A nonparametric dependence measure for random variables based on the one-to-one correspondence between Copulas and Markov operators

Trutschnig, Wolfgang
European Centre for Soft Computing
Research Unit for Intelligent Data Analysis and Graphical Models
Edificio de Investigación, Calle Gonzalo Gutiérrez Quirós S/N
33600 Mieres, Spain
wolfgang.trutschnig@softcomputing.es

Introduction

Considering the uniform distance d_∞ on the space $\mathcal{C}$ of two-dimensional copulas yields a compact metric space $(\mathcal{C}, d_\infty)$ in which the family of shuffles of the minimum copula M are dense (see [5], [10], [11]). If $A \in \mathcal{C}$ is a shuffle of M, μ_A denotes the corresponding doubly stochastic measure and X, Y are random variables on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with $\mathcal{P}^{X \otimes Y} = \mu_A$, then X and Y are mutually completely dependent (see [11]) and knowing X implies knowing Y and vice versa. Consequently the product copula Π (describing complete unpredictability) can be approximated arbitrary well by mutually completely predictable copulas with respect to d_∞. In other words, d_∞ does not 'distinguish between different types of statistical dependence' (see [10]) and dependence measures which are continuous w.r.t. d_∞ like Schweizer and Wolff's σ (see [11] and [14]) seem somehow unnatural.
Using the one-to-one correspondence between copulas and Markov operators on $L^1([0,1], \mathcal{B}([0,1]), \lambda)$ allows to consider the topology $\mathcal{O}_\mathcal{M}$ on $\mathcal{C}$ which is induced by the strong operator topology on the space $\mathcal{M}$ of Markov operators (see [3], [10], [12]). Since the topology that the weak operator topology on $\mathcal{M}$ induces on $\mathcal{C}$ coincides with the topology induced by d_∞ (see [12]) it is straightforward to see that $\mathcal{O}_\mathcal{M}$ is finer than $\mathcal{O}_{d_\infty}$. Rewriting the Markov operators in terms of regular conditional distributions (Markov kernels) a L^1-type metric D_1 on $\mathcal{C}$ based on the conditional distribution functions was defined in [15] and shown to have (amongst others) the following properties:

(P1) $D_1(A, B) \leq 1/2$ for all $A, B \in \mathcal{C}$
(P2) D_1 is a metrization of $\mathcal{O}_\mathcal{M}$
(P3) The metric space $(\mathcal{C}, D_1)$ is complete and separable
(P4) $D_1(A, \Pi) \leq 1/3$ for all $A \in \mathcal{C}$ with equality if and only if A is a deterministic[1] copula

Hence, according to point (P4), in contrast to d_∞, all 'deterministic' copulas have maximum D_1-distance to Π and Π can not be approximated by such copulas w.r.t. D_1. The dependence measure $\tau_1 : \mathcal{C} \to [0,1]$ induced by D_1 is therefore naturally defined by $\tau_1(A) := 3\, D_1(A, \Pi)$.
In the current paper we will take a look to the L^2-versions of D_1 and τ_1 and show that the new metric D_2 and the new dependence measure τ_2 also exhibit various good properties (similar to the ones of D_1 and τ_1 respectively). Before doing so we will collect some notation and preliminaries in the next section.

Notation and preliminaries

Throughout the paper $\mathcal{C}$ will denote the family of all *two-dimensional copulas*. For every copula $A \in \mathcal{C}$ the corresponding doubly stochastic measure will be denoted by μ_A, the family of all these μ_A

[1] a copula supported only on the graph of a Lebesgue-measure-preserving transformation S on $[0,1]$, see Definition 1

by $\mathcal{P}_\mathcal{C}$. M will denote the minimum copula, Π the product copula and W the lower Fréchet-Hoeffding bound. For properties of copulas see [6] and [11]. d_∞ will denote the uniform metric on $\mathcal{C}$, i.e.

