

Article

Some Technical Remarks on Negations of Discrete Probability Distributions and Their Information Loss

Ingo Klein

Department of Statistics and Econometrics, Friedrich-Alexander Universität Erlangen-Nürnberg, Lange Gasse 20, D-90403 Nürnberg, Germany; ingo.klein@fau.de

Abstract: Negation of a discrete probability distribution was introduced by Yager. To date, several papers have been published discussing generalizations, properties, and applications of negation. The recent work by Wu et al. gives an excellent overview of the literature and the motivation to deal with negation. Our paper focuses on some technical aspects of negation transformations. First, we prove that independent negations must be affine-linear. This fact was established by Batyrshin et al. as an open problem. Secondly, we show that repeated application of independent negations leads to a progressive loss of information (called monotonicity). In contrast to the literature, we try to obtain results not only for special but also for the general class of ϕ -entropies. In this general framework, we can show that results need to be proven only for Yager negation and can be transferred to the entire class of independent (=affine-linear) negations. For general ϕ -entropies with strictly concave generator function ϕ , we can show that the information loss increases separately for sequences of odd and even numbers of repetitions. By using a Lagrangian approach, this result can be extended, in the neighbourhood of the uniform distribution, to all numbers of repetition. For Gini, Shannon, Havrda–Charvát (Tsallis), Rényi and Sharma–Mittal entropy, we prove that the information loss has a global minimum of 0. For dependent negations, it is not easy to obtain analytical results. Therefore, we simulate the entropy distribution and show how different repeated negations affect Gini and Shannon entropy. The simulation approach has the advantage that the entire simplex of discrete probability vectors can be considered at once, rather than just arbitrarily selected probability vectors.

Keywords: negation; Gini entropy; Shannon entropy; Havrda–Charvát (Tsallis) entropy; ϕ -entropy; Rényi entropy; Sharma–Mittal entropy; (h, ϕ) -entropy; Dirichlet distribution; Monte Carlo simulation

MSC: 62E10; 94A16



Citation: Klein, I. Some Technical Remarks on Negations of Discrete Probability Distributions and Their Information Loss. *Mathematics* **2022**, *10*, 3893. <https://doi.org/10.3390/math10203893>

Academic Editor: Christophe Chesneau

Received: 19 September 2022

Accepted: 18 October 2022

Published: 20 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In knowledge-based systems, terms with n categories can be characterized by probability distributions. Let us consider the term “conservative” and assume that we know how the conservative population is distributed among the three categories “right-wing conservative”, “conservative” and “liberal-conservative”. Can we learn anything about the non-conservative population from this distribution by looking at the negation of the original distribution? If so, does this not necessarily involve a loss of information, because the original distribution makes no explicit statement about non-conservatism? Can the loss of information be measured by any information measure? These are questions that are still the subject of intense debate that starts with the seminal work of Yager [1]. He proposed to define negation of a probability distribution by subtracting the probability distribution from 1 and distributing the sum of negated probabilities equally among the n categories. The equal distribution can be motivated by Dempster–Shafer theory and a maximal entropy argument. Technically, the equal distribution means that the negation of one category’s probability must not depend on the probabilities of the other categories. Batyrshin et al. [2] call a negation with this property “independent”. Yager negation and all

other affine-linear negations are independent. Batyrshin [3] stated this as an open problem to show that independent negations must be affine-linear. One of our technical remarks concerns the solution of this open problem.

In particular, for Yager negation, information content and information loss have been measured by different entropies. Yager [1] discussed Gini entropy. This entropy is very popular because no complicated calculations are required. However, the question arises whether the results obtained for Gini entropy can be transferred to other entropies. Therefore, Gao and Deng [4,5] have considered Shannon and Havrda–Charvát entropy. Zhang et al. [6] also studied Havrda–Charvát entropy and Wu et al. [7] Shannon entropy for exponential negations. Srivastava and Maheswari [8] introduced a new kind of entropy tailored for Yager negation. Their proposal is also based on Shannon entropy. All authors concluded that negation leads to an information loss. They use a Lagrangian approach to reach this conclusion. Because the articles mentioned mostly deal with the application of negations in various scientific fields, the more technical aspects seem to be of less interest. In particular, the sufficient conditions for optimization have not been investigated. Therefore, our technical remarks are intended to complete the proofs and generalise the results to the class of ϕ -entropies with strictly concave generating function ϕ .

Negations can be applied not only to the original probability vector, but also to the negation of a probability vector because negation again gives a probability distribution. For example, consider a sequence of recursive application of negation of length $k + 1$. Then, Yager [1] showed for Yager negation and Batyrshin et al. [3] proved for general affine-linear negations that the $k + 1$ -times recursively applied negation is given by a recursion relation that updates the k -times recursively repeated negation by the uniform distribution. Convergence against the uniform distribution with $k \rightarrow \infty$ is easy to show. We ask whether this updating rule also applies to entropy of negation. In the tradition of Yager's work, we focus on Gini entropy as the preferred information measure. We show that Gini entropy of $k + 1$ -times repeated independent negations is a convex combination of the Gini entropy of the uniform distribution and the Gini entropy of the k -times recursively repeated negations. It is well known that the uniform distribution maximises the Gini entropy and represents the point of minimum information, maximum dispersion, or maximum uncertainty. Therefore, this updating formula ensures that negation leads to an information loss as suspected in the introductory discussion of "conservatism".

The work of Gao and Deng [4,5], Zhang et al. [6] and Wu et al. [7] illustrates the entropy behaviour in terms of negation by numerical examples. Typical probability vectors were selected and Yager or exponential negation were applied to this vector. As an alternative, we propose a numerical procedure to compare ϕ -entropies for all possible probability vectors. The numerical way is to draw probability vectors of size n from the Dirichlet distribution and to simulate the entropy distribution. This procedure is particularly recommended for dependent negations. In our opinion, there seems to be no analytical way to discuss the behaviour of dependent negations and the corresponding ϕ -entropies. For all negations considered, the entropy distribution is more or less concentrated below the entropy's maximum value. What we can learn is that negations that lie above Yager negation (like exponential negation discussed in [7]) give more concentrated and peaked entropy distributions and negations that lie below the Yager negation (like the Tsallis negation with parameter $1/2$ [6]) give more spread and less peaked distributions. This confirms the statement of Wu et al. [7] that recursively repeated exponential negations converge faster to the uniform distribution than recursively repeated Yager negations.

The specific aims of our paper are as follows.

1. It will be proven that independent negations have to be affine-linear.
2. It will be proven that the uniform distribution maximizes all ϕ -entropies with strictly concave generating function ϕ .
3. It will be proven that the Yager negation minimizes any ϕ -entropy in the class of affine-linear negations, with the consequence that the information loss has to be discussed only for Yager negation.

4. It will be proven that the information loss, measured by the difference of ϕ -entropies, increases separately for odd and even sequences of repetition numbers of Yager negation.
5. It will be proven that the uniform distribution yields a local minimum of the information loss produced by the Yager negation for each ϕ -entropy.
6. It will be proven that the uniform distribution yields the global minimum of the information loss produced by the Yager negation for Gini, paired Shannon, Shannon and Havrda–Charvát entropy.
7. An explicit formula for Yager negation’s information loss in the case of the Gini entropy will be given.
8. It will be shown that results concerning the information loss of Havrda–Charvát entropy can be transferred to Rényi and Sharma–Mittal entropies by using the concept of (h, ϕ) -entropies with h strictly increasing.
9. An impression of how the information loss behaves for dependent negations when analytical results do not seem to be available will be given.

The paper is organised along the lines of the abovementioned objectives. After some definitions in the first section, we show in the second section that independent negations must be affine-linear. The third section introduces ϕ -entropies and shows that they are maximized for the uniform distribution. Moreover, the Yager negation will be identified as a minimum ϕ entropy representation of the class of affine-linear negations. The remainder of the section discusses the information loss of Yager negation for general ϕ -entropies. Approaches based on the strict concavity of ϕ as well as the Lagrangian approach will be considered. In the fourth section, we elaborate on the results concerning the information loss for some prominent ϕ -entropies. The impression could be given that all ϕ -entropies behave similarly with respect to a negation. Therefore, we present some examples of entropies wherein negations behave differently. In the fifth section, we discuss the information loss for (h, ϕ) -entropies with strictly increasing function h . In the sixth section, we present the results of a simulation study for the information loss of dependent negations. The seventh section summarizes the main results. Three proofs refer to the same bordered Hessian matrix. For this reason, this matrix is dealt with in Appendix A. Three proofs are moved out to Appendix B to improve the readability.

2. Definitions

Let

$$\mathbf{I}_n = \left\{ (p_1, \dots, p_n) \in [0, 1]^n \mid \sum_{i=1}^n p_i = 1 \right\} \tag{1}$$

be the probability simplex of size $n \geq 1$. It contains all discrete probability distributions with support of size n . A probability transformation maps a discrete probability distribution to a new discrete probability distribution with support of the same size.

Definition 1.

1. A probability transformation \mathbf{T} maps \mathbf{I}_n into \mathbf{I}_n such that

$$\mathbf{T}(p_1, \dots, p_n) = (T_1(p_1, \dots, p_n), \dots, T_n(p_1, \dots, p_n)) \tag{2}$$

with $T_i(p_1, \dots, p_n) \geq 0, i = 1, 2, \dots, n$ and $\sum_{i=1}^n T_i(p_1, \dots, p_n) = 1$ for all $(p_1, \dots, p_n) \in \mathbf{I}_n$.

2. The probability transformation \mathbf{T} is independent, if there exists a function $T : [0, 1] \rightarrow [0, 1]$ such that

$$\mathbf{T}(p_1, \dots, p_n) = (T(p_1), \dots, T(p_n)) \tag{3}$$

for all $(p_1, \dots, p_n) \in \mathbf{I}_n$.

For an independent transformation, the function T_i depends exactly on the i -th component p_i and not on the other probabilities $p_j, j \neq i$ for $j = 1, 2, \dots, n$.

Three examples, to be discussed in detail later, are the exponential, the cosine, and the affine-linear transformation. The first two are not and the last is independent.

Example 1.

1. In Wu et al. [7] an exponential probability transformation with

$$T_i(p_1, \dots, p_n) = \frac{e^{-p_i}}{\sum_{l=1}^n e^{-p_l}}, \quad i = 1, 2, \dots, n \tag{4}$$

was considered.

2. Another example could be the cosinus transformation

$$T_i(p_1, \dots, p_n) = \frac{1 + \cos(\pi p_i)}{n + \sum_{l=1}^n \cos(\pi p_l)}, \quad i = 1, 2, \dots, n. \tag{5}$$

3. Affine-linear probability transformations are given by

$$T_i(p_1, \dots, n) = \frac{a + bp_i}{na + b} =: T(p_i) \tag{6}$$

for a and b such that $0 \leq T(p_i) \leq 1, i = 1, 2, \dots, n$.

