Stochastic Processes and their Applications 20 (1985) 181-196
North-Holland

SYSTEMS WEAKENED BY FAILURES

Ilkka NORROS

Department of Mathematics, University of Helsinki, Hallituskatu 15, 00100 Helsinki 10, Finland

Received 3 January 1984
Revised 8 February 1985

The ideas of a dynamic approach to the analysis of multivariate life length distributions, introduced in Arjas (1981a) and Arjas and Norros (1984), are developed further. Basic definitions are given in terms of prediction processes. Properties of martingales jumping downwards at failure times are studied. Finally, the special case of a general multivariate exponential distribution is considered.

prediction process * martingale * association * stochastic order * marked point process * multivariate exponential * weakened by failures

0. Introduction

Consider a system consisting of k components and starting its operation at time 0. Suppose that the components have finite positive life lengths $S_i > 0$, $i = 1, \ldots, k$, and failed components are not replaced by new ones. We allow two or more S_i's to be equal with positive probability. For example, several components might be destroyed simultaneously by an external shock.

Analysing systems of this kind is equivalent to studying probability distributions on the interior of the positive orthant $\mathbb{R}_+^k$. Traditionally, conditions for mutual dependence of the life lengths S_i have been formulated in terms of the distribution function $F(s_1, \ldots, s_k) = P(S_1 \leq s_1, \ldots, S_k \leq s_k)$. Arjas [2] suggested a different, 'dynamical' approach, where the failure times S_i were considered as events in a stochastic process, the *failure process*. This makes it possible to apply notions and results from the theory of stochastic processes. Especially, it becomes possible to study *causal* effects the destruction of one or more components has on the residual life distribution of the others.

In [4], Arjas and the author introduced the concept of a *system weakened by failures*. In this paper, this notion is analyzed further and simple conditions for it are given in the case of a generalized multivariate exponential distribution, where the failure process is a time homogeneous Markov process.

Notation. We write $S = (S_1, \ldots, S_k)$. For any $t \geq 0$, denote $X_t = \{i: S_i > t,\ i = 1, \ldots, k\}$. Then $X = (X_t)_{t \geq 0}$ is a stochastic process with values in the set $\mathcal{I}_0 = \{I: I \subseteq \{1, \ldots, k\}\}$. We consider $\mathcal{I}_0$ as partially ordered through inclusion.

As the underlying probability space we take the path space $\Omega = \{x\colon x$ is a decreasing, right continuous function $\mathbb{R}_+ \to \mathcal{T}_0$ with $x(\infty) = \emptyset\}$. Let X_t be the coordinate process on Ω, that is, $X_t(x) = x(t)$. The map defined above which connects **S** with the process X is one-to-one from $\mathbb{R}_+^k$ onto Ω.

Let $(\mathcal{F}_t)_{t\geq 0}$ be the natural history (filtration) on Ω, that is, $\mathcal{F}_t = \sigma(X_s\colon s \leq t)$. When we have a probability measure P on Ω, we suppose that the σ-fields $\mathcal{F}_t$ are completed with all P-null sets. Dellacherie's 'usual conditions' are then satisfied, that is, the history $(\mathcal{F}_t)$ is right continuous and $\mathcal{F}_0$ contains all null sets.

Partial order relations. If s, $t \in \mathbb{R}_+^k$, we write $s \leq t$ if $s_i \leq t_i$ for all i. If $x, y \in \Omega$, we write $x \leq y$ if $x(t) \subseteq y(t)$ for all t. Then the bijection $s \to x$ is also an order isomorphism.

We compare probability measures on $\mathbb{R}_+^k$ through the *stochastic order*: we call P smaller than Q and write $P \leq Q$ if $\int f\,\mathrm{d}P \leq \int f\,\mathrm{d}Q$ for all increasing bounded functions $f\colon \mathbb{R}_+^k \to \mathbb{R}$. (We tacitly assume also the Borel measurability of the functions f.) It can be shown (see, e.g., Arjas [2]) that the class of test functions f in the definition above can be reduced in various ways without altering its content. Sufficient classes of test functions are, among others,

(i) increasing indicators,

(ii) indicators of sets of the form $\bigcup_{i=1}^{n} \{s\colon s_1 > q_j^i, \ldots, s_k > q_j^i\}$, where all the q_j^i's are rational,

(iii) continuous, increasing and bounded functions.

Note that the class (ii) is denumerable. Sets with increasing indicators are called *upper*, their complements *lower sets.*

A probability measure P on $\mathbb{R}_+^k$ is called *associated* if, for all increasing bounded f and g, $\mathrm{Cov}(f, g) = \int fg\,\mathrm{d}P - \int f\,\mathrm{d}P \int g\,\mathrm{d}P \geq 0$. Here again the class of test functions f and g can be reduced to any of the classes (i)–(iii) above. (For a systematic treatment of association see, e.g., Barlow and Proschan [5].)

For other partially ordered spaces, the concepts of stochastic order and association are defined analogously.

1. The prediction process and the residual prediction process

In the 'dynamical approach', conditional distributions with respect to the past of the process at time t (that is, w.r.t. $\mathcal{F}_t$) play a fundamental role. For a while, let t be fixed. It is well known that there exists a regular conditional distribution (r.c.d.) μ_t of S given $\mathcal{F}_t\colon \mu_t(A) = P(S \in A \mid \mathcal{F}_t)$.

If E is a topological space, denote by $\mathcal{P}(E)$ the set of all probability measures on the Borel sets of E with the topology of weak convergence of measures. It is easy to see that the r.c.d. $\mu_t(\cdot)$, considered as a function $\Omega \to \mathcal{P}(\mathbb{R}_+^k)$, is a $\mathcal{P}(\mathbb{R}_+^k)$-valued random variable and for all $h \in C(\mathbb{R}_+^k)$, $\int h\,\mathrm{d}\mu_t = E[h(S) \mid \mathcal{F}_t]$ a.s.

Consider again the failure process (X_t). For each time t we have the prognosis μ_t for S (and, equivalently, for $(X_t)_{t\geq 0}$). Thus we are led to study μ_t as a $\mathcal{P}(\mathbb{R}^k_+)$-valued stochastic process. The following important theorem from Aldous [1] shows that this process has a regular modification. To formulate the general results, we replace S by a more general random variable with values in a Polish space.

