Survey on (Some) Nonparametric and Robust Multivariate Methods

Robert Serfling*

University of Texas at Dallas

June 2007

Abstract

Rather than attempt an encyclopedic survey of nonparametric and robust multivariate methods, we limit to a manageable scope by focusing on just two leading and pervasive themes, *descriptive statistics* and *outlier identification*. We set the stage with some *perspectives*, and we conclude with a look at some *open issues and directions*.

A variety of questions are raised. Is nonparametric inference the goal of nonparametric methods? Are nonparametric methods more important in the multivariate setting than in the univariate case? Should multivariate analyis be carried out componentwise, or with full dimensionality, or pairwise? Do multivariate depth, outlyingness, quantile, and rank functions represent different methodological approaches? Can we have a coherent series of nonparametric multivariate descriptive measures for location, spread, skewness, kurtosis, etc., that are robust and accommodate heavy tailed multivariate data? Can nonparametric outlier identifiers be defined that do not require ellipsoidal contours? What makes an outlier identifier itself robust against outliers? Does outlyingness of a data point with respect to location estimation differ from its outlyingness with respect to dispersion estimation? How do univariate L-functionals extend to the multivariate setting? Does the transformation-retransformation approach pose any issues? How might we conceptualize multivariate descriptive measures and outlier identification methods with respect to arbitrary data spaces, for applications such as functional data analysis, shape fitting, and text analysis?

AMS 2000 Subject Classification: Primary 62G10 Secondary 62H99.

Key words and phrases: nonparametric; robust; multivariate; descriptive measures; outlier identification.

*Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: `serfling@utdallas.edu`. Website: `www.utdallas.edu/~serfling`.

1 Introduction

This paper is based on my exposition for *Tilastopäivät 2007*. My main goal was to stimulate or provoke new ideas and fresh discussions, rather than to exposit topics with infallible accuracy. To confine to a manageable scope, We focus on just two themes, *Descriptive Statistics* and *Outlier Identification*, each pervasive in Nonparametric and Robust Multivariate Inference. First, however, we set the stage with a few *Perspectives*. We conclude with some *Open Issues and Directions*. Due to space considerations, citations are sparse. Source papers and references are on my website. And now let us proceed in the spirit of ...

> "Don't walk in front of me, I may not follow. Don't walk behind me, I may not lead. Just walk beside me and be my friend."
>
> – *Albert Camus*

2 A Few Perspectives

A few perspectives on nonparametric and robust multivariate inference are presented. These are intended for relaxed consideration.

2.1 The Goal of Nonparametric Methods is ... *Robust Parametric Modeling!*

Applications using data are best served through appropriate *parametric models*, which offer the most complete explanations. Due, however, to our uncertainty about how to model our uncertainty, the practical goal becomes *"robust" parametric modeling*. In all applications of statistical science, the fitting of such models plays the role of an *ideal endpoint*. Hence nonparametric statistics should be carried out not as an end in itself, but rather as an intermediate process oriented to, and leading toward, the ultimate goal of *robust parametric modeling*.

2.2 Multivariate Analysis Needs ... *Nonparametrics!*

Parametric modeling of multivariate data, compared to the univariate case, enjoys relatively few tractable models (although the situation is improving – see, for example, Tõnu Kollo's talk on skewed multivariate distributions). Rather by default, the *normal* model is *central*, because of the wealth of multivariate statistical theory available exclusively for it.. On the other hand, some of this nice theory requires the dimension d not to exceed the sample size n. To avoid assuming normality as a default, *nonparametric* approaches are thus even more significant in the multivariate case. *Nonparametric multivariate methods need much more development*, however. This is the challenge!

2.3 Is Multivariate Analysis Best Done ... Componentwise? Pairwise? In Full Dimensionality? *Pairwise!*

In higher dimension, the central region of a probability distribution contains little probability mass. Samples tend to fall in the tails, leaving most "local" neighborhoods empty (see Scott, 1992, for examples and discussion). Thus experience and intuition based on low dimensions are not accurate guides to the nature and handling of higher dimensional distributions and data.

Further, data in $\mathbb{R}^d$ typically has *structure of lesser dimension*, but *dimension reduction techniques* provide only partial solutions, *computional complexity* reaches practical limitations as d increases, and *visualization* so effective in the 1st and 2nd dimensions does not extend well to higher dimension. Also, for purposes of *description, modeling, or inference* with higher dimensional distributions in a way that takes into account the "*heavy tailed*" feature, even univariate methods are few and offer little as starting points.

We may not know how to think properly about high dimensional descriptive measures, even if easily formulated as natural extensions of univariate versions that (we believe) we understand. How should we think of "kurtosis" in higher dimension? While our grasp of high dimensional distributions is limited, we do, however, understand *bivariate* distributions very well. As a *partial solution*, therefore, we might pursue *increased emphasis on pairwise treatment of variables.* We already study dispersion through *covariance and correlation matrices*, and use *matrix plots (of pairwise scatterplots)* for visualization. Can the pairwise approach be developed into a *solution* instead of an *apology*? Should we develop and learn to use a *more comprehensive* pairwise approach that includes handling of *skewness* and *kurtosis*?

2.4 How Best to "Survey" Nonparametric and Robust Multivariate Methods? *With Focus, Rather than Encyclopedically!*

Statistical science enjoys many approaches: *modeling*, *descriptive measures*, *density and curve estimation*, *empirical likelihood*, *order statistics*, *outlier detection*, *quantiles*, *signs and ranks*, *smoothing*, *Bayesian methods*, *learning*, *mining*, *clustering*, *PCA*, *discriminant analysis*, *spatial methods*, ... $\to \infty$. In all of these arise notions of "robust" and "nonparametric", which, however, have become loose terms and no longer distinguishing qualifications. Nowadays, *every procedure* put forth is "robust" and/or "nonparametric", and *every statistician* is an avowed practitioner of these arts! Success has dissolved these notions! Yet ... each has *core features, concepts, and approaches*, and one can look at some of these, rather than catalog all problem settings and their associated robust or nonparametric methods.

2.5 Multivariate Depth, Outlyingness, Rank, and Quantile Functions are ... *Equivalent!*

Depth, outlyingness, quantile, and rank functions may be viewed as *equivalent methodologies in nonparametric multivariate analysis.* We "prove" this by making *the right definitions*!

Depth Functions. Given a cdf F on $\mathbb{R}^d$, a *depth function* $D(\boldsymbol{x}, F)$ provides an associated *center-outward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$. For $d \geq 2$, lack of linear order is compensated for by *orienting to a "center".* *Higher* depth represents greater "*centrality*", *maximum* depth defines "center". Depth functions yield *nested contours* and *ignore multimodality.* (In contrast, the likelihood is not a depth. It does not in general measure centrality or outlyingness, its interpretation has no global perspective, it is sensitive to multimodality, and the point of maximality is not interpretable as a "center". For F uniform on $[0, 1]^d$, for example, the likelihood function yields no contours at all.)

Some general treatments of depth functions are found in Liu, Parelius, and Singh (1999), Zuo and Serfling (2000), and Serfling (2006).

Outlyingness Functions. For F on $\mathbb{R}^d$, an *outlyingness function* $O(\boldsymbol{x}, F)$ provides an associated *center-outward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$ with *higher* values representing greater "outlyingness". "Outlyingness" and "depth" ("centrality") are equivalent *inversely*. Given $O(\boldsymbol{x}, F)$, with $\alpha_F = \sup_{\boldsymbol{x}} O(\boldsymbol{x}, F)$ finite (otherwise use O/(1+O)), the depth function $D(\boldsymbol{x}, F) = \alpha_F - O(\boldsymbol{x}, F)$ is equivalent, with values also in $[0, \alpha_F]$. A similar construction defines O from D. These differ only by a constant and generate the *same contours*, but each has its own unique appeal and role.

Quantile Functions. *Univariate quantiles* are viewed as *boundary points* demarking specified *lower and upper fractions* of the population. Each $x \in \mathbb{R}$ has a *quantile interpretation.* For *quantile-based* inference in $\mathbb{R}^d$, $d \geq 2$, it is convenient and natural to orient to a "center" that should represent a notion of multidimensional *median* and agree with a "point of symmetry" (if any). The center then serves as the *starting point* for a *median-oriented* formulation of multivariate quantiles.

The usual univariate quantile function $F^{-1}(p)$, $0 < p < 1$, transforms via $u = 2p - 1$ to $Q(u, F)$ $= F^{-1}\left(\frac{1+u}{2}\right)$, $-1 < u < 1$ (merely a re-indexing). The median is $Q(0, F) = M_F$, say. For $u \neq 0$, the index u through its *sign* indicates direction to $Q(u, F)$ from M_F and through its *magnitude* $|u| = |2F(x) - 1|$ measures outlyingness of $x = Q(u, F)$. For $|u| = c \in (0, 1)$, the "contour" $\left\{F^{-1}\left(\frac{1-c}{2}\right), F^{-1}\left(\frac{1+c}{2}\right)\right\}$ demarks the upper and lower tails of equal probability weight $\frac{1-c}{2}$. The enclosed "central region" has probability weight $|u|$.

To extend to a median-oriented formulation in $\mathbb{R}^d$, a quantile function attaches to each point $\boldsymbol{x}$ a *quantile representation* $\boldsymbol{Q}(\boldsymbol{u}, F)$ indexed by $\boldsymbol{u}$ in the unit ball $\mathbb{B}^{d-1}(\boldsymbol{0})$, with *nested* contours $\{\boldsymbol{Q}(\boldsymbol{u}, F) : \|\boldsymbol{u}\| = c\}$, $0 \leq c < 1$. For $\boldsymbol{u} = \boldsymbol{0}$, $\boldsymbol{Q}(\boldsymbol{0}, F) = \boldsymbol{M}_F$ denotes the most central point, interpreted as a *d-dimensional median.* For $\boldsymbol{u} \neq \boldsymbol{0}$, the index $\boldsymbol{u}$ represents *direction* in some sense: for example, direction to $\boldsymbol{Q}(\boldsymbol{u}, F)$ from $\boldsymbol{M}_F$, or *expected* direction to $\boldsymbol{Q}(\boldsymbol{u}, F)$ from random $\boldsymbol{X} \sim F$. The *magnitude* $\|\boldsymbol{u}\|$ represents an *outlyingness parameter*, higher values corresponding to more extreme points. The contours represent equivalence classes of points of equal outlyingness.

