STOCHASTIC PRECEDENCE AND MINIMA AMONG DEPENDENT VARIABLES

EMILIO DE SANTIS, YAAKOV MALINOVSKY, AND FABIO SPIZZICHINO

Abstract. The notion of stochastic precedence between two random variables emerges as a relevant concept in several fields of applied probability. When one consider a vector of random variables $X_1, ..., X_n$, this notion has a preeminent role in the analysis of minima of the type $\min_{j \in A} X_j$ for $A \subset \{1, \ldots n\}$. In such an analysis, however, several apparently controversial aspects can arise (among which phenomena of "non-transitivity"). Here we concentrate attention on vectors of non-negative random variables with absolutely continuous joint distributions, in which a case the set of the multivariate conditional hazard rate (m.c.h.r.) functions can be employed as a convenient method to describe different aspects of stochastic dependence. In terms of the m.c.h.r. functions, we first obtain convenient formulas for the probability distributions of the variables $\min_{j \in A} X_j$ and for the probability of events $\{X_i = \min_{j \in A} X_j\}$. Then we detail several aspects of the notion of stochastic precedence. On these bases, we explain some controversial behavior of such variables and give sufficient conditions under which paradoxical aspects can be excluded. On the purpose of stimulating active interest of readers, we present several comments and pertinent examples.

Keywords: Multivariate Conditional Hazard Rates, Non-transitivity, aggregation/marginalization paradoxes, "small" variables, initially time–homogeneous models, time–homogeneous load sharing models.

AMS MSC 2010: 60K10, 60E15, 91B06.

1. Introduction

Let us consider a vector of non-negative random variables $\mathbf{X} = (X_1, ..., X_n)$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We write $[n]$ for the set $\{1, 2, \ldots, n\}$. For $i \neq j \in [n]$, one says that X_i *is smaller than* X_j *in the stochastic precedence* whenever the inequality

$$\mathbb{P}(X_i \le X_j) \ge \mathbb{P}(X_j \le X_i)$$

holds true; this condition will be denoted by $X_i \preceq_{sp} X_j$.

This notion of comparison is clearly very natural and it is of actual interest for some applications. In fact, it had been considered several times in the literature, possibly under a variety of different terms. In the last few years, in particular, this property has been attracting more and more interest in different applied contexts; see e.g. references [1, 5, 8, 13, 21].

Several controversial or apparently counter-intuitive aspects have been however pointed out, since a long time. In particular one can meet aspects of *non-transitivity* and other related phenomena which we will refer to as *aggregation/marginalization paradoxes.* See in particular [3, 8] and the references cited therein. More generally, it there exists a very wide literature concerning with controversial and counter-intuitive aspects related with non-transitivity, in mathematics and probability (see e.g. [15, 25, 34, 35]). In the fields of economics, statistics, social choices, as it is well-known, the interest toward these topics is enormous and the literature considering such subjects has a very long tradition, see, in particular, [6, 14, 24] and references cited therein. In our analysis, it is important to be aware of the relations and similarities among all such contexts.

The aspects concerning with aggregation/marginalization paradoxes can be seen as related to the literature on the theme of Simpson's paradoxes (see e.g. [4, 26, 30]). Specifically concerning the topic of stochastic precedence, several examples and counter-examples about controversial aspects can be found in the analysis of occurrence times for "words" in random sampling of letters from an alphabet (see e.g. [7, 9, 15, 17]). This field is also related to the analysis of stochastic comparisons for hitting times for Markov chains see e.g. [10, 11] and references therein.

Going to the specific purposes of this paper, we notice that it can be useful to understand situations where the paradoxical phenomena of stochastic precedence are to be expected or, on the contrary, where they can be excluded. We point out that many of such phenomena emerge in the case of stochastic dependence among the random variables under consideration. It is relevant, in this respect, to pay attention to the way in which stochastic dependence is described. Here we limit our attention to the cases, when the joint probability distribution of $(X_1, ..., X_n)$ is absolutely continuous and can thus be described in terms of the joint probability density.

More in particular we consider non-negative random variables, in which case a possible tool for describing the joint probability law and the type of stochastic dependence can be based on the family of *multivariate conditional hazard rate* functions. See e.g. [27] and also the reviews within the more recent papers [29, 32, 33]. This tool is different, but equivalent to the one based on the joint density function. In fact, there are well-known formulas that, at least in principle, allow one to derive the m.c.h.r. functions from the knowledge of the joint density and viceversa. But the two types of descriptions completely differ in their abilities to highlight different aspects of stochastic dependence. Here, we aim to point out that, for non-negative variables, the description based upon the m.c.h.r. functions can reveal a useful one to understand some aspects of stochastic precedence and related issues.

The structure of the paper is described as follows.

In the next Section 2 we give some basic notation and definitions, and preliminary results concerning the minimum among several non-negative random variables in the jointly absolutely continuous case. In particular we recall basic definitions and facts about the system of the m.c.h.r. functions. Section 3 will be devoted to the notion of stochastic precedence and related controversial aspects. In Section 4 we analyze some different conditions on the variables $X_1, ..., X_n$, that exclude the occurrence of some of such paradoxical situations.

2. Notation, basic definitions and preliminary results

In this section, we give basic definitions and we show some preliminary results about the minimum among random variables. In particular we analyze the role of multivariate conditional hazard rates. For a given non-negative, absolutely continuous, random variable X, we denote by $r(t)$ the ordinary hazard rate (or failure rate) of it:

$$r(t) := \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(X \in (t, t+\Delta t) | X > t)}{\Delta t},$$

To start our discussion, we recall a very simple and useful result concerning the minimum of several independent, exponentially distributed, random variables.

Let $\Upsilon_1, ..., \Upsilon_n$ denote n independent random variables, distributed according to exponential distributions with parameters $\lambda_1, ..., \lambda_n$, respectively, and set

$$\Upsilon_{1:n} := \min\{\Upsilon_1, ..., \Upsilon_n\}.$$

Then we can state (see e.g. [22], Chp. 2) the following result.

Lemma 1. *For any $t > 0$ and $j \in [n]$, the following identities hold*

$$\mathbb{P}(\Upsilon_{1:n} = \Upsilon_j, \Upsilon_{1:n} > t) = \mathbb{P}(\Upsilon_{1:n} = \Upsilon_j)\mathbb{P}(\Upsilon_{1:n} > t), \tag{1}$$

$$\mathbb{P}(\Upsilon_{1:n} = \Upsilon_j) = \frac{\lambda_j}{\sum_{s=1}^n \lambda_s}, \tag{2}$$

$$\mathbb{P}(\Upsilon_{1:n} > t) = \exp\left\{-t \sum_{s=1}^n \lambda_s\right\}. \tag{3}$$

We now want to show (see Proposition 1) in which sense this result can be extended to the random variable $X_{1:n} := \min\{X_1, \dots, X_n\}$, where $X_1, \dots, X_n$ are not necessarily independent nor exponentially distributed. We maintain however the condition of absolute continuity for the joint probability distribution and the joint density function will be denoted by $f_{\mathbf{X}}$. The latter condition in particular implies the *no-tie property*

$$\mathbb{P}(X_i = X_j) = 0, \tag{4}$$

for any $i, j = 1, \dots, n$, with $i \neq j$, which will be of basic importance all along the paper.

