Stochastic comparisons of cumulative entropies[*]

Antonio Di Crescenzo[†]
Dipartimento di Matematica
Università di Salerno
Via Ponte don Melillo
I-84084 Fisciano (SA), Italy
Email: adicrescenzo@unisa.it

Maria Longobardi
Dipartimento di Matematica e Applicazioni
Università di Napoli Federico II
Via Cintia
I-80126 Napoli, Italy
Email: maria.longobardi@unina.it

March 31, 2012

Abstract

The cumulative entropy is an information measure which is alternative to the differential entropy and is connected with a notion of reliability theory. Indeed, the cumulative entropy of a random lifetime X can be expressed as the expectation of its mean inactivity time evaluated at X. After a brief rieview of its main properties, in this paper we relate the cumulative entropy to the cumulative inaccuracy, and provide some inequalities based on suitable stochastic orderings. We also show a characterization property of the dynamic version of the cumulative entropy. In conclusion, a stochastic comparison between the empirical cumulative entropy and the empirical cumulative inaccuracy is investigated.

Keywords: Stochastic orders, Cumulative entropy, Mean inactivity time, Cumulative inaccuracy, Dynamic cumulative entropy, Empirical cumulative entropy, Sample spacings.

AMS Subject Classification: 94A17; 60E15.

1 Introduction

In the last 40 years stochastic orders have attracted an increasing number of authors, who used them in several areas of probability and statistics, with applications in many fields, such as

[*]This paper is dedicated to Moshe Shaked in admiration to his most profound contributions on stochastic orders.

[†]corresponding author

reliability theory, queueing theory, survival analysis, operations research, mathematical finance, risk theory, management science, biomathematics. Indeed, stochastic orders are often invoked not only to provide useful bounds and inequalities, but also to compare stochastic systems. A landmark in this area is the book by Shaked and Shanthikumar (2007), which represents an essential reference for a large number of researchers dealing with stochastic orderings. To give an idea of its broad impact we notice that up to now it has received more than 2000 citations in the literature.

The aim of this paper is twofold: to give a brief review on the properties of an information measure recently introduced by the authors, and to provide some new results, including simple examples of applications of stochastic orders to related notions of information theory.

It is well known that the basic way to establish if one random variable is "larger" than another is based on the comparison of their distributions functions. Formally, given two random variables X and Y, we say that X is smaller than Y in the usual stochastic order, denoted by $X \leq_{\rm st} Y$, if and only if

$$\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)] \tag{1}$$

for all increasing functions $\phi : \mathbb{R} \to \mathbb{R}$ for which the expectations exist (see Section 1.A.1 of Shaked and Shanthikumar, 2007). Equivalently, $X \leq_{\rm st} Y$ if and only if $\mathbb{P}(X \leq t) \geq \mathbb{P}(Y \leq t)$ for all $t \in \mathbb{R}$. Another stochastic order that will be used in this paper is the decreasing convex order, denoted by $X \leq_{\rm dcx} Y$, which holds if, and only if, Eq. (1) is true for all decreasing convex functions $\phi : \mathbb{R} \to \mathbb{R}$ for which the expectations exist. We remark that the notion of dcx-order is counterintuitive, in the sense that if $X \leq_{\rm dcx} Y$, then X is "larger" than Y in some stochastic sense (see Section 4.A.1 of Shaked and Shanthikumar, 2007).

Let us now recall some preliminary notions of information theory. The concept of entropy was introduced by Claude Shannon (1948) as a measure of the uncertainty associated with a discrete random variable. Formally, for a random variable X with possible values $\{x_1, \ldots, x_n\}$ and probability mass function $p(\cdot)$, the entropy is given by

$$H(X) = -\mathbb{E}[\log_b p(X)] = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i), \tag{2}$$

where b, the base of the logarithm, is usually equal to 2, e, or 10. Entropy is the minimum descriptive complexity of a random variable X, in the sense that it quantifies the expected value of the information contained in a realization of X. For a thorough description of its role in coding theory, compression schemes and other fields of information theory see Cover and Thomas (1991), for instance. A comprehensive description of information-theoretic methodologies, based on focal measures such as Shannon entropy and Kullback-Leibler information, is given in Ebrahimi *et al.* (2010).

A suitable extension of the Shannon entropy to the absolutely continuous case is the so-called *differential entropy*, which is the shift-independent functional given by

$$H(X) = -\mathbb{E}[\log f_X(X)] = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x) \, \mathrm{d}x, \tag{3}$$

where $\log = \log_e$, and where $f_X(x)$ is the probability density function of an absolutely continuous random variable X having support in $\mathbb{R}$. However, although the analogy between definitions (2) and (3), the differential entropy is an inaccurate extension of the Shannon discrete entropy. Indeed, the latter is not invariant under changes of variables and can even become negative.

Various alternatives for the entropy of a continuous distribution have been proposed in the literature. In Section 4 of Jaynes (1963) the following notion is suggested:

$$H_m(X) = -\int_{-\infty}^{\infty} f_X(x) \log \frac{f_X(x)}{m(x)} \, \mathrm{d}x, \tag{4}$$

where $m(x)$ is a suitable invariant measure. More recently, the "measure problem" involving Eq. (4) has been faced in Maynar and Trizac (2011). Another example of information notion is due to Schroeder (2004), who proposed a measure that, unlike entropy, can be easily and consistently extended to the continuous probability distributions on interval $[a, b]$, and unlike differential entropy is always positive and invariant with respect to linear transformations of coordinates. A "length biased" shift-dependent measure of uncertainty that stems from the differential entropy is the weighted entropy (see Di Crescenzo and Longobardi, 2006),

$$H^w(X) = -\mathbb{E}[X \log f_X(X)] = -\int_0^{+\infty} x \, f_X(x) \log f_X(x) \, \mathrm{d}x, \tag{5}$$

which assigns larger weights to larger values of a non-negative random variable X.

