

## Article

# (Generalized) Maximum Cumulative Direct, Residual, and Paired $\Phi$ Entropy Approach

Ingo Klein <sup>1,\*</sup> and Monika Doll <sup>2</sup>
<sup>1</sup> Department of Statistics and Econometrics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany

<sup>2</sup> GfK SE, Nordwestring 101, 90419 Nürnberg, Germany; monika.doll@fau.de

\* Correspondence: ingo.klein@fau.de

Received: 18 December 2019; Accepted: 9 January 2020; Published: 12 January 2020

**Abstract:** A distribution that maximizes an entropy can be found by applying two different principles. On the one hand, Jaynes (1957a,b) formulated the maximum entropy principle (MaxEnt) as the search for a distribution maximizing a given entropy under some given constraints. On the other hand, Kapur (1994) and Kesavan and Kapur (1989) introduced the generalized maximum entropy principle (GMaxEnt) as the derivation of an entropy for which a given distribution has the maximum entropy property under some given constraints. In this paper, both principles were considered for cumulative entropies. Such entropies depend either on the distribution function (direct), on the survival function (residual) or on both (paired). We incorporate cumulative direct, residual, and paired entropies in one approach called cumulative  $\Phi$  entropies. Maximizing this entropy without any constraints produces an extremely *U*-shaped (=bipolar) distribution. Maximizing the cumulative entropy under the constraints of fixed mean and variance 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, which is in contrast to minimum or no information. In the literature, to date, only a few maximum entropy distributions for cumulative entropies have been derived. In this paper, we extended the results to well known flexible distributions (like the generalized logistic distribution) and derived some special distributions (like the skewed logistic, the skewed Tukey  $\lambda$  and the extended Burr XII distribution). The generalized maximum entropy principle was applied to the generalized Tukey  $\lambda$  distribution and the Fechner family of skewed distributions. Finally, cumulative entropies were estimated such that the data was drawn from a maximum entropy distribution. This estimator will be applied to the daily S&P500 returns and time durations between mine explosions.

**Keywords:** cumulative entropy; maximum entropy distribution; generalized Tukey  $\lambda$  distribution; generalized logistic distribution; skewed logistic distribution; skewed Tukey  $\lambda$  distribution; skewed normal distribution; Weibull distribution; extended Burr XII distribution

## 1. Introduction

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

$$E_S(f) = - \int f(x) \ln f(x) dx. \quad (1)$$

Maximizing (1) with respect to  $f$  under the constraint of observed power or  $L$ -moments gives maximum entropy (ME) densities (see e.g., [1,2]). This ME solution represents a distributional model which is compatible with the minimum information given by the fixed constraints. The task of deriving ME densities is important, as they are the only reasonable distribution to use for estimation, as lower entropy distributions would mean to assume information that we do not possess. However,

(1) has some shortcomings in order to be a good measure of information, as it could, for example be negative for special densities [3]. Nevertheless, much of the literature has been concerned with the ME task, not only for classical differential entropy, but also for the cumulative residual, the cumulative, cumulative paired entropy, and further entropies with modifications of the generating functions. Thus, in the following section, a short literature review is given.

Substituting the density by the survival function in (1) leads to the cumulative residual (Shannon) entropy. Rao et al. [4] were the first who discussed this new entropy. The discussion occurred in the context of reliability theory, where only random variables with non-negative support are of importance while the survival function is the natural distributional concept. The solution of the maximum entropy task under power moment constraints has already been discussed by [4,5]. They used the log-sum inequality to derive the ME solution instead of the usual approach based on the Lagrange–Euler equations. The exponential and more generally, the Weibull distribution, are solutions to these special ME tasks. In the following years, several authors [6–17] also focused on cumulative residual entropies. In particular, Drissi et al. [10] were concerned with the ME problem. They considered random variables with support  $\mathbb{R}$  and derived the logistic distribution as the ME solution under the additional constraint that the ME solution has to be symmetric. DiCrescenzo [18] applied (1) to the distribution function and called the result ‘cumulative entropy’. Based on early results of DeLuca and Pal [19,20] in the fuzzy set theory concerning membership functions, Li et al. [21] defined a further entropy concept for so-called uncertainty variables which are similar but not identical to random variables. The main idea here is to consider, in (1), the distribution function as well as the survival function. The obvious corresponding ME task was discussed by [22–24]. There are many studies in the literature concerned with the generalization of (1). Thus, some authors [25–27] modified the entropy generating function. General generating functions were considered by [28–30] for the definition of the related concept of  $f$ -divergence [31]. Sometimes  $f$ -divergences are also called  $\phi$ -divergences [32]. Zografos [33] generalized the cumulative residual Shannon entropy in a similar way. Klein et al. [34] combined the ME task known from uncertainty theory with the use of general entropy generating functions. They derived the Tukey  $\lambda$  distribution as an ME distribution when the entropy generating function of [25] is applied together with the distribution and the survival function. They introduced the term ‘cumulative paired entropy’ analogue to the paired  $\phi$  entropy introduced by [35]. Recent publications [36,37] applied the Havrda–Charvát approach to the survival function under the name ‘Cumulative Tsallis entropy’ of order  $\alpha$ . This term refers to the famous paper of [38], where he gives a physical foundation for the Havrda–Charvát approach.

In this paper, the first research question was to clarify what kind of information cumulative entropies really measure. Therefore, we introduce the concept of ‘contradictory information’ in contrast to ‘no information’. As a second research question, we want to unify the diverse approaches of cumulative entropies and their maximization. For this purpose, general cumulative  $\Phi$  entropies will be introduced. All known variants of cumulative entropies are special cases of this general class. Then, after deriving two general formulas for ME quantile functions under some moment restrictions, we apply these formulas to derive ME distributions for new cumulative entropies (like the cumulative Mielke(r) entropy) as well as to identify the cumulative entropy for some flexible families of distributions that allow for skewness (like the generalized Tukey  $\lambda$  or the generalized logistic distribution). As a byproduct, we find some new families of distributions (like a special skewed Tukey  $\lambda$  distribution and a generalized Weibull distribution). The results are summarized in a table and discussed in detail in the Appendix A. The last research question starts with observed data and tries to estimate the cumulative entropy in such a way that the data come from the corresponding ME distribution. This gives an alternative to non-parametric estimation of density functions or distribution functions.

This paper is organized in line with these research questions. Section 2 starts with the discussion of contradictory information and cumulative entropies in principle. In Section 3, we introduce general cumulative  $\Phi$  entropies and prove general results for ME distributions for cumulative  $\Phi$  entropies

under different constraints. In Section 4, we propose an estimator for the ME generating function. Finally, we apply this estimator to real datasets. In the Appendix A, we apply the theoretical results to seven families of cumulative  $\Phi$  entropies (MaxEnt) or families of distributions (GMaxEnt).

## 2. What does Maximizing Cumulative Direct, Residual, and Paired Shannon Entropies Mean?

In this section, we first discuss the concept of ‘contradictory information’ in contrast to no or minimum information and determine that contradictory information corresponds with  $U$ -shaped/bipolar distributions. Then, we learn that maximizing cumulative paired entropies best reflects this situation by comparing the results with those of maximizing differential, cumulative residual, and cumulative direct entropies. Next, we see that the cumulative residual and cumulative direct entropies do not correspond to a  $U$ -shaped distribution if the support of a random variable is only non-negative. Overall, in this section, the focus is on Shannon entropies. However, all insights can be transferred to arbitrary cumulative entropies immediately.

The traditional ME approach starts with the result that the uniform distribution has minimum information (= maximum entropy) under the constraint that the area under the density sums up to one. However, there is another concept of maximum entropy in fuzzy set [19] and uncertainty theory [24,39]. Transferring this concept to probability theory, maximum uncertainty represents the fact that an event  $A$  with probability  $0 < P(A) < 1$  and its complementary event  $\bar{A}$  with probability  $P(\bar{A}) = 1 - P(A)$  have identical probability. This means that  $P(A) = 1/2$ . Since the Shannon entropy

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

is maximized for  $P(A) = 1/2$ , this kind of entropy could serve as the basis for 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

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

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

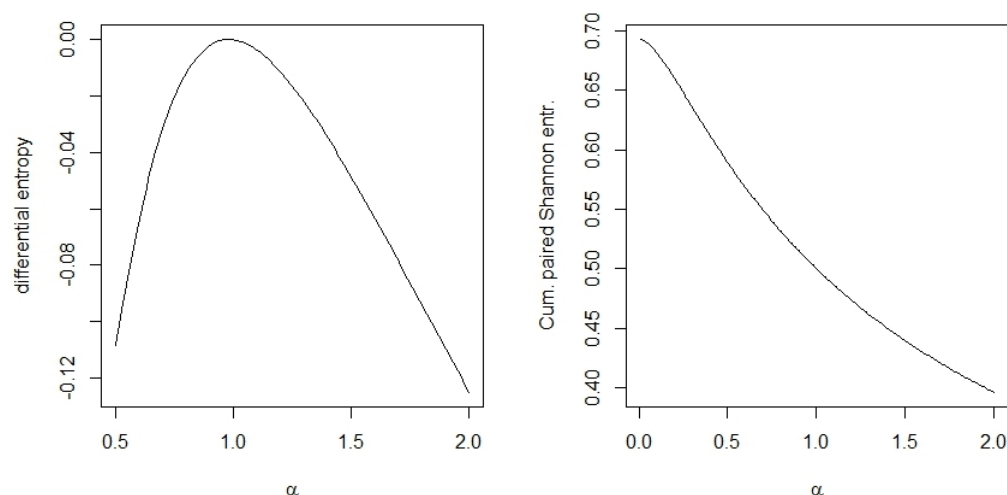
$$\begin{aligned} CPE_S(F) &= -\int F(x) \log F(x) dx - \int (1 - F(x)) \log(1 - F(x)) dx \\ &= -\int_0^1 (u \log u + (1 - u) \log(1 - u)) q(u) du \end{aligned} \quad (2)$$

with probability integral transformation  $u = F(x)$ , quantile function  $Q(u) = F^{-1}(u)$ , and 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 differential Shannon entropy ( $E_S$ ) takes a value of  $\ln 2$  for all bipolar distributions, independently of how large the distance between the two mass points is. Rao [5] identified this property as an important advantage of cumulative entropies over the differential entropy. The different behaviors of differential (Shannon) entropy and cumulative Shannon entropy are illustrated in Example 1.

**Example 1.** We consider the symmetric beta distribution with density

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

while parameter  $\alpha \in (0.1, 2)$ . This range allows almost bipolar distributions ( $\alpha = 0.1$ ), uniform distributions ( $\alpha = 1$ ), and bell-shaped distributions ( $\alpha = 2$ ). Figure 1 compares the values of the differential entropy and the cumulative paired Shannon entropy for this range of the parameter  $\alpha$ . We see 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  $\alpha$ .



**Figure 1.** Differential entropy and cumulative paired Shannon entropy for the symmetric beta distribution with several parameter values of  $\alpha$ .

As we do not want to assume information that we do not possess, we perform the ME task to search for densities that are relying on maximum entropy. The densities that are based on minimum information are the only ones that could reasonably be used. In this paper, we propose to rely on bipolar distributions, as they provide contradictory information which is even less useful than minimum or no information for prediction.

The following examples intend to explain where bipolar distributions appear in real situations and how this bipolarity affects the predictability of a random variable  $X$ .

**Example 2.** In an opinion poll survey, individuals are asked to judge their political belief on a continuous left–right scale. 0 (100) symbolizes an extremely left (right) political view. The survey’s result maximizes the cumulative paired Shannon entropy if half of the people state to be extremely left ( $= 0$ ), and the other half state to be extremely right ( $= 100$ ). This is a situation of maximum uncertainty regarding how to predict the political view of an individual person.

**Example 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. However, 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 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 over buying the product? In this situation, buying would therefore mean receiving an either excellent or very bad product. This is a situation in which it is most complicated to predict the customer’s decision.

