu_4(f)$. Dyson's result does not assume symmetry and is one of the few to suggest a relationship between the skewness and kurtosis of a distribution; this relationship receives little attention because of the common practice of restricting the discussion of kurtosis to symmetric distributions only. Balanda (1986), Balanda and MacGillivray (1987), and MacGillivray and Balanda (1987) each considered kurtosis in asymmetric distributions; the later paper discussed this relationship in detail.

An error commonly associated with kurtosis is that the sign of γ_2 compares the value of the density at the center with that of the corresponding normal density. Kaplansky (1945) gave four examples of standardized (mean 0 and variance 1) symmetric distributions that, when compared with the standard normal, show there is no logical connection between the value of the density of the standardized distribution at the center and the sign of γ_2.

Finucan (1964) "rediscovers the original interpretation of kurtosis as an indicator of a prominent peak and tail on the density curve" (p. 111), claiming that the incorrectly simplified version of this interpretation as peakedness led to the types of errors discussed by Kaplansky (1945). Finucan claimed that the quantity β_2 measures what is best described as peakedness combined with tailedness or lack of shoulders, and proved that if f and g are symmetric with mean 0 and common variance and the graph of $[g(x) - f(x)]$ goes through a peak–trough–peak pattern as $|x|$ increases, then $\beta_2(f) \leq \beta_2(g)$. This result was mentioned without proof by Fisher (1925) and is essentially Dyson's result in the symmetric case. Figure 1 contains two standardized symmetric densities satisfying the Dyson–Finucan condition. Finucan suggested that this pattern be taken as the common explanation of high kurtosis and hoped that some further explanation may be found for the exceptions.

The Dyson–Finucan condition involves crossings of standardized densities. Marsagalia, Marshall, and Proschan (1965) gave further results concerning the relationship between the number of crossings of the absolute moments of two standardized symmetric distributions and the number of crossings of their densities. In particular, they proved that if two such densities f and g satisfy the Dyson–Finucan condition then (provided the absolute moments are finite) (a) $v_s(f) > v_s(g)$ if $0 < s < 2$ and (b) $v_s(f) < v_s(g)$ if $s < 0$ or $s > 2$, where $v_s(h)$ is the sth absolute moment of h. The result is an immediate consequence of the variation-diminishing properties of totally positive functions (discussed by Karlin 1968), and more general comparisons can be obtained using the ideas of positivity (MacGillivray 1985).

Ali (1974), using generalizations of the stochastic ordering, proved that if two standardized symmetric random variables X, Y have the property that $|Y|$ is fourth-degree stochastically larger than $|X|$, then $\beta_2(X) \leq \beta_2(Y)$. If X, Y satisfy the Dyson–Finucan condition, then $|Y|$ is third-degree stochastically larger than $|X|$, a slightly stronger condition. Other results are given, as well as the following example, which demonstrates how γ_2 can be a misleading measure of nonnormality. For $k = 2, 3, \ldots$, let F_k denote the mixture

$$F_k(x) = [1 - 1/(k^2 - 1)]\Phi(x) + [1/(k^2 - 1)]\Phi(x/k),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

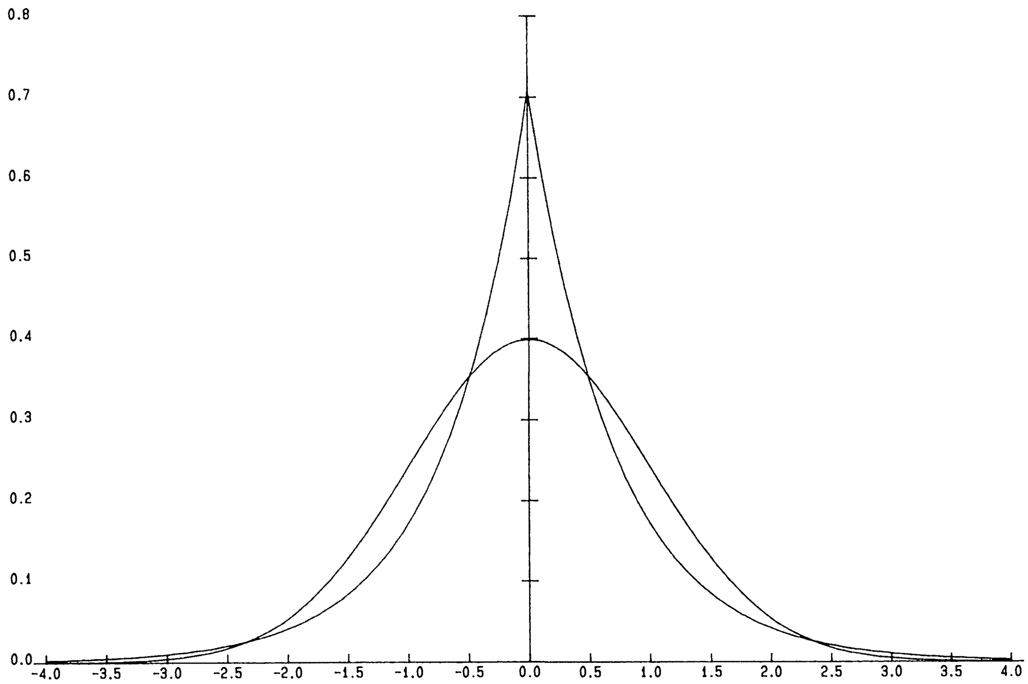

Figure 1. Standardized Symmetric Densities f, g Satisfying the Dyson–Finucan Condition. The standardized densities of the double-exponential and Normal distributions are plotted.

The sequence converges in distribution (uniformly in x) to the standard normal distribution as $k \to \infty$, and $\gamma_2(F_k) = 3(k^2 - 2)/4 \to \infty$ as $k \to \infty$. Thus F_k is uniformly approximated with increasing accuracy by the standard normal distribution, and $\gamma_2(F_k)$ grows without limit. Under suitable regularity conditions, this phenomenon cannot occur for quantile-based measures of kurtosis that have been used by a number of research workers (see Sec. 3). Ali observed that large γ_2 can arise from tailedness without peakedness about the mean and noted that a number of the exceptional cases given by Dyson (1943) and Kaplansky (1945) fall into this category. On the basis of these observations, Ali concluded (erroneously) that β_2 "measures only the tailedness of a symmetric distribution" (p. 543). If distributions cross more than the required minimum number of times, the value of β_2 cannot be predicted without more information. It is the failure to recognize this that causes most of the mistakes and problems in interpreting β_2.

