Research Article

Applied Stochastic Models in Business and Industry

Received 12 December 2013, Revised 14 January 2015, Accepted 2 February 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2116

Some properties and applications of cumulative Kullback–Leibler information

Antonio Di Crescenzo[a*†] and Maria Longobardi[b]

The cumulative Kullback–Leibler information has been proposed recently as a suitable extension of Kullback–Leibler information to the cumulative distribution function. In this paper, we obtain various results on such a measure, with reference to its relation with other information measures and notions of reliability theory. We also provide some lower and upper bounds. A dynamic version of the cumulative Kullback–Leibler information is then proposed for past lifetimes. Furthermore, we investigate its monotonicity property, which is related to some new concepts of relative aging. Moreover, we propose an application to the failure of nanocomponents. Finally, in order to provide an application in image analysis, we introduce the empirical cumulative Kullback–Leibler information and prove an asymptotic result. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: cumulative entropy; cumulative residual entropy; proportional reversed hazard model; relative aging; stochastic order

1. Introduction

Information measures have a fundamental role in informational sciences and other applied fields. As well known, the basic notion describing the information content in a nonnegative absolutely continuous random variable is the differential entropy, also known as Shannon information measure. However, because it may be not suitable for some applications (see, for instance, Section 5 of Schroeder [1]), various authors introduced other suitable information measures, such as the cumulative residual entropy (see Rao *et al.* [2]) and the cumulative entropy (see Di Crescenzo and Longobardi [3] and Navarro *et al.* [4]). The notion of differential entropy has been extended in the past to the relative entropy, which is a discrepancy measure between two distributions. It is also called Kullback–Leibler information (see Kullback and Leibler [5]).

Our research is stimulated by the need of describing the uncertainty content in dynamic stochastic models involving truncated distributions. Indeed, in the last two decades, wide attention has been given by several authors to measure uncertainty of truncated random variables, due to their relevance in applied fields such as reliability theory and survival analysis. One of the first papers in this context is due to Muliere *et al.* [6], who investigated the residual entropy function and the right truncated entropy function. Quite strangely, to a certain extent, this contribution passed unnoticed. The aforementioned measures were investigated later independently by Ebrahimi and Pellerey [7], Ebrahimi [8], and Di Crescenzo and Longobardi [9]. Along these lines, several recent studies involve dynamic information measures based on truncated random variables. See, for instance, Kapodistria and Psarrakos [10], centered on the residual lifetime and its relation with the cumulative residual entropy. Other results on dynamic discrimination measures based on the Kullback–Leibler information have been provided by Khorashadizadeh *et al.* [11], Kumar *et al.* [12], Navarro *et al.* [13], and Taneja *et al.* [14]. Moreover, recently, Park *et al.* [15] defined a discrepancy measure named cumulative Kullback–Leibler information, which can be viewed as the analog of the Kullback–Leibler information with reference to the cumulative distribution function.

The first aim of this paper is to study some properties of the cumulative Kullback–Leibler information. We also investigate a dynamic version of such discrimination measure that is relevant for stochastic models whose uncertainty is related to the past. It is defined in terms of the cumulative Kullback–Leibler information for past lifetimes. Such a measure also involves other dynamic information measures, as the cumulative inaccuracy of past lifetimes and the dynamic cumulative

[a]*Dipartimento di Matematica, Università di Salerno, Via Giovanni Paolo II, n. 132, I-84084 Fisciano (SA), Italy*
[b]*Dipartimento di Matematica e Applicazioni, Università di Napoli Federico II, Via Cintia, I-80126 Napoli, Italy*
**Correspondence to: Antonio Di Crescenzo, Dipartimento di Matematica, Università di Salerno, Via Giovanni Paolo II, n. 132, I-84084 Fisciano (SA), Italy.*
[†]*E-mail: adicrescenzo@unisa.it*

entropy, as well as the mean past lifetimes. Furthermore, we provide two applications. In the first case, we employ the cumulative Kullback–Leibler information to the design of nanocomponents. In the second case, we introduce the empirical cumulative Kullback–Leibler information and study its asymptotic properties for measuring the discrepancies in the gray levels of digitalized images.

This is the plan of the paper: In Section 2, we analyze some properties of the new notion, such as the effect of linear transformations, and its behavior under the proportional reversed hazard model. In Section 3, we develop certain inequalities and bounds for the cumulative Kullback–Leibler information. In Section 4, we investigate the dynamic version of the cumulative Kullback–Leibler information for the past lives of random lifetimes. In Section 5, we provide some results on notions of relative aging and study monotonicity properties of the dynamic cumulative Kullback–Leibler information. Finally, the preannounced applications of the cumulative Kullback–Leibler information to the failure of nanocomponents and in image analysis are given in Sections 6 and 7, respectively.

For readability convenience, the proofs of the theoretical results have been included in the appendix. Moreover, throughout the paper , we denote by $[X|B]$ any random variable having the same distribution of X conditional on B.

2. Cumulative Kullback–Leibler information

We first recall some notions of information theory. Let X and Y be nonnegative absolutely continuous random variables having probability density functions (pdf's) $f(x)$ and $g(x)$, respectively.

- The differential entropy of X is defined by

$$H_X = -\mathbb{E}[\log(f(X))] = -\int_0^{+\infty} f(t)\,\log(f(t))\,\mathrm{d}t, \tag{1}$$

  where 'log' means natural logarithm, $0\log 0 = 0$ by convention. We recall that H_X describes the 'uniformity' of the distribution of X, that is, how the distribution spreads over its domain, and is irrespective of the locations of concentration.
- The cumulative residual entropy of X is defined by

$$\mathcal{E}(X) = -\int_0^{+\infty} \overline{F}(t)\log(\overline{F}(t))\,\mathrm{d}t, \tag{2}$$

  where $\overline{F}(t) = \mathbb{P}(X > t)$ is the survival function of X.
- The cumulative residual entropy of X is defined by

$$\mathcal{CE}(X) = -\int_0^{+\infty} F(t)\log(F(t))\,\mathrm{d}t, \tag{3}$$

  where $F(t) = \mathbb{P}(X \leqslant t)$ is the distribution function of X. The measures defined in (2) and (3) constitute the cumulative residual and the cumulative analogs of the differential entropy given in (1), respectively.
- The relative entropy of X and Y is defined by

$$I_{X,Y} = \int_0^{+\infty} f(t)\log\left(\frac{f(t)}{g(t)}\right)\mathrm{d}t. \tag{4}$$

Equation (4) defines a discrepancy measure between two distributions and is also called Kullback–Leibler information (see Kullback and Leibler [5]). Note that $I_{X,Y}$ measures the inefficiency of assuming that the pdf of a stochastic model is g when the true pdf is f.

In the following, for any random variable T with cumulative distribution function F_T, we denote its left-hand point and right-hand point by

$$l_T = \inf\left\{t \in \mathbb{R} : F_T(t) > 0\right\}, \qquad r_T = \sup\left\{t \in \mathbb{R} : F_T(t) < 1\right\}.$$

Let us now consider a discrimination measure between two random variables X and Y, with distribution functions $F(t)$ and $G(t)$, introduced recently by Park *et al.* [15].

Definition 2.1
Let X and Y be random variables with finite means and having the same left-hand points $l = l_X = l_Y$. The cumulative Kullback–Leibler information of X and Y is defined by

$$\mathcal{C}_{\mathrm{KL}}(X,Y) = \int_l^{\max\{r_X, r_Y\}} F(t) \log\left(\frac{F(t)}{G(t)}\right) \mathrm{d}t + \mathbb{E}(X) - \mathbb{E}(Y), \tag{5}$$

provided that the integral in the right-hand side is finite.

The measure given in (5) can be seen as a suitable extension of the Kullback–Leibler information to the cumulative distribution function. It is suitably rescaled in order to be nonnegative. Indeed, because $\log(x) \leqslant x - 1$ for all $x > 0$, we have $\mathcal{C}_{\mathrm{KL}}(X,Y) \geqslant 0$. Specifically, $\mathcal{C}_{\mathrm{KL}}(X,Y) = 0$ if and only if $F(u) = G(u)$ almost everywhere. Otherwise, $\mathcal{C}_{\mathrm{KL}}(X,Y) \neq \mathcal{C}_{\mathrm{KL}}(Y,X)$, as shown in the following example.

Example 2.1
If X_a and X_b have uniform distribution over $(0,a)$ and $(0,b)$, respectively, then $\mathcal{C}_{\mathrm{KL}}(X_a, X_b) \neq \mathcal{C}_{\mathrm{KL}}(X_b, X_a)$ for $a \neq b$, because from (5), we have

$$\mathcal{C}_{\mathrm{KL}}(X_a, X_b) = \begin{cases} \dfrac{a}{2}\log\left(\dfrac{a}{b}\right) + \dfrac{b-a}{2}, & \text{if } 0 < a < b, \\ \dfrac{(a-b)^2}{4a}, & \text{if } 0 < b < a. \end{cases}$$

Other examples are given in Table I.

