ILLINOIS JOURNAL OF MATHEMATICS
Volume 36, Number 4, Winter 1992

COPULAS AND MARKOV PROCESSES

BY

WILLIAM F. DARSOW, BAO NGUYEN AND ELWOOD T. OLSEN

1. Introduction

In this paper we study Markov processes using the copulas of A. Sklar [1], [2]. A 2-copula is a function C: $[0,1]^2 \to [0,1]$ satisfying:

(i) (Boundary conditions)

$$C(0, x_2) = C(x_1, 0) = 0 \quad \text{for all } x_1, x_2 \in [0,1]$$

and

$$C(x_1, 1) = x_1 \quad \text{and} \quad C(1, x_2) = x_2 \quad \text{for all } x_1, x_2 \in [0,1].$$

(ii) (Monotonicity condition)

$$C(x_1, x_2) + C(y_1, y_2) - C(x_1, y_2) - C(y_1, x_2) \geq 0$$

for all $x_1, x_2, y_1, y_2 \in [0,1]$ satisfying $x_1 \leq y_1$ and $x_2 \leq y_2$.

These conditions imply the continuity of C. Copulas are of interest because they link joint distributions to one-dimensional marginal distributions. Sklar showed that for any real valued random variables X_1 and X_2 with joint distribution F_{12} there is a copula C such that

$$F_{12}(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{1.1}$$

where F_1 and F_2 denote the cumulative distribution functions of X_1 and X_2 respectively. The copula C is said to *connect* X_1 and X_2 or to be a copula *of* X_1 and X_2. In the other direction, for any distribution functions F_1 and F_2 and any copula C, the function defined on the right hand side of (1.1) is a two dimensional distribution whose margins are F_1 and F_2. Copulas thus capture all of the information concerning the dependence structure of random variables irrespective of their distributions and so provide a natural frame-

Received December 14, 1990
1991 Mathematics Subject Classification. Primary 60G07; Secondary 60J05, 60J25.

work for many investigations. Here, we make use of copulas to investigate a certain type of dependence structure—the conditional independence condition satisfied by random variables in a Markov process.

If the random variables X_1 and X_2 are continuous with joint distribution F_{12}, then the copula C is uniquely determined by (1.1). If, however, the random variables are not continuous, the copula C is not unique; in this case, the values of the copula are uniquely determined at points (x_1, x_2) where x_k is in the range of F_k, $k = 1, 2$, and a copula C for which the expression above holds can be obtained by interpolating the values at these points in any manner consistent with the defining properties of a copula. Interpolation which is linear in each place ("bilinear interpolation") works, and we adopt the convention that bilinear interpolation is always used to fill in values at other points. With this convention, we can refer to *the* copula of X_1 and X_2.

For $m \geq 3$ an m-copula is a function C: $[0,1]^m \to [0,1]$ satisfying:

(i) (Boundary conditions)

  (a) $C(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_m) = 0$ for all i and for all $x_1, \ldots, x_m$;

  (b) The function

$$(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m) \to C(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_m)$$

  is an $(m-1)$-copula for all i.

(ii) (Monotonicity condition)

$$\sum_{V \in R} \operatorname{sgn}(V) C(V) \geq 0$$

for all rectangles R of the form $R = \prod_{i=1}^m [x_i, y_i]$, $x_i \leq y_i$. Here, the sum is over all vertices $V = (\varepsilon_1, \ldots, \varepsilon_m)$ of the rectangle, where $\varepsilon_i = x_i$ or y_i, and

$$\operatorname{sgn}(V) = \begin{cases} -1, & \text{if the number of } x_i\text{'s among the coordinates of } V \text{ is odd,} \\ 1, & \text{otherwise.} \end{cases}$$

Again, these conditions imply the continuity of C. Sklar's basic theorem, referred to above, states in this context that if $X_1, \ldots, X_m$ are real valued random variables with joint distribution $F_{1\ldots m}$ then there exists an m-copula C such that for all $x_1, \ldots, x_m$,

$$F_{1\ldots m}(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)) \tag{1.2}$$

where F_i is the cumulative distribution function of X_i. Conversely, for any distribution functions $F_1, \ldots, F_m$ and any m-copula C, the right hand side of

(1.2) defines an m-dimensional joint distribution function whose one-dimensional marginals are $F_1, \ldots, F_m$.

If the random variables in (1.2) are all continuous, the m-copula C in (1.2) is uniquely defined; otherwise it is uniquely determined at points $(x_1, \ldots, x_m)$ where x_k is in the range of F_k, $k = 1, \ldots, m$, and as before can be obtained at other points by interpolation. Here, m-linear interpolation works, and we adopt the convention that it is always used. For discussion of these issues, see e.g., [1]–[4].

We begin Section 2 with a list of some key properties of 2-copulas.

This paper is divided roughly into two parts. In the first part (Sections 2, 3 and 4), we define a product, which we shall call the $*$ operation, on copulas and discuss its interpretation in the context of Markov processes. The $*$ operation is defined and its basic properties are given in Section 2 of the paper. The $*$ operation on copulas has a natural interpretation in Markov processes; it permits the conditional independence of random variables to be described in terms of a binary operation—the $*$ operation—on the copulas of the process. Section 3 investigates this interpretation. In particular, the $*$ operation on copulas corresponds in a natural way to the operation on transition probabilities contained in the Chapman-Kolmogorov equations. It leads, however, to a technique for constructing Markov processes which is different from the conventional technique; in particular, once copulas have been specified satisfying the $*$ product analog of the Chapman-Kolmogorov equations, the marginal distributions can all be specified at will, subject to a continuity condition. In the conventional technique, once transition probabilities satisfying the Chapman-Kolmogorov equations are specified, it remains to give a single marginal distribution, which can be viewed as the initial data for the process. Section 4 gives examples.

In the second part (Sections 5 to 11) we define Markov algebras, the algebraic entities which the $*$ operation on the set of copulas naturally suggests, and explore some of their properties.

2. A product for 2-copulas

We state first some properties of copulas which we will need below. Let $\mathscr{C}$ denote the set of all copulas on $[0,1]^2$, and let $C \in \mathscr{C}$. Then:

1. For any x and ξ satisfying $0 \leq x \leq \xi \leq 1$, the function

$$\eta \to C(\xi, \eta) - C(x, \eta) \tag{2.1}$$

is non-decreasing. Similarly, for any y and η satisfying $0 \leq y \leq \eta \leq 1$, the function

$$\xi \to C(\xi, \eta) - C(\xi, y) \tag{2.2}$$

is non-decreasing. Both of these results are immediate consequences of the monotonicity condition.

2. Taking first $\eta = 0$ and then $\eta = 1$ in (2.1) and combining the inequalities, we obtain

$$0 \leq C(\xi, \eta) - C(x, \eta) \leq \xi - x \tag{2.3}$$

for all x and $\xi \in [0, 1]$ satisfying $0 \leq x \leq \xi \leq 1$ and for all $\eta \in [0, 1]$. Similarly, for all y and $\eta \in [0, 1]$ satisfying $0 \leq y \leq \eta \leq 1$ and for all $\xi \in [0, 1]$,

$$0 \leq C(\xi, \eta) - C(\xi, y) \leq \eta - y. \tag{2.4}$$

3. For all x, ξ, y and $\eta \in [0, 1]$,

$$|C(\xi, \eta) - C(x, y)| \leq |x - \xi| + |y - \eta|. \tag{2.5}$$

This is an immediate consequence of (2.3) and (2.4). It follows from (2.5) that copulas are Lipschitz continuous with Lipschitz constant equal to 1.

4. For all $\xi \in [0, 1]$ the function $\eta \to C(\xi, \eta)$ is non-decreasing. Similarly for all $\eta \in [0, 1]$ the function $\xi \to C(\xi, \eta)$ is non-decreasing. These results are special cases of (2.1) and (2.2), respectively, obtained by taking $x = y = 0$. Thus, the x- and y-sections of a copula are non-decreasing functions.

5. Let $C_{,1}$, $C_{,2}$ and $C_{,12}$ denote the partial derivatives $\partial C/\partial x$, $\partial C/\partial y$, and $\partial^2 C/\partial x\, \partial y$, respectively. Since monotonic functions are differentiable almost everywhere, it follows that for given y the partial derivative $C_{,1}(x, y)$ exists for almost all x and

$$0 \leq C_{,1}(x, y) \leq 1 \quad \text{a.s.} \tag{2.6}$$

Similarly, for given x the partial derivative $C_{,2}(x, y)$ exists for almost all y and

$$0 \leq C_{,2}(x, y) \leq 1 \quad \text{a.s.} \tag{2.7}$$

These facts are immediate consequences of (2.3) and (2.4). The two statements above hold almost surely with respect to Lebesgue measure on $[0, 1]$ for each y, in the case of (2.6), and for each x, in the case of (2.7).

6. By similar reasoning, the functions (2.1) and (2.2) have derivatives almost everywhere, and

$$[C(\xi,\eta) - C(x,\eta)]_{,2} \geq 0 \quad \text{if } x \leq \xi \tag{2.8}$$

and

$$[C(\xi,\eta) - C(\xi,y)]_{,1} \geq 0 \quad \text{if } y \leq \eta. \tag{2.9}$$

It follows that the functions $\eta \to C_{,1}(x,\eta)$ and $\xi \to C_{,2}(\xi, y)$ are defined and non-decreasing almost everywhere.

7. The set $\mathscr{C}$ of all 2-copulas is a compact and convex subset of the space of all continuous real valued functions defined on the unit square under the topology of uniform convergence. It follows that, in $\mathscr{C}$, pointwise convergence implies uniform convergence. These facts are easy to prove; we omit the arguments.

We observe that properties analogous to those set forth above hold for m-copulas; we omit their statement, but we will use some of them later.

Three copulas arise repeatedly

$$\begin{aligned} M(x,y) &= \min(x,y), \\ P(x,y) &= xy, \\ W(x,y) &= \max(x+y-1,0). \end{aligned}$$

We leave it to the reader to verify that these functions satisfy the monotonicity and boundary conditions for 2-copulas. It can easily be shown that for any $C \in \mathscr{C}$

$$W \leq C \leq M$$

where the inequality is the usual pointwise partial ordering for continuous functions.

The copulas M, P and W have the following stochastic interpretations: Random variables X_1 and X_2 are connected by P if and only if they are independent. Continuous random variables X_1 and X_2 are connected by M (W) if and only if X_2 is a.s. a non-decreasing (non-increasing) function of X_1. Thus, P corresponds to independence whereas M and W correspond to species of deterministic dependence. We will see later that deterministic dependence of two random variables can be characterized in terms of the algebraic properties of their copula (Theorem 11.1).

We now define the product operation on copulas, which is central to this paper. Consider A, B in $\mathscr{C}$. For x, y in [0, 1], set

$$(A * B)(x,y) = \int_0^1 A_{,2}(x,t) B_{,1}(t,y)\, dt. \tag{2.10}$$

Since A and B are absolutely continuous in each place, by (2.5), the integral in (2.10) exists.

THEOREM 2.1. *$A * B$ is in $\mathscr{C}$.*

Proof. Let $C = A * B$ and let x, ξ, y and η be any numbers in $[0,1]$ satisfying $0 \leq x \leq \xi \leq 1$ and $0 \leq y \leq \eta \leq 1$. Then

$$
\begin{aligned}
&C(x,y) + C(\xi,\eta) - C(x,\eta) - C(\xi,y) \\
&\quad = \int_0^1 [A(\xi,t) - A(x,t)]_{,2}[B(t,\eta) - B(t,y)]_{,1}\, dt \\
&\quad \geq 0
\end{aligned}
$$

by (2.8) and (2.9). Thus, C exhibits the monotonicity property. The boundary conditions are also easily verified. ∎

THEOREM 2.2. *As a binary operation on $\mathscr{C}$ the $*$ operation is right and left distributive over convex combinations.*

Proof. This is clear from the definition of the $*$ operation in (2.10). ∎

Let C be any copula. By direct calculation, the $*$ products of C with P, M and W are as follows:

$$
\begin{aligned}
P * C &= C * P = P, \\
M * C &= C * M = C, \\
(W * C)(x,y) &= y - C(1-x,y), \\
(C * W)(x,y) &= x - C(x,1-y).
\end{aligned}
$$

In particular, P is a null element in $\mathscr{C}$ and M is an identity.

The remainder of this section is devoted to showing that the $*$ operation is associative. Our argument proceeds by way of two preliminary results. The first concerns a continuity property of the $*$ product and has independent interest.