2.2 Consideration of a Single Distribution

Chissom (1970) adopted an approach different from those of the aforementioned authors. By progressively modifying the shape of a single (discrete) distribution, each time noting the effect on β_2, Chissom attempted to describe those shape characteristics that affect the value of β_2. Although Chissom agreed with Ali that the tails of a distribution can drastically affect the kurtosis value, he reminded us that it also depends on the peak and that the tendency toward bimodality may also be important.

Darlington (1970) noted that $\beta_2(X) - 1 = \text{var}(Z_x^2)$, where $Z_x = (X - \mu_x)/\sigma_x$, and argued that $\gamma_2(X)$ measures the clustering of the Z_x values about ± 1 and hence is best described as a measure of unimodality versus bimodality, with a small value of $\beta_2(X)$ suggesting that X displays a strong tendency toward bimodality ("bimodality" here is taken to be clustering about $\mu - \sigma$ and $\mu + \sigma$). In a vague sense this interpretation in terms of tendency toward bimodality is consistent with Finucan's (1964) interpretation. If a distribution displays a tendency toward bimodality, then it can be thought of as having "strong shoulders" and thus, in Finucan's sense, low kurtosis.

The problem with this interpretation lies, of course, in the use of vague, undefined terms such as "tendency toward bimodality" and "lack of shoulders"; Hildebrand (1971) gave two examples highlighting this difficulty. Hildebrand first considered symmetric beta distributions with densities

$$f(x; a) = [\Gamma(2\alpha)/\Gamma^2(\alpha)]\, x^{\alpha-1}(1 - x^{\alpha-1}), \qquad 0 < x < 1,$$

where $\alpha > 0$. Here $\gamma_2(\alpha) = -6/(2\alpha + 3)$. If $\alpha < 1$ then the distribution is bimodal and $\gamma_2(\alpha) < -1.2$. As $\alpha \to 0$, $\gamma_2(\alpha) \to -2$ and the distribution approaches the two-point binomial. If $\alpha = 1$ then $\gamma_2(1) = -1.2$ and the distribution is uniform (nonmodal), whereas when $\alpha \to \infty$ the distribution approaches normality and $\gamma_2(\alpha) \to 0$. This family, then, is consistent with Darlington's interpretation. On the other hand, the family of double-gamma distributions with densities

$$f(x; \alpha, \beta) = [\beta^\alpha/2\Gamma(\beta)]\, |x|^{\alpha-1} \exp(-\beta|x|)$$

for all x, where α and β are both positive, is inconsistent with Darlington's interpretation. The values of γ_2 are given by $\gamma_2(\alpha, \beta) = [(\alpha+3)(\alpha+2)/(\alpha+1)\alpha] - 3$, a decreasing function of α. If $\alpha < 1$ the distribution is unimodal and $\gamma_2(\alpha, \beta) > 3$. If $\alpha = 1$ then f is the double exponential density and $\gamma_2 = 3$, whereas if $\alpha > 1$ the distribution is bimodal and $\gamma_2(\alpha, \beta)$ ranges from 3 to the limiting value -2 (being 0 at $\alpha = [1 + 13^{1/2}]/2$). This family, then, contains bimodal distributions with values of γ_2 ranging from -2 to 3.

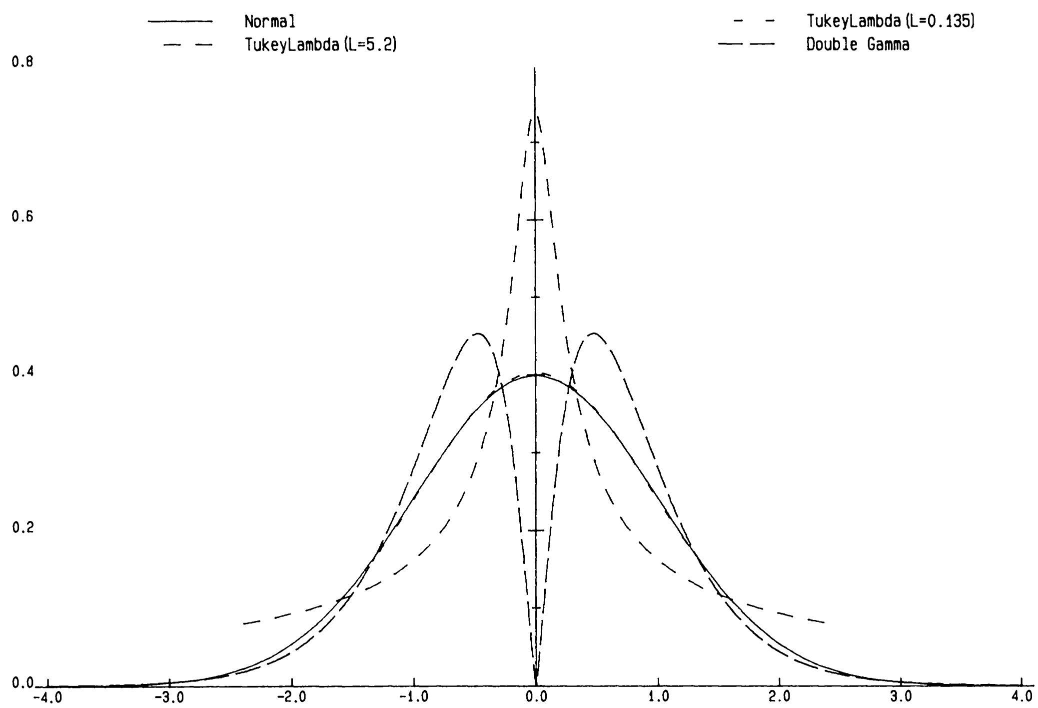

Figure 2. Standardized Symmetric Densities With $\gamma_2 = 0$: Standard Normal Distribution; Symmetric Tukey Lambda Distribution With $\lambda = .135$; Symmetric Tukey Lambda Distribution With $\lambda = 5.2$; Double Gamma Distribution With $\alpha = (1 + 13^{1/2})/2$.

