Consistency of the likelihood depth estimator for the correlation coefficient

Liesa Denecke* and Christine H. Müller
Fakultät Statistik
Technische Universität Dortmund
44221 Dortmund
Germany

March 30, 2012

Abstract

Denecke and Müller (2011) presented an estimator for the correlation coefficient based on likelihood depth for Gaussian copula and Denecke and Müller (2012) proved a theorem about the consistency of general estimators based on data depth using uniform convergence of the depth measure. In this article, the uniform convergence of the depth measure for correlation is shown so that consistency of the correlation estimator based on depth can be concluded. The uniform convergence is shown with the help of the extension of the Glivenko-Cantelli Lemma by Vapnik-C̃ervonenkis classes.

Keywords: consistency, data depth, Gaussian copula, likelihood depth, parametric estimation, correlation coefficient

AMS Subject classification: 62G35, 62H20, 62G07

1 Introduction

Different notions of data depth were presented to generalize the median to multivariate data and more complex situations, see e.g. Tukey (1975), Liu (1990) and Mosler (2002). Now

*Corresponding Author, Email: denecke@statistik.tu-dortmund.de

there exists a broad class of applications as e.g. in Lin and Chen (2006), Li and Liu (2008), Romanazzi (2009), López-Pintado and Romo (2009), López-Pintado et al. (2010), Hu et al. (2009), just to mention some recent results. Rousseeuw and Hubert (1999) developed depth notions via the nonfit and Mizera (2002) extended this approach to general quality function. Using the likelihood function as a quality function leads to likelihood depth, see Mizera and Müller (2004). Estimators maximizing the likelihood depth are robust alternatives to the nonrobust maximum likelihood estimators (MLE). The estimator based on likelihood depth is as flexible as the MLE and can be used in many situations. While the MLE is very sensitive to changes in the underlying distribution, the estimator based on likelihood depth is not. In particular, these estimators show high robustness against contaminations with other distributions, see e.g. Denecke and Müller (2011). Denecke and Müller (2012) proved a high breakdown point and consistency of estimators and tests based on a general depth notion including likelihood depth for one-dimensional parameters. Thereby consistency of a depth estimator is shown under uniform convergence of the depth measure.

Using likelihood depth, Denecke and Müller (2011) developed robust estimators for the parameters of copulas. Applying this approach to the Gaussian copula led to a new robust estimator of correlation since the parameter ρ of the Gaussian copula is the classical correlation parameter. However, the proof of its consistency is difficult since uniform convergence of the depth measure must be shown. In this paper, the proof of this uniform convergence and thus of the consistency of the new correlation estimator is given for $\rho \neq 0$, i.e. for the dependent case.

We start in Section 2 with a very short introduction of the likelihood depth for a one-dimensional parameter. The theorem about consistency of the maximum depth estimator under uniform convergence of the depth measure is given and the application of the Theorem of Vapnik-Červonenkis for providing uniform convergence of likelihood depth is presented. Section 3 presents the new estimator for correlation based on likelihood depth. In Section 4, the uniform convergence and thus the consistency is proved with the Theorem of Vapnik-Červonenkis. Finally, a small data example in Section 5 shows that the new estimator behaves similar to the correlation estimator based on the Minimum Covariance Determinant (MCD) given in Rousseeuw and Leroy (1987).

2 Consistency of estimators based on likelihood depth

Let be $Z_1, \ldots, Z_N$ i.i.d. with distribution P_θ and density $f_\theta : \mathbb{R}^p \to \mathbb{R}$ where $\theta \in \Theta$ is an unknow parameter of an ideal distribution P_θ. We consider here only the case $\theta \in \Theta \subset \mathbb{R}$. Realizations of $Z_1, \ldots, Z_N$ are $z_1, \ldots, z_N \in \mathbb{R}^p$. Crucial for the definition of likelihood depth of a parameter $\theta \in \Theta$ in a sample $z_* = (z_1, \ldots, z_N)$ are the sets

