On entropy, divergence and recent reconsiderations: A bridge to Copula

Conference in memory of Charalambos Charalambides (1945-2024)

Athens, October 5 - 6, 2024

Kostas Zografos

Department of Mathematics

University of Ioannina, Greece

Overview

- **1. Measures of Information**
- **2. Applications of the Measures in Statistics: A brief exposition**
- **3. Cumulative/survival entropies and divergences**
- **4. Information-theory & Copula-theory: Extreme Value Copulas**
- **5. Epilogue-Conclusions**

1. Measures of Information (m.o.i)

Many areas in science and engineering (information theory, comunication theory, signal processing, probability, statistics, statistical learning, data science, etc.) have been greatly benefited from m.o.i.

$$\mathbf{m.o.i} \quad \rightarrow \quad \begin{cases} \text{Entropy-type,} \\ \text{Fisher-type,} \\ \text{Divergence.} \end{cases} \qquad \text{(cf. Ferentinos and Papaioannou, 1981)}$$

Let a random quantity X and let for simplicity f (or f_θ) denotes the pd.f., F (or F_θ) the respective c.d.f. and $p = (p_1, ..., p_m)$ (or $p(\theta)$) the p.d. in the discrete case, $\theta \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d, d \geq 1$.

The **main representatives** of each type of measures are:

$$\mathbf{representative} \ \rightarrow \begin{cases} \mathcal{E}_{Sh}(p) = -\sum\limits_{i=1}^{m} p_i \ln p_i, \ \mathcal{E}_{Sh}(f) = -\int\limits_{\mathbb{R}} f(x) \ln f(x) dx, \ \text{(Shannon 1948)} \\ \mathcal{I}_X^{Fi}(\theta) = \int\limits_{\mathbb{R}} f_\theta(x) \left(\frac{d}{d\theta} \ln f_\theta(x)\right)^2 dx, \ \text{(Fisher, 1922)} \\ \mathcal{D}_0(f, g) = \int\limits_{\mathbb{R}} f(x) \ln \frac{f(x)}{g(x)} dx. \ \text{(Kullback-Leibler, 1951, Kullback, 1959)} \end{cases}$$

m.o.i obey a lot of properties, of axiomatic or operational type: For Instance,

- $\mathcal{E}_{Sh}(p)$ is maximized when $p = (p_1, ..., p_m)$, $p_i = 1/m$ (discrete uniform - the most uncertain distribution - all outcomes are equally likely to occur).

- If $T = T(X)$ is a measurable transformation of the data X, then

$$\mathcal{I}_T^{Fi}(\theta) \leq \mathcal{I}_X^{Fi}(\theta) \text{ with equality if-f } T \text{ is sufficient.}$$

- $\mathcal{D}_0(f, g) \geq 0$ with equality if-f $f = g$.

Interpretation (I) of m.o.i.:

$$\mathbf{I} \quad \rightarrow \quad \begin{cases} \textbf{Entropy-type} \rightarrow \; \text{The amount of uncertainty/information about} \\ \qquad\qquad\qquad\quad \text{the outcome of a random experiment} \\ \textbf{Fisher-type} \rightarrow \; \text{The amount of information about the unknown parameter } \theta \\ \textbf{Divergence} \rightarrow \; \text{The amount of information for discrimination between } f \text{ and } g, \\ \qquad\qquad\qquad\quad \text{or } \textbf{pseudo-distance} \text{ between } f \text{ and } g. \end{cases}$$

1.1 Extensions of Shannon Entropy

$$\begin{aligned}
\mathcal{E}_{R,a}(f) &= \frac{1}{1-a}\ln\int_{\mathbb{R}} f^a(x)dx,\ a>0, a\neq 1,\ \text{(Rényi, 1961)},\ \lim_{a\to 1}\mathcal{E}_{R,a}(f)=\mathcal{E}_{Sh}(f).\\
\mathcal{E}_{Ts,a}(f) &= \frac{1}{a-1}\left(1-\int_{\mathbb{R}} f^a(x)dx\right),\ a>0, a\neq 1,\ \text{(Tsallis, 1988)},\ \lim_{a\to 1}\mathcal{E}_{Ts,a}(f)=\mathcal{E}_{Sh}(f)\\
\mathcal{I}(f) &= \int_{\mathbb{R}} f^a(x)dx,\ a>0,\ \textbf{information generating function}\text{, Golomb (1966).}\\
\mathcal{J}(f) &= -\frac{1}{2}\int_{\mathbb{R}} f^2(x)d\mu,\ \textbf{Extropy}\text{ (Frank et al., 2015, Qiu,2017, Qiuand and Jia,2018).}\\
\mathcal{E}_{\phi}(f) &= -\int_{\mathbb{R}}\phi(f(x))dx,\ \text{for a convex function } \phi \text{ (Burbea and Rao, 1982).}\\
\mathcal{E}_{\phi}^{h}(f) &= h\left(\int_{\mathbb{R}}\phi(f(x))dx\right),\ \phi \text{ concave, } h \text{ differentiable \& increasing, (Salicru et al. 1993).}\\
\mathcal{S}(f) &= -2(d/d\alpha)\mathcal{E}_{R,a}(f)|_{a=1}=Var[\ln f(X)],\ \text{(Song, 2001, Zografos, 2008).}
\end{aligned}$$

1.2 Extensions of Kullback-Leibler divergence

$$\mathcal{D}_{\chi^2}(f,g) = \int_{\mathbb{R}} g(x)\left(1-\frac{f(x)}{g(x)}\right)^2 dx, \text{ (Kagan, 1963, Pearson's, 1900 chi-square).}$$

$$\mathcal{D}_{\chi^a}(f,g) = \int_{\mathbb{R}} g(x)\left|1-\frac{f(x)}{g(x)}\right|^a dx,\ a \geq 1,\ \text{(Vajda, 1973). For } a=1, \textit{ total variation} \text{ (Saks, 1937).}$$

$$\mathcal{D}_{M,a}(f,g) = \int_{\mathbb{R}} (f^a(x)-g^a(x))^{1/a}\, dx, 0<a<1,\ \text{(Matusita, 1964). For } a=1/2,$$

$$\mathcal{D}_H(f,g) = \mathcal{D}_{M,\frac{1}{2}}(f,g) = \tfrac{1}{2}\int_{\mathbb{R}} \left(\sqrt{f(x)}-\sqrt{g(x)}\right)^2 dx, \text{ Hellinger distance,}$$

$$\mathcal{D}_{M,\frac{1}{2}}(f,g) = 2-2\int_{\mathbb{R}} \sqrt{f(x)g(x)}dx = 2-2\rho_{\frac{1}{2}}(f,g),$$

$$\rho_a(f,g) = \int_{\mathbb{R}} f^a(x)g^{1-a}(x)dx, 0<a<1, \textit{ affinity} \text{ of } f,\ g,$$

$$\mathcal{D}_{Ba}(f,g) = -\ln \int_{\mathbb{R}} \sqrt{f(x)g(x)}dx,\ \text{(Bhattacharyya, 1943).}$$

$$\mathcal{D}_{R,a}(f,g) = \frac{1}{a-1}\ln\int_{\mathbb{R}} f^a(x)g^{1-a}(x)dx, a>0, a\neq 1,\ \text{(Rényi, 1961)},\ \lim_{a\to 1}\mathcal{D}_{R,a}(f,g) = \mathcal{D}_0(f,g).$$

1.3 Csiszár's phi-divergence

After Rényi's divergence, **the broad class of ϕ-divergence between two densities f and g** introduced by Csiszár (1963, 1967) and independently by Ali and Silvey (1966), or Morimoto (1963) according to Harremoës and Vajda (2011). This omnipresent measure is defined by

$$\mathcal{D}_{\phi}(f,g)=\int_{\mathbb{R}} g(x)\phi\left(\frac{f(x)}{g(x)}\right)dx.$$

$\phi:(0,\infty)\to\mathbb{R}$ is a real valued convex function (Csiszár, 1963, 1967 and Pardo, 2006) belonging to the class of functions:

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

$\mathcal{D}_{\phi}(f,g)$ has a wide range of applications because it's a **measure of quasi-distance** or **a measure of statistical distance** between f and g since it's non-negativity and satisfies the identity of indiscernibles property (terminology by Weller-Fahy et al., 2015),

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

It's not **symmetric** for $\phi\in\boldsymbol{\Phi}$. It becomes symmetric for functions ϕ_*, $\phi_*(u)=\phi(u)+u\phi\left(\frac{1}{u}\right)$, $\phi\in\boldsymbol{\Phi}$ (Liese and Vajda, 1987, Vajda, 1995). It doesn't obey the **triangular inequality**, in general. A discussion is provided in Liese and Vajda (2008), Vajda (2009).

