STOCHASTICALLY MONOTONE DEPENDENCE

BY BENNETT BRADY AND NOZER D. SINGPURWALLA[1]

U.S. Nuclear Regulatory Commission and George Washington University

The notion of monotone dependence, which has played a key role in reliability theory, is generalized to that of "stochastically monotone dependence." The idea here is that since two lifelengths are dependent or independent based on the disposition of a conditioning variable, they are unconditionally stochastically dependent or independent. A measure of stochastic dependence is introduced and the measure used for comparing the correlations of pairs of random variables which can now be described as being "highly stochastically correlated" or "weakly stochastically correlated." Extensions to the multivariate case are possible and the ideas illustrated via examples. This paper is expository; its purpose is to propose a natural idea and to explore its ramifications.

1. Introduction and Motivation. An important, though little noticed, principle of probability theory is that the notions of dependence and independence are conditional, the conditioning being done on some observable or unobservable quantity, say Θ. It is common to think of Θ as a "parameter" and this is the point of view that we adopt. A consequence of the above is that unconditionally the notions of dependence and independence must be stochastic. That is, one should not make an unqualified judgment that lifelengths X_1 and X_2 are dependent or independent—rather one may talk in terms of the *probability that they are dependent or independent.* This is contrary to current thinking although the literature on artificial intelligence [cf. Pearl (1989)] appears to be taking cognizance of this fact. In this paper, we explore the ramifications of the above formulation, and in the sequel raise questions pertaining to the everyday used notions of covariance and correlation.

By way of some motivation, consider a system of two components with lifelengths X_1 and X_2, operating in an environment which is characterized by an

[1]Research supported by Grant DAAL 03–87–K–0056, the U.S. Army Research Office, Contract N00014–85–K–0202, Project NR 042–372 Office of Naval Research, and the Air Force Office of Scientific Research, Grant AFOSR–89–0381.

AMS 1980 subject classifications. Primary 60K10; secondary 62N05.

Key words and phrases. Biometry, multivariate lifelengths, reliability.

Discussions with Henry Block and Allan Sampson have helped clarify several issues; the operational scheme following Example 2.1 is a consequence of such discussions.

abstract (idealized and unobservable) parameter $\Theta \in \mathbf{R}$. Suppose that I_1, I_2, and I_3 partition the real line $\mathbf{R}$ such that $I_1 \cup I_2 \cup I_3 = \mathbf{R}$, and suppose that when $\Theta \in I_1$, the operating environment is classified as being "average" or "normal" whereas when $\Theta \in I_2$ of I_3 the operating environment is classified as being "mild" or "harsh," respectively. Now it is possible to conceive of a situation in which X_1 is independent of X_2—denoted henceforth as $X_1 \perp\!\!\!\perp X_2$—whenever $\Theta \in I_1$, and that X_1 and X_2 are positively or negatively dependent whenever $\Theta \in I_2$ or I_3 respectively. In any particular application, the exact disposition of Θ will be unknown or will change—from the point of view of an analyst—and so the nature of dependence between X_1 and X_2 is stochastic, depending on the probability that Θ belongs to I_1, I_2, or I_3.

Other scenarios which motivate the thesis of this paper arise from the biological sciences in which the nature of dependence between the lifelengths of two organs depends on the stochastic behavior of a conditioning covariate, such as the "life-style" of an individual.

Whereas the rationalization of positive dependence under common environmental conditions is relatively straightforward, see for example Lindley and Singpurwalla (1986), the rationalization of independence and negative dependence, particularly the latter, is more difficult. One possible argument is to suppose that under harsh conditions there may be a tendency to devote more resources and maintenance to the more important components of the system with the result that such components perform better than expected than those components which receive less attention. Such a policy would result in negative dependence.

Latent variable methods (cf. Holland and Rosenbaum (1986)) consider the concept of the distribution of a set of random variables given the latent variable. Such models consist of a set of "manifest variables" and the "latent or parametric variable." The manifest variables which are real or integer valued can be observed directly while the latent variable is unobservable. A basic assumption of the model is that the manifest variables are conditionally independent given the latent variable. Certain classes of latent variable models imply that the manifest variables exhibit stronger forms of positive dependence with the latent variable. However, these models do not incorporate the notion that the dependence of the manifest variables may change with a change in the latent variable.

