ON A SIMPLE METHOD OF MULTIVARIATE DISTRIBUTION ESTIMATION DEFINED BY COPULAS

Alexander Andronov

Riga Technical University
Kalku 1, Riga, LV-1658, Latvia
E-mail: Aleksandrs.Andronovs@rtu.lv

The paper deals with a simple procedure for an estimation of multivariate distribution defined by copulas. Some efficiency criterion of the estimation is suggested and the distribution optimising these criteria is used as an estimate. The simulation study shows that such an approach gives good estimates.

Keywords: *copulas, estimation, asymptotic unbiasness, gradient method*

1. Introduction

The application of multivariate distributions will entail great difficulties [4, 8]. The so-called *copulas* are often used to avoid the problems [1, 2, 3, 6, 9]. Distribution function $C(u_1,u_2,...,u_d)=P\{U_1\le u_1,U_2\le u_2,...,U_d\le u_d\}$ is called *copula* if the marginal distributions of all components $U_1,U_2,...,U_d$ are uniform on [0, 1]. The following fact is basic [7]: any multivariate distribution function $F(x_1,x_2,...,x_d)=P\{X_1\le x_1,X_2\le x_2,...,X_d\le x_d\}$ can be presented uniquely via distribution function of its component $F_i(x_i)=P\{X_i\le x_i\}$ by corresponding copula *C*:

$$F(x_1,x_2,...,x_n)=C\big(F_1(x_1),F_2(x_2),...,F_d(x_d)\big). \qquad (1)$$

A crucial moment here is an estimation of copula and marginal distributions. We consider this problem for the following assumptions. 1) A finite family of potential copulas $\Psi=\{C_1,C_2,...,C_m\}$ is formed. Note that some copulas can contain unknown vector-values parameters, θ_ν for copulas C_ν.
2) We choose some efficiency criteria (an objective function) of a smoothing (for example, the least squares residual sum).

Now we use the following ordinary procedure. 1) On basis of given samples $X^{(j)}=\{X_{j,1},X_{j,2},...,X_{j,d}\}$, *j* = 1, 2, …, *n*, we estimate marginal distribution functions and get corresponding estimates $\tilde{F}_i\ (x_i)=P^*\{X_i\le x_i\}$, *i* = 1, 2, …, *d*. 2) If it is necessary, unknown parameter θ_ν for each copula C_ν is estimated. 3) The criteria values are calculated for all copulas. The copula with the best criteria value is chosen finally.

Let us concentrate our attention on a concrete copula *C*. We assume, the distribution $F(x_1,x_2,...,x_d)$ has continuous probability density function $f(x_1,x_2,...,x_d)$, with the continuous density function $f_i(x_i)$ for X_i. Then, the copula *C* has the continuous probability density function too:

$$c(u_1,u_2,...,u_d)=f\big(F_1^{-1}(u_1),F_2^{-1}(u_2),...,F_d^{-1}(u_d)\big)\left(\prod_{i=1}^{d}f_i(F_i^{-1}(u_i))\right)^{-1},\ \forall u_i\in[0,\ 1]. \qquad (2)$$

We consider the case when copula *C* depends on unknown (in general, vector-valued) parameter $\theta=(\theta_1,\theta_2,...,\theta_m)\in\Omega$, where Ω is a parametrical space. We wish to estimate this parameter. For that firstly, we estimate distribution function $F_i(x_i)$ for all *i* = 1, 2,…, *d* by using ordinary statistical procedures; let $\tilde{F}_i(x_i)$ means a corresponding estimate. Then, we choose an approximation criterion to be as follows.

For each observation $(x_1^{(j)},x_2^{(j)},...,x_d^{(j)})$, *j* = 1, 2, …, *n*, we calculate a value $F^{(j)}*(x_1^{(j)},x_2^{(j)},...,x_d^{(j)})$ of the empirical distribution at the point $(x_1^{(j)},x_2^{(j)},...,x_d^{(j)})$ that has been calculated without the *j*-th observation.