1.1. Theorem. *Let E be a Polish space and let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration on the probability space $(\Omega, \mathcal{F}, P)$ satisfying Dellacherie's 'usual conditions'. Let X be an E-valued random variable. Then there exists a $\mathcal{P}(E)$-valued cadlag process $\mu = (\mu_t)_{t\in[0,\infty]}$ such that for any $t \in [0, \infty]$, μ_t is a r.c.d. of X given $\mathcal{F}_t$. This process is uniquely determined up to indistinguishability. ('Cadlag' means: right continuous with left limits.)*

Proof. 1°. Let K_n be a sequence of compact subsets of E such that

$$P(X \notin K_n) \leq 2^{-2n} \quad \text{for all } n.$$

Let h_n be a sequence of functions in $C(E)$ such that for all n, $0 \leq h_n \leq 1$ and the h_n's separate the points of $\mathcal{P}(E)$. The following lemma is a straightforward application of Prokhorov's theorem:

Lemma. *Let $\mu_j \in \mathcal{P}(E)$ and suppose*

(i) *for all n the limit $\lim_j \int h_n \, \mathrm{d}\mu_j = a_n$ exists;*

(ii) $\lim_n \overline{\lim}_j \mu_j(K_n^c) = 0$.

Then there is a measure $\mu \in \mathcal{P}(E)$ such that $\mu_j \to \mu$ and $\int h_n \, \mathrm{d}\mu = a_n$ for all n.

2°. For every nonnegative rational r, let σ_r be a regular conditional distribution of X given $\mathcal{F}_r$. Then σ_r can also be considered as a $\mathcal{P}(E)$-valued random variable. Let M_t^n be a cadlag version of the martingale $P(X \notin K_n \mid \mathcal{F}_t)$. By a well known martingale maximal inequality,

$$P\left(\sup_t M_t^n \geq 2^{-n}\right) \leq 2^n P(X \notin K_n) \leq 2^{-n}.$$

Denote $N_1 = \{\lim_n \sup_t M_t^n = 0\}^c$. Then $P(N_1) = 0$.

Then let H_t^n be a cadlag version of the martingale $E[h_n(X) \mid \mathcal{F}_t]$. Denote

$$N_2 = \left\{\exists r \in \mathbb{Q}_+,\ n \in \mathbb{N} \text{ such that } \sigma_r(K_n^c) \neq M_r^n \text{ or } \int h_n \, \mathrm{d}\sigma_r \neq H_r^n\right\}.$$

Then $P(N_2) = 0$.

3°. Now fix $t \in \mathbb{R}_+$ and let $r_j \downarrow t$, $r_j \in \mathbb{Q}$, $r_j > t$. In the complement of $N_1 \cup N_2$ the sequence $\sigma_{r_j}(\omega)$ satisfies the conditions of the lemma in 1°. Thus there exists a limit measure $\mu_{t+}(\omega)$ satisfying

$$\int h_n \, \mathrm{d}\mu_{t+}(\omega) = H_{t+}^n(\omega) = H_t^n(\omega) \quad \text{for each } n.$$

Correspondingly, a left limit measure $\mu_{t-}(\omega)$ is defined in the complement of $N_1 \cup N_2$.

Choose $\mu_t(\omega) = \mu_{t+}(\omega)$ on $(N_1 \cup N_2)^c$ and $\mu_t(\omega) =$ constant on $N_1 \cup N_2$. It is easy to see that for each t, μ_t is a r.c.d. of X given $\mathscr{F}_t$.

4°. The uniqueness of μ follows from the uniqueness (mod P) of the r.c.d.'s $P(X \in \cdot \mid \mathscr{F}_t)$ and from the cadlag property. □

1.2. Definition. The process μ in Theorem 1.1 is called the *prediction process* of the random variable X with respect to the history $(\mathscr{F}_t)$.

The idea of a prediction process was introduced by Knight [13]. Aldous [1] developed a somewhat different approach, which we follow here. In the next proposition we have listed some basic properties of the prediction process:

1.3. Proposition. *Under the assumptions of Theorem* 1.1, *the prediction process* μ *has the following properties*:

(i) *For any* $(\mathscr{F}_t)$*-stopping time* T, μ_T *is a regular conditional distribution of* X *given* $\mathscr{F}_T$.

(ii) *For any positive predictable stopping time* T, μ_{T-} *is a r.c.d. of* X *given* $\mathscr{F}_{T-}$.

(iii) *Denote* $\mu_t^- = \mu_{t-}$. *Then the process* μ^- *is a predictable* $\mathscr{P}(E)$*-valued process such that for any predictable stopping time* T, μ_T^- *is a r.c.d. of* X *given* $\mathscr{F}_{T-}$.

(iv) *If* $\mathscr{F}_\infty \subseteq \sigma(X)$ *(modulo null sets), then* μ *is a (nonhomogeneous) strong Markov process.*

Proof. (i) is easy to prove by approximating T from above by stopping times taking only countably many values. (ii) follows from (i) and the martingale convergence theorem, and (iii) is an immediate corollary to (ii).

In order to show (iv), let T be any finite stopping time, $t > 0$ and B any measurable subset of $\mathscr{P}(E)$. Then there exists a measurable function $f: E \to \mathbb{R}$ such that $1_{\{\mu_{T+t} \in B\}} = f(X)$ a.s. Now $P(\mu_{T+t} \in B \mid \mathscr{F}_T) = \int f(x)\mu_T(\mathrm{d}x)$. Thus $P(\mu_{T+t} \in B \mid F_T)$ is $\sigma(\mu_T)$-measurable. □

Note that in the case of the failure process as described above, the condition in (iv) is satisfied since the fundamental σ-algebra is that generated by $\boldsymbol{S}$.