$$T_{pos}^\theta := \{z \in \mathbb{R}^p; \tfrac{\partial}{\partial\theta} \ln f_\theta(z) \geq 0\}, \quad T_{neg}^\theta := \{z \in \mathbb{R}^p; \tfrac{\partial}{\partial\theta} \ln f_\theta(z) \leq 0\},$$
$$T_0^\theta := \{z \in \mathbb{R}^p; \tfrac{\partial}{\partial\theta} \ln f_\theta(z) = 0\}$$

and the quantities

$$\lambda_N^+(\theta, z_*) := \tfrac{1}{N}\sharp\{n;\ z_n \in T_{pos}^\theta\}, \quad \lambda_N^-(\theta, z_*) := \tfrac{1}{N}\sharp\{n;\ z_n \in T_{neg}^\theta\},$$
$$\lambda_N^0(\theta, z_*) := \tfrac{1}{N}\sharp\{n;\ z_n \in T_0^\theta\}.$$

Then the likelihood depth of a parameter $\theta \in \Theta$ in a sample $z_* = (z_1, \ldots, z_N)$ is defined by

$$d_L(\theta, z_*) = \lambda_N^0(\theta, z_*) + \min(\lambda_N^+(\theta, z_*), \lambda_N^-(\theta, z_*)),$$

i.e. the likelihood depth is calculated by counting the observations z_n, $n = 1, \ldots, N$, for which $\frac{\partial}{\partial\theta} \ln f_\theta(z_n)$ is positive, negative and zero respectively, see e.g. Denecke an Müller (2011). The maximum likelihood depth estimator $\tilde{\theta}_N$ for the parameter θ is the one in the parameter-space Θ that has maximum likelihood depth, i.e.

$$\tilde{\theta}_N(z_*) \in \arg\max_{\theta \in \Theta} d_L(\theta, z_*).$$

Müller and Denecke (2012) pointed out, that the maximum likelihood depth estimator is biased if

$$p_{\theta,\theta} := P_\theta(T_{pos}^\theta) \neq \tfrac{1}{2}.$$

In these cases, they show that the estimator converges to a shifted $s(\theta) \neq \theta$ that is given by the equation

$$p_{\theta,s(\theta)} := P_\theta(T_{pos}^{s(\theta)}) = \tfrac{1}{2}.$$

Denecke and Müller also showed that the corrected maximum depth estimator $\widehat{\theta}_N(z_*) = s^{-1}(\tilde{\theta}_N(z_*))$ is a consistent estimator under some regularity conditions:

Proposition 1. *Let P_{θ_0} be the underlying distribution, $\lambda^+_{\theta_0}(\theta) = P_{\theta_0}(T^\theta_{pos})$, and $\lambda^-_{\theta_0}(\theta) = P_{\theta_0}(T^\theta_{neg})$. If*

a) $\lambda^\pm_N(\cdot, Z_*)$ converges uniformly almost surely to $\lambda^\pm_{\theta_0}(\cdot)$,

b) s^{-1} is continuous,

c) and for all $\varepsilon > 0$ there exists $\delta > 0$, such that

$$\min(\lambda^+_{\theta_0}(\theta), \lambda^-_{\theta_0}(\theta)) < \frac{1}{2} - \delta \text{ for all } \theta \text{ with } |s(\theta_0) - \theta| > \varepsilon,$$

then the maximum depth estimator $\tilde{\theta}_N$ converges to $s(\theta_0)$ almost surely and the corrected maximum depth estimator $s^{-1}(\tilde{\theta}_N)$ converges to θ_0.

Hence crucial for the consistency is the uniform convergence of $\lambda^\pm_N$. This can be shown by a generalization of the Glivenko-Cantelli-Lemma. The generalization is the Theorem of Vapnik-C̃ervonenkis based Vapnik-C̃ervonenkis classes, see e.g. van der Vaart and Wellner (1996). The definition of a Vapnik-C̃ervonenkis class can be found in van der Vaart and Wellner (1996), Section 2.6:

Definition 1. *Let be $\mathcal{C}$ a collection of subsets of a set $\mathcal{X}$. An arbitrary set of n points $\{x_1, \ldots, x_n\}$ posses 2^n subsets. $\mathcal{C}$ picks out a certain subset from $\{x_1, \ldots, x_n\}$, if this can be formed as a set of the form $C \cap \{x_1, \ldots, x_n\}$ for $C \in \mathcal{C}$. $\mathcal{C}$ is said to shatter $\{x_1, \ldots, x_n\}$ if each of the 2^n subsets can be picked out. The VC-index $V(\mathcal{C})$ of a class $\mathcal{C}$ is the smallest n for which no set of size n is shattered by $\mathcal{C}$. Formally this means*

$$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) := \sharp\{C \cap \{x_1, \ldots, x_n\}; C \in \mathcal{C}\},$$
$$V(\mathcal{C}) := \inf\{n; \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) < 2^n\}.$$

A collection $\mathcal{C}$ of measurable sets C is called VC-class, if its index is finite.

A corollary of the Theorem of Vapnik-C̃ervonenkis is then:

Corollary 1. *If $\{T^\theta_{pos};\ \theta \in \Theta\}$ and $\{T^\theta_{neg};\ \theta \in \Theta\}$ are VC-classes, then $\lambda^\pm_N(\cdot, Z_*)$ converges uniformly almost surely to $\lambda^\pm_{\theta_0}(\cdot)$.*

If

$$T^{\theta}_{pos} \subsetneq T^{\theta'}_{pos} \text{ and } T^{\theta'}_{neg} \subsetneq T^{\theta}_{neg} \text{ for all } \theta < \theta', \tag{1}$$

or

$$T^{\theta}_{pos} \subsetneq T^{\theta'}_{pos} \text{ and } T^{\theta'}_{neg} \subsetneq T^{\theta}_{neg} \text{ for all } \theta > \theta'. \tag{2}$$

then the Vapnik-$\tilde{\text{C}}$ervonenkis index of $\{T^{\theta}_{pos};\ \theta \in \Theta\}$ as well as of $\{T^{\theta}_{neg};\ \theta \in \Theta\}$ is two. But this is not satisfied for the likelihood depth of the Gaussian copula.

3 Estimator for the correlation coefficient

In this section we present the estimator for the correlation coefficient ρ based on likelihood depth for the bivariate Gaussian copula. The bivariate Gaussian copula is given by a bivariate normal distribution where the marginals have standard normal distribution. We asume here that the original data $(u_1, v_1), \ldots, (u_N, v_N)$ are realizations of i.i.d. random variables $(U_1, V_1), \ldots, (U_N, V_N)$ with assumed bivariate normal distribution. To achieve that the marginal distributions are standard normal distributions, (U_n, V_n) are standardized to $Z_n = (X_n, Y_n)$ so that X_n and Y_n have standard normal distribution. In applications the standardization is done by estimating the means and the variances of U_n and V_n. But for deriving the maximum likelihood depth estimator it is assumed that these means and variances are known. Then the derivative of the log-likelihood function of the standardized variables $Z_n = (X_n, Y_n)$ at $z = (x, y) \in \mathbb{R}^2$ is

$$\frac{\partial}{\partial \rho} \ln f_\rho(x, y) = \frac{-\rho y^2 + (1 + \rho^2)\, x\, y + \rho - \rho^3 - \rho x^2}{(1 - \rho^2)^2}$$

(see Denecke and Müller 2011). The next step is to check whether the maximum likelihood depth estimator is biased, therefore the values for $p_{\rho,\rho} = P_\rho(T^{\rho}_{pos})$ are calculated. To determine $p_{\rho,\rho}$ for a fixed ρ, we need the boundaries of T^{ρ}_{pos}, which are given by the zeros of $\frac{\partial}{\partial \rho} \ln f_\rho(x, y)$.

For $\rho = 0$ we have

$$\frac{\partial}{\partial \rho} \ln f_\rho(x, y) = \frac{-0 \cdot y^2 + (x + 0^2 \cdot x)y + 0 - 0^3 - 0 \cdot x^2}{(1 - 0^2)^2} = xy.$$

This means that $\frac{\partial}{\partial\rho} \ln f_\rho(x,y) < 0$ if and only if x and y have different sign so that the probability that a data lies inside the region T^ρ_{pos} is $\frac{1}{2}$. Thus, the parameter with maximum depth is not asymptotically biased for $\rho = 0$. However, the situation changes for $\rho \neq 0$.