Further, let $F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) = C_\theta\left(\widetilde{F}_1(x_1^{(j)}), \widetilde{F}_2(x_2^{(j)}), \ldots, \widetilde{F}_d(x_d^{(j)})\right)$ be our hypothetical distribution function value as a function of θ at the same point.

Then, we estimate parameter θ under a condition of a minimization of the objective function

$$R(\theta) = \sum_{j=1}^{n} \omega\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) \psi\left(F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) - F^{(j)} * \left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right)\right), \tag{3}$$

where $\omega(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)})$ is a weighted function,

$\psi(y)$ is some convex function [0, 1] $\to$ [0, 1], such that $\psi(0) = 0$.

Now we have an optimisation task:

to minimize sum (3) with respect to parameter $\theta \in \Omega$. (4)

2. First-Order Necessary Condition for Optimality

To optimise objective function R we use the first-order necessary condition for optimality [5]. We write down an expression for a gradient of (3):

$$\frac{\partial}{\partial \theta} R(\theta) = \sum_{j=1}^{n} \omega\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) \frac{\partial}{\partial z} \psi\left(z - F^{(j)} * \left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right)\right)\Big|_{z = F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right)} \times$$
$$\times \frac{\partial}{\partial \theta} F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right), \tag{5}$$

where

$$\frac{\partial}{\partial \theta} F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) = \frac{\partial}{\partial \theta} C_\theta\left(\widetilde{F}_1\left(x_1^{(j)}\right), \widetilde{F}_2\left(x_2^{(j)}\right), \ldots, \widetilde{F}_d\left(x_d^{(j)}\right)\right). \tag{6}$$

In a partial case, objective function $\psi(y)$ is quadratic: $\psi(y) = y^2$, $\omega(\ldots) = 1$. Then,

$$\frac{\partial}{\partial \theta} R(\theta) = \frac{\partial}{\partial \theta} \sum_{j=1}^{n} \left(F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) - F^{(j)} * \left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right)\right)^2 =$$
$$= 2\sum_{j=1}^{n} \left(F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) - F^{(j)} * \left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right)\right) \frac{\partial}{\partial \theta} F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right).$$

As an estimate of interest $\widetilde{\theta}$ we take the root of an equation

$$\sum_{j=1}^{n} \left(F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) - F^{(j)} * \left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right)\right) \frac{\partial}{\partial \theta} F_\theta\left(x_1^{(j)}, x_2^{(j)}, \ldots, x_d^{(j)}\right) = 0. \tag{7}$$

As we show in Appendix 2, under some regularity conditions on function $F_\theta(x_1, x_2, \ldots, x_d)$ the suggested approach gives asymptotically unbiased estimate of θ.

Now we consider a special case of the Gumbel copula.

3. A Special Case: Gumbel Copula

Gumbel copula with scalar parameter $\theta \in [0, \infty]$ has a form

$$C_\theta(u_1, u_2, \ldots, u_d) = \exp\left(-\left(\sum_{i=1}^{d} (-\ln u_i)^\theta\right)^{\frac{1}{\theta}}\right), \quad u_i \in [0,1] \;\; \forall i. \tag{8}$$

We have such an expression for the derivative respect to parameter θ:

$$\frac{\partial}{\partial \theta} C_\theta(u_1, u_2, ..., u_d) = \left(\sum_{i=1}^{d} (-\ln u_i)^\theta \right)^{\frac{1}{\theta}} \times$$
$$\times \ln\left(\sum_{i=1}^{d} (-\ln u_i)^\theta \right) \left(\theta^{-2} \sum_{i=1}^{d} (-\ln u_i)^\theta \ln(-\ln u_i) \right) \exp\left(\left(-\sum_{i=1}^{d} (-\ln u_i)^\theta \right)^{\frac{1}{\theta}} \right). \tag{9}$$

Further, we will use the quadratic form of criteria *R* and the expression (7), where

$\frac{\partial}{\partial \theta} F_\theta\left(x_1^{(j)}, x_2^{(j)}, ..., x_d^{(j)}\right)$ is calculated by (6) and (9).

Now we can use a gradient method for the optimisation of the quadratic objective function *R*.