Let us analyze the effect of linear transformations on the cumulative Kullback–Leibler information.

Proposition 2.1
Let X and Y be random variables with finite means and having the same left-hand points; for $a, c > 0$ and $b \geqslant 0$, we have

$$\begin{aligned} \mathcal{C}_{\mathrm{KL}}(aX+b, cY+b) &= a\,\mathcal{C}_{\mathrm{KL}}\left(X, \frac{c}{a}Y\right) + a\,\mathbb{E}(X) - c\,\mathbb{E}(Y) \\ &= c\,\mathcal{C}_{\mathrm{KL}}\left(\frac{a}{c}X, Y\right) + a\,\mathbb{E}(X) - c\,\mathbb{E}(Y). \end{aligned}$$

For any pair of random variables X and Y having the same left-hand points l, the *cumulative inaccuracy* is defined by (see Di Crescenzo and Longobardi [16])

$$K(X,Y) = -\int_l^{\max\{r_X, r_Y\}} F(t)\log(G(t))\,\mathrm{d}t, \tag{6}$$

provided that the integral in the right-hand side of (6) is finite. This measure is the analogous of the cumulative residual inaccuracy studied in Taneja and Kumar [17].

Remark 2.1
For any pair of random variables X and Y with finite means, having the same left-hand points, and such that $\mathcal{C}_{\mathrm{KL}}(X,Y)$ is finite, from (5), we have

$$\mathcal{C}_{\mathrm{KL}}(X,Y) = K(X,Y) - \mathcal{CE}(X) + \mathbb{E}(X) - \mathbb{E}(Y), \tag{7}$$

Table I. Examples of cumulative Kullback–Leibler information.

$F_{X_i}(x)$, $i = 1,2$	$\mathcal{C}_{\mathrm{KL}}(X_1, X_2)$
$\exp\left\{-c_i x^{-\gamma_i}\right\}$ $x > 0$, $c_i > 0$, $\gamma_i > 1$	$\frac{c_1^{1/\gamma_1}}{\gamma_1^2}\left[\Gamma\left(-\frac{1}{\gamma_1}\right) + c_1^{-\gamma_2/\gamma_1} c_2 \gamma_1 \Gamma\left(\frac{\gamma_2 - 1}{\gamma_1}\right)\right] + c_1^{1/\gamma_1}\,\Gamma\left(1 - \frac{1}{\gamma_1}\right) - c_2^{1/\gamma_2}\,\Gamma\left(1 - \frac{1}{\gamma_2}\right)$
$\exp\left\{c_i(1 - x^{-1})\right\}$ $0 < x < 1$, $c_i > 0$	$(c_1 - c_2)\left[1 - (1 + c_1)\mathrm{e}^{c_1}\Gamma(0, c_1)\right] + c_1 \mathrm{e}^{c_1}\,\Gamma(0, c_1) - c_2 \mathrm{e}^{c_2}\,\Gamma(0, c_2)$

Here, as usual, $\Gamma(\cdot)$ is the Euler gamma function and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function.

where $\mathcal{CE}(X)$ and $K(X,Y)$ are defined in (3) and (6), respectively.

The reversed hazard rate function of a nonnegative absolutely continuous random variable X, representing a random lifetime, is $\tau_X(t) = f(t)/F(t)$, for $t > 0$.

Remark 2.2
If X and Y are nonnegative absolutely continuous random variables, then $\mathcal{CE}(X)$ and $K(X,Y)$ satisfy (see Proposition 3.1 of [3], and Proposition 8.2.1 of [16]):

$$\mathcal{CE}(X) = \mathbb{E}\left[T_X^{(2)}(X)\right], \qquad K(X,Y) = \mathbb{E}\left[T_Y^{(2)}(X)\right], \tag{8}$$

where

$$T_X^{(2)}(t) := -\int_t^{+\infty} \log(F(s))\,\mathrm{d}s = \int_t^{+\infty}\int_s^{+\infty} \tau_X(x)\,\mathrm{d}x\,\mathrm{d}s, \qquad t > 0, \tag{9}$$

and $T_Y^{(2)}(t)$ is similarly defined.

For any nonnegative absolutely continuous random variable Y, let us now set

$$T_Y(t) := -\log(G(t)) = \int_t^{+\infty} \tau_Y(x)\,\mathrm{d}x, \qquad t \in (0,+\infty). \tag{10}$$

This function is involved in the following representation of the cumulative Kullback–Leibler information in terms of the cumulative entropies of X and Y. Such a representation is based on the probabilistic mean value theorem due to Di Crescenzo [18].

Proposition 2.2
If X and Y are nonnegative absolutely continuous random variables, with finite unequal means, with the same left-hand points, and such that $X \leqslant_{st} Y$ or $Y \leqslant_{st} X$, then

$$C_{\mathrm{KL}}(X,Y) = \mathcal{CE}(Y) - \mathcal{CE}(X) + \left\{\mathbb{E}\left[T_Y(Z)\right] - 1\right\}[\mathbb{E}(Y) - \mathbb{E}(X)], \tag{11}$$

where Z is a nonnegative absolutely continuous random variable with pdf

$$f_Z(z) = \frac{G(z) - F(z)}{\mathbb{E}(X) - \mathbb{E}(Y)}, \qquad 0 < z < +\infty.$$

Table II. Some discrimination measures, for X and Y exponentially distributed with parameters 1 and μ, respectively, where $BC(X,Y) = \int_{-\infty}^{+\infty}\sqrt{f(x)g(x)}\,\mathrm{d}x$ is the Bhattacharya coefficient, $\gamma \simeq 0.577216$ is the Euler's constant and $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

$I_{X,Y}$ (Kullback–Leibler information)	$\frac{1}{\mu} - 1 + \log(\mu)$
$C_{\mathrm{KL}}(X,Y)$ (cumulative Kullback–Leibler information)	$2 - \gamma - \frac{1}{\mu} - \mu + \frac{1}{6}(\mu-1)\pi^2 - \psi(\mu)$
$d_{CM}(X,Y) = \int_{-\infty}^{+\infty}[F(x) - G(x)]^2\,\mathrm{d}F(x)$ (Cramer-von Mises distance)	$\frac{2(\mu-1)^2}{6+15\mu+6\mu^2}$
$d_{R,2}(X,Y) = \log\left[\int_{-\infty}^{+\infty}(f(x))^2(g(x))^{-1}\,\mathrm{d}x\right]$ (Renyi divergence of order 2)	$\log\left(\frac{\mu^2}{2\mu-1}\right)$
$d_E(X,Y) = 2\int_{-\infty}^{+\infty}[F(x) - G(x)]^2\,\mathrm{d}x$ (energy distance)	$\frac{(\mu-1)^2}{\mu+1}$
$d_H(X,Y) = \frac{1}{2}\int_{-\infty}^{+\infty}\left[\sqrt{f(x)} - \sqrt{g(x)}\right]^2\mathrm{d}x$ (Hellinger distance)	$1 - \frac{2\sqrt{\mu}}{\mu+1}$
$d_B(X,Y) = -\log[BC(X,Y)]$ (Bhattacharya distance)	$\log\left(\frac{\mu+1}{2\sqrt{\mu}}\right)$

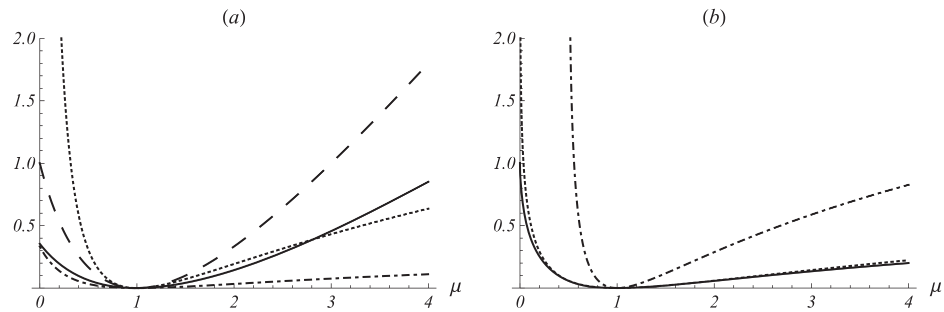

Figure 1. Plot of measures given in Table II for $\mu \in [0, 2]$, with (a) $C_{\rm KL}(X, Y)$ (line), $d_{CM}(X, Y)$ (dot-dashed), $d_E(X, Y)$ (dashed), $I_{X,Y}$ (dotted) and (b) $d_H(X, Y)$ (line), $d_{R,2}(X, Y)$ (dot-dashed), $d_B(X, Y)$ (dotted).

In the next example, the cumulative Kullback–Leibler information is compared with other suitable discrimination measures. Further examples of distances can be found in Finkelstein [19].