Moors (1986) noted that bimodal distributions can have large kurtosis and argued that the value of β_2 measures the dispersion around the values $\mu - \sigma$ and $\mu + \sigma$. Because this can arise either from a concentration of probability mass around μ or in the tails of the distribution, Moors argued that Darlington's description of this in terms of bimodality is incorrect.

Ruppert (1987) used the influence function to investigate the effect of a small amount of two-point contamination on the value of β_2 and other kurtosis measures; he also pointed out that this is an extension of Darlington's approach.

The value of β_2 is affected by so many different aspects of a distribution that Kendall and Stuart (1977) concluded the words *leptokurtic* and *platykurtic* are best regarded as describing the sign of γ_2 rather than the shape of the density. It seems that because of the averaging process involved in its definition, a given value of β_2 can correspond to several different distributional shapes. Figure 2 contains a number of standardized symmetric densities with $\beta_2 = 3$. Although Curve 3 has finite support (and thus short tails) it is a good approximation to the Normal distribution. Curve 4 is bimodal whereas curve 2, although it has infinite support and is unimodal, is considerably more peaked than the standard normal distribution.

3. ALTERNATIVE MEASURES OF KURTOSIS, PEAKEDNESS, AND TAIL WEIGHT

Although β_2 is a poor measure of the kurtosis, peakedness, or tail weight of a distribution, these concepts nevertheless play an important role in both descriptive and inferential statistics. This has led some authors to propose alternative measures. Most are quantile-based and together form a haphazardly constructed collection of alternatives rather than a coherent alternative approach to the standardized fourth central moment. They do, however, recognize a number of the different formalizations of the concepts involved.

One large class of alternative measures is based on the idea that if X is a symmetric random variable with median m_x, then the skewness properties of the positive random variable $|X - m_x|$ represent the kurtosis properties of X. If $m(\cdot)$ is a measure of skewness, then $m(|X - m_x|)$ is used as a measure of kurtosis for X and alternative kurtosis measures can thus be generated from existing skewness measures. Using this idea, Groeneveld and Meeden (1984) proposed a number of alternative measures of kurtosis that have natural interpretations for symmetric distributions in terms of the movement of probability mass from the shoulders of a distribution into its center or tails. They suggested that, for each α in (0, ¼), the quantity $\beta_2(\alpha, H)$, defined by

$$\beta_2(\alpha, H) = \frac{H^{-1}(.75+\alpha) + H^{-1}(.75-\alpha) - 2H^{-1}(.75)}{H^{-1}(.75+\alpha) - H^{-1}(.75-\alpha)},$$

measures the kurtosis of the symmetric distribution H. The quantity $\beta_2(\alpha, H)$ is the value of a measure of skewness (MacGillivray 1986) applied to $|X - m_x|$, where X has distribution H. Referring to Figure 3, if $\beta_2(\alpha, H)$ is large then, relative to the quartiles, there has been a shift of mass into the center or tails of H. These measures lie in the interval $(-1, 1)$, U-shaped distributions have negative kurtosis, and the uniform distributions have zero kurtosis. Groeneveld and Meeden (1984) proposed other measures, and we refer the reader to their article for further details. Groeneveld and Meeden have been, to our knowledge, the only authors to propose alternative measures for kurtosis that cannot be considered in terms of just peakedness or just tail weight. Their measures involve both peakedness and tail weight as components of kurtosis, whereas the measures we discuss next deal separately with peakedness or tail weight. As we argue later, a better understanding of distributional shape through partial orderings on distributions involves the si-

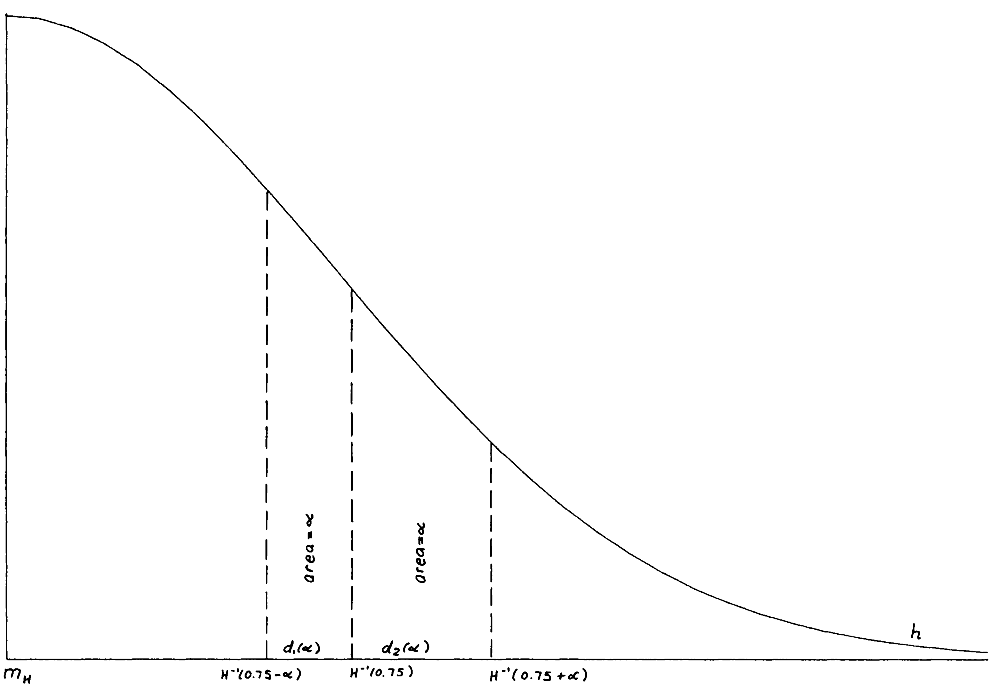

Figure 3. Groeneveld and Meeden's (1984) Measures of Kurtosis. $\beta_2(\alpha, H)$ is the (scaled) difference $[d_2(\alpha) - d_1(\alpha)]/[d_1(\alpha) + d_2(\alpha)]$.

multaneous consideration of these two concepts.