Moreover, the "cumulative residual entropy" is defined as (see Rao *et al.*, 2004)

$$\mathcal{E}(X) = -\int_{-\infty}^{+\infty} \overline{F}_X(x) \log \overline{F}_X(x) \, \mathrm{d}x, \tag{6}$$

where $\overline{F}_X(x) = \mathbb{P}(X > x)$ is the cumulative residual distribution, or survival function, of a random variable X. Various applications of (6) are given in Asadi and Zohrevand (2007), Wang and Vemuri (2007) and Wang *et al.* (2003a), (2003b).

In Section 2 we recall an information measure, named "cumulative entropy", defined by substituiting the survival function $\overline{F}_X(x)$ with the distribution function of X in Eq. (6). Evaluations of the cumulative entropy for some distributions over finite and infinite domains are explicitly given. We also present various properties of such measure. In particular, we relate the cumulative entropy to the cumulative inaccuracy, and recall that it can be expressed as the expectation of the mean inactivity time evaluated at X. Section 3 is devoted to provide some bounds and inequalities involving the cumulative entropy, for which use of stochastic orders is made. In Section 4 the dynamic version of the cumulative entropy is recalled, and a characterization property is provided. Finally, in Section 5 we illustrate some features of a simple estimator of the cumulative entropy based on the sample spacings. The empirical cumulative inaccuracy is also introduced, and a stochastic comparison between such empirical measures is provided.

Note that throughout this paper, the terms "increasing" and "decreasing" are used in non-strict sense.

2 Cumulative entropy

An information measure similar to (6) is the cumulative entropy, defined as (Di Crescenzo and Longobardi, 2009a)

$$\mathcal{CE}(X) = -\int_{-\infty}^{+\infty} F_X(x) \log F_X(x) \, \mathrm{d}x, \tag{7}$$

where $F_X(x) = \mathbb{P}(X \leq x)$ is the cumulative distribution function of a random variable X. The measure $\mathcal{CE}(X)$ is defined similarly to the differential entropy (3). However, since the argument of the logarithm is a probability, we have

$$0 \leq \mathcal{CE}(X) \leq +\infty,$$

whereas $H(X)$ may be negative. Moreover, $\mathcal{CE}(X) = 0$ if and only if X is a constant. From (6) and (7) it follows that the cumulative entropy and the cumulative residual entropy are related by the following relation (Di Crescenzo and Longobardi, 2012):

$$\mathcal{E}(X) + \mathcal{CE}(X) = \int_{-\infty}^{+\infty} h(x) \, \mathrm{d}x,$$

where

$$h(x) = -[F_X(x) \log F_X(x) + \overline{F}_X(x) \log \overline{F}_X(x)], \qquad x \in \mathbb{R}$$

is the partition entropy of X evaluated at x (see Bowden, 2010).

The cumulative entropy is evaluated in Table 1 for various examples of even probability density functions of standard random variables.

We point out that if $Y = aX + b$, with $a \in \mathbb{R}$, $a \neq 0$ and $b \in \mathbb{R}$, then

$$\mathcal{CE}(Y) = |a| \cdot \begin{cases} \mathcal{CE}(X) & \text{if } a > 0, \\ \mathcal{E}(X) & \text{if } a < 0. \end{cases}$$

$f_X(x)$	support	$\mathcal{CE}(X)$
$\frac{9\sqrt{3}}{10\sqrt{10}}x^2,$	$-\sqrt{5/3} < x < \sqrt{5/3}$	0.789790
$\frac{1}{2\sqrt{3}},$	$-\sqrt{3} < x < \sqrt{3}$	0.866025
$\frac{3}{4\sqrt{5}}\left(1-\frac{x^2}{5}\right),$	$-\sqrt{5} < x < \sqrt{5}$	0.885688
$\frac{15}{784\sqrt{7}}\left(7-x^2\right)^2,$	$-\sqrt{7} < x < \sqrt{7}$	0.892215
$\frac{1}{6}\left(\sqrt{6}-\lvert x\rvert\right),$	$-\sqrt{6} < x < \sqrt{6}$	0.892953
$\frac{3}{20\sqrt{10}}\left(\lvert x\rvert-\sqrt{10}\right)^2,$	$-\sqrt{10} < x < \sqrt{10}$	0.900979
$\frac{1}{\sqrt{2}}\,\mathrm{e}^{-\sqrt{2}\,\lvert x\rvert},$	$-\infty < x < \infty$	0.901835
$\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2},$	$-\infty < x < \infty$	0.903197

Table 1: Cumulative entropy for some standard random variables with even densities.

$F_X(x)$	support: $0 < x < +\infty$	$\mathcal{CE}(X)$
$\frac{\Gamma\left(3, \frac{2}{x}\right)}{\Gamma(3)}$	(inverse-gamma distribution)	0.474543
$\frac{x^3(20+x(15+x(6+x)))}{(1+x)^6}$	(beta prime distribution)	0.556511
$\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\log x+0.5\log 2}{\sqrt{2\log 2}}\right)$	(lognormal distribution)	0.565746
$1-e^{-x}$	(exponential distribution)	0.644934

Table 2: Cumulative entropy for some non-negative variables with mean 1 and variance 1.