Both situations of Examples 2 and 3 can be characterized by the term ‘contradictory information’ in contrast to minimum or no information. In general, information is able to reduce uncertainty. However, contradictory information is implicitly defined by the fact that it increases uncertainty and provides a high chance for a wrong decision. (Anti-information is a related, but less formal concept introduced by the information scientist J. Verhoeff [40].) Therefore, as bipolar distributions

lead to contradictory information, it is an important task to consider entropies that will be maximized by a bipolar distribution if there are no constraints. Thus, we propose to use cumulative paired entropies to cover contradictory information. Example 1 already showed that the differential entropy does not embody contradictory information. In the following section, we compare the information provided by using cumulative residual and cumulative direct entropies in contrast to using cumulative paired entropies. Rao et al. [4] introduced cumulative residual entropies as a Shannon entropy, where the density is substituted by the survival function. Then, [5–17] discussed this cumulative residual Shannon entropy:

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

DiCrescenzo and Longobardi [41] applied the Shannon entropy to a distribution function and called it cumulative entropy. This gave the formula

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

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

What does maximum entropy mean for the cases of cumulative residual and cumulative direct Shannon entropy? 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 bipolar. However, this bipolarity is less extreme than in the symmetric case. This is due to that fact that it holds  $P(X = a) = 1 - 1/e$  and  $P(X = b) = 1/e$ . Therefore, there is a preference for the alternative  $a$  that makes the prediction of  $X$  easier than in the symmetric case. However, there is still somewhat contradictory information rather than information. Regarding (4), the probabilities for  $a$  and  $b$  have to be interchanged to get a maximum  $CDE_S$  distribution. The following example illustrates this for a beta distribution with parameters  $\alpha$  and  $\beta$ .

**Example 4.** Let  $X$  be beta distributed with density

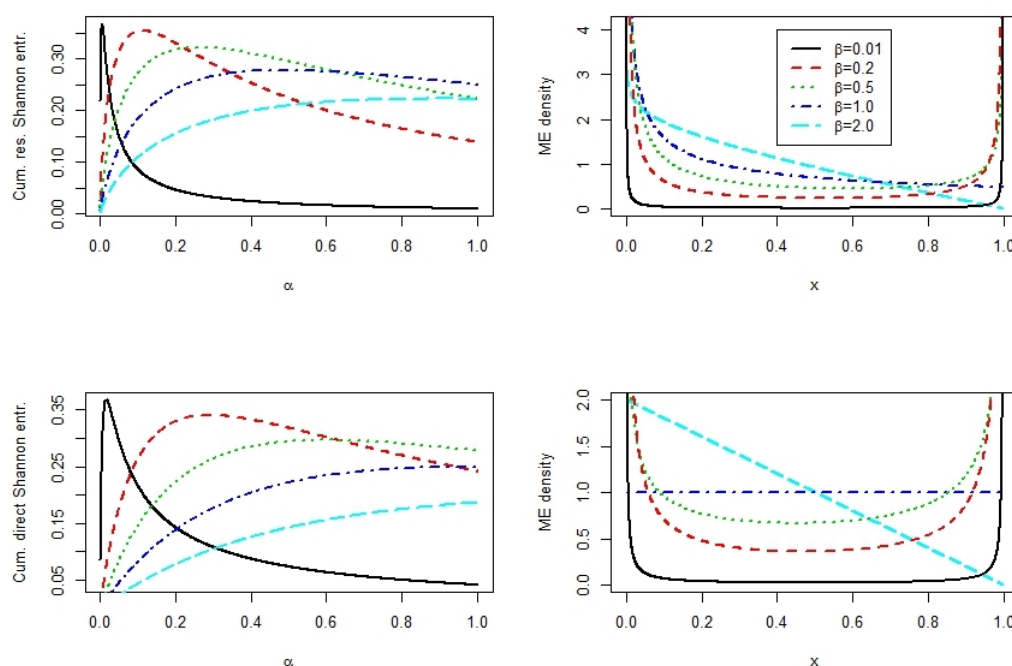
$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1, \alpha, \beta > 0.$$

In the following section, we fix  $\beta$  and compute  $\alpha$  such that (3) or (4) will be maximized. In Table 1  $\alpha_{cre}$  and  $\alpha_{cde}$  denote the corresponding maximum values. Moreover, this table also contains the maximum values of  $CRE_S$  and  $CDE_S$ . We see that the maximum is attained for small values of  $\alpha$  and  $\beta$ , denoting a slightly asymmetric U-shaped beta distribution.

**Table 1.** Maximum entropy beta distributions with several parameter values of  $\beta$ .

$\beta$	$\alpha_{cre}$	Max. $CRE_S$	$\alpha_{cde}$	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

Figure 2 illustrates the maximum  $CRE_S$  and  $CDE_S$  beta distributions for the parameter settings displayed in Table 1.



**Figure 2.** Cumulative residual and cumulative direct Shannon entropy for the asymmetric beta distribution with parameter values  $\alpha$  such that  $CRE_S$  ( $CDE_S$ ) is maximized for given  $\beta$ .

To date, the support has been  $\mathbb{R}$ . However, as, e.g., in the reliability theory, the focus is on random variables with only non-negative support. Thus, it is of importance to also discuss this situation. When only considering a random variable with non-negative support and for the ME quantile function  $Q$  holds  $Q(0) = 0$ , maximizing  $CRE_S$  or  $CDE_S$  gives a distribution which is no longer  $U$ -shaped, and the maximum entropy situation no longer corresponds with contradictory information. We illustrate this in Example 5 using a special beta distribution. The parameter  $\beta$  will be set to 1 such that  $Q(0) = 0$ .

**Example 5.** Let  $X$  be beta distributed with density

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

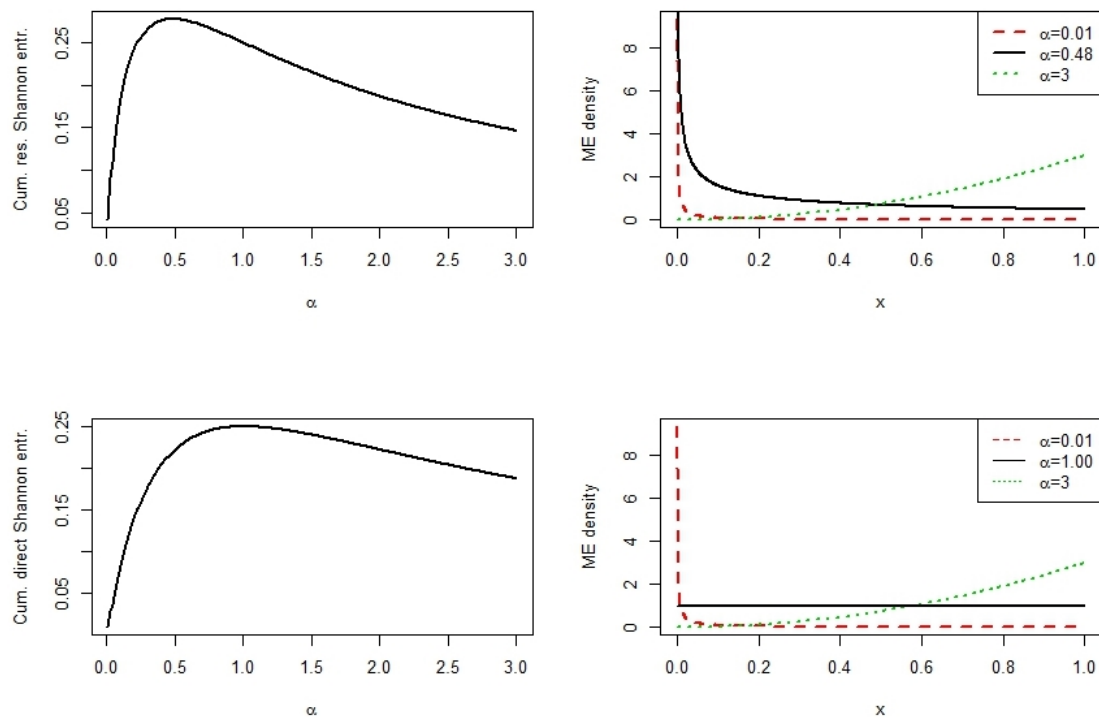
Table 2 displays the values for  $CRE_S$  and  $CDE_S$  for certain values of  $\alpha$ . For  $\alpha = 0.01$ , we get an extremely right skewed and for  $\alpha = 3$  an extremely left skewed distribution. For  $\alpha = 0.48$  ( $\alpha = 1$ ),  $CRE_S$  ( $CDE_S$ ) attains its maximum.

**Table 2.** Cumulative residual and cumulative direct Shannon entropy for the beta distribution with several parameter values  $\alpha$  and  $\beta = 1$ .

$\alpha$	$CRS_S$	$CDE_S$
0.01	0.1836	0.0826
0.48	0.2779	0.2191
1.00	0.2500	0.2500
3.00	0.1464	0.1876

Figure 3 displays, on the top row, the maximum  $CRE_S$  distribution at  $\alpha = 0.48$  and shows that this is an arrangement between an extremely right skewed ( $\alpha = 0.01$ ) and an extremely left skewed ( $\alpha = 3$ ) distribution. On the bottom row, we see the maximum  $CDE_S$  distribution at  $\alpha = 1$ .





**Figure 3.** Cumulative residual and cumulative direct Shannon entropy and density for the beta distribution with several parameter values  $\alpha$  and  $\beta = 1$ .

The question we raised in the section title on what maximizing cumulative direct, residual, and paired Shannon entropies means can be answered by the conclusion that maximizing these entropies leads 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. This *U*-shaped distribution corresponds to contradictory information. Examples 2 and 3 showed that this kind of information is even less useful for prediction and estimation than minimum or no information. Therefore, those distributions are the only reasonable distributions to consider to not assume information that we do not possess.

In the following section, we unify the diverse approaches of cumulative entropies and introduce the general class of cumulative  $\Phi$  entropies. Then, we derive two general formulas for ME quantile functions under some restrictions.

### 3. Maximum Cumulative $\Phi$ Entropy Distributions

In the following, we will first introduce the general class of cumulative  $\Phi$  entropies that incorporates and generalizes well-known entropies. Then, we will derive general formulas for maximum entropy distributions for this new class regarding arbitrary support as well as non-negative support.

#### 3.1. General Class of Cumulative $\Phi$ Entropies

In this section, we incorporate cumulative direct, residual, and paired entropies into one approach. Additionally, instead of focusing on the Shannon case, we allow for a general so-called entropy generating function  $\phi$ , which has to be non-negative and concave on  $[0, 1]$ . In general, but not mandatory  $\phi$  has a maximum in the interval  $[0, 1]$ . Hence, the corresponding cumulative  $\phi$  entropies are the cumulative paired  $\phi$  entropy

$$CPE_{\phi}(F) = \int_{\mathbb{R}} \phi(F(x)) + \phi(1 - F(x)) dx = \int_0^1 (\phi(u) + \phi(1 - u)) q(u) du, \quad (5)$$

the cumulative residual  $\phi$  entropy

$$CRE_{\phi}(F) = \int_{\mathbb{R}} \phi(1 - F(x))dx = \int_0^1 \phi(1 - u)q(u)du, \quad (6)$$

and the cumulative direct  $\phi$  entropy with

$$CDE_{\phi}(F) = \int \phi(F(x))dx = \int_0^1 \phi(u)q(u)du. \quad (7)$$

To cover all three cases into one approach, we consider a general concave entropy generating function  $\Phi$  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,

$$CE_{\Phi}(F) = \int_0^1 \Phi(u)du$$

will be called cumulative  $\Phi$  entropy. For the maximum entropy task, the objective is now to maximize this cumulative  $\Phi$  entropy with respect to  $F$  under distinct constraints and to search for the distribution that maximizes this entropy. At first, we consider cumulative  $\Phi$  entropies in a situation with fixed mean and variance. The restriction to these two moments can be explained by the fact that higher moments lead to equations for the ME quantile function which cannot be solved explicitly or the solution does not exist. Then, 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 relations between the fixed mean and the fixed  $k$ -th power moment.

