

Cumulative Paired ϕ -Entropy

Ingo Klein ^{*,†}, Benedikt Mangold [†] and Monika Doll

Department of Statistics and Econometrics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lange Gasse 20, Nürnberg 90403, Germany; benedikt.mangold@fau.de (B.M.); monika.doll@fau.de (M.D.)

* Correspondence: ingo.klein@fau.de; Tel.: +49-911-530-2290; Fax: +49-911-530-2277

† These authors contributed equally to this work.

Academic Editor: Adom Giffin

Received: 13 April 2016; Accepted: 24 June 2016; Published: 1 July 2016

Abstract: A new kind of entropy will be introduced which generalizes both the differential entropy and the cumulative (residual) entropy. The generalization is twofold. First, we simultaneously define the entropy for cumulative distribution functions (cdfs) and survivor functions (sfs), instead of defining it separately for densities, cdfs, or sfs. Secondly, we consider a general “entropy generating function” ϕ , the same way Burbea et al. (IEEE Trans. Inf. Theory 1982, 28, 489–495) and Liese et al. (Convex Statistical Distances; Teubner-Verlag, 1987) did in the context of ϕ -divergences. Combining the ideas of ϕ -entropy and cumulative entropy leads to the new “cumulative paired ϕ -entropy” (CPE_ϕ). This new entropy has already been discussed in at least four scientific disciplines, be it with certain modifications or simplifications. In the fuzzy set theory, for example, cumulative paired ϕ -entropies were defined for membership functions, whereas in uncertainty and reliability theories some variations of CPE_ϕ were recently considered as measures of information. With a single exception, the discussions in the scientific disciplines appear to be held independently of each other. We consider CPE_ϕ for continuous cdfs and show that CPE_ϕ is rather a measure of dispersion than a measure of information. In the first place, this will be demonstrated by deriving an upper bound which is determined by the standard deviation and by solving the maximum entropy problem under the restriction of a fixed variance. Next, this paper specifically shows that CPE_ϕ satisfies the axioms of a dispersion measure. The corresponding dispersion functional can easily be estimated by an L -estimator, containing all its known asymptotic properties. CPE_ϕ is the basis for several related concepts like mutual ϕ -information, ϕ -correlation, and ϕ -regression, which generalize Gini correlation and Gini regression. In addition, linear rank tests for scale that are based on the new entropy have been developed. We show that almost all known linear rank tests are special cases, and we introduce certain new tests. Moreover, formulas for different distributions and entropy calculations are presented for CPE_ϕ if the cdf is available in a closed form.

Keywords: ϕ -entropy; absolute mean deviation; cumulative residual entropy; measure of dispersion; generalized maximum entropy principle; Tukey’s λ distribution; ϕ -regression; L -estimator; linear rank test

1. Introduction

The ϕ -entropy

$$E_\phi(F) = \int_{\mathbb{R}} \phi(f(x)) dx, \quad (1)$$

where f is a probability density function and ϕ is a strictly concave function, was introduced by [1]. If we set $\phi(u) = -u \ln u$, $u \in [0, 1]$, we get Shannon’s differential entropy as the most prominent special case.

Shannon et al. [2] derived the “entropy power fraction” and showed that there is a close relationship between Shannon entropy and variance. In [3], it was demonstrated that Shannon’s

differential entropy satisfies an ordering of scale and thus is a proper measure of scale (MOS). Recently, the discussion in [4] has shown that entropies can be interpreted as a measure of dispersion. In the discrete case, minimal Shannon entropy means maximal certainty about the random outcome of an experiment. A degenerate distribution minimizes the Shannon entropy as well as the variance of a discrete quantitative random variable. For such a degenerate distribution, Shannon entropy and variance both take the value 0. However, there is an important difference between the differential entropy and the variance when discussing a discrete quantitative random variable with support $[a, b]$. The differential entropy is maximized by a uniform distribution over $[a, b]$, while the variance is maximal if both interval bounds a and b have the probability mass of 0.5 (cf. [5]). A similar result holds for a discrete random variable with a finite number of realizations. Therefore, it is doubtful that Equation (1) is a true measure of dispersion.

We propose to define the ϕ -entropy for cumulative distribution functions (cdfs) F and survivor functions (sf) $1 - F$ instead of for density functions f . Throughout the paper, we define $\bar{F} := 1 - F$. By applying this modification we get

$$CPE_{\phi}(F) = \int_{\mathbb{R}} \phi(F(x)) + \phi(\bar{F}(x)) dx, \quad (2)$$

where cdf F is absolutely continuous, CPE means “cumulative paired entropy”, and ϕ is the “entropy generating function” defined on $[0, 1]$ with $\phi(0) = \phi(1) = 0$. We will assume that ϕ is concave on $[0, 1]$ throughout most of this paper. In particular, we will show that Equation (2) satisfies a popular ordering of scale and attains its maximum if the domain is an interval $[a, b]$, while a, b occur with a probability of $1/2$. This means that Equation (2) behaves like a proper measure of dispersion.

In addition, we generalize results from the literature, focusing on the Shannon case with $\phi(u) = -u \ln u$, $u \in [0, 1]$ (cf. [6]), the cumulative residual entropy

$$CRE(F) = - \int_{\mathbb{R}^+} \bar{F}(x) \ln \bar{F}(x) dx \quad (3)$$

(cf. [7]), and the cumulative entropy

$$CE(F) = - \int_{\mathbb{R}} F(x) \ln F(x) dx \quad (4)$$

(cf. [8,9]). In the literature, this entropy is interpreted as a measure of information rather than dispersion without any clarification on what kind of information is considered.

A first general aim of this paper is to show that entropies can rather be interpreted as measures of dispersion than as measures of information. A second general aim is to demonstrate that the entropy generating function ϕ , the weight function J in L -estimation, the dispersion function d which serves as a criterion for minimization in robust rank regression, and the scores-generating function φ_1 are closely related.

Specific aims of this paper are:

1. To show that the cdf-based entropy Equation (2) originates in several distinct scientific areas.
2. To demonstrate the close relationship between Equation (2) and the standard deviation.
3. To derive maximum entropy (ME) distributions under simple and more complex restrictions and to show that commonly known as well as new distributions solve the ME principle.
4. To derive the entropy maximized by a given distribution under certain restrictions.
5. To formally prove that Equation (2) is a measure of dispersion.
6. To propose an L -estimator for Equation (2) and derive its asymptotic properties.
7. To use Equation (2) in order to obtain new related concepts measuring the dependence of random variables (such as mutual ϕ -information, ϕ -correlation, and ϕ -regression).
8. To apply Equation (2) to get new linear rank tests for the comparison of scale.

The paper is structured in the same order as these aims. After this introduction, in the second section we give a short review of the literature that is concerned with Equation (2) or related measures. The third section begins by summarizing reasons for defining entropies for cdfs and sfs instead of defining them for densities. Next, some equivalent characterizations of Equation (2) are given, provided the derivative of ϕ exists. In the fourth section, we use the Cauchy–Schwarz inequality to derive an upper bound for Equation (2), which provides sufficient conditions for the existence of CPE . In addition, more stringent conditions for the existence are directly proven. In the fifth section, the Cauchy–Schwarz inequality allows to derive ME distributions if the variance is fixed. For more complicated restrictions we attain ME distributions by solving the Euler–Lagrange conditions. Following the generalized ME principle (cf. [10]), we change the perspective and ask which entropy is maximized if the variance and the population’s distribution is fixed. The sixth section is of key importance because the properties of Equation (2) as a measure of dispersion is analyzed in detail. We show that Equation (2) satisfies an often applied ordering of scale by [3], is invariant with respect to translations and equivariant with respect to scale transformations. Additionally, we provide certain results concerning the sum of independent random variables. In the seventh section, we propose an L -estimator for CPE_ϕ . Some basic properties of this estimator like influence function, consistency, and asymptotic normality are shown. In the eighth section, we introduce several new statistical concepts based on CPE_ϕ , which are generalizing divergence, mutual information, Gini correlation, and Gini regression. Additionally, we show that new linear rank tests for dispersion can be based on CPE_ϕ . The known linear rank tests like the Mood- or the Ansari-Bradley tests are special cases of this general approach. However, in this paper we exclude most of the technical details for they will be presented in several accompanying papers. In the last section we compute Equation (2) for certain generating functions ϕ and some selected families of distributions.

2. State of the Art—An Overview

Entropies are usually defined on the simplex of probability vectors, which are summing up to one (cf. [2,11]). Until now it has been rather usual to calculate the Shannon entropy not for vectors of probability or probability density functions f , but for distribution functions F . The corresponding Shannon entropy is given by

$$CPE_S(F) = - \int_{\mathbb{R}} F(x) \ln F(x) + \bar{F}(x) \ln \bar{F}(x) dx. \quad (5)$$

Nevertheless, we have identified five scientific disciplines directly or implicitly working with an entropy based on distribution functions or survivor functions:

1. Fuzzy set theory,
2. Generalized ME principle,
3. Theory of dispersion of ordered categorical variables,
4. Uncertainty theory,
5. Reliability theory.

2.1. Fuzzy Set Theory

To the best of our knowledge, Equation (5) was initially introduced by [12]. However, they did not consider the entropy for a cdf F . Instead, they were concerned with a so-called membership function μ_A that quantifies the degree to which a certain element x of a set Ω belongs to a subset $A \subseteq \Omega$. Membership functions were introduced by [13] within the framework of the “fuzzy set theory”.

It is important to note that if all elements of Ω are mapped to the value $1/2$, maximum uncertainty about x belonging to a set A will be attained.

This main property is one of the axioms of membership functions. In the aftermath of [12] numerous modifications to the term “entropy” have been made and axiomatizations of the membership functions have been stated (see, e.g., the overview in [14]).

Finally, those modifications proceeded parallel to a long history of extensions and parametrizations of the term entropy for probability vectors and densities. It began with [15] up to [16,17], who provided a superstructure of those generalizations consisting of a very general form of the entropy, including the ϕ -entropy Equation (1) as a special case. Burbea et al. [1] introduced the term ϕ -entropy. If both $\phi(x)$ and $\phi(1-x)$ appeared in the entropy, as in the Fermi-Dirac entropy (cf. [18], p. 191), they used the term “paired” ϕ -entropy.

2.2. Generalized Maximum Entropy Principle

Regardless of the debate in the fuzzy set theory and the theory of measurement of dispersion, Kapur [10] showed that a growth model with a logistic growth rate is yielded as the solution of maximizing Equation (5) under two simple constraints. This provides an example for the “generalized maximum entropy principle” postulated by Kesavan et al. [19]. In addition to that, the simple ME principle introduced by [20,21] derives a distribution which maximizes an entropy given certain constraints. Furthermore, the generalization of [19] consists of determining the ϕ -entropy, which is maximized given a distribution and some constraints. Finally, they used a slightly modified formula Equation (5). The cdf had to be replaced by a monotonically increasing function with logistic shape.

2.3. Theory of Dispersion

Irrespectively of the discussion on membership functions in the fuzzy set theory and the proposals of generalizing the Shannon entropy, Leik [22] discussed a measure of dispersion for ordered categorical variables with a finite number k of categories $x_1 < x_2 < \dots < x_k$. His measure is based on the distance between the $k-1$ -dimensional vectors of cumulated frequencies $(F_1, F_2, \dots, F_{k-1})$ and $(1/2, 1/2, \dots, 1/2)$. Both vectors only coincide if the extreme categories x_1 and x_k appear with same frequency. This represents the case of maximal dispersion. Consider

$$CPE_{\phi}(F) = \sum_{i=1}^{k-1} (\phi(F_i) + \phi(1 - F_i)) \quad (6)$$

as discrete version of Equation (2). Setting $\phi(u) = \min\{u, 1-u\}$, we get the measure of Leik as a special case of Equation (6) up to a change of sign. Vogel et al. [23] considered $\phi(u) = -u \ln(u)$ and the Shannon variation of Equation (6) as measure of dispersion for ordered categorical variables. Numerous modifications of Leik’s measure of dispersion have been published. In [24–29], the authors implicitly used $\phi(u) = 1/4 - (u - 1/2)^2$ or equivalently $\phi(u) = u(1-u)$. Most of the discussion was conducted in the journal “Perceptual and Motor Skills”. For a recent overview of measuring dispersion including ordered categorical variables see, e.g., [30]. Instead of dispersion, some articles are concerned with related concepts for ordered categorical variables, like bipolarization and inequality (cf. [31–35]). A class of measures of dispersion for ordered categorical variables with a finite number of categories that is similar to Equation (6) has been introduced by Klein and Yager [36,37] independently of each other. They had obviously not been aware of the discussion in “Perceptual and Motor Skills”. Both authors gave axiomatizations to describe which functions ϕ will be appropriate for measuring dispersion. However, at least Yager [37] recognized the close relationship between those measures and the general term “entropy” in the fuzzy set theory. He introduced the term “dissonance” to more precisely characterize measures of dispersion for ordered categorical variables. In the language of information theory, maximum dissonance describes an extreme case in which there is still some information. But this information is extremely contradictory. As an example, we could ask in the field of product evaluation to what degree information, which states that 50 percent of the recommendations are extremely good and at the same time 50 percent are extremely bad, is useful to make a purchase decision. This is an important difference to the Shannon entropy, which is maximal if there is no information at all, i.e., all categories occur with same probability.

Bowden [38] defines the location entropy function $h(x) = -F(x) \ln F(x) + \bar{F}(x) \ln \bar{F}(x)$, given a value of x . He emphasizes the possibility to construct measures of spread and symmetry based on this function. To the best of our knowledge, Bowden [38] is the only one to mention the application of cumulated paired Shannon entropy to continuous distributions so far.

2.4. Uncertainty Theory

Reference ([6] (first edition 2004)) can be considered the founder of the uncertainty theory. This theory is concerned with formalizing data consisting of expert opinions rather than formalizing data gathered by repeating a random experiment. Liu slightly modified the Kolmogoroff axioms of probability theory to receive an uncertainty measure, following which he defined uncertain variables, uncertainty distribution functions, and moments of uncertain variables. Liu argued that “an event is the most uncertain if its uncertainty measure is 0.5, because the event and its complement may be regarded as ‘equally likely’ ” ([6], p. 14). Liu’s maximum uncertainty principle states: “For any event, if there are multiple reasonable values that an uncertain measure may take, the value as close to 0.5 as possible is assigned to the event” [6] (p. 14). Similar to the fuzzy set theory, the distance between the uncertainty distribution and the value 0.5 can be measured by the Shannon-type entropy Equation (5). Apparently for the first time in the third edition of 2010, he explicitly calculated Equation (5) for several distributions (e.g., the logistic distribution) and derived upper bounds. He applied the ME principle to uncertainty distributions. The preferred constraint is to predetermine values of mean and variance ([6], p. 83ff.). In this case, the logistic distribution maximizes Equation (5). In this context, the logistic distribution plays the same role in uncertainty theory as the Gaussian distribution in probability theory. The Gaussian distribution maximizes the differential entropy, given values for mean and variance. Therefore, in uncertainty theory the logistic distribution is called “normal The authors of distribution”. [39] provided Equation (5) as a function of the quantile function. In addition to that, the authors of [40] chose $\phi(u) = u(1 - u)$, $u \in [0, 1]$, as entropy generating function and derived the ME distribution as a discrete uniform distribution, which is concentrated on the endpoints of the compact domain $[a, b]$ if no further restrictions are assumed. Popoviciu [5] attained the same distribution by maximizing the variance. Chen et al. [41] introduced cross entropies and divergence measures based on general functions ϕ . Further literature on this topic is provided by [42–44].

2.5. Reliability Theory

Entropies also play a prominent role in reliability theory. They were initially introduced in the fields of hazard rates and residual lifetime distributions (cf. [45]). In addition, the authors of [46,47] introduced the cumulative residual entropy Equation (3), discussed its properties, and derived the exponential and the Weibull distribution by an ME principle, given the coefficient of variation. This work went into detail on the advantage of defining entropy via survivor functions instead of probability density functions. Rao et al. [46] refer to the extensive criticism on the differential entropy by [48]. Moreover, Zografos et al. [49] generalized the Shannon-type cumulative residual entropy to an entropy of the Rényi type. Furthermore, Drissi et al. [50] considered random variables with general support. They also presented solutions for the maximization of Equation (3), provided that more general restrictions are considered. Similar to [51], they identified the logistic distribution to be the ME distribution, given mean, variance, and a symmetric form of the distribution function.

Di Crescenzo et al. [9] analyzed Equation (4) for cdfs and discussed its stochastic properties. Sunoj et al. [52] plugged the quantile function into the Shannon-type entropy Equation (4) and presented expressions if the quantile function possesses a closed form, but not the cdf. In recent papers an empirical version of Equation (3) is used as goodness-of-fit test (cf. [53]).

Additionally, CRE and CE are applied to the distribution function of the residual lifetime $(X - t | X > t)$ and the inactivity time $(t - X | X < t)$ (cf. [54]). This can directly be generalized to the CPE framework.

Moreover, Psarrakos et al. [55] provides an interesting alternative generalization of the Shannon case. In this paper we focus on the class of concave functions ϕ . Special extensions to non-concave functions will be subject to future research.

This brief overview shows that different disciplines are accessing an entropy based on distribution functions. The contributions of the fuzzy set theory, the uncertainty theory, and the reliability theory all have the exclusive consideration of continuous random variables in common. The discussions about entropy in reliability theory on the one hand and fuzzy set theory and uncertainty theory, respectively, on the other hand were conducted independently of each other without even noticing the results of the other disciplines. However, Liu's uncertainty theory benefits from the discussion in the fuzzy set theory. In the theory of dispersion of ordered categorical variables the authors do not appear to be aware of their implicit use a concept of entropy. Nevertheless the situation is somewhat different to that of the other areas since only discrete variables were discussed. Kiesel's dissertation [56] provides a theory of measures of the form Equation (6) with numerous applications. However, an intensive discussion of Equation (2) is missing and will be provided here.

3. Cumulative Paired ϕ -Entropy for Continuous Variables

3.1. Definition

We focus on absolute continuous cdfs F with density functions f . The set of all those distribution functions is called \mathcal{F} . We call a function "entropy generating function" if it is non-negative and concave on the domain $[0, 1]$ with $\phi(0) = \phi(1) = 0$. In this case, $\phi(u) + \phi(1 - u)$ is a symmetric function with respect to $1/2$.

Definition 1. The functional $CPE_\phi : \mathcal{F} \rightarrow \mathbb{R}_0^+$ with

$$CPE_\phi(F) = \int_{\mathbb{R}} \phi(F(x)) + \phi(\bar{F}(x)) dx \quad (7)$$

is called cumulative paired ϕ -entropy for $F \in \mathcal{F}$ with entropy generating function ϕ .

Up to now, we assumed the existence of CPE_ϕ . In the following section we will discuss some sufficient criteria ensuring the existence of CPE_ϕ . If X is a random variable with cdf F , we occasionally use the notation $CPE_\phi(X)$ instead.

Next, some examples of well established concave entropy generating functions ϕ and corresponding cumulative paired ϕ -entropies will be given.

1. Cumulative paired α -entropy CPE_α : Following [57], let ϕ be given by

$$\phi(u) = u \frac{u^{\alpha-1} - 1}{1 - \alpha}, \quad u \in [0, 1],$$

for $\alpha > 0$. The corresponding so-called cumulative paired α -entropy is

$$CPE_\alpha(F) = \int_{\mathbb{R}} F(x) \frac{F(x)^{\alpha-1} - 1}{1 - \alpha} + \bar{F}(x) \frac{\bar{F}(x)^{\alpha-1} - 1}{1 - \alpha} dx. \quad (8)$$

2. Cumulative paired Gini entropy CPE_G : For $\alpha = 2$ we get

$$CPE_G(F) = 2 \int_{\mathbb{R}} F(x) \bar{F}(x) dx \quad (9)$$

as a special case of CPE_α .