In view of the preceding arguments it is necessary to reconsider the various notions of monotone dependence and their resulting bounds and inequalities. In this paper we define a new concept of dependence between random variables. Two variables are not unconditionally dependent or independent but are probably dependent or independent, depending on the disposition of the conditioning variable.

2. Stochastic Monotone Dependence. Let A be a σ-field of events generated by a sample space $\mathcal{X}$ and $\mathbf{P}$ be a family of probability measures defined on A_i, $i = 1, 2, \ldots$, the elements of A. Let $\mathbf{X}$ and $\mathbf{Y}$ be two vector valued random variables, of dimension p and q respectively, defined on $\mathcal{X}$; assume for now that $p = q \geq 1$.

The notation $((\mathbf{X} \perp\!\!\!\perp \mathbf{Y})|\boldsymbol{\theta})$ means that $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is an s-dimensional vector of parameters. Without any loss of generality assume that $s = 1$.

2.1 Stochastic Dependence and Independence. Suppose we have a partial ordering on $\mathbf{R}^p$ such that for vectors $\mathbf{a} = (a_1, a_2, \ldots, a_p)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_p)$ in $\mathbf{R}^p$

$$\mathbf{a} \leq \mathbf{b} \text{ means } a_i \leq b_i, \quad i = 1, 2, \ldots, p$$

and suppose that the elements A_i are open upper sets, i.e., A_i is an upper set if $\mathbf{a} \in A_i$, and $\mathbf{a} \leq \mathbf{b}$ implies $\mathbf{b} \in A_i$ (Shaked (1982)).

DEFINITION 2.1. The random vectors $\mathbf{X}$ and $\mathbf{Y}$ are independent given $\boldsymbol{\theta}$, denoted by $\{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y})|\boldsymbol{\theta}\}$ if

$$P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta}\} = P\{\mathbf{X} \in A_i | \boldsymbol{\theta}\}, \ \ \forall A_i,\ A_j,\ \boldsymbol{\theta}, \text{ and any } P \in \mathbf{P}.$$

Suppose that $\boldsymbol{\theta}$ takes values in $\mathbf{R}$, and suppose that the Borel σ-field generated by $\mathbf{R}$ is endowed with a family of probability measures $\tilde{\mathbf{P}}$. Let I_1, I_2, and I_3 be members of the Borel σ-field generated by $\mathbf{R}$ and let $\tilde{P} \in \tilde{\mathbf{P}}$.

DEFINITION 2.2. The random vector $\mathbf{X}$ is $\theta \in I_1$ *conditionally independent of* $\mathbf{Y}$ and $\boldsymbol{\theta} \notin I_1$ *conditionally dependent on* Y, denoted by $\{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y})|\boldsymbol{\theta} \in I_1,\ \neq\}$, if

i) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \in I_1\} = P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \in I_1\}$,

ii) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \notin I_1\} \neq P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \notin I_1\}$, and $\forall A_i,\ A_j,\ \boldsymbol{\theta}$.

DEFINITION 2.3. The random vectors $\mathbf{X}$ and $\mathbf{Y}$ are unconditionally independent, denoted by $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y})$ if

$$\nexists \boldsymbol{\theta} \ni P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta}\} \neq P\{\mathbf{X} \in A_i | \boldsymbol{\theta}\}, \ \ \forall A_i,\ A_j.$$

The subjective nature of the notion of independence is revealed in Definition 2.3 if one interprets $\nexists \boldsymbol{\theta}$ as being the nonexistence—to a probability assessor—of a $\boldsymbol{\theta}$.

LEMMA 2.1. *If* $\{\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \boldsymbol{\theta} \in I_1,\ \neq\}$, *and if* $\{\boldsymbol{\theta} \in I_1\}$ *is the only event for which* $\mathbf{X}$ *and* $\mathbf{Y}$ *are conditionally independent, then*

$$P\{\mathbf{X} \perp\!\!\!\perp \mathbf{Y}\} = \tilde{P}\{\boldsymbol{\theta} \in I_1\}.$$

PROOF. From Definition 2.3 we see that

$$\begin{aligned} P\{\mathbf{X} \perp\!\!\!\perp \mathbf{Y}\} &= P\{\nexists \boldsymbol{\theta} \ni P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta}\} \neq P\{\mathbf{X} \in A_i | \boldsymbol{\theta}\}\} \\ &= 1 - \{P\{\nexists \boldsymbol{\theta} \ni P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta}\} \neq P\{\mathbf{X} \in A_i | \boldsymbol{\theta}\}\} \\ &= 1 - (1 - \Pi(\boldsymbol{\theta})) = \Pi(\boldsymbol{\theta}), \end{aligned}$$

where $\Pi(\boldsymbol{\theta}) = \tilde{P}(\boldsymbol{\theta} \in I_1)$.

