Università Degli Studi Di Napoli
Federico II

Dottorato di Ricerca

in

Matematica e Applicazioni
XXXV ciclo

New Developments in Reliability Theory: from Measures of Uncertainty to Load-Sharing Models through Aging Properties and Failure Times of Systems

Francesco Buono

Anno Accademico 2021/2022

Contents

Introduction

The reliability theory is a branch of applied mathematics which studies the lifetimes of devices, items or organisms by regarding them as random variables. The first attempts of using statistical methods in the context of quality control of industrial processes can be found at the beginning of the 1930s although the birth of the reliability theory dates back to the World War II during which the quality of missiles was analyzed by a probabilistic point of view. After World War II, the development of new technologies and more complicated systems has contributed to further applications of the reliability theory around the world which has since become a discipline of increasing interest. For further details on the historical background of the reliability theory one may refer to Hoyland and Rausand [55].

In this work, many topics related to the reliability theory are addressed. The topics can be traced back to three macro-areas: the measures of uncertainty, the hazard and the reversed hazard rate functions and the problem of predictions from censored data. Certainly, there will be topics that will be involved on several occasions, such as stochastic orders and coherent systems, aimed at linking the different macro-areas together. More precisely, this work is organized as follows.

In Chapter 1, some fundamental notions of the reliability theory are presented. The concepts introduced there are involved and used throughout all the other chapters. In particular, the definitions and some properties of the hazard rate function and of the reversed hazard rate function are given. Then, the definitions and the interrelations among some of the most important stochastic orders are presented. Finally, some useful definitions and properties of conditional distributions and coherent systems are studied.

In Chapter 2, it is performed the study of some new measures of uncertainty in the classical probability theory. Before doing this, some well-known formulations of entropy are recalled and a particular attention is devoted to the cumulative versions of entropy and their connection with the moments of order statistics. This connection has been introduced in the paper entitled ***On cumulative entropies in terms of moments of order statistics*** by Narayanaswamy Balakrishnan, Francesco Buono and Maria Longobardi, published in 2022 in *Methodology and Computing in Applied Probability*, vol. 24, pp. 345–359. Then, the chapter proceeds by focusing attention on the extropy, the measure of uncertainty dual of entropy, and by defining some new formulations of it. In particular, the past extropy, the weighted extropy, the interval extropy and Tsallis extropy are introduced and studied. The results about these measures of uncertainty are based on the papers ***On weighted extropies*** by Narayanaswamy Balakrishnan, Francesco Buono and Maria Longobardi, published in 2022 in *Communications in Statistics - Theory and Methods*, vol. 51, pp. 6250–6267, ***On Tsallis extropy with an application to pattern recognition*** by Narayanaswamy Balakrishnan, Francesco Buono and Maria Longobardi, published in 2022 in *Statistics & Probability Letters*, vol. 180, p. 109241, ***Interval extropy and weighted interval extropy*** by Francesco Buono, Osman Kamari and Maria Longobardi published in 2021 in *Ricerche di matematica*, DOI: 10.1007/s11587-021-00678-x, and ***On extropy of past lifetime distribution*** by Osman Kamari and Francesco Buono published in 2021 in *Ricerche di Matematica*, vol. 70, pp. 505–515. Finally, some properties of the varentropy are introduced and a study on the past varentropy is performed. This last part of the chapter is based on the paper ***Varentropy of past lifetimes*** by Francesco Buono, Maria Longobardi and Franco Pellerey, published in 2022 in *Mathematical Methods of Statistics*, vol. 31, pp. 57–73.

Chapter 3 is dedicated to the study of some new measures of uncertainty in the context of Dempster-Shafer theory of evidence, which is a generalization of the classical probability theory. First, some general concepts of this theory are revised and then new measures of discrimination are presented, namely Deng extropy, fractional Deng entropy and extropy and the unified formulation of entropy. Finally, these measures of uncertainty are applied to classification problems and several examples are given. Most of the results presented in this chapter are based on the papers ***A unified formulation of entropy and its application*** by Narayanaswamy Balakrishnan, Francesco Buono and Maria Longobardi, published in 2022 in *Physica A: Statistical Mechanics and its Applications*, vol. 596, p. 127214, ***A dual measure of uncertainty: The Deng Extropy*** by Francesco Buono and Maria Longobardi published in 2020 in *Entropy*, vol. 22, p. 582, and ***Fractional Deng entropy and extropy and some applications*** by Mohammad Reza Kazemi, Saeid Tahmasebi, Francesco Buono and Maria Longobardi published in 2021 in *Entropy*, vol. 23, p. 623.

In Chapter 4, the extensions of the hazard rate function and of the reversed hazard rate function to the multivariate case are analyzed. More precisely, these functions are defined by taking into account the possibility of observing a dynamic history up to a fixed time and so they are known as multivariate conditional hazard and reversed hazard rate functions. In the first part of the chapter, the definition and some properties of the multivariate conditional hazard rate functions are presented. Then, the main properties of a well-known dependence model, namely the Load-Sharing model with its time-homogeneous version, are pointed out. In the second part of the chapter, based on the paper ***Multivariate reversed hazard rates and inactivity times of systems*** by Francesco Buono, Emilio De Santis, Maria Longobardi and Fabio Spizzichino, published in 2022 in *Methodology and Computing in Applied Probability*, vol. 24, pp. 1987–2008, the multivariate conditional reversed hazard rate functions are defined and studied.

Chapter 5 is devoted to the study of the aging intensity functions. First, the classical concepts of aging and reversed aging intensity functions are recalled and a recent generalization of the aging intensity function is also presented. Then, based on the paper ***On generalized reversed aging intensity functions*** by Francesco Buono, Maria Longobardi and Magdalena Szymkowiak, published in 2022 in *Ricerche di matematica*, vol. 71, pp. 85–108, the generalized version of the reversed aging intensity function is introduced, some of its properties are studied and stochastic comparisons are performed. The last part of the chapter is strictly related to the developments of Chapter 4. In fact, the multivariate conditional version of the aging intensity function is studied and some interesting properties are given in the case of the Load-Sharing models. Most of the results presented in the last part of the chapter, are given in the paper ***Multivariate conditional aging intensity functions and load-sharing models*** by Francesco Buono, published in 2022 *Hacettepe Journal of Mathematics and Statistics*, vol. 51, pp. 1710–1722.

Chapter 6 is based on the studies that I have carried out during my visiting period at Universidad de Murcia under the supervision of Prof. Jorge Navarro. The main task of this chapter is the prediction of future failures from censored data. In the first part, it is performed a study based on quantile regression techniques with a special regard to the Proportional Hazard Rate model. Most of the results related to this analysis are given in the paper ***Predicting future failure times by using quantile regression*** by Jorge Navarro and Francesco Buono, published in 2022 in *Metrika*, DOI: 10.1007/s00184-022-00884-z. In the second part, the problem of predictions is studied for a generalization of the Load-Sharing model, namely the order dependent Load-Sharing model. Furthermore, an algorithm to simulate random vectors from it is presented. The results related to the simulations and predictions of the order dependent Load-Sharing model are based on the paper ***Simulations and predictions of future values in the time-homogeneous load-sharing model*** by Francesco Buono and Jorge Navarro which is submitted for publication.

Finally, I would like to thank my PhD advisor Prof. Maria Longobardi for guiding and supporting me over these years.

Chapter 1

Reliability theory

The reliability of an item can be treated as its ability to perform a required function, under given environmental and operational conditions, for a stated period of time. There are several tools to measure the reliability, depending on the particular case, and they are strictly related to some functions of interest in this field. In fact, the reliability of an item can be expressed in terms of the number of failures per time unit, the mean time to failure, the probability that the item is available at time t or the probability that the item does not fail in the interval $(0, t]$ and so on. Those aspects can be easily analyzed through some functions of interest such as the survival and the hazard rate functions.

With the purpose of studying the reliability of a device, it is possible to proceed in several and different ways. In fact, it is possible to regard the entire device as a unique structure characterized by a random lifetime and then study the corresponding distribution properties, or to study how the different components of the system affect its behavior. In the latter case, the assumption of the hypothesis of coherence, i.e., the study of coherent systems, brings out some interesting properties. Moreover, dealing with the random lifetimes of the devices, the measures of uncertainty and discrimination find many key applications.

In this chapter, some primary functions of reliability theory, as the hazard and reversed hazard rate functions, are introduced and some of their properties are discussed. In relation with inspection times, some random variables of interest, as the residual and past lifetimes and the inactivity time, are then discussed. Furthermore, useful concepts related to coherent systems and stochastic orders are displayed.

1.1 The hazard rate function

Let X be a non-negative absolutely continuous random variable with probability density function (pdf) f, cumulative distribution function (cdf) F and survival function $\overline{F}$. In reliability theory, the hazard rate function of X has found many key applications. The hazard rate function (or failure rate function) of X at x, for x such that $\overline{F}(x) > 0$, is defined as

$$\begin{aligned} r(x) &= \lim_{\Delta x \to 0^+} \frac{\mathbb{P}(x < X \leq x + \Delta x | X > x)}{\Delta x} \\ &= \frac{1}{\overline{F}(x)} \lim_{\Delta x \to 0^+} \frac{\mathbb{P}(x < X \leq x + \Delta x)}{\Delta x} = \frac{f(x)}{\overline{F}(x)}, \end{aligned} \tag{1.1}$$

where the last equality follows from the assumption of absolute continuity. In fact, it has to be noted that the hazard rate function can be defined also without the assumption of absolute continuity, in particular for discrete distributions, and not necessarily for non-negative random variables. Here, with the purpose of involving this function in reliability analysis, we will focus on the case in which the random variable describes a lifetime and hence we will require the non-negativity and absolute continuity assumptions. The hazard rate $r(x)$ can be interpreted as the rate of instantaneous failure occurring immediately after the time point x, given that the unit has survived up to time x.

The hazard rate function characterizes the distribution in the sense that it uniquely determines the survival function, and then the cdf and the pdf, as stated in the following theorem, see Barlow and Proschan [11].

Theorem 1.1. *Let X be a non-negative absolutely continuous random variable with survival function $\overline{F}$ and hazard rate function r. Then, $\overline{F}$ and r are related by*

$$\overline{F}(x) = \exp\left(-\int_0^x r(u)\ du\right), \quad x \in (0, +\infty). \tag{1.2}$$

It is well known that the exponential distribution is the unique absolutely continuous probability distribution with the lack of memory property. This means that if $X \sim Exp(\lambda)$, $\lambda > 0$, then, for all $x, y > 0$,

$$\mathbb{P}(X > x + y | X > y) = \mathbb{P}(X > x).$$

By the lack of memory property, every instant is like the beginning of a new random period, which has the same distribution regardless of how much time has already elapsed. This property of the exponential distribution is explained also in terms of the hazard rate function. In fact, the exponential distribution is the unique absolutely continuous probability distribution with a constant hazard rate function, $r(x) = \lambda$, $x > 0$. Then, the rate of instantaneous failure does not change by time and the failures are not caused by aging or degradation.

In reliability theory, it is of wide interest to study the monotonic properties of the hazard rate function. For instance, there are several materials or devices which wear out with time and hence the class of distributions with an increasing hazard rate has broad applications. Moreover, also the class of distributions with decreasing hazard rate is of high interest due to the debugging of complex systems and the phenomenon of work hardening of some materials. Hence, it is useful to give the following definitions.

Definition 1.1. A random variable X is IFR (increasing failure rate) if, and only if, the hazard rate function $r(x)$ is increasing in x.

Definition 1.2. A random variable X is DFR (decreasing failure rate) if, and only if, the hazard rate function $r(x)$ is decreasing in x.

It has to be noted that, without requiring a strictly monotonic behavior, the exponential distribution belongs to both classes of distributions, IFR and DFR. In the following theorem, alternative ways to establish if a random variable is IFR or DFR are given by the use of the survival function.

Theorem 1.2. *Let X be a non-negative absolutely continuous random variable with survival function $\overline{F}$ and hazard rate function r. Then, X is IFR (DFR) if, and only if, $\log \overline{F}(x)$ is concave (convex) in x.*

The distributions in the classes IFR and DFR hold some nice and useful properties. For instance, an IFR random variable has finite moments of all orders and a DFR distribution has decreasing probability density function. The following theorem provides bounds based on quantiles for the survival function of IFR distributions.

Theorem 1.3. *Let X be IFR with survival function $\overline{F}$, cdf F and let ξ_p be such that $F(\xi_p) = p$, i.e., ξ_p is the p-th quantile. Then,*

$$\begin{aligned} &\overline{F}(x) \geq \exp(-\alpha x), \ \ \textit{if } 0 < x \leq \xi_p, \\ &\overline{F}(x) \leq \exp(-\alpha x), \ \ \textit{if } x \geq \xi_p, \end{aligned}$$

with $\alpha = -\frac{\log(1-p)}{\xi_p}$.

The following theorem gives an important lower bound for the survival function of X in the IFR case. In particular, the lower bound is given in terms of the survival functions of an exponential random variable with the same mean of X and of a degenerate distribution concentrating at the mean of X.

Theorem 1.4. *Let X be a non-negative absolutely continuous random variable with survival function $\overline{F}$ and finite mean μ. If X is IFR, then*

$$\begin{aligned} &\overline{F}(x) \geq \exp\left(-\frac{x}{\mu}\right), \ \ \textit{if } 0 < x < \mu, \\ &\overline{F}(x) \geq 0, \ \ \textit{if } x \geq \mu. \end{aligned}$$

Since not each distribution exhibits a defined monotonic behavior, it is of interest to study if such a behavior is presented at least in average and hence the following definitions are given.

Definition 1.3. A non-negative random variable X is IFRA (increasing in failure rate average) if, and only if, $\frac{\int_0^x r(t)\ dt}{x}$ is increasing in $x > 0$.

Definition 1.4. A non-negative random variable X is DFRA (decreasing in failure rate average) if, and only if, $\frac{\int_0^x r(t)\ dt}{x}$ is decreasing in $x > 0$.

Another important class of distributions in reliability theory is given by the bathtub shaped failure rate (BFR) ones. The name of such distributions comes from the shape of the graph of the function r. In fact, the hazard rate function decreases at first, then remains constant for a period and finally it increases. Those distributions are of wide interest since these different periods correspond to three phases of the life of the system. At first, there is the early life in which failures are caused by manufacturing defects. Then, the useful life in which failures are caused by chance. Finally, the wear out phase in which the hazard is increasing due to the aging of the system. In the following, the formal definition of BFR distributions is given and some properties are presented, see Lai et al. [68] for further details.

Definition 1.5. A non-negative random variable X with hazard rate function r is said to be BFR if there exists $x_0 > 0$, named change point, such that r is non-increasing in $(0, x_0)$ and non-decreasing in $(x_0, +\infty)$

In the above definition, the part in which the hazard rate is constant does not emerge explicitly. Hence, it can be useful to consider the following equivalent definition.

Definition 1.6. A non-negative random variable X with hazard rate function r is said to be BFR if there exist $0 \leq x_1 \leq x_2 < +\infty$ such that

(1) r is strictly decreasing in $(0, x_1)$;

(2) r is constant in (x_1, x_2);

(3) r is strictly increasing in $(x_2, +\infty)$.

Remark 1.1. For $x_1 = x_2 = 0$, BFR distributions reduce to IFR ones. On the contrary, if $x_1 = x_2 \to +\infty$, BFR distributions reduce to DFR ones. Furthermore, if $x_1 = x_2$, i.e., the interval in which r is constant degenerates, it is more appropriate to talk about U-shaped hazard rate functions.

We give a further equivalent definition of BFR distributions based on the connection between the concepts of IFR and DFR and the concavity or convexity of the function $\log \overline{F}$ expressed in Theorem 1.2.

Definition 1.7. A non-negative random variable X with survival function $\overline{F}$ is said to be BFR if there exists $x_0 > 0$ such that $-\log \overline{F}(x)$ is concave in $(0, x_0)$ and convex in $(x_0, +\infty)$.

In the following proposition, it is presented a comparison between the survival functions of a BFR distribution and an exponential distribution. It is a comparison in the usual stochastic order which will be introduced in Section 1.3.

Proposition 1.1. *Let X be a non-negative BFR random variable with survival function $\overline{F}$ and change point x_0 and let Y be exponentially distributed with mean $\frac{1}{r(x_0)}$ with survival function $\overline{G}$. Then, $\overline{F}(x) \leq \overline{G}(x)$, for all $x \in (0, +\infty)$.*

Example 1.1. Let X be a non-negative random variable with quadratic hazard rate function

$$r(x) = \alpha + \beta x + \gamma x^2,$$

where $\alpha \geq 0$, $\beta < 0$ and $\gamma > 0$. X is BFR, and more precisely with U-shaped hazard rate function, since r is decreasing in $\left(0, -\frac{\beta}{2\gamma}\right)$ and increasing in $\left(-\frac{\beta}{2\gamma}, +\infty\right)$. Furthermore, the random variable X^* with hazard rate function $r^*(x) = \exp[r(x)]$ is BFR since its failure rate is obtained by composing r with a strictly increasing function.

The above example shows how it can be important to consider random variables which exhibit a relation among their hazard rate functions. In this perspective, one the most important classes of models is given by the proportional hazard rate models (PHRM) which were introduced by Cox [31]. The family of absolutely continuous random variables $\{X_\theta : \theta > 0\}$ follows a proportional hazard rate model (PHRM) if there exists a non-negative random variable X with survival function $\overline{F}$ and pdf f such that the survival function $\overline{F}_\theta$ of X_θ is expressed as

$$\overline{F}_\theta(x) = \left[\overline{F}(x)\right]^\theta, \quad x > 0.$$

Furthermore, the corresponding probability density functions are related by

$$f_\theta(x) = \theta f(x) \left[\overline{F}(x)\right]^{\theta - 1}, \quad x > 0,$$

and hence the hazard rate function of X_θ is expressed as

$$r_\theta(x) = \theta r(x), \quad x > 0, \tag{1.3}$$

from which the model name follows. From the relation in (1.3), it readily follows that the hazard rate functions of the random variables X_θ, $\theta > 0$, have the same monotonic properties of r. Hence, if X is IFR, DFR or BFR, X_θ will belong to the same aging class.

1.2 The reversed hazard rate function

The reversed hazard rate function is in a certain sense the dual function of the hazard rate. It is related to the probability which the failure occurs immediately before the point of evaluation given that the unit does not survive up to that point. For a non-negative and absolutely continuous random variable X with pdf f and cdf F, the reversed hazard rate function is defined as

$$\begin{aligned} q(x) &= \lim_{\Delta x \to 0^+} \frac{\mathbb{P}(x - \Delta x < X \leq x | X \leq x)}{\Delta x} \\ &= \frac{1}{F(x)} \lim_{\Delta x \to 0^+} \frac{\mathbb{P}(x - \Delta x < X \leq x)}{\Delta x} = \frac{f(x)}{F(x)}, \end{aligned} \tag{1.4}$$

where the last equality follows by the assumption of absolute continuity.

As the hazard rate function, the reversed hazard rate function is a useful tool since it characterizes the distribution uniquely as presented in the following theorem.

Theorem 1.5. *Let X be a non-negative and absolutely continuous random variable with cdf F and reversed hazard rate q. Then, F and q are related by*

$$F(x) = \exp\left(-\int_x^{+\infty} q(t)\ dt\right),\ x \in (0, +\infty). \tag{1.5}$$

It follows that, unlike the hazard rate fucntion, the reversed hazard rate function of a non-negative random variable cannot be constant over the entire support.

The hazard rate and the reversed hazard rate functions are strictly related and it is possible to obtain one from the other as presented in the following theorem by Finkelstein [45].

Theorem 1.6. *Let X be a non-negative and absolutely continuous random variable with hazard rate function r and reversed hazard rate function q. Then, r and q are related by*

$$q(x) = \frac{r(x)}{\exp\left(\int_0^x r(t)dt\right) - 1}, \quad x \in (0, +\infty). \tag{1.6}$$

From (1.6), it readily follows the reversed hazard rate function of $X \sim Exp(\lambda)$,

$$q(x) = \frac{\lambda}{\exp(\lambda x) - 1}, \quad x > 0,$$

which is decreasing in x. Moreover, from (1.6), it can be argued that if the hazard rate function is decreasing then the reversed hazard rate function is also decreasing. About the monotonicity of the reversed hazard rate function it is essential to mention the following result given by Block et al. [19].

Theorem 1.7. *There does not exist a non-negative random variable having increasing reversed hazard rate function on its interval of support.*

As a generalization of the PHRM model, now we present the proportional reversed hazard rate model (PRHRM) introduced and studied by Gupta et al. [52]. The family of absolutely continuous random variables $\{X_\theta : \theta > 0\}$ follows a proportional reversed hazard rate model if there exists a non-negative random variable X with cdf F, pdf f and reversed hazard rate function q such that the cdf of X_θ is expressed as

$$F_\theta(x) = [F(x)]^\theta, \quad x > 0.$$

Then, the pdf of X_θ is related to the one of X by

$$f_\theta(x) = \theta[F(x)]^{\theta-1}f(x),$$

and the reversed hazard rate of X_θ is expressed as

$$q_\theta(x) = \theta q(x), \quad x > 0. \tag{1.7}$$

In the following, two important functions related to the hazard rate and the reversed hazard rate functions are introduced, for further properties on these functions see [86]. Let X be a non-negative random variable. The log-odds of X is defined for $x > 0$ as the natural logarithm of the ratio between the cdf and the survival function of X,

$$LO_X(x) = \log \frac{F(x)}{\overline{F}(x)}. \tag{1.8}$$

Furthermore, the log-odds rate of X is defined for $x > 0$ as the derivative of the log-odds of X,

$$LOR_X(x) = \frac{d}{dx} LO_X(x).$$

If X is an absolutely continuous random variable, then the log-odds rate is expressed in different ways by involving the hazard rate and reversed hazard rate functions

$$LOR_X(x) = \frac{f(x)}{F(x)\overline{F}(x)} = \frac{r(x)}{F(x)} = \frac{q(x)}{\overline{F}(x)}, \tag{1.9}$$

and the following relation holds

$$r(x) + q(x) = LOR(x).$$

1.3 Stochastic orders

In this section, some of the most important stochastic orders are presented. For more details on these concepts, one can see Shaked and Shanthikumar [106]. The stochastic orders are useful tools to compare random variables from different perspectives. The simplest way of comparing two distributions is by the comparison of the associated means. However, such a comparison is based on only two numbers and often it is not so much informative. Moreover, the mean sometimes does not exist and there may be situations in which two distributions have the same mean. Furthermore, we often have more information and it is fruitful to use them. In the following definition, some of the most important classical stochastic orders are recalled.

Definition 1.8. Let X and Y be non-negative absolutely continuous random variables with survival functions $\overline{F}$, $\overline{G}$, pdf's f, g, hazard rate functions r_X, r_Y and reversed hazard rate functions q_X, q_Y, respectively. Then,

1. X is smaller than Y in the usual stochastic order, denoted by $X \leq_{st} Y$, if $\overline{F}(x) \leq \overline{G}(x)$ for all $x > 0$ or, equivalently, $F(x) \geq G(x)$ for all $x > 0$;

2. X is smaller than Y in the likelihood ratio ordering, denoted by $X \leq_{lr} Y$, if $\frac{g(x)}{f(x)}$ is increasing in $x > 0$;

3. X is smaller than Y in the hazard rate order, denoted by $X \leq_{hr} Y$, if $r_X(x) \geq r_Y(x)$ for all $x > 0$;

4. X is smaller than Y in the reversed hazard rate order, denoted by $X \leq_{rh} Y$, if $q_X(x) \leq q_Y(x)$ for all $x > 0$;

5. X is smaller than Y in the dispersive order, denoted by $X \leq_{disp} Y$, if $f(F^{-1}(u)) \geq g(G^{-1}(u))$ for all $u \in (0,1)$, where F^{-1} and G^{-1} are right continuous inverses of F and G, respectively;

6. X is smaller than Y in the convex transform order, denoted by $X \leq_c Y$, if $G^{-1}(F(x))$ is a convex function on the support of X;

7. X is smaller than Y in the star order, denoted by $X \leq_* Y$, if $\frac{G^{-1}F(x)}{x}$ is increasing in $x > 0$;

8. X is smaller than Y in the superadditive order, denoted by $X \leq_{su} Y$, if $G^{-1}(F(x+t)) \geq G^{-1}(F(x)) + G^{-1}(F(t))$ for $x > 0, t > 0$;

9. X is smaller than Y in the log-odds rate order, denoted by $X \leq_{LOR} Y$, if $LOR_X(x) \geq LOR_Y(x)$ for all $x > 0$.

Different stochastic orders are often related each other in the sense that a stochastic order may imply a different one without additional assumptions, whereas sometimes specific conditions are required to obtain an implication. In the following propositions, some relations among the stochastic orders given in Definition 1.8 are presented.

Proposition 1.2. *Let X and Y be non-negative and absolutely continuous random variables. Then,*

$$X \leq_{hr} Y \quad \Longrightarrow \quad X \leq_{st} Y,$$
$$X \leq_{rh} Y \quad \Longrightarrow \quad X \leq_{st} Y.$$

Proposition 1.3. *Let X and Y be non-negative and absolutely continuous random variables. Then,*

$$X \leq_{lr} Y \quad \Longrightarrow \quad X \leq_{hr} Y,$$
$$X \leq_{lr} Y \quad \Longrightarrow \quad X \leq_{rh} Y,$$

and, by Proposition 1.2, $X \leq_{lr} Y \implies X \leq_{st} Y$.

Proposition 1.4. *Let X and Y be non-negative and absolutely continuous random variables. Then,*

$$X \leq_{disp} Y \quad \Longrightarrow \quad X \leq_{st} Y.$$

Proposition 1.5. *Let X and Y be non-negative and absolutely continuous random variables. Then,*

$$X \leq_{c} Y \implies X \leq_{*} Y \implies X \leq_{su} Y.$$

Without any additional assumption, the reversed implications of the relations given in Proposition 1.2 cannot be obtained. In order to obtain such a result, it is necessary to give an assumption related to the log-odds rate order, as presented in the following proposition.

Proposition 1.6. *Let X and Y be non-negative and absolutely continuous random variables such that $X \leq_{LOR} Y$. Then,*

$$X \leq_{st} Y \iff X \leq_{hr} Y.$$

Furthermore, if $X \geq_{LOR} Y$, then

$$X \leq_{st} Y \iff X \leq_{rh} Y.$$

By taking in consideration some monotonic properties of the hazard rate function, it is possible to obtain some connections between the hazard rate and the dispersive orders.

Proposition 1.7. *Let X and Y be non-negative and absolutely continuous random variables.*

(a) If $X \leq_{hr} Y$ and X or Y is DFR, then $X \leq_{disp} Y$;

(b) If $X \leq_{disp} Y$ and X or Y is IFR, then $X \leq_{hr} Y$.

The star order and the dispersive order are related as presented in the following proposition.

Proposition 1.8. *Let X and Y be non-negative and absolutely continuous random variables. Then,*

$$X \leq_{*} Y \iff \log X \leq_{disp} \log Y,$$

or, equivalently,

$$X \leq_{disp} Y \iff e^X \leq_{*} e^Y.$$

By considering an additional assumption, the superadditive order (and then, by virtue of Proposition 1.5, also the star and the convex orders) implies the dispersive order as presented in the following proposition.

Proposition 1.9. *Let X and Y be non-negative and absolutely continuous random variables such that $X \leq_{st} Y$. Then,*

$$X \leq_{su} Y \implies X \leq_{disp} Y.$$

1.4 Residual and past lifetimes, inactivity time

In this section, some useful conditional distributions are recalled and some of their properties are discussed. These distributions are related to the possibility of making an inspection at a fixed time. Suppose that X is the random lifetime of a device. It may happen that making an inspection at time t, the device is found working and then it is of interest to study the corresponding residual lifetime $X_t = [X - t|X > t]$. It is a non-negative random variable whose cdf and survival function are expressed as

$$\begin{aligned} F_{X_t}(x) &= \mathbb{P}(X - t \leq x|X > t) = \frac{F_X(x+t) - F_X(t)}{\overline{F}_X(t)}, \quad x \in (0, +\infty) \\ \overline{F}_{X_t}(x) &= \frac{\overline{F}_X(x+t)}{\overline{F}_X(t)}, \quad x \in (0, +\infty), \end{aligned}$$

where $F_X(\cdot)$ and $\overline{F}_X(\cdot)$ are the cdf and survival function of X, respectively. Moreover, the pdf and the hazard rate function of X_t are given by

$$\begin{aligned} f_{X_t}(x) &= \frac{f_X(x+t)}{\overline{F}_X(t)}, \quad x \in (0, +\infty) \\ r_{X_t}(x) &= \frac{f_X(x+t)}{\overline{F}_X(x+t)} = r_X(x+t), \quad x \in (0, +\infty), \end{aligned}$$

where $f_X(\cdot)$ and $r_X(\cdot)$ are the pdf and hazard rate function of X, respectively.

Since X_t is a random variable, it is possible to evaluate its mean that is an important concept in reliability theory, known as mean residual life, i.e., $mrl(t) = \mathbb{E}(X_t)$. If X is a non-negative random variable with finite mean, the mean residual life can be expressed as

$$mrl(t) = \frac{\int_t^{+\infty} \overline{F}_X(x)\, dx}{\overline{F}_X(t)}. \tag{1.10}$$

Moreover, the mean residual life and the hazard rate function are connected through the following relation

$$mrl(t) = \int_t^{+\infty} \exp\left(-\int_t^x r_X(u)\, du\right) dx, \tag{1.11}$$

and, by using (1.10), it readily follows

$$r_X(t) = \frac{mrl'(t) + 1}{mrl(t)}, \tag{1.12}$$

see Ebrahimi [43] for further details. Furthermore, it is possible to introduce a stochastic order by comparing the mean residual lives, as presented in the following definition (see Shaked and Shanthikumar [106]).

Definition 1.9. Let X and Y be non-negative absolutely continuous random variables with mean residual life functions mrl_X and mrl_Y, respectively. Then, X is smaller than Y in the mean residual life order, denoted by $X \leq_{mrl} Y$, if $mrl_X(t) \leq mrl_Y(t)$ for all $t > 0$.

By using (1.11), a connection between the hazard rate and the mean residual life orders follows.

Proposition 1.10. *Let X and Y be non-negative and absolutely continuous random variables. Then,*

$$X \leq_{hr} Y \implies X \leq_{mrl} Y.$$

Conversely, it may happen that at the time of inspection t the device is out of order. In this scenario, it is of interest to study two random variables: the past lifetime and the inactivity time. They are strictly connected but essentially different since they focus attention on two different aspects. In fact, the past lifetime is defined as ${}_tX = [X|X \leq t]$ and the inactivity time as $X_{[t]} = [t - X|X \leq t]$, so that the inactivity time consider the time elapsed between the failure and the inspection time whereas the past lifetime is related to the random lifetime with a left censoring.

The past lifetime is a random variable with support $(0, t)$ and whose cdf and pdf are expressed as

$$\begin{aligned} F_{{}_tX}(x) &= \frac{F_X(x)}{F_X(t)} \qquad x \in (0,t), \\ f_{{}_tX}(x) &= \frac{f_X(x)}{F_X(t)} \qquad x \in (0,t). \end{aligned}$$

Furthermore, the mean of ${}_tX$, known as mean past lifetime, is given by

$$\tilde{\mu}_X(t) = \int_0^t \left(1 - \frac{F_X(x)}{F_X(t)}\right) dx = t - \frac{1}{F_X(t)} \int_0^t F_X(x)dx, \tag{1.13}$$

see Di Crescenzo and Longobardi [37] for further details.

The inactivity time has the same support of the past lifetime and the pdf and cdf are expressed as

$$\begin{aligned} F_{X_{[t]}}(x) &= \frac{F_X(t) - F_X(t-x)}{F_X(t)} \qquad x \in (0,t), \\ f_{X_{[t]}}(x) &= \frac{f_X(t-x)}{F_X(t)} \qquad x \in (0,t). \end{aligned}$$

In the applications, it is of great interest the mean inactivity time, see, for instance Finkelstein [45], that is expressed as

$$m_X(t) = \frac{1}{F_X(t)} \int_0^t F_X(x)dx = t - \tilde{\mu}_X(t). \tag{1.14}$$

Assuming that $m_X(t)$ is differentiable, it readily follows

$$q(t) = \frac{1 - m'_X(t)}{m_X(t)}, \tag{1.15}$$

and then, by using (1.5) it is possible to obtain the cdf in function of the mean inactivity time

$$F(t) = \exp\left(- \int_t^{+\infty} \frac{1 - m'_X(x)}{m_X(x)} \, dx \right), \ t \in (0, +\infty).$$

1.5 Coherent systems

In this section, some general definitions and results on coherent systems are presented. For a more detailed discussion one may refer to Navarro [82]. The systems are one of the most important concepts in reliability theory. A system is a structure made up of components which determine its functioning. Here, we suppose to have only two possible states for the systems and the components, the functioning state represented by 1 and the failure state represented by 0, and then we talk about binary systems. The behavior of a system is governed by a function, known as structure function. For a system with n components, the structure function is defined as

$$\phi : \{0, 1\}^n \to \{0, 1\},$$

where $\phi(x_1, \ldots, x_n)$ represents the state of the system and x_i the state of the i-th component, $i = 1, \ldots, n$. It is reasonable to require that the function ϕ satisfies some properties. In fact it is reasonable to assume that if all the components work then the system works and if all the components do not work then the system does not work. Moreover, it is rational to require that if a broken component is replaced by a functioning one, the behavior of the system cannot get worse. The last sentence refers to the increasing property of the function ϕ, where we mean that if $x_1 \le y_1, \ldots, x_n \le y_n$, then $\phi(x_1, \ldots, x_n) \le \phi(y_1, \ldots, y_n)$. These assumptions conduct to the definition of semi-coherent system.

Definition 1.10. A system S with structure function ϕ is semi-coherent if

1. $\phi(0, \ldots, 0) = 0$;

2. $\phi(1, \ldots, 1) = 1$;

3. ϕ is increasing.

In order to give the definition of coherent system, an additional assumption is needed. In fact, semi-coherent system may have irrelevant components. A component is said to be irrelevant if the state of the system is independent of the state of that component, in the sense that

$$\phi(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = \phi(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n), \tag{1.16}$$

for all $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \{0, 1\}$.

Definition 1.11. A system S with structure function ϕ is coherent if

1. ϕ is increasing;

2. ϕ is strictly increasing in each variable in at last one point.

Obviously, all the coherent systems are semi-coherent but the converse is not true. For example, a system with 2 components and structure function $\phi(x_1, x_2) = x_1$ is semi-coherent but not coherent since it has an irrelevant component.

It has to be noted that some systems are equivalent under permutations in the sense that there exists a permutation $\sigma \in \mathcal{P}$, where $\mathcal{P}$ is the set of permutation over n elements, such that two structure functions ϕ_1 and ϕ_2 are related by $\phi_1(x_1, \ldots, x_n) = \phi_2(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. In this case, we write $\phi_1 \sim \phi_2$.

A coherent system can be determined by the sets of components that assure that the system works when these components work or by the set of components that assure the system fails when these components fail.

Definition 1.12. A non-empty set $P \subseteq [n] = \{1, \ldots, n\}$ is a path set of a system ϕ if $\phi(x_1, \ldots, x_n) = 1$ when $x_i = 1$ for all $i \in P$. A path set is a minimal path set if it does not contain other path sets. A non-empty set $C \subseteq [n]$ is a cut set of ϕ if $\phi(x_1, \ldots, x_n) = 0$ when $x_i = 0$ for all $i \in C$. A cut set is a minimal cut set if it does not contain other cut sets.

A system is completely determined by its minimal path (or cut) sets, as presented in the following proposition.

Proposition 1.11. *The non-empty sets $P_1, \ldots, P_r \subseteq [n]$ are the minimal path (or cut) sets of a coherent system if and only if the two following properties hold:*

1. *P_i is not contained in P_j for all $i \neq j$;*

2. *$P_1 \cup \cdots \cup P_r = [n]$.*

The above proposition was applied by Navarro and Rubio [85] to determine all the coherent systems of order n with a recursive algorithm.

Dealing with a system ϕ, it is always possible to consider its dual ϕ^D. The structure function of a system and its dual are related, for all $x_1, \ldots, x_n \in \{0, 1\}$, by

$$\phi^D(x_1, \ldots, x_n) = 1 - \phi(1 - x_1, \ldots, 1 - x_n). \tag{1.17}$$

A system and its dual are strictly related and share many properties. Naturally, if ϕ is a coherent system then ϕ^D is a coherent system and $(\phi^D)^D = \phi$. Moreover, a set is a path set of ϕ if and only if it is a cut set of ϕ^D, and a set is a cut set of ϕ if and only if it is a path set of ϕ^D. The same equivalences hold when considering minimal cut sets and minimal path sets.

Two of the most well-known systems are the series and the parallel ones, which are one the dual of the other. The series system of order n has a structure function defined as

$$\phi_{1:n}(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n).$$

The parallel system of order n has a structure function defined as

$$\phi_{n:n}(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n).$$

Series and parallel systems may be considered as special cases of k-out-of-n systems. The structure function of a k-out-of-n system is expressed as

$$\phi_{n-k+1:n}(x_1, \ldots, x_n) = \begin{cases} 1, & \text{if } x_1 + \cdots + x_n \geq k, \\ 0, & \text{if } x_1 + \cdots + x_n < k, \end{cases} \tag{1.18}$$

for $k = 1, \ldots, n$. A k-out-of-n system works if at least k of its n components work. So series systems are n-out-of-n systems and parallel systems are 1-out-of-n systems. The minimal path sets of k-out-of-n systems are all the sets with exactly k elements and then they have $\binom{n}{k}$ minimal path sets.

One of the most known ways to describe the structure function of a system is in terms of minimal path sets or minimal cut sets as presented in the following theorem.

Theorem 1.8. *Let ϕ be a coherent system of order n and let $P_1, \ldots, P_r$ and $C_1, \ldots, C_s$ be its minimal path sets and minimal cut sets, respectively. Then,*

$$\begin{aligned} \phi(x_1, \ldots, x_n) &= \max_{i=1,\ldots,r} \min_{j \in P_i} x_j, \\ \phi(x_1, \ldots, x_n) &= \min_{i=1,\ldots,s} \max_{j \in C_i} x_j, \end{aligned}$$

for all $(x_1, \ldots, x_n) \in \{0, 1\}^n$.

The expressions in Theorem 1.8 hold since a coherent system works if and only if at least one of the series systems obtained from its minimal path sets works and fails if and only if at least one of the parallel systems obtained from its minimal cut sets fails.

The signature of a system is an index which, although less general than a structure function, has the virtues of being both quite manageable and easily interpreted. In the following definition the signature of a system is introduced, see Samaniego [100] for further details on this concept. The definition in based on the concept of order statistics. We recall that, if we have n independent and identically distributed (IID) random variables $X_1, \ldots, X_n$, we can introduce the order statistics $X_{k:n}$, $k = 1, \ldots, n$. The k-th order statistic is equal to the k-th smallest value from the sample. The cdf of $X_{k:n}$ can be given in terms of the cdf of the parent distribution as

$$F_{k:n}(x) = \sum_{j=k}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j},$$

and the pdf of $X_{k:n}$ is expressed as

$$f_{k:n}(x) = \binom{n}{k} k [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x).$$

By choosing $k = 1$ and $k = n$, we get the smallest and largest order statistics, respectively. Their cdf and pdf are given by

$$\begin{aligned} F_{1:n}(x) &= 1 - [1 - F(x)]^n, \qquad & f_{1:n}(x) &= n[1 - F(x)]^{n-1} f(x), \\ F_{n:n}(x) &= [F(x)]^n, & f_{n:n}(x) &= n[F(x)]^{n-1} f(x). \end{aligned}$$

Definition 1.13. Let ϕ be a coherent system of order n. Assume that the lifetimes of the components of the system are independent and identically distributed according to the continuous distribution F. The signature of the system, denoted by s is an n-dimensional probability vector whose i-th element s_i is equal to the probability that the i-th component failure causes the system to fail. In brief, $s_i = \mathbb{P}(T = X_{i:n})$, where T is the failure time of the system and $X_{i:n}$ is the i-th order statistic of the n component failure times, i.e., the time of the i-th component failure.

From the definition, it readily follows that the signature of the series system is $(1, 0, \ldots, 0)$ and the signature of the parallel system is $(0, \ldots, 0, 1)$. More in general, the signature of k-out-of-n system is a vector with all zeros and a one in the $n - k + 1$-th position. The signature is a useful tool to describe the lifetime of a system as presented in the following proposition.

Proposition 1.12. *Let T be the lifetime of a coherent system with independent and identically distributed component lifetimes $X_1, \ldots, X_n$ having a common continuous distribution function F. Then,*

$$\overline{F}_T(t) = \sum_{i=1}^{n} s_i \overline{F}_{i:n}(t), \tag{1.19}$$

for all t, where $s = (s_1, \ldots, s_n)$ is the signature of the system.

Chapter 2

New measures of uncertainty in the classical probability theory

The measure of the uncertainty associated to a random variable, which may represent the lifetime of an item or a human being, is a task of great and increasing interest. Since the pioneering work of Shannon [108], in which the concept of Shannon entropy was defined as the average level of information or uncertainty related to a random event, several measures of uncertainty with different purposes have been defined and studied. Among them, one of the most interesting is the extropy, considered as the dual measure of Shannon entropy, defined by Lad et al. [67]. In this chapter, we will first give an overview of some well-known measures of uncertainty and provide a connection between cumulative entropies and the moments of order statistics (see Balakrishnan, Buono and Longobardi [6]). Then, based on Balakrishnan, Buono and Longobardi [7, 8], Buono, Kamari and Longobardi [24], Buono, Longobardi and Pellerey [26] and Kamari and Buono [58], new measures of uncertainty will be presented and their properties will be explained.

2.1 Some formulations of entropy

Let X be a discrete random variable with support $S = \{x_1, \dots, x_N\}$ and with corresponding probability vector $\mathbf{p} = (p_1, \dots, p_N)$, i.e., $\mathbb{P}(X = x_i) = p_i$, for $i = 1, \dots, N$. In 1948, Shannon [108] introduced a measure of information related to the information content and the uncertainty about an event associated with a discrete random variable. This measure, known as Shannon entropy, is defined as

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i,$$

where log is the natural logarithm. For the sake of simplicity, sometimes for discrete random variables in place of $H(X)$ we will use $H(\mathbf{p})$, and similarly with other measures of discrimination. The concept of entropy has since been generalized in different ways. Analogous to the discrete case, the Shannon entropy has been defined in the continuous case as

$$H(X) = \mathbb{E}[IC(X)] = \mathbb{E}[-\log f(X)] = -\int_0^{+\infty} f(x) \log f(x) dx, \tag{2.1}$$

where X is a non-negative random variable with probability density function f and $IC(X) = -\log f(X)$ denotes the information content of X, which can be also understood as the self-information or "surprise" associated with the possible outcomes of X. It is also known as differential entropy. Although the definitions are similar, the entropy is always non-negative in the discrete case while it could be negative in the continuous case. In the literature, different versions of entropy have been introduced. As a measure of information, the Shannon entropy is position-free, i.e., a random variable X has the same Shannon entropy of $X + b$, for any $b \in \mathbb{R}$. To avoid this problem, the concept of weighted entropy has been introduced (see Di Crescenzo and Longobardi [38]) as

$$H^w(X) = -\mathbb{E}[X \log f(X)] = -\int_0^{+\infty} x f(x) \log f(x) dx.$$

To measure the uncertainty about the residual lifetime of X at time t, Ebrahimi [43] introduced the residual entropy as

$$H(X_t) = -\int_t^{+\infty} \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)} dx, \tag{2.2}$$

where $\overline{F}$ is the survival function of X, and it is the differential entropy of the residual lifetime X_t. It is also possible to study the uncertainty about the past lifetime by introducing the past entropy as

$$H({}_tX) = -\int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx, \tag{2.3}$$

where F is the cumulative distribution function of X, and it is the differential entropy of the past lifetime ${}_tX$, see Di Crescenzo and Longobardi [37].

Among the different generalizations of Shannon entropy, the Tsallis entropy [118] has attracted considerable attention. For a discrete random variable, the Tsallis entropy $S_\alpha(X)$ is defined as

$$S_\alpha(X) = \frac{1}{\alpha - 1}\left(1 - \sum_{i=1}^N p_i^\alpha\right), \tag{2.4}$$

where $\alpha > 0$ and $\alpha \neq 1$. It is a generalization of Shannon entropy since it simply follows

$$\lim_{\alpha \to 1} S_\alpha(X) = H(X).$$

The differential version of the Tsallis entropy is defined as well in the following way

$$S_\alpha(X) = \frac{1}{\alpha - 1}\left[1 - \int_0^{+\infty} f^\alpha(x) dx\right]; \qquad \alpha \neq 1, \quad \alpha > 0.$$

Another interesting generalization of Shannon entropy was given by Ubriaco [119] which defined a new entropy based on fractional calculus as follows:

$$H_q(X) = \sum_{i=1}^{n} p_i[-\log p_i]^q, \quad 0 < q \leq 1. \tag{2.5}$$

It is known as fractional entropy and it is concave, positive and non-additive. Moreover, for $q = 1$, the fractional entropy reduces to the Shannon entropy. From a physical sense, it also satisfies Lesche and thermodynamic stability.

Lad et al. [67] introduced the extropy, a measure of uncertainty, as a dual version of the entropy. For a discrete random variable X, the extropy $J(X)$ is defined as

$$J(X) = -\sum_{i=1}^{N}(1-p_i)\log(1-p_i), \tag{2.6}$$

and it is always non-negative. An important connection between the entropy and the extropy emerges by analyzing their sum. In fact, the sum of entropy and extropy can be expressed as

$$H(X) + J(X) = \sum_{i=1}^{n} H(p_i, 1-p_i) = \sum_{i=1}^{n} J(p_i, 1-p_i), \tag{2.7}$$

where $H(p_i, 1-p_i) = J(p_i, 1-p_i) = -p_i \log p_i - (1-p_i)\log(1-p_i)$ are the entropy and the extropy of a discrete random variable which support has cardinality two and whose probability mass function vector is $(p_i, 1-p_i)$. We remark that the entropy and the extropy coincide for the variables whose support has cardinality two. Moreover, for a non-negative absolutely continuous random variable X, the extropy is defined as

$$J(X) = -\frac{1}{2}\mathbb{E}\left[f(X)\right] = -\frac{1}{2}\int_0^{+\infty} f^2(x)dx.$$

2.1.1 On cumulative entropies in terms of the moments of order statistics

Among the generalizations of the entropy, a great interest is devoted to the cumulative versions since they not require any assumptions on the probability density function of the random variable. In this section, we recall the definitions of two cumulative entropies: the Cumulative Residual Entropy (CRE) defined in Rao et al. [95] as

$$\mathcal{E}(X) = -\int_0^{+\infty} \overline{F}(x)\log\overline{F}(x)dx,$$

and the Cumulative Entropy (CE) introduced in Di Crescenzo and Longobardi [39] as

$$\mathcal{CE}(X) = -\int_0^{+\infty} F(x)\log F(x)dx. \tag{2.8}$$

Now, based on the results of Balakrishnan, Buono and Longobardi [6], we will provide a connection of the above measures with the moments of order statistics and we will present some bounds.

Let X be a random variable with finite expectation μ. The Cumulative Residual Entropy (CRE) of X can also be written in terms of order statistics as

$$\begin{aligned}
\mathcal{E}(X) &= -\int_0^{+\infty}(1-F(x))\log(1-F(x))dx \\
&= -x(1-F(x))\log(1-F(x))\big|_0^{+\infty} - \int_0^{+\infty} x\log(1-F(x))f(x)dx - \int_0^{+\infty} xf(x)dx \\
&= \int_0^{+\infty} x[-\log(1-F(x))]f(x)dx - \mu = \int_0^{+\infty} x\left[\sum_{n=1}^{+\infty}\frac{F(x)^n}{n}\right]f(x)dx - \mu \\
&= \sum_{n=1}^{+\infty}\frac{1}{n(n+1)}\mu_{n+1:n+1} - \mu, \qquad (2.9)
\end{aligned}$$

where $\mu_{n+1:n+1} = \mathbb{E}(X_{n+1:n+1})$, provided that $\lim_{x\to+\infty} -x(1-F(x))\log(1-F(x))$ exists and CRE is finite. In this case, the previous limit is equal to 0. Note that (2.9) can be rewritten as

$$\mathcal{E}(X) = \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{n+1}\right)\mu_{n+1:n+1} - \mu. \qquad (2.10)$$

Remark 2.1. We want to emphasize that, under the assumptions made, the steps in (2.9) are correct. The improper integral can be written as

$$\lim_{t\to+\infty}\int_0^t x \lim_{N\to+\infty}\sum_{n=1}^{N}\frac{F(x)^n}{n}f(x)dx. \qquad (2.11)$$

Hence, the sequence $S_N(x) = \sum_{n=1}^{N}\frac{F(x)^n}{n}$ is increasing and converges pointwise to the continuous function $-\log(1-F(x))$ for each $x \in [0,t]$ and, by applying Dini's theorem for uniform convergence [14], the convergence is uniform. Thus, (2.11) can be written as

$$\lim_{t\to+\infty}\lim_{N\to+\infty}\sum_{n=1}^{N}\int_0^t x\frac{F(x)^n}{n}f(x)dx. \qquad (2.12)$$

In order to apply Moore-Osgood theorem for the iterated limit [116], we have to show that

$$\lim_{t\to+\infty}\sum_{n=1}^{N}\int_0^t x\frac{F(x)^n}{n}f(x)dx = \sum_{n=1}^{N}\frac{1}{n(n+1)}\mu_{n+1:n+1}$$

converges pointwise for each fixed N, and this is satisfied if X has finite mean. Hence, by applying Moore-Osgood theorem for the iterated limit, (2.12) can be written as

$$\lim_{N\to+\infty}\lim_{t\to+\infty}\sum_{n=1}^{N}\int_0^t x\frac{F(x)^n}{n}f(x)dx = \sum_{n=1}^{+\infty}\frac{1}{n(n+1)}\mu_{n+1:n+1}.$$

In the following example, we use (2.10) to evaluate the CRE for the standard exponential distribution.

Example 2.1. Consider the standard exponential distribution with pdf $f(x) = e^{-x}$, $x > 0$. It is known that

$$\mu = 1 \quad \text{and} \quad \mu_{n:n} = 1 + \frac{1}{2} + \cdots + \frac{1}{n},$$

see [3] for further details. Hence, from (2.10), it follows

$$\begin{aligned}
\mathcal{E}(X) &= \left(1 - \frac{1}{2}\right)\mu_{2:2} + \left(\frac{1}{2} - \frac{1}{3}\right)\mu_{3:3} + \left(\frac{1}{3} - \frac{1}{4}\right)\mu_{4:4} + \cdots - \mu \\
&= (\mu_{2:2} - \mu) + \frac{1}{2}(\mu_{3:3} - \mu_{2:2}) + \frac{1}{3}(\mu_{4:4} - \mu_{3:3}) + \cdots = \frac{1}{2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \ldots \\
&= \sum_{n=1}^{+\infty} \frac{1}{n(n+1)} = 1.
\end{aligned}$$

In analogy with (2.9), it is also possible to provide a connection of the Cumulative Entropy (CE) with the moments of order statistics. More precisely, the CE of X, with finite expectation μ, can also be rewritten in terms of the mean of the minimum order statistic. From (2.8), we easily obtain

$$\begin{aligned}
\mathcal{CE}(X) &= -xF(x)\log F(x)\Big|_0^{+\infty} + \int_0^{+\infty} x \log F(x) f(x)dx + \int_0^{+\infty} x f(x)dx \\
&= \int_0^{+\infty} x \log[1 - (1 - F(x))] f(x)dx + \mu = -\int_0^{+\infty} x \sum_{n=1}^{+\infty} \frac{(1 - F(x))^n}{n} f(x)dx + \mu \\
&= -\sum_{n=1}^{+\infty} \frac{1}{n(n+1)} \mu_{1:n+1} + \mu, \qquad (2.13)
\end{aligned}$$

provided that $\lim_{x \to +\infty} -xF(x)\log F(x)$ exists and CE is finite. Note that (2.13) can be also written as

$$\mathcal{CE}(X) = -\sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right) \mu_{1:n+1} + \mu. \qquad (2.14)$$

In the following example, we apply (2.14) to the standard exponential distribution.

Example 2.2. For the standard exponential distribution, it is known that

$$\mu_{1:n} = \frac{1}{n},$$

and so from (2.14), by the use of Euler's identity, we get

$$\begin{aligned}
\mathcal{CE}(X) &= -\sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right) \frac{1}{n+1} + 1 = -\sum_{n=1}^{+\infty} \frac{1}{n(n+1)} + \sum_{n=1}^{+\infty} \frac{1}{(n+1)^2} + 1 \\
&= \frac{\pi^2}{6} - 1.
\end{aligned}$$

Now, based on the expressions given in (2.9) and (2.13), we provide some bound for the CRE and the CE. With this purpose, let Z denote the standard version of the random variable X, i.e.,

$$Z = \frac{X - \mu}{\sigma},$$

where σ is the standard deviation of X. By construction, the relation between a variable and its standard version holds for order statistics and so we have

$$Z_{k:n} = \frac{X_{k:n} - \mu}{\sigma},$$

for $k = 1, \dots, n$. Hence, the mean of $X_{k:n}$ and the mean of $Z_{k:n}$ are directly related and, in particular, for the largest order statistic, we have

$$\mathbb{E}(Z_{n:n}) = \frac{\mu_{n:n} - \mu}{\sigma}.$$

Note that this formula also holds by considering a generalization of the random variable Z with an arbitrary location parameter μ in place of the mean, and an arbitrary scale parameter σ in place of the standard deviation.

Consider a sample with parent distribution Z such that $\mathbb{E}(Z) = 0$ and $\mathbb{E}(Z^2) = 1$. Hartley and David [54] and Gumbel [51] have then shown that

$$\mathbb{E}(Z_{n:n}) \le \frac{n-1}{\sqrt{2n-1}}. \tag{2.15}$$

The inequality in (2.15) is known as Hartley-David-Gumbel bound. Then, by using the Hartley-David-Gumbel bound for a parent distribution with mean μ and variance σ^2, we get

$$\mu_{n:n} = \sigma \mathbb{E}(Z_{n:n}) + \mu \le \sigma \frac{n-1}{\sqrt{2n-1}} + \mu. \tag{2.16}$$

Theorem 2.1. *Let X be a random variable with mean μ and variance σ^2. Then, an upper bound for the CRE of X is given as*

$$\mathcal{E}(X) \le \sum_{n=1}^{+\infty} \frac{\sigma}{(n+1)\sqrt{2n+1}} \simeq 1.21\ \sigma. \tag{2.17}$$

Proof. From (2.9) and (2.16), we get

$$\begin{aligned} \mathcal{E}(X) &= \sum_{n=1}^{+\infty} \frac{1}{n(n+1)} \mu_{n+1:n+1} - \mu \le \sum_{n=1}^{+\infty} \frac{1}{n(n+1)} \left(\sigma \frac{n}{\sqrt{2n+1}} + \mu \right) - \mu \\ &= \sum_{n=1}^{+\infty} \frac{\sigma}{(n+1)\sqrt{2n+1}} \simeq 1.21\ \sigma, \end{aligned}$$

which is the upper bound given in (2.17). $\square$

Remark 2.2. If X is a non-negative random variable, we have $\mu_{n+1:n+1} \ge 0$ and $\mu_{1:n+1} \ge 0$, for all $n \in \mathbb{N}$. For this reason, by using finite series approximations in (2.9) and (2.13), we get bounds for $\mathcal{E}(X)$ and $\mathcal{CE}(X)$ as

$$\begin{aligned} \mathcal{E}(X) &\ge \sum_{n=1}^{m} \frac{1}{n(n+1)} \mu_{n+1:n+1} - \mu, \\ \mathcal{CE}(X) &\le -\sum_{n=1}^{m} \frac{1}{n(n+1)} \mu_{1:n+1} + \mu, \end{aligned}$$

for all $m \in \mathbb{N}$.

In the following theorem, we provide a lower bound of CE for decreasing failure rate distributions (DFR).

Theorem 2.2. *Let X be DFR. Then, a lower bound for $\mathcal{CE}(X)$ is given as*

$$\mathcal{CE}(X) \geq \mu - \sqrt{\frac{\mu^{(2)}}{2}}\left(2 - \frac{\pi^2}{6}\right),$$

where $\mu^{(2)} = \mathbb{E}(X^2)$ is the second moment of X.

Proof. Let X be DFR. From Theorem 12 of Rychlik [99], if in a sample of size n

$$\delta_{j,n} = \sum_{k=1}^{j} \frac{1}{n+1-k} \leq 2 \quad j \in \{1, \dots, n\},$$

then

$$\mu_{j:n} \leq \frac{\delta_{j,n}}{\sqrt{2}}\sqrt{\mu^{(2)}}.$$

For $j = 1$, we have $\delta_{1,n} = \frac{1}{n} \leq 2$ for all $n \in \mathbb{N}$, so that

$$\mu_{1:n} \leq \frac{\sqrt{\mu^{(2)}}}{\sqrt{2}\, n}.$$

Then, from (2.13), we get the following lower bound for $\mathcal{CE}(X)$:

$$\mathcal{CE}(X) \quad \geq \quad -\sum_{n=1}^{+\infty} \frac{1}{n(n+1)^2}\sqrt{\frac{\mu^{(2)}}{2}} + \mu = \mu - \sqrt{\frac{\mu^{(2)}}{2}}\left(2 - \frac{\pi^2}{6}\right).$$

□

Remark 2.3. Note that we cannot provide an analogous bound for $\mathcal{E}(X)$ since $\delta_{n,n} \leq 2$ is not fulfilled for $n \geq 4$.

The connection with the moments of order statistics can be done also for the weighted versions of the CRE and CE. Moreover, by using these relations, it is possible to compute estimations of the cumulative entropies, see Balakrishnan, Buono and Longobardi [6] for further details.

2.2 Past extropy

In this section based on the results given in Kamari and Buono [58], we study the extropy for ${}_tX$, where X is a non-negative random variable. This measure is known as the past extropy and is defined as

$$J\left({}_tX\right) = -\frac{1}{2}\int_0^{+\infty} f^2_{{}_tX}(x)dx = -\frac{1}{2F^2(t)}\int_0^t f^2(x)dx. \tag{2.18}$$

It is clear that $J\left({}_{+\infty}X\right) = J(X)$ and that the past extropy is non-positive, i.e., $J({}_tX) \leq 0$. Moreover, it can be expressed also by using the reversed hazard rate function as

$$J\left({}_tX\right) = \frac{-q^2(t)}{2f^2(t)} \int_0^t f^2(x)dx.$$

Example 2.3. a) If $X \sim Exp(\lambda)$, then $J\left({}_tX\right) = -\frac{\lambda}{4}\frac{1+e^{-\lambda t}}{1-e^{-\lambda t}}$ for $t > 0$. This shows that the past extropy of exponential distribution is an increasing function of t.

b) If $X \sim U(0,b)$, then $J\left({}_tX\right) = -\frac{1}{2t}$.

c) If X has power distribution with parameter $\alpha > 0$, i.e., $f(x) = \alpha x^{(\alpha-1)}$, $0 < x < 1$, then $J\left({}_tX\right) = \frac{-\alpha^2}{2(2\alpha-1)t}$.

d) If X has Pareto distribution with parameters $\theta > 0, x_0 > 0$, i.e., $f(x) = \frac{\theta}{x_0}\frac{x_0^{\theta+1}}{x^{\theta+1}}$, $x > x_0$, then $J\left({}_tX\right) = \frac{\theta^2}{2(2\theta+1)(t^\theta - x_0^\theta)^2}\left[\frac{x_0^{2\theta}}{t} - \frac{t^{2\theta}}{x_0}\right]$.

The past extropy can be considered as a measure dual to the residual extropy, defined in Qiu and Jia [94] to study the uncertainty about the residual lifetime X_t as

$$J\left(X_t\right) = -\frac{1}{2}\int_0^{+\infty} f_{X_t}^2(x)dx = -\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty} f^2(x)dx.$$

There is a functional relation between past extropy and residual extropy expressed as follows:

$$J(X) = F^2(t)J\left({}_tX\right) + \overline{F}^2(t)J\left(X_t\right), \quad \text{for all} \quad t > 0.$$

Definition 2.1. A random variable is said to be increasing (decreasing) in past extropy if $J\left({}_tX\right)$ is an increasing (decreasing) function of t.

Proposition 2.1. *$J\left({}_tX\right)$ is increasing (decreasing) if and only if $J\left({}_tX\right) \leq (\geq)\frac{-1}{4}q(t)$.*

Proof. By differentiating with respect to t in (2.18), we get

$$\frac{d}{dt}J\left({}_tX\right) = -2q(t)J\left({}_tX\right) - \frac{1}{2}q^2(t).$$

Then, the past extropy $J\left({}_tX\right)$ is increasing if, and only if,

$$2q(t)J\left({}_tX\right) + \frac{1}{2}q^2(t) \leq 0,$$

but the reversed hazard rate $q(t)$ is non-negative and so this is equivalent to

$$J\left({}_tX\right) \leq -\frac{1}{4}q(t).$$

□

Theorem 2.3. *The past extropy $J\left({}_tX\right)$ of X is uniquely determined by the reversed hazard rate $q(t)$.*

Proof. From (2.18) we get

$$\frac{d}{dt}J\left({}_tX\right) = -2q(t)J\left({}_tX\right) - \frac{1}{2}q^2(t).$$

This is a linear differential equation of order one and it is solved by

$$J\left({}_tX\right) = e^{-2\int_{t_0}^{t} q(s)ds}\left[J\left({}_{t_0}X\right) - \int_{t_0}^{t}\frac{1}{2}q^2(s)e^{2\int_{t_0}^{s} q(y)dy}ds\right],$$

where by the use of the boundary condition $J\left({}_{+\infty}X\right) = J(X)$, we obtain

$$J\left({}_tX\right) = e^{2\int_{t}^{+\infty} q(s)ds}\left[J\left(X\right) + \int_{t}^{+\infty}\frac{1}{2}q^2(s)e^{-2\int_{s}^{+\infty} q(y)dy}ds\right].$$

□

By using the usual stochastic order and the likelihood ratio order, we show a result related to the monotinicity of $J\left({}_tX\right)$. In order to do this, we recall that $X \leq_{st} Y$ if and only if $\mathbb{E}(\varphi(Y)) \leq (\geq)\mathbb{E}(\varphi(X))$ for any decreasing (increasing) function φ.

Theorem 2.4. *Let X be a random variable with cdf F and pdf f. If $f\left(F^{-1}(x)\right)$ is decreasing in $x \geq 0$, then $J\left({}_tX\right)$ is increasing in $t \geq 0$.*

Proof. Let U_t be a random variable with uniform distribution on $(0, F(t))$ with pdf $g_t(x) = \frac{1}{F(t)}$ for $x \in (0, F(t))$. Then, by (2.18) we have

$$\begin{aligned} J\left({}_tX\right) &= -\frac{1}{2F^2(t)}\int_0^{F(t)} f\left(F^{-1}(u)\right)du = -\frac{1}{2F(t)}\int_0^{F(t)} g_t(u)f\left(F^{-1}(u)\right)du \\ &= -\frac{1}{2F(t)}\mathbb{E}\left[f\left(F^{-1}(U_t)\right)\right]. \end{aligned}$$

Let $0 \leq t_1 \leq t_2$. If $0 < x \leq F(t_1)$, then $\frac{g_{t_1}(x)}{g_{t_2}(x)} = \frac{F(t_2)}{F(t_1)}$ is a non-negative constant. If $F(t_1) < x \leq F(t_2)$, then $\frac{g_{t_1}(x)}{g_{t_2}(x)} = 0$. Therefore $\frac{g_{t_1}(x)}{g_{t_2}(x)}$ is decreasing in $x \in (0, F(t_2))$, which implies $U_{t_1} \leq_{lr} U_{t_2}$. Hence $U_{t_1} \leq_{st} U_{t_2}$ and so

$$0 \leq \mathbb{E}\left[f\left(F^{-1}(U_{t_2})\right)\right] \leq \mathbb{E}\left[f\left(F^{-1}(U_{t_1})\right)\right],$$

by using the assumption that $f\left(F^{-1}(U_t)\right)$ is a decreasing function. Since $0 \leq \frac{1}{F(t_2)} \leq \frac{1}{F(t_1)}$ then

$$J\left({}_{t_1}X\right) = -\frac{1}{2F(t_1)}\mathbb{E}\left[f\left(F^{-1}(U_{t_1})\right)\right] \leq -\frac{1}{2F(t_2)}\mathbb{E}\left[f\left(F^{-1}(U_{t_2})\right)\right] = J\left({}_{t_2}X\right).$$

□

Remark 2.4. Let X be a random variable with cdf $F(x) = x^2$, for $x \in (0,1)$. Then $f\left(F^{-1}(x)\right) = 2\sqrt{x}$ is increasing in $x \in (0,1)$. However $J\left({}_tX\right) = -\frac{2}{3t}$ is increasing in $t \in (0,1)$. So the condition in Theorem 2.4 that $f\left(F^{-1}(x)\right)$ is decreasing in x is sufficient but not necessary.

Let $X_1, X_2, \dots, X_n$ be continuous and IID random variables with cdf F representing the lifetimes of n components in a parallel system. Let $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ be the ordered lifetimes of the components. The past extropy of $X_{n:n}$ is given by

$$J\left({}_tX_{n:n}\right) = -\frac{n^2}{2(F(t))^{2n}} \int_0^t f^2(x)[F(x)]^{2n-2}dx.$$

Theorem 2.5. *If X has an increasing pdf f on $[0,T]$, with $T > t$, then $J\left({}_tX_{n:n}\right)$ is decreasing in $n \geq 1$.*

Proof. The pdf of $(X_{n:n}|X_{n:n} \leq t)$ can be expressed as

$$g^t_{n:n}(x) = \frac{nf(x)F^{n-1}(x)}{F^n(t)}, \ x \leq t.$$

Note that

$$\frac{g^t_{2n-1:2n-1}(x)}{g^t_{2n+1:2n+1}(x)} = \frac{2n-1}{2n+1} \cdot \frac{F^2(t)}{F^2(x)}$$

is decreasing in $x \in [0,t]$ and so $(X_{2n-1:2n-1}|X_{2n-1:2n-1} \leq t) \leq_{lr} (X_{2n+1:2n+1}|X_{2n+1:2n+1} \leq t)$ which implies $(X_{2n-1:2n-1}|X_{2n-1:2n-1} \leq t) \leq_{st} (X_{2n+1:2n+1}|X_{2n+1:2n+1} \leq t)$. If f is increasing on $[0,T]$, then

$$\mathbb{E}\left[f\left(X_{2n-1:2n-1}\right)|X_{2n-1:2n-1} \leq t\right] \leq \mathbb{E}\left[f\left(X_{2n+1:2n+1}\right)|X_{2n+1:2n+1} \leq t\right].$$

By using the definition of the past extropy, it readily follows

$$\begin{aligned}
J\left({}_tX_{n:n}\right) &= -\frac{n^2}{2F^{2n}(t)} \int_0^t f^2(x)F^{2n-2}(x)dx \\
&= \frac{-n^2}{2(2n-1)F(t)} \int_0^t \frac{(2n-1)F^{2n-2}(x)f(x)}{F^{2n-1}(t)} f(x)dx \\
&= \frac{-n^2}{2(2n-1)F(t)} \mathbb{E}\left[f\left(X_{2n-1:2n-1}\right)|X_{2n-1:2n-1} \leq t\right],
\end{aligned}$$

and then

$$\begin{aligned}
\frac{J\left({}_tX_{n:n}\right)}{J\left({}_tX_{n+1:n+1}\right)} &= \frac{n^2}{(n+1)^2} \frac{2n-1}{2n+1} \frac{\mathbb{E}\left[f\left(X_{2n-1:2n-1}\right)|X_{2n-1:2n-1} \leq t\right]}{\mathbb{E}\left[f\left(X_{2n+1:2n+1}\right)|X_{2n+1:2n+1} \leq t\right]} \\
&\leq \frac{\mathbb{E}\left[f\left(X_{2n-1:2n-1}\right)|X_{2n-1:2n-1} \leq t\right]}{\mathbb{E}\left[f\left(X_{2n+1:2n+1}\right)|X_{2n+1:2n+1} \leq t\right]} \leq 1.
\end{aligned}$$

Since the past extropy of a random variable is non-positive, we have $J\left({}_tX_{n:n}\right) \geq J\left({}_tX_{n+1:n+1}\right)$ and the proof is completed. □

Example 2.4. Let X be a random variable distribuited as a Weibull with two parameters, $X \sim W2(\alpha, \lambda)$, with pdf $f(x) = \lambda\alpha x^{\alpha-1} \exp(-\lambda x^{\alpha})$. For $\alpha > 1$ this pdf has a maximum point $T = \left(\frac{\alpha-1}{\lambda\alpha}\right)^{\frac{1}{\alpha}}$. Let us consider the case in which X has a Weibull distribution with parameters $\alpha = 2$ and $\lambda = 1$, $X \sim W2(2,1)$, so that $T = \frac{\sqrt{2}}{2}$. The hypothesis of Theorem 2.5 are satisfied for $t = 0.5 < T = \frac{\sqrt{2}}{2}$. Moreover, it can be also graphically observed that the result in Theorem 2.5 does not hold for the smallest order statistic by considering again $t = 0.5$.

In the case in which X has an increasing pdf on $[0, T]$ with $T > t$ we give a lower bound for $J({}_tX)$.

Proposition 2.2. *If X has an increasing pdf f on $[0, T]$, with $T > t$, then $J({}_tX) \geq -\frac{q(t)}{2}$.*

Proof. From the definition of the past extropy we get

$$J({}_tX) = -\frac{1}{2F^2(t)} \int_0^t f^2(x)dx = \frac{-f(t)}{2F(t)} + \frac{1}{2F^2(t)} \int_0^t F(x)f'(x)dx \geq -\frac{q(t)}{2}.$$

□

Example 2.5. Let $X \sim W2(2,1)$, as in Example 2.4, so we know that its pdf is increasing in $[0, T]$ with $T = \frac{\sqrt{2}}{2}$. The hypothesis of Proposition 2.2 are satisfied for $t < T = \frac{\sqrt{2}}{2}$. Figure 2.1 shows that the function $-\frac{q(t)}{2}$ is a lower bound for the past extropy. We remark that the theorem gives information only for $t \in [0, T]$, in fact for larger values of t the function $-\frac{q(t)}{2}$ could not be a lower bound anymore, as shown in Figure 2.1.

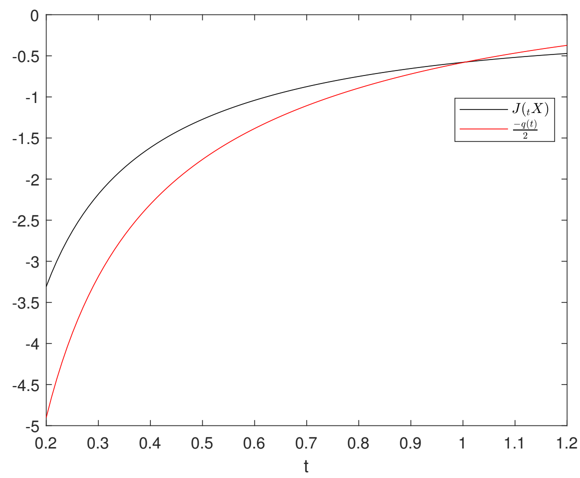

Figure 2.1: $J({}_tX)$ (in black) and $-\frac{q(t)}{2}$ (in red) of a $W2(2,1)$.

