Università degli Studi di Salerno

DIPARTIMENTO DI FISICA "E.R. CAIANIELLO"
Corso di Dottorato in Matematica, Fisica e Applicazioni
XXXIII Ciclo

TESI DI DOTTORATO IN

Properties and applications of pdf-related information measures and distributions

Settore Scientifico-Disciplinare MAT/06 Probabilità e Statistica Matematica

Dottorando:
Luca Paolillo

Coordinatore del Dottorato:
Prof. Carmine Attanasio

Tutor:
Prof. Antonio Di Crescenzo

Anno Accademico 2019-2020

Abstract

The study of the information measures gives rise to different measures according to the contests in which it is applied. In the contest of reliability theory and survival analysis, an ever-growing interest is given by the entropy applied to continuous random variables.This quantity gives the expectation of the information content and is known as differential entropy. Another quantity, the differential varentropy is the variance of the information content. Differential entropy and differential varentropy are mainly applied to the study of brand-new item.

Other measures of interests in reliability contests are the dynamical measures of information. In this thesis a particular attention is devoted to residual entropy and residual varentropy, that are the expectation and the variance of the information content of a residual lifetime distribution. They can be very useful in the cases in which the item has a finite age. In particular, the residual varentropy is a largely unexplored subject and a focus on this quantity constitute the central part of the thesis.

Stimulated by the need of describing useful notions related to the quantity described above, we introduce the 'pdf-related distributions'. These are defined in terms of transformation of absolutely continuous random variables through their own probability density functions. We investigate their main characteristics, with reference to the general form of the distribution, the quantiles, and some related notions of reliability theory. This allows us to obtain a characterization of the uniform distribution based on pdf-related distributions of exponential, Laplace and power type as well. We also face the problem of stochastic comparing the pdf-related distributions by resorting to suitable stochastic orders. Finally, the given results are used to analyse properties and to compare some useful information measures, such as the differential entropy and the varentropy.

This work of thesis covers different arguments in the contest of information for continuous random variables. Firstly, mathematical properties of residual varentropy are discussed, such as conditions for which it is constant or monotonic and the determination of the upper and the lower bound. Another theoretical aspect that will be discussed concerns the properties of entropy and varentropy for stochastically ordered distributions. In addition, some applications of residual varentropy are proposed. The first, the proportional hazards model gives an example of application of the varentropy in the context of reliability theory and survival analysis. The second is an application to stochastic process. More specifically, residual varentropy is applied to the first passage-time of an Ornstein–Uhlenbeck jump-diffusion process. A kernel-type estimation of the residual varentropy is finally proposed, as a further application to real data. The last part of the thesis concerns the "covarentropy", that is a new measure introduced in order to study the correlations between the information contents of two random variables.

Contents

Introduction

The information theory has fundamental contribution in communication theory, but also in different areas as statistical physics, computer science, statistical inference. The pioneer was Shannon [68] who introduced the entropy, also called Shannon entropy, that was already used in thermodynamics. Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If X is a discrete variable defined on a support $S_X = \{x_i; i \in I\}$ and $IC(X)$ is the information content is defined as the random variable

$$IC(X) = -\log p(X) \tag{1}$$

In the (1) the quantity $p(x) = \mathbb{P}(X = x)$ is the discrete density function of X calculated at point x. In the information theory the information content is the number of bits needed to represent X by a coding scheme that minimizes the average code lengths (see [68]). The Shannon entropy is defined as

$$H(X) = \mathbb{E}[IC(X)] = -\sum_{i \in I} p(x_i) \log p(x_i) \tag{2}$$

where "log" is the natural logarithm, we set $0 \log 0 = 0$ by convention and we denote with $\mathbb{E}[\cdot]$ the expectation. In information theory, entropy is the minimum descriptive complexity of a random variable and can be introduced together with its related quantities as for example relative entropy and mutual information (for a comprehensive discussion see Cover and Thomas [17]) in order to give an extensive interpretation to problems in data information and communication.

In mathematics, the fundamental quantities of information theory are defined as functionals of probability distributions. In turn, they characterize the behaviour of long sequences of random variables and allow us to estimate the probability of rare events (large deviation theory) and to find the best error exponent in hypothesis test.

Kolmogorov, Chaitin and Solomonoff put forth the idea that the complexity of a string of data can be defined by the length of the shortest binary program for computing the string. Thus the complexity is the minimal description length. Kolmogorov complexity lays the foundation for the theory of the descriptive complexity and is approximately equal to the Shannon entropy if the sequence is drawn at random from a distribution that has the given entropy.

In addition to the computational complexity, we need also to know the distribution about where the information arises. Supplementary measures for distributional spread and asymmetry can be derived in terms of effective concentration of entropic mass, such the given distribution is formally equivalent to bipolar outcomes, for example, as 'good' or 'bad'. This kind of measure is object of investigation when the partition theory is examined (for a comprehensive discussion see Bowden [14]) and allows to compare the different distributions from an information point of view. An alternative way to carry out this analysis consists in studying the dispersion around the mean value of the information content. This measure is called "varentropy" and it is able to highlight some characteristic of the distributions, as we will show in the course of this thesis.

The comparison of the information content of two random variables can also be done in a bivariate distribution contest. For such a measure we can introduce a new kind of measure that gives the covariance of the information content of two random variables and is called "covarentropy". The covarentropy is an information measure that takes into account the presence of correlations between random variables and is the object of Chapter 4.

If X is absolutely continuous with probability density function (pdf) $f(t)$, we can introduce the random variable

$$IC(X) = -\log f(X) \tag{3}$$

The Equation (3) is often referred as the random information content of X. $IC(X)$ is the natural counterpart of the number of bits needed to represent X in the discrete case by a coding scheme that minimizes the average length code (see [68]). The *differential entropy* $H(X)$ of a continuous random variable X with a density $f(x)$ is a very common uncertainty measure and is defined as

$$H(X) := \mathbb{E}[IC(X)] = -\int_{-\infty}^{\infty} f(x) \log f(x) \, \mathrm{d}x, \tag{4}$$

As we see from (4), differential entropy does not depend on the random variable but only on its pdf. Intuitively, $H(X)$ measures the expected uncertainty contained in $f(x)$ about the predictability of an outcome of X. $H(X)$ may or may not exist (in the Lebesgue sense). The differential entropy is also related to the evaluation of the size of the smallest set containing the realizations of typical random samples taken from X (see Chapter 9 of [17]). Moreover, as in the discrete case, the differential entropy depends only on the probability density function of the random variable. When the differential entropy exists, differently from the entropy of discrete random variables, it takes values in the extended real line $[-\infty, \infty]$. Other incongruities have been pointed out in various investigations (see for instance, [19] and [65]). In information theory, large attention is given to the so-called entropy power of a continuous random variable X, which is a positive quantity expressed in terms of $H(X)$. Rather than in stochastic modeling, it is usually adopted to compare the differential entropy of a sum of independent random variables with their differential entropies, and with the entropy of a suitable random variables (see Chapter 16 of [17] and [52] also for its connection to the Fisher information). Hence, the entropy power is useful to analyze stochastic systems governed by unbounded random variables that are comparable to random variables.

The problem of the independence between random variables is very crucial in the study of the information measures. In Probability Theory and Statistics two events are said independent when “the occurrence of one event does not increase neither decrease the probability of the other event occurring” (T. Bayes, 1763). Today, while the notion of “independence” is well-formalized, the complementary notion of “dependence” is much less well understood. The investigation about the “measures” and the dependence “models” was started with the pioneering works of Pearson (at the end of XIX century) about correlation coefficients and their variations and is going on nowadays. These investigations have been developed according to three different approaches:

- measures of association giving a numerical evaluation of the “degree of dependence” between the random variables;
- multivariate (especially Gaussian) models that quantify the relations of dependence in a functional form;
- stochastic orderings that provide a criterium to compare random objects.

Along this lines of research, the project of thesis was oriented to the study of the properties of new indexes appropriate to the measure of the variability in random phenomena and to individuate new instruments in order to compare random variables. The investigation is motivated by the need to obtain theoretical instruments in order to evaluate their dependence. As a starting point, we individuate mathematical properties of the variance of residual lifetimes. The problem of varentropy of residual lifetime or residual varentropy is introduced initially as a dynamical measure of the variability of the information content of the stochastic system that are conditioned to survival. Our attention is devoted to disclose properties of varentropy of residual lifetimes. We give special attention to the conditions such that it is constant. We also discuss the effect of linear transformations, we provide suitable lower and upper bounds. Moreover, we focus on certain applications involving the proportional hazards model and the first-passage times for Ornstein–Uhlenbeck jump-diffusion processes. Finally, as an example of application of the varentropy to real data, we deal with the problem of the kernel estimation of the residual varentropy. Another problem related to the distribution of the entropic mass is addressed making use of effective probabilistic and statistical tools for assessing the above mentioned information measures, introducing the so-called 'pdf-related distributions'. These are constructed by means of transformation of absolutely continuous random variables through their own pdf's.

In order to consider these aspects, in the second part of the thesis we investigate the main properties of the pdf-related distributions, with special reference to (i) the general form of their distribution functions, (ii) the connections with various notions of interest in reliability theory (such as the residual lifetime), (iii) the determination of quantiles. We analyse pdf-related distributions generated by general distributions. However, special attention is devoted to the case when the underlying distributions possess monotone or unimodal pdf's. Here, unimodality refers to strictly decreasing pdf's or to pdf's that are first strictly increasing and then strictly decreasing.

The special form that is involved in their definition, allows to tackle the problem of comparing the pdf-related distributions by resorting to suitable stochastic orders, such as the usual stochastic order, the dispersive order, the convex trasform order, the star order and the kurtosis order. Hence, in this framework the main results involve both location and variability orderings, and are obtained by techniques based on mapping of quantiles, convexity and rearrangement of pdf's.

As main application, the pdf-related distributions are finalized to construct a stochastic framework aimed to compare the basic information measures for absolutely continuous random variables.

The last part of the thesis concerns the covariance between information contents. The investigation is carried out both for discrete and continuous random variable. In order to acquire more insight on this quantity, we focus on the relation between covarentropy and covariance of random variables.

The thesis is organized as follows. We start with a mathematical background (Chapter

1), where we discuss about random lifetime distributions and residual entropy (Section 1.1), kernel density estimation (Section 1.2) and Ehrenfest model (Section 1.3). In Chapter 2 we introduce the varentropy and in particular the varentropy of residual lifetimes. Firstly, we investigate the mathematical properties (Section 2.2), constant residual varentropy conditions and some bounds for the residual varentropy. The applications (Section 2.4) are the proportional hazard model and the reliability analysis of series system, the Ornstein-Uhlenbeck jump-diffusion process arising form the Ehrenfest model subject to catastrophes and a kernel approach for the estimation of the residual varentropy. Chapter 3 gives an analysis of entropy and varentropy as related to stochastic comparisons between pdf-related random variables. As a starting point, a description of quantiles for unimodal distributions (Section 3.2) is given in order to introduce the pdf-related distributions (Section 3.3). Some useful results about the stochastic order are given in Section 3.4. In Section 3.5 the previous notion are applied to information content and in Section 3.6 some conditions are given in order to compare entropy and varentropy. Finally, in Chapter 4 the covarentropy of two random variables is discussed. This chapter includes general theorems that provide conditions in order to give a connection between covarentropy and covariance but also some applications to specific distributions. The analysis is carried out both for discrete (Section 4.2) and for absolutely continuous random variables (Section 4.3).

The investigation concerning this thesis was part of the Innovative Doctorates with Industrial characterization funded by European Social Fund PON 2014/2020 and performed in collaboration with Nexsoft S.p.A. (Salerno) and Departamento de Estadística e Investigación Operativa of Universidad de Cádiz (Spain).

Throughout the thesis we refer to the following papers.

- Di Crescenzo, A., & Paolillo, L. Analysis and applications of the residual varentropy of random lifetimes. *Probab. Engrg. Inform. Sci.* pp. 1–19. doi:

10.1017/S0269964820000133 (2020).

- Di Crescenzo, A., Paolillo, L., & Suárez-Llorens, A. Stochastic comparisons, differential entropy and varentropy for distributions induced by probability density functions. 21 pp. (2021). *arXiv:2103.11038v1*. Submitted for publication.

Chapter 1

Mathematical background

The aim of this chapter is to give some mathematical instruments that are useful to the study of mathematical properties and applications of the residual varentropy. Section 1 is concerned about the random lifetimes, that are useful to a statistical description of the life of an item that can be also conditional to the survival until a given time t (residual lifetime distributions) and the properties of the entropy of residual lifetime distributions, or alternatively, residual entropy. Section 2 is focused on the description of the kernel density estimation. In Section 3 a brief description of the continuous time Eherefest model with catastrophes is given. This model is one of the applications to the residual varentropy that we will analyse in Chapter 2.

1.1 Notions on lifetimes distributions and residual entropy

Let X be an absolutely nonnegative continuous random variable with pdf $f(x)$ and let the interval $S_X \subseteq \mathbb{R}$ be its support. Let $F(x) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$, be the cumulative distribution function (cdf) and $\overline{F}(x) = 1-F(x)$ the complementary distribution function, also known as survival function, then the residual lifetime X_t is given by

$$X_t = [X - t | X > t], \qquad t \in S_X. \tag{1.1}$$

where $S_X = (0, b)$, for $0 < b \leq \infty$. Here, notation $[X|B]$ is adopted for a random variable whose distribution is identical to that of X conditional on B. Clearly, X_t denotes the system lifetime conditioned to the survival of the system at time t. In order to give an interpretation in the contest of reliability theory, let us consider a system (such as an item or a living organism) that starts its activity at time 0 and works regularly up to its failure time. We can introduce a random variable X that describes the random lifetime of such a system. Hence, as a consequence of (1.1), the residual lifetime X_t has the significance of the system lifetime conditioned to the survival of the system at time t. Useful applications of residual lifetime distributions in actuarial science can be found in Sachlas and Papaiannou [64].

If X is a nonnegative random lifetime, with support $S_X = (0, b)$, for $0 < b \leq \infty$, then the random variable X_t given by (1.1) describes the residual lifetime of X at age t. Then, the cdf, survival function and pdf of X_t, for $x \in (0, b - t)$ and $t \in S_X$ are given by:

$$F_t(x) = \mathbb{P}(X_t \leq x) = \frac{F(x+t) - F(t)}{\overline{F}(t)},$$

$$\overline{F}_t(x) = 1 - F_t(x) = \frac{\overline{F}(x+t)}{\overline{F}(t)}, \tag{1.2}$$

$$f_t(x) = \frac{\mathrm{d}F_t(x)}{\mathrm{d}x} = \frac{f(x+t)}{\overline{F}(t)}. \tag{1.3}$$

The conventional approach used to characterize the failure distribution of X is either by its (instantaneous) hazard rate function

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} = \lim_{h \to 0^+} \frac{1}{h} \mathbb{P}[X \leq t + h | X > t], \qquad t \in S_X, \tag{1.4}$$

or by its mean residual lifetime function, defined as

$$m(t) = \mathbb{E}(X_t) = \mathbb{E}[X - t | X > t] = \frac{1}{\overline{F}(t)} \int_t^\infty \overline{F}(x)\,\mathrm{d}x, \qquad t \in S_X. \tag{1.5}$$

We recall also the cumulative hazard rate function of X,

$$\Lambda(t) = -\log \overline{F}(t) = \int_0^t \lambda(x)\,\mathrm{d}x, \qquad t \in S_X, \tag{1.6}$$

which plays a relevant role in numerous contexts.

Moreover, it is known that each of the functions $\overline{F}$, λ and m uniquely determines the other two. More specifically, for $t \in S_X$ we have

$$\overline{F}(t) = \exp\left\{-\int_0^t \lambda(x)\,\mathrm{d}x\right\} = \frac{m(0)}{m(t)} \exp\left\{-\int_0^t \frac{1}{m(x)}\,\mathrm{d}x\right\}, \qquad \lambda(t) = \frac{m'(t)+1}{m(t)}.$$

Another useful quantity is the variance of X_t (called also "variance residual life function") that is

$$\sigma^2(t) := \mathrm{Var}(X_t) = \mathrm{Var}[X - t \mid X > t] = \frac{2}{\overline{F}(t)} \int_t^\infty \mathrm{d}x \int_x^\infty \overline{F}(y)\,\mathrm{d}y - [m(t)]^2, \tag{1.7}$$

with $t \in S_X$ and $m(t)$ defined in (1.5). For instance, see Gupta [36] for characterization results and properties of $\sigma^2(t)$. The vitality function of X is

$$\delta(t) := \mathbb{E}[X | X > t] = m(t) + t, \qquad t \in S_X. \tag{1.8}$$

Namely, since X denotes the random lifetime of a system, $\delta(t)$ can be interpreted as the average life span of a system whose age exceeds t. The generalized hazard rate of X (see Schweizer and Szech [66]) can be expressed by

$$\lambda_\alpha(t) = \frac{f(t)}{[\overline{F}(t)]^{1+\alpha}}, \qquad \alpha \in \mathbb{R}, t \in S_X. \tag{1.9}$$

and reduces to λ when $\alpha = 0$. Clearly, recalling (1.4) one has $\lambda_0(t) = \lambda(t)$ for all t. Other parameterizations of $\lambda_\alpha(t)$ have been treated in Bieniek and Szpak [12] as a special case of the generalized failure rate defined by Barlow and van Zwet [7]. Further forms of generalized hazard rates have been considered in the past. For instance, Lariviere and Porteus [47], and Maoui *et al.* [54] considered $t\,\lambda(t)$ as generalized hazard rate. Moreover, a different version has been treated in Li and Tewari [49].

1.1.1 An application: Cox proportional hazards model

Consider a family of absolutely continuous nonnegative random variables $\{X^{(a)}; a > 0\}$, where the survival function and the pdf of $X^{(a)}$ are expressed, respectively, as

$$\overline{F}^{(a)}(t) = \mathbb{P}[X^{(a)} > t] = [\overline{F}(t)]^a, \quad f^{(a)}(t) = a[\overline{F}(t)]^{a-1} f(t), \qquad t > 0, \tag{1.10}$$

where $\overline{F}(t)$ a suitable baseline survival function and $f(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\overline{F}(t)$ the associated pdf. As a function of (1.10) we can write the hazard function as

$$\lambda^{(a)}(t) = \frac{f^{(a)}(t)}{\overline{F}^{(a)}(t)} = a\,\lambda(t), \qquad t > 0, \tag{1.11}$$

where

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}.$$

and moreover applying Equation 1.6

$$\Lambda^{(a)}(t) = -\log \overline{F}^{(a)}(t) = a\,\Lambda(t), \qquad t > 0. \tag{1.12}$$

As we can see from Equations (1.11) and (1.12) the hazard rate and the cumulative hazard function are proportional to their respective baseline function. Due to this characteristic model is known as the proportional hazards model, see Cox [18].

Let us consider a system of n devices connected in series, whose component i has survival functions $\overline{F}_i(t)$ and pdf $f_i(t)$ for all $t > 0$. Let $\overline{F}_S(t)$ the survival function of a series system, then we have

$$\overline{F}_S(t) = \prod_{i=1}^{n} \overline{F}_i(t), \qquad t > 0 \qquad i = 1, \ldots, n.$$

As a consequence of (1.6) we can easily prove that the hazard rate of the system $\lambda_S(t)$ can be written in terms of the hazard rate of each component $\lambda_i(t) = \frac{f_i(t)}{\overline{F}_i(t)}$ as

$$\lambda_S(t) = \sum_{i=1}^{n} \lambda_i(t), \qquad t > 0.$$

When the components are identical

$$\overline{F}_i(t) \equiv \overline{F}(t), \qquad f_i(t) \equiv f(t), \qquad \lambda_i(t) = \frac{f(t)}{\overline{F}(t)} \equiv \lambda(t), \qquad t > 0 \qquad i = 1, ..., n.$$

where and then

$$\overline{F}_S(t) = [\overline{F}(t)]^n, \qquad t > 0$$

and

$$\lambda_S(t) = n\lambda(t), \qquad t > 0$$

1.1.2 Residual entropy

Now, we assume that X is a nonnegative absolutely continuous random variable that describes the random lifetime of such a system. Hence, $H(X)$ is a suitable measure of uncertainty of the failure time. However, the use of $H(X)$ is adequate for a brand new system, whereas is somewhat unrealistic whenever the initial age of the considered system is non-zero. To this aim, Ebrahimi [27], Ebrahimi and Pellerey [29] and Muliere *et al.* [57] proposed to use the residual lifetime X_t in place of X in the expression (4). Let X

be an absolutely continuous random variable, then the quantity

$$H(X_t) = \mathbb{E}[IC(X_t)] = -\int_t^\infty \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)}\,\mathrm{d}x, \qquad t \in S_X, \tag{1.13}$$

is called entropy of residual lifetime or residual entropy, for short. After the component has survived up to time t, $H(X_t)$ basically measures the expected uncertainty contained in the conditional density of $(X-t)$ given that $X > t$ about the predictability of the remaining lifetime. More information about residual entropy are given in reff. [27, 28, 29, 57].

1.1.3 Alternative forms of the residual entropy

Alternative forms of the residual entropy (1.13) (see Equations 2.2 of Ebrahimi [27]) are given by

$$H(X_t) = -\Lambda(t) - \frac{1}{\overline{F}(t)} \int_t^\infty f(x) \log f(x)\,\mathrm{d}x, \tag{1.14a}$$

$$H(X_t) = 1 - \frac{1}{\overline{F}(t)} \int_t^\infty f(x) \log \lambda(x)\,\mathrm{d}x, \tag{1.14b}$$

where $\lambda(x)$ is the failure rate function of X (1.4) and $\Lambda(t)$ is the cumulative hazard function of X (1.6).

1.1.4 Residual entropy for linear transformations

Let X and Y two random lifetime distributions related by a linear transformation according to the relation

$$Y = aX + b, \qquad a > 0, \quad b \geq 0, \tag{1.15}$$

then the residual entropy of X and Y are related by (see Eq. (2.6) of Ebrahimi and Pellerey [29])

$$H(Y_t) = H\left(X_{\frac{t-b}{a}}\right) + \log a, \qquad \forall\, t. \tag{1.16}$$

1.1.5 Derivative of the residual entropy

The monotonicity of the residual varentropy was object of investigation by different authors (see for example Ebrahimi [27] and Belzunce *et al.* [11]). In order to study the monotonicity of the residual entropy it can be useful to express its derivative with respect to t in terms of other known quantities. Let us denote with g' the derivative of a given function g. As the residual entropy $H(X_t)$ can be viewed as a function of t, say $\Phi(t)$, the derivative $H'(X_t)$ will be the derivative $\Phi'(t)$. The result found for $H'(X_t)$ is the following. For the residual entropy of a random variable X we have (see, e.g. Eq. (2.4) of [27])

$$H'(X_t) = \lambda(t)[H(X_t) - 1 + \log \lambda(t)]. \tag{1.17}$$

where $\lambda(t)$ is the hazard rate of X (see Equation (1.4)). A consequence of increasing residual entropy distributions was provided by Belzunce *et al.*[11]. They proved that if X has an absolutely continuous distribution and an increasing residual entropy $H(X_t)$, then the underlying distribution is uniquely determined.

1.1.6 Weighted version of the residual entropy

A weighted version of the residual entropy (1.13) is given by (see Di Crescenzo and Longobardi [21] for details)

$$\begin{aligned} H^w(X_t) &= -\int_t^\infty x\,\frac{f(x)}{\overline{F}(t)}\log\frac{f(x)}{\overline{F}(t)}\,\mathrm{d}x \\ &= -\frac{1}{\overline{F}(t)}\int_t^\infty x\,f(x)\log f(x)\mathrm{d}x - \frac{\Lambda(t)}{\overline{F}(t)}\int_t^\infty x\,f(x)\mathrm{d}x, \qquad t \in S_X. \end{aligned} \tag{1.18}$$

1.2 Kernel density estimation

Starting from a random sample $X_1, X_2, \ldots, X_n$ from a population with density function f a kernel density estimator is given by the quantity

$$\widehat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad x \in \mathbb{R},$$

or the leave-one-out kernel estimator obtained excluding X_i in the calculation of the density function:

$$\widehat{f}_i(X_i) = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n} \frac{1}{h} K\Big(\frac{X_i - X_j}{h}\Big), \qquad i = 1, 2, \ldots, n. \tag{1.19}$$

where h is a smoothing parameter or bandwidth and K is the kernel function. K is a smooth symmetric kernel function satisfying

$$\int K(x)\ \mathrm{d}x = 1$$

$$\int xK(x)\ \mathrm{d}x = 0$$

$$\int x^2K(x)\ \mathrm{d}x < \infty$$

1. Uniform

$$K(x) = \frac{1}{2}\mathbb{1}_{\{|x|<1\}}$$

2. Triangular

$$K(x) = (1 - |x|)\mathbb{1}_{\{|x|<1\}}$$

3. Gaussian

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$$

4. Parabolic

$$K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{\{|x|<1\}}.$$

Let $X_1, \ldots, X_n$ be a random sample from a population with cdf $F(x)$ and pdf $f(x)$. Two typical estimators $\widehat{R}(t)$ of the survival function $\overline{F}(t)$ are:

(i) Empirical estimator

$$\widehat{R}(t) = \widehat{R}^e(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>t\}}. \tag{1.20}$$

(ii) Kernel estimator (cf. Azzalini [6], for instance)

$$\widehat{R}(t) = \widehat{R}_h(t) := \frac{1}{n}\sum_{i=1}^{n} S\Big(\frac{t - X_i}{h}\Big). \tag{1.21}$$

where $h > 0$ is a smoothing parameter, named bandwidth, and

$$S(t) = \int_t^{\infty} K(x)\,\mathrm{d}x \tag{1.22}$$

is the survival function associated to the kernel function $K(x)$.

1.3 Continuous-time Ehrenfest model with catastrophes

Let us consider a system subject to catastrophes, described by a stationary Markov chain $\{M(t), t \geq 0\}$ defined by a stationary Markov chain $\{M(t),\, t \geq 0\}$ defined on the state-space $S = \{-N, -N+1, \ldots, -1, 0, 1, \ldots, N\}$, with N a positive integer. Let us suppose that the catastrophes occur according to a Poisson process with intensity ξ. Denoting by

$$r(k, n) = \lim_{h\to 0^+} \frac{1}{h}\mathbb{P}[M(t+h) = n|M(t) = k], \qquad k, n \in S$$

the transition rate function of $M(t)$, we assume that the allowed transitions occur according to the following scheme (for a complete discussion see Dharmaraja *et al.* [20]):

$$r(-1,0) = \lambda(N+1) + \xi, \qquad r(n,n+1) = \lambda(N-n), \qquad \forall n \in S \setminus \{-1\}; \tag{1.23}$$

$$r(1,0) = \mu(N+1) + \xi, \qquad r(n,n-1) = \mu(N+n), \qquad \forall n \in S \setminus \{-1\}; \tag{1.24}$$

$$r(n,0) = \xi, \qquad n \in S \setminus \{0\}. \tag{1.25}$$

with $\lambda, \mu, \xi > 0$. Hence, $\{M(t), t \geq 0\}$ is a time-homogeneous continuous-time Markov process with transition rates (1.23), (1.24) and (1.25), and is defined on the state-space S, where S is an irreducible class. Equation (1.25) defines the catastrophes rate, the effect of each catastrophe being the instantaneous transition to state 0.

Due to (1.23), (1.24) and (1.25), for all $j, n \in S$ and $t \geq 0$ the transition probabilities

$$p_{j,n}(t) = \mathbb{P}[M(t) = n | M(0) = j] \qquad (j, n \in S)$$

satisfies the following differential-difference equations:

$$\begin{cases} \frac{d}{dt} p_{j,0}(t) = -[N(\lambda+\mu)+\xi] p_{j,0}(t) + (N+1)\lambda p_{j,-1}(t) + (N+1)\mu p_{j,1}(t) + \xi, \\ \frac{d}{dt} p_{j,n}(t) = -[(N-n)\lambda + (N+n)\mu + \xi] \\ \qquad\qquad p_{j,n}(t) + (N+n+1)\mu\, p_{j,n+1}(t), \qquad n = \pm 1, \pm 2, \ldots, \pm(N-1) \\ \frac{d}{dt} p_{j,-N}(t) = -(2N\lambda + \xi) p_{j,-N}(t) + \mu\, p_{j,-N+1}(t), \\ \frac{d}{dt} p_{j,N}(t) = -(2N\mu + \xi) p_{j,N}(t) + \lambda\, p_{j,N-1}(t). \end{cases} \tag{1.26}$$

The initial condition for system (1.26) is expressed in terms of the Kroenecker's delta:

$$p_{j,n}(0) = \delta_{j,n} = \begin{cases} 1, & n = j \\ 0, & \text{otherwise.} \end{cases} \tag{1.27}$$

With a typical scaling in oder to obtain a diffusion approximation of the discrete-time

Ehrenfest model leading to Ornstein–Uhlenbeck process it can be done a jump-diffusion approximation for the process $M(t)$. We first rename the parameters related to the birth and death process, given in Equations (1.24) and (1.23), by setting

$$\lambda = \frac{\alpha}{2} + \frac{\gamma}{2}\epsilon, \qquad \mu = \frac{\alpha}{2} - \frac{\gamma}{2}\epsilon, \tag{1.28}$$

with $\varepsilon > 0$, $\alpha > 0$ and

$$-\frac{\alpha}{\epsilon} < \gamma < \frac{\alpha}{\epsilon}$$

For all $t \geq 0$, consider the position $M_\epsilon^*(t)\epsilon$, so that $\{M_\epsilon(t), t \geq 0\}$ is a continuous-time stochastic process with state-space $\{-N\epsilon, -N\epsilon+\epsilon, \ldots, -\epsilon, 0, \epsilon, \ldots, N\epsilon-\epsilon, N\epsilon\}$ and transient probabilities, for $j, n \in S$ and $t \geq 0$,

$$\begin{aligned} p_{j,n}^*(t) &:= \mathbb{P}\left[M_\epsilon^*(t) = n\epsilon | M_\epsilon^*(0) = j\epsilon\right] \\ &= \mathbb{P}\left[n\epsilon \leq M_\epsilon^*(t) < (n+1)\epsilon | M_\epsilon^*(0) = j\epsilon\right] \equiv p_{j,n}(t). \end{aligned}$$

Under suitable limit conditions the scaled process $M^*(t)$ converges weakly to a jump-diffusion process $\{X(t); t \geq 0\}$ having state-space $\mathbb{R}$ and transition density

$$f(x,t|y) = \frac{\partial}{\partial x}\mathbb{P}[X(t) \leq x | X(0) = y], \qquad t \geq 0.$$

Indeed, with reference to system (1.26), we make use of (1.29) and assume that $p_{j,n}^*(t) \simeq f(x,t|y)\epsilon$ for ϵ close to 0, with $x = n\epsilon$ and $y = j\epsilon$, and expand f as Taylor series, with

$$\epsilon \to 0^+, \qquad N \to +\infty, \qquad N\epsilon \to +\infty, \qquad N\epsilon^2 \to \nu > 0. \tag{1.29}$$

Hence, under limits (1.29), from the first and the second equation of (1.26) we obtain the following partial diffusion equation, with $x \in \mathbb{R}$, $y \in \mathbb{R}$, $t \geq 0$:

$$\frac{\partial}{\partial t}f(x,t|y) = \frac{\partial}{\partial x}\{(\alpha x - \gamma\nu)f(x,t|y)\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{\alpha\nu f(x,t|y)\} - \xi f(x,t|y) + \xi\delta(x), \tag{1.30}$$

whereas from third and fourth equation of (1.26) we have

$$\lim_{x \pm \infty} f(x,t|y) = 0,$$

for $t \geq 0$ and $y \in \mathbb{R}$. Moeover, initial condition (1.27) gives the following delta-Dirac initial condition:

$$\lim_{t \to 0^+} f(x,t|y) = \delta(x-y). \tag{1.31}$$

We remark that Equation (1.30) is the Fokker–Planck equation for a temporally homogeneous jump-diffusion process $\{X(t),\, t \geq 0\}$ with state-space $\mathbb{R}$, having linear drift and constant infinitesimal variance. The jumps occur with constant rate ξ, and each jump makes $X(t)$ instantly attain the state 0.

In the sequence, for simplicity we set

$$\beta = \frac{\gamma \nu}{\alpha}. \tag{1.32}$$

We point out that if $\xi \to 0^+$ then (1.30) yields the Fokker–Planck equation of an Ornstein–Uhlenbeck process on $\mathbb{R}$, denoted by $\{\widetilde{X}(t), \quad t \geq 0\}$, with initial condition (1.31) , and having drift and infinitesimal variance

$$A_1(x) = -\alpha(x-\beta), \qquad A_2(x) = \alpha\nu,$$

with $\alpha > 0$, $\beta \in \mathbb{R}$ and $\nu > 0$. This process has state-space $\mathbb{R}$ and transition density denoted as

$$\widetilde{f}(x,t|y) = \frac{\partial}{\partial x}\mathbb{P}\left[\widetilde{X}(t) \leq x | \widetilde{X}(0) = y\right], \qquad t \geq 0.$$

We remark that, due to (1.28) and (1.32), the special case $\beta = 0$ arises when birth and death rates λ and μ are equal. In this case the drift of the approximating process becomes $A_1(t) = -\alpha x$ so that $\widetilde{X}(t)$ has an equilibrium point in the state 0.

Let us denote by

$$\widetilde{T}_y = \inf\{t \geq 0 : \widetilde{X}(t) = 0\}. \qquad y \in \mathbb{R} \setminus \{0\},$$

the first-passage time (FPT) of $\widetilde{X}(t)$ through 0, with $\widetilde{X}(0) = y$, and let $\widetilde{g}(0, t|y)$ be the corresponding density. For $y \in \mathbb{R} \setminus \{0\}$, the first-passage time density of $\tilde{X}(t)$ from y to 0 is given by:

$$\widetilde{g}(0, t|y) = \frac{2\alpha|y|e^{-\alpha t}}{\sqrt{\pi\nu}\,(1 - e^{-2\alpha t})^{3/2}} \exp\left\{-\frac{y^2 e^{-2\alpha t}}{\nu(1 - e^{-2\alpha t})}\right\}, \qquad y \neq 0, t > 0. \tag{1.33}$$

with initial condition

$$\widetilde{g}(0, 0|y) = \lim_{t\to 0} \widetilde{g}(0, t|y) = 0$$

Finally, accounting for the transition density of $X(t)$, $f(x, t|y)$ can be related to $\widetilde{f}(x, t|y)$ through the following relation

$$f(x, t|y) = e^{-\xi t} \widetilde{f}(x, t|y) + \xi \int_0^t e^{-\xi t} \widetilde{f}(x, \tau|0)\, \mathrm{d}\tau, \qquad x, y, \in \mathbb{R}, \qquad t > 0. \tag{1.34}$$

Making use of (1.33) and (1.34) in the case $\beta = 0$ we obtain

$$g(0, t|y) = e^{-\xi t}\, \widetilde{g}(0, t|y) + \xi\, e^{-\xi t} \mathrm{Erf}\left(|y|e^{-\alpha t} \left[\nu(1 - e^{-2\alpha t})\right]^{-1/2}\right), \qquad t \geq 0, y \neq 0, \tag{1.35}$$

where where $\widetilde{g}(0, t|y)$ is given in (1.33) and $\mathrm{Erf}(\cdot)$ is the error function (see Eq. 7.1.1 of [1]) defined by

$$\mathrm{Erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, \mathrm{d}t, \tag{1.36}$$

The initial condition of (1.35) is given by

$$g(0, 0|y) = \lim_{t\to 0} g(0, t|y) = \xi.$$

The FPT pdf (1.35) deserves interest in the realm of stochastic processes with stochastic reset (see, for instance, Kusmierz *et al.* [46] and Pal [61]).