$$d_\infty(A,B) := \max_{(x,y)\in[0,1]^2} |A(x,y) - B(x,y)|.$$

For every $d \geq 1$ $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-field in $\mathbb{R}^d$, $\mathcal{B}([0,1]^d)$ the Borel σ-field in $[0,1]^d$, and λ^d the d-dimensional Lebesgue measure. In case $d = 1$ we will also simply write $\mathcal{B}([0,1])$ and λ. If X, Y are real-valued random variables on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ then we will write $\mathcal{P}^{X\otimes Y}$ for their joint distribution and $\mathcal{P}^X, \mathcal{P}^Y$ for the distributions of X and Y. $\mathbf{E}(Y|X)$ will denote the *conditional expectation of* Y *given* X. Since by definition $\mathbf{E}(Y|X)$ is $\mathcal{A}_\sigma(X)$-measurable there exists a measurable function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbf{E}(Y|X) = g \circ X$ holds $\mathcal{P}$-almost surely; we will write $\mathbf{E}(Y|X = x) = g(x)$ and call g a *version of the conditional expectation of* Y *given* X. A *Markov kernel* from $\mathbb{R}$ to $\mathcal{B}(\mathbb{R})$ is a mapping $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ such that $x \mapsto K(x,B)$ is measurable for every fixed $B \in \mathcal{B}(\mathbb{R})$ and $B \mapsto K(x,B)$ is a probability measure for every fixed $x \in \mathbb{R}$. A Markov kernel $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ is called *regular conditional distribution of* Y *given* X if for every $B \in \mathcal{B}(\mathbb{R})$

$$K(X(\omega), B) = \mathbb{E}(\mathbf{1}_B \circ Y|X)(\omega) \tag{1}$$

holds $\mathcal{P}$-a.s. It is well known that for each pair (X,Y) of real-valued random variables a regular conditional distribution $K(\cdot,\cdot)$ of Y given X exists, that $K(\cdot,\cdot)$ is unique $\mathcal{P}^X$-a.s. (i.e. unique for $\mathcal{P}^X$-almost all $x \in \mathbb{R}$) and that $K(\cdot,\cdot)$ only depends on $\mathcal{P}^{X\otimes Y}$. Hence, given $A \in \mathcal{C}$ we will denote (a version of) the regular conditional distribution of Y given X by $K_A(\cdot,\cdot)$ and refer to $K_A(\cdot,\cdot)$ simply as *regular conditional distribution of* A. Note that for every $A \in \mathcal{C}$, its conditional regular distribution $K_A(\cdot,\cdot)$, and Borel sets $E, F \in \mathcal{B}([0,1])$ we have

$$\int_F K_A(x,E)\, d\lambda(x) = \mu_A(F \times E), \tag{2}$$

so in particular

$$\int_{[0,1]} K_A(x,E)\, d\lambda(x) = \lambda(E). \tag{3}$$

For more details and properties of conditional expectation and regular conditional distributions see [8], [9], [1], [2].
A linear operator T on $L^1([0,1], \mathcal{B}([0,1]), \lambda)$ is called *Markov operator* (see [3],[10], [12]) if it fulfils the following three properties:

1. T is positive, i.e. $T(f) \geq 0$ whenever $f \geq 0$
2. $T(\mathbf{1}_{[0,1]}) = \mathbf{1}_{[0,1]}$
3. $\int_{[0,1]} (Tf)(x) d\lambda(x) = \int_{[0,1]} f(x) d\lambda(x)$

The class of all Markov operators on $L^1([0,1], \mathcal{B}([0,1]), \lambda)$ will be denoted by $\mathcal{M}$. It is straightforward to see that the operator norm of T is one, i.e. $\|T\| := \sup\{\|Tf\|_1 : \|f\|_1 \leq 1\} = 1$ holds. According to [3] and [12] *there is a one-to-one correspondence between* $\mathcal{C}$ *and* $\mathcal{M}$ - in fact, the mappings $\Phi : \mathcal{C} \to \mathcal{M}$ and $\Psi : \mathcal{M} \to \mathcal{C}$, defined by

$$\begin{aligned} \Phi(A)(f)(x) \ &: = (T_A f)(x) := \frac{d}{dx} \int_{[0,1]} A_{,2}(x,t) f(t) d\lambda(t), \\ \Psi(T)(x,y) \ &: = A_T(x,y) := \int_{[0,x]} (T\mathbf{1}_{[0,y]})(t) d\lambda(t) \end{aligned} \tag{4}$$

for every $f \in L^1([0,1])$ and $(x,y) \in [0,1]^2$ ($A_{,2}$ denoting the partial derivative w.r.t. y), fulfil $\Psi \circ \Phi = id_\mathcal{C}$ and $\Phi \circ \Psi = id_\mathcal{M}$. Note that in case of $f := \mathbf{1}_{[0,y]}$ we have $(T_A \mathbf{1}_{[0,y]})(x) = A_{,2}(x,y)$ λ-a.s. (the a.s.