3. Cumulative paired Shannon entropy CPE_S : Set $\phi(u) = -u \ln u$, $u \in [0, 1]$. Thus,

$$CPE_S(F) = - \int_{\mathbb{R}} F(x) \ln F(x) + \bar{F}(x) \ln \bar{F}(x) dx \quad (10)$$

gives the entropy which was already mentioned in the introduction. It is a special case of CPE_α for $\alpha \rightarrow 1$.

4. Cumulative paired Leik entropy CPE_L : The function

$$\phi(u) = \min\{u, 1-u\} = \frac{1}{2} - \left|u - \frac{1}{2}\right|, \quad u \in [0, 1],$$

represents the limiting case of a linear concave function ϕ . The measure of dispersion proposed by [22] implicitly makes use of ϕ such that we call

$$CPE_L(F) = 2 \int_{\mathbb{R}} \min\{F(x), \bar{F}(x)\} dx \quad (11)$$

cumulative paired Leik entropy.

Figure 1 gives an impression of the previously mentioned generating functions ϕ .

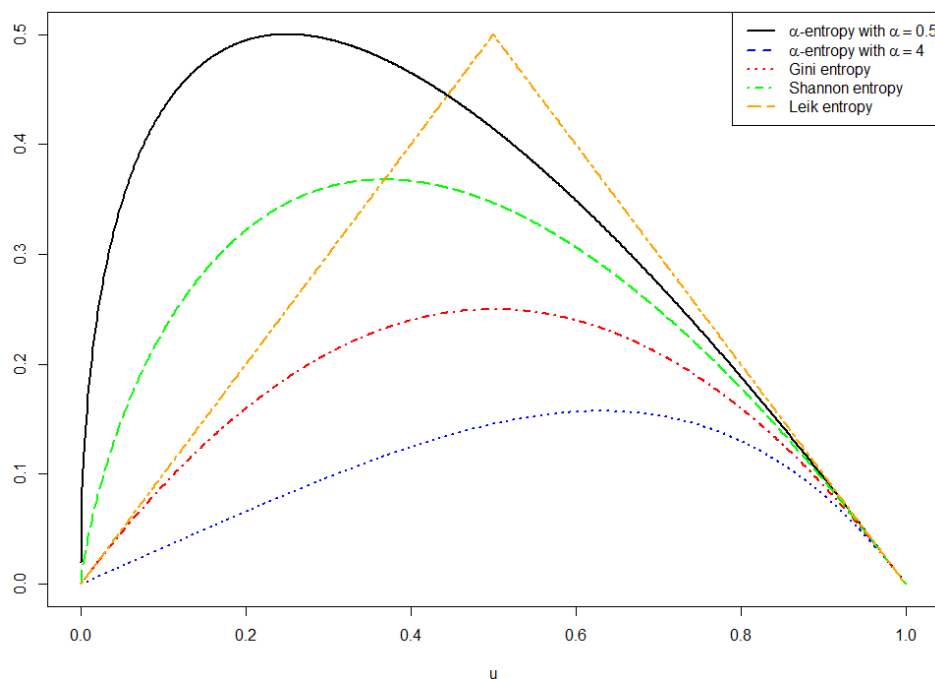


Figure 1. Some entropy generating functions ϕ .

3.2. Advantages of Entropies Based on Cdfs

The authors of [46,47] list several reasons for better defining an entropy for distribution functions rather than defining it for density functions. Starting point is the well-known critique of Shannon's differential entropy $-\int f(x) \ln f(x) dx$ that was expressed by several authors like [48,58] and (p. 58f) in [59].

Transferred to cumulative paired entropies, the advantages of entropies based on distribution functions (cf. [46]) are as follows:

1. CPE_ϕ is based on probabilities and has a consistent definition for both discrete and continuous random variables.
2. CPE_ϕ is always non-negative.
3. CPE_ϕ can easily be estimated by the empirical distribution function. This estimation is strongly consistent, due to the strong consistency of the empirical distribution function.

Problems of the differential entropy are occasionally discussed in case of grouped data, at which the usual Shannon entropy is calculated for each group probability. With an increasing amount of groups, the Shannon entropy not only does not converge to the respective differential entropy, but it even diverges (cf., e.g., (p. 54) in ([59], (p. 239) in [60]). In the next section we will show that the discrete version of CPE_ϕ converges to CPE_ϕ as the number of groups approaches infinity.

3.3. CPE_ϕ for Grouped Data

First, we show the notation for characterizing grouped data. The interval $[\tilde{x}_0, \tilde{x}_k]$ is divided into k subintervals with limits $\tilde{x}_0 < \tilde{x}_1 < \dots < \tilde{x}_{k-1} < \tilde{x}_k$. The range of each group is called $\Delta x_i = \tilde{x}_i - \tilde{x}_{i-1}$ for $i = 1, 2, \dots, k$. Let X be a random variable with absolute continuous distribution function F , which is only known at the limits of each group. The probabilities of each group are denoted by $p_i = F(\tilde{x}_i) - F(\tilde{x}_{i-1})$, $i = 1, 2, \dots, k$. X^* is the random variable whose distribution function F^* is yielded by linear interpolation of the values of F at the limits of successive groups. Finally, X^* is the result of adding an independent, uniformly distributed random variable to X . It holds that

$$F^*(x) = F(\tilde{x}_{i-1}) + \frac{p_i}{\Delta x_i}(x - \tilde{x}_{i-1}) \quad \text{if } \tilde{x}_{i-1} < x \leq \tilde{x}_i \quad (12)$$

for $x \in \mathbb{R}$, $F^*(x) = 0$ for $x \leq \tilde{x}_0$ and $F^*(x) = 1$ for $x > \tilde{x}_k$.

Let X^* denote the respective random variable of F^* . The probability density function f^* of X^* is defined by $f^*(x) = p_i/\Delta x_i$ for $\tilde{x}_{i-1} < x \leq \tilde{x}_i$, $i = 1, 2, \dots, k$.

Lemma 1. Let ϕ be an entropy generating function with antiderivative S_ϕ . The paired cumulative ϕ -entropy of the distribution function in Equation (12) is given as follows:

$$CPE_\phi(X^*) = \sum_{i=1}^k \frac{\Delta x_i}{p_i} (S_\phi(F(\tilde{x}_i)) - S_\phi(F(\tilde{x}_{i-1})) + S_\phi(\bar{F}(\tilde{x}_{i-1})) - S_\phi(\bar{F}(\tilde{x}_i))). \quad (13)$$

Proof. For $x \in (\tilde{x}_{i-1}, \tilde{x}_i]$, we have

$$F^*(x) = a_i + b_i x \quad \text{with } b_i = \frac{p_i}{\Delta x_i} \text{ and } a_i = F(\tilde{x}_{i-1}) - b_i \tilde{x}_{i-1}$$

with $a_i + b_i \tilde{x}_{i-1} = F(\tilde{x}_{i-1})$, $a_i + b_i \tilde{x}_i = F(\tilde{x}_i)$, $1 - a_i - b_i \tilde{x}_{i-1} = \bar{F}(\tilde{x}_{i-1})$, and $1 - a_i - b_i \tilde{x}_i = \bar{F}(\tilde{x}_i)$, $i = 1, 2, \dots, k$. With $y = a_i + b_i x$ and $dx = 1/b_i dy$ we have

$$\begin{aligned} CPE_\phi(X^*) &= \sum_{i=1}^k \int_{\tilde{x}_{i-1}}^{\tilde{x}_i} \phi(a_i + b_i x) + \phi(1 - a_i - b_i x) dx \\ &= \sum_{i=1}^k \frac{1}{b_i} \int_{F(\tilde{x}_{i-1})}^{F(\tilde{x}_i)} \phi(y) + \phi(1 - y) dy = \sum_{i=1}^k \frac{\Delta x_i}{p_i} \left(\int_{F(\tilde{x}_{i-1})}^{F(\tilde{x}_i)} \phi(y) dy - \int_{\bar{F}(\tilde{x}_{i-1})}^{\bar{F}(\tilde{x}_i)} \phi(y) dy \right) \\ &= \sum_{i=1}^k \frac{\Delta x_i}{p_i} \left(\int_{F(\tilde{x}_{i-1})}^{F(\tilde{x}_i)} \phi(y) dy + \int_{\bar{F}(\tilde{x}_i)}^{\bar{F}(\tilde{x}_{i-1})} \phi(y) dy \right) \\ &= \sum_{i=1}^k \frac{\Delta x_i}{p_i} (S_\phi(F(\tilde{x}_i)) - S_\phi(F(\tilde{x}_{i-1})) + S_\phi(\bar{F}(\tilde{x}_{i-1})) - S_\phi(\bar{F}(\tilde{x}_i))). \quad \square \end{aligned}$$

Considering this result, we can easily prove the convergence property for $CPE_\phi(X^*)$:

Theorem 1. Let ϕ be a generating function with antiderivative S_ϕ and let F be a continuous distribution function of the random variable X with support $[a, b]$. X^* is the corresponding random variable for grouped data with $\Delta x = (b - a)/k$, $k > 0$. Then the following holds:

$$CPE_\phi(X^*) \rightarrow \int_a^b \phi(F(x)) + \phi(\bar{F}(x)) dx \text{ for } k \rightarrow \infty.$$

Proof. Consider equidistant classes with $\Delta x_i = \Delta x = (b - a)/k$, $i = 1, 2, \dots, k$. Subsequently, Equation (13) results in

$$CPE_\phi(X^*) = \sum_{i=1}^k \left(\frac{S_\phi(F(\tilde{x}_i)) - S_\phi(F(\tilde{x}_{i-1}))}{F(\tilde{x}_i) - F(\tilde{x}_{i-1})} + \frac{S_\phi(\bar{F}(\tilde{x}_{i-1})) - S_\phi(\bar{F}(\tilde{x}_i))}{F(\tilde{x}_i) - F(\tilde{x}_{i-1})} \right) \Delta x. \quad (14)$$

With $k \rightarrow \infty$ we have $\Delta x \rightarrow 0$ such that for F continuous we get $F(\tilde{x}_i) - F(\tilde{x}_{i-1}) \rightarrow 0$. The antiderivative S_ϕ has the derivative ϕ almost everywhere such that with $k \rightarrow \infty$

$$\sum_{i=1}^k \frac{S_\phi(F(\tilde{x}_i)) - S_\phi(F(\tilde{x}_{i-1}))}{F(\tilde{x}_i) - F(\tilde{x}_{i-1})} \Delta x \rightarrow \int_a^b \phi(F(x)) dx.$$

An analogue argument holds for the second term of Equation (14). \square

In addition to this theoretical result we get the following useful expressions for CPE_ϕ for grouped data and a specific choice of ϕ as Corollary 1 shows:

Corollary 1. Let ϕ be s.t.

$$\phi(u) = \begin{cases} -u \ln u & \text{for } \alpha = 1 \\ -u^{\frac{1-\alpha}{1-\alpha}} & \text{for } \alpha \neq 1, \end{cases}$$

where $u \in [0, 1]$. Then for $\alpha = 1$

$$CPE_S(X^*) = -\frac{1}{2} \sum_{i=1}^k \frac{\Delta x_i}{p_i} \left(F(\tilde{x}_i)^2 \ln F(\tilde{x}_i) - F(\tilde{x}_{i-1})^2 \ln F(\tilde{x}_{i-1}) \right) - \frac{1}{2} \sum_{i=1}^k \frac{\Delta x_i}{p_i} \left(\bar{F}(\tilde{x}_{i-1})^2 \ln \bar{F}(\tilde{x}_{i-1}) - \bar{F}(\tilde{x}_i)^2 \ln \bar{F}(\tilde{x}_i) \right) + \frac{1}{2} (\tilde{x}_k - \tilde{x}_0)$$

and for $\alpha \neq 1$

$$CPE_\alpha(X^*) = \frac{1}{1-\alpha} \sum_{i=1}^k \frac{\Delta x_i}{p_i} \left(\frac{1}{\alpha+1} \left(F(\tilde{x}_i)^{\alpha+1} - F(\tilde{x}_{i-1})^{\alpha+1} \right) + \bar{F}(\tilde{x}_{i-1})^{\alpha+1} - \bar{F}(\tilde{x}_i)^{\alpha+1} \right) - (\tilde{x}_k - \tilde{x}_0).$$

Proof. Using the antiderivatives

$$S_\alpha(u) = \begin{cases} -\frac{1}{2} u^2 \ln u + \frac{1}{4} u^2 & \text{for } \alpha = 1 \\ \frac{1}{1-\alpha} \left(\frac{1}{\alpha+1} u^{\alpha+1} - \frac{1}{2} u^2 \right) & \text{for } \alpha \neq 1, \end{cases}$$

since $p_i = F(\tilde{x}_i) - F(\tilde{x}_{i-1})$, it holds that

$$\begin{aligned} & \frac{1}{p_i} \left(F(\tilde{x}_i)^2 - F(\tilde{x}_{i-1})^2 + \bar{F}(\tilde{x}_{i-1})^2 - \bar{F}(\tilde{x}_i)^2 \right) \\ &= \frac{(F(\tilde{x}_i) - F(\tilde{x}_{i-1}))(F(\tilde{x}_i) + F(\tilde{x}_{i-1}))}{F(\tilde{x}_i) - F(\tilde{x}_{i-1})} \\ &+ \frac{(\bar{F}(\tilde{x}_{i-1}) - \bar{F}(\tilde{x}_i))(\bar{F}(\tilde{x}_{i-1}) + \bar{F}(\tilde{x}_i))}{F(\tilde{x}_i) - F(\tilde{x}_{i-1})} = 2 \end{aligned}$$

for $i = 1, 2, \dots, k$. The results follow immediately. \square

3.4. Alternative Representations of CPE_ϕ

In case $\phi(0) = \phi(1) = 0$ holds and ϕ is differentiable, one can provide several alternative representations of CPE_ϕ in addition to Equation (7). These alternative representations will be useful in the following to find conditions ensuring the existence of CPE_ϕ and to find some simple estimators.

Proposition 1. Let ϕ be a non-negative and differentiable function on the domain $[0, 1]$ with derivative ϕ' and $\phi(0) = \phi(1) = 0$. In this case, for $F \in \mathcal{F}$ with quantile function $F^{-1}(u)$, density function f , and quantile density function $q(u) = 1/f(F^{-1}(u))$, for $u \in [0, 1]$, the following holds:

$$CPE_\phi(F) = \int_0^1 \phi(u) + \phi(1-u)q(u)du, \quad (15)$$

$$CPE_\phi(F) = \int_0^1 (\phi'(1-u) - \phi'(u))F^{-1}(u)du, \quad (16)$$

$$CPE_\phi(F) = \int_{\mathbb{R}} x(\phi'(\bar{F}(x)) - \phi'(F(x)))f(x)dx. \quad (17)$$

Proof. Apply probability integral transform $U = F(X)$ and partial integration. \square

Due to $\phi(0) = \phi(1) = 0$ it holds that

$$\int_0^1 (\phi'(1-u) - \phi'(u))du = 0.$$

This property supports the understanding of CPE_ϕ being a covariance for which the Cauchy-Schwarz inequality gives an upper bound:

Corollary 2. Let ϕ be a non-negative and differentiable function on the domain $[0, 1]$ with derivative ϕ' and $\phi(0) = \phi(1) = 0$. Then if U is uniformly distributed on $[0, 1]$ and $X \sim F$:

$$CPE_\phi(F) = \text{Cov}(\phi'(1-U) - \phi'(U), F^{-1}(U)), \quad (18)$$

$$CPE_\phi(F) = \text{Cov}(\phi'(\bar{F}(X)) - \phi'(F(X)), X). \quad (19)$$

Proof. Let $\mu = E[X]$, then since $E[\phi'(1-U) - \phi'(U)] = 0$,

$$\begin{aligned} CPE_\phi(F) &= \int_0^1 (\phi'(1-u) - \phi'(u))F^{-1}(u)du \\ &= \int_0^1 (\phi'(1-u) - \phi'(u))(F^{-1}(u) - \mu)du. \quad \square \end{aligned}$$

Depending on the context, we switch between these alternative representations of CPE_ϕ .

4. Sufficient Conditions for the Existence of CPE_ϕ

4.1. Deriving an Upper Bound for CPE_ϕ

The Cauchy–Schwarz inequality for Equations (18) and (19), respectively, provides an upper bound for CPE_ϕ if the variance $\sigma^2 = E[(F^{-1}(u) - \mu)^2]$ exists and

$$\int_0^1 (\phi'(1-u) - \phi'(u))^2 du < \infty \quad (20)$$

holds. The existence of the upper bound simultaneously ensures the existence of CPE_ϕ .

Proposition 2. Let ϕ be a non-negative and differentiable function on the domain $[0, 1]$ with derivative ϕ' and $\phi(0) = \phi(1) = 0$. If Equation (20) holds, then for $X \sim F$ with $\text{Var}(X) < \infty$ and quantile function F^{-1} , we have

$$CPE_\phi(F) \leq \sqrt{E((\phi'(1-U) - \phi'(U))^2) E((F^{-1}(U) - \mu)^2)} \quad (21)$$

$$CPE_\phi(F) \leq \sqrt{E((\phi'(\bar{F}(X)) - \phi'(F(X)))^2) \sigma^2}. \quad (22)$$

Proof. The statement follows from

$$\begin{aligned} \left(E((\phi'(1-U) - \phi'(U))(F^{-1}(U) - \mu)) \right)^2 &\leq \int_0^1 (\phi'(1-u) - \phi'(u))^2 du \\ &\quad \times E((F^{-1}(U) - \mu)^2). \quad \square \end{aligned}$$

Next, we consider the upper bound for the cumulative paired α -entropy:

Corollary 3. Let X be a random variable having a finite variance. Then

$$CPE_\alpha(X) \leq \sigma \left| \frac{\alpha}{1-\alpha} \right| \sqrt{2 \left(\frac{1}{2\alpha-1} - B(\alpha, \alpha) \right)} \quad (23)$$

for $\alpha > 1/2$, $\alpha \neq 1$ with

$$CPE_S(X) \leq \frac{\pi\sigma}{\sqrt{3}} \quad (24)$$

for $\alpha = 1$.

Proof. For $\phi(u) = u(u^{\alpha-1} - 1)/(1-\alpha)$ and $\phi'(u) = (\alpha u^{\alpha-1} - 1)/(1-\alpha)$, $u \in [0, 1]$, we have

$$\begin{aligned} \int_0^1 (\phi'(1-u) - \phi'(u))^2 du &= \left(\frac{\alpha}{1-\alpha} \right)^2 \left(\int_0^1 (u^{\alpha-1} - (1-u)^{\alpha-1})^2 du \right) \\ &= 2 \left(\frac{\alpha}{1-\alpha} \right)^2 \left(\int_0^1 u^{2(\alpha-1)} du \right. \\ &\quad \left. - 2 \int_0^1 u^{\alpha-1} (1-u)^{\alpha-1} du \right) \\ &= 2 \left(\frac{\alpha}{1-\alpha} \right)^2 \left(\frac{1}{2\alpha-1} - B(\alpha, \alpha) \right). \end{aligned}$$

$\alpha > 1/2$ is required for the existence of $CPE_\alpha(X)$. For $\alpha = 1$ we have $\phi(u) = -u \ln u$ and $\phi'(u) = -\ln u - 1$, $u \in [0, 1]$, such that

$$\int_0^1 (\phi'(1-u) - \phi'(u))^2 du = \int_0^1 \left(\ln \left(\frac{1-u}{u} \right) \right)^2 du = \frac{\pi^2}{3}. \quad \square$$

In the framework of uncertainty theory, the upper bound for the paired cumulative Shannon entropy was derived by [51] (see also [6], p. 83). For $\alpha = 2$ we get the upper bound for the paired cumulative Gini entropy

$$CPE_G(X) \leq \sigma \frac{2}{\sqrt{3}}. \quad (25)$$

This result has already been proven for non-negative uncertainty variables by [40]. Finally, one yields the following upper bound for the paired cumulative Leik entropy:

Corollary 4. *Let X be a random variable with existing variance. Then*

$$CPE_L[X] \leq 2\sigma. \quad (26)$$

Proof. Use

$$\int_0^1 (\text{sign}(u - 1/2) - \text{sign}(1/2 - u))^2 du = 4$$

to get the result. \square

4.2. Stricter Conditions for the Existence of CPE_α

So far, we only considered sufficient conditions for an existing variance. Following the arguments in [46,50], which were used for the special case of cumulative residual and residual Shannon entropy, one can derive stricter sufficient conditions for the existence of CPE_α .