Horn (1983) suggested that, for $0 < p < 1/4$, the quantity

$$mt_P(h) = 1 - p/[h(m_H)(H^{-1}(.5+p) - m_H)]$$

be used as a measure of peakedness for a symmetric unimodal density h. Rogers and Tukey (1972) used $mt_P(h)$ for $p > 1/4$ as measures of tail weight. Rosenberger and Gasko (1983), however, rejected these as measures of tail weight, arguing that they were too sensitive to the central part of a distribution to be used for tail comparisons. Referring to Figure 4, if $mt_P(H)$ is large for p close to 0, then h looks like a spike at the center. These measures essentially refer to the slope of the density near the median. If, for example, the graph of h exhibits a plateau around the median (albeit very high), then $mt_P(H) = 0$ for p close to 0. Each measure takes values between 0 and 1 for symmetric unimodal distributions, 0 if the density is constant around the center. Using these measures, Horn ranked the Normal, t_6, Cauchy, and double-exponential distributions in order of increasing peakedness and suggested that the peakedness of the central t distributions decreases as the degrees of freedom increase. Most would agree with these conclusions.

Another family of quantile-based measures that has appeared in the literature is the following. For a distribution H and for $0 \leq p < 1/2$, let

$$t_P(H) = \frac{H^{-1}(.5+p) - H^{-1}(.5-p)}{H^{-1}(.75) - H^{-1}(.25)} .$$

Sometimes $t_P(H)$ is standardized to be 1 for the normal distribution. The standardized version $st_P(H)$ is related to the pth pseudovariance $\mathrm{PV}_P(H)$ by

$$st_P(H) = \mathrm{PV}_P(H)^{1/2}/\mathrm{PV}_{.25}(H)^{1/2} .$$

Pseudovariances $\mathrm{PV}_P(\cdot)$ are robust estimators of scale and were discussed by Andrews et al. (1972) for example. Extensions of these were discussed by Balanda (1986) and Ruppert (1987). Particular values of $t_P(H)$ have been used in a variety of contexts:

1. Crow and Siddiqui (1967) used $t_{.45}(H)$ to rank, in order of increasing tail thickness, the (symmetric) distributions included in a comparative study of location estimators. Their measure suggested that, in order of increasing tail thickness, the distributions could be ranked as uniform, parabolic, triangular, Normal, double exponential, and Cauchy. Where appropriate, this coincides with the ranking suggested by the value of β_2. Note that, although the double-exponential distribution is more peaked than the Cauchy distribution (in Horn's sense), the Cauchy has heavier tails [in terms of $t_{.45}(H)$].

2. Rosenberger and Gasko (1983) used $st_{.49}(H)$ as an index of tail weight of a symmetric distribution H, arguing that $st_P(H)$ (for $1/4 \leq p < 1/2$) measures how the extreme portion of a distribution spreads out relative to the width of the center (this measure being standardized to be 1 for a normal distribution). Rosenberger and Gasko also used $st_{.49}(H)$ to order the distributions included in a comparative study of location estimators and, where appropriate, their ranking agrees with that of Crow and Siddiqui (1967).

3. Heavy tail weight is often the most important aspect of nonnormality, and Andrews et al. (1972) used $t_{.49}(H)$ as an index of nonnormality to assess the distribution of the estimators included in the Princeton Robustness Study.

4. Parzen (1979) proposed that sample versions of $\log(t_P(H))$ be compared with the values of $\log(t_P(\Phi))$ (where Φ is the standard normal distribution function) in diagnostic tests for nonnormal tails in H.

5. Hogg (1974) proposed adaptive location estimators that used statistics like

$$Q = [U(.2) - L(.2)]/[U(.5) - L(.5)]$$

as the selector [where $U(d)$ and $L(d)$ denote the average of the largest and smallest $100d\%$ of the sample]. Such sta-

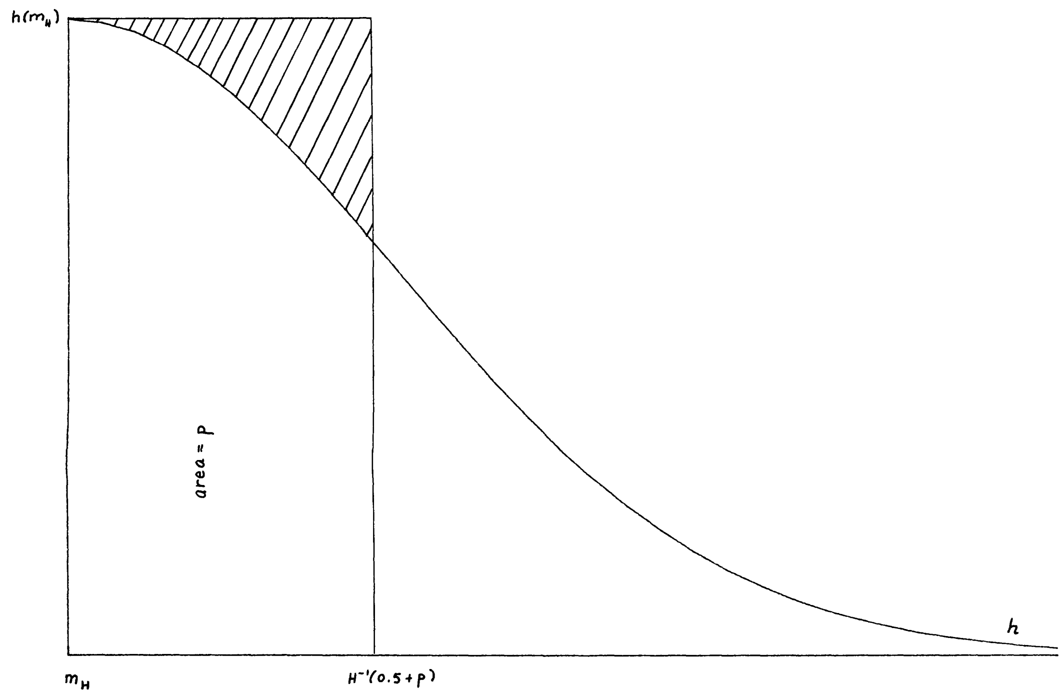

Figure 4. Horn's Peakedness Measure. $mt_p(H)$ is the ratio of the area of the shaded region to the area of the enclosing rectangle.

tistics are sample analogs of the measures

$$t_{\tau,\delta}(H) = \frac{U_\tau(H) - L_\tau(H)}{U_\delta(H) - L_\delta(H)}, \qquad 0 < \tau < \delta \le 1/2,$$

where $U_\varepsilon(H) = \mathrm{E}[X|X > H^{-1}(1-\varepsilon)]$, $L_\varepsilon(H) = \mathrm{E}[X|X < H^{-1}(\varepsilon)]$, and $X \sim H$. Originally the sample kurtosis was used, but Hogg later found that measures like Q were more appropriate for detecting heavy tails.