existence of the partial derivative follows from the fact that for every fixed y the mapping $x \mapsto A(x, y)$ is absolutely continuous since copulas are Lipschitz continuous, see [11], [13], [7]). According to [10] $T_A f$ is a version of the conditional expectation of $f \circ Y$ given X, i.e.

$$(T_A f)(x) = \mathbb{E}(f \circ Y | X = x) \tag{5}$$

holds λ-a.s. The metric D_1 mentioned in the Introduction and its induced dependence measure $\tau : \mathcal{C} \to [0, 1]$ are defined by

$$D_1(A, B) := \int_{[0,1]} \int_{[0,1]} |K_A(x, [0, y]) - K_B(x, [0, y])| d\lambda(x)\, d\lambda(y) \tag{6}$$

$$\tau_1(A) := 3\, D_1(A, \Pi). \tag{7}$$

The metric D_2 and its induced dependence measure τ_2

Throughout the rest of this paper we will consider the L^2-version D_2 of D_1, which is defined as follows:

$$D_2^2(A, B) := \int_{[0,1]} \int_{[0,1]} (K_A(x, [0, y]) - K_B(x, [0, y]))^2 d\lambda(x)\, d\lambda(y) \tag{8}$$

It is straightforward to show that D_2 is a metric on $\mathcal{C}$. Hölder's inequality implies $D_1(A, B) \leq D_2(A, B)$ for all $A, B \in \mathcal{C}$, so the topology D_2 induced by D_2 on $\mathcal{C}$ is at least as fine as the one induced by D_1. Furthermore obviously $D_2^2(A, B) \leq D_1(A, B)$ holds, so altogether we have

$$D_1(A, B) \leq D_2(A, B) \leq \sqrt{D_1(A, B)}, \tag{9}$$

which shows that D_1 and D_2 induce the same topology on $\mathcal{C}$. As a consequence, using the facts about D_1 mentioned in the Introduction, D_2 is a metrization of the topology $\mathcal{O}_{\mathcal{M}}$ too.
To simplify notation we will write

$$\Psi_{A,B}(y) := \int_{[0,1]} (K_A(x, [0, y]) - K_B(x, [0, y]))^2 d\lambda(x) \tag{10}$$

for all $A, B \in \mathcal{C}$. The following first result holds:

Lemma 1 *For every pair $A, B \in \mathcal{C}$ the function $\Psi_{A,B}$, defined according to (10), is Lipschitz continuous with Lipschitz constant 4 and fulfils $\Psi_{A,B}(y) \leq \min\{2y, 2(1 - y)\}$ for every $y \in [0, 1]$. Moreover there exist copulas $A, B \in \mathcal{C}$ for which equality $\Psi_{A,B}(y) = \min\{2y, 2(1 - y)\}$ holds for all $y \in [0, 1]$.*

Proof: Suppose that $E \in \mathcal{B}([0, 1])$, then using (3) and applying Scheffé's theorem (see [4]) we get

$$\begin{aligned} \int_{[0,1]} (K_A(x, E) - K_B(x, E))^2 d\lambda(x) &\leq \int_{[0,1]} |K_A(x, E) - K_B(x, E)| d\lambda(x) \\ &= 2 \int_G K_A(x, E) - K_B(x, E)\, d\lambda(x) \\ &\leq 2 \int_{[0,1]} K_A(x, E)\, d\lambda(x) = 2\lambda(E) \end{aligned}$$

whereby $G = \{x \in [0, 1] : K_A(x, E) > K_B(x, E)\}$. Since $K_A(\cdot, E^c) = 1 - K_A(\cdot, E)$ holds, considering $E = [0, y]$ implies the desired inequality. Straightforward calculations show that in case of the copulas M and W we get $\Psi_{M,W}(y) = \min\{2y, 2(1 - y)\}$ for every $y \in [0, 1]$.

