Probability in the Engineering and Informational Sciences (2023), **37**:1 294–321
doi:10.1017/S0269964822000031

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

On reconsidering entropies and divergences and their cumulative counterparts: Csiszár's, DPD's and Fisher's type cumulative and survival measures

Konstantinos Zografos

Department of Mathematics, University of Ioannina, 451 10 Ioannina, Greece. E-mail: kzograf@uoi.gr.

Keywords: Cumulative divergence, Cumulative entropy, Cumulative Fisher information, Mutual information, Shannon entropy, Survival divergence, Survival entropy, Survival Fisher information, Tsallis entropy

AMS Subject Classification: 62B10, 94A17, 62E10, 62N99, 60E99

Abstract

This paper concentrates on the fundamental concepts of entropy, information and divergence to the case where the distribution function and the respective survival function play the central role in their definition. The main aim is to provide an overview of these three categories of measures of information and their cumulative and survival counterparts. It also aims to introduce and discuss Csiszár's type cumulative and survival divergences and the analogous Fisher's type information on the basis of cumulative and survival functions.

1. Introduction

Measures of entropy, information and divergence have a long history and they hold a prominent position in the scientific life and literature. Some of them, like Shannon entropy, Shannon [81], Fisher information measure, Fisher [35], and Kullback–Leibler divergence, Kullback and Leibler [49] and Kullback [48], have played a prominent role in the development of many scientific fields. Since the early 1960s and in light of the above-mentioned omnipresent and prominent universal quantities, there has been increased interest in the definition, the study, the axiomatic characterization and the applications of measures which formulate and express: (i) the amount of information or uncertainty about the outcome of a random experiment, (ii) the amount of information about the unknown characteristics of a population or about the unknown parameters of the respective distribution that drives the population or, (iii) the amount of information for discrimination between two distributions or between the respective populations which are driven by them. A classification of the measures of information into these three broad categories and a tabulation and discussion of their main properties and applications is provided in the publications by Ferentinos and Papaioannou [34], Papaioannou [62], Zografos *et al.* [99], Vajda [89], Soofi [83,84], Papaioannou [63], Cover and Thomas [26], Pardo [65], among many other citations in the field. The increasing interest of the scientific community in measures of information and the numerous applications of these quantities in several disciplines and contexts, in almost all the fields of science and engineering and first of all in probability and statistics, have contributed to the development of the field of statistical information theory, a field and a terminology which was initiated, to the best of our knowledge, in the title of the monograph by Kullback *et al.* [50]. The numerous applications of the measures of information in the area of probability and statistics have led to an enormous number of papers, monographs and books, like that by Csiszár and Körner [30], Kullback *et al.* [50], Liese and Vajda [51], Read and Cressie [74], Vajda [89], Arndt [5], Pardo [65], Basu *et al.* [14], among others, which have been published after the seminal monograph by Kullback [48]. A huge number of statistical techniques and methods have been

introduced and developed on the basis of entropies and divergences. These techniques are presented in the above monographs and the bibliography cited in them.

A resurgence of the interest in the definition and the development of new ideas and information theoretic measures is signaled by the paper of Rao *et al.* [73] and a subsequent paper by Zografos and Nadarajah [98] where entropy type measures are defined on the basis of the cumulative distribution function or on the basis of the respective survival function. These papers were the basis for an increasing interest for the definition of informational measures in the context of the paper by Rao *et al.* [73]. In this direction, entropy and divergence type measures are introduced and studied, in the framework of the cumulative distribution function or in terms of the respective survival function, in the papers by Rao [72], Zografos and Nadarajah [98], Di Crescenzo and Longobardi [31], Baratpour and Rad [12], Park *et al.* [66] and the subsequent papers by Di Crescenzo and Longobardi [32], Klein *et al.* [46], Asadi *et al.* [6,7], Park *et al.* [67], Klein and Doll [45], among many others. In these and other treatments, entropy type measures and Kullback-Leibler type divergences have been mainly received the attention of the authors. Entropy and divergence type measures are also considered in the papers by Klein *et al.* [46], Asadi *et al.* [6], Klein and Doll [45] by combining the cumulative distribution function and the survival function. However, to the best of our knowledge, it seems that there has not yet appeared in the existing literature a definition of the broad class of Csiszár's type ϕ-divergences or a definition of the density power divergence in the framework which was initiated in the paper by Rao *et al.* [73]. In addition, to the best of our knowledge, there is no an analogous formulation of Fisher's measure of information, as by product of this type of divergences. This paper aims to bridge this gap.

In the context described above, this paper is structured as follows. The next section provides a short review of measures of entropy, divergence and Fisher's type in the classic setup. A similar review is provided in Section 3 for the concepts of cumulative and survival entropies and divergences and the proposed in the existing literature respective measures. Section 4 is devoted to the definition of Csiszár's type ϕ-divergences in terms of the cumulative and the survival function. The density power divergence type is also defined in the same setup. Section 5 is concentrated to the definition of Fisher measure of information in terms of the cumulative distribution function and the survival function. The content of the paper is summarized in the last section where some conclusions and directions of a future work are also presented.

2. A short review on entropies and divergences

To present some of the measures of entropy and divergence, which will be mentioned later, consider the probability space $(\mathcal{X}, \mathcal{A}, P)$, and let a σ-finite measure μ on the same space with $P \ll \mu$. Denote by f the respective Radon–Nikodym derivative $f = dP/d\mu$. The Shannon entropy, Shannon [81], is defined by

$$\mathcal{E}_{Sh}(f) = -\int_{\mathcal{X}} f(x) \ln f(x) d\mu, \tag{1}$$

and it is a well known and broadly applied quantity, as its range of applications is extending from thermodynamics to algorithmic complexity, including a fundamental usage in probability and statistics (cf. [26]). Two well-known extensions of Shannon entropy have been introduced by Rényi [75] and Tsallis [88], as follows:

$$\mathcal{E}_{R,\alpha}(f) = \frac{1}{1-\alpha} \ln \int_{\mathcal{X}} f^{\alpha}(x) d\mu, \quad \alpha > 0,\ \alpha \neq 1, \tag{2}$$

and

$$\mathcal{E}_{Ts,\alpha}(f) = \frac{1}{\alpha - 1} \left(1 - \int_{\mathcal{X}} f^{\alpha}(x) d\mu\right), \quad \alpha > 0,\ \alpha \neq 1, \tag{3}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

respectively. It is easily seen that $\lim_{\alpha\to 1}\mathcal{E}_{R,\alpha}(f) = \mathcal{E}_{Sh}(f)$ and $\lim_{\alpha\to 1}\mathcal{E}_{Ts,\alpha}(f) = \mathcal{E}_{Sh}(f)$. All the above presented measures have been initially defined in the discrete case. Although the concept of entropy has been introduced by the second law of thermodynamics, Shannon has defined $\mathcal{E}_{Sh}$, in the discrete case, as a measure of the information transmitted in a communication channel. Shannon entropy is analogous of the thermodynamic entropy and from a probabilistic or statistical point of view it is a measure of uncertainty before the implementation of a random experiment regarding its final result. This interpretation is based on the fact that Shannon entropy is maximized subject to the most uncertain distribution, the univariate discrete uniform distribution. Hence, Shannon entropy quantifies uncertainty relative to the uniform distribution and this is the fundamental characteristic of all proper entropy measures. Rényi's measure $\mathcal{E}_{R,\alpha}$ extends Shannon's entropy while Tsallis' entropy $\mathcal{E}_{Ts,\alpha}$ has motivated by problems in statistical mechanics and it is related to the α-order entropy of Havrda and Charvát [41], cf. also Pardo [65] p. 20 Table 1.1. It should be emphasized, at this point, that $\mathcal{E}_{R,\alpha}$ and $\mathcal{E}_{Ts,\alpha}$ are functionally related, for $\alpha = 2$, with the extropy of a random variable X which receives a great attention the last decade (cf. [36,69,70]). The extropy of a random variable X with an absolutely continuous distribution function F and respective probability density function f is defined by $J(f) = -\frac{1}{2}\int_{\mathcal{X}} f^2(x)d\mu$ and it has been introduced in the statistical literature as the complement of Shannon entropy. This measure is directly connected with Onicescu [61] information energy defined by $E(f) = -2J(f)$. It is also noted that $\int_{\mathcal{X}} f^{\alpha}(x)d\mu$, in (2) and (3), defines the Golomb [37] information function (cf. also [39]) which is still used nowadays (cf. [44]). Last, we have to mention that Rényi's measure $\mathcal{E}_{R,\alpha}$ is the basis for the definition by Song [82] of a general measure of the shape of a distribution. Song's measure is defined by $\mathcal{S}(f) = -2(d/d\alpha)\mathcal{E}_{R,\alpha}(f)|_{\alpha=1} = \mathrm{Var}[\ln f(X)]$ and it has been applied and studied inside the family of elliptically contoured distributions in Zografos [97] and Batsidis and Zografos [15]. The measure $\mathcal{S}(f)$ above is the varentropy used in Kontoyiannis and Verdú [47] and Arikan [4], where the measure is defined in terms of conditional distributions.

Burbea and Rao [20] have extended (1) and (3) by introducing the ϕ-entropy functional

$$\mathcal{E}_{\phi}(f) = -\int_{\mathcal{X}} \phi(f(x))d\mu, \tag{4}$$

where ϕ is a convex real function which satisfies suitable conditions. Shannon's and Tsallis' entropies are obtained as particular cases of $\mathcal{E}_{\phi}$ for specific choices of the convex function ϕ and more precisely for $\phi(u) = u\ln u$ and $\phi(u) = (\alpha-1)^{-1}(u^{\alpha}-u)$, $\alpha > 0, \alpha \neq 1, u > 0$, respectively. Observe that Rényi's measure $\mathcal{E}_R$ does not directly follow from ϕ-entropy functional. This point led Salicru *et al.* [79] to define the (h,ϕ)-entropy which unified all the existing, at that time, entropy measures. Based on Pardo [65] p. 21, the (h,ϕ)-entropy is defined as follows:

$$\mathcal{E}^{h}_{\phi}(f) = h\left(\int_{\mathcal{X}} \phi(f(x))d\mu\right), \tag{5}$$

where either $\phi : (0,\infty) \to \mathbb{R}$ is concave and $h : \mathbb{R} \to \mathbb{R}$ is differentiable and increasing, or $\phi : (0,\infty) \to \mathbb{R}$ is convex and $h : \mathbb{R} \to \mathbb{R}$ is differentiable and decreasing. Table 1.1 in p. 20 of Pardo [65] lists important entropy measures obtained from (5) for particular choices of the functions ϕ and h.

Following the above short overview on the most historic measures of entropy, lets now proceed to a short review on measures of divergence between probability distributions. Consider the measurable space $(\mathcal{X},\mathcal{A})$, and two probability measures P and Q on this space. Let a σ-finite measure μ on the same space with $P \ll \mu$ and $Q \ll \mu$. Denote by f and g the respective Radon–Nikodym derivatives, $f = dP/d\mu$ and $g = dQ/d\mu$. The most historic measure of divergence is the well-known Kullback–Leibler divergence (cf. [48,49]) which is defined by

$$\mathcal{D}_0(f:g) = \int_{\mathcal{X}} f(x)\ln\left(\frac{f(x)}{g(x)}\right)d\mu. \tag{6}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

Intuitively speaking, $\mathcal{D}_0$ expresses the information, contained in the data, for discrimination between the underlined models f and g. Several interpretations of $\mathcal{D}_0$ are discussed in the seminal paper by Soofi [83]. In this context, $\mathcal{D}_0$ quantifies the expected information, contained in the data, for discrimination between the underlined models f and g in favor of f. This interpretation is based on the Bayes Theorem (cf. [48] pp. 4–5). Moreover, $\mathcal{D}_0$ measures loss or gain of information. It has been interpreted as a measure of loss of information when one of the two probability density functions represents an ideal distribution and $\mathcal{D}_0$ measures departure from the ideal; (e.g., f is the unknown "true" data-generating distribution and g is a model utilized for the analysis). However, following [83] p. 1246, $\mathcal{D}_0$ in (6) "is often used just as a measure of divergence between two probability distributions rather than as a meaningful information quantity in the context of the problem being discussed." The above defined Kullback–Leibler divergence $\mathcal{D}_0$ satisfies the non-negativity property, that is, $\mathcal{D}_0(f : g) \geq 0$, with equality if and only if the underlined densities are coincide, $f = g$, a.e. (cf. [48] p. 14).

Rényi [75] has extended the above measure by introducing and studying the information of order α by the formula:

$$\mathcal{D}_{R,\alpha}(f : g) = (1/(\alpha - 1)) \ln \int_{\mathcal{X}} f^{\alpha}(x) g^{1-\alpha}(x) d\mu, \quad \alpha > 0, \ \alpha \neq 1. \tag{7}$$

This measure is related with Kullback–Leibler divergence by the limiting behavior, $\lim_{\alpha \to 1} \mathcal{D}_{R,\alpha}(f : g) = \mathcal{D}_0(f : g)$. After Rényi's divergence, the broad class of ϕ-divergence between two densities f and g has been introduced by Csiszár [28,29] and independently by Ali and Silvey [1]. Some authors (see, e.g., [40]) mention that ϕ-divergence has been also independently introduced by Morimoto [58]. This omnipresent measure is defined by

$$\mathcal{D}_{\phi}(f : g) = \int_{\mathcal{X}} g(x) \phi \left(\frac{f(x)}{g(x)} \right) d\mu, \tag{8}$$

for two Radon–Nikodym derivatives f and g on the measurable space $\mathcal{X}$. $\phi : (0, \infty) \to \mathbb{R}$ is a real-valued convex function satisfying conditions which ensure the existence of the above integral. Based on Csiszár [28,29] and Pardo [65] p. 5, it is assumed that the convex function ϕ belongs to the class of functions

$$\Phi = \left\{ \phi : \phi \text{ is strictly convex at 1, with } \phi(1) = 0, 0\phi\left(\frac{0}{0}\right) = 0, 0\phi\left(\frac{u}{0}\right) = \lim_{v \to \infty} \frac{\phi(v)}{v} \right\}. \tag{9}$$

In order to be (8) useful in statistical applications, the class Φ is enriched with the additional assumption $\phi'(1) = 0$ (cf. [65] p. 5). Csiszár's ϕ-divergence owes its wide range of applications to the fact that it can be considered as a measure of quasi-distance or a measure of statistical distance between two probability densities f and g since it obeys the non-negativity and identity of indiscernibles property, a terminology which is conveyed by Weller-Fahy *et al.* [93] and it is formulated by

$$\mathcal{D}_{\phi}(f : g) \geq 0 \text{ with equality if and only if } f(x) = g(x), \quad \text{a.e.} \tag{10}$$

Csiszár's ϕ-divergence is not symmetric for each convex function $\phi \in \Phi$ but it can become symmetric if we restrict to the convex functions ϕ_*, defined by $\phi_*(u) = \phi(u) + u\phi(1/u)$, for $\phi \in \Phi$ (cf. [51,90] p. 14 p. 23 Theorem 4). This measure does not obey the triangular inequality, in general, while a discussion about this property and its satisfaction by some measures of divergence is provided in Liese and Vajda [52], Vajda [91]. Several well known in the literature divergences can be obtained from $\mathcal{D}_{\phi}(f : g)$, given in (8) above, for specific choices of the convex function $\phi \in \Phi$. We mention only the Kullback–Leibler divergence (6), which is obtained from (8) for $\phi(u) = u \ln u$ or $\phi(u) = u \ln u + u - 1, u > 0$ (see [49] or [48]) and the Cressie and Read λ-power divergence or the I_{α}-divergence of Liese and Vajda [51], obtained from (8) for $\phi(u) = \phi_{\lambda}(u) = (u^{\lambda+1} - u - \lambda(u - 1))/\lambda(\lambda + 1), \lambda \neq 0, -1, u > 0$ (see [27,51,74])

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

and defined by

$$\mathcal{D}_\lambda(f:g)=\frac{1}{\lambda(\lambda+1)}\left(\int_\mathcal{X} g(x)\left(\frac{f(x)}{g(x)}\right)^{\lambda+1}d\mu-1\right),\quad -\infty<\lambda<+\infty,\ \lambda\neq 0,-1. \tag{11}$$

This measure is also related to $\mathcal{D}_0$ by means of the limiting behavior $\lim_{\lambda\to 0}\mathcal{D}_\lambda(f:g)=\mathcal{D}_0(f:g)$ and $\lim_{\lambda\to -1}\mathcal{D}_\lambda(f:g)=\mathcal{D}_0(g:f)$. Cressie and Read λ-power divergence $\mathcal{D}_\lambda(f:g)$ is closely related to Rényi's divergence $\mathcal{D}_{R,\alpha}(f:g)$, in the sense

$$\mathcal{D}_{R,\alpha}(f:g)=\frac{1}{\alpha-1}\ln\left[\alpha(\alpha-1)\mathcal{D}_{\alpha-1}(f:g)+1\right],$$

in view of (7) and (11). It is easy to see that Rényi's divergence is not included in the family of ϕ-divergence. This point led Menéndez *et al.* [56] to define the (h,ϕ)-divergence which unified all the existing divergence measures. Based on Pardo [65] p. 8, the (h,ϕ)-divergence is defined as follows:

$$\mathcal{D}^h_\phi(f:g)=h\left(\int_\mathcal{X} g(x)\phi\left(\frac{f(x)}{g(x)}\right)d\mu\right),$$

where h is a differentiable increasing real function mapping from $[0,\phi(0)+\lim_{t\to\infty}(\phi(0)/t)]$ onto $[0,\infty)$. Special choices of the functions h and ϕ lead to particular divergences, like Rényi's, Sharma-Mittal, Bhattacharyya, and they are tabulated in p. 8 of Pardo [65].

