arXiv:math/0703489v1 [math.ST] 16 Mar 2007

ON WEIGHTED RESIDUAL AND PAST ENTROPIES*

Antonio Di Crescenzo
Dipartimento di Matematica e Informatica
Università di Salerno
Via Ponte don Melillo
I-84084 Fisciano (SA), Italy
adicrescenzo@unisa.it

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

Abstract

We consider a "length-biased" shift-dependent information measure, related to the differential entropy in which higher weight is assigned to large values of observed random variables. This allows us to introduce the notions of "weighted residual entropy" and "weighted past entropy", that are suitable to describe dynamic information of random lifetimes, in analogy with the entropies of residual and past lifetimes introduced in [9] and [6], respectively. The obtained results include their behaviors under monotonic transformations.

AMS Classification: 62N05, 62B10

Key words and phrases: Life distributions; entropy; residual lifetime, past lifetime.

1 Introduction

It is well known that knowledge and use of schemes for information coding and transmission play a relevant role in understanding and modeling certain aspects of biological systems features, such as neuronal activity. Since the pioneering contributions by Shannon [20] and Wiener [22] numerous efforts have been devoted to enrich and extend the underlying information theory. Various measures of uncertainty introduced in the past have been recently invoked in order to deal with information in the context of theoretical neurobiology (see, for instance, Johnson and Glantz [17]). In addition, recent articles have thoroughly explored the use of information measures for absolutely continuous non-negative random variables, that appear to be suitable to describe random lifetimes (see [14] and [5], for instance). Here, wide use is made of Shannon entropy, that is also applied to residual and past lifetimes (cf., for instance, [9] and [6]). However, use of this type of entropy has the drawback of being position-free. In other terms, such an information measure does not take into account the values of the random variable but only its probability density. As a consequence, a random variable X possesses the same Shannon entropy as $X+b$, for any $b \in \mathbb{R}$.

To come up with a mathematical tool whose properties are similar to those of Shannon entropy, however without being position-free, we introduce the notions of "weighted residual entropy" and "weighted past entropy". They are finalized to describe dynamic information of random lifetimes. In Section 2 we provide some basic notions on weighted entropy,

*Paper appeared in Sci. Math. Jpn. 64 (2006), no. 2, 255–266.

complemented by some examples. Section 3 is devoted to the presentation of properties and of results on weighted residual and past entropies, whose behaviors under monotonic transformations are studied in Section 4.

Throughout this paper, the terms “increasing” and “decreasing” are used in non-strict sense. Furthermore, we shall adopt the following notations:

X: an absolutely continuous non-negative honest random variable (representing for instance the random lifetime of a system or of a living organism, or the interspike intervals in a model of neuronal activity);

$f(x)$: the probability density function (pdf) of X;

$\mathcal{S} = (0, \nu)$, with $\nu \leq +\infty$: the support of $f(x)$;

$F(x) = \Pr(X \leq x)$: the cumulative distribution function of X;

$\overline{F}(x) = 1 - F(x)$: the survival function of X;

$\lambda(x) = f(x)/\overline{F}(x)$: the hazard function, or failure rate, of X;

$\tau(x) = f(x)/F(x)$: the reversed hazard rate function of X;

$[Z|B]$: any random variable whose distribution is identical to the conditional distribution of Z given B.

2 Weighted entropy

The differential entropy of X, or Shannon information measure, is defined by

$$H := -\operatorname{E}[\log f(X)] = -\int_0^{+\infty} f(x)\, \log f(x)\, \mathrm{d}x, \tag{1}$$

where here “log” means natural logarithm. The integrand function on the right-hand-side of (1) depends on x only via $f(x)$, thus making H shift-independent. Hence, H stays unchanged if, for instance, X is uniformly distributed in (a, b) or $(a+h, b+h)$, whatever $h \in \mathbb{R}$. However, in certain applied contexts, such as reliability or mathematical neurobiology, it is desirable to deal with shift-dependent information measures. Indeed, knowing that a device fails to operate, or a neuron to release spikes in a given time-interval, yields a relevantly different information from the case when such an event occurs in a different equally wide interval. In some cases we are thus led to resort to a shift-dependent information measure that, for instance, assigns different measures to such distributions.

In agreement with Belis and Guiaşu [3] and Guiaşu [16], in this paper we shall refer to the following notion of *weighted entropy*:

$$H^w := -\operatorname{E}[X \log f(X)] = -\int_0^{+\infty} x\, f(x)\, \log f(x)\, \mathrm{d}x, \tag{2}$$

or equivalently:

$$H^w = -\int_0^{+\infty} \mathrm{d}y \int_y^{+\infty} f(x)\, \log f(x) \mathrm{d}x.$$

Recalling Taneja [21] we point out that the occurrence of an event removes a double uncertainty: a qualitative uncertainty related to its probability of occurrence, and a quantitative uncertainty concerning its value or its usefulness. The factor x, in the integral on the right-hand-side of (2), may be viewed as a weight linearly emphasizing the occurrence of the event $\{X = x\}$. This yields a "length biased" shift-dependent information measure assigning greater importance to larger values of X. The use of weighted entropy (2) is also motivated by the need, arising in various communication and transmission problems, of expressing the "usefulness" of events by means of an information measure given by $H^w = \mathrm{E}[I(u(X), f(X))]$, where $I(u(x), f(x))$ satisfies the following properties (see Belis and Guiaşu [3]):

$$I(u(x), f(x)g(x)) = I(u(x), f(x)) + I(u(x), g(x)), \qquad I(\lambda u(x), f(x)) = \lambda I(u(x), f(x)).$$

The relevance of weighted entropies as a measure of the average amount of *valuable* or *useful* information provided by a source has also been emphasized by Longo [18].