Theorem 2. *If $E(|X|^p) < \infty$ for $p > 1$, then $CPE_\alpha < \infty$ for $\alpha > 1/p$.*

Proof. To prepare the proof we first note that

$$u \frac{u^{\alpha-1} - 1}{1 - \alpha} \leq -u \ln u \leq u \frac{u^{\beta-1} - 1}{1 - \beta} \leq 1 - u \quad (27)$$

holds for $0 < \beta < 1 < \alpha$ and $0 \leq u \leq 1$.

The second fact required for the proof is that

$$\int_0^\infty \bar{F}(x) dx < \infty \quad \text{and} \quad \int_{-\infty}^0 F(x) dx < \infty \quad (28)$$

if $E(X) < \infty$, because

$$E(X) = \int_0^\infty \bar{F}(x) dx + \int_{-\infty}^0 F(x) dx.$$

Third, it holds that

$$P(-X \geq y) \leq P(|X| \geq y) \quad \text{for } y > 0, \quad (29)$$

because

$$\begin{aligned} P(|X| \geq y) &= 1 - P(|X| < y) = 1 - (P(X < y) - P(X \leq -y)) \\ &= 1 - P(X < y) + P(X \leq -y) \\ &= 1 - P(X < y) + P(-X \geq y). \end{aligned}$$

CPE_α consists of four indefinite integrals:

$$\begin{aligned} CPE_\alpha = & \int_0^\infty F(x) \frac{F(x)^{\alpha-1} - 1}{1-\alpha} dx + \int_{-\infty}^0 \bar{F}(x) \frac{\bar{F}(x)^{\alpha-1} - 1}{1-\alpha} dx \\ & + \int_{-\infty}^0 F(x) \frac{F(x)^{\alpha-1} - 1}{1-\alpha} dx + \int_0^\infty \bar{F}(x) \frac{\bar{F}(x)^{\alpha-1} - 1}{1-\alpha} dx. \end{aligned}$$

It must be shown separately that these integrals converge.

The convergence of the first two terms follows directly from the existence of $E(X)$. With Equations (27) and (28) we have for $\alpha > 0$

$$\int_0^\infty F(x) \frac{F(x)^{\alpha-1} - 1}{1-\alpha} dx \leq \int_0^\infty \bar{F}(x) dx < \infty$$

and

$$\int_{-\infty}^0 \bar{F}(x) \frac{\bar{F}(x)^{\alpha-1} - 1}{1-\alpha} dx \leq \int_{-\infty}^0 F(x) dx < \infty.$$

For the third term we have to demonstrate that

$$\int_{-\infty}^0 F(x) \frac{F(x)^{\alpha-1} - 1}{1-\alpha} dx < \infty$$

for $\alpha > 1/p$. If $p > 1$, there is a β with $1/p < \beta < 1$ while $\beta < \alpha$. With Equation (27) it is for $-\infty < x \leq 0$ that

$$F(x) \frac{F(x)^\alpha - 1}{1-\alpha} \leq F(x) \frac{F(x)^\beta - 1}{1-\beta} \leq \frac{1}{1-\beta} F(x)^\beta$$

because $1 - \beta > 0$.

With $F(x) = P(X \leq x) = P(-X \geq -x)$ there exists

$$\frac{1}{1-\beta} F(x)^\beta \begin{cases} \leq \frac{1}{1-\beta} & \text{for } 0 \leq -x \leq 1 \\ = \frac{1}{\beta-1} P(-X \geq -x)^\beta \leq \frac{1}{\beta-1} P(|X| \geq -x)^\beta & \text{for } 1 < -x < \infty. \end{cases}$$

For $p > 0$ the transformation $g(y) = y^p$ is monotonically increasing for $y > 1$. Using the Markov inequality we get

$$P(|X| \geq y) \geq \frac{E[|X|^p]}{y^p}.$$

Putting these results together, we attain

$$\int_{-\infty}^0 F(x) \frac{F(x)^{\alpha-1} - 1}{1-\alpha} dx \leq \frac{1}{1-\beta} + \frac{1}{1-\beta} \int_1^\infty \frac{E[|X|^p]^\beta}{y^{p\beta}} dy < \infty$$

for $\beta > 1/p$ (and thus for $p\beta > 1$) and due to $\int_1^\infty 1/y^q dy < \infty$ for $q > 1$.

It remains to show the convergence of the fourth term:

$$\int_0^\infty \bar{F}(x) \frac{\bar{F}(x)^{\alpha-1} - 1}{1-\alpha} dx < \infty$$

for $\alpha > 1/p$. For $p > 1$, there is a β with $1/p < \beta < 1$ and $\beta < \alpha$. Due to Equation (27) and $1 - \beta > 0$ for $0 \leq x < \infty$ it is true that

$$\bar{F}(x) \frac{\bar{F}(x)^\alpha - 1}{1-\alpha} \leq \bar{F}(x) \frac{\bar{F}(x)^\beta - 1}{1-\beta} \leq \frac{1}{1-\beta} \bar{F}(x)^\beta.$$

With $\bar{F}(x) = P(X > x)$ we have

$$\frac{1}{1-\beta} \bar{F}(x)^\beta \begin{cases} \leq \frac{1}{1-\beta} & \text{for } 0 \leq x \leq 1 \\ = \frac{1}{\beta-1} P(X \geq x)^\beta \leq \frac{1}{\beta-1} P(|X| \geq x)^\beta & \text{for } 1 < x < \infty. \end{cases}$$

Now, the Markov inequality gives

$$P(|X| \geq y) \geq \frac{E(|X|^p)}{y^p}.$$

In summary, we have

$$\int_{-\infty}^0 \bar{F}(x) \frac{\bar{F}(x)^{\alpha-1} - 1}{1-\alpha} dx \leq \frac{1}{1-\beta} + \frac{1}{1-\beta} \int_1^\infty \frac{E[|X|^p]^\beta}{y^{p\beta}} dy < \infty$$

for $\beta > 1/p$ and by $\int_1^\infty 1/y^q dy < \infty$ for $q > 1$. This completes the proof. \square

Following Theorem 2, depending on the number of existing moments, specific conditions for α arise in order to ensure the existence of CPE_α :

1. If the variance of X exists (i.e., $p = 2$), $CPE_\alpha(X)$ exists for $\alpha > 1/2$.
2. For $p > 1$, $E[|X|^p] < \infty$ is sufficient for the existence of CPE_S (i.e., $\alpha = 1$).
3. For $p = 1$, $E[|X|^p] < \infty$ is sufficient for the existence of CPE_G (i.e., $\alpha = 2$).

5. Maximum CPE_ϕ Distributions

5.1. Maximum CPE_ϕ Distributions for Given Mean and Variance

Equality in the Cauchy–Schwarz inequality gives a condition under which the upper bound is attained. This is the case if an affine linear relation between $F^{-1}(U)$ respectively X and $\phi'(1 - U) - \phi'(U)$ respectively $\phi'(\bar{F}(x)) - \phi'(F(X))$ exists with probability 1. Since the quantile function is monotonically increasing, such an affine linear function can only exist if $\phi'(1 - u) - \phi'(u)$ is monotonic as well (de- or increasing). This implies that ϕ needs to be a concave function on $[0, 1]$. In order to derive a maximum CPE_ϕ distribution under the restriction that mean and variance are given, one may only consider concave generating functions ϕ .

We summarize this obvious but important result in the following Theorem:

Theorem 3. Let ϕ be a non-negative and differentiable function on the domain $[0, 1]$ with derivative ϕ' and $\phi(0) = \phi(1) = 0$. Then F is the maximum CPE_ϕ distribution with prespecified mean μ and variance σ^2 of $X \sim F$ iff a constant $b \in \mathbb{R}$ exists such that

$$P\left(F^{-1}(U) - \mu = \frac{\sigma}{\sqrt{E((\phi'(1-U) - \phi'(U))^2)}} (\phi'(1-U) - \phi'(U))\right) = 1.$$

Proof. The upper bound of the Cauchy–Schwarz inequality will be attained if there are constants $a, b \in \mathbb{R}$ such that the first restriction equals

$$P\left(F^{-1}(U) = a + b(\phi'(1-U) - \phi'(U))\right) = 1.$$

The property $\phi(0) = \phi(1) = 0$ leads to $E(\phi'(1-U) - \phi'(U)) = 0$ such that

$$\mu = \int_0^1 F^{-1}(u) du = a + b \int_0^1 (\phi'(1-u) - \phi'(u)) du = a.$$

This means that there is a constant $b \in \mathbb{R}$ with

$$P\left(F^{-1}(U) - \mu = b(\phi'(1-U) - \phi'(U))\right) = 1.$$

The second restriction postulates that

$$\sigma^2 = \int_0^1 (F^{-1}(u) - \mu)^2 du = b^2 E\left((\phi'(1-U) - \phi'(U))^2\right).$$

ϕ is concave on $[0, 1]$ with

$$-\phi''(1-u) - \phi''(u) \leq 0, \quad u \in [0, 1].$$

Therefore, $\phi'(1-u) - \phi'(u)$ is monotonically increasing. The quantile function is also monotonically increasing such that b has to be positive. This gives

$$b = \frac{\sigma}{\sqrt{E((\phi'(1-U) - \phi'(U))^2)}}. \quad \square$$

The quantile function of the Tukey's λ distribution is given by

$$Q(u, \lambda) = \frac{1}{\lambda}(u^\lambda - (1-u)^\lambda), \quad u \in [0, 1], \quad \lambda \neq 0.$$

Its mean and variance are

$$\mu = 0 \quad \text{and} \quad \sigma^2 = \frac{2}{\lambda^2} \left(\frac{1}{2\lambda+1} - B(\lambda+1, \lambda+1) \right).$$

The domain is given by $[-1/\lambda, 1/\lambda]$ for $\lambda > 0$.

By discussing the paired cumulative α -entropy, one can prove the new result that the Tukey's λ distribution is the maximum CPE_α distribution for prespecified mean and variance. Tukey's λ distribution takes on the role of the Student- t distribution if one changes from the differential entropy to CPE_α (cf. [61]).

Corollary 5. *The cdf F maximizes CPE_α for $\alpha > 1/2$ under the restrictions of a given mean μ and given variance σ^2 iff F is the cdf of the Tukey λ distribution with $\lambda = \alpha - 1$.*

Proof. For $\phi(u) = u(u^{\alpha-1} - 1)/(1 - \alpha)$, $u \in [0, 1]$, we have

$$\begin{aligned} \int_0^1 (\phi'(1-u) - \phi'(u))^2 du &= \left(\frac{\alpha}{1-\alpha} \right)^2 \int_0^1 ((1-u)^{\alpha-1} - u^{\alpha-1})^2 du \\ &= 2 \left(\frac{\alpha}{1-\alpha} \right)^2 \left(\frac{1}{2\alpha-1} - B(\alpha, \alpha) \right) \end{aligned}$$

for $\alpha > 1/2$. As a consequence, the constant b is given by

$$b = \frac{1}{\sqrt{2}} \sigma \left| \frac{1-\alpha}{\alpha} \right| \left(\frac{1}{2\alpha-1} - B(\alpha, \alpha) \right)^{-1/2},$$

and the maximum CPE_α distribution results in

$$\begin{aligned} F^{-1}(u) - \mu &= \frac{\sigma}{\sqrt{2}} \left| \frac{1-\alpha}{\alpha} \right| \left(\frac{1}{2\alpha-1} - B(\alpha, \alpha) \right)^{-1/2} \frac{\alpha}{1-\alpha} \left((1-u)^{\alpha-1} - u^{\alpha-1} \right) \\ &= \sigma \frac{|\alpha-1|}{\sqrt{2}} \left(\frac{1}{2\alpha-1} - B(\alpha, \alpha) \right)^{-1/2} \frac{(u^{\alpha-1} - (1-u)^{\alpha-1})}{\alpha-1}. \end{aligned}$$

F^{-1} can easily be identified as the quantile function of a Tukey's λ distribution with $\lambda = \alpha - 1$ and $\alpha > 1/2$. \square

For the Gini case ($\alpha = 2$), one obtains the quantile function of a uniform distribution

$$F^{-1}(u) = \mu + \sigma \sqrt{\frac{1}{2}} \sqrt{6} (2u - 1) = \mu + \sigma \sqrt{3} (2u - 1), \quad u \in [0, 1],$$

with domain $[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]$. This maximum CPE_G distribution corresponds essentially to the distribution derived by Dai et al. [40].

The fact that the logistic distribution is the maximum CPE_S distribution, provided mean and variance are given, was derived by Chen et al. [51] in the framework of uncertainty theory and by ([50], p. 4) in the framework of reliability theory. Both proved this result using Euler–Lagrange equations. In the interest of completeness, we provide an alternative proof via the upper bound of the Cauchy–Schwarz inequality.

Corollary 6. *The cdf F maximizes CPE_S under the restrictions of a known mean μ and a known variance σ^2 iff F is the cdf of a logistic distribution.*

Proof. Since

$$\int_0^1 \left(\ln \left(\frac{1-u}{u} \right) \right)^2 du = \frac{\pi^2}{3}, \quad (30)$$

one receives

$$F^{-1}(u) - \mu = \frac{\sigma}{\pi/\sqrt{3}} \ln \left(\frac{1-u}{u} \right), \quad u \in [0, 1].$$

Inverting gives the distribution function of the logistic distribution with mean μ and variance 1:

$$F(x) = \frac{1}{1 + \exp \left(-\frac{\pi}{\sqrt{3}} \frac{x-\mu}{\sigma} \right)}, \quad x \in \mathbb{R}. \quad \square$$

As a last example we consider the cumulative paired Leik entropy CPE_L .

Corollary 7. *The cdf F maximizes CPE_L under restrictions of a known mean μ and a known variance σ^2 iff for F holds*

$$F(x) = \begin{cases} 0 & \text{for } x < \mu - \sigma \\ 1/2 & \text{for } \mu - \sigma \leq x < \mu + \sigma \\ 1 & \text{for } x \geq \mu + \sigma. \end{cases}$$

Proof. From $\phi(u) = \min\{u, 1-u\}$ and $\phi'(u) = \text{sign}(1/2 - u)$, $u \in [0, 1]$, follows that

$$F^{-1}(u) - \mu = \sigma \text{sign}(u - 1/2), \quad u \in [0, 1]. \quad \square$$

Therefore, the maximization of CPE_L with given mean and variance leads to a distribution whose variance is maximal on the interval $[\mu - \sigma, \mu + \sigma]$.

5.2. Maximum CPE_ϕ Distributions for General Moment Restrictions

Drissi et al. [50] discuss general moment restrictions of the form

$$\int_{-\infty}^{\infty} c_i(x) f(x) dx = \int_0^1 c_i(F^{-1}(U)) du = k_i, \quad i = 1, 2, \dots, k, \quad (31)$$

for which the existence of the moments is assumed. By using Euler–Lagrange equations they show that

$$\bar{F}(x) = \frac{1}{1 + \exp(\sum_{i=1}^r \lambda_i c'_i(x))}, \quad x \in \mathbb{R},$$

maximizes the residual cumulative entropy $-\int_{\mathbb{R}} \bar{F}(x) \ln \bar{F}(x) dx$ under constraints Equation (31). Moreover, they demonstrated that the solution needs to be symmetric with respect to μ . Here, λ_i , $i = 1, 2, \dots, k$, are the Lagrange parameters which are determined by the moment restrictions, provided a solution exists. Rao et al. [47] shows that for distributions with support \mathbb{R}^+ the ME distribution is given by

$$\bar{F}(x) = \exp \left(- \sum_{i=1}^r \lambda_i c'_i(x) \right), \quad x > 0$$

if the restrictions Equation (31) are again required.

One can easily examine the shape of a distribution which maximizes the cumulative paired ϕ -entropy under the constraints Equation (31). This maximum CPE_{ϕ} distribution can no longer be derived by the upper bound of the Cauchy–Schwarz inequality if $i > 2$. One has to solve the Euler–Lagrange equations for the objective function

$$\int_0^1 (\phi'(u) - \phi'(1-u)) F^{-1}(u) du - \sum_{i=1}^k \lambda_i (c_i(F^{-1}(u)) - k_i) \quad (32)$$

with Lagrange parameters λ_i , $i = 1, 2, \dots, k$. The Euler–Lagrange equations lead to the optimization problem

$$\sum_{i=1}^k \lambda_i c'_i(F^{-1}(u)) = \phi'(1-u) - \phi'(u), \quad u \in [0, 1], \quad (33)$$

for $i = 1, 2, \dots, k$. Once again there is a close relation between the derivative of the generating function and the quantile function, provided a solution of the optimization problem Equation (32) exists.

The following example shows that the optimization problem Equation (32) leads to a well-known distribution if constraints are chosen carefully in case of a Shannon-type entropy.

Example 1. The power logistic distribution is defined by the distribution function

$$F(x) = \frac{1}{1 + \exp(-\lambda \operatorname{sign}(x)x^{\gamma})}, \quad x \in \mathbb{R},$$

for $\gamma > 0$. The corresponding quantile function is

$$F^{-1}(u) = \left(\frac{1}{\lambda} \right)^{1/\gamma} \operatorname{sign}(u - 1/2) \left| \ln \left(\frac{1-u}{u} \right) \right|^{1/\gamma}, \quad u \in [0, 1].$$

This quantile function is also solution of Equation (33) given $\phi(u) = -u \ln u$, $u \in [0, 1]$, under the constraint $E[|X|^{\gamma+1}] = c$. The maximum of the cumulative paired Shannon entropy under the constraint $E[|X|^{\gamma+1}] = c$ is given by

$$\begin{aligned} CPE_S(X) &= \int_0^1 \ln \left(\frac{1-u}{u} \right) \left(\frac{1}{\lambda} \right)^{1/\gamma} \operatorname{sign}(u - 1/2) \cdot \left| \ln \left(\frac{1-u}{u} \right) \right|^{1/\gamma} du \\ &= \left(\frac{1}{\lambda} \right)^{1/\gamma} \int_0^1 \left| \ln \left(\frac{1-u}{u} \right) \right|^{(\gamma+1)/\gamma} du = \lambda E(|X|^{\gamma+1}). \end{aligned}$$

Setting $\gamma = 1$ leads to the familiar result for the upper bound of CPE_S given the variance.

5.3. Generalized Principle of Maximum Entropy

Kesavan et al. [19] introduced the generalized principle of an ME problem which describes the interplay of entropy, constraints, and distributions. A variation of this principle is the aim of finding an entropy that is maximized by a given distribution and some moment restrictions.

This problem can easily be solved for CPE_ϕ if mean and variance are given, due to the linear relationship between $\phi'(1-u) - \phi'(u)$ and the quantile function $F^{-1}(u)$ of the maximum CPE_ϕ distribution provided by the Cauchy-Schwarz inequality. However, it is a precondition for $F^{-1}(u)$ that $\phi'(1-u) - \phi'(u)$ is strictly monotonic on $[0, 1]$ in order to be a quantile function. Therefore, the concavity of $\phi(u)$ and the condition $\phi(0) = \phi(1) = 0$ are of key importance.

We demonstrate the solution to the generalized principle of the maximum entropy problem for the Gaussian and the Student- t distribution.