Chapter 2

Residual varentropy of random lifetimes and applications

The aim of this chapter is to discuss properties and applications of residual varentropy with particular attention to the residual lifetime distribution or residual varentropy. After an introduction to the meaning and the main properties of the varentropy, the investigation of some mathematical properties of the residual varentropy is object of Section 1. Finally some applications are proposed in Section 2.

2.1 Varentropy

The varentropy is the variance of the information content and is used in various applications of information theory, such as for the estimation of the performance of optimal block-coding schemes. Recent contributions on the varentropy can be found in various papers by Arikan [3], Bobkov and Madiman [13], Fradelizi *et al.* [32], Kontoyiannis and Verdú [43], [44], [71]. Most of such results have been aimed to mathematical properties or to applications in information theory. However, it should be pointed out that such information measures often deserve interest in other fields, such as reliability and survival analysis. See, for instance Nanda and Chowdhury [59] for a recent comprehensive review on the Shannon's entropy and its applications in various fields. Several investigations have been oriented in the past to assess the information content of stochastic systems with special attention to dynamic measures related to the residual lifetime, the past lifetime, the inactivity time and their suitable generalizations. However, no efforts have been dedicated to the analysis of the variance of the information content in dynamic contexts.

Bobkov and Madiman [13] investigated a relevant problem concerning the concentration of the information content around the entropy in high dimensions, when the pdf of X is log-concave[1]. Restricting our attention to the one-dimensional case, hereafter we focus on a relevant quantity related to the concentration of $IC(X)$ around $H(X)$, namely the so-called varentropy of X. The *varentropy of a random lifetime* X is defined as the variance of the information content of X, i.e.

$$\begin{aligned} V(X) &:= \mathrm{Var}[IC(X)] = \mathrm{Var}[\log f(X)] = \mathbb{E}[(IC(X))^2] - [H(X)]^2 \\ &= \int_{-\infty}^{\infty} f(x)[\log f(x)]^2 \,\mathrm{d}x - \left[\int_{-\infty}^{\infty} f(x)\log f(x)\,\mathrm{d}x\right]^2. \end{aligned} \tag{2.1}$$

The varentropy thus measures the variability in the information content of X. The relevance of this measure has been pointed out in various investigations, especially from Fradelizi *et al.* [32] that start from the concept of varentropy of a random variable X and use it to find an optimal varentropy bound for log-concave distributions. Furthermore, a sharp uniform bound on varentropy for log-concave distributions is found in the work of Madiman [53]. An alternative way to calculate a bound for varentropy is discussed in the article due to Goodarzi *et al.* [34] where the authors use some concepts of reliability theory. The generalization from log-concave to convex measures has been studied in the work of Li *et al.* [48] where a bound on the varentropy for convex measures is discussed. Other works that deal with the bounds of the varentropy in the contest of source coding are due to Arikan [3], that has analized the case of the polar transform showing that varentropy decreases to zero asymptotically as the transform size increases. In studies on the lossless source code it is possible to relate varentropy to the dispersion of the source code,

[1] We say that X has a log-concave density if its pdf is such that $\log f(x)$ is a concave function on $(0,\infty)$. Equivalently, we say that X is ILR (increasing in likelihood ratio).

as shown in the papers by Kontoyiannis and Verdú [43], [44], [71]. Specifically, together with the entropy rate, the varentropy rate serves to tightly approximate the fundamental nonasymptotic limits of fixed-to-variable compression for all but very small block lengths.

We remark that, due to (4) and (2.1), both the entropy and varentropy do not depend on the realization of X but only on its pdf f.

The *varentropy of a discrete random variable* X taking values in the set $\{x_i; i \in I\}$ is expressed as

$$V(X) = \mathrm{Var}[IC(X)] = \sum_{i\in I} \mathbb{P}(X = x_i)[\log \mathbb{P}(X = x_i)]^2 - [H(X)]^2. \tag{2.2}$$

Hereafter we analyze an illustrative example related to the study of varentropy for a three-valued random variable.

Example 2.1 Let X be a discrete random variable such that, for a fixed $h > 0$,

$$\mathbb{P}(X = h) = p, \qquad \mathbb{P}(X = 0) = 1 - p - q, \qquad \mathbb{P}(X = -h) = q, \tag{2.3}$$

with $0 \le q \le 1 - p \le 1$. Thus, from (2) we have that the entropy expressed in natural units is

$$H(X; p, q) = -p \log p - (1 - p - q) \log (1 - p - q) - q \log q, \tag{2.4}$$

Thus, from (2.2) we have

$$V(X; p, q) = p(\log p)^2 + (1 - p - q)[\log (1 - p - q)]^2 + q(\log q)^2 - [H(X; p, q)]^2. \tag{2.5}$$

Figure 2.1 shows the varentropy given in (2.5) as a function of (p, q). Clearly, it confirms the symmetry property $V(X; p, q) = V(X; q, p)$. We can see that the varentropy vanishes in the following 7 cases: $(p, q, 1-p-q) = (0, 0, 1)$, $(0, 1, 0)$, $(1, 0, 0)$, $(0.5, 0.5, 0)$, $(0.5, 0, 0.5)$, $(0, 0.5, 0.5)$, $(1/3, 1/3, 1/3)$. Moreover, the maximum of $V(X; p, q)$ is attained for $(p, q, 1-p-q) = (0.06165, 0.06165, 0.8767)$, $(0.8767, 0.06165, 0.06165)$, $(0.06165, 0.8767, 0.06165)$.

Now consider a system based on the superposition of three Gaussian signals. Namely we deal with a random variable, say Y, whose pdf is a mixture of Gaussian densities with unity variance and mean given by h, 0, $-h$ according to the probability law specified in (2.3). Hence, for $x \in \mathbb{R}$ one has

$$f_Y(x) = (2\pi)^{-1/2}\left[pe^{-(x-h)^2/2} + (1-p-q)e^{-x^2/2} + qe^{-(x+h)^2/2}\right]. \tag{2.6}$$

Figure 2.2 shows some instances of the corresponding varentropy as a function of h, determined numerically by means of (2.1). It can be shown that $V(Y)$ is not monotonic in h; moreover it reaches large values for the choices of (p,q) that maximize $V(X;p,q)$ and for large values of h.

The relevance of the entropy in information theory and other disciplines is very well known, whereas the varentropy has attracted less attention. Nevertheless, the latter plays a relevant role in the assessment of the statistical significance of entropy. Specifically, in the discrete case the entropy (2) represents the expected number of symbols, in natural base, required to code an event produced by a source of information governed by the probability distribution of X. In this case the varentropy (2.2) measures the variability related to such a coding. In other terms, if two sources of information have the same entropy, than the number of digits required in the average to code two sequences produced by such sources is the same and is proportional to $H(X)$. However, the number of digits required for a single observed sequence in the average is closer to the expected one for the source having the smallest varentropy. Hence, $V(X)$ measures how much the entropy is meaningful in the coding of sequences of symbols generated by X.

Example 2.2 Let Y be a Bernoulli random variable having distribution $\mathbb{P}(Y=0) = 1-\theta$, $\mathbb{P}(Y=1)=\theta$, with $0 \le \theta \le 1$. By means of numerical calculations it is easy to see that for $\theta \approx 0.337009$ one has $H(Y) \approx 0.639032$ and $V(Y) \approx 0.1023$. For the distribution considered in the Example 2.1, if $p=q=0.1$ from (2.4) and (2.5) we have $H(X) \approx 0.639032$ and $V(X) \approx 0.691852$, respectively. Hence, the considered random

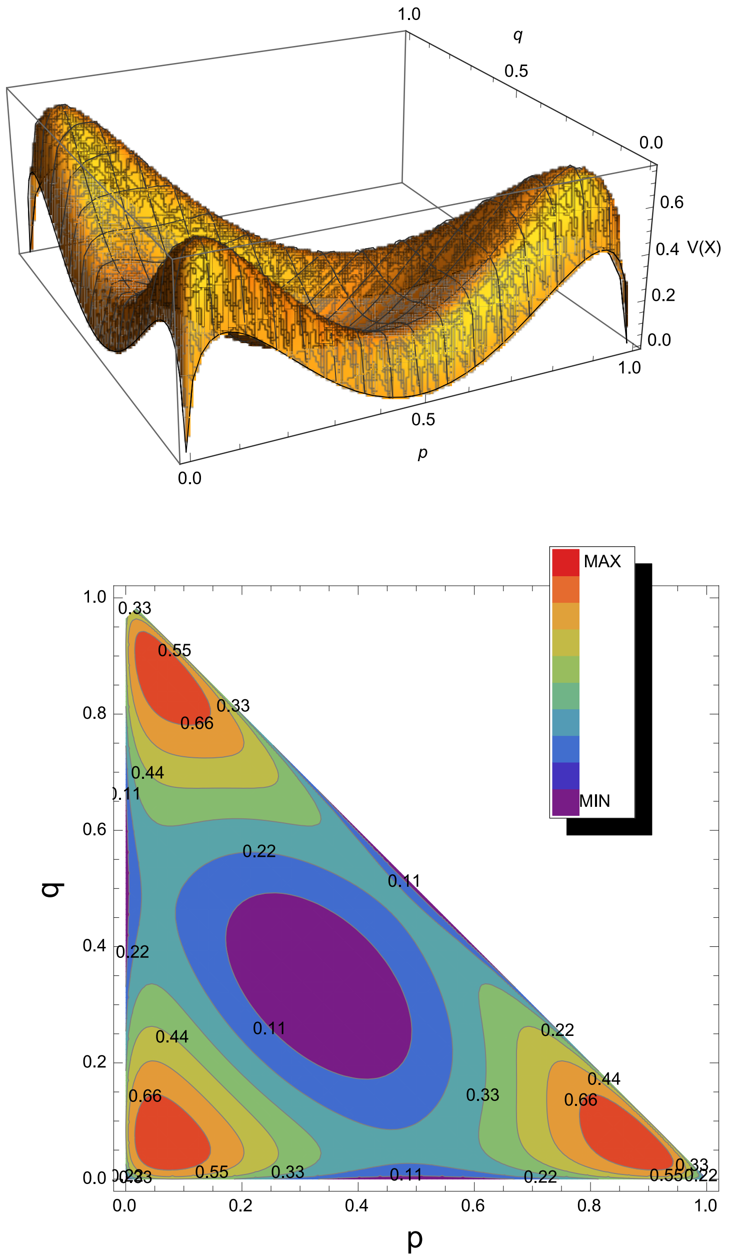


Figure 2.1: Plots of varentropy (2.5); top: 3D plot; bottom: contourplot.

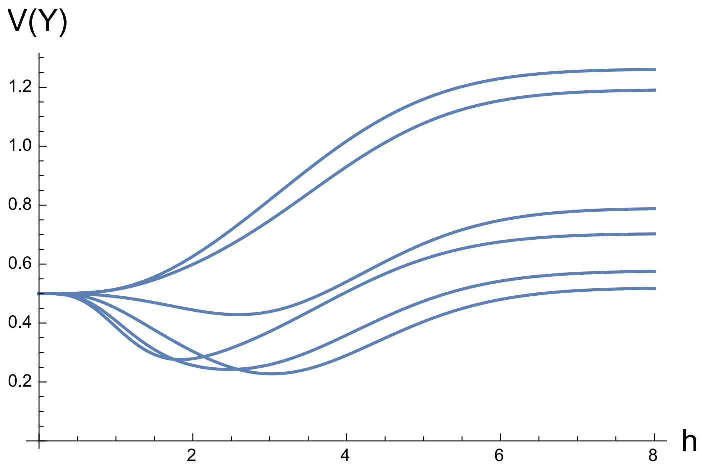


Figure 2.2: The varentropy corresponding to pdf (2.6) for $p = q = 0.06165$, 0.1, 0.2, 0.45, 0.4, 0.3 (from top to bottom for large values of h).

variables have the same entropy, but the varentropy of X is larger. This implies that the coding procedure is much more reliable for sequences generated by Y.

In Table 2.1 the differential entropy and differential varentropy[2] have been reported in the case of uniform, exponential, normal, logistic and Cauchy distributions. Here the set

$$S_X := \{x \in \mathbb{R} \ : \ F(x) > 0\}$$

is the support of the given distribution of X. We note that for these distributions differential varentropy has a fixed constant value that does not depend on the parameters of the distributions.

2.2 Residual varentropy

Recalling that the varentropy of a random lifetime X is defined in (2.1), in ref. [24] the notion of varentropy is extended to the residual lifetime considered in (1.1). Namely, recalling Equation (1.3), for $t \in S_X$

Definition 2.1 The *varentropy of residual lifetime distribution* or *residual varentropy* of

[2]The differential varentropy for Cauchy Distribution in Table 2.1 was estimated through numerical calculations.

Table 2.1: Differential entropy and differential varentropy for selected distributions.

Distribution	Pdf $f(x)$	Differential entropy $H(X)$	Differential varentropy $V(X)$
Uniform $S_X = (0, \theta)$	$\frac{1}{\theta}$	$\log \theta$	0
Exponential $S_X = (0, \infty)$	$\lambda e^{-\lambda x}$, $\lambda > 0$	$1 - \log \lambda$	1
Normal $S_X = \mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mu \in \mathbb{R}$, $\sigma > 0$	$\frac{1}{2} \log (2\pi e \sigma^2)$	$\frac{1}{2}$
Logistic $S_X = \mathbb{R}$	$\frac{e^{-\frac{x-\mu}{\sigma}}}{1+e^{-\frac{x-\mu}{\sigma}}}$, $\mu \in \mathbb{R}$, $\sigma > 0$	$\log \sigma + 2$	$4 - \frac{\pi^2}{3}$
Cauchy $S_X = \mathbb{R}$	$\frac{1}{\pi} \frac{y_0}{(x-x_0)^2 + y_0^2}$, $x_0 \in \mathbb{R}$, $y_0 > 0$	$\log (4\pi y_0)$	3.28987...

a random lifetime X_t is defined as

$$V(X_t) := \mathrm{Var}[IC(X_t)] = \int_t^\infty \frac{f(x)}{\overline{F}(t)} \left(\log \frac{f(x)}{\overline{F}(t)} \right)^2 \mathrm{d}x - [H(X_t)]^2 \tag{2.7}$$

An alternative expression to the varentropy of a residual lifetime distribution is the following.

Proposition 2.1 *For all $t \in S_X$ the residual varentropy (2.8) has the following alternative form*

$$V(X_t) = \frac{1}{\overline{F}(t)} \int_t^\infty f(x) \left(\log f(x) \right)^2 \mathrm{d}x - [\Lambda(t) + H(X_t)]^2 \tag{2.8}$$

where $\Lambda(t)$ is given in (1.6), and $H(X_t)$ is provided in (1.13) and (1.14).

Proof. Applying the properties of the logarithm to Equation (2.8) we obtain

$$\begin{aligned} \int_t^\infty \frac{f(x)}{\overline{F}(t)} \left(\log \frac{f(x)}{\overline{F}(t)} \right)^2 \mathrm{d}x &= \int_t^\infty \frac{f(x)}{\overline{F}(t)} \big([\log f(x)]^2 + [\log \overline{F}(t)]^2 \\ &- 2 \log f(x) \log \overline{F}(t) \big) \, \mathrm{d}x \\ &= \frac{1}{\overline{F}(t)} \int_t^\infty f(x) (\log f(x))^2 \, \mathrm{d}x + [\log \overline{F}(t)]^2 \\ &- 2 \frac{\log \overline{F}(t)}{\overline{F}(t)} \int_t^\infty f(x) \log f(x) \, \mathrm{d}x \end{aligned} \tag{2.9}$$

and applying Equation (1.13) for the last two terms of Equation (2.9) it follows that

$$\begin{aligned}[\log \overline{F}(t)]^2 - 2\frac{\log \overline{F}(t)}{\overline{F}(t)} \int_t^{\infty} f(x) \log f(x) \, \mathrm{d}x &= \Lambda(t)^2 - 2\Lambda(t)(H(X_t) + \Lambda(t)) \\ &= -[\Lambda(t) + H(t)]^2 + [H(X_t)]^2\end{aligned}$$

from which applying (1.6) we get

$$\begin{aligned}V(X_t) &= \int_t^{\infty} \frac{f(x)}{\overline{F}(t)} \left(\log \frac{f(x)}{\overline{F}(t)}\right)^2 \mathrm{d}x - [H(X_t)]^2 \\ &= \frac{1}{\overline{F}(t)} \int_t^{\infty} f(x)(\log f(x))^2 \, \mathrm{d}x - [H(X_t) + \Lambda(t)]^2.\end{aligned}$$

As a dynamical measure of the variance of the information content, the residual varentropy was a topic of interest in the last years. Goodarzi *et al.* applied residual varentropy in the study of the upper bounds for the variance of functions of the inactivity time (see ref. [33]) and in the variance of functions of residual life of random variables (see ref. [35]). Other alternative definitions are applications to the order statistics (see ref. [50]) and to the Tsallis varentropy (see ref. [51]).

2.3 Mathematical properties of residual varentropy

In the following some results involving mathematical properties of residual varentropy are shown. These results are obtained through the analysis done in ref. [24]. In particular we illustrate some properties related to the constant residual varentropy, varentropy for linear transformations and upper and lower boundaries of the residual varentropy.

2.3.1 Constant residual varentropy

Making use of Eq. (2.8) we can show in Table 2.2 some examples in which the residual varentropy is constant.

Table 2.2: Selected distributions with constant varentropy.

Distribution	Pdf $f(x)$	Residual entropy $H(X_t)$	Residual varentropy $V(X_t)$
Uniform $S_X=(0,\theta)$	$\frac{1}{\theta}$	$\log(\theta-t)$	0
Exponential $S_X=(0,\infty)$	$\lambda e^{-\lambda x}$, $\lambda>0$	$1-\log\lambda$	1
Triangular $S_X=(0,1)$	$2(1-x)$	$\frac{1}{2}+\log\frac{1-t}{2}$	$\frac{1}{4}$

In the following we determine the conditions for which the residual varentropy is costant. To this aim we first obtain an expression of its derivative.

Proposition 2.2 *For all $t\in S_X$ the derivative of the residual varentropy is*

$$V'(X_t)=\lambda(t)\left\{V(X_t)-[H(X_t)+\log\lambda(t)]^2\right\}. \tag{2.10}$$

Proof. By differentiating both sides of Eq. (2.8), and recalling (1.4) we have

$$\begin{aligned} V'(X_t)=\lambda(t)&\left\{\frac{1}{\overline{F}(t)}\int_t^\infty f(x)[\log f(x)]^2\,\mathrm{d}x-[\log f(t)]^2\right\}\\ &-2[\Lambda(t)+H(X_t)][\lambda(t)+H'(X_t)], \qquad t\in D. \end{aligned} \tag{2.11}$$

Then, making use of Eqs. (1.17) and (2.8), from (2.11) we get

$$\begin{aligned} V'(X_t)=\lambda(t)&\left\{V(X_t)+[\Lambda(t)+H(X_t)]^2-[\log f(t)]^2\right.\\ &\left.-2[\Lambda(t)+H(X_t)][H(X_t)+\log\lambda(t)]\right\}, \qquad t\in D. \end{aligned}$$

Hence, due to (1.6), after some calculations we obtain Eq. (2.10).

As a consequence of Proposition 2.2 we can now provide some useful results involving the residual varentropy, the residual entropy, the hazard rate and the varentropy of a lifetime X.

Theorem 2.1 *Let X have a pdf such that $f(t) > 0$ for all $t \in (0, r)$, with $r \in (0, \infty]$. Let $c \in \mathbb{R}$, the following statements are equivalent*

(i) The residual varentropy is constant, such that

$$V(X_t) = c^2, \qquad \forall t \in [0, r), \tag{2.12}$$

(ii) The generalized hazard rate (1.9) of X is constant, such that

$$\lambda_{c-1}(t) = e^{c-H(X)}, \qquad t \in [0, r), \tag{2.13}$$

(iii)

$$H(X_t) + \log \lambda(t) = c, \qquad \forall t \in (0, r), \tag{2.14}$$

Proof. It is immediate to show that (i) implies (iii). In fact, since $f(t) > 0$ for all $t \in (0, r)$, the assumption (2.12) immediately gives (2.14), due to (2.10).

In order to show that (iii) implies (i) let us assume that the condition (2.14) is satisfied, then Eq. (2.10) becomes

$$V'(X_t) = \lambda(t) \left[V(X_t) - c^2 \right], \qquad t \in (0, r), \tag{2.15}$$

with initial condition $V(X_t)|_{t=0} = V(X)$. Equation (2.15) can be solved applying formula

$$x'(t) = a(t)x(t) + b(t) \qquad x(0) = x_0$$

that has solution

$$x(t) = e^{\int_0^t a(s) \mathrm{d}s} \, x_0 + \int_0^t e^{\int_s^t a(u) \mathrm{d}u} \, b(s) \, \mathrm{d}s$$

from which if $x(t) = V(t)$, $a(t) = \lambda(t)$ and $b(t) = -c^2 \lambda(t)$ and $x_0 = V(X)$ we have

$$V(t) = e^{\int_0^t \lambda(s) \mathrm{d}s} \, V(X) - c^2 \int_0^t e^{\int_s^t \lambda(u) \mathrm{d}u} \, \lambda(s) \, \mathrm{d}s, \qquad t \in [0, r). \tag{2.16}$$

Applying (1.6) to (2.16) we have

$$V(X_t) = \frac{V(X)}{\overline{F}(t)} - \frac{c^2}{\overline{F}(t)} \int_0^t \overline{F}(s)\lambda(s)\ \mathrm{d}s, \qquad t \in [0, r)$$

and applying Equation (1.4) we have

$$\int_0^t \overline{F}(s)\lambda(s)\mathrm{d}s = \int_0^t f(s)\mathrm{d}s = F(t) = 1 - \overline{F}(t), \qquad t \in [0, r)$$

and so we obtain the equation

$$V(X_t) = c^2 + \frac{V(X) - c^2}{\overline{F}(t)} \tag{2.17}$$

In order to show (2.12), without loss of generality, let us assume that the support of X is $S_X = (0, \infty)$. The differential varentropy of X can be obtained from Eq. (2.8) for which we have

$$V(X) = \int_0^\infty f(x)[\log f(x)]^2\ \mathrm{d}x - [H(X)]^2 \tag{2.18}$$

where $H(X)$ is the differential entropy of X (cf. (4)). By the hypothesis (2.14), we have

$$\log f(x) = c \log \overline{F}(x) + c - H(X). \tag{2.19}$$

If we substitute (2.19) into the integral at the right hand side of (2.18) we get

$$\int_0^\infty f(x)[\log f(x)]^2\ \mathrm{d}x = \int_0^\infty f(x)\left[c\log\overline{F}(x) + c - H(X)\right]^2\ \mathrm{d}x. \tag{2.20}$$

Expanding the integrand at the right hand side of (2.20) we obtain

$$\begin{aligned}\int_0^\infty f(x)[\log f(x)]^2\ \mathrm{d}x &= c^2 \int_0^\infty f(x)[\log \overline{F}(x)]^2\ \mathrm{d}x \\ &+ 2c(c - H(X)) \int_0^\infty f(x) \log \overline{F}(x)\ \mathrm{d}x + (c - H(X))^2.\end{aligned} \tag{2.21}$$

Applying integration by part we get the following results:

$$\int_0^\infty f(x)\log\overline{F}(x)\ \mathrm{d}x = -\int_0^\infty f(x)\ \mathrm{d}x = -1 \tag{2.22}$$

$$\int_0^\infty f(x)[\log\overline{F}(x)]^2\ \mathrm{d}x = -2\int_0^\infty f(x)\log\overline{F}(x)\ \mathrm{d}x = 2 \tag{2.23}$$

from which substituting (2.22) and (2.23) into (2.21), after a bit simplification we obtain

$$\int_0^\infty f(x)[\log f(x)]^2\ \mathrm{d}x = [H(X)]^2 + c^2 \tag{2.24}$$

Finally, if we substitute back the result (2.24) into (2.18) we get

$$V(X) = c^2$$

from which applying (2.17) we get immediately $V(X_t) = c^2$ for all $t \in S_X$.

Let us show that (iii) implies (ii). Assume that the Eq. (2.13) is fulfilled. Making use of (1.4) and (1.14a), we have

$$H(X_t) + \log\lambda(t) = \log f(t) + \frac{1}{\overline{F}(t)}\left\{H(X) + \int_0^t f(x)\log f(x)\,\mathrm{d}x\right\}. \tag{2.25}$$

From the assumption (2.13) it is not hard to see that

$$\int_0^t f(x)\log f(x)\,\mathrm{d}x = -F(t)\,H(X) - c\,\overline{F}(t)\log\overline{F}(t).$$

Hence, due to Eqs. (2.13) and (2.25) we have

$$H(X_t) + \log\lambda(t) = H(X) + \log\frac{f(t)}{[\overline{F}(t)]^c} = c,$$

so that (2.14) holds.

Finally, let us prove that (ii) implies (iii). In fact, rearranging Eq. (1.17), we have

$$H(X_t) + \log \lambda(t) = \frac{H'(X_t)}{\lambda(t)} + 1,$$

so that, due to Eq. (2.14), one has

$$H'(X_t) = (c-1)\lambda(t), \qquad t \in (0, r).$$

By integration over $[0, t]$, and recalling (1.6), one obtains

$$H(X_t) - H(X) = (c-1)\Lambda(t), \qquad t \in [0, r).$$

Comparing the latter identity with Eq. (2.14) and in virtue of (1.6), we have

$$c - \log \lambda(t) - H(X) = (1-c) \log \overline{F}(t) \qquad t \in [0, r),$$

and applying (1.4) we get

$$\log f(t) - c \log \overline{F}(t) = \log \frac{f(t)}{[\overline{F}(t)]^c} = c - H(X) \qquad t \in [0, r)$$

which gives immediately relation (2.13) by virtue of (1.9).

Remark 2.1 (i) It is worth pointing out that, due to Theorem 3.1 of Asadi and Ebrahimi [5], the condition expressed in Eq. (2.14) is fulfilled if and only if X has a generalized Pareto distribution, with survival function

$$\overline{F}(t) = \left(\frac{b}{at+b}\right)^{\frac{1}{a}+1}, \qquad t \geq 0, \tag{2.26}$$

for $a > -1$ and $b > 0$. The generalized Pareto distribution is a flexible statistical model which is employed in several research areas, such as statistical physics, econophysics and social sciences, since its distribution possesses a tail of general form. Specifically, it includes the exponential distribution ($a \to 0$), the Pareto distribution ($a > 0$, with heavy tail), and the power distribution ($-1 < a < 0$, with bounded support). An intuitive reason leading to the above result is due to the property that the generalized Pareto distribution is the only family of distributions whose mean residual function (1.5) is linear (see Hall and Wellner [38]). Indeed, for the survival function (2.26) we have $m(t) = at + b$, with hazard rate function $\lambda(t) = \frac{1+a}{at+b}$. For a recent characterization of this distribution in the context of shape functionals see Arriaza *et al.* [4].

(ii) A special case arises from (2.26) in the limit as $a \to \infty$ and $b \to \infty$, with $\frac{a}{b} \to \lambda > 0$, by which the pdf and the survival function of X are given respectively by

$$f(t) = \frac{\lambda}{(1+\lambda t)^2}, \qquad \overline{F}(t) = \frac{1}{1+\lambda t}, \qquad t \in [0, \infty).$$

In this case X has a modified Pareto distribution that describes the first arrival time in a Geometric counting process with parameter $\lambda > 0$ (cf. Section 2.2 of [26], for instance). From Eq. (1.9) it immediately follows that the generalized hazard rate of X is a constant for $\alpha = 1$, i.e. $\lambda_1(t) \equiv \lambda$. As a consequence, Eq. (2.13) is fulfilled for $c = 2$ and $H(X) = 2 - \log \lambda$. From Theorem 2.1 we thus obtain the (increasing) residual entropy,

$$H(X_t) = 2 - \log \frac{\lambda}{1+\lambda t}, \qquad t \geq 0,$$

and the corresponding constant residual varentropy, $V(X_t) = 4$. It is worth pointing out that in this special case the mean residual lifetime is infinite. Hence, for such a stochastic model the residual entropy and the residual varentropy provide useful information even if the mean residual lifetime is not finite.

The following example is concerning a family of distributions for which the residual varentropy exhibits different behaviors.

Example 2.3 Let $X_{\lambda,k}$ have Weibull distribution, with pdf

$$f_{\lambda,k}(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \qquad x > 0, \tag{2.27}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. Recall that this family of distributions includes special cases of interest, such as the exponential distribution (for $k = 1$) and the Rayleigh distribution (for $k = 2$). The expression of the residual varentropy is reported in Appendix A (cf. Eq. (A.10)). The behavior of the pdf (2.27) and of the corresponding residual varentropy is visualized in Fig. 2.3 for some choices of the shape parameter. It can be seen that the residual varentropy is decreasing, constant, increasing, non monotonic for $k = 0.5$, 1, 1.5, 3.5 respectively.

2.3.2 Linear transformations

Let us now analyze the effect of linear transformations to the residual varentropy.

Proposition 2.3 *Let X and Y be related by (1.15). Hence, for their residual varentropies we have:*

$$V(Y_t) = V\left(X_{\frac{t-b}{a}}\right) \qquad \forall\ t. \tag{2.28}$$

Proof. Clearly, from (1.15) we have that the cdf's and the pdf's of Y and X are related by $F_Y(x) = F_X\left(\frac{x-b}{a}\right)$ and $f_Y(x) = \frac{1}{a}\, f_X\left(\frac{x-b}{a}\right)$. Hence, recalling (2.7) and (1.16), it is not hard to see that

$$V(Y_t) = \int_{\frac{t-b}{a}}^{\infty} \frac{f_X(x)}{\overline{F}_X(\frac{t-b}{a})}\left[\log\frac{f_X(x)}{\overline{F}_X(\frac{t-b}{a})} - \log a\right]^2 \mathrm{d}x - \left[H\left(X_{\frac{t-b}{a}}\right) + \log a\right]^2.$$

that can be expanded obtaining

$$\begin{aligned}V(Y_t) = &\int_{\frac{t-b}{a}}^{\infty} \frac{f_X(x)}{\overline{F}_X(\frac{t-b}{a})}\left[\log\frac{f_X(x)}{\overline{F}_X(\frac{t-b}{a})}\right]^2 \mathrm{d}x\\ &- 2\log a\int_{\frac{t-b}{a}}^{\infty} \frac{f_X(x)}{\overline{F}_X(\frac{t-b}{a})}\log\frac{f_X(x)}{\overline{F}_X(\frac{t-b}{a})}\mathrm{d}x + (\log a)^2 - \left[H\left(X_{\frac{t-b}{a}}\right) + \log a\right]^2.\end{aligned}$$

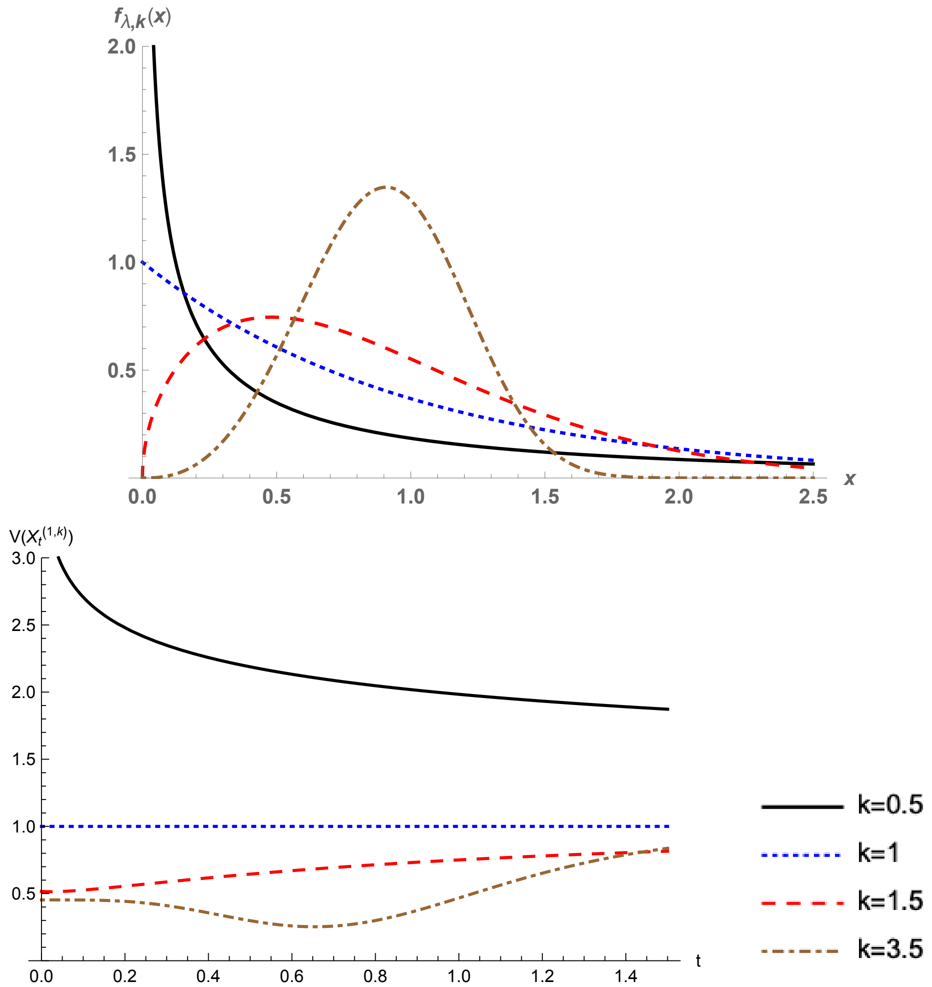


Figure 2.3: (top) Weibull pdf, given in (2.27), and (bottom) residual varentropy for $\lambda = 1$ and various choices of k (as indicated in the label).

The thesis (2.28) follows finally from Equation (1.13).

2.3.3 Bounds

We conclude this section by discussing some bounds to the residual varentropy.

First, we provide a lower bound for $V(X_t)$.

Theorem 2.2 *Let X_t be a residual lifetime as defined in (1.1), and assume that the corresponding mean residual lifetime $m(t)$ and variance residual lifetime $\sigma^2(t)$ are finite (cf. (1.5) and (1.7), respectively). Then, for all $t \in D$,*

$$V(X_t) \geq \sigma^2(t)\,(\mathbb{E}[w_t'(X_t)])^2 \tag{2.29}$$

where the function $w_t(x)$ is defined by

$$\sigma^2(t)\,w_t(x)\,f_t(x) = \int_0^x [m(t) - z]\,f_t(z)\,\mathrm{d}z, \qquad x > 0,$$

with $f_t(x)$ given in the (1.3).

