Central regions and dependency[1]

by Karl Mosler, Universität zu Köln[2]

Revised December 2002

Abstract

The paper introduces an approach to the ordering of dependence which is based on central regions. A d-variate probability distribution is described by a nested family of sets, called *central regions*. Those regions are affine equivariant, compact and star-shaped and concentrate about a properly defined center. They can be seen as level sets of a *depth function*. Special cases are Mahalanobis, zonoid, and likelihood regions. A d-variate distribution is called more dependent than another one if the volume of each central region is smaller with the first distribution. This dependence order is characterized by an inequality between determinants of certain parameter matrices if either (i) F and G are arbitrary distributions and the central regions are Mahalanobis or (ii) F and G belong to an elliptical family of distributions and the central regions are arbitrary. If the regions are zonoid regions, the dependence order implies the ordering of lift zonoid volumes. Alternatively, the dependence order is applied to the copulae of the given distributions. Generalized correlation indices are proposed which are increasing with the dependence orders.

Keywords: Dependence order, generalized correlation, lift zonoid volume, data depth, trimmed regions.

1 Introduction

Modeling and measuring dependency is an important task in many fields of applied probability and statistics, e.g. in analyzing actuarial

[1]Paper presented to IWAP2002, Caracas, January 2002. Thanks are to Alfred Müller and Marco Scarsini for discussions. I thank Richard Hoberg for computing assistance and a referee for his very helpful remarks.

[2]Karl Mosler, Seminar für Wirtschafts- und Sozialstatistik, Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany, mosler@statistik.uni-koeln.de

or financial risks. Various partial orders, indicating that one distribution is more dependent than another one, have been developed in the literature (Block et al., 1990; Joe, 1997; Müller and Stoyan, 2002). In this paper we investigate a new approach to the ordering of dependence which is based on the comparison of central regions.

A probability distribution in d-space can be described by central regions, that is a family of nested sets which include a properly defined center and whose size and shape reflect the scale and dependency of the distribution. For example, if the distribution is starshaped unimodal (Dharmadhikari and Joag-Dev, 1988) and Lebesgue continuous, each density level set forms a central region, and the center is the mode.

In this paper we consider central regions that are defined by an affine invariant depth function like the Mahalanobis depth, the zonoid depth, or the halfspace depth. The zonoid depth has been investigated by Koshevoy and Mosler (1997b). For a unified investigation of various other depths, see Zuo and Serfling (2000a). In Section 2 a general definition of central regions is given which relates to proper notions of depth. In particular, Mahalanobis central regions and zonoid central regions are considered. Section 3 investigates an order which rests on the comparison of central region volumes. Zuo and Serfling (2000b) have introduced this order as an order of *scatter* of multivariate distributions and investigated it for halfspace trimmed regions. Here we use the order with general central regions as an order of *dependency* of multivariate distributions. Our dependence order coincides with an inequality between certain parameter matrices if either (i) F and G are arbitrary distributions and the central regions are Mahalanobis or (ii) F and G belong to some elliptical family of distributions and the central regions are arbitrary. Special indices of dependence are obtained by choosing the volume of a single central region or by aggregating such volumes. When the depth is the zonoid depth, the dependence order implies the ordering of lift zonoid volumes. In Section 4 the same approach is applied to the copulae of the given distributions instead of the distributions themselves, which makes the ordering invariant to increasing transforms. Section 5 contains final remarks.

2 Central regions and depth

Consider a set $\mathcal{P}$ of d-variate probability distributions and a family $\mathcal{R}$ of functions D_α that map $F \in \mathcal{P}$ to a measurable set $D_\alpha(F) \subset \mathbb{R}^d$,

$0 < \alpha \le \alpha_1$. Denote by $F_{\boldsymbol{X}}$ the distribution of a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$.

Definition 1 (Family of central regions)
$\mathcal{R} = (D_\alpha)_{\alpha \in]0,\alpha_1]}$ *is called a* family of starshaped central regions *if it is nested,* $D_\alpha(F) \subset D_\beta(F)$ *for* $\alpha > \beta$ *and* $F \in \mathcal{P}$*, and if, for every* α*,* D_α *is*

- *affine equivariant,* $D_\alpha(F_{\boldsymbol{XA}+\boldsymbol{c}}) = D_\alpha(F_{\boldsymbol{X}})\boldsymbol{A} + \boldsymbol{c}$ *for any* $d \times d$ *matrix* $\boldsymbol{A}$ *having full rank and any point*[3] $\boldsymbol{c} \in \mathbb{R}^d$*,*
- *bounded,*
- *closed,*
- *starshaped*[4] *about every point in* D_{α_1}*.*

In addition, $D_0(F)$ is defined as the convex hull of the support of F. A family of starshaped central regions $\mathcal{R}$ gives rise to a *depth function*,

$$\text{depth}(\boldsymbol{y}|F) = \max\{\alpha : \boldsymbol{y} \in D_\alpha\}\,, \quad F \in \mathcal{P}, \quad \boldsymbol{y} \in \mathbb{R}^d\,,$$

which is

- affine invariant, $\text{depth}(\boldsymbol{yA} + \boldsymbol{c}|F_{\boldsymbol{XA}+\boldsymbol{c}}) = \text{depth}(\boldsymbol{y}|F_{\boldsymbol{X}})$,
- vanishing at infinity,
- upper semi-continuous,
- decreasing on rays from any $\boldsymbol{y}_0 \in \mathit{argmax}\, \text{depth}(\boldsymbol{y}|F)$.