Depth functions induce quantile functions. For $D(\boldsymbol{x}, F)$ having nested contours enclosing the "median" $\boldsymbol{M}_F$ and bounding "central regions" $\{\boldsymbol{x} : D(\boldsymbol{x}, F) \geq \alpha\}$, $\alpha > 0$, the depth contours induce $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0})$, endowing each $\boldsymbol{x} \in \mathbb{R}^d$ with a *quantile representation.* For $\boldsymbol{x} = \boldsymbol{M}_F$, denote it by $\boldsymbol{Q}(\boldsymbol{0}, F)$. For $\boldsymbol{x} \neq \boldsymbol{M}_F$, denote it by $\boldsymbol{Q}(\boldsymbol{u}, F)$ with $\boldsymbol{u} = p\boldsymbol{v}$, where p is the probability weight of the central region with $\boldsymbol{x}$ on its boundary and $\boldsymbol{v}$ is the unit vector toward $\boldsymbol{x}$ from $\boldsymbol{M}_F$. In this case, $\boldsymbol{u}$ indicates direction toward $\boldsymbol{Q}(\boldsymbol{u}, F)$ from $\boldsymbol{M}_F$, and $\|\boldsymbol{u}\|$ is the *probability weight* of the central region with $\boldsymbol{Q}(\boldsymbol{u}, F)$ on its boundary and thus represents an outlyingness parameter.

Centered Rank Functions. A *centered rank function* in $\mathbb{R}^d$ takes values in $\mathbb{B}^{d-1}(\boldsymbol{0})$ with the origin assigned to a chosen multivariate median $\boldsymbol{M}_F$, i.e., with $\boldsymbol{R}(\boldsymbol{M}_F, F) = \boldsymbol{0}$. A "directional rank" is associated with any $\boldsymbol{x} \neq \boldsymbol{M}_F$: the vector $\boldsymbol{R}(\boldsymbol{x}, F)$ provides a "direction" in some sense, and $\|\boldsymbol{R}(\boldsymbol{x}, F)\|$ provides a "rank" measuring outlyingness of $\boldsymbol{x}$. For testing $H_0 : \boldsymbol{M}_F = \boldsymbol{\theta}_0$, the sample version of $\boldsymbol{R}(\boldsymbol{\theta}_0, F)$ provides a natural test statistic, extending the univariate sign test. (For the univariate $R(x, F) = 2F(x) - 1$, the *sign* gives "direction" (from the median) and the magnitude provides the "rank").

Quantiles and ranks are thus equivalent inversely. Given $\boldsymbol{Q}(\cdot, F)$, its *inverse function* $\boldsymbol{Q}^{-1}(\boldsymbol{x}, F)$,

$\boldsymbol{x} \in \mathbb{R}^d$, i.e., the solution $\boldsymbol{u}$ of the equation $\boldsymbol{x} = \boldsymbol{Q}(\boldsymbol{u}, F)$, is interpretable as a *centered rank function* whose magnitude $\|\boldsymbol{u}\| = \|\boldsymbol{Q}^{-1}(\boldsymbol{x}, F)\|$ measures the *outlyingness* of $\boldsymbol{x}$. Conversely, a given $\boldsymbol{R}(\cdot, F)$ generates a corresponding *quantile function* as *its* inverse: for $\boldsymbol{u}$ in $\mathbb{B}^{d-1}(\boldsymbol{0})$, $\boldsymbol{Q}(\boldsymbol{u}, F)$ is the solution $\boldsymbol{x}$ of the equation $\boldsymbol{R}(\boldsymbol{x}, F) = \boldsymbol{u}$.

Conclusion. Depth, outlyingness, quantiles, and ranks in $\mathbb{R}^d$ are equivalent, via: equivalence of $D(\boldsymbol{x}, F)$ and $O(\boldsymbol{x}, F)$, equivalence of $\boldsymbol{Q}(\boldsymbol{u}, F)$ and $\boldsymbol{R}(\boldsymbol{x}, F)$, and the links $O(\boldsymbol{x}, F) = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ $(= \|\boldsymbol{u}\|)$ and $D(\boldsymbol{x}, F) \to \boldsymbol{Q}(\boldsymbol{u}, F)$.

In sum: The *inverse* $\boldsymbol{R}(\boldsymbol{x}, F) = \boldsymbol{Q}^{-1}(\boldsymbol{x}, F)$ of a *quantile function* $\boldsymbol{Q}(\boldsymbol{u}, F)$ is interpreted as a *centered rank function*, whose *magnitude* is interpreted as an *outlyingness function* $O(\boldsymbol{x}, F)$, which in turn generates a *depth function* $D(\boldsymbol{x}, F)$. each of D, O, $\boldsymbol{Q}$, and $\boldsymbol{R}$ can generate the others, and in considering any one of them we implicitly are discussing *all* of them. We may start with one most convenient to a given context and pass to the one of focal interest for a given purpose. For example, as noted already, a depth function D induces both an outlyingness function O and a quantile function $\boldsymbol{Q}$ whose inverse is a centered rank function $\boldsymbol{R}$.

Another example is *the spatial quantile function* (Dudley and Koltchinski, 1992, and Chaudhuri, 1996), the solution $\boldsymbol{x} = \boldsymbol{Q}(\boldsymbol{u}, F)$ of $E\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X}) = \boldsymbol{u}$, with $\boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0})$, $\boldsymbol{S}(\boldsymbol{y}) = \boldsymbol{y}/\|\boldsymbol{y}\|$ $(\boldsymbol{S}(\boldsymbol{0}) = \boldsymbol{0})$, the vector sign function in $\mathbb{R}^d$, and $\boldsymbol{M}_F = \boldsymbol{Q}(\boldsymbol{0}, F)$, the *spatial median.* The corresponding *centered rank function* (Möttönen and Oja, 1995) is $\boldsymbol{R}(\boldsymbol{x}, F) = E\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\}$ and *depth function* (Vardi and Zhang, 2000) is $D(\boldsymbol{x}, P) = 1 - \|E\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\|$, for $\boldsymbol{x} \in \mathbb{R}^d$. Here $\boldsymbol{u}$ is *not* direction from $\boldsymbol{M}_F$ to $\boldsymbol{Q}(\boldsymbol{u}, F)$, but rather *expected direction* from $\boldsymbol{X} \sim F$, and $\|\boldsymbol{u}\| = \|E\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\|$ measures *outlyingness*, but not as probability weight of a central region. (See also Serfling, 2004, Möttönen, Oja, and Tienari, 1997, Möttönen, Oja, and Serfling, 2005, and Zhou and Serfling, 2007b).

3 Descriptive Measures in $\mathbb{R}^d$

Section 3.1 comments on nonparametric descriptive features, starting with *symmetry*, which is not *measured* but rather *perceived*, and ending with *kurtosis.* Then we look at a few general approaches, emphasizing (Section 3.2) the classical moment and comoment approach and (Section 3.5) a new "L-comoments" approach accommodating heavy-tailed data. Section 3.6 questions how to formulate multivariate L-functionals.

3.1 Nonparametric Descriptive Features – A Brief Look

Symmetry. Approved by *Nature*, symmetry has been a concept in human thought from ancient times: in *aesthetics*, *mathematics and science*, *philosophy*, and *poetry.* In statistics, symmetry is a *hypothesis to be tested*, with a multitude of definitions in the multivariate case. In the setting of nonparametric inference, a broad notion of symmetry is sought. In choosing a multivariate depth or quantile function, its "center" should agree with the "point of symmetry" (if any) of F.

Location. In $\mathbb{R}$ *the mean*, in $\mathbb{R}^d$ the *mean vector.* Or, in $\mathbb{R}$ *the median*, in $\mathbb{R}^d$ *various notions.* Or, in $\mathbb{R}$ *the trimmed mean (an L-functional)*, in $\mathbb{R}^d$ *depth-weighted trimming. Question: how best to formulate multivariate L-functionals in general?*

DISPERSION, SPREAD. In $\mathbb{R}$ *the variance*, in $\mathbb{R}^d$ the *covariance matrix* (under 2nd moment assumptions). Or, in $\mathbb{R}$ *the Gini mean difference*, in $\mathbb{R}^d$ *a Gini covariance matrix* (available under 1st moment assumptions – see §3.5). Or, in $\mathbb{R}$ *the trimmed variance (an L-functional)*, in $\mathbb{R}^d$ *a depth-weighted trimmed covariance matrix*. Or, in $\mathbb{R}$ *the interquartile range*, in $\mathbb{R}^d$ *the volume of a depth-based central region of probability 1/2*. Or, last but not least, in $\mathbb{R}^d$ a *scatter matrix functional* based on the *rank covariance matrix*, e.g., the covariance matrix of the centered rank function vector $\boldsymbol{R}(\boldsymbol{X}, F)$ defined as the gradient of the criterion function defining the *Oja median*.

SKEWNESS OR ASYMMETRY. In $\mathbb{R}$ a *difference of two real-valued location measures*, in $\mathbb{R}^d$ a *difference of two vector-valued location measures*. Or, in $\mathbb{R}$ *the standardized 3rd central moment*, in $\mathbb{R}^d$ *scalar extensions* (details to come). Or, in $\mathbb{R}$ *the 3rd L-moment (a particular L-functional)*, in $\mathbb{R}^d$ *the matrix of 3rd L-comoments (also L-functionals)* – see §3.5. (*Question: is there a multivariate L-functional extension of the univariate 3rd L-moment?*) Alternatively, a *skewness functional*, defined in $\mathbb{R}$ as a *difference of differences of selected quantiles* (Oja, 1981), and in $\mathbb{R}^d$ as a *difference of volumes of depth-based central regions of specified probabilities*.

