



EXTROPY: CHARACTERIZATIONS AND DYNAMIC VERSIONS

ABDOLSAEED TOOMAJ,* *Gonbad Kavous University*

MAJID HASHEMPOUR,** *University of Hormozgan*

NARAYANASWAMY BALAKRISHNAN,*** *McMaster University*

Abstract

Several information measures have been proposed and studied in the literature. One such measure is extropy, a complementary dual function of entropy. Its meaning and related aging notions have not yet been studied in great detail. In this paper, we first illustrate that extropy information ranks the uniformity of a wide array of absolutely continuous families. We then discuss several theoretical merits of extropy. We also provide a closed-form expression of it for finite mixture distributions. Finally, the dynamic versions of extropy are also discussed, specifically the residual extropy and past extropy measures.

Keywords: Shannon's entropy; extropy; Kullback–Leibler information; relative extropy; dynamic extropy; mixture distribution

2020 Mathematics Subject Classification: Primary 62N05

Secondary 62F10

1. Introduction

The entropy measure of a probability distribution, as introduced in the pioneering work of Shannon [25], has found key applications in numerous fields. In information theory, it is used as a measure of uncertainty associated with a random phenomenon. If X is an unknown but observable quantity with a finite discrete range of possible values $\{x_1, \dots, x_n\}$ with an associated probability mass function vector $\mathbf{p}_n = (p_1, \dots, p_n)$, where $p_i = P(X = x_i)$, $i = 1, \dots, n$, such that $\sum_{i=1}^n p_i = 1$, the Shannon entropy measure, denoted by $H(X) = H(\mathbf{p}_n)$, equals $-\sum_{i=1}^n p_i \log p_i$, where $\log(\cdot)$ denotes the natural logarithm. As mentioned in [17], this measure has a complementary dual, termed *extropy*, which is also a very useful notion. The extropy measure, denoted by $J(X) = J(\mathbf{p}_n)$, is defined as $-\sum_{i=1}^n (1 - p_i) \log(1 - p_i)$ in the discrete case. Just as for entropy, extropy can also be interpreted as a measure of the amount of uncertainty associated with the distribution of X . It can be seen that the entropy and extropy of a binary distribution are identical, but, for $n \geq 3$, the entropy is greater than the extropy; see, e.g., [17]. As with entropy, the maximum extropy distribution is also the uniform distribution, and both measures are invariant with respect to permutations of their mass functions, while they behave quite differently in their assessments of the refinement of a distribution.

Received 21 April 2022; revision received 9 January 2023.

* Postal address: Faculty of Basic Sciences and Engineering, Department of Mathematics and Statistics, Gonbad Kavous University, Gonbad Kavous, Iran. Emails: ab.toomaj@gonbad.ac.ir, ab.toomaj@gmail.com

** Postal address: Department of Statistics, Faculty of Basic Sciences, University of Hormozgan, Bandar Abbas, Iran. Email: ma.hashempour@hormozgan.ac.ir

*** Postal address: Department of Mathematics and Statistics, McMaster University, Hamilton, ON L8S 4K1, Canada. Email: bala@mcmaster.ca

© The Author(s), 2023. Published by Cambridge University Press on behalf of Applied Probability Trust.

If X is an absolutely continuous non-negative random variable having probability density function (PDF) $f(x)$ with support S , the Shannon differential entropy is then defined as

$$H(X) = H(f) = - \int_S f(x) \log f(x) dx, \tag{1.1}$$

provided the integral is finite with mean value $\mu = \mathbb{E}[X]$ and variance $\sigma^2(X) = \text{Var}(X) = \mathbb{E}[X - \mu]^2$. It is known that $\sigma^2(X) < \infty$ implies that $\mathbb{E}(|X|) < \infty$, so $H(X) < \infty$, but the converse is not true. Also, $H(X)$ can be finite when $\mathbb{E}(|X|)$ is not (e.g. the Cauchy distribution). It is important to mention that information-theoretic methodologies are useful in many problems and so have received considerable attention in the literature; see, for example, [6, 8, 9, 14, 26, 29] and the references therein.

Shannon motivated the measure in (1.1) by arguing that refining the categories for a discrete quantity X , with diminishing probabilities in each, yields this analogous definition in the limit. This motivated [17] to introduce the dual notion of extropy for a continuous random variable. As pointed out in [17], for large n the extropy measure can be approximated by

$$J(\mathbf{p}_n) \approx 1 - \frac{\Delta x}{2} \sum_{i=1}^n f^2(x_i) \Delta x,$$

where $\Delta x = (x_n - x_1)/(n - 1)$ for any specific n . Thus, the measure of differential extropy for a continuous PDF can be well defined, via the limit of $J(\mathbf{p}_n)$ as n increases, in the following form:

$$J(X) = J(f) = \lim_{\Delta x \rightarrow 0} \frac{[J(\mathbf{p}_n) - 1]}{\Delta x} = -\frac{1}{2} \int_S f^2(x) dx. \tag{1.2}$$

Another useful expression of it can be given in terms of the hazard rate and reversed hazard rate functions. For an absolutely continuous non-negative random variable X with survival function $\bar{F}(x) = 1 - F(x)$, and hazard rate and reversed hazard rate functions $\lambda(x) = f(x)/\bar{F}(x)$ and $\tau(x) = f(x)/F(x)$, respectively, the extropy can be expressed as

$$J(f) = -\frac{1}{4} \mathbb{E}_{12}[\lambda(X)] = -\frac{1}{4} \mathbb{E}_{22}[\tau(X)], \tag{1.3}$$

where \mathbb{E}_{12} and \mathbb{E}_{22} denote expectations with respect to the PDFs

$$f_{12}(x) = 2f(x)\bar{F}(x), \quad x > 0, \tag{1.4}$$

$$f_{22}(x) = 2f(x)F(x), \quad x > 0, \tag{1.5}$$

respectively. The densities in (1.4) and (1.5) are in fact the densities of minima and maxima of two independent and identically distributed (i.i.d.) random variables [1]. Several properties and statistical applications of the extropy in (1.2) were discussed in [17]. Moreover, [28] studied relations between extropy and variational distance, and determined the distribution that attains the minimum or maximum extropy among all distributions within a given variation distance from any given probability distribution. Qiu and Jia in [22] explored the residual extropy properties of order statistics and record values, while [21] proposed two estimators of extropy and used them to develop goodness-of-fit tests for the standard uniform distribution. In the present work, we carry out a detailed study of extropy and its various properties, including its dynamic versions.

The rest of this paper is organized as follows. In Section 2 we describe some preliminary details on information divergence, and equilibrium and weighted distribution, and also mention some well-known distributional orders that are most pertinent for all the results developed here. In Section 3, some differences and similarities between entropy, extropy, and variance are pointed out, and they are then applied to a wide family of distributions. In Section 4, characterizations based on maximum extropy and minimum relative extropy criteria for probability models based on moment constraints are presented. In Section 5, monotonicity properties of the dynamic residual and past entropies are established. Finally, some concluding remarks are provided in Section 6.

2. Preliminaries

We briefly describe here information divergence, and equilibrium and weighted distributions, and then mention some well-known distributional orders that are essential for all the subsequent developments. Throughout, X and Y will denote non-negative random variables with absolutely continuous cumulative distribution functions (CDFs) $F(x)$ and $G(x)$, survival functions $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(x) = 1 - G(x)$, and PDFs $f(x)$ and $g(x)$, respectively. When the considered random variables are not non-negative, it will be mentioned explicitly.

