J. Appl. Prob. **43,** 48–59 (2006)
Printed in Israel

POSITIVE DEPENDENCE AND WEAK CONVERGENCE

A. COLANGELO,* *Università dell'Insubria*
A. MÜLLER,** *Universität Karlsruhe*
M. SCARSINI,*** *Università di Torino*

Abstract

A more general definition of MTP_2 (multivariate total positivity of order 2) probability measure is given, without assuming the existence of a density. Under this definition the class of MTP_2 measures is proved to be closed under weak convergence. Characterizations of the MTP_2 property are proved under this more general definition. Then a precise definition of conditionally increasing measure is provided, and closure under weak convergence of the class of conditionally increasing measures is proved. As an application we investigate MTP_2 properties of stationary distributions of Markov chains, which are of interest in actuarial science.

Keywords: Affiliation; MTP_2; conditionally increasing in sequence; conditionally increasing

2000 Mathematics Subject Classification: Primary 60E15
Secondary 60B10

1. Introduction

Starting with the work of Kimeldorf and Sampson (1987), (1989), several concepts and orderings of positive dependence have been studied axiomatically. One of the axioms that is usually considered is closure under weak convergence. Many positive dependence concepts and orders are known to satisfy this axiom. For instance, this is the case for positive quadrant dependence, association, supermodular dependence, RCSI (right corner set increasing), LCSI (left corner set increasing), RTIS (right tail increasing in sequence), and LTDS (left tail increasing in sequence); see Colangelo *et al.* (2005), Kimeldorf and Sampson (1989), and the references therein.

The purpose of this paper is to study the behavior of some strong positive dependence concepts, such as MTP_2 (multivariate total positivity of order 2), CIS (conditionally increasing in sequence), and CI (conditionally increasing). These concepts are not directly checkable unless the measure is concentrated on a finite number of atoms, because otherwise it would be necessary to verify an infinite number of inequalities. Therefore, closure under weak convergence is an important property, as it might help in the verification of MTP_2 or CI by allowing one to find an approximating sequence of discrete measures with finitely many atoms.

Received 13 April 2005; revision received 28 October 2005.
* Postal address: Dipartimento di Economia, Università dell'Insubria, Via Monte Generoso 71, I-21100 Varese, Italy. Email address: antonio.colangelo@uninsubria.it
** Postal address: Institut für Wirtschaftstheorie und Operations Research, Universität Karlsruhe, Geb. 20.21, D-76128 Karlsruhe, Germany. Email address: mueller@wior.uni-karlsruhe.de
*** Postal address: Dipartimento di Statistica e Matematica Applicata, Università di Torino, Piazza Arbarello 8, I-10122 Torino, Italy. Email address: marco.scarsini@unito.it

In order to study the weak convergence property in some generality, we cannot use the usual definition of MTP_2, which requires the existence of a density with respect to some product measure. For instance, the measure associated to the upper Fréchet bound, which is the most positive dependent multivariate distribution (in any possible sense), does not have a density with respect to a product measure if the marginals are continuous. However, it is easy to find a sequence of measures that have MTP_2 densities and converge weakly to the upper Fréchet bound; for instance, a sequence of MTP_2 normal distributions with fixed marginals and correlation coefficients converging to 1.

Thus motivated, in this paper we provide a more general definition of MTP_2 and show that it coincides with the usual one when a density exists with respect to a product measure. We then show that, even in this more general setting, MTP_2 is equivalent to affiliation (see Milgrom and Weber (1982)). For a general treatment of MTP_2 and related concepts, the reader is referred to Karlin and Rinott (1980), Milgrom and Weber (1982), Joe (1997), and Colangelo *et al.* (2005).

When dealing with CIS and CI, a new definition, which does not make use of conditional distributions, will be used. This definition is more formally sound than the usual one, and is proved to be equivalent to a definition that is, in turn, more suitable for dealing with weak convergence.

For properties of CIS and CI, see, e.g. Tukey (1958), Lehmann (1966), Barlow and Proschan (1975), Alam and Wallenius (1976), Joe (1997), and Müller and Scarsini (2001). In the literature one can also find results stated under the assumption of MTP_2 when in fact only CI is required in the proofs. As an example we mention the proof of the Simes conjecture in Sarkar (1998).

The paper is organized as follows. In Section 2 we state axioms of positive dependence notions for multivariate distributions. In Sections 3 and 4 we show that MTP_2 and CI fulfill all these axioms if they are properly defined. In particular, we show that they are closed with respect to weak convergence. In Section 5 these results are applied to a problem that is relevant in actuarial science, and the stationary distribution of a Markov chain is proved to be MTP_2 under appropriate conditions.

2. Axioms of positive dependence notions

Kimeldorf and Sampson (1989) introduced a list of desirable properties for a bivariate notion of positive dependence to have. The following generalization of these axioms to higher dimensions has been described by Pellerey and Semeraro (2003) and Colangelo *et al.* (2005).

In the following, we write $\boldsymbol{X} = (X_1, \dots, X_d) \sim \mu$ if the d-variate random vector $\boldsymbol{X}$ has distribution μ. For any k-tuple $I = (i_1, \dots, i_k) \subset \{1, \dots, d\}$ we denote by μ_I the distribution of $\boldsymbol{X}_I = (X_{i_1}, \dots, X_{i_k})$. Let Δ_d be the set of all d-variate distributions, and let '$\xrightarrow{\mathrm{w}}$' denote weak convergence of measures. Throughout the paper the terms 'increasing' and 'decreasing' are used in the weak sense.