If $h \in C(E)$, then the process $\int h \,\mathrm{d}\mu_t$ is a cadlag version of the martingale $E[h(X) \mid \mathscr{F}_t]$. This follows directly from the definition of weak convergence, since μ_t is cadlag w.r.t. the weak topology. On the other hand, it is clear that noncontinuous functions h can be found such that there exist paths of the process $\int h \,\mathrm{d}\mu_t$ that are not cadlag. However, as Aldous mentioned in [1], this can happen only on a null set:

1.4. Proposition. *For any bounded measurable function* $h: E \to \mathbb{R}$, *the process* $Y_t = \int h \,\mathrm{d}\mu_t$ *is a martingale with a.s. cadlag paths. Moreover,* Y_{t-} *is indistinguishable from the process* $\int h \,\mathrm{d}\mu_{t-}$.

Proof. Denote by M a cadlag version of the martingale $E[h(X)\,|\,\mathscr{F}_t]$. By Proposition 1.3(i), for any finite stopping time T, $Y_T = M_T$ a.s. But μ is optional, and the map $\rho \to \int h\,\mathrm{d}\rho$ from $\mathscr{P}(E)$ to $\mathbb{R}$ is measurable, so that Y is optional. Since M is also optional, it follows that Y and M are indistinguishable.

The second assertion is proved analogously. The process (μ_{t-}) is left continuous and thus predictable. It follows that the process $(\int h\,\mathrm{d}\mu_{t-})$ is predictable. On the other hand, it is well known that (Y_{t-}) is the predictable projection of Y. Thus the indistinguishability of (Y_{t-}) and $(\int h\,\mathrm{d}\mu_{t-})$ follows from Proposition 1.3(ii). □

Now we turn back to the life length vector $\boldsymbol{S}$ and the failure process (X_t). Let μ be the prediction process of $\boldsymbol{S}$.

Remark. As an alternative, we could study the prediction process of X, equipping Ω with the Skorohod D-metric. This would be equivalent to considering S with a slightly finer topology on $\mathbb{R}_+^k$. For the aims of this paper, however, the normal topology is more convenient.

For a given positive t, the information contained in μ_t can be split into two parts: the information up to time t, which is exact, and the future, for which only a distribution is known. In some contexts it is better to have the past 'cut off' from μ_t. Therefore we define also a 'residual prediction process', which consists of the 'future part' of μ.

For $\boldsymbol{S} \in \mathbb{R}_+^k$, $t \in \mathbb{R}_+$ denote $(\boldsymbol{s}-t)^+ = ((s_1-t)^+, \ldots, (s_k-t)^+)$ and $\boldsymbol{s} \wedge t = ((s_1 \wedge t), \ldots, (s_k \wedge t))$. Then we can write $\boldsymbol{S} = (\boldsymbol{S} \wedge t) + (\boldsymbol{S}-t)^+$. Define a mapping $\psi : \mathscr{P}(\mathbb{R}_+^k) \times \mathbb{R}_+ \to \mathscr{P}(\mathbb{R}_+^k)$ by

$$\int h\,\mathrm{d}\psi(\rho, t) = \int h((\boldsymbol{s}-t)^+)\,\mathrm{d}\rho(\boldsymbol{s}), \quad h \in C(\mathbb{R}_+^k).$$

The proof of the following lemma is straightforward:

1.5. Lemma. *ψ is continuous.*

Now we define:

1.6. Definition. The *residual prediction process* of $\boldsymbol{S}$ is the $\mathscr{P}(\mathbb{R}_+^k)$-valued process $\nu_t = \psi(\mu_t, t)$.

Clearly ν_T is a r.c.d. of $(\boldsymbol{S}-T)^+$ given $\mathscr{F}_T$, where T is any finite stopping time. This shows also that μ is indistinguishable from the process $\delta_{\boldsymbol{S}\wedge t} * \nu_t$, where $*$ means convolution and δ is the Dirac measure. Since convolution on $\mathbb{R}_+^k$ is continuous with respect to weak convergence and since $\boldsymbol{S} \wedge t$ is a continuous process, we have $\mu_{t-} = \delta_{\boldsymbol{S}\wedge t} * \nu_{t-}$. It follows that μ and ν have exactly the same jump times.

The prediction process μ and the residual prediction process ν each have their advantages. The important property of μ is that the integral processes $\int f\,\mathrm{d}\mu_t$ are martingales. On the other hand, it can be shown that ν is a *time homogeneous strong Markov process.* In fact, the residual prediction process can be regarded as a modification of Knight's definition of a prediction process, which also implies the strong Markov property (see [13]).

1.7. Proposition. *The residual prediction process ν is a time homogeneous strong Markov process.*

Proof. Let T be any finite stopping time and $t>0$ arbitrary. We want to show that the conditional distribution of ν_{T+t} given $\mathscr{F}_T$ is a function of ν_T only, and, moreover, a function which does not depend on T. It suffices to show this for the conditional distribution of $\int h\,\mathrm{d}\nu_{T+t}$ given $\mathscr{F}_T$, for every function $h\in C(\mathbb{R}_+^k)$. So let h be fixed.

Denote $g_t(s)=s+\infty 1_{\{s>t\}}$ and $f_t(s)=(g_t(s_1),\ldots,g_t(s_k))$. Then $\mathscr{F}_{T+t}=\mathscr{F}_T\vee\sigma(f_t((S-T)^+))$. Denote also $h_t(s)=h((s-t)^+)$, and let ξ_n be a sequence of finite measurable partitions of $\mathbb{R}_+^k$ such that $\sigma(\xi_n)$ increases towards the Borel σ-field of $\mathbb{R}_+^k$.

Now define the jointly measurable function $\beta:\mathbb{R}_+^k\times\mathscr{P}(\mathbb{R}_+^k)\to\mathbb{R}$ by (taking $0/0=0$)

$$\beta(u,\rho)=\overline{\lim_n}\sum_{A\in\xi_n}1_A(u)\frac{1}{\rho(f_t^{-1}(A))}\int_{f_t^{-1}(A)}h_t(s)\rho(\mathrm{d}s).$$

By the martingale convergence theorem, it is easy to see that

$$\begin{aligned}\beta(f_t((S-T)^+),\nu_T)&\overset{\text{a.s.}}{=}E[h_t((S-T)^+)\,|\,\mathscr{F}_T\vee\sigma(f_t((S-T)^+))]\\&\overset{\text{a.s.}}{=}E[h((S-(T+t))^+)\,|\,\mathscr{F}_{T+t}]\overset{\text{a.s.}}{=}\int h\,\mathrm{d}\nu_{T+t}.\end{aligned}$$

Note that β depends on t and h, but not on T.