Example 2.2
Let X and Y be exponentially distributed with parameters 1 and μ, respectively. In Table II, we evaluate the Kullback–Leibler information and the cumulative Kullback–Leibler information obtained via Equations (4) and (5), and compare them with other discrimination criteria. See Rényi [20] and Frery *et al.* [21] for their definitions. Such measures are also shown in Figure 1 as functions of μ. As expected, all measures are close to 0 when μ approaches 1 and increase monotonically when μ goes away from 1.

We conclude this section by showing a connection with the proportional reversed hazard model. In such a model, the distribution functions of X_θ and X are related by

$$F_\theta(x) = [F(x)]^\theta \qquad \forall x \in \mathbb{R}, \qquad \theta > 0. \tag{12}$$

See, for instance, Di Crescenzo [22], Gupta and Gupta [23], Sankaran and Gleeja [24] for various results about this model.

Proposition 2.3
If X and X_θ are random variables satisfying the proportional reversed hazard model expressed by (12), then

$$C_{\rm KL}(X, X_\theta) = (\theta - 1)\, \mathcal{CE}(X) + \mathbb{E}(X) - \mathbb{E}(X_\theta).$$

3. Inequalities and bounds

Lemma 1 of Ebrahimi *et al.* [25] shows that the differential entropies of two stochastically ordered random variables are suitably related. Hereafter, we provide a similar inequality concerning the cumulative entropy function given in (3). To this aim, and in order to perform some comparisons, we recall some well-known stochastic orders (see Shaked and Shanthikumar [26]). Given two random variables X and Y having distribution functions $F(t)$ and $G(t)$, then X is said to be larger than Y in the

– usual stochastic order (denoted by $X \geqslant_{st} Y$) if $F(t) \leqslant G(t)$ for all $t \in \mathbb{R}$,
– reversed hazard rate order (denoted by $X \geqslant_{rh} Y$) if $F(s)\, G(t) \leqslant F(t)\, G(s)$ for all $s \leqslant t$.

Proposition 3.1
If $X \leqslant_{st} Y$, then

$$\mathcal{CE}(X) - \mathbb{E}(X) \leqslant \mathcal{CE}(Y) - \mathbb{E}(Y).$$

Let us now provide some lower bounds for the cumulative Kullback–Leibler information.

Proposition 3.2
(i) If X and Y are absolutely continuous random variables taking values in $[0, r]$, with r finite, then

$$C_{\rm KL}(X, Y) \geqslant [r - \mathbb{E}(X)] \log\left(\frac{r - \mathbb{E}(X)}{r - \mathbb{E}(Y)}\right) + \mathbb{E}(X) - \mathbb{E}(Y). \tag{13}$$

(ii) If X and Y are random variables with finite means, and having the same left-hand points $l = l_X = l_Y$, then

$$\mathcal{C}_{\mathrm{KL}}(X,Y) \geqslant \frac{3}{2}\int_l^{\max\{r_X,r_Y\}} \frac{[F(u)-G(u)]^2}{F(u)+2G(u)}\,\mathrm{d}u. \tag{14}$$

Making use of inequality $x\log(x) \geqslant x-1$, $x>0$, it is not hard to see that the right-hand side of (13) is nonnegative, and that it vanishes when $\mathbb{E}(X)=\mathbb{E}(Y)$.

The following example shows that there are no implications between the bounds given in Proposition 3.2. Hereafter, we denote by $\mathbf{1}_A$ the indicator function of A.

Example 3.1
Let X be uniformly distributed in $(0,1)$ and let Y have distribution function

$$G(x) = \frac{x}{10}\,\mathbf{1}_{\{0\leqslant x<a\}} + \left[\frac{a}{10}+\left(1-\frac{a}{10}\right)\frac{x-a}{1-a}\right]\mathbf{1}_{\{a\leqslant x<1\}} + \mathbf{1}_{\{x\geqslant 1\}},$$

with $0<a<1$. Figure 2 shows the cumulative Kullback–Leibler information and the bounds obtained in Proposition 3.2 as functions of $a \in (0,1)$.

Hereafter, we obtain some upper bounds for the cumulative Kullback–Leibler information.

Proposition 3.3
Let X and Y be random variables with finite means and having the same left-hand points $l = l_X = l_Y$.

(i) Then,

$$\mathcal{C}_{\mathrm{KL}}(X,Y) \leqslant \int_l^{\max\{r_X,r_Y\}} F(u)\left[\frac{F(u)}{G(u)}-1\right]\mathrm{d}u + \mathbb{E}(X)-\mathbb{E}(Y), \tag{15}$$

provided that the integral in the right-hand side is finite.

(ii) If $X \geqslant_{st} Y$, we have

$$\mathcal{C}_{\mathrm{KL}}(X,Y) \leqslant \frac{1}{2}\int_l^{\max\{r_X,r_Y\}} F(u)\left[\frac{F(u)}{G(u)}-1\right]\left[3-\frac{F(u)}{G(u)}\right]\mathrm{d}u + \mathbb{E}(X)-\mathbb{E}(Y) \tag{16}$$

and

$$\mathcal{C}_{\mathrm{KL}}(X,Y) \leqslant -2.45678\int_l^{\max\{r_X,r_Y\}} F(u)\left[1-\frac{F(u)}{G(u)}\right]^2\mathrm{d}u + \mathbb{E}(X)-\mathbb{E}(Y). \tag{17}$$

We remark that, under assumption $X \geqslant_{st} Y$, the bound shown in (16) is better than the bound provided in (15), in the sense that the right-hand side of (16) is smaller than the right-hand side of (15).

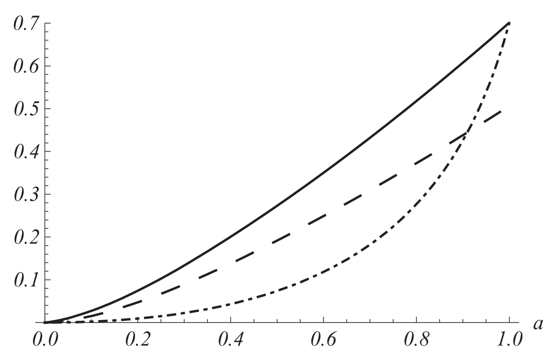

Figure 2. Plot of the cumulative Kullback–Leibler information (line), lower bound (13) (dashed curve), and lower bound (14) (dot-dashed curve) of Example 3.1, for $0<a<1$.

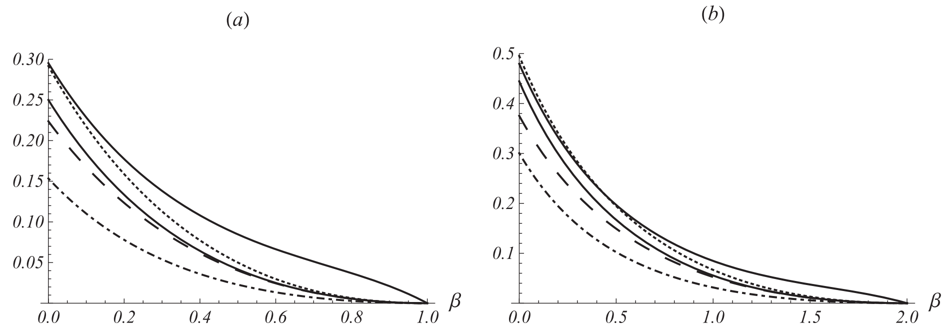

Figure 3. Plot of the cumulative Kullback–Leibler information (18) (lower line), lower bound (13) (dot-dashed curve), lower bound (14) (dashed curve), upper bound (16) (dotted curve) and upper bound (17) (upper line) of Example 3.2, for $0 < \beta \leqslant \alpha$, with $\alpha = 1$ (case *a*) and $\alpha = 2$ (case *b*).

Example 3.2
Let X and Y be random variables having power distribution functions $F(x) = x^\alpha$ and $G(x) = x^\beta$, $0 \leqslant x \leqslant 1$, with $\alpha > 0$ and $\beta > 0$. They satisfy the proportional reversed hazard model $G(x) = [F(x)]^\theta$, $x \in \mathbb{R}$, with $\theta = \beta/\alpha$. From (5), noting that $\mathbb{E}(X) = \alpha/(\alpha+1)$ and $\mathbb{E}(Y) = \beta/(\beta+1)$, we have

$$\mathcal{C}_{\mathrm{KL}}(X,Y) = \frac{(\alpha-\beta)^2}{(1+\alpha)^2(1+\beta)}. \tag{18}$$

For brevity, we omit the expressions of the bounds evaluated previously and provide some plots in Figure 3. We note that in case (b), the upper bounds intersect each other, and thus, there are no implications between them.