It is evident that a notion of positive dependence is uniquely determined by a subset $\mathcal{P}_d^+ \subset \Delta_d$. One of the weakest reasonable concepts of positive dependence is *positive quadrant dependence*, introduced by Lehmann (1966). The distribution μ of a bivariate random vector $\boldsymbol{X}$ is said to be positive quadrant dependent if $\mathrm{P}(X_1 \le x_1,\ X_2 \le x_2) \ge \mathrm{P}(X_1 \le x_1)\,\mathrm{P}(X_2 \le x_2)$ for any $x_1, x_2 \in \mathbb{R}$. We denote by $\mathcal{P}_{2,\mathrm{PQD}}^+$ the set of all bivariate distributions that are positive quadrant dependent. By $\mathcal{F}_d^+ \subset \Delta_d$ we denote the subclass of all upper Fréchet bounds, i.e. the set of distributions of random vectors satisfying

$$\mathrm{P}(X_1 \le x_1, \dots, X_d \le x_d) = \min_{i=1,\dots,d} \mathrm{P}(X_i \le x_i) \quad \text{for all } x_1, \dots, x_d, \tag{1}$$

and by $\mathcal{J}_d$ the set of all distributions with independent marginals.

https://doi.org/10.1239/jap/1143936242 Published online by Cambridge University Press

What follows is a list of desirable properties that any multivariate positive dependence notion corresponding to the set $\mathcal{P}_d^+ \subset \Delta_d$ should have. It is borrowed from Colangelo *et al.* (2005).

B1. If $\mu \in \mathcal{P}_d^+$ then $\mu_{(i,j)} \in \mathcal{P}_{2,\mathrm{PQD}}^+$ for all $i, j \in \{1, \dots, d\}$ with $i < j$.

B2. $\mathcal{F}_d^+ \subseteq \mathcal{P}_d^+$.

B3. $\mathcal{J}_d \subseteq \mathcal{P}_d^+$.

B4. If $(X_1, \dots, X_d) \sim \mu \in \mathcal{P}_d^+$ then $(\phi_1(X_1), \dots, \phi_d(X_d)) \sim \nu \in \mathcal{P}_d^+$ for all increasing functions $\phi_1, \dots, \phi_d \colon \mathbb{R} \to \mathbb{R}$.

B5. If $(X_1, \dots, X_d) \sim \mu \in \mathcal{P}_d^+$ then $(X_{i_1}, \dots, X_{i_d}) \sim \nu \in \mathcal{P}_d^+$ for all permutations $(i_1, \dots, i_d)$ of $(1, \dots, d)$.

B6. If $\{\mu_n, n \geq 1\} \subseteq \mathcal{P}_d^+$ is such that $\mu_n \xrightarrow{\mathrm{w}} \mu$, then $\mu \in \mathcal{P}_d^+$.

B7. If $(X_1, \dots, X_d) \sim \mu \in \mathcal{P}_d^+$ then $\boldsymbol{X}_I \sim \nu \in \mathcal{P}_{\mathrm{card}(I)}^+$ for all $I \subseteq \{1, \dots, d\}$.

Many known concepts of dependence, like PUOD (positive upper orthant dependence), PLOD (positive lower orthant dependence), positive supermodular dependence, and positive association, have the properties B1–B7 (see Colangelo *et al.* (2005) for details and references). In the next sections we will show that this also holds for MTP_2 and CI if they are properly defined.

3. MTP_2 measures

The usual definition of MTP_2 probability measure on a product space assumes the existence of a density with respect to a product measure; see, e.g. Karlin and Rinott (1980), Milgrom and Weber (1982), and Müller and Stoyan (2002). However, when studying positive dependence it is fundamental to consider situations in which a probability measure is not dominated by a product measure. This holds especially for the important special case of the upper Fréchet bound. We will therefore give a general definition of MTP_2 measure that does not require the existence of a density, and will prove some characterization results in this greater generality. For the bivariate case, some of the ideas used below can be traced back to Block *et al.* (1982).

Let us first recall the concept of MTP_2 as considered in the seminal paper of Karlin and Rinott (1980). A partially ordered set $(L, \leq)$ is called a *lattice* if for every $\boldsymbol{x}, \boldsymbol{y} \in L$ we have $\boldsymbol{x} \vee \boldsymbol{y}, \boldsymbol{x} \wedge \boldsymbol{y} \in L$, where $\boldsymbol{x} \vee \boldsymbol{y}$ is the unique smallest element of the set $\{\boldsymbol{z} \colon \boldsymbol{x} \leq \boldsymbol{z},\ \boldsymbol{y} \leq \boldsymbol{z}\}$ and $\boldsymbol{x} \wedge \boldsymbol{y}$ is the unique largest element of the set $\{\boldsymbol{z} \colon \boldsymbol{z} \leq \boldsymbol{x},\ \boldsymbol{z} \leq \boldsymbol{y}\}$. For subsets $A, B \subset L$ we write

$$A \vee B = \{\boldsymbol{z} \colon \boldsymbol{z} = \boldsymbol{x} \vee \boldsymbol{y},\ \boldsymbol{x} \in A,\ \boldsymbol{y} \in B\},$$
$$A \wedge B = \{\boldsymbol{z} \colon \boldsymbol{z} = \boldsymbol{x} \wedge \boldsymbol{y},\ \boldsymbol{x} \in A,\ \boldsymbol{y} \in B\}.$$

Definition 1. Let L be a lattice. A nonnegative function $g \colon L \to \mathbb{R}$ is MTP_2 if

$$g(\boldsymbol{x})g(\boldsymbol{y}) \leq g(\boldsymbol{x} \vee \boldsymbol{y})g(\boldsymbol{x} \wedge \boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in L$.