The transformations (4), (5) and the affine-linear transformation (6) with $b < 0$ are negative transformations (called negations) in the following sense.

Definition 2.

1. The probability transformation $\mathbf{N} : \mathbf{I}_n \rightarrow \mathbf{I}_n$ with

$$\mathbf{N}(p_1, \dots, p_n) = (N_1(p_1, \dots, p_n), \dots, N_n(p_1, \dots, p_n))$$

is called a negation, if

$$p_i \leq p_j \implies N_i(p_1, \dots, p_n) \geq N_j(p_1, \dots, p_n) \tag{7}$$

for all $i, j = 1, 2, \dots, n$ and $(p_1, \dots, p_n) \in \mathbf{I}_n$.

2. For independent negations \mathbf{N} there exists a function $N : [0, 1] \rightarrow [0, 1]$ with

$$\mathbf{N}(p_1, \dots, p_n) = (N(p_1), \dots, N(p_n)), \quad (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{8}$$

N will be called a negator [3].

We will highlight two affine-linear negations that play a central role in what follows.

Example 2. An affine-linear negator is given by

$$N(p) = \frac{a + bp}{na + b}, \quad 0 \leq p \leq 1$$

with $b \leq 0$. Yager [1] discussed the special case $b = -1$ which is now called Yager negator. For $b = 0$, we obtain the uniform negator characterizing the uniform distribution $(1/n, \dots, 1/n)$.

The transformations (4), (5) and (6) are of the form

$$T_i(p_1, \dots, p_n) = \frac{f(p_i)}{\sum_{l=1}^n f(p_l)}, \quad (p_1, \dots, p_n) \in \mathbf{I}_n \tag{9}$$

generated by a function $f : [0, 1] \rightarrow (0, \infty)$. The division by $\sum_{l=1}^n f(p_l)$ ensures that the image of the transformation again results in a discrete probability distribution.

If f has a negative slope, f generates a negation

$$\mathbf{N}(\mathbf{p}) = \left(\frac{f(p_1)}{\sum_{l=1}^n f(p_l)}, \dots, \frac{f(p_n)}{\sum_{l=1}^n f(p_l)} \right), \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{10}$$

Even though (9) is a function of the whole vector of probabilities, Batyrshin [3] defines

$$N(p_i) := \frac{f(p_i)}{\sum_{l=1}^n f(p_l)}, \quad (p_1, \dots, p_n) \in \mathbf{I}_n \tag{11}$$

and calls the N negator as well. This small formal incorrectness simplifies the notation and will be used in the following. Furthermore, we speak of an independent negator if the corresponding negation is independent. This formulation also follows Batyrshin [3].

3. Independence and Linearity

The negator (11) can only be independent, if

$$\sum_{j=1}^n f(p_j) = \sum_{j=1}^n f(q_j)$$

holds for $(p_1, \dots, p_n), (q_1, \dots, q_n) \in \mathbf{I}_n$. This means that there is constant $c \neq 0$ with

$$\sum_{l=1}^n f(p_l) = c, \quad (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{12}$$

The constant c can be characterized more precisely if the special discrete probability distribution $(1, 0, \dots, 0)$ is inserted into (12):

$$c = f(1) + (n - 1)f(0), \quad \text{resp. } 1 = N(1) + (n - 1)N(0). \tag{13}$$

Batyrshin et al. [2] and Batyrshin [3] showed many properties of independent negators and of affine-linear negators with generating function $f(p) = a + bp$ with $a = f(0)$ and $b = f(1) - f(0)$. It is easy to verify that affine-linear negators are independent. Batyrshin et al. [2] identified this as an open problem to show that the converse is also true. The following theorem gives a proof that independent negators must be affine-linear. For simplicity, we assume that f is continuous. The proof can be based on weaker assumptions.

Theorem 1. *Let $f : [0, 1] \rightarrow [0, \infty)$ be a continuous function. Then the unique independent negator is generated by*

$$f(p) = f(0) + (f(1) - f(0))p, \quad 0 \leq p \leq 1 \tag{14}$$

with $f(1) < f(0)$.

Proof. From (12) and (13) it follows that

$$\sum_{l=1}^n f(p_l) = f(1) + (n - 1)f(0)$$

resp.

$$\sum_{l=1}^{n-1} f(p_l) = f(1) + (n - 1)f(0) - f(p_n).$$

Consider $n = 2$ with $p_1 = p$ and $p_2 = 1 - p$, this results in

$$f(1 - p) = f(1) + f(0) - f(p). \tag{15}$$

For general n , it is $\sum_{l=1}^{n-1} p_l = 1 - p_n$ and by inserting (15) for $p = \sum_{l=1}^{n-1} p_l$ we get

$$\begin{aligned} \sum_{l=1}^{n-1} f(p_l) &= f(1) - (n - 1)f(0) - f\left(1 - \sum_{l=1}^{n-1} p_l\right) \\ &= f(1) - (n - 1)f(0) + f\left(\sum_{l=1}^{n-1} p_l\right) - f(1) + f(0), \end{aligned}$$

such that

$$\sum_{l=1}^{n-1} (f(p_l) - f(0)) = f\left(\sum_{l=1}^{n-1} p_l\right) - f(0).$$

Define $g(p) := f(p) - f(0)$, and then we get the Cauchy functional equation

$$g\left(\sum_{l=1}^{n-1} p_l\right) = \sum_{l=1}^{n-1} g(p_l)$$

for $n = 2, 3, \dots$. If f and therefore g are continuous, it is well-known that the Cauchy functional equation has the unique solution ([9], p. 51) $g(p) = kp$. This gives $f(p) = f(0) + kp$ for $0 \leq p \leq 1$ with k determined by

$$\sum_{l=1}^n f(p_l) = nf(0) + k = (n - 1)f(0) + f(1)$$

as $k = f(1) - f(0)$. The unique independent negator is generated by

$$f(p) = f(0) + (f(1) - f(0))p, \quad 0 \leq p \leq 1.$$

□

The corresponding affine-linear negator generated by (14) is

$$N(p) = \frac{f(0) + (f(1) - f(0))p}{(n - 1)f(0) + f(1)}, \quad 0 \leq p \leq 1. \tag{16}$$

The following collection of properties for affine-linear negators has already been proven by Batyrshin et al. [2]. These properties are needed to show some results concerning the entropy of negations. Set

$$A := N(0) = \frac{f(0)}{(n - 1)f(0) + f(1)} \quad \text{and} \quad B = N(1) - N(0) = \frac{f(1) - f(0)}{(n - 1)f(0) + f(1)}$$

and then $N(p) = A + Bp$ for $0 \leq p \leq 1$.

Remark 1.

1. $-1/(n - 1) \leq B \leq 0$.
2. $N(p) \in [0, 1/n]$ for $p \geq 1/n$, $N(p) \in [1/n, 1/(n - 1)]$ for $p \leq 1/n$. This means that, depending on n , an independent negator takes on values in a very small interval.
3. $N(p)$ can be equivalently represented as the convex combination of the constant (or uniform) negator $N_U(p) = 1/n$ and the Yager negator $N_Y(p) = (1 - p)/(n - 1)$ for $0 \leq p \leq 1$. This means, there exists an $\alpha \in [0, 1]$ such that

$$N(p) = \alpha N_U(p) + (1 - \alpha)N_Y(p), \quad 0 \leq p \leq 1. \tag{17}$$

For an affine-linear negator N , there is an alternative representation as difference between the negator and the uniform distribution:

$$N(p) - \frac{1}{n} = B\left(p - \frac{1}{n}\right), \quad 0 \leq p \leq 1. \tag{18}$$

Yager [1] already considered the repeated use of negation. According to Batyrshin [3], $N^{(k+1)}(p)$ denotes the negator after $k + 1$ -times repeated use of N . By using (18) and by induction it is easy to show

$$N^{(k+1)}(p) - \frac{1}{n} = B^{k+1}\left(p - \frac{1}{n}\right) = B\left(N^{(k)}(p) - \frac{1}{n}\right), \quad 0 \leq p \leq 1, \tag{19}$$

or, equivalently,

$$N^{(k+1)}(p) = (1 - B^{k+1})\frac{1}{n} + B^{k+1}p = (1 - B)\frac{1}{n} + BN^{(k)}(p), \quad 0 \leq p \leq 1 \tag{20}$$

for $k = 0, 1, 2, 3, \dots$ [2]. The recursion relation (20) starts with $B^{(0)}(p) = p$. From Remark 1, we know that $|B| < 1$, such that $B^k \rightarrow 0$ for $k \rightarrow \infty$. This means that $(N^k(p) - 1/n)_{k=1,2,\dots}$ is an alternating sequence converging to 0.

4. Information Loss for Independent Negations and General ϕ -Entropies

4.1. ϕ -Entropies

Yager [1] decided to discuss the Gini entropy based on its simple formula. This leads to the question whether the results proven for the Gini entropy also apply to other entropies. In the literature, the Shannon and the so-called Havrda–Charvát (or Tsallis) entropy have been the focus of discussion. These entropies are special cases of a broader class of entropies. In a recent paper, Ilić et al. [10]) gave an excellent overview of what they call “generalized entropic forms”. For our purposes, it is sufficient to consider ϕ -entropies introduced by Burbea and Rao [11] and (h, ϕ) -entropies introduced by Salicrú et al. [12].

Definition 3. Let $\phi : [0, 1] \rightarrow [0, \infty)$ be strictly concave on $[0, 1]$ with $\phi''(p) < 0$ for $0 \leq p \leq 1$. Then,

$$H_\phi(\mathbf{p}) = \sum_{i=1}^n \phi(p_i), \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n \tag{21}$$

is called ϕ -entropy with generating function ϕ .

Examples for ϕ -entropies are

Example 3.

1. Gini (or quadratic) entropy: $\phi(u) = u(1 - u)$ [13–16].
2. Shannon entropy: $\phi(u) = -u \ln u$ [17].
3. Havrda–Charvát (or Daróczy or Tsallis) entropy: $\phi(u) = -u(u^{q-1} - 1)/(q - 1)$ [18–20].
4. Paired Shannon entropy: $\phi(u) = -u \ln u - (1 - u) \ln(1 - u)$ [11].
5. Paired Havrda–Charvát (or Tsallis) entropy: $\phi(u) = -u(u^{q-1} - 1)/(q - 1) - (1 - u)((1 - u)^{q-1} - 1)/(q - 1)$ [11].
6. Modified paired Shannon entropy: $\phi(u) = -u \ln u - (1/(n - 1) - u) \ln(1/(n - 1) - u)$ [8].
7. Leik entropy: $\phi(u) = \min\{u, 1 - u\}$ [21,22].
8. Entropy introduced by Shafee: $\phi(u) = -u^\alpha \ln u$ [23].