Finally, writing $\beta(u,\rho)=\beta_\rho(u)$, define the mapping $\alpha:\mathscr{P}(\mathbb{R}_+^k)\to\mathscr{P}(\mathbb{R}_+^k)$ by

$$\alpha(\rho)(B)=\rho(f_t^{-1}(\beta_\rho^{-1}(B)))=\int 1_{\{\beta(f_t(u),\rho)\in B\}}\rho(\mathrm{d}u).$$

Then $\alpha(\nu_T)(B)$ is a version of $P(\int h\,\mathrm{d}\nu_{T+t}\in B\,|\,\mathscr{F}_T)$. Thus $\alpha(\nu_T)$ is a r.c.d. of $\int h\,\mathrm{d}\nu_{T+t}$ given $\mathscr{F}_T$. □

2. The martingale representation theorem

To apply martingale calculus, we recall the notation of Arjas and Norros [4]. Let

$$T_0=0,\quad J_0=\emptyset;\qquad T_n=\inf\{S_i\colon 1\le i\le k,\ S_i>T_{n-1}\},$$

$$J_n=\{i\colon S_i=T_n\},\qquad n\ge 1.$$

We assume that $P(S_i = 0) = 0$ for all i. Then $\{T_n, J_n\}$ is a marked point process with mark space $\mathcal{T} = \mathcal{T}_0 \backslash \{\emptyset\}$. (If the probability that two or more components fail simultaneously is zero, then the mark space can be reduced to $\{1, \ldots, k\}$.)

Denote further

$$U_I = \inf\{T_n : J_n = I\}, \quad N_t(I) = 1_{\{U_I \leq t\}}, \qquad t \geq 0, \quad I \in \mathcal{I}.$$

No two of the single jump processes $N(I)$ can jump simultaneously (and, unless $k = 1$, some of them never jump).

Let $A(I)$ be the $(\mathcal{F}_t)$-compensator of $N(I)$, and denote by $M(I)$ the martingale $N(I) - A(I)$. We recall the well-known martingale representation theorem (see, e.g., Jacod [10], Brémaud and Jacod [6]):

2.1. Theorem. *Let M be any locally square integrable local martingale with respect to the (internal!) history $(\mathcal{F}_t)$. The there exist predictable processes $C(I)$, $I \in \mathcal{I}$, such that*

$$M_t = \sum_{I \in \mathcal{I}} \int_0^t C_s(I) \, \mathrm{d}M_s(I) + M_0.$$

Using the representation theorem we can prove the following fact about the prediction process:

2.2. Proposition. *The prediction process (and the residual prediction process) jumps exactly at the jump times of the fundamental martingales $M(I)$. More precisely, for any finite stopping time T,*

$$\{\mu_T \neq \mu_{T-}\} \overset{\text{a.s.}}{=} \bigcup_I \{\Delta M_T(I) \neq 0\}.$$

Proof. We first show the inclusion '$\supseteq$'. Let I be fixed. In order to make everything bounded, let T_n be an increasing sequence of stopping times such that $|M_{T_n}(I)| \leq n$, $U_I \leq \sup T_n$ and $U_I < \sup T_n$ on the set $\{U_I < \infty\}$. For each n, choose a bounded measurable function h^n such that $h^n(S) = M_{T_n}(I)$ a.s. Then, by Proposition 1.4,

$$\{\Delta M_T(I) \neq 0\} \overset{\text{a.s.}}{=} \bigcup_n \left\{ \int h^n \, \mathrm{d}\mu_T \neq \int h^n \, \mathrm{d}\mu_{T-} \right\} \subseteq \{\mu_T \neq \mu_{T-}\}.$$

The inclusion '$\subseteq$' follows from Theorem 2.1. Indeed, for any countable, convergence-determining set of functions $h_n \in C(\mathbb{R}_+^k)$

$$\{\mu_T \neq \mu_{T-}\} = \bigcup_n \left\{ \int h_n \, \mathrm{d}\mu_T \neq \int h_n \, \mathrm{d}\mu_{T-} \right\} \overset{\text{a.s.}}{\subseteq} \bigcup_I \{\Delta M_T(I) \neq 0\},$$

where the last inclusion follows from Theorem 2.1. $\square$

More use of the representation theorem is made in the next section.

3. Systems weakened by failures

One motivation for considering multivariate life distributions with positive dependence between the component life lengths is the idea that the components of a system can often be assumed to 'support' each other. For example, the destruction of one component often increases the stress on the others. In that case, the *event* that one component fails has an immediate *causal* influence on the expected life lengths of the remaining components. Our definition of a system weakened by failures is a formalization of this aspect of failure process dynamics.

3.1. Definition. The system S is *weakened by failures* (WBF), if the prediction process μ jumps downwards at failure times, that is, $\mu_{S_i} \leq \mu_{S_i-}$ a.s. for every $i = 1, \ldots, k$.

Remark. This definition differs a little, not only in form but also in content, from the corresponding definition in Arjas and Norros [4]. The latter worked only in the case of continuous compensators $A(I)$, while a rather sophisticated extra condition, 'monotonically weakened by failures', was used in the general case. However, there is no implication in either direction between 'monotonically weakened by failures' and Definition 3.1.

It is important to note that the definition of a system weakened by failures could have also been made in terms of the residual prediction process:

3.2. Proposition. *S is weakened by failures if and only if the residual prediction process jumps downwards at failure times, that is, $\nu_{S_i} \leq \nu_{S_i-}$ a.s. for every $i = 1, \ldots, k$.*

Proof. The assertion follows directly from the equation

$$\mu_t = \delta_{S \wedge t} * \nu_t. \qquad \square$$

To proceed, we need certain results concerning square integrable martingales on $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$. In the martingale representation theorem, the predictable 'coefficient' processes are not determined uniquely. The following lemma, which turns out to be very useful, deals with the choice of 'suitable' coefficient processes $C(I)$.