KURTOSIS, TAILWEIGHT, AND PEAKEDNESS. We may think of *kurtosis* as a location- and scale-free measure of the degree to which the probability mass of a distribution is diminished in the "shoulders" and increased in the center (*peakedness*) or in the tails (*tailweight*). More precisely, it concerns the structure of the distribution in the "middle" of *the region that falls between, and links, the center and the tails*. The "middle" of this connecting region represents the "shoulders" of the distribution. In this way, *kurtosis*, *peakedness* and *tailweight* are regarded as *distinct*, although *interrelated*, features of a distribution. (Some writers treat kurtosis and tailweight as equivalent and inverse to peakedness.)

In the *univariate* case, the "shoulders" are represented sometimes by $\mu \pm \sigma$ and sometimes by *the 1st and 3rd quartiles*. The *moment-based* version of univariate kurtosis, given by $\kappa = E\{(X-\mu)^4\}/\sigma^4 = \mathrm{Var}\{(\frac{X-\mu}{\sigma})^2\} + (E\{(\frac{X-\mu}{\sigma})^2\})^2 = \mathrm{Var}\{(\frac{X-\mu}{\sigma})^2\} + 1$, measures dispersion of $(\frac{X-\mu}{\sigma})^2$ about its mean 1, or equivalently the dispersion of X about the points $\mu \pm \sigma$, which are viewed as the "shoulders". That is, κ measures in a location- and scale-free sense the *dispersion of probability mass away from the shoulders*, toward either the center or the tails or both.

In $\mathbb{R}^d$ this generalizes to $E\{[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^2\}$, the 4th moment of the Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}$, which measures the dispersion of the squared Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}$ about its mean d, or equivalently the dispersion of $\boldsymbol{X}$ about the points on the ellipsoid $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d$, which surface thus comprises the "shoulders" of the distribution. Higher kurtosis arises when probability mass is diminished near the shoulders and greater either near $\boldsymbol{\mu}$ (greater peakedness) or in the tails (greater tailweight) or both.

An alternative "moment-type" approach is given in $\mathbb{R}$ by *the 4th L-moment (a particular L-functional)* and in $\mathbb{R}^d$ *the matrix of 4th L-comoments (also L-functionals)* (details in §3.5). *Question: is there a multivariate L-functional extension of the univariate 4th L-moment?*

Quantile-based approaches have been formulated. In $\mathbb{R}$, a *kurtosis functional*

$$k_F(p) = \frac{v_F(\frac{1}{2}-\frac{p}{2}) + v_F(\frac{1}{2}+\frac{p}{2}) - 2\, v_F(\frac{1}{2})}{v_F(\frac{1}{2}+\frac{p}{2}) - v_F(\frac{1}{2}-\frac{p}{2})}$$

is defined as a curve for $0 \le p < 1$, where $v_F(p) = F^{-1}(\frac{1}{2}+\frac{p}{2}) - F^{-1}(\frac{1}{2}-\frac{p}{2})$. For each p, $k_F(p)$ is a *scaled difference of interquantile ranges*. Extension to $\mathbb{R}^d$ is obtained using the same expression for $k_F(p)$, but now with $v_F(p)$ the *volume* of a *depth-based central region* of probability weight p. Thus $k_F(p)$ is a *scaled difference of differences of volumes of depth-based central regions of specified probabilities*. The boundary of the central region of probability 1/2 represents the "shoulders", separating a "central part" of the distribution, from a complementary "tail part". Hence $k_F(p)$ measures the relative volumetric difference between equiprobable regions within and without the shoulders, which are defined by shifting the probability weight parameter 1/2 of the shoulders by equal amounts $p/2$ toward the center and toward the tails.

Dependence. The classical dependence measure in $\mathbb{R}^d$ is the $d \times d$ *correlation matrix*, available under 2nd moment assumptions. Alternatively, *a Gini correlation matrix* is derived from the Gini covariance matrix and available under only 1st moment assumptions (details in §3.5). Alternatively, *rank correlations* provide another attractive approach that is receiving increasing interest and application.

3.2 Moments (Classical Approach)

Multivariate statistical analysis relies heavily on *moment assumptions of 2nd order and higher*. There is, however, increasing interest in modeling with *heavy tailed distributions*. It is desirable to describe multivariate *dispersion, skewness, and kurtosis*, and to have *correlational analysis*, under just *1st order (or even lower) moment assumptions*. These are goals in *both parametric and nonparametric settings*.

Moments. For univariate X having mean μ and finite central moments $\mu_k = E(X-\mu)^k$, $2 \le k \le K$, the classical scaled central moments for $k \ge 3$ are given by $\psi_k = \mu_k/(\mu_2)^{k/2}$. The cases $k = 3$ and 4, respectively, are the classical *skewness* and *kurtosis* coefficients.

For $\boldsymbol{X}$ in $\mathbb{R}^d$, μ extends to a *vector* $\boldsymbol{\mu}$ and μ_2 to a (covariance) *matrix* $\boldsymbol{\Sigma}$. These play a central role in classical multivariate analysis. *Scalar* notions of multivariate skewness and kurtosis, respectively, are due to Mardia (1970): $E\{[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^3\}$ and $E\{[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^2\}$. Are there *matrix* extensions?

Comoments. A further classical approach involves *comoment matrices* (not well known, however). For a d-vector $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})'$ with marginal means μ_i and central moments $\mu_k^{(i)}$, $k \ge 2$, $i \le i \le d$, the *covariance* of $X^{(i)}$ and $X^{(j)}$ can, of course, be expressed as $\mathrm{Cov}\{X^{(i)}, X^{(j)}\} = \mathrm{Cov}\{X^{(i)} - \mu_i, X^{(j)} - \mu_j\}$. An analogous *coskewness of* $X^{(i)}$ *with respect to* $X^{(j)}$ is defined as $\xi_{3\,[ij]} = \mathrm{Cov}\{X^{(i)} - \mu_i, (X^{(j)} - \mu_j)^2\}$ and a *cokurtosis of* $X^{(i)}$ *with respect to* $X^{(j)}$ by $\xi_{4\,[ij]} = \mathrm{Cov}\{X^{(i)} - \mu_i, (X^{(j)} - \mu_j)^3\}$. In general, a *kth order central comoment matrix* is given by

$$(\xi_{k\,[ij]}) = \left(\mathrm{Cov}\{X^{(i)} - \mu_i, (X^{(j)} - \mu_j)^{k-1}\}\right).$$

This contributes to *pairwise* multivariate analysis. Note that the *covariance operator* is involved in all of these.

For order $k \ge 3$, the comoment matrices are *asymmetric*, and that is *good!* The symmetry when $k = 2$ is merely an artifact of the definition rather than a feature necessarily desired for comoments in general. In any case, for higher orders one can obtain symmetric versions by, for example, signed versions of $\sqrt{\xi_{k\,[ij]}\,\xi_{k\,[ji]}}$, but the ordered pairs $(\xi_{k\,[ij]}, \xi_{k\,[ji]})$ carry greater information while still simple and thus are preferred. We may regard the classical covariance matrix as an unfortunate exception!

The comoments arise in portfolio optimization in finance (Rubinstein, 1973, Christie-David and Chaudhry, 2001, and Jurczenko, Maillet, and Merlin, 2005). Skewness relates to asymmetric volatility and downside risk of a portfolio, kurtosis to high volatility and uncertainty in returns. The covariance of a security and the "market return" measures the security's contribution to portfolio diversification, and coskewness and cokurtosis measure its contributions to overall skewness and kurtosis. In further developing the role of the Capital Asset Pricing Model in financial risk analysis, there is increasing interest in including higher order comoments and, accordingly, increasing concern about *high sensitivity to the more extreme values* in the data. Can a useful notion of kth comoment be defined under just 1st moment assumptions? Can an asymmetric version of the 2nd comoment (covariance) be defined? Can we have *correlation analysis*, either symmetric or asymmetric, with only 1st moment finite? We return to these questions in §3.5.

3.3 Expectations of Volumes (Oja, 1983)

In a landmark paper of great impact, Oja (1983) introduced a quite novel departure from moment methods, a family of alternative series of measures for location, skewness, kurtosis, etc., in $\mathbb{R}^d$. For each $\alpha > 0$, a *location* measure is defined by the vector $\boldsymbol{\mu}_\alpha(F)$ minimizing

$$E\{[\Delta(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d, \boldsymbol{\mu}_\alpha(F))]^\alpha\},$$

where $\Delta(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1})$ denotes the volume of the d-dimensional simplex determined by the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1}$. The case $\alpha = 1$ gives the now well-known and popular *Oja median.*

Associated measures are given for *scatter* by $\sigma_\alpha(F) = \sqrt[\alpha]{E\{[\Delta(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d, \boldsymbol{\mu}_\alpha(F))]^\alpha\}}$, for *skewness* by either $(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ or $E\{\Delta(\boldsymbol{X}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)\}/\sigma_2$, and for *kurtosis* by

$$E\{[\Delta(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d, \boldsymbol{\mu}_2)]^4\}/(E\{[\Delta(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d, \boldsymbol{\mu}_2)]^2\})^2.$$

3.4 Miscellaneous Unsystematic Efforts

These half-baked efforts are too numerous to list!

3.5 Matrices of Univariate L-Functionals

A recent new approach that contributes toward pairwise multivariate methodology is based on *L-comoments* (Serfling and Xiao, 2007), which extend the univariate *L-moments* (Hosking, 1990). Like the central moments, "L-moments" provide a coherent series of measures of spread, skewness, kurtosis, etc, and, like the central comoments, "L-comoments" measure covariation, coskewness, cokurtosis, etc., in some fashion. All of these measures are defined, however, under merely 1st moment assumptions, the L-moments being particular *L-functionals* and the L-comoments being L-functionals in terms of *concomitants.* Sample L-moments and L-comoments are less sensitive to extreme observations than their classical analogues. Also, for all orders including $k = 2$, the L-comoments are *asymmetric.*

Univariate L-moments

Definition. With standard order statistic notation $X_{1:k} \leq X_{2:k} \leq \ldots \leq X_{k:k}$ for a sample of size k, the *kth L-moment* is given by

$$\lambda_k = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} E(X_{k-j:k})$$

In particular, $\lambda_1 = E(X_{1:1})$ (= the *mean*), $\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2})$ (= one-half *Gini's mean difference*), $\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3})$ (measuring *skewness*), and $\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})$ (measuring *kurtosis*).

