ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

An invitation to coupling and copulas: With applications to multisensory modeling

Hans Colonius

University of Oldenburg, Oldenburg, Germany

HIGHLIGHTS

- Gives an introduction to the stochastic concepts of coupling and copula.
- Demonstrates their role in modeling reaction times.
- Presents an example from modeling multisensory integration.
- Gives pointers to the advanced literature on coupling and copula and to the relevant multisensory literature.

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Coupling
Copula
Multisensory
Race model
Time window of integration model

ABSTRACT

This paper presents an introduction to the stochastic concepts of *coupling* and *copula*. Coupling means the construction of a joint distribution of two or more random variables that need not be defined on one and the same probability space, whereas a copula is a function that joins a multivariate distribution to its one-dimensional margins. Their role in stochastic modeling is illustrated by examples from multisensory perception. Pointers to more advanced and recent treatments are provided.

1. Introduction

The concepts of *coupling* and *copula* refer to two related areas of probability and statistics whose importance for mathematical psychology has arguably been ignored so far. This paper gives an elementary, non-rigorous introduction to both concepts. Moreover, applications of both concepts to modeling in a multisensory context are presented. Briefly, *coupling* means the construction of a joint distribution of two or more random variables that need not be defined on one and the same probability space, whereas a *copula* is a function that joins a multivariate distribution to its one-dimensional margins.

The theory of copulas has stirred a lot of interest in recent years in several areas of statistics, including finance, mainly for the following reasons (see e.g., Joe, 2015): it allows one (i) to study the structure of stochastic dependency in a "scale-free" manner, i.e., independent of the specific marginal distributions, and (ii) to construct families of multivariate distributions with specified properties. We will demonstrate in the final section how copulas can be used to test models of multisensory integration and to derive measures of the amount of multisensory integration occurring in a given context.

In psychology, a need for the theory of coupling emerges from the following observations. Empirical data are typically collected under various experimental conditions, e.g. speed vs. accuracy instructions in a reaction time (RT) experiment or different numbers of targets or nontargets in a visual search paradigm. Data collected under a particular experimental condition are considered realizations of some random variable with respect to an underlying probability (sample) space. For example, let T_k be the random variable denoting the time to detect a target in the presence of k nontargets (distractors). Now consider random variable T_{k+1}, the time to detect a target in the presence of $k + 1$ targets. Whenever response times T_k and T_{k+1} are collected in different trials (or blocks of trials), there is no empirical coupling scheme, so these two random variables are "stochastically unrelated", i.e., they do not refer to outcomes defined on one and the same underlying sample space.

Alternatively, if the target is presented first under k and then under $k + 1$ nontargets, one immediately following the other, and responses are recorded in pairs at any given trial n, this imposes a coupling on T_k and T_{k+1}, that is the existence of a bivariate distribution with marginal distributions equal to the distributions of T_k and T_{k+1}. As explicated in the theory of "contextuality-by-default" being developed by E. N. Dzhafarov and J. Kujala

E-mail address: hans.colonius@uni-oldenburg.de.
URL: http://www.uni-oldenburg.de/en/hans-colonius/.

http://dx.doi.org/10.1016/j.jmp.2016.02.004

Nomenclature

$\mathbb{R}$	Set of real numbers
Ω	Set of elementary outcomes ω
$\mathcal{F}$	Sigma-algebra
$\mathbb{P}$	Probability measure on $(\Omega, \mathcal{F})$
$\mathcal{B}$	Borel sigma algebra
P	Probability on $(\mathbb{R}, \mathcal{B})$
$\overset{d}{\leq}$	Dominated in distribution
a.s.	Almost surely
$X \leq Y$	$X(\omega) \leq Y(\omega)$ for all ω (X, Y random variables)
$\mathbf{a} \leq \mathbf{b}$	$a_i \leq b_i$ for $i = 1, \ldots, n$
Ran$F$	Range (image) of function F
$\bar{F}$	Survival distribution corresponding to distribution F
$\tilde{C}$	Dual of copula C
C^*	Co-copula of copula C
δ_C	Diagonal section of copula C
ρ	Spearman's rho
τ	Kendall's tau
r	Pearson's linear correlation
$\overset{d}{=}$	Equal in distribution
Cov	Covariance
E	Expected value

(see Dzhafarov and Kujala, 2016, in this issue) such a coupling represents an empirically defined meaning of "togetherness" of the responses.

Note that even if no "natural" coupling exists because responses cannot co-occur, a coupling satisfying some desirable properties can always be constructed. This will be illustrated in Section 4 with an example from multisensory modeling, coupling unisensory reaction times to visual and auditory stimuli presented in separate blocks of trials. Moreover, stochastic unrelatedness does not rule out numerically comparing averaged data under separate conditions, or even the entire distributions functions. Specifically, as shown in the next section, one can turn a statement about an ordering of two stochastically unrelated distributions into a statement about point-wise ordering of the corresponding random variables on a common probability space.

2. Coupling

We begin with a few common definitions.[1] Let X and Y be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in $(\mathbb{R}, \mathcal{B}, P)$. "Equality" of random variables can be interpreted in different ways. Random variables X and Y are *equal in distribution* iff they are governed by the same probability measure:

$$X \overset{d}{=} Y \iff P(X \in A) = P(Y \in A), \quad \text{for all } A \in \mathcal{B}.$$

X and Y are *point-wise equal* iff they agree for almost all elementary events[2]:

$$X \overset{a.s.}{=} Y \iff P(\{\omega \mid X(\omega) = Y(\omega)\}) = 1.$$

Let X be a real-valued random variable with distribution function $F(x)$. Then the *quantile function* of X is defined as

$$Q(u) = F^{-1}(u) = \inf\{x \mid F(x) \geq u\}, \quad 0 \leq u \leq 1. \tag{1}$$

[1] If not indicated otherwise, most of the material on coupling in this section is taken from the monograph by Thorisson (2000).

[2] Here, a.s. is for "almost surely".

Table 1
Joint distribution of two Bernoulli random variables.

		$\hat{X}_q$		
		0	1	
$\hat{X}_p$	0	$1-q$	$q-p$	$1-p$
	1	0	p	p
		$1-q$	q	1

For every $-\infty < x < +\infty$ and $0 < u < 1$, we have

$$F(x) \geq u \quad \text{if, and only if,} \quad Q(u) \leq x.$$

Thus, if there exists x with $F(x) = u$, then $F(Q(u)) = u$ and $Q(u)$ is the smallest value of x satisfying $F(x) = u$. If $F(x)$ is continuous and strictly increasing, $Q(u)$ is the unique value x such that $F(x) = u$.