3.3. Lemma. *Let M be a locally square integrable local martingale with $M_0 = 0$. Suppose that $\Delta M_{S_i} \leq 0$ a.s. for all i. Then the predictable coefficient processes $C(I)$ in the representation*

$$M_t = \sum_{I \in \mathcal{I}} \int_0^t C_s(I)\, \mathrm{d}M_s(I)$$

can be chosen so that, for all I, $C(I) \leq D \leq 0$, where $D_t = \sum_I C_t(I) \Delta A_t(I)$. (The notation '$\Delta M_t$' means the jump of the process M at time t: $\Delta M_t = M_t - M_{t-}$.)

Proof. By Theorem 3.75 of Jacod [11] there exist predictable processes $Y(I)$ which are the 'predictable conditional expectations' of the jumps ΔM_{U_I}. In terms of Arjas and Norros [4], $Y_t(I) = Z_t(I) - M_{t-}$ and $Y_t(I)$ has the interpretation that it tells 'what would be the jump ΔM_t if there would be a failure of pattern I at time t'. Since the jumps ΔM_{U_I} are nonpositive, the $Y(I)$'s can also be assumed to be nonpositive everywhere.

By the theorem mentioned above, we may choose in the representation theorem

$$C_t(I) = Y_t(I) + 1_{\{B_t<1\}} \frac{1}{1-B_t} \bar{Y}_t,$$

where $B = \sum \Delta A(J)$ and $\bar{Y} = \sum Y(J)\Delta A(J)$. Moreover, the processes $Y(I)$ can be chosen so that $\bar{Y}_t = 0$ whenever $B_t = 1$. Then (with $0/0 = 0$)

$$D = \bar{Y} + \frac{B}{1-B} \bar{Y} = \frac{1}{1-B} \bar{Y} \leq 0$$

and

$$D - C(I) = \frac{1}{1-B} \bar{Y} - Y(I) - \frac{1}{1-B} \bar{Y} = -Y(I) \geq 0. \qquad \square$$

From Lemma 3.3 we immediately obtain a simple criterion for a martingale to be zero (and for two martingales to be equal):

3.4. Proposition. *Let M be a locally square integrable martingale such that $M_0 = 0$ and $\Delta M_{S_i} = 0$ for all $i = 1, \ldots, k$. Then $M = 0$.*

Proof. Applying Lemma 3.3 for the martingales M and $-M$ yields that the coefficient processes $C(I)$ in the representation theorem 2.1 can be chosen to be all identically zero. $\square$

From Lemma 3.3 we get also the following result, which is central in our positive dependence considerations. This result was, in a slightly different form, implicitly presented also in Arjas and Norros [4].

3.5. Proposition. *Let M and M' be square integrable martingales such that for all $i = 1, \ldots, k$, $\Delta M_{S_i} \leq 0$ and $\Delta M'_{S_i} \leq 0$. Then* $\mathrm{Cov}(M_\infty, M'_\infty) \geq 0$.

Proof. We may assume that $M_0 = M'_0 = 0$. Let M and M' have the representations

$$M_t = \sum_I \int_0^t C_s(I)\, \mathrm{d}M_s(I), \qquad M'_t = \sum_I \int_0^t C'_s(I)\, \mathrm{d}M_s(I).$$

By Lemma 3.3, we may assume that, for all I,

$$C(I) \leq \sum_J C(J)\Delta A(J) = D \leq 0 \quad \text{and} \quad C'(I) \leq \sum_J C'(J)\Delta A(J) = D' \leq 0.$$

By the rules of Stieltjes stochastic calculus,

$$\mathrm{Cov}(M_\infty, M'_\infty) = E\langle M, M'\rangle_\infty = E\sum_I\sum_J \int_0^\infty C_s(I)C'_s(J)\,\mathrm{d}\langle M(I), M(J)\rangle_s.$$

Since

$$\langle M(I), M(J)\rangle_t = \int_0^t (\delta_{IJ} - \Delta A_s(I))\,\mathrm{d}A_s(J),$$

we can write

$$\begin{aligned}\langle M, M'\rangle_t &= \sum_I \int_0^t C_s(I)C'_s(I)\,\mathrm{d}A_s(I) - \sum_{s\le t}\sum_I\sum_J C_s(I)C'_s(J)\Delta A_s(I)\Delta A_s(J)\\ &= \sum_I \int_0^t C_s(I)C'_s(I)\,\mathrm{d}A^c_s(I) + \sum_{s\le t}\left[\sum_I C_s(I)C'_s(I)\Delta A_s(I) - D_sD'_s\right],\end{aligned}$$

where $A^c(I)$ means the continuous part of $A(I)$. (For a summary of calculation rules used above, see Arjas and Norros [4].)

In the last expression, the first sum is nonnegative, because $C(I)C'(I)\ge 0$. But so is the second, since for any finite predictable S, $D_SD'_S = \sum_I C_S(I)\Delta A_S(I)D'_S \le \sum_I C_S(I)C'_S(I)\Delta A_S(I)$. Thus we have shown that $\langle M, M'\rangle$ is in fact increasing. □

The next proposition gives equivalent conditions for Definition 3.1:

3.6. Proposition. *Let $\mathscr{H}$ be a set of increasing bounded functions $\mathbb{R}^k_+ \to \mathbb{R}$ such that, for all $\rho_1, \rho_s \in \mathscr{P}(\mathbb{R}^k_+)$,*

$$\rho_1 \le \rho_2 \Leftrightarrow \forall h\in\mathscr{H} \int h\,\mathrm{d}\rho_1 \le \int h\,\mathrm{d}\rho_2.$$

For $h\in\mathscr{H}$ denote $M^h_t = \int h\,\mathrm{d}\mu_t$, where μ is the prediction process. Then the following conditions are equivalent:

(i) *S is weakened by failures;*
(ii) *for all $h\in\mathscr{H}$, $\Delta M^h_{S_i} \le^{\text{a.s.}} 0$ for all $i = 1, \dots, k$;*
(iii) *for all $h\in\mathscr{H}$, the coefficients $C^h(I)$ in the representation*

$$M^h_t = \sum_I \int_0^t C^h_s(I)\,\mathrm{d}M_s(I) + M^h_0$$

can be chosen so that $C^h(I) \le \sum_J C^h(J)\Delta A(J) = D^h \le 0$.