We respectively denote by $f_{(1)}$, $\bar{F}_{(1)}(t)$, $h_{(1)}(t)$, $H_{(1)}(t)$, the probability density function, survival function, hazard rate function and cumulative hazard function of $X_{1:n}$. Namely

$$h_{(1)}(t) := \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(X_{1:n} \in (t, t+\Delta t] | X_{1:n} > t)}{\Delta t},$$

$$H_{(1)}(t) := \int_0^t h_{(1)}(s) ds,$$

$$\bar{F}_{(1)}(t) := e^{-H_{(1)}(t)},$$

$$f_{(1)}(t) := h_{(1)}(t) e^{-H_{(1)}(t)}. \tag{5}$$

In view of the assumption of absolute continuity, we can define the following limits, for $j = 1, \dots, n$

$$\gamma_j(t) := \lim_{\Delta t \to 0^+} \mathbb{P}(X_j = X_{1:n} | X_{1:n} \in (t, t+\Delta t]) = \mathbb{P}(X_j = X_{1:n} | X_{1:n} = t), \tag{6}$$

$$\mu_j(t) := \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}(X_j \leq t + \Delta t | X_{1:n} > t). \tag{7}$$

We notice that, in the case of regular conditional probabilities,

$$\sum_{j=1}^{n} \gamma_j(t) = 1. \tag{8}$$

Furthermore,

$$\mu_j(t) = h_{(1)}(t)\gamma_j(t). \tag{9}$$

In fact, since

$$\mathbb{P}(\{X_j \leq t + \Delta t\} \cap \{X_{1:n} > t\}) = \mathbb{P}(\{X_j \leq t + \Delta t\} \cap \{X_{1:n} \in (t, t + \Delta t]\})$$

we can write

$$\begin{aligned} h_{(1)}(t)\gamma_j(t) &= \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(\{X_j = X_{1:n}\} \cap \{X_{1:n} \in (t, t + \Delta t]\})}{\mathbb{P}(X_{1:n} \in (t, t + \Delta t])} \frac{\mathbb{P}(X_{1:n} \in (t, t + \Delta t])}{\Delta t \mathbb{P}(X_{1:n} > t)} \\ &= \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(\{X_j < t + \Delta t\} \cap \{X_{1:n} > t\})}{\Delta t \mathbb{P}(X_{1:n} > t)} = \mu_j(t). \end{aligned}$$

The following two results will have a key role in the next discussion.

Proposition 1. *With the notation introduced above, the following properties hold.*

a) *For any $t \geq 0$ one has*

$$H_{(1)}(t) = \sum_{i=1}^{n} \int_0^t \mu_i(s) ds. \tag{10}$$

b) *For any Borel set $B \in \mathcal{B}(\mathbb{R}_+)$, one can write*

$$\mathbb{P}(X_i = X_{1:n}, X_{1:n} \in B) = \int_B \mu_i(s) e^{-H_{(1)}(s)} ds. \tag{11}$$

Proof. From (8) and (9) one has $h_{(1)}(s) = \sum_{i=1}^{n} \mu_i(s)$ and then (10).Taking into account the positions (5) and (9) we obtain

$$\mathbb{P}(X_j = X_{1:n}, X_{1:n} \in B) = \int_B f_{(1)}(s) \mathbb{P}(X_j = X_{1:n} | X_{1:n} = s) ds =$$

$$\int_B h_{(1)}(s) e^{-H_{(1)}(s)} \gamma_i(s) ds, \tag{12}$$

that is equal to (11). □

Denote by Λ the Lebesgue measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$.

Theorem 1. *The following two statements are equivalent:*

a) *For each $B \in \mathcal{B}(\mathbb{R}_+)$ and $i, j = 1, \ldots n$,*

$$\mathbb{P}(X_{1:n} \in B, X_i = X_{1:n}) \leq \mathbb{P}(X_{1:n} \in B, X_j = X_{1:n});$$

b) $\Lambda(\{t \in \mathbb{R}_+ : \mu_i(t) > \mu_j(t)\}) = 0$.

Proof. The implication b) $\Rightarrow$ a) is immediate in view of the identity (11).

The implication a) $\Rightarrow$ b) is proved by contradiction.

Assume, in fact, $\Lambda(\{t \in \mathbb{R}_+ : \mu_i(t) > \mu_j(t)\}) > 0$ then, by continuity of probability measures, there exists $\varepsilon > 0$ such that $\Lambda(\{t \in \mathbb{R}_+ : \mu_i(t) > \mu_j(t) + \varepsilon\}) > 0$. Therefore, by setting

$$B = \{t \in \mathbb{R}_+ : \mu_i(t) > \mu_j(t) + \varepsilon\}$$

one obtains

$$\mathbb{P}(X_i = X_{1:n}, X_{1:n} \in B) - \mathbb{P}(X_j = X_{1:n}, X_{1:n} \in B) \geq \varepsilon \int_B e^{-H_{(1)}(s)} ds > 0.$$

□

It is convenient, at this step, to recall the definition of m.c.h.r. functions for the non-negative random variables $X_1, ...X_n$. We denote by $X_{1:n}, ..., X_{n:n}$ the corresponding order statistics. For $A \subseteq [n]$ with $|A| > 1$, set

$$X_{1:A} := \min_{i \in A} X_i.$$

In particular, we obtain

$$X_{1:[n]} := X_{1:n} = \min_{1 \leq j \leq n} X_j.$$

In the following definition for a given subset $I \subset [n]$ we will consider the random variable $X_{1:\tilde{I}}$, where the symbol $\tilde{I}$ denotes the complementary set $[n] \setminus I$.

Definition 1. *For a fixed index $j \in [n]$, an ordered set $I = (i_1, ..., i_k) \subset [n]$ with $j \notin I$, and an ordered sequence $0 < t_1 < ... < t_k$, the Multivariate Conditional Hazard Rate function $\lambda_j(t|I; t_1, ..., t_k)$ is defined as follows:*

$$\lambda_j(t|I; t_1, ..., t_k) := \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}(X_j \leq t + \Delta t | X_{i_1} = t_1, ..., X_{i_k} = t_k, X_{1:\tilde{I}} > t). \tag{13}$$

Furthermore, one puts

$$\lambda_j(t|\emptyset) := \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}(X_j \leq t + \Delta t | X_{1:n} > t). \tag{14}$$

For what specifically concerns the position in (14), we must notice that we reobtain nothing else than the functions defined in (9); more precisely

$$\mu_j(t) = \lambda_j(t|\emptyset). \tag{15}$$

For this reason, the symbol $\mu_j(t)$ will not be used anymore, from now on.