A probability measure μ on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ is usually called MTP_2 if it has an MTP_2 density with respect to a dominating product measure. Notice that the assumption of the dominating measure being a product measure is crucial, as for every probability measure one can find a dominating measure such that the corresponding density is MTP_2, since the density of a probability measure with respect to itself is identically 1 and, therefore, MTP_2. On the other hand, it is easy to see that if there is an MTP_2 density with respect to some dominating product measure, then there is also an MTP_2 density with respect to any other dominating product measure. Therefore, the dominating product measure can without loss of generality be chosen as, for instance, the product of the marginals.

The upper Fréchet bound, however, does not have a density with respect to any product measure if the marginals are continuous. As any reasonable concept of dependence should include the upper Fréchet bound, there is an interest in finding a density-free definition of MTP_2 that satisfies properties B2 and B6. The following definition serves this purpose.

Definition 2. A probability measure μ on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ is MTP_2 if

$$\mu(A)\mu(B) \leq \mu(A \vee B)\mu(A \wedge B) \tag{2}$$

for all $A, B \in \mathrm{Bor}(\mathbb{R}^d)$.

Whenever μ admits a density f with respect to some product measure, Theorem 3.10.14 of Müller and Stoyan (2002) implies that the probability measure μ is MTP_2 according to Definition 2 if and only if its density f is MTP_2 according to Definition 1. We should point out that Definition 2 is not directly checkable unless the measure μ is concentrated on a finite number of atoms, because otherwise it would be necessary to verify an inequality over a continuum of sets A and B. Theorem 1, below, shows that the MTP_2 property of μ can be established by finding a sequence of MTP_2 measures converging weakly to μ.

Milgrom and Weber (1982) introduced the strongly related concept of *affiliation*, which also does not rely on the existence of densities. To define it we must recall the notion of an *upper set*. A set $U \subset \mathbb{R}^d$ is called *upper* if $\boldsymbol{x} \in U$ and $\boldsymbol{y} \geq \boldsymbol{x}$ imply that $\boldsymbol{y} \in U$. We call $\mathcal{U}_d$ the class of upper sets in $\mathrm{Bor}(\mathbb{R}^d)$.

Definition 3. A probability measure μ on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ is *affiliated* if

$$\mu(A \cap B \mid L) \geq \mu(A \mid L)\mu(B \mid L)$$

for all sets $A, B \in \mathcal{U}_d$ and for all sublattices $L \in \mathrm{Bor}(\mathbb{R}^d)$ such that $\mu(L) > 0$, where, for any such set L, the conditional measure $\mu(\cdot \mid L)$ is defined as

$$\mu(A \mid L) := \frac{\mu(A \cap L)}{\mu(L)}, \qquad A \in \mathrm{Bor}(\mathbb{R}^d).$$

In order to state the following theorem we need some preliminary definitions. Given two sets $C, D \subset \mathbb{R}^d$, we say that $C < D$ if, for all $\boldsymbol{c} \in C$ and $\boldsymbol{d} \in D$, we have $\boldsymbol{c} < \boldsymbol{d}$ (i.e. $c_i < d_i$ for $i = 1, \dots, d$). The sets C and D are *comparable* if either $C < D$ or $D < C$. For $i = 1, \dots, d$, let A_i and B_i be intervals in $\mathbb{R}$. The intervals $A = \times_{i=1}^d A_i$ and $B = \times_{i=1}^d B_i$ in $\mathbb{R}^d$ are called *strongly disjoint* if A_i and B_i are disjoint for all $i \in \{1, \dots, d\}$. We call $\mathcal{I}_d$ the class of (possibly unbounded) intervals in $\mathbb{R}^d$ of type

$$A = \mathop{\times}_{i=1}^{d} (a_i, c_i].$$

Notice that $\mathcal{I}_d$ is a product lattice.

https://doi.org/10.1239/jap/1143936242 Published online by Cambridge University Press

The following theorem provides the required characterizations of the MTP_2 property.

Theorem 1. *Let μ be a probability measure on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$. The following characterizations of μ are equivalent:*

(a) *μ is MTP_2;*

(b) *μ satisfies (2) for all strongly disjoint noncomparable (possibly unbounded) intervals $A, B \in \mathcal{I}_d$;*

(c) *μ satisfies (2) for all strongly disjoint noncomparable closed intervals A and B;*

(d) *μ is affiliated;*

(e) *there exists a sequence of probability measures μ_n such that $\mu_n \xrightarrow{\mathrm{w}} \mu$ and, for each $n \in \mathbb{N}$, the measure μ_n has an MTP_2 density with respect to a product measure.*

Proof. (b) $\Rightarrow$ (a): Let $\mathcal{A}$ be the smallest algebra containing $\mathcal{I}_d$ and let $A, B \in \mathcal{A}$. Then

$$A = \bigcup_{k=1}^{n_A} \bigtimes_{i=1}^{d} (a_i^{(k)}, c_i^{(k)}], \qquad B = \bigcup_{k=1}^{n_B} \bigtimes_{i=1}^{d} (b_i^{(k)}, d_i^{(k)}]$$

for some $n_A, n_B \in \mathbb{N}$. For $i \in \{1, \dots, d\}$, let

$$\begin{aligned} \mathcal{C}_i &= \{a_i^{(k_A)}, c_i^{(k_A)}, b_i^{(k_B)}, d_i^{(k_B)} \colon 1 \le k_A \le n_A, 1 \le k_B \le n_B\} \\ &=: \{\alpha_i^{(j)} \colon \alpha_i^{(1)} \le \alpha_i^{(2)} \le \cdots \le \alpha_i^{(m_i)}\}. \end{aligned}$$