Proposition 3. Let ϕ , Φ and Φ^{-1} be the density, the cdf and the quantile function of a standard Gaussian random variable. The Gaussian distribution is the maximum CPE_ϕ distribution for a given mean μ and variance σ^2 for CPE_ϕ with entropy generating function

$$\phi(u) = \phi(\Phi^{-1}(u)), \quad u \in [0, 1].$$

Proof. With

$$\phi'(u) = \frac{\phi'(\Phi^{-1}(u))}{\phi(\Phi^{-1}(u))} = -\Phi^{-1}(u), \quad u \in [0, 1],$$

the condition for the maximum CPE_ϕ distribution with mean μ and variance σ^2 becomes

$$F^{-1}(u) - \mu = \frac{\sigma}{\sqrt{\int_0^1 (2\Phi^{-1}(u))^2 du}} 2\Phi^{-1}(u), \quad u \in [0, 1].$$

By substituting $\int_0^1 (2\Phi^{-1}(u))^2 du = 4$, it follows that

$$F^{-1}(u) - \mu = \sigma\Phi^{-1}(u), \quad u \in [0, 1],$$

such that F^{-1} is the quantile function of a Gaussian distribution with mean μ and variance σ^2 . \square

An analogue result holds for the Student- t distribution with k degrees of freedom. In this case, the main difference to the Gaussian distribution is the fact that the entropy generating function possesses no closed form but is obtained by numerical integration of the quantile function.

Corollary 8. Let t_k respectively t_k^{-1} be the cdf respectively the quantile function of a Student- t distribution with k degrees of freedom for $k > 2$. $\mu + \frac{k}{k-2}t_k^{-1}$ is the maximum CPE_ϕ quantile function for a given mean μ and variance σ^2 iff

$$\phi(u) = \sqrt{\frac{k-2}{k}} \int_0^u t_k^{-1}(p) dp, \quad u \in [0, 1].$$

Proof. Starting with

$$\phi'(u) = -\sqrt{\frac{k-2}{k}} t_k^{-1}(u), \quad u \in [0, 1],$$

and the symmetry of the t_k distribution around μ , we get the condition

$$F^{-1}(u) - \mu = \frac{\sigma}{\sqrt{\int_0^1 (2t_k^{-1}(u))^2 du}} 2\sqrt{k-2}t_k^{-1}(u), \quad u \in [0, 1].$$

With $\int_0^1 (t_k^{-1}(u))^2 du = k/(k-2)$ we get the quantile function of the t distribution with k degrees of freedom and mean μ :

$$F^{-1}(u) - \mu = \sigma \frac{k-2}{k} t_k^{-1} = t_k^{-1}(u), \quad u \in [0,1]. \quad \square$$

Figure 2 shows the shape of the entropy generating function ϕ for several distributions generated by the generalized ME principle.

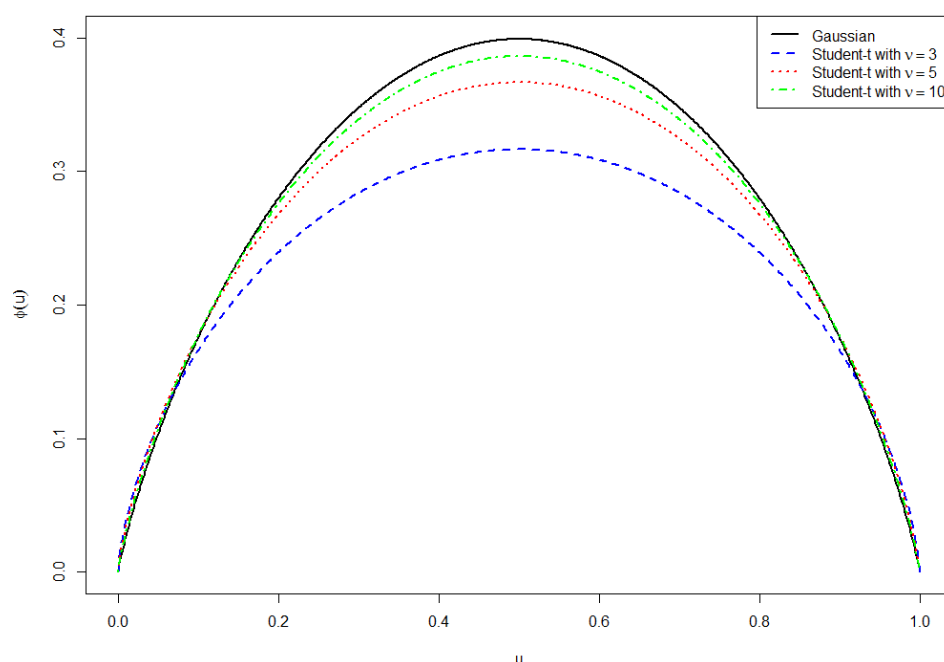


Figure 2. Several entropy generating functions ϕ derived from the generalized maximum entropy (ME) principle.

6. CPE_ϕ as a Measure of Scale

6.1. Basic Properties of CPE_ϕ

The cumulative residual entropy (CRE) introduced by [46], the generalized cumulative residual entropy (GCRE) of [50], and the cumulative entropy (CE) discussed by [8,9], have always been interpreted as measures of information. However, all these approaches do not explain which kind of information was considered. In contrast to this interpretation as measures of information, Oja [3] proved that the differential entropy satisfies a special ordering of scale and has certain meaningful properties of measures of scale. In [4], the authors discussed the close relationship between differential entropy and variance. In the discrete case the Shannon entropy can be interpreted as a measure of diversity, which is a concept of dispersion if there is no ordering and no distance between the realizations of a random variable. In this section, we will clarify the important role which the variance plays for the existence of CPE_ϕ .

Therefore, we intend to provide a deeper insight in CPE_ϕ as a proper MOS. We start by showing that CPE_ϕ has typical properties of an MOS. In detail, a proper MOS should always be non-negative and attain its minimal value 0 for a degenerated distribution. If a finite interval $[a, b]$ is considered as support, an MOS should attain its maximum if a and b occur with probability $1/2$. CPE_ϕ possesses all these properties as shown in the next proposition.

Proposition 4. Let $\phi : [0,1] \rightarrow \mathbb{R}$ with $\phi(u) > 0$ for $u \in (0,1)$ and $\phi(0) = \phi(1) = 0$. Let X be a random variable with support D and CPE_ϕ is assumed to exist. Then the following properties hold:

1. $CPE_\phi(X) \geq 0$.
2. $CPE_\phi(X) = 0$ iff there exists an x^* with $\mathbb{P}(X = x^*) = 1$.
3. $CPE_\phi(X)$ attains its maximum iff there exist a, b with $-\infty < a < b < \infty$ such that $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 1/2$.

Proof.

1. Follows from the non-negativity of ϕ .
2. If there is an $x^* \in \mathbb{R}$ with $\mathbb{P}(X = x^*) = 1$, then $F_X(x) = 0$ and $\bar{F}_X(x) \in \{0, 1\}$ for all $x \in \mathbb{R}$. Due to $\phi(0) = \phi(1) = 0$ follows $\phi(F_X(x)) = \phi(\bar{F}_X(x)) = 0$ for all $x \in \mathbb{R}$.
Set $CPE_\phi(X) = 0$. Due to the non-negativity of the integrand of CPE_ϕ , $\phi(F_X(x)) + \phi(\bar{F}_X(x)) = 0$ must hold for $x \in \mathbb{R}$. Since $\phi(u) > 0, 0 < u < 1$, it follows $F_X(x), \bar{F}_X(x) \in \{0, 1\}$ for $x \in [0, 1]$.
3. Let $CPE_\phi(X)$ have a finite maximum. Since $\phi(u) + \phi(1 - u)$ has a unique maximum at $u = 1/2$, the maximum of $CPE_\phi(X)$ is

$$2 \int_D \phi(1/2) du = 2\phi(1/2) \int_D du.$$

In order to attain the assumed finite maximum, the support D has to be a finite interval $[a, b]$. Here, $2\phi(1/2)(b - a)$ is the maximum. Now, it is sufficient to construct a distribution with support $[a, b]$ that attains this maximum. Set

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ 1/2 & \text{for } a \leq x \leq b \\ 1 & \text{for } x \geq b, \end{cases}$$

then $CPE_\phi(F) = \int_a^b \phi(F(x)) + \phi(\bar{F}(x)) dx = 2\phi(1/2)(b - a)$. Therefore, F is CPE_ϕ -maximal.

To prove the other direction of statement 3 we consider an arbitrary distribution G with survival function \bar{G} and support $[a, b]$. Due to $\phi(0) = \phi(1) = 0$ and $\phi(u) + \phi(1 - u) \leq 2\phi(1/2)$, it holds that

$$CPE_\phi(G) = \int_a^b \phi(G(x)) + \phi(\bar{G}(x)) dx \leq 2\phi(1/2)(b - a) = CPE_\phi(F). \quad \square$$

6.2. CPE_ϕ and Oja's Axioms for Measures of Scale

Oja ([3] p. 159) defined a MOS as follows:

Definition 2. Let \mathcal{F} be a set of continuous distribution functions and \preceq an appropriate ordering of scale on \mathcal{F} . $T : \mathcal{F} \rightarrow \mathbb{R}$ is called MOS, if

1. $T(aX + b) = |a|T(X)$ for all $a, b \in \mathbb{R}, F \in \mathcal{F}$.
2. $T(X_1) \leq T(X_2)$, for $X_1 \sim F_1, X_2 \sim F_2, F_1, F_2 \in \mathcal{F}$ with $F_1 \preceq F_2$.

Oja [3] discussed several orderings of scale. He showed in particular that Shannon entropy and variance satisfy a partial quantile based ordering of scale, which has been discussed by [62]. Referring to [63,64] criticized that this ordering and the location-scale family of distributions focused by Oja [3] were too restrictive. He discussed a more general nonparametric model of dispersion based on a more general ordering of scale (cf. [65,66]). In line with [4], we focus on the scale ordering proposed by [62].

Definition 3. Let F_1, F_2 be continuous cdfs with respective quantile functions F_1^{-1} and F_2^{-1} . F_2 is said to be more spread out than F_1 ($F_1 \preceq_1 F_2$) if

$$F_2^{-1}(u) - F_2^{-1}(v) \geq F_1^{-1}(u) - F_1^{-1}(v) \text{ for all } 0 < u < v < 1. \quad (34)$$

If F_1, F_2 are absolutely continuous with density functions f_1, f_2 , \preceq_1 can be characterized equivalently by the property that $F_2^{-1}(F_1^{-1}(x)) - x$ is monotonically non-decreasing or

$$f_1(F_1^{-1}(u)) \geq f_2(F_2^{-1}(u)), \quad u \in [0, 1] \quad (35)$$

(cf. [3], p. 160).

Next, we show that CPE_ϕ is an MOS in the sense of [3]. This following lemma examines the behavior of CPE_ϕ with respect to affine linear transformations, referring to the first axiom of Definition 2:

Lemma 2. *Let F be the cdf of the random variable X . Then*

$$CPE_\phi(aX + b) = |a|CPE_\phi(X).$$

Proof. For $Y = aX + b$, it follows that

$$\int_{-\infty}^{\infty} \phi(P(Y \leq y)) dy = \begin{cases} \int_{-\infty}^{\infty} P\left(X \leq \frac{y-b}{a}\right) dy & \text{for } a > 0 \\ \int_{-\infty}^{\infty} P\left(X > \frac{y-b}{a}\right) dy & \text{for } a < 0 \end{cases}.$$

Substitution of $x = (y - b)/a$ with $dy = adx$ gives

$$\int_{-\infty}^{\infty} \phi(P(Y \leq y)) dy = \begin{cases} a \int_{-\infty}^{\infty} P(X \leq x) dx & \text{for } a > 0 \\ -a \int_{-\infty}^{\infty} P(X > x) dx & \text{for } a < 0 \end{cases}.$$

Likewise, we have that

$$\int_{-\infty}^{\infty} \phi(P(Y > y)) dy = \begin{cases} a \int_{-\infty}^{\infty} P(X > x) dx & \text{for } a > 0 \\ -a \int_{-\infty}^{\infty} P(X \leq x) dx & \text{for } a < 0 \end{cases},$$

such that

$$CPE_\phi(aX + b) = |a|CPE_\phi(X). \quad \square$$

In order to satisfy the second axiom of Oja's definition of a measure of scale, CPE_ϕ has to satisfy the ordering of scale \preceq . This is shown by the following lemma:

Lemma 3. *Let F_1 and F_2 be continuous cdfs of the random variables X_1 and X_2 with $F_1 \preceq_1 F_2$. Then the following holds:*

$$CPE_\phi(X_1) \leq CPE_\phi(X_2).$$

Proof. One can show with $u = F_i(x)$ that

$$CPE_\phi(F_i) = \int_0^1 \phi(u) \frac{1}{f_i(F_i^{-1}(u))} du + \int_0^1 \phi(1-u) \frac{1}{f_i(F_i^{-1}(u))} du$$

for $i = 1, 2$. Therefore,

$$\begin{aligned} CPE_\phi(F_1) - CPE_\phi(F_2) &= \int_0^1 \phi(u) \left(\frac{1}{f_1(F_1^{-1}(u))} - \frac{1}{f_2(F_2^{-1}(u))} \right) du \\ &\quad + \int_0^1 \phi(1-u) \left(\frac{1}{f_1(F_1^{-1}(u))} - \frac{1}{f_2(F_2^{-1}(u))} \right) du. \end{aligned}$$

If $F_1 \preceq_1 F_2$ and hence $f_1(F_1^{-1}(u)) \geq f_2(F_2^{-1}(u))$ for $u \in [0, 1]$, it follows that $CPE_\phi(F_1) - CPE_\phi(F_2) \leq 0$. \square

As a consequence of Lemma 2 and Lemma 3, CPE_ϕ is an MOS in the sense of [3]. Thus, not only variance, differential entropy, and other statistical measures have the properties of measures of scale, but also CPE_ϕ .

6.3. CPE_ϕ and Transformations

Ebrahimi et al. ([4] p. 323), the authors considered cdf F_1, F_2 on domain D_1, D_2 and density functions f_1, f_2 , which are connected via $F_2(x) = F_1(g^{-1}(x))$, $x \in D_1$, via a differentiable transformation $g : D_1 \rightarrow D_2$, that is $F_2(y) = F_1(g(y))$ respectively $f_2(y) = f_1(g^{-1}(y)) |dg^{-1}(y)/dy|$ for $y \in D_1$. Thus, they demonstrated for Shannon's differential entropy H that the transformation only affects the difference:

$$H(f_2) = H(f_1) - \int_{D_2} \ln \left| \frac{dg^{-1}(y)}{dy} \right| f_2(y) dy.$$

For CPE_ϕ , one gets a less explicit relationship between $CPE_\phi(F_2)$ and $CPE_\phi(F_1)$:

$$CPE_\phi(F_2) = \int_{D_1} (\phi(F_1(y)) + \phi(\bar{F}_1(y))) \frac{dg^{-1}(y)}{dy}.$$

Transformations with $|g'(y)| \geq 1$, $y \in D_2$, are of special interest since these transformations do not diminish measures of scale. In Theorem 1, Ebrahimi et al. [4] showed that $F_1 \preceq_1 F_2$ holds if $|g'(y)| \geq 1$ for $y \in D_2$. Hence, no MOS can be diminished by this specific transformation, especially neither Shannon entropy nor CPE_ϕ .

Ebrahimi et al. [4] considered the special transformation $g(x) = ax + b$, $x \in D_1$. They showed that Shannon's differential entropy is moved additively by this transformation, which is not expected for an MOS. Furthermore, the standard deviation is changed by the factor $|a|$, which is also true for CPE_ϕ as shown in Lemma 2.

6.4. CPE_ϕ for Sums of Independent Random Variables

As is generally known, variance and differential entropy behave additively for the sum of independent random variables X and Y . More general entropies such as the Rényi or the Havrda & Charvát entropy are only subadditive (cf. [18], p. 194).

Neither the property of additivity nor the property of subadditivity could be shown for cumulative paired ϕ -entropies. Instead, they possess the maximum property if ϕ is a concave function on $[0, 1]$. This means that, for two independent variables X and Y , $CPE_\phi(X + Y)$ is lower-bounded by the maximum of the two individual entropies $CPE_\phi(X)$ and $CPE_\phi(Y)$. This result was shown by [46] for the cumulative residual Shannon entropy. The following Theorem generalizes this result, while the proof partially follows Theorem 2 of [46].

Theorem 4. Let X and Y be independent random variables and ϕ a concave function on the interval $[0, 1]$ with $\phi(0) = \phi(1) = 0$. Then we have

$$CPE_\phi(X + Y) \geq \max \{CPE_\phi(X), CPE_\phi(Y)\}. \quad (36)$$

Proof. Let X and Y be independent random variables with distribution functions F_X, F_Y and densities f_X, f_Y . Using the convolution formula, we immediately get

$$P(X + Y \leq t) = \int_{-\infty}^{\infty} F_X(t - y) f_Y(y) dy = E_Y[F_X(t - Y)], \quad t \in \mathbb{R}. \quad (37)$$

Applying Jensen's inequality for a concave function ϕ to Equation (37) results in

$$E_Y[\phi(F_X(t - Y))] \geq \phi(E_Y[F_X(t - Y)]) \quad (38)$$

and

$$E_Y [\phi(\bar{F}_X(t - Y))] \geq \phi(E_Y [\bar{F}_X(t - Y)]). \quad (39)$$

The existence of the expectation is assumed. To prove the Theorem, we begin with

$$CPE_\phi[X + Y] = \int_{-\infty}^{\infty} \phi(E_Y[F_X(t - Y)]) + \phi(E_Y[\bar{F}_X(t - Y)]) dt.$$

By using Equations (38) and (39), setting $z = t - y$, and exchanging the order of integration, one yields

$$\begin{aligned} CPE_\phi[X + Y] &\geq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(F_X(t - y)) + \phi(\bar{F}_X(t - y)) dt f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(F_X(z)) + \phi(\bar{F}_X(z)) dz f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \phi(F_X(z)) + \phi(\bar{F}_X(z)) dz = CPE_\phi[X]. \quad \square \end{aligned}$$

In the context of uncertainty theory, Liu [6] considered a different definition of independence for uncertain variables leading to the simpler additivity property

$$CPE_\phi(X + Y) = CPE_\phi(X) + CPE_\phi(Y) \quad (40)$$

for independent uncertain variables X and Y .

7. Estimation of CPE_ϕ

Beirlant et al. [67] presented an overview of differential entropy estimators. Essentially, all proposals are based on the estimation of a density function f inheriting all typical problems of nonparametric estimation of a density function. Among others, the problems are biasedness, choice of a kernel, and optimal choice of the smoothing parameter (cf. [68], p. 215ff.). However, CPE_ϕ is based on cdf F for which several natural estimators with desirable stochastic properties, derived from the Theorem of Glivenko and Cantelli (cf. [69], p. 61), exist. For a simple random sample (X_1, \dots, X_n) , independently distributed random variables with identical distribution function F , the authors of [8,9] estimated F using the empirical distribution function $F_n(x) = \frac{1}{n}I(X_i \leq x)$ for $x \in \mathbb{R}$. Moreover, they showed for the cumulative entropy $CE(F) = -\int_{\mathbb{R}} F(x) \ln F(x) dx$ that the estimator $CE(F_n)$ is consistent for $CE(F)$ (cf. [8]). In particular, for F being the distribution function of a uniform distribution, they provided the expected value of the estimator and demonstrated that the estimator is asymptotically normal. For F being the cdf of an exponential distribution, they additionally derived the variance of the estimator.