Finally, to see Lipschitz continuity, suppose that $s > t$, then

$$\begin{aligned}|\Psi_{A,B}(s) - \Psi_{A,B}(t)| &\leq \Big| \int_{[0,1]} (K_A(x,(0,s]) - K_B(x,(0,s]))^2 \, d\lambda(x) - \\ &\qquad \int_{[0,1]} (K_A(x,(0,t]) - K_B(x,(0,t]))^2 \, d\lambda(x) \Big| \\ &\leq 2 \int_{[0,1]} |K_A(x,(t,s]) - K_B(x,(t,s])| \, d\lambda(x) \leq 4(s-t). \blacksquare\end{aligned}$$

Definition 1 *A copula $A \in \mathcal{C}$ is called* deterministic *(or* completely dependent*) if there exists a λ-preserving transformation $S : [0,1] \to [0,1]$ such that $K(x,E) := \mathbf{1}_E(Sx) = \delta_{Sx}(E)$ is a regular conditional distribution of A. The class of all deterministic copulas will be denoted by $\mathcal{C}_d$.*

Remark 1 *It is easy to see that Definition 1 is equivalent to the condition that μ_A has support only on the graph of S.*

As mentioned before D_1 and D_2 induce the same topology - nevertheless the following lemma holds:

Lemma 2 *D_1 and D_2 are not equivalent metrics.*

Proof: We will start by calculating $D_1(A_1, A_2)$ and $D_2(A_1, A_2)$ for $A_1, A_2 \in \mathcal{C}_d$. Suppose that S_1, S_2 are the corresponding λ-preserving transformations, then

$$\begin{aligned}D_2^2(A_1,A_2) &= \int_{[0,1]} \int_{[0,1]} (\mathbf{1}_{[0,y]}(S_1x) - \mathbf{1}_{[0,y]}(S_2x))^2 d\lambda(x) d\lambda(y) \\ &= \int_{[0,1]} \int_{[0,1]} |\mathbf{1}_{[0,y]}(S_1x) - \mathbf{1}_{[0,y]}(S_2x)| d\lambda(x) d\lambda(y) = D_1(A_1,A_2) \\ &= \|S_1 - S_2\|_1,\end{aligned}$$

which, in particular, implies that the second inequality in (9) can not be improved. For every $n \in \mathbb{N}$ define an interval-exchange transformation (see [5]) $S_n : [0,1] \to [0,1]$ as follows (see Figure 1):

$$S_n(x) = \begin{cases} x + (1 - \frac{1}{2^n}) & \text{if } x \in (0\,,\,\frac{1}{2^n}] \\ x - (1 - \frac{1}{2^n}) & \text{if } x \in (1 - \frac{1}{2^n}\,,\,1] \\ x & \text{otherwise} \end{cases}$$

Furthermore let S denote the identity on $[0,1]$ and $M, A_1, A_2 \ldots$ the corresponding deterministic copulas in $\mathcal{C}_d$. Then we get

$$D_2^2(A_n, M) = D_1(A_n, M) = \|S_n - S\|_1 = 2\int_{[0,\frac{1}{2^n}]} \Big(1 - \frac{1}{2^n}\Big) d\lambda(x) = \frac{1}{2^{n-1}}\Big(1 - \frac{1}{2^n}\Big).$$

This shows that the quotient $\frac{D_2(A_n,M)}{D_1(A_n,M)}$ is unbounded in n, so D_1 and D_2 can not be equivalent metrics. $\blacksquare$

Next we will show the D_2-versions of (P2) and (P4) in the Introduction.

Theorem 1 *The metric D_2 only assumes values in $[0, \sqrt{1/2}]$. Furthermore, given $A \in \mathcal{C}$, we have $D_2^2(A, \Pi) \leq 1/6$ with equality if and only if $A \in \mathcal{C}_d$.*

Proof: The proof is easier than the proof of the corresponding result for D_1 (see [15]) since we are working in L^2 instead of L^1. The fact that $D_2^2(A,B) \leq 1/2$ is a direct consequence of Lemma 1. To prove the second part of the theorem we may proceed as follows: Fix $A \in \mathcal{C}$, $y \in [0,1]$ and define a random variable $Z_y : ([0,1], \mathcal{B}([0,1], \lambda) \to [0,1]$ by

