Discussion Papers in Economics

No. 25/2017

(Generalized) Maximum Cumulative Direct, Paired, and Residual ϕ Entropy Principle

Ingo Klein
University of Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/

(Generalized) Maximum Cumulative Direct, Paired, and Residual ϕ Entropy Principle

Ingo Klein*

Friedrich Alexander University Erlangen-Nürnberg (FAU), Germany

December 12, 2017

Jaynes (1957a,b) formulates the maximum entropy (ME) principle as the search for a distribution maximizing a given entropy under some given constraints. Kapur (1984) and Kesavan & Kapur (1989) introduce the generalized maximum entropy principle as the derivation of an entropy for which a given distribution has the maximum entropy property under some given constraints. Both principles will be considered for cumulative entropies. Such entropies depend either on the distribution (direct) or on the survival function (residual) or on both (paired). Maximizing this entropy without any constraint gives a extremely U-shaped (= bipolar) distribution. Under the constraint of fixed mean and variance. maximizing the cumulative entropy tries to transform a distribution in the direction of a bipolar distribution as far as it is allowed by the constraints. A bipolar distribution represents so-called contradictory information in contrast to minimum or no information. Only a few maximum entropy distributions for cumulative entropies have already been derived in the literature. We extend the results to well-known flexible distributions (like the generalized logistic distribution) and derive some special distributions (like the skewed logistic, the skewed Tukey λ and the extended Burr XII distribution). The generalized maximum entropy principle will be applied to the generalized Tukey λ distribution and the Fechner family of skewed distributions. At last, cumulative entropies will be estimated such that the data was drawn from a ME distribution. This estimator will be applied to the daily S&P500 returns and the time duration between mine explosions.

Keywords: Cumulative entropy, maximum entropy distribution, generalized Tukey λ distribution, generalized logistic distribution, skewed logistic distribution, skewed Tukey λ distribution, skewed normal distribution, Weibull distribution, extended Burr XII distribution

*ingo.klein@fau.de

1 Introduction

For a continuous random variable with density f, the classical differential entropy is defined by

$$E_S = -\int f(x) \ln f(x) dx. \tag{1}$$

This measure has some shortcomings in order to be a good measure of information (cf. Schroeder (2004)). For example, it could be negative for special densities. Nevertheless, maximizing (1) with respect to f under the constraint of observed power or L-moments gives maximum entropy (ME) densities (cf. Park & Bera (2008), Hosking (2005)). This ME solution represents a distributional model which is compatible with the minimum information given by the observed constraints. Substituting in (1) the density by the survival function, leads to the cumulative residual (Shannon) entropy. Rao et al. (2004) are the first who discuss this new entropy. The discussion happens in the context of reliability theory. For this context, only random variables with non-negative support are of importance, and the survival function is the natural distributional concept. Rao et al. (2004) and Rao (2005) already discuss the solution of the maxiumum entropy task under power moment constraints. They use the log-sum inequality to derive the ME solution instead of the usual approach based on the Lagrange-Euler equations. The exponential, and more general the Weibull distribution are solutions of this special ME tasks. In the following years, Asadi et al. (2007), Drissi et al. (2008), Kumar & Taneja (2011), Navarro et al. (2010), Baratpour (2010), Baratpour et al. (2012), Psarrakos & Navarro (2013), Chamany & Baratpour (2014) publish papers with the focus on cumulative residual entropies. Especially Drissi et al. (2008) are concerned with the ME problem. They consider random variables with support $\mathbf{R}$ and derive the logistic distribution as ME solution under the additional constraint that the ME solution has to be symmetric. Most recent papers are again focussed on cumulative residual entropies of the Shannon type (cf. Minimol (2017), Mirali et al. (2017), Navarro & Psarrakos (2017) and Sankaran & Sunoj (2017)). Di Crescenzo & Longobardi (2009a), (2009b) apply (1) to the distribution function and call the result "cumulative entropy". Based on early results in the fuzzy set theory (cf. De Luca & Termini (1972), Pal & Bedzek (1994))) concerning membership functions, Li & Liu (2008) define a further entropy concept for so-called uncertainty variables which are similar but not identical to random variables. The main idea is now to consider in (1) the distribution function as well as the survivor function. The obvious corresponding ME task is discussed by Chen & Dai (2011), Liu (2015) and Dai (2017). There is a vast literature concerned with the generalization of (1). Havrda & Charvát (1967), Kapur (1994), Rényi (1961) modify the entropy generating function. General generating function are considered by Csiszár (1963), Morimoto (1963) and Ali & Silvey (1966) for the definition of the related concept of f-divergence (cf. Liese & Vajda (2006)). Sometimes f divergences will be called ϕ-divergences (cf. Cressie & Pardo (2002)). It is obvious that Zografos & Nadarajah (2005) generalize the cumulative residual Shannon entropy in a similar way. Klein et al. (2016) combine the ME task known from uncertainty theory with the use of general entropy generating functions. They derive the Tukey λ distribution as a ME distribution if the entropy generating function of Havrda & Charvát will be applied together to the distribution and the survivor function. They introduce the term "cumulative paired entropy" analogue to the paired ϕ entropy introduced by Burbea & Rao (1993). Recent publications (cf. Kumar (2017), Rajesh & Sunoj (2017)) apply the Havrda & Charvát approach to the survivor function under the name "Cumulative Tsallis entropy" of order α. This term refers to Tsallis famous paper from 1988, where he gives a physical foundation for the Havrda & Charvát approach.

The first purpose of our paper is to unify the diverse approaches of cumulative entropies and their maximization. For this purpose, general cumulative Φ entropies will be introduced. All known variants of cumulative entropies will be special cases of this general class. The ME task in reliability and uncertainty theory is mainly focused on technical reasons. As a second research question, we want to clarify what kind of information cumulative entropies really measure. Therefore, we introduce the concept of "contradictory information" in contrast to "no information". After deriving two general formulas for ME quantile functions under some moment restrictions we apply this formulas to derive ME distributions for new cumulative entropies (like the cumulative Mielke entropy) or to identify the cumulative entropy for some flexible family of distributions (like the generalized Tukey λ or the generalized logistic distribution) allowing for skewness. As a byproduct, we find some new family of distributions (like a special skewed Tukey λ distribution and a generalized Weibull distribution). The last research question starts with the observed data and tries to estimate the cumulative entropy such that the data comes from the corresponding ME distribution. This gives an alternative to the non-parametric estimation of the density or the distribution function.

The paper is organized parallel to this research questions. Section 2 starts with the principal discussion of cumulative entropies. Here we concentrate on the Shannon case. All insights can be transferred to other cumulative entropies immediately. In section 3, we introduce general cumulative Φ entropies and prove general result for ME distributions for cumulative Ψ entropies under different constraints. In section 4, we apply this results to eight families of entropies or families of distributions. In the fourth section, we propose an estimator for the ME generating function. We apply this estimator to real data sets.

2 What does maximizing cumulative direct, paired and residual Shannon entropies mean?

The traditional maximum entropy approach starts with the result that the uniform distribution has minimum information (= maximal entropy) under the constraint that the area under the density sums up to one. In fuzzy set and uncertainty theory, there is another concept of maximal entropy. Transferred to probability theory, maximum uncertainty represents the fact that an event A with probability $0 < P(A) < 1$ and the complementary event $\overline{A}$ with probability $P(\overline{A}) = 1 - P(A)$ are equal-probable. This means $P(A) = 1/2$.

Since the Shannon entropy

$$-P(A)\log P(A) - (1 - P(A))\log P(A)$$

is maximized for $P(A) = 1/2$, this kind of entropy could serve as the basis of an uncertainty measure. For a continuous random variable X, the ensemble of events $(X \leq x)$, $x \in \mathbb{R}$ such that $0 < P(X \leq x) < 1$ can be considered. It is obvious to measure the amount of uncertainty of X by

$$-\int P(X \leq x)\log P(X \leq x)dx - \int (1 - P(X \leq x))\log(1 - P(X \leq x))dx.$$

The integration area is the whole real line $\mathbb{R}$. We set $0\log 0 = 0$. Let F be the cumulative distribution function of X the cumulative paired Shannon entropy is defined by

$$\begin{aligned} CPE_S(F) &= -\int F(x)\log F(x)dx - \int (1 - F(x))log(1 - F(x))dx \\ &= -\int_0^1 (u\log u + (1-u)\log(1-u))q(u)du \end{aligned} \tag{2}$$

with the probability integral transformation $u = F(x)$, the quantile function $Q(u) = F^{-1}(u)$, and the quantile density $q(u) = dQ(u)/du = 1/f(Q(u))$ for $u \in [0, 1]$. f denotes the density of X. If X has a compact support $[a, b]$, $CPE_S(F)$ attains its maximum for $F(x) = 1/2$ for $a \leq x < b$. This corresponds to a so-called bipolar distribution with $P(X = a) = P(X = b) = 1/2$. For this bipolar distribution, $CPE_S(F) = \ln 2(b - a)$, $a < b$ holds. Therefore, the cumulative paired Shannon entropy increases with $b - a$. In contrast to this, the classical Shannon entropy takes a value of $\ln 2$ for all bipolar distributions, independent how great the distance between the two mass points is. Wang et al. (2004) identify this property as an important advantage of cumulative entropies over the differential entropy.

Example 2.1. *To illustrate the different behaviour of the differential entropy and the cumulative Shannon entropy, we consider the symmetric beta distribution with density*

$$f(x;\alpha) = \frac{1}{B(\alpha,\alpha)}x^{\alpha-1}(1-x)^{\alpha-1},\ \ 0 < x < 1,\ \alpha > 0$$

with parameter $\alpha \in (0.1, 2)$*. This range allows almost bipolar distributions (*$\alpha = 0.1$*), uniform distributions (*$\alpha = 1$*), as well as bell-shaped distributions (*$\alpha = 2$*). Figure 1 compares the values of the differential entropy and the cumulative paired Shannon entropy for this range of parameter values. What we learn is that the differential entropy is non-positive everywhere and attains its maximum for the uniform distribution (*$\alpha = 1$*). In contrast to this, the cumulative paired Shannon entropy starts with the maximum value for a bipolar distribution and decreases monotonically with an increase of the parameter* α*.*

The following examples try to explain how bipolar distributions can appear in real

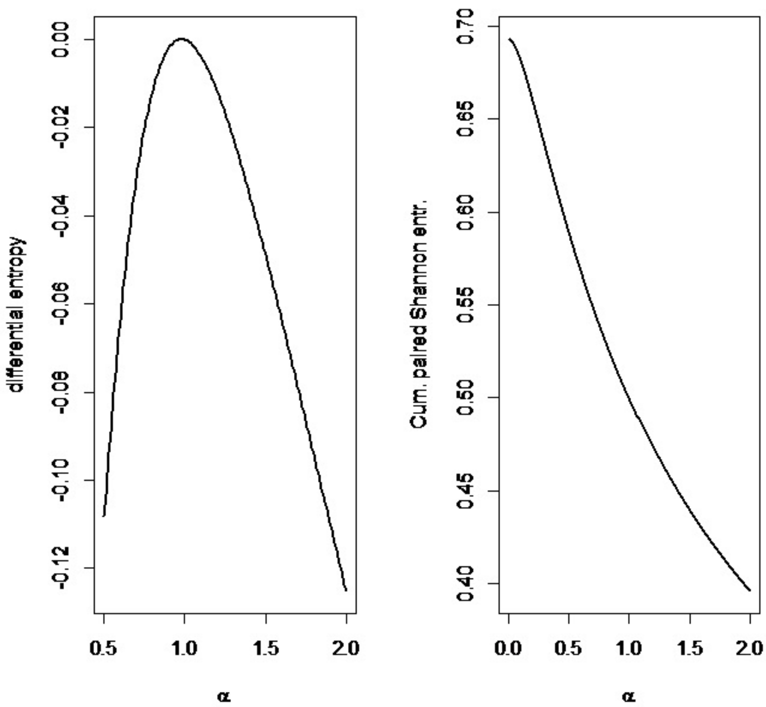

Figure 1: Differential entropy and cumulative paired Shannon entropy for symmetric beta distribution for several parameter values of α.

situations and what this bipolarity means for the predictability of the random variable X.

Example 2.2. *In an opinion poll survey, persons will be asked to judge their political belief on a continuous left-right scale.* 0 *(*100*) symbolizes an extremely left (right) position. The survey results maximizes the cumulative paired Shannon entropy if half of the people say, that they are extremely left* (= 0)*, and the other half, that they are extremely right* (= 100)*. This means a situation of maximal uncertainty, how to predict the political orientation of an individual person.*

Example 2.3. *If the task is to judge a product on a Likert scale with five ordered categories, the uniform distribution means that no category will be favored by the majority of the voters. But, there could be a result of the voting that is still more confusing than the uniform distribution. What can we learn from the extreme situation that the half of the voters give their vote to the best and the other half to the worst category? What does this mean for a new customer thinking to buy the product? Buying means to receive an excellent or a really bad product. This is the situation where it is most complicate to predict what the customer will really do.*

Both situations can be characterized by the term "contradictory information" in contrast to minimal or no information. Information is able to reduce uncertainty. Contradictory information is implicitly defined by the fact that it increases uncertainty, and gives great chance for a wrong decision. [1] Therefore, it is a important task to consider entropies that will be maximized by a bipolar distribution if there are no constraints. In general, cumulative paired entropies are introduced to cover this contradictory information.

Rao et al. (2004) introduce cumulative residual entropies as a Shannon entropy, where the density is substituted by the survivor function. In the following years, Rao (2005), Asadi et al. (2007), Drissi et al. (2008), Kumar & Taneja (2011), Navarro et al. (2010), Baratpour (2010), Baratpour et al. (2012), Psarrakos & Navarro (2013), Chamany & Baratpour (2014), Minimol (2017), Mirali et al. (2017), Navarro & Psarrakos (2017), Sankaran & Sunoj (2017)) published papers with the focus on the cumulative residual Shannon entropy

$$CRE_S(F) = -\int_{infty}^{\infty} (1 - F(x)) \ln(1 - F(x)) dx. \tag{3}$$

The only exception is Di Crescenzo & Longobardi (2009a), (2009b). They apply the Shannon entropy to distribution function and call it cumulative entropy. This gives the formula

$$CDE_S(F) = -\int_{-\infty}^{\infty} F(x) \ln F(x) dx. \tag{4}$$

We will call (4) cumulative direct Shannon entropy for a better distinction to the cumulative paired and the cumulative residual Shannon entropy.

