Stochastic Orders and Decision under Risk
IMS Lecture Notes – Monograph Series (1991)

STOCHASTIC ORDER AND MARTINGALE DYNAMICS IN MULTIVARIATE LIFE LENGTH MODELS: A REVIEW

BY ELJA ARJAS AND ILKKA NORROS

University of Oulu and Technical Research Centre of Finland

The purpose of this paper is to review some ideas connected to aging and dependence, in the context of technical reliability. The dynamic aspects of these notions are stressed throughout. The review is based mainly on the authors' own work during the past decade, but it connects very closely with some recent results by Shaked and Shanthikumar. New definitions, results and examples are also presented.

1. Introduction and Mathematical Preliminaries. In this paper we review some notions of aging and dependence which arise naturally in the context of engineering reliability. These notions are based on "ordinary" stochastic order of multivariate distributions in the positive orthant. But the given definitions differ in two important respects from the standard comparison of multivariate distributions with respect to stochastic order: they are *dynamic* and *conditional*.

The first characterization means that *time* becomes a key element of our analysis. Time is of course present in every meaningful notion of aging. But it is equally basic in every causality reasoning, and therefore also enters our modeling of dependence.

The second characterization emphasizes the role of *information*, which corresponds to the observed behavior of the considered device in the past and forms a natural basis on which its future behavior can be predicted.

From a mathematical point of view our approach to modeling aging and dependence can be seen as a particular application of the modern stochastic calculus and martingale theory for point processes. As a consequence, our presentation is somewhat unusual in the reliability tradition. However, we

---

AMS 1980 Subject Classification: primary 60G55; secondary 60G25, 90B25

Key words and phrases: Marked point process, prediction process, reliability, positive dependence.

maintain that the concepts introduced are very intuitive and easily interpreted in an actual engineering context. For the necessary mathematical background we refer to the monographs Liptser and Shiryayev (1978), Brémaud (1981) and Karr (1986), and to the recent review article Arjas (1989).

Throughout this paper we consider "a device consisting of k parts" ($k \geq 1$), and denote the corresponding k-vector of life lengths by $\mathbf{S} = (S_1, \ldots, S_k)$. Rather than viewing $\mathbf{S}$ as a random point in $\mathbf{R}_+^k$, we consider the corresponding part failures sequentially in the order in which they occur in calendar time. Suppose for convenience that all parts are in a working state at time $t = 0$. We then arrive at an alternative description of $\mathbf{S}$ in terms of a *marked point process* (MPP) $(T, J) := \{(T_n, J_n);\ n \geq 1\}$, where

$$\begin{aligned} &T_1 = \inf\{S_i :\ 1 \leq i \leq k\} \quad \text{and} \quad J_1 = \{i :\ 1 \leq i \leq k,\ S_i = T_1\}, \\ &\quad \ldots \\ &T_n = \inf\{S_i :\ 1 \leq i \leq k,\ S_i > T_{i-1}\} \quad \text{and} \quad J_n = \{i :\ 1 \leq i \leq k,\ S_i = T_n\}. \end{aligned} \tag{1.1}$$

T_n is then the nth smallest of the time epochs at which one or more of the k parts fail, and the corresponding mark J_n is simply the list of parts which fail at T_n. We follow the convention that $\inf \emptyset = \infty$, and let $J_n = \emptyset$ if $T_n = \infty$. The set of all possible marks (for finite T_n) is then

$$\mathcal{J} = \{I :\ \emptyset \neq I \subseteq \{1, 2, \ldots, k\}\}.$$

We use the (partial) order of inclusion for the elements of $\mathcal{J}$. It is natural to call (T, J) the *failure process*, and consider it simply on the canonical space

$$\begin{aligned} \Omega := \{(t_n, I_n) :\ &0 \leq t_1 \leq t_2 \leq \cdots \nearrow \infty; \\ &t_n < \infty \Rightarrow t_{n+1} > t_n,\ I_n \in \mathcal{J};\ t_n = \infty \Rightarrow I_n = \emptyset\} \end{aligned}$$

of marked point sequences. We denote by

$$N_t(I) = \sum_{n \geq 1} 1_{\{T_n \leq t, J_n = I\}},\ t \geq 0,\ I \in \mathcal{J},$$

the corresponding counting processes. Clearly $N(I) = \{N_t(I) :\ t \geq 0\}$ counts "one" at T_n if $J_n = I$, and remains zero if there is no such T_n.

Apart from the final Section 6, we assume that the level of information regarding the behavior of the considered device corresponds exactly to observing when its parts fail. Mathematically this corresponds to conditioning the prediction made at time t on the σ-field generated by the pre-t part failures, i.e., on

$$\mathcal{F}_t = \sigma\{N_s(I) :\ s \leq t,\ I \in \mathcal{J}\}. \tag{1.2}$$

As a convention we assume that the null-sets of $\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t$ are included already in $\mathcal{F}_0$.

The notion of *hazard* can now be formulated in terms of the $(\mathcal{F}_t)$-compensators $A(I) = \{A_t(I) : t \geq 0\}$ of the processes $N(I)$, $I \in \mathcal{J}$. For the definition of a compensator we refer to the general references on martingale theory mentioned before. However, for our purposes it will be sufficient that each process $A(I)$ can be viewed as the cumulative hazard which is specific to failure pattern I and based on knowing the previous part failures, in the sense that the following interpretation can be made:

$$dA_t(I) = P(dN_t(I) = 1 \mid \mathcal{F}_{t-}).$$

(Note that this is a purely heuristic formula, it does not hold literally for a fixed t.) We call $\{A(I) : I \in \mathcal{J}\}$ the *hazard process* of S (Arjas (1981b)).

