IWQW

Institut für Wirtschaftspolitik und Quantitative
Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 07/2015

Cumulative Paired ϕ-Entropy

Ingo Klein
University of Erlangen-Nürnberg

Benedikt Mangold
University of Erlangen-Nürnberg

ISSN 1867-6707

Cumulative Paired ϕ-Entropy

Ingo Klein* Benedikt Mangold†

July 27, 2015

A new kind of entropy will be introduced generalizing both the differential entropy and the cumulative (residual) entropy. The generalization is twofold. Firstly, we define the entropy for cumulative distribution functions (cdf) and survivor functions (sf) simultaneously instead of densities, cdf or sf alone. Secondly, we consider a general 'entropy generating function' ϕ like Burbea & Rao (1982) or Liese & Vajda (1987) in the context of ϕ-divergences. Combining the ideas of a ϕ-entropy and a cumulative entropy gives the new 'cumulative paired ϕ-entropy' (CPE_ϕ). With some modifications or simplifications this new entropy has already been discussed in at least four scientific disciplines. In the fuzzy set theory cumulative paired ϕ-entropies were defined for membership functions. A discrete version serves as a measure of dispersion for ordered categorial variables. More recently, uncertainty and reliability theory considered some variants as a measure of information. With only one exception the discussions seem to happen independently of each other. We consider CPE_ϕ only for continuous cdf and show that CPE_ϕ is rather a measure of dispersion than a measure of information. At first, 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 that the variance is fixed. We cannot only reproduce the central role of the logistic distribution in entropy maximization. We derive Tukey's λ distribution as the solution of an entropy maximization problem as well. Secondly, it will be shown explicitly that CPE_ϕ fulfills the axioms of a dispersion measure. The corresponding dispersion functional can easily be estimated by an L-estimator with all its known asymptotical properties. CPE_ϕ are the starting point for several related concepts like mutual ϕ-information, ϕ-correlation and ϕ-regression which generalize Gini correlation and Gini regression. We give a short introduction into all of these related concepts. Also linear rank tests for scale can be developed based on the new entropy. We show that almost all known tests are special cases and introduce some new tests. In the literature Shannon's differential entropy has been calculated for a lot of distributions. The formulas were presented explicitly. We have done the same for CPE_ϕ if the cdf is available in a closed form.

Keywords: ϕ-entropy, differential entropy, absolute mean deviation, cumulative residual entropy, cumulative entropy, measure of dispersion, measure of polarization, generalized maximum entropy principle, Tukey's λ distribution, power logistic distribution, ϕ-dependence, ϕ-regression, L-estimator, linear rank test

*ingo.klein@fau.de
†benedikt.mangold@fau.de

1 Introduction

Burbea & Rao (1982) introduced the ϕ-entropy

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

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

Recently, the discussion in Ebrahimi et al. (1999) has shown that entropies can be interpreted as a measure of dispersion. Shannon (1948) derived the 'entropy power fraction' and showed that there is a close relationship between Shannon entropy and variance. Oja (1981) illustrated that Shannon's differential entropy holds a ordering of scale and is therefore a proper measure of scale. In the discrete case, minimal Shannon entropy means maximal certainty about the random outcome of an experiment. A degenerate distribution minimizes the Shannon entropy. This is also true for the variance of a discrete quantitative random variable. For this 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 $[a,b]$. Differential entropy will be maximized by the uniform distribution over $[a,b]$ and the variance is maximal if both interval bounds a and b have the probability mass of 0.5 (cf. Popoviciu (1935)). A similar result holds for a discrete random variable with a finite number of realizations. Therefore, one can doubt that (1) is a true measure of dispersion.

We propose to define the ϕ-entropy for cumulative distribution function (cdf) F and survivor function (sf) $1-F$ instead of density functions f. With this modification we get

$$CPE_\phi(F) = -\int_{\mathbb{R}} \left(\phi(F(x) + \phi(1-F(x))\right) dx \tag{2}$$

with an absolutely continuous cdf F. CPE stands for 'cumulative paired entropy' and ϕ is the 'entropy generating function' defined on $[0,1]$ with $\phi(0) = \phi(1) = 0$. Throughout this paper we will almost always assume that ϕ is concave on $[0,1]$. In particular, we will show that (2) holds a prominent ordering of scale and attains its maximum if the domain is an interval $[a,b]$ and a, b occur with probability 1/2. This means that (2) behaves like a proper measure of dispersion. We want to generalize results from the literature focussing on the Shannon case with $\phi(u) = -u\ln u$, $u \in [0,1]$ (see Liu (2015)) and the cumulative residual entropy

$$CRE(F) = -\int_{\mathbb{R}+} (1-F(x))\ln(1-F(x))dx \tag{3}$$

(cf. Wang et al. (2003)) or

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

(cf. Di Crescenzo & Longobardi (2009 a,b)). In this literature the entropy still has the interpretation as a measure of information rather than dispersion without any clarification what kind of information has been considered.

As a first general aim this paper wants to shown that entropies can be rather interpreted as measures of dispersion than as measures of information. The second general aim is to demonstrate that 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.

Special aims of the paper are:

1. To show that the cdf based entropy (2) has roots in several distinct scientific areas.

2. To demonstrate the close relationship between (2) and the standard deviation.

3. To derive maximum entropy distributions under simple and more complicated restrictions, and to show that well-known and new distributions solve the maximum entropy principle.

4. To invert the question and to derive the entropy maximized by a given distribution under some restrictions.

5. To prove formally that (2) is a measure of dispersion.

6. To propose a L-estimator for (2) and to derive its asymptotically properties.

7. To use (2) to get new related concepts measuring the dependence of random variables (like mutual ϕ-information, ϕ-correlation and ϕ-regression).

8. To apply (2) to get new linear rank tests for the comparison of scale.

The paper is structured parallel to these aims. In the second section we give a short review of the literature concerned with (2) or related measures. The third section starts with a summary of the reasons why there is an advantage to define entropies for cdf and sf instead of densities. Some equivalent characterization of (2) will be given if the derivative of ϕ exists. In the fourth section we use the Cauchy-Schwarz inequalitty to derive an upper bound for (2). This upper bound gives sufficient conditions for the existence of CPE. More stringent conditions for the existence will be proven directly. In the fifth section the Cauchy-Schwarz inequality allows to derive maximum entropy distributions if the variance is fixed. For more complicated restrictions we derive the maximum entropy distributions by solving the Euler-Lagrange conditions. According to the generalized maximum entropy principle (cf. Kapur (1983)) we change the perspective and ask which entropy will be maximized if the variance and the population's distribution is fixed. The sixth section is of central importance because the properties of (2) as a measure of dispersion will be investigated in detail. It will be shown that it satisfies an often applied ordering of scale, is invariant with respect to translations and equivariant with respect to scale transformations. Additionally, we provide some results concerning the sum of independent random variables. In the seventh section a L estimator for CPE_ϕ will be proposed. Some basic properties of this estimator like the influence function, consistency and asymptotic normality are shown. Based on CPE_ϕ in the eight section we introduce several new statistical concepts generalizing divergence, mutual information, Gini correlation and Gini regression. Additionally, new linear rank tests for dispersion can base on CPE_ϕ. The known linear rank tests like Mood- or Ansari-Bradley test are special cases of this general approach. In this paper we leave off most of the technical details which will be presented in several accompanying papers. In the last section we compute (2) for some special 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 summing up to one (cf. Hartley (1928), Shannon (1948)). Until now it is rather unusual 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

$$CPE_S(F) = -\int_{\mathbb{R}} (F(x) \ln F(x) + (1 - F(x)) \ln(1 - F(x))dx. \quad (5)$$

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

1. Fuzzy set theory,

2. generalized maximum entropy principle,

3. theory of dispersion of ordered categorial variables,

4. uncertainty theory,

5. reliability theory.

Fuzzy set theory

To the best of our knowledge (5) has been introduced the first time by De Luca & Termini (1972). However, they did not consider the entropy for a cdf F. Instead, they were concerned with a so-called membership function μ_A quantifying the degree of which a certain element x of a set Ω belongs to a subset $A \subseteq \Omega$. Membership functions have been introduced by Zadeh (1968) within the framework of the 'fuzzy set theory'. It is important, that maximum uncertainty to belong to a set A should be attained if all elements of Ω will be mapped to the value $1/2$. This main property is one of the axioms for membership functions. In the aftermath of De Luca & Termini (1972) numerous modifications to the term entropy have been made and axiomatizations of the membership functions have been stated (see e. g. the overview in Pal & Bezdek (1994)).

Ultimately those modifications proceeded parallel to a long history of extensions and parametrizations of the term entropy for vectors of probability and densities starting with Renyi (1961) up to Esteban & Morales (1995) or Cichocki & Amari (2010), who provided a superstructure of those generalizations consisting of a very general form of the entropy, including the ϕ-entropy (1) as a special case. Burbea & Rao (1982) introduced the term ϕ-entropy. If both $\phi(x)$ and $\phi(1-x)$ appear in the entropy as in the Fermi-Dirac entropy (cf. Arndt (2004), p. 191) they used the term 'paired' ϕ-entropy.

Generalized maximum entropy principle

Regardless of the debate in the fuzzy set theory and the theory of measurement of dispersion Kapur (1983) considered a growth model with logistic growth rate. He showed that this growth model is yielded as solution of maximizing (5) under two simple constraints. This gives an example for the 'generalized maximum entropy principle' postulated by Kesavan & Kapur (1989). The simple maximum entropy principle introduced by Jaynes (1957a,b) derives a distribution which maximizes an entropy given some constraints. The generalization of Kesavan & Kapur (1989) consists of determining the ϕ-entropy which is maximized given a distribution and some constraints. Ultimately, they used formula (5) with a small modification. The cdf had to be replaced by a monotone increasing function having a logistic shape.

Theory of dispersion

Independently from the discussion of membership functions in the fuzzy set theory and the proposals to generalize the Shannon entropy, Leik (1966) discussed a measure of dispersion for ordered categorial variables with a finite number k of categories $x_1 < x_2 < \ldots < x_k$. His measure is based on the distance between the $k-1$-dimensional vectors of cumulated frequencies $(F_1, F_2, \ldots, F_{k-1})$ and $(1/2, 1/2, \ldots, 1/2)$. Both vectors coincide if only the both extreme categories x_1 and x_k appear with the same frequency. This represents the case of maximal dispersion. Consider

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

as a discrete version of (2). Setting $\phi(u) = \min\{u, 1-u\}$ we get the measure of Leik as a special case of (6) up to a change of sign. Vogel & Dobbener (1982) considered $\phi(u) = -u\ln(u)$ and the Shannon variant of (6) as a measure of dispersion for ordered categorial variables. Numerous modifications of Leik's measure of dispersion have been published. Kvålseth (1989), Berry & Mielke (1992a,b, 1994) and Blair & Lacy (1996, 2000) implicitly use $\phi(u) = 1/4 - (u - 1/2)^2$ or equivalently $\phi(u) = u(1-u)$. The discussion took place mainly in the journal 'Perceptual and Motor Skills'. For a recent overview on measuring dispersion, including ordered categorial variables see f.e. Gadrich et al. (2014). Instead of dispersion, some articles are concerned with related concepts like bipolarization and inequality for ordered categorial variables (cf. Allison & Forster (2004), Zheng (2008), Abul Naga & Yalcin (2008), Zheng (2011) and Apouey & Silver (2013)). A class of measures of dispersion for ordered categorial variables with a finite number of categories that is similar to (6) had been introduced independently of each other by Klein (1999) and Yager (2001). They were obviously not 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 (2001) recognized the close relationship between those measures and the general term of entropy in the 'fuzzy set' theory. He introduced the term 'dissonance' to characterize more precisely what dispersion for ordered categorial variables measures. In the language of information theory maximum dissonance means that there is still some information in this extreme case. But this information is extremely contradictory. Let us give an example from product evaluation. What can we learn for our decision to buy a product if 50 percent of the recommendations are extremely good as well as extremely bad? This is an important difference to the property of the Shannon entropy which is maximal if there is no information at all. This means that all categories occur with the same probability.

