The Annals of Statistics
1997, Vol. 25, No. 5, 1984–1997

SKEWNESS FOR MULTIVARIATE DISTRIBUTIONS: TWO APPROACHES

BY JEAN AVÉROUS AND MICHEL MESTE

Paul Sabatier University

This paper presents two approaches for qualitative, quantitative and comparative concepts of skewness to be defined with respect to the spatial median for multivariate distributions. They extend the known quantile-based notions defined for real distributions. The main tool for such extensions consists of a family of central parts that provide suitable generalizations of the real interquantile intervals.

1. Introduction. Real skewed distributions have been extensively studied in the literature, for they offer adequate models for many examples from various fields. Theoretical developments for the description of these distributions from the viewpoint of skewness can be found in many papers [see, e.g., Oja (1981), Grœneveld and Meeden (1984), MacGillivray (1986), MacGillivray and Balanda (1988), Avérous and Meste (1990), Benjamini and Krieger (1996)]. Comparative, qualitative and quantitative aspects are taken into account by these authors, defining, respectively, orderings, classes of distributions and functional or single-valued measures. These orderings, classes and measures must be location and scale free, and a measure must be monotone with respect to a skewness ordering. The usual classes are associated with a given skewness ordering $\prec$ by comparing F with $\overline{F}$ [where $\overline{F}(x) = 1 - F(-x)$]: F is said to be skew to the right if and only if $\overline{F} \prec F$.

The very strong convex ordering introduced by van Zwet (1964) is the reference ordering for the skewness structure of continuous univariate distributions: F will be called less skew to the right than G ($F \prec G$) if and only if $G^{-1} \circ F$ is convex. This ordering does not refer to any location or scale parameter, and thus may be used to study skewness with respect to any center. Weaker orderings have been defined to enlarge the family of comparable distributions, particularly involving a weakening of the convexity of $G^{-1} \circ F$. In general, they need the selection of a location parameter and the choice of a scaling technique, and then they appear as orderings of skewness with respect to the given center [see Oja (1981), MacGillivray (1986)]. In this paper, we are interested in generalizing skewness with respect to the median M_F.

Although the skewness concept is expressed, as are other descriptive concepts, by a stochastic ordering, it can be noted that several advances in skewness description arose from the definition of various classes and mea-

Received February 1993; revised March 1997.
AMS 1991 *subject classifications*. Primary 62H05; secondary 60E05.
Key words and phrases. Skewness, tailweight, spatial median, central parts, interquantile intervals, orderings.

sures, without referring to a particular ordering. This is the case for the class of skewed distributions defined by Doksum (1975) in the following two equivalent ways.

The first uses interquantile intervals by means of the symmetry function $x \to \theta_F(x) = \frac{1}{2}[x - \overline{F}^{-1}(F(x))]$, giving the center of the interquantile interval with bounds x and $F^{-1}(1 - F(x))$. Doksum defines qualitative right skewness for F by the positivity of the functional $x \to \theta_F(x) - M_F$. The second uses intervals centered at M_F; F is skewed to the right if $1 - F(M_F + t) \geq F(M_F - t)$ for any positive t. This means that, for right and left tails symmetric with respect to M_F, the weight of the right tail is heavier than that of the left.

Note that the functionals used in qualitative exploration of skewness need to be scaled to lead to a skewness ordering. However, an ordering so defined may not necessarily be classified in an already recognized structure. Moreover, for a given scaling, two distributions equivalent for such an ordering may not have equivalent shapes. This point will be discussed in Section 2.

The usual skewness parameters are defined as measures of skewness with respect to a given center. For instance, the third standardized moment μ_3/σ^3 measures skewness against the mean. Other parameters based on the interquantile intervals $[F^{-1}(\alpha), F^{-1}(1 - \alpha)]$ measure skewness with respect to the median. A detailed study of the relationships between quantitative, qualitative and comparative concepts of skewness will be found in MacGillivray (1986).

For multidimensional distributions, an extension of the quantile-based approach first needs to define multidimensional "quantiles" or interquantile domains [a comprehensive review of the difficulties of such an extension will be found in Barnett (1976)]. In fact, only the quantitative aspects have been studied, by introducing several real-valued generalizations of μ_3/σ^3, measuring asymmetry with respect to the mean [see Mardia (1970), Malkovitch and Afifi (1973), Srivastava (1984)]. The only vector-valued skewness parameter was proposed by Oja (1983). Finally, some investigation of the different kinds of symmetry has been made by Blough (1989), using a projection pursuit method.

In this paper, after a brief survey of the literature about skewness for univariate or multivariate distributions, we present a multivariate generalization of each of the two above-mentioned approaches of Doksum. In these extensions, the spatial median is the reference center. Its properties have been extensively studied by Kemperman (1987). In order to extend to $\mathbb{R}^n$ the first approach, we shall use in Section 3 a family of balls generalizing the interquantile intervals. For this, we use a particular characterization of these intervals introduced in Avérous and Meste (1990). We extend the second approach using the family of balls centered at the spatial median, and introducing an appropriate concept of $\mathbb{L}_1$-weight for directed tails. Unlike the real case, these extensions produce two different classes. Finally, an example of a family of distributions possessing a given skewness is proposed, within the framework of the second approach.