4. A Simulation Study

We investigate an efficiency of the suggested approach by using simulation. For that we fix two-dimensional Gumbel copula (d = 2) with known scalar parameter θ and normal distributions with parameters μ_i, σ_i, i = 1, 2, as the marginal distributions. We estimate parameters θ and { μ_i, σ_i } on base of n-sample $X^{(j)} = \{X_{j,1}, X_{j,2}\}$, j = 1, 2, …, n, using the suggested procedures.

A simulation study works as follows. We perform r simulation runs. Each run contains the following steps. We generate sequences $X_{j,1}, X_{j,2}$, j = 1, 2, … ,n, of random variables with prescribed distributions and parameters θ, μ_i, σ_i (see Appendix 1). Then, for each i = 1, 2 we compute estimates μ_i^*, σ_i^* on base of sample $X_{(i)} = \{X_{1,i}, X_{2,i}, ..., X_{n,i}\}$ and get the estimate of distribution function $\widetilde{F}_i\ (x) = P^*\{X_i \le x\}$. Further, we realize a gradient procedure and find estimate $\widetilde{\theta}$ that minimizes *the* sum (3). Repeating such a run r times we get a sequence of the estimates $\{\widetilde{\theta}_1, \widetilde{\theta}_2, ..., \widetilde{\theta}_r\}$. It allows us to evaluate such properties of estimate $\widetilde{\theta}$ as the bias, the variance and so on.

Let us consider the following data: $\mu_1 = 1,\ \ \mu_2 = 2,\ \ \sigma_1 = 0.2,\ \ \sigma_2 = 0.5$. As θ parameter's true values θ_0 have been considered the following: θ_0 = 1.5, 2, 2.5 and 3. The sample size n = 600. The received graphics (see Figure 1) show how the sum (3) depends on parameter θ if a true value of θ is θ_0. We have the following estimates $\widetilde{\theta}$ for θ_0 : $\widetilde{\theta}$ = 1.64 for θ_0 = 1.5, $\widetilde{\theta}$ = 1.92 for θ_0 = 2, $\widetilde{\theta}$ = 2.59 for θ_0 = 2.5, $\widetilde{\theta}$ = 2.92 for θ_0 = 3. We can conclude that the sum's (3) minimum gives a well estimate $\widetilde{\theta}$ of θ_0.

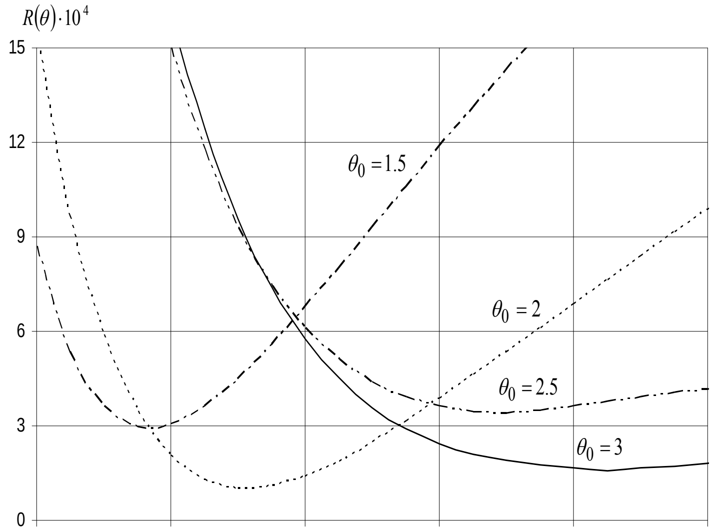

Fig. 1. *Realization of the Objective Function R(θ)*

Further, $\tilde{\theta}$ tends to θ_0 when sample size *n* increases (see Figure 2, where θ_0 = 2): $\tilde{\theta}$ = 2.14 for *n* = 100, $\tilde{\theta}$ = 1.58 for *n*= 200, $\tilde{\theta}$ = 1.88 for *n* = 300, $\tilde{\theta}$ = 2.10 for θ_0 = 400.