THEOREM 2.3. *Consider A_n, B in $\mathscr{C}$ such that $A_n \to A$. Then, $A_n * B \to A * B$ and $B * A_n \to B * A$.*

Proof. We prove $A_n * B \to A * B$; the proof of the other conclusion is analogous. Consider $\varepsilon > 0$. Fix x and y in $[0,1]$ and write $g(t) = B(t,y)$ and $f_n(t) = A(x,t) - A_n(x,t)$. Clearly g' and f_n' are in $L^\infty([0,1])$, by (2.6) and

(2.7); in particular, $\|f_n'\|_\infty \le 2$. There is a non-zero step function

$$\phi = \sum_{i=1}^{k} a_i \chi_i$$

such that $\|g' - \phi\|_1 < \varepsilon$ where $0 = x_0 < x_1 < \cdots < x_k = 1$ and χ_i is the characteristic function of $[x_{i-1}, x_i]$. Since $f_n(x_i) \to 0$ for each i, there is an N such that $n \ge N$ implies

$$|f_n(x_i) - f_n(x_{i-1})| \le \frac{\varepsilon}{\sum_{i=1}^{k} |a_i|}$$

for all i. Then, when $n \ge N$,

$$\begin{aligned}\left|\int_0^1 g'(t) f_n'(t)\,dt\right| &\le \left|\int_0^1 [g'(t) - \phi(t)] f_n'(t)\,dt\right| + \left|\int_0^1 \phi(t) f_n'(t)\,dt\right| \\ &\le 2\varepsilon + \sum_{i=1}^{k} |a_i| \cdot \left|\int_{x_{i-1}}^{x_i} f_n'(t)\,dt\right| \\ &= 2\varepsilon + \sum_{i=1}^{k} |a_i| \cdot |f_n(x_i) - f_n(x_{i-1})| \\ &\le 3\varepsilon.\end{aligned}$$

This yields the desired result. ■

This theorem says that the $*$ product is continuous in each place. We will see later (Theorem 7.6) that it is not jointly continuous.

A copula $A \in \mathscr{C}$ induces a probability measure α on the Borel sets in $[0,1]^2$ via the assignment

$$\alpha(R) = A(x, y) - A(x, \eta) - A(\xi, y) + A(\xi, \eta) \tag{2.11}$$

for all rectangles $R = [x, \xi] \times [y, \eta] \subset [0,1]^2$. By the monotonicity condition, the measure of every rectangle, and therefore of every Borel set, is nonnegative, and $\alpha([0,1]^2) = 1$. It is not true that every probability measure α on $[0,1]^2$ is induced by a copula A in the manner of (2.11). For a measure α to be induced by a copula, it must spread mass in a manner consistent with the boundary conditions on a copula; that is, it must be true that for all x and y,

$$\alpha([0, x] \times [0, 1]) = x \quad \text{and} \quad \alpha([0, 1] \times [0, y]) = y.$$

It is easy to see that these conditions are both necessary and sufficient for α to be induced by a copula. It is sometimes useful to construct a copula A

with desired properties by starting with a measure α satisfying the consistency conditions above. This construction is used in the following lemma.

LEMMA 2.1. *The set of copulas whose induced measures are absolutely continuous with respect to Lebesgue measure is dense in the set $\mathscr{C}$ of all copulas.*

Proof (Sherwood [9]). Let a copula A and a number $\varepsilon > 0$ be given. We want to construct a copula B, satisfying $\|B - A\| < \varepsilon$, whose induced measure β is absolutely continuous with respect to Lebesgue measure. Choose an integer $N > 1$ such that $N > 2/\varepsilon$. Cut $[0,1]^2$ into N^2 congruent squares S_{ij}, $i, j = 1, 2, \dots N$. Let B be the copula whose induced measure β satisfies

$$\beta(S_{ij}) = \alpha(S_{ij})$$

and whose mass is spread uniformly in S_{ij}. In this expression α denotes the measure induced by the copula A. Clearly β is absolutely continuous with respect to Lebesgue measure. We leave to the reader the task of verifying that β is induced by a copula. Observe that for any $(x, y) \in [0,1]^2$, there is a corner (x_i, y_j) of some square S_{ij} such that

$$|x - x_i| + |y - y_j| \le \frac{1}{N}.$$

Then

$$\begin{aligned}
|A(x,y) - B(x,y)| &\le |A(x,y) - A(x_i, y_j)| + |A(x_i, y_j) - B(x_i, y_j)| \\
&\quad + |B(x_i, y_j) - B(x,y)| \\
&\le 2\big(|x - x_i| + |y - y_j|\big) \\
&< \varepsilon.
\end{aligned}$$

Here, we have used (2.5) and the fact that by construction $|A(x_i, y_j) - B(x_i, y_j)| = 0$. ■

It is a corollary of the proof of Lemma 2.1 that given a copula A and a number $\varepsilon > 0$ we can always find a copula B satisfying $\|A - B\| < \varepsilon$, the density of whose induced measure β with respect to Lebesgue measure is a linear combination of χ_{ij}'s, where χ_{ij} is the characteristic function of the set S_{ij} constructed in the proof of the theorem. We denote the density as the Radon-Nikodym derivative $d\beta/d\mu$ where μ is Lebesgue measure. It is then easy to see that in this case $d\beta/d\mu$ is bounded and

$$\frac{d\beta}{d\mu} = B_{,12} = B_{,21} \quad \text{a.s.} \tag{2.12}$$

We make use of this fact in the proof of the associativity theorem.

THEOREM 2.4. *The binary operation* $*$ *is associative.*

Proof. Let A, B and C be copulas. We want to show that

$$A * (B * C) = (A * B) * C.$$

By Theorem 2.3 and Lemma 2.1 it suffices to consider B in $\mathscr{C}$ for which the doubly stochastic measure β induced by B is absolutely continuous with respect to Lebesgue measure. Furthermore, we may assume that (2.12) holds, that is, that $B_{,12}$ and $B_{,21}$ exist almost everywhere, are bounded and integrable, and are equal almost everywhere. Fix x and y and set $f(t) = A(x, t)$ and $g(s) = C(s, y)$. Then

$$\begin{aligned}
[A * (B * C)](x, y) &= \int_0^1 f'(t) \frac{d}{dt}\left(\int_0^1 B_{,2}(t, s) g'(s)\, ds\right) dt \\
&= \int_0^1 \int_0^1 f'(t) B_{,12}(t, s) g'(s)\, ds\, dt \\
&= \int_0^1 g'(s) \frac{d}{ds}\left(\int_0^1 f'(t) B_{,1}(t, s)\, dt\right) ds \\
&= [(A * B) * C](x, y)
\end{aligned}$$

by Fubini's theorem. ■

We turn now to a probabilistic interpretation of the $*$ operation.

3. The $*$ product and Markov processes

Let X and Y be random variables defined on the same probability space, and let C be their copula. The conditional expectations $E(I_{X<x}|Y)$ and $E(I_{Y<y}|X)$ are closely related to the copula, and this fact is basic to the interpretation we give here of the $*$ product.

THEOREM 3.1. *If random variables X and Y have the copula C, then*

$$E(I_{X<x}|Y)(\omega) = C_{,2}(F_X(x), F_Y(Y(\omega))) \quad \text{a.s.} \tag{3.1}$$

and

$$E(I_{Y<y}|X)(\omega) = C_{,1}(F_X(X(\omega)), F_Y(y)) \quad \text{a.s.} \tag{3.2}$$

Before proving the theorem, we present an argument for random variables X and Y whose distribution functions are continuous and strictly increasing.

In this case, we can write for the conditional probability $P(X < x|Y) = E(I_{X<x}|Y)$:

$$\begin{aligned} P(X < x|Y = y) &= \lim_{\Delta y \to 0} P(X < x|y < Y \le y + \Delta y) \\ &= \lim_{\Delta y \to 0} \frac{F_{XY}(x, y + \Delta y) - F_{XY}(x, y)}{F_Y(y + \Delta y) - F_Y(y)} \\ &= \lim_{\Delta y \to 0} \frac{C(F_X(x), F_Y(y + \Delta y)) - C(F_X(x), F_Y(y))}{F_Y(y + \Delta y) - F_Y(y)} \\ &= C_{,2}(F_X(x), F_Y(y)) \end{aligned}$$

wherever the derivative in the last expression exists.

Proof of Theorem 3.1. Let $\sigma(Y)$ denote the inverse images of the Borel sets under Y. Since the function $\omega \to C_{,2}(F_X(x), F_Y(Y(\omega)))$ is measurable with respect to $\sigma(Y)$, we need only show that for all $A \in \sigma(Y)$ and for all x,

$$\int_A C_{,2}(F_X(x), F_Y(Y(\omega))\, dP(\omega) = \int_A I_{X<x}(\omega)\, dP(\omega). \tag{3.3}$$

This will prove (3.1); (3.2) is proved analogously. In fact, it is sufficient to consider $A = Y^{-1}((-\infty, a])$ in (3.3). Then

$$\begin{aligned} \text{LHS of (3.3)} &= \int_{-\infty}^{a} C_{,2}(F_X(x), F_Y(\xi))\, dF_Y(\xi) \\ &= \int_0^{F_Y(a)} C_{,2}(F_X(x), \eta)\, d\eta \\ &= C(F_X(x), F_Y(a)) \\ &= \text{RHS of (3.3)}. \end{aligned}$$

The second equality above clearly holds when F_Y is continuous. When it is not, we argue as follows. Let t_k range over the points of discontinuity of F_Y and let $[b_k, c_k]$ be the corresponding jump interval in the values of F_Y. Let F_Y^* be a quasi-inverse of F_Y satisfying $F_Y(F_Y^*(s)) = s$ when $s \notin [b_k, c_k]$ for any k and $F_Y(F_Y^*(s)) = b_k$ when $s \in (b_k, c_k)$. Then, the second equality above holds if and only if

$$C_{,2}(F_X(x), b_k) = \frac{C(F_X(x), c_k) - C(F_X(x), b_k)}{c_k - b_k}$$

for all x and k. This condition is guaranteed to hold because of the linear interpolation convention previously adopted (see above in Introduction); the condition does not, however, imply the linear interpolation convention. In any event, with this condition satisfied, and F_Y^* as above, $C_{,2}(F_X(x), F_Y(F_Y^*(s)) = C_{,2}(F_X(x), s)$ for almost every s. It follows that

$$\begin{aligned}\int_{-\infty}^{a} C_{,2}(F_X(x), F_Y(t))\, dF_Y(t) &= \int_0^{F_Y(a)} C_{,2}(F_X(x), F_Y(F_Y^*(s))\, ds \\ &= \int_0^{F_Y(a)} C_{,2}(F_X(x), s)\, ds.\end{aligned}$$

The first equality holds by Lebesgue's definition of the Lebesgue-Stieltjes integral (see, e.g., [13], [14]). ■

We observe that the proof of the foregoing theorem uses the linear interpolation convention in an essential and apparently unavoidable way. We shall return to this point later.

It follows directly from Theorem 3.1 and the definition of the $*$ product that, if X, Y and Z are random variables and X and Z are conditionally independent given Y, then

$$C_{XZ} = C_{XY} * C_{YZ}.$$

The converse statement need not be true, however (see Theorem 3.3 below). Thus, we have a stochastic interpretation of the $*$ product, and also an essential limitation on that interpretation. We will explore the interpretation further in the context of Markov processes.

First, some preliminaries. If we have a two-place function $g(x, y)$ and we set $f(x) = g(x, y)$ for fixed y, we shall use the notation

$$\int_a^b h(x) g(dx, y)$$

for the Stieltjes integral

$$\int_a^b h(x)\, df(x).$$

We will need the following intermediate result in the proof below of the theorem which interprets the $*$ product in the context of Markov processes:

LEMMA 3.1. *Let A and B be copulas. Then for almost all x,*

$$\frac{\partial}{\partial x}\int_0^1 A_{,1}(t, y) B_{,2}(x, t)\, dt = \int_0^1 A_{,1}(t, y) B_{,1}(x, dt). \tag{3.4}$$

Proof. Let ϕ be any C^∞ function which vanishes at 0 and 1. Then

$$\int_0^1 \phi(x)\left(\int_0^1 A_{,1}(t,y) B_{,1}(x,dt)\right) dx$$
$$= -\int_0^1 \phi'(x)\left(\int_0^x \int_0^1 A_{,1}(t,y) B_{,1}(\xi,dt)\, d\xi\right) dx$$
$$= -\int_0^1 \phi'(x)\left(\int_0^1 A_{,1}(t,y) B(x,dt)\right) dx$$
$$= -\int_0^1 \phi'(x)\left(\int_0^1 A_{,1}(t,y) B_{,2}(x,t)\, dt\right) dx$$
$$= \int_0^1 \phi(x) \frac{\partial}{\partial x}\left(\int_0^1 A_{,1}(t,y) B_{,2}(x,t)\, dt\right) dx$$

where the second step must be justified. Then (3.4) follows immediately, since the result above holds for all ϕ. To justify the second step, observe that the maps

$$f \to \int_0^x \left(\int_0^1 f(t) B_{,1}(u,dt)\right) du \quad \text{and} \quad f \to \int_0^1 f(t) B(x,dt)$$

are both positive bounded linear functionals defined on the set of all right continuous step functions on $[0,1]$. By direct calculation they are equal for all such functions f. Since right continuous step functions are dense in $L^1([0,1])$, the linear functionals defined above extend uniquely to linear functionals which are equal for all functions $f \in L^1([0,1])$. This completes the proof. ∎

Now let X_t, $t \in T$, denote a real stochastic process, that is, a sequence of real valued random variables indexed by $t \in T$, where T is some set of real numbers. We will call the process *continuous* if X_t is a continuous random variable for all $t \in T$. We follow the convention of replacing the subscript t by i in case the index set T is discrete; the subscript t may denote an element of either a continuous or a discrete set. The symbols F_t and F_{st} denote the distribution function of X_t and the joint distribution of X_s and X_t, respectively.