Proof. If X is an absolutely continuous random variable with pdf $f(x)$, mean μ and variance σ^2, then (cf. [15])

$$\mathrm{Var}(g(X)) \geq \sigma^2\,(\mathbb{E}[w(X)g'(X)])^2, \tag{2.30}$$

where the function $w(x)$ is defined by

$$\sigma^2\,w(x)\,f(x) = \int_0^x (\mu - z)\,f(z)\,\mathrm{d}z, \qquad x > 0,$$

Hence, by taking X_t as reference, with $g(x) = -\log f(x)$ and integrating by parts, similarly as Equation (3.9) of Goodarzi *et al.* [34] we obtain (2.29).

Let us observe that the inequality in (2.30) holds if and only if X is exponentially distributed.

Hereafter we determine suitable upper bounds to the residual varentropy, thus providing conditions on its finiteness.

Theorem 2.3 *Given a random lifetime X with log-concave pdf $f(x)$, then*

$$V(X_t) \leq 1, \qquad \textit{for all } t \in D.$$

Proof. We note that if $f(x)$ is log-concave, then also $f_t(x)$ is log-concave due to (1.3). Hence, the proof is a direct consequence of Theorem 2.3 of Fradelizi et al. [32], which states that the varentropy of a random lifetime with log-concave pdf is not greater than 1.

The following bound for the residual varentropy is expressed in terms of the residual entropy of X (cf. (1.13)) and weighted residual entropy of X (cf. (1.18)).

Theorem 2.4 *If X is a random lifetime such that its pdf satisfies*

$$e^{-\alpha x-\beta} \leq f(x) \leq 1 \qquad \forall x \geq 0, \tag{2.31}$$

with $\alpha > 0$ and $\beta \geq 0$, then for all $t \geq 0$

$$V(X_t) \leq \alpha[\Lambda(t)\delta(t) + H^w(X_t)] + \beta[\Lambda(t) + H(X_t)] - [\Lambda(t) + H(X_t)]^2. \tag{2.32}$$

where $\delta(t)$ is the vitality function (1.8) and $\Lambda(t)$ is the cumulative hazard function (1.6).

Proof. From Eq. (2.8), due to (2.31) one has

$$V(X_t) \leq -\frac{1}{\overline{F}(t)} \int_t^{\infty} (\alpha x + \beta) f(x) \log f(x) \mathrm{d}x - [\Lambda(t) + H(X_t)]^2, \qquad t \geq 0. \tag{2.33}$$

We note that Eqs. (1.5) and (1.8) give

$$\int_t^\infty x\, f(x)\mathrm{d}x = \overline{F}(t)\delta(t), \qquad t \geq 0,$$

where $\delta(t)$ is the vitality function Hence, recalling (1.6) and (1.8), Eq. (1.18) implies:

$$\int_t^\infty x\, f(x)\log f(x)\mathrm{d}x = -\overline{F}(t)[\Lambda(t)\delta(t) + H^w(X_t)], \qquad t \geq 0. \tag{2.34}$$

Moreover, from (1.14a) we have

$$\int_t^\infty f(x)\log f(x)\mathrm{d}x = -\overline{F}(t)[\Lambda(t) + H(X_t)], \qquad t \geq 0. \tag{2.35}$$

Finally, substituting (2.34) and (2.35) in (2.33) we immediately obtain the inequality (2.32).

2.4 Some applications of residual varentropy

In this section some applications of the residual varentropy are considered. The first two applications are illustrated in ref. [24]. They are the residual varentropy for proportional hazard rates model and the first-passage-time problem of an Ornstein–Uhlenbeck jump-diffusion process which arises as limit of the continuous-time Ehrenfest model. Finally, as a further application, a kernel estimation is proposed in order to study of a given sample of data.

2.4.1 Proportional hazards model

Let us now address the problem of evaluating the residual varentropy for the model (1.10). From (1.14a) it is not hard to see that the residual entropy of $X^{(a)}$ is expressed as

$$\begin{aligned} H(X_t^{(a)}) &= -\Lambda^{(a)}(t) - \frac{1}{[\overline{F}(t)]^a} \int_t^{\infty} f^{(a)}(x) \log f^{(a)}(x)\, \mathrm{d}x \\ &= -a\,\Lambda(t) - \frac{1}{[\overline{F}(t)]^a} \int_0^{[\overline{F}(t)]^a} \ell(y;a)\, \mathrm{d}y, \qquad t>0, \end{aligned} \tag{2.36}$$

with $y = [\overline{F}(x)]^a$, and where

$$\ell(y;a) := \log \Big\{ a\, y^{1-1/a} f[\overline{F}^{-1}(y^{1/a})] \Big\}, \qquad 0<y<1. \tag{2.37}$$

Hence, recalling (2.8), from (1.12) and (2.36) after some calculations we obtain the residual varentropy of $X^{(a)}$, for $t>0$:

$$\begin{aligned} V(X_t^{(a)}) &= \frac{\int_t^{\infty} f^{(a)}(x)[\log f^{(a)}(x)]^2\, \mathrm{d}x}{[\overline{F}(t)]^a} - \left[\frac{\int_t^{\infty} f^{(a)}(x) \log f^{(a)}(x)\, \mathrm{d}x}{[\overline{F}(t)]^a} \right]^2 \\ &= \frac{1}{[\overline{F}(t)]^a} \int_0^{[\overline{F}(t)]^a} [\ell(y;a)]^2 \mathrm{d}y - \left\{ \frac{1}{[\overline{F}(t)]^a} \int_0^{[\overline{F}(t)]^a} \ell(y;a)\, \mathrm{d}y \right\}^2. \end{aligned} \tag{2.38}$$

Making use of Eqs. (1.4) and (1.6), one has $f(x) = \lambda(x) e^{-\Lambda(x)}$, so that the function introduced in (2.37) can be rewritten also as follows:

$$\ell(y;a) = \log \Big\{ a y \lambda\Big(\Lambda^{-1}\Big(-\frac{1}{a}\log y\Big)\Big)\Big\}.$$

An application can be immediately given to series systems.

Example 2.4 Consider a system composed of n units in series and characterized by i.i.d. random lifetimes $X_1, \ldots, X_n$. Let the survival function of each unit be denoted with $\overline{F}(t) = \mathbb{P}(X_i > t)$. Since the system lifetime is given by $X^{(n)} = \min\{X_1, \ldots, X_n\}$, the model of series system satisfies the proportional hazards model specified in (1.10), for $a = n \in \mathbb{N}$.

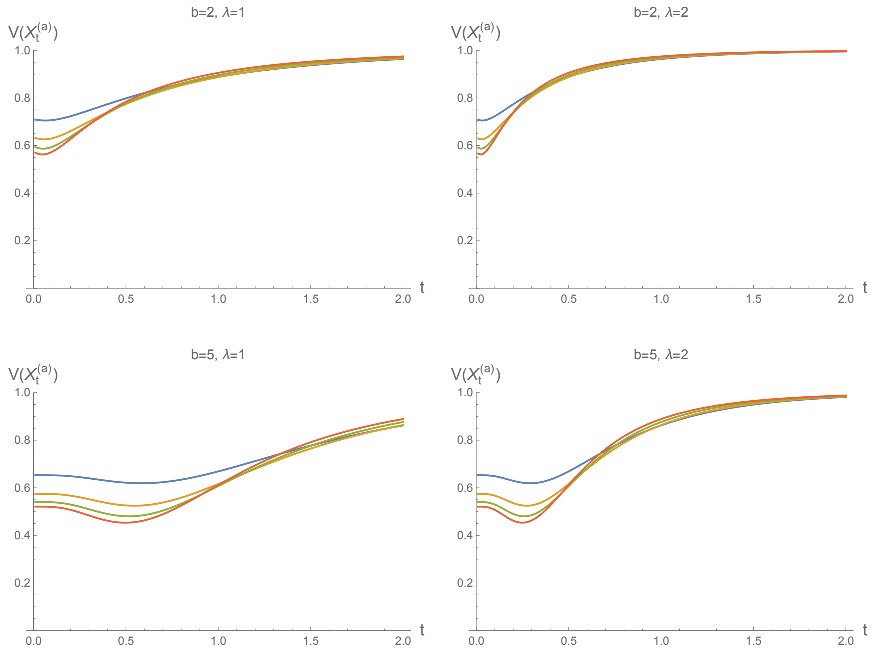


Figure 2.4: The residual varentropy of $X_t^{(a)}$ for the series system of Example 2.4, for $a = n = 1, 2, 3, 4$ (from top to bottom) and for b and λ as indicated.

For an illustrative example we assume that the random lifetimes X_i have generalized exponential distribution with survival function $\overline{F}(t) = 1 - (1 - e^{-\lambda t})^b$, $t \geq 0$, for $b > 0$. (We recall that this distribution plays a role in the construction of probabilistic models for damped random motions with finite velocities [22]). From (2.37) thus we have[3]

$$\ell(y; a) = \log \left\{ ab\lambda y^{1-\frac{1}{a}} (1 - y^{\frac{1}{a}})^{1-\frac{1}{b}} \left[1 - (1 - y^{\frac{1}{a}})^{\frac{1}{b}} \right] \right\}, \qquad 0 < y < 1.$$

From Eq. (2.38) we come to the residual varentropy of the system lifetime $X^{(n)}$. The expression of $V(X_t^{(a)})$ cannot be obtained in closed form, but it can be evaluated via numerical computations. Figure 2.4 shows some plots of the residual varentropy for some choices of $a = n$. It is clear that the varentropy decreases when the number of units grows, and approaches the value 1 when t becomes larger.

Example 2.5 Under the proportional hazards model, Eq. (2.38) can be used to construct time-varying reference sets for the information content of the residual lifetime

[3]We corrected a misprinting in the result obtained in Example 4.1 of [24].

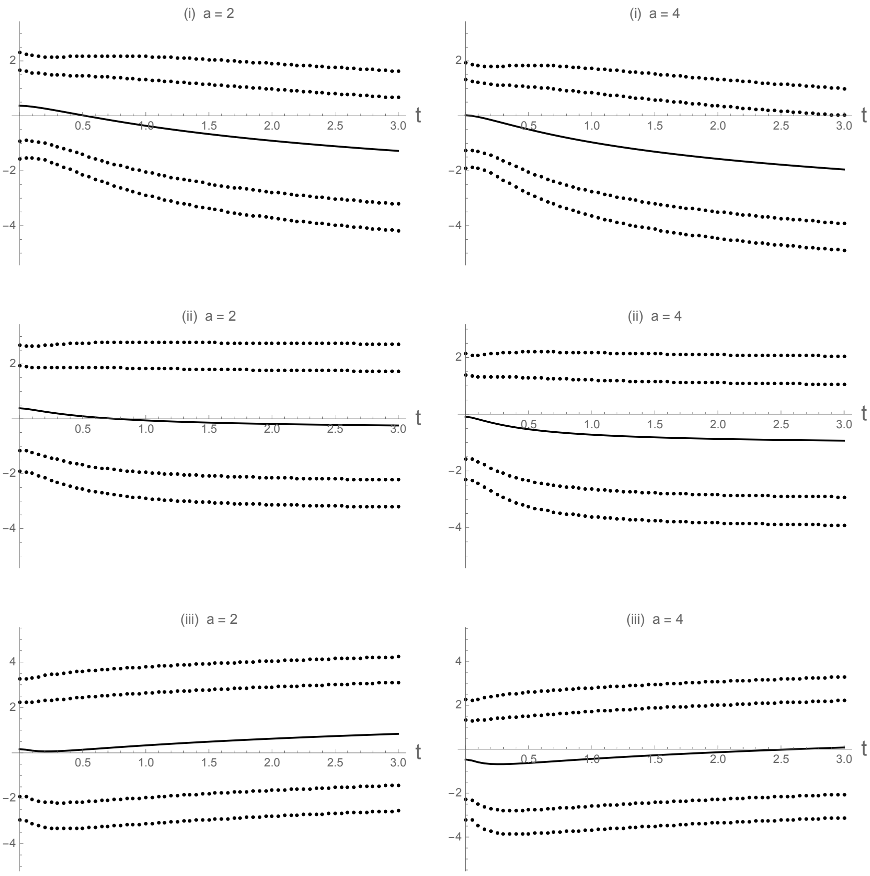


Figure 2.5: Residual entropy $H(X_t^{(a)})$ (full) and extremes of the intervals (2.39) (dotted) with $a = 2$ (left) and $a = 4$ (right), for the following baseline pdfs:
(i) (Weibull) $f(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \exp\{-\left(\frac{t}{\lambda}\right)^k\}$, $t > 0$, for $k = 2$, $\lambda = \frac{2}{\pi}$;
(ii) (gamma) $f(t) = \frac{1}{\theta}\left(\frac{t}{\theta}\right)^{r-1} \exp\left\{-\frac{t}{\theta}\right\} \frac{1}{\Gamma(r)}$, $t > 0$, for $r = 2$, $\theta = \frac{1}{2}$;
(iii) (lognormal) $f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left\{-\frac{(\log t-\mu)^2}{2\sigma^2}\right\}$, $t > 0$, for $\mu = -\frac{1}{2}$, $\sigma = 1$.

(1.1). Specifically, we determine intervals of the form

$$H(X_t^{(a)}) \pm k\sqrt{V(X_t^{(a)})} = \mathbb{E}[IC(X_t^{(a)})] \pm k\sqrt{\mathrm{Var}[IC(X_t^{(a)})]}, \qquad k = 2, 3 \tag{2.39}$$

for suitable baseline distributions (Weibull, gamma and lognormal). Since closed forms are not available, we illustrate such results with some graphics given in Figure 2.5. For comparison purposes, the relevant parameters are chosen in order that the baseline distributions have unity means.

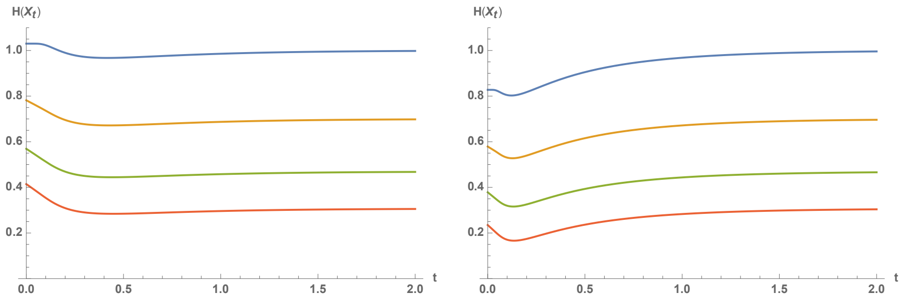


Figure 2.6: Residual entropy for the FPT pdf (1.35), when $y = 1$, $\alpha = 1$, $\nu = 1$ (left), $\nu = 2$ (right), and $\xi = 0, 0.35, 0.7, 1$ (from top to bottom).

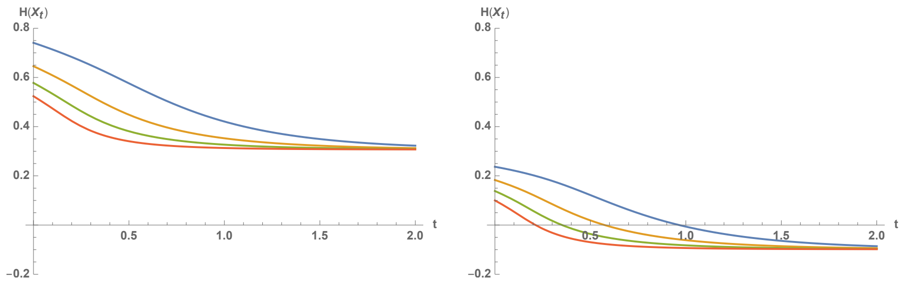


Figure 2.7: Same as Figure 2.6, for $\xi = 1$ (left), $\xi = 2$ (right), and $\nu = 0.15, 0.3, 0.45, 0.6$ (from top to bottom).

2.4.2 First-passage times of an Ornstein–Uhlenbeck jump-diffusion process

Starting from the analysis of the first-passage times of an Ornstein–Uhlenbeck jump-diffusion process (see Section 1.3) we can investigate the residual entropy and residual varentropy associated to the first-passage transition density without catastrophes (1.33) or with catastrophes (1.35). In order to analyse the relevant information content, Figures 2.6 and 2.7 show some instances of the residual entropy related to pdf (1.35), whereas the corresponding residual varentropy is provided in figures 2.8 and 2.9. It is shown that the residual entropy is decreasing in ξ and in ν; moreover it tends to a constant when t grows, such limit being decreasing in ξ and constant in ν. The residual varentropy exhibits a different behavior, since it is decreasing in ξ and is increasing in ν for sufficiently large values of t. Moreover, it tends to an identical limit when t grows. This latter property is confirmed by extensive computations performed for various choices of the parameters.

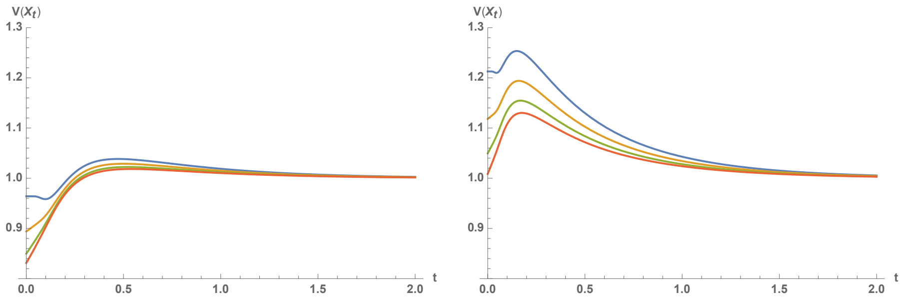


Figure 2.8: Residual varentropy for the same cases of Figure 2.6, with $\xi = 0,\ 0.35,\ 0.7,\ 1$ (from top to bottom).

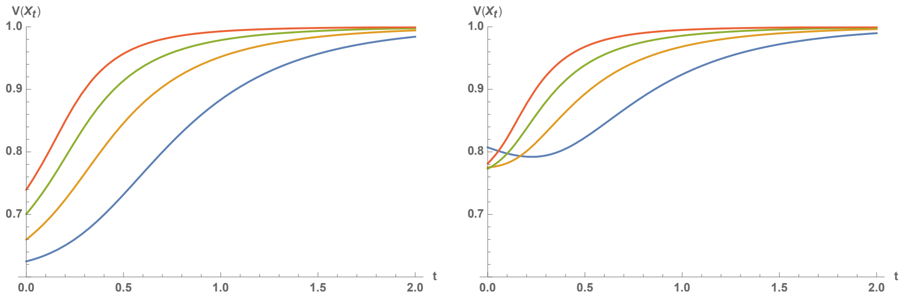


Figure 2.9: Residual varentropy for the same cases of Figure 2.7, with $\nu = 0.15,\ 0.3,\ 0.45,\ 0.6$ (from bottom to top).

2.4.3 Kernel estimation of the residual varentropy

This section is finalized to obtain suitable estimates of the residual varentropy. A classical approach in this area is centered on kernel type estimation. The kernel estimation of the entropy and of other information measures has been treated by many authors (see, for instance, Hall and Morton [37] and Belzunce *et al.* [10]). Along the lines of the previous investigations, we purpose to adopt this criterion to estimate the residual varentropy. Recalling Eqs. (2.8) and (1.14a) it is now convenient to express such measure as follows, for $t \in D$:

$$\begin{aligned} V(X_t) &= \frac{1}{\overline{F}(t)} \int_t^{\infty} f(x) \left[\log f(x)\right]^2 \mathrm{d}x - \left[\frac{1}{\overline{F}(t)} \int_t^{\infty} f(x) \log f(x) \, \mathrm{d}x\right]^2 \\ &= \frac{I_2(t)}{\overline{F}(t)} - \left[\frac{I_1(t)}{\overline{F}(t)}\right]^2, \end{aligned} \tag{2.40}$$

where

$$I_r(t) := \int_t^{\infty} f(x)[\log f(x)]^r \, \mathrm{d}x, \qquad r = 1, 2. \tag{2.41}$$

Hereafter we consider two suitable criteria in order to estimate the quantities given in (2.41). These will be employed to construct suitable estimates of the residual varentropy.

(i) Hybrid estimator

$$\widehat{I}_r^e(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>t\}} \left[\log \widehat{f}_i(X_i)\right]^r, \qquad r = 1,2$$

where $\widehat{f}_i(X_i)$ is the leave-one-out kernel estimator of the pdf $f(x)$ (1.19). Hence, in this case the estimate of the residual varentropy (2.40) is provided by

$$\widehat{V}^e(X_t) = \frac{\widehat{I}_2^e(t)}{\widehat{R}^e(t)} - \left[\frac{\widehat{I}_1^e(t)}{\widehat{R}^e(t)}\right]^2.$$

where $\widehat{R}^e(t)$ is the empirical estimator of the survival function (1.20).

(ii) Kernel estimator

$$\begin{aligned}\widehat{I}_{h,r}(t) &= \int_t^\infty \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h}K\Big(\frac{x-X_i}{h}\Big) \left[\log \widehat{f}_i(X_i)\right]^r \mathrm{d}x \\ &= \frac{1}{n}\sum_{i=1}^{n} S\Big(\frac{t-X_i}{h}\Big)\left[\log \widehat{f}_i(X_i)\right]^r, \qquad r = 1,2.\end{aligned}$$

where $S(x)$ is the survival function associated to $K(x)$ (cf. (1.22)). Consequently, now a kernel estimator of $V(X_t)$ is given by

$$\widehat{V}_h(X_t) = \frac{\widehat{I}_{h,2}(t)}{\widehat{R}_h(t)} - \left[\frac{\widehat{I}_{h,1}(t)}{\widehat{R}_h(t)}\right]^2. \tag{2.42}$$

where $\widehat{R}^h(t)$ is the kernel estimator of the survival function (1.21).

Example 2.6 Let us now consider an application to a set of data consisting in the following 63 observations of the strengths of 1.5 cm glass fibers, originally obtained by workers at the UK National Physical Laboratory (cf. Smith *et al.* [70] or Merovci *et al.* [55]):

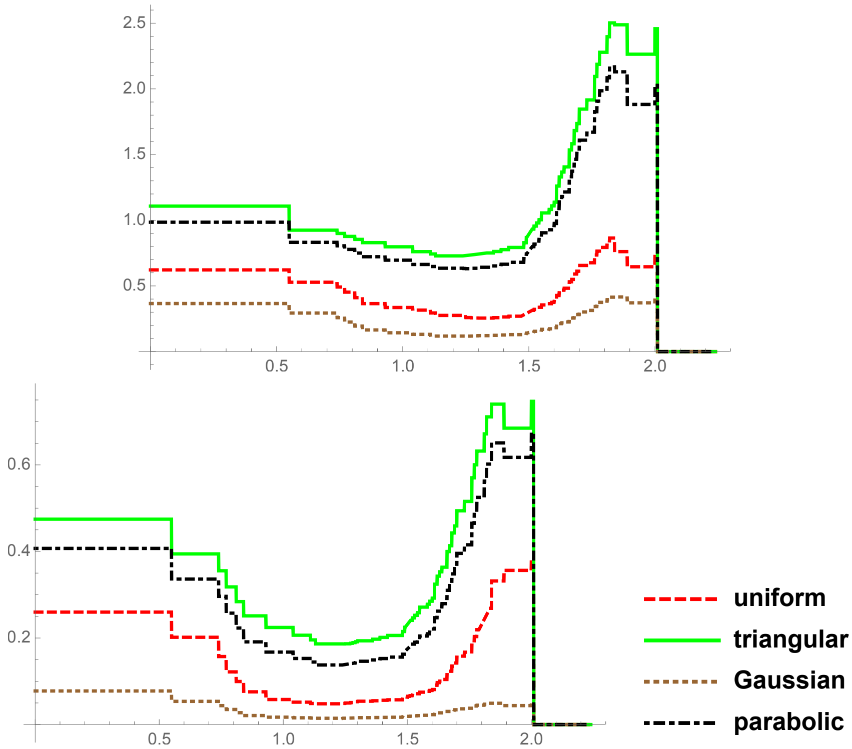


Figure 2.10: Estimates of the residual varentropy obtained from (2.42) for $h = 0.25$ (left) and $h = 0.5$ (right), with the indicated choices of kernel functions.

0.55, 0.74, 0.77, 0.81, 0.84, 0.93, 1.04, 1.11, 1.13, 1.24, 1.25, 1.27, 1.28, 1.29, 1.30, 1.36, 1.39, 1.42, 1.48, 1.48, 1.49, 1.49, 1.50, 1.50, 1.51, 1.52, 1.53, 1.54, 1.55, 1.55, 1.58, 1.59, 1.60, 1.61, 1.61, 1.61, 1.61, 1.62, 1.62, 1.63, 1.64, 1.66, 1.66, 1.66, 1.67, 1.68, 1.68, 1.69, 1.70, 1.70, 1.73, 1.76, 1.76, 1.77, 1.78, 1.81, 1.82, 1.84, 1.84, 1.89, 2.00, 2.01, 2.24.

The kernel estimator of the residual varentropy has been obtained from such data by means of Eq. (2.42), with four different choices of the kernel function. The curves obtained in these cases are shown in Figure 2.10. They have similar bathtub shape, with vertical shifts according to the kernel function adopted in the estimation.

Example 2.6 shows that the shape of $\widehat{V}_h(X_t)$ is sensible to the choice of K and h. A comprehensive investigation on their optimal choice is a suitable object of future studies.

Chapter 3

Stochastic comparisons and connections with entropy and varentropy

In this chapter we explore some properties of differential entropy and varentropy by focusing on their implications in the context of stochastic comparisons. In order to do this in the first part we concentrate our efforts to study the quantiles for an unimodal distribution (Section 2) and we introduce the notion of pdf-related distribution (Sections 3 and 4). In the second part of the chapter, these notions are applied in order to compare distributions induced by probability density function. This investigation will lead us to obtain the determination of the quantiles of the information content (Section 5) that will determine some conditions in order to compare differential entropy and varentropy (Section 6) and residual varentropy (Section 7) of random variable distributions.

3.1 Introduction

Let X be an absolutely continuous random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let the interval $S_X \subseteq \mathbb{R}$ be its support. The results of Sections 1-3 that are analysed in ref. [25] follow from the hypothesis that the pdf $f(x)$ of X satisfies one of the following assumptions:

(i) $f(x)$ is strictly decreasing for all $x \in S_X$, with $S_X = (m, b)$, with $m < b \leq \infty$,

(ii) $f(x)$ is strictly increasing for $x \leq m$ and strictly decreasing for $x \geq m$, with $m \in S_X$.

Henceforth, X is unimodal with mode $m \in S_X$.

3.2 Quantiles for unimodal distributions

In the following, for all $u \in (0, 1)$ we shall denote by x_u the lower quantile of a random variable X, such that $F(x_u) = u$, i.e.

$$x_u = F^{-1}(u), \qquad u \in (0, 1),$$

where

$$F^{-1}(\alpha) := \sup\{x \in S_X \colon F(x) \leq \alpha\}, \qquad \alpha \in (0, 1) \tag{3.1}$$

denotes the right-continuous inverse of the distribution function F. In the following the function (3.1) will be said quantile function of X. Clearly, $\overline{x}_u := x_{1-u}$ is the upper quantile, and $x_{1/2}$ is the median of the distribution of X.

We recall that the distribution of X is symmetric if there exists a constant m such that $F(m - x) = \overline{F}(m + x)$ for all $x \in S_X$, equivalently if $F(x_u) = \overline{F}(\overline{x}_u)$ for all $u \in (0, 1)$. Moreover, if X is absolutely continuous, a sufficient condition for the symmetry is $f(m - x) = f(m + x)$ for all $x \in S_X$, and an equivalent condition is (cf. Lemma 1 of Fashandi and Ahmadi [30]) $f(x_u) = f(\overline{x}_u)$, for almost all $u \in (0, 1)$.

In the following, given a pdf f having support S_X, we define

$$\mathrm{Im}^+(f) := \{y > 0 \colon \exists\, x \in S_X \text{ s.t. } f(x) = y\}.$$

Definition 3.1 Let X be an absolutely continuous random variable having support $S_X = (a, b)$ and a unimodal pdf f with mode m. Let $y \in \mathrm{Im}^+(f)$.

(i) If $f(x)$ is continuous and strictly monotone on (a, b), then the *inverse* of f at y is denoted by

$$l_y := f^{-1}(y). \tag{3.2}$$

(ii) If $f(x)$ is continuous in all the support S_X and is strictly increasing for $x \le m$ and strictly decreasing for $x \ge m$, then the *lower inverse* of f at y is given in (3.2), where in this case f^{-1} denotes the inverse of the restriction of f to $S_X \cap (-\infty, m]$. Moreover, the *upper inverse* of f at y is denoted by

$$u_y := f_2^{-1}(y) \tag{3.3}$$

where in this case f_2^{-1} denotes the inverse of the restriction of f to $S_X \cap [m, \infty)$.

Remark 3.1 If, in the case (ii) of Definition 3.1, f is symmetric in m then

$$u_y = 2m - f^{-1}(y) = 2m - l_y.$$

Remark 3.2 Under the assumptions specified in the case (ii) of Definition 3.1 we have

$$f(u_y) = f(l_y) \quad \text{and} \quad F(l_y) = \overline{F}(u_y).$$

3.2.1 Unimodal residual lifetime distributions and quantiles

Now we generalize the above concepts to the case of residual lifetimes.

We first consider some properties of the residual lifetime when the underlying pdf is unimodal.

Proposition 3.1 *Let X be an absolutely continuous random variable with support $S_X = (0, b)$, for $0 < b \le \infty$, and having a unimodal pdf $f(x)$ with mode m. Let X_t be defined as in (1.1), for $t \in S_X$, and with pdf (1.3).*

(i) If $t < m$, then $f_t(x)$ is unimodal with mode $m_t = m - t$.

(ii) If $t \geq m$, then $f_t(x)$ is decreasing for all $x \in (0, b-t)$.

The proof is straightforward and thus is omitted.

We can now recover the expression of the quantile function of the residual lifetime.

Proposition 3.2 *If X is an absolutely continuous random variable having distribution function $F(x)$ and support $S_X = (0, b)$, for $0 < b \leq \infty$, then the quantile function associated with the residual lifetime X_t, for $t \in S_X$ can be expressed as*

$$F_t^{-1}(p) = F^{-1}(1-(1-p)\overline{F}(t)) - t, \qquad p \in (0,1).$$

Proof. From Equation (1.2), imposing $F_t(x) = p \in (0,1)$ we have $\overline{F}(x+t) = (1-p)\overline{F}(t)$, for $x \in (0, b-t)$. Hence, one has $F(x+t) = 1 - \overline{F}(x+t) = 1-(1-p)\overline{F}(t)$, which gives the thesis.

If the pdf $f(x)$ is symmetrical and unimodal, then the median and the mode of X coincide. However, even in this case the median and the mode of X_t are not generally coincident.

3.3 Pdf-related distributions

The object of the present section is a special type of distribution, named "pdf-related distribution" that is illustrated in ref. [25]. This particular type of random variable is defined by means of a transformation expressed by the pdf of a given baseline random variable.

Definition 3.2 Let X be an absolutely continuous random variable with support $S_X = (a, b)$, and having pdf $f(x)$. Then, $f(X)$ is named as the pdf-related random variable of X. For any $y \in \mathrm{Im}^+(f)$ the distribution function of $f(X)$ is defined as

$$K(y) := \mathbb{P}(f(X) \leq y). \qquad (3.4)$$

In the following proposition, which refers to the distribution function of pdf-related distributions, we use the notation adopted in Eqs. (3.2) and (3.3).

Proposition 3.3 *Let X be an absolutely continuous random variable with support $S_X = (a,b)$, and having pdf $f(x)$ that is continuous in all the support S_X.*

(a) If $f(x)$ is strictly decreasing, then

$$K(y) = \overline{F}(l_y) = \overline{F}(f^{-1}(y)), \qquad y \in \mathrm{Im}^+(f).$$

(b) If $f(x)$ is strictly increasing, then

$$K(y) = F(l_y) = F(f^{-1}(y)), \qquad y \in \mathrm{Im}^+(f).$$

(c) If $f(x)$ is strictly increasing for $x \le m$ and strictly decreasing for $x \ge m$, then

$$K(y) = F(l_y) + \overline{F}(u_y) = F(f^{-1}(y)) + \overline{F}(f_2^{-1}(y)), \qquad y \in \mathrm{Im}^+(f).$$

The proof easily follows from Definition 3.1 and Eq. (3.4). Clearly, in the case (c) of Proposition 3.3 if the pdf $f(x)$ is symmetric then

$$K(y) = 2F(l_y) = 2\overline{F}(u_y), \qquad y \in \mathrm{Im}^+(f). \tag{3.5}$$

Example 3.1 From the case (a) of Proposition 3.3 it is not hard to see that if X has pdf

$$f(x) = \left(1 - \left(1 - \frac{1}{\alpha}\right) x\right)^{1/(\alpha-1)}, \qquad 0 < x < \frac{\alpha}{\alpha - 1},$$

for $\alpha > 1$, then $K(y) = y^\alpha$, for $0 \le y \le 1$.

It is worth pointing out that a pdf-related distribution is not necessarily continuous. Indeed, for instance, if X is uniformly distributed on (a,b), then clearly $f(X)$ is degenerate in $(b-a)^{-1}$.

Hereafter we obtain a characterization of the $(0,1)$-uniform distribution based on pdf-related distributions of exponential and Laplace type.

Proposition 3.4 *Under the assumptions of Proposition 3.3, we have that the pdf-related random variable $f(X)$ is uniform on $(0,1)$ if and only if*
(a) $f(x)=e^{a-x}$ for $x\in(a,\infty)$ and $f(x)=0$ otherwise, when f is strictly decreasing in (a,∞), for $a\in\mathbb{R}$;
(b) $f(x)=e^{x-b}$ for $x\in(-\infty,b)$ and $f(x)=0$ otherwise, when f is strictly increasing in $(-\infty,b)$, for $b\in\mathbb{R}$;
(c) $f(x)=e^{2(x-m)}$ for $x\in(-\infty,m]$ and $f(x)=e^{2(m-x)}$ for $x\in[m,\infty)$, when f is strictly increasing in $(-\infty,m]$ and is strictly decreasing in $[m,\infty)$, for $m\in\mathbb{R}$.

Proof. In the following we will use the relation

$$K(y)=y=f(x). \tag{3.6}$$

If $f(X)$ is uniform in $(0,1)$ it is immediate to conclude from (3.4) that

$$K(y)=\begin{cases}0 & \text{if } y\le 0\\ y & \text{if } 0<y<1\\ 1 & \text{if } y\ge 1\end{cases} \tag{3.7}$$

(a) If $f(x)$ is strictly decreasing for all $x\in(a,\infty)$, applying Proposition 3.3 we have

$$K(y)=\overline{F}\left(l_y\right)=\overline{F}\left(f^{-1}(y)\right)=\overline{F}\left(f^{-1}(f(x))\right)=\overline{F}(x)=1-F(x)$$

and from Equation (3.6)

$$K(y)=f(x)\Longleftrightarrow 1-F(x)=f(x)\Longleftrightarrow F'(x)+F(x)=1. \tag{3.8}$$

Solving the last differential equation in (3.8) with initial condition $F(a)=1$ we obtain

$F(x) = 1 - e^{a-x}$ for $x \in (a, \infty)$ and

$$K(y) = \overline{F}(x) = e^{a-x}, \qquad x \in (a, \infty)$$

whose corresponding pdf is

$$f(x) = \frac{d}{dx}(1 - e^{a-x}) = e^{a-x}, \qquad x \in (a, \infty)$$

From Equations (3.7) and (3.6) trivially follows that outside of the interval $y \in (0, 1)$ that corresponds to the interval $x \in (a, \infty)$ the pdf of $f(X)$, as the pdf of X vanishes.

