Adv. Appl. Prob. **23**, 46–63 (1991)
Printed in N. Ireland

ON CONDITIONAL STOCHASTIC ORDERING OF DISTRIBUTIONS

LUDGER RÜSCHENDORF,* *University of Münster*

Abstract

Conditional stochastic ordering is concerned with the stochastic ordering of a pair of probability measures conditional on certain subsets or sub-σ-algebras. Some basic results of conditional stochastic ordering were proved by Whitt. We extend some of Whitt's results and prove a basic relation between stochastic ordering conditional on subsets and stochastic ordering conditional on σ-algebras. In the second part of the paper we consider the ordering of conditional expectations. There are several different formulations of this problem motivated by different types of applications.

MONOTONE LIKELIHOOD RATIO; MAXIMAL INEQUALITIES; ASSOCIATION

1. Introduction

The problem of conditional stochastic ordering of distributions is a natural and well-motivated problem, and one can find in the literature a lot of examples and applications for the conditional stochastic ordering. To mention a few of these applications let X, Y be two life lengths. Instead of comparing only X and Y with respect to the stochastic ordering $\leqq_{st}$, say $X \leqq_{st} Y$, in survival analysis it is more useful to compare the conditional life length $P^{X|X\geqq t} \leqq_{st} P^{Y|Y\geqq t}$ for $t \geqq t_0$. Here $P^{X|X\geqq t}$ denotes the conditional distribution of X given that $X \geqq t$.

In reliability, let X, Y be two systems whose reliability is to be compared. Typically, their behaviour depends on the state of (a part of) the remaining components say Z. So a useful comparison between X and Y is to compare the conditional reliability probabilities, say for example $P^{X|Z\geqq u} \leqq_{st} P^{Y|Z\geqq u}$, where u is a certain vector in $\{0, 1\}^k$ or even the conditional distributions $P^{X|Z} \leqq_{st} P^{Y|Z}$ a.s. In some applications it will be difficult to specify the whole conditional distributions. In these cases it might be more natural to compare only the conditional expectations, say $E(X \mid Z) \leqq_{st} E(Y \mid Z)$ or the conditional variances. If X is a vector random variable, then it might be of interest to compare only the conditional distribution functions or conditional survival functions, e.g. $P(X \geqq u \mid Z) \leqq P(Y \geqq u \mid Z)$ a.s., $u \in \mathbb{R}^k$, which by some well-known results implies corresponding comparisons of the system reliability. A weaker form of comparison (for $k = 2$) is to consider the conditional covariances $E(X_1X_2 \mid Z) \leqq_{st} E(Y_1Y_2 \mid Z)$, or the covariances of the conditional expectations $E(X_1 \mid Z)E(X_2 \mid Z) \leqq_{st} E(Y_1 \mid Z)E(Y_2 \mid Z)$.

A general class of models where such questions arise is latent variable models,

Received 5 June 1989; revision received 1 December 1989.
* Postal address: Institut für Mathematische Statistik, Westfalische Wilhelms-Universität, Einsteinstrasse 62, W-4400 Münster, Germany.

which have a wide range of applications. In Holland and Rosenbaum [5] some of the applications to psychometrics, population genetics, factor analysis, response models and reliability are discussed. In particular the question of conditional stochastic dependence (positive or negative) is discussed in their paper. Questions of positive stochastic dependence are a special case of conditional stochastic ordering. Consider e.g. the positive orthant dependence (POD) of two random variables X, Y defined by $P(X \geqq u, Y \geqq v) \geqq P(X \geqq u)P(Y \geqq v)$. This relation is equivalent to $P(X \geqq u \mid Y \geqq v) \geqq P(X \geqq u)$ i.e. to the conditional stochastic ordering relation $P^{X|Y \geqq v} \geqq_{\mathrm{st}} P^X$ (we can write formally also $P^X = P^{X'|Y \geqq v}$, where X', Y are independent and X' has the same distribution as X).

Some general results on conditional stochastic ordering were derived in several papers of Whitt [12], [13], [14]. Conditional stochastic ordering has been considered in the literature in particular for the stochastic ordering with respect to monotone functions and for variability and dispersion type orderings (cf. [12], [13], [14], [7], [2]). We shall discuss some general results relating the ordering of elementary conditional probabilities to the ordering of the conditional distributions and derive some bounds in problems which are motivated by the applications discussed above.

The framework of this paper is the following. On the Borel space $(E, \mathscr{A})$ let $P, Q \in M^1(E, \mathscr{A})$—the set of probability measures—and let $U \subset \mathscr{L}_+(\mathscr{A})$ be a subset of the non-negative measurable real functions. The *U-ordering* $\leqq_U$ on $M^1(E, \mathscr{A})$ is defined by

$$(1) \qquad P \leqq_U Q \quad \text{if} \quad \int h\,dP \leqq \int h\,dQ \qquad \text{for all } h \in U.$$

Let $\mathscr{E} \subset \mathscr{A}$ be a subset of measurable sets. Then one defines the *conditional U-ordering* $\leqq_{U,\mathscr{E}}$ by

$$(2) \qquad P \leqq_{U,\mathscr{E}} Q \quad \text{if} \quad P_A \leqq_U Q_A \qquad \text{for all } A \in \mathscr{E},$$

where $P_A(B) := P(A \cap B)/P(A)$ is the elementary conditional distribution for $P(A) > 0$, while we define $\int h\,dP_A = 0$ for $P(A) = 0$, $h \in U$. If $\mathscr{E} = \{E\}$, then $\leqq_{U,\mathscr{E}}$ is identical to $\leqq_U$. If $\mathscr{E} = A\mathscr{A}$ for some $A \in \mathscr{A}$, then $P \leqq_{U,\mathscr{E}} Q$ is equivalent to $P_A \leqq_{U,\mathscr{A}} Q_A$.

In Section 2 we derive some general results for the conditional U-ordering. Section 3 contains some applications to the conditional stochastic ordering with respect to monotone functions. Finally in Section 4 we consider the question of stochastic ordering of conditional expectations and correlations.

2. Conditional ordering of distributions

For a sub-σ-algebra $\mathscr{B} \subset \mathscr{A}$ let $P_{\mathscr{B}}$ denote the conditional distribution of P given $\mathscr{B}$. Define

$$(3) \qquad P_{\mathscr{B}} \leqq_U Q_{\mathscr{B}}$$

if for any $h \in U$ there are versions h_P of $\int h\, dP_{\mathscr{B}} = E_P(h \mid \mathscr{B})$ and h_Q of $E_Q(h \mid \mathscr{B})$ such that $h_P \leqq h_Q[P+Q]$, i.e. $h_P \leqq h_Q$ a.s. with respect to $P+Q$.

Lemma 1. If $(B_n)_{n \in \mathbb{N}}$ is a measurable partition of E and $\mathscr{B} = \sigma(\{B_n; n \in \mathbb{N}\})$, then it holds that

$$P_{B_n} \leqq_U Q_{B_n}, \qquad n \in \mathbb{N} \Leftrightarrow P_{\mathscr{B}} \leqq_U Q_{\mathscr{B}}. \tag{4}$$

Proof. The proof is obvious from the formula

$$\int h\, dP_{\mathscr{B}} = \sum_{n=1}^{\infty} \left(\int h\, dP_{B_n}\right) 1_{B_n}, \qquad \int h\, dQ_{\mathscr{B}} = \sum_{n=1}^{\infty} \left(\int h\, dQ_{B_n}\right) 1_{B_n}. \tag{5}$$

To extend Lemma 1 to general sub-σ-algebras we need the following lemma. An index set T with a partial ordering $\leqq$ is called upwards filtering, if for any $s, t \in T$ there exists an element $u \in T$ with $s \leqq u$ and $t \leqq u$. Let $\mathscr{L}(\mathscr{A})(\mathscr{L}_+(\mathscr{A}))$ denote the set of all (non-negative) $\mathscr{A}$-measurable real functions. Let $\underset{P}{\rightarrow}$ denote convergence in probability.