Moreover, if U is a standard uniform random variable (i.e., defined on $[0, 1]$), then $X = Q(U)$ has its distribution function as $F(x)$; thus, any distribution function can be conceived as arising from the uniform distribution transformed by $Q(u)$.

2.1. Definition and examples

Note to the reader: We enumerate Definitions, Theorems, Examples sequentially, so Definition 1 is followed by Example 2, and so on.

Definition 1. A *coupling* of a collection of random variables $\{X_i, i \in \mathrm{I}\}$, with I denoting some index set, is a family of jointly distributed random variables

$$(\hat{X}_i : i \in \mathrm{I}) \quad \text{such that } \hat{X}_i \overset{d}{=} X_i, \ i \in \mathrm{I}.$$

Note that the joint distribution of the $\hat{X}_i$ need not be the same as that of X_i; in fact, the X_i may not even have a joint distribution because they need not be defined on a common probability space. However, the family $(\hat{X}_i : i \in \mathrm{I})$ has a joint distribution with the property that its marginals are equal to the distributions of the individual X_i variables. The individual $\hat{X}_i$ is also called a *copy of* X_i.

Example 2 (*Coupling Two Bernoulli Random Variables*)**.** Let X_p, X_q be Bernoulli random variables, i.e.,

$$P(X_p = 1) = p \quad \text{and} \quad P(X_p = 0) = 1 - p,$$

and X_q defined analogously. Assume $p < q$; we can couple X_p and X_q as follows:

Let U be a uniform random variable on $[0, 1]$, i.e., for $0 \leq a < b \leq 1$,

$$P(a < U \leq b) = b - a.$$

Define

$$\hat{X}_p = \begin{cases} 1, & \text{if } 0 < U \leq p; \\ 0, & \text{if } p < U \leq 1; \end{cases} \qquad \hat{X}_q = \begin{cases} 1, & \text{if } 0 < U \leq q; \\ 0, & \text{if } q < U \leq 1. \end{cases}$$

Then U serves as a common source of randomness for both $\hat{X}_p$ and $\hat{X}_q$. Moreover, $\hat{X}_p \overset{d}{=} X_p$ and $\hat{X}_q \overset{d}{=} X_q$, and $\mathrm{Cov}(\hat{X}_p, \hat{X}_q) = p(1-q)$. The joint distribution of $(\hat{X}_p, \hat{X}_q)$ is presented in Table 1 and illustrated in Fig. 1.

A simple, though somewhat fundamental, coupling is the following:

Example 3 (*Quantile Coupling*)**.** Let X be a random variable with distribution function F, that is,

$$P(X \leq x) = F(x), \quad x \in \mathbb{R}.$$

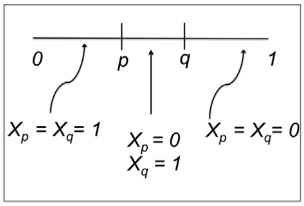

Fig. 1. Joint distribution of two Bernoulli random variables.

Let U be a uniform random variable on $[0, 1]$. Then, for random variable $\hat{X} = F^{-1}(U)$,

$$P(\hat{X} \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x), \quad x \in \mathbb{R},$$

that is, $\hat{X}$ is a copy of X, $\hat{X} \stackrel{d}{=} X$. Thus, letting F run over the class of all distribution functions (using the same U), yields a coupling of all differently distributed random variables, the *quantile coupling*. One can show that quantile coupling consists of positively correlated random variables.

2.2. Strassen's theorem, coupling event inequality, and maximal coupling

An important application of quantile coupling is in reducing a stochastic order between random variables to a pointwise (a.s.) order: Let X and X' be two random variables with distribution functions F and G, respectively. If there is a coupling $(\hat{X}, \hat{X}')$ of X and X' such that $\hat{X}$ is *pointwise dominated* by $\hat{X}'$, that is

$$\hat{X} \leq \hat{X}' \quad \text{(almost surely)},$$

then $\{\hat{X} \leq x\} \supseteq \{\hat{X}' \leq x\}$, which implies

$$P(\hat{X} \leq x) \geq P(\hat{X}' \leq x),$$

and thus

$$F(x) \geq G(x), \quad x \in \mathbb{R}.$$

Then X is said to be *stochastically dominated* (or *dominated in distribution*) by X':

$$X \stackrel{d}{\leq} X'.$$

But one can show (Thorisson, 2000, p. 4) that the other direction also holds: *stochastic* domination implies *pointwise* domination. Thus, we have a (simple) version of *Strassen's theorem* (Strassen, 1965):

Theorem 4 (*Strassen, 1965*)**.** *Let X and X' be random variables. Then*

$$X \stackrel{d}{\leq} X'$$

if, and only if, there is a coupling $(\hat{X}, \hat{X}')$ of X and X' such that a.s.

$$\hat{X} \leq \hat{X}'.$$

The following question is the starting point of many convergence and approximation results obtainable from coupling arguments. Let X and X' be two random variables with non-identical distributions. How can one construct a coupling of X and X', $(\hat{X}, \hat{X}')$, such that $P(\hat{X} = \hat{X}')$ is maximal across all possible couplings? Here we follow the slightly more general formulation in Thorisson (2000) but limit presentation to the case of discrete random variables (the continuous case being completely analogous).

Definition 5. Suppose $(\hat{X}_i : i \in \mathrm{I})$ is a coupling of X_i, $i \in \mathrm{I}$, and let C be an event such that if it occurs, then all the $\hat{X}_i$ coincide, that is,

$$C \subseteq \{\hat{X}_i = \hat{X}_j, \text{ for all } i, j \in \mathrm{I}\}.$$

Such an event is called a *coupling event*.

Assume all the X_i take values in a finite or countable set E with $P(X_i = x) = p_i(x)$, for $x \in E$. For all $i, j \in \mathrm{I}$ and $x \in E$, we clearly have

$$P(\hat{X}_i = x, C) = P(\hat{X}_j = x, C) \leq p_j(x),$$

and thus for all $i \in \mathrm{I}$ and $x \in E$,

$$P(\hat{X}_i = x, C) \leq \inf_{j \in \mathrm{I}} p_j(x).$$

Summing over $x \in E$ yields the basic *coupling event inequality*:

$$P(C) \leq \sum_{x \in E} \inf_{j \in \mathrm{I}} p_j(x). \tag{2}$$

As an example, consider again the case of two discrete random variables X and X' with coupling $(\hat{X}, \hat{X}')$, and set $C = \{\hat{X} = \hat{X}'\}$. Then

$$P(\hat{X} = \hat{X}') \leq \sum_x \min\{P(X = x), P(X' = x)\}. \tag{3}$$

Interestingly, it turns out that, at least in principle, one can always construct a coupling such that the above coupling event inequality (2) holds with identity. Such a coupling is called *maximal* and C a *maximal coupling event*.