In Section 3.2, we consider the situation where mean and variance are fixed and in Section 3.3, the situation with the additional requirement of  $Q(0) = 0$ .

### 3.2. General Results for Arbitrary Support

The focus in this section is on the situation where mean and variance are fixed and the support is arbitrary. First, the maximum cumulative  $\Phi$  entropy principle and then the generalized maximum cumulative  $\Phi$  entropy principle are introduced. General formulas for ME quantile functions are provided.

#### 3.2.1. Maximum Cumulative $\Phi$ Entropy Approach

In this section, for a given entropy and fixed constraints, general formulas for ME distributions are derived. This maximum cumulative  $\Phi$  entropy approach follows the maximum entropy principle in the sense of [42,43]. The following theorem provides a general formula for the ME quantile function  $Q$ .

**Theorem 1.** Let  $CE_{\Phi}$  be the cumulative  $\Phi$  entropy with concave entropy generating function  $\Phi$  such that the derivative  $\Phi'$  exists i.e., is quadratic integrable over  $[0, 1]$ , and  $|\Phi(0)| < \infty$ ,  $|\Phi(1)| < \infty$  hold. Then, the maximum CE distribution under the constraints of fixed mean  $\mu$  and variance  $\sigma^2$  is given by the quantile function

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

**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

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



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

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

with  $l_1$  and  $l_2$  denoting the Lagrange parameters. The Euler–Lagrange equation gives

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

Solving this equation leads to the quantile function

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

$l_1$  and  $l_2$  are determined by the moments  $\mu$  and  $\sigma^2$ . 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

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

From

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

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 $\Phi$ Entropy Approach

In this section, for a given quantile function  $Q$ , the corresponding generating function  $\Phi$  of the cumulative  $\Phi$  entropy will be derived. This generalized maximum cumulative  $\Phi$  entropy approach follows the generalized maximum entropy principle formulated by [44]. We also use formula (8) for

this approach. 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  $X$ . Thus, the partial mean function  $\mu(u)$  is given by

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

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 that  $Q$  is the ME quantile function under the constraints of given mean  $\mu$  and variance  $\sigma^2$ .

**Corollary 1.** *Let  $Q$  be a quantile function. The entropy generating function  $\Phi$ , such that  $Q$  is ME under the constraints 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

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

□

Hence,  $-\Phi(u)/u$  is the conditional mean of  $X$  given  $X \leq Q(u)$  for  $u \in [0, 1]$ . It holds  $\mu(0) = 0$  and  $\mu(1) = \mu$  such that  $\Phi(0) = 0$  and  $\Phi(1) = -\mu$ . The partial mean function  $\mu(u)$  therefore has a special role. As  $\mu(u)$  sums up the values  $x$  of  $X$  weighted with the density  $f(x)$  until the  $u$ -quantile of  $X$ , this addition gives constant values until the median quantile for an extremely  $U$ -shaped distribution. Thereafter, the value will be changed one time and stays again constant. Thus, the heavier the tails of a distribution, the steeper the entropy-generating function  $\Phi(u)$  at  $u = 0$  and  $u = 1$ . This leads to a large value for the derivative  $\Phi'(u)$  at  $u = 0$  and  $u = 1$ . If the support is  $\mathbb{R}$ , then  $\lim \Phi'(0) = \infty$  and  $\lim \Phi'(u) = -\infty$ . In line with the generalized maximum entropy principle, we will use (11) to derive  $\Phi$  such that a given distribution has the ME property under the constraints of fixed mean and variance. In Section 5, based on (11), we will propose an estimator for  $\Phi$ .

### 3.3. General Results for Non-Negative Support

To date, in the literature, the ME task was mainly considered for lifetime distributions with the special property that the support is  $(0, \infty)$ . Therefore, in this section, the focus is on the situation where next to the constraints of fixed mean and variance also the support is restricted to  $(0, \infty)$ . Similar to Section 3.2, the maximum cumulative  $\Phi$  entropy principle and then the generalized maximum cumulative  $\Phi$  entropy principle will be introduced in this situation and general formulas for ME quantile functions will be provided.

#### 3.3.1. Maximum Cumulative $\Phi$ Entropy Approach

In this section, for given entropy and constraints, a general formula for ME distributions will be derived, following the maximum cumulative  $\Phi$  entropy approach, while the support of the ME distribution is  $(0, \infty)$ , which 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  $\mu$  and a fixed  $k$ -th power moment  $\mu'_k$ ,  $k > 1$ . The following theorem shows how to derive the ME quantile function under these three constraints. For an ME solution to be existent, it requires a special relationship between the fixed moments  $\mu$  and  $\mu'_k$ .

**Theorem 2.** Let  $\Phi$  be a concave function on  $[0, 1]$  with derivative  $\Phi'$  such that  $-\Phi'(u) + \Phi'(0)$  is monotonically increasing. Then, the ME quantile function under the constraints of given mean and  $k$ -th power moment  $\mu'_k$  is

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

if

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

Otherwise there is no solution of the ME task.

**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

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

$kl_2$  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  $\mu^k$  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 $\Phi$ Entropy Approach

The generalized maximum cumulative  $\Phi$  entropy approach for random variables with non-negative support and  $Q(0) = 0$  remains to be discussed. We start with the knowledge of the

quantile function  $Q$  to derive the corresponding generating function  $\Phi$  of the cumulative  $\Phi$  entropy such that  $Q$  is the ME quantile function for  $\Phi$  under the constraints  $Q(0) = 0$  and fixed mean  $\mu$  and fixed  $k$ -th power moment  $\mu'_k$ . Therefore, we introduce a special partial mean function.  $\mu_{k-1}(u)$  denotes the partial  $(k-1)$ -th power mean function with

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

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

**Corollary 2.** *Let  $Q$  be a quantile function. The entropy generating function  $\Phi$ , such that  $Q$  is ME under the constraints  $Q(0) = 0$ , fixed mean and fixed variance, is given by*

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

**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

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

□

In Section 5, we use (14) to estimate  $\Phi$  from a data set such that the data are generated by the corresponding ME distribution under the constraints of  $Q(0) = 0$  and fixed mean and fixed  $k$ -th power moment.

#### 4. Applications

In this section, we give an overview about some ME distributions for cumulative entropies applying the results of Section 3. For some choices of  $\Phi$ , the problem of the ME task has already been solved. In the following section, we consider further choices of  $\Phi$  with a focus on those that lead 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 displays an overview of several entropy generating functions and the corresponding ME distributions. The table is divided by the situation where mean and variance are fixed, by the distinction of the MaxEnt and the GMaxEnt task, and by the situation with the additional requirement of  $Q(0) = 0$ . Moreover, while cases no. 1 to no. 4 require symmetry of the ME distribution, cases no. 5 to no. 12 allow for skewness of the ME distribution.  $f_N$  and  $F_N^{-1}$  denote the density and the quantile function of the standard normal distribution. We try to assign well known terms to the cumulative entropies generated by the respective  $\Phi$ . For the solution of the GMaxEnt task (no. 9 and 10), such terms are not available. The second column refers to the Appendix where the cases are discussed in detail.

**Table 3.** Entropy generating functions with corresponding maximum entropy distributions.

No.	App.		$\Phi$	ME distr.
Fixed mean and variance, without $Q(0) = 0$				
1		Shannon	$-u \ln u - (1-u) \ln(1-u)$	logistic
2		Havrda-Charvát	$\frac{u(u^{\alpha-1}-1)/(1-\alpha)}{+(1-u)((1-u)^{\alpha-1}-1)/(1-\alpha)}$	Tukey $\lambda$
3	Appendix A.1	Leik	$2(1/2 -  u - 1/2 )$	bimodal
4	Appendix A.1	Gini	$2(1/4 - (u - 1/2)^2)$	uniform
5	Appendix A.1	Mielke	$2((1/2)^r -  u - 1/2 ^r)$	symm. beta
6	Appendix A.2	Havrda-Charvát like	$-\frac{B(2-\alpha, \alpha)}{\alpha-1} \left( \beta(u; 2-\alpha, \alpha) - \frac{u}{B(2-\alpha, \alpha)} \right)$	general. logistic
7	Appendix A.3	non-symm. Shannon	$-\alpha_1 u \ln u - \alpha_2 (1-u) \ln(1-u)$	skewed logistic
8	Appendix A.3	non-symm. Havrda-Charvát	$\alpha_1 u \frac{u^{\alpha_1}-1}{1-\alpha_1} + \alpha_2 (1-u) \frac{(1-u)^{\alpha_2}-1}{1-\alpha_2}$	skewed Tukey $\lambda$
9	Appendix A.4	GMaxEnt	$\frac{u(u^{\alpha_1-1}-1)/(1-\alpha_1)}{+(1-u)((1-u)^{\alpha_2-1}-1)/(1-\alpha_2)}$	general. Tukey $\lambda$
10	Appendix A.5	GMaxEnt	$\frac{2\gamma^3/(1+\gamma^2)f_N(F_N^{-1}(u)) \cdot I(u \leq \gamma^2/(1+\gamma^2))}{+2/(\gamma(1+\gamma^2)) \left( F_N(F_N^{-1}(0.5((1+\gamma^2)u+1-\gamma^2))) \right.}$ $\left. -1/\sqrt{2\pi} \right) \cdot I(u > \gamma^2/(1+\gamma^2))$	skewed normal
Fixed mean and $k$ -th moment, with $Q(0) = 0$				
11	Appendix A.6	Shannon	$-(1-u) \ln(1-u)$	Weibull
12	Appendix A.7	Havrda-Charvát	$-(1-u)((1-u)^{\alpha-1}-1)/(\alpha-1)$	ext. Burr XII

Some of the results presented in Table 3 are already known from the literature. These are the solutions of no. 1 [24], no. 2 [34], no. 3 [34], no. 4 [23,45] and no. 11 [5]. The remaining cases state new results which are discussed in the Appendix for all readers interested in flexible statistical distributions. The general finding can best be illustrated by solutions no. 1 and no. 2. The ME distributions are the logistic and the Tukey  $\lambda$  distribution. Solving the ME task for the classical differential and the Havrda-Charvát (or Tsallis) entropy given fixed mean and variance results in the normal and the  $t$ - or  $r$ -distribution [46,47]. The difference is easy to explain. The cumulative entropy pulls the ME distribution as much as possible (limited by the restrictions) towards a  $U$ -shaped distribution. This leads to distributions with heavier tails (logistic instead of normal, Tukey  $\lambda$  instead of  $t$  or  $r$ ).

There are a lot of entropy-generating functions well-known from physics which could also be considered in the context of cumulative entropies. It is easy to show that the results of Theorem 1 and Theorem 2 can be applied to, e.g., the generating functions of the Rényi [27], the Kaniadakis [48], or the Hanel-Thurner entropy [49], to mention only a few. Another comment deals with the concept of skewness. Some families of distributions have natural parameters of skewness. If the members of these families have closed expressions for the quantile function, Corollary 1 can be applied directly to derive the function  $\Phi$  for the corresponding cumulative entropy (GMaxEnt task). This is the reason why we focus on the generalized Tukey  $\lambda$  distribution. It is worth noting that again, a kind of non-symmetric cumulative Havrda-Charvát entropy appears as solution (see no 9). Other families of skewed distributions will be defined by modifying a given symmetric distribution. The Fechner approach as well as the still more popular Azzalini approach proceed in such a way. The Fechner approach introduces skewness by splitting the scale parameter for the positive and the negative halves of the underlying symmetric distribution. This leads to a corresponding splitting of the quantile function. Corollary 1 can again easily be applied to solve the GMaxEnt task as long as the quantile function is available in a manageable form. The solution for the normal distribution is given by

solution no. 10. For the more popular Azzalini approach [50,51], this is not the case. Therefore, we omit to discuss the GMaxEnt task for this concept to generate skewed distributions.