Other features of $\mathcal{CE}(X)$, such as properties of its two-dimensional version, and a normalized cumulative entropy defined as $\mathcal{NCE}(X) = \mathcal{CE}(X)/\mathbb{E}(X)$ for $0 < \mathbb{E}(X) < +\infty$, were discussed in Di Crescenzo and Longobardi (2009a).

Table 2 shows the cumulative entropy of some non-negative random variables having unity mean and variance.

We notice that an extension of the cumulative entropy has been proposed by Abbasnejad (2011), namely the failure entropy of order α defined as

$$\mathcal{FE}_\alpha(X) = -\frac{1}{\alpha-1}\log\int_0^{+\infty} F_X^\alpha(x)\,\mathrm{d}x,$$

for $\alpha > 0$, $\alpha \neq 1$.

Furthermore, we recall that a weighted version of the cumulative entropy has been defined recently as (Misagh *et al.*, 2011)

$$\mathcal{CE}^w(X) = -\int_0^{+\infty} x\, F_X(x) \log F_X(x)\, \mathrm{d}x,$$

in analogy with the weighted entropy (5).

2.1 Connections to reliability theory

Let us now recall various connections between the cumulative entropy and concepts of reliability theory.

Let X be a non-negative random variable that represents the random lifetime of a reliability system. Denote by $[X \,|\, B]$ a random variable whose distribution is identical to that of X conditional on an event B. The residual lifetime $[X - t \,|\, X > t]$, $t > 0$, describes the time lenght between the failure time X and the inspection time t, given that at time t the system is still active. One of the most used functions to describe the aging process of a system is the mean residual life of X, given by

$$\mathrm{mrl}(t) = \mathbb{E}[X - t \,|\, X > t] = \frac{1}{\overline{F}_X(t)} \int_t^{+\infty} \overline{F}_X(x)\, \mathrm{d}x, \qquad \forall t \geq 0 : \overline{F}_X(t) > 0, \tag{8}$$

which uniquely determines the distribution function of X. Its properties in the description of systems composed by finite mixtures are pinpointed in Navarro and Hernandez (2008). Properties of the mean residual life function in a renewal process and relationships with other relevant functions of reliability theory are examined in Nair and Sankaran (2010).

Information measures have been proposed in the past as a tool to explore the information content in random lifetimes. We recall Ebrahimi and Pellerey (1995), where a new partial ordering among life distributions in terms of their uncertainties is introduced and is used to assess the notion of a "better system". See also Ebrahimi (1996), and Ebrahimi and Kirmani (1996), where a direct approach to measure uncertainty in the residual lifetime distribution has been addressed. Further developments involving new properties of the proposed measure in connection to order statistics and record values are then derived in Asadi and Ebrahimi (2000).

Theorem 2.1 of Asadi and Zohrevand (2007) shows that the cumulative residual entropy (6) can be expressed in terms of (8) as

$$\mathcal{E}(X) = \mathbb{E}[\mathrm{mrl}(X)]. \tag{9}$$

A similar result holds for the cumulative entropy. We recall that, given that at time t a system has been found inactive, $[t - X \,|\, X \leq t]$, $t > 0$, describes the inactivity time of the system, i.e.

the time elapsing between the inspection time t and the failure time X. The inactivity time is thus dual to the residual lifetime $[X - t \,|\, X > t]$. The mean inactivity time of X, given by

$$\tilde{\mu}_X(t) = \mathbb{E}[t - X \,|\, X \le t] = \frac{1}{F_X(t)} \int_0^t F_X(x)\, \mathrm{d}x, \qquad \forall t \ge 0 : \; F_X(t) > 0, \tag{10}$$

has been studied in reliability theory by Ahmad and Kayid (2005), Ahmad *et al.* (2005), Misra *et al.* (2008), for instance. Similarly to (9), Theorem 3.1 of Di Crescenzo and Longobardi (2009a) shows that the cumulative entropy can be expressed as the expectation of the mean inactivity time evaluated at X, i.e.

$$\mathcal{CE}(X) = \mathbb{E}[\tilde{\mu}_X(X)]. \tag{11}$$

We recall that the reversed hazard rate of a random lifetime X is given by (see Block *et al.*, 1998)

$$\tau_X(t) = \frac{\mathrm{d}}{\mathrm{d}t} \log F_X(t) = \frac{f_X(t)}{F_X(t)}, \qquad t > 0 : \; F_X(t) > 0. \tag{12}$$

The following decreasing convex function is defined as a double integral of the reversed hazard rate:

$$T_X^{(2)}(x) = -\int_x^{+\infty} \log F_X(z)\, \mathrm{d}z = \int_x^{+\infty} \left[\int_z^{+\infty} \tau_X(u)\, \mathrm{d}u\right] \mathrm{d}z, \qquad x \ge 0. \tag{13}$$

Its derivative is strictly related to the distribution function of X. Indeed, from Eq. (13) we have

$$\dot{T}_X^{(2)}(x) := \frac{\mathrm{d}}{\mathrm{d}x} T_X^{(2)}(x) = \log F_X(x) = -\int_z^{+\infty} \tau_X(u)\, \mathrm{d}u. \tag{14}$$

We recall that Proposition 3.1 of Di Crescenzo and Longobardi (2009a) provides the following alternative expression of the cumulative entropy of X:

$$\mathcal{CE}(X) = \mathbb{E}\left[T_X^{(2)}(X)\right], \tag{15}$$

with $T_X^{(2)}$ defined in (13).