A process X_t, $t \in T$ is called a *Markov process* if for all finite index sets $t_1, \dots, t_n$ and $t \in T$ satisfying $t_1 < t_2 < \cdots < t_n < t$,

$$(3.5) \qquad E\left(I_{X_t < \lambda} | X_{t_1}, \dots, X_{t_n}\right) = E\left(I_{X_t < \lambda} | X_{t_n}\right).$$

The interpretation of the $*$ operation is clarified in the following theorem.

THEOREM 3.2. *Let X_t, $t \in T$, be a real stochastic process, and for each $s, t \in T$ let C_{st} denote the copula of the random variables X_s and X_t. The following are equivalent*:

(i) *The transition probabilities $P(s, x, t, A) = P(X_t \in A \mid X_s = x)$ of the process satisfy the Chapman-Kolmogorov equations* [7], [8]

$$P(s, x, t, A) = \int_{-\infty}^{\infty} P(u, \xi, t, A) P(s, x, u, d\xi) \tag{3.6}$$

for all Borel sets A, for all $s < t$ in T, for all $u \in (s, t) \cap T$ and for almost all $x \in R$.

(ii) *For all $s, u, t \in T$ satisfying $s < u < t$,*

$$C_{st} = C_{su} * C_{ut}. \tag{3.7}$$

Proof. To see that (ii) implies (i), observe first that since $A \to P(s, x, t, A)$ is a probability measure for all $s < t$ and almost all x, it is sufficient to verify the Chapman-Kolmogorov equations (3.6) for Borel sets A of the form $A = (-\infty, a)$ and that for sets A of this form the transition probabilities are given in terms of the copulas by

$$P(s, x, t, A) = C_{st,1}(F_s(x), F_t(a)) \quad \text{a.s.} \tag{3.8}$$

by Theorem 3.1. Thus, for sets A of the form $A = (-\infty, a)$, we have for almost all x,

$$\begin{aligned} &\int_{-\infty}^{\infty} P(u, \xi, t, A) P(s, x, u, d\xi) \\ &= \int_{-\infty}^{\infty} C_{ut,1}(F_u(\xi), F_t(a)) C_{su,1}(F_s(x), F_u(d\xi)) \\ &= \int_0^1 C_{ut,1}(\eta), F_t(a)) C_{su,1}(F_s(x), d\eta)) \\ &= \frac{\partial}{\partial \zeta} \int_0^1 C_{ut,1}(\eta, F_t(a)) C_{su,2}(\zeta, \eta)\, d\eta \Big|_{\zeta = F_s(x)} \\ &= (C_{su} * C_{ut})_{,1}(F_s(x), F_t(a)). \end{aligned} \tag{3.9}$$

But by (3.7) $C_{su} * C_{ut} = C_{st}$; substituting this in the last expression yields (3.6), since the last quantity above is $P(s, x, t, A)$ by (3.8). Lemma 3.1 was used in the third step of this argument.

Conversely, if the Chapman-Kolmogorov equations hold, then the first expression in (3.9) is equal to $C_{st,1}(F_s(x), F_t(a))$ for almost all x, by (3.8), so that (3.7) holds.

If the random variables of the process are continuous, this completes the proof. If not there is one more detail to attend to. We assert that if C_{su} and C_{ut} obey the linear interpolation convention referred to in the Introduction, then so does $C_{su} * C_{ut}$. Verification of this is straightforward but tedious, and we omit details. This fact and (3.9) yield the desired conclusion also in this case. ■

Satisfaction of the Chapman-Kolmogorov equations is a necessary but not sufficient condition for a Markov process. We can also state a sufficient condition in terms of copulas. To do so, we first define a generalization of the $*$ product.

Let A be an m-copula and let B be an n-copula. Define $A \star B$: $[0,1]^{m+n-1} \to [0,1]$ via

$$(3.10)\quad \begin{aligned} &A \star B(x_1, \dots, x_{m+n-1}) \\ &\quad = \int_0^{x_m} A_{,m}(x_1, \dots, x_{m-1}, \xi) B_{,1}(\xi, x_{m+1}, \dots, x_{m+n-1})\, d\xi. \end{aligned}$$

Observe that if $m = n = 2$, the $\star$ and $*$ products are related by

$$A * B(x, y) = A \star B(x, 1, y).$$

By arguments similar to those used in Section 2 it is readily verified that $A \star B$ is an $(m + n - 1)$-copula and that the $\star$ product is distributive over convex combinations, is associative (in the sense that $(A \star B) \star C = A \star (B \star C)$) and is continuous in each place.

THEOREM 3.3. *A real valued stochastic process X_t, $t \in T$ is a Markov process if and only if for all positive integers n and for all $t_1, \dots, t_n \in T$ satisfying $t_k < t_{k+1}$, $k = 1, \dots, n-1$,*

$$(3.11)\quad C_{t_1 \dots t_n} = C_{t_1 t_2} \star C_{t_2 t_3} \star \cdots \star C_{t_{n-1} t_n}$$

where $C_{t_1 \dots t_n}$ is the copula of $X_{t_1}, \dots, X_{t_n}$ and $C_{t_k t_{k+1}}$ is the copula of X_{t_k} and $X_{t_{k+1}}$.

Proof. We have to show that (3.11) above implies the conditional independence property (3.5) of a Markov process and vice versa.

For notational convenience we write F_1 for F_{t_1}, C_{12} for $C_{t_1 t_2}$ and so forth.

Observe first that if t_1, t_2, and $t_3 \in T$ satisfy $t_1 < t_2 < t_3$, then the conditional independence property (3.5) for $n = 2$ holds if and only if

$$(3.12)\quad E\big(I_{X_1 < \mu_1} I_{X_3 < \mu_3} | X_2\big) = E\big(I_{X_1 < \mu_1} | X_2\big) E\big(I_{X_3 < \mu_3} | X_2\big) \quad \text{a.s.}$$

To see that (3.5) implies (3.12), observe that for any Borel set B,

$$\begin{aligned}
\int_{X_2^{-1}(B)} I_{X_1<\mu_1} I_{X_3<\mu_3}\, dP &= \int_{X_2^{-1}(B)\cap X_1^{-1}((-\infty,\mu_1))} I_{X_3<\mu_3}\, dP \\
&= \int_{X_2^{-1}(B)\cap X_1^{-1}((-\infty,\mu_1))} E\left(I_{X_3<\mu_3}|X_1, X_2\right) dP \\
&= \int_{X_2^{-1}(B)\cap X_1^{-1}((-\infty,\mu_1))} E\left(I_{X_3<\mu_3}|X_2\right) dP \\
&= \int_{X_2^{-1}(B)} I_{X_1<\mu_1} E\left(I_{X_3<\mu_3}|X_2\right) dP \\
&= \int_{X_2^{-1}(B)} E\left(I_{X_1<\mu_1} E\left(I_{X_3<\mu_3}|X_2\right)|X_2\right) dP \\
&= \int_{X_2^{-1}(B)} E\left(I_{X_1<\mu_1}|X_2\right) E\left(I_{X_3<\mu_3}|X_2\right) dP.
\end{aligned}$$

This yields (3.12). The converse is proved similarly; in fact, one need only read the chain of equalities above in different order.

Integrating both sides of (3.12) over $X_2^{-1}((-\infty, \mu_2))$ yields

$$\begin{aligned}
F_{123}(\mu_1, \mu_2, \mu_3) &= \int_{-\infty}^{\mu_2} C_{12,2}(F_1(\mu_1), F_2(\xi)) C_{23,1}(F_2(\xi)), F_3(\mu_3))\, dF_2(\xi) \\
&= \int_0^{F_2(\mu_2)} C_{12,2}(F_1(\mu_1), \eta) C_{23,1}(\eta, F_3(\mu_3))\, d\eta \\
&= C_{12} \star C_{23}(F_1(\mu_1), F_2(\mu_2), F_3(\mu_3)).
\end{aligned}$$

This yields (3.11) for the case $n = 3$, if X_1, X_2 and X_3 are continuous. If they are not continuous, we assert that if C_{12} and C_{23} obey the linear interpolation convention stated in the introduction, then so does $C_{12} \star C_{23}$. Verification of this is straightforward but tedious, and we omit the details. This fact and the preceding equation imply the desired result even if the process is not continuous.

Conversely, suppose (3.11) holds for the case $n = 3$; we want to show that for all Borel sets $B \in \sigma(X_2)$,

$$\int_B I_{X_1<\mu_1} I_{X_3<\mu_3}\, dP = \int_B E\left(I_{X_1<\mu_1}|X_2\right) E\left(I_{X_3<\mu_3}|X_2\right) dP. \tag{3.13}$$

As usual, it suffices to verify that this is the case when B has the form

$B = X_2^{-1}((-\infty, \mu_2))$, and for sets of this form, it follows by Theorem 3.1 that

$$\begin{aligned}\text{LHS of (3.13)} &= C_{123}(F_1(\mu_1), F_2(\mu_2), F_3(\mu_3)) \\ &= \int_0^{F_2(\mu_2)} C_{12,2}(F_1(\mu_1), \xi) C_{23,1}(\xi, F_3(\mu_3))\, d\xi \\ &= \text{RHS of (3.13)}.\end{aligned}$$

The argument for $n > 3$ proceeds by similar reasoning and an induction. We give only the idea of the proof. Let $t_1 < \cdots < t_n < t$. (Note the slightly altered notation; this simplifies somewhat the necessary accounting in subscripts.) We have from the conditional independence condition

$$E\left(I_{X_t < \mu} | X_1, \dots, X_n\right) = E\left(I_{X_t < \mu} | X_n\right) \quad \text{a.s.}$$

Therefore, schematically,

$$\begin{aligned}&C(F_1(\mu_1), \dots, F_n(\mu_n), F_t(\mu)) \\ &\quad = \int_{A_1 \cap \dots \cap A_n} C_{nt,1}(F_n(X_n(\omega)), F_t(\mu))\, dP(\omega) \\ &\quad = \int_{-\infty}^{\mu_1} \cdots \int_{-\infty}^{\mu_n} C_{nt,1}(F_n(\xi_n), F_t(\mu))\, dF_{1\dots n}(\xi_1, \dots, \xi_n) \\ &\quad = \int_0^{F_1(\mu_1)} \cdots \int_0^{F_n(\mu_n)} C_{nt,1}(\eta_n, F_t(\mu)) C_{1\dots n}(d\eta_1, \dots, d\eta_n) \\ &\quad = \int_0^{F_n(\mu_n)} C_{nt,1}(\eta_n, F_t(\mu)) C_{1\dots n}(F_1(\mu_1), \dots, F_{n-1}(\mu_{n-1}), d\eta_n) \\ &\quad = C_{1\dots n} \star C_{nt}(F_1(\mu_1), \dots, F_n(\mu_n), F_t(\mu)).\end{aligned}$$

The details are similar to those given above for the case $n = 3$, and the converse is also similar. ■

Remarks. 1. It was observed above that satisfaction of the Chapman-Kolmogorov equations is not sufficient to guarantee that a process is Markov. It is easy to see this from the copula viewpoint: We can explicitly construct a family of m-copulas different from those given in (3.11) which are compatible with the 2-copulas of a Markov process, and thus with the Chapman-Kolmogorov equations. For example, let the random variables in a stochastic process be pairwise independent, so that every 2-copula of the process is P. Since $P * P = P$, the Chapman-Kolmogorov equations are satisfied. In this case it is easy to verify that the formula (3.11) returns the m-fold product for all $m = 3, 4, \dots$; it follows that the only Markov process with pairwise independent random variables is a process in which all finite subsets of random variables in the process are independent. But there are many 3-copulas which coincide with P on each of the faces of $[0,1]^3$; for example, it is easy to check that for $|\alpha| < 1$,

$$(3.14) \qquad C_\alpha(x, y, z) = xyz + \alpha x(1-x)y(1-y)z(1-z)$$

is a copula. We use this copula in place of (3.13) for some set of three random variables in the process, say X_1, X_2 and X_3. We then decree that each finite subset not containing all three shall be independent, whereas the m-copula of a set containing all three shall have the form

$$C_{1\dots m}(x_1, \dots, x_m) = C_\alpha(x_1, x_2, x_3) P_{4\dots m}(x_4, \dots, x_m)$$

where C_α is as above. If we now specify marginal distributions, all of the finite dimensional joint distributions of the process are obtained by composing the copulas in the family constructed above with the marginal distributions. We leave it to the reader to verify that the resulting joint distributions satisfy the compatibility conditions of Kolmogorov's fundamental theorem. [6], [7], [8]. Now apply Kolmogorov's theorem to obtain a stochastic process with the specified joint distributions. Since the random variables in the process are pairwise independent by construction, the Chapman-Kolmogorov equations are satisfied. But since the copula C of the three random variables singled out is inconsistent with (3.11), the process is not a Markov process.