Remark 1. *The limits considered in the above definition make sense in view of the assumption of absolute continuity and the quantity $\lambda_j(t|I; t_1, ..., t_k)$ can be seen as the failure intensity, at time t, associated to the conditional distribution of the variable X_j, given the observation of the dynamic history*

$$\mathfrak{h}_t =: \{X_{i_1} = t_1, ..., X_{i_k} = t_k, X_{1:\tilde{I}} > t\}. \tag{16}$$

The functions $\lambda_j(t|I; t_1, ..., t_k)$ and $\lambda_j(t|\emptyset)$ can be computed in terms of the joint density function $f_{\mathbf{X}}$. On the other hand, based on the knowledge of the functions $\lambda_j(t|I; t_1, ..., t_k)$

and $\lambda_j(t|\emptyset)$, one can recover the function $f_{\mathbf{X}}$. In fact, the following formula holds for $0 < x_1 < ... < x_n$

$$f_{\mathbf{X}}(x_1, ..., x_n) = \lambda_1(x_1|\emptyset) \exp\left\{-\sum_{j=1}^{n} \int_0^{x_1} \lambda_j(u|\emptyset)du\right\} \times$$
$$\times \lambda_2(x_2|\{1\}; x_1) \exp\left\{-\sum_{j=2}^{n} \int_{x_1}^{x_2} \lambda_j(u|\{1\}; x_1)du\right\} \times ...$$
$$\times \lambda_{k+1}(x_{k+1}|\{1, ..., k\}; x_1, ..., x_k) \exp\left\{-\sum_{j=k+1}^{n} \int_{x_k}^{x_{k+1}} \lambda_j(u|\{1, ..., k\}; x_1, ..., x_k)du\right\} \times ...$$
$$\times \lambda_n(x_n|\{1, ..., n-1\}; x_1, ..., x_{n-1}) \exp\left\{-\int_{x_{n-1}}^{x_n} \lambda_n(u|\{1, ..., n-1\}; x_1, ..., x_{n-1})du\right\}. \tag{17}$$

Similar expressions hold when $x_1, ..., x_n$ are such that $x_{\pi(1)} < ... < x_{\pi(n)}$, for some permutation π. For proofs, details, and for general aspects see [27], [28], *and the review paper* [29].

Remark 2. *In the reliability field, the variables $X_1, ..., X_n$ are interpreted as the random lifetimes of n components in a system. From the identities* (9) *and* (15) *one immediately obtains the relation*

$$\lambda_j(t|\emptyset) = h_{(1)}(t)\gamma_j(t),$$

tying the specific m.c.h.r. function $\lambda_j(t|\emptyset)$ with the conditional probability $\gamma_j(t)$ defined in (6) *and with the univariate failure rate of the minimum $X_{1:n}$, i.e. the lifetime of the series system.*

In the frame of system reliability, indexes of importance of a component are generally relevant notions. In that context, the conditional probabilities $\lambda_j(t|\emptyset)$ and $\gamma_j(t)$ are related with the Barlow-Proschan indexes of importance of stochastically dependent components (see [2, 16, 18, 19]*) in the special case when the system is a series. An integral expression for Barlow-Proschan index in a series system can be obtained by specializing a result given in* [19]. *Such an expression is alternative to the integral one given in* (12) *above, with $B = \mathbb{R}_+$. In* [19], *the validity of expression* (8) *have been pointed out for general coherent systems.*

We notice that some arguments in [19] *might also be applied to conditional probabilities of the type $\mathbb{P}(X_j = X_{k+1:n}|X_{k+1:n} = t; \mathfrak{h}_t)$ which are related with conditional importance indexes for surviving components in a series system surviving at time t.*

Remark 3. *Let us look at the conditional distribution of the residual lifetimes of components surviving at time t given the dynamic history $\mathfrak{h}_t$ in* (16). *Similarly to what noticed for $\lambda_j(t|\emptyset)$ in the remark above, and conditioning upon the observation $\mathfrak{h}_t$, one can obtain an expression for the m.c.h.r. function $\lambda_j(t|I; t_1, ..., t_k)$ in terms of the conditional probability*

$$\mathbb{P}(X_j = X_{k+1:n}|X_{k+1:n} = t; \mathfrak{h}_t)$$

and of the conditional univariate failure rate of $X_{k+1:n}$ (namely, the residual lifetime of the series system made with the components surviving at time t).

On the other hand, the m.c.h.r. functions $\lambda_j(t|I;t_1,...,t_k)$ can be used to describe the conditional distribution of the residual lifetimes of components, given $\mathfrak{h}_t$. Such a description will be presented in formula (25)*. It can be interesting to compare it with the alternative expression that can be given in terms of copula-based representations of joint distributions of lifetimes (see in particular* [12] *and* [20]*).*

As a direct corollary of Proposition 1 we obtain that, for any vector of dependent variables, probabilities of events related to the behavior of their minimum are equal to probabilities of corresponding events for a vector of independent variables. We point out that, in the case of independence, the function $\lambda_j(\cdot|\emptyset)$ coincides with the ordinary failure rate functions $r_j(\cdot)$ and we can more precisely state the following results.

Proposition 2. *Let $(X_1,...,X_n)$ be a vector with m.c.h.r. functions $\lambda_j(t|\emptyset)$ and take independent random variables $Z_1,...,Z_n$, with ordinary failure rate functions r_j given by*

$$r_j(t) := \lim_{\Delta t \to 0^+} \frac{\mathbb{P}\left(Z_j < t+\Delta t | Z_j > t\right)}{\Delta t} = \lambda_j(t|\emptyset).$$

Then

$$\mathbb{P}(X_i = X_{1:n},\ X_{1:n} \in B) = \mathbb{P}(Z_i = Z_{1:n},\ Z_{1:n} \in B) \tag{18}$$

for any $i \in [n]$ and any Borel set B.

Proof. In order to prove (18) it is enough to apply, to both the vectors of random variables $(X_1,\dots,X_n)$ and $(Z_1,\dots,Z_n)$, item b) of Proposition 1. □

For our purposes it is also useful to specialize Proposition 1 and Proposition 2 to the limiting case $B = [0,\infty)$. Thus we obtain that the analysis of the minimum among several random variables can be reduced to the case of independent variables by means of the functions $\lambda_1(s|\emptyset),\dots,\lambda_n(s|\emptyset)$ by considering the following formulas.

$$\mathbb{P}\left(X_{1:n} > t\right) = \exp\left\{-\sum_{i=1}^{n}\int_0^t \lambda_i(s|\emptyset)ds\right\} = \exp\left\{-\sum_{i=1}^{n}\int_0^t r_i(s)ds\right\} = \mathbb{P}\left(Z_{1:n} > t\right), \tag{19}$$

$$\begin{aligned}\mathbb{P}(X_i = X_{1:n}) &= \int_0^\infty \lambda_i(s|\emptyset)\exp\{-\sum_{i=1}^{n}\lambda_i(s|\emptyset)\}ds \\ &= \int_0^\infty r_i(s)\exp\{-\sum_{i=1}^{n} r_i(s)\}ds = \mathbb{P}(Z_i = Z_{1:n})\end{aligned} \tag{20}$$

for any $i \in [n]$.

Before continuing, the method of m.c.h.r. functions for describing the behaviour of the minimum among dependent variables will now be further demonstrated by means of some relevant examples.

Example 1. *(The case of exchangeability) When $X_1,...,X_n$ are exchangeable, then the dependence of $\lambda_j(t|\emptyset)$ on the index j is obviously dropped, namely for a suitable function $\lambda(\cdot|\emptyset)$ and for $j = 1,...,n$, $t > 0$,*

$$\lambda_j(t|\emptyset) = \lambda(t|\emptyset). \tag{21}$$

Thus we obtain

$$\mathbb{P}\left(X_{1:n} > t\right) = \exp\left\{-n\int_0^t \lambda(s|\emptyset)ds\right\}, \tag{22}$$

$$\mathbb{P}\left(X_{1:n} = X_j\right) = \frac{1}{n}. \tag{23}$$

Notice that the same identities do hold even if $X_1, ..., X_n$ *are not exchangeable, provided the above condition* (21) *holds.*

Example 2. *(The case of conditional independence and identical exponential distribution). Let* Θ *be a non-negative random variable with distribution* Π_Θ *and let* $X_1, ..., X_n$ *be conditionally independent and exponentially distributed given* Θ*, i.e.*

$$\mathbb{P}\left(X_1 > x_1, ..., X_n > x_n\right) = \int_0^\infty \exp\{-\theta\sum_{i=1}^n x_i\}\Pi_\Theta\left(d\theta\right).$$