What does maximal entropy mean in this cases? The entropy generating function $-(1-u)\ln(1-u)$ attains its maximum for $u = 1 - 1/e = 0.632 > 0.5$. If the support is $[a, b]$, the maximum CRE_S distribution is again bipolar. But this bipolarity is less extreme than in the symmetric case. It holds $P(X = a) = 1 - 1/e$ and $P(X = b) = 1/e$. There is a preference for the alternative a, what makes the prediction of X easier than in the symmetric case. But, there is still a kind of contradictory information instead of information. For (4), the probabilities for the a and b have to be interchanged to ge a maximal CDE_S distribution.

The following example illustrates for a beta distribution with parameters α and β.

Example 2.4. *Let X be beta distributed with the density*

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \;\; 0 < x < 1, \; \alpha, \beta > 0.$$

In the following we fix β and compute α such that (3) or (4) will be maximized. α_{cre} and α_{ce} denote the corresponding maximal values in the following table 1. This table contains also the maximal values of CRE_S and CE_S. We see that the maximum will be attained for small values of α and β. This means a slightly asymmetric U-shaped beta distribution.

Figure 2 illustrates the maximum CRE_S and CE_S beta distributions for this parameter

[1] Anti-information is a related, but less formal concept introduced by the information scientist J. Verhoeff (1984).

β	α_{cre}	Max. CRE_S	α_{ce}	Max. CDE_S
0.01	0.006	0.3678	0.017	0.3677
0.20	0.111	0.3543	0.290	0.3408
0.50	0.259	0.3222	0.595	0.2970
1.00	0.482	0.2778	1.000	0.2500
2.00	0.905	0.2226	1.000	0.1869

Table 1: ME beta distributions for different values of β.

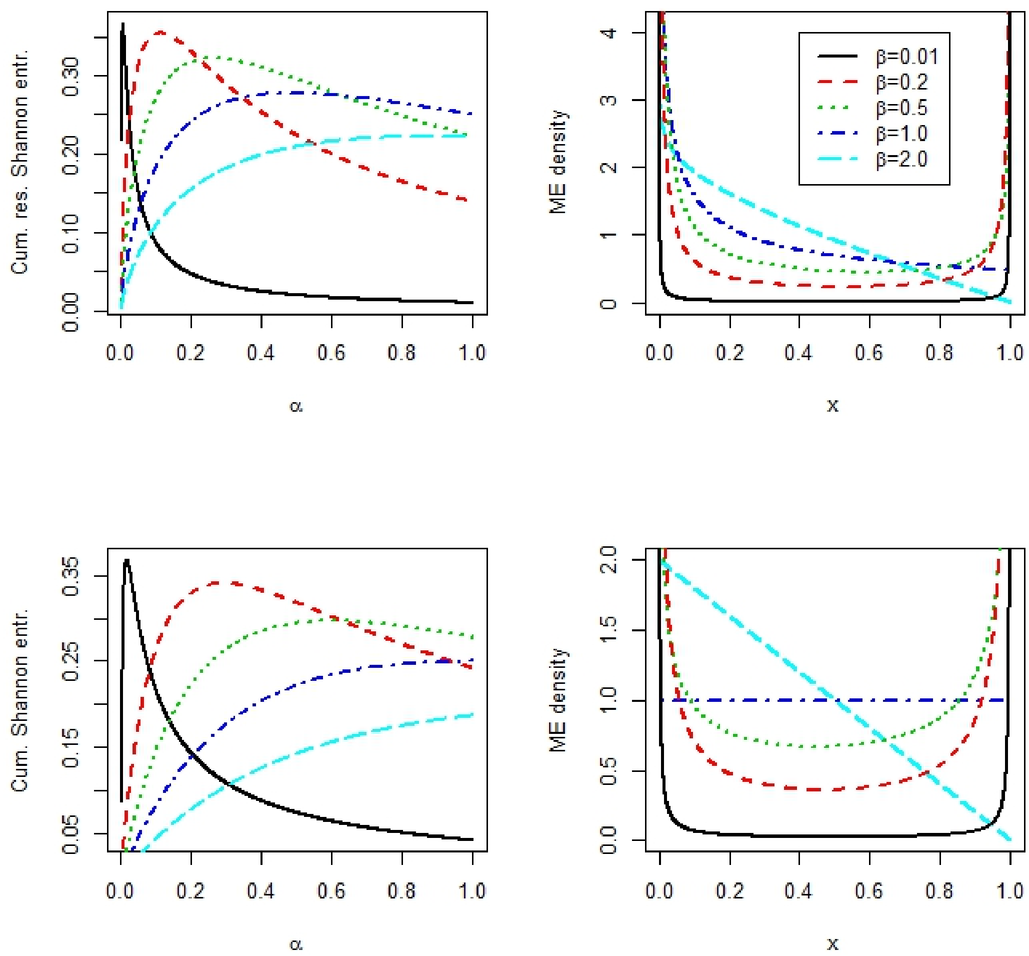

Figure 2: Cumulative residual and cumulative direct Shannon entropy of the asymmetric beta distribution with parameter α such that CRE_S (CE_S) is maximized for given β

setting.

We get a completely different result if we consider only random variable with non-negative support and $Q(0) = 0$ (c.f. Rao et al (2004)). Such distributions are f.e. in the focus of the reliability theory. Maximizing CRE_S or CE_S under the constraint $Q(0) = 0$ gives a distribution which is no longer U-shaped, and the maximal entropy situation no longer corresponds with contradictory information. We illustrate this again by a special beta distribution. The parameter β will be set to 1 such that $Q(0) = 0$.

Example 2.5. *Let X be beta distributed with density*

$$f(x;\alpha) = \frac{1}{\alpha}x^{\alpha-1}, \quad 0 < x < 1, \alpha > 0.$$

Table 2 reports the values for CRE_S and CE_S for selected values of α. For $\alpha = 0.48$ ($\alpha = 1$), CRE_S (CE_S is maximal.

α	CRS_S	CDE_S
0.01	0.1836	0.0826
0.48	0.2779	0.2191
1.00	0.2500	0.2500
3.00	0.1464	0.1876

Table 2: Cumulative residual and cumulative Shannon entropy for a beta distribution with parameters α and $\beta = 1$.

In figure 3, we see that the maximal CRE_S distribution is a compromise between an extremely right skewed ($\alpha = 0.01$) and an extremely left skewed ($\alpha = 3$) distribution.

The question we raised in the section title can be answered by the conclusion that maximizing direct, paired and residual cumulative Shannon entropies lead to a more or less skewed U-shaped distribution as long as there are no special constraints (like $Q(0) = 0$) which are able to prevent this. In the following, we will identify the impact of additional constraints concerning f.e. the mean, the variance or (k-th power moment on the maximum entropy and the generalized maximum entropy task. This has do be done in general as well as under the constraint $Q(0) = 0$.

3 Maximum cumulative entropy distributions

3.1 General class of cumulative Φ entropies

We want to incorporate cumulative direct, paired, and residual cumulative entropies in one approach. Additionally, we want to abstract from the Shannon case and allow for a general so-called entropy generating function ϕ. ϕ has to be a non-negative and concave function on $[0, 1]$. In general, but not mandatory ϕ has a maximum in the interval $[0, 1]$. The corresponding cumulative ϕ entropies are the cumulative direct ϕ entropy with

$$CDE_\phi(F) = \int \phi(F(x))dx = \int_0^1 \phi(u)q(u)du, \tag{5}$$

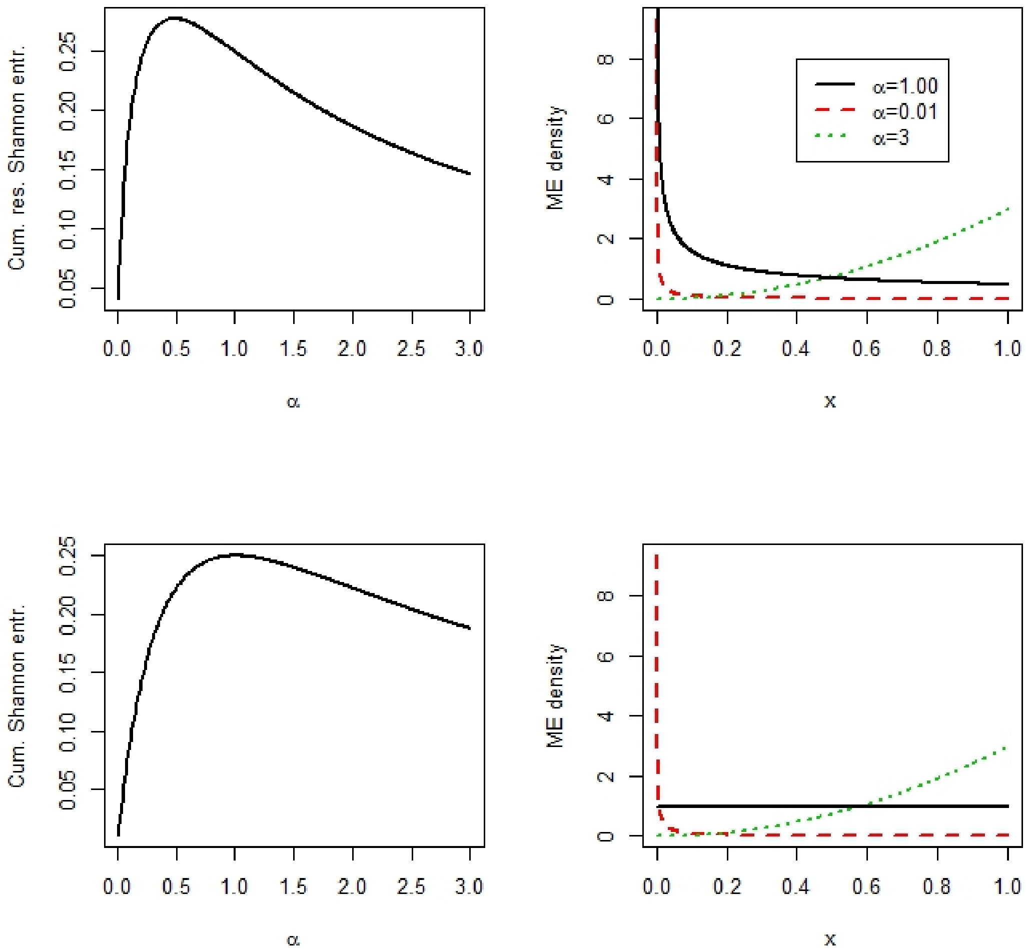

Figure 3: Cumulative residual and cumulative direct Shannon entropies and the density function of a beta distribution with several α and $\beta = 1$.

the cumulative paired ϕ entropy

$$CPE_\phi(F) = \int_{\mathbb{R}} \phi(F(x)) + \phi(1 - F(x))dx = \int_0^1 (\phi(u) + \phi(1-u))q(u)du \tag{6}$$

and the cumulative residual ϕ entropy

$$CRE_\phi(F) = \int_{\mathbb{R}} \phi(1 - F(x))dx = \int_0^1 \phi(1-u)))q(u)du. \tag{7}$$

To cover all three cases, we consider a general concave entropy generating function Φ such that $\Phi(u) = \phi(u)$ or $\Phi(u) = \phi(u) + \phi(1-u)$ or $\Phi(u) = \phi(1-u)$, $u \in [0, 1]$. Then,

$$CE_\Phi(F) = \int_{\mathbf{R}} \Phi(u)du$$

will be called cumulative Φ entropy. The objective is to maximize this cumulative Φ entropy with respect to F under distinct constraints. At first, we consider cumulative Φ entropies in the situation where mean and variance are fixed. The restriction to this both moments can be explained by that fact that higher moments lead to equations for the ME quantile function which cannot be solved explicitly or the solution does not exist. At second, we discuss the same task with the additional requirement that $Q(0) = 0$. This leads to the fact that the solution can only exist for special proportions of the fixed mean and the fixed k-th power moment.

For some choices of Φ, the problem has already been solved. Others will be considered in the following. Thereby, we will focus on Φ leading to well-known distributions. With the ME principle, it is no problem to generate completely new distributions. But this will not be the objective of this paper.

Table 3 gives an overview of results already known from the literature and other results that will be shown in this paper. f_N and F_N^{-1} denote the density and the quantile function of the standard normal distribution.

no.	Φ	ME distribution
	Without $Q(0) = 0$ and fixed mean and variance	
1	$-u\ln u - (1-u)\ln(1-u)$	logistic
2	$u(u^{\alpha-1}-1)/(1-\alpha)$ $+(1-u)((1-u)^{\alpha-1}-1)/(1-\alpha)$	Tukey λ
3	$2(1/2 - \lvert u-1/2\rvert)$	extremely bimodal
4	$2(1/4 - (u-1/2)^2$	uniform
5	$2((1/2)^r - \lvert u-1/2\rvert^r$	symmetric beta
6	$u(u^{\alpha_1-1}-1)/(1-\alpha_1)$ $+(1-u)((1-u)^{\alpha_2-1}-1)/(1-\alpha_2)$	general. Tukey λ
7	$-\frac{B(2-\alpha,\alpha)}{\alpha-1}\left(\beta(u;2-\alpha,\alpha) - \frac{u}{B(2-\alpha,\alpha)}\right)$	general. logistic
8	$-\alpha_1 u\ln u - \alpha_2(1-u)\ln(1-u)$	skewed logistic
9	$\alpha_1 u\frac{u^\alpha-1}{1-\alpha} + \alpha_2(1-u)\frac{(1-u)^\alpha-1}{1-\alpha}$	skewed Tukey λ
10	$2\gamma^3/(1+\gamma^2)f_N(F_N^{-1}(u))I(u \le \gamma^2/(1+\gamma^2))$ $+2/(\gamma(1+\gamma^2))\left(f_N(F_N^{-1}(0.5((1+\gamma^2)u+1-\gamma^2)))\right.$ $\left.-1/\sqrt{2\pi}\right)\cdot I(u > \gamma^2/(1+\gamma^2))$	skewed normal
	With $Q(0) = 0$ and fixed mean and k-th moment	
11	$-(1-u)\ln(1-u)$	Weibull distribution
12	$-(1-u)((1-u)^{\alpha-1}-1)/(\alpha-1)$	extended Burr XII

Table 3: Some entropy generating functions with corresponding ME distributions

Already known are the solutions of number (1) (cf. Liu (2012)), (2) (cf. Klein et al. (2016)), (3) (cf. Klein et al. (2016), (4) (cf. Dai & Chen (2012), Dai (2017)) and (11) (cf. Rao (2005)). The remaining cases state new results. The cases 5 to 12 allow for asymmetry of the ME distribution. For (6) and (10) the generalized maximum entropy principle will be applied. This means that the ME distribution and constraints are fixed and the corresponding entropy will be derived.