1.3.1 Cressie and Read λ-power divergence

It is defined (Cressie and Read, 1984, Liese and Vajda, 1987, Read and Cressie, 1988)

$$\mathcal{D}_\lambda(f,g) = \frac{1}{\lambda(\lambda+1)}\left(\int_{\mathbb{R}} g(x)\left(\frac{f(x)}{g(x)}\right)^{\lambda+1} dx - 1\right), \quad -\infty < \lambda < +\infty, \ \lambda \neq 0, -1,$$

and it is obtained from Csiszár's ϕ-divergence for

$$\phi(u) = \phi_\lambda(u) = \frac{u^{\lambda+1} - u - \lambda(u-1)}{\lambda(\lambda+1)}, \lambda \neq 0, -1, u > 0.$$

For $\lambda = 0$ and $\lambda = -1$ it is defined by

$$\lim_{\lambda \to 0} \mathcal{D}_\lambda(f,g) = \mathcal{D}_0(f,g) \text{ and } \lim_{\lambda \to -1} \mathcal{D}_\lambda(f,g) = \mathcal{D}_0(g,f)$$

It was defined to unify and study the existing **chi-square multinomial goodness-of-fit tests** , such as Pearson's chi-square ($\lambda = 1$), the log-likelihood ratio statistic ($\lambda \to 0$), modified versions of these two statistics and the Freeman and Tukey statistic ($\lambda = -1/2$).

The authors suggest the use of $\lambda = 2/3$ as an alternative to $\mathcal{D}_0(f,g)$ and $\mathcal{D}_1(f,g)$.

1.4 Basu-Harris-Hjort-Jones, 1998, Biometrika - density power divergence (DPD)

$$d_a(f,g) = \int_{\mathbb{R}} \left\{ g(x)^{1+a} - \left(1+\frac{1}{a}\right) g(x)^a f(x) + \frac{1}{a} f(x)^{1+a} \right\} dx, \;\; a > 0.$$

For all $a \geq 0$,

$$\boldsymbol{d_a(f,g) \geq 0}, \textbf{ with equality, if and only if, } \boldsymbol{f(x) = g(x), a.e.x}.$$

For $a = 0$, it is defined by

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

For $a = 1$, it reduces to the $\boldsymbol{L_2}$ **distance** $L_2(f,g) = \int_{\mathbb{R}} (f(x) - g(x))^2 \, dx$.

It is a special case of the so-called **Bregman divergence (**for $T(u) = u^{1+a}$ we get a times $d_a(f,g)$),

$$\int_{\mathbb{R}} \left[T(f(x)) - T(g(x)) - \{f(x) - g(x)\} T'(g(x)) \right] dx.$$

The meaning of the tuning parameter a:

It controls the trade off between robustness and asymptotic efficiency of the parameter estimates which are the minimizers of this family of divergences (cf. Basu et al., 2011, p. 297).

1.5 Handling more than two distributions

Generalized f-divergences have been introduced under the name of ***f*-dissimilarity** by Gyorfi and Nemetz (1977, 1978),

$$D_f(f_1, ..., f_k) = \int f(f_1(x), ..., f_k(x))dx,$$

where f is a real convex, continuous and homogeneous function.

Particular Cases: *Affinity* of Toussaint (1974), $f(x_1, ..., x_k) = -\prod_{i=1}^{k} x_i^{a_i}$, $a_i > 0$, $\sum_{i=1}^{k} a_i = 1$,

$$\rho_a(f_1, ..., f_k) = -\int f_1^{a_1}(x) ... f_k^{a_k}(x))dx.$$

Affinity of Matusita: For $a_i = \frac{1}{k}$, $i = 1, ..., k$.

f-dissimilarity leads to Csiszár's ϕ-divergence for $k = 2$ and $f(x_1, x_2) = x_2 \phi\left(\frac{x_1}{x_2}\right)$, $x_1, x_2 > 0$.

If f is strictly convex, then (cf. Gyorfi and Nemetz, 1978)

$$D_f(f_1, ..., f_k) \geq f(1, ..., 1) \text{ with equality if-f } f_1 = ... = f_k.$$

1.6 Extensions of Fisher's measure

$$\begin{aligned}
\mathcal{I}_X^{V}(\theta) &= \int\limits_{\mathbb{R}} f_\theta(x)\left|\frac{d}{d\theta}\ln f_\theta(x)\right|^a dx,\ a \geq 1 \quad \text{(Vajda, 1973)} \\
\mathcal{I}_X^{Mat}(\theta) &= \left(\int\limits_{\mathbb{R}} f_\theta(x)\left|\frac{d}{d\theta}\ln f_\theta(x)\right|^a dx\right)^{1/a},\ a \geq 1 \quad \text{(Mathai, 1967)} \\
\mathcal{I}_X^{Bo}(\theta) &= \left(\int\limits_{\mathbb{R}} f_\theta(x)\left|\frac{d}{d\theta}\ln f_\theta(x)\right|^{\frac{s}{s-1}} dx\right)^{s-1},\ s > 1 \quad \text{(Boekee, 1977)}
\end{aligned}$$

They obey the *maximal-invariance property*: If $T = T(X)$ is a measurable transformation of the data X, then

$$\mathcal{I}_T^{Fi}(\theta) \leq \mathcal{I}_X^{Fi}(\theta) \text{ with equality if-f } T \text{ is sufficient.}$$

Fisher information number: If $\boldsymbol{\theta}$ **is a location parameter** in the model $f(x;\theta)$, $x \in \mathbb{R}$, $\theta \in \boldsymbol{\Theta} \subseteq \mathbb{R}$, $f(x;\theta) = h(x-\theta)$, then

$$\mathcal{J}^{Fi}(f) = \int\limits_{\mathbb{R}} h(x)\left(\frac{d}{dx}\ln h(x)\right)^2 dx = -\int_{\mathbb{R}} h(x)\frac{d^2}{dx^2}\ln h(x)dx.$$

It is widely used in different areas, such as in statistics, in functional analysis, statistical physics, in signal processing, etc. (cf. Mayer-Wolf, 1990, Carlen, 1991, Papaioannou and Ferentinos 2005, Bobkov et al., 2014, Walker, 2016, Toranzo et al., 2018, Choi et al., 2021). The multivariate version is analogous and it also received the attention of researches nowadays, cf. Yao, et al. (2019) in dimension reduction and Zografos (1998, 2000) in formulating multivariate dependence.

Connection with Divergences: for a parametric family $f(x;\theta)$, $x \in \mathbb{R}$, $\theta \in \boldsymbol{\Theta} \subseteq \mathbb{R}$,

$$\begin{aligned}
\lim_{\delta \to 0} \frac{1}{\delta^2} \mathcal{D}_0(f(x;\theta), f(x;\theta+\delta)) &= \mathcal{I}_f^{Fi}(\theta), \ \theta \in \boldsymbol{\Theta}, \\
\lim_{\delta \to 0} \frac{1}{\delta^2} \mathcal{D}_\phi(f(x;\theta+\delta), f(x;\theta)) &= \frac{\phi''(1)}{2} \mathcal{I}_f^{Fi}(\theta), \ \theta \in \boldsymbol{\Theta}.
\end{aligned}$$

Fisher's measure of information within a second-order approximation **is the discrimination information between two distributions** that belong to the same parametric family.

Fisher information can be used to define distance between densities in a parametric probability space. It devised by the 25-year-old C. R. Rao in 1945, who was introduced differential geometry into statistical inference, opening up the burgeoning field now called **information geometry**.

In Summary: **m.o.i.** are defined by means of **density functions**, they are omnipresent quantities, which play an important role, the last eight decades, in probability and statistics but also to many other fields of science and engineering.

Closing this review on entropy and divergences:

Interesting **generalized** and **unified classes of divergences** have been proposed in the literature by:

Tsairidis et al. (1996,..., 2000) in *cencored data* , Mattheou and Karagrigoriou (2009), Vonta ang Karagrigoriou (2010) in *reliability* , Sachlas and Papaioannou (2010,..., 2014) in *insurance* , Stummer and Vajda (2012), Broniatowski and Stummer (2019, 2022), among many others.

The notion of divergence has been extended to a **local setting** and the respective **local divergences** have been used to develop *statistical inference* and *model selection* techniques *in a local setting* (cf. Avlogiaris, Micheas and Zografos (2016a,b, 2019)).

Measures of Entropy or slight modifications of them are considered and used as indices of:

Diveristy (C. R. Rao, 1982, 1986, Ricotta, 2006, Rajaram, 2017), **Measures of the Shape which are used in developing goodness-of-fit tests** (Song, 2001, Zografos, 2008, Kontoyiannis and Verdu, 2013 and Arikan, 2016), **Risk Measures** (Pichler and Schlotter, 2020), ...

2. Applications of the measures in Statistics

Applications of Shannon entropy or other entropy measures include, among many others, encoding information sources, compressing data (e.g., for ZIP files), encoding channels, quantification of the ecological diversity of ecosystems, as well as error detection and correction.

Applications of Entropy in Probability and Statistics:

→ **Limit Theorems**, such as *Central Limit Theorem* (CLT) by exploiting information theoretic properties of Entropy/Fisher Information like *entropy power inequality*, etc.

cf. Oliner Johnson (2004). *Information Theory and the CLT*, ICP.

→ **Goodness-of-fit Tests**,developed on the basis of *Maximum Entropy Principle* and **spacings-type non-parametric estimators of Shannon** or other entropy.

cf. Vasicek, 1976, Dudewicz and an der Meulen, 1981, Jammalamadaka Rao et al. (1984,..., 2024), Arizono and Ohta, 1989, Ebrahimi et al, 1992, ..., Chaji and Zografos, 2019, Girardin and Lequesne, 2019, Leonenko et al., 2021, ...