4. A dynamic measure for past lifetimes

We consider two nonnegative random variables X and Y, representing the lifetimes of two items, and having support $(0, r_X)$ and $(0, r_Y)$, respectively. For $t > 0$, let $X_{(t)} = [X|X \leqslant t]$ and $Y_{(t)} = [Y|Y \leqslant t]$ be their respective past lifetimes. These notions deserve interest in applied fields such as survival analysis, reliability theory, and actuarial science. For each $t > 0$, we denote the distribution functions of $X_{(t)}$ and $Y_{(t)}$ by

$$F_{(t)}(x) = \frac{F(x)}{F(t)}, \qquad G_{(t)}(x) = \frac{G(x)}{G(t)}, \qquad 0 \leqslant x \leqslant t, \tag{19}$$

respectively. See Sunoj *et al.* [27], Vonta and Karagrigoriou [28], and Sachlas and Papaioannou [29] for various results and applications of past lifetimes.

In analogy with the discrimination measure between past lives given in Definition 1.1 of Di Crescenzo and Longobardi [30], we now introduce a dynamic version of the cumulative Kullback–Leibler information referred to the past lifetimes $X_{(t)}$ and $Y_{(t)}$. It involves the mean past lifetimes of X and Y, which are respectively

$$\mu_X(t) = \mathbb{E}\left[X_{(t)}\right] = \int_0^t \left[1 - \frac{F(x)}{F(t)}\right] \mathrm{d}x, \qquad \mu_Y(t) = \mathbb{E}\left[Y_{(t)}\right] = \int_0^t \left[1 - \frac{G(y)}{G(t)}\right] \mathrm{d}y, \tag{20}$$

for all $t \in \mathcal{D}_{X,Y}$, where $\mathcal{D}_{X,Y} := (0, \min\{r_X, r_Y\})$. In agreement with Definition 2.1, we now consider the following.

Definition 4.1
The cumulative past Kullback–Leibler information of two random lifetimes X and Y, for any $t \in \mathcal{D}_{X,Y}$ is defined by

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) := \int_0^t \frac{F(x)}{F(t)} \log\left(\frac{F(x)}{F(t)}\frac{G(t)}{G(x)}\right) \mathrm{d}x + \mu_X(t) - \mu_Y(t), \tag{21}$$

where $\mu_X(t)$ and $\mu_Y(t)$ are given in (20).

Roughly speaking, $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$ measures the closeness of the distributions of the two lifetimes, when the items are inspected at time $t > 0$, and both are found failed.

Remark 4.1
From Definitions 2.1 and 4.1, it is easy to see that $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$ is nonnegative for all $t \in \mathcal{D}_{X,Y}$, and the following limits hold

$$\lim_{t\to 0^+} \mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) = 0, \qquad \lim_{t\to \sup \mathcal{D}_{X,Y}} \mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) = \mathcal{C}_{\mathrm{KL}}(X, Y).$$

Remark 4.2
Let $t \in \mathcal{D}_{X,Y}$. In analogy with Equation (7), from (21), it follows that the cumulative past Kullback–Leibler information of X and Y can be expressed as

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) = K\left[X_{(t)}, Y_{(t)}\right] - \mathcal{CE}\left(X_{(t)}\right) + \mu_X(t) - \mu_Y(t), \tag{22}$$

where, due to (6) and (19), $K[X_{(t)}, Y_{(t)}]$ is the cumulative inaccuracy of the past lifetimes, that is,

$$K\left[X_{(t)}, Y_{(t)}\right] := -\int_0^t \frac{F(x)}{F(t)} \log\left(\frac{G(x)}{G(t)}\right) \mathrm{d}x, \tag{23}$$

and where

$$\mathcal{CE}\left(X_{(t)}\right) = -\int_0^t \frac{F(x)}{F(t)} \log\left(\frac{F(x)}{F(t)}\right) \mathrm{d}x, \tag{24}$$

is the dynamic cumulative entropy of X, introduced in Section 5 of Di Crescenzo and Longobardi [3].

Note that the cumulative measure given in (23) is an analogous of the dynamic measure of inaccuracy between two past lifetime distributions proposed in Equation (9) of Kumar *et al.* [12].

Let us now obtain some lower and upper bounds for $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$. The next result immediately follows from Proposition 3.2, and thus, the proof is omitted.

Proposition 4.1
For all $t \in \mathcal{D}_{X,Y}$, the following lower bounds hold

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) \geqslant \left[t - \mu_X(t)\right] \log\left(\frac{t-\mu_X(t)}{t-\mu_Y(t)}\right) + \mu_X(t) - \mu_Y(t), \tag{25}$$

and

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) \geqslant \frac{1}{2}\int_0^t \frac{\left[F_{(t)}(x) - G_{(t)}(x)\right]^2}{\frac{1}{3}F_{(t)}(x) + \frac{2}{3}G_{(t)}(x)} \mathrm{d}x. \tag{26}$$

Proposition 4.2
For all $t \in \mathcal{D}_{X,Y}$, the following upper bound holds

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) \leqslant \int_0^t \left[\frac{F_{(t)}^2(u)}{G_{(t)}(u)} - F_{(t)}(u)\right] \mathrm{d}u + \mu_X(t) - \mu_Y(t). \tag{27}$$

Moreover, if $X \geqslant_{rh} Y$, then

$$\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)}) \leqslant \frac{1}{2}\int_0^t F_{(t)}(u)\left[\frac{F_{(t)}(u)}{G_{(t)}(u)} - 1\right]\left[3 - \frac{F_{(t)}(u)}{G_{(t)}(u)}\right] \mathrm{d}u + \mu_X(t) - \mu_Y(t), \tag{28}$$

and

$$\mathcal{C}_{\mathrm{KL}}(X, Y) \leqslant -2.45678\int_0^t F_{(t)}(u)\left[1 - \frac{F_{(t)}(u)}{G_{(t)}(u)}\right]^2 \mathrm{d}u + \mu_X(t) - \mu_Y(t). \tag{29}$$

Example 4.1
Let X and Y have distribution functions

$$F(x) = \exp\left\{-c_1/x^2\right\}, \qquad G(x) = \exp\left\{-c_2/x^2\right\}, \qquad x > 0,$$

respectively, with $c_1, c_2 > 0$. Hence, from Equations (20) and (21), for $t \in \mathcal{D}_{X,Y}$, we have

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) = (c_1 - c_2)\left[\frac{1}{t} - \frac{\sqrt{\pi}\,\exp\left\{c_1/t^2\right\}\,\left(2c_1 + t^2\right)\operatorname{erfc}\left\{\sqrt{c_1}/t\right\}}{2\sqrt{c_1}t^2}\right] \tag{30}$$
$$+ \sqrt{c_1\pi}\,\exp\left\{c_1/t^2\right\}\,\operatorname{erfc}\{\sqrt{c_1}/t\} - \sqrt{c_2\pi}\,\exp\left\{c_2/t^2\right\}\,\operatorname{erfc}\left\{\sqrt{c_2}/t\right\},$$

where erfc$(\cdot)$ is the complementary error function (see Gradshteyn and Ryzhik [31], p. 888). If $c_1 > c_2$, then $X \geqslant_{rh} Y$, and thus, inequalities provided in (28) and (29) hold. Figure 4 shows the plots of the bounds given in Propositions 4.1 and 4.2 for a case when $c_1 > c_2$.

In the following counterexample, we show that the assumption $X \geqslant_{rh} Y$ is necessary to obtain inequalities expressed in (28) and (29).

Counterexample 4.1
Let X be uniformly distributed in (0, 1), and let Y have distribution function

$$G(x) = \left(\frac{4}{3}x\right)^{\alpha}\mathbf{1}_{\{0\leqslant x<1/3\}} + \left(\frac{4}{9}\right)^{\alpha}\mathbf{1}_{\{1/3\leqslant x<2/3\}} + x^{2\alpha}\mathbf{1}_{\{2/3\leqslant x<1\}} + \mathbf{1}_{\{x\geqslant 1\}},$$

with $\alpha > 0$. Hence, it is not hard to see that if $0.5 < \alpha < 1$, then $X \not\geqslant_{rh} Y$. Moreover, in this case, one can show that both the functions given in the right-hand side of (28) and (29) intersect $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$ for suitable choices of α (see an example in Figure 5).

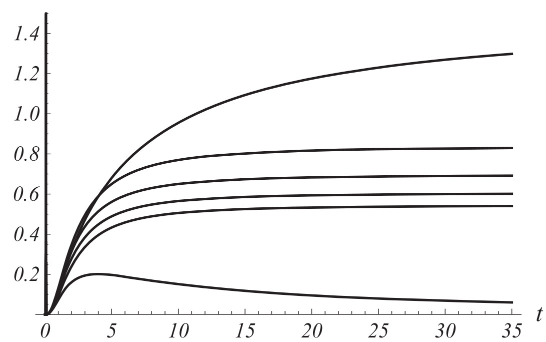

Figure 4. The upper bounds given in (29), (27), and (28), the cumulative past Kullback–Leibler information shown in (30), and the lower bounds provided in (26) and (25) for Example 4.1 (from top to bottom) for $c_1 = 5$ and $c_2 = 1$.