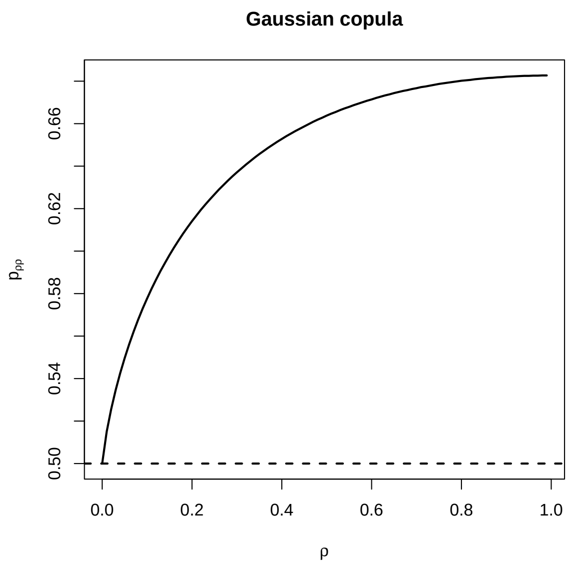

Figure 1: Plot of $(\rho, p_{\rho,\rho})$.

Since the cases $\rho > 0$ and $\rho < 0$ are completely similar, we now consider only $\rho > 0$. In Denecke and Müller (2011) it was shown that the zeros of $\frac{\partial}{\partial\rho} \ln f_\rho(x,y)$ are

$$v_+(x,\rho) = \frac{\rho^2 x + x + \sqrt{\rho^4 x^2 - 2\rho^2 x^2 + x^2 - 4\rho^4 + 4\rho^2}}{2\rho}$$

and

$$v_-(x,\rho) = \frac{\rho^2 x + x - \sqrt{\rho^4 x^2 - 2\rho^2 x^2 + x^2 - 4\rho^4 + 4\rho^2}}{2\rho}$$

so that

$$T^\rho_{pos} = \{z = (x,y); v_-(x,\rho) \leq y \leq v_+(x,\rho)\}$$

and

$$p_{\rho_0,\rho} = P_{\rho_0}(T^\rho_{pos}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \left(\Phi\left(\frac{v_+(x,\rho) - \rho_0 x}{\sqrt{1-\rho_0^2}} \right) - \Phi\left(\frac{v_-(x,\rho) - \rho_0 x}{\sqrt{1-\rho_0^2}} \right) \right) dx,$$

where Φ denotes the one-dimensional standard normal distribution function. Furthermore we have $v_{-}(x,\rho) < x < v_{+}(x,\rho)$, see also Denecke (2010). The values of $p_{\rho,\rho}$ in Figure 1 were calculated by numerical integration. The graphic shows that the probability $p_{\rho,\rho}$ differs from $\frac{1}{2}$, so that the maximum likelihood depth estimator is biased.

The bias function s given by $P_\rho(T_{pos}^{s(\rho)}) = \frac{1}{2}$ as well as the bias correction function s^{-1} given by $P_{s^{-1}(\rho)}(T_{pos}^{\rho}) = \frac{1}{2}$ have no explicit form. The function s^{-1} was approximated in Denecke and Müller (2011) numerically by

$$s^{-1}(\rho) = -1.24101\,\rho^3 + 3.68702\,\rho^2 - 1.4546\,\rho + 0.00857$$

for $\rho > 0$ so that the new estimator for the correlation ρ was defined by

$$\widehat{\rho}(z_*) \;=\; \begin{cases} -1.24101\,\tilde{\rho}^3 + 3.68702\,\tilde{\rho}^2 - 1.4546\,\tilde{\rho} + 0.00857, & \text{if } \tilde{\rho} \geq 0.461, \\ 1.24101\,\tilde{\rho}^3 - 3.68702\,\tilde{\rho}^2 + 1.4546\,\tilde{\rho} - 0.00857, & \text{if } \tilde{\rho} \leq -0.461, \\ 0, & \text{else,} \end{cases} \qquad (3)$$

where $\tilde{\rho} = \arg\max_\rho d_L(\rho, z_*)$. The three cases are caused by the fact that $\lambda_0^+(s(0)) = P_0(T_{pos}^{s(0)}) = \frac{1}{2}$ has three solutions, namely $s(0) = 0$, $s(0) \approx 0.461$, and $s(0) \approx -0.461$.