Proposition 6 (*Maximal Coupling*)**.** *Suppose X_i, $i \in \mathrm{I}$, are discrete random variables taking values in a finite or countable set E. Then there exists a maximal coupling, that is, a coupling with coupling event C such that*

$$P(C) = \sum_{x \in E} \inf_{i \in \mathrm{I}} p_i(x).$$

Proof. Put

$$c := \sum_{x \in E} \inf_{i \in \mathrm{I}} p_i(x) \quad \text{(the maximal coupling probability)}.$$

If $c = 0$, take the $\hat{X}_i$ independent and $C = \emptyset$. If $c = 1$, take the $\hat{X}_i$ identical and $C = \Omega$, the sample space. For $0 < c < 1$, these couplings are mixed as follows: Let J, V, and W_i, $i \in \mathrm{I}$, be independent random variables such that J is Bernoulli distributed with $P(J = 1) = c$ and, for $x \in E$,

$$P(V = x) = \frac{\inf_{i \in \mathrm{I}} p_i(x)}{c}$$

$$P(W_i = x) = \frac{p_i(x) - c\,P(V = x)}{1 - c}.$$

Define, for each $i \in \mathrm{I}$,

$$\hat{X}_i = \begin{cases} V, & \text{if } J = 1, \\ W_i, & \text{if } J = 0. \end{cases} \tag{4}$$

Then

$$\begin{aligned} P(\hat{X}_i = x) &= P(V = x)P(J = 1) + P(W_i = x)P(J = 0) \\ &= P(X_i = x). \end{aligned}$$

Thus, the $\hat{X}_i$ are a coupling of the X_i, $C = \{J = 1\}$ is a coupling event, and it has the desired value c. □

The representation (4) of the X_i is known as *splitting representation*.

We conclude the treatment of coupling with a variation on the theme of maximal coupling: Given two random variables, what is a measure of closeness between them when an appropriate coupling is applied to make them as close to being identical as possible?

The *total variation distance* between two probability distributions μ and ν on Ω is defined as

$$\|\mu - \nu\|_{TV} := \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|,$$

for all Borel sets A. Thus, the distance between μ and ν is the maximum difference between the probabilities assigned to a single event by the two distributions. Using the coupling inequality, it can be shown that

$$\|\mu - \nu\|_{TV} = \inf\{P(X \neq Y) \mid (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}. \quad (5)$$

A splitting representation analogous to the one in the previous proof assures that a coupling can be constructed so that the infimum is obtained (see Levin, Peres, & Wilmer, 2008, pp. 50–52).

3. Copulas

Frechét (1951) studied the following problem, formulated here for the bivariate case: Given the distribution functions F_X and F_Y of two random variables X and Y defined on the same probability space, what can be said about the class $\mathcal{G}(F_X, F_Y)$ of the bivariate distribution functions whose marginals are F_X and F_Y?

Obviously, the class $\mathcal{G}(F_X, F_Y)$ is non-empty since it contains the case of X and Y being independent. Let $F(x, y)$ be a joint distribution function for (X, Y). To each pair of real numbers (x, y), we can associate three numbers: $F_X(x)$, $F_Y(y)$, and $F(x, y)$. Note that each of these numbers lies in the interval $[0, 1]$. In other words, each pair (x, y) of real numbers is mapped to a point $(F_X(x), F_Y(y))$ in the unit square $[0, 1] \times [0, 1]$, and this ordered pair in turn corresponds to a number $F(x, y)$ in $[0, 1]$.

$$(x, y) \mapsto (F_X(x), F_Y(y)) \mapsto F(x, y) = C(F_X(x), F_Y(y)),$$

$$\mathbb{R} \times \mathbb{R} \longrightarrow [0, 1] \times [0, 1] \xrightarrow{C} [0, 1].$$

Then the mapping C is an example of a copula (it "couples" the bivariate distribution with its marginals).

3.1. Definition, examples, and Sklar's Theorem

A straightforward definition for any finite dimension n is the following:

Definition 7. A function $C : [0, 1]^n \longrightarrow [0, 1]$ is called n-dimensional *copula* if there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a vector of standard uniform random variables $(U_1, \ldots, U_n)$ such that

$$C(u_1, \ldots, u_n) = P(U_1 \le u_1, \ldots, U_n \le u_n),$$
$$u_1, \ldots, u_n \in [0, 1].$$

Clearly, any copula is a distribution function. There is an alternative, analytical definition of copula based on the fact that distribution functions can be characterized as functions satisfying certain conditions, without reference to a probability space.

Definition 8. An n-dimensional copula C is a function on the unit n-cube $[0, 1]^n$ that satisfies the following properties:

1. the range of C is the unit interval $[0, 1]$;
2. $C(\mathbf{u})$ is zero for all $\mathbf{u}$ in $[0, 1]^n$ for which at least one coordinate is zero (groundedness);
3. $C(\mathbf{u}) = u_k$ if all coordinates of $\mathbf{u}$ are 1 except the kth one;
4. C is *n-increasing*, that is, for every $\mathbf{a} \le \mathbf{b}$ in $[0, 1]^n$ ($\le$ defined componentwise) the volume assigned by C to the n-box $[\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is nonnegative.

One can show that groundedness and the n-increasing property are sufficient to define a proper distribution function. Moreover, copulas are *uniformly continuous* and all their *partial derivatives* exist almost everywhere, which is a useful property especially for computer simulations (for proofs see, e.g., Durante & Sempi, 2016).

A very simple copula is the following:

Example 9 (*The Independence Copula*)**.** For independent standard uniform random variables $U_1, \ldots, U_n$ and $\mathbf{U} = (U_1, \ldots, U_n)$

$$P(\mathbf{U} \le \mathbf{u}) = C(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i$$

is a copula, called the *independence copula*.

There are two further copulas of special importance:

Example 10 (*The Comonotonicity Copula*)**.** For U uniformly distributed on $[0, 1]$, consider the random vector $\mathbf{U} = (U_1, \ldots, U_n)$. Then, for any $\mathbf{u} \in [0, 1]^n$,

$$P(\mathbf{U} \le \mathbf{u}) = P(U \le \min\{u_1, \ldots, u_n\}) = \min\{u_1, \ldots, u_n\}$$

is a copula, called the *comonotonicity copula*.

Example 11 (*The Countermonotonicity Copula*)**.** For U uniformly distributed on $[0, 1]$, consider the random vector $\mathbf{U} = (U, 1 - U)$. Then, for any $\mathbf{u} \in [0, 1]^2$,

$$P(\mathbf{U} \le \mathbf{u}) = P(U \le u_1, 1 - U \le u_2) = \max\{0, u_1 + u_2 - 1\}$$

is a copula, called the *countermonotonicity copula*.