2.1. Information divergence

The Kullback–Leibler (KL) discrimination information between two densities f and g is defined as

$$K(f : g) = d(f||g) = \int_{\mathcal{S}} \log \frac{f(x)}{g(x)} dF(x) \geq 0, \tag{2.1}$$

provided the integral is finite, and it requires f to be absolutely continuous with respect to g . This condition is necessary, but not sufficient, for the finiteness of (2.1). The equality holds in (2.1) if and only if $f(x) = g(x)$ almost everywhere. The KL discrimination information between any distribution with a PDF f and the uniform PDF f^* on a common bounded support \mathcal{S} is given by [8]

$$d(f||f^*) = H(f^*) - H(f) \geq 0, \tag{2.2}$$

where $H(f^*) = \log \|\mathcal{S}\|$, with $\|\mathcal{S}\|$ denoting the size of the support $\|\mathcal{S}\|$. We recall that X is smaller than Y in the entropy order (denoted by $X \leq_e Y$) if and only if $H(X) \leq H(Y)$. From (2.2), for two distributions with PDFs f and g on a common bounded $\|\mathcal{S}\|$, $X \leq_e Y$ if and only if $d(f||f^*) \geq d(g||f^*)$. The case of unbounded $\|\mathcal{S}\|$ can be interpreted similarly in terms of (2.2).

A natural problem of interest is to determine a distribution, within a class of probability distributions $\Omega = \{f\}$, that minimizes $d(f||g)$ for a given g , referred to as the reference distribution. The classical minimum discrimination information (MDI) formulation is defined in terms of moment constraints: $\Omega = \Omega_\theta$, defined by all distributions with $\mathbb{E}_f[T_j(X)] = \theta_j < \infty$, $j = 1, \dots, J$, where θ_j is a constant and $T_j(x)$ is a measurable statistic. For this problem, the MDI theorem [16] gives the form of the MDI PDF $f^*(x) \in \Omega_\theta$, a formula for the MDI function $d(f||g)$, and a formula for the recovery of moment constraint parameters. With a single moment constraint, for example, the MDI theorem concerning $\min_f d(f||g)$ subject to $\mathbb{E}_f[T(X)] = \theta$, $\int_0^\infty f(x) dx = 1$, gives the solution as $f^*(x) = g(x)C_\lambda e^{\tau T(x)}$, where $\tau > 0$ is the Lagrange multiplier and C_λ is a normalizing constant. For further applications of the MDI model, see [4, 5] and the references therein.

2.2. Equilibrium distribution

Recall that the limiting distribution of the excess time (or the forward recurrence time) in a renewal process (or in shock models) results in the so-called equilibrium distribution. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of independent non-negative random variables representing inter-arrival times between shocks. Further, suppose these random variables have an identical CDF $F(t)$, with finite mean μ . Also, let X_1 have a possibly different CDF $F_1(t)$, with finite mean $\mu_1 = \mathbb{E}[X_1]$. Both distribution functions $F_1(t)$ and $F(t)$ are non-degenerate at $t = 0$, i.e. $F_1(0) = F(0) = 0$. For $S_n = \sum_{i=1}^n X_i$, $n \in \mathbb{N}$, with $S_0 \equiv 0$, let $N(t) = \max\{n : S_n \leq t\}$ represent the number of renewals during $(0, t]$. Let $\gamma(t)$ be the excess time in a stochastic process or residual lifetime at time t , i.e. $\gamma(t) = S_{N(t)+1} - t$. From the elementary renewal theorem, the distribution of the equilibrium random variable \tilde{X}_e is known to be

$$\tilde{F}_e(x) = \lim_{t \rightarrow \infty} \mathbb{P}[\gamma(t) \leq x] = \frac{1}{\mu} \int_0^x \bar{F}(u) du, \quad x > 0,$$

and the corresponding PDF is $\tilde{f}_e(x) = \bar{F}(x)/\mu$, $x > 0$ [19]. The equilibrium distribution is the asymptotic distribution of the time since the last renewal at time t and the waiting time until the next renewal.

Weighted distributions have found many applications; see, for example, [20] and the references therein. For a variable X with PDF f and a non-negative real function w , let

$$f^w(x) = \frac{w(x)f(x)}{\mathbb{E}[w(X)]}, \quad x > 0,$$

be the PDF of the associated weighted random variable X^w , provided $\mathbb{E}[w(X)]$ is positive and finite. Note that the equilibrium random variable X_e is a weighted random variable obtained from X with $w(x) = 1/\lambda(x)$, where λ is the failure rate function of X .

2.3. Stochastic orders

Aging notions and stochastic orders, as discussed in [24], have found several important uses in many disciplines. We mention below some key aging concepts and stochastic orders, which are most pertinent for the developments here. Throughout, the terms ‘increasing’ and ‘decreasing’ are used in a non-strict sense.

Let X have the hazard rate and reversed hazard rate functions $\lambda_X(x) = f(x)/\bar{F}(x)$ and $\tau_X(x) = f(x)/F(x)$, respectively. Similarly, let Y have the hazard rate and reversed hazard rate functions $\lambda_Y(x) = g(x)/\bar{G}(x)$ and $\tau_Y(x) = g(x)/G(x)$, respectively. Then, in the present work, we use the following notions: the decreasing reversed failure rate (DRFR) property; the increasing/decreasing failure rate (IFR/DFR) properties; the usual stochastic order (denoted by $X \leq_{st} Y$); hazard rate order (denoted by $X \leq_{hr} Y$); dispersive order (denoted by $X \leq_d Y$); convex order (denoted by $X \leq_{cx} Y$). For their informal definitions and properties, we refer the readers to [24]. In Table 1, we present the implications of these orders in terms of random variables X and Y .

3. Results on extropy

Differential entropy is a measure of the disparity of the PDF $f(x)$ from the uniform distribution. Indeed, it measures uncertainty in the sense of the utility of using $f(x)$ in place of the ultimate uncertainty of the uniform distribution [10]. Variance measures the average of

TABLE 1. Distributional orders and their implications.

Distributional order	Notation	Implications [condition]
Stochastic order	$X \leq_{st} Y$	$X \leq_e Y$ [if Y is DFR]
Dispersive order	$X \leq_d Y$	$X \leq_e Y$ and $X \leq_{st} Y$
Convex order	$X \leq_{cx} Y$	$X \leq_e Y$ [if g is log-concave]
Hazard rate order	$X \leq_{hr} Y$	$X \leq_{st} Y$ $X \leq_d Y$ [if X or Y is DFR].

distances of outcomes of a probability distribution from its mean. Because entropy is a complementary dual of entropy, it is also a measure of the disparity of the PDF $f(x)$ from the uniform distribution. Even though entropy, extropy, and variance are all measures of dispersion and uncertainty, the lack of a simple relationship between orderings of a distribution by the three measures arises from some substantial and subtle differences. For example, the differential entropy of random variable X takes values in $[-\infty, \infty]$, while extropy takes values in $[-\infty, 0)$. Moreover, $J(f) < H(f)$ due to the fact that $2x \log x < x^2$ for all $x > 0$.

In terms of mathematical properties, both entropy and extropy are non-negative in the discrete case. Moreover, in this case, $H(\mathbf{p})$ and $J(\mathbf{p})$ are invariant under one-to-one transformations of X . In the continuous case, neither the entropy nor the extropy is invariant under one-to-one transformations of X . Let $\phi(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ be a one-to-one function and $Y = \phi(X)$. It is known that $H(Y) = H(X) - \mathbb{E}[\log J_\phi(Y)]$ [8], where $J_\phi(Y) = |d\phi^{-1}(Y)/dY|$ is the Jacobian of the transformation. As $f_Y(y) = f_X(\phi^{-1}(y))|1/(\phi'(\phi^{-1}(y)))|$, we readily find that

$$\begin{aligned}
 J(Y) &= -\frac{1}{2} \int_0^\infty f_Y^2(y) dy = -\frac{1}{2} \int_0^\infty f_X^2(\phi^{-1}(y)) \left[\frac{1}{\phi'(\phi^{-1}(y))} \right]^2 dy \\
 &= -\frac{1}{2} \int_0^\infty \frac{f_X^2(u)}{\phi'(u)} du \\
 &= J(X) - \frac{1}{2} \mathbb{E} \left[\left(\frac{1}{\phi'(X)} - 1 \right) f(X) \right].
 \end{aligned}$$

However, there is no such direct relationship with the standard deviation. Furthermore, for any $a > 0$ and $b \in \mathbb{R}$,

$$H(aX + b) = H(X) + \log a,$$

$$\sigma(aX + b) = a\sigma(X),$$

$$J(aX + b) = \frac{1}{a}J(X),$$

which means that they are all position-free but scale-dependent. The following theorem extends the impact of scale on the extropy of a random variable to more general transformations. This result is similar to [7, Theorem 1] for the differential entropy and variance, and we therefore do not present its proof. First, we recall that X is smaller than Y in the extropy order (denoted by $X \leq_{ex} Y$) if and only if $J(X) \leq J(Y)$.