Lemma 2. Let $(T, \leqq)$ be an upwards filtering index set and let $f_t, g_t, f, g \in \mathscr{L}(\mathscr{A})$, $f_t \underset{P}{\rightarrow} f$, $g_t \underset{Q}{\rightarrow} g$. If $f_t \leqq g_t[P]$, then there exist versions $\tilde{f}$ of f (with respect to P) and $\tilde{g}$ of g (with respect to Q) such that

$$\tilde{f} \leqq \tilde{g}[P+Q].$$

Proof. Assume at first that $P \ll Q$. By the $\varepsilon - \delta$ characterization of continuity we have: $\forall \varepsilon' > 0 : \exists \varepsilon > 0$, $\varepsilon \leqq \varepsilon'$, such that $Q(A) < \varepsilon$ implies $P(A) < \varepsilon'$. By assumption for $\delta > 0$ there exists a $t_{\varepsilon,\delta} \in T$ such that $Q(|g_t - g| > \delta) < \varepsilon'$ for $t \geqq t_{\varepsilon,\delta}$. By continuity, therefore, $P(|g_t - g| > \delta) < \varepsilon$ for $t \geqq t_{\varepsilon,\delta}$. Similarly, there exists $\bar{t}_{\varepsilon,\delta} \in T$ such that for $t \geqq \bar{t}_{\varepsilon,\delta}: P(|f_t - f| > \delta) < \varepsilon$. For $t \geqq t_{\varepsilon,\delta}$ and $t \geqq \bar{t}_{\varepsilon,\delta}$ we have $P(f > g + 2\delta) \leqq P(|f - f_t| > \delta) + P(|g - g_t| > \delta) \leqq 2\varepsilon$. So we have $f \leqq g[P]$. With $A := \{dP/dQ = 0\}$ and $\tilde{f} := f 1_{A^c} + g 1_A$ it holds that $f_t \underset{P}{\rightarrow} \tilde{f}$ and $\tilde{f} \leqq g[P+Q]$.

In the general case let $\mu := P + Q$ and let

$$\frac{dP}{dQ} = \frac{dP/d\mu}{dQ/d\mu}$$

denote the generalized likelihood ratio. Arguing with the continuity part of P with respect to Q we obtain as in the first part of the proof the existence of some versions $\tilde{f}$ of f (with respect to P) and $\tilde{g}$ of g (with respect to Q) satisfying $\tilde{f} \leqq \tilde{g}[P+Q]$.

Theorem 3. For $U \subset \mathscr{L}^1_+(\{P, Q\})$ and a sub-σ-algebra $\mathscr{B} \subset \mathscr{A}$:

$$P \leqq_{U,\mathscr{B}} Q \quad \text{implies that} \quad P_{\mathscr{B}} \leqq_U Q_{\mathscr{B}}. \tag{6}$$

Proof. Define $T := \{t = \sigma(\{B_i\}); (B_i) \subset \mathscr{B}$ is a measurable partition of $E\}$. T is upwards filtering and by the L^1-martingale convergence theorem (cf. [9], p. 95) for $f \in U$ it holds that: $f_t := E_P(f \mid t) \rightarrow E_P(f \mid \mathscr{B})$ in $\mathscr{L}^1(P)$ and $g_t := E_Q(f \mid t) \rightarrow E_Q(f \mid \mathscr{B})$ in $\mathscr{L}^1(Q)$ implying that $f_t \underset{P}{\rightarrow} E_P(f \mid \mathscr{B})$, $g_t \underset{Q}{\rightarrow} E_Q(f \mid \mathscr{B})$. Since by Lemma 1, $f_t = E_P(f \mid t) \leqq E_Q(f \mid t) = g_t[P]$ we obtain from Lemma 2, $E_P(f \mid \mathscr{B}) \leqq E_Q(f \mid t)[P]$, where we identify $E_P(t \mid \mathscr{B})$ with a suitable version $\tilde{f}$ as in Lemma 2.

Remark 1.

(a) The converse direction in Theorem 3 is generally false. Take for example $\mathcal{B} = \mathcal{A}$; then for any $h \in \mathcal{L}_+(\mathcal{A})$, $E_P(h \mid \mathcal{B}) \leqq E_Q(h \mid \mathcal{B})$ holds, i.e. $P_{\mathcal{B}} \leqq_U Q_{\mathcal{B}}$ is trivial but $P \leqq_{U,\mathcal{B}} Q$ obviously is not. The valid 'equivalence' formulation of Theorem 3 is: $P \leqq_{U,\mathcal{B}} Q$ if and only if $P_{\mathcal{C}} \leqq_U Q_{\mathcal{C}}$ for all sub-σ-algebras $\mathcal{C} \subset \mathcal{B}$. This equivalence follows from Lemma 1.

(b) There is the following extension of Theorem 3 to certain generating systems. Let $U \subset \mathcal{L}^1_+(\{P, Q\})$ and let $\mathcal{E} \subset \mathcal{A}$ satisfy the following two conditions:

1. $B_1, B_2 \in \mathcal{E}$ implies that $B_1 \cap B_2$, $B_1 \cap B_2^c \in \mathcal{E}$;
2. There exists a sequence $E_n \in \mathcal{E}$ with $E_n \uparrow E$.

If $P \leqq_{U,\mathcal{E}} Q$, then $P_{\mathcal{B}} \leqq_U Q_{\mathcal{B}}$ with $\mathcal{B} := \sigma(\mathcal{E})$.

Proof. Define $T := \{t = \sigma(\{B_i\});\ (B_i) \subset \mathcal{E}$ is a measurable partition of $E\}$. Then by assumptions (a), (b) $T \neq \emptyset$, T is upwards filtering with $\sigma\big(\bigcup_{t\in T} t\big) = \mathcal{B}$. Now the proof is analogous to that of Theorem 3.

Proposition 4.

(a) If P is a mixture, $P = \int_\Theta P_\theta \, d\mu(\theta)$, $\mu \in M^1(\Theta, \mathcal{C})$, and $P_\theta \leqq_{U,\mathcal{E}} Q$ for all $\theta \in \Theta$, then $P \leqq_{U,\mathcal{E}} Q$.

(b) If $T:(E, \mathcal{A}) \to (E', \mathcal{A}')$, $P_y := P(\cdot \mid T = y)$ is the conditional distribution, then $P_y \leqq_{U,\mathcal{E}} Q$ for P^T almost all $y \in E'$ implies that $P \leqq_{U,\mathcal{E}} Q$.

Proof.

(a) For $h \in U$ the following holds:

$$\int h \, dP_A = \frac{1}{P(A)} \int_A h \, dP = \frac{1}{P(A)} \int_\Theta \left(\int_A h \, dP_\theta \right) d\mu(\theta)$$
$$= \frac{1}{P(A)} \int_\Theta P_\theta(A) \left(\int h \, dP_{\theta,A} \right) d\mu(\theta) \leqq Q(A).$$

(b) follows from (a).

Part (a) of Proposition 4 is a generalization of Theorem 2.4 in [12]. There are some elementary bounds useful for conditional ordering.

Proposition 5. For $h \in U \subset \mathcal{L}_+(\mathcal{A})$, $\mathcal{B} \subset \mathcal{A}$ holds:

(a) If $A, B \in \mathcal{A}$, $A \cap B = \emptyset$, then

$$\min\left(\int h \, dP_A, \int h \, dP_B \right) \leqq \int h \, dP_{A \cup B} \leqq \max\left(\int h \, dP_A, \int h \, dP_B \right). \tag{7}$$

(b) With $h_{P,B} := \operatorname{ess\,inf}_P E_P(h \mid \mathcal{B}) 1_B$, $h_P^B := \operatorname{ess\,sup}_P E_P(h \mid \mathcal{B}) 1_B$ and $h_{Q,B}$, h_Q^B defined similarly it holds that:

$$h_{Q,B} - h_P^B \leqq \int h \, dQ_B - \int h \, dP_B \leqq h_Q^B - h_{P,B}. \tag{8}$$

(c) If $L = dQ/dP$ is the generalized likelihood ratio, $L_B := \text{ess sup}_P (L1_B)$, then for $h \in U$:

(9) $$\int h \, dQ_B \leqq \left(\frac{P(B)}{Q(B)} L_B\right) \int h \, dP_B + \int h 1_{\{L=\infty\}} \, dQ_B.$$

(d) If $P_{\mathscr{B}} \leqq_U Q_{\mathscr{B}}$ and $h \in U$ implies $E_Q(h \mid \mathscr{B}) 1_B \in U$ for $B \in \mathscr{E} \subset \mathscr{B}$, then $\int_B h \, dP \leqq \int_B h \, dQ$, $B \in \mathscr{E}$.