Given two random lifetimes X and Y having distribution functions F_X and F_Y defined on $(0, \infty)$, let us now introduce the "cumulative inaccuracy"

$$K[F_X, F_Y] = -\int_0^{+\infty} F_X(u) \log F_Y(u)\, \mathrm{d}u, \tag{16}$$

as the cumulative analog of the measure of inaccuracy due to Kerridge (1961). Denoting the reversed hazard rate of Y as τ_Y, we set

$$T_Y^{(2)}(x) = -\int_x^{+\infty} \log F_Y(z)\, \mathrm{d}z = \int_x^{+\infty} \left[\int_z^{+\infty} \tau_Y(u)\, \mathrm{d}u\right] \mathrm{d}z, \qquad x \ge 0. \tag{17}$$

Hereafter we give a probabilistic meaning of the cumulative inaccuracy in terms of (13) and (17).

Proposition 2.1 *For non-negative absolutely continuous random variables X and Y, having distribution functions F_X and F_Y, we have*

$$K[F_X, F_Y] = \mathbb{E}\left[T_Y^{(2)}(X)\right], \qquad K[F_Y, F_X] = \mathbb{E}\left[T_X^{(2)}(Y)\right]. \tag{18}$$

The proof of Proposition 2.1 is omitted, being similar to that of Proposition 3.1 of Di Crescenzo and Longobardi (2009a).

We now aim to provide a connection between the information measures $\mathcal{CE}(X)$ and $K[\cdot,\cdot]$. Let X and Y be the random lifetimes of two systems which have finite unequal means and satisfy $X \geq_{\rm st} Y$ or $Y \geq_{\rm st} X$. Proposition 3.2 of Di Crescenzo and Longobardi (2009a) shows that if X is absolutely continuous and $\mathbb{E}[\tilde{\mu}_X(Y)]$ is finite, then

$$\mathcal{CE}(X) = \mathbb{E}[\tilde{\mu}_X(Y)] + \mathbb{E}[\tilde{\mu}'_X(Z)]\,[\mathbb{E}(X) - \mathbb{E}(Y)], \tag{19}$$

where $\tilde{\mu}'_X(t) = 1 - \tau_X(t)\tilde{\mu}_X(t)$, for all $t > 0$ such that $F_X(t) > 0$, and where Z has probability density function

$$f_Z(x) = \frac{F_Y(x) - F_X(x)}{\mathbb{E}(X) - \mathbb{E}(Y)}, \qquad x \geq 0. \tag{20}$$

Hereafter we state an identity similar to (19).

Proposition 2.2 *Let X and Y be non-negative random variables with finite unequal means and satisfying $X \geq_{\rm st} Y$ or $Y \geq_{\rm st} X$, with X absolutely continuous. If $K[F_Y, F_X]$ is finite, then*

$$\mathcal{CE}(X) = K[F_Y, F_X] + \mathbb{E}\left[\dot{T}_X^{(2)}(Z)\right][\mathbb{E}(X) - \mathbb{E}(Y)], \tag{21}$$

where $\dot{T}_X^{(2)}(\cdot)$ is given in (14), and where the Z is an absolutely continuous non-negative random variable having probability density function (20).

Proof. It follows from identity (15), from the second of (18) and from the probabilistic analogue of the mean value theorem given in Di Crescenzo (1999). ■

3 Inequalities and stochastic comparisons

In this section we shall focus on upper and lower bounds for the cumulative entropy and on some stochastic comparisons.

In Di Crescenzo and Longobardi (2009a) it has been proved that if X is a non-negative random variable, then

(i) $\mathcal{CE}(X) \geq C\, \mathrm{e}^{H(X)}$, where $C = \exp\left\{\int_0^1 \log(x\,|\log x|)\,\mathrm{d}x\right\} = 0.2065$;

(ii) $\mathcal{CE}(X) \geq \int_0^{+\infty} F(x)\,\overline{F}(x)\,\mathrm{d}x$;

(iii) $\mathcal{CE}(X) \geq -\int_{\mu}^{+\infty} \log F(z)\,\mathrm{d}z$;

(iv) $\mathcal{CE}(X) \leq \mathbb{E}(X)$;

(v) $\mathcal{CE}(X) \leq \mathrm{e}^{-1}\, b$;

(vi) $\mathcal{CE}(X) \leq [b - \mathbb{E}(X)] \left|\log\left(1 - \frac{\mathbb{E}(X)}{b}\right)\right|$,

where bounds (v) and (vi) hold if X takes values in $[0, b]$, with b finite. The latter inequality can be generalized by means of the log-sum inequality (see, for instance, Rao, 2005). Indeed, Proposition 1 of Di Crescenzo and Longobardi (2010) states that if X and Y take values in $[0, b]$, with b finite, and if $X \geq_{\rm st} Y$, then

$$\mathcal{CE}(X) \leq \mathcal{CE}(Y) + [b - \mathbb{E}(X)] \left|\log \frac{b - \mathbb{E}(X)}{b - \mathbb{E}(Y)}\right|. \tag{22}$$

We remark that the inequality given in Eq. (22) is tighter than that given in Proposition 4.5 of Di Crescenzo and Longobardi (2009a), which holds under the same assumption, that is $X \geq_{\rm st} Y$. When the stochastic ordering between X and Y is reversed, the following result holds.