The comonotonicity copula is often denoted as

$$M_n(u_1, \ldots, u_n) = \min\{u_1, \ldots, u_n\}$$

and is also called the *upper Fréchet–Hoeffding bound copula*. Similarly, the function

$$W_n(u_1, \ldots, u_n) = \max\{u_1 + \cdots + u_n - (n - 1), 0\}$$

is called the *lower Fréchet–Hoeffding bound copula* for $n = 2$, but it is not a copula for $n > 2$. The reason for the latter statement is that W_n for $n \ge 3$ is in general not a proper distribution function (see below Section 3.5). Importantly, any copula obeys the *Fréchet–Hoeffding bounds*:

Theorem 12 (*Frechét, 1951*)**.** *If C is any n-dimensional copula, then for every $\mathbf{u} \in [0, 1]^n$,*

$$W_n(\mathbf{u}) \le C(\mathbf{u}) \le M_n(\mathbf{u}).$$

Proof. See, e.g., Durante and Sempi (2016), p.27.

Although the Fréchet–Hoeffding lower W_n is never a copula for $n \ge 3$, it is the best possible lower bound in the following sense:

Theorem 13. *For any $n \ge 3$ and any $\mathbf{u} \in [0, 1]^n$, there is an n-dimensional copula (which depends on $\mathbf{u}$) such that*

$$C(\mathbf{u}) = W_n(\mathbf{u}).$$

For the proof, see Nelsen (2006), p.48. The following famous theorem laid the foundation of many subsequent studies (for a proof, see Nelsen, 2006, Theorem 2.10.9).

Theorem 14 (*Sklar's Theorem, 1959*)**.** *Let $F(x_1, \ldots, x_n)$ be an n-variate distribution function with margins $F_1(x_1), \ldots, F_n(x_n)$; then there exists an n-copula $C : [0, 1]^n \longrightarrow [0, 1]$ that satisfies*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

If all univariate margins $F_1, \ldots, F_n$ are continuous, then the copula is unique. Otherwise, C is uniquely determined on $\text{Ran}F_1 \times \text{Ran}F_2 \times \cdots \times \text{Ran}F_n$.

If $F_1^{-1}, \ldots, F_n^{-1}$ are the quantile functions of the margins, then for any $(u_1, \ldots, u_n) \in [0, 1]^n$

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)).$$

Please cite this article in press as: Colonius, H., An invitation to coupling and copulas: With applications to multisensory modeling. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.004

Copulas with discrete margins have also been defined, but their treatment is less straightforward (for a review, see Pfeifer & Nešlehová, 2004). Sklar's theorem shows that copulas remain invariant under strictly increasing transformations of the underlying random variables. It is possible to construct a wide range of multivariate distributions by separately choosing the marginal distributions and a suitable copula.

Example 15 (*Bivariate Exponential*)**.** For $\delta > 0$, the distribution

$$F(x, y) = \exp\{-[e^{-x} + e^{-y} - (e^{\delta x} + e^{\delta y})^{-1/\delta}]\},\quad -\infty < x, y, < +\infty,$$

with margins $F_1(x) = \exp\{-e^{-x}\}$ and $F_2(y) = \exp\{-e^{-y}\}$ corresponds to the copula

$$C(u, v) = uv \exp\{[(-\log u)^{-\delta} + (-\log v)^{-\delta}]^{-1/\delta}\},$$

an example of the class of *bivariate extreme value copulas* characterized by $C(u^t, v^t) = C^t(u, v)$, for all $t > 0$.

3.2. Copula density and pair copula constructions (vines)

If the probability measure associated with a copula C is absolutely continuous (with respect to the Lebesgue measure on $[0, 1]^n$), then there exists a *copula density* $c : [0, 1]^n \longrightarrow [0, \infty]$ almost everywhere unique such that

$$C(u_1, \ldots, u_n) = \int_0^{u_1} \cdots \int_0^{u_n} c(v_1, \ldots, v_n)\, dv_n \ldots dv_1,\quad u_1, \ldots, u_n \in [0, 1].$$

Such an absolutely continuous copula is n times differentiable and

$$c(u_1, \ldots, u_n) = \frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_n} C(u_1, \ldots, u_n),\quad u_1, \ldots, u_n \in [0, 1].$$

For example, the independence copula is absolutely continuous with density equal to 1:

$$\Pi(u_1, \ldots, u_n) = \prod_{k=1}^{n} u_k = \int_0^{u_1} \cdots \int_0^{u_n} 1\, dv_n \ldots dv_1.$$

When the density of a distribution $F_{12}(x_1, x_2)$ exists, it can be written as

$$f_{12}(x_1, x_2) = f_1(x_1) f_2(x_2)\, c_{12}(F_1(x_1), F_2(x_2)),$$

with c_{12} the copula density of the copula of $F_{12}(x_1, x_2)$. This equation shows how independence is "distorted" by copula density c whenever c is different from 1. Moreover, this yields an expression for the conditional density of X_1 given $X_2 = x_2$:

$$f_{1|2}(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1). \tag{6}$$

This is the starting point of a recent, important approach to constructing high-dimensional dependency structures from pairwise dependencies ("vine copulas"). Note that a multivariate density of dimension n can be decomposed as follows, here taking the case for $n = 3$:

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1).$$

Applying the decomposition in Eq. (6) to each of these terms yields,

$$\begin{aligned} f_{2|1}(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \\ f_{3|12}(x_3|x_1, x_2) &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) f_{3|2}(x_3|x_2) \\ f_{3|2}(x_3|x_2) &= c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3), \end{aligned}$$

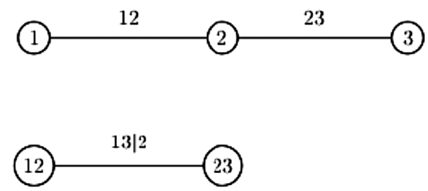

Fig. 2. Graphical illustration of the decomposition in Eq. (7). A line in the graph corresponds to the indices of a copula linking two distributions, unconditional in the upper graph, conditional in the lower graph.

resulting in the "regular vine tree" representation

$$\begin{aligned} f(x_1, x_2, x_3) &= f_3(x_3) f_2(x_2) f_1(x_1) \quad \text{(marginals)} \\ &\times c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \\ &\quad \text{(unconditional pairs)} \\ &\times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \quad \text{(conditional pair)}. \end{aligned} \tag{7}$$

In order to visualize this structure, in particular for larger n, one defines a sequence of trees (acyclic undirected graphs), a simple version of it is depicted in Fig. 2.

