View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Archivio della ricerca - Università degli studi di Napoli Federico II

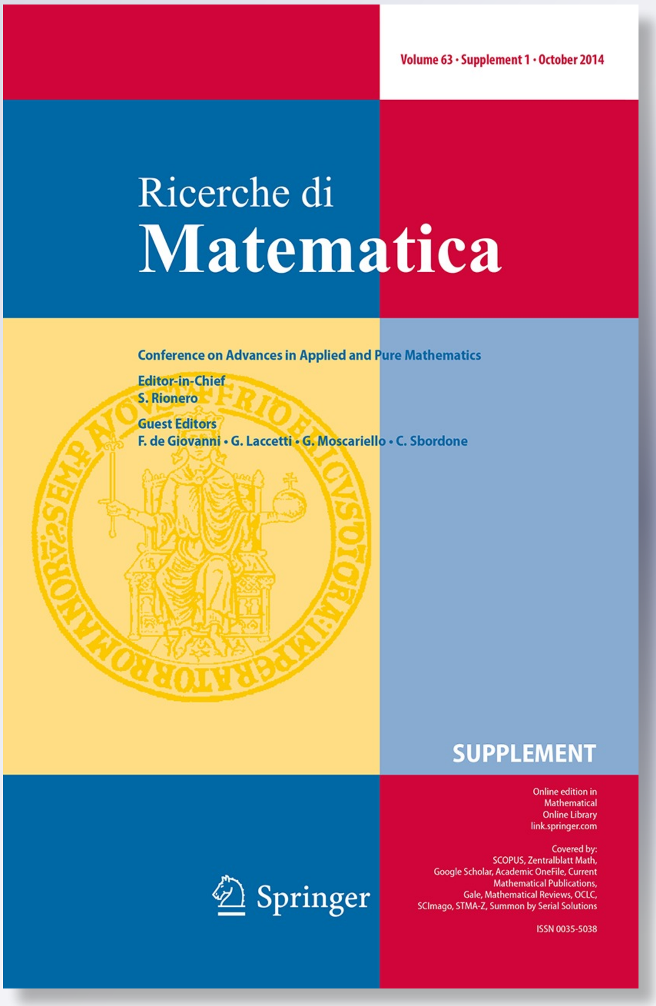
Cumulative measures of information and stochastic orders

Maria Longobardi

Ricerche di Matematica

ISSN 0035-5038
Volume 63
Supplement 1

Ricerche mat. (2014) 63:209-223
DOI 10.1007/s11587-014-0212-x



Springer

Your article is protected by copyright and all rights are held exclusively by Università degli Studi di Napoli "Federico II". This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Springer

Author's personal copy

Ricerche mat. (2014) 63 (Suppl 1):S209–S223
DOI 10.1007/s11587-014-0212-x

Cumulative measures of information and stochastic orders

Maria Longobardi

Received: 4 July 2014 / Revised: 1 August 2014 / Published online: 20 August 2014
© Università degli Studi di Napoli "Federico II" 2014

Abstract In this paper we present some known results on cumulative measures of information, study their properties and relate these definitions to concepts of reliability theory. We give some relations of these measures of discrimination with some well-known stochastic orders and with the relative reversed hazard rate order. We investigate also a stochastic comparison among the empirical cumulative measures that can be related to the cumulative measures. Large part of this paper is a survey article; however, in the last section, we define a new measure of discrimination between residual lifetimes and study some of its properties.

Keywords Cumulative residual entropy · Differential entropy · Empirical cumulative entropy · Kullback–Leibler discrimination measure · Reversed hazard rate

Mathematics Subject Classification Primary 94A17; Secondary 62N05 · 60E15

1 Introduction

Entropy is a baseline concept in information theory and it was introduced by Claude Shannon in 1948 (see [25]) as a measure of the uncertainty associated to a discrete random variable. If X is a r.v. with values $\{x_1, \ldots, x_n\}$ and probability mass function p, the Shannon entropy is:

Communicated by Salvatore Rionero.

Dedicated to the memory of Luigi Maria Ricciardi, master of science and life.

M. Longobardi (✉)
Dipartimento di Matematica e Applicazioni, Università di Napoli Federico II,
Via Cintia, 80126 Naples, Italy
e-mail: maria.longobardi@unina.it

Author's personal copy

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

In the classical approach to Information Theory, when X is a non-negative absolutely continuous random variable with support $S = (0, u), u \leq +\infty$, and $f_X(x)$, $F_X(x)$ and $\overline{F}_X(x)$ are the probability density function, the cumulative distribution function and the survival function of X, respectively, the Shannon information measure of X, also called the differential entropy of X, is defined by

$$H_X = E[-\log f(X)] = -\int_0^{+\infty} f(x) \log f(x) \, dx \tag{1}$$

where 'log' means natural logarithm and $0 \log 0 = 0$ by convention. It is well-known that H_X measures the "uniformity" of the distribution of X, i.e. how the distribution spreads over its domain, and is irrespective of the locations of concentration. Observe that H_X is not invariant under changes of variables and can even be negative.