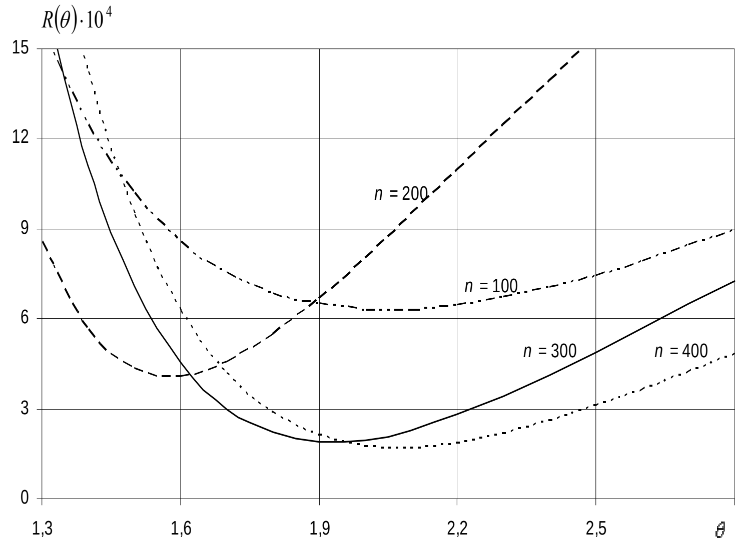

Fig. 2. *Realization of the Objective Function R(θ)*

If a sample size *n* is small then the estimates are worse. So, Figure 3 shows a scattering plot for $\tilde{\theta}$ when θ_0 = 1.5, *n* = 20 and the value of $\tilde{\theta}$ is calculated for 15 various samples with numbers *k* = 1, 2, …, 15.

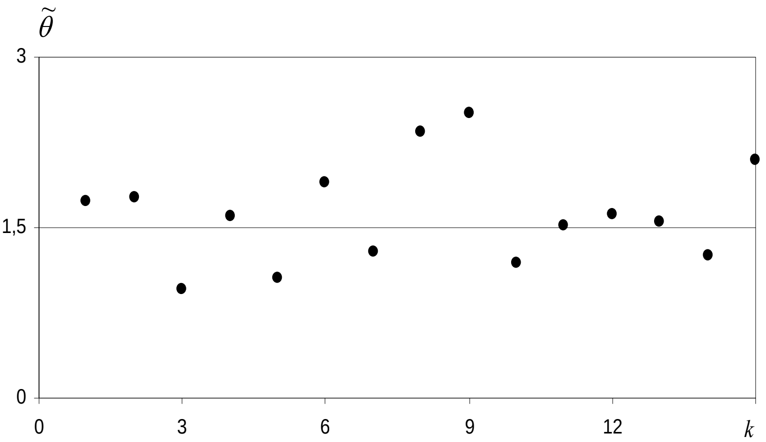

Fig. 3. *Scattering Plot of the Estimate* $\tilde{\theta}$ *for* $\theta_0 = 1.5$

5. Conclusions

A simple method for an estimation of multivariate distributions is described. It is supposed that the distributions are defined by copulas. An efficient of the suggested method has been investigated both analytically and by simulation. The analytical result states that the suggested method gives asymptotically unbiased estimate. The simulation study states the same result for the case of Gumbel copula.

Appendix 1. Random number generation for Gumbel copula

We use a well-known approach [2, 3, 6, 9] for generating random variables $X_1, X_2, ..., X_d$ with the distribution (1). At first, random variables $U_1, U_2, ..., U_d$ with the distribution C are generated. Then, we compute $X_1, X_2, ..., X_d$ from an equation $U_i = F_i(X_i),\ i = 1,2,...,d$.

So, we need to generate $U_1, U_2, ..., U_d$. The usual way is the following. Let $C_k(u_1, u_2, ..., u_k) = C(u_1, u_2, ..., u_k, 1,\ 1,\ ...,\ 1)$ be the marginal distribution of *k* first components $U_1, U_2, ..., U_k$, *k* < *d*. Obviously, $C_1(u) = u$, $\frac{\partial}{\partial u} C_1(u) = 1$, *0* < *u* < 1. Let $C_k(u_k | u_1, u_2, ..., u_{k-1})$ be a conditional distribution of U_k on condition $\{U_1 = u_1, U_2 = u_2, ..., U_{k-1} = u_{k-1}\}$. Those distributions are calculated recurrently.