Proposition 3.1 *If X and Y are non-negative random variables such that $X \leq_{\rm st} Y$, then*

$$K[F_Y, F_X] \leq \mathcal{CE}(X) \leq K[F_X, F_Y].$$

Proof. Since, by assumption, $F_X(t) \geq F_Y(t)$ for all $t \in \mathbb{R}$, the proof follows from Eqs. (7) and (16). ■

We remark that $X \leq_{\rm st} Y$ does not imply $\mathcal{CE}(X) \leq \mathcal{CE}(Y)$.

Proposition 3.2 *If X and Y are non-negative random variables such that $X \leq_{\rm dcx} Y$, then*

$$\mathcal{CE}(X) \leq K[F_Y, F_X].$$

Proof. Recalling the definition of decreasing convex order, Eq. (15) and the second of (18), the proof follows noting that (13) is a decreasing convex function. ■

We notice that Proposition 3.2 substitutes Proposition 4.6 of Di Crescenzo and Longobardi (2009a).

Example 3.1 Let X and Y have distribution functions $F_X(x) = \exp\{-c\,x^{-\gamma}\}$, $x > 0$, and $F_Y(x) = \exp\{-d\,x^{-\gamma}\}$, $x > 0$, with $c > 0$, $d > 0$ and $\gamma > 1$. From (13) and (17) we have

$$T_X^{(2)}(x) = \frac{c}{\gamma - 1}\, x^{-\gamma+1}, \qquad T_Y^{(2)}(x) = \frac{d}{\gamma - 1}\, x^{-\gamma+1}, \qquad x > 0.$$

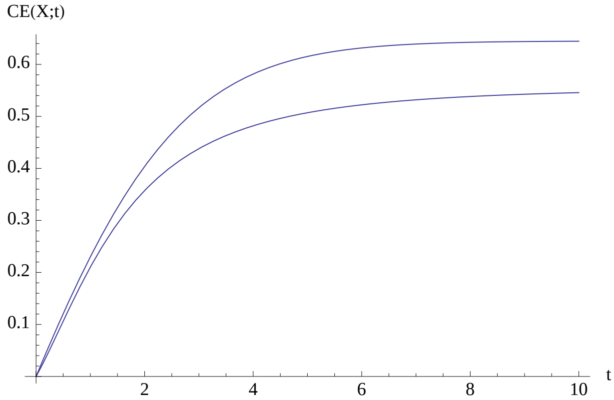

Figure 1: Dynamic cumulative entropy for the beta prime distribution (lower curve) and the exponential distribution given in Table 2.

Hence, making use of (15) and the second of (18) we obtain

$$\mathcal{CE}(X) = \frac{c^{1/\gamma}}{\gamma}\,\Gamma\left(1-\frac{1}{\gamma}\right), \qquad K[F_Y, F_X] = \frac{c d^{-1+1/\gamma}}{\gamma}\,\Gamma\left(1-\frac{1}{\gamma}\right).$$

It immediately follows that if $c \geq d$, i.e. $X \leq_{\rm dcx} Y$, then $\mathcal{CE}(X) \leq K[F_Y, F_X]$, in agreement with Proposition 3.2.

We conclude this section by recalling two further inequalities stated in Di Crescenzo and Longobardi (2009a):

– If X and Y are non-negative and independent random variables, then

$$\max\{\mathcal{CE}(X), \mathcal{CE}(Y)\} \leq \mathcal{CE}(X+Y).$$

– If $X_1, X_2, \ldots, X_n$ are non-negative i.i.d. random variables, then

$$\mathcal{CE}(n\,X_1) \geq \mathcal{CE}(\max\{X_1, X_2, \ldots, X_n\}).$$

4 Dynamic cumulative entropy

Dynamic information measures are often employed in system reliability to describe the effect of the age t on the uncertainty in random lifetimes. For instance, we recall the residual entropy (Ebrahimi, 1996) and the past entropy (Di Crescenzo and Longobardi, 2002), defined as the differential entropy of $[X \,|\, X > t]$ and of $[X \,|\, X \leq t]$, respectively.

The dynamic cumulative residual entropy was proposed by Asadi and Zohrevand (2007) as the cumulative residual entropy of $[X \mid X > t]$, given by

$$\mathcal{E}(X;t) = -\int_t^{+\infty} \frac{\overline{F}_X(x)}{\overline{F}_X(t)} \log \frac{\overline{F}_X(x)}{\overline{F}_X(t)} \, \mathrm{d}x, \qquad t \geq 0. \tag{23}$$

Similarly to (23), the "dynamic cumulative entropy" was defined in Di Crescenzo and Longobardi (2009a) as the cumulative entropy of $[X \mid X \leq t]$, namely

$$\mathcal{CE}(X;t) = -\int_0^t \frac{F_X(x)}{F_X(t)} \log \frac{F_X(x)}{F_X(t)} \, \mathrm{d}x, \qquad t > 0: \; F_X(t) > 0.$$