Table 3 only contains special choices of the entropy generating functions  $\Phi$ . The main question is how to know  $\Phi$ . The answer could be given by an axiomatic approach or empirically. The starting point for the axiomatic approach are fundamental requirements with a plausible and general accepted interpretation in the considered scientific discipline. Such axiomatizations are available for the differential and the Tsallis entropy. A recent publication on this topic is e.g., [52]. In the context of cumulative entropies, we can go back to approaches in the fuzzy set theory. In this theory, measures of indefiniteness will be axiomatized (see [19,20,53]). The axioms are directly applicable to cumulative entropies. (The discussion of alternative entropies, skewness and axiomatic approaches is based on valuable comments of two anonymous referees.) In the following section, we do not want to discuss the axiomatic approach further. Instead, in the next section, we will focus on how to estimate the entropy generating function  $\Phi$ .

## 5. Estimating the Entropy Generating Function

Can we learn something from data about the entropy generating function  $\Phi$  for which the data generating distribution is an ME distribution under the constraints of given mean and variance? The entropy generating function  $\Phi$  is given by the partial mean function

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

Therefore, we can estimate this partial mean function to get an estimator for  $\Phi$ .

Let  $X_1, \dots, X_n$  be identically and stochastically 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

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

for the entropy generating function  $\Phi$ .

We demonstrate the usefulness of this estimator by the following examples.

**Example 6.** The data set consists of the S&P500 standardized daily logarithmic returns from 10-05-2012 to 10-04-2017. (The data are available from <https://fred.stlouisfed.org/series/SP500>.) 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 4, 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 chose the number of degrees of freedom by trail and error, but ML estimation gives a value not far away from 4.

We can see that by estimating the entropy generating function  $\Phi$  by the partial mean function, the density of the S&P500 standardized daily logarithmic returns can be fitted quite well.

In the following example, we consider a situation with non-negative support. We know from (14) that for a non-negative random variable with  $Q(0) = 0$  and fixed mean  $\mu$  and fixed  $k$ -th power moment  $\mu'_k$  the entropy generating function  $\Phi$  is given by

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

For this entropy generating function  $Q$  is an ME quantile function.



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

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

An estimator for the entropy generating function  $\Phi$  is given by

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

We will show that this estimator works well for a real data set and the Weibull distribution. Therefore, we need the partial  $(k-1)$ -th power mean function for the Weibull distribution with shape parameter  $r$  and scale parameter  $\lambda$ . For this distribution, it holds

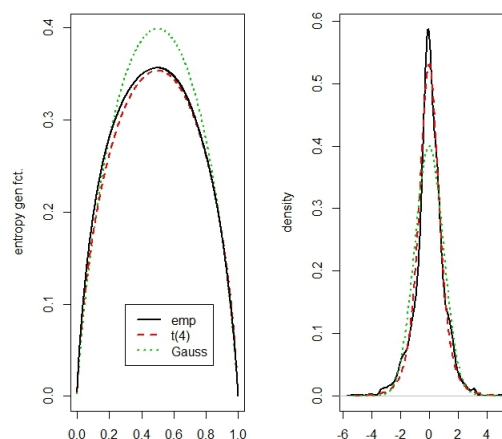
$$\mu_{k-1}(u) = \int_0^u (\lambda(-(1-v)^r))^{k-1} dv = \lambda^{k-1} \Gamma \left( \frac{k-1}{r} + 1 \right) \Gamma \left( -\ln(1-v); \frac{k-1}{r} + 1, 1 \right)$$

for  $u \in [0, 1]$ .  $\Gamma(x; a, b)$  denotes the distribution function of a  $\Gamma$  distribution with shape parameter  $a$  and scale parameter  $\beta$ . The corresponding entropy generating function  $\Phi$  such that this Weibull distribution is  $CE_\Phi$  maximum under  $Q(0) = 0$  and the constraints of fixed mean  $\mu$  and fixed  $k$ -th 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

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

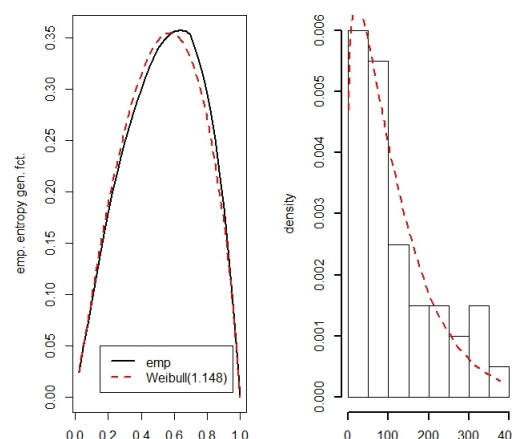


**Figure 4.** Estimated entropy generating function and estimated density for the S&P500 standardized daily log returns from 10-05-2012 to 10-04-2017.

**Example 7.** Let  $X$  be a random variable representing the duration in days between two explosions in the mines of a specific region. From [54], we get the following dataset 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 5, 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 relatively small sample size.



**Figure 5.** Estimated entropy generating function and estimated density for a data set with time intervals between 41 mine explosions.

Further work will be conducted 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. However, this will be investigated in future research.

## 6. Conclusions

To be able to estimate and predict while not using information that we do not possess, it is important to derive maximum entropy distributions. Maximizing Shannon's differential entropy under different moment constraints is a well-known task. Without any constraints, the differential entropy will be maximized by a uniform distribution representing the situation of no information. However, an extremely bimodal (=bipolar) distribution represents a situation of so-called contradictory information since an event and its complement can happen with equal probability. In this situation, it is extremely hard to make a forecast, even harder than for a uniformly distributed random variable. Hence, this paper claims that contradictory information is even less useful than minimum or no information as it increases uncertainty and provides a high chance for a wrong decision. Such a bipolar distribution is covered by maximizing a cumulative entropy instead of the differential entropy without any constraints. Such a cumulative entropy depends either on the distribution function (direct), on the survival function (residual) or on both (paired). Under the constraints 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 maximization problem leads to symmetric ME distributions like the logistic and the Tukey  $\lambda$  distribution [21,34]. So far, other ME distributions were found for the cumulative paired Leik and Gini entropy [23,34,45]. There are two different principles to derive maximum entropy distributions. The maximum entropy principle in the sense of [42,43] is the task to derive an ME distribution for a given entropy and fixed constraints. The generalized maximum entropy approach formulated by [44] uses a given ME distribution for which the corresponding generating function of the cumulative entropy will be derived. In this paper, we will apply both approaches for the cumulative  $\Phi$  entropy, which generalizes the cumulative paired entropy in several ways and thus introduced the maximum cumulative  $\Phi$  entropy approach and the generalized maximum cumulative  $\Phi$  entropy approach. Moreover, we regarded situations with different constraints. First, we considered a situation with arbitrary support and given mean and variance and second a situation with non-negative support and the additional constraint of  $Q(0) = 0$  for the ME quantile function. This was done, as 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 these additional constraints, we derived ME distributions for fixed mean and  $k$ -th power moment. For the situation with arbitrary support and given mean and variance, we introduced the cumulative paired Mielke(r) entropy and derived the ME distributions. The results already known for the cumulative paired Leik and Gini entropy are included for  $r = 1$  and  $r = 2$ . Then, starting with a natural generalization of the derivative of the entropy generating function known from the logistic distribution, we derived as ME distribution the generalized logistic distribution (GLO) immediately. Considering a linear combination of entropy generating functions led to new ME distributions with skewness properties. Here, we derived the skewed logistic distribution and the skewed Tukey  $\lambda$  distribution in line with [55]. Next, using the generalized maximum  $\Phi$  entropy approach, we derived an entropy generating function such that a pre-specified skewed distribution is an ME distribution. The generalized Tukey  $\lambda$  distribution served as an example. Now, we considered Fechner's proposal to define different values of a scale parameter for both halves of a distribution for getting skewed distributions. Again, we derived the corresponding entropy generating function. The skewed normal distribution served as an illustrative example. Then, we focused on the situation where the support of the ME distribution is restricted to  $(0, \infty)$ , while using the maximum cumulative  $\Phi$  entropy approach. Here, we derived as ME distribution for the cumulative residual Shannon entropy the Weibull distribution and for the cumulative residual Havrda-Charvát entropy the extended Burr XII distribution. Finally, we proposed an estimator for the cumulative  $\Phi$  entropy generating function representing all the properties of the underlying ME data generating distribution. This gives an alternative to non-parametric estimation of density functions or distribution functions. The usefulness of this estimator was demonstrated for two real data sets.

**Author Contributions:** I.K. conceived the new maximum entropy concepts, investigated its properties, applied it to several distributions and wrote an initial version of the manuscript. M.D. contributed by mathematical and linguistic revision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to thank Paul van Staden for the hint to Hosking's work about the generalized logistic distribution and three anonymous reviewers for their constructive criticism, which helped to improve the presentation of this paper significantly.

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

## Appendix A

In Appendixes A.1–A.5 the situation will be considered where mean and variance are fixed and the support is arbitrary. Then, in Appendixes A.6 and A.7 the support of the ME distribution has to be non-negative and the property  $Q(0) = 0$  will be required as an additional constraint.

Furthermore, we will use the maximum cumulative  $\Phi$  entropy approach in Appendixes A.1–A.3 as well as in Appendixes A.6 and A.7. In Appendixes A.4 and A.5 we will use the generalized maximum cumulative  $\Phi$  entropy approach.

#### Appendix A.1. Cumulative Paired Mielke( $r$ ) Entropy and the Symmetric Beta Distribution

Mielke [56] as well as Mielke and Johnson [57] discussed two-sample linear rank tests for alternatives of scale with score generating function

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

This family of tests, parameterized by  $r$ , includes the well known test from [58] for  $r = 1$  and [59] for  $r = 2$ . Mood [59] also derived the symmetric distribution corresponding to an asymptotically optimal linear rank test with this score generating function. Similar to [34], there is a simple relationship between the score generation functions of two sample linear rank tests for scale alternatives and the entropy generating function  $\Phi$  of a corresponding cumulative paired  $\Phi$  entropy. In the case [56,57] considered, 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 these cases.  $\Phi$  is differentiable and the derivative is 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 section, we talk about the cumulative paired Mielke( $r$ ) entropy. We show that the symmetric double beta distribution maximizes  $CPE_{M(r)}$ ,  $r > 1$ , if mean and variance are known (see Table 3 no. 5). For  $r = 1$ , we get an extremely bimodal distribution. In both cases, the support of the ME distribution is a closed interval. The four-parameter double beta distribution is defined by

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

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

**Corollary A1.** *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 constraints of known mean  $\mu$  and known variance  $\sigma^2$  is a double beta distribution with parameters  $a = 1/(r-1)$ ,  $b = 1$  and support  $[\mu - \sigma\sqrt{2r-1}, \mu + \sigma\sqrt{2r-1}]$  for  $r > 1$ . We get an extremely bimodal distribution with support  $\{\mu - \sigma, \mu + \sigma\}$  for  $r = 1$ .*

**Proof.** The derivative of  $\Phi$  is given by

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

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

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

with the support given by

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

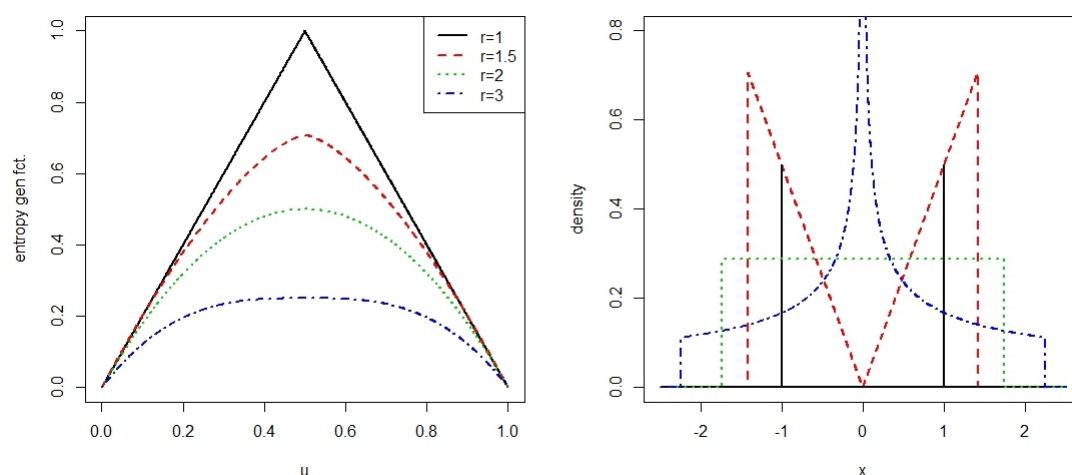
for  $r \neq 1$  and  $\{\mu - \sigma, \mu + \sigma\}$  for  $r = 1$ . For  $r \geq 1$  the corresponding distribution function is

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

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 beta 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 an extremely bimodal distribution with mass points  $\mu - \sigma$  and  $\mu + \sigma$ .  $\square$



**Figure A1.** Entropy generating function and density for the double beta distribution with parameter values  $r = 1, 1.5, 2, 3$ .