2. In the conventional approach, one specifies a Markov process by giving the initial distribution F_{t_0} and a family of transition probabilities $P(s, x, t, A)$ satisfying the Chapman-Kolmogorov equations. In our approach, one specifies a Markov process by giving all of the marginal distributions and a family of 2-copulas satisfying (3.7). Ours is accordingly an alternative approach to the study of Markov processes which is different in principle from the conventional one. Holding the transition probabilities of a Markov process fixed and varying the initial distribution necessarily varies all of the marginal distributions, but holding the copulas of the process fixed and varying the initial distribution does not affect any other marginal distribution.

3. We comment here on the essential use of the linear interpolation convention in the proof of Theorem 3.1. Observe that if we have a family of copulas satisfying (3.7), and we specify discontinuous marginal distributions, we can still obtain a family of compatible finite dimensional distributions via (3.11) and therefore we can apply Kolmogorov's theorem to obtain a stochastic process with the given marginal distributions and copulas. But if the copulas and marginal distributions are such that the linear interpolation convention is not satisfied, the interpretation of the partial derivatives of the copulas as conditional expectations (Theorem 3.1) may fail, so that (3.7) apparently does not guarantee satisfaction of the Chapman-Kolmogorov

equations. We conclude that the approach outlined here permits continuous Markov processes to be specified by giving copulas satisfying (3.7) and any continuous marginal distributions, but that, in constructing Markov processes whose random variables are not continuous, including Markov chains, the selection of the copulas satisfying (3.7) and of the discontinuous marginal distributions is coupled, due to the requirement that linear interpolation must obtain.

4. Examples

In this section we give several examples of Markov processes, or rather of ways to construct or specify Markov processes, using copulas.

Example 4.1. Let T be the set of all non-negative integers. Choose any 2-copula C and set $C_{mn} = C^{n-m}$ where $m \leq n$ and the latter symbol denotes the $(n-m)$-fold $*$ product of C with itself (C^0 is by convention the copula M). The resulting family of copulas clearly satisfies (3.7), so that a Markov process is specified by supplying a sequence F_n of continuous marginal distributions. Processes constructed in this manner are analytically similar to Markov chains, and some of the theorems concerning Markov chains have direct analogs—including analogous proofs—for processes constructed in this manner. See, for example, Theorem 7.1 below. These processes need not be stationary, however, since the marginal distributions may vary with n.

Example 4.2. The preceding can be generalized as follows. Let T be the set of integers. To each k in T assign any 2-copula C_k. Then, for $m \leq n$ in T, set

$$C_{mn} = \begin{cases} M, & \text{if } m = n, \\ C_m * C_{m+1} * \cdots * C_{n-1}, & \text{if } m < n. \end{cases}$$

This yields a Markov process upon assigning a continuous distribution function to each element of T.

Example 4.3. One can calculate the copulas for a known Markov process, and then specify a new process by way of the same family of copulas and new marginal distributions. In this manner, for example, one can obtain a Brownian motion process with non-Gaussian marginal distributions. The transition probabilities for Brownian motion are given by

$$P(s, x, t, (-\infty, y]) = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{y} \exp\bigl(-(u-x)^2/2(t-s)\bigr)\, du$$

for $0 < s < t$. Since

$$P(s, x, t, (-\infty, y]) = C_{st,1}(F_s(x), F_t(y)),$$

we can integrate the expression above to obtain

$$(4.1)\quad C_{st}(F_s(x), F_t(y)) = \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{x} \left(\int_{-\infty}^{y} \exp(-(u-v)^2/2(t-s))\, du \right) dF_s(v)$$

for $0 < s < t$. Let

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)\, du.$$

If we assume that $X_0 = 0$ a.s., then $F_t(x) = G(x/\sqrt{t})$ for $t > 0$. Substituting this in (4.1) yields, after simplification

$$C_{st}(x, y) = \int_0^x G\left(\frac{\sqrt{t}\, G^{-1}(y) - \sqrt{s}\, G^{-1}(u)}{\sqrt{t-s}} \right) du.$$

This is a family of copulas which must, by construction, satisfy (3.7). Now any desired continuous non-Gaussian marginals can be specified to obtain a Brownian motion process with non-Gaussian marginals.

Example 4.4. A number of one-parameter families of copulas are known [4], [5], [11], [15]. Sometimes the copulas in a family satisfy

$$(4.2)\qquad C_\alpha * C_\beta = C_{\alpha\cdot\beta}$$

either directly or after a suitable change of parameter. Then we can set $\alpha_{st} = \exp(s - t)$ for $s \le t$ and also set $C_{st} = C_{\alpha_{st}}$; in this case (4.2) implies (3.7).

For example,

$$C_\alpha(x, y) = xy + 3\alpha xy(1-x)(1-y)$$

is a family of copulas for $-\frac{1}{3} \le \alpha \le \frac{1}{3}$, which, by direct calculation, satisfies (4.2). Defining C_{st} as above, we observe that $\lim_{t\to\infty} C_{st} = P$ for this family, so that the random variables in a Markov process constructed using this family become more nearly independent as $|s - t|$ gets larger.

Let W, P and M be as defined in the introduction. Since any convex combination of copulas is a copula, it follows that C_α given by

$$(4.3)\qquad C_\alpha = \frac{\alpha^2(1-\alpha)}{2} W + (1-\alpha^2) P + \frac{\alpha^2(1+\alpha)}{2} M$$

is a copula when $-1 \le \alpha \le 1$. Note that random variables whose connecting copula is C_α in this family are nearly independent if $|\alpha|$ is small, strongly positively correlated if α is near 1 and strongly negatively correlated if α is near -1. By direct calculation the copulas C_α of (4.3) satisfy (4.2) so that if we define C_{st} in the manner indicated above, we have once again $C_{rs} * C_{st} = C_{rt}$. We many now specify a Markov process by assigning any continuous marginal distributions.

Example 4.5. M.J. Frank has used the copulas W, P and M to construct a large class of families of copulas satisfying (3.7). Let α, β: $[0,\infty) \times [0,\infty) \to [0,1]$ be such that $\alpha, \beta \geq 0$ and $\alpha + \beta \leq 1$. Define

$$C_{st} = \alpha(s,t)W + (1 - \alpha(s,t) - \beta(s,t))P + \beta(s,t)M$$

for $0 \leq s \leq t < \infty$. It is easily checked that $C_{rs} * C_{st} = C_{rt}$ if and only if

$$\beta(r,s)\alpha(s,t) + \alpha(r,s)\beta(s,t) = \alpha(r,t) \tag{4.4a}$$

$$\alpha(r,s)\alpha(s,t) + \beta(r,s)\beta(s,t) = \beta(r,t). \tag{4.4b}$$

Set $f(s) = \alpha(0,s)$ and $g(s) = \beta(0,s)$. Then setting $r = 0$ in (4.4) we obtain the conditions

$$g(s)\alpha(s,t) + f(s)\beta(s,t) = f(t)$$
$$f(s)\alpha(s,t) + g(s)\beta(s,t) = g(t).$$

This system is readily solved to obtain

$$\alpha(s,t) = \frac{f(t)g(s) - f(s)g(t)}{g(s)^2 - f(s)^2}$$

$$\beta(s,t) = \frac{g(t)g(s) - f(t)f(s)}{g(s)^2 - f(s)^2}$$

$$\alpha(s,t) + \beta(s,t) = \frac{f(t) + g(t)}{f(s) + g(s)}.$$

It is easily verified that the following four conditions are sufficient (but not necessary) conditions for (4.4) and the other conditions on α and β to hold:

(i) f: $[0,\infty) \to [0,1]$ is increasing, with $f(0) = 0$;
(ii) g: $[0,\infty) \to [0,1]$ is decreasing, with $g(0) = 1$;
(iii) $f < g$ and $\sup f \leq \inf g$;
(iv) $f + g$ is decreasing.

Note that $\alpha + \beta = 1$ if and only if $f + g = 1$.

We give two simple examples of families of copulas obtained via the construction above.

First, let $f = 0$ and let g be any strictly decreasing function satisfying $g(0) = 1$ and $\lim_{s\to\infty} g(s) = 0$. Then

$$C_{st} = \left(1 - \frac{g(t)}{g(s)}\right)P + \frac{g(t)}{g(s)}M$$

for $0 \le s \le t$. Here, as above, $\lim_{t\to\infty} C_{st} = P$.

Second, let

$$f(t) = \frac{t}{2(1+t)} \quad \text{and} \quad g(t) = \frac{2+t}{2(1+t)}.$$

Then

$$C_{st} = \frac{t-s}{2(1+t)}W + \frac{2+t+s}{2(1+t)}M$$

for $0 \le s \le t$. In this case, $\lim_{t\to\infty} C_{st} = (W+M)/2$.

Example 4.6. A Markov chain is a Markov processes X_n, $n \in T$, for which T is the nonnegative integers, and in which the random variables X_n take values in a finite set of positive integers. The most thoroughly studied Markov chains are the stationary ones (stationary here means that the transition probabilities are all the same). In this example, we show how to associate a matrix of transition probabilities for a link in a Markov chain with the copula of the corresponding link. This is a case where the random variables are not continuous, so we will need to rely on the linear interpolation convention stated in the Introduction to obtain a unique such copula.

Let X_n denote a Markov chain taking values in the set $\{1,2,3\}$. Denote a matrix of transition probabilities of the process by Q, where $Q_{kl} = P(X_n = l | X_m = k)$, $n > m$. We want to calculate the corresponding copula $C = C_{mn}$. Let p_k denote the probability that $X_m = k$ (this probability can be computed from the initial distribution and the transition probabilities of prior states in the process). Write $\mathbf{p} = (p_1, p_2, p_3)$. Observe that the probability that $X_n = k$ is then $(\mathbf{p}Q)_k$. Observe also that the distribution function F_m is given by

$$\begin{aligned} F_m(1) &= p_1, \\ F_m(2) &= p_1 + p_2, \\ F_m(3) &= 1. \end{aligned}$$

It follows from Theorem 3.1 above that

$$C(F_m(x), F_n(y)) = \int_{-\infty}^{x} P(X_n \le y | X_m = \xi)\, d\xi.$$

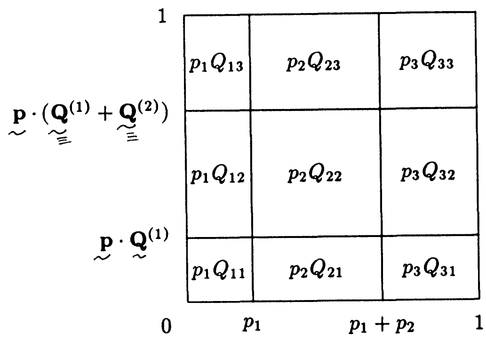

FIG. 1 Measure assigned by the copula C corresponding to the stochastic matrix Q to nine rectangles in the unit square.

The copula C corresponding to Q can be calculated explicitly from the information above and this formula, with the help of the linear interpolation convention. The measure induced by the resulting copula is depicted in Figure 1.

In the figure, $\mathbf{Q}^{(k)}$ denotes the kth column of Q. The measure of each of nine rectangles is indicated in the figure; mass is spread uniformly on each of the nine rectangles, by reason of the linear interpolation convention. The values of the copula at the vertices of the rectangles can easily be determined from the figure and the boundary conditions on copulas.

A principal fact to observe here is that the copula of Figure 1 depends not only on Q but also on the probabilities p_k. This says that if one varies the initial distribution holding the transition probabilities of the process fixed, the copulas of the process vary. This statement is complementary to an observation made above at the end of Section 3.

Example 4.7. Fix $a > 0$ and let C and E be copulas satisfying $C * E = E * C = C$ and $E * E = E$. Many copulas satisfying these conditions can be found; see Section 8 below. For $t > 0$ let

$$C_t = e^{-at}\left(E + \sum_{n=1}^{\infty} \frac{a^n t^n}{n!} C^n\right). \tag{4.5}$$

It is straightforward to verify that C_t is a copula and that $C_{s+t} = C_s C_t$ for all $s, t > 0$. The system $\{C_{st}: 0 \le s < t\}$ where $C_{st} = C_{t-s}$ gives rise to a continuous process analogous to that in Example 4.1. It is an open question whether or not every one parameter semigroup $t \to C_t$, $t > 0$ of copulas has the form (4.5).

5. Markov algebras

We showed in section 2 above that for any copulas A and B, the product $A * B$ is a copula, that the $*$ product is associative, that the product distributes over convex combinations, and that the product is continuous in each place. Recall also that P and M are null and unit elements, respectively, for the $*$ product, and that the set $\mathscr{C}$ of all copulas is a compact and convex set in the Banach space $C^0([0, 1]^2)$ of continuous functions on the unit square under the uniform norm. Observe finally that if A is a copula, then A^T defined by $A^T(x, y) = A(y, x)$ is also a copula. This motivates the following definition:

DEFINITION 5.1. A *Markov algebra* is a compact convex subset $\mathscr{A}$ of a real Banach space on which a product $(\xi, \eta) \to \xi\eta$ is defined which is associative, which distributes over convex combinations, which is continuous in each place (but not necessarily jointly continuous), and which possesses unit and null elements.

This definition is not a purely axiomatic one, since it requires that there be an underlying Banach space. Compare [20], in which the term "Markov algebra" is used in the context of Markov chains with a different but related meaning.