Clearly $m_i \le 2n_A + 2n_B$. If we let $R_i^j := (\alpha_i^{(j)}, \alpha_i^{(j+1)}]$ then we can write

$$A = \bigsqcup_{\boldsymbol{j} \in \tilde{A}} \bigtimes_{i=1}^{d} R_i^{j_i} =: \bigsqcup_{\boldsymbol{j} \in \tilde{A}} R^{\boldsymbol{j}}, \qquad B = \bigsqcup_{\boldsymbol{j} \in \tilde{B}} \bigtimes_{i=1}^{d} R_i^{j_i} =: \bigsqcup_{\boldsymbol{j} \in \tilde{B}} R^{\boldsymbol{j}},$$

where $\tilde{A} = \{\boldsymbol{j} \colon R^{\boldsymbol{j}} \subset A\}$, $\tilde{B} = \{\boldsymbol{j} \colon R^{\boldsymbol{j}} \subset B\}$, and '$\bigsqcup$' indicates disjoint union.

On the space $\Omega' = \bigtimes_{i=1}^{d} \{1, 2, \dots, m_i - 1\}$, define the discrete probability measure Q as having the density

$$f_Q(j_1, j_2, \dots, j_n) = \frac{1}{\beta} \mu(R^{\boldsymbol{j}}),$$

where

$$\beta = \sum_{\boldsymbol{j} \in \Omega'} \mu(R^{\boldsymbol{j}}).$$

It is not difficult to see that f_Q is an MTP_2 function; in fact,

$$\begin{aligned} f_Q(\boldsymbol{i} \wedge \boldsymbol{j}) f_Q(\boldsymbol{i} \vee \boldsymbol{j}) &= \frac{1}{\beta^2} \mu(R^{\boldsymbol{i} \wedge \boldsymbol{j}}) \mu(R^{\boldsymbol{i} \vee \boldsymbol{j}}) \\ &= \frac{1}{\beta^2} \mu(R^{\boldsymbol{i}} \wedge R^{\boldsymbol{j}}) \mu(R^{\boldsymbol{i}} \vee R^{\boldsymbol{j}}) \\ &\ge \frac{1}{\beta^2} \mu(R^{\boldsymbol{i}}) \mu(R^{\boldsymbol{j}}) \\ &= f_Q(\boldsymbol{i}) f_Q(\boldsymbol{j}). \end{aligned}$$

By Karlin and Rinott (1980, Corollary 2.1), it follows that

$$Q(\tilde{A})Q(\tilde{B}) \le Q(\tilde{A} \vee \tilde{B})Q(\tilde{A} \wedge \tilde{B}).$$

Therefore,

$$\mu(A)\mu(B) = \beta^2 Q(\tilde{A})Q(\tilde{B}) \le \beta^2 Q(\tilde{A} \vee \tilde{B})Q(\tilde{A} \wedge \tilde{B}) = \mu(A \vee B)\mu(A \wedge B).$$

Let

$$\mathcal{G}_1 = \{A \in \mathrm{Bor}(\mathbb{R}^d) \colon \mu(A)\mu(B) \le \mu(A \vee B)\mu(A \wedge B),\ B \in \mathcal{A}\}.$$

Since $\mathcal{A} \subset \mathcal{G}_1$ and $\mathcal{G}_1$ is a monotone class, by the monotone class lemma (Shiryaev (1996, Theorem 1, p. 141)) we have $\mathcal{G}_1 = \mathrm{Bor}(\mathbb{R}^d)$.

Now define

$$\mathcal{G}_2 = \{B \in \mathrm{Bor}(\mathbb{R}^d) \colon \mu(A)\mu(B) \le \mu(A \vee B)\mu(A \wedge B),\ A \in \mathcal{G}_1\}.$$

The same argument proves that $\mathcal{G}_2 = \mathrm{Bor}(\mathbb{R}^d)$.

(a) $\Rightarrow$ (c): This implication is trivial.

(c) $\Rightarrow$ (b): Any interval in $\mathcal{I}_d$ can be obtained as the limit of an increasing sequence of closed intervals. Intervals in the different sequences are disjoint and noncomparable.

We have thus proved the equivalence of (a), (b), and (c). This equivalence will be used in the proof of Theorem 2, below.

(a) $\Rightarrow$ (d): This implication has been shown in Müller and Stoyan (2002, Theorem 3.10.14).

(d) $\Rightarrow$ (b): Let A and B be strongly disjoint noncomparable intervals with positive μ-measure. Given a set C, denote by C' the smallest upper superset of C.

The set L defined as

$$L = A \cup B \cup (A \vee B) \cup (A \wedge B)$$

is a lattice. Observe that

$$\begin{aligned} A' \cap L &= A \cup (A \vee B), \\ B' \cap L &= B \cup (A \vee B), \\ A' \cap B' \cap L &= A \vee B. \end{aligned}$$

Since μ is affiliated, we have

$$\mu(A' \cap B' \cap L)\mu(L) \ge \mu(A' \cap L)\mu(B' \cap L),$$

which is equivalent to

$$\mu(A \vee B)\mu(L) \ge \mu(A \cup (A \vee B))\mu(B \cup (A \vee B)), \tag{3}$$

which in turn is equivalent to

$$\mu(A \wedge B)\mu(A \vee B) \ge \mu(A)\mu(B). \tag{4}$$

To see this, notice that the sets A, B, $A \vee B$, and $A \wedge B$ are disjoint, and, using the abbreviations

$$a = \mu(A), \qquad b = \mu(B), \qquad c = \mu(A \vee B), \qquad d = \mu(A \wedge B),$$

that (3) can be rewritten as $c(a+b+c+d) \ge (a+c)(b+d)$, which is equivalent to $cd \ge ab$. This, however, is just (4).