On the reverse, a depth function that satisfies these restrictions determines a family of starshaped central regions,

$$D_\alpha(F) = \{\boldsymbol{y} : \text{depth}(\boldsymbol{y}|F) \ge \alpha\}\,.$$

See Zuo and Serfling (2000a) and Dyckerhoff (2002) for inquiries into general notions of depth.

---

[3]Points in $\mathbb{R}^d$ are rows.

[4]The set $S \subset \mathbb{R}^d$ is *starshaped about* some given *point* $\boldsymbol{x} \in \mathbb{R}^d$ if for every $\boldsymbol{y} \in S$ the straight line connecting $\boldsymbol{x}$ and $\boldsymbol{y}$ lies in S. Note that S is convex iff S is starshaped about every $\boldsymbol{x} \in S$.

2.1 Special central regions and depths

Let F be a given d-variate probability distribution. Many special depth notions have been developed. Among the most useful are the Mahalanobis, the halfspace, and the zonoid depth.

2.1.1 Mahalanobis regions

For $\boldsymbol{y} \in \mathbb{R}^d$ and a distribution F having second moments, define the *Mahalanobis depth* (e.g. Liu et al. (1999))

$$\text{depth}^{Mah}(\boldsymbol{y}|F) = \frac{1}{1 + (\boldsymbol{y} - \mu_F)\Sigma_F^{-1}(\boldsymbol{y} - \mu_F)'},$$

where μ_F and Σ_F are the expectation vector and the covariance matrix of F. The *Mahalanobis family* of central regions consists of ellipsoids

$$D_\alpha^{Mah}(F) = \left\{ \boldsymbol{y} \,:\, (\boldsymbol{y} - \mu_F)\Sigma_F^{-1}(\boldsymbol{y} - \mu_F)' \leq \frac{1}{\alpha} - 1 \right\}, \quad 0 < \alpha \leq 1\,.$$

The most central Mahalanobis region, with $\alpha = 1$, consists of the expectation only, which is the unique deepest point.

Example 1 (Bivariate normal distribution) 50 observations have been simulated from a bivariate normal distribution with standard marginals and correlation $\rho = 0.5$. Figure 1 exhibits the Mahalanobis regions for $\alpha = 0.1, 0.2, \ldots, 0.9$. Figure 2 does the same for 500 bivariate normal observations.

Example 2 (Bivariate exponential distribution) Figure 3 shows Mahalanobis regions for 50 bivariate observations simulated from a Marshall-Olkin distribution. Note that (X_1, X_2) has a *Marshall-Olkin distribution* if $(X_1, X_2) = (\min\{Z_1, Z_3\}, \min\{Z_2, Z_3\})$, and Z_1, Z_2, Z_3 are exponentially distributed, $Z_i \sim Exp(\lambda_i)$, and independent. Then X_1 and X_2 have exponential marginal distributions, $X_1 \sim Exp(\lambda_1 + \lambda_3)$ and $X_2 \sim Exp(\lambda_2 + \lambda_3)$. The random variables X_1 and X_2 are dependent and their correlation coefficient amounts to $\rho = \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$.

It can be seen from the definition that, for given α, the Mahalanobis α-region is *consistent*: The sample version converges in Hausdorff distance to the Mahalanobis α-region of the underlying distribution. This is also illustrated in Figures 1–2.

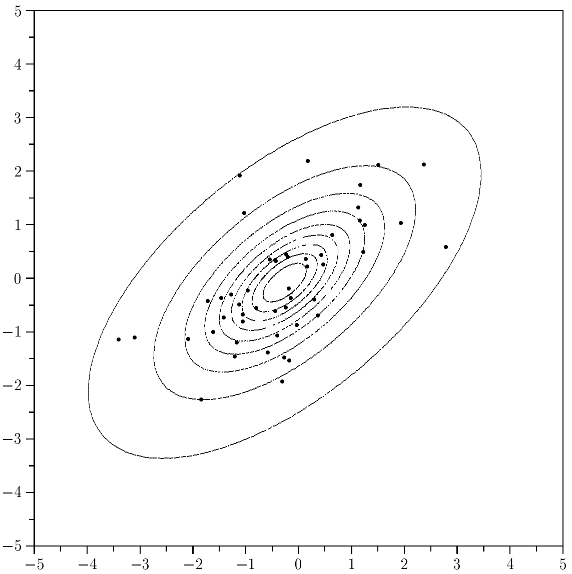

Figure 1: Mahalanobis regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 50$ bivariate normal observations, $(X_1, X_2) \sim N(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho = 0.5)$.

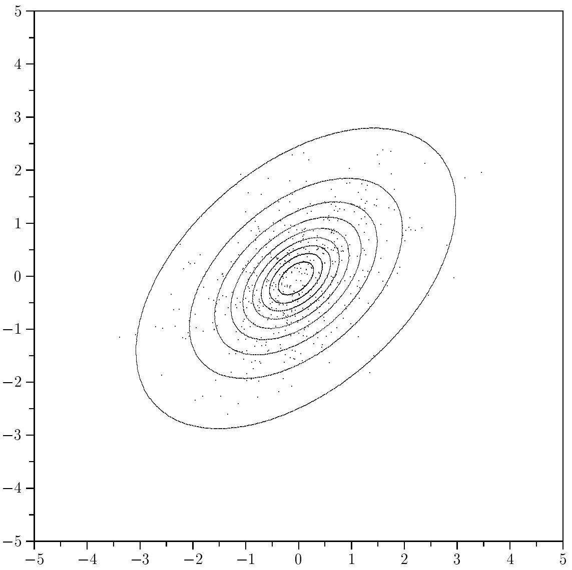

Figure 2: Mahalanobis regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 500$ bivariate normal observations, $(X_1, X_2) \sim N(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho = 0.5)$.