In the following theorem, we show that the past extropy of the largest order statistic can uniquely characterize the underlying distribution. The proof is based on the following lemma.

Lemma 2.1. *Let X and Y be non-negative random variables such that $J\left(X_{n:n}\right)=J\left(Y_{n:n}\right)$, for all $n\geq 1$. Then $X\overset{d}{=}Y$.*

Proof. By the definition of the extropy, $J\left(X_{n:n}\right)=J\left(Y_{n:n}\right)$ holds if, and only if,

$$\int_0^{+\infty} F_X^{2n-2}(x)f_X^2(x)dx=\int_0^{+\infty} F_Y^{2n-2}(x)f_Y^2(x)dx,$$

that is equivalent to

$$\int_0^{+\infty} F_X^{2n-2}(x)q_X(x)dF_X^2(x)=\int_0^{+\infty} F_Y^{2n-2}(x)q_Y(x)dF_Y^2(x).$$

By using $u=F_X^2(x)$ in the left hand side of the above equation and $u=F_Y^2(x)$ in the right hand side, we have

$$\int_0^1 u^{n-1}q_X\left(F_X^{-1}(\sqrt{u})\right)du=\int_0^1 u^{n-1}q_Y\left(F_Y^{-1}(\sqrt{u})\right)du,$$

or, equivalently,

$$\int_0^1 u^{n-1}\left[q_X\left(F_X^{-1}(\sqrt{u})\right)-q_Y\left(F_Y^{-1}(\sqrt{u})\right)\right]du=0\quad\text{for all}\quad n\geq 1.$$

Then, by Lemma 3.1 of Qiu [93] we get $q_X\left(F_X^{-1}(\sqrt{u})\right)=q_Y\left(F_Y^{-1}(\sqrt{u})\right)$ for all $u\in(0,1)$. By taking $\sqrt{u}=v$, we have $q_X\left(F_X^{-1}(v)\right)=q_Y\left(F_Y^{-1}(v)\right)$ and so $f_X\left(F_X^{-1}(v)\right)=f_Y\left(F_Y^{-1}(v)\right)$ for all $v\in(0,1)$. This is equivalent to $(F_X^{-1})'(v)=(F_Y^{-1})'(v)$, i.e., $F_X^{-1}(v)=F_Y^{-1}(v)+C$, for all $v\in(0,1)$ where C is a constant. But for $v=0$ we have $F_X^{-1}(0)=F_Y^{-1}(0)=0$ and so $C=0$. □

Theorem 2.6. *Let X and Y be two non-negative random variables with cumulative distribution functions $F(x)$ and $G(x)$, respectively. Then F and G belong to the same family of distributions if and only if for $t\geq 0$ and $n\geq 1$,*

$$J\left({}_tX_{n:n}\right)=J\left({}_tY_{n:n}\right).$$

Proof. We only need to prove the sufficiency. $J\left({}_tX_{n:n}\right)$ is the past extropy for $X_{n:n}$ but it is also the extropy for the variable ${}_tX_{n:n}$. So by Lemma 2.1 we get ${}_tX\overset{d}{=}{}_tY$. Then $\frac{F(t-x)}{F(t)}=\frac{G(t-x)}{G(t)}$, for $x\in(0,t)$. If there exists t' such that $F(t')\neq G(t')$ then in $(0,t')$ $F(x)=\alpha G(x)$ with $\alpha\neq 1$. But for all $t>t'$, there exists $x\in(0,t)$ such that $t-x=t'$ and so $F(t)\neq G(t)$ and, as in the previous step, we have $F(x)=\alpha G(x)$ for $x\in(0,t)$. By letting t to $+\infty$ a contradiction occurs since F and G are both cumulative distribution functions and their common limit is 1. □

2.3 Weighted extropy

The extropy introduced by Lad et al. [67] is a shift-independent information measure just as the entropy. In analogy with the weighted entropy and to efficiently model statistical data, a new measure of information, named weighted extropy, was proposed in Balakrishnan, Buono and Longobardi [8]. It is defined as

$$J^w(X) = -\frac{1}{2}\mathbb{E}[Xf(X)] = -\frac{1}{2}\int_0^{+\infty} xf^2(x)dx. \tag{2.19}$$

We now present two examples of distributions with the same extropy, but different weighted extropy. By the first example, it is possible to observe that the weighted extropy is indeed shift-dependent.

Example 2.6. Let X and Y be two random variables such that $X \sim U(0,b)$, $Y \sim U(a, a+b)$, where $a, b > 0$. We have $f_X(x) = \frac{1}{b}$, for $x \in (0,b)$, and $f_Y(y) = \frac{1}{b}$, for $y \in (a, a+b)$, and then

$$J(X) = -\frac{1}{2}\int_0^b \frac{1}{b^2}dx = -\frac{1}{2b}, \qquad J(Y) = -\frac{1}{2}\int_a^{a+b} \frac{1}{b^2}dy = -\frac{1}{2b},$$

i.e., X and Y have the same extropy. But, they have different weighted extropy:

$$\begin{aligned} J^w(X) &= -\frac{1}{2}\int_0^b x\frac{1}{b^2}dx = -\frac{1}{2b^2}\frac{b^2}{2} = -\frac{1}{4}, \\ J^w(Y) &= -\frac{1}{2}\int_a^{a+b} y\frac{1}{b^2}dy = -\frac{1}{2b^2}\frac{(a+b)^2-a^2}{2} = -\frac{b^2+2ab}{4b^2} = -\frac{b+2a}{4b}, \end{aligned}$$

and so if $a \neq 0$, i.e., X and Y are not identically distributed, then $J^w(X) \neq J^w(Y)$.

Example 2.7. Let X be a random variable with piecewise constant pdf

$$f(x) = \sum_{k=1}^n c_k \mathbf{1}_{[k-1,k)}(x),$$

where $c_k \geq 0$, $k = 1, \ldots, n$, $\sum_{k=1}^n c_k = 1$, and $\mathbf{1}_{[k-1,k)}(x)$ is the indicator function of x in the interval $[k-1, k)$. Then, the extropy and the weighted extropy of X are

$$\begin{aligned} J(X) &= -\frac{1}{2}\int_0^n \sum_{k=1}^n c_k^2 \mathbf{1}_{[k-1,k)}(x)dx = -\frac{1}{2}\sum_{k=1}^n \int_{k-1}^k c_k^2 dx = -\frac{1}{2}\sum_{k=1}^n c_k^2, \\ J^w(X) &= -\frac{1}{2}\int_0^n x\sum_{k=1}^n c_k^2 \mathbf{1}_{[k-1,k)}(x)dx = -\frac{1}{2}\sum_{k=1}^n \int_{k-1}^k xc_k^2 dx \\ &= -\frac{1}{2}\sum_{k=1}^n c_k^2 \frac{k^2-(k-1)^2}{2} = -\frac{1}{4}\sum_{k=1}^n c_k^2(2k-1). \end{aligned}$$

Since we obtain different distributions through a permutation of $c_1, \ldots, c_n$, we observe that they have the same extropy, but different weighted extropy (except in some special cases).

We now present an example of distributions with the same weighted extropy, but different extropy.

Example 2.8. Let X be a random variable such that $X \sim U(0,b)$, with $b > 0$. By Example 2.6, we have $J(X) = -\frac{1}{2b}$ and $J^w(X) = -\frac{1}{4}$, and so the extropy depends on b while the weighted extropy does not. Thus, if we consider $Y \sim U(0,a)$, with $a > 0$ and $a \neq b$, we have random variables X and Y with the same weighted extropy, but different extropy.

In the following example, the weighted extropy is evaluated for some well-known distributions.

Example 2.9. (a) Let X be exponentially distributed with parameter λ. Then,

$$J^w(X) = -\frac{1}{2}\int_0^{+\infty} x\lambda^2 e^{-2\lambda x}dx = \frac{1}{2}\left[\frac{\lambda x}{2}e^{-2\lambda x}\Big|_0^{+\infty} - \int_0^{+\infty}\frac{\lambda}{2}e^{-2\lambda x}dx\right] = -\frac{1}{8}.$$

(b) Let X be uniformly distributed over (a,b). Then,

$$J^w(X) = -\frac{1}{2}\int_a^b x\frac{1}{(b-a)^2}dx = -\frac{1}{2(b-a)^2}\frac{b^2-a^2}{2} = -\frac{1}{4}\frac{b+a}{b-a}.$$

Note that in this case the weighted extropy can be expressed as the product

$$J^w(X) = J(X)\mathbb{E}(X),$$

where $\mathbb{E}(X) = \frac{a+b}{2}$ and $J(X) = -\frac{1}{2(b-a)}$. Then, $J(X) \le J^w(X)$ if, and only if, $\mathbb{E}(X) \le 1$, since the extropy and the weighted extropy are non-positive.

(c) Let X be gamma distributed with parameters α and β, and with pdf

$$f(x) = \begin{cases} \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Then, the weighted extropy of X is expressed as

$$\begin{aligned} J^w(X) &= -\frac{1}{2}\int_0^{+\infty} x\frac{x^{2\alpha-2}e^{-2x/\beta}}{\beta^{2\alpha}\Gamma^2(\alpha)}dx \\ &= -\frac{1}{2\beta^{2\alpha}\Gamma^2(\alpha)}\int_0^{+\infty} x^{2\alpha-1}e^{-2x/\beta}dx = -\frac{1}{2^{2\alpha+1}}\frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)}, \end{aligned}$$

and it is free of the scale parameter β.

(d) Let X be beta distributed with parameters α and β, and with pdf

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise,} \end{cases}$$

where

$$B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

is the complete beta function. Then, the weighted extropy is evaluated as

$$J^w(X) \;=\; -\frac{1}{2}\int_0^1 x\frac{x^{2\alpha-2}(1-x)^{2\beta-2}}{B^2(\alpha,\beta)}dx=-\frac{1}{2}\frac{B(2\alpha,2\beta-1)}{B^2(\alpha,\beta)},$$

if $2\beta-1>0$, i.e., $\beta>\frac{1}{2}$, but if $0<\beta\leq\frac{1}{2}$ the weighted extropy is not finite, $J^w(X)=-\infty$.

Remark 2.5. Let us now focus the attention on the integrability of $xf^2(x)$ on the support of X. If the support is unbounded, i.e., $(a,+\infty)$, with $a\geq 0$, and the function is bounded, we have to investigate the behaviour at infinity. As $\int_a^{+\infty}f(x)dx=1$, we have $\lim_{x\to+\infty}f(x)=0$ and $f(x)$ is infinitesimal of higher order with respect to $\frac{1}{x^{1+\varepsilon}}$, for $x\to+\infty$, for some $\varepsilon>0$. Then, $xf^2(x)$ is infinitesimal of higher order with respect to $\frac{1}{x^{1+2\varepsilon}}$, for $x\to+\infty$ and so it is integrable, i.e., the weighted extropy is finite. If the support and the density are unbounded, we also have to study the behaviour at a. Suppose $a>0$. If $\lim_{x\to a^+}f(x)=+\infty$, by normalization condition, we know that $f(x)$ is infinity of lower order with respect to $\frac{1}{(x-a)^{1-\varepsilon}}$, for $x\to a^+$, for some $0<\varepsilon<1$. Hence, $xf^2(x)$ is infinity of lower order with respect to $\frac{1}{(x-a)^{2-2\varepsilon}}$, for $x\to a^+$, and so is integrable if $\varepsilon\in\left(\frac{1}{2},1\right)$. If $a=0$, by normalization condition, we know that $f(x)$ is infinity of lower order with respect to $\frac{1}{x^{1-\varepsilon}}$, for $x\to 0^+$, for some $0<\varepsilon<1$. Hence, $xf^2(x)$ is infinity of lower order with respect to $\frac{1}{x^{1-2\varepsilon}}$, for $x\to 0^+$, and it is integrable. If the support is bounded and f is bounded, then $xf^2(x)$ is bounded and integrable. If the support is bounded and f is unbounded, then we can refer to the previous cases. Observe that if the support is $(0,+\infty)$, the weighted extropy is always finite.

In the following proposition, we study weighted extropy under monotone transformations.

Proposition 2.3. *Let $Y=\Phi(X)$, where Φ is strictly monotone and differentiable, with derivative Φ'. Then, we have*

$$J^w(Y)=\begin{cases}-\frac{1}{2}\int_0^{+\infty}\frac{\Phi(x)}{\Phi'(x)}f_X^2(x)dx, & \textit{if } \Phi \textit{ is strictly increasing}\\ -\frac{1}{2}\int_0^{+\infty}\frac{\Phi(x)}{|\Phi'(x)|}f_X^2(x)dx, & \textit{if } \Phi \textit{ is strictly decreasing.}\end{cases} \tag{2.20}$$

Proof. From (2.19), we have

$$J^w(Y)=-\frac{1}{2}\int_0^{+\infty}x\frac{f_X^2(\Phi^{-1}(x))}{(\Phi'(\Phi^{-1}(x)))^2}dx.$$

Let Φ be strictly increasing. Then, with a change of variable in the above integral, it follows

$$J^w(Y)=-\frac{1}{2}\int_0^{+\infty}\frac{\Phi(x)}{\Phi'(x)}f_X^2(x)dx,$$

giving the first expression in (2.20). If Φ is strictly decreasing, the second expression in (2.20) can be similarly obtained. $\square$

Remark 2.6. If in Theorem 2.3 we consider the increasing transformation $\Phi(X) = F_X(X)$, then

$$J^w(Y) = -\frac{1}{2}\int_0^{+\infty} F_X(x)f_X(x)dx = -\frac{1}{4},$$

which agrees with the result for $U(0,1)$ distribution, since it is known in this case that the probability integral transformation $Y = F_X(X)$ is $U(0,1)$.

In the following proposition, it is recalled how extropy changes under linear transformations.

Proposition 2.4. *Let X be a non-negative absolutely continuous random variable, and $Y = aX + b$, with $a > 0, b \geq 0$. Then, $J(Y) = \frac{1}{a}J(X)$.*

Remark 2.7. In Proposition 2.4, if we choose $a = 1$, we get the known property that extropy is invariant under translations.

In the following corollary of Proposition 2.3, we discuss how weighted extropy behaves under linear transformations.

Corollary 2.1. *Let X be a non-negative absolutely continuous random variable, and $Y = aX + b$, with $a > 0, b \geq 0$. Then, $J^w(Y) = J^w(X) + \frac{b}{a}J(X)$.*

Remark 2.8. In Corollary 2.1, if we choose $b = 0$, we see that the weighted extropy is invariant for proportional random variables, as observed in Example 2.8 for the case of uniform distribution.

In the following proposition, we give bounds for random variables with support $(0, b)$, with finite b, or $(a, +\infty)$, with $a > 0$. The proof is straightforward and hence it is omitted.

Proposition 2.5. *(i) Let X be a continuous random variable with support $(0, b)$, $b < +\infty$. Then, $J^w(X) \geq bJ(X)$;*

(ii) Let X be a continuous random variable with support $(a, +\infty)$, $a > 0$. Then, $J^w(X) \leq aJ(X)$.

In the following theorem, it is provided a lower bound for the weighted extropy of the sum of two independent random variables.

Theorem 2.7. *Let X and Y be two non-negative independent random variables with densities f_X and f_Y, respectively. Then,*

$$J^w(X + Y) \geq -2\{J(X)J^w(Y) + J^w(X)J(Y)\}. \tag{2.21}$$

Proof. As X and Y are non-negative independent random variables, the density function of $Z = X + Y$ is given, for $z > 0$, by

$$f_Z(z) = \int_0^z f_X(x)f_Y(z - x)dx.$$

Hence, the weighted extropy of Z is obtained as

$$J^w(Z) = -\frac{1}{2}\int_0^{+\infty} z\left[\int_0^z f_X(x)f_Y(z-x)dx\right]^2 dz.$$

By using Jensen's inequality, we get

$$\begin{aligned} J^w(Z) &\geq -\frac{1}{2}\int_0^{+\infty} z\int_0^z f_X^2(x)f_Y^2(z-x)dxdz = -\frac{1}{2}\int_0^{+\infty} f_X^2(x)\int_x^{+\infty} zf_Y^2(z-x)dzdx \\ &= -\frac{1}{2}\int_0^{+\infty} f_X^2(x)\int_0^{+\infty}(z+x)f_Y^2(z)dzdx = \int_0^{+\infty} f_X^2(x)(J^w(X)+xJ(Y))dx \\ &= -2J(X)J^w(Y) - 2J(Y)J^w(X), \end{aligned}$$

as required. □

Remark 2.9. Note that if X and Y are independent and identically distributed, by Theorem 2.7 it follows

$$J^w(X+Y) \geq -4J(X)J^w(X).$$

2.3.1 Weighted residual and past extropies

In this section, we analyze some properties of the weighted residual extropy and the weighted past extropy, see Balakrishnan, Buono and Longobardi [8].

Definition 2.2. Let X be a non-negative absolutely continuous random variable. For all t in the support of the pdf f, we define

(i) the weighted residual extropy of X at time t as

$$J^w(X_t) = -\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty} xf^2(x)dx; \tag{2.22}$$

(ii) the weighted past extropy of X at time t as

$$J^w({}_tX) = -\frac{1}{2F^2(t)}\int_0^t xf^2(x)dx. \tag{2.23}$$

Remark 2.10. The definition in Equation (2.22) is in conformance with other definitions of residual entropies in the literature (see, for instance, Di Crescenzo and Longobardi [38] and Sekeh et al. [103]). Moreover, we can refer to (2.22) as the weighted residual extropy of the first type and introduce the weighted residual extropy of the second type as

$$J^{w*}(X_t) = -\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty}(x-t)f^2(x)dx.$$

These measures are related by a simple relationship involving the residual extropy

$$J^{w*}(X_t) = J^w(X_t) - tJ(X_t).$$

We remark that the use of the first or the second type of weighted residual extropy is based on the way in which we want to give a weight to the observations. More precisely, in the first case, we take into account the time t in the weight, while in the second case we use as weight the time elapsed between t and the value assumed by X.

Remark 2.11. Note that

$$\lim_{t\to 0^+} J^w(X_t) = \lim_{t\to+\infty} J^w({}_tX) = J^w(X).$$

In the following lemma, the first derivatives of the weighted residual extropy and the weighted past extropy are evaluated.

Lemma 2.2. *Let X be a non-negative absolutely continuous random variable with weighted residual extropy $J^w(X_t)$ and weighted past extropy $J^w({}_tX)$. Then,*

(i) $\frac{d}{dt} J^w(X_t) = 2r(t)\left[J^w(X_t) + \frac{t\ r(t)}{4}\right]$, *where $r(t)$ is the hazard rate function of X;*

(ii) $\frac{d}{dt} J^w({}_tX) = -2q(t)\left[J^w({}_tX) + \frac{t\ q(t)}{4}\right]$, *where $q(t)$ is the reversed hazard rate function of X.*

Proof. *(i)* By the definitions of weighted residual extropy and hazard rate function, we have

$$\frac{d}{dt} J^w(X_t) = -\frac{1}{\overline{F}^3(t)} f(t) \int_t^{+\infty} x f^2(x)dx + \frac{1}{2\overline{F}^2(t)} t f^2(t) = 2r(t)\left[J^w(X_t) + \frac{t\ r(t)}{4}\right];$$

(ii) By the definition of weighted past extropy and reversed hazard rate function, we have

$$\frac{d}{dt} J^w({}_tX) = \frac{1}{F^3(t)} f(t) \int_0^t x f^2(x)dx - \frac{1}{2F^2(t)} t f^2(t) = -2q(t)\left[J^w({}_tX) + \frac{t\ q(t)}{4}\right].$$

□

Remark 2.12. We may ask if the weighted residual extropy could be constant over the entire support of a non-negative absolutely continuous random variable. In this regard, if $J^w(X_t)$ is constant, then for all $t > 0$, we have

$$J^w(X_t) + \frac{t\ r(t)}{4} = 0,$$

from which

$$\int_t^{+\infty} x f^2(x)dx = \frac{t}{2} f(t)\overline{F}(t).$$

By differentiating both sides of the above expression, it follows

$$2tf^2(t) = f(t)\overline{F}(t) + tf'(t)\overline{F}(t).$$

It is known that $r'(t) = \frac{f'(t)}{\overline{F}(t)} + r^2(t)$ and so dividing by $\overline{F}^2(t)$ both sides of the above equality, it follows

$$r'(t) = -\frac{r(t)}{t} + 3r^2(t),$$

which is a Bernoulli differential equation with initial condition $r(t_0) = r_0 > 0$, $t_0 > 0$. Upon solving this differential equation, we get

$$r(t) = \frac{1}{t(C - 3\log t)},$$

where $C = \frac{1}{t_0 r_0} + 3\log t_0$. Since the hazard rate function is non-negative, this condition is satisfied if and only if $t \leq e^C/3$. Hence, the weighted residual extropy cannot be constant over $(0, +\infty)$.

In the following proposition, a connection among the weighted extropy, the weighted residual extropy and the weighted past extropy is provided. The proof is straightforward and hence it is omitted.

Proposition 2.6. *The weighted extropy, the weighted residual extropy and the weighted past extropy satisfy the following relationship:*

$$J^w(X) = F^2(t) J^w({}_tX) + \overline{F}^2(t) J^w(X_t).$$

Theorem 2.8. *If X is a non-negative absolutely continuous random variable and if $J^w(X_t)$ is increasing in $t > 0$, then $J^w(X_t)$ uniquely determines the distribution of X.*

Proof. By Lemma 2.2, we have

$$\frac{d}{dt} J^w(X_t) = 2r(t) \left[J^w(X_t) + \frac{t\, r(t)}{4} \right].$$

Consider the function

$$g(x) = 2x \left[J^w(X_t) + \frac{tx}{4} \right] - \frac{d}{dt} J^w(X_t).$$

We know that $g(r(t)) = 0$ and $g(0) = -\frac{d}{dt} J^w(X_t) \leq 0$ since $J^w(X_t)$ is increasing in $t > 0$. Moreover, $\lim_{x \to +\infty} g(x) = +\infty$. If we obtain the derivative of $g(x)$, we observe that there is only one point at which it vanishes; in fact,

$$\frac{d}{dx} g(x) = 2J^w(X_t) + tx$$

and so

$$\frac{d}{dx} g(x) = 0 \Longleftrightarrow x = -\frac{2}{t} J^w(X_t) \; (\geq 0).$$

Then, $g(x) = 0$ has a unique solution and it is $r(t)$ which uniquely determines the distribution, so that $J^w(X_t)$ uniquely determines the distribution as well. $\square$

Theorem 2.9. *If X is a non-negative absolutely continuous random variable and if $J^w({}_tX)$ is decreasing in $t>0$, then $J^w({}_tX)$ uniquely determines the distribution of X.*

Proof. By Lemma 2.2, we have

$$\frac{d}{dt}J^w\left({}_tX\right) = -2q(t)\left[J^w\left({}_tX\right) + \frac{t\ q(t)}{4}\right].$$

Consider the function

$$h(x) = 2x\left[J^w\left({}_tX\right) + \frac{tx}{4}\right] + \frac{d}{dt}J^w\left({}_tX\right).$$

We know that $h(q(t)) = 0$ and $h(0) = \frac{d}{dt}J^w\left({}_tX\right) \leq 0$ being $J^w({}_tX)$ decreasing in $t>0$. Moreover, $\lim_{x\to+\infty} h(x) = +\infty$. If we obtain the derivative of $h(x)$, we observe that there is only one point at which it vanishes; in fact,

$$\frac{d}{dx}h(x) = 2J^w({}_tX) + tx$$

and so

$$\frac{d}{dx}h(x) = 0 \Longleftrightarrow x = -\frac{2}{t}J^w({}_tX)\ (\geq 0).$$

Then, $h(x)=0$ has a unique solution and it is $q(t)$ which uniquely determines the distribution, so that $J^w({}_tX)$ uniquely determines the distribution as well. □

In the following two propositions, we provide bounds for the weighted residual extropy and the weighted past extropy under the monotonicity of hazard rate and reversed hazard rate functions.

Proposition 2.7. *If the hazard rate function $r(t)$ is increasing, then*

$$J^w(X_t) \leq t\,r^2(t)J_s\left(X_t\right), \tag{2.24}$$

where $J_s\left(X_t\right)$ is the dynamic survival extropy defined by Sathar and Nair [101] as

$$J_s\left(X_t\right) = -\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty}\overline{F}^2(x)dx.$$

Proof. By the definition of the weighted residual extropy, it follows

$$J^w\left(X_t\right) = -\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty} xf^2(x)dx = -\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty} xr^2(x)\overline{F}^2(x)dx.$$

As the hazard rate function is increasing by assumption, we obtain

$$\begin{aligned}
-\frac{1}{2\overline{F}^2(t)}\int_t^{+\infty} xr^2(x)\overline{F}^2(x)dx &\leq -\frac{r^2(t)}{2\overline{F}^2(t)}\int_t^{+\infty} x\overline{F}^2(x)dx \\
&\leq -t\frac{r^2(t)}{2\overline{F}^2(t)}\int_t^{+\infty}\overline{F}^2(x)dx = t\,r^2(t)J_s\left(X_t\right),
\end{aligned}$$

as required. □

Proposition 2.8. *If the reversed hazard rate function $q(t)$ is decreasing, then*

$$J^w({}_tX) \leq \frac{1}{4}\left(\frac{1}{2} - t\ q(t)\right). \tag{2.25}$$

Proof. By the definition of the weighted past extropy, we have

$$J^w\left({}_tX\right) = -\frac{1}{2F^2(t)}\int_0^t xf^2(x)dx = -\frac{1}{2F^2(t)}\int_0^t xq(x)F(x)f(x)dx.$$

Integration by parts gives

$$J^w\left({}_tX\right) = -\frac{1}{2F^2(t)}\left[t\ q(t)\frac{F^2(t)}{2} - \int_0^t \left(q(x) + x\ q'(x)\right)\frac{F^2(x)}{2}dx\right].$$

Furthermore, by using the assumption of monotonicity of the reversed hazard rate function, it follows

$$J^w\left({}_tX\right) \leq -\frac{t\ q(t)}{4} + \frac{1}{8}$$

yielding (2.25). □

In the following proposition, we discuss weighted residual extropy and weighted past extropy under monotone transformation.

Proposition 2.9. *Let $Y = \Phi(X)$, where Φ is strictly monotone and differentiable, with derivative Φ'. Then, for all $t > 0$, we have*

$$J^w(Y_t) = \begin{cases} -\frac{1}{2\overline{F}^2_X(\Phi^{-1}(t))}\int_{\Phi^{-1}(t)}^{+\infty}\frac{\Phi(x)}{\Phi'(x)}f_X^2(x)dx, & \textit{if } \Phi \textit{ is strictly increasing} \\ -\frac{1}{2F^2_X(\Phi^{-1}(t))}\int_0^{\Phi^{-1}(t)}\frac{\Phi(x)}{|\Phi'(x)|}f_X^2(x)dx, & \textit{if } \Phi \textit{ is strictly decreasing} \end{cases} \tag{2.26}$$

and

$$J^w({}_tY) = \begin{cases} -\frac{1}{2F^2_X(\Phi^{-1}(t))}\int_0^{\Phi^{-1}(t)}\frac{\Phi(x)}{\Phi'(x)}f_X^2(x)dx, & \textit{if } \Phi \textit{ is strictly increasing} \\ -\frac{1}{2\overline{F}^2_X(\Phi^{-1}(t))}\int_{\Phi^{-1}(t)}^{+\infty}\frac{\Phi(x)}{|\Phi'(x)|}f_X^2(x)dx, & \textit{if } \Phi \textit{ is strictly decreasing.} \end{cases} \tag{2.27}$$

Proof. From (2.22), we have

$$J^w(Y_t) = -\frac{1}{2\overline{F}^2_X(\Phi^{-1}(t))}\int_t^{+\infty} x\frac{f_X^2(\Phi^{-1}(x))}{(\Phi'(\Phi^{-1}(x)))^2}dx.$$

Now, let Φ be strictly increasing. Then, with a change of variable in the above integral, it follows

$$J^w(Y_t) = -\frac{1}{2\overline{F}^2_X(\Phi^{-1}(t))}\int_{\Phi^{-1}(t)}^{+\infty}\frac{\Phi(x)}{\Phi'(x)}f_X^2(x)dx,$$

giving the first expression in (2.26). If Φ is strictly decreasing, the second expression in (2.26) is similarly obtained. The proof of (2.27) is quite similar and is therefore omitted. □

2.4 Interval extropy

Recently, there has been growing attention to study uncertainty measures for doubly truncated random variables which are widely applied in several fields as survival analysis and reliability engineering. In survival analysis, if the lifetime of the item falls in an interval (t_1, t_2), information about lifetime between these two points (also named doubly truncated failure time) is studied, see, for instance, Betensky and Martin [17] and Khorashadizadeh et al. [63]. Then, the random variable $(X \mid t_1 < X < t_2)$ is introduced with pdf $f_{X_{t_1,t_2}}(x) = \frac{f(x)}{F(t_2)-F(t_1)}$ and cdf $F_{X_{t_1,t_2}}(x) = \frac{F(x)-F(t_1)}{F(t_2)-F(t_1)}$, $t_1 < x < t_2$. With this motivation, Sunoj et al. [111] introduced the interval entropy to measure uncertainty in truncated random variable $(X \mid t_1 < X < t_2)$ as follows

$$H(t_1, t_2) = -\int_{t_1}^{t_2} \frac{f(x)}{F(t_2) - F(t_1)} \log \frac{f(x)}{F(t_2) - F(t_1)}\, dx. \tag{2.28}$$

If $t_2 \to +\infty$, then $H(t_1, t_2)$ tends to the residual entropy (2.2). Moreover, if $t_1 \to 0$, then $H(t_1, t_2)$ tends to the past entropy (2.3). Several other properties of the interval entropy were studied by Misagh and Yari [74]. Furthermore, the weighted interval entropy was introduced by Misagh and Yari [73] for doubly truncated random variable $(X \mid t_1 < X < t_2)$ as

$$IH^w(t_1, t_2) = -\int_{t_1}^{t_2} x \frac{f(x)}{F(t_2) - F(t_1)} \log \frac{f(x)}{F(t_2) - F(t_1)}\, dx.$$

In analogy with the interval entropy and the weighted interval entropy, in Buono, Kamari and Longobardi [24] we have introduced the concepts of interval extropy and weighted interval extropy for doubly truncated random variables and have studied some of the properties presented in this section.

Let the random variable $(X \mid t_1 < X < t_2)$ represent the lifetime of a unit which fails between t_1 and t_2 where $(t_1, t_2) \in D = \{(u, v) \in R_+^2 : F(u) < F(v)\}$. The extropy for the doubly truncated random variable is defined as

$$IJ(t_1, t_2) = IJ(X \mid t_1 < X < t_2) = -\frac{1}{2(F(t_2) - F(t_1))^2} \int_{t_1}^{t_2} f^2(x)\, dx, \tag{2.29}$$

and it is an extension of extropy named interval extropy. In (2.29) the dependence of X in the expression $IJ(t_1, t_2)$ has been omitted, but when it is necessary we denote by $IJ_X(t_1, t_2)$ the interval extropy of X to distinguish it from the interval extropy of another random variable.

Remark 2.13. Note that $IJ(0, t_2) = J({}_{t_2}X)$, $IJ(t_1, +\infty) = J(X_{t_1})$ and $IJ(0, +\infty) = J(X)$ are the past extropy, the residual extropy and the extropy, respectively.

Example 2.10. Let $X \sim Exp(\lambda)$, $\lambda > 0$. Based on (2.29), for $0 < t_1 < t_2 < +\infty$, the interval extropy of X is given by

$$IJ(t_1, t_2) = \frac{-1}{2(e^{-\lambda t_1} - e^{-\lambda t_2})^2} \int_{t_1}^{t_2} \lambda^2 e^{-2\lambda x} dx = -\frac{\lambda}{4} \cdot \frac{e^{-\lambda t_2} + e^{-\lambda t_1}}{e^{-\lambda t_1} - e^{-\lambda t_2}}.$$

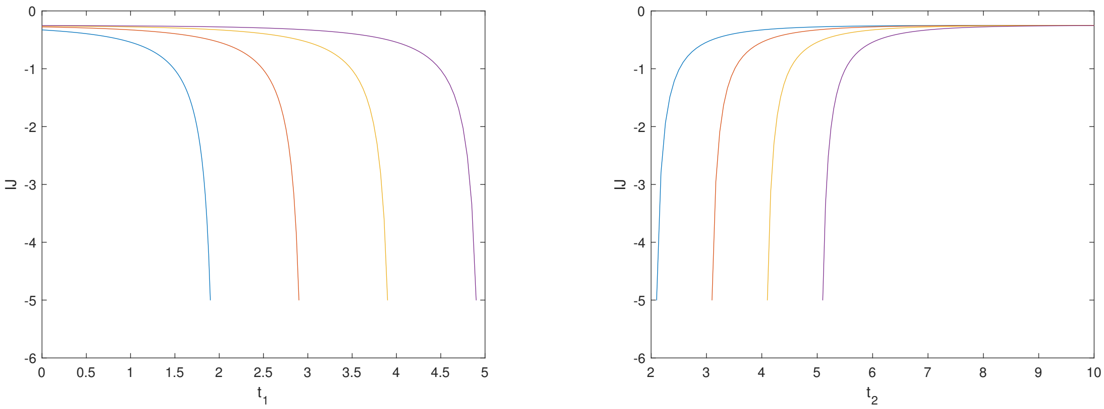

Figure 2.2: Plot of $IJ(t_1, t_2)$ in Example 2.10 as a function of t_1 (left) or t_2 (right) fixing the other one with $t_i = 2$ (blue), 3 (red), 4 (yellow) and 5 (violet), $i = 1, 2$.

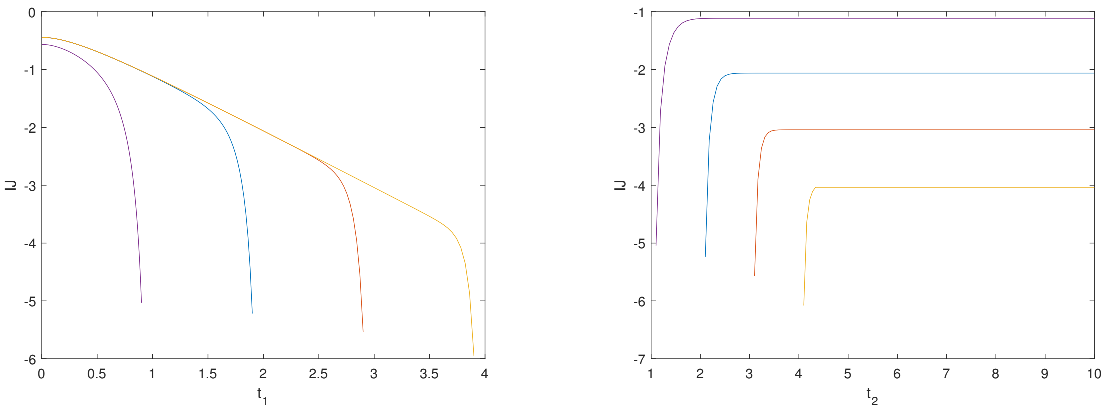

Figure 2.3: Plot of $IJ(t_1, t_2)$ in Example 2.11 as a function of t_1 (left) or t_2 (right) fixing the other one with $t_i = 1$ (violet), 2 (blue), 3 (red) and 4 (yellow), $i = 1, 2$.

In Figure 2.2, the interval extropy is plotted as a function of t_1 for fixed t_2 (Figure 2.2, left) and vice versa (Figure 2.2, right) for $\lambda = 1$.

Example 2.11. Let X follow the Weibull distribution, $W2(\alpha, \lambda)$, with parameters $\alpha = \lambda = 2$, $X \sim W2(2, 2)$. The cdf and the pdf of X are expressed as

$$F(x) = 1 - \exp(-2x^2), \quad f(x) = 4x \exp(-2x^2), \quad x \in (0, +\infty).$$

Since the expression of the interval extropy is not given in terms of elementary functions, in Figure 2.3, the interval extropy is plotted as a function of t_1 for fixed t_2 (Figure 2.3, left) and vice versa (Figure 2.3, right). From Figure 2.3, right, we observe an asymptotic behavior of the interval extropy as $t_2 \to +\infty$ towards $-t_1$, i.e., when the interval extropy $IJ(t_1, t_2)$ reduces

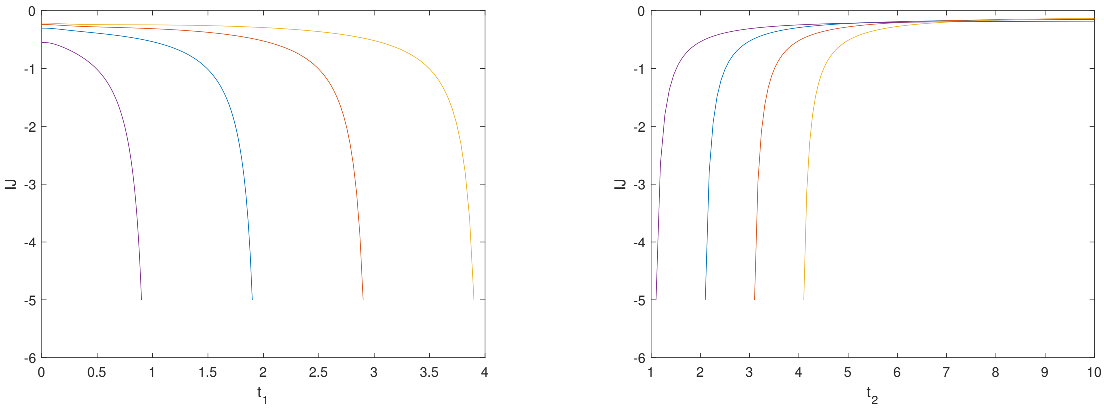

Figure 2.4: Plot of $IJ(t_1, t_2)$ in Example 2.12 as a function of t_1 (left) or t_2 (right) fixing the other one with $t_i = 1$ (violet), 2 (blue), 3 (red) and 4 (yellow), $i = 1, 2$.

to the residual extropy $J(X_{t_1})$. In fact, the residual extropy of X in t can be derived as

$$
\begin{aligned}
J(X_t) &= -\frac{1}{2\exp(-4t^2)}\int_t^{+\infty} 16x^2\exp(-4x^2)dx \\
&= -t - \frac{1}{\exp(-4t^2)}\int_t^{+\infty}\exp(-4x^2)dx \\
&= -t - \frac{1}{2\sqrt{2}\exp(-4t^2)}\int_{2\sqrt{2}t}^{+\infty}\exp\left(-\frac{y^2}{2}\right)dy = -t - \frac{\sqrt{\pi}}{2}\frac{\overline{F}_Z(2\sqrt{2}t)}{\exp(-4t^2)},
\end{aligned}
$$

where $\overline{F}_Z(\cdot)$ is the survival function of $Z \sim N(0,1)$.

Example 2.12. Let X follow the Lognormal distribution, $Lognormal(\mu, \sigma^2)$, with parameters $\mu = 0,\ \sigma^2 = 1,\ X \sim Lognormal(0,1)$. The pdf of X is expressed as

$$f(x) = \frac{1}{x\sqrt{2\pi}}\exp\left(-\frac{\log^2 x}{2}\right), \qquad x \in (0, +\infty).$$

The expression of the interval extropy is not given in terms of elementary functions, hence it is plotted as a function of t_1 for fixed t_2 (Figure 2.4, left) and vice versa (Figure 2.4, right).

Based on the above examples, it could seem that the interval extropy is always decreasing with respect to t_1 and always increasing with respect to t_2. In the following, we provide two counterexamples to show that the interval extropy can be non-monotonous with respect to t_1 and t_2.

Example 2.13. Let X be a random variable with support $S = (a, +\infty)$, $a > 0$ and cdf $F(x) = 1 - \left(\frac{a}{x}\right)^b$, $b > 0$. The interval extropy of X can be obtained as

$$IJ(t_1, t_2) = -\frac{1}{2\left[\left(\frac{a}{t_1}\right)^b - \left(\frac{a}{t_2}\right)^b\right]^2}\int_{t_1}^{t_2}\frac{b^2 a^{2b}}{x^{2b+2}}dx = \frac{b^2\left(t_1^{2b+1} - t_2^{2b+1}\right)}{2(2b+1)t_1 t_2\left(t_2^b - t_1^b\right)^2}.$$

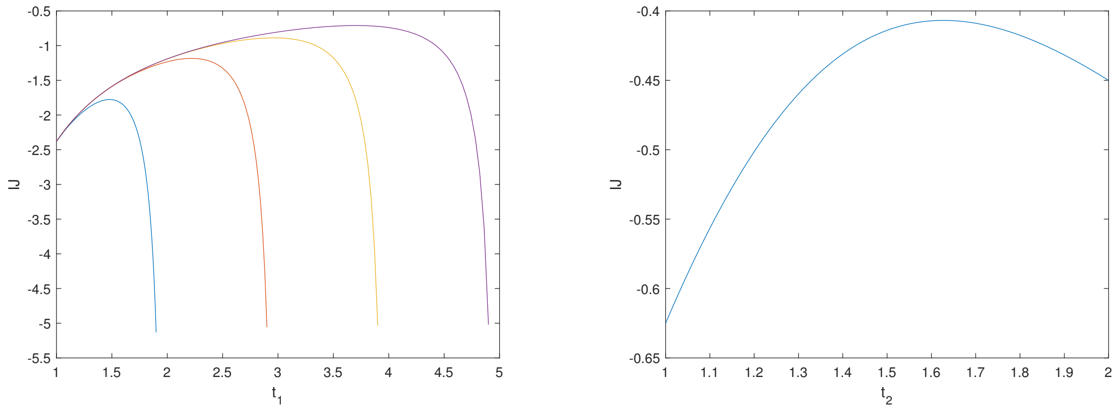

Figure 2.5: Plot of IJ in Example 2.13 as a function of t_1 fixing $t_2 = 2$ (blue), 3 (red), 4 (yellow) and 5 (violet) (left). Plot of IJ in Example 2.14 as a function of $t_2 \in (1, 2)$ fixing $t_1 = 0.1$ (right).

Consider the case $a = 1$ and $b = 10$. In Figure 2.5, left, the interval extropy of X is plotted as a function of t_1 for fixed different values of t_2 and we can observe that it is initially increasing and then decreasing with respect to t_1.

Example 2.14. Let X be a random variable with cdf and pdf respectively defined as

$$F(x) = \begin{cases} \exp\left(-\frac{1}{2} - \frac{1}{x}\right), & \text{if} \quad x \in (0, 1] \\ \exp\left(-2 + \frac{x^2}{2}\right), & \text{if} \quad x \in [1, 2) \end{cases}$$

$$f(x) = \begin{cases} \exp\left(-\frac{1}{2} - \frac{1}{x}\right) \frac{1}{x^2}, & \text{if} \quad x \in (0, 1] \\ \exp\left(-2 + \frac{x^2}{2}\right) x, & \text{if} \quad x \in [1, 2). \end{cases}$$

In Figure 2.5, right, the interval extropy is plotted as a function of $t_2 \in (1, 2)$ with fixed $t_1 = 0.1$ and we can observe a non-monotonic behavior.

In the following proposition, the connection among some different versions of extropy is displayed. The proof is straightforward and hence it is omitted.

Proposition 2.10. *Let X be a non-negative random variable. For all $0 < t_1 < t_2 < +\infty$ the extropy can be decomposed as follows:*

$$J(X) = F^2(t_1)J({}_{t_1}X) + (F(t_2) - F(t_1))^2 IJ(t_1, t_2) + \overline{F}^2(t_2)J(X_{t_2}), \tag{2.30}$$

i.e., the extropy is a function of past extropy, residual extropy and interval extropy.

Definition 2.3. Let X be a non-negative and absolutely continuous random variable with cdf F and pdf f. The Generalized Failure Rate (GFR) functions of X in t_1 and t_2 (with $F(t_2) - F(t_1) > 0$) are defined in [80] as

$$h_i(t_1, t_2) = \frac{f(t_i)}{F(t_2) - F(t_1)}, \qquad i = 1, 2. \tag{2.31}$$

An upper bound in terms of GFR is obtained for the interval extropy in the following proposition.

Proposition 2.11. *Let X be an absolutely continuous non-negative random variable. If the interval extropy is increasing in t_2, then*

$$IJ(t_1,t_2) \le -\frac{h_2(t_1,t_2)}{4}. \tag{2.32}$$

Proof. By differentiating the interval extropy with respect to t_2, we have

$$\frac{\partial IJ(t_1,t_2)}{\partial t_2} = -\frac{h_2^2(t_1,t_2)}{2} - 2h_2(t_1,t_2)IJ(t_1,t_2). \tag{2.33}$$

If $IJ(t_1,t_2)$ is increasing in t_2, then (2.33) implies (2.32). □

In the following proposition, we analyze the effect of linear transformations on the interval extropy. Note that this result could be generalized to monotonic transformations but, in those cases, we do not obtain a formula of interest, in the sense that the interval extropy of the transformed random variable is not expressed in terms of the one of the original random variable.

Proposition 2.12. *Let X be a non-negative and absolutely continuous random variable and let $Y = aX + b$ where $a > 0$ and $b \ge 0$. The interval extropy of Y is given in terms of the interval extropy of X as*

$$IJ_Y(t_1,t_2) = \frac{1}{a}IJ_X\left(\frac{t_1-b}{a}, \frac{t_2-b}{a}\right), \tag{2.34}$$

where $t_1, t_2 \in S_Y$.

Proof. The cdf and the pdf of Y are expressed in terms of F_X and f_X as

$$F_Y(x) = F_X\left(\frac{x-b}{a}\right), \quad f_Y(x) = \frac{1}{a}f_X\left(\frac{x-b}{a}\right).$$

Hence, the interval extropy of Y is expressed as

$$\begin{aligned}
IJ_Y(t_1,t_2) &= -\frac{1}{2\left(F_X\left(\frac{t_2-b}{a}\right) - F_X\left(\frac{t_1-b}{a}\right)\right)^2}\int_{t_1}^{t_2} \frac{1}{a^2} f_X^2\left(\frac{x-b}{a}\right) dx \\
&= -\frac{1}{2\left(F_X\left(\frac{t_2-b}{a}\right) - F_X\left(\frac{t_1-b}{a}\right)\right)^2}\int_{\frac{t_1-b}{a}}^{\frac{t_2-b}{a}} \frac{1}{a} f_X^2(x)\, dx \\
&= \frac{1}{a}IJ_X\left(\frac{t_1-b}{a}, \frac{t_2-b}{a}\right),
\end{aligned}$$

which completes the proof. □

In the following theorem, we give a characterization of the exponential distribution based on the interval extropy.

Theorem 2.10. *Let X be a random variable with support $(0,+\infty)$, differentiable and strictly positive pdf f and cdf F. Then, X is exponentially distributed if, and only if, for all (t_1,t_2) such that $0<t_1<t_2<+\infty$, the following relation holds*

$$IJ(t_1,t_2) = -\frac{1}{4}\left[h_1(t_1,t_2)+h_2(t_1,t_2)\right]. \tag{2.35}$$

Proof. Suppose $X \sim Exp(\lambda)$. By Example 2.10, the interval extropy of X is given by

$$IJ(t_1,t_2) = -\frac{\lambda}{4}\cdot\frac{e^{-\lambda t_2}+e^{-\lambda t_1}}{e^{-\lambda t_1}-e^{-\lambda t_2}}.$$

Moreover, the GFR functions of X are expressed as

$$h_1(t_1,t_2)=\frac{\lambda e^{-\lambda t_1}}{e^{-\lambda t_1}-e^{-\lambda t_2}}, \qquad h_2(t_1,t_2)=\frac{\lambda e^{-\lambda t_2}}{e^{-\lambda t_1}-e^{-\lambda t_2}},$$

and then the first part of the proof is completed.

Conversely, suppose (2.35) holds. Then, by making explicit the interval extropy and GFR functions, we obtain

$$-\frac{1}{2(F(t_2)-F(t_1))^2}\int_{t_1}^{t_2} f^2(x)\,dx = -\frac{f(t_1)+f(t_2)}{4(F(t_2)-F(t_1))},$$

from which it follows

$$\int_{t_1}^{t_2} f^2(x)\,dx = \frac{1}{2}(F(t_2)-F(t_1))(f(t_1)+f(t_2)). \tag{2.36}$$

By differentiating both sides of (2.36) with respect to t_1, we get

$$-f^2(t_1) = -\frac{1}{2}f(t_1)(f(t_1)+f(t_2))+\frac{1}{2}f'(t_1)(F(t_2)-F(t_1)),$$

which reduces to

$$-f^2(t_1)+f(t_1)f(t_2) = f'(t_1)(F(t_2)-F(t_1)). \tag{2.37}$$

By differentiating both sides of (2.37) with respect to t_2, we obtain

$$f(t_1)f'(t_2) = f'(t_1)f(t_2),$$

which is equivalent to

$$\frac{f'(t_1)}{f(t_1)} = \frac{f'(t_2)}{f(t_2)},$$

i.e., the ratio is constant for $x>0$,

$$\frac{f'(x)}{f(x)} = A. \tag{2.38}$$

Hence, by integrating both sides of (2.38) from 0 to t, we get

$$f(t) = f(0)\, e^{At},$$

and in order to satisfy the condition of normalization for the pdf f, we need $A = -f(0)$, i.e., f is the pdf of an exponential distribution. $\square$

It is possible to introduce the weighted version of the interval extropy, namely the weighted interval extropy. Let X be a non-negative absolutely continuous random variable. For all t_1 and t_2 such that $(t_1, t_2) \in D = \{(u, v) \in R_+^2 : F(u) < F(v)\}$ the weighted interval extropy of X is defined as

$$IJ^w(t_1, t_2) = -\frac{1}{2(F(t_2) - F(t_1))^2} \int_{t_1}^{t_2} x f^2(x)\, dx. \tag{2.39}$$

In analogy with the interval extropy, it is possible to study some properties of the weighted interval extropy and provide bounds connected with the generalized failure rate functions. For further details, see Buono, Kamari and Longobardi [24].

2.5 Tsallis extropy

In this section, by resorting to Balakrishnan, Buono and Longobardi [7], the definition of the Tsallis extropy, dual to the Tsallis entropy in (2.4), is given. It is defined to preserve a relationship similar to the one about the sum of Shannon entropy and extropy in (2.7).

Definition 2.4. Let X be a discrete random variable with support $S = \{x_1, \dots, x_N\}$ and with probability vector $\mathbf{p} = (p_1, \dots, p_N)$, and let $\alpha > 0$, $\alpha \neq 1$. Then, the Tsallis extropy of X, $JS_\alpha(X)$, is defined as

$$JS_\alpha(X) = \frac{1}{\alpha - 1}\left(N - 1 - \sum_{i=1}^{N}(1 - p_i)^\alpha\right). \tag{2.40}$$

Remark 2.14. By using the normalization condition, the Tsallis entropy can be expressed as

$$S_\alpha(X) = \frac{1}{\alpha - 1}\sum_{i=1}^{N} p_i(1 - p_i^{\alpha-1}).$$

Hence, it seems that we can simply introduce the Tsallis extropy as

$$\frac{1}{\alpha - 1}\sum_{i=1}^{N}(1 - p_i)\left(1 - (1 - p_i)^{\alpha-1}\right) = \frac{1}{\alpha - 1}\left(\sum_{i=1}^{N}(1 - p_i) - \sum_{i=1}^{N}(1 - p_i)^\alpha\right), \tag{2.41}$$

i.e., by replacing all the occurrences of p_i by $(1 - p_i)$. Although the expression is the same of the one given in (2.40), we will show in Proposition 2.15 that the above definition has a deeper meaning as it preserves the invariance property about the sum of entropy and extropy in (2.7).

Proposition 2.13. *The Tsallis extropy in non-negative.*

Proof. Consider the expression of Tsallis extropy in the LHS of (2.41). For $\alpha > 1$, the function $h(x) = x^{\alpha-1}$ is increasing in $x > 0$, and so $1 - (1-p_i)^{\alpha-1} \geq 0$, and the Tsallis extropy is non-negative. For $0 < \alpha < 1$, the function $h(x) = x^{\alpha-1}$ is decreasing in $x > 0$, that is $1 - (1-p_i)^{\alpha-1} \leq 0$, and then the Tsallis extropy is non-negative due to the multiplicative factor $\frac{1}{\alpha-1}$ being negative. □

In the following proposition, we show that the Tsallis extropy reduces to the extropy in (2.6) when α goes to 1. Notice that this is a classical property of Tsallis and Shannon entropies.

Proposition 2.14. *Let X be a discrete random variable with finite support S and with corresponding probability vector $\mathbf{p}$. Then,*

$$\lim_{\alpha \to 1} JS_\alpha(X) = J(X). \tag{2.42}$$

Proof. From (2.40) and by using L'Hôpital's rule, we obtain

$$\begin{aligned} \lim_{\alpha \to 1} JS_\alpha(X) &= \lim_{\alpha \to 1} \frac{1}{\alpha - 1} \left(N - 1 - \sum_{i=1}^{N} (1-p_i)^\alpha \right) \\ &= -\lim_{\alpha \to 1} \sum_{i=1}^{N} (1-p_i)^\alpha \log(1-p_i) = -\sum_{i=1}^{N} (1-p_i) \log(1-p_i) = J(X). \end{aligned}$$

□

Before discussing the sum of Tsallis entropy and extropy, we need the following lemma about random variables with support of cardinality two.

Lemma 2.3. *Let X be a discrete random variable taking on two values with corresponding probabilities $(p, 1-p)$. Then,*

$$JS_\alpha(X) = S_\alpha(X). \tag{2.43}$$

Proof. From (2.40), with $N = 2$, it follows

$$\begin{aligned} JS_\alpha(X) &= \frac{1}{\alpha - 1} \left[1 - (1-p)^\alpha - \{1 - (1-p)\}^\alpha \right] \\ &= \frac{1}{\alpha - 1} \{1 - p^\alpha - (1-p)^\alpha\} = S_\alpha(X). \end{aligned}$$

□

Proposition 2.15. *Let X be a discrete random variable with finite support S and with probability vector $\mathbf{p}$. Then,*

$$S_\alpha(X) + JS_\alpha(X) = \sum_{i=1}^{N} S_\alpha(p_i, 1-p_i) = \sum_{i=1}^{N} JS_\alpha(p_i, 1-p_i), \tag{2.44}$$

where $S_\alpha(p_i, 1-p_i)$ and $JS_\alpha(p_i, 1-p_i)$ are the Tsallis entropy and extropy of a discrete random variable taking on two values with corresponding probabilities $(p_i, 1-p_i)$.

Proof. We have to prove only the first equality, since the second one is given by Lemma 2.3. From (2.4) and (2.40), we have

$$\begin{aligned}
S_\alpha(X) + JS_\alpha(X) &= \frac{1}{\alpha-1}\left\{N - \sum_{i=1}^{N} p_i^\alpha - \sum_{i=1}^{N}(1-p_i)^\alpha\right\} \\
&= \frac{1}{\alpha-1}\sum_{i=1}^{N}\{1 - p_i^\alpha - (1-p_i)^\alpha\} = \sum_{i=1}^{N} S_\alpha(p_i, 1-p_i),
\end{aligned}$$

which completes the proof. □

In the following proposition, we observe that the Tsallis entropy and extropy coincide for $\alpha = 2$.

Proposition 2.16. *Let X be a discrete random variable with finite support S of cardinality N. Then, $S_2(X) = JS_2(X)$.*

Proof. From (2.40), by choosing $\alpha = 2$, we obtain

$$\begin{aligned}
JS_2(X) &= \frac{1}{2-1}\left(N - 1 - \sum_{i=1}^{N}(1-p_i)^2\right) = N - 1 - \sum_{i=1}^{N}(1 + p_i^2 - 2p_i) \\
&= N - 1 - N - \sum_{i=1}^{N} p_i^2 + 2\sum_{i=1}^{N} p_i = 1 - \sum_{i=1}^{N} p_i^2 = S_2(X).
\end{aligned}$$

□

In the following theorem, we show that the Tsallis entropy is always greater than the Tsallis extropy for $\alpha < 2$ and that the reverse inequality holds for $\alpha > 2$.

Theorem 2.11. *For any discrete random variable X with support of cardinality $N \geq 3$, we have*

$$\begin{aligned}
S_\alpha(X) &\geq JS_\alpha(X) \quad \textit{if} \;\; 0 < \alpha < 2, \\
S_\alpha(X) &\leq JS_\alpha(X) \quad \textit{if} \;\; \alpha > 2.
\end{aligned}$$

Proof. Note that for $\alpha = 1$ we mean the limit case in which we obtain the well-known result about entropy and extropy. Consider the difference between Tsallis entropy and extropy

$$S_\alpha(X) - JS_\alpha(X) = \frac{1}{\alpha-1}\left[2 - N - \sum_{i=1}^{N} p_i^\alpha + \sum_{i=1}^{N}(1-p_i)^\alpha\right].$$

Then, consider the Lagrange function L defined as

$$L = S_\alpha(X) - JS_\alpha(X) + \lambda\left(\sum_{i=1}^{N} p_i - 1\right),$$

for which the partial derivatives with respect to p_i are

$$\frac{\partial L}{\partial p_i} = \frac{-\alpha}{\alpha - 1}\left(p_i^{\alpha-1} + (1-p_i)^{\alpha-1}\right) + \lambda,$$

which vanish if, and only if,

$$p_i^{\alpha-1} + (1-p_i)^{\alpha-1} = C, \tag{2.45}$$

being C a constant. Consider, for $0 \le x \le 1$, the function $h(x) = x^{\alpha-1} + (1-x)^{\alpha-1}$, such that $h(x) = h(1-x)$ and $h(0) = h(1) = 1$. The function h has a minimum at $x = \frac{1}{2}$, if $\alpha > 2$ or $0 < \alpha < 1$, and a maximum at the same point for $1 < \alpha < 2$. Then, to satisfy both (2.45) and the normalization condition, we have only two possibilities. The first one is given by choosing one p_i equal to 1 and all the others equal to 0, whereas the second one is given by $p_i = \frac{1}{N}$, $i = 1, \ldots, N$. These are the cases in which the difference between Tsallis entropy and extropy takes the maximum and the minimum values. In the first case, we have $S_\alpha(X) - JS_\alpha(X) = 0$ whereas in the second one

$$\begin{aligned} S_\alpha(X) - JS_\alpha(X) &= \frac{1}{\alpha-1}\left[2 - N - \sum_{i=1}^{N}\frac{1}{N^\alpha} + \sum_{i=1}^{N}\left(1-\frac{1}{N}\right)^\alpha\right] \\ &= \frac{2N^{\alpha-1} - N^\alpha - 1 + (N-1)^\alpha}{(\alpha-1)N^{\alpha-1}}. \end{aligned} \tag{2.46}$$

Consider the numerator of (2.46) as a function of α, $g(\alpha) = 2N^{\alpha-1} - N^\alpha - 1 + (N-1)^\alpha$. Then,

$$g'(\alpha) = N^{\alpha-1}(\log N)(2-N) + (N-1)^\alpha \log(N-1),$$

which is non-negative if, and only if,

$$\alpha \le \frac{\log\left[\frac{N}{N-2}\frac{\log(N-1)}{\log N}\right]}{\log\left(\frac{N}{N-1}\right)} = G(N).$$

We have that, for $N \ge 3$,

$$1 < G(N) < 2 \Longleftrightarrow \frac{N-2}{N-1} < \frac{\log(N-1)}{\log N} < \frac{N(N-2)}{(N-1)^2}, \tag{2.47}$$

which holds as can be seen in Figure 2.6.

Then, the function g has a maximum between 1 and 2 and $g(1) = g(2) = 0$. Hence, we have $g(\alpha) > 0$ if $1 < \alpha < 2$ and $g(\alpha) < 0$ if $0 < \alpha < 1$ or $\alpha > 2$. By recalling the definition of g and (2.46), we obtain that the difference between Tsallis entropy and extropy for uniform distribution is greater than 0 if $0 < \alpha < 1$ or $1 < \alpha < 2$ and less than 0 if $\alpha > 2$. Hence, $S_\alpha(X) - JS_\alpha(X)$ has minimum of 0 and maximum for the uniform distribution if $0 < \alpha < 2$ and vice versa if $\alpha > 2$. □

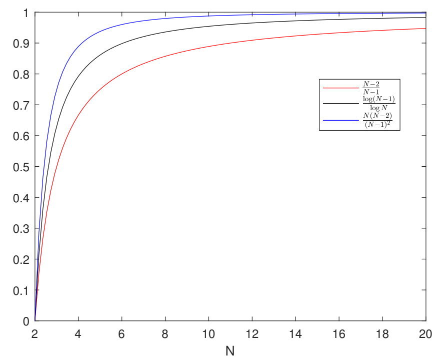

Figure 2.6: Plot of the functions on the RHS of (2.47) in red, black and blue, respectively.

2.5.1 The maximum Tsallis extropy

Dealing with a measure of information, it is useful to know its maximum value. The Tsallis extropy reaches its maximum value if the random variable X is uniformly distributed, as established in the following theorem.

Theorem 2.12. *Let X be a discrete random variable with finite support S of cardinality N, and let $\alpha > 0$, $\alpha \neq 1$. Then, X has maximum Tsallis extropy for fixed N and α if, and only if, it is uniformly distributed.*

Proof. Let N and α be fixed. Then, we need to maximize the function of N variables given by

$$JS_\alpha(\mathbf{p}) = \frac{1}{\alpha-1}\left(N-1-\sum_{i=1}^{N}(1-p_i)^\alpha\right),$$

subject to the normalization condition

$$\sum_{i=1}^{N} p_i = 1. \tag{2.48}$$

Consider the Lagrange function defined by

$$JS^*_\alpha(\mathbf{p}) = \frac{1}{\alpha-1}\left(N-1-\sum_{i=1}^{N}(1-p_i)^\alpha\right) + \lambda\left(\sum_{i=1}^{N} p_i - 1\right),$$

whose partial derivatives with respect to p_i, $i = 1, \ldots, N$, are

$$\frac{\partial JS^*_\alpha(\mathbf{p})}{\partial p_i} = \frac{\alpha}{\alpha-1}(1-p_i)^{\alpha-1} + \lambda.$$

Then, we obtain the stationary points as

$$\frac{\alpha}{\alpha-1}(1-p_i)^{\alpha-1} + \lambda = 0 \Longleftrightarrow p_i = 1 - \left(\frac{(1-\alpha)\lambda}{\alpha}\right)^{\frac{1}{\alpha-1}} = K,$$

where K is a constant. In order to satisfy the condition in (2.48), we need $K = \frac{1}{N}$, and then $\mathbf{p}$ becomes the probability mass function vector of a discrete uniform distribution. $\square$

Remark 2.15. Let X be a discrete random variable uniformly distributed over $\{1, \dots, N\}$. Then, the maximum Tsallis extropy is given by

$$\begin{aligned} JS_\alpha(X) &= \frac{1}{\alpha-1}\left\{N-1-\sum_{i=1}^{N}\left(1-\frac{1}{N}\right)^\alpha\right\} \\ &= \frac{1}{\alpha-1}\left\{N-1-\frac{(N-1)^\alpha}{N^{\alpha-1}}\right\} = \frac{N-1}{\alpha-1}\frac{\{N^{\alpha-1}-(N-1)^{\alpha-1}\}}{N^{\alpha-1}}. \end{aligned} \tag{2.49}$$

Theorem 2.13. *The Tsallis extropy is less than* 1.

Proof. To prove the statement, we show that the Tsallis extropy of a discrete uniform distribution increases to 1 as the size of the support N increases. Let X_N be a discrete random variable uniformly distributed over a finite support of size N. From (2.49), the corresponding Tsallis extropy is given by

$$JS_\alpha(X_N) = \frac{1}{\alpha-1}\left(N-1-\frac{(N-1)^\alpha}{N^{\alpha-1}}\right).$$

Consider the function

$$g(N) = N-1-\frac{(N-1)^\alpha}{N^{\alpha-1}}, \tag{2.50}$$

we show that it increases for $\alpha > 1$ and decreases for $0 < \alpha < 1$. In this way, we will prove that $JS_\alpha(X_N)$ is increasing in N. By treating g as a function of a continuous variable N, its first derivative is given as

$$g'(N) = 1-\frac{(N-1)^{\alpha-1}}{N^\alpha}(\alpha+N-1) = \frac{N^\alpha-(N-1)^{\alpha-1}(\alpha+N-1)}{N^\alpha},$$

whose sign is determined by

$$N^\alpha-(N-1)^{\alpha-1}(\alpha+N-1) = N^\alpha-(N-1)^\alpha-\alpha(N-1)^{\alpha-1},$$

which, by the mean value theorem, is equal to

$$\alpha(N-1+\varepsilon)^{\alpha-1}-\alpha(N-1)^{\alpha-1},$$

for some $\varepsilon \in (0,1)$. Hence, by using that the function $h(x) = x^{\alpha-1}$ is increasing in $x > 0$ for $\alpha > 1$ and decreasing for $0 < \alpha < 1$, we get the monotonicity of $g(N)$ in (2.50).

Now, we evaluate the limit of $JS_\alpha(X_N)$ as N tends to infinity and we obtain

$$\begin{aligned} \lim_{N\to+\infty}\frac{N-1}{\alpha-1}\frac{(N^{\alpha-1}-(N-1)^{\alpha-1})}{N^{\alpha-1}} &= \lim_{N\to+\infty}\frac{N-1}{\alpha-1}\left[1-\left(1-\frac{1}{N}\right)^{\alpha-1}\right] \\ &= \lim_{N\to+\infty}\frac{N-1}{N} = 1. \end{aligned}$$

Finally, by using the result in Theorem 2.12 about the maximum Tsallis extropy, we conclude that the Tsallis extropy is less than 1 for any discrete random variable. $\square$

Corollary 2.2. *For any discrete random variable X, we have*

$$0 \leq JS_\alpha(X) < 1.$$

Proof. The result follows by Proposition 2.13 and Theorem 2.13. □

2.6 Past varentropy

As already mentioned, it follows from (2.1) that the Shannon entropy represents the expectation of the (random) information content $IC(X)$. But for different purposes (see, for instance, Bobkov and Madiman [20]), it is also possible to consider its variance, in order to evaluate the concentration of the information content around the entropy $H(X)$. Thus, we can be interested in the varentropy of X (also known as minimal coding variance of X, whenever X is discrete), defined as

$$\begin{aligned} V_e(X) &= \mathrm{Var}[IC(X)] = \mathrm{Var}[-\log f(X)] \\ &= \mathrm{Var}[\log f(X)] = \mathbb{E}\left[(\log f(X))^2\right] - [H(X)]^2 \\ &= \int_0^{+\infty} f(x)[\log f(x)]^2 dx - \left[\int_0^{+\infty} f(x)\log f(x) dx\right]^2. \end{aligned} \tag{2.51}$$

In recent literature, several papers deal with properties and applications of the varentropy, such as Madiman and Wang [72], Arikan [2] and references therein. Furthermore, by the entropy and the varentropy, it is possible to define reference intervals for the information content $IC(X)$ as

$$\mathbb{E}[IC(X)] \pm k\sqrt{\mathrm{Var}[IC(X)]} = H(X) \pm k\sqrt{V_e(X)} \tag{2.52}$$

for suitable choices of k. In the statistics field, such intervals can be used to evaluate the uncertainty about likelihood estimates.

We remind that the Shannon entropy, as well as the varentropy, provides a measure of information for the random lifetime of an item which is new, when X represents its lifetime. For such reason, different time dependent versions of this measure have been proposed in the context of reliability and survival analysis, as the residual entropy recalled in (2.2). In analogy, it can be defined a dynamic version of the varentropy, useful to evaluate the concentration of the information content in residual lifetimes. This is the residual varentropy, studied in details in Di Crescenzo and Paolillo [40] and Paolillo et al. [91], defined as

$$\begin{aligned} V_e(X_t) &= \mathrm{Var}[IC(X_t)] = \mathrm{Var}[-\log f_{X_t}(X_t)] \\ &= \mathrm{Var}[\log f_{X_t}(X_t)] = \mathbb{E}\left[(\log f_{X_t}(X_t))^2\right] - [H(X_t)]^2 \\ &= \int_t^{+\infty} \frac{f(x)}{\overline{F}(t)}\left[\log\frac{f(x)}{\overline{F}(t)}\right]^2 dx - \left[\int_t^{+\infty} \frac{f(x)}{\overline{F}(t)}\log\frac{f(x)}{\overline{F}(t)} dx\right]^2. \end{aligned}$$

A large number of studies in reliability theory deals with past lifetime, that is the random variable conditioned on the fact that the failure occurs before a specified inspection time t. In many situations, uncertainty can refer to the past instead of the future. Recall that given the absolutely continuous random lifetime X, having support $\mathcal{S} \subseteq \mathbb{R}^+$, cdf F and pdf f, its past lifetime at time $t \in \mathcal{S}$ is the variable ${}_tX = (X|X \leq t)$. The corresponding past entropy has been recalled in (2.3) as

$$H({}_tX) = \mathbb{E}[IC({}_tX)] = \mathbb{E}[-\log f_{{}_tX}({}_tX)] = -\int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx. \tag{2.53}$$

In this section based on Buono, Longobardi and Pellerey [26], the past varentropy is defined and some of its properties are studied. The past varentropy can be defined, for all $t \in \mathcal{S}$, as

$$\begin{aligned} V_e({}_tX) &= \mathrm{Var}[IC({}_tX)] = \mathrm{Var}[-\log f_{{}_tX}({}_tX)] \\ &= \mathrm{Var}[\log f_{{}_tX}({}_tX)] = \mathbb{E}\left[(\log f_{{}_tX}({}_tX))^2\right] - [H({}_tX)]^2 \\ &= \int_0^t \frac{f(x)}{F(t)} \left[\log \frac{f(x)}{F(t)}\right]^2 dx - \left[\int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx\right]^2. \end{aligned} \tag{2.54}$$

It can be expressed in a different way by using the reversed hazard rate function q given in (1.4) and the cumulative reversed hazard rate function defined as

$$Q(t) = \int_t^{+\infty} q(x)dx = -\log F(t), \tag{2.55}$$

see, for instance, Li and Li [69]. Also, note that the past entropy can be expressed as

$$\begin{aligned} H({}_tX) &= -Q(t) - \frac{1}{F(t)} \int_0^t f(x) \log f(x) dx \\ &= 1 - \frac{1}{F(t)} \int_0^t f(x) \log q(x) dx, \end{aligned} \tag{2.56}$$

as shown in Di Crescenzo and Longobardi [37]. Thus, by (2.54) and (2.56), it follows for $t \in \mathcal{S}$

$$\begin{aligned} V_e({}_tX) &= \int_0^t \frac{f(x)}{F(t)} \left[\log \frac{f(x)}{F(t)}\right]^2 dx - [H({}_tX)]^2 \\ &= \frac{1}{F(t)} \int_0^t f(x)(\log f(x))^2 dx + \log^2 F(t) - \frac{2\log F(t)}{F(t)} \int_0^t f(x) \log f(x) dx - [H({}_tX)]^2 \\ &= \frac{1}{F(t)} \int_0^t f(x)(\log f(x))^2 dx + (Q(t))^2 + \frac{2Q(t)}{F(t)} \int_0^t f(x) \log f(x) dx - [H({}_tX)]^2 \\ &= \frac{1}{F(t)} \int_0^t f(x)(\log f(x))^2 dx + (Q(t))^2 - 2Q(t)\left[Q(t) + H({}_tX)\right] - [H({}_tX)]^2 \\ &= \frac{1}{F(t)} \int_0^t f(x)(\log f(x))^2 dx - (Q(t) + H({}_tX))^2. \end{aligned} \tag{2.57}$$

Remark 2.16. As well as, when t tends to the supremum of the support $\mathcal{S}$, u_X, the past entropy tends to Shannon entropy, also the past varentropy reduces to the varentropy, i.e.,

$\lim_{t\to u_X} V_e({}_tX) = V_e(X)$. Now, consider the case in which t tends to the infimum of the support, l_X. If the pdf f of X is differentiable and such that

$$\lim_{t\to l_X^+} f(t) \neq 0 \quad \text{and} \quad \lim_{t\to l_X^+} f'(t) \neq +\infty, \tag{2.58}$$

then $\lim_{t\to l_X^+} V_e({}_tX) = 0$. In fact, from (2.57) the past varentropy can be expressed as

$$V_e({}_tX) = \frac{F(t)\int_{l_X}^t f(x)\,(\log f(x))^2\,dx - \left(\int_{l_X}^t f(x)\log f(x)dx\right)^2}{F^2(t)},$$

and by using L'Hôpital's rule twice, it readily follows

$$\begin{aligned}
\lim_{t\to l_X^+} V_e({}_tX) &= \lim_{t\to l_X^+}\left(\frac{f'(t)F(t)}{f^2(t)}\log f(t) - \frac{f'(t)}{f^2(t)}\int_{l_X}^t f(x)\log f(x)dx\right)\\
&= \lim_{t\to l_X^+}\frac{f'(t)}{f^2(t)}\int_{l_X}^t \frac{F(x)f'(x)}{f(x)}dx = 0,
\end{aligned}$$

where the last equality depends on the assumptions in (2.58).