The above representation of S in terms of a failure process (T, J) is obviously completely general, in the sense that to every value of S in $\overline{\mathbf{R}}_+^k$ there corresponds a uniquely defined sample path of (T, J) with mutually disjoint I_n's, and conversely. On the other hand, it is possible (Jacod (1975)), and often most convenient in practice, to specify the law of (T, J) by specifying the corresponding $(\mathcal{F}_t)$-hazards $A(I)$, $I \in \mathcal{J}$. However, before considering such constructions explicitly we want to distinguish between different ways in which the hazard process can behave. We use the following three criteria:

(i) The first distinction is whether the compensators $A(I)$, $I \in \mathcal{J}$, are all continuous or not. In the former case we use the code *con*, whereas if discontinuities are allowed we write *dis*. Note that the latter class comprises all compensators; the *con*-class corresponds to the model where, as in the Poisson process, the exact failure times cannot be predicted with positive probability from the preceding history. A further subclass of *con* is formed by those compensators which are absolutely continuous, admitting the representation

$$A_t(I) = \int_0^t \lambda_s(I) ds.$$

The process $\lambda(I) = \{\lambda_t(I) : t \geq 0\}$ can then be viewed as the $(\mathcal{F}_t)$-based I-specific intensity (or *hazard rate*).

(ii) The second distinction we make is whether the parts in the device fail always one by one, or whether simultaneous failures of two or more parts are possible (e.g., because of a common cause). In the former case only the compensators $A(\{i\})$, $1 \leq i \leq k$, are not identically zero. We use the code *sin* for this class, and *mul* for the general class.

(iii) In the *dis*-class where the compensators can jump, it is still relevant to distinguish, mainly for technical reasons, whether two or more of the $A(I)$'s can jump at the same time epoch. If not, we use the code *sep*, denoting the general class by *com*.

Combining (i)–(iii), we have the following six classes of distributions of $\mathbf{S}$: *(con, sin)*, *(con, mul)*, *(dis, sin, sep)*, *(dis, sin, com)*, *(dis, mul, sep)*, and *(dis, mul, com)*. The last class contains all distributions.

We now want to formalize the idea of information, in the form of observing the part failures, which was mentioned at the beginning. We call a finite subset H of $\mathbf{R}_+ \times \mathcal{J}$ a *history set* if it is such that $(t, I) \in H$ and $(t, I') \in H$ imply $I = I'$ (Norros (1985)). We denote by $\mathbf{H}$ the space of all history sets endowed with the Borel σ-field $\mathcal{H}$ generated by a natural topology. We denote by $\mathcal{P}(\Omega)$ (resp. $\mathcal{P}(\mathcal{H})$) the set of probability measures on $(\Omega, \mathcal{F}_\infty)$ (resp. on $(\mathbf{H}, \mathcal{H})$) equipped with the topology of weak convergence of measures.

For history sets H we define the operations $c_t()$, $c()$ and $d_t()$ as follows:

$$\begin{aligned} c_t(H) &= \bigcup\{I : \exists s \leq t : (s, I) \in H\}, \\ c(H) &= \bigcup\{I : \exists t : (t, I) \in H\}, \\ d_t(H) &= \{(s, I) \in H : s \leq t\}. \end{aligned}$$

Clearly $c_t = c \circ d_t$. We introduce a partial order of history sets in the following way: $H \leq H'$ if $c_t(H) \supseteq c_t(H')$ for all $t \geq 0$. In other words, $H \leq H'$ if component failures described in H occur earlier (= not later) than the corresponding failures in H'.

The $\mathbf{H}$-valued pre-t histories of (T, J) correspond to the *history process* $\{H_t;\ t \geq 0\}$, defined by

$$H_t(\omega) = \{(T_n(\omega), J_n(\omega)) :\ T_n(\omega) \leq t\},$$

and its left continuous version defined by

$$H_{t-}(\omega) = \{(T_n(\omega), X_n(\omega)) :\ T_n(\omega) < t\}.$$

It then follows by (A2,T34) of Brémaud (1981) that for each $(\mathcal{F}_t)$-predictable process $Y = \{Y_t;\ t \geq 0\}$ there exists a non-random $\mathcal{R}_+ \otimes \mathcal{H}$-measurable function $(t, H) \mapsto Y^*(t \mid H)$ such that the process

$$Y_t^*(\omega) = Y^*(t \mid H_{t-}(\omega)),\ t \geq 0,$$

is indistinguishable from Y. For simplicity, we drop "*" from Y^* from now on.

In this way we can define the *compensator function family* of a multivariate life length distribution to be a family of functions $a_t(I \mid H)$ satisfying $a_t(I \mid H_t) = A_t(I)$ a.s. and the consistency condition

$$a_t(I \mid H) = a_t(I \mid d_{t-}(H)).$$

In particular, we can use this construction for defining the failure pattern specific intensities $\lambda(I)$ by using functions $(t, H) \mapsto \lambda_t(I \mid H)$ and the requirement that $\lambda_t(I)(\omega) = \lambda_t(I \mid H_{t-}(\omega))$. As mentioned before, these functions will also specify uniquely the distribution of (T, J) (and of $\mathbf{S}$).