In this case one has (for details see e.g. [31]*)*

$$\lambda_j(t|\emptyset) = \mathbb{E}(\Theta|X_{1:n} > t) = \int_0^\infty \theta\Pi_\Theta\left(d\theta|X_{1:n} > t\right),$$

where $\Pi_\Theta\left(\cdot|X_{1:n} > t\right)$ *denotes the a posteriori distribution of* Θ*, given the observation* $X_{1:n} > t$*. Moreover*

$$\mathbb{P}\left(X_{1:n} > t\right) = \int_0^\infty \exp\{-n\theta t\}\Pi_\Theta\left(d\theta\right)$$

and since $X_1, ..., X_n$ *are, in particular, exchangeable*

$$\mathbb{P}\left(X_{1:n} = X_j\right) = \frac{1}{n}.$$

Example 3. *The following case can be considered as a generalization of the case of independent, exponential, variables: consider dependent random variables* $\Upsilon_1, ..., \Upsilon_n$ *such that for* $j = 1, ..., n$*, the ratio*

$$\frac{\lambda_j(t|\emptyset)}{\sum_{i=1}^n \lambda_i(t|\emptyset)}$$

does not depend on the variable t*. In such a case, the identities (1) and (2) hold.*

Example 4. *A more special class of survival models generalizing the case of independent, exponential, variables is the one of* time-homogeneous load-sharing *models, characterized by the condition*

$$\lambda_j(t|\emptyset) = r_j\left(\emptyset\right), \lambda_j(t|I; t_1, ..., t_k) = r_j\left(I\right),$$

for a suitable family of constants $\{r_j\left(\emptyset\right); r_j\left(I\right); j \in [n], I \subset [n], i \notin I\}$.

Several theoretical and applied aspects of such survival models have been studied in different fields and, in particular, in the reliability literature. See e.g. [33] *and references cited therein.*

By limiting attention to this class of models, useful examples can be constructed for different types of properties related with the arguments of this paper. In particular, in the case of time-homogeneous load-sharing model we obtain from Proposition 1

$$\mathbb{P}\left(X_{1:n}=X_j, X_{1:n}>t\right)=r_j(\emptyset)\exp\{-t\sum_{i=1}^{n} r_i\left(\emptyset\right)\}.$$

Still considering time-homogeneous load-sharing models, it is also useful recalling attention on the following property of conditional distribution of the residual lifetimes

$$X_{j_1}-t,...,X_{j_{n-k}}-t$$

given the observation of a dynamic history $\mathfrak{h}_t$ as in (16). Of course, conditionally on $\mathfrak{h}_t$, the joint distribution of $\left(X_{j_1}-t,...,X_{j_{n-k}}-t\right)$ is generally absolutely continuous if the one of $(X_1,...,X_n)$ is such. Furthermore it is a time-homogeneous load-sharing model if joint distribution of $(X_1,...,X_n)$ is such and one has the simple relation

$$\widehat{r_j}(\emptyset)=r_j(I), j\in\widetilde{I}.$$

From Proposition 1 we obtain, for $j\in\widetilde{I}$,

$$\mathbb{P}\left(X_{\tilde{I}}=X_l, X_{1:\tilde{I}}>t+s|\mathfrak{h}_t\right)=r_l(I)\exp\left\{-s\sum_{j\in\tilde{I}} r_j\left(I\right)\right\}.$$

Denote by $J_1,J_2,...J_k$ the random indices such that

$$X_{1:n}=X_{J_1},...,X_{k:n}=J_k.$$

By applying the product formula of conditional probabilities, we thus can also obtain that the joint density function

$$f_{X_{1:n},...,X_{k:n},J_1,...,J_k}(t_1,t_2,\ldots,t_k,j_1,j_2,\ldots,j_k)$$

of $(X_{1:n},\ldots,X_{k:n},J_1,\ldots,J_k)$, $k=1,\ldots,n$, with respect to the product of k-dimensional Lebesgue measure on $[0,\infty)^k$ and k-dimensional counting measure on $[n]^k$ is the product of terms of the form

$$r_{j_{h+1}}(\{j_1,\ldots,j_h\})\exp\{-(t_{h+1}-t_h)\sum_{l\neq j_1,\ldots,j_h} r_l\left(\{j_1,\ldots,j_h\}\right)\}. \tag{24}$$

Using once again Proposition 1, the argument presented above can easily be extended to the case of an arbitrary absolutely continuous model, characterized in terms of its m.c.h.r. functions.

First of all we notice that, conditionally on a dynamic history $\mathfrak{h}_t$, the joint distribution of residual lifetimes $\left(X_{j_1}-t,...,X_{j_{n-k}}-t\right)$ is characterized by the m.c.h.r. functions

$$\widehat{\lambda}_j^{(\mathfrak{h}_t)}(t|\emptyset)=\lambda_j(t|I;t_1,\ldots,t_h),\ \ j\in\tilde{I}. \tag{25}$$

We can thus state the following proposition.

Proposition 3. *The joint density function*

$$f_{X_{1:n},\dots,X_{k:n},J_1,\dots,J_k}(t_1,t_2,\dots,t_k,j_1,j_2,\dots,j_k)$$

of $(X_{1:n},\dots,X_{k:n},J_1,\dots,J_k)$*,* $k=1,\dots,n$*, with respect to the product of* k*-dimensional Lebesgue measure on* $[0,\infty)^k$ *and counting measure on* $[n]^k$ *is the product of terms of the form*

$$\lambda_{j_{h+1}}(t_{h+1};\{j_1,\dots,j_h\};t_1,\dots,t_h)\exp\{-(t_{h+1}-t_h)\sum_{l\neq j_1,\dots,j_h}\lambda_l\,(t_{h+1};\{j_1,\dots,j_h\};t_1,\dots,t_h)\}. \tag{26}$$

As Proposition 2 shows, the factors in (26) can be replaced, at any step, by corresponding factors related with independent variables whose distribution are affected by the past observations.

For time-homogeneous load–sharing models, the factors in (26) reduce to those in (24). The concept of time-homogeneous load–sharing models can be extended in a natural way to the non–homogeneous case. For such a case, the specific form of the above result has been given in [23].

Before concluding this section, we also recall attention on a further aspect of m.c.h.r. functions. For $m<n$, m.c.h.r. functions are generally different from the corresponding m.c.h.r. functions associated to the marginal distribution of the vector $(X_1,\dots,X_m)$.

3. Controversial aspects of stochastic precedence

Let Y_1 and Y_2 be two random variables. We remind from the Introduction that Y_1 stochastically precedes Y_2 if $\mathbb{P}(Y_1\le Y_2)\ge\mathbb{P}(Y_2\le Y_1)$. Under the no-tie condition, this definition is equivalent to $\mathbb{P}(Y_1\le Y_2)\ge\frac{1}{2}$. This will be written $Y_1\preceq_{sp}Y_2$. The previous formula (20) in particular provides us with a simple characterization of stochastic precedence when Y_1 and Y_2 are two non-negative random variables. In fact, letting

$$\lambda_i(t|\emptyset):=\lim_{\Delta t\to0^+}\frac{1}{\Delta t}\mathbb{P}(Y_i\le t+\Delta t|Y_{1:n}>t)$$

for $i=1,2$, we can write

$$Y_1\preceq_{sp}Y_2\Leftrightarrow\int_0^{+\infty}\lambda_1(s|\emptyset)e^{-H_{(1)}(s)}ds\ge\frac{1}{2} \tag{27}$$

where $H_{(1)}(t)=\int_0^t[\lambda_1(s|\emptyset)+\lambda_2(s|\emptyset)]\,ds$.