Notice that the entropy-generating functions of the Leik entropy is only concave and not strictly concave, and the entropy-generating function of the entropy introduced by

Shafee is strictly concave only for $1/2 \leq \alpha \leq 1$. We will discuss both entropies in detail in the Examples 4 and 5.

A basic axiom that all entropies should satisfy is that they must be maximal for the uniform distribution $p_i = 1/n, i = 1, 2, \dots, n$. This axiom is part of the Shannon–Khinchin axioms that justify Shannon entropy. However, we do not choose an axiomatic approach. To show that the definition of an entropy is useful, we have at least to prove that the uniform distribution maximizes this entropy. This will be done for ϕ -entropies. It seems useful to unify the partially incomplete proofs given separately in the literature for individual entropies [7,8].

Theorem 2. Let $\phi : [0, 1] \rightarrow [0, \infty)$ be strictly concave on $[0, 1]$ with $\phi''(p) < 0$ for $0 \leq p \leq 1$ and H_ϕ the corresponding ϕ entropy. Then,

$$H_\phi(\mathbf{p}) \leq H_\phi(1/n, \dots, 1/n), \quad \mathbf{p} \in \mathbf{I}_n$$

and $n = 2, 3, \dots$

Proof. Consider the Lagrangian function

$$L(\mathbf{p}, \lambda) = \sum_{i=1}^n \phi(p_i) + \lambda \left(\sum_{i=1}^n p_i - 1 \right), \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n \tag{22}$$

and the derivatives

$$\frac{\partial L(\mathbf{p}, \lambda)}{\partial p_i} = \phi'(p_i) + \lambda$$

resp.

$$\frac{\partial^2 L(\mathbf{p}, \lambda)}{\partial p_i \partial p_j} = \begin{cases} \phi''(p_i) & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

for $i, j = 1, 2, \dots, n$ and $\mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n$. The necessary condition for an optimum (for a stationary point) is $\phi'(p_i) = -\lambda, i = 1, 2, \dots, n$. ϕ' has to be strictly decreasing because $\phi''(p) < 0, 0 \leq p \leq 1$. This means that the inverse function $(\phi')^{-1}$ exists. It is

$$p_i = p_j = (\phi')^{-1}(\lambda), \quad i = 1, 2, \dots, n.$$

All probabilities are identical and add up to 1 such that $p_i = 1/n, i = 1, 2, \dots, n$. The uniform distribution is the only stationary point of the Lagrangian function (22). To show that this stationary point belongs to a global maximum of (22) we consider the bordered Hessian matrix (see Lemma A1) with $x_i := \phi''(p_i) < 0, i = 1, 2, \dots, n$. Let Λ_m denote the determinant of the upper left $m \times m$ -matrix for $m = 3, 4, \dots, n + 1$. For a local maximum, we have to show that $(-1)^{m-1} \Lambda_m > 0$ for $m = 3, 4, \dots, n + 1$ [24] (p. 203). In the appendix (see Lemma A1), we prove that

$$\Lambda_m = - \sum_{I_{m-2}} \prod_{j=1}^{m-2} \phi''(p_i)$$

with

$$I_{m-2} = \{(i_1, \dots, i_{m-2}) \in \{1, 2, \dots, m\}^{m-1} | i_1 < i_2 < \dots < i_{m-2}\}$$

for $m = 3, 4, \dots, n + 1$. For m odd, it is $\prod_{j=1}^{m-2} \phi''(p_i) < 0$ and $(-1)^{m-1} (-\prod_{j=1}^{m-2} \phi''(p_i)) > 0$. For m even, we get $\prod_{j=1}^{m-2} \phi''(p_i) > 0$ and $(-1)^{m-1} (-\prod_{j=1}^{m-2} \phi''(p_i)) > 0$ such that $(-1)^{m-1} \Lambda_m > 0$ for $m = 3, 4, \dots, n + 1$. Therefore, the uniform distribution is not only a stationary point, but also the point where (22) has a global maximum. \square

The importance of the fact that the entropy-generating function is strictly concave shall be illustrated by two examples.

Example 4. We consider the Leik entropy [21,22]

$$H_L(\mathbf{p}) = \sum_{i=1}^n \min\{p_i, 1 - p_i\} = \sum_{i=1}^n \frac{1}{2} - \left| p_i - \frac{1}{2} \right|, \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \quad (23)$$

The generating function corresponding to (23) is $\phi(u) = \min\{u, 1 - u\}$, $0 \leq u \leq 1$. ϕ is concave but not strictly concave. If $p_i < 1/2$ for $i = 1, 2, \dots, n$, then it is $H_L(p_1, \dots, p_n) = 1$. For independent negators N we know from Remark 1 that $N(p) \leq 1/(n - 1) \leq 1/2$, $0 \leq p \leq 1$ and $n > 2$. This means that the Leik entropy cannot distinguish between the information content of different independent negations and different numbers of repetitions of independent negations.

Example 5. Let us consider an entropy introduced by Shafee [23] as

$$-\sum_{i=1}^n p_i^\alpha \ln p_i, \quad (p_1, \dots, p_n) \in \mathbf{I}_n. \quad (24)$$

This entropy is part of the family of Sharma–Taneja–Mittal entropies [10] and a simple generalization of Shannon entropy ($\alpha = 1$). The entropy generating function is

$$\phi(u) = -u^\alpha \ln u \quad 0 \leq u \leq 1$$

with derivatives

$$\phi'(u) = -u^{\alpha-1}(\alpha \ln u + 1), \quad 0 \leq u \leq 1$$

and

$$\phi''(u) = -u^{\alpha-2}(\alpha(\alpha - 1) \ln u + 2\alpha - 1), \quad 0 \leq u \leq 1.$$

ϕ is strictly concave for $1/2 \leq \alpha \leq 1$.

Now, we want to show that (24), depending on n , has a local maximum or a local minimum for $(1/n, \dots, 1/n)$. If we solve $\phi''(p^*) = 0$ then

$$p^* = e^{-1/\alpha-1/(\alpha-1)}.$$

Consider $\alpha < 1/2$. Then it is $p^* < 1$ with $\phi''(p) < 0$ for $p < p^*$ and $\phi''(p) > 0$ for $p > p^*$.

If n is large enough such that $1/n < p^*$, then $\phi''(1/n) < 0$ and, following the proof of Theorem 2, $(1/n, \dots, 1/n)$ gives a local maximum of (24). For $\alpha = 0.2$, we get $p^* = 0.0235$ such that $1/n < 0.0235$ for $n > 42$.

For smaller n with $1/n > p^*$, it is $\phi''(1/n) > 0$. We again consider the bordered Hessian matrix (see Lemma A1). Λ_m denotes the determinant of the upper left $m \times m$ -matrix for $m = 3, 4, \dots, n + 1$. For a local minimum, we have to show that $(-1)\Lambda_m > 0$ for $m = 3, 4, \dots, n + 1$ [24] (p. 203) with

$$\Lambda_m = -\sum_{I_{m-2}} \prod_{j=1}^{m-2} \phi''(1/n)$$

and

$$I_{m-2} = \{(i_1, \dots, i_{m-2}) \in \{1, 2, \dots, m\}^{m-1} | i_1 < i_2 < \dots < i_{m-2}\}$$

for $m = 3, 4, \dots, n + 1$. With $\phi''(1/n) > 0$, we immediately see that $-1\Lambda_m > 0$ for $m = 3, 4, \dots, n + 1$.

For $\alpha < 1$, $\phi''(p) < 0$ for $p > p^*$ and $\phi''(p) > 0$ for $p < p^*$. By using this property, a similar argument can be applied to show that the uniform distribution can be a point at which (24) has either a local maximum or a local minimum.

The important overall result is that the uniform distribution does not maximize (24) for all n .

Let \mathbf{N} be a negation, and then we want to check the property of monotonicity

$$H_\phi(\mathbf{N}(\mathbf{p})) \geq H_\phi(\mathbf{p}) \quad (25)$$

in a most general setting. In other words, this means that the information loss caused by the negation

$$H_\phi(\mathbf{N}(\mathbf{p})) - H_\phi(\mathbf{p}) \tag{26}$$

is non-negative.

From (25) we can conclude

$$H_\phi(\mathbf{N}(\mathbf{p})) \geq H_\phi(\mathbf{p}) \implies H_\phi(\mathbf{N}^{(k+1)}(\mathbf{p})) \geq H_\phi(\mathbf{N}^{(k)}(\mathbf{p})), \quad k = 1, 2, \dots$$

$\mathbf{N}(\mathbf{p})$ can be considered as a new vector of probabilities for which (25) holds and so on.

Table 1 gives an overview about the results concerning (25) that will be proven in the following. Without loss of generality, we can restrict the discussion to Yager negations as will be shown in the next section.

Table 1. Property of monotonicity for several entropies H .

Entropy	Source	Result	Method
ϕ	Theorem 4	$H(\mathbf{N}_Y^{(2)}(\mathbf{p})) \geq H(\mathbf{p})$	strict concavity
ϕ	Theorem 5	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$ for $\mathbf{p} = (1/n, \dots, 1/n)$	Lagrange
paired Shannon	Theorem 6	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	Lagrange
Shannon	Theorem 7	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	Lagrange
Havrda-Charvát	Theorem 8	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	Lagrange
Gini	Theorem 10	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	updating formula
Leik	Example 6	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	$H(\mathbf{N}_Y(\mathbf{p})) = 1$
Rényi	Theorem 11	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	transformation
Sharma-Mittal	Theorem 12	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	transformation
Uffink	Remark 2	$H(\mathbf{N}_Y(\mathbf{p})) \geq H(\mathbf{p})$	transformation

4.2. Yager Negation Minimizes ϕ -Entropies

In Remark 1 we quote a result proven by [2]. Any affine-linear negator can be represented as a convex combination of the uniform and the Yager negator. This means that uniform and Yager negators are the vertices for each independent negator. The uniform negator is known to maximize any ϕ entropy. Therefore, it is not surprising that the convex combination can be used to show that the Yager negator $N_Y(p) = (1 - p)/(n - 1)$ minimizes any ϕ -entropy with strictly concave generating function ϕ in the class of all affine-linear negators.