An alternative expression of $\mathcal{CE}(X;t)$ is given by

$$\mathcal{CE}(X;t) = -\frac{1}{F_X(t)} \int_0^t F_X(x) \log F_X(x) \, \mathrm{d}x + \tilde{\mu}_X(t) \log F_X(t), \qquad t > 0: \; F_X(t) > 0, \tag{24}$$

where $\tilde{\mu}_X(t)$ is the mean inactivity time defined in (10). We remark that $\mathcal{CE}(X;t)$ is non-negative for all t, with

$$\lim_{t \to 0^+} \mathcal{CE}(X;t) = 0, \qquad \lim_{t \to b^-} \mathcal{CE}(X;t) = \mathcal{CE}(X),$$

for any random variable X with support $(0, b)$, with $b \leq +\infty$. Figure 1 shows two cases where $\mathcal{CE}(X;t)$ is increasing in t. An instance of absolutely continuous distribution whose dynamic cumulative entropy is not increasing for all t is provided in Example 6.2 of Di Crescenzo and Longobardi (2009a). This paper also provides various properties of $\mathcal{CE}(X;t)$, such as lower and upper bounds, and the following two representations as conditional means:

$$\mathcal{CE}(X;t) = \mathbb{E}[\tilde{\mu}_X(X) \mid X \leq t], \qquad t > 0,$$

and, when X is a absolutely continuous,

$$\mathcal{CE}(X;t) = \mathbb{E}[T_X^{(2)}(X;t) \mid X \leq t], \qquad t > 0,$$

where

$$T_X^{(2)}(x;t) = -\int_x^t \log \frac{F(z)}{F(t)} \, \mathrm{d}z, \qquad t \geq x \geq 0. \tag{25}$$

Hereafter we give a characterization result for $\mathcal{CE}(X;t)$. To this purpose, we recall that (see Theorem 6.1 of Di Crescenzo and Longobardi, 2009a) $\mathcal{CE}(X;t)$ is increasing in t if and only if $\mathcal{CE}(X;t) \leq \tilde{\mu}_X(t)$ for all $t > 0$ such that $F_X(t) > 0$.

Proposition 4.1 *If X is a non-negative absolutely continuous random variable and if $\mathcal{CE}(X;t)$ is increasing for all $t \geq 0$, then $\mathcal{CE}(X;t)$ uniquely determines $F_X(t)$.*

Proof. Differentiating (24) we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{CE}(X;t) = \tau_X(t)\,[\tilde{\mu}_X(t) - \mathcal{CE}(X;t)]\,, \tag{26}$$

where $\tau_X(t)$ is given in (12). Hence, for any fixed t, the reversed hazard rate $\tau_X(t)$ is a positive solution of equation $g(x) = 0$, where

$$g(x) := x\,[\tilde{\mu}_X(t) - \mathcal{CE}(X;t)] - \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{CE}(X;t).$$

The assumption that $\mathcal{CE}(X;t)$ is increasing in t yields $\mathcal{CE}(X;t) \leq \tilde{\mu}_X(t)$ for all t, so that $\lim_{x\to+\infty} g(x) = +\infty$ and $g(0) \leq 0$, due to (26). Therefore, $g(x) = 0$ has a unique positive solution. Consequently $\tau_X(t)$, and hence $F_X(x)$, is uniquely determined by $\mathcal{CE}(X;t)$ under the assumption that such function is increasing in t. ■

We remark that Corollary 6.1 of Di Crescenzo and Longobardi (2009a) shows that $\mathcal{CE}(X;t)$ is increasing for all $t \geq 0$ if $\tilde{\mu}(t)$ is increasing for all $t \geq 0$. Such paper presents other results on the cumulative entropy, such as characterizations involving identities $\mathcal{CE}(X;t) = c\,\tilde{\mu}_X(t)$ and $\mathcal{CE}(X;t) = c\,\mu_X(t)$, where $\mu_X(t) = \mathbb{E}[X \mid X \leq t]$ denotes the mean past lifetime of X. See also Section 4 of Navarro *et al.* (2010) for related results, such as an extension of a characterization of the power distribution that involves the cumulative entropy.

5 Empirical cumulative entropy

Let $X_1, X_2, \ldots, X_n$ be a random sample of non-negative, absolutely continuous i.i.d. random variables. A suitable estimator of $\mathcal{CE}(X)$ is the "empirical cumulative entropy", proposed in Section 7 of Di Crescenzo and Longobardi (2009a) as

$$\mathcal{CE}(\hat{F}_n) = -\int_0^{+\infty} \hat{F}_n(x)\log\hat{F}_n(x)\,\mathrm{d}x, \tag{27}$$

where

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i \leq x\}}, \qquad x \in \mathbb{R}$$

is the empirical distribution of the sample. Denoting by $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ the sample order statistics, and by

$$U_1 = X_{(1)}, \qquad U_i = X_{(i)} - X_{(i-1)}, \qquad i = 2, 3, \ldots, n$$

the corresponding sample spacings, it is not hard to prove that the empirical cumulative entropy can be expressed as

$$\mathcal{CE}(\hat{F}_n) = -\sum_{j=1}^{n-1} U_{j+1}\,\frac{j}{n}\log\frac{j}{n}. \tag{28}$$

Eq. (28) shows that the empirical cumulative entropy is a positive linear combination of the sample spacings $U_2, \ldots, U_n$, where the outer spacings U_2 and U_n possess small weights, whereas the larger weight is given to the spacing U_{j+1} such that j is close to $\mathrm{e}^{-1}\, n \approx 0.3679\, n$. Eq. (28) gives asymmetric weights to the sample spacings, so that the empirical cumulative entropy is asymmetric to the right. It is thus appropriate to measure variability in right-skewed distributions. A case study on neuronal firing data is provided in Di Crescenzo and Longobardi (2012).