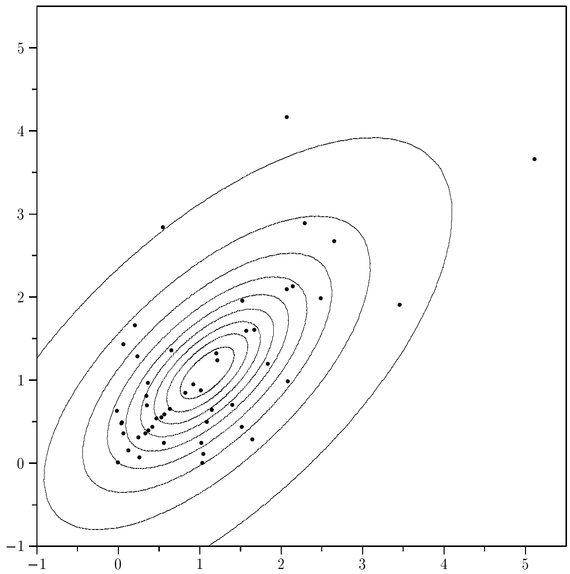

Figure 3: Mahalanobis regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 50$ bivariate exponential observations (Marshall–Olkin distribution), $(X_1, X_2) = (\min\{Z_1, Z_3\}, \min\{Z_2, Z_3\})$, Z_1, Z_2 and Z_3 independently $Exp(1/2)$–distributed.

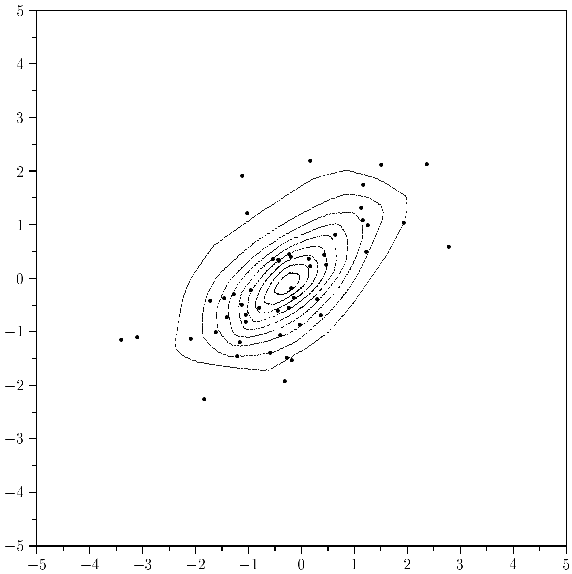

Figure 4: Zonoid regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 50$ bivariate normal observations, $(X_1, X_2) \sim N(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho = 0.5)$.

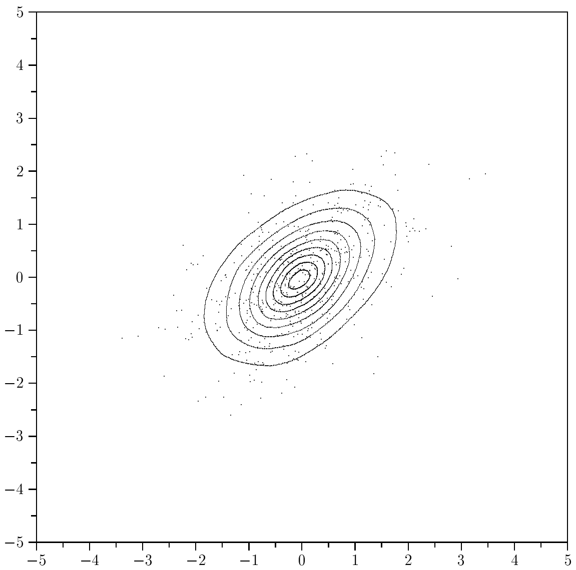

Figure 5: Zonoid regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 500$ bivariate normal observations, $(X_1, X_2) \sim N(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \rho = 0.5)$.

2.1.2 Zoonid regions

If F has finite expectation,

$$D_\alpha^{zon}(F) = \left\{ \frac{1}{\alpha} \int_{\mathbb{R}^d} \boldsymbol{x}\, g(\boldsymbol{x})\, dF(\boldsymbol{x}) \ : \ g : \mathbb{R}^d \to [0,1] \ \text{ measurable and } \int_{\mathbb{R}^d} g(\boldsymbol{x})\, dF(\boldsymbol{x}) = \alpha \right\}, \tag{1}$$

$0 < \alpha \leq 1$, are the *zonoid central regions* of F. As with the Mahalanobis regions, the zonoid region at $\alpha = 1$ is a singleton that contains the expectation as the unique deepest point.