3.2 General results for arbitrary support

3.2.1 Maximum cumulative Φ entropy task

The following theorem gives a general formula for the ME quantile function Q. It is the solution of the maximum entropy principle in the sense of Jaynes (1987a,b). This means that the entropy and the constraints are fixed and the ME distribution will be derived.

Theorem 3.1. *Let CE_Φ be the cumulative Φ entropy with concave entropy generating function Φ such that the derivative Φ' exists a.e., is quadratic integrable over $[0,1]$, and $|\Phi(0)| < \infty$, $|\Phi(1)| < \infty$ hold . Then, the maximum CE distribution under the constraint of fixed mean μ and variance σ^2 is given by the quantile function*

$$Q(u) = \mu + \sigma \frac{-\Phi'(u) + (\Phi(1) - \Phi(0))}{\sqrt{\int_0^1 \Phi'(u)^2 du + (\Phi(1) - \Phi(0))^2}}. \tag{8}$$

Proof: The objective function

$$\int \Phi(F(x))du = \int_0^1 \Phi(u)q(u)du$$

has to be maximized under the restrictions of fixed

$$\mu = \int_0^1 Q(u)du \text{ and } \sigma^2 + \mu^2 = \int_0^1 Q(u)^2 du$$

with respect to the quantile function Q and the quantile density q. This leads to the Lagrange function

$$\begin{aligned}\mathcal{L}(q, Q, \lambda_1, \lambda_2) = & \int_0^1 \Phi(u)q(u)du - l_1 \left(\int_0^1 Q(u)du - \mu \right) \\ & -l_2 \left(\int_0^1 Q(u)^2 du - \sigma^2 + \mu^2 \right).\end{aligned}$$

The Euler-Lagrange equation gives

$$\frac{d}{du}\frac{\partial \mathcal{L}}{\partial q} - \frac{\partial \mathcal{L}}{\partial Q} = \Phi'(u) + l_1 + 2l_2 Q(u) \stackrel{!}{=} 0.$$

l_1 and l_2 denote the Lagrange parameter. Solving this equation leads to the quantile function

$$Q(u) = \frac{1}{2l_2}\left(-\Phi'(u) - l_1\right).$$

l_1 and l_2 are determined by the moments μ and σ^2. Rearranging

$$\mu = \frac{1}{2l_2}\int_0^1 (-\Phi'(u) - l_1)du = -\frac{1}{2l_2}(\Phi(1) - \Phi(0)) - \frac{l_1}{2l_2}$$

leads to

$$l_1 = -2l_2\mu - (\Phi(1) - \Phi(0))$$

and

$$Q(u) = \mu - \frac{1}{2l_2}\left(\Phi'(u) - (\Phi(1) - \Phi(0))\right). \qquad (9)$$

From

$$\begin{aligned}
\mu^2 + \sigma^2 &= \int_0^1 Q(u)^2 du \\
&= \mu^2 + \frac{1}{l_2}\mu\left(\int_0^1 \Phi'(u) - (\Phi(1) - \Phi(0))du\right) \\
&\quad + \frac{1}{4l_2^2}\int_0^1 \left(\Phi'(u) - (\Phi(1) - \Phi(0))\right)^2 du \\
&= \mu^2 + \frac{1}{4l_2^2}\left(\int_0^1 \Phi'(u)^2 du - 2\int_0^1 \Phi'(u)du(\Phi(1) - \Phi(0)) + (\Phi(1) - \Phi(0))^2\right) \\
&= \mu^2 + \frac{1}{4l_2^2}\left(\int_0^1 \Phi'(u)^2 du - (\Phi(1) - \Phi(0))^2\right).
\end{aligned}$$

Solving with respect to l_2 leads to

$$l_2 = \frac{1}{2\sigma}\sqrt{\int_0^1 \Phi'(u)^2 du - (\Phi(1) - \Phi(0))^2}.$$

Inserting l_2 into (9) gives the quantile function (8). $\square$

3.2.2 Generalized maximum cumulative Φ entropy task

Starting with a quantile function Q, the formula (8) can also be used to derive the corresponding generating function Φ of the cumulative Φ entropy. This procedure follows the generalized maximum entropy principle formulated by Kesavan & Kapur (1989). For a simpler notation we introduce the partial mean function. Let X be the random variable corresponding to Q and f be the density of X. This function is given by

$$\begin{aligned}
\mu(u) &= \int_0^u Q(v)dv = \int_{-\infty}^{Q(u)} xf(x)dx = -E(X|X \leq Q(u))P(X \leq Q(u)) \qquad (10) \\
&= uE(X|X \leq Q(u)) \;\; u \in [0,1].
\end{aligned}$$

Obviously, $\mu(0) = 0$ and $\mu(1) = \mu$ hold.

The following corollary states that the negative of the partial mean function determines the entropy generating function such hat Q is the ME quantile function under the constraint of given mean μ and variance σ^2.

Corollary 3.1. *Let Q be a quantile function. The entropy generating function Φ such that Q is ME under the constraint of given mean and variance is given by $\Phi(u) = -\mu(u)$, $u \in [0,1]$.*

Proof: Setting $\Phi'(u) = -Q(u)$, $u \in [0,1]$ gives

$$\Phi(u) = -\int_0^u Q(v)dv = -\mu(u) \quad u \in [0,1]. \quad \square \tag{11}$$

Hence, $-\Phi(u)/u$ is the conditional mean of X given $X \leq Q(u)$ fur $u \in [0,1]$. It holds $\mu(0) = 0$ and $\mu(1) = \mu$ such that $\Phi(0) = 0$ and $\Phi(1) = -\mu$. The special role of the partial mean function $\mu(u)$ is easy to explain. $\mu(u)$ sums up the values x of X weighted with the density $f(x)$ until the u-quantile of X. For an extremely U-shaped distribution this addition gives constant values until the the median quantile. Thereafter, the value will be changed one time and stays again constant. The heavier the tails of a distribution are, the steeper the entropy generating function $\Phi(u)$ at $u = 0$ and $u = 1$ is. This leads to a large value for the derivative $\Phi'(u)$ at $u = 0$ and $u = 1$. If the support is $\mathbf{R}$, then $\lim \Phi'(0) = \infty$ and $\lim \Phi'(u) = -\infty$. We will use (11) to derive Φ such that a given distribution has the ME property under the constraint of fixed mean and variance. Based on (11), in the last section we propose an estimator for Φ.

3.3 General results for non-negative support

3.3.1 Maximum cumulative Φ entropy task

We restrict the support of the ME distribution to $(0, \infty)$. This means that $Q(0) = 0$ holds for the ME quantile function Q. From this fact, we get an additional constraint for the ME task. As further constraints, we consider a fixed mean μ and a fixed k-th power moment μ_k', $k > 1$. The following theorem shows how to derive the ME quantile function under this three constraints. The ME solution requires a special relationship between the fixed moments μ and μ_k'. Otherwise, there is no solution of the ME task.

Theorem 3.2. *Let Φ a be concave function on $[0,1]$ with derivative Φ' such that $-\Phi'(u)+\Phi'(0)$ is monotone increasing. Then the ME quantile function under the constraint of given mean and k-th power moment μ_k' is*

$$Q(u) = \frac{(-\Phi'(u) + \Phi'(0))^{1/(k-1)}}{\int_0^1 (-\Phi'(u) + \Phi'(0))^{1/(k-1)} du}\mu. \tag{12}$$

This solution is only valid, if

$$\frac{\mu_k'}{\mu^k} = \frac{\int_0^1 (-\Phi'(u) + \Phi'(0))^{k/(k-1)} du}{\left(\int_0^1 (-\Phi'(u) + \Phi'(0))^{1/(k-1)} du\right)^k}. \tag{13}$$

Proof: Due to The Euler-Lagrange equation it is

$$\Phi'(u) + l_1 + kl_2 Q(u)^{k-1} \stackrel{!}{=} 0, \quad u \in [0,1].$$

The constraint $Q(0) = 0$ leads to $l_1 = -\Phi'(0)$ and

$$Q(u) = ((-\Phi'(u) + \Phi'(0))/(kl_2))^{1/(k-1)}, \;\; u \in [0, 1].$$

kl_2 can be derived from

$$\mu = \int_0^1 Q(u)du = \left(\frac{1}{kl_2}\right)^{1/(k-1)} \int_0^1 (-\Phi'(u) + \Phi'(0))^{1/(k-1)}du$$

as

$$kl_2 = \left(\frac{1}{\mu}\right)^{k-1} \left(\int_0^1 (-\Phi'(u) + \Phi'(0))^{1/(k-1)}du\right)^{k-1}.$$

Inserting kl_2 into $Q(u)$ gives (12) immediately.

There is the third constraint

$$\mu'_k = \int_0^1 Q(u)^k du = \frac{\int_0^1 (-\Phi'(u) + \Phi'(0))^{k/(k-1)}du}{\left(\int_0^1 (-\Phi'(u) + \Phi'(0))^{1/(k-1)}du\right)^k} \mu^k.$$

Dividing μ^k on both sides gives (13). $\square$.

In the most popular case mean and variance are fixed. This means $k = 2$ and

$$Q(u) = \frac{-\Phi'(u) + \Phi(0)}{\int_0^1 (-\Phi'(u) + \Phi'(0))du} \mu = \frac{-\Phi'(u) + \Phi'(0)}{-\Phi(1) + \Phi(0) + \Phi'(0)} \mu$$

and

$$\begin{aligned} \frac{\mu'_2}{\mu^2} &= \frac{\int_0^1 (-\Phi'(u) + \Phi'(0))^2 du}{\left(\int_0^1 (-\Phi'(u) + \Phi'(0))du\right)^2} \\ &= \frac{\int_0^1 \Phi'(u)^2 du - 2(\Phi(1) - \Phi(0))\Phi'(0) + \Phi'(0)^2}{(\Phi(1) - \Phi(0))^2 - 2(\Phi(1) - \Phi(0))\Phi'(0) + \Phi'(0)^2} \end{aligned}$$

3.3.2 Generalized maximum cumulative Φ entropy task

It remains to discuss the generalized maximum entropy principle for random variables with non-negative support and $Q(0) = 0$. We start with the knowledge of the quantile function Q to derive the corresponding generating function Φ of the cumulative Φ entropy such that Q is the ME quantile function for Φ under the constraints $Q(0) = 0$ and fixed mean μ and fixed k-th power moment μ'_k. It is again helpful to introduce a special partial mean function. $\mu_{k-1}(u)$ denotes the partial $(k-1)$-th power mean function with

$$\mu_{k-1}(u) = uE[X^{k-1}|X \leq Q(u)], \;\; u \in [0, 1]$$

for $k = 2, 3, \ldots$. This partial $(k-1)$-th power moment function is an important part of the entropy generating function as the following corollary shows.

Corollary 3.2. *Let Q be a quantile function. The entropy generating function Φ, such* *that Q is ME under the constraints $Q(0) = 0$, fixed mean and fixed variance, is given by*

$$\Phi(u) = \mu'_{k-1}u - \mu_{k-1}(u), \quad u \in [0,1]. \tag{14}$$

Proof: Let X be the random variable corresponding to Q and f be the density of X. From

$$Q(u) = (-\Phi'(u) - \Phi'(0))^{1/(k-1)}, \quad u \in [0,1]$$

we get

$$-\Phi(u) + \Phi'(0)u = \int_0^u Q(v)^{k-1}dv \equiv \mu_{k-1}(u), \quad u \in [0,1].$$

It is easy to verify that $\Phi(0) = 0$. Under the assumption $\Phi(1) = 0$ it is $\Phi'(0) = \mu'_{k-1}$ and

$$\Phi(u) = \mu'_{k-1}u - \mu_{k-1}(u), \quad u \in [0,1]. \quad \square \tag{15}$$

In the last section we will use (15) to estimate Φ from a data set such that the data is generated by the corresponding ME distribution under the constraints of $Q(0) = 0$ and fixed mean and fixed k-the power moment.

4 Applications

4.1 Cumulative paired Mielke(r) entropy

Mielke (1972) and Mielke & Johnson (1973) discuss two-sample linear rank tests for alternatives of scale with score generating function

$$\left| u - \frac{1}{2} \right|^r, \quad u \in [0,1], \ r > 0.$$

This family of tests, parameterized by r, includes the well-known test from Ansari & Bradley (1960) for $r = 1$ and Mood (1954) for $r = 2$. He derived the symmetric distribution for an asymptotically optimal with this score generating function. Similar to Klein et al. (2016), there is a simple relationship between the score generation functions of two sample linear rank tests for scale alternatives and the entropy generating function Φ of a corresponding cumulative paired Φ entropy. In the case Mielke considers, we get the entropy generating function

$$\Phi(u) = 2\left(\left(\frac{1}{2}\right)^r - \left| u - \frac{1}{2} \right|^r \right) \quad u \in [0,1].$$

This function is strictly concave on $[0,1]$ for $r > 1$ and concave for $r = 1$. Therefore, we will only consider this cases. Φ is differentiable and the derivative twice integrable. It holds $\Phi(1/2) = (1/2)^(r) \geq \Phi(u)$, $u \in [0,1]$ and $\Phi(1) = \Phi(0) = 0$. In the following, we abbreviate the cumulative paired Mielke(r) entropy by $CPE_M(r)$). Now, we will show that the double β distribution maximizes $CPR_{M(r)}$, $r > 1$ if mean and variance are known.

For $r = 1$, we get an extremely U-shaped distribution. In both cases the support of the ME distribution is a closed interval. The four-parameter double β distribution is defined by

$$f(x; a, b, c, d) = \frac{1}{2B(a,b)} \frac{1}{d} \left(\frac{x-c}{d} \right)^{a-1} \left(1 - \frac{x-c}{d} \right)^{b-1}, \; c - d \leq x \leq c + d$$

with $a, b, d > 0$, $c \in \mathbf{R}$.