Proof. (i) ⇔ (ii): By Proposition 1.4, for all $h\in\mathscr{H}$, $\Delta M^h_{S_i} =_{\text{a.s.}} \int h\,\mathrm{d}\mu_{S_i} - \int h\,\mathrm{d}\mu_{S_i-}$. The assertion follows immediately. (ii) ⇔ (iii): The assertion follows from Lemma 3.3. □

From condition (iii) above we can make the following observation:

3.7. Theorem. *If S is weakened by failures, then the prediction process μ is increasing between the failure times S_i.*

Thus the paths of the prediction process of a system weakened by failures have a *piecewise monotone* structure with respect to the stochastic order. Upwards jumps occur only at the times at which one or more of the compensators $A(I)$ have a jump.

We conclude this section with the following result, which was, with a slightly modified definition of a system weakened by failures, presented in Arjas and Norros [4].

3.8. Theorem. *If S is weakened by failures, then it is associated.*

Proof. Let f and g be any increasing bounded functions $\mathbb{R}_+^k \to \mathbb{R}$. Denote $M_t^f = \int f \, d\mu_t$ and $M_t^g = \int g \, d\mu_t$. By assumption, the martingales M^f and M^g satisfy the assumptions of Proposition 3.5. Thus

$$\mathrm{Cov}(f(S), g(S)) = \mathrm{Cov}(M_\infty^f, M_\infty^g) \geq 0. \qquad \square$$

4. A connection between aging and dependence

The relationships between notions of aging and those of dependence have been discussed to a certain extent in reliability literature. For example, Harris [12] included a positive dependence condition to his definition of multivariate increasing hazard rate; this was needed in the proofs of closure properties. Brindley and Thompson [7] showed that their concept of multivariate increasing failure rate is compatible with both positive and negative dependence between the component life lengths.

Arjas [2] gave a definition for multivariate increasing failure rate with respect to a history $(\mathcal{F}_t)$ (MIFR/$(\mathcal{F}_t)$), which was based on the stochastic process approach. In the case of the internal history, his definition can be given in terms of the residual prediction process:

4.1. Definition. The component life length vector S is said to have a distribution with *multivariate increasing failure rate* (MIFR) if the residual prediction process ν is a decreasing process.

The intuitive meaning of this definition is that whatever happens at the component level, the residual life lengths become always stochastically shorter. Since a system is weakened by failures whenever its residual prediction process decreases at failure times (Proposition 3.2), MIFR trivially implies WBF. As a corollary to Theorem 3.8 we then get the following result:

4.2. Theorem. *If S is MIFR (in the sense of Definition 4.1), then it is associated.*

Thus, Arjas' definition for MIFR (w.r.t. the internal history) implies a form of positive dependence between the component life lengths.

There is a certain formal analogy between Theorem 4.2 and a well known result of Fortuin, Kasteleyn and Ginibre (see [9]). They showed that for a probability measure on a finite lattice, the 'logarithmic convexity' condition

$$P(\{a \vee b\})P(\{a \wedge b\}) \geq P(\{a\})P(\{b\})$$

implies association. Also MIFR is a kind of logarithmic convexity condition; for example, it implies the convexity of the compensators $A(I)$ (see Arjas [3]).

5. An application: A multivariate exponential distribution

As an application we consider the case that the $\mathscr{I}_0$-valued process X is a time homogeneous Markov process. Then the compensators $A(I)$ are linear between failure times, with derivatives (that is, stochastic intensities) that are functions of the present state X_t only.

For brevity, write $\mathscr{I}$ instead of $\mathscr{I}_0$. Now let $\lambda_K(I)$ be the intensity for failure pattern I when the state of the process X (that is, the set of operating components) is K. In usual Markov process terminology, $\lambda_K(I)$ is the transition intensity from state K to state $K \setminus I$. (We assume $\lambda_K(\emptyset) = 0$.)

Since we want to compare this kind of measures for different initial states, we allow the initial state X_0 to be any fixed element of $\mathscr{I}$.

We call the corresponding distribution of S *the multivariate exponential distribution* (MED). This model contains as special cases several multivariate exponential distributions that have appeared in the literature, e.g. that of Marshall and Olkin (see [14]). In a MED, the times between failures are exponential, and at a given time, the set of still functioning components contains all information from the past that is relevant for predictions of the future. In general, the lower-dimensional distributions do not have these properties, especially, the one-dimensional distributions need not be exponential. The bivariate case has been studied at least in Sagsveen [15].

In a construction below we need the well known Strassen representation theorem for stochastic ordering. Since Strassen's original paper [16] is rather abstract and difficult to read, we present here Strassen's theorem and its proof in a form which is restricted to our case. As a lemma we need an application of the separating hyperplane theorem.

5.1. Lemma. *Let M be a convex and closed set of probability measures on a Polish space E. Then, for any probability measure P on E, $P \in M$ if and only if $\int f \, dP \leq \sup_{Q \in M} \int f \, dQ$ for all $f \in C(E)$.*

5.2. Theorem (Strassen). *Let P and Q be probability measures on the subsets of $\mathscr{I}$.*

Then $P \leqslant Q$ *(in the sense of stochastic order) if and only if there exists a transition probability matrix* $\Pi = (\pi(I,J))_{I,J\in\mathcal{I}}$ *such that*

(i) Π *is 'upwards', that is,* $\pi(I,J) > 0$ *implies* $I \subseteq J$;

(ii) $Q = P\Pi$, *that is,* $Q(\{J\}) = \sum_I P(\{I\})\pi(I,J)$ *for all* $J \in \mathcal{I}$.

Proof. The 'if' part is easy to check. So assume $P \leqslant Q$. Denote $M = \{P'$: there exists an upwards kernel Π such that $P' = P\Pi\}$. It is easy to see that M is convex and closed. We have to show that $Q \in M$. To apply Lemma 5.1, let f be any function $\mathcal{I} \to \mathbb{R}$.