Uncertainty theory

Liu (2015) (first edition 2004) can be considered as the founder of the uncertainty theory. This theory is concerned with the formalization of data coming from expert opinions and not from the repetition of a random experiment. Liu modified the Kolmogoroff axioms of probability theory slightly to receive an uncertainty measure. Starting with this uncertainty measure he defined uncertainty variables, uncertainty distribution functions and moments of uncertainty variables. The experts shall answer questions concerning the uncertainty distribution function. Liu argued that 'an event is the most uncertain if its uncertain measure is 0.5 because the event and its complement may be regarded as "equally likely" (Liu (2015), p. xiv). Liu's maximum uncertainty principle states: 'For any event, if there are multiple reasonable values that an uncertain measure may take, the the value as close to 0.5 as possible is assigned to the event" (Liu (2015), p. xiv.). Similar to the fuzzy set theory, the distance between the uncertainty distribution and the value 0.5 will be measured by the Shannon-type entropy (5). Apparently for the first time in the third edition of 2010 he calculates (5) for several distributions (e. g. the logistic distribution) explicitly and derived upper bounds. He applied the maximum entropy principle to uncertainty distributions. The preferred constraint is to predetermine the values of the mean and the variance (Liu (2015), p. 83ff.). In this case the logistic distribution maximizes (5). In this sense the logistic distribution plays the same role as the Gaussian distribution in probability theory. The Gaussian distribution maximizes the differential entropy given values for mean and variance. Therefore in the uncertainty theory the logistic distribution will be called "normal distribution". Dai & Chen (2012) provided (5) as a function of the quantile function. Dai (2012) chooses $\phi(u) = u(1-u)$, $u \in [0,1]$ as entropy generating function and derives the maximum entropy distribution as a discrete uniform distribution concentrated on the endpoints of the compact domain $[a,b]$ if no further restrictions are assumed. Popiviciu (1935) got the same distribution as a result by maximizing the variance. Chen, Kar & Ralescu (2012) introduced cross entropies and divergence measures based on general functions ϕ.

Reliability theory

Entropies also play a prominent role in reliability theory. They were introduced for hazard-rates and residual lifetime distributions (cf. Ebrahimi (1996)). Rao et al. (2004) and Rao (2005) introduced the cumulative residual entropy (3), discussed its properties and derive the exponential and Weibull distribution by a maximum entropy 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. (2004) were referring to the massive criticism on the differential entropy by Schroeder (2004). Zografos & Nadaraja (2005) generalize the Shannon-type cumulative residual entropy to an entropy of the Rényi-type. Drissi et al. (2008) considered random variables with general support. They also give solutions for the maximization of (3) if more general restrictions are considered. Similar to Chen & Dai (2011) they identified the logistic distribution to be of maximum entropy given the mean, the variance and that the distribution has to be symmetric. Di Crescenzo & Longobardi (2009) analyzed (4) for cdf. They were engaged in the estimation of (4) and discussed the stochastic properties of the estimators. Sunoj & Sankaran (2012) plugged the quantile function into the Shannon-type entropy (4) and yielded expressions for the case that not the cdf but the quantile function has a closed form. In recent papers an empirical version of (3) is used as a goodness-of-fit test (cf. Zardasht et al. (2015)).

This brief overview shows that there are different disciplines accessing to an entropy based on distribution functions. The contributions of the fuzzy set theory, the uncertainty theory and the reliability theory have all in common that they consider continuous random variables exclusively. The discussion about entropy in reliability theory on the one hand and fuzzy set theory respectively uncertainty theory on the other hand took place independently without noticing 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 categorial variables the authors do not seem to be aware that they implicitly use a concept of entropy. Nevertheless the situation is somewhat different to the other areas since only discrete variables were discussed. Kiesl's dissertation (2003) provides a theory of measures of the form (6) with numerous applications. However, a intensive discussion of (2) is missing and shall be provided here.

3 Cumulative paired ϕ-entropy for continuous variables

3.1 Definition

We focus on an absolute continuous cdf F with density function f. The set of all those distribution functions is called $\mathcal{F}$. We call a function 'entropy generating function' if it is nonnegative 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 3.1. *The functional $CPE_\phi : \mathcal{F} \to \mathbb{R}_0^+$ with*

$$CPE_\phi(F) = \int_{\mathbb{R}} \left(\phi(F(x)) + \phi(1-F(x))\right) dx \tag{7}$$

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

In the next section we will discuss some sufficient criteria ensuring the existence of CPE_ϕ. Until then we assume that CPE_ϕ exists. If X is a random variable with cdf F we occasionally use the notation $CPE_\phi(X)$.

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

1. Cumulative paired α-entropy CPE_α: Following Havrda & Charvát (1967) 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

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

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

$$CPE_G(F) = 2\int_{\mathbb{R}} F(x)(1-F(x))dx. \tag{9}$$

   as a special case of CPE_α.

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

$$CPE_S(F) = -\int_{\mathbb{R}} \left(F(x)\ln F(x) + (1-F(x))\ln(1-F(x))\right) dx \tag{10}$$

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

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 Leik (1966) makes implicitly use of ϕ such that we call

$$CPE_L(F) = 2\int_{\mathbb{R}} \min\{F(x), 1-F(x)\}dx. \tag{11}$$

cumulative paired Leik entropy.

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

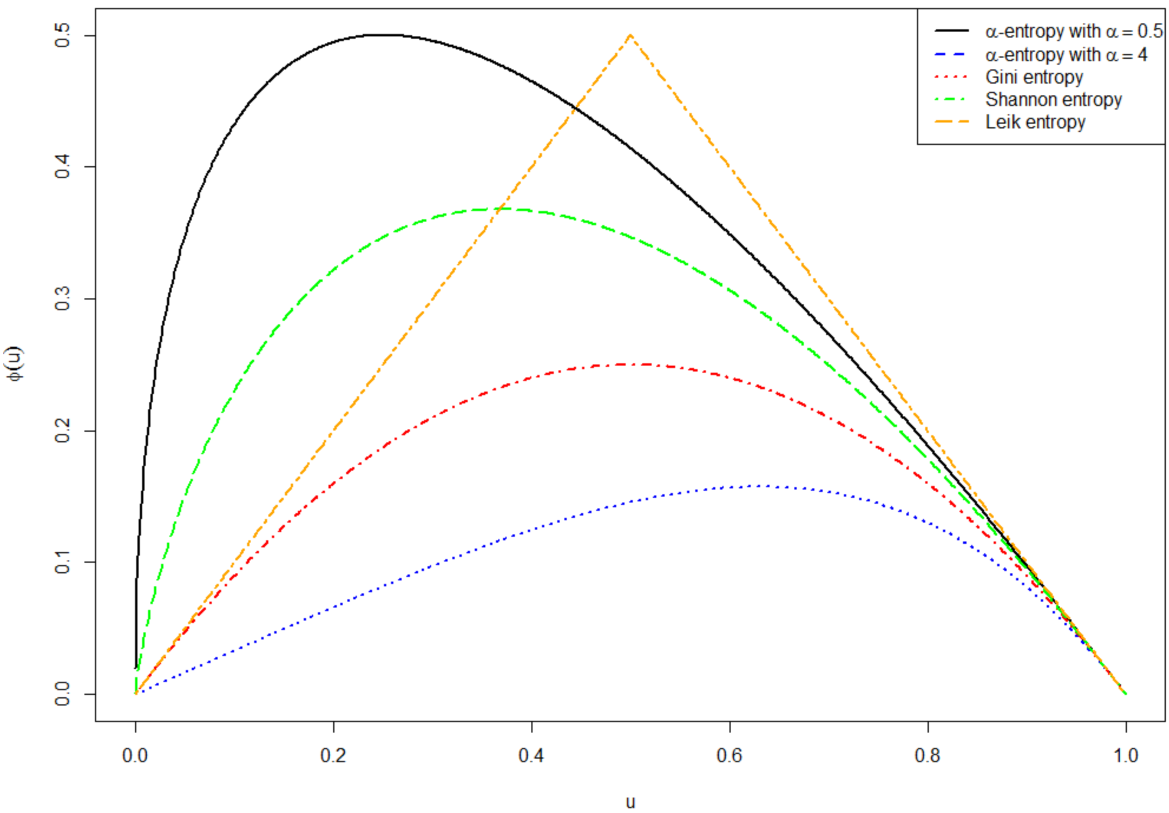

Figure 1: Some entropy generating functions ϕ

3.2 Advantages of entropies based on the cdf

Rao et al. (2004) and Rao (2005) list several reasons why an entropy should better be defined for distribution functions rather than density functions. Starting point is the well-known critique of Shannon's differential entropy $-\int f(x) \ln f(x) dx$ by several authors like Jumarie (1990), Kapur (1994), p. 58f. and Schroeder (2004).

Transferred to cumulative paired entropies the advantages of entropies based on distribution functions (see Rao et al. (2004)) are:

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

The problems of the differential entropy are occasionally discussed in the case of grouped data, where the usual Shannon entropy is calculated for the probabilities of each group. As the amount of groups is growing, the Shannon entropy does not converge to the respective differential entropy but is even divergent (cf. e. g. Cover & Thomas (1991), p. 239, Kapur (1994), p. 54). In the next section we will show that the discrete version of CPE_ϕ converges to CPE_ϕ as the number of groups is growing to infinity.

3.3 CPE_ϕ for grouped data

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

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

for $x \in \mathbb{R}$, respectively $F^*(x) = 0$ 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 $\tilde{x}_{i-1} < x \leq \tilde{x}_i$, $i = 1, ..., k$.

Lemma 3.1. *Let ϕ be a entropy generating function with antiderivative S_ϕ. The paired cumulative ϕ-entropy of the distribution function in (12) is given as follows:*

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

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

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

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

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

□

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

Theorem 3.1. *Let ϕ be a generating function with antiderivative S_ϕ and F 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:*

$$CPE_\phi(X^*) \to \int_a^b \left(\phi(F(x)) + \phi(1 - F(x)\right) dx \ \text{ for } k \to \infty.$$

Proof. Consider equidistant classes with $\Delta x_i = \Delta x = (b - a)/k$, $i = 1, ..., k$. In this case the equation (13) results in

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

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

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

An analogue argument holds for the second term of (14). □

In addition to this theoretic result we get some useful expressions for CPE_ϕ for grouped data and a special choice of ϕ as the following corollary shows:

Corollary 3.1. *Let ϕ be s. t.*

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

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

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

and for $\alpha \neq 1$

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

Proof. With the antiderivatives

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

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

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

for $i = 1, ..., k$, from which the results follow immediately. □

3.4 Alternative representations of CPE_ϕ

If $\phi(0) = \phi(1) = 0$ holds and ϕ is differentiable, one can provide several alternative representations of CPE_ϕ in addition to (7). Later this alternative representations will be useful to find conditions insuring the existence of CPE_ϕ and to find some simple estimators.

Proposition 3.1. *Let ϕ be a non-negative and differentiable function on the domain $[0,1]$ with derivative ϕ' and $\phi(0) = \phi(1) = 0$. Then 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:*

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

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

$$CPE_\phi(F) = \int_{\mathbb{R}} x(\phi'(1-F(x)) - \phi'(F(x)))f(x)dx. \tag{17}$$

Proof. Apply probability integral transformation $U = F(X)$ and partial integration. □

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

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

This property helps to represent CPE_ϕ as a covariance for which the Cauchy-Schwarz inequality gives an upper bound.

Corollary 3.2. *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 $X \sim F$:*

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

$$CPE_\phi(F) = Cov(\phi'(1-F(X)) - \phi'(F(X)), X). \tag{19}$$

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

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

□

Depending on the context we will switch between this 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 (18) resp. (19) provides an upper bound for CPE_ϕ if the variance $\sigma^2 = \mathbb{E}[(F^{-1}(u)-\mu)^2]$ exists and

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

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

Proposition 4.1. *Let ϕ be a non-negative and differentiable function on the domain $[0,1]$ with derivative ϕ' and $\phi(0)=\phi(1)=0$. If (20) holds, then for $X \sim F$ with $Var(X) < \infty$ and quantile function F^{-1} we have*

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

$$CPE_\phi(F) \leq \sqrt{E\left((\phi'(1-F(X))-\phi'(F(X)))^2\right) \sigma^2}. \tag{22}$$

Proof. The statement follows from

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

□

We consider the upper bound for the cumulative paired α-entropy.

Corollary 4.1. *Let X be a random variable having a finite variance. Then*

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

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

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

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

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

$\alpha > 1/2$ 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

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

□

In the framework of uncertainty theory the upper bound for the paired cumulative Shannon entropy has been derived by Chen & Dai (2011) (see also Liu (2015), p. 83). For $\alpha = 2$ we get the upper bound for the paired cumulative Gini entropy

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

This result has already been proved for non-negative uncertainty variables by Dai (2012). Finally, one yields the following upper bound for the paired cumulative Leik entropy.

Corollary 4.2. *Let X be a random variable with existing variance. Then*

$$CPE_L[X] \leq 2\sigma \tag{26}$$

Proof. Use

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

to get the result. □

4.2 Stronger conditions of existence for CPE_α

Up to now we only considered the sufficient condition that the variance exists. Following arguments in Rao et al. (2004) and Drissi et al. (2008) used for the special case of the cumulative residual and the residual Shannon entropy one can derive stricter sufficient conditions for the existence of CPE_α.