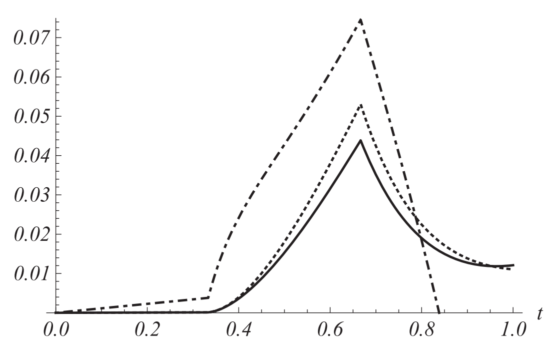

Figure 5. The cumulative past Kullback–Leibler information (line), the right-hand side of (28) (dotted curve), and the right-hand side of (29) (dot-dashed curve) when $\alpha = 0.95$ for Counterexample 4.1.

5. Relative aging and monotonicity property

Let us now provide some results on relative aging that will be used in the succeeding text. To this purpose, consider a nonnegative random variable X, with support $(0, +\infty)$, denoting a random lifetime. We recall that X is said to be increasing (decreasing) failure rate, in short IFR (DFR), if the survival function $\overline{F}(t)$ is logconcave (logconvex), that is, if the hazard rate function $\lambda_X(t) := f(t)/\overline{F}(t)$ is increasing (decreasing) in $t > 0$.

Given two nonnegative random variables X and Y describing random lifetimes, in reliability theory, it is of interest to study the reliability of one variable relative to the other. For instance, we recall that X is said to be aging faster than Y if the random variable $Z = T_Y(X)$ is IFR, where $T_Y(t)$ is defined in Equation (10) (see Definition 1 of Sengupta and Deshpande [32]). This fact introduces a notion of relative aging, in the sense that X is aging faster (slower) than Y if and only if the ratio of hazard rates $\lambda_X(t)/\lambda_Y(t)$ is increasing (decreasing) in t. See also the contribution by Rowell and Siegrist [33] on this topic.

We now consider an analogous concept, which involves the ratio of reversed hazard rates of X and Y.

Proposition 5.1
Given two nonnegative random variables X and Y, the following statements are equivalent:

(i) $Z = T_X(Y)$ is IFR (DFR);
(ii) $T_Y \circ T_X^{-1}$ is convex (concave);
(iii) $\tau_X(t)/\tau_Y(t)$ is increasing (decreasing) in t.

We remark that condition (iii) of Proposition 5.1 is involved in point (b) of Theorem 1.C.4 of Shaked and Shanthikumar [26], where it is stated that if $X \leqslant_{rh} Y$ and $\tau_X(t)/\tau_Y(t)$ is decreasing in t, then X is smaller than Y in the likelihood ratio order.

Condition (iii) of Proposition 5.1 will be used in Proposition 5.2. It is not hard to see that such a condition defines a stochastic order (see Rezaei *et al.* [34]).

Let us now investigate the monotonicity property of $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$.

Theorem 5.1
Let X and Y be nonnegative absolutely continuous random variables, and let $t \in \mathcal{D}_{X,Y}$. Then, $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$ is increasing in t if and only if

$$\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right) \leqslant B(t) \qquad \forall t \in \mathcal{D}_{X,Y}, \tag{31}$$

where

$$B(t) := \left[\frac{\tau_Y(t)}{\tau_X(t)} - 1\right]\left[\mu_Y(t) - \mu_X(t)\right].$$

In the following proposition, we find sufficient conditions such that $B(t)$ is increasing. The proof easily follows from the definitions of the involved notions, and thus is omitted.

Proposition 5.2
Let $\tau_X(t)/\tau_Y(t)$ be decreasing in t, and let $\mu_Y(t) - \mu_X(t)$ be increasing in t. Then, $B(t)$ is increasing in t.

Example 5.1
Let $F(t) = \left(\frac{t}{r}\right)^{\alpha}$ and $G(t) = \left(\frac{t}{r}\right)^{\beta}$, for $0 \leqslant t \leqslant r$. If $0 < \alpha < \beta$, then the assumptions of Proposition 5.2 are satisfied. In this case, it is easy to check that $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$ and $B(t)$ are both linearly increasing in t.

6. Application to failure of nanocomponents

The design of nanocomponents with specific physical properties often requires the assessing of items reliability. In this framework, a current problem is related to the contiguous nanocrack. With reference to Ebrahimi [35], we recall that a nanocrack is a region of a nanocomponent where there is no bonding between adjacent atoms of the lattices. A contiguous nanocrack is formed when the accumulated number of broken bonds for the crack reaches a certain threshold. Denoting by b such threshold, and by S the random stress level, then, according to the Maxwell–Boltzmann distribution of thermal energies of atoms, the net probability of permanent bond breakage at the given stress level $S = s$ is $2\exp(-\alpha)\sinh(\beta\gamma s)$, $s \geqslant 0$, where α, β, and γ are suitable parameters. Specifically, $\alpha = \frac{Q}{kT}$ and $\beta = \frac{1}{kT}$, where T is the absolute temperature, Q is the activation energy, and k is the Boltzmann constant, whereas γ is an appropriately chosen constant (see Ebrahimi

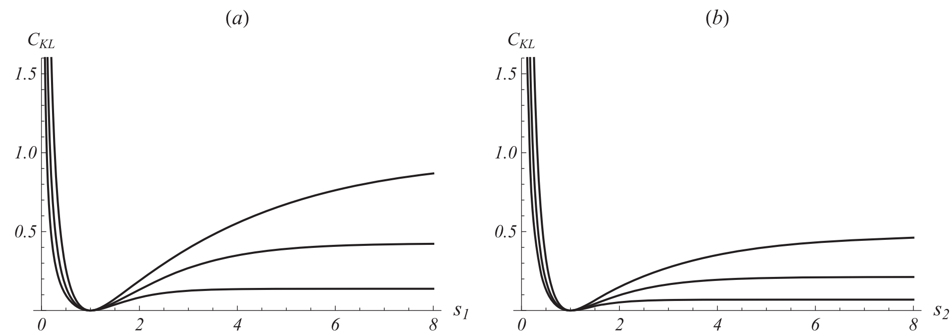

Figure 6. Plot of the cumulative Kullback–Leibler information given in (33) for $e^{\alpha}/c_0^* = 1$, $b = 1$, and $\beta\gamma = 0.5, 1, 2$ (from top to bottom), with (a) $0 < s_1 \leqslant 8$ and $s_2 = 1$, (b) $0 < s_2 \leqslant 8$ and $s_1 = 1$.

[35] for details). Hence, if X denotes the load duration needed for the accumulated number of broken bonds to reach the threshold b, then the distribution function of $[X|S = s]$ is (cf. Equation (1) of Ebrahimi [35])

$$F_{X|S}(x|s) = \left(\frac{x}{r(s)}\right)^b, \qquad 0 \leqslant x \leqslant r(s) := \frac{e^{\alpha}}{c_0^*}\,\mathrm{csch}(\beta\gamma s), \tag{32}$$

for $b \geqslant 1$, $s > 0$, $\alpha > 0$, $\beta\gamma > 0$, and where $c_0^* > 0$ is an appropriately chosen constant. In this section, we adopt the notation X_s to represent the random variable $[X|S = s]$, that is, the load duration leading to the formation of a contiguous nanocrack conditional on the stress level $S = s$. Aiming to study the dependence between pairs of random variables belonging to the family $\{X_s; s > 0\}$, from (5) and (32), we obtain the following expression of the cumulative Kullback–Leibler information:

$$\mathcal{C}_{\mathrm{KL}}\left(X_{s_1}, X_{s_2}\right) = \begin{cases} \dfrac{b^2}{b+1}\, r(s_1) \log\left(\dfrac{r(s_1)}{r(s_2)}\right) - b\,\mathcal{B}_b(s_1, s_2), & 0 < s_2 \leqslant s_1, \\ \dfrac{b}{(b+1)^2}\, r(s_1) \left\{ \left[\dfrac{r(s_2)}{r(s_1)}\right]^{b+1} - 1 \right\} + \mathcal{B}_b(s_1, s_2), & 0 < s_1 \leqslant s_2, \end{cases} \tag{33}$$

where

$$\mathcal{B}_b(s_1, s_2) = \frac{b\left[r(s_1) - r(s_2)\right]}{b+1}.$$

Figure 6 shows some plots of $\mathcal{C}_{\mathrm{KL}}(X_{s_1}, X_{s_2})$ for various choices of the involved parameters.