EXAMPLE. Of particular interest is the *Markovian case* where the hazard rates depend on the past history only through the current configuration of parts down: whenever two history sets H and H' are such that $c_t(H) = c_t(H')$ for a considered time t, we have $\lambda_t(I \mid H) = \lambda_t(I \mid H')$. Then $c(H_t)$ is a Markov process on the state space $\mathcal{J}_0 = \mathcal{J} \cup \{\emptyset\}$. The so called Freund (1961) model and the multivariate exponential model of Marshall and Olkin (1967) are further special cases of this, with $\lambda_t(I \mid H)$ not depending on t. In our examples below we also restrict ourselves to the time homogeneous case, and write $\lambda(I \mid J)$ instead of $\lambda(I \mid H)$ when $c(H) = J$. Thus the second argument in $\lambda(I \mid J)$ is the set of failed components.

2. Monotonicity Conditions for the Prediction Processes. The notion of a *prediction process* was introduced by Knight (1975). Aldous (1981) developed a somewhat different approach which was applied in Norros (1985) and is followed here. We denote by $\mathcal{P}(\mathbf{R}_+^k)$ the space of all probability measures on the Borel sets of $\mathbf{R}_+^k$, endowed with the topology of weak convergence. $\mathcal{P}(\mathbf{R}_+^k)$ is also a Polish space. We often consider $\mathcal{P}(\mathbf{R}_+^k)$ as a partially ordered space, equipped with the usual stochastic order relation. (In later sections we consider also some other order relations defined on $\mathcal{P}(\mathbf{R}_+^k)$ or on a subset of it.)

THEOREM 2.1. *There exists a $\mathcal{P}(\mathbf{R}_+^k)$-valued cadlag process μ such that for any stopping time T, μ_T is a regular version of the conditional probability $P(\mathbf{S} \in \cdot \mid \mathcal{F}_T)$.*

The proof can be found in Aldous (1981), and it is reproduced in Norros (1985). In this paper we call the process μ of Theorem 2.1 simply *the prediction process.*

The prediction process μ_t bears in itself a complete description of the past, on the level of the made observations. Sometimes it is more convenient to have this degenerate part of the conditional distribution cut off from the prediction. Denote $(\mathbf{s} - t)^+ = ((s_1 - t)^+, \ldots, (s_k - t)^+)$. We define the *residual* *prediction process* ν_t as the measure valued cadlag process satisfying

$$\nu_T(\cdot) = P((\mathrm{S} - T)^+ \in \cdot \mid \mathcal{F}_T)$$

for every finite stopping time T. The residual prediction process is a time homogeneous strong Markov process.

Some interesting notions of aging and dependence can be defined as monotonicity conditions for the prediction processes. Arjas (1981a) generalized the class of IFR (Increasing Failure Rate) distributions in the following way (here we restrict ourselves to the case where $\mathcal{F}_t$ is the internal history):

DEFINITION 2.2. S is multivariate IFR (MIFR) if the residual prediction process ν_t is a decreasing process with respect to the stochastic order on the space $\mathcal{P}(\mathbf{R}_+^k)$.

Thus the MIFR distributions can be characterized by the following intuitive property: whatever happens in the internal history, the prediction of the remaining lifetimes becomes worse with increasing age.

This definition is meaningful for any class of distributions, but it actually implies that the compensators are continuous, except for possible final jumps of size 1. In fact, MIFR implies convexity of the compensator functions, although the converse does not hold (see Arjas (1981b)). Thus, all MIFR distributions lie in the class *(dis,mul,sep)*.

A positive dependence condition in the same spirit was introduced in Arjas and Norros (1984) (in the final form in Norros (1985)):

DEFINITION 2.3. S is *weakened by failures* (WBF) if the prediction process μ_t decreases (with respect to stochastic order) at failure times:

$$\mu_{S_i} \leq \mu_{S_i-} \quad \text{for all} \quad i.$$

It is (almost) evident that this definition is equivalent to the corresponding condition for the residual prediction process: $\nu_{S_i} \leq \nu_{S_i-}$ for all i. In other words, S is WBF if any part failure reduces, in the sense of stochastic order, the remaining life of the parts still alive. We then have the following trivial but interesting implication, an example of an aging condition implying positive dependence:

THEOREM 2.4. *MIFR implies WBF.*

WBF is a meaningful notion for completely general life length distributions, and the following result holds:

THEOREM 2.5. *WBF implies association.*

The proof can be found in Norros (1985) (it was first proven in Arjas and Norros (1984) with a slightly different definition of WBF). The proof is based on the compensator processes and the integral representation theorem for point process martingales.

EXAMPLE 2.6. Consider the time homogeneous Markovian case defined at the end of Section 1. For any function $\alpha : \mathcal{J}_0 \to \mathbf{R}_+$ denote by $\hat{\alpha}$ the measure on $\mathcal{J}_0$ with point mass function α. For two such functions, say α and β, we write $\hat{\alpha} \leq \hat{\beta}$ if $\hat{\alpha}(U) \leq \hat{\beta}(U)$ for each upper set (with respect to inclusion) $U \subseteq \mathcal{J}_0$.