In Figure A1, 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 ME density is bipolar (extremely bimodal) (see Table 3 no. 3). Increasing  $r$  leads to a bimodal distribution ( $r = 1.5$ ), a uniform distribution ( $r = 2$ ) (see Table 3 no. 4), up to a very leptokurtic distribution with singularity at 0 ( $r = 3$ ). With increasing  $r$ , the entropy generating functions become more and more platycurtic. 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 distributional characteristics can be learned from the form of the entropy generating function. Klein et al. [34] discussed the cumulative paired Leik entropy and Dai and Chen [45] the cumulative paired Gini entropy. Both are embedded in the class of cumulative paired Mielke( $r$ ) entropy for  $r = 1$  and  $r = 2$ .

### Appendix A.2. Cumulative Paired Shannon Entropy and the Generalized Logistic Distribution

Formula (8) can also be used by starting with derivative of the entropy generating function, instead of with the entropy generating function itself. Such a derivative could be

$$\Phi'(u) = \phi\left(\frac{1-u}{u}\right), \quad u \in [0, 1] \quad (\text{A1})$$

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

$$\Phi'(u) = \ln\left(\frac{1-u}{u}\right) \quad u \in [0, 1]. \quad (\text{A2})$$

In this special case,  $\Phi'$  belongs to the cumulative paired Shannon entropy and the corresponding ME distribution is the logistic distribution under the constraints of fixed mean and variance [21,24,34,39].

As an extension, we can consider

$$\Phi'(u) = \frac{((1-u)/u)^{\alpha-1} - 1}{\alpha - 1}, \quad u \in [0, 1], \quad \alpha \neq 1. \quad (\text{A3})$$

For  $\alpha \rightarrow 1$  we get (A2).

**Corollary A2.** The density of the ME distribution for the cumulative entropy with derivative (A3) is given by

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

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

**Proof.**  $\Phi'$  determines the ME quantile function via (8). This means

$$Q(u) = -\frac{((1-u)/u)^{\alpha-1} - 1}{\alpha - 1}, \quad u \in [0, 1]. \quad (\text{A4})$$

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  $i = 1, 2, \dots, k$ . 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.  $\square$



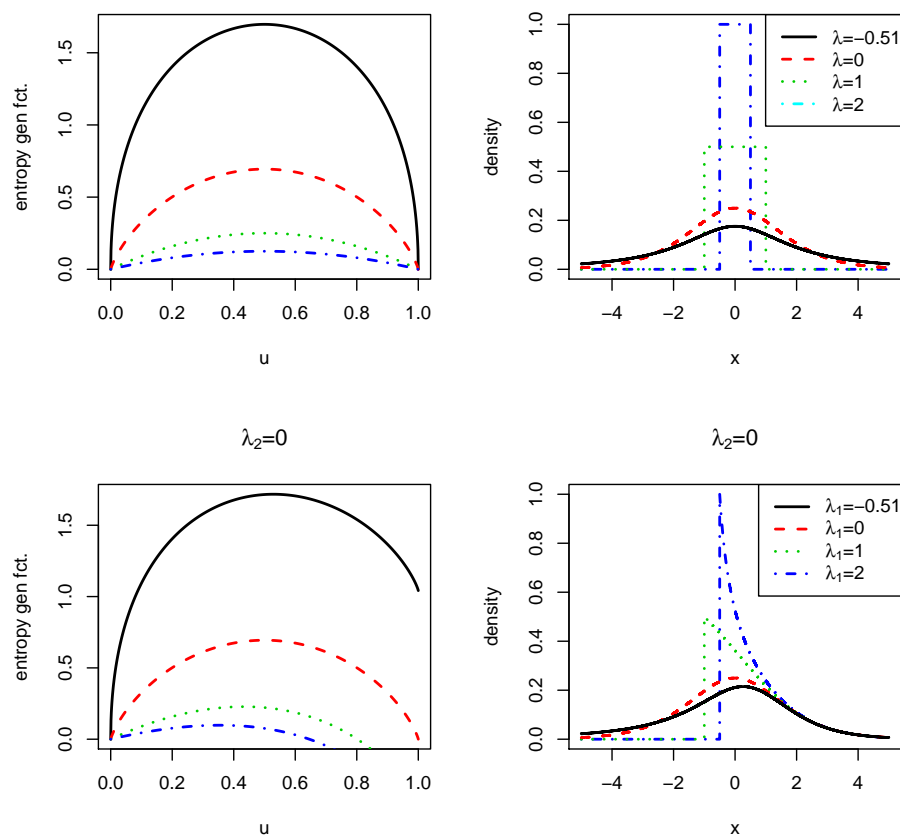
The quantile function (A4) has first been introduced by [60] under the term ‘generalized logistic distribution’ (GLO). Further discussions of the GLO can be found in [61,62]. The generalization of the GLO is due to the fact that skewness will be introduced into the logistic distribution by the parameter  $\alpha$ . The support also depends on the parameter  $\alpha$ . For  $\alpha \rightarrow 1$ , we get the support  $\mathbb{R}$  and as ME distribution 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 skewed to the left. For  $\alpha < 1$ , we get the support  $[-1/(1 - \alpha), \infty)$  and the ME distribution is skewed to the right.

A working paper written by [63] gives an overview about other generalizations of the logistic distribution that are completely different to (A4). Nassar and Elmasry as well as Tripathi et al. [64,65] are also concerned with generalizations of the logistic distribution.

In this paper, as we started with the derivative of the entropy generating function, the entropy generating function  $\Phi$  belonging to (A3) remains to be identified. By simple integration, we get the partial mean 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  $\beta(\cdot; a, b)$  as the distribution function of the  $\beta(a, b)$  distribution (see Table 3 no. 6). 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 A2.** Entropy generating function and density for the generalized logistic (GLO) distribution with parameter values  $\alpha = 0.6, 1, 1.2, 1.4$ .

Figure A2 displays  $\Phi$  as well as the corresponding density of the GLO distribution. The logistic distribution serves as a reference ( $\alpha = 1$ ). The other settings show left skewed ( $\alpha = 0.6$ ) and two right skewed distributions ( $\alpha = 1.2$  and  $\alpha = 1.4$ ).

### Appendix A.3. Asymmetric Entropy Generating Functions and 'New' Skewed Logistic and Tukey $\lambda$ Distributions

Again, we apply the maximum cumulative  $\Phi$  entropy approach where the entropy generating function  $\Phi$  is fixed and we search for the corresponding ME distribution.

In [34] entropy generating functions of the symmetric form

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

with  $\varphi$  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  $\lambda$  distribution as ME distributions under the constraints of given mean and variance (see Table 3 no. 1, no. 2).

In the following section, we consider the more general case that

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

with  $\varphi_i, i = 1, 2$  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  $\varphi_i$  concave on  $[0, 1], i = 1, 2$ . If  $\varphi$  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 a scale parameter. In the following, we are interested in skewed ME distributions and thus this case will not be considered.

Examples for asymmetric functions are

$$\varphi(u) = \begin{cases} -u \frac{u^\lambda - 1}{\lambda} & \text{for } u \in [0, 1], \lambda \neq 0 \\ -u \ln u & \text{for } u \in [0, 1], \lambda = 0 \end{cases} \quad (\text{A5})$$

We search 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 gives the corresponding ME quantile function.

**Corollary A3.** Let  $\Phi(u) = \alpha\varphi(u) + (1 - \alpha)\varphi(u), u \in [0, 1], \alpha \in [0, 1]$  with  $\varphi$  from (A5) the entropy generating function. Then the corresponding ME quantile function is

$$Q(u) = \begin{cases} (2\alpha - 1) + (\lambda + 1) \left( \alpha \frac{u^\lambda - 1}{\lambda} - (1 - \alpha) \frac{(1-u)^\lambda - 1}{\lambda} \right) & \text{for } u \in [0, 1], \lambda \neq 0 \\ (2\alpha - 1) + \alpha \ln u - (1 - \alpha) \ln(1 - u) & \text{for } u \in [0, 1], \lambda = 0 \end{cases} \quad (\text{A6})$$

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

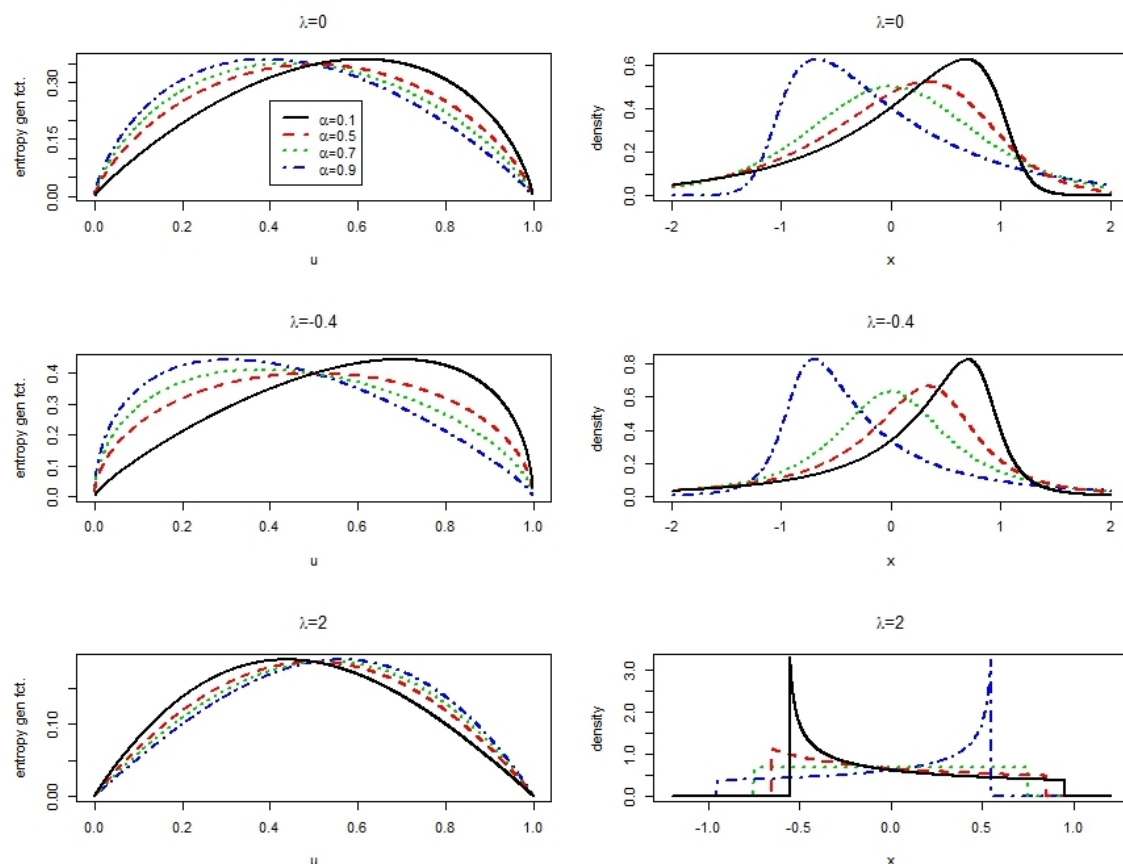
$\lambda$	$\alpha$	support
$\in (-1/2, 0]$	$= 0$	$[-1, \infty)$
	$= 1$	$(-\infty, 1]$
	$\in (0, 1)$	$(-\infty, \infty)$
$\in (0, \infty)$	$= 0$	$[1, -1 + (\lambda + 1)/\lambda]$
	$= 1$	$[1 - (\lambda + 1)/\lambda, 1]$
	$\in (0, 1)$	$[2\alpha - 1 - \alpha(\lambda + 1)/\lambda, 2\alpha - 1 + (1 - \alpha)(\lambda + 1)/\lambda]$