DEFINITION 5.2. A Markov algebra is *symmetric* if it possesses a continuous operation $\eta \to \eta^T$ satisfying $(\eta^T)^T = \eta$, $(\lambda\eta + (1-\lambda)\xi)^T = \lambda\eta^T + (1-\lambda)\xi^T$ and $(\eta\xi)^T = \xi^T\eta^T$.

By remarks above, we have the following theorem:

THEOREM 5.1. *The set $\mathscr{C}$ of all 2-copulas is a symmetric Markov algebra under $*$ and ^{T} as previously defined. The unit and null elements are given by*

$$M(x, y) = \min(x, y) \quad \textit{and} \quad P(x, y) = xy$$

respectively.

There is one other interesting example, or rather class of examples, of Markov algebras, in addition to copulas under the $*$ product. A stochastic matrix is a matrix whose entries are nonnegative and each of whose rows sums to 1; the set of $n \times n$ stochastic matrices will be denoted $\mathscr{S}^n$. A doubly stochastic matrix is a matrix whose entries are nonnegative and each of whose rows and columns sums to 1; the set of $n \times n$ doubly stochastic matrices will be denoted $\mathscr{D}^n$. Clearly each of these sets is convex. It is easy to verify also that each is closed under matrix multiplication and that each forms a compact set in R^{n^2}. Observe also that the (matrix) transpose of a doubly stochastic matrix is doubly stochastic. $\mathscr{D}^n$ possesses a unit element (the $n \times n$ identity matrix, which we shall here denote M_n) and a null element (the $n \times n$ matrix each of whose entries is $1/n$, which we shall here denote P_n). Therefore, $\mathscr{D}^n$, like $\mathscr{C}$, is a symmetric Markov algebra. $\mathscr{S}^n$, on the other hand, is not symmetric; it also does not possess a null element. To see this, observe first that if $Q \in \mathscr{S}^n$ has all rows equal, then $PQ = Q$ for all $P \in \mathscr{S}^n$. It follows that $\mathscr{S}^n$ possesses a multiplicity of right null elements; but then it possesses no left null element, since by a standard algebraic argument, every right null element must equal every left null element, which is impossible if there is a multiplicity of right null elements. That $\mathscr{S}^n$ is not symmetric follows as a corollary—the transpose of a right null element must be a left null element, so there can be no transpose operation satisfying the symmetry axiom.

It is $\mathscr{S}^n$ which plays a central role in the theory of Markov chains. The set $\mathscr{C}$ of copulas under the $*$ operation is a sort of generalization of the set of singly stochastic matrices under matrix multiplication, but as the argument above makes clear, it has some additional algebraic structure not present in $\mathscr{S}^n$, namely a (unique) null element and a transpose operation.

The goal of the remainder of the paper is to investigate the algebraic properties of the Markov algebra $\mathscr{C}$ and to interpret those properties in the context of Markov processes. We begin with the algebras $\mathscr{D}^n$.

6. The algebras $\mathscr{D}^n$

We discuss here properties of the algebras $\mathscr{D}^n$. In later sections, we will address the question of whether the algebra $\mathscr{C}$ has similar properties.

An element $A \in \mathscr{D}^n$ is *extreme* if $A = \lambda B + (1 - \lambda)C$ and $\lambda \in (0, 1)$ implies $B = C = A$.

THEOREM 6.1. *The extreme points of $\mathscr{D}^n$ are the $n \times n$ permutation matrices.*

This is a well known theorem, attributed to Birkhoff and von Neumann. [10].

THEOREM 6.2. *An element Q of $\mathscr{D}^n$ has an inverse in $\mathscr{D}^n$ if and only if Q is extreme. In this case, the inverse of Q is Q^T.*

Proof. By the previous theorem, if Q is extreme, it is a permutation matrix, and therefore possesses an inverse, which is its transpose. On the other hand, if Q possesses an inverse and $Q = \lambda A + (1 - \lambda)B$ for some $\lambda \in (0, 1)$, then necessarily

$$M = Q^{-1}Q = \lambda Q^{-1}A + (1 - \lambda)Q^{-1}B.$$

Since M is extreme, it follows that $Q^{-1}A = Q^{-1}B = M$, and therefore that $A = B = Q$. Thus, Q is extreme, and by Theorem 6.1 above it is therefore a permutation matrix whose inverse is necessarily Q^T. ∎

Thus, the invertible elements of $\mathscr{D}^n$ are the extreme elements. Since $\mathscr{D}^n$ is a compact and convex set in R^{n^2}, it follows from the Krein-Milman theorem that every element of $\mathscr{D}^n$ is a convex combination of permutation matrices.

An element $Q \in \mathscr{D}^n$ is *idempotent* if $Q^2 = Q$.

THEOREM 6.3. *The number of idempotents in $\mathscr{D}^n$ is finite. Each of them is symmetric. Moreover, $E \in \mathscr{D}^n$ is idempotent if and only if there is an integer $k \geq 1$ and there are integers $m_j \geq 1$, $j = 1, \dots, k$ whose sum is n and there is an $n \times n$ permutation matrix Q, such that*

$$E = Q^T \begin{pmatrix} P_{m_1} & 0 & \cdots & 0 \\ 0 & P_{m_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_{m_k} \end{pmatrix} Q$$

where P_{m_j} denotes the null element in $\mathscr{D}^{m_j}$.

Proof. Since this theorem does not seem to be well known, we outline a proof. Let $E \in \mathscr{D}^n$ be idempotent but not null. If E has a column with every entry positive then $E^m \to P_n$, by a basic result concerning Markov chains. [7, p. 173]. But then E idempotent implies $E = P_n$, contrary to hypothesis. Consequently, each column has a zero entry. Let $\mathscr{I}_k$ denote the indices j such that $E_{jk} \neq 0$. Then since E is idempotent, we have for all indices j,

$$E_{jk} = \sum_{l=1}^{n} E_{jl}E_{lk} = \sum_{l \in \mathscr{I}_k} E_{jl}E_{lk}.$$

Since E is doubly stochastic, we obtain

$$1 = \sum_{j \in \mathscr{I}_k} E_{jk} = \sum_{l \in \mathscr{I}_k} \left(\sum_{j \in \mathscr{I}_k} E_{jl} \right) E_{lk} \leq \sum_{l \in \mathscr{I}_k} E_{lk} = 1.$$

If $\sum_{j \in \mathscr{I}_k} E_{jl} < 1$ for any $l \in \mathscr{I}_k$, the foregoing relation would imply $1 < 1$, a contradiction. Therefore, necessarily $\mathscr{I}_l \subset \mathscr{I}_k$ for all $l \in \mathscr{I}_k$. It follows from this (by an argument we omit) that $k \in \mathscr{I}_k$, that is, that necessarily $E_{kk} \neq 0$ for any idempotent E and all $k = 1, \dots, n$. It also follows that there is a permutation matrix Q such that

$$E = Q^T \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} Q$$

where A is $m \times m$ and $m = \#\mathscr{I}_k$. Since the partitioned matrix on the right is doubly stochastic, the sum of the entries in C is $n - m$, and the sum of the entries in B is therefore necessarily zero. It follows that B is the zero matrix and hence that A and C are both idempotents. The proof can be completed by induction on n; the theorem is true by direct calculation for $n = 2$. ■

It is easy to show that the only invertible idempotent (and therefore the only extreme idempotent) is M_n. There is a natural partial ordering of idempotents: $E \preccurlyeq F$ if and only if E and F commute, and $EF = FE = E$. It is associated with a pointwise ordering of diagonal elements:

THEOREM 6.4. *If $E, F \in \mathscr{D}^n$ are idempotents and $E \preccurlyeq F$, then $E_{kk} \le F_{kk}$ for $k = 1, \dots, n$. Partial converse: If E and F are commuting idempotents and $[EF]_{kk} = E_{kk}$ for $k = 1, \dots, n$, then $E \preccurlyeq F$.*

Proof. The proof of this theorem is similar to the proof of Theorems 8.5 and 8.6 below.

The *center* of $\mathscr{D}^n$ is the set of elements of $\mathscr{D}^n$ that commute with all elements of $\mathscr{D}^n$.

THEOREM 6.5. *The center of $\mathscr{D}^n$ is the interval $[P_n, M_n] = \{(1 - \lambda)P_n + \lambda M_n \colon 0 \le \lambda \le 1\}$.*

Proof. Clearly any element of $[P_n, M_n]$ is in the center. Suppose that A is in the center. Fix $i \ne j$ and let Q be the permutation matrix that interchanges i and j. Then since $QA = AQ$, necessarily

$$\begin{aligned} A_{ik} &= A_{jk}, \quad k \ne i, j, \\ A_{ki} &= A_{kj}, \quad k \ne i, j, \\ A_{ij} &= A_{ji}, \\ A_{ii} &= A_{jj}. \end{aligned}$$

Upon varying i and j it is easy to see that any two diagonal entries are the same and that any two off diagonal entries are the same. Thus, A has the required form. ■

We do not assert that this is an exhaustive list of interesting properties of the algebras $\mathscr{D}^n$. One can easily construct ideals in $\mathscr{D}^n$ (sets closed under convex combinations and under multiplication on one side by arbitrary elements of $\mathscr{D}^n$); we can prove no interesting facts about the ideals and so will not mention them further. Also, $\mathscr{D}^n$ contains divisors of zero for at least some values of n; for example, if

$$E_1 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \qquad E_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

then E_1 and E_2 are idempotents (by Theorem 6.3) and $E_1E_2 = E_2E_1 = P_4$, by direct computation. We can make no general statements, however, about the divisors of zero in the algebras $\mathscr{D}^n$.

The properties above are not exhaustive, but we shall take them as a framework for analyzing the algebra $\mathscr{C}$ of interest.

7. Invertible and extreme elements of $\mathscr{C}$

Henceforth, we denote the $*$ product of A and $B \in \mathscr{C}$ by AB.

Extreme elements in $\mathscr{C}$ are defined analogously to extreme elements in $\mathscr{D}^n$. For $A, B \in \mathscr{C}$, if $AB = M$ we say that A is a *left inverse* of B and B is a *right inverse* of A. If $C \in \mathscr{C}$ has both a left inverse L and a right inverse R, then necessarily $L = R$ and the common value is the *inverse* of C; in this case we say C is *invertible*. (Observe that by known properties of matrices, an element $A \in \mathscr{D}^n$ has a left inverse if and only if it has a right inverse, so it was unnecessary to make the distinction made here in discussion of the algebras $\mathscr{D}^n$.)

THEOREM 7.1. *C in $\mathscr{C}$ has a left inverse if and only if for each $y \in [0,1]$, $C_{,1}(x, y) = 0$ or 1 for almost all $x \in [0,1]$; and if C has a left inverse, then C^T is a left inverse of C. C has a right inverse if and only if, for each $x \in [0,1]$, $C_{,2}(x, y) = 0$ or 1 for almost all $y \in [0,1]$; and if C has a right inverse, then C^T is a right inverse.*

Proof. We prove the first statement only; the second follows from the first by taking transposes. Suppose that for each y, $C_{,1} = 0$ or 1 for almost all x. Since for almost all x the map $y \to C_{,1}(x, y)$ is nondecreasing (see Section 2 above), it follows that for $u \le v$, $C_{,1}(x,u)C_{,1}(x,v) = C_{,1}(x,u)$ for almost all

x. Hence,

$$\begin{aligned} C^T C(x,y) &= \int_0^1 C^T_{,2}(x,t) C_{,1}(t,y)\,dt \\ &= \int_0^1 C_{,1}(t,x) C_{,1}(t,y)\,dt \\ &= \int_0^1 C_{,1}(t,\min\{x,y\})\,dt \\ &= \min\{x,y\} \\ &= M(x,y). \end{aligned}$$

Thus, C has a left inverse, and C^T is a left inverse of C.

For the converse, suppose $LC = M$. Then for all y

$$\begin{aligned} y &= \int_0^1 L_{,2}(y,t) C_{,1}(t,y)\,dt \\ &\le \left(\int_0^1 L_{,2}(y,t)^2\,dt\right)^{1/2} \left(\int_0^1 C_{,1}(t,y)^2\,dt\right)^{1/2} \\ &\le y^{1/2}\left(\int_0^1 C_{,1}(t,y)^2\,dt\right)^{1/2} \\ &\le y^{1/2}\left(\int_0^1 C_{,1}(t,y)\,dt\right)^{1/2} \\ &= y^{1/2} y^{1/2} \\ &= y. \end{aligned}$$

This uses Schwartz's inequality and the fact that the first partial derivatives of a copula are sandwiched between 0 and 1 almost certainly (see Section 2 above). It follows that equality must hold at each step in the foregoing chain, so that, from lines 3 and 4, for all $y > 0$,

$$\int_0^1 \left[C_{,1}(t,y) - C_{,1}(t,y)^2\right] dt = 0.$$

Since the integrand in this expression is almost certainly positive, it follows that for all $y > 0$, $C_{,1}(x,y) = 0$ or 1 for almost all x, as required. When $y = 0$ $C_{,1}(x,y) = 0$ for all x, by the boundary condition satisfied by C. This completes the proof. ∎

THEOREM 7.2. *Any element of $\mathscr{C}$ that possesses a left or right inverse is extreme.*

Proof. Suppose C is left invertible. Then by Theorem 7.1 there exist disjoint sets U and V such that $U \cup V = [0,1]^2$, $C_{,1} = 0$ almost everywhere on U and $C_{,2} = 1$ almost everywhere on V. Suppose $C = \lambda A + (1-\lambda)B$ for some $\lambda \in (0,1)$. Then $C_{,1} = \lambda A_{,1} + (1-\lambda)B_{,1}$ almost everywhere. Since $0 \le A_{,1}, B_{,1} \le 1$ almost everywhere (see Section 2 above), necessarily $A_{,1} = B_{,1} = 0$ almost everywhere on U and $A_{,1} = B_{,1} = 1$ almost everywhere on V. It follows that $A_{,1} = B_{,1} = C_{,1}$ almost everywhere, and thus, upon integration, that $A = B = C$. Therefore, C is indeed extreme. ■

THEOREM 7.3. *Left and right inverses in $\mathscr{C}$ are unique.*

Proof. Suppose C has left inverses A and B. Then $\frac{1}{2}A + \frac{1}{2}B$ is also a left inverse of C and is therefore, by Theorem 7.2, extreme. Consequently $A = B$. ■

Remarks. 1. An element of $\mathscr{C}$ may have an inverse on one side but not on the other. Here is a 1-parameter family of copulas that are right invertible but not left invertible.