A "length biased" shift-dependent measure of uncertainty that stems from the differential entropy is the weighted entropy (see Di Crescenzo and Longobardi [6])

$$H^w(X) = -E[X \log f_X(X)] = -\int_0^{+\infty} x \, f_X(x) \log f_X(x) \, dx \tag{2}$$

If X describes the random lifetime of a biological system, such as an organism or a cell, then $X_t = [X - t \,|\, X > t]$ describes the remaining (residual) lifetime of the system at age t, with distribution function

$$F_{X_t}(x) = \frac{F_X(x+t) - F_X(t)}{\overline{F}_X(t)}$$

Hence, if the system has survived up to time t, the uncertainty about the remaining lifetime is measured by means of the differential entropy of X_t. The mean residual lifetime is

$$mrl_X(t) = E(X_t) = E(X - t \,|\, X > t) = \frac{1}{\overline{F}(t)} \int_t^{+\infty} \overline{F}(x) \, dx, \quad t \in \mathcal{S}.$$

The random variable $X_{(t)} = [X \,|\, X \leq t]$ describes the past lifetime of the system at age t with distribution function

$$F_{X_{(t)}}(x) = \frac{F_X(x)}{F_X(t)}$$

and mean past lifetime

$$\mu_X(t) = E[X_{(t)}] = \int_0^t \left[1 - \frac{F_X(x)}{F_X(t)}\right] dx, \quad t \in \mathcal{S}.$$

Author's personal copy

Its uncertainty was defined in 2002 by Di Crescenzo and Longobardi [5] and called past entropy.

In reliability theory, the duration of the time between an inspection time t and the failure time X, given that at time t the system has been found failed, is called inactivity time and is represented by the random variable $[t - X \mid X \le t]$, $t > 0$, with mean inactivity time

$$\tilde{\mu}_X(t) = E[t - X \mid X \le t] = \frac{1}{F(t)} \int_0^t F(x)\,\mathrm{d}x$$

For further results about these concepts in reliability theory see [4,13,14,16,18,19].

The hazard rate function $\lambda_X(t)$ and the reversed hazard rate function $\tau_X(t)$ of X are important biometric functions, and are defined as the ratio of the density of X to the survival function of X and the ratio of the density of X to the distribution function of X, respectively, i.e.

$$\lambda_X(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log \overline{F}_X(t) = \frac{f_X(t)}{\overline{F}_X(t)}$$

and

$$\tau_X(t) = \frac{\mathrm{d}}{\mathrm{d}t} \log F_X(t) = \frac{f_X(t)}{F_X(t)}$$

for every t such that $\overline{F}_X(t) > 0$ and $F_X(t) > 0$. For further properties of hazard rate and reversed hazard rate functions see [2,15,19].

The following decreasing convex function is defined as a double integral of the reversed hazard rate, for $x \ge 0$:

$$T_X^{(2)}(x) = -\int_x^{+\infty} \log F_X(z)\,\mathrm{d}z = \int_x^{+\infty} \left[\int_z^{+\infty} \tau_X(u)\,\mathrm{d}u \right] \mathrm{d}z \tag{3}$$

Its derivative is strictly related to the distribution function of X. Indeed, we have

$$\dot{T}_X^{(2)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} T_X^{(2)}(x) = \log F_X(x) = -\int_z^{+\infty} \tau_X(u)\,\mathrm{d}u \tag{4}$$

In the last 50 years stochastic orders have attracted many authors of different areas of probability and statistics, because orders are invoked to give bounds and inequalities and also comparisons among stochastic systems. The basic idea is to establish if a random variable is larger than another one. Let us recall some stochastic orders (see Shaked and Shanthikumar [24]). Let X and Y be continuous random variables having support (l_X, r_X) and (l_Y, r_Y), respectively, with $-\infty \le l_X < r_X \le \infty$ and $-\infty \le l_Y < r_Y \le \infty$. Then

- $X \le_{\text{st}} Y$ (in the usual stochastic order) if and only if

$$E[\phi(X)] \le E[\phi(Y)] \tag{5}$$

Author's personal copy

for all increasing functions $\phi : \mathbf{R} \to \mathbf{R}$ for which the expectations exist. Equivalently, $X \le_{st} Y$ if and only if $\mathbf{P}(X \le t) \ge \mathbf{P}(Y \le t)$ for all $t \in \mathbf{R}$

- $X \le_{lr} Y$ (in likelihood ratio order) if and only if $f_X(t)/f_Y(t)$ is decreasing in $t \in (l_X, r_Y)$;
- $X \le_{hr} Y$ (in hazard rate order) if and only if $\lambda_X(t) \ge \lambda_Y(t)$, for every $t \in (-\infty, \max\{r_X, r_Y\})$ or, equivalently, if and only if $\overline{F}_X(t)/\overline{F}_Y(t)$ is decreasing in $t \in (-\infty, \max\{r_X, r_Y\})$;
- $X \le_{rh} Y$ (in the reversed hazard rate order) if and only if $\tau_X(t) \le \tau_Y(t)$, $\forall t \in (\min\{l_X, l_Y\}, \infty)$ or, equivalently

$$\frac{F_X(t)}{F_Y(t)} \quad \text{is decreasing in } t \in (\min\{l_X, l_Y\}, \infty). \tag{6}$$

- $X \le_{dcx} Y$ (in the decreasing convex order) if and only if (5) is true for all decreasing convex functions $\phi : \mathbf{R} \to \mathbf{R}$ for which the expectations exist.