**Proof.** The claim follows immediately by calculating the derivative  $\Phi'$  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 an asymmetric ME quantile function determined by the negative of the derivative of  $\Phi$ . In the case of  $\alpha = 1/2$ , we get the the logistic distribution and the well-known Tukey  $\lambda$  distribution. Therefore, (A6) can be considered as the quantile function of a skewed logistic distribution or a skewed Tukey  $\lambda$  distribution (see Table 3 no. 7, no. 8). Both seem to be a new alternative to the already discussed generalized Tukey  $\lambda$  and the GLO-distribution. However, in his Ph.D. thesis [55] already introduced these alternatives. In his joint paper with King [66], he investigated the properties of the skewed logistic distribution intensively. Now, we know how the cumulative paired  $\Phi$  entropy looks like for which these distributions are ME distributions under the constraints of known mean and variance. [55] and [66] derived explicit expressions for the ML estimators for the location parameter, the scale parameter, the parameter  $\alpha$ , and the parameter  $\lambda$  as well as their asymptotic standard errors. Here, skewness and kurtosis are controlled by the latter parameters jointly. More precisely, [55] discussed the quantile function

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

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 A3.** Entropy generating function and density for the skewed logistic ( $\lambda = 0$ ) and the skewed Tukey  $\lambda$  distribution for  $\lambda = -0.4, 2$  and  $\alpha = 0.1, 0.5, 0.7, 0.9$ .

Figure A3 gives an impression of the different entropy characterizing functions  $\Phi$  and the density of the skewed logistic ( $\lambda = 0$ ) and the skewed Tukey  $\lambda$  distribution ( $\lambda = -0.4$ ). The symmetric logistic and Tukey  $\lambda$  distribution ( $\alpha = 1/2$ ) serve as a point of reference. The parameter setting  $\alpha = 0.1$  results in left skewed logistic and Tukey  $\lambda$  distributions. For  $\alpha = 0.7$  and  $\alpha = 0.9$ , we get a right skewed distribution. Moreover, in Figure A3, the parameter  $\lambda$  was set to 2.0 to display a compact support and an extremely skewed distribution.

#### Appendix A.4. Skewed Tukey $\lambda$ Distribution and Entropy Generating Function $\Phi$

Now, we will apply the generalized maximum cumulative  $\Phi$  entropy approach. This means that we interpret formula (8) as a situation where a quantile function  $Q$  is fixed and we derive the entropy generating function  $\Phi$  such that  $Q$  represents the corresponding ME distribution. We want to apply this approach to a quantile function considered by [67]. They introduced the so-called generalized Tukey  $\lambda$  distribution with the help of the quantile function

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

Here,  $\lambda_4$  is a location parameter,  $\lambda_3$  is a scale parameter and  $\lambda_1, \lambda_2$  determine the distribution's skewness. Freimer et al. [67] showed 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 the entropy generating function for which the generalized Tukey  $\lambda$  distribution is an ME distribution under the constraints of fixed mean and variance.

**Corollary A4.** Let  $Q$  be the quantile function of the generalized Tukey  $\lambda$  distribution. The entropy generating function  $\Phi$ , such that  $Q$  is the ME quantile function under the constraints of known mean and variance, is given by

$$\Phi(u) = -\frac{u^{\lambda_1+1} - (\lambda_1 + 1)u}{\lambda_1(\lambda_1 + 1)} - \frac{(1-u)^{\lambda_2+1} - (\lambda_2 + 1)(1-u)}{\lambda_2(\lambda_2 + 1)} - \frac{1}{\lambda_2 + 1} \quad (\text{A7})$$

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

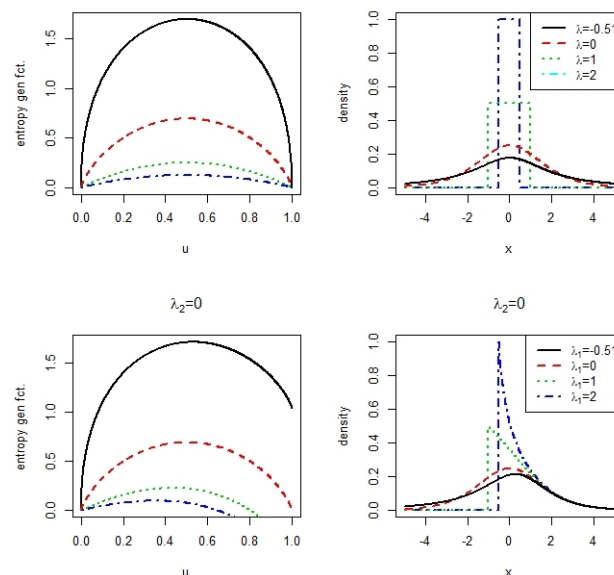
**Proof.** It is  $\mu(u) = -\int_0^u Q(v)dv = -\Phi(u)$ . For  $\mu(u)$  we get

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

□

The setting  $\lambda = \lambda_1 = \lambda_2$  results in the symmetric Tukey  $\lambda$  distribution (see Table 3 no. 9). Klein et al. [34] identified 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 (A7). In the upper two panels of Figure A4, we see the corresponding entropy generating function and the ME density for several choices of  $\lambda$ . The range of distributional properties for the Tukey  $\lambda$  distribution is much richer than it is for the ME distributions belonging to the cumulative paired Mielke(r)-entropy. For negative values of  $\lambda$ , the support is the whole real line  $\mathbb{R}$  since the entropy generating function has a non-finite derivative at 0 and 1. For  $\lambda = 0$ , we get the logistic distribution. For positive values of  $\lambda$  the support is a compact interval.  $\lambda = 1$  and  $\lambda = 2$  give uniform distributions.



**Figure A4.** Entropy generating function and density for the Tukey  $\lambda$  ( $\lambda = 0.51, 0, 1, 2$ ) and the generalized Tukey  $\lambda$  distribution ( $\lambda_1 = 0.51, 0, 1, 2$ ;  $\lambda_2 = 0$ ).

To demonstrate the consequences of skewness, the lower panels of Figure A4 show the entropy generating function and the density for  $\lambda_1 \neq \lambda_2$ . We set  $\lambda_2 = 0$  and vary only the values of  $\lambda_1$ . Now, for  $\lambda_1 \neq 0$  it is  $\Phi(1) \neq 0$ . This can be explained by the skewness of the distribution.  $\Phi$  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. [68] gave an excellent overview about the properties, parameter estimation, and applications of the generalized Tukey  $\lambda$  distribution. We recommend to study the long list of references in their paper. Also [55] discussed all these aspects in detail in his Ph.D. thesis from the University of Pretoria. He mentioned several applications in distinct scientific fields ranging from actuarial science over finance up to supply chain planning (pp. 127). Chalabi et al. [68] focused on applications in finance. King and MacGillivray [69] discussed the ordering properties of skewness and kurtosis for the generalized  $\lambda$  distribution. Concerned with estimation in this family were [70–72].

#### Appendix A.5. Fechner Approach of Skewness and Entropy Generating Function $\Phi$

There are many other proposals to introduce a skewness parameter into a symmetric distribution next to those presented in Appendix A.3. One proposal can be traced back to [73], p. 295. He proposed to use different values of a scale parameter for the left and the right half of a symmetric distribution. Klein and Fischer [74] showed that such a split of the scale parameter leads to a skewed distribution such that the skewness parameter attains the skewness ordering of [75]. Arellano-Valle et al. [76] picked up the Fechner proposal and introduced one parameter of skewness  $\gamma$  by considering functions  $f_1(\gamma)$  and  $f_2(\gamma)$  as different scale parameters for both halves of a symmetric distribution. This approach

includes the proposal made by [77] with  $f_1(\gamma) = 1/\gamma$  and  $f_2(\gamma) = \gamma$ ,  $\gamma > 0$ . Let  $f$  be a density that is symmetric around 0. Then a density with skewness parameter  $\gamma$  is given by

$$f(x, \gamma) = \frac{2\gamma}{1+\gamma^2} (f(x/\gamma)I(x < 0) + f(x\gamma)I(x \geq 0)) \quad (\text{A8})$$

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

$$F(x, \gamma) = \begin{cases} 2\gamma^2/(1+\gamma^2)F(x/\gamma, \gamma=1) & \text{for } x < 0 \\ 2/(1+\gamma^2)(F(0, \gamma=1)(\gamma^2-1) + F(x\gamma, \gamma=1)) & \text{for } x \geq 0 \end{cases} \quad (\text{A9})$$

In this section, we follow the generalized maximum cumulative  $\Phi$  entropy approach. This means to identify the entropy generating function  $\Phi$  such that (A8) is an ME distribution under the constraints of fixed mean and variance.

**Corollary A5.** Let  $f$  be a density, which is symmetric around 0 and  $Q$  the corresponding quantile function. Constraints are given by fixing mean and variance. Then, (A8) is the density of the ME distribution with entropy generating function

$$\Phi(u) = \begin{cases} -2\gamma^3/(1+\gamma^2) \int_0^u Q(v)dv & \text{for } u \leq \gamma^2/(1+\gamma^2) \\ -2/(\gamma(1+\gamma^2)) \int_{0.5}^{0.5(1+\gamma^2)u+0.5(1-\gamma^2)} Q(z)dz & \text{for } u > \gamma^2/(1+\gamma^2) \end{cases} \quad (\text{A10})$$

**Proof.** By inverting the distribution function (A9), we get the quantile function

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

Summing up this quantile function leads to the partial mean function

$$\mu(u, \gamma) = \int_0^u Q(v, \gamma)dv = \begin{cases} 2\gamma^3/(1+\gamma^2) \int_0^u Q(v)dv & \text{for } u \leq \gamma^2/(1+\gamma^2) \\ 2/(\gamma(1+\gamma^2)) \int_{0.5}^{0.5(1+\gamma^2)u+0.5(1-\gamma^2)} Q(z)dz & \text{for } u > \gamma^2/(1+\gamma^2) \end{cases} \quad (\text{A11})$$

The negative of the partial mean function determines the entropy generating function  $\Phi$  such that (A5) is an ME distribution under the constraints of given mean and variance.  $\square$

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

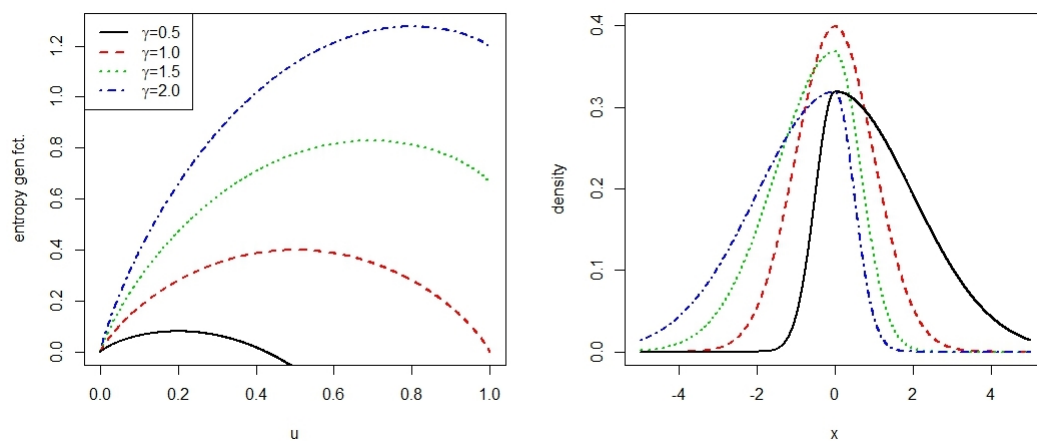
**Example A1.** Let  $f_N$  ( $F_N^{-1}$ ) be the density (quantile function) of the standard normal distribution. From  $\int_0^u F_N^{-1}(v)dv = f_N(F_N^{-1}(u))$ ,  $u \in [0, 1]$  we get

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

(see Table 3 no. 10).