Finally, we discuss a somewhat different approach. Parzen (1979) defined a density-quantile tail exponent based on the limiting behavior of the density-quantile function $h(H^{-1}(p))$ as p approached 0 or 1. For a tail-monotone density h (Parzen 1979), the limit

$$a(H) = -\lim_{p\to 1} \frac{(1-p)h'(H^{-1}(p))}{(h(H^{-1}(p)))^2}$$

is finite and is called the tail exponent of H. If p is close to 1, then $-[\log(h(H^{-1}(p)))]' \approx a(H)/(1-p)$ and $h(H^{-1}(p)) \approx (1-p)^{a(H)}$. Thus $a(H)$ describes the tail behavior of H and $a(F) \le a(G)$ iff $G^{-1}(F(x))$ is convex for all sufficiently large x. Parzen classified distributions according to the value of a: $a < 1$—short tails (or limited type); $a = 1$—medium tails (or exponential type); $a > 1$—long tails (or Cauchy type). The names in parentheses are used in classical extreme-value theory (Gumbel 1962) to classify the types of distributions leading to the three possible limiting distributions of $X_{(n)}$, the largest order statistic in a random sample of size n. Parzen (1980) proved the classification based on the value of $a(H)$ coincides with the extreme-value classification and advocated tail exponents as easily applied criteria for determining the extreme-value distribution corresponding to a specified distribution. Examples of Parzen's classification include short tailed [uniform, $f(x) = c(1-x)^{c-1}$ for $0 < x < 1$]; medium tailed [exponential, extreme value, logistic, Normal, Weibull]; and long tailed [Cauchy, Pareto, slash (1/uniform), symmetric Tukey lambda with $\lambda < 1$]. Although the exponent above is a right-tail index, a left-tail index can be similarly defined in terms of the limiting behavior of $h(H^{-1}(p))$ as p approaches 0 from above. Although the left and right indexes agree in the symmetric case, this approach has the advantage that the left and right tails of an asymmetric distribution can be considered separately.

Schuster (1984) refined Parzen's classification using the limiting value of the reciprocal of the hazard function and gave an interpretation in terms of the limiting size of the extreme spacings in a random sample from the distribution.

Bickel and Lehmann (1975) suggested that measures of kurtosis, peakedness, and tail weight should be ratios of scale measures in the sense that both numerator and denominator should preserve their spread ordering. Except for Parzen's tail exponent, all of the preceding measures are of this form.

4. KURTOSIS AS A VAGUE CONCEPT AND THE ORDERING-BASED APPROACH

The aforementioned works have identified a general shape characteristic that can be called kurtosis. All are consistent with the definition of kurtosis as the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails. In particular, this definition implies that peakedness and tail weight are best viewed as components of kurtosis, since any movement of mass from the shoulders into the tails must be accompanied by a movement of mass into the center if the scale is to be left unchanged. This definition is necessarily vague because the movement can be formalized in many ways. Specifically, the formalization depends on the scaling technique used to make it scale free and the position chosen for the shoulders. The measure of location used is not important in the symmetric case, as they all coincide with the center of symmetry.

The various measures discussed use different scaling techniques and positioning of shoulders. The scaling techniques used include (a) the standard deviation (in the definition of β_2), (b) *central density matching* using the inverse of the density at the median as the scale measure (in the definition of Horn's peakedness measure), (c) *interquartile matching* using the interquartile range [in the definitions of $t_P(H)$ and $st_P(H)$], and (d) matching techniques involving different distributional distances (in the definition of Hogg's selector statistic). The shoulders above were placed around the quartiles [in Groeneveld and Meeden's (1984) measure of kurtosis] and around $\mu - \sigma$ and $\mu + \sigma$ [Darlington's (1970) and Moors's (1986) interpretation of β_2], and they can be considered to coincide at the median in Horn's (1983) peakedness measure. In the latter case there is no movement of mass into the center, since peakedness corresponds to the density falling away from that center.

Different scaling techniques and positioning of the shoulders give rise to different formalizations of kurtosis, and its components' peakedness and tail weight. For example, Horn's peakedness corresponds to a spike at the center, whereas Groeneveld and Meeden's measures correspond to a persistence of mass around the center compared with the quartiles. These different formalizations have been used in practice, and it seems preferable to accept kurtosis as a vague concept with the definition already given and develop a coherent structure of such formalizations rather than to concentrate only on β_2.

We have only discussed alternative measures; however, the measure-based approach has been criticized recently. For example, van Zwet (1964) recorded two serious reservations about the use of β_2:

1. Many of the comparisons made are meaningless. Any two distributions with finite fourth moments, for example, can be compared using β_2, whereas one feels there are pairs of such distributions that are totally incomparable in this regard.

2. Very few applications of general interest have arisen.

These difficulties regarding β_2 apply to any other single-parameter representation and arise because a single value usually corresponds to many different distributional shapes.

Many of the measures discussed in Section 3 are families of measures indexed by a range of p values, and some authors suggest that a plot of these measures against p is required to fully describe the concept being discussed. This implicitly identifies an underlying ordering and leads to the ordering-based approach proposed by van Zwet. Rather than measure the kurtosis of a single distribution, we define partial orderings $<<$ in such a way that $F << G$ means, in some sense, that G has greater kurtosis than F. Some of the orderings called kurtosis orderings in the literature are not scale free and are in fact used as scale orderings in other contexts. For example, Birnbaum (1948) used a scale ordering to indicate peakedness, whereas Doksum (1969) used Bickel and Lehmann's (1975) spread ordering to indicate heavy tails. The value of β_2 is dimensionless, and we believe that any discussion of kurtosis should, therefore, be location and scale free, where location and scale free means invariant under linear transformations of the random variables involved. Orderings $<<$ are defined so that $F << G$ means, in some location- and scale-free sense, that G has greater mass in the center and tails than does F. Measures of kurtosis are then restricted to be location- and scale-free functionals of distributions that preserve one of these orderings. We believe that a kurtosis measure should not be used without first identifying the ordering underlying it and that a measure should not be used to make comparisons within a family of distributions unless that family is totally ordered by the underlying ordering. It is only in these circumstances that the measure genuinely summarizes a kurtosis property in a meaningful way. In this approach it is important to identify the weakest ordering underlying measures that have been proposed in the literature.