Recently, a new stochastic order, called relative reversed hazard rate order, was given in [23]. The random variable X is said to be smaller than Y in relative reversed hazard rate order ($X \le_{rrh} Y$) if and only if $\frac{\tau_X(t)}{\tau_Y(t)}$ is decreasing in $t > 0$.

In this paper we summarize some results recently obtained on various information measures, and provide a new dynamic cumulative measure of discrimination that will be the object of future investigation.

In Sect. 2 we recall the definition of the cumulative entropy, an information measure defined by substituting the probability density function $f_X(x)$ with the distribution function $F_X(x)$ in (1), we relate this measure to the cumulative inaccuracy, give some results in reliability theory and obtain some bounds and inequalities using stochastic orders. Moreover, we give the definition of a dynamic version of the cumulative entropy (see [7–11]).

In Sect. 3 we study the cumulative Kullback–Leibler measure of discrimination, a dynamic version of this measure and study some relations with inaccuracy (see [12]).

In Sect. 4 we write the empirical cumulative entropy and the empirical cumulative KL measure and their relation with the empirical cumulative inaccuracy. We give also a theorem in which this measure is related to the relative reversed hazard rate order (see [7–12]).

Finally, in Sect. 5 we define a new cumulative discrimination measure for residual lifetimes, providing some properties and an example.

2 Cumulative entropies

Various alternatives for the entropy of a continuous distribution have been proposed in the literature. The cumulative residual entropy (CRE) of a random lifetime X is given by

Author's personal copy

$$\mathcal{E}(X) = -\int_0^{+\infty} \overline{F}(x)\,\log \overline{F}(x)\,\mathrm{d}x = E[mrl(X)] \tag{7}$$

(see [22] and, for applications, [1]). For further results see [20].

Di Crescenzo and Longobardi [7] defined a new information measure similar to $\mathcal{E}(X)$, that is useful to measure information on the inactivity time of a system. This measure is called cumulative entropy and is defined as

$$\mathcal{CE}(X) = -\int_0^{+\infty} F(x)\,\log F(x)\,\mathrm{d}x \tag{8}$$

In this definition the argument of the logarithm is a probability.

We remark that:

- $\mathcal{CE}(X) = 0$ iff X is a constant;
- If $Y = aX + b$, with $a \neq 0$ and $a, b \in \mathbf{R}$, then

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

For a non-negative random variable X, with mean inactivity time $\tilde{\mu}(t)$ and cumulative entropy $\mathcal{CE}(X) < +\infty$, we have (see [7])

$$\mathcal{CE}(X) = E[\tilde{\mu}(X)] = E\left[T_X^{(2)}(X)\right] \tag{9}$$

The cumulative entropy and the cumulative residual entropy are related by the following relation:

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

where

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

is the partition entropy of X evaluated at x defined in [3] (for details, see [10]).

Let us summarize some bounds of the cumulative entropy in the following theorem:

Theorem 1 *Let X be a non-negative random variable. Then*

(i) $\mathcal{CE}(X) \geq C\,\mathrm{e}^{H(X)}$, *where* $C = \exp\{\int_0^1 \log(x\,|\log x|)\,\mathrm{d}x\} = 0.2065$;
(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) $\mathcal{CE}(X) \leq E(X)$;
(v) $\mathcal{CE}(X) \leq \mathrm{e}^{-1}\,b$;

Author's personal copy

(vi) $\mathcal{CE}(X) \le [b - E(X)] \left| \log \left(1 - \frac{E(X)}{b}\right) \right|$, *where the equality holds if X takes values* 0 *and* b *with probability* e^{-1} *and* $1 - e^{-1}$, *respectively, in which case* $\mathcal{CE}(X) = e^{-1} b$.

The bounds (v) and (vi) hold if X takes values in $[0, b]$, *with b finite.*

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

$$K[F_X, F_Y] = -\int_0^{+\infty} F_X(u) \log F_Y(u)\, \mathrm{d}u \tag{10}$$

as the cumulative analog of the measure of inaccuracy due to Kerridge ([17]).

We give now a probabilistic meaning of the cumulative inaccuracy in terms of $T_Y^{(2)}(X)$ and $T_X^{(2)}(Y)$ and a connection between $K[F_X, F_Y]$ and $\mathcal{CE}(X)$ (see [11]).

Proposition 1 *For non-negative absolutely continuous random variables X and Y, we have*

$$K[F_X, F_Y] = E\left[T_Y^{(2)}(X)\right], \quad K[F_Y, F_X] = E\left[T_X^{(2)}(Y)\right] \tag{11}$$

Proposition 2 *Let X and Y be non-negative random variables with finite unequal means and satisfying* $X \ge_{\mathrm{st}} Y$ *or* $Y \ge_{\mathrm{st}} X$, *with X absolutely continuous. If* $K[F_Y, F_X]$ *is finite, then*

$$\mathcal{CE}(X) = K[F_Y, F_X] + E\left[\dot{T}_X^{(2)}(Z)\right][E(X) - E(Y)], \tag{12}$$

where Z is an absolutely continuous non-negative random variable having probability density function

$$f_Z(x) = \frac{F_Y(x) - F_X(x)}{E(X) - E(Y)}, \quad x \ge 0 \tag{13}$$

The following theorem proved in [9] is a generalization of (vi) of Theorem 1:

Theorem 2 *Let X and Y be random variables that take values in* $[0, b]$, *where* $b < +\infty$, *with finite means* $E(X)$ *and* $E(Y)$, *respectively, and such that* $X \ge_{st} Y$. *Then*

$$\mathcal{CE}(X) \le \mathcal{CE}(Y) + [b - E(X)] \left| \log \frac{b - E(X)}{b - E(Y)} \right| \tag{14}$$

In the following theorem, we obtain a connection between our measures of discrimination and stochastic orders (see [11]).