Hence, we have proved the equivalence of (a)–(d).

(e) $\Rightarrow$ (c): This implication follows from Theorem 2, below.

(c) $\Rightarrow$ (e): For a fixed $n \in \mathbb{N}$, consider the lattice

$$L_n = \left\{ \left(\frac{k_1}{2^n}, \dots, \frac{k_d}{2^n} \right),\ (k_1, \dots, k_d) \in \mathbb{Z}^d \right\},$$

which in fact is a product lattice. Partition $\mathbb{R}^d$ into intervals of type

$$\mathop{\times}_{i=1}^{d} \left[\frac{k_i}{2^n}, \frac{k_i + 1}{2^n} \right), \qquad (k_1, \dots, k_d) \in \mathbb{Z}^d.$$

The left-hand endpoint of each of these intervals is in the lattice L_n. Consider a sequence of probability measures μ_n that discretize μ by concentrating the μ-mass of each of the above intervals on its left-hand endpoint. It is clear that every μ_n has an MTP_2 density (with respect to the counting measure on the lattice L_n, which obviously is a product measure), and that $\mu_n \xrightarrow{\text{w}} \mu$.

This completes the proof.

Theorem 2. *If $\{\mu_n\}_{n \in \mathbb{N}}$ is a sequence of MTP_2 probability measures, and if $\mu_n \xrightarrow{\text{w}} \mu$, then μ is MTP_2.*

Proof. By the equivalence between parts (a) and (c) of Theorem 1, we know that μ is MTP_2 if and only if (2) holds for all closed, strongly disjoint noncomparable intervals A and B.

There exist two decreasing sequences of strongly disjoint noncomparable open intervals, $\{A_m\}$ and $\{B_m\}$, such that $A_m \searrow A$ and $B_m \searrow B$. Hence, $(A_m \wedge B_m) \searrow (A \wedge B)$ and $(A_m \vee B_m) \searrow (A \vee B)$. It is always possible to choose such sequences in such a way that the μ-mass of the boundaries of all the involved sets is 0.

Since, for every $n \in \mathbb{N}$, μ_n is MTP_2, for every $m \in \mathbb{N}$ we have

$$\mu_n(A_m)\mu_n(B_m) \le \mu_n(A_m \vee B_m)\mu_n(A_m \wedge B_m).$$

Since, for every m, the sets A_m, B_m, $(A_m \vee B_m)$, and $(A_m \wedge B_m)$ are μ-continuity sets, by the portmanteau theorem (see Billingsley (1999, Theorem 2.1, p. 16)) we have, for every m,

$$\mu(A_m)\mu(B_m) \le \mu(A_m \vee B_m)\mu(A_m \wedge B_m).$$

By letting $m \to \infty$ we obtain the result.

The idea of the proof of Theorem 2 has been used before to show the closure under weak convergence of related stochastic orderings (see Lemma 4.6 of Kimeldorf and Sampson (1987) and Theorem 5.8 of Müller (1997)).

Part (e) of Theorem 1 and Theorem 2 together imply that the set $\mathcal{P}^+_{d,\text{MTP}_2}$ of all probability measures fulfilling the MTP_2 definition (2) is just the weak closure of the set of probability measures having an MTP_2 density. As a consequence we have the following result.

Theorem 3. *The set $\mathcal{P}^+_{d,\text{MTP}_2}$ has the properties B1–B7.*

Proof. For property B6, the proof follows from Theorem 2. For B2, see Theorem 3.10.15 of Müller and Stoyan (2002). The remaining properties are easy to verify.

4. Conditionally increasing measures

In this section we provide a precise definition of, and characterize, conditionally increasing measures. As a by-product we obtain closure under weak convergence of the class of conditionally increasing measures.

Usually, a random vector $(X_1, \ldots, X_d)$ is said to be *conditionally increasing in sequence* (CIS) if

$$\mathrm{P}(X_k > t \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) \tag{5}$$

is an increasing function of $x_1, \ldots, x_{k-1}$ for all $k = 2, \ldots, d$, and is said to be *conditionally increasing* (CI) if any permutation of the components is CIS.

In order to state a more general definition, not relying on conditional distributions, we require the concept of a cylinder. Given a set $A \in \mathrm{Bor}(\mathbb{R}^k)$, with $k < d$, we call $\hat{A}$ the cylinder $A \times \mathbb{R}^{d-k}$.

Definition 4. A probability measure μ on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ is *conditionally increasing in sequence (CIS)* if, for all $k \in \{1, \ldots, d-1\}$, for all $A, B \in \mathrm{Bor}(\mathbb{R}^k)$ such that $A < B$, and for all sets $U \in \mathcal{U}_d$, we have

$$\mu(U \cap \hat{A})\mu(\hat{B}) \leq \mu(U \cap \hat{B})\mu(\hat{A}). \tag{6}$$

The measure μ is *conditionally increasing (CI)* if $\mu\pi^{-1}$ is CIS for every permutation function π.

As for Definition 2, we should point out a problem of checkability of Definition 4. Theorem 5, below, shows how to establish the CIS property of a measure μ via a sequence of measures that converges weakly to μ and is such that each member of the sequence has the CIS property.

Notice that, for the sets in (6), we have

$$(U \cap \hat{A}) \vee \hat{B} \subseteq U \cap \hat{B},$$
$$(U \cap \hat{A}) \wedge \hat{B} \subseteq \hat{A};$$

therefore, an MTP_2 measure is always CI.