We write down the corresponding procedure for Gumbel copula (8). At first,

$$C_2(u_2 | u_1) = \frac{\partial}{\partial u_1} C(u_2, u_1) / \frac{\partial}{\partial u_1} C_1(u_1) =$$

$$= \exp\left(-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta \right)^{\frac{1}{\theta}} \right) \theta^{-1} \left((-\ln u_1)^\theta + (-\ln u_2)^\theta \right)^{\frac{1}{\theta}-1} \theta (-\ln u_1)^{\theta-1} \frac{1}{u_1} =$$

$$= \frac{1}{u_1} \exp\left(-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta \right)^{\frac{1}{\theta}} \right) \left((1 + \left(\frac{\ln u_2}{\ln u_1} \right)^\theta \right)^{\frac{1}{\theta}-1}.$$

Therefore,

$$C_2(u_2 | u_1) = \frac{1}{u_1} \exp\left(-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta \right)^{\frac{1}{\theta}} \right) \left((1 + \left(\frac{\ln u_2}{\ln u_1} \right)^\theta \right)^{\frac{1}{\theta}-1}, \quad 0 \le u_2 \le 1. \tag{10}$$

Also, the generation procedure is as such. We generate two independent uniformly distributed on [0, 1] random variables U_1 and *R*. Further, with respect to U_2, it is necessary to solve the equation $C_2(U_2 | U_1) = R$, i.e.

$$R = \frac{1}{U_1} \exp\left(-\left((-\ln U_1)^\theta + (-\ln U_2)^\theta \right)^{\frac{1}{\theta}} \right) \left(1 + \left(\frac{\ln U_2}{\ln U_1} \right)^\theta \right)^{\frac{1}{\theta}-1}$$

or

$$\ln R = \left(-(-\ln U_1) \left(1 + \left(\frac{\ln U_2}{\ln U_1} \right)^\theta \right)^{\frac{1}{\theta}} \right) + \frac{1-\theta}{\theta} \ln\left(1 + \left(\frac{\ln U_2}{\ln U_1} \right)^\theta \right) - \ln U_1.$$

Let $\mathrm{V} = 1 + \left(\frac{\ln U_2}{\ln U_1} \right)^\theta$, then we have the equation

.

$$\ln R = \left(-V^{\frac{1}{\theta}} (-\ln U_1) \right) + \frac{1-\theta}{\theta} \ln(V) - \ln U_1,$$

$$V = \exp\left\{ \frac{\theta}{1-\theta} \left(\ln(R U_1) - V^{\frac{1}{\theta}} \ln U_1 \right) \right\}.$$

If the solution *V* has been solved, it is possible to find U_2 :

$$U_2 = \exp((V-1)^{\frac{1}{\theta}} \ln U_1). \tag{11}$$

Now we can repeat the above-considered approach to generate random variable U_3. At first, the conditional probability density function of U_2 has a form:

$$\frac{\partial}{\partial u_2}C_2(u_2|u_1)=\frac{\partial}{\partial u_2}\exp\left(-\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}}\right)\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}-1}(-\ln u_1)^{\theta-1}\frac{1}{u_1}=$$

$$=(-\ln u_1)^{\theta-1}\frac{1}{u_1}\exp\left(-\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}}\right)\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}-1}\times$$

$$\times\left((-\ln u_2)^{\theta-1}\frac{1}{u_2}\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}-1}+\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{-1}(\theta-1)(-\ln u_2)^{\theta-1}\frac{1}{u_2}\right)=$$

$$=(-\ln u_1)^{\theta-1}(-\ln u_2)^{\theta-1}\frac{1}{u_1u_2}\exp\left(-\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}}\right)\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}-2}\times$$

$$\times\left(\left((-\ln u_1)^\theta+(-\ln u_2)^\theta\right)^{\frac{1}{\theta}}+(\theta-1)\right).$$