Theorem 3.1. *Let X be a random variable with PDF $f(x)$, and $Y = \phi(X)$, where $\phi: (0, \infty) \rightarrow (0, \infty)$ is a function with a continuous derivative $\phi'(x)$ in the support of X such that $\mathbb{E}(Y^2) < \infty$. If $|\phi'(x)| \geq 1$ for all x in the support of X , then $X \leq_{\text{ex}} Y$.*

It is known that the Shannon entropy of the sum of two independent random variables is larger than both their individual entropies. In a similar manner, the following theorem presents the corresponding result for extropy.

Theorem 3.2. *If X and Y are two absolutely continuous independent random variables, then $J(X + Y) \geq \max\{J(X), J(Y)\}$.*

Proof. Let X and Y be two absolutely continuous independent random variables with CDFs F and G , and PDFs f and g , respectively. Then, using the convolution formula and setting $Z = X + Y$, we immediately obtain

$$f_Z(z) = \int_{-\infty}^{\infty} g(y)f(z - y) \, dy = \mathbb{E}_Y[f(z - Y)], \quad z \in \mathbb{R}.$$

Now, applying Jensen’s inequality for the convex function x^2 to this result, we get $(\mathbb{E}_Y[f(z - Y)])^2 \leq \mathbb{E}_Y[f^2(z - Y)]$, $z \in \mathbb{R}$. Then, by integrating both sides of this inequality with respect to z from $-\infty$ to ∞ , we obtain

$$\begin{aligned} J(X + Y) &\geq -\frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f^2(z - y) \, dy \, dz \\ &= -\frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f^2(u) \, du \, dy \quad (\text{letting } u = z - y) \\ &= J(X). \end{aligned}$$

The proof is completed by using similar arguments for the random variable Y . □

We recall that the two-dimensional version of the Shannon differential entropy in (1.1) is $H(X, Y) = -\mathbb{E}[\log f(X, Y)]$. If X and Y are independent, then it is evident that $H(X, Y) = H(X) + H(Y)$. However, extropy has a distinctly different property in this regard. Indeed, defining the two-dimensional version of the differential extropy in (1.2) as $J(X, Y) = -\frac{1}{2}\mathbb{E}[f(X, Y)]$, if X and Y are independent, then $J(X, Y) = -2J(X)J(Y)$.

In analogy to (2.1), the relative extropy in a density $f(\cdot)$ relative to $g(\cdot)$ is defined as [17]

$$d^c(f||g) = \frac{1}{2} \int_{\mathcal{S}} [f(x) - g(x)]^2 \, dx \geq 0,$$

provided the integral is finite. The equality holds if and only if $f(x) = g(x)$ almost everywhere. The relative extropy can then be represented as

$$d^c(f||g) = 2J(f, g) - J(f) - J(g), \tag{3.1}$$

where $J(f, g) = -\frac{1}{2}\mathbb{E}[f(Y)] = -\frac{1}{2}\mathbb{E}[g(X)]$ is the inaccuracy measure of f with respect to g or vice versa. As pointed out in [17], the relative extropy between any distribution with a PDF f and the uniform PDF f^* on a common bounded support \mathcal{S} is given by $d^c(f||f^*) = J(f^*) - J(f) \geq 0$, where $J(f^*) = -1/(2\|\mathcal{S}\|)$. So, by this result, for two distributions with PDFs f and g on a common bounded $\|\mathcal{S}\|$, we have $X \leq_{\text{ex}} Y$ if and only if

$d^c(f||f^*) \geq d^c(g||f^*)$. The case of unbounded $\|\mathcal{S}\|$ can be interpreted in a similar manner. We now present some implications of the stochastic and convex orderings for two distributions by means of extropy.

Theorem 3.3. *Let X and Y be two non-negative random variables with PDFs $f(x)$ and $g(x)$, respectively. If $X \leq_{st} Y$ and Y is DFR, then $X \leq_{ex} Y$.*

Proof. Let Y be DFR with $X \leq_{st} Y$. Thus, we have

$$\int_0^\infty g^2(x) dx \leq \int_0^\infty f(x)g(x) dx \leq \left(\int_0^\infty g^2(x) dx \right)^{\frac{1}{2}} \left(\int_0^\infty f^2(x) dx \right)^{\frac{1}{2}}. \tag{3.2}$$

The first inequality in (3.2) is obtained by noting that $X \leq_{st} Y$ implies $\mathbb{E}_X[g(X)] > \mathbb{E}_Y[g(Y)]$ as g is a decreasing function because Y is DFR. The second inequality is obtained by using the Cauchy–Schwarz inequality. Making use of (1.2) and (3.2), we obtain the required result. \square

The following theorem gives implications of the convex order under some condition for the same ordering of the two models by extropy.

Theorem 3.4. *Under the conditions of Theorem 3.3, if $X \leq_{cx} Y$ and $g(x)$ is a concave function, then $X \leq_{ex} Y$.*

Proof. From the non-negativity of relative extropy in (3.1), we get

$$J(f) + J(g) \leq 2J(f, g). \tag{3.3}$$

Let $g(x)$ be a concave function, so $-g(x)$ is a convex function. By applying the definition of convex order, the assumption $X \leq_{cx} Y$ implies that

$$2J(f, g) = - \int_0^\infty f(x)g(x) dx \leq - \int_0^\infty g^2(x) dx = 2J(g). \tag{3.4}$$

Then, from (3.4) and (3.3), the desired result follows. \square

For convenience, we present in Table 2 the expressions for extropy, entropy, and standard deviation of some common distributions.

3.1. Extropy of finite mixture distributions

The entropy of mixture distributions has been studied by many authors, including [12, 23, 27]. Here, we derive a closed-form expression for the extropy of finite mixture distributions. Let $X_i, i = 1, \dots, n$, be a collection of n absolutely continuous independent random variables. Further, let $f_i(\cdot)$ be the PDF of X_i and $\mathbf{P} = (p_1, \dots, p_n)$ be the mixing probabilities. Then, the PDF of a finite mixture random variable X_p is given by

$$f_p(x) = \sum_{i=1}^n p_i f_i(x), \quad x \in \mathbb{R}, \tag{3.5}$$

where $\sum_{i=1}^n p_i = 1, p_i \geq 0$. Using the algebraic identity

$$\left(\sum_{i=1}^n a_i \right)^2 = \sum_{i=1}^n a_i^2 + 2 \sum_{i < j} a_i a_j,$$

TABLE 2. Extropy, entropy, and standard deviation of some common distributions.