2. Skewness with respect to the median for real distributions. Let X be a real random variable with a continuous and strictly increasing c.d.f. F. The extension to the general case needs only minor modifications. In this section, we consider skewness with respect to the median M_F, the corresponding weight of the left (resp. right) tail $(-\infty, x)$ [resp. $(x, +\infty)$] being $F(x)$ [resp. $1 - F(x)$]. Then the median appears as the point that defines two equally weighted tails. More generally, for any $\alpha \in (0, \frac{1}{2}]$, the closed interquantile interval $[F^{-1}(\alpha), F^{-1}(1-\alpha)]$ has the same property.

Two functionals that have arisen in univariate qualitative skewness are the following [see Doksum (1975)]:

$$\alpha \to m_F(\alpha) = \left[F^{-1}(\alpha) + F^{-1}(1-\alpha)\right]/2, \qquad \alpha \in (0,1)$$

$$x \to \theta_F(x) = \tfrac{1}{2}\left[x - \overline{F}^{-1}(F(x))\right], \qquad x \in \mathbb{R}^n$$

where $\overline{F}$ is the c.d.f. of $-X$. A qualitative weak skewness to the right is then defined by the positivity of $m_F - M_F$ or equivalently by that of $\theta_F - M_F$. A stronger skewness is defined by "m_F is nonincreasing on $(0, 1/2)$" or equivalently, "θ_F is nonincreasing for $x \le M_F$ and nondecreasing for $x \ge M_F$." The strongest definition, associated with the van Zwet skewness ordering, is based on the concavity of the mapping $\overline{F}^{-1} \circ F$ or equivalently on the convexity of θ_F.

These definitions of qualitative skewness can be expressed equivalently by the properties of the mapping

$$\lambda \to c_F(\lambda),$$

which associates with any $\lambda \ge 0$ the center of the interquantile interval with length 2λ [see Avérous and Meste (1990)]. When it is unique, $c_F(\lambda)$ is the location M-parameter solution of the minimization problem:

$$\inf_c \int_{\mathbb{R}} \rho_\lambda(x - c)\, dF(x)$$

where $\rho_\lambda(y) = (|y| - \lambda)\mathbb{1}_{(\lambda, +\infty)}(|y|)$. The three previous concepts of qualitative skewness are equivalently defined, respectively, by $c_F - M_F \ge 0$, c_F nondecreasing and c_F both convex and nondecreasing. The functionals θ_F, m_F, c_F provide three parametrizations for the centers of the interquantile intervals. Their links can be visualized by means of the following diagram [Avérous and Meste (1990)]:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\;F\;} & (0,1) \\ {\scriptstyle\theta_F}\downarrow & \swarrow {\scriptstyle m_F} & \downarrow{\scriptstyle\lambda_F} \\ \mathbb{R} & \xleftarrow{\;c_F\;} & \mathbb{R}_+ \end{array},$$

where

$$\lambda_F(\alpha) = \left|F^{-1}(\alpha) - F^{-1}(1-\alpha)\right|/2.$$

The functional c_F does not make use of the quantile function F^{-1} and thus allows a generalization to the multidimensional case, given in Section 4.

Another approach is to compare the weights of the right and left tails symmetrically located with respect to the median. The condition

$$1 - F(M_F + x) - F(M_F - x) \geq 0, \qquad x \geq 0, \tag{1}$$

has been given by Doksum (1975) to define a qualitative skewness to the right. Doksum has shown that this positivity is equivalent to those of $\theta_F - M_F$. Van Zwet (1979) has given (1) as a sufficient condition for the inequality: mean(F) ≥ median(F) ≥ mode(F).

The functionals used to define the previous concept of weak skewness are not scale free and so a scaling technique is required for their use in defining a skewness ordering. However, a skewness ordering compatible in the following sense with a qualitative skewness,

$$\overline{F} \prec F \Leftrightarrow F \text{ is skew to the right},$$

does not need to derive from one of the previous functionals, especially if it must take place in a recognized structure (and particularly to be weaker than the van Zwet ordering). MacGillivray has defined the following weak skewness ordering, where f, g are the probability density functions of F and G:

$$\begin{aligned} F \prec_2^m G &\Leftrightarrow \left(G^{-1} \circ F(x) - M_G\right) g(M_G) \geq (x - M_F) f(M_F) \\ &\Leftrightarrow \left[G^{-1}(u) - M_G\right] g(M_G) \geq \left[F^{-1}(u) - M_F\right] f(M_F) \end{aligned} \tag{2}$$

$$\text{for } x \in I_F \text{ and } u \in (0,1).$$

This ordering, which corresponds to the usual stochastic ordering on centered and scaled distributions, possesses the property of shape equivalence: two equivalent distributions are equal up to a change of location and scale. Moreover, as shown by MacGillivray (1986), the choice of $1/f(M_F)$ as scaling parameter is necessary for this ordering to be weaker than the van Zwet one. Another ordering for the weak skewness has been proposed by Grœneveld and Meeden (1984), based on the functional

$$u \to \gamma_F(u) = \frac{m_F(u) - M_F}{\lambda_F(u)}, \qquad u \in (0, 0.5]$$

introduced by David and Johnson (1956). This ordering is defined by

$$F \prec_{2,\gamma}^m G \quad \Leftrightarrow \quad \gamma_F(u) \leq \gamma_G(u)$$

for $u \in (0, 1/2)$. It is weaker than the ordering $\prec_2^m$, but does not have the property of shape equivalence. Let us note that the scaling technique chosen to get the skewness functional γ_F from m_F uses a dispersion functional instead of a real dispersion parameter.