Proposition 6. For $P \in M^1(E, \mathscr{A})$, $h \in \mathscr{L}_+(\mathscr{A})$ and $A \in \mathscr{A}$ with $P(A) \geqq P(h \geqq t) > 0$ the following hold:

(a) $P_A^h \leqq_{\text{st}} P_{\{h \geqq t\}}^h$, $\leqq_{\text{st}}$ denoting stochastic order.

(b) $\int h \, dP_A \leqq \int h \, dP_{\{h \geqq t\}}$.

Proof. (a)

$$P_A(h \geqq u) = \frac{P((h \geqq u) \cap A)}{P(A)} \leqq \frac{P(h \geqq u)}{P(A)} \leqq \frac{P(h \geqq u)}{P(h \geqq t)} = P_{\{h \geqq t\}}(h \geqq u)$$

for $u > t$. For $u < t$ this inequality holds trivially. This implies part (b) by a well-known integration by parts formula.

Remark 2. If h is monotonically non-decreasing on $E = \mathbb{R}^1$, then $\{h \geqq t\}$ is of the form $[u, \infty)$ or (u, ∞) with $u = h^-(t)$ the generalized inverse, i.e. the right open intervals have the highest concentration. If h is unimodal, symmetric around zero, then $\{h \geqq t\}$—a symmetric interval around zero—has the highest concentration.

3. Conditional stochastic order

For $E = \mathbb{R}^k$ and $P, Q \in M^1(\mathbb{R}^k, \mathbb{B}^k)$ define the *MLR-ordering*

(10) $$P \leqq_r Q,$$

if there exist versions p, q of the densities of P, Q with respect to a dominating measure μ, such that $q(x)/p(x)$ is non-decreasing on the support of $P + Q$. For $\mathscr{E} \subset \mathscr{A} = \mathbb{B}^1$ and $U \subset \mathscr{L}_+(\mathbb{R}^1, \mathbb{B}^1)$ the set of all non-decreasing, non-negative functions define the *$\mathscr{E}$-conditional* stochastic order by

(11) $$P \leqq_{\mathscr{E},\text{st}} Q \quad \text{if} \quad P \leqq_{U',\mathscr{E}} Q \quad \text{with} \quad U' = U \cap \mathscr{L}^1(\{P, Q\}).$$

If $\mathscr{E} = \mathbb{B}^1$, then $\leqq_{\mathscr{E},\text{st}}$ is called *uniform conditional stochastic order* (UCSO) in the notation of Whitt [12]).

It is well known that $P \leqq_r Q$ implies $P \leqq_{\text{st}} Q$ for $k = 1$. Furthermore, the ordering $\leqq_r$ is 'stable' with respect to conditioning, i.e. $P \leqq_r Q$ implies $P_A \leqq_r Q_A$ if $P(A) > 0$, $Q(A) > 0$ (this idea of finding sufficient conditions for $\leqq_U$ which are stable with respect to conditioning is one general idea in conditional ordering). As a consequence one obtains part (a) of the following theorem which is due to Whitt [12], Theorem 1.1. Part (b) follows from (a) and Theorem 3.

Theorem 7. For $P, Q \in M^1(\mathbb{R}^1, \mathbb{B}^1)$ with $P \leqq_r Q$ holds:
(a) $P \leqq_{\mathbb{B}^1,\mathrm{st}} Q$;
(b) $P_{\mathscr{B}} \leqq_{\mathrm{st}} Q_{\mathscr{B}}$ for all sub-σ-algebras $\mathscr{B} \subset \mathbb{B}^1$.

The following result shows that the notion of uniform conditional stochastic order on $\mathbb{R}^k$ is too strong, amplifying p. 117 of [12]. Let $f\mu$ denote the measure with density f with respect to μ.

Theorem 8. Let $P, Q \in M^1(\mathbb{R}^k, \mathbb{B}^k)$, $P = f\mu$, $Q = g\mu$, μ σ-finite equivalent to $P + Q$ and $P \leqq_{\mathbb{B}^k,\mathrm{st}} Q$. Then:
(a) $\dfrac{f(x)}{f(y)} = \dfrac{g(x)}{g(y)} [\mu \otimes \mu]$ on $\{(x, y): x \nleqq y \text{ and } y \nleqq x\}$.
(b) $P \leqq_r Q$.

Proof.
(a) Let $\mathscr{B}_n = \sigma\{B_{i,n}\}$, $n \in \mathbb{N}$, be the increasing system of σ-algebras, generated by pairwise disjoint dyadic intervals $\left(\dfrac{i-1}{2^n}, \dfrac{i}{2^n}\right]$, $i = (i_1, \cdots, i_k) \in \mathbb{Z}^k$. For $x \in \mathbb{R}^k$, $h \in \mathbb{N}$ let $x \in B_{i_x,n} =: B_{x,n}$. For (x, y) such that neither $x \leqq y$ nor $y \leqq x$, it holds for $n \geqq n_0$ that $\{z \in \mathbb{R}^k; \exists w \in B_{x,n} \text{ with } z \leqq w\} \cap B_{y,n} = \emptyset$ and $\{z \in \mathbb{R}^k; \exists w \in B_{y,n} \text{ with } z \leqq w\} \cap B_{x,n} = \emptyset$. With $A_n := B_{x,n} \cup B_{y,n}$ the assumption $P_{A_n} \leqq_{\mathrm{st}} Q_{A_n}$ for $n \geqq n_0$ implies that

$$\frac{P(B_{y,n})}{P(B_{x,n}) + P(B_{y,n})} \leqq \frac{Q(B_{y,n})}{Q(B_{x,n}) + Q(B_{y,n})}, \qquad n \geqq n_0, \tag{12}$$

and also

$$\frac{P(B_{x,n})}{P(B_{x,n}) + P(B_{y,n})} \leqq \frac{Q(B_{x,n})}{Q(B_{x,n}) + Q(B_{y,n})}, \qquad n \geqq n_0,$$

i.e.

$$\frac{P(B_{x,n})}{P(B_{y,n})} = \frac{Q(B_{x,n})}{Q(B_{y,n})}, \qquad n \geqq n_0.$$

By the martingale convergence theorem

$$f_n := E_\mu(f \mid \mathscr{B}_n) = \sum \left(\int_{B_{i,n}} \frac{dP}{d\mu} d\mu_{B_{i,n}}\right) 1_{B_{i,n}} = \sum \frac{P(B_{i,n})}{\mu(B_{i,n})} 1_{B_{i,n}} \to f = \frac{dP}{d\mu} [\mu].$$

Similarly,

$$g_n = E_\mu(g \mid \mathscr{B}_n) = \sum \frac{Q(B_{i,n})}{\mu(B_{i,n})} \to g = \frac{dQ}{d\mu} [\mu].$$

This implies that

$$\frac{f_n(x)}{f_n(y)} \to \frac{f(x)}{f(y)} [(P + Q) \otimes (P + Q)].$$

Since by (12)

$$\frac{f_n(x)}{f_n(y)} = \frac{P(B_{x,n})}{P(B_{y,n})} = \frac{Q(B_{x,n})}{Q(B_{y,n})} = \frac{g_n(x)}{g_n(y)}$$

it follows that

$$\frac{f(x)}{f(y)} = \frac{g(x)}{g(y)} \qquad [\mu \otimes \mu].$$

(b) The proof of (b) is similar now using that for $x \leqq y$, $x \neq y$, $B_{x,n} \leqq B_{y,n}$ (elementwise) for $n \geqq n_0$ implying that $\frac{g(x)}{f(x)} \leqq \frac{g(y)}{f(y)} [\mu \otimes \mu]$ a.s. on $\{x \leqq y\}$.

Remark 3. Part (a) of Theorem 8 essentially implies that for equivalent distributions uniform conditional stochastic order implies that the common support of P, Q is totally ordered. Whitt [12], Theorem 1.3, gives a valid proof of part (b) for continuous densities. In the general case in Whitt's proof it is not clear that the continuous approximating densities can be chosen in a way such that the conditional stochastic order is preserved.

Let $\mathscr{L}$ denote the set of all lattices in $\mathbb{R}^k$ and define for $P, Q \ll \mu = \bigotimes_{i=1}^k \mu_i$ with densities p, q:

(13) $$P \leqq_{t_p} Q \quad \text{if} \quad p(x)q(y) \leqq p(x \wedge y)q(x \vee y) \qquad \text{for all } x, y$$

(cf. Karlin and Rinott [6]). Then $\leqq_{t_p}$ is stable by conditioning

(14) $$P \leqq_{t_p} Q \Rightarrow P_A \leqq_{t_p} Q_A \qquad \text{for all } A \in \mathscr{L}$$

and, as $\leqq_{t_p} \rightarrow \leqq_{\text{st}}$ (cf. 1.19 of Karlin and Rinott [6]),

(15) $$P \leqq_{t_p} Q \Rightarrow P \leqq_{\mathscr{L},\text{st}} Q,$$

(cf. Theorem 2.3 of Whitt [13]).