For $0 \le \lambda \le 1$, let

$$C_\lambda(x,y) = \begin{cases} y, & \text{if } y \le \lambda x, \\ \lambda x, & \text{if } \lambda x < y \le 1 - (1-\lambda)x, \\ x + y - 1, & \text{if } 1 - (1-\lambda)x < y \le 1 \end{cases}$$

which can be pictured as indicated in Figure 2.

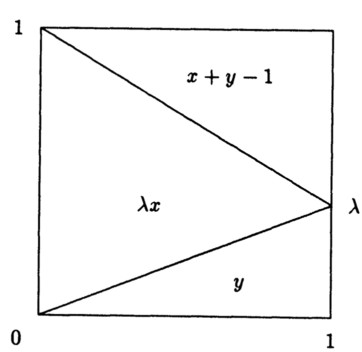

FIG. 2 Value of $C_\lambda(x, y)$ in three regions in the unit square.

Each C_λ is in $\mathscr{C}$. In particular $C_0 = W$ and $C_1 = M$. For $0 < \lambda < 1$, $C_{,2} = 0$ or 1 *a.s.*, but $C_{,1} = \lambda$ on a set of positive measure. Thus, the C_λ are right but not left invertible when $0 < \lambda < 1$.

2. Consider the analogy with $\mathscr{D}^n$. In $\mathscr{C}$ there are elements which are invertible on one side only; this is not true in $\mathscr{D}^n$. In $\mathscr{C}$ all left or right invertible elements are extreme; this is true in $\mathscr{D}^n$. It is an open question whether all extreme elements in $\mathscr{C}$ are necessarily either left or right invertible; in $\mathscr{D}^n$, an extreme element is necessarily invertible.

3. Observe that if A has a left inverse then the map $C \to ACA^T$ is a continuous homomorphism of $\mathscr{C}$. It maps M to AA^T, which is equal to M only if A has a right inverse. (We do not require that a homomorphism map the null element into the null element or the unit into the unit; this would be too restrictive; we require only that products, convex combinations and transposes be preserved. The image of the null element is then necessarily a null element in the range, but not necessarily of the algebra of which the range is a subset, and similarly for the unit. Other properties of Markov algebra homomorphisms are easy to figure out, and we omit discussion of them.) Observe that the homomorphism here constructed is necessarily injective for all left invertible copulas A but that it is onto only if A is also right invertible. Thus, the homomorphism establishes a one to one correspondence between $\mathscr{C}$ and a proper subset of itself which has the same algebraic structure, if A is left but not right invertible.

The Krein-Milman theorem applies to $\mathscr{C}$ as well as to $\mathscr{D}^n$, and we can conclude that $\mathscr{C}$ is the closure of the convex hull of its extreme points. We can actually prove a stronger result: that the invertible elements of $\mathscr{C}$ are dense in $\mathscr{C}$. To prove this, we need first to describe a useful construction.

Let $\mathscr{P}$: $0 = x_0 < x_1 < \cdots < x_n = 1$ be a partition of $[0,1]$, and let σ be a permutation of $\{1, 2, \ldots, n\}$. Let A be a copula with induced measure α. Let σ shuffle the vertical strips

$$V_k = [x_{k-1}, x_k) \times [0,1]$$

and carry along the measure α restricted to each of the strips. This will yield a doubly stochastic measure α_σ on $[0,1]^2$. More precisely, set $\Delta x_k = x_k - x_{k-1}$ and let $\mathscr{P}_\sigma$: $0 = u_0 < \cdots < u_n = 1$ be the partition for which $\Delta u_k = \Delta x_{\sigma(k)}$. Define α_σ via

$$\begin{aligned}(7.1)\quad & \alpha_\sigma(B \cap [u_{k-1}, u_k) \times [0,1]) \\ & \qquad = \alpha\big(\big[(x_{\sigma(k)} - u_k)\hat{\mathbf{e}}_1 + B\big] \cap [x_{\sigma(k)-1}, x_{\sigma(k)}) \times [0,1]\big)\end{aligned}$$

for all Borel sets B. Here $\hat{\mathbf{e}}_1$ denotes the unit vector along the x-axis. The copula A_σ induced by α_σ is called a *horizontal shuffle* of A. Similarly, there is a *vertical shuffle* A^σ of A which equals $((A^T)_\sigma)^T$. This definition is adapted from Mikusinski et al. [12], but it is not quite identical to their definition.

THEOREM 7.4. *Every vertical and horizontal shuffle of a left (right) invertible copula is left (right) invertible.*

Proof. This is an immediate consequence of the characterization in Theorem 7.1. Let A be left invertible, so that for all y, $A_{,1}(x, y) = 0$ or 1 for almost all x. Then for all y and almost all $x \in (u_{k-1}, u_k)$ we have (using (7.1))

$$\begin{aligned} A_{\sigma,1}(x, y) &= \frac{\partial}{\partial x}\alpha_\sigma([0, x] \times [0, y]) \\ &= \frac{\partial}{\partial x}\alpha_\sigma([u_{k-1}, x] \times [0, y]) \\ &= \frac{\partial}{\partial x}\alpha\big([x_{\sigma(k)-1}, x_{\sigma(k)-1} + (x - uk - 1] \times [0, y]\big) \\ &= A_{,1}\big(x_{\sigma(k)-1} + (x - u_{k-1}), y\big) \\ &= 0 \text{ or } 1. \end{aligned}$$

The desired result follows. If on the other hand A is right invertible, so that for all x, $A_{,2}(x, y) = 0$ or 1 for almost all y, then for all $x \in [u_{k-1}, u_k)$ and almost all y,

$$\begin{aligned} A_{\sigma,2}(x, y) &= \frac{\partial}{\partial y}\alpha_\sigma([0, x] \times [0, y]) \\ &= \frac{\partial}{\partial y}\left(\alpha_\sigma([u_{k-1}, x] \times [0, y]) + \sum_{j=1}^{k-1}\alpha_\sigma([u_{j-1}, u_j] \times [0, y])\right) \\ &= \frac{\partial}{\partial y}\left(\alpha\big([x_{\sigma(k)-1}, x_{\sigma(k)-1} + (x - u_{k-1})] \times [0, y]\big)\right. \\ &\qquad \left. + \sum_{j=1}^{k-1}\alpha\big([x_{\sigma(j)-1}, x_{\sigma(j)}] \times [0, y]\big)\right) \\ &= A_{,2}\big(x_{\sigma(k)-1} + (x - u_{k-1}), y\big) - A_{,2}(x_{\sigma(k)-1}, y) \\ &\quad + \sum_{j=1}^{k-1}\left(A_{,2}(x_{\sigma(j)}, y) - A_{,2}(x_{\sigma(j)-1}, y)\right) \\ &= \text{sum of 0's, 1's and } (-1)\text{'s lying necessarily in } [0, 1] \\ &= 0 \text{ or } 1. \end{aligned}$$

Again, the desired result follows. ■

THEOREM 7.5. *Shuffles of M are dense in $\mathscr{C}$.*

Proof. For a proof of this theorem using copulas, see Mikusinski et al. [12]. The theorem can be formulated and proved without using copulas; see Vitale [18], [19].

As a corollary of Theorem 7.5 we have:

THEOREM 7.6. *The $*$ product is not jointly continuous on $\mathscr{C}$.*

Proof. By Theorems 7.4 and 7.5 there is a sequence $C_n \to P$ for which each C_n is invertible. Then $C_n^T \to P^T = P$, since the transpose operation is continuous. Thus, if the $*$ product were jointly continuous, we would have $C_n^T C_n \to P^2 = P$. But $C_n^T C_n = M$ for all n and $M \neq P$. ■

8. Idempotents in $\mathscr{C}$

An *idempotent* in $\mathscr{C}$ is any 2-copula E for which $E^2 = E$. Constructions introduced in this and in the next section guarantee the existence of idempotents in $\mathscr{C}$ other than M and P. They arise in another way.

THEOREM 8.1. *For C in $\mathscr{C}$, define $\hat{C}_n$ via*

$$\hat{C}_n = 1/n \sum_{k=1}^{n} C^k.$$

The sequence $\hat{C}_n$ converges uniformly, and its limit E is idempotent. Furthermore, $EC = CE = E$.

Proof. The proof parallels the proof of the analogous theorem for stochastic matrices given in [7, p. 175]. ■

Our results for idempotents in $\mathscr{C}$ are fragmentary. The questions we address first are those raised by the characterization of idempotents of $\mathscr{D}^n$ in Theorem 6.3. First, are the idempotents in $\mathscr{C}$ necessarily symmetric? This remains an open question. Second, is there an analog of the direct sum decomposition of Theorem 6.3, applicable to idempotents in $\mathscr{C}$? There is an analog of the direct sum construction, and it has interest independent of the application we will make of it here. We digress to introduce this construction; we shall return to the question concerning characterization of idempotents below.

The construction is called the ordinal sum construction, and it is well known. [4], [16], [17]. In this context, a partition of $[0, 1]$ denotes a finite or countable family $\{(a_n, b_n)\}$ of disjoint intervals the union of whose closures is

$[0,1]$. To each n assign a 2-copula C_n. The function C defined by

$$C(x,y) = \begin{cases} a_n + (b_n - a_n) C_n\left(\dfrac{x - a_n}{b_n - a_n}, \dfrac{y - a_n}{b_n - a_n}\right), & \text{if } (x,y) \in [a_n, b_n]^2, \\ M(x,y), & \text{otherwise} \end{cases}$$

is a 2-copula, called the *ordinal sum* of the copulas C_n with respect to the partition $\{(a_n, b_n)\}$.

THEOREM 8.2. *A 2-copula C has an ordinal sum decomposition with respect to a partition $\{(a_n, b_n)\}$ of $[0,1]$ if and only if the end points of the intervals (a_n, b_n) are fixed points of the map $x \to C(x,x)$ for all n. In this case, the functions C_n defined by*

$$C_n(x,y) = \frac{C(a_n + (b_n - a_n)x, a_n + (b_n - a_n)y) - a_n}{b_n - a_n}$$

are copulas and C is their ordinal sum with respect to the partition.

Proof. The proof is nearly the same as the proof of the well-known analogous theorem for associative binary operations. See [4, 16]. ∎

THEOREM 8.3. *Suppose A and B are 2-copulas that are ordinal sums of 2-copulas A_n and B_n with respect to the same partition of* $[0,1]$. *Then, AB is an ordinal sum of the A_nB_n with respect to that partition.*

Proof. Induction; the theorem is true when there are two subintervals in the partition by direct calculation. ∎

It follows from Theorem 8.3 that an ordinal sum of idempotents is idempotent. It is also true that if E is idempotent and Q is right invertible, then Q^TEQ is idempotent. The question raised by Theorem 6.3 can now be formulated more precisely: Can every idempotent be decomposed in the form Q^TEQ where Q is right invertible and E is an ordinal sum of copies of P and M? This is an open question; we conjecture that the answer is yes.

The ordinal sum construction yields a continuous one-parameter family of idempotents in $\mathscr{C}$ connecting P and M, which we will use later. Consider the intervals $(0, 1-\lambda)$ and $(1-\lambda, 1)$ for $0 < \lambda < 1$. With respect to these there is an ordinal sum of P and M:

$$E_\lambda(x,y) = \begin{cases} xy/(1-\lambda), & \text{if } 0 \le x, y \le 1-\lambda, \\ M(x,y), & \text{otherwise.} \end{cases}$$

Also let $E_0 = P$ and $E_1 = M$. From Theorem 8.3 it follows that E_λ is idempotent for $0 \neq \lambda \neq 1$.

THEOREM 8.4. *For* $0 \neq \mu \neq \lambda \neq 1$, $\|E_\lambda - E_\mu\| \neq \lambda - \mu$ *and* $E_\lambda E_\mu = E_\mu E_\lambda = E_\mu$.