Theorem 3 *Let X and Y be non-negative random variables. If* $X \le_{\mathrm{st}} Y$, *then*

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

Author's personal copy

If $X \le_{\rm dcx} Y$, *then*

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

We remark that $X \le_{\rm st} Y$ does not imply in general $\mathcal{CE}(X) \le \mathcal{CE}(Y)$.

If we consider the random variable past lifetime $[X \mid X \le t]$, it is interesting to consider a measure of uncertainty related to the past. The dynamic cumulative entropy is defined, for any random variable X with support $(0, b)$, for $b \le +\infty$, as the cumulative entropy of $[X \mid X \le t]$, namely

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

Let us recall that:

- $\mathcal{CE}(X;t)$ is non-negative for all t
- $\lim_{t\to 0^+} \mathcal{CE}(X;t) = 0, \quad \lim_{t\to b^-} \mathcal{CE}(X;t) = \mathcal{CE}(X)$.

Remark also that

$$\mathcal{CE}(X;t) = E[\tilde{\mu}_X(X) \mid X \le t] = E[T_X^{(2)}(X;t) \mid X \le t], \quad t > 0,$$

where

$$T_X^{(2)}(x;t) = -\int_x^t \log \frac{F(z)}{F(t)}\, \mathrm{d}z, \quad t \ge x \ge 0 \tag{15}$$

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

3 Cumulative Kullback–Leibler information measure

In this section we recall a new measure of information and its properties recently studied by Di Crescenzo and Longobardi [12]. The notion of differential entropy has been extended to the relative entropy of two distributions, which is a discrepancy measure between the distributions of two non-negative absolutely continuous random variables X and Y with distribution functions F_X and F_Y, respectively, called Kullback–Leibler information measure:

$$I_{X,Y} = \int_0^{+\infty} f_X(u) \log \frac{f_X(u)}{f_Y(u)}\, \mathrm{d}u \tag{16}$$

$I_{X,Y}$ measures the inefficiency of assuming that the pdf is f_Y when the true pdf is f_X.

Let X and Y be random variables with finite means and with left-hand-points

$$l_X = inf\{t \in R : F_X(t) > 0\} \quad \text{and} \quad l_Y = inf\{t \in R : F_Y(t) > 0\}$$

Author's personal copy

and right-hand-points

$$r_X = sup\{t \in R : F_X(t) < 1\} \quad \text{and} \quad r_Y = sup\{t \in R : F_Y(t) < 1\},$$

respectively. Suppose that $l_X = l_Y = l$. The cumulative KL information of X and Y is defined as

$$\mathcal{C}_{KL}(X,Y) = \int_l^{max\{r_X,r_Y\}} F_X(u) \log \frac{F_X(u)}{F_Y(u)} \, \mathrm{d}u + E(X) - E(Y) \tag{17}$$

provided that the integral in the right-hand-side is finite (see [21]).

Remark that $\mathcal{C}_{KL}(X,Y) \geq 0$ and $\mathcal{C}_{KL}(X,Y) = 0$ if and only if $F_X(u) = F_Y(u)$ almost everywhere. The measure $\mathcal{C}_{KL}(X,Y)$ is strictly related to the cumulative entropy of X. In fact,

$$\mathcal{C}_{KL}(X,Y) = K[F_X, F_Y] - \mathcal{CE}(X) + E(X) - E(Y) \tag{18}$$

where $K[F_X, F_Y] = -\int_l^{max\{r_X,r_Y\}} F_X(u) \log F_Y(u) \, \mathrm{d}u$ and $\mathcal{CE}(X)$ is defined in (8).

It is interesting to compare the cumulative KL information with other suitable discrimination measures like Cramer–von Mises distance, Renyi divergence of order $\alpha > 0$, energy distance, Hellinger distance and Bhattacharya distance. To this aim, we give the following example, which is a particular case of Example 2.2 in [12].

Example 1 If X and Y have exponential distributions with parameters 1 and 2, respectively, we have:

$$\begin{aligned} &\mathcal{C}_{KL}(X,Y) = -\gamma - \tfrac{1}{2} + \tfrac{1}{6}\pi^2 - \psi(2), && I_{X,Y} = -\tfrac{1}{2} + \log 2, \\ &d_{CM}(X,Y) = \frac{1}{30}, && d_{R,2}(X,Y) = \log \tfrac{4}{3}, \\ &d_{R,0.5}(X,Y) = 2\log \tfrac{3}{2\sqrt{2}}, && d_E(X,Y) = \tfrac{1}{3}, \\ &d_H(X,Y) = 1 - \tfrac{2\sqrt{2}}{3}, && d_B(X,Y) = \log \tfrac{3}{2\sqrt{2}} \end{aligned} \tag{19}$$

where $\gamma \simeq 0.577216$ is the Euler's constant and $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

The next five results on the cumulative KL discrimination measure are taken from [12].