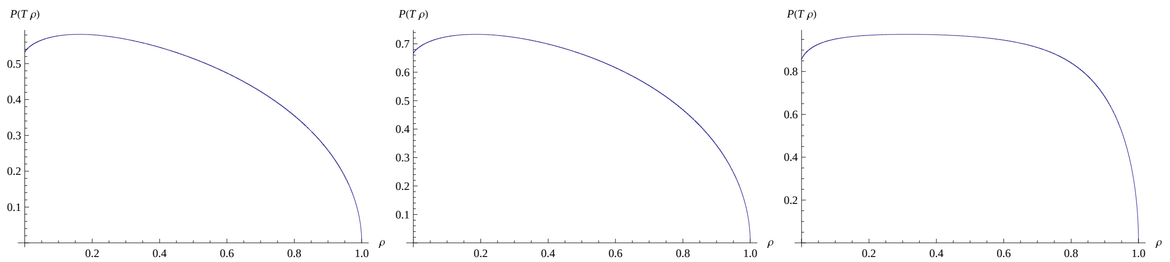

Figure 2: $\lambda_{\rho_0}^+(\rho)$ for $\rho_0 = 0.1, 0.5$ and 0.9

4 Consistency of the correlation estimator

Figure 2 shows that the solutions $s(\rho)$ and $s^{-1}(\rho)$ of $\lambda_\rho^+(s(\rho)) = P_\rho(T_{pos}^{s(\rho)}) = \frac{1}{2}$ and $\lambda_{s^{-1}(\rho)}^+(\rho) = P_{s^{-1}(\rho)}(T_{pos}^{\rho}) = \frac{1}{2}$, respectively, are unique for $\rho > 0$. In particular it holds $\lambda_{\rho_0}^-(\rho) = 1 - \lambda_{\rho_0}^+ < \frac{1}{2}$ for $\rho < s(\rho_0)$ and $\lambda_{\rho_0}^+(\rho) < \frac{1}{2}$ for $\rho > s(\rho_0)$. An analogous result holds for $\rho < 0$. The functions s and s^{-1} are also continuous since $p_{\rho_0,\rho} = P_{\rho_0}(T_{pos}^{\rho})$ is continuous differentiable in both arguments. Hence conditions b) and c) of Proposition 1 are satisfied.

However Figure ??????????????????????? shows that neither condition (1) nor condition (2) is true. But we have the following Theorem:

Theorem 1. *$\{T^\rho_{pos};\ 0<\rho\leq 1\}$, $\{T^\rho_{neg};\ 0<\rho\leq 1\}$, $\{T^\rho_{pos};\ -1\leq\rho<0\}$, and $\{T^\rho_{neg};\ -1\leq \rho<0\}$ are VC-classes, each with VC-index less than 7.*

Proof: Because of symmetry, we regard only $0<\rho\leq 1$. We already elaborated in Section 3 that $T^\rho_{pos}=\{(x,y)\in\mathbb{R}^2; v_-(x,\rho)\leq y\leq v_+(x,\rho)\}$ holds with

$$v_\pm(x,\rho)=\frac{1}{2\rho}\left(x(\rho^2+1)\pm\sqrt{x^2(1-\rho^2)^2-4\rho^2(\rho^2-1)}\right).$$

Since the density $f_\rho(x,y)$ for the Gaussian copula (the bivariate normal distribution with means equal to 0 and variances equal to 1) is symmetric in x and y, it holds

$$(x,y)\in T^\rho_{pos}\Leftrightarrow(y,x)\in T^\rho_{pos}\Leftrightarrow(-x,-y)\in T^\rho_{pos}\Leftrightarrow(-y,-x)\in T^\rho_{pos}.$$

Thus, for checking $(x,y)\in T^\rho_{pos}$, we can transform (x,y) to $(\tilde{x},\tilde{y})$, such that $\tilde{x}\geq 0$ and $\tilde{y}\leq\tilde{x}$. Then $(x,y)\in T^\rho_{pos}$, iff $\tilde{y}\geq v_-(\tilde{x},\rho)$, as $\tilde{y}\leq v_+(\tilde{x},\rho)$ is always true because $\tilde{y}\leq\tilde{x}\leq v_+(\tilde{x},\rho)$. Because of this, it is sufficient to consider points (x,y) with $x\geq 0$ and $y\leq x$.