Define a function $\gamma : \mathcal{I} \to \mathcal{I}$ so that, for every I, $\gamma(I) \supseteq I$ and

$$f(\gamma(I)) = \max_{J\supseteq I} f(J).$$

Let $\Pi = (\pi(I,J))$ be the deterministic transition matrix defined by $\pi(I,J) = \delta_{\gamma(I),J}$. Now $f \circ \gamma \geqslant f$ and $f \circ \gamma$ is a decreasing function. Thus we have

$$\int f\,\mathrm{d}Q \leqslant \int f \circ \gamma\,\mathrm{d}Q \leqslant \int f \circ \gamma\,\mathrm{d}P = \int f\,\mathrm{d}(P\Pi) = \sup_{P'\in M} \int f\,\mathrm{d}P',$$

and the theorem is proved. □

As a corollary we get the following result for finite (not necessarily probability) measures on the subsets of $\mathcal{I}$.

For a function $\alpha : \mathcal{I} \to \mathbb{R}_+$, denote by $\hat{\alpha}$ the measure with point mass function α, that is, $\hat{\alpha}(\mathcal{M}) = \sum_{I\in\mathcal{M}} \alpha(I)$ for $\mathcal{M} \subseteq \mathcal{I}$. For two such functions, say α and β, we write $\hat{\alpha} \leqslant \hat{\beta}$ if $\hat{\alpha}(V) \leqslant \hat{\beta}(V)$ for each upper set $V \subseteq \mathcal{I}$.

5.3. Corollary. *Let* α *and* β *be functions from* $\mathcal{I}$ *into* $\mathbb{R}_+$. *Then the following are equivalent*:

(i) $\hat{\alpha} \leqslant \hat{\beta}$;

(ii) *there exists an upwards stochastic matrix* $(\pi(I,J))$ *such that, for all* J, $\sum_I \alpha(I)\pi(I,J) \leqslant \beta(J)$.

Proof. Again the (ii)⇒(i) part is immediate. If $\sum \alpha(I) = \sum \beta(I)$, then (i)⇒(ii) follows directly from the theorem. If $\sum \alpha(I) < \sum \beta(I)$, first increase the value $\alpha(\phi)$ so that an equality of the total masses is achieved. Then apply the theorem, and finally remove again the extra mass from ϕ to obtain the asserted inequality. □

In order to apply our results to the MED, we have to find conditions which guarantee stochastic dominance between two such distributions. So let P and Q be two MED's with respective intensity functions $\kappa_K(I)$ and $\lambda_K(I)$ and initial states X_0 and Y_0. We need the following notation: For $\phi \neq I \subseteq L \subseteq K \in \mathcal{I}$, denote $\lambda_L^K(I) = \sum_{J\subseteq K\setminus L} \lambda_K(I \cup J)$. Thus $\lambda_L^K(I)$ is the total intensity, when the set of still operating components is K, for a failure whose intersection with L is I.

5.4. Proposition. *Let $P(Q)$ be a MED with intensity function $\kappa_K(I)$ $(\lambda_K(I))$ and initial state X_0 (Y_0). Suppose*

(i) $X_0 \supseteq Y_0$;

(ii) *for any* K, L *with* $K \supseteq L$, $\hat{\kappa}_L^K \leq \hat{\lambda}_L$.

Then $P \geq Q$.

Proof. We prove the proposition by the familiar technique, showing that processes X and Y, with respective distributions P and Q, can be defined on the same probability space so that $X \supseteq Y$ everywhere.

By assumption (ii) and Corollary 5.3, for all pairs $K \supseteq L$, there exists a transition matrix $(\pi_L^K(I, J))_{I,J \in \mathcal{I}}$ such that $\sum_{J \supseteq I} \pi_L^K(I, J) = 1$ and $\sum_I \kappa_L^K(I)\pi_L^K(I, J) \leq \lambda_L(J)$. Denote

$$a(L, K) = \sum_{J \cap L = \emptyset} \kappa_K(J), \qquad b(L, K) = \sum_{J \cap L \neq \emptyset} \kappa_K(J),$$

$$c(L, K, J) = \lambda_L(J) - \sum_I \kappa_L^K(I)\pi_L^K(I, J), \qquad c(L, K) = \sum_J c(L, K, J),$$

$$d(L, K) = a(L, K) + b(L, K) + c(L, K);$$

$$p_L^K(J) = \frac{\kappa_K(J)}{a(L, K)}, \quad J \cap K = \emptyset, \qquad r_L^K(J) = \frac{\kappa_K(J)}{b(L, K)}, \quad J \cap K \neq \emptyset,$$

$$q_L^K(J) = \frac{c(L, K, J)}{c(L, K)}.$$

Now consider the following stochastic mechanism. Suppose the processes X and Y have been generated up to time T, and at time T a failure is obtained (or $T = 0$). Let $X_T = K$, $Y_T = L$, and suppose $K \supseteq L$ (induction hypothesis). Then the next failure time and pattern are determined as follows:

1. Choose the waiting time W until the next failure from the distribution $\text{Exp}(d(L, K))$.

2. Give a random variable U the value 1, 2 or 3 with probabilities $a(L, K)/d(L, K)$, $b(L, K)/d(L, K)$ and $c(L, K)/d(L, K)$ respectively. (1 means failure in process X only, 2 in both X and Y and 3 in Y only.)

3. Choose the next failure pattern as follows:

if $U = 1$, take J from the distribution p_L^K and set $X_{T+W} = K \backslash J$, $Y_{T+W} = L$;

if $U = 2$, take I with the distribution r_L^K and set $X_{T+W} = K \backslash I$; then choose J with the distribution $\pi_L^K(I, \cdot)$ and set $Y_{T+W} = Y \backslash J$;

if $U = 3$, take J from the distribution q_L^K and set $X_{T+W} = K$, $Y_{T+W} = L \backslash J$.

