Computational Statistics and Data Analysis 55 (2011) 2724–2738

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust estimators and tests for bivariate copulas based on likelihood depth

Liesa Denecke *, Christine H. Müller

Chair of Statistics with Application in Engineering Sciences, Faculty of Statistics, TU Dortmund, 44221 Dortmund, Germany

ARTICLE INFO

Article history:
Received 19 November 2009
Received in revised form 29 March 2011
Accepted 5 April 2011
Available online 13 April 2011

Keywords:
Copula
Gaussian copula
Gumbel copula
Data depth
Likelihood depth
Simplicial depth
Parametric estimation
Test
Robustness against contamination

ABSTRACT

Estimators and tests based on likelihood depth for one-parametric copulas are given. For the Gaussian and Gumbel copulas, it is shown that the maximum depth estimators are biased. They can be corrected and the new estimators are robust against contamination. For testing, simplicial likelihood depth is considered. Because of the bias of the maximum depth estimator, simplicial likelihood depth is not a degenerated U-statistic so that easily asymptotic α-level tests can be derived for arbitrary hypotheses. Tests are in particular investigated for the one-sided alternatives. Simulation studies for the Gaussian and Gumbel copulas show that the power of the first test is rather good, but the latter one has to be improved, which is also done here. The new tests are robust against contamination.

1. Introduction

The copula model has a variety of applications because it models dependence structures. For example in finance, in the analysis of credit risks, the insolvency of several debtors at the same time or for insurances the risk of appearance of different claims at the same time have to be modeled to insure solvency of the bank and insurance, respectively, all the time. Copulas can also be used in the simulation of technical production processes to model the occurrence of coupled failures. Some applications of copulas can be found in Aas (2004), Andersen (2005), Cízek et al. (2005) and Dobrić and Schmid (2005) or Malvergne and Sornette (2006). For an introduction to copulas, see for example Joe (1997) or Nelsen (2006).

Different estimation procedures for copulas were introduced. Parametric, semi-parametric and nonparametric methods are proposed. Most of the parametric and semi-parametric methods are two-stage estimations, as presented in Andersen (2005), Genest et al. (1995) and Hoff (2007) or Kim et al. (2007) for example. Here in most cases as a first step the margins are estimated by parametric or nonparametric methods, then an estimation procedure for the parameter of the copula is presented; see also Lawless and Yilmaz (2011). An example for a nonparametric estimation model for the copula is the empirical copula; see e.g. Capéraà et al. (1997). Goodness-of-fit-tests can be found, for e.g. in Dobrić and Schmid (2005), Fermian (2005) or Panchenko (2005).

In this work, we derive estimators and tests for one-parametric two-dimensional copulas via likelihood depth and simplicial likelihood depth. Likelihood depth and simplicial likelihood depth are rather general notions of data depth, and were first used by Mizera and Müller (2004) and Müller (2005). They extended the half space depth of Tukey (1975) and

* Corresponding author. Tel.: +49 2317553114; fax: +49 2317553454.
E-mail addresses: denecke@statistik.tu-dortmund.de, liesadenecke@googlemail.com (L. Denecke).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.04.005

the simplicial depth of Liu (1988, 1990) which led to outlier robust generalizations of the median for multivariate data. They belong to a broad class of depth notions introduced and studied in the last 20 years, see e.g. Rousseeuw and Hubert (1999), Zou and Serfling (2000a,b) and Mizera (2002), and the book of Mosler (2002). Although likelihood depth is based on a parametric approach, it can lead to distribution-free estimators and tests as Mizera and Müller (2004) demonstrated for location-scale estimation and Müller (2005) for regression. Müller (2005) also showed that simplicial likelihood depth is in particular appropriate for testing since it is an U-statistic. Thereby rather general hypotheses can be tested and the resulting tests are outlier robust.

Copulas are often given by distributional assumptions on the form of the copula. These distributional assumptions for the copula will be used here to define likelihood depth and simplicial likelihood depth for copulas. The approach is demonstrated for the Gaussian copula and the Gumbel copula for two dimensions which are based on one parameter only. However, the approach can also be used for other one-parametric copulas.