Corollary 4.1. *The distribution F maximizing*

$$\int_0^1 2 \left(\left(\frac{1}{2} \right)^r - \left| F(x) - \frac{1}{2} \right|^r \right) dx$$

under the constraint of known mean μ and known variance σ^2 is a double β-distribution with parameters $a = 1/(r-1)$, $b = 1$ and support $[\mu - \sigma\sqrt{2r-1}, \mu + \sqrt{2r-1}\sigma]$ for $r > 1$ an extremely bimodal distribution with support $\{\mu - \sigma, \mu + \sigma\}$ for $r = 1$.

Proof: The derivative of Φ is given by

$$\Phi'(u) = -2r|u - 1/2|^{r-1} \text{sign}(u - 1/2), \;\; u \in [0,1], \; r \geq 1.$$

In (8) we need the expected value of $\Phi'(U)^2$ for $U \sim R(0,1)$. This expected value is given by

$$\begin{aligned} E[\Phi'(U)^2] &= 4r^2 \int_0^1 (u - 1/2)^{2r-2} du = 4r^2 \int_{-1/2}^{1/2} v^{2r-2} dv \\ &= 4r^2 2 \int_0^{1/2} v^{2r-2} dv = \frac{4r^2}{2r-1} (1/2)^{2r}. \end{aligned}$$

With $\Phi(0) = \Phi(1) = 0$ and according to (8) the quantile function is given by

$$Q(u) - \mu = \sigma\sqrt{2r-1}|2u - 1|^{r-1} \text{sign}(u - 1/2), \;\; u \in [0,1]$$

with the support given by

$$[\mu - \sigma\sqrt{2r-1}, \mu + \sigma\sqrt{2r-1}].$$

for $r \neq 1$ and $\{\mu - \sigma, \mu + \sigma\}$ for $r = 1$. For $r \geq 1$ the corresponding distribution function is

$$F(x) = 1/2 \left(1 + \text{sign}(x - \mu) \left(\frac{|x - \mu|}{\sigma\sqrt{2r-1}} \right)^{1/(r-1)} \right)$$

for $x \in [\mu - \sigma\sqrt{2r-1}, \mu + \sigma\sqrt{2r-1}]$. The corresponding density is

$$f(x) = 1/2 \frac{1}{\sigma\sqrt{2r-1}} \frac{1}{r-1} \left| \frac{x - \mu}{\sigma\sqrt{2r-1}} \right|^{1/(r-1)-1}$$

for $x \in [\mu - \sigma\sqrt{2r-1}, \mu + \sigma\sqrt{2r-1}]$. This is a β distribution with parameters $a = 1/(r-1,$

$b = 1$ and support $x \in [\mu - \sigma\sqrt{2r-1}, \mu + \sigma\sqrt{2r-1}]$. For $r = 1$ we get the extremely bi-modal distribution with mass points $\mu - \sigma$ and $\mu + \sigma$. $\square$

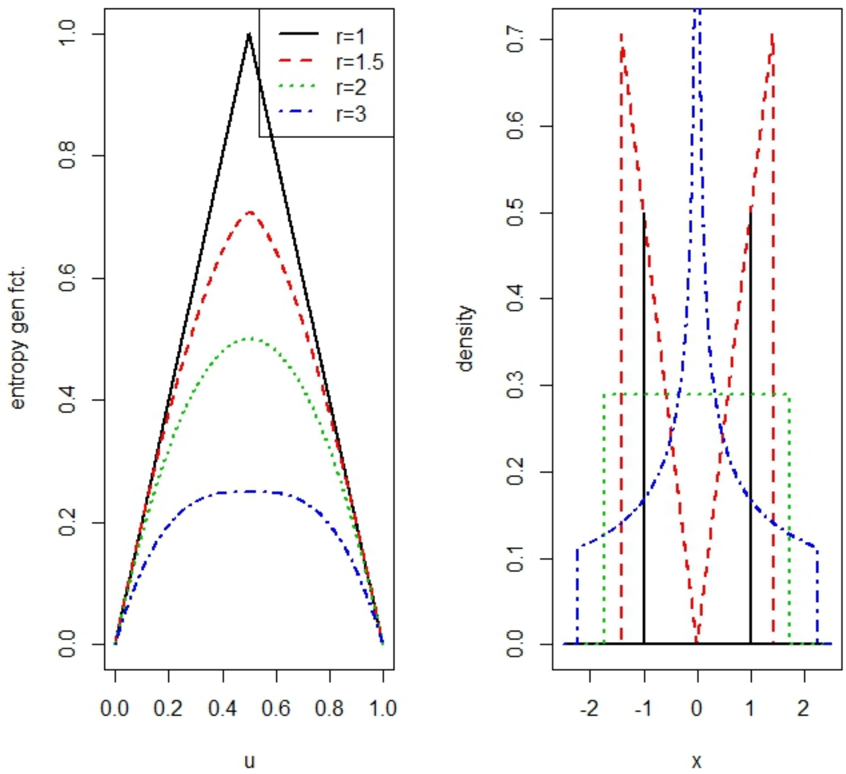

Figure 4: Entropy generating and density of the double beta distribution with parameter $r = 1, 2, 3, 4$.

In figure 4, the entropy generating functions and ME densities demonstrate the impact of different settings for the parameter r. For $r = 1$, the entropy generating function is a triangle, such that the density is bipolar. Increasing r leads to a bi-modal distribution ($r = 2$), an uniform distribution ($r = 3$), up to an very leptokurtic distribution with singularity at 0 ($r = 4$). With increasing r, the entropy generating functions becomes more and more flat. The entropy generating function starts at 0 and 1 with an absolute value of the derivative which is smaller than $1/2$. This characterizes the compact support of all ME distributions for the cumulative paired Mielke(r) entropy. All the distributional characteristics can be learned from the form of the entropy generating function.

Klein et al. (2016) discuss the cumulative paired Leik entropy and Dai & Chen (2012) the cumulative paired Gini entropy. Both are embedded in the class of cumulative paired Mielke(r) entropy for $r = 1$ and $r = 2$.

4.2 Generalized Tukey λ distribution

The second possibility to interpret (8) is to fix a quantile function and derive the entropy generating function Φ such that Q represents the corresponding ME distribution. This procedure has been called the generalized cumulative Φ entropy approach. We want to apply this approach to a quantile function considered by Freimer et al. (1988). They introduced the so-called generalized Tukey λ distribution with the help of the quantile function

$$Q(u) = \lambda_4 + \frac{1}{\lambda_3}\left(\frac{u^{\lambda_1}-1}{\lambda_1} - \frac{(1-u)^{\lambda_2}-1}{\lambda_2}\right), \quad u \in [0,1].$$

λ_4 is a location, λ_3 is a scale parameter and λ_1, λ_2 determine the distribution's skewness.

Freimer et al. (1988) show that the k-th moment exists iff $\min(\lambda_1, \lambda_2 > -1/k$. This means that the variance exists for $\min(\lambda_1, \lambda_2) > -1/2$. Simple calculations lead to formulas for the mean

$$\mu = \lambda_4 + \frac{1}{\lambda_3}\left(\frac{1}{\lambda_1+1} - \frac{1}{\lambda_2+1}\right)$$

and the variance

$$\begin{aligned}\sigma^2 = \frac{1}{\lambda_3}^2 &\left(\frac{1}{(2\lambda_1+1)(\lambda_1+1)^2} + \frac{1}{(2\lambda_2+1)(\lambda_2+1)^2}\right.\\ &\left.-\frac{2}{\lambda_1\lambda_2}\left(B(\lambda_1+1,\lambda_2) + \frac{1}{(\lambda_1+1)(\lambda_2+1)}\right)\right).\end{aligned}$$

In the following corollary, we show how the entropy generating functions look like for which the generalized Tukey Λ distribution is a maximum entropy distribution under the constraint of fixed mean and variance.

Corollary 4.2. *Let Q be the quantile function of the generalized Tukey λ distribution. The entropy generating function Φ such that Q is the ME quantile function under the constraint of known mean and variance is given by*

$$\Phi(u) = -\frac{u^{\lambda_1+1}-(\lambda_1+1)u}{\lambda_1(\lambda_1+1)} - \frac{(1-u)^{\lambda_2+1}-(\lambda_2+1)(1-u)}{\lambda_2(\lambda_2+1)} - \frac{1}{\lambda_2+1} \tag{16}$$

for $u \in [0,1]$, $\lambda_1 > 0$, $\lambda_2 > 0$.

Proof: It is $\mu(u) = -\int_0^u Q(v)du = -\Phi(u)$. For $\mu(u)$ we get

$$\begin{aligned}\mu(u) &= \int_0^u \frac{v^{\lambda_1}-1}{\lambda_1}dv - \int_0^u \frac{(1-v^{\lambda_2})-1}{\lambda_2}dv\\ &= \frac{u^{\lambda_1+1}-(\lambda_1+1)u}{\lambda_1(\lambda_1+1)} + \frac{(1-u)^{\lambda_2+1}-(\lambda_2+1)(1-u)}{\lambda_2(\lambda_2+1)} + \frac{1}{\lambda_2+1}. \quad \square\end{aligned}$$

The setting $\lambda = \lambda_1 = \lambda_2$ results in the symmetric Tukey λ distribution. Klein et al. (2016) identify this distribution as ME distribution for the entropy generating function $\Phi(u) = -u(u^\lambda - 1)/\lambda - (1-u)^\lambda - 1)/\lambda$, $u \in [0,1]$. This is (up to constant $1/(\lambda+1)$) identical with the entropy generating function (16). In the upper panel of figure 5, we

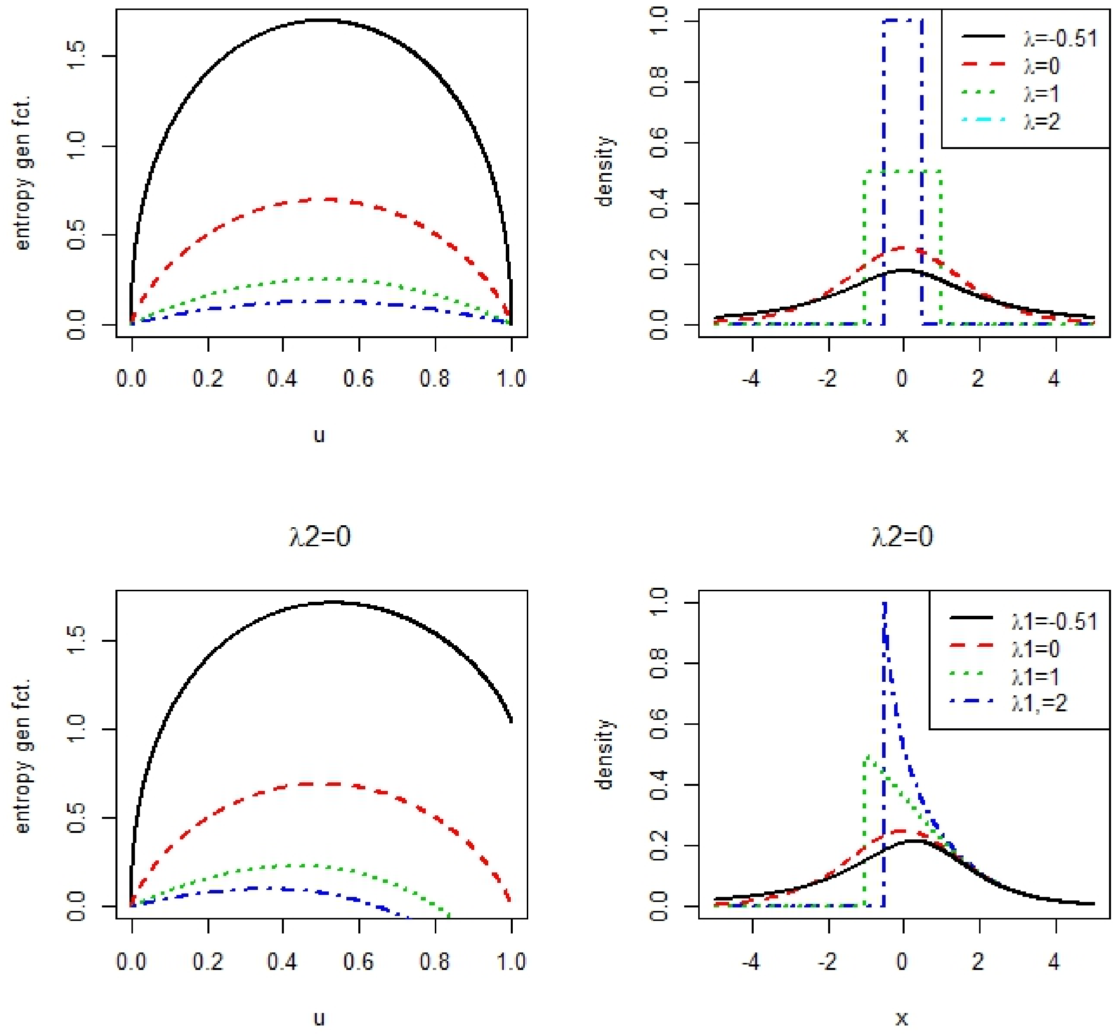

Figure 5: Entropy generating function and density for the Tukey λ ($\lambda = 0.51, 0, 1, 2$) and the generalized Tukey λ distribution ($\lambda_1 = 0.51, 0, 1, 2,\ \lambda_2 = 0$).

see the corresponding entropy generating function and the ME density for several choices of λ. The range of distributional properties for the Tukey λ distribution is much richer than it is for the ME distributions belonging to the cumulative paired Mielke(r)-entropy. For negative values of λ, the support is the whole real line $\mathbb{R}$ since the entropy generating function has an non-finite derivative at 0 and 1. For $\lambda = 0$, we get the logistic distribution. For positive values of λ the support is a compact interval. $\lambda = 1$ and $\lambda = 2$ give uniform distributions.

To demonstrate the consequence of skewness, the lower panel of figure 5 shows the entropy generating function and the density for $\lambda_1 \neq \lambda_2$. We set $\lambda_2 = 0$ and vary only the values of λ_1. Now, for $\lambda_1 \neq 0$ it is $\Phi(1) \neq 0$. This can be explained by the skewness of the distribution. Φ represents the cumulative mean function. For a skewed distribution, summing up the quantile function over the positive part does not exactly compensate the sum of the quantile function over the negative part. This latter part can be smaller for a left-skewed and greater for a right-skewed distribution.