Theorem 4.1. *If $E(|X|^p) < \infty$ for $p > 1$ then $CPE_\alpha < \infty$ for $\alpha > 1/p$*

Proof. To prepare the proof first we notice that

$$u\frac{u^{\alpha-1}-1}{1-\alpha} \leq -u\ln u \leq u\frac{u^{\beta-1}-1}{1-\beta} \leq 1-u \tag{27}$$

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

The second fact we need for the proof is

$$\int_0^\infty (1-F(x))dx < \infty \text{ und } \int_{-\infty}^0 F(x)dx < \infty \tag{28}$$

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

$$E(X) = \int_0^\infty (1-F(x))dx + \int_{-\infty}^0 F(x)dx.$$

Thirdly it holds that

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

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

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

It has to be shown separately that these integrals converge.

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

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

and

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

For the third term we have to show 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 $1/p < \beta < 1$ and $\beta < \alpha$. With (27) it is for $-\infty < x \leq 0$

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

because $1-\beta > 0$.

With $F(x) = P(X \leq x) = P(-X \geq -x)$ it holds

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

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

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

Putting this results together we get that

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

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

It remains to show the convergence of the fourth term:

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

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

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

With $1-F(x) = P(X > x)$ we have

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

Now the Markov inequality gives

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

To summarize it is

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

for $\beta > 1/p$ und by $\int_1^{\infty} 1/y^q dy < \infty$ for $q > 1$. This completes the proof. □

If the mean or the variance exist theorem 4.1 results in concrete conditions for α in order to insure 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. $\mathbb{E}[|X|^p] < \infty$ for $p > 1$ is sufficient for the existence of CPE_S (i. e. $\alpha = 1$).

3. $\mathbb{E}[|X|^p] < \infty$ for $p = 1$ 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 there exists an affine linear relation between $F^{-1}(U)$ resp. X and $\phi'(1-U)-\phi'(U)$ resp. $\phi'(1-F(X))-\phi'(F(X))$ with probability equal to 1. Since the quantile function is monotone increasing, such a affine linear function can only exist if $\phi'(1-u)-\phi'(u)$ is monotone 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 distributions under the restriction that the mean and the variance are given one may consider only concave generating functions ϕ.

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

Theorem 5.1. *Let ϕ be a non-negative and differentiable function on the domain $[0,1]$ with derivative ϕ' and $\phi(0) = \phi(1) = 0$. F is the maximum CPE_ϕ distribution with prespecified mean μ and variance σ^2 of $X \sim F$ iff there exists a constant $b \in \mathbb{R}$ such that*

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

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

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

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

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

ϕ is concave on $[0,1]$ with

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

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

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

□

The quantile function of the Tukey λ distribution is given by

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

mean and variance are

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

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

Discussing the paired cumulative α-entropy one can prove the new result that the Tukey λ distribution is the maximum CPE_α distribution for prespecified mean and variance. Tukey's λ distribution assumes the role of the Student-t distribution if one changes from the differential entropy to CPE_α (c.f. Kapur (1988)).

Corollary 5.1. *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 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

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

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 maximum CPE_α distribution results in

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

for $\alpha>1/2$. F^{-1} can easily be identified as the quantile function of a Tukey λ distribution with

$\lambda = \alpha - 1$ and $\alpha > 1/2$. $\square$

For the Gini case $\alpha = 2$ one yields the quantile function of an 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 (2012).

The result that the logistic distribution is the maximum CPE_S distribution if the mean and the variance are given was derived by Chen & Dai (2011) in the framework of uncertainty theory and by Drissi et al. (2008), p. 4 in the framework of reliability theory. Both proved the result with the help of Euler-Lagrange equations. Only for completeness we provide an alternative proof via the upper bound of the Cauchy-Schwarz inequality.

Corollary 5.2. *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

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

one gets

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

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

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

$\square$

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

Corollary 5.3. *The cdf F maximizes CPE_L under the restriction of a known mean μ and known variance σ^2 iff for F holds*

$$F(x) = \begin{cases} 0 & \textit{für } x < \mu - \sigma \\ 1/2 & \textit{für } \mu - \sigma \le x < \mu + \sigma \\ 1 & \textit{für } x \ge \mu + \sigma \end{cases}$$

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

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

$\square$

This means that 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. (2008) discusses general moment restrictions of the form

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

where the existence of the moments is assumed. By use of the Euler-Lagrange equations they show, that

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

maximizes the residual cumulative entropy $-\int_{\mathbb{R}}(1 - F(x) \ln(1 - F(x)))\mathrm{d}x$ under the constraint (31) and that the solution needs to be symmetric with respect to μ. λ_i, $i =, ..., k$ are the Lagrange parameters determined by the moment restrictions provided a solution exists. Rao (2005) shows that for distributions with support $\mathbb{R}^+$ the maximum entropy distribution is given by

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

if again the restrictions (31) are demanded.

It is an obvious task to examine the shape of a distribution that maximizes the cumulative paired ϕ-entropy under the constraints (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

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

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

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

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

The following example shows, that the optimization problem (32) leads to useful distribution if the constraints will be chosen carefully in the case of a Shannon-type entropy.

Example 5.1. *The power logistic distribution is defined by the distribution function*

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

for $\gamma > 0$. The corresponding quantile function reads as

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

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

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

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 & Kapur (1989) introduced the generalized principle of maximum entropy problem which describes the interplay of entropy, constraints and distributions. A variant of this principle is to find an entropy which will be maximized by a given distribution and some moment restrictions.

This problem can be solved easily for CPE_ϕ if the 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 monotone 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 central importance.

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

Proposition 5.1. *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 given mean μ and variance σ^2 for CPE_ϕ with entropy generating function*

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

Proof. With

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

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

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

Substituting $\int_0^1 (2\Phi^{-1}(u))^2 du = 4$ one gets

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

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

An analogue result holds for the Student-t distribution with k degrees of freedom. The main difference to the Gaussian distribution is that the entropy generating function has no closed form but is given by numerical integration of the quantile function.

Corollary 5.4. *Let t_k and t_k^{-1} be the cdf and 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 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

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

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:

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

□

The following figure 3 gives an impression of the shape of the entropy generating function ϕ for several distributions generated by the generalized maximum entropy principle.

6 CPE_ϕ as a measure of scale

6.1 Basic properties of CPE_ϕ

The cumulative residual entropy (CRE) introduced by Rao et al. (2004), the generalized cumulative residual entropy (GRCE) of Drissi et al. (2008) and the cumulative entropy (CE) discussed by Creszenzo & Longobardi (2009) have always been interpreted as a measure of information.

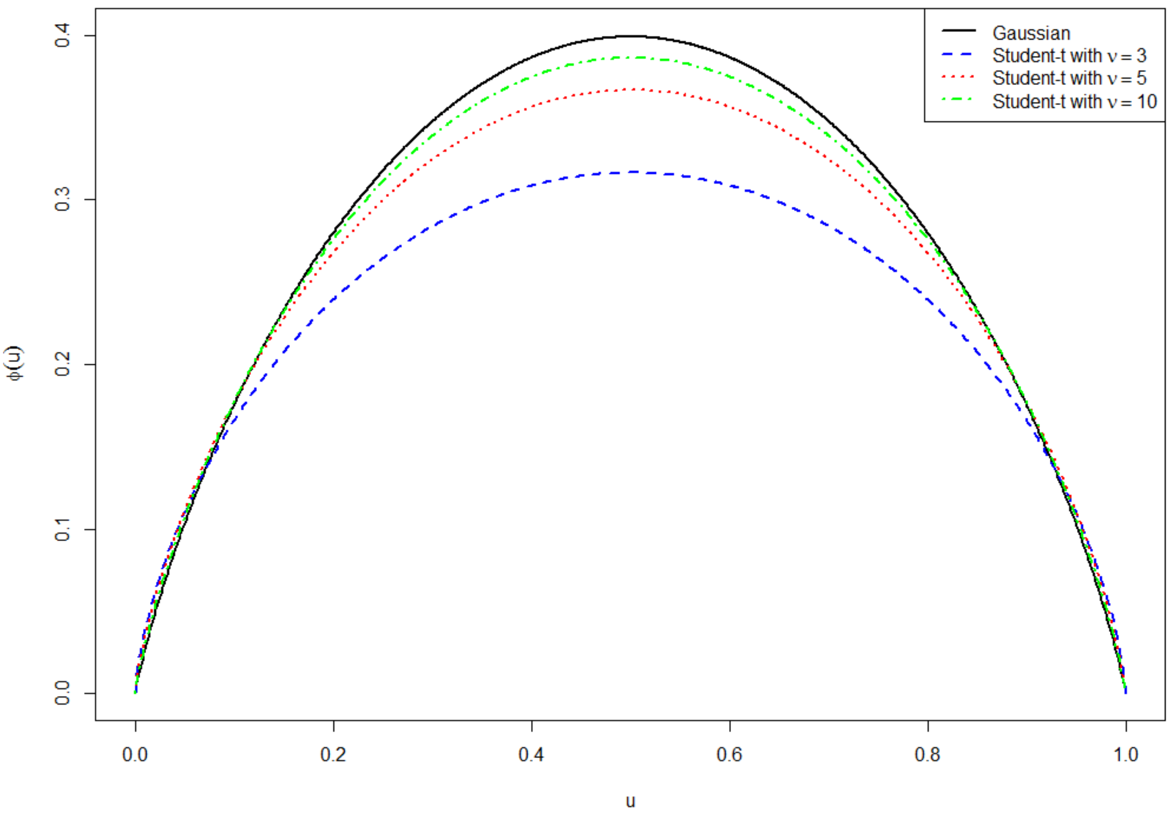

Figure 2: Several entropy generating functions ϕ derived from the generalized maximum entropy principle

However, all these approaches do not explain what kind of information will be considered. In contrast to this interpretation as a measure of information, Oja (1981) proved that the differential entropy holds a special ordering of scale and has some meaningful properties of a measure of scale. Ebrahimi et al. (1999) discussed the close relationship between differential entropy and variance. In the discrete case the Shannon entropy has an interpretation of a measure of diversity, which is a concept of dispersion when there is no ordering and no distance between the realizations of a random variable. In the last section (see lemma 4.1 and lemma 6.1) we will clarifying the important role the variance plays for the existence of CPE_ϕ.

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

Proposition 6.1. *Let $\phi : [0,1] \to \mathbb{R}$ with $\phi(u) > 0$ for $u \in (0,1)$ and $\phi(0) = \phi(1) = 0$. X shall be a random variable with support D. CPE_ϕ will be assumed to exist. Then the following properties hold:*

1. $CPE_\phi(X) \geq 0$

2. $CPE_\phi(X) = 0$ iff there exists x^* with $\mathbb{P}(X = x^*) = 1$.

3. $CPE_\phi(X)$ attains it 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 $x^* \in \mathbb{R}$ with $\mathbb{P}(X = x^*) = 1$ then $F_X(x) = 0$ and $1 - F_X(x) \in \{0, 1\}$ for all $x \in \mathbb{R}$. Due to $\phi(0) = \phi(1) = 0$ it follows $\phi(F_X(x)) = \phi(1 - F_X(x)) = 0$ for all $x \in \mathbb{R}$.

   Let now be $CPE_\phi(X) = 0$. Because of the non-negativity of the integrand $\phi(F_X(x)) + \phi(1 - F_X(x)) = 0$ must hold for $x \in \mathbb{R}$. Since $\phi(u) > 0$, $0 < u < 1$ it follows $F_X(x)$, $1 - F_X(x) \in \{0, 1\}$ 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 $[a, b]$. In this case the maximum is $2\phi(1/2)(b - a)$. Now, it suffices to construct a distribution with support $[a, b]$ that attains this maximum. Set

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

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

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

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

□

6.2 CPE_ϕ and Oja's axioms for measures of scale

Oja (1981), p. 159 defined a measure of scale as follows:

Definition 6.1. *Let $\mathcal{F}$ be a set of continuous distribution functions and $\preceq$ an appropriate ordering of scale on $\mathcal{F}$. $T : \mathcal{F} \to \mathbb{R}$ is called measure of scale, if*

1. $T(aX + b) = |a|T(X)$ for all $a, b \in \mathbb{R}$, $F \in \mathcal{F}$,

2. $T(X_1) \leq T(X_2)$, if $X_1 \sim F_1$, $X_2 \sim F_2$, $F_1, F_2 \in \mathcal{F}$ with $F_1 \preceq F_2$.

Oja (1981) discussed several orderings of scale. He showed in particular that Shannon entropy and variance hold a partial quantile based ordering of scale which has been discussed by Bickel & Lehmann (1976). Burger (1993) criticized, referencing to Behnen & Neuhaus (1989), that this ordering and the location-scale family of distributions focused by Oja (1981) were too restrictive. He discussed a more general nonparametric model of dispersion based on a more general ordering of scale (see Bickel & Lehmann (1979), Pfanzagl (1985)). Like Ebrahimi et al. (1999) we focus on the scale ordering proposed by of Bickel & Lehmann (1976).

Definition 6.2. *Let F_1, F_2 be continuous cdf with respective quantile function F_1^{-1} and F_2^{-1}. F_2 is called more spread out as F_1 ($F_1 \preceq_1 F_2$) if*

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

If F_1 resp. F_2 is absolute continuous with density functions f_1 resp. f_2 $\preceq_1$ can be characterized equivalently by the property that $F_2^{-1}\left(F_1^{-1}(x)\right) - x$ is monotone non-decreasing or

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

(cf. Oja, p. 160).