We will close this short exposition on measures of divergence between probability distributions with a presentation of the density power divergence introduced by Basu *et al.* [13], in order to develop and study robust estimation procedures on the basis of this new family of divergences. For two Radon–Nikodym derivatives f and g, the density power divergence (DPD) between f and g was defined in Basu *et al.* [13], cf. also Basu *et al.* [14], by

$$d_a(f:g)=\int_\mathcal{X}\left\{g(x)^{1+a}-\left(1+\frac{1}{a}\right)g(x)^a f(x)+\frac{1}{a}f(x)^{1+a}\right\}d\mu, \tag{12}$$

for $a>0$, while for $a=0$, it is defined by

$$\lim_{a\to 0} d_a(f:g)=\mathcal{D}_0(f:g).$$

For $a=1$, (12) reduces to the L_2 distance $L_2(f,g)=\int_\mathcal{X}(f(x)-g(x))^2d\mu$. It is also interesting to note that (12) is a special case of the so-called Bregman divergence

$$\int_\mathcal{X}\left[T(f(x))-T(g(x))-\{f(x)-g(x)\}T'(g(x))\right]d\mu.$$

If we consider $T(l)=l^{1+a}$, we get a times $d_a(f:g)$. The density power divergence depends on the tuning parameter a which controls the trade off between robustness and asymptotic efficiency of the parameter estimates which are the minimizers of this family of divergences (cf. [14] p. 297). Based on Theorem 9.1 of this book, $d_a(f:g)$ represents a genuine statistical distance for all $a\geq 0$, that is, $d_a(f:g)\geq 0$ with equality, if and only if, $f(x)=g(x)$, a.e. x. The proof of this result is provided in Basu *et al.* [14] p. 301 for the case $a>0$. The case $a=0$ follows from a similar property which is proved and obeyed by $\mathcal{D}_0(f:g)=\int_\mathcal{X} f(x)\ln(f(x)/g(x))d\mu$ (cf. [48] Thm. 3.1 p. 14). For more details about this family of divergence measures, we refer to Basu *et al.* [14].

Closing this review on divergences, interesting generalized and unified classes of divergences have been recently proposed in the literature by Stummer and Vajda [86] and Broniatowski and Stummer [19] while extensions in the case of discrete, non-probability vectors with applications in insurance can be found in Sachlas and Papaioannou [76,77]. Csiszár's ϕ-divergence has been also recently extended

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

to a local setup by Avlogiaris *et al.* [8] and the respective local divergences have been used to develop statistical inference and model selection techniques in a local setting (cf. [9,10]).

The third category of measures of information is that of parametric or Fisher's type measures of information (cf. [34,62,63]). Fisher information measure is the main representative of this category of measures and it is well known from the theory of estimation and the Cramér–Rao inequality. Fisher information measure is defined by

$$\mathcal{I}_f^{Fi}(\theta) = \int_{\mathcal{X}} f(x;\theta) \left(\frac{d}{d\theta} \ln f(x;\theta) \right)^2 d\mu, \tag{13}$$

where $f(x;\theta)$ is the Radon–Nikodym derivative of a parametric family of probability measures $P_\theta \ll \mu$ on the measurable space $(\mathcal{X}, \mathcal{A})$, while the parameter $\theta \in \Theta \subseteq \mathbb{R}$. The measure defined above is a fundamental quantity in the theory of estimation, connected, in addition, with the asymptotic variance of the maximum likelihood estimator, subject to a set of suitable regularity assumptions (cf. [23] p. 311 326). Subject to the said conditions, the following representation of $\mathcal{I}_f^{Fi}(\theta)$,

$$\mathcal{I}_f^{Fi}(\theta) = -\int_{\mathcal{X}} f(x;\theta) \frac{d^2}{d\theta^2} \ln f(x;\theta) d\mu, \tag{14}$$

has a prominent position in the literature as it provides an easy way to get the expression of Fisher information measure, in some applications. From an information theoretic point of view, $\mathcal{I}_f^{Fi}$ formulates and it expresses the amount of information contained in the data about the unknown parameter θ. Several extensions of (13) have been appeared in the bibliography of the subject while this measure obeys nice information theoretic and statistical properties (cf. [62]).

Besides Fisher's information measure (13), another quantity is widely used in different areas, such as in statistics and in functional analysis (cf. [18,22,55] and references appeared therein). This measure is defined by

$$\mathcal{J}^{Fi}(f) = \int_{\mathbb{R}} h(x) \left(\frac{d}{dx} \ln h(x) \right)^2 dx, \tag{15}$$

where, without any loss of generality, h is a density with $x \in \mathbb{R}$ (cf. [85] p. 102) and $\mathcal{J}^{Fi}(f)$ coincides with (13) when $f(x;\theta) = h(x-\theta)$, that is, when the parameter θ is a location parameter, in the considered model $f(x;\theta)$, $x \in \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}$. Papaioannou and Ferentinos [64] have studied the above measure, calling it Fisher information number, and the authors provided with an alternative expression of it, $\mathcal{J}_*^{Fi}(f) = -\int_{\mathbb{R}} h(x)(d^2/dx^2) \ln h(x) dx$, as well. The above measure is not so well known in the statistics literature, however it receives the attention of researchers and it is connected with several results in statistics, in statistical physics, in signal processing (cf., e.g., [25,87,92], and references therein). The multivariate version of (15) is analogous and it also received the attention of researches nowadays. We refer to the recent work by Yao *et al.* [94] and references therein, while the multivariate version has been exploited in Zografos [95,96] for the definition of measures of multivariate dependence.

Based on Soofi [83] p. 1246, Fisher's measure of information within a second-order approximation is the discrimination information between two distributions that belong to the same parametric family and differ infinitesimally over a parameter space. More precisely, subject to the standard regularity conditions of estimation theory (cf. [23] p. 311 326), stated also on pp. 26–27 of the monograph by Kullback [48], Fisher information measure $\mathcal{I}_f^{Fi}$ is connected with Kullback–Leibler divergence $\mathcal{D}_0$, defined in (6), by the next equality derived in the monograph of Kullback [48] p. 28,

$$\lim_{\delta \to 0} \frac{1}{\delta^2} \mathcal{D}_0(f(x;\theta) : f(x;\theta+\delta)) = \mathcal{I}_f^{Fi}(\theta), \quad \theta \in \Theta, \tag{16}$$

while similar connections of Fisher information measure with other divergences, obtained from Csiszár's divergence in (8), have been derived in Ferentinos and Papaioannou [34]. The limiting relationship

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

between Kullback–Leibler divergence and Fisher information, formulated in (16), can be easily extended to the case of Csiszár's ϕ-divergence (8). In this context, it can be easily proved (cf. [78]) that

$$\lim_{\delta\to 0}\frac{1}{\delta^2}\mathcal{D}_\phi(f(x;\theta+\delta):f(x;\theta))=\frac{\phi''(1)}{2}\mathcal{I}_f^{Fi}(\theta),\quad \theta\in\Theta. \tag{17}$$

To summarize this section, it was presented above the most representative measures of statistical information theory which play an important role, the last seven decades, not only to the fields of probability and statistics but also to many other fields of science and engineering. Interesting analogs of the above measures on the basis of the cumulative function or on the basis of the survival functions occupies a significant part of the respective literature the last 17 years and this line of research work is outlined in the next section.

3. A short review on cumulative entropies and cumulative Kullback–Leibler information

To present some of the measures of cumulative entropy and cumulative divergence, suppose in this section that X is a non-negative random variable with distribution function F and respective survival function $\bar{F}(x)=1-F(x)$. Among the huge amount of extensions or analogs of Shannon entropy, defined in (1), the cumulative residual entropy is a notable and worthwhile recent analog. Rao *et al.* [73], in a pioneer paper, introduced the cumulative residual entropy with a functional similarity with Shannon's [81] omnipresent entropy measure. Rao's *et al.* [73] measure is defined by

$$\mathrm{CRE}(F)=-\int_0^{+\infty}\bar{F}(x)\ln\bar{F}(x)dx, \tag{18}$$

where $\bar{F}(x)=1-F(x)$ is the cumulative residual distribution or the survival function of a non-negative random variable X. A year later, Zografos and Nadarajah [98] provided a timely elaboration of Rao *et al.* [73] measure and they have defined the survival exponential entropies by

$$M_\alpha(F)=\left(\int_0^{+\infty}\bar{F}^\alpha(x)dx\right)^{\frac{1}{1-\alpha}},\quad \alpha>0,\ \alpha\neq 1, \tag{19}$$

where, again, $\bar{F}(x)=1-F(x)$ is the survival function of a non-negative random variable X. The quantity $M_\alpha(F)$, defined by (19), asymptotically coincides with the exponential function of the cumulative residual entropy $\mathrm{CRE}(F)$, suitably scaled in the following sense

$$\lim_{\alpha\to 1}M_\alpha(F)=\exp\left\{-\frac{\mathrm{CRE}(F)}{\int_0^{+\infty}\bar{F}(x)dx}\right\}.$$

Moreover, the logarithmic function of $M_\alpha(F)$ leads to an analogous quantity to that of Rényi entropy (2) (cf. [98]). The analogous of $\mathrm{CRE}(F)$ Tsallis' [88] measure (3) has been recently considered in the papers by Sati and Gupta [80], Calì *et al.* [21], Rajesh and Sunoj [71] and the references appeared therein. It has a similar functional form as that of $\mathcal{E}_{Ts}(f)$ in (3), given by

$$\mathrm{CRE}_{Ts,\alpha}(F)=\frac{1}{\alpha-1}\left(1-\int_0^{+\infty}\bar{F}^\alpha(x)dx\right),\quad \alpha>0,\ \alpha\neq 1,$$

while letting $\alpha\to 1$, $\lim_{\alpha\to 1}\mathrm{CRE}_{Ts,\alpha}(F)=\mathrm{CRE}(F)$. Asadi *et al.* [6] and Rajesh and Sunoj [71] introduced an alternative measure of $\mathrm{CRE}_{Ts,\alpha}(F)$, as follows,

$$CRh_\alpha(F)=\frac{1}{\alpha-1}\int_0^{+\infty}(\bar{F}(x)-\bar{F}^\alpha(x))dx,\quad \alpha>0,\ \alpha\neq 1, \tag{20}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

and letting $\alpha \to 1$, $\lim_{\alpha \to 1} CRh_\alpha(F) = \text{CRE}(F)$, defined by (18). Moreover, it is easy to see that for $\alpha = 2$, the entropy type functional $CRh_2(F)$ coincides with Gini's index, multiplied by the expected value of the random variable associated with F (cf. [6] p. 1037). Shannon and other classic measures of entropy quantify uncertainty relative to the uniform distribution, as it was mentioned previously. This is not the case for cumulative residuals entropies, like that in (18). Following the exposition in Asadi *et al.* [6] p. 1030, the so-called by them generalized entropy functional, (20), is a measure of concentration of the distribution. That is, it is non-negative and equals zero if and only if the distribution is degenerate. Moreover, strictly positive values of $\text{CRE}(F)$ in (18) does not indicate departure from the perfect concentration toward the perfect uncertainty about prediction of random outcomes from the distribution. Rao's *et al.* [73] measure is an example for making distinction between a measure of concentration and a measure of uncertainty (every measure of concentration is not necessarily a measure of uncertainty).

Some years later, Di Crescenzo and Longobardi [31] define the cumulative entropy, in analogy with the cumulative residual entropy of Rao *et al.* [73]. The cumulative entropy is defined by

$$\text{CE}(F) = -\int_0^{+\infty} F(x) \ln F(x) dx, \tag{21}$$

where F is the distribution function, associated to a non-negative random variable X. It is clear that $\text{CRE}(F) \geq 0$ and $\text{CE}(F) \geq 0$.

Chen *et al.* [24] and some years later Klein *et al.* [46] and Klein and Doll [45], in their interesting papers, have unified and extended the cumulative residual entropy (18) and the cumulative entropy (21). Based on Klein and Doll [45] p. 8, the cumulative Φ^* entropy is defined by,

$$\text{CE}_{\Phi^*}(F) = \int_{-\infty}^{+\infty} \Phi^*(F(x)) dx, \tag{22}$$

where Φ^* is a general concave entropy generating function such that $\Phi^*(u) = \varphi(1-u)$ or $\Phi^*(u) = \varphi(u)$ leads, respectively, to the cumulative residual φ entropy and the cumulative φ entropy. The entropy generating function φ is a non-negative and concave real function defined on $[0, 1]$. The above measure is analogous with Burbea and Rao's [20] ϕ-entropy $\mathcal{E}_\phi(f)$, defined in (4). It is, moreover, clear that $\text{CRE}(F)$, in (18), and $\text{CE}(F)$, in (21), are, respectively, special cases of $\text{CE}_{\Phi^*}(F)$, in (22), for $\Phi^*(u) = \varphi(1-u)$ or $\Phi^*(u) = \varphi(u)$, with $\varphi(x) = -x \ln x$, $x \in (0, 1]$. The cumulative Φ^* entropy $\text{CE}_{\Phi^*}(F)$, inspired by Klein and Doll [45], is a broad family of measures of cumulative residual entropy and cumulative entropy and special choices of the concave function φ lead to interesting particular entropies, like that appeared in Table 3 of p. 13, in Klein and Doll [45]. An interesting special case of (22) is obtained for $\Phi^*(x) = \varphi(x) = (x^\alpha - x)/(1-a)$, $x \in (0, 1]$, $\alpha > 0$, $\alpha \neq 1$, a concave function that leads the cumulative Φ^* entropy in (22) to coincide with the entropy type measure (20), above, of Asadi *et al.* [6] and the measure of equation (6) in the paper by Rajesh and Sunoj [71].