Example 5. *(The case of independence). Let* X_1,X_2 *be two independent, non-negative, random variables with absolutely continuous distributions characterized by the hazard rate functions* $r_1(t),r_2(t)$*, respectively. Then, in view of the characterization in* (27)*, one has*

$$X_1\preceq_{sp}X_2\Leftrightarrow\int_0^{\infty}r_1(t)e^{-\int_0^t[r_1(s)+r_2(s)]ds}dt\ge\frac{1}{2}.$$

When X_1,X_2 *are independent and exponential with parameters* r_1,r_2*, the condition* $X_1\preceq_{sp}X_2$ *simply becomes* $r_1\ge r_2$ *and so we get the hazard rate ordering.*

Example 6. *(The case of conditional independence and exponentiality). Similarly to the previous Example 2, consider now the case when Θ is a non-negative random variable with distribution Π_Θ and X_1, X_2 are conditionally independent given Θ, with*

$$\mathbb{P}(X_1 > t|\Theta = \theta) = \exp\{-c_1\theta t\},$$
$$\mathbb{P}(X_2 > t|\Theta = \theta) = \exp\{-c_2\theta t\},$$

where c_1, c_2 are two fixed positive numbers. In this case, one has

$$\lambda_i(t|\emptyset) = c_i\mathbb{E}(\Theta|X_{1:n} > t).$$

Thus, we are in the case of Example 3 and the condition $X_1 \preceq_{sp} X_2$ becomes $c_1 \geq c_2$.

Remark 4. *It is immediate to see that, in the case of stochastic independence, the condition $X_1 \preceq_{sp} X_2$ is implied by the condition that X_1 precedes X_2 in the usual stochastic ordering (written $X_1 \preceq_{st} X_2$), namely*

$$\mathbb{P}(X_1 > t) \leq \mathbb{P}(X_2 > t), \forall t > 0.$$

This implication is not valid anymore, when the condition of independence is dropped; see e.g. the discussion and counter-examples in [8, 10]. *The characterization in* (27) *can help us to easily understand the logic on which counter-examples may be built up.*

Remark 5. *The relation of stochastic precedence does not generally satisfy the transitivity property. In fact, it is possible to show examples where, for three real-valued random variables X_1, X_2, X_3, the following conditions simultaneously hold:*

$$\mathbb{P}(X_1 < X_2) > \frac{1}{2},\ \mathbb{P}(X_2 < X_3) > \frac{1}{2},\ \mathbb{P}(X_3 < X_1) > \frac{1}{2} \tag{28}$$

Possibly under different languages, this topic has been often considered in the literature and famous examples have been given (see e.g. [3, 14, 15, 25, 34, 35]*). In this respect, we point out that, for the case of non-negative variables, examples in discrete-time can be easily converted into examples in continuous-time.*

We notice furthermore that the possibility of (28) *is obviously excluded when X_1, X_2, X_3 are independent variables, satisfying the property*

$$X_1 \preceq_{st} X_2 \preceq_{st} X_3.$$

We now introduce the following notation to point out a further aspect, of stochastic precedence, which may appear controversial at first glance.

The probability $\mathbb{P}(X_{1:n} = X_j)$ will be denoted by α_j. More generally, for $A \subseteq [n]$ with $|A| > 1$, $j \in A$, we set

$$\alpha_j^{[A]} := \mathbb{P}(X_{1:A} = X_j). \tag{29}$$

For $A \subseteq [n]$ with $|A| > 1$, $i, j \in A$, set

$$X_i \preceq_{sp}^{[A]} X_j \tag{30}$$

if

$$\alpha_i^{[A]} \geq \alpha_j^{[A]}.$$

In words the relation (30) says that the random variable X_i has, with respect to X_j, a greater (or equal) probability to be the minimum among all the variables X_k with $k \in A \subset [n]$. Such a relation can be of interest in several contexts where it is important to detect which variable has a greater probability to be a minimum among a given set of random variables. For instance, consider the collection, denoted by $[n]$, of all the horses registred in a racecourse. Let $A \subset [n]$ be the set of those horses scheduled to take part in a specific race. The bookmakers are called to compare the probability of victory only among the horses belonging to A. Thus they are interested in the probabilities $\alpha_i^{[A]}$'s, for $i \in A$, and in the relations (30). It is natural to think of others contexts, in particular Economics and Reliability theory, in which definition (30) could be relevant, see e.g. Example 7 below.

Consider now two non-disjoint subsets $A, B \subseteq [n]$ and two elements $i, j \in A \cap B$.
We notice that the inequalities

$$X_i \preceq_{sp}^{[A]} X_j, X_j \preceq_{sp}^{[B]} X_i$$

can simultaneously hold (see also, e.g., [24]). In particular, it can happen that, for an element $l \notin A$, we can have

$$\alpha_i^{[A]} > \alpha_j^{[A]}, \alpha_i^{[A \cup l]} < \alpha_j^{[A \cup l]]}.$$

and, for three different elements $i, j, l \in [n]$,

$$X_i \preceq_{sp} X_j, X_j \preceq_{sp}^{[\{i,j,l\}]} X_i.$$

Example 8 in the next section shows a case where the latter situation arises. We will refer to this type of circumstances as to an *aggregation/marginalization paradox*.

Example 7. *Let T denote the lifetime of a coherent system made with n binary components, whose lifetimes are denoted by $X_1, ..., X_n$. As it is very well known, T can be written in the form*

$$T = \max_{k=1,\dots,K} \min_{h \in P_k} X_h \tag{31}$$

where $P_1, ..., P_K$ are the minimal path sets of the system (see e.g. [2]). Any P_k is a series system and Equation (31) says that the system can be written as a parallel of series system. Using the notation defined in (29), the quantity $\alpha_h^{[P_k]}$ can be seen as the Barlow-Proschan importance index of the component h w.r.t. the k-th minimal path set. See also Remark 2 in Section 2. For two components i and j both belonging to two different minimal path sets $P_{k'}$ and $P_{k''}$, it may in general happen that $\alpha_i^{[P_{k'}]} < \alpha_j^{[P_{k'}]}$ and $\alpha_i^{[P_{k''}]} > \alpha_j^{[P_{k''}]}$. Such a situation may appear rather controversial, taking into account that both $P_{k'}$ and $P_{k''}$ are series systems.

4. A simplifying scenario

Let the random vector $\mathbf{X} = (X_1, \dots, X_n)$ be given. An issue of interest in our study is the identification of the variables that are small according to the following definition

Definition 2. (i) *We say that X_i is* weekly small *with respect to (w.r.t.) $\mathbf{X} = (X_1, X_2, \dots, X_n)$ if $\alpha_i \geq \alpha_j$ for $j = 1, \dots, n$.*
(ii) *We say that X_i is* small *w.r.t. $\mathbf{X}$ if X_i is weekly small w.r.t. $\mathbf{X}$ and there exists j such that $\alpha_i > \alpha_j$.*

Notice that a weakly small element always exists whereas the existence of a small element is not guaranteed. However, no small element exists if and only if $\alpha_1 = \cdots = \alpha_n = 1/n$.