Theorem 4 *Let X and Y be absolutely continuous random variables taking values in $[0, r]$, with r finite. Then,*

$$\mathcal{C}_{KL}(X,Y) \geq [r - E(X)] \log \frac{r - E(X)}{r - E(Y)} + E(X) - E(Y) \tag{20}$$

The right-hand-side of (20) is non-negative, and it vanishes when $E(X) = E(Y)$.

In the next theorem we recall some lower and upper bounds for $\mathcal{C}_{KL}(X,Y)$.

Author's personal copy

Theorem 5 *Let X and Y be random variables with finite means. Then*

$$\mathcal{C}_{KL}(X,Y) \geq \frac{1}{2}\int_l^{max\{r_X,r_Y\}} \frac{[F_X(u)-F_Y(u)]^2}{\frac{1}{3}F_X(u)+\frac{2}{3}F_Y(u)}\,\mathrm{d}u \tag{21}$$

and

$$\mathcal{C}_{KL}(X,Y) \leq \int_l^{\max\{r_X,r_Y\}} \left[\frac{F_X^2(u)}{F_Y(u)} - F_X(u)\right]\mathrm{d}u + E(X) - E(Y),$$

provided that the integrals in the right-hand-sides are finite. Moreover, if $X \geq_{st} Y$, we have

$$\mathcal{C}_{KL}(X,Y) \leq \int_l^{\max\{r_X,r_Y\}} \frac{F_X(u)}{2}\left[\frac{F_X(u)}{F_Y(u)}-1\right]\left[3-\frac{F_X(u)}{F_Y(u)}\right]\mathrm{d}u + E(X) - E(Y)$$

and

$$\mathcal{C}_{KL}(X,Y) \leq -\alpha\int_l^{\max\{r_X,r_Y\}} F_X(u)\left[1-\frac{F_X(u)}{F_Y(u)}\right]^2\mathrm{d}u + E(X) - E(Y),$$

with $\alpha = 2.45678$.

We recall also a dynamic measure for past lifetimes (see [12]). The cumulative past KL information of two random lifetimes X and Y, having supports $(0, r_X)$ and $(0, r_Y)$, respectively, is defined as

$$\mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) = \int_0^t \frac{F_X(x)}{F_X(t)} \log\left(\frac{F_X(x)}{F_X(t)}\frac{F_Y(t)}{F_Y(x)}\right)\mathrm{d}x + \mu_X(t) - \mu_Y(t),$$

for every $t \in \mathcal{D}_{X,Y} = (0, min\{r_X, r_Y\})$.

Note that $\mathcal{C}_{KL}(X_{(t)}, Y_{(t)})$ measures how the distributions of the past lifetimes $X_{(t)}$ and $Y_{(t)}$ are close, when the items at time $t > 0$ are inspected and both are found failed. This measure satisfies:

- $\mathcal{C}_{KL}(X_{(t)}, Y_{(t)})$ is non-negative for all $t \in \mathcal{D}_{X,Y}$
- $\lim_{t\to 0^+} \mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) = 0$ and $\lim_{t\to +\infty} \mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) = \mathcal{C}_{KL}(X, Y)$.

The cumulative past KL information of X and Y can be expressed as

$$\mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) = K[F_{X_{(t)}}, F_{Y_{(t)}}] - \mathcal{CE}(X;t) + \mu_X(t) - \mu_Y(t), \quad \text{for } t \in \mathcal{D}_{X,Y} \tag{22}$$

where $K[F_{X_{(t)}}, F_{Y_{(t)}}]$ introduces the cumulative inaccuracy of the past lifetimes, i.e.

$$K[F_{X_{(t)}}, F_{Y_{(t)}}] = -\int_0^t \frac{F_X(x)}{F_X(t)} \log\frac{F_Y(x)}{F_Y(t)}\,\mathrm{d}x \tag{23}$$

(see [12]).

Author's personal copy

The following result gives some bounds for this measure.

Theorem 6 *For all* $t \in \mathcal{D}_{X,Y}$

$$\mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) \geq [t - \mu_X(t)] \log \frac{t - \mu_X(t)}{t - \mu_Y(t)} + \mu_X(t) - \mu_Y(t) \tag{24}$$

and

$$\mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) \geq \frac{1}{2} \int_0^t \frac{[F_{X_{(t)}}(x) - F_{Y_{(t)}}(x)]^2}{\frac{1}{3} F_{X_{(t)}}(x) + \frac{2}{3} F_{Y_{(t)}}(x)} \,\mathrm{d}x \tag{25}$$

Moreover, if $X \geq_{rh} Y$, *then*

$$\mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) \leq \mu_X(t) - \mu_Y(t) \tag{26}$$

A non-negative random variable X is said to be increasing (decreasing) failure rate, in short IFR (DFR), if $\overline{F}_X(t) = \mathbb{P}(X > t)$ is logconcave (logconvex), i.e. if the hazard rate function $\lambda_X(t)$ is increasing (decreasing) in $t > 0$. Given two non-negative random variables X and Y describing random lifetimes, in reliability theory it is of interest to study the reliability of one variable with respect to the other.