Proof. If $0 < \mu < \lambda$, clearly $E_\lambda(1-\mu, 1-\mu) = 1 - \mu$. Consequently, by Theorem 8.2, E_λ has an ordinal sum decomposition with respect to the partition $(0, 1-\mu)$, $(1-\mu, 1)$. It is readily verified that it is the ordinal sum of

$$C_1 = E_{(\lambda-\mu)/(1-\mu)} \quad \text{and} \quad C_2 = M$$

with respect to this partition. It follows at once, from Theorem 8.3, that $E_\lambda E_\mu = E_\mu E_\lambda = E_\mu$.

For the inequality of the first part,

$$E_\lambda(x, y) - E_\mu(x, y) = \begin{cases} \dfrac{(\lambda-\mu)xy}{(1-\lambda)(1-\mu)}, & \text{if } 0 \le x, y \le 1 - \lambda, \\ M(x, y) - xy/(1-\mu), & \text{if } 1 - \lambda < x, y \le 1 - \mu, \\ 0, & \text{otherwise}. \end{cases}$$

It remains to show that the absolute value of each of these is no larger than $\lambda - \mu$. The last is trivial. The first, at once, since $x \le 1 - \lambda$ and $y \le 1 - \lambda < 1 - \mu$. For the middle, there is no loss to assume that $x \le y$. Then

$$\left| M(x, y) - \frac{xy}{1-\mu} \right| = x\left| 1 - \frac{y}{1-\mu} \right| = \frac{x}{1-\mu}|1 - \mu - y|.$$

But $y \le 1 - \mu$ so that it suffices to show that $x(1 - \mu - y) \le (1-\mu)(\lambda - \mu)$ and, thus, that $1 - \mu - y \le \lambda - \mu$ which holds since $1 - y \le \lambda$. ∎

We give one additional construction using idempotents and the ordinal sum. Let $\{(a_n, b_n)\}_{n \in \mathscr{I}}$ be a partition of $[0, 1]$, and let $\mathscr{I}_1$ and $\mathscr{I}_2$ be disjoint index sets such that $\mathscr{I} = \mathscr{I}_1 \cup \mathscr{I}_2$. To each index $n \in \mathscr{I}_1$ assign an idempotent E_n. Define a map $\Psi : \mathscr{C} \to \mathscr{C}$ as follows: $\Psi(C)$ is the ordinal sum of the E_n's, corresponding to indices in $\mathscr{I}_1$, and copies of C itself, corresponding to indices in $\mathscr{I}_2$. By Theorem 8.3, this assignment is a homomorphism; it is injective, so long as $\mathscr{I}_2$ is not empty; it is onto only if $\mathscr{I}_1$ is empty and the partition collapses to the single interval $(0, 1)$.

There is a partial ordering $\preccurlyeq$ for idempotents in $\mathscr{C}$ which is identical to that for idempotents in $\mathscr{D}^n$. For idempotents E and F define $E \preccurlyeq F$ if and only if $EF = FE = E$. It is clear that $P \preccurlyeq E \preccurlyeq M$ for all idempotents E. Note that the idempotents E_λ of the preceding theorem with respect to $\preccurlyeq$ are order isomorphic to $[0, 1]$.

For C in $\mathscr{C}$ define the *diagonal* δ_C by

$$\delta_C(x) = C(x, x)$$

for all x in $[0, 1]$.

THEOREM 8.5. *If E and F are symmetric idempotents and $E \preccurlyeq F$, then $\delta_E \leq \delta_F$.*

Proof. We have by assumption that $E^2 = E$, $F^2 = F$, $EF = FE = E$, $E^T = E$, and $F^T = F$, whereupon

$$\begin{aligned}\delta_F(x) - \delta_E(x) &= F^2(x, x) - 2FE(x, x) + E^2(x, x) \\ &= \int_0^1 \left[F_{,2}(x, t)^2 - 2F_{,2}(x, t)E_{,2}(x, t) + E_{,2}(x, t)^2\right] dt \\ &= \int_0^1 [F_{,2}(x, t) - E_{,2}(x, t)]^2 \, dt \\ &\geq 0\end{aligned}$$

as needed. ■

There is a partial converse to this.

THEOREM 8.6. *Let E and F be commuting symmetric idempotents such that $\delta_{EF} = \delta_E$. Then, $E \preccurlyeq F$.*

Proof. Under the assumptions on E and F it follows at once that EF is also a symmetric idempotent and that $EF \preccurlyeq E, F$. Also, $EFE = E^2F = EF = (EF)^2$ so that

$$\int_0^1 (EF)_{,2}(x, t)[E_{,1}(t, x) - (EF)_{,1}(t, x)] \, dt = 0. \tag{8.1}$$

But we also have $\delta_{EF} = \delta_E$ so that $\delta_{E^2F} = \delta_{E^2}$. Consequently

$$\int_0^1 E_{,2}(x, t)[E_{,1}(t, x) - (EF)_{,1}(t, x)] \, dt = 0. \tag{8.2}$$

Subtract (8.1) from (8.2) and use the fact that E and EF are symmetric to get

$$\int_0^1 [E_{,1}(t, x) - (EF)_{,1}(t, x)]^2 \, dt = 0$$

so that $E_{,1} = (EF)_{,1}$ almost everywhere, whereupon $E = EF$. ■

As a corollary to this there is:

THEOREM 8.7. *There is only one symmetric idempotent whose diagonal is the function $x \to x^2$, and it is P.*

Proof. Suppose E is idempotent and $\delta_E(x) = x^2$ for all x in $[0, 1]$. Since $EP = PE = P$, then $\delta_{EP}(x) = \delta_P(x) = x^2 = \delta_E(x)$ for all x, whence $\delta_{EP} = \delta_E$. By the preceeding theorem, then, $E = EP = P$. ■

Note that M is the only 2-copula whose diagonal is the identity map.

9. The center of $\mathscr{C}$

The *center* of $\mathscr{C}$ is the set of elements of $\mathscr{C}$ that commute with all the elements of $\mathscr{C}$. We showed in Theorem 6.5 that the center of $\mathscr{D}_n$ is the interval

$$[P_n, M_n] = \{(1-t)P_n + tM_n : 0 \le t \le 1\}.$$

In this section, we prove the analogous result for $\mathscr{C}$. The proofs are the same, except that the permutation matrices are replaced by shuffles of M. We shall need a lemma of interest in its own right.

Recall the definition in section 7 of a horizontal shuffle A_σ of a copula A with respect to a partition $\mathscr{P}$: $0 = x_0 < x_1 < \cdots < x_n = 1$ and a permutation σ. The partition $\mathscr{P}$ and σ induce a second partition $\mathscr{P}'$: $0 = u_0 < u_1 < \cdots < u_n = 1$ of $[0, 1]$ for which $\Delta u_k = \Delta x_{\sigma(k)}$. For $u_{k-1} \le x < u_k$ and $0 \le y \le 1$,

$$(9.1) \qquad A_\sigma(x, y) = A\big(x_{\sigma(k)-1} + (x - u_{k-1}), y\big) - A(x_{\sigma(k)-1}, y) + \sum_{j<k} \big[A(x_{\sigma(j)}, y) - A(x_{\sigma(j)-1}, y)\big].$$

LEMMA 9.1. *For any partition $\mathscr{P}$: $0 = x_0 < \cdots < x_n = 1$ of $[0, 1]$ and any permutation σ,*

$$A_\sigma = M_\sigma A$$

$$A^\sigma = AM^\sigma$$

for all $A \in \mathscr{C}$.

Proof. Note that the second statement follows from the first upon taking transposes. For $x \in [u_{k-1}, u_k)$,

$$\begin{aligned} M_\sigma A(x, y) &= \int_0^1 \Big[M_{,2}\big(x_{\sigma(k)-1} + (x - u_{k-1}), t\big) - M_{,2}\big(x_{\sigma(k)-1}, t\big) \\ &\qquad + \sum_{j<k} \big(M_{,2}(x_{\sigma(j)}, t) - M_{,2}(x_{\sigma(j)-1}, t)\big)\Big] A_{,1}(t, y)\, dt \\ &= A\big(x_{\sigma(k)-1} + (x - u_{k-1}), y\big) - A(x_{\sigma(k)-1}, y) \\ &\quad + \sum_{j<k} \big(A(x_{\sigma(j)}, y) - A(x_{\sigma(j)-1}, y)\big) \\ &= A_\sigma(x, y). \end{aligned}$$

The first and last equalities are obtained by applying (9.1) to M and A. The second holds because $M_{,2}(x, y) = \chi_{[0,x]}(y)$ for almost all y. ∎

THEOREM 9.1. *The center of $\mathscr{C}$ is the interval $[P, M] = \{(1-t)\{+tM: 0 \le t \le 1\}$.*

Proof. The proof is analogous to that of Theorem 6.5. Clearly $(1-t)P + tM$ is in the center for $0 \le t \le 1$. Suppose that A is in the center. Consider any (possibly degenerate) partition

$$\mathscr{P}: 0 = x_0 \le x_1 < x_2 \le x_3 < x_4 \le x_5 = 1$$

for which $x_4 - x_3 = x_2 - x_1 = \delta > 0$ and $\delta \le \frac{1}{4}$. Let σ be the 2,4 transposition of $\{1, 2, \ldots, 5\}$. It is readily verified that in this case $M^\sigma = (M_\sigma)^T = M_\sigma$. Therefore, by Lemma 9.1, $A_\sigma = A^\sigma$, since A lies in the center. Consequently $\alpha_\sigma = \alpha^\sigma$, where α_σ and α^σ are the doubly stochastic measure induced by A_σ and A^σ, respectively. Let α denote the doubly stochastic measure induced by A. Let $R_{ij} = [x_{i-1}, x_i) \times [x_{j-1}, x_j)$ and set $\varepsilon = x_3 - x_1$. It follows readily that for any Borel set B,

$$\begin{aligned} \alpha(B) &= \alpha(B + \varepsilon\hat{\mathbf{e}}_2), \quad \text{if } B \subset R_{12} \cup R_{32} \cup R_{52} \\ \alpha(B) &= \alpha(B + \varepsilon\hat{\mathbf{e}}_1), \quad \text{if } B \subset R_{21} \cup R_{23} \cup R_{25} \\ \alpha(R_{24}) &= \alpha(R_{42}) \\ \alpha(R_{22}) &= \alpha(R_{44}). \end{aligned}$$

It follows from this, upon varying x_1 and ε for fixed $\delta \le \frac{1}{4}$, that any two squares of side δ with sides parallel to those of $[0, 1]^2$, which do not intersect the line $y = x$ except possibly at a vertex, have the same α-measure and also that any two squares of side δ with opposite vertices on the line have the same α-measure.

Let S be a square of side δ with sides parallel to those of $[0,1]^2$, which does not intersect the line $y = x$, and let T be a square of side δ with opposite vertices on this line. Let λ and ν be such that $\alpha(S) = \lambda\delta^2$ and $\alpha(T) = \nu\delta^2$. Suppose that $\delta = 1/m$ for some integer m. Since $\alpha([0,1]^2) = 1$, it is necessarily true that

$$m(m-1)\lambda\delta^2 + m\nu\delta^2 = 1.$$

Solving this for ν we get

$$\nu = \lambda + (1-\lambda)/\delta$$

and thus

$$\alpha(T) = (1-\lambda)\delta + \lambda\delta^2. \tag{9.2}$$

It is not difficult to show that the proportionality constant λ is independent of δ and that

$$\alpha(R) = \lambda\mu(R) \tag{9.3}$$

for any rectangle R which does not intersect the diagonal except possibly at a vertex, where μ denotes Lebesgue measure. A can be calculated from (9.2) and (9.3). If $x \le y$,

$$\begin{aligned} A(x,y) &= \alpha([0,x] \times [0,y]) \\ &= \alpha([0,x] \times [0,x]) + \alpha([0,x] \times (x,y]) \\ &= (1-\lambda)x + \lambda x^2 + \lambda x(y-x) \\ &= \lambda xy + (1-\lambda)x. \end{aligned}$$

Similarly, if $y \le x$ then $A(x,y) = \lambda xy + (1-\lambda)y$. Consequently, $A = \lambda P + (1-\lambda)M$. ■

10. Homomorphisms

Let $\mathscr{A}$ and $\mathscr{B}$ be symmetric Markov algebras. A *homomorphism* is a map Ψ: $\mathscr{A} \to \mathscr{B}$ which preserves the algebraic structure; that is, Ψ maps convex combinations into convex combinations, products into products, and, if $\mathscr{A}$ and $\mathscr{B}$ are both symmetric, transposes into transposes. As previously noted, we do not require that Ψ map the null element into the null element or the unit into the unit. Thus, if E is a symmetric idempotent in $\mathscr{B}$, the constant map Ψ which assigns E to each element of $\mathscr{A}$ is a homomorphism, which we shall call a trivial homomorphism. We have seen examples of classes of nontrivial homomorphisms Ψ: $\mathscr{C} \to \mathscr{C}$ in Sections 7 and 8 above. Recall that the nontrivial homomorphisms so constructed were one-to-one maps. We have no example of a nontrivial Markov algebra which is not one-to-one, and we conjecture that there are none.

In this section, we construct a homomorphism from $\mathscr{D}^n \to \mathscr{C}$. Then we prove two theorems concerning nonexistence of homomorphisms.