Therefore

$$\frac{\partial}{\partial u_2}C_2(u_2|u_1)=\frac{1}{u_1u_2}\exp\left(-\left(\sum_{i=1}^{2}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}}\right)\left((-\ln u_1)^{\theta-1}(-\ln u_2)^{\theta-1}\right)\times$$
$$\times\left(\sum_{i=1}^{2}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}-2}\left(\left(\sum_{i=1}^{2}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}}+\theta-1\right). \qquad (12)$$

The conditional distribution functions of U_2 and U_3 have the following form:

$$C_{2,3}(u_2,u_3|u_1)=\frac{1}{u_1}\exp\left(-\left(\sum_{i=1}^{3}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}}\right)\left(\left(\sum_{i=1}^{3}(-\ln(u_i))^\theta\right)^{\frac{1}{\theta}-1}\left((-\ln u_1)^{\theta-1}\right)\right). \qquad (13)$$

It allows us to calculate the conditional distribution function of U_3 under given U_1 and U_2 :

$$C_3(u_3|u_2,u_1)=\frac{\partial}{\partial u_2}C_{2,3}(u_2,u_3|u_1)/\frac{\partial}{\partial u_2}C_2(u_2|u_1)=$$

$$\frac{1}{u_1u_2}\exp\left(-\left(\sum_{i=1}^{3}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}}\right)\left((-\ln u_1)^{\theta-1}(-\ln u_2)^{\theta-1}\right)\times$$

$$\times\left(\sum_{i=1}^{3}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}-2}\left(\left(\sum_{i=1}^{3}(-\ln u_i)^\theta\right)^{\frac{1}{\theta}}+\theta-1\right)\left(\frac{\partial}{\partial u_2}C_2(u_2|u_1)\right)^{-1}.$$

Also the generation procedure of U_3 is as such. Let

$$W=\left(\sum_{i=1}^{3}(-\ln U_i)^\theta\right)^{\frac{1}{\theta}}.$$

We need, with respect to *W*, to solve the equation $\widetilde{C}_3(W|U_2,U_1)=R$, where

$$\widetilde{C}_3(W|U_2,U_1)=\frac{1}{U_1U_2}\exp(-W)\left((-\ln U_1)^{\theta-1}(-\ln U_2)^{\theta-1}\right)W^{1-2\theta}(W+\theta-1)\left(\frac{\partial}{\partial_2}C_2(U_2|U_1)\right)^{-1}.$$

Further,

$$W^{\theta}=\sum_{i=1}^{3}\left(-\ln U_i)^{\theta}\right) \quad ,-\ln U_3=\left(W^{\theta}-\sum_{i=1}^{2}\left(-\ln U_i)^{\theta}\right)\right)^{\frac{1}{\theta}}.$$

Finally, we have

$$U_3=\exp\left\{-\left(W^{\theta}-\sum_{i=1}^{2}\left(-\ln U_i)^{\theta}\right)\right)^{\frac{1}{\theta}}\right\}.$$

Appendix 2. Asymptotic unbiasness of the estimate for univariate parameter θ

Theorem. Let θ_0 be a true value of parameter θ of the distribution function $F_{\theta}(x_1,x_2,...,x_d)$ and $\widetilde{\theta}$ be an estimate of θ_0 that can be computed as a root of equation (7) on base of *n*-sample. If $F_{\theta}(x_1,x_2,...,x_d)$ is twice continuously differentiable with respect to θ on open set Ω $(\theta_0\in\Omega)$ and the expectation $E\left(\left(\frac{\partial}{\theta}F_{\theta}\left(X_1^{(j)},X_2^{(j)},...,X_d^{(j)}\right)\right)^2\right)$ exists for $\theta\in\Omega$, then

$$E\widetilde{\theta}-\theta_0=O\left(\frac{1}{n}\right).$$

Proof. By doing Taylor series expansion around θ_0, we have for $\theta\in\Omega$:

$$F_{\theta}(x_1,...,x_d)=F_{\theta_0}(x_1,...,x_2)+(\theta-\theta_0)\frac{\partial}{\partial\theta}F_{\theta_0}(x_1,...,x_d)+o(\theta-\theta_0).$$