For I, K and L in $\mathcal{J}_0$ such that $I \neq \emptyset$, $K \subseteq L$ and $I \cap L = \emptyset$ denote by

$$\lambda_L(I \mid K) = \sum_{J \subseteq L \setminus K} \lambda(I \cup J \mid K)$$

the total intensity, when the set of failed components is K, for the event that the next failure pattern is the union of I and possibly some subset of L. Let $\mathbf{S}$ and $\mathbf{T}$ be Markovian systems with the same set of parts, having intensity functions $\kappa(I \mid K)$ and $\lambda(I \mid K)$, respectively. The following results were proven in Norros (1985):

(i) If the implication

$$K \subseteq L \;\Rightarrow\; \hat{\kappa}_L(\cdot \mid K) \leq \hat{\lambda}(\cdot \mid L),$$

holds for all L and K, then $\mathbf{S} \geq_{\mathrm{st}} \mathbf{T}$.

(ii) If $K \subseteq L$ implies $\hat{\lambda}_L(\cdot \mid K) \leq \hat{\lambda}(\cdot \mid L)$, then $\mathbf{T}$ is MIFR (and, consequently, WBF).

(iii) If there are no multiple failures, a sufficient condition for WBF is

$$K \subseteq L, i \notin L \Rightarrow \; \lambda(\{i\} \mid K) \leq \lambda(\{i\} \mid L)$$

for all i, K and L.

(iv) The conditions in (i) and (ii) are also necessary if all transition intensities or, respectively, all intensities of single failures, are positive. (The latter case is not considered explicitly in Norros (1985), but the proof follows by the same technique as in Proposition 5.7 of that paper.)

We conclude this section by yet another dependence condition, called *strong supportivity*, which has not appeared in the literature before. Let $m_t(H)$ be a measurable function with values in $\mathcal{P}(\mathbf{R}_+^k)$, defined for non-negative t and history sets H, such that $m_t(H_t)$ is indistinguishable from μ_t. As in the definition of the compensator function family (Section 1), we require that $m_t(H)$ satisfies the consistency condition $m_t(H) = m_t(d_t(H))$. We define:

DEFINITION 2.7. S is called *strongly supportive* if $H \leq H'$ implies $m_t(H) \leq m_t(H')$ for all $t > 0$ and all history sets H and H'.

It is evident that a strongly supportive system (distribution) is weakened by failures. We return to this new notion again in Section 5.

3. The Compensator Representation of the Class *(dis,mul,sep).* In this section we present a generalization of the *compensator representation* of life length vectors, first studied in Norros (1986) and independently, with the name "multivariate hazard construction", in Shaked and Shanthikumar (1987b).

Consider first a (univariate) random lifetime S with a continuous distribution function F. Denote the corresponding survival function by $\overline{F} = 1 - F$. It is easily checked that $-\ln \overline{F}(S)$ has the 1-exponential distribution. In the class *(con,sin)* we have the following analogous multivariate result, first proven in Meyer (1971):

THEOREM 3.1. *Let* S *be in (con,sin), and denote*

$$X_i = A_{S_i}(\{i\}).$$

Then the X_i's are independent 1-exponential random variables.

If F is not continuous, A_S can not have an exponential distribution, of course. However, we can represent S as a function of a 1-exponential variable as follows.

Notation. If $f(t)$ is a right continuous increasing function with jumps ≤ 1, we write

$$-\int_0^t \ln(1 - df(s)) \underset{\text{def}}{=} f^c(t) + \sum_{s \leq t}(-\ln(1 - \Delta f(s))),$$

where $\Delta f(t) = f(t) - f(t-)$ and $f^c(t) = f(t) - \sum_{s \leq t} \Delta f(s)$. For continuous functions this notation is motivated by the "infinitesimal" formula $-\ln(1 - dx) = dx$.

Let now a_t be the compensator function of S, so that $A_t = a_{t \wedge S}$, and define

$$b_t = -\int_0^t \ln(1 - da_s),\ b_x^* = \inf\{t :\ b_t > x\}.$$

Let X have the 1-exponential distribution. Then b_X^* has the same distribution as S.

In fact, b_t is simply $-\ln P(S > t)$. Dropping the "minus logarithm" we would have represented S in the more familiar way as b_T^*, where T is uniformly distributed on $[0,1]$. However, we want to work with exponential random variables because of Theorem 3.1, and because we want to have the compensator functions as our starting point rather than the usual distribution functions.

Let us now turn to the multivariate case. Let $a_t(I \mid H)$ be the compensator function of a class *(dis,mul,sep)* distribution. Define

$$\begin{aligned} b_t(I \mid H) &= -\int_0^t \ln(1 - da_s(I \mid H)), \\ b_x^*(I \mid H) &= \inf\{t:\ b_t(I \mid H) > x\}. \end{aligned}$$

We call the functions $b_t(i \mid H)$ *b-functions.* As usually, we let $\inf \emptyset = \infty$. For arbitrary $\mathbf{x} \in \mathbf{R}_+^{\mathcal{J}}$, $\mathbf{x} = (x_J)$, we define inductively the mapping $\mathbf{x} \mapsto \Psi^*(\mathbf{x}) = \mathbf{s} = (s_1, \ldots, s_k)$:

$$\begin{aligned} &t_{n+1} = \inf\{b_{x_I}^*(I \mid \{(t_p, J_p):\ p = 1, \ldots, n\}):\ I \cap (J_1 \cup \cdots \cup J_n) = \emptyset\}; \\ &\text{if } t_{n+1} < \infty, \text{ then} \\ &J_{n+1} = \text{that } I \text{ at which the minimum is obtained, and} \\ &s_i = t_{n+1} \text{ for } i \in J_{n+1}. \end{aligned} \tag{3.2}$$

Since common jumps of the compensators are forbidden in the class *(dis, mul, sep)*, the mapping is uniquely determined for almost every $\mathbf{x}$.