Using the parametrization of the interquantile intervals by their length, Avérous and Meste (1990) gave the following ordering, which is also compatible with the weak qualitative skewness

$$F \prec_{sk} G \quad \Leftrightarrow \quad \mathscr{S}_F \leq \mathscr{S}_G,$$

where

$$\mathscr{S}_F(\lambda) = \frac{c_F(\lambda\sigma_F) - c_F(0)}{\sigma_F},$$

σ_F being a given arbitrary scale parameter.

Here also, choosing $\sigma_F = 1/f(M_F)$ allows ordering weaker than $\prec_2^m$. As the ordering $\prec_{2,\gamma}^m$, it does not have the property of shape equivalence. However, the extension of $\mathscr{S}_F$ will be used in Section 3 to define a weak skewness for multivariate distributions.

For the quantitative aspects, various measures of skewness monotone for the weak skewness, with respect to the median, have been introduced. For instance, the parameter $(q_u + q_l - 2M_F)/(q_u - q_l)$, where q_u and q_l are the upper and lower quartiles, was one of the first quantile-based measures. This parameter, which can be also written $(m_F(1/4) - m_F(1/2))/\lambda_F(1/4)$, is no more than a summary of the skewness functional γ_F [see Benjamini and Krieger (1996) for other skewness measures based on quantiles]. They provide alternatives to the usual third standardized moment μ_3/σ^3, which is monotone for the convex ordering of van Zwet, but not for the ordering $\prec_2^m$.

3. Skewness and asymmetry for multivariate distributions.

3.1. *Existing quantitative notions.* Suppose X is a random vector from some multivariate distribution function F on $\mathbb{R}^n$. Rather than skewness, asymmetry has been investigated, in particular by means of various real asymmetry measures, generalizing the real parameters. In particular, Mardia (1970) defined

$$\beta_{1,n} = \mathbb{E}\Big[\big\{(X-\mu)'\Sigma^{-1}(Y-\mu)\big\}^3\Big],$$

where X and Y are i.i.d. random vectors with mean μ and covariance matrix Σ. Other generalizations of the usual third standardized moment have been proposed [see Srivastava (1984), Malkovitch and Afifi (1973), Móri, Rohatgi and Székely (1994)]. Oja (1983) generalized the Pearson measure of skewness by the parameter $(\mu_2 - \mu_1)'\Sigma^{-1}(\mu_2 - \mu_1)$, where μ_1 is the Oja median and μ_2 is a generalized mean, $\mu_2 - \mu_1$ giving the direction of skewness. Chaudhuri (1996) suggested an asymmetry measure based on a geometric approach of multivariate quantiles.

Blough (1989) used a projection pursuit approach to define a location region, generalizing the location interval of Doksum, whose length measures the asymmetry of a real distribution. More precisely, let $\mathscr{O}$ be the set of orthogonal transformations on $\mathbb{R}^n$ and let I_G be the univariate location interval for G defined by

$$I_G = \big[\inf\{m_G(u);\, u \in (0, 0.5]\}, \sup\{m_G(u);\, u \in (0, 0.5]\}\big].$$

Then, after F has been standardized to remove location, scale and linear correlations, the location region of F is defined by

$$B_F = \bigcap_{T\in\mathscr{O}} T^{-1}(D_{F_T})$$

where F_T is the distribution function of $T(X)$ and D_G is the Cartesian product of the location intervals I_{G_i} of the marginal distributions G_i from G. If the distribution of X is symmetric with respect to a hyperplane H, the width of B_F in the direction of the orthogonal complement of H is null. Though this approach also may be used to detect directions of strong skewness, it seems less adapted to give a description of directed skewness or to construct skewed distributions possessing assigned skewness properties.

As pointed out by several authors, no partial ordering or classes have been proposed for qualitative multivariate skewness generalizing skewness to the right for real distributions. The following extensions to multivariate distributions of the two approaches described in Section 2 allow us to fill the gap by defining criteria of weak qualitative skewness and some corresponding orderings.