A discussion on $\mathcal{CE}(\hat{F}_n)$ in the case of random samples from uniform distribution and exponential distribution is given in Di Crescenzo and Longobardi (2009a). Moreover, the following asymptotic results have been proved:

(i) the standardized empirical cumulative entropy converges in distribution to a standard normal variable as $n \to +\infty$ (Di Crescenzo and Longobardi, 2009a);

(ii) $\mathcal{CE}(\hat{F}_n) \to \mathcal{CE}(X)$ a.s. as $n \to +\infty$ (Di Crescenzo and Longobardi, 2009b).

We note that by use of identity $-u \log u \leq 1-u$, $0 < u < 1$, from (27) the following relation follows:

$$\mathcal{CE}(\hat{F}_n) \leq \overline{X} \quad \text{a.s.},$$

where $\overline{X}$ is the sample mean.

Let us now consider another random sample $Y_1, Y_2, \ldots, Y_n$ of non-negative, absolutely continuous i.i.d. random variables, and denote its empirical cumulative entropy by

$$\mathcal{CE}(\hat{G}_n) = -\int_0^{+\infty} \hat{G}_n(y) \log \hat{G}_n(y)\, \mathrm{d}y,$$

where $\hat{G}_n(y)$ is the empirical distribution of the sample. Moreover, in analogy with (16), we define the empirical cumulative inaccuracy as

$$K[\hat{F}_n, \hat{G}_n] = -\int_0^{+\infty} \hat{F}_n(u) \log \hat{G}_n(u)\, \mathrm{d}u.$$

It can be expressed as:

$$K[\hat{F}_n, \hat{G}_n] = -\sum_{j=1}^{n-1} \int_{Y_{(j)}}^{Y_{(j+1)}} \hat{F}_n(u) \log \frac{j}{n}\, \mathrm{d}u, \tag{29}$$

where $Y_{(1)} < Y_{(2)} < \ldots < Y_{(n)}$ are the order statistics of the new sample. Let us denote by

$$N_j = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq Y_{(j)}\}}, \qquad j = 1, 2, \ldots, n,$$

the number of random variables of the first sample that are less than or equal to the j-th order statistic of the second sample. Moreover, we rename by $X_{j,1} < X_{j,2} < \ldots$ the random variables of the first sample belonging to $(Y_{(j)}, Y_{(j+1)}]$, if any. From the above positions we thus have

$$\int_{Y_{(j)}}^{Y_{(j+1)}} \hat{F}_n(u)\,\mathrm{d}u = \frac{N_j}{n}\left[Y_{(j+1)} - Y_{(j)}\right] + \frac{1}{n}\sum_{r=1}^{N_{j+1}-N_j}\left[Y_{(j+1)} - X_{j,r}\right],$$

so that Eq. (29) becomes

$$K[\hat{F}_n, \hat{G}_n] = -\frac{1}{n}\sum_{j=1}^{n-1}\left[N_{j+1}Y_{(j+1)} - N_j Y_{(j)} - \sum_{r=1}^{N_{j+1}-N_j} X_{j,r}\right]\log\frac{j}{n}.$$

Cleary, $K[\hat{G}_n, \hat{F}_n]$ can be obtained by symmetry.

In analogy to Proposition 3.1, hereafter we show that if the random variables of the two samples are stochastically ordered, then the empirical cumulative entropy and the empirical cumulative inaccuracies are suitably ordered.

Proposition 5.1 *If random variables X_i and Y_i satisfy condition $X_i \leq_{\rm st} Y_i$, then*

$$K[\hat{G}_n, \hat{F}_n] \leq_{\rm st} \mathcal{CE}(X) \leq_{\rm st} K[\hat{F}_n, \hat{G}_n].$$

Proof. Since $X_i \leq_{\rm st} Y_i$, from Theorem 1.A.3 of Shaked and Shanthikumar (2007) we have that $\mathbf{1}_{\{X_i \leq x\}} \geq_{\rm st} \mathbf{1}_{\{Y_i \leq x\}}$, and thus $\hat{F}_n(x) \geq_{\rm st} \hat{G}_n(x)$, for all $x \in \mathbb{R}$. The proof then follows from the definitions of the involved notions. ■

Acknowledgments

The corresponding author is grateful to Fabio Spizzichino for introducing Moshe Shaked to him in 1996, which started the collaboration behind the Di Crescenzo and Shaked (1996) paper.

Research performed under partial support by MIUR-PRIN 2008 "Mathematical models and computation methods for information processing and transmission in neuronal systems subject to stochastic dynamics".

References

Abbasnejad M. (2011) Some characterization results based on dynamic survival and failure entropies. *Comm. Korean Stat. Society* **18**, 1-12.

Ahmad, I.A., Kayid, M., 2005. Characterizations of the RHR and MIT orderings and the DRHR and IMIT classes of life distributions. *Probab. Eng. Inform. Sci.* **19**, 447–461.

Ahmad, I.A., Kayid, M., Pellerey, F., 2005. Further results involving the MIT order and IMIT class. *Probab. Eng. Inform. Sci.* **19**, 377–395.

Asadi, M., Ebrahimi, N., 2000. Residual entropy and its characterizations in terms of hazard function and mean residual life function. *Stat. Prob. Lett.* **49**, 263-269.

Asadi, M., Zohrevand, Y., 2007. On the dynamic cumulative residual entropy. *J. Stat. Plann. Infer.* **137**, 1931–1941.