(b) If $f(x)$ is strictly increasing for all $x \in (a, \infty)$, applying Proposition 3.3 we have

$$K(y) = F(l_y) = F\left(f^{-1}(y)\right) = F\left(f^{-1}(f(x))\right) = F(x)$$

and from Equation (3.6)

$$K(y) = f(x) \iff F(x) = f(x) \iff F'(x) - F(x) = 0. \tag{3.9}$$

Solving the last differential equation in (3.9) with initial condition $F(b) = 0$ we obtain $F(x) = e^{x-b}$ and

$$K(y) = F(x) = e^{x-b}, \qquad x \in (-\infty, b)$$

whose corresponding pdf is

$$f(x) = \frac{d}{dx}(e^{x-b}) = e^{x-b}, \qquad x \in (-\infty, b)$$

and $f(x)$ vanishes outside the interval $x \in (-\infty, b)$.

(c) In this case $f(x)$ is symmetric and unimodal, so we can apply (3.5) and we obtain

$$K(y) = 2F(l_y) = 2F\left(f^{-1}(y)\right) = 2F\left(f^{-1}(f(x))\right) = 2F(x) \qquad x \in (-\infty, m) \tag{3.10}$$

and

$$\begin{aligned} K(y) &= 2F\left(u_y\right) = 2F\left(f_2^{-1}(y)\right) \\ &= 2\overline{F}\left(f_2^{-1}(f(x))\right) = 2(1-F(x)) = 2-2F(x). \qquad x \in (m,\infty). \end{aligned} \tag{3.11}$$

Solving differential equations (3.10) and (3.11) with boundary condition $F(m) = \frac{1}{2}$ we obtain, in a similar way as for cases (a) and (b), respectively, that $f(x) = e^{2(x-m)}$ for $x \in (-\infty, m]$ and $f(x) = e^{2(m-x)}$ for $x \in [m, \infty)$.

A characterization of the power distribution is given in the following proposition. Let us recall that a random variable with power distribution in $(0,1)$ has the following cdf

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^\alpha & \text{if } 0 < x < 1, \alpha > 0 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Proposition 3.5 *Let X be an absolutely continuous random variable having support $S_X = (a,b)$ and pdf $f(x)$ continuous over all the support S_X. The pdf-related variable $f(X)$ has power distribution in $(0,1)$ if and only if*
(a) $f(x) = \sqrt[\alpha-1]{\frac{\alpha-1}{\alpha}(a-x)+1}$ for $x \in (a,\infty)$ and $f(x) = 0$ otherwise, when f is strictly decreasing in (a,∞), for $a \in \mathbb{R}$,
(b) $f(x) = \sqrt[\alpha-1]{\frac{\alpha-1}{\alpha}(x-b)+1}$ for $x \in (-\infty,b)$ and $f(x) = 0$ otherwise, when f is strictly increasing in $(-\infty,b)$, for $b \in \mathbb{R}$;
(c) $f(x) = \begin{cases} \sqrt[\alpha-1]{2\frac{\alpha-1}{\alpha}(x-m)+1} & , x \in (-\infty, m] \\ \sqrt[\alpha-1]{2\frac{\alpha-1}{\alpha}(m-x)+1} & , x \in [m, +\infty) \end{cases}$ when f is strictly increasing in $(-\infty,m]$ and strictly decreasing in $[m,\infty)$, per $m \in \mathbb{R}$ with $0 < \alpha < 1$.

Proof. The proof follows in a similar way as the proof of Proposition 3.4. The distribution function of $f(X)$ is given straightforwardly applying (3.4)

$$K(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ y^{\alpha} & \text{se } 0 < y < 1, \alpha > 0 \\ 1 & \text{se } y \geq 1 \end{cases} \tag{3.12}$$

Applying Proposition 3.3 for the interval $y \in (0,1)$ we can determine the following differential equations according to the monotonicity of the function $f(x)$.

(a) $f(x)$ is decreasing in (a, ∞), then applying Proposition 3.3 and Equation (3.12) the following chain of equivalences are given

$$K(y) = y^{\alpha} = (f(x))^{\alpha} \Leftrightarrow 1 - F(x) = (f(x))^{\alpha} \Leftrightarrow 1 - F(x) = F'(x)^{\alpha},\ x \in (a, \infty). \tag{3.13}$$

Solving the last differential equation in (3.13) with initial condition $F(a) = 0$ we get

$$F(x) = 1 - \left[\frac{\alpha - 1}{\alpha}(a - x) + 1\right]^{\frac{\alpha}{\alpha-1}} \qquad x \in (a, \infty). \tag{3.14}$$

It can be proved straightforwardly (but the proof is here omitted) that (3.14) is a distribution function for $\alpha \in (0,1)$ and the pdf is given by

$$f(x) = F'(x) = \begin{cases} \sqrt[\alpha-1]{\frac{\alpha-1}{\alpha}(x - b) + 1} & , x \in (-\infty, b) \\ 0 & , \text{ otherwise} \end{cases} , 0 < \alpha < 1$$

(b) $f(x)$ is increasing in $(-\infty, b)$, then applying again Proposition 3.3 and Equation (3.12) the following chain of equivalences are given

$$K(y) = y^{\alpha} = (f(x))^{\alpha} \Leftrightarrow F(x) = (f(x))^{\alpha} \Leftrightarrow F(x) = F'(x)^{\alpha}, \qquad x \in (-\infty, b). \tag{3.15}$$

Solving the last differential equation (3.15) with initial condition $F(a) = 0$ we get

$$F(x) = \left[\frac{\alpha - 1}{\alpha}(x - b) + 1\right]^{\frac{\alpha}{\alpha-1}}, \qquad x \in (-\infty, b). \tag{3.16}$$

Also in this second case Equation (3.16) is a distribution function for $\alpha \in (0,1)$ and the pdf is given by

$$f(x) = F'(x) \begin{cases} \sqrt[\alpha-1]{\frac{\alpha-1}{\alpha}(x-b)+1} & , x \in (-\infty, b) \\ 0 & , \text{ altrimenti} \end{cases} , 0 < \alpha < 1$$

(c) $f(x)$ is increasing in $(-\infty, m)$ and decreasing in $(m, +\infty)$. Accounting for the symmetry and unimodality of the pdf, with a similar strategy as in cases (a) and (b) we have that the cdf and the pdf are respectively

$$F(x) = \left[2^{\frac{1}{\alpha}-1} \left(2\frac{\alpha-1}{\alpha}(x-m)+1 \right) \right]^{\frac{\alpha}{\alpha-1}} \qquad x \in (-\infty, m) \tag{3.17}$$

and

$$F(x) = 1 - \frac{1}{2} \left[2\frac{\alpha-1}{\alpha}(m-x)+1 \right]^{\frac{\alpha}{\alpha-1}} \qquad x \in (m, +\infty) \tag{3.18}$$

are distribution functions for $\alpha \in (0,1)$ and

$$f(x) = F'(x) = \begin{cases} \sqrt[\alpha-1]{2\frac{\alpha-1}{\alpha}(x-m)+1} & , x \in (-\infty, m] \\ \sqrt[\alpha-1]{2\frac{\alpha-1}{\alpha}(m-x)+1} & , x \in [m, +\infty) \end{cases} , 0 < \alpha < 1 \, .$$

In the following, when the pdf f has support $S_X = (a, b)$, we adopt the following notation: $f(a) = \lim_{x \to a^+} f(x)$ and $f(b) = \lim_{x \to b^-} f(x)$.

Hereafter we determine the cdf of the pdf-related random variable of the residual lifetime defined in (1.1), denoted as

$$K_t(y) := \mathbb{P}(f_t(X_t) \le y) = \mathbb{P}(f(X) \le y\overline{F}(t) \,|\, X > t), \tag{3.19}$$

with the last identity due to (1.3).

Proposition 3.6 *Let X be an absolutely continuous random variable defined on a support $S_X = (0, b)$.*

(a) If the pdf f is continuous and strictly decreasing in (t_0, b) for a given $t_0 \in (0, b)$, then for all $t \in (t_0, b)$ one has

$$K_t(y) = \frac{\overline{F}(f^{-1}(y\overline{F}(t)))}{\overline{F}(t)}, \qquad y \in \mathrm{Im}^+(f_t) = \left(\frac{f(b)}{\overline{F}(t)}, \lambda(t)\right), \tag{3.20}$$

where f^{-1} denotes the inverse of the restriction to (t_0, b) of f.

(b) If the pdf f is continuous and strictly increasing in (t_0, b) for a given $t_0 \in (0, b)$, then for all $t \in (t_0, b)$ one has

$$K_t(y) = 1 - \frac{\overline{F}(f^{-1}(y\overline{F}(t)))}{\overline{F}(t)}, \qquad y \in \mathrm{Im}^+(f_t) = \left(\lambda(t), \frac{f(b)}{\overline{F}(t)}\right), \tag{3.21}$$

where f^{-1} denotes the inverse of the restriction to (t_0, b) of f.

(c) Let the pdf f be continuous in S_X, unimodal and symmetric, strictly increasing for $x \in (0, m]$ and strictly decreasing for $x \in [m, b)$, with $m = b/2$. Then, for all $t \in (0, m]$ one has

$$K_t(y) = \begin{cases} \dfrac{F(f^{-1}(y\overline{F}(t)))}{\overline{F}(t)}, & y \in \left(\frac{f(0)}{\overline{F}(t)}, \lambda(t)\right) \\[2ex] \dfrac{2F(f^{-1}(y\overline{F}(t))) - F(t)}{\overline{F}(t)}, & y \in \left[\lambda(t), \frac{f(m)}{\overline{F}(t)}\right) \end{cases} \tag{3.22}$$

where f^{-1} is the inverse of the restriction to $(0, m]$ of f.

Proof. In the case (a), from (3.19) for all $t \in (t_0, b)$ we have

$$K_t(y) = \mathbb{P}(X \geq f^{-1}(y\overline{F}(t)) \,|\, X > t), \qquad y \in \mathrm{Im}^+(f_t).$$

Eq. (3.20) then follows from the assumption that f is strictly decreasing in (t_0, b). In the case (b), from (3.19) one similarly has, for all $t \in (t_0, b)$,

$$K_t(y) = \frac{\mathbb{P}(t < X \leq f^{-1}(y\overline{F}(t))}{\overline{F}(t)}, \qquad y \in \mathrm{Im}^+(f_t),$$

this giving Eq. (3.21) due to the assumption that f is strictly increasing in (t_0, b). In the case (c), when $y \in \left(\frac{f(0)}{\overline{F}(t)}, \lambda(t)\right)$ the function $K_t(y)$ can be obtained similarly as in case (a), whereas when $y \in \left[\lambda(t), \frac{f(m)}{\overline{F}(t)}\right)$ we have

$$K_t(y) = \mathbb{P}(t < X \leq f^{-1}(y\overline{F}(t)) \,|\, X > t) + \mathbb{P}(b - f^{-1}(y\overline{F}(t)) \leq X \leq b \,|\, X > t).$$

The expression given in (3.22) thus follows by noting that $\overline{F}(b-x) = F(x)$ for all $0 \leq x \leq b$.

Clearly, the cdf given in (3.22) is continuous in $y = \lambda(t)$, with

$$K_t(\lambda(t)) = \frac{F(t)}{\overline{F}(t)}, \qquad t \in (0, m],$$

where the right-hand-side is the odds function of X evaluated at $t \leq m$.

Example 3.2 Let X be a Pareto-type random variable having pdf and cdf given respectively by

$$f(x) = \frac{1}{(1+x)^2}, \qquad x \in (0, +\infty), \qquad F(x) = \frac{x}{1+x}, \qquad x \in [0, +\infty).$$

Since f is strictly decreasing for all $x \in (0, +\infty)$, with inverse $f^{-1}(y) = y^{-1/2} - 1$, $y \in (0, 1)$, from (3.20) we have that, for all $t > 0$

$$K_t(y) = (y(1+t))^{1/2}, \qquad y \in \left(0, \frac{1}{1+t}\right).$$

Example 3.3 Let X be a random variable having support $S_X = (0, 2)$, with pdf and cdf given respectively by

$$f(x) = \frac{b}{3} + \frac{1}{2} - b(x-1)^2, \qquad F(x) = \left(\frac{b}{3} + \frac{1}{2}\right)x - \frac{b}{3}(x-1)^3 - \frac{b}{3}, \qquad x \in S_X,$$

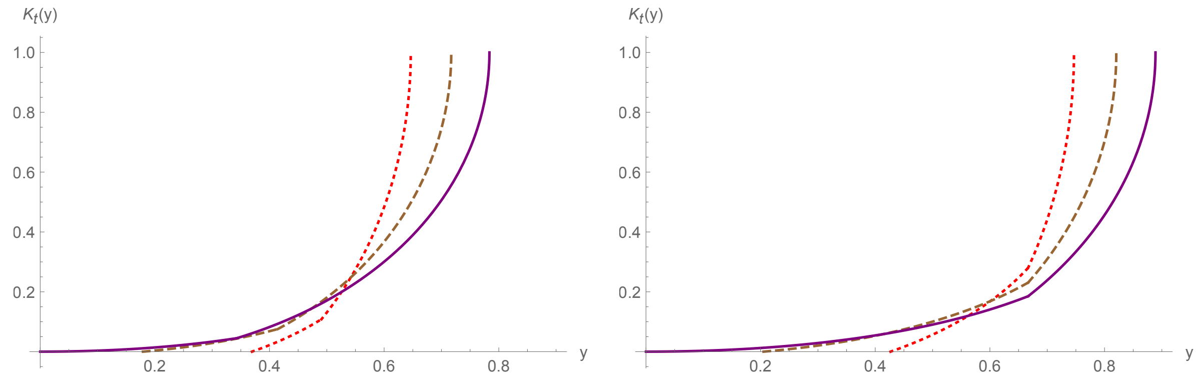


Figure 3.1: The cdf $K_t(y)$ of Example 3.3 for $t = 0.25$ (left) and for $t = 0.5$ (right), with $b = 0.25$ (dotted), $b = 0.5$ (dashed) and $b = 0.75$ (full).

with $b \in \left(0, \frac{3}{4}\right]$. It satisfies the assumptions of Case (c) of Proposition 3.6, with mode $m = 1$ and inverse

$$f^{-1}(y) = 1 - \sqrt{\frac{1}{b}\left(\frac{b}{3} + \frac{1}{2} - y\right)}, \qquad y \in \left(\frac{1}{2} - \frac{2b}{3}, \frac{1}{2} + \frac{b}{3}\right).$$

The cdf $K_t(y)$ can be easily obtained from (3.22). We omit the details for brevity, and show some instances in Figure 3.1.

A further case of interest is considered hereafter. For a given absolutely continuous random variable X with support $S_X = (a, b)$, and having pdf f, let us now analyze the survival function of $f(t + X_t)$. For all $y \in \mathrm{Im}^+(f)$ and $t \in (a, b)$, recalling Eq. (1.1) we have

$$\overline{G}_t(y) := \mathbb{P}(f(t + X_t) > y) = \mathbb{P}\left(f(X) > y \,|\, X > t\right). \tag{3.23}$$

Proposition 3.7 *Let X be an absolutely continuous random variable with support $S_X = (a, b)$, and let its pdf f be continuous and unimodal with mode m, such that $f(x)$ is strictly increasing for $x \leq m$ and strictly decreasing for $x \geq m$. Let $y \in \mathrm{Im}^+(f)$; denoting by F the cdf of X we have* *(a) for $t \in (a, m]$ and $f(a) < f(b) < f(t)$*

$$\overline{G}_t(y) = \begin{cases} 1 & \textit{if} \;\; f(a) < y < f(b) \\ \dfrac{F\left(u_y\right) - F(t)}{\overline{F}(t)} & \textit{if} \;\; f(b) < y \leq f(t) \\ \dfrac{F\left(u_y\right) - F\left(l_y\right)}{\overline{F}(t)} & \textit{if} \;\; f(t) \leq y \leq f(m) \end{cases}$$

(b) for $t \in (a, m]$ and $f(a) < f(t) < f(b)$

$$\overline{G}_t(y) = \begin{cases} 1 & \textit{if} \;\; f(a) < y \leq f(t) \\ \dfrac{1 - F(l_y)}{\overline{F}(t)} & \textit{if} \;\; f(t) \leq y < f(b) \\ \dfrac{F\left(u_y\right) - F(t)}{\overline{F}(t)} & \textit{if} \;\; f(b) < y \leq f(m); \end{cases}$$

(c) for $t \in (a, m]$ and $f(b) \leq f(a)$

$$\overline{G}_t(y) = \begin{cases} \dfrac{F(u_y) - F(t)}{\overline{F}(t)} & \textit{if} \;\; f(b) < y < f(a) \quad \textit{or} \quad f(a) < y \leq f(t) \\ \dfrac{F(u_y) - F(l_y)}{\overline{F}(t)} & \textit{if} \;\; f(t) \leq y \leq f(m); \end{cases}$$

(d) for $t \in [m, b)$

$$\overline{G}_t(y) = \frac{F(u_y) - F(l_y)}{\overline{F}(t)} \qquad \textit{if} \;\; f(b) < y \leq f(t).$$

Proof. According to Eq. (3.23) we can consider the following cases.

(a) for $t \in (a, m]$ and $f(a) < f(b) < f(t)$:

$$\overline{G}_t(y) = \mathbb{P}\left(t \leq X < b \,|\, X > t\right) = 1, \qquad y \in (f(a), f(b)),$$

$$\overline{G}_t(y) = \mathbb{P}\left(t \leq X \leq u_y \,|\, X > t\right) = \frac{F(u_y) - F(t)}{\overline{F}(t)}, \qquad y \in (f(b), f(t)],$$

$$\overline{G}_t(y) = \mathbb{P}\left(l_y \leq X \leq u_y \,|\, X > t\right) = \frac{F(u_y) - F(l_y)}{\overline{F}(t)}, \qquad y \in [f(t), f(m)];$$

(b) for $t \in (a, m]$ and $f(a) < f(t) < f(b)$:

$$\overline{G}_t(y) = \mathbb{P}\,(t \le X < b \,|\, X > t) = 1, \qquad y \in (f(a), f(b)),$$

$$\overline{G}_t(y) = \mathbb{P}\,(l_y \le X < b \,|\, X > t) = \frac{\overline{F}(l_y)}{\overline{F}(t)}, \qquad y \in (f(b), f(t)],$$

$$\overline{G}_t(y) = \mathbb{P}\,(l_y \le X \le u_y \,|\, X > t) = \frac{F(u_y) - F(l_y)}{\overline{F}(t)}, \qquad y \in [f(t), f(m)];$$

(c) for $t \in (a, m]$ and $f(b) \le f(a)$:

$$\overline{G}_t(y) = \mathbb{P}\,(t \le X \le u_y \,|\, X > t) = \frac{F(u_y) - F(t)}{\overline{F}(t)}, \qquad y \in (f(b), f(a)) \cup (f(a), f(t)];$$

$$\overline{G}_t(y) = \mathbb{P}\,(l_y \le X \le u_y \,|\, X > t) = \frac{F(u_y) - F(l_y)}{\overline{F}(t)}, \qquad y \in [f(t), f(m)];$$

(d) for $t \in [m, b)$:

$$\overline{G}_t(y) = \mathbb{P}\,(t \le X \le u_y \,|\, X > t) = \frac{F(u_y) - F(t)}{\overline{F}(t)}, \qquad y \in (f(b), f(t)].$$

The proof is thus completed.

Remark 3.3 With reference to Proposition 3.7, recalling Definition 3.1 we have

$$\frac{d}{dy}F(u_y) = \frac{y}{f'(u_y)}, \qquad \frac{d}{dy}F(l_y) = \frac{y}{f'(l_y)},$$

so that we can obtain the pdf of $f(t + X_t)$, denoted by $g_t(y) := -\frac{\partial \overline{G}_t(y)}{\partial y}$, as follows:

- for $t < m$ and $f(a) \le f(b) < f(t)$

$$g_t(y) = \begin{cases} -\dfrac{y}{\overline{F}(t)\, f'(u_y)} & \text{if } \; f(b) < y \le f(t) \\[2ex] -\dfrac{y}{\overline{F}(t)} \left(\dfrac{1}{f'(u_y)} - \dfrac{1}{f'(l_y)} \right) & \text{if } \; f(t) < y \le f(m), \end{cases} \tag{3.24}$$

- for $t < m$ and $f(a) < f(t) < f(b)$

$$g_t(y) = \begin{cases} \dfrac{y}{\overline{F}(t)\, f'(l_y)} & \text{if } \ f(b) < y \leq f(t) \\ -\dfrac{y}{\overline{F}(t)} \left(\dfrac{1}{f'(u_y)} - \dfrac{1}{f'(l_y)} \right) & \text{if } \ f(t) \leq y \leq f(m), \end{cases} \tag{3.25}$$

- for $t < m$ and $f(b) \leq f(a)$

$$g_t(y) = \begin{cases} -\dfrac{y}{\overline{F}(t)\, f'(u_y)} & \text{if } \ f(b) < y < f(a) \ \vee \ f(a) < y \leq f(t) \\ -\dfrac{y}{\overline{F}(t)} \left(\dfrac{1}{f'(u_y)} - \dfrac{1}{f'(l_y)} \right) & \text{if } \ f(t) \leq y \leq f(m), \end{cases} \tag{3.26}$$

- for $t \geq m$

$$g_t(y) = -\frac{y}{\overline{F}(t)\, f'(u_y)} \qquad \text{if } \ f(b) < y \leq f(t).$$

We note that $f'(u_y) < 0 < f'(l_y)$. Moreover, if X has a symmetric pdf, i.e. $f(u_y) = f(l_y)$ then $f'(u_y) = -f'(l_y)$, so that the last expressions in the right-hand-sides of Eqs. (3.24)-(3.26) are simplified.

Hereafter we discuss the expression of the survival function (3.23) and the corresponding pdf as a function of t.

Remark 3.4 If $f(a) < f(b)$ then we have the following cases:

(a) for $y \in (f(a), f(b))$

$$\overline{G}_t(y) = \frac{\overline{F}(l_y)}{\overline{F}(t)}, \qquad g_t(y) = \frac{y}{f'(l_y)\overline{F}(t)}, \qquad a < t \leq l_y$$

(b) for $y \in (f(b), f(m))$

$$\overline{G}_t(y) = \begin{cases} \dfrac{F(u_y) - F(l_y)}{\overline{F}(t)}, & a < t \leq l_y \\ \dfrac{F(u_y) - F(t)}{\overline{F}(t)}, & l_y \leq t \leq u_y \end{cases}$$

$$g_t(y) = \begin{cases} -\dfrac{y}{\overline{F}(t)}\Big(\dfrac{1}{f'(u_y)} - \dfrac{1}{f'(l_y)}\Big), & a < t \leq l_y \\ -\dfrac{y}{f'(l_y)\overline{F}(t)}, & l_y \leq t \leq u_y. \end{cases}$$

If $f(b) \leq f(a)$ then we have:

(a) when $f(b) < f(a)$, for $y \in (f(b), f(a))$

$$\overline{G}_t(y) = \frac{F(u_y) - F(t)}{\overline{F}(t)}, \qquad g_t(y) := -\frac{y}{f'(u_y)\overline{F}(t)}, \qquad a < t \leq u_y$$

(b) for $y \in (f(a), f(m)]$

$$\overline{G}_t(y) = \begin{cases} \dfrac{F(u_y) - F(l_y)}{\overline{F}(t)}, & a < t \leq l_y \\ \dfrac{F(u_y) - F(t)}{\overline{F}(t)}, & l_y \leq t \leq u_y \end{cases}$$

$$g_t(y) = \begin{cases} -\dfrac{y}{\overline{F}(t)}\Big(\dfrac{1}{f'(u_y)} - \dfrac{1}{f'(l_y)}\Big), & a < t \leq l_y \\ -\dfrac{y}{f'(u_y)\overline{F}(t)}, & l_y \leq t \leq u_y. \end{cases}$$

As in Remark 3.3, some expressions are simplified in the symmetric case for $f'(u_y) = -f'(l_y)$.

3.3.1 Determination of quantiles

Generally the determination of quantiles for a pdf-related distribution is not an easy task. Nevertheless, in some cases this problem can be solved, as reported hereafter.

Proposition 3.8 *Let X be an absolutely continuous random variable with support $S_X = (0, b)$ and having a symmetric and unimodal pdf $f(x)$. Then the lower quantile and the upper quantile of $f(X)$ are given respectively by*

$$\xi_u = f\left(F^{-1}\left(\frac{u}{2}\right)\right), \qquad \xi_{1-u} = f\left(F^{-1}\left(1-\frac{u}{2}\right)\right), \qquad u \in (0,1). \tag{3.27}$$

Proof. Recalling (3.4), from Proposition 3.3 we have

$$K(y) = F(l_y) + \overline{F}(u_y), \qquad y \in \mathrm{Im}^+(f).$$

For the symmetry, we can apply Remark 3.2 so that

$$K(y) = 2F(l_y) = 2\overline{F}(u_y).$$

Hence, applying Eqs. (3.2) and (3.3) we obtain the lower and upper quantiles of $f(X)$ given in (3.27).

The following result refers to the cdf of the the pdf-related random variable of the residual lifetime (1.1), that is given in (3.19).

Proposition 3.9 *Let X be an absolutely continuous random variable with support $S_X = (0, b)$ and having a strictly monotone pdf $f(x)$.*

(a) If $f(x)$ is strictly decreasing, then for all $t \in S_X$ and $0 < p < 1$,

$$K_t(y) = 1 - p, \qquad y \in \mathrm{Im}^+(f_t)$$

if and only if

$$y = f_t\left(F_t^{-1}(p)\right) = \frac{1}{\overline{F}(t)} f\left(F^{-1}\left(1-(1-p)\overline{F}(t)\right)\right).$$

(b) If $f(x)$ is strictly increasing, then for all $t \in S_X$ and $0 < q < 1$,

$$K_t(y) = q, \qquad y \in S_X$$

if and only if

$$y = f_t\left(F_t^{-1}(q)\right) = \frac{1}{\overline{F}(t)} f\left(F^{-1}\left(1-(1-q)\overline{F}(t)\right)\right).$$

Proof. After straightforward calculations, in both cases the result follows making use of Proposition 3.6 and Proposition 3.2.

With reference to the cdf (3.4), a straightforward consequence of Proposition 3.9 is given hereafter.

Corollary 3.1 *Under the assumptions of Proposition 3.9,*
(a) if $f(x)$ is strictly decreasing, then for $0 < p < 1$,

$$K(y) = 1-p \qquad \textit{if and only if} \qquad y = f\left(F^{-1}(p)\right);$$

(b) if $f(x)$ is strictly increasing, then for $0 < q < 1$,

$$K(y) = q \qquad \textit{if and only if} \qquad y = f\left(F^{-1}(q)\right).$$

3.4 Results based on stochastic orders

In this section some results involving stochastic orders and quantiles are reported. This results are part of the analysis made in ref. [25] that is finalized to obtain stochastic comparisons between pdf-related random variables. For a comprehensive review on the background we refer to Shaked and Shantikumar [67]. We first recall notions that will be used in the following to establish an ordering between the varentropy of different random variables. Hereafter, the notation F^{-1} refers to the right-continuous inverse introduced in (3.1). Moreover, $Me = x_{1/2}$ is the median of the distribution of X.

Definition 3.3 Let X and Y be random variables having cdfs $F(x)$ and $G(x)$, respectively. Then,

(i) X is stochastically smaller than Y, and write $X \le_{st} Y$, if $F(x) \ge G(x)$ for all $x \in \mathbb{R}$, or equivalently if

$$F^{-1}(u) \le G^{-1}(u), \qquad \text{for all } \; u \in (0,1);$$

(ii) X is smaller than Y in the dispersive order, and write $X \le_{disp} Y$, if

$$F^{-1}(\beta) - F^{-1}(\alpha) \le G^{-1}(\beta) - G^{-1}(\alpha), \qquad \text{for all } \; 0 < \alpha \le \beta < 1.$$

(iii) X is smaller than Y in the convex transform order, with X and Y nonegative, and write $X \le_c Y$, if $G^{-1}(F(x))$ is convex in x on the support of $F(x)$;

(iv) X is smaller than Y in the star order, with X and Y nonegative, and write $X \le_* Y$, if G^{-1} is starshaped in x, or equivalently

$$\frac{G^{-1}(F(x))}{x} \qquad \text{is increasing in } x \text{ on the support of } F(x);$$

(v) X is smaller than Y in kurtosis order, and write $X \le_k Y$, if X and Y symmetric and $G^{-1}(F(x))$ is concave for all $x < Me$, or, equivalently, $G^{-1}(F(x))$ is convex for all $x > Me$ (see, for example, Arriaza *et al.* [4]).

3.4.1 Mapping of quantiles

Given two random variables X and Y having cdfs $F(x)$ and $G(x)$, respectively, let us define the mapping function

$$\phi(x) := G^{-1}(F(x)). \tag{3.28}$$

Clearly, this well-known function maps the quantiles of X to the quantile of Y, since $Y =_{st} \phi(X)$. The following result follows from Theorem 1.A.17 of [67].

Remark 3.5 The random variables X and Y satisfy $X \leq_{st} Y$ if and only if, for all x in the support of X, we have

$$\phi(x) \geq x.$$

The following result is related to the dispersive order, and involves the function (3.28).

Remark 3.6 For continuous random variables X and Y we have that $X \leq_{\text{disp}} Y$ if and only if $Y =_{\text{st}} \psi(X)$ for some function ψ which satisfies (cf. Section 3.B.1 of [67])

$$\psi(x_2) - \psi(x_1) \geq x_2 - x_1 \quad \text{whenever } x_1 \leq x_2. \tag{3.29}$$

It is easy to verify that if $\psi(x)$ is equal to the function given in (3.28), then condition (3.29) is satisfied.

3.4.2 Stochastic orders and convexity

Let us now recall a characterization of the convex transform order.

Remark 3.7 Let X and Y be nonnegative random variables with cdfs $F(x)$ and $G(x)$, and pdfs $f(x)$ and $g(x)$, respectively. Then, (cf. Section 4.B.2 of [67])

$$X \leq_c Y \quad \Longleftrightarrow \quad \frac{f(F^{-1}(p))}{g(G^{-1}(p))} \quad \text{is increasing for all } \ p \in (0,1).$$

We also recall two results involving the star order.

Remark 3.8 Let X and Y be nonnegative random variables with cdfs $F(x)$ and $G(x)$, respectively. Then, (cf. Section 4.B.1 of [67])

$$X \leq_* Y \quad \Longleftrightarrow \quad \frac{G^{-1}(p)}{F^{-1}(p)} \quad \text{is increasing in } p \in (0,1).$$

Remark 3.9 Let X and Y be nonnegative random variables. Then, (cf. Theorem 4.B.1 of [67])

$$X \leq_* Y \quad \Longleftrightarrow \quad \log X \leq_{disp} \log Y.$$

We are now able perform the comparison of pdf-related distributions as defined in Section 3.3 in terms of the star order.

Theorem 3.1 *Let X and Y be absolutely continuous random variables having pdfs $f(x)$ and $g(x)$, respectively. Then,*

$$f(X) \leq_* g(Y)$$

if and only if

$$\frac{\phi(x)}{x} \quad \text{is increasing in } x \quad \text{for all } x \text{ in the support of } f(X),$$

where

$$\phi(x) := S^{-1}(K(x)),$$

for

$$K(x) = \mathbb{P}(f(X) \leq x) \qquad \text{and} \qquad S(x) = \mathbb{P}(g(Y) \leq x). \tag{3.30}$$

Proof. The proof is a consequence of the point (iv) of Definition 3.3 applied to the random variables $f(X)$ and $g(Y)$.

From the point (iv) of Definition 3.3 one has that both relations $U \leq_* V$ and $V \leq_* U$ hold simultaneously, and in the following we shall write $U =_* V$, if the random variables U and V have proportional quantile functions. In this case they belong to the same scale family of distributions (see, for instance, Section 4.1 of Di Crescenzo *et al.* [23]). Let us now investigate this relation for the pdf-related distributions connected by affine transformations.

Theorem 3.2 *Let X be an absolutely continuous random variable with pdf $f(x)$, and let $Y = aX + b$, $a > 0$, have pdf $g(x)$. Then, one has*

$$f(X) =_* g(Y). \tag{3.31}$$

Proof. For the cdfs (3.30), recalling that $g(x) = \frac{1}{a} f\left(\frac{x-b}{a}\right)$, for all x in the support of Y we have

$$S(x) = \mathbb{P}\left(\frac{1}{a} f\left(\frac{Y-b}{a}\right) \leq x\right) = K(ax).$$

Hence, one has

$$S^{-1}(p) = \frac{1}{a} K^{-1}(p), \qquad \text{for all } p \in (0,1),$$

so that

$$\frac{\phi(x)}{x} = \frac{S^{-1}(K(x))}{x} = \frac{1}{a} \qquad \text{for all } x \text{ in the support of } f(X).$$

This implies relation (3.31) by Theorem 3.1.

The notion of rearrangement of a function has played a key role in many inequalities in literature. It is described in the seminal book of Hardy *et al.* [40] and has been studied in the context of entropy, randomness, majorization and dispersion in Hickey [41], [42] and Fernández-Ponce and Suárez-Llorens [31], among other authors. We next recall the decreasing rearrangement of a density function and some of its important properties.

Definition 3.4 Let X be an absolutely continuous random variable having a pdf $f(x)$. The decreasing rearrangement of $f(x)$, denoted by $f^*(x)$, is given by

$$f^*(x) = \sup\{c : m(c) > x\}, \qquad x > 0,$$

where $m(c) = \mu\{t : f(t) > c\}$, μ denoting the Lebesgue measure.

It is well known, see [39], that the decreasing rearrangement satisfies the following identity (see, Hardy *et al.* [39])

$$\int_0^{\infty} f^*(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

Then f^* can be considered as a pdf of a particular random variable that we will denote by X^*.

The following implication is well-known (see Hardy *et al.* [39], Theorems 9 and 10). Let X and Y be absolutely continuous random variables having pdf f and g, respectively. Then

$$\int_0^t f^*(x)dx \geq \int_0^t g^*(x)dx, \quad \forall t > 0 \iff \int_{-\infty}^{\infty} u(f(x))dx \leq \int_{-\infty}^{\infty} u(g(x))dx, \qquad (3.32)$$

for all continuous and concave functions u. From Definition 3.3, the expressions in (3.32) are also equivalent to $X^* \leq_{st} Y^*$. It is worth mentioning that $X^* \leq_{st} Y^*$ is interpreted as continuous majorisation in Hickey [41].

It seems that $X^* \leq_{st} Y^*$ implies a kind of stochastic comparison between $f(X)$ and $g(Y)$. Rewriting the right side of expression (3.32), we obtain that $X^* \leq_{st} Y^*$ is equivalent to

$$\mathbb{E}\left[\frac{u(f(X))}{f(X)}\right] \leq \mathbb{E}\left[\frac{u(g(Y))}{g(Y)}\right]$$

for all continuous and concave functions u. The latter relation provides a class of measures of entropy. Indeed, as it is interpreted in Hickey [41], just taking $u(x) = -x\log(x)$, one has that if $X^* \leq_{st} Y^*$ then $H(X) \leq H(Y)$, due to (4).