Now, we want to show that CPE_ϕ is a measure of scale in the sense of Oja (1981). This first lemma investigates the behavior of CPE_ϕ with respect to affine-linear transformations which refers to the first axiom of definition 6.1.

Lemma 6.1. *Let F be the cdf of the random variable X. Then*

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

Proof. Let $Y = aX + b$, then

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

Substitution of $x = (y - b)/a$ with $\mathrm{d}y = a\mathrm{d}x$ gives

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

At the same time is

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

such that

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

□

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

Lemma 6.2. *Let F_1, F_2 be continuous cdf of the random variables X_1 and X_2 with $F_1 \preceq_1 F_2$. Then the following holds:*

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

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,

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

If $F_1 \preceq_1 F_2$ and hence $f_1\left(F_1^{-1}(u)\right) \ge f_2\left(F_2^{-1}(u)\right)$ for $u \in [0,1]$, it follows that $CPE_\phi(F_1) - CPE_\phi(F_2) \le 0$. □

As a consequence of lemma 6.1 and lemma 6.2 CPE_ϕ is a measure of scale in the sense of Oja (1981). CPE_ϕ shares this properties with the variance, the differential entropy a many other statistical measures.

6.3 CPE_ϕ and transformations

Ebrahimi et al. (1999), p. 323, considered the cdf F_1 resp. F_2 on domain D_1 resp. D_2 and density functions f_1 resp. f_2 which are connected via $F_2(x) = F_1\left(g^{-1}(x)\right)$, $x \in D_1$, a differentiable transformation $g : D_1 \to D_2$, that is $F_2(y) = F_1(g(y))$ resp. $f_2(y) = f_1\left(g^{-1}(y)\right) \left|\mathrm{d}g^{-1}(y)/\mathrm{d}y\right|$ for $y \in D_1$. Ebrahimi et al. (1999) showed 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 $CPE_\phi(F_1)$:

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

Of special interest are transformations with $|g'(y)| \ge 1$, $y \in D_2$ since such a transformation does not diminish a measure of scale. In theorem 1 Ebrahimi et al. (1999) show that $F_1 \preceq_1 F_2$ holds if $|g'(y)| \ge 1$ for $y \in D_2$. Hence, every measure of scale cannot be diminished by this special transformation. Especially, this is true for the Shannon entropy H and for CPE_ϕ.

Ebrahimi et al. (1999) 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 what we would expect from a measure of scale. The standard deviation is changed by the factor $|a|$, the same is true for CPE_ϕ as we have shown in lemma 6.1.

6.4 CPE_ϕ for the sum 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 like the Rényi or the Havrda & Charvát entropy are only subadditive (c.f. Arndt (2004), p. 194).

The property of additivity or just subadditivity cannot be shown for cumulative paired ϕ-entropies. Instead, they possess a maximum property if ϕ is a concave function on $[0,1]$. This means for two independent variables X and Y that $CPE_\phi(X+Y)$ is bounded from below by the maximum of the two individual entropies $CPE_\phi(X)$ and $CPE_\phi(Y)$. This result has been shown by Rao et al. (2004) for the case of the cumulative residual Shannon entropy. The following theorem generalizes this result. The proof follows partially Rao et al. (2004), theorem 2.

Theorem 6.1. *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*

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

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

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

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

$$E_Y\left[\phi(F_X(t-Y))\right] \geq \phi\left(E_Y\left[F_X(t-Y)\right]\right) \tag{38}$$

and

$$E_Y\left[\phi(1-F_X(t-Y))\right] \geq \phi\left(E_Y\left[1-F_X(t-Y)\right]\right). \tag{39}$$

The existence of the expectation is assumed. To prove the theorem's statement we start with

$$CPE_\phi[X+Y] = \int_{-\infty}^{\infty} \left(\phi\left(E_Y\left[F_X(t-Y)\right]\right) + \phi\left(E_Y\left[1-F_X(t-Y)\right]\right)\right) dt.$$

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

$$\begin{aligned} CPE_\phi[X+Y] &\geq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\phi\left(F_X(t-y)\right) + \phi\left(1-F_X(t-y)\right)\right) dt f_Y(y) dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\phi\left(F_X(z)\right) + \phi\left(1-F_X(z)\right)\right) dz f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left(\phi\left(F_X(z)\right) + \phi\left(1-F_X(z)\right)\right) dz = CPE_\phi[X]. \end{aligned}$$

□

Liu (2015) considered a different definition of independence for uncertainty variables leading to the simpler additivity property

$$CPE_\phi(X+Y) = CPE_\phi(X) + CPE_\phi(Y) \tag{40}$$

for independent uncertainty variables X and Y.

7 Estimation of CPE_ϕ

Beirlant et al. (1997) gave an overview for estimators of the differential entropy. Ultimately, all proposals are based on the estimation of a density function f inheriting all typical problems of non-parametric estimaton of a density function. Among others, those problems are biasedness, the choice of a kernel and the optimal choice of the smoothing parameter (cf. Büning & Trenkler (1994), p. 215 ff.). However, CPE_ϕ is based on the cdf F for which several natural estimators with desirable stochastic properties derived from the theorem of Glivenko & Cantelli (cf. Serfling (1980), p. 61) exist. For a simple random sample sample $(X_1, ..., X_n)$ of identically and independently distributed random variables with common distribution function F Di Crescenzo & Longobardi (2009a,b) estimate F by the empirical distribution function $F_n(x) = \frac{1}{n}\mathrm{I}(X_i \leq x)$ for $x \in \mathbb{R}$ and They showed for the cumulative entropy $CE(F) = -\int_{\mathbb{R}} F(x) \ln F(x) \mathrm{d}x$ that the estimator $CE(F_n)$ is consistent for $CE(F)$ (cf. Di Crescenzo & Longobardi (2009a)). In particular, if F is the distribution function of a uniform distribution they provided the expected value of the estimator and showed that the estimator is asymptotically normal. If F is cdf of an exponential distribution they derived additionally the variance of the estimator.

In the following we generalize the estimation approach of Di Crescenzo & Longobardi by embedding it into the well-established theory of L-estimators (cf. e. g. Huber (1981), p. 55ff.). If ϕ is differentiable, CPE_ϕ can be reprsented as the covariance between the random variable X and $\phi'(1 - F(X)) - \phi'(F(X))$

$$CPE_\phi(F) = \mathbb{E}\left(X\left(\phi'(1 - F(X)) - \phi'(F(X))\right)\right). \tag{41}$$

An unbiased estimator for this covariance is

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

with

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

This results in an L-estimator $\sum_{i=1}^{n} J(i/(n+1))X_{n:i}$ with $J(u) = \phi'(1-u) - \phi'(u)$, $u \in (0, 1)$. Applying known results for the influence function of L estimators (cf. Huber (1981)) we get for the influence function of CPE_ϕ

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

In particular, the derivative is

$$\frac{dIF(x;CPE_\phi,F)}{dx} = \phi'(1-F(x))-\phi'(F(x))), \quad x\in\mathbb{R}. \quad (44)$$

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

Example 7.1. *Starting with the derivative*

$$\frac{dIF(x;CPE_S,F)}{dx} = \phi'(1-F(x))-\phi'(F(x)) = \ln\left(\frac{F(x)}{1-F(x)}\right) = x, \quad x\in\mathbb{R}$$

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 asymptotical and robustness behavior. The integration constant C has to be determined such that $\mathbb{E}\left[IF(x;CPE_S,F)\right]=0$:

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

Example 7.2. *Using the Gini entropy CPE_G and the logistic distribution function F we have*

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

Integration gives the influence function

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

With numerical integration we get $C=-1.2741$.

Example 7.3. *For $\phi(u)=u(u^{\alpha-1}-1)/(1-\alpha)$ the derivative of the influence function is given by*

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

Integration leads to the influence function

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

with

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

Under certain conditions (see Jurečková & Sen (1996), p. 143) concerning J, resp. ϕ and F L estimators are consistent and asymptotical normal. This means for the cumulative paired ϕ-entropy

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

with the asymptotical variance

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

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

Example 7.4. *For the cumulative paired Shannon entropy it holds*

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

since

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

Example 7.5. *In the Gini-case we get*

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

since by numerical integration

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

It is known that L-estimators have a remarkable bias for small sample sizes. Following Parr & Schucany (1982) the bias can be reduced via the Jackknife method. Obviously, the asymptotical distribution can be used to construct approximate confidence intervals or for a hypothesis tests in the one or two sample case. Huber (1981), p. 116ff. discussed asymptotically efficient L-estimators for a parameter of scale θ. In Klein & Mangold (2015a) it will be investigated 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

There are several statistical concepts closely related to cumulative paired ϕ-entropies. These concepts generalize some results known from the literature. We start with the cumulative paired ϕ-divergence which has been discussed the first time by Chen et al. (2012) under the name 'generalized cross entropy'. They concentrated on uncertainty variables and not on random variables as we will do. 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 the variance and allows to define directed measures of dependence for X and Y. The third concept again has to do with dependence. More precisely, we introduce a new family of correlation coefficients measuring the strength of a monotone relationship between X and Y. In this approach well-known coefficients like the Gini correlation can be imbedded. The fourth concept treats the problem of linear regression. CPE_ϕ can serve as general measure of dispersion which has to be minimized to estimate the regression coefficients. This approach will be identified as special case of rank based regression or R regression. The entropy generating function ϕ decides about the robustness properties of the rank based estimator. Asymptotics can be derived from the 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, from the theory of linear rank tests can be transferred 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 which are 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 (see f.e. Pardo (2006), p. 5). Therefore, in the following we consider $-\phi$ with ϕ concave.

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

Definition 8.1. *Let X and Y be two random variables with cdf F_X and F_Y. Then the cumulative paired ϕ-divergence of X and Y is given by*

$$CPD_\phi(X,Y) = -\int_{-\infty}^{\infty} \left(F_Y(x)\phi\left(\frac{F_X(x)}{F_Y(x)}\right) + (1-F_Y(x))\phi\left(\frac{1-F_X(x)}{1-F_Y(x)}\right)\right) dx. \tag{45}$$

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

Example 8.1. *1. Consider $\phi(u) = -u\ln u$, $u \in [0,\infty)$. Then we get the cumulative paired Shannon divergence*

$$CPD_S(X,Y) = \int_{-\infty}^{\infty} \left(F_X(x)\ln\left(\frac{F_X(x)}{F_Y(x)}\right) + (1-F_X(x))\ln\left(\frac{1-F_X(x)}{1-F_Y(x)}\right)\right) dx.$$

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

$$\begin{aligned} CPD_\alpha(X,Y) &= \frac{1}{\alpha - 1}\left(\int_{-\infty}^{\infty} \left(F_X(x)^\alpha F_Y(x)^{1-\alpha}\right.\right. \\ &\quad \left.\left. +(1-F_X(x))^\alpha (1-F_Y(x))^{1-\alpha} - 1\right) dx\right). \end{aligned}$$

3. *For* $\alpha = 2$ *we get the special case of a cumulative paired Gini divergence*

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

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

$$CPD_L(X,Y) = \int_{-\infty}^{\infty} \left(\left|F_X(x) - \frac{1}{2}F_Y(x)\right| + \left|F_X(x) - \frac{1}{2}(1+F_Y(x))\right| - \frac{1}{2}\right) dx.$$

CPD_S is equivalent to the Anderson-Darling functional (cf. Anderson & Darling (1952)), and has been used by Berk & Jones (1979) for a goodness-of-fit test if F_X represents the empirical distribution. Also CPD_S serves for a goodness-of-fit test (cf. Donoho & Jin (2004)).

Based on work from Csiszár (1963), Ali & Silvey (1966), Cressie & Read (1984) and Liese & Vajda (1987) Jager & Wellner (2007) discuss a general function ϕ_α

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

Up to multiplicative constant ϕ_α includes all the above mentioned examples. Additionally, the Hellinger distance is a special case for $\alpha = 1/2$ that leads to the cumulative paired Hellinger divergence:

$$CPD_H(X,Y) = 2\int_{-\infty}^{\infty} \left(\left(\sqrt{F_X(x)} - \sqrt{F_Y(x)}\right)^2 + \left(\sqrt{1-F_X(x)} - \sqrt{1-F_Y(x)}\right)^2\right) dx.$$

For a strictly concave function ϕ Chen et al. (2012) proved that $CPE_\phi(X,Y) \geq 0$ and $CPE_\phi(X,Y) = 0$ iff X and Y have identical distribution. In this sense the cumulative paired ϕ-divergence can serve as a kind of distance between distribution functions. As an application Chen et al. (2012) 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 fixed. If F_Y is an empirical distribution and F_X has an unknown vector of parameters θ CPD_ϕ can be minimized to get point estimator for θ (cf. Parr & Schucany (1980)). The large class of goodness-of-fit tests based on CPD_ϕ discussed by Jager & Wellner (2007) has already been mentioned.