Theorem 3. Let $\phi : [0, 1] \rightarrow (0, \infty)$ be strictly concave, $\mathbf{N}(\mathbf{p}) = (N(p_1), \dots, N(p_n))$ an independent (= affine-linear) negation with negator N and

$$\mathbf{N}_Y(\mathbf{p}) = \left(\frac{1 - p_1}{n - 1}, \dots, \frac{1 - p_n}{n - 1} \right) \tag{27}$$

the Yager negation for $\mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n$. Then, the following applies:

$$H_\phi(\mathbf{N}(\mathbf{p})) \geq H_\phi(\mathbf{N}_Y(\mathbf{p})), \quad \mathbf{p} \in \mathbf{I}_n. \tag{28}$$

Proof. From the strict concavity of ϕ follows for $0 \leq p \leq 1$:

$$\phi(N(p)) = \phi\left(\alpha \frac{1}{n} + (1 - \alpha) \frac{1 - p}{n - 1}\right) \geq \alpha \phi\left(\frac{1}{n}\right) + (1 - \alpha) \phi\left(\frac{1 - p}{n - 1}\right).$$

This implies that the ϕ -entropy of \mathbf{N} is greater than the convex combination of the ϕ -entropies of the uniform and the Yager negator:

$$H_\phi(\mathbf{N}(\mathbf{p})) \geq \alpha H_\phi\left(\frac{1}{n}, \dots, \frac{1}{n}\right) + (1 - \alpha) H_\phi(\mathbf{N}_Y(\mathbf{p})).$$

The maximum property of the uniform negator leads to

$$H_\phi\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \geq H_\phi(\mathbf{N}_Y(\mathbf{p}))$$

such that

$$H_\phi(\mathbf{N}(\mathbf{p})) \geq H_\phi(\mathbf{N}_Y(\mathbf{p})), \quad \mathbf{p} \in \mathbf{I}_n$$

follows. \square

From Theorem 3, we can conclude that the property of monotonicity (25) must only be investigated for the Yager negation. Monotonicity for the Yager negation implies monotonicity for any independent (=affine-linear) negation:

Corollary 1. *Let $\phi : [0, 1] \rightarrow (0, \infty)$ be strictly concave, \mathbf{N} an independent (=affine-linear) negation. Then, it is*

$$H_\phi(\mathbf{N}_Y(\mathbf{p})) \geq H_\phi(\mathbf{p}) \implies H_\phi(\mathbf{N}(\mathbf{p})) \geq H_\phi(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n. \tag{29}$$

Proof. Let $H_\phi(\mathbf{N}_Y(\mathbf{p})) \geq H_\phi(\mathbf{p})$, $\mathbf{p} \in \mathbf{I}_n$. From (28) follows

$$H_\phi(\mathbf{N}(\mathbf{p})) \geq H_\phi(\mathbf{N}_Y(\mathbf{p})) \geq H_\phi(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n.$$

\square

4.3. ϕ -Entropy and Yager Negation

In the literature, in addition to Gini entropy, Havrda–Charvát (or Tsallis) entropy was considered for Yager negation [4–6]. Srivastava and Maheshwari [8] discussed a modified version of the Shannon entropy and Gao and Deng [4] Shannon entropy for Yager negation. We want to investigate whether the property of monotonicity (25) applies to all ϕ -entropies. Without further assumptions on the entropy-generating function ϕ , we can only prove a weaker version of (25), as Theorem 4 shows.

Theorem 4. *Let $\phi : [0, 1] \rightarrow (0, \infty)$ be strictly concave. Then, it holds*

$$H_\phi(\mathbf{N}_Y(\mathbf{N}_Y(\mathbf{p}))) \geq H_\phi(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n.$$

Proof. The Yager negation is affine-linear with $B = -1/(n - 1)$. From (20), we get for $0 \leq p \leq 1$

$$\begin{aligned} N_Y^{(2)}(p) &= \frac{1}{n}(1 - B) + BN_Y(p) \\ &= \frac{1}{n}(1 - B) + B\left(\frac{1}{n}(1 - B) + Bp\right) \\ &= \frac{1}{n}(1 - B) + \frac{1}{n}(B - B^2) + B^2p \\ &= (1 - B^2)\frac{1}{n} + B^2p. \end{aligned}$$

Due to strict concavity of ϕ and $B^2 \leq 1$ (see Remark 1) it is

$$\phi\left((1 - B^2)\frac{1}{n} + B^2p\right) \geq (1 - B^2)\phi\left(\frac{1}{n}\right) + B^2\phi(p), \quad 0 \leq p \leq 1.$$

Inserting into the entropy formula (21) gives

$$\begin{aligned} H_\phi(\mathbf{N}_Y^{(2)}(\mathbf{p})) &= \sum_{i=1}^n \phi(N_Y^{(2)}(p_i)) \\ &= \sum_{i=1}^n \phi\left((1 - B^2)\frac{1}{n} + B^2 p_i\right) \\ &\geq (1 - B^2) \sum_{i=1}^n \phi\left(\frac{1}{n}\right) + B^2 \sum_{i=1}^n \phi(p_i). \end{aligned}$$

Again, we have $\sum_{i=1}^n \phi\left(\frac{1}{n}\right) \geq \sum_{i=1}^n \phi(p_i)$ such that

$$H_\phi(\mathbf{N}_Y^{(2)}(\mathbf{p})) \geq H_\phi(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n.$$

□

The difference to (25) is that ϕ -entropy increases separately for sequences of odd numbers and sequences of even numbers of repeated uses of negation. This means that $H_\phi(\mathbf{N}^{(1)}(\mathbf{p})) \leq H_\phi(\mathbf{N}^{(3)}(\mathbf{p})) \leq H_\phi(\mathbf{N}^{(5)}(\mathbf{p}))$ and $H_\phi(\mathbf{N}^{(2)}(\mathbf{p})) \leq H_\phi(\mathbf{N}^{(4)}(\mathbf{p})) \leq H_\phi(\mathbf{N}^{(6)}(\mathbf{p}))$ and so on. We have no general proof that, for example, $H_\phi(\mathbf{N}^{(1)}(\mathbf{p})) \leq H_\phi(\mathbf{N}^{(2)}(\mathbf{p}))$ holds. This result is not surprising considering that only an odd number of applications of a negation lead back to a negation. Both the original probability vector (= identical transformation) and all even numbers of repeated applications of a negation are non-negative transformations.

4.4. Lagrangian Approach

To see whether the property of monotonicity (25) does not hold only for sequences of odd and even k separately, we choose a Lagrangian approach with the aim to show that the uniform distribution is a stationary point where the information loss

$$H_\phi(\mathbf{N}_Y(\mathbf{p})) - H_\phi(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n \tag{30}$$

is minimal. The minimum value is 0, because $\mathbf{N}_Y(1/n, \dots, 1/n) = (1/n, \dots, 1/n)$. Then (30) must be non-negative and the property of monotonicity (25) is satisfied in the neighbourhood of the uniform distribution.

Theorem 5. Let $n > 1$, ϕ be twice differentiable and $\phi''(p) < 0$. Then

$$L(\mathbf{p}; \lambda) = H_\phi(\mathbf{N}_Y(\mathbf{p})) - H_\phi(\mathbf{p}) + \lambda \left(\sum_{i=1}^n p_i - 1 \right). \tag{31}$$

has a local minimum for $\mathbf{p} = (1/n, \dots, 1/n)$.

Proof. The necessary condition of optimality is

$$\frac{\partial L(\mathbf{p}, \lambda)}{\partial p_i} = -\frac{1}{n-1} \phi' \left(\frac{1-p_i}{n-1} \right) - \phi'(p_i) + \lambda = 0, \quad i = 1, 2, \dots, n.$$

Set

$$g(p) := -\frac{1}{n-1} \phi' \left(\frac{1-p}{n-1} \right) - \phi'(p), \quad 0 \leq p \leq 1, \tag{32}$$

and then the necessary condition of optimality means that

$$g(p_i) = g(p_j), \quad i, j = 1, 2, \dots, n. \tag{33}$$

One solution of (33), and therefore stationary point of (31), is $\mathbf{p} = (p_1, \dots, p_n)$ with $p_i = p_j, i, j = 1, 2, \dots, n$. With $\sum_{i=1}^n p_i = 1$ it is $p_i = 1/n, i = 1, 2, \dots, n$. To show that the uniform distribution is the point where (31) has a local minimum, we need to investigate the second derivatives of (31):

$$\frac{\partial^2 L(\mathbf{p}, \lambda)}{\partial p_i^2} = g'(p_i) = \left(\frac{1}{n-1}\right)^2 \phi''\left(\frac{1-p_i}{n-1}\right) - \phi''(p_i), \quad i = 1, 2, \dots, n$$

and

$$\frac{\partial^2 L(\mathbf{p}, \lambda)}{\partial p_i \partial p_j} = 0, \quad i \neq j, i, j = 1, 2, \dots, n.$$

For $p_i = 1/n$ we get

$$x_i := \left. \frac{\partial^2 L(\mathbf{p}, \lambda)}{\partial p_i^2} \right|_{p_i=1/n} = \left(\left(\frac{1}{n-1}\right)^2 - 1 \right) \phi''\left(\frac{1}{n}\right) > 0.$$

Again, consider the bordered Hessian matrix in the appendix (see Lemma A1). Let Λ_m denote the determinant of the upper left $m \times m$ -matrix of the bordered Hessian matrix for $m = 3, 4, \dots, n + 1$. For a local minimum we have to show that $-1\Lambda_m > 0$ for $m = 3, 4, \dots, n + 1$ [24] (p. 203). From

$$-1\Lambda_m = -\left(-\sum_{I_{m-2}} \prod_{j=1}^{m-2} x_{i_j} \right) > 0$$

with

$$I_{m-2} = \{(i_1, \dots, i_{m-2}) \in \{1, 2, \dots, m\}^{m-2} | i_1 < i_2 < \dots < i_{m-2}\}$$

for $m = 3, 4, \dots, n + 1$, we can conclude that (31) has a local minimum for $\mathbf{p} = (1/n, \dots, 1/n)$. \square

The question is whether there are more points where (31) has a local minimum. If not, the uniform distribution characterizes a global minimum of (31). Because the minimum is 0, the difference (30) is non-negative for all $\mathbf{p} \in \mathbf{I}_n$ and the ϕ -entropy has the property of monotonicity (25). Two criteria are taken into account. If one of these criteria is satisfied, there is no further local minimum.

1. The first criterion is that the function (32) is strictly monotone on $[0, 1]$. In this case, one can conclude from $g(p_i) = g(p_j)$ that $p_i = p_j = 1/n, i = 1, 2, \dots, n$. In Section 5.1, it will be shown that this criterion can be applied to paired Shannon entropy.
2. The second criterion allows (32) to be non-monotone. In principle, there could be other candidates (p_1^*, \dots, p_n^*) satisfying (33). If we can show that these candidates violate the restriction that the probabilities add to 1, we are left with the uniform distribution as the only point where (31) has a local minimum. This criterion can be applied for Shannon and Havrda–Charvát entropy (see Sections 5.2 and 5.3).