In the following, some examples of evaluation of the past entropy and the past varentropy are given by using (2.53) and (2.57).

- Let X be a random variable with uniform distribution over $(0, b)$, $X \sim U(0, b)$, $b > 0$. Hence, for $t \in (0, b)$ we have

$$H({}_tX) = \log t, \qquad V_e({}_tX) = 0.$$

- Let X be a random variable with exponential distribution, $X \sim Exp(\lambda)$, for $\lambda > 0$. Then, for $t > 0$ we have

$$\begin{aligned}
H({}_tX) &= 1 + \log\left(\frac{1-e^{-\lambda t}}{\lambda}\right) - \frac{\lambda t e^{-\lambda t}}{1-e^{-\lambda t}}\\
V_e({}_tX) &= 1 - \frac{\lambda^2 t^2 e^{-\lambda t}}{(1-e^{-\lambda t})^2}.
\end{aligned}$$

  The plots of past entropy and past varentropy are shown in Figure 2.7 for different choices of λ. Note that $\lim_{t\to 0^+} V_e({}_tX) = 0$, as expected, since the exponential distribution satisfies the assumptions given in (2.58) for any value of λ.

- Let X be a random variable such that $f(x) = 2x$ and $F(x) = x^2$, $x \in (0, 1)$. Hence, for $t \in (0, 1)$ we have

$$H({}_tX) = \frac{1}{2} + \log\frac{t}{2}, \qquad V_e({}_tX) = \frac{1}{4}.$$

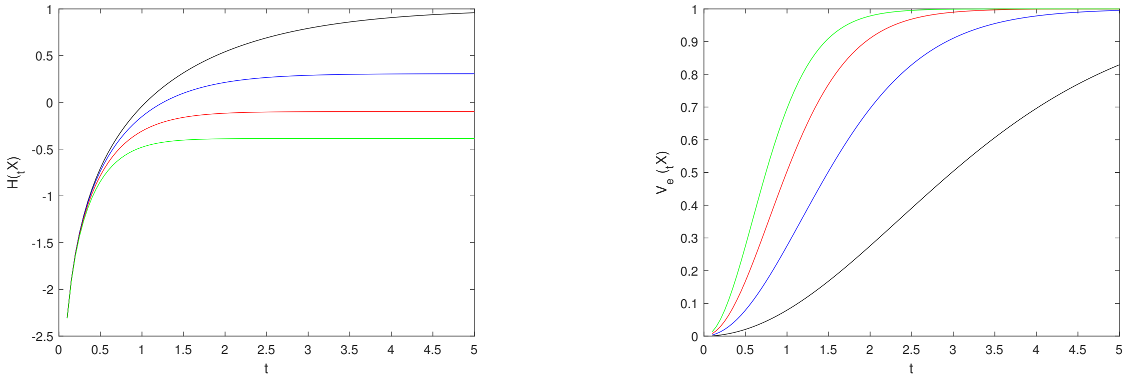

Figure 2.7: Plots of past entropy (left) and past varentropy (right) of exponential distribution with parameter $\lambda = 1, 2, 3, 4$ (black, blue, red and green, respectively).

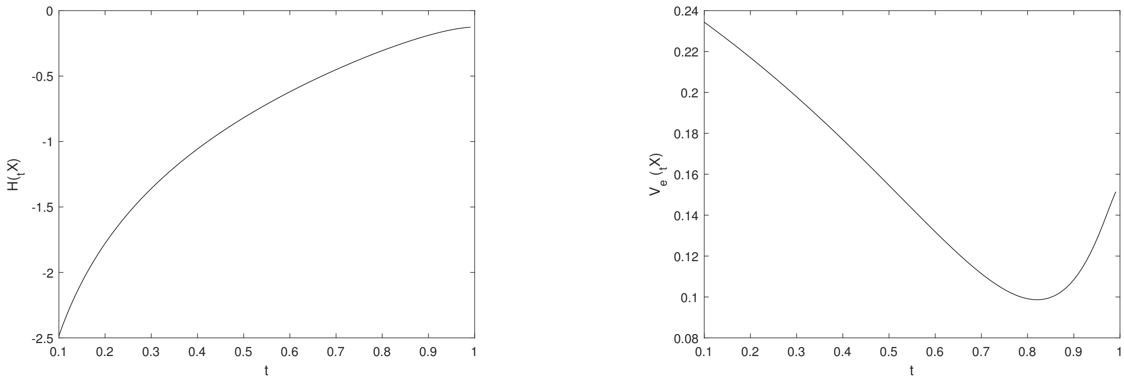

Figure 2.8: Plots of past entropy (left) and past varentropy (right) of $X \sim Beta(2, 2)$.

- Let X be a random variable with $Beta(2, 2)$ distribution, i.e., $f(x) = 6x(1-x)$ and $F(x) = 3x^2 - 2x^3$, $x \in (0, 1)$. Hence, for $t \in (0, 1)$ we have

$$H({}_tX) = \frac{1}{t^2/2 - t^3/3}\left[\left(\frac{t^2}{2} - \frac{t^3}{3}\right)\log\left(\frac{6(1-t)}{t(3-2t)}\right) + \frac{2}{9}t^3 - \frac{1}{3}t^2 - \frac{1}{6}t - \frac{1}{6}\log(1-t)\right]$$

$$\begin{aligned} V_e({}_tX) = \frac{1}{t^2/2 - t^3/3}&\left[\left(\frac{t^2}{2} - \frac{t^3}{3}\right)\log^2\left(\frac{6(1-t)}{t(3-2t)}\right) + \frac{1}{3}\left(\frac{4}{3}t^3 - 2t^2 - t\right)\log\left(\frac{6(1-t)}{t(3-2t)}\right)\right. \\ &+\frac{1}{9}\left(-\frac{8}{3}t^3 + 4t^2 + 8t + 5\log(1-t)\right) - \frac{1}{3}\log\left(\frac{1}{t^2/2 - t^3/3}\right)\log(1-t) \\ &\left.-\frac{1}{6}\log^2(1-t) - \frac{\pi^2}{18} + \frac{1}{3}Li_2(1-t)\right)\Bigg] - [H({}_tX)]^2 \end{aligned}$$

  where Li_2 is the Spence's function or dilogarithm function, see Morris [75]. The plots of these past entropy and past varentropy are shown in Figure 2.8.

- Let X be a random variable with cdf $F(x) = 1 - \left(\frac{b-x}{b}\right)^{\alpha}$, for $x \in (0, b) \subseteq \mathbb{R}^+$ and $\alpha > 0$.

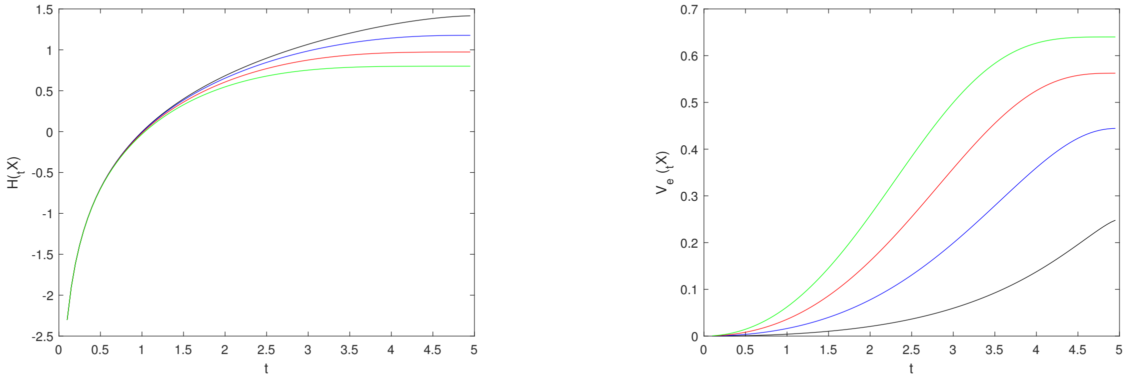

Figure 2.9: Plots of past entropy (left) and past varentropy (right) of X with cdf $F(x) = 1 - \left(\frac{b-x}{b}\right)^{\alpha}$ for $b = 5$ and $\alpha = 2, 3, 4, 5$ (black, blue, red and green, respectively).

Then, for $t \in (0, b)$

$$H({}_tX) = \frac{b^\alpha}{b^\alpha - (b-t)^\alpha} \log\left(\frac{\alpha b^{(\alpha-1)}}{b^\alpha - (b-t)^\alpha}\right) - \frac{(b-t)^\alpha}{b^\alpha - (b-t)^\alpha} \log\left(\frac{\alpha (b-t)^{(\alpha-1)}}{b^\alpha - (b-t)^\alpha}\right) - \frac{\alpha - 1}{\alpha}$$

$$V_e({}_tX) = \left(\frac{\alpha-1}{\alpha}\right)^2 - \frac{b^\alpha (b-t)^\alpha}{[b^\alpha - (b-t)^\alpha]^2} \log^2\left[\left(\frac{b}{b-t}\right)^{\alpha-1}\right].$$

The plots of this past entropy and of the corresponding past varentropy are shown in Figure 2.9.

Note that the past varentropy is constant in two of the cases described above, increasing in one case, and non-monotone in the other one. Thus, monotonicity of the varentropy is not always guaranteed. We recall that, if the reversed hazard rate is decreasing in t, then the past entropy is increasing in t (see Di Crescenzo and Longobardi [37], Proposition 2.2). However, the monotonicity of the reversed hazard rate is not a sufficient condition for the monotonicity of the past varentropy, as shown for the $Beta(2, 2)$ distribution whose past varentropy is not monotone but whose reversed hazard rate $q(t) = 6(1-t)/(3t - 2t^2)$ is decreasing. For this reason, conditions for the monotonicity of $V_e({}_tX)$ and an implicit formula for the derivative of the past varentropy are now described. Before stating the following result, we remind the definition of two stochastic comparisons orders used in the proof. Given the random variables X_1 and X_2 with distributions F_1 and F_2, respectively, we say that X_1 is smaller than X_2 in the convex transform order, $X_1 \le_c X_2$, if $F_2^{-1}(F_1(x))$ is convex on the support of F_1. We say that X_1 is smaller than X_2 in the star order, $X_1 \le_* X_2$, if $F_2^{-1}(F_1(x))/x$ is increasing on the support of F_1. See Shaked and Shanthikumar [106] for further details.

Proposition 2.17. *Let X be a random lifetime with an absolutely continuous cdf F and a* *strictly decreasing [increasing] pdf f. If the ratio*

$$\frac{f(F^{-1}(pF(s)))}{f(F^{-1}(pF(t)))} \tag{2.59}$$

is increasing in $p \in (0,1)$ for all $s \leq t$, then the corresponding past varentropy $V_e({}_tX)$ is increasing [decreasing] in $t \in \mathcal{S}$.

Proof. Recall that, for any $t \in \mathcal{S}$, the past lifetime ${}_tX$ has pdf $f_{{}_tX}(x) = f(x)/F(t)$ and cdf $F_{{}_tX}(x) = F(x)/F(t)$, with $x \leq t$. Then, the corresponding quantile function is $F^{-1}_{{}_tX}(p) = F^{-1}(pF(t))$, for $p \in (0,1)$. Also observe that, for $s \leq t$,

$$\frac{f_{{}_sX}(F^{-1}_{{}_sX}(p))}{f_{{}_tX}(F^{-1}_{{}_tX}(p))} = \frac{f(F^{-1}(pF(s)))}{F(s)} \cdot \frac{F(t)}{f(F^{-1}(pF(t)))} = \frac{f(F^{-1}(pF(s)))}{f(F^{-1}(pF(t)))} \cdot \frac{F(t)}{F(s)},$$

where the latter is increasing in p by assumption (2.59). Then, it follows ${}_sX \leq_c {}_tX$ (see Remark 4.3 in Paolillo et al. [91]). Observe that, since f is decreasing [increasing] by assumption, also $f_{{}_sX}$ and $f_{{}_tX}$ are decreasing [increasing]. Thus, by the equivalence pointed out in Remark 4.6 in Paolillo et al. [91], one also has $f_{{}_sX}({}_sX) \leq_* f_{{}_tX}({}_tX)$ $[f_{{}_sX}({}_sX) \geq_* f_{{}_tX}({}_tX)]$, which implies $V_e({}_sX) \leq V_e({}_tX)$ $[V_e({}_sX) \geq V_e({}_tX)]$ by Theorem 5.2 in the same paper. □

It is easy to verify, for example, that exponential distributions satisfy the assumptions of Proposition 2.17 for any value of λ. The following result provides an implicit formula for the derivative of the past varentropy, useful to describe distributions with constant varentropy.

Proposition 2.18. *For all $t \in \mathcal{S}$, the derivative of the past varentropy is*

$$V_e'({}_tX) = -q(t)\left[V_e({}_tX) - (H({}_tX) + \log q(t))^2\right].$$

Proof. First observe that by differentiating both sides of (2.56) we get the following expression for the derivative of the past entropy:

$$H'({}_tX) = q(t)[1 - H({}_tX) - \log q(t)]. \tag{2.60}$$

Consider now (2.57). By differentiating both sides it follows

$$V_e'({}_tX) = \frac{q(t)}{F(t)}\int_0^t f(x)(\log f(x))^2 dx + q(t)(\log f(t))^2 - 2(Q(t) + H({}_tX))(-q(t) + H'({}_tX)). \tag{2.61}$$

Hence, recalling (2.60) and (2.57), from (2.61) we get

$$V_e'({}_tX) = -q(t)\left[V_e({}_tX) + (Q(t) + H({}_tX))^2 - \log^2 f(t) - 2(Q(t) + H({}_tX))(H({}_tX) + \log q(t))\right]$$

and, after straightforward calculations, the proof is completed. □

From Proposition 2.18 one can obtain conditions such that absolutely continuous distributions, with continuous densities, have a corresponding constant past varentropy. Consider first the case of random variables having support $\mathcal{S} = [0, 1]$.

Proposition 2.19. *Let X have support $\mathcal{S} = [0, 1]$. Then, its varentropy $V_e({}_tX)$ is constant if, and only if, X has cdf*

$$F(x) = x^{\alpha}, \qquad x \in [0, 1], \tag{2.62}$$

where $\alpha > 0$ is a parameter. In this case, $V_e({}_tX) = (1 - 1/\alpha)^2$ for all $t \in [0, 1]$.

Proof. If X has the cdf defined in (2.62), then the pdf is given by

$$f(x) = \alpha x^{\alpha-1}, \qquad x \in (0, 1),$$

and, for $t \in (0, 1)$, the past varentropy is

$$V_e({}_tX) = \int_0^t \frac{\alpha x^{\alpha-1}}{t^{\alpha}} \left[\log\left(\frac{\alpha x^{\alpha-1}}{t^{\alpha}}\right)\right]^2 dx - \left[\int_0^t \frac{\alpha x^{\alpha-1}}{t^{\alpha}} \log\left(\frac{\alpha x^{\alpha-1}}{t^{\alpha}}\right) dx\right]^2.$$

By the change of variable $y = \left(\frac{x}{t}\right)^{\alpha}$, we get

$$\begin{aligned} V_e({}_tX) &= \int_0^1 \left[\log\left(\frac{\alpha y^{(\alpha-1)/\alpha}}{t}\right)\right]^2 dy - \left[\int_0^1 \log\left(\frac{\alpha y^{(\alpha-1)/\alpha}}{t}\right) dy\right]^2 \\ &= \log^2\left(\frac{\alpha}{t}\right) - 2\left(\frac{\alpha-1}{\alpha}\right)\log\left(\frac{\alpha}{t}\right) + 2\left(\frac{\alpha-1}{\alpha}\right)^2 - \left[\log\left(\frac{\alpha}{t}\right) - \frac{\alpha-1}{\alpha}\right]^2. \end{aligned}$$

Thus $V_e({}_tX)$ is constant and equal to $(1 - 1/\alpha)^2$. It follows now, from Proposition 2.18, that

$$(H({}_tX) + \log q(t))^2 = (1 - 1/\alpha)^2,$$

and so

$$|H({}_tX) + \log q(t)| = |1 - 1/\alpha|, \qquad \forall t \in [0, 1]. \tag{2.63}$$

Since the pdf f is continuous by assumption, then also q and $H(X_t)$ are continuous. Thus, $H({}_tX) + \log q(t)$ is continuous in $t \in [0, 1]$, so that equality (2.63) implies

$$H({}_tX) + \log q(t) = c, \qquad \forall t \in [0, 1]. \tag{2.64}$$

for some $c \in \mathbb{R}$. As shown in Kundu et al. [66], Theorem 2.1, there exist only three families of distributions for which (2.64) is satisfied. Two of them have infinite support on the left, i.e., of the form $(-\infty, b]$, for $b \in \mathbb{R}$ (thus they cannot be distributions of random lifetimes), and the only one having support entirely contained in $\mathbb{R}^+$ (and in $[0, 1]$ in particular) is the one defined in (2.62). $\square$

To generalize the above result to random lifetimes having different supports, we can use the following proposition, based on the past varentropy under linear transformations. We recall that if $Y = aX + b$ for $a > 0$ and $b \geq 0$, then, as shown in Di Crescenzo and Longobardi [37], the past entropies of X and Y are related by

$$H({}_tY) = H\left({}_{\frac{t-b}{a}}X\right) + \log a \qquad \forall t. \tag{2.65}$$

Proposition 2.20. *Let $Y = aX + b$, with $a > 0$ and $b \geq 0$. Then, their past varentropies are related by*

$$V_e({}_tY) = V_e\left({}_{\frac{t-b}{a}}X\right), \qquad \forall t. \tag{2.66}$$

Proof. From $Y = aX + b$ we know that $F_Y(x) = F_X\left(\frac{x-b}{a}\right)$ and $f_Y(x) = \frac{1}{a} f_X\left(\frac{x-b}{a}\right)$. Hence, from (2.54) and (2.65), we get

$$V_e({}_tY) = \int_0^{\frac{t-b}{a}} \frac{f_X(x)}{F_X\left(\frac{t-b}{a}\right)} \left(\log \frac{\frac{1}{a} f_X(x)}{F_X\left(\frac{t-b}{a}\right)}\right)^2 dx - \left(H\left({}_{\frac{t-b}{a}}X\right) + \log a\right)^2. \tag{2.67}$$

By writing

$$\log \frac{\frac{1}{a} f_X(x)}{F_X\left(\frac{t-b}{a}\right)} = \log \frac{f_X(x)}{F_X\left(\frac{t-b}{a}\right)} - \log a,$$

and developing the two squares in (2.67), the statement easily follows. □

From Propositions 2.19 and 2.20, the following corollary immediately follows.

Corollary 2.3. *Let X be an absolutely continuous random lifetime with continuous pdf f. Then, its varentropy $V_e({}_tX)$ is constant if, and only if, X has cdf in the family*

$$F(x) = \left(\frac{x-b}{a}\right)^\alpha, \qquad x \in [b, a+b], \tag{2.68}$$

for a parameter $\alpha > 0$.

A generalization of Proposition 2.20 will now be stated. Let ϕ be a differentiable and strictly monotone function and let $Y = \phi(X)$ for a given X. It has been shown in Di Crescenzo and Longobardi [37] that the past entropies of X and Y are related by

$$H({}_tY) = \begin{cases} H\left({}_{\phi^{-1}(t)}X\right) + \mathbb{E}[\log \phi'(X) | X < \phi^{-1}(t)], & \text{if } \phi \text{ is strictly increasing,} \\ H\left(X_{\phi^{-1}(t)}\right) + \mathbb{E}[\log(-\phi'(X)) | X > \phi^{-1}(t)], & \text{if } \phi \text{ is strictly decreasing.} \end{cases} \tag{2.69}$$

Similar results can be proved for the past varentropy.

Proposition 2.21. *Let $Y = \phi(X)$, where ϕ is a differentiable and strictly monotone function. Then, if ϕ is strictly increasing, for the past varentropy of Y we have*

$$\begin{aligned} V_e({}_tY) &= V_e\left({}_{\phi^{-1}(t)}X\right) - 2\mathbb{E}\left[\log \frac{f_X(X)}{F_X(\phi^{-1}(t))} \log \phi'(X) \middle| X < \phi^{-1}(t)\right] \\ &\quad + \mathrm{Var}[\log \phi'(X) | X < \phi^{-1}(t)] - 2H\left({}_{\phi^{-1}(t)}X\right) \mathbb{E}[\log \phi'(X) | X < \phi^{-1}(t)], \end{aligned} \tag{2.70}$$

whereas, if ϕ is strictly decreasing

$$\begin{aligned}
V_e({}_tY) &= V_e\left(X_{\phi^{-1}(t)}\right) - 2\mathbb{E}\left[\log \frac{f_X(X)}{\overline{F}_X(\phi^{-1}(t))} \log(-\phi'(X)) \middle| X > \phi^{-1}(t)\right] \\
&+ \mathrm{Var}[\log(-\phi'(X))|X > \phi^{-1}(t)] - 2H\left(X_{\phi^{-1}(t)}\right)\mathbb{E}[\log(-\phi'(X))|X > \phi^{-1}(t)].
\end{aligned} \tag{2.71}$$

Proof. Suppose that ϕ is strictly increasing. From $Y = \phi(X)$ we know that $F_Y(x) = F_X\left(\phi^{-1}(x)\right)$ and $f_Y(x) = \frac{f_X(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))}$. Hence, by using (2.54) and (2.69), we get

$$\begin{aligned}
V_e({}_tY) &= \int_0^{\phi^{-1}(t)} \frac{f_X(x)}{F_X\left(\phi^{-1}(t)\right)} \left(\log \frac{f_X\left(x\right)}{F_X\left(\phi^{-1}(t)\right)} - \log \phi'(x)\right)^2 dx \\
&\quad - \left[H\left({}_{\phi^{-1}(t)}X\right) + \mathbb{E}[\log \phi'(X)|X < \phi^{-1}(t)]\right]^2.
\end{aligned}$$

Then, the result follows by developing the two squares in the above equality and observing that

$$\begin{aligned}
\int_0^{\phi^{-1}(t)} \frac{f_X(x)}{F_X\left(\phi^{-1}(t)\right)} \left(\log \phi'(x)\right)^2 dx &- \mathbb{E}^2[\log \phi'(X)|X < \phi^{-1}(t)] \\
&= \mathrm{Var}[\log \phi'(X)|X < \phi^{-1}(t)].
\end{aligned}$$

The proof is similar if ϕ is strictly decreasing and hence it is omitted. □

Example 2.15. The Inverted Exponential distribution (*invExp*), introduced as a lifetime model in Lin et al. [70], has been considered by many authors in reliability studies (see, for instance, Krishna and Kumar [65] and references therein). The past varentropy of an inverted exponential distribution can be obtained by using Proposition 2.21. Consider $X \sim Exp(\lambda)$ and $Y = \phi(X) = 1/X$ so that ϕ is strictly decreasing and $Y \sim invExp(\lambda)$. The past varentropy of Y can be evaluated by using (2.71)

$$\begin{aligned}
V_e({}_tY) &= V_e\left(X_{1/t}\right) - 2\mathbb{E}\left[\log \frac{\lambda e^{-\lambda X}}{e^{-\lambda/t}} \log\left(\frac{1}{X^2}\right) \middle| X > \frac{1}{t}\right] \\
&+ \mathrm{Var}\left[\log\left(\frac{1}{X^2}\right) \middle| X > \frac{1}{t}\right] - 2H\left(X_{1/t}\right)\mathbb{E}\left[\log\left(\frac{1}{X^2}\right) \middle| X > \frac{1}{t}\right].
\end{aligned}$$

The residual entropy and the residual varentropy for the exponential distribution are

$$H(X_t) = 1 - \log \lambda, \quad V_e(X_t) = 1,$$

and then the past varentropy of Y is expressed as

$$\begin{aligned}
V_e({}_tY) &= 1 - 2\mathbb{E}\left[\log \frac{\lambda e^{-\lambda X}}{e^{-\lambda/t}} \log\left(\frac{1}{X^2}\right) \middle| X > \frac{1}{t}\right] \\
&+ \mathrm{Var}\left[\log\left(\frac{1}{X^2}\right) \middle| X > \frac{1}{t}\right] - 2(1 - \log \lambda)\mathbb{E}\left[\log\left(\frac{1}{X^2}\right) \middle| X > \frac{1}{t}\right].
\end{aligned}$$

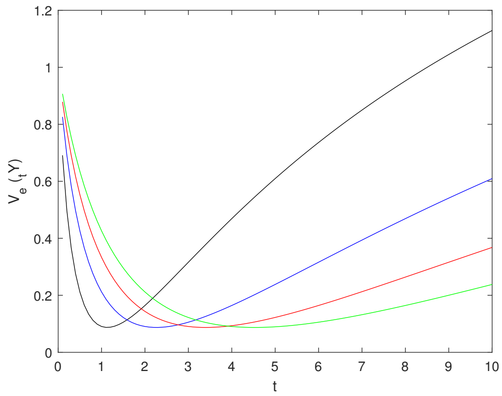

Figure 2.10: Plots of past varentropies of inverse exponential distributions with parameter $\lambda = 1, 2, 3, 4$ (black, blue, red and green, respectively).

With several calculations, the above expression reduces to

$$\begin{aligned} V_e(_tY) &= -3 + \frac{4\lambda}{t}\log\frac{1}{t^2} + \frac{8t}{\lambda} + \left(8 - \frac{4\lambda}{t}\right)\frac{1}{e^{-\lambda/t}}Ei\left(-\frac{\lambda}{t}\right) - \frac{4}{e^{-2\lambda/t}}Ei^2\left(-\frac{\lambda}{t}\right) \\ &\quad + \frac{4}{e^{-\lambda/t}}\log\frac{1}{t^2}Ei\left(-\frac{\lambda}{t}\right) - \frac{4}{\lambda e^{-\lambda/t}}\int_{1/t}^{+\infty}\frac{\log x^2}{x^2}e^{-\lambda x}dx, \end{aligned}$$

where $Ei(\cdot)$ is the exponential integral function (see Gautschi and Gahill [48]). The plot of this past varentropy is shown in Figure 2.10 for different choices of λ.

In the following, we give some bounds for the past varentropy. A very simple upper bound can be provided for a large class of distributions, as stated in the following proposition.

Proposition 2.22. *Let X be a non-negative random variable with support $\mathcal{S}$ and log-concave pdf f. Then*

$$V_e(_tX) \le 1 \qquad \textit{for all} \;\; t \in \mathcal{S}.$$

Proof. Observe that if $f(x)$ is log-concave, then also $f_{_tX}(x) = \frac{f(x)}{F(t)}$ is log-concave. From Theorem 2.3 of Fradelizi et al. [47], we know that if X has a log-concave pdf, then $V_e(X) \le 1$ and the proof is completed. □

For example, the pdf $f(x) = 6x(1-x)$, $x \in [0,1]$ of $X \sim Beta(2,2)$ is logconcave, so that the past varentropy of X is always smaller than 1, as confirmed by its plot shown in Figure 2.8, right. However, by comparing this bound with the plot of $V_e(_tX)$, one can immediately observe that it is a really large bound. Better upper bounds can be provided, for any X, by using results available in the literature. Recall that, for a random lifetime X, the inactivity time at t is defined as $X_{[t]} = (t - X|X \le t) = t - {_tX}$. The following upper bound for $\mathrm{Var}[-\log f_{X_{[t]}}(X_{[t]})]$ has been proved in Goodarzi et al. [50], Proposition 1, making use of an upper bound for variances proved in Cacoullos and Papathanasiou [29]:

$$\mathrm{Var}[-\log f_{X_{[t]}}(X_{[t]})] \leq \mathbb{E}\left[\frac{\eta^2(t-X_{[t]})}{q(t-X_{[t]})}\left(m_X(t-X_{[t]})-m_X(t)+X_{[t]}\right)\right] \tag{2.72}$$

for all $t \in \mathcal{S}$, where $\eta(x) = -f'(x)/f(x)$ is the eta function and $m_X(x) = x - \tilde{\mu}_X(x)$ is the mean inactivity time given in (1.14), and where $\tilde{\mu}_X(x)$ is defined in (1.13). Now observe that

$$\begin{aligned}
V_e(X_{[t]}) &= \mathrm{Var}[-\log f_{X_{[t]}}(X_{[t]})] = \mathbb{E}[\log^2 f_{X_{[t]}}(X_{[t]})] - \left[H(X_{[t]}))\right]^2 \\
&= \int_0^t \frac{f(t-x)}{F(t)}\log^2\left(\frac{f(t-x)}{F(t)}\right)dx - \left[\int_0^t \frac{f(t-x)}{F(t)}\log\left(\frac{f(t-x)}{F(t)}\right)dx\right]^2 \\
&= \int_0^t \frac{f(x)}{F(t)}\log^2\left(\frac{f(x)}{F(t)}\right)dx - \left[\int_0^t \frac{f(x)}{F(t)}\log\left(\frac{f(x)}{F(t)}\right)dx\right]^2 \\
&= \mathrm{Var}[-\log f_{{}_tX}({}_tX)] \\
&= V_e({}_tX).
\end{aligned}$$

Thus, by recalling that $m_X(x) = x - \tilde{\mu}_X(x)$ and $X_{[t]} = t - {}_tX$, from (2.72) one gets the upper bound

$$V_e({}_tX) \leq \mathbb{E}\left[\frac{\eta^2({}_tX)}{q({}_tX)}\left(\tilde{\mu}(t)-\tilde{\mu}({}_tX)\right)\right] \quad \forall\, t \in \mathcal{S}.$$

A lower bound for the past varentropy can also be proved. In order to do this, we recall the definition of the variance past lifetime function $\tilde{\nu}_X^2$

$$\tilde{\nu}_X^2(t) = \mathrm{Var}({}_tX) = \mathrm{Var}(X|X \leq t) = \frac{1}{F(t)}\int_0^t x^2 f(x)dx - (\tilde{\mu}_X(t))^2, \quad t \in \mathcal{S}.$$

Note that, for every $t \in \mathcal{S}$ the variance past lifetime function $\tilde{\nu}_X^2(t)$ is the same as the variance of the inactivity time $X_{[t]}$, see Kandil et al. [59] for details and properties of the variance of the inactivity time function.

Proposition 2.23. *Let ${}_tX$ be the past lifetime of X at time t, and let the mean past lifetime $\tilde{\mu}_X(t)$ and the variance past lifetime $\tilde{\nu}_X^2(t)$ be finite for all $t \in \mathcal{S}$. Then*

$$V_e({}_tX) \geq \tilde{\nu}_X^2(t)\left[\mathbb{E}(\omega_t'({}_tX))\right]^2,$$

where the function $\omega_t(x)$ is defined by solving the equation

$$\tilde{\nu}_X^2(t)\omega_t(x)f_{{}_tX}(x) = \int_0^x (\tilde{\mu}_X(z)-z)f_{{}_tX}(z)dz, \quad x \in \mathcal{S}. \tag{2.73}$$

Proof. Recall that if X is a random variable with pdf f, mean μ_X and variance σ_X^2, then

$$\mathrm{Var}[g(X)] \geq \sigma^2\left[\mathbb{E}(\omega(X)g'(X))\right]^2, \tag{2.74}$$

where $\omega(x)$ is defined by $\sigma^2\omega(x)f(x) = \int_0^x(\mu - z)f(z)dz$ (see Cacoullos and Papathanasiou [30]). Hence, in (2.74) by choosing $g(x) = -\log f_{tX}(x)$ and ${}_tX$ as X, it follows

$$\operatorname{Var}(-\log f_{tX}({}_tX)) \geq \tilde{\nu}_X^2(t)\left[\mathbb{E}\left(\omega_t({}_tX)\frac{f'_{tX}({}_tX)}{f_{tX}({}_tX)}\right)\right]^2. \tag{2.75}$$

By differentiating both sides of (2.73), we have

$$\omega(x)\frac{f'_{tX}(x)}{f_{tX}(x)} = \frac{\tilde{\mu}_X(x) - x}{\tilde{\nu}_X^2(t)} - \omega'_t(x),$$

and then, from (2.75),

$$\begin{aligned} V({}_tX) &= \operatorname{Var}(-\log f_{tX}({}_tX)) \\ &\geq \tilde{\nu}_X^2(t)\left[\mathbb{E}\left(\frac{\tilde{\mu}_X(t) - {}_tX}{\tilde{\nu}_X^2(t)} - \omega'({}_tX)\right)\right]^2 = \tilde{\nu}_X^2(t)\left[\mathbb{E}(\omega'_t({}_tX))\right]^2. \end{aligned}$$

□

2.6.1 Past varentropy and parallel systems

When the past varentropy $V_e({}_tX)$ of a random lifetime X is available, in some cases it is possible to easily compute the past varentropy of another lifetime Y whose distribution is a transformation of that one of X. An example is given by the scale model: the family of random variables $\{X^{(a)} : a > 0\}$ follows a Scale model if there exists a non-negative random variable X with cdf F and pdf f such that $X^{(a)}$ has distribution $F^{(a)}(t) = F(at)$ for all t, where $a > 0$ is the parameter of the model. Some examples are the exponential, Weibull (with a fixed shape parameter) and Pareto (with a fixed shape parameter) distributions. Hence, from Proposition 2.20 it immediately follows

$$V_e({}_tX^{(a)}) = V_e\left({}_{at}X\right), \qquad \forall t.$$

A more interesting case is when the family of random variables $\{X^{(a)} : a > 0\}$ follows the Proportional Reversed Hazard Rate model introduced in Chapter 1. Then, there exists a non-negative random variable X with cdf F and pdf f such that

$$F^{(a)}(t) = \mathbb{P}(X^{(a)} \leq t) = [F(t)]^a, \qquad f^{(a)}(t) = a[F(t)]^{(a-1)}f(t), \quad t \in \mathcal{S}, \tag{2.76}$$

being $F^{(a)}$ and $f^{(a)}$ the cdf and the pdf of $X^{(a)}$, respectively. The relation between the reversed hazard rates of $X^{(a)}$ and X has already been given in (1.7). Moreover, we note that the cumulative reversed hazard rate function of $X^{(a)}$ is expressed as

$$Q_{X^{(a)}}(t) = -\log F^{(a)}(t) = aQ(t).$$

The proportional reversed hazard rate model finds applications, for example, in analysis of parallel systems. In fact, if we have a system composed by n units in parallel and characterized by IID lifetimes $X_1, \ldots, X_n$ with cdf F, then the lifetime of the system is given by $X^{(n)} = \max\{X_1, \ldots, X_n\}$. Hence $F_{X^{(n)}}(t) = [F(t)]^n$, so that the system satisfies the PRHR model (2.76) with $a = n$. In the following examples, it is highlighted the behavior of the past varentropy when it refers to the lifetime of a parallel system with IID components. First, we evaluate the past entropy of $X^{(a)}$ and the past varentropy of $X^{(a)}$, for an arbitrary $a > 0$,

$$H\left({}_tX^{(a)}\right) = -Q_{X^{(a)}}(t) - \frac{1}{[F(t)]^a}\int_0^t f^{(a)}(x)\log f^{(a)}(x)dx = -aQ(t) - \frac{1}{[F(t)]^a}\int_0^{[F(t)]^a} \gamma(y;a)dy$$

with the change of variable $y = [F(x)]^a$, and where $\gamma(y;a) = \log\left[ay^{1-1/a}f(F^{-1}(y^{1/a}))\right]$. Hence, the past varentropy of $X^{(a)}$ is given as

$$\begin{aligned} V_e\left({}_tX^{(a)}\right) &= \frac{1}{[F(t)]^a}\int_0^t f^{(a)}(x)(\log f^{(a)}(x))^2dx - \left[\frac{1}{[F(t)]^a}\int_0^t f^{(a)}(x)\log f^{(a)}(x)dx\right]^2 \\ &= \frac{1}{[F(t)]^a}\int_0^{[F(t)]^a}[\gamma(y;a)]^2dy - \left[\frac{1}{[F(t)]^a}\int_0^{[F(t)]^a}\gamma(y;a)dy\right]^2. \end{aligned}$$

Example 2.16. Let us consider the case in which X has a Pareto type II distribution with cdf $F(t) = t/(1+t)$ and pdf $f(t) = 1/(1+t)^2$, for $t \geq 0$. Then, $\gamma(y;a) = \log\left[ay^{1-1/a}(1-y^{1/a})^2\right]$, so that

$$\begin{aligned} H\left({}_tX^{(a)}\right) &= a\log\left(\frac{t}{1+t}\right) - \frac{1}{[t/(1+t)]^a}\int_0^{[t/(1+t)]^a}\gamma(y;a)dy, \\ V_e\left({}_tX^{(a)}\right) &= \frac{1}{[t/(1+t)]^a}\int_0^{[t/(1+t)]^a}[\gamma(y;a)]^2dy - \left[\frac{1}{[t/(1+t)]^a}\int_0^{[t/(1+t)]^a}\gamma(y;a)dy\right]^2. \end{aligned}$$

If a is an integer, i.e., $X^{(a)}$ represents the lifetime of a parallel system of a number a of IID components, we obtain the past entropies and past varentropies shown in Figure 2.11 (for different integer values of a). It is interesting to observe that both the past entropies and the past varentropies intersect each other for different values of a: for small values of the time t one has the smaller past entropies and larger past varentropies when the number of components in parallel is large, and vice versa for large values of the time t. It means, for example, that in the long run (for large values of the inspection time t) the uncertainty of the information content of the past lifetime of a parallel system reduces as the number of components in the system increases (and vice versa for small t).

The same can be observed if X has an exponential distribution with parameter λ. In this case, $\gamma(y;a) = \log\left[\lambda ay^{1-1/a}(1-y^{1/a})\right]$, so that

$$\begin{aligned} H\left({}_tX^{(a)}\right) &= a\log\left(1-e^{-\lambda t}\right) - \frac{1}{[1-e^{-\lambda t}]^a}\int_0^{[1-e^{-\lambda t}]^a}\gamma(y;a)dy, \\ V_e\left({}_tX^{(a)}\right) &= \frac{1}{[1-e^{-\lambda t}]^a}\int_0^{[1-e^{-\lambda t}]^a}[\gamma(y;a)]^2dy - \left[\frac{1}{[1-e^{-\lambda t}]^a}\int_0^{[1-e^{-\lambda t}]^a}\gamma(y;a)dy\right]^2. \end{aligned}$$

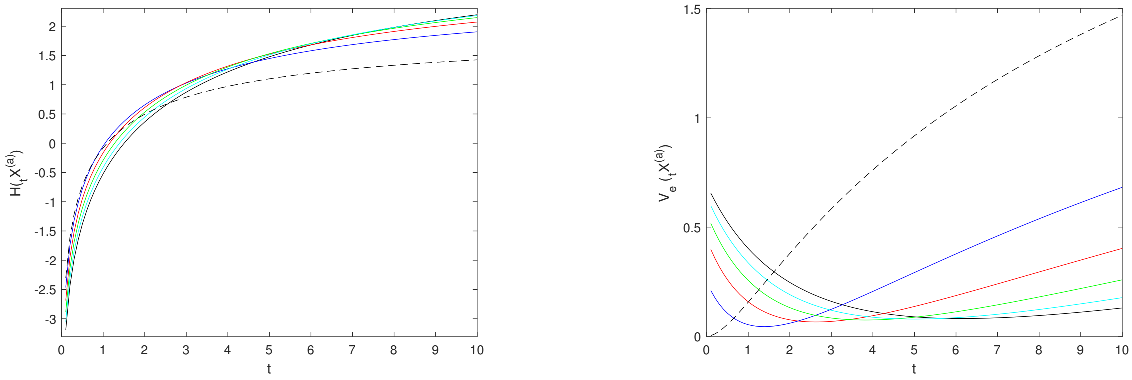

Figure 2.11: Plots of the past entropy (left) and the past varentropy (right) of Pareto type II PRHR model for $a = 1$ (dashed line) and $a = 2, 3, 4, 5, 6$ (blue, red, green, cyan and black, respectively).

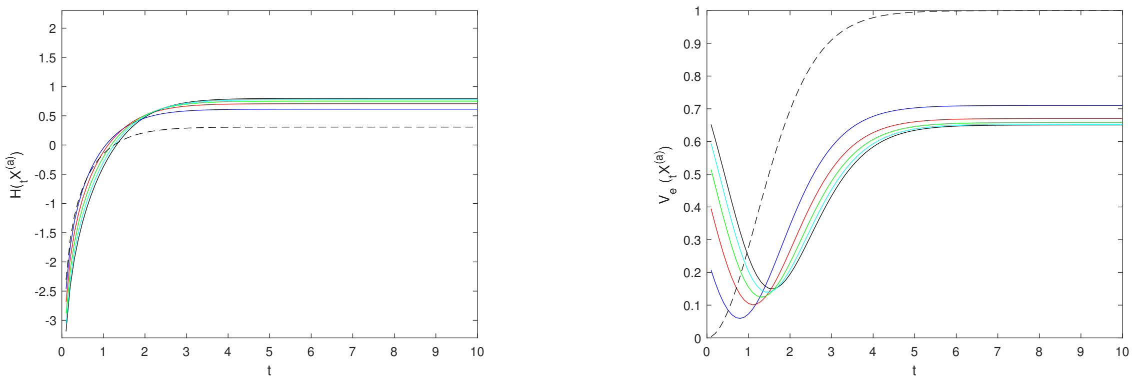

Figure 2.12: Plots of the past entropy (left) and the past varentropy (right) of exponential PRHR model for $a = 1$ (dashed line) and $a = 2, 3, 4, 5, 6$ (blue, red, green, cyan and black, respectively).

The plots of $H\left({}_tX^{(a)}\right)$ and $V_e\left({}_tX^{(a)}\right)$, for different integer values of a and with $\lambda = 2$, are shown in Figure 2.12. As for the Pareto type II case, both the past entropies and the past varentropies intersect each other for different values of a, having a similar behavior.

There exists a family for which the past varentropies do not intersect, which is the family discussed in Proposition 2.19, whose varentropies are constant. Let X_α be a lifetime having support $\mathcal{S} = [0, 1]$ and cdf $F_\alpha(x) = x^\alpha$, for $x \in \mathcal{S}$. Then, the corresponding parallel system with n IID components has distribution $F_{n\alpha}(x) = x^{n\alpha}$ for $x \in \mathcal{S}$, which is still in the family of distribution having constant past varentropy. Thus, $V_e({}_tX_{n\alpha}) = (1 - 1/(n\alpha))^2$ for all $t \in \mathcal{S}$, and obviously these past varentropies do not intersect as n varies in $\mathbb{N}^+$. This is another interesting property of such a family of distributions.

Chapter 3

New measures of uncertainty in Dempster-Shafer theory of evidence

Dempster [34] and Shafer [104] introduced a theory to study uncertainty. Their theory of evidence is a generalization of the classical probability theory. In Dempster-Shafer theory (DST) of evidence, an uncertain event with a finite number of alternatives is considered, and a mass function over the power set of the alternatives (i.e., a degree of confidence to all of its subsets) is defined. If we give positive mass only to singletons, a discrete probability distribution is obtained. Through DST it is possible to describe situations in which there is less specific information. In this chapter, some concepts and definitions of DST are given. Then, based on Balakrishnan, Buono and Longobardi [5], Buono and Longobardi [25] and Kazemi, Tahmasebi, Buono and Longobardi [62], new measures of uncertainty in DST will be studied and applications to classifications problems will be given.

3.1 Dempster-Shafer theory of evidence

We start this section by describing a simple example given in [35] to explain how DST extends the classical probability theory. Consider two boxes, A and B, such that in A there are only red balls and in B there are only green balls. The number of balls in each box is unknown. A ball is picked randomly from one of these two boxes. The box A is selected with probability $p_A = 0.6$ and the box B is selected with probability $p_B = 1 - p_A = 0.4$. Hence, the probability of picking up a red ball is 0.6, $\mathbb{P}(R) = 0.6$, whereas the probability of picking up a green ball is 0.4, $\mathbb{P}(G) = 0.4$. Suppose now that in box B there are green and red balls with rates unknown. The box A is still selected with probability $p_A = 0.6$ and the box B with probability $p_B = 0.4$. In this case, we cannot obtain the probability of picking up a red ball. To analyze this problem, we can use DST to express the uncertainty. In particular, we choose a mass function m such that, $m(R) = 0.6$ and $m(R, G) = 0.4$. In the following, we recall some of the basic notions in

DST.

Let X be a frame of discernment (FOD), i.e., a set of mutually exclusive and collectively exhaustive events indicated by $X = \{\theta_1, \theta_2, \ldots, \theta_{|X|}\}$. The power set of X is indicated by 2^X and it has cardinality $2^{|X|}$. A function $m : 2^X \to [0,1]$ is called a mass function or a basic probability assignment (BPA) if

$$m(\emptyset) = 0 \ \text{ and } \ \sum_{A \in 2^X} m(A) = 1.$$

If $m(A) \neq 0$ implies $|A| = 1$ then m is also a probability mass function, i.e., BPAs generalize discrete random variables. Moreover, the elements A such that $m(A) > 0$, are known as focal elements.

In DST, there are different indices to evaluate the degree of belief in a subset of the FOD. Among them, here we recall the definitions of belief function, plausibility function and pignistic probability transformation (PPT). The belief function and the plausibility function are, respectively, defined as

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B).$$

Note that the plausibility of A can be expressed also as one minus the sum of the masses of all sets whose intersection with A is empty. Moreover, both the belief and the plausibility vary from zero to one and the belief is always less than or equal to the plausibility. Given a BPA, we can evaluate for each focal element the pignistic probability transformation (PPT) which represents a point estimate of belief and can be determined as [109]

$$PPT(A) = \sum_{B : A \subseteq B} \frac{m(B)}{|B|}. \tag{3.1}$$

If we have a weight or a reliability of an evidence, represented by a coefficient $\alpha \in [0,1]$, we can use it to generate a new BPA m^α in the following way (see [104])

$$m^\alpha(A) = \begin{cases} \alpha \, m(A), & \text{if } A \subset X \\ \alpha \, m(X) + (1 - \alpha), & \text{if } A = X. \end{cases} \tag{3.2}$$

Further, if we have two BPAs m_1, m_2 for a frame of discernment X, we can introduce a new BPA m^* for X by using the Dempster rule of combination, see [34]. We define $m^*(A)$, $A \subseteq X$, in the following way

$$m^*(A) = \begin{cases} 0, & \text{if } A = \emptyset \\ \frac{\sum_{B,C \subseteq X : B \cap C = A} m_1(B) m_2(C)}{1-K}, & \text{if } A \neq \emptyset \end{cases} \tag{3.3}$$

where $K = \sum_{B,C \subseteq X : B \cap C = \emptyset} m_1(B) m_2(C)$. Note that if $K > 1$, we cannot apply the Dempster rule of combination.

Recently, several measures of discrimination and uncertainty have been proposed in the literature and in the context of the Dempster-Shafer evidence theory. Among them, one of the most important is known as Deng entropy. The Deng entropy was introduced in [35] for a BPA m as

$$E_d(m) = -\sum_{A\subseteq X: m(A)>0} m(A)\log_2\left(\frac{m(A)}{2^{|A|}-1}\right). \tag{3.4}$$

This entropy is similar to Shannon entropy and they coincide if the BPA is also a probability mass function. The term $2^{|A|}-1$ represents the potential number of states in A. For a fixed value of $m(A)$, as the cardinality of A increases, $2^{|A|}-1$ increases and then also Deng entropy does.

In the literature, several properties of Deng entropy have been studied and other measures of uncertainty based on Deng entropy have been introduced. Other relevant measures of uncertainty and information known in the Dempster-Shafer theory of evidence are, for example, Hohle's confusion measure, Yager's dissonance measure and Klir and Ramer's discord measure. For a detailed review on the measures of uncertainty defined in DST, one may refer to Deng [36]. Here, as it will be useful in the sequel, we recall the definition of the Tsallis-Deng entropy, which was defined by Liu et al. [71] as

$$SD_\alpha(m) = \frac{1}{\alpha-1}\sum_{A\subseteq X: m(A)>0} m(A_i)\left[1-\left(\frac{m(A_i)}{2^{|A_i|}-1}\right)^{\alpha-1}\right], \tag{3.5}$$

where $\alpha > 0$ and $\alpha \neq 1$. Note that Tsallis-Deng entropy reduces to Deng entropy when α goes to 1.

3.2 Deng extropy

In this section, we analyze some properties of the dual measure of Deng entropy, namely Deng extropy, which was introduced and studied in Buono and Longobardi [25]. The definition was given in order to preserve the property about the sum of the entropy and the extropy in (2.7). More precisely, for a BPA m over a FOD X, the Deng extropy is defined by

$$EX_d(m) = -\sum_{A\subset X: m(A)>0} (1-m(A))\log_2\left(\frac{1-m(A)}{2^{|A^c|}-1}\right), \tag{3.6}$$

where A^c is the complementary set of A in X and $|A^c| = |X|-|A|$. In the following proposition, in analogy with the property given in (2.7), we study the sum of Deng entropy and Deng extropy.

Proposition 3.1. *Let m be a BPA for a frame of discernment X. Then*

$$E_d(m) + EX_d(m) = \sum_{A \subset X : m(A) > 0} E_d(m_A^*) - m(X) \log_2 \left(\frac{m(X)}{2^{|X|} - 1} \right) \tag{3.7}$$

$$= \sum_{A \subset X : m(A) > 0} EX_d(m_A^*) - m(X) \log_2 \left(\frac{m(X)}{2^{|X|} - 1} \right), \tag{3.8}$$

with the convention $0 \log 0 = 0$, where m_A^* is a BPA on X defined as

$$m_A^*(B) = \begin{cases} m(A), & \textit{if } B = A \\ 1 - m(A), & \textit{if } B = A^c \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. By using the definition of the BPA m_A^*, it follows

$$E_d(m_A^*) = -m(A) \log_2 \frac{m(A)}{2^{|A|} - 1} - (1 - m(A)) \log_2 \frac{1 - m(A)}{2^{|A^c|} - 1}$$

$$EX_d(m_A^*) = -(1 - m(A)) \log_2 \frac{1 - m(A)}{2^{|A^c|} - 1} - m(A) \log_2 \frac{m(A)}{2^{|A|} - 1},$$

so that they are equal. Hence, for every $A \subset X$ such that $m(A) > 0$, $E_d(m_A^*)$ (or $EX_d(m_A^*)$) gives the corresponding addend of $E_d(m) + EX_d(m)$. The only exception is given by X, which could give a contribution in the left hand side of Equation (3.7) if $m(X) > 0$, and for this reason there is an extra term in the right side of Equations (3.7) and (3.8). □

Now, some examples of evaluation of Deng extropy and the corresponding entropy are given.

Example 3.1. Given a frame of discernment X, $a \in X$ and a BPA m such that $m(\{a\}) = m(a) = 1$, we have

$$EX_d(m) = -(1-1) \log_2 \frac{1-1}{2^{|X|-1} - 1} = 0,$$

$$E_d(m) = -\log_2 1 = 0.$$

In this case, Deng entropy coincides with Deng extropy and they are both equal to 0.

Example 3.2. Given a frame of discernment $X = \{a, b, c\}$ and a BPA m such that $m(a) = m(b) = m(c) = \frac{1}{3}$, we have

$$EX_d(m) = -\frac{2}{3} \log_2 \left(\frac{2}{9} \right) \cdot 3 = -2 \log_2 \frac{2}{9},$$

$$E_d(m) = -\frac{1}{3} \log_2 \left(\frac{1}{3} \right) \cdot 3 = -\log_2 \frac{1}{3}.$$

Example 3.3. Given a frame of discernment X with cardinality n and a BPA m such that $m(i) = \frac{1}{n}$, for $i =\ 1, \dots, n$, we have

$$EX_d(m) = -n\left(1 - \frac{1}{n}\right)\log_2\left(\frac{1-\frac{1}{n}}{2^{n-1}-1}\right) = (n-1)\left[\log_2\left(\frac{n}{n-1}\right) + \log_2(2^{n-1}-1)\right],$$

$$E_d(m) = \log_2(n),$$

which are increasing in $n \in \mathbb{N}$.

Example 3.4. Consider a frame of discernment $X = \{1, 2, \dots, 15\}$ and a BPA m such that $m(3, 4, 5) = 0.05$, $m(6) = 0.05$, m(A)=0.8, $m(X) = 0.1$. The values for the Deng extropy and entropy are obtained in Table 3.1 for different choices of A.

Table 3.1: The values of the Deng extropy and the Deng entropy as A changes.

A	Deng Extropy	Deng Entropy
$\{1\}$	28.104	2.6623
$\{1, 2\}$	27.904	3.9303
$\{1, 2, 3\}$	27.704	4.9082
$\{1, \dots, 4\}$	27.504	5.7878
$\{1, \dots, 5\}$	27.304	6.6256
$\{1, \dots, 6\}$	27.104	7.4441
$\{1, \dots, 7\}$	26.903	8.2532
$\{1, \dots, 8\}$	26.702	9.0578
$\{1, \dots, 9\}$	26.500	9.8600
$\{1, \dots, 10\}$	26.295	10.661
$\{1, \dots, 11\}$	26.086	11.462
$\{1, \dots, 12\}$	25.866	12.262
$\{1, \dots, 13\}$	25.621	13.062
$\{1, \dots, 14\}$	25.304	13.862

Kang and Deng [60] studied the problem of the maximum Deng entropy. They find out that the maximum Deng entropy on a frame of discernment X with cardinality $|X|$ is attained if and only if the BPA m is defined as

$$m(A_i) = \frac{2^{|A_i|}-1}{\sum_{j=1}^{2^{|X|}-1}(2^{|A_j|}-1)}, \quad i = 1, \dots, 2^{|X|}-1,$$

where A_i, $i = 1, \dots, 2^{|X|} - 1$, are all non-empty elements of 2^X. Hence, the value of the maximum Deng entropy is given by

$$E_d^* = -\sum_{i=1}^{2^{|X|}-1} \frac{2^{|A_i|}-1}{\sum_{j=1}^{2^{|X|}-1}(2^{|A_j|}-1)} \log_2 \frac{\frac{2^{|A_i|}-1}{\sum_{j=1}^{2^{|X|}-1}(2^{|A_j|}-1)}}{2^{|A_i|}-1} = \log_2 \sum_{i=1}^{2^{|X|}-1}(2^{|A_i|}-1).$$

In analogy with this result, about the Deng extropy we do not obtain the maximum value but an upper bound. In the following theorem, we provide conditions to obtain an upper bound for Deng extropy with a fixed number of focal elements and a fixed value for $m(X)$.

Theorem 3.1. *Let m be a BPA for a frame of discernment X. An upper bound for Deng extropy with fixed values of $m(X)$ and N number of focal elements different from X, $N = |\mathcal{N}| = |\{A \subset X : m(A) > 0\}|$, is assumed in correspondence of the fictitious BPA m (in the sense that some values of $m(A)$ may be negative) such that*

$$m(A) = 1 - \frac{N - (1 - m(X))}{\sum_{B \in \mathcal{N}}(2^{|B^c|}-1)}(2^{|A^c|}-1), \quad A \in \mathcal{N}.$$

The value of the upper bound is

$$EX_d^* = -[N - (1 - m(X))] \log_2 \frac{N - (1 - m(X))}{\sum_{A \in \mathcal{N}}(2^{|A^c|}-1)}.$$

Proof. Suppose $m(X) = 0$. We will prove that in this case that the upper bound is

$$EX_d^* = -(N-1) \log_2 \frac{N-1}{\sum_{A \in \mathcal{N}}(2^{|A^c|}-1)},$$

and that it is attained in correspondence of the fictitious BPA defined by

$$m(A) = 1 - \frac{N-1}{\sum_{B \in \mathcal{N}}(2^{|B^c|}-1)}(2^{|A^c|}-1), \quad A \in \mathcal{N}. \tag{3.9}$$

We want to maximize

$$EX_d = -\sum_{A \in \mathcal{N}}(1 - m(A)) \log_2 \frac{1 - m(A)}{2^{|A^c|}-1}$$

subject to the condition

$$\sum_{A \in \mathcal{N}} m(A) = 1.$$

Note that we are not requiring any assumption about the sign of $m(A)$. Then, the Lagrange function can be defined as

$$(EX_d)_0 = -\sum_{A \in \mathcal{N}}(1 - m(A)) \log_2 \frac{1 - m(A)}{2^{|A^c|}-1} + \lambda \left(\sum_{A \in \mathcal{N}} m(A) - 1 \right).$$

Thus the gradient can be computed, and for $A \in \mathcal{N}$ we have

$$\frac{\partial (EX_d)_0}{\partial m(A)} = \log_2 \frac{1 - m(A)}{2^{|A^c|}-1} + \log_2 e + \lambda = 0,$$

where $\log_2 e + \lambda$ does not depend on $m(A)$. By vanishing all the partial derivatives, it follows

$$\frac{1 - m(A)}{2^{|A^c|} - 1} = K, \quad A \in \mathcal{N},$$

where K is a constant and so

$$m(A) = 1 - K\left(2^{|A^c|} - 1\right), \quad A \in \mathcal{N}. \tag{3.10}$$

By summing over $A \in \mathcal{N}$, we get

$$1 = \sum_{A \in \mathcal{N}} \left(1 - K\left(2^{|A^c|} - 1\right)\right),$$

and then

$$K = \frac{N - 1}{\sum_{A \in \mathcal{N}} \left(2^{|A^c|} - 1\right)}.$$

Therefore, by Equation (3.10) we deduce

$$m(A) = 1 - \frac{N - 1}{\sum_{B \in \mathcal{N}} \left(2^{|B^c|} - 1\right)} \left(2^{|A^c|} - 1\right), \quad A \in \mathcal{N},$$

i.e., Equation (3.9). Finally, for the Deng extropy related to this fictitious BPA, we get

$$\begin{aligned} EX_d(m) &= -\sum_{A \in \mathcal{N}} \frac{N - 1}{\sum_{B \in \mathcal{N}} \left(2^{|B^c|} - 1\right)} \left(2^{|A^c|} - 1\right) \log_2 \frac{N - 1}{\sum_{B \in \mathcal{N}} \left(2^{|B^c|} - 1\right)} \\ &= -(N - 1) \log_2 \frac{N - 1}{\sum_{A \in \mathcal{N}} (2^{|A^c|} - 1)} = EX_d^* \end{aligned}$$

and the proof is completed. □

3.3 Fractional Deng entropy and extropy

In recent years, great attention has been given to fractional calculus. For this reason, several authors have studied various fractional entropies from the idea that they satisfy physical conditions of stability. In order to obtain an analog of (2.5) in the context of Dempster-Shafer theory of evidence, the concepts of fractional Deng entropy and fractional Deng extropy have been introduced and studied in Kazemi, Tahmasebi, Buono and Longobardi [62]. In this section, some of their properties will be pointed out.

Definition 3.1. Let m be a BPA on a FOD X. The Fractional Deng Entropy (FDEn) of m is defined as

$$E_d^q(m) = \sum_{A \subseteq X : m(A) > 0} m(A) \left[-\log_2 \left(\frac{m(A)}{2^{|A|} - 1}\right)\right]^q, \quad 0 < q \leq 1. \tag{3.11}$$

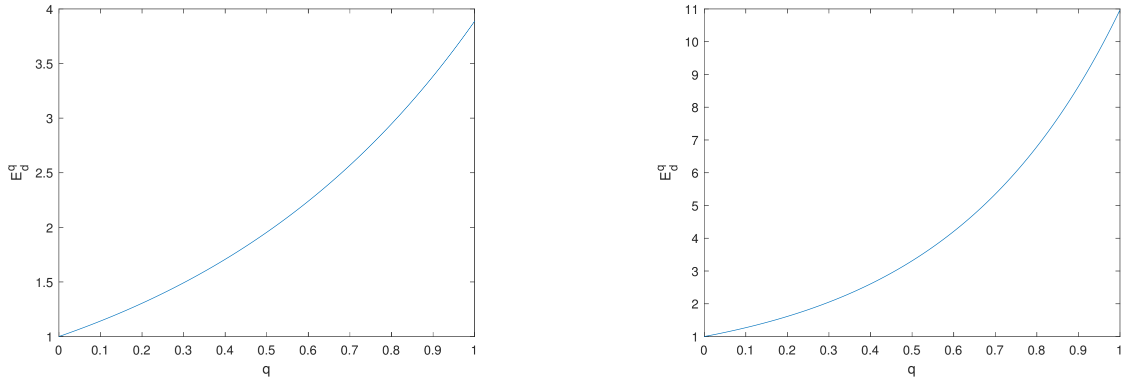

Figure 3.1: Plot of $E_d^q(m_2)$ in Example 3.5 as a function of q (left). Plot of $E_d^q(m)$ in Example 3.6 as a function of q (right).

Note that the fractional entropy and FDEn are identical when the BPA assigns a positive mass only to singletons. Moreover, we remark that if there exists $A \subseteq X$ such that $m(A) > 0$ and $|A| > 1$, we cannot evaluate the fractional entropy.

Example 3.5. Given a FOD $X = \{a, b, c\}$, for a mass function $m_1(a, b, c) = 1$, we have

$$E_d^q(m_1) = [\log_2 7]^q .$$

For another BPA $m_2(a) = m_2(b) = m_2(c) = m_2(a, b) = m_2(a, c) = m_2(b, c) = m_2(a, b, c) = \frac{1}{7}$, we obtain

$$E_d^q(m_2) = \frac{3}{7}\left([\log_2 7]^q + [\log_2 21]^q\right) + \frac{1}{7}[\log_2 49]^q. \tag{3.12}$$

The plot of the FDEn in (3.12) as a function of $q \in (0, 1]$ is given in Figure 3.1 (left). From Figure 3.1 (left), it is seen that $E_d^q(m)$ is increasing in q and the maximum is achieved for $q = 1$, i.e., when the FDEn reduces to Deng entropy.

Example 3.6. Assume that the FOD is $X = \{a_1, a_2, \ldots, a_{20}\}$. For a mass function $m(\{a_1, a_2, \ldots, a_{10}\}) = 0.4$, $m(\{a_{11}, a_{12}, \ldots, a_{20}\}) = 0.6$, we obtain

$$E_d^q(m) = 0.4\left[-\log_2\left(\frac{0.4}{2^{10}-1}\right)\right]^q + 0.6\left[-\log_2\left(\frac{0.6}{2^{10}-1}\right)\right]^q .$$

The plot of this FDEn is given in Figure 3.1 (right). From Figure 3.1 (right), it is seen that $E_d^q(m)$ is increasing in q, and the maximum is achieved when FDEn reduces to Deng entropy.

Example 3.7. Let us consider a FOD $X = \{a, b, c\}$ and a BPA m such that $m(a) = p_a$ and $m(a, b) = r_a$, where $r_a = 1 - p_a$. The FDEn is given by

$$E_d^q(m) = p_a\left[\log_2(1/p_a)\right]^q + r_a\left[\log_2(3/r_a)\right]^q .$$

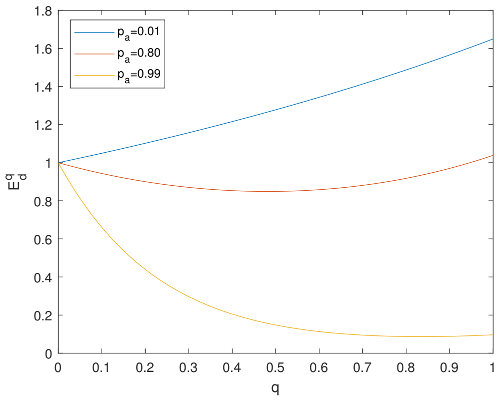

Figure 3.2: Plot of $E_d^q(m)$ in Example 3.7 as a function of q for different values of p_a.

In Figure 3.2, the plots of $E_d^q(m)$ for $p_a \in \{0.01, 0.8, 0.99\}$ are given. It is seen that for $p_a = 0.01$, $p_a = 0.80$ and $p_a = 0.99$, the plot of $E_d^q(m)$ is increasing, upside-down bathtub shaped and decreasing, respectively.

In the above examples, it is seen that the function $E_d^q(m)$ cannot be a concave function, and it can be increasing, decreasing and upside-down bathtub shape. Furthermore, the supremum FDEn is achieved when q is near to the boundary of interval $(0, 1]$. Therefore, we can state and prove the following theorem.

Theorem 3.2. *Let m be a non-degenerate BPA on a FOD X and $q \in (0, 1]$. Then, the supremum FDEn as a function of q is attained for $q \in \{0, 1\}$ and the infimum is attained in the extremes of interval $(0, 1)$, or it is a minimum assumed in a unique $q_0 \in (0, 1)$.*

Proof. By noting that for fixed $x > 0$ the function $g(p) = x^p$ is a convex function of p we can conclude that the FDEn is a strictly convex function of q. Hence, we have three possible scenarios. In the first one, the FDEn is strictly increasing in q and hence it assumes the maximum value for $q = 1$, i.e., when it reduces to Deng entropy, and the infimum is 1 by the normalization condition. In the second scenario, the FDEn is strictly decreasing; hence, the supremum is 1 and the minimum is assumed for $q = 1$. In the third case, there is a unique stationary point in $(0, 1)$, it is an absolute minimum, whereas the supremum is given by $\max\{1, E_d(m)\}$. $\square$

In the following theorem, we study the maximum FDEn for a fixed value of q. This is an important issue in the theory of measures of uncertainty, as remarked, for instance, in the previous section.

Theorem 3.3. *Let X be a FOD, $q \in (0, 1]$ and m be a BPA, which assigns positive mass to* *each non-empty subset of X. The maximum FDEn is attained if the BPA m is defined as*

$$m(A) = \frac{2^{|A|} - 1}{\sum_{B \subseteq X}(2^{|B|} - 1)}, \quad A \subseteq X. \tag{3.13}$$

Proof. For a fixed $q \in (0, 1]$ the FDEn is given by (3.11) as

$$E_d^q(m) = \sum_{\emptyset \neq A \subseteq X} m(A) \left[-\log_2\left(\frac{m(A)}{2^{|A|} - 1}\right)\right]^q. \tag{3.14}$$

We have to maximize (3.14) subject to the constraint (normalization condition)

$$\sum_{\emptyset \neq A \subseteq X} m(A) = 1. \tag{3.15}$$

In order to use the method of Lagrange multipliers, we have to compute the partial derivatives of the function

$$\tilde{E}_d^q = \sum_{\emptyset \neq A \subseteq X} m(A) \left[-\log_2\left(\frac{m(A)}{2^{|A|} - 1}\right)\right]^q + \lambda \left(\sum_{\emptyset \neq A \subseteq X} m(A) - 1\right)$$

with respect to $m(A)$. By differentiating $\tilde{E}_d^q$ with respect to $m(A)$, it follows

$$\begin{aligned} \frac{\partial \tilde{E}_d^q}{\partial m(A)} &= \left[-\log_2\left(\frac{m(A)}{2^{|A|} - 1}\right)\right]^q - q\log_2(e)\left[-\log_2\left(\frac{m(A)}{2^{|A|} - 1}\right)\right]^{q-1} + \lambda \\ &= \left[-\log_2\left(\frac{m(A)}{2^{|A|} - 1}\right)\right]^{q-1}\left[-q\log_2(e) - \log_2\left(\frac{m(A)}{2^{|A|} - 1}\right)\right] + \lambda. \end{aligned}$$

In order to vanish all the partial derivatives of $\tilde{E}_d^q$, the ratio $\frac{m(A)}{2^{|A|}-1} = K$ has to be invariant with respect to A. In fact, the function

$$g(z) = [-\log_2(z)]^{q-1}[-q\log_2(e) - \log_2(z)]$$

is strictly decreasing in $z \in (0, 1)$ since

$$g'(z) = \frac{q\log_2(e)}{z}[-\log_2(z)]^{q-2}[(q-1)\log_2(e) + \log_2(z)]$$

and $z < e^{1-q}$. Hence, by the constraint (3.15), we derive

$$K = \frac{1}{\sum_{B \subseteq X}(2^{|B|} - 1)}$$

and the BPA m, which maximizes the FDEn, is given in (3.13). □

Example 3.8. Based on the result of Theorem 3.3, let us evaluate the maximum FDEn for a FOD of cardinality 3, $X = \{a, b, c\}$. In this case, the BPA given in (3.13) is expressed as

$$\begin{aligned} &m(a) = m(b) = m(c) = \frac{1}{19}, \\ &m(a, b) = m(a, c) = m(b, c) = \frac{3}{19}, \quad m(X) = \frac{7}{19}. \end{aligned}$$

Then, the maximum FDEn is given by

$$E_d^q(m) = [\log_2(19)]^q.$$

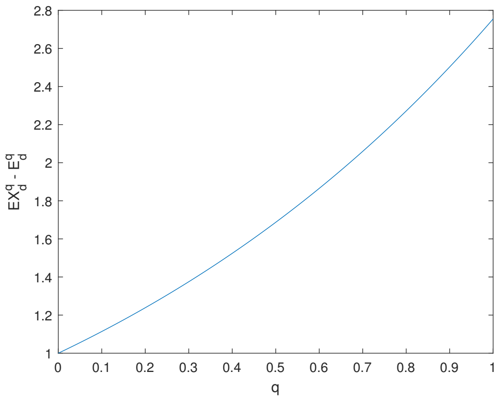

Figure 3.3: Plot of $EX_d^q(m) - E_d^q(m)$ in Example 3.9 as a function of q.