The following are examples of pairs of distributions that possess same differential entropies but unequal weighted entropies.

Example 2.1 Let X and Y be random variables with densities

$$f_X(t) = \begin{cases} 2t & \text{if } 0 < t < 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(t) = \begin{cases} 2(1-t) & \text{if } 0 < t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Their differential entropies are identical (see Example 1.1 of [6], where a misprint has to be noticed):

$$H_X = H_Y = \frac{1}{2} - \log 2.$$

Hence, the expected uncertainties for f_X and f_Y on the predictability of the outcomes of X and Y are identical. Instead, $H_X^w < H_Y^w$, since

$$H_X^w = -\int_0^1 x\, 2x\, \log 2x \,\mathrm{d}x = \frac{2}{9} - \frac{2}{3}\log 2,$$

$$H_Y^w = -\int_0^1 2x\,(1-x)\log(1-x)\,\mathrm{d}x = \frac{1}{9}.$$

Hence, even though $H_X = H_Y$, the expected weighted uncertainty contained in f_X on the predictability of the outcome of X is smaller than that of f_Y on the predictability of the outcome of Y.

Example 2.2 Consider the piece-wise constant pdf

$$f_X(x) = \sum_{k=1}^{n} c_k\, \mathbf{1}_{\{k-1 \le x < k\}} \qquad \left(c_k \ge 0,\ \ k = 1, 2, \ldots, n; \quad \sum_{k=1}^{n} c_k = 1 \right).$$

Its differential entropy is $H_X = -\sum_{k=1}^{n} c_k \log c_k$ while its weighted entropy is

$$H_X^w = -\sum_{k=1}^{n} k\, c_k \log c_k - \frac{1}{2} H_X. \tag{3}$$

Note that any new density obtained by permutation of $c_1, c_2, \ldots, c_n$ has the same entropy H_X, whereas its weighted entropy is in general different from (3).

Hereafter we recall some properties of differential entropy (1):
(i) for $a, b > 0$ it is $H_{aX+b} = H_X + \log a$.
(ii) if X and Y are independent, then

$$H_{(X,Y)} = H_X + H_Y,$$

where

$$H_{(X,Y)} := -\,\mathrm{E}[\log f(X,Y)] = -\int_0^{+\infty}\int_0^{+\infty} f(x,y)\,\log f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

is the bidimensional version of (1).

Proposition 2.1, whose proof is omitted being straightforward, shows the corresponding properties of H^w similar to *(i)* and *(ii)*, but the essential differences emphasize the role of the mean value in the evaluation of the weighted entropy.

Proposition 2.1 *The following statements hold:*
(i) *for* $a, b > 0$ *it is* $H^w_{aX+b} = a[H^w_X + \mathrm{E}(X)\log a] + b(H_X + \log a)$.
(ii) *if* X *and* Y *are independent, then*

$$H^w_{(X,Y)} = \mathrm{E}(Y)\,H^w_X + \mathrm{E}(X)\,H^w_Y,$$

where $H^w_{(X,Y)} := -\,\mathrm{E}[X\,Y\,\log f(X,Y)]$.

Let us now evaluate the weighted entropy of some random variables.

Example 2.3 *(a)* X is exponentially distributed with parameter $\lambda > 0$. Then,

$$H^w = -\int_0^{+\infty} x\,\lambda e^{-\lambda x}\,\log\left(\lambda e^{-\lambda x}\right)\mathrm{d}x = \frac{2-\log\lambda}{\lambda}. \tag{4}$$

(b) X is uniformly distributed over $[a, b]$. Then

$$H^w = -\int_a^b x\,\frac{1}{b-a}\,\log\frac{1}{b-a}\,\mathrm{d}x = \frac{a+b}{2}\,\log(b-a). \tag{5}$$

It is interesting to note that in this case the weighted entropy can be expressed as the product

$$H^w = \mathrm{E}(X)\,H. \tag{6}$$

(c) X is Gamma distributed with parameters α and β:

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

There holds:

$$\begin{aligned} H^w &= -\int_0^{+\infty}\frac{x^\alpha e^{-\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}\,\log\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}\,\mathrm{d}x\\ &= \alpha\beta\,\log[\beta^\alpha\Gamma(\alpha)] - \alpha(\alpha-1)\beta\left\{\log\beta + \psi_0(\alpha+1)\right\} + \alpha(\alpha+1)\beta, \end{aligned}$$

where

$$\psi_0(x) := \frac{\mathrm{d}}{\mathrm{d}x}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)} \tag{7}$$

and where we have used the following identities, with $\alpha > -1$ and $\beta > 0$:

$$\int_0^{+\infty} x^\alpha e^{-\frac{x}{\beta}}\,\mathrm{d}x = \left(\frac{1}{\beta}\right)^{-1-\alpha} \Gamma(1+\alpha),$$

$$\int_0^{+\infty} x^\alpha e^{-\frac{x}{\beta}}\,\log x\,\mathrm{d}x = \left(\frac{1}{\beta}\right)^{-1-\alpha} \Gamma(1+\alpha)\left\{\log\beta + \psi_0(1+\alpha)\right\}.$$

(d) X is Beta distributed with parameters α and β:

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

with

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

Then,

$$\begin{aligned} H^w &= \frac{\log B(\alpha,\beta)}{B(\alpha,\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} - \frac{(\alpha-1)}{B(\alpha,\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\left\{\psi_0(\alpha+1) - \psi_0(\alpha+\beta+1)\right\} \\ &- \frac{(\beta-1)}{B(\alpha,\beta)}\frac{\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\left\{\psi_0(\beta)\Gamma(\alpha+1) - \alpha\Gamma(\alpha)\psi_0(\alpha+\beta+1)\right\}, \end{aligned}$$

where $\psi_0(x)$ is defined in (7).

Remark 2.1 We point out that different distributions may have identical weighted entropies. For instance, if X is uniformly distributed over $[0,1]$ and Y is exponentially distributed with mean e^{-2}, then from (4) and (5) we have

$$H_X^w = H_Y^w = 0.$$

Another example is the following: if X is uniformly distributed in $[0,2]$ and Y is exponentially distributed with parameter $\lambda = 1.93389$ (such that $\log\lambda + \lambda\log 2 = 2$), then from (4) and (5) one obtains:

$$H_X^w = H_Y^w = \log 2.$$

Remark 2.2 There exist random variables having negative arbitrarily large weighted entropy. For instance, if X is uniformly distributed over $[a,b]$, with $b > 0$, then from (5) we have

$$\lim_{a\to b^-} H^w = \lim_{a\to b^-} \frac{a+b}{2}\log(b-a) = -\infty.$$

Remark 2.3 Notice that in general H^w can be either larger or smaller than H. For instance, if X is uniformly distributed over $[a,b]$, from (6) it follows $H^w \geq H$ when $\mathrm{E}(X) \geq 1$, and $H^w \leq H$ when $\mathrm{E}(X) \leq 1$.

Remark 2.4 If the support of X is $[0,\nu]$, with ν finite, the following upper bound for the weighted entropy of X holds:

$$H^w \leq \mu\log\frac{\nu^2}{2\mu}, \qquad \text{where } \mu = \mathrm{E}(X) \in [0,\nu]. \tag{8}$$

This follows via the continuous version of Jensen's inequality. The maximum of $b(\mu) := \mu \log \frac{\nu^2}{2\mu}$ is attained at $\mu = \mu_M$, where

$$\mu_M = \begin{cases} \dfrac{\nu^2}{2e} & \text{if } \nu < 2e, \\ \nu & \text{if } \nu \geq 2e, \end{cases} \qquad \text{with} \quad b(\mu_M) = \begin{cases} \dfrac{\nu^2}{2e} & \text{if } \nu < 2e, \\ \nu \log \dfrac{\nu}{2} & \text{if } \nu \geq 2e. \end{cases}$$

If, in particular, X is uniformly distributed over $[0, \nu]$, then (8) holds with the equal sign since $H^w = \frac{\nu}{2} \log \nu$ and $\mu = \nu/2$.

3 Weighted residual and past entropies

The residual entropy at time t of a random lifetime X was introduced by Ebrahimi [9] and defined as:

$$\begin{aligned} H(t) &= -\int_t^{+\infty} \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)} \, \mathrm{d}x \\ &= \log \overline{F}(t) - \frac{1}{\overline{F}(t)} \int_t^{+\infty} f(x) \log f(x) \, \mathrm{d}x \\ &= 1 - \frac{1}{\overline{F}(t)} \int_t^{+\infty} f(x) \log \lambda(x) \, \mathrm{d}x, \end{aligned} \tag{9}$$

for all $t \in \mathcal{S}$. We note that $H(t)$ is the differential entropy of the residual lifetime of X at time t, i.e. of $[X \,|\, X > t]$. Various result on $H(t)$ have been the object of recent researches (see [1], [4], [10], [11], [12]). The past entropy at time t of X is defined by Di Crescenzo and Longobardi [6] as:

$$\overline{H}(t) = -\int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} \, \mathrm{d}x \,, \qquad t \in \mathcal{S}, \tag{10}$$

where $f(x)/F(t)$, $0 < x < t$, is the pdf of the past lifetime $[X \,|\, X \leq t]$. Hence, given for instance that an item has been found failing at time t, $\overline{H}(t)$ measures the uncertainty about its past life. We remark that $\overline{H}(t)$ can also be viewed as the entropy of the inactivity time $[t - X \,|\, X \leq t]$.