5. Information Loss for Special Entropies

5.1. Paired Shannon Entropy and Yager Negation

The paired Shannon entropy

$$H_{pS}(\mathbf{p}) = -\sum_{i=1}^n p_i \ln p_i - \sum_{i=1}^n (1-p_i) \ln(1-p_i) \quad \mathbf{p} = (p_1, \dots, p_n) \tag{34}$$

is given by a generating function ϕ being twice differentiable such that

$$\begin{aligned} \phi(u) &= -u \ln u - (1-u) \ln(1-u), \quad \phi'(u) = -\ln u + \ln(1-u), \\ \phi''(u) &= -1/(u(1-u)) < 0, \quad 0 < u < 1. \end{aligned}$$

As announced, we investigate the monotonicity of the function (32). The proof is given in Appendix B.

Lemma 1. For $\phi(p) = -p \ln p, 0 < p < 1$, the function (32) is given by

$$g(p) = -\frac{n-2}{n-1} \ln(1-p) - \frac{1}{n-1} \ln(n-2+p) + \ln p, \quad 0 < p < 1$$

and strictly increasing on $(0, 1)$.

Then it follows that for the paired Shannon entropy and the Yager negation, the condition of monotonicity holds.

Theorem 6. Let H_{pS} be the paired Shannon entropy and N_Y be the Yager negation. Then, it holds that

$$H_{pS}(N_Y(\mathbf{p})) - H_{pS}(\mathbf{p}) \geq 0, \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{35}$$

Proof. From Theorem 5, we know that the uniform distribution is the point where the difference $H_{pS}(N_Y(\mathbf{p})) - H_{pS}(\mathbf{p})$ has a local minimum with value 0 under the restriction $\sum_{i=1}^n p_i = 1$. By Lemma 1, g is strictly increasing, so $g(p_i) = g(p_j)$ can only hold for $p_i = p_j = 1/n, i, j = 1, 2, \dots, n$. This means that there are no other local minima and the difference has a global minimum for the uniform distribution. \square

Srivastava and Maheshwari [8] discussed a modified version of paired Shannon entropy. They considered the generating function

$$\phi(u) = -u \ln u - \left(\frac{1}{n-1} - u\right) \ln\left(\frac{1}{n-1} - u\right), \quad 0 < u < 1.$$

Let N_Y be Yager negator, and then this modification is motivated by the fact that $N_Y(p) \leq 1/(n-1), 0 \leq p \leq 1$. The maximal upper bound 1 can only be assumed for $n = 2$. However, there are at least two drawbacks to this choice for ϕ . The first is that this entropy depends on the length n of the probability vector. A variation of n means to define a new entropy. The second concerns the property of monotonicity (25). The entropy of the Yager negation has to be compared with the entropy of the original probability vector $\mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n$:

$$-\sum_{i=1}^n p_i \ln p_i - \sum_{i=1}^n \left(\frac{1}{n-1} - p_i\right) \ln\left(\frac{1}{n-1} - p_i\right).$$

For $p_i > 1/(n-1)$, this entropy difference is not well-defined.

5.2. Shannon Entropy and Yager Negation

The Shannon entropy is given by

$$H_S(\mathbf{p}) = -\sum_{i=1}^n p_i \ln p_i \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{36}$$

The corresponding generating function is twice differentiable with

$$\phi(u) = -u \ln u, \quad \phi'(u) = -\ln u, \quad \phi''(u) = -1/u < 0, \quad 0 < u < 1.$$

Again, we consider the function (32). Unlike Lemma 1, this function is no longer strictly increasing as the following lemma shows. Again, you can find the proof in Appendix B.

Lemma 2. Let $n > 2$ and $\phi(p) = -p \ln p, 0 < p < 1$.

The function (32) is given by

$$g(p) = \frac{1}{n-1} \ln(1-p) + \ln p - \frac{1}{n-1} \ln \frac{1}{n-1} + \frac{n}{n-1}, \quad 0 < p < 1.$$

g is strictly increasing on the interval $(0, (n-1)/n)$ and strictly decreasing on $((n-1)/n, 1)$. Let $p_i, p_j \in (0, 1)$ with $g(p_i) = g(p_j)$, then $p_i = p_j$ for $i, j = 1, 2, \dots, n$.

With this result, we can conclude that Shannon entropy satisfies the property of monotonicity (25).

Theorem 7. Let H_S denote Shannon entropy. Then, it holds that

$$H_S(\mathbf{N}_Y(\mathbf{p})) - H_S(\mathbf{p}) \geq 0, \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{37}$$

Proof. Again, we learn from Theorem 5 that the uniform distribution is the point where the difference $H_P(\mathbf{N}_Y(\mathbf{p})) - H_P(\mathbf{p})$ has a local minimum. From Lemma 2 we conclude that $g(p_i) = g(p_j)$ can only be if $p_i = p_j = 1/n, i, j = 1, 2, \dots, n$. This means that there are no other local minima, and the difference (37) has a global minimum for the uniform distribution. \square

The Lagrangian approach has already been proposed by Gao and Deng [4,5] to show that the condition of monotonicity holds for the Shannon entropy. But their proof does not seem to be quite complete. The sufficient condition of local minimum was not considered. The discussion of the global minimum is also missing. This research gap is now filled by the Theorem 7.

5.3. Havrda–Charvát Entropy and Yager Negation

Havrda–Charvát (or Tsallis) entropy [18,20] is given by

$$H_{HC;q}(\mathbf{p}) = \frac{1}{q-1} \sum p_i (1 - p_i^{q-1}), \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n \tag{38}$$

for $q > 0$ and $q \neq 1$. The generating function is

$$\phi(u) = \frac{1}{q-1} u (1 - u^{q-1}) \tag{39}$$

with

$$\phi'(u) = \frac{1}{q-1} (1 - qu^{q-1}) \quad \text{and} \quad \phi''(u) = -qu^{q-2}$$

for $0 < u < 1$ and $q \neq 1$. $\phi''(u)$ is negative on $(0, 1)$ for $q > 0$. We again consider the function (32). The function (32) is not strictly increasing. One can apply a similar reasoning as in Lemma 2 by distinguishing three different ranges for the parameter q . For the proof, see Appendix B.

Lemma 3. Let $n > 2$ and $q > 0, q \neq 1$. The function (32) is

$$g(p) = \frac{q}{q-1} \left(\frac{1}{(n-1)^q} (1-p)^{q-1} + p^{q-1} \right), \quad 0 < p < 1.$$

and, if $g(p_i) = g(p_j)$, then $p_i = p_j$ for $i, j = 1, 2, \dots, n$.

For Havrda–Charvát entropy and Yager negation the property of monotonicity (25) is satisfied.

Theorem 8. Let $H_{HC;q}$ be Havrda–Charvát entropy with $q \neq 1$. Then, it holds

$$H_{HC;q}(\mathbf{N}_Y(\mathbf{p})) - H_{HC;q}(\mathbf{p}) \geq 0, \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{40}$$

Proof. We can use the same arguments as in the proof of Theorem 7. \square

There again are some results proven by Gao and Deng [4,5] concerning Havrda–Charvát entropy. With Theorem 8, we fill a gap in their reasoning.

5.4. Gini Entropy and Independent Negations

Yager [1] showed that Gini entropy H_G of \mathbf{p} cannot be larger than Gini entropy of $\mathbf{N}_Y(\mathbf{p})$. Therefore, the property of monotonicity has already been proven for Gini entropy. We want to show that Gini entropy plays a special role because it is the only entropy such that the negation’s information loss can be calculated explicitly. To see this, we show that Gini entropy of \mathbf{N}_Y is a convex combination of Gini entropy of the uniform distribution and Gini entropy of \mathbf{p} . This means that in every repetition Gini entropy will be updated by Gini entropy of the uniform distribution. To get the information loss (26) not only for Yager negation, we consider a general independent (= affine-linear) negation \mathbf{N} .

Theorem 9. Let \mathbf{N} be an independent (= affine-linear) negation with negator $N, B = N(1) - N(0)$ and $\mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n$. For Gini entropy H_G of $\mathbf{N}^{(k+1)}(\mathbf{p}) = (N^{(k+1)}(p_1), \dots, N^{(k+1)}(p_n))$ applies

$$H_G(\mathbf{N}^{(k+1)}(\mathbf{p})) = (1 - B^2)H_G\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) + B^2H_G(\mathbf{N}^{(k)}(\mathbf{p})). \tag{41}$$

for $k = 0, 1, 2, \dots$

Proof. From (20) we get

$$N^{(k+1)}(p) = (1 - B)\frac{1}{n} + BN^{(k)}(p), \quad 0 \leq p \leq 1$$

and

$$1 - N^{(k+1)}(p) = (1 - B)\frac{n-1}{n} + B(1 - N^{(k)}(p)), \quad 0 \leq p \leq 1$$

such that for the Gini entropy of $\mathbf{N}^{(k+1)}(\mathbf{p})$ holds

$$\begin{aligned} H_G(\mathbf{N}^{(k+1)}(\mathbf{p})) &= \sum_{i=1}^n N^{(k+1)}(p_i)(1 - N^{(k+1)}(p_i)) \\ &= \frac{n-1}{n}(1 - B)^2 + \frac{1}{n}(1 - B)(n-1)B \\ &\quad + \frac{n-1}{n}(1 - B)B + B^2 \sum_{i=1}^n N^{(k)}(p_i)(1 - N^{(k)}(p_i)) \\ &= \sum_{i=1}^n \frac{1}{n} \left(1 - \frac{1}{n}\right) (1 - 2B + B^2 + B - B^2 + B - B^2) \\ &\quad + B^2 \sum_{i=1}^n \sum_{i=1}^n N^{(k)}(p_i)(1 - N^{(k)}(p_i)) \\ &= (1 - B^2)H_G\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) + B^2H_G(\mathbf{N}^{(k)}(\mathbf{p})). \end{aligned}$$

\square

A simple consequence of this theorem is that the Gini entropy cannot increase from one repetition to the next. The information loss is given by

$$H_G(\mathbf{N}^{(k+1)}(\mathbf{p})) - H_G(\mathbf{N}^{(k)}(\mathbf{p})) = (1 - B^2) \left(H_G\left(\frac{1}{n}, \dots, \frac{1}{n}\right) - H_G(\mathbf{N}^{(k)}(\mathbf{p})) \right) \tag{42}$$

for $\mathbf{p} \in \mathbf{I}_n$. It is really a loss because the uniform distribution maximizes Gini entropy such that the difference of Gini entropies is non-negative. This proves the following theorem without the help of the Lagrangian approach.