A strengthening of Definition 2.2 is given next.

DEFINITION 2.4. The random vector $\mathbf{X}$ is $\boldsymbol{\theta} \in I_1$ *conditionally independent of* $\mathbf{Y}$, and is $\boldsymbol{\theta} \notin I_1$ *conditionally positively (negatively) dependent on* $\mathbf{Y}$, denoted $\{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}) | (\boldsymbol{\theta} \in I_1,\ > (<))\}$, if

i) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \in I_1\} = P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \in I_1\}$, and $\forall A_i,\ A_j,\ \boldsymbol{\theta}$,

ii) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \notin I_1\} \geq (\leq) P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \notin I_1\}$.

For convenience we denote $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \notin I_1\} \geq P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \notin I_1\}$ by $\mathbf{X} \perp\!\!\!\perp^{(+)} \mathbf{Y}$, and by $\mathbf{X} \perp\!\!\!\perp^{(-)} \mathbf{Y}$ when the above inequality is reversed.

LEMMA 2.2. *If* $\{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}) | \boldsymbol{\theta} \in I_1,\ > (<)\}$, *and* $\boldsymbol{\theta}$ *is unique, then*

$$\begin{aligned} &P\{\mathbf{X} \perp\!\!\!\perp \mathbf{Y}\} = \Pi(\boldsymbol{\theta}) \ \textit{and} \\ &P\{\mathbf{X} \perp\!\!\!\perp^{(+)} \mathbf{Y}\} \ \textit{or} \ P\{\mathbf{X} \perp\!\!\!\perp^{(-)} \mathbf{Y}\} = 1 - \Pi(\boldsymbol{\theta}). \end{aligned}$$

PROOF. Follows from Lemma 2.1 and the fact that it is not possible to have both $P\{\mathbf{X} \perp\!\!\!\perp^{(+)} \mathbf{Y}\}$ and $P\{\mathbf{X} \perp\!\!\!\perp^{(-)} \mathbf{Y}\}$.

A further strengthening of Definition 2.4 is given next.

DEFINITION 2.5. The random vector $\mathbf{X}$ is $\boldsymbol{\theta} \in I_1$ *conditionally independent of* $\mathbf{Y}$, and is $\boldsymbol{\theta} \in I_2(I_3)$ *conditionally positively (negatively) dependent on* $\mathbf{Y}$, denoted by $\{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}) | (\boldsymbol{\theta} \in I_1,\ > \boldsymbol{\theta} \in I_2,\ < \boldsymbol{\theta} \in I_3)\}$, if

i) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \in I_1\} = P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \in I_1\}$,

ii) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \in I_2\} \geq P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \in I_2\}$, and

iii) $P\{\mathbf{X} \in A_i | \mathbf{Y} \in A_j,\ \boldsymbol{\theta} \in I_3\} \leq P\{\mathbf{X} \in A_i | \boldsymbol{\theta} \in I_3\}, \quad \forall A_i,\ A_j,\ \boldsymbol{\theta}$.

We may now state

LEMMA 2.3. *If* $\Pi_i(\boldsymbol{\theta}) = \tilde{P}\{\boldsymbol{\theta} \in I_i\}$, *then*

$$\begin{aligned} P\{\mathbf{X} \perp\!\!\!\perp \mathbf{Y}\} &= \Pi_1(\boldsymbol{\theta}) \\ P\{\mathbf{X} \perp\!\!\!\perp^{(+)} \mathbf{Y}\} &= \Pi_2(\boldsymbol{\theta}), \text{ and} \\ P\{\mathbf{X} \perp\!\!\!\perp^{(-)} \mathbf{Y}\} &= \Pi_3(\boldsymbol{\theta}). \end{aligned}$$

2.2. Equivalent Conditions and Definitions. Assume that $p = q = 1$. Definition 2.5 can also be stated in terms of the joint and the marginal distribution functions of X and Y. Let

$$\begin{aligned} F(x, y|\boldsymbol{\theta}) &= P\{X \leq x,\ Y \leq y|\boldsymbol{\theta}\} \\ G(x|\boldsymbol{\theta}) &= P\{X \leq x|\boldsymbol{\theta}\}, \text{ and} \\ H(y|\boldsymbol{\theta}) &= P\{Y \leq y|\boldsymbol{\theta}\}. \end{aligned}$$