Now, in analogy with FDEn, we introduce the definition of the fractional version of Deng extropy (3.6) given in Kazemi, Tahmasebi, Buono and Longobardi [62].

Definition 3.2. Let m be a BPA on a FOD X. The Fractional Deng Extropy (FDEx) of m is defined as

$$EX_d^q(m) = \sum_{A \subset X : m(A) > 0} (1 - m(A)) \left[-\log_2 \left(\frac{1 - m(A)}{2^{|A^c|} - 1} \right) \right]^q, \quad 0 < q \leq 1. \tag{3.16}$$

Example 3.9. Assume that the FOD is $X = \{a, b, c\}$. For a mass function $m(a) = m(b) = m(c) = \frac{1}{3}$, the associated FDEn and FDEx are obtained as

$$E_d^q(m) = [\log_2 3]^q, \quad EX_d^q(m) = 2 \left[\log_2 \frac{9}{2} \right]^q.$$

In Figure 3.3, the plot of the difference $EX_d^q(m) - E_d^q(m)$ is given. Note that $EX_d^q(m) - E_d^q(m)$ is an increasing function of q, and this function is greater than 1. Thus, for $q \in (0, 1]$, the FDEx is greater than the FDEn. Furthermore, $EX_d^q(m)$ is increasing in q and the maximum is achieved for $q = 1$, i.e., when FDEx reduces to Deng extropy.

Example 3.10. Let us consider a FOD $X = \{a, b, c\}$. For a mass function $m(a) = m(b) = m(c) = m(a, b) = m(a, c) = m(b, c) = m(a, b, c) = \frac{1}{7}$, we obtain

$$EX_d^q(m) = \frac{18}{7} \left([\log_2 7 - 1]^q + [\log_2 7 - \log_2 6]^q \right).$$

In Figure 3.4 (left), the plot of $EX_d^q(m)$ is given. One can see that as a function of q, it has a convex parabolic shape and the maximum is achieved when it reduces to Deng extropy.

Example 3.11. Consider the FOD $X = \{a_1, a_2, \ldots, a_{20}\}$. For a mass function $m(\{a_1, a_2, \ldots, a_{10}\}) = 0.4$, $m(\{a_{10}, a_{11}, \ldots, a_{20}\}) = 0.6$, we obtain

$$EX_d^q(m) = 0.6 \left[-\log_2 \left(\frac{0.6}{2^{10} - 1} \right) \right]^q + 0.4 \left[-\log_2 \left(\frac{0.4}{2^{10} - 1} \right) \right]^q.$$

In this case, FDEx and FDEn are equal, see Example 3.6.

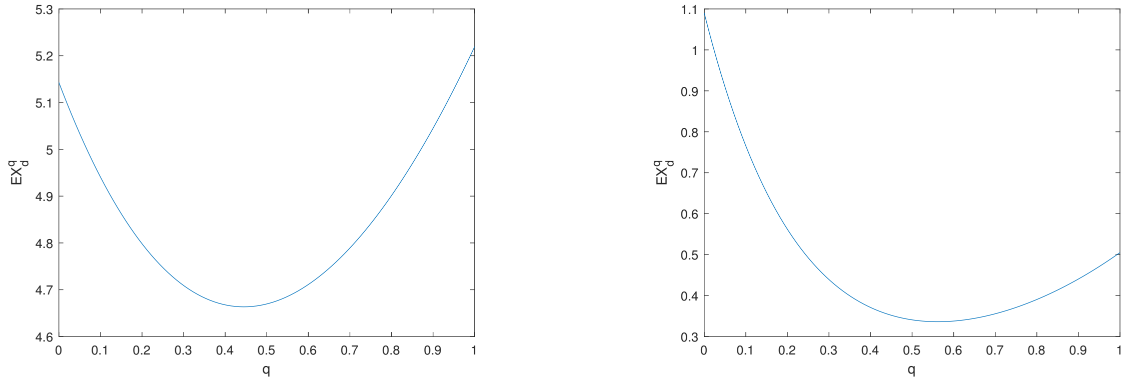

Figure 3.4: Plots of $EX_d^q(m)$ in Example 3.10 (left) and in Example 3.12 (right) as a function of q.

Example 3.12. Given a FOD $X = \{a, b, c\}$ and a BPA m such that $m(a) = 0.9$, $m(a, b) = 0.01$ and $m(X) = 0.09$, we have

$$EX_d^q(m) = 0.1\,[\log_2 30]^q + 0.99 \left[\log_2 \frac{100}{99}\right]^q .$$

In Figure 3.4 (right), the plot of $EX_d^q(m)$ is given. Note that as a function of q, it has a convex parabolic shape and that the maximum is achieved when q goes to zero.

Similar to FDEn, in the above examples, it is seen that the function $EX_d^q(m)$ cannot be a concave function and it can be increasing, decreasing and upside-down bathtub shape. Furthermore, the supremum FDEx is achieved when q is near the boundary of interval $(0, 1]$. The following theorem is readily obtained.

Theorem 3.4. *Let m be a non-degenerate BPA on a FOD X and $q \in (0, 1]$. Then, the supremum FDEx as a function of q is attained for $q \in \{0, 1\}$ and the infimum is attained in the extremes of interval $(0, 1)$ or it is a minimum assumed in a unique $q_0 \in (0, 1)$.*

Proof. The proof is similar to that of Theorem 3.2. In this case, the supremum is given by $\max\{N - 1 + m(X), EX_d(m)\}$, where N is the number of focal elements different from X. $\square$

In analogy with Theorem 3.3, we obtain an upper bound for the maximum FDEx with a fixed value of q.

Theorem 3.5. *Let X be a FOD, $q \in (0, 1]$ and let m be a BPA that assigns positive mass to each non-empty subset of X. For a fixed value of $m(X)$, an upper bound for the FDEx is assumed in correspondence of the fictitious BPA $\tilde{m}$ such that $\tilde{m}(X) = m(X)$ and*

$$\tilde{m}(A) = 1 - \frac{2^{|X|} - 3 + m(X)}{\sum_{\emptyset \neq B \subset X} (2^{|B^c|} - 1)} \left(2^{|A^c|} - 1\right), \quad \emptyset \neq A \subset X. \tag{3.17}$$

Proof. The proof is similar to the one given for Theorem 3.3. After establishing that $\frac{1-m(A)}{2^{|A^c|}-1} = K$ has to be invariant with respect to A, in order to satisfy the condition of normalization, we get

$$1 - m(A) = K\left(2^{|A^c|} - 1\right)$$

and, by summing over $A \subset X$

$$K = \frac{2^{|X|} - 3 + m(X)}{\sum_{\emptyset \neq A \subset X} (2^{|A^c|} - 1)}.$$

Hence, the BPA which maximizes the FDEx is given in (3.17). We remark that it is a fictitious BPA, in the sense that $\tilde{m}(A)$ may be negative for some subsets of X. □

Example 3.13. Based on the result of Theorem 3.5, let us evaluate the upper bound for FDEx in the case $|X| = 3$ with fixed $m(X)$. There are three subsets of cardinality one and three of cardinality two, and then the upper bound is given by

$$\begin{aligned} U &= 3 \cdot \frac{3(5+m(X))}{12}\left[-\log_2\left(\frac{5+m(X)}{12}\right)\right]^q + 3 \cdot \frac{5+m(X)}{12}\left[-\log_2\left(\frac{5+m(X)}{12}\right)\right]^q \\ &= (5+m(X))\left[\log_2\left(\frac{12}{5+m(X)}\right)\right]^q. \end{aligned}$$

3.4 Unified formulation of entropy

In this section, based on Balakrishnan, Buono and Longobardi [5], a general formulation of entropy is proposed. It depends on two parameters and includes Shannon, Tsallis and fractional entropy, all as special cases. This measure of information is named fractional Tsallis entropy. Then, the corresponding entropy in the context of Dempster-Shafer theory of evidence is proposed and referred to as fractional version of Tsallis-Deng entropy.

The fractional Tsallis entropy of X is defined as

$$S_\alpha^q(X) = \frac{1}{\alpha - 1}\sum_{i=1}^N p_i(1 - p_i^{\alpha-1})(-\log p_i)^{q-1}, \tag{3.18}$$

where $\alpha > 0$, $\alpha \neq 1$ and $0 < q \leq 1$. As mentioned above, a distinct advantage of this definition is that it includes fractional, Tsallis and Shannon entropies, all as special cases.

Remark 3.1. The fractional Tsallis entropy in (3.18) is always non-negative. This is due to the fact that $1 - p_i^{\alpha-1}$ is positive for $\alpha > 1$ and negative for $0 < \alpha < 1$, and so the sum in (3.18) has a definite sign and it is the same as that of $\alpha - 1$.

Proposition 3.2. *The fractional Tsallis entropy in* (3.18) *coincides with Tsallis entropy* (2.4) *when* $q = 1$.

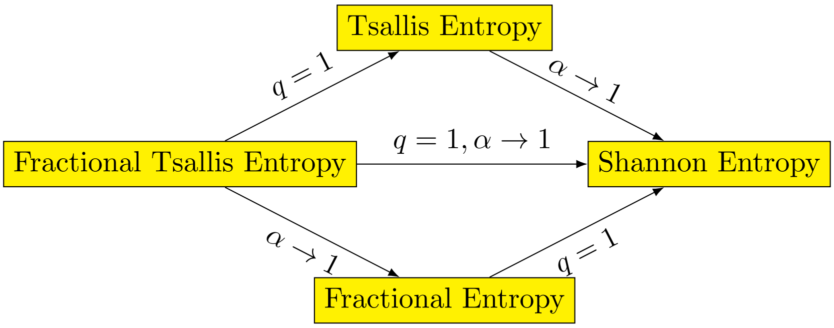

Figure 3.5: Relationships among different entropies in classical probability theory.

Proof. When $q = 1$, from (3.18) it follows

$$S_\alpha^1(X) = \frac{1}{\alpha - 1}\sum_{i=1}^{N} p_i(1 - p_i^{\alpha-1}) = \frac{1}{\alpha - 1}\left[1 - \sum_{i=1}^{N} p_i^\alpha\right] = S_\alpha(X),$$

as required. □

Proposition 3.3. *As α tends to 1, fractional Tsallis entropy converges to fractional entropy* (2.5).

Proof. Taking the limit as α tends to 1 in (3.18), and by using L'Hôpital's rule, we obtain

$$\begin{aligned} \lim_{\alpha\to1} S_\alpha^q(X) &= \lim_{\alpha\to1} \frac{1}{\alpha - 1}\sum_{i=1}^{N} p_i(1 - p_i^{\alpha-1})(-\log p_i)^{q-1} \\ &= \lim_{\alpha\to1} \sum_{i=1}^{N} p_i\left(-p_i^{\alpha-1}\right)\log p_i(-\log p_i)^{q-1} \\ &= \lim_{\alpha\to1} \sum_{i=1}^{N} p_i^\alpha(-\log p_i)^q = \sum_{i=1}^{N} p_i(-\log p_i)^q = H_q(X), \end{aligned}$$

as required. □

Corollary 3.1. *If both parameters of the fractional Tsallis entropy tend to* 1*, then the fractional Tsallis entropy in* (3.18) *converges to Shannon entropy, i.e.,*

$$\lim_{\alpha,q\to1} S_\alpha^q(X) = H(X).$$

To summarize the results given in Propositions 3.2 and 3.3 and Corollary 3.1, the relationships among different kinds of entropies are depicted in the form of a schematic diagram, in Figure 3.5.

Example 3.14. Let X be uniformly distributed over a support S of cardinality N. When N changes, the values of fractional Tsallis entropy are computed for different choices of α and q, and these are presented in Table 3.2. From Table 3.2, we observe that fractional Tsallis

Table 3.2: Values of fractional Tsallis entropy as N changes, for different choices of q and α.

N	$q=0.5$, $\alpha=0.5$	$q=0.5$, $\alpha=2$	$q=1$, $\alpha=2$	$q=0.3$, $\alpha=0.75$	$q=0.3$, $\alpha=1.5$	$q=0.5$, $\alpha=5$
2	0.9950	0.6006	0.5	0.9782	0.7571	0.2815
3	1.3968	0.6360	0.6667	1.1837	0.7914	0.2356
4	1.6986	0.6370	0.75	1.3182	0.7956	0.2115
5	1.9487	0.6306	0.8	1.4200	0.7923	0.1967
6	2.1657	0.6226	0.8333	1.5207	0.7868	0.1866
7	2.3596	0.6145	0.8571	1.5727	0.7807	0.1791
8	2.5359	0.6068	0.875	1.6336	0.7745	0.1733
9	2.6985	0.5997	0.8889	1.6877	0.7685	0.1686
10	2.8499	0.5931	0.9	1.7364	0.7628	0.1647

entropy does not always exhibit the same monotonic behavior as a function of N. For the uniform distribution, the fractional Tsallis entropy is given by

$$S_\alpha^q\left(\frac{1}{p},\dots,\frac{1}{p}\right)=\frac{1}{\alpha-1}\left(1-\frac{1}{N^{\alpha-1}}\right)(\log N)^{q-1}.$$

By treating N as continuous and upon differentiating with respect to N, we obtain

$$\begin{aligned}\frac{\partial S_\alpha^q}{\partial N} &= \frac{(\log N)^{q-2}}{N^\alpha}\frac{1}{\alpha-1}\left[(\alpha-1)\log N+(q-1)(N^{\alpha-1}-1)\right]\\ &\overset{sgn}{=} \frac{1}{\alpha-1}\left[(\alpha-1)\log N+(q-1)(N^{\alpha-1}-1)\right].\end{aligned}\tag{3.19}$$

Then, by observing that

$$(\alpha-1)\log N+(q-1)(N^{\alpha-1}-1)=\log N^{\alpha-1}+(q-1)(N^{\alpha-1}-1),$$

we see that we need to consider the function

$$\rho(v)=\log v+(q-1)(v-1),\quad v>0,$$

which is increasing for $v<\frac{1}{1-q}$. Here, v represents $N^{\alpha-1}$ which is in $(0,1)$ for $\alpha\in(0,1)$ and greater than $2^{\alpha-1}$ for $\alpha>1$. As $\frac{1}{1-q}>1$, with $q\in(0,1)$, ρ is increasing in $(0,1)$ and reaches the maximum value for $v=1$, that is, $\rho(1)=0$. Hence, for $\alpha\in(0,1)$, the sign in (3.19) is given by the ratio of two negative quantities and so fractional Tsallis entropy is increasing in N. For $q=1$, it is easy to observe by (3.19) that it is increasing in N regardless of α. Finally, for $\alpha>1$ and $q\in(0,1)$, there are two possible scenarios. In fact, the fractional Tsallis entropy may be always decreasing in N, or simply definitely decreasing as seen in Table 3.2.

Theorem 3.6. *The supremum of the fractional Tsallis entropy, as a function of $q\in(0,1]$, is attained at one of the extremes of the interval, and the infimum is attained at one of the extremes of the interval or it is a minimum at a unique $q_0\in(0,1)$.*

Proof. The fractional Tsallis entropy is a convex function of q. Hence, it can be strictly increasing, strictly decreasing or decreasing up to $q_0 \in (0,1)$, and then increasing. In the first case, the infimum is attained at 0 and the maximum at $q = 1$. In the second case, the minimum is reached at $q = 1$ and the supremum at 0. In the last case, the minimum is attained at q_0 and the supremum is reached at one of the extremes of the interval $(0,1)$. □

3.4.1 Fractional version of Tsallis-Deng entropy

The fractional version of Tsallis-Deng entropy for a BPA m is defined as (see Balakrishnan, Buono and Longobardi [5])

$$SD_\alpha^q(m) = \frac{1}{\alpha-1} \sum_{A\subseteq X: m(A)>0} m(A) \left[1 - \left(\frac{m(A)}{2^{|A|}-1}\right)^{\alpha-1}\right] \left(-\log \frac{m(A)}{2^{|A|}-1}\right)^{q-1}, \tag{3.20}$$

where $\alpha > 0$, $\alpha \neq 1$, $0 < q \leq 1$. In analogy with the fractional Tsallis entropy, this is a general expression of entropy as it includes several versions of entropy measure both in the context of DST and in the classical probability theory viewpoint.

Remark 3.2. In analogy with Remark 3.1, fractional version of Tsallis-Deng entropy (3.20) is non-negative too.

Proposition 3.4. *When $q = 1$, the fractional version of Tsallis-Deng entropy in* (3.20) *is equal to Tsallis-Deng entropy* (3.5).

Proof. Upon taking $q = 1$ in (3.20), we conclude

$$SD_\alpha^1(m) = \frac{1}{\alpha-1} \sum_{A\subseteq X: m(A)>0} m(A) \left[1 - \left(\frac{m(A)}{2^{|A|}-1}\right)^{\alpha-1}\right] = SD_\alpha(m),$$

as required. □

Proposition 3.5. *As α tends to 1, the fractional version of Tsallis-Deng entropy in* (3.20) *reduces to the fractional Deng entropy* (3.11).

Proof. Upon letting α tend to 1 in (3.20), and by using L'Hôpital's rule, it follows

$$\begin{aligned}
\lim_{\alpha\to1} SD_\alpha^q(m) &= \lim_{\alpha\to1} \sum_{A\subseteq X: m(A)>0} m(A) \left[-\left(\frac{m(A)}{2^{|A|}-1}\right)^{\alpha-1}\right] \left(-\log \frac{m(A)}{2^{|A|}-1}\right)^{q-1} \log \frac{m(A)}{2^{|A|}-1} \\
&= \lim_{\alpha\to1} \sum_{A\subseteq X: m(A)>0} m(A) \left(\frac{m(A)}{2^{|A|}-1}\right)^{\alpha-1} \left(-\log \frac{m(A)}{2^{|A|}-1}\right)^{q} \\
&= \sum_{A\subseteq X: m(A)>0} m(A) \left(-\log \frac{m(A)}{2^{|A|}-1}\right)^{q} = E_d^q(m),
\end{aligned}$$

as required. □

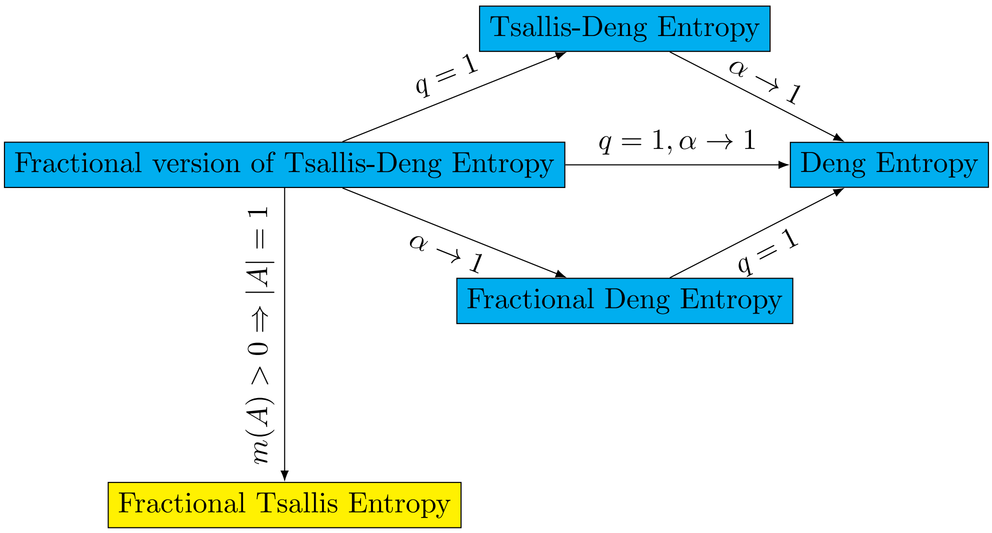

Figure 3.6: Relationships among different entropies in DST (blue) and in classical probability theory (yellow).

Corollary 3.2. *When both parameters α and q in* (3.20) *tend to* 1*, the fractional version of Tsallis-Deng entropy in* (3.20) *converges to Deng entropy* (3.4)*, i.e.,*

$$\lim_{\alpha,q\to 1} SD_{\alpha}^{q}(m) = E_d(m).$$

Remark 3.3. If the BPA m is a discrete probability distribution, then for each focal element $|A| = 1$; in this case, the fractional version of Tsallis-Deng entropy reduces to fractional Tsallis entropy (3.18), i.e.,

$$SD_{\alpha}^{q}(m) = \frac{1}{\alpha-1} \sum_{A\subseteq X:m(A)>0} m(A)\left[1-(m(A))^{\alpha-1}\right](-\log m(A))^{q-1} = S_{\alpha}^{q}(X),$$

where X is a discrete random variable with probability vector m.

To summarize the results given in Propositions 3.4 and 3.5, Corollary 3.2 and Remark 3.3, the relationships among different kinds of entropies are depicted in the form of a schematic diagram, see Figure 3.6.

Theorem 3.7. *The supremum of the fractional version of Tsallis-Deng entropy in* (3.20)*, as a function of $q \in (0,1]$, is attained at one of the extremes of the interval, and the infimum is attained at one of the extremes of the interval or it is a minimum at a unique $q_0 \in (0,1)$.*

Proof. The proof is similar to that of Theorem 3.6 and is therefore omitted. □

Example 3.15. Consider a frame of discernment $X = \{1, 2, \ldots, 15\}$ and a BPA m such that $m(3, 4, 5) = 0.05$, $m(6) = 0.05$, m(A)=0.8, $m(X) = 0.1$. When A changes, the values of the fractional version of Tsallis-Deng entropy in (3.20) are computed for different choices of α and q, and these are presented in Table 3.3.

Table 3.3: Values of the fractional version of Tsallis-Deng entropy as A changes, for different choices of q and α.

A	$q=0.5$ $\alpha=0.5$	$q=0.5$ $\alpha=2$	$q=1$ $\alpha=2$	$q=0.3$ $\alpha=0.75$	$q=0.3$ $\alpha=1.5$	$q=0.5$ $\alpha=5$
$\{1\}$	35.1571	0.4165	0.3571	2.3353	0.5823	0.2698
$\{1,2\}$	34.0606	0.5881	0.7838	2.8415	0.7361	0.1929
$\{1,2,3\}$	34.8845	0.5590	0.9057	3.1505	0.7156	0.1556
$\{1,\dots,4\}$	35.8694	0.5202	0.9545	3.4400	0.6793	0.1367
$\{1,\dots,5\}$	37.1288	0.4854	0.9765	3.7409	0.6415	0.1244
$\{1,\dots,6\}$	38.7866	0.4558	0.9870	4.0676	0.6056	0.1156
$\{1,\dots,7\}$	41.0020	0.4310	0.9921	4.4306	0.5726	0.1087
$\{1,\dots,8\}$	43.9887	0.4100	0.9946	4.8390	0.5428	0.1032
$\{1,\dots,9\}$	48.0383	0.3921	0.9959	5.3023	0.5160	0.0986
$\{1,\dots,10\}$	53.5510	0.3767	0.9965	5.8305	0.4919	0.0946
$\{1,\dots,11\}$	61.0779	0.3633	0.9968	6.4350	0.4704	0.0913
$\{1,\dots,12\}$	71.3801	0.3515	0.9970	7.1288	0.4512	0.0883
$\{1,\dots,13\}$	85.5090	0.3411	0.9971	7.9267	0.4339	0.0857
$\{1,\dots,14\}$	104.9202	0.3317	0.9971	8.8461	0.4183	0.0833

Example 3.16. In this example, we consider a well-known BPA defined for all $A \subseteq X$ as

$$m^*(A) = \frac{2^{|A|}-1}{\sum_{B\subseteq X}\left(2^{|B|}-1\right)},$$

so that

$$\frac{m^*(A)}{2^{|A|}-1} = \frac{1}{\sum_{B\subseteq X}\left(2^{|B|}-1\right)} = K,$$

where $K \in (0,1)$ is a constant. The interest of this BPA is due to the fact that it gives a degree of belief to each non-empty subset of the frame of discernment and the mass of a focal element depends only on its cardinality. Then, the fractional version of Tsallis-Deng entropy can be evaluated for the BPA m^* as

$$\begin{aligned} SD_\alpha^q(m^*) &= \frac{1}{\alpha-1}\sum_{A\subseteq X}\left(2^{|A|}-1\right)K\left[1-K^{\alpha-1}\right](-\log K)^{q-1} \\ &= \frac{1}{\alpha-1}\left[1-K^{\alpha-1}\right](-\log K)^{q-1}. \end{aligned} \tag{3.21}$$

It is a decreasing function in α since the partial derivative with respect to α of the function in (3.21) has the same sign as that of the function

$$g(x) = x - 1 - x\log x$$

which is non-positive for each $x > 0$. Based on Proposition 3.5, $\alpha = 1$ is a removable discontinuity and then the supremum of $SD_\alpha^q(m^*)$ is attained for $\alpha \to 0^+$ and it is given by

$$\lim_{\alpha\to 0^+}\frac{1}{\alpha-1}\left[1-K^{\alpha-1}\right](-\log K)^{q-1} = \left[K^{-1}-1\right](-\log K)^{q-1}.$$

3.5 Application to classification problems

In this section, the measures of uncertainty described in this chapter and Tsallis extropy given in Chapter 2 are applied to classification problems. More precisely, two datasets are considered: the Iris dataset and the wine dataset both given in [41]. We start with the classification problems related to the Iris dataset. In the first example, we apply Tsallis extropy to define a classification technique, see Balakrishnan, Buono and Longobardi [7].

Example 3.17. The objective is to classify among three classes of flowers: Iris Setosa (Se), Iris Versicolor (Ve) and Iris Virginica (Vi). The dataset consists of 150 samples, with 50 in each class. The characteristics measured for each flower are: the sepal length in cm (SL), the sepal width in cm (SW), the petal length in cm (PL), the petal width in cm (PW) and the class (one of Se, Ve and Vi). We select 40 samples for each kind of Iris and then we find a sample of max-min value to generate a model of interval numbers, as shown in Table 3.4. Each element of the dataset can be regarded as an unknown test sample. Suppose the selected sample data is (6.1, 3.0, 4.9, 1.8, Vi).

Table 3.4: The interval numbers of the statistical model.

Item	**SL**	**SW**	**PL**	**PW**
Se	[4.4,5.8]	[2.3,4.4]	[1.0,1.9]	[0.1,0.6]
Ve	[4.9,7.0]	[2.0,3.4]	[3.0,5.1]	[1.0,1.7]
Vi	[4.9,7.9]	[2.2,3.8]	[4.5,6.9]	[1.4,2.5]

We then generate four discrete probability distributions using the method proposed by Kang et al. [61] based on the similarity of interval numbers. Given two intervals $A = [a_1, a_2]$ and $B = [b_1, b_2]$, their similarity $S(A, B)$ is defined as

$$S(A, B) = \frac{1}{1 + \gamma \ D(A, B)}, \tag{3.22}$$

where $\gamma > 0$ is the coefficient of support (we have used $\gamma = 5$), and $D(A, B)$, the distance between intervals A and B, is defined in [117] as

$$D^2(A, B) = \left[\left(\frac{a_1 + a_2}{2} \right) - \left(\frac{b_1 + b_2}{2} \right) \right]^2 + \frac{1}{3} \left[\left(\frac{a_2 - a_1}{2} \right)^2 + \left(\frac{b_2 - b_1}{2} \right)^2 \right]. \tag{3.23}$$

To generate probability distributions, the intervals given in Table 3.4 are used for interval A and for interval B we use singletons given by the selected sample. For each one of the four characteristics measured, we get three values of similarity and then we obtain a discrete probability distribution by normalizing them (see Table 3.5). Then, we evaluate the Tsallis

Table 3.5: Probability distributions based on Kang's method.

	SL	**SW**	**PL**	**PW**
$\mathbb{P}(Se)$	0.3058	0.2748	0.1391	0.1563
$\mathbb{P}(Ve)$	0.4148	0.3516	0.3801	0.3737
$\mathbb{P}(Vi)$	0.2794	0.3736	0.4808	0.4700

Table 3.6: Tsallis extropy for different choices of α.

	SL	**SW**	**PL**	**PW**
$\alpha = 0.5$	0.8941	0.8965	0.8715	0.8759
$\alpha = 0.7$	0.8560	0.8592	0.8267	0.8324
$\alpha = 1.5$	0.7245	0.7291	0.6781	0.6871
$\alpha = 2$	0.6564	0.6613	0.6050	0.6150

extropy of these probability distributions, as presented in Table 3.6, wherein we have used different values for the parameter α (0.5, 0.7, 1.5 and 2).

We use the Tsallis extropies in Table 3.6 to generate other probability distributions. Observe that the higher the extropy, the higher the uncertainty, and so it would be reasonable to give more weight to observations related to characteristics with lower Tsallis extropy. We refer to the obtained Tsallis extropies as $JS_\alpha(SL)$, $JS_\alpha(SW)$, $JS_\alpha(PL)$, $JS_\alpha(PW)$. Due to the monotonicity of the exponential function, we choose as baseline weight the function $w(x) = e^{-x}$, and we can then obtain the weights ω by normalization. For example, for the sepal length, we obtain the weight as

$$\omega(SL) = \frac{e^{-JS_\alpha(SL)}}{e^{-JS_\alpha(SL)} + e^{-JS_\alpha(SW)} + e^{-JS_\alpha(PL)} + e^{-JS_\alpha(PW)}}.$$

The values of the weights are listed in Table 3.7 for different choices of the parameter α. Then, we determine a final probability distribution in the following way: for each kind of flower, we have four probabilities, one for a specific characteristic; we multiply the probabilities given in Table 3.5 by the corresponding weights and then sum the values relating to the same class. For example, the probability of the class Iris Setosa is obtained as follows:

$$\mathbb{P}(Se) = 0.3058 \cdot \omega(SL) + 0.2748 \cdot \omega(SW) + 0.1391 \cdot \omega(PL) + 0.1563 \cdot \omega(PW).$$

Thus, by choosing $\alpha = 0.5$, we obtain the final probability distribution to be

$$\mathbb{P}(Se) = 0.2182, \quad \mathbb{P}(Ve) = 0.3800, \quad \mathbb{P}(Vi) = 0.4018,$$

Table 3.7: The weights for different choices of α.

	$\omega(SL)$	$\omega(SW)$	$\omega(PL)$	$\omega(PW)$
$\alpha = 0.5$	0.2476	0.2470	0.2533	0.2522
$\alpha = 0.7$	0.2469	0.2461	0.2542	0.2528
$\alpha = 1.5$	0.2450	0.2439	0.2567	0.2544
$\alpha = 2$	0.2445	0.2433	0.2574	0.2548

and then the decision is that the selected flower belongs to the class with the higher probability, Iris Virginica, i.e., we thus made the correct decision. In this manner, we tested all 150 samples for different values of α, and observed that the overall recognition rate of this method based on the Tsallis extropy to be 94.66%.

In the second example, we consider again the Iris dataset and apply Deng extropy to solve the classification problem, see Buono and Longobardi [25]. Then, we make some comparisons with the recognition rates obtained in Example 3.17.

Example 3.18. In this example, we consider a classification problem based on the dataset studied in Example 3.17. We compare a method proposed by Kang et al. [61] with methods based on the use of Deng extropy and Tsallis extropy. The Iris dataset is useful to introduce the application of the generation of BPAs based on the Deng extropy in the classification of the kind of flowers. In analogy with Table 3.4, by selecting 40 samples for each kind of Iris and then by using the sample of max–min value, we generate a model of interval numbers, as shown in Table 3.8. Suppose the selected sample data is (6.1, 3.0, 4.9, 1.8, Iris Virginica).

Table 3.8: The interval numbers of the statistical model.

Item	**SL**	**SW**	**PL**	**PW**
Se	[4.4,5.8]	[2.3,4.4]	[1.0,1.9]	[0.1,0.6]
Ve	[4.9,7.0]	[2.0,3.4]	[3.0,5.1]	[1.0,1.7]
Vi	[4.9,7.9]	[2.2,3.8]	[4.5,6.9]	[1.4,2.5]
Se, Ve	[4.9,5.8]	[2.3,3.4]	–	–
Se, Vi	[4.9,5.8]	[2.3,3.8]	–	–
Ve, Vi	[4.9,7.0]	[2.2,3.4]	[4.5,5.1]	[1.4,1.7]
Se, Ve, Vi	[4.9,5.8]	[2.3,3.4]	–	–

Four BPAs are generated with a method proposed by Kang et al. [61] based on the similarity of interval numbers given in (3.22). In order to generate BPAs, the intervals given in Table 3.8 are used as interval A and as interval B the singletons given by the selected sample are employed. Again, we have chosen $\gamma = 5$. For each one of the four properties, we get seven values of similarity and then we get a BPA by normalizing them (see Table 3.9). Hence, we evaluate the Deng extropy of these BPAs, as shown in the bottom row of Table 3.9. We obtain a combined BPA by using the Dempster rule of Combination (3.3). The type of unknown sample is determined by the combined BPA. From Equation (3.1), we get the maximum value of PPT. Hence, Kang's method assigns to the sample the type Iris Versicolor and it does not make the right decision.

Table 3.9: BPAs based on Kang's method, Deng extropy and final fusion result.

Item	**SL**	**SW**	**PL**	**PW**	**Combined BPA**
$m(Se)$	0.1098	0.1018	0.0625	0.1004	0.0059
$m(Ve)$	0.1703	0.1303	0.1839	0.2399	0.4664
$m(Vi)$	0.1257	0.1385	0.1819	0.3017	0.4656
$m(Se, Ve)$	0.1413	0.1663	0.0000	0.0000	0.0000
$m(Se, Vi)$	0.1413	0.1441	0.0000	0.0000	0.0000
$m(Ve, Vi)$	0.1703	0.1527	0.5719	0.3580	0.0620
$m(Se, Ve, Vi)$	0.1413	0.1663	0.0000	0.0000	0.0000
Deng extropy	5.2548	5.2806	5.1636	4.9477	

Next, we use the Deng extropies given in Table 3.9 to generate other BPAs. We refer to these extropies as $EX_d(SL)$, $EX_d(SW)$, $EX_d(PL)$, $EX_d(PW)$. We use these values as significance of each sample property to evaluate the corresponding weight. For the sample property sepal length we have

$$\omega(SL) = \frac{e^{-EX_d(SL)}}{e^{-EX_d(SL)} + e^{-EX_d(SW)} + e^{-EX_d(PL)} + e^{-EX_d(PW)}}.$$

We divide each weight by the maximum of the weights and use these values as discounting coefficients to generate new BPAs, as shown in Equation (3.2), see Table 3.10. Again, a combined BPA is obtained by using the Dempster rule of combination. The type of unknown sample is determined by the combined BPA. Hence, the method based on the Deng extropy can make the right recognition.

We tested all 150 samples and we get that the global recognition rate of the method based on the Deng extropy is 94%. Then, we compare the results obtained with those based on the method by Kang et al. [61] and with the ones obtained in Example 3.17 by using Tsallis extropy, see Table 3.11.

Table 3.10: The modified BPAs based on the Deng extropy and final fusion result.

Item	SL	SW	PL	PW	Combined BPA
$m(Se)$	0.0808	0.0730	0.0504	0.1004	0.0224
$m(Ve)$	0.1252	0.0934	0.1482	0.2399	0.4406
$m(Vi)$	0.0925	0.0993	0.1465	0.3017	0.4451
$m(Se, Ve)$	0.1039	0.1192	0.0000	0.0000	0.0000
$m(Se, Vi)$	0.1039	0.1033	0.0000	0.0000	0.0000
$m(Ve, Vi)$	0.1252	0.1095	0.4608	0.3580	0.0919
$m(Se, Ve, Vi)$	0.3684	0.4023	0.1942	0.0000	0.0000

Table 3.11: The recognition rates of different methods.

Item	Se	Ve	Vi	Overall
Kang's method	100%	96%	84%	93.33%
Method based on Deng extropy	100%	96%	86%	94%
Method based on Tsallis extropy	100%	98%	86%	94.66%

In the following example, we apply the fractional Tsallis-Deng entropy to the classification problem for the Iris dataset, see Balakrishnan, Buono and Longobardi [5].

Example 3.19. Consider the Iris dataset analyzed in Examples 3.17 and 3.18. By using the method of max-min values, the model of interval numbers is obtained and is presented in Table 3.12. Suppose the selected instance is $(6.3, 2.7, 4.9, 1.8)$. It belongs to the kind Iris Virginica and then our aim is to classify it correctly. Four BPAs, one for each attribute, are generated by using the similarity of interval numbers as above. Without any additional information, the final BPA is determined by giving the same weight to each attribute, i.e., by summing the four values that are related to a focal element and then dividing by four. In this way, we get the final BPA as presented in Table 3.13.

In order to discriminate among classes, we evaluate the PPT of singleton classes for the BPA given in Table 3.13 and the results are

$$PPT(Se) = 0.1826, \quad PPT(Ve) = 0.4131, \quad PPT(Vi) = 0.4043.$$

Thus, the focal element with the highest PPT is the type Iris Versicolor, and would therefore be our final decision, which is not the correct one in this case. We now try to improve the

Table 3.12: The model of interval numbers.

Class	SL	SW	PL	PW
Se	$[4.3, 5.8]$	$[2.3, 4.4]$	$[1.0, 1.9]$	$[0.1, 0.6]$
Ve	$[4.9, 7.0]$	$[2.0, 3.4]$	$[3.0, 5.1]$	$[1.0, 1.8]$
Vi	$[4.9, 7.9]$	$[2.2, 3.8]$	$[4.5, 6.9]$	$[1.4, 2.5]$
Se, Ve	$[4.9, 5.8]$	$[2.3, 3.4]$	–	–
Se, Vi	$[4.9, 5.8]$	$[2.3, 3.8]$	–	–
Ve, Vi	$[4.9, 7.0]$	$[2.2, 3.4]$	$[4.5, 5.1]$	$[1.4, 1.8]$
Se, Ve, Vi	$[4.9, 5.8]$	$[2.3, 3.4]$	–	–

Table 3.13: Final BPA.

Class	Final BPA
$m(Se)$	0.0872
$m(Ve)$	0.1891
$m(Vi)$	0.1861
$m(Se, Ve)$	0.0759
$m(Se, Vi)$	0.0643
$m(Ve, Vi)$	0.3215
$m(Se, Ve, Vi)$	0.1759

method by using the fractional version of Tsallis-Deng entropy in (3.20). Fix the values of $q = 0.5$ and $\alpha = 0.5$. The fractional version of Tsallis-Deng entropy of BPAs obtained by using the similarity of interval numbers is then evaluated and the corresponding results are shown in Table 3.14.

Since a greater value of SD_α^q means a higher uncertainty, it is reasonable to give more weight to the attributes with lower SD_α^q. In this case, we define the weights by normalizing to 1 the exponential function of fractional versions of Tsallis-Deng entropies multiplied by minus one. The weights so determined are reported in Table 3.15.

Based on the weights in Table 3.15, a weighted version of the final BPA is obtained, as given in Table 3.16. Then, based on the weighted BPA in Table 3.16, we compute the PPT of the singleton classes to be

$$PPT(1) = 0.1156, \quad PPT(2) = 0.4360, \quad PPT(3) = 0.4485.$$

Table 3.14: Fractional versions of Tsallis-Deng entropies of BPAs based on similarity of interval numbers.

Attribute	**SL**	**SW**	**PL**	**PW**
SD_α^q	3.6184	3.7153	2.1226	2.0932

Table 3.15: The weights of attributes based on fractional version of Tsallis-Deng entropy.

Attribute	**SL**	**SW**	**PL**	**PW**
Weight	0.0912	0.0828	0.4069	0.4191

Thus, the focal element with the highest PPT is the type Iris Virginica, and would therefore be our final decision, which is indeed the correct one in this case.

Table 3.16: Final weighted BPA.

Class	**Final Weighted BPA**
Se	0.0825
Ve	0.2050
Vi	0.2194
Se, Ve	0.0262
Se, Vi	0.0224
Ve, Vi	0.4182
Se, Ve, Vi	0.0262

In Table 3.17, the recognition rates of the non-weighted method and methods based on fractional version of Tsallis-Deng entropy are presented for different choices of q and α.

In the following two examples, we consider classification problems related to the wine dataset. In the first example, the fractional Deng entropy and extropy are applied to define a classification rule, see Kazemi, Tahmasebi, Buono and Longobardi [62].

Example 3.20. In this example, we apply FDEn and FDEx to a classification problem. We analyze a dataset given in [41] about typical qualities of Italian wines. The dataset is composed of 178 instances and, for each one, thirteen attributes are given. The instances of the dataset

Table 3.17: The recognition rate for different choices of q and α.

Non-Weighted Method	q	α	**Fractional Tsallis-Deng Method**
94%	0.5	0.5	96.67%
	0.25	0.6	94.67%
	0.8	0.75	96.67%
	0.75	2	94%
	1	5	94%

are divided into three classes of wine: class 1, class 2 and class 3. We use six attributes to discriminate for each instance the correct class. In particular, the attributes involved in the example are: Alcohol, Malic acid, Ash, OD280/OD315 of diluted wines (OD), Color intensity (CI) and Proline. We use the method of max-min values to generate a model of interval numbers. In particular, for a fixed attribute, we study the interval of variability in a single class, and then we intersect the intervals of more classes. The model of interval numbers is shown in Table 3.18.

Table 3.18: The model of interval numbers.

Class	**Alcohol**	**Malic Acid**	**Ash**	**OD**	**CI**	**Proline**
1	$[12.850, 14.830]$	$[1.3500, 4.0400]$	$[2.0400, 3.2200]$	$[2.5100, 4.0000]$	$[3.5200, 8.9000]$	$[680, 1680]$
2	$[11.030, 13.860]$	$[0.7400, 5.8000]$	$[1.3600, 3.2300]$	$[1.5900, 3.6900]$	$[1.2800, 6.0000]$	$[278, 985]$
3	$[12.200, 14.340]$	$[1.2400, 5.6500]$	$[2.1000, 2.8600]$	$[1.2700, 2.4700]$	$[3.8500, 13.0000]$	$[415, 880]$
$1, 2$	$[12.850, 13.860]$	$[1.3500, 4.0400]$	$[2.0400, 3.2200]$	$[2.5100, 3.6900]$	$[3.5200, 6.0000]$	$[680, 985]$
$1, 3$	$[12.850, 14.340]$	$[1.3500, 4.0400]$	$[2.1000, 2.8600]$	$-$	$[3.8500, 8.9000]$	$[680, 880]$
$2, 3$	$[12.200, 13.860]$	$[1.2400, 5.6500]$	$[2.1000, 2.8600]$	$[1.5900, 2.4700]$	$[3.8500, 6.0000]$	$[415, 880]$
$1, 2, 3$	$[12.850, 13.860]$	$[1.3500, 4.0400]$	$[2.1000, 2.8600]$	$-$	$[3.8500, 6.0000]$	$[680, 880]$

Suppose the selected instance is $(13.860, 1.5100, 2.6700, 3.1600, 3.3800, 410)$. By the dataset, we know that the selected instance belongs to class 2, and our purpose is to classify it in the right way. We generate six BPAs, one for each attribute, by using the method proposed by Kang et al. [61] based on the similarity of interval numbers given in (3.22). Hence, for each attribute, we get seven values of similarity and by normalizing them, we get six BPAs, as reported in Table 3.19.

Without any additional information, we can evaluate a final BPA by giving the same weight to each attribute, i.e., by summing the six values related to a focal element and then dividing by six, see Table 3.20. Then, based on the BPA in Table 3.20, we evaluate the PPT of the

Table 3.19: BPAs based on Kang's method.

Class	Alcohol	Malic Acid	Ash	OD	CI	Proline
$m(1)$	0.1699	0.1685	0.1416	0.2700	0.0967	0.0623
$m(2)$	0.0715	0.1095	0.0897	0.1732	0.2088	0.1700
$m(3)$	0.1244	0.1083	0.1568	0.1126	0.0562	0.1877
$m(1,2)$	0.1675	0.1685	0.1416	0.3168	0.1889	0.1187
$m(1,3)$	0.1860	0.1685	0.1568	0.0000	0.0939	0.1368
$m(2,3)$	0.1132	0.1083	0.1568	0.1273	0.1777	0.1877
$m(1,2,3)$	0.1675	0.1685	0.1568	0.0000	0.1777	0.1368

Table 3.20: Final BPA.

Class	Final BPA
$m(1)$	0.1515
$m(2)$	0.1371
$m(3)$	0.1243
$m(1,2)$	0.1837
$m(1,3)$	0.1237
$m(2,3)$	0.1452
$m(1,2,3)$	0.1345

classes obtaining

$$PPT(1) = 0.3500, \quad PPT(2) = 0.3464, \quad PPT(3) = 0.3036.$$

Hence, the focal element with the highest PPT is class 1, and it would be our final hypothesis without making the correct decision.

Next, we try to improve the described method by using FDEn. Let us fix the value $q = 0.6$. We evaluate the FDEn of BPAs given in Table 3.19 and then we obtain the results that are listed in Table 3.21.

Since a higher value of FDEn means a higher uncertainty, we can give more weight to the attributes with lower FDEn. In particular, we define the weights by normalizing to 1 the reciprocal values of fractional Deng entropies. We obtain the weights presented in Table 3.22.

Based on the weights in Table 3.22, we get a weighted version of the final BPA, as shown

Table 3.21: Fractional Deng entropies of BPAs in Table 3.19.

Attribute	**Alcohol**	**Malic Acid**	**Ash**	**OD**	**CI**	**Proline**
$FDEn$	2.2684	2.2658	2.2638	1.8801	2.2494	1.4378

Table 3.22: The weights of attributes based on FDEn.

Attribute	**Alcohol**	**Malic Acid**	**Ash**	**OD**	**CI**	**Proline**
Weight	0.1472	0.1473	0.1474	0.1775	0.1484	0.2322

in Table 3.23. Finally, based on the BPA in Table 3.23, we evaluate the PPT of the singleton

Table 3.23: Final weighted BPA.

Class	**Final Weighted BPA**
$m(1)$	0.1474
$m(2)$	0.1411
$m(3)$	0.1293
$m(1,2)$	0.1822
$m(1,3)$	0.1210
$m(2,3)$	0.1483
$m(1,2,3)$	0.1307

classes and we get

$$PPT(1) = 0.3426, \quad PPT(2) = 0.3499, \quad PPT(3) = 0.3075.$$

Hence, the focal element with the highest PPT is class 2, so it is our final hypothesis and we make the correct decision. Along the same lines, we can use FDEx. In Table 3.24, we give the recognition rates of the non-weighted method and methods based on FDEn and FDEx for different choices of the parameter q.

In the last example, we consider the fractional Tsallis-Deng entropy to discriminate among classes in the wine dataset, see Balakrishnan, Buono and Longobardi [5].

Example 3.21. In this example, we consider again the classification problem based on the dataset about typical qualities of Italian wines described in Example 3.20. The model of interval

Table 3.24: The recognition rates.

Non-Weighted Method	q	FDEn Method	FDEx Method
93.26%	0.5	94.38%	93.26%
	0.6	94.94%	93.26%
	1	94.38%	93.26%

numbers is the one given in Table 3.18. Suppose the selected instance is $(12.330, 1.1000, 2.2800, 1.6700, 3.2700, 680)$. It belongs to Class 2 and our aim is to classify it correctly. Six BPAs, one for each attribute, are generated by using the method based on the similarity of interval numbers proposed by Kang et al. [61]. For each attribute, we obtain seven values of similarity and by normalizing them, six BPAs are computed and are presented in Table 3.25.

Table 3.25: BPAs based on Kang's method.

Class	Alcohol	Malic Acid	Ash	OD	CI	Proline
$m(1)$	0.1008	0.1645	0.1180	0.1062	0.0977	0.0394
$m(2)$	0.1785	0.1153	0.1098	0.1455	0.2137	0.1084
$m(3)$	0.1381	0.1132	0.1635	0.3259	0.0575	0.1646
$m(1,2)$	0.1445	0.1645	0.1180	0.1171	0.1859	0.1292
$m(1,3)$	0.1191	0.1645	0.1635	0.0000	0.0949	0.1969
$m(2,3)$	0.1745	0.1132	0.1635	0.3053	0.1751	0.1646
$m(1,2,3)$	0.1445	0.1645	0.1635	0.0000	0.1751	0.1969

Without any additional information, the final BPA is determined by giving the same weight to each attribute, i.e., by summing the six values that are related to a focal element and then dividing by six. In this way, we get the final BPA as presented in Table 3.26.

Now, based on the BPA in Table 3.26, the PPT of the singleton classes are computed and the values which we obtain are

$$PPT(1) = 0.2846, \quad PPT(2) = 0.3551, \quad PPT(3) = 0.3603.$$

Thus, the focal element with the highest PPT is Class 3, and would therefore be our final decision, which is not the correct one in this case.

Now, we try to improve the described method by using the fractional version of Tsallis-Deng entropy in (3.20). Fix the values of $q = 0.5$ and $\alpha = 4$. The fractional version of Tsallis-Deng

Table 3.26: Final BPA.

Class	Final BPA
$m(1)$	0.1045
$m(2)$	0.1452
$m(3)$	0.1605
$m(1,2)$	0.1432
$m(1,3)$	0.1232
$m(2,3)$	0.1827
$m(1,2,3)$	0.1408

entropy of BPAs given in Table 3.25 is then evaluated and the corresponding results are listed in Table 3.27.

Table 3.27: Fractional versions of Tsallis-Deng entropies of BPAs in Table 3.25.

Attribute	Alcohol	Malic Acid	Ash	OD	CI	Proline
SD_α^q	0.2085	0.2058	0.2057	0.2464	0.2087	0.2032

Since a greater value of SD_α^q represents a higher uncertainty, it is reasonable to give more weight to the attributes with lower SD_α^q. Specifically, we define the weights by normalizing to 1 the reciprocal values of the fourth power of fractional versions of Tsallis-Deng entropies. The weights so determined are reported in Table 3.28.

Table 3.28: The weights of attributes based on fractional version of Tsallis-Deng entropy.

Attribute	Alcohol	Malic Acid	Ash	OD	CI	Proline
Weight	0.1745	0.1838	0.1843	0.0895	0.1741	0.1938

Based on the weights in Table 3.28, a weighted version of the final BPA is obtained, as given in Table 3.29. Then, based on the weighted BPA in Table 3.29, we compute the PPT of the singleton classes as

$$PPT(1) = 0.2956, \quad PPT(2) = 0.3533, \quad PPT(3) = 0.3510.$$

Hence, the focal element with the highest PPT is Class 2, and would therefore be our final decision, which is indeed the correct one in this case.

Table 3.29: Final weighted BPA.

Class	**Final Weighted BPA**
$m(1)$	0.1037
$m(2)$	0.1438
$m(3)$	0.1431
$m(1,2)$	0.1451
$m(1,3)$	0.1358
$m(2,3)$	0.1711
$m(1,2,3)$	0.1542

In Table 3.30, the recognition rates of the non-weighted method and of the method based on the fractional version of Tsallis-Deng entropy are presented for different choices of q and α.

Table 3.30: The recognition rates for different choices of q and α.

Non-Weighted Method	$\boldsymbol{q}$	$\boldsymbol{\alpha}$	**Fractional Tsallis-Deng Method**
93.26%	0.5	4	93.82%
	0.6	3	93.82%
	0.1	0.8	91.57%
	1	5	93.26%

Chapter 4

Multivariate hazard rate functions

The hazard rate and the reversed hazard rate functions have been defined in the univariate case in Chapter 1. In a similar way, these function can be defined for multivariate distributions. If a random vector is considered, it is of great interest to analyze the case in which its components are interdependent. In this framework, two new versions of the hazard rate and the reversed hazard rate functions find their collocation and will be presented in this chapter. These functions are known as multivariate conditional hazard rate functions and multivariate conditional reversed hazard rate functions where the word "conditional" is related to the fact that they depend on something that is observed. The first ones have been defined by Shaked and Shanthikumar [105, 107] whereas the second ones have been introduced and studied in Buono, De Santis, Longobardi and Spizzichino [23].

4.1 Multivariate conditional hazard rate functions

The multivariate conditional hazard rate functions are useful tools to describe dependence models. By regarding the components of the vector as the lifetimes of the components in a system, their definition is related to the probability that one of the components fails immediately after a fixed time t given that it survived up to that time and given a dynamic observed history. The dynamic observed history consists in which components failed before time t and at what time their failures occur. Before analyzing the n-dimensional case, the bivariate case, which was first studied by Cox [31], is presented.

4.1.1 The bivariate case

Let (X_1, X_2) be a non-negative and absolutely continuous random vector. It is clear that in this case there are only two possible scenarios to take in consideration: X_1 fails before than X_2 and vice versa. We remark that the absolutely continuity assumption guarantees that $\mathbb{P}(X_1 = X_2) = 0$ and then we have not to consider the case in which X_1 and X_2 assume the same value. Up to the random time $X_{1:2}$, X_1 and X_2 have a proper hazard rate function but, when the first failure occurs, the other random variable may change its hazard rate due to the dependency. Hence, the following four functions are introduced and named multivariate conditional hazard rate functions

$$\lambda_1(t|\emptyset) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_1 \leq t + \Delta t | X_1 > t, X_2 > t\right), \ t \geq 0, \tag{4.1}$$

$$\lambda_2(t|\emptyset) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_2 \leq t + \Delta t | X_1 > t, X_2 > t\right), \ t \geq 0, \tag{4.2}$$

$$\lambda_1(t|2;t_2) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_1 \leq t + \Delta t | X_1 > t, X_2 = t_2\right), \ t \geq t_2 \geq 0, \tag{4.3}$$

$$\lambda_2(t|1;t_1) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_2 \leq t + \Delta t | X_1 = t_1, X_2 > t\right), \ t \geq t_1 \geq 0. \tag{4.4}$$

As mentioned above, $\lambda_1(t|\emptyset)$ and $\lambda_2(t|\emptyset)$ describe the hazard of X_1 and X_2, respectively, when they have both survived up to time t, whereas $\lambda_1(t|2;t_2)$ describes the hazard of X_1 at time t given that it has survived up to time t and that X_2 failed at a time $t_2 \leq t$. The interpretation of $\lambda_2(t|1;t_1)$ is analogous to that of $\lambda_1(t|2;t_2)$.

Remark 4.1. If X_1 and X_2 are independent, then $\lambda_1(t|\emptyset) = \lambda_1(t|2;t_2)$, for all $t_2 \geq 0$, and they coincide with the univariate hazard rate function of X_1, $r_{X_1}(t)$, for all $t \geq 0$. Analogously, $\lambda_2(t|\emptyset) = \lambda_2(t|1;t_1) = r_{X_2}(t)$, for all $t_1 \geq 0$ and $t \geq 0$.

The multivariate conditional hazard rate functions can be determined by the joint probability density function f, the joint cumulative distribution function F and the joint survival function $\overline{F}$ of (X_1, X_2) by the following relations

$$\lambda_1(t|\emptyset) = \frac{-\frac{\partial}{\partial t_1}\overline{F}(t_1,t)|_{t_1=t}}{\overline{F}(t,t)}, \ t \geq 0, \tag{4.5}$$

$$\lambda_2(t|\emptyset) = \frac{-\frac{\partial}{\partial t_2}\overline{F}(t,t_2)|_{t_2=t}}{\overline{F}(t,t)}, \ t \geq 0, \tag{4.6}$$

$$\lambda_1(t|2;t_2) = \frac{f(t,t_2)}{-\frac{\partial}{\partial t_2}\overline{F}(t,t_2)}, \ t > t_2 \geq 0, \tag{4.7}$$

$$\lambda_2(t|1;t_1) = \frac{f(t_1,t)}{-\frac{\partial}{\partial t_1}\overline{F}(t_1,t)}, \ t > t_1 \geq 0. \tag{4.8}$$

Conversely, it is possible to determinate the joint probability density function of (X_1, X_2) by Equations (4.1)–(4.4) as

$$f(t_1,t_2) = \begin{cases} \exp\left(-\int_0^{t_1}(\lambda_1(u|\emptyset)+\lambda_2(u|\emptyset))du\right)\lambda_1(t_1|\emptyset)\exp\left(-\int_{t_1}^{t_2}\lambda_2(u|1;t_1)du\right)\lambda_2(t_2|1;t_1), \\ \hfill \text{if } 0 \leq t_1 \leq t_2, \\ \exp\left(-\int_0^{t_2}(\lambda_1(u|\emptyset)+\lambda_2(u|\emptyset))du\right)\lambda_2(t_2|\emptyset)\exp\left(-\int_{t_2}^{t_1}\lambda_1(u|2;t_2)du\right)\lambda_1(t_1|2;t_2), \\ \hfill \text{if } 0 \leq t_2 \leq t_1. \end{cases}$$

In order to have a probability density function in the above relation, the functions $\lambda_1(\cdot|\emptyset), \lambda_2(\cdot|\emptyset)$, $\lambda_1(\cdot|2;t_2), \lambda_2(\cdot|1;t_1)$ have to satisfy some conditions. More precisely, it is necessary that $\int_0^{+\infty} \lambda_1(u|\emptyset) + \lambda_2(u|\emptyset) du = +\infty$, $\int_{t_1}^{+\infty} \lambda_2(u|1;t_1)du = +\infty$ for all $t_1 > 0$ and $\int_{t_2}^{+\infty} \lambda_1(u|2;t_2)du = +\infty$ for all $t_2 > 0$. In the following example, the multivariate conditional hazard rate functions of a well-known bivariate model are computed.

Example 4.1. Let us consider the Gumbel's type I bivariate exponential distribution with parameter $\theta \in [0,1]$. It has attracted the interest of researchers since it has a wide range of applications including competing risks, extreme values, failure times, regional analyses of precipitation, and reliability (see Nadarajah and Kotz [78]). The joint cumulative distribution function is expressed as

$$F(x,y) = 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)}, \quad x, y \geq 0,$$

and the joint density and survival functions are respectively given by

$$f(x,y) = e^{-(x+y+\theta xy)} \left[(1+\theta x)(1+\theta y) - \theta\right],$$

$$\overline{F}(x,y) = e^{-(x+y+\theta xy)}.$$

Then, the multivariate conditional hazard rate functions can be evaluated as

$$\lambda_1(t|\emptyset) = 1 + \theta t, \qquad \lambda_2(t|\emptyset) = 1 + \theta t, \tag{4.9}$$

$$\lambda_1(t|2;t_2) = \frac{(1+\theta t)(1+\theta t_2) - \theta}{1+\theta t}, \tag{4.10}$$

$$\lambda_2(t|1;t_1) = \frac{(1+\theta t)(1+\theta t_1) - \theta}{1+\theta t}, \tag{4.11}$$

where, since the two components are exchangeable, $\lambda_1(t|\emptyset) = \lambda_2(t|\emptyset)$ and $\lambda_1(t|2;z) = \lambda_2(t|1;z)$. If $\theta = 0$ the components are independent and exponentially distributed with mean 1 and all the above functions are constantly equal to 1 as the (univariate) hazard rate function of the standard exponential distribution.

4.1.2 The n-dimensional case

Let $X_1, \ldots, X_n$ be non-negative random variables with an absolutely continuous joint distribution. For a fixed index $j \in [n] = \{1, \ldots, n\}$ and $I = \{i_1, \ldots, i_k\} \subset [n]$ with $j \notin I$, and an ordered sequence $0 \leq t_1 \leq \cdots \leq t_k$, the multivariate conditional hazard rate (m.c.h.r.) function $\lambda_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ is defined as follows [107]:

$$\lambda_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_j \leq t + \Delta t \,\middle|\, X_{i_1} = t_1, \ldots, X_{i_k} = t_k, \min_{h \notin I} X_h > t \right). \tag{4.12}$$

Furthermore, we use the notation

$$\begin{aligned}\lambda_j(t|\emptyset) &= \lim_{\Delta t\to 0^+}\frac{1}{\Delta t}\mathbb{P}\left(X_j \le t+\Delta t \,\middle|\, \min_{h\in[n]} X_h > t\right)\\ &= \lim_{\Delta t\to 0^+}\frac{1}{\Delta t}\mathbb{P}\left(X_j \le t+\Delta t\,|X_{1:n} > t\right). \qquad (4.13)\end{aligned}$$

From Equation (4.12), it readily follows that the function $\lambda_j(t|i_1,\dots,i_k;t_1,\dots,t_k)$ describes the hazard of X_j at time t given an observed history, from 0 to t, in which the failures of components $X_{i_1},\dots,X_{i_k}$ have been observed at times $t_1,\dots,t_k$, respectively. Moreover, the function $\lambda_j(t|\emptyset)$ in Equation (4.13) describes the hazard rate of X_j when all the components assume a value greater than t and is generally known in literature as initial failure rate or risk-specific failure rate.

Remark 4.2. If the random variables $X_1,\dots,X_n$ are independent, then the multivariate conditional hazard rate functions degenerate into the classical hazard rate function, in the sense that $\lambda_j(t|i_1,\dots,i_k;t_1,\dots,t_k) = r_j(t)$ for all $t>0$ regardless of the indices $i_1,\dots,i_k$ and failure times $t_1,\dots,t_k$, where $r_j(\cdot)$ is the hazard rate function of X_j. Furthermore, if the random variables are exchangeable, the multivariate conditional hazard rate functions do not depend on j and $i_1,\dots,i_k$ but only on the cardinality of $I=\{i_1,\dots,i_k\}$ and the failure times $t_1,\dots,t_k$. Then, in the exchangeable case, Equations (4.12) and (4.13) become

$$\lambda_j(t|I;t_1,\dots,t_k) = \lambda^{(k)}(t|t_1,\dots,t_k), \qquad \lambda_j(t|\emptyset) = \lambda^{(0)}(t),$$

for $k=|I|\in\{1,2,\dots,n-1\}$ and $0\le t_1\le t_2\le\cdots\le t_k\le t$.

As in the bivariate case, the joint probability density function of $(X_1,\dots,X_n)$ can be determined and computed in terms of the multivariate conditional hazard rate functions. In fact, for $0\le t_1\le t_2\le\cdots\le t_n$, we have

$$\begin{aligned} f_{X_1,\dots,X_n}(t_1,\dots,t_n) &= \lambda_1(t_1|\emptyset)\exp\left[-\int_0^{t_1}\left(\sum_{j=1}^n \lambda_j(u|\emptyset)\right)du\right]\cdot\\ &\cdot\lambda_2(t_2|1;t_1)\exp\left[-\int_{t_1}^{t_2}\left(\sum_{j=2}^n \lambda_j(u|1;t_1)\right)du\right]\cdot\dots\\ &\cdot\lambda_{k+1}(t_k|1,\dots,k;t_1,\dots,t_k)\exp\left[-\int_{t_k}^{t_{k+1}}\left(\sum_{j=k+1}^n \lambda_j(u|1,\dots,k;t_1,\dots,t_k)\right)du\right]\cdot\dots\\ &\cdot\lambda_n(t_n|1,\dots,n-1;t_1,\dots,t_{n-1})\exp\left[-\int_{t_{n-1}}^{t_n}\lambda_n(u|1,\dots,n-1;t_1,\dots,t_{n-1})du\right].\end{aligned} \qquad (4.14)$$

Similar expressions hold when $t_1, \dots, t_n$ are such that $t_{\pi(1)} \leq \dots \leq t_{\pi(n)}$ for some permutation π of the set $\{1, \dots, n\}$. For details on the proof of this expression, one may refer to Shaked and Shanthikumar [105].

The multivariate conditional hazard rate functions are a useful tool to study the minimum among dependent random variables as shown in De Santis et al. [33]. In that paper, the authors proved that, for any vector of dependent random variables, the probabilities of the events related to the behavior of the minimum are equal to the probabilities of the same events for a vector of independent random variables. About the multivariate conditional hazard rate functions, it is possible to state the following result.

Proposition 4.1. *Let $(X_1, \dots, X_n)$ be a random vector with absolutely continuous joint distribution and m.c.h.r. functions $\lambda_j(\cdot|\emptyset)$. Let $Z_1, \dots, Z_n$ be independent random variables with (univariate) hazard rate functions $r_j(\cdot)$ such that*

$$r_j(t) = \lambda_j(t|\emptyset).$$

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

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

As a consequence of Proposition 4.1, in the particular case $B = [0, +\infty)$ it readily follows

$$\mathbb{P}(X_i = X_{1:n}) = \mathbb{P}(Z_i = Z_{1:n}), \quad i = 1, \dots, n.$$

Example 4.2. If $X_1, \dots, X_n$ are exchangeable, as stated in Remark 4.2, we have $\lambda_j(t|\emptyset) = \lambda^{(0)}(t)$, i.e., the random variables $Z_1, \dots, Z_n$ in Proposition 4.1 are independent and identically distributed. Then, as expected, it follows $\mathbb{P}(X_i = X_{1:n}) = \frac{1}{n}$, $i = 1, \dots, n$.

4.2 Load-Sharing models

The multivariate conditional hazard rate functions are efficient tools to describe the joint distribution and they are also useful to study interesting models. In this section, some properties of load-sharing models, and the corresponding time-homogeneous version, will be described. These models are efficiently determined by some conditions on the m.c.h.r. functions. In fact, if the m.c.h.r. functions do not depend on the failure times of the components, $t_1, \dots, t_{|I|}$ and on the order of $i_1, \dots, i_{|I|}$, then we have a Load-Sharing model. In this case, the current hazard of a working component only depends on the time of evaluation t and on the set of working components. Moreover, if in addition the m.c.h.r. functions do not depend also on the time of evaluation t, then, they are constant functions, we talk about Time-Homogeneous Load-Sharing models. In particular, Time-Homogeneous Load-Sharing models can be seen as a natural generalization of independent and exponentially distributed random variables. For a review of general properties of Load-Sharing (LS) models and Time-Homogeneous Load-Sharing (THLS) models see Ross [97], Schechner [102] and Spizzichino [110]. In the following, the formal definitions of LS and THLS models are given.

Definition 4.1. Let $(X_1, \ldots, X_n)$ be a random vector with absolutely continuous joint distribution. It is distributed according to a Load-Sharing model (LS) if, for any $i_1, \ldots, i_k$ and $j \in [n] \setminus I$, where $I = \{i_1, \ldots, i_k\}$, there exist functions $\mu_j(t|I)$ such that, for all $0 \leq t_1 \leq \cdots \leq t_k \leq t$,

$$\lambda_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = \mu_j(t|I).$$

Furthermore, a load-sharing model is time-homogeneous (THLS) if there exist non-negative numbers $\mu_j(I)$ and $\mu_j(\emptyset)$ such that, for any $t > 0$,

$$\begin{aligned} \mu_j(t|I) &= \mu_j(I), \\ \lambda_j(t|\emptyset) &= \mu_j(\emptyset). \end{aligned}$$

Remark 4.3. Notice that the joint distribution of n independent and exponential variables (non-necessarily identically distributed) is a special case of THLS.

Of course, under the assumption of a THLS model, the expression of the joint probability density function given in Equation (4.14) simplifies considerably and reduces to

$$\begin{aligned} f_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \mu_1(\emptyset) \exp\left[-t_1 \sum_{j=1}^{n} \mu_j(\emptyset)\right] \cdot \mu_2(\{1\}) \exp\left[-(t_2 - t_1) \sum_{j=2}^{n} \mu_j(\{1\})\right] \cdot \\ \ldots \cdot \mu_{k+1}(\{1, \ldots, k\}) \exp\left[-(t_{k+1} - t_k) \sum_{j=k+1}^{n} \mu_j(\{1, \ldots, k\})\right] \cdot \ldots \\ \cdot \mu_n(\{1, \ldots, n-1\}) \exp\left[-(t_n - t_{n-1}) \mu_n(\{1, \ldots, n-1\})\right], \end{aligned}$$

where $t_1 \leq t_2 \leq \cdots \leq t_n$.

Dealing with a THLS model, the following quantities are of great interest

$$\begin{aligned} M(I) &= \sum_{h \notin I} \mu_h(I), \\ \rho_j(I) &= \frac{\mu_j(I)}{M(I)}, \end{aligned}$$

since they are very useful in the study of the order statistics of $(X_1, \ldots, X_n)$ as stated in the following proposition by Spizzichino [110].

Proposition 4.2. *Let $(X_1, \ldots, X_n)$ be distributed according to a THLS model with parameters $\mu_j(\emptyset), \mu_j(I)$ and let π be a fixed permutation of $[n]$. Then, for $r = 1, 2, \ldots, n-1$*

$$\mathbb{P}(X_{1:n} = X_{\pi(1)}, \ldots, X_{r:n} = X_{\pi(r)}) = \\ \rho_{\pi(1)}(\emptyset)\rho_{\pi(2)}(\{\pi(1)\})\rho_{\pi(3)}(\{\pi(1), \pi(2)\}) \cdot \cdots \cdot \rho_{\pi(r)}(\{\pi(1), \pi(2), \ldots, \pi(r-1)\})$$

and

$$\mathbb{P}(X_{1:n} = X_{\pi(1)}, \ldots, X_{n:n} = X_{\pi(n)}) = \\ \rho_{\pi(1)}(\emptyset)\rho_{\pi(2)}(\{\pi(1)\})\rho_{\pi(3)}(\{\pi(1), \pi(2)\}) \cdot \cdots \cdot \rho_{\pi(n-1)}(\{\pi(1), \pi(2), \ldots, \pi(n-2)\}).$$

In the above proposition, the case in which r is equal to n is separated by the others to emphasize that, once chosen the first $n-1$ order statistics on n components, the last one is determined with probability equal to one.

In order to state the following result, let us denote by $\Lambda^{(r)}$ a vector $(\lambda_1, \ldots, \lambda_r) \in \mathbb{R}_+^r$ and by $\overline{G}_{\Lambda^{(r)}}(t)$ the survival function of the random variable $S_r = \sum_{s=1}^r \Gamma_s$, where $\Gamma_1, \ldots, \Gamma_r$ are independent random variables with exponential distribution of parameter $\lambda_1, \ldots, \lambda_r$, respectively. Moreover, for π permutation of $[n]$ and $r \in [n]$, we place

$$\Lambda^{(r)}(\pi) = (M(\emptyset), M(\{\pi(1)\}), \ldots, M(\{\pi(1), \ldots, \pi(r-1)\})).$$

We have the following proposition by Spizzichino [110].

Proposition 4.3. *Let $(X_1, \ldots, X_n)$ be distributed according to a THLS model with parameters $\mu_j(\emptyset), \mu_j(I)$. Then, for any $t > 0$ and $j \in [n]$,*

$$\mathbb{P}(X_{1:n} > t | X_{1:n} = X_j) = \exp(-tM(\emptyset)),$$

and for any permutation π of $[n]$ and $k = 2, \ldots, n$,

$$\mathbb{P}(X_{k:n} > t | X_{1:n} = X_{\pi(1)}, \ldots, X_{k-1:n} = X_{\pi(k-1)}, X_{k:n} = X_{\pi(k)}) = \overline{G}_{\Lambda^{(k)}(\pi)}(t). \tag{4.15}$$

4.3 Multivariate conditional reversed hazard rate functions

In this section, we introduce the dual version of m.c.h.r. functions and THLS model, namely the multivariate conditional reversed hazard rate functions and the reversed time-homogeneous load-sharing model. The reversed multivariate conditional hazard rate functions extend the one-dimensional notion of reversed hazard rate of a single non-negative random variable and have a related role in the study of the behavior of the maximum value among inter-dependent lifetimes. Moreover, the class of reversed load-sharing models can be seen as natural extensions to the multivariate case of the univariate inverse exponential distributions. The results of this section are based on Buono, De Santis, Longobardi and Spizzichino [23].