In the way that classical Shannon entropy (1) has motivated the definition of the cumulative entropy (21), in a completely similar manner, Kullback and Leibler [49] divergence (6) has motivated the definition of the cumulative Kullback–Leibler information and the cumulative residual Kullback–Leibler information, by the work of Rao [72], Baratpour and Rad [12], Park *et al.* [66] and the subsequent papers by Di Crescenzo and Longobardi [32] and Park *et al.* [67], among others. In these and other treatments, the cumulative Kullback–Leibler information and the cumulative residual Kullback–Leibler information are defined, respectively, by

$$\text{CKL}(F : G) = \int_{\mathbb{R}} F(x) \ln \left(\frac{F(x)}{G(x)} \right) dx + \int_{\mathbb{R}} [G(x) - F(x)] dx, \tag{23}$$

and

$$\text{CRKL}(F : G) = \int_{\mathbb{R}} \bar{F}(x) \ln \left(\frac{\bar{F}(x)}{\bar{G}(x)} \right) dx + \int_{\mathbb{R}} [\bar{G}(x) - \bar{F}(x)] dx, \tag{24}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

for two distribution functions F and G with respective survival functions $\bar{F}$ and $\bar{G}$. It is clear that if the random quantities X and Y, associated with F and G, are non-negative, then $\int_0^{+\infty} \bar{F}(x)dx$ and $\int_0^{+\infty} \bar{G}(x)dx$ are equal to $E(X)$ and $E(Y)$, respectively. It should be mentioned at this point that Asadi *et al.* [2] defined, in Subsection 3.2, a Kullback–Leibler type divergence function between two non-negative functions P_1 and P_2 which provides a unified representation of the measures (6), (23) and (24), with P_i, $i = 1, 2$, being probability density function, cumulative distribution function and survival function, respectively. Based on Asadi *et al.* [7], for non-negative random variables X and Y, associated with F and G, respectively,

$$\int_0^{+\infty} [G(x) - F(x)]dx = \int_0^{+\infty} [G(x) - 1 + 1 - F(x)]dx = \int_0^{+\infty} [\bar{F}(x) - \bar{G}(x)]dx = E(X) - E(Y),$$

and (23), (24) are simplified as follows,

$$\mathrm{CKL}(F : G) = \int_0^{+\infty} F(x) \ln\left(\frac{F(x)}{G(x)}\right) dx + [E(X) - E(Y)], \tag{25}$$

and

$$\mathrm{CRKL}(F : G) = \int_0^{+\infty} \bar{F}(x) \ln\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dx + [E(Y) - E(X)]. \tag{26}$$

Based on Baratpour and Rad [12] and Park *et al.* [66],

$$\mathrm{CKL}(F : G) \geq 0, \quad \mathrm{CRKL}(F : G) \geq 0 \text{ with equality if and only if } F(x) = G(x), \text{ a.e. } x. \tag{27}$$

This is an important property because (27) ensures that $\mathrm{CKL}(F : G)$ and $\mathrm{CRKL}(F : G)$ can be used, in practice, as pseudo distances between the underling probability distributions. More generally, non-negativity and identity of indiscernibles, formulated by (27), is a desirable property of each newly defined measure of divergence because it expands its horizon in applications in formulating and solving problems in statistics and probability theory, among many other potential areas. The counter-example, that follows, illustrates the necessity of the last integrals of the right-hand side of (23) and (24), so as (27) to be valid.

Example 1. *The analogs of Kullback–Leibler divergence (6), in case of cumulative and survival functions, would be* $\int_{-\infty}^{+\infty} F(x) \ln(F(x)/G(x))dx$ *or* $\int_{-\infty}^{+\infty} \bar{F}(x) \ln(\bar{F}(x)/\bar{G}(x))dx$, *respectively. Consider two exponential distributions with survival functions* $\bar{F}(x) = e^{-\lambda x}$, $x > 0$, $\lambda > 0$ *and* $\bar{G}(x) = e^{-\mu x}$, $x > 0$, $\mu > 0$. *Then, it is easy to see that,*

$$\int_0^{+\infty} \bar{F}(x) \ln \frac{\bar{F}(x)}{\bar{G}(x)} dx = \frac{\mu - \lambda}{\lambda^2}.$$

It is clear that for $\mu < \lambda$, *the second quantity, formulated in terms of the survival functions, does not obey the non-negativity property (27). Moreover, for* $\lambda = 3$ *and* $\mu = 2$, *numerical integration leads to* $\int_0^{+\infty} F(x) \ln(F(x)/G(x))dx = 0.1807$, *while for* $\lambda = 1$ *and* $\mu = 2$, *the same measure is negative,* $\int_0^{+\infty} F(x) \ln(F(x)/G(x))dx = -0.4362$. *Therefore, the analogs of Kullback–Leibler divergence (6), in case of cumulative and survival functions, do not always satisfy the non-negativity property which is essential for applications of a measure of divergence, in practice.*

Moreover, this counter-example underlines that the analog of (8), in case of cumulative and survival functions, of the form

$$\int G(x)\phi\left(\frac{F(x)}{G(x)}\right) dx \quad \text{or} \quad \int \bar{G}(x)\phi\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dx, \tag{28}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

obtained from (8) by replacing the densities by cumulative distributions and survival functions does not always lead to non-negative measures, something which is a basic prerequisite for a measure of divergence. Baratpour and Rad [12] and Park *et al.* [66] have defined Kullback–Leibler type cumulative and survival divergences, by (23) and (24), as the analog of Kullback–Leibler classic divergence (6), which should obey the non-negativity property. In this direction, they have exploited a well-known property of the logarithmic function, namely $x\ln(x/y) \geq x - y$, for $x, y > 0$, and they defined Kullback–Leibler type divergences (23) and (24) by moving the right-hand side of the logarithmic inequality to the left-hand side.

Continuing the critical review of the cumulative and survival divergences, the cumulative paired ϕ-divergence of Definition 4, p. 26 in the paper by Klein *et al.* [46], can be considered as an extension of the above divergences $\mathrm{CKL}(F : G)$ and $\mathrm{CRKL}(F : G)$, defined by (23) and (24), in a completely similar manner that the survival and cumulative entropies (18) and (21) have been unified and extended to the cumulative Φ^* entropy, given by (22). Working in this direction, Klein *et al.* [46] p. 26 of 45, in their recent paper, have defined the cumulative paired ϕ-divergence for two distributions, by generalizing the cross entropy of Chen *et al.* [24] p. 56, as follows,

$$\mathrm{CPD}_{\phi}(F : G) = \int_{-\infty}^{+\infty} \left(G(x)\phi\left(\frac{F(x)}{G(x)}\right) + \bar{G}(x)\phi\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)\right) dx, \tag{29}$$

where F and G are distribution functions and $\bar{F} = 1 - F, \bar{G} = 1 - G$ are the respective survival functions. ϕ is again a real convex function defined on $[0, \infty]$ with $\phi(0) = \phi(1) = 0$ and satisfying additional conditions, like that of the class Φ in (9) (cf. [46]). Klein *et al.* [46] have discussed several properties of $\mathrm{CPD}_{\phi}(F : G)$ and they have been presented particular measures, obtained from $\mathrm{CPD}_{\phi}(F : G)$ for special cases of the convex function ϕ. A particular case is that obtained for $\phi(u) = u\ln u, u > 0$, and the resulting cumulative paired Shannon divergence

$$\mathrm{CPD}_{S}(F : G) = \int_{-\infty}^{+\infty} \left(F(x)\ln\left(\frac{F(x)}{G(x)}\right) + \bar{F}(x)\ln\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)\right) dx. \tag{30}$$

This measure is the cross entropy, introduced and studied previously in the paper by Chen *et al.* [24] and it is also considered in the paper by Park *et al.* [67] under the terminology general cumulative Kullback–Leibler (GCKL) information. $\mathrm{CPD}_{S}(F : G)$, defined by (30), obeys the non-negativity and identity of indiscernibles property, similar to that of Eq. (27) (cf. [24]). However, there is not a rigorous proof of non-negativity of $\mathrm{CPD}_{\phi}(F : G)$, defined by (29), to the best of our knowledge. The non-negativity property is quite necessary for a measure of divergence between probability distributions as it supports and justifies the use of such a measure as a measure of quasi-distance between the respective probability distributions, and hence, this property makes up the benchmark in developing information theoretic methods in statistics. The cumulative divergences of (29) and (30) depend on both, the cumulative function and the survival function. However, this dependence on both functions may cause problems, in practice, in cases where one of the two functions is not so tractable. Exactly this notion, that is a possible inability of the above divergences to work in practice in cases of complicated cumulative or survival functions, was the motivation point in order to try to define Csiszár's type cumulative and survival ϕ-divergence in a complete analogy to the classic divergence of Csiszár, defined by (8).

4. Csiszár's ϕ-divergence type cumulative and survival measures

The main aim of the section is to introduce Csiszár's type ϕ-divergence where the cumulative distribution function and the survival function will be used in place of probability density functions in (8). To proceed in this direction, a first thought is to define a Csiszár's type ϕ-divergence that resembles (8), by replacing the densities f and g by the respective distributions F and G or the respective survival functions $\bar{F}$ and $\bar{G}$. However, such a clear reasoning does not always lead to divergences which obey, in all the cases,

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

the non-negativity property, as it was shown in the previous motivating counter-example. To overcome this problem, motivated by the above described procedure of Baratpour and Rad [12] and Park *et al.* [66] and the use by them of a classic logarithmic inequality, we will proceed to a definition of Csiszár's type cumulative and survival ϕ-divergence, as a non-negative analog of the classic one defined by (8), by suitably applying the well-known Jensen inequality (cf., e.g., [60]). This is the theme on the next proposition.

To formulate Jensen's type inequality in the framework of cumulative and survival functions, following standard arguments (cf. [16]), consider the d-dimensional Euclidean space $\mathbb{R}^d$ and denote by $\mathcal{B}^d$ the σ-algebra of Borel subsets of $\mathbb{R}^d$. For two probability measures P_X and P_Y on $(\mathbb{R}^d, \mathcal{B}^d)$ and two d-dimensional random vectors $X = (X_1, \ldots, X_d)$ and $Y = (Y_1, \ldots, Y_d)$, let F and G denote, respectively, the joint distribution functions of X and Y, defined by, $F(x_1, \ldots, x_d) = P_X(X_1 \leq x_1, \ldots, X_d \leq x_d)$ and $G(y_1, \ldots, y_d) = P_Y(Y_1 \leq y_1, \ldots, Y_d \leq y_d)$, for $(x_1, \ldots, x_d) \in \mathbb{R}^d$ and $(y_1, \ldots, y_d) \in \mathbb{R}^d$. In a similar manner, the respective multivariate survival functions are defined by $\bar{F}(x_1, \ldots, x_d) = P_X(X_1 > x_1, \ldots, X_d > x_d)$ and $\bar{G}(y_1, \ldots, y_d) = P_Y(Y_1 > y_1, \ldots, Y_d > y_d)$. Let also a convex function ϕ defined in the interval $(0, +\infty)$ and satisfying the assumptions of p. 299 of Csiszár [29] (cf. also the class Φ, defined by (9)). The next proposition formulates, in a sense, Lemma 1.1 on p. 299 of Csiszár [29] in terms of cumulative and survival functions.

Proposition 1.

(a) Let F and G are two cumulative distribution functions. Then, for $\alpha = \int_{\mathbb{R}^d} F(x)dx \,/ \int_{\mathbb{R}^d} G(x)dx$,

$$\int_{\mathbb{R}^d} G(x)\phi\left(\frac{F(x)}{G(x)}\right)dx \geq \phi(\alpha)\int_{\mathbb{R}^d} G(x)dx,$$

and the sign of equality holds if $F(x) = G(x)$, on $\mathbb{R}^d$. Moreover, if ϕ is strictly convex at $\alpha = \int_{\mathbb{R}^d} F(x)dx \,/ \int_{\mathbb{R}^d} G(x)dx$ and equality holds in the above inequality, then $F(x) = \alpha G(x)$, on $\mathbb{R}^d$.

(b) Let $\bar{F}$ and $\bar{G}$ denote two survival functions. Then, for $\bar{\alpha} = \int_{\mathbb{R}^d} \bar{F}(x)dx / \int_{\mathbb{R}^d} \bar{G}(x)dx$,

$$\int_{\mathbb{R}^d} \bar{G}(x)\phi\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)dx \geq \phi\left(\bar{\alpha}\right)\int_{\mathbb{R}^d} \bar{G}(x)dx,$$

and the sign of equality holds if $\bar{F}(x) = \bar{G}(x)$, on $\mathbb{R}^d$. Moreover, if ϕ is strictly convex at $\bar{\alpha} = \int_{\mathbb{R}^d} \bar{F}(x)dx / \int_{\mathbb{R}^d} \bar{G}(x)dx$ and equality holds in the above inequality, then $\bar{F}(x) = \bar{\alpha}\bar{G}(x)$, on $\mathbb{R}^d$.

Proof. The proof is based on the proof of the classic Jensen's inequality and it closely follows the proof of Lemma 1.1 of Csiszár [29] p. 300. It is presented here in the context of distribution and survival functions for the sake of completeness. We will present the proof of part (a) because the proof of part (b) is quite similar and it is omitted. Following Csiszár [29] p. 300, first, one may assume that $\int_{\mathbb{R}^d} F(x)dx > 0$ and $\int_{\mathbb{R}^d} G(x)dx > 0$. Otherwise, the statement is true because of the conventions which define the class Φ of the convex functions ϕ, in (9). By the convexity of ϕ, it is valid

$$\phi(u) \geq \phi(\alpha) + b(u - \alpha), \quad 0 < u < +\infty,$$

with b, $b < +\infty$, equals, for example, to the arithmetic mean of the right and left derivatives of $\phi(u)$ at the point $\alpha = \int_{\mathbb{R}^d} F(x)dx / \int_{\mathbb{R}^d} G(x)dx$. Replacing $u = F(x)/G(x)$, we obtain for $G(x) > 0$,

$$G(x)\phi\left(\frac{F(x)}{G(x)}\right) \geq G(x)\phi(\alpha) + b(F(x) - \alpha G(x)). \tag{31}$$

According to the conventions that define the class Φ, the above inequality holds even for $G(x) = 0$, because the convexity of ϕ leads to $b \leq \lim_{u\to\infty} \phi(u)/u$. Integrating both sides of (31) over $\mathbb{R}^d$, it is

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

obtained

$$\int_{\mathbb{R}^d} G(x)\phi\left(\frac{F(x)}{G(x)}\right)dx \geq \phi(\alpha)\int_{\mathbb{R}^d} G(x)dx,$$

and the first part of the assertion in (a) has been proved because of $\int_{\mathbb{R}^d}(F(x)-\alpha G(x))dx = 0$ by the definition of α.