Chalabi et al. (2012) give an excellent overview about the properties, parameter estimation and applications of the generalized Tukey λ distribution. We recommend to study the long list of references in their paper. Also van Staden (2013) discusses all this aspects in detail in his Ph.D. thesis from the University of Pretoria. Van Staden mentions several applications in distinct scientific fields ranging from actuarial science over finance until supply chain planning (pp. 127). Chalabi et al. (2012) focus on applications in finance. King & McGillivray (1999) discuss the ordering properties of skewness and kurtosis for the generalized λ distribution. Su (2007), Su (2012) and van Staden & Loots (2009) are concerned with estimation in this family.

4.3 Generalized logistic distribution

A further way to use formula (8) is to start not with an entropy generating function but with its derivative. Such a derivative could be

$$\Phi'(u) = \Phi\left(\frac{1-u}{u}\right), \quad u \in [0,1] \tag{17}$$

where Φ is strictly increasing on $[0,1]$. An example is $\Phi(x) = \ln x$, $x > 0$ such that

$$\Phi'(u) = \ln\left(\frac{1-u}{u}\right) \quad u \in [0,1]. \tag{18}$$

In this special case, Φ belongs to the cumulative paired Shannon entropy and the corresponding ME distribution is the logistic distribution under the constraint of fixed mean and variance (cf. Liu (2007), Li & Liu (2008), Liu (2015), Klein et al. (2016)).

A plausible extension is to consider

$$\Phi'(u) = \frac{((1-u)/u)^{\alpha-1} - 1}{\alpha - 1}, \quad u \in [0,1], \ \alpha \neq 1. \tag{19}$$

For $\alpha \to 1$ we get (18).

Corollary 4.3. *The density of the ME distribution for the cumulative entropy with derivative (19) is given by*

$$f(x) = \frac{1-(\alpha-1)x)^{1/(\alpha-1)-1}}{\left(1+(1-(\alpha-1)x)^{1/(\alpha-1)}\right)^2}$$

for $x \in (-\infty, 1/(\alpha-1))$ *and* $1 < \alpha < 3/2$ *and* $x \in (-1/(\alpha-1), \infty)$ *if* $1/2 < \alpha < 1$.

Proof: Φ' determines the ME quantile function via (8). This means

$$Q(u) = -\frac{((1-u)/u)^{\alpha-1} - 1}{\alpha - 1}, \quad u \in [0,1]. \tag{20}$$

The support is given by

$$\text{supp}(F) = (Q(0), Q(1)) = \begin{cases} (-\infty, 1/(\alpha-1)) & \text{for } \alpha > 1 \\ (1/\alpha - 1, \infty) & \text{for } \alpha < 1 \end{cases}$$

The considered ME task requires the existence of the variance. The k-th power mean

$$\begin{aligned} \int_0^1 Q(u)^k du &= \left(\frac{1}{\alpha-1}\right)^k \int_0^1 \left(1 - u^{1-\alpha}(1-u)^{\alpha-1}\right)^k du \\ &= \left(\frac{1}{\alpha-1}\right)^k \sum_{i=0}^{k} \binom{k}{i} (1-)^{k-i} B((1-\alpha)(k-i)+1, (\alpha-1)(k-i)+1) \end{aligned}$$

exists for $(1-\alpha)(k-i)+1 > 0$ and $(\alpha-1)(k-i)+1) > 0$ for $i = 1, 2, \ldots, k$. This means $1 - 1/k < \alpha < 1 + 1/k$. Therefore, the variance exists for $1/2 < \alpha < 3/2$.

Solving $Q(u) = x$ with respect to u gives

$$F(x) = \left(1 + (1 - (\alpha-1)x)^{1/(\alpha-1)}\right)^{-1} \quad x \in \text{supp}(F)$$

and differentiating the distribution function delivers the postulated density. $\square$.

The quantile function (20) has been discussed by Shabri et al. (2011). They introduce the term "generalized logistic distribution" (GLO). The generalization is due to the fact that skewness will be introduced into the logistic distribution by the parameter α. The support also depends on the parameter α. For $\alpha \to 1$, we result in the support $\mathbf{R}$ and the logistic distribution. This is the only symmetric distribution in this GLO class. For $\alpha > 1$, the support is $(-\infty, 1/(\alpha-1)$ and the GLO distribution is left-skewed. For $\alpha < 1$, we get the support $[-1/(1-\alpha), \infty)$ and a right-skewed distribution.

A working paper written by Morais & Cordeiro (no year) gives an overview about other generalizations of the logistic distribution that are completely different to (20). Nassar & Elmasry (2012) and Tripathi et al. (2017) are also concerned with generalizations of the logistic distribution.

It remains to identify the entropy generating function Φ belonging to (19). By simple integration, we get the partial sum function

$$\begin{aligned} \mu(u) &= \frac{1}{\alpha-1}\left(\int_0^u ((1-p)/p)^{\alpha-1} dp - u\right) \\ &= \frac{1}{\alpha-1}\left(\int_0^u p^{1-\alpha}(1-p)^{\alpha-1} dp - u\right) \\ &= \frac{B(2-\alpha,\alpha)}{\alpha-1}\left(\beta(u; 2-\alpha, \alpha) - \frac{u}{B(2-\alpha,\alpha)}\right) \end{aligned}$$

with $\beta(.; a, b)$ as the distribution function of the $\beta(a, b)$ distribution. Then, $\Phi(u) = -\mu(u)$, $u \in [0, 1]$ with $\Phi(1/2) = 0$ and $\Phi(1) = -\mu = /1(\alpha-1)(1 - B(2-\alpha, \alpha)$ for $1/2 < \alpha < 3/2$, $\alpha \neq 1$ hold.

Figure 6 demonstrates Φ together with the density of the GLO distribution. The logistic distributions serves as a reference ($\alpha = 1$). The other settings show left-skewed ($\alpha = 0.6$) and two right-skewed distributions ($\alpha = 1.2$ and $\alpha = 1.4$).

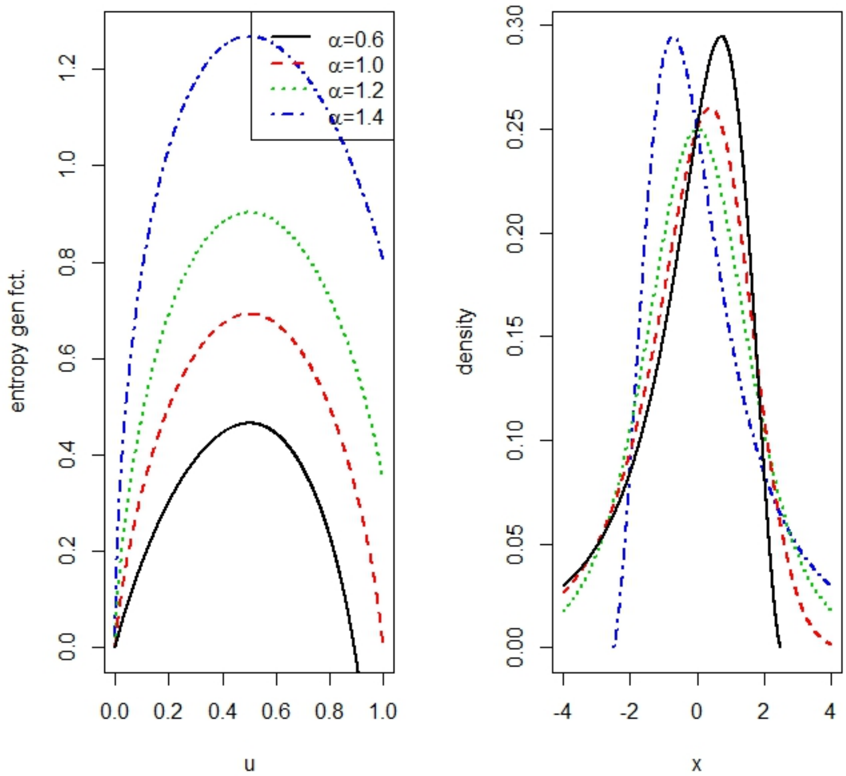

Figure 6: Entropy generating function and density for the generalized logistic (GLO) distribution with parameter $\alpha = 0.6, 1, 1.2, 1.4$.

4.4 "New" skewed logistic and Tukey λ distributions

In Klein et al. (2016) entropy generating functions of the symmetric form

$$\Phi(u) = \varphi(u) + \varphi(1-u), \quad u \in [0,1]$$

with φ concave on $[0,1]$ have been discussed. Again, we quote the result that for $\varphi(u) = -u \ln u$ we get the logistic and for $\varphi(u) = u(u^{\alpha-1} - 1)/(1-\alpha)$ the Tukey λ distribution as ME distributions under the constraint of given mean and variance.

In the following, we consider the more general case that

$$\Phi(u) = \varphi_1(u) + \varphi_2(1-u), \quad u \in [0,1]$$

with φ_i, $i = 1, 2$ concave on $[0, 1]$. More special is the form

$$\Phi(u) = \alpha_1\varphi(u) + \alpha_2\varphi(1-u), \quad u \in [0, 1],$$

where $\alpha_i > 0$, and φ_i concave on $[0, 1]$, $i = 1, 2$. If φ is symmetric around $1/2$, then $\Phi(u) = (\alpha_1 + \alpha_2)\varphi(u)$, $u \in [0, 1]$. $\alpha_1 + \alpha_2$ only plays the role of an scale parameter. This case will be not considered in the following since we are interested in skewed ME distributions.

Examples for asymmetric functions are

$$\varphi(u) = \begin{cases} -u\frac{u^\lambda - 1}{\lambda} & \text{for } u \in [0, 1],\ \lambda \neq 0 \\ -u \ln u & \text{for } u \in [0, 1],\ \lambda = 0 \end{cases} \tag{21}$$

We look for ME distributions in this special case. To get only one skewness parameter, we set $\alpha_1 + \alpha_2 = 1$ such that $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$ for $\alpha \in [0, 1]$. The following corollary informs about the corresponding ME quantile function.

Corollary 4.4. *Let $\Phi(u) = \alpha\varphi(u) + (1-\alpha)\varphi(u)$, $u \in [0, 1]$, $\alpha \in [0, 1]$ with φ from (21) the entropy generating function. Then the corresponding ME quantile function is*

$$Q(u) = \begin{cases} (2\alpha - 1) + (\lambda + 1)\left(\alpha\frac{u^\lambda - 1}{\lambda} - (1-\alpha)\frac{(1-u)^\lambda - 1}{\lambda}\right) & \textit{for } u \in [0, 1],\ \lambda \neq 0 \\ (2\alpha - 1) + \alpha \ln u - (1-\alpha)\ln(1-u) & \textit{for } u \in [0, 1],\ \lambda = 0 \end{cases} \tag{22}$$

with support depending on α and $\lambda > -1/2$

λ	α	*support*
$(-1/2, 0]$	0	$[-1, \infty)$
	1	$(-\infty, 1]$
	$(0, 1)$	$(-\infty, \infty)$
$(0, \infty)$	0	$([1, -1 + (\lambda+1)/\lambda]$
	1	$[1 - (\lambda+1)/\lambda, 1]$
	$(0, 1)$	$[2\alpha - 1 - \alpha(\lambda+1)/\lambda, 2\alpha - 1 + (1-\alpha)(\lambda+1)/\lambda]$

Proof: The claim follows immediately by calculating the derivative Φ' and the fact that $Q(u)) = -\Phi'(u)$, $u \in [0, 1]$. The variance exists for $\lambda > -1/2$. The support is easily verified by calculating $Q(0)$ and $Q(1)$. $\square$

For $\alpha \neq 1/2$, the entropy generating function gives different weight to small and large values of u which leads to a asymmetric ME quantile function determined by the negative of the derivative of Φ. In the case of $\alpha = 1/2$, we get the well-known Tukey λ and the logistic distribution. Therefore, (4.4) can be considered as the quantile function of a skewed Tukey λ distribution or an skewed logistic distribution. Both seems to be a new alternative to the already discussed generalized Tukey λ and the GLO-distribution. But this is not completely true. In his Ph.D. thesis, van Staden (2013) introduce already this alternatives. In his joint paper with King (2014) he investigates the properties of the skewed logistic distribution intensively. Now, we know how the cumulative paired Φ entropy looks like for which this distributions are ME distributions under the constraint of known mean and variance. van Staden (2013) and van Staden & King (2015) derive explicit expressions for the ML estimators for the location, the scale parameter and the parameter α and λ and their asymptotic standard errors. The latter parameters together control skewness and kurtosis. King % van Staden (2014) fit the skewed logistic distribution to a data set from biology.