Key features. *Finiteness for all orders, if mean finite. Determination of the distribution, if mean exists. Representation as expected value of linear function of order statistics, any sample size* $n \geq k$*. Mutual orthogonality of weight functions in L-functional representations of different L-moments. L-statistic structure of sample versions. U-statistic structure of sample versions. Unbiasedness of sample versions.*

Key representations. Let F denote the relevant distribution. An *L-functional representation* for the kth L-moment is given by $\lambda_k = \int_0^1 F^{-1}(t)\, P_{k-1}^*(t)\, dt$, using the *shifted Legendre system* of orthogonal polynomials, $P_k^*(u) = \sum_{j=0}^k p_{k,j}^* u^j$, $0 \leq u \leq 1$, with $p_{k,j}^* = (-1)^{k-j}\binom{k}{j}\binom{k+j}{j}$.

Another important representation expresses λ_k in terms of *covariance*:

$$\lambda_k = \begin{cases} E(X), & k = 1; \\ \mathrm{Cov}\{X, P_{k-1}^*(F(X))\}, & k \geq 2, \end{cases}$$

For example, $\lambda_2 = \mathrm{Cov}(X, 2F(X) - 1)$, the covariance of X and its *centered rank*.

Trimmed versions. One approach (Elamir and Seheult, 2003) to a notion of trimmed L-moment increases the conceptual sample size for defining the kth L-moment from k to $k + t_1 + t_2$ and uses the k order statistics remaining after trimming the t_1 smallest and t_2 largest observations in the conceptual sample. For $(t_1, t_2) \neq (0, 0)$, the "TL-moments" exist under weaker moment assumptions (satisfied by the Cauchy distribution, for example) and eliminate the influence of the most extreme observations. For example, with $(t_1, t_2) = (1, 1)$, the 1st TL-moment is $\lambda_1^{(1,1)} = E(X_{2:3})$, the expected value of the median from a sample of size 3.

A multivariate extension: L-comoments

Definition. As a multivariate extension of L-moments, Serfling and Xiao (2007) simultaneously extend the covariance representation for L-moments and the definition of the kth order central comoment of $X^{(i)}$ with respect to $X^{(j)}$, defining the *kth order L-comoment* of $X^{(i)}$ with respect to $X^{(j)}$ as

$$\lambda_{k\,[ij]} = \mathrm{Cov}(X^{(i)}, P_{k-1}^*(F_j(X^{(j)}))),\ k \geq 2.$$

The case $k = 2$ is in fact already known and studied, as the "*Gini covariance*" (Schechtman and Yitzhaki, 1987, Yitzhaki and Olkin, 1991, and Olkin and Yitzhaki, 1992).

Representation in terms of concomitants. For a bivariate sample $(X_m^{(i)}, X_m^{(j)})$, $1 \le m \le n$, denote the ordered $X_m^{(j)}$-values by $X_{1:n}^{(j)} \le X_{2:n}^{(j)} \le \cdots \le X_{n:n}^{(j)}$. Then (David and Nagaraja, 2003) the $X_m^{(i)}$ paired with $X_{r:n}^{(j)}$ is called the *concomitant* of $X_{r:n}^{(j)}$. Denote it by $X_{[r:n]}^{(ij)}$. It turns out that

$$\lambda_{k\,[ij]} = k^{-1} \sum_{r=0}^{k-1} (-1)^r \binom{k-1}{r} E(X_{[k-r:k]}^{(ij)}).$$

identical to the defining formula for λ_k except with $E(X_{k-r:k})$ replaced by $E(X_{[k-r:k]}^{(ij)})$.

L-correlation. The *L-correlation* of $X^{(i)}$ with respect to $X^{(j)}$ is defined by appropriate scaling of the 2nd L-comoment of $X^{(i)}$ with respect to $X^{(j)}$,

$$\rho_{[ij]} = \frac{\lambda_{2\,[ij]}}{\lambda_2^{(i)}},$$

and thus is defined under a 1st moment assumption on $X^{(i)}$. Note that $\rho_{[ij]}$ and $\rho_{[ji]}$ need not be equal.

Like Pearson correlation, the L-correlation lies in the interval $[-1, 1]$. We have

Proposition 1. $|\rho_{[ij]}| \le 1$, *with equality if* $X^{(i)}$ *is a monotone function of* $X^{(j)}$.

As an L-moment analogue of the decomposition of the variance of a sum in terms of variances of summands, the 2nd L-moment of a sum may be expressed as a weighted sum of the individual 2nd L-moments, in which the (asymmetric) L-correlations arise naturally as coefficients. We have

Proposition 2. *For a sum* $S_n = Y_1 + \cdots + Y_n$ *of univariate* $Y_1, \ldots, Y_n$, *we have (without dependence restrictions)*

$$\lambda_2(S_n) = \sum_{i=1}^{n} \rho_{[Y_i, S_n]}\, \lambda_2(Y_i).$$

Under certain conditions satisfied in many models, the formulas for the L-correlation $\rho_{[ij]}$ and the Pearson correlation ρ_{ij} agree. This permits standard correlational analysis under 2nd order assumptions to be coherently extended down to 1st order. We have

Proposition 3. *Assume (1) linear regression of* $X^{(i)}$ *on* $X^{(j)}$, $E(X^{(i)} \,|\, X^{(j)}) = a + bX^{(j)}$, *and (2) affine equivalence of* $X^{(i)}$ *and* $X^{(j)}$, $X^{(j)} \stackrel{d}{=} \theta + \eta X^{(i)}$. *Then*

$$\rho_{[ij]} = b\eta = \rho_{ij}$$

holds under 2nd moment assumptions, with the first equality valid as well under just 1st moment assumptions.

In particular, for the multivariate normal model, $\rho_{[ij]} = b\eta = (\sigma_{ij}/\sigma_j^2)(\sigma_j/\sigma_i) = \rho_{ij}$. Proposition 3 also applies to certain multivariate Pareto models, for example.

L-comoment matrices. For $k \ge 2$, the *kth order L-comoment matrix* is the $d \times d$ matrix of kth L-comoments taken over the pairs $(X^{(i)}, X^{(j)})$, $1 \le i, j \le d$: $\mathbf{\Lambda}_k = (\lambda_{k\,[ij]})$ (possibly asymmetric).

These matrices exist (finite) under merely 1st moment assumptions on the components of $\boldsymbol{X}$. We call $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Lambda}_3$, and $\boldsymbol{\Lambda}_4$, respectively, the *L-covariance*, *L-coskewness*, and *L-cokurtosis* matrices. In practice, *scaled* versions are used, as with correlation and with the general univariate case.

TRIMMED L-COMOMENTS. Trimmed versions of L-comoments are obtained by extending our definition of L-comoments to a definition of "TL-comoments", as a direct analogue of the extension of univariate L-moments to univariate TL-moments.

L-COMOMENT MATRICES. Packages for computation of L-moments, L-comoments, and trimmed versions are available. Links are given at `www.utdallas.edu/~serfling`.

Alternative approaches to multivariate L-moments

APPROACH 1

Can we define multivariate L-moments as *multivariate L-functionals*? For this purpose, let us rexpress the *L-functional representation* for L-moments in median-oriented form via $u = 2t - 1$, with $Q(u, F) = F^{-1}((1+u)/2)$ and with the *standard* Legendre polynomials $P_k(u) = P_k^*((1+u)/2)$ on $[-1, 1]$, obtaining

$$\lambda_k = \frac{1}{2}\int_{-1}^{1} Q(u, F)\, P_{k-1}(u)\, du,$$

which is meaningful for $d = 1$. Now seek to extend to $\mathbb{R}^d$:

1. Replace the interval $[-1, 1]$ by the unit ball,
2. Replace $Q(u, F)$ by a choice of $\boldsymbol{Q}(\boldsymbol{u}, F)$,
3. Define $P_k(\boldsymbol{u})$ in some sense,
4. Define "product" $\boldsymbol{Q}(\boldsymbol{u}, F)P_{k-1}(\boldsymbol{u})$.

This falls within the problem of defining multivariate L-functionals in general (see §3.6).

APPROACH 2

Perhaps a projection-pursuit approach can be developed. Given $\boldsymbol{X}$ in $\mathbb{R}^d$ with distribution F, fpr each unit vector $\boldsymbol{v}$ project $\boldsymbol{X}$ onto the line in direction $\boldsymbol{v}$. For the univariate distribution $F_{\boldsymbol{v}'\boldsymbol{X}}$ of this projection, denote the k-th L-moment by $\lambda_k(F_{\boldsymbol{v}'\boldsymbol{X}})$. Then construct the corresponding k-th L-moment $\lambda_k(F)$ by integrating $\lambda_k(F_{\boldsymbol{v}'\boldsymbol{X}})$ over all $\boldsymbol{v}$, via polar coordinates. The orthogonality properties survive, and the interpretations survive. Note also that the series $\{\lambda_k(F_{\boldsymbol{v}'\boldsymbol{X}}), k \geq 1\}$ determines $F_{\boldsymbol{v}'\boldsymbol{X}}$, each $\boldsymbol{v}$, and (Rényi, 1952) the collection $\{F_{\boldsymbol{v}'\boldsymbol{X}}$, all unit $\boldsymbol{v}\}$ determines F.