8.2 Mutual cumulative ϕ-information

Let X and Y again be random variables with cdf 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

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

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 then the total variation $CPE_\phi(Y)$. More precisely holds:

1. $E_X\left(CPE_\phi(Y|X)\right) \leq CPE_\phi(Y)$.

2. $E_X\left(CPE_\phi(Y|X)\right) = CPE_\phi(Y)$ if X and Y are stochastically independent.

3. If ϕ is strictly concave and $E_X\left(CPE_\phi(Y|X)\right) = CPE_\phi(Y)$, X and Y are stochastically independent random variables.

We will 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 'cumulative paired mutual ϕ-information' $MCPI_\phi$ (cf. Rao et al. (2004) using the term 'cross entropy' and Drissi et al. (2008), p. 3). $MCPI_\phi$ is the analogue of the transinformation which is defined for Shannon's differential entropy (c.f. Cover & Thomas (1991), p. 20f.) In contrast to the transinformation $MCPI_\phi$ is not symmetric. This means that $MCPI_\phi(X,Y) = MCPI_\phi(Y,X)$ must not hold in general.

Cumulative paired mutual ϕ-information acts as a starting point for two directed measures of the strength of ϕ-dependence between X and Y. $DCPD$ means 'directed (measure) of cumulative paired ϕ-dependence. The first one is

$$DCPD^1_\phi(X \to Y) = \frac{MCPI_\phi(X,Y)}{CPE_\phi(Y)}$$

and the second one is

$$DCPD^2_\phi(X \to Y) = \frac{CPE_\phi(Y)^2 - E_X(CPE_\phi(Y|X)^2)}{CPE_\phi(Y)^2}.$$

Both expressions measure the relative decrease of variation of Y if X is known. The domain is $[0,1]$. The lower bound 0 will be taken if Y and X are independent. 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 distributions $F_{Y|X}(y|x)$ has to be degenerated. This means that for every $x \in D_X$ there is exactly one $y* \in D_Y$ with $P(Y = y^*|X = x) = 1$. In this sense there is a perfect association between X and Y. The next example illustrates the concepts and shows why it is useful to consider both types of measures of dependence.

Example 8.2. *Let (X, Y) follow a bivariate standard Gaussian distribution with $E(X) = E(Y) = 0$, $Var(X) = Var(Y) = 1$ and $Cov(X, Y) = \rho$, $-1 < \rho < 1$. Note that X and Y follow a univariate standard Gaussian whereas $X+Y$ follows a univariate Gaussian distribution with mean 0 and variance $2(1+\rho)$. From this one can conclude that*

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

Plugging this quantile functions into the definition 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 $\rho \to -1$ the cumulative paired ϕ-entropy behaves like the variance or the standard deviation. All measures approach 0 for $\rho \to -1$. This means 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.

More precisely, CPE_ϕ behaves more 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 or smaller than the sum of the individual standard deviations. This is also true for CPE_ϕ.

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

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

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

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

Like the variance of the $Y|x$ CPE_ϕ does not depend on x in the case of a bivariate Gaussian distribution. This implies that the internal variation is also $\sqrt{1-\rho^2}CPE_\phi(Y)$. For $\rho \to 1$ 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 this case X and Y are independent.

The two measures of directed cumulative ϕ-dependence are

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

and

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

This means that ρ completely determines the value for both measures of directed dependence. In the case that 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 a bivariate distribution function with uniform marginals for two random variables U and V with support $[0, 1]$. For the details concerning copulas see f.e. Nelsen (1999).

Example 8.3. *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. Nelsen (1999), p. 68). With

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

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

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

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

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

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

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

It is well-known for the FGM copula that only a small range of dependence can be covered (cf. Nelsen (1999), p. 129).

Hall et al. (1999) discussed several methods for estimating a conditional distribution. The results can used for an estimator of the mutual ϕ-information and for the two directed measures of dependence. This will be the task of further research.

8.3 ϕ-correlation

Schechtman & Yitzhaki (1987) introduced Gini correlations of two random variables X and Y with distribution functions F_X and F_Y as

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

The numerator equals to 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 identical distributed random variables X_1 and X_2.

Gini's mean difference coincides with the cumulative paired Gini-entropy $CPE_\phi(X)$ in the following sense:

$$Cov(X, F_X(X)) = 4CPE_G(X) = 4 \int_{-\infty}^{\infty} X(\phi'(1 - F_X(X)) - \phi'(F_X(X)))dx.$$

Gini correlations can be generalized to ϕ-correlations based on CPE_ϕ instead of the Gini mean difference. Let X, Y be two random variables, $CPE_\phi(X)$ and $CPE_\phi(Y)$ are the corresponding cumulative paired ϕ entropies. Then

$$\Gamma_\phi(X, Y) = \frac{E(X(\phi'(1 - F_Y(Y)) - \phi'(F_Y(Y)))}{CPE_\phi(X)} \tag{49}$$

and

$$\Gamma_\phi(Y, X) = \frac{E(Y(\phi'(1 - F_X(X)) - \phi'(F_X(X)))}{CPE_\phi(Y)} \tag{50}$$

are called ϕ-correlations of X and Y. Since $E(\phi'(1 - F_Y(Y)) - \phi'(F_Y(Y)) = 0$ the numerator is the covariance between X and $\phi'(1 - F_Y(Y)) - \phi'(F_Y(Y))$,

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

Example 8.4. *The setting $\phi(u) = u(1 - u)$, $u \in [0, 1]$ leads to Gini correlation because*

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

and $E(X(\phi'(1 - F_X(X)) - \phi'(F_X(X))) = 4Cov(X, F_X(X))$.

The second example considers the new Shannon correlation.

Example 8.5. *Set $\phi(u) = -u \ln u$, $u \in [0, 1]$ then we get the Shannon correlation*

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

If Y follows a logistic distribution with $F_Y(y) = 1/(1 + e^{-y})$, $y \in \mathbb{R}$ then $\ln(F_Y(y)/(1 - F_Y(y)) = y$. Inserting this result leads to

$$\Gamma_S(X, Y) = \frac{E(XY)}{CPE_S(X)}.$$

From (30) we know that $CPE_S(X) = \pi/\sqrt{3}$ if X is logistic distributed. In this special case we get

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

In the following example we introduce the α-correlation.

Example 8.6. *For $\phi(u) = u(u^\alpha - 1)/(1-\alpha)$, $u \in [0,1]$ we get the α-correlation*

$$\Gamma_\alpha(X,Y) = \frac{E(X\frac{\alpha}{1-\alpha}((F_Y(Y)^{\alpha-1} - (1-F_Y(Y)^{\alpha-1})))}{CPE_\alpha(X)}.$$

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

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

Schechtman & Yitzhaki (1987), (1999) and Yitzhaki (2003) proved that Gini correlations have many desirable properties. In the following we give an overview over all properties which can be transferred to ϕ-correlations. For proofs and further details we refer to Klein & Mangold (2015c).

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

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

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

Example 8.7. *In the Gini case it is $\phi''(u) + \phi''(1-u) = -4$. This leads to*

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

Example 8.8. *In the Shannon case it is $\phi''(u) + \phi''(1-u) = -1/(u(1-u))$ such that*

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

The following basic properties of ϕ-correlations can be easily checked with the arguments applied by Schechtman & Yitzhaki (1987):

1. $\Gamma_\phi(X,Y) \in [-1,1]$
2. $\Gamma_\phi(X,Y) = 1$ ($\Gamma_\phi(X,Y) = -1$) if there is strictly increasing (decreasing) transformation g such that $X = g(Y)$.
3. If g is monotone, 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 depend on ϕ if a bivariate Gaussian will be considered. The same is true for ϕ-correlations as will be demonstrated in the following example.

Example 8.9. *Let (X, Y) be a bivariate standard Gaussian random variable with Pearson correlation coefficient ρ. In this case, all ϕ-correlations coincide with ρ as the following consideration shows: With $E(X|y) = \rho y$ it is*

$$\begin{aligned} Cov(X, \phi'(1-F_Y(Y)) - \phi'(F_Y(Y)) &= E_Y E_{X|Y}(X|Y)(\phi'(1-F_Y(Y)) - \phi'(F_Y(Y))) \\ &= \rho E_Y(Y(\phi'(1-F_Y(Y)) - \phi'(F_Y(Y))) \\ &= \rho CPE_\phi(Y) = \rho CPE_\phi(X). \end{aligned}$$

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

Weighted sums of random variables appear f.e. in portfolio optimization. The diversification effect concerns negative correlations between the return of asset. With this effect the risk of a portfolio can be significant smaller that the sum of the individual risks. It is the question whether cumulative paired ϕ-entropies can serve as a risk measure as well. Therefore we have to investigate the diversification effect for CPE_ϕ.

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

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

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, $i, j = 1, 2, \ldots, k$. A similar result was given by Yitzhaki (2003) for the Gini correlation without proof. Let $Y = \sum_{i=1}^k a_i X_i$ and set $D_{iy} = \Gamma_\phi(X_i, Y) - \Gamma_\phi(Y, X_i)$, for $i = 1, 2, \ldots, k$. Then the following decomposition of the square of $CPE_\phi(Y)$ holds:

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

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 für $i, j = 1, 2, \ldots, k$.

Schechtman & Yitzhaki (1987) also introduced an estimator for the Gini correlation and derived its asymptotic distribution. For the proof it is helpful that numerator of the Gini correlation can be represented as an 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 Klein & Mangold (2015c).

8.4 ϕ-regression

Based on the Gini correlation, Olkin & Yitzhaki (1992) discovered parallels between the traditional OLS-approach in regressions analysis and the minimization of the covariance between the error term ε in a linear regression model

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

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 $Cov(\varepsilon, F_\varepsilon)$ stands in the center of this new approach for regression analysis. Olkin & Yitzhaki (1992) noticed that this approach was already known under the name 'rank based regression' or shortly 'R regression' in robust statistics. In robust regression analysis a more general optimization criteria $Cov(\varepsilon, \varphi(F_\varepsilon)$ has been considered where φ denotes a strictly increasing score function (cf. Hettmansperger (1984), p. 233). The choice $\varphi(u) = 1 - 2u$ leads to the Gini mean difference. This 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 the the Gini regression to a so-called ϕ-regression based on the criteria function

$$CPE_\phi(\varepsilon) = Cov(\varepsilon, \phi'(1 - F_\varepsilon(\varepsilon)) - \phi'(F_\varepsilon)) \tag{52}$$

which has to minimized to get α and β. Therefore, cumulative paired ϕ-entropies are special cases of the dispersion function Jaeckel (1972) and Jureckova (1971) proposed as optimization criteria for R regression. More precisely, R estimation proceeds in two steps. In the first step

$$d_\phi(\beta) = CPE_\phi(y - X\beta) \tag{53}$$

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

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

Kloke & McKean (2012) and McKean & Kloke (2014) gave an overview about recent developments in rank based regression. We will apply their main results to ϕ-regression. Hettmansperger & McKean (2011) showed that for the influence function of $\hat{\beta}_\phi$ holds

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

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

The scale parameter τ_ϕ is given by

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

From the influence function we can see easily that $\hat{\beta}_\phi$ is asymptotically normal in the following sense:

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

For $\phi'(1-u) - \phi'(u)$ bounded, Koul et al. (1987) proposed a consistent estimator $\hat{\tau}_\phi$ for the scale parameter τ_ϕ. This asymptotical 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 an measure of determination (see Kloke & McKean (2012)).

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 own ϕ functions into rank based regression (cf. Kloke & McKean (2012)). With the help of this option and the dataset 'telephone' with several outliers available in "Rfit") we compare the fit of the Shannon regression ($\alpha \to 1$), the Leik regression, the α-regression for several values of α with the OLS regression. On the left of figure 3 the original data, the OLS and the Shannon regression are displayed. On the right side outliers are excluded to get a more detailed impression about the differences between the ϕ-regressions.