The next step is to show that for every $z=(x,y)$ there are only finitely many disjunct intervals $[\rho_{i_1},\rho_{i_2}]$, $0<\rho_{i_1}<\rho_{i_2}\leq 1$, such that $z\in T^\rho_{pos}$ for $\rho\in[\rho_{i_1},\rho_{i_2}]$ and $z\notin T^\rho_{pos}$ outside of the intervals. That is true, if $v_-(x,\cdot)$ takes every value only a finite time, i.e. $v_-(x,\cdot)$ has the slope zero only for a finite number of values. Therefore we regard the derivative of $v_-(x,\cdot)$. For $0<\rho<1$, it holds

$$\begin{aligned}\frac{\partial}{\partial\rho}v_-(x,\rho) &= \frac{x(\rho^2-1)\sqrt{x^2(1-\rho^2)^2-4\rho^2(\rho^2-1)}+x^2(1-\rho^2)^2-4\rho^2(\rho^2-1)}{2\rho^2\sqrt{x^2(1-\rho^2)^2-4\rho^2(\rho^2-1)}}\\ &\quad+\frac{2x^2\rho^2-2x^2\rho^4+8\rho^4-4\rho^2}{2\rho^2\sqrt{x^2(1-\rho^2)^2-4\rho^2(\rho^2-1)}}.\end{aligned}$$

$\frac{\partial}{\partial\rho}v_-(x,\rho)=0$ is true, iff

$$x(\rho^2-1)\sqrt{x^2(1-\rho^2)^2-4\rho^2(\rho^2-1)}+x^2+2x^2\rho^4+4\rho^4=0,$$

which is equivalent to $-4\rho^2x^4+2\rho^4x^4-4\rho^6x^4-3\rho^8x^4-12\rho^4x^2+12\rho^6x^2-8\rho^8x^2+4\rho^2x^2+16\rho^8=0$. This is a polynomial with degree 8 for ρ, so it has at most 8 zeros, especially the number of zeros is finite. This means for every $z = (x, y)$ with $x \geq 0, y \leq x$ that there are at most $l = 9$ intervals $[\rho_{i_1}, \rho_{i_2}]$ such that $z \in T^{\rho}_{pos}$ for $\rho \in [\rho_{i_1}, \rho_{i_2}]$.

Now we show that $V(\mathcal{C}) < 7$. Let be $\{z_1, \ldots, z_7\}$ with $z_k = (x_k, y_k)$, where it is enough to consider $x_k \geq 0, y_k \leq x_k$, $k = 1, \ldots, 7$, as discussed above. We already stated that for every z there are at most $l = 9$ intervals $[\rho_{i_1}, \rho_{i_2}]$ such that $z \in T^{\rho}_{pos}$ for $\rho \in [\rho_{i_1}, \rho_{i_2}]$, $1 \leq i \leq 9$. Every interval has 2 endpoints, thus there are at most $2 \cdot 9$ endpoints for every z. The first point z_1 divides the interval $[0, 1]$ into maximal $2 \cdot 9 + 1$ subsets. Every point, that is added, increases the number of subsets of $[0, 1]$ by at most $2 \cdot 9$. All in all we get at most $7 \cdot 2 \cdot 9 + 1 = 127$ subsets. To shatter the seven points, there are $2^7 = 128$ subsets needed. Therefore not all possible subsets of $\{z_1, \ldots, z_7\}$ are picked out. Hence the VC-index of $\{T^{\rho}_{pos};\ 0 < \rho \leq 1\}$ is less than 7. Similar proofs provide also the same VC-index of $\{T^{\rho}_{neg};\ 0 < \rho \leq 1\}$, $\{T^{\rho}_{pos};\ -1 \leq \rho < 0\}$, and $\{T^{\rho}_{neg};\ -1 \leq \rho < 0\}$. $\square$