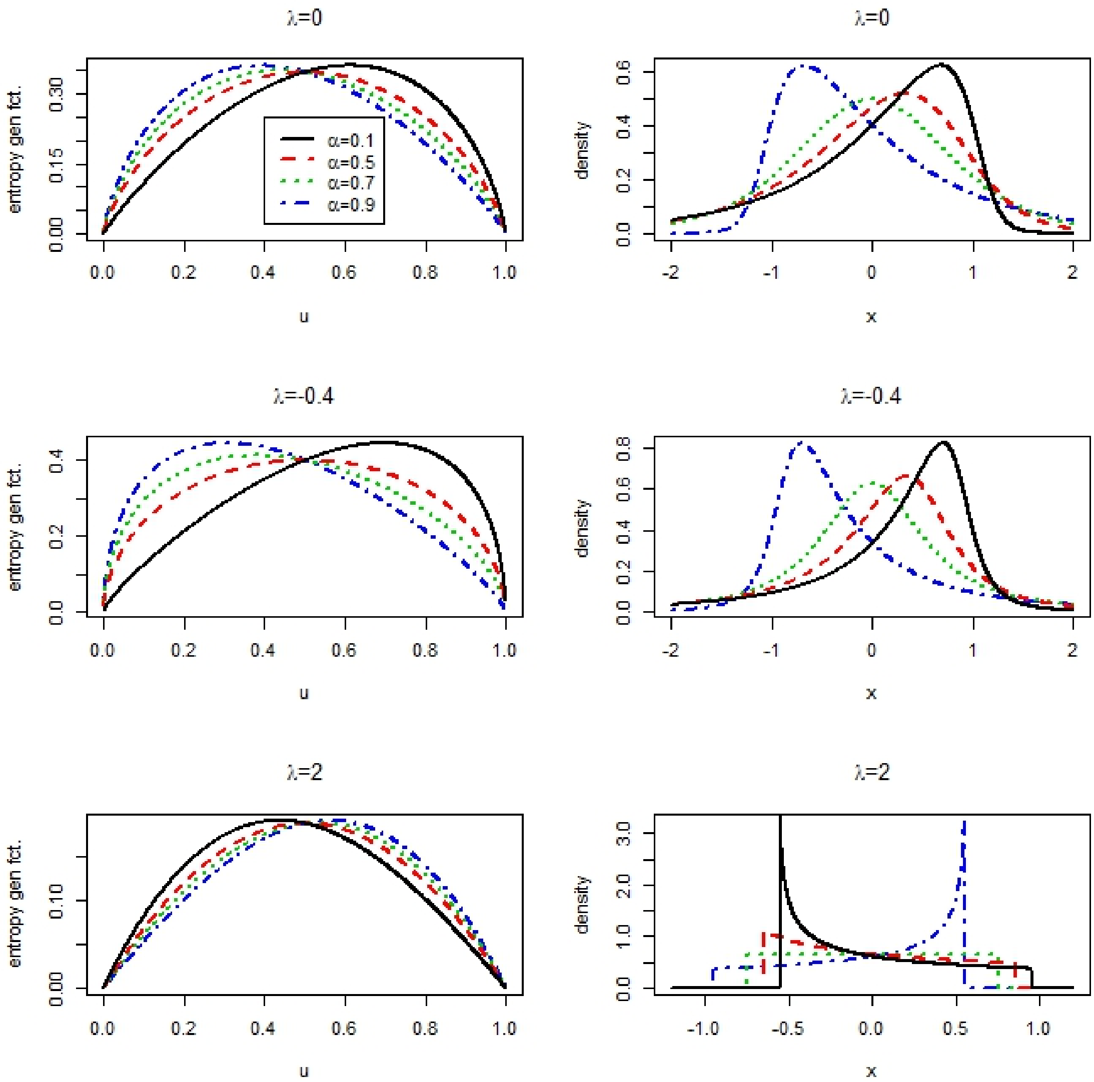

Figure 7: Entropy generating function and density for the skewed logistic ($\lambda = 0$) and the skewed Tukey λ distribution for $\lambda = -0.4, 2$ and $\alpha = 0.1, 0.5, 0.7, 0.9$.

More precisely, van Staden (2013) discuss the quantile function

$$Q(u) == \begin{cases} a + b\left(\alpha\frac{u^\lambda - 1}{\lambda} - (1-\alpha)\frac{(1-u)^\lambda - 1}{\lambda}\right) & \text{for } u \in [0,1],\ \lambda \neq 0 \\ a + b\left(\alpha \ln u - (1-\alpha)\ln(1-u)\right) & \text{for } u \in [0,1],\ \lambda = 0 \end{cases}$$

with a location parameter a and a scale parameter $b > 0$. With this terminology Q belongs to the setting $a = 2\alpha - 1$ and $b = \lambda + 1$.

Figure 7 gives an impression of the different entropy characterizing function Φ and the density of the skewed logistic ($\lambda = 0$) and the skewed Tukey λ distribution ($\lambda = -0.4$. The symmetric logistic and Tukey λ distribution ($\alpha = 1/2$) serve as a point of reference. The parameter setting $\alpha = 0.1$ results in left-skewed logistic and Tukey λ distributions. For $\alpha = 0.7$ and $\alpha = 0.9$, we get a right-skewed distribution. At last, in figure 7, the parameter λ was set to 2.0. This means a compact support and an extremely skewed distribution.

4.5 Fechner approach of skewness

There are many other proposals to introduce a skewness parameter into a symmetric distribution. One proposal can be traced back to Fechner (1897), pp. 295. He proposes to use different values of a scale parameter for the left and the right half of a symmetric distribution. Klein & Fischer (2006) show that such a split of the scale parameter leads to a skewed distribution in the sense that the skewness parameter attains the skewness ordering of van Zwet (1964). Arrelano-Valle et al. (2006) picks up the Fechner proposal and introduce one parameter of skewness γ by considering functions $f_1(\gamma)$ and $f_2(\gamma)$ as different scale parameters for the both halves of a symmetric distribution. This approach includes the proposal made by Fernandez et al. (1995) with $f_1(\gamma) = 1/\gamma$) and $f_2(\gamma) = \gamma$, $\gamma > 0$. Let f be the density being symmetric around 0. Then, a density with skewness parameter γ is given by

$$f(x,\gamma) = \frac{2\gamma}{1+\gamma^2}\left(f(x/\gamma)I(x<0) + f(x\gamma)I(x \geq 0)\right) \tag{23}$$

for $x \in \mathbb{R}$, $\gamma > 0$. The corresponding distribution function is

$$F(x,\gamma) = \begin{cases} 2\gamma^2/(1+\gamma^2)F(x/\gamma, \gamma = 1) & \text{for } x < 0 \\ 2/(1+\gamma^2)(F(0,\gamma=1)(\gamma^2 - 1) + F(x\gamma, \gamma = 1)) & \text{for } x \geq 0 \end{cases} \tag{24}$$

We want to follow the generalized cumulative Φ approach. This means to identify the entropy generating function Φ such that (23) is a ME distribution under the constraint of fixed mean and variance.

Corollary 4.5. *Let f be a density symmetric around* 0 *and Q the corresponding quantile function. Constraints are given by fixing mean and variance. Then, (23) is the density of* *the ME distribution with entropy generating function*

$$\Phi(u) = \begin{cases} -2\gamma^3/(1+\gamma^2)\int_0^u Q(v)dv \\ \text{for } u \le \gamma^2/(1+\gamma^2) \\ -2/(\gamma(1+\gamma^2))\int_{0.5}^{0.5(1+\gamma^2)u+0.5(1-\gamma^2)} Q(z)dz \\ \text{for } u > \gamma^2/(1+\gamma^2) \end{cases}$$

Proof: By inverting the distribution function (24), we get the quantile function

$$Q(u,\gamma) = \begin{cases} \gamma Q((1+\gamma^2)/(2\gamma^2)u, \gamma=1) & \text{for } u < \gamma^2/(1+\gamma^2) \\ 1/\gamma Q((1+\gamma^2)/2u, \gamma=1) + (1-\gamma^2)/2 & \text{for } u \ge \gamma/(1+\gamma^2) \end{cases}$$

Summing up this quantile function leads to the partial mean function

$$\mu(u,\gamma) = \int_0^u Q(v,\gamma)dv = \begin{cases} 2\gamma^3/(1+\gamma^2)\int_0^u Q(v)dv \\ \text{for } u \le \gamma^2/(1+\gamma^2) \\ 2/(\gamma(1+\gamma^2))\int_{0.5}^{0.5(1+\gamma^2)u+0.5(1-\gamma^2)} Q(z)dz \\ \text{for } u > \gamma^2/(1+\gamma^2) \end{cases}$$

The negative of the partial mean function determines the entropy generating function Φ such that (8) is a ME distribution under the constraint of given man and variance. □

As an illustrating example, we consider the skewed normal distribution.

Example 4.1. *Let f_N (F_N^{-1}) be the density (quantile function) of the standard normal distribution. From $\int_0^u F_N^{-1}(v)dv = f_N(F_N^{-1}(u))$, $u \in [0,1]$ we get*

$$\mu(u,\gamma) = \begin{cases} 2\gamma^3/(1+\gamma^2) f_N(F_N^{-1}(u)) \text{ for } u \le \gamma^2/(1+\gamma^2) \\ 2/(\gamma(1+\gamma^2))\left(f_N(F_N^{-1}(0.5((1+\gamma^2)u+1-\gamma^2)) - 1/\sqrt{2\pi}\right) \\ \text{for } u > \gamma^2/(1+\gamma^2) \end{cases}$$

Figure 8 shows the entropy generating function and the density of the skewed normal distribution for different values of the skewness parameter γ. The symmetric standard normal distribution ($\gamma = 1$) serves as a point of reference.

Up to now, we there is no constraint on the ME distribution's support f.e. in the form $Q(0) = 0$. We could always apply theorem 3.1. In the following two subsections, the support of the ME distribution has to be non-negative and the property $Q(0) = 0$ will be required as an additional constraint for the ME task. Due to the non-negativity, it is possible to derive ME distributions under the constraint of a given mean and more general under a given k-th power moment for $k \ge 2$.

4.6 Cumulative residual Shannon entropy and the Weibull distribution

We consider ME distributions with $Q(0) = 0$. We can apply theorem 3.2 to get the ME distribution if the mean μ and the k-th moment are fixed. At first, we consider the

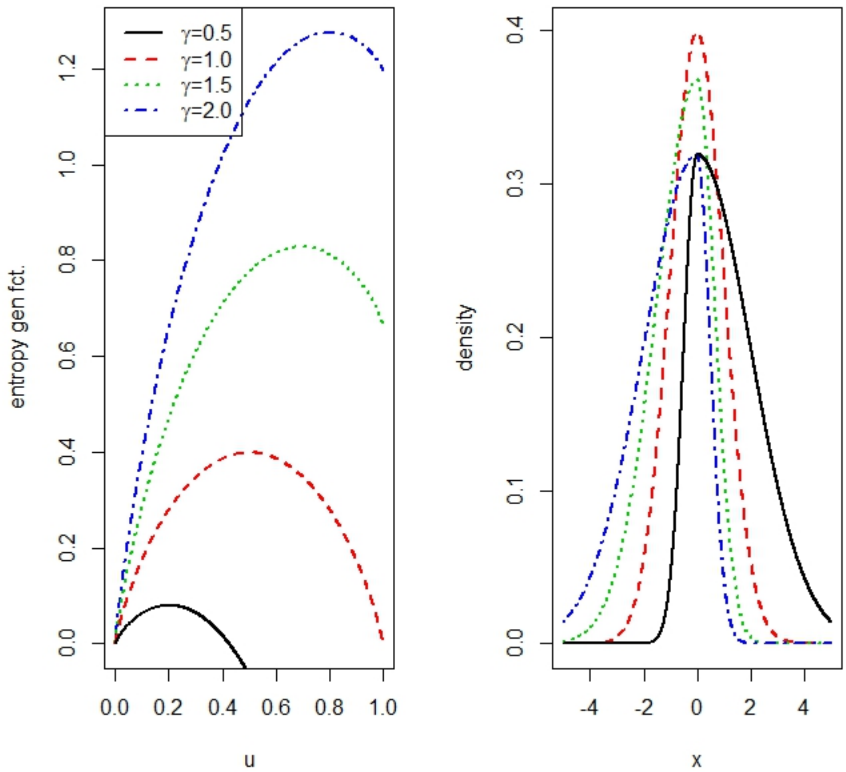

Figure 8: Entropy generating function and density for the skewed normal distribution with parameter values $\gamma = 0.5, 1.0, 1.2, 2$.

entropy generating function $\Phi(u) = -(1-u)\ln(1-u)$, $u \in [0, 1]$ with $\Phi(1) = \Phi(0) = 0$ and $\Phi'(u) = \ln(1-u) - 1$, $u \in [0, 1]$ such that $\Phi'(0) = -1$ and $-\Phi'(u) + \Phi'(0) = -\ln(1-u)$, $u \in [0, 1]$ of the cumulative residual Shannon entropy and receive the Weibull distribution as ME solution. In a modified form, this result is known from Rao (2005).

Corollary 4.6. *Let $\Phi(u) = -(1-u)\ln(1-u)$, $u \in [0, 1]$ the entropy generating function and the constraints are that the mean μ and the k-th power moment μ'_k are fixed. Then the corresponding ME distribution is given by a Weibull distribution with scale parameter $\lambda = \mu/\Gamma(1 + 1/(k-1))$ and shape parameter $r = (k-1)$ if*

$$\frac{\mu'_k}{\mu^k} = \frac{\Gamma(1 + k/(k-1))}{\Gamma(1 + 1/(k-1))^k}. \quad \square$$

Proof: The Weibull distribution is defined by a quantile function

$$Q_W(u) = \lambda(-\ln(1-u))^{1/r}, \quad u \in [0, 1], \; r > 0.$$

The support is given by $[Q_W(0, \gamma) = 0, Q_W(1, \gamma) = \infty)$ and the mean is $\lambda\Gamma(1+1/r)$. From

(12) we get the quantile function

$$Q(u) = \frac{(-\ln(1-u))^{1/(k-1)}}{\int_0^1 (-\ln(1-u))^{1/(k-1)} du} \mu, \quad u \in [0,1].$$

This means that $Q(u)$ is proportional to the quantile function Q_W of a Weibull distribution with shape parameter $r = (k-1)$. With

$$\int_0^1 (-\ln(1-u))^{1/(k-1)} du = \Gamma\left(1 + 1/(k-1)\right)$$

and $\lambda \equiv \mu/\Gamma(1 + 1/(k-1))$ we get the wanted quantile function of a Weibull distribution with scale parameter λ and shape parameter $r = (k-1)$. It is easy to verify that

$$\frac{\mu'_k}{\mu^k} = \frac{\Gamma(1 + k/(k-1))}{\Gamma(1 + 1/(k-1))^k}.$$

For $k = 2$ we get an exponential distribution with scale parameter $\lambda = \mu$ if $\mu'_2/\mu^2 = \Gamma(1+2)/\Gamma(2)^2 = 2$.

It is an obvious task to substitute the generating function of the cumulative residual Shannon entropy by the more general generating function of the cumulative residual Havrda-Charvát entropy. This will lead to a generalized Weibull distribution, also known as extended Burr XII distribution) as will be shown in the next subsection.

4.7 Cumulative residual Havrda-Charvát entropy and the extended Burr XII distribution

The entropy generating function of the cumulative residual Shannon entropy can be generalized to

$$\Phi(u) = (1-u)((1-u)^{\alpha-1} - 1)/(1-\alpha), \quad u \in [0,1], \ \alpha \neq 1. \tag{25}$$

Again we get $\Phi(1) = \Phi(0) = 0$. The derivative Φ' is

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

such that $\Phi'(0) = 1$ and

$$-\Phi'(u) + \Phi'(0) = \frac{1}{1-\alpha}\left(\alpha((1-u)^{\alpha-1} - 1)\right), \quad u \in [0,1].$$

Now, we look for the corresponding ME distribution if $Q(0) = 0$ and μ, μ'_k are fixed. The following corollary states that this ME distribution generalizes the Weibull distribution. For the sake of simple notation, we introduce a two-parametric generalization of the complete Γ function as

$$\Gamma_2(r, \lambda) = \int_0^1 \left(-\frac{u^\lambda - 1}{\lambda}\right)^{r-1} du, \quad r > 0.$$

Notice that $\Gamma_2(r,\lambda) \to \Gamma(r)$ for $\lambda \to 0$.