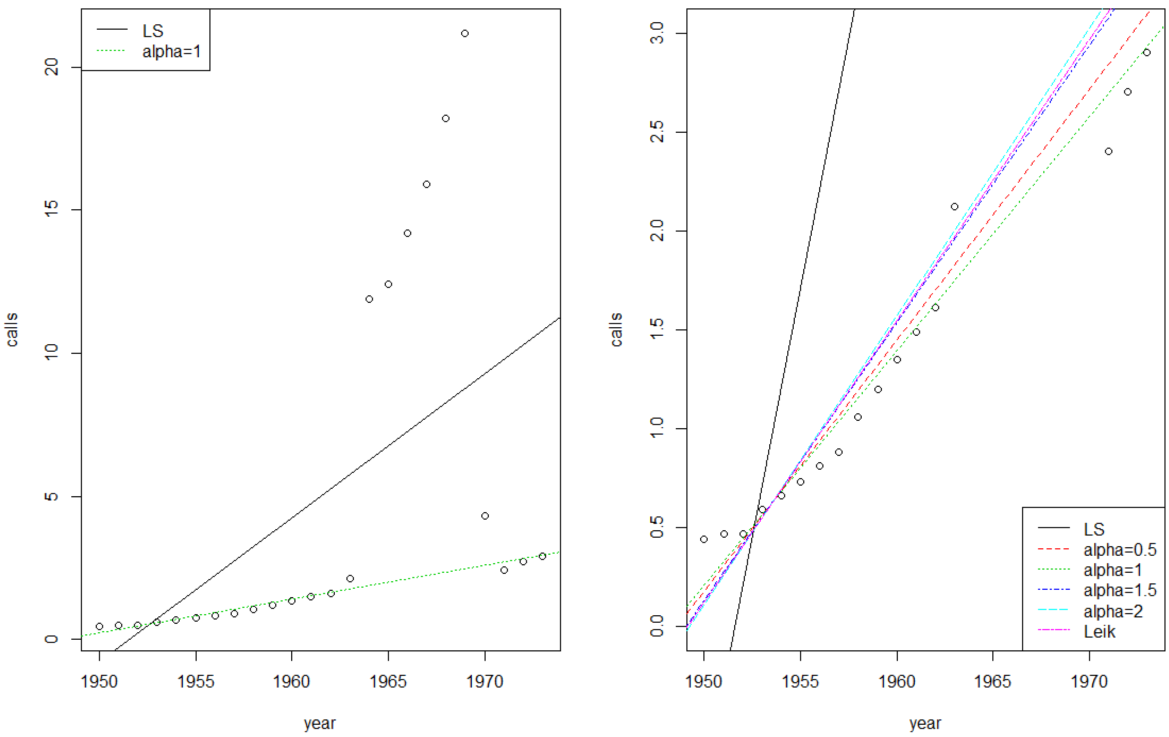

Figure 3: ϕ-regression fit for the number of calls in the 'telephone' data set.

In comparison with the very sensitive OLS regression all rank based regression techniques behave similar. McKean & Kloke (2014) gave an asymptotic efficient estimator for τ_ϕ if the error distribution is known. This procedure also determines the entropy generating function ϕ. If the error distribution is unknown, but some information with respect to skewness and leptokurtosis is available, a data-driven (= adaptive) procedure was proposed by McKean & Kloke (2014).

8.5 Two sample rank test on dispersion

Based on CPE_ϕ the linear rank statistics

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

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

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

Example 8.10. *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 & Bradley (1960) suggest the statistic

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

as a two sample test for alternatives of scale (cf. Hájek et al. (1999), p. 104). Apparently, we have $S_{AB} = 1/2(n+m+1)CPE_L(R)$.

Example 8.11. *Let $\phi(u) = 1/4 - (u - 1/2)^2$, $u \in [0, 1]$. Then 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 Mood (1954) up to a affine linear relation (cf. Büning & Trenkler (1994), p. 149f.). This test statistic is given by $S_M = \sum_{i=1}^{n}(R_i - (n+m+1)/2)^2$, so the resulting relation is given by

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

Ultimately, the scores of the Mood test are 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 8.12. *Let $\phi(u) = 1/2 - 1/2(sign(|u - 1/2| - 1/4) + 1)$, $u \in [0, 1]$ which is not concave on the interval [0,1]. Then we have*

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

which is identical to the quantile test statistic for alternatives of scale up to a affine linear relation (Hájek et al. (1999), 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 Hájek et al. (1999). The asymptotic distribution under the null hypothesis is needed to come to an approximate test decision given a significance level α, the asymptotic distribution under the alternative hypothesis for an approximate evaluation of the test power respectively the choice of the required sample size in order to reassure a given effect size.

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

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

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

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

(cf. Hájek et al. (1999), p. 194, theorem 1 and p. 195, lemma 1).

The asymptotic normality of the Ansari-Bradley test and the Mood test are well-known. Therefore, we provide a new linear rank test based on cumulative paired Shannon entropy CPE_S (so-called 'Shannon'-test) with $\phi(u) = -u\ln u$, $u \in [0,1]$ in the following example.

Example 8.13. *With $\phi(u) = -u\ln u$, $u \in [0,1]$ and $\bar{\phi} = 1/4$ we have*

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

and

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

Under the null hypothesis of identical scale the centered statistic 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

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

for $\sigma > 0$ and $\sigma \neq 1$. 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 $\min(n, m) \to \infty$ and $\ln \sigma I(f_0) mn/(n+m) \to b^2$ 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

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

This result follows immediately from Hájek et al. (1999), p. 267, theorem 1 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

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

This simplifies the variance of the asymptotic normal distribution.

The asymptotic normality of the test statistic of Ansari-Bradley and the Mood test under the alternative hypothesis have been examined intensively (cf. i. e. Mood (1954) and Klotz (1961)). Therefore, we concentrate on the new Shannon test.

Example 8.14. *Set $\phi(u) = -u \ln u$, $u \in [0, 1]$. 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 $I_1(f_0) = 1$. Then 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 has been evaluated by numerical integration. Then, under the alternative (58)

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

Finally, one can discuss the asymptotic efficiency of linear rank tests based on cumulative paired

ϕ-entropy. If f_0 is the true density and

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

then ρ_1^2 gives the wanted asymptotic efficiency (cf. Hájel et al. (1999), p. 317).

The asymptotic efficiency of the Ansari-Bradley test (respectively the asymptotic equivalent Siegel-Tukey test) and the Mood test have been investigated by Klotz (1961), Basu & Woodworth (1967) and Shiraishi (1986). The asymptotic relative efficiency (ARE) with respect to the traditional F-test for differences in scale for two Gaussian distributions has been discussed by Mood (1954). This asymptotic relative efficiency between Mood test and F-test for differences in scale have been derived by Sukhatme (1956). We concentrate again on the new Shannon-test.

Example 8.15. *The Klotz test is asymptotically efficient for the Gaussian distribution. With* $\int_0^1 (\Phi^{-1}(u)^2 - 1)^2 du = 2$,

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

gives the asymptotic efficiency of the new Shannon test.

Finally, we want to compare the asymptotic efficiency of the Shannon test for a distribution insuring asymptotic efficiency of the Ansari-Bradley test.

Example 8.16. *The Ansari-Bradley test statistic* S_{AB} *which is asymptotically efficient for the double log-logistic distribution with density function* f_0 *(cf. Hájek et al. (1999), p. 104). The Fisher information is given by*

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

Further we have

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

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

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

This two examples show that the Shannon test behaves relatively well even if the underlying distribution has moderate tails like the Gaussian or heavy tails like the double log-logistic distribution. An asymptotic efficient linear rank tests corresponds to a distribution and a scores generating function φ_1 from which we can derive a entropy generating function ϕ and a cumulative paired ϕ-entropy. This relationship will be further investigated in Klein & Mangold (2015b).

9 CPE_G, CPE_S and CPE_L for selected distribution functions

In the following, we want to derive closed form expressions for some cumulative paired ϕ-entropies. In some sense we want to mimic the procedure of Ebrahimi et al. (1999), p. 326. Table 1 of their paper contains many formulas of the differential entropy for the most popular statistical distributions. Several of those distributions will also be considered in the following. Since cumulative entropies depend on the distribution function or equivalently on the quantile function we concentrate 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 has only a multiplicative effect on CPE_ϕ and parameter of location has no effect at all. 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 could not establish a closed form of CPE_S and CPE_G. Therefore, we applied numerical integration in this case as well. In the following we use

- the incomplete Gamma function

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

- the incomplete beta function

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

- the Digamma function

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

Uniform distribution Let X have the standard uniform distribution. Then we have

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

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

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

Triangular distribution with parameter c Let X have a triangular distribution with density function

$$f(x) = \begin{cases} 2/cx & \text{für } 0 < x < c \\ 2/(1-c)(1-x) & \text{für } c \le x < 1 \end{cases}$$

Then we have

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

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

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

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

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

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

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

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

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

Pareto distribution Let X have the Pareto distribution with probability function $F_X(x) = 1 - x^c$ for $x > 1$, $c > 0$, then we have

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

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

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

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

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

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

As can be seen in figure 4, the heavy tails of the Student-t distribution result in a higher value for the CPE as in comparison to the Gaussian distribution.

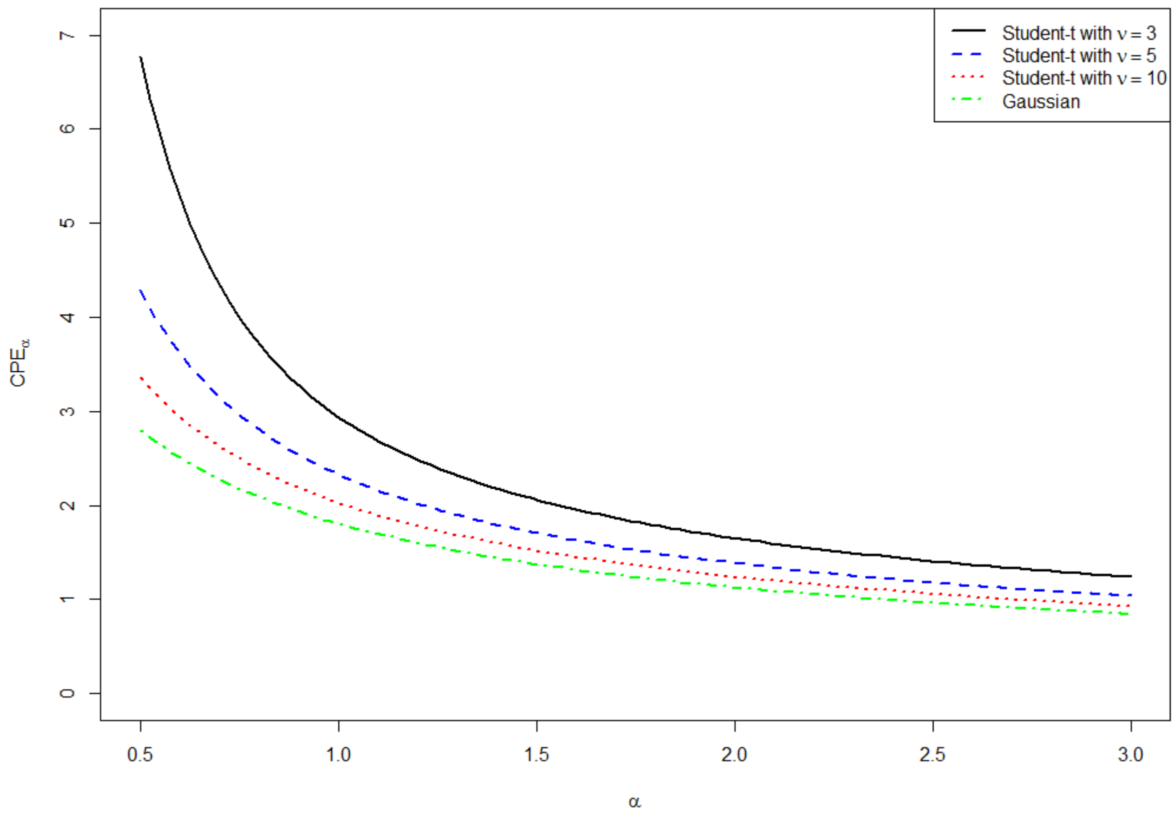

Figure 4: CPE_α, $\alpha \in [0.5, 3]$ for the standard Gaussian and the the Student-t distribution

10 Summary

A new kind of entropy has been introduced generalizing the Shannon's differential entropy. The main difference to the usual discussion of entropies is that it is defined for distribution functions and not for densities. It was shown that such a definition has a long tradition in several scientific disciplines like fuzzy set theory, reliability theory and more recently uncertainty theory. With only one exception in all disciplines the concepts have been discussed independently. Also the theory of dispersion measures for ordered categorial variables refers to distribution function based measures without noticing that at least implicitly a kind of entropy has been used. With the help of the Cauchy-Schwarz inequality we can show that there is a close relationship between the new kind of entropy named cumulative paired ϕ-entropy and the standard deviation. More precisely, the standard deviation gives an upper limit for the entropy. Additionally, the Cauchy-Schwarz inequality helps to derive maximum entropy distributions if constraints specify the values of mean and variance. For the cumulative paired Shannon entropy the logistic distribution adopts the central role the Gaussian distribution has if one wants to maximize the differential entropy. As a new result we have shown that Tukey's λ distribution is a maximum entropy distribution if the entropy generating function ϕ is used which is known from the Harvda & Charvát entropy. Some new distributions can be derived if more general constraints were considered. Changing perspective allows to determine the entropy that will be maximized by a certain distribution if f.e. mean and variance are known. In this context the Gaussian distribution gives a simple solution.