Applications of Divergences in Estimation and Testing

Applications are based by considering divergences as *Statistical Distances* or *pseudo-distances.*

2.1 Minimum Kullback-Leibler Divergence (KLD) and ML Estimators

Minimum Distance Estimation has a long history: Wolfowitz (1953, 1957), Matusita (1953), ..., Beran (1977),..., Lindsay (1994),..., Basu et al (1998, 2011), ..., Pyne et al. (2022), Ghosh (2022), ...

Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ be i.i.d. replications of $\boldsymbol{X}$ which are described by the **true but unknown distribution** g.

Suppose that $\{f_{\boldsymbol{\theta}} = f(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$ is a parametric identifiable family of candidate distributions to describe the data.

The *Maximum Likelihood Estimator (mle)* ,

$$\widehat{\boldsymbol{\theta}}_n^{mle} = \arg\max_{\boldsymbol{\theta}\, \in\Theta} \left\{ \log \prod_{i=1}^{n} f(\boldsymbol{x}_i;\boldsymbol{\theta}) \right\} \overset{\text{large } n}{\simeq} \arg\min_{\boldsymbol{\theta}\, \in\Theta} \mathcal{D}_0(g, f_{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\theta}}_n^{KL}.$$

Then, **the MLE coinsides with Minimum KLD Estimator of $\boldsymbol{\theta}$**, obtained by minimizing KLD between the true model g and the model which is adopted to describe the data, $f(\cdot;\boldsymbol{\theta})$.

Minimum Density Power Divergence Estimators (MDPDE)

MLE are efficient and consistent but they fail in robustness. This was motivated Basu, Harris, Hjort and Jones, 1998, to introduce MDPDE, by minimizinng the *Density Power Divergence*

$$d_a(g, f_{\boldsymbol{\theta}}) = \int\limits_{\mathbb{R}^m} \left\{ f(\boldsymbol{x};\boldsymbol{\theta})^{1+a} - \left(1+\frac{1}{a}\right) g(\boldsymbol{x}) f(\boldsymbol{x};\boldsymbol{\theta})^{a} + \frac{1}{a} g(\boldsymbol{x})^{1+a} \right\} d\boldsymbol{x}, \ a > 0.$$

The MDPDE of $\boldsymbol{\theta}$ is defined,

$$\widehat{\boldsymbol{\theta}}_n^a = \arg \min_{\boldsymbol{\theta} \in \Theta} d_a(g, f_{\boldsymbol{\theta}}).$$

The tuning parameter a: *controls the trade off between robustness and asymptotic efficiency* of the estimators which are the minimizers of this family of divergences.

When $a \longrightarrow 0, \lim_{a \to 0} d_a(g, f_{\boldsymbol{\theta}}) = \mathcal{D}_0(g, f_{\boldsymbol{\theta}})$. **Then,** for $a \longrightarrow 0$, $\widehat{\boldsymbol{\theta}}_n^a = \widehat{\boldsymbol{\theta}}_n^{KL} \overset{\text{large } n}{\simeq} \widehat{\boldsymbol{\theta}}_n^{mle}$.

Properties of MDPDE $\widehat{\boldsymbol{\theta}}_n^a$:

- MDPDE $\widehat{\boldsymbol{\theta}}_n^a$ is a **Consistent Estimator** of $\boldsymbol{\theta}$.
- MDPDE $\widehat{\boldsymbol{\theta}}_n^a$ is **Asymptotically Normal.**
- **Robustness of $\widehat{\boldsymbol{\theta}}_n^a$**: observed in the **influence function** of the estimators. Through simulations, MDPDE outperforms the MLE in the presence of outlying observations at higher values of a.

2.2. Applications of Divergences in Testing Statistical Hypotheses

Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ be i.i.d. from the **identifiable** family $\{f_{\boldsymbol{\theta}} = f(\cdot; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$. Let,

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ versus } H_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \text{ for a } \textbf{specific } \boldsymbol{\theta}_0 \in \Theta \subseteq \mathbb{R}^p.$$

The Procedure: Based on Wald's ∼ 1943 ideas, let an estimator of $\boldsymbol{\theta}$, say $\widehat{\boldsymbol{\theta}}_n^a$.
If $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true, then $\widehat{\boldsymbol{\theta}}_n^a$ is expected to be close of $\boldsymbol{\theta}_0$. Hence, a *divergence* between $f_{\widehat{\boldsymbol{\theta}}_n^a} = f(\cdot; \widehat{\boldsymbol{\theta}}_n^a)$ and $f_{\boldsymbol{\theta}_0} = f(\cdot; \boldsymbol{\theta}_0)$ will be **small**. This is in **favour** of H_0.
Then**, Large Values** of a *divergence* $(f_{\widehat{\boldsymbol{\theta}}_n^a}, f_{\boldsymbol{\theta}_0})$ supports **rejection** of H_0.

A test statistic can be based on **a divergence**, say the Csiszár's **ϕ-divergence** $\mathcal{D}_\phi(f_{\widehat{\boldsymbol{\theta}}_n^a}, f_{\boldsymbol{\theta}_0})$ or $d_a\left(f_{\widehat{\boldsymbol{\theta}}_n^a}, f_{\boldsymbol{\theta}_0}\right)$, between $f(\cdot; \widehat{\boldsymbol{\theta}}_n^a)$ and the H_0 model $f(\cdot; \boldsymbol{\theta}_0)$. This **information theoretic procedure** provides an **intuitive formulation and solution** of the testing of hypotheses problem.
Under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\frac{2n}{\phi''(1)} \mathcal{D}_\phi\left(f_{\widehat{\boldsymbol{\theta}}_n^{a=0}}, f_{\boldsymbol{\theta}_0}\right) \underset{n \to \infty}{\overset{\mathcal{L}}{\to}} \chi_p^2 \quad \textbf{or} \quad 2n d_a\left(f_{\widehat{\boldsymbol{\theta}}_n^a}, f_{\boldsymbol{\theta}_0}\right) \underset{n \to \infty}{\overset{\mathcal{L}}{\to}} \sum_{i=1}^{r} w_i \chi_{i,1}^2,$$

w_i eigenvalues of matrices depending on the **score function** $\boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\boldsymbol{x}; \boldsymbol{\theta})$.

For $a = 0$, $2n d_a(f_{\widehat{\boldsymbol{\theta}}_n^a}, f_{\boldsymbol{\theta}_0})$ is asymptotically equivalent to the classical *LR test statistic.*

2.3. Advanced Applications of Divergences in Estimation and Testing

(i) Robust DPD-based tests for composite hypotheses:

$$H_0: \quad \boldsymbol{\theta} \in \ \Theta_0 = \{\boldsymbol{\theta} \in \ \Theta : \boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{0}_p\} \subseteq \Theta, \quad \text{against } H_a: \quad \boldsymbol{\theta} \notin \ \Theta_0.$$

Ref: Basu, Chakraborty, Ghosh and Pardo, 2022, *JMVA*, 50th Anniversary Jubilee Edition.

(ii) Estimation and testing on independent, not identically distributed observations:

Ref: Castilla, Jaenada and Pardo, 2022, *IEEE Trans. Inform. Theory*.

(iii) On distance-type Gaussian estimation: , Zhang (2019, *JMVA*)

Ref: Castilla and Zografos, 2022, *JMVA*, 50th Anniversary Jubilee Edition.

Felipe, Jaenada, Miranda and Pardo, *Mathematics*, 2023.

(iv) MDPD Estimation, Testing and Model Selection in a Composite Likelihood Framework:

Ref: A series of papers by: Castilla, Martin, Pardo and Zografos, 2018-2021.

(v) MDPD Estimation, Testing and Model Selection in a local setting by using local divergences:

Ref: A series of papers by: Avlogiaris, Micheas and Zografos, 2018-2021.

(vi) *Ordinal Response Models* (Pyne et al., 2022), *Finite Markov Chains* (Ghosh, 2022), *Interval-monitored step-stress experiment* (Balakrishanan et al., 2024), ...

2.4. Measures of Dependence - Tests of Independence

Let r.v. $\boldsymbol{X} = (X_1, X_2, ..., X_n)^T$,with joint density $f(x_1, ..., x_n)$ and marginals $f_{X_i}(x_i)$, $i = 1, ..., n$. Let, $f_0(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$ the joint density under independence.

Mutual Information:

$$\mathcal{MI}(X_1, X_2, ..., X_n) = \mathcal{D}_0(f, f_0) = \int_{\mathbb{R}^n} f(x_1, ..., x_n) \log \frac{f(x_1, ..., x_n)}{f_{X_1}(x_1) \cdots f_{X_n}(x_n)} dx_1 ... dx_n.$$

(cf. Linfoot, 1957, Kullback, 1959,...., Joe, 1989, ..., Geenens and de Michaux, 2022, De Keyser and Gijbels, 2023, ...). More generaly,

$$\mathcal{D}_\phi(f, f_0) = \int_{\mathbb{R}^n} f_0(x_1, ..., x_n) \phi\left(\frac{f(x_1, ..., x_n)}{f_0(x_1, ..., x_n)}\right) dx_1 ... dx_n = \int_{\mathbb{R}^n} f_0(\boldsymbol{x}) \phi\left(\frac{f(\boldsymbol{x})}{\prod_{i=1}^{n} f_{X_i}(x_i))}\right) d\boldsymbol{x},$$

defines a distance-type measure between f and f_0 and it serves as a broad **measure of dependence** because it obeys **Renyi's (1959) type axioms** , like the following (cf. Micheas and Zografos, 2006, *JMVA*).