Family and PDF	$J(X)$	$\sigma(X)$	$H(X)$
Normal(α, β)	$f(x) = 1/(\beta\sqrt{2\pi}) \exp\{-(1/(2\beta^2))(x - \alpha)^2\}$	$x \in \mathbb{R}, \beta > 0, \alpha \in \mathbb{R}$	
	$-\frac{1}{4\beta\sqrt{\pi}}$	β	$\log \beta + \frac{1}{2} \log(2\pi e)$
Gumbel(α, β)	$f(x) = (1/\beta) \exp\{-(x - \alpha)/\beta - \exp\{-(x - \alpha)/\beta\}\}$	$x \in \mathbb{R}, \beta > 0, \alpha \in \mathbb{R}$	
	$-\frac{1}{8\beta}$	$\frac{\pi\beta}{\sqrt{6}}$	$\log \beta + 1 + \gamma^\dagger$
Laplace(α, β)	$f(x) = (1/2\beta) \exp\{- x - \alpha /\beta\}$	$x \in \mathbb{R}, \beta > 0, \alpha \in \mathbb{R}$	
	$-\frac{1}{8\beta}$	$\sqrt{2}\beta$	$\log \beta + \log(2e)$
Uniform(α, β)	$f(x) = 1/(\beta - \alpha)$	$\alpha < x < \beta$	
	$-\frac{1}{2(\beta - \alpha)}$	$\frac{\beta - \alpha}{\sqrt{12}}$	$\log(\beta - \alpha)$
Gamma(α, β)	$f(x) = 1/(\beta\Gamma(\alpha))(x/\beta)^{\alpha-1} e^{-x/\beta}$	$x \geq 0, \alpha > 0, \beta > 0$	
	$-\frac{\Gamma(\alpha - \frac{1}{2})}{4\beta\sqrt{\pi}\Gamma(\alpha)}, \alpha > \frac{1}{2}$	$\beta\sqrt{\alpha}$	$\log \beta + \log \Gamma(\alpha) + (1 - \alpha)\psi(\alpha) + \alpha$
Weibull(α, β)	$f(x) = (\alpha/\beta)(x/\beta)^{\alpha-1} e^{-(x/\beta)^\alpha}$	$x \geq 0, \alpha > 0, \beta > 0$	
	$-\frac{\alpha 2^{1/\alpha} \Gamma(2 - (1/\alpha))}{8\beta}, \alpha > \frac{1}{2}$	$\beta\sqrt{\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)}$	$\log(\beta/\alpha) + ((\alpha - 1)\gamma)/\alpha + 1^\dagger$

TABLE 2. Continued.

Family and PDF	$J(X)$	$\sigma(X)$	$H(X)$
Beta(α, β)	$f(x) = (1/B(\alpha, \beta))x^{\alpha-1}(1-x)^{\beta-1}$ $-\frac{B(2\alpha-1, 2\beta-1)}{2B^2(\alpha, \beta)}, \alpha, \beta > \frac{1}{2}$	$0 \leq x \leq 1, \alpha > 0, \beta > 0$ $\sqrt{\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}}$	$\log B(\alpha, \beta) - (\alpha-1)\varphi_1(\alpha) - (\beta-1)\varphi_2(\beta)^*$
Generalized Pareto(α, β)	$f(x) = (1/\beta)(1 + (\alpha x/\beta))^{-(1/\alpha)-1}$ $-\frac{1 - \beta^{2/\alpha}(\alpha + \beta)^{-(2/\alpha)-1}}{2(\alpha + 2)}$	$x \geq 0$ if $\alpha > 0$; $x + (\beta/\alpha) \leq 0$ if $\alpha \leq 0$ $\frac{\beta}{(1-\alpha)\sqrt{1-2\alpha}}, \alpha < \frac{1}{2}$	$\log \beta + \alpha + 1$

[†] $\gamma = 0.5773 \dots$

* $\varphi_1(\alpha) = \psi(\alpha) - \psi(\alpha + \beta), \varphi_2(\beta) = \psi(\beta) - \psi(\alpha + \beta)$

from (1.2) and (3.5) we readily get

$$J(f_p) = \sum_{i=1}^n p_i^2 J(f_i) + 2 \sum_{i < j} p_i p_j J(f_i, f_j), \tag{3.6}$$

where $J(f_i, f_j) = -\frac{1}{2} \mathbb{E}[f_i(X_j)]$ is the inaccuracy measure of f_i with respect to f_j . It is evident that the expression in (3.6) is easy to compute, but there is no such expression for the entropy. Thus, this seems to be one advantage of extropy over entropy. We now present the following example as an illustration of the above result.

Example 3.1. Let $f_m = 0.5[N(\mu, \sigma^2) + N(-\mu, \sigma^2)]$ denote a mixed Gaussian distribution. It is clear that this distribution is obtained by just splitting a Gaussian distribution $N(0, \sigma^2)$ into two parts, centering one half about $+\mu$ and the other half about $-\mu$, and consequently has a mean of zero and variance $\sigma_m^2 = \sigma^2 + \mu^2$. In this case, [18] provided an analytical expression for signal entropy in situations when the corrupting noise source is mixed Gaussian, since a mixed Gaussian distribution is often considered as a noise model in many signal processing applications. From (3.6), $J(f_m) = 0.25\{J(f_1) + J(f_2)\} + 0.5J(f_1, f_2)$, and in this case it can be shown that

$$J(f_1) = J(f_2) = -\frac{1}{4\sqrt{\pi\sigma^2}}, \quad J(f_1, f_2) = -\frac{1}{4\sqrt{\pi\sigma^2}} e^{-\mu^2/\sigma^2}.$$

Thus, we obtain

$$J(f_m) = -\frac{1}{8\sqrt{\pi\sigma^2}} \{1 + e^{-\mu^2/\sigma^2}\},$$

which can be seen as an increasing function of σ^2 . Figure 1 shows $J(f_m)$ as a function of σ^2 with respect to various values of μ . The plots show that $J(f_m)$ is increasing with respect to both σ^2 and μ .

4. Characterizations based on maximum extropy

The maximum entropy (ME) principle is an extension of Laplace’s principle of insufficient reason for assigning probabilities. Both principles stipulate distributing the probability uniformly when the only available information is the support of the distribution \mathcal{S} . When additional information is available, the ME principle stipulates distributing the probability close to the uniform distribution while preserving the relevant information. In a similar vein, the maximum extropy (MEX) can be regarded as an extension of Laplace’s principle of insufficient reason for assigning probabilities. Let us consider the moment class of distributions

$$\Omega_\theta = \{f : \mathbb{E}[T_j(X)] = \theta_j < \infty, j = 0, 1, \dots, J\}, \tag{4.1}$$

where the $T_j(x)$ are integrable with respect to f and $T_0(x) = \theta_0 = 1$ is the normalizing factor. Then, the objective is to find f^* that maximizes $J(f)$ subject to a set of moment constraints defined in (4.1).

Theorem 4.1. Let Ω_θ be as defined in (4.1), with $T_j(x), j = 1, \dots, J$, being integrable functions with respect to f and $T_0(x) = \theta_0 = 1$ being the normalizing factor. Then, MEX is attained by the distribution with PDF

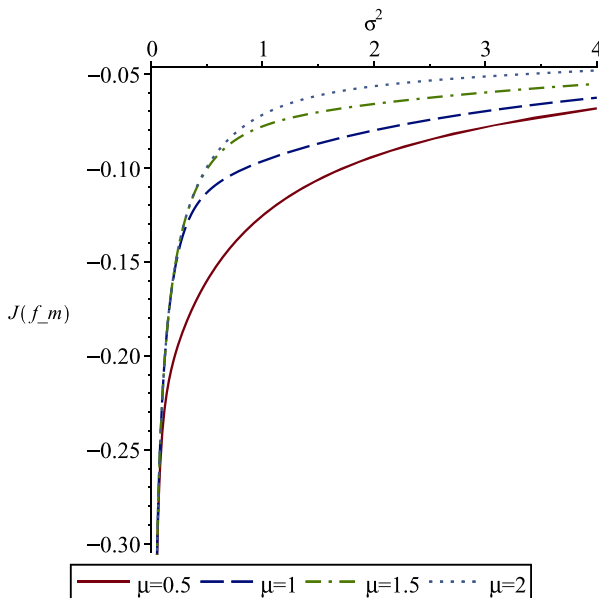


FIGURE 1. The entropy of mixture distribution $J(f_m)$ as a function of σ^2 given in Example 3.1.

$$f^*(x) = \arg \max_{f \in \Omega_\theta} J(f) = \sum_{j=0}^J \lambda_j T_j(x), \tag{4.2}$$

where $(\lambda_0, \lambda_1, \dots, \lambda_J)$ are Lagrange multipliers such that $\lambda_0 = 0$ when \mathcal{S} is unbounded.