3.3. Survival copula, co-copula, dual and diagonal section of a copula

Whenever it is more convenient to describe the multivariate distribution of a random vector $(X_1, \ldots, X_n)$ by means of its survival distribution, i.e.,

$$\bar{F}(x_1, \ldots, x_n) = P(X_1 > x_1, \ldots, X_n > x_n),$$

its *survival copula* can be introduced such that the analog of Sklar's theorem holds, with

$$\bar{F}(x_1, \ldots, x_n) = \bar{C}(\bar{F}_1(x_1), \ldots, \bar{F}_n(x_n)),$$

where the $\bar{F}_i$, $i = 1, \ldots, x_n$, are the marginal survival distributions. In the continuous case there is a one-to-one correspondence between the copula and its survival copula. For $n = 2$, this is

$$C(u_1, u_2) = \bar{C}(1 - u_1, 1 - u_2) + u_1 + u_2 - 1.$$

For the general case, we refer to (Mai & Scherer, 2012, pp. 20–21).

Two other functions closely related to copulas and survival copulas are useful in the response time modeling context. The *dual of a copula* C is the function $\tilde{C}$ defined by $\tilde{C}(u, v) = u + v - C(u, v)$ and the *co-copula* is the function C^* defined by $C^*(u, v) = 1 - C(1 - u, 1 - v)$. Neither of these is a copula, but when C is the (continuous) copula of a pair of random variables X and Y, the dual of the copula and the co-copula each express a probability of an event involving X and Y:

$$\tilde{C}(F_X(x), F_Y(y)) = P(X \le x \text{ or } Y \le y)$$

and

$$C^*(1 - F_X(x), 1 - F_Y(y)) = P(X > x \text{ or } Y > y).$$

Finally, for standard uniform random variables $U_1, \ldots, U_n$ and the corresponding copula $C(u_1, \ldots, u_n)$, its *diagonal section* is defined as $\delta_C(u) = C(u, \ldots, u)$. Durante and Sempi (2016, p. 69) state necessary and sufficient conditions for a function $\delta : [0, 1] \to [0, 1]$ to be the diagonal section of some copula. Later, the concept of diagonal section will be central for the example presented in Section 4.

3.4. Copulas with singular components

If the probability measure associated with a copula C has a singular component, then the copula also has a singular component which can often be detected by finding points $(u_1, \ldots, u_n) \in [0, 1]^n$, where some (existing) partial derivative of the copula has a point of discontinuity. A standard example is the comonotonicity copula

$$M_n(u_1, \ldots, u_n) = \min(u_1, \ldots, u_n),$$

where the partial derivatives have a point of discontinuity;

$$\frac{\partial}{\partial u_k} M_n(u_1, \ldots, u_n) = \begin{cases} 1, & \text{if } u_k < \min(u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n), \\ 0, & \text{if } u_k > \min(u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n). \end{cases}$$

The probability measure associated with $M_n(u_1, \ldots, u_n)$ assigns all mass to the diagonal of the unit n-cube $[0, 1]^n$ ("perfect positive dependence").

3.5. Copulas and extremal dependence

Here we take a closer look at how copulas relate to stochastic dependency. For $n = 2$, Theorem 12 reduces to

Example 16 (*Fréchet–Hoeffding Copula*)**.** Let $C(u, v)$ be a 2-copula; then, for $u, v \in [0, 1]$,

$$W_2(u, v) \equiv \max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} \equiv M_2(u, v),$$

and M and W are also copulas, the *upper* and *lower Fréchet–Hoeffding copula*.

We have seen (Examples 10 and 11) that M_2 and W_2 are bivariate distribution functions of the random vectors (U, U) and $(U, 1 - U)$, respectively, where U is a standard uniform random variable. In this case, we say that M_2 (comonotonicity copula) describes *perfect positive dependence* and W_2 (countermonotonicity copula) describes *perfect negative dependence*. For U and V standard uniform random variables whose joint distribution is the copula M_2, then $P(U = V) = 1$; and if the copula is W_2, then $P(U + V = 1) = 1$.

If X and Y are random variables with joint distribution function $H(x, y)$ and margins $F_X(x)$ and $F_Y(y)$, then it is easy to show (e.g. Joe, 2015, p. 47) that, for all $x, y \in \mathbb{R}$,

$$\max\{F_X(x) + F_Y(y) - 1, 0\} \leq H(x, y) \leq \min\{F_X(x), F_Y(y)\} \quad (8)$$

$F^-(x, y) = \max\{F_X(x) + F_Y(y) - 1, 0\}$ is called the *lower* and $F^+(x, y) = \min\{F_X(x), F_Y(y)\}$ the *upper Fréchet–Hoeffding bound*, respectively, or *Fréchet bound*, for short. What can be said about random variables X and Y when their joint distribution H equals one of its Fréchet–Hoeffding bounds?

If both margins F_X and F_Y are continuous then, by Sklar's theorem, the copulas corresponding to H are unique and the Fréchet bounds F^+ and F^- represent perfect positive and negative dependence, respectively. In the discrete case, the bounds can also sometimes represent perfect dependence, but not with any generality (see Example 2.9 in Joe, 2015).

3.6. Linear measures of dependence

The most widely known and used dependence measure is Pearson's linear correlation,

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(X)}}.$$

Given finite variances, it is a measure of linear dependence that takes values in the range $[-1, 1]$. An obvious disadvantage of r in the context of copulas, in addition to requiring finite variances, is that it depends on the marginal distributions; thus, it is not invariant under strictly increasing nonlinear transformations of the variables, whereas copulas are. Embrechts, McNeil, and Straumann (2002) mention two "pitfalls" in dealing with linear correlation: (1) assuming that marginal distributions and correlation determine the joint distributions, and (2) assuming that, given marginal distributions F_X and F_Y, all linear correlations between -1 and 1 can be attained through suitable specification of the joint distribution of X and Y. Counterexamples to both assumptions abound in the copula literature (e.g., Embrechts, Lindskog, & McNeil, 2003; Embrechts et al., 2002; Joe, 2015; Nelsen, 2006).

Proposition 17 (*Hoeffding, 1940*)**.** *Let X and Y have finite (nonzero) variances with an unspecified dependence structure. Then*

1. *the set of possible linear correlations is a closed interval $[r_{\min}, r_{\max}]$ and for the extremal correlations $r_{\min} < 0 < r_{\max}$ holds;*
2. *the extremal correlation $r = r_{\min}$ is attained if and only if X and Y are countermonotonic; $r = r_{\max}$ is attained if and only if X and Y are comonotonic.*
3. *$r_{\min} = -1$ if and only if X and $-Y$ are of the same type and $r_{\max} = 1$ if and only if X and Y are of the same type (X and Y are* ofthesametype *if we can find $a > 0$ and $b \in \mathbb{R}$ so that $Y \stackrel{d}{=} aX + b$).*

The proof of this proposition presented in Embrechts et al., 2002, (p. 24) starts by recalling the Fréchet–Hoeffding bounds (Eq. (8)),

$$F^- = \max\{F_X(x) + F_Y(y) - 1, 0\} \leq H(x, y) \leq \min\{F_X(x), F_Y(y)\} = F^+.$$

Inserting the upper and the lower bound into Hoeffding's identity (e.g. Shea, 1983)

$$\text{Cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [H(x, y) - F_X(x) F_Y(y)] \, dx \, dy, \quad (9)$$

immediately gives the set of possible correlations (see Embrechts et al., 2002, (ibid.)) for the complete proof of the proposition).