5. EXISTING ORDERINGS ON SYMMETRIC DISTRIBUTIONS

The various conditions (discussed in Sec. 2.1) on F and G that guarantee $\beta_2(F) \leq \beta_2(G)$ may be thought of as defining relations on certain classes of distributions. They do not, however, play a major role in the ordering-based approach, because the primary interest there lay only in the standardized fourth central moment and not in the definition of general orderings on distributions or the measures that may preserve them. All existing orderings are weakenings of van Zwet's (1964) ordering $\leq_S$.

5.1 Van Zwet's Ordering

Van Zwet (1964) introduced, for the class of symmetric distributions, an ordering $\leq_S$ defined by $F \leq_S G$ iff $R_{F,G}(x) = G^{-1}(F(x))$ is convex for $x > m_F$, where m_F is the point of symmetry of F. Since the distributions are assumed symmetric, $R_{F,G}(x)$ is convex for $x > m_F$ iff it is concave for $x < m_F$. $F \leq_S G$ holds iff a random variable X with distribution F can be mapped to a random variable Y with distribution G by an increasing concave–convex function about the median. Van Zwet argued that if this is the case then, in the transformation of F to G, there is a contraction of the middle and an extension of the ends of the F scale. Moreover, this deformation increases toward the middle and ends, so intuitively G has a greater concentration of mass around its median and in its tails than does F.

Van Zwet (1964) showed that U-shaped $\leq_S$ uniform $\leq_S$ Normal $\leq_S$ logistic $\leq_S$ double exponential and logistic $\leq_S$ Cauchy. Although the double-exponential and Cauchy distributions are not $\leq_S$ comparable, they can be compared (Balanda 1987) using the orderings alluded to in Section 6. Both the family of double-gamma distributions and the family of symmetric beta distributions are totally ordered by $\leq_S$, so kurtosis comparisons within these families can be based on the value of β_2, which preserves the ordering. Hildebrand's (1971) examples (see Sec. 2.2), however, show that β_2 is totally inadequate as a description of the shape of individual members. These difficulties arise because the ordering is preserved by the standardized even central moments and all of the measures discussed in Section 3, reflecting many different kurtosis formalizations.

Van Zwet (1964) gave several applications of this ordering. For example, the asymptotic relative efficiency of Wilcoxon's two-sample test to the normal scores test and the relative efficiency of the sample median to the sample mean are no smaller for G than they are for F if $F \leq_S G$. We refer the reader to van Zwet (1964) for further examples. Like most researchers, van Zwet considered kurtosis as a property only of symmetric distributions, even though measures of kurtosis are used for asymmetric distributions. Since its introduction more than 20 years ago, no extension of this ordering to the asymmetric case has been suggested. Balanda (1986), Balanda and MacGillivray (1987), and MacGillivray and Balanda (1987) proposed several such extensions.

Van Zwet's ordering is probably the strongest that needs to be considered in descriptive work, although several stronger orderings have appeared (Hettmansperger and Keenan 1975). The remaining existing kurtosis orderings are weaker than $\leq_S$ and belong to one of two branches leading from $\leq_S$. These branches correspond to the two characterizations of van Zwet's convexity condition on which weakenings can be based: (a) $R_{F,G}(x)$ is convex for $x > m_F$ iff, for each $x_0 > m_F$,

$$[R_{F,G}(x) - R_{F,G}(x_0)]/(x - x_0)$$

is increasing for $x > m_F$ (with the reverse for $x < m_F$ following from the symmetry of F and G). (b) $R_{F,G}(x)$ is either concave for $x > m_F$ or convex for $x > m_F$ iff, for all c and d, the graphs of $y = R_{F,G}(x)$ and $y = cx + d$ cross each other at most twice for $x > m_F$.

If characterization (a) is used, we obtain the orderings of Lawrence (1975) and Loh (1984). Characterization (b) leads to the orderings defined by Oja (1981).

5.2 Lawrence's Ordering

Using the concept of a star-shaped function, Lawrence (1966) defined the ordering $\leq_R$ by

$$F \leq_R G \text{ iff } [R_{F,G}(x) - m_G]/(x - m_F) \quad \text{is increasing for } x > m_F$$

(equivalently, decreasing for $x < m_F$). $F \leq_R G$ holds iff $R_{F,G}(x)$ is star shaped for $x > m_F$, and we refer the reader to Bruckner and Ostrow (1962) for properties of such functions. Star-shaped functions have been used extensively in reliability theory to describe the concept of "wearout" and were discussed further by Barlow and Proschan (1966) and Barlow, Marshall, and Proschan (1969).

Lawrence (1975) showed that $\leq_R$ is preserved by the standardized even central moments. The measures $t_P(H)$, $st_P(H)$, and $t_{\gamma,\delta}(H)$ discussed in Section 3 also preserve $\leq_R$. The ordering has a number of applications. Rivest (1982) investigated various families of distributions and proved the following:

1. The family of central t distributions is totally ordered by $\leq_R$, with kurtosis decreasing with increasing degrees of freedom.
2. The family of Tukey models H_k, with distribution functions $H_k(x) = (1-\varepsilon)H(x) + \varepsilon H(x/k)$ [where $\varepsilon < ½$ is fixed, H is symmetric about 0, and $H(cx)$ has the monotone likelihood property] is totally ordered by $\leq_R$. Kurtosis increases with k.
3. If F_α is the distribution of a symmetric stable law with exponent α, then $\delta < \gamma \Rightarrow F_\gamma \leq_R F_\delta$.

Bickel and Lehmann (1975) showed that the ordering is preserved by the asymptotic relative efficiencies of trimmed means, and Doksum (1969) used $\leq_R$ to investigate the power of two-sample monotone rank tests with translation alternatives. Lawrence (1975) obtained stochastic comparisons between combinations of order statistics arising from $\leq_R$-ordered distributions. These results, of course, also hold for $\leq_S$, since it is a stronger ordering.