3.6 Multivariate L-Functionals

We wish to make sense of functionals of F defined as weighted integrals of quantiles,

$$L(F) = \int_{\mathbb{B}^{d-1}} \boldsymbol{Q}(\boldsymbol{u}, F)\, J(\boldsymbol{u})\, d\boldsymbol{u}.$$

If $J(\boldsymbol{u})$ is *scalar*, then the product is scalar multiplication and $L(F)$ is a *vector*. In this case, if $J(\boldsymbol{u})$ equals $J_0(\|\boldsymbol{u}\|)$, i.e., is *constant on contours* of $\boldsymbol{Q}$, then this is identical with a form of *depth-weighted* averaging, $\int \boldsymbol{x}\, W(D(\boldsymbol{x}, F))\, d\boldsymbol{R}(\boldsymbol{x}, F)$, which we note is not, however, the same as $\int \boldsymbol{x}\, W(D(\boldsymbol{x}, F))\, dF(\boldsymbol{x})$ except when $d = 1$. If, on the other hand, $J(\boldsymbol{u})$ is a *vector*, then with *inner product* $L(F)$ is *scalar*, and with *Kronecker product* $L(F)$ is *matrix-valued.*

Thus the form of a multivariate L-functional may be *scalar*, *vector*, or *matrix-valued.* All are useful possibilities for investigation in general along with identification of special cases already studied in the literature. Also, more generally, we wish to investigate the above with $J(\boldsymbol{u})\, d\boldsymbol{u}$ replaced by $dK(\boldsymbol{u})$ for suitable measures K on $\mathbb{B}^{d-1}$.

3.7 Multiple Location and Scatter Measures (Oja, 2007)

Another novel *matrix-based* approach to skewness and kurtosis uses two location measures and two scatter matrix measures (Oja talk, this conference).

3.8 The Ultimately Perfected Method (Oja, 2031)

Let us anticipate and look forward to the development of an ultimately perfected approach toward nonparametric descriptive measures (Oja, 2031).

4 Outlier Identification

4.1 Outlyingness Functions in $\mathbb{R}^d$

The idea of outliers has a long tradition

"Whoever knows the ways of Nature will more easily notice her deviations; and, on the other hand, whoever knows her deviations will more accurately describe her way."

– Francis Bacon, 1620

(Borrowed from Billor, Hadi and Velleman, 2000)

"I do not condemn in every case the principle of rejecting an observation, indeed I approve it wherever in the course of observation an accident occurs which in itself raises an immediate scruple in the mind of the observer before he has considered the event and compared it with the other observations. If there is no such reason for dissatisfaction I think that each and every observation should be admitted whatever its quality so long as the observer is conscious that he has taken every care."

– Daniel Bernoulli, 1777

(Borrowed from Finney, 2006)

"In almost every true series of observations, some are found, which differ so much from the others as to indicate some abnormal source of error not contemplated in the theoretical discussions, and the introduction of which into the investigations can only serve ... to perplex and mislead the inquirer."

– B. Pierce, 1852

(Borrowed from Barnett and Lewis, 1994)

But ... in higher dimension, the first problem is *identification* of the "outliers":

> "As Gnanadesikan and Kettenring (1972) remark, a multivariate outlier no longer has a simple manifestation as an observation which 'sticks out at the end' of the sample. The sample has no 'end'! But, notably in bivariate data, we may still perceive an observation as suspiciously aberrant from the data mass ..."
>
> – *Barnett and Lewis, 1994*

In the univariate case detection is easy, and likewise in the bivariate case via scatter plots. But in higher dimension we need *algorithmic approaches*, i.e., *outlyingness functions*. And SPSS offers to do it for you!!! Their advertisement in *Amstat News*, 2007:

> **Quickly find multivariate outliers**
> Prevent outliers from skewing analyses when you use the Anomaly Detection Procedure. This procedure searches for unusual cases based on deviations from similar cases and gives reasons for such deviations. You can flag outliers by creating a new variable. Once you have identified unusual cases, you can further examine them and determine if they should be included in your analyses.

Unabashed, however, let us explore some approaches, starting with our basic notion: Given a cdf F on $\mathbb{R}^d$, an *outlyingness function* $O(\boldsymbol{x}, F)$ provides an associated *center-outward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$ with *higher* values representing greater "outlyingness".

Approaches toward univariate outlyingness functions

"Tail probability" approach. Starting with $D(x, F) = \min\{F(x), 1-F(x)\}$ ($\leq 1/2$), we obtain $O(x, F) = \frac{1}{2} - D(x, F) = \frac{1}{2}\,|2F(x) - 1|$.

"Sign function" (rank) approach. With $S(\cdot)$ the sign function, we define

$$O(x, F) = |ES(x - X)| = |2F(x) - 1| = 1 - 2\min\{F(x), 1 - F(x)\}.$$

"Scaled deviation" approach. With $\mu(F)$ and $\sigma(F)$ given location and spread measures (e.g., Med(F) and MAD(F) as in Mosteller and Tukey, 1977), define

$$O(x, F) = \left| \frac{x - \mu(F)}{\sigma(F)} \right| \quad (< \infty).$$

"M-functional" approach. Let ψ be a nondecreasing function on $\mathbb{R}$ with $\psi(-\infty) < 0$, $\psi(0) = 0$, and $\psi(\infty) > 0$. [The *M-functional* associated with ψ is the solution t of $E_F\psi(X - t) = \int \psi(x - t)dF(x) = 0$.] Then take $O(x, F) = |E_F\,\psi(\frac{X-x}{\sigma(F)})|$. Special cases:

a) For $\psi(x) = x$, we obtain the preceding "scaled deviation" approach.

b) For $\psi(x) = \text{sign}(x) = \pm 1$ according as $x > 0$ or $x \leq 0$, we obtain the preceding "sign function (rank)" approach.

c) A "Huberized" version of a) is given by

$$\psi(x) = x\mathbf{1}\{|x| \leq c\} + c\mathbf{1}\{x > c\} - c\mathbf{1}\{x < c\}.$$

d) A smoothing of a) is given by

$$\psi(x) = \frac{1 - e^{-cx}}{1 + e^{-cx}}.$$

Generalized scaled deviation approach. Let g be a function on $\mathbb{R}$. With $\mu(F)$ and $\sigma(F)$ given location and spread measures, define $O(x, F) = g(\frac{x-\mu(F)}{\sigma(F)})$. Special cases:

a) For $g(x) = |x|$, we obtain the preceding "scaled deviation" approach.

b) $g(x) = x^2$.

c) $g(x) = (1-x^2)^2$, $|x| \leq 1$ (and 0 elsewhere), the Mosteller/Tukey "biweight" (bisquare weight).

d) $g(x) = x\mathbf{1}\{|x| \leq c\} + c\mathbf{1}\{x > c\} - c\mathbf{1}\{x < c\}$, a Huberization of a).

Two general forms. Consider

(I) $O(x, F) = g(\frac{x-\mu(F)}{\sigma(F)})$, with $\mu(F)$ and $\sigma(F)$ location and spread funcftionals defined on F, respectively, and

(II) $O(x, F) = \mu_F(g(\frac{x-X}{\sigma(F)}))$, where $\mu_F(\cdot)$ denotes a location measure taken with respect to the cdf F for random X, and $\sigma(F)$ is a spread functional defined on F.

These two forms *include all of the preceding cases*, some from Zhang (2002) and some new. It is of interest to *study systematically and comprehensively* the behavior of sample versions in this general setting.

Extension to outlyingness functions in $\mathbb{R}^d$

Projection pursuit approach. Given a *univariate* outlyingness function $O_1(\cdot, \cdot)$ and $\boldsymbol{X} \sim F$ on $\mathbb{R}^d$, define

$$O_d(\boldsymbol{x}, F) = \sup_{\|\boldsymbol{u}\|=1} O_1(\boldsymbol{u}'\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{X}}), \ \boldsymbol{x} \in \mathbb{R}^d.$$

For both the *tail-probability* and *sign* $O_1(\cdot, \cdot)$, this leads to the *halfspace (Tukey) depth*. For the Mosteller-Tukey *scaled deviaton* $O_1(\cdot, \cdot)$, this gives the outlyingness function $O_d(\boldsymbol{x}, F)$ of Donoho and Gasko (1992) and the equivalent *projection depth*.

Substitution approach. In the *univariate sign* $O_1(\cdot, \cdot)$, insert the d-dimensional sign function (i.e., unit vector function) $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{x}/\|\boldsymbol{x}\|$, $\boldsymbol{x} \in \mathbb{R}^d$, and obtain the *spatial* outlyingness function, $O_d(\boldsymbol{x}, F) = \|E\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\|$, $\boldsymbol{x} \in \mathbb{R}^d$. In the *univariate scaled deviation* $O_1(\cdot, \cdot)$, substitute multivariate location and spread measures and obtain a *Mahalanobis type* $O_d(\boldsymbol{x}, F)$.

Approaches toward location estimation in $\mathbb{R}^d$ via sample outlyingness

Minimize sample outlyingness (equivalently maximize sample depth). To estimate the maximum-$D(\cdot, F)$ point $\boldsymbol{\theta}$, minimize $O_d(\hat{\boldsymbol{\theta}}, \widehat{F}_n)$ with respect to $\hat{\boldsymbol{\theta}}$. When based on a *projection-pursuit* type of outlyingness function, this is a *minimax* approach. In the univariate case, it includes *M-estimation.*

Use outlyingness-downweighted averages. For a real nonincreasing weight function $w(\cdot)$, define $w_i = w(O_d(\boldsymbol{X_i}, \widehat{F}_n))$ and estimate location by the weighted average

$$\hat{\boldsymbol{\theta}} = \frac{\sum_1^n w_i \boldsymbol{X}_i}{\sum_1^n w_i},$$

which downweights outlying observations. This includes the well-known Mosteller-Tukey-Stahel-Donoho approach.

Location estimation in $\mathbb{R}^d$ via projection-pursuit sample outlyingness: some details

Step 1. Choose a *univariate* sample outlyingness function defined on a data set $\mathbb{Y}_n = (Y_1, \ldots, Y_n)$ in $\mathbb{R}$. In particular, given translation and scale equivariant univariate location and scale statistics $\mu(\cdot)$ and $\sigma(\cdot)$ defined on $\mathbb{Y}_n$, let us take

$$O_{1n}(y, \mathbb{Y}_n) = \left|\frac{y - \mu(\mathbb{Y}_n)}{\sigma(\mathbb{Y}_n)}\right| = \left|\mu\left(\frac{Y_1 - y}{\sigma(\mathbb{Y}_n)}, \ldots, \frac{Y_n - y}{\sigma(\mathbb{Y}_n)}\right)\right|.$$

Example 1 (Mosteller and Tukey, 1977). Use $\mu(\mathbb{Y}_n) = \text{Med}\{Y_1, \ldots, Y_n\}$ along with $\sigma(\mathbb{Y}_n) = \text{Med}\{|Y_1 - \mu(\mathbb{Y}_n)|, \ldots, |Y_n - \mu(\mathbb{Y}_n)|\}$ (or the Tyler, 1994, modification).