Block, H.W., Savits, T.H., Singh, H., 1998. The reversed hazard rate function. *Prob. Engin. Inform. Sci.* **12**, 69-90.

Bowden, R.J., 2010. Information, measure shifts and distribution metrics. *Statistics* **46**, 249-262.

Cover, T.M., Thomas. J.A., 1991. *Elements of Information Theory.* Wiley, New York.

Di Crescenzo, A., 1999. A probabilistic analogue of the mean value theorem and its applications to reliability theory. *J. Appl. Prob.* **36**, 706–719.

Di Crescenzo, A., Longobardi, M., 2002. Entropy-based measure of uncertainty in past lifetime distributions. *J. Appl. Prob.* **39**, 434–440.

Di Crescenzo, A., Longobardi, M., 2006. On weighted residual and past entropies. *Scientiae Math. Japon.* **64**, 255–266.

Di Crescenzo, A., Longobardi, M., 2009a. On cumulative entropies. *J. Stat. Plann. Infer.* **139**, 4072-4087.

Di Crescenzo, A., Longobardi, M., 2009b. On cumulative entropies and lifetime estimations. In: *Methods and Models in Artificial and Natural Computation*, IWINAC 2009, *Lecture Notes in Computer Science*, Vol. 5601 (Mira J., Ferrandez J.M., Alvarez Sanchez J.R., Paz F., Toledo J. eds.), 132-141. Springer-Verlag, Berlin.

Di Crescenzo, A., Longobardi, M., 2010. More on cumulative entropy. In: *Cybernetics and Systems 2010*, (Trappl R. ed.), 181-186. Austrian Society for Cybernetic Studies, Vienna.

Di Crescenzo, A., Longobardi, M., 2012. Neuronal data analysis based on the empirical cumulative entropy. In: *Computer Aided Systems Theory*, EUROCAST 2011, Part I, *Lecture Notes in Computer Science*, Vol. 6927 (Moreno-Diaz R., Pichler F., Quesada-Arencibia A. eds.), 72-79. Springer-Verlag, Berlin.

Di Crescenzo, A., Shaked, M., 1996. Some applications of the Laplace transform ratio order. *Arab J. Math. Sci.* **2**, 121-128.

Ebrahimi, N., 1996. How to measure uncertainty in the residual life time distribution. *Sankhyä, Ser. A* **58**, 48–56.

Ebrahimi, N., Kirmani, S.N.U.A., 1996. Some results on ordering of survival functions through uncertainty. *Stat. Prob. Lett.* **29**, 167-176.

Ebrahimi, N., Pellerey, F., 1995. New partial ordering of survival functions based on the notion of uncertainty. *J. Appl. Prob.* **32**, 202–211.

Ebrahimi, N., Soofi, E.S., Soyer, R., 2010. Information measures in perspective. *Internat. Stat. Review* **78**, 383-412.

Jaynes, E.T., 1963. Information theory and statistical mechanics" In: *Statistical Physics*, 1962 Brandeis Lectures in Theoretical Physics, Vol. 3. (Ford K. ed.), 181-218. Benjamin, New York.

Kerridge, D.F., 1961. Inaccuracy and Inference. *J. Royal Stat. Soc., Ser. B* **23**, 184–194.

Maynar, P., Trizac, E., 2011. Entropy of continuous mixtures and the measure problem. *Phys. Rev. Lett.* **106**, 160603 [4 pages]

Misagh, F., Panahi, Y., Yari, G.H., Shahi, R. 2011. Weighted cumulative entropy and its estimation. IEEE International Conference on Quality and Reliability (ICQR), 2011. 477-480.

Misra, N., Gupta, N., Dhariyal, I.D., 2008. Stochastic properties of residual life and inactivity time at a random time. *Stoch. Models* **24**, 89-102.

Navarro, J., del Aguila, Y., Asad M., 2010. Some new results on the cumulative residual entropy. *J. Stat. Plann. Infer.* **140**, 310-322.

Navarro, J., Hernandez P.J., 2008. Mean residual life functions of finite mixtures, order statistics and coherent systems. *Metrika* **67**, 277-298.

Rao, M., 2005. More on a new concept of entropy and information. *J. Theor. Probab.* **18**, 967–981.

Rao, M., Chen, Y., Vemuri, B.C., Wang, F., 2004. Cumulative residual entropy: a new measure of information. *IEEE Trans. Inf. Theory* **50**, 1220–1228.

Schroeder, M.J., 2004. An alternative to entropy in the measurement of information. *Entropy* **6**, 388–412.

Shaked, M., Shanthikumar, J.G., 2007. *Stochastic Orders*, Springer, New York.

Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Techn. J.* **27**, 279-423.

Unnikrishnan Nair, N., Sankaran, P.G., 2010. Properties of a mean residual life function arising from renewal theory. *Naval Res. Logist.* **57**, 373-379.

Wang, F., Vemuri, B.C., 2007. Non-rigid multi-modal image registration using cross-cumulative residual entropy. *Intern. J. Computer Vision* **74**, 201–215.

Wang, F., Vemuri, B.C., Rao, M., Chen, Y., 2003a. A new & robust information theoretic measure and its application to image alignment. In: *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, Vol. 2732, p. 388–400. Springer-Verlag, Heidelberg.

Wang, F., Vemuri, B.C., Rao, M., Chen, Y., 2003b. Cumulative residual entropy, a new measure of information & its application to image alignment. In: *Proceedings on the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, Vol. 1, p. 548–553. IEEE Computer Society.