Then the inclusion $X_{T+W} \supseteq Y_{T+W}$ holds, and it is easily seen that the algorithm above gives correct transition intensities to both processes X and Y. Thus we have proved the proposition. □

From Proposition 5.4 we obtain a sufficient condition for a MED to be weakened by failures:

5.5. Proposition. *Let P be MED with intensity function $\lambda_K(I)$. Suppose that for all K, L with $K \supseteq L$, $\hat{\lambda}_L^K \leq \hat{\lambda}_L$. Then P is WBF.*

Proof. We prove the proposition by showing that the residual prediction process of the system decreases at failure times. By the Markov property, the value of the residual prediction process at time t is MED with intensity function $\lambda_K(I)$ and initial state X_t. Thus we only have to show that if we consider two MED's with intensity function $\lambda_K(I)$ and initial states K, L such that $K \supseteq L$, then the distribution with smaller initial state is stochastically smaller. But this follows directly from Proposition 5.4. □

The condition in Proposition 5.5 has a simple intuitive meaning: having fewer operating components implies higher failure intensities for the remaining ones. In the special case that the components fail only one at a time, the condition is simply the following:

$$\text{for all } i \in L \subseteq K, \quad \lambda_K(\{i\}) \leq \lambda_L(\{i\}).$$

Next we shall discuss the necessity of our conditions for stochastic order and WBF by a multivariate exponential distribution. First we note that in Proposition 5.4, the condition is necessary for $K = X_0$, $L = Y_0$:

5.6. Proposition. *Let P (Q) be a MED with intensity function κ (λ) and initial state K (L). Suppose that $P \geq Q$. Then $\hat{\kappa}_L^K \leq \hat{\lambda}_L$.*

Proof. Let $V \subseteq \mathcal{I}$ be any upper set. Denote $a = \hat{\kappa}_L^K(V)$, $b = \hat{\lambda}_L(V)$. We have to show that $a \leq b$. Denote $V' = \{I \in V: I \subseteq L\}$ and $V'' = \{I \cup J: I \in V', J \subseteq L^c\}$. It is easy to see that V'' is an upper set and $a = \hat{\kappa}_L^K(V'')$ and $b = \hat{\lambda}_L(V'')$.

Assume the contrary: $a > b$. Let X be the canonical failure process and consider P and Q as distributions for X. Now, by the differential definition of intensity,

$$P(X_\varepsilon^c \in V'') = P(X_0^c \cup J_1 \in V'', T_1 \leq \varepsilon < T_2) + o(\varepsilon)$$

$$= P(J_1 \in V'', T_1 \leq \varepsilon < T_2) + o(\varepsilon) = a\varepsilon + o(\varepsilon)$$

and, correspondingly, $Q(X_\varepsilon^c \in V'') = b\varepsilon + o(\varepsilon)$. ($T_n$ and J_n are defined at the beginning of Section 2.) Thus $P(X_\varepsilon^c \in V'') > Q(X_\varepsilon^c \in V'')$ when ε is sufficiently small. But $\{X_\varepsilon^c \in V''\}$ is a lower set (on the canonical space Ω, or, through the natural identification, on $\mathbb{R}_+^k$). Thus we get a contradiction with the assumption $P \geq Q$. □

The condition in Proposition 5.5 can now be shown to be necessary in the case where all transitions are possible.

5.7. Proposition. *Let P be MED with intensity function $\lambda_K(I)$. Suppose that $\lambda_K(I) > 0$ for all K, I such that $\phi \neq I \subseteq K$. Then P is WBF if and only if $\hat{\lambda}_L^K \leq \hat{\lambda}_L$ for all L, K such that $L \subseteq K$.*

Proof. The 'if' part is Proposition 5.5. The 'only if' part follows from Proposition 5.6, since for any K, L such that $K \supset L$, the transition $K \to L$ occurs with positive probability: thus the MED with intensity function $\lambda_J(I)$ and initial state K must be bigger than that with initial state L. □

Remarks. 1. The Markov property of X implies that the residual prediction process is constant between failure times. Thus, for a MED, WBF is equivalent to MIFR.

2. By Theorem 3.8 and Proposition 5.5, the condition $\hat{\lambda}_L^K \leq \hat{\lambda}_L$ implies association for MED. Though this condition is to a certain extent also necessary for WBF (Proposition 5.7), it is not necessary for association. For example, Sagsveen [15] proved that in the bivariate case, MED is associated if

$$\lambda_{\{1\}}(\{1\}) \geq \lambda_{\{1,2\}}(\{1,2\}) + \lambda_{\{1,2\}}(\{1\}) \quad \text{and} \quad \lambda_{\{2\}} \geq \lambda_{\{1,2\}}(\{2\}).$$

Acknowledgement

The author is very indebted to Professor Elja Arjas for numerous valuable discussions. Thanks are due also to one of the referees for a shorter proof for Lemma 3.3.

References

[1] D. Aldous, Weak convergence and the general theory of processes, incomplete draft of monograph, 1981.
[2] E. Arjas, A stochastic process approach to multivariate reliability systems: Notions based on conditional stochastic order, Math. Oper. Res. 6 (1981) 263-276.
[3] E. Arjas, The failure and hazard processes in multivariate reliability systems, Math. Oper. Res. 6 (1981) 551-562.
[4] E. Arjas and I. Norros, Life lengths and association: A dynamic approach, Math. Oper. Res. 9(1) (1984) 151-158.
[5] R.E. Barlow and F. Proschan, Statistical theory of reliability and life testing (Holt, Rinehart and Winston, New York, 1975).
[6] P. Brémaud and J. Jacod, Processus ponctuels et martingales: résultats récents sur la modélisation et le filtrage, Adv. Appl. Prob. 9 (1977) 362-476.
[7] E.C. Brindley and W.A. Thompson, Jr., Dependence and aging aspects of multivariate survival, JASA 67 (1972) 822-830.
[8] C. Dellacherie, Capacités et processus stochastiques (Springer-Verlag, Berlin, 1972).
[9] C.M. Fortuin, P.W. Kasteleyn and J. Ginibre, Correlation inequalities on some partially ordered sets, Commun. Math. Phys. 22 (1971) 89-103.
[10] J. Jacod, Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales, Z. Wahrsch. 34 (1975) 225-244.
[11] J. Jacod, Calcul Stochastique et Problèmes de Martingales, Lect. Notes in Math. 714 (Springer-Verlag, Berlin, 1979).
[12] R. Harris, A multivariate definition for increasing hazard rate, Ann. Math. Stat. 41 (1970) 713-717.
[13] F. Knight, A predictive view of continuous time processes, Ann. Prob. 3 (1975) 573-596.
[14] A.W. Marshall and I. Olkin, A multivariate exponential distribution, JASA 62 (1967) 30-44.
[15] A. Sagsveen, Interaction in Markov chains, graduate thesis in statistics at the Oslo University, 1982.
[16] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Stat. 36 (1965) 423-439.