Various dynamic information functions have been recently introduced to measure discrepancies between residual lifetime distributions [13] and between past lifetime distributions [7], as well as to measure dependence between two residual lifetimes [8]. In order to introduce new shift-dependent dynamic information measures, we now make use of (2) to define two weighted entropies for residual lifetimes and past lifetimes that are the weighted version of entropies (9) and (10).

Definition 3.1 *For all $t \in \mathcal{S}$,*
(i) the weighted residual entropy at time t of a random lifetime X is the differential weighted entropy of $[X \,|\, X > t]$:

$$H^w(t) := -\int_t^{+\infty} x \, \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)} \, \mathrm{d}x; \tag{11}$$

(ii) the weighted past entropy at time t of a random lifetime X is the differential weighted entropy of $[X \,|\, X \leq t]$:

$$\overline{H}^w(t) := -\int_0^t x \, \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} \, \mathrm{d}x. \tag{12}$$

We notice that

$$\lim_{t\to+\infty} H^w(t) = \lim_{t\to 0^+} \overline{H}^w(t) = H^w.$$

In addition, due to (11), the weighted residual entropy can be rewritten as

$$H^w(t) = -\frac{1}{\overline{F}(t)} \int_t^{+\infty} x\, f(x) \log f(x)\, \mathrm{d}x + \frac{\log \overline{F}(t)}{\overline{F}(t)} \int_t^{+\infty} x\, f(x)\, \mathrm{d}x. \tag{13}$$

The second integral in (13) can be calculated by noting that

$$\int_t^{+\infty} x\, f(x)\, \mathrm{d}x = t\,\overline{F}(t) + \int_t^{+\infty} \overline{F}(y)\, \mathrm{d}y. \tag{14}$$

Furthermore, an alternative way of writing $H^w(t)$ is the following:

$$\begin{aligned} H^w(t) &= -\int_t^{+\infty} \mathrm{d}x \int_0^x \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)}\, \mathrm{d}y \\ &= -\int_0^t \mathrm{d}y \int_t^{+\infty} \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)}\, \mathrm{d}x - \int_t^{+\infty} \mathrm{d}y \int_y^{+\infty} \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)}{\overline{F}(t)}\, \mathrm{d}x. \end{aligned}$$

Recalling (9), we thus obtain:

$$H^w(t) = t\, H(t) + \int_t^{+\infty} H(y)\, \mathrm{d}y. \tag{15}$$

From (15) it easily follows

$$\frac{\mathrm{d}}{\mathrm{d}t} H^w(t) = t\, \frac{\mathrm{d}}{\mathrm{d}t} H(t). \tag{16}$$

In analogy with Theorem 1 of Belzunce *et al.* [5], the following characterization result holds.

Theorem 3.1 *If X has an absolutely continuous distribution function $F(t)$ and if $H(t)$ is increasing for all $t \in \mathcal{S}$, then $H^w(t)$ uniquely determines $\overline{F}(t)$.*

Proof. From (13) and (14) we have:

$$\int_t^{+\infty} x\, f(x) \log f(x)\, \mathrm{d}x = \log \overline{F}(t) \left[t\,\overline{F}(t) + \int_t^{+\infty} \overline{F}(y)\, \mathrm{d}y \right] - H^w(t)\, \overline{F}(t).$$

Differentiating both sides we obtain:

$$-t\, f(t)\, \log f(t) = -t\, f(t)\, \log \overline{F}(t) - t\, f(t) - \frac{f(t)}{\overline{F}(t)} \int_t^{+\infty} \overline{F}(y)\, \mathrm{d}y + f(t)\, H^w(t) - \overline{F}(t)\, \frac{\mathrm{d}}{\mathrm{d}t} H^w(t).$$

Hence, due to (16) and recalling the hazard function $\lambda(t) = f(t)/\overline{F}(t)$, it follows:

$$t\, \lambda(t) [1 - \log \lambda(t)] = \lambda(t)\, H^w(t) - \lambda(t) \int_t^{+\infty} \frac{\overline{F}(y)}{\overline{F}(t)}\, \mathrm{d}y - t\, \frac{\mathrm{d}}{\mathrm{d}t} H(t).$$

Then, for any fixed $t \in \mathcal{S}$, $\lambda(t)$ is a positive solution of equation $g(x) = 0$, where

$$g(x) := x \left[t\,(1 - \log x) - H^w(t) + \int_t^{+\infty} \frac{\overline{F}(y)}{\overline{F}(t)}\, \mathrm{d}y \right] + t\, \frac{\mathrm{d}}{\mathrm{d}t} H(t). \tag{17}$$