From (33), it is easy to check that $\mathcal{C}_{\mathrm{KL}}(X_s, X_s) = 0$, $s > 0$. Moreover, the following limits hold:

$$\lim_{s_1 \to 0} \mathcal{C}_{\mathrm{KL}}(X_{s_1}, X_{s_2}) = \lim_{s_2 \to 0} \mathcal{C}_{\mathrm{KL}}(X_{s_1}, X_{s_2}) = +\infty,$$

$$\lim_{s_1 \to +\infty} \mathcal{C}_{\mathrm{KL}}(X_{s_1}, X_{s_2}) = \frac{ab^2}{b+1}\,\mathrm{csch}(cs_2), \qquad \lim_{s_2 \to +\infty} \mathcal{C}_{\mathrm{KL}}(X_{s_1}, X_{s_2}) = \frac{ab^2}{(b+1)^2}\,\mathrm{csch}(cs_1).$$

This shows that the discrepancy between X_{s_1} and X_{s_2} measured through the cumulative Kullback–Leibler information tends to become infinitely large when the stress level s_1 or s_2 goes to 0, whereas $\mathcal{C}_{\mathrm{KL}}(X_{s_1}, X_{s_2})$ tends to a finite value when s_1 or s_2 goes to infinity. Hence, consider the load durations leading to the formation of a contiguous nanocrack conditional on two different stress levels $S = s_1$ and $S = s_2$. Such load durations tend to be very different when the stress level s_1 or s_2 is close to 0. On the contrary, if the stress level s_1 or s_2 grows, then the discrepancy between the load durations does not grow indefinitely because it tends to a constant. This information may help engineers to design nanocomponents. For instance, aiming to affect the load duration that causes a contiguous nanocrack, the knowledge that the baseline stress level is large implies that it is not useful to reduce the stress slightly.

7. Application in image analysis

Information measures are often employed in image analysis as well as in other applied fields. For instance, we recall that the cumulative entropy and its conditional version have been successfully exploited by Nguyen *et al.* [36] to perform cluster analysis on multidimensional data. Moreover, with reference to the cumulative residual entropy defined in (2), Wang *et al.* [37] proposed the empirical version of such a measure to solve the alignment problem in image analysis. We also recall the papers by Contreras-Reyes and Arellano-Valle [38] and Frery *et al.* [21], where useful methods of information theory are proposed and applied to investigate images obtained from real sensors. Specifically, their studies involve the flexible families of multivariate skew-normal distribution, and Wishart distribution, respectively.

On the line of some previous investigations, in this section, we focus on a suitable application of the cumulative Kullback–Leibler information in image analysis. To this purpose, we first address the problem of estimating the cumulative Kullback–Leibler information by means of an empirical measure of discrimination. Consider two independent random samples $\mathcal{X} = (X_i;\ i = 1, 2, \dots, n)$ and $\mathcal{Y} = (Y_j;\ j = 1, 2, \dots, m)$, where X_i and Y_j are independent copies of X and Y, respectively. According to (5), we give the following.

Definition 7.1
Let X and Y be absolutely continuous and nonnegative. The empirical cumulative Kullback–Leibler information of X and Y is defined by

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

where

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

are the empirical distributions of the samples, and where $\overline{X}_n$ and $\overline{Y}_m$ are the sample means.

In order to obtain a computationally effective formula for $\mathcal{C}_{\mathrm{KL}}(\hat{F}_n, \hat{G}_m)$, we introduce the order statistics of the first random sample, $X_{(1)} \leqslant X_{(2)} \leqslant \dots \leqslant X_{(n)}$. Similarly, we denote by $Y_{(1)} \leqslant Y_{(2)} \leqslant \dots \leqslant Y_{(m)}$ the order statistics of $\mathcal{Y}$, and set $Y_{(0)} = -\infty$.

Remark 7.1
In agreement with (7), it is not hard to see that the empirical cumulative Kullback–Leibler information can be rewritten as

$$\mathcal{C}_{\mathrm{KL}}\left(\hat{F}_n, \hat{G}_m\right) = K\left[\hat{F}_n, \hat{G}_m\right] - \mathcal{CE}\left(\hat{F}_n\right) + \overline{X}_n - \overline{Y}_m. \tag{34}$$

Note that

(i) $\mathcal{CE}(\hat{F}_n)$ is the empirical cumulative entropy (see Di Crescenzo and Longobardi [3])

$$\begin{aligned}\mathcal{CE}(\hat{F}_n) &= -\int_0^{+\infty} \hat{F}_n(x) \log\left(\hat{F}_n(x)\right) \mathrm{d}x \\ &= -\sum_{i=1}^{n-1} U_{i+1}\, \frac{i}{n} \log\left(\frac{i}{n}\right),\end{aligned} \tag{35}$$

where $U_1 = X_{(1)}$, $U_i = X_{(i)} - X_{(i-1)}$, for $i = 2, 3, \dots, n$, are the sample spacings of the first sample;

(ii) $K[\hat{F}_n, \hat{G}_m]$ is the empirical cumulative inaccuracy (see Di Crescenzo and Longobardi [16], for the case $n = m$)

$$\begin{aligned}K\left[\hat{F}_n, \hat{G}_m\right] &= -\int_0^{+\infty} \hat{F}_n(u) \log\left(\hat{G}_m(u)\right) \mathrm{d}u \\ &= -\sum_{j=1}^{m-1} \int_{Y_{(j)}}^{Y_{(j+1)}} \hat{F}_n(u) \log\left(\frac{j}{m}\right) \mathrm{d}u \\ &= \frac{1}{n}\sum_{j=1}^{m-1}\left[\sum_{r=1}^{N_{j+1}-N_j} X_{j,r} + N_j Y_{(j)} - N_{j+1} Y_{(j+1)}\right] \log\left(\frac{j}{m}\right),\end{aligned} \tag{36}$$

where

$$N_j := \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leqslant Y_{(j)}\}}, \qquad j = 1, 2, \dots, m,$$

denotes the number of random variables of the sample $\mathcal{X}$ that are less than or equal to the j-th order statistic of $\mathcal{Y}$, and where we rename by $X_{j,r}$, $r = 1, 2, \dots, (N_{j+1} - N_j)$, the random variables of the first sample belonging to the interval $(Y_{(j)}, Y_{(j+1)}]$, if any.

The following asymptotic result holds.

Proposition 7.1
Let X and Y be nonnegative random variables. If X is in L^p for some $p > 1$, and $X \geqslant_{st} Y$ then the empirical cumulative Kullback–Leibler information of X and Y converges to the cumulative Kullback–Leibler information of X and Y, that is,

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

Example 7.1
Let us now adopt the empirical cumulative Kullback–Leibler information to measure the discrepancies in the gray levels of the digitalized images shown in Figure 7. The pictures are formed by 1600 cells; the gray level of each cell being measured by a real number ranging from 0 (black) to 1 (white). Figure 8 shows the gray levels arranged in increasing order. Hence, a

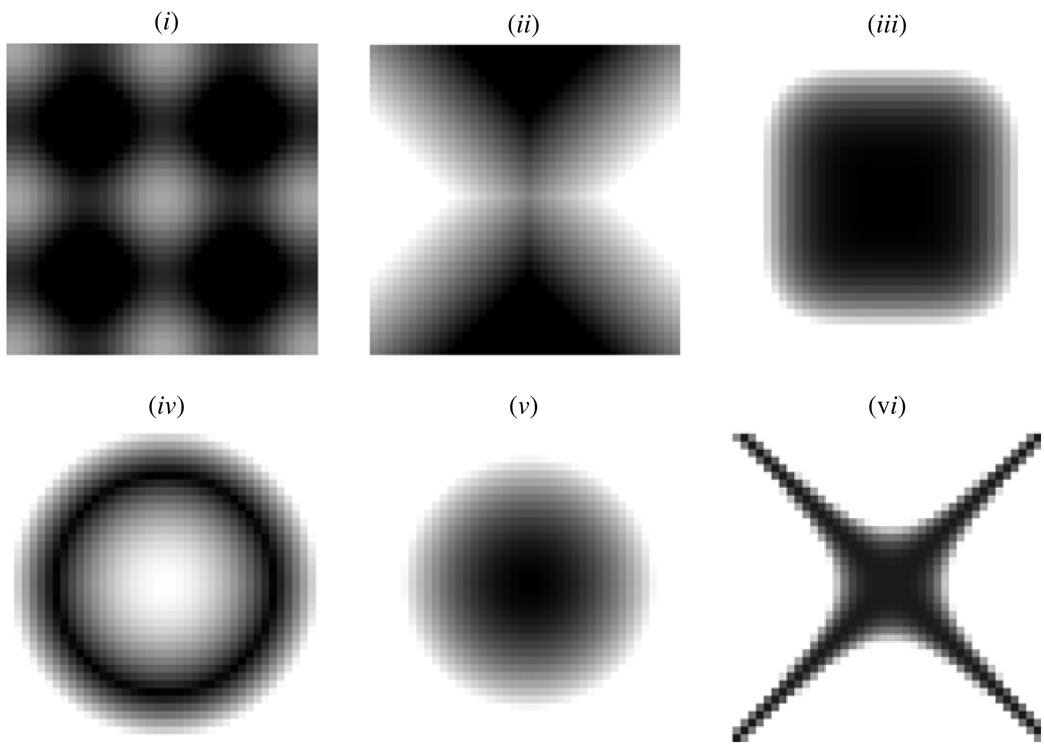

Figure 7. Images studied in Section 7.