In Section 2, the basic concepts of likelihood depth are given and specified for the case of one unknown parameter $\theta \in \Theta \subset \mathbb{R}$. Also the maximum likelihood depth estimator is defined. Under regularity conditions it is a consistent estimator for the "deepest point" of the population. But the deepest point of P_θ can be different from θ, in this case the maximum likelihood depth estimator is an asymptotically biased estimator for θ, but a correction of the bias is given. Moreover, the definitions of Gaussian copula and Gumbel copula are given. Section 3 provides the main results for estimating the parameter θ of a Gaussian copula and Gumbel copula via likelihood depth and simplicial likelihood depth. The resulting estimators are biased but can be corrected. They are robust against contamination.

Tests for general hypotheses about the parameter θ are derived in Section 4.1 via simplicial likelihood depth. Since the maximum likelihood depth estimator is biased, simplicial likelihood depth is not a degenerated U-statistic as is the case for most simplicial depth notions. Hence its asymptotic distribution is simply the normal distribution so that asymptotic α-level tests can be derived easily. Simulation studies show that these tests have a reasonable power for testing $H_0 : \theta \le \theta_0$ for the Gaussian copula and the Gumbel copula. In particular, the test for the Gaussian copula parameter ρ is as powerful as the classical Fisher–Samiuddin test. But the power is bad for testing $H_0 : \theta \ge \theta_0$ because of the bias of the underlying estimator. Therefore an improvement of the tests is proposed which leads to rather powerful tests. All new tests show also high robustness against contamination.

2. Preliminaries

2.1. Likelihood depth and related estimators

Let $Z_1, \ldots, Z_N$ be i.i.d. with density f_θ, $\theta \in \Theta \subset \mathbb{R}^q$. The likelihood function at parameter θ and observation z_n will be denoted by $L(\theta, z_n) := f_\theta(z_n)$. Now we are able to define global likelihood depth similar to Mizera (2002), Mizera and Müller (2004) and Müller (2005):

Definition 1. The global likelihood depth of a parameter θ within observations $z_1, \ldots, z_N$ is the minimal number m of $z_{i_1}, \ldots, z_{i_m}$, such that θ is a likelihood nonfit within $\{z_1, \ldots, z_N\} \setminus \{z_{i_1}, \ldots, z_{i_m}\}$, which means, one can find $\theta' \neq \theta$ such that $L(\theta', z_n) > L(\theta, z_n)$ for every $z_n \in \{z_1, \ldots, z_N\} \setminus \{z_{i_1}, \ldots, z_{i_m}\}$.

In large datasets the calculation of global likelihood depth can be complicated. Mizera (2002) and Mizera and Müller (2004) defined tangent likelihood depth and Müller (2005) introduced simplicial likelihood depth, which are easier to handle.

Definition 2.

(i) Tangent likelihood depth of θ within $z_* := (z_1, \ldots, z_N)^T$ is

$$d_T(\theta, z_*) := \frac{1}{N} \inf_{u \neq 0} \sharp\{n;\, u^T h_n'(\theta) \le 0\}$$

where $h_n(\theta) := \ln(L(\theta, z_n))$ and $h_n'(\theta)$ is the vector of the partial derivatives of $h_n(\theta)$ for $\theta = (\theta_1, \ldots, \theta_q)$ (especially for $\theta \in \mathbb{R}$, $h_n'(\theta) = \frac{\partial}{\partial\theta} \ln f_\theta(z_n)$).

(ii) Simplicial likelihood depth of θ within observations $z_* := (z_1, \ldots, z_N)^T$ is defined as

$$d_S(\theta, z_*) := \binom{N}{q+1}^{-1} \cdot \sharp\{\{n_1, \ldots, n_{q+1}\} \subset \{1, \ldots, N\};\, d_T(\theta, (z_{n_1}, \ldots, z_{n_{q+1}})) > 0\},$$

where q is the dimension of θ.

(iii) 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.

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