If $\delta_\phi(X_1, X_2, ..., X_n) = \delta_\phi(\boldsymbol{X}) = \mathcal{D}_\phi(f, f_0) - \phi(1)$, then:

(A1) $\delta_\phi(\boldsymbol{X})$ is defined for each $X_1, X_2, ..., X_n$, when X_i, $i = 1, ..., n$, is not a constant with probability one.

(A2) $\delta_\phi(\boldsymbol{X})$ is symmetric in its argument.

(A3) $\phi(1) \leq \mathcal{D}_\phi(f, f_0) \leq \phi(0) + \lim_{u \to +\infty} \frac{\phi(u)}{u}$, and hence

$$0 \leq \delta_\phi(\boldsymbol{X}) \leq \gamma = \phi(0) - \phi(1) + \lim_{u \to +\infty} \frac{\phi(u)}{u},$$

Notice that axiom (A3) is satisfied by the measure $\mathcal{D}_\phi(f, f_0)$, if and only if $\phi(1) = 0$.

(A4) If the function ϕ is strictly convex at 1 then $\delta_\phi(\boldsymbol{X}) = 0$ if and only if the random variables $X_1, X_2, ..., X_n$ are independent.

(A5) If the function ϕ is strictly convex at 1 and $\gamma = \phi(0) - \phi(1) + \lim_{u \to +\infty} \frac{\phi(u)}{u} < +\infty$, then $\delta_\phi(\boldsymbol{X}) = \gamma$ if and only if the random variables $X_1, X_2, ..., X_n$ are completely dependent.

(A6) $\delta_\phi(\boldsymbol{X})$ is invariant under one-to-one transformations $\boldsymbol{T}(\boldsymbol{X})$ of $\boldsymbol{X}$, for any selection of φ.

(A7) In the bivariate normal case with Pearson's correlation coefficient ρ, $\delta_\phi(\boldsymbol{X})$ is an increasing function of $|\rho|$.

(A8) Let $\phi(x) = x \log x$. Let also $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$. Then, $\delta_\phi(\boldsymbol{X}) \geq \delta_\phi(\boldsymbol{X}_1) + \delta_\phi(\boldsymbol{X}_2)$. (**Super-Additivity Property**: cf. Carlen, 1991, Zolotarev, 1991, Micheas and Zografos, 2006, Blumentritt and Schmid, 2012).

Testing Independence:

$$T_{ind} = \widehat{\mathcal{MI}}(X_1, X_2, ..., X_n) = \int\limits_{\mathbb{R}^n} \widehat{f}(\boldsymbol{x}) \log \frac{\widehat{f}(\boldsymbol{x})}{\widehat{f}_{X_1}(x_1) \cdots \widehat{f}_{X_n}(x_n)} d\boldsymbol{x}.$$

(cf. ..., Zeng, et al. 2018, Zhang, 2019, Geenens and de Michaux, 2022 and referenses therein)

Remark: (a) Mutual Information and Copula density

$$\mathcal{MI}(\boldsymbol{X}) = \mathcal{D}_0(f, f_0) = \int\limits_{\mathbb{R}^n} f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{f_{X_1}(x_1) \cdots f_{X_n}(x_n)} d\boldsymbol{x} = \int\limits_{[0,1]^n} c(\boldsymbol{u}) \log c(\boldsymbol{u}) d\boldsymbol{u},$$

or

$\mathcal{MI}(\boldsymbol{X}) = -\mathcal{E}_{Sh}(c)$, c is the **copula density** of $\boldsymbol{X}$.

(cf. Micheas, 1996-Thesis, Ma and Sun, 2008, 2011, Geenens and de Michaux, 2022-*JASA*)

(b) De Keyser and Gijbels, 2024, *JMVA*, propose copula-based dependence quantification between multiple groups of random variables of possibly different sizes via the family of ϕ-divergence.

Another way to bridge Information Theory and Copula Theory:

→→→→ **Cumulative Entropy and Cumulative Divergence-type Measures**

→→ which may be initiated a new period in the development of *statistical information theory*.

3.1 Cumulative/Survival-type Entropies and Divergences

Rao et al. (2004, IEEE *Tran. Inf. Theory*) introduced the **cumulative residual entropy** with a functional similarity with Shannon's (1948) entropy $\left(\mathcal{E}_{Sh}(f) = -\int\limits_{\mathbb{R}} f(x) \ln f(x) dx\right)$, by

$$CRE(F) = -\int\limits_{0}^{+\infty} \overline{F}(x) \ln \overline{F}(x) dx, \ \overline{F}(x) = 1 - F(x).$$

Zografos and Nadarajah (2005, IEEE *Tran. Inf. Theory*) provided a timely elaboration of Rao et al. (2004) measure, the **survival exponential entropies**, by

$$M_\alpha(F) = \left(\int_0^{+\infty} \overline{F}^{\alpha}(x) dx\right)^{\frac{1}{1-\alpha}}, \alpha > 0, \alpha \neq 1, \text{ and } \lim_{\alpha \to 1} M_\alpha(F) = \exp\left\{-\frac{CRE(F)}{\int_0^{+\infty} \overline{F}(x) dx}\right\}$$

Di Crescenzo and Longobardi (2009, *JSPI*), define the **cumulative entropy**, like CRE, by

$$CE(F) = -\int\limits_{0}^{+\infty} F(x) \ln F(x) dx.$$

Several other measures based on F or $\overline{F}$ have appeared (cf. Park et al. 2012, Sati and Gupta 2015, Asadi et al. 2017, Calì et al. 2017, Rajesh and Sunoj 2019, ...).

In complete analogy with Burbea and Rao (1982) ϕ-entropy, $\mathcal{E}_{\phi}(f) = -\int_{\mathbb{R}} \phi(f(x))dx$, ϕ convex, Chen et al. (2012), Klein et al. (2016) and Klein and Doll (2020) have unified and extended the CRE and the CE. Klein and Doll (2020), define the **cumulative Φ^* entropy** by,

$$CE_{\Phi^*}(F) = \int_{-\infty}^{+\infty} \Phi^*(F(x))dx,$$

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 $[0, 1]$. $CRE(F)$ and $CE(F)$ are special cases of $CE_{\Phi^*}(F)$, for $\Phi^*(u) = \varphi(1-u)$ or $\Phi^*(u) = \varphi(u)$, with $\varphi(x) = -x \ln x$, $x \in (0, 1]$.

3.1.1 Similarities-Differences of Classical Entropies and Entropies based on Cdfs

Let $\mathcal{E}_{Sh}(f_X) = -\int f_X(x) \ln f_X(x)dx$ classical Shannon entropy and the cumulative entropy $CE(F_X) = -\int F_X(x) \ln F_X(x)dx$. Then (cf. Di Crescenzo and Longobardi, 2009),

▲ $CE(F) \geq 0$ which does not hold for $\mathcal{E}_{Sh}(f)$.

▲ Let $Y = \alpha X + \beta$, $\alpha > 0$, $\beta \geq 0$. Then,
$\mathcal{E}_{Sh}(f_Y) = \mathcal{E}_{Sh}(f_X) + \log \alpha$, and $CE(F_Y) = \alpha CE(F_X)$.

▲ If X and Y are independent, then $\mathcal{E}_{Sh}(f_{X,Y}) = \mathcal{E}_{Sh}(f_X) + \mathcal{E}_{Sh}(f_Y)$, while

$$CE(F_{X,Y}) = \left(\int_0^\infty F_Y(y)dy\right) CE(F_X) + \left(\int_0^\infty F_X(x)dx\right) CE(F_Y).$$

3.1.2 Advantages of Entropies Based on Cdfs (cf. Klein et al, 2016)

▲ CE is based on probabilities and has a consistent definition for both discrete and continuous random variables.

▲ CE is always non-negative.

▲ CE can easily be estimated by the empirical distribution function. This estimation is strongly consistent, due to the strong consistency of the empirical distribution function.

3.2 The Cumulative and the Cumulative Residual Kullback-Leibler information

They are *direct extensions of the classic Kullback-Leibler divergence* $\left(\mathcal{D}_0(f,g)=\int\limits_{\mathbb{R}} f(x)\ln\frac{f(x)}{g(x)}dx\right)$ while the integral of the right hand-side ensures the non-negativity.

$$CKL(F,G)=\int\limits_{\mathbb{R}} F(x)\ln\left(\frac{F(x)}{G(x)}\right)dx+\int\limits_{\mathbb{R}}[G(x)-F(x)]dx,$$

$$CRKL(\overline{F},\overline{G})=\int\limits_{\mathbb{R}} \overline{F}(x)\ln\left(\frac{\overline{F}(x)}{\overline{G}(x)}\right)dx+\int\limits_{\mathbb{R}}[\overline{G}(x)-\overline{F}(x)]dx,$$

(cf. Baratpour and Rad 2012, Park et al. 2012, Di Crescenzo and Longobardi 2015, Park et al. 2018, among others).

Based on $\ln x \leq x-1$, they are **non-negative**,

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

This property supports the use of $CKL(F,G)$ and $CRKL(\overline{F},\overline{G})$ as **pseudo distances** between the underling distributions.