3.2. *Median balls approach.* Let X be a random vector with probability distribution (p.d.) P on $\mathbb{R}^n$. Among the various generalizations of the real median [see Small (1990)], we use here the spatial median c_0 induced by the usual Euclidean norm ($\|\cdot\|$), that is defined as a solution of the minimization problem:

$$\inf_c \int_{\mathbb{R}^n} (\|x - c\| - \|x\|)\, dP(x).$$

In the following, we consider only skewness with respect to the spatial median. To generalize the first approach of Section 2, we use a generalization of the real interquantile intervals given by the following family of central parts, called median balls [see Avérous and Meste (1993)]. Let $B(c, \lambda)$ be the closed ball with center c and radius λ ($\lambda \geq 0$), and let

$$\rho_\lambda(y) = (\|y\| - \lambda)\, \mathbb{1}_{(\lambda, \infty)}(\|y\|),$$

$$f(c, \lambda) = \int_{\mathbb{R}^n} (\rho_\lambda(x - c) - \rho_0(x))\, dP(x).$$

DEFINITION 1. A median ball for P is a ball $B(c_P(\lambda), \lambda)$ with center $c_P(\lambda)$ the solution of

$$\inf_c f(c, \lambda).$$

For the sake of simplicity, $c_P(\lambda)$ will be also denoted by c_λ. The centers c_λ are defined for any distribution P and any $\lambda \geq 0$, and c_λ is unique when the support of P is not included in the union of a ball $B(c, \lambda)$ and a straight line containing c (this last condition will be assumed in the following).

The median ball with null radius is nothing more than the spatial median and the centers c_λ can be interpreted as M-parameters [Haberman (1989)]. The equivariance, robustness and convergence properties of these centers are those of the spatial median [Avérous and Meste (1993)].

As it is usually done for the M-parameters, the centers c_λ can be characterized by the directional derivatives of the mapping $f(\cdot, \lambda)$. More precisely, if we denote by $\mathscr{S}^{n-1}$ the unit sphere (i.e., the surface of the unit ball) in $\mathbb{R}^n$, let $y \to \Psi_{y,\lambda}(h)$ be the derivative of the mapping $y \to \rho_\lambda(y)$ in the direction $h \in \mathscr{S}^{n-1}$. Then we have the following.

PROPOSITION 1. *The centers c_λ can be characterized by*

$$\int_{\mathbb{R}^n} \Psi_{c_\lambda - x,\lambda}(h)\, dP(x) \geq 0 \qquad \forall\, h \in \mathscr{S}^{n-1}. \tag{3}$$

PROOF. The convexity of $f(\cdot, \lambda)$ for any $\lambda \geq 0$ implies that c_λ is characterized by

$$\forall\, h \in \mathscr{S}^{n-1}, \qquad \lim_{t \downarrow 0} \int_{\mathbb{R}^n} \frac{1}{t}\left[\rho_\lambda(x - c - th) - \rho_\lambda(x - c)\right] dP(x) \geq 0.$$

We have $\forall (c, x) \in (\mathbb{R}^n)^2$, $\forall\, h \in \mathscr{S}^{n-1}$, $\forall\, t \in \mathbb{R}$,

$$\left|\rho_\lambda(x - c - th) - \rho_\lambda(x - c)\right| \leq t\|h\|$$

so that, using the bounded convergence theorem and the relation

$$\Psi_{c-x,\lambda}(h) = \Psi_{x-c,\lambda}(-h),$$

we obtain

$$\int_{\mathbb{R}^n} \Psi_{c-x,\lambda}(h)\, dP(x) = \lim_{t \downarrow 0} \int_{\mathbb{R}^n} \frac{1}{t}\left[\rho_\lambda(x - c - th) - \rho_\lambda(x - c)\right] dP(x),$$

and the result follows. □

Let $D_\lambda(h)$ be the set $\{y \in \mathbb{R}^n; \langle y, h\rangle > 0,\ \|y\| > \lambda\}$ (where $\langle\cdot,\cdot\rangle$ denotes the usual scalar product). Then,

$$\Psi_{y,\lambda}(h) = \frac{\langle y, h\rangle}{\|y\|} \mathbb{1}_{\overline{D_\lambda(h) \cup D_\lambda(-h)}}(y).$$

If the probability of any sphere is null, then the previous characterization becomes

$$\int_{\mathbb{R}^n} \Psi_{c_\lambda - x,\lambda}(h)\, dP(x) = 0$$

or equivalently

$$\int_{c_\lambda + D_\lambda(h)} \frac{|\langle c_\lambda - x, h\rangle|}{\|c_\lambda - x\|}\, dP(x) = \int_{c_\lambda + D_\lambda(-h)} \frac{|\langle c_\lambda - x, h\rangle|}{\|c_\lambda - x\|}\, dP(x)$$

for any h in $\mathscr{S}^{n-1}$. This last characterization leads to the interpretation of the median ball $B(c_\lambda, \lambda)$ as the ball that determines, for any direction h, two opposite tails: $c_\lambda + D_\lambda(h)$ and $c_\lambda + D_\lambda(-h)$, which are equally weighted if the weight of a tail $c_\lambda + D_\lambda(h)$ is defined as

$$W(c_\lambda + D_\lambda(h)) = \int_{c_\lambda + D_\lambda(h)} \frac{|\langle c_\lambda - x, h\rangle|}{\|c_\lambda - x\|}\, dP(x). \tag{4}$$

For a continuous distribution on the real line, the weight of a tail reduces to its probability, and the family of tails is no more than the family of right and left open tails with equal probabilities, determined by the closed interquantile intervals parametrized by their half-length.