For any non-negative random variable Y, consider the function

$$T_Y(t) = -\log F_Y(t) = \int_t^{+\infty} \tau_Y(x)\,\mathrm{d}x, \quad t > 0.$$

This introduces a notion of relative aging, in the sense that X is ageing faster than Y if and only if the random variable $T_Y(X)$ is IFR, or equivalently if and only if the ratio of hazard rates $\lambda_X(t)/\lambda_Y(t)$ is increasing in t. Here, for $t > 0$, the hazard rate function of $Z = T_X(Y)$ is

$$\lambda_Z(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log \overline{F}_Z(t) = -\tau_Y(T_X^{-1}(t)) \frac{\mathrm{d}}{\mathrm{d}t} T_X^{-1}(t) = \frac{\tau_Y(T_X^{-1}(t))}{\tau_X(T_X^{-1}(t))}.$$

We can relate the concept of ageing in reliability theory to the relative reversed hazard rate order. In fact:

Proposition 3 *The following statements are equivalent:*

(i) $Z = T_X(Y)$ *is IFR (DFR);*
(ii) $T_Y \circ T_X^{-1}$ *is convex (concave);*
(iii) $X \geq_{\mathrm{rrh}} (\leq_{\mathrm{rrh}}) Y$ (*in the relative reversed hazard rate order*).

It is easy to prove that if $X \leq_{rh} Y$ and $X \leq_{\mathrm{rrh}} Y$, then $X \leq_{\mathrm{lr}} Y$.

Recall now a monotonicity property.

Author's personal copy

Theorem 7 *Let X and Y be non-negative absolutely continuous random variables, and let $t \in \mathcal{D}_{X,Y}$. Then $\mathcal{C}_{KL}(X_{(t)}, Y_{(t)})$ is increasing in t if and only if*

$$\mathcal{C}_{KL}(X_{(t)}, Y_{(t)}) \leq \frac{1}{\tau_X(t)} \{[\tau_Y(t) - \tau_X(t)]\,[\mu_Y(t) - \mu_X(t)]\} \tag{27}$$

4 Empirical cumulative measures

Let $X_1, X_2, \dots, X_n$ be non-negative, absolutely continuous, independent and identically distributed (i.i.d.) random variables, that form a random sample drawn from a population having distribution function $F_X(x)$. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the order statistics of the random sample and $U_1 = X_{(1)}$, $U_i = X_{(i)} - X_{(i-1)}$ $(i = 2, 3, \dots, n)$ be the sample spacings. In this section we state some results obtained in [7–12]. The cumulative entropy can be estimated by means of the empirical cumulative entropy which is defined as

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

where

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

is the empirical distribution of the sample. The empirical cumulative entropy can be expressed as

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

where U_2 and U_n possess small weights, whereas the larger weight is given to U_{j+1} such that j is close to $\mathrm{e}^{-1}\,n$. Thus we note from (28) that the empirical cumulative entropy is appropriate to measure variability in right-skewed distributions.

Observe that the standardized empirical cumulative entropy converges in distribution to a standard normal variable as $n \to +\infty$ and $\mathcal{CE}(\hat{F}_n) \to \mathcal{CE}(X)$ a.s. as $n \to +\infty$. Moreover, note that

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

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

Consider now another random sample $Y_1, Y_2, \dots, Y_n$ 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.

Author's personal copy

Moreover, in analogy with (10), we define the empirical cumulative inaccuracy as

$$K[\hat{F}_n,\hat{G}_n] = -\int_0^{+\infty} \hat{F}_n(u)\log\hat{G}_n(u)\,\mathrm{d}u = -\sum_{j=1}^{n-1}\int_{Y_{(j)}}^{Y_{(j+1)}} \hat{F}_n(u)\log\frac{j}{n}\,\mathrm{d}u \tag{29}$$

where $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ are the order statistics of the new sample. Set

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

Moreover, rename by $X_{j,1} < X_{j,2} < \cdots$ the random variables of the first sample belonging to $(Y_{(j)}, Y_{(j+1)}]$, if any. From the above positions we thus have that (29) becomes

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

If X_i and Y_i satisfy condition $X_i \le_{\mathrm{st}} Y_i$, then

$$K[\hat{G}_n,\hat{F}_n] \le_{\mathrm{st}} \mathcal{CE}(X) \le_{\mathrm{st}} K[\hat{F}_n,\hat{G}_n],$$

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

We address the problem of estimating the cumulative KL information by means of an empirical measure of discrimination. The empirical cumulative KL information is defined as

$$\mathcal{C}_{KL}(\hat{F}_n,\hat{G}_m) = -\int_0^{+\infty} \hat{F}_n(x)\,\log\frac{\hat{F}_n(x)}{\hat{G}_m(x)}\,\mathrm{d}x + \overline{X} - \overline{Y},$$

where $\overline{X}$ and $\overline{Y}$ are the sample means.