According to (7), we must find a root of the equation

$$0=\sum_{j=1}^{n}\left(F_{\theta}\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)-F^{(j)}*\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)\right)\frac{\partial}{\partial\theta}F_{\theta}\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)=$$
$$=\sum_{j=1}^{n}\left(F_{\theta_0}\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)+(\theta-\theta_0)\frac{\partial}{\partial\theta}F_{\theta_0}\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)-F^{(j)}*\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)\right)\times$$
$$\times\frac{\partial}{\partial\theta}F_{\theta}\left(x_1^{(j)},x_2^{(j)},...,x_d^{(j)}\right)+o(\theta-\theta_0).$$

As an estimate of interest $\widetilde{\theta}$ we take the root of this equation. We wish to investigate a bias of this estimate. For that aim we use notation *X* for a random variable instead of *x*. A covariance between total estimate $\widetilde{\theta}$ and single member $F_{\theta}(X_1^{(j)},X_2^{(j)},...,X_d^{(j)})$ is very small, precisely, of order $O\left(\frac{1}{n}\right)$. Take an expectation with respect to the distribution F_{θ_0} we get:

$$E\sum_{j=1}^{n}\left(\widetilde{\theta}-\theta_0\right)\left(\frac{\partial}{\partial\theta}F_{\theta_0}\left(X_1^{(j)},...,X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta}F_{\widetilde{\theta}}\left(X_1^{(j)},...,X_d^{(j)}\right)\right)=$$
$$=\sum_{j=1}^{n}\left(E\left(\widetilde{\theta}-\theta_0\right)E\left(\frac{\partial}{\partial\theta}F_{\theta_0}\left(X_1^{(j)},...,X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta}F_{\widetilde{\theta}}\left(X_1^{(j)},...,X_d^{(j)}\right)\right)+O\left(\frac{1}{n}\right)\right)=$$
$$=\left(E\widetilde{\theta}-\theta_0\right)\sum_{j=1}^{n}E\left(\frac{\partial}{\partial\theta}F_{\theta_0}\left(X_1^{(j)},...,X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta}F_{\widetilde{\theta}}\left(X_1^{(j)},...,X_d^{(j)}\right)\right)+O(1).$$

Note, that the expectation under the sum exists according to theorem's conditions and Schwarz inequality

$$\left(E\left(\left(\frac{\partial}{\theta} F_{\theta_0}\left(X_1^{(j)}, X_2^{(j)}, ..., X_d^{(j)}\right)\right)\left(\frac{\partial}{\theta} F_{\theta}\left(X_1^{(j)}, X_2^{(j)}, ..., X_d^{(j)}\right)\right)\right)\right)^2 \leq$$

$$\leq E\left(\left(\frac{\partial}{\theta} F_{\theta_0}\left(X_1^{(j)}, X_2^{(j)}, ..., X_d^{(j)}\right)\right)^2\right) E\left(\left(\frac{\partial}{\theta} F_{\theta}\left(X_1^{(j)}, X_2^{(j)}, ..., X_d^{(j)}\right)\right)^2\right).$$

Further,

$$\left(E\tilde{\theta} - \theta_0\right)\sum_{j-1}^{n} E\left(\frac{\partial}{\partial\theta} F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right) + O(1) =$$

$$= \sum_{j-1}^{n} E\left\{\left(F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right) - F^{(j)} * \left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right\} + E\left(o\left(\tilde{\theta} - \theta_0\right)\right). \tag{14}$$

Let $x^{(j)} = \left(x_1^{(j)}, x_2^{(j)}, ..., x_d^{(j)}\right)$, $X^{(j)} = \left(X_1^{(j)}, X_2^{(j)}, ..., X_d^{(j)}\right)$. We have

$$F^{(j)} * \left(x^{(j)}\right) = \frac{1}{n-1}\sum_{i \neq j} \delta\left(X^{(i)}, x^{(j)}\right),$$

where

$$\delta\left(x^{(i)}, x^{(j)}\right) = \begin{cases} 1 \; if \; x^{(i)} \leq x^{(j)}, \\ 0 \; otherwise. \end{cases}$$