Define P to be MTP_2 if $P \leqq_{t_p} P$. Whitt proves also the following partial converse of (15). If P or Q is MTP_2, then:

(16) $$P \leqq_{t_p} Q \Leftrightarrow P \leqq_{\mathscr{L},\text{st}} Q \Leftrightarrow P \leqq_r Q.$$

P is called associated if $\int fg\,dP \geqq (\int f\,dP)(\int g\,dP)$ for all non-decreasing f, g such that the integrals exist. We have the following characterization of association.

Lemma 9. Let $P \in M^1(\mathbb{R}^k, \mathbb{B}^k)$, then P is associated iff for any $Q \in M^1(\mathbb{R}^k, \mathbb{B}^k)$, $Q \ll P$ and $P \leqq_r Q$ implies that $P \leqq_{\text{st}} Q$.

Proof. Association of P is equivalent to the condition that $\int fg\,dP \geqq \int f\,dP \int g\,dP$ for all $f, g \geqq 0$ non-decreasing integrable with respect to P. Defining $h = g/\int g\,dP$ and $Q = hP$ the measure with density h with respect to P we see that P is associated iff $\int f\,dQ \geqq \int f\,dP$ for all $Q = hP$ with $h\uparrow$.

Proposition 10. Let $\mathscr{E} \subset \mathbb{B}^k$, and let P be associated, conditionally on $\mathscr{E}$ (i.e. P_A is associated for $A \in \mathscr{E}$). If $P \ll Q$ and $P \leqq_r Q$, then $P \leqq_{\mathscr{E},\text{st}} Q$.

Proof. Since $\leqq_r$ is stable with respect to conditioning, Proposition 10 follows from Lemma 9.

If P is MTP_2, then P is associated conditionally on $\mathscr{L}$. But the MTP_2 property is much stronger than association (consider e.g. the case of normal distributions where association is equivalent with positive correlation while the MTP_2 property is equivalent to the condition that the inverse of the covariance matrix Σ^{-1} is an M-matrix). Note that Proposition 10 does not need the assumption of dominatedness by a product measure which is crucial for the application of the t_p-ordering and the MTP_2 property.

4. Ordering of conditional expectations

If one is not able to compare the whole conditional distributions, it may be possible in some cases to compare or bound conditional expectations $E(X \mid \mathscr{B})$, conditional covariances $E(XY \mid \mathscr{B})$ (assuming $EX = EY = 0$) or products of conditional expectations $E(X \mid \mathscr{B})E(Y \mid \mathscr{B})$.

For a real random variable $X \in \mathscr{L}^1(P)$ on $(\Omega, \mathscr{A}, P)$ define the Hardy–Littlewood maximal function

$$H(t) = E(X \mid X > t). \tag{17}$$

Then H is right continuous, monotonically non-decreasing on the support of P^X and

$$EX \leqq H(t) \leqq \operatorname*{ess\,sup}_P X. \tag{18}$$

By Proposition 6,

$$\sup\left\{\int X\, dP_A;\, A \in \mathscr{A},\, P(A) \geqq P(X \geqq t)\right\} = H(t). \tag{19}$$

The function $H(t)$ has an interesting property also for the formulation of sharp bounds for the set of conditional expectations $\{E(X \mid \mathscr{B});\, \mathscr{B} \subset \mathscr{A}\}$. Define

$$M(u) := \inf_{x<u} \frac{E(X-x)_+}{u-x}.$$

Theorem 11. (cf. Blackwell and Dubins [1], Dubins and Gilat [3], Meilijson and Nadas [8]).

(a) For any sub-σ-algebra $\mathscr{B} \subset \mathscr{A}$,

$$P(E(X \mid \mathscr{B}) \geqq u) \leqq M(u). \tag{20}$$

(b) For $u \in (EX, \operatorname{ess\,sup}_P X]$ define $x_0 := H^{-1}(u) = \inf\{t: H(t) \geqq u\}$, then:

$$P(X > x_0) \leqq \frac{E(x-x_0)_+}{u-x_0} \leqq P(X \geqq x_0) = P(H(X) \geqq u). \tag{21}$$

(c) $E(X \mid \mathscr{B}) \leqq_{\mathrm{st}} H(X)$ for any $\mathscr{B} \subset \mathscr{A}$.

(d) $M(u) = P(H(X) \geqq u) = \sup\{P(E(X \mid \mathscr{B}) \geqq u);\, \mathscr{B} \subset \mathscr{A}\}$, $\forall u$.

Remark 4.

(a) The proof of the crucial inequality in Theorem 11 is simple. By the Markov and Jensen inequality with $\varphi(t) := (t-x)_+$,

$$P(E(X \mid \mathscr{B}) \geqq u) \leqq P(\varphi(E(X \mid \mathscr{B})) \geqq \varphi(u)) \leqq \frac{E\varphi(E(X \mid \mathscr{B}))}{u-x} \leqq \frac{E\varphi(X)}{u-x} = \frac{E(X-x)_+}{u-x}.$$

(b) From representation results for the convex ordering $<_c$ (the ordering with respect to convex functions) it is known (cf. [3], [10], [11]) that

$$\begin{aligned} \{P^{E(X\mid\mathscr{B})}; \mathscr{B} \subset \mathscr{A}\} &= \{Q \in M^1(\mathbb{R}^1, \mathbb{B}^1); Q <_c P^X\} \\ &= \{P^Y; (Y, X) \text{ is a martingale}\}. \end{aligned} \tag{22}$$

Therefore, Theorem 11 gives the least upper bound (with respect to $\leqq_{st}$) for the set of all probability measures which are smaller than P^X with respect to $<_c$. Trivially, the bound in Theorem 11 is also valid and sharp for $\{Q \in M^1(\mathbb{R}^1, \mathbb{B}^1); Q <_m P^X\}$, where $<_m$ is the ordering with respect to convex monotone functions (cf. [9]). For $p > 1$ it follows from the Doob inequality that

$$\|H(X)\|_p \leqq \frac{p}{p-1} \|X\|_p \tag{23}$$

(cf. [3]). It is interesting to remind that Blackwell and Dubins [1] showed that the upper bound is even valid (and sharp) for the maxima of martingales in countable time implying in particular sharp prophet inequalities for martingale sequences.

Later on in this section we shall improve the bounds in (20), (21) and (c) and (d) of Theorem 11 under additional restrictions on the sub-σ-algebras. These restrictions are partially motivated by the following consideration of bounds for $E(XY \mid \mathscr{B})$ respectively for $E(X \mid \mathscr{B})E(Y \mid \mathscr{B})$ for sub-σ-algebras $\mathscr{B} \subset \mathscr{A}$ for real random variables X, Y.

We first note that integrability of XY, equivalently, of $E(XY \mid \mathscr{B})$ does generally not imply integrability of $E(X \mid \mathscr{B})E(Y \mid \mathscr{B})$ and conversely. The reason is that the choice of $\mathscr{B}$ may introduce strong dependence entailing that large values of X are combined with large values of Y.

Example 1.

(a) Let $X, Y \in \mathscr{L}^1(P)$ be independent, identically distributed, then $XY \in \mathscr{L}^1(P)$ and $EXY = EXEY = (EX)^2$. Let $T := X + Y$ and $\mathscr{B} := \sigma(T)$—the generated σ-algebra. Then $E(X \mid \mathscr{B}) = E(X \mid T) = \frac{1}{2}T = E(Y \mid T)$ and $E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) = \frac{1}{4}(X+Y)^2$. The product of the conditional expectations is, therefore, integrable if and only if $X, Y \in \mathscr{L}^2(P)$ and then it holds that

$$(EX)^2 = EXY \leqq EE(X \mid \mathscr{B})E(Y \mid \mathscr{B}) = \tfrac{1}{2}EX^2 + \tfrac{1}{2}(EX)^2. \tag{24}$$

(b) Similarly, if $X, Y \in \mathscr{L}^1(P)$ are i.i.d. with d.f. F symmetric around zero and $D := X - Y$, then $E(X \mid D) = \frac{1}{2}(X-Y)$, $E(Y \mid X-Y) = -\frac{1}{2}(X-Y)$ and, therefore,

$$E(X \mid D)E(Y \mid D) = -\tfrac{1}{4}(X-Y)^2, \tag{25}$$

which is integrable if and only if $X \in \mathscr{L}^2(P)$.