We call the function Ψ^* the *compensator representation* of $\mathbf{S}$. The reason we use this term is, as is easy to see by the memoryless property of the exponential distribution, that $\Psi^*(\mathbf{X})$ has the same distribution as $\mathbf{S}$ if $\mathbf{X} = (X_J)_{J \in \mathcal{J}}$ is a vector of independent 1-exponential random variables.

The reasoning goes as follows. First, it is obvious that a copy of S can be generated proceeding in time inductively from one failure to the next, choosing always the next failure time and failure pattern according to the conditional distribution where the conditioning is based on the history up to the previous failure.

Second, at each step the conditional distribution of the time to the next failure and the next failure pattern can be produced as the minimum of *independent* "competing risks" (see Arjas and Greenwood (1981)). At this point, the assumption that the compensators do not have *common* jumps is crucial.

Third, the survival functions of these competing risks are exponential functions of minus the increments (with respect to the starting point) of the corresponding b-functions. Thus the "competing" failure times, from which the minimum is chosen, can be generated using independent 1-exponential random variables and the b-functions as shown above in the case of one component.

Fourth, the "fine point" of the construction is that *the same* exponential variables can be used through all steps. Indeed, we can think that at each step only the result, that is, the exponential giving the shortest time to the corresponding failure pattern, is revealed. As regards the others, it is known only that they are greater than certain values, namely those attained by the corresponding b-functions so far. But, by the memoryless property of the exponential distribution, this gives no information about the following step. Thus, the "unused" exponentials are "as new" at the beginning of each step.

In Norros (1986), a condition was given which implied the monotonicity of Ψ^* in the class of *(con,sin)* distributions. The next example shows that the representation need not be monotone if multiple failures are allowed, even when there is a strong positive dependence between the coordinates of $\mathbf{S}$.

EXAMPLE 3.3. Consider a Markovian system with $k = 2$, and let the intensities be

$$\lambda(\{1\}) \mid \emptyset) = \lambda(\{2\} \mid \emptyset) = \lambda(\{1,2\} \mid \emptyset) = 1,$$
$$\lambda(\{1\} \mid \{2\}) = \lambda(\{2\} \mid \{1\}) = 2.$$

The system is obviously WBF. But take $x_1 = 1$, $x_2 = 5$, $x_{12} = 2$, and $x'_1 = 3$, $x'_2 = 5$, $x'_{12} = 2$. Then $\mathbf{x} < \mathbf{x}'$, but $\Psi^*(\mathbf{x}) = (1,3)$ and $\Psi^*(\mathbf{x}') = (2,2)$. Thus Ψ^* is not monotone.

The simplicity of this example indicates that, at least in the context of stochastic order, the compensator representation is a useful notion only for distributions without multiple failures.

We now extend the notion of *supportivity*, introduced in Norros (1986), to the case where multiple failures and simultaneous jumps of compensators are not allowed but where the compensators need not be continuous.

DEFINITION 3.4. A distribution of the class *(dis,sin,sep)* is called *supportive* if for all t

$$H \le H',\ i \notin c(H) \implies b_t(\{i\} \mid H) \ge b_t(\{i\} \mid H').$$

The usefulness of the notion of supportivity is based on the one hand on its practical verifiability (compared, for example, to verifying the WBF property, or association) and, on the other hand, on the following fact:

THEOREM 3.5. *If* $\mathbf{S}$ *is supportive, then* Ψ^* *is componentwise increasing.*

PROOF. The condition appearing in Definition 3.4 is the same as in Norros (1986) except that it is formulated for b-functions instead of compensator functions, and the same holds for the construction (3.2). It follows that the proof of the corresponding Proposition 3.3 in Norros (1986) is applicable when compensator functions are replaced by b-functions. ∎

Counterexample 4.9 of Shaked and Shanthikumar (1987b) shows that the implication of Theorem 3.5 is strict (even in the class *(con,sin)*).

Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a set of independent 1-exponential random variables, and let $\Psi^*(\mathbf{X})$ be the compensator representation of a random vector of class *(dis, sin, sep)*. Since a set of independent random variables is always associated, Theorem 3.5 has the following corollary.

COROLLARY 3.6. *If* S *is supportive, then it is associated.*

We show in Section 5 that supportivity implies even a stronger positive dependence condition, WBF.

EXAMPLE 3.7. Consider again the Markovian case and assume that only single component failures are possible. It is shown in Norros (1986) that the system is supportive if and only if $\lambda(\{i\} \mid K)$ is increasing in K. This is the same condition as the sufficient (and, at least when all single failure intensities are positive, necessary) condition for WBF, mentioned in Example 2.6.

4. Conditions for Stochastic Order. Let S have a distribution in the class *(dis,sin,sep)*. We now extend to this class the definition of cumulative hazard ordering of Shaked and Shanthikumar (1990, Section 2).