Then Definition 2.5 is equivalent to:

DEFINITION 2.6. The random vector $\mathbf{X}$ is $\boldsymbol{\theta} \in I_1$ *conditionally independent of* $\mathbf{Y}$ and is $\boldsymbol{\theta} \in I_2(I_3)$ *conditionally positively (negatively) quadrant dependent on* $\mathbf{Y}$, denoted by $\{(\mathbf{X} \perp\!\!\!\perp \mathbf{Y})|\boldsymbol{\theta} \in I_1,\ > \boldsymbol{\theta} \in I_2,\ < \boldsymbol{\theta} \in I_3)\}$, if

i) $F(x, y|\boldsymbol{\theta} \in I_1) = G(x|\boldsymbol{\theta} \in I_1)H(y|\boldsymbol{\theta} \in I_1)$,

ii) $F(x, y|\boldsymbol{\theta} \in I_2) \geq G(x|\boldsymbol{\theta} \in I_2)H(y|\boldsymbol{\theta} \in I_2)$, and

iii) $F(x, y|\boldsymbol{\theta} \in I_3) \leq G(x|\boldsymbol{\theta} \in I_3)H(y|\boldsymbol{\theta} \in I_3)$.

Using a lemma by Hoeffding (1940), we are able to state Lemma 2.5.

LEMMA 2.4. *(Hoeffding) If F denotes the joint, and G and H the marginal distribution functions of X and Y respectively, then*

$$E(XY) - E(X)E(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F(x, y) - G(x)H(y)]dx\, dy,$$

provided the expectations exist.

LEMMA 2.5. *If conditions i), ii), and iii) of Definition 2.6 hold and if the conditional expectations $E(XY|\boldsymbol{\theta})$, $E(X|\boldsymbol{\theta})$ and $E(Y|\boldsymbol{\theta})$ exist, then Definition 2.6 implies that*

i) $E(XY|\boldsymbol{\theta} \in I_1) = E(X|\boldsymbol{\theta} \in I_1)E(Y|\boldsymbol{\theta} \in I_1)$,

ii) $E(XY|\boldsymbol{\theta} \in I_2) \geq E(X|\boldsymbol{\theta} \in I_2)E(Y|\boldsymbol{\theta} \in I_2)$, *and*

iii) $E(XY|\boldsymbol{\theta} \in I_3) \leq E(X|\boldsymbol{\theta} \in I_3)E(Y|\boldsymbol{\theta} \in I_3)$.

It follows then that

LEMMA 2.6. $\{(X \perp\!\!\!\perp Y) | \boldsymbol{\theta} \in I_1, > \boldsymbol{\theta} \in I_2, < \boldsymbol{\theta} \in I_3\} \Longrightarrow$

i) $COV(X, Y | \boldsymbol{\theta} \in I_1) = 0$,

ii) $COV(X, Y | \boldsymbol{\theta} \in I_2) \geq 0$, *and*

iii) $COV(X, Y | \boldsymbol{\theta} \in I_3) \leq 0$.

EXAMPLE 2.1. As an example illustrating the intent of Lemma 2.6, suppose that X and Y have a bivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and covariance

$$\sum = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where ρ is the coefficient of correlation. Then ρ is our conditioning variable and as is well known $\{(X \perp\!\!\!\perp Y) | \rho = 0\}$, and X and Y have positive (negative) dependence when $\rho > (<) 0$.

To describe the operational implication of the above, suppose that we are asked to make a prediction of X when $Y = y$ has been observed. To simplify matters suppose that $\boldsymbol{\mu} = (0, 0)$ and that $P(\rho \neq 0) = \Pi_1$. Assuming that the penalty of poor prediction is described by the squared error loss, we would specify $E(X|y,\ \rho = 0) = 0$, or $E(X|y,\ \rho \neq 0) = \rho\sigma_1 y/\sigma_2$. Operationally, we would toss a coin whose probability of heads is Π_1, and bet on $\rho\sigma_1 y/\sigma_2$ if the coin lands heads, and on 0 if it lands tails.