Proof. The aim is to maximize $J(f) = -\frac{1}{2} \int_0^\infty f^2(x) dx$ subject to the constraints $\int_{\mathcal{S}} T_j(x) f(x) dx = \theta_j, j = 0, 1, \dots, J$, where \mathcal{S} may be bounded or unbounded, the $T_j(x)$ are integrable with respect to g , and $T_0(x) = \theta_0 = 1$ is the normalizing factor. The requirement is then equivalent to maximizing

$$\int_{\mathcal{S}} \left(-\frac{1}{2} f^2(x) + \lambda_0 f(x) + \sum_{j=1}^J \lambda_j T_j(x) f(x) \right) dx,$$

where $(\lambda_0, \lambda_1, \dots, \lambda_J)$ are Lagrange multipliers such that $\lambda_0 = 0$ when \mathcal{S} is unbounded. The Lagrangian is similar to the ME problem in terms of f , so taking derivatives gives the solution as in (4.2) [13]. Because the function $-\frac{1}{2}x^2, x > 0$, is concave, the solution is unique. \square

Theorem 4.1 readily provides the following characterizations of some well-known distributions.

Corollary 4.1.

- (i) The uniform distribution in $[0, 1]$ is the MEX model in the class of distributions with no constraint.
- (ii) A distribution with PDF $f(x) = 2(2 - 3\theta) + 6(2\theta - 1)x, 0 < x < 1$, is the MEX model in the class of distributions with finite expectation, $\mathbb{E}(X) = \theta$.

- (iii) The exponential distribution with PDF $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, is the MEX model in the class of distributions with finite moment $\mathbb{E}(e^{-\lambda X}) = 1$.
- (iv) The Weibull distribution with PDF $f(x) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}$, $x \geq 0$, $\alpha > 0$, is the MEX model in the class of distributions with finite moment $\mathbb{E}(e^{-\lambda X^\alpha}) = 1$.
- (v) A distribution with PDF $f(x) = e^{-x}(3 - 2x)$, $x > 0$, is the MEX model in the class of distributions with finite moments $\mathbb{E}(e^{-X}) = \mathbb{E}(Xe^{-X}) = 1$.

Corollary 4.1 leads us to derive the equilibrium distribution model as a MEX model as follows.

Theorem 4.2. *The solution to the constrained maximization problem*

$$\max_f J(f) \quad \text{subject to } \mathbb{E}[\bar{F}(X)] = 1, \quad \int_0^\infty f(x) dx = 1 \tag{4.3}$$

is the equilibrium distribution of the renewal time with PDF $f^*(x) = \tilde{f}_e(x)$, $x > 0$.

Proof. For an unbounded random variable, Theorem 4.1 gives the solution to (4.3) as $f^*(x) = \lambda_1 \bar{F}(x)$, where $\lambda_1^{-1} = \int_0^\infty \bar{F}(x) dx = \mu$. Thus, $f^*(x)$ is the PDF of the equilibrium distribution for the renewal time. □

Let us assume that $\psi(x)$ is an increasing differentiable function with $\psi'(x) = \phi(x) \geq 0$. Denote by \tilde{X}_ϕ the weighted version of \tilde{X}_e with PDF

$$\tilde{f}_\phi(x) = \frac{\phi(x)\bar{F}(x)}{\mathbb{E}[\psi(X)]} \tag{4.4}$$

for $x > 0$, where $\mathbb{E}[\psi(X)] = \int_S \phi(x)\bar{F}(x) dx$, provided it exists. The following theorem then generalizes Theorem 4.2.

Theorem 4.3. *The solution to the constrained maximization problem*

$$\max_f J(f) \quad \text{subject to } \mathbb{E}[\phi(X)\bar{F}(X)] = \theta_\phi, \quad \int_S f(x) dx = 1$$

is the weighted version of \tilde{X}_ϕ with PDF $f^*(x) = \tilde{f}_\phi(x)$, $x \in S$, in (4.4).

Proof. The proof is similar to that of Theorem 4.2 and is therefore omitted for the sake of brevity. □

Note that when $\phi(x) \equiv 1$, Theorem 4.3 reduces to Theorem 4.2. Moreover, the PDF $f^*(x) = 2x\bar{F}(X)/\mathbb{E}[X^2]$, $x \in S$, is the MEX model in the class of distributions with $\phi(x) = x$.

In an analogous manner, we can consider the minimum relative extropy (MREX). In this case, we seek the distribution in a class of probability distributions $\Omega = \{f\}$ that minimizes $d^c(f||g)$ for a given g , called the reference distribution, in terms of moment constraints. In this regard, we have the following theorem.

Theorem 4.4. *Let Ω_θ be as defined in (4.1), with $T_j(x)$, $j = 1, 2, \dots, J$, being integrable functions with respect to f and $T_0(x) = \theta_0 = 1$ being the normalizing factor. Then, MEX is attained by the distribution with PDF*

$$f^*(x) = \arg \min_{f \in \Omega_\theta} d^c(f||g) = g(x) - \sum_{j=0}^J \lambda_j T_j(x), \tag{4.5}$$

where $(\lambda_0, \lambda_1, \dots, \lambda_J)$ are Lagrangian multipliers such that $\lambda_0 = 0$ when S is unbounded.

Proof. The aim is to minimize $d^c(f||g) = \frac{1}{2} \int_S [f(x) - g(x)]^2 dx$ subject to the constraints $\int_S T_j(x)f(x) dx = \theta_j, j = 0, 1, \dots, J$, where S may be bounded or unbounded, the $T_j(x)$ are integrable with respect to g , and $T_0(x) = \theta_0 = 1$ is the normalizing factor. The requirement is then equivalent to minimizing $\int_S \psi(x) dx$, where

$$\psi(x) = \frac{1}{2}f^2(x) + \frac{1}{2}g^2(x) - f(x)g(x) + \lambda_0 f(x) + \sum_{j=1}^J \lambda_j T_j(x)f(x),$$

such that $(\lambda_0, \lambda_1, \dots, \lambda_J)$ are Lagrangian multipliers and $\lambda_0 = 0$ when S is unbounded. The Lagrangian is similar to the MEX problem in terms of f , so taking the derivatives gives the solution as in (4.5). Because the function $\frac{1}{2}x^2, x > 0$, is convex, the solution is unique. \square

5. Results on residual and past entropy

Let X be a non-negative random variable representing the lifetime of a unit, and let $t \geq 0$ denote its current age. Then, our interest now is the residual lifetime $S_t = \{x : x > t\}$. At age t , the PDF of the residual lifetime, $X_t = [X - t | X \geq t]$, is $f(x; t) = f(x)/\bar{F}(t), x \geq t > 0$. In this situation, the residual entropy is given by [21]

$$J(X; t) = -\frac{1}{2} \int_{S_t} f^2(x; t) dx = -\frac{1}{2} \int_t^\infty \left[\frac{f(x)}{\bar{F}(t)} \right]^2 dx = -\frac{1}{2} \mathbb{E}_t[f(X; t)], \tag{5.1}$$

where \mathbb{E}_t is the expectation with respect to the residual density $f(x; t)$. The residual entropy takes values in $[-\infty, 0)$ and it identifies with the entropy of $[X | X > t]$. Another useful expression can be given as

$$J(X; t) = -\frac{1}{4} \mathbb{E}_{12,t}[\lambda(X)], \tag{5.2}$$

where $\mathbb{E}_{12,t}$ is the expectation with respect to the residual density of $f_{12}(x)$ as defined in (1.4).

The question of whether $J(X; t)$ characterizes the lifetime distribution is answered partially in the following theorem.