With $0 \leq \lambda \leq 1$, the mixture $\lambda F^- + (1 - \lambda) F^+$ has correlation

$$r = \lambda r_{\min} + (1 - \lambda) r_{\max}$$

and can be used to construct joint distributions with marginals F_X and F_Y and with arbitrary correlations $r \in [r_{\min}, r_{\max}]$.

Finally, the following example shows that small (linear) correlations cannot be interpreted as implying weak dependence between random variables.

Example 18 (*Embrechts et al., 2002*)**.** Let X be distributed as Lognormal$(0, 1)$ and Y as Lognormal$(0, \sigma^2)$, $\sigma > 0$. Note that X and Y are not of the same type although $\log X$ and $\log Y$ are. From Proposition 17, comonotonicity and countermonotonicity require that $r_{\max} = r(e^Z, e^{\sigma Z})$ and $r_{\min} = r(e^Z, e^{-\sigma Z})$, respectively, with Z standard-normally distributed. Analytic calculation then yields

$$r_{\min} = \frac{e^{-\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}$$

and

$$r_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}.$$

Letting $\sigma \to \infty$, both correlations converge to zero.

Thus, it is possible to have a random vector (X, Y) where the correlation is almost zero, even though X and Y are comonotonic or countermonotonic and therefore have the strongest kind of dependency possible.

Please cite this article in press as: Colonius, H., An invitation to coupling and copulas: With applications to multisensory modeling. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.004

3.7. Copula-based measures of dependence

There are several alternatives to the linear correlation coefficient when the latter is inappropriate or misleading. Two important ones are *Kendall's tau* and *Spearman's rho*.

For a vector (X, Y), Kendall's tau is defined as

$$\tau(X, Y) = P[(X - \tilde{X})(Y - \tilde{Y}) > 0] - P[(X - \tilde{X})(Y - \tilde{Y}) < 0],$$

where $(\tilde{X}, \tilde{Y})$ is an independent copy of (X, Y). Hence Kendall's tau is simply the probability of concordance minus the probability of discordance. When (X, Y) is continuous with copula C, then it can be shown (Nelsen, 2006, p. 159) that

$$\tau(X, Y) = \tau_C = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1.$$

Note that the integral above can be interpreted as the expected value of the function $C(U, V)$ of the standard uniform random variables U and V whose joint distribution is C, i.e.,

$$\tau_C = 4\,\mathrm{E}[C(U, V)] - 1.$$

In the case of three known bivariate distribution functions, Kendall's tau gives a necessary condition for compatibility, i.e., for the existence of a trivariate distribution function with the given bivariate marginals. Specifically,

Proposition 19 (*Joe, 1997, p. 76*)**.** *Let $F \in \mathcal{G}(F_{12}, F_{13}, F_{23})$, the class of trivariate distributions with marginals F_{12}, F_{13}, F_{23} and suppose F_{jk}, $j < k$, are continuous. Let $\tau_{jk} = \tau_{kj}$ be the value of Kendall's tau for F_{jk}, $j \neq k$. Then the inequality*

$$-1 + |\tau_{ij} + \tau_{jk}| \leq \tau_{ik} \leq 1 - |\tau_{ij} - \tau_{jk}|$$

holds for all permutations (i, j, k) of $(1, 2, 3)$ and the bounds are sharp.

Thus, if the above inequality does not hold for some (i, j, k), then the three bivariate margins are not compatible. Sharpness follows from the special trivariate normal case: Kendall's tau for the bivariate normal is $\tau = (2/\pi) \arcsin(\rho)$, so that the inequality becomes

$$-\cos\left(\tfrac{1}{2}\pi(\tau_{12} + \tau_{23})\right) \leq \sin\left(\tfrac{1}{2}\pi\tau_{13}\right) \leq \cos\left(\tfrac{1}{2}\pi(\tau_{12} - \tau_{23})\right),$$

with $(i, j, k) = (1, 2, 3)$.

Let us now turn to Spearman's rho, $\rho(X, Y)$. For three independent random vectors (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) with a common joint distribution H (whose margins are F and G, and X_i and Y_i, $i = 1, 2, 3$, are copies of X and Y, respectively), Spearman's rho is defined as

$$\begin{aligned}\rho(X, Y) = 3\{&P[(X_1 - X_2)(Y_1 - Y_3) > 0] \\ &- P[(X_1 - X_2)(Y_1 - Y_3) < 0]\}.\end{aligned}$$

Note that (X_1, Y_1) and (X_2, Y_3) is pair of vectors with the same marginals, but (X_1, Y_1) has distribution function H, while the components of (X_2, Y_3) are independent. For X and Y continuous with copula C, one can show (Nelsen, 2006, p. 167) that

$$\rho(X, Y) = \rho_C = 12 \iint_{[0,1]^2} uv\; dC(u, v) - 3$$

and, moreover, that Spearman's rho is identical to the linear correlation coefficient between the random variables $U = F(X)$ and $V = G(Y)$:

$$\begin{aligned}\rho(X, Y) = \rho_C &= 12\,\mathrm{E}[UV] - 3 \\ &= \frac{\mathrm{E}[UV] - 1/4}{1/12} = \frac{\mathrm{E}[UV] - \mathrm{E}[U]\mathrm{E}[V]}{\sqrt{\mathrm{Var}[U]}\sqrt{\mathrm{Var}[V]}}.\end{aligned}$$

Finally, the relation between Kendall's tau and Spearman's rho has been investigated early on. It has been shown (e.g. Kruskal, 1958) that always

$$-1 \leq 3\tau - 2\rho \leq 1,$$

where τ is Kendall's tau and ρ is Spearman's rho.

4. Example application: multisensory modeling

While the applications discussed here refer to the context of multisensory modeling, similar examples can more generally be found in any context where processing of information takes place in two or more channels (e.g. the visual search paradigm mentioned in the introduction). We do not strive to be exhaustive in presenting multisensory applications, nor do we try to treat them in depth. Rather, our goal is to encourage further work in an area that is quickly gaining importance in psychology, neuroscience, and related areas.