Since cumulative paired ϕ-entropy and the variance are related that closely we investigated whether the cumulative paired ϕ-entropy is a proper measure of scale. We show that it fulfills the axioms Oja has introduced for a measure of scale. Several further properties concerning the behavior under transformations or the sum of independent random variables have been proven. We give a first impression how to estimate the entropy. Based on cumulative paired ϕ-entropy we introduce some new concepts like ϕ-divergence, mutual ϕ-information and ϕ-correlation. Also ϕ-regression and linear rank tests for scale alternatives can be considered. For some popular distributions with cdf or quantile function in closed form and for some cumulative paired ϕ-entropies formulas have been derived.

References

1. Abul Naga, R.H. & Yalcin, T. (2008). Inequality measurement for ordered response health data. *Journal of Health Economics* **27**, 1614-1625.

2. Ali, S.M. & Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B* **28**, 131-142.

3. Allison, R.A. & Foster, J.E. (2004). Measuring health inequality using qualitative data. *Journal of Health Economics* **23**, 505-424.

4. Anderson, T.W. & Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193-212.

5. Ansari, A.R. & Bradley, R.A. (1960). Rank-sum tests for dispersion. *Annals of Mathematical Statistics* **31**, 1147-1189.

6. Apouey, B. & Silber, J. (2013). Inequality and bi-polarization in socioeconomic status and health: ordinal approaches. *Reseach and Economic Inequality* **21**, 77-109.

7. Arndt, C. (2004). *Information Measures.* Springer-Verlag, Heidelberg.

8. Basu, A.P. & Woodworth, G. (1967). A note on nonparametric tests for scale. *Annals of Mathematical Staistics* **38**, 274-277.

9. Beirlant, J., Dudewicz, E.J., Györfi, L. & van der Meulen, E.C. (1997). Nonparametric entropy estimation: an overview. *International Journal mof Mathematical and Statistical Science* **6**, 17-39.

10. Behnen, K. & Neuhaus, G. (1989). *Rank Tests with Estimated Scores and their Applications.* Teubner-Verlag, Stuttgart.

11. Berk, R.H. & Jones, D:H. (1979). Goodness-of-fit statistics that dominate the Kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **47**, 47-59.

12. Berry, K.J. & Mielke, P.W. (1992a). Assessment of variation in ordinal data. *Perceptual and Motor Skills* **74**, 63-66.

13. Berry, K.J. & Mielke, P.W. (1992b). Indices of ordinal variation. *Perceptual and Motor Skills* **74**, 576-578.

14. Berry, K.J. & Mielke, P.W. (1994). A test of significance for the index of ordinal variation. *Perceptual and Motor Skills* **79**, 1291-1295.

15. Bickel, P.J. & Lehmann, E.L. (1976). Descriptive statistics for nonparametric models: III. Dispersion. *Annals of Statistics* **5**, 1139-1158.

16. Bickel, P.J. & Lehmann, E.L. (1979). Descriptive statistics for nonparametric models: IV. Spread. In: Jurečková, J. (ed.). *Contributions to Statistics.* Academic Press, New York, 33-40.

17. Blair, J. & Lacy, M.G. (1996). Measures of variation for ordinal data. *Perceptual Motor Skills* **82**, 411-418.

18. Blair, J. & Lacy, M.G. (2000). Statistics of ordinal variation. *Sociological Methods Research* **28**, 251-280.

19. Büning, H. & Trenkler, G. (1994). *Nichtparametrische statistische Methoden.* de Gruyter, Berlin.

20. Burbea, J. & Rao, C.R. (1982). On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory* **28**, 489-495.

21. Burger, H.U. (1993). Dispersion orderings with applications to nonparametric tests. *Statistics & Probability Letters* **16**, 1-9.

22. Chen, X. & Dai, W. (2011). Maximum entropy principle for uncertain variables. *International Journal of Fuzzy Systems* **13**, 232-236.

23. Chen, X., Kar, S. & Ralescu, D.A. (2012). Cross-entropy measure of uncertain variables. *Information Sciences* **201**, 53-60.

24. Cichocki, A. & Amari., S. (2010). Families of alpha- beta- and gamma-divergences: flexible and robust measures of similarities. *Entropy* **12**, 1532-1568.

25. Cover, Th. M. & Thomas, J.A. (1991). *Elements of Information Theory.* Jon WiLey & Sons, New York.

26. Cressie, N. & Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B* **46**, 440.464.

27. Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8**, 85-108.

28. Dai, W. (2012). Maximum entropy principle for quadratic entropy of uncertain variables.

29. Dai, W. & Chen, X. (2012). Entropy of function of uncertain variables. *Mathematical and*

Computer Modelling **55**, 754-760.

30. De Luca, A. & Termini, S. (1972). A definition of a nonprobabilistic entropy in the setting of fuzzy set theory. *Information and Control* **20**, 301-312.

31. Di Crescenzo, A. & Longobardi, M. (2009a). On cumulative entropies and lifetime estimation. In: Mira. J.M. et al. (eds) *IWINAC 2009, Part I, LNCS 5601.* Springer, Berlin, 132-141.

32. Di Crescenzo, A. & Longobardi, M. (2009b). On cumulative entropies. *Journal of Statistical Planning and Inference* **139**, 4072-4087.

33. Donoho, D. & Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics* bf 32, 962-994.

34. Drissi, N., Chonavel, Th. & Boucher, J.M. (2008). Generalzed cumulaitve residual entropy distributions with unrestricted supports. *Research Letters in Signal Processing* **11**, 1-5.

35. Ebrahimi, B.R. (1996). How to measure uncertainty in the residual lifetime distribution. *Sankhya, Series A* **58**, 48-56.

36. Ebrahimi, N., Massoumi, E. & Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics* **90**, 317-336.

37. Esteban, M.D. & Morales, D. (1995). A suumary on entropy statistics. *Kybernetika* **31**, 337-346.

38. Gadrich, T., Bashkansky, E. & Zitikas, R. (2015). Assessing variation: a unifying approach for all scales of measurement. *Quality & Quantity* **49**, 1145-1167.

39. Gini, C. (1912). *Variabilità e mutabilità: contributo alla distribuzioni e delle relazioni statistiche.* Bologna.

40. Hájek, J., Šidek, Z. & Sen, P.K. (1999). *Theory of Rank Tests.* Academic Press, San Diego.

41. Hall, P., Wolff, R.C. & Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94**, 154-163.

42. Hartley, R.V.L. (1928). Transmission of information. *Bell Systems Technical Journal* **7**, 535-563.

43. Havrda, J. & Charvát, F. (1967). Quantification method of classification processes. Concept of structural α-entropy. *Kybernetika* **3**, 30-35.

44. Hettmansperger, Th.P. (1984). *Statistical Inference based on Ranks.* John Wiley & Sons, New York

45. Hettmansperger, Th.P. & McKean, J.W. (2011). *Robust Nonparametric Statistical Methods.* Chapman & Hall, New York.

46. Huber, P.J. (1981). *Robust Statistics.* Wiley & Sons, New York.

47. Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *Annals of Mathematical Statistics* **43**, 1449-1458.

48. Jager, L. & Wellner, J.A. (2007). Goodness-of-fis tests via phi-divergences. *The Annals of Statistics* **35**, 2018-2053.

49. Jaynes, E.T. (1957a). Information theory and statistical mechanics. *Physical Review* **106**, 620-630.

50. Jaynes, E.T. (1957b). Information theory and statistical mechanics. *Physical Review* **108**, 171-190.

51. Jumarie, G. (1990). *Relative Information: Theories and Applications* Springer-Verlag, Berlin.

52. Jurečková, J. (1971). Nonparametric estimate of regresion coefficients. *Annals of Mathematical Statistics* **42**, 1328-1338.

53. Jurečková, J. & Sen, P.K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* John Wiley & Sons, New York.

54. Jurečková, J. & Picek, J. (2006). *Robust Statistical Methods with R.* Chapman & Hall/CRC, Boca Raton.

55. Kapur, J.N. (1983). Derivation of logistic law of population growth from maximum entropy principle. *National Academy Science Letters* **6**, 429-433.

56. Kapur, J.N. (1988). Generalized Cauchy and Students distributions as maximum entropy distributions. *Proceedings of the National Academy of Sciences India* **58**, 235-246.

57. Kapur, J.N. (1994). *Measures of Information and their Applications.* New Age International Publishers, New Delhi.

58. Kesavan, H.K. & Kapur, J.N. (1989). The generalized maximum entropy principle. *IEEE Transactions on Systems, Man, and Cybernetics* **19**, 1042-1052.

59. Kiesl, H. (2003). *Ordinale Streuungsmaße* JOSEF-EUL-Verlag, Köln.

60. Klein, I. (1999). Rangordnungsstatistiken als Verteilungsmaßzahlen für ordinalskalierte Merkmale: I. Streuungsmessung. *Diskussionspapier 27/1999.* Lehrstuhl für Statistik und Ökonometrie, Universität Erlangen-Nürnberg.

61. Klein, I. & Mangold, B. (2015a). Cumulative paired ϕ-entropies - Estimation and Robustness. To appear.

62. Klein, I. & Mangold, B. (2015b). Cumulative paired ϕ-entropies - New rank tests for dispersion. To appear.

63. Klein, I. & Mangold, B. (2015c). ϕ-correlation and ϕ-regression. To appear.

64. Kloke, J.D. & McKean, J.W. (2012) Rfit: Rank-based estimation for linear models. *The R Journal* **4**, 57-64.

65. Klotz, J. (1961). Nonparametric tests for scale. *Annals of Mathematical Statistics* **33**, 498-512.

66. Koul, H.L., Sievers, G.L. & McKean, J.W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scnadinavian Journal of Statistics* **14**, 131-141.

67. Kvålseth, T.O. (1989). Nominal versus ordinal variation. *Perceptual and Motor Skills* **69**, 234.

68. Leik, R.K. (1966). A measure of ordinal consensus. *Pacific Sociological Review* **9**, 85-90.

69. Lerman, R. & Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. *Economic Letters* **15**, 363-368.

70. Liese, F. & Vajda I. (1987). *Convex Statistical Distances.* Teubner, Leipzig.

71. Liu, B. (2015). *Uncertainty Theory, 5th ed..* http://orsc.edu.cn/liu/ut.pdf.

72. McKean, J.W. & Kloke, J.D. (2014). Efficient and adaptive rank-based fits for linear models with skew-normal errors. *Journal of Statistical Distributions and Applications*, 1-18.

73. Mood, A.M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Annals of Mathematical Statistics* **25**, 514-522.

74. Nelsen, R.B. (1999). *An Introduction to Copulas.* Springer, New York.

75. Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics* **8**, 154-168.

76. Olkin, I. & Yitzhaki, S. (1992). Gini regression analysis. *International Statistical Review* **60**, 185-196.

77. Pal, N.R. & Bezdek, J.C. (1994). Measuring fuzzy uncertainty. *IEEE Transactions on Fuzzy Systems* **2**, 107-118.

78. Pardo, L. (2006). *Statistical Inferences based on Divergence Measures.* Chapman & Hall, Boca Raton, Fl.

79. Parr, W.C. & Schucany, W.R. (1980). Minimum distance and robust estimation. *Journal of the American Statistical Association* **75**, 616-624.

80. Parr, W.C. & Schucany, W.R. (1982). Jackknifing l-statistics with smooth weight functions. *Journal of the American Statistical Association* **77**, 629-638.

81. Pfanzagl, J. (1985). *Asymptotic Expansions for General Statistical Models.* Springer-Verlag, New York.

82. Popoviciu, T. (1935). Sur les équations algébraique ayant toutes leurs racines réelles. *Mathemtica (Cluj)* **9**, 129-145.

83. Rao, M., Chen, Y. Vemuri, B.C. & Wang, F. (2004). Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory* **50**, 1220-1228.

84. Rao, M. (2005). More on a new concept of entropy and information. *Journal of Theoretical Probability* **18**, 967-981

85. Rényi, A. (1961). On measures of entropy and information. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 547-561.

86. Schechtman, E. & Yitzhaki, S. (1987). A measure of association based on Gini's mean difference. *Communications in Statistics - Theory and Methods* **16**, 207-231.

87. Schechtman, E. & Yitzhaki, S. (1999). On the proper bounds of the Gini correlation. *Economics Letters* **63**, 133-138.

88. Schechtman, E. & Yitzhaki, S. (2003). A family of correlation coefficients based on extended Gini. *Journal of Economic Inequality* **1**, 129-146.

89. Schroeder, M.J. (2004). An alternative to entropy in the measurement of information. *Entropy* **6**, 388-412.

90. Serfling, R.J. (1980). *Approximation Theorems in Mathematical Statistics.* Wiley & Sons, New York,

91. Shannon, C.E. (1948). A mathematical theory oc communication. *Bell System Technical Journal* **27**, 379-423.

92. Shiraishi, T.-A. (1986). The asymptotic power of rank tests under scale-alternatives including contaminated distributions. *Annals of Mathematical Statistics* **38**, 513-522.