Actually, the quantities $(\alpha_i : i = 1, \ldots, n)$ can be computed using (20). However, for some models, the determination of small variables may be rather complicate.

In this section we analyze situations where the scenario is simplified. First of all, we notice that, in the case when we deal with only two random variables, the property of being weakly small is actually equivalent to stochastic precedence.

Let us now consider the case when $n > 2$. As the following example shows, the circumstance that X_1 stochastically precedes all the variables $X_2, ..., X_n$ does not imply (and is not implied by) the condition that X_1 is small w.r.t. $\mathbf{X}$.

Example 8. *Let X_1, X_2, X_3 be three independent random variables where, for $\varepsilon \in (0, \frac{1}{2})$, X_1 is the degenerate random variable $\frac{1}{2} - \varepsilon$ and where $X_2, X_3 \sim U(0,1)$. For ε small enough, the r.v. X_2 and X_3 are small w.r.t. (X_1, X_2, X_3). Indeed*

$$\mathbb{P}(X_1 < \min\{X_2, X_3\}) = \left(\frac{1}{2} + \varepsilon\right)^2 \cong \frac{1}{4}$$

and

$$\mathbb{P}(X_2 < \min\{X_1, X_3\}) = \mathbb{P}(X_3 < \min\{X_1, X_2\}) = \frac{1}{2}\left[1 - \mathbb{P}(X_1 < \min\{X_2, X_3\})\right] =$$

$$= \frac{1}{2}\left[1 - \left(\frac{1}{2} + \varepsilon\right)^2\right] \cong \frac{3}{8}.$$

On the other hand, we obviously have

$$\mathbb{P}(X_1 < X_2) = \mathbb{P}(X_1 < X_3) = \frac{1}{2} + \varepsilon.$$

Of course, checking the stochastic precedence of a random variable X_1 with respect to a set of other variables is typically much easier than checking the property of it being small. In this respect, the following two definitions are of interest in our analysis.

Given $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, for any $i = 1, \ldots, n$ we denote by $V_{[i]}$ the set of indexes defined as

$$V_{[i]} = \left\{ j \in [n] : \mathbb{P}(X_i < X_j) \geq \frac{1}{2} \right\},$$

and set

$$\mathbf{X}_A = (X_i : i \in A).$$

Definition 3. *We say that X_i is* pair-determined *in $\mathbf{X}$ if for any subset $A \subset V_{[i]}$ the random variable X_i is weakly small w.r.t. $\mathbf{X}_{A \cup \{i\}}$.*

The applied meaning of the above definition can be appreciated by thinking of a betting situation, where $X_1, ..., X_n$ are hitting times until the first occurrence of competing events (such as in horse-racing) and where different players are expected to bet on them. A player, betting on X_i, wins when $X_i = X_{1:n}$, namely it is convenient to bet on X_i when X_i is small w.r.t. $\mathbf{X}$. In such a context, the pair-determined property guarantees that the choice of betting on X_i is justified all the times that only elements X_j with $j \in V_{[i]}$ take part in the competition.

A simple case when all the variables are pair-determined is given in the next Example 9. A case where not all the variables are pair-determined can, on the contrary, be found in Example 8.

Example 9. *Consider a triple X_1, X_2, X_3 such that*

$$X_1 \preceq_{sp} X_2, X_2 \preceq_{sp} X_3, X_3 \preceq_{sp} X_1$$

and the inequalities are understood in a "strict" sense. Thus we have that each single variable is trivially pair-determined since we have

$$V_{[1]} = \{2\}, V_{[2]} = \{3\}, V_{[3]} = \{1\}.$$

Reminding the definition, given above, of the symbol $\preceq_{sp}^{[A]}$, we now present the following

Definition 4. *The vector $\mathbf{X} = (X_1, \dots, X_n)$ is* ordered by pairs *when $X_i \preceq_{sp} X_j$ implies that*

$$X_i \preceq_{sp}^{[A]} X_j$$

for any $A \subset [n]$ and $i, j \in A$.

The ordered by pairs property is indeed rather strong and has a number of implications as shown next.

Proposition 4. *If $\mathbf{X} = (X_1, \dots, X_n)$ is ordered by pairs then the following properties hold:*

a) *For any $i \in [n]$, X_i has the pair-determined property;*
b) *If $X_i \preceq_{sp} X_j$ and $X_j \preceq_{sp} X_k$ then $X_i \preceq_{sp}^{[A]} X_k$, for any $A \subset [n]$ such that $i, k \in A$;*
c) *$X_1 \preceq_{sp} X_j$, for $j = 2, \dots, n$ if and only if X_1 is weakly small w.r.t. $\mathbf{X}_A$ for any $A \subset [n]$ such that $1 \in A$.*

Proof. a) To fix ideas we assume that the variables are indexed in such a way that

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n.$$

Namely, for $i < j$, one has $X_i \preceq_{sp}^{[n]} X_j$. By taking into account the ordered by pairs property for $\mathbf{X}$, one has $X_i \preceq_{sp} X_j$. Thus,

$$V_{[i]} = \{i + 1, \dots, n\}$$

for $i = 1, \dots, n - 1$ and $V_{[n]} = \emptyset$. Fix now A such that $i \in A$ and $A \subset V_{[i]} \cup \{i\}$. By applying again the ordered by pairs property we obtain

$$\alpha_i^{[A]} \geq \alpha_k^{[A]},$$

for any $k \in A$. Thus X_i is weakly small in $X_{A \cup \{i\}}$. Whence we can conclude that $\mathbf{X}$ has the pair–determined property.

b) By hypothesis, $X_i \preceq_{sp} X_j$ and $X_j \preceq_{sp} X_k$. Then the property of ordered by pairs yields

$$X_i \preceq_{sp}^{[B]} X_j, X_j \preceq_{sp}^{[B]} X_k$$

where $B = \{i, j, k\}$. Namely, $\alpha_i^{[B]} \geq \alpha_j^{[B]} \geq \alpha_k^{[B]}$. Then, by applying again the property of ordered by pairs we obtain that for any A with $i, k \in A$ one has $X_i \preceq_{sp}^{[A]} X_k$.

The proof of c) is similar to the above and it can be omitted. □

As an immediate consequence of Proposition 4, one obtains that both non–transitivity and aggregation/marginalization paradoxes can be excluded under the ordered by pairs property. Thus Example 9 shows a case where the latter property fails even if all the variables are pair determined.

Remark 6. *Concerning Barlow-Proschan importance indexes of components in a coherent system, we can cite a simple application of Proposition 4. Let the vector $(X_1, ..., X_n)$ of components' lifetimes have the ordered by pairs property and compare two components i, j. Then the condition $X_i \preceq_{sp} X_j$ implies that the component i has a greater importance index, within any path set, than the component j. See also Example 7.*

The following results give some sufficient conditions for the pair–determined, ordered by pairs, or weakly small properties.

Lemma 2. *Let Y_1, Y_2, Z be independent random variables with $Y_1 \preceq_{st} Y_2$. Then*

$$\mathbb{P}\left(Y_1 \leq \min\left(Y_2, Z\right)\right) \geq \mathbb{P}\left(Y_2 \leq \min\left(Y_1, Z\right)\right).$$

Proof. Denote by f_{Y_i} the marginal density function of Y_i, for $i = 1, 2$ and by f_Z the one of Z.

$$\begin{aligned}\mathbb{P}\left(Y_1 \leq \min\left(Y_2, Z\right)\right) &= \int_0^{+\infty} \mathbb{P}\left(Y_1 \leq \min\left(Y_2, Z\right) | Z = \xi\right) f_Z\left(\xi\right) d\xi = \\ &= \int_0^{+\infty} \mathbb{P}\left(Y_1 \leq \min\left(Y_2, \xi\right)\right) f_Z\left(\xi\right) d\xi, \end{aligned} \tag{32}$$

where the second identity follows by the assumption of stochastic independence.