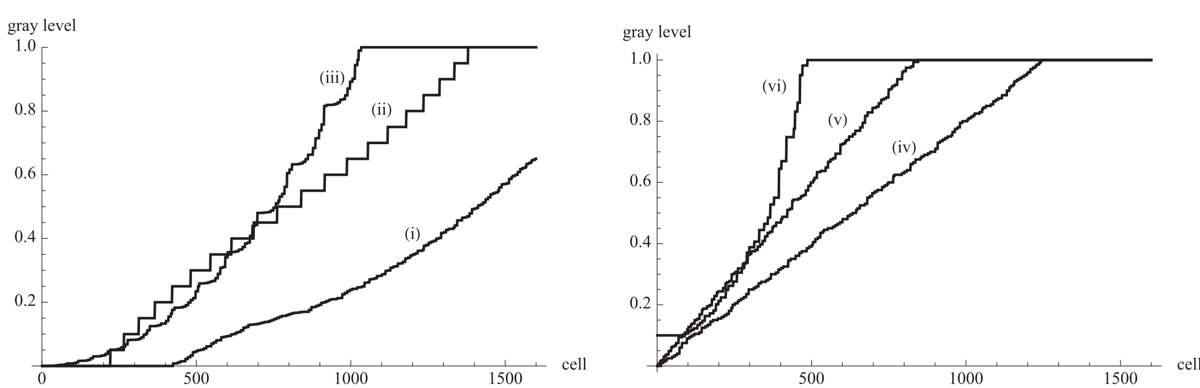

Figure 8. Ordered gray levels of the images in Figure 7.

Table III. Mean gray levels of the images of Figure 7.

Image	Mean gray level
(*i*)	0.205311
(*ii*)	0.5
(*iii*)	0.562977
(*iv*)	0.607250
(*v*)	0.738219
(*vi*)	0.802999

Table IV. $C_{\mathrm{KL}}(\hat{F}_n, \hat{G}_m)$ for the images of Figure 7.

$X \setminus Y$	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)	(*vi*)
(*i*)	0	0.0937469	0.122712	0.218842	0.414844	0.402350
(*ii*)	0.0755055	0	0.0110273	0.0244186	0.0984964	0.109691
(*iii*)	0.0987752	0.0100067	0	0.0216801	0.0688432	0.0583435
(*iv*)	0.149475	0.0182159	0.0168919	0	0.0285125	0.0640901
(*v*)	0.258760	0.0729513	0.0510091	0.0248633	0	0.00282724
(*vi*)	0.332174	0.127517	0.0899049	0.0662241	0.0151126	0

graph close to 0, such as case (i) of Figure 8, corresponds to a picture with many black cells. Moreover, the mean gray level of the six images are presented in Table III. The empirical cumulative Kullback–Leibler information of the gray levels for the pairs of images are evaluated by means of Equation (34), for $n = m = 1600$ (Table IV). The maximum discrepancy is found between images (i) and (v), whereas the minimum value of $C_{\mathrm{KL}}(\hat{F}_n, \hat{G}_m)$ is attained for images (v) and (vi).

Example 7.1 provides a suitable application of the empirical cumulative Kullback–Leibler information in image analysis. In this context, starting from the gray levels of images in random samples $\mathcal{X}$ and $\mathcal{Y}$, higher values of $C_{\mathrm{KL}}(\hat{F}_n, \hat{G}_m)$ correspond to marked differences in the nuances of the figures. Therefore, an automatic procedure that evaluates the empirical cumulative Kullback–Leibler information may allow to categorize images on the base of their gray (or color) levels.

8. Concluding remarks

The present paper is centered on the study of the cumulative Kullback–Leibler information. This is an asymmetric measure of discrimination for pairs of random variables. We have first investigated some properties, inequalities, and bounds of this measure. The analysis also focused on its dynamic version, which is specially suited for past lifetimes. A monotonicity property of its upper bound has been related to a quite new notion of relative aging. In conclusion, in order to shade some light on the practical use of this measure, we considered two applications:
(i) The first is about the failures of nanocomponents, with reference to a stochastic model for the load duration needed for the number of broken bonds that leads to a contiguous nanocrack. The cumulative Kullback–Leibler information in this case may give some advice on the role of the stress level involved in this model.
(ii) In the second case, the empirical cumulative Kullback–Leibler information is applied in images analysis. After providing a convergence result, we adopt this empirical measure to compare the gray level in some images. Differently from various papers mentioned in Section 7, our approach is nonparametric. The extension to cases involving specific families of distributions is too heavy in this context, and thus will be the object of future investigation.

Appendix

Proof of Proposition 2.1
It immediately follows from Definition 2.1, and recalling that the distribution function of $aX + b$ is $F\left(\frac{x-b}{a}\right)$, $x \in \mathbb{R}$. $\square$

Proof of Proposition 2.2
From Theorem 4.1 of [18], for a measurable and differentiable function $h(t)$ such that $\mathbb{E}[h(Y)]$ and $\mathbb{E}[h(X)]$ are finite, and its derivative $h'(t)$ is measurable and Riemann-integrable of the interval $[x, y]$ for all $y \geqslant x \geqslant 0$, we have $\mathbb{E}[h(Y)] - E[h(X)] = \mathbb{E}[h'(Z)][\mathbb{E}(Y) - \mathbb{E}(X)]$, under the given assumptions. Hence, by choosing $h(t) = T_Y^{(2)}(t)$, and thus $h'(t) = -T_Y(t)$

due to (9) and (10), and recalling the identities given in (8), we obtain (see also Proposition 8.2.2 of Di Crescenzo and Longobardi [16]):

$$\mathcal{CE}(Y) - K(X,Y) = -\mathbb{E}\left[T_Y(Z)\right][\mathbb{E}(Y) - \mathbb{E}(X)].$$

Hence, Equation (11) follows by use of (7). □

Proof of Proposition 2.3
Because of Equation (12), recalling Equations (3) and (6), we have $K(X, X_\theta) = \theta\, \mathcal{CE}(X)$. The proof then immediately follows from (7). □

Proof of Proposition 3.1
From Equation (7), recalling that $\mathcal{C}_{\mathrm{KL}}(X,Y) \geqslant 0$, we have

$$\mathcal{CE}(X) - \mathbb{E}(X) \leqslant K(X,Y) - \mathbb{E}(Y).$$

Proposition 3.1 of Di Crescenzo and Longobardi [16] states that assumption $X \leqslant_{st} Y$ implies $K(X,Y) \leqslant \mathcal{CE}(Y)$. The proof then follows. □

Proof of Proposition 3.2
The proof of (13) immediately follows applying the well-known log-sum inequality (see, for instance, Cover and Thomas [39]) to the pdf's of X and Y. Because

$$x\log(x) \geqslant x - 1 + \frac{3(x-1)^2}{2(x+2)} \qquad \text{for all } x > 0,$$

the proof of (14) follows from Equation (5), by taking $x = F(u)/G(u)$. □

Proof of Proposition 3.3
Recalling that $\log(x) \leqslant x - 1$, $\forall x > 0$, the proof of (15) follows from Definition 2.1. From the following inequalities

$$\log(x) \leqslant \frac{1}{2}(x-1)(3-x), \qquad \log(x) \leqslant -2.45678(1-x)^2, \qquad 0 < x \leqslant 1,$$

and the definition of $\mathcal{C}_{\mathrm{KL}}(X,Y)$, we easily obtain the proof of (16) and (17). □

Proof of Proposition 4.2
We note that $X_{(t)} \geqslant_{st} Y_{(t)}$ for all $t \in \mathcal{D}_{X,Y}$ if, and only if, $X \geqslant_{rh} Y$. The proof thus follows from Proposition 3.3. □

Proof of Proposition 5.1
The equivalence between (i) and (ii) has been shown in Theorem 5.1 of Di Crescenzo [22]. Noting that the hazard rate function of $Z = T_X(Y)$ is

$$\lambda_Z(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\log(\overline{F}_Z(t)) = -\tau_Y\left(T_X^{-1}(t)\right)\frac{\mathrm{d}}{\mathrm{d}t}T_X^{-1}(t) = \frac{\tau_Y\left(T_X^{-1}(t)\right)}{\tau_X\left(T_X^{-1}(t)\right)}, \qquad t > 0,$$

the equivalence between (i) and (iii) immediately follows. □

Proof of Theorem 5.1
From (23) we have

$$\frac{\mathrm{d}}{\mathrm{d}t}K\left[X_{(t)}, Y_{(t)}\right] = \tau_Y(t)\left[t - \mu_X(t)\right] - \tau_X(t)\,K\left[X_{(t)}, Y_{(t)}\right],$$

whereas (see Section 6, proof of Theorem 6.1, Equation (32) of Di Crescenzo and Longobardi [3])

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{CE}\left(X_{(t)}\right) = \tau_X(t)\left[t - \mu_X(t) - \mathcal{CE}\left(X_{(t)}\right)\right].$$