Theorem 5.1. *Let X be a non-negative random variable with CDF F which is differentiable and has a continuous PDF f over S_t . If $f(x)$ is strictly decreasing over S_t , then $J(X; t)$ uniquely determines F .*

Proof. As $f(x)$ is strictly decreasing over S_t , we obtain

$$\mathbb{E}_t[f(X; t)] < \lambda_X(t). \tag{5.3}$$

Now, let us consider two random variables X and Y and suppose that $J(X; t) = J(Y; t)$, i.e.

$$\int_t^\infty \left[\frac{f(x)}{\bar{F}(t)} \right]^2 dx = \int_t^\infty \left[\frac{g(x)}{\bar{G}(t)} \right]^2 dx \quad \text{for all } t \in S_t.$$

Taking derivatives on both sides with respect to t we get

$$\lambda_X^2(t) - 2\lambda_X(t)\mathbb{E}_t[f(X; t)] = \lambda_Y^2(t) - 2\lambda_Y(t)\mathbb{E}_t[g(Y; t)].$$

Now, suppose there exists a $t^* \in \mathcal{S}_t$ such that $\lambda_Y(t^*) \neq \lambda_X(t^*)$. Rearranging the terms and letting $\mathbb{E}_{t^*}[f(X; t^*)] = \mathbb{E}_{t^*}[g(Y; t^*)]$, we obtain

$$\lambda_Y^2(t^*) - \lambda_X^2(t^*) = 2[\lambda_Y(t) - \lambda_X(t^*)]\mathbb{E}_{t^*}[f(X; t^*)] \quad \text{for all } t^* \in \mathcal{S}_t,$$

or equivalently

$$\lambda_Y(t^*) + \lambda_X(t^*) = 2\mathbb{E}_{t^*}[f(X; t^*)] \quad \text{for all } t^* \in \mathcal{S}_t. \tag{5.4}$$

Without loss of generality, let $\lambda_Y(t^*) > \lambda_X(t^*)$. Then, using (5.4), we obtain $\mathbb{E}_{t^*}[f_{t^*}(X)] > \lambda_X(t^*)$, which is a contradiction to the condition in (5.3). For the case when $\lambda_Y(t^*) < \lambda_X(t^*)$, the contradiction is obtained in terms of $\mathbb{E}_{t^*}[g_{t^*}(Y)] > \lambda_Y(t^*)$, which completes the proof of the theorem. \square

It is important to mention that the above theorem is applicable to a large class of distributions that include monotone densities [4]. The following theorem relates the dynamic extropy and hazard rate orderings.

Theorem 5.2. *Let X and Y be two non-negative continuous random variables having CDFs F and G , PDFs f and g , and hazard rate functions λ_X and λ_Y , respectively. If $X \leq_{hr} Y$ and either X or Y is DFR, then $J(X; t) \leq J(Y; t)$.*

Proof. Let X_t and Y_t denote the residual lifetime variables with densities f_t and g_t , respectively. The condition $X \leq_{hr} Y$ implies that $X_{12,t} \leq Y_{12,t}$ in the usual stochastic order, where X_{12} and Y_{12} have survival functions $\bar{F}^2(x)$ and $\bar{G}^2(x)$, respectively. If we assume that X is DFR, then $\mathbb{E}_{12,t}[\lambda_X(X)] \geq \mathbb{E}_{12,t}[\lambda_X(Y)] \geq \mathbb{E}_{12,t}[\lambda_Y(Y)]$. From (5.2), we get $J(X; t) \leq J(Y; t)$. If Y is DFR, then, using a similar argument, we again obtain $J(X; t) \leq J(Y; t)$. Hence, the theorem. \square

Example 5.1. Let X be an absolutely continuous non-negative random variable with PDF $f(x)$ and survival function $\bar{F}(x)$. Further, let $0 \equiv X_0 \leq X_1 \leq X_2 \leq \dots$ denote the epoch times of a non-homogeneous Poisson process (NHPP) with intensity function $\lambda(x)$, $x \geq 0$, where X_1 has the same distribution as X . Let $T_n = X_n - X_{n-1}$, $n \in \mathbb{N}$, denote the length of the n th inter-epoch interval (or inter-occurrence time). Denoting by $\bar{F}_n(x)$ the survival function of X_n , $n \in \mathbb{N}$, we have [2]

$$\bar{F}_n(x) = \bar{F}(x) \sum_{k=0}^{n-1} \frac{\Lambda^k(x)}{k!}, \quad x \geq 0.$$

From [24, Example 1.B.13], it is known that $T_n \leq_{hr} T_{n+1}$. On the other hand, for all $n \geq 1$, if X is DFR then T_n is DFR due to [11, Theorem 5]. As a result, Theorem 5.2 implies that $J(T_n; t) \leq J(T_{n+1}; t)$.

We now propose two new classes of life distributions by combining the notions of extropy and aging.

Definition 5.1. Let X be an absolutely continuous random variable with PDF f . Then, we say that X has increasing/decreasing dynamic extropy (IDEX/DDEX) if $J(X; t)$ is increasing/decreasing.

Roughly speaking, if a unit has a CDF that belongs to the class of DDEX, then as the unit ages, the conditional probability density function becomes more informative. The following theorem gives a relationship between these classes and the well-known increasing (decreasing) failure rate classes of life distributions.

Theorem 5.3. *For an absolutely continuous random variable X with PDF f , if X is IFR (DFR), then X is DDEX (IDEX).*

Proof. We prove it for IFR; the DFR case can be handled in an analogous manner. Suppose X is IFR; then, from (5.2), for $t > 0$ we get

$$J'(X; t) = \frac{\lambda_{12}(t)}{4} \left[\lambda(t) - \int_t^\infty \lambda(x) \frac{f_{12}(x)}{F_{12}(t)} dx \right] \leq \frac{\lambda_{12}(t)}{4} [\lambda(t) - \lambda(t)] = 0,$$

where $\lambda_{12}(t)$ is the hazard rate function of X_{12} . From this, we see that $J(X; t)$ is decreasing in t , i.e. X is DDEX. □

Another important class of life distributions is the class of increasing failure rate in average (IFRA) distributions. Recall that X is IFRA if $H(x)/x$ is increasing in x , where $H(x) = -\log \bar{F}(x)$ denotes the cumulative hazard function. The following example shows that there is no relationship between the proposed class and the IFRA class of life distributions.

Example 5.2. Consider the random variable X with survival function

$$\bar{F}(t) = \begin{cases} 1 & \text{if } 0 \leq t < 2, \\ e^{2-t} & \text{if } 2 \leq t < 3, \\ e^{-1} & \text{if } 3 \leq t < 4, \\ e^{7-2t} & \text{if } t \geq 4. \end{cases}$$

Figure 2 presents the residual extropy and the function $H(t)/t$, from which we observe that X is not an IFRA distribution. Moreover, it is easy to verify that, in this case,

$$J(X; t) = \begin{cases} -\frac{1}{4}(e^{2t-6} + 1) & \text{if } 2 \leq t < 3, \\ -\frac{1}{2} & \text{if } t \geq 3. \end{cases}$$

The plot of the residual extropy in Fig. 2 shows that the random variable X is DDEX. This example also shows that DDEX does not imply the IFR property.

The connection between the extropy residual life functions of two random variables and the proportional hazard model is explored in the following theorem.