A strengthening of Lemma 2.6 is

LEMMA 2.7. *Let f and g be nondecreasing functions of X and Y, respectively. If X and Y satisfy Definition 2.6, then*

i) $COV(f(X), g(Y) | \boldsymbol{\theta} \in I_1) = 0$,

ii) $COV(f(X), g(Y) | \boldsymbol{\theta} \in I_2) \geq 0$, *and*

iii) $COV(f(X), g(Y) | \boldsymbol{\theta} \in I_3) \leq 0$.

PROOF. This follows by an extension of a proof due to Lehmann (1966).

When ii) holds, we shall say that X and Y are *conditionally $\theta \in I_2$ positively associated*; and when iii) holds, we shall say that X and Y are *conditionally $\theta \in I_3$ negatively associated.*

Our motivation for a consideration of the above material is the introduction of the notion of stochastic covariance and correlation. This is discussed next.

2.3. Stochastic Linear Dependence. The notions of covariance and correlation appear in everyday use of probability and statistics. In sample theory statistics, the covariance and correlation have been viewed as fixed but unknown quantities. In Bayesian statistics, all unknown quantities are to be assigned a prior distribution to reflect one's uncertainty about them. To see this, note that the conditional nature of Definition 2.6, and its implications prompt us to generalize the notion of covariance and make it stochastic. The situation here is not unlike that of hierarchical modelling [c.f. Good (1983)].

Suppose that $p = q = s = 1$, so that $\mathbf{X} = X$, $\mathbf{Y} = Y$, and $\boldsymbol{\theta} = \theta$. Then from Definition 2.6, it follows that COV $(X, Y|\theta \in (I_1 \cup I_2)) \geq 0$, where

$$\begin{aligned} \mathrm{COV}\,(X,Y|\theta \in (I_1 \cup I_2)) &= E\,(XY|\theta \in (I_1 \cup I_2)) \\ &-E\,(X|\theta \in (I_1 \cup I_2))\,E\,(Y|\theta \in (I_1 \cup I_2)). \end{aligned}$$

Using an argument analogous to that of Lemma 2.1, we see that unconditionally $P\{\mathrm{COV}(X,Y) \geq 0\} = P\{\theta \in (I_1 \cup I_2)\}$. Thus a prior distribution on the covariance would depend on the nature of the parameterization of the probability model for X and Y, and our uncertainty about the disposition of the parameter. If $\theta \in I_1$, or if $\theta \in I_2$, that is, if we judge $(X \perp\!\!\!\perp Y)$ or $(X \perp\!\!\!\perp^{(+)} Y)$, and if θ is unique, then $P\{\mathrm{COV}(X,Y) \geq 0\} = 1$. Thus $P\{\mathrm{COV}(X,Y) \geq 0\}$ gives us a measure of the strength of the linear relationship between X and Y.

The above motivates us to consider the quantity $\Pi(\alpha) = P\{|\mathrm{COV}(X,Y)| \geq \alpha\}$, $\alpha \geq 0$, for characterizing the strength of linear dependence between X and Y.

DEFINITION 2.7. To simplify matters, suppose that $E(X|\theta) = E(Y|\theta) = 0$ and that VAR $(X|\theta)$ = VAR $(Y|\theta) = 1$, for all values of θ: then COV $(X,Y) = \rho(X,Y)$, the correlation coefficient between X and Y. Let $\Pi(\alpha) = P\{|\mathrm{COV}(X,Y)| \geq \alpha\} = P\{|\rho(X,Y)| \geq \alpha\}$; we refer to $\Pi(\alpha)$ as the *correlation survival function.*

It is clear that $\Pi(\alpha) \downarrow \alpha$, $\Pi(0) = 1$, and $\Pi(1^+) = 0$. A plot of $\Pi(\alpha)$ versus α, $0 \leq \alpha \leq 1$, for $\Pi(\alpha) = 1 - \alpha$, is given on the following page.

EXAMPLE 2.2. As an example of the above, let X and Y be binary with $P\{X = 1\} = p_x$, $P\{Y = 1\} = p_y$, and $P\{X = 1,\ Y = 1\} = p_{xy}$. Clearly, $P\{X = 1,\ Y = 0\} = p_x - p_{xy}$, $P\{X = 0,\ Y = 1\} = p_y - p_{xy}$, $P\{X = 0,\ Y = 0\} = 1 - p_x - p_y + p_{xy}$, and thus $0 \leq p_{xy} \leq \min(p_x, p_y)$.