Suppose now that ϕ is strictly convex at $\alpha = \int_{\mathbb{R}^d} F(x)dx/\int_{\mathbb{R}^d} G(x)dx$ and that $\int_{\mathbb{R}^d} G(x)\phi(F(x)/G(x))dx \geq (\int_{\mathbb{R}^d} G(x)dx)\phi(\int_{\mathbb{R}^d} F(x)dx/\int_{\mathbb{R}^d} G(x)dx)$. Taking into account that ϕ is strictly convex at $u = \alpha$, the inequality in (31) is strict, except for $F(x) = \alpha G(x)$. □

The above result provides with lower bounds for the integrals (28), which constitute straightforward analogs of Csiszár's ϕ-divergence, defined by (8). The said lower bounds, if they will be moved on the left-hand side of the inequalities of the previous proposition, can be exploited in order to define, in the sequel, non-negative Csiszár's type ϕ-divergences by means of cumulative distribution functions and survival functions.

Definition 1. *Let the cumulative distribution functions F and G. The cumulative Csiszár's type ϕ-divergence between F and G is defined by*

$$\mathcal{CD}_\phi(F:G) = \int_{\mathbb{R}^d} G(x)\phi\left(\frac{F(x)}{G(x)}\right)dx - \left(\int_{\mathbb{R}^d} G(x)dx\right)\phi\left(\frac{\int_{\mathbb{R}^d} F(x)dx}{\int_{\mathbb{R}^d} G(x)dx}\right), \tag{32}$$

where $\phi : (0,\infty) \to \mathbb{R}$ is a real-valued convex function and $\phi \in \Phi$, defined by (9).

Definition 2. *Let the survival functions $\bar{F}$ and $\bar{G}$. The survival Csiszár's type ϕ-divergence between $\bar{F}$ and $\bar{G}$ is defined by*

$$\mathcal{SD}_\phi(\bar{F}:\bar{G}) = \int_{\mathbb{R}^d} \bar{G}(x)\phi\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)dx - \left(\int_{\mathbb{R}^d} \bar{G}(x)dx\right)\phi\left(\frac{\int_{\mathbb{R}^d} \bar{F}(x)dx}{\int_{\mathbb{R}^d} \bar{G}(x)dx}\right), \tag{33}$$

where $\phi : (0,\infty) \to \mathbb{R}$ is a real-valued convex function and $\phi \in \Phi$, defined by (9).

However, the main aim is the definition of Csiszár's type ϕ-divergences on the basis of distribution and survival functions, which will obey the non-negativity and the identity of indiscernibles property, a property which will support applications of the proposed measures as quasi-distances between distributions. The quantities $\mathcal{CD}_\phi(F,G)$ and $\mathcal{SD}_\phi(\bar{F},\bar{G})$, defined above, are non-negative in view of the previous proposition. It remains to prove the identity of indiscernibles property which is the theme of the next proposition.

Proposition 2. *The measures $\mathcal{CD}_\phi(F,G)$ and $\mathcal{SD}_\phi(\bar{F},\bar{G})$, defined by (32) and (33), obey the non-negativity and the identity of indiscernibles property, that is,*

$$\mathcal{CD}_\phi(F:G) \geq 0 \text{ with equality if and only if } F(x) = G(x), \quad \text{on } \mathbb{R}^d,$$
$$\mathcal{SD}_\phi(\bar{F}:\bar{G}) \geq 0 \text{ with equality if and only if } \bar{F}(x) = \bar{G}(x), \quad \text{on } \mathbb{R}^d,$$

for the convex function ϕ being strictly convex at the points α and $\bar{\alpha}$ of the previous proposition.

Proof. If $F(x) = G(x)$, on $\mathbb{R}^d$, then the assertion follows from the fact that the convex function ϕ belongs to the class Φ and therefore $\phi(1) = 0$. For functions ϕ which are strictly convex at α and $\bar{\alpha}$, if the sign of equality holds in the inequalities of parts (a) and (b) of the previous proposition, then $F(x) = \alpha G(x)$, on $\mathbb{R}^d$ and $\bar{F}(x) = \bar{\alpha}\bar{G}(x)$, on $\mathbb{R}^d$. Given that F and G are cumulative distribution

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

functions and based on Billingsley [16] p. 260, $F(x) \to 1$ and $G(x) \to 1$, if $x_i \to +\infty$ for each i and $F(x) \to 0$, $G(x) \to 0$, if $x_i \to -\infty$ for some i (the other coordinates held fixed). Moreover, taking into account that the multivariate survival functions $\bar{F}$ and $\bar{G}$ are functionally related with the corresponding cumulative functions F and G (cf. [42] p. 27), we conclude that $\bar{F}(x) \to 1$, $\bar{G}(x) \to 1$, if $x_i \to -\infty$ for each i. All these relationships between cumulative and survival functions lead to the conclusion that $\alpha = \bar{\alpha} = 1$ and then they are coincide, that is, $F(x) = G(x)$ and $\bar{F}(x) = \bar{G}(x)$, on $\mathbb{R}^d$. Therefore, the lower bounds, derived in the proposition, are attained if and only if the underlined cumulative and survival functions coincide for the convex function ϕ being strictly convex at the points α and $\bar{\alpha}$. This completes the proof of the proposition. □

In the sequel, the interest is focused in product measures, obtained from (32) and (33) for particular choices of the convex function ϕ.

4.1. Kullback–Leibler type cumulative and survival divergences and mutual information

At a first glance, if Csiszar's type cumulative and survival divergences, defined by (32) and (33) above, will be applied for the convex function $\phi(u) = u \log u$, $u > 0$, or $\phi(u) = u \log u + u - 1$, $u > 0$, then they do not lead to the respective Kullback–Leibler divergences, defined by (23) and (24). An application of (32) and (33) for $\phi(u) = u \ln u$, $u > 0$, leads to the measures

$$\begin{aligned} C\mathcal{D}_{\mathrm{KL}}(F : G) &= \int_{\mathbb{R}^d} F(x) \ln\left(\frac{F(x)}{G(x)}\right) dx - \left(\int_{\mathbb{R}^d} F(x)dx\right) \ln\left(\int_{\mathbb{R}^d} F(x)dx \Big/ \int_{\mathbb{R}^d} G(x)dx\right), \\ S\mathcal{D}_{\mathrm{KL}}(\bar{F} : \bar{G}) &= \int_{\mathbb{R}^d} \bar{F}(x) \ln\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dx - \left(\int_{\mathbb{R}^d} \bar{F}(x)dx\right) \ln\left(\int_{\mathbb{R}^d} \bar{F}(x)dx \Big/ \int_{\mathbb{R}^d} \bar{G}(x)dx\right), \end{aligned} \tag{34}$$

respectively. Based on the elementary logarithmic inequality, $x \ln(x/y) \geq x - y$, for $x, y > 0$, and on Eqs. (23) and (24) it is immediate to see that

$$\begin{aligned} C\mathcal{D}_{\mathrm{KL}}(F : G) &\leq \int_{\mathbb{R}^d} F(x) \ln\left(\frac{F(x)}{G(x)}\right) dx + \int_{\mathbb{R}^d} [G(x) - F(x)]dx = \mathrm{CKL}(F : G), \\ S\mathcal{D}_{\mathrm{KL}}(\bar{F} : \bar{G}) &\leq \int_{\mathbb{R}^d} \bar{F}(x) \ln\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dx + \int_{\mathbb{R}^d} [\bar{G}(x) - \bar{F}(x)]dx = \mathrm{CRKL}(F : G), \end{aligned} \tag{35}$$

where $\mathrm{CKL}(F : G)$ and $\mathrm{CRKL}(F : G)$ are the measures (23) and (24) defined by Rao [72], Baratpour and Rad [12], Park *et al.* [66], among others. It is clear, in view of (35), that the measures $\mathrm{CKL}(F : G)$ and $\mathrm{CRKL}(F : G)$ over evaluate the divergence or the quasi-distance between the distribution of two random variables, as it is formulated and expressed by the respective Kullback–Leibler type cumulative distribution functions or the survival functions, defined by (34).

Csiszár's type and Kullback–Leibler's type survival divergences can be expressed in terms of expected values if we restrict ourselves to the univariate case, $d = 1$. Indeed, if we focus again on non-negative random variables X and Y with respective survival functions $\bar{F}$ and $\bar{G}$, then $S\mathcal{D}_\phi(\bar{F}, \bar{G})$ of (33) is formulated as follows:

$$S\mathcal{D}_\phi(\bar{F} : \bar{G}) = \int_0^{+\infty} \bar{G}(x) \phi\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dx - (EY)\, \phi\left(\frac{EX}{EY}\right), \tag{36}$$

and for the special choice $\phi(u) = u \log u, u > 0$, (34) leads to

$$S\mathcal{D}_{\mathrm{KL}}(\bar{F} : \bar{G}) = \int_0^{+\infty} \bar{F}(x) \ln\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right) dx - (EX) \ln\left(\frac{EX}{EY}\right). \tag{37}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

It should be noted at this point that Asadi *et al.* [6], in their Lemma 2.1, formulated a general divergence measure by moving the right-hand side of their inequality to the left-hand side. The defined by Asadi *et al.* [6] measure includes the divergence in (37) as a limiting case.

The survival analogs of Csiszár's and Kullback–Leibler's divergences can be expressed in terms of expected values, in view of (36) or (37). The implication of this point is shown in the next example.

Example 2. *Park et al. [66] considered the standard exponential distribution* $\bar{F}(x) = e^{-x}, x > 0$ *and the Weibull distribution* $\bar{G}(x) = e^{-x^k}$, $x > 0$, $k > 0$, *with scale parameter* 1 *and shape parameter* k. *It is well known that the mean of these distributions exist and they are equal to* $E(X) = 1$ *and* $E(Y) = \Gamma(1+1/k)$, *where* Γ *denotes the complete gamma function. In this context, based on (25), Park's et al. [66] cumulative Kullback–Leibler information (*CKL*) between* F *and* G *can be easily obtained because* $E(X) - E(Y) = 1 - \Gamma(1 + 1/k)$, *while the integral* $\int_0^{+\infty} F \ln(F/G)dx$ *can be numerically evaluated for specific values of the shape parameter* $k > 0$. *On the other hand, elementary algebraic manipulations lead that*

$$\int_0^{+\infty} \bar{F}(x) \ln \frac{\bar{F}(x)}{\bar{G}(x)} dx = -E(X) + E(X^k) = -1 + k! = -1 + \Gamma(k+1),$$

by taking into account that the simple moment of order k *of the standard exponential distribution is* $E(X^k) = k! = \Gamma(k+1)$. *Therefore, based on (26), Park's et al. [66] cumulative residual KL information (*CRKL*) between* $\bar{F}$ *and* $\bar{G}$ *is given by*

$$\text{CRKL}(F : G) = -1 + k! + \Gamma\left(1 + \frac{1}{k}\right) - 1 = \Gamma(k+1) + \Gamma\left(1 + \frac{1}{k}\right) - 2.$$

Let's now derive the measures $\mathcal{CD}_{\text{KL}}(F : G)$ *and* $\mathcal{SD}_{\text{KL}}(\bar{F} : \bar{G})$, *formulated by (34) or (37) for non-negative random variables, as is the case of the random variables* X *and* Y, *above. It is easy to see that the integrals* $\int_0^{+\infty} F(x)dx$ *and* $\int_0^{+\infty} G(x)dx$ *do not convergence, and therefore,* $\mathcal{CD}_{\text{KL}}(F : G)$ *in (34) is not defined for this specific choice of* F *and* G. *On the other hand,* $\mathcal{SD}_{\text{KL}}(\bar{F} : \bar{G})$ *is derived in a explicit form by using (37) and it is given by*

$$\mathcal{SD}_{\text{KL}}(\bar{F} : \bar{G}) = \int_0^{+\infty} \bar{F}(x) \ln \frac{\bar{F}(x)}{\bar{G}(x)} dx - (EX) \ln \frac{EX}{EY} = \int_0^{+\infty} \bar{F}(x) \ln \frac{\bar{F}(x)}{\bar{G}(x)} dx + \ln(EY),$$

or

$$\mathcal{SD}_{\text{KL}}(\bar{F}, \bar{G}) = -1 + \Gamma(k+1) + \ln\Gamma\left(1 + \frac{1}{k}\right).$$

The classic Kullback–Leibler divergence between the standard exponential distribution with density function $f(x) = e^{-x}, x > 0$ *and the Weibull distribution with scale parameter equal to* 1 *and density function* $g(x) = kx^{k-1}e^{-x^k}, x > 0, k > 0$, *is defined by*

$$\mathcal{D}_0(f : g) = \int_0^{+\infty} f(x) \ln \frac{f(x)}{g(x)} dx.$$

Simple algebraic manipulations lead to

$$\begin{aligned}
\int_0^{+\infty} f(x) \ln f(x) dx &= -\int_0^{+\infty} x f(x) dx = -E(X) = -1, \\
\int_0^{+\infty} f(x) \ln g(x) dx &= \int_0^{+\infty} e^{-x}(\ln k + (k-1)\ln x - x^k) dx \\
&= \ln k + (k-1)E_f(\ln X) - E_f(X^k).
\end{aligned}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

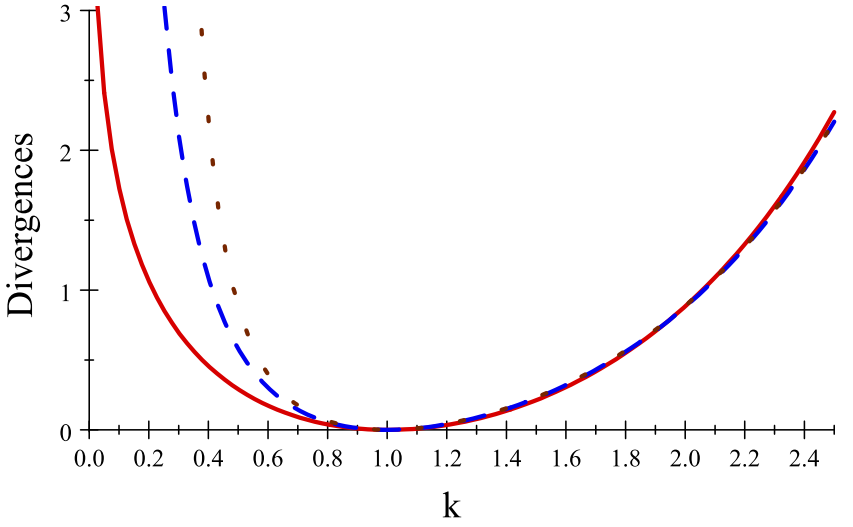

Figure 1. *Plot of divergences* $\mathcal{D}_0$ *(red-solid),* CRKL *(brown-dots) and* $\mathcal{SD}_{\mathrm{KL}}$ *(blue-dash).*

Taking into account that $E_f(X^k) = k! = \Gamma(k+1)$ *and* $\int_0^{+\infty} e^{-x}(\ln x)dx = -0.57722$,

$$\int_0^{+\infty} f(x)\ln g(x)dx = \ln k - 0.577\,22(k-1) - \Gamma(k+1),$$

and hence

$$\mathcal{D}_0(f,g) = -1 - \ln k + 0.577\,22(k-1) + \Gamma(k+1).$$

Figure 1 shows the plot of $\mathcal{D}_0(f : g)$ *(red-solid),* $\mathrm{CRKL}(F : G)$ *(brown-dots) and* $\mathcal{SD}_{\mathrm{KL}}(\bar{F} : \bar{G})$ *(blue-dash).*

We observe from this figure that all the considered divergences attain their minimum value 0 *at* $k = 1$ *because in this case the standard exponential model coincides with the Weibull model with scale parameter and shape parameter equal to one. For values of* k *greater than* 1*, all the measures almost coincide. The plots are also in harmony with inequality (35).*

Mutual information is closely related and it is obtained from Kullback–Leibler divergence, defined by (6). It has its origins, to the best of our knowledge, in a paper by Linfoot [54] and it has received a great attention in the literature as it has been connected with a huge literature on topics of statistical dependence. It has been used for the definition of measures of dependence, which have been broadly applied to develop tests of independence (cf., [17,26,33,38,57], among many others). Mutual information is, in essence, the Kullback–Leibler divergence $\mathcal{D}_0$, in (6), between the joint distribution of d random variables and the distribution of these random variables subject to the assumption of their independence.