Note that $\lim_{x\to+\infty} g(x) = -\infty$ and $g(0) = t\,\frac{\mathrm{d}}{\mathrm{d}t}H(t) \geq 0$. Furthermore, from (17) we have:

$$\frac{\mathrm{d}}{\mathrm{d}x}g(x) = -t\,\log x - H^w(t) + \int_t^{+\infty} \frac{\overline{F}(y)}{\overline{F}(t)}\,\mathrm{d}y,$$

so that $\frac{\mathrm{d}}{\mathrm{d}x}g(x) = 0$ if and only if

$$x = \exp\left\{-\frac{1}{t}\left[H^w(t) - \int_t^{+\infty} \frac{\overline{F}(y)}{\overline{F}(t)}\,\mathrm{d}y\right]\right\}.$$

Therefore, $g(x) = 0$ has a unique positive solution so that $\lambda(t)$, and hence $\overline{F}(t)$, is uniquely determined by $H^w(t)$ under assumption $\frac{\mathrm{d}}{\mathrm{d}t}H(t) \geq 0$. This concludes the proof. ■

Remark 3.1 We note that, by virtue of (16), if $H(t)$ is increasing [decreasing], then also $H^w(t)$ is increasing [decreasing].

In order to attain a decomposition of the weighted entropy, similar to that given in Proposition 2.1 of Di Crescenzo and Longobardi [6], for a random lifetime X possessing finite mean $\mathrm{E}(X)$ we recall that the *length-biased distribution function* and the *length-biased survival function* are defined respectively as

$$(18)\qquad F^*(t) = \frac{1}{\mathrm{E}(X)}\int_0^t x\,f(x)\,\mathrm{d}x, \qquad \overline{F}^*(t) = \frac{1}{\mathrm{E}(X)}\int_t^{+\infty} x\,f(x)\,\mathrm{d}x, \qquad t \in \mathcal{S}.$$

These functions characterize weighted distributions that arise in sampling procedures where the sampling probabilities are proportional to sample values. (See Section 3 of Belzunce *et al.* [5] for some results on uncertainty in length-biased distributions. Moreover, see Navarro *et al.* [19], Bartoszewicz and Skolimowska [2] and references therein for characterizations involving weighted distributions).

Theorem 3.2 *For a random lifetime X having finite mean $\mathrm{E}(X)$, for all $t \in \mathcal{S}$ the weighted entropy can be expressed as follows:*

$$H^w = \mathrm{E}(X)\left\{-F^*(t)\,\log F(t) - \overline{F}^*(t)\,\log \overline{F}(t)\right\} + F(t)\,\overline{H}^w(t) + \overline{F}(t)\,H^w(t).$$

Proof. Recalling Eqs. (2), (11) and (12) we have:

$$\begin{aligned} H^w &= -F(t)\int_0^t x\,\frac{f(x)}{F(t)}\log f(x)\,\mathrm{d}x - \overline{F}(t)\int_t^{+\infty} x\,\frac{f(x)}{\overline{F}(t)}\log f(x)\,\mathrm{d}x \\ &= -\log F(t)\int_0^t x\,f(x)\,\mathrm{d}x - \log \overline{F}(t)\int_t^{+\infty} x\,f(x)\,\mathrm{d}x + F(t)\,\overline{H}^w(t) + \overline{F}(t)\,H^w(t). \end{aligned}$$

The proof then follows from (18). ■

In order to provide a lower bound for the weighted residual entropy of a random lifetime X, let us introduce the following conditional mean value:

$$(19)\ \delta(t) := \mathrm{E}(X \mid X > t) = \frac{1}{\overline{F}(t)}\int_t^{+\infty} x\,f(x)\,\mathrm{d}x = t + \frac{1}{\overline{F}(t)}\int_t^{+\infty} \overline{F}(x)\,\mathrm{d}x, \qquad t \in \mathcal{S}.$$

Theorem 3.3 *If the hazard function $\lambda(t)$ is decreasing in $t \in \mathcal{S}$, then*

$$H^w(t) \geq -\delta(t)\, \log \lambda(t), \qquad t \in \mathcal{S}. \tag{20}$$

Proof. From (11) we have:

$$H^w(t) = -\frac{1}{\overline{F}(t)} \int_t^{+\infty} x\, f(x)\, \log \lambda(x)\, \mathrm{d}x - \frac{1}{\overline{F}(t)} \int_t^{+\infty} x\, f(x)\, \log \frac{\overline{F}(x)}{\overline{F}(t)}\, \mathrm{d}x, \qquad t \in \mathcal{S}.$$

Since $\log \frac{\overline{F}(x)}{\overline{F}(t)} \leq 0$ for $x \geq t$ and, by assumption, $\log \lambda(x) \leq \log \lambda(t)$, there holds:

$$\begin{aligned} H^w(t) &\geq -\frac{1}{\overline{F}(t)} \int_t^{+\infty} x\, f(x)\, \log \lambda(x)\, \mathrm{d}x \\ &\geq -\frac{\log \lambda(t)}{\overline{F}(t)} \int_t^{+\infty} x\, f(x)\, \mathrm{d}x. \end{aligned}$$

The proof then follows by recalling (19). ■

In the following example we consider the case of constant hazard function.