Example 2 (Zhang, 2002). For some function $g(\cdot)$, use $\mu(\mathbb{Y}_n) = n^{-1}\sum_{i=1}^n g(Y_i)$.

Example 3. For some function $g(\cdot)$, use $\mu(\mathbb{Y}_n) = \text{Med}\{g(Y_1), \ldots, g(Y_n)\}$,

Step 2. Via projection-pursuit, construct a sample outlyingness function defined on data $\mathbb{X}_n = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ in $\mathbb{R}^d$:

$$O_{dn}(\boldsymbol{x}, \mathbb{X}_n) = \sup_{\|\boldsymbol{u}\|=1} O_{1n}(\boldsymbol{u}'\boldsymbol{x}, \boldsymbol{u}'\mathbb{X}_n),\ \boldsymbol{x} \in \mathbb{R}^d.$$

Example 2.* With $O_{1n}(\cdot, \cdot)$ as in Example 2 using $g(z) = \pm 1$ according as $z \geq 0$ or < 0, i.e., with $O_{1n}(y, \mathbb{Y}_n) = |n^{-1}\sum_1^n \text{sign}(Y_i - y)|$, we obtain the *halfspace (Tukey)* outlyingness: $O_{dn}(\boldsymbol{x}, \mathbb{X}_n)$ $= 1 - 2n^{-1}\inf_{\|\boldsymbol{u}\|=1}\sum_{i=1}^n \mathbf{1}\{\boldsymbol{u}'\boldsymbol{X}_i < \boldsymbol{u}'\boldsymbol{x}\}$.

Example: Univariate case of Example 2. Connection with univariate M-estimation. Let $O_{1n}(\cdot, \cdot)$ be as in Example 2 using for $g(\cdot)$ a ψ function as in classical *M-estimation.* Then minimization of $O_{1n}(y, \mathbb{Y}_n) = |\mu(\frac{Y_1 - y}{\sigma(\mathbb{Y}_n)}, \ldots, \frac{Y_n - y}{\sigma(\mathbb{Y}_n)})|$ with $\mu(\mathbb{Y}_n) = n^{-1}\sum_{i=1}^n \psi(Y_i)$ can be carried out by solving the estimating equation $O_{1n}(\theta, \mathbb{Y}_n) = 0$ to find θ, which corresponds to M-estimation for location. That is, univariate M-estimation may be characterized as a particular case of minimum-outlyingness estimation.

Some questions

Does "outlyingness" of an observation $\boldsymbol{X}_i$ perhaps depend on the problem? The *minimum outlyingness* estimators seem to be for estimation of a *location* measure, i.e., a "center" defined as a *multidimensional median.* Likewise, the *outlyingness-downweighted averages of observations* seem to estimate *location.* That is, the outlyingness (and their corresponding depth functions) discussed so far are *location-oriented*, i.e., oriented to *location inference problems.*

Can "outlyingness" also be oriented to scale, regression, skewness, kurtosis, etc.? We return to this question later and see that one solution is to *define depth and outlyingness in the parameter space.*

4.2 Nonparametric Outlier Identification in $\mathbb{R}^d$

The outlier identification problem

Without loss of generality, let $D(\boldsymbol{x}, F)$ take values within the interval $[0, 1]$. Define "α outlier regions" associated with F through a given depth $D(\boldsymbol{x}, F)$, by

$$out(\alpha, F) = \{\boldsymbol{x} : D(\boldsymbol{x}, F) < \alpha\},\ 0 < \alpha < 1.$$

THE GOAL. For given threshhold α, classify all points $\boldsymbol{x}$ of $\mathbb{R}^d$ as belonging to $out(\alpha, F)$ or not.

THE METHOD. Estimate $out(\alpha, F)$ based on a data set $\mathbb{X}_n$ with sample df $\widehat{F}_n$, by

$$out(\alpha, \widehat{F}_n) = \{\boldsymbol{x} : D(\boldsymbol{x}, \widehat{F}_n) < \alpha\}.$$

THE KEY ISSUE. If $D(\boldsymbol{x}, \widehat{F}_n)$ itself is sensitive to outliers, then so is $out(\alpha, \widehat{F}_n)$ and it cannot serve as a reliable outlier identifier. We need to confine to *robust* choices of $out(\alpha, \widehat{F}_n)$.

Breakdown point criteria for choice of depth function

Masking occurs if outliers are misidentified as *nonoutliers. Masking breakdown* occurs if contaminants in the sample can cause points of *arbitrarily extreme $D(\cdot, F)$-outlyingness* to be classified as *nonoutliers.* The *Masking Breakdown Point* (MBP) is the minimal fraction of contaminants in the total sample which can cause masking breakdown when allowed to be placed *arbitrarily.*

Swamping occurs if nonoutliers become misidentified as *outliers. Swamping breakdown* occurs if contaminants cause classification of nonoutliers as outliers at *arbitrarily extreme $D(\cdot, F)$-outlyingness* levels. The *Swamping Breakdown Point* (SBP) is the minimal fraction of contaminants in the total sample which can cause swamping breakdown when allowed to be placed *arbitrarily.*

MBP and SBP for three depth functions

depth	*MBP*	*SBP*
halfspace	$\alpha/2$	1
spatial	$\alpha/2$	1
projection	1/2	1/2

(Dang and Serfling, 2005)

For the *halfspace* MBP, α is restricted to $(0, 2/3)$, imposing the upper bound $1/3$ on the MBP. For the *spatial* MBP, however, α has range $(0, 1)$, permitting MBP $\uparrow 1/2$.

Influence function criteria for choice of depth function

The halfspace, projection, and spatial depth functions have *bounded* influence functions, under typical conditions. This feature makes outlyingness of a point less sensitive to gross contamination of the data – i.e., less sensitive to the outliers themselves.

Also, the spatial depth function has *smoother* influence function than leading competitors. This feature makes outlyingness of a point less sensitive to small perturbations of the data.

Equivariance criteria for choice of depth function

The desired equivariance is given by

$$Q_{\boldsymbol{AX}+\boldsymbol{b}}\left(\frac{\boldsymbol{Au}}{\|\boldsymbol{Au}\|}\|\boldsymbol{u}\|\right) = \boldsymbol{A}Q_{\boldsymbol{X}}(\boldsymbol{u}) + \boldsymbol{b},\ \boldsymbol{u} \in \mathbb{B}^{d-1},$$

for any affine transformation $\boldsymbol{x} \mapsto \boldsymbol{Ax} + \boldsymbol{b}$ transforming $\boldsymbol{X}$. In terms of a data cloud in $\mathbb{R}^d$, the sample quantile function becomes transformed and re-indexed in the manner prescribed by this equation if the cloud of observations becomes translated, or rescaled, or rotated about the origin, or reflected about a $(d-1)$-dimensional hyperplane through the origin.

For the *spatial* quantile function, however, this holds for *shift*, em orthogonal, and *homogeneous scale* transformations, but fails under *heterogenous scale* transformations. In practical terms, for the spatial quantile function, *the outlyingness of a point is not invariant under heterogeneous scale changes.* Various solutions, for example transformation-retransformation or modification of the objective function, impose their own issues. In some cases it may be satisfactory to *transform the coordinate variables to have similar scales*, at the outset of data analysis.

Summary for three depth functions using three criteria

Depth	*MBP*	*SBP*	*Influence Fcn*	*Equivariance*
halfspace	$\alpha/2$	1	bdd, with jumps	affine
spatial	$\alpha/2$	1	bdd, smooth	orthogonal
projection	1/2	1/2	bdd, smooth	affine

4.3 Depth and Outlyingness in Parameter Space

Depth and outlyingness for arbitrary inference settings

We address the question of how to generalize depth and outlyingness to arbitrary inference settings for data in $\mathbb{R}^d$ (and beyond). Let us suppose that the inference problem using a data set $\mathbb{X}_n$ from a distribution F is defined by a target parameter $\theta(F)$, which may concern location, scale, skewness, kurtosis, or a regression line or hyperplane, for example.

A general approach. Define a *data-based* function of θ giving for each point θ its "outlyingness" in Θ when considered as an estimate of $\theta(F)$. Then take as estimator $\hat{\theta}$ the minimum-outlyingness element of Θ.

Analogy with maximum likelihood. We are accustomed to evaluating a density function $f(x;\theta)$, $x \in \mathcal{X}$, at an observed outcome $x = \mathbb{X}_n$ and then considering the data-based "likelihood function" $L(\theta) = f(\mathbb{X}_n;\theta)$, $\theta \in \Theta$, as an objective function of θ to be maximized in the parameter space Θ.

General formulation in arbitrary data and parameter spaces

Consider a "parameter" space Θ of arbitrary type and "data" $\mathbb{X}_n = (X_1, \ldots, X_n)$ for values X_i in an arbitrary space $\mathcal{X}$.

Goal. Find point $\theta \in \Theta$ which is a "*best fit*" to the set of data points $\mathbb{X}_n$, or optionally to those points remaining after *identification and elimination of outliers*. In short: *identify the "outliers", decide what to do about them, and then complete the "fitting"*.

Maximal depth approach. Employ as the *relevant criterion function* a *data-based* depth function $D(\theta, \mathbb{X}_n)$, $\theta \in \Theta$, defined on the parameter space Θ, with "best fit" equivalent to *maximal depth*.

Issue. Passing from a "*relevant criterion function*" to an *appropriate depth function*.