In the following, we generalize the estimation approach of [8] by embedding it into the well-established theory of L -estimators (cf. [70], p. 55ff.). If ϕ is differentiable, then CPE_ϕ can be represented as the covariance between the random variable X and $\phi'(\bar{F}(X)) - \phi'(F(X))$:

$$CPE_\phi(F) = E(X(\phi'(\bar{F}(X)) - \phi'(F(X)))) \quad (41)$$

An unbiased estimator for this covariance is

$$\begin{aligned}
 CPE_{\phi}(F_n) &= \frac{1}{n} \sum_{i=1}^n X_i (\phi'(1 - F_n(X_i)) - \phi'(F_n(X_i))) \\
 &= \frac{1}{n} \sum_{i=1}^n X_{n:i} (\phi'(1 - F_n(X_{n:i})) - \phi'(F_n(X_{n:i}))) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\phi' \left(1 - \frac{i}{n+1} \right) - \phi' \left(\frac{i}{n+1} \right) \right) X_{n:i} \\
 &= \sum_{i=1}^n c_{ni} X_{n:i}
 \end{aligned} \tag{42}$$

where

$$c_{ni} = \frac{1}{n} \left(\phi' \left(1 - \frac{i}{n+1} \right) - \phi' \left(\frac{i}{n+1} \right) \right), \quad i = 1, 2, \dots, n.$$

This results in an L -estimator $\sum_{i=1}^n J(i/(n+1))X_{n:i}$ with $J(u) = \phi'(1-u) - \phi'(u)$, $u \in (0, 1)$. By applying known results for the influence functions of L -estimators (cf. [70]), we get for the influence function of CPE_{ϕ} :

$$\begin{aligned}
 IF(x; CPE_{\phi}, F) &= \int_0^1 \frac{u}{f(F^{-1}(u))} (\phi'(1-u) - \phi'(u)) du \\
 &\quad - \int_{F(x)}^1 \frac{1}{f(F^{-1}(u))} (\phi'(1-u) - \phi'(u)) du.
 \end{aligned} \tag{43}$$

In particular, the derivative is

$$\frac{dIF(x; CPE_{\phi}, F)}{dx} = \phi'(\bar{F}(x)) - \phi'(F(x)), \quad x \in \mathbb{R}. \tag{44}$$

This means that the influence function will be completely determined by the antiderivative of $\phi'(F(x))$. The following examples demonstrate that the influence function of CPE_{ϕ} can easily be calculated if the underlying distribution F is logistic. We consider the Shannon, the Gini, and the α -entropy cases.

Example 2. *Beginning with the derivative*

$$\frac{dIF(x; CPE_S, F)}{dx} = \phi'(\bar{F}(x)) - \phi'(F(x)) = \ln \left(\frac{F(x)}{\bar{F}(x)} \right) = x, \quad x \in \mathbb{R},$$

we arrive at

$$IF(x, CPE_S, F) = \frac{1}{2}x^2 + C, \quad x \in \mathbb{R}.$$

The influence function is not bounded and proportional to the influence function of the variance, which implies that variance and CPE_S have a similar asymptotic and robustness behavior. The integration constant C has to be determined such that $E[IF(x; CPE_S, F)] = 0$:

$$C = -\frac{1}{2}E(X^2) = -\frac{1}{2} \frac{\pi^2}{3} = -\frac{\pi^2}{6}.$$

Example 3. *Using the Gini entropy CPE_G and the logistic distribution function F we have*

$$\begin{aligned}
 \frac{dIF(x; CPE_G, F)}{dx} &= \phi'(\bar{F}(x)) - \phi'(F(x)) = 2(2F(x) - 1) \\
 &= 2 \frac{e^x - 1}{e^x + 1} = 2 \tanh \left(\frac{x}{2} \right), \quad x \in \mathbb{R}.
 \end{aligned}$$

Integration gives the influence function

$$IF(x, CPE_G, F) = 4 \ln \left(\cosh \left(\frac{x}{2} \right) \right) + C, \quad x \in \mathbb{R}.$$

By applying numerical integration we get $C = -1.2741$.

Example 4. For $\phi(u) = u(u^{\alpha-1} - 1)/(1 - \alpha)$ the derivative of the influence function is given by

$$\begin{aligned} \frac{dIF(x; CPE_\alpha, F)}{dx} &= \phi'(\bar{F}(x)) - \phi'(F(x)) = \frac{\alpha}{1 - \alpha} \frac{1 - e^{(\alpha-1)x}}{(1 + e^x)^{\alpha-1}} \\ &= \frac{\alpha}{1 - \alpha} \left(\frac{1}{(1 + e^x)^{\alpha-1}} - \frac{1}{(1 + e^{-x})^{\alpha-1}} \right), \quad x \in \mathbb{R}. \end{aligned}$$

Integration leads to the influence function

$$\begin{aligned} IF(x; CPE_\alpha, F) &= {}_2F_1(\alpha, \alpha; \alpha + 1; -e^{-x}) \frac{e^{\alpha x}}{\alpha} \left(1 + \left(\frac{e^{-x} + 1}{e^x + 1} \right)^\alpha \right) \\ &\quad + \frac{1}{\alpha - 1} \left(\frac{1 + e^x + e^{(\alpha-1)x}}{(e^x + 1)^\alpha} - 1 \right), \end{aligned}$$

where

$${}_2F_1(\alpha, \alpha; \alpha + 1; -e^{-x}) = \alpha \int_0^1 t^{\alpha-1} (1 + te^{-x})^{-\alpha} dt + C, \quad x \in \mathbb{R}.$$

Under certain conditions (cf. [71], p. 143) concerning J , or ϕ and F , L -estimators are consistent and asymptotically normal. So, the cumulative paired ϕ -entropy is

$$CPE_\phi(F_n) \sim_{asy} N \left(CPE_\phi(F), \frac{1}{n} A(F, CPE_\phi) \right)$$

with asymptotic variance

$$A(F, CPE_\phi) = \text{Var}(IF(X; CPE_\phi(F), F)) = \int_{-\infty}^{\infty} \left(\int_{F(x)}^1 \frac{\phi'(1-u) - \phi'(u)}{f(F^{-1}(u))} du \right)^2 f(x) dx.$$

The following examples consider the Shannon and the Gini case for which the condition that is sufficient to guarantee asymptotic normality can easily be checked. We consider again the cdf F of the logistic distribution.

Example 5. For the cumulative paired Shannon entropy it holds that

$$CPE_S(F_n) \sim_{asy} N \left(CPE_S(F), \frac{4}{45} \pi^4 \right)$$

since

$$A(F, L) = \text{Var}(IF(X; CPE_\phi(F), F)) = \frac{1}{4} \text{Var}(X^2) = \frac{1}{4} (E(X^4) - E(X^2)^2) = \frac{4}{45} \pi^4.$$

Example 6. In the Gini case we get

$$CPE_G(F_n) \sim_{asy} N(CPE_G(F), 2.8405)$$

since by numerical integration

$$A(F, L) = \int_{-\infty}^{\infty} \left(4 \ln \left(\cosh \left(\frac{x}{2} \right) \right) - 1.2274 \right)^2 \frac{e^{-x}}{(1 + e^{-x})^2} dx = 2.8405.$$

It is known that L -estimators have a remarkable small-sample bias. Following [72], the bias can be reduced by applying the Jackknife method. It is well-known that asymptotical distributions can be used to construct approximate confidence intervals as well as that they can be applied for hypothesis tests in the one- or two-sample case. ([70], p. 116ff.) discussed asymptotic efficient L -estimators for a parameter of scale θ . Klein et al. [73] examine how the entropy generating function ϕ will be determined by the requirement that $CPE_\phi(F_n)$ has to be asymptotically efficient.

8. Related Concepts

Several statistical concepts are closely related to cumulative paired ϕ -entropies. These concepts generalize some results which are known from literature. We begin with the cumulative paired ϕ -divergence that was discussed for the first time by [41], who called it “generalized cross entropy”. Their focus was on uncertain variables, whereas ours is on random variables. The second concept generalizes mutual information, which is defined for Shannon’s differential entropy, to mutual ϕ -information. We consider two random variables X and Y . The task is to decompose $CPE_\phi(Y)$ into two kinds of variation such that the so-called external variation measures how much of $CPE_\phi(Y)$ can be explained by X . This procedure mimics the well-known decomposition of variance and allows to define directed measures of dependence for X and Y . The third concept deals with dependence. More precisely, we introduce a new family of correlation coefficients that measure the strength of a monotonic relationship between X and Y . Well-known coefficients like the Gini correlation can be embedded in this approach. The fourth concept treats the problem of linear regression. CPE_ϕ can serve as general measure of dispersion that has to be minimized to estimate the regression coefficients. This approach will be identified as a special case of rank-based regression or R regression. Here, the robustness properties of the rank-based estimator can directly be derived from the entropy generating function ϕ . Moreover, asymptotics can be derived from theory of rank-based regression. The last concept we discuss applies CPE_ϕ to linear rank tests for the difference of scale. Known results, especially concerning the asymptotics, can be transferred from the theory of linear rank tests to this new class of tests. In this paper, we only sketch the main results and focus on examples. For a detailed discussion including proofs we refer to a series of papers by Klein and Mangold ([73–75]), which are currently work in progress.

8.1. Cumulative Paired ϕ -Divergence

Let ϕ be a concave function defined on $[0, \infty]$ with $\phi(0) = \phi(1) = 0$. Additionally, we need $0\phi(0/0) = 0$. In the literature, ϕ -divergences are defined for convex functions ϕ (cf., e.g., [76], p. 5). Consequently, we consider $-\phi$ with ϕ concave.

The cumulative paired ϕ -divergence for two random variables is defined as follows.

Definition 4. Let X and Y be two random variables with cdfs F_X and F_Y . Then the cumulative paired ϕ -divergence of X and Y is given by

$$CPD_\phi(X, Y) = - \int_{-\infty}^{\infty} F_Y(x) \phi \left(\frac{F_X(x)}{F_Y(x)} \right) + \bar{F}_Y(x) \phi \left(\frac{\bar{F}_X(x)}{\bar{F}_Y(x)} \right) dx. \quad (45)$$

The following examples introduce cumulative paired ϕ -divergences for the Shannon, the α -entropy, the Gini, and the Leik cases:

Example 7.

1. Considering $\phi(u) = -u \ln u$, $u \in [0, \infty)$, we obtain the cumulative paired Shannon divergence

$$CPD_S(X, Y) = \int_{-\infty}^{\infty} F_X(x) \ln \left(\frac{F_X(x)}{F_Y(x)} \right) + \bar{F}_X(x) \ln \left(\frac{\bar{F}_X(x)}{\bar{F}_Y(x)} \right) dx.$$

2. Setting $\phi(u) = u(u^{\alpha-1} - 1)/(1 - \alpha)$, $u \in [0, \infty)$, leads to the cumulative paired α -divergence

$$CPD_{\alpha}(X, Y) = \frac{1}{\alpha - 1} \left(\int_{-\infty}^{\infty} F_X(x)^{\alpha} F_Y(x)^{1-\alpha} + \bar{F}_X(x)^{\alpha} \bar{F}_Y(x)^{1-\alpha} - 1 \right) dx.$$

3. For $\alpha = 2$ we receive as a special case the cumulative paired Gini divergence

$$\begin{aligned} CPD_G(X, Y) &= \int_{-\infty}^{\infty} \left(\frac{F_X(x)^2}{F_Y(x)} + \frac{\bar{F}_X(x)^2}{\bar{F}_Y(x)} - 1 \right) dx \\ &= \int_{-\infty}^{\infty} \frac{(F_X(x) - F_Y(x))^2}{F_Y(x) \bar{F}_Y(x)} dx. \end{aligned}$$

4. The choice $\phi(u) = 1/2 - |u - 1/2|$, $u \in [0, 1]$, leads to the cumulative paired Leik divergence

$$\begin{aligned} CPD_L(X, Y) &= \int_{-\infty}^{\infty} -F_Y(x) \left(\frac{1}{2} - \left| \frac{F_X(x)}{F_Y(x)} - \frac{1}{2} \right| \right) - \bar{F}_Y(x) \left(\frac{1}{2} - \left| \frac{\bar{F}_X(x)}{\bar{F}_Y(x)} - \frac{1}{2} \right| \right) dx \\ &= \int_{-\infty}^{\infty} -\frac{1}{2} + \left| F_X(x) - \frac{1}{2} F_Y(x) \right| + \left| \frac{1}{2} + \frac{1}{2} F_Y(x) - F_X(x) \right| dx \\ &= \int_{-\infty}^{\infty} -\frac{1}{2} + \left| F_X(x) - \frac{1}{2} F_Y(x) \right| + \left| F_X(x) - \frac{1}{2} (1 + F_Y(x)) \right| dx \end{aligned}$$

CPD_S is equivalent to the Anderson-Darling functional (cf. [77]) and has been used by [78] for a goodness-of-fit test, where F_X represents the empirical distribution. Likewise, CPD_S serves as a goodness-of-fit test (cf. [79]).

Further work in this area with similar concepts was done by [80,81], using the notation cumulative residual Kullback-Leiber (CRKL) information and cumulative Kullback-Leiber (CKL) information.

Based on work from [82–85] a general function ϕ_{α} was discussed by [86]:

$$\phi_{\alpha}(u) = \begin{cases} (\alpha - 1 - \alpha u + u^{\alpha})/(\alpha(1 - \alpha)) & \text{for } \alpha \neq 0, 1 \\ -u(\ln u - 1) - 1 & \text{for } \alpha = 1 \\ \ln u - u + 1 & \text{for } \alpha = 0. \end{cases}$$

Up to a multiplicative constant, ϕ_{α} includes all of the aforementioned examples. In addition, the Hellinger distance is a special case for $\alpha = 1/2$ that leads to the cumulative paired Hellinger divergence:

$$CPD_H(X, Y) = 2 \int_{-\infty}^{\infty} \left(\sqrt{F_X(x)} - \sqrt{F_Y(x)} \right)^2 + \left(\sqrt{\bar{F}_X(x)} - \sqrt{\bar{F}_Y(x)} \right)^2 dx.$$

For a strictly concave function ϕ , Chen et al. [41] proved that $CPE_{\phi}(X, Y) \geq 0$ and $CPE_{\phi}(X, Y) = 0$ iff X and Y have identical distributions. Thus, the cumulative paired ϕ -divergence can be interpreted as a kind of a distance between distribution functions. As an application, Chen et al. [41] mentioned the “minimum cross-entropy principle”. They proved that X follows a logistic distribution if CPD_S is minimized, given that Y is exponentially distributed and the variance of X is fixed. If F_Y is an empirical distribution and F_X has an unknown vector of parameters θ , CPD_{ϕ} can be minimized to attain a point estimator for θ (cf. [87]). The large class of goodness-of-fit tests based on CPD_{ϕ} , discussed by Jager et al. [86], has already been mentioned.

8.2. Mutual Cumulative ϕ -Information

Let X and Y again be random variables with cdfs F_X , F_Y , density functions f_X , f_Y , and the conditional distribution function $F_{Y|X}$. D_X and D_Y denote the supports of X and Y . Then we have

$$CPE_{\phi}(Y|x) = \int_{-\infty}^{\infty} \phi(F_{Y|X}(y|x)) dy + \int_{-\infty}^{\infty} \phi(1 - F_{Y|X}(y|x)) dy, \quad (46)$$

which is the variation of Y given $X = x$. Averaging with respect to x leads to the internal variation

$$E_X(CPE_{\phi}(Y|X)) = \int_{-\infty}^{\infty} CPE_{\phi}(Y|x) f_X(x) dx. \quad (47)$$

For a concave entropy generating function ϕ , this internal variation cannot be greater than the total variation $CPE_{\phi}(Y)$. More precisely, it holds:

1. $E_X(CPE_{\phi}(Y|X)) \leq CPE_{\phi}(Y)$.
2. $E_X(CPE_{\phi}(Y|X)) = CPE_{\phi}(Y)$ if X and Y are independent.
3. If ϕ is strictly concave and $E_X(CPE_{\phi}(Y|X)) = CPE_{\phi}(Y)$, X and Y are independent random variables.

We consider the non-negative difference

$$MCPI_{\phi}(X, Y) := CPE_{\phi}(Y) - E_X(CPE_{\phi}(Y|X)). \quad (48)$$

This expression measures the part of the variation of Y that can be explained by the variable X (= external variation) and shall be named “mutual cumulative paired ϕ -information” $MCPI_{\phi}$ (cf. Rao et al. [46] using the term “cross entropy”, (p. 3) in [50]). $MCPI_{\phi}$ is equivalent to the transinformation that is defined for Shannon’s differential entropy (cf. [60], p. 20f.). In contrast to transinformation, $MCPI_{\phi}$ is not symmetric, so $MCPI_{\phi}(X, Y) = MCPI_{\phi}(Y, X)$ is not true in general.

Cumulative paired mutual ϕ -information is the starting point for two directed measures of strength of ϕ -dependence between X and Y , namely “directed (measure) of cumulative paired ϕ -dependence”, $DCPD$. The first one is

$$DCPD_{\phi}^1(X \rightarrow Y) = \frac{MCPI_{\phi}(X, Y)}{CPE_{\phi}(Y)}$$

and the second one is

$$DCPD_{\phi}^2(X \rightarrow Y) = \frac{CPE_{\phi}(Y)^2 - E_X(CPE_{\phi}(Y|X)^2)}{CPE_{\phi}(Y)^2}.$$

Both expressions measure the relative decrease in variation of Y if X is known. The domain is $[0, 1]$. The lower bound 0 is taken if Y and X are independent, while the upper bound 1 corresponds to $E_X(CPE_{\phi}(Y|X)) = 0$. In this case, from $\phi(u) > 0$ for $0 < u < 1$ and $\phi(0) = \phi(1) = 0$, we can conclude that the conditional distribution $F_{Y|X}(y|x)$ has to be degenerated. Thus, for every $x \in D_X$ there is exactly one $y^* \in D_Y$ with $P(Y = y^*|X = x) = 1$. Therefore, there is a perfect association between X and Y . The next example illustrates these concepts and demonstrates the advantage of considering both types of measures of dependence.

Example 8. Let (X, Y) follow a bivariate standard Gaussian distribution with $E(X) = E(Y) = 0$, $\text{Var}(X) = \text{Var}(Y) = 1$, and $\text{Cov}(X, Y) = \rho$, $-1 < \rho < 1$. Note that X and Y follow univariate standard Gaussian distributions, whereas $X + Y$ follows a univariate Gaussian distribution with mean 0 and variance $2(1 + \rho)$. Considering this, one can conclude that

$$F_X^{-1}(u) = F_Y^{-1}(u) = \Phi^{-1}(u), \quad F_{X+Y}^{-1}(u) = \sqrt{2(1 + \rho)}\Phi^{-1}(u), \quad u \in [0, 1].$$

By plugging this quantile function into the defining equation of the cumulative paired ϕ -entropy one yields

$$CPE_{\phi}(X + Y) = \sqrt{2(1 + \rho)}CPE_{\phi}(X) \leq CPE_{\phi}(X) + CPE_{\phi}(Y).$$

For $\rho \rightarrow -1$, the cumulative paired ϕ -entropy behaves like the variance or the standard deviation. All measures approach 0 for $\rho \rightarrow -1$, such that CPE_ϕ can be used as a measure of risk since the risk can be completely eliminated in a portfolio with perfectly negative correlated returns of assets. To be more precise, it is to say that CPE_ϕ rather behaves like the standard deviation than the variance.

For $\rho = 0$, the variance of the sum equals the sum of the variances, but the standard deviation of the sum is equal to or smaller than the sum of the individual standard deviations. This is also true for CPE_ϕ .

In case of the bivariate standard Gaussian distribution, $Y|x$ is Gaussian as well with mean ρx and variance $1 - \rho^2$ for $x \in \mathbb{R}$ and $-1 < \rho < 1$. Therefore, the quantile function of $Y|x$ is

$$F_{Y|x}^{-1}(u) = \rho x + \sqrt{1 - \rho^2} \Phi^{-1}(u), \quad u \in [0, 1].$$

Using this quantile function, the cumulative paired ϕ -entropy for the conditional random variable $Y|x$ is

$$CPE_\phi(Y|x) = \sqrt{1 - \rho^2} \int_0^1 (\phi'(1 - u) - \phi'(u)) \Phi^{-1}(u) du = \sqrt{1 - \rho^2} CPE_\phi(Y).$$

Just like the variance of $Y|x$, CPE_ϕ does not depend on x in case of a bivariate Gaussian distribution. This implies that the internal variation is $\sqrt{1 - \rho^2} CPE_\phi(Y)$, as well.