For a noncontinuous distribution on $\mathbb{R}$, a median M does not necessarily define two tails with equal probabilities, but can be characterized by

$$P((-\infty, M]) \geq P((M, +\infty)) \quad \text{and} \quad P((-\infty, M)) \leq P([M, +\infty)).$$

Such a characterization, which uses open and closed tails, remains valid for the interquantile intervals and is extended to the median balls in the multivariate case by the characterization (3) of c_λ. This characterization may be equivalently written

$$W\big(c_\lambda + \overline{D_\lambda(h)}\big) \geq W\big(c_\lambda + D_\lambda(-h)\big)$$

for any h in $\mathscr{S}^{n-1}$.

The links between this tail weight and the probability of the median balls are given by the following proposition (where $\mathscr{B}(\cdot,\cdot)$ denotes the usual beta function and μ the uniform probability on $\mathscr{S}^{n-1}$).

PROPOSITION 2. *If the probability of any sphere is null, then*

$$\int_{\mathscr{S}^{n-1}} W(c_\lambda + D_\lambda(h))\, d\mu(h) = \frac{2}{(n-1)\mathscr{B}\big(\frac{1}{2}, (n-1)/2\big)} P[B^c(c_\lambda, \lambda)].$$

PROOF. We have

$$\begin{aligned}
&\int_{\mathscr{S}^{n-1}} W(c_\lambda + D_\lambda(h))\, d\mu(h) \\
&\quad = \int_{\mathscr{S}^{n-1}} \left[\int_{\mathbb{R}^n} \frac{|\langle c_\lambda - x, h\rangle|}{\|c_\lambda - x\|} \mathbb{1}_{\mathbb{R}_+^*}(\langle x - c_\lambda, h\rangle) \mathbb{1}_{(\lambda, +\infty)}(\|x - c_\lambda\|)\, dP(x) \right] d\mu(h) \\
&\quad = \int_{\mathbb{R}^n} \mathbb{1}_{(\lambda, +\infty)}(\|x - c_\lambda\|) \left[\int_{\mathscr{S}^{n-1}} \left| \left\langle \frac{c_\lambda - x}{\|c_\lambda - x\|}, h \right\rangle \right| \mathbb{1}_{\mathbb{R}_+^*}(\langle x - c_\lambda, h\rangle)\, d\mu(h) \right] dP(x)
\end{aligned}$$

by the Fubini theorem. The integral on $\mathscr{S}^{n-1}$ being independent of $c_\lambda - x$, the result follows. □

According to the previous results, the behavior of c_λ can be used, just as in the real case, to define qualitative and comparative skewness in the direction of a vector h, which generalize the weak skewness to the right with respect to the median. For such a proposal, the choice of a scaling technique will be important, the interpretation of the resultant ordering being highly related to it. As usual, this choice depends on the wanted equivariance or invariance properties of this ordering. As the reference center is here the spatial median, which is not affine equivariant, it seems justified to require the scaling technique to ensure the following properties for an ordering $\prec_h$ [where $\theta(P)$ denotes the probability induced by the transformation θ from $\mathbb{R}^n$ to $\mathbb{R}^n$]:

1. For any translation or homothety θ,

$$P \prec_h Q \;\Rightarrow\; \theta(P) \prec_h \theta(Q);$$

2. For any orthogonal transformation ρ,

$$P \prec_h Q \;\Rightarrow\; \rho(P) \prec_{\rho(h)} \rho(Q).$$

For a quantile-based scaling technique, MacGillivray (1986) has pointed out that "$F^{-1}(1-u) - F^{-1}(u)$ and $\mathbb{E}|X - M_F|$ are perhaps more appealing than $1/f(M_F)$, but the last parameter is essentially the scale measure that arises in the ordering as the van Zwet ordering is gradually weakened." In the multidimensional continuous case, this remark leads to using as a privileged scale parameter the inverse of the density at the spatial median. An extension of $\mathbb{E}|X - M_F|$ would be given by $\mathbb{E}\|X - c_0\|$.

Let $\tilde{P}$ be a standardized version of P obtained by centering at the spatial median and scaling as previously mentioned and let $c_P(\lambda)$ denote the center of the median ball with radius λ for P. Then, for any probability distribution P, the mapping $\lambda \to c_{\tilde{P}}(\lambda)$ can be considered as a *skewness function* for P. In the same way, for any unit vector h, the mapping $\lambda \to \langle c_{\tilde{P}}(\lambda), h\rangle$ will be considered as a *function of skewness in the direction h* for P.

By the interpretation and the equivariance properties of the mapping $P \to c_P(\lambda)$ [Avérous and Meste (1993)], these previous mappings allow qualitative, comparative and quantitative concepts of skewness to be introduced.

DEFINITION 2. The distribution P is defined to be weakly skew in the direction h if and only if $\langle c_{\tilde{P}}(\lambda), h\rangle \geq 0$ for any λ (this property will be denoted SK_h).