3.3 Csiszár's ϕ-divergence type cumulative and survival measures

Starting from the necessity to be non-negative, it is defined (Zografos, 2023, PEIS) by

$$\begin{aligned}
\mathcal{CD}_\phi(F,G) &= \int\limits_{\mathbb{R}} G(x)\phi\left(\frac{F(x)}{G(x)}\right)dx - \left(\int_{\mathbb{R}} G(x)dx\right)\phi\left(\frac{\int_{\mathbb{R}} F(x)dx}{\int_{\mathbb{R}} G(x)dx}\right), \\
\mathcal{SD}_\phi(\overline{F},\overline{G}) &= \int\limits_{\mathbb{R}} \overline{G}(x)\phi\left(\frac{\overline{F}(x)}{\overline{G}(x)}\right)dx - \left(\int\limits_{\mathbb{R}} \overline{G}(x)dx\right)\phi\left(\frac{\int_{\mathbb{R}} \overline{F}(x)dx}{\int_{\mathbb{R}} \overline{G}(x)dx}\right),
\end{aligned}$$

where $\phi : (0,\infty) \to \mathbb{R}$ is a real valued convex function, $\phi \in \mathbf{\Phi}$ as in the case of classic Csiszár's ϕ-divergence, $\mathcal{D}_\phi(f,g) = \int\limits_{\mathbb{R}} g(x)\phi\left(\frac{f(x)}{g(x)}\right)dx$.

They satisfy,

$$\begin{aligned}
\mathcal{CD}_\phi(F,G) &\geq 0 \text{ with equality if and only if } F(x) = G(x), \text{ on } \mathbb{R}, \\
\mathcal{SD}_\phi(\overline{F},\overline{G}) &\geq 0 \text{ with equality if and only if } \overline{F}(x) = \overline{G}(x), \text{ on } \mathbb{R}.
\end{aligned}$$

Special choices of the convex function ϕ lead to particular divergences, like **Cressie and Read cumulative/survival divergence**.

3.4 Density power divergence type cumulative and survival divergences

$$\begin{aligned}\mathcal{C}d_a(F,G) &= \int_{\mathbb{R}} \left\{G(x)^{1+a} - \left(1+\frac{1}{a}\right) G(x)^a F(x) + \frac{1}{a}\, F(x)^{1+a}\right\} dx, \;\; a > 0, \\ \mathcal{S}d_a(\overline{F},\overline{G}) &= \int_{\mathbb{R}} \left\{\overline{G}(x)^{1+a} - \left(1+\frac{1}{a}\right) \overline{G}(x)^a \overline{F}(x) + \frac{1}{a}\, \overline{F}(x)^{1+a}\right\} dx, \;\; a > 0.\end{aligned}$$

$\mathcal{C}d_a(F,G)$ and $\mathcal{S}d_a(\overline{F},\overline{G})$ are **non-negative**, for all $a > 0$ and they are equal to 0 if and only if the underline cumulative distributions F and G, or the respective survival functions $\overline{F}$ and $\overline{G}$ are coincide.

Moreover,

$$\begin{aligned}\lim_{a\to 0} \mathcal{C}d_a(F,G) &= CKL(F,G), \\ \lim_{a\to 0} \mathcal{S}d_a(\overline{F},\overline{G}) &= CRKL(F,G),\end{aligned}$$

for the limiting measures $CKL(F,G)$ and $CRKL(F,G)$ the cumulative versions of Kullback-Leibler divergence.

Remark: Cumulative and Survival Fisher's type Measure: A series of papers by Balakrishnan and Kharazmi, from 2021-2024, Zografos, 2023 and references therein.

3.5 Concluding Remark on the two types of m.o.i.

Classical m.o.i. have a long history, of more than eight decades, and they have been studied, characterized and successfully applied in a huge number of scientific fields.

M.o.i. based on the cumulative distribution function or the survival function have a **shorter live** of two decades. Their **detailed study is in progress** and their **interpretation** is not still so clear, in my view.

Both types of measures obey nice properties, and they guarantee their applications in probability, statistics, reliability engineering and computer vision, among many others, however, **they are characterized by some differences**, too.

Which of the two types of measures should be prefered, in practice?

There is no clear answer to this question and as is it is usually the case in science, **any scientific tool has its own distinct place and its own distinct role**. Thus, both approaches must **co-exist**, **complement** each other and **motivate** each other.

Trying to characterize cumulative measures, this talk will conclude with a study/application of cumulative measures to copula theory, where the cumulative distribution function dominates over the density function.

4. Information-theory & Copula-theory

Copula (Nelsen, 2006): A function $C : [0,1]^2 \longrightarrow [0,1]$ such that,
(i) $C(u,0) = 0 = C(0,v)$, $u, v \in [0,1]$,
(ii) $C(u,1) = u$ or $C(1,v) = v$, $u, v \in [0,1]$,
(iii) If u_1, v_1 and $u_2, v_2 \in [0,1]$, with $u_1 \leq u_2$ and $v_1 \leq v_2$, then

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Sklar (1959): Copula links marginal distributions to form multivariate distributions.
If (X, Y) is a r.v. with joint d.f. $F_{X,Y}$ and marginals F_X and F_Y, then there is a copula function C such that,

$$F_{X,Y}(x,y) = C\left(F_X(x), F_Y(y)\right).$$

C is itself a **bivariate distribution function** with marginals $U(0,1)$.

Let's Concentrate on: The most representative Cumulative Entropy and Divergence

Cumulative Shannon-type Entropy: $CE(F) = -\int_{0}^{+\infty} F(x) \ln F(x) dx,$

DP Divergence-type For $a > 0$,

Cumulative Divergence: $\mathcal{C}d_a(F,G) = \int_{\mathbb{R}} \left\{G(x)^{1+a} - \left(1+\frac{1}{a}\right) G(x)^a F(x) + \frac{1}{a}\, F(x)^{1+a}\right\} dx.$

How these measures are translated in terms of copulas?

If C is a copula function, then the **Shannon-type copula entropy** is (cf. Zografos, 2024),

$$\mathcal{C}\mathcal{E}_{Sh}(C) = -\int_{0}^{1}\int_{0}^{1} C(u,v) \ln C(u,v) du dv.$$

If C_1 and C_2 are copulas, then the **density power type copula divergence** between C_1, C_2 is,

$$\mathcal{C}d_a(C_1,C_2) = \int_{0}^{1}\int_{0}^{1} \left\{C_2^{1+a}(u,v) - \left(1+\frac{1}{a}\right) C_2^{a}(u,v) C_1(u,v) + \frac{1}{a}\, C_1^{1+a}(u,v)\right\} du dv, \ \ a > 0.$$

In the sequel, $\mathcal{C}\mathcal{E}_{Sh}(C)$ and $\mathcal{C}d_a(C_1,C_2)$ are studied in the context of *Extreme Value Copulas*.

Remark: **(i)** $\mathcal{C}d_a(C_1, C_2)$ serves as a quasi-distance between the underlined copulas C_1 and C_2. It can be used to introduce robust statistical procedures if one of them will be replaced by its empirical counterpart.

(ii) If C_2 is the *independence copula,* $\Pi(u,v) = uv$, $0 < u, v \leq 1$, $(u,v) \neq (1,1)$, the empirical version of $\mathcal{C}d_a(C, \Pi)$ can be the basis of the test statistic for the delopment of a test of independence.

Extreme Value Copulas (EVC)

A bivariate copula is an extreme value copula if and only if

$$C_A(u,v) = \exp\left[\ln(uv) \cdot A\left\{\frac{\ln v}{\ln(uv)}\right\}\right], \quad 0 < u, v \leq 1, \ (u,v) \neq (1,1).$$

$A : [0,1] \to [1/2, 1]$ is a convex function such that $A(0) = A(1) = 1$,

max $\boldsymbol{\{t, 1-t\}}$ (complete positive dependence-comonotonicity) $\leq \boldsymbol{A(t)} \leq$**1** (independence).

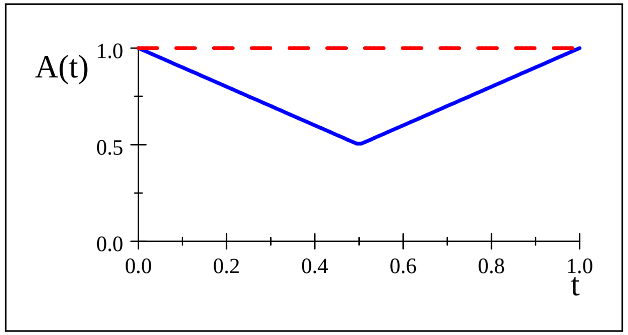


Figure 1: Pickands dependence function $A(t)$

Remark: **EVC** coincide with the set of copulas of *extreme-value distributions*, that is, the class of limit distributions with nondegenerate margins of $\left(\frac{M_{n,1}-b_{n,1}}{a_{n,1}}, \frac{M_{n,2}-b_{n,2}}{a_{n,2}}\right)$ with $M_{n,j} = \underset{i=1,\dots,n}{max}\{X_{ij}\}$, for a sample of size n of 2-dimensional random vectors $\boldsymbol{X}^{(i)} = (X_{i1}, X_{i2}), i = 1, \dots, n$, where $b_{n,j} \in \mathbb{R}$ are centering constants and $a_{n,j} > 0$ are scaling constants, $j = 1, 2$.