Let us consider a vector of n non-negative random variables $X_1, \ldots, X_n$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume the joint probability distribution of $X_1, \ldots, X_n$ is absolutely continuous and so ties among $X_1, \ldots, X_n$ have probability zero, i.e.,

$$P\left(X_{1:n} < \ldots < X_{n:n}\right) = 1. \tag{4.16}$$

In the following definition, for any fixed positive number t, the set I must be interpreted as the set of indices associated to the variables which take values greater than t. Correspondingly, $[n] \smallsetminus I$ is the set of indices of the variables which take values less than or equal to t.

Definition 4.2. For a vector $(i_1, \ldots, i_k)$, where $i_1 \neq \ldots \neq i_k \in [n]$, let us set $I \equiv \{i_1, \ldots, i_k\} \subset [n]$. For $j \notin I$ and an ordered sequence $0 \leq t \leq t_k \leq \cdots \leq t_1$ the multivariate conditional reversed hazard rate (m.c.r.h.r.) function $\tau_j(t|I; t_1, \ldots, t_k)$ is defined as follows:

$$\tau_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_j \geq t - \Delta t \,\middle|\, X_{i_1} = t_1, \ldots, X_{i_k} = t_k, \max_{h \in [n] \smallsetminus I} X_h \leq t\right).$$

Furthermore, we use the notation

$$\begin{aligned}\tau_j(t|\emptyset) &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_j \geq t - \Delta t \,\middle|\, \max_{h \in [n]} X_h \leq t\right) \\ &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_j \geq t - \Delta t \,|X_{n:n} \leq t\,\right).\end{aligned}$$

When necessary to distinguish between different vectors of lifetimes, notations as $\tau_j^{(\mathbf{X})}(t|\emptyset)$, $\lambda_j^{(\mathbf{X})}(t|\emptyset)$ in place of $\tau_j(t|\emptyset)$, $\lambda_j(t|\emptyset)$ and $\tau_j^{(\mathbf{X})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$, $\lambda_j^{(\mathbf{X})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ in place of $\tau_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$, $\lambda_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ will be used.

Remark 4.4. If $X_1, \ldots, X_n$ are independent, $\tau_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ does not depend on the indices $i_1, \ldots, i_k$ and the times $t_1, \ldots, t_k$, for $j \notin \{i_1, \ldots, i_k\}$. In this case, $\tau_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ coincides with the classical, univariate, reversed hazard rate function $q_j(t)$ of X_j.

The information contained in the family of the m.c.r.h.r. functions allows to analyze different type of properties of the order statistics $X_{1:n}, \ldots, X_{n:n}$ of n random variables $X_1, X_2, \ldots, X_n$. In particular, the knowledge of the m.c.r.h.r. functions will be relevant when studying the behavior of the maximum order statistic $X_{n:n}$.

Let us introduce some notations which will be useful in the following propositions. We respectively denote by $k_{(n)}, K_{(n)}, F_{(n)}, f_{(n)}$, the past intensity function (the reversed hazard rate function), the integrated past intensity function, the distribution function and the probability density function of $X_{n:n}$. Namely,

$$\begin{aligned} &k_{(n)}(t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \mathbb{P}\left(X_{n:n} \geq t - \Delta t \,|X_{n:n} \leq t\,\right), \qquad K_{(n)}(t) = \int_t^{+\infty} k_{(n)}(s) ds, \\ &F_{(n)}(t) = e^{-K_{(n)}(t)}, \qquad\qquad f_{(n)}(t) = k_{(n)}(t) e^{-K_{(n)}(t)}. \end{aligned} \tag{4.17}$$

In view of the assumption of absolute continuity and following the analogy with the definition of the m.c.h.r. functions, we can define the following limits for $j = 1, \ldots, n$

$$\delta_j(t) = \lim_{\Delta t \to 0^+} \mathbb{P}(X_j = X_{n:n} | X_{n:n} \in (t - \Delta t, t]) = \mathbb{P}(X_j = X_{n:n} | X_{n:n} = t) \tag{4.18}$$

and we notice that

$$\sum_{j=1}^{n} \delta_j(t) = 1. \tag{4.19}$$

Moreover,

$$\begin{aligned} k_{(n)}(t)\delta_j(t) &= \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(X_j = X_{n:n}, X_{n:n} \in (t - \Delta t, t])}{\mathbb{P}(X_{n:n} \in (t - \Delta t, t])} \frac{\mathbb{P}(X_{n:n} \in (t - \Delta t, t])}{\Delta t \; \mathbb{P}(X_{n:n} \leq t)} \\ &= \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(X_j > t - \Delta t, X_{n:n} \leq t)}{\Delta t \; \mathbb{P}(X_{n:n} \leq t)} = \tau_j(t|\emptyset). \end{aligned} \tag{4.20}$$

By taking into account (4.19) and (4.20), and by performing an integration from t to $+\infty$ we immediately get the following result.

Proposition 4.4. *For any $t \geq 0$ we have*

$$k_{(n)}(t) = \sum_{j=1}^{n} \tau_j(t|\emptyset), \quad K_{(n)}(t) = \int_t^{+\infty} \sum_{j=1}^{n} \tau_j(s|\emptyset) ds. \tag{4.21}$$

The role of the functions $\tau_1(t|\emptyset), \ldots, \tau_n(t|\emptyset)$ in the study of the properties of the statistic $X_{n:n}$ is described by the following result.

Proposition 4.5. *For any $t \geq 0$ and $j = 1, \ldots, n$, we have*

$$\mathbb{P}(X_j = X_{n:n}, X_{n:n} \leq t) = \int_0^t \tau_j(s|\emptyset) e^{-K_{(n)}(s)} ds.$$

Proof. Taking into account (4.17), (4.18) and (4.20) we obtain

$$\begin{aligned} \mathbb{P}(X_j = X_{n:n}, X_{n:n} \leq t) &= \int_0^t f_{(n)}(s) \mathbb{P}(X_j = X_{n:n} | X_{n:n} = s) ds \\ &= \int_0^t k_{(n)}(s) e^{-K_{(n)}(s)} \delta_j(s) ds = \int_0^t \tau_j(s|\emptyset) e^{-K_{(n)}(s)} ds, \end{aligned}$$

that is the thesis. □

As an immediate consequence of the previous proposition we note that, for $j = 1, \ldots, n$, the probability $\mathbb{P}(X_j = X_{n:n}, X_{n:n} \leq t)$ only depends on the functions $\tau_1(t|\emptyset), \ldots, \tau_n(t|\emptyset)$, and the following result follows.

Proposition 4.6. *Take n independent random variables $Z_1, \ldots, Z_n$ with reversed hazard rate functions $q_j(t)$ and let $(X_1, \ldots, X_n)$ be a vector with m.c.r.h.r. functions $\tau_j^{(\mathbf{X})}(t|\emptyset)$ such that*

$$\tau_j^{(\mathbf{X})}(t|\emptyset) = q_j(t), \; j = 1, \ldots, n. \tag{4.22}$$

Then, for any $j \in [n]$ and for any $t \geq 0$

$$\mathbb{P}(X_j = X_{n:n}, X_{n:n} \leq t) = \mathbb{P}(Z_j = Z_{n:n}, Z_{n:n} \leq t). \tag{4.23}$$

Proof. In view of independence, the m.c.r.h.r. functions $\tau_j^{(\mathbf{Z})}(t|\emptyset)$ for the vector $Z_1, \ldots, Z_n$ respectively coincide with the univariate reversed hazard rate functions $q_j(t)$. Then, the thesis follows immediately by applying Proposition 4.5 to both the vectors $(X_1, \ldots, X_n)$ and $(Z_1, \ldots, Z_n)$. $\square$

We also notice that the multivariate conditional reversed hazard rate functions $\tau_j^{(\mathbf{X})}$'s of variables $X_1, \ldots, X_n$ are strictly related to the m.c.h.r. functions $\lambda_j^{(\mathbf{Y})}$'s of the variables $Y_1 = 1/X_1, \ldots, Y_n = 1/X_n$, as stated in the following proposition.

Proposition 4.7. *Let $X_1, \ldots, X_n$ be absolutely continuous random variables and let $Y_i = 1/X_i$, for $i = 1, \ldots, n$. Then, we have*

$$\tau_j^{(\mathbf{X})}(t|\emptyset) = \frac{1}{t^2}\lambda_j^{(\mathbf{Y})}\left(\frac{1}{t}\middle|\,\emptyset\right) \tag{4.24}$$

and, for $j \notin I = \{i_1, \ldots, i_k\} \subset [n]$, $0 \le t \le t_k \le \cdots \le t_1$

$$\tau_j^{(\mathbf{X})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = \frac{1}{t^2}\lambda_j^{(\mathbf{Y})}\left(\frac{1}{t}\middle|\, i_1, \ldots, i_k; \frac{1}{t_1}, \ldots, \frac{1}{t_k}\right). \tag{4.25}$$

Proof. The stated result can be obtained as follows

$$\begin{aligned}
\tau_j^{(\mathbf{X})}(t|\emptyset) &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(X_j \ge t - \Delta t\,|X_{n:n} \le t\right) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(\frac{1}{Y_j} \ge t - \Delta t \middle| Y_{1:n} \ge \frac{1}{t}\right) \\
&= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(Y_j \le \frac{1}{t - \Delta t} \middle| Y_{1:n} \ge \frac{1}{t}\right) \\
&= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(Y_j \le \frac{1}{t} + \frac{\Delta t}{t(t - \Delta t)} \middle| Y_{1:n} \ge \frac{1}{t}\right) \\
&= \lim_{\Delta t \to 0^+} \frac{1}{t(t - \Delta t)} \cdot \frac{t(t - \Delta t)}{\Delta t}\mathbb{P}\left(Y_j \le \frac{1}{t} + \frac{\Delta t}{t(t - \Delta t)} \middle| Y_{1:n} \ge \frac{1}{t}\right) = \frac{1}{t^2}\lambda_j^{(\mathbf{Y})}\left(\frac{1}{t}\middle|\,\emptyset\right).
\end{aligned}$$

Similarly, for $j \notin I = \{i_1, \ldots, i_k\} \subset [n]$, $0 \le t \le t_k \le \cdots \le t_1$

$$\begin{aligned}
\tau_j^{(\mathbf{X})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k) &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(X_j \ge t - \Delta t \middle| X_{i_1} = t_1, \ldots, X_{i_k} = t_k, \max_{h \notin I} X_h \le t\right) \\
&= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(Y_j \le \frac{1}{t - \Delta t} \middle| Y_{i_1} = \frac{1}{t_1}, \ldots, Y_{i_k} = \frac{1}{t_k}, \min_{h \notin I} Y_h \ge \frac{1}{t}\right) \\
&= \frac{1}{t^2}\lambda_j^{(\mathbf{Y})}\left(\frac{1}{t}\middle|\, i_1, \ldots, i_k; \frac{1}{t_1}, \ldots, \frac{1}{t_k}\right).
\end{aligned}$$

$\square$

4.3.1 Reversed Load-Sharing models

We start this section by reminding the definition of inverse exponential distribution. Let us consider Y distributed as an exponential random variable, $Y \sim Exp(\lambda)$, then $X = 1/Y \sim$

$invExp(\lambda)$ is an inverse exponential random variable. For $t > 0$, the cdf, pdf and reversed hazard rate function of X are respectively given by

$$F_X(t) = \overline{F}_Y\left(\frac{1}{t}\right) = e^{-\lambda/t}, \quad f_X(t) = \frac{1}{t^2} f_Y\left(\frac{1}{t}\right) = \frac{\lambda}{t^2} e^{-\lambda/t}, \quad q_X(t) = \frac{1}{t^2} r_Y\left(\frac{1}{t}\right) = \frac{\lambda}{t^2}. \tag{4.26}$$

The inverse exponential distribution and some of its generalizations have found many key applications in several contexts, such as medicine, survival analysis of patients and of devices (see [76, 90]). The following result is related to the behavior of the maximum $X_{n:n}$ among independent variables distributed according to inverse exponential distributions.

Proposition 4.8. *Let $X_1, \dots, X_n$ be independent random variables, respectively distributed according to inverse exponential distributions with parameters $\lambda_1, \dots, \lambda_n$. Then, the following identities hold:*

$$\begin{aligned}
&\mathbb{P}(X_{n:n} = X_j, X_{n:n} \le t) = \mathbb{P}(X_{n:n} = X_j)\mathbb{P}(X_{n:n} \le t), \text{ for any } t > 0,\\
&\mathbb{P}(X_{n:n} = X_j) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i},\\
&\mathbb{P}(X_{n:n} \le t) = e^{-\frac{1}{t}\sum_{i=1}^n \lambda_i}.
\end{aligned}$$

Proof. Consider the event $(X_{n:n} \le t)$. In view of independence among variables, we have

$$\mathbb{P}(X_{n:n} \le t) = \mathbb{P}(X_1 \le t, \dots, X_n \le t) = \mathbb{P}(X_1 \le t) \cdots \mathbb{P}(X_n \le t) = e^{-\frac{1}{t}\sum_{i=1}^n \lambda_i}.$$

Let us then consider the event $(X_{n:n} = X_j, X_{n:n} \le t)$. By Proposition 4.5, we have

$$\mathbb{P}(X_{n:n} = X_j, X_{n:n} \le t) \;=\; \mathbb{P}(X_j \le t \text{ and } X_j > X_i, i \ne j) = \int_0^t \frac{\lambda_j}{s^2} e^{-\frac{\lambda_j}{s}} \mathbb{P}(X_i \le s, i \ne j) ds.$$

By taking into account the independence among $X_1, \dots, X_n$, we then obtain

$$\mathbb{P}(X_{n:n} = X_j, X_{n:n} \le t) \;=\; \int_0^t \frac{\lambda_j}{s^2} e^{-\frac{\lambda_j}{s}} \prod_{i=1, i\ne j}^n e^{-\frac{\lambda_i}{s}} ds = \int_0^t \frac{\lambda_j}{s^2} \prod_{i=1}^n e^{-\frac{\lambda_i}{s}} ds = \frac{\lambda_j e^{-\frac{1}{t}\sum_{i=1}^n \lambda_i}}{\sum_{i=1}^n \lambda_i}.$$

Hence, the events $(X_{n:n} \le t)$ and $(X_{n:n} = X_j)$ are independent and so we get

$$\mathbb{P}(X_{n:n} = X_j) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}.$$

□

Remark 4.5. We highlight that Proposition 4.8 is analogous to a well known result concerning with the minimum among variables and with exponential distributions. See e.g. Theorem 2.3.3 in [89]: Let $Y_1, \dots, Y_n$ be independent random variables such that $Y_j \sim Exp(\lambda_j)$, then

$$\begin{aligned}
&\mathbb{P}(Y_{1:n} = Y_j, Y_{1:n} > t) = \mathbb{P}(Y_{1:n} = Y_j)\mathbb{P}(Y_{1:n} > t), \text{ for any } t > 0,\\
&\mathbb{P}(Y_{1:n} = Y_j) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i},\\
&\mathbb{P}(Y_{1:n} > t) = e^{-\frac{1}{t}\sum_{i=1}^n \lambda_i}.
\end{aligned}$$

We have preferred to provide a direct proof of Proposition 4.8, even though a proof can be easily obtained from the above result. In fact, it is sufficient to remind that if $X_1, \ldots, X_n$ are independent and $X_j \sim invExp(\lambda_j)$, then $Y_1 = 1/X_1, \ldots, Y_n = 1/X_n$ are independent and $Y_j \sim Exp(\lambda_j)$ and, furthermore, the following equivalences hold

$$\begin{aligned} Y_{1:n} = Y_j &\Leftrightarrow 1/X_{n:n} = 1/X_j \Leftrightarrow X_{n:n} = X_j, \\ Y_{1:n} > t &\Leftrightarrow 1/X_{n:n} > t \Leftrightarrow X_{n:n} < \frac{1}{t}. \end{aligned}$$

Definition 4.3. We say that the random vector $(X_1, \ldots, X_n)$ is distributed according to a Reversed Load-Sharing model (RLS) if, for $i_1, \ldots, i_k \in [n]$, $j \notin I = \{i_1, \ldots, i_k\}$, the m.c.r.h.r. functions $\tau_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ does not depend on the order of $i_1, \ldots, i_k$ and on $t_1, \ldots, t_k$, for all $0 \le t \le t_k \le \cdots \le t_1$, i.e.,

$$\tau_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = \tau_j(t|I).$$

We now concentrate attention on a special subclass of reversed load sharing models. Let us consider a vector $(Y_1, \ldots, Y_n)$ distributed according to a THLS model with parameters $\mu_j(\emptyset), \mu_j(I)$. Then $(X_1, \ldots, X_n)$, defined by $X_j = 1/Y_j$, for $j = 1, \ldots, n$, is such that the m.c.r.h.r. functions are expressed, by using (4.24) and (4.25), as

$$\begin{aligned} \tau_j^{(\mathbf{X})}(t|\emptyset) &= \frac{1}{t^2}\lambda_j^{(\mathbf{Y})}\left(\frac{1}{t}\middle|\,\emptyset\right) = \frac{1}{t^2}\mu_j(\emptyset), \\ \tau_j^{(\mathbf{X})}(t|I) &= \frac{1}{t^2}\lambda_j^{(\mathbf{Y})}\left(\frac{1}{t}\middle|\,I\right) = \frac{1}{t^2}\mu_j(I). \end{aligned} \tag{4.27}$$

By recalling the formula of the reversed hazard rate in (4.26), we observe, in particular, that the reversed m.c.h.r. functions of vectors of independent, inverse-exponentially distributed random variables satisfy the identities in (4.27) and we give the following definition.

Definition 4.4. The random vector $(X_1, \ldots, X_n)$ is distributed according to a Reversed Time Homogeneous Load-Sharing model (RTHLS) if, it is a RLS and, in addition, for $I \subset [n]$ and $j \in [n] \setminus I$, the m.c.r.h.r. functions are expressed as

$$\tau_j(t|I) = \frac{c_j(I)}{t^2},$$

where $c_j(I) \ge 0$.

We emphasize that the vector $(X_1, \ldots, X_n)$ is distributed according to a RTHLS model if, and only if, the vector $(Y_1, \ldots, Y_n)$ is distributed according to a THLS model, where $Y_j = 1/X_j$, $j = 1, \ldots, n$. Furthermore, the RTHLS models can be seen as natural generalizations of the case of independent variables with inverse exponential distributions and, in particular, they inherit several remarkable properties of them. If $(X_1, \ldots, X_n)$ follows a RTHLS model, then we set, for $j \in [n]$, $I \subset [n]$, $j \notin I$,

$$N(I) = \sum_{h \notin I} c_h(I) \tag{4.28}$$

$$\eta_j(I) = \frac{\tau_j(t|I)}{\sum_{h \notin I} \tau_h(t|I)} = \frac{c_j(I)}{\sum_{h \notin I} c_h(I)} = \frac{c_j(I)}{N(I)}. \tag{4.29}$$

We note furthermore that the parameters $c_j(\emptyset), c_j(I)$ of a RTHLS model for variables $X_1, \ldots,$ X_n actually coincide with the parameters of the THLS model for the reciprocal variables $Y_j = 1/X_j$ $(j = 1, \ldots, n)$, i.e.,

$$c_j(\emptyset) = \mu_j(\emptyset), \quad c_j(I) = \mu_j(I). \tag{4.30}$$

The following result points out the appropriate way to extend Proposition 4.8 from the independent case to the case of RTHLS models.

Proposition 4.9. *Let $(X_1, \ldots, X_n)$ be distributed according to a reversed time-homogeneous load-sharing model with parameters $c_j(I)$. Then, the following identities hold:*

$$\begin{aligned}
&\mathbb{P}(X_{n:n} = X_j, X_{n:n} \le v) = \mathbb{P}(X_{n:n} = X_j)\mathbb{P}(X_{n:n} \le v), \ \textit{for any } v > 0,\\
&\mathbb{P}(X_{n:n} = X_j) = \eta_j(\emptyset),\\
&\mathbb{P}(X_{n:n} \le v) = e^{-\frac{N(\emptyset)}{v}}.
\end{aligned}$$

Proof. By applying Proposition 4.5 to the present case, we obtain

$$\begin{aligned}
\mathbb{P}(X_{n:n} = X_{\pi(n)}, X_{n:n} \le v) &= \int_0^v \tau_{\pi(n)}(s|\emptyset) e^{-\int_s^{+\infty} \sum_{i=1}^n \tau_i(w|\emptyset) dw} ds\\
= \int_0^v \frac{c_{\pi(n)}(\emptyset)}{s^2} e^{-\int_s^{+\infty} \frac{\sum_{i=1}^n c_i(\emptyset)}{w^2} dw} ds &= \int_0^v \frac{c_{\pi(n)}(\emptyset)}{s^2} e^{-\frac{\sum_{i=1}^n c_i(\emptyset)}{s}} ds\\
= \frac{c_{\pi(n)}(\emptyset)}{\sum_{i=1}^n c_i(\emptyset)} e^{-\frac{\sum_{i=1}^n c_i(\emptyset)}{v}} &= \eta_{\pi(n)}(\emptyset) e^{-\frac{N(\emptyset)}{v}}.
\end{aligned}$$

□

Now, we introduce the discrete random variables $J_1, \ldots, J_n$, where $J_h = j$ if $X_{h:n} = X_j$. Other basic aspects of RTHLS models can be better understood by writing, for $k = 1, \ldots, n$, the joint pdf $f_{X_{n:n},\ldots,X_{k:n},J_n,\ldots,J_k}(t_n, \ldots, t_k; j_n, \ldots, j_k)$ of $(X_{n:n}, \ldots, X_{k:n}; J_n, \ldots, J_k)$ with respect to the product between the k-dimensional Lebesgue measure and an appropriate counting measure. For $t_1 \le t_2 \le \cdots \le t_n$, we can write in this respect

$$f_{X_{n:n},\ldots,X_{k:n},J_n,\ldots,J_k}(t_n, \ldots, t_k; j_n, \ldots, j_k)$$

$$= f_{X_{n:n},J_n}(t_n; j_n) \times f_{X_{n-1:n},J_{n-1}}(t_{n-1}; j_{n-1}|t_n; j_n) \times \cdots \times f_{X_{k:n},J_k}(t_k; j_k|t_n, \ldots, t_{k+1}; j_n, \ldots, j_{k+1}).$$

Taking into account both the meaning of the m.c.r.h.r. functions and the definition of RTHLS models, the above equation takes the form

$$f_{X_{n:n},\ldots,X_{k:n},J_n,\ldots,J_k}(t_n,\ldots,t_k;j_n,\ldots,j_k)$$

$$= \frac{c_{j_n}(\emptyset)}{t_n^2}\exp\left\{-\int_{t_n}^{+\infty}\frac{1}{u^2}\sum_{i=1}^{n}c_i(\emptyset)\right\}du\times\frac{c_{j_{n-1}}(\{j_n\})}{t_{n-1}^2}\exp\left\{-\int_{t_{n-1}}^{t_n}\frac{1}{u^2}\left(\sum_{i\in[n]\setminus\{j_n\}}c_i(\{j_n\})\right)du\right\}$$

$$\times\cdots\times\frac{c_{j_k}(\{j_n,\ldots,j_{k+1}\})}{t_k^2}\exp\left\{-\int_{t_k}^{t_{k+1}}\frac{1}{u^2}\left(\sum_{i\in[n]\setminus\{j_n,\ldots,j_{k+1}\}}^{n}c_i(\{j_n,\ldots,j_{k+1}\})\right)du\right\}. \quad (4.31)$$

Hence, by using (4.28), we have

$$f_{X_{n:n},\ldots,X_{k:n},J_n,\ldots,J_k}(t_n,\ldots,t_k;j_n,\ldots,j_k)$$

$$= \frac{c_{j_n}(\emptyset)}{t_n^2}\exp\left\{-\int_{t_n}^{+\infty}\frac{1}{u^2}N(\emptyset)du\right\}\times\frac{c_{j_{n-1}}(\{j_n\})}{t_{n-1}^2}\exp\left\{-\int_{t_{n-1}}^{t_n}\frac{1}{u^2}N\left(\{j_n\}\right)du\right\}\times\ldots$$

$$\times\frac{c_{j_k}(\{j_n,\ldots,j_{k+1}\})}{t_k^2}\exp\left\{-\int_{t_k}^{t_{k+1}}\frac{1}{u^2}N\left(\{j_n,\ldots,j_{k+1}\}\right)du\right\}. \quad (4.32)$$

4.3.2 Applications to the Inactivity Times of coherent systems

Let S be a coherent system whose components lifetimes $X_1,\ldots,X_n$ are jointly distributed according to a RTHLS model, let T_S be the its lifetime and $\hat{T}_{v,S}=v-T_S$ its inactivity time at time v. The purpose of this section is to compute the conditional probability

$$\mathbb{P}(\hat{T}_{v,S}\geq t|X_{n:n}\leq v).$$

Namely, we look for the conditional distribution of the inactivity time of the system, conditional on the detailed information that all the components are down at time v. In this perspective, we will in particular employ the following results which are respectively dual to results valid for the ordinary THLS models, as presented in [110] or to results presented in [33].

First of all notice that, in view of Proposition 4.9, the conditional distribution of the maximum order statistic $X_{n:n}$ given the event $(X_{n:n}=X_j)$ coincides with an inverse exponential distribution whose parameter is $N(\emptyset)$. More precisely we can state the following result.

Proposition 4.10. *We have, for any $v>t>0$ and for any $j\in[n]$*

$$\mathbb{P}(v-X_{n:n}\geq t|X_{n:n}=X_j,X_{n:n}\leq v)=\exp\left(-\frac{t\ N(\emptyset)}{v(v-t)}\right). \quad (4.33)$$

Proof. From Proposition 4.5 and Equations (4.28)–(4.29), we have

$$\begin{aligned}
&\mathbb{P}(v - X_{n:n} \geq t | X_{n:n} = X_j, X_{n:n} \leq v) = \mathbb{P}(X_{n:n} \leq v - t | X_{n:n} = X_j, X_{n:n} \leq v) \\
&= \frac{\mathbb{P}(X_{n:n} \leq v - t, X_{n:n} = X_j)}{\mathbb{P}(X_{n:n} \leq v, X_{n:n} = X_j)} = \frac{\int_0^{v-t} \tau_j(s|\emptyset) e^{-\int_s^{+\infty} \sum_{h=1}^n \tau_h(w|\emptyset)dw} ds}{\int_0^{v} \tau_j(s|\emptyset) e^{-\int_s^{+\infty} \sum_{h=1}^n \tau_h(w|\emptyset)dw} ds} \\
&= \frac{\eta_j(\emptyset) \exp\left(-\frac{N(\emptyset)}{v-t}\right)}{\eta_j(\emptyset) \exp\left(-\frac{N(\emptyset)}{v}\right)} = \exp\left(-\frac{t\, N(\emptyset)}{v(v-t)}\right),
\end{aligned}$$

which completes the proof of (4.33). □

Proposition 4.11. *Let $(X_1, \dots, X_n)$ be distributed according to a reversed time homogeneous load-sharing model with parameters $c_j(I)$, $I \subset [n], j \in [n] \setminus I$. Let us fix $v > 0$. We have for $k = 1, \dots, n$,*

$$\begin{aligned}
&\mathbb{P}(X_{n:n} = X_{j_n}, X_{n-1:n} = X_{j_{n-1}}, \dots, X_{k:n} = X_{j_k}, X_{n:n} \leq v) \\
&= \eta_{j_n}(\emptyset)\eta_{j_{n-1}}(\{j_n\}) \cdots \eta_{j_k}(\{j_n, j_{n-1}, \dots, j_{k+1}\}) \exp\left\{-\frac{N(\emptyset)}{v}\right\}.
\end{aligned}$$

Proof. By plugging the identity

$$\int_a^b \frac{A}{u^2} du = A\left(\frac{1}{a} - \frac{1}{b}\right),$$

for $0 < a < b$ and $A > 0$, within formula (4.32), we can write

$$\begin{aligned}
f_{X_{n:n},\dots,X_{1:n},J_n,\dots,J_1}(t_n, \dots, t_1; j_n, \dots, j_1) &= \frac{c_{j_n}(\emptyset) \cdot c_{j_{n-1}}(\{j_n\}) \cdot \ldots \cdot c_{j_1}(\{j_n, \dots, j_2\})}{t_n^2 \cdot t_{n-1}^2 \cdot \ldots \cdot t_1^2} \times \\
\times \exp\left\{-\left[N(\emptyset)\frac{1}{t_n} + N(\{j_n\})\left(\frac{1}{t_{n-1}} - \frac{1}{t_n}\right) + \dots + N(\{j_n, \dots, j_2\})\left(\frac{1}{t_1} - \frac{1}{t_2}\right)\right]\right\} \\
&= \frac{c_{j_n}(\emptyset) \cdot c_{j_{n-1}}(\{j_n\}) \cdot \ldots \cdot c_{j_1}(\{j_n, \dots, j_2\})}{t_n^2 \cdot t_{n-1}^2 \cdot \ldots \cdot t_1^2} \times \\
\times \exp\Bigg\{-\Bigg[\frac{1}{t_n}\left[N(\emptyset) - N(\{j_n\})\right] + \frac{1}{t_{n-1}}\left[N(\{j_n\}) - N(\{j_n, j_{n-1}\})\right] + \dots \\
+ \frac{1}{t_2}\left[N(\{j_n, \dots, j_3\}) - N(\{j_n, \dots, j_2\})\right] + \frac{1}{t_1} N(\{j_n, \dots, j_2\})\Bigg]\Bigg\}.
\end{aligned}$$

By properly integrating the joint density function of $(X_{n:n}, \dots, X_{1:n}; J_n, \dots, J_1)$ over the appropriate domain, we get

$$\mathbb{P}\left(X_{n:n} = X_{j_n}, \dots, X_{k:n} = X_{j_k}, X_{n:n} \leq v\right)$$

$$= c_{j_n}(\emptyset) \cdot \ldots \cdot c_{j_k}(\{j_n, \dots, j_{k+1}\}) \sum_{j_1 \neq \dots \neq j_{k-1} \neq j_k \neq j_{k+1} \neq \dots \neq j_n} c_{j_{k-1}}(\{j_n, \dots, j_k\}) \cdot \ldots \cdot c_{j_1}(\{j_n, \dots, j_2\}) \times$$

$$\times \int_0^v dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} \frac{1}{t_n^2 \cdot t_{n-1}^2 \cdot \ldots \cdot t_1^2} \exp\left\{ -\left[\frac{1}{t_n} [N(\emptyset) - N(\{j_n\})] + \right.\right.$$

$$\left.\left. \cdots + \frac{1}{t_2}[N(\{j_n, \ldots, j_3\}) - N(\{j_n, \ldots, j_2\})] + \frac{1}{t_1} N(\{j_n, \ldots, j_2\}) \right] \right\} dt_1$$

$$= c_{j_n}(\emptyset) \cdot \ldots \cdot c_{j_k}(\{j_n, \ldots, j_{k+1}\}) \sum_{j_1 \neq \ldots \neq j_{k-1} \neq j_k \neq j_{k+1} \neq \ldots \neq j_n} c_{j_{k-1}}(\{j_n, \ldots, j_k\}) \cdot \ldots \cdot c_{j_1}(\{j_n, \ldots, j_2\}) \times$$

$$\int_0^v \frac{\exp\left\{ -\left[\frac{1}{t_n} [N(\emptyset) - N(\{j_n\})] \right] \right\}}{t_n^2} dt_n \cdots \int_0^{t_3} \frac{\exp\left\{ -\left[\frac{1}{t_2} [N(\{j_n, \ldots, j_3\}) - N(\{j_n, \ldots, j_2\})] \right] \right\}}{t_2^2} \times$$

$$\times \int_0^{t_2} \frac{\exp\left\{ -\frac{1}{t_1} N(\{j_n, \ldots, j_2\}) \right\}}{t_1^2} dt_1.$$

Now, by taking into account the identity

$$\int_0^{t_2} \frac{\exp\left\{ -\frac{1}{t_1} N(\{j_n, \ldots, j_2\}) \right\}}{t_1^2} dt_1 = \frac{\exp\left\{ -\frac{1}{t_2} N(\{j_n, \ldots, j_2\}) \right\}}{N(\{j_n, \ldots, j_2\})},$$

we obtain

$$\mathbb{P}\left(X_{n:n} = X_{j_n}, \ldots, X_{k:n} = X_{j_k}, X_{n:n} \leq v \right)$$

$$= c_{j_n}(\emptyset) \cdot \ldots \cdot c_{j_k}(\{j_n, \ldots, j_{k+1}\}) \sum_{j_1 \neq \ldots \neq j_{k-1} \neq j_k \neq j_{k+1} \neq \ldots \neq j_n} c_{j_{k-1}}(\{j_n, \ldots, j_k\}) \cdot \ldots \cdot c_{j_1}(\{j_n, \ldots, j_2\}) \times$$

$$\times \int_0^v \frac{\exp\left\{ -\left[\frac{1}{t_n} [N(\emptyset) - N(\{j_n\})] \right] \right\}}{t_n^2} dt_n \cdots \int_0^{t_3} \frac{\exp\left\{ -\frac{1}{t_2} [N(\{j_n, \ldots, j_2\}) \right\}}{N(\{j_n, \ldots, j_2\}) \cdot t_2^2} dt_2.$$

Continuing so on, it follows

$$\mathbb{P}\left(X_{n:n} = X_{j_n}, \ldots, X_{k:n} = X_{j_k}, X_{n:n} \leq v \right) = \frac{c_{j_n}(\emptyset) \cdot \ldots \cdot c_{j_k}(\{j_n, \ldots, j_{k+1}\})}{N(\{j_n\}) \cdot \ldots \cdot N(\{j_n, \ldots, j_{k+1}\})} \times$$

$$\times \sum_{j_1 \neq \ldots \neq j_{k-1} \neq j_k \neq j_{k+1} \neq \ldots \neq j_n} \frac{c_{j_{k-1}}(\{j_n, \ldots, j_k\}) \cdot \ldots \cdot c_{j_1}(\{j_n, \ldots, j_2\})}{N(\{j_n, \ldots, j_k\}) \cdot \ldots \cdot N(\{j_n, \ldots, j_2\})} \times$$

$$\times \int_0^v \frac{\exp\left\{ -\left[\frac{1}{t_n} [N(\emptyset) - N(\{j_n\})] \right] \right\} \exp\left\{ -\frac{1}{t_n} N(\{j_n\}) \right\}}{t_n^2} dt_n$$

$$= c_{j_n}(\emptyset) \cdot \eta_{j_{n-1}}(\{j_n\}) \cdot \ldots \cdot \eta_{j_k}(\{j_n, \ldots, j_{k+1}\}) \frac{1}{N(\emptyset)} \exp\left\{ -\frac{N(\emptyset)}{v} \right\}$$

$$= \eta_{j_n}(\emptyset) \cdot \eta_{j_{n-1}}(\{j_n\}) \cdot \ldots \cdot \eta_{j_k}(\{j_n, \ldots, j_{k+1}\}) \exp\left\{ -\frac{N(\emptyset)}{v} \right\}.$$

□

Remark 4.6. As pointed out in [33], an important property of THLS models is the following: conditioning on the event $\{Y_{1:n} > t\}$ the joint distribution of the residual lifetimes $Y_j - t$, $j = 1, \ldots, n$, is the same of the original variables. By using this property, a different proof of Proposition 4.11 based on THLS models is presented in [23].

Denote by $\overline{G}_{\lambda_1,\dots,\lambda_r}$ the survival function of the distribution obtained as convolution of r exponential distributions with parameters $\lambda_1,\dots,\lambda_r$, respectively. Also the next result can easily be obtained by resorting to THLS models by recalling the notation introduced in (4.28).

Proposition 4.12. *Let $(X_1,\dots,X_n)$ be distributed according to a reversed time-homogeneous load-sharing model. We have, for any $v>u>0$ and $k=1,\dots,n$,*

$$\mathbb{P}\left(X_{k:n}<u|X_{n:n}=X_{j_n},\dots,X_{k:n}=X_{j_k},X_{n:n}\le v\right)=\overline{G}_{N(\emptyset),\dots,N(\{j_n,\dots,j_{k+1}\})}\left(\frac{1}{u}-\frac{1}{v}\right). \tag{4.34}$$

Proof. Let us consider the variables $Y_j=1/X_j$, $j=1,\dots,n$. Then, $(Y_1,\dots,Y_n)$ follows a THLS model with the same parameters of the RTHLS model associated to $(X_1,\dots,X_n)$. Thus, we have

$$\begin{aligned}
&\mathbb{P}\left(X_{k:n}<u|X_{n:n}=X_{j_n},\dots,X_{k:n}=X_{j_k},X_{n:n}\le v\right)\\
&=\mathbb{P}\left(Y_{n-k+1:n}>\frac{1}{u}\middle|Y_{1:n}=Y_{j_n},\dots,Y_{n-k+1:n}=Y_{j_k},Y_{1:n}>\frac{1}{v}\right)\\
&=\mathbb{P}\left(Y_{n-k+1:n}>\frac{1}{u}-\frac{1}{v}\middle|Y_{1:n}=Y_{j_n},\dots,Y_{n-k+1:n}=Y_{j_k}\right)=\overline{G}_{N(\emptyset),\dots,N(\{j_n,\dots,j_{k+1}\})}\left(\frac{1}{u}-\frac{1}{v}\right),
\end{aligned}$$

where the last equality follows by Proposition 4 of [110]. □

Several properties about reliability characteristics of coherent systems can be obtained by assuming a THLS model for the components' lifetimes. In particular a special formula is obtained for the computation of the survival function of the lifetime T_S of a given system S, in terms of appropriate convolutions of exponential distributions (see [110]). In the following, we will show a dual result under the assumption of a RTHLS model.

Let us consider a system formed by n components $C_1,\dots,C_n$, whose lifetimes are non-negative random variables $X_1,\dots,X_n$. Assume the joint probability distribution of $X_1,\dots,X_n$ is absolutely continuous and so ties among $X_1,\dots,X_n$ have probability zero. Let us denote by T_S the lifetime of the system and by $\hat{T}_{v,S}$ the inactivity time at time v, namely $\hat{T}_{v,S}=v-T_S$. Let $\mathcal{P}_n$ be the set of permutation of $\{1,\dots,n\}$ and let B_k be the subset of $\mathcal{P}_n$ composed with the elements π such that the event $\{X_{n:n}=X_{\pi(n)},\dots,X_{k:n}=X_{\pi(k)}\}$ implies that the system fails at the k-th failure at component level, i.e.

$$B_k=\{\pi\in\mathcal{P}_n:\ \text{if } X_{n:n}=X_{\pi(n)},\dots,X_{k:n}=X_{\pi(k)} \text{ then } E_k\},$$

where, for $k=1,\dots,n$, E_k is the event $E_k=\{T_S=X_{k:n}\}$. The events E_k are strictly related to the structure of the system. They are also connected with the concepts of signature (Definition 1.13) and dual signature (see [100] for further details).

Proposition 4.13. *Let S be a system formed by n components whose lifetimes are non-negative random variables $X_1, \ldots, X_n$ distributed according to a reversed time-homogeneous load-sharing model and let $\hat{T}_{v,S}$ be the inactivity time of the system at time v. We have, for $0 < t < v$,*

$$\begin{aligned}
\mathbb{P}(\hat{T}_{v,S} \geq t | X_{n:n} \leq v) \quad = \quad & \sum_{k=1}^{n} \sum_{\pi \in B_k} \overline{G}_{N(\emptyset), \ldots, N(\{\pi(n), \ldots, \pi(k+1)\})} \left(\frac{t}{v(v-t)} \right) \cdot \\
& \cdot \eta_{\pi(n)}(\emptyset) \cdots \eta_{\pi(2)}(\{\pi(n), \pi(n-1), \ldots, \pi(3)\}).
\end{aligned}$$

Proof. Taking into account that $\{B_1, \ldots, B_n\}$ is a partition of $\mathcal{P}_n$, we write

$$\begin{aligned}
\mathbb{P}(\hat{T}_{v,S} \geq t | X_{n:n} \leq v) = & \sum_{\pi \in \mathcal{P}_n} \mathbb{P}(\hat{T}_{v,S} \geq t | X_{n:n} = X_{\pi(n)}, \ldots, X_{1:n} = X_{\pi(1)}, X_{n:n} \leq v) \cdot \\
& \cdot \mathbb{P}(X_{n:n} = X_{\pi(n)}, \ldots, X_{1:n} = X_{\pi(1)} | X_{n:n} \leq v) \\
= & \sum_{k=1}^{n} \sum_{\pi \in B_k} \mathbb{P}(T_S \leq v - t | X_{n:n} = X_{\pi(n)}, \ldots, X_{1:n} = X_{\pi(1)}, X_{n:n} \leq v) \cdot \\
& \cdot \mathbb{P}(X_{n:n} = X_{\pi(n)}, \ldots, X_{1:n} = X_{\pi(1)}) \\
= & \sum_{k=1}^{n} \sum_{\pi \in B_k} \mathbb{P}(X_{k:n} \leq v - t | X_{n:n} = X_{\pi(n)}, \ldots, X_{1:n} = X_{\pi(1)}, X_{n:n} \leq v) \cdot \\
& \cdot \mathbb{P}(X_{n:n} = X_{\pi(n)}, \ldots, X_{1:n} = X_{\pi(1)}) \\
= & \sum_{k=1}^{n} \sum_{\pi \in B_k} \overline{G}_{N(\emptyset), \ldots, N(\{\pi(n), \ldots, \pi(k+1)\})} \left(\frac{t}{v(v-t)} \right) \cdot \\
& \cdot \eta_{\pi(n)}(\emptyset) \cdots \eta_{\pi(2)}(\{\pi(n), \pi(n-1), \ldots, \pi(3)\}).
\end{aligned}$$

□

In the following, based on the result of Proposition 4.13, we give an example of evaluation of the probability distributions of the inactivity times for two different systems.

Example 4.3. Let us consider a coherent system S with three components X_1, X_2, X_3. The structure of the system in displayed in Figure 4.1 and its lifetime T_S is described as

$$T_S = \max\{X_1, \min\{X_2, X_3\}\}.$$

Let X_1, X_2, X_3 be distributed according to a RTHLS model with the parameters given as follows: for $j = 1, 2, 3$ and $i \neq j$

$$c_j(\emptyset) = 1, \quad c_j(\{i\}) = 1, \quad c_2(\{1,3\}) = c_3(\{1,2\}) = 1, \quad c_1(\{2,3\}) = \varepsilon,$$

where ε is a positive number, close to 0.

We want to apply the result of Proposition 4.13 to evaluate the distribution of the inactivity time of the system. In order to do this, we need to establish how the partition $\{B_1, B_2, B_3\}$ of

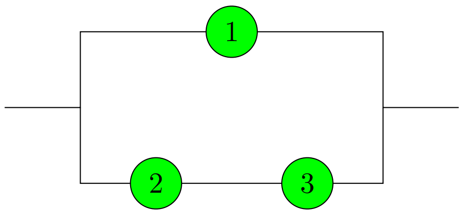

Figure 4.1: The structure of the system S in Example 4.3.

$\mathcal{P}_3$ is composed. Here, we have

$$B_1 = \emptyset, \quad B_2 = \{(1,2,3),(1,3,2),(2,1,3),(3,1,2)\}, \quad B_3 = \{(2,3,1),(3,2,1)\}.$$

Then, about the inactivity time of the system, we have

$$\begin{aligned} \mathbb{P}(\hat{T}_{v,S} \geq t | X_{n:n} \leq v) &= \sum_{k=2}^{3} \sum_{\pi \in B_k} \overline{G}_{N(\emptyset),\ldots,N(\{\pi(3),\ldots,\pi(k+1)\})} \left(\frac{t}{v(v-t)} \right) \eta_{\pi(3)}(\emptyset) \eta_{\pi(2)}(\{\pi(3)\}) \\ &= \sum_{\pi \in B_2} \overline{G}_{N(\emptyset),N(\{\pi(3)\})} \left(\frac{t}{v(v-t)} \right) \eta_{\pi(3)}(\emptyset) \eta_{\pi(2)}(\{\pi(3)\}) + \\ &\quad + \sum_{\pi \in B_3} \overline{G}_{N(\emptyset)} \left(\frac{t}{v(v-t)} \right) \cdot \eta_{\pi(3)}(\emptyset) \eta_{\pi(2)}(\{\pi(3)\}). \end{aligned}$$

By recalling the identities (4.28)–(4.29), the related coefficients are described as follows. Regardless of $\pi \in \mathcal{P}_3$, we have

$$N(\emptyset) = 3, \quad N(\{\pi(3)\}) = 2, \quad \eta_{\pi(3)}(\emptyset) = 1/3, \quad \eta_{\pi(2)}(\{\pi(3)\}) = 1/2.$$

Then, we conclude

$$\begin{aligned} \mathbb{P}(\hat{T}_{v,S} \geq t | X_{n:n} \leq v) &= 4\overline{G}_{3,2} \left(\frac{t}{v(v-t)} \right) \cdot \frac{1}{6} + 2\overline{G}_3 \left(\frac{t}{v(v-t)} \right) \cdot \frac{1}{6} \\ &= \frac{2}{3} \overline{G}_{3,2} \left(\frac{t}{v(v-t)} \right) + \frac{1}{3} \exp \left(-3 \frac{t}{v(v-t)} \right). \end{aligned}$$

Concerning the computation of the convolution of several exponential distributions one can of course resort to a quite wide literature, see e.g. [1] and references therein.

With the same RTHLS model described above for the components' lifetimes, we now switch to the system $\widetilde{S}$ dual of S (see Figure 4.2 for the structure), whose lifetime is

$$T_{\widetilde{S}} = \min\{X_1, \max\{X_2, X_3\}\}.$$

In this case, the partition $\{B_1, B_2, B_3\}$ of $\mathcal{P}_3$ is given by

$$B_1 = \{(1,2,3),(1,3,2)\}, \quad B_2 = \{(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}, \quad B_3 = \emptyset.$$

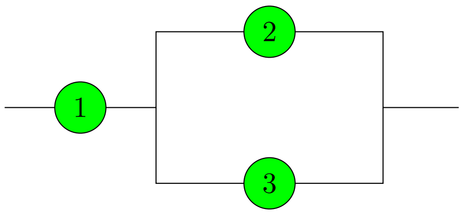

Figure 4.2: The structure of the system $\widetilde{S}$ in Example 4.3.

Hence, about the inactivity time of the system $\widetilde{S}$, we have

$$\begin{aligned}\mathbb{P}(\hat{T}_{v,\widetilde{S}} \geq t | X_{n:n} \leq v) &= \sum_{\pi \in B_1} \overline{G}_{N(\emptyset),N(\{\pi(3)\}),N(\{\pi(3),\pi(2)\})}\left(\frac{t}{v(v-t)}\right)\eta_{\pi(3)}(\emptyset)\eta_{\pi(2)}(\{\pi(3)\}) + \\ &\quad + \sum_{\pi \in B_2} \overline{G}_{N(\emptyset),N(\{\pi(3)\})}\left(\frac{t}{v(v-t)}\right) \cdot \eta_{\pi(3)}(\emptyset)\eta_{\pi(2)}(\{\pi(3)\}).\end{aligned}$$

The parameters of the form $N(\emptyset)$ and $N(\{i\})$ $(i = 1, 2, 3)$ have already been computed above. For what concerns the parameters of the form $N(\{i_1, i_2\})$, with $i_1 \neq i_2$, the structure of this system entails that we only need to consider $N(\{2, 3\}) = \varepsilon$. Then, we get

$$\begin{aligned}\mathbb{P}(\hat{T}_{v,\widetilde{S}} \geq t | X_{n:n} \leq v) &= 2\overline{G}_{3,2,\varepsilon}\left(\frac{t}{v(v-t)}\right) \cdot \frac{1}{6} + 4\overline{G}_{3,2}\left(\frac{t}{v(v-t)}\right) \cdot \frac{1}{6} \\ &= \frac{1}{3}\overline{G}_{3,2,\varepsilon}\left(\frac{t}{v(v-t)}\right) + \frac{2}{3}\overline{G}_{3,2}\left(\frac{t}{v(v-t)}\right).\end{aligned}$$

Chapter 5

Aging intensity functions

The aging intensity function is a useful tool to describe reliability properties of a random variable. It is a function described in terms of the hazard rate, the probability density and the survival functions and it can be applied in order to make comparisons among different distributions. The aging intensity function does not characterize uniquely the distribution and, in order to overcome this issue, a generalization has been introduced. Moreover, it is possible to study and analyze a function dual to the aging intensity function, namely the reversed aging intensity function, based on the reversed hazard rate function. The aim of this chapter is twofold. First, based on the results in Buono, Longobardi and Szymkowiak [27], a generalization of the reversed aging intensity function in introduced and studied. Then, by using the m.c.h.r. functions studied in Chapter 4, an extension of the (univariate) aging intensity function to the multivariate case is provided as presented in Buono [22].

5.1 Aging intensity and reversed aging intensity functions

The notion of aging intensity (AI) function has been introduced in [56] as the ratio of instantaneous failure rate, or hazard rate, and a baseline failure rate. For a non-negative and absolutely continuous random variable X with probability density function (pdf) f, survival function $\overline{F}$ and hazard rate function r, the Aging Intensity (AI) function is defined as

$$L(t) = \frac{r(t)}{\frac{1}{t}\int_0^t r(x)dx} = \frac{-tf(t)}{\overline{F}(t)\log \overline{F}(t)}, \tag{5.1}$$

i.e., $L(t)$ is the ratio of the instantaneous failure rate $r(t)$ to the average failure rate in the interval $(0,t)$ and expresses the units average aging behavior. It analyzes the aging property quantitatively, in the sense that the larger the aging intensity, the stronger the tendency of aging. The survival function and the aging intensity function are strictly related each other. More precisely, the failure rate function, or the survival function, uniquely determines the AI function but not conversely. The AI function of a non-negative random variable determines a family of survival functions through the relation presented in the following theorem by [113].

Theorem 5.1. *Let X be a non-negative and absolutely continuous random variable with survival function $\overline{F}$ and aging intensity function L. Then, $\overline{F}$ and L are related, for all $a \in (0, +\infty)$, by the relation*

$$\overline{F}(t) = \exp\left[\log k \exp\left(\int_a^t \frac{L(x)}{x} dx\right)\right], \quad t \in (0, +\infty),$$

where $k = \overline{F}(a)$. Moreover, a non-negative function L defined on $(0, +\infty)$, and such that, for a fixed $a \in (0, +\infty)$, $\lim_{t\to 0^+} \int_t^a \frac{L(x)}{x} dx = +\infty$, $\lim_{t\to +\infty} \int_a^t \frac{L(x)}{x} dx = +\infty$, determines a family of absolutely continuous survival functions by the relation

$$\overline{F}_k(t) = \exp\left[\log k \exp\left(\int_a^t \frac{L(x)}{x} dx\right)\right], \quad t \in (0, +\infty),$$

by varying the parameter $k \in (0, 1)$ and it is the aging intensity function for such survival functions.

There are several families of distributions in which the parameter k reduces to be one of the model. For instance, if X follows the Weibull distribution, $X \sim W2(\alpha, \lambda)$, with survival function $\overline{F}(t) = \exp(-\lambda t^\alpha)$, $t > 0$, then the AI function is constant and expressed as $L(t) = \alpha$. Then, $L(t) = \alpha$ determines the subfamily of the family of the Weibull distributions with fixed parameter α and varying parameter $\lambda > 0$. Based on these considerations, it is possible to use the shape of an estimated aging intensity function in order to discover the underlying distribution of some data. A survey of characterization results based on AI functions for different types of Weibull distributions is presented in [49].

The evaluation of the AI function for some well known models is presented in [56] where it is also introduced the notion of average aging intensity in order to study models characterized by quasi-constant failure rate. Some properties of AI functions are presented in [79] where, in particular, a new stochastic order (aging intensity order) based on the AI functions is defined. More precisely, by observing that the larger the aging intensity, the stronger the tendency to aging, a random variable X is said to be smaller than another random variable Y in the AI order, denoted by $X \leq_{AI} Y$, if $L_X(t) \geq L_Y(t)$, for all $t > 0$. In the following theorem, some equivalent conditions to the definition of aging intensity order are given.

Theorem 5.2. *Let X and Y be non-negative and absolutely continuous random variables. Then, the following conditions are equivalent:*

(1) $X \leq_{AI} Y$;

(2) the ratio $\frac{\int_0^t r_X(x)dx}{\int_0^t r_Y(x)dx}$ is increasing in $t > 0$;

(3) the ratio $\frac{\log \overline{F}_X(t)}{\log \overline{F}_Y(t)}$ *is increasing in* $t > 0$.

Furthermore, another important property of the aging intensity order is given by the fact that if X is IFRA and Y is DFRA then $X \leq_{AI} Y$.

The study of the aging intensity functions has been extended to some simply systems by preserving the assumption of independence. In [18], the authors have proved that if X is the lifetime of a series system formed by n independent components, then the aging intensity function of X satisfies

$$\min_{1 \leq i \leq n} L_{X_i}(t) \leq L_X(t) \leq \max_{1 \leq i \leq n} L_{X_i}(t),$$

where $L_{X_i}(\cdot)$ is the AI function of the i-th component. About parallel systems, they proved that if X and Y are the lifetimes of parallel systems with n and m independent and identically distributed components then, for $n > m$, $X \leq_{AI} Y$. However, the possibility of considering a mode of dependence among components is not yet foreseen and, in this perspective, the necessity of defining a more general form of aging intensity functions emerges.

Another kind of aging intensity is known as reversed aging intensity (RAI) function, it shares some properties with the aging intensity function and is based on the reversed hazard rate function. The reversed aging intensity function $\breve{L}(t)$ is defined, for $t > 0$, as follows [96]

$$\breve{L}(t) = \frac{-tf(t)}{F(t)\log F(t)} = \frac{-tq(t)}{\log F(t)}, \tag{5.2}$$

and it is the analogous for the future of the aging intensity function. The reversed aging intensity function can be expressed also in a different way by observing that the cumulative reversed hazard rate function defined as

$$\breve{R}(t) = \int_t^{+\infty} q(x)dx = \log F(t)\big|_{x=t}^{x\to+\infty} = -\log F(t), \tag{5.3}$$

can be treated as the total amount of failures accumulated after the time point t. Note that this function has been already introduced in (2.55) with a different notation. Here we prefer to use this different notation to be consistent with the other functions involved in this context. So $\breve{H}(t) = \frac{1}{t}\breve{R}(t)$, being the proportion between the total amount of failures accumulated after the time point t and the time t for which the unit has still survived, can be considered as the baseline value of the reversed hazard rate. Then, (5.2) can be written as

$$\breve{L}(t) = \frac{tq(t)}{\breve{R}(t)} = \frac{q(t)}{\breve{H}(t)},$$

and so the reversed aging intensity function, defined as the ratio of the instantaneous reversed hazard rate q to the baseline value of the reversed hazard rate $\breve{H}$, expresses the units average aging behavior: the higher the reversed aging intensity (it means the higher the instantaneous reversed hazard rate, and the smaller the total amount of failures accumulated after the time point t), the weaker the tendency of aging. By using the reversed aging intensity function it is possible to obtain a result analogous to Theorem 5.1, in the sense that there is a relation between the cumulative distribution function and the reversed aging intensity function and a reversed aging intensity function determines a family of absolutely continuous distributions.

The aging intensity function of X and the reversed aging intensity function of $1/X$ are strictly connected by the following relation

$$L_X\left(\frac{1}{t}\right) = \breve{L}_{\frac{1}{X}}(t). \tag{5.4}$$

Moreover, in analogy with the aging intensity order, in [96] it is defined an order based on the comparison among reversed aging intensity functions, namely the reversed aging intensity order. A random variable X is said to be less than or equal to Y in the reversed aging intensity (RAI) order, $X \leq_{RAI} Y$, if $\breve{L}_X(t) \leq \breve{L}_Y(t)$ for all $t > 0$. In analogy with Theorem 5.2, there are equivalent conditions to this definition which are related to the cumulative reversed hazard rates and cumulative distribution functions. The aging intensity order and the reversed aging intensity order are strictly connected by the following theorem.

Theorem 5.3. *Let X and Y be non-negative and absolutely continuous random variables. Then, $\frac{1}{X} \geq_{RAI} \frac{1}{Y} \iff X \leq_{AI} Y$.*

5.1.1 Generalized aging intensity functions

It is important to remark that the aging intensity function and the reversed aging intensity function do not determine uniquely the distribution. In order to overcome this problem, a new version of aging intensity, known as generalized aging intensity function, has been defined and studied by Szymkowiak [114].

Let X be a non-negative and absolutely continuous random variable with cdf F and pdf f, and let G be a strictly increasing cdf with pdf g. The G–generalized aging intensity function of X is defined as

$$L_{G,F}(t) = \frac{tf(t)}{g(G^{-1}(F(t)))G^{-1}(F(t))}.$$

A very interesting case, because it provides intuitive results, is the one in which the distribution function G is the distribution function of a generalized Pareto distribution.

Definition 5.1. A random variable X_α follows a generalized Pareto distribution with parameter $\alpha \in \mathbb{R}$ if the distribution function W_α is expressed as (see [92]):

$$W_\alpha(t) = \begin{cases} 1-(1-\alpha t)^{\frac{1}{\alpha}}, & \text{for } \begin{cases} t>0, & \text{if } \alpha<0 \\ 0<t<\frac{1}{\alpha}, & \text{if } \alpha>0 \end{cases} \\ 1-\exp(-t), & \text{for } t>0 \text{ if } \alpha=0. \end{cases}$$

Note that for $\alpha = 0$ it is the distribution function of an exponential random variable with parameter 1. From the distribution function it is possible to obtain the quantile and the density function as

$$W_\alpha^{-1}(t) = \begin{cases} \frac{1}{\alpha}[1-(1-t)^\alpha], & \text{for } 0<t<1, \text{ if } \alpha \neq 0 \\ -\log(1-t), & \text{for } 0<t<1, \text{ if } \alpha = 0 \end{cases}$$

$$w_\alpha(t) = \begin{cases} (1-\alpha t)^{\frac{1-\alpha}{\alpha}}, \quad \text{for } \begin{cases} t>0, & \text{if } \alpha<0 \\ 0<t<\frac{1}{\alpha}, & \text{if } \alpha>0 \end{cases} \\ \exp(-t), \qquad \text{for } t>0, \text{ if } \alpha=0. \end{cases}$$

By using the generalized Pareto distribution, it is possible to define an interesting class of generalized aging intensity functions. They are known as α–generalized aging intensity functions and are defined by

$$L_\alpha(t) = \begin{cases} \frac{\alpha t(1-F(t))^{\alpha-1}f(t)}{1-(1-F(t))^\alpha}, & \text{for } t>0, \text{ if } \alpha \neq 0 \\ \frac{-tf(t)}{(1-F(t))\log(1-F(t))}, & \text{for } t>0, \text{ if } \alpha = 0. \end{cases} \tag{5.5}$$

As mentioned above, those functions are of great interest since, for $\alpha > 0$, they determine the distribution function uniquely as stated in the following theorem by Szymkowiak [114].

Theorem 5.4. *Let X be a non-negative and absolutely continuous random variable with cdf F and α–generalized aging intensity function L_α, with $\alpha > 0$. Then, F and L_α are related by the relation*

$$F(t) = 1 - \left[1 - \exp\left(-\int_t^{+\infty} \frac{L_{\alpha,F}(x)}{x} dx\right)\right]^{\frac{1}{\alpha}}, \quad t \in (0,+\infty).$$

Moreover, a non-negative function L defined on $(0,+\infty)$, and such that, for a fixed $a \in (0,+\infty)$, $\lim_{t\to 0^+} \int_t^a \frac{L(x)}{x}dx = +\infty$, $\lim_{t\to+\infty} \int_a^t \frac{L(x)}{x}dx < +\infty$, determines, for $\alpha > 0$, a unique absolutely continuous distribution function by the relation

$$\begin{aligned} F(t) &= W_\alpha\left(\frac{1}{\alpha}\exp\left(-\int_t^{+\infty}\frac{L(x)}{x}dx\right)\right) \\ &= 1 - \left[1 - \exp\left(-\int_t^{+\infty} \frac{L(x)}{x} dx\right)\right]^{\frac{1}{\alpha}}, \quad t \in (0,+\infty), \end{aligned}$$

and it is the α–generalized aging intensity function for that cdf.

However, in the case $\alpha < 0$, we have a result analogous to Theorem 5.1 and then also in that case an α–generalized aging intensity function determines a family of distributions. Moreover, it is possible to introduce stochastic orders based on the comparisons among α–generalized aging intensity functions. These orders are known as α aging intensity orders and we have $X \leq_{\alpha AI} Y$ if, and only if, $L_{\alpha,X}(t) \geq L_{\alpha,Y}(t)$, $\forall t \in (0,+\infty)$.

5.2 Generalized reversed aging intensity functions

In this section, the concept of generalized reversed aging intensity function is defined and some of its properties are discussed. The results of this section are based on Buono, Longobardi and Szymkowiak [27]. Let W_0 be the distribution function of an exponential random variable with parameter 1, $W_0(t) = 1 - \exp(-t)$, $t > 0$, so $\breve{R}(t) = W_0^{-1}(1 - F(t))$, where $\breve{R}$ is defined in (5.3). In fact, $W_0^{-1}(t) = -\log(1-t)$ and then

$$W_0^{-1}(1 - F(t)) = -\log F(t) = \breve{R}(t).$$

Replacing W_0 by a strictly increasing cdf G with pdf g, it is possible to generalize the concepts of reversed hazard rate function, cumulative reversed hazard rate function and reversed aging intensity function. The generalization of the hazard rate function was introduced by Barlow and Zwet [12, 13].

Definition 5.2. Let X be a non-negative and absolutely continuous random variable with cdf F. Let G be a strictly increasing cdf with pdf g. We define the G–generalized cumulative reversed hazard rate function, $\breve{R}_G$, the G–generalized reversed hazard rate function, q_G, the G–generalized reversed aging intensity function, $\breve{L}_G$, of X as

$$\breve{R}_G(t) = G^{-1}(1 - F(t)), \tag{5.6}$$

$$q_G(t) = -\frac{d\breve{R}_G(t)}{dt} = \frac{f(t)}{g(G^{-1}(1 - F(t)))}, \tag{5.7}$$

$$\breve{L}_G(t) = \frac{tq_G(t)}{\breve{R}_G(t)} = \frac{tf(t)}{g(G^{-1}(1 - F(t)))G^{-1}(1 - F(t))}. \tag{5.8}$$

Again, a very interesting generalization is the one in which G is the cdf of a generalized Pareto distribution. Let X be a non-negative and absolutely continuous random variable with cdf F and pdf f. Then, it is possible to determine the W_α–generalized cumulative reversed hazard rate and the W_α–generalized reversed hazard rate functions:

$$\breve{R}_{W_\alpha}(t) = W_\alpha^{-1}(1 - F(t)) = \begin{cases} \frac{1}{\alpha}[1 - F^\alpha(t)], & \text{for } t > 0, \text{ if } \alpha \neq 0 \\ -\log F(t), & \text{for } t > 0, \text{ if } \alpha = 0 \end{cases}$$

$$q_{W_\alpha}(t) = -\frac{d\breve{R}_\alpha(t)}{dt} = F^{\alpha-1}(t)f(t), \text{ for } t > 0.$$

For the sake of simplicity, those functions can be, respectively, indicated by $\breve{R}_\alpha$, q_α and we can refer to them as the α–generalized cumulative reversed hazard rate function and the α–generalized reversed hazard rate function. Note that the 1–generalized reversed hazard rate function is equal to the density function. In fact, the density function gives a first rough illustration of the aging tendency of the random variable by its monotonicity. Moreover, the 0–generalized reversed hazard rate function is equal to the usual reversed hazard rate function.

From these functions, it is possible to introduce the α–generalized reversed aging intensity function

$$\breve{L}_\alpha(t) = \frac{q_\alpha(t)}{\frac{1}{t}\breve{R}_\alpha(t)} = \begin{cases} \frac{\alpha t F^{\alpha-1}(t) f(t)}{1-F^\alpha(t)}, & \text{for } t>0, \text{ if } \alpha \neq 0 \\ \frac{-tf(t)}{F(t)\log F(t)}, & \text{for } t>0, \text{ if } \alpha = 0. \end{cases} \tag{5.9}$$

The α–generalized reversed aging intensity function describes the relationship between the instantaneous value of the α–generalized reversed hazard rate function $q_\alpha(t)$ and the baseline value of the α–generalized reversed hazard rate function $\frac{1}{t}\breve{R}_\alpha(t)$. The higher the α–generalized reversed aging intensity function (it means the higher the actual value of the α–generalized reversed hazard rate function with respect to its baseline value), the weaker the tendency of aging. Moreover, the α–generalized reversed aging intensity function can be treated as the elasticity (see [112]), except for the sign, of the α–generalized cumulative reversed hazard rate function, i.e., it indicates how much the function $\breve{R}_\alpha$ changes if t changes by a small amount.

Remark 5.1. The 0–generalized reversed aging intensity function is equal to the usual reversed aging intensity function. For $\alpha = 1$, we have

$$\breve{L}_1(t) = \frac{tf(t)}{\overline{F}(t)},$$

i.e., it is the negative of the elasticity of the survival function $\overline{F}$ (they are equal in modulus). For $\alpha = n \in \mathbb{N}$, we have

$$\breve{L}_n(t) = \frac{ntF^{n-1}(t)f(t)}{1-F^n(t)},$$

where the denominator is the survival function of the largest order statistic for a sample of n IID variables, while the numerator is t multiplied by the density of this order statistic. So $\breve{L}_n$ can be considered as the negative of the elasticity for the survival function of the largest order statistic. For $\alpha = -1$, we have

$$\breve{L}_{-1}(t) = \frac{-t(F(t))^{-2}f(t)}{1-(F(t))^{-1}} = \frac{tf(t)}{F(t)(1-F(t))},$$

and so

$$\breve{L}_{-1}(t) = tLOR_X(t) = L_{-1}(t),$$

where LOR_X is the log-odds rate of X, see (1.9).

The next proposition analyzes the monotonicity of α–generalized reversed aging intensity functions with respect to the parameter α.

Proposition 5.1. *Let X be a non-negative and absolutely continuous random variable with cdf F and pdf f. Then, the α–generalized reversed aging intensity function is decreasing with respect to $\alpha \in \mathbb{R}$, $\forall t \in (0, +\infty)$.*

Proof. For some $c \in (0,1)$, consider the function $h_c(\alpha) = \frac{\alpha c^\alpha}{1-c^\alpha}$, for $\alpha \neq 0$. Then,

$$\frac{dh_c(\alpha)}{d\alpha} = \frac{c^\alpha(1-c^\alpha+\log c^\alpha)}{(1-c^\alpha)^2}.$$

That derivative is negative because $c^\alpha \in (0,+\infty)$ and the function $k(t) = 1-t+\log t$ is negative for $t>0$ and different from 1. In fact, $k(1)=0$ and 1 is the maximum point for this function. So h_c is decreasing in $(-\infty,0)\cup(0,+\infty)$. Defining the extension for continuity in 0 of h_c,

$$h_c(0) = \lim_{\alpha\to 0} h_c(\alpha) = \lim_{\alpha\to 0}\frac{\alpha c^\alpha}{1-c^\alpha} = \lim_{\alpha\to 0}\frac{c^\alpha+\alpha c^\alpha\log c}{-c^\alpha\log c} = -\frac{1}{\log c},$$

it is possible to say that h_c is decreasing in $\mathbb{R}$.

Fixing $c = F(t)$, with $t>0$, and multiplying $h_{F(t)}(\alpha)$ by the positive factor $\frac{tf(t)}{F(t)}$ we get that the function

$$\frac{tf(t)}{F(t)}h_{F(t)}(\alpha) = \begin{cases} \frac{\alpha t F^{\alpha-1}(t)f(t)}{1-F^\alpha(t)}, & \text{if } \alpha\neq 0 \\ \frac{-tf(t)}{F(t)\log F(t)}, & \text{if } \alpha = 0 \end{cases} = \breve{L}_\alpha(t)$$

is decreasing in α as t is fixed. □

In the following theorem we show that, for $\alpha>0$, the distribution function of a non-negative and absolutely continuous random variable is defined by the α–generalized reversed aging intensity function and that, under some conditions, a function can be considered as the α–generalized reversed aging intensity function for a unique random variable.

Theorem 5.5. *Let X be a non-negative and absolutely continuous random variable with cdf F and let $\breve{L}_\alpha$ be its α–generalized reversed aging intensity function with $\alpha>0$. Then, F and $\breve{L}_\alpha$ are related, for all $a\in(0,+\infty)$, by the relationship*

$$F(t) = \left[1-\exp\left(-\int_0^t \frac{\breve{L}_\alpha(x)}{x}\,dx\right)\right]^{\frac{1}{\alpha}}, \quad t\in(0,+\infty). \tag{5.10}$$

Moreover, a function $\breve{L}$ defined on $(0,+\infty)$ and satisfying, for $a\in(0,+\infty)$, the following conditions:

(1) $0\leq \breve{L}(t)+\infty$, for all $t\in(0,+\infty)$;

(2) $\lim_{t\to 0^+}\int_t^a \frac{\breve{L}(x)}{x}\,dx < +\infty$;

(3) $\lim_{t\to+\infty}\int_a^t \frac{\breve{L}(x)}{x}\,dx = +\infty$;

determines, for $\alpha>0$, a unique absolutely continuous cdf F by

$$\begin{aligned} F(t) &= 1-W_\alpha\left(\frac{1}{\alpha}\exp\left(-\int_0^t\frac{\breve{L}(x)}{x}\,dx\right)\right) \\ &= \left[1-\exp\left(-\int_0^t\frac{\breve{L}(x)}{x}\,dx\right)\right]^{\frac{1}{\alpha}}, \quad t\in(0,+\infty), \end{aligned} \tag{5.11}$$

and it is the α–generalized reversed aging intensity function for that cdf.

Proof. Fix the distribution function F with respective density function f, and put $\alpha > 0$. From the definition of $\breve{L}_\alpha$ it is possible to obtain

$$\frac{\breve{L}_\alpha(x)}{x} = \frac{\alpha F^{\alpha-1}(x) f(x)}{1 - F^\alpha(x)}, \; x \in (0, +\infty).$$

By integrating both members between 0 and t, we get

$$\int_0^t \frac{\breve{L}_\alpha(x)}{x}\, dx = \int_0^t \frac{\alpha F^{\alpha-1}(x) f(x)}{1 - F^\alpha(x)}\, dx = -\log(1 - F^\alpha(t)),$$

therefore

$$1 - F^\alpha(t) = \exp\left(-\int_0^t \frac{\breve{L}_\alpha(x)}{x}\, dx\right),$$

and so we get (5.10).