93. Sukhatme, B.V. (1957). On certain two-sample nonparametricn tests for variances. *Annals of Mathematical Statistics* **28**, 188-194.

94. Sunoj, S.M. & Sankaran, P.G. (2012). Quantile based entropy function. *Statistics and Probability Letters* **82**, 1049-1053.

95. Vogel, H. & Dobbener, R. (1982). Ein Streuungsmaß für komparative Merkmale. *Jahrbücher für Nationalökonomie und Statistik* **197**, 145-157.

96. Wang, F., Vemuri, B.C., Rao, M. & Chen, Y. (2003). A new & robust information theoretic measure and its application to image alignment. *Proceedings of the 18th International Conference on Information Processing in Medical Imaging (IPMI 03). Lecture Notes in Computer Science (vol. 2732).* Springer-Verlag, Berlin, 388-400.

97. Yager, R.R. (2001). Dissonance - a measure of variability for ordinal random variables. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **9**, 39-53.

98. Yitzhaki, S. (2003). Gini's mean difference: a superior measure of variability for non-normal distributions. *Metron*, 285-316.

99. Yitzhaki, S. & Schechtman, E. (2005). The properties of the extended Gini measures of variability and inequality. *Metron* **LXIII**, 401-433.

100. Zadeh, L. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications* **23**, 421-427.

101. Zardasht, V., Parsi, S. & Mousazadeh, M. (2015). On empirical cumulative residual entropy and a goodness-of-fit test for exponentiality. *Statistical Papers* DOI: 10.1007/s00362-014-0603-9.

102. Zheng, B. (2008). Measuring inequality with ordinal data: a note. *Research in Economic Inequality* **16**, 177-188.

103. Zheng, B. (2011). A new approach to measure socioeconomic inequality in health. *Journal of Economic Inequality* **9**, 555-577.

104. Zografos, K. & Nadarajah, S. (2005). Survival exponential entropies. *IEEE Transactions on Information Theory* **51**, 1239-1246.

Diskussionspapiere 2015
Discussion Papers 2015

01/2015 **Seebauer, Michael**: Does Direct Democracy Foster Efficient Policies? An Experimental Investigation of Costly Initiatives

02/2015 **Bünnings, Christian, Schmitz, Hendrik, Tauchmann, Harald and Ziebarth, Nicolas R.**: How Health Plan Enrollees Value Prices Relative to Supplemental Benefits and Service Quality

03/2015 **Schnabel, Claus:** United, yet apart? A note on persistent labour market differences between western and eastern Germany

04/2015 **Dieckmann, Anja, Fischbacher, Urs, Grimm, Veronika, Unfried, Matthias, Utikal, Verena and Valmasoni, Lorenzo:** Trust and Beliefs among Europeans: Cross-Country Evidence on Perceptions and Behavior

05/2015 **Grimm, Veronika, Utikal, Verena and Valmasoni, Lorenzo**: In-group favoritism and discrimination among multiple out-groups

06/2015 **Reichert, Arndt R., Tauchmann, Harald and Wübker, Ansgar**: Weight Loss and Sexual Activity in Adult Obese Individuals: Establishing a Causal Link

Diskussionspapiere 2014
Discussion Papers 2014

01/2014 **Rybizki, Lydia**: Learning cost sensitive binary classification rules accounting for uncertain and unequal misclassification costs

02/2014 **Abbiati, Lorenzo, Antinyan, Armenak and Corazzini, Lucca:** Are Taxes Beautiful? A Survey Experiment on Information, Tax Choice and Perceived Adequacy of the Tax Burden

03/2014 **Feicht, Robert, Grimm, Veronika and Seebauer, Michael**: An Experimental Study of Corporate Social Responsibility through Charitable Giving in Bertrand Markets

04/2014 **Grimm, Veronika, Martin, Alexander, Weibelzahl, Martin and Zoettl, Gregor**: Transmission and Generation Investment in Electricity Markets: The Effects of Market Splitting and Network Fee Regimes

05/2014 **Cygan-Rehm, Kamila and Riphahn, Regina**: Teenage Pregnancies and Births in Germany: Patterns and Developments

06/2014 **Martin, Alexander and Weibelzahl, Martin:** Where and when to Pray? - Optimal Mass Planning and Efficient Resource Allocation in the Church

07/2014 **Abraham, Martin, Lorek, Kerstin, Richter, Friedemann and Wrede, Matthias**: Strictness of Tax Compliance Norms: A Factorial Survey on the Acceptance of Inheritance Tax Evasion in Germany

08/2014 **Hirsch, Boris, Oberfichtner, Michael and Schnabel Claus**: The levelling effect of product market competition on gender wage discrimination

09/1014 **Mangold, Benedikt:** Plausible Prior Estimation

10/2014 **Gehrke, Britta:** Fiscal Rules and Unemployment

11/2014 **Gehrke, Britta** and **Yao, Fang**: Phillips Curve Shocks and Real Exchange Rate Fluctuations: SVAR Evidence

12/2014 **Mäder, Miriam, Müller, Steffen, Riphahn, Regina and Schwientek, Caroline**: Intergenerational transmission of unemployment - evidence for German sons

13/2014 **Casal, Sandro, Ploner, Matteo and Sproten, Alec N.**: Fostering the Best Execution Regime. An Experiment about Pecuniary Sanctions and Accountability in Fiduciary Money Management

14/2014 **Lochner, Benjamin**: Employment protection in dual labor markets - Any amplification of macroeconomic shocks?

15/2014 **Herrmann, Klaus, Teis, Stefan and Yu, Weijun**: Components of Intraday Volatility and Their Prediction at Different Sampling Frequencies with Application to DAX and BUND Futures

Diskussionspapiere 2013
Discussion Papers 2013

01/2013 **Wrede, Matthias**: Rational choice of itemized deductions

02/2013 **Wrede, Matthias**: Fair Inheritance Taxation in the Presence of Tax Planning

03/2013 **Tinkl, Fabian**: Quasi-maximum likelihood estimation in generalized polynomial autoregressive conditional heteroscedasticity models

04/2013 **Cygan-Rehm, Kamila**: Do Immigrants Follow Their Home Country's Fertility Norms?

05/2013 **Ardelean, Vlad and Pleier, Thomas**: Outliers & Predicting Time Series: A comparative study

06/2013 **Fackler, Daniel and Schnabel, Claus**: Survival of spinoffs and other startups: First evidence for the private sector in Germany, 1976-2008

07/2013 **Schild, Christopher-Johannes:** Do Female Mayors Make a Difference? Evidence from Bavaria

08/2013 **Brenzel, Hanna, Gartner, Hermann and Schnabel Claus**: Wage posting or wage bargaining? Evidence from the employers' side

09/2013 **Lechmann, Daniel S. and Schnabel Claus**: Absence from work of the self-employed: A comparison with paid employees

10/2013 **Bünnings, Ch. and Tauchmann, H.**: Who Opts Out of the Statutory Health Insurance? A Discrete Time Hazard Model for Germany

Diskussionspapiere 2012
Discussion Papers 2012

01/2012 **Wrede, Matthias**: Wages, Rents, Unemployment, and the Quality of Life

02/2012 **Schild, Christopher-Johannes:** Trust and Innovation Activity in European Regions - A Geographic Instrumental Variables Approach

03/2012 **Fischer, Matthias:** A skew and leptokurtic distribution with polynomial tails and characterizing functions in closed form

04/2012 **Wrede, Matthias**: Heterogeneous Skills and Homogeneous Land: Segmentation and Agglomeration

05/2012 **Ardelean, Vlad**: Detecting Outliers in Time Series

Diskussionspapiere 2011
Discussion Papers 2011

01/2011 **Klein, Ingo, Fischer, Matthias and Pleier, Thomas:** Weighted Power Mean Copulas: Theory and Application

02/2011 **Kiss, David:** The Impact of Peer Ability and Heterogeneity on Student Achievement: Evidence from a Natural Experiment

03/2011 **Zibrowius, Michael:** Convergence or divergence? Immigrant wage assimilation patterns in Germany

04/2011 **Klein, Ingo and Christa, Florian:** Families of Copulas closed under the Construction of Generalized Linear Means

05/2011 **Schnitzlein, Daniel**: How important is the family? Evidence from sibling correlations in permanent earnings in the US, Germany and Denmark

06/2011 **Schnitzlein, Daniel**: How important is cultural background for the level of intergenerational mobility?

07/2011 **Steffen Mueller**: Teacher Experience and the Class Size Effect - Experimental Evidence

08/2011 **Klein, Ingo:** Van Zwet Ordering for Fechner Asymmetry

09/2011 **Tinkl, Fabian and Reichert Katja:** Dynamic copula-based Markov chains at work: Theory, testing and performance in modeling daily stock returns

10/2011 **Hirsch, Boris and Schnabel, Claus:** Let's Take Bargaining Models Seriously: The Decline in Union Power in Germany, 1992 – 2009

11/2011 **Lechmann, Daniel S.J. and Schnabel**, **Claus** : Are the self-employed really jacks-of-all-trades? Testing the assumptions and implications of Lazear's theory of entrepreneurship with German data

12/2011 **Wrede, Matthias**: Unemployment, Commuting, and Search Intensity

13/2011 **Klein, Ingo**: Van Zwet Ordering and the Ferreira-Steel Family of Skewed Distributions

Diskussionspapiere 2010
Discussion Papers 2010

01/2010 **Mosthaf, Alexander, Schnabel, Claus and Stephani, Jens:** Low-wage careers: Are there dead-end firms and dead-end jobs?

02/2010 **Schlüter, Stephan and Matt Davison:** Pricing an European Gas Storage Facility using a Continuous-Time Spot Price Model with GARCH Diffusion

03/2010 **Fischer, Matthias, Gao, Yang and Herrmann, Klaus:** Volatility Models with Innovations from New Maximum Entropy Densities at Work

04/2010 **Schlüter, Stephan and Deuschle, Carola:** Using Wavelets for Time Series Forecasting – Does it Pay Off?

05/2010 **Feicht, Robert and Stummer, Wolfgang:** Complete closed-form solution to a stochastic growth model and corresponding speed of economic recovery.

06/2010 **Hirsch, Boris and Schnabel, Claus:** Women Move Differently: Job Separations and Gender.

07/2010 **Gartner, Hermann, Schank, Thorsten and Schnabel, Claus:** Wage cyclicality under different regimes of industrial relations.

08/2010 **Tinkl, Fabian**: A note on Hadamard differentiability and differentiability in quadratic mean.

Diskussionspapiere 2009
Discussion Papers 2009

01/2009 **Addison, John T. and Claus Schnabel:** Worker Directors: A German Product that Didn't Export?

02/2009 **Uhde, André and Ulrich Heimeshoff:** Consolidation in banking and financial stability in Europe: Empirical evidence

03/2009 **Gu, Yiquan and Tobias Wenzel:** Product Variety, Price Elasticity of Demand and Fixed Cost in Spatial Models

04/2009 **Schlüter, Stephan:** A Two-Factor Model for Electricity Prices with Dynamic Volatility

05/2009 **Schlüter, Stephan and Fischer, Matthias**: A Tail Quantile Approximation Formula for the Student t and the Symmetric Generalized Hyperbolic Distribution

06/2009 **Ardelean, Vlad:** The impacts of outliers on different estimators for GARCH processes: an empirical study

07/2009 **Herrmann, Klaus:** Non-Extensitivity versus Informative Moments for Financial Models - A Unifying Framework and Empirical Results

08/2009 **Herr, Annika:** Product differentiation and welfare in a mixed duopoly with regulated prices: The case of a public and a private hospital

09/2009 **Dewenter, Ralf, Haucap, Justus and Wenzel, Tobias**: Indirect Network Effects with Two Salop Circles: The Example of the Music Industry

10/2009 **Stuehmeier, Torben and Wenzel, Tobias**: Getting Beer During Commercials: Adverse Effects of Ad-Avoidance

11/2009 **Klein, Ingo, Köck, Christian and Tinkl, Fabian**: Spatial-serial dependency in multivariate GARCH models and dynamic copulas: A simulation study

12/2009 **Schlüter, Stephan**: Constructing a Quasilinear Moving Average Using the Scaling Function

13/2009 **Blien, Uwe, Dauth, Wolfgang, Schank, Thorsten and Schnabel, Claus:** The institutional context of an "empirical law": The wage curve under different regimes of collective bargaining

14/2009 **Mosthaf, Alexander, Schank, Thorsten and Schnabel, Claus:** Low-wage employment versus unemployment: Which one provides better prospects for women?

Diskussionspapiere 2008
Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl:** Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl:** Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm:** Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner:** The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias:** On the Incentives to Form Strategic Coalitions in ATM Markets

06/2008 **Herrmann, Klaus:** Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

07/2008 **Klein, Ingo and Michael Grottke:** On J.M. Keynes' "The Principal Averages and the Laws of Error which Lead to Them" - Refinement and Generalisation