Theorem 10. Consider an independent (= affine-linear) negation \mathbf{N} . Then, for Gini entropy H_G of \mathbf{N} applies

$$H_G(\mathbf{N}(\mathbf{p})) \geq H_G(\mathbf{N}(\mathbf{p})). \tag{43}$$

This means that Gini entropy also satisfies the property of monotonicity (25).

5.5. Information Loss If ϕ Is Not Strictly Concave

The discussion so far seems to give the impression that negation leads to an information loss for all ϕ -entropies. To show that this impression is misleading, let us again consider Leik entropy (23) and the entropy (24).

Example 6. The information loss of Yager negation measured by Leik entropy is given by

$$\sum_{i=1}^n \min \left\{ \frac{1-p_i}{n-1}, 1 - \frac{1-p_i}{n-1} \right\} - \sum_{i=1}^n \min \{ p_i, 1 - p_i \}, \quad (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{44}$$

The information loss is

$$1 - \sum_{i=1}^n \min \{ p_i, 1 - p_i \} \geq 0, \quad (p_1, \dots, p_n) \in \mathbf{I}_n.$$

This means that the property of monotonicity (25) is indeed satisfied. However, applying Yager negation twice or more often cannot increase the information loss. This property is not desirable.

Example 7. Remember the entropy (24). By counterexamples we are able show to that the difference

$$- \sum_{i=1}^n \left(\frac{1-p_i}{n-1} \right)^\alpha \ln \frac{1-p_i}{n-1} + \sum_{i=1}^n p_i^\alpha \ln p_i \tag{45}$$

can be negative for suitable choices of α and p_1, \dots, p_n . For example, consider $\alpha = 2.1$, $n = 3$ and the probability vector $(0.001, 0.510, 0.489)$. Then the difference (45) is -0.01035 . To see that this can also happen for larger n , we choose $n = 5$ and $\mathbf{p} = (0.1169, 0.0638, 0.3386, 0.0034, 0.4773)$. The corresponding difference (45) is -0.02348 .

Example 7 shows that negations do not automatically lead to a loss of information for every entropy. It is the rule, but there are exceptions.

6. Strictly Increasing Relationship between Entropies

6.1. (h, ϕ) -Entropies

In the following lemma, we state the fact that the property of monotonicity (25) can be transferred from a entropy H_1 to an entropy H_2 , if H_1 and H_2 are related by an strictly increasing transformation h .

Lemma 4. Let \mathbf{N} be a negation and H_1, H_2 two entropies such that $H_2(\mathbf{p}) = h(H_1(\mathbf{p}))$ for $\mathbf{p} \in \mathbf{I}_n$ with $h : \mathbb{R} \rightarrow \mathbb{R}$ strictly increasing. Then, it holds

$$H_1(\mathbf{N}(\mathbf{p})) \geq H_1(\mathbf{p}) \implies H_2(\mathbf{N}(\mathbf{p})) \geq H_2(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n.$$

Proof. The statement immediately follows from the fact that h is strictly increasing. \square

In addition to ϕ -entropies, there are many other entropies. We have already mentioned the overview given by Ilić et al. [10]. By Lemma 4, the property of monotonicity (25) can be transferred from ϕ -entropies to entropies which are strictly increasing functions of ϕ -entropies. This leads us to (h, ϕ) -entropies.

Salicrú et al. [12] generalized ϕ -entropies to (h, ϕ) -entropies

$$H_{h,\phi}(\mathbf{p}) = h\left(\sum_{i=1}^n \phi(p_i)\right), \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n, \tag{46}$$

where either ϕ is strictly concave and h is strictly increasing or ϕ is strictly convex and h is strictly decreasing. For $h(x) = x, x \geq 0$ we get the class of ϕ -entropies.

In [10,12], it was shown that the famous Rényi entropy and the Sharma–Mittal entropy are (h, ϕ) -entropies with suitable chosen h and ϕ . Both do not belong to the class of ϕ -entropies, but are closely related to Havrda–Charvát entropy by a strictly increasing function h .

6.2. Rényi Entropy, Havrda–Charvát Entropy and Yager Negation

Rényi [25] introduced the entropy

$$H_{R;\alpha}(\mathbf{p}) = \frac{1}{1-\alpha} \ln \sum_{i=1}^n p_i^\alpha, \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n \tag{47}$$

with parameter $\alpha > 0$. For $\alpha \rightarrow 1$ we get Shannon entropy. Rényi entropy (47) does not belong to the class of ϕ -entropies. Nevertheless, Rényi entropy satisfies the property of monotonicity (25) for Yager negation.

Theorem 11. Let $H_{R;\alpha}$ denote Rényi entropy. Then, it holds

$$H_{R;\alpha}(\mathbf{N}_Y(\mathbf{p})) \geq H_{R;\alpha}(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n$$

for $\alpha > 0$.

Proof. There is a functional relationship h between Rényi and Havrda–Charvát entropy given by

$$H_{HC;\alpha}(\mathbf{p}) = h(H_{R;\alpha}(\mathbf{p})), \quad \mathbf{p} \in \mathbf{I}_n$$

with

$$h(x) = \frac{1}{1-\alpha} \left(e^{(1-\alpha)x} - 1 \right), \quad x \geq 0$$

where $\alpha > 0$. According to

$$h'(x) = e^{(1-\alpha)x} > 0, \quad x > 0$$

for $\alpha > 0, h$ is strictly increasing. In Theorem 8 we proved that Havrda–Charvát entropy has the property of monotonicity (25). By using Lemma 4 we see that Rényi entropy also satisfies the property of monotonicity (25). \square

6.3. Sharma–Mittal Entropy, Havrda–Charvát Entropy and Yager Negation

We can go a step further and consider Sharma–Mittal entropy [10,26] with two parameters α and β :

$$H_{SM;\alpha,\beta}(\mathbf{p}) = \frac{1}{1-\beta} \left(\left(\sum_{i=1}^n p_i^\alpha \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \quad \mathbf{p} = (p_1, \dots, p_n) \in \mathbf{I}_n. \tag{48}$$

Special cases are Shannon ($\alpha = \beta = 1$), Rényi ($\beta = 1$) and Havrda–Charvát ($\alpha = \beta$) entropy. Again, we can find a strictly increasing functional relation h between Sharma–Mittal entropy and Havrda–Charvát entropy to prove the following theorem.

Theorem 12. *Let $H_{SM;\alpha,\beta}$ be Sharma–Mittal entropy. Then, it follows that*

$$H_{SM;\alpha,\beta}(\mathbf{N}(\mathbf{p})) \geq H_{SM;\alpha,\beta}(\mathbf{p}), \quad \mathbf{p} \in \mathbf{I}_n$$

for $\alpha > 0$ with $\alpha \neq 1$ and $\beta \neq 1$.

Proof. It is

$$\sum_{i=1}^n p_i^\alpha = (1-\alpha)H_{HC;\alpha}(\mathbf{p}) + 1 > 0, \quad \mathbf{p} \in \mathbf{I}_n. \tag{49}$$

Substituting $\sum_{i=1}^n p_i^\alpha$ in (48) gives

$$H_{SM;\alpha,\beta}(\mathbf{p}) = \frac{1}{1-\beta} \left(((1-\alpha)H_{HC;\alpha}(\mathbf{p}) + 1)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \quad \mathbf{p} \in \mathbf{I}_n.$$

Differentiating Sharma–Mittal entropy with regard to Havrda–Charvát entropy and using (49) leads to

$$\frac{\partial H_{SM;\alpha,\beta}(\mathbf{p})}{\partial H_{HC;\alpha}(\mathbf{p})} = ((1-\alpha)H_{HC;\alpha}(\mathbf{p}) + 1)^{\alpha-\beta} = \left(\sum_{i=1}^n p_i^\alpha \right)^{\alpha-\beta} > 0, \quad \mathbf{p} \in \mathbf{I}_n.$$

Therefore, h is strictly increasing and the property of monotonicity (25) is valid for Sharma–Mittal entropy and Yager negation. \square

Remark 2. *This results can be generalized to entropies of the form*

$$f \left(\left(\sum_{i=1}^n p_i^\alpha \right)^{1/(1-\alpha)} \right), \quad (p_1, \dots, p_n) \in \mathbf{I}_n \tag{50}$$

with f strictly increasing on $[0, \infty)$ and $\alpha > 0$. This class (50) was considered by Uffink [27] and intensively discussed in a recent paper of Jizba and Korbel [28]. Similar to Theorem 11 and Theorem 12, $\sum_{i=1}^n p_i^\alpha$ can be substituted by Havrda–Charvát entropy such that (50) is a strictly increasing function of Havrda–Charvát entropy. This means that (50) also satisfies the property of monotonicity (25).

7. ϕ -Entropy in the Dependent Case

Wu et al. [7] used the rudimentary arguments of Gao and Deng [4,5] to show that Gini entropy of the exponential negation cannot be smaller than Gini entropy of the original probability vector. For dependent negators, the bordered Hessian matrix is still more complex such that it is doubtful whether a proof that (30) is non-negative will be possible.

We consider negators of the form (9) with generating function f decreasing on $[0, 1]$. Some examples are given in the introduction and will be repeated here.

1. Yager negator [1] N_1 with

$$f_1(p) = 1 - p, \quad 0 \leq p \leq 1,$$

2. exponential negator [7] N_2 with

$$f_2(p) = e^{-p}, \quad 0 \leq p \leq 1,$$

3. cosinus negator N_3 with

$$f_3(p) = 0.5(\cos(\pi p)), \quad 0 \leq p \leq 1,$$

4. and square root negator N_4 (special case of the Tsallis negator discussed by Zhang et al. [6]):

$$f_4(p) = 1 - \sqrt{p}, \quad 0 \leq p \leq 1.$$

f_2, f_3 and f_4 are generating functions for dependent negations. For dependent negations, it is not easy to prove properties of their entropies. Wu et al. [7] discussed the exponential negation. They considered the Shannon entropy for this negator. In their numerical examples, they compared Yager und exponential negators for different concrete probability vectors and a different number of categories n . Their general result is that exponential negator converges faster to the uniform distribution than Yager negator.

We choose another approach to get an idea how Gini and Shannon entropies behave for dependent negators and how fast the convergence is. We obtain 1,000,000 different probability vectors p_1, \dots, p_n by drawing random samples from a Dirichlet distribution. For each probability vector, the entropy (Gini or Shannon) is calculated. From the resulting 1,000,000 entropy values we estimate the entropy density. The same procedure is applied for the probability vectors transformed by the negators $N_i, i = 1, 2, 3, 4$. We use the random generator `rdirichlet` from the R-package "MCMCpack" of [29] and density estimation by the standard R routine "density". Walley [30] discussed the Dirichlet distribution as a model for probabilities. The Dirichlet distribution has a vector of hyperparameters $(\alpha_1, \dots, \alpha_n)$ that has to be chosen before starting the simulation. Following [30], the decision falls on a noninformative prior setting with $\alpha = i = 1, i = 1, 2, \dots, n$. Note that other choices affect the graphical representation, but not the general results.