For $\rho \rightarrow 1$, the bivariate distribution becomes degenerated and the internal variation consequently approaches 0. The mutual cumulative paired ϕ -information is given by

$$MCPI_\phi(X, Y) = CPE_\phi(Y) - E_Y(CPE_\phi(Y|X)) = (1 - \sqrt{1 - \rho^2}) CPE_\phi(Y).$$

$MCPI_\phi$ takes the value 0 if and only if $\rho^2 = 0$, in which case X and Y are independent.

The two measures of directed cumulative ϕ -dependence for this example are

$$DCPD_\phi^1(X \rightarrow Y) = \frac{MCPI_\phi(X, Y)}{CPE_\phi(Y)} = 1 - \sqrt{1 - \rho^2}$$

and

$$DCPD_\phi^2(X \rightarrow Y) = \frac{CPE_\phi(Y)^2 - E_X(CPE_\phi(Y|X)^2)}{CPE_\phi(Y)^2} = \rho^2.$$

ρ completely determines the values for both measures of directed dependence. Provided the upper bound 1 will be attained, there is a perfect linear relation between Y and X .

As a second example we consider the dependence structure of the Farlie-Gumbel-Morgenstern copula (FGM copula). For the sake of brevity, we define a copula C as bivariate distribution function with uniform marginals for two random variables U and V with support $[0, 1]$. For details concerning copulas see, e.g., [88].

Example 9. Let

$$C_{U,V}(u, v) = uv + \theta u(1 - u)v(1 - v), \quad u, v \in [0, 1], \theta \in [-1, 1],$$

be the FGM copula (cf. [88], p. 68). With

$$C_{U|V}(u|v) = \frac{\partial C(u, v)}{\partial v} = u + \theta u(1 - u)(1 - 2v)$$

it holds for the conditional cumulative ϕ -entropy of U given $V = v$ that

$$CPE_\phi(C_{U|V}) = \int_0^1 \phi(1 - u - \theta u(1 - u)(1 - 2v)) + \phi(u + \theta u(1 - u)(1 - 2v)) du.$$

To get expressions in closed form we consider the Gini case with $\phi(u) = u(1-u)$, $u \in [0, 1]$. After some simple calculations we have

$$CPE_G(C_{U|V}) = \frac{1}{3} - \frac{\theta^2}{15}(1-2v)^2, \quad v \in [0, 1].$$

Averaging over the uniform distribution of V leads to the internal variation

$$E(CPE_G(C_{U|V})) = \frac{1}{3} - \frac{\theta^2}{45}.$$

With $CPE_G(U) = 1/3$, the mutual cumulative Gini information and the directed cumulative measure of Gini dependence are

$$MCI_G(V \rightarrow U) = \frac{\theta^2}{45} \quad \text{and} \quad DCPD_G^1(V \rightarrow U) = \frac{\theta^2}{15}.$$

It is well-known that only a small range of dependence can be covered by the FGM copula (cf. [88], p. 129).

Hall et al. [89] discussed several methods for estimating a conditional distribution. The results can be used for estimating the mutual ϕ -information and the two directed measures of dependence. This will be the task of future research.

8.3. ϕ -Correlation

Schechtman et al. [90] introduced Gini correlations of two random variables X and Y with distribution functions F_X and F_Y as

$$\Gamma_G(X, Y) = \frac{\text{Cov}(X, F_Y(Y))}{\text{Cov}(X, F_X(X))} \quad \text{and} \quad \Gamma_G(Y, X) = \frac{\text{Cov}(Y, F_X(X))}{\text{Cov}(Y, F_Y(Y))}.$$

The numerator equals $1/4$ of the Gini mean difference

$$\Delta_X = E_{X_1} E_{X_2} [|X_1 - X_2|],$$

where the expectation is calculated for two independent and with F_X identically distributed random variables X_1 and X_2 .

Gini's mean difference coincides with the cumulative paired Gini entropy $CPE_G(X)$ in the following way:

$$\text{Cov}(X, F_X(X)) = 4CPE_G(X) = 4 \int_{-\infty}^{\infty} X(\phi'(\bar{F}_X(X)) - \phi'(F_X(X))) dx.$$

Therefore, in the same way that Gini's mean difference can be generalized to the Gini correlation, CPE_ϕ can be generalized to the ϕ -correlation.

Let X, Y be two random variables and let $CPE_\phi(X), CPE_\phi(Y)$ be the corresponding cumulative paired ϕ -entropies, then

$$\Gamma_\phi(X, Y) = \frac{E(X(\phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y))))}{CPE_\phi(X)} \quad (49)$$

and

$$\Gamma_\phi(Y, X) = \frac{E(Y(\phi'(\bar{F}_X(X)) - \phi'(F_X(X))))}{CPE_\phi(Y)} \quad (50)$$

are called ϕ -correlations of X and Y . Since $E(\phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y))) = 0$, the numerator is the covariance between X and $\phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y))$.

The first example verifies that the Gini correlation is a proper special case of the ϕ -correlation.

Example 10. The setting $\phi(u) = u(1 - u)$, $u \in [0, 1]$, leads to the Gini correlation, because

$$\begin{aligned} E(X(\phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y)))) &= 2E(X(2F_Y(Y) - 1)) = 4E(X(F_Y(Y) - 1/2)) \\ &= 4E((X - E(X))(F_Y(Y) - 1/2)) = 4\text{Cov}(X, F_Y(Y)) \end{aligned}$$

and

$$E(X(\phi'(\bar{F}_X(X)) - \phi'(F_X(X)))) = 4\text{Cov}(X, F_X(X)).$$

The second example considers the new Shannon correlation.

Example 11. Set $\phi(u) = -u \ln u$, $u \in [0, 1]$, then we get the Shannon correlation

$$\Gamma_S(X, Y) = \frac{E(X \ln(F_Y(Y)/(1 - F_Y(Y))))}{\text{CPE}_S(X)}.$$

If Y follows a logistic distribution with $F_Y(y) = 1/(1 + e^{-y})$, $y \in \mathbb{R}$, then $\ln(F_Y(y)/\bar{F}_Y(y)) = y$. Considering this, we get

$$\Gamma_S(X, Y) = \frac{E(XY)}{\text{CPE}_S(X)}.$$

From Equation (30) we know that $\text{CPE}_S(X) = \pi/\sqrt{3}$ if X is logistically distributed. In this specific case we get

$$\Gamma_S(X, Y) = \sqrt{3} \frac{E(XY)}{\pi}.$$

In the following example we introduce the α -correlation.

Example 12. For $\phi(u) = u(u^{\alpha-1} - 1)/(1 - \alpha)$, $u \in [0, 1]$, we get the α -correlation

$$\Gamma_\alpha(X, Y) = \frac{E(X \frac{\alpha}{1-\alpha} (F_Y(Y)^{\alpha-1} - \bar{F}_Y(Y)^{\alpha-1}))}{\text{CPE}_\alpha(X)}.$$

For $F_Y(y) = 1/(1 + e^{-y})$, $y \in \mathbb{R}$, we get

$$\Gamma_S(X, Y) = \frac{\alpha}{(1 - \alpha)\text{CPE}_S(X)} E \left(X \left(\left(\frac{1}{1 + e^{-Y}} \right)^{\alpha-1} - \left(\frac{1}{1 + e^Y} \right)^{\alpha-1} \right) \right).$$

The authors of [90–92] proved that Gini correlations possess many desirable properties. In the following we give an overview of all properties which can be transferred to ϕ -correlations. For proofs and further details we refer to [75].

We start with the fact that ϕ -correlations also have a copula representation since for the covariance holds

$$\text{Cov}(X, F_Y(Y)) = - \int_0^1 \int_0^1 (C(u, v) - uv) \frac{1}{f(F_X^{-1}(u))} (\phi''(1 - v) + \phi''(v)) du dv.$$

The following examples demonstrate the copula representation for the Gini and the Shannon correlation.

Example 13. In the Gini case it is $\phi''(u) + \phi''(1 - u) = -4$. This leads to

$$\text{Cov}(X, F_Y(Y)) = 4 \int_0^1 \int_0^1 (C_{X,Y}(u, v) - uv) \frac{1}{f_X(F_X^{-1}(u))} du dv.$$

Example 14. In the Shannon case, $\phi''(u) + \phi''(1-u) = -1/(u(1-u))$ such that

$$\text{Cov}\left(X, \ln \frac{F_Y(Y)}{\bar{F}_Y(Y)}\right) = \int_0^1 \int_0^1 \frac{C_{X,Y}(u,v) - uv}{u(1-u)} \frac{1}{f_X(F_X^{-1}(u))} dudv.$$

The following basic properties of ϕ -correlations can easily be checked with the arguments applied by [90]:

1. $\Gamma_\phi(X, Y) \in [-1, 1]$.
2. $\Gamma_\phi(X, Y) = 1$ (-1) if there is a strictly increasing (decreasing) transformation g such that $X = g(Y)$.
3. If g is monotonic, then $\Gamma_\phi(X, Y) = \Gamma_\phi(X, g(Y))$.
4. If g is affin-linear, then $\Gamma_\phi(X, Y) = \Gamma_\phi(g(X), Y)$.
5. If X and Y are independent, then $\Gamma_{X,Y} = \Gamma(Y, X) = 0$.
6. If $a + bX$ and $c + dY$ are exchangeable for some constants $a, b, c, d \in \mathbb{R}$ with $b, d > 0$, then $\Gamma_\phi(X, Y) = \Gamma_\phi(Y, X)$.

In the last subsection we have seen that two directed measures of ϕ -dependence do not rely on ϕ if a bivariate Gaussian distribution is considered. The same holds for ϕ -correlations as will be demonstrated in the following example.

Example 15. Let (X, Y) be a bivariate standard Gaussian random variable with Pearson correlation coefficient ρ . Thus, all ϕ -correlations coincide with ρ as the following consideration shows:

With $E(X|Y) = \rho Y$ it is

$$\begin{aligned} \text{Cov}(X, \phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y))) &= E_Y E_{X|Y}(X|Y)(\phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y))) \\ &= \rho E_Y(Y(\phi'(\bar{F}_Y(Y)) - \phi'(F_Y(Y)))) \\ &= \rho CPE_\phi(Y) = \rho CPE_\phi(X). \end{aligned}$$

Dividing this by $CPE_\phi(X)$ yields the result.

Weighted sums of random variables appear for example in portfolio optimization. The diversification effect concerns negative correlations between the returns of assets. Thus, the risk of a portfolio can be significantly smaller than the sum of the individual risks. Now, we analyze whether cumulative paired ϕ -entropies can serve as a risk measure as well. Therefore, we have to examine the diversification effect for CPE_ϕ .

First, we display the total risk $CPE_\phi(Y)$ as a weighted sum of individual risks. Essentially, the weights need to be the ϕ -correlations of the individual returns with the portfolio return: Let $Y = \sum_{i=1}^k a_i X_i$, then it holds that

$$CPE_\phi(Y) = \sum_{i=1}^k a_i \Gamma_\phi(X_i, Y) CPE_\phi(X_i). \quad (51)$$

For the diversification effect the total risk $CPE_\phi(Y)$ has to be displayed as a function of the ϕ -correlations between X_i and X_j , $i, j = 1, 2, \dots, k$. A similar result was provided by [92] for the Gini correlation without proof. Let $Y = \sum_{i=1}^k a_i X_i$ and set $D_{iy} = \Gamma_\phi(X_i, Y) - \Gamma_\phi(Y, X_i)$, $i = 1, 2, \dots, k$, then the following decomposition of the square of $CPE_\phi(Y)$ holds:

$$\begin{aligned} CPE_\phi(Y)^2 - CPE_\phi(Y) \sum_{i=1}^k a_i D_{iy} CPE_\phi(X_i) \\ = \sum_{i=1}^k a_i^2 CPE_\phi(X_i)^2 + \sum_{i=1}^k \sum_{j \neq i}^k a_i a_j CPE_\phi(X_i) CPE_\phi(X_j) \Gamma_\phi(X_i, X_j). \end{aligned}$$

This is similar to the representation for the variance of Y , where $\Gamma_\phi(X_i, X_j)$ takes the role of the Pearson correlation and $CPE_\phi(X_i)$ the role of the standard deviation for $i, j = 1, 2, \dots, k$.

Schechtman et al. [90] also introduced an estimator for the Gini correlation and derived its asymptotic distribution. For the proof it is useful to note that the numerator of the Gini correlation can be represented as a U -statistic. For the general case of the ϕ -correlation it is necessary to derive the influence function and to calculate its variance. This will be done in [75].

8.4. ϕ -Regression

Based on the Gini correlation Olkin et al. [93] considered the traditional ordinary least squares (OLS) approach in regression analysis

$$Y_i = \alpha + x_i' \beta + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where Y is the dependent variable and x is the independent variable. They modified it by minimizing the covariance between the error term ε in a linear regression model and the ranks of ε with respect to α and β . Ranks are the sample analogue of the theoretical distribution function F_ε , such that the Gini mean difference $Cov(\varepsilon, F_\varepsilon)$ is the center of this new approach for regression analysis. Olkin et al. [93] noticed that this approach is already known as “rank based regression” or short “ R regression” in robust statistics. In robust regression analysis the more general optimization criteria $Cov(\varepsilon, \varphi(F_\varepsilon))$ has been considered, where φ denotes a strictly increasing score function (cf. [94], p. 233). The choice $\varphi(u) = 1 - 2u$ leads to the Gini mean difference, which is the scores generating function of the Wilcoxon scores. The rank based regression approach with general scores generating function $\varphi(u) = \phi'(1 - u) - \phi'(u)$, $u \in [0, 1]$, is equivalent to the generalization of the Gini regression to a so-called ϕ -regression based on the criteria function

$$CPE_\phi(\varepsilon) = Cov(\varepsilon, \phi'(1 - F_\varepsilon(\varepsilon)) - \phi'(F_\varepsilon)), \quad (52)$$

which has to be minimized to obtain α and β . Therefore, cumulative paired ϕ -entropies are special cases of the dispersion function that [95,96] proposed as optimization criteria for R regression. More precisely, R estimation proceeds in two steps. In the first step

$$d_\phi(\beta) = CPE_\phi(y - X\beta) \quad (53)$$

has to be minimized with respect to β . Let $\hat{\beta}_\phi$ denote this estimator. In the second step α will be estimated separately by

$$\hat{\alpha}_\phi = \text{med}_i(y_i - x_i' \hat{\beta}_\phi). \quad (54)$$

The authors of [97,98] gave an overview of recent developments in rank based regression. We will apply their main results to ϕ -regression. In [99], the authors showed that the following property holds for the influence function of $\hat{\beta}_\phi$:

$$IF(x_0, y_0; \hat{\beta}_\phi, F_{Y,X}) = \tau_\phi((X'X)/n)^{-1} (\phi'(\bar{F}_\varepsilon(y_0)) - \phi'(F_\varepsilon(y_0))) x_0,$$

where (x_0', y_0) represents an outlier. ϕ' determines the influence of an outlier in the dependent variable on the estimator $\hat{\beta}_\phi$.

The scale parameter τ_ϕ is given by

$$\tau_\phi = - \left(\int_0^1 (\phi'(1 - u) - \phi'(u)) \frac{f'_\varepsilon(F_\varepsilon^{-1}(u))}{f_\varepsilon(F_\varepsilon^{-1}(u))} du \right)^{-1}.$$

The influence function shows that $\hat{\beta}_\phi$ is asymptotically normal:

$$\hat{\beta}_\phi \sim_{asy} N\left(\beta, \tau_\phi^2 (X'X)^{-1}\right). \quad (55)$$

For $\phi'(1-u) - \phi'(u)$ bounded, Koul et al. [100] proposed a consistent estimator $\hat{\tau}_\phi$ for the scale parameter τ_ϕ . This asymptotic property can again be used to construct approximate confidence limits for the regression coefficients, to derive a Wald test for the general linear hypothesis, to derive a goodness-of-fit test, and to define a measure of determination (cf. [97]).

Gini regression corresponds to $CPE_G(\varepsilon, F_\varepsilon(\varepsilon))$. In the same way we can derive from $CPE_S(\varepsilon, F_\varepsilon(\varepsilon))$ the new Shannon regression, from $CPE_\alpha(\varepsilon, F_\varepsilon(\varepsilon))$ the α -regression, and from $CPE_L(\varepsilon, F_\varepsilon(\varepsilon))$ the Leik regression.

The R package “Rfit” has the option to include individual ϕ -functions into rank based regression (cf. [97]). Using this option and the dataset “telephone”, which is available with several outliers in “Rfit”, we compare the fit of the Shannon regression ($\alpha \rightarrow 1$), the Leik regression, and the α -regression (for several values of α) with the OLS regression. Figure 3 shows on the left the original data, the OLS, and the Shannon regression, while on its right side outliers were excluded to get a more detailed impression of the differences between the ϕ -regressions.

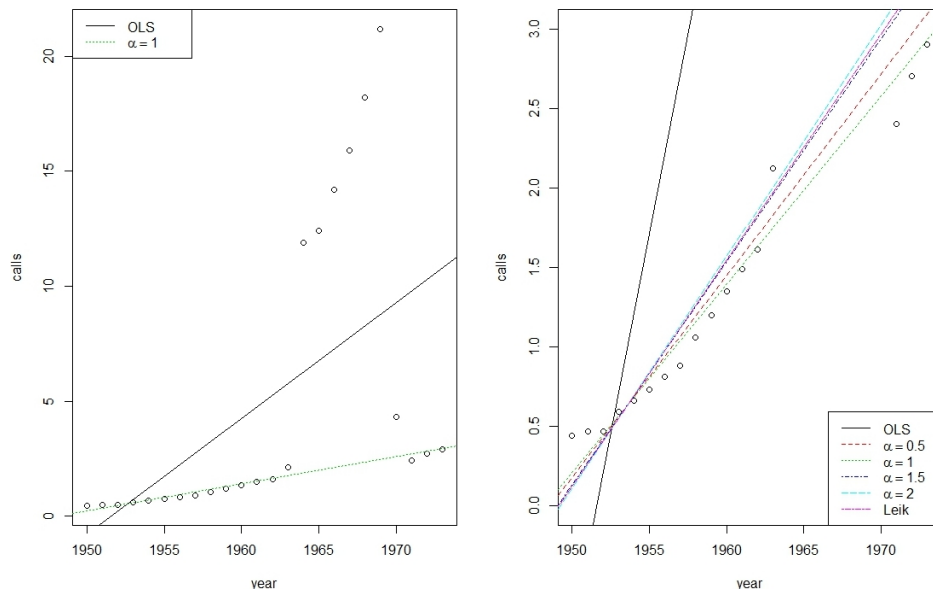


Figure 3. ϕ -regression fit for the number of calls in the “telephone” data set.

In comparison with the very sensitive OLS regression all rank based regression techniques behave similarly. In case of a known error distribution, McKean et al. [98] showed an asymptotically efficient estimator for τ_ϕ . This procedure also determines the entropy generating function ϕ . In case of an unknown error distribution but some available information with respect to skewness and leptokurtosis, a data-driven (adaptive) procedure was proposed by them.

8.5. Two-Sample Rank Test on Dispersion

Based on CPE_ϕ the linear rank statistics

$$CPE_\phi(R) = \sum_{i=1}^n \phi\left(\frac{R_i}{n+m+1}\right) + \phi\left(1 - \frac{R_i}{n+m+1}\right) \quad (56)$$

can be used as a test statistic for alternatives of scale, where R_1, R_2, \dots, R_n are the ranks of X_1, X_2, \dots, X_n in the pooled sample $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$. All random variables are assumed to be independent.