Similarly

$$\mathbb{P}\left(Y_2 \leq \min\left(Y_1, Z\right)\right) = \int_0^{+\infty} \mathbb{P}\left(Y_2 \leq \min\left(Y_1, \xi\right)\right) f_Z\left(\xi\right) d\xi. \tag{33}$$

Now, for any $\xi > 0$, we can write

$$\mathbb{P}\left(Y_1 \leq \min\left(Y_2, \xi\right)\right) = \int_0^{+\infty} f_{Y_2}\left(y\right) \left[\int_0^{\min(y,\xi)} f_{Y_1}\left(x\right) dx\right] dy, \tag{34}$$

and

$$\mathbb{P}\left(Y_2 \leq \min\left(Y_1, \xi\right)\right) = \int_0^{+\infty} f_{Y_1}\left(x\right) \left[\int_0^{\min(x,\xi)} f_{Y_2}\left(y\right) dy\right] dx. \tag{35}$$

For any $\xi > 0$, the functions

$$\rho_\xi(u) := \int_0^{\min(u,\xi)} f_{Y_1}\left(x\right) dx, \quad \sigma_\xi(u) := \int_0^{\min(u,\xi)} f_{Y_2}\left(x\right) dx$$

are non-decreasing function w.r.t. $u > 0$, and

$$\rho_\xi(u) \geq \sigma_\xi(u)$$

in view of the assumption $Y_1 \preceq_{st} Y_2$. The same assumption then puts us in a position to conclude

$$\mathbb{P}\left(Y_1 \leq \min\left(Y_2, \xi\right)\right) = \int_0^{+\infty} f_{Y_2}\left(u\right) \rho_\xi(u) du \geq$$

$$\geq \int_0^{+\infty} f_{Y_2}\left(u\right) \sigma_\xi(u) du \geq \int_0^{+\infty} f_{Y_1}\left(u\right) \sigma_\xi(u) du = \mathbb{P}\left(Y_1 \leq \min\left(Y_2, \xi\right)\right).$$

Whence the thesis is obtained by recalling equations (32), (33) and by integrating the functions in (34) and (35) with respect to the variable ξ. □

We can now obtain a simple sufficient condition, in the case of independent random variables, ensuring that a single random variable X_1 is simultaneously pair-determined and weakly small in $\mathbf{X}$.

Proposition 5. *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a vector of independent random variables such that $X_1 \preceq_{st.} X_i$, for $i = 2, \ldots, n$. Then X_1 is pair-determined in $\mathbf{X}$ and it is weakly small w.r.t. $\mathbf{X}$. Moreover, if the random variables are not identically distributed then X_1 is small w.r.t. $\mathbf{X}$.*

Proof. Fix j and a set $A \subset [n]$ such that $1, j \in A$. We recall the notation $X_{1:A\setminus\{1,j\}} = \min_{\ell \in A, \ell \neq 1, j} X_\ell$. From the assumption that $X_1 \preceq_{st} X_j$ and from the above Lemma 2, we immediately get

$$\alpha_1^{[A]} = \mathbb{P}(X_1 \leq X_j \wedge X_{1:A\setminus\{1,j\}}) \geq \mathbb{P}(X_j \leq X_1 \wedge X_{1:A\setminus\{1,j\}}) \geq \alpha_j^{[A]}.$$

The proof can be concluded by recalling Definition 3. □

Proposition 6. *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a vector of independent random variables such that $X_{i-1} \preceq_{st.} X_i$, for $i = 2, \ldots, n$. Then $\mathbf{X}$ is ordered by pairs.*

Proof. Fix $i < j$ and a set $A \subset [n]$ such that $i, j \in A$. From the assumption that $X_i \preceq_{st} X_j$ and from the above Lemma 2, we immediately get

$$\alpha_i^{[A]} = \mathbb{P}(X_i \leq X_j \wedge X_{1:A\setminus\{i,j\}}) \geq \mathbb{P}(X_j \leq X_i \wedge X_{1:A\setminus\{i,j\}}) \geq \alpha_j^{[A]}.$$

The proof can be concluded by recalling Definition 4. □

We now pass to consider the case of non-independent random variables and focus attention on the family of the m.c.h.r.'s. First of all we have the following simple conclusion.

Proposition 7. *If $\lambda_1(t|\emptyset) \geq \lambda_j(t|\emptyset)$, for any $t \geq 0$ and $j \geq 2$, then X_1 is weakly small w.r.t. $\mathbf{X} = (X_1, \ldots, X_n)$.*

Proof. It is an immediate consequence of formula (20). □

In our analysis a simplifying condition is the one of "initial-time-homogeneity". We say that $\mathbf{X} = (X_1, \ldots, X_n)$ is *initially time homogeneous* if there exist constants $(\beta_j : j = 1, \ldots, n)$ such that $\lambda_j(t|\emptyset) = \beta_j$.

Remark 7. *Let us focus attention on the independent random variables $Z_1, \dots, Z_n$ introduced in Proposition 2 and considered in equation* (19) *and* (20)*. If $\mathbf{X} = (X_1, \dots, X_n)$ is initially time-homogeneous then $Z_1, \dots, Z_n$ are exponentially distributed.*

In view of the above remark, an obvious corollary of Proposition 7 is the following one

Corollary 1. *If $\mathbf{X} = (X_1, \dots, X_n)$ is initially time homogeneous, then X_j is weakly small w.r.t. $\mathbf{X}$ if and only if $\beta_j = \max_{i=1,\dots,n} \beta_i$;*

The following simple result shows a sufficient condition for the property of being ordered by pairs.

Proposition 8. *Assume that $\mathbf{X}$ is initially-time-homogeneous, with $\beta_i \neq \beta_j$ for all $i \neq j$, and that the condition $\beta_\ell > \beta_j$ implies $\lambda_\ell(t|I; t_1, \dots, t_{|I|}) \geq \lambda_j(t|I; t_1, \dots, t_{|I|})$, for any $I \subset [n]$, $t_1, \dots, t_{|I|}$ and $t > t_{|I|}$. Then the condition of ordered by pairs holds true.*

Proof. Without any loss of generality we can consider that

$$\beta_1 > \beta_2 > \dots > \beta_n. \tag{36}$$

In order to obtain the thesis we must prove that, for any $j < \ell$ and any $A \subset [n]$ such that $\ell, j \in A$, $X_j \preceq^{[A]} X_\ell$. We then consider the "marginal" m.c.h.r function

$$\lambda_i^{[A]}(t|\emptyset) := \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(X_i \in (t, t+\Delta t)|X_{1:A} > t)}{\Delta t}.$$

By taking into account the definition of the $\lambda_i(t|I, t_1, \dots, t_{|I|})$ we obtain

$$\lambda_i^{[A]}(t|\emptyset) = \sum_{I \subset A^c} \int_{[0,t]^{|I|}} \lambda_i(t; I, t_1, \dots, t_{|I|}) f_{\mathbf{X}_I}(t_1, \dots, t_{|I|}|X_{1:A} > t) dt_1 \dots dt_{|I|}. \tag{37}$$

By (37) and (36) we obtain that

$$\lambda_j^{[A]}(t|\emptyset) \geq \lambda_\ell^{[A]}(t|\emptyset)$$

where $j, \ell \in A$ with $j < \ell$. Then by formula (20) follows the thesis. $\square$

As a special case of initially-time-homogeneous models, we find the time-homogeneous load-sharing models mentioned in Section 2. Even if such a condition is very restrictive, this class of models is relevant in that it can still be seen as a generalization of the condition of independence and exponentiality. It can be interesting to specialize to these cases the preceding results about initially-time-homogeneous models.