DEFINITION 4.1. Let S and $\overline{\mathrm{S}}$ be two life length vectors with the same number of components and with distributions in the class *(dis,sin,sep)*. Denote the corresponding b-functions by $b_t(\{i\} \mid H)$ and $\bar{b}_t(\{i\} \mid H)$. We say that S *is less than* $\overline{\mathrm{S}}$ *in the cumulative hazard ordering*, and denote this by $\mathrm{S} \leq_{ch} \overline{\mathrm{S}}$, if

$$H \leq H' \text{ and } i \notin c(H) \;\Rightarrow\; b_t(\{i\}) \mid H) \geq \bar{b}_t(\{i\} \mid H')$$

for any i, H and H'.

The relation $\leq_h$ is transitive, but it is obviously reflexive only in the class of supportive distributions.

The following result was proven by Shaked and Shanthikumar (1987a) (in the absolutely continuous case):

THEOREM 4.2. *If* $\mathrm{S} \leq_{ch} \overline{\mathrm{S}}$, *then* $\mathrm{S} \leq_{st} \overline{\mathrm{S}}$.

PROOF. Using the construction 3.2, the S_i's and $\overline{S}_i$'s can be generated inductively by using the same exponential variables $X_i = X_{\{i\}}$. It is easy to see that then $S_i \leq \overline{S}_i$ for all i. ∎

This result may be useful in proving that certain non-continuous multivariate distributions are stochastically ordered. Note that although the cumulative hazard ordering is a proper (reflexive) order relation only in the class of supportive distributions, supportivity is not required in Theorem 4.2.

A third order relation for multivariate life length distributions is the TP_2 ordering of Karlin and Rinott (1980). Shaked and Shanthikumar (1990) chose this ordering as the multivariate generalization of the likelihood ratio ordering and denoted it by $\leq_{lr}$. The definition is as follows.

DEFINITION 4.3. Assume that $\mathbf{S}$ and $\mathbf{T}$ have density functions f and g, respectively. We say that $\mathbf{S}$ *is less than* $\mathbf{T}$ *in the likelihood ordering*, and denote this by $\mathbf{S} \leq_{lr} \mathbf{T}$, if

$$f(\mathbf{s})g(\mathbf{t}) \leq f(\mathbf{s} \wedge \mathbf{t})g(\mathbf{s} \vee \mathbf{t})$$

for all $\mathbf{s}, \mathbf{t} \in \mathbf{R}_{+}^{k}$. If some of the S_i's are *identically* zero, we set the same definition for the marginal densities of the positive components.

This relation is reflexive exactly in the class of the so called multivariate TP_2 (MTP_2) distributions. Shaked and Shanthikumar (1990) proved:

THEOREM 4.4. *If* $\mathbf{S} \leq_{lr} \mathbf{T}$, *then* $\mathbf{S} \leq_{ch} \mathbf{T}$.

Since the likelihood ratio ordering is defined only for absolutely continuous distributions, our extension of the definition of cumulative hazard ordering does not generalize this theorem.

5. Relationships Between Positive Dependence Notions. Consider a distribution of class *(dis,sin,sep)*. Assume that $a_t(\{i\} \mid H)$ is supportive. Consider the conditional distributions ν_{S_i} and ν_{S_i-}. It is obvious that their b-function families are

$$(t, H) \mapsto b_{S_i+t}(\{j\} \mid (H + S_i) \cup H_{S_i}) - b_{S_i}(\{j\} \mid H_{S_i})$$

and

$$(t, H) \mapsto b_{S_i+t}(\{j\} \mid (H + S_i) \cup H_{S_i-}) - b_{S_i}(\{j\} \mid H_{S_i-}),$$

respectively. (The notation is: $H + t = \{(s + t, \{i\}) : (s, \{i\}) \in H\}$.) Now, Theorem 4.2 implies that $\nu_{S_i} \leq \nu_{S_i-}$, so that we have the following result.

THEOREM 5.1.*If a class (dis,sin,sep) distribution is supportive, then it is WBF.*

The next example shows that the implication is strict.

EXAMPLE 5.2. Consider a system with two parts. Let S_1 have constant intensity 1 (it follows the 1-exponential distribution), and let the intensity of

S_2 be defined as follows:

$$\lambda_t(\{2\} \mid \emptyset) = 0,$$
$$\lambda_t(\{2\} \mid \{(s,1)\}) = s \vee 1.$$

Then $a_t(\{2\} \mid (s,1)) = (s \vee 1)(t-s)^+$, which is increasing in s when $s \in (1, t/2)$, so that the system is not supportive. On the other hand, we have for all s and x such that $s < x$

$$\frac{P(S_2 > x \mid S_1 > s)}{P(S_2 > x \mid S_1 = s)} > \frac{P(S_1 > x \mid S_1 > s)}{P(S_2 > x \mid S_1 = s)} = e^{(s-1)^+(x-s)} \geq 1,$$

which is enough to show that the system is WBF.

Shaked and Shanthikumar (1990) showed that MTP_2 implies supportivity. Their argument consists of the following two steps:

(i) MTP_2 implies that the prediction process decreases in likelihood ratio ordering at failure times;

(ii) Supportivity is equivalent to the following: the prediction process decreases at failure times w.r.t. the cumulative hazard ordering.

Their idea is to recast the definitions of MTP_2 and supportivity into the form of WBF, where only jumps of the prediction process at failure times are considered. Then, the only difference of the notions is the order relation that is required between the predictions before and after the jumps. Note that in (i) the implication is only one way. In (ii), the equivalence is rather trivial.

The following two theorems offer an alternative route from MTP_2 to WBF.