Moreover, an ordering of skewness in the direction h could be given by:

$$P \prec_h Q \text{ iff} \left\langle c_{\tilde{Q}}(\lambda) - c_{\tilde{P}}(\lambda), h\right\rangle \geq 0 \quad \text{for any } \lambda.$$

In the real case, the previous qualitative notion reduces to the weak skewness to the right (or to the left) [Doksum (1975)] that is characterized by the sign of the function θ_F, or $m_F - M_F$ (see Section 2).

Moreover, the ordering introduced in Definition 2 reduces for real distributions to the ordering $\prec_{sk}$ already defined in Avérous and Meste (1990), which has been shown to be weaker than the ordering $\prec_2^m$ given by MacGillivray (1986), if we standardize with the inverse of the density at the median.

It is clear that, as in the real case, asking the function of skewness in the direction h to be increasing instead of positive will lead to a stronger concept of multivariate skewness.

For the quantitative aspects, the previous approach allows some vector-valued parameters giving a direction of skewness to be introduced, the scalar product of these vectors with a direction h measuring the skewness in the direction h. A simple way to do this is to consider

$$\beta_\mu(P) = \int_{\mathbb{R}_+} c_{\tilde{P}}(\lambda)\, d\mu(\lambda),$$

where μ is a given probability measure on $\mathbb{R}_+$ independent of P. An example of β_μ is given if we take for μ a Dirac measure ($\mu = \delta_{\lambda_0}$), generalizing the skewness parameter $\gamma_F(0.25) = m_{\tilde{F}}(0.25)$ introduced by Yule (1911), where $\tilde{F}$ is a particular standardized version of F.

For a real r.v. X with continuous p.d. P, we have $\mathbb{E}(X) = \int_{\mathbb{R}_+} c_P(\lambda)\, dQ_P(\lambda)$, where $Q_P(\lambda) = P[c_P(\lambda) - \lambda, c_P(\lambda) + \lambda]$. Then the usual Pearson parameter $(\mathbb{E}(X) - \mathrm{Med}(X))/\sigma(X)$ [where $\sigma(X)$ is the standard deviation of X], is nothing more than $\int_{\mathbb{R}_+} c_{\tilde{P}}(\lambda)\, dQ_{\tilde{P}}(\lambda)$, where $\tilde{P}$ is the probability distribution of $(X - \mathrm{Med}(X))/\sigma(X)$. As mentioned by van Zwet (1964), this parameter is not admissible for the convex ordering. Standardizing by $1/f(M_F)$ allows obtaining the admissibility to the weak skewness $\prec_2^m$ [see MacGillivray (1986)], and then to the convex ordering.

For real distributions, the qualitative and comparative concepts are usually linked by the following property: P is skew to the right if and only if the probability $\overline{P}$ induced from P by $x \to -x$ is less skew to the right than P. Moreover, a skewness parameter must be positive for distributions that are skew to the right. This is extended by the following proposition, the proof of which is straightforward:

PROPOSITION 3. *For any distribution P and any direction h, P is weakly skew in the direction h if and only if $\overline{P} \prec_h P$. Moreover, if P is weakly skew in the direction h, then $\langle \beta_\mu(P), h \rangle \geq 0$ for any probability measure μ.*

In the previous approach, the skewness is based on the departure of c_λ from c_0. The following characterization, in terms of conditional expectations, of the probability distributions that satisfy $c_\lambda = c_0$ for any λ makes this approach more explicit and clarifies its interpretation.

PROPOSITION 4. *For any distribution P that does not charge any sphere, and any X with p.d. P,*

$$c_P(\lambda) = c_P(0), \qquad \forall \lambda \geq 0 \;\Leftrightarrow\; \mathbb{E}^{\|X - c_P(0)\|}(X) = c_P(0) \quad a.s.$$

PROOF. Let P be a probability distribution such that $c_P(0) = 0$ (which does not restrict the generality), and such that any sphere centered at 0 has a null probability. Then, if X is a r.v. with p.d. P, we have

$$\begin{aligned} c_P(\lambda) = 0, \qquad \forall \lambda \geq 0 \;&\Leftrightarrow\; \int \mathbb{1}_{[\lambda, +\infty)}(\|x\|) \frac{x}{\|x\|}\, dP(x) = 0 \qquad \forall \lambda \geq 0 \\ &\Leftrightarrow\; \forall \lambda \geq 0, \qquad \mathbb{E}\left[\mathbb{1}_{[\lambda, +\infty)}(\|X\|) \frac{1}{\|X\|} \mathbb{E}^{\|X\|}(X) \right] = 0, \\ &\Leftrightarrow\; \mathbb{E}^{\|X\|}(X) = 0 \quad \text{a.s.} \end{aligned}$$

and the result follows. □

The previous proposition shows that the property $c_P(\lambda) = c_P(0)$, $\forall\ \lambda \geq 0$ can be considered as defining a weakened version of centrosymmetry, based on the fixity of the means of the conditional distributions on the spheres centered on $c_P(0)$. This property also implies that

$$\mathbb{E}(X) = c_P(0)$$

and more generally

$$\mathbb{E}\left[(X - c_P(0))\|X - c_0\|^k\right] = 0$$

for any integer k such that $\mathbb{E}((\|X\|^{k+1}) < +\infty$.