(c) Let $A_1, \cdots, A_k \in \mathscr{A}$, $k \leqq \infty$, be a partition of Ω and $\mathscr{B} = \sigma(A_1, A_2, \cdots, A_k)$. Then

$$E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) = \sum_{i=1}^{k} P(A_i)^{-2}\Big(\int_{A_i} X\,dP\Big)\Big(\int_{A_i} Y\,dP\Big)1_{A_i} \tag{26}$$

is integrable if X, Y are integrable. So XY is not necessarily integrable.

(d) Some implications are valid in more special cases:

(d1) If $p, q \in [1, \infty]$ are conjugate, $\frac{1}{p} + \frac{1}{q} = 1$ and $X \in \mathscr{L}^p$, $Y \in \mathscr{L}^q$, then $E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) \in \mathscr{L}^1$ for any $\mathscr{B}$.

(d2) If $\mathscr{B}$ is large in the sense that $\mathscr{B} \supset \sigma(X)$ or $\mathscr{B} \supset \sigma(Y)$, then $XY \in \mathscr{L}^1$ if and only if $E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) \in \mathscr{L}^1$.

Define for a fixed sub-σ-algebra $\mathscr{B} \subset \mathscr{A}$ the conditional distribution functions

$$F_{\mathscr{B}}(x) := P(X \leqq x \mid \mathscr{B}), \qquad G_{\mathscr{B}}(x) := P(Y \leqq x \mid \mathscr{B}) \quad \text{and} \quad H_{\mathscr{B}}(x, y) := P(X \leqq x, Y \leqq y \mid \mathscr{B}). \tag{27}$$

We now vary the dependence structure but fix the conditioning σ-algebra $\mathscr{B}$.

Proposition 12.

(a) $E(XY \mid \mathscr{B}) = E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) = \int\int (H_{\mathscr{B}}(x, y) - F_{\mathscr{B}}(x)G_{\mathscr{B}}(y)\,dx\,dy$ a.s. with respect to P.

(b) $\int_0^1 F_{\mathscr{B}}^{-1}(u)G_{\mathscr{B}}^{-1}(1-u)\,du \leqq E(XY \mid \mathscr{B}) \leqq \int_0^1 F_{\mathscr{B}}^{-1}(u)G_{\mathscr{B}}^{-1}(u)\,du$.

Proof.

(a) This follows from the covariance representation formula of Hoeffding applied to the conditional distributions.

(b) For the conditional distributions the Frechét bounds imply that

$$(F_{\mathscr{B}}(x) + G_{\mathscr{B}}(y) - 1)_+ \leqq H_{\mathscr{B}}(x, y) \leqq \min(F_{\mathscr{B}}(x), G_{\mathscr{B}}(y)) \text{ for } x, y \in \mathbb{R}^1. \tag{28}$$

Again by the Hoeffding formula this implies the corresponding inequality for the expectations. Since $(F_{\mathscr{B}}(x) + G_{\mathscr{B}}(y) - 1)_+$, $\min(F_{\mathscr{B}}(x), G_{\mathscr{B}}(y))$ are the d.f.'s of $(F_{\mathscr{B}}^{-1}(U), G_{\mathscr{B}}^{-1}(1-U))$ or $(F_{\mathscr{B}}^{-1}(U), G_{\mathscr{B}}^{-1}(U))$, where U is uniformly distributed on $[0, 1]$, (b) follows.

Remark 5.

(a) In particular, the condition

$$H_{\mathscr{B}}(x, y)(\gtreqless)F_{\mathscr{B}}(x)G_{\mathscr{B}}(y) \qquad \text{for } \lambda^2 \text{ a.a. } x, y \tag{29}$$

i.e. conditional positive (negative) quadrant dependence, implies that

$$E(XY \mid \mathscr{B})(\gtreqless)E(X \mid \mathscr{B})E(Y \mid \mathscr{B}). \tag{30}$$

Part (b) of Proposition 12 gives exactly the range of possible values of $E(XY \mid \mathscr{B})$ if the conditional distributions $F_{\mathscr{B}}$, $G_{\mathscr{B}}$ are fixed. The case of conditional independence 'corresponds' to

$$E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) = \left(\int_0^1 F_{\mathscr{B}}^{-1}(u)\,du\right)\left(\int_0^1 G_{\mathscr{B}}^{-1}(u)\,du\right) \tag{31}$$

lying somewhere between the bounds and each intermediate point is possible. In particular, the conditions $-\infty < E(\int_0^1 F_{\mathscr{B}}^{-1}(u)G_{\mathscr{B}}^{-1}(1-u)\,du)$ and $E(\int_0^1 F_{\mathscr{B}}^{-1}(u)G_{\mathscr{B}}^{-1}(u)\,du) < \infty$, imply that $XY \in \mathscr{L}^1$.

(b) A useful condition to imply that $H_{\mathscr{B}}(x, y) \geqq F_{\mathscr{B}}(x)F_{\mathscr{B}}(y)$ is to apply the multivariate total positivity (MTP_2) ordering (cf. (13)) to the conditional distributions. If $f_{\mathscr{B}}(x, y)$ is a density of $H_{\mathscr{B}}$ with respect to $\mu = \mu_1 \otimes \mu_2$, then the condition

$$f_{\mathscr{B}}(z_1)f_{\mathscr{B}}(z_2) \leqq f_{\mathscr{B}}(z_1 \vee z_2)f_{\mathscr{B}}(z_1 \wedge z_2) \tag{32}$$

for all $z_1, z_2 \in \mathbb{R}^2$, implies that $H_{\mathscr{B}}$ is associated and, therefore, $H_{\mathscr{B}}$ is positive quadrant dependent i.e. $H_{\mathscr{B}}(x, y) \geqq F_{\mathscr{B}}(x)G_{\mathscr{B}}(y)$, $\forall x, y \in \mathbb{R}^1$.

Example 2. Let X be distributed on $(0, \infty)$ and let $Y := 1/X$. For any sub-σ-algebra $\mathscr{B}$ of the Borel σ-algebra,

$$G_{\mathscr{B}}(y) = P(Y \leqq y \mid \mathscr{B}) = P\left(X \geqq \frac{1}{y} \,\middle|\, \mathscr{B}\right) = 1 - F_{\mathscr{B}}\left(\left(\frac{1}{y}\right)-\right) \tag{33}$$

(the minus sign denoting the left-hand limit).

For any $\frac{1}{y} \leqq x$, $H_{\mathscr{B}}(x, y) = P(X \leqq x, Y \leqq y \mid \mathscr{B}) = P\left(\frac{1}{y} \leqq X \leqq x \mid \mathscr{B}\right) = F_{\mathscr{B}}(x) - F_{\mathscr{B}}\left(\left(\frac{1}{y}\right)-\right) \leqq F_{\mathscr{B}}(x)\left(1 - F_{\mathscr{B}}\left(\left(\frac{1}{y}\right)-\right)\right) = F_{\mathscr{B}}(x)G_{\mathscr{B}}(y)$. Therefore, for any $\mathscr{B} \subset \mathscr{A}$, X, Y are conditional negative quadrant dependent. By Proposition 12

$$\begin{aligned} E(X \mid \mathscr{B})E(Y \mid \mathscr{B}) &= 1 - \iint (H_{\mathscr{B}}(x, y) - F_{\mathscr{B}}(x)G_{\mathscr{B}}(y))\,dx\,dy \\ &= 1 - \iint_{\{(1/y) \leqq x\}} F_{\mathscr{B}}\left(\left(\frac{1}{y}\right)-\right)(1 - F_{\mathscr{B}}(x))\,dx\,dy \\ &\quad + \iint_{\{(1/y) \leqq x\}} F_{\mathscr{B}}(x)\left(1 - F_{\mathscr{B}}\left(\left(\frac{1}{y}\right)-\right)\right)dx\,dy \\ &\geqq E(XY \mid \mathscr{B}) = 1. \end{aligned} \tag{34}$$