Hence, it is worth mentioning that the rearrangement captures in some sense the degree of randomness in a density function. Now we wonder if the rearrangement is also related with the concept of varentropy. From now on, we will assume that the density functions have no flat zones or, equivalently, the rearrangements are strictly decreasing. Next we provide a result that we will use later on.

Theorem 3.3 *Let X and Y be absolutely continuous random variables having pdf f and* *g, respectively, with no flat zones. Then*

$$X^* \leq_c Y^* \quad \Longleftrightarrow \quad f(X) \leq_* g(Y).$$

Proof. From Eq. (2.1) of Hickey [41] we have that the distribution function of X^* can be expressed as

$$F_{X^*}(t) = \mathbb{P}(X^* \leq t) = \int_0^t f^*(x)\,\mathrm{d}x = \int_{\{x\in\mathbb{R}:\ f(x)>f^*(t)\}} f(x)\,\mathrm{d}x = \mathbb{P}(f(X) > f^*(t)).$$

The use of strict inequality in the set $\{x \in \mathbb{R} : f(x) > f^*(x)\}$ follows from the assumption of f having no flat zones. Therefore, one has

$$F_{f(X)}\left(f^*(t)\right) = 1 - F_{X^*}(t), \qquad t \in \mathbb{R}.$$

Because f^* is strictly decreasing and thus invertible, we have

$$F^{-1}_{f(X)}(p) = f^*\left(F^{-1}_{X^*}(1-p)\right), \qquad 0 < p < 1, \tag{3.33}$$

where $F^{-1}_{X^*}$ denotes the right-continuous inverse of F_{X^*}. Then, recalling Remark 3.8 the stochastic relation

$$f(X) \leq_* g(Y)$$

is satisfied if and only if

$$\frac{F^{-1}_{g(Y)}(p)}{F^{-1}_{f(X)}(p)} = \frac{g^*\left(G^{-1}_{Y^*}(1-p)\right)}{f^*\left(F^{-1}_{X^*}(1-p)\right)}, \qquad \text{is increasing in } p \in (0,1).$$

Finally, this is equivalent to $X^* \leq_c Y^*$ by virtue of Remark 3.7.

Example 3.4 From expression (3.33) and taking in account the inverse probability integral transform, we obtain that $f(X) =_{st} f^*(X^*)$. Therefore, it is clear that random variables having the same rearrangements have also the same differential entropy and varentropy. Here we provide an illustrative example. Let X and Y be random variables taking values in $[-1,1]$, with density functions f and g given by

$$f(x) = 1 - |x|, \quad -1 \leq x \leq 1, \qquad g(x) = |x|, \quad -1 \leq x \leq 1,$$

respectively. It is easy to compute that

$$m_f(c) = m_g(c) = \mu\{x : g(x) > c\} = 2(1-c)I_{[0,1]}(c).$$

Then $f^*(x) = g^*(x) = (1 - x/2)I_{[0,2)}(x)$. Therefore, we have that $f^*(X^*) =_{st} g^*(Y^*)$ and thus we can conclude that $H(X) = H(Y)$ and $V(X) = V(Y)$. It is worth mentioning that g is not unimodal.

Remark 3.10 Let X and Y be nonnegative random variables having common support $S = (a,b)$, and having pdfs $f(x)$ and $g(x)$, respectively. Then,

$$X \leq_c Y \qquad \Longrightarrow \qquad \begin{cases} f(X) \leq_* g(Y), & \text{if } f \text{ and } g \text{ are strictly decreasing} \\ f(X) \geq_* g(Y), & \text{if } f \text{ and } g \text{ are strictly increasing.} \end{cases}$$

The following remarks are straightforward and they will be useful later on.

Remark 3.11 If X has a symmetric and unimodal pdf $f(x)$, then $X^* =_{st} 2|X - Me|$ where Me is the median of X.

Remark 3.12 If X has an unimodal and strictly decreasing pdf in the interval $(a, +\infty)$, then $f^*(x) = f(x)$ for all $x \in \mathbb{R}$, i.e. $X^* =_{st} X$.

With the expression $Y =_{st} X$ we intend that X and Y are identically distributed. A possible comparison in terms of the kurtosis order is provided by Remark 3.11 through the following theorem (see Oja [60]).

Theorem 3.4 *Let X and Y be symmetric unimodal random variables having median zero. $X \leq_k Y$ if and only if $|X| \leq_c |Y|$.*

3.4.3 Pdf-related distributions and variability orders

The following result allows to relate the kurtosis order between two symmetric and unimodal pdf and the star order between the respective pdf-related distributions.

Theorem 3.5 *Let X and Y be random variables having symmetric and unimodal pdfs f and g, respectively. If $X \leq_k Y$ then $f(X) \leq_* g(Y)$.*

Proof. Since changes in location do not affect kurtosis we can assume that $Me(X) = Me(Y) = 0$. Therefore from Theorem 3.4 and using Remark 3.11 we obtain that $X \leq_k Y$ if and only if $X^* \leq_c Y^*$. The result follows easily just applying Theorem 3.3.

3.5 Quantiles of the information content

An important aspect in the study of entropy and varentropy for an absolutely continuous variable X is the determination the distribution of $IC(X)$ and $IC(X_t)$. In this respect, it can be useful to introduce the cdf and the quantile of the information content $IC(X)$ (cf. (3)) and $IC(X_t)$ (cf. (1.1)).

Definition 3.5 Let X be an absolutely continuous random variable with support $S_X = (0, b)$, and having pdf $f(x)$ that is continuous in all the support S_X.

(i) The cdf of the information content of X, that is $IC(X)$ is given by

$$M(y) := \mathbb{P}(IC(X) \leq y) = \mathbb{P}(-\log f(X) \leq y) = 1 - K(e^{-y}), \qquad y \in \text{Im}^+(-\log f) \tag{3.34}$$

where $K(x)$ is the cdf of $f(X)$ defined in (3.4) .

(ii) The cdf of the information content of X_t, that is $IC(X_t)$ is given by

$$M_t(y) := \mathbb{P}(-\log f_t(X_t) \leq y) = \mathbb{P}(f(X_t) \geq e^{-y}\, \overline{F}(t)) = 1 - K_t(e^{-y}),$$
$$y \in \text{Im}^+(-\log f_t)$$

where $K_t(x)$ is the cdf of $f_t(X_t)$ defined in (3.19) .

In the following proposition a connection between the log-quantile and the quantile of a pdf-related distribution has been provided.

Proposition 3.10 *Let X be an absolutely continuous random variable having pdf $f(x)$, then the inverse of the log-quantile and of the quantile of $f(X)$ is given by*

$$M^{-1}(q) = -\log K^{-1}(1-q) \qquad \forall q \in (0,1). \tag{3.35}$$

Proof. Applying (3.34) for all $q \in (0,1)$

$$q = M(y) = 1 - K(e^{-y}), \qquad y \in Im^+(-\log f)$$

implies that

$$y = -\log K^{-1}(1-q) \qquad q \in (0,1),$$

from which we have the thesis.

In the following examples two applications involving the determination of quantile and log-quantile of a pdf-related distribution are illustrated.

Example 3.5 Let $X \sim \text{Exp}(\lambda)$, with $\lambda > 0$. Let $K(y)$ the cdf of $f(X)$. The expression for $K(y)$ can be obtained applying case (a) of Proposition 3.3 from which we have

$$K(y) = \overline{F}(f^{-1}(y)) = \overline{F}\left(\frac{\log\left(\frac{\lambda}{y}\right)}{\lambda}\right) = \int_{\frac{\log\left(\frac{\lambda}{y}\right)}{\lambda}}^{\infty} \lambda e^{-\lambda x}\, \mathrm{d}x = \frac{y}{\lambda}, \qquad y \in [0,\infty) \tag{3.36}$$

and by substituting (3.36) into (3.34) we get

$$M(y) = 1 - K(e^{-y}) = 1 - \frac{e^{-y}}{\lambda}, \qquad y \in [0,\infty). \tag{3.37}$$

Finally, the quantile function of $IC(X)$ is obtained from the inverse of the (3.36) applying Equation (3.35)

$$M^{-1}(u) = -\log\left[K^{-1}(1-u)\right] = -\log\left[(1-u)\lambda\right], \qquad u \in (0,1). \tag{3.38}$$

Example 3.6 Let $X \sim \text{Norm}(\mu, \sigma)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$. Let $K(y)$ the cdf of $f(X)$. In order to have an expression for the cdf of $M(y)$ (3.34) it is convenient to calculate the survival function $1 - K(y)$ where $K(y)$ can be obtained applying case (c) of Proposition 3.3

$$\begin{aligned} 1 - K(y) &= F\left(\mu + \sqrt{2}\sigma\sqrt{\log\left(\frac{1}{\sqrt{2\pi}\sigma y}\right)}\right) - F\left(\mu - \sqrt{2}\sigma\sqrt{\log\left(\frac{1}{\sqrt{2\pi}\sigma y}\right)}\right) \\ &= \text{Erf}\left(\sqrt{\log\left(\frac{1}{\sqrt{2\pi}\sigma y}\right)}\right), \qquad y \in \left(0, \frac{1}{\sqrt{2\pi}\sigma}\right] \end{aligned} \tag{3.39}$$

where $F(x)$ is the cdf of the Normal distribution

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x, \qquad x \in \mathbb{R}.$$

and $\text{Erf}(\cdot)$ is the error function defined in (1.36). Applying (3.39) to obtain (3.34) we get

$$M(y) = 1 - K(e^{-y}) = \text{Erf}\left(\sqrt{y + \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)}\right), \qquad y \in \left[\log\left(\sqrt{2\pi}\sigma\right), +\infty\right). \tag{3.40}$$

and inverting (3.40)

$$M^{-1}(u) = \left[\text{Erf}^{-1}(u)\right]^2 + \log\left(\sqrt{2\pi}\sigma\right), \qquad u \in (0,1). \tag{3.41}$$

where $\text{Erf}^{-1}(\cdot)$ is the inverse error function.[1] Therefore the distribution of $IC(X)$ when

[1] If $\theta = \text{Erf}(x)$, then the inverse error function is $x = \text{Erf}^{-1}(\theta)$ and it is defined using McLaurin series (cf. Eq. (1.3) of [16])

$$\text{Erf}^{-1}(\theta) = u + \frac{1}{3}u^3 + \frac{7}{30}u^5 + \frac{127}{630}u^7 + \frac{4369}{22680}u^9 + \dots$$

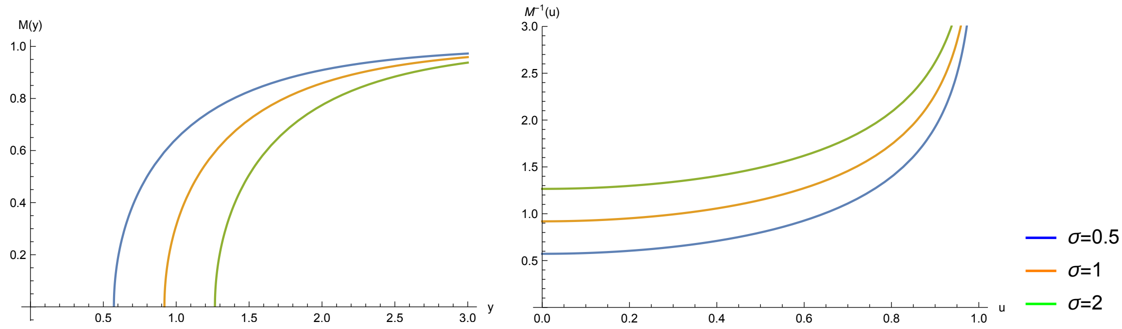


Figure 3.2: (left) cdf of IC(X) when X has a Normal distribution, given in (3.40), and (right) quantile of $IC(X)$, given in (3.41) for various choices of σ (as indicated in the label).

X is a Normal distribution $N(\mu, \sigma)$ does not depend on μ. In the figure 3.2 the cdf (3.40) and the quantile $M^{-1}(u)$ (cf. (3.35)) are plotted as a function of the parameter σ.

3.6 Differential entropy, varentropy and stochastic orders

In this we will use section some concepts related to perform some comparisons for the differential entropy and the differential varentropy. The first part of the section concerns results that are analysed in ref. [25] and are finalized to illustrate the effect of the usual stochastic ordering between pdf-related random variables on the differential entropy of the underlying random variables.

Theorem 3.6 *Let X and Y be absolutely continuous random variables having pdfs $f(x)$ and $g(x)$, respectively, and having finite differential entropy. If $f(X) \leq_{st} g(Y)$, then*

$$H(X) \geq H(Y).$$

Proof. From the hypothesis $f(X) \leq_{st} g(Y)$ it follows that $-\log f(X) \geq_{st} -\log g(Y)$, so that $\mathbb{E}[IC(X)] \geq \mathbb{E}[IC(Y)]$. The thesis then follows from Eq. (4).

where $u = \frac{1}{2}\pi^{1/2}(1-\theta)$.

Remark 3.13 When X and Y have cdf's F and G, and pdf's f and g, respectively, then $X \leq_{disp} Y$ if and only if $f(F^{-1}(u)) \geq g(G^{-1}(u))$ for all $u \in (0,1)$, see (3.B.11) in [67]. Just considering the inverse probability integral transformation and Theorem 1.A.1 in [67], if $X \leq_{disp} Y$ holds then $f(X) \geq_{st} g(Y)$ and using Theorem 3.6 we obtain that $H(X) \leq H(Y)$. On the other hand, it is well known that if X is IFR then $X_{t_2} \leq_{disp} X_{t_1}$ for all $t_1 \leq t_2$, see Belzunce *et al.* [9] and Pellerey and Shaked [63]. From the previous arguments it is apparent that if X is IFR then $H(X_t)$ is decreasing when t increases. This result is well known in the literature, see Ebrahimi [27]. A similar result holds for DFR distributions by exchanging all inequalities and replacing decreasing for increasing in the residual differential entropy.

A similar result can be stated for the differential varentropy when the pdf-related random variables are compared in the star order.

Theorem 3.7 *Let X and Y be absolutely continuous random variables having pdf $f(x)$ and $g(x)$, respectively, and having finite differential varentropy. If $f(X) \leq_* g(Y)$, then*

$$V(X) \leq V(Y).$$

Proof. Recalling Remark 3.9 one has that the assumption $f(X) \leq_* g(Y)$ is equivalent to $\log f(X) \leq_{disp} \log g(Y)$. The latter relation implies that $\mathrm{Var}[\log f(X)] \leq \mathrm{Var}[\log g(Y)]$ (see, for instance, Section 3.B.2 of [67]). The thesis then follows from (2.1).

With reference to equation (3), we denote by

$$L(x) = \mathbb{P}(IC(X) \leq x) \qquad \text{and} \qquad M(x) = \mathbb{P}(IC(Y) \leq x) \tag{3.42}$$

the cdfs of the information contents $IC(X) = -\log f(X)$ and $IC(Y) = -\log g(Y)$, respectively (cf. Equation (3.34)). Then, in the following proposition we show some stochastic comparisons between the information contents of X and Y, where a role is played by the mapping function related to these cdfs.

Remark 3.14 Let X and Y be absolutely continuous random variables having pdf $f(x)$ and $g(x)$, respectively, and let $\varphi(x) := M^{-1}(L(x))$ be the mapping function related to the cdfs (3.42).
(i) From Remark 3.5 we have that $IC(X) \leq_{st} IC(Y)$ if and only if $\varphi(x) \geq x$ for all x in the support of X. Moreover, due to Theorem 1.A.3(a) of [67] these conditions are equivalent to $f(X) \geq_{\text{st}} g(Y)$.
(ii) From Remark 3.6 one has $IC(X) \leq_{disp} IC(Y)$ if and only if $\varphi'(x) \geq 1$ for all x in the support of X.
(iii) Due to Theorem 3.4 of Alimohammadi *et al.* [2], if $IC(X) \geq_{st} IC(Y)$ and $IC(X) \leq_{disp} IC(Y)$, then $f(X) \leq_{\text{disp}} g(Y)$.

Hereafter we show further comparison results for the differential varentropy.

Theorem 3.8 *Let X and Y be symmetric and unimodal random variables with median zero. If $X \leq_k Y$ then*

$$V(X) \leq V(Y).$$

Proof. The result follows easily from Theorem 3.4 and 3.5.

Remark 3.15 Theorem 3.8 provides many possible comparisons. We find in Arriaza *et al.* [4] a compilation of many unimodal symmetric distributions that are ordered in the kurtosis sense. For example, we know that the classical normal, logistic and Cauchy distributions satisfy that normal $\leq_k$ logistic $\leq_k$ Cauchy, Therefore, just applying Theorem 3.8 we obtain[2] that $V(\text{normal}) \leq V(\text{logistic}) \leq V(\text{Cauchy})$.

[2]This result is obviously confirmed by the direct calculations (see Table 2.1 and Appendix A).

Proposition 3.11 *If* X *and* Y *are absolutely continuous random variables having strictly decreasing pdfs in the interval* $(a, +\infty)$ *and if* $X \leq_c Y$*, then*

$$V(X) \leq V(Y).$$

Proof. Without lack of generality we assume that $a = 0$. By Definition 3.4 we have that X^* is identically distributed as X. Hence, by Theorem 3.3 the assumption $X^* \leq_c Y^*$ is equivalent to $f(X) \leq_* g(Y)$, where $f(x)$ and $g(x)$ denote respectively the pdfs of X and Y. We thus obtain $V(X) \leq V(Y)$ from Theorem 3.7.

In the last part of this section we present two examples involving the exponential and the normal families of distribution. In these examples a comparison of differential entropy and differential varentropy is given making use of Theorems 3.6 and 3.7.

Example 3.7 Let $X \sim \text{Exp}(\lambda_X)$ and $Y \sim \text{Exp}(\lambda_Y)$. From (3.37) the cdf of $IC(X)$ and $IC(Y)$ are given by $L(y) = 1 - \frac{e^{-y}}{\lambda_X}$ and $M(y) = 1 - \frac{e^{-y}}{\lambda_Y}$. Then, applying Equation (3.38) the mapping function (3.37) is given by

$$\varphi(x) = M^{-1}(L(x)) = -\log(1 - L(x)\lambda_Y) = x + \log\left(\frac{\lambda_X}{\lambda_Y}\right), \qquad x \in [0, \infty).$$

If $\lambda_X \leq \lambda_Y$ the condition $\varphi(x) \geq x$ is satisfied for all x in the support of X. Hence, for the point (i) of Remark 3.14 we have $IC(X) \leq_{st} IC(Y)$ and for the Theorem 3.6, we have $H(X) \geq H(Y)$. Moreover, because $\varphi'(x) = 1$ for all x in the support of X, for the point (ii) of Remark 3.14 we have $IC(X) =_{disp} IC(Y)$ or alternatively $f(X) =_* g(Y)$ implying that, for Theorem 3.7, $V(X) = V(Y)$.

Example 3.8 Let $X \sim \text{Norm}(\mu_X, \sigma_X)$ and $Y \sim \text{Norm}(\mu_Y, \sigma_Y)$. From (3.40) the cdf of $IC(X)$ and $IC(Y)$ are given respectively by

$$L(y) = \text{Erf}\left(\sqrt{y + \log\left(\frac{1}{\sqrt{2\pi}\sigma_X}\right)}\right)$$

$$M(y) = \mathrm{Erf}\left(\sqrt{y + \log\left(\frac{1}{\sqrt{2\pi}\sigma_Y}\right)}\right)$$

where $\mathrm{Erf}(\cdot)$ is the error function defined in (1.36). Then, applying Equation (3.41) the mapping function (3.37) is given by

$$\varphi(x) = M^{-1}(L(x)) = [\mathrm{Erf}^{-1}(L(x))]^2 + \log\left(\sqrt{2\pi}\sigma_Y\right) = x + \log\left(\frac{\sigma_Y}{\sigma_X}\right) \tag{3.43}$$

where $\mathrm{Erf}^{-1}(\cdot)$ is the inverse error function (see footnote 1 at page 74). Observing from (3.43) that the condition $\sigma_X \geq \sigma_Y$ implies that $\varphi(x) \geq x$ is satisfied for all x in the support of X, according to the Remark 3.14 and to the Theorem 3.6 we have $IC(X) \leq_{st} IC(Y)$ and $H(X) \geq H(Y)$. Moreover, the condition $\varphi'(x) = 1$ for all x in the support of X implies for Remarks 3.7 and 3.14 that we have $IC(X) =_{disp} IC(Y)$ and $V(X) = V(Y)$.

3.6.1 Application to the residual varentropy

In the following theorems, that are given in [25], a condition for which the residual varentropy is monotonic and an upper bound are provided.

Theorem 3.9 *Let X be a nonnegative absolutely continuous random lifetime with support S_X having a cdf F and a pdf $f(x)$. Let X_t be its residual lifetime at time t, as defined in (1.1). Let us suppose that exists $t_0 \in S_X$ such the pdf of X_t, defined in (1.3), is strictly decreasing $\forall t \geq t_0$. If the ratio*

$$\frac{f(F^{-1}(1-(1-p)u))}{f(F^{-1}(1-(1-p)v))}$$

increases (decreases) in $p \in (0,1)$, $\forall\, 0 < v < u < 1$. Then $V(X_t)$ increases (decreases) in $t \geq t_0$.

Proof. Given t_1 and t_2 such that $t_0 < t_1 < t_2$ and using Proposition 3.11 we just need to check that $X_{t_1} \leq_c (\geq_c) X_{t_2}$. From the characterization of the convex transfor order given in Remark 3.7, the expression of the pdf of the residual life given in (1.3) and the expression of the inverse of the distribution function of the residual life given in Proposition 3.2 we obtain that

$$X_{t_1} \leq_c (\geq_c) X_{t_2} \iff \frac{f(F^{-1}(1-(1-p)\overline{F}(t_1)))}{f(F^{-1}(1-(1-p)\overline{F}(t_2)))} \text{ is increasing (decreasing) in } p \in (0,1).$$

The result follows easily from the hypothesis assumption just considering $v = \overline{F}(t_2)$ and $u = \overline{F}(t_1)$.

Example 3.9 Let $X \sim \text{Weibull}(k, \lambda)$ be a Weibull distribution with shape parameter k and scale parameter λ. From the expression of the cdf of X given by $F(x) = 1 - \exp(-(x/\lambda)^k)$, it is a straightforward matter to compute the ratio

$$\frac{f(F^{-1}(1-(1-p)u))}{f(F^{-1}(1-(1-p)v))} = \frac{u}{v}\left(\frac{\ln((1-p)u)}{\ln((1-p)v)}\right)^{\frac{k-1}{k}}$$

for all $0 < v < u < 1$. It follows easily that the above ratio is increasing (decreasing) when $k > 1$ ($k < 1$) and it is constant for $k = 1$. It is well-known that Weibull distributions are always unimodal having decreasing density functions after the mode. Then, the pdfs of the residual lives after the mode satisfy the conditions of Theorem 3.9, see Proposition 3.1. Then, $V(X_t)$ increases (decreases) $\forall t \geq$ mode, for $k > 1$ ($k < 1$). On the other hand, the Weibull distribution is IFR (DFR) for $k > 1$ ($k < 1$). Then using Remark 3.13 we obtain that the differential entropy and the varentropy of the residual lives have opposite directions of growth depending on the shape parameter k. Of course, the case $k = 1$ is just the exponential distribution where the residual lives have both differential entropy and varentropy constant. Some examples of the varentropy of the residual lives in the Weibull distributions are shown computationally in figure 2.3.

We recall that in Theorem 2.3 it has been proved that if a random lifetime X is ILR, i.e. its pdf is logconcave, then the residual varentropy (2.8) is such that $V(X_t) \leq 1$ for all t in the support of X. The condition of log-concave density means that the pdf belongs to the class of strong unimodal densities. Hereafter we show a similar result making use of some of the above results. To this aim we will make use of IFR random variables. A random variable is IFR if its hazard rate (1.4) is increasing.

Theorem 3.10 *Let X be a nonnegative absolutely continuous random lifetime with support S_X, and let X_t be its residual lifetime at time t, as defined in (1.1). If, for a given $t \in S_X$, X_t is IFR and its pdf (1.3) is strictly decreasing, then*

$$V(X_t) \leq 1.$$

Proof. Since, for a given $t \in S_X$, the residual lifetime X_t is IFR, then for Theorem 4.B.11 of [67] one has that $X_t \leq_c \text{Exp}$, where Exp denotes an exponential random variable. Moreover, by assumption X_t has a strictly decreasing pdf $f_t(x)$ for all $x \geq 0$, so that $X_t \leq_c \text{Exp}$ if and only if $X_t^* \leq_c \text{Exp}^*$ for all $t \in S_X$, where X_t^* and Exp^* are the rearrangement of X and Exp, respectively. Due to Theorem 3.3, the last stochastic inequality is satisfied if and only if

$$f_{X_t}(X_t) \leq_* g(\text{Exp}),$$

where g is the pdf of Exp. Hence, making use of Theorem 3.7 we find

$$V(X_t) \leq V(\text{Exp}).$$

Finally, the thesis follows recalling that $V(\text{Exp}) = 1$ (see table 2.1).

Remark 3.16 From Proposition 3.1, we have that the conditions of Theorem 3.10 easily hold for all unimodal IFR distributions when $t \geq m$, where m is the mode of X. For example, this is the case of the residual lives of the Weibull distribution for $k > 1$, as it is described in Example 3.9. After the mode the varentropy increases but it has an upper bound equal to 1.

Chapter 4

Correlations in the information of bivariate distributions

In this chapter we will show how the study of the correlations in the information of bivariate distributions can be carried on making use of the concept of covarentropy. We will study the property of covarentropy both for discrete (Section 2) and for continuous random variables (Section 3). In particular we are interested to the investigation of relations between covarentropy and covariance. This will be done for different choices of families of bivariate distributions as normal, exponential, log-normal and Gamma distributions.

4.1 Introduction

Let us suppose that X and Y are two random variables. If X and Y are discrete with joint probability function $p(x, y)$, the joint information content is given by the random variable

$$IC(X, Y) = -\log p(X, Y)$$

If X and Y are absolutely continuous with joint pdf $f(x, y)$, then the joint information content is given by the random variable

$$IC(X, Y) = -\log f(X, Y).$$

Let us assume that X and Y are discrete random variables having probability functions $p_X(x)$ and $p_Y(y)$, respectively and joint probability $p(x, y)$. If X and Y are independent, then $p(x, y) = p_X(x)p_Y(y)$ that implies

$$IC(X, Y) = IC(X) + IC(Y)$$

where $IC(X)$ and $IC(Y)$ are the information content of X and Y, respectively (cf. (1)). Similarly, let us consider two absolutely continuous random variables X and Y having pdfs $f_X(x)$ and $f_Y(y)$, respectively and joint pdf $f(x, y)$. If X and Y are independent then $f(x, y) = f_X(x)f_Y(y)$ implying that the the information content of two random variables is the sum of the information contents of the two random variables.

With the name *covarentropy* we will denote the covariance of $IC(X)$ and $IC(Y)$ or alternatively the variance of $IC(X, Y)$. We will write the covarentropy of X and Y using the symbol $\mathrm{Cov}\mathcal{E}(X, Y)$. The covarentropy is thus a measure of the joint variability of the information contents of X and Y and is directly proportional to the correlation between their information contents.

4.2 Discrete random variables

In this section we will consider X and Y as two discrete random variables, having joint probability function $p(x, y)$ and whose marginal distributions X and Y have probability function $p_X(x)$ and $p_Y(y)$, respectively. Let us now give a definition of the covarentropy of two discrete random variables.

Definition 4.1 Let X and Y be two discrete random variables, the *covarentropy* of X

and Y is defined as

$$\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &:= \mathrm{Cov}[IC(X), IC(Y)] \\
&= \mathbb{E}[IC(X)IC(Y)] - \mathbb{E}[IC(X)] \cdot \mathbb{E}[IC(Y)] \\
&= \sum_x \sum_y p(x,y)\{\log p_X(x) \log p_Y(y)\} - H(X)H(Y) \qquad (4.1)
\end{aligned}$$

where $IC(X)$ and $IC(Y)$ are the information contents of X and Y, respectively (cf. (1)) and $H(X)$ and $H(Y)$ their respective the Shannon entropies (cf. (2)).

4.2.1 Independent random variables

Let us show that if X and Y are independent random variables, then the covarentropy vanishes.

Theorem 4.1 *Let X and Y be two discrete random variables, if X and Y are independent, then*

$$\mathrm{Cov}\mathcal{E}(X,Y) = 0 \qquad (4.2)$$

Proof. Suppose that X and Y are independent, then we have

$$p(x,y) = p_X(x) \cdot p_Y(y) \qquad \forall x, y.$$

This implies that applying (4.1)

$$\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \sum_x \sum_y p(x,y) \log p_X(x) \log p_Y(y) \\
&- \sum_x p_X(x) \log p_X(x) \sum_y p_Y(y) \log p_Y(y) \\
&= \sum_x \sum_y (p_X(x) \cdot p_Y(y)) \log p_X(x) \log p_Y(y) \\
&- \sum_x p_X(x) \log p_X(x) \sum_y p_Y(y) \log p_Y(y) = 0
\end{aligned}$$

Example 4.1 Let X and Y be two discrete random variables having the following joint probability distribution

$x \backslash y$	0	1	$p_X(x)$
0	θ	$\frac{1}{2}-\theta$	$\frac{1}{2}$
1	$\frac{1}{2}-\theta$	θ	$\frac{1}{2}$
$p_Y(y)$	$\frac{1}{2}$	$\frac{1}{2}$	1

where $\theta \in [0, \frac{1}{2}]$. The covariance of X and Y can be calculated as a function of the parameter θ:

$$\mathrm{Cov}(X,Y) = \sum_x \sum_y x\ y\ p(x,y) - \sum_x x\ p_X(x) \sum_y y\ p_Y(y) = \theta - \frac{1}{4}$$

The correlation coefficient of X and Y is therefore

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\theta - \frac{1}{4}}{\frac{1}{4}} = 4\theta - 1$$

Observing Figure 4.1 we see that the correlation coefficient vanishes for $\theta = \frac{1}{4}$, for greater values we find positive correlation while for smaller values negative correlation. The entropy of X and the entropy of Y can be easily calculated according to the Equation (2)

$$H(X) = H(Y) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = \log 2$$

The covarentropy can be found from Equation (4.1)

$$\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \mathrm{Cov}[IC(X), IC(Y)] = \mathbb{E}[IC(X)\cdot IC(Y)] - \mathbb{E}[IC(X)]\cdot\mathbb{E}[IC(Y)] \\
&= \sum_x \sum_y \log p_X(x) \log p_Y(y)\ p(x,y) - H(X)\cdot H(Y) \\
&= (2\theta + 1 - 2\theta)\left[\log\frac{1}{2}\right]^2 - \left[\log\frac{1}{2}\right]^2 = 0
\end{aligned}$$

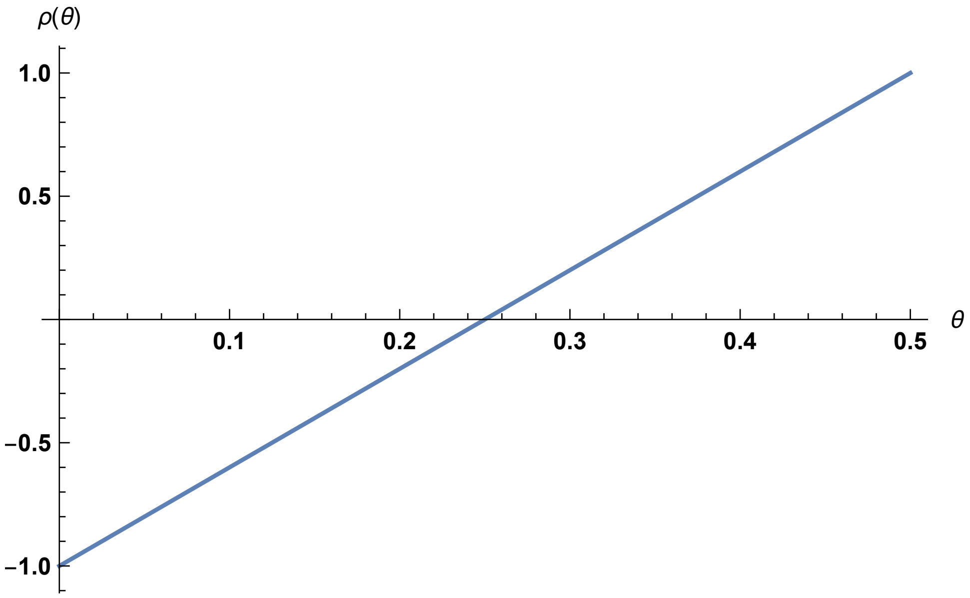


Figure 4.1: Correlation coefficient vs. θ parameter.