Let $\breve{L}$ be a function defined on $(0, +\infty)$ and satisfying, for $a \in (0, +\infty)$, the conditions *(1)*, *(2)*, *(3)*. We show that $1 - W_\alpha\left(\frac{1}{\alpha}\exp\left(-\int_0^t \frac{\breve{L}(x)}{x}\, dx\right)\right) = F(t)$ defines a cdf of a non-negative and absolutely continuous random variable. In fact, from *(2)* it follows $\lim_{t\to 0^+} F(t) = 0$, whereas from *(3)* we obtain $\lim_{t\to+\infty} F(t) = 1$. Since W_α is increasing, $\alpha > 0$ and the exponential function is increasing, in order to show that it is an increasing function we have to prove that $-\int_0^t \frac{\breve{L}(x)}{x}\, dx$ is a decreasing function in t, i.e., $\int_0^t \frac{\breve{L}(x)}{x}\, dx$ is increasing in t, but this is immediate since the integrand is non-negative and as t increases, the integration interval widens. Since W_α, the exponential function, the multiplication for a scalar and the indefinite integral $t \mapsto \int_0^t \frac{\breve{L}(x)}{x} dx$ are continuous functions, we have a continuous function. In order to obtain the absolute continuity of F, it suffices to observe that the derivative

$$F'(t) = -\frac{1}{\alpha}\left[1 - \exp\left(-\int_0^t \frac{\breve{L}(x)}{x} dx\right)\right]^{\frac{1}{\alpha}-1} \exp\left(-\int_0^t \frac{\breve{L}(x)}{x} dx\right)\left(-\frac{\breve{L}(t)}{t}\right)$$

is non-negative in $t > 0$. Finally, F and $\breve{L}$ are related by the same relationship found in the first part of the theorem and so $\breve{L}$ is the α–generalized reversed aging intensity function for that cdf. $\square$

Remark 5.2. If $\breve{L}$ is the α–generalized reversed aging intensity function, with $\alpha > 0$, of a non-negative and absolutely continuous random variable X, it satisfies conditions *(1)*, *(2)*, *(3)* of Theorem 5.5. In fact, from (5.9) we observe that $\breve{L}$ is non-negative for $t \in (0, +\infty)$ and we have

$$\begin{aligned}
\lim_{t\to 0^+} \int_t^a \frac{\breve{L}(x)}{x}\, dx &= \lim_{t\to 0^+} \int_t^a \frac{\alpha F^{\alpha-1}(x) f(x)}{1 - F^\alpha(x)}\, dx = \lim_{t\to 0^+} -\log\frac{1 - F^\alpha(a)}{1 - F^\alpha(t)} < +\infty, \\
\lim_{t\to+\infty} \int_a^t \frac{\breve{L}(x)}{x}\, dx &= \lim_{t\to+\infty} \int_a^t \frac{\alpha F^{\alpha-1}(x) f(x)}{1 - F^\alpha(x)}\, dx = \lim_{t\to+\infty} -\log\frac{1 - F^\alpha(t)}{1 - F^\alpha(a)} = +\infty.
\end{aligned}$$

If we have some data it is possible to obtain an estimation of both cdf and α–generalized reversed aging intensity function. So it may happen that the shape of an α–generalized reversed aging intensity function is easier to recognize than that of the cdf.

In analogy with Theorem 5.5, there is a similar result about α–generalized reversed aging intensity functions with $\alpha < 0$. Nevertheless, it is essentially different since in this case the α–generalized reversed aging intensity does not determine a unique cdf but a family of distributions, as in the classical case. The result is stated in the following theorem whose proof is omitted being similar to that of Theorem 5.5.

Theorem 5.6. *Let X be a non-negative and absolutely continuous random variable with cdf F and let $\breve{L}_\alpha$ be its α–generalized reversed aging intensity function with $\alpha < 0$. Then, F and $\breve{L}_\alpha$ are related, for all $a \in (0, +\infty)$, by the relationship*

$$F(t) = \left[1 - (1 - F^\alpha(a)) \exp\left(-\int_a^t \frac{\breve{L}_\alpha(x)}{x}\, dx\right)\right]^{\frac{1}{\alpha}}, \quad t \in (0, +\infty). \tag{5.12}$$

Moreover, a function $\breve{L}$ defined on $(0, +\infty)$ and satisfying, for $a \in (0, +\infty)$, the following conditions:

(1) $0 \le \breve{L}(t) < +\infty$, *for all* $t \in (0, +\infty)$;

(2) $\lim_{t\to 0^+} \int_t^a \frac{\breve{L}(x)}{x}\, dx = +\infty$;

(3) $\lim_{t\to +\infty} \int_a^t \frac{\breve{L}(x)}{x}\, dx = +\infty$;

determines, for $\alpha < 0$ and $k \in (0, +\infty)$, a family of absolutely continuous distribution functions F_k by

$$\begin{aligned} F_k(t) &= 1 - W_\alpha\left(k \exp\left(-\int_a^t \frac{\breve{L}(x)}{x}\, dx\right)\right) \\ &= \left[1 - k\alpha \exp\left(-\int_a^t \frac{\breve{L}(x)}{x}\, dx\right)\right]^{\frac{1}{\alpha}}, \quad t \in (0, +\infty), \end{aligned} \tag{5.13}$$

and it is the α–generalized reversed aging intensity function for those distribution functions.

Remark 5.3. The expression $W_\alpha\left(k \exp\left(-\int_a^t \frac{\breve{L}(x)}{x}\, dx\right)\right)$ depends only on the parameter $k \in (0, +\infty)$ being fictitious the dependence on $a \in (0, +\infty)$. In fact, replacing a by $b \in (0, +\infty)$ we get

$$\begin{aligned} W_\alpha\left(k \exp\left(-\int_b^t \frac{\breve{L}(x)}{x}\, dx\right)\right) &= W_\alpha\left(k \exp\left(-\int_b^a \frac{\breve{L}(x)}{x}\, dx\right) \exp\left(-\int_a^t \frac{\breve{L}(x)}{x}\, dx\right)\right) \\ &= W_\alpha\left(k_1 \exp\left(-\int_a^t \frac{\breve{L}(x)}{x}\, dx\right)\right), \end{aligned}$$

where $k_1 = k \exp\left(-\int_b^a \frac{\breve{L}(x)}{x}\, dx\right) > 0$.

Remark 5.4. If $\breve{L}$ is the α–generalized reversed aging intensity function, with $\alpha < 0$, of a non-negative and absolutely continuous random variable X, it satisfies conditions *(1)*, *(2)*, *(3)* of Theorem 5.6. In fact, from (5.9) we observe that $\breve{L}$ is non-negative for $t \in (0, +\infty)$ and we have

$$\lim_{t\to 0^+} \int_t^a \frac{\breve{L}(x)}{x}\, dx = \lim_{t\to 0^+} \int_t^a \frac{\alpha F^{\alpha-1}(x) f(x)}{1-F^{\alpha}(x)}\, dx = \lim_{t\to 0^+} -\log \frac{1-F^{\alpha}(a)}{1-F^{\alpha}(t)} = +\infty,$$

$$\lim_{t\to +\infty} \int_a^t \frac{\breve{L}(x)}{x}\, dx = \lim_{t\to +\infty} \int_a^t \frac{\alpha F^{\alpha-1}(x) f(x)}{1-F^{\alpha}(x)}\, dx = \lim_{t\to +\infty} -\log \frac{1-F^{\alpha}(t)}{1-F^{\alpha}(a)} = +\infty.$$

Remark 5.5. If $\breve{L}$ is a function that satisfies conditions *(1)*, *(2)*, *(3)* of Theorem 5.6 then it determines, for $\alpha = 0$ and $k \in (0, +\infty)$, a family of absolutely continuous distribution functions F_k by the relationship

$$F_k(t) = 1 - W_0\left(k \exp\left(-\int_a^t \frac{\breve{L}(x)}{x}\, dx\right)\right) = \exp\left[-k\, \exp\left(-\int_a^t \frac{\breve{L}(x)}{x} dx\right)\right],$$

$t \in (0, +\infty)$, and it is the 0–generalized reversed aging intensity function (i.e., the reversed aging intensity function) for these distribution functions. This follows from Corollary 4 in [113] and by (5.4).

Example 5.1. Let us consider $\breve{L}_\alpha(t) = A > 0$, for $t > 0$. It can be a constant α–generalized reversed aging intensity function for $\alpha \le 0$, since for $\alpha > 0$ it does not satisfy the hypothesis of Theorem 5.5. For $\alpha = 0$ it determines a family of inverse two-parameter Weibull distributions by

$$F_k(t) = \exp\left[-k\left(\frac{1}{t}\right)^A\right], \qquad t \in (0, +\infty), \tag{5.14}$$

where k is a non-negative parameter. For $\alpha < 0$, it determines a family of continuous distributions by

$$F_k(t) = \left[1 + k\left(\frac{1}{t}\right)^A\right]^{\frac{1}{\alpha}}, \qquad t \in (0, +\infty) \tag{5.15}$$

where k is a non-negative parameter.

Example 5.2. Let us consider $\breve{L}_\alpha(t) = A + Bt$, for $t > 0$ where $A, B > 0$. It can be a linear α–generalized reversed aging intensity function for $\alpha \le 0$, since for $\alpha > 0$ it does not satisfy the hypothesis of Theorem 5.5. For $\alpha = 0$, it determines a family of continuous distributions by

$$F_k(t) = \exp\left[-k\left(\frac{1}{t}\right)^A \exp(-B\, t)\right], \qquad t \in (0, +\infty) \tag{5.16}$$

where k is a non-negative parameter. For $\alpha < 0$, it determines a family of continuous distributions by

$$F_k(t) = \left[1 + k\left(\frac{1}{t}\right)^A \exp(-B\,t)\right]^{\frac{1}{\alpha}}, \qquad t \in (0, +\infty) \tag{5.17}$$

where k is a non-negative parameter.

Example 5.3. Let us consider $\breve{L}_\alpha(t) = Bx$, for $t > 0$, where $B > 0$. It can be a linear α–generalized reversed aging intensity function for $\alpha > 0$ since it satisfies the hypothesis of Theorem 5.5. It determines a unique continuous distribution function by

$$F(t) = [1 - \exp(-Bt)]^{\frac{1}{\alpha}}, \qquad t \in (0, +\infty), \tag{5.18}$$

i.e., an exponentiated exponential distribution (see [53]). Note that for $\alpha = 1$ this is the cdf of an exponential random variable with parameter B. So if X has 1–generalized reversed aging intensity function $\breve{L}_1(t) = Bt$, for $t > 0$ and $B > 0$, then $X \sim Exp(B)$.

5.2.1 α–generalized reversed aging intensity orders

In this section we study the family of the α–generalized reversed aging intensity orders. In the following, we use the notation $L_{\alpha,X}$ to indicate the α–generalized aging intensity function of the random variable X and $\breve{L}_{\alpha,X}$ to indicate the α–generalized reversed aging intensity function of the random variable X. In the next proposition we show a useful relationship between $L_{\alpha,X}$ and $\breve{L}_{\alpha,\frac{1}{X}}$.

Proposition 5.2. *Let X be a non-negative and absolutely continuous random variable and let $\frac{1}{X}$ be its inverse. Then, the following equality holds*

$$L_{\alpha,X}\left(\frac{1}{t}\right) = \breve{L}_{\alpha,\frac{1}{X}}(t), \ \ t \in (0, +\infty). \tag{5.19}$$

Proof. We obtain an expression for the distribution function and the density function of the random variable $\frac{1}{X}$ through X, for $t > 0$ we have

$$\begin{aligned} F_{\frac{1}{X}}(t) &= \mathbb{P}\left(\frac{1}{X} \le t\right) = \mathbb{P}\left(X \ge \frac{1}{t}\right) = 1 - F_X\left(\frac{1}{t}\right), \\ f_{\frac{1}{X}}(t) &= \frac{1}{t^2} f_X\left(\frac{1}{t}\right). \end{aligned}$$

If $\alpha = 0$, by (5.4) we have, for $t > 0$,

$$\breve{L}_{0,\frac{1}{X}}(t) = \breve{L}_{\frac{1}{X}}(t) = L_X\left(\frac{1}{t}\right) = L_{0,X}\left(\frac{1}{t}\right).$$

If $\alpha \neq 0$, we have, for $t > 0$,

$$\breve{L}_{\alpha,\frac{1}{X}}(t) = \frac{\alpha t (F_{\frac{1}{X}}(t))^{\alpha-1} f_{\frac{1}{X}}(t)}{1 - (F_{\frac{1}{X}}(t))^\alpha} = \frac{\alpha \frac{1}{t}\left(1 - F_X\left(\frac{1}{t}\right)\right)^{\alpha-1} f_X\left(\frac{1}{t}\right)}{1 - (1 - F_X\left(\frac{1}{t}\right))^\alpha} = L_{\alpha,X}\left(\frac{1}{t}\right).$$

$\square$

Definition 5.3. Let X and Y be non-negative and absolutely continuous random variables and let α be a real number. We say that X is smaller than Y in the α–generalized reversed aging intensity order, $X \leq_{\alpha RAI} Y$, if and only if $\breve{L}_{\alpha,X}(t) \leq \breve{L}_{\alpha,Y}(t)$, $\forall t \in (0, +\infty)$.

In the next lemma we show a relationship between the αRAI order and the αAI order.

Lemma 5.1. *Let X and Y be non-negative and absolutely continuous random variables and let α be a real number. We have $X \leq_{\alpha RAI} Y$ if, and only if, $\frac{1}{X} \geq_{\alpha AI} \frac{1}{Y}$.*

Proof. We have $X \leq_{\alpha RAI} Y$ if and only if $\breve{L}_{\alpha,X}(t) \leq \breve{L}_{\alpha,Y}(t)$, $\forall t \in (0, +\infty)$. By Proposition 5.2 this is equivalent to $L_{\alpha,\frac{1}{X}}\left(\frac{1}{t}\right) \leq L_{\alpha,\frac{1}{Y}}\left(\frac{1}{t}\right)$, $\forall t \in (0, +\infty)$, i.e., $\frac{1}{X} \geq_{\alpha AI} \frac{1}{Y}$. □

Remark 5.6. For particular choices of the real number α we get some connections with other stochastic orders. Obviously, the reversed aging intensity order coincides with the 0–generalized reversed aging intensity order. For $\alpha = 1$ we have shown in Remark 5.1 that $\breve{L}_{1,X}(t) = tr_X(t)$, so we get a relationship with the hazard rate order. In fact,

$$\begin{aligned} X \leq_{hr} Y \Leftrightarrow r_X(t) \geq r_Y(t), \forall t > 0 \Leftrightarrow tr_X(t) \geq tr_Y(t), \forall t > 0 \\ \Leftrightarrow \breve{L}_{1,X}(t) \geq \breve{L}_{1,Y}(t), \forall t > 0 \Leftrightarrow X \geq_{1RAI} Y. \end{aligned}$$

For $\alpha = -1$ we have shown in Remark 5.1 that $\breve{L}_{-1,X}(t) = tLOR_X(t) = L_{-1,X}(t)$, so we get a connection with the log-odds rate order. In fact,

$$\begin{aligned} X \leq_{LOR} Y \Leftrightarrow LOR_X(t) \geq LOR_Y(t), \forall t > 0 \Leftrightarrow tLOR_X(t) \geq tLOR_Y(t), \forall t > 0 \\ \Leftrightarrow \breve{L}_{-1,X}(t) \geq \breve{L}_{-1,Y}(t), \forall t > 0 \Leftrightarrow X \geq_{-1RAI} Y. \end{aligned}$$

Moreover, we have $X \geq_{-1RAI} Y \Leftrightarrow X \leq_{-1AI} Y$ so they are dual relations. By Lemma 5.1 we get the following series of equivalences

$$X \leq_{-1RAI} Y \Leftrightarrow \frac{1}{X} \leq_{-1RAI} \frac{1}{Y} \Leftrightarrow X \geq_{-1AI} Y \Leftrightarrow \frac{1}{X} \geq_{-1AI} \frac{1}{Y}.$$

For $\alpha = n \in \mathbb{N}$, we have shown in Remark 5.1 that

$$\breve{L}_{n,X}(t) = \frac{nt(F_X(t))^{n-1} f_X(t)}{1 - (F_X(t))^n} = tr_{X_{n:n}}(t),$$

so there is a connection with the largest order statistic and the hazard rate order. In fact

$$\begin{aligned} X_{n:n} \leq_{hr} Y_{n:n} \Leftrightarrow r_{X_{n:n}}(t) \geq r_{Y_{n:n}}(t), \forall t > 0 \Leftrightarrow tr_{X_{n:n}}(t) \geq tr_{Y_{n:n}}(t), \forall t > 0 \\ \Leftrightarrow \breve{L}_{n,X}(t) \geq \breve{L}_{n,Y}(t), \forall t > 0 \Leftrightarrow X \geq_{nRAI} Y. \end{aligned}$$

Proposition 5.3. *Let X and Y be non-negative and absolutely continuous random variables such that $X \leq_{st} Y$, i.e., $F_X(t) \geq F_Y(t)$ for all $t > 0$.*

(1) If there exists $\beta \in \mathbb{R}$ such that $X \leq_{\beta RAI} Y$, then for all $\alpha < \beta$ we have $X \leq_{\alpha RAI} Y$;

(2) If there exists $\beta \in \mathbb{R}$ such that $X \geq_{\beta RAI} Y$, then for all $\alpha > \beta$ we have $X \geq_{\alpha RAI} Y$.

Proof. (1). From $X \leq_{\beta RAI} Y$ and Lemma 5.1 we have $\frac{1}{X} \geq_{\beta AI} \frac{1}{Y}$. Moreover, from $X \leq_{st} Y$ we get $\frac{1}{X} \geq_{st} \frac{1}{Y}$ so by Proposition 4 in [114] we obtain that $\forall \alpha < \beta$ $\frac{1}{X} \geq_{\alpha AI} \frac{1}{Y}$, i.e., $X \leq_{\alpha RAI} Y$. The proof of part (2) is analogous. □

Proposition 5.4. *Let X and Y be non-negative and absolutely continuous random variables.*

(1) If there exists $\beta \in \mathbb{R}$ such that for all $\alpha < \beta$ we have $X \geq_{\alpha RAI} Y$, then $X \geq_{rh} Y$;

(2) If there exists $\beta \in \mathbb{R}$ such that for all $\alpha > \beta$ we have $X \leq_{\alpha RAI} Y$, then $X \geq_{st} Y$.

Proof. (1). From $X \geq_{\alpha RAI} Y$ and Lemma 5.1 we have $\frac{1}{X} \leq_{\alpha AI} \frac{1}{Y}$, $\forall \alpha < \beta$. So with the use of Proposition 5 in [114] we obtain $\frac{1}{X} \leq_{hr} \frac{1}{Y}$, i.e., $X \geq_{rh} Y$. The proof of part (2) is analogous. □

Corollary 5.1. *Let X and Y be non-negative and absolutely continuous random variables.*

(1) $X \leq_{st} Y$ and $X \geq_{LOR} Y \Rightarrow X \leq_{\alpha RAI} Y$ for all $\alpha \in (-\infty, -1)$;

(2) $X \leq_{st} Y$ and $X \leq_{LOR} Y \Rightarrow X \geq_{\alpha RAI} Y$ for all $\alpha \in (-1, +\infty)$.

Proof. (1). We have $X \geq_{LOR} Y \Leftrightarrow X \leq_{-1RAI} Y$ so the proof is completed with the use of Proposition 5.3. The proof of part (2) is analogous. □

5.2.2 Application of α–generalized reversed aging intensity function in data analysis

It is a difficult task to recognize the lifetime data distribution by analyzing only the shapes of their pdf and cdf estimators. But sometimes, the corresponding α–generalized reversed aging intensity function for a properly chosen α can have a relatively simple form, and it can be easily recognized with the use of the respective reversed aging intensity estimate.

For some distribution F with support $(0, +\infty)$, we obtain a natural estimator of the α–generalized reversed aging intensity function

$$\widehat{\breve{L}}_{\alpha}(t) = \begin{cases} \frac{\alpha\, t\, \widehat{f}(t)[\widehat{F}(t)]^{\alpha-1}}{1-[\widehat{F}(t)]^{\alpha}} & \text{for} \quad t > 0, \quad \alpha \neq 0 \\ -\frac{t\, \widehat{f}(t)}{\widehat{F}(t)\ln[\widehat{F}(t)]} & \text{for} \quad t > 0, \quad \alpha = 0, \end{cases} \tag{5.20}$$

where $\widehat{f}$ denotes a non-parametric estimate of the unknown density function f and $\widehat{F}(t) = \int_0^t \widehat{f}(x)dx$ represents the corresponding distribution function estimate. The proposed estimation of the aging intensity function is possible if we assume that the data follow an absolutely continuous distribution with support $(0, +\infty)$ and if the non-parametric estimate of its density function exists. Moreover, larger sample sizes generally lead to increased precision of estimation. We perform our study for both the generated and the real data.

In the following example we consider an application of the estimator (5.20) for $\alpha = -1$ to verify the hypothesis that some simulated data come from the family of inverse log-logistic distributions.

Example 5.4. Our goal is to check if a member of the inverse log-logistic distributions $invLLog(\gamma, \lambda)$ with the distribution function given by

$$F_{\gamma,\lambda}(t) = \left[1 + \left(\frac{\lambda}{t}\right)^{\gamma}\right]^{-1}, \quad t \in (0, +\infty), \tag{5.21}$$

for some unknown positive parameters of the shape γ and the scale λ, is the parent distribution of a random sample $X_1, \ldots, X_N$.

From Example 5.1 we know that for distribution function (5.21), the -1–generalized reversed aging intensity function is constant and equal to $\breve{L}_{-1}(t) = \gamma$. So, we check if the respective reversed aging intensity estimator (5.20) is indeed an accurate approximation of a constant function. We use the following procedure to obtain N independent random variables $X_1, \ldots, X_N$ with $invLLog(\gamma, \lambda)$ lifetime distribution. First, we generate standard uniform random variables $U_1, \ldots, U_N$ by using function `random` of *MATLAB*. Then, by applying the inverse transform technique with $F_{\gamma,\lambda}(t) = \left[1 + \left(\frac{\lambda}{t}\right)^{\gamma}\right]^{-1}$, we get $Y_i = F^{-1}_{\gamma,\lambda}(1 - U_i) = \lambda \left(\frac{1}{1-U_i} - 1\right)^{-\frac{1}{\gamma}}$, $i = 1, \ldots, N$, with the inverse log-logistic distribution $invLLog(\gamma, \lambda)$. In this way, by applying the function `random` with the `seed`$= 88$, we generate $N = 1000$ independent inverse log-logistic random variables with the shape parameter $\gamma = 4$, and the scale parameter $\lambda = 0.5$. To calculate the reversed aging intensity estimator (5.20), we apply a kernel density estimator [21], given in *MATLAB* `ksdensity` function,

$$\widehat{f}(t) = \frac{1}{N\,h} \sum_{j=1}^{N} K\left(\frac{t - X_j}{h}\right), \tag{5.22}$$

with a chosen normal kernel smoothing function and a selected bandwidth $h = 0.05$. Then, the kernel estimator of the distribution function is equal to

$$\widehat{F}(t) = \frac{1}{N} \sum_{j=1}^{N} I\left(\frac{t - X_j}{h}\right),$$

where $I(t) = \int_{-\infty}^{t} K(x)dx$. The obtained -1–generalized reversed aging intensity function estimate (5.20) is

$$\widehat{\breve{L}}_{-1}(t) = \frac{t\,\widehat{f}(t)}{\widehat{F}(t)\left[1 - \widehat{F}(t)\right]} = \frac{t\,\frac{1}{N\,h}\sum_{j=1}^{N} K\left(\frac{t-X_j}{h}\right)}{\frac{1}{N}\sum_{j=1}^{N} I\left(\frac{t-X_j}{h}\right)\left[1 - \frac{1}{N}\sum_{j=1}^{N} I\left(\frac{t-X_j}{h}\right)\right]}. \tag{5.23}$$

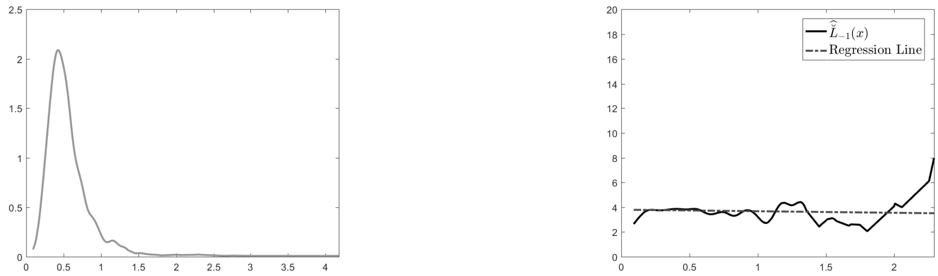

Figure 5.1: Density estimator $\widehat{f}(x)$ for the data from Example 5.4 (left). $\widehat{\breve{L}}_0(x)$ and adjusted regression line for the data from Example 5.4 (right).

For our simulated data, the plot of the density estimator (5.22) is presented in Figure 5.1, left. Analyzing the plot, it is not easy to decide if the density function belongs to the inverse log-logistic family. We can observe that the plot of respective estimator (5.23) of -1–generalized reversed aging intensity function $\widehat{\breve{L}}_{-1}(t)$ (see Figure 5.1, right), oscillates around a constant function, especially after removing few outlying values at the right-end. This gives us the motivation to accept our hypothesis that an inverse log-logistic distribution is the parent distribution of the generated sample.

To justify our intuitive decision, we propose to carry out the following more formal statistical procedure. First, we calculate the least squares estimate of the intercept which for our data is equal to $\widehat{\gamma} = 3.7990$. Next, we put it into the log-likelihood function, and determine maximum likelihood estimator (MLE) of parameter λ by maximizing it. The problem resolves into finding the solution of the equation

$$\sum_{i=1}^{N} \frac{1}{\left(\frac{x_i}{\lambda}\right)^{\widehat{\gamma}} + 1} = \frac{N}{2},$$

that is $\widehat{\lambda} = 0.4957$. Note that the estimators $\widehat{\gamma}$ and $\widehat{\lambda}$ based on the empirical -1–generalized reversed aging intensity are quite precise (see Table 5.1).

Table 5.1: Parameters of $invLLog(\gamma, \lambda)$

	γ	λ
Theoretical parameters	4	0.5
Estimators	3.7990	0.4957

Finally, by the chi-square goodness-of-fit test we check if the data really fit the inverse log-logistic distribution. For this purpose, we apply the function `histogram`, available in `MATLAB` and group the data into $k = 20$ classes of observations lying into intervals $[x_j, x_{j+1}) = [x_j, x_j + \Delta x)$, $j = 1, \dots, k$, of length $\Delta x = 0.21$. The classes, together with their empirical frequencies $N_j = N_j(X_1, \ldots, X_N)$ and theoretical frequencies based on the inverse log-logistic distribution with parameters replaced by the estimators $n_j = N\left[F_{\widehat{\gamma},\widehat{\lambda}}(x_{j+1}) - F_{\widehat{\gamma},\widehat{\lambda}}(x_j)\right]$, are presented in Table 5.2.

Table 5.2: Grouped data and respective values of empirical and theoretical frequency

class	$[x_j, x_{j+1})$	N_j	n_j
1	0.0000-0.2100	26	36.8543
2	0.2100-0.4200	322	310.6691
3	0.4200-0.6300	371	365.5653
4	0.6300-0.8400	150	168.0593
5	0.8400-1.0500	68	64.2263
6	1.0500-1.2600	29	26.5322
7	1.2600-1.4700	15	12.2550
8	1.4700-1.6800	5	6.2413
9	1.6800-1.8900	3	3.4409
10	1.8900-2.1000	3	2.0223
11	2.1000-2.3100	0	1.2522
12	2.3100-2.5200	1	0.8095
13	2.5200-2.7300	2	0.5425
14	2.7300-2.9400	1	0.3749
15	2.9400-3.1500	0	0.2660
16	3.1500-3.3600	0	0.1931
17	3.3600-3.5700	0	0.1430
18	3.5700-3.7800	0	0.1078
19	3.7800-3.9900	0	0.0826
20	3.9900-4.200	1	0.0641

Furthermore, with *MATLAB* function `chi2gof` we determine the value of chi-square statistics $\chi^2 = 9.3209$ with $\nu = 7$ degrees of freedom (automatically joining together the last twelve classes with low frequencies) and the respective p-value, $p = 0.2304$. It means that for a given significance level less than 0.2304 we do not reject the hypothesis that the considered data follow the inverse log-logistic distribution.

Next, we present an example with real data. Analyzing its estimated α–generalized reversed aging intensity we could assume that the data follow the adequate distribution.

Example 5.5. The real data (see Data Set 6.2 in [77]) concern failure times of 20 components: 0.067 0.068 0.076 0.081 0.084 0.085 0.085 0.086 0.089 0.098 0.098 0.114 0.114 0.115 0.121 0.125

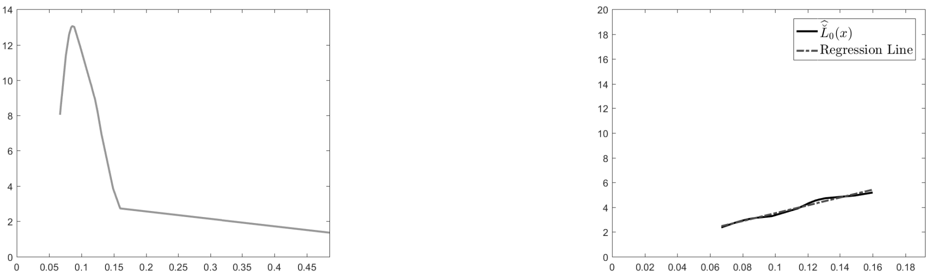

Figure 5.2: Kernel density estimator $\widehat{f}$ for the data from Example 5.5 (left). $\widehat{\breve{L}}_0(x)$ and adjusted regression line for the data from Example 5.5 (right).

0.131 0.149 0.160 0.485.

For the given data, the plot of the normal kernel density estimator [21], obtained by *MATLAB* function `ksdensity` with a returned bandwidth $h = 0.0147$, is presented in Figure 5.2, left. An analysis of the graph does not enable us to recognize the distribution of the data. To identify the data distribution we propose to estimate the 0–generalized reversed aging intensity by (5.20)

$$\widehat{\breve{L}}_0(t) = -\frac{t\,\widehat{f}(t)}{\widehat{F}(t)\ln[\widehat{F}(t)]}, \quad t \in (0, +\infty).$$

The plot of the estimator $\widehat{\breve{L}}_0(t)$, Figure 5.2, right, can be treated as oscillating around a linear function, especially after removing one outlying value at the right-end. This motivates us to state the hypothesis that data follow an inverse modified Weibull distribution (see Example 5.2) with distribution function

$$F_{\gamma,\lambda,\delta}(t) = \exp\left[-\left(\frac{\lambda}{t}\right)^{\gamma} \exp(-\delta\, t)\right], \qquad t \in (0, +\infty), \tag{5.24}$$

and 0–generalized reversed aging intensity function

$$\breve{L}_0(t) = \delta\, t + \gamma, \qquad t \in (0, +\infty).$$

Moreover, we provide the following procedure. First, we determine the least squares estimates $\widehat{\gamma} = 0.3441$ and $\widehat{\delta} = 31.6785$ of the intercept and the slop of linear $\breve{L}_0$, respectively. Then we determine MLE of parameter λ

$$\widehat{\lambda} = \left(\frac{N}{\sum_{i=1}^{N} \frac{\exp(-\widehat{\delta}\, x_i)}{(x_i)^{\widehat{\gamma}}}}\right)^{\frac{1}{\widehat{\gamma}}}$$

which maximizes the likelihood function. Here, we obtain $\widehat{\lambda} = 549.9663$. Then, to check if the data fit the inverse modified Weibull distribution we use (adequate for few data) the Kolmogorov-Smirnov goodness-of-fit test (available in *MATLAB* function `kstest`), we determine statistics $K = 0.1496$ and p-value of the test equal to $p = 0.7072$. Then, for a given significance level less than 0.7072 we do not reject the hypothesis that the considered data follow the inverse modified Weibull distribution.

5.3 Multivariate conditional aging intensity functions

In this section, the concept of aging intensity function is extended to the multivariate case by the use of the multivariate conditional hazard rate functions. Some properties of these functions are studied and a focus on the bivariate case is performed. Finally, the multivariate conditional aging intensity functions are studied for the time-homogeneous load-sharing model and a study on the comparison among surviving components in a system is provided. The results of this section are given in Buono [22].

Based on the definition of multivariate conditional hazard rate functions, here we introduce the multivariate conditional aging intensity functions. We observe that these functions will depend on the dynamic history, observed up to the calendar time t, for the random vector $(X_1, \dots, X_n)$. In particular, it is defined only for the components surviving at t and it depends on the failure times of the components which have failed before t. The given definition entails that the multivariate conditional aging intensity functions establish the tendency to aging of random variables and allows us to make comparisons among surviving components at a fixed time. Moreover, it is possible to observe that the tendency of aging of a component proceeds continuously with the exception of the failure times of the other components when it may undergo a sudden variation due to the stochastic dependence of components. Hence, it is of great interest to study the continuity of multivariate conditional aging intensity functions.

Definition 5.4. Let $(X_1, \dots, X_n)$ be a random vector whose components are non-negative random variables with an absolutely continuous joint distribution. For $i_1, \dots, i_k \in [n]$, the Multivariate Conditional Aging Intensity (MCAI) function is defined as

$$L_j(t|i_1, \dots, i_k; t_1, \dots, t_k) = \frac{\lambda_j(t|i_1, \dots, i_k; t_1, \dots, t_k)}{\frac{1}{t}\sum_{h=0}^{k}\int_{t_h}^{t_{h+1}} \lambda_j(x|i_1, \dots, i_h; t_1, \dots, t_i)dx} \tag{5.25}$$

where $0 \equiv t_0 < t_1 < t_2 < \cdots < t_k < t_{k+1} \equiv t$, $j \notin I = \{i_1, \dots, i_k\}$ and $\min_{l \notin I} X_l > t$. In the case in which $I = \emptyset$, the MCAI fucntion can be expressed as

$$L_j(t|\emptyset) = \frac{\lambda_j(t|\emptyset)}{\frac{1}{t}\int_0^t \lambda_j(x|\emptyset)dx}. \tag{5.26}$$

Remark 5.7. If $X_1, \dots, X_n$ are independent, then MCAI functions reduce to the classical aging intensity functions since in this case the multivariate conditional hazard rates are equal to the hazard rates independently of I. In fact,

$$\begin{aligned}\lambda_j(t|i_1, \dots, i_k; t_1, \dots, t_k) &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(X_j \le t + \Delta t \,\middle|\, X_{i_1} = t_1, \dots, X_{i_k} = t_k, \min_{h \notin I} X_h > t\right) \\ &= \lim_{\Delta t \to 0^+} \frac{1}{\Delta t}\mathbb{P}\left(X_j \le t + \Delta t \,|X_j > t\right) = r_j(t).\end{aligned}$$

Then, from (5.25), we get

$$L_j(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = \frac{r_j(t)}{\frac{1}{t}\sum_{i=0}^{k}\int_{t_i}^{t_{i+1}} r_j(x)dx} = \frac{r_j(t)}{\frac{1}{t}\int_0^t r_j(x)dx} = L_j(t).$$

In a similar manner, we get $L_j(t|\emptyset) = L_j(t)$.

In the following example, we present the computation of the MCAI functions $L_j(t|\emptyset)$ for simply models characterized by multivariate conditional hazard rate functions $\lambda_j(t|\emptyset)$ proportional to a power of t and we observe that the MCAI functions are independent of the constant of proportionality.

Example 5.6. Let us evaluate the MCAI functions $L_j(t|\emptyset)$ for a model in which $\lambda_j(t|\emptyset) = at$, $a > 0$. From (5.26) we have

$$L_j(t|\emptyset) = \frac{at}{\frac{1}{t}\int_0^t axdx} = 2,$$

and it is independent on a. Moreover, if $\lambda_j(t|\emptyset) = at^b$, where $a, b > 0$, then we get

$$L_j(t|\emptyset) = \frac{at^b}{\frac{1}{t}\int_0^t ax^b dx} = b + 1,$$

that is again independent on a.

In the following proposition, we generalize the property shown in Example 5.6 concerning the comparison between two models with proportional multivariate conditional hazard rate functions. This circumstance can be interpreted as a generalization of the classical proportional hazard rate model introduced in [31] about the univariate case. Generally, it is of interest to consider distributions with proportional hazard rates to obtain models which preserve the monotonicity properties of a fixed hazard rate function.

Proposition 5.5. *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_n)$ *be two random vectors. If there exists a constant* $a > 0$ *such that* $\lambda_j^{(\mathbf{Y})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = a\lambda_j^{(\mathbf{X})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k)$ *for all* $i_1, \ldots, i_k$, $0 < t_1 < \cdots < t_k < t$, $j \notin I = \{i_1, \ldots, i_k\}$, *then*

$$L_j^{(\mathbf{X})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k) = L_j^{(\mathbf{Y})}(t|i_1, \ldots, i_k; t_1, \ldots, t_k). \tag{5.27}$$

In the following proposition, we study the continuity of the MCAI functions associated to a fixed component. The critical points are the ones in which the other components fail. In fact, if we consider a time between two consecutive failures, the expression of the MCAI function is given in (5.25) without changing the parameters and then its continuity is guaranteed by the continuity of multivariate conditional hazard rate functions that is assured under the assumption of absolutely continuous joint distribution.

Proposition 5.6. *Let $(X_1,\dots,X_n)$ be a random vector with non-negative components. Then, for $j\notin I=\{i_1,\dots,i_k\}$,*

$$\lim_{t\to t_k^+} L_j(t|i_1,\dots,i_k;t_1,\dots,t_k)=\lim_{t\to t_k^-} L_j(t|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})$$

if, and only if

$$\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)=\lambda_j(t_k|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1}).$$

Proof. From the definition of MCAI functions, we have

$$\begin{aligned}
&L_j(t|i_1,\dots,i_k;t_1,\dots,t_k)\\
&=\frac{t\lambda_j(t|i_1,\dots,i_k;t_1,\dots,t_k)}{\sum_{r=0}^{k-1}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx+\int_{t_k}^{t}\lambda_j(x|i_1,\dots,i_k;t_1,\dots,t_k)dx},\\
&L_j(t|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})\\
&=\frac{t\lambda_j(t|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})}{\sum_{r=0}^{k-2}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx+\int_{t_{k-1}}^{t}\lambda_j(x|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})dx},
\end{aligned}$$

and then by taking the limits we obtain

$$\begin{aligned}
\lim_{t\to t_k^+} L_j(t|i_1,\dots,i_k;t_1,\dots,t_k)= &= \frac{t_k\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)}{\sum_{r=0}^{k-1}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx},\\
\lim_{t\to t_k^-} L_j(t|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1}) &= \frac{t_k\lambda_j(t_k|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})}{\sum_{r=0}^{k-1}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx}.
\end{aligned}$$

Then, the above limits coincide if, and only if,

$$\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)=\lambda_j(t_k|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1}).$$

□

From the above proposition, we can conclude that the jumps of the MCAI functions, i.e., changes in the aging tendency, may occur only at the failure times of other components. The failure of a component may then produce a shock for a different component. However, not necessarily a component is affected by the failure of another one. For instance, if the components are independent the continuity of the MCAI functions is guaranteed also under failures. In the following proposition, an expression for the size of the jump discontinuity is given.

Proposition 5.7. *Let $(X_1,\dots,X_n)$ be a random vector with non-negative components. Let $t_1,\dots,t_k$ be the failure times of the components $i_1,\dots,i_k$, respectively. Then, the size of the jump discontinuity at t_k of the MCAI function of component $j\notin I=\{i_1,\dots,i_k\}$, is given by*

$$\begin{aligned}
&L_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)-L_j(t_k|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})\\
&=\frac{\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)-\lambda_j(i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})}{\frac{1}{t_k}\sum_{r=0}^{k-1}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx}. \qquad (5.28)
\end{aligned}$$

Proof. From the definition of MCAI functions, we have

$$
\begin{aligned}
&L_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k) - L_j(t_k|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1}) \\
&= \frac{\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)}{\frac{1}{t_k}\sum_{r=0}^{k}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx} - \frac{\lambda_j(i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})}{\frac{1}{t_k}\sum_{r=0}^{k-1}\int_{t_r}^{t_{r+1}}\lambda_j(x|i_1,\dots,i_r;t_1,\dots,t_r)dx},
\end{aligned}
$$

hence, by observing that t_{k+1} is the point of evaluation of the MCAI function and then it is equal to t_k, the ratios in the above equation have a common denominator and the thesis follows. □

The result of Proposition 5.7 agrees with that of Proposition 5.6 since if $\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k) = \lambda_j(t_k|i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})$ then the size of the jump is zero. Moreover, from the expression given in (5.28), it follows that the sign of the jump is determined by the difference $\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k)-\lambda_j(i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})$, since the denominator in (5.28) is positive. Hence, the jump is upward if $\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k) > \lambda_j(i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})$, i.e., if the failure of the component i_k at time t_k increases the hazard of component j, and downward if $\lambda_j(t_k|i_1,\dots,i_k;t_1,\dots,t_k) < \lambda_j(i_1,\dots,i_{k-1};t_1,\dots,t_{k-1})$.

5.3.1 The bivariate case

In the applications, there are several situations in which a model can be described by two random variables with a certain mode of dependence. Hence, it is of interest to specialize the concept of MCAI functions for bivariate distributions. In the literature, it has been already presented a definition of the bivariate aging intensity function (see [115]). We remark that this definition is different from the one considered here since it is based on the failure rates gradient defined in [57]. For a random vector (X_1, X_2) with joint survival function $\overline{F}(\cdot,\cdot)$, the failure rates gradient is defined as $(r_1(t_1,t_2), r_2(t_1,t_2))$ where

$$
r_1(t_1,t_2) = -\frac{\partial}{\partial t_1}\log\overline{F}(t_1,t_2), \quad r_2(t_1,t_2) = -\frac{\partial}{\partial t_2}\log\overline{F}(t_1,t_2). \tag{5.29}
$$

Hence, the bivariate aging intensity functions defined in [115] are

$$
\mathcal{L}_1(t_1,t_2) = \frac{r_1(t_1,t_2)}{\frac{1}{t_1}\int_0^{t_1} r_1(x,t_2)dx}, \quad \mathcal{L}_2(t_1,t_2) = \frac{r_2(t_1,t_2)}{\frac{1}{t_2}\int_0^{t_2} r_1(t_1,x)dx}. \tag{5.30}
$$

As one can see, the above definition does not take in account the possibility of observing a dynamic history. In the following, based on [22], we extend the concept of bivariate aging intensity by considering stochastic dependence and the possibility of observing a dynamic history. For a random vector of dimension two, (X_1, X_2), we have to consider four aging intensity functions depending on how many variables and which ones assume a value greater than t. If $X_1 > t$ and $X_2 = t_2 < t$, we consider

$$
L_1(t|2;t_2) = \frac{t\lambda_1(t|2;t_2)}{\int_0^{t_2}\lambda_1(x|\emptyset)dx + \int_{t_2}^{t}\lambda_1(x|2;t_2)dx},
$$

if $X_2 > t$ and $X_1 = t_1 < t$, we have

$$L_2(t|1;t_1) = \frac{t\lambda_2(t|1;t_1)}{\int_0^{t_1} \lambda_2(x|\emptyset)dx + \int_{t_1}^{t} \lambda_2(x|1;t_1)dx},$$

and if $X_1, X_2 > t$, we consider

$$L_j(t|\emptyset) = \frac{t\lambda_j(t|\emptyset)}{\int_0^t \lambda_j(x|\emptyset)dx}, \quad j = 1,2.$$

In the case in which $t_1 \geq t_2$, the joint pdf $f(t_1,t_2)$ can be expressed in terms of the multivariate conditional hazard rate functions as

$$f(t_1,t_2) = \lambda_2(t_2|\emptyset)\lambda_1(t_1|2;t_2)\exp\left[-\int_0^{t_2} (\lambda_1(u|\emptyset) + \lambda_2(u|\emptyset))\, du - \int_{t_2}^{t_1} \lambda_1(u|2;t_2)du\right]. \quad (5.31)$$

From (5.31) we get

$$\log\left(\frac{f(t_1,t_2)}{\lambda_2(t_2|\emptyset)\lambda_1(t_1|2;t_2)}\right) = -\int_0^{t_2} (\lambda_1(u|\emptyset) + \lambda_2(u|\emptyset))\, du - \int_{t_2}^{t_1} \lambda_1(u|2;t_2)du,$$

and then

$$\int_0^{t_2} \lambda_1(u|\emptyset)du + \int_{t_2}^{t_1} \lambda_1(u|2;t_2)du = -\left[\int_0^{t_2} \lambda_2(u|\emptyset)du + \log\left(\frac{f(t_1,t_2)}{\lambda_2(t_2|\emptyset)\lambda_1(t_1|2;t_2)}\right)\right],$$

where we can observe that the LHS is the denominator of $L_1(t|2;t_2)$. Hence, we can express $L_1(t|2;t_2)$ in a different way as

$$L_1(t|2;t_2) = \frac{-t\lambda_1(t|2;t_2)}{\int_0^{t_2} \lambda_2(u|\emptyset)du + \log\left(\frac{f(t,t_2)}{\lambda_2(t_2|\emptyset)\lambda_1(t|2;t_2)}\right)}. \quad (5.32)$$

Then, by taking into account the relations among the multivariate conditional hazard rate functions, the joint density function and the joint survival function (see (4.5)–(4.8)), the MCAI function can be written as

$$L_1(t|2;t_2) = \frac{\frac{tf(t,t_2)}{\frac{\partial}{\partial t_2}\overline{F}(t,t_2)}}{\int_0^{t_2} \frac{-\frac{\partial}{\partial t_2}\overline{F}(u,t_2)\big|_{t_2=u}}{\overline{F}(u,u)}du + \log\left(\overline{F}(t_2,t_2)\frac{\frac{\partial}{\partial t_2}\overline{F}(t,t_2)}{\frac{\partial}{\partial v}\overline{F}(t_2,v)\big|_{v=t_2}}\right)}. \quad (5.33)$$

In a similar manner, about $L_1(t|\emptyset)$, we get the following expression

$$L_1(t|\emptyset) = \frac{-t\frac{\frac{\partial}{\partial t_1}\overline{F}(t_1,t)\big|_{t_1=t}}{\overline{F}(t,t)}}{\int_0^t \frac{-\frac{\partial}{\partial t_1}\overline{F}(t_1,u)\big|_{t_1=u}}{\overline{F}(u,u)}du}. \quad (5.34)$$

The above expressions in (5.33)–(5.34) are useful in the sense that they give the MCAI functions without involving the m.c.h.r. functions which may be of difficult evaluation and they are based only on the joint probability density and survival functions. In the following example, the MCAI functions are obtained for a family of bivariate distributions by applying (5.33)–(5.34).

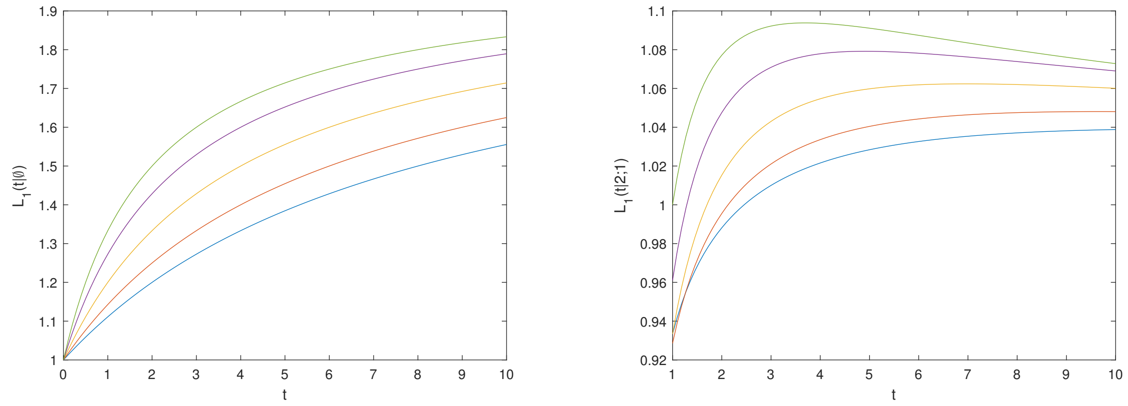

Figure 5.3: Plot of $L_1(t|\emptyset)$ (left) and $L_1(t|2;1)$ (right) with $\theta = 0.25$ (blue), $1/3$ (red), 0.5 (yellow), 0.75 (violet) and 1 (green).

Example 5.7. Let us obtain the MCAI functions of a well-known bivariate distribution, the Gumbel's type I bivariate exponential distribution with parameter $\theta \in [0,1]$ (see Example 4.1, where the expression of the joint distribution, density and survival functions are given). About the failure rates gradient (5.29), we have

$$r_1(x,y) = 1 + \theta y, \quad r_2(x,y) = 1 + \theta x,$$

and then from (5.30) we obtain

$$\mathcal{L}_1(x,y) = 1, \quad \mathcal{L}_2(x,y) = 1. \tag{5.35}$$

Now, we aim to compute the bivariate aging intensity functions based on m.c.h.r. functions. We use (5.33) to evaluate the aging intensity function $L_1(t|2;t_2)$ and we get

$$L_1(t|2;t_2) = \frac{-t\left[(1+\theta t)(1+\theta t_2) - \theta\right]}{(1+\theta t)\left(\frac{\theta t_2^2}{2} - t - \theta t t_2 + \log\left(\frac{1+\theta t}{1+\theta t_2}\right)\right)}. \tag{5.36}$$

If $\theta = 0$ we are in the independent case and (5.36) reduces to $L_1(t|2;t_2) = 1$ as the aging intensity function of the exponential distribution is equal to 1. By using (5.34) we can express $L_1(t|\emptyset)$ as

$$L_1(t|\emptyset) = \frac{2(1+\theta t)}{2+\theta t}. \tag{5.37}$$

In Figure 5.3 we plot the aging intensity functions related to component 1 for different choices of θ. For $L_1(t|2;t_2)$ we choose the value $t_2 = 1$ and so the function are plotted for $t \geq 1$.

The size of the jump at time t_2 for the MCAI function of component 1 is given by

$$\begin{aligned}
L_1(t_2|2;t_2) - L_1(t_2|\emptyset) &= \frac{-t_2\left[(1+\theta t_2)^2-\theta\right]}{(1+\theta t_2)\left(\frac{\theta t_2^2}{2}-t_2-\theta t_2^2\right)} - \frac{2(1+\theta t_2)}{2+\theta t_2} \\
&= \frac{2[(1+\theta t_2)^2-\theta]}{(1+\theta t_2)(2+\theta t_2)} - \frac{2(1+\theta t_2)}{2+\theta t_2} \\
&= \frac{-2\theta}{(1+\theta t_2)(2+\theta t_2)},
\end{aligned}$$

i.e., it is a negative jump with the exception of the case $\theta = 0$ in which there are the independence and the continuity of the MCAI function.

5.3.2 Aging intensities for Load-Sharing models

In this section, we focus attention on a generalization of the time-homogeneous load-sharing (THLS) models which was recently introduced in Foschi et al. [46].

Definition 5.5. Let $(X_1,\dots,X_n)$ be a random vector with absolutely continuous joint distribution. It is distributed according to an Order Dependent Load-Sharing model (ODLS) if, for any $i_1,\dots,i_k \in [n]$ and $j \notin I = \{i_1,\dots,i_k\}$, there exist functions $\mu_j(t|i_1,\dots,i_k)$ such that, for all $0 \le t_1 \le \dots \le t_k \le t$,

$$\lambda_j(t|i_1,\dots,i_k;t_1,\dots,t_k) = \mu_j(t|i_i\dots,i_k).$$

Furthermore, an order dependent load-sharing model is time-homogeneous (ODTHLS) when there exist non-negative numbers $\mu_j(i_1,\dots,i_k)$ and $\mu_j(\emptyset)$ such that, for any $t>0$ and any $j \notin I$,

$$\begin{aligned}
&\mu_j(t|i_1,\dots,i_k) = \mu_j(i_1,\dots,i_k), \\
&\lambda_j(t|\emptyset) = \mu_j(\emptyset).
\end{aligned}$$

Remark 5.8. If for any non-empty set $I \subset [n]$ and any $j \notin I$, the function $\mu_j(t|i_1,\dots,i_k)$ is invariant under permutations of $i_1,\dots,i_k$, then the ODLS model reduces to a LS model. In the same way, if for any non-empty set $I \subset [n]$ and any $j \notin I$ the number $\mu_j(i_1,\dots,i_k)$ is invariant under permutations of $i_1,\dots,i_k$, then the ODTHLS model reduces to a THLS model.

Let $(X_1,\dots,X_n)$ be distributed according to an ODTHLS model, then the MCAI functions can be expressed as

$$\begin{aligned}
L_j(t|i_1,\dots,i_k;t_1,\dots,t_k) &= \frac{t\mu_j(i_1,\dots,i_k)}{\sum_{h=0}^k \int_{t_h}^{t_{h+1}} \mu_j(i_1,\dots,i_h)dx} = \frac{t\mu_j(I)}{\sum_{h=0}^k (t_{h+1}-t_h)\mu_j(i_1,\dots,i_h)} \\
&= \frac{t\mu_j(i_1,\dots,i_k)}{t_1\mu_j(\emptyset)+(t_2-t_1)\mu_j(i_1)+\dots+(t_k-t_{k-1})\mu_j(i_1,\dots,i_{k-1})+(t-t_k)\mu_j(i_1,\dots,i_k)}.
\end{aligned}$$

In the case in which $X_{1:n} > t$ we have to consider $L_j(t|\emptyset)$ that is given by

$$L_j(t|\emptyset) = \frac{\mu_j(\emptyset)}{\frac{1}{t}t\mu_j(\emptyset)} = 1. \tag{5.38}$$

Proposition 5.8. *$L_j(t|\emptyset) = 1$ for all $t > 0$, $j = 1, \dots, n$ if, and only if, $\lambda_j(t|\emptyset)$ is constant for all j with respect to t.*

Proof. If $\lambda_j(t|\emptyset)$ is constant for all j, then we have shown in (5.38) that the MCAI functions related to the empty set are constant and equal to 1. Conversely, let us suppose $L_j(t|\emptyset) = 1$ for all $t > 0$, $j = 1, \dots, n$. Then, from (5.26) we have

$$\frac{\lambda_j(t|\emptyset)}{\frac{1}{t}\int_0^t \lambda_j(x|\emptyset)dx} = 1$$

and so it follows

$$\lambda_j(t|\emptyset) = \frac{1}{t}\int_0^t \lambda_j(x|\emptyset)dx.$$

By the mean value theorem for definite integrals we get $\lambda_j(t|\emptyset) = \lambda_j(\tilde{t}|\emptyset)$, where $\tilde{t} \in (0, t)$. Hence, $\lambda_j(t|\emptyset)$ cannot be strictly monotone in an arbitrary small interval and then it has to be constant. $\square$

In the following theorem, a characterization of ODTHLS models is given in terms of MCAI functions. In particular, in this case the MCAI functions are constant and equal to 1 or hyperbolas.

Theorem 5.7. *Let $(X_1, \dots, X_n)$ be a random vector with non-negative components. Then, $(X_1, \dots, X_n)$ is distributed according to an ODTHLS model if, and only if, the MCAI functions can be expressed as*

$$\begin{aligned} L_j(t|\emptyset) &= 1, \\ L_j(t|i_1, \dots, i_k; t_1, \dots, t_k) &= \frac{t}{(t - t_k) + C(i_1, \dots, i_k; t_1, \dots, t_k)}, \end{aligned}$$

where $C(i_1, \dots, i_k; t_1, \dots, t_k) = \frac{t_1\mu_j(\emptyset)+(t_2-t_1)\mu_j(i_1)+\cdots+(t_k-t_{k-1})\mu_j(i_1,\dots,i_{k-1})}{\mu_j(I)} > 0$ is constant with respect to t.

Proof. If $(X_1, \dots, X_n)$ is distributed according to an ODTHLS model, then $L_j(t|\emptyset) = 1$ and

$$\begin{aligned} &L_j(t|i_1, \dots, i_k; t_1, \dots, t_k) \\ &= \frac{t\mu_j(i_1, \dots, i_k)}{t_1\mu_j(\emptyset) + (t_2 - t_1)\mu_j(i_1) + \cdots + (t_k - t_{k-1})\mu_j(i_1, \dots, i_{k-1}) + (t - t_k)\mu_j(i_1, \dots, i_k)} \\ &= \frac{t}{(t - t_k) + \frac{t_1\mu_j(\emptyset)+(t_2-t_1)\mu_j(i_1)+\cdots+(t_k-t_{k-1})\mu_j(i_1,\dots,i_{k-1})}{\mu_j(i_1,\dots,i_k)}}, \end{aligned}$$

and then, by letting $C(i_1,\ldots,i_k;t_1,\ldots,t_k) = \frac{t_1\mu_j(\emptyset)+(t_2-t_1)\mu_j(i_1)+\cdots+(t_k-t_{k-1})\mu_j(i_1,\ldots,i_{k-1})}{\mu_j(i_1,\ldots,i_k)}$, we get the result.

Conversely, by Proposition 5.8, if $L_j(t|\emptyset) = 1$ for all $t > 0$, $j = 1,\ldots,n$, then $\lambda_j(t|\emptyset) = \mu_j(\emptyset)$ is constant for all j. Let us now consider the case in which $|I| = 1$. We have

$$L_j(t|i;t_1) = \frac{t\lambda_j(t|i;t_1)}{\int_0^{t_1}\lambda_j(x|\emptyset)dx + \int_{t_1}^t \lambda_j(x|i;t_1)dx} = \frac{t\lambda_j(t|i;t_1)}{t_1\mu_j(\emptyset) + \int_{t_1}^t \lambda_j(x|i;t_1)dx},$$

and then by the assumptions

$$\frac{t\lambda_j(t|i;t_1)}{t_1\mu_j(\emptyset) + \int_{t_1}^t \lambda_j(x|i;t_1)dx} = \frac{t}{(t-t_1) + C(i;t_1)}. \tag{5.39}$$

From (5.39), we get

$$(t-t_1)\lambda_j(t|i;t_1) + C(i;t_1)\lambda_j(t|i;t_1) = t_1\mu_j(\emptyset) + \int_{t_1}^t \lambda_j(x|i;t_1)dx, \tag{5.40}$$

and, by differentiating both sides of (5.40) with respect to t, we obtain

$$\lambda_j(t|i;t_1) + \lambda_j'(t|i;t_1) + C(i;t_1)\lambda_j'(t|i;t_1) = \lambda_j(t|i;t_1),$$

that is equivalent to

$$(1 + C(i;t_1))\lambda_j'(t|i;t_1) = 0. \tag{5.41}$$

By observing that $C(i;t_1) > 0$, in order to satisfy (5.41), $\lambda_j(t|i;t_1)$ has to be constant, $\lambda_j(t|i;t_1) = \mu_j(i)$. Moreover, if in (5.40) we take the limit $t \to t_1^+$ we get

$$C(i;t_1) = \frac{t_1\mu_j(\emptyset)}{\mu_j(i)}.$$

By induction, we consider the case in which $|I| > 1$ and obtain

$$\frac{t\mu_j(i_1,\ldots,i_k)}{t_1\mu_j(\emptyset) + (t_2-t_1)\mu_j(i_1) + \cdots + (t_k-t_{k-1})\mu_j(i_1,\ldots,i_{k-1}) + \int_{t_k}^t \lambda_j(x|i_1,\ldots,i_k;t_1,\ldots,t_k)dx}$$
$$= \frac{t}{(t-t_k) + C(i_1,\ldots,i_k;t_1,\ldots,t_k)}.$$

By following the same steps of the case $|I| = 1$, we conclude that $\lambda_j(t|i_1,\ldots,i_k;t_1,\ldots,t_k)$ has to be constant,

$$\lambda_j(t|i_1,\ldots,i_k;t_1,\ldots,t_k) = \mu_j(i_1,\ldots,i_k),$$
$$C(i_1,\ldots,i_k;t_1,\ldots,t_k) = \frac{t_1\mu_j(\emptyset) + (t_2-t_1)\mu_j(i_1) + \cdots + (t_k-t_{k-1})\mu_j(i_1,\ldots,i_{k-1})}{\mu_j(i_1,\ldots,i_k)}.$$

□

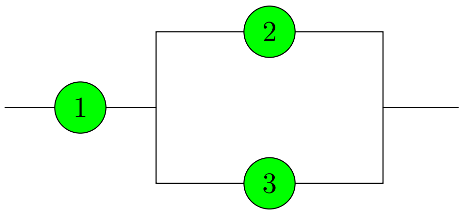

Figure 5.4: The structure of the system S.

It is of interest to study what happens for the MCAI functions of surviving components in the failure time of other ones. From Proposition 5.7, the sign of the size of the jump for ODTHLS model is determined by the difference $\mu_j(i_1,\ldots,i_k)-\mu_j(i_1,\ldots,i_{k-1})$, and, in particular, the continuity of the MCAI function is given by the condition $\mu_j(i_1,\ldots,i_k)=\mu_j(i_1,\ldots,i_{k-1})$.

In the following, we show an application of MCAI functions by considering coherent systems whose lifetimes are distributed according to ODTHLS models. We use the MCAI functions to make comparisons among surviving components and to discover which component ages faster than the others.

Let us consider a coherent system S formed by three components X_1, X_2, X_3. The structure of the system is displayed in Figure 5.4 and the lifetime T_S is described as

$$T_S=\min\{X_1,\max\{X_2,X_3\}\}.$$

Let us suppose that the component 2 failed at time t_1 and that at time $t>t_1$ the components 1 and 3 are still working, i.e., the system is still working. Moreover, (X_1,X_2,X_3) is distributed according to an ODTHLS model and the parameters of interest are expressed as

$$\mu_1(\emptyset)=2,\quad \mu_3(\emptyset)=1,\quad \mu_1(2)=2,\quad \mu_3(2)=2.$$

The aging intensity functions of components 1 and 3 at time t are expressed as

$$\begin{aligned}L_1(t|2;t_1) &= \frac{2}{\frac{1}{t}[2t_1+2(t-t_1)]}=1,\\ L_3(t|2;t_1) &= \frac{2}{\frac{1}{t}[t_1+2(t-t_1)]}=\frac{2t}{2t-t_1}.\end{aligned}$$

Then, we have

$$L_3(t|2;t_1)>L_1(t|2;t_1)\Leftrightarrow \frac{2t}{2t-t_1}>1\Leftrightarrow 2t>2t-t_1\Leftrightarrow t_1>0,$$

and so the component 3 suffers more than 1 the failure of component 2 by aging faster. Moreover, we can observe that the MCAI function of component 1 is constantly equal to 1 and hence continuous also at time t_1, in fact $\mu_1(\emptyset)=\mu_1(2)$. Furthermore, since $\mu_3(2)>\mu_3(\emptyset)$, we expect an upward jump for the MCAI function of component 3 at time t_1, that is

$$L_3(t_1|2;t_1)-L_3(t_1|\emptyset) = \frac{2t_1}{2t_1-t_1}-1=1.$$

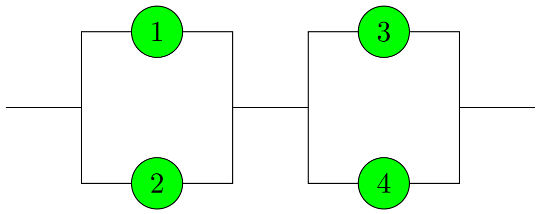

Figure 5.5: The structure of the system S^*.

We can perform comparisons among surviving components without fixing the values of parameters. In this case, about the aging intensities, we have

$$\begin{aligned} L_1(t|2;t_1) &= \frac{t\mu_1(2)}{\mu_1(\emptyset)t_1 + \mu_1(2)(t-t_1)}, \\ L_3(t|2;t_1) &= \frac{t\mu_3(2)}{\mu_3(\emptyset)t_1 + \mu_3(2)(t-t_1)}. \end{aligned}$$

Then, we can compare the aging intensities as

$$\begin{aligned} L_3(t|2;t_1) > L_1(t|2;t_1) &\Leftrightarrow \frac{t\mu_3(2)}{\mu_3(\emptyset)t_1 + \mu_3(2)(t-t_1)} > \frac{t\mu_1(2)}{\mu_1(\emptyset)t_1 + \mu_1(2)(t-t_1)} \\ &\Leftrightarrow \frac{\mu_3(2)}{\mu_3(\emptyset)t_1 + \mu_3(2)(t-t_1)} > \frac{\mu_1(2)}{\mu_1(\emptyset)t_1 + \mu_1(2)(t-t_1)} \\ &\Leftrightarrow \frac{\mu_3(\emptyset)t_1 + \mu_3(2)(t-t_1)}{\mu_3(2)} < \frac{\mu_1(\emptyset)t_1 + \mu_1(2)(t-t_1)}{\mu_1(2)} \\ &\Leftrightarrow \frac{\mu_3(\emptyset)t_1}{\mu_3(2)} + (t-t_1) < \frac{\mu_1(\emptyset)t_1}{\mu_1(2)} + (t-t_1) \\ &\Leftrightarrow \frac{\mu_3(\emptyset)}{\mu_3(2)} < \frac{\mu_1(\emptyset)}{\mu_1(2)}. \end{aligned}$$

We can observe that the comparison is not dependent on t_1 and t. As we will see with a further example, when we compare the aging intensities of ODTHLS components the dependence of t is always lost whereas if the number of failed components is greater than one, the times of failure will be involved in the comparisons.

Let us consider a coherent system S^* formed by four components X_1, X_2, X_3, X_4, with structure displayed in Figure 5.5, and whose lifetime T_{S^*} is described as

$$T_{S^*} = \min\{\max\{X_1, X_2\}, \max\{X_3, X_4\}\}.$$

Suppose that the component 2 failed at time t_1, the component 4 failed at time $t_2 > t_1$ and that at time $t > t_2$ the components 1 and 3 are still working, that is $T_{S^*} > t$. Moreover, (X_1, X_2, X_3, X_4) is distributed according to an ODTHLS model. The aging intensity functions of components 1 and 3 at time t are given as

$$\begin{aligned} L_1(t|2,4;t_1,t_2) &= \frac{t\mu_1(2,4)}{\mu_1(\emptyset)t_1 + \mu_1(2)(t_2-t_1) + \mu_1(2,4)(t-t_2)}, \\ L_3(t|2,4;t_1,t_2) &= \frac{t\mu_3(2,4)}{\mu_3(\emptyset)t_1 + \mu_3(2)(t_2-t_1) + \mu_3(2,4)(t-t_2)}. \end{aligned}$$

Then, the aging intensities can be compared as

$$L_3(t|2,4;t_1,t_2) > L_1(t|2,4;t_1,t_2)$$

$$\Leftrightarrow \frac{\mu_3(2,4)}{\mu_3(\emptyset)t_1 + \mu_3(2)(t_2-t_1) + \mu_3(2,4)(t-t_2)} > \frac{\mu_1(2,4)}{\mu_1(\emptyset)t_1 + \mu_1(2)(t_2-t_1) + \mu_1(2,4)(t-t_2)}$$

$$\Leftrightarrow t_1\frac{\mu_3(\emptyset)}{\mu_3(2,4)} + (t_2-t_1)\frac{\mu_3(2)}{\mu_3(2,4)} < t_1\frac{\mu_1(\emptyset)}{\mu_1(2,4)} + (t_2-t_1)\frac{\mu_1(2)}{\mu_1(2,4)}.$$

Chapter 6

Predicting future failure times from censored data

One of the most relevant topics in Probability and Statistics during the last decades has been the use of censored data. When in a study one works with a sample of several lifetimes, it is common to have censored data, i.e., just the exact values of the first r failures (or survival times). The other values are censored (type II censored data). This approach is of interest both in Survival and Reliability studies. An interesting review of the different situations about ordered and censored data was made in Cramer [32].

Several tools have been developed to use censored data. Also, some procedures have been studied to predict the unknown future failure times by using the exact values of the r early failures. The main results can be seen in [9, 15, 44] and in the references therein. Recently, Barakat et al. [10] and Bdair and Raqab [16] proposed two methods based on different pivotal quantities for samples of independent and identically distributed (IID) lifetimes with a common mixture of two exponential distributions.

In this chapter, we extend these results considering the IID case with the more general Proportional Hazard Rate (PHR) Cox model and also the case of dependent samples. In both cases, to provide such predictions we will use quantile regression (QR) techniques that can also be used to get prediction bands for them, see [64, 84]. The results are given in Navarro and Buono [83]. Then, some properties of the order dependent time-homogeneous load-sharing model are proposed and an algorithmic procedure to simulate samples from this model is explained. The problem of the predictions of the future failure times in a sample from censored data is analyzed for this model (see Buono and Navarro [28]).

6.1 Predictions by using quantile regression

6.1.1 Independent data

Let $X_1,\dots,X_n$ be a sample of independent and identically distributed (IID) random variables with a common absolutely continuous distribution function F and with a probability density function (pdf) f. Let $\overline{F}$ be the reliability (or survival) function and let $X_{1:n}<\cdots<X_{n:n}$ be the associated ordered data (order statistics).

Consider $X_1,\dots,X_n$ as lifetimes (or survival times) of some items. In practice, sometimes we just have the first r early failure times $X_{1:n},\dots,X_{r:n}$ for some $r<n$. Then, we want to predict the remaining lifetimes $X_{r+1:n},\dots,X_{n:n}$ from the early failures.

The results obtained in this section are based on the following proposition extracted from Bdair and Raqab [16] and the well known Markov property of the order statistics (see [4]).

Proposition 6.1. *Let $W_{r,s:n}=\overline{F}(X_{s:n})/\overline{F}(X_{r:n})$ for $1\le r<s\le n$. Then the distributions of the conditional random variables $(W_{r,s:n}|X_{1:n}=x_1,\dots,X_{r:n}=x_r)$ and $(W_{r,s:n}|X_{r:n}=x_r)$ coincide with a beta distribution of parameters $n-s+1$ and $s-r$.*

Proof. The distributions coincide from Theorem 2.4.3 in [4]. From expression (2.4.3) in that book (p. 23), the pdf of $(X_{s:n}|X_{r:n}=x_r)$ is

$$f_{s|r:n}(x_s|x_r)=c\left(\frac{\overline{F}(x_r)-\overline{F}(x_s)}{\overline{F}(x_r)}\right)^{s-r-1}\left(\frac{\overline{F}(x_s)}{\overline{F}(x_r)}\right)^{n-s}\frac{f(x_s)}{\overline{F}(x_r)}$$

for $1\le r<s\le n$ and $x_r<x_s$, where c is the normalizing constant. On the other hand, if $\overline{G}$ is the reliability function of $(W_{r,s:n}|X_{r:n}=x_r)$, we get

$$\begin{aligned}\overline{G}\left(\frac{\overline{F}(x_s)}{\overline{F}(x_r)}\right)&=\mathbb{P}\left(W_{r,s:n}>\frac{\overline{F}(x_s)}{\overline{F}(x_r)}\middle|\,X_{r:n}=x_r\right)\\&=\mathbb{P}\left(\frac{\overline{F}(X_{s:n})}{\overline{F}(X_{r:n})}>\frac{\overline{F}(x_s)}{\overline{F}(x_r)}\middle|\,X_{r:n}=x_r\right)=\mathbb{P}\left(X_{s:n}<x_s|X_{r:n}=x_r\right).\end{aligned}$$

Therefore, its pdf $g=-\overline{G}'$ satisfies

$$g\left(\frac{\overline{F}(x_s)}{\overline{F}(x_r)}\right)\frac{f(x_s)}{\overline{F}(x_r)}=f_{s|r:n}(x_s|x_r)$$

and so, by using the preceding expression for $f_{s|r:n}$, we obtain

$$g(w)=c(1-w)^{s-r+1}w^{n-s}$$

for $0<w<1$. Therefore, $(W_{r,s:n}|X_{r:n}=x_r)$ has a beta distribution with parameters $n-s+1$ and $s-r$. □

It is possible to use the preceding proposition to get the median regression curve to estimate $X_{s:n}$ from $X_{r:n}=x$ (or from $X_{1:n}=x_1,\dots,X_{r:n}=x_r$) as stated in the following proposition.