REMARK 1. Although it has been developed with the usual Euclidean norm, the methodology may be applied for any norm on $\mathbb{R}^n$. Then the privileged weight would be associated with the directional derivatives of the chosen norm.

REMARK 2. Unfortunately, the explicit expressions for the centers c_λ of the median balls, just as those of the spatial medians, are rarely available. Nevertheless, the convexity of $f(\cdot, \lambda)$ (in fact the strict convexity under weak conditions) makes the computations no more difficult than those of the spatial median.

3.3. *Tailweight approach.* As specified in Section 2, a second approach to study skewness for real distributions is based on the comparison of the weights of tails located symmetrically with respect to the median. The previous introduction of "oriented" tails and of a weight for these tails associated in the $\mathbb{L}_1$-sense (for $\|\cdot\|$) allows a generalization of this second approach to be proposed. In particular, the weak skewness to the right defined by (1) can be easily extended by considering, for a direction h, the difference between the weights of the two directed tails $c_0 + D_\lambda(h)$ and $c_0 + D_\lambda(-h)$:

$$W[c_0 + D_\lambda(h)] - W[c_0 + D_\lambda(-h)] = \mathbb{E}\left\{\left\langle \frac{X - c_0}{\|X - c_0\|}, h\right\rangle \mathbb{1}_{(\lambda, +\infty)}(\|X - c_0\|)\right\}.$$

This difference is equal to $-\langle \nabla f(c_0, \lambda), h\rangle$, where $\nabla f(\cdot, \lambda)$ denotes the gradient of $f(\cdot, \lambda)$. In the following, $\nabla f(c_0, \lambda)$ will be denoted by $\mu_P(\lambda)$:

$$\mu_P(\lambda) = \mathbb{E}\left\{\frac{X - c_0}{\|X - c_0\|} \mathbb{1}_{(\lambda, +\infty)}(\|X - c_0\|)\right\}.$$

Here also, the mapping $\lambda \to \mu_{\tilde{P}}(\lambda)$ (where $\tilde{P}$ is a standardized version of P as mentioned in Section 3.2) may be considered as a *skewness function*. If P does not charge the spatial median c_0, we have $\mu_P(0) = 0$ and $\lim_{\lambda \to +\infty} \mu_P(\lambda) = 0$.

Then qualitative and comparative concepts of weak skewness in the direction of h are given by the following definition.

DEFINITION 3. P will be called $*$-weakly skew in the direction h iff:

$$(5)\quad (SK_h^*) \qquad \langle \mu_{\tilde{P}}(\lambda), h \rangle \geq 0 \qquad \forall \lambda \geq 0.$$

Moreover, Q will be defined to be more $*$-weakly skew in the direction h than P, and we will note $P \prec_h^* Q$ iff

$$(6)\qquad \langle \mu_{\tilde{Q}}(\lambda) - \mu_{\tilde{P}}(\lambda), h \rangle \geq 0 \qquad \forall \lambda \geq 0.$$

Just as for the weak skewness (Proposition 3), P is $*$-weakly skew in the direction h if and only if $\overline{P} \prec_h^* P$.

It is easy to show that when P does not charge any sphere, we have

$$\mu_P(\lambda) = 0, \quad \forall \lambda \geq 0 \;\Leftrightarrow\; c_P(\lambda) = c_0 \quad \forall \lambda \geq 0,$$

which points out that both median balls and tailweight approaches describe the departure from the same family of weakly symmetric distributions.

For the quantitative aspects, several vector-valued parameters giving a direction of skewness can be constructed from the mapping $\lambda \to \mu_P(\lambda)$, just as in the median ball approach. Such a parameter is given by $\mu_P(\lambda_0)$ where λ_0 is any particular value of λ. A more interesting choice is given by:

$$\int_0^{+\infty} \mu_P(\lambda)\, d\lambda = \mathbb{E}(X) - c_0.$$

This parameter, which becomes location and scale free when a standardized version of X is used, is similar to those proposed by Oja (1983). A norm of such a vector-valued parameter gives an asymmetry coefficient similar to those of Mardia (1970).

Let us remark that $\mu_P(\lambda)$ can also be written

$$\mathbb{E}\left\{ \mathbb{1}_{(\lambda, +\infty)}(\|X - c_0\|) \frac{1}{\|X - c_0\|} \mathbb{E}^{\|X - c_0\|}(X - c_0) \right\},$$

which points out that the $*$-weak skewness of P considered in this section depends heavily on the means of the conditional distributions given $\|X - c_0\| = \rho$, as a function of ρ. This interpretation is particularly useful to define families of distributions that are $*$-weakly skew. An example of such a family follows.