If $F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}$ is an empirical distribution giving probability $\frac{1}{n}$ to each of (not necessarily different) points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$, the zonoid central regions (1) specialize to

$$D_\alpha^{zon}(F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}) = \left\{ \sum_{i=1}^n \lambda_i \boldsymbol{x}_i : \sum_{i=1}^n \lambda_i = 1, \ \ 0 \leq \lambda_i \leq \frac{1}{n\alpha} \text{ for all } i \right\}.$$

In this case, $D_\alpha^{zon}(F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n})$ equals the convex hull of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ when $0 < \alpha \leq \frac{1}{n}$. For larger α the zonoid region may be seen as a 'restricted

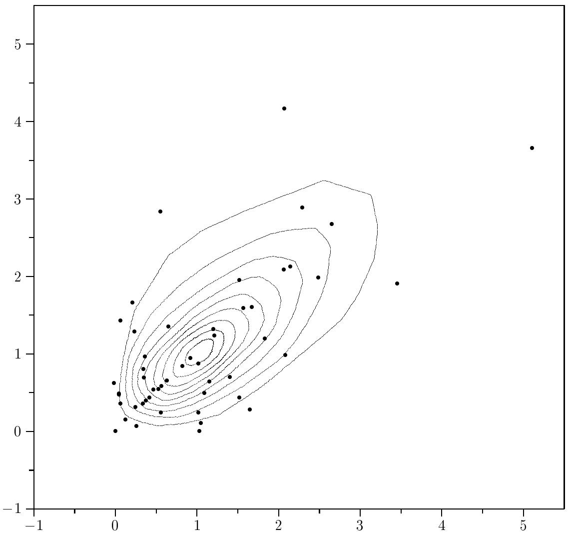

Figure 6: Zoonid regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 50$ bivariate exponential observations (Marshall–Olkin distribution), $(X_1, X_2) = (\min\{Z_1, Z_3\}, \min\{Z_2, Z_3\})$, Z_1, Z_2 and Z_3 independently $Exp(1/2)$–distributed.

convex hull', that is, all convex combinations in which the weight of each $\boldsymbol{x}_i$ is bounded above by $\frac{1}{n\alpha}$.

Figures 4 and 5 exhibit zonoid regions for 50 (500) bivariate normal observations (Example 1) with correlation 0.5. Zonoid regions for 50 (500) bivariate observations from a Marshall-Olkin distribution (Example 2) are seen in Figures 6 and 7.

The zonoid central regions are convex, compact sets. They depend continuously on α and F (Koshevoy and Mosler, 1997a).

These two (and any other) depths are related to some *center* which consists of the deepest point or points. Note that they do not assume unimodality nor is their use restricted to unimodal distributions.

Observe that, for a distribution which is uniform on the unit sphere, the zonoid central regions are *balls around the origin;* the same applies for every spherical distribution. A family of central regions which, for any spherical distribution, are balls around the origin, is said to satisfy the 'balls under sphericity' condition. The condition holds for Mahalanobis regions and for virtually all other special notions of depth based central regions that have been proposed in the literature.

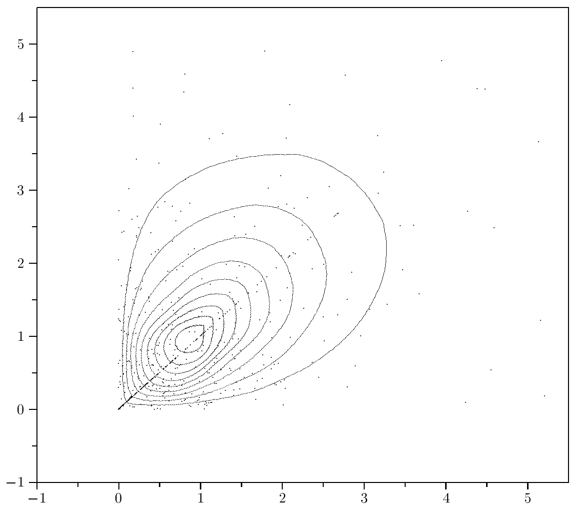

Figure 7: Zonoid regions ($\alpha = 0.1, \ldots, 0.9$) for $n = 500$ bivariate exponential observations (Marshall–Olkin distribution), $(X_1, X_2) = (\min\{Z_1, Z_3\}, \min\{Z_2, Z_3\})$, Z_1, Z_2 and Z_3 independently $Exp(1/2)$–distributed.

An important question is, whether a family of depth central regions characterizes a distribution in a unique way. The answer is clearly 'no' for the Mahalanobis regions, but 'yes' for zonoid regions and arbitrary distributions that have finite expectation. It is also 'yes' for other notions of central regions and some restricted set of distributions, like the *halfspace regions* (which are based on the *halfspace depth*) and distributions that are either discrete or continuous with compact support; see, e.g., Mosler (2002, Sec 4.3).

3 Dependence ordering by volumes

The basic problem of dependence ordering is: Given two d-variate distributions that have the same univariate marginals, which one, if any, is more dependent than the other? In other words, given d random variables, a dependence order specifies a condition under which the variables are considered to be more or less dependent. In this section the volume of a central region is considered as a measure of dependency and a dependence ordering among distributions is defined by comparing all such volumes.

The idea is illustrated in Table 1. For a bivariate normal distribution with standard marginals and correlation coefficient $\rho = 0.0, 0.1, \ldots, 0.9$, Table 1 contains the volumes of Mahalanobis and zonoid regions. The regions are based on 500 observations. The volumes differ between the Mahalanobis and the zonoid depth, but both volumes decrease monotonically when the correlation ρ increases.