Figure A5 shows the entropy generating function and the density of the skewed normal distribution for different values of the skewness parameter  $\gamma$ . The symmetric standard normal distribution ( $\gamma = 1$ ) serves as a point of reference.





**Figure A5.** Entropy generating function and density for the skewed normal distribution with parameter values  $\gamma = 0.5, 1.0, 1.5, 2$ .

#### Appendix A.6. Cumulative Residual Shannon Entropy and the Weibull Distribution

Up to now, there was no constraint on the ME distribution's support e.g., in the form  $Q(0) = 0$ . Therefore, we could always apply Theorem 1. However, as in the literature the ME task was considered mainly for situations with support  $[0, \infty)$ , we will consider this situation in the following two Sections. Thus, in this Section and Appendix A.7, 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 constraints of a given mean and more general under a given  $k$ -th power moment for  $k \geq 2$ .

At first, following the maximum cumulative  $\Phi$  entropy approach, we consider the 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 [5].

**Corollary A6.** Let  $\Phi(u) = -(1-u)\ln(1-u)$ ,  $u \in [0, 1]$ , be the entropy generating function and the constraints are that the mean  $\mu$  and the  $k$ -th power moment  $\mu'_k$  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

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

(see Table 3 no. 11).

**Proof.** The Weibull distribution is defined by a quantile function

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

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

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

and  $\lambda \equiv \mu/\Gamma(1 + 1/(k-1))$  we get the searched for quantile function of a Weibull distribution with scale parameter  $\lambda$  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  $\mu'_2/\mu^2 = \Gamma(1 + 2)/\Gamma(2)^2 = 2$ .

#### Appendix A.7. Cumulative Residual Havrda & Charvát Entropy and the Extended Burr XII Distribution

Now, we 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. Again, following the maximum cumulative  $\Phi$  entropy approach, the task is to search for the ME distribution while  $\Phi$  is given.

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

$$\Phi(u) = (1-u)((1-u)^{\alpha-1} - 1)/(1-\alpha), \quad u \in [0,1], \quad \alpha \neq 1. \quad (\text{A12})$$

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

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

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

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

Now, we look for the corresponding ME distribution if  $Q(0) = 0$  and  $\mu, \mu'_k$  are fixed. The following corollary states that this ME distribution generalizes the Weibull distribution (see Table 3 no. 12). For the sake of simple notation, we introduce a two-parametric generalization of the complete  $\Gamma$  function as

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

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

**Corollary A7.** Let  $\Phi(u) = (1-u)((1-u)^{\alpha-1} - 1)/(1-\alpha)$ ,  $u \in [0,1]$ ,  $\alpha \neq 1$  be the entropy generating function and the constraints are that the mean  $\mu$  and the  $k$ -th power moment  $\mu'_k$  are fixed. Then the corresponding ME distribution is given by the quantile function

$$Q(u) = \frac{\left( \frac{1}{1-\alpha} ((1-u)^{\alpha-1} - 1) \right)^{1/(k-1)}}{\Gamma_2(1 + 1/(k-1), \alpha - 1)} \mu, \quad u \in [0,1], \quad \alpha \neq 1 \quad (\text{A13})$$

if the following relation holds between  $\mu$  and  $\mu'_k$ :

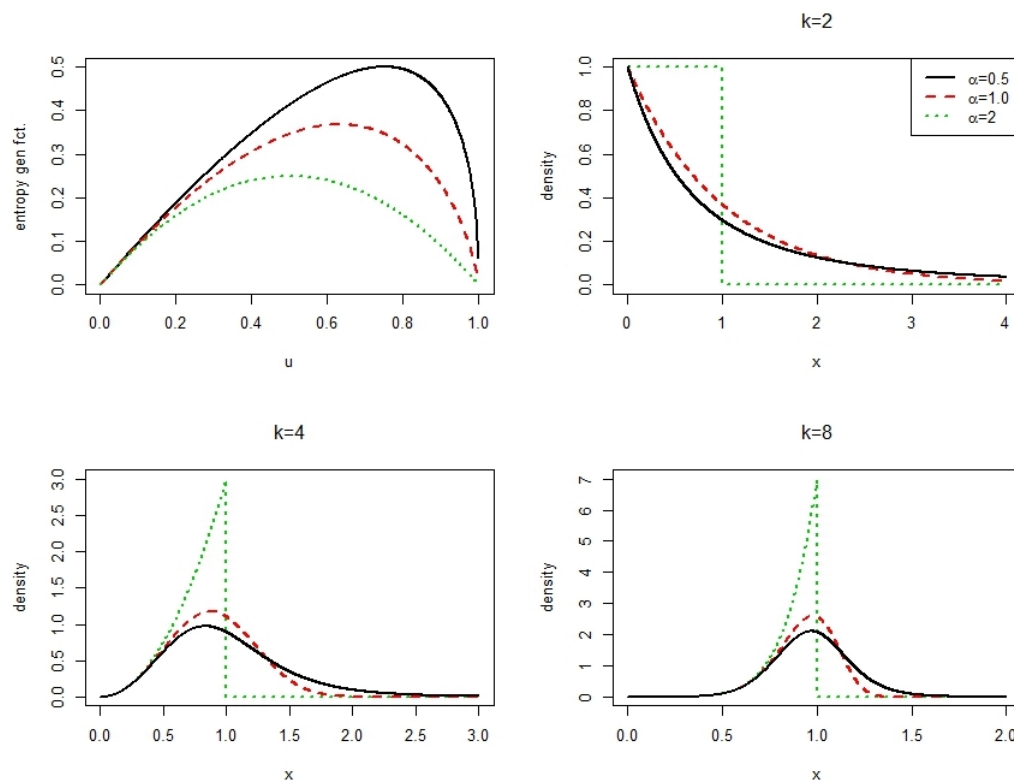
$$\frac{\mu'_k}{\mu^k} = \frac{\Gamma_2(1 + k/(k-1), \alpha - 1)}{\Gamma_2(1 + 1/(k-1), \alpha - 1)^k}. \quad (\text{A14})$$

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 A6 shows the entropy generating function (A12) as well as ME densities under the constraint of a fixed power moment  $\mu'_k$  for  $k = 2, 4, 8$ . The setting  $k = 2$  leads to a density with a shape similar to an exponential distribution. The constraint (A14) refers to a fixed coefficient of variation [4].



**Figure A6.** Entropy generating function and densities for the extended Burr XII distribution with parameter values  $\alpha = 0.5, 1, 2$  and  $k = 2, 4, 8$ .

**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.  $\square$

Mudholkar et al. [78] introduced the quantile function (A13) with a different parametrization and called the corresponding distribution ‘generalized Weibull distribution’. Due to the fact that the Burr XII distribution [79] is a special case (in our parametrization for  $\alpha = 2$ ), [80] use the term ‘extended Burr XII distribution’ (EBXII). A more recent paper that discussed this distribution is [81]. They were

concerned with parameter estimation (see also [82–84]). Applications are lifetimes of devices with a bathtub hazard rate [78], flood frequency, and duration analysis [81–83].

## References

- Hosking, J.R.M. Distributions with maximum entropy subject to constraints on their  $L$ -moments or expected order statistics. *J. Stat. Plan. Inference* **2007**, *137*, 2870–2891. [\[CrossRef\]](#)
- Park, S.Y.; Bera, A.K. Maximum entropy autoregressive conditional heteroskedasticity model. *J. Econom.* **2008**, *150*, 219–230. [\[CrossRef\]](#)
- Schroeder, M.J. An alternative to entropy in the measurement of information. *Entropy* **2004**, *6*, 388–412. [\[CrossRef\]](#)
- Rao, M.; Chen, Y.; Vemuri, B.C.; Wang, F. Cumulative residual entropy: A new measure of information. *IEEE Trans. Inf. Theory* **2004**, *50*, 1220–1228. [\[CrossRef\]](#)
- Rao, M. More on a new concept of entropy and information. *J. Theor. Probab.* **2005**, *18*, 967–981. [\[CrossRef\]](#)
- Asadi, M.; Zohrevand, Y. On the dynamic cumulative residual entropy. *J. Stat. Plan. Inference* **2007**, *137*, 1931–1941. [\[CrossRef\]](#)
- Baratpour, S. Characterizations based on cumulative residual entropy of first-order statistics. *Commun. Stat. Theory Methods* **2010**, *39*, 3645–3651. [\[CrossRef\]](#)
- Baratpour, S.; Habibi, R.A. Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. *Commun. Stat. Theory Methods* **2012**, *39*, 3645–3651. [\[CrossRef\]](#)
- Chamany, A.; Baratpour, S. A dynamic discrimination information based on cumulative residual entropy and its properties. *Commun. Stat. Theory Methods* **2014**, *43*, 1041–1049. [\[CrossRef\]](#)
- Drissi, N.; Chonavel, T.; Boucher, J.M. Generalized cumulative residual entropy distributions with unrestricted supports. *Res. Lett. Signal Process.* **2008**, *11*, 1–5. [\[CrossRef\]](#)
- Kumar, V.; Taneja, H.C. Some characterizations results on generalized cumulative residual entropy measures. *Stat. Probab. Lett.* **2011**, *81*, 1072–1077. [\[CrossRef\]](#)
- Minimol, S. On generalized dynamic cumulative past entropy measures. *Commun. Stat. Theory Methods* **2017**, *46*, 2816–2822. [\[CrossRef\]](#)
- Mirali, M.; Baratpur, S.; Fakoor, V. On weighted cumulative residual entropy. *Commun. Stat. Theory Methods* **2017**, *46*, 2857–2869. [\[CrossRef\]](#)
- Navarro, J.; del Aguila, Y.; Asadi, M. Some new results on the cumulative residual entropy. *J. Stat. Plan. Inference* **2010**, *140*, 310–322. [\[CrossRef\]](#)
- Navarro, J.; Psarrakos, G. Characterizations based on generalized cumulative residual entropy functions. *Commun. Stat. Theory Methods* **2017**, *46*, 1247–1260. [\[CrossRef\]](#)
- Psarrakos, G.; Navarro, J. Generalized cumulative residual entropy and record values. *Metrika* **2013**, *76*, 623–640. [\[CrossRef\]](#)
- Sankaran, P.G.; Sunoj, S.M. Quantile-based cumulative entropies. *Commun. Stat. Theory Methods* **2017**, *46*, 805–814. [\[CrossRef\]](#)
- Di Crescenzo, A.; Longobardi, M. On cumulative entropies. *J. Stat. Plan. Inference* **2009**, *139*, 4072–4087. [\[CrossRef\]](#)
- De Luca, A.; Termini, S. A definition of a nonprobabilistic entropy measure of fuzzy entropy. *Inf. Control* **1972**, *20*, 301–312. [\[CrossRef\]](#)
- Pal, N.R.; Bezdek, J.C. Measuring fuzzy uncertainty. *IEEE Trans. Fuzzy Syst.* **1994**, *2*, 107–118. [\[CrossRef\]](#)
- Li, P.; Liu, B. Entropy of credibility distributions for fuzzy variables. *IEEE Trans. Fuzzy Syst.* **2008**, *16*, 123–129.
- Chen, X.; Dai, W. Maximum entropy principle for uncertain variables. *Int. J. Fuzzy Syst.* **2011**, *13*, 232–236.
- Dai, W. Quadratic entropy of uncertain variables. *Soft Comput.* **2018**, *22*, 5699–2706. [\[CrossRef\]](#)
- Liu, B. *Uncertainty Theory*; Springer: New York, NY, USA, 2015.
- Havrda, J.; Charvát, F. Quantification method of classification processes. Concept of structural  $\alpha$ -entropy. *Kybernetika* **1967**, *3*, 30–35.
- Kapur, J.N. *Measures of Information and their Applications*; Wiley Eastern Limited: New Delhi, India, 1994.
- Rényi, A. On measures of entropy and information. *Fourth Berkeley Symp. Math. Stat. Probab.* **1961**, 547–561.
- Ali, S.M.; Silvey, S.D. A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc. Ser. B* **1966**, *28*, 131–142. [\[CrossRef\]](#)