Theorem 5.4. *Let X and Y be two absolutely continuous non-negative random variables with survival functions $\bar{F}(t)$ and $\bar{G}(t)$, and hazard rate functions $\lambda_X(t)$ and $\lambda_Y(t)$, respectively. Further, let $\theta(t)$ be a non-negative increasing function such that $\lambda_Y(t) = \theta(t)\lambda_X(t)$, $t > 0$, and $0 \leq \theta(t) \leq 1$. Then, if $J(X; t)$ is a decreasing function of t , $J(Y; t)$ is also decreasing in t , provided $\lim_{t \rightarrow \infty} (\bar{G}(t)/\bar{F}(t)) < \infty$.*

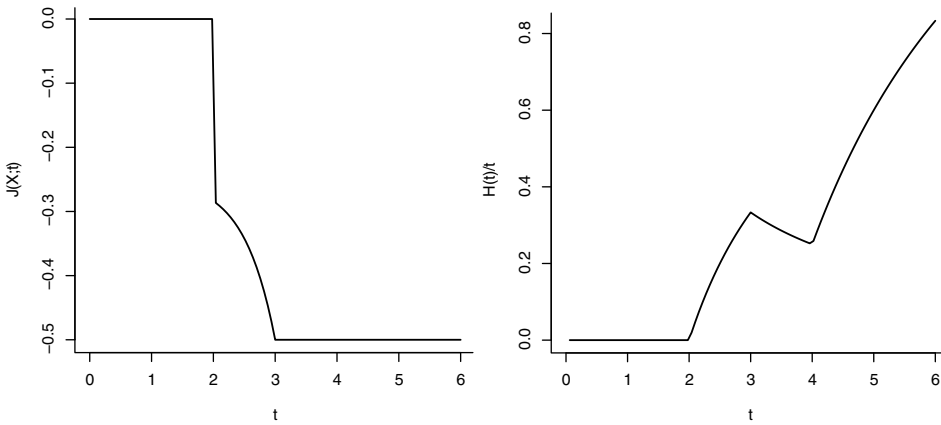


FIGURE 2. The residual entropy (left panel) and the ratio of the hazard function with respect to t (right panel) given in Example 5.2.

Proof. From (5.2), $J(Y; t)$ is decreasing in t if and only if $\mathbb{E}_{12,t}[\lambda_Y(Y)]$ is increasing in t . Let us set $m_1(t) = \mathbb{E}_{12,t}[\lambda_X(X)]$ and $m_2(t) = \mathbb{E}_{12,t}[\lambda_Y(Y)]$. Then, because $m_1'(t) = \lambda_{12}^X(t)[m_1(t) - \lambda_X(t)]$ and $m_2'(t) = \lambda_{12}^Y(t)[m_2(t) - \lambda_Y(t)]$, $m_2(t)$ is increasing in t if $m_2(t) \geq \lambda_Y(t) = \theta(t)\lambda_X(t)$, which holds if $m_2(t) \geq \theta(t)m_1(t)$, $t > 0$. Define the function $\varphi(t)$ as

$$\varphi(t) = \bar{G}_{12}(t)[\theta(t)m_1(t) - m_2(t)].$$

We now prove that $\varphi(t) \leq 0$. Differentiating $\varphi(t)$ with respect to t and then performing some algebraic manipulations, we obtain

$$\begin{aligned} \varphi'(t) &= -g_{12}(t)(\theta(t)m_1(t) - m_2(t)) + \bar{G}_{12}(t)\{\theta'(t)m_1(t) + \theta(t)m_1'(t) - m_2'(t)\} \\ &= -g_{12}(t)(\theta(t)m_1(t) - m_2(t)) \\ &\quad + \bar{G}_{12}(t)\{\theta'(t)m_1(t) + 2\theta(t)\lambda_X(t)(m_1(t) - \lambda_X(t)) - 2\lambda_Y(t)(m_2(t) - \lambda_Y(t))\} \\ &= \bar{G}_{12}(t)\{\theta'(t)m_1(t) + 2\theta(t)\lambda_X(t)m_1(t) - 2\theta(t)\lambda_X^2(t) - 2\theta(t)\lambda_Y(t)m_1(t) + 2\theta(t)\lambda_X(t)\lambda_Y(t)\} \\ &= \bar{G}_{12}(t)\{\theta'(t)m_1(t) + 2\theta(t)(m_1(t) - \lambda_X(t))(\lambda_X(t) - \lambda_Y(t))\}. \end{aligned}$$

From the assumptions that $\theta(t)$ and $m_1(t)$ are increasing in t , we get $\varphi'(t) > 0$, i.e. $\varphi(t)$ is increasing in t . Now, as $\lim_{t \rightarrow \infty} (\bar{G}(t)/\bar{F}(t)) < \infty$, we get

$$\lim_{t \rightarrow \infty} \varphi(t) = \lim_{t \rightarrow \infty} \left\{ \left(\frac{\bar{G}(t)}{\bar{F}(t)} \right)^2 \int_t^\infty \theta(t)\lambda_X(x)f_{12}(x) dx \right\} - \lim_{t \rightarrow \infty} \left\{ \int_t^\infty \lambda_Y(x)g_{12}(x) dx \right\} = 0.$$

Hence, $\varphi(t) \leq 0$ for any t , i.e. $m_1(t) \leq m_2(t)$, which completes the proof of the theorem. □

Consider a parallel system with n units having lifetimes X_1, \dots, X_n , which are i.i.d. absolutely continuous random variables with CDF $F(x)$. The corresponding system lifetime is $X_{n:n} = \max\{X_1, \dots, X_n\}$, whose CDF is $F_{n:n}(x) := \mathbb{P}(X_{n:n} \leq x) = [F(x)]^n$, $x \geq 0$. Then, we have the following theorem, which gives the closure property of DDEX distributions under the

formation of parallel systems. Its proof is similar to that of [3, Theorem 2.3], so we do not present it here.

Theorem 5.5. *Let X_1, \dots, X_n be a set of i.i.d. random variables with CDF F , PDF f , hazard rate function λ , and decreasing residual entropy $J(X; t)$. If $J(X_{n:n}; t)$ denotes the residual entropy of the n th-order statistic among X_1, \dots, X_n , then $J(X_{n:n}; t)$ is also decreasing.*

Let X be a non-negative random variable representing the lifetime of a unit, and $t \geq 0$ denote its current age. It is then of interest to examine the inactivity time of the item with support $\mathcal{S}_{[t]} = \{x: x \leq t\}$. At age t , the PDF of the inactivity time, $X_{[t]} = [t - X | X \leq t]$, is given by $f(x; [t]) = f(x)/F(t)$, $0 < x \leq t$. Then, the past entropy is defined as

$$\tilde{J}(X; [t]) = -\frac{1}{2} \int_{\mathcal{S}_{[t]}} f^2(x; [t]) dx = -\frac{1}{2} \int_0^t \left[\frac{f(x)}{F(t)} \right]^2 dx = -\frac{1}{2} \mathbb{E}_{[t]}[f(X; [t])],$$

where $\mathbb{E}_{[t]}$ is the expectation with respect to the inactivity density, $f(x; [t])$. In analogy with (5.1), the past entropy also takes values in $[-\infty, 0)$ and it identifies with the entropy of $[X | X \leq t]$; see, e.g., [15]. By using (1.3), another useful expression for it can be given as

$$\tilde{J}(X; [t]) = -\frac{1}{4} \mathbb{E}_{22,t}[\tau(X)], \tag{5.5}$$

where $\mathbb{E}_{22,t}$ is the expectation with respect to the inactivity density of $f_{22}(x)$ defined in (1.5). We now propose a new class of life distributions based on the notion of past entropy.

Definition 5.2. We say that X has increasing past entropy (IPEX) if $\tilde{J}(X; [t])$ is increasing in $t > 0$.

The expression in (5.5) is useful in examining the behavior of past entropy in terms of the behavior of the reversed failure rate, as done in the following theorem.

Theorem 5.6. *For an absolutely continuous non-negative random variable X with PDF f , if X is DRFR, then X is IPEX.*

Proof. If X is DRFR, then $\tau(t)$ is decreasing in t , so

$$\tilde{J}'(X; [t]) = \frac{\tau_{12}(t)}{4} \left[\int_0^t \tau(x) \frac{f_{22}(x)}{F_{22}(t)} dx - \tau(t) \right] \geq \frac{\tau_{12}(t)}{4} \{\tau(t) - \tau(t)\} = 0,$$

where $\tau_{22}(t)$ is the reversed hazard rate function of X_{22} . Hence, the theorem. □

The following example demonstrates the usefulness of Theorem 5.6 in recognizing some IPEX distributions.