Using Corollary 1, the condition a) of Proposition 1 is also satisfied so that we have:

Theorem 2. *The corrected maximum likelihood-depth estimator $\widehat{\rho}$ given by (3) is a strongly consistent estimator for $\rho \neq 0$.*

Since we have $P_0(T^0_{pos}) = P_0(T^r_{pos}) = P_0(T^{-r}_{pos}) = \frac{1}{2}$ for $r \approx 0.461$, the consistency does not hold for $\rho = 0$.

5 Example

As a data example we use the data set `Animals2` of the R-package "robustbase". A data frame with average brain and body weights for 62 species of land mammals and three others. It is a union of the mammals data set of Weisberg (1985) and the animals data set of Rousseeuw and Leroy (1987). A scatterplot of the log-data is given in Figure 3. We see that there are three outlying points. To calculate the correlation between the logarithm of brain and body weights we use Pearson's correlation coefficient, the robust minimum covariance determinant estimator (MCD), see Rousseeuw and Leroy (1987), and the corrected maximum likelihood depth estimator (MLD). For calculating the MLD estimator, the data are standardized with the arithmetic mean and the standard deviation. Although these estimators are not robust ????????????, MLD and MCD give the same result, 0.956, what

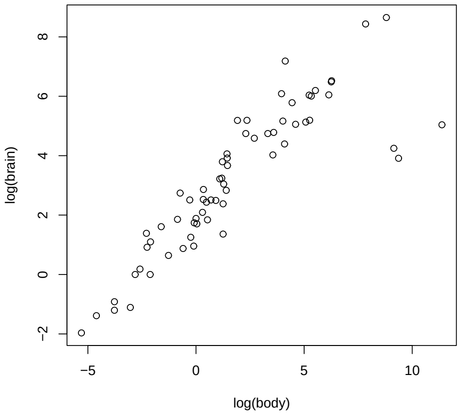

Figure 3: Scatterplot of the Animals2 data.

reflects the high correlation of the majority of the data. In contrast, the correlation coefficient of Pearson, 0.875, is influenced by the three outliers.

References

Denecke L. (2010). Estimators and tests based on likelihood-depth with application to Weibull distribution, Gaussian and Gumbel copula. *Ph. D. thesis*, University of Kassel, urn:nbn:de:hebis:34-2010101234699.

Denecke, L. and Müller, Ch.H. (2011). Robust estimators and tests for copulas based on likelihood depth. *Computational Statistics and Data Analysis* 55, 2724-2738.

Denecke L. and Müller, Ch.H. (2012). Consistency and robustness of tests and estimators based on depth. To appear in *J. Statist. Plann. Inference*.

Hu, Y., Wang, Y., Wu, Y., Li, Q. and Hou, C. (2009). Generalized Mahalanobis depth in the reproducing kernel Hilbert space. *Statistical Papers* 52, 511-522.

Li, J. and Liu, R.Y. (2008). Multivariate spacings based on data depth: I. Construction of nonparametric multivariate tolerance regions. *Ann. Statist.* 36, 1299-1323.

Lin, L. and Chen, M. (2006). Robust estimating equation based on statistical depth. *Statistical Papers* 47, 263-278.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *J. Am. Stat. Assoc.* 104, 718-734.

López-Pintado, S., Romo, J. and Torrente, A. (2010). Robust depth-based tools for the analysis of gene expression data. *Biostatistics* 11, 254-264.

Mizera, I. (2002). On depth and deep points: a calculus. *Ann. Statist.* 30, 1681-1736.

Mizera, I. and Müller, Ch.H. (2004). Location-Scale Depth. *J. Am. Stat. Assoc.* 99, 949-989.

Müller, Ch.H. (2005). Depth estimators and tests based on the likelihood principle with applications to regression. *J. Multivariate Anal.* 95, 153-181.

Rousseeuw, P.J. and Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* 94, 388-433.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, 57, New York.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes, With Applications to Statistics*. Springer, New York.