Suppose that p_x and p_y are specified, but the disposition of p_{xy} is unknown. Then it can be verified that

$$(\rho(X,Y)|p_{xy}) = \frac{(p_{xy} - p_x p_y)}{\sqrt{p_x p_y (1 - p_x)(1 - p_y)}}$$

from which it follows that

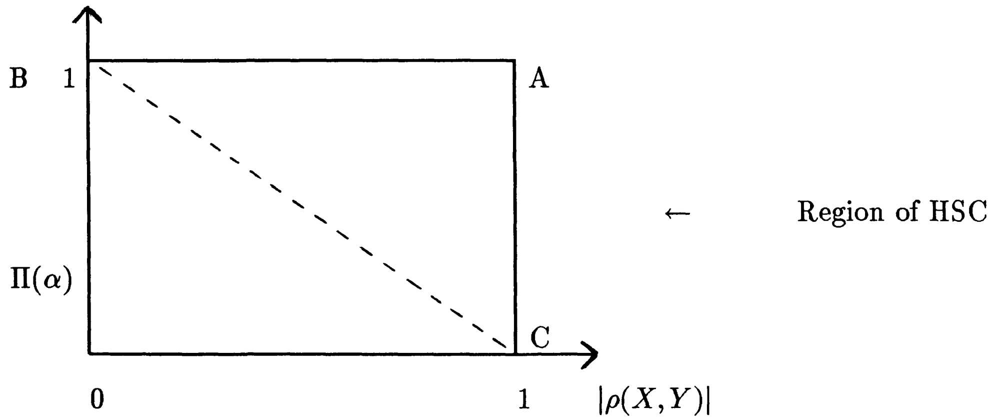

Figure 2.1. The correlation survival function when $\Pi(\alpha) = 1 - \alpha$

$$\Pi(\alpha) = 1 - \{\Delta(p_x p_y + \alpha\sqrt{p_x p_y (1-p_x)(1-p_y)}) \\ -\Delta(p_x p_y - \alpha\sqrt{p_{xy}(1-p_x)(1-p_y)})\}$$

where $\Delta(\cdot)$ is the cumulative distribution function of p_{xy}, for $0 \leq x \leq \min(p_x, p_y)$.

To characterize the strength of linear dependence between X and Y via $\Pi(\alpha)$, we first note that the strongest case for linear dependence is when $P\{\rho(X,Y) = 1\} = 1$; that is, when we are absolutely sure that the values of X and Y match perfectly. When this happens $P\{|\rho(X,Y)| \geq \alpha\} = \Pi(\alpha) = 1$, for all $0 \leq \alpha \leq 1$; that is, the correlation survival curve is the locus BAC in Figure 2.1. The worst case for linear dependence is when $P\{|\rho(X,Y)| \geq \alpha\} = 0$ for all $\alpha > 0$. For this case the correlation survival rate is the locus BOC. Thus, any correlation survival function which is closer to the locus BAC is to be preferred, in the sense that the variables are more highly correlated, to the one which is closer to the locus BOC. The above considerations prompt us to suggest the following as a plausible criteria for describing a high stochastic correlation.

DEFINITION 2.8. Random variables X and Y are said to be *highly stochastically correlated* (HSC) if

$$P\{|\mathrm{COV}(X,Y)| \geq \alpha\} = P\{|\rho(X,Y)| \geq \alpha\} \geq 1 - \alpha,$$

$0 \leq \alpha \leq 1$; otherwise they are *weakly stochastically correlated* (WSC).

EXAMPLE 2.3. Suppose that in Example 2.2, that p_{xy} has a uniform distribution on $[0, \min(p_x, p_y)]$. Then,

$$\Pi(\alpha) = 1 - \frac{(2\alpha\sqrt{p_x p_y(1-p_x)(1-p_y)})}{\min(p_x, p_y)}.$$

The binary variables X and Y are HSC if $\Pi(\alpha) \geq 1 - \alpha$; i.e., if

$$\frac{1}{2} \geq \frac{\sqrt{p_x p_y(1-p_x)(1-p_y)}}{\min(p_x, p_y)}.$$

In order to compare the strength of linear dependence between two pairs of random variables, we introduce the following definition.

DEFINITION 2.9. Random variables (X, Y) are *stochastically more (less) correlated* than (X^1, Y^1) if

$$P\{|\rho(X,Y)| \geq \alpha\} \geq (\leq) P\{|\rho(X^1,Y^1)| \geq \alpha\}, \quad 0 \leq \alpha \leq 1.$$

We shall denote the above writing $\rho(X,Y) \overset{\text{st}}{\geq} (\overset{\text{st}}{\leq}) \rho(X^1, Y^1)$.