In the behavioral version of the *multisensory paradigm*, one distinguishes two different conditions:

Unimodal condition: a stimulus of a single modality (visual, auditory, tactile) is presented and the participant is (i) asked to respond (e.g., by button press or eye movement) as quickly as possible upon detecting the stimulus (*reaction time task*), or (ii) to indicate whether or not a stimulus of that modality was detected (*detection task*).

Bi- or trimodal condition: stimuli from two or three modalities are presented (nearly) simultaneously and the participant is (i) asked to respond as quickly as possible upon detecting a stimulus of any modality (*redundant signals RT task*), or (ii) to indicate whether or not a stimulus of any modality was detected (*redundant signals detection task*).

We refer to $\mathcal{V}$, $\mathcal{A}$, $\mathcal{T}$ as the unimodal context where visual, auditory, or tactile stimuli are presented, resp. Similarly, $\mathcal{VA}$ denotes a bimodal (visual–auditory) context, etc. For each stimulus, or crossmodal stimulus combination, we observe samples from a random variable representing the reaction time measured in any given trial. Let $F_V(t)$, $F_A(t)$, $F_{VA}(t)$ denote the (theoretical) distribution functions of reaction time in a unimodal visual, auditory, or a visual–auditory context, respectively, when a specific stimulus (combination) is presented.[3] Analogously, we define the probabilities for indicator functions in a detection task: $p_V = P(detection|V)$, $p_A = P(detection|A)$, and $p_{VA} = P(detection|VA)$.

Note that, from a modeling point of view, each context $\mathcal{V}$, $\mathcal{A}$, or $\mathcal{VA}$ refers to a different sample space and σ-algebra. Therefore, no probabilistic coupling between the (reaction time) random variables in these different conditions necessarily exist. A common assumption, often not stated explicitly, is that there does exist a coupling between visual and auditory RT, for example, such that the margins of the coupling, i.e., of a bivariate distribution $\hat{H}_{VA}$, are equal to the distributions F_V and F_A (see, e.g., Colonius, 1990, and Luce, 1986, pp. 129–130).

Assuming such a coupling exists, a multisensory model should specify how F_{VA} relates to the bivariate distribution $\hat{H}_{VA}$. In principle, this could be tested empirically. However, in the multisensory RT paradigm described above, the marginals of $\hat{H}_{VA}$ are not observable, only the distribution function of RTs in the bimodal context, F_{VA}, is. Therefore, testing for the existence of a coupling requires an additional assumption about how $\hat{H}_{VA}$ and F_{VA} are related. The model studied most often is the so-called *race model*.

Example 20 (*The Race Model*)**.** Let V and A be the random reaction times in unimodal conditions $\mathcal{V}$ and $\mathcal{A}$, with distribution functions $F_V(t)$ and $F_A(t)$, resp. Assume a coupling exists, i.e., a bivariate distribution function $\hat{H}_{VA}$ for $(\hat{V}, \hat{A})$ such that

$$V \stackrel{d}{=} \hat{V} \quad \text{and} \quad A \stackrel{d}{=} \hat{A};$$

[3] For simplicity, we write $F_V(t)$, etc., instead of $F_{\mathcal{V}}(t)$.

assume bimodal RT is determined by the "winner" of the race between the modalities:

$$F_{VA}(t) = P(\hat{V} \le t \text{ or } \hat{A} \le t).$$

Then

$$F_{VA}(t) = F_V(t) + F_A(t) - \hat{H}_{VA}(t, t). \tag{10}$$

The function $\hat{H}_{VA}(s, t) = C(F_V(s), F_A(t))$ is clearly a copula and, from Sklar's theorem, it is unique assuming continuous unimodal distribution functions. Moreover, we have the upper and lower Fréchet copulas (Example 16) such that

$$\max\{F_V(s) + F_A(t) - 1, 0\} \le \hat{H}_{VA}(s, t) \le \min\{F_V(s), F_A(t)\}. \tag{11}$$

Taking the diagonal sections of these copulas (i.e., setting $s = t$ throughout), inserting $F_{VA}(t)$ from Eq. (10), and rearranging yields the "race model inequality" (Miller, 1982):

$$\max\{F_V(t), F_A(t)\} \le F_{VA}(t) \le \min\{F_V(t) + F_A(t), 1\}, \quad t \ge 0.$$

The upper bound corresponds to maximal negative dependence between $\hat{V}$ and $\hat{A}$, the lower bound to maximal positive dependence. Empirical violation of the upper bound (occurring only for small enough t) is interpreted as evidence against the race mechanism ("bimodal RT faster than predictable from unimodal conditions"), but it may also be evidence against the coupling assumption (Colonius & Diederich, 2006).

Example 21 (*Time Window of Integration Model, Colonius & Diederich, 2004*)**.** The time window of integration (*TWIN*) model distinguishes a first stage where unimodal neural activations race against each other, and a subsequent stage of converging processes that comprise neural integration of the input and preparation of a response. Multisensory integration occurs only if all peripheral processes of the first stage terminate within a given temporal interval, the "time window of integration". Total reaction time in the crossmodal (visual–auditory) condition is the sum of first and second stage processing times:

$$RT_{VA} = W_1 + W_2, \tag{12}$$

where W_1 and W_2 are two random variables on the same probability space. Letting I denote the random event that integration occurs and I^c its complement, W_1 and W_2 are assumed to be *conditionally independent*, conditioning on either I or I^c. This implies that any dependency between the processing stages is solely generated by the event of integration or its complement. The distribution of the pair (W_1, W_2) then is

$$H(w_1, w_2) = \pi H_I(w_1, w_2) + (1 - \pi) H_{I^c}(w_1, w_2), \tag{13}$$

where H_I and H_{I^c} denote the conditional distributions of W_1 and W_2 with respect to I and I^c, respectively, and $\pi = P(I)$.

By conditional independence, H_I and H_{I^c} can be written as products of their marginal distributions,

$$H_I(w_1, w_2) = F_I(w_1) G_I(w_2) \quad \text{and}$$
$$H_{I^c}(w_1, w_2) = F_{I^c}(w_1) G_{I^c}(w_2),$$

where F and G refer to the first and second stage (conditional) distributions, respectively. Inserting into Eq. (13) yields

$$H(w_1, w_2) = \pi F_I(w_1) G_I(w_2) + (1 - \pi) F_{I^c}(w_1) G_{I^c}(w_2). \tag{14}$$

The covariance between W_1 and W_2, computed using Hoeffding's identity (Eq. (9)), equals

$$\mathrm{Cov}(W_1, W_2) = \pi(1 - \pi)\{\mathrm{E}(W_1|I^C) - \mathrm{E}(W_1|I)\} \times \{\mathrm{E}(W_2|I^C) - \mathrm{E}(W_2|I)\}, \tag{15}$$

showing that the dependence between the stage processing times can be positive, negative, or zero.[4]

Bibliographic and historical notes

The origins of the theory of probabilistic coupling have been traced back to the work of Wolfgang Döblin[5] in the late 1930s (see Döblin, 1938; Lindvall, 1991), but interest in the theory waxed and waned for a long time, and it has only recently become part of some standard texts in probability theory (e.g. Gut, 2013). For an advanced treatment of coupling theory see Levin et al. (2008), Lindvall (2002) and Thorisson (2000).