Proposition 6.2. *The median regression curve to estimate $X_{s:n}$ from $X_{r:n} = x$ is*

$$m(x) = \overline{F}^{-1}\left(q_{0.5}\overline{F}(x)\right), \tag{6.1}$$

where $q_{0.5}$ is the median of a beta distribution with parameters $n-s+1$ and $s-r$.

Proof. From the expressions obtained in the proof of Proposition 6.1 we get that

$$\mathbb{P}\left(X_{s:n} < x_s \mid X_{r:n} = x_r\right) = 0.5$$

is equivalent to

$$\overline{G}\left(\frac{\overline{F}(x_s)}{\overline{F}(x_r)}\right) = 0.5,$$

where $\overline{G}$ is the reliability function of a beta distribution with parameters $n-s+1$ and $s-r$. This expression leads to $\overline{F}(x_s) = q_{0.5}\overline{F}(x_r)$. Therefore, the first expression is equivalent to $x_s = \overline{F}^{-1}(q_{0.5}\overline{F}(x_r))$ which gives the expression for the median regression curve. □

Along the same lines, it is possible to determine quantile predictions bands for these predictions by using the quantiles of a beta distribution. If we want to get a prediction interval of size $\gamma = \beta - \alpha$, where $\alpha, \beta, \gamma \in (0,1)$ and q_α and q_β are the respective quantiles of the above beta distribution, we use that

$$\mathbb{P}\left(\overline{F}^{-1}\left(q_\beta\overline{F}(x)\right) \le X_{s:n} \le \overline{F}^{-1}\left(q_\alpha\overline{F}(x)\right) | X_{r:n} = x\right) = \gamma. \tag{6.2}$$

For instance, the centered 90% prediction band is obtained with $\beta = 0.95$ and $\alpha = 0.05$ as

$$C_{90} = \left[\overline{F}^{-1}\left(q_{0.95}\overline{F}(x)\right), \overline{F}^{-1}\left(q_{0.05}\overline{F}(x)\right)\right].$$

Sometimes it is preferred to use bottom (or lower) prediction bands starting at $X_{r:n} = x$. For example, the bottom 90% prediction band is obtained with $\beta \to 1$ and $\alpha = 0.1$ as

$$B_{90} = \left[x, \overline{F}^{-1}\left(q_{0.1}\overline{F}(x)\right)\right].$$

As it will be pointed out in the examples, these prediction bands represent better the uncertainty in the prediction of $X_{s:n}$ from $X_{r:n}$. In particular, the area of these bands will increase with $s-r$. We recall that the quantiles q_z of a beta distribution (including the median) are available in many statistical programs (for example, in R, q_z is obtained with the code `qbeta(z,a,b)`, with $a = n-s+1$ and $b = s-r$ in our case).

In the following proposition the exponential distribution is characterized in terms of its quantile regression curves. In fact, it is the unique distribution with quantile regression curves that are lines with slope equal to one.

Proposition 6.3. *Let X be a non-negative and absolutely continuous random variable with survival function $\overline{F}$ and $m_\alpha(x) = \overline{F}^{-1}(q_\alpha \overline{F}(x))$ quantile regression curve of order α. Then,*

$$m_\alpha(x) = x + c_\alpha \tag{6.3}$$

for all $x > 0$ and $\alpha \in (0,1)$ if, and only if, X is exponentially distributed, where c_α is a constant depending on α.

Proof. Suppose X is exponentially distributed with parameter λ. Then, $\overline{F}(x) = e^{-\lambda x}$ and $\overline{F}^{-1}(x) = -\frac{1}{\lambda}\log x$. Therefore, the quantile regression curve of order α is given by

$$m_\alpha(x) = \overline{F}^{-1}(q_\alpha \overline{F}(x)) = \overline{F}^{-1}\left(q_\alpha e^{-\lambda x}\right) = x - \frac{1}{\lambda}\log q_\alpha.$$

Conversely, suppose Equation (6.3) holds. By choosing $x = 0$, we get $\overline{F}^{-1}(q_\alpha) = c_\alpha$, or, equivalently, $q_\alpha = \overline{F}(c_\alpha)$. By applying $\overline{F}$ to both sides of (6.3), it readily follows

$$\overline{F}(x) = \frac{\overline{F}(x + c_\alpha)}{q_\alpha},$$

that is

$$\mathbb{P}(X > x) = \frac{\mathbb{P}(X > x + c_\alpha)}{\mathbb{P}(X > c_\alpha)} = \mathbb{P}(X > x + c_\alpha | X > c_\alpha),$$

for all $x > 0$, i.e., the memoryless property which characterizes the exponential distribution. □

In practice, the common reliability function $\overline{F}$ is unknown. Sometimes, it can be estimated from historical data sets by using non-parametric estimators. In these cases, we just replace in the preceding expressions the exact unknown reliability function $\overline{F}$ with its estimation. In other cases, we may have a model for it. Consider $\overline{F}_\theta$ with an unknown parameter θ. We can use $X_{1:n} = x_1, \ldots, X_{r:n} = x_r$ to estimate θ. The associated likelihood function is

$$\ell(\theta) = \frac{n!}{(n-r)!}\overline{F}_\theta^{n-r}(x_r)\prod_{i=1}^{r} f_\theta(x_i)$$

(see, e.g., [4] or (5) in [16]). By maximizing this function we obtain a good estimation for θ.

Assume the Proportional Hazard Rate (PHR) Cox model with $\overline{F}_\theta = \overline{F}_0^\theta$, where $\overline{F}_0$ is a known baseline reliability function and $\theta > 0$ is an unknown risk parameter. Then,

$$\ell(\theta) = \frac{n!}{(n-r)!}\theta^r \overline{F}_0^{(n-r)\theta}(x_r)\prod_{i=1}^{r}\overline{F}_0^{\theta-1}(x_i)\prod_{i=1}^{r} f_0(x_i)$$

and so $L(\theta) = \log \ell(\theta)$ can be written as

$$L(\theta) = K + r\log\theta + (n-r)\theta\log\overline{F}_0(x_r) + (\theta - 1)\sum_{i=1}^{r}\log\overline{F}_0(x_i),$$

where K is a constant with respect to θ. Hence, its derivative is

$$L'(\theta) = \frac{r}{\theta} + (n-r)\log \overline{F}_0(x_r) + \sum_{i=1}^{r} \log \overline{F}_0(x_i)$$

and the maximum likelihood estimator (MLE) for θ is

$$\widehat{\theta} = \frac{r}{-(n-r+1)\log \overline{F}_0(x_r) - \sum_{i=1}^{r-1} \log \overline{F}_0(x_i)}. \tag{6.4}$$

Thus, to get the predictions and the prediction bands for $X_{s:n}$, we just replace in the preceding expressions $\overline{F}$ with $\overline{F}_{\widehat{\theta}}$.

Some well known distributions are included in the PHR model. For example, if $\overline{F}_0(t) = e^{-t}$ for $t \geq 0$ (Exponential model), then (6.4) leads to

$$\widehat{\theta} = \frac{r}{(n-r+1)x_r + \sum_{i=1}^{r-1} x_i}. \tag{6.5}$$

Analogously, the mean $\mu = 1/\theta$ can be estimated by

$$\widehat{\mu} = \frac{1}{\widehat{\theta}} = \frac{n-r+1}{r} x_r + \frac{1}{r}\sum_{i=1}^{r-1} x_i$$

(a well known result, see e.g. [32]). Along the same lines, $\overline{F}_0(t) = 1/(1+t)$ for $t \geq 0$ (Pareto type II model), leads to

$$\widehat{\theta} = \frac{r}{(n-r+1)\log(1+x_r) + \sum_{i=1}^{r-1} \log(1+x_i)}. \tag{6.6}$$

Finally, note that the prediction regions obtained from different quantiles can be used to get bivariate box plots and fit tests for the assumed reliability function $\overline{F}$ (or $\overline{F}_\theta$), see [81]. For example, to get equal expected values, we consider the regions:

$$R_1 = \left[x, \overline{F}^{-1}\left(q_{0.75}\overline{F}(x)\right)\right], \qquad R_2 = \left[\overline{F}^{-1}\left(q_{0.75}\overline{F}(x)\right), \overline{F}^{-1}\left(q_{0.50}\overline{F}(x)\right)\right],$$
$$R_3 = \left[\overline{F}^{-1}\left(q_{0.50}\overline{F}(x)\right), \overline{F}^{-1}\left(q_{0.25}\overline{F}(x)\right)\right], \qquad R_4 = \left[\overline{F}^{-1}\left(q_{0.25}\overline{F}(x)\right), +\infty\right].$$

We remark that if $\overline{F}$ is correct, then $\mathbb{P}(X_{s:n} \in R_i | X_{r:n} = x) = 1/4$ for $i = 1, 2, 3, 4$.

If we have many values for $X_{r:n}$ and $X_{s:n}$, we could consider these regions for these fixed values of r, s and n. If we just have some values from a sample of size n, we could consider the regions for different values of r and s or use resampling methods. In all these cases we use the Pearson statistic

$$T = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i},$$

where O_i and E_i are the observed and expected data in each region, being $E_i = N/4$ if we have N data and we assume that $\overline{F}$ is correct (null hypothesis). Under this assumption, the asymptotic distribution of T as $N \to +\infty$ is a chi-squared distribution with 3 degrees of freedom (2 if we use the MLE of the parameter θ). Under this null hypothesis, the associated P-value will be $\mathbb{P}(\chi^2_3 > T)$. Some illustrative examples will be provided later.

6.1.2 Dependent data

Assume that we have a sample $X_1, \dots, X_n$ of identically distributed (ID) random variables with a common distribution function F and that the data might have some kind of dependency. Usually, the dependency is due to the fact that they share the same environment (for example, when they are components of the same system). The dependency will be modeled with a copula function C that is used to write their joint distribution by Sklar's theorem as

$$\mathbf{F}(x_1, \dots, x_n) = \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = C(F(x_1), \dots, F(x_n))$$

for all $x_1, \dots, x_n$. See Nelsen [88] and Durante and Sempi [42] for the main properties of copulas.

We assume that $\mathbf{F}$ is absolutely continuous and we again consider the ordered data $X_{1:n} < \dots < X_{n:n}$ obtained from $X_1, \dots, X_n$. Suppose we just have the first r data $X_{1:n} < \dots < X_{r:n}$ and we want to predict $X_{s:n}$ for $s > r$.

Under dependency, the order statistics do not satisfy the Markov property, i.e., the distributions of $(X_{s:n}|X_{1:n} = x_1, \dots, X_{r:n} = x_r)$ and $(X_{s:n}|X_{r:n} = x_r)$ do not coincide. To get bivariate plots, we will use the second one and will predict $X_{s:n}$ from $X_{r:n} = x$ for $r < s$. The other data can just be used to estimate the unknown parameters in the model. In this case, we might have unknown parameters both in F and in C.

To get the predictions, we will use a distortion representation for the joint distribution of the random vector $(X_{r:n}, X_{s:n})$ as proposed in [84]. We remark that the results for the case $n = 2$, $r = 1$ and $s = 2$ (paired ordered data) were obtained there. Here, the procedure is similar but, the expressions obtained for $n = 4$ will be more complicated. They are based on the following two facts.

The first one is that there exists a distortion function D (which depends on r, s, n and C), such that the joint distribution $G_{r,s:n}$ of $(X_{r:n}, X_{s:n})$ can be written as

$$G_{r,s:n}(x, y) = \mathbb{P}(X_{r:n} \le x, X_{s:n} \le y) = D_{r,s:n}(F(x), F(y))$$

for all x, y. The distortion function $D_{r,s:n}$ is a continuous bivariate distribution function with support included in the set $[0, 1]^2$. This representation is similar to the classical copula representation but here $D_{r,s:n}$ is not a copula and F does not coincide with the marginal distributions (i.e., the distributions of $X_{r:n}$ and $X_{s:n}$).

The second fact is that, from the results obtained in [84], we can obtain the median regression curve and the associated prediction bands to predict $X_{s:n}$ from $X_{r:n}$. The result can be stated as follows (see Proposition 7 in [84]). We use the notation $\partial_i D_{r,s:n}$ for the partial derivative of $D_{r,s:n}$ with respect to its ith variable.

Proposition 6.4. *If we assume that both F and $D_{r,s:n}$ are absolutely continuous, then the* *conditional distribution of* $X_{s:n}$ *given* $X_{r:n} = x$ *is*

$$G_{s|r:n}(y|x) = \frac{\partial_1 D_{r,s:n}(F(x), F(y))}{\partial_1 D_{r,s:n}(F(x), 1)} \tag{6.7}$$

for $x < y$ *such that* $\partial_1 D_{r,s:n}(F(x), v) > 0$ *and* $\lim_{v \to 0^+} \partial_1 D_{r,s:n}(F(x), v) = 0$.

Sometimes, it is better to use the reliability functions instead of the distribution functions and we have similar results for them. The joint reliability function of $X_1, \ldots, X_n$ can be written as

$$\overline{\mathbf{F}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 > x_1, \ldots, X_n > x_n) = \widehat{C}(\overline{F}(x_1), \ldots, \overline{F}(x_n))$$

for all $x_1, \ldots, x_n$, where $\overline{F} = 1 - F$ and $\widehat{C}$ is another copula called *survival copula*. $\widehat{C}$ can be obtained from C (and vice versa). Analogously, the joint reliability function of $\overline{G}_{r,s:n}$ of $(X_{r:n}, X_{s:n})$ can be written as

$$\overline{G}_{r,s:n}(x, y) = \mathbb{P}(X_{r:n} > x, X_{s:n} > y) = \widehat{D}_{r,s:n}(\overline{F}(x), \overline{F}(y)) \tag{6.8}$$

for all x, y. The distortion function $\widehat{D}_{r,s:n}$ is also a continuous bivariate distribution function with support included in the set $[0,1]^2$. It depends on r, s, n and C (or $\widehat{C}$). From this expression, the conditional reliability function can be obtained as expressed in the following proposition.

Proposition 6.5. *If we assume that both* $\overline{F}$ *and* $\widehat{D}_{r,s:n}$ *are absolutely continuous, then the conditional reliability function of* $X_{s:n}$ *given* $X_{r:n} = x$ *is*

$$\overline{G}_{s|r:n}(y|x) = \frac{\partial_1 \widehat{D}_{r,s:n}(\overline{F}(x), \overline{F}(y))}{\partial_1 \widehat{D}_{r,s:n}(\overline{F}(x), 1)} \tag{6.9}$$

for $x < y$ *such that* $\partial_1 \widehat{D}_{r,s:n}(\overline{F}(x), v) > 0$ *and* $\lim_{v \to 0^+} \partial_1 \widehat{D}_{r,s:n}(\overline{F}(x), v) = 0$.

The preceding expressions can be used to solve the general case in which we want to predict $X_{s:n}$ from $X_{r:n}$ for $1 \leq r < s \leq n$. To show the procedure, we choose different cases for $n = 4$. In all these cases we assume that the joint distribution of (X_1, X_2, X_3, X_4) is exchangeable (EXC), that is, it does not change if we permute them. This is equivalent to the assumption that they are ID and C (or $\widehat{C}$) is exchangeable. In the first case, we choose $r = 1$ and $s = 2$.

Proposition 6.6. *If both* $\overline{F}$ *and* $\widehat{C}$ *are absolutely continuous and* $\widehat{C}$ *is EXC, then the conditional reliability function of* $X_{2:4}$ *given* $X_{1:4} = x$ *is*

$$\overline{G}_{2|1:4}(y|x) = \frac{\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(y), \overline{F}(y), \overline{F}(y))}{\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x))} \tag{6.10}$$

for all $x < y$ *such that* $\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x)) > 0$ *and* $\lim_{v \to 0^+} \partial_1 \widehat{C}(\overline{F}(x), v, v, v) = 0$.

Proof. The joint reliability function $\overline{G}_{1,2:4}$ of $(X_{1:4}, X_{2:4})$ satisfies

$$\overline{G}_{1,2:4}(x,y) = \mathbb{P}(X_{1:4} > x, X_{2:4} > y) = \mathbb{P}(X_{1:4} > x)$$

for all $x \geq y$, where

$$\mathbb{P}(X_{1:4} > x) = \mathbb{P}(X_1 > x, X_2 > x, X_3 > x, X_4 > x) = \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x)).$$

Analogously, for $x < y$, we get

$$\begin{aligned}\overline{G}_{1,2:4}(x,y) &= \mathbb{P}(X_{1:4} > x, X_{2:4} > y) = \mathbb{P}(X_{1:4} > x, \max_{i=1,\dots,r} X_{P_i} > y)\\ &= \mathbb{P}\left(\cup_{i=1}^{r}(\{X_{P_i} > y\} \cap \{X_{1:4} > x\})\right),\end{aligned}$$

where $X_{P_i} = \min_{j\in P_i} X_j$ and $P_1, \dots, P_r$ are all the minimal path sets of $X_{2:4}$ (see, for instance, [82], p. 23). In this case they are all the subsets of $\{1,2,3,4\}$ with cardinality 3, and so $r = \binom{4}{3} = 4$. Hence, by applying the inclusion-exclusion formula and by using the exchangeable assumption we get

$$\begin{aligned}\overline{G}_{1,2:4}(x,y) &= \sum_{i=1}^{4} \mathbb{P}\left(X_{P_i} > y, X_{1:4} > x\right) - \sum_{i<j} \mathbb{P}\left(X_{P_i\cup P_j} > y, X_{1:4} > x\right)\\ &+ \sum_{i<j<k} \mathbb{P}\left(X_{P_i\cup P_j\cup P_k} > y, X_{1:4} > x\right) - \mathbb{P}\left(X_{P_1\cup P_2\cup P_3\cup P_4} > y, X_{1:4} > x\right)\\ &= 4\mathbb{P}\left(X_1 > x, X_2 > y, X_3 > y, X_4 > y\right) - 3\mathbb{P}\left(X_1 > y, X_2 > y, X_3 > y, X_4 > y\right)\\ &= 4\widehat{C}\left(\overline{F}(x), \overline{F}(y), \overline{F}(y), \overline{F}(y)\right) - 3\widehat{C}\left(\overline{F}(y), \overline{F}(y), \overline{F}(y), \overline{F}(y)\right)\end{aligned}$$

for $x < y$. Therefore, (6.8) holds for

$$\widehat{D}_{1,2:4}(u,v) = \begin{cases} \widehat{C}(u,u,u,u) & \text{for} \quad 0 \leq u \leq v \leq 1;\\ 4\widehat{C}(u,v,v,v) - 3\widehat{C}(v,v,v,v) & \text{for} \quad 0 \leq v < u \leq 1.\end{cases}$$

Hence

$$\partial_1\widehat{D}_{1,2:4}(u,v) = \begin{cases} 4\partial_1\widehat{C}(u,u,u,u) & \text{for} \quad 0 \leq u \leq v \leq 1;\\ 4\partial_1\widehat{C}(u,v,v,v) & \text{for} \quad 0 \leq v < u \leq 1.\end{cases}$$

Finally, we use (6.9) to get (6.10). □

In the following propositions, we provide the expressions for the other cases. As the proofs are similar, they are omitted. Note that in Proposition 6.10 we use (6.7) instead of (6.9).

Proposition 6.7. *If both $\overline{F}$ and $\widehat{C}$ are absolutely continuous and $\widehat{C}$ is EXC, then the conditional reliability function of $X_{3:4}$ given $X_{1:4} = x$ is*

$$\overline{G}_{3|1:4}(y|x) = \frac{3\partial_1\widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(y), \overline{F}(y)) - 2\partial_1\widehat{C}(\overline{F}(x), \overline{F}(y), \overline{F}(y), \overline{F}(y))}{\partial_1\widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x))} \tag{6.11}$$

for all $x < y$ such that $\partial_1\widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x)) > 0$ and $\lim_{v\to 0^+} 3\partial_1\widehat{C}(\overline{F}(x), \overline{F}(x), v, v) - 2\partial_1\widehat{C}(\overline{F}(x), v, v, v) = 0$.

Proposition 6.8. *If both $\overline{F}$ and $\widehat{C}$ are absolutely continuous and $\widehat{C}$ is EXC, then the conditional reliability function of $X_{4:4}$ given $X_{1:4} = x$ is*

$$\overline{G}_{4|1:4}(y|x) = \frac{A_{4|1:4}(x,y)}{\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x))}, \tag{6.12}$$

for all $x < y$ such that $\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x)) > 0$ and $\lim_{y\to+\infty} A_{4|1:4}(x,y) = 0$, where

$$\begin{aligned} A_{4|1:4}(x,y) &= 3\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(y)) - 3\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(y), \overline{F}(y)) \\ &\quad + \partial_1 \widehat{C}(\overline{F}(x), \overline{F}(y), \overline{F}(y), \overline{F}(y)). \end{aligned}$$

Proposition 6.9. *If both $\overline{F}$ and $\widehat{C}$ are absolutely continuous and $\widehat{C}$ is EXC, then the conditional reliability function of $X_{4:4}$ given $X_{2:4} = x$ is*

$$\overline{G}_{4|2:4}(y|x) = \frac{A_{4|2:4}(x,y)}{\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), 1) - \partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x))} \tag{6.13}$$

for all $x < y$ such that $\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), 1) - \partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(x)) > 0$ and $\lim_{y\to+\infty} A_{4|2:4}(x,y) = 0$, where

$$\begin{aligned} A_{4|2:4}(x,y) &= 2\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(y), 1) - 2\partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(x), \overline{F}(y)) \\ &\quad - \partial_1 \widehat{C}(\overline{F}(x), \overline{F}(y), \overline{F}(y), 1) + \partial_1 \widehat{C}(\overline{F}(x), \overline{F}(x), \overline{F}(y), \overline{F}(y)). \end{aligned}$$

Proposition 6.10. *If both F and C are absolutely continuous and C is EXC, then the conditional distribution function of $X_{4:4}$ given $X_{3:4} = x$ is*

$$G_{4|3:4}(y|x) = \frac{\partial_1 C(F(x), F(x), F(x), F(y)) - \partial_1 C(F(x), F(x), F(x), F(x))}{\partial_1 C(F(x), F(x), F(x), 1) - \partial_1 C(F(x), F(x), F(x), F(x))},$$

for all $x < y$ such that $\partial_1 C(F(x), F(x), F(x), 1) - \partial_1 C(F(x), F(x), F(x), F(x)) > 0$.

Now, we show how to get predictions from more than one early failures. If we want to predict $X_{3:4}$ from $X_{1:4} = x$ and $X_{2:4} = y$, we need a distortion representation for their joint reliability function as

$$\overline{G}_{1,2,3}(x,y,t) = \mathbb{P}(X_{1:4} > x, X_{2:4} > y, X_{3:4} > t) = \widehat{D}(\overline{F}(x), \overline{F}(y), \overline{F}(t)),$$

where $\overline{F}$ is the common reliability function of X_1, X_2, X_3, X_4. Then their joint pdf is

$$g_{1,2,3}(x,y,t) = f(x)f(y)f(t)\ \partial_{1,2,3}\widehat{D}(\overline{F}(x), \overline{F}(y), \overline{F}(t)),$$

where $f = -\overline{F}'$. Analogously, the joint reliability function of $X_{1:4}$ and $X_{2:4}$ is

$$\overline{G}_{1,2}(x,y) = \mathbb{P}(X_{1:4} > x, X_{2:4} > y) = \widehat{D}(\overline{F}(x), \overline{F}(y), 1)$$

and their joint pdf

$$g_{1,2}(x,y) = f(x)f(y)\ \partial_{1,2}\widehat{D}(\overline{F}(x), \overline{F}(y), 1).$$

Hence, the pdf of $(X_{3:4}|X_{1:4}=x, X_{2:4}=y)$ is

$$g_{3|1,2}(t|x,y)=\frac{g_{1,2,3}(x,y,t)}{g_{1,2}(x,y)}=f(t)\frac{\partial_{1,2,3}\widehat{D}(\overline{F}(x),\overline{F}(y),\overline{F}(t))}{\partial_{1,2}\widehat{D}(\overline{F}(x),\overline{F}(y),1)}$$

and the reliability function is

$$\overline{G}_{3|1,2}(t|x,y)=\frac{\partial_{1,2}\widehat{D}(\overline{F}(x),\overline{F}(y),\overline{F}(t))}{\partial_{1,2}\widehat{D}(\overline{F}(x),\overline{F}(y),1)},$$

for $x<y<t$, whenever $\partial_{1,2}\widehat{D}(\overline{F}(x),\overline{F}(y),1)>0$ and $\lim_{v\to 0^+}\partial_{1,2}\widehat{D}(\overline{F}(x),\overline{F}(y),v)=0$. If the survival copula $\widehat{C}$ is EXC, then

$$\widehat{D}(u,v,w)=12\widehat{C}(u,v,w,w)-12\widehat{C}(u,w,w,w)-6\widehat{C}(v,v,w,w)+7\widehat{C}(w,w,w,w)$$

and

$$\partial_{1,2}\widehat{D}(u,v,w)=12\widehat{C}(u,v,w,w),$$

for $1>u>v>w>0$. Analogously,

$$\widehat{D}(u,v,1)=4\widehat{C}(u,v,v,v)-3\widehat{C}(v,v,v,v)$$

and

$$\partial_{1,2}\widehat{D}(u,v,1)=12\widehat{C}(u,v,v,v),$$

for $1>u>v>0$. Hence,

$$\overline{G}_{3|1,2}(t|x,y)=\frac{\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(t),\overline{F}(t))}{\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(y),\overline{F}(y))} \tag{6.14}$$

for $x<y<t$ such that $\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(y),\overline{F}(y))>0$ and

$$\lim_{t\to+\infty}\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(t),\overline{F}(t))=0.$$

The prediction is obtained by solving (numerically) $\overline{G}_{3|1,2}(t|x,y)=0.5$ for given values of x and y. The prediction intervals are obtained in a similar manner.

Proceeding as above we can also obtain the expression to predict $X_{4:4}$ from $X_{1:4}=x$ and $X_{2:4}=y$ as

$$\overline{G}_{4|1,2}(t|x,y)=\frac{2\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(y),\overline{F}(t))-\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(t),\overline{F}(t))}{\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(y),\overline{F}(y))} \tag{6.15}$$

for $x<y<t$ such that $\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(y),\overline{F}(y))>0$ and

$$\lim_{t\to+\infty}2\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(y),\overline{F}(t))-\partial_{1,2}\widehat{C}(\overline{F}(x),\overline{F}(y),\overline{F}(t),\overline{F}(t))=0.$$

6.1.3 Examples

First, we illustrate the IID case with simulated samples.

Example 6.1. We simulate a sample of size $n = 20$ from a standard exponential distribution. The ordered (rounded) sample values obtained are

0.00599 0.02454 0.04600 0.07663 0.08168 0.14609 0.24391 0.72400 1.30312 1.37244

1.37962 1.54357 1.71278 2.22949 2.24561 2.56783 2.61441 2.80786 3.90280 7.68743

If we want to predict $X_{2:20}$ from $X_{1:20}$ by assuming that $\overline{F}$ is known (or that it is estimated from a preceding sample), we use the quantile regression curve given in (6.1)

$$m(x) = \overline{F}^{-1}(q_{0.5}\overline{F}(x)) = x - \log(q_{0.5}) = x + 0.03648$$

where $q_{0.5} = 0.96418$ is the median of a beta distribution with parameters $n - s + 1 = 19$ and $s - r = 1$. Thus, we get the prediction for $X_{2:20}$ as

$$\widehat{X}_{2:20} = m(X_{1:20}) = m(0.00599) = 0.00599 + 0.03648 = 0.04247.$$

The real value is $X_{2:20} = 0.02454$. The 90% and 50% prediction intervals for this prediction are obtained from (6.2) as $C_{90} = [0.00869, 0.16366]$ and $C_{50} = [0.02113, 0.07895]$. The real value belongs to both the intervals.

To see better what happens with these predictions we simulate $N = 100$ predictions of this kind, that is, 100 samples of size 20. The data are plotted in Figure 6.1, left. There we can see that the prediction bands represent very well the dispersion of the majority of data (except some extreme values). In this sample, C_{50} contains 51 values and C_{90} contains 90 while 5 values are above the upper limit and 5 are below the bottom limit. Of course, if we test $H_0 : \overline{F}$ is correct vs $H_0 : \overline{F}$ is not correct, by using the four regions R_1, R_2, R_3, R_4, we get the observed values: $25, 30, 21, 24$ and the T statistic value is

$$T = \frac{(25-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(21-25)^2}{25} + \frac{(24-25)^2}{25} = 1.68.$$

Thus, the P-value $P = \mathbb{P}(\chi^2_3 > 1.68) = 0.64139$ leads to the acceptance of the exponential distribution (as expected). In practice, it is not easy to perform this test because we need the first two values of several samples with the same size ($n = 20$ in this example).

We do the same in Figure 6.1, right, with $n = 20$, $r = 12$ and $s = 13$. About the initial sample, the prediction for $X_{13:20} = 1.71278$ from $X_{12:20} = 1.54357$ obtained with the median curve

$$m(x) = \overline{F}^{-1}(q_{0.5}\overline{F}(x)) = x - \log(q_{0.5}) = x + 0.08664$$

is $m(1.54357) = 1.63021$, where $q_{0.5} = 0.91700$ is the median of a beta distribution with parameters $n - s + 1 = 8$ and $s - r = 1$. The prediction intervals for this prediction are

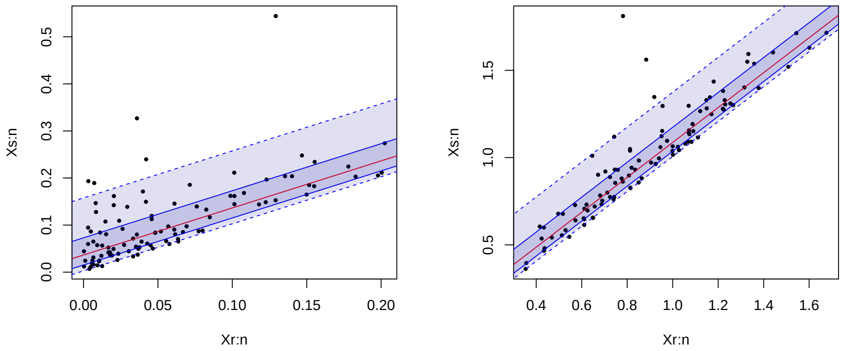

Figure 6.1: Scatterplots of a simulated sample from $(X_{r:n}, X_{s:n})$ for $n = 20$, $r = 1$ and $s = 2$ (left) and $r = 12$ and $s = 13$ (right) for the exponential distribution in Example 6.1 jointly with the theoretical median regression curves (red) and 50% (dark grey) and 90% (light grey) prediction bands.

$C_{90} = [1.549986, 1.918041]$ and $C_{50} = [1.579534, 1.716861]$. Both intervals contain the real value. The 100 repetitions of this case are plotted in Figure 6.1, right. We remind that for the exponential distribution all the curves are lines with slope one (Proposition 6.3). Here, we have 92 values in C_{90}, 56 in C_{50} and 8 values out of C_{90} (4 above and 4 below). The T statistic is 2.24 and its associated P-value 0.52411 leads again to accept the (real) distribution F.

The predictions will be worse for more distant future values (i.e., the dispersion will be greater). To show this, in Figure 6.2 we plot the prediction bands for $r = 12$, $s = 14$ (left) and $s = 20$ (right). However, of course, the coverage probabilities of these regions will be the same. The predictions obtained in the initial sample are $\widehat{X}_{14:20} = 1.76813$ and $\widehat{X}_{20:20} = 4.03254$, with prediction intervals $C_{90} = [1.59107, 2.17974]$ and $C_{90} = [2.70722, 6.59642]$, respectively. The real values are $X_{14:20} = 2.22949$ and $X_{20:20} = 7.68743$ and both are out of the interval C_{90}.

In Figure 6.3 we plot the predictions for $X_{s:20}$ (red line) jointly with the limits of the 90% (dashed blue lines) and 50% (continuous blue line) prediction intervals in the initial simulated sample from $X_{12:20}$ (left) for $s = 13, \ldots, 20$ and from the preceding data $X_{s-1:20}$ (right) for $s = 2, \ldots, 20$. In the left plot 2-out-of-8 exact points do not belong to the 90% prediction intervals while 5 are out of the 50% prediction intervals (the expected values are $8 \cdot 0.1 = 0.8$ and $8 \cdot 0.5 = 4$, respectively). In the right plot 4-out-of-19 points do not belong to 90% prediction intervals while 11 do not belong to the 50% prediction intervals (we expect 1.9 and 9.5, respectively).

In practice, we do not know the exact distribution. If we just assume the exponential model $\overline{F}(t) = e^{-\theta t}$ for $t \geq 0$, with an unknown parameter $\theta > 0$, we can use (6.5) to estimate θ. With

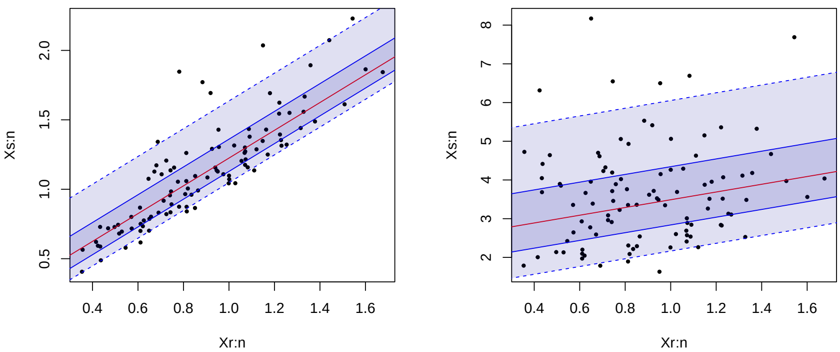

Figure 6.2: Scatterplots of a simulated sample from $(X_{r:n}, X_{s:n})$ for $n = 20$, $r = 12$ and $s = 14$ (left) and $s = 20$ (right) for the exponential distribution in Example 6.1 jointly with the theoretical median regression curves (red) and 50% (dark grey) and 90% (light grey) prediction bands.

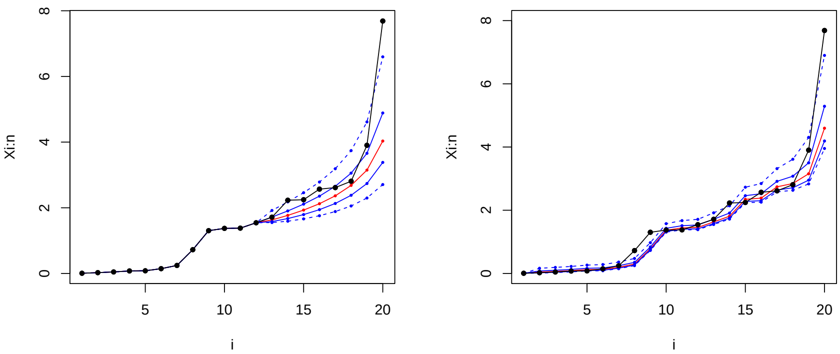

Figure 6.3: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 20$, $r = 12$ and $s = 13, \ldots, 20$ (left) and $r = 1, \ldots, 19$ and $s = r + 1$ (right) for the exponential distribution in Example 6.1. The black points are the exact values and the blue lines are the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals.

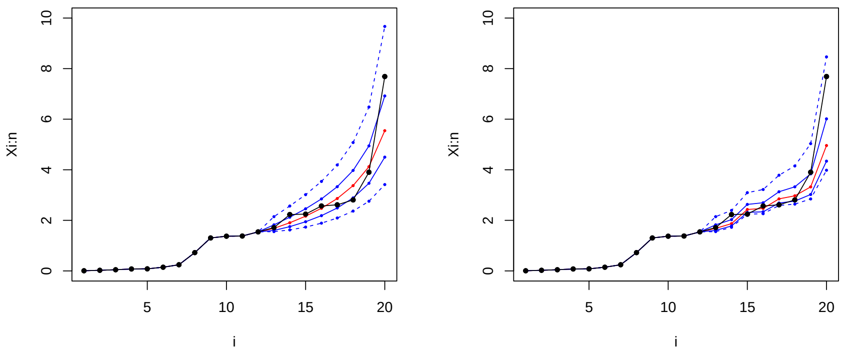

Figure 6.4: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 20$, $r = 12$ and $s = 13, \ldots, 20$ (left) and $r = 12, \ldots, 19$ and $s = r + 1$ (right) for the exponential distribution in Example 6.1 estimating θ at $X_{r:20}$. The black points are the exact values and the blue lines are the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals.

the above sample and $r = 12$ we get

$$\widehat{\theta} = \frac{12}{9X_{12:20} + \sum_{i=1}^{11} X_{i:20}} = 0.62188.$$

The exact value is $\theta = 1$. Replacing the exact reliability function $\overline{F}(t) = e^{-t}$ with $\overline{F}_{\widehat{\theta}}(t) = e^{-0.62188t}$, we can obtain predictions for $X_{s:20}$ from $X_{12:20}$ as above. For example, for $s = 13$ we get the prediction $\widehat{X}_{13:20} = 1.6829$ for $X_{13:20} = 1.71278$. The estimated prediction intervals are $\widehat{C}_{90} = [1.55388, 2.14572]$ and $\widehat{C}_{50} = [1.60140, 1.82222]$ and both contain the exact value. Since we have estimated the parameter, we do not know the exact coverage probabilities for these intervals. The predictions from $X_{12:20}$ for $X_{s:20}$ and $s = 13, \ldots, 20$ are plotted in Figure 6.4, left. The blue lines represent the prediction intervals. Note that all the exact values belong to the 90% intervals (dashed blue lines) and that three of them do not belong to the 50% intervals (continuous blue lines).

We do the same in Figure 6.4, right, but, in this case, $X_{r+1:20}$ is predicted from all the preceding values $X_{1:20}, \ldots, X_{r:20}$ for $r = 12, \ldots, 19$ which are used to estimate θ. The estimations obtained for θ and the predictions for $X_{s:20}$ are given in Table 6.1. Note that the estimations for θ are similar. The MLE for θ from the complete sample (which is not available under our assumptions) is $\widehat{\theta} = 20/(X_1 + \cdots + X_{20}) = 0.6113249$ which is very similar to our estimations for $r \geq 12$ (although the exact value is $\theta = 1$). In practice, when we work with real data, the stability of these predictions might confirm the assumed parametric model. Note that all the exact values belong to the 90% prediction intervals while 4 do not belong to the 50% prediction intervals (as expected). Surprisingly, the estimations obtained from $X_{12:20}$ seem to be better

Table 6.1: Predicted values $\widehat{X}_{r+1:n}$ and centered prediction intervals $C_{50} = [l_r, u_r]$ and $C_{90} = [L_n, U_n]$ for $X_{r+1:n}$ from $X_{r:n}$ in a standard exponential distribution; $\widehat{\theta}$ is the estimate of θ at $X_{r:n}$ for $r = 12, \ldots, 19$.

r	$\widehat{\theta}$	L_r	l_r	$\widehat{X}_{r+1:n}$	$X_{r+1:n}$	u_r	U_r
12	0.62188	1.55388	1.60140	1.68290	1.71278	1.82222	2.14572
13	0.62954	1.72442	1.77807	1.87007	2.22949	2.02736	2.39258
14	0.57692	2.24431	2.31260	2.42974	2.24561	2.62998	3.09493
15	0.61567	2.26227	2.33906	2.47078	2.56783	2.69595	3.21877
16	0.61599	2.58865	2.68459	2.84915	2.61441	3.13047	3.78366
17	0.64982	2.640723	2.76198	2.96997	2.80786	3.32553	4.15110
18	0.67312	2.845958	3.02155	3.32274	3.90280	3.83762	5.03313
19	0.65673	3.980903	4.34085	4.95825	7.68743	6.01370	8.46438

than that obtained from the preceding values but it has to be pointed out that the lengths of the intervals in the first case are greater than the ones obtained in the second.

In the following example we analyze a real data set by assuming that the original (not ordered) data values are independent (the ordered values are always dependent).

Example 6.2. Let us study the real data set considered in [16]. They represent ordered lifetimes of 20 electronic components. The complete sample is

0.03	0.12	0.22	0.35	0.73	0.79	1.25	1.41	1.52	1.79
1.80	1.94	2.38	2.40	2.87	2.99	3.14	3.17	4.72	5.09

Assume that we just know the first 12 failure times and that we want to predict the future failures. If we assume an exponential distribution with unknown failure rate θ, then we estimate it from (6.5) as

$$\widehat{\theta} = \frac{r}{(n-r+1)x_r + \sum_{i=1}^{r-1} x_i} = \frac{12}{9 \cdot 1.94 + 0.03 + \cdots + 1.8} = 0.43684.$$

From this value we obtain the predictions and prediction intervals given in Figure 6.5, left. For example, the prediction for $X_{13:20} = 2.38$ is $\widehat{X}_{13:20} = 2.13834$ with prediction intervals $C_{90} = [1.95468, 2.797216]$ and $C_{50} = [2.02232, 2.33668]$. The exact value belongs to C_{90} but not to C_{50}. The predictions for the last values are not very good. However, all the exact values belong to the 90% prediction intervals and only 4 out of 8 of them do not belong to the 50% prediction intervals (as expected). Note that this plot is similar to the plot obtained in Figure 6.3, left, with a sample of size 20 from an exponential distribution. If we count the data in

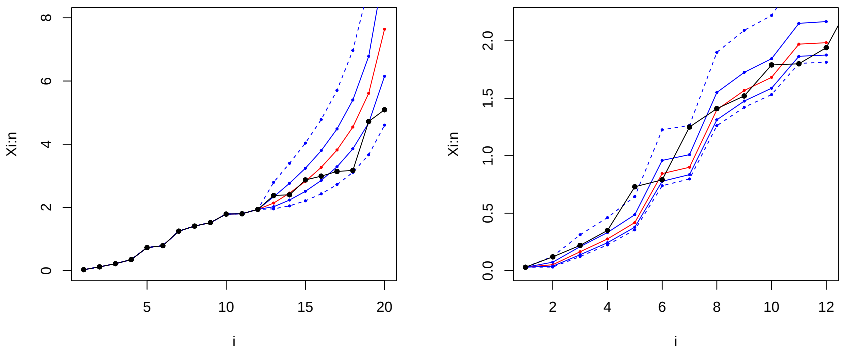

Figure 6.5: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 20$, $r = 12$ and $s = 13, \dots, 20$ (left) and $s = r+1$ and $r = 1, \dots, 11$ (right) for the real data set in Example 6.2 estimating θ under an exponential model. The black points are the exact values and the blue lines are the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals.

the four regions R_1, R_2, R_3, R_4, we get the observed data $3, 3, 1, 1$ and the Pearson T statistic value is

$$T = \frac{(3-2)^2}{2} + \frac{(3-2)^2}{2} + \frac{(1-2)^2}{2} + \frac{(1-2)^2}{2} = 2.$$

We can approximate its distribution with a chi-squared distribution with 2 degrees of freedom (since we have estimated a parameter), and then its associated P-value is $\mathbb{P}(\chi_2^2 > 2) = 0.36788$. So the exponential model cannot be rejected with the complete sample (by using this test).

To check the model with the first 12 values we could estimate θ and predict $X_{r+1:20}$ from $X_{r:20}$ for $r = 1, \dots, 11$. The predictions can be seen in Figure 6.5, right. The estimations for θ are

1.66667	0.86580	0.72993	0.63291	0.40323	0.45113
0.35461	0.36664	0.38894	0.38300	0.41969	0.43684

These estimations are stable from $r = 5$ to $r = 12$. The MLE estimation with the complete sample is $\widehat{\theta} = 0.51666$. As we can see in the figure the predictions are accurate. Two and six exact points do not belong to the 90% and 50% prediction intervals, respectively. These numbers are close to the expected values ($0.1 \cdot 11 = 1.1$ and $0.5 \cdot 11 = 5.5$). So the exponential model cannot be rejected by using these first 12 values (the P-value obtained with the four regions is 0.20374).

In the same framework, assume a Pareto type II distribution with unknown parameter θ. Then, we estimate it from (6.6) as

$$\widehat{\theta} = \frac{r}{(n-r+1)\log(1+x_r) + \sum_{i=1}^{r-1}\log(1+x_i)} = 0.74311.$$

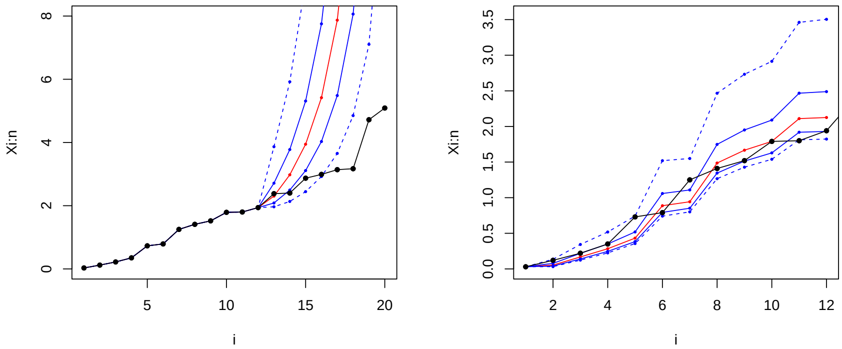

Figure 6.6: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 20$, $r = 12$ and $s = 13, \ldots, 20$ (left) and $s = r+1$ and $r = 1, \ldots, 11$ (right) for the real data set in Example 6.2 estimating θ under a Pareto type II model. The black points are the exact values and the blue lines are the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals.

From this value we obtain the predictions and prediction intervals given in Figure 6.6, left. The prediction obtained with the Pareto model for $X_{13:20} = 2.38$ is $\widehat{X}_{13:20} = 2.30358$. The predictions for the last values are very bad. In fact, 4 of the exact values do not belong to the 90% prediction interval and only 1 of them belongs to the 50% prediction interval. If we count the data in the four regions R_1, R_2, R_3, R_4, we get the observed data $7, 0, 1, 0$ and the Pearson T statistic value is

$$T = \frac{(7-2)^2}{2} + \frac{(0-2)^2}{2} + \frac{(1-2)^2}{2} + \frac{(0-2)^2}{2} = 17.$$

By using a chi-squared distribution with 2 degrees of freedom, the P-value is $\mathbb{P}(\chi^2_2 > 17) = 0.00020$ and so the Pareto type II model is rejected with the complete sample. Let us see what happens if we check the model with the first 12 values by estimating θ and predicting $X_{r+1:20}$ from $X_{r:20}$ for $r = 1, \ldots, 11$. The predictions can be seen in Figure 6.6, right. The estimations obtained for θ are

1.61099	0.91625	0.80597	0.73482	0.53125	0.60464
0.53333	0.57068	0.61839	0.63802	0.70023	0.74311

In the figure, we can see that one and seven exact points do not belong to the 90% and 50% prediction intervals, respectively. In this case, the P-value obtained with the four regions is 0.06843 and the Pareto type II model can be rejected by using these 12 values.

Now, we consider an example of four dependent data values. They can represent the values in a small sample but they could also be the lifetimes of the four engines in a plane, i.e., a case in which it is very important to predict the future failure times.

Example 6.3. First we consider the case $r = 1$, $s = 2$ and $n = 4$, i.e., we want to predict $X_{2:4}$ from $X_{1:4} = x$. Assume that (X_1, X_2, X_3, X_4) has the following Farlie-Gumbel-Morgenstern (FGM) survival copula

$$\widehat{C}(u_1, u_2, u_3, u_4) = u_1u_2u_3u_4 + \theta u_1u_2u_3u_4(1 - u_1)(1 - u_2)(1 - u_3)(1 - u_4) \tag{6.16}$$

for $u_1, u_2, u_3, u_4 \in [0, 1]$ and $\theta \in [-1, 1]$. The independent case is obtained when $\theta = 0$. Then,

$$\partial_1\widehat{C}(u_1, u_2, u_3, u_4) = u_2u_3u_4 + \theta u_2u_3u_4(1 - 2u_1)(1 - u_2)(1 - u_3)(1 - u_4)$$

and

$$\lim_{v\to 0^+} \partial_1\widehat{C}(\overline{F}(x), v, v, v) = \lim_{v\to 0^+} v^3 + \theta(1 - 2\overline{F}(x))v^3(1 - v)^3 = 0$$

for all x. Hence, from (6.10), we get

$$\overline{G}_{2|1:4}(y|x) = \frac{\overline{F}^3(y) + \theta\overline{F}^3(y)F^3(y)(1 - 2\overline{F}(x))}{\overline{F}^3(x) + \theta\overline{F}^3(x)F^3(x)(1 - 2\overline{F}(x))}$$

for all $x \le y$ such that $\overline{F}^3(x) + \theta\overline{F}^3(x)F^3(x)(1 - 2\overline{F}(x)) > 0$.

Unfortunately, we cannot obtain an explicit expression for its inverse. However, we can plot the level curves of this function to get the plots of the median regression curve (level 0.5) and the limits of the centered prediction regions C_{90} (levels $0.05, 0.95$) and C_{50} (levels $0.25, 0.75$). They are plotted in Figure 6.7, left, jointly with the values obtained from 100 samples of size $n = 4$ from a standard exponential distribution and a FGM survival copula with $\theta = 1$. The (rounded) ordered values obtained in the first sample are

0.07086 0.32313 0.88360 1.66760.

The method used to generate these sample values will be explained in Remark 6.1. Note that the data values are perfectly represented by these prediction regions. In the plot we also provide the curves for $\theta = 0$ (green lines) and $\theta = -1$. As we can see the changes are really minor since the FGM copula gives a weak dependence relation. The curves might be more different in other dependence models (copulas).

Now, assume that the parameters in the model are unknown. Taking into account the preceding comments, instead of estimating θ, we could just plot the curves for the extremes values $\theta = -1, 1$. The exact curves will be between them. So we just need to estimate the parameter $\lambda = 1$ of the exponential distribution. For this purpose, in practice, we have just the sample minimum $X_{1:4}$. Its reliability function is

$$\overline{F}_{1:4}(t) = \widehat{C}(\overline{F}(t), \overline{F}(t), \overline{F}(t), \overline{F}(t)) = \overline{F}^4(t) + \theta\overline{F}^4(t)F^4(t)$$

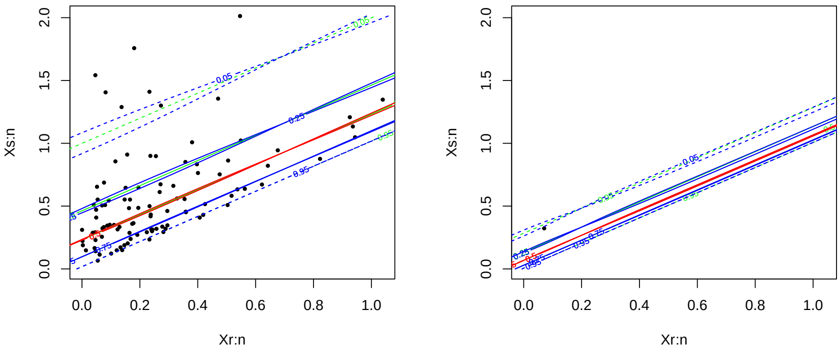

Figure 6.7: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 4$, $r = 1$ and $s = 2$ for $\theta = -1, 0, 1$ (left) jointly with the values (black point) from 100 simulated samples from a standard exponential distribution and a FGM survival copula with $\theta = 1$ (see Example 6.3). The blue lines represent the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals. The green lines are the curves for the independent case. In the right plot we have the curves when the mean of the exponential distribution is estimated from the minimum $X_{1:4}$ in the first sample.

for $t \geq 0$. Hence, if $\overline{F}(t) = \exp(-t/\mu)$ with $\mu = 1/\lambda$, then the mean of $X_{1:4}$ is

$$\begin{aligned} E(X_{1:4}) &= \int_0^{+\infty} (e^{-4t/\mu} + \theta e^{-4t/\mu}(1 - e^{-t/\mu})^4) dt \\ &= \int_0^{+\infty} (1+\theta)e^{-4t/\mu} - 4\theta e^{-5t/\mu} + 6\theta e^{-6t/\mu} - 4\theta e^{-7t/\mu} + \theta e^{-8t/\mu} dt \\ &= \mu \left(\frac{1+\theta}{4} - \frac{4\theta}{5} + \theta - \frac{4\theta}{7} + \frac{\theta}{8} \right). \end{aligned}$$

Therefore, μ can be estimated by

$$\widehat{\mu} = \frac{X_{1:4}}{\theta + \frac{1+\theta}{4} - \frac{4\theta}{5} - \frac{4\theta}{7} + \frac{\theta}{8}}.$$

For $\theta = 1$, we get $\widehat{\mu} = 3.94366 X_{1:4}$ and for $\theta = -1$, $\widehat{\mu} = 4.05797 X_{1:4}$. In our first simulated sample, we obtain the value $X_{1:4} = 0.07086$ and so $\widehat{\mu} \in [0.27945, 0.28755]$. As we are assuming that θ is unknown, we can use the average of these two estimations to approximate μ with 0.2835. By using this value, we get the curves plotted in Figure 6.7, right. Although the estimation for $\mu = 1$ is very bad (since we just have one data point) and the curves are far from the exact ones (plotted in Figure 6.7, left), note that the value $X_{2:4}$ belongs to the 90% prediction interval obtained from $X_{1:4}$.

Along the same lines and by using the same copula, we can consider other cases. If we

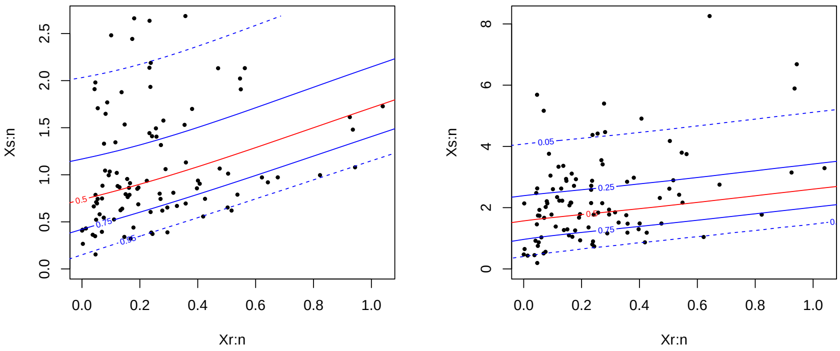

Figure 6.8: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 4$, $r = 1$, $s = 3$ (left) and $s = 4$ (right) for $\theta = 1$ jointly with the values (black point) from 100 simulated samples from a standard exponential distribution and a FGM survival copula with $\theta = 1$ (see Example 6.3). The blue lines represent the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals.

want to predict $X_{3:4}$ from $X_{1:4} = x$, by using (6.11), we get

$$\overline{G}_{3|1:4}(y|x) = \frac{A_{3|1:4}(x,y)}{\overline{F}^3(x) + \theta\overline{F}^3(x)F^3(x)(1-2\overline{F}(x))}$$

for all $x \leq y$ such that $\overline{F}^3(x) + \theta\overline{F}^3(x)F^3(x)(1-2\overline{F}(x)) > 0$, where

$$\begin{aligned} A_{3|1:4}(x,y) = {} & 3\overline{F}(x)\overline{F}^2(y) + 3\theta\overline{F}(x)\overline{F}^2(y)F(x)F^2(y)(1-2\overline{F}(x)) \\ & - 2\overline{F}^3(y) - 2\theta\overline{F}^3(y)F^3(y)(1-2\overline{F}(x)). \end{aligned}$$

As in the preceding case, we plot the level curves of this function to get the plots of the median regression curve (level 0.5) and the limits of the centered prediction regions C_{90} (levels $0.05, 0.95$) and C_{50} (levels $0.25, 0.75$). They are plotted in Figure 6.8, left, jointly with the values from 100 samples of size $n = 4$ from a standard exponential distribution and a FGM survival copula with $\theta = 1$. Here, we do not plot also the curves for $\theta = 0$ and $\theta = -1$ since the changes are again really minor.

Furthermore, if we want to predict $X_{4:4}$ from $X_{1:4} = x$, by using (6.12), we get

$$\overline{G}_{4|1:4}(y|x) = \frac{A_{4|1:4}(x,y)}{\overline{F}^3(x) + \theta\overline{F}^3(x)F^3(x)(1-2\overline{F}(x))}$$

for all $x \leq y$ such that $\overline{F}^3(x) + \theta\overline{F}^3(x)F^3(x)(1-2\overline{F}(x)) > 0$, where

$$\begin{aligned} A_{4|1:4}(x,y) = {} & 3\overline{F}^2(x)\overline{F}(y) + 3\theta\overline{F}^2(x)\overline{F}(y)F^2(x)F(y)(1-2\overline{F}(x)) - 3\overline{F}(x)\overline{F}^2(y) \\ & - 3\theta\overline{F}(x)\overline{F}^2(y)F(x)F^2(y)(1-2\overline{F}(x)) + \overline{F}^3(y) + \theta\overline{F}^3(y)F^3(y)(1-2\overline{F}(x)). \end{aligned}$$

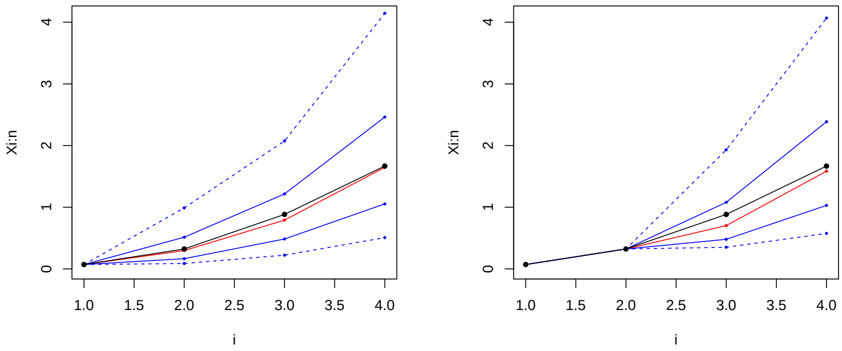

Figure 6.9: Predictions (red) for $X_{s:n}$ from $X_{r:n}$ for $n = 4$, $r = 1$ and $s = 2, 3, 4$ (left) jointly with the exact values (black points) from a simulated samples from a standard exponential distribution and an FGM survival copula with $\theta = 1$ (see Example 6.3). The blue lines represent the limits for the 50% (continuous lines) and the 90% (dashed lines) prediction intervals. In the right plot we can see the predictions for $X_{3:4}$ and $X_{4:4}$ from $X_{1:4}$ and $X_{2:4}$.

We plot the level curves of this function to get the plots of the median regression curve and the limits of the centered prediction regions C_{90} and C_{50}. They are given in Figure 6.8, right, jointly with the values from 100 samples of size $n = 4$ from a standard exponential distribution and a FGM survival copula with $\theta = 1$. Again, we do not plot the curves for $\theta = 0$ and $\theta = -1$ since the changes are really minor.

Proceeding as above, we can predict $X_{2:4}, X_{3:4}$ and $X_{4:4}$ by using the median regression curve and we can obtain the limits of the centered prediction regions C_{90} and C_{50}. In the first sample, the prediction obtained for $X_{2:4} = 0.32313$ from $X_{1:4} = 0.70086$ is $\widehat{X}_{2:4} = 0.29708$ with prediction intervals $C_{50} = [0.16582, 0.51384]$ and $C_{90} = [0.08788, 0.98928]$. Analogously, the prediction for $X_{3:4} = 0.88360$ from $X_{1:4} = 0.70086$ is $\widehat{X}_{3:4} = 0.79118$ with prediction intervals $C_{50} = [0.48511, 1.21566]$ and $C_{90} = [0.22238, 2.07476]$. Finally, the prediction for $X_{4:4} = 1.6676$ from $X_{1:4}$ is $\widehat{X}_{4:4} = 1.64681$ with prediction intervals $C_{50} = [1.05428, 2.46201]$ and $C_{90} = [0.50708, 4.14529]$. In Figure 6.9, left, we plot these predictions (red) for $X_{2:4}, X_{3:4}, X_{4:4}$ from $X_{1:4}$ jointly with the exact values (black points) in the first simulated sample from a standard exponential distribution and a FGM survival copula with $\theta = 1$.

We can use more than one data point to predict future failures. We can predict $X_{3:4}$ from $X_{1:4} = 0.07086$ and $X_{2:4} = 0.32313$ by using (6.14). With the FGM copula we get

$$\overline{G}_{3|1,2}(t|x, y) = \frac{1 + \theta(1 - 2\overline{F}(x))(1 - 2\overline{F}(y))F^2(t)}{1 + \theta(1 - 2\overline{F}(x))(1 - 2\overline{F}(y))F^2(y)} \cdot \frac{\overline{F}^2(t)}{\overline{F}^2(y)}$$

for $x < y \leq t$. By solving $\overline{G}_{3|1,2}(t|x, y) = 0.5$ we get the prediction $\widehat{X}_{3:4} = 0.70208$ for

$X_{3:4} = 0.8836$. Analogously, we obtain the prediction intervals $C_{50} = [0.47975, 1.07961]$ and $C_{90} = [0.3509, 1.93089]$. In a similar way, we can predict $X_{4:4} = 1.58455$ from $X_{1:4}$ and $X_{2:4}$ by using (6.15) obtaining $\widehat{X}_{4:4} = 1.58455$, $C_{50} = [1.02971, 2.38660]$ and $C_{90} = [0.57592, 4.06791]$. The predictions are plotted in Figure 6.9, right.

Remark 6.1. Let us see how to generate a sample (U_1, U_2, U_3, U_4) from the survival copula $\widehat{C}$. Then, the sample from (X_1, X_2, X_3, X_4) with a common reliability $\overline{F}$ is obtained as $(\overline{F}(U_1), \overline{F}(U_2), \overline{F}(U_3), \overline{F}(U_4))$. The joint distribution function of (U_1, U_2, U_3, U_4) is given in (6.16). Then, its joint pdf is

$$\widehat{c}(u_1, u_2, u_3, u_4) = 1 + \theta(1 - 2u_1)(1 - 2u_2)(1 - 2u_3)(1 - 2u_4)$$

for $u_1, u_2, u_3, u_4 \in (0, 1)$. The joint distribution function of (U_1, U_2, U_3) is

$$\widehat{C}_{1,2,3}(u_1, u_2, u_3) = \widehat{C}(u_1, u_2, u_3, 1) = u_1 u_2 u_3$$

and so its joint pdf is $\widehat{c}_{1,2,3}(u_1, u_2, u_3) = 1$ for $u_1, u_2, u_3 \in (0, 1)$. They are IID and can be simulated just as independent uniform random variables. The conditional pdf of U_4 given $U_1 = u_1, U_2 = u_2, U_3 = u_3$ is obtained as

$$\widehat{c}_{4|1,2,3}(u_4 \mid u_1, u_2, u_3) = \widehat{c}(u_1, u_2, u_3, u_4)$$

for $u_1, u_2, u_3, u_4 \in (0, 1)$. Therefore, its distribution function is

$$\begin{aligned}
\widehat{C}_{4|1,2,3}(u_4 \mid u_1, u_2, u_3) &= \int_0^{u_4} \widehat{c}(u_1, u_2, u_3, z)dz \\
&= \int_0^{u_4} (1 + \theta(1 - 2u_1)(1 - 2u_2)(1 - 2u_3)(1 - 2z))dz \\
&= u_4 + \theta(1 - 2u_1)(1 - 2u_2)(1 - 2u_3)(u_4 - u_4^2)
\end{aligned}$$

for $u_4 \in [0, 1]$. To get its inverse function we solve $\widehat{C}_{4|1,2,3}(u_4 \mid u_1, u_2, u_3) = q$, for $0 < q < 1$, which leads to

$$\widehat{C}^{-1}_{4|1,2,3}(q \mid u_1, u_2, u_3) = \frac{1 + A - \sqrt{A^2 + 1 + 2A(1 - 2q)}}{2A}$$

when $A \neq 0$, where $A = \theta(1-2u_1)(1-2u_2)(1-2u_3)$ (the other solution of the quadratic equation does not belong to the interval $[0, 1]$). In the simulation, as U_1, U_2, U_3 are independent random numbers in $(0, 1)$, the event $A = 0$ has probability zero.

6.1.4 Simulation study

In this section we show a simulation study to estimate the coverage probabilities of the prediction regions when we estimate the parameter in the PHR model for samples of IID random variables.

Table 6.2: Number of $X_{s:20}$ in $\widehat{C}_{90}$ and $\widehat{C}_{50}$ by varying s and N in the exponential model.

	$\widehat{C}_{90}$			$\widehat{C}_{50}$		
s \ N	500	1000	10000	500	1000	10000
13	446	890	8869	259	512	4875
14	449	890	8705	233	483	4665
20	426	853	8309	210	441	4236

Table 6.3: Number of $X_{s:20}$ in $\widehat{C}_{90}$ and $\widehat{C}_{50}$ by varying s and N in the Pareto type II model.

	$\widehat{C}_{90}$			$\widehat{C}_{50}$		
s \ N	500	1000	10000	500	1000	10000
13	439	880	8812	247	472	4841
14	435	858	8711	251	491	4787
20	414	838	8373	202	416	4260

First, assume the exponential model with parameter $\theta = 1$, $\overline{F}(t) = e^{-t}$ for $t \geq 0$. We generate N samples of size 20 and, by supposing that the parameter $\theta > 0$ is unknown, we use $X_{12:20}$ to predict $X_{s:20}$, $s = 13,\ 14,\ 20$. For each sample we use (6.5) with $r = 12$ to estimate θ, and we obtain $\widehat{\theta}_j$, $j = 1, \ldots, N$. Replacing the exact survival function $\overline{F}(t) = e^{-t}$ with $\overline{F}_{\widehat{\theta}_j}(t) = e^{-\widehat{\theta}_j}$, $j = 1, \ldots, N$, we can obtain predictions for $X_{s:20}$ from $X_{12:20}$ for each simulated sample. Our purpose is to estimate the coverage probabilities for the estimated prediction intervals $\widehat{C}_{90}$ and $\widehat{C}_{50}$ varying s and N. The results are listed in Table 6.2.

Furthermore, we perform the same study by choosing as baseline distribution for the PHR model the Pareto type II distribution, $\overline{F}(t) = 1/(1+t)^{\theta}$ for $t \geq 0$. We choose $\theta = 2$. For each sample we use (6.6) with $r = 12$ to estimate θ, and we obtain $\widehat{\theta}_j$, $j = 1, \ldots, N$. Replacing the exact survival function $\overline{F}(t) = 1/(1+t)^2$ with $\overline{F}_{\widehat{\theta}_j}(t) = 1/(1+t)^{\widehat{\theta}_j}$, $j = 1, \ldots, N$, we obtain predictions for $X_{s:20}$ from $X_{12:20}$ for each simulated sample. The results about the coverage probabilities of the estimated prediction intervals are given in Table 6.3. In both cases, the coverage probabilities are a little bit below of the expected values (for the exact model), especially when we predict the last value $X_{20:20}$ from $X_{12:20}$. Note that in both models, we have some extreme upper values (especially in the Pareto model).

6.2 Order dependent Load-Sharing models

In this section, some properties of load-sharing and time-homogeneous load-sharing models are extended to the order dependent version proposed in Foschi et al. [46] and recalled in Definition 5.5. Then, two objectives are pursued. Firstly, a method for simulating samples from the proposed models is discussed and an algorithmic procedure is proposed. Secondly, a study on the predictions of future values from these models is performed. The analysis about predictions is carried out by assuming different levels of knowledge about the sample. Confidence intervals are obtained as well by using convolutions of exponential distributions. Finally, it is explored the problem of predicting the lifetime of a coherent system whose components are distributed according to an order dependent time-homogeneous load-sharing model. The results of this section are given in Buono and Navarro [28].

Under the assumption of an ODTHLS model, the expression of the joint probability density function given in Equation (4.14) simplifies considerably and reduces to

$$
\begin{aligned}
f(t_1,\dots,t_n) = \mu_1(\emptyset)\exp\left[-t_1\sum_{j=1}^{n}\mu_j(\emptyset)\right]\cdot \mu_2(1)\exp\left[-(t_2-t_1)\sum_{j=2}^{n}\mu_j(1)\right]\cdot \\
\dots\ \cdot \mu_{k+1}(1,\dots,k)\exp\left[-(t_{k+1}-t_k)\sum_{j=k+1}^{n}\mu_j(1,\dots,k)\right]\cdot\dots \\
\cdot\, \mu_n(1,\dots,n-1)\exp\left[-(t_n-t_{n-1})\mu_n(1,\dots,n-1)\right],
\end{aligned}
$$

for $t_1 \le t_2 \le \dots \le t_n$. Similar expressions hold when $t_1,\dots,t_n$ are such that $t_{\pi(1)} \le \dots \le t_{\pi(n)}$ for some permutation π of the set $\{1,\dots,n\}$.

Dealing with an ODTHLS model, the following quantities are of great interest

$$
M(i_1,\dots,i_k) = \sum_{h\notin\{i_1,\dots,i_k\}} \mu_h(i_1,\dots,i_k), \tag{6.17}
$$

$$
\rho_j(i_1,\dots,i_k) = \frac{\mu_j(i_1,\dots,i_k)}{M(i_1,\dots,i_k)}, \tag{6.18}
$$

since they are useful in the study of the order statistics of $(X_1,\dots,X_n)$ as stated in the following version of Proposition 4.2 adapted here to the general model.