THEOREM 5.3. *If* S *is* MTP_2*, then it is strongly supportive.*

PROOF. Formula (3.6) in Shaked and Shanthikumar (1990) is shown to follow from $\mathbf{X} \leq_{lr} \mathbf{Y}$. Choosing $\mathbf{X} = \mathbf{Y}$ and taking into account that $\leq_{lr}$ implies stochastic order we get the assertion of the theorem. ∎

Recall from Section 2 that strong supportivity implies the WBF property. Actually the following conclusion holds:

THEOREM 5.4. *Let* S *be in the class (con,sin). If* S *is strongly supportive, then it is supportive.*

PROOF. Fix i and t. We denote $U_t = \{\mathbf{s} : s_i \leq t\}$. U_t is a lower set. Let

$$\phi_t(H) = \begin{cases} 0 & \text{if there is an } s \leq t \text{ such that } (s, \{i\}) \in H, \\ 1 & \text{otherwise.} \end{cases}$$

Let $t_{jn} = j/2^n$. Identifying ω with $H = H_\infty$, we may write

$$A_t^{(n)}(i)(H) = \sum_{j:\ t_{jn}<t} P(S_i \in (t_{jn}, t_{j+1,n}] \mid \mathcal{F}_{t_{jn}})$$
$$= \sum_{j:\ t_{jn}<t} \phi_{t_{jn}}(H) \cdot m_{t_{jn}}(H)(U_{t_{j+1,n}}),$$

where $m_t(H)$ is the measure-valued function appearing in Definition 2.7. It is well known that $A_t^{(n)}(H)$ converges to $A_t(H)$ stochastically (Murali-Rao (1969)), and taking a subsequence we may assume that the convergence holds for almost every H. Taking H and H' from this set and satisfying $H \leq H'$, $i \notin c_t(H)$, gives the assertion. Indeed, $m_s(H)(U_t)$ is decreasing in H, and the condition $i \notin c_t(H)$ implies that $\phi_s(H) = 1$ for $s \leq t$. Thus an inequality holds in each term of the sum, and the desired inequality between compensator functions is obtained at the limit. ∎

Our next example shows that the implication in Theorem 5.4 is strict.

EXAMPLE 5.5. Consider a system with two parts. Let S_1 have constant intensity 1 (it follows the 1-exponential distribution), and let the intensity of S_2 be defined as follows:

$$\lambda_t(\{2\} \mid \emptyset) = 2 \cdot 1_{(2,3]}(t) + 1_{(3,\infty)}(t)$$
$$\lambda_t(\{2\} \mid \{(s,1)\}) = \begin{cases} \lambda_t(\{2\} \mid \emptyset), & s \geq 1, \\ 1_{(1,\infty)}(t), & s < 1. \end{cases}$$

It is easy to check that $a_t(\{2\} \mid \{(s,1)\})$ is decreasing in s, so that the system is obviously supportive. However,

$$m_2(\{(0.5,1)\})(I\!R \times (3,\infty)) = e^{-1} > e^{-2} = m_2(\{(1.5,1)\})(I\!R \times (3,\infty)),$$

so that the system is not strongly supportive.

6. Histories and Information. In the above analysis of aging and dependence we have restricted ourselves to the history $(\mathcal{F}_t)$ which is generated by the part failures. This may have led the reader to think that this choice is in some sense the only possible. Rather the opposite is true, however. In the present framework a history represents cumulating knowledge or information, and the resulting conditional distributions are viewed simply as quantitative expressions of the remaining uncertainty regarding the exact values of the part life lengths. The histories can be arbitrary (subject to certain technical conditions), different histories then giving rise to different compensators and prediction processes. More, or less, detailed knowledge regarding the events in the past typically changes the prediction made about the future, and thereby also has the potential of influencing the ways in which aging and dependence can be characterized.

Obviously every elapsed time unit of a part's life brings the part closer to its failure, by the same amount. However, when the time of failure is not known, the predicted remaining life can become even longer, in the sense of stochastic order, as the age increases. In practice this can be often explained by the heterogeneity of the population of parts, and a corresponding selection mechanism which tends to leave the strong parts alive while the weak ones fail.

In order to keep the following discussion in the simplest possible terms we consider only two particular levels of information, called *part level* and *system level*. The former is the history $(\mathcal{F}_t)$ which was considered above. To define the latter, suppose that the k parts give rise to l (monotone) systems, say $\phi_1, \dots, \phi_l$, with respective life lengths $\tau_1, \dots, \tau_l$. It is well known (e.g. Barlow and Proschan (1975)) that each τ_j can be expressed as a simple increasing function of $\mathbf{S}$. Note that we are here not assuming any form of independence: originally the part life lengths can be statistically dependent, the systems may have parts in common, and two or more systems can fail at the same time, possibly of a common cause.

We can now consider the counting processes corresponding to system failures, defined simply by $N_t(\phi_j) = 1_{\{\tau_j \le t\}}$, $t \ge 0$, and the corresponding system level history $(\mathcal{G}_t)$ where

$$\mathcal{G}_t = \sigma\{N_s(\phi_j);\ 1 \le j \le l,\ s \le t\}.$$

(Again $\mathcal{G}_0$ should contain all nullsets of $\mathcal{G}_\infty = \bigvee_{t \ge 0} \mathcal{G}_t$.)

It is easy to see that the inclusion $\mathcal{G}_t \subseteq \mathcal{F}_t$ holds for all $t \ge 0$. Intuitively this corresponds to the property that the pre-t behavior of the parts determines completely that of the systems.