The following definitions will be needed for the characterization result. Given a set $A \in \mathrm{Bor}(\mathbb{R}^d)$ and an $\boldsymbol{s} \in \mathbb{R}^k$, we define the section of A as

$$A_{\boldsymbol{s}} = \{\boldsymbol{x} \in \mathbb{R}^{d-k} : (\boldsymbol{s}, \boldsymbol{x}) \in A\}. \tag{7}$$

Given a probability measure μ on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$, we define $\mu^{(k)}$, its marginal distribution on the first k components, by

$$\mu^{(k)}(A) = \mu(A \times \mathbb{R}^{d-k}).$$

Definition 5. A *stochastic kernel* on $\mathbb{R}^k \times \mathbb{R}^m$ is a function $K : \mathbb{R}^k \times \mathrm{Bor}(\mathbb{R}^m) \to [0, 1]$ such that, for all $A \in \mathrm{Bor}(\mathbb{R}^m)$, the function $K(\cdot, A)$ is $\mathrm{Bor}(\mathbb{R}^k)$-measurable and, for all $\boldsymbol{x} \in \mathbb{R}^k$, the function $K(\boldsymbol{x}, \cdot)$ is a probability measure on $\mathrm{Bor}(\mathbb{R}^m)$. We define the measure $\mu^{(k)} * K$ on $(\mathbb{R}^{k+m}, \mathrm{Bor}(\mathbb{R}^{k+m}))$ by

$$\mu^{(k)} * K(A \times B) = \int_A K(t, B)\mu^{(k)}(\mathrm{d}t).$$

A kernel K is said to be *stochastically increasing* if

$$K(\boldsymbol{s}, U) \leq K(\boldsymbol{t}, U)$$

for all $U \in \mathcal{U}_m$ and for all $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^k$ such that $\boldsymbol{s} \leq \boldsymbol{t}$.

Notice that if K is a stochastically increasing kernel, $D_1, D_2 \in \mathcal{U}_m$ are such that $D_1 \subseteq D_2$, and $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^k$ are such that $\boldsymbol{s} \le \boldsymbol{t}$, then

$$K(\boldsymbol{s}, D_1) \le K(\boldsymbol{s}, D_2) \le K(\boldsymbol{t}, D_2). \tag{8}$$

Theorem 4. *Let μ be a probability measure on $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$. The following characterizations of μ are equivalent:*

(a) *μ is CIS;*

(b) *for every $k \in \{1, \dots, d-1\}$, there exists a stochastically increasing kernel $K^{k,1}$ on $\mathbb{R}^k \times \mathbb{R}$ such that $\mu^{(k+1)} = \mu^{(k)} * K^{k,1}$;*

(c) *for all $k \in \{1, \dots, d-1\}$, for all intervals $A, B \in \mathcal{I}_k$ such that $A < B$, and for all sets $V \in \mathcal{U}_1$, we have*

$$\mu^{(k+1)}(A \times V)\mu^{(k)}(B) \le \mu^{(k+1)}(B \times V)\mu^{(k)}(A).$$

Proof. (a) $\Rightarrow$ (c): Let $U = \mathbb{R}^k \times V \times \mathbb{R}^{d-k-1}$ in (6). Then (a) clearly implies (c).

(c) $\Rightarrow$ (b): From (c) it follows that, for every $k \in \{1, \dots, d-1\}$, there exists a kernel $K^{k,1}$ such that, for all $A, B \in \mathrm{Bor}(\mathbb{R}^k)$ with $A < B$ and for all sets $V \in \mathcal{U}_1$, we have

$$\int_A K^{k,1}(\boldsymbol{s}, V)\mu^{(k)}(\mathrm{d}\boldsymbol{s}) \int_B \mu^{(k)}(\mathrm{d}\boldsymbol{t}) \le \int_B K^{k,1}(\boldsymbol{t}, V)\mu^{(k)}(\mathrm{d}\boldsymbol{t}) \int_A \mu^{(k)}(\mathrm{d}\boldsymbol{s}).$$

By Tonelli's theorem, this is equivalent to

$$\int_{A\times B} K^{k,1}(\boldsymbol{s}, V)\mu^{(k)} \otimes \mu^{(k)}(\mathrm{d}\boldsymbol{s} \times \mathrm{d}\boldsymbol{t}) \le \int_{A\times B} K^{k,1}(\boldsymbol{t}, V)\mu^{(k)} \otimes \mu^{(k)}(\mathrm{d}\boldsymbol{s} \times \mathrm{d}\boldsymbol{t}). \tag{9}$$

Consider the set $\mathcal{X} := \{(\boldsymbol{s}, \boldsymbol{t})\colon \boldsymbol{s} < \boldsymbol{t},\ \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^k\}$. Denote by $\mathcal{C}$ the class of sets $A \times B \in \mathrm{Bor}(\mathbb{R}^{2k})$ such that $A, B \in \mathcal{I}_k$ and $A < B$. The class $\mathcal{C}$ is a semi-ring of subsets of $\mathcal{X}$, and generates its Borel σ-field. Furthermore, $\mathcal{X}$ is a countable union of sets in $\mathcal{C}$. By Billingsley (1995, Corollary 2, p. 169), (9) holds for all Borel sets in $\mathcal{X}$. Billingsley (1995, Theorem 16.10, p. 213) implies that, for all $\boldsymbol{s} < \boldsymbol{t}$,

$$K^{k,1}(\boldsymbol{s}, V) \le K^{k,1}(\boldsymbol{t}, V) \quad \mu^{(k)}\text{-almost surely.}$$