Definition 2.9 provides a basis for comparing the strength of the linear dependence between (X, Y) and (X^1, Y^1) when their respective correlation survival functions do not cross. To characterize linear dependence when the correlation survival functions do cross, we need the following.

DEFINITION 2.10. Random variables (X, Y) are *more (less) correlated in expectation* than (X^1, Y^1) if

$$\int_0^1 \Pi_{X,Y}(\alpha)\, d\alpha \geq (\leq) \int_0^1 \Pi_{X^1,Y^1}(\alpha)\, d\alpha$$

where

$$\Pi_{X,Y}(\alpha) = P\{|\rho(X,Y)| \geq \alpha\} \text{ and } \Pi_{X^1,Y^1}(\alpha) = P\{|\rho(X^1,Y^1)| \geq \alpha\}.$$

We shall denote the above by writing $\rho(X,Y) \overset{E}{\geq} (\overset{E}{\leq}) \rho(X^1, Y^1)$. It is obvious from the above that

PROPOSITION 2.1. $\rho(X,Y) \overset{\text{st}}{\geq} (\overset{\text{st}}{\leq}) \rho(X^1, Y^1) \Longrightarrow \rho(X,Y) \overset{E}{\geq} (\overset{E}{\leq}) \rho(X^1, Y^1)$.

EXAMPLE 2.4. The survival function of Marshall and Olkin's bivariate exponential distribution is given as

$$\bar{F}(x, y) = e^{-\lambda_1 x - \lambda_2 y - \lambda \max(x,y)},$$

where we suppose λ_1 and λ_2 known but λ is unknown. Then $\{X \perp\!\!\!\perp Y | \lambda = 0, > \lambda \in (0, \infty)\}$; i.e., X and Y are independent if $\lambda = 0$, and otherwise always positive dependent.

The coefficient of correlation is

$$\rho(X,Y|\lambda) = \lambda/(\lambda_1 + \lambda_2 + \lambda),$$

from which it follows that the correlation survival function is

$$\Pi(\alpha) = P\{\rho(X,Y) \geq \alpha\} = 1 - \Delta(\frac{\alpha}{1-\alpha}[\lambda_1 + \lambda_2])$$

where $\Delta(\cdot)$ is the cumulative distribution function of λ, for $\lambda \geq 0$.

If λ has an exponential distribution with parameter λ', then the correlation survival function is

$$\Pi(\alpha) = \exp\{-\frac{\alpha\lambda'(\lambda_1 + \lambda_2)}{1-\alpha}\}.$$

So X and Y are HSC if

$$\exp\{-\frac{\alpha\lambda'(\lambda_1 + \lambda_2)}{1-\alpha}\} \geq 1 - \alpha,$$

that is, if

$$\lambda' \leq -\frac{1-\alpha}{\alpha(\lambda_1 + \lambda_2)} \log(1-\alpha), \quad \text{for all } \alpha, \quad 0 \leq \alpha \leq 1.$$

An extension of the above ideas to the multivariate case together with notions of higher order stochastic dependence are discussed in Brady (1988).

REFERENCES

BRADY, B.M. (1988). Stochastic positive and negative dependence. Doctoral dissertation. Department of Operations Research, The George Washington University, Washington, DC 20052.

GOOD, I.J. (1983). *Good Thinking.* University of Minnesota Press, Minneapolis, MN.

HOEFFDING, W. (1940). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–375.

HOLLAND, P.W. and ROSENBAUM, P.R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14**, No. 4, 1523–1543.

LEHMANN, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1152.

LINDLEY, D.V. and SINGPURWALLA, N.D. (1986). Multivariate distributions for lifelengths of components of a system sharing a common environment. *J. Appl. Prob.* **23** No. 2.

PEARL, J. (1989). *Probabilistic Reasoning in Intelligent System: Networks of Plausible Inference.* Morgan Kauffman Publishers, Inc., San Mateo, CA.

SHAKED, M. (1982). A general theory of some positive dependence notions. *J. Multi. Anal.* **12** 199–218.

U.S. NUCLEAR REGULATORY COMMISSION
WASHINGTON, D.C. 20555

DEPARTMENT OF OPERATIONS RESEARCH
GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC 20052