In this context, consider the d-dimensional Euclidean space $\mathbb{R}^d$ and denote by $\mathcal{B}^d$ the σ-algebra of Borel subsets of $\mathbb{R}^d$. For a probability measure P_X on $(\mathbb{R}^d, \mathcal{B}^d)$ and a d-dimensional random vector $X = (X_1, \ldots, X_d)$, let F_X be the joint distribution function of X, defined by, $F_X(x_1, \ldots, x_d) = P_X(X_1 \leq x_1, \ldots, X_d \leq x_d)$. Let now denote by P_X^0 the probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$ under the assumption of independence of the components X_i of the random vector $X = (X_1, \ldots, X_d)$, that is P_X^0 is product measure $P_X^0 = P_{X_1} \times \ldots \times P_{X_d}$, where P_{X_i} are probability measures on $(\mathbb{R}, \mathcal{B})$ and $F_{X_i}(x_i) = P_{X_i}(X_i \leq x_i)$, $x_i \in \mathbb{R}$, is the marginal distribution function of X_i, $i = 1, \ldots, d$. In this setting, the joint distribution of $X = (X_1, \ldots, X_d)$, under the assumption of independence, is defined by $F_X^0(x_1, \ldots, x_d) = \prod_{i=1}^d F_{X_i}(x_i)$, for $(x_1, \ldots, x_d) \in \mathbb{R}^d$. If f_X and f_X^0 are the respective joint densities of $X = (X_1, \ldots, X_d)$, then the

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

classic mutual information is defined by

$$\mathcal{MI}(X) = \mathcal{D}_0(f_X : f_X^0) = \int_{\mathbb{R}^d} f_X(x) \ln \frac{f_X(x)}{f_X^0(x)} dx = \int_{\mathbb{R}^d} f_X(x) \ln \frac{f_X(x)}{f_{X_1}(x_1) \dots f_{X_d}(x_d)} dx. \tag{38}$$

The measure (38) satisfies the non-negativity and identity of indiscernibles property, $\mathcal{MI}(X) \geq 0$, with equality if and only if $f_X(x) = \prod_{i=1}^d f_{X_i}(x_i)$, that is, if and only if $X_1, \dots, X_d$ are independent. Hence, the above measure is ideal to formulate the degree of stochastic dependence between the components of $X = (X_1, \dots, X_d)$ and to serve, therefore, as a measure of stochastic dependence. An empirical version of (38) can be also used as a test statistic in testing independence of the components of $X = (X_1, \dots, X_d)$.

Mutual information can be defined in terms of cumulative and survival functions by using $\mathcal{CD}_{\mathrm{KL}}$ and $\mathcal{SD}_{\mathrm{KL}}$ of (34). Then, the cumulative mutual information and the survival mutual information are defined by,

$$\begin{aligned} \mathcal{CMI}(X) &= \int_{\mathbb{R}^d} F_X(x) \ln \left(\frac{F_X(x)}{F_X^0(x)} \right) dx - \left(\int_{\mathbb{R}^d} F_X(x) dx \right) \ln \left(\int_{\mathbb{R}^d} F_X(x) dx \Big/ \int_{\mathbb{R}^d} F_X^0(x) dx \right), \\ \mathcal{SMI}(X) &= \int_{\mathbb{R}^d} \bar{F}_X(x) \ln \left(\frac{\bar{F}_X(x)}{\bar{F}_X^0(x)} \right) dx - \left(\int_{\mathbb{R}^d} \bar{F}(x) dx \right) \ln \left(\int_{\mathbb{R}^d} \bar{F}(x) dx \Big/ \int_{\mathbb{R}^d} \bar{F}_X^0(x) dx \right), \end{aligned} \tag{39}$$

where F_{X_i} is the marginal distribution function of X_i, $i = 1, \dots, d$, while $F_X^0(x) = \prod_{i=1}^d F_{X_i}(x_i)$, $\bar{F}_X^0(x) = \prod_{i=1}^d [1 - F_{X_i}(x_i)]$, are used to denote the cumulative and the respective survival function under the assumption of independence of the components of $X = (X_1, \dots, X_d)$. It is immediate to see that the cumulative and survival mutual information $\mathcal{CMI}(X)$ and $\mathcal{SMI}(X)$, of (39), are non-negative and they attain their minimum value 0 if and only if $F_X(x) = F_X^0(x) = \prod_{i=1}^d F_{X_i}(x_i)$ and $\bar{F}_X(x) = \bar{F}_X^0(x) = \prod_{i=1}^d [1 - F_{X_i}(x_i)]$. Hence, $\mathcal{CMI}(X)$ and $\mathcal{SMI}(X)$ express on how close is $F_X(x)$ with $\prod_{i=1}^d F_{X_i}(x_i)$ and $\bar{F}_X(x)$ with $\prod_{i=1}^d [1 - F_{X_i}(x_i)] = \prod_{i=1}^d \bar{F}_{X_i}(x_i)$, respectively. Based, moreover, on the fact that equality $F_X(x) = F_X^0(x) = \prod_{i=1}^d F_{X_i}(x_i)$ is equivalent to independence of $X_1, \dots, X_d$, the cumulative mutual information $\mathcal{CMI}(X)$ can be also used as a measure of dependence and its empirical version can be also serve as an index to develop tests of independence. The same is true for the measure $\mathcal{SMI}(X)$ in the bivariate, $d = 2$, case. Cumulative and survival mutual information $\mathcal{CMI}(X)$ and $\mathcal{SMI}(X)$ can be generalized, by using (32) and (33), to produce Csiszár's type cumulative and survival mutual ϕ-divergences in a way, similar to that of Micheas and Zografos [11] who have extended the classic mutual information (38) to the classic Csiszár's ϕ-divergence between the joint distribution and the similar one under the assumption of independence.

Next example presents the measures $\mathcal{CMI}(X)$ and $\mathcal{SMI}(X)$ on the basis of a well-known bivariate distribution, the Farlie-Gumbel-Morgenstern (FGM) bivariate distribution (cf. [11], and references therein).

Example 3. *Let the FGM bivariate distribution of a random vector* (X_1, X_2), *with the joint cumulative distribution function,*

$$F_{X_1,X_2}(x_1, x_2) = x_1 x_2 [1 + \theta(1 - x_1)(1 - x_2)], \quad 0 < x_1, x_2 < 1, \ -1 \leq \theta \leq 1,$$

and the joint probability density function,

$$f_{X_1,X_2}(x_1, x_2) = 1 + \theta(1 - 2x_1)(1 - 2x_2), \quad 0 < x_1, x_2 < 1, \ -1 \leq \theta \leq 1.$$

The marginal distributions are uniform $X_1 \sim U(0, 1)$ *and* $X_2 \sim U(0, 1)$ *and the correlation coefficient is* $\rho = \rho(X_1, X_2) = \theta/3$, *which clearly ranges from* $-\frac{1}{3}$ *to* $\frac{1}{3}$. *For FGM family of bivariate distributions,*

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

it can be easily seen that the last term of the right-hand side of (39) is obtained in an analytic form and it is given by,

$$\left(\int_0^1\int_0^1 F_{X_1,X_2}(x_1,x_2)dx_1dx_2\right)\ln\frac{\int_0^1\int_0^1 F_{X_1,X_2}(x_1,x_2)dx_1dx_2}{\int_0^1\int_0^1 F_{X_1}(x_1)F_{X_2}(x_2)dx_1dx_2}=\left(\frac{1}{4}+\frac{\theta}{36}\right)\ln\left(1+\frac{\theta}{9}\right).$$

The first term of the right-hand side of (39), $\int_0^1\int_0^1 F_{X_1,X_2}\ln(F_{X_1,X_2}/F_{X_1}F_{X_2})dx_1dx_2$*, and the classic mutual information (38) can be numerically evaluated for several values of the dependence parameter* θ*,* $-1\le\theta\le 1$*. Moreover, based on Nelsen [59] p. 32 or Joe [42] pp. 27–28,*

$$\bar{F}_{X_1,X_2}(x_1,x_2)=1-F_{X_1}(x_1)-F_{X_2}(x_2)+F_{X_1,X_2}(x_1,x_2),$$

and therefore, the survival function of the FGM family of bivariate distributions is given by,

$$\bar{F}_{X_1,X_2}(x_1,x_2)=(1-x_1-x_2+x_1x_2)(1+\theta x_1x_2),\ 0<x_1,x_2<1\ \textit{and}\ -1\le\theta\le 1,$$

while the survival function $\bar{F}^0_{X_1,X_2}(x_1,x_2)$*, under the assumption of independence of* X_1*,* X_2 *is* $\bar{F}^0_{X_1,X_2}(x_1,x_2)=(1-x_1)(1-x_2)$*. The table presents the values of the correlation coefficient* $\rho(X_1,X_2)$*, the mutual information* $\mathcal{MI}(X_1,X_2)$*, the cumulative and the survival mutual information* $\mathcal{CMI}(X_1,X_2)$ *and* $\mathcal{SMI}(X_1,X_2)$*, for several values of the dependence parameter* θ*.*

	ρ, $\mathcal{MI}$, $\mathcal{CMI}$ and $\mathcal{SMI}$ for values of the dependence parameter θ							
θ	-1	-0.75	-0.5	-0.25	0.25	0.5	0.75	1
ρ	-0.33333	-0.25	-0.16667	-0.083333	-0.083333	0.16667	0.25	0.33333
$\mathcal{MI}$	0.06	0.032455	0.014109	0.0034855	0.0034855	0.014109	0.032455	0.06
$\mathcal{CMI}$	0.002426	0.001274	0.000534	0.00012663	0.0001153	0.000442	0.000957	0.001638
$\mathcal{SMI}$	0.002426	0.001277	0.000537	0.0001228	0.0001143	0.000441	0.000955	0.001641

Observe that all the measures decrease and they approach their minimum value as the value of the dependence parameter θ *approaches independence* ($\theta=0$)*. The correlation coefficient* ρ *and the mutual information* $\mathcal{MI}$ *are symmetric, something which is not obeyed by the cumulative mutual information* $\mathcal{CMI}$ *and the survival mutual information* $\mathcal{SMI}$*. The correlation coefficient captures negative dependence but the other informational measures do not discriminate between positive and negative dependence. Last, it is interesting to note that the quantity* $\int_0^1\int_0^1 F_{X_1,X_2}\ln(F_{X_1,X_2}/F_{X_1}F_{X_2})dx_1dx_2$ *can take positive or negative values, for instance, it is equal to* -0.00672 *for* $\theta=-0.25$*, or it is equal to* 0.01471 *for* $\theta=0.5$*. The same is also true for the quantity* $\int_0^1\int_0^1 \bar{F}_{X_1,X_2}\ln(\bar{F}_{X_1,X_2}/\bar{F}_{X_1}\bar{F}_{X_2})dx_1dx_2$ *which is equal to* -0.01866 *for* $\theta=-0.75$ *and equal to* 0.01471 *for* $\theta=0.5$*. This point justifies the definition of the cumulative and survival mutual information by (39) which ensure non-negativity of the measures.*

4.2. Cressie and Read type cumulative and survival divergences

Let now consider the convex function $\phi(u)=\phi_\lambda(u)=(u^{\lambda+1}-u-\lambda(u-1))/\lambda(\lambda+1)$, $\lambda\neq 0,-1$, $u>0$, which leads Csiszár's ϕ-divergence, defined by (8), to Cressie and Read [27] and Read and Cressie [74] power divergence. A straight application of (32) and (33) for this specific choice of the convex function ϕ leads to the Cressie and Read type cumulative and survival divergences, which are

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

defined as follows:

$$\mathcal{CD}_\lambda(F:G)=\frac{1}{\lambda(\lambda+1)}\left(\int_{\mathbb{R}^d}G(x)\left(\frac{F(x)}{G(x)}\right)^{\lambda+1}dx-\left(\int_{\mathbb{R}^d}G(x)dx\right)\right.$$
$$\left.\times\left(\int_{\mathbb{R}^d}F(x)dx\,\Big/\int_{\mathbb{R}^d}G(x)dx\right)^{\lambda+1}\right), \tag{40}$$

and

$$\mathcal{SD}_\lambda(\bar{F}:\bar{G})=\frac{1}{\lambda(\lambda+1)}\left(\int_{\mathbb{R}^d}\bar{G}(x)\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)^{\lambda+1}dx-\left(\int_{\mathbb{R}^d}\bar{G}(x)dx\right)\right.$$
$$\left.\times\left(\int_{\mathbb{R}^d}\bar{F}(x)dx\,\Big/\int_{\mathbb{R}^d}\bar{G}(x)dx\right)^{\lambda+1}\right),$$

for $\lambda\in\mathbb{R},\lambda\neq 0,-1$. The last measures can be formulated in terms of expected values if we concentrate on non-negative random variables X and Y with respective survival functions $\bar{F}$ and $\bar{G}$. In this frame, $\mathcal{SD}_\lambda(\bar{F},\bar{G})$ is simplified as follows,

$$\mathcal{SD}_\lambda(\bar{F}:\bar{G})=\frac{1}{\lambda(\lambda+1)}\left(\int_0^{+\infty}\bar{G}(x)\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)^{\lambda+1}dx-E(Y)\left(\frac{E(X)}{E(Y)}\right)^{\lambda+1}\right),\quad \lambda\in\mathbb{R},\ \lambda\neq 0,-1.$$

The previous propositions supply $\mathcal{CD}_\lambda(F,G)$ and $\mathcal{SD}_\lambda(\bar{F},\bar{G})$ with non-negativity and the identity of indiscernibles property. Cressie and Read's [27] type cumulative and survival divergences (40) are not defined for $\lambda=-1$ and $\lambda=0$. When the power λ approaches these values, then $\mathcal{CD}_\lambda(F,G)$ and $\mathcal{SD}_\lambda(\bar{F},\bar{G})$ are reduced to the respective Kullback–Leibler divergences, in the limiting sense that follows and can be easily proved,

$$\lim_{\lambda\to 0}\mathcal{CD}_\lambda(F:G)=\mathcal{CD}_{\mathrm{KL}}(F:G)\quad\text{and}\quad\lim_{\lambda\to 0}\mathcal{SD}_\lambda(\bar{F}:\bar{G})=\mathcal{SD}_{\mathrm{KL}}(\bar{F}:\bar{G}),$$
$$\lim_{\lambda\to -1}\mathcal{CD}_\lambda(F:G)=\mathcal{CD}_{\mathrm{KL}}(G:F)\quad\text{and}\quad\lim_{\lambda\to -1}\mathcal{SD}_\lambda(\bar{F}:\bar{G})=\mathcal{SD}_{\mathrm{KL}}(\bar{G}:\bar{F}).$$