(b) $\Rightarrow$ (a): For sets $U \in \mathcal{U}_d$ and $A, B \in \mathrm{Bor}(\mathbb{R}^k)$, such that $A < B$, the following holds, where $U_{\boldsymbol{s}}$ is the section of U, as defined in (7):

$$\begin{aligned}
\mu(U \cap \hat{A})\mu(\hat{B}) &= \int_A K^{k,1} * K^{k+1,1} * \cdots * K^{d-1,1}(\boldsymbol{s}, U_{\boldsymbol{s}})\mu^{(k)}(\mathrm{d}\boldsymbol{s}) \int_B \mu^{(k)}(\mathrm{d}\boldsymbol{t}) \\
&= \int_{A\times B} K^{k,1} * K^{k+1,1} * \cdots * K^{d-1,1}(\boldsymbol{s}, U_{\boldsymbol{s}})\mu^{(k)} \otimes \mu^{(k)}(\mathrm{d}\boldsymbol{s} \times \mathrm{d}\boldsymbol{t}) \\
&\le \int_{A\times B} K^{k,1} * K^{k+1,1} * \cdots * K^{d-1,1}(\boldsymbol{t}, U_{\boldsymbol{t}})\mu^{(k)} \otimes \mu^{(k)}(\mathrm{d}\boldsymbol{s} \times \mathrm{d}\boldsymbol{t}) \\
&= \int_B K^{k,1} * K^{k+1,1} * \cdots * K^{d-1,1}(\boldsymbol{t}, U_{\boldsymbol{t}})\mu^{(k)}(\mathrm{d}\boldsymbol{t}) \int_A \mu^{(k)}(\mathrm{d}\boldsymbol{s}) \\
&= \mu(U \cap \hat{B})\mu(\hat{A}).
\end{aligned}$$

The inequality follows from Kamae *et al.* (1977, Proposition 1) and from (8).

https://doi.org/10.1239/jap/1143936242 Published online by Cambridge University Press

Notice that the equivalence of parts (a) and (b) of Theorem 4 shows that our definition of CIS, given in (6), is equivalent to the traditional one, mentioned in (5).

We are now ready to state a convergence theorem for sequences of CIS probability measures.

Theorem 5. *If $\{\mu_n\}_{n\in\mathbb{N}}$ is a sequence of CIS probability measures, and if $\mu_n \xrightarrow{\mathrm{w}} \mu$, then μ is CIS.*

Proof. First we notice that in part (c) of Theorem 4 we can replace the intervals $A, B \in \mathcal{I}_d$ and sets $V \in \mathcal{U}_1$ with closed intervals. There exist three decreasing sequences of open intervals, $\{A_m\}$ and $\{B_m\}$ in $\mathbb{R}^k$ and $\{V_m\}$ in $\mathbb{R}$, such that, for all m, $A_m < B_m$ and $A_m \searrow A$, $B_m \searrow B$, and $V_m \searrow V$. Hence, $A_m \times V_m \searrow A \times V$ and $B_m \times V_m \searrow B \times V$.

For all m we can choose the sets A_m, B_m, and V_m in such a way that the probability mass of their boundaries is 0. Since μ_n is CIS for every $n \in \mathbb{N}$, for every $m \in \mathbb{N}$ we have

$$\mu_n^{(k+1)}(A_m \times V_m)\mu_n^{(k)}(B_m) \le \mu_n^{(k+1)}(B_m \times V_m)\mu_n^{(k)}(A_m).$$

Since $A_m \times V_m$ and $B_m \times V_m$ are $\mu^{(k+1)}$-continuity sets, the portmanteau theorem (see Billingsley (1999)) implies that

$$\mu^{(k+1)}(A_m \times V_m)\mu^{(k)}(B_m) \le \mu^{(k+1)}(B_m \times V_m)\mu^{(k)}(A_m).$$

By letting $m \to \infty$ we obtain the result.

Corollary 1. *If $\{\mu_n\}_{n\in\mathbb{N}}$ is a sequence of CI probability measures, and if $\mu_n \xrightarrow{\mathrm{w}} \mu$, then μ is CI.*

Proof. It is enough to recall that any permutation function is continuous, and to apply Theorem 5.

In order to verify that CI has property B2, note that the support of any measure μ that satisfies (1) is a chain (a linearly ordered subset of $\mathbb{R}^d$). Therefore, for all sets $U \in \mathcal{U}_d$ and $A, B \in \mathrm{Bor}(\mathbb{R}^k)$, such that $A < B$, if $\mu(U \cap \hat{A}) > 0$ then $\mu(\hat{B}) = \mu(U \cap \hat{B})$. Thus, for such measures μ, (6) holds.

As it is easy to check that the other properties, B1 and B3–B7, obtain, we thus have the following result.

Theorem 6. *The set $\mathcal{P}^+_{d,\mathrm{CI}}$ of all CI probability measures has the properties B1–B7.*

5. An application to Markov chains

Assume that the sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ is a homogenous Markov chain, with state space $\mathbb{R}^d$, characterized by its initial distribution $\pi_0(A) = \mathrm{P}(\boldsymbol{X}_0 \in A)$ and its transition probability measure $Q(\boldsymbol{x}, \mathrm{d}\boldsymbol{y})$, given by $Q(\boldsymbol{x}, A) = \mathrm{P}(\boldsymbol{X}_{n+1} \in A \mid \boldsymbol{X}_n = \boldsymbol{x})$. We further assume that there is a product measure μ on $\mathbb{R}^d$ such that π_0 and $Q(\boldsymbol{x}, \cdot)$, $\boldsymbol{x} \in \mathbb{R}^d$, have densities f_0 and $q(\boldsymbol{x}, \cdot)$, $\boldsymbol{x} \in \mathbb{R}^d$, with respect to μ. The following result then holds.