Example: location estimation in $\mathbb{R}^d$

Here $\Theta = \mathbb{R}^d$ and we consider $\mathbb{R}^d$-valued data $\mathbb{X}_n = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$. A depth function on $\mathbb{R}^d$ $D(\boldsymbol{x}, F)$, $\boldsymbol{x} \in \mathbb{R}^d$, is maximized at a "center" interpreted as a location parameter. Its sample analogue $D(\boldsymbol{x}, \widehat{F}_n)$, $\boldsymbol{x} \in \mathbb{R}^d$, induces by reinterpretation a depth on the possible values for the location parameter, $\widetilde{D}(\theta, \mathbb{X}_n)$, $\boldsymbol{\theta} \in \Theta$ $(= \mathbb{R}^d)$, with maximal $\widetilde{D}$-depth equivalent to maximal D-depth. Here $\widetilde{D} = D$, $\theta = \mathbb{R}^d$, and $\widehat{F}_n \leftrightarrow \mathbb{X}_n$. Thus location estimation by maximizing sample depth on $\mathbb{R}^d$ is equivalent to maximizing sample depth in the parameter space, since the data space and the parameter space are identical.

Example: dispersion estimation

Here $\Theta = \{\text{covariance matrices } \boldsymbol{\Sigma}\}$. For univariate data $\mathbb{Y}_n = (Y_1, \ldots, Y_n)$, define outlyingness $O_{1n}(\sigma, \mathbb{Y}_n)$ of a scale parameter σ by $|\mu(\frac{|Y_i - m(\mathbb{Y}_n)|}{\sigma}, 1 \le i \le n)|$ with $\mu(\cdot)$ and $m(\cdot)$ univariate location statistics, or by $|\mu(\frac{|Y_i - Y_j|}{\sigma}, 1 \le i < j \le n)|$, eliminating the need for $m(\cdot)$. Then, for $\mathbb{R}^d$-valued data $\mathbb{X}_n = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, define outlyingness of a covariance matrix $\boldsymbol{\Sigma}$ via the projection-pursuit approach, obtaining $O_{dn}(\boldsymbol{\Sigma}, \mathbb{X}_n) = \sup_{\|\boldsymbol{u}\|=1} O_{1n}(\sqrt{\boldsymbol{u}'\boldsymbol{\Sigma}\boldsymbol{u}},\ \boldsymbol{u}'\mathbb{X}_n)$. Minimization of $O_{dn}(\boldsymbol{\Sigma}, \mathbb{X}_n)$ with respect to $\boldsymbol{\Sigma}$ then yields a *"maximum dispersion depth estimator"*.

Example: regression

Here $\Theta = \{\text{hyperplanes}\}$. For $\mathbb{R}^d$-valued data $\mathbb{X}_n = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, define the depth of a hyperplane h as *the minimum fraction of observations in* $\mathbb{X}_n$ *whose removal makes* h *a "nonfit"*. We make "fit" and "nonfit" precise in the more general setting considered next.

Example: shape fitting in computational geometry

Shape-fitting problems arise not only in regression but also, for example, in *computer vision*, *machine learning*, and *data mining*. This is the realm of *computational geometry*, with goals such as finding the *point*, *line*, *hypersphere*, *sphere*, or *cylinder* best fitting a point set, i.e., the *minimum radius sphere*, *minimum width cylinder*, *smallest width slab*, *minimum radius spherical shell*, or *minimum radius cylindrical shell* enclosing the point set. The key question: given a *family of shapes* Θ, a set of input points $\mathbb{X}_n$, and a *fitting criterion* $\beta(\theta, \mathbb{X}_n)$ to be minimized, which shape fits best? A major open problem in the general case is *identification and handing of outliers*. Only *ad hoc* solutions in special cases have been given. A *general approach* is needed.

A general depth function approach to shape fitting with outliers

To address the outlier problem, the notion of "depth" has a natural role, but as a function of "fits" rather than of input data points. Above we have discussed depth and outlyingness defined on location, dispersion, and regression parameter spaces.

INFORMAL NOTION. In objective optimization with a space of fits Θ, define the *depth* of $\theta \in \Theta$ as *the minimum fraction of input observations whose removal makes θ a "nonfit"*, i.e., *inadmissible*. Take the "deepest" or "maximal depth" fit as the "best fit". In the *location estimation* problem, for example, points outside the convex hull of the data are "nonfits" and eliminated from consideration.

Formulation of shape fitting criteria

Let the shape-fitting criterion $\beta(\theta, \mathbb{X}_n)$ be a function of *pointwise* functions $\beta(X_i, \theta, \alpha(\mathbb{X}_n))$, $1 \leq i \leq n$, defined on the input data space and Θ, and measuring the closeness of data points to fits θ.

Example A. With the data space given by $\mathbb{R}^d$ and the shape space Θ given by some class of subsets in $\mathbb{R}^d$, the pointwise objective function $\beta(\boldsymbol{x}, \theta) = \min_{\boldsymbol{y} \in \theta} \|\boldsymbol{x} - \boldsymbol{y}\|$ represents the minimum Euclidean distance from $\boldsymbol{x}$ to a point in the set θ. Then $\beta(\theta, \mathbb{X}_n) = \max_{1 \leq i \leq n} \beta(\boldsymbol{X}_i, \theta)$ represents the *maximal distance* from a point in $\mathbb{X}_n$ to the nearest point in θ. Thus, for example, fitting a line to $\mathbb{X}_n$ may be represented as finding the minimum radius enclosing cylinder θ.

*Example A**. Alternatively, in the previous example, replace *maximum* by *sum*: $\beta(\theta, \mathbb{X}_n) = \sum_{1 \leq i \leq n} g(\beta(\boldsymbol{X}_i, \theta))$. With $g(\cdot)$ the *identity function*, this gives the *total distance* from the points of $\mathbb{X}_n$ to θ as the optimization criterion. Thus, for example, fitting a line to $\mathbb{X}_n$ yields the *principal components regression line*, and for *location problems* with θ representing a point, $\beta(\boldsymbol{x}, \theta)$ reduces to $\|\boldsymbol{x} - \boldsymbol{\theta}\|$ and yields the *spatial median* as the optimal fit. With $g(\cdot)$ an *increasing* function and θ now representing a hyperplane, other objective functions arising in regression problems are obtained.

Example B: Clustering. The *k-center problem* involves the objective function of Example A. The *k-median problem* involves the objective function of Example A* with $g(\cdot)$ the identity function. This is a special case of the *facility location problem*.

Example C: Projective Clustering. The *approximate k-flat problem* involves the objective function of Example A, with θ a k-flat in $\mathbb{R}^d$. The *j approximate k-flat problem* involves the maximum-type objective function with the pointwise function $\beta(\boldsymbol{x}, \theta) = \min_{1 \leq i \leq j} \min_{\boldsymbol{y} \in \mathcal{F}_i} \|\boldsymbol{x} - \boldsymbol{y}\|$, where θ is a vector $(\mathcal{F}_1, \ldots, \mathcal{F}_j)$ of j k-flats in $\mathbb{R}^d$.

Data-based depth functions on fits θ

As indicated informally above, we might define the *depth* $D_n(\theta)$ of the fit θ as the *minimum fraction of input data points whose removal makes the fit* θ *"inadmissible"*, i.e., makes θ is uniformly strictly improvable by some other fit $\tilde{\theta}$: $\beta(x, \tilde{\theta}, \alpha(\widetilde{\mathbb{X}}_n)) < \beta(x, \theta, \alpha(\mathbb{X}_n))$ for each x belonging to the set of unremoved data points $\widetilde{\mathbb{X}}_n$. This extends Rousseeuw and Hubert (1999), Mizera (2002), and Zhang (2002), and includes our examples already considered. For a special case of $\beta(\cdot, \cdot)$, the *maximal depth* point generalizes the *halfspace median*.

Alternatively, define the depth of θ by $D_n^*(\theta) = (1 + \sum_{1 \le i \le n} \beta(X_i, \theta, \alpha(\mathbb{X}_n)))^{-1}$. Now, for a special case, the maximal depth point generalizes the *spatial median*.

The key outlier problem in shape fitting

From depth and outlyingness defined on *fits*, pass to

1. *identification of outliers in the input*, i.e., identify those points which cause the fit to be significantly less than "optimal",
2. *optional removal* of selected outliers,
3. *optimal fit* to the remaining points.

One approach towards a solution:

residuals analysis

A general "residuals approach" for depth-based outlier identification and handling in "shape fitting" problems

1. Obtain an initial *robust* fit θ_n^* as a *maximum depth fit*, employing for example a suitable version of either $D_n(\theta)$ or $D_n^*(\theta)$.
2. Define an *outlyingness measure* for each input data point, using the *outlyingness function* induced on the input space via $O_n(x) = \beta(x, \theta_n^*, \alpha(\mathbb{X}_n))$.
3. Set a threshold for *removal of outliers* among $X_1, \ldots, X_n$ and after optional removal of outliers *optimize using the reduced input*, thereby obtaining a *robust optimal fit* θ_n.
4. Carry out *confirmatory residual analysis* with θ_n as reference point.
5. *Iterate* these steps until a satisfactory "optimum" is attained.

5 Some Open Issues and Directions

Project 1: Unification of existing methods in nonparametric multivariate analysis and systematic formulation of new ones

Quantile, *rank*, *depth* and *outlyingness* functions provide very different methodologies arising from quite distinct points of view. Each approach has independently developed its special domain of application, e.g.

quantiles: median, bagplot, descriptive statistics
ranks: test statistics
depths: volume functional, scale curve, depth-weighted trimmed means
outlyingness: outlier identification

But, as we have seen, these functions are all closely interrelated. *Systematic exploitation of these connections can unify existing methods and also lead to new ones.*

Project 2: Formulation of depth functions on function spaces

Applications such as *nonparametric curve estimation* or *functional MRI*, for example, involve data in *function spaces*. How to characterize the *median* of the data? How to characterize the *central half* of the data? Or central 90%? How to define *depth* and *outlyingness* of data in a function space?

One approach. Define and investigate *spatial* depth in a Banach space $\mathcal{B}$, using the *"sign" function* $S(x) = \frac{x}{\|x\|}$, $x \in \mathcal{B}$.

Project 3: More generally, formulation of depth and quantiles on abstract data spaces

?