$$Z_y(x) := K_A(x, [0,y]).$$

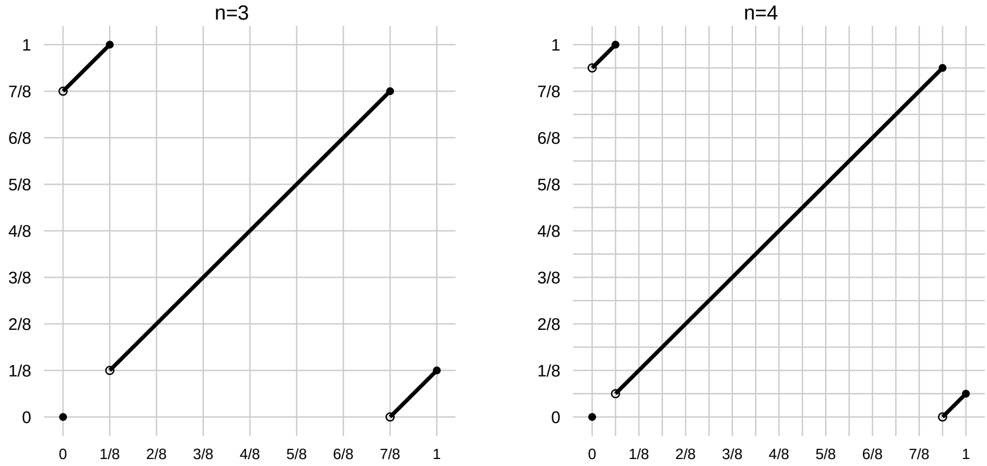

Figure 1: Interval exchange transformations S_n used in the proof of Lemma 2

Then

$$\mathbf{E}(Z_y) = \int_{[0,1]} K_A(x,[0,y])d\lambda(x) = \mu_A([0,1]\times[0,y]) = y$$

as well as

$$\Psi_{A,\Pi}(y) = \int_{[0,1]} (K_A(x,[0,y]) - y))^2 d\lambda(x) = \mathbf{V}(Z_y)$$

follows. Hence, since Z_y is $[0,1]$-valued, $\Psi_{A,\Pi}(y)$ becomes maximal if and only if $Z_y \sim Binomial(y,1)$, in which case $\Psi_{A,\Pi}(y) = y(1-y)$ holds. As a consequence $D_2^2(A,B) \le \int_{[0,1]} y(1-y)d\lambda(y) = 1/6$. Assume now that $D_2^2(A,\Pi) = 1/6$ for some $A \in \mathcal{C}$. We want to show that $A \in \mathcal{C}_d$ holds. Due to (Lipschitz-) continuity we have $\Psi_{A,\Pi}(y) = y(1-y)$ for all $y \in [0,1]$, so $Z_y(x) := K_A(x,[0,y])$ fulfils $Z_y \sim Binomial(y,1)$. For every $y \in [0,1]$ setting $E_y := \{x : Z_y(x) = 1\} \in \mathcal{B}([0,1])$ implies $\lambda(E_y) = y$ as well as $Z_y(x) = K_A(x,[0,y]) = \mathbf{1}_{E_y}(x)$ for λ-almost every $x \in [0,1]$. Consequently we can find a measurable set $M \subseteq [0,1]$ fulfilling $\lambda(M) = 1$ such that for every $x \in M$ we have $K_A(x,[0,y]) = \mathbf{1}_{E_y}(x)$ for every $y \in [0,1] \cap \mathbb{Q}$. Define a transformation $S : [0,1] \to [0,1]$ by

$$Sx := \mathbf{1}_M(x) \inf\{y \in \mathbb{Q} \cap [0,1] : K_A(x,[0,y]) = 1\}.$$

Using right-continuity of distribution functions it follows that on M we have $K_A(x,[0,y_0]) = 1$ if and only if $Sx \le y_0$, i.e. if $\mathbf{1}_{[0,y_0]}(Sx) = 1$. This implies that S is measurable since

$$\{x \in [0,1] : Sx \le y_0\} = M^c \cup \{x \in M : K_A(x,[0,y_0]) = 1\} \in \mathcal{B}([0,1])$$

holds for every $y_0 \in [0,1]$. Furthermore

$$\lambda^S([0,y_0]) = \lambda(\{x \in [0,1] : K_A(x,[0,y_0]) = 1)\} = \lambda(E_{y_0}) = y_0,$$

so S is also λ-preserving. Since on M $K_A(x,[0,y_0]) = \mathbf{1}_{[0,y_0]}(Sx) = \delta_{Sx}([0,y_0])$ holds we have $K_A(x,E) = \delta_{Sx}(E)$ for every Borel set E which shows that $(x,E) \mapsto \delta_{Sx}(E)$ is a regular conditional distribution of A. Finally, if $A \in \mathcal{C}_d$ is a deterministic copula with corresponding λ-preserving transformation S then it follows that

$$D_2^2(A,\Pi) = \int_{[0,1]}\int_{[0,1]} (\mathbf{1}_{[0,y]}(Sx) - y)^2 d\lambda(x)\, d\lambda(y) = \int_{[0,1]}\int_{[0,1]} (\mathbf{1}_{[0,y]}(x) - y)^2 d\lambda(x)\, d\lambda(y) = 1/6.$$