Theorem 7. *If $f_0 \colon \mathbb{R}^d \to [0, \infty)$ and $q : \mathbb{R}^{2d} \to [0, \infty)$ are MTP$_2$ functions, then*

(a) *the joint distribution of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is MTP$_2$ for any n;*

(b) *the marginal distribution π_n of $\boldsymbol{X}_n$ is MTP$_2$;*

(c) *if, moreover, π_n converges to a stationary distribution π, then π is MTP$_2$.*

Proof. Parts (a) and (b) follow from basic properties of MTP_2 functions; see Propositions 3.2–3.4 of Karlin and Rinott (1980). Part (c) then is a direct consequence of Theorem 2.

The MTP_2 properties concerning the finite-time behavior of Markov chains as described in part (a) of Theorem 7 are well known, especially in the case of finite state space; see, e.g. Kijima (1997). The proof of the MTP_2 property of a stationary distribution, however, is new, and requires our generalized definition of MTP_2, which does not require the existence of a density with respect to a product measure.

In fact, this application of the weak convergence property of MTP_2 inspired us to consider the topic. This question was posed to us by Michel Denuit (personal communication), who was interested in an actuarial application to bonus–malus systems. We give a short description of the context. In the automobile insurance business one uses experience rating to find a fair premium for each individual policy holder. This is done using a so-called bonus–malus system. If a policy holder has no claims in the previous year he goes down in the bonus–malus scale, and if he has one or more accidents he goes up in it, i.e. the level L_{t+1} in year $t+1$ is a function of the level L_t in year t and of the number of claims N_t in year t. It is usually assumed that the numbers, $N_1, N_2, \dots$, of claims are independent, identically distributed random variables having a Poisson distribution with parameter θ, which a priori is unknown. The insurance company only knows the distribution of the random variable Θ from which the parameter θ is drawn. It is easy to see that the bivariate process (Θ, L_t), $t \in \mathbb{N}_0$, is a homogenous Markov chain. It is natural to assume that Θ and L_0 are independent (typically L_0 will be constant) and, thus, MTP_2. It follows from Theorem 7 that the stationary distribution of the bivariate Markov chain is MTP_2 if the transition density has this property. As Θ is constant, the assumption of an MTP_2 transition density reduces to the assumption that the function

$$(\theta, i, j) \mapsto \mathrm{P}(L_{t+1} = j \mid L_t = i,\ \Theta = \theta) \quad \text{is } \mathrm{MTP}_2. \tag{10}$$

Notice that in the simplest case, with $L_{t+1} = N_t$, this assumption holds because (Θ, N_t) is MTP_2. It also holds if L_t is a moving average of the sequence (N_t), i.e. if $L_{t+1} = \alpha L_t + (1-\alpha)N_t$ for some α, $0 < \alpha < 1$. In this case, Proposition 3.7 of Karlin and Rinott (1980) is applicable.

If (10) holds, and if (Θ, L) is a pair of random variables having the stationary distribution of the corresponding Markov chain, then $\mathrm{E}[\Theta \mid L = \ell]$ is an increasing function of ℓ. This follows from the fact that the MTP_2 property implies the CI property. This means that the so-called *Bayesian relativities* are an increasing function of the level of the bonus–malus system. This is a desirable property of the system (see Borgan *et al.* (1981) and Norberg (1976) for more details on this topic).

Acknowledgement

The authors thank an anonymous referee for helpful comments.

References

Alam, K. and Wallenius, K. T. (1976). Positive dependence and monotonicity in conditional distributions. *Commun. Statist. Theory Meth.* **5,** 525–534.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.

Billingsley, P. (1995). *Probability and Measure*, 3rd edn. John Wiley, New York.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.

Block, H. W., Savits, T. H. and Shaked, M. (1982). Some concepts of negative dependence. *Ann. Prob.* **10,** 765–772.

Borgan, Ø., Hoem, J. M. and Norberg, R. (1981). A nonasymptotic criterion for the evaluation of automobile bonus systems. *Scand. Actuarial J.* **1981,** 165–178.
Colangelo, A., Scarsini, M. and Shaked, M. (2005). Some notions of multivariate positive dependence. *Insurance Math. Econom.* **37,** 13–26.
Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
Kamae, T., Krengel, U. and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Prob.* **5,** 899–912.
Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10,** 467–498.
Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Chapman and Hall, London.
Kimeldorf, G. and Sampson, A. R. (1987). Positive dependence orderings. *Ann. Inst. Statist. Math.* **39,** 113–128.
Kimeldorf, G. and Sampson, A. R. (1989). A framework for positive dependence. *Ann. Inst. Statist. Math.* **41,** 31–45.
Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37,** 1137–1153.
Milgrom, P. R. and Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica* **50,** 1089–1122.
Müller, A. (1997). Stochastic orders generated by integrals: a unified study. *Adv. Appl. Prob.* **29,** 414–428.
Müller, A. and Scarsini, M. (2001). Stochastic comparison of random vectors with a common copula. *Math. Operat. Res.* **26,** 723–740.
Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley, Chichester.
Norberg, R. (1976). A credibility theory for automobile bonus systems. *Scand. Actuarial J.* **1976,** 92–107.
Pellerey, F. and Semeraro, P. (2003). A positive dependence notion based on the supermodular order. Res. Rep. 6, Dipartimento di Matematica, Politecnico di Torino.
Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Ann. Statist.* **26,** 494–504.
Shiryaev, A. N. (1996). *Probability*, 2nd edn. Springer, New York.
Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators. *Ann. Math. Statist.* **29,** 588–592.