Example 4. *Let consider again the random variables X and Y of the previous example with distribution functions and survival functions* $F(x)=1-e^{-x}$, $\bar{F}(x)=e^{-x}$ *and* $G(x)=1-e^{-x^k}$, $\bar{G}(x)=e^{-x^k}$, $x>0,k>0$. *It is easy to see that* $\mathcal{CD}_\lambda(F:G)$ *is not obtained in an explicit form for this specific choice of F and G. In respect to* $\mathcal{SD}_\lambda(\bar{F}:\bar{G})$, *elementary algebraic manipulations entail that*

$$\int_0^{+\infty}\bar{G}(x)\left(\frac{\bar{F}(x)}{\bar{G}(x)}\right)^{\lambda+1}dx=\int_0^{+\infty}e^{-x^k}\left(\frac{e^{-x}}{e^{-x^k}}\right)^{\lambda+1}dx=\int_0^{+\infty}e^{-(\lambda+1)x+\lambda x^k}dx,$$

and the last integral can be numerically evaluated for specific values of the power λ *and the shape parameter k. Taking into account that* $E(X)=1$ *and* $E(Y)=\Gamma(1+(1/k))$, *the Cressie and Read type survival divergence is given by,*

$$\mathcal{SD}_\lambda(\bar{F}:\bar{G})=\frac{1}{\lambda(\lambda+1)}\left(\int_0^{+\infty}e^{-(\lambda+1)x+\lambda x^k}dx-\Gamma^{-\lambda}\left(1+\frac{1}{k}\right)\right),\quad \lambda\in\mathbb{R},\ \lambda\neq 0,-1,\ k>0.$$

4.3. Density power divergence type cumulative and survival divergences

A straightforward application of d_a, given by (12), leads to the cumulative and survival counterparts of d_a, which are defined in the sequel. Let F and G denote the cumulative distribution functions of the

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

random vectors X and Y and $\bar{F}$ and $\bar{G}$ denote the respective survival functions. Then, the cumulative and survival density power type divergences, are defined by,

$$\mathcal{C}d_a(F:G) = \int_{\mathbb{R}^d} \left\{ G(x)^{1+a} - \left(1+\frac{1}{a}\right) G(x)^a F(x) + \frac{1}{a}\, F(x)^{1+a} \right\} dx, \quad a > 0, \tag{41}$$

and

$$\mathcal{S}d_a(\bar{F}:\bar{G}) = \int_{\mathbb{R}^d} \left\{ \bar{G}(x)^{1+a} - \left(1+\frac{1}{a}\right) \bar{G}(x)^a \bar{F}(x) + \frac{1}{a}\, \bar{F}(x)^{1+a} \right\} dx, \quad a > 0.$$

The above divergences, $\mathcal{C}d_a(F:G)$ and $\mathcal{S}d_a(\bar{F}:\bar{G})$, are non-negative, for all $a > 0$. They are equal to 0 if and only if the underline cumulative distributions F and G, or the respective survival functions $\bar{F}$ and $\bar{G}$ are coincide. The proof of this assertion is immediately obtained by following the proof of Theorem 9.1 of Basu *et al.* [14]. It is seen that the case $a = 0$ is excluded from the definition of $\mathcal{C}d_a(F:G)$ and $\mathcal{S}d_a(\bar{F}:\bar{G})$ in (41). It can be easily shown that $\lim_{a\to 0} \mathcal{C}d_a(F:G) = \mathrm{CKL}(F:G)$ and $\lim_{a\to 0} \mathcal{S}d_a(\bar{F}:\bar{G}) = \mathrm{CRKL}(F:G)$, where the limiting measures $\mathrm{CKL}(F:G)$ and $\mathrm{CRKL}(F:G)$ have been defined by (23) and (24), respectively.

5. Fisher's type cumulative and survival information

Fisher's information measure, defined by (13), is a key expression which is connected with important results in mathematical statistics. It is related to the Kullback–Leibler divergence, in a parametric framework, and this relation is formulated in (16). A natural question is raised at this point: How would be defined Fisher's type measure in terms of a distribution function or in terms of a survival function? We will try to give an answer to this question motivated by the limiting connection of classic Csiszár's ϕ-divergence and Fisher's information, formulated by (16) and (17). This is also based on similar derivations in Section 3 of Park *et al.* [66] and the recent work by Kharazmi and Balakrishnan [43]. To formulate the definition, consider the d-dimensional Euclidean space $\mathbb{R}^d$ and denote by $\mathcal{B}^d$ the σ-algebra of Borel subsets of $\mathbb{R}^d$. Let a parametric family of probability measures P_θ on $(\mathbb{R}^d, \mathcal{B}^d)$, depending on an unknown parameter θ, belonging to the parameter space $\Theta \in \mathbb{R}$. For a d-dimensional random vector $X = (X_1, \ldots, X_d)$, let F_θ denotes the joint distribution function of X, with $F_\theta(x_1, \ldots, x_d) = P_\theta(X_1 \le x_1, \ldots, X_d \le x_d)$ and let $\bar{F}_\theta$ denotes the joint survival function of X, with $\bar{F}_\theta(x_1, \ldots, x_d) = P_\theta(X_1 > x_1, \ldots, X_d > x_d)$, for $(x_1, \ldots, x_d) \in \mathbb{R}^d, \theta \in \Theta \in \mathbb{R}$.

Motivated by the limiting behavior, cf. (17), between the classic Fisher information and Csiszár's ϕ-divergence, it is investigated, in the next proposition, an analogous behavior of the cumulative and survival Csiszár's type ϕ-divergences, defined by (32) and (33).

Proposition 3. *Let a parametric family of joint distribution functions $F_\theta(x)$, $x \in \mathbb{R}^d$ and $\theta \in \Theta \subseteq \mathbb{R}$. Let also $\bar{F}_\theta(x)$ be the corresponding survival function. Then, the cumulative and survival Csiszár's type ϕ-divergences, defined by (32) and (33), are characterized by the following limiting behavior,*

$$\lim_{\delta\to 0} \frac{1}{\delta^2} \mathcal{C}\mathcal{D}_\phi(F_{\theta+\delta}, F_\theta) = \frac{\phi''(1)}{2} \left\{ \int_{\mathbb{R}^d} F_\theta(x) \left(\frac{d}{d\theta} \ln F_\theta(x) \right)^2 dx \right.$$
$$\left. - \left(\int_{\mathbb{R}^d} F_\theta(x) dx \right) \left(\frac{d}{d\theta} \ln \int_{\mathbb{R}^d} F_\theta(x) dx \right)^2 \right\},$$
$$\lim_{\delta\to 0} \frac{1}{\delta^2} \mathcal{S}\mathcal{D}_\phi(\bar{F}_{\theta+\delta}, \bar{F}_\theta) = \frac{\phi''(1)}{2} \left\{ \int_{\mathbb{R}^d} \bar{F}_\theta(x) \left(\frac{d}{d\theta} \ln \bar{F}_\theta(x) \right)^2 dx \right.$$
$$\left. - \left(\int_{\mathbb{R}^d} \bar{F}_\theta(x) dx \right) \left(\frac{d}{d\theta} \ln \int_{\mathbb{R}^d} \bar{F}_\theta(x) dx \right)^2 \right\},$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

for $\theta \in \Theta \subseteq \mathbb{R}$ and subject to the additional assumptions $\phi'(1) = 0$, for $\phi \in \Phi$, defined by (9).

Proof. We will only sketch the proof for $C\mathcal{D}_\phi$ because the other one follows in a completely similar manner. Following Pardo [65] p. 411, for $w(\theta+\delta,\theta) = \int F_\theta \phi(F_{\theta+\delta}/F_\theta)dx$, a second-order Taylor expansion of $w(\theta^*,\theta)$ around $\theta^* = \theta$ at $\theta^* = \theta + \delta$ gives

$$w(\theta+\delta,\theta) = w(\theta,\theta) + (\theta+\delta-\theta)\frac{d}{d\theta^*}w(\theta^*,\theta)|_{\theta^*=\theta} + \frac{1}{2}(\theta+\delta-\theta)^2\frac{d^2}{d(\theta^*)^2}w(\theta^*,\theta)|_{\theta^*=\theta} + O(\delta^3), \tag{42}$$

where $w(\theta,\theta) = 0$ in view of $\phi(1) = 0$ and

$$\frac{d}{d\theta^*}w(\theta^*,\theta)|_{\theta^*=\theta} = \int \phi'\left(\frac{F_{\theta^*}(x)}{F_\theta(x)}\right)\frac{d}{d\theta^*}F_{\theta^*}(x)dx\Big|_{\theta^*=\theta} = \phi'(1)\int \frac{d}{d\theta}F_\theta(x)dx = 0, \tag{43}$$

taking into account that $\phi'(1) = 0$. On the other hand,

$$\frac{d^2}{d(\theta^*)^2}w(\theta^*,\theta)|_{\theta^*=\theta} = \int \phi''\left(\frac{F_{\theta^*}(x)}{F_\theta(x)}\right)\frac{1}{F_\theta(x)}\left(\frac{d}{d\theta^*}F_{\theta^*}(x)\right)^2 dx\Big|_{\theta^*=\theta} + \int \phi'\left(\frac{F_{\theta^*}(x)}{F_\theta(x)}\right)\frac{d^2}{d(\theta^*)^2}F_{\theta^*}(x)dx\Big|_{\theta^*=\theta},$$

and taking into account that $\phi'(1) = 0$,

$$\frac{d^2}{d(\theta^*)^2}w(\theta^*,\theta)|_{\theta^*=\theta} = \int \phi''(1)\frac{1}{F_\theta(x)}\left(\frac{d}{d\theta}F_\theta(x)\right)^2 dx = \phi''(1)\int F_\theta(x)\left(\frac{d}{d\theta}\ln F_\theta(x)\right)^2 dx. \tag{44}$$

Based on (42), (43) and (44),

$$w(\theta+\delta,\theta) = \int_{\mathbb{R}^d} F_\theta(x)\phi\left(\frac{F_{\theta+\delta}(x)}{F_\theta(x)}\right)dx = \frac{1}{2}\delta^2\phi''(1)\int_{\mathbb{R}^d} F_\theta(x)\left(\frac{d}{d\theta}\ln F_\theta(x)\right)^2 dx + O(\delta^3). \tag{45}$$

Following exactly the same argument for the function $w^*(\theta+\delta,\theta) = \phi\left(\int F_{\theta+\delta}dx \,/\, \int F_\theta dx\right)$, we obtain

$$w^*(\theta+\delta,\theta) = \phi\left(\frac{\int F_{\theta+\delta}(x)dx}{\int F_\theta(x)dx}\right) = \frac{1}{2}\delta^2\phi''(1)\left(\frac{d}{d\theta}\ln\int F_\theta(x)dx\right)^2 + O(\delta^3)$$

and then

$$\left(\int F_\theta(x)dx\right)\phi\left(\frac{\int F_{\theta+\delta}(x)dx}{\int F_\theta(x)dx}\right) = \frac{1}{2}\delta^2\phi''(1)\left(\int F_\theta(x)dx\right)\left(\frac{d}{d\theta}\ln\int F_\theta(x)dx\right)^2 + O(\delta^3). \tag{46}$$

Based on (32), (45) and (46),

$$C\mathcal{D}_\phi(F_{\theta+\delta},F_\theta) = \frac{\phi''(1)}{2}\delta^2\left\{\int_{\mathbb{R}^d} F_\theta(x)\left(\frac{d}{d\theta}\ln F_\theta(x)\right)^2 dx - \left(\int_{\mathbb{R}^d} F_\theta(x)dx\right)\left(\frac{d}{d\theta}\ln\int_{\mathbb{R}^d} F_\theta(x)dx\right)^2\right\} + O(\delta^3),$$

which leads to the desired result. □

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

Based on the previous Proposition and in complete analogy with (17), which connects the classic Fisher information with Csiszár's ϕ-divergence, we state the definition of the Fisher's type cumulative and survival information.

Definition 3. *For a parametric family of joint distribution functions $F_\theta(x)$ with corresponding survival function $\bar{F}_\theta(x)$, $x \in \mathbb{R}^d$ and $\theta \in \Theta \subseteq \mathbb{R}$, the Fisher's type cumulative and survival measures of information are defined by*

$$\mathcal{CI}_F^{Fi}(\theta) = \int_{\mathbb{R}^d} F_\theta(x) \left(\frac{d}{d\theta} \ln F_\theta(x)\right)^2 dx - \left(\int_{\mathbb{R}^d} F_\theta(x)dx\right) \left(\frac{d}{d\theta} \ln \int_{\mathbb{R}^d} F_\theta(x)dx\right)^2, \tag{47}$$

$$\mathcal{SI}_F^{Fi}(\theta) = \int_{\mathbb{R}^d} \bar{F}_\theta(x) \left(\frac{d}{d\theta} \ln \bar{F}_\theta(x)\right)^2 dx - \left(\int_{\mathbb{R}^d} \bar{F}_\theta(x)dx\right) \left(\frac{d}{d\theta} \ln \int_{\mathbb{R}^d} \bar{F}_\theta(x)dx\right)^2, \tag{48}$$

for $\theta \in \Theta \subseteq \mathbb{R}$.

Remark 1.

(a) *Observe that the above defined Fisher's type cumulative and survival measures are not analogous of the classic one defined by means of probability density functions, such as the measure $\mathcal{I}_f^{Fi}(\theta)$, defined by (13). It was expected because the cumulative and survival ϕ-divergences, (32) and (33), which are used to define the Fisher's type measures of the above definition, are not analogous of the classic Csiszár's ϕ-divergence (8) for the reasons provided in the previous subsections. More precisely, because the analogous expressions of classic divergences, which are obtained by replacing densities with cumulative and survival functions may lead to negative quantities, something which was shown in the counter example, of Section 3.*

(b) *Fisher's type survival measure $\mathcal{SI}_F^{Fi}(\theta)$ has an alternative expression, in terms of expected values, if we restrict ourselves to the univariate case $d = 1$. Indeed, if we focus again on a non-negative random variable X with survival function $\bar{F}_\theta$, then $\mathcal{SI}_F^{Fi}(\theta)$ of (48) is formulated as follows:*

$$\mathcal{SI}_F^{Fi}(\theta) = \int_0^\infty \bar{F}_\theta(x) \left(\frac{d}{d\theta} \ln \bar{F}_\theta(x)\right)^2 dx - (E_\theta(X)) \left(\frac{d}{d\theta} \ln E_\theta(X)\right)^2, \quad \theta \in \Theta \subseteq \mathbb{R}. \tag{49}$$

(c) *Fisher's type cumulative and survival measures of (47) and (48) can be extended to the multiparameter case $\theta \in \Theta \subseteq \mathbb{R}^m$. In this case, the extensions of $\mathcal{CI}_F^{Fi}(\theta)$ and $\mathcal{SI}_F^{Fi}(\theta)$ will be $m \times m$ symmetric matrices, but their exposition is outside the scopes of this paper.*

(d) *The above defined Fisher's type measures should be non-negative. It is true. The proof of non-negativity of $\mathcal{CI}_F^{Fi}$ and $\mathcal{SI}_F^{Fi}$, in (47) and (48), is immediately obtained in view of the last proposition. Indeed, $\mathcal{CD}_\phi(F_{\theta+\delta}, F_\theta)$ and $\mathcal{SD}_\phi(\bar{F}_{\theta+\delta}, \bar{F}_\theta)$ are non-negative, while $\phi''(1) \geq 0$ because ϕ is a convex function. Therefore, $\mathcal{CI}_F^{Fi}$ and $\mathcal{SI}_F^{Fi}$ are non-negative as the limits of non-negative functions.*

The Fisher's type cumulative and survival measures of the previous definition have an alternative representation which is formulated in the next proposition. The representation of the proposition is the analogous of the representation (14) of the classic Fisher information measure.