While Proposition 12 gives relevant bounds if one fixes the sub-σ-algebra $\mathscr{B} \subset \mathscr{A}$ but varies the dependence structure between X and Y, we now fix the joint distribution $P^{(X,Y)}$ but vary the σ-algebras $\mathscr{B} \subset \mathscr{A}$. Define

$$\begin{aligned} M(X, Y) &:= \sup_{\mathscr{B} \subset \mathscr{A}} E(E(X \mid \mathscr{B})E(Y \mid \mathscr{B})) = \sup_{\mathscr{B} \subset \mathscr{A}} EXE(Y \mid \mathscr{B}), \\ m(X, Y) &:= \inf_{\mathscr{B} \subset \mathscr{A}} E(E(X \mid \mathscr{B})E(Y \mid \mathscr{B})). \end{aligned} \tag{35}$$

Remark 5. The problem of determining M and m is related to a problem considered by Dubins and Pitman [4], namely deriving (sharp) upper bounds for $E\max_{1\leqq i\leqq n} E(X \mid \mathcal{B}_i)$, $0\leqq X\leqq 1$, $EX=p$, the bounds depending only on p and n. In this paper one also finds a motivation for this kind of problem. Sharp lower bounds for $E\max_{1\leqq i\leqq n} E(X \mid \mathcal{B}_i)$ are trivial as are sharp upper bounds for $E\prod_{i=1}^n E(X \mid \mathcal{B}_i)$. The question of lower bounds for $E\prod_{i=1}^n E(X \mid \mathcal{B}_i)$ remains open (cf. [4], p. 224).

Proposition 13. If $E(X \mid X+Y)=E(Y \mid X+Y)$, then

$$M(X, Y)=\tfrac{1}{4}E(X+Y)^2, \tag{36}$$

which is attained for $\mathcal{B}=\sigma(X+Y)$.

Proof. By the geometric–arithmetic mean and the Jensen inequality for any $\mathcal{B}\subset\mathcal{A}$, $E(X \mid \mathcal{B})E(Y \mid \mathcal{B})\leqq\frac{1}{2}(E(X \mid \mathcal{B})+E(Y \mid \mathcal{B}))^2=\frac{1}{4}(E(X+Y \mid \mathcal{B}))^2\leqq\frac{1}{4}E((X+Y)^2 \mid \mathcal{B})$. Therefore, $EE(X \mid \mathcal{B})E(Y \mid \mathcal{B})\leqq\frac{1}{4}E(X+Y)^2$. By our assumption $E(X \mid X+Y)=E(Y \mid X+Y)=\frac{1}{2}(X+Y)$, and, therefore, equality holds.

Remark 7.

(a) If $P^{(X,Y)}=P^{(Y,X)}$, then $E(X \mid X+Y)=E(Y \mid X+Y)$.

(b) The σ-algebra $\mathcal{B}=\sigma(X+Y)$ does not always produce large conditional correlation. For example, let $P^{(X,Y)}\{(1,0)\}=P^{(X,Y)}\{(0,2)\}=\frac{1}{2}$, then: $E(X \mid X+Y)=1_{(X+Y=1)}$, $E(Y \mid X+Y)=21_{(X+Y=2)}$ and, therefore, $E(X \mid X+Y)E(Y \mid X+Y)=0$, while $EXEY=\frac{1}{2}$, which is attained for the trivial σ-algebra $\mathcal{B}=\{\phi, \Omega\}$.

(c) If $X=U$ is uniformly distributed on $[0,1]$ and $Y=1-U$, then $X+Y=1$ and the proof of Proposition 13 yields:

$$M(U, 1-U)=\tfrac{1}{4} \text{ with equality for } \mathcal{B}=\{\phi, [0,1]\}. \tag{37}$$

Proposition 14. Assume that $X, Y\geqq 0$, $Y=\varphi(U)$, $X=\psi(U)$.

(a) If φ, ψ are monotonically non-decreasing, then $M(X, Y)=EXY$.

(b) If φ is monotonically non-increasing, ψ monotonically non-decreasing, then $m(X, Y)=EXY$.

Proof.

(a) A well-known rearrangement argument shows that $E(\psi(U)\varphi(U) \mid \mathcal{B})\geqq E(\psi(U) \mid \mathcal{B})E(\varphi(U) \mid \mathcal{B})$ and, therefore, $EE(X \mid \mathcal{B})E(Y \mid \mathcal{B})\leqq EXY$.

(b) is proved analogously.

Proposition 15. Let $X, Y\geqq 0$ and $\varphi:[0,\infty)\to\mathbb{R}_+$ be measurable, $\varphi(0)=0$, φ monotonically non-decreasing with generalized inverse φ^{-1}. Define the Young functions $\phi(x):=\int_0^x\varphi(u)\,du$, $\psi(x):=\int_0^x\varphi^{-1}(u)\,du$, then

(a) $A:=EE(X \mid \mathcal{B})E(Y \mid \mathcal{B})\leqq B:=E\phi(X)+E\psi(E(Y \mid \mathcal{B}))\leqq C:=E\phi(X)+E\psi(Y)$.

(b) $A=B$ if and only if $E(Y \mid \mathcal{B})=\varphi(X)$.

(c) If $Y=\varphi(X)$, then $M(X, Y)=E\phi(X)+E\psi(Y)$.

Proof. Since $EE(X \mid \mathscr{B})E(Y \mid \mathscr{B}) = EXE(Y \mid \mathscr{B})$, (a) and (b) follow from the Young inequality. (c) follows from (a), (b) and the Jensen inequality.

The results and examples so far suggest to look for σ-algebras $\mathscr{B}$ such that $E(Y \mid \mathscr{B}) = \varphi(X)$, $\varphi\uparrow$. But it is easy to construct examples where an 'optimal' $\mathscr{B}$ is not of this type. From Theorem 11 and (28) we obtain the following general bound for the survival function of $(X, E(Y \mid \mathscr{B}))$.

Proposition 16. With $H(y) := E(Y \mid Y \geqq y)$:

(a) $P(X \geqq x, E(Y \mid \mathscr{B}) \geqq y) \leqq P(F_X^{-1}(U) \geqq x, F_{H(Y)}^{-1}(U) \geqq y) = \min\{\bar{F}_X(x), \bar{F}_{H(Y)}(y)\}$, $x, y \in \mathbb{R}^1$; U denoting a random variable distributed uniformly on $(0, 1)$, $\bar{F}_X(x)$, $\bar{F}_{H(Y)}$ denoting the survival functions of X, $H(Y)$.

(b) $EXE(Y \mid \mathscr{B}) \leqq \int_0^1 F_X^{-1}(u) F_{H(Y)}^{-1}(u)\, du$.

Proof. $P(X \geqq x, \quad E(Y \mid \mathscr{B}) \geqq y) \leqq \sup\{\mu([(x, y), \infty)); \quad \mu \in M(P^X, P^{E(Y \mid \mathscr{B})})\} \leqq \sup\{\mu[(x, y), \infty); \mu \in M(P^X, P^{H(Y)})\}$, where $M(P, Q)$ denotes the measures with marginals P, Q; the last inequality following from Theorem 11. Now (a) follows from the Frechét bounds, while (b) is a consequence of (a).

The bounds in Proposition 16 for $EE(X \mid \mathscr{B})E(Y \mid \mathscr{B})$ are typically too large. By Propositions 14 and 15, they are quite good if X and Y are approximately similarly ordered. This leads to the question whether for the determination of $M(X, Y)$ one can restrict to sub-σ-algebras $\mathscr{B}$ such that X and $E(Y \mid \mathscr{B})$ are similarly ordered.

The next example shows that $M(X, Y)$ in general is not identical to

$$(38) \quad \tilde{M}(X, Y) := \sup\{EXE(Y \mid \mathscr{B}); \mathscr{B} \subset \mathscr{A}, X \text{ and } E(Y \mid \mathscr{B}) \text{ are similarly ordered}\}.$$

Example 3. Let $(\Omega, \mathscr{A}, P) = ([0, 1), \ [0, 1)\mathbb{B}^1, \lambda^1)$ and $X =_{[\frac{1}{3}, \frac{2}{3})} + 3\, 1_{[\frac{2}{3}, 1)}$, $Y = 3\, 1_{[0, \frac{1}{3})} + 1_{[\frac{1}{3}, \frac{2}{3})}$. Then $M(X, Y) = EXE(Y \mid \mathscr{B}^*) = \frac{11}{6}$, where $\mathscr{B}^* = \sigma\{[0, \frac{1}{3}) \cup [\frac{2}{3}, 1)\}$ while the optimal monotone solution is $\tilde{\mathscr{B}} = \{\phi, [0, 1)\}$ which yields $\tilde{M}(X, Y) = EXE(Y \mid \tilde{\mathscr{B}}) = \frac{16}{9} < \frac{11}{6} = M(X, Y)$.