Proposition 6.11. *Let $(X_1,\dots,X_n)$ be distributed according to an ODTHLS model with parameters $\mu_j(\emptyset), \mu_j(i_1,\dots,i_k)$ and let π be a fixed permutation of $[n]$. Then, for $r = 1,2,\dots,n-1$*

$$
\begin{aligned}
&\mathbb{P}(X_{1:n} = X_{\pi(1)},\dots,X_{r:n} = X_{\pi(r)}) = \\
&\quad \rho_{\pi(1)}(\emptyset)\rho_{\pi(2)}(\pi(1))\rho_{\pi(3)}(\pi(1),\pi(2))\cdot\dots\cdot\rho_{\pi(r)}(\pi(1),\dots,\pi(r-1))
\end{aligned}
$$

and

$$\mathbb{P}(X_{1:n} = X_{\pi(1)}, \dots, X_{n:n} = X_{\pi(n)}) =$$
$$\rho_{\pi(1)}(\emptyset)\rho_{\pi(2)}(\pi(1))\rho_{\pi(3)}(\pi(1), \pi(2)) \cdot \dots \cdot \rho_{\pi(n-1)}(\pi(1), \dots, \pi(n-2)).$$

Proof. The proof follows by the product rule of probability and in analogy with [110]. There, the parameters as $\rho_{\pi(k)}(\pi(1), \dots, \pi(k-1))$ are independent of the order of the indices $\pi(1), \dots,$ $\pi(k-1)$, but all the passages still hold by considering them as an ordered sequence. For instance, $\rho_{\pi(k)}(\pi(1), \dots, \pi(k-1))$ is related to the behavior of the k-th order statistic, given that $X_{1:n} = X_{\pi(1)}, \dots, X_{k-1:n} = X_{\pi(k-1)}$, and the importance of the order of the indices emerges. □

In order to state the following result, let us denote by $\Lambda^{(r)}$ a vector $(\lambda_1, \dots, \lambda_r) \in \mathbb{R}_+^r$ and by $\overline{G}_{\Lambda^{(r)}}(t)$ the survival function of the random variable $S_r = \sum_{s=1}^r \Gamma_s$, where $\Gamma_1, \dots, \Gamma_r$ are independent random variables with exponential distributions of parameters (hazard rates) $\lambda_1, \dots, \lambda_r$, respectively. Moreover, for π permutation of $[n]$ and $r \in [n]$, we place

$$\Lambda^{(r)}(\pi) = (M(\emptyset), M(\pi(1)), \dots, M(\pi(1), \dots, \pi(r-1))).$$

Then, we have the following proposition which is the adapted version to the ODTHLS model of Proposition 4.3.

Proposition 6.12. *Let $(X_1, \dots, X_n)$ be distributed according to an ODTHLS model with parameters $\mu_j(\emptyset)$ and $\mu_j(i_1, \dots, i_k)$. Then, for any $t > 0$ and $j \in [n]$,*

$$\mathbb{P}(X_{1:n} > t | X_{1:n} = X_j) = \exp(-tM(\emptyset)),$$

and for any permutation π of $[n]$ and $k = 2, \dots, n$,

$$\mathbb{P}(X_{k:n} > t | X_{1:n} = X_{\pi(1)}, \dots, X_{k-1:n} = X_{\pi(k-1)}, X_{k:n} = X_{\pi(k)}) = \overline{G}_{\Lambda^{(k)}(\pi)}(t). \tag{6.19}$$

To prove Proposition 6.12, an important property of interarrival times of ODTHLS models follows. In fact, in [110] it is observed that conditioning on the event $(X_{1:n} = X_{\pi(1)}, \dots,$ $X_{k:n} = X_{\pi(k)})$, the interarrival times $X_{1:n}, X_{2:n} - X_{1:n}, \dots, X_{k:n} - X_{k-1:n}$ can be seen as independent random variables, exponentially distributed with parameters $M(\emptyset), M(\pi(1)), \dots,$ $M(\pi(1), \dots, \pi(k-1))$, respectively. By using this fact, we can prove the following results which are key tools for predictions and simulations.

Remark 6.2. From Proposition 6.12 it readily follows the independence of the events $(X_{1:n} >$

t) and $(X_{1:n} = X_j)$. In fact

$$\begin{aligned}\mathbb{P}(X_{1:n} > t) &= \sum_{j=1}^{n} \mathbb{P}(X_{1:n} = X_j)\mathbb{P}(X_{1:n} > t | X_{1:n} = X_j) \\ &= \exp(-tM(\emptyset)) \sum_{j=1}^{n} \mathbb{P}(X_{1:n} = X_j) \\ &= \exp(-tM(\emptyset)).\end{aligned}$$

Furthermore, we note that $M(\emptyset), M(\pi(1)), \ldots, M(\pi(1), \ldots, \pi(k-1))$ do not depend on $\pi(k)$ and then the interarrival times $X_{1:n}, X_{2:n} - X_{1:n}, \ldots, X_{k:n} - X_{k-1:n}$, can be seen as independent random variables, exponentially distributed with parameters $M(\emptyset), M(\pi(1)), \ldots, M(\pi(1), \ldots, \pi(k-1))$, respectively, given $(X_{1:n} = X_{\pi(1)}, \ldots, X_{k-1:n} = X_{\pi(k-1)})$. Hence, under this conditioning event, from (6.19) the distribution of $X_{k:n}$ is a convolution of k exponential distributions.

6.2.1 Simulation of ODTHLS model

Let $(X_1, \ldots, X_n)$ be a random vector satisfying the ODTHLS model with parameters $\mu_j(\emptyset)$ and $\mu_j(i_1, \ldots, i_k)$, $I = \{i_1, \ldots, i_k\} \subset [n]$ and $j \notin I$. There are $n!$ ways to order the variables $X_1, \ldots, X_n$ and, for each one, the probability that such an order corresponds to the sequence given by the order statistics is given in Proposition 6.11. These probabilities depend on the parameters of the model, hence, once they are fixed, it is possible to choose one of the permutations by a random generation. For instance, it is possible to simulate the random choice between all the permutations by generating a uniform number in $(0, 1)$.

Suppose the permutation π is randomly selected according to these probabilities. Hence, we have $X_{1:n} = X_{\pi(1)}, \ldots, X_{n:n} = X_{\pi(n)}$. Then, by Proposition 6.12, given that the minimum is assumed in $X_{\pi(1)}$, it is distributed as an exponential random variable with parameter $M(\emptyset)$. Actually, from Remark 6.2 we note that, by the properties of ODTHLS models, the distribution of the minimum is not affected by which is the random variable in which it is assumed. Then, it is possible to simulate the minimum by a random generator of an exponential random variable with parameter $M(\emptyset)$.

Let k be a natural number between 2 and n and suppose we have already simulated $X_{1:n}, \ldots, X_{k-1:n}$. Now, by using that conditioning on the event $(X_{1:n} = X_{\pi(1)}, \ldots, X_{k-1:n} = X_{\pi(k-1)})$, the interarrival times $X_{1:n}, X_{2:n} - X_{1:n}, \ldots, X_{k:n} - X_{k-1:n}$ can be seen as independent random variables, exponentially distributed with parameters $M(\emptyset), M(\pi(1)), \ldots, M(\pi(1), \ldots, \pi(k-1))$, respectively, the interarrival time $X_{k:n} - X_{k-1:n}$ can be simulated by generating an exponential number with parameter $M(\pi(1), \ldots, \pi(k-1))$. Then, the simulation of $X_{k:n}$ is given by summing this exponential number with the simulation of $X_{k-1:n}$ obtained in the previous step.

We want to emphasize that, once the permutation is fixed, the interarrival times can be generated all at the same time. If we denote by Z_j, $j \in [n]$, the j-th interarrival time, i.e., $Z_1 = X_{1:n}$, $Z_2 = X_{2:n} - X_{1:n}, \ldots, Z_n = X_{n:n} - X_{n-1:n}$ obtained for the permutation π, then the simulation of the k-th order statistic is given as $X_{k:n} = \sum_{j=1}^{k} Z_j$. Finally, by using the permutation π, the simulated values for $X_1, \ldots, X_n$ are obtained as $X_{\pi(1)} = X_{1:n}, \ldots, X_{\pi(n)} = X_{n:n}$.

The algorithm procedure can be summarized as follows.

Algorithm 6.1.

Step 1. Choose π according to the probabilities given in Proposition 6.11.

Step 2. Simulate n independent exponential distributions $Z_1, \ldots, Z_n$ with respective parameters $M(\emptyset), M(\pi(1)), \ldots, M(\pi(1), \ldots, \pi(n-1))$.

Step 3. Compute $X_{k:n} = Z_1 + \cdots + Z_k$, for $k = 1, \ldots, n$.

Step 4. Compute $X_{\pi(k)} = X_{k:n}$, for $k = 1, \ldots, n$.

Note that this algorithm can also be applied to simulate samples from THLS models since they are particular models of ODTHLS ones. Let us see an example.

Example 6.4. Let (X_1, X_2, X_3) be distributed according to an ODTHLS model with parameters defined as follows

$$\begin{aligned}
&\mu_1(\emptyset) = 1, \quad \mu_1(2) = 2, \quad \mu_1(3) = 1, \quad \mu_1(2,3) = \mu_1(3,2) = 3,\\
&\mu_2(\emptyset) = 2, \quad \mu_2(1) = 1, \quad \mu_2(3) = 3, \quad \mu_2(1,3) = \mu_2(3,1) = 2,\\
&\mu_3(\emptyset) = 2, \quad \mu_3(1) = 2, \quad \mu_3(2) = 1, \quad \mu_3(1,2) = \mu_3(2,1) = 2.
\end{aligned}$$

We note that it is also a THLS model since $\mu_i(j,k) = \mu_i(k,j)$ for all distinct i, j and k. Hence, from (6.17) and (6.18) we have

$$\begin{aligned}
&M(\emptyset) = 5, \quad M(1) = 3, \quad M(2) = 3, \quad M(3) = 4,\\
&M(1,2) = M(2,1) = 2, \quad M(1,3) = M(3,1) = 2, \quad M(2,3) = M(3,2) = 3,
\end{aligned}$$

from which

$$\begin{aligned}
&\rho_1(\emptyset) = \frac{1}{5}, \quad \rho_2(\emptyset) = \frac{2}{5}, \quad \rho_3(\emptyset) = \frac{2}{5}, \rho_2(1) = \frac{1}{3}, \quad \rho_3(1) = \frac{2}{3},\\
&\rho_1(2) = \frac{2}{3}, \quad \rho_3(2) = \frac{1}{3}, \rho_1(3) = \frac{1}{4}, \quad \rho_2(3) = \frac{3}{4},
\end{aligned}$$

and, naturally,

$$\rho_1(2,3) = \rho_1(3,2) = \rho_2(1,3) = \rho_2(3,1) = \rho_3(1,2) = \rho_3(2,1) = 1.$$

For $n = 3$ there are six possible permutations and from Proposition 6.11 the corresponding probabilities are given as

$$\mathbb{P}(X_{1:3} = X_1, X_{2:3} = X_2, X_{3:3} = X_3) = \frac{1}{15},$$
$$\mathbb{P}(X_{1:3} = X_1, X_{2:3} = X_3, X_{3:3} = X_2) = \frac{2}{15},$$
$$\mathbb{P}(X_{1:3} = X_2, X_{2:3} = X_1, X_{3:3} = X_3) = \frac{4}{15},$$
$$\mathbb{P}(X_{1:3} = X_2, X_{2:3} = X_3, X_{3:3} = X_1) = \frac{2}{15},$$
$$\mathbb{P}(X_{1:3} = X_3, X_{2:3} = X_1, X_{3:3} = X_2) = \frac{1}{10},$$
$$\mathbb{P}(X_{1:3} = X_3, X_{2:3} = X_2, X_{3:3} = X_1) = \frac{3}{10}.$$

By generating a uniform number in $(0, 1)$ and accordingly to the probabilities given above, the permutation $(2, 1, 3)$ is chosen. Hence, three exponential numbers are generated with parameters $M(\emptyset) = 5$, $M(2) = 3$, and $M(2, 1) = 2$, respectively. In this way, the simulated interarrival times are 0.17166, 0.14498, 0.25606 and then the simulated values of the order statistics of our model are, respectively, 0.17166, $0.31663 = 0.17166 + 0.14498$, $0.57270 = 0.31663 + 0.25606$. Finally, since we have fixed the permutation $(2, 1, 3)$, these values represent a simulation of X_2, X_1 and X_3, respectively.

6.2.2 Predictions

Let $(X_1, \ldots, X_n)$ follow an ODTHLS model with parameters $\mu_j(\emptyset)$ and $\mu_j(i_1, \ldots, i_k)$, $I = \{i_1, \ldots, i_k\} \subset [n]$ and $j \notin I$. Let us suppose to know the values and the variables corresponding to $X_{1:n}, X_{2:n}, \ldots, X_{k:n}$, for $k < n$. Our first purpose is to predict the next failure time $X_{k+1:n}$. As observed above under this model, given $(X_{1:n} = X_{\pi(1)}, \ldots, X_{k:n} = X_{\pi(k)}, X_{k+1:n} = X_{\pi(k+1)})$, the interarrival times $X_{1:n}, X_{2:n} - X_{1:n}, \ldots, X_{k:n} - X_{k-1:n}, X_{k+1:n} - X_{k:n}$ can be seen as independent random variables, exponentially distributed with parameters $M(\emptyset), M(\pi(1)), \ldots, M(\pi(1), \ldots, \pi(k-1)), M(\pi(1), \ldots, \pi(k))$, respectively. Hence, conditioning on our information, the interarrival time $Z_{k+1} = X_{k+1:n} - X_{k:n}$ is exponential with parameter $M(\pi(1), \ldots, \pi(k))$ and its value can be estimated by its median. If we denote by $m_{M(\pi(1),\ldots,\pi(k))} = \frac{\log 2}{M(\pi(1),\ldots,\pi(k))}$ the median of an exponential distribution with parameter $M(\pi(1), \ldots, \pi(k))$, then the estimation of $X_{k+1:n}$ is given by

$$\hat{X}_{k+1:n} = \mathfrak{m}(t_k) = t_k + m_{M(\pi(1),\ldots,\pi(k))} = t_k + \frac{\log 2}{M(\pi(1), \ldots, \pi(k))},$$

where t_k is the value assumed for the k-th order statistic. An alternative way of predicting $X_{k+1:n}$ is based on the mean of the exponential distribution with parameter $M(\pi(1), \ldots, \pi(k))$

and this prediction is obtained as

$$\tilde{X}_{k+1:n} = t_k + \frac{1}{M(\pi(1),\ldots,\pi(k))}.$$

Here, we prefer to use the prediction given by the median since we are interested in constructing confidence intervals for our predictions. In fact, if we want to get a confidence $\gamma = \beta - \alpha$, where $\alpha, \beta, \gamma \in (0,1)$ and q_α and q_β are the respective quantiles of the exponential distribution with parameter $M(\pi(1),\ldots,\pi(k))$, then we use that

$$\mathbb{P}\left(t_k + q_\alpha \le X_{k+1:n} \le t_k + q_\beta | X_{1:n} = X_{\pi(1)},\ldots,X_{k:n} = X_{\pi(k)} = t_k\right) = \gamma.$$

For example, the centered 90% confidence band is obtained with $\beta = 0.95$ and $\alpha = 0.05$ as

$$C_{90} = [t_k + q_{0.05}, t_k + q_{0.95}] = \left[t_k - \frac{\log(0.95)}{M(\pi(1),\ldots,\pi(k))}, t_k - \frac{\log(0.05)}{M(\pi(1),\ldots,\pi(k))}\right].$$

Note that the prediction $\mathfrak{m}(t_k)$ always belongs to that interval.

In a different scenario, we suppose to know how the realizations of the variables are ordered up to a certain index, for instance k, without knowing the assumed values. Hence, we have information just about $\pi(1),\ldots,\pi(k)$, but not on $t_1,\ldots,t_k$, and our purpose is to predict $X_{k+1:n}$. Two reasonable ways of predicting it are given by estimating each interarrival time through the median or the mean and then provide the estimate of $X_{k+1:n}$ as

$$\hat{X}_{k+1:n} = m_{M(\emptyset)} + m_{M(\pi(1))} + \cdots + m_{M(\pi(1),\ldots,\pi(k))}, \tag{6.20}$$

$$\tilde{X}_{k+1:n} = \frac{1}{M(\emptyset)} + \frac{1}{M(\pi(1))} + \cdots + \frac{1}{M(\pi(1),\ldots,\pi(k))}, \tag{6.21}$$

where the first is the prediction based on the median and the second on the mean. Moreover, by observing that

$$X_{k+1:n} = X_{1:n} + (X_{2:n} - X_{1:n}) + \cdots + (X_{k+1:n} - X_{k:n}),$$

another option to predict $X_{k+1:n}$ is to obtain the median of the convolution given above, i.e., the convolution of $k+1$ independent exponential distributions with parameters $M(\emptyset)$, $M(\pi(1)),\ldots,M(\pi(1),\ldots,\pi(k))$. Note that the mean of that convolution coincides with (6.21).

Now, let us suppose $k < n-1$ and that our purpose is to predict $X_{k+2:n}$. By the assumptions of the model, the value of $X_{k+2:n}$ will depend on which path will be traversed to move from $X_{k:n}$ to $X_{k+2:}$, i.e., on which of the $n-k$ available alternatives will be assumed for $X_{k+1:n}$. For $j \notin \{\pi(1),\ldots,\pi(k)\}$, we have

$$\mathbb{P}(X_{k+1:n} = X_j | X_{1:n} = X_{\pi(1)},\ldots,X_{k:n} = X_{\pi(k)}) = \rho_j(\pi(1),\ldots,\pi(k)). \tag{6.22}$$

Proceeding as above, we can predict the value of $X_{k+1:n}$, namely $\hat{X}_{k+1:n}$. Then, by using this value and the median regression, we can predict the value of $X_{k+2:n}$ in $n-k$ different ways depending on which variable is the $(k+1)$-th order statistic. Hence, for each $j \notin \{\pi(1),\ldots,\pi(k)\}$, we obtain a prediction of $X_{k+2:n}$, namely $\hat{X}^{(j)}_{k+2:n}$. Finally, based on (6.22), the final prediction of $X_{k+2:n}$ is obtained by the weighted mean of all $n-k$ predictions as

$$\hat{X}_{k+2:n} = \sum_{j \notin \{\pi(1),\ldots,\pi(k)\}} \rho_j(\pi(1),\ldots,\pi(k)) \hat{X}^{(j)}_{k+2:n},$$

or, equivalently, since the values $\rho_j(\pi(1),\ldots,\pi(k))$ sum to one,

$$\begin{aligned} \hat{X}_{k+2:n} &= \hat{X}_{k+1:n} + \sum_{j \notin \{\pi(1),\ldots,\pi(k)\}} \rho_j(\pi(1),\ldots,\pi(k)) m_{M(\pi(1),\ldots,\pi(k),j)} \\ &= \hat{X}_{k+1:n} + \sum_{j \notin \{\pi(1),\ldots,\pi(k)\}} \rho_j(\pi(1),\ldots,\pi(k)) \frac{\log 2}{M(\pi(1),\ldots,\pi(k),j)}. \end{aligned}$$

We remark that the above prediction can be done also in terms of the mean. In order to set a different prediction and to obtain the related confidence bands, we state the following result.

Proposition 6.13. *Let $(X_1,\ldots,X_n)$ follow an ODTHLS model with parameters $\mu_j(\emptyset)$ and $\mu_j(i_1,\ldots,i_k)$, $I = \{i_1,\ldots,i_k\} \subset [n]$ and $j \notin I$. Let π be a fixed permutation of $[n]$ and $k < n-1$. Then,*

$$\begin{aligned} &\mathbb{P}(X_{k+2:n} - X_{k:n} > t | X_{1:n} = X_{\pi(1)} = t_1, \ldots, X_{k:n} = X_{\pi(k)} = t_k) \\ &= \sum_{j \neq \pi(1),\ldots,\pi(k)} \rho_j(\pi(1),\ldots,\pi(k)) \overline{G}_{\Upsilon_j^{(k)}(\pi)}(t), \end{aligned} \tag{6.23}$$

where $\overline{G}_{\Upsilon_j^{(k)}(\pi)}(t)$ is the survival function of the random variable Y_1+Y_2, being Y_1 and Y_2 independent random variables with exponential distributions of parameters (hazard rates) $M(\pi(1), \ldots,\pi(k))$ and $M(\pi(1),\ldots,\pi(k),j)$, respectively.

Proof. The result follows by the law of total probability and by Proposition 6.12 observing that $X_{k+2:n} - X_{k:n}$ can be seen as the sum of two independent interarrival times, $X_{k+2:n} - X_{k:n} = (X_{k+2:n} - X_{k+1:n}) + (X_{k+1:n} - X_{k:n})$ with exponential distributions. □

Conditioning on the observed history, the interarrival time $X_{k+2:n} - X_{k:n}$ is a mixture of $n-k$ distributions which are sums of two independent exponential distributions, not necessarily with the same parameters. We refer for example to (4.8) and (4.9) in [87] for the analytical expressions of the survival functions of such distributions, see also [98] p. 299. In particular, it is necessary to distinguish between the case in which the exponential distributions have the same parameter or not. If Y_1 and Y_2 are independent and exponentially distributed with parameters λ and μ, $\lambda \neq \mu$, then the survival function of $Y = Y_1 + Y_2$ is

$$\overline{F}_Y(t) = \frac{\mu}{\mu-\lambda} e^{-\lambda t} - \frac{\lambda}{\mu-\lambda} e^{-\mu t}, \tag{6.24}$$

for $t \geq 0$. In the case $\lambda = \mu$, the survival function of Y is given, for $t \geq 0$, as

$$\overline{F}_Y(t) = (1+\lambda t)e^{-\lambda t}. \tag{6.25}$$

The median of such distributions can also be a good prediction for $X_{k+2:n}$. Numerical methods should be used to get that median from (6.23). Then, if we want to get a confidence $\gamma = \beta - \alpha$, where $\alpha, \beta, \gamma \in (0,1)$ and q_α and q_β are the respective quantiles of the distribution given in Proposition 6.13, we use that

$$\mathbb{P}\left(t_k + q_\alpha \le X_{k+2:n} \le t_k + q_\beta | X_{1:n} = X_{\pi(1)}, \ldots, X_{k:n} = X_{\pi(k)}, X_{k:n} = t_k\right) = \gamma.$$

Remark 6.3. By proceeding in this way, it is possible to estimate each $X_{s:n}$ for $s > k$. As seen above, with the increase of s there will be more terms in the convolutions. In particular, by supposing to know in which variables are assumed $X_{1:n}, \ldots, X_{k:n}$ and the corresponding values, the estimation of $X_{s:n}$ will be based on the sum of $\frac{(n-k)!}{(n-s+1)!}$ terms. Moreover, it is also possible to construct confidence bands by giving a result similar to Proposition 6.13. In this case, we will need distributions constructed as the sum of $s-k$ independent exponential distributions. Such distributions have been studied in [1].

Example 6.5. Let us consider the ODTHLS model given in Example 6.4 and suppose that the realization of the sample is the one that we have simulated there, i.e., $X_1 = 0.31663$, $X_2 = 0.17166$ and $X_3 = 0.57270$. Suppose we just know $X_{1:3} = X_2 = 0.17166$ and we want to predict $X_{2:3}$ and $X_{3:3}$. Proceeding as described above, the mean and the median predictions of $X_{2:3} = 0.31663$ are

$$\tilde{X}_{2:3} = X_{1:3} + \frac{1}{M(2)} = 0.50499$$

and

$$\hat{X}_{2:3} = \mathfrak{m}(X_{1:3}) = X_{1:3} + \frac{\log 2}{M(2)} = 0.40270,$$

respectively. Furthermore, we obtain the centered 90% and 50% confidence bands as

$$C_{90} = \left[X_{1:3} - \frac{\log(0.95)}{M(2)}, X_{1:3} - \frac{\log(0.05)}{M(2)}\right] = [0.18875, 1.17023]$$

and $C_{50} = [0.26755, 0.63375]$. In this case, the exact value of $X_{2:3}$ belongs to both regions. Once $X_{2:3}$ has been predicted, proceeding as described above, also $X_{3:3}$ can be predicted. In this case the prediction of $X_{3:3} = 0.57270$ is given by

$$\hat{X}_{3:3} = \hat{X}_{2:3} + \rho_1(2)\frac{\log 2}{M(2,1)} + \rho_3(2)\frac{\log 2}{M(2,3)} = 0.40270 + \frac{2}{3}\cdot\frac{\log 2}{2} + \frac{1}{3}\cdot\frac{\log 2}{3} = 0.71077.$$

From Proposition 6.13 we get a different prediction for $X_{3:3}$. We have

$$\overline{G}_{3|1}(t) = \mathbb{P}(X_{3:3} - X_{1:3} > t | X_{1:3} = X_2 = 0.17166) = \rho_1(2)\overline{G}_{Y_{1,1}+Y_{1,2}}(t) + \rho_3(2)\overline{G}_{Y_{2,1}+Y_{2,2}}(t),$$

where $Y_{1,1}$ and $Y_{1,2}$ are independent and exponentially distributed with parameters $M(2) = 3$ and $M(2,1) = 2$, respectively, and $Y_{2,1}$ and $Y_{2,2}$ are independent and exponentially distributed

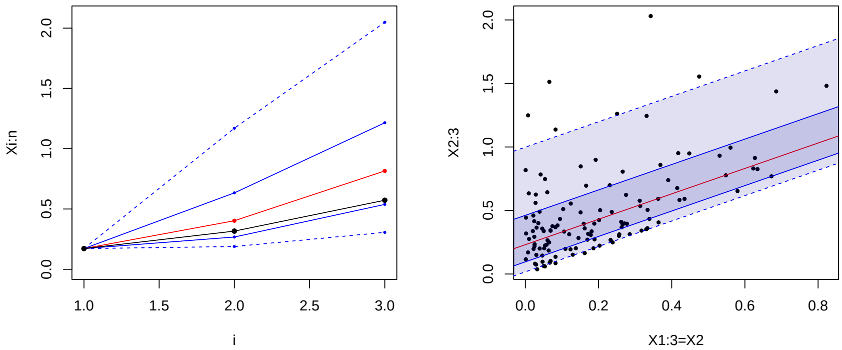

Figure 6.10: Predictions (red) for $X_{s:3}$ from $X_{1:3}$ for $s = 2, 3$ jointly with the exact values (black points) for a simulated sample from an ODTHLS model (see Example 6.5). The blue lines represent the limits for the 50% (continuous lines) and the 90% (dashed lines) confidence intervals (left). Scatterplots of a simulated sample from $(X_{1:3}, X_{2:3})$, for the case $X_{1:3} = X_2$, for the ODTHLS model in Example 6.5 jointly with the theoretical median regression curve (red) and 50% (dark grey) and 90% (light grey) confidence bands (right).

with parameters $M(2) = 3$ and $M(2,3) = 3$, respectively. Hence, by referring to the analytical expressions in (6.24) and (6.25), we obtain

$$\overline{G}_{3|1}(t) = \rho_1(2)\frac{M(2)e^{-M(2,1)t} - M(2,1)e^{-M(2)t}}{M(2) - M(2,1)} + \rho_3(2)(1 + M(2)t)e^{-M(2)t},$$

where the second term is related to the sum of two independent exponential distributions with the same parameter $M(2) = M(2,3) = 3$. Hence, by solving $\overline{G}_{3|1}(t) = 0.5$ we obtain a prediction for the difference $X_{3:3} - X_{1:3}$ that is 0.64409, from which

$$\hat{X}_{3:3} = t_1 + 0.64409 = 0.81575.$$

By resolving $\overline{G}_{3|1}(t) = \alpha$, for $\alpha = 0.05, 0.25, 0.75, 0.95$, we obtain the 90% and 50% centered confidence bands as $C_{90} = [0.30639, 2.04858]$ and $C_{50} = [0.53811, 1.21520]$. Observe that $X_{3:3} = 0.57270$ belongs to both regions. In Figure 6.10, left, we plot these predictions (red) for $X_{2:3}, X_{3:3}$ from $X_{1:3}$ jointly with the exact values (black points) and the confidence bands.

To see better what happens with these predictions we simulate $N = 300$ predictions of this kind, that is, 300 samples of size 3. Let us consider the case in which we predict $X_{2:3}$ from $X_{1:3}$. In order to give the results in a more readable way, we group them in three classes based on which is the component corresponding to the minimum order statistic. The data are plotted in Figures 6.10, right, and 6.11. There we can see that the confidence bands represent very well the dispersion of the majority of data (except some extreme values). In these samples,

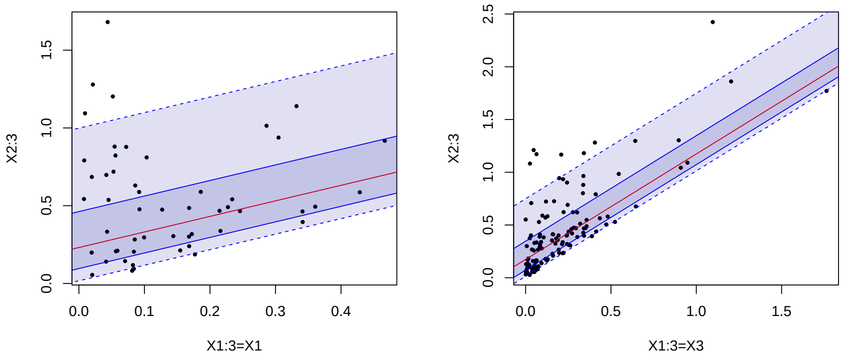

Figure 6.11: Scatterplots of a simulated sample from $(X_{1:3}, X_{2:3})$ for the ODTHLS model in Example 6.5 jointly with the theoretical median regression curves (red) and 50% (dark grey) and 90% (light grey) confidence bands for the case $X_{1:3} = X_1$ (left) and $X_{1:3} = X_3$ (right).

the minimum is assumed in X_1, X_2 and X_3 for 52, 122 and 126 times, respectively. These values are consistent with the expected ones given by $\rho_1(\emptyset) \cdot 300 = 60$, $\rho_2(\emptyset) \cdot 300 = 120$ and $\rho_3(\emptyset) \cdot 300 = 120$. If the minimum is assumed in X_1, C_{50} contains 24 data and C_{90} contains 45 while 4 data are above the upper limit and 3 are below the bottom limit. If the minimum is assumed in X_2, C_{50} contains 69 data and C_{90} contains 109 while 6 data are above the upper limit and 7 are below the bottom limit. If the minimum is assumed in X_3, C_{50} contains 58 data and C_{90} contains 112 data while 7 data are above the upper limit and 7 are below the bottom limit. Note that the confidence bands depend on which component fails first.

Now, suppose we know the minimum is assumed by X_2 and we have no information about its value. Then, as described in (6.20) and (6.21), the predictions for the first and the second order statistics based on the median (left) and the mean (right) are given by

$$\hat{X}_{1:3} = \frac{\log 2}{M(\emptyset)} = 0.13863, \qquad \tilde{X}_{1:3} = \frac{1}{M(\emptyset)} = 0.2,$$
$$\hat{X}_{2:3} = \frac{\log 2}{M(\emptyset)} + \frac{\log 2}{M(2)} = 0.36968, \qquad \tilde{X}_{2:3} = \frac{1}{M(\emptyset)} + \frac{1}{M(2)} = 0.53333.$$

Moreover, the prediction of $X_{2:3}$ can be obtained also by the median of the convolution $X_{1:3} + (X_{2:3} - X_{1:3})$. In fact, given that $X_{1:3} = X_2$, these interarrival times are independent and exponential with parameters $M(\emptyset) = 5$ and $M(2) = 3$ and the survival function of their convolution is given by (6.24). The median of such a distribution can be numerically computed and gives another prediction for $X_{2:3}$ as 0.44139. Of course, if we use the mean we get again the value 0.53333.

Furthermore, if we know that the first and the second order statistics are assumed in X_2 and X_1, respectively, the maximum $X_{3:3}$ can be predicted by the median and the mean, respectively, as

$$\hat{X}_{3:3} = \frac{\log 2}{M(\emptyset)} + \frac{\log 2}{M(2)} + \frac{\log 2}{M(2,1)} = 0.71625,$$

and

$$\tilde{X}_{3:3} = \frac{1}{M(\emptyset)} + \frac{1}{M(2)} + \frac{1}{M(2,1)} = 1.03333.$$

In addition, we can obtain the prediction of $X_{3:3}$ based on the convolution $Y = X_{1:3} + (X_{2:3} - X_{1:3}) + (X_{3:3} - X_{2:3})$, given that $X_{1:3} = X_2, X_{2:3} = X_1$. The interarrival times are independent and have exponential distributions with parameters $M(\emptyset) = 5$, $M(2) = 3$ and $M(2,1) = 2$. The survival function of this convolution can be obtained by specializing the result of [1] to the case of three exponential distributions with different parameters and it is expressed, for $t \geq 0$, as

$$\begin{aligned} \overline{G}_Y(t) \quad = \quad & \frac{M(2)M(2,1)}{(M(2)-M(\emptyset))(M(2,1)-M(\emptyset))} e^{-M(\emptyset)t} \\ & + \frac{M(\emptyset)M(2,1)}{(M(\emptyset)-M(2))(M(2,1)-M(2))} e^{-M(2)t} \\ & + \frac{M(\emptyset)M(2)}{(M(\emptyset)-M(2,1))(M(2)-M(2,1))} e^{-M(2,1)t}. \end{aligned} \tag{6.26}$$

The median of such a distribution can be numerically computed and gives another prediction for $X_{3:3}$ as $X^*_{3:3} = 0.90225$. Note that $\overline{G}_Y$ can also be used to get the confidence intervals for that prediction. We have $C_{90} = [0.26708, 2.24684]$ and $C_{50} = [0.57337, 1.35021]$. The exact value 0.57270 belongs to C_{90} but does not belong to C_{50}.

Next, suppose we have even less information and we just know that the first and the second order statistics are assumed by X_1 and X_2 but we have not the possibility to establish which one is $X_{1:3}$ or $X_{2:3}$. There are two possible scenarios corresponding to the permutations $(1,2,3)$ and $(2,1,3)$. Conditioning on the information $X_{3:3} = X_3$, it follows

$$\begin{aligned} \mathbb{P}(X_{1:3} = X_1, X_{2:3} = X_2 | X_{3:3} = X_3) \quad &= \quad \frac{1}{5}, \\ \mathbb{P}(X_{1:3} = X_2, X_{2:3} = X_1 | X_{3:3} = X_3) \quad &= \quad \frac{4}{5}. \end{aligned}$$

Thus, the predictions for the first, second and third order statistics are obtained as

$$\begin{aligned} \hat{X}_{1:3} &= \frac{\log 2}{M(\emptyset)} = 0.13863, \\ \hat{X}_{2:3} &= \frac{\log 2}{M(\emptyset)} + \frac{1}{5} \cdot \frac{\log 2}{M(1)} + \frac{4}{5} \cdot \frac{\log 2}{M(2)} = 0.36968, \\ \hat{X}_{3:3} &= \frac{\log 2}{M(\emptyset)} + \frac{1}{5}\left(\frac{\log 2}{M(1)} + \frac{\log 2}{M(1,2)}\right) + \frac{4}{5}\left(\frac{\log 2}{M(2)} + \frac{\log 2}{M(2,1)}\right) = 0.71625. \end{aligned}$$

In this case, we have obtained the same predictions of the case in which we know that $X_{1:3} = X_2$ and $X_{2:3} = X_1$, but this is only due to the assumptions $M(1) = M(2)$ and $M(1,2) = M(2,1)$

and the same holds for the predictions based on the mean or on the convolutions. In fact, if we consider the same model except for $\mu_3(1,2) = 3$ (which implies $M(1,2) = 3 \neq M(2,1) = 2$), then the median prediction of $X_{3:3}$ knowing that $X_{1:3} = X_2$ and $X_{2:3} = X_1$ is still 0.71625, but without knowing which one between X_1 and X_2 is the minimum and which one the second order statistic, the prediction of $X_{3:3}$ becomes $\hat{X}_{3:3} = 0.69315$. Under these assumptions, the mean prediction of $X_{3:3}$ is $\tilde{X}_{3:3} = 1$.

Finally, we obtain a prediction based on convolutions by giving a weight of 0.2 and 0.8, respectively, to the medians of the convolutions of independent and exponential distributions with parameters $M(\emptyset) = 5, M(1) = 3, M(1,2) = 3$ and $M(\emptyset) = 5, M(2) = 3, M(2,1) = 2$. The survival function of the latter is equal to the one given in (6.26), whereas the former has a different expression since two of the three parameters coincide. In particular, from [1], the convolution of three independent exponential distributions of parameters $M(\emptyset), M(1), M(1)(= M(1,2))$ has the following survival function, for $t \geq 0$,

$$\begin{aligned} \overline{G}(t) \quad = \quad & \frac{M(1)^2}{(M(1) - M(\emptyset))^2} e^{-M(\emptyset)t} - \frac{M(\emptyset)M(1)}{(M(\emptyset) - M(1))^2} e^{-M(1)t} \\ & + \frac{M(\emptyset)M(1)}{M(\emptyset) - M(1)} t e^{-M(1)t} + \frac{M(\emptyset)}{M(\emptyset) - M(1)} e^{-M(1)t}, \end{aligned} \tag{6.27}$$

and its median is 0.76649. Hence, the prediction for $X_{3:3}$ based on the convolutions is given as

$$X^*_{3:3} = \frac{1}{5} \cdot 0.76649 + \frac{4}{5} \cdot 0.90225 = 0.87510.$$

Moreover, a different prediction for $X_{3:3}$ can be obtained by using the median of the mixture of the survival functions given in (6.26) and (6.27) with 0.8 and 0.2 as weights, respectively. The prediction obtained in this way is $X^*_{3:3} = 0.87229$ and its advantage is that we can give the confidence regions. The centered 90% and 50% confidence bands are $[0.25848, 2.17710]$ and $[0.55452, 1.30560]$, respectively. Note that the exact value $X_{3:3} = 0.57270$ belongs to both the regions.

In the following example, we analyze the problem of the predictions dealing with an ODTHLS model for which we need to estimate the parameters.

Example 6.6. Let (X_1, X_2, X_3) be distributed according to an ODTHLS model with parameters defined as follows

$$\begin{aligned} &\mu_1(\emptyset) = 1, \quad \mu_1(2) = 2, \quad \mu_1(3) = 1, \quad \mu_1(2,3) = 3, \quad \mu_1(3,2) = 1, \\ &\mu_2(\emptyset) = 2, \quad \mu_2(1) = 1, \quad \mu_2(3) = 3, \quad \mu_2(1,3) = 2, \quad \mu_2(3,1) = 1, \\ &\mu_3(\emptyset) = 2, \quad \mu_3(1) = 2, \quad \mu_3(2) = 1, \quad \mu_3(1,2) = 2, \quad \mu_3(2,1) = 1. \end{aligned}$$

Hence, we have

$$\begin{aligned} &M(\emptyset) = 5, \quad M(1) = 3, \quad M(2) = 3, \quad M(3) = 4, \\ &M(1,2) = 2, \quad M(2,1) = 1, \quad M(1,3) = 2, \quad M(3,1) = 1, \quad M(2,3) = 3, \quad M(3,2) = 1. \end{aligned}$$

Suppose we do not know the parameters of the model and we have historical data related to $N = 300$ samples. For those samples we know how X_1, X_2 and X_3 are ordered and their values. Then, we can estimate the parameters of the model through the values of interarrival times. Since the minimum is distributed as an exponential distribution with parameter $M(\emptyset)$, it can be estimated as

$$\hat{M}(\emptyset) = \frac{N}{\sum_{i=1}^{N} X_{1:3}^{(i)}} = 5.19128,$$

where $X_{1:3}^{(i)}$ is the minimum in the i-th sample.

In order to estimate the other parameters, we need to group the data by the corresponding permutation. Let $\pi_1 = (1,2,3)$, $\pi_2 = (1,3,2)$, $\pi_3 = (2,1,3)$, $\pi_4 = (2,3,1)$, $\pi_5 = (3,1,2)$ and $\pi_6 = (3,2,1)$ and define $\mathcal{P}_j$ as the set of the observed samples ordered according to π_j, $j = 1, 2, \ldots, 6$.

By recalling $Z_2 = X_{2:3} - X_{1:3}$, the estimations of $M(1)$, $M(2)$ and $M(3)$ are obtained as

$$\begin{aligned}
\hat{M}(1) &= \frac{|\mathcal{P}_1 \cup \mathcal{P}_2|}{\sum_{i \in \mathcal{P}_1 \cup \mathcal{P}_2} Z_2^{(i)}} = 2.48951, \\
\hat{M}(2) &= \frac{|\mathcal{P}_3 \cup \mathcal{P}_4|}{\sum_{i \in \mathcal{P}_3 \cup \mathcal{P}_4} Z_2^{(i)}} = 3.34077, \\
\hat{M}(3) &= \frac{|\mathcal{P}_5 \cup \mathcal{P}_6|}{\sum_{i \in \mathcal{P}_5 \cup \mathcal{P}_6} Z_2^{(i)}} = 4.10161.
\end{aligned}$$

Finally, about the parameters $M(h,k)$, $h,k = 1,2,3$, $h \neq k$, by using $Z_3 = X_{3:3} - X_{2:3}$, we have

$$\begin{aligned}
\hat{M}(1,2) &= \frac{|\mathcal{P}_1|}{\sum_{i \in \mathcal{P}_1} Z_3^{(i)}} = 2.67262, & \hat{M}(1,3) &= \frac{|\mathcal{P}_2|}{\sum_{i \in \mathcal{P}_2} Z_3^{(i)}} = 2.11041, \\
\hat{M}(2,1) &= \frac{|\mathcal{P}_3|}{\sum_{i \in \mathcal{P}_3} Z_3^{(i)}} = 0.96048, & \hat{M}(2,3) &= \frac{|\mathcal{P}_4|}{\sum_{i \in \mathcal{P}_4} Z_3^{(i)}} = 3.89834, \\
\hat{M}(3,1) &= \frac{|\mathcal{P}_5|}{\sum_{i \in \mathcal{P}_5} Z_3^{(i)}} = 0.91519, & \hat{M}(3,2) &= \frac{|\mathcal{P}_6|}{\sum_{i \in \mathcal{P}_6} Z_3^{(i)}} = 0.82732.
\end{aligned}$$

Suppose we know everything about the first and second order statistics, i.e., the value and the corresponding component, and we want to predict the maximum order statistic. Then, we can predict the interarrival time by using the quantile regression with the estimated parameters $M(h,k)$. We repeat this procedure for the 300 samples and the results are presented in Figure 6.12 where they are grouped by the different permutations. In order to compare with the predictions based on the fully knowledge of the model, in the figures it is also plotted the theoretical median regression line, whereas the theoretical confidence bands are omitted for the readability of the plots. Moreover, since the parameters have been estimated, here the theoretical coverage percentage of the confidence bands is not exactly 50% or 90% and we refer

Table 6.4: Percentage of exact data in $\hat{C}_{90}^{\pi_j}$ and $\hat{C}_{50}^{\pi_j}$, $j \in \{1, \dots, 6\}$, and weighted average in Example 6.6.

Real coverage in π_j	$\hat{C}_{90}^{\pi_j}$	$\hat{C}_{50}^{\pi_j}$
$j = 1$	88.23%	64.70%
$j = 2$	97.14%	51.43%
$j = 3$	89.77%	54.54%
$j = 4$	97.06%	32.35%
$j = 5$	85.71%	45.71%
$j = 6$	90.11%	47.25%
Weighted average	91.00%	49.00%

to them as $\hat{C}_{50}^{\pi_j}$ and $\hat{C}_{90}^{\pi_j}$, $j \in \{1, \dots, 6\}$. The percentage of exact data in these regions are listed in Table 6.4.

In the last example, we consider a coherent system whose components are distributed according to an ODTHLS model and, by using the observed hystory, we obtain predictions for the lifetime of the system.

Example 6.7. Let us consider a coherent system formed by four components X_1, X_2, X_3, X_4 and whose lifetime T is described as

$$T = \min\{\max\{X_1, X_2\}, \max\{X_3, X_4\}\},$$

with structure displayed in Figure 6.13. Suppose (X_1, X_2, X_3, X_4) is distributed according to an ODTHLS model and assume that $X_{1:4} = X_1 = t_1$. We want to predict the lifetime of the system. The parameters of the model are (we give just the ones interesting for our purposes)

$$\begin{aligned}
&\mu_1(\emptyset) = 4, \quad \mu_2(\emptyset) = 1, \quad \mu_3(\emptyset) = 1, \quad \mu_4(\emptyset) = 2, \\
&\mu_2(1) = 1, \quad \mu_2(1,3) = 2, \quad \mu_2(1,4) = 2, \quad \mu_2(1,3,4) = 2, \quad \mu_2(1,4,3) = 3, \\
&\mu_3(1) = 3, \quad \mu_3(1,2) = 3, \quad \mu_3(1,4) = 3, \quad \mu_3(1,2,4) = 1, \quad \mu_3(1,4,2) = 2, \\
&\mu_4(1) = 2, \quad \mu_4(1,2) = 3, \quad \mu_4(1,3) = 1, \quad \mu_4(1,2,3) = 3, \quad \mu_4(1,3,2) = 2.
\end{aligned}$$

By knowing the first failure and the structure of the system, we deduce that T will be equal to the second order statistic if it is assumed by X_2 whereas it will be the third order statistic if the second failure is assumed by X_3 or X_4. Hence, we obtain a prediction for the lifetime of the system by using the predictions of the second and third order statistics appropriately weighted. More precisely, from Proposition 6.11, the weight for the prediction of the second order statistic will be

$$\mathbb{P}(X_{2:4} = X_2 | X_{1:4} = X_1) = \rho_2(1) = \frac{\mu_2(1)}{\mu_2(1) + \mu_3(1) + \mu_4(1)} = \frac{1}{6}.$$

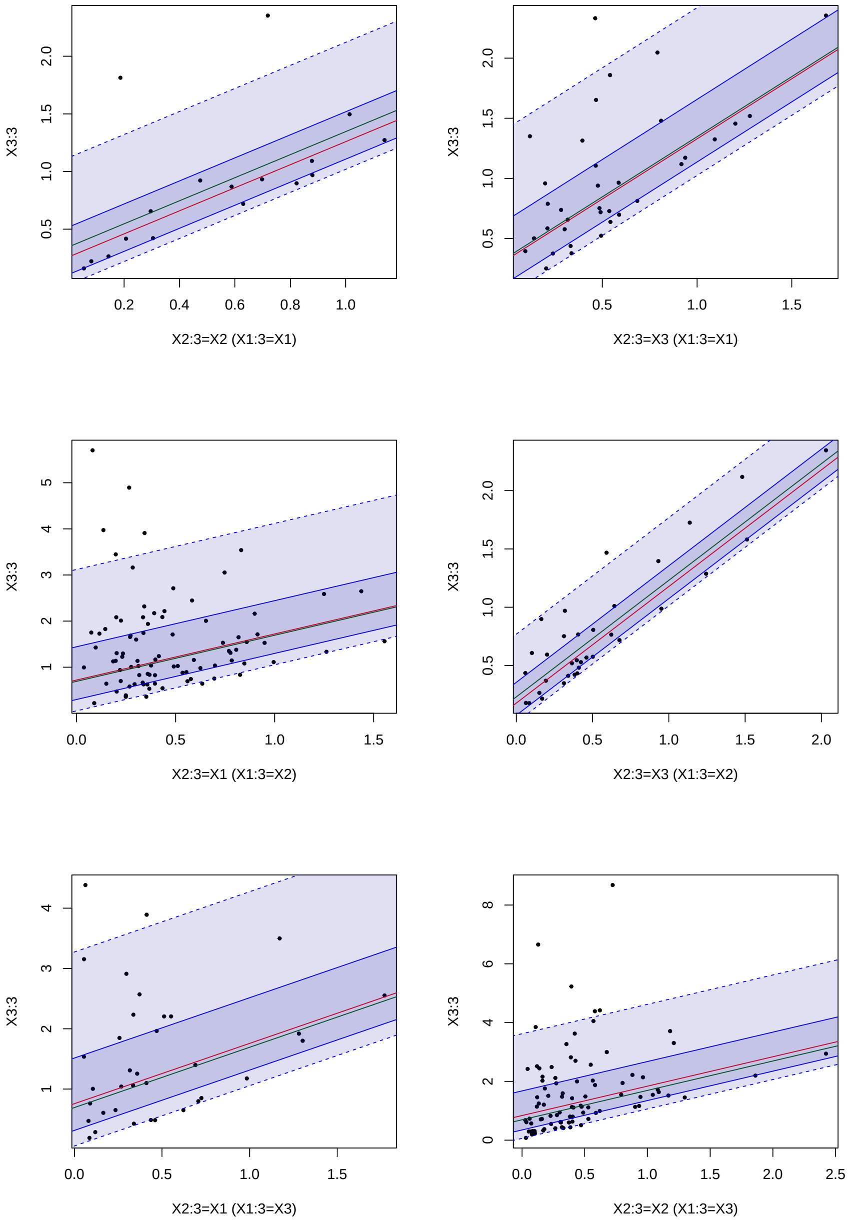

Figure 6.12: Scatterplots of a simulated sample from $(X_{2:3}, X_{3:3})$ for the ODTHLS model in Example 6.6 jointly with the median regression curves (red) and 50% (dark grey) and 90% (light grey) confidence bands obtained by estimating the parameters and with the theoretical median regression curve (green).

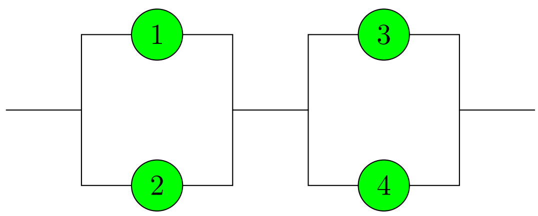

Figure 6.13: The structure of the system in Example 6.7.

About the third order statistic, we have to consider two different predictions, one for the case $X_{2:4} = X_3$ and one for $X_{2:4} = X_4$. The corresponding weights are

$$\mathbb{P}(X_{2:4} = X_3 | X_{1:4} = X_1) = \rho_3(1) = \frac{1}{2},$$
$$\mathbb{P}(X_{2:4} = X_4 | X_{1:4} = X_1) = \rho_4(1) = \frac{1}{3}.$$

Hence, the prediction for the lifetime of the system is obtained as

$$\hat{T} = \rho_2(1) \cdot \hat{X}_{2:4} + \rho_3(1) \cdot \hat{X}^{(3)}_{3:4} + \rho_4(1) \cdot \hat{X}^{(4)}_{3:4},$$

where $\hat{X}^{(j)}_{3:4}$, $j = 3, 4$, denotes the prediction of the third order statistic given $(X_{1:4} = X_1 = t_1, X_{2:4} = X_j)$.

Consider the following (simulated) realization of the sample

$$X_{1:4} = X_1 = 0.10728,\ X_{2:4} = X_3 = 0.17977,\ X_{3:4} = X_2 = T = 0.35048,\ X_{4:4} = X_4 = 0.99044.$$

Suppose we know only $X_{1:4} = X_1 = 0.10728$, hence, by proceeding as described above we obtain

$$\begin{aligned}
\hat{X}_{2:4} &= t_1 + \frac{\log 2}{M(1)} = 0.22281, && M(1) = \mu_2(1) + \mu_3(1) + \mu_4(1) = 6, \\
\hat{X}^{(3)}_{3:4} &= \hat{X}_{2:4} + \frac{\log 2}{M(1,3)} = 0.45386, && M(1,3) = \mu_2(1,3) + \mu_4(1,3) = 3, \\
\hat{X}^{(4)}_{3:4} &= \hat{X}_{2:4} + \frac{\log 2}{M(1,4)} = 0.36144, && M(1,4) = \mu_2(1,4) + \mu_3(1,4) = 5,
\end{aligned}$$

from which it follows the prediction for $T = 0.35048$ as

$$\hat{T} = \frac{1}{6} \cdot 0.22281 + \frac{1}{2} \cdot 0.45386 + \frac{1}{3} \cdot 0.36144 = 0.38454.$$

If the system does not fail at $X_{2:4}$, i.e. the second order statistic is assumed by X_3 or X_4, and we just know that $t_2 = X_{2:4} = 0.17977$, then the prediction for the lifetime of the system will be

$$\hat{T} = t_2 + \frac{3}{5} \cdot \frac{\log 2}{M(1,3)} + \frac{2}{5} \cdot \frac{\log 2}{M(1,4)} = 0.37385,$$

or, by using the median of the mixture of two exponential distributions with parameters $M(1,3)$ and $M(1,4)$ and weights 0.6 and 0.4, respectively, $\hat{T} = 0.36645$. If we also know that $X_{2:4} = X_3$, then the prediction will be

$$\hat{T} = t_2 + \frac{\log 2}{M(1,3)} = 0.41082.$$

In both cases we can obtain confidence bands for the predictions. In the first one, we have a mixture of two exponential distributions and in the second one an exponential distribution with parameter 3. The confidence bands in the first case, for $\hat{T} = 0.36645$, are $C_{90} = [0.19329, 1.04527]$ and $C_{50} = [0.25621, 0.56174]$, and in the second case, for $\hat{T} = 0.41082$, we have $C_{90} = [0.19687, 1.17835]$ and $C_{50} = [0.27566, 0.64187]$.

Bibliography

[1] Akkouchi, M. (2008). On the convolution of exponential distributions. *Journal of the Chungcheong Mathematical Society*, **21**, 501–510.

[2] Arikan, E. (2016). Varentropy decreases under polar transform. *IEEE Transactions on Information Theory*, **62**, 3390–3400.

[3] Arnold, B.C., Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics.* Springer, New York, NY.

[4] Arnold, B.C., Balakrishnan, N., Nagaraja, H.N. (2008). *A First Course in Order Statistics.* SIAM, Philadelphia.

[5] Balakrishnan, N., Buono, F., Longobardi, M. (2022). A unified formulation of entropy and its application. *Physica A: Statistical Mechanics and its Applications*, **596**, 127214.

[6] Balakrishnan, N., Buono, F., Longobardi, M. (2022). On Cumulative Entropies in Terms of Moments of Order Statistics. *Methodology and Computing in Applied Probability*, **24**, 345–359.

[7] Balakrishnan, N., Buono, F., Longobardi, M. (2022). On Tsallis extropy with an application to pattern recognition. *Statistics & Probability Letters*, **180**, 109241.

[8] Balakrishnan, N., Buono, F., Longobardi, M. (2022). On weighted extropies. *Communications in Statistics - Theory and Methods*, **51**, 6250–6267.

[9] Barakat, H.M., El-Adll, M.E., Aly, A.E. (2011). Exact prediction intervals for future exponential lifetime based on random generalized order statistics. *Computers & Mathematics with Applications*, **61**, 1366–1378.

[10] Barakat, H.M., Khaled, O.M., Ghonem, H.A. (2022). Predicting future lifetime for mixture exponential distribution. *Communications in Statistics - Simulation and Computation*, **51**, 3533–3552.

[11] Barlow, R.E., Proschan, F.J. (1996). *Mathematical Theory of Reliability.* Philadelphia: Society for Industrial and Applied Mathematics.

[12] Barlow, R.E., Zwet, W.R. (1969). Asymptotic Properties of Isotonic Estimators for the Generalized Failure Rate Function. Part I: Strong Consistency. *Berkeley: University of California*, **31**, 159–176.

[13] Barlow, R.E., Zwet, W.R. (1969). Asymptotic Properties of Isotonic Estimators for the Generalized Failure Rate Function. Part II: Asymptotic Distributions. *Berkeley: University of California*, **34**, 69–110.

[14] Bartle, R.G., Sherbert, D.R. (2000). *Introduction to Real Analysis,* Third Edition. John Wiley & Sons, New York, NY.

[15] Basiri, E., Ahmadi, J., Raqab, M.Z. (2016). Comparison among non-parametric prediction intervals of order statistics. *Communications in Statistics - Theory and Methods*, **45**, 2699–2713.

[16] Bdair, O.M., Raqab, M.Z. (2022). Prediction of future censored lifetimes from mixture exponential distribution. *Metrika*, **85**, 833–857.

[17] Betensky, R.A., Martin, E.C. (2003). Commentary: Failure-rate functions for doubly truncated random variables. *IEEE Transactions on Reliability*, **52**(1), 7–8.

[18] Bhattacharjee, S., Nanda, A.K., Misra, S.K. (2013). Reliability analysis using ageing intensity function. *Statistics & Probability Letters*, **83**, 1364–1371.

[19] Block, H.W., Savits, T.H., Singh, H. (1998). The reversed hazard rate function. *Probability in the Engineering and Informational Sciences*, **12**, 69–90.

[20] Bobkov, S., Madiman, M. (2011). Concentration of the information in data with log-concave distributions. *Annals of Probability*, **39**, 1528–1543.

[21] Bowman, A.W., Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford University Press Inc., New York.

[22] Buono, F. (2022). Multivariate conditional aging intensity functions and load-sharing models. *Hacettepe Journal of Mathematics and Statistics*, **51**, 1710–1722.

[23] Buono, F., De Santis, E., Longobardi, M., Spizzichino, F. (2022). Multivariate reversed hazard rates and inactivity times of systems. *Methodology and Computing in Applied Probability*, **24**, 1987–2008.

[24] Buono, F., Kamari, O., Longobardi, M. (2021). Interval extropy and weighted interval extropy. *Ricerche di matematica*, DOI: 10.1007/s11587-021-00678-x.

[25] Buono, F., Longobardi, M. (2020). A dual measure of uncertainty: The Deng Extropy. *Entropy*, **22**, 582.

[26] Buono, F., Longobardi, M., Pellerey, F. (2022). Varentropy of past lifetimes. *Mathematical Methods of Statistics*, **31**, 57–73.

[27] Buono, F., Longobardi, M., Szymkowiak, M. (2022). On Generalized Reversed Aging Intensity Functions. *Ricerche di matematica*, **71**, 85–108.

[28] Buono, F., Navarro, J. (2022). Simulations and predictions of future values in the time-homogeneous load-sharing model. Submitted.

[29] Cacoullos, T., Papathanasiou, V., (1985). On upper bounds for the variance of functions of random variables. *Statistics & Probability Letters*, **3**, 175–184.

[30] Cacoullos, T., Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statistics & Probability Letters*, **7**(5), 351–356.

[31] Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 187–220.

[32] Cramer, E. (2021). Ordered and censored lifetime data in reliability: An illustrative review. *WIREs Computational Statistics*, e1571. DOI: 10.1002/wics.1571.

[33] De Santis, E., Malinovsky, Y., Spizzichino, F.L. (2021). Stochastic precedence and minima among dependent variables. *Methodology and Computing in Applied Probability*, **23**, 187–205.

[34] Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics* **38**, 325–339.

[35] Deng, Y. (2016). Deng entropy. *Chaos Solitons Fractals* **91**, 549–553.

[36] Deng, Y. (2020). Uncertainty measure in evidence theory. *Science China Information Sciences*, **63**, 210201.

[37] Di Crescenzo, A., Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied Probability*, **39**, 434–440.

[38] Di Crescenzo, A., Longobardi, M. (2006). On weighted residual and past entropies. *Scientiae Mathematicae Japonicae*, **64**(2), 255–266.

[39] Di Crescenzo, A., Longobardi, M. (2009). On cumulative entropies. *Journal of Statistical Planning and Inference*, **139**, 4072–4087.

[40] Di Crescenzo, A., Paolillo, L. (2021). Analysis and applications of the residual varentropy of random lifetimes. *Probability in the Engineering and Informational Sciences*, **35**(3), 680–698.

[41] Dua, D., Graff, C. (2019). UCI Machine Learning Repository. Available online: http://archive.ics.uci.edu/ml.

[42] Durante, F., Sempi, C. (2016). *Principles of Copula Theory.* CRC/Chapman & Hall, London.

[43] Ebrahimi, N. (1996). How to measure uncertainty in the residual life time distribution. *Sankhyā: Series A*, **58**, 48–56.

[44] El-Adll, M.E. (2011). Predicting future lifetime based on random number of three parameters Weibull distribution. *Computers & Mathematics with Applications*, **81**, 1842–1854.

[45] Finkelstein, M.S. (2002). On the reversed hazard rate. *Reliability Engineering and System Safety*, **78**, 71–75.

[46] Foschi, R., Nappo, G., Spizzichino, F.L. (2021). Diagonal sections of copulas, multivariate conditional hazard rates and distributions of order statistics for minimally stable lifetimes. *Dependence Modeling*, **9**, 394–423.

[47] Fradelizi, M., Madiman, M., Wang, L. (2016). Optimal Concentration of Information Content for Log-Concave Densities. *In: Houdré C., Mason D., Reynaud-Bouret P., Rosinski J. (eds) High Dimensional Probability VII. Progress in Probability*, **71**, 45–60.

[48] Gautschi, W., Gahill, W.F. (1972). Exponential Integral and Related Functions. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. (M. Abramowitz and I. A. Stegun, eds.). New York: Dover.

[49] Giri, R.L., Nanda, A.K., Dasgupta, M., Misra, S.K., Bhattacharjee, S. (2021). On ageing intensity function of some Weibull models. *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2021.1910845.

[50] Goodarzi, F., Amini, M., Mohtashami Borzadaran, G.R. (2016). On upper bounds for the variance of functions of the inactivity time. *Statistics & Probability Letters*, **117**, 62–71.

[51] Gumbel, E.J. (1954). The maxima of the mean largest value and of the range. *The Annals of Mathematical Statistics*, **25**, 76–84.

[52] Gupta, R.C., Gupta, P.L., Gupta, R.D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics - Theory and Methods*, **27**(4), 887–904.

[53] Gupta, R.D., Kundu, D. (2001). Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, **43**, 117–130.

[54] Hartley, H.O., David, H.A. (1954). Universal bounds for mean range and extreme observation. *The Annals of Mathematical Statistics*, **25**, 85–99.

[55] Hoyland, A., Rausand, M. (2009). *System reliability theory: models and statistical methods*. Vol. 420. John Wiley & Sons.

[56] Jiang, R., Ji, P., Xiao, X. (2003). Aging property of unimodal failure rate models. *Reliability Engineering & System Safety*, **79**, 113–116.

[57] Johnson, N.L., Kotz, S. (1975). A vector multivariate hazard rate. *Journal of Multivariate Analysis*, **5**, 53–66.

[58] Kamari, O., Buono, F. (2021). On extropy of past lifetime distribution. *Ricerche di Matematica*, **70**, 505–515.

[59] Kandil, A. M., Kayid, M., Mahdy, M. (2011). Variance Inactivity Time Function and its Reliability Properties. *The 45th Annual Conference on Statistics, Computer Science and operations Research ISRR*, Cairo-Egypt, 94–113.

[60] Kang, B., Deng, Y. (2019). The Maximum Deng Entropy. *IEEE Access* **7**, 120758–120765.

[61] Kang, B.Y., Li, Y., Deng, Y., Zhang, Y.J., Deng, X.Y. (2012) Determination of basic probability assignment based on interval numbers and its application. *Dianzi Xuebao (Acta Electronica Sinica),* **40**, 1092–1096.

[62] Kazemi, M.R., Tahmasebi, S., Buono, F., Longobardi, M. (2021). Fractional Deng entropy and extropy and some applications. *Entropy*, **23**, 623.

[63] Khorashadizadeh, M., Rezaei Roknabadi, A.H., Mohtashami Borzadaran, G.R. (2012). Characterizations of lifetime distributions based on doubly truncated mean residual life and mean past to failure. *Communications in Statistics - Theory and Methods*, **41**(6), 1105–1115.

[64] Koenker R. (2005). *Quantile Regression*. Cambridge University Press.

[65] Krishna, H., Kumar, K (2013). Reliability estimation in generalized inverted exponential distribution with progressively type II censored sample. *Journal of Statistical Computation and Simulation*, **83** (6), 1007–1019.

[66] Kundu, C., Nanda, A., Maiti, S. (2010). Some distributional results through past entropy. *Journal of Statistical Planning and Inference* **140**, 1280–1291.

[67] Lad, F., Sanfilippo, G., Agrò, G. (2015). Extropy: complementary dual of entropy. *Statistical Science*, **30**, 40–58.

[68] Lai, C.D., Xie, M., Murthy, D.N.P. (2001). Bathtub-shaped failure rate life distributions. *Handbook of Statistics*, Elsevier, **20**, 69–104.

[69] Li, X., Li, Z. (2008). A Mixture Model of Proportional Reversed Hazard Rate. *Communications in Statistics - Theory and Methods*, **37**(18), 2953–2963.

[70] Lin, C.T., Duran, B.S., Lewis, T.O. (1989). Inverted gamma as life distribution. *Microelectronics Reliability.* **29**(4), 619–626.

[71] Liu, F., Gao, X., Deng, Y. (2019). Generalized belief entropy and its application in identifying conflict evidence. *IEEE Access*, **7**, 126625–126633.

[72] Madiman, M., Wang, L. (2014). An optimal varentropy bound for log-concave distributions. *International Conference on Signal Processing and Communications* (SPCOM), Bangalore, DOI: 10.1109/SPCOM.2014.6983953.

[73] Misagh, F., Yari, G.H. (2011). On weighted interval entropy. *Statistics and Probability Letters*, **81**(2), 188–194.

[74] Misagh, F., Yari, G.H. (2012). Interval entropy and informative distance. *Entropy*, **14**(3), 480-490.

[75] Morris, R. (1979). The Dilogarithm Function of a Real Argument. *Mathematics of Computation*, **33**, 778–787.

[76] Murty, A.S.R., Naikan, V.N.A. (1996). Reliability strength design through inverse distributions–exponential and Weibull cases. *Reliability Engineering & System Safety*, **54**, 77–82.

[77] Murthy, D.N.P., Xie, M., Jiang, R. (2004). *Weibull Models.* Hoboken: Wiley-Interscience.

[78] Nadarajah, S., Kotz, S. (2006). Reliability for some bivariate exponential distributions. *Mathematical Problems in Engineering*, DOI: 10.1155/MPE/2006/41652.

[79] Nanda, A.K., Bhattacharjee, S., Alam, S.S. (2007). Properties of aging intensity function. *Statistics & Probability Letters* **77**, 365–373.

[80] Navarro, J., Ruiz, J.M. (1996). Failure-rate function for doubly-truncated random variables. *IEEE Transactions on Reliability*, **4**, 685–690.

[81] Navarro, J. (2020). Bivariate box plots based on quantile regression curves. *Dependence Modeling*, **8**, 132–156.

[82] Navarro, J. (2022). *Introduction to System Reliability Theory.* Springer, Cham, DOI: 10.1007/978-3-030-86953-3.

[83] Navarro, J., Buono, F. (2022). Predicting future failure times by using quantile regression. *Metrika*, DOI: 10.1007/s00184-022-00884-z.

[84] Navarro, J., Calì, C., Longobardi, M., Durante, F. (2022). Distortion representations of multivariate distributions. *Statistical Methods & Applications*, **31**, 925–954.

[85] Navarro, J., Rubio, R. (2010). Computations of signatures of coherent systems with five components. *Communications in Statistics - Simulation and Computation*, **39**, 68–84.

[86] Navarro, J., Ruiz, J.M., Del Aguila, Y. (2008) Characterizations and ordering properties based on log-odds functions. *Statistics: A Journal of Theoretical and Applied Statistics*, **42**(4), 313–328.

[87] Navarro, J., Sarabia, J. (2022). Copula representations for the sum of dependent risks: models and comparisons. *Probability in the Engineering and Informational Sciences*, **36**, 320–340.

[88] Nelsen, R.B. (2006). *An Introduction to Copulas.* Springer, New York.

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

[90] Oguntunde, P., Adejumo, A.O., Owoloko, A. (2017). Application of Kumaraswamy Inverse Exponential Distribution to Real Lifetime Data. *International Journal of Applied Mathematics and Statistics*, **56**, 34–47.

[91] Paolillo, L., Di Crescenzo, A., Suárez-Llorens, A. (2021). Stochastic Comparisons, Differential Entropy and Varentropy for Distributions Induced by Probability Density Functions. *arXiv:2103.11038v1.*

[92] Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, **3**, 119–131.

[93] Qiu, G. (2017). The extropy of order statistics and record values. *Statistics & Probability Letters*, **120**, 52–60.

[94] Qiu, G., Jia, K. (2018). The residual extropy of order statistics. *Statistics & Probability Letters*, **133**, 15–22.

[95] Rao, M., Chen, Y., Vemuri, B., Wang, F. (2004). Cumulative residual entropy: A new measure of information. *IEEE Transactions on Information Theory*, **50**, 1220–1228.

[96] Rezaei, M., Khalef, V.A. (2014). On the Reversed Average Intensity Order. *Journal of Statistical Research of Iran,* **11**, 25–39.

[97] Ross, S.M. (1984). A model in which component failure rates depend on the working set. *Naval research logistics quarterly*, **31**, 297–300.

[98] Ross, S. (2007). *Introduction to probability models*, 9th ed. New York: Academic Press.

[99] Rychlik, T. (2001). *Projecting Statistical Functionals.* Springer, New York, NY.

[100] Samaniego, F.J. (2007). *System Signatures and Their Applications in Engineering Reliability.* New York, NY: Springer Science + Business Media. International Series in Operations Research and Management Science; Vol. 110.

[101] Sathar, A.E.I., Nair, D.R. (2021). On dynamic survival extropy. *Communications in Statistics - Theory and Methods*, **50**, 1295–1313.

[102] Schechner, Z. (1984). A load-sharing model: the linear breakdown rule. *Naval research logistics quarterly*, **31**, 137–144.

[103] Sekeh, S.Y., Mohtashami Borzadaran, G.R., Rezaei Roknabadi, A.H. (2014). Some results based on a version of the generalized dynamic entropies. *Communications in Statistics - Theory and Methods*, **43**(14), 2989–3006.

[104] Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press: Princeton, NJ, USA.

[105] Shaked, M., Shanthikumar, J.G. (1988). Multivariate conditional hazard rates and the MIFRA and MIFR properties. *Journal of Applied Probability*, **25**, 150–168.

[106] Shaked, M., Shanthikumar, J.G. (2007). *Stochastic Orders*, Springer: New York, NY, USA.

[107] Shaked, M., Shanthikumar, J.G. (2015). Multivariate conditional hazard rate functions - an overview. *Applied Stochastic Models in Business and Industry*, **31**, 285–296.

[108] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423.

[109] Smets, P. (2000). Data fusion in the transferable belief model. In Proceedings of the Third International Conference on Information Fusion, Paris, France, 10–13 July 2000, **1**, PS21–PS33.

[110] Spizzichino, F.L. (2018). Reliability, signature, and relative quality functions of systems under time-homogeneous load-sharing models. *Applied Stochastic Models in Business and Industry*, **35**, 158–176.

[111] Sunoj, S.M., Sankaran, P.G., Maya, S.S. (2009). Characterizations of Life Distributions Using Conditional Expectations of Doubly (Interval) Truncated Random Variables. *Communications in Statistics - Theory and Methods*, **38**(9), 1441–1452.

[112] Sydsaeter, K., Hammond, P. (2012). *Essential Mathematics for Economic Analysis.* London: Pearson Education.

[113] Szymkowiak, M. (2018). Characterizations of Distributions through Aging Intensity. *IEEE Transactions on Reliability*, **67**, 446–458.

[114] Szymkowiak, M. (2018). Generalized Aging Intensity Functions. *Reliability Engineering & System Safety* **178**, 198–208.

[115] Szymkowiak, M. (2020). *Lifetime Analysis by Aging Intensity Functions.* Springer, DOI: 10.1007/978-3-030-12107-5.

[116] Taylor, A.E. (2010). *General Theory of Functions and Integration.* Dover Publications, Mineola, NY.

[117] Tran, L., Duckstein, L. (2002). Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems*, **130**, 331–341.

[118] Tsallis, C. (1988) Possible generalization of Boltzmann-Gibbs statistic. *Journal of Statistical Physics*, **52**, 479–487.

[119] Ubriaco, M.R. (2009). Entropies based on fractional calculus. *Physics Letters A*, **373**, 2516–2519.