Project 4: More elaborate criterion functions for shape fitting

Recalling details from §4.3, let the shape-fitting criterion $\beta(\theta, \mathbb{X}_n)$ be a function of *setwise* functions defined on the input data space and Θ by $\beta(\{X_{i_1}, \ldots, X_{i_m}\}, \theta, \alpha(\mathbb{X}_n))$, measuring closeness of a *set* $\{X_{i_1}, \ldots, X_{i_m}\}$ of data points to a given fit θ, for all $\binom{n}{m}$ m-sets $\{X_{i_1}, \ldots, X_{i_m}\}$. For example, $\beta(\cdot, \theta, \alpha(\mathbb{X}_n))$ might depend upon $\{X_{i_1}, \ldots, X_{i_m}\}$ through $h(X_{i_1}, \ldots, X_{i_m})$ for some "kernel" h as arises in U-statistic applications. This involves *multivariate U-quantiles* (Zhou and Serfling, 2007a,b). In *location estimation*, for example, let h take the m-wise average of arguments (for $m = 2$, a Hodges-Lehmann approach).

Project 5: Formulation of multivariate L-functionals and formal theory of sample versions

Pursue the details laid out in §3.6.

Project 6: Improved equivariance properties for spatial quantiles

As noted in §4.2, the spatial quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$ is not fully affine equivariant, and likewise for the sample version $\boldsymbol{Q}_n(\boldsymbol{u})$. Modified sample versions produced via *transformation-retransformation* are fully affine equivariant (Chakraborty, Chaudhuri, and Oja, 1998, Randles, 2000, Chakraborty, 2001, and Hettmansperger and Randles, 2002). It estimates, however, not the population spatial quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$ but rather a *random* quantile function $\boldsymbol{Q}_n^{(\mathrm{TR})}(\boldsymbol{u}, F)$ which depends not only upon F but also upon an arbitrary subset of $d+1$ observations selected from the sample. Of interest is an investigation of properties of $\boldsymbol{Q}_n^{(\mathrm{TR})}(\boldsymbol{u}, F)$: *its relationship to* $\boldsymbol{Q}(\boldsymbol{u}, F)$, *its relationship to* $\boldsymbol{Q}_n(\boldsymbol{u})$, *its robustness*, *its performance in outlier identification*, etc.

Another approach to achieve an affinely equivariant spatial quantile function is to modify the objective function formulation of $\boldsymbol{Q}(\boldsymbol{u}, F)$ by replacing $\|\boldsymbol{X} - \boldsymbol{\theta}\|$ in the objective function by $\sqrt{(\boldsymbol{X} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\theta})}$, with $\boldsymbol{\Sigma}$ the covariance matrix of $\boldsymbol{X}$. This extends an approach suggested by Isogai (1985) and Rao (1988) in treating the spatial median (see also Zuo, 2004). Of course, one may question whether the resulting coordinate system produced by so standardizing the coordinate variables has appeal from a geometric standpoint. In any case, practical implementation requires a consistent and affine equivariant estimator of $\boldsymbol{\Sigma}$, preferably one that is also robust.

What are the possibilities for development of an **affinely equivariant version of the spatial quantile function** *that retains all of the favorable properties of the original version?*

Project 7: Studies in L-moments and L-comoments

- Extent to which the L-moments, $k \geq 1$, and L-comoments, $k \geq 2$, characterize a bivariate distribution.
- Asymptotics under moments less than 2.
- L-moments and L-comoments approach for stationary multivariate time series analysis.
- L-moments and L-comoments approach for multivariate test procedures (e.g., testing the hypothesis of "central symmetry").
- Simulation studies of small sample performance.
- Effectiveness of the L-comoments and related coefficients as *data summaries*.
- Multivariate extension of univariate regional frequency analysis (Hosking and Wallis, 1997).
- Trimmed L-comoments and other trimming notions.
- Comparisons with competing estimators.

Project 8: Robustness studies for sample versions

BP's and IF's for everything discussed! And other robustness criteria ...

Project 9: Asymptotic and bootstrap results for sample versions

... for everything discussed!

**how to view nine projects in a
"big picture"**

1								
					8			
						9		
				2				
								7
		5						
			4					
	6							
							3	

Miscellaneous other projects ...

See **http://www.utdallas.edu/∼serfling** for recent efforts with potential continuations ... for anyone interested ... *welcome!*

In development: Serfling and Zuo (2008), *Depth and Quantile Functions in Nonparametric Multivariate Analysis.* (Please send your input!)

Even better, *keep an eye on Hannu's website!*

Many other directions and connectons to be pursued:

> *jackknife, smoothing, maxbias curves, descriptive measures beyond* $\mathbb{R}^d$*, PCA, clustering, semiparametric inference, computational issues, mixtures, data mining, regression quantiles, ...*

Acknowledgments

The author thanks G. L. Thompson, Hannu Oja, Ron Randles, and many others including anonymous commentators, for very thoughtful and helpful remarks. The author also especially thanks Hannu and the organizers of *Tilastopäivät 2007* for the kind opportunity to participate and present ideas. Finally, support by the USA National Science Foundation Grants DMS-0103698 and CCF-0430366 is gratefully acknowledged.

References

[1] Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd edition. John Wiley & Sons, New York.

[2] Billor, N., Hadi, A. S., and Velleman, P. F. (2000). BACON: Blocked Adaptive Computationally-Efficient Outlier Nominators. *Computational Statistics & Data Analysis* **34** 279–298.

[3] Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics* **53** 380–403.

[4] Chakraborty, B., Chaudhuri, P., and Oja, H. (1998). Operating transformation and retransformation on spatial median and angle test. *Statistica Sinica* **8** 767–784.

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

[6] Christie-David, R. and Chaudhry, M. (2001). Coskewness and cokurtosis in futures markets. *Journal of Empirical Finance* **8** 55–81.

[7] Dang, X. and Serfling, R. (2005). Nonparametric depth-based outlier identifiers, and robustness properties. Preprint.

[8] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*, 3rd Edition. Wiley, New York.

[9] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estinates based on halfspace depth and projected outlyingness. *Annals of Statistics* **20** 1803–1827.

[10] Dudley, R. M. and Koltchinskii, V. I. (1992). The spatial quantiles.

[11] Elamir, E. A. and Seheult, A. H. (2003). Trimmed L-moments. *Computational Statistics and Data Analysis* **43** 299–314.

[12] Finney, D. J. (2006). Calibration guidelines challenge outlier practices. *The American Statistician* **60** 309–314.

[13] Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28** 81–124.

[14] Hettmansperger, T. P. and Randles, R. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860.

[15] Hosking, J. R. M. (1990). L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society*, *Series B* **52** 105–124.

[16] Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

[17] Isogai, T. (1985). Some extension of Haldane's multivariate median and its application. *Annals of Institute of Statistical Mathematics* **37** 289–301.

[18] Jurczenko, E., Maillet, B., and Merlin, P. (2005). Hedge funds portfolio selection with higher-order moments: a non-parametric mean-variance-skewness-*kurtosis* efficient frontier. working paper.

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

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

[21] Mizera, I. (2002). On depth and deep points: a calculus. *Annals of Statistics* **30** 1681–1736.

[22] Mosteller, C. F. and Tukey, J. W. (1977). *Data Analysis and Regression.* Addison-Wesley, Reading, Mass.

[23] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics* **5** 201–213.

[24] Möttönen, J., Oja, H., and Serfling, R. (2005). Multivariate generalized spatial signed-rank methods. *Journal of Statistical Research* **39** 25–48.

[25] Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Annals of Statistics* **25** 542–552.

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

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

[28] Oja, H. (2007). Multivariate data analysis based on two location vectors and two scatter matrices. Presentation, *Tilastopäivät 2007*.

[29] Oja, H. (2031). Presentation, *Tilastopäivät 2031*.

[30] Olkin, I. and Yitzhaki, S. (1992). Gini regression analysis. *International Statistical Review* **60** 185–196.

[31] Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Assocation* **95** 1263–1268.

[32] Rao, C. R. (1988). Methodology based on the L_1 norm in statistical inference. *Sankhyā Series A* **50** 289–313.

[33] Rényi, A. (1952). On projections of probability distributions. *Acta Math. Acad. Sci. Hung.* **3** 131–142.

[34] Rousseeuw, P. J. and Hubert, M. (1999). Regression depth (with discussion). *Journal of the American Statistical Association* **94** 388–433.

[35] Rubinstein, M. E. (1973). The fundamental theorem of parameter-preference security valuation. *Journal of Financial and Quantitative Analysis* **8** 61–69.

[36] Schechtman, E. and Yitzhaki, S. (1987). A measure of association based on Gini's mean difference. *Communications in Statistics – Theory and Methods* **16** 207–231.

[37] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York.

[38] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference* **123** 259–278.

[39] Serfling, R. (2006). Depth functions in nonparametric multivariate inference. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications* (R. Y. Liu, R. Serfling, and D. L. Souvaine, eds.), pp. 1–16. American Mathematical Society DIMACS Book Series, Volume 72, 2006.

[40] Serfling, R. and Xiao, P. (2007). A contribution to multivariate L-moments: L-comoment matrices. *Journal of Multivariate Analysis*, 2007, to appear

[41] Serfling, R. and Zuo, Y. (2008) *Depth and Quantile Functions in Nonparametric Multivariate Analysis.* In preparation for Springer.

[42] Tyler, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Annals of Statistics* **22** 1024–1044.

[43] Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proceedings of National Academy of Science USA* **97** 1423–1426.

[44] Yitzhaki, S. and Olkin, I. (1991). Concentration indices and concentration curves. In *Stochastic Orders and Decision under Risk* (K. Mosler and M. Scarsini, eds.), pp. 380–392. IMS Lecture Notes – Monograph Series, Volume 19, Hayward, California.

[45] Zhang, J. (2002). Some extensions of Tukey's depth function. *Journal of Multivariate Analysis* **82** 134–165.

[46] Zhou, W. and Serfling, R. (2007a). Multivariate spatial U-quantiles: a Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator. *Journal of Statistical Planning and Inference*, to appear.

[47] Zhou, W. and Serfling, R. (2007b). Generalized multivariate rank test statistics via spatial U-quantiles. *Statistics and Probability Letters*, to appear.

[48] Zuo, Y. (2004). Robustness of weighted L^p-depth and L^p-median. *Allgemeines Statistisches Archiv* **86** 1–20.

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