Corollary 4.7. *Let $\Phi(u) = (1-u)((1-u)^{\alpha-1}-1)/(1-\alpha)$, $u \in [0,1]$, $\alpha \neq 1$ the entropy generating function and the constraints are that the mean μ and the k-the power moment μ'_k are fixed. Then the corresponding ME distribution is given by the quantile function*

$$Q(u) = \frac{\left(\frac{1}{1-\alpha}\left((1-u)^{\alpha-1}-1\right)\right)^{1/(k-1)}}{\Gamma_2(1+1/(k-1),\alpha-1)}\mu \quad u \in [0,1],\ \alpha \neq 1 \tag{26}$$

if the following relation holds between μ and μ'_k:

$$\frac{\mu'_k}{\mu^k} = \frac{\Gamma_2(1+k/(k-1),\alpha-1)}{\Gamma_2(1+1/(k-1),\alpha-1)^k}. \tag{27}$$

Proof: With

$$\int_0^1 \left(\frac{1}{1-\alpha}\left(\alpha((1-u)^{\alpha-1}-1)\right)\right)^{1/(k-1)} du = \alpha^{1/(k-1)}\Gamma_2(1+1/(k-1),\alpha-1)$$

the assertion follows immediately. □

Mudholkar et al. (1996) introduce the quantile function (26) with a different parametrization and call the corresponding distribution "generalized Weibull distribution". Due to the fact that the Burr XII distribution (cf. Burr (1942)) is a special case (in our parametrization for $\alpha = 2$), Shao et al. (2004) use the term "extended Burr XII distribution" (EBXII). A recent paper, discussing this distribution, is Ganora & Laio (2015). They are concerned with parameter estimation (see also Shao (2004), Nadaraja & Kotz (2006), Usta, (2013)). Applications are lifetimes of devices with a bathtub hazard rate (cf. Mudholkar et al. (1996)) and flood frequency and duration analysis (cf. Shao et al. (2004), Nadarajah & Kotz (2006), Ganora & Laio (2015)).

The EBXII density is given by

$$f(x) = \frac{k-1}{\lambda}\frac{(x/\lambda)^{k-2}}{(1+(1-\alpha)(x/\lambda)^{k-1})^{1-1/(\alpha-1)}}, \quad x > 0,$$

with $\lambda = \mu/\Gamma_2(1+1/(k-1),\alpha-1)$, $k > 1$, $\alpha \neq 1$. The support is $(0, \lambda(1/(\alpha-1)^{1/(k-1)})$ for $\alpha > 1$ and $(0,\infty)$ for $\alpha < 1$).

Figure 9 shows the entropy generating function (25) as well as ME densities under the constraint of a fixed power moment μ'_k for $k = 2, 4, 8$. The setting $k = 2$ leads to density with a shape similar to an exponential distribution. The constraint (27) means a fixed coefficient of variation (cf. Rao et al. (2004)).

5 Estimating the entropy generating function

Can we learn something from data about the entropy generating function Φ for which the data generating distribution is a ME distribution under the constraint of given mean and

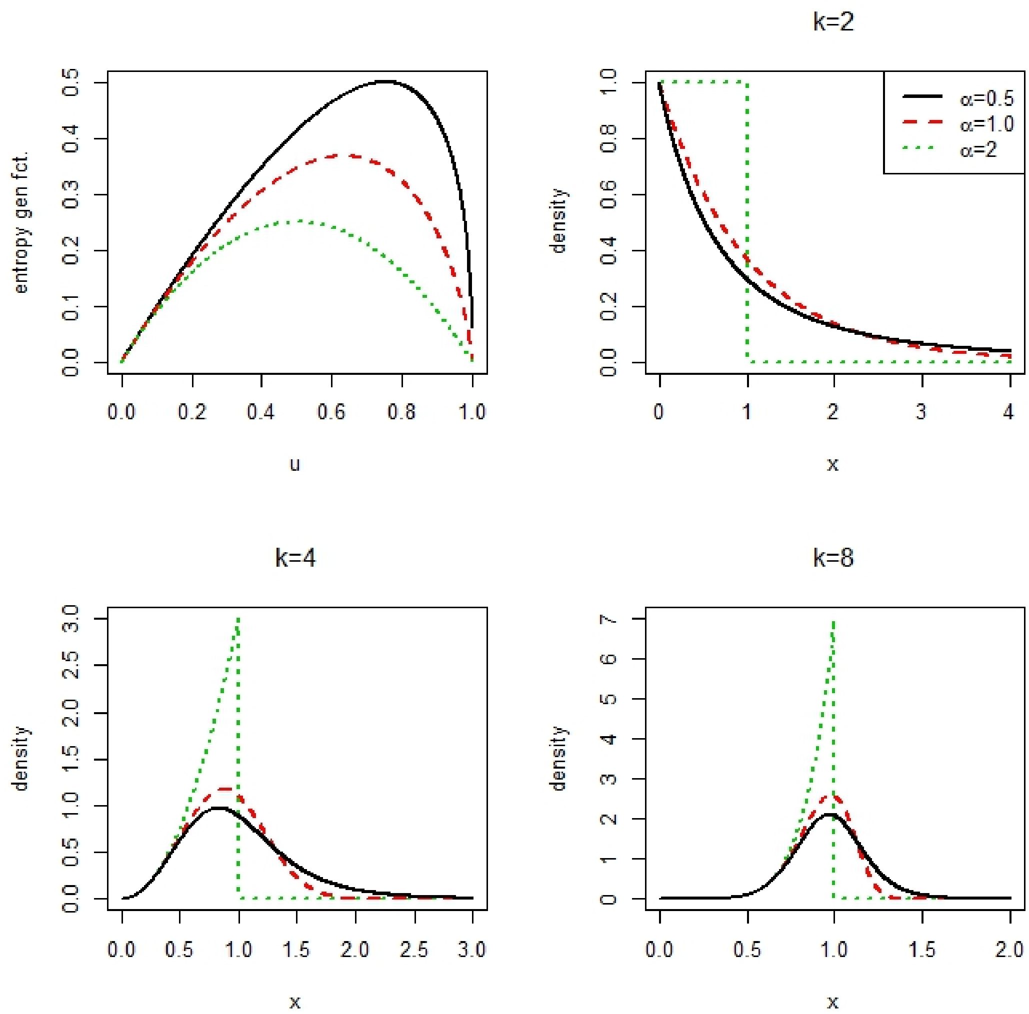

Figure 9: Entropy generating function and extended Burr XII density for $\alpha = 0.5, 1, 2$ and $k = 2, 4, 8$.

variance? The entropy generating function Φ is given by the partial mean function

$$\mu(u) = uE[X|X \leq Q(u)] \;\; u \in [0, 1].$$

Therefore, it is obvious to estimate this partial mean function to get an estimator for Φ.

Let $X_1, \dots, X_n$ be identical and stochastic independent distributed random variables. $X_{(n:1)}, \dots, X_{(n:n)}$ denote the corresponding sequence of order statistics. For a fixed value $u \in [0, 1]$ such that $nu \in \{1, 2, \dots, n\}$ we consider an estimator of the form

$$\hat{\Phi}(u) = -\hat{\mu}(u) = -u \frac{1}{nu} \sum_{i=1}^{nu} X_{(n:i)}, \;\; nu \in \{1, 2, \dots, n\}$$

for the entropy generating function Φ.

We would like to demonstrate the usefulness of this estimator by an example.

Example 5.1. *The first data set consists of the S&P500 standardized daily logarithmic returns from 10-05-2012 to 10-04-2017* [2]*. This gives 1256 data points. We have to notice that the mean and the variance are fixed to the values* 0 *and* 1*. In figure 10, we compare the estimated entropy generating function* $(-\hat{\mu}(u))$ *with the entropy generating functions of the standardized* t *distribution with* 4 *degrees of freedom and the standard normal distribution. Standardizing gives also the mean value* 0 *and the variance* 1 *for the* t *distribution. The entropy generating function of the* t *distribution must be calculated by numerical integration. We choose the number of degrees of freedom by trail and error. But, ML estimation gives a value not far away from* 4.

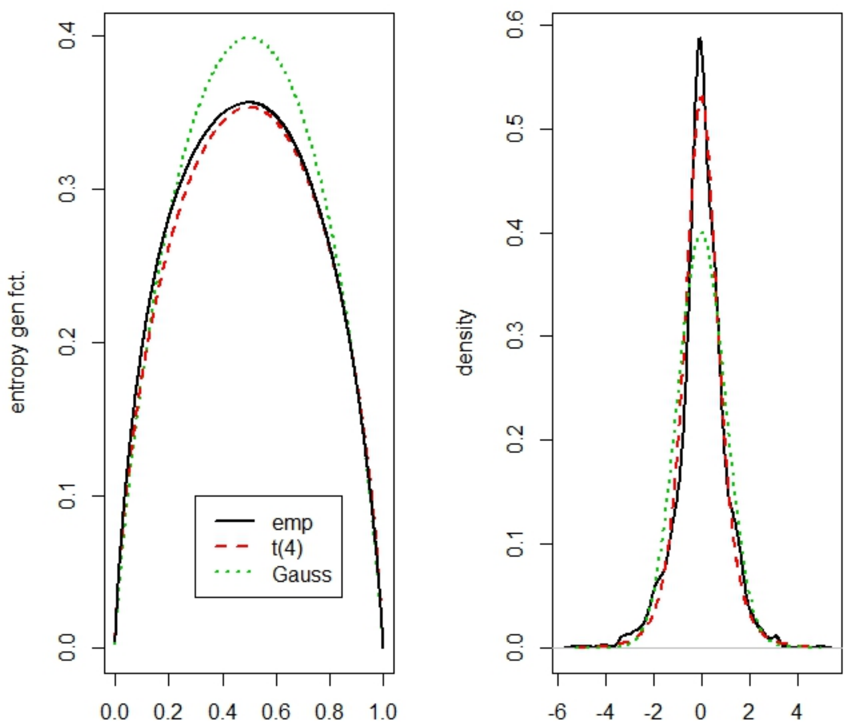

Figure 10: Estimated entropy generating function and estimated density for the S&P500 standardized daily log returns from from 10-05-2012 to 10-04-2017.

We know from (15) that for a non-negative random variable with $Q(0) = 0$ and fixed mean μ and fixed k-th power moment μ'_k the entropy generating function Φ is given by

$$\Phi(u) = \mu'_{k-1}u - \mu_{k-1}(u) = \mu'_{k-1}u - uE[X^{k-1}|X \leq Q(u)].$$

[2]The data is available from https://fred.stlouisfed.org/series/SP500

For this entropy generating function. Q is a ME quantile function.

To get an estimator for Φ. it is only necessary to estimate the $(k-1)$-th power mean μ'_{k-1} and the partial $(k-1)$-th power mean function $\mu_{k-1}(u)$. For a fixed value $u \in [0,1]$ (such that $nu \in \{1,2,\ldots,n\}$), a natural estimator for the partial $(k-1)$-th power mean function is

$$\hat{\mu}_{k-1}(u) = u\frac{1}{nu}\sum_{i=1}^{nu} X_{(n:i)}^{k-1}, \quad nu \in \{1,2,\ldots,n\}.$$

An estimator for the entropy generating function Φ is given by

$$\hat{\Phi}(u) = u\left(\frac{1}{n}\sum_{i=1}^{n} X_{(n:i)}^{k-1} - \frac{1}{nu}\sum_{i=1}^{nu} X_{(n:i)}^{k-1}\right), \quad nu \in \{1,2,\ldots,n\}.$$

We want to demonstrate that this estimator works really good for a real data set and the Weibull distribution. To do this, we need the partial $(k-1)$-th power mean function for the Weibull distribution with shape parameter r and scale parameter λ. For this distribution, it holds

$$\mu_{k-1}(u) = \int_0^u (\lambda(-(1-v)^r))^{k-1}\,dv = \lambda^{k-1}\Gamma\left(\frac{k-1}{r}+1\right)\Gamma\left(-\ln(1-v);\frac{k-1}{r}+1,1\right)$$

for $u \in [0,1]$. $\Gamma(x;a,b)$ denotes the distribution function of a Γ distribution with shape parameter a and scale parameter β. The corresponding entropy generating function Φ such that this Weibull distribution is CE_Φ maximal under $Q(0)=0$ and the constraints of fixed mean μ and fixed k-the power moment is

$$\Phi(u) = u\lambda^{k-1}\Gamma\left(1+\frac{k-1}{r}\right)\left(1-\Gamma\left(-\ln(1-u);\frac{k-1}{r}+1,1\right)\right), \quad u \in [0,1].$$

k determines the shape parameter r by the relation

$$\frac{\mu'_k}{\mu^k} = \frac{\Gamma(1+k/r)}{\Gamma(1+1/r)^k}.$$

Example 5.2. *Let X be a random variable representing the duration in days between two explosions in the mines of a specific region. From Schlittgen (1996), we get the following data set with the duration between 41 mine explosions:*

378	36	15	31	215	11	137	4	15	72
96	124	50	120	203	176	55	93	59	315
59	61	1	13	189	345	20	81	286	114
108	188	233	28	22	61	78	99	326	275

We set $k=2$. This means that for every potential ME distribution $\mu'_2/\mu^2 = 1.762$ has to hold. This implies $r = 1.148$ for the shape parameter r of the Weibull distribution. In figure 11, the estimated entropy generating function is compared with $\mu_{k-1}(u)$ for this

Weibull distribution. The fit seems to be rather good in view of the relative small sample size.