Proposition: **(i) Shannon's** type extreme value copula entropy,

$$\mathcal{CE}_{Sh}(C_A) = -\int\limits_{[0,1]^2} C_A(u,v) \ln C_A(u,v)dudv = 2\int\limits_0^1 \frac{A(t)}{[1+A(t)]^3}dt.$$

(ii) Tsallis' type extreme value copula entropy,

$$\mathcal{CE}_{Ts,\lambda}(C_A) = \frac{1}{\lambda-1}\int\limits_{[0,1]^2} \left[C_A(u,v) - C_A^{\lambda}(u,v)\right] dudv, \;\; \lambda > 0, \lambda \neq 1,$$

where the *information generating type function*, is given by

$$\mathcal{CI}_{\lambda}(C_A) = \int\limits_{[0,1]^2} C_A^{\lambda}(u,v)dudv = \int\limits_0^1 \frac{1}{[1+\lambda A(t)]^2}dt, \;\; \lambda > 0.$$

Remark: (i) $\mathcal{CE}_{Sh}(C_A)$ and $\mathcal{CE}_{Ts,\lambda}(C_A)$ can be thought as indices which are strongly related to the dependence in the bivariate EV case. Moreover,

$$\lim_{\lambda \to 1} \mathcal{CE}_{Ts,\lambda}(C_A) = \mathcal{CE}_{Sh}(C_A).$$

(ii) For $\lambda = 1$, the generating type function $\mathcal{CI}_\lambda(C_A)$ is simplified as follows,

$$\mathcal{CI}_1(C_A) = \int_{[0,1]^2} C_A(u,v)dudv = \int_0^1 \frac{1}{[1+A(t)]^2}dt.$$

It is connected with *Spearman's rho correlation coefficient* by

$$\rho_S(A) = 12 \int_{[0,1]^2} C_A(u,v)dudv - 3 = 12\mathcal{CI}_1(C_A) - 3 = 12\int_0^1 \frac{1}{[1+A(t)]^2}dt - 3.$$

(iii) $\mathcal{CI}_\lambda(C_A)$ is a type of *generating function* because its derivative, in respect to λ, at $\lambda = 1$, generates $\mathcal{CE}_{Sh}(C_A)$ in the sense that

$$(d/d\lambda)\mathcal{CI}_\lambda(C_A)|_{\lambda=1} = -\mathcal{CE}_{Sh}(C_A)$$

(iv) It is also the basis for the definition of Rényi's type extreme value copula entropy of the form

$$\mathcal{CE}_R(C_A) = (1-\lambda)^{-1} \log \mathcal{CI}_\lambda(C_A), \lambda > 0, \lambda \neq 1.$$

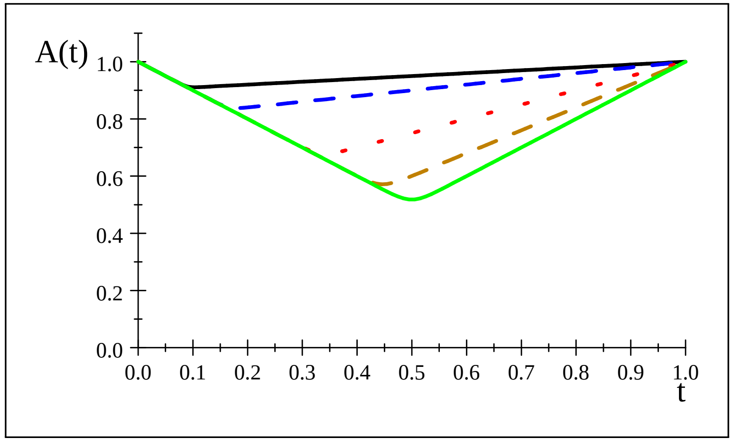
Example (Asymmetric Tawn extreme value copula): Pickands dependence function:

$$A^{Tawn}(t) = (1-\psi_1)(1-t)+(1-\psi_2)t+\left[(\psi_1(1-t))^\theta + (\psi_2 t)^\theta\right]^{1/\theta}, \ 0 \le \psi_1, \psi_2 \le 1, \theta \ge 1.$$

If $\psi_1 = \psi_2 = \Psi$, with $0 \le \Psi \le 1$, $A^{Tawn}(t)$ is related with *Gumbel copula* $A^G(t) = \left[(1-t)^\theta + t^\theta\right]^{1/\theta}$, $\theta \ge 1$, by the formula,

$$A^{Tawn}(t) = 1 + \Psi\left(A^G(t) - 1\right),$$

The figure includes the plots of $A^{Tawn}(t)$ with $\theta = 20$, $\psi_2 = 1$, for different values of ψ_1: $\psi_1 = 0.1$ (black-solid), $\psi_1 = 0.2$ (blue-dash), $\psi_1 = 0.5$ (red-dots), $\psi_1 = 0.8$ (brown-dash) and $\psi_1 = 1$ (green-solid), the last one is corresponding to symmetry ($\psi_1 = \psi_2 = 1$, Gumbel copula).



Next table evaluates Shannon type measure $\mathcal{CE}_{Sh}(C_A)$ with $\theta = 20$, $\psi_2 = 1$, for different values of ψ_1.

	$\mathcal{CE}_{Sh}$ for $\theta = 20$, $\psi_2 = 1$ and different ψ_1					
ψ_1	0.001	0.1	0.2	0.5	0.8	1
$\mathcal{CE}_{Sh}(C_A)$	0.25006	0.25566	0.26032	0.26997	0.27549	0.27777

Observe that $\mathcal{CE}_{Sh}(C_A)$ increases as the value of ψ_1 increases and $A^{Tawn}(t)$ is moving from the case of independence to the case of **complete positive dependence** in the sense of **comonotonicity** (*one variable is essentially, almost surely, an increasing function of the other* (Nelsen, 2006, p. 32) and symmetry.

Similar is the behavior of $\mathcal{CE}_{Sh}(C_A)$ when $\theta = 20$ and the role of ψ_1 and ψ_2 is reversed (cf. Zografos, 2024 where an additional example for Marshall-Olkin EVC is given).

4.1 Properties of the measures in EVC

Proposition 1 *The measures $\mathcal{CE}_{Sh}(C_A)$, $\mathcal{CI}_\lambda(C_A)$ and $\mathcal{CI}_1(C_A)$ are maximized for $A(t) = \max\{t, 1-t\}$ and they are minimized for $A(t) = 1$, for $t \in [0,1]$, while their range of values is* *as follows,*

$$\begin{aligned} 0.25 &\leq \mathcal{CE}_{Sh}(C_A) \leq 0.277778, \\ \frac{1}{(1+\lambda)^2} &\leq \mathcal{CI}_\lambda(C_A) \leq \frac{2}{(1+\lambda)(2+\lambda)}, \;\; \lambda > 0, \\ \frac{1}{4} &\leq \mathcal{CI}_1(C_A) \leq \frac{1}{3}. \end{aligned}$$

Properties of extreme value copula entropies

Properties based on Scarsini's (1984) axioms for a measure of concordance.

Start, with monotonicity of the measures with respect to the **concordance ordering** of copulas. (informally, "large" values of one variable tend to be associated with "large" values of the other).

Let C_{A_1} and C_{A_2} are EVC, then $C_{A_1} \prec C_{A_2}$ if $C_{A_1}(u,v) \leq C_{A_2}(u,v)$, $0 < u,v \leq 1$, $(u,v) \neq (1,1)$.

Proposition 2 *If C_{A_1} and C_{A_2} are EVC with $C_{A_1} \prec C_{A_2}$, then,* **the measures preserve the concordance ordering**,

$$\mathcal{CE}_{Sh}(C_{A_1}) \leq \mathcal{CE}_{Sh}(C_{A_2}) \text{ and } \mathcal{CI}_\lambda(C_{A_1}) \leq \mathcal{CI}_\lambda(C_{A_2}),\ \lambda > 0,\ \lambda \neq 1.$$

Continuity of the measures with respect to pointwise convergence: Let $(X_n, Y_n), n = 1, 2, \ldots$ be a sequence of continuous random variables with EVC C_{A_n} which converges pointwise to an extreme value copula C_A. In this setting,

$$C_{A_n}(u,v) = \exp\left[\ln(uv) \cdot A_n\left\{\frac{\ln v}{\ln(uv)}\right\}\right], \quad 0 < u, v \leq 1, \ (u,v) \neq (1,1),$$

and

$$C_A(u,v) = \exp\left[\ln(uv) \cdot A\left\{\frac{\ln v}{\ln(uv)}\right\}\right], \quad 0 < u, v \leq 1, \ (u,v) \neq (1,1),$$

for Pickands dependence functions A_n and A. Then, **Shannon and Tsallis' measures are continuous with respect to pointwise convergence of copulas**.