Example 3.1 For an exponential distribution with parameter $\lambda > 0$, the weighted residual entropy is given by

$$H^w(t) = -\int_t^{+\infty} x\, \frac{\lambda e^{-\lambda x}}{e^{-\lambda t}}\, \log \frac{\lambda e^{-\lambda x}}{e^{-\lambda t}}\, \mathrm{d}x = t + \frac{2}{\lambda} - \left(t + \frac{1}{\lambda}\right) \log \lambda, \qquad t \geq 0. \tag{21}$$

Recalling that for an exponential r.v. $\lambda(t) = \lambda$ and $\delta(t) = t + \frac{1}{\lambda}$, it is easily seen that (20) is fulfilled.

Let us now discuss some properties of the weighted past entropy. From (12) we have:

$$\overline{H}^w(t) = -\frac{1}{F(t)} \int_0^t x\, f(x) \log f(x)\, \mathrm{d}x + \frac{\log F(t)}{F(t)} \int_0^t x\, f(x)\, \mathrm{d}x, \qquad t \in \mathcal{S},$$

where, similarly to (14), it is

$$\int_0^t x\, f(x)\, \mathrm{d}x = t\, F(t) - \int_0^t F(y)\, \mathrm{d}y. \tag{22}$$

Alternatively,

$$\begin{aligned} \overline{H}^w(t) &= -\int_0^t \mathrm{d}x \int_0^x \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)}\, \mathrm{d}y \\ &= t\, \overline{H}(t) - \int_0^t \overline{H}(y)\, \mathrm{d}y, \end{aligned} \tag{23}$$

where $\overline{H}(t)$ is the past entropy given in (10). We note that Eq. (23) gives:

$$\frac{\mathrm{d}}{\mathrm{d}t} \overline{H}^w(t) = t\, \frac{\mathrm{d}}{\mathrm{d}t} \overline{H}(t), \qquad t \in \mathcal{S}.$$

Example 3.2 The weighted past entropy of an exponentially distributed random variable with parameter $\lambda > 0$ is instead given by

$$\overline{H}^w(t) = \frac{1}{1-e^{-\lambda t}} \left[\frac{2}{\lambda} - \frac{2}{\lambda}e^{-\lambda t} - 2t\, e^{-\lambda t} - \lambda t^2\, e^{-\lambda t} + \left(\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda t} - t\, e^{-\lambda t}\right) \log \frac{1-e^{-\lambda t}}{\lambda}\right]$$

for $t > 0$.

In order to obtain upper bounds for the weighted past entropy let us now recall the definition of mean past lifetime:

$$\mu(t) := \mathrm{E}(X|X \le t) = \int_0^t x\, \frac{f(x)}{F(t)}\, \mathrm{d}x = t - \frac{1}{F(t)} \int_0^t F(y)\, \mathrm{d}y. \tag{24}$$

We incidentally note that this is related to the Bonferroni Curve by $\mu(t) = \mathrm{E}(X)\, B_F[F(t)]$ (see Giorgi and Crescenzi [15]).

Theorem 3.4 *(i) For all $t \in \mathcal{S}$, it is*

$$\overline{H}^w(t) \le \mu(t) \log \frac{t^2}{2\mu(t)}. \tag{25}$$

(ii) If $\tau(t)$ is decreasing in $t \in \mathcal{S}$, then

$$\overline{H}^w(t) \le \int_0^t x\, \tau(x)\, \mathrm{d}x - \mu(t)\, [1 + \log \tau(t)]. \tag{26}$$

Proof. Eq. (25) is an immediate consequence of (8). Furthermore, from (12) we have:

$$\overline{H}^w(t) = -\frac{1}{F(t)} \int_0^t x\, f(x)\, \log \tau(x)\, \mathrm{d}x - \frac{1}{F(t)} \int_0^t x\, f(x)\, \log \frac{F(x)}{F(t)}\, \mathrm{d}x.$$

Since $\tau(t)$ is decreasing in $t \in \mathcal{S}$, we have $\log \tau(x) \ge \log \tau(t)$ for $0 < x < t$. Moreover, recalling that $\log x < x - 1$ for $x > 0$, we obtain:

$$\overline{H}^w(t) \le -\frac{\log \tau(t)}{F(t)} \int_0^t x\, f(x)\, \mathrm{d}x + \frac{1}{F(t)} \int_0^t x\, f(x) \left[\frac{F(t)}{F(x)} - 1\right] \mathrm{d}x.$$

By (22) and (24) we finally come to (26). ■

Remark 3.2 If X is uniformly distributed on $(0, \nu)$, the weighted past entropy is:

$$\overline{H}^w(t) = \frac{t}{2} \log t, \qquad 0 < t < \nu. \tag{27}$$

Hence, since $\mu(t) = t/2$, Eq. (25) is satisfied with the equality sign for all $t \in (0, \nu)$.

We shall now introduce two new classes of distributions based on monotonicity properties of the weighted entropies.