Mahalanobis region volumes										
α	Correlation ρ									
0.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.10	27.20	27.07	26.64	25.94	24.91	23.51	21.69	19.35	16.22	11.76
0.20	12.07	12.01	11.82	11.50	11.05	10.43	9.62	8.58	7.20	5.21
0.30	7.03	7.00	6.89	6.71	6.44	6.07	5.61	5.00	4.19	3.04
0.40	4.51	4.49	4.42	4.30	4.13	3.90	3.60	3.21	2.69	1.95
0.50	3.00	2.99	2.94	2.86	2.75	2.59	2.39	2.13	1.78	1.30
0.60	2.00	1.99	1.96	1.90	1.83	1.72	1.59	1.42	1.19	0.86
0.70	1.28	1.27	1.25	1.22	1.17	1.11	1.02	0.91	0.76	0.55
0.80	0.74	0.74	0.73	0.71	0.68	0.64	0.59	0.53	0.44	0.32
0.90	0.33	0.33	0.32	0.31	0.30	0.28	0.26	0.23	0.19	0.14
Zonoid region volumes										
α	Correlation ρ									
	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.10	9.00	8.96	8.82	8.59	8.25	7.80	7.20	6.42	5.38	3.90
0.20	5.65	5.63	5.54	5.39	5.17	4.88	4.51	4.02	3.37	2.45
0.30	3.88	3.87	3.80	3.70	3.56	3.36	3.10	2.76	2.32	1.68
0.40	2.70	2.69	2.64	2.57	2.47	2.33	2.15	1.92	1.61	1.16
0.50	1.83	1.83	1.80	1.75	1.68	1.59	1.47	1.31	1.09	0.79
0.60	1.19	1.19	1.17	1.14	1.09	1.03	0.95	0.85	0.71	0.52
0.70	0.70	0.70	0.70	0.67	0.65	0.61	0.56	0.50	0.42	0.30
0.80	0.35	0.34	0.34	0.33	0.32	0.30	0.28	0.24	0.21	0.15
0.90	0.11	0.11	0.10	0.10	0.10	0.09	0.08	0.07	0.06	0.05

Table 1: Volumes of Mahalanobis and zonoid regions for $n = 500$ bivariate normal observations with $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$ and several ρ and α.

Let $\mathcal{R}$ be a family of starshaped central regions and consider two d-variate distributions F and G which have the same univariate marginals.

Definition 2 (Central regions dependence order)
Distribution G is more dependent *than distribution F with respect to $\mathcal{R}$, $F \preceq_{\mathcal{R}} G$, if*

$$\mathrm{vol}_d D_\alpha(F) \geq \mathrm{vol}_d D_\alpha(G) \quad \textit{for every } \alpha\,.$$

Theorem 1 *If $\mathcal{R}$ is the Mahalanobis family, then*

$$F \preceq_{\mathcal{R}} G \qquad \Leftrightarrow \qquad \det \mathrm{P}_F \geq \det \mathrm{P}_G\,, \tag{2}$$

where P_F and P_G are the correlation matrices of F and G.

Proof. The Mahalanobis central region $D_\alpha(F)$ of a distribution F is an ellipsoid around the expectation of F. Its shape is determined by the covariance matrix Σ_F and its volume equals $\sqrt{\det \Sigma_F}$ multiplied by a factor that depends on α but not on the distribution F. Thus $F \preceq_{\mathcal{R}} G$ iff $\det \Sigma_F \geq \det \Sigma_G$. As the marginals are equal, the latter inequality is equivalent to $\det \mathrm{P}_F \geq \det \mathrm{P}_G$. □

Clearly, in the bivariate case (Table 1), we have $\det(\mathrm{P}_F) = 1 - \rho_F^2$, and (2) becomes the order of correlation coefficients, $F \preceq_{\mathcal{R}} G$ iff $\rho_F \leq \rho_G$
.

A distribution $F_{\boldsymbol{X}}$ is a *spherical distribution*, denoted $F_{\boldsymbol{X}} \in \mathcal{S}(\psi)$, if $\boldsymbol{X} =_{st} R\, \boldsymbol{U}$ with $\boldsymbol{U}$ being uniformly distributed on the unit sphere and R having distribution function ψ on the real line. (The symbol $=_{st}$ denotes equality in distribution.) $F_{\boldsymbol{X}}$ is an *elliptical distribution around* $\boldsymbol{a}$, denoted $F_{\boldsymbol{X}} = \mathcal{E}(\boldsymbol{a}, \boldsymbol{BB}', \psi)$, if $\boldsymbol{X} =_{st} \boldsymbol{a} + \boldsymbol{ZB}$, where $\boldsymbol{a} \in \mathbb{R}^d, \boldsymbol{B} = (b_{jl}) \in \mathbb{R}^{k \times d}$ with rank k and $\boldsymbol{Z}$ is spherical, $\boldsymbol{Z} \sim \mathcal{S}(\psi)$. For $F = \mathcal{E}(\boldsymbol{a}, \boldsymbol{BB}', \psi)$ and $G = \mathcal{E}(\boldsymbol{a}, \boldsymbol{CC}', \psi)$ denote

$$\Theta_F = diag(b_{11}^{-1}, \ldots, b_{dd}^{-1})\boldsymbol{BB}' diag(b_{11}^{-1}, \ldots, b_{dd}^{-1})\,,$$

and similarly Θ_G.

Theorem 2 *Consider a family of starshaped central regions that satisfies the 'balls under sphericity' condition. If F and G are elliptical, $F \in \mathcal{E}(\mu, \boldsymbol{BB}', \psi)$, $G \in \mathcal{E}(\mu, \boldsymbol{CC}', \psi)$, then*

$$F \preceq_{\mathcal{R}} G \qquad \Leftrightarrow \qquad \det \Theta_F \geq \det \Theta_G\,.$$

Zuo and Serfling (2000b, Th 4.4) present a similar result (with scatter matrices) for their scatter order and the halfspace depth.