Example 5.3.

- (i) Let X be an exponential random variable with PDF $f(x) = \lambda e^{-\lambda x}$ for $x > 0$, $\lambda > 0$. The RFR function of X is $\tau(x) = \lambda[1 - e^{-\lambda x}]^{-1}$. We can easily check that $\tau(x)$ is decreasing in x , and so, according to Theorem 5.6, X is IPEX.
- (ii) Let X have an inverse Weibull distribution with CDF $F(x) = \exp[-(1/\sigma x)^\lambda]$, $x > 0$, $\sigma, \lambda > 0$. The RFR function is $\tau(x) = \lambda \sigma^{-\lambda} x^{-(1+\lambda)}$, which is decreasing in x . Hence, X is IPEX.

Theorem 5.7. *Let X and Y be two absolutely continuous non-negative random variables with reversed hazard rate functions $\tau_X(t)$ and $\tau_Y(t)$, respectively. Further, let $\theta(t)$ be a non-negative increasing function such that $\tau_Y(t) = \theta(t)\tau_X(t)$, $t > 0$, and $0 \leq \theta(t) \leq 1$. Then, if $\tilde{J}(X; [t])$ is a decreasing function of t , $\tilde{J}(Y; [t])$ is also decreasing in t , provided $\lim_{t \rightarrow 0} (G(t)/F(t)) < \infty$.*

Note that this theorem connects the past extropy of two random variables to the known proportional reversed hazard rates model.

6. Concluding remarks

Many information measures have been studied in the literature. For instance, entropy functions are used to measure the uncertainty in a random variable. If these entropy functions are applied to residual lifetime or past lifetime (or inactivity time) variables, then we obtain dynamic measures of uncertainty that can measure the aging process. We have provided several results on extropy, which is a complementary dual function of entropy. Some similarities between entropy, extropy, and variance have been discussed. In spite of some agreements between these measures, there are some notable differences as well. For example, many well-known families of distributions have been characterized as the unique maximum entropy and extropy solutions, while no such characterization is available in terms of variance. It needs to be mentioned that there is no universal relationship between entropy, extropy, and variance orderings of distributions. One advantage of extropy as compared to other measures is that it yields an expression for finite mixture distributions, while no such expression is available for closed-form entropy and variance measures. We have shown that extropy information ranks uniformity of a wide variety of absolutely continuous distributions. We have then elaborated on some theoretical merits of extropy and presented several results about the associated characterizations and also its dynamic versions. The most important advantage of extropy is that it is easy to compute, and it will therefore be of great interest to explore its important potential applications in developing goodness-of-fit tests and inferential methods.

Acknowledgements

A. Toomaj was partially supported by a grant from Gonbad Kavous University. We thank the Editor-in-Chief and anonymous reviewers for their useful comments and suggestions on an earlier version of this manuscript which led to this improved version.

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] ARNOLD, B. C., BALAKRISHNAN, N. AND NAGARAJA, H. N. (2008). *A First Course in Order Statistics*. SIAM, Philadelphia, PA.
- [2] ARNOLD, B. C., BALAKRISHNAN, N. AND NAGARAJA, H. N. (2011). *Records*. Wiley, Chichester.
- [3] ASADI, M. AND EBRAHIMI, N. (2000). Residual entropy and its characterizations in terms of hazard function and mean residual life function. *Statist. Prob. Lett.* **49**, 263–269.

- [4] ASADI, M., EBRAHIMI, N., HAMEDANI, G. AND SOOFI, E. S. (2004). Maximum dynamic entropy models. *J. Appl. Prob.* **41**, 379–390.
- [5] ASADI, M., EBRAHIMI, N., HAMEDANI, G. AND SOOFI, E. S. (2005). Minimum dynamic discrimination information models. *J. Appl. Prob.* **42**, 643–660.
- [6] ASADI, M., EBRAHIMI, N., KHARAZMI, O. AND SOOFI, E. S. (2018). Mixture models, Bayes Fisher information, and divergence measures. *IEEE Trans. Inf. Theory* **65**, 2316–2321.
- [7] EBRAHIMI, N., MAASOUMI, E. AND SOOFI, E. S. (1999). Ordering univariate distributions by entropy and variance. *J. Econometrics* **90**, 317–336.
- [8] EBRAHIMI, N., SOOFI, E. S. AND SOYER, R. (2010). Information measures in perspective. *Int. Statist. Rev.* **78**, 383–412.
- [9] EBRAHIMI, N., SOOFI, E. S. AND ZAHEDI, H. (2004). Information properties of order statistics and spacings. *IEEE Trans. Inf. Theory* **50**, 177–183.
- [10] GOOD, I. (1968). Utility of a distribution. *Nature* **219**, 1392–1392.
- [11] GUPTA, R. C. AND KIRMANI, S. (1988). Closure and monotonicity properties of nonhomogeneous Poisson processes and record values. *Prob. Eng. Inf. Sci.* **2**, 475–484.
- [12] HILD, K. E., PINTO, D., ERDOGMUS, D. AND PRINCIPE, J. C. (2005). Convolutional blind source separation by minimizing mutual information between segments of signals. *IEEE Trans. Circuits Systems* **52**, 2188–2196.
- [13] JAYNES, E. T. (1982). On the rationale of maximum-entropy methods. *Proc. IEEE* **70**, 939–952.
- [14] KHARAZMI, O. AND BALAKRISHNAN, N. (2021). Cumulative residual and relative cumulative residual Fisher information and their properties. *IEEE Trans. Inf. Theory* **67**, 6306–6312.
- [15] KRISHNAN, A. S., SUNOJ, S. AND UNNIKRISHNAN NAIR, N. (2020). Some reliability properties of extropy for residual and past lifetime random variables. *J. Korean Statist. Soc.* **49**, 457–474.
- [16] KULLBACK, S. (1997). *Information Theory and Statistics*. Courier Corporation, North Chelmsford, MA.
- [17] LAD, F. *et al.* (2015). Extropy: Complementary dual of entropy. *Statist. Sci.* **30**, 40–58.
- [18] MICHALOWICZ, J. V., NICHOLS, J. M. AND BUCHOLTZ, F. (2008). Calculation of differential entropy for a mixed Gaussian distribution. *Entropy* **10**, 200–206.
- [19] NAKAGAWA, T. (2006). *Maintenance Theory of Reliability*. Springer, New York.
- [20] NANDA, A. K. AND JAIN, K. (1999). Some weighted distribution results on univariate and bivariate cases. *J. Statist. Planning Infer.* **77**, 169–180.
- [21] QIU, G. AND JIA, K. (2018). Extropy estimators with applications in testing uniformity. *J. Nonparametric Statist.* **30**, 182–196.
- [22] QIU, G. AND JIA, K. (2018). The residual extropy of order statistics. *Statist. Prob. Lett.* **133**, 15–22.
- [23] ROHDE, G., NICHOLS, J., BUCHOLTZ, F. AND MICHALOWICZ, J. (2007). Signal estimation based on mutual information maximization. In *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*. IEEE, New York, pp. 597–600.
- [24] SHAKED, M. AND SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. Springer, New York.
- [25] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423.
- [26] SOOFI, E. S., EBRAHIMI, N. AND HABIBULLAH, M. (1995). Information distinguishability with application to analysis of failure data. *J. Amer. Statist. Assoc.* **90**, 657–668.
- [27] TAN, Y., TANTUM, S. L. AND COLLINS, L. M. (2004). Cramèr–Rao lower bound for estimating quadrupole resonance signals in non-Gaussian noise. *IEEE Sig. Proc. Lett.* **11**, 490–493.
- [28] YANG, J., XIA, W. AND HU, T. (2019). Bounds on extropy with variational distance constraint. *Prob. Eng. Inf. Sci.* **33**, 186–204.
- [29] YUAN, A. AND CLARKE, B. (1999). An information criterion for likelihood selection. *IEEE Trans. Inf. Theory* **45**, 562–571.