If $X = \sum_{j=1}^K x_j 1_{B_j}$ is a discrete random variable, $B_j \in \mathscr{A}$, then for an optimal solution $\mathscr{B}^*$ satisfying $M(X, Y) = EXE(Y \mid \mathscr{B}^*)$ we can restrict to sub-σ-algebras of $\sigma\{B_j, j = 1, \cdots, K\}$.

Proposition 17. If $X = \sum_{j=1}^K x_j 1_{B_j}$ is discrete, $K \leqq \infty$ and $\mathscr{B}^* = \sigma\{A_1, A_2, \cdots, A_m\}$, (A_i) pairwise disjoint, is an optimal σ-algebra satisfying $EXE(Y \mid \mathscr{B}^*) = M(X, Y)$, then for $i < j \leqq M$

$$(39) \quad \begin{gathered} E(X \mid A_i) < E(X \mid A_j) \quad \text{implies} \quad E(Y \mid A_i) \leqq E(Y \mid A_j), \\ \text{i.e.} \quad E(X \mid \mathscr{B}^*),\ E(Y \mid \mathscr{B}^*) \text{ are similarly ordered.} \end{gathered}$$

Proof. Let $p_i = P(A_i) > 0$, $a_i = E(X \mid A_i)$, $b_i = E(Y \mid A_i)$ and assume that $a_i < a_j$

and $b_i < b_j$, without loss of generality assume that $i=1$, $j=2$. Then

$$E(X \mid A_1 \cup A_2) = \frac{1}{p_1+p_2}(p_1a_1+p_2a_2),$$
$$E(Y \mid A_1 \cup A_2) = \frac{1}{p_1+p_2}(p_1b_1+p_2b_2).$$

Therefore,

$$\begin{aligned}(p_1+p_2)&E(X \mid A_1\cup A_2)E(Y \mid A_1 \cup A_2)\\ &= \frac{1}{p_1+p_2}(p_1a_1+p_2a_2)(p_1b_1+p_2b_2)\\ &= \frac{1}{p_1+p_2}(p_1^2a_1b_1+p_1p_2b_2+p_1p_2b_1a_2+p_2^2a_2b_2)\\ &> \frac{1}{p_1+p_2}[p_1^2a_1b_1+p_1p_2(a_1b_1+a_2b_2)+p_2^2a_2b_2]\\ &= \frac{1}{p_1+p_2}[(p_1+p_2)p_1a_1b_1+(p_1+p_2)p_2a_2b_2]\\ &= E[E(X \mid A_1)E(Y \mid A_1)1_{A_1}+E(X \mid A_2)E(Y \mid A_2)1_{A_2}],\end{aligned}$$

the inequality following from $a_1b_1+a_2b_2 < a_1b_2+a_2b_1$. With $\tilde{\mathscr{B}} := \sigma(A_1 \cup A_2, A_3, \cdots, A_M)$ a contradiction to the optimality of $\mathscr{B}^*$ follows.

It is not difficult but a bit technical to extend Proposition 17 to the general non-discrete case. We next evaluate $\tilde{M}(X, Y)$ explicitly, in this way sharpening the bound of Theorem 11 considerably under the additional hypothesis that X and $E(Y \mid \mathscr{B})$ are similarly ordered.

Assume in the first case that X, Y are defined on $[0,1)$ with $P=\lambda^1$ and assume that X is monotonically increasing. We also assume that $a_1 = 0 < a_2 < \cdots < a_n = 1$ is a partition of $[0,1)$ such that $Y(u) = y_j$ for $u \in [a_j, a_{j+1})$. Define a sequence $j_1 > j_2 > \cdots > j_k = 1$ by:

$$\begin{aligned}
&j_1 := \inf\left\{j \leqq n: E(Y \mid [a_j, a_n)) = \sup_i E(Y \mid [a_i, a_n))\right\}\\
(40)\quad &j_2 := \inf\left\{j < j_1: E(Y \mid [a_j, a_{j_1})) = \sup_{i<j_1} E(Y \mid [a_i, a_{j_1}))\right\} \quad \text{if} \quad j_1 > 1\\
&\vdots\\
&j_l := \inf\left\{j < j_{l-1}: E(Y \mid [a_j, a_{j_{l-1}})) = \sup_{i<j_{l-1}} E(Y \mid [a_i, a_{j_{l-1}}))\right\}, \quad \text{if} \quad j_{l-1} > 1\\
&j_k := 1,
\end{aligned}$$

define $\mathscr{B}^* = \sigma\{[a_{j_1}, a_n), [a_{j_2}, a_{j_1}), \cdots, [0, a_{j_{k-1}})\}$ and

$$(41)\qquad Y^* = E(Y \mid \mathscr{B}^*) = \sum_{r=1}^{k} y_r^* 1_{[a_{j_r}, a_{j_{r-1}})} \quad \text{with} \quad y_r^* := E(Y \mid [a_{j_r}, a_{j_{r-1}})).$$

Y^* is monotonically non-decreasing and Y^* has the following uniform optimality property independently from the special non-decreasing X.

Theorem 18. If $\tilde{Y} = E(Y \mid \tilde{\mathscr{B}})$ is monotonically non-decreasing on $[0, 1]$ and Y^* is defined as in (41), then

$$\tilde{Y}P \leqq_{\text{st}} Y^*P, \tag{42}$$

$\tilde{Y}P$ denoting the measure with density $\tilde{Y}$ with respect to $P = \lambda^1_{|[0,1]}$. Equivalently,

$$\int X\tilde{Y}\, dP \leqq \int XY^*\, dP = \tilde{M}(X, Y) \tag{43}$$

for all $X: [0, 1] \to \mathbb{R}_+$ monotonically non-decreasing.

Proof. Since $\tilde{Y} = E(Y \mid \tilde{\mathscr{B}})$ is monotonically non-decreasing, we can assume without loss of generality that $\tilde{\mathscr{B}} = \sigma\{[a_{s_i}, a_{s_{i+1}}), 1 \leqq i \leqq r-1\}$, where $1 = s_1 < s_2 < \cdots < s_{r+1} = n$. By definition of Y^* it holds that

$$\int_{a_{j_i}}^1 Y^*(u)\, dP(u) = \int_{a_{j_i}}^1 Y(v)\, dP(v), \qquad 1 \leqq i \leqq k,$$

and, furthermore, for any $u \in [a_{j_{i+1}}, a_{j_i})$ it holds that

$$\int_u^{a_{j_i}} Y^*(v)\, dP(v) \geqq \int_u^{a_{j_i}} Y(v)\, dP(v), \tag{44}$$

since

$$\frac{1}{a_{j_i} - u} \int_u^{a_{j_i}} Y(v)\, dP(v) \leqq \frac{1}{a_{j_i} - a_{j_{i+1}}} \int_{a_{j_{i+1}}}^{a_{j_i}} Y(v)\, dP(v) = Y^*(v)$$

for $v \in [a_{j_{i+1}}, a_{j_i})$. Therefore, in particular for $u = a_{s_i}$ it holds that

$$\int_{a_{s_i}}^1 Y^*(v)\, dP(v) \geqq \int_{a_{s_i}}^1 Y(v)\, dP(v) = \int_{a_{s_i}}^1 \tilde{Y}(v)\, dP(v). \tag{45}$$

This implies that $\int_u^1 Y^*(v)\, dP(v) \geqq \int_u^1 \tilde{Y}(v)\, dP(v)$ as follows from the following argument. If $a_{j_{i+1}} < a_{s_l} < u < v < a_{j_i} < a_{s_{l+1}}$ and

$$\int_u^1 \tilde{Y}(v)\, dP(v) > \int_u^1 Y^*(v)\, dP(v),$$

then in case that $\tilde{Y}(v) > Y^*(v)$ we also would have that

$$\int_{a_{s_l}}^1 \tilde{Y}(v)\, dP(v) > \int_{a_{s_l}}^1 Y^*(v)\, dP(v)$$

in contradiction to (45). But if $\tilde{Y}(v) \leqq Y^*(v)$, then also $\tilde{Y}(v) = \tilde{Y}(v') \leqq Y^*(v')$ for $a_{j_i} \leqq v' < a_{s_{l+1}}$ and we would have

$$\int_{a_{s_{l+1}}}^1 \tilde{Y}(v)\, dP(v) > \int_{a_{s_{l+1}}}^1 Y^*(v)\, dP(v),$$

again a contradiction to (45). In the case that $a_{j_{i+1}} < a_{s_l} < u < a_{s_{l+1}} < a_{j_i}$ we can argue in a similar way. This completes the proof.