Proof. By assumption the central regions of a spherical distribution are balls around the origin. Since central regions are, by definition, affine equivariant, each α-region of an elliptical distribution $\mathcal{E}(\mu, \boldsymbol{BB}', \psi)$ is an ellipsoid around μ with shape given by $\boldsymbol{BB}'$ and volume $\sqrt{\det \boldsymbol{BB}'}$ times a factor depending on α only. As the marginals are the same, $c_{jj} = b_{jj}$, obtain $F \preceq_{\mathcal{R}} G$ iff $\det \boldsymbol{BB}' \geq \det \boldsymbol{CC}'$ iff $\det \Theta_F \geq \det \Theta_G$. □

The dependent exponential distribution of Example 2 is no elliptical distribution. Table 2 exhibits, for different values of the correlation coefficient ρ, the volumes of central regions in the Mahalanobis and the zonoid sense. It is seen that the volume decreases with increasing ρ, unless ρ is very close to 0 and α is small.

Joe (1997) *postulates* that a dependence order $\preceq$ should possess the following properties:

1. preorder (reflexive and transitive),
2. antisymmetric,
3. continuous w.r.t. weak convergence of distributions,
4. invariant to marginalization,
5. invariant to permutations of the components,
6. invariant to increasing transforms of the components,
7. consistent with the bivariate concordance order,[5]
8. maximal at the upper Fréchet bound.[6]

Note that Postulate 8 is implied by 7. It can be shown that Postulates 1–5 are satisfied with the Mahalanobis dependence order and the zonoid dependence order, but Postulates 6–7 are not. Neither the weaker Postulate 8 is fulfilled.

[5] Two distributions are ordered in the *bivariate concordance order* if all two-dimensional marginals are ordered in the lower orthant order, that is their bivariate distribution functions are pointwise ordered.

[6] Given the univariate marginals $F_1, \ldots, F_d$, the upper Fréchet bound is $F^+(\boldsymbol{x}) = \min_j F_j(x_j)$. The bivariate concordance order attains its maximum at the upper Fréchet bound.

Mahalanobis region volumes										
α	Correlation ρ									
0.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.10	28.87	29.78	28.73	28.42	26.69	24.57	21.85	18.13	14.56	9.83
0.20	12.83	13.22	12.74	12.61	11.84	10.91	9.69	8.04	6.45	4.36
0.30	7.41	7.70	7.42	7.35	6.89	6.35	5.64	4.68	3.76	2.54
0.40	4.77	4.94	4.77	4.71	4.42	4.08	3.62	3.00	2.41	1.62
0.50	3.21	3.29	3.17	3.14	2.95	2.71	2.41	2.00	1.60	1.08
0.60	2.10	2.19	2.11	2.09	1.96	1.80	1.60	1.33	1.06	0.72
0.70	1.36	1.40	1.35	1.34	1.26	1.15	1.03	0.85	0.68	0.46
0.80	0.80	0.81	0.78	0.78	0.73	0.67	0.60	0.49	0.40	0.27
0.90	0.35	0.36	0.35	0.34	0.32	0.29	0.26	0.22	0.17	0.12
Zonoid region volumes										
α	Correlation ρ									
0.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.10	8.60	9.15	9.07	8.80	8.49	7.84	6.89	5.76	4.40	2.39
0.20	5.30	5.42	5.16	4.96	4.63	4.09	3.48	2.78	1.94	0.98
0.30	3.57	3.55	3.32	3.13	2.85	2.46	2.04	1.60	1.08	0.53
0.40	2.46	2.40	2.21	2.05	1.84	1.57	1.29	1.02	0.68	0.33
0.50	1.67	1.62	1.48	1.37	1.22	1.04	0.86	0.67	0.45	0.22
0.60	1.08	1.06	0.97	0.90	0.81	0.69	0.57	0.45	0.30	0.15
0.70	0.65	0.64	0.60	0.56	0.51	0.44	0.37	0.29	0.20	0.10
0.80	0.32	0.33	0.31	0.30	0.28	0.25	0.21	0.17	0.12	0.06
0.90	0.10	0.11	0.11	0.10	0.10	0.09	0.08	0.07	0.05	0.03

Table 2: Volumes of Mahalanobis and zonoid regions for $n = 500$ bivariate exponential observations (Marshall-Olkin distribution) with $X_1, X_2 \sim Exp(1)$ and several α and $\rho = corr(X_1, X_2)$.

4 Copula dependence ordering

It is clear that our above dependence orders do not satisfy Joe's Postulate 6. To obtain dependence orderings that are invariant to increasing transforms of the components, we apply the previous dependence orders to the copulae of two distributions instead of the distributions themselves. For given $\boldsymbol{X}$ denote

$$\boldsymbol{X}^0 = (F_1(X_1), \ldots, F_d(X_d)),$$

where F_j is the distribution function of the marginal X_j. The distribution of $\boldsymbol{X}^0$ is mentioned as the *standardized distribution* of the random vector $\boldsymbol{X}$. Define $\boldsymbol{Y}^0$ similarly. Note that each univariate marginal of $\boldsymbol{X}^0$ is uniformly distributed on $[0, 1]$, and the same holds for the marginals of $\boldsymbol{Y}^0$.

Definition 3 (Copula dependence order)
Distribution $F_{\boldsymbol{Y}}$ is more copula-dependent *than distribution $F_{\boldsymbol{X}}$ with respect to $\mathcal{R}$, $F_{\boldsymbol{X}} \preceq^0_{\mathcal{R}} F_{\boldsymbol{Y}}$, if*

$$\text{vol}_d D_\alpha(F_{\boldsymbol{X}^0}) \geq \text{vol}_d D_\alpha(F_{\boldsymbol{Y}^0}) \quad \textit{holds for every } \alpha\,.$$

Then the copula dependence order $\preceq^0_{\mathcal{R}}$ satisfies Joe's Postulates 1 and 3–6, but not 2 and 7.