The empirical cumulative KL information can be rewritten as

$$\mathcal{C}_{KL}(\hat{F}_n,\hat{G}_m) = K[\hat{F}_n,\hat{G}_m] - \mathcal{CE}(\hat{F}_n) + \overline{X} - \overline{Y}.$$

It is easy to prove that, if X and Y are non-negative random variables such that X is in L^p for some $p > 1$ and $X \ge_{st} Y$, then the empirical cumulative KL information of X and Y converges to the cumulative KL information of X and Y, i.e.

$$\mathcal{C}_{KL}(\hat{F}_n,\hat{G}_m) \to \mathcal{C}_{KL}(X,Y) \quad \text{a.s. as } n\to+\infty \text{ and } m\to+\infty.$$

For other details see [7,8,12].

Author's personal copy

5 The cumulative residual KL measure of discrimination

The aim of this last section is the introduction of a new measure of discrimination. The cumulative residual KL information of two random lifetimes X and Y is defined as

$$\mathcal{C}_{KL}(X_t, Y_t) = \int_t^{+\infty} \frac{F_X(x+t) - F_X(t)}{\overline{F}_X(t)} \log\left(\frac{F_X(x+t) - F_X(t)}{F_Y(x+t) - F_Y(t)} \frac{\overline{F}_Y(t)}{\overline{F}_X(t)}\right) \mathrm{d}x + mrl_X(t) - mrl_Y(t) \tag{30}$$

for any $t \in \mathcal{G}_{X,Y} = \{t > 0 : F_X(t) < 1, \quad F_Y(t) < 1\}$. We can write this measure also as

$$\mathcal{C}_{KL}(X_t, Y_t) = \int_t^{+\infty} F_{X_t}(x) \log \frac{F_{X_t}(x)}{F_{Y_t}(x)} \mathrm{d}x + mrl_X(t) - mrl_Y(t).$$

Notice that this measure is the cumulative KL information of the residual lifetimes of X and Y; it measures the closeness of the distributions of the two residual lifetimes, when the items at time $t > 0$ are inspected and both are found alive.

This measure satisfies:

- $\mathcal{C}_{KL}(X_t, Y_t)$ is non-negative for all $t \in \mathcal{G}_{X,Y}$,
- $\lim_{t \to 0^+} \mathcal{C}_{KL}(X_t, Y_t) = \mathcal{C}_{KL}(X, Y)$, and $\lim_{t \to +\infty} \mathcal{C}_{KL}(X_t, Y_t) = 0$.

Now we give an example of computation of this measure for a particular choice of X and Y.

Example 2 Let X and Y have uniform distribution over $(0, a)$ and $(0, b)$, respectively. Then

$$\mathcal{C}_{KL}(X_t, Y_t) = \begin{cases} \frac{a+t}{2} \log \frac{b-t}{a-t} + \frac{a-b}{2}, & \text{if } 0 < a < b \\ \frac{b^2-t^2}{2(a-t)} \log \frac{b-t}{a-t} + \frac{a-b}{2}, & \text{if } 0 < b < a \end{cases}$$

The measure $\mathcal{C}_{KL}(X_t, Y_t)$ is not symmetric with respect to a and b, that is $\mathcal{C}_{KL}(X_t, Y_t) \neq \mathcal{C}_{KL}(Y_t, X_t)$.

Note that $\mathcal{C}_{KL}(X_t, Y_t) \geq 0$, and $\mathcal{C}_{KL}(X_t, Y_t) = 0$ if and only if $F_X(u) = F_Y(u)$ almost everywhere. Also this measure of discrimination has a meaning in reliability theory. In fact, the following proposition can be easily proved.

Proposition 4 *The cumulative residual KL information of X and Y can be expressed as*

$$\mathcal{C}_{KL}(X_t, Y_t) = K[F_{X_t}, F_{Y_t}] - \mathcal{CE}(X_t) + mrl_X(t) - mrl_Y(t), \quad \text{for } t \in \mathcal{G}_{X,Y} \tag{31}$$

Author's personal copy

where $K[F_{X_t}, F_{Y_t}]$ introduces a new cumulative measure that is the cumulative inaccuracy of the residual lifetimes of X_t and Y_t, i.e.

$$K[F_{X_t}, F_{Y_t}] = -\int_t^{+\infty} F_{X_t}(x) \log F_{Y_t}(x)\, \mathrm{d}x \tag{32}$$

and

$$\mathcal{CE}(X_t) = -\int_t^{+\infty} F_{X_t}(x) \log F_{X_t}(x)\, \mathrm{d}x \tag{33}$$

is the cumulative entropy referred to the residual lifetime X_t, for $t \in \mathcal{G}_{X,Y}$, provided that the integrals in (32) and (33) are finite.

For this measure, the following lower bound holds.

Proposition 5 *For all $t \in \mathcal{G}_{X,Y}$,*

$$\mathcal{C}_{KL}(X_t, Y_t) \geq \frac{1}{2} \int_t^{+\infty} \frac{[F_{X_t}(u) - F_{Y_t}(u)]^2}{\frac{1}{3} F_{X_t}(u) + \frac{2}{3} F_{Y_t}(u)}\, \mathrm{d}u. \tag{34}$$

Proof The proof is similar to those of (21) and (25) and follows easily from the fact that the function

$$h(x) = x \log x - x + 1 - \frac{1}{2} \frac{(x-1)^2}{1 + \frac{1}{3}(x-1)}$$

is non-negative for all $x > 0$. Applying this result to (30), with $x = \frac{F_{X_t}(u)}{F_{Y_t}(u)}$, we obtain the statement.