Proposition 3 *If $(X_n, Y_n), n = 1, 2, \ldots$ is a sequence of continuous random variables with EVC C_{A_n} which converges pointwise to an EVC C_A, then*

$$\begin{aligned}
\lim_{n\to\infty} \mathcal{CE}_{Sh}(C_{A_n}) &= \mathcal{CE}_{Sh}(C_A), \\
\lim_{n\to\infty} \mathcal{CI}_{\lambda}(C_{A_n}) &= \mathcal{CI}_{\lambda}(C_A), \ \lambda \geq 1, \\
\lim_{n\to\infty} \mathcal{CE}_{Ts,\lambda}(C_{A_n}) &= \mathcal{CE}_{Ts,\lambda}(C_A), \ \lambda \geq 1.
\end{aligned}$$

Similar continuity property is obeyed by Spearman's rho (cf. Nelsen, 2006, p. 169) which is directly connected with $\mathcal{CI}_\lambda(C_{A_n})$ and $\mathcal{CI}_\lambda(C_A)$, for $\lambda = 1$.

Invariance of copulas, under strictly increasing transformations of the corresponding continuous random variables:

Proposition 4 *Let X and Y are continuous random variables with copula C_{XY} and $\alpha(X)$ and $\beta(Y)$ are almost surely strictly monotone functions on range of X and Y respectively, then,*

$$\mathcal{CE}_{Sh}\left(C_A^{\alpha,\beta}\right) = \mathcal{CE}_{Sh}(C_A),\ \mathcal{CE}_{Ts,\lambda}\left(C_A^{\alpha,\beta}\right) = \mathcal{CE}_{Ts,\lambda}(C_A),\ \mathcal{CI}_\lambda\left(C_A^{\alpha,\beta}\right) = \mathcal{CI}_\lambda(C_A),$$

where the superscript α, β in the notation of $C_A^{\alpha,\beta}$ in the measures is used to denote the respective measure based on the EVC of $\alpha(X)$ and $\beta(Y)$.

Remark: **(i)** The above properties support the conclusion that **the measures introduced here are more related to concordance measures (like Spearman's rho) than to extent of information they contain.**

(ii) EVC are positively quadrant dependent (PQD) and the measures studied here are monotone with respect to the concordance ordering of the respective EVC. So, small values of these measures correspond to less concordant or less PQD EVC.

(iii) $\mathcal{CE}_{Sh}(C_A)$ is maximized for $A(t) = \max\{t, 1-t\}$ and the maximum value of $\mathcal{CE}_{Sh}$ is

$$2\int_0^1 \frac{\max\{t,1-t\}}{[1+\max\{t,1-t\}]^3} dt = 0.277778.$$

That is, $\mathcal{CE}_{Sh}(C_A)$ is maximized in the case of complete positive dependence, in the sense of comonotonicity.

4.2 Divergences in the EVC setting

Let A_i, $i = 1, 2$, be two Pickands dependence functions with associated bivariate extreme value copulas

$$C_{A_i}(u,v) = \exp\left[\ln(uv) \cdot A_i\left\{\frac{\ln v}{\ln(uv)}\right\}\right], \quad 0 < u, v \leq 1, \ (u,v) \neq (1,1), \ i = 1, 2.$$

Proposition 5 **(a)** **Csiszar's** *type ϕ-divergence between EVC C_{A_1} and C_{A_2},*

$$\mathcal{CD}_\phi(C_{A_1}, C_{A_2}) = \int_{[0,1]^2} \left(\ln \frac{1}{v}\right) v^{\frac{1-t+A_2(t)}{t}} \frac{1}{t^2} \phi\left(v^{\frac{A_1(t)-A_2(t)}{t}}\right) dv dt - \frac{\rho_S(A_2)+3}{12} \phi\left(\frac{\rho_S(A_1)+3}{\rho_S(A_2)+3}\right).$$

(b) *The* **Kullback-Leibler** *type extreme value copulas divergence is given by,*

$$\mathcal{CD}_{KL}(C_{A_1}, C_{A_2}) = 2\int_0^1 \frac{A_2(t)-A_1(t)}{[1+A_1(t)]^3} dt - \frac{\rho_S(A_1)+3}{12} \ln\left(\frac{\rho_S(A_1)+3}{\rho_S(A_2)+3}\right).$$

(c) *For $\lambda \in \mathbb{R}, \lambda \neq 0, -1$, the* **Cressie-Read** *λ-power type extreme value copulas divergence is given by*

$$\mathcal{CD}_\lambda(C_{A_1}, C_{A_2}) = \frac{1}{\lambda(\lambda+1)} \left(\int_0^1 \frac{1}{[1+A_1(t)+\lambda(A_1(t)-A_2(t))]^2} dt - \frac{\rho_S(A_2)+3}{12} \left(\frac{\rho_S(A_1)+3}{\rho_S(A_2)+3}\right)^{\lambda+1} \right).$$

(d) *For $a > 0$, the* **density power type extreme value copulas divergence** *is given by,*

$$\begin{aligned}\mathcal{C}d_a(C_{A_1}, C_{A_2}) \;=\;& \int_0^1 \frac{1}{[1+(1+a)A_2(t)]^2}dt + \frac{1}{a}\int_0^1 \frac{1}{[1+(1+a)A_1(t)]^2}dt \\ & - \left(1+\frac{1}{a}\right)\int_0^1 \frac{1}{[1+A_1(t)+aA_2(t)]^2}dt.\end{aligned}$$

(e) *The $CKL(C_{A_1}, C_{A_2})$ extreme value copulas divergence is given by,*

$$CKL(C_{A_1}, C_{A_2}) = \lim_{a\to 0}\mathcal{C}d_a(C_{A_1}, C_{A_2}) = 2\int_0^1 \frac{A_2(t)-A_1(t)}{[1+A_1(t)]^3}dt + \frac{1}{12}\left(\rho_S(A_2)-\rho_S(A_1)\right).$$

Remark: **(i)** All the above measures are non-negative and they attain their minimum value if and only if the underline copulas are coincide. Hence, *all these divergence type measures can by considered as quasi-distances or statistical distances between the underlined copulas.*

(ii) The value $A(t) = 1$ corresponds to *independence*. Then, anyone of these measures for $A_2(t) = 1$, lead to *a measure of the distance from independence*. Hence, *their empirical versions can be applied to develop broad classes of tests of independence.*

(iii) Cases (b) and (e) of the proposition concentrate to *two different forms of Kullback-Leibler's type EVC* divergences $\mathcal{CD}_{KL}(C_{A_1}, C_{A_2})$ and $CKL(C_{A_1}, C_{A_2})$. The first is a special case of Csiszar's type measure and the second a special case of the density power measure (cf. Zografos, 2023 for details)

Example (Gumbel extreme value copula):

Let *Gumbel copula* with $A_1(t) = [t^\theta + (1-t)^\theta]^{1/\theta}$, $\theta \geq 1$ and the *independence copula* $\Pi(u,v) = uv$ with $A_2(t) = 1$. Then, **Kullback-Leibler** type measure **between Gumbel and Independence copula** is:

$$\begin{aligned}
\mathcal{CD}_{KL}(C_{A_1}, \Pi) &= 2\int_0^1 \frac{1 - A_1(t)}{[1 + A_1(t)]^3} dt - \frac{\rho_S(A_1) + 3}{12} \ln\left(\frac{\rho_S(A_1) + 3}{\rho_S(A_2) + 3}\right) \\
&= 2\int_0^1 \frac{1 - [t^\theta + (1-t)^\theta]^{1/\theta}}{\left[1 + [t^\theta + (1-t)^\theta]^{1/\theta}\right]^3} dt - \frac{\rho_S(A_1) + 3}{12} \ln\left(\frac{\rho_S(A_1) + 3}{3}\right),
\end{aligned}$$

$$\rho_S(A_1) = 12 \int_{[0,1]^2} C_{A_1}(u,v)dudv - 3 = 12\int_0^1 \frac{1}{[1 + A_1(t)]^2} dt - 3 \text{ and } \rho_S(A_2) = 0.$$

Table gives values of $\mathcal{CD}_{KL}$ for values of the *dependence parameter* θ. Observe that the minimum value of $\mathcal{CD}_{KL}$, equal to 0, is attained in **the case of independence, $\boldsymbol{\theta = 1}$**, which is expected.

	$\mathcal{CD}_{KL}$ for values of θ						
	$\theta = 1$	$\theta = 1.5$	$\theta = 2$	$\theta = 3$	$\theta = 8$	$\theta = 15$	$\theta = 50$
$\mathcal{CD}_{KL}(C_{A_1}, \Pi)$	0	0.0025	0.0056	0.0095	0.0141	0.0149	0.0152

5. Epilogue-Conclusions

- ♦ Review of classic measures of information (m.o.i): Entropy - Divergence - Fisher-type.
- ♦ Properties of the measures - applications.
- ♦ Recent reconsiderations of the classical measures, based on the d.f.
- ♦ Cumulative entropy and divergence provide a link with copula theory.
- ♦ The reconsiderd measures was studies in the frame of EVC.

Open Problems - Future work

Let i.i.d. observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of a random vector $\boldsymbol{X}$. The dependence of the components of $\boldsymbol{X}$ is formulated from a true but unknown copula C with respective copula density c and let $\{C_\theta(\cdot) = C_\theta(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p, p \geq 1\}$ is a parametric identifiable family of candidate copulas to describe the observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and $\{c_\theta(\cdot) = c_\theta(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p, p \geq 1\}$ the respective family of copula densities.

(I) Following the previous presentation on EVC, it is open:

▲ The application of Cumulative-type m.o.i. inside broad families of copulas (elliptical copulas, Archimedean copulas, etc.) and the Characterization, the Interpretation of the Cumulative-type m.o.i. for these families of copulas.