Acknowledgements. We would like to thank an anonymous Referee for valuable comments and suggestions which, in particular, led us to add Remarks 2, 3, and 6. Most of the results had been presented at the IWAP conference held in Budapest (Hungary), June 2018. E.D.S. and F.S. acknowledge partial support of Ateneo Sapienza Research Projects "*Dipendenza, disuguaglianze e approssimazioni in modelli stocastici*" (2015), "*Processi stocastici: Teoria e applicazioni*" (2016), and "*Simmetrie e Disuguaglianze in Modelli Stocastici*" (2018). Y.M. would like to express his gratitude to coathors for their invitation and support during his visit at Department of Mathematics, Sapienza University of Rome, in January 2018.

References

[1] M. A. Arcones and F. J. Samaniego. On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint. *Ann. Statist.*, 28(1):116–150, 2000.

[2] R. E. Barlow and F. Proschan. *Statistical theory of reliability and life testing*. Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London, 1975. Probability models, International Series in Decision Processes, Series in Quantitative Methods for Decision Making.

[3] C. R. Blyth. Some probability paradoxes in choice from among random alternatives. *J. Amer. Statist. Assoc.*, 67:366–373, 373–381, 1972.

[4] C. R. Blyth. Simpson's paradox and mutually favorable events. *J. Amer. Statist. Assoc.*, 68:746, 1973.

[5] P. J. Boland, H. Singh, and B. Cukic. The stochastic precedence ordering with applications in sampling and testing. *J. Appl. Probab.*, 41(1):73–82, 2004.

[6] S. J. Brams and P. C. Fishburn. *Approval voting*. Birkhäuser, Boston, Mass., 1983.

[7] R. Chen and A. Zame. On fair coin-tossing games. *J. Multivariate Anal.*, 9(1):150–156, 1979.

[8] E. De Santis, F. Fantozzi, and F. Spizzichino. Relations between stochastic orderings and generalized stochastic precedence. *Probab. Engrg. Inform. Sci.*, 29(3):329–343, 2015.

[9] E. De Santis and F. Spizzichino. First occurrence of a word among the elements of a finite dictionary in random sequences of letters. *Electron. J. Probab.*, 17:no. 25, 9, 2012.

[10] E. De Santis and F. Spizzichino. Some sufficient conditions for stochastic comparisons between hitting times for skip-free Markov chains. *Methodol. Comput. Appl. Probab.*, 18(4):1021–1034, 2016.

[11] E. De Santis and F. Spizzichino. Usual and stochastic tail orders between hitting times for two Markov chains. *Appl. Stoch. Models Bus. Ind.*, 32(4):526–538, 2016.

[12] F. Durante and R. Foschi. Dependence of exchangeable residual lifetimes subject to failure. *Appl. Math. Comput.*, 235:502–511, 2014.

[13] M. Finkelstein, N. K. Hazra, and J. H. Cha. On optimal operational sequence of components in a warm standby system. *J. Appl. Probab.*, 55(4):1014–1024, 2018.

[14] P. C. Fishburn. *The theory of social choice*. Princeton University Press, Princeton, N.J., 1973.

[15] L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, 30(2):183–208, 1981.

[16] S. Iyer. The Barlow-Proschan importance and its generalizations with dependent components. *Stochastic Process. Appl.*, 42(2):353–359, 1992.

[17] S.-Y. R. Li. A martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Ann. Probab.*, 8(6):1171–1176, 1980.

[18] J.-L. Marichal and P. Mathonet. On the extensions of Barlow-Proschan importance index and system signature to dependent lifetimes. *J. Multivariate Anal.*, 115:48–56, 2013.

[19] P. Miziuła and J. Navarro. Birnbaum importance measure for reliability systems with dependent components. *IEEE Transactions on Reliability*, 68(2):439–450, 2019.

[20] J. Navarro and F. Durante. Copula-based representations for the reliability of the residual lifetimes of coherent systems with dependent components. *J. Multivariate Anal.*, 158:87–102, 2017.

[21] J. Navarro and R. Rubio. Comparisons of coherent systems using stochastic precedence. *TEST*, 19(3):469–486, 2010.

[22] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[23] T. Rychlik and F. Spizzichino. Probabilistic properties of components and systems lifetimes under linear breakdown models. *Submitted*.

[24] D. G. Saari. A chaotic exploration of aggregation paradoxes. *SIAM Rev.*, 37(1):37–52, 1995.

[25] R. P. Savage, Jr. The paradox of nontransitive dice. *Amer. Math. Monthly*, 101(5):429–436, 1994.

[26] M. Scarsini and F. Spizzichino. Simpson-type paradoxes, dependence, and ageing. *J. Appl. Probab.*, 36(1):119–131, 1999.

[27] M. Shaked and J. G. Shanthikumar. Dynamic construction and simulation of random vectors. In *Topics in statistical dependence (Somerset, PA, 1987)*, volume 16 of *IMS Lecture Notes Monogr. Ser.*, pages 415–433. Inst. Math. Statist., Hayward, CA, 1990.

[28] M. Shaked and J. G. Shanthikumar. *Stochastic orders and their applications*. Probability and Mathematical Statistics. Academic Press, Inc., Boston, MA, 1994.

[29] M. Shaked and J. G. Shanthikumar. Multivariate conditional hazard rate functions—an overview. *Appl. Stoch. Models Bus. Ind.*, 31(3):285–296, 2015.

[30] E. H. Simpson. The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. Ser. B.*, 13:238–241, 1951.

[31] F. Spizzichino. *Subjective probability models for lifetimes*, volume 91 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[32] F. Spizzichino. Some remarks on multivariate conditional hazard rates and dependence modeling. In *Recent advances in probability and statistics*, volume 12 of *Lect. Notes Semin. Interdiscip. Mat.*, pages 255–271. Semin. Interdiscip. Mat. (S.I.M.), Potenza, 2015.

[33] F. Spizzichino. Reliability, signature, and relative quality functions of systems under time-homogeneous load-sharing models. *Appl. Stoch. Models Bus. Ind.*, 35(2):158–176, 2019.

[34] H. Steinhaus and S. Trybula. On a paradox in applied probabilities. *Bull. Acad. Polon. Sci*, 7(67-69):108, 1959.

[35] S. Trybula. Cyclic random inequalities. *Zastos. Mat.*, 10:123–127, 1969.

University of Rome La Sapienza, Department of Mathematics Piazzale Aldo Moro, 5, I-00185, Rome, Italy
E-mail address: `desantis@mat.uniroma1.it`

Department of Mathematics and Statistics University of Maryland, Baltimore County 1000 Hilltop Circle Baltimore, MD 21250
E-mail address: `yaakovm@umbc.edu`

University of Rome La Sapienza Piazzale Aldo Moro, 5, I-00185, Rome, Italy
E-mail address: `fabio.spizzichino@fondazione.uniroma1.it`