According to Durante and Sempi (2010, 2016), the history of copulas may be traced back to Fréchet (1951). However, the term *copula* and the theorem bearing his name was introduced by Abe Sklar (Sklar, 1959). Comprehensive treatments of copula theory are Durante and Sempi (2016) and Jaworski, Durante, Härdle, and Rychlik (2010). Joe (1997, 2015), Mai and Scherer (2012) and Nelsen (2006) focus on the simulation aspects, and Denuit, Dhaene, Goovaerts, and Kaas (2005) and Rüschendorf (2013) emphasize the actuarial and financial risks background of the theory. A compact and application oriented introduction is Trivedi and Zimmer (2005).

To our knowledge, the first application of the concepts of coupling and copula (without using those terms) to reaction time modeling is Colonius (1990). Early investigations of the race model inequality include Colonius (1986), Diederich (1992); Diederich and Colonius (1987), Miller (1982) and Ulrich and Giray (1986). An application of coupling/copula concepts to multisensory detection is Colonius (2015).

Acknowledgments

This work was supported by DFG (German Science Foundation) SFB/TRR-31 (Project B4) and the DFG Cluster of Excellence EXC 1077/1 Hearing4all. Helpful comments by Phil Smith and an anonymous reviewer are gratefully acknowledged.

References

Colonius, H. (1986). Measuring channel dependence in separate activation models. *Perception and Psychophysics*, *40*, 251–255.
Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, *34*, 253–275.
Colonius, H. (2015). Behavioral measures of multisensory integration: bounds on bimodal detection probability. *Brain Topography*, *28*, 1–4.
Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*, 1000–1009.
Colonius, H., & Diederich, A. (2006). The race model inequality: interpreting a geometric measure of the amount of violation. *Psychological Review*, *113*, 148–154.
Denuit, M., Dhaene, J., Goovaerts, M., & Kaas, R. (2005). *Actuarial theory for dependent risks*. John Wiley and Sons, Ltd..
Diederich, A. (1992). Probability inequalities for testing separate activation models of divided attention. *Perception and Psychophysics*, *52*, 714–716.
Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research*, *49*, 23–29.
Döblin, W. (1938). Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d' états. *Revues Mathématiques de l'Union Interbalkanique*, *2*, 77–105.
Durante, F., & Sempi, C. (2010). Copula theory: An introduction. In *Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (Eds.), Copula theory and its applications: Proceedings of the workshop held in Warsaw, 25–26 September 2009* (pp. 3–31).
Durante, F., & Sempi, C. (2016). *Principles of copula theory*. Boca Raton, FL: CRC Press.
Dzhafarov, E., & Kujala, J. (2016). Context-content systems of random variables: the contextuality-by-default theory. *Journal of Mathematical Psychology*, (in press). Available as: arXiv:1511.03516v1.

[4] Investigation of Kendall's tau and Spearman's rho for this model will be pursued elsewhere.

[5] Son of Alfred Döblin (1878–1957), an important German writer well known for his novel *Berlin Alexanderplatz*.

Please cite this article in press as: Colonius, H., An invitation to coupling and copulas: With applications to multisensory modeling. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.004

Embrechts, P., Lindskog, F., & McNeil, A. (2003). Modelling dependence with copulas and applicatiions to risk management. In S. Rachev (Ed.), *Handbook of heavy tailed distributions in finance* (pp. 329–384) Elsevier, (Chapter 8).

Embrechts, P., McNeil, A., & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In M. Dempster (Ed.), *Risk management: Value at risk and beyond* (pp. 176–223). Cambridge: Cambridge University Press.

Frechét, M. (1951). Sur les tableaux de corrélation dont les marges sont donnée. *Annales de l'Université de Lyon. Section A*, *14*, 53–77.

Gut, A. (2013). *Probability: A graduate course* (2nd ed.). New York, NY: Springer.

Hoeffding, W. (1940). Maßstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, *5*, 181–233.

Jaworski, P., Durante, F., Härdle, W., & Rychlik, T. (Eds.) (2010). *Lecture notes in statistics—Proceedings*: *Vol. 198*. *Copula theory and its applications: Proceedings of the workshop held in Warsaw, 25–26 September 2009*. Springer.

Joe, H. (1997). *Multivariate models and dependence concepts*. London, UK: Chapman & Hall.

Joe, H. (2015). *Dependence modeling with copulas*. Boca Raton, FL: CRC Press, Taylor & Francis.

Kruskal, W. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, *53*, 814–861.

Levin, D. A., Peres, Y., & Wilmer, E. L. (2008). *Markov chains and mixing times*. Providence, RI: American Mathematical Society.

Lindvall, T. (1991). W. Doeblin 1915–1940. *The Annals of Probability*, *19*, 929–934.

Lindvall, T. (2002). *Lectures on the coupling method* (Reprint of Wiley 1992 ed.). Mineola, NY: Dover.

Luce, R. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Mai, J., & Scherer, M. (2012). *Series in quantitative finance*: *Vol. 4*. *Simulating copulas: Stochastic models, sampling algorithms and applications*. Imperial College Press.

Miller, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.

Nelsen, R. B. (2006). *Lecture notes in statistics*: *Vol. 139*. *An introduction to copulas* (2nd ed.). New York, NY: Springer-Verlag.

Pfeifer, D., & Nešlehová, J. (2004). Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin*, *34*, 333–360.

Rüschendorf, L. (2013). *Mathematical risk analysis*. New York, NY: Springer-Verlag.

Shea, G. (1983). Hoeffding's lemma. In S. Kotz, N. Johnson, & C. Read (Eds.), *Encyclopedia of statistical science* (pp. 648–649). New York, NY: Wiley.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, *8*, 229–231.

Strassen, V. (1965). The exisistence of probability measures with given marginals. *Annals of Mathematical Statistics*, *36*, 423–439.

Thorisson, H. (2000). *Coupling, stationarity, and regeneration*. New York, NY: Springer Verlag.

Trivedi, P., & Zimmer, D. (2005). Copula modeling: an introduction for practitioners. *Foundations and Trends in Econometrics*, *1*, 1–111.

Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception and Psychophysics*, *39*, 248–254.

Please cite this article in press as: Colonius, H., An invitation to coupling and copulas: With applications to multisensory modeling. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.02.004