For our new measure it is possible to obtain other lower and upper bounds similar to those studied for the cumulative KL information and the cumulative past KL information. It is our intention to investigate the properties and the applications of the cumulative residual KL discrimination measure in a next research.

Acknowledgments The author thanks the referee for helpful comments and Prof. Antonio Di Crescenzo for useful suggestions which lead to substantial improvements.

References

1. Asadi, M., Zohrevand, Y.: On the dynamic cumulative residual entropy. J. Stat. Plann. Infer. **137**, 1931–1941 (2007)
2. Barlow, R.E., Proschan, F.: Mathematical Theory of Reliability. Wiley, New York (1965)
3. Bowden, R.J.: Information, measure shifts and distribution metrics. Statistics **46**, 249–262 (2010)
4. Di Crescenzo, A.: A probabilistic analogue of the mean value theorem and its applications to reliability theory. J. Appl. Prob. **36**, 706–719 (1999)
5. Di Crescenzo, A., Longobardi, M.: Entropy-based measure of uncertainty in past lifetime distributions. J. Appl. Prob. **39**, 434–440 (2002)

Author's personal copy

6. Di Crescenzo, A., Longobardi, M.: On weighted residual and past entropies. Scientiae Math. Japon. **64**, 255–266 (2006)
7. Di Crescenzo, A., Longobardi, M.: On cumulative entropies. J. Stat. Plann. Infer. **139**, 4072–4087 (2009)
8. Di Crescenzo, A., Longobardi, M.: On cumulative entropies and lifetime estimations. In: Mira J., Ferrandez J.M., Alvarez Sanchez J.R., Paz F., Toledo J. (eds.) Methods and Models in Artificial and Natural Computation, IWINAC 2009, Lecture Notes in Computer Science, vol. 5601, pp. 132–141. Springer, Berlin (2009)
9. Di Crescenzo, A., Longobardi, M.: More on cumulative entropy. In: Trappl, R. (ed.) Cybernetics and Systems 2010, pp. 181–186. Austrian Society for Cybernetic Studies, Vienna (2010)
10. Di Crescenzo, A., Longobardi, M.: Neuronal data analysis based on the empirical cumulative entropy. In: Moreno-Diaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) Computer Aided Systems Theory, EUROCAST 2011, Part I, Lecture Notes in Computer Science, vol. 6927, pp. 72–79. Springer, Berlin (2012)
11. Di Crescenzo, A., Longobardi, M.: Stochastic comparisons of cumulative entropies. In: Li, H., Li, X. (eds.) Stochastic Orders in Reliability and Risk, In Honor of Professor Moshe Shaked. Lecture Notes in Statistics, vol. 208, pp. 167–182. Springer, New York (2013)
12. Di Crescenzo, A., Longobardi, M.: Some properties and applications of cumulative Kullback–Leibler information. Appl. Stoch. Mod. Bus. Ind. (2014, to appear)
13. Ebrahimi, N.: How to measure uncertainty in the residual life time distribution. Sankhyä Ser. A **58**, 48–56 (1996)
14. Ebrahimi, N., Pellerey, F.: New partial ordering of survival functions based on the notion of uncertainty. J. Appl. Prob. **32**, 202–211 (1995)
15. Gupta, R.C., Gupta, R.D.: Proportional reversed hazard rate model and its applications. J. Stat. Plann. Infer. **137**, 3525–3536 (2007)
16. Kayid, M., Ahmad, I.A.: On the mean inactivity time ordering with reliability applications. Probab. Eng. Inform. Sci. **18**, 395–409 (2004)
17. Kerridge, D.F.: Inaccuracy and inference. J. R. Stat. Soc. Ser. B **23**, 184–194 (1961)
18. Misra, N., Gupta, N., Dhariyal, I.D.: Stochastic properties of residual life and inactivity time at a random time. Stoch. Models **24**, 89–102 (2008)
19. Nanda, A.K., Singh, H., Misra, N., Paul, P.: Reliability properties of reversed residual lifetime. Commun. Stat. Theory Methods **32**, 2031–2042 (2003) (with correction in Comm. Statist. Theory Methods **33**, 991–992 (2004))
20. Navarro, J., del Aquila, Y., Asadi, M.: Some new results on the cumulative residual entropy. J. Stat. Plann. Infer. **140**, 310–322 (2010)
21. Park, S., Rao, M., Shin, D.W.: On cumulative residual Kullback–Leibler information. Stat. Prob. Lett. **82**, 2025–2032 (2012)
22. Rao, M., Chen, Y., Vemuri, B.C., Wang, F.: Cumulative residual entropy: a new measure of information. IEEE Trans. Inf. Theory **50**, 1220–1228 (2004)
23. Rezaei, M., Gholizadeh, B., Izadkhah, S.: On relative reversed hazard rate order. Commun. Stat. Theory Meth. (to appear)
24. Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Springer, New York (2007)
25. Shannon, C.E.: A mathematical theory of communication. Bell System Tech. J. **27**, 279–423 (1948)