Proposition 4. *For a parametric family of joint distribution functions $F_\theta(x)$ and survival functions $\bar{F}_\theta(x)$, $x \in \mathbb{R}^d$, $\theta \in \Theta \subseteq \mathbb{R}$, and under the assumption of interchanging the integral and the derivative sign*

$$\mathcal{CI}_F^{Fi}(\theta) = -\int_{\mathbb{R}^d} F_\theta(x) \left(\frac{d^2}{d\theta^2} \ln F_\theta(x)\right) dx + \frac{d^2}{d\theta^2} i(\theta) - i(\theta) \left(\frac{d}{d\theta} \ln i(\theta)\right)^2, \tag{50}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

$$\mathcal{SI}_F^{Fi}(\theta) = -\int_{\mathbb{R}^d} \bar{F}_\theta(x)\left(\frac{d^2}{d\theta^2}\ln \bar{F}_\theta(x)\right)dx + \frac{d^2}{d\theta^2}\bar{i}(\theta) - \bar{i}(\theta)\left(\frac{d}{d\theta}\ln \bar{i}(\theta)\right)^2, \tag{51}$$

with $i(\theta) = \int_{\mathbb{R}^d} F_\theta(x)dx$ *and* $\bar{i}(\theta) = \int_{\mathbb{R}^d} \bar{F}_\theta(x)dx$, *for* $\theta \in \Theta \subseteq \mathbb{R}$. *Moreover, for a non-negative random variable* X *with survival function* $\bar{F}_\theta$, $\bar{i}(\theta) = E_\theta(X)$ *and*

$$\mathcal{SI}_F^{Fi}(\theta) = -\int_0^\infty \bar{F}_\theta(x)\left(\frac{d^2}{d\theta^2}\ln \bar{F}_\theta(x)\right)dx + \frac{d^2}{d\theta^2}E_\theta(X) - E_\theta(X)\left(\frac{d}{d\theta}\ln E_\theta(X)\right)^2. \tag{52}$$

Proof. The proof is obtained by standard algebraic manipulations, similar to that in Kharazmi and Balakrishnan [43] p. 6307. □

An analogous to (15) respesentation of Fisher's type cumulative and survival measures in the case of a location parameter θ is formulated in the next proposition.

Proposition 5. *Let a random variable* X *with distribution function* $F_\theta(x)$ *and survival function* $\bar{F}_\theta(x) = 1 - F_\theta(x)$, $x \in \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}$. *Suppose, moreover, that the parameter* θ *in the considered models is a location parameter. Then, under the assumption of interchanging the integral and the derivative sign*

$$\mathcal{CI}_F^{Fi}(X) = \mathcal{CI}_F^{Fi}(F) = \int_{\mathbb{R}} F(x)\left(\frac{d}{dx}\ln F(x)\right)^2 dx - i^{-1}(F), \tag{53}$$

$$\mathcal{SI}_F^{Fi}(X) = \mathcal{SI}_F^{Fi}(\bar{F}) = \int_{\mathbb{R}} \bar{F}(x)\left(\frac{d}{dx}\ln \bar{F}(x)\right)^2 dx - \bar{i}^{-1}(\bar{F}), \tag{54}$$

where F *is a distribution function such that* $F_\theta(x) = F(x-\theta)$ *and* $\bar{F}_\theta(x) = \bar{F}(x-\theta) = 1 - F(x-\theta)$, $x \in \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}$, *with* $i(F) = \int_{\mathbb{R}} F(x)dx$ *and* $\bar{i}(\bar{F}) = \int_{\mathbb{R}} \bar{F}(x)dx$.

Proof. Taking into account that θ is a location parameter, the c.d.f. $F_\theta(x)$ depends only on $x - \theta$, that is, $F_\theta(x) = F(x-\theta)$, $x \in \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}$. Then,

$$\frac{d}{d\theta}\ln F_\theta(x) = \frac{d}{d\theta}\ln F(x-\theta) = -\frac{(d/d\theta)F(x-\theta)}{F(x-\theta)} = -\frac{F'(x-\theta)}{F(x-\theta)}.$$

Therefore,

$$\begin{aligned}\int_{\mathbb{R}} F_\theta(x)\left(\frac{d}{d\theta}\ln F_\theta(x)\right)^2 dx &= \int_{\mathbb{R}} F(x-\theta)\left(\frac{F'(x-\theta)}{F(x-\theta)}\right)^2 dx = \int_{\mathbb{R}} F(x)\left(\frac{F'(x)}{F(x)}\right)^2 dx \\ &= \int_{\mathbb{R}} F(x)\left(\frac{d}{dx}\ln F(x)\right)^2 dx.\end{aligned}$$

In a similar manner,

$$\begin{aligned}\frac{d}{d\theta}\ln\int_{\mathbb{R}} F_\theta(x)dx &= \frac{\int_{\mathbb{R}}(d/d\theta)F(x-\theta)dx}{\int_{\mathbb{R}} F(x-\theta)dx} = -\frac{\int_{\mathbb{R}} F'(x)dx}{\int_{\mathbb{R}} F(x)dx} \\ &= -\frac{\int_{\mathbb{R}}(d/dx)F(x)dx}{\int_{\mathbb{R}} F(x)dx} = -\frac{1}{i(F)} = -i^{-1}(F),\end{aligned}$$

and

$$\int_{\mathbb{R}} F_\theta(x)dx = \int_{\mathbb{R}} F(x-\theta)dx = \int_{\mathbb{R}} F(x)dx = i(F).$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

The proof of (53) follows by compining all the above intermediate results and similar is the proof of (54). □

The example that follows clarifies the measures defined above in a specific distribution.

Example 5. *Consider a random variable X with exponential distribution with mean* $E(X) = \theta$, $\theta > 0$, *and* $F_\theta(x) = 1 - e^{-x/\theta}$, $x > 0$. *Then, based on Example 3.1 of Park et al. [43],*

$$-\int_0^\infty F_\theta(x)\left(\frac{d^2}{d\theta^2}\ln F_\theta(x)\right)dx = \frac{2(\zeta(3)-1)}{\theta} \simeq \frac{0.4041}{\theta},$$
$$-\int_0^\infty \bar{F}_\theta(x)\left(\frac{d^2}{d\theta^2}\ln \bar{F}_\theta(x)\right)dx = \frac{2}{\theta},$$

where $\zeta(3)$ *is the Riemann zeta function of order three. Moreover,* $(d/d\theta)i(\theta) = \int_0^\infty (d/d\theta)(1 - e^{-x/\theta})dx = -1$. *Then, a straightforward application of (50) and (52) entails that,*

$$\mathcal{CI}_F^{Fi}(\theta) = \frac{2(\zeta(3)-1)}{\theta} \simeq \frac{0.4041}{\theta} \quad \text{and} \quad \mathcal{SI}_F^{Fi}(\theta) = \frac{1}{\theta}.$$

It is immediate to see that

$$\mathcal{CI}_F^{Fi}(\theta) < \mathcal{SI}_F^{Fi}(\theta).$$

Taking into account that $\mathcal{CD}_\phi(F_{\theta+\delta}, F_\theta)$ *and* $\mathcal{SD}_\phi(\bar{F}_{\theta+\delta}, \bar{F}_\theta)$ *are related with* $\mathcal{CI}_F^{Fi}(\theta)$ *and* $\mathcal{SI}_F^{Fi}(\theta)$, *respectively, then, in light of the previous Proposition 3, it is clear that for small values of* $\delta, \delta > 0, \mathcal{CD}_\phi(F_{\theta+\delta}, F_\theta)$ *appears to be smaller than* $\mathcal{SD}_\phi(\bar{F}_{\theta+\delta}, \bar{F}_\theta)$. *Then, based on Park et al. [66], Example 3.1, we can say that the survival divergence* $\mathcal{SD}_\phi$ *is most sensitive to some departures from exponentiality in comparison with the cumulative divergence* $\mathcal{CD}_\phi$ *and this conclusion has its origins on the comparison of the respective cumulative and survival Fisher's type information.*

Park *et al.* [66] and Kharazmi and Balakrishnan [43], concentrated in the univariate case, introduced the cumulative residual Fisher information, that is Fisher's type measure which is defined by

$$\mathcal{CI}(\theta) = \int_{\mathbb{R}} \bar{F}_\theta(x)\left(\frac{d}{d\theta}\ln \bar{F}_\theta(x)\right)^2 dx, \quad \theta \in \Theta \subseteq \mathbb{R}.$$

A special case of $\mathcal{CI}(\theta)$, when the parameter θ in the considered model is a location parameter, is presented in Kharazmi and Balakrishnan [43] and it is defined by

$$\mathcal{CI}(X) = \mathcal{CI}(\bar{F}) = \int_{\mathbb{R}} \bar{F}(x)\left(\frac{d}{dx}\ln \bar{F}(x)\right)^2 dx.$$

This measure is called the cumulative residual Fisher information measure of $\bar{F}$ for the parameter θ and it is quite analogous with the alternative form of classic Fisher's measure, given in (15). Kharazmi and Balakrishnan [43] proceed further to the study of properties of these measures.

Based on (48) and focus in the univariate case, $d = 1$, there is a strong relationship between Kharazmi and Balakrishnan's [43] measure $\mathcal{CI}(\theta)$ with the measure $\mathcal{SI}_F^{Fi}(\theta)$, namely,

$$\mathcal{SI}_F^{Fi}(\theta) = \mathcal{CI}(\theta) - \left(\int_{\mathbb{R}} \bar{F}_\theta(x)dx\right)\left(\frac{d}{d\theta}\ln \int_{\mathbb{R}} \bar{F}_\theta(x)dx\right)^2. \tag{55}$$

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

Moreover, there is a strong relation between Kharazmi and Balakrishnan [43] cumulative residual Fisher information measure $C\mathcal{I}(\bar{F})$ with the measure $\mathcal{SI}_F^{Fi}(\bar{F})$, defined by (54), namely,

$$\mathcal{SI}_F^{Fi}(\bar{F}) = C\mathcal{I}(\bar{F}) - \bar{\imath}^{-1}(\bar{F}), \tag{56}$$

with $\bar{\imath}(\bar{F}) = \int_{\mathbb{R}} \bar{F}(x)dx$. In case of a non-negative random variable X, (55) and (56) are simplified as follows,

$$\mathcal{SI}_F^{Fi}(\theta) = C\mathcal{I}(\theta) - E(X)\left(\frac{d}{d\theta}\ln E(X)\right)^2, \tag{57}$$

and

$$\mathcal{SI}_F^{Fi}(\bar{F}) = C\mathcal{I}(\bar{F}) - [E(X)]^{-1}. \tag{58}$$

Based on (57) and (58), it is, therefore, expected that the survival Fisher's type information $\mathcal{SI}_F^{Fi}(\theta)$ and $\mathcal{SI}_F^{Fi}(\bar{F})$, defined here, obeys similar properties as those in Kharazmi and Balakrishnan [43]. Two of the said properties are formulated in the next proposition.

Proposition 6.

(a) Let X be a continuous non-negative variable with absolutely continuous survival function $\bar{F}$ and hazard function $r_F(x) = f(x)/\bar{F}(x)$, $x > 0$. Then,

$$\mathcal{SI}_F^{Fi}(\bar{F}) = E[r_F(X)] - [E(X)]^{-1}.$$

(b) Let X_a be a variable with proportional hazard model corresponding to baseline non-negative variable X with survival function $\bar{F}$. Then, the survival function of X_a is given by $\bar{F}^a$, it depends on the proportionality parameter a and the survival Fisher's type information is given by,

$$\mathcal{SI}_F^{Fi}(a) = \frac{2}{a^2}\xi_2(X_a) - (E(X_a))\left(\frac{d}{da}\ln E(X_a)\right)^2,$$

with $\xi_n(X) = \int_0^{+\infty} \bar{F}(x)([-\ln \bar{F}(x)]^n/n!)dx$.

This proposition is strongly related with Theorems 1 and 2 of Kharazmi and Balakrishnan [43] and its proof is a straightforward application of (57) and (58) and the respective proofs of this cited paper. $\xi_n(X)$ is the generalized cumulative residual entropy, defined in Psarrakos and Navarro [68].

The paper by Kharazmi and Balakrishnan [43] discusses also connections between the cumulative residual entropy (18) and the cumulative residual Fisher information measure $C\mathcal{I}(\bar{F})$, presented above. This connection is achieved by a De Bruijn's type identity which is formulated by means of the above-mentioned measures which are defined on the basis of survival functions, insted of densities, which is the frame of the classic De Bruijn identity, first studied in information theory by Stam [85]. Based on (56) and the similarity between Kharazmi and Balakrishnan [43] measure $C\mathcal{I}(\bar{F})$ and the introduced here measure $\mathcal{SI}_F^{Fi}(\bar{F})$, it can been formulated a De Bruijn's type identity by means of $\mathcal{SI}_F^{Fi}(\bar{F})$, something which can be obtained in view of Theorem 5 in Kharazmi and Balakrishnan [43] and (56), above. De Bruijn identity, which concentrates and it is applied to a Gaussian noise channel, is a classic result that receives the attention nowadays. Generalizations of this identity are provided in several frames and for various generalizations of the noise channel (cf., [25,87], and references therein).

Concluding this section, it is devoted to the definition of Fisher's type measures which are defined as limiting cases of divergences, which are based on the cumulative distribution and the respective survival function. Fisher's initial measure of information (13) is a universal quantity and it is meaningful in terms of the bound it provides for the variance of an unbiased estimator and for chaotic systems, among other applications. So, establishing limiting relationship between the Fisher information and divergence

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

provides insights about these measures in various application problems; examples include Kullback's [48] interpretation of information divergence in terms of the widely used Fisher information in statistics, Lindley's [53] interpretation of the mutual information in terms of the Fisher information for evaluation of experiments, and Soofi's [83] interpretation of the Fisher information in terms of divergence in the chaotic systems with noise due to the difference between two initial values. Based on this discussion, it is clear that Fisher's measure and the measures of information, in general, are meaningful and they are successfully used, in practice, in various applications, in almost all fields of science and engineering. It is, therefore, expected the same to be valid for the newly defined measures in terms of cumulative and survival functions, initiated by the work of Rao *et al.* [73]. Although they have a short presence in the respective literature, they have received considerable attention by the research community and they have already been used to various application problems (cf., e.g., the recent papers by [2,3]).

6. Conclusion

This paper aimed to summarize parts of the enormous existing literature and to provide a short review on the most known entropies and divergences and their cumulative and survival counterparts. Searching the literature, it seemed that there is not yet appeared a definition of the broad class of Csiszár's type ϕ-divergences or a definition of the density power divergence of Basu *et al.* [13] on the basis of cumulative and survival functions, a framework which was initiated in the paper by Rao *et al.* [73]. There is no, in addition, an analogous formulation of Fisher's measure of information, on the basis of cumulative and survival functions, as limiting case of similar Csiszár's type ϕ-divergence. This gap aims to bridge the present paper. Therefore, the main aim of this work is to fill the gap and to introduce Csiszár's type ϕ-divergences, density power type divergence and Fisher's type information measure by means of cumulative distribution functions and survival functions. The measures introduced here are based on distribution functions which always exist, while their classic counterparts are defined on the basis of probability density functions, which do not always exist or they are complicated in some disciplines and contexts.

Classic measures of divergence have been broadly used to present and to develop statistical inference by exploiting the non-negativity and identity of indiscernibles property (10) which permits their use as pseudo distances between probability distribution. At least two monographs, to the best of our knowledge, are devoted to this line of research (cf. [14,65]) providing robust statistical procedures. The introduced here measures obey a similar property and they can be also used as pseudo distances between distributions, especially in cases where the underlined densities are not so tractable while the respective distribution or survival functions can be managed, easily. Moreover, they depend on the distribution and the survival functions, and hence, there is more flexibility in defining their empirical counterparts so as to be used as a type of loss function to formulate and develop estimation or as test statistics in testing statistical hypotheses, in several disciplines and contexts. The usefulness of the introduced here measures for the development of information theoretic statistical inference is hoped to be the subject of future work.