Some of the linear rank statistics which are well-known from the literature are special cases of Equation (56) as will be shown in the following examples:

Example 16. Let $\phi(u) = 1/2 - |u - 1/2|$, $u \in [0, 1]$, then we have

$$CPE_L(R) = 2 \sum_{i=1}^n \left(\frac{1}{2} - \left| \frac{R_i}{n+m+1} - \frac{1}{2} \right| \right).$$

Ansari et al. [101] suggest the statistic

$$S_{AB} = \sum_{i=1}^n \left(\frac{1}{2}(n+m+1) - \left| R_i - \frac{1}{2}(n+m+1) \right| \right) \quad (57)$$

as a two-sample test for alternatives of scale (cf. [102], p. 104). Apparently, we have $S_{AB} = 1/2(n+m+1)CPE_L(R)$.

Example 17. Let $\phi(u) = 1/4 - (u - 1/2)^2$, $u \in [0, 1]$. Consequently, we have

$$CPE_G(R) = \frac{n}{2} - 2 \sum_{i=1}^n \left(\frac{R_i}{n+m+1} - \frac{1}{2} \right)^2,$$

which is identical to the test statistic suggested by [103] up to an affine linear relation (cf. [68], p. 149f.). This test statistic is given by $S_M = \sum_{i=1}^n (R_i - (n+m+1)/2)^2$, thus, the resulting relation is given by

$$CPE_\phi(R) = \frac{n}{2} - 2(n+m+1)^2 S_M.$$

In the following, the scores of the Mood test will be generated by the generating function of CPE_G .

Dropping the requirement of concavity of ϕ , one finds analogies to other well-known test statistics.

Example 18. Let $\phi(u) = 1/2 - 1/2(\text{sign}(|u - 1/2| - 1/4) + 1)$, $u \in [0, 1]$, which is not concave on the interval $[0, 1]$, we have

$$CPE_\phi(R) = n - \sum_{i=1}^n \left(\text{sign} \left(\left| \frac{R_i}{n+m+1} - \frac{1}{2} \right| - \frac{1}{4} \right) + 1 \right),$$

which is identical to the quantile test statistic for alternatives of scale up to an affine linear relation ([102], p. 105).

The asymptotic distribution of linear rank tests based on CPE_ϕ can be derived from the theory of linear rank test, as discussed in [102]. The asymptotic distribution under the null hypothesis is needed to be able to make an approximate test decision given a significance level α . The asymptotic distribution under the alternative hypothesis is needed for an approximate evaluation of the test power and the choice of the required sample size in order to ensure a given effect size, respectively.

We consider the centered linear rank statistic

$$\overline{CPE}_\phi(R) = CPE_\phi(R) - \frac{2n}{n+m} \sum_{i=1}^{n+m} \phi \left(\frac{i}{n+m+1} \right).$$

Under the null hypothesis of identical scale parameters and the assumption that

$$\int_0^1 (\phi(u) - \bar{\phi})^2 + (\phi(u) - \bar{\phi})(\phi(1-u) - \bar{\phi}) du > 0,$$

where $\bar{\phi} = \int_0^1 \phi(u) du$, the asymptotical distribution of $\overline{CPE}_\phi(R)$ is given by

$$\overline{CPE}_\phi(R) \sim_{asy} N\left(0, \frac{2nm}{n+m} \int_0^1 \left(\phi(u) - \bar{\phi}\right)^2 + (\phi(u) - \bar{\phi})(\phi(1-u) - \bar{\phi})\right)$$

(cf. [102], p. 194, Theorem 1 and p. 195, Lemma 1).

The property of asymptotic normality of the Ansari-Bradley test and the Mood test is well-known. Therefore, we provide a new linear rank test based on cumulative paired Shannon entropy CPE_S (so-called “Shannon”-test) in the following example:

Example 19. With $\phi(u) = -u \ln u$, $u \in [0, 1]$, and $\bar{\phi} = 1/4$ we have

$$\int_0^1 (\phi(u) - \bar{\phi})^2 du = \int_0^1 \phi(u)^2 du - \frac{1}{16} = \int_0^1 u^2 (\ln u)^2 du - \frac{1}{16} = \frac{2}{27} - \frac{1}{16} = \frac{5}{432}$$

and

$$\begin{aligned} \int_0^1 (\phi(u) - \bar{\phi})(\phi(1-u) - \bar{\phi}) du &= \int_0^1 \phi(u)\phi(1-u) du - \frac{1}{16} \\ &= \int_0^1 u(1-u) \ln u \ln(1-u) du - \frac{1}{16} \\ &= \frac{37 - 3\pi^2}{108} - \frac{1}{16} = \frac{121 - 12\pi^2}{432}. \end{aligned}$$

Under the null hypothesis of identical scale, the centered linear rank statistic $\overline{CPE}_S(R)$ is asymptotically normal with variance

$$\frac{nm}{n+m} \frac{63 - 6\pi^2}{108}.$$

If the alternative hypothesis H_1 for a density function f_0 is given by

$$f(x_1, \dots, x_{n+m}; \sigma) = \prod_{i=1}^n \frac{1}{\sigma} f_0\left(\frac{x_i}{\sigma}\right) \prod_{i=n+1}^{n+m} f_0(x_i) \quad (58)$$

for $\sigma > 0$ and $\sigma \neq 1$, then set

$$\varphi_1(u; f_0) = -1 - F_0^{-1}(u) \frac{f_0'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}$$

and assume $I(f_0) = \int_0^1 \varphi_1(u; f_0)^2 du > 0$. If $\min(n, m) \rightarrow \infty$ and $\ln \sigma I(f_0) mn / (n+m) \rightarrow b^2$ with $0 < b^2 < \infty$, $\overline{CPE}_\phi(R)$ is asymptotically normal distributed with mean

$$-\frac{n}{n+m} \ln \sigma \frac{mn}{n+m} \int_0^1 \phi(u) \varphi_1(u; f_0) + \phi(1-u) \varphi_1(u; f_0) du$$

and variance

$$\frac{2nm}{n+m} \int_0^1 \left(\phi(u) - \bar{\phi}\right)^2 + (\phi(u) - \bar{\phi})(\phi(1-u) - \bar{\phi}) du.$$

This result follows immediately from [102], p. 267, Theorem 1, together with the Remark on, p. 268.

If f_0 is a symmetric distribution, $\varphi_1(u; f_0) = \varphi_1(1-u; f_0)$, $u \in [0, 1]$, holds such that

$$\int_0^1 (2\bar{\phi} - \phi(u) - \phi(1-u)) \varphi_1(u; f_0) du = -2 \int_0^1 \phi(u) \varphi_1(u; f_0) du.$$

This simplifies the variance of the asymptotic normal distribution.

Since the asymptotic normality of the test statistic of the Ansari-Bradley test and the Mood test under the alternative hypothesis have been examined intensely (cf., e.g., [103,104]), we focus in the following example on the new Shannon test:

Example 20. Set $\phi(u) = -u \ln u$, $u \in [0, 1]$ and let f_0 be the density function of a standard Gaussian distribution, such that $\varphi_1(u; f_0) = -1 + \Phi^{-1}(u)^2$ and $I_1(f_0) = 1$. As a consequence, we have

$$-2 \int_0^1 (-u \ln u) (\Phi^{-1}(u)^2 - 1) du = 0.240,$$

and

$$\int_0^1 (1/2 + u \ln u + (1-u) \ln(1-u))^2 du = \frac{63 - 6\pi^2}{108},$$

where the integrals have been evaluated by numerical integration. Then under the alternative Equation (58):

$$\overline{CPE}_S(R) \sim_{asy} N \left(0.240 \frac{n}{n+m} \ln \sigma \frac{mn}{n+m}, \frac{63 - 6\pi^2}{108} \frac{2nm}{n+m} \right).$$

Hereafter, one can discuss the asymptotic efficiency of linear rank tests based on cumulative paired ϕ -entropy. If f_0 is the true density and

$$\rho_1 = \frac{\int_0^1 (\phi u) \varphi_1(u; f_0) + \phi(1-u) \varphi_1(u; f_0) du}{\sqrt{\int_0^1 \varphi_1(u; f_0)^2 du \int_0^1 (\phi(u) - \bar{\phi})^2 + (\phi(u) - \bar{\phi})(\phi(1-u) - \bar{\phi}) du}},$$

then ρ_1^2 gives the desired asymptotic efficiency (cf. [102], p. 317).

The asymptotic efficiency of the Ansari-Bradley test (and the asymptotic equivalent Siegel-Tukey test, respectively) and the Mood test have been analyzed by [104–106]. The asymptotic relative efficiency (ARE) with respect to the traditional F -test for differences in scale for two Gaussian distributions has been discussed by [103]. This asymptotic relative efficiency between Mood test and F -test for differences in scale has been derived by [107]. Once more, we focus on the new Shannon-test.

Example 21. The Klotz test is asymptotically efficient for the Gaussian distribution. With $\int_0^1 (\Phi^{-1}(u)^2 - 1)^2 du = 2$,

$$\rho_1^2 = \frac{0.24^2}{(63 - 6\pi^2)/108 \times 2} = 0.823$$

gives the asymptotic efficiency of the new Shannon test.

Using a distribution that ensures the asymptotic efficiency of the Ansari-Bradley test, we compare the asymptotic efficiency of the Shannon test to the one of the Ansari-Bradley test.

Example 22. The Ansari-Bradley test statistic S_{AB} is asymptotically efficient for the double log-logistic distribution with density function f_0 (cf. [102], p. 104). The Fisher information is given by

$$\int_0^1 \varphi_1(u; f_0)^2 du = \int_0^1 (2|2u-1| - 1)^2 du = 4 \int_0^1 (2u-1)^2 du - 1 = \frac{1}{3}.$$

Furthermore, we have

$$\begin{aligned} \int_0^1 \varphi_1(u; f_0) (2\bar{\phi} - \phi(u) - \phi(1-u)) du &= \int_0^1 \varphi_1(u; f_0) \left(\frac{1}{2} + u \ln u + (1-u) \ln(1-u) \right) du \\ &= 2 \int_0^1 |2u-1| (u \ln u + (1-u) \ln(1-u)) du = 0.102, \end{aligned}$$

such that the asymptotic efficiency of the Shannon-test for f_0 is

$$\rho_1^2 = \frac{0.102^2}{1/3 \times (63 - 2\pi^2)/108} = 0.892.$$

These two examples show that the Shannon test has a rather good asymptotic efficiency, even if the underlying distribution has moderate tails similar to the Gaussian distribution or heavy tails like the double log-logistic distribution. Asymptotic efficient linear rank tests correspond to a distribution and a scores generating function φ_1 , from which we can derive an entropy generating function ϕ and a cumulative paired ϕ -entropy. This relationship will be further examined in [74].

9. Some Cumulative Paired Entropies for Selected Distribution Functions

In the following, we derive closed form expressions for some cumulative paired ϕ -entropies. We mimic the procedure of ([4], p. 326) to some degree. Table 1 of their paper contains multiple formulas of the differential entropy for the most popular statistical distributions. Several of these distributions will also be considered in the following. Since cumulative entropies depend on the distribution function or equivalently on the quantile function, we focus on families of distributions for which these functions have a closed form expression. Furthermore, we only discuss standardized random variables since the parameter of scale only has a multiplicative effect on CPE_ϕ and the parameter of location has no effect. For the standard Gaussian distribution we provide the value of CPE_S by numerical integration rounded to two decimal places since the probability function has no explicit form. For the Gumbel distribution however, there is a closed form expression for the distribution function – nevertheless, we were unable to establish a closed form of CPE_S and CPE_G . Therefore, we applied numerical integration in this case as well. In the following, next to the Gamma function $\Gamma(a)$ and the Beta function $B(a, b)$, we use

- the incomplete Gamma function

$$\Gamma(x; a) = \int_0^x y^{a-1} e^{-y} dy \quad \text{for } x > 0, a > 0,$$

- the incomplete Beta function

$$B(x; a, b) = \int_0^x u^{a-1} (1-u)^{b-1} du \quad \text{for } 0 < x < 1, a, b > 0,$$

- and the Digamma function

$$\psi(a) = \frac{d}{da} \ln \Gamma(a), \quad a > 0.$$

9.1. Uniform Distribution

Let X have the standard uniform distribution. Then we have

$$CPE_S(X) = \frac{3}{2}, \quad CPE_G(X) = \frac{1}{3}, \quad CPE_L(X) = \frac{1}{2}, \quad CPE_\alpha(X) = \frac{1}{\alpha + 1}.$$

9.2. Power Distribution

Let X have the Beta distribution on $[0, 1]$ with parameter $\alpha > 0$ and $b = 1$, i.e., density function $f_X(x) = ax^{a-1}$ for $x \in [0, 1]$, then we have

$$\begin{aligned} CPE_S(X) &= \frac{a}{(a+1)^2} + \psi\left(\frac{a+1}{a}\right) - \frac{a+1}{a}\psi\left(\frac{a+2}{a}\right) + \frac{1}{a}\psi(1), \\ CPE_G(X) &= \frac{2a}{(1+a)(1+2a)}, \quad CPE_L(X) = \frac{a}{a+1} \left(1 - \left(\frac{1}{2}\right)^{1/a}\right), \\ CPE_\alpha(X) &= \frac{1}{a(1-\alpha)} B\left(\frac{1}{a}, \alpha+1\right) - \frac{\alpha a}{(1-\alpha)(1+\alpha a)}. \end{aligned}$$

9.3. Triangular Distribution with Parameter c

Let X have a triangular distribution with density function

$$f(x) = \begin{cases} 2x/c & \text{for } 0 < x < c \\ 2(1-x)/(1-c) & \text{for } c \leq x < 1. \end{cases}$$

Then the following holds:

$$\begin{aligned} CPE_S(X) &= \frac{\pi^2}{6} + \ln 2(1 - \ln 2), \\ CPE_G(X) &= \frac{2}{3} (c^2 + (1-c)^2) - \frac{2}{5} (c^3 + (1-c)^3), \\ CPE_L(X) &= \frac{1}{3}(2-c) - \frac{3-\sqrt{2}}{3\sqrt{2}}\sqrt{1-c}, \\ CPE_\alpha(X) &= \frac{1}{1-\alpha} \left(\frac{2}{2\alpha+1} (c^{\alpha+1} + (1-c)^{\alpha+1}) \right. \\ &\quad \left. + \sqrt{c} B\left(c; \frac{1}{2}, \alpha+1\right) + \sqrt{1-c} B\left(1-c; \frac{1}{2}, \alpha+1\right) - 2 \right). \end{aligned}$$

9.4. Laplace Distribution

Let X follow the Laplace distribution with density function $f_X(x) = 1/2 \exp(-|x|)$ for $x \in \mathbb{R}$, then we have

$$\begin{aligned} CPE_S(X) &= \frac{\pi^2}{6} + \ln 2(1 - \ln 2), \quad CPE_G(X) = \frac{3}{2}, \quad CPE_L(X) = 2, \\ CPE_\alpha(X) &= \frac{4}{\alpha-1} \left(\frac{1}{2}\right)^{\alpha-1} \left(\frac{1}{\alpha-1} - \frac{1}{2\alpha}\right). \end{aligned}$$

9.5. Logistic Distribution

Let X follow the logistic distribution with distribution function $F_X(x) = 1/(1 + \exp(-x))$ for $x \in \mathbb{R}$, then we have

$$\begin{aligned} CPE_S(X) &= \frac{\pi^2}{3}, \quad CPE_G(X) = 2, \quad CPE_L(X) = 4 \ln 2, \\ CPE_\alpha &= \frac{2}{\alpha-1} (\psi(\alpha) - \psi(1)). \end{aligned}$$

9.6. Tukey λ Distribution

Let X follow the Tukey λ distribution with quantile function $F^{-1}(U) = 1/\lambda (u^\lambda - (1-u)^{1-\lambda})$ for $0 \leq u \leq 1$ and $\lambda > -1$. Then the following holds:

$$\begin{aligned} CPE_S(X) &= \frac{2}{(\lambda+1)^2} \left(1 + \left(1 + \frac{1}{\lambda} \right) ((\lambda+1)\psi(\lambda+1) - \psi(\lambda+2) - \psi(1)) \right), \\ CPE_G(X) &= \frac{4}{\lambda+1} \left(1 + \frac{1}{\lambda} \right), \\ CPE_L(X) &= 2 \left(\frac{1}{\lambda+1} \left(\frac{1}{2} \right)^{\lambda+1} + B\left(\frac{1}{2}; 2, \lambda\right) \right), \\ CPE_\alpha(X) &= 2 \frac{1}{1-\alpha} \left(\frac{\lambda^3 - \lambda\alpha - 2(\lambda+\alpha)}{\lambda^2(\lambda+1)(\lambda+\alpha)} + B(\alpha+1, \lambda) \right). \end{aligned}$$

9.7. Weibull Distribution

Let X follow the Weibull distribution with distribution function $F_X(x) = 1 - e^{-x^c}$ for $x > 0, c > 0$, then we have

$$\begin{aligned} CPE_S(X) &= \frac{1}{c} \Gamma\left(\frac{1}{c}\right) \left(1 + \sum_{i=1}^{\infty} \frac{1}{i!} \left(\left(\frac{1}{i} \right)^{1/c} - \left(\frac{1}{i+1} \right)^{1/c} \right) \right), \\ CPE_G(X) &= \frac{2}{c} \left(\Gamma\left(\frac{1}{c}\right) - \frac{1}{2} \Gamma\left(\frac{1}{2c}\right) \right), \\ CPE_L(X) &= 2 \left((\ln 2)^{1/c} + \frac{1}{c} \left(\Gamma\left(\frac{1}{c}\right) - 2\Gamma\left(\ln 2; \frac{1}{c}\right) \right) \right), \\ CPE_\alpha(X) &= \frac{1}{c} \Gamma\left(\frac{1}{c}\right) \left(\frac{1}{\alpha^{1/c}} + \sum_{i=1}^{\infty} \binom{\alpha}{i} (-1)^i i^{-1/c} \right). \end{aligned}$$

9.8. Pareto Distribution

Let X follow the Pareto distribution with distribution function $F_X(x) = 1 - x^{-c}$ for $x > 1, c > 1$, then we have

$$\begin{aligned} CPE_S(X) &= \frac{1}{c-1} \psi\left(2 - \frac{1}{c}\right) + \psi\left(1 - \frac{1}{c}\right) - \frac{c}{c-1} \psi(1) + \frac{4}{c}, \\ CPE_G(X) &= \frac{2c}{(c-1)(2c-1)}, \\ CPE_L(X) &= 2 \frac{1}{c-1}, \\ CPE_\alpha(X) &= \frac{1}{1-\alpha} \left(\frac{c(1-\alpha)}{(c\alpha-1)(c-1)} - \frac{1}{c} B\left(\alpha, 1 - \frac{1}{c}\right) \right). \end{aligned}$$

9.9. Gaussian Distribution

By means of numerical integration we calculated the following values for the standard Gaussian distribution:

$$CPE_S(X) = 1.806, \quad CPE_G(X) = 1.128, \quad CPE_L(X) = 1.596.$$

CPE_α for $\alpha \in [0.5, 3]$ and the standard Gaussian distribution can be seen in Figure 4.