This completes the proof. ■

Remark 2 *Another possibility to prove Theorem 1 would be to embed $\mathcal{C}$ in $L^2([0,1]^2, \mathcal{B}([0,1]^2), \lambda^2)$ via the mapping $\epsilon : A \mapsto (H_A : (x,y) \mapsto K_A(x,[0,y]))$, show that $\epsilon(\mathcal{C})$ is a closed convex subset of $L^2([0,1]^2, \mathcal{B}([0,1]^2), \lambda^2)$ and that the extreme points of $\epsilon(\mathcal{C})$ are exactly the deterministic copulas.*

Using Theorem 1 we can finally define the dependence measure $\tau_2 : \mathcal{C} \to [0,1]$ by

$$\tau_2(A) := \sqrt{6}\, D_2(A, \Pi), \qquad A \in \mathcal{C}. \tag{11}$$

Remark 3 *Looking at the definition of D_2 the dependence measure $\tau_2(A)$ can, up to a scalar, be interpreted as expected L^2-distance between the conditional distribution function of A and the distribution function of the uniform distribution $\mathcal{U}_{[0,1]}$.*

Reformulating Theorem 1 in terms of τ_2 immediately yields

Proposition 1 *Suppose that $A \in \mathcal{C}$ and let τ_2 be defined according to (11). Then $\tau_2(A) \in [0,1]$. Furthermore $\tau_2(A) = 1$ if and only if $A \in \mathcal{C}_d$, i.e. it is exactly the class of deterministic copulas that is assigned maximum dependence measure.*

We have already discussed all D_2-versions of the points (P1)-(P4) except (P3). Nevertheless, the proof for the corresponding result for D_1 can be modified easily to show that the metric space $(\mathcal{C}, D_2)$ is separable and complete too.

References

[1] R. Bhattacharya, E.C. Waymire: *A Basic Course in Probability Theory*, Springer Verlag New York 2007

[2] H. Bauer: *Wahrscheinlichkeitstheorie*, de Gruyter, Berlin New York 2002 (Fifth Edition)

[3] W.F. Darsow, B. Nguyen, E.T. Olsen: Copulas and Markov processes, *Illinois Journal of Mathematics*, Volume 36, Number 4, 600-642 (1992)

[4] L. Devroye: *A Course in Density Estimation*, Birkhäuser, Boston Basel Stuttgart, 1987

[5] F. Durante, P. Sarkoci, C. Sempi: Shuffles of copulas, *Journal of Mathematical Analysis and Applications*, 352, 914-921 (2009)

[6] F. Durante, C. Sempi: Copula theory: an introduction, in P. Jaworski, F. Durante, W. Härdle, T. Rychlik, Eds., *Copula theory and its applications, Lecture Notes in Statistics*–Proceedings, vol 198, pp. 1–31, Springer, Berlin (2010)

[7] E. Hewitt, K. Stromberg: *Real and Abstract Analysis*, Springer Verlag, Berlin Heidelberg, 1965

[8] O. Kallenberg: *Foundations of modern probability*, Springer Verlag, New York Berlin Heidelberg, 1997

[9] A. Klenke: *Probability Theory - A Comprehensive Course*, Springer Verlag Berlin Heidelberg 2007

[10] X. Li, P. Mikusinski, M.D. Taylor: Strong approximation of copulas, *Journal of Mathematical Analysis and Applications* **255**, 608-623 (1998)

[11] R.B. Nelsen: *An Introduction to Copulas*, Springer, New York, 2006

[12] E.T. Olsen, W.F. Darsow, B. Nguyen: Copulas and Markov operators, in *Proceedings of the Conference on Distributions with Fixed Marginals and Related Topics, IMS Lecture Notes*, Monograph Series Vol. 28, pp. 244-259 (1996)

[13] W. Rudin: *Real and Complex Analysis*, McGraw-Hill International Editions, Singapore, 1987

[14] B. Schweizer, E. F. Sklar: On nonparametric measures of dependence for random variables, *Annals of Statistics*, Volume 9, No. 4, 879-885 (1981)

[15] W. Trutschnig: On a strong metric on the space of copulas and its induced dependence measure, submitted to the *Journal of Mathematical Analysis and Applications*, current status: under review