Figures 8 and 9 show the Mahalanobis and zonoid regions for 50 observations simulated from a bivariate standardized Marshall-Olkin distribution. Table 3 shows the volumes of zonoid regions for 500 observations from standardized Marshall-Olkin distributions for different values of the correlation ρ. It is seen from the table that all these volumes decrease as ρ increases.

MAHALANOBIS REGION VOLUMES										
α	Correlation ρ									
0.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.10	0.71	0.71	0.69	0.66	0.64	0.59	0.54	0.47	0.37	0.20
0.20	0.53	0.52	0.50	0.48	0.45	0.40	0.34	0.27	0.19	0.10
0.30	0.39	0.38	0.36	0.34	0.31	0.27	0.22	0.17	0.12	0.06
0.40	0.28	0.27	0.25	0.23	0.21	0.18	0.15	0.11	0.08	0.04
0.50	0.20	0.19	0.17	0.16	0.14	0.19	0.10	0.08	0.05	0.02
0.60	0.13	0.12	0.11	0.10	0.09	0.08	0.06	0.05	0.03	0.02
0.70	0.07	0.07	0.07	0.06	0.06	0.05	0.04	0.03	0.02	0.01
0.80	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01
0.90	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00

Table 3: Volumes of zonoid trimmed regions for $n = 500$ standardized Marshall–Olkin observations, $(1 - \exp(-X_1), 1 - \exp(-X_2))$, $X_1, X_2 \sim Exp(1)$, for several α and ρ.

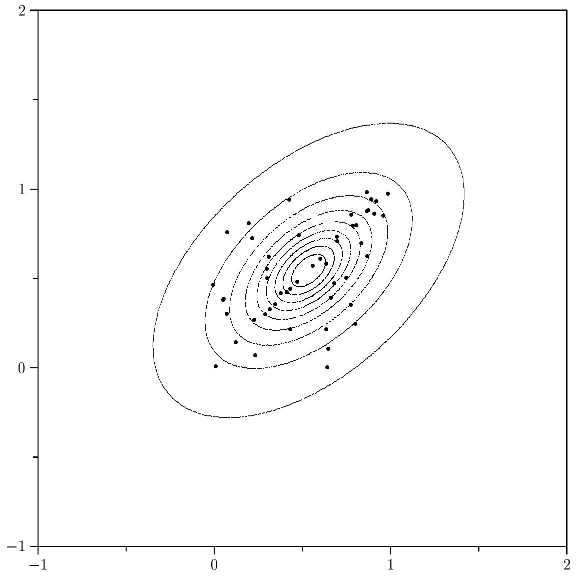

Figure 8: Mahalanobis regions for $n = 50$ standardized Marshall–Olkin observations, $(1 - \exp(-X_1), 1 - \exp(-X_2))$, $X_1, X_2 \sim Exp(1)$, $\rho = 1/3$.

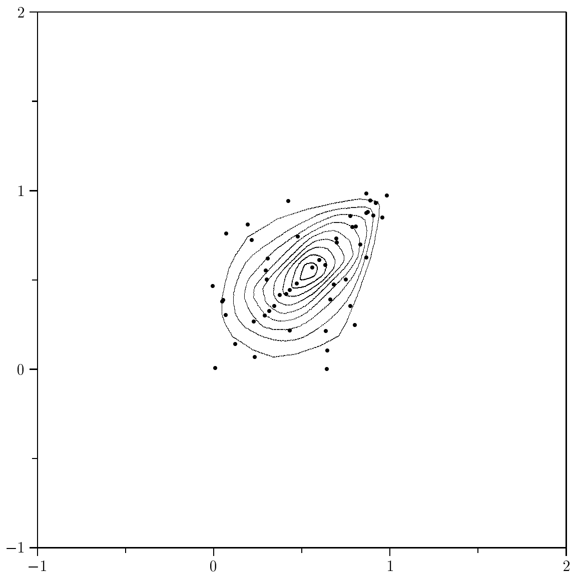

Figure 9: Zonoid regions for $n = 50$ standardized Marshall–Olkin observations, $(1 - \exp(-X_1), 1 - \exp(-X_2))$, $X_1, X_2 \sim Exp(1)$, $\rho = 1/3$.

5 Dependence indices

Indices consistent with the above dependence orders are easily constructed. Let F be a distribution with correlation matrix P_F and let F_{ind} be the independent version of F, that is, the distribution with the same but independent marginals. Consider, for some fixed α^*, the ratio

$$GC_{\alpha^*}(F) = \frac{\mathrm{vol}_d D_{\alpha^*}(F_{ind})}{\mathrm{vol}_d D_{\alpha^*}(F)}\,. \tag{3}$$

If either both distributions are elliptical or the regions are Mahalanobis, then

$$GC_{\alpha^*}(F) = \frac{1}{\sqrt{\det \mathrm{P}_F}} \quad \text{resp.} \quad GC_{\alpha^*}(F) = \frac{1}{\sqrt{\det \Theta_F}}$$

holds independently of α^*. By this, for a given family of starshaped central regions and with a properly chosen α^*, the ratio GC_{α^*} in (3) can be mentioned as the *generalized correlation coefficient* of a distribution F.