5.3 Loh's Ordering

Loh (1982) introduced an ordering $\leq_T$ implied by $\leq_R$ and defined by $F \leq_T G$ iff

$$g(m_G)[G^{-1}(.5+p)-m_G] \geq f(m_F)[F^{-1}(.5+p-m_F)]$$

for $0 \leq p < ½$. Loh noted that $F \leq_T G$ holds iff the random variable $|g(m_G)(Y-m_G)|$ is stochastically larger than $|f(m_F)(X-m_F)|$, where X and Y have distributions F and G, respectively. A well-known property of stochastically ordered random variables then implies that if $\Omega[\cdot]$ is a positive function symmetric about 0 and increasing on $(0,\infty)$, then $E_H[\Omega[h(m_H)(X-m_H)]]$ is a measure for $\leq_T$ (where X has distribution H). In particular, $E_H[[h(m_H)\cdot(X-m_H)]^{2r}]$ is a measure for $\leq_T$ for each positive integer r; so although the usual standardized, even central moments do not preserve $\leq_T$, it is preserved by these alternatively standardized moments. Horn's (1983) peakedness measures also preserve $\leq_T$ for all p. Although $\beta_2(\alpha, H)$ does not preserve $\leq_R$ or $\leq_T$, it does preserve many of the weak orderings alluded to in Section 6. Loh (1984) used $\leq_T$ to characterize some families of failure-rate distributions and to obtain bounds for certain asymptotic relative efficiencies over these families.

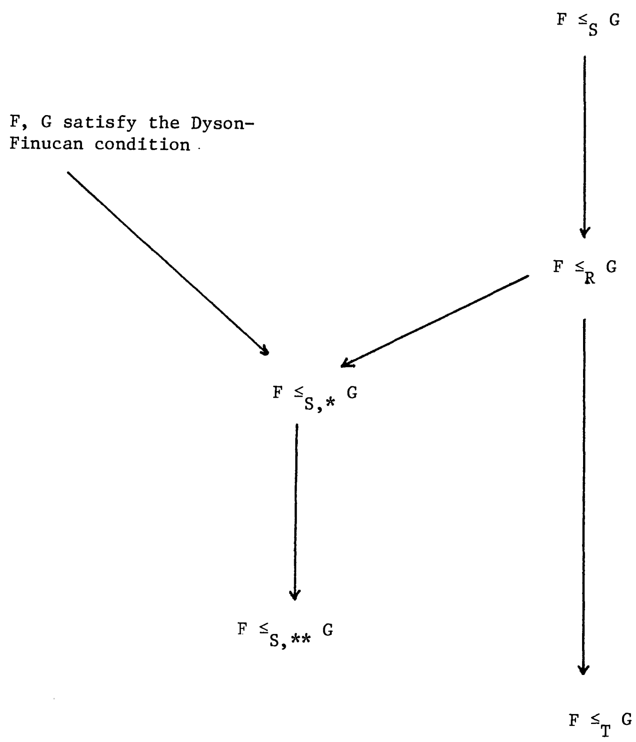

Figure 5. The Relationships Between Existing Kurtosis Orderings on Symmetric Distributions.

5.4 Oja's Orderings

Using characterization (b) of van Zwet's convexity condition (see Sec. 5.1), Oja (1981) developed two moment-based weakenings of $\leq_S$. If F and G have finite means μ_F and μ_G and finite standard deviation σ_F and σ_G, Oja (1981) defined $\leq_{S,*}$ by $F \leq_{S,*} G$ iff there exists x_1 and x_2 such that

$$\begin{aligned} F(x) &\leq G(\sigma_G x/\sigma_F - \sigma_G \mu_F/\sigma_F) \\ &\qquad\qquad \text{for } x < x_1 \text{ or } \mu_F < x < x_2 \\ &\geq G(\sigma_G x/\sigma_F - \sigma_G \mu_F/\sigma_F) \\ &\qquad\qquad \text{for } x_1 < x < \mu_F \text{ or } x_2 < x. \end{aligned}$$

Oja defined a further ordering $\leq_{S,**}$ by $F \leq_{S,**} G$ iff there exist c, d, x_1, x_2, and x_3 with $x_1 < x_2 < x_3$ such that

$$\begin{aligned} F(x) &\leq G(cx+d) \quad \text{for } x < x_1 \text{ or } x_2 < x < x_3 \\ &\geq G(cx+d) \quad \text{for } x_1 < x < x_2 \text{ or } x_3 < x. \end{aligned}$$

Oja (1981) proved that the standardized even central moments preserve $\leq_{S,*}$ and that if standardized f and g satisfy the Dyson–Finucan condition then $F \leq_{S,*} G$. We refer the reader to Oja (1981) for further discussion of these orderings.

Figure 5 summarizes the relationships between these orderings. All of the orderings $<<$ are location and scale free and have the property that $F << G$ and $G << F$ hold simultaneously (for symmetric F and G) iff there exist constants a, b such that $G(x) = F(ax+b)$ for all x. The orderings $\leq_S$, $\leq_R$, and $\leq_T$ are transitive and thus induce partial orderings on location-scale families of symmetric distributions. Oja's orderings are not transitive, so they cannot be used to make meaningful comparisons within families of more than two distributions. This difficulty arises essentially because the mean and the variance (moment based) are taken to be measures of location and scale. If quantile-based measures are chosen instead, this problem does not arise and the obtained orderings are transitive.

6. CURRENT WORK

We have argued that kurtosis should be viewed as a vague concept best formalized using partial orderings on distributions and measures that preserve them. Only a few orderings have been defined to date, however, and these only on symmetric distributions. Consequently, the weakest ordering underlying several of the measures discussed in Section 3 have not been identified, and the notion of kurtosis in asymmetric distributions and its relationship with skewness have not been discussed. These problems need further attention.

We are considering these problems in some current work. The work defines a structure of location- and scale-free partial orderings on arbitrary distributions. The structure consists of hierarchies of orderings of varying strengths, and each hierarchy corresponds to a formalization of kurtosis arising from the use of a particular scaling technique, positioning of shoulders, and, in the asymmetric case, measure of location. We consider extensions of van Zwet's (1964) ordering to the asymmetric case, and we investigate the relationship between skewness and kurtosis. Interested readers are referred to the appropriate references.