Moreover, from (20), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_X(t) = \tau_X(t)[t - \mu_X(t)].$$

Hence, recalling Equation (22), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)}) = \left[\tau_Y(t) - \tau_X(t)\right]\left[\mu_Y(t) - \mu_X(t)\right] - \tau_X(t)\,\mathcal{C}_{\mathrm{KL}}\left(X_{(t)}, Y_{(t)}\right).$$

It thus follows that $\mathcal{C}_{\mathrm{KL}}(X_{(t)}, Y_{(t)})$ is increasing in t if and only if the inequality given in Equation (31) holds. □

Proof of Proposition 7.1
By the law of large numbers, we have $\overline{X}_n \to \mathbb{E}(X)$ a.s. as $n \to +\infty$, and $\overline{Y}_m \to \mathbb{E}(Y)$ a.s. as $m \to +\infty$. Moreover, under the given assumptions $\mathcal{CE}(\hat{F}_n) \to \mathcal{CE}(X)$ a.s. as $n \to +\infty$ (see Proposition 2 of Di Crescenzo and Longobardi [40]). Similarly, in the same line of Theorem 9 of Rao *et al.* [2] making use of assumption $F(t) \leqslant G(t)$ for all $t \in \mathbb{R}$, one can prove that $K[\hat{F}_n, \hat{G}_m] \to K(X, Y)$ a.s. as $n \to +\infty$ and $m \to +\infty$. The proof thus follows immediately from the relations given in (7) and (34). □

Acknowledgements

The authors warmly thank the associate editor and two anonymous reviewers for several helpful comments that lead to substantial improvements and acknowledge the partial support by Project FARO (University of Naples Federico II), GNCS-INdAM, Regione Campania (legge 5), and Project 'AMALFI: Metodologie Avanzate per l'AnaLisi e la gestione della Internet del Futuro' (University of Turin/Compagnia San Paolo).

References

1. Schroeder MJ. An alternative to entropy in the measurement of information. *Entropy* 2004; **6**:388–412.
2. Rao M, Chen Y, Vemuri BC, Wang F. Cumulative residual entropy: a new measure of information. *IEEE Transactions in Information Theory* 2004; **50**:1220–1228.
3. Di Crescenzo A, Longobardi M. On cumulative entropies. *Journal of Statistical Planning and Inference* 2009; **139**:4072–4087.
4. Navarro J, del Aguila Y, Asadi M. Some new results on the cumulative residual entropy. *Journal of Statistical Planning and Inference* 2010; **140**:310–322.
5. Kullback S, Leibler RA. On information and sufficiency. *Annals of Mathematical Statistics* 1951; **22**:79–86.
6. Muliere P, Parmigiani G, Polson NG. A note on the residual entropy function. *Probability in the Engineering and Informational Sciences* 1993; **7**:413–420.
7. Ebrahimi N, Pellerey F. New partial ordering of survival functions based on the notion of uncertainty. *Journal of Applied Probability* 1995; **32**:202–211.
8. Ebrahimi N. How to measure uncertainty in the residual life time distribution. *Sankhyā-The Indian Journal of Statistics Series A* 1996; **58**:48–56.
9. Di Crescenzo A, Longobardi M. Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied Probability* 2002; **39**:434–440.
10. Kapodistria S, Psarrakos G. Some extensions of the residual lifetime and its connection to the cumulative residual entropy. *Probability in the Engineering and Informational Sciences* 2012; **26**:129–146.
11. Khorashadizadeh M, Rezaei Roknabadi AH, Mohtashami Borzadaran GR. Doubly truncated (interval) cumulative residual and past entropy. *Statistics and Probability Letters* 2013; **83**:1464–1471.
12. Kumar V, Taneja HC, Srivastava R. A dynamic measure of inaccuracy between two past lifetime distributions. *Metrika* 2011; **74**:1–10.
13. Navarro J, Sunoj SM, Linu MN. Characterizations of bivariate models using dynamic Kullback-Leibler discrimination measures. *Statistics and Probability Letters* 2011; **81**:1594–1598.
14. Taneja HC, Kumar V, Srivastava R. A dynamic measure of inaccuracy between two residual lifetime distributions. *International Mathematical Forum* 2009; **4**:1213–1220.
15. Park S, Rao M, Shin DW. On cumulative residual Kullback-Leibler information. *Statistics and Probability Letters* 2012; **82**:2025–2032.
16. Di Crescenzo A, Longobardi M. Stochastic comparisons of cumulative entropies. In *Stochastic Orders in Reliability and Risk, In Honor of Professor Moshe Shaked*, Li H, Li X (eds), Lecture Notes in Statistics 208. Springer: New York, 2013; 167–182.
17. Taneja HC, Kumar V. On dynamic cumulative residual inaccuracy measure. In *Proceedings of the World Congress on Engineering*, Ao SI, Gelman L, Hukins DWL, Hunter A, Korsunsky AM (eds), vol. 1. WCE: London, UK, 2012; 153–156.
18. Di Crescenzo A. A probabilistic analogue of the mean value theorem and its applications to reliability theory. *Journal of Applied Probability* 1999; **36**:706–719.
19. Finkelstein MS. On some measures and distances for positive random variables. *Applied Stochastic Models in Business and Industry* 2003; **19**:133–146.
20. Rényi A. *Probability Theory*, North-Holland Series in Applied Mathematics and Mechanics, vol. 10. North-Holland Publishing Co.; American Elsevier Publishing Co., Inc.: Amsterdam-London; New York, 1970.

21. Frery AC, Nascimento ADC, Cintra RJ. Information theory and image understanding: an application to polarimetric SAR imagery. *Chilean Journal of Statistics* 2011; **2**:81–100.
22. Di Crescenzo A. Some results on the proportional reversed hazards model. *Statistics and Probability Letters* 2000; **50**:313–321.
23. Gupta RC, Gupta RD. Proportional reversed hazard rate model and its applications. *Journal of Statistical Planning and Inference* 2007; **137**:3525–3536.
24. Sankaran PG, Gleeja VL. Proportional reversed hazard and frailty models. *Metrika* 2008; **68**:333–342.
25. Ebrahimi N, Soofi ES, Soyer R. When are observed failures more informative than observed survivals? *Naval Research Logistics* 2013; **60**: 102–110.
26. Shaked M, Shanthikumar JG. *Stochastic Orders*. Springer: New York, 2007.
27. Sunoj SM, Sankaran PG, Nanda AK. Quantile based entropy function in past lifetime. *Statistics and Probability Letters* 2013; **83**:366–372.
28. Vonta F, Karagrigoriou A. Generalized measures of divergence in survival analysis and reliability. *Journal of Applied Probability* 2010; **47**: 216–234.
29. Sachlas A, Papaioannou T. Residual and past entropy in actuarial science and survival models. *Methodology and Computing in Applied Probability* 2014; **16**:79–99.
30. Di Crescenzo A, Longobardi M. A measure of discrimination between past lifetime distributions. *Statistics and Probability Letters* 2004; **67**: 173–182.
31. Gradshteyn IS, Ryzhik IM. *Table of Integrals, Series, and Products* (7th edn). Elsevier/Academic Press: Amsterdam, 2007.
32. Sengupta D, Deshpande JV. Some results on the relative ageing of two life distributions. *Journal of Applied Probability* 1994; **31**:991–1003.
33. Rowell G. Siegrist K. Relative aging of distributions. *Probability in the Engineering and Informational Sciences* 1998; **12**:469–478.
34. Rezaei M, Gholizadeh B, Izadkhah S. On relative reversed hazard rate order. *Communications in Statistics - Theory and Methods* 2015; **44**: 300–308.
35. Ebrahimi N. Assessing the reliability of a nanocomponent using nanocrack growth. *Applied Stochastic Models in Business and Industry* 2014; **30**:294–302.
36. Nguyen HV, Müller E, Vreeken J, Keller F, Böhm K. CMI: An information-theoretic contrast measure for enhancing subspace cluster and outlier detection. In *Proceedings of the SIAM International Conference on Data Mining (SDM'13)*. SIAM: Austin, Texas, USA, 2013; 1–9.
37. Wang F, Vemuri BC, Rao M, Chen Y. Cumulative residual entropy, a new measure of information & its application to image alignment. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, vol. 1. IEEE Computer Society Press: Silver Spring, MD, 2003; 548–553.
38. Contreras-Reyes JE, Arellano-Valle RB. Kullback–Leibler divergence measure for multivariate skew-normal distributions. *Entropy* 2012; **14**:1606–1626.
39. Cover TM, Thomas JA. *Elements of Information Theory*. John Wiley and Sons: New York, 1991.
40. Di Crescenzo A, Longobardi M. On cumulative entropies and lifetime estimations. In *IWINAC 2009*, Mira J, Ferrández JM, Álvarez JR, de la Paz F, Javier Toledo F (eds), Part I, Lecture Notes in Computer Sciences 5601. Springer-Verlag: Berlin, 2009; 132–141.