Example 4.2 Let X and Y be two discrete random variables having the following joint probability distribution

$x \backslash y$	0	1	$p_X(x)$
0	θ	$1-p-\theta$	$1-p$
1	$1-p-\theta$	$2p-1+\theta$	p
$p_Y(y)$	$1-p$	p	1

If we fix $\theta \geq 0$ for $p < \frac{1}{2}$ the condition for which each term of the distribution is positive implies that $1-p \geq \theta$. In analogy with the previous case if we suppose that $\theta \geq \frac{1}{2}$ we find again that $1-p \geq \theta$. So that we have

$$\max\{0, 1-2p\} \leq \theta \leq 1-p, \tag{4.3}$$

that is individuated by the region highlighted in the yellow area of figure 4.2. The covari-

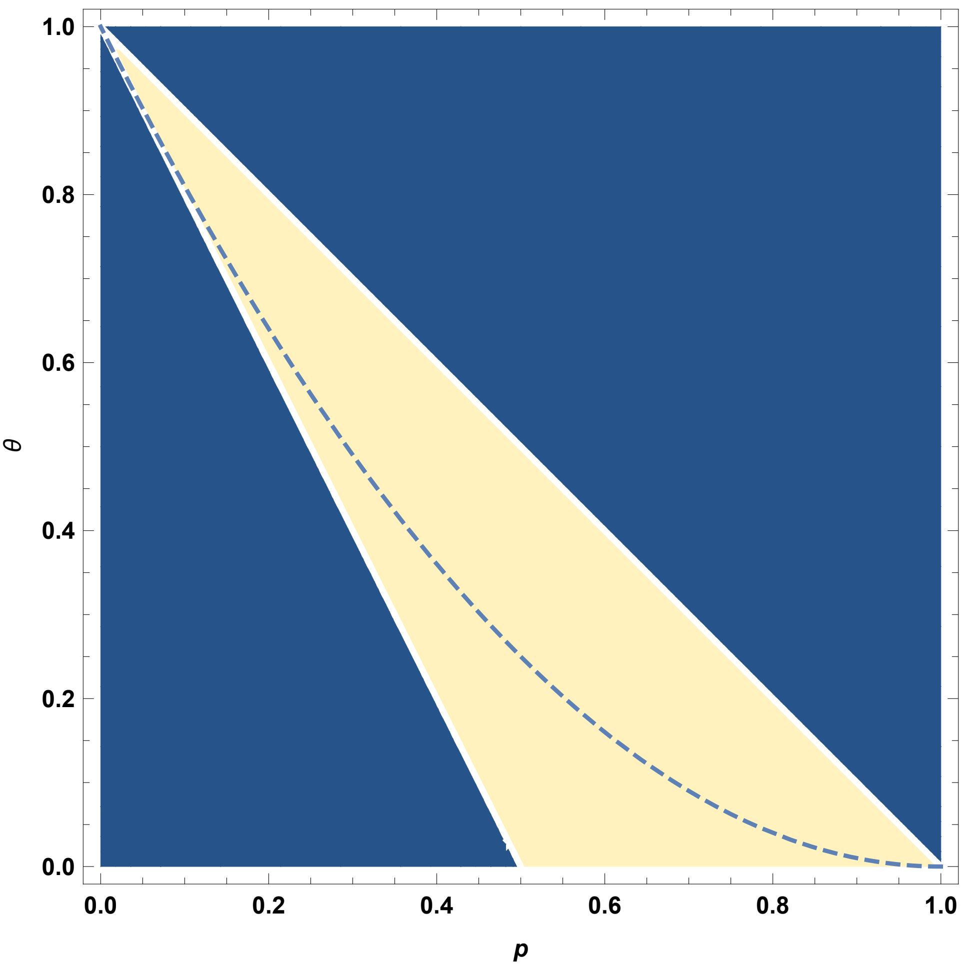


Figure 4.2: Area of the region for which inequality (4.3) is verified (yellow area). The dashed line corresponds to the condition of vanishing covariance (4.4).

ance of X and Y has the following expression

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \sum_x \sum_y xy\ p(x,y) - \sum_x x\ p_X(x) \sum_y y\ p_Y(y) \\
&= \theta + 2p - 1 - p^2 = \theta - (p^2 - 2p + 1) = \theta - (1-p)^2
\end{aligned}
$$

The condition of zero covariance is

$$
\theta = (1-p)^2 \tag{4.4}
$$

that is represented by the parabola in Figure 4.2. The entropy of the random variable X is calculated through the relation

$$
H(X) = -\sum_x p_X(x) \log p_X(x) = -(1-p)\log(1-p) - p \log p
$$

An analogous calculation can be done for Y that gives

$$
H(Y) = -\sum_y p_Y(y) \log p_Y(y) = -(1-p)\log(1-p) - p \log p
$$

Finally, applying (4.1) we can calculate covarentropy of X and Y

$$
\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \sum_x \sum_y \log p_X(x) \log p_Y(y)\ p(x,y) - H(X) \cdot H(Y) \\
&= \theta[\log(1-p)]^2 + 2(1-p-\theta)\log p\ \log(1-p) + (\theta + 2p - 1)\ (\log p)^2 \\
&- (1-p)^2\ [\log(1-p)]^2 - p^2\ (\log p)^2 - 2p\ (1-p)\log p\ \log(1-p) \\
&= [\theta - (1-p)^2]\ [\log(1-p)]^2 + [\theta - (1-p)^2]\ (\log p)^2 \\
&- 2[\theta - (1-p)^2]\log p\ \log(1-p) = [\theta - (1-p)^2]\ [\log p - \log(1-p)]^2 \\
&= [\theta - (1-p)^2] \left(\log \frac{p}{1-p} \right)^2
\end{aligned} \tag{4.5}
$$

Eq. (4.5) vanishes for $\theta = (1-p)^2$ and $p = \frac{1}{2}$. The first coincides with the condition of zero covariance, while the last condition was already treated in Example 4.1.

4.2.2 Uniform distributions

Another remarkable property of the covarentropy concerns the case in which one of them is uniform, as showed in the following theorem.

Theorem 4.2 *Let X and Y be two discrete random variables. If X is uniform, then*

$$\mathrm{Cov}\mathcal{E}(X,Y) = 0$$

Proof. Suppose that X is uniform, that is

$$p_X(x) = \frac{1}{n} \tag{4.6}$$

for $x \in \{x_1, x_2 \ldots, x_n\}$. Then, substituting (4.6) into (4.1) we have

$$\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \sum_x \sum_y p(x,y) \log p_X(x) \log p_Y(y) \\
&- \sum_x p_X(x) \log p_X(x) \sum_y p_Y(y) \log p_Y(y) \\
&= \sum_y \log p_Y(y) \log \frac{1}{n} \sum_x p(x,y) - \log \frac{1}{n} \, [-H(Y)] \\
&= \sum_y \log p_Y(y) \; \log \frac{1}{n} \, p_Y(y) - \log n \; H(Y) \\
&= \log \frac{1}{n} \sum_y p_Y(y) \log p_Y(y) - \log \frac{1}{n} H(Y) \\
&= \log n \; H(Y) - \log n \; H(Y) = 0.
\end{aligned}$$

4.2.3 Covarentropy for uncorrelated random variables

If the random variables X and Y are uncorrelated, it is not generally true that the covarentropy is zero. We can construct an example to show this property.

Example 4.3 Let X and Y be two discrete random variables having the following joint distribution

y / x	0	1	$p_X(x)$
-1	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$
$p_Y(y)$	$\frac{3}{4}$	$\frac{1}{4}$	1

The covariance is zero. In fact

$$\mathrm{Cov}(X;Y) = \sum_x \sum_y x\ y\ p(x,y) - \sum_x x\ p_X(x) \sum_y y\ p_Y(y) = 0 - \left(-\frac{1}{4} + \frac{1}{4}\right) \left(\frac{1}{4}\right) = 0$$

Therefore the variables X and Y are uncorrelated. Let us show that the covarentropy is not zero. The entropy of X and the entropy of Y is given applying (2):

$$\begin{aligned} H(X) &= -\sum_x p_X(x) \log p_X(x) = -\frac{1}{4} \log \frac{1}{4} - \frac{1}{2} \log \frac{1}{2} \\ & \quad -\frac{1}{4} \log \frac{1}{4} = 3 \left(\frac{1}{2} \log 2\right) = \frac{3}{2} \log 2 \\ H(Y) &= -\sum_y p_Y(y) \log p_Y(y) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} \\ &= \frac{3}{4} \log \frac{4}{3} + \frac{1}{4} \log 4 \end{aligned}$$

The covarentropy can be calculated form (4.1)

$$
\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \mathbb{E}[IC(X)\cdot IC(Y)] \\
&- \mathbb{E}[IC(X)]\cdot\mathbb{E}[IC(Y)] = \sum_x\sum_y \log p_X(x)\log p_Y(y)\ p(x,y) \\
&- H(X)\cdot H(Y) = \frac{1}{4}\log 4\ \log\frac{4}{3} + \frac{1}{4}\log 2\ \log\frac{4}{3} + \frac{1}{4}\log 2\ \log 4 \\
&+ \frac{1}{4}\log 4\ \log\frac{4}{3} - \left(\frac{3}{4}\log\frac{4}{3} - \frac{1}{4}\log 4\right)\frac{3}{2}\log 2 \\
&= \frac{1}{4}\left[\frac{1}{2}\log 2\ \log\frac{4}{3} + 5\ (\log 2)^2\right] \cong 0.625
\end{aligned}
$$

We conclude that even if the correlation between the variables X and Y is zero, the covarentropy is not necessarily zero.

4.3 Absolutely continuous random variables

Let us now consider X and Y as two absolutely continuous random variables, having joint probability pdf $f(x,y)$ and whose the marginal distributions X and Y have pdfs $f_X(x)$ and $f_Y(y)$, respectively. Let us now give a definition of the covarentropy of two continuous random variables.

Definition 4.2 Let X and Y two absolutely continuous random, the *covarentropy* of X and Y is defined as the quantity

$$
\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \mathbb{E}[IC(X)\cdot IC(Y)] - \mathbb{E}[IC(X)]\cdot\mathbb{E}[IC(Y)] \\
&= \int_{\mathbb{R}^2} \log f_X(y)\log f_Y(x)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y - H(X)\cdot H(Y)
\end{aligned}
\tag{4.7}
$$

where $IC(X)$ and $IC(Y)$ are the information contents of X and Y, respectively (cf. (3)) and $H(X)$ and $H(Y)$ their respective the differential entropies (cf. (4)).

4.3.1 Covarentropy for independent random variables

In analogy with the discrete random variables, also for continuous random variables the covarentropy vanishes if the random variables are independent.

Theorem 4.3 *Let X and Y be independent absolutely continuous random variables, then*

$$\mathrm{Cov}\mathcal{E}(X,Y) = 0.$$

Proof. Suppose that X and Y are independent, so that

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall x \; \forall y.$$

This implies that for (4.7)

$$\begin{aligned} \mathrm{Cov}\mathcal{E}(X,Y) &= \int_{\mathbb{R}^2} \log f_Y(y) \log f_X(x) \; f(x,y) \; \mathrm{d}x \; \mathrm{d}y - H(X) \cdot H(Y) \\ &= \int_{\mathbb{R}^2} \log f_Y(y) \log f_X(y) \; f_X(x) f_Y(y) \; \mathrm{d}x \; \mathrm{d}y - H(X) \cdot H(Y) = 0 \end{aligned} \tag{4.8}$$

4.3.2 Uniform distributions

Theorem 4.4 *Let X and Y two absolutely continuous random variables, if X is uniform, then*

$$\mathrm{Cov}\mathcal{E}(X,Y) = 0$$

Proof. Suppose that X is uniform in an interval $[a,b]$ with a and b real constants such that $-\infty < a < b < \infty$, the pdf of uniform distribution is the function

$$f_X(x) = \frac{1}{b-a}, \qquad a < x < b, \tag{4.9}$$

then substituting (4.9) into (4.7) we have

$$\begin{aligned}
\mathrm{Cov}\mathcal{E}(X,Y) &= \int_{[a,b]\times\mathbb{R}} \log f_X(y) \log f_Y(x)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y - H(X)\cdot H(Y) \\
&= \log\left(\frac{1}{b-a}\right) \int_{[a,b]\times\mathbb{R}} \log f_Y(y)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y - \log(b-a)\ H(Y) \\
&= -\log(b-a) \int_{\mathbb{R}} \log f_Y(y)\ f_Y(y)\ \mathrm{d}y - \log(b-a)\ H(Y) \\
&= \log(b-a)\ H(Y) - \log(b-a)\ H(Y) = 0
\end{aligned}$$

4.3.3 Gumbel exponential distributions

Let us consider two random variables X and Y having joint pdf

$$f(x,y) = [(\mu+\theta x)(\lambda+\theta y)-\theta]\ \exp(-\lambda x - \mu y - \theta xy), \qquad x>0, y>0, \tag{4.10}$$

with $\lambda > 0$, $\mu > 0$ and $0 \le \theta \le 1$. Equation (4.10) describes the *Gumbel exponential pdf* (cf. [58]). The marginals of the bivariate distribution (X,Y) are the exponential distributions $\mathrm{Exp}(\lambda)$ and $\mathrm{Exp}(\mu)$ and so $H(X)$ and $H(Y)$ are obtained by (4) and are given by

$$H(X) = -\int_0^\infty f_X(x)\log f_X(x)\ \mathrm{d}x = 1-\log\lambda, \tag{4.11a}$$

$$H(Y) = -\int_0^\infty f_Y(x)\log f_Y(x)\ \mathrm{d}x = 1-\log\mu. \tag{4.11b}$$

The covariance of X and Y is given by

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \int_{[0,+\infty)\times[0,+\infty)} \left(x-\frac{1}{\lambda}\right)\left(y-\frac{1}{\mu}\right) f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \frac{1}{\lambda\mu} + \frac{1}{\theta}\, e^{\frac{\lambda\mu}{\theta}}\ \Gamma\left(0, \frac{\lambda\mu}{\theta}\right)
\end{aligned} \tag{4.12}$$

where $\Gamma(z,x)$ is the incomplete Gamma function (see Eq. 6.5.3 of [1]), defined as

$$\Gamma(z,x) = \int_x^\infty t^{z-1}\, e^{-t}\, \mathrm{d}t \tag{4.13}$$

The covarentropy is given applying (4.7)

$$\begin{aligned}\mathrm{Cov}\mathcal{E}(X,Y) &= \int_{[0,+\infty)\times[0,+\infty)} \log f_X(x)\, \log f_Y(y)\, f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ & -H(X)\cdot H(Y) = -1 + \frac{\lambda\mu}{\theta}\, e^{\frac{\lambda\mu}{\theta}}\, \mathrm{Ei}\left(-\frac{\lambda\mu}{\theta}\right)\end{aligned} \tag{4.14}$$

where $\mathrm{Ei}(x)$ is the exponential integral function of x (see Equation 5.1.2 of [1]), defined as

$$\mathrm{Ei}(x) = -\int_{-x}^{+\infty} \frac{e^t}{t}\, \mathrm{d}t \tag{4.15}$$

In the case where $\lambda = \frac{1}{\mu}$ the covarentropy coincides with the covariance, while in the other cases they differ each other (see Fig. 4.3). In fact, the following theorem can be stated.

Theorem 4.5 *Let X and Y be two random variables having Gumbel exponential distribution with correlation parameter $\theta \neq 0$. Then*

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}\mathcal{E}(X,Y) \iff \lambda = \frac{1}{\mu}.$$

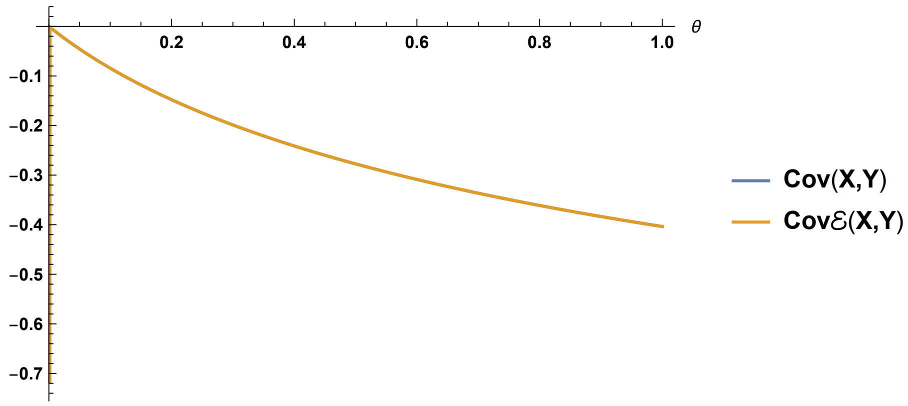
Proof. Substituting (4.13) into (4.12) we have:

$$\mathrm{Cov}(X,Y) = -\frac{1}{\lambda\mu} + \frac{1}{\theta}\, e^{\frac{\lambda\mu}{\theta}} \int_{-\frac{\lambda\mu}{\theta}}^{\infty} \frac{e^{-t}}{t}\, \mathrm{d}t. \tag{4.16}$$

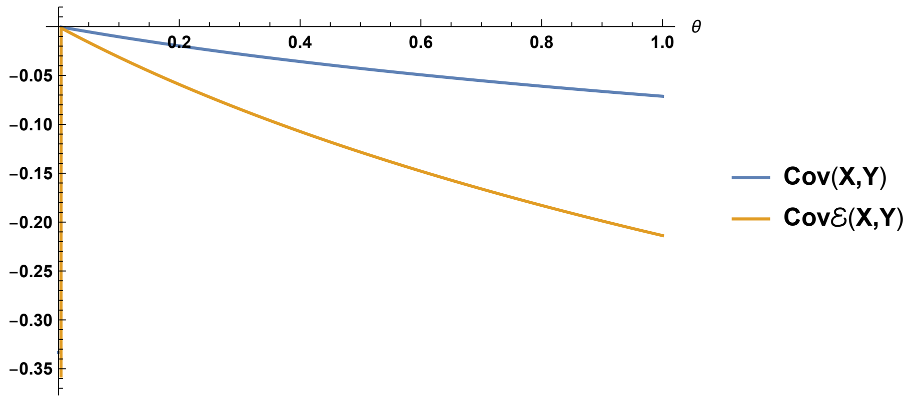
Moreover, substituting (4.15) into (4.14) we have

$$\mathrm{Cov}\mathcal{E}(X,Y) = -1 - \frac{1}{\theta}\, e^{\frac{\lambda\mu}{\theta}} \int_{-\frac{\lambda\mu}{\theta}}^{\infty} \frac{e^{-t}}{t}\, \mathrm{d}t. \tag{4.17}$$

Comparing (4.16) and (4.17) we have that $\mathrm{Cov}(X,Y) = \mathrm{Cov}\mathcal{E}(X,Y)$ is satisfied for all θ



(a) $\lambda = \frac{1}{\mu}$



(b) $\lambda = 1$; $\mu = 3$

Figure 4.3: Covariance and covarentropy calculated for the Gumbel exponential pdf (4.10) for different choices of λ and μ.

in the interval $(0, 1]$ if and only if $\lambda = \frac{1}{\mu}$.

Theorem 4.6 *Let X and Y be two random variables having Gumbel exponential distribution with correlation parameter $\theta = 0$. Then*

$$\text{Cov}(X, Y) = \text{Cov}\mathcal{E}(X, Y) = 0.$$

Proof. It is sufficient to observe that for $\theta = 0$ Gumbel pdf (4.10) has the expression

$$f(x, y) = \mu\lambda \exp(-\lambda x - \mu y) = f_X(x) f_Y(y) \qquad \text{for all } x, y > 0. \tag{4.18}$$

Hence, as a consequence of (4.18), we have that X and Y are independent random variables and therefore they have vanishing covariance and covarentropy.

4.3.4 McKay's bivariate Gamma distributions

Let us consider two random variables X and Y having joint pdf

$$f(x, y) = \frac{a^{p+q}}{\Gamma(p)\,\Gamma(q)} x^{p-1} (y - x)^{q-1} e^{-ay} \qquad x > 0, y > x \tag{4.19}$$

with $p > 0$, $q > 0$. The function $\Gamma(z)$ is the Gamma function of the parameter q (see Eq. 6.1.1 of [1]) defined as

$$\Gamma(z) = \int_0^{+\infty} t^{z-1}\, e^{-t}\, \mathrm{d}t \tag{4.20}$$

Eq. (4.19) gives *McKay's bivariate Gamma pdf* (cf. Chapter 48, Sec. 2.1 of [45]). The marginals are the Gamma pdf's

$$f_X(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} \qquad x > 0,$$

$$f_Y(y) = \frac{a^{p+q}}{\Gamma(p+q)} y^{p+q-1} e^{-ay} \qquad y > 0,$$

with expectations $\mu_X = \frac{p}{a}$ and $\mu_Y = \frac{p+q}{a}$ and variances $\sigma_X^2 = \frac{p}{a^2}$ and $\sigma_Y^2 = \frac{p+q}{a^2}$, respectively. The expression of the covariance is

$$\mathrm{Cov}(X,Y) = \int_{[0,+\infty)\times[x,+\infty)} (x-\mu_X)(y-\mu_Y)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y = \frac{p}{a^2} \tag{4.21}$$

The correlation coefficient is therefore

$$r = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\ \sigma_Y} = \sqrt{\frac{p}{p+q}} = \left(1 + \frac{p}{q}\right)^{-\frac{1}{2}}$$

and depends only on the ratio $\frac{p}{q}$. The entropy of the marginal distributions is

$$H(X) = -\int_0^{\infty} f_X(x)\ \log f_X(x)\ \mathrm{d}x = p + \log\left(\frac{\Gamma(p)}{a}\right) - (-1+p)\ \psi^{(0)}(p)$$

where $\psi^{(m)}(x)$ is the PolyGamma function of order m calculated at x (see Eq. 6.4.1 of [1]) defined by

$$\psi^{(m)}(x) := \frac{d^{m+1}}{dx^{m+1}} \log\Gamma(x) \tag{4.22}$$

where $\Gamma(x)$ is the Gamma function defined in Eq. (4.20). Similarly,

$$\begin{aligned} H(Y) &= -\int_0^{\infty} f_Y(y)\ \log f_Y(y)\ \mathrm{d}y \\ &= p + q + \log\left(\frac{\Gamma(p+q)}{a}\right) - (-1+p+q)\ \psi^{(0)}(p) \end{aligned}$$

The covarentropy can be calculated as a function of $\frac{p}{q}$ with q fixed. The covarentropy is given applying (4.7).

$$\mathrm{Cov}\mathcal{E}(X,Y) = \int_{[0,+\infty)\times[x,+\infty)} \log f_X(x) \log f_Y(y)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y - H(X)\cdot H(Y) \tag{4.23}$$

that can be solved numerically. The covarentropy compared with the covariance has been plotted in Figure 4.4. The covariance is linear in $\frac{p}{q}$ with the slope increasing when q increases. The covarentropy is a non-linear decreasing function of q. It approaches to be linear as the parameter q decreases to zero.

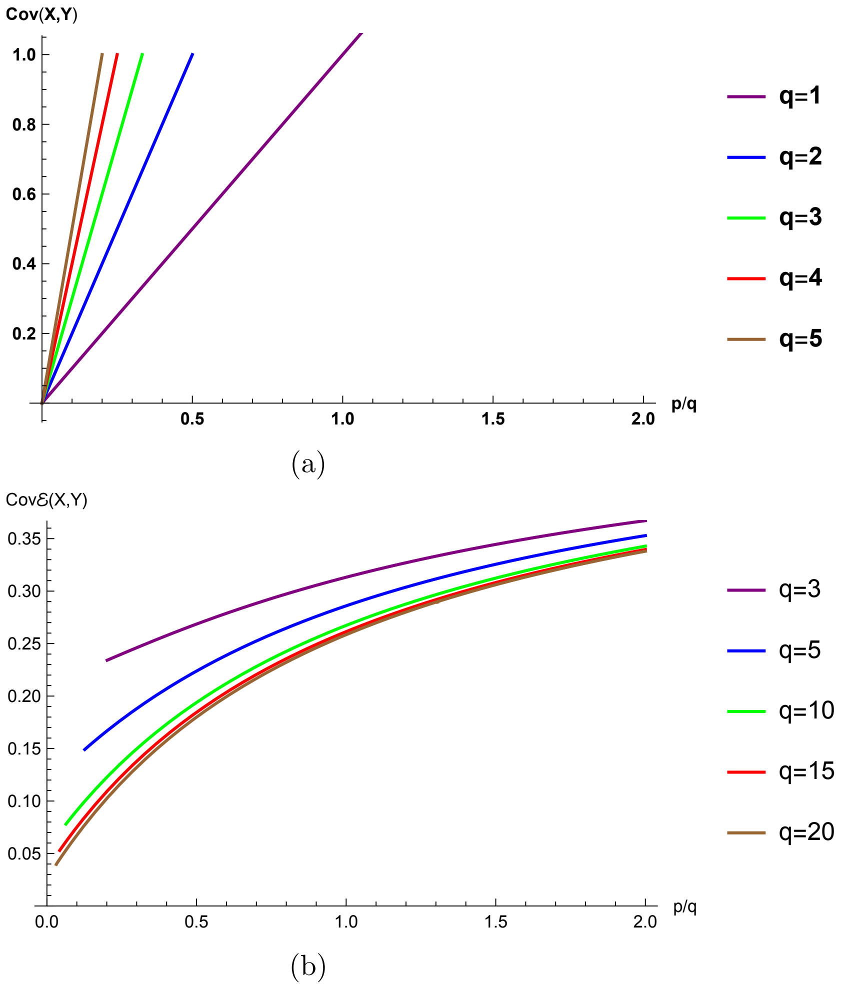


Figure 4.4: (a) Covariance and (b) covarentropy with varying p/q calculated for the Gamma bivariate pdf (4.19) for different choices of q (see Equations (4.21) and (4.23)).

4.3.5 Normal bivariate distributions

Let us consider two random variables X and Y having joint pdf

$$f(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_X\sigma_Y}$$
$$\times \quad \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right]$$
$$x, y \in \mathbb{R} \tag{4.24}$$

with $\mu_X \in \mathbb{R}$, $\mu_Y \in \mathbb{R}$, $\sigma_X > 0$ and $\sigma_Y > 0$. Eq. (4.24) gives *normal bivariate pdf* (cf. Chapter 46 of [45]). The marginals are the normal distributions

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}, \qquad x \in \mathbb{R}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}, \qquad y \in \mathbb{R}$$

with expectations μ_X and μ_Y and the variances σ_X^2 and σ_Y^2, respectively. The entropy calculated for the variables X and Y are obtained by (4)

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x) \; \mathrm{d}x = \frac{1}{2} \log\left(2\pi e \sigma_X^2\right)$$

and

$$H(Y) = -\int_{-\infty}^{\infty} f_Y(x) \log f_Y(x) \; \mathrm{d}x = \frac{1}{2} \log\left(2\pi e \sigma_Y^2\right)$$

The covariance of X and Y is

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty} (x-\mu_Y)(y-\mu_Y) \; f(x,y) \; \mathrm{d}y \; \mathrm{d}x = \rho \; \sigma_X \; \sigma_Y. \tag{4.25}$$

Finally we calculate the covarentropy of X and Y applying (4.7)

$$\mathrm{Cov}\mathcal{E}(X,Y) = \int_{\mathbb{R}^2} \log f_X(x) \log f_Y(y) \; f(x,y) \; \mathrm{d}x \; \mathrm{d}y - H(X) \cdot H(Y) = \frac{\rho^2}{2} \tag{4.26}$$

As a direct consequence of (4.26) we can state the following proposition.

Proposition 4.1 *Let X and Y be two distributions having Normal bivariate pdf (4.24) with mean values μ_X and μ_Y and variances σ_X and σ_Y, respectively. Then $Cov\mathcal{E}(X,Y)$ is independent of μ_X, μ_Y, σ_X and σ_Y.*

We plot the dependence of the covariance and of the covarentropy when the parameter ρ varies (see fig. 4.5).

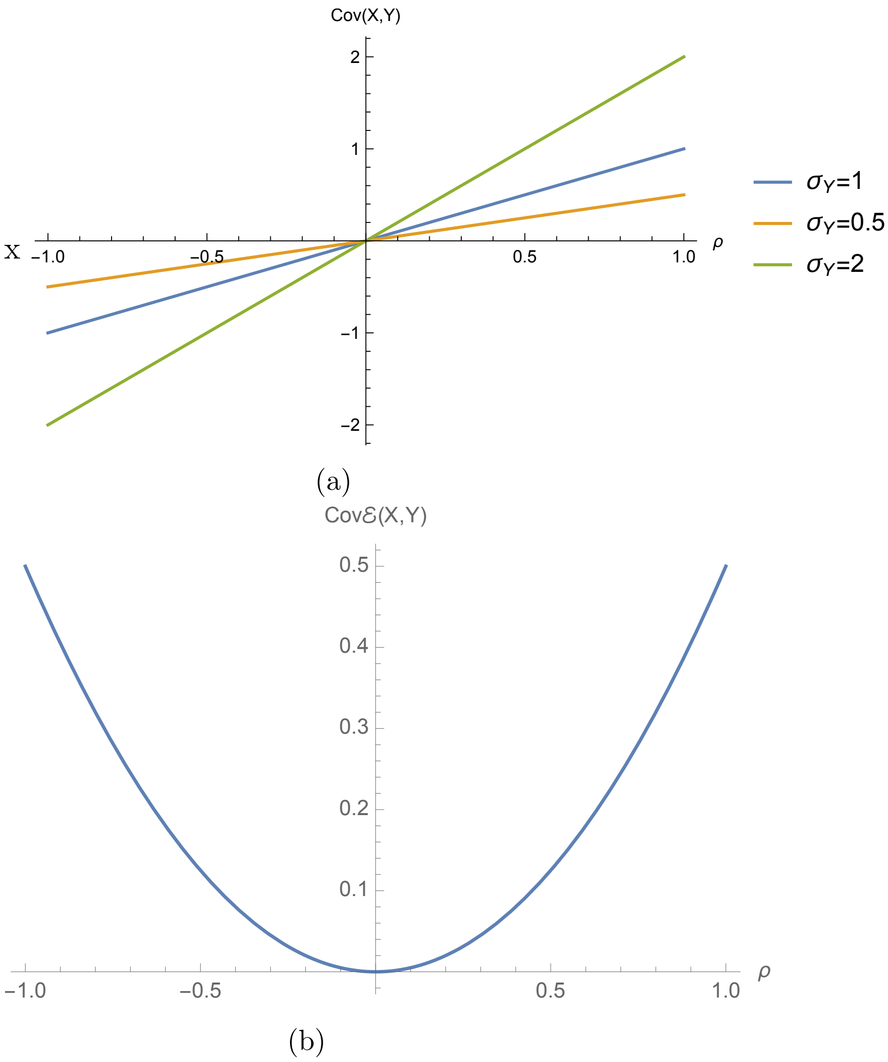


Figure 4.5: Covariance and covarentropy for a bivariate normal distribution.
(a) Covariance as a function of ρ when σ_X is fixed to 1 and for different choices of σ_Y (see Eq. (4.25)).
(b) Covarentropy for a bivariate normal distribution as a function of ρ (see Eq. (4.26)).

4.3.6 Gumbel–Malik–Abraham logistic distributions

Let us consider a continuous random vector (X, Y) with pdf

$$f(x,y) = \frac{2\,e^{-\frac{x-\mu_X}{\sigma_X}}\;e^{-\frac{y-\mu_Y}{\sigma_Y}}}{\sigma_X\sigma_Y\left(1+e^{-\frac{x-\mu_X}{\sigma_X}}+e^{-\frac{y-\mu_Y}{\sigma_Y}}\right)^3}, \qquad x, y \in \mathbb{R} \tag{4.27}$$

with $\mu_X \in \mathbb{R}$ and $\mu_Y \in \mathbb{R}$, $\sigma_X > 0$ and $\sigma_Y > 0$. Eq. (4.27) gives the *Gumbel–Malik–Abraham logistic bivariate pdf* (cf. Chapter 51, Sec. 2 of [45]). The marginals are the logistic distributions

$$f_X(x) = \frac{e^{-\frac{x-\mu_X}{\sigma_X}}}{\left(1+e^{-\frac{x-\mu_X}{\sigma_X}}\right)^2}, \qquad x \in \mathbb{R},$$

$$f_Y(y) = \frac{e^{-\frac{y-\mu_Y}{\sigma_Y}}}{\left(1+e^{-\frac{y-\mu_Y}{\sigma_Y}}\right)^2}, \qquad x \in \mathbb{R},$$

with expectation values μ_X and μ_Y and the variances σ_X^2 and σ_Y^2, respectively.

The entropy of the marginal distributions are obtained by (4)

$$H(X) = -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)\ \mathrm{d}x = 2 + \log\sigma_X \tag{4.28a}$$

and

$$H(Y) = -\int_{-\infty}^{\infty} f_Y(x)\log f_Y(x)\ \mathrm{d}x = 2 + \log\sigma_Y. \tag{4.28b}$$

The covariance of X and Y gives

$$\mathrm{Cov}(X,Y) = \int_{\mathbb{R}^2} (x-\mu_X)\ (y-\mu_Y)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y = \frac{\pi^2}{6}\sigma_X\sigma_Y \tag{4.29}$$

while the covarentropy is given by expression (4.7)

$$\begin{aligned}\mathrm{Cov}\mathcal{E}(X,Y) &= \int_{\mathbb{R}^2} \log f_X(x)\log f_Y(y)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\ &- H(X)\cdot H(Y) = \frac{5\pi^2}{6} - 8 \simeq 0.22\end{aligned} \tag{4.30}$$

As a direct consequence of (4.30) we can state the following proposition.

Proposition 4.2 *Let X and Y be two distributions having Gumbel-Malik-Abraham pdf (4.27) with mean values μ_X and μ_Y and variances σ_X and σ_Y, respectively. Then $Cov\mathcal{E}(X,Y)$ is independent of μ_X, μ_Y, σ_X and σ_Y.*

4.3.7 Farlie–Gumbel–Morgenstern distributions

Let X and Y be two absolutely continuous random variables having cdfs $F_X(x)$ and $F_Y(y)$ and pdfs $f_X(x)$ and $f_Y(y)$, respectively. Then the *Farlie–Gumbel–Morgenstern* (FGM) distribution is the bivariate distribution with joint pdf

$$f(x,y) = f_X(x) f_Y(y)[1 + \alpha(2F_X(x) - 1)(2F_Y(y) - 1)], \qquad x \in D_X, y \in D_Y \tag{4.31}$$

with $\alpha \in [-1,1]$ is a correlation parameter (see Chapter 44, Sec. 13 of [45]). If $f_X(x)$ and $f_Y(y)$ are exponential pdfs, then the joint pdf (4.31) reduces to

$$f(x,y) = \lambda\, \mu \exp(-\lambda x - \mu y)[1 + \alpha(2e^{-\lambda x} - 1)(2e^{-\mu y} - 1)] \qquad x > 0, y > 0 \tag{4.32}$$

with $\lambda > 0$, $\mu > 0$. and $-1 \leq \alpha \leq 1$. Eq. (4.32) gives the *Farlie–Gumbel–Morgenstern exponential pdf* (cf. Equation 47.14 of [45]). The marginal distributions are exponentials with parameters λ and μ, respectively. The covariance of X and Y can be written in terms of the parameters α, λ and μ and its calculation gives

$$\mathrm{Cov}(X,Y) = \int_0^\infty \left(x - \frac{1}{\lambda}\right) \left(y - \frac{1}{\mu}\right) f(x,y)\, \mathrm{d}x\, \mathrm{d}y = \frac{\alpha}{4\lambda\mu}. \tag{4.33}$$

The covarentropy of X and Y has the following expression

$$\begin{aligned} \mathrm{Cov}\mathcal{E}(X,Y) &= \int_0^\infty \log f_X(x) \log f_Y(y)\; f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ &- H(X) \cdot H(Y) \end{aligned} \tag{4.34}$$

where $H(X)$ and $H(Y)$ are given by (4.11). If we solve the integrals we obtain that

$$\mathrm{Cov}(X,Y) = \frac{\alpha}{4\lambda\mu}.$$

and

$$\mathrm{Cov}\mathcal{E}(X,Y) = \frac{\alpha}{4}.$$

A special case is given when $\lambda = \frac{1}{\mu}$ (see Figure 4.6).

Theorem 4.7 *Let X and Y be two random variables having exponential FGM pdf (4.32) with correlation parameter $\alpha \neq 0$. Then*

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}\mathcal{E}(X,Y) \iff \lambda = \frac{1}{\mu}$$

Proof. The proof is immediate.