[*Received June 1987. Revised October 1987.*]

REFERENCES

Ali, M. M. (1974), "Stochastic Ordering and Kurtosis Measure," *Journal of the American Statistical Association*, 69, 543–545.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), "Robust Estimates of Location: Survey and Advances," technical report, Princeton University Press, Princeton, NJ.

Balanda, K. P. (1986), "Kurtosis, Peakedness and Tailweight," unpublished M.Sc. thesis, University of Queensland, Dept. of Mathematics.

——— (1987), "Kurtosis Comparison of the Cauchy and Double Exponential Distributions," *Communications in Statistics—Theory and Methods*, 16, 579–592.

Balanda, K. P., and MacGillivray, H. L. (1987), "Kurtosis and Spread," unpublished manuscript.

Barlow, R. E., Marshall, A. W., and Proschan, F. (1969), "Some Inequalities for Starshaped and Convex Functions," *Pacific Journal of Mathematics*, 29, 19–42.

Barlow, R. E., and Proschan, F. (1966), "Inequalities for Linear Combinations of Order Statistics From Restricted Families," *Annals of Mathematical Statistics*, 37, 1574–1592.

Bickel, P. J., and Lehmann, E. L. (1975), "Descriptive Statistics for Non-Parametric Models (Introduction and Location)," *The Annals of Statistics*, 3, 1038–1069.

Birnbaum, Z. W. (1948), "On Random Variables With Comparable Peakedness," *Annals of Mathematical Statistics*, 37, 1593–1601.

Bruckner, A. M., and Ostrow, E. (1962), "Some Function Classes Related to the Class of Convex Functions," *Pacific Journal of Mathematics*, 12, 1203–1215.

Chissom, B. S. (1970), "Interpretation of the Kurtosis Statistic," *The American Statistician*, 24, 19–23.

Crow, E. L., and Siddiqui, M. M. (1967), "Robust Estimation of Location," *Journal of the American Statistical Association*, 62, 353–389.

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

Doksum, K. A. (1969), "Starshaped Transformations and the Power of Rank Tests," *Annals of Mathematical Statistics*, 40, 1167–1176.

Dyson, F. J. (1943), "A Note on Kurtosis," *Journal of the Royal Statistical Society*, Ser. B, 106, 360–361.

Finucan, H. M. (1964), "A Note on Kurtosis," *Journal of the Royal Statistical Society*, Ser. B, 26, 111–112.

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, London: Oliver & Boyd.

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

Gumbel, E. J. (1962), "Statistical Theory of Extreme Values (Main Results)," in *Contributions to Order Statistics*, eds. A. E. Sarhan and B. G. Greenberg, New York: John Wiley, pp. 56–93.

Hettmansperger, T. P., and Keenan, M. A. (1975), "Tailweight, Statistical Inference and Families of Distributions—A Brief Survey," in *Statistical Distributions in Scientific Work* (Vol. 1), ed. G. P. Patil, Dordrecht, Holland: D. Reidel, pp. 161–172.

Hildebrand, D. K. (1971), "Kurtosis Measures Bimodality?" *The American Statistician*, 25, 42–43.

Hogg, R. V. (1974), "Adaptive Robust Procedures," *Journal of the American Statistical Association*, 69, 909–921.

Horn, P. S. (1983), "A Measure for Peakedness," *The American Statistician*, 37, 55–56.

Kaplansky, I. (1945), "A Common Error Concerning Kurtosis," *Journal of the American Statistical Association*, 40, 259.

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

Kendall, M. G., and Stuart, A. (1977), *The Advanced Theory of Statistics* (Vol. 1: Distribution Theory), London: Charles W. Griffin.

Lawrence, M. J. (1966), "Inequalities and Tolerance Limits for *s*-Ordered Distributions," Technical Report 66-37, University of California, Berkeley, Operations Research Center.

——— (1975), "Inequalities of *s*-Ordered Distributions," *The Annals of Statistics*, 3, 413–428.

Loh, W. J. (1982), "Tail Orderings on Symmetric Distributions With Statistical Applications," unpublished Ph.D. thesis, University of California, Berkeley, Graduate Division.

——— (1984), "Bounds on AREs for Restricted Classes of Distributions Defined via Tail Orderings," *The Annals of Statistics*, 12, 685–701.

MacGillivray, H. L. (1985), "A Crossing Theorem for Distribution Functions and Their Moments," *Bulletin of the Australian Mathematical Society*, 31, 413–419.

——— (1986), "Skewness and Asymmetry: Measures and Orderings," *The Annals of Statistics*, 14, 994–1011.

MacGillivray, H. L., and Balanda, K. P. (1987), "The Relationships Between Skewness and Kurtosis," unpublished manuscript.

Marsaglia, G., Marshall, A. W., and Proschan, F. (1965), "Moment Crossings as Related to Density Crossings," *Journal of the Royal Statistical Society*, Ser. B, 27, 91–93.

Moors, J. J. A. (1986), "The Meaning of Kurtosis: Darlington Reexamined," *The American Statistician*, 40, 283–284.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.

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

Parzen, E. (1979), "Nonparametric Statistical Data Modelling" (with discussion), *Journal of the American Statistical Association*, 74, 105–121.

——— (1980), "Quantile Functions, Convergence in Quantile and Extreme Value Distribution Theory," Technical Report B-3, Texas A&M University, Institute of Statistics.

Pearson, K. R. (1905), "Skew Variation, a Rejoinder," *Biometrika*, 4, 169–212.

Rivest, L. P. (1982), "Products of Random Variables and Star Shaped Ordering," *The Canadian Journal of Statistics*, 10, 219–223.

Rogers, W. H., and Tukey, J. W. (1972), "Understanding Some Long-Tailed Symmetrical Distributions," *Statistica Neerlandica*, 26, 211–226.

Rosenberger, J. L., and Gasko, M. (1983), "Comparing Location Estimators: Trimmed Means, Medians and Trimean," in *Understanding Robust and Exploratory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, New York: John Wiley, pp. 297–338.

Ruppert, D. (1987), "What Is Kurtosis?" *The American Statistician*, 41, 1–5.

Schuster, E. F. (1984), "Classification of Probability Laws by Tail Behavior," *Journal of the American Statistical Association*, 79, 936–939.

Van Zwet, W. R. (1964), "Convex Transformations of Random Variables," Mathematics Centre Tract 7, Mathematisch Centrum, Amsterdam.