Definition 3.2 *A random lifetime X will be said to have*
(i) decreasing [increasing] weighted uncertainty residual life (DWURL) [IWURL] if and only if $H^w(t)$ is decreasing [increasing] in $t \in \mathcal{S}$;
(ii) decreasing [increasing] weighted uncertainty past life (DWUPL) [IWUPL] if and only if $\overline{H}^w(t)$ is decreasing [increasing] in $t \in \mathcal{S}$.

Let X be a random variable uniformly distributed on $(0, \nu)$. Since $H^w(t) = \frac{t+\nu}{2} \log(b-t)$, $0 < t < \nu$, X is DWURL if and only if $0 < \nu \le e$, and it can never be IWURL. Moreover, from (27) X is DWUPL if and only if $0 < \nu \le \frac{1}{e}$, and it can never be IWUPL. Finally, by virtue of (21), an exponential distribution with parameter λ is DWURL (IWURL) if and only if $\lambda \ge e$ $(0 < \lambda \le e)$.

4 Monotonic transformations

In this section we study the weighted residual entropy and the weighted past entropy under monotonic transformations. Similarly to Proposition 2.4 of Di Crescenzo and Longobardi [6], we have:

Theorem 4.1 *Let $Y = \phi(X)$, with ϕ strictly monotonic, continuous and differentiable, with derivative ϕ'. Then, for all $t \in \mathcal{S}$*

$$H_Y^w(t) = \begin{cases} H^{w,\phi}\left(\phi^{-1}(t)\right) + \mathrm{E}\left\{\phi(X)\log\phi'(X)\,|\,X > \phi^{-1}(t)\right\}, & \phi \text{ strictly increasing} \\ \overline{H}^{w,\phi}\left(\phi^{-1}(t)\right) + \mathrm{E}\left\{\phi(X)\log[-\phi'(X)]\,|\,X \le \phi^{-1}(t)\right\}, & \phi \text{ strictly decreasing} \end{cases} \tag{28}$$

and

$$\overline{H}_Y^w(t) = \begin{cases} \overline{H}^{w,\phi}\left(\phi^{-1}(t)\right) + \mathrm{E}\left\{\phi(X)\log\phi'(X)\,|\,X \le \phi^{-1}(t)\right\}, & \phi \text{ strictly increasing} \\ H^{w,\phi}\left(\phi^{-1}(t)\right) + \mathrm{E}\left\{\phi(X)\log[-\phi'(X)]\,|\,X > \phi^{-1}(t)\right\}, & \phi \text{ strictly decreasing,} \end{cases} \tag{29}$$

where

$$H^{w,\phi}(t) = -\frac{1}{\overline{F}_X(t)}\int_t^{+\infty}\phi(x)\,f_X(x)\log\frac{f_X(x)}{\overline{F}_X(t)}\,\mathrm{d}x,$$

$$\overline{H}^{w,\phi}(t) = -\frac{1}{F_X(t)}\int_t^{+\infty}\phi(x)\,f_X(x)\log\frac{f_X(x)}{F_X(t)}\,\mathrm{d}x.$$

Proof. From (11) we have

$$H_Y^w(t) = -\int_t^{+\infty} y\,\frac{f_X(\phi^{-1}(y))}{P(Y>t)}\left|\frac{\mathrm{d}}{\mathrm{d}y}\phi^{-1}(y)\right|\log\left\{\frac{f_X(\phi^{-1}(y))}{P(Y>t)}\left|\frac{\mathrm{d}}{\mathrm{d}y}\phi^{-1}(y)\right|\right\}\mathrm{d}y.$$

Let ϕ be strictly increasing. By setting $y = \phi(x)$, we obtain

$$H_Y^w(t) = -\int_{\phi^{-1}(t)}^{\phi^{-1}(+\infty)}\phi(x)\,\frac{f_X(x)}{\overline{F}_X(\phi^{-1}(t))}\log\left\{\frac{f_X(x)}{\overline{F}_X(\phi^{-1}(t))}\left|\frac{\mathrm{d}}{\mathrm{d}x}\phi(x)\right|^{-1}\right\}\mathrm{d}x,$$

or

$$H_Y^w(t) = -\int_{\phi^{-1}(t)}^{+\infty}\frac{\phi(x)f_X(x)}{\overline{F}_X(\phi^{-1}(t))}\log\frac{f_X(x)}{\overline{F}_X(\phi^{-1}(t))}\,\mathrm{d}x + \int_{\phi^{-1}(t)}^{+\infty}\frac{\phi(x)f_X(x)}{\overline{F}_X(\phi^{-1}(t))}\log\left|\frac{\mathrm{d}}{\mathrm{d}x}\phi(x)\right|\mathrm{d}x,$$

giving the first of (28). If ϕ is strictly decreasing we similarly obtain:

$$H_Y^w(t) = -\int_{\phi^{-1}(+\infty)}^{\phi^{-1}(t)}\frac{\phi(x)\,f_X(x)}{F_X(\phi^{-1}(t))}\log\frac{f_X(x)}{F_X(\phi^{-1}(t))}\,\mathrm{d}x + \int_0^{\phi^{-1}(t)}\frac{\phi(x)\,f_X(x)}{F_X(\phi^{-1}(t))}\log\left|\frac{\mathrm{d}}{\mathrm{d}x}\phi(x)\right|\mathrm{d}x,$$

i.e., the second of (28). The proof of (29) is analogous. ■

According to Remark 2.3 of [6] we note that when $Y = \phi(X)$ is distributed as X (like as for certain Pareto and beta-type distributions), Theorem 4.1 yields useful identities that allow to express $H_X^w(t)$ and $\overline{H}_X^w(t)$ in terms of $H^{w,\phi}\left(\phi^{-1}(t)\right)$ or $\overline{H}^{w,\phi}\left(\phi^{-1}(t)\right)$, depending on the type of monotonicity of ϕ.