Another case is when $f_X(x)$ and $f_Y(y)$ are logistic with parameters μ_X and μ_Y, σ_X and σ_Y, respectively, then the joint pdf (4.31) reduces to

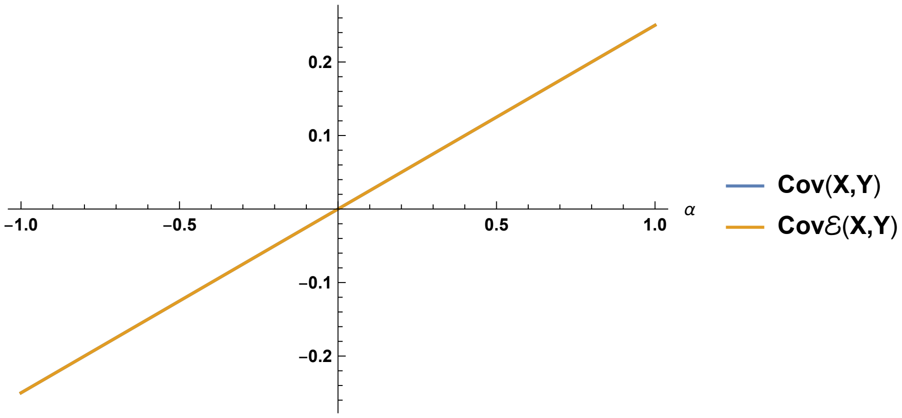
$$f(x,y) = \frac{e^{-z_X - z_Y}}{\sigma_X\sigma_Y(1+e^{-z_X})^3(1+e^{-z_y})^3} \cdot [1+\alpha+(e^{-z_X}+e^{-z_y})(1-\alpha)+e^{-z_X-z_Y}(1+\alpha)] \quad (4.35)$$

where

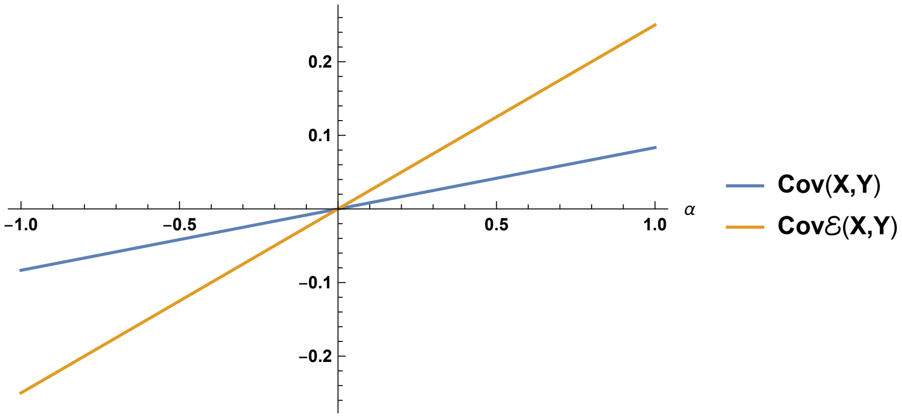
$$z_X = \frac{x-\mu_X}{\sigma_X}, \qquad z_Y = \frac{y-\mu_Y}{\sigma_Y} \qquad x \in \mathbb{R}, y \in \mathbb{R}.$$

Equation (4.35) gives the *Farlie-Gumbel-Morgenstern logistic bivariate pdf* (cf. Chapter 51, Sec. 4 of [45]). The covariance of X and Y is given by

$$\mathrm{Cov}(X,Y) = \int_{\mathbb{R}^2} (x-\mu_X)\ (y-\mu_Y)\ f(x,y)\ \mathrm{d}x\ \mathrm{d}y = \alpha\ \sigma_X\sigma_Y$$



(a) $\lambda = \frac{1}{\mu}$



(b) $\lambda = 1$; $\mu = 3$

Figure 4.6: Covariance and covarentropy calculated for a FGM exponential distribution for different choices of λ and μ (cf. Equations (4.33) and (4.34)).

tht is, the covariance is linear in α. The covarentropy of X and Y gives the result

$$\begin{aligned} \mathrm{Cov}\mathcal{E}(X,Y) &= \int_{\mathbb{R}} \log f_X(x) \int_{\mathbb{R}} \log f_Y(y)\; f(x,y)\; \mathrm{d}y\; \mathrm{d}x \\ &- \; H(X) \cdot H(Y) = 0. \end{aligned} \tag{4.36}$$

where the values of $H(X)$ and $H(Y)$ are given by the Eqs. (4.28). The result (4.36) can be synthesized by the following proposition.

Proposition 4.3 *Let X and Y be two logistic FGM distributions, then* $\mathrm{Cov}\mathcal{E}(X,Y)$ *vanishes for all* $\alpha \in [0,1]$.

Conclusions

In this thesis we discussed some mathematical properties and applications of information measures for continuous random variables. Among these quantities this thesis concerns entropy, the varentropy and the covarentropy. The differential entropy (4) is largely used in information theory and other related areas, being the analogue of the Shannon entropy for a continuous random variable. It constitutes the expected value of the information content (3), whereas its variance is given by the varentropy (2.1) and the covariance of two information contents is given by the covarentropy. The varentropy is useful to assess the effectiveness of the differential entropy as a measure of the information content of a random system, while the covarentropy provide the correlations of the information content of two random variables.

Motivated by possible application in reliability theory and survival analysis, in the first part of the thesis we investigated the residual varentropy, i.e. the varentropy of the residual lifetime distribution. Together with the residual entropy, this measure allows to analyse the dynamical information content of time-varying systems conditional on being active at current time. We discussed various properties, with connections to the generalized hazard rate, the effect of linear transformations, and a suitable lower bound that involves the variance residual life function. We also addressed the use of the residual varentropy in connection with classical distributions, and within some applications concerning the proportional hazards model and the first-passage time problem of an Ornstein-Uhlenbeck jump-diffusion process with catastrophes. Finally we applied kernel varentropy estimation to obtain residual varentropy from a given set of data.

In the second part of the thesis we obtained some results about the residual entropy and varentropy making use of stochastic comparisons, such as usual stochastic order, dispersive order, star order and kurtosis order. In order to do this, we concentrated our attention on the pdf-related, that are the distribution induced by probability density function. The investigation of their properties led us to study quantile functions under suitable assumptions of monotonicity and unimodality. Other properties of pdf-related distributions in connection to stochastic order were studied making use of mapping of quantiles and using the properties of the decreasing rearrangement of a given pdf. The final result is the comparison of differential entropy and differential varentropy of random lifetimes.

In the final part of the thesis we obtained some results for the covarentropy. The covarentropy was applied both to discrete and continuous distribution and different results involving the independence and the correlation of random variables were proved. Also specific examples were illustrated for different bivariate distributions: Gumbel exponential, Normal bivariate,Gumbel–Malik-Abraham logistic, Farlie–Gumbel–Morgenstern exponential and logistic. In some of the cases we noted relations between covariance and covarentropy, in other the covarentropy is independent of the parameters of the distribution.

Future developments will be oriented to applications of the information measures to other stochastic models of interest (such as order statistics, spacings, record values, inaccuracy measures based on the relevation transform and its reversed version), and to provide further improvements to the empirical version of the residual varentropy. Other developments will involve also the application of the stochastic order, for which the future investigation in this field will be oriented to the study of residual entropy and residual varentropy. Finally, future developments can be oriented to analyze the dependence of the covarentropy on the copula of the random vector (X, Y), and to compare the covarentropy of vectors with identical marginals but with different copulas.

Appendix A

Entropy and varentropy for selected distributions

In this appendix some results about entropy and varentropy for ordinary and residual lifetime distributions are given.

A.1 Uniform distributions

$X_{a,b} \sim U(a,b)$ $(a, b > 0,\ a < b)$

Pdf

$$f_{a,b}(x) = \frac{1}{b-a} \qquad x \in S_X = (a,b) \tag{A.1}$$

Survival function

$$\overline{F}_{a,b}(t) = \int_t^b \frac{1}{b-a}\,\mathrm{d}x = \frac{b-t}{b-a}, \qquad t \in [a,b]$$

Differential entropy

$$H(X_{a,b}) = -\int_a^b \frac{1}{b-a}\log\left(\frac{1}{b-a}\right)\,\mathrm{d}x = \log(b-a)$$

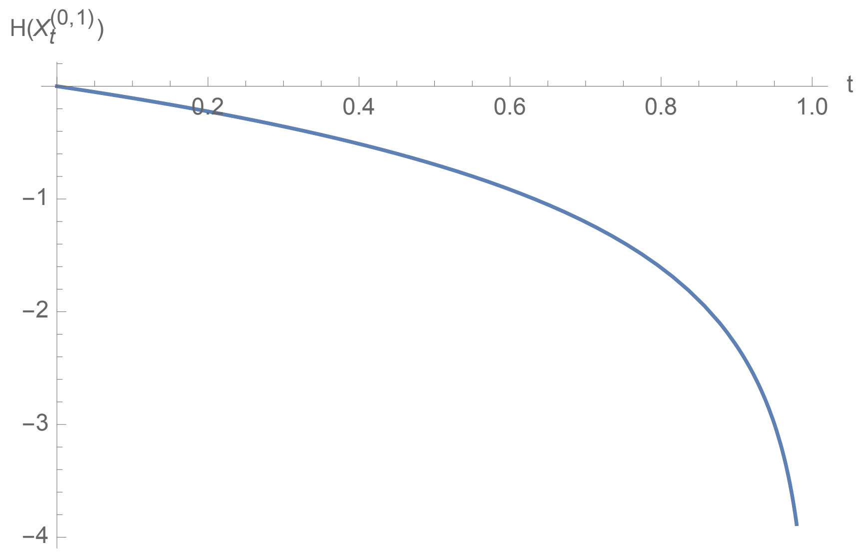


Figure A.1: Residual entropy of a distribution having a uniform pdf (A.1) with $a = 0$ and $b = 1$ (cf. Eq. (A.2)).

Differential varentropy

$$
\begin{aligned}
V(X_{a,b}) &= \int_a^b \frac{1}{b-a}\left[\log\left(\frac{1}{b-a}\right)\right]^2 \mathrm{d}x - [\log(b-a)]^2 \\
&= [\log(b-a)]^2 - [\log(b-a)]^2 = 0
\end{aligned}
$$

Residual entropy

$$
H(X_t^{(a,b)}) = -\int_t^b \frac{1}{b-t}\log\left(\frac{1}{b-t}\right) \mathrm{d}x = \log(b-t) \qquad t \in S_X \tag{A.2}
$$

Residual varentropy

$$
\begin{aligned}
V(X_t^{(a,b)}) &= -\int_t^b \frac{1}{b-t}\left[\log\left(\frac{1}{b-t}\right)\right]^2 \mathrm{d}x - [\log(b-t)]^2 \\
&= [\log(b-t)]^2 - [\log(b-t)]^2 = 0 \qquad t \in S_X
\end{aligned}
$$

A.2 Decreasing exponential distributions

$X_\lambda \sim \mathrm{Exp}(\lambda)$ $(\lambda > 0)$

Pdf

$$f_\lambda(x) = \lambda e^{-\lambda x}, \qquad x \in S_X = (0, \infty)$$

Survival function

$$\overline{F}_\lambda(t) = \int_t^\infty e^{-\lambda x}\ \mathrm{d}x = e^{-\lambda t}, \qquad t \in [0, \infty)$$

Differential entropy

$$H(X_\lambda) = -\int_0^\infty e^{-\lambda x}(-\lambda x + \log\lambda)\ \mathrm{d}x = 1 - \log\lambda$$

Differential varentropy

$$V(X_\lambda) = \int_0^\infty e^{-\lambda x}(-\lambda x + \log\lambda)^2\ \mathrm{d}x - (1 - \log\lambda)^2 = 1$$

Residual entropy

$$\begin{aligned} H(X_t^{(\lambda)}) &= -\int_t^\infty \lambda e^{-\lambda(x-t)}[-\lambda(x-t) + \log\lambda]\ \mathrm{d}x \\ &= -\int_0^\infty e^{-\lambda x}(-\lambda x + \log\lambda)\ \mathrm{d}x = H(X) = 1 - \log\lambda \qquad t \in S_X \end{aligned}$$

Residual varentropy

$$\begin{aligned} V(X_t^{(\lambda)}) &= \int_t^\infty \lambda e^{-\lambda(x-t)}[-\lambda(x-t) + \log\lambda]^2\ \mathrm{d}x - (1 - \log\lambda)^2 \\ &= \int_0^\infty e^{-\lambda x}(-\lambda x + \log\lambda)^2\ \mathrm{d}x - (1 - \log\lambda)^2 = V(X) = 1 \qquad t \in S_X \end{aligned}$$

A.3 Triangular distribution

$X \sim \text{Triang}(0, 1)$

Pdf

$$f(x) = 2(1-x), \qquad x \in S_X = (0,1) \tag{A.3}$$

Survival function

$$\overline{F}(t) = \int_t^1 2(1-x)\ \mathrm{d}x = (1-t)^2, \qquad t \in [0,1]$$

Differential entropy

$$H(X) = -\int_0^1 2(1-x)\log(1-2x)\ \mathrm{d}x = \frac{1}{2} - \log\left(\frac{1}{2}\right)$$

Differential varentropy

$$V(X) = \int_0^1 2(1-x)\left\{\log[2(1-x)]\right\}^2\ \mathrm{d}x - \left[\frac{1}{2} - \log\left(\frac{1}{2}\right)\right]^2 = \frac{1}{4}$$

Residual entropy

$$H(X_t) = -\int_t^1 \frac{2(1-x)}{(1-t)^2}\log\left[\frac{2(1-x)}{(1-t)^2}\right]\mathrm{d}x = \frac{1}{2} + \log\left(\frac{1-t}{2}\right) \qquad t \in S_X \tag{A.4}$$

Residual varentropy

$$V(X_t) = \int_t^1 \frac{2(1-x)}{(1-t)^2}\left\{\log\left[\frac{2(1-x)}{(1-t)^2}\right]\right\}^2\ \mathrm{d}x - \left[\frac{1}{2} + \log\left(\frac{1-t}{2}\right)\right]^2 = \frac{1}{4} \qquad t \in S_X$$

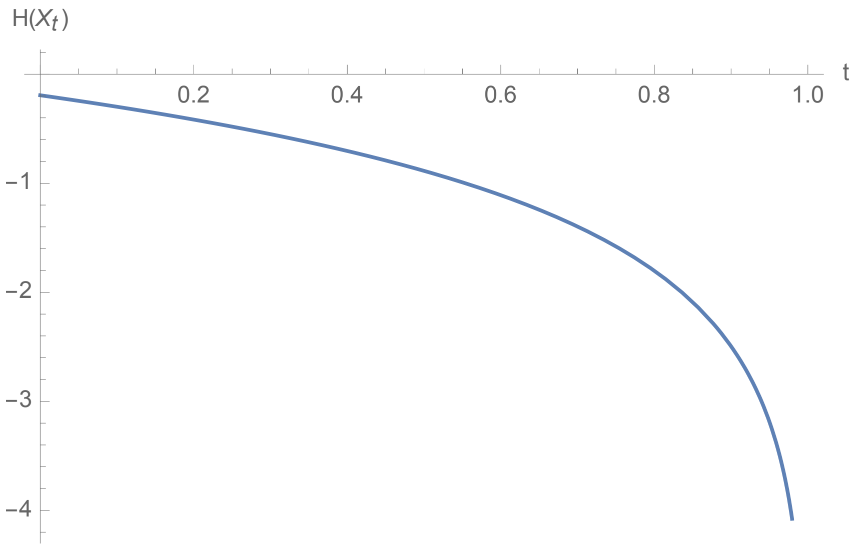


Figure A.2: Residual entropy of a distribution having a triangular pdf (A.3) (cf. Eq. (A.4)).

A.4 Normal distributions

$X_{\mu,\sigma} \sim \mathrm{Norm}(\mu, \sigma)$, ($\mu \in \mathbb{R}$ and $\sigma > 0$)

Pdf

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in S_X = \mathbb{R} \tag{A.5}$$

Survival function

$$\overline{F}_{\mu,\sigma}(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, \mathrm{d}x = \frac{1}{2}\,\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right), \qquad t \in S_X$$

The complementary error function $\mathrm{Erfc}(\cdot)$ is defined as (cf. Eq. 7.1.2 of [1])

$$\mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2}\, \mathrm{d}t = 1 - \mathrm{Erf}(x)$$

Differential entropy

$$\begin{aligned} H(X_{\mu,\sigma}) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left[-\log\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x \\ &= -\log\left(\sqrt{2\pi e}\sigma\right) \end{aligned}$$

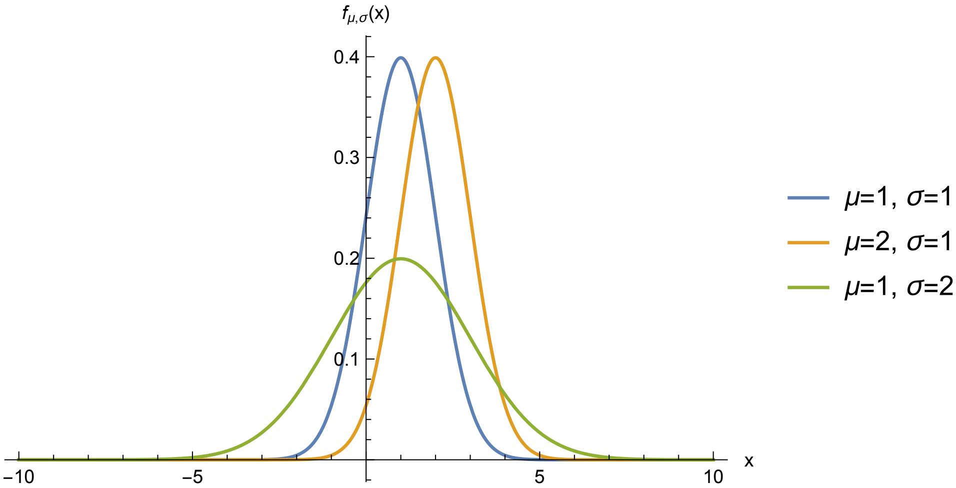


Figure A.3: Pdf of a normal distribution (A.5) for different choices of the parameters μ and σ (as indicated in the labels).

Differential varentropy

$$\begin{aligned} V(X_{\mu,\sigma}) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left[-\log\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]^2 \mathrm{d}x \\ &- \left[\log\left(\sqrt{2\pi e}\sigma\right)\right]^2 = \frac{1}{2} \end{aligned}$$

Residual entropy

$$\begin{aligned} H(X_t^{(\mu,\sigma)}) &= -\frac{2\int_t^{\infty}\left(-\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)\frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x}{\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)} \\ &+ \log\left[\frac{1}{2}\,\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right] \\ &= \frac{(t-\mu)e^{-\frac{(t-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma\,\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)} + \log\left(\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right) + \log\sigma + \frac{1}{2}\left(1+\log\left(\frac{\pi}{2}\right)\right) \quad t \geq 0 \end{aligned} \tag{A.6}$$

Residual varentropy

$$
\begin{aligned}
V(X_t^{(\mu,\sigma)}) &= \frac{2\int_t^{\infty}\left(-\log\sigma-\frac{(x-\mu)^2}{2\sigma^2}-\frac{1}{2}\log(2\pi)\right)^2 \frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx}{\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)} \\
&- \frac{4\left(\int_t^{\infty}\left(-\log\sigma-\frac{(x-\mu)^2}{2\sigma^2}-\frac{1}{2}\log(2\pi)\right)\frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx\right)^2}{\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2} \\
&+ \log\left[\frac{1}{2}\ \mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)\right] \\
&= \frac{1}{4}\left(\frac{(t-\mu)e^{-\frac{(t-\mu)^2}{\sigma^2}}\left(\sqrt{2\pi}e^{\frac{(t-\mu)^2}{2\sigma^2}}\left(\sigma^2+(t-\mu)^2\right)\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)+2\sigma(\mu-t)\right)}{\pi\sigma^3\mathrm{Erfc}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2}+2\right) \\
&\qquad t\geq 0
\end{aligned}
\tag{A.7}
$$

A.5 Weibull distributions

$X_{k,\lambda}\sim \mathrm{Weibull}(k,\lambda)$

Pdf

$$
f_{k,\lambda}(x)=\frac{k}{\lambda}\,x^{k-1}\,e^{-\left(\frac{x}{\lambda}\right)^k},\qquad x\in S_X=(0,+\infty) \tag{A.8}
$$

Survival function

$$
\overline{F}_{k,\lambda}(t)=\int_t^{\infty} f_{k,\lambda}(x)\ \mathrm{d}x=e^{-\left(\frac{t}{\lambda}\right)^k},\qquad t\in[0,+\infty)
$$

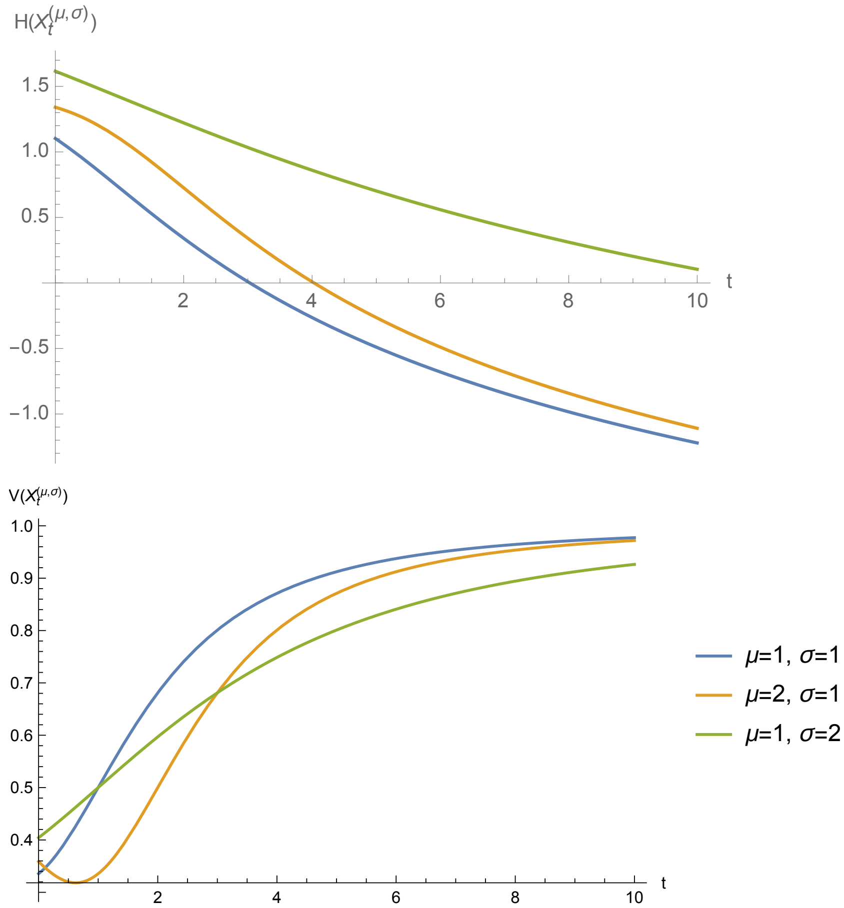


Figure A.4: (top) Residual entropy and (bottom) residual varentropy for the normal distributions in Fig. A.3 (cf. Eqs. (A.6) and (A.7)).

Differential entropy

$$\begin{aligned} H(X_{k,\lambda}) &= -\int_0^{\infty} \frac{k}{\lambda}\, x^{k-1}\, e^{-\left(\frac{x}{\lambda}\right)^k} \left[\log k - k\log\lambda + (k-1)\log x - \left(\frac{x}{\lambda}\right)^k\right]\, \mathrm{d}x \\ &= -\frac{\gamma}{k} - \log k + \log\lambda + \gamma + 1 \end{aligned}$$

The constant $\gamma = 0.57721\ldots$ is the Eulero–Mascheroni constant.

Differential varentropy

$$\begin{aligned} V(X_{k,\lambda}) &= \int_0^{\infty} \frac{k}{\lambda}\, x^{k-1}\, e^{-\left(\frac{x}{\lambda}\right)^k} \left[\log k - k\log\lambda + (k-1)\log x - \left(\frac{x}{\lambda}\right)^k\right]^2\, \mathrm{d}x \\ &- \left[-\frac{\gamma}{k} - \log k + \log\lambda + \gamma + 1\right]^2 \\ &= \frac{(\pi^2-6)\,(k-2)k + \pi^2}{6k^2} \end{aligned}$$

Residual entropy

$$\begin{aligned} H(X_t^{(k,\lambda)}) &= -\left(\frac{t}{\lambda}\right)^k \\ &- \frac{1}{e^{-\left(\frac{t}{\lambda}\right)^k}} \int_t^{\infty} \frac{k}{\lambda}\, x^{k-1}\, e^{-\left(\frac{x}{\lambda}\right)^k} \left[\log k - k\log\lambda + (k-1)\log x - \left(\frac{x}{\lambda}\right)^k\right]\, \mathrm{d}x \\ &= \frac{(k-1)e^{\left(\frac{t}{\lambda}\right)^k}\mathrm{Ei}\left(-\left(\frac{t}{\lambda}\right)^k\right)}{k} + k\log\lambda - k\log t - \log k + \log t + 1 \\ & \qquad t \in S_X \end{aligned} \tag{A.9}$$

The exponential integral function $\mathrm{Ei}(\cdot)$ is defined in Eq. (4.15)).

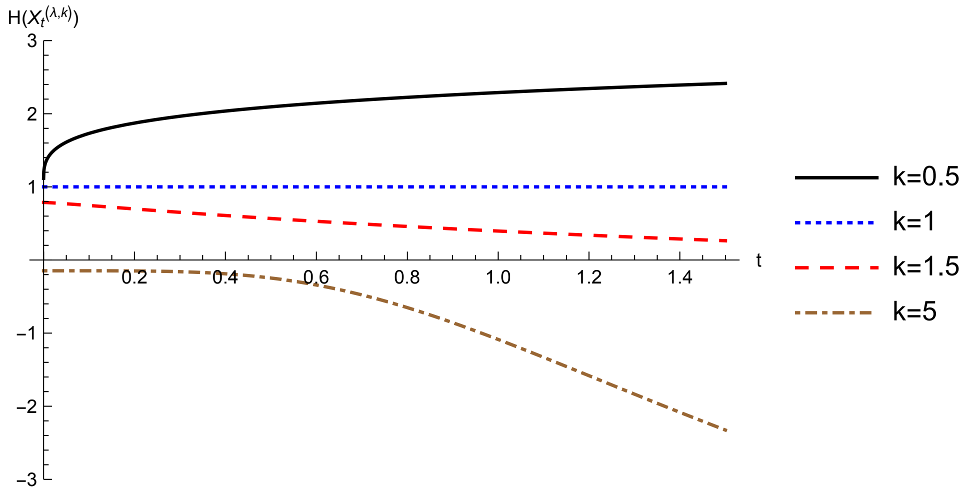


Figure A.5: Residual entropy of the Weibull distributions of Fig. 2.3 (cf. Eq. (A.9)).

Residual varentropy

$$
\begin{aligned}
V(X_t^{(k,\lambda)}) &= \frac{1}{e^{-\left(\frac{t}{\lambda}\right)^k}} \int_t^{\infty} \frac{k}{\lambda}\, x^{k-1}\, e^{-\left(\frac{x}{\lambda}\right)^k} \left[\log k - k\log\lambda + (k-1)\log x - \left(\frac{x}{\lambda}\right)^k\right]^2 \,\mathrm{d}x \\
&- \left[-\frac{\gamma}{k} - \log k + \log\lambda + \gamma + 1 - \left(\frac{t}{\lambda}\right)^k\right]^2 \\
&= \frac{\lambda^{-2k}}{6k^2}\Bigg\{ -6k^2\lambda^{2k} + 6\gamma^2k^2\lambda^{2k}e^{\left(\frac{t}{\lambda}\right)^k} + 12k\lambda^{2k} - 12\gamma^2k\lambda^{2k}e^{\left(\frac{t}{\lambda}\right)^k} + 6\gamma^2\lambda^{2k}e^{\left(\frac{t}{\lambda}\right)^k} \\
&+ 12ke^{\left(\frac{t}{\lambda}\right)^k}(\lambda t)^k \mathrm{Ei}\left(-\left(\frac{t}{\lambda}\right)^k\right) + \pi^2k^2\lambda^{2k}e^{\left(\frac{t}{\lambda}\right)^k} + \pi^2\lambda^{2k}e^{\left(\frac{t}{\lambda}\right)^k} - 2\pi^2k\lambda^{2k}e^{\left(\frac{t}{\lambda}\right)^k} \\
&- 12k^2e^{\left(\frac{t}{\lambda}\right)^k}(\lambda t)^k \mathrm{Ei}\left(-\left(\frac{t}{\lambda}\right)^k\right) - 6\lambda^{2k}e^{2\left(\frac{t}{\lambda}\right)^k}\mathrm{Ei}\left(-\left(\frac{t}{\lambda}\right)^k\right)^2 \\
&+ 24ke^{\left(\frac{t}{\lambda}\right)^k}(\lambda t)^k\,{}_3F_3\left(1,1,1;2,2,2;-\left(\frac{t}{\lambda}\right)^k\right) \\
&+ 12k\lambda^{2k}e^{2\left(\frac{t}{\lambda}\right)^k}\mathrm{Ei}\left(-\left(\frac{t}{\lambda}\right)^k\right)^2 - 12e^{\left(\frac{t}{\lambda}\right)^k}(\lambda t)^k\,{}_3F_3\left(1,1,1;2,2,2;-\left(\frac{t}{\lambda}\right)^k\right) \\
&- 6k^2\lambda^{2k}e^{2\left(\frac{t}{\lambda}\right)^k}\mathrm{Ei}\left(-\left(\frac{t}{\lambda}\right)^k\right)^2 + 12\gamma k\lambda^{2k}(\log t)e^{\left(\frac{t}{\lambda}\right)^k} \\
&- 12k^2e^{\left(\frac{t}{\lambda}\right)^k}(\lambda t)^k\,{}_3F_3\left(1,1,1;2,2,2;-\left(\frac{t}{\lambda}\right)^k\right) \\
&- 12k^3\lambda^{2k}(\log t)^2e^{\left(\frac{t}{\lambda}\right)^k} + 6k^2\lambda^{2k}(\log t)^2e^{\left(\frac{t}{\lambda}\right)^k} + 12\gamma k^3\lambda^{2k}(\log t)e^{\left(\frac{t}{\lambda}\right)^k} \\
&+ 24\gamma k^2\lambda^{2k}(\log t)e^{\left(\frac{t}{\lambda}\right)^k} + 6k^4\lambda^{2k}(\log t)^2e^{\left(\frac{t}{\lambda}\right)^k} + 24\gamma k^2\lambda^{2k}(\log\lambda)e^{\left(\frac{t}{\lambda}\right)^k} \\
&- 12\gamma k\lambda^{2k}(\log\lambda)e^{\left(\frac{t}{\lambda}\right)^k} - 12\gamma k^3\lambda^{2k}(\log\lambda)e^{\left(\frac{t}{\lambda}\right)^k} + 24k^3\lambda^{2k}(\log\lambda)(\log t)e^{\left(\frac{t}{\lambda}\right)^k} \\
&- 12k^2\lambda^{2k}(\log\lambda)(\log t)e^{\left(\frac{t}{\lambda}\right)^k} + 6k^4\lambda^{2k}(\log\lambda)^2e^{\left(\frac{t}{\lambda}\right)^k} - 12k^3\lambda^{2k}(\log\lambda)^2e^{\left(\frac{t}{\lambda}\right)^k} \\
&+ 6k^2\lambda^{2k}(\log\lambda)^2e^{\left(\frac{t}{\lambda}\right)^k} - 12k^4\lambda^{2k}(\log\lambda)(\log t)e^{\left(\frac{t}{\lambda}\right)^k}\Bigg\} \qquad t \in S_X \qquad \text{(A.10)}
\end{aligned}
$$

The generalized hypergeometric function ${}_p\,\mathrm{F}_q(\cdot;\cdot)$ is defined as (see Section 2.1 of [69])

$$
{}_p\,\mathrm{F}_q\left(a_1,\ldots,a_p;b_1,\ldots,b_q;x\right) := \sum_{k=0}^{\infty} \frac{(a_1)_k\cdots(a_p)_k}{(b_1)_k\cdots(b_q)_k}\frac{x^k}{k!}
$$

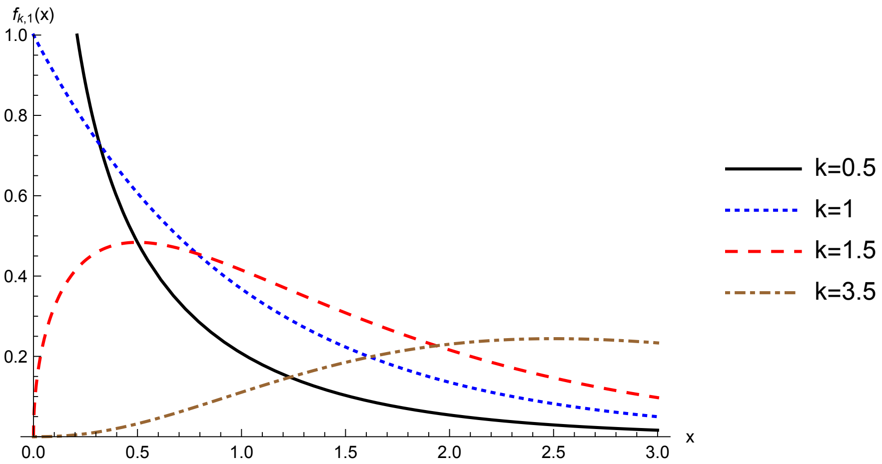


Figure A.6: Gamma pdf (A.11) for some choices of k (as indicated in the labels) and for θ fixed to 1.

where

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)} = a(a+1)\cdots(a+k-1)$$

A.6 Gamma distributions

$X_{k,\theta} \sim \text{Gamma}(k,\theta)$ ($k > 0$ and $\theta > 0$)

Pdf

$$f_{k,\theta}(x) = \frac{x^{k-1}\, e^{-\frac{x}{\theta}}}{\theta^k\, \Gamma(k)}, \qquad x \in S_X = (0,\infty) \tag{A.11}$$

$\Gamma(\cdot)$ is the Gamma function defined in Eq. (4.20).

Survival function

$$\overline{F}_{k,\theta}(t) = \int_t^\infty \frac{x^{k-1}\, e^{-\frac{x}{\theta}}}{\theta^k\, \Gamma(k)}\, \mathrm{d}x = \frac{\Gamma(k,\frac{t}{\theta})}{\Gamma(k)}, \qquad t \geq 0$$

$\Gamma(\cdot,\cdot)$ is the Gamma incomplete function defined in Eq. (4.13).

Differential entropy

$$\begin{aligned} H(X_{k,\theta}) &= -\int_0^\infty \frac{x^{k-1}\ e^{-\frac{x}{\theta}}}{\theta^k\ \Gamma(k)} \left(-k\log\theta + (k-1)\log x - \log\Gamma(k) - \frac{x}{\theta}\right) dx \\ &= \log\theta + k - (k-1)\psi^{(0)}(k) + \log\Gamma(k) \end{aligned}$$

The PolyGamma function of order m $\psi^{(m)}(\cdot)$ is defined in (4.22).