29. Csiszár, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar. Tud. Akad. Mat. Kut. Int. Kozl.* **1963**, *8*, 85–108.
30. Morimoto, T. Markov processes and the H-theorem. *J. Phys. Soc. Jpn.* **1963**, *18*, 328–331. [\[CrossRef\]](#)
31. Liese, F.; Vajda, I. On divergences and information in statistics and information theory. *IEEE Trans. Inf. Theory* **2006**, *52*, 4394–4412. [\[CrossRef\]](#)
32. Cressie, N. A.; Pardo, L. Phi-divergence statistic. In *Encyclopedia of Environmetrics*; El-Shaarawi, A.H., Piegorisch, W.W., Eds.; John Wiley and Sons: New York, NY, USA, 2002; Volume 3, pp. 1551–1555.
33. Zografos, K.; Nadarajah, S. Survival exponential entropies. *IEEE Trans. Inf. Theory* **2005**, *51*, 1239–1246. [\[CrossRef\]](#)
34. Klein, I.; Mangold, B.; Doll, M. Cumulative paired  $\phi$  entropies. *Entropy* **2016**, *18*, 248. [\[CrossRef\]](#)
35. Burbea, J.; Rao, C.R. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 489–495. [\[CrossRef\]](#)
36. Kumar, V. Characterization results based on dynamic Tsallis cumulative residual entropy. *Commun. Stat. Theory Methods* **2017**, *46*, 8343–8354. [\[CrossRef\]](#)
37. Rajesh, G.; Sunoj, S.M. Some properties of cumulative Tsallis entropy of order  $\alpha$ . *Stat. Pap.* **2019**, *60*, 933–943. [\[CrossRef\]](#)
38. Tsallis, C. Possible generalization of Boltzmann. Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [\[CrossRef\]](#)
39. Liu, B. A survey of entropy of fuzzy variables. *J. Uncertain Syst.* **2007**, *1*, 4–13.
40. Verhoeff, J. Recreative Informatica. In Proceedings of the Speech on the Occasion of the 21th Dies Natalis of the Erasmus University Rotterdam, Rotterdam, The Netherlands, 9 November 1984.
41. Di Crescenzo, A.; Longobardi, M. On cumulative entropies and lifetime estimation. In *Methods and Models in Artificial and Natural Computation*; Mira, J.M., Ferrández, J.M., Alvarez-Sanchez, J.-P., Paz, F., Toledo, J., Eds.; IWINAC Part I, LNCS 5601; Springer: Berlin, Germany, 2009; pp. 132–141.
42. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [\[CrossRef\]](#)
43. Jaynes, E.T. Information theory and statistical mechanics II. *Phys. Rev.* **1957**, *108*, 171–190. [\[CrossRef\]](#)
44. Kesavan, H.K.; Kapur, J.N. The generalized maximum entropy principle. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 1042–1052. [\[CrossRef\]](#)
45. Dai, W.; Chen, X. Entropy of function of uncertain variables. *Math. Comput. Model.* **2012**, *55*, 754–760. [\[CrossRef\]](#)
46. De Sousa, A.M.C.; Tsallis, C. Student's  $t$  and  $r$ -distributions: Unified derivation from an entropic variational principle. *Physica A* **1997**, *236*, 52–57. [\[CrossRef\]](#)
47. Kapur, J.N. Generalized Cauchy and Student's distribution as maximum-entropy distributions. *Proc. Natl. Acad. Sci. India* **1988**, *58*, 235–246.
48. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **2002**, *66*, 056125. [\[CrossRef\]](#)
49. Hanel, R.; Thurner, S.T. Generalized Boltzmann factors and the maximum entropy principle: Entropies for complex systems. *Physica A* **2007**, *380*, 109–114. [\[CrossRef\]](#)
50. Azzalini, A. A class of distributions which includes the Normal Ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
51. Contreras-Reyes, J.E. Rényi entropy and complexity measure for skew-gaussian distributions and related families. *Phys. A Stat. Mech. Appl.* **2015**, *433*, 84–91. [\[CrossRef\]](#)
52. Jizba, P.; Korbel, J. Maximum entropy principle in statistical inference: Case for non-Shannonian entropies. *Phys. Rev. Lett.* **2019**, *122*, 12601. [\[CrossRef\]](#)
53. Ebanks, B.R. On measures of fuzziness and their representations. *J. Math. Anal. Appl.* **1983**, *94*, 24–37. [\[CrossRef\]](#)
54. Schlittgen, R. *Einführung in Die Statistik*; Oldenbourg: München, Germany, 1996.
55. van Staden, P.J. Modeling of Generalized Families of Probability Distributions in the Quantile Statistical Universe. Ph.D. Thesis, Department of Statistics, University of Pretoria, Pretoria, South Africa, 2013.
56. Mielke, P.W., Jr. Asymptotic behavior of two-sample tests based on powers of ranks for detecting scale and location alternatives. *J. Am. Stat. Assoc.* **1972**, *67*, 850–854. [\[CrossRef\]](#)
57. Mielke, P.W., Jr.; Johnson, E.S. Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests. *Mon. Weather Rev.* **1973**, *101*, 701–707. [\[CrossRef\]](#)
58. Ansari, A.R.; Bradley, R.A. Rank-sum tests for dispersion. *Ann. Math. Stat.* **1960**, *31*, 1147–1189. [\[CrossRef\]](#)

59. Mood, A.M. On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Stat.* **1954**, *25*, 514–522. [[CrossRef](#)]
60. Hosking, J.R.M. *The Theory of Probability Weighted Moments*; Research Report RC12210; IBM Research: Yorktown Heights, NY, USA, 1986.
61. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on L-Moments*; Cambridge University Press: Cambridge, UK, 1997.
62. Shabri, A.; Jemain, A.A. Fitting the generalized logistic distribution by LQ-Moments. *Appl. Math. Sci.* **2011**, *5*, 2663–2676.
63. Morais, A.L.; Cordeiro, G.M. A Four-Parameter Generalized Logistic Distribution. 2010. Available online: <http://www.ime.unicamp.br/sinape/sites/default/files/Lemos-Cordeiro.pdf> (accessed on 11 January 2020).
64. Nassar, M.M.; Elmasry, A. A study of the generalized logistic distributions. *J. Egypt. Math. Soc.* **2012**, *20*, 126–133. [[CrossRef](#)]
65. Tripathi, Y.M.; Mahto, A.K.; Dey, S. Efficient estimation of the PDF and the CDF of a generalized logistic distribution. *Ann. Data Sci.* **2017**, *4*, 63–81. [[CrossRef](#)]
66. van Staden, P.J.; King, R.A.R. The quantile-based skew logistic distribution. *Stat. Probab. Lett.* **2014**, *96*, 109–116. [[CrossRef](#)]
67. Freimer, M.; Kollia, G.; Mudholkar, G.S.; Lin, C.T. A study of the generalized Tukey lambda family. *Commun. Stat. Theory Methods* **1988**, *17*, 3547–3567. [[CrossRef](#)]
68. Chalabi, Y.; Scott, J.; Wuertz, D. Flexible Distribution Modeling with the Generalized Lambda Distribution. 2012. Available online: <https://mpira.ub.uni-muenchen.de/id/eprint/43333> (accessed on 11 January 2020).
69. King, R.A.R.; MacGillivray, H.L. A starship estimation method for the generalized  $\lambda$  distributions. *Aust. N. Z. J. Stat.* **1999**, *41*, 353–374. [[CrossRef](#)]
70. Su, S. Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R. *J. Stat. Softw.* **2007**, *21*, 1–17. [[CrossRef](#)]
71. Su, S. GLDEX: Fitting Single and Mixture of Generalized Lambda Distribution (RS and FMKL) Using Discretized and Maximum Likelihood Methods. R Package Version 1.044. 2012. Available online: <http://CRAN.R-project.org/package=GLDEX> (accessed on 11 January 2020).
72. van Staden, P.J.; Loots, M.T. Method of  $L$  moment estimation for the generalized lambda distribution. In Proceedings of the Third Annual Applied Statistics and Research Collaboration Research Conference, Callaghan, Australia, 7–8 December 2009.
73. Fechner, G.T. *Kollektivmaßlehre*; Engelmann: Leipzig, Germany, 1897.
74. Klein I.; Fischer, M. Skewness by splitting the scale parameter. *Commun. Stat. Theory Methods* **2006**, *35*, 1159–1171. [[CrossRef](#)]
75. van Zwet, W.R. *Convex Transformations of Random Variables*; Mathematical Centre Tracts No. 7; Mathematical Centre: Amsterdam, The Netherlands, 1964.
76. Arellano-Valle, R.B.; Gomez, H.W.; Quintana, F.A. Statistical inference for a general class of asymmetric distributions. *J. Stat. Plan. Inference* **2005**, *128*, 427–443. [[CrossRef](#)]
77. Fernandez, C.; Osiewalski, J.; Steel, M.F.J. Modelling and inference with  $\nu$ -spherical distributions. *J. Am. Stat. Assoc.* **1995**, *90*, 1331–1340.
78. Mudholkar, G.S.; Srivastava, D.K.; Kollia, G.D. A generalization of the Weibull distribution with applications to the analysis of survival data. *J. Am. Stat. Assoc.* **1996**, *91*, 1575–1583. [[CrossRef](#)]
79. Burr, I.W. Cumulative frequency functions. *Ann. Math. Stat.* **1942**, *13*, 215–232. [[CrossRef](#)]
80. Shao, Q. Notes on the maximum likelihood estimation for the three-parameter Burr XII distribution. *Comput. Stat. Data Anal.* **2004**, *45*, 675–687. [[CrossRef](#)]
81. Ganora, D.; Liao, F. Hydrological applications of the Burr distribution: Practical method for parameter estimation. *J. Hydrol. Eng.* **2015**, *20*, 04015024. [[CrossRef](#)]
82. Nadarajah, S.; Kotz, S. Discussion of models for extremes using the extended three parameter Burr XII system with applications to the flood frequency analysis. *Hydrol. Sci. J.* **2006**, *51*, 1203–1204. [[CrossRef](#)]



83. Shao, Q.; Wong, H.; Xia, J.; Ip, W. Models for extremes using the extended three-parameter Burr XII system with application to flood frequency analysis. *Hydrol. Sci. J.* **2004**, *49*, 685–702. [[CrossRef](#)]
84. Usta, I. Different estimation methods for the parameters of the extended Burr XII distribution. *J. Appl. Stat.* **2013**, *40*, 397–414. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).