So in the class of monotonically non-decreasing elements $E(Y \mid \tilde{\mathcal{B}})$ we found one element $Y^* = E(Y \mid \mathcal{B}^*)$ such that Y^*P is the largest one with respect to stochastic order. If we denote $Y^*(v) = y_j^*$ for $v \in [a_j, a_{j+1})$, $1 \leqq j \leqq n$, and similarly $\tilde{Y}(v) = \tilde{y}_j$ for any $\tilde{Y} = E(Y \mid \tilde{\mathcal{B}})$, then we get the following corollary.

Corollary 19. Assume that $P([a_j, a_{j+1})) = \dfrac{1}{n}$, $1 \leqq j \leqq n-1$, and define $A_\uparrow := \{\tilde{y} = (\tilde{y}_1, \cdots, \tilde{y}_n);\ \exists \tilde{\mathcal{B}}$ with $\tilde{Y} = E(Y \mid \tilde{\mathcal{B}})\uparrow,\ \tilde{Y} \mid [a_j, a_{j+1}) = \tilde{y}_j\}$. Then with respect to Schur-ordering $\leqq_{\text{sch}}$, $y^* = (y_1^*, \cdots, y_n^*)$ is the largest element of $A_\uparrow$.

Obviously the smallest element of $A_\uparrow$ with respect to $\leqq_{\text{sch}}$ is $(\bar{y}, \cdots, \bar{y})$ with $\bar{y} := EY$. It is now also obvious how to construct decreasing elements $\tilde{Y} = E(Y \mid \tilde{\mathcal{B}})$ and how to minimize

(45) $$\int X\tilde{Y}\, dP \geqq \int XY_*\, dP, \quad \text{with} \quad Y_* = E(Y \mid \mathcal{B}_*),$$

with respect to this class, namely, by defining the sequence $0 \leqq j_1 \leqq j_2 \cdots$ in (46) increasing from the left to the right.

We next extend our construction to the case that $X = \sum_{j=1}^n x_j 1_{B_j}$ is finite discrete on $(\Omega, \mathcal{A}, P)$, $x_1 \leqq \cdots \leqq x_n$ and $y_j := E(Y \mid X = x_j) = E(Y \mid B_j)$ arranges Y in the order of X by conditioning. Define

$$j_1 := \inf \{j \leqq n: E(Y \mid X \geqq x_j) = \sup_{1 \leqq i \leqq n} E(Y \mid X \geqq x_i)\}$$

(47) $$j_2 := \inf \{j < j_1: E(Y \mid x_j \leqq X < x_{j_1}) = \sup_{i < j_1} E(Y \mid x_i \leqq X < x_{j_1}) \text{ if } j_1 > 1$$

$$\vdots$$

$$j_k = 1.$$

Defining

(48) $$\mathcal{B}^* = \sigma\{X \in [x_{j_1}, x_n], X \in [x_{j_2}, x_{j_1}), \cdots\}, \qquad Y^* := E(Y \mid \mathcal{B}^*),$$

then Y^* and X are similarly ordered, i.e. $X(w) < X(w') \Rightarrow Y^*(w) \leqq Y^*(w')$ and $Y^*(w) < Y^*(w') \Rightarrow X(w) \leqq X(w')$. With the same arguments as in the proof of Theorem 18 we obtain the following.

Theorem 19. If $Y = E(Y \mid \tilde{\mathcal{B}})$ is similarly ordered as X, then for all j:

(49) $$\int_{X \geqq x_j} \tilde{Y}\, dP \leqq \int_{X \geqq x_j} Y^*\, dP$$

and, in particular,

(50) $$\int X\tilde{Y}\, dP \leqq \int XY^*\, dP = \tilde{M}(X, Y).$$

By a limiting argument one can extend the construction of Y^* to non-discrete random variables X using the inequalities:

$$\left|\sup_{\mathscr{B}} E(YE(X \mid \mathscr{B})) - \sup_{\mathscr{B}} E(YE(\tilde{X} \mid \mathscr{B}))\right| \leqq \sup_{\mathscr{B}} |EYE(X - \tilde{X} \mid \mathscr{B})|$$
$$\leqq \varepsilon E\,|Y| \quad \text{if} \quad |X - \tilde{X}| \leqq \varepsilon.$$

But since this construction is not very explicit it may be more useful to construct a somewhat larger majorant which is not necessarily monotonically non-decreasing. For X, Y real and $t \in \mathbb{R}^1$ define

(51) $$G(t) = \sup_{s<t} E(Y \mid s \leqq X \leqq t);$$

then we have the following quite obvious result.

Proposition 20. If $X, Y \geqq 0$, and $\tilde{Y} = E(Y \mid \mathscr{B})$ is similarly ordered as X, then for all x

(52) $$\int_{X \geqq x} \tilde{Y}\, dP \leqq \int_{X \geqq x} G(X)\, dP.$$

In particular,

(53) $$\int X\tilde{Y}\, dP \leqq \int XG(X)\, dP.$$

Again there are also antitone versions of (51). Obviously, $G(X) \geqq X^*$ in the cases where we have defined X^*. On the other hand $G(X)$ is smaller than the 'sharp' stochastic bound $H(X)$ considered in Proposition 16, or Theorem 11 respectively. So in the case where we can assume that the ordering of $E(Y \mid \mathscr{B})$ is the same as the ordering of X we obtain a strong improvement of the crude stochastic bounds in Proposition 16 (or Theorem 11), while in the general case (without restricting the conditional information $\mathscr{B}$) our proposed construction method seems to be quite good if the orderings of X, Y are not 'too much' in an opposite direction i.e. if the correlation is not very negative. These kind of assumptions seem to be realistic, e.g. in examples concerning the reliability of systems where the positive dependence of two components should not be too much destroyed by additional information on the system.

Acknowledgement

I thank the referee for several suggestions concerning the organization of the manuscript.

References

[1] BLACKWELL, D. AND DUBINS, L. E. (1963) A converse to the dominated convergence theorem. *Illinois J. Math.* **7**, 508–514.

[2] Block, H. W. and Sampson, A. R. (1988) Conditionally ordered distributions. *J. Multivariate Anal.* **27,** 91–104.

[3] Dubins, L. E. and Gilat, D. (1978) On the distribution of the maxima of martingales. *Trans. Amer. Math. Soc.* **68,** 337–338.

[4] Dubins, L. E. and Pitman, J. (1980) A maximal inequality for skew fields. *Z. Wahrscheinlichkeitsth.* **52,** 219–227.

[5] Holland, P. W. and Rosenbaum, P. R. (1986) Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14,** 1523–1543.

[6] Karlin, S. and Rinott, Y. (1980) Classes of orderings of measures and related correlation inequalities. I Multivariate totally positive functions. *J. Multivariate Anal.* **10,** 467–498.

[7] Keilson, J. and Sumita, U. (1982) Uniform stochastic ordering and related inequalities. *Canad. J. Statist.* **10,** 181–198.

[8] Meilijson, I. and Nadas, A. (1979) Convex majorization with an application to the length of critical paths. *J. Appl. Prob.* **16,** 671–677.

[9] Neveu, J. (1975) *Discrete Parameter Martingales.* North-Holland, Amsterdam.

[10] Rüschendorf, L. (1981) Ordering of distributions and rearrangement of functions. *Ann. Prob.* **9,** 276–283.

[11] Strassen, V. (1966) The existence of probability measures with given marginals. *Ann. Math. Statist.* **36,** 423–439.

[12] Whitt, W. (1980) Uniform conditional stochastic order. *J. Appl. Prob.* **17,** 112–123.

[13] Whitt, W. (1982) Multivariate monotone likelihood ratio and uniform conditional stochastic order. *J. Appl. Prob.* **19,** 695–701.

[14] Whitt, W. (1985) Uniform conditional variability ordering of probability distributions. *J. Appl. Prob.* **22,** 619–633.