For any notion of central regions and given distributions F and G that have the same marginals, holds

$$\begin{aligned} F \preceq_{\mathcal{R}} G \quad &\Leftrightarrow \quad \mathrm{vol}_d D_\alpha(F) \geq \mathrm{vol}_d D_\alpha(G) \quad \text{for all } \alpha \\ &\Leftrightarrow \quad GC_{\alpha^*}(F) \leq GC_{\alpha^*}(G) \quad \text{for all } \alpha^*\,. \end{aligned} \tag{4}$$

Hence, for any α^*, the generalized correlation coefficient is consistent with the central regions dependence order $\preceq_{\mathcal{R}}$. A generalized correlation coefficient that is consistent with the copula dependence order $\preceq^0_{\mathcal{R}}$ is defined in an obvious similar way.

Instead of picking some α^*, the volumes can be aggregated over α. In particular, we propose the weighted index

$$\int_0^1 \alpha^d \ \mathrm{vol}_d D_\alpha(F)\ d\alpha\,,$$

where each D_α receives weight α^d. With zonoid central regions, this index is equal to the lift zonoid volume. Observe that $F \preceq_{\mathcal{R}} G$ with zonoid central regions implies that the lift zonoid of F has a larger volume than that of G.

An aggregate version of the generalized correlation of F is obtained by the ratio of lift zoonoid volumes,

$$GC_{lz}(F) = \frac{\int_0^1 \alpha^d \ \mathrm{vol}_d D_\alpha(F_{ind}) \ d\alpha}{\int_0^1 \alpha^d \ \mathrm{vol}_d D_\alpha(F) \ d\alpha}$$

where the D_α are zonoid regions and, again, F_{ind} is the independent version of F. By (4), the coefficient $GC_{lz}(F)$ is consistent with the zonoid regions order of F.

6 Final remarks

The approach of this paper is in the spirit of Liu et al. (1999) who use notions of depth and trimmed regions to describe the *location* and *scatter* of multivariate distributions. However, these authors do not consider dependence. They employ central regions and parametrize them by their *probability content*, which is different from the approach chosen here. Here we have measured *dependence* by an order of central region *volumes*. The same order has been previously used by Zuo and Serfling (2000b) for comparing the scatter of distributions.

In fact, our dependence orders measure *departures from independence* in either direction and do not distinguish whether the dependence is in the positive or negative. By relating a given dependent distribution to the independent distribution that has the same marginals, a *generalized correlation coefficient* is obtained as the ratio of two central region volumes. For this, neither the existence of second moments is assumed nor the dependence restricted to some linear interrelationship among components.

To compare distributions with *unequal marginals* with respect to their dependence, the copula dependence order $\preceq^0_{\mathcal{R}}$ may be used as it is, and the non-copula dependence order $\preceq_{\mathcal{R}}$ after proper standardization with respect to location and scale.

When the above dependence orders are applied to data, volumes of central regions have to be calculated for empirical distributions. This is most simple for the Mahalanobis ellipsoids, but can also be done in reasonable time for the zonoid regions. When determining these central regions from data, that is, from an empirical distribution, P_F and Θ_F may be estimated in a way which is robust against outliers. See, e.g., Rousseeuw's minimum volume ellipsoid, or the rank correlation matrix and center rank vector by Hettmansperger et al. (1998) and Oja (1999).

In comparing empirical distributions, rather often a dependence order fails because it is violated by a few 'extreme' data points which lie far outside the 'main bulk' of the data. A natural variant of our dependence order is the following order, which is finer than (i.e. implied by) $\preceq_{\mathcal{R}}$: Define $F \preceq_{\mathcal{R},\alpha_0} G$ if

$$\mathrm{vol}_d D_\alpha(F) \geq \mathrm{vol}_d D_\alpha(G) \quad \text{for every } \ \alpha > \alpha_0 \, .$$

The parameter α_0 controls how coarse or fine the order should be.

It suggests itself to consider also the order that is defined by the *set inclusion* of all central regions. This order, obviously, implies the order of volumes. However, the set inclusion order may be too strong. With the zonoid regions it amounts to the set inclusion of all lift zonoids and becomes trivial, which means that only identical distributions are ordered. Dall'Aglio and Scarsini (2001) define a dependence order by the set inclusion of zonoids (not lift zonoids); their order yields a linear dependence order which implies the concordance order.

References

Block, H., Sampson, A. and Savits, T., eds. (1990). *Topics in Statistical Dependence.* Institute of Mathematical Statistics, Hayward, California.

Dall'Aglio, M. and Scarsini, M. (2001). When Lorenz met Lyapunov. *Statistics and Probability Letters* **54**, 101–105.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, Convexity, and Applications.* Academic Press, Boston.

Dyckerhoff, R. (2002). Inference based on data depth. Chapter 5, in K. Mosler, *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*, Springer, New York.

Hettmansperger, T., Möttönen, J. and Oja, H. (1998). Affine invariant multivariate rank tests for several samples. *Statistica Sinica* **8**, 785–800.

JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

KOSHEVOY, G. and MOSLER, K. (1997a). Multivariate Gini indices. *Journal of Multivariate Analysis* **60**, 252–276.

KOSHEVOY, G. and MOSLER, K. (1997b). Zonoid trimming for multivariate distributions. *Annals of Statistics* **25**, 1998–2017.

LIU, R., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statistics* **27**, 783–858. With discussion.

MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*. Springer, New York.

MÜLLER, A. and STOYAN, D. (2002). *Comparison Methods for Stochastic Models and Risks*. J. Wiley, New York.

OJA, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates. *Scandinavian Journal of Statistics* **26**, 319–343.

ZUO, Y. and SERFLING, R. (2000a). General notions of statistical depth function. *Annals of Statistics* **28**, 461–482.

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