Now suppose that we have been able to characterize the behavior of the k parts (i.e., the distribution of $\mathbf{S}$) in terms of the aging and dependence concepts introduced earlier. We consider the following two questions:

(1) Do the corresponding properties hold when the vector $\mathbf{S}$ is replaced by the vector $\boldsymbol{\tau}$? In other words, we want to characterize the behavior of the systems instead of parts.

(2) Do such properties hold when the part level knowledge $(\mathcal{F}_t)$ is replaced by $(\mathcal{G}_t)$?

The answer to the first question turns out to be positive, to the second generally negative.

That the first question gets a positive answer is a simple consequence of the fact that all increasing functions of τ are also increasing functions of S. For example, if S is IFR (resp. WBF, or (strongly) supportive) relative to the history $(\mathcal{F}_t)$, τ has the same properties. In particular, the IFR property is "closed under formation of monotone systems", provided that the part level history $(\mathcal{F}_t)$ is used throughout.

On the other hand, the situation is different if the history is changed as well. The well-known example of two parts with independent exponential life lengths in parallel, showing that the classical notion of IFR is not closed under formation of monotone systems (Barlow and Proschan (1975)), serves here as a counterexample as well: it proves that IFR-property relative to $(\mathcal{F}_t)$ does not imply IFR-property relative to $(\mathcal{G}_t)$. This follows from the obvious facts: (a) if S_1 and S_2 are independent exponential part life lengths, then $\mathbf{S} = (S_1, S_2)$ is IFR relative to $(\mathcal{F}_t)$, and (b) a univariate life length is IFR in the classical sense if and only if it is IFR relative to its internal history.

We remark that certain aging properties which are weaker than IFR can remain valid when the history is changed. An easy example is the multivariate NBU (New Better than Used) concept introduced in Arjas (1981a).

A fairly elaborate counterexample is needed to show that the WBF property can be lost in a history change (Arjas and Norros (1991)). A more systematic analysis of the effects of history change is outside the scope of this review; we refer to Arjas (1991), Arjas and Norros (1991) and Arjas, Haara and Norros (1989) for such results.

REFERENCES

ALDOUS, D. (1981) Weak convergence and the general theory of processes. Incomplete draft of a monograph.

ARJAS, E. (1981a). A stochastic process approach to multivariate reliability systems: notions based on conditional stochastic order. *Math. Oper. Res.* **6**, 263–276.

ARJAS, E. (1981b). The failure and hazard processes in multivariate reliability systems. *Math. Oper. Res.* **6**, 551–562.

ARJAS, E. (1989). Survival models and martingale dynamics. *Scand. J. Statist.* **16**, 177–225. (With discussion.)

ARJAS, E. and GREENWOOD, P. (1981). Competing risks and independent minima: a marked point process approach. *Adv. in Appl. Probab.* **13**, 669–680.

ARJAS, E., HAARA, P., and NORROS, I. (1991). On filtering the histories of a partially observed marked point process. *Stoch. Proc. Appl.* To appear.

ARJAS, E. and NORROS, I. (1984). Life lengths and association: a dynamic approach. *Math. Oper. Res.* **9**, 151–158.

ARJAS, E. and NORROS, I. (1988). A martingale approach to reliability theory: on the role of filtration in the model specification. *Proc. of the first Finnish-Soviet Symp. on Probab. Theor. and Math. Statist., Ann. Acad. Sci. Fennicae, Series A. I. Math.*, Vol. 13, 183–189.

ARJAS, E. and NORROS, I. (1989). Change of life distribution via a hazard transformation: an inequality with application to minimal repair. *Math. Oper. Res.* **14**, 355–361.

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

BRÉMAUD, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer Verlag, New York.

JACOD, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **34**, 235–253.

KARR, A.F. (1986). *Point Processes and Their Statistical Inference.* Dekker, New York.

KNIGHT, F. (1975). A predictive view of continuous time processes. *Ann. Probab.* **3**, 573–596.

LIPTSER, R.N. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes II. Applications.* Springer Verlag, New York.

MEYER, P.A. (1971). *Demonstration Simplifiée d'un Theorème de Knight.* Séminaire de Probabilités V, Springer Verlag, Berlin.

MURALI-RAO, K. (1969). On decomposition theorems of Meyer. *Math. Scand.* **24**, 66–78.

NORROS, I. (1985). Systems weakened by failures. *Stoch. Proc. Appl.* **20**, 181–196.

NORROS, I. (1986). A compensator representation of multivariate life length distributions, with applications. *Scand. J. Statist.* **13**, 99–112.

SHAKED, M. and SHANTHIKUMAR, J. G. (1987a). Multivariate hazard rates and stochastic ordering. *Adv. in Appl. Probab.* **19**, 123–137.

SHAKED, M. and SHANTHIKUMAR, J. G. (1987b). The multivariate hazard construction. *Stoch. Proc. Appl.* **24**, 241–258.

SHAKED, M. and SHANTHIKUMAR, J. G. (1990). Multivariate stochastic orderings and positive dependence in reliability theory. *Math. Oper. Res.* **15**, 545–552.

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
UNIVERSITY OF OULU
LINNANMAA
SF-90570 OULU, FINLAND

TECHNICAL RESEARCH CENTRE OF FINLAND
TELECOMMUNICATIONS LABORATORY
SF-02150 ESPOO, FINLAND