Further,

$$E\left(F^{(j)} * \left(X^{(j)}\right) \middle| X^{(j)} = x^{(j)}\right) = \frac{1}{n-1} E\left(\sum_{i \neq j} \delta\left(X^{(i)}, x^{(j)}\right)\right) = \frac{1}{n-1}\left(\sum_{i \neq j} F_{\theta_0}\left(x^{(j)}\right)\right) = F_{\theta_0}\left(x^{(j)}\right).$$

Therefore,

$$E\left\{\left(F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right) - F^{(j)} * \left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right\} =$$

$$= EE\left\{\left(F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right) - F^{(j)} * \left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right) \middle| X^{(j)} = x^{(j)}\right\} =$$

$$= E\left\{\left(F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right) - E\left[F^{(j)} * \left(X_1^{(j)}, ..., X_d^{(j)}\right) \middle| X^{(j)} = x^{(j)}\right]\right)\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right\} =$$

$$= E\left\{\left(F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right) - F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right\} = 0.$$

Now, we keep in mind that all random vectors $X^{(j)} = \left(X_1^j, ..., X_d^j\right)$ are independent and identically distributed, therefore instead of (14) we have

$$\left(E\tilde{\theta} - \theta_0\right)\sum_{j-1}^{n} E\left(\frac{\partial}{\partial\theta} F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right) + O(1) = E\left(o\left(\tilde{\theta} - \theta_0\right)\right),$$

$$\left(E\tilde{\theta} - \theta_0\right) nE\left(\frac{\partial}{\partial\theta} F_{\theta_0}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)}, ..., X_d^{(j)}\right)\right) = O(1) + E\left(o\left(\tilde{\theta} - \theta_0\right)\right).$$

Finally,

$$E\tilde{\theta} - \theta_0 = \frac{O(1)}{nE\left(\left(\frac{\partial}{\partial\theta} F_{\theta_0}\left(X_1^{(j)},\dots,X_d^{(j)}\right)\right)\left(\frac{\partial}{\partial\theta} F_{\tilde{\theta}}\left(X_1^{(j)},\dots,X_d^{(j)}\right)\right)\right)} = O\left(\frac{1}{n}\right). \quad \square$$

References

1. Abegaz, F., Naik-Nimbalkar, U.V. Modelling statistical dependence of Markov chains via copula models, *Journal of Statistical Planning and Inference*, **138**, 2008, pp. 1131-1146.
2. Embrechts, P. Some Selected Papers – http://www.math.ethz.ch/~baltes/ftp/papers.html
3. Embrechts, P., Lindskog, F., McNeil, A. *Modelling Dependence with Copulas and Applications to Risk Management (Chapter 8): Handbook of Heavy Tailed Distributions in Finance.* Amsterdam: Elsevier, 2003, pp. 329-384.
4. Gentlel, J. E. *Elements of Computational Statistics*. New York: Springer, 2002. 420 p.
5. Nocedal, J., Wright, S. J. *Numerical optimisation, 2nd edition.* New York: Springer, 2006. 664 p.
6. Pettere, G. Stochastic Risk Capital Model for Insurance Company. In: *Applied Stochastic Models and Data Analysis: Proceedings of the XII-th International Conference, Chania, Crete, Greece, May 29, 30, 31 and June 1, 2007.* Chania, Greece, 2007.
7. Sklar, A. Fonctions de répartition à *n* dimensions et leurs marges, *Publ. Inst Statist. Univ. Paris,* Vol. 8, 1959, pp. 229-231.
8. Srivastava, M.S. *Methods of Multivariate Statistics, 2nd edition.* New York: John Wiley and Sons Inc, 2002. 698 p.
9. The 8th Tartu Conference on MULTIVARIATE STATISTICS; The 6th Conference on MULTIVARIATE DISTRIBUTIONS – http://www.ms.ut.ee/tartu07/presentations