Figures 1 and 2 show the results for the Gini resp. the Shannon entropy. The number of categories is in both cases $n = 5$. For a higher number of categories, the results are even more pronounced. The top left panel presents the generating functions of the four negators. The top right panel shows the density estimation for the entropy of the original (non-transformed) probability vector. The bottom left panel compares the estimated entropy densities for the Yager negation applied once ($k = 1$) and twice ($k = 2$). The bottom right panel presents a comparison of the estimated entropy densities when the four negators are applied exactly once.

We can see that there is not much difference between the results for the Gini and the Shannon entropy. For $n = 5$, the Gini entropy has the maximum value 0.8 and the Shannon entropy the maximum value $\ln 5 = 1.609$. This explains the range of the abscissae.

The main results are:

1. A single application of the negation already results in a very concentrated and strongly peaked entropy distribution. This confirms the fact stated in Remark 1 that the range of negations is very narrow.
2. The double or multiple use of a negation leads to a distribution that resembles a singular distribution concentrated at the entropy's maximum value. Therefore, the convergence rate to the uniform distribution is very high.
3. The convergence rate is even higher when we consider negations with generating function f with $f(p) \geq f_1(p)$ for $0 \leq p \leq 1$. The reverse is also true. Negations with generating function $f(p) \leq f_1(p), 0 \leq p \leq 1$ give lower convergence rates.

- In the interval $[0, 1/2]$, the generating function of the cosine negation is greater and in $[1/2, 1]$ smaller than the generating function of the Yager negation. Nevertheless, the cosine negation produces entropy distributions looking similar to the entropy distribution of the Yager negation. The entropy formula seems to eliminate the difference between the cosine and the Yager negator.

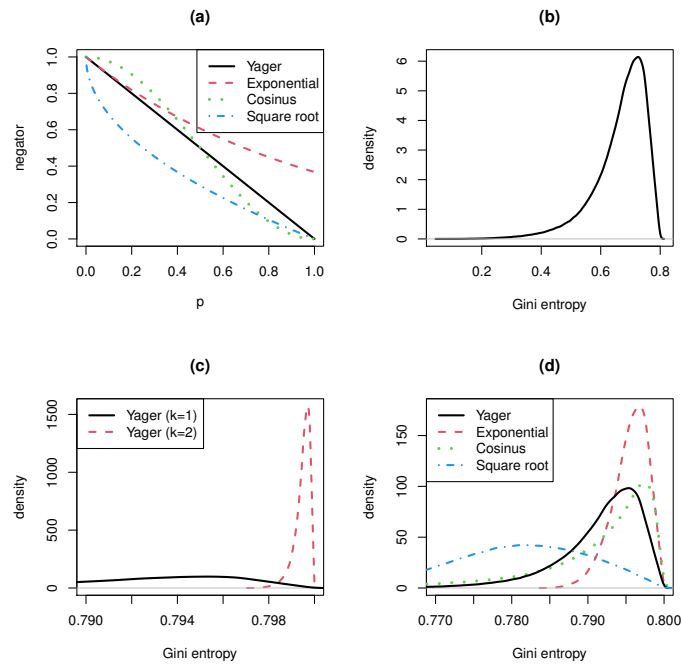


Figure 1. Gini entropy for different negations and $n = 5$. (a) Generating functions of negators. (b) Density of Gini entropy. (c) Density of Gini entropy for Yager negation applied once ($k = 1$) and twice ($k = 2$). (d) Density of Gini entropy for four negators.

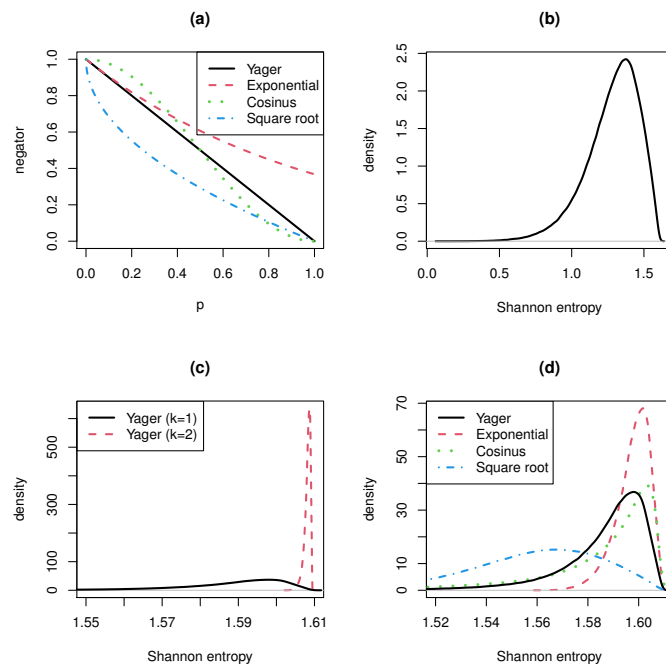


Figure 2. Shannon entropy for different negations and $n = 5$. (a) Generating functions of negators. (b) Density of Shannon entropy. (c) Density of Shannon entropy for Yager negation applied once ($k = 1$) and twice ($k = 2$). (d) Density of Shannon entropy for four negators.

8. Conclusions

First, we can prove that independent negations must be affine-linear. Due to Batyrshin [2] this was an open problem. The property of monotonicity means that the entropy cannot decrease by applying negation to a probability vector. This is equivalent to the fact that negations imply an information loss. We show that the property of monotonicity is satisfied for ϕ -entropies and Yager negation. It is sufficient to consider Yager negation because we can prove that the monotonicity of entropies of Yager negation can be transferred to all affine-linear (=independent) negations. We try to prove the monotonicity of ϕ -entropies by means of strict concavity of the entropy generating function. This procedure is only partially successful. The monotonicity holds for odd and even sequences for the number of negation repetitions k separately. Alternatively, following some examples from the literature, we try to prove monotonicity by a Lagrangian approach based on the difference of the ϕ -entropies for Yager negation and the original probability vector. For general ϕ -entropies, this approach is also only partially successful. We can prove monotonicity in the neighbourhood of the uniform distribution as the point where the difference of the ϕ -entropies has a local minimum. To show monotonicity for all probability vectors, we have to consider concrete ϕ -entropies like the Gini, the Shannon or the Havrda–Charvát (Tsallis) entropy. For the Gini entropy, it is not necessary to use the Lagrangian approach. The Gini entropy of an affine-linear negation can be represented as a convex combination of maximal Gini entropy and Gini entropy of the original probability vector. This updating formula can be applied to arbitrary k -times the negation is repeated. This leads to a sequence of non-decreasing values for Gini entropy converging towards the maximum value generated by the uniform distribution. Such an argument does not seem to apply for Shannon entropy. For this reason, we again take up the Lagrangian approach and show that the uniform distribution is the unique point at which the Lagrange function of Shannon entropy difference has a global minimum. The same can be shown for the difference of Havrda–Charvát (Tsallis) entropies. This means that the property of monotonicity is valid for Gini, Shannon and Havrda–Charvát (Tsallis) entropy. (h, ϕ) -entropies generalize ϕ -entropies. If h is a strictly increasing function and the condition of monotonicity holds for a ϕ -entropy, we show that the property of monotonicity is also satisfied for (h, ϕ) -entropy. With this argument, the property of monotonicity can easily be checked for Rényi and Sharma–Mittal entropy. For dependent negations, it is not easy to get analytical results. Therefore, we simulate the entropy distribution and show how different repeated negations affect Gini and Shannon entropy. The simulation approach has the advantage that the whole simplex of discrete probability vectors can be considered at once and not just arbitrarily selected probability vectors. When Yager negation is used as a point of reference, we see that negations with generating functions larger (smaller) than the generating function of Yager negation produce larger (smaller) information loss and a (lower) higher speed of convergence to the uniform distribution.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: I thank Christian Weiss for introducing me to the concept of negation and two anonymous reviewers for their helpful comments.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

The Lagrangian approach is used to prove that the uniform distribution gives a global maximum for ϕ -entropies and a local minimum for the difference of the ϕ -entropy of the Yager negation and the ϕ -entropy of the original probability vector, in each case under the restriction that the probabilities add up to 1. The corresponding bordered Hessian matrices and their determinants are considered in Lemma A1.

Lemma A1. Let $(x_1, \dots, x_n) \in \mathbb{R}^n$ be a vector with $x_i \neq 0$ for $i = 1, 2, \dots, n$ and

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & x_1 & 0 & 0 & \dots & 0 \\ 1 & 0 & x_2 & 0 & \dots & 0 \\ 1 & 0 & 0 & x_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & x_n \end{bmatrix} \tag{A1}$$

a $(n + 1) \times (n + 1)$ -matrix. Λ_m denotes the determinant of the upper left $m \times m$ -matrix for $m = 3, 4, \dots, n + 1$. Then,

$$\Lambda_m = - \sum_{I_{m-2}} \prod_{j=1}^{m-2} x_{i_j}$$

with

$$I_{m-2} = \{(i_1, \dots, i_{m-2}) \in \{1, 2, \dots, m - 1\}^{m-2} | i_1 < i_2 < \dots < i_{m-2}\}$$

for $m = 3, 4, \dots, n + 1$.