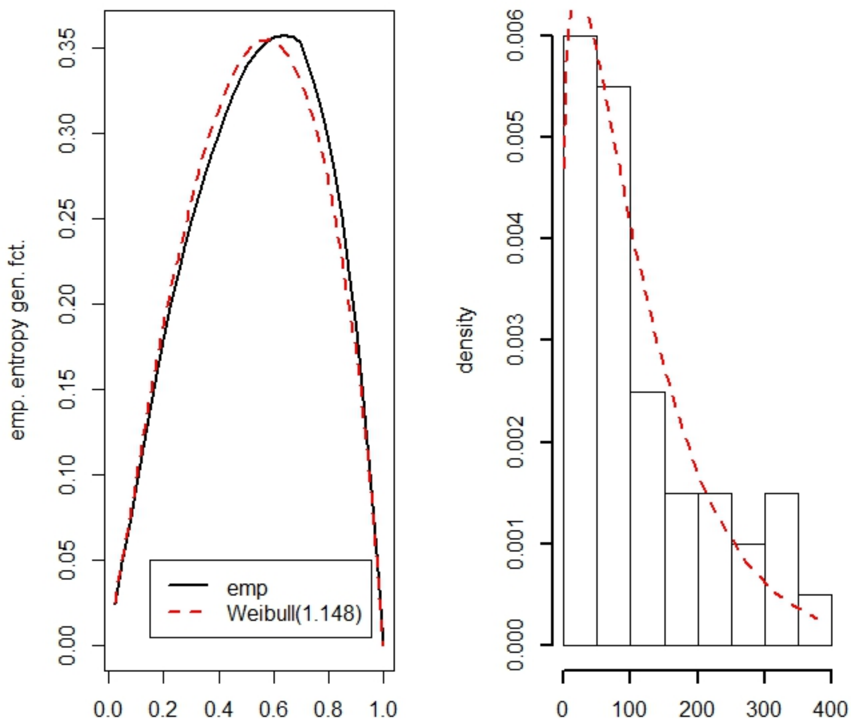

Figure 11: Estimated entropy generating function and estimated density for a data set with time intervals between 41 mine explosions.

Further work will be to estimate the number of degrees or parameters of other flexible distributions by minimizing the distance between the easy to calculate empirical entropy generating function, and the entropy generating function of the distribution, we suppose the data could be generated from. The advantage of this procedure could be that the empirical entropy generating function is rather smooth. Therefore, minimizing the distance between the entropy generating functions could be more accurate than considering the distance between the empirical quantile functions, a density estimator or the empirical distribution functions and the corresponding theoretical counterpart. But this will investigated in future research.

6 Conclusion

Maximizing Shannon's differential entropy under different moment constraints is a well-known task. Without any constraint, the differential entropy will be maximized by an uniform distribution representing the situation of no information. Instead of the differential entropy, a cumulative entropy can be used. This entropy depends either on the distribution (direct) or on the survival function (residual) or on both (apired). Maximizing this entropy without any constraint gives a extremely bi-modal (= bipolar) distribution. This distribution represents so-called contradictory information since an event and the complement can happen with equal probability. In this situation, it is extremely hard to make a forecast, even harder than for an uniformly distributed random variable. Under the constraint of fixed mean and variance, maximizing the cumulative entropy tries to transform a distribution in the direction of a bipolar distribution as far as it is allowed by the constraints. For so-called cumulative paired, entropies and the constraints that mean and variance are known, solving the maximizing problem lead to symmetric ME distributions like logistic and the Tukey λ distribution (cf. Li & Liu (2008), Klein et al. (2016)). Other ME distributions were found for the cumulative paired Leik and Gini entropy (cf. Klein (2016 et al., Dai & Chen (2012), Dai (2017)). We generalize the cumulative paired entropy in several ways. First, we introduce the cumulative paired Mielke(r) entropy and derive the ME distributions. The results already known for the cumulative paired Leik and Gini entropy are included for $r = 1$ and $r = 2$. Then, we turn around the research question, consider the generalized maximum entropy approach and derive the entropy generating function such that a pre-specified skewed distribution is a ME distribution. The generalized λ distribution serves as an example. The generalized logistic distribution (GLO) is another example for an ME distribution. Starting with a natural generalization of the derivative of the entropy generating function known from the logistic distribution, we arrive at the GLO immediately. Considering a linear combination of entropy generating functions leads to new ME distributions with skewness properties. Fechner's proposal to define different values of a scale parameter for both halves of a distribution also gives skewed distributions. Again, we derive the corresponding entropy generating function. The skewed normal distribution serves as an illustrative example. In the literature, the ME task was considered mainly for lifetime distributions with the special property that the support is $[0, \infty)$, and it holds $Q(0) = 0$. Under this additional constraint, we derive the ME distribution for fixed mean and k-the power moment. This gives the extended Burr XII distribution. At last, we propose an estimator for the cumulative entropies generating function representing all the properties of the underlying ME data generating distribution. The usefulness of this estimator can be demonstrated for two real data sets.

References

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

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

3. Arellano-Valle R.B., Gomez H.W. & Quintana, F.A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* **128**, 427- 443.

4. Asadi, M. & Zohrevand, Y. (2007). On the dynamic cumulative residual entropy. *Journal of Statistical Planning and Inference* **137**, 1931-1941.

5. Baratpour, S. (2010). Characterizations based on cumulative residual entropy of first-order statistics. *Communications in Statistics - Theory and Methods* **39**, 3645-3651.

6. Baratpour, S. & Habibi, Rad, A. (2012). Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. *Communications in Statistics - Theory and Methods* **39**, 3645-3651.

7. 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

8. Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics* **13**, 215–232.

9. Chalabi, Y., Scott, J. & Wuertz, D. (2012). Flexible distribution modeling with the generalized lambda distribution. https://mpra.ub.uni-muenchen.de/id/eprint/43333.

10. Chamany, A. & Baratpour, S. (2014). A dynamic discrimination information based on cumulative residual entropy and its properties. *Communications in Statistics - Theory and Methods* **43**, 1041-1049.

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

12. Cressie, N. A. & Pardo, L. (2002). Phi-divergence statistic. In A. H. El-Shaarawi & W. W. Piegorsch (Eds.), *Encyclopedia of Environmetrics, Vol 3*, Wiley, New York, 1551-1555.

13. Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten. *Magyar. Tud. Akad. Mat. Kutato Int. Kozl.* **8**, 85–108.

14. Dai, W. (2017). Quadratic entropy of uncertain variables *Soft Computing* doi/10.1007/s00500-017-2602-y.

15. Dai, W. & Chen X (2012) Entropy of function of uncertain variables. *Mathematical and Computer Modelling* **55**, 754-760.

16. De Luca, A. & Termini, S. (1972) A definition of a nonprobabilistic entropy measure of fuzzy entropy. *Information and Control* **20**, 301–312.

17. Di Crescenzo, A. & Longobardi, M. (2009a) On cumulative entropies and lifetime estimation. In: *Methods and Models in Artificial and Natural Computation* Mira, J.M. et al.; Eds. IWINAC Part I, LNCS 5601. Springer-Verlag, Berlin, Germany, 132-141.

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

19. Drissi, N., Chonavel, T. & Boucher, J.M. (2008) Generalized cumulative residual entropy distributions with unrestricted supports. *Research Letters in Signal Processing* **11**, 1–5.

20. Fechner, G.T (1897). *Kollektivmaßlehre.* Engelmann, Leipzig.

21. Fernandez C., Osiewalski J. & Steel, M.F.Jj. (1995): Modelling and inference with ν-spherical distributions. *Journal of American Statistical Association* **90**, 1331-1340.

22. Freimer, M., Kollia, G, Mudholkar, G.S. & Lin, C.T. (1988). A study of the generalized Tukey lambda family. *Communications in Statistics - Theory and Methods* **17**, 3547-3567.

23. Ganora, D. & Liao, F. (2015). Hydrological applications of the Burr distribution: Practical method for parameter estimation. *Journal of Hydrologic Engineering* **20**, doi/10.1061/%28ASCE%29HE.1943-5584.0001203.

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

25. Hosking, J.R.M. (2007). Distributions with maximum entropy subject to constraints their L-moments or expected order statistics. *Journal of Statistical Planning and Inference* **137**, 2870-2891.

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

27. Jaynes, E.T. (1987b) Information theory and statistical mechanics. II. *Physical Review* **108**, 171-190.

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

29. Kapur, J.N. (1994). *Measure of Information and their Applications.* Wiley Eastern Limited, New Delhi.

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

31. King, R.A.R. & MacGillivray, H.L. (1999). A starship estimation method for the generalized λ distributions. *Australian and New Zealand Journal of Statistics* **41**, 353-374.

32. Klein I. & Fischer, M. (2006). Skewness by splitting the scale parameter. *Communications in Statistics, Theory and Methods* **35**, 1159-1171.

33. Klein, I., Mangold, B. & Doll, M. (2016). Cumulative paired ϕ entropies. *Entropy* **18**, 248.

34. Kumar, V. (2017). Characterization results based on dynamic Tsallis cumulative residual entropy. *Communications in Statistics - Theory and Methods* **46**, 8343-8354.

35. Kumar, V. & Taneja. H.C. (2011). Some characterizations results on generalized cumulative residual entropy measures. *Statistics and Probability Letters* **81**, 1072-1077.

36. Lai, C.-D. (2014). *Generalized Weibull Distributions.* Springer, New York,

37. Li, P. & Liu, B. (2008) Entropy of credibility distributions for fuzzy variables. *IEEE Transactions on Fuzzy Systems* **16**, 123–129.

38. Liese, F.; Vajda, I. (2006). On divergences and information in statistics and information theory. *IEEE Transactions on Information Theory* **52**, 4394–4412.

39. Liu B. (2007) A survey of entropy of fuzzy variables. *Journal of Uncertain Systems* **1**, 4–13.

40. Liu, B. (2015). *Uncertainty Theory.* Springer, New York.

41. Mielke, P.W. Jr. (1972). Asymptotic beavior of two-sample tests based on powers of ranks for detecting scale and location alternatives. *Journal of the American Statistical Association* **67**, 850-854.

42. Mielke, P.W. Jr. & Johnson, E.S. (1973). Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests. *Monthly Weather Review* **101**, 701–707.

43. Minimol, S. (2017). On generalized dynamic cumulative past entropy measures. *Communications in Statistics - Theory and Methods* **46**, 2816-2822.

44. Mirali, M, Baratpur, S. & Fakoor, V. (2017). ON weighted cumulative residual entropy. *Communications in Statistics - Theory and Methods* **46**, 2857-2869.

45. Morais, A.L. & Cordeiro, G.M. (no year). A four-parameter generalized logistic distribution. http://www.ime.unicamp.br/sinape/sites/default/files/Lemos-Cordeiro.pdf.

46. Morimoto, T. (1963). Markov processes and the H-theorem. *Journal of the Physical Society of Japan* **18**, 328–331.

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

48. Mudholkar, G.S, Srivastava, D.K. & Kollia, G.D. (1996). A generalization of the Weibull distribution with applications to the analysis of survival data. *Journal of the American Statistical Association* **91**, 1575-1583.

49. Nadarajah, S. & Kotz, S. (2006). Discussion of models for extremes using the extended three parameter Burr XII system with applications to the flood frequency analysis. *Hydrological Sciences Journal* **51**, 1203-1204.

50. Nassar, M.M. & Elmasry, A. (2012). A study of the generalized logistic distribution.s. *Journal of the Egyptian Mathematical Society* **20**, 126-133.

51. Navarro, J., del Aguila, Y. & Asadi, M. (2010). Some new results on the cumulative residual entropy. *Journal of Statistical Planning and Inference* **136**, 3659-3674.

52. Navarro, J. & Psarrakos, G. (2017). Characterizations based on generalized cumulative residual entropy functions. *Communications in Statistics - Theory and Methods* **46**, 1247-1260.

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

54. Park, S.Y. & Bera, A.K. (2008). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics* **150**, 219-230.

55. Psarrakos, G. & Navarro, J. (2013). Generalized cumulative residual entropy and record values. *Metrika* **76**, 623-640.

56. Rajesh, G. & Sunoj, S.M. (2016). Some properties of cumulative Tsallis entropy of order α. *Statistical Papers* https://doi.org/10.1007/s00362-016-0855-7

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

58. 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.

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

60. Sankaran, P.G. & Sunoj, S.M. (2017). Quantile-based cumulative entropies. *Communications in Statistics - Theory and Methods* **46**, 805-814.

61. Schlittgen, R. (1996). Einführung in die Statistik. Oldenbourg-Verlag, München.

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

63. Shabri, A. Jemain, A.A. (2011). Fitting the generalized logistic distribution by LQ-Moments. *Applied Mathematical Sciences* **5**, 2663 - 2676

64. Shao, Q. (2004). Notes on the maximum likelihood estimation for the three-parameter Burr XII distribution. *Computation Statistics and Data Analysis* **45**, 675-687.

65. Shao, Q., Wong, H. Xia, J. & Pp, W. (2004). Models for extremes using the extended three-parameter Burr XII system with application to flood frequency analysis. **Hydrological Sciences Journal 49**, 685-702.

66. Su, S. (2007). Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R. *Journal of Statistical Software* **21**, 1-17.

67. Su, S. (2012). **GLDEX:** Fitting single and mixture of generalized lambda distribution (RS and FMKL) using discretized and maximum likelihood methods. R package version 1.044. URL http://CRAN.R-project.org/.

68. Tripathi, Y.M., Mahto, A.K. & Dey, S. (2017). Efficient estimation of the PDF and the CDF of a generalized logistic distribution. *Annals of Data Science* **4**, 63-81.

69. Tsallis, C. (1988). Possible generalization of Boltzmann.Gibbs statistics. *Journal of Statistical Physics* **52**, 479-487.

70. Usta, I. (2013). Different estimation methods for the parameters of the extended Burr XII distribution. *Journal of Applied Statistics* **40**, 397-414.

71. van Staden, P.J. (2013). *Modeling of Generalized Families of Probability Distributions in the Quantile Statistical Universe. Ph.D.thesis* Department of Statistics, University of Pretoria, Pretoria, South Africa.

72. van Staden, P.J. & Loots, M.T. (2009). Method of L moment estimation for the generalized lambda distribution. In: *Proceedings of the Third Annual Applied Statistics and Research Collaboration Research Conference* University of Newcastle, Australia.

73. van Staden, P.J. & King, R.A.R. (2014). The quantile-based skew logistic distribution. *Statistics and Probability Letters* **96**, 109-116.

74. van Zwet, W.R. (1964). *Convex Transformations of Random Variables.* Mathematical Centre Tracts No. 7. Mathematical Centre, Amsterdam

75. Verhoeff, J. (1984). *Recreatieve Informatica. Speech on the Occasion of the 21th Dies Natalis of the Erasmus University Rotterdam*, Rotterdam.

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