Remark 4.1 Due to Theorem 4.1, for all $a > 0$ and $t > 0$ there holds:

$$H^w_{aX}(t) = a\, H^w\left(\frac{t}{a}\right) + \delta\left(\frac{t}{a}\right) a\, \log a,$$

$$\overline{H}^w_{aX}(t) = a\, \overline{H}^w\left(\frac{t}{a}\right) + \mu\left(\frac{t}{a}\right) a\, \log a.$$

Furthermore, for all $b > 0$ and $t > b$ one has:

$$H^w_{X+b}(t) = H^w(t-b) + b\, H(t-b),$$

$$\overline{H}^w_{X+b}(t) = \overline{H}^w(t-b) + b\, \overline{H}(t-b).$$

Acknowledgments

This work has been performed under partial support by MIUR (PRIN 2005), by G.N.C.S.-INdAM and by Campania Region.

References

[1] Asadi, M. and Ebrahimi, N. (2000), Residual entropy and its characterizations in terms of hazard function and mean residual life function. *Stat. Prob. Lett.* **49**, 263–269.

[2] Bartoszewicz, J. and Skolimowska, M. (2006), Preservation of classes of life distributions and stochastic orders under weighting. *Stat. Prob. Lett.* **76**, 587–596.

[3] Belis, M. and Guiaşu, S. (1968), A quantitative-qualitative measure of information in cybernetic systems. *IEEE Trans. Inf. Th.* **IT-4**, 593–594.

[4] Belzunce, F., Guillamon, A., Navarro, J. and Ruiz, J.M. (2001), Kernel estimation of residual entropy. *Commun. Statist. – Theory Meth.* **30**, 1243–1255.

[5] Belzunce, F., Navarro, J., Ruiz, J.M. and del Aguila, Y. (2004), Some results on residual entropy function. *Metrika* **59**, 147–161.

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

[7] Di Crescenzo, A. and Longobardi, M. (2004), A measure of discrimination between past lifetime distributions. *Stat. Prob. Lett.* **67**, 173–182.

[8] Di Crescenzo, A., Longobardi M. and Nastro, A. (2004), On the mutual information between residual lifetime distributions. In: *Cybernetics and Systems 2004*, Vol. 1 (R. Trappl, Ed.), p. 142–145. Austrian Society for Cybernetic Studies, Vienna. ISBN 3-85206-169-5.

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

[10] Ebrahimi, N. (1997), Testing whether lifetime distribution is decreasing uncertainty. *J. Stat. Plann. Infer.* **64**, 9–19.

[11] Ebrahimi, N. (2000), The maximum entropy method for lifetime distributions. *Sankhyä, Ser. A* **62**, 236–243.

[12] Ebrahimi, N. and Kirmani, S.N.U.A. (1996), Some results on ordering of survival functions through uncertainty. *Stat. Prob. Lett.* **29**, 167–176.

[13] Ebrahimi, N. and Kirmani, S.N.U.A. (1996), A measure of discrimination between two residual lifetime distributions and its applications. *Ann. Inst. Stat. Math.* **48**, 257–265.

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

[15] Giorgi, G.M. and Crescenzi, M. (2001), A look at the Bonferroni inequality measure in a reliability framework. *Statistica* **LXI** (4), 571–583.

[16] Guiaşu, S. (1986), Grouping data by using the weighted entropy. *J. Statist. Plann. Infer* **15**, 63–69.

[17] Johnson, D.H. and Glantz, R.M. (2004), When does interval coding occur? *Neurocomputing* **59-60**, 13–18.

[18] Longo, G. (1976), A noiseless coding theorem for sources having utilities. *SIAM J. Appl. Math.* **30**, 739–748.

[19] Navarro, J., del Aguila, Y. and Ruiz, J.M. (2001), Characterizations through reliability measures from weighted distributions. *Stat. Papers* **42**, 395–402.

[20] Shannon, C.E. (1948), A mathematical theory of communication. *Bell System Tech. J.* **27**, 279–423.

[21] Taneja, I.J. (1990), On generalized entropy with applications. In: *Lectures in Applied Mathematics and Informatics* (L.M. Ricciardi, Ed.), p. 107–169. Manchester Univ. Press, Manchester.

[22] Wiener, N. (1948, 2nd Ed. 1961), *Cybernetics*, The MIT Press and Wiley, New York.