Differential varentropy

$$\begin{aligned} V(X_{k,\theta}) &= \int_0^\infty \frac{x^{k-1}\ e^{-\frac{x}{\theta}}}{\theta^k\ \Gamma(k)} \left(-k\log\theta + (k-1)\log x - \log\Gamma(k) - \frac{x}{\theta}\right)^2 dx \\ &- [\log\theta + k - (k-1)\psi^{(0)}(k) + \log\Gamma(k)]^2 = -k + (k-1)^2\psi^{(1)}(k) + 2 \end{aligned}$$

Residual entropy

$$\begin{aligned} H(X_t^{(k,\theta)}) &= -\frac{\Gamma(k)}{\Gamma\left(k,\frac{t}{\theta}\right)} \int_t^\infty \frac{x^{k-1}\ e^{-\frac{x}{\theta}}}{\theta^k\ \Gamma(k)} \left(-k\log\theta + (k-1)\log x - \log\Gamma(k) - \frac{x}{\theta}\right) dx \\ &+ \log\Gamma(k,\theta) - \log\Gamma(k) \\ &= \frac{1}{\Gamma\left(k,\frac{t}{\theta}\right)} - (k-1)G_{2,3}^{3,0}\left(\begin{array}{c} 1,1 \\ 0,0,k \end{array}\middle| \frac{t}{\theta}\right) \\ &+ \Gamma\left(k+1,\frac{t}{\theta}\right) + (k-1)\Gamma(k)\left(-\log\theta - \log\left(\frac{t}{\theta}\right) + \log t\right) \\ &+ \Gamma\left(k,\frac{t}{\theta}\right)\left(k\log\theta + \log\left(\Gamma\left(k,\frac{t}{\theta}\right)\right) - k\log t + \log t\right) \qquad t \in S_X \end{aligned}$$

Meijer's G-function $G_{p,q}^{m,n}(\cdot|\cdot)$ (see section 5.3 of [8]) is defined as

$$G_{p,q}^{m,n}\left(\begin{array}{c} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{array}\middle| z\right) = \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j - s)\prod_{j=1}^n \Gamma(1 - a_j + s)}{\prod_{j=m+1}^q \Gamma(1 - b_j + s)\prod_{j=n+1}^p \Gamma(a_j - s)} z^s ds \qquad \text{(A.12)}$$

and is a complex line integral. The path L of integration of (A.12) is determined in a specific way for which we remind to section 5.3 of ref. [8].

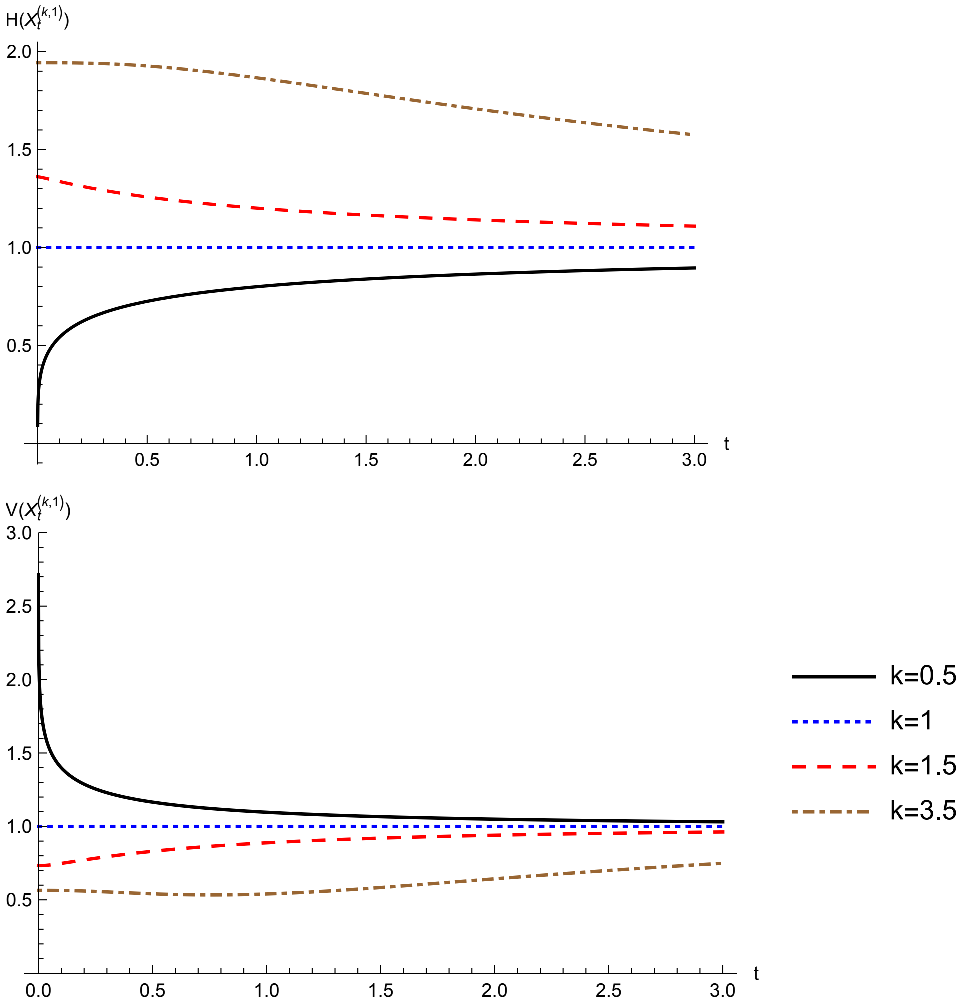


Figure A.7: Residual entropy (top) and residual varentropy (bottom) for the Gamma distributions of Fig. A.6.

Residual varentropy

$$V(X_t^{(k,\theta)}) = \frac{\Gamma(k)}{\Gamma\left(k,\frac{t}{\theta}\right)} \int_t^{\infty} \frac{x^{k-1}\, e^{-\frac{x}{\theta}}}{\theta^k\, \Gamma(k)} \left(-k\log\theta + (k-1)\log x - \log\Gamma(k) - \frac{x}{\theta}\right)^2 dx$$
$$- \left(\frac{\Gamma(k)}{\Gamma\left(k,\frac{t}{\theta}\right)} \int_t^{\infty} \frac{x^{k-1}\, e^{-\frac{x}{\theta}}}{\theta^k\, \Gamma(k)} \left(-k\log\theta + (k-1)\log x - \log\Gamma(k) - \frac{x}{\theta}\right) dx\right)^2 \quad t \in S_X \tag{A.13}$$

It was not possible to do an analytical computation of (A.13). Nevertheless, we give a numerical plot in the bottom side of Fig. A.7.

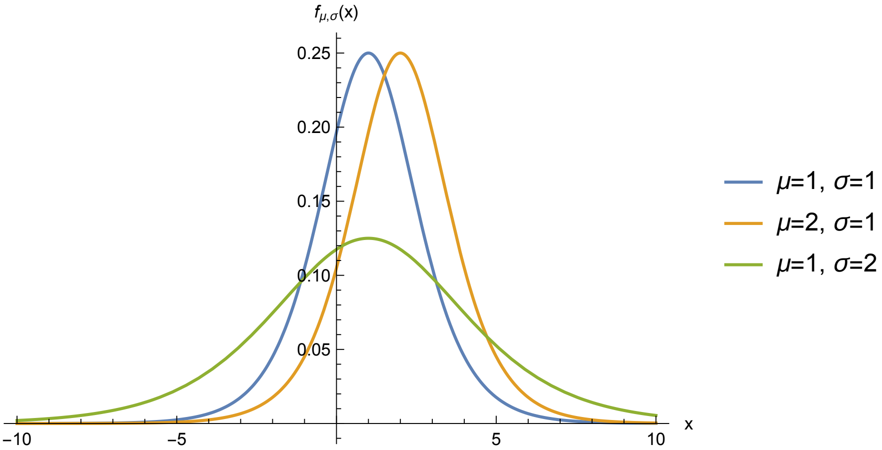


Figure A.8: Logistic pdf (A.14) for some choices of μ and σ (as indicated in the label).

A.7 Logistic distributions

$X_{\mu,\sigma} \sim \text{Logistic}(\mu, \sigma)$ ($\mu > 0$ and $\sigma > 0$)

Pdf

$$f_{\mu,\sigma}(x) = \frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{-\frac{x-\mu}{\sigma}}\right)^2}, \qquad x \in S_X = \mathbb{R} \tag{A.14}$$

Survival function

$$\overline{F}_{\mu,\sigma}(t) = \int_{\frac{t-\mu}{\sigma}}^{\infty} \frac{e^{-z}}{(1+e^{-z})^2}\, \mathrm{d}z = \frac{e^{-\frac{t-\mu}{\sigma}}}{1+e^{-\frac{t-\mu}{\sigma}}}, \qquad t \in S_X$$

Differential entropy

$$\begin{aligned} H(X_{\mu,\sigma}) &= -\int_{-\infty}^{\infty} \frac{e^{-z}}{(1+e^{-z})^2} \left[-z - \log\sigma - 2\log\left(1+e^{-z}\right)\right]\, \mathrm{d}z \\ &= 2 + \log\sigma \end{aligned}$$

Differential varentropy

$$\begin{aligned} V(X_{\mu,\sigma}) &= \int_{-\infty}^{\infty} \frac{e^{-z}}{(1+e^{-z})^2} \left[-z - \log\sigma - 2\log\left(1+e^{-z}\right)\right]^2\, \mathrm{d}z - (2+\log\sigma)^2 \\ &= 4 - \frac{\pi^2}{3} \end{aligned}$$

Residual entropy

$$\begin{aligned} H(X_t^{(\mu,\sigma)}) &= -\frac{t-\mu}{\sigma} - \left(1+e^{\frac{t-\mu}{\sigma}}\right)\int_{\frac{t-\mu}{\sigma}}^{\infty} \frac{e^{-z}}{(1+e^{-z})^2}\left[-z-\log\sigma - 2\log\left(1+e^{-z}\right)\right]\,\mathrm{d}z \\ &= \frac{1}{\sigma}\Bigg[-\mu+2\sigma+(t-\mu)e^{\frac{t-\mu}{\sigma}} \\ &+ \sigma\left(\log\sigma + \left(-e^{\frac{t-\mu}{\sigma}}-1\right)\log\left(e^{\frac{t-\mu}{\sigma}}+1\right)+\log\left(e^{\frac{\mu-t}{\sigma}}+1\right)\right)+t\Bigg] \qquad t\geq 0 \end{aligned} \tag{A.15}$$

Residual varentropy

$$\begin{aligned} V(X_t^{(\mu,\sigma)}) &= \left(1+e^{\frac{t-\mu}{\sigma}}\right)\int_{\frac{t-\mu}{\sigma}}^{\infty} \frac{e^{-z}}{(1+e^{-z})^2}\left[-z-\log\sigma-2\log\left(1+e^{-z}\right)\right]^2\,\mathrm{d}z \\ &= \frac{e^{-\frac{2\mu}{\sigma}}}{3\sigma^2} \\ &\times \Bigg\{-3(t-\mu)^2e^{\frac{2t}{\sigma}} + e^{\frac{\mu+t}{\sigma}}\left(13\pi^2\sigma^2 - 24i\pi\sigma(t-\mu)-21(t-\mu)^2\right) \\ &+ e^{\frac{2\mu}{\sigma}}\left(\left(12+13\pi^2\right)\sigma^2-24i\pi\sigma(t-\mu)-18(t-\mu)^2\right) \\ &- 3\sigma\left(e^{\mu/\sigma}+e^{t/\sigma}\right)\Bigg[\sigma\left(e^{t/\sigma}-e^{\mu/\sigma}\right)\left[\log\left(e^{\frac{t-\mu}{\sigma}}+1\right)\right]^2 \\ &- 2(t-\mu)\left(e^{\mu/\sigma}+e^{t/\sigma}\right)\log\left(e^{\frac{t-\mu}{\sigma}}+1\right) \\ &- 4e^{\mu/\sigma}(\mu-4i\pi\sigma-t)\log\left(e^{\frac{\mu-t}{\sigma}}+1\right)\Bigg] + 6\sigma^2e^{\mu/\sigma}\left(e^{\mu/\sigma}+e^{t/\sigma}\right) \\ &\times \left[\mathrm{Li}_2\left(-e^{\frac{t-\mu}{\sigma}}\right)+2\mathrm{Li}_2\left(-e^{\frac{\mu-t}{\sigma}}\right)-4\left(\mathrm{Li}_2\left(1+e^{\frac{t-\mu}{\sigma}}\right)+\mathrm{Li}_2\left(1+e^{\frac{\mu-t}{\sigma}}\right)\right)\right] \\ &\qquad t\geq 0 \end{aligned}$$

The dilogarithm function $\mathrm{Li}_2(\cdot)$ is defined as (see [56])

$$\mathrm{Li}_2(z) := -\int_0^x \frac{\mathrm{d}z}{z}\log(1-z)$$

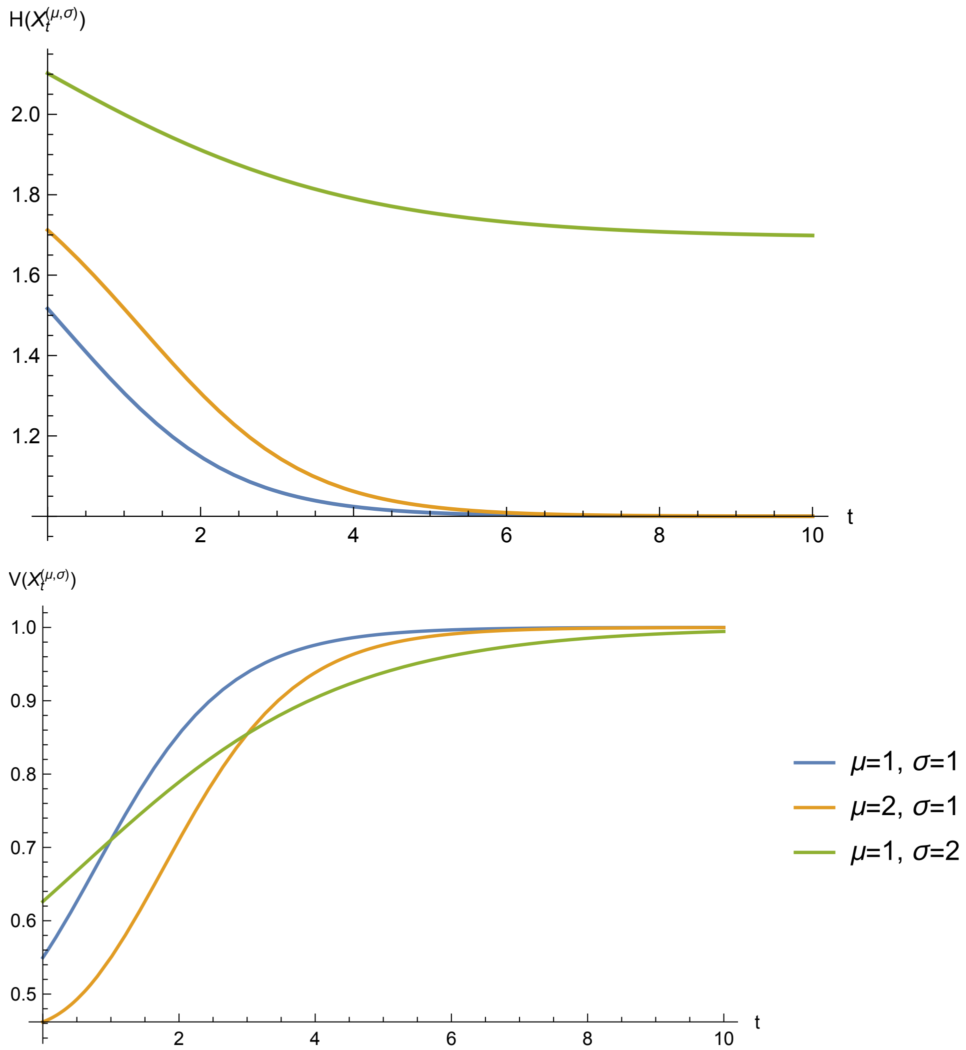


Figure A.9: Residual entropy (top) and residual varentropy (bottom) for the logistic distributions of Fig. A.8 (cf. Eqs. (A.15) and (A.16)).

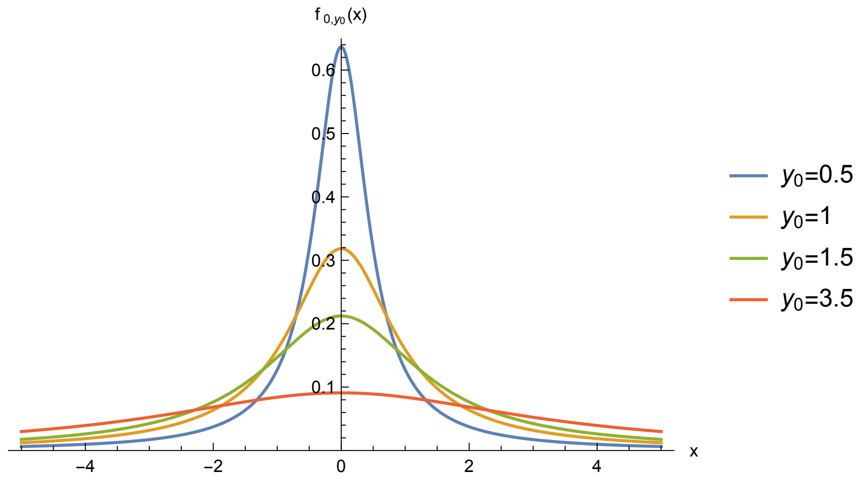


Figure A.10: Cauchy pdf (A.16) for $x_0 = 0$ and for some choices of y_0 (as indicated in the label).

A.8 Cauchy distributions

$X_{x_0,y_0} \sim \text{Cauchy}(x_0, y_0)$ $(x_0 \in \mathbb{R}, y_0 > 0)$

Pdf

$$f_{x_0,y_0}(x) = \frac{1}{\pi} \frac{y_0}{(x-x_0)^2 + y_0^2}, \qquad x \in S_X = \mathbb{R} \tag{A.16}$$

Survival function

$$\overline{F}_{x_0,y_0}(t) = \int_{t-x_0}^{\infty} \frac{1}{\pi} \frac{y_0}{z^2 + y_0^2} \, \mathrm{d}z = -\frac{1}{\pi} \arctan\left(\frac{t - x_0}{y_0}\right) + \frac{1}{2}, \qquad t \in S_X$$

Differential entropy

$$H(X_{x_0,y_0}) = -\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{y_0}{z^2 + y_0^2} \left(-\log \pi + \log y_0 - \log\left(z^2 + y_0^2\right)\right) \, \mathrm{d}z = \log\left(4\pi y_0\right)$$

Differential varentropy (numerical estimation)

$$V(X_{x_0,y_0}) = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{y_0}{z^2+y_0^2}\left(-\log\pi + \log y_0 - \log\left(z^2+y_0^2\right)\right)^2 \, \mathrm{d}z - [\log(4\pi y_0)]^2 = 3.28987\ldots$$

Residual entropy

$$\begin{aligned}
H(X_t^{(x_0,y_0)}) &= -\log\pi - \log y_0 - \log\left[(t-x_0)^2+y_0^2\right] \\
&- \left(\frac{1}{\pi}\frac{y_0}{(t-x_0)^2+y_0^2}+\frac{1}{2}\right)^{-1} \\
&\times \int_{t-x_0}^{\infty} \frac{1}{\pi}\frac{y_0}{z^2+y_0^2}\left(-\log\pi + \log y_0 - \log\left(z^2+y_0^2\right)\right) \, \mathrm{d}z \qquad t \geq 0
\end{aligned} \tag{A.17}$$

Residual varentropy

$$\begin{aligned}
V(X_t^{(x_0,y_0)}) &= \left(\frac{1}{\pi}\frac{y_0}{(t-x_0)^2+y_0^2}+\frac{1}{2}\right)^{-1} \\
&\times \int_{t-x_0}^{\infty} \frac{1}{\pi}\frac{y_0}{z^2+y_0^2}\left(-\log\pi + \log y_0 - \log\left(z^2+y_0^2\right)\right)^2 \, \mathrm{d}z \\
&- \left[\left(\frac{1}{\pi}\frac{y_0}{(t-x_0)^2+y_0^2}+\frac{1}{2}\right)^{-1}\right. \\
&\times \left.\int_{t-x_0}^{\infty} \frac{1}{\pi}\frac{y_0}{z^2+y_0^2}\left(-\log\pi + \log y_0 - \log\left(z^2+y_0^2\right)\right) \, \mathrm{d}z\right]^2 \qquad t \geq 0
\end{aligned} \tag{A.18}$$

The numerical plot of residual entropy and residual varentropy of Cauchy distributions is given in Fig. A.11.

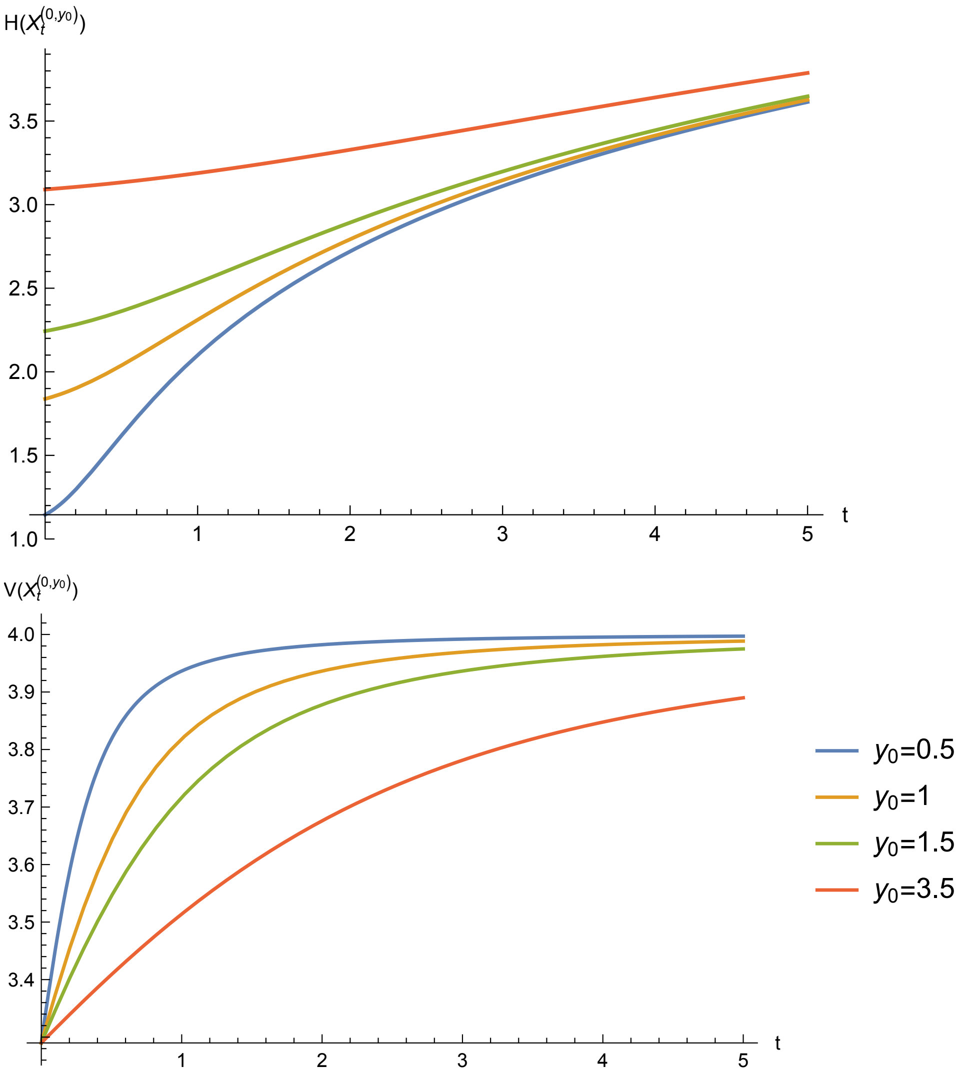


Figure A.11: Residual entropy (top) and residual varentropy (bottom) for the Cauchy distributions of Fig. A.10 (cf. Eqs. (A.17) and (A.18)).

Bibliography

[1] Abramovitz, M., & Stegun, I. Handbook of mathematical functions. *Dover*, New York (1972).

[2] Alimohammadi, M., Esna-Ashari, M., & Cramer, E. On dispersive and star orderings of random variables and order statistics. *Stat. Prob. Lett.* pp. 1–7 (2021).

[3] Arikan, E. Varentropy decreases under polar transform. *IEEE Transactions on Information Theory* 62:3390–3400 (2016).

[4] Arriaza, A., Di Crescenzo, A., Sordo, M. A., & Suárez Llorens, A. Shape measures based on the convex transform order. Metrika 82:99–124 (2019).

[5] Asadi, M., & Ebrahimi, N. Residual entropy and its characterizations in terms of hazard function and mean residual life function. *Statistics and Probability Letters* 49:263–269 (2000).

[6] Azzalini, A. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* 68:326-328 (1981).

[7] Barlow, R.E., & van Zwet, W. Asymptotic properties of isotonic estimators for the generalized failure rate function. I. Strong consistency. In M.L. Puri (ed.), *Nonparametric techniques in statistical inference.* London, Cambridge University Press, pp. 159-176 (1970).

[8] Bateman, H. Higher trascendental functions, Vol. 1. *McGraw Hill*, New York (1953)

[9] Belzunce, F., Candel & J., Ruiz, J. M. Dispersive Ordering and Characterizations of Aging Classes. *Stat. Prob. Lett.* 28:321-327 (1996).

[10] Belzunce, F., Guillamón, A., Navarro, J., & Ruiz, J. M. Kernel estimation of residual entropy. *Comm. Stat. Theory Meth.* 30:1243–1255 (2001).

[11] Belzunce, F., Navarro, J., Ruiz, J.M., & De Aguila, J. Some results on residual entropy function. *Metrica* 59: 147–161 (2004).

[12] Bieniek, M., & Szpak, M. Sharp bounds for the mean of the total time on test for distributions with increasing generalized failure rate. *Statistics* 52:818–828 (2018).

[13] Bobkov, S., & Madiman, M. Concentration of the information in data with log-concave distributions. *Annals of Probability* 39:1528–1543 (2011).

[14] Bowden, R. The information theory of comparisons. *Springer* (2018).

[15] Cacoullos, T., & Papathanasiou, V. Characterizations of distributions by variance bounds. *Stat. Prob. Lett.* 7:351–356 (1989).

[16] Carlitz, L. The inverse of the error function. *Pacific Journal of Mathematics* Vol. 13. 2:459–470 (1963).

[17] Cover, T.M., & Thomas, J.A. Elements of Information Theory. *J. Wiley & Sons*, New York (1991).

[18] Cox, D.R. The analysis of exponentially distributed lifetimes with two types of failure. *Journal of the Royal Statistical Society. Series B* 21:411-421 (1959).

[19] Cufaro Petroni, N. Entropy and its discontents: A note on definitions. *Entropy* 16:4044-4059 (2014).

[20] Dharmaraja, S., Di Crescenzo, A., Giorno, V., & Nobile, A.G. A continuous-time Ehrenfest model with catastrophes and its jump-diffusion approximation. *Journal of Statistical Physics* 161:326–345 (2015).

[21] Di Crescenzo, A., & Longobardi, M. On weighted residual and past entropies. *Scientiae Mathematicae Japonicae* 64:255–266 (2006).

[22] Di Crescenzo, A., & Martinucci, B. A damped telegraph random process with logistic stationary distribution. *Journal of Applied Probability* 47:84–96 (2010).

[23] Di Crescenzo, A., Martinucci, B. & Mulero, J. A quantile-based probabilistic mean value theorem. *Probab. Engrg. Inform. Sci.*, 30:261–280 (2016).

[24] Di Crescenzo, A., & Paolillo, L. Analysis and applications of the residual varentropy of random lifetimes. *Probab. Engrg. Inform. Sci.* pp. 1–19. doi: 10.1017/S0269964820000133 (2020).

[25] Di Crescenzo, A., Paolillo, L., & Suárez-Llorens, A. Stochastic comparisons, differential entropy and varentropy for distributions induced by probability density functions. 21 pp. (2021). *arXiv:2103.11038v1*. Submitted for publication.

[26] Di Crescenzo, A., & Pellerey, F. Some results and applications of geometric counting processes. *Methodology and Computing in Applied Probability* 21:203–233 (2019).

[27] Ebrahimi, N. How to measure uncertainty in the residual life time distribution. *Sankhyā. The Indian Journal of Statistics. Series A* 58:48–56 (1996).

[28] Ebrahimi, N., & Kirmani, S.N.U.A. Some results on ordering of survival functions through uncertainty, *S*tat. Prob. Lett. 29:167–176 (1996).

[29] Ebrahimi, N., & Pellerey, F. New partial ordering of survival functions based on the notion of uncertainty. *Journal of Applied Probability* 32:202–211 (1995).

[30] Fashandi, M., & Ahmadi, J. Characterizations of symmetric distributions based on Rényi entropy. *Stat. Prob. Lett.* 82:798–804 (2012).

[31] Fernández-Ponce J.M., & Suárez-Llorens A. An aging concept based on majorization. *Prob. Engin. Inf. Sci.* 17:107–117 (2003).

[32] Fradelizi, M., Madiman, M., & Wang, L. Optimal concentration of information content for log-concave densities. In Houdré, C., Mason, D., Reynaud-Bouret, P., & Rosiński, J. (eds.), High Dimensional Probability VII. *Progress in Probability*, Cham, Springer, pp. 45-60 (2016).

[33] Goodarzi, F., Amini, M., & Borzadaran, G.R.M. On upper bounds for the variance of functions of the inactivity time. *Applications of Mathematics* 117:62–71 (2016).

[34] Goodarzi, F., Amini, M., & Borzadaran, G.R.M. Characterizations of continous distributions through inequalities involving the expected values of selected functions. *Statistics and Probability Letters* 62:493–507 (2017).

[35] Goodarzi, F., Amini, M., & Borzadaran, G.R.M. Some results on upper bounds for the variance of functions of the residual life random variables. *Journal of computation and applied mathematics* 320:30–42 (2017).

[36] Gupta, R.C. Variance residual life function in reliability studies. *Metron* 64:343–355 (2006).

[37] Hall, P., & Morton, S.C., On the estimation of entropy. *Ann. Inst. Statist. Math.* 45:69–88 (1993).

[38] Hall, W.J., & Wellner, J.A. Mean residual life. In M. Csörgö, D.A. Dawson, J.N.K. Rao & A.K.Md.E. Saleh (eds.), *Statistics and Related Topics*, North-Holland, pp. 169–184 (1981).

[39] Hardy, G.H., Littlewood, & J.E., Polya, G. Some simple inequalities satisfied by convex functions. *Messenger of Math.* 58:145–152 (1929).

[40] Hardy, G.H., Littlewood, & J.E., Polya, G. Inequalities. *Cambridge University Press*, Cambridge (1934).

[41] Hickey, R. J. Continuous majorization and randomness *J. Appl. Prob.* 21:924–929 (1984).

[42] Hickey, R.J. Concepts of dispersion in distributions: a comparative note *J. Appl. Prob.* 23:914–921 (1986).

[43] Kontoyiannis, I., & Verdú, S. Optimal lossless compression: source varentropy and dispersion. *IEEE International Symposium on Information Theory*, Istanbul, pp. 1739–1743 (2013).

[44] Kontoyiannis, I. & Verdú, S. Optimal lossless data compression: non-asymptotics and asymptotics. *IEEE Transactions on Information Theory* 60:777–795 (2014).

[45] Kotz S., Balakrishnan N. & Jonson N.L. Continuous Multivariate Distributions Vol. 1, *Wiley & Sons* (2000).

[46] Kusmierz, L., Majumdar, S.N., Sabhapandit, S., & Schehr, G. First order transition for the optimal search time of Lévy flights with resetting. *Physical Review Letters* 113:220602 (2014).

[47] Lariviere, M.A., & Porteus, E.L. Selling to the newsvendor: an analysis of price-only contracts. *Manufacturing & Service Operations Management* 3:293–305 (2001).

[48] Li, J., Fradelizi, M., & Madiman, M. Information concentration for convex measures. *IEEE International Symposium on Information Theory*, Barcelona, pp. 1128–1132 (2016).

[49] Li, Z., & Tewari, A. Beyond the hazard rate: more perturbation algorithms for adversarial multi-armed bandits. *Journal of Machine Learning Research* 18:1–24 (2018).

[50] Maadani, S., Mohtashami Borzadaran, G. R. & Rezaei Roknabadi, A. H. A new generalized varentropy and its properties. *Ural Mathematical Journal* 6, 1:114–129 (2020).

[51] Maadani, S., Mohtashami Borzadaran, G. R. & Rezaei Roknabadi, A. H. Varentropy of order statistics and some stochastic comparisons. *Communications in Statistics - Theory and Methods* 6, 16 pp. (2021).

[52] Madiman, M., & Barron, A. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Transactions on Information Theory* 53:2317–2329 (2007).

[53] Madiman, M., & Wang, L. An optimal varentropy bound for log-concave distributions. *International Conference on Signal Processing and Communications* (SPCOM), Bangalore , 1 p. (2014).

[54] Maoui, I., Ayhan, H., & Foley, R. Congestion-dependent pricing in a stochastic service system. *Advances in Applied Probability* 39:898–921 (2007).

[55] Merovci, F., Khaleel, M.A. Ibrahim, N.A. & Shitan, M. The beta Burr type X distribution properties with application. *SpringerPlus* 5, 697, 18 pp. (2016).

[56] Morris, R. The dilogarithm function of a real argument *Mathematics of Computation* Vol. 33, 146:778-787 (1979).

[57] Muliere, P., Parmigiani, G., & Polson, N.G. A note on the residual entropy function. *Probability in the Engineering and Informational Sciences* 7:413–420 (1993).

[58] Muraleedharan Nair, K.R. & Unnikrishnan Nair, N. On characterizing the bivariate exponential and geometric distributions. *Ann. Inst. Statist. Math.* 40, 2:267–271 (1988).

[59] Nanda, A.K., & Chowdhury, S. Shannon's entropy and its generalizations towards statistics. Reliability and information science during 1948-2018. 18 pp. (2019). *arXiv:1901.09779v1*

[60] Oja, H. Descriptive Statistics for Multivariate Distributions. *Statistics & Probability Letters* 1:327–332 (1983).

[61] Pal, A. Diffusion in a potential landscape with stochastic resetting. *Physical Review E* 91:012113 (2015).

[62] Parsa, M., Di Crescenzo, A., & Jabbari, H. Analysis of reliability systems via Gini-type index. *European Journal of Operational Research* 264:340–353 (2018).

[63] Pellerey, F. & Shaked, M. Characterizations of the IFR and DFR Aging Notions by Means of the Dispersive Order. *Statistics and Probability Letters*, 33:389–393 (1997).

[64] Sachlas, A., & Papaioannou, T. Residual and past entropy in actuarial science and survival models. *Methodology and Computing in Applied Probability* 16:79–99 (2014).

[65] Schroeder, M.J. An alternative to entropy in the measurement of information. *Entropy* 6:388–412 (2004).

[66] Schweizer, N., & Szech, N. A quantitative version of Myerson regularity. *Working Paper Series in Economics 76, Karlsruhe Institute of Technology (KIT)*, Department of Economics and Business Engineering, 18 pp. (2015).

[67] Shaked, M., & Shantikumar, J.G. Stochastic Orders. *Springer* (2007).

[68] Shannon, C.E. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423,623–656 (1948).

[69] Slater, L.J. Generalized hypergeometric functions. *Cambridge University Press* (1966).

[70] Smith, R.L., & Naylor, J.C. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *J. Roy. Statist. Soc. Ser. C* 36:358–369 (1987).

[71] Verdú, S., & Kontoyiannis, I. Lossless data compression rate: asymptotics and non-asymptotics. *In 46th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, 6 pp. doi: 10.1109/CISS.2012.6310950 (2012).

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”

UNIONE EUROPEA
Fondo Sociale Europeo

Ministero dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE 2014 - 2020