Acknowledgment. The author is thankful to the editor and an anonymous reviewer for the careful reading of a previous version of the paper and for valuable comments and suggestions which have considerably improved this work.

References

[1] Ali, S.M. & Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society Series B* 28: 131–142.
[2] Ardakani, O.M., Ebrahimi, N., & Soofi, E.S. (2018). Ranking forecasts by stochastic error distance, information and reliability measures. *International Statistical Review* 86: 442–468.
[3] Ardakani, O.M., Asadi, M., Ebrahimi, N., & Soofi, E.S. (2020). MR plot: a big data tool for distinguishing distributions. *Statistical Analysis and Data Mining* 13: 405–418.
[4] Arikan, E. (2016). Varentropy decreases under the polar transform. *IEEE Transactions on Information Theory* 62: 3390–3400.
[5] Arndt, C. (2001). *Information measures. Information and its description in science and engineering*. Berlin: Springer-Verlag.

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

[6] Asadi, M., Ebrahimi, N., & Soofi, E.S. (2017). Connections of Gini, Fisher, and Shannon by Bayes Risk under proportional hazards. *Journal of Applied Probability* 54: 1027–1050.
[7] Asadi, M., Ebrahimi, N., & Soofi, E.S. (2018). Optimal hazard models based on partial information. *European Journal of Operational Research* 270: 723–733.
[8] Avlogiaris, G., Micheas, A., & Zografos, K. (2016). On local divergences between two probability measures. *Metrika* 79: 303–333.
[9] Avlogiaris, G., Micheas, A., & Zografos, K. (2016). On testing local hypotheses via local divergence. *Statistical Methodology* 31: 20–42.
[10] Avlogiaris, G., Micheas, A., & Zografos, K. (2019). A criterion for local model selection. *Sankhya A* 81: 406–444.
[11] Balakrishnan, N. & Lai, C.-D. (2009). *Continuous bivariate distributions*, 2nd ed. Dordrecht: Springer.
[12] Baratpour, S. & Rad, H.A. (2012). Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. *Communications in Statistics - Theory and Methods* 41: 1387–1396.
[13] Basu, A., Harris, I.R., Hjort, N.L., & Jones, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika* 85: 549–559.
[14] Basu, A., Shioya, H., & Park, C. (2011). *Statistical inference. The minimum distance approach*. Monographs on Statistics and Applied Probability, 120. Boca Raton, FL: CRC Press.
[15] Batsidis, A. & Zografos, K. (2013). A necessary test of fit of specific elliptical distributions based on an estimator of Song's measure. *The Journal of Multivariate Analysis* 113: 91–105.
[16] Billingsley, P. (1995). *Probability and measure*, 3rd ed. New York: John Wiley & Sons, Inc.
[17] Blumentritt, T. & Schmid, F. (2012). Mutual information as a measure of multivariate association: analytical properties and statistical estimation. *Journal of Statistical Computation and Simulation* 82: 1257–1274.
[18] Bobkov, S.G., Gozlan, N., Roberto, C., & Samson, P.-M. (2014). Bounds on the deficit in the logarithmic Sobolev inequality. *Journal of Functional Analysis* 267: 4110–4138.
[19] Broniatowski, M. & Stummer, W. (2019). Some universal insights on divergences for statistics, machine learning and artificial intelligence. In *Geometric structures of information. Signals and Communication Technology*. Springer, pp. 149–211.
[20] Burbea, J. & Rao, C.R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *The Journal of Multivariate Analysis* 12: 575–596.
[21] Calì, C., Longobardi, M., & Ahmadi, J. (2017). Some properties of cumulative Tsallis entropy. *Physica A* 486: 1012–1021.
[22] Carlen, E.A. (1991). Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *Journal of Functional Analysis* 101: 194–211.
[23] Casella, G. & Berger, R.L. (1990). *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Pacific Grove, CA: Duxbury Press.
[24] Chen, X., Kar, S., & Ralescu, D.A. (2012). Cross-entropy measure of uncertain variables. *Information Science* 201: 53–60.
[25] Choi, M.C.H., Lee, C., & Song, J. (2021). Entropy flow and De Bruijn's identity for a class of stochastic differential equations driven by fractional Brownian motion. *Probability in the Engineering and Informational Sciences* 35: 369–380.
[26] Cover, T.M. & Thomas, J.A. (2006). *Elements of information theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].
[27] Cressie, N. & Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society Series B* 46: 440–464.
[28] Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten. *Magyar Tudományos Akadémia. Matematikai KutatóIntézetének Közleményei* 8: 85–108.
[29] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 2: 299–318.
[30] Csiszár, I. & Körner, J. (1986). *Information theory, coding theorems for discrete memoryless systems*. Budapest: Akadémiai Kiadó.
[31] Di Crescenzo, A. & Longobardi, M. (2009). On cumulative entropies. *Journal of Statistical Planning and Inference* 139: 4072–4087.
[32] Di Crescenzo, A. & Longobardi, M. (2015). Some properties and applications of cumulative Kullback-Leibler information. *Applied Stochastic Models in Business and Industry* 31: 875–891.
[33] Ebrahimi, N., Jalali, N.Y., & Soofi, E.S. (2014). Comparison, utility, and partition of dependence under absolutely continuous and singular distributions. *The Journal of Multivariate Analysis* 131: 32–50.
[34] Ferentinos, K. & Papaioannou, T. (1981). New parametric measures of information. *Information and Control* 51: 193–208.
[35] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, A* 222: 309–368.
[36] Frank, L., Sanfilippo, G., & Agró, G. (2015). Extropy: complementary dual of entropy. *Statistical Science* 30: 40–58.
[37] Golomb, S. (1966). The information generating function of a probability distribution. *IEEE Transactions on Information Theory* 12: 75–77.
[38] Guha, A., Biswas, A., & Ghosh, A. (2021). A nonparametric two-sample test using a general φ-divergence-based mutual information. *Statistica Neerlandica* 75: 180–202.
[39] Guiasu, S. & Reischer, C. (1985). The relative information generating function. *Information Science* 35: 235–241.

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

[40] Harremoës, I. & Vajda, P. (2011). On pairs of f -divergences and their joint range. *IEEE Transactions on Information Theory* 57: 3220–3225.
[41] Havrda, J. & Charvát, F. (1967). Quantification method of classification processes. Concept of structural α-entropy. *Kybernetika* 3: 30–35.
[42] Joe, H. (2015). *Dependence modeling with copulas*. Boca Raton, FL: CRC Press.
[43] Kharazmi, O. & Balakrishnan, N. (2021). Cumulative residual and relative cumulative residual Fisher information and their properties. *IEEE Transactions on Information Theory* 67: 6306–6312.
[44] Kharazmi, O. & Balakrishnan, N. (2021). Jensen-information generating function and its connections to some well-known information measures. *Statistics & Probability Letters* 170: 108995.
[45] Klein, I. & Doll, M. (2020). (Generalized) maximum cumulative direct, residual, and paired Φ entropy approach. *Entropy* 22: Paper No. 91, 33 pp.
[46] Klein, I., Mangold, B., & Doll, M. (2016). Cumulative paired ϕ-entropy. *Entropy* 18: Paper No. 248, 45 pp.
[47] Kontoyiannis, I. & Verdú, S. (2013). Optimal lossless compression: source varentropy and dispersion. In *Proceedings of IEEE International Symposium on Information Theory*, Istanbul, Turkey, pp. 1739–1743.
[48] Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.
[49] Kullback, S. & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22: 79–86.
[50] Kullback, S., Keegel, J.C., & Kullback, J.H. (1987). *Topics in statistical information theory*. Lecture Notes in Statistics, 42. Berlin: Springer-Verlag.
[51] Liese, F. & Vajda, I. (1987). *Convex statistical distances*. Teubner Texts in Mathematics, Band 95. Leipzig: Teubner.
[52] Liese, F. & Vajda, I. (2008). f -divergences: sufficiency, deficiency and testing of hypotheses. In N.S. Barnett & S.S. Dragomir (eds), *Advances in inequalities from probability theory & statistics*. New York: Nova Science Publishers, pp. 113–149.
[53] Lindley, D. (1961). The use of prior probability distributions in statistical inference and decision. In J. Neyman (ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. I. Berkeley: University of California Press, pp. 453–468.
[54] Linfoot, E.H. (1957). An informational measure of correlation. *Information and Control* 1: 85–89.
[55] Mayer-Wolf, E. (1990). The Cramér-Rao functional and limiting laws. *Annals of Probability* 18: 840–850.
[56] Menéndez, M.L., Morales, D., Pardo, L., & Salicrú, M. (1995). Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: a unified study. *Statistical Papers* 36: 1–29.
[57] Micheas, A. & Zografos, K. (2006). Measuring stochastic dependence using ϕ-divergence. *The Journal of Multivariate Analysis* 97: 765–784.
[58] Morimoto, T. (1963). Markov processes and the h-theorem. *Journal of the Physical Society of Japan* 18: 328–331.
[59] Nelsen, R.B. (2006). *An introduction to copulas*, 2nd ed. Springer Series in Statistics. New York: Springer.
[60] Niculescu, C.P. & Persson, L.E. (2006). *Convex functions and their applications. A contemporary approach*. CMS Books in Mathematics. New York: Springer.
[61] Onicescu, O. (1966). Energie informationelle. *Comptes Rendus de l'Académie des Sciences Paris, Series A* 263: 841–842.
[62] Papaioannou, T. (1985). Measures of information. In S. Kotz & N.L. Johnson (eds), *Encyclopedia of statistical sciences*, vol. 5. New York: John Wiley & Sons, pp. 391–397.
[63] Papaioannou, T. (2001). On distances and measures of information: a case of diversity. In C.A. Charalambides, M.V. Koutras, & N. Balakrishnan (eds), *Probability and statistical models with applications*. London: Chapman & Hall, pp. 503–515.
[64] Papaioannou, T. & Ferentinos, K. (2005). On two forms of Fisher's measure of information. *Communications in Statistics - Theory and Methods* 34: 1461–1470.
[65] Pardo, L. (2006). *Statistical inference based on divergence measures*. Boca Raton, FL: Chapman & Hall/CRC.
[66] Park, S., Rao, M., & Shin, D.W. (2012). On cumulative residual Kullback-Leibler information. *Statistics & Probability Letters* 82: 2025–2032.
[67] Park, S., Alizadeh Noughabi, H., & Kim, I. (2018). General cumulative Kullback-Leibler information. *Communications in Statistics - Theory and Methods* 47: 1551–1560.
[68] Psarrakos, G. & Navarro, J. (2013). Generalized cumulative residual entropy and record values. *Metrika* 76: 623–640.
[69] Qiu, G. (2017). The extropy of order statistics and record values. *Statistics & Probability Letters* 120: 52–60.
[70] Qiu, G. & Jia, K. (2018). The residual extropy of order statistics. *Statistics & Probability Letters* 133: 15–22.
[71] Rajesh, G. & Sunoj, S.M. (2019). Some properties of cumulative Tsallis entropy of order α. *Statistical Papers* 60: 583–593.
[72] Rao, M. (2005). More on a new concept of entropy and information. *Journal of Theoretical Probability* 18: 967–981.
[73] Rao, M., Chen, Y., Vemuri, B.C., & Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory* 50: 1220–1228.
[74] Read, T.R.C. & Cressie, N. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
[75] Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, pp. 547–561.
[76] Sachlas, A. & Papaioannou, T. (2010). Divergences without probability vectors and their applications. *Applied Stochastic Models in Business and Industry* 26: 448–472.
[77] Sachlas, A. & Papaioannou, T. (2014). Residual and past entropy in actuarial science and survival models. *Methodology and Computing in Applied Probability* 16: 79–99.
[78] Salicrú, M. (1994). Measures of information associated with Csiszár's divergences. *Kybernetika* 30: 563–573.

https://doi.org/10.1017/S0269964822000031 Published online by Cambridge University Press

[79] Salicrú, M., Menéndez, M.L., Morales, D., & Pardo, L. (1993). Asymptotic distribution of (h, ϕ)-entropies. *Communications in Statistics - Theory and Methods* 22: 2015–2031.
[80] Sati, M.M. & Gupta, N. (2015). Some characterization results on dynamic cumulative residual Tsallis entropy. *Journal of Probability and Statistics*: Art. ID 694203, 8 pp.
[81] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656.
[82] Song, K.-S. (2001). Rényi information, loglikelihood and an intrinsic distribution measure. *Journal of Statistical Planning and Inference* 93: 51–69.
[83] Soofi, E.S. (1994). Capturing the intangible concept of information. *Journal of the American Statistical Association* 89: 1243–1254.
[84] Soofi, E.S. (2000). Principal information theoretic approaches. *Journal of the American Statistical Association* 95: 1349–1353.
[85] Stam, A.J. (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* 2: 101–112.
[86] Stummer, W. & Vajda, I. (2012). On Bregman distances and divergences of probability measures. *IEEE Transactions on Information Theory* 58: 1277–1288.
[87] Toranzo, I.V., Zozor, S., & Brossier, J.-M. (2018). Generalization of the de Bruijn identity to general ϕ-entropies and ϕ-Fisher informations. *IEEE Transactions on Information Theory* 64: 6743–6758.
[88] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52: 479–487.
[89] Vajda, I. (1989). *Theory of statistical inference and information*. Dordrecht, Netherlands: Kluwer Academic Publishers.
[90] Vajda, I. (1995). Information theoretic methods in statistics. In *Research Report No. 1834. Academy of Sciences of the Czech Republic*. Prague: Institute of Information Theory and Automation.
[91] Vajda, I. (2009). On metric divergences of probability measures. *Kybernetika* 45: 885–900.
[92] Walker, S.G. (2016). Bayesian information in an experiment and the Fisher information distance. *Statistics & Probability Letters* 112: 5–9.
[93] Weller-Fahy, D.J., Borghetti, B.J., & Sodemann, A.A. (2015). A survey of distance and similarity measures used within network intrusion anomaly detection. *IEEE Communications Surveys & Tutorials* 17: 70–91.
[94] Yao, W., Nandy, D., Lindsay, B.G., & Chiaromonte, F. (2019). Covariate information matrix for sufficient dimension reduction. *Journal of the American Statistical Association* 114: 1752–1764.
[95] Zografos, K. (1998). On a measure of dependence based on Fisher's information matrix. *Communications in Statistics - Theory and Methods* 27: 1715–1728.
[96] Zografos, K. (2000). Measures of multivariate dependence based on a distance between Fisher information matrices. *Journal of Statistical Planning and Inference* 89: 91–107.
[97] Zografos, K. (2008). On Mardia's and Song's measures of kurtosis in elliptical distributions. *The Journal of Multivariate Analysis* 99: 858–879.
[98] Zografos, K. & Nadarajah, S. (2005). Survival exponential entropies. *IEEE Transactions on Information Theory* 51: 1239–1246.
[99] Zografos, K., Ferentinos, K., & Papaioannou, T. (1986). Discrete approximations to the Csiszár, Rényi, and Fisher measures of information. *Canadian Journal of Statistics* 14: 355–366.

Cite this article: Zografos K (2023). On reconsidering entropies and divergences and their cumulative counterparts: Csiszár's, DPD's and Fisher's type cumulative and survival measures. *Probability in the Engineering and Informational Sciences* **37**, 294–321. https://doi.org/10.1017/S0269964822000031