(II) There is a direct relationship between **m.o.i. (classical or cumulative)** and a copula density $c_\theta(\cdot) = c_\theta(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}$, or a copula function $C_\theta(\cdot) = C_\theta(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}$, respectively, Then:

▲ Information theoretic methods of estimation and testing can be developed directly to the family $c_\theta(\cdot)$, by using classical m.o.i, or to the family $C_\theta(\cdot)$, by using cumulative m.o.i.

This talk was based on:

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

Zografos, K. (2024). On Entropy and Divergence Type Measures of Bivariate Extreme Value Copulas: Accepted - July 2024. REVSTAT-*Statistical Journal*.
Retrieved from https://revstat.ine.pt/index.php/REVSTAT/article/view/676

and REFERENCES appeared therein.

Arndt, C. (2001). *Information measures. Information and its description in science and engineering*. Springer-Verlag, Berlin.

Asadi, M., Ebrahimi, N. and Soofi, E. S. (2017). Connections of Gini, Fisher, and Shannon by Bayes Risk under proportional hazards. *Journal of Applied Probability* **54**, 1027-1050.

Avlogiaris, G., Micheas, A. and Zografos, K. (2019). A criterion for local model selection. *Sankhya* A **81**, 406-444.

Baratpour, S. and Rad, H. A. (2012). Testing Goodness-of-Fit for Exponential Distribution Based on Cumulative Residual Entropy. *Comm. Statist. Theory Methods* **41**, 1387-1396.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **85**, 549-559.

Basu, A., Shioya, H. and Park, C. (2011). *Statistical inference. The minimum distance approach*. Monographs on Statistics and Applied Probability, 120. CRC Press, Boca Raton, FL.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, LtdBillingsley, P. (1986). *Probability and measure*. Second edition. John Wiley & Sons, Inc., New York.

Blumentritt, T. and Schmid, F. (2012). Mutual information as a measure of multivariate association: analytical properties and statistical estimation. *J. Stat. Comput. Simul*. **82**, 1257-1274.

Broniatowski, M. and Stummer, W. (2019). Some universal insights on divergences for statistics, machine learning and artificial intelligence. *Geometric structures of information*, 149-211, Signals Commun. Technol., Springer.

Burbea, J. and Rao, C. R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J. Multivariate Anal*. **12**, 575-596.

Chen, X., Kar, S. and Ralescu, D. A. (2012). Cross-entropy measure of uncertain variables. *Inform. Sci*. **201**, 53-60.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc., Ser. B* **46**, 440-464.

De Keyser, S. and Gijbels, I. (2024). Hierarchical variable clustering via copula-based divergence measures between random vectors. *International Journal of Approximate Reasoning* **165**, 109090.

Drouet Mari, D. and Kotz, S. (2001). *Correlation and Dependence*. Imperial College Press, London.

Durante, F. and Mesiar, R. (2010). L^{∞}-measure of non-exchangeability for bivariate extreme value and Archimax copulas. *Journal of Mathematical Analysis and Applications* **369**, 610-615.

Edwards, H. H., Mikusinski, P. and Taylor, M. D. (2005). Measures of concordance determined by D4–invariant measures on $(0, 1)^2$. *Proc. Amer. Math. Soc.* **133**, 1505-1513.

Eschenburg, P. (2013). Properties of extreme-value copulas. Diplomarbeit, Technische Universität München. Available online: https://mediatum.ub.tum.de/1145695?i (accessed on 15 February 2023).

Ferentinos, K. and Papaioannou, T. (1981). New parametric measures of information. *Inform. and Control* **51**, 193-208.

Fuchs, S. (2014). Multivariate copulas: transformations, symmetry, order and measures of concordance. *Kybernetika* **50**(5), 725-743.

Fuchs, S. and K. D. Schmidt (2014). Bivariate copulas: transformations, asymmetry and measures of concordance. *Kybernetika* **50**(1), 109-125.

Fuchs, S. (2016). Copula–Induced Measures of Concordance. *Depend. Model.* **4**, 205-214.

Ghoudi, K., Khoudraji, A. and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extremes bidimensionnelles. *Canad. J. Statist.* **26**, 187-197.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of integrals, series, and products*. Seventh edition. Elsevier/Academic Press, Amsterdam.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. *Copula theory and its applications*, 127-145, Lect. Notes Stat. Proc., 198, Springer, Heidelberg, 2010.

Guiasu, S. and Reischer, C. (1985). The relative information generating function. *Inform. Sci.* **35**, 235-241.

Golomb, S. (1966). The information generating function of a probability distribution. *IEEE Trans. Inform. Theory* **12**, 75-77.

Joe, H. (2015). *Dependence modeling with copulas*. Boca Raton, FL. Chapman & Hall/CRC.

Kamnitui, N., Genest, C., Jaworski, P. and Trutschnig, W. (2019). On the size of the class of bivariate extreme-value copulas with a fixed value of Spearman's rho or Kendall's tau. *J. Math. Anal. Appl.* **472**, 920-936.

Kharazmi, O. and Balakrishnan, N. (2021b). Cumulative residual and relative cumulative residual Fisher information and their properties. *IEEE Trans. Inform. Theory* **67**, 6306-6312.

Klein, I., Mangold, B. and Doll, M. (2016). Cumulative paired ϕ-entropy. *Entropy* **18**, Paper No. 248, 45 pp.

Klein, I. and Doll, M. (2020). (Generalized) maximum cumulative direct, residual, and paired Φ entropy approach. *Entropy* **22**, Paper No. 91, 33 pp.

Liebscher, E. (2014). Copula-based dependence measures. *Depend. Model.* **2**, 49-64.

Liese, F. and Vajda, I. (1987). *Convex statistical distances*. Teubner Texts in Mathematics, Band 95. Teubner, Leipzig.

Linfoot, E. H. (1957). An informational measure of correlation. *Information and Control* **1**, 85-89.

Ma, J. and Sun, Z. (2008). Mutual information is copula entropy. arXiv:0808.0845v1 [cs.IT], 6 Aug 2008.

Micheas, A. and Zografos, K. (2006). Measuring stochastic dependence using ϕ-divergence. *J. Multivariate Anal.* **97**, 765-784.

Nair, N. U. and Sunoj, S. M. (2023). Survival Copula Entropy and Dependence in Bivariate Distributions. *REVSTAT-Statistical Journal*. https://doi.org/00.00000/revst

Nelsen, R. B. (2002). Concordance and copulas: a survey. In C. M. Cuadras, J. Fortiana, and J. A. Rodriguez-Lallena (Eds.), *Distributions with Given Marginals and Statistical Modelling*, pp. 169–177. Kluwer Academic Publishers, Dordrecht.

Nelsen, R. B. (2006). *An introduction to copulas*. Second edition. Springer Series in Statistics. Springer, New York.

Papaioannou, T. (1985). Measures of information, In *Encyclopedia of Statistical Sciences* (Eds., S. Kotz and N. L. Johnson), 5, pp. 391–397, John Wiley & Sons, New York.

Papaioannou, T. (2001). On distances and measures of information: a case of diversity, In *Probability and Statistical Models with Applications*, (Eds., C. A. Charalambides, M. V. Koutras, and N. Balakrishnan), pp. 503–515, Chapman & Hall, London.

Pardo, L. (2006). *Statistical inference based on divergence measures*. Boca Raton, FL. Chapman & Hall/CRC.

Park, S., Rao, M. and Shin, D. W. (2012). On cumulative residual Kullback-Leibler information. *Statist. Probab. Lett.* **82**, 2025-2032.

Rajesh, G. and Sunoj, S. M. (2019). Some properties of cumulative Tsallis entropy of order α. *Statist. Papers* **60**, 583-593.

Rao, M., Chen, Y., Vemuri, B. C. and Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE Trans. Inform. Theory* **50**, 1220-1228.

Read, T. R. C. and Cressie, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York.

Scarsini, M. (1984). On measures of concordance. *Stochastica* **8**(3), 201-218.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27**, 379-423, 623-656.

Song, K.-S. (2001). Rényi information, loglikelihood and an intrinsic distribution measure. *J. Statist. Plann. Inference* **93**, 51-69.

Soofi, E. S. (1994). Capturing the intangible concept of information. *J. Amer. Statist. Assoc.* **89**, 1243-1254.

Soofi, E. S. (2000). Principal information theoretic approaches. *J. Amer. Statist. Assoc.* **95**, 1349-1353.

Tawn, J. A. (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika* **77**, 245-253.

Weller-Fahy, D. J., Borghetti, B. J. and Sodemann, A. A. (2015). A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection. *IEEE Communications Surveys & Tutorials* **17**, 70-91.

Zografos, K. and Nadarajah, S. (2005). Survival exponential entropies. *IEEE Trans. Inform. Theory* **51**, 1239-1246.

Zografos, K. (2008). On Mardia's and Song's measures of kurtosis in elliptical distributions. *J. Multivariate Anal.* **99**, 858-879.

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.

Zografos, K. (2024). On Entropy and Divergence Type Measures of Bivariate Extreme Value Copulas: Accepted - July 2024. REVSTAT-*Statistical Journal*. Retrieved from https://revstat.ine.pt/index.php/REVSTAT/article/view/676

Thank you for your attention