Proof. The statement is true for $m = 3$: Laplace expansion [24] (p. 292) along the third row gives

$$\Lambda_3 = 1(-x_1) + x_2(-1) = -x_1 - x_2 = \sum_{I_1} \prod_{j=1}^1 x_{i_j}$$

with

$$I_1 = \{i_1 \in \{1, 2\}^1\} = \{1, 2\}.$$

Assume that the statement is true for m . Laplace expansion of the upper left $m + 1 \times m + 1$ -matrix along the $m + 1$ -th row gives

$$\Lambda_{m+1} = -1 \prod_{i=1}^{m-1} x_i + x_m \Lambda_m$$

Inserting

$$\Lambda_m = -x_1 x_2 \dots x_{m-2} - x_1 x_2 \dots x_{m-3} x_{m-1} - \dots - x_2 x_3 \dots x_{m-1}$$

leads to

$$\Lambda_{m+1} = -x_1 x_2 \dots x_{m-1} - x_1 x_2 \dots x_{m-3} x_{m-1} x_m - \dots - x_2 x_3 \dots x_{m-1} x_m$$

such that

$$\Lambda_{m+1} = - \sum_{I_{m-1}} \prod_{j=1}^{m-2} x_{i_j}$$

with

$$I_{m-1} = \{(i_1, \dots, i_{m-1}) \in \{1, 2, \dots, m\}^{m-1} | i_1 < i_2 < \dots < i_{m-1}\}.$$

□

Appendix B

Proof of Lemma 1. It is

$$\begin{aligned} g(p) &= -\frac{1}{n-1}\phi'\left(\frac{1-p}{n-1}\right) - \phi'(p) \\ &= \frac{1}{n-1}\left(-\ln\left(\frac{1-p}{n-1}\right)\right) + \ln\left(\frac{n-2+p}{n-1}\right) + \ln p - \ln(1-p) \\ &= -\frac{n-2}{n-1}\ln(1-p) - \frac{1}{n-1}\ln(n-2+p) + \ln p \end{aligned}$$

for $0 < p < 1$. After some simple calculations, the derivative of g is

$$g'(p) = \frac{n-2}{p(1-p)(n-2+p)} > 0, \quad 0 < p < 1$$

for $n > 2$. Therefore, g is strictly increasing on $(0, 1)$. \square

Proof of Lemma 2. Consider the derivative

$$g'(p) = \frac{1}{n-1} \frac{1}{1-p} + \frac{1}{p} = n \frac{(n-1)/n - p}{(n-1)p(1-p)}$$

for $0 < p < 1$. Then it is $g'(p) > 0$ for $p < (n-1)/n$ and $g'(p) < 0$ for $p > (n-1)/n$. Let $p > (n-1)/n$. It is $1-p < p$ and $\ln(1-p) < \ln p$ such that

$$g(p) - g(1-p) = \frac{n-2}{n-1}(\ln p - \ln(1-p)) > 0$$

for $n > 2$. Let $p_i \neq p_j$ and $p_i > (n-1)/n$ with $g(p_i) = g(p_j)$ then

$$g(p_j) = g(p_i) > g(1-p_i).$$

g is strictly increasing on $[0, (n-1)/n]$. $g(p_j) > g(1-p_i)$ means $p_j > 1-p_i$ and $p_i + p_j > 1$. This contradicts the fact that p_i and p_j are probabilities. Therefore, it must be $p_i = p_j$, if $g(p_i) = g(p_j)$. \square

Proof of Lemma 3. $g(p)$ follows immediately by inserting ϕ' . The derivative is

$$g'(p) = q\left(-\frac{1}{(n-1)^q}(1-p)^{q-2} + p^{q-2}\right), \quad 0 < p < 1, \quad q \neq 2$$

with $g'(p^*) = 0$ for

$$p^* = \frac{1}{1 + (n-1)^{q/(q-2)}}, \quad 0 < p < 1, \quad q \neq 2.$$

For $q = 2$, we get the Gini entropy already discussed by Yager [1].

- Case 1: $0 < q < 2$. We have

$$g'(p) = \begin{cases} > 0 & \text{for } p < p^* \\ = 0 & \text{for } p = p^* \\ < 0 & \text{for } p > p^* \end{cases}$$

It is

$$p^* = \frac{1}{1 + (1/(n-1))^{q/(2-q)}} = \frac{(n-1)^{q/(2-q)}}{1 + (n-1)^{q/(2-q)}} > 1/2 \quad \text{for } n > 2.$$

- Subcase 1.1: $1 < q < 2$. For $p > 1 - p$ it holds

$$g(p) - g(1 - p) = \left(1 - \left(\frac{1}{n-1}\right)^q\right) \left(p^{q-1} - (1-p)^{q-1}\right) > 0.$$

Consider $p_i > p^* > 1/2$ such that $p_i > 1 - p_i$ and $p_j \neq p_i$ with $g(p_i) = g(p_j)$. We have $g(p_j) = g(p_i) > g(1 - p_i)$ such that $p_j > 1 - p_i$ and $p_i + p_j > 1$.

- Subcase 1.2: $0 < q < 1$: First, reformulate $g(p)$ as

$$\begin{aligned} g(p) &= -\frac{q}{1-q} \left(\frac{1}{(n-1)^q}\right) \left(\left(\frac{1}{1-p}\right)^{1-q} + \left(\frac{1}{p}\right)^{1-q}\right) \\ &= \frac{q/(1-q)}{p^{1-q}(1-p)^{1-q}} \left(-\frac{1}{(n-1)^q}\right) \left(p^{1-q} - (1-p)^{1-q}\right). \end{aligned}$$

This gives for $g(p) - g(1 - p)$ and $p > 1 - p$

$$g(p) - g(1 - p) = \frac{q/(1-q)}{p^{1-q}(1-p)^{1-q}} \left(1 - \frac{1}{(n-1)^q}\right) \left(p^{1-q} - (1-p)^{1-q}\right) > 0.$$

The rest follows the arguments from subcase 1.1.

- Case 2: $q > 2$ means that $q/(q-2) > 0$ holds. Then, $p^* < 1/2$ since $(n-1)^{q/(q-2)} > 1$ for $n > 2$. Now, we have

$$\frac{\partial g(p)}{\partial p} \begin{cases} < 0 & \text{for } p < p^* \\ = 0 & \text{for } p = p^* \\ > 0 & \text{for } p > p^* \end{cases}$$

From this and $p < 1 - p$ it follows

$$g(p) - g(1 - p) = \frac{q}{q-1} \left(1 - \frac{1}{(n-1)^q}\right) \left(p^{1-q} - (1-p)^{1-q}\right) < 0.$$

Consider $p_i < p^* < 1/2$ such that $p_i < 1 - p_i$ and $p_j \neq p_i$ with $g(p_i) = g(p_j)$. Because g is strictly decreasing in $(0, p^*)$ we have $g(p_j) = g(p_i) < g(1 - p_i)$ such that $p_j > 1 - p_i$ and $p_i + p_j > 1$. This contradicts the fact that p_i and p_j are probabilities.

□

References

1. Yager, R. On the maximum entropy negation of a probability distribution. *IEEE Trans. Fuzzy Syst.* **2014**, *23*, 1899–1902. [\[CrossRef\]](#)
2. Batyrshin, I.; Villa-Vargas, L.A.; Ramirez-Salinas, M.A.; Salinas-Rosales, M.; Kubysheva, N. Generating negations of probability distributions. *Soft Comput.* **2021**, *25*, 7929–7935. [\[CrossRef\]](#)
3. Batyrshin, I. Contracting and involutive negations of probability distributions. *Mathematics* **2021**, *9*, 2389. [\[CrossRef\]](#)
4. Gao, X.; Deng, Y. The generalization negation of probability distribution and its application in target recognition based on sensor fusion. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1–8. [\[CrossRef\]](#)
5. Gao, X.; Deng, Y. The negation of basic probability assignment. *IEEE Access* **2019**, *7*, 107006–107014. [\[CrossRef\]](#)
6. Zhang, J.; Liu, R.; Zhang, J.; Kang, B. Extension of Yager’s negation of a probability distribution based on Tsallis entropy. *Int. J. Intell. Syst.* **2020**, *35*, 72–84. [\[CrossRef\]](#)
7. Wu, Q.; Deng, Y.; Xiong, N. Exponential negation of a probability distribution. *Soft Comput.* **2022**, *26*, 2147–2156. [\[CrossRef\]](#)
8. Srivastava, A.; Maheshwari, S. Some new properties of negation of a probability distribution. *Int. J. Intell. Syst.* **2018**, *33*, 1133–1145. [\[CrossRef\]](#)
9. Aczél, J. *Vorlesungen über Funktionalgleichungen und ihre Anwendungen*; Birkhauser: Basel, Switzerland, 1961.
10. Ilić, V.; Korbel, J.; Gupta, S.; Scarfone, A. An overview of generalized entropic forms. *Europhys. Lett.* **2021**, *133*, 50005. [\[CrossRef\]](#)
11. Burbea, J.; Rao, C. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 489–495. [\[CrossRef\]](#)
12. Salicrú, M.; Menéndez, M.; Morales, D.; Pardo, L. Asymptotic distribution of (h, Φ) -entropies. *Commun. Stat. Theory Methods* **1993**, *22*, 2015–2031. [\[CrossRef\]](#)

13. Gini, C. *Variabilità e Mutabilità: Contributo alla Distribuzioni e delle Relazioni Statistiche*; Tipografia di Paolo Cuppin: Bologna, Italy, 1912.
14. Onicescu, O. Théorie de l'information énergie informationelle. *Comptes Rendus l'Academie Sci. Ser. AB* **1966**, *263*, 841–842.
15. Vajda, I. Bounds on the minimal error probability and checking a finite or countable number of hypotheses. *Inf. Transm. Probl.* **1968**, *4*, 9–17.
16. Rao, C. Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.* **1982**, *21*, 24–43. [[CrossRef](#)]
17. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
18. Havrda, J.; Charvát, F. Quantification method of classification processes. Concept of structural α -entropy. *Kybernetika* **1967**, *3*, 30–35.
19. Daróczy, Z. Generalized information functions. *Inf. Control.* **1970**, *16*, 36–51. [[CrossRef](#)]
20. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
21. Leik, R. A measure of ordinal consensus. *Pac. Sociol. Rev.* **1966**, *9*, 85–90. [[CrossRef](#)]
22. Klein, I.; Mangold, B.; Doll, M. Cumulative paired ϕ -entropy. *Entropy* **2016**, *18*, 248. [[CrossRef](#)]
23. Shafee, F. Lambert function and a new non-extensive form of entropy. *IMA J. Appl. Math.* **2007**, *72*, 785–800. [[CrossRef](#)]
24. Mosler, K.; Dyckerhoff, R.; Scheicher, C. *Mathematische Methoden für Ökonomen*; Springer: Berlin, Germany, 2009.
25. Rényi, A. On measures of entropy and information. In *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
26. Sharma, B.; Mittal, D. New nonadditive measures of entropy for discrete probability distributions. *J. Math. Sci.* **1975**, *10*, 28–40.
27. Uffink, J. Can the maximum entropy principle be explained as a consistency requirement? *Stud. Hist. Philos. Sci. Part B Stud. Hist. Philos. Mod. Phys.* **1995**, *26*, 223–261. [[CrossRef](#)]
28. Jizba, P.; Korbel, J. When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics. *Phys. Rev. E* **2020**, *101*, 042126. [[CrossRef](#)] [[PubMed](#)]
29. Martin, A.; Quinn, K.; Park, J. MCMCpack: Markov Chain Monte Carlo in R. *J. Stat. Softw.* **2011**, *42*, 22. [[CrossRef](#)]
30. Walley, P. Inferences From multinomial data: Learning about a bag of marbles (with discussion). *J. R. Stat. Soc. Ser. B* **1996**, *58*, 3–57.