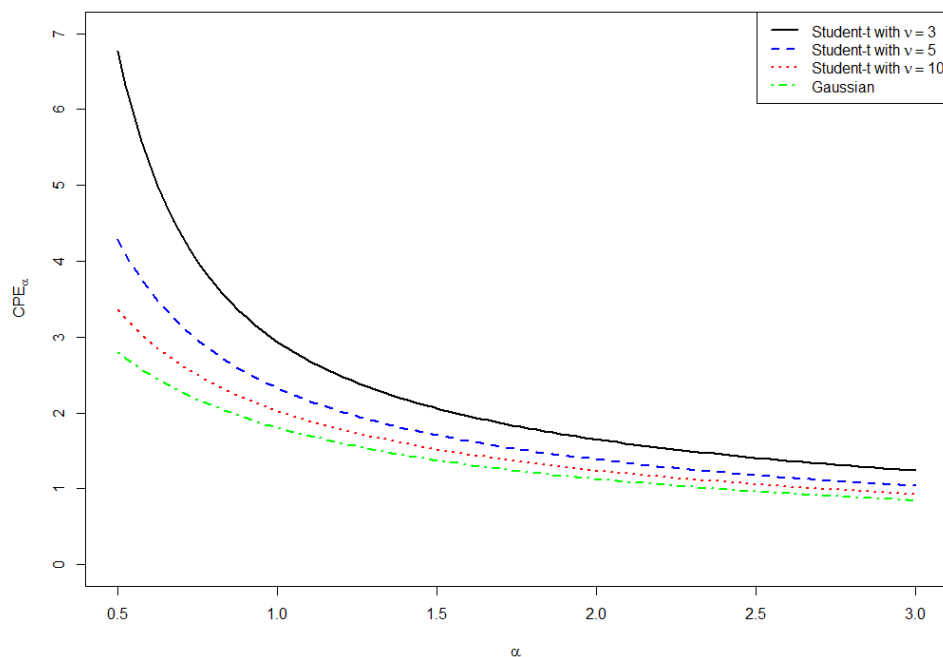


Figure 4. CPE_α , $\alpha \in [0.5, 3]$ for the standard Gaussian and the Student- t distribution.

9.10. Student- t Distribution

By means of numerical integration and for $\nu = 3$ degrees of freedom we calculated the following values for the Student- t distribution

$$CPE_S(X) = 2.947, \quad CPE_G(X) = 3.308, \quad CPE_L(X) = 2.205.$$

As can be seen in Figure 4, the heavy tails of the Student- t distribution result in a higher value for CPE_α as compared with the Gaussian distribution.

10. Conclusions

A new kind of entropy has been introduced that generalizes Shannon's differential entropy. The main difference to the previous discussion of entropies is the fact that the new entropy is defined for distribution functions instead of density functions. This paper shows that this definition has a long tradition in several scientific disciplines like fuzzy set theory, reliability theory, and more recently in uncertainty theory. With only one exception within all the disciplines, the concepts had been discussed independently. Along with that, the theory of dispersion measures for ordered categorical variables refers to measures based on distribution functions, without realizing that implicitly some sort of entropies are applied. Using the Cauchy-Schwarz inequality, we were able to show the close relationship between the new kind of entropy named cumulative paired ϕ -entropy and the standard deviation. More precisely, the standard deviation yields an upper limit for the new entropy. Additionally, the Cauchy-Schwarz inequality can be used to derive maximum entropy distributions provided that there are constraints specifying values of mean and variance. Here, the logistic distribution takes on the same key role for the cumulative paired Shannon entropy which the Gaussian distribution takes by maximizing the differential entropy. As a new result we have demonstrated that Tukey's λ distribution is a maximum entropy distribution if using the entropy generating function ϕ which is known from the Harvda and Charvát entropy. Moreover, some new distributions can be derived by considering more general constraints. A change in perspective allows to determine the entropy that will be maximized by a certain distribution if, e.g., mean and variance are known. In this context the Gaussian distribution gives a simple solution. Since cumulative paired ϕ -entropy

and variance are closely related, we have investigated whether the cumulative paired ϕ -entropy is a proper measure of scale. We show that it satisfies the axioms which were introduced by Oja for measures of scale. Several further properties, concerning the behavior under transformations or the sum of independent random variables, have been proven. Consequently, we have given first insights on how to estimate the new entropy. In addition, based on cumulative paired ϕ -entropy we have introduced new concepts like ϕ -divergence, mutual ϕ -information, and ϕ -correlation. ϕ -regression and linear rank tests for scale alternatives were considered as well. Furthermore, formulas have been derived for some popular distributions with cdf or quantile function in closed form and for certain cumulative paired ϕ -entropies.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive criticism, which helped to improve the presentation of this paper significantly. Furthermore, we would like to thank Michael Grottko for helpful advises.

Author Contributions: Ingo Klein conceived the new entropy concept, investigated its properties and wrote an initial version of the manuscript. Benedikt Mangold cooperated especially by checking, correcting and improving the mathematical details including the proofs. He examined the entropy's properties by simulation. Monika Doll contributed by mathematical and linguistic revision. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Burbea, J.; Rao, C.R. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 489–495.
- Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
- Oja, H. On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Stat.* **1981**, *8*, 154–168.
- Ebrahimi, N.; Massoumi, E.; Soofi, E.S. Ordering univariate distributions by entropy and variance. *J. Econometr.* **1999**, *90*, 317–336.
- Popoviciu, T. Sur les équations algébrique ayant toutes leurs racines réelles. *Mathematica* **1935**, *9*, 129–145. (In French)
- Liu, B. Uncertainty Theory. Available online: <http://orsc.edu.cn/liu/ut.pdf> (accessed on 27 June 2016).
- Wang, F.; Vemuri, B.C.; Rao, M.; Chen, Y. *A New & Robust Information Theoretic Measure and Its Application to Image Alignment: Information Processing in Medical Imaging*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2732, pp. 388–400.
- Di Crescenzo, A.; Longobardi, M. On cumulative entropies and lifetime estimation. In *Methods and Models in Artificial and Natural Computation*; Mira, J.M., Ed.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 132–141.
- Di Crescenzo, A.; Longobardi, M. On cumulative entropies. *J. Stat. Plan. Inference* **2009**, *139*, 4072–4087.
- Kapur, J.N. Derivation of logistic law of population growth from maximum entropy principle. *Natl. Acad. Sci. Lett.* **1983**, *6*, 429–433.
- Hartley, R. Transmission of information. *Bell Syst. Tech. J.* **1928**, *7*, 535–563.
- De Luca, A.; Termini, S. A definition of a nonprobabilistic entropy in the setting of fuzzy set theory. *Inf. Control* **1972**, *29*, 301–312.
- Zadeh, L. Probability measures of fuzzy events. *J. Math. Anal. Appl.* **1968**, *23*, 421–427.
- Pal, N.R.; Bezdek, J.C. Measuring fuzzy uncertainty. *IEEE Trans. Fuzzy Syst.* **1994**, *2*, 107–118.
- Rényi, A. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Oakland, CA, USA, 1961; pp. 547–561.
- Esteban, M.D.; Morales, D. A summary on entropy statistics. *Kybernetika* **1995**, *31*, 337–346.
- Cichocki, A.; Amari, S. Families of alpha-beta- and gamma-divergences: Flexible and robust measures of similarities. *Entropy* **2010**, *12*, 1532–1568.
- Arndt, C. *Information Measures*; Springer: Berlin/Heidelberg, Germany, 2004.
- Kesavan, H.K.; Kapur, J.N. The generalized maximum entropy principle. *IEEE Trans. Syst. Man Cyber.* **1989**, *19*, 1042–1052.
- Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
- Jaynes, E.T. Information theory and statistical mechanics. II. *Phys. Rev.* **1957**, *108*, 171–190.
- Leik, R.K. A measure of ordinal consensus. *Pac. Sociol. Rev.* **1966**, *9*, 85–90.

23. Vogel, H.; Dobbener, R. Ein Streuungsmaß für komparative Merkmale. *Jahrbücher für Nationalökonomie und Statistik* **1982**, *197*, 145–157. (In German)
24. Kvålseth, T.O. Nominal versus ordinal variation. *Percept. Mot. Skills* **1989**, *69*, doi:10.2466/pms.1989.69.1.234.
25. Berry, K.J.; Mielke, P.W. Assessment of variation in ordinal data. *Percept. Motor Skills* **1992**, *74*, 63–66.
26. Berry, K.J.; Mielke, P.W. Indices of ordinal variation. *Percept. Motor Skills* **1992**, *74*, 576–578.
27. Berry, K.J.; Mielke, P.W. A test of significance for the index of ordinal variation. *Percept. Motor Skills* **1994**, *79*, 291–295.
28. Blair, J.; Lacy, M.G. Measures of variation for ordinal data. *Percept. Motor Skills* **1996**, *82*, 411–418.
29. Blair, J.; Lacy, M.G. Statistics of ordinal variation. *Sociol. Methods Res.* **2000**, *28*, 251–280.
30. Gadrich, T.; Bashkansky, E.; Zitikas, R. Assessing variation: A unifying approach for all scales of measurement. *Qual. Quant.* **2015**, *49*, 1145–1167.
31. Allison, R.A.; Foster, J.E. Measuring health inequality using qualitative data. *J. Health Econ.* **2004**, *23*, 505–524.
32. Zheng, B. Measuring inequality with ordinal data: A note. *Res. Econ. Inequal.* **2008**, *16*, 177–188.
33. Abul Naga, R.H.; Yalcin, T. Inequality measurement for ordered response health data. *J. Health Econ.* **2008**, *27*, 1614–1625.
34. Zheng, B. A new approach to measure socioeconomic inequality in health. *J. Econ. Inequal.* **2011**, *9*, 555–577.
35. Apouey, B.; Silber, J. Inequality and bi-polarization in socioeconomic status and health: Ordinal approaches. *Res. Econ. Inequal.* **2013**, *21*, 77–109.
36. Klein, I. Rangordnungsstatistiken als Verteilungsmaßzahlen für ordinalskalierte Merkmale: I. Streuungsmessung. In *Diskussionspapiere des Lehrstuhls für Statistik und; Ökonometrie der Universität: Erlangen-Nürnberg, Germany*, 1999; Volume 27. (In German)
37. Yager, R.R. Dissonance—A measure of variability for ordinal random variables. *Int. J. Uncertain. Fuzzin. Knowl. Based Syst.* **2001**, *9*, 39–53.
38. Bowden, R.J. Information, measure shifts and distribution metrics. *Statistics* **2012**, *46*, 249–262.
39. Dai, W. Maximum entropy principle for quadratic entropy of uncertain variables. Available online: <http://orisc.edu.cn/online/100314.pdf> (accessed on 27 June 2016).
40. Dai, W.; Chen, X. Entropy of function of uncertain variables. *Math. Comput. Model.* **2012**, *55*, 754–760.
41. Chen, X.; Kar, S.; Ralescu, D.A. Cross-entropy measure of uncertain variables. *Inf. Sci.* **2012**, *201*, 53–60.
42. Yao, K.; Gao, J.; Dai, W. Sine entropy for uncertain variables. *Int. J. Uncertain. Fuzzin. Knowl. Based Syst.* **2013**, *21*, 743–753.
43. Yao, K.; Ke, H. Entropy operator for membership function of uncertain set. *Appl. Math. Comput.* **2014**, *242*, 898–906.
44. Ning, Y.; Ke, H.; Fu, Z. Triangular entropy of uncertain variables with application to portfolio selection. *Soft Comput.* **2015**, *19*, 2203–2209.
45. Ebrahimi, N. How to measure uncertainty in the residual lifetime distribution. *Sankhya Ser. A* **1996**, *58*, 48–56.
46. Rao, M.; Chen, Y.; Vemuri, B.C.; Wang, F. Cumulative residual entropy: A new measure of information. *IEEE Trans. Inf. Theory* **2004**, *50*, 1220–1228.
47. Rao, M. More on a new concept of entropy and information. *J. Theor. Probabil.* **2005**, *18*, 967–981.
48. Schroeder, M.J. An alternative to entropy in the measurement of information. *Entropy* **2004**, *6*, 388–412.
49. Zografos, K.; Nadarajah, S. Survival exponential entropies. *IEEE Trans. Inf. Theory* **2005**, *51*, 1239–1246.
50. Drissi, N.; Chonavel, T.; Boucher, J.M. Generalized cumulative residual entropy distributions with unrestricted supports. *Res. Lett. Signal Process.* **2008**, *2008*, doi:10.1155/2008/790607.
51. Chen, X.; Dai, W. Maximum entropy principle for uncertain variables. *Int. J. Fuzzy Syst.* **2011**, *13*, 232–236.
52. Sunoj, S.M.; Sankaran, P.G. Quantile based entropy function. *Stat. Probabil. Lett.* **2012**, *82*, 1049–1053.
53. Zardasht, V.; Parsi, S.; Mousazadeh, M. On empirical cumulative residual entropy and a goodness-of-fit test for exponentiality. *Stat. Pap.* **2015**, *56*, 677–688.
54. Navarro, J.; del Aguila, Y.; Asadi, M. Some new results on the cumulative residual entropy. *J. Stat. Plan. Inference* **2010**, *140*, 310–322.
55. Psarrakos, G.; Navarro, J. Generalized cumulative residual entropy and record values. *Metrika* **2013**, *76*, 623–640.
56. Kiesel, H. *Ordinale Streuungsmaße*; JOSEF-EUL-Verlag: Köln, Germany, 2003. (In German)
57. Havrda, J.; Charvát, F. Quantification method of classification processes. Concept of structural α -entropy. *Kybernetika* **1967**, *3*, 30–35.

58. Jumarie, G. *Relative Information: Theories and Applications*; Springer: Berlin/Heidelberg, Germany, 1990.
59. Kapur, J.N. *Measures of Information and their Applications*; New Age International Publishers: New Delhi, India, 1994.
60. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 1991.
61. Kapur, J.N. Generalized Cauchy and Students distributions as maximum entropy distributions. *Proc. Natl. Acad. Sci. India* **1988**, *58*, 235–246.
62. Bickel, P.J.; Lehmann, E.L. Descriptive statistics for nonparametric models: III. Dispersion. *Ann. Stat.* **1976**, *5*, 1139–1158.
63. Behnen, K.; Neuhaus, G. *Rank Tests with Estimated Scores and their Applications*; Teubner-Verlag: Stuttgart, Germany, 1989.
64. Burger, H.U. Dispersion orderings with applications to nonparametric tests. *Stat. Probabil. Lett.* **1993**, *16*, doi:10.1016/0167-7152(93)90114-X.
65. Bickel, P.J.; Lehmann, E.L. Descriptive statistics for nonparametric models: IV. Spread. In *Contributions to Statistics*; Jurečková, J., Ed.; Academic Press: New York, NY, USA, 1979; pp. 33–40.
66. Pfanzagl, J. *Asymptotic Expansions for General Statistical Models*; Springer: New York, NY, USA, 1985.
67. Beirlant, J.; Dudewicz, E.J.; Györfi, L.; van der Meulen, E.C. Nonparametric entropy estimation: An overview. *Int. J. Math. Stat. Sci.* **1997**, *6*, 17–39.
68. Büning, H.; Trenkler, G. *Nichtparametrische Statistische Methoden*; de Gruyter: Berlin, Germany, 1994.
69. Serfling, R.J. *Approximation Theorems in Mathematical Statistics*; John Wiley & Sons: New York, NY, USA, 1980.
70. Huber, P.J. *Robust Statistics*; John Wiley & Sons: New York, NY, USA, 1981.
71. Jurečková, J.; Sen, P.K. *Robust Statistical Procedures: Asymptotics and Interrelations*; John Wiley & Sons: New York, NY, USA, 1996.
72. Parr, W.C.; Schucany, W.R. Jackknifing L-statistics with smooth weight functions. *J. Am. Stat. Assoc.* **1982**, *77*, 629–638.
73. Klein, I.; Mangold, B. Cumulative paired ϕ -entropies—Estimation and Robustness. Unpublished work. 2016.
74. Klein, I.; Mangold, B. Cumulative paired ϕ -entropies and two sample linear rank tests for scale alternatives. Unpublished work, 2016.
75. Klein, I.; Mangold, B. ϕ -correlation and ϕ -regression. Unpublished work, 2016.
76. Pardo, L. *Statistical Inferences based on Divergence Measures*; Chapman & Hall: Boca Raton, FL, USA, 2006.
77. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Stat.* **1952**, *23*, 193–212.
78. Berk, R.H.; Jones, D.H. Goodness-of-fit statistics that dominate the Kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **1979**, *47*, 47–59.
79. Donoho, D.; Jin, J. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Stat.* **2004**, *32*, 962–994.
80. Park, S.; Rao, M.; Shin, D.W. On cumulative residual Kullback–Leibler information. *Stat. Probabil. Lett.* **2012**, *82*, 2025–2032.
81. Di Crescenzo, A.; Longobardi, M. Some properties and applications of cumulative Kullback–Leibler information. *Appl. Stoch. Models Bus. Ind.* **2015**, *31*, 875–891.
82. Liese, F.; Vajda, I. *Convex Statistical Distances*; Teubner-Verlag: Leipzig, Germany, 1987.
83. Csiszár, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1963**, *8*, 85–108. (In German)
84. Ali, S.M.; Silvey, S.D. A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc. Ser. B* **1966**, *28*, 131–142.
85. Cressie, N.; Read, T. Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B* **1984**, *46*, 440–464.
86. Jager, L.; Wellner, J.A. Goodness-of-fit tests via phi-divergences. *Ann. Stat.* **2007**, *35*, 2018–2053.
87. Parr, W.C.; Schucany, W.R. Minimum distance and robust estimation. *J. Am. Stat. Assoc.* **1980**, *75*, 616–624.
88. Nelsen, R.B. *An Introduction to Copulas*; Springer: New York, NY, USA, 1999.
89. Hall, P.; Wolff, R.C.; Yao, Q. Methods for estimating a conditional distribution function. *J. Am. Stat. Assoc.* **1999**, *94*, 154–163.
90. Schechtman, E.; Yitzhaki, S. A measure of association based on Gini’s mean difference. *Commun. Stat. Theory Methods* **1987**, *16*, 207–231.
91. Schechtman, E.; Yitzhaki, S. On the proper bounds of the Gini correlation. *Econ. Lett.* **1999**, *63*, 133–138.

92. Yitzhaki, S. Gini's mean difference: A superior measure of variability for non-normal distributions. *Metron* **2003**, *61*, 285–316.
93. Olkin, I.; Yitzhaki, S. Gini regression analysis. *Int. Stat. Rev.* **1992**, *60*, 185–196.
94. Hettmansperger, T.P. *Statistical Inference Based on Ranks*; John Wiley & Sons: New York, NY, USA, 1984.
95. Jaeckel, L.A. Estimating regression coefficients by minimizing the dispersion of residuals. *Ann. Math. Stat.* **1972**, *43*, 1449–1458.
96. Jurečková, J. Nonparametric estimate of regression coefficients. *Ann. Math. Stat.* **1971**, *42*, 1328–1338.
97. Kloke, J.D.; McKean, J.W. Rfit: Rank-based estimation for linear models. *R J.* **2012**, *4*, 57–64.
98. McKean, J.W.; Kloke, J.D. Efficient and adaptive rank-based fits for linear models with skew-normal errors. *J. Stat. Distrib. Appl.* **2014**, *1*, doi:0.1186/s40488-014-0018-0.
99. Hettmansperger, T.P.; McKean, J.W. *Robust Nonparametric Statistical Methods*; Chapman & Hall: New York, NY, USA, 2011.
100. Koul, H.L.; Sievers, G.L.; McKean, J. An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Stat.* **1987**, *14*, 131–141.
101. Ansari, A.R.; Bradley, R.A. Rank-sum tests for dispersion. *Ann. Math. Stat.* **1960**, *31*, 142–149.
102. Hájek, J.; Šidák, Z.; Sen, P.K. *Theory of Rank Tests*; Academic Press: San Diego, CA, USA, 1999.
103. Mood, A.M. On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Stat.* **1954**, *25*, 514–522.
104. Klotz, J. Nonparametric tests for scale. *Ann. Math. Stat.* **1961**, *33*, 498–512.
105. Basu, A.P.; Woodworth, G. A note on nonparametric tests for scale. *Ann. Math. Stat.* **1967**, *38*, 274–277.
106. Shiraishi, T.A. The asymptotic power of rank tests under scale-alternatives including contaminated distributions. *Ann. Math. Stat.* **1986**, *38*, 513–522.
107. Sukhatme, B.V. On certain two-sample nonparametric tests for variances. *Ann. Math. Stat.* **1957**, *28*, 188–194.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).