EXAMPLE. Let X be a random vector on $\mathbb{R}_2$ with a spatial median c_0 that will be assumed to be 0. Let $\{e_1, e_2\}$ be an orthogonal basis, so that $X = R\cos\Theta e_1 + R\sin\Theta e_2$. It is supposed that R has any density g_1 on $\mathbb{R}_+$ and that the conditional distribution of Θ given $R = \rho$ is a von Mises distribution with mean direction depending on ρ, of the form $\mu + \delta(\rho)$, and positive concentration κ. The density of this conditional distribution is given by

$$g^{\rho}(\theta) = [2\pi I_0(\kappa)]^{-1} \exp\{\kappa\cos(\theta - \mu - \delta(\rho))\}, \qquad \theta \in (-\pi, \pi).$$

Here $I_n(\cdot)$ is the modified Bessel function of the first kind and order n [see Watson (1983)], μ is a fixed parameter and δ a given mapping on $\mathbb{R}_+$. The spatial median c_0 being characterized by $\mathbb{E}((X - c_0)/\|X - c_0\|) = 0$, a necessary and sufficient condition for the spatial median of X to be fixed at 0 is given by

$$\mathbb{E}[e^{i\delta(R)}] = 0.$$

A simple way to satisfy this condition is to take $\delta = 2\pi G_1$ where G_1 is the distribution function of R, so that $\delta(R)$ is uniformly distributed on $[-\pi, \pi]$. Then, for such conditional distributions, we have

$$2\pi \frac{I_0(\kappa)}{I_1(\kappa)} \mu_P(\lambda) = \left[\sin \mu - \sin(\mu + 2\pi G_1(\lambda)), -\cos \mu + \cos(\mu + 2\pi G_1(\lambda))\right]$$

so that the distribution of X is $*$-weakly skew in the direction of the unit vector h corresponding to $\mu - \pi/2$. It follows that within this family, it is possible to obtain distributions that are $*$-weakly skew in the direction of any fixed unit vector. When κ tends to 0, the distribution of X tends to the spherical distribution such that $\|X\|$ has the density g_1. Moreover, if we consider $\|\mathbb{E}(X) - c_0\|$ to measure global asymmetry with respect to the spatial median, then this parameter, which is proportional to $(I_1/I_0)(\kappa)$, increases with κ. If we choose for g_1 the density of the chi-distribution, which is the one of the norm of a standard Gaussian vector, the Gaussian distribution appears as the limit, when κ tends to 0, of the distribution of X.

Conclusion. The methods proposed in this paper can be considered as part of the description of probability distributions "in the $\mathbb{L}_1$-sense." They can be used both for theoretical and empirical distributions (e.g., for an exploratory purpose). The methodology presented herein, based on a family of central parts defined with the proximity criterion corresponding to the spatial median, may be extended to introduce skewness with respect to other location parameters (in particular, any *M*-location parameter).

Acknowledgment. We thank the referees for their criticisms and useful comments.

REFERENCES

AVÉROUS, J. and MESTE, M. (1990). Location, skewness and tailweight in L_s-sense: a coherent approach. *Statistics* **21** 57–74.

AVÉROUS, J. and MESTE, M. (1993). Median balls: an extension of the interquantile intervals to multivariate distributions. Unpublished manuscript.

BARNETT, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A* **139** 318–355.

BENJAMINI, Y. and KRIEGER, A. M. (1996). Concepts and measures for skewness with data-analytic implications. *Canad. J. Statist.* **24** 131–140.

BLOUGH, D. K. (1989). Multivariate symmetry via projection pursuit. *Ann. Inst. Statist. Math.* **41** 461–475.

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.

DAVID, F. N. and JOHNSON, N. L. (1956). Some test of significance with ordered variables. *J. Roy. Statist. Soc. Ser. B* **18** 1–20.

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

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

HABERMAN, S. J. (1989). Concavity and estimation. *Ann. Statist.* **17** 1631–1661.

KEMPERMAN, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical Data Analysis Based on the L_1-norm and Related Methods* (Y. Dodge, ed.) 217–230. North-Holland, Amsterdam.

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

MACGILLIVRAY, H. L. and BALANDA, K. P. (1988). The relationships between skewness and kurtosis. *Austral. J. Statist.* **30** 319–337.

MALKOVITCH, J. F. and AFIFI, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68** 176–179.

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

MÓRI, T. F., ROHATGI, V. and SZÉKELY, G. J. (1994). On multivariate skewness and kurtosis. *Pub. Inst. Statist. Univ. Paris* **38** 101–108.

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

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332.

SMALL, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.* **58** 263–277.

SRIVASTAVA, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statist. Probab. Lett.* **2** 263–267.

VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Math. Centrum, Amsterdam.

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

WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.

YULE, G. U. (1911). *Introduction to the Theory of Statistics*. Griffin, London.

UNIVERSITÉ PAUL SABATIER
LABORATOIRE DE STATISTIQUE ET PROBABILITÉS
UNITE MIXTE DE RECHERCHE C5583
ROUTE DE NARBONNE
31062 TOULOUSE CEDEX 4
FRANCE
E-MAIL: averous@cict.fr