We define a map Ψ: $\mathscr{D}^n \to \mathscr{C}$ as follows: Let $A \in \mathscr{D}^n$. Divide $[0,1]^2$ into n^2 congruent squares with vertices (x_i, x_j), $i, j = 0, 1, \ldots, n$, where $x_k = k/n$. Let χ_{ij} denote the characteristic function of the square $S_{ij} = [x_{i-1}, x_i) \times [x_{j-1}, x_j)$. Define Ψ via

$$[\Psi(A)](x, y) = n\int_0^x \int_0^y \sum_{i,j=1}^{n} A_{ij}\chi_{ij}(s, t)\, ds\, dt.$$

It is easy to verify that Ψ is a continuous homomorphism of $\mathscr{D}^n$ into $\mathscr{C}$. Observe that Ψ maps P_n to P and M_n to an ordinal sum of copies of P. Note also that if E is any symmetric idempotent copula with square integrable second derivatives, and if the characteristic functions χ_{ij} in the definition of Ψ are replaced by

$$\eta_{ij}(x, y) = \begin{cases} E_{,12}\big(n(x - x_i), n(y - x_j)\big), & \text{if } (x, y) \in S_{ij}, \\ 0, & \text{otherwise} \end{cases}$$

then we again obtain a homomorphism of $\mathscr{D}^n \to \mathscr{C}$, since

$$\int_0^1 \eta_{ij}(u, t)\eta_{kl}(t, v)\, dt = \frac{1}{n}\delta_{jk}\eta_{il}(u, v).$$

This is verified by direct calculation, using the fact that E is idempotent. The condition that E have well behaved second derivatives can be dropped at the cost of introducing more complicated notation, but it is essential that E be a symmetric idempotent. Basically, therefore, to each symmetric idempotent in $\mathscr{C}$ corresponds a continuous homomorphism Ψ: $\mathscr{D}^n \to \mathscr{C}$.

There are, however, no homomorphisms going the other way, or at least no continuous nontrivial ones.

THEOREM 10.1. *Any continuous homomorphism of $\mathscr{C}$ into $\mathscr{D}^n$ is trivial.*

Proof. Suppose Ψ: $\mathscr{C} \to \mathscr{D}^n$ is a continuous homomorphism. Let

$$\{E_\lambda : 0 \le \lambda \le 1\}$$

be the continuous one-parameter family of idempotents of Theorem 8.4. Then, the composite map $\lambda \to E_\lambda \to \Psi(E_\lambda)$ is continuous. Since each $\Psi(E_\lambda)$

is idempotent, since there are only finitely many idempotents in $\mathscr{D}^n$, and since $\mathscr{D}^n$ is topologically connected, then $\Psi(E_\lambda) = \Psi(E_\mu)$ for all λ, μ in $[0, 1]$. In particular, $\Psi(P) = \Psi(M)$. Whereupon, $\Psi(A) = \Psi(AM) = \Psi(A)\Psi(M) = \Psi(A)\Psi(P) = \Psi(AP) = \Psi(P)$ for all A in $\mathscr{C}$. That is, Ψ is trivial. ■

There are likewise no nontrivial homomorphisms $\mathscr{D}^n \to \mathscr{D}^m$ when $n > m$, or at least no surjective ones.

THEOREM 10.2. *For $n > m > 1$ there is no homomorphism of $\mathscr{D}^n$ onto $\mathscr{D}^m$.*

Proof. Suppose $\Psi: \mathscr{D}^n \to \mathscr{D}^m$ is a homomorphism. We assume that Ψ is onto, so that, as is easy to verify, $\Psi(M_n) = M_m$ and $\Psi(P_n) = P_m$. Let Ψ_0 be the restriction of Ψ to the group G_n of permutation matrices in $\mathscr{D}^n$. It is readily checked that a surjective homomorphism maps invertible elements onto invertible elements; it follows that Ψ_0 is a group homomorphism taking values in G_m. Since $m < n$, the kernel N of Ψ_0 is a non-trivial normal subgroup of G_n. Let k be the number of elements in N and let u in $\mathscr{D}^n$ be the convex combination

$$u = \frac{1}{k} \sum_{h \in N} h.$$

Then

$$\Psi(u) = \frac{1}{k} \sum_{h \in N} \Psi_0(h) = M_m. \tag{10.1}$$

Since N is a normal subgroup of G_n, necessarily $gu = ug$ for all g in G_n. It follows that u is in the center of $\mathscr{D}^n$, so that, by Theorem 6.5, $u = \lambda P_n + (1 - \lambda)M_n$ for some $\lambda \in [0, 1]$. It is easily checked that u is idempotent. It is also easily checked that the only convex combinations of P_n and M_n which are idempotent occur when $\lambda = 0$ or 1. But since N is a nontrivial normal subgroup u necessarily has nonzero off diagonal entries, so that necessarily $\lambda \neq 0$. It follows that $\lambda = 1$ and therefore that $u = P_n$. Thus, (10.1) says that $\Psi(P_n) = M_m$. Since $m > 1$ by hypothesis, $M_m \neq P_m$ and we have a contradiction. ■

11. Interpretation

We conclude this paper with some observations on the interpretation of the algebraic concepts of $\mathscr{C}$ in the context of Markov processes.

Invertible and Extreme Copulas. Extreme stochastic matrices correspond to deterministic links in a Markov chain, in the sense that if the transition matrix $Q_{n,n+1}$ is extreme and the state at n is known with probability 1, then the state at $n+1$ is known with probability 1, since all transition probabilities are 0 or 1.

Left invertible copulas also correspond to deterministic links in a Markov process, but the argument is more circumspect. We will say that random variables X_s and X_t in a Markov process are deterministicly related if $t \geq s$ and there is a Borel function f such that $X_t = f(X_s)$ almost surely. This is true for continuous random variables if and only if the copula C_{st} is left invertible. We state the result as a formal theorem:

THEOREM 11.1. *Let X_1 and X_2 be continuous random variables. The following are equivalent*:

(a) *There is a Borel function f such that* $X_2 = f(X_1)$ a.s.
(b) *For all x,* $E(I_{X_2<x}|X_1) = 0$ *or* 1 a.s.
(c) *The copula C_{12} is left invertible.*

Proof. Statements (b) and (c) are equivalent by Theorems 7.1 and 3.1, using the fact that the random variables are continuous. We will complete the proof by showing that statements (a) and (b) are equivalent.

Suppose $E(I_{X_2<x}|X_1) = 0$ or 1 almost certainly for all x. Then for all Borel sets B,

$$\begin{aligned}\int_{X_1^{-1}(B)} E\big(I_{X_2<x}|X_1\big) I_{X_2<x}\, dP &= \int_{X_1^{-1}(B)} E\Big(E\big(I_{X_2<x}|X_1\big) I_{X_2<x}|X_1\Big)\, dP \\ &= \int_{X_1^{-1}(B)} E\big(I_{X_2<x}|X_1\big)^2\, dP \\ &= \int_{X_1^{-1}(B)} E\big(I_{X_2<x}|X_1\big)\, dP \\ &= \int_{X_1^{-1}(B)} I_{X_2<x}\, dP \\ &= \int_{X_1^{-1}(B)} I^2_{X_2<x}\, dP.\end{aligned}$$

It follows that

$$\begin{aligned}0 &= \int_{X_1^{-1}(B)} \Big(I^2_{X_2<x} - 2E\big(I_{X_2<x}|X_1\big) I_{X_2<x} + E\big(I_{X_2<x}|X_1\big)^2\Big)\, dP \\ &= \int_{X_1^{-1}(B)} \big(I_{X_2<x} - E\big(I_{X_2<x}|X_1\big)\big)^2\, dP\end{aligned}$$

which implies that $E(I_{X_2<x}|X_1) = I_{X_2<x}$ almost certainly for all x. But then X_2 is measureable with respect to the completion of the σ algebra of sets $\sigma(X_1)$ generated by X_1, so that statement (a) holds.

If statement (a) holds, then $\sigma(X_2) \subset \sigma(X_1)$ so that $E(I_{X_2<x}|X_1) = I_{X_2<x}$ a.s. This implies (b). ∎

Observe that right invertibility of the copula C_{st} is neither necessary nor sufficient for X_t and X_s to be deterministically related. In fact, we can use a right invertible copula which is not left invertible to construct a Markov process which has a curious property. Suppose that C is right invertible but not left invertible, and let $E = C^T C$. Suppose all copulas in a Markov process C_{uv} are invertible for $u < v \le s$ and $t \le u < v$ where $s < t$. Set

$$\begin{aligned} C_{uv} &= C_{ut} * C, \quad u \le s < v < t \\ C_{uv} &= E, \quad s < u < v < t \\ C_{uv} &= C^T * C_{tv}, \quad s < u < t \le v. \end{aligned}$$

It is easy to verify that the entire process satisfies (3.7) and therefore the Chapman-Kolmogorov equations. Observe that $C_{st} = CC^T = M$, so that X_s and X_t are deterministically related. But if $s < u < v < t$, $C_{uv} = E$ is not left invertible, so in the interval (s, t), the process is not deterministic. The right invertible copula C permits the intrusion of a bubble of randomness in an otherwise deterministic Markov process.

The interpretation of Theorems 7.4 and 7.5 is also interesting; cf. [12], [18], [19]. These results imply that any link in a Markov process can be approximated to arbitrary accuracy by a deterministic link. Thus, if one constructs a discrete Markov process by setting the copula of each adjacent pair of random variables equal to some non-invertible copula C and then approximates C closely enough by an invertible copula A, one can apparently expect the corresponding deterministic process to exhibit behavior similar to that exhibited by the original nondeterministic process; certainly adjacent pairs of random variables in the associated Markov process can be expected to exhibit similar behavior. The joint behavior of distant pairs may not be so similar, because of the fact (Theorem 7.6) that the $*$ product is not jointly continuous.

Idempotents. Observe that a Markov process all of whose marginal distributions F are the same and all of whose copulas C are the same consists of interchangeable random variables if

$$C^2 = C$$

since this condition implies that $C^n = C$ for all n, and the higher order joint distributions are all identical by reason of the Markov condition. The converse is also true: if the random variables in a Markov process are interchangeable, then necessarily $C^2 = C$, and all of the marginal distributions are the same. By DeFinetti's theorem [6] interchangeable random variables are conditionally independent. Thus, we think of idempotents in $\mathscr{C}$ as corresponding to conditionally independent Markov processes.

References

1. A. Sklar, *Fonctions de répartition à n dimensions et leur marges*, Publ. Inst. Statist. Univ. Paris, vol. 8 (1959), pp. 229–231.
2. ______, *Random variables, joint distribution functions and copulas*, Kybernetika, vol. 9 (1973), pp. 449–460.
3. B. Schweizer and A. Sklar, *Operations on distribution functions not derivable from operations on random variables*, Studia Math., vol. 52 (1974), pp. 43–52.
4. ______, *Probabalistic metric spaces*, Elsevier, New York, 1983; especially Chapter 5, Associativity, and Chapter 6, Copulas.
5. R. B. Nelson, *Copulas and associativity*, preprint, 1990.
6. Y. S. Chow and H. Teicher, *Probability theory: Independence interchangeability martingales*, Springer, New York, 1978.
7. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.
8. L. Arnold, *Stochastic differential equations*, Wiley, New York, 1973.
9. H. Sherwood, Oral presentation in Illinois Institute of Technology copula seminar, 1988.
10. V. Losert, *Counterexamples to some conjectures about doubly stochastic measures*, Pacific J. Math., vol. 99 (1982), pp. 387–97.
11. C. Genest and R. MacKay, *Copules archimédiennes et familles de lois bidimensionelles dont les marges sont données*, Canad. J. Statist., vol. 14 (1986), pp. 145–159.
12. P. Mikusinski, H. Sherwood and M. Taylor, *Shuffles of Min*, preprint, 1990.
13. E. W. Hobson, *The theory of functions of a real variable and the theory of Fourier's series*, Vol. 1, Dover, New York, 1957, pp. 662 ff.
14. E. Asplund and L. Bungart, *A first course in integration*, Holt Rinehart, New York, 1966, 345–352.
15. N. L. Johnson and S. Kotz, *Propriétés de dépendence des distributions itérées, généralisées à deux variables Farlie-Gumbel-Morgenstern*, C. R. Acad. Sci. Paris, vol. 285A (1977), pp. 277–281.
16. B. Schweizer and A. Sklar, *Associative functions and abstract semigroups*, Publ. Math. Debrecen, vol. 10 (1963), pp. 69–81.
17. A. B. Paalman-DeMiranda, "Topological semigroups" in *Mathematical Centre Tracts*, No. 11, Mathematisch Centrum, Amsterdam, 1964.
18. R. A. Vitale, "Stochastic dependence and a class of degenerate distributions," in *Topics in statistical dependence*, Block, H., Sampson, A. R., and Savits, T., eds., Institute of Mathematical Statistics Lecture Notes and Monograph Series, to appear.
19. ______, *Approximation by mutually completely dependent processes*, preprint, 1990.
20. E. Hille, *Functional analysis and semi-groups*, American Mathematical Society, New York, 1948, pp. 423ff.