

Article

Some Theoretical Foundations of Bare-Simulation Optimization of Some Directed Distances between Fuzzy Sets Respectively Basic Belief Assignments

Michel Broniatowski ¹  and Wolfgang Stummer ^{2,*} 

¹ Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, 4 Place Jussieu, 75252 Paris, France; michel.broniatowski@sorbonne-universite.fr

² Department of Mathematics, Friedrich-Alexander-Universität Erlangen–Nürnberg, Cauerstrasse 11, 91058 Erlangen, Germany

* Correspondence: stummer@math.fau.de

Abstract: It is well known that in information theory—as well as in the adjacent fields of statistics, machine learning and artificial intelligence—it is essential to quantify the dissimilarity between objects of uncertain/imprecise/inexact/vague information; correspondingly, constrained optimization is of great importance, too. In view of this, we define the dissimilarity-measure-natured *generalized φ -divergences* between fuzzy sets, ν -rung orthopair fuzzy sets, extended representation type ν -rung orthopair fuzzy sets as well as between those fuzzy set types and vectors. For those, we present how to tackle corresponding constrained minimization problems by appropriately applying our recently developed dimension-free *bare (pure) simulation method*. An analogous program is carried out by defining and optimizing *generalized φ -divergences* between (rescaled) basic belief assignments as well as between (rescaled) basic belief assignments and vectors.

Keywords: generalized φ -divergences; fuzzy sets; basic belief assignments



Citation: Broniatowski, M.; Stummer, W. Some Theoretical Foundations of Bare-Simulation Optimization of Some Directed Distances between Fuzzy Sets Respectively Basic Belief Assignments. *Entropy* **2024**, *26*, 312. <https://doi.org/10.3390/e26040312>

Academic Editors: Abhik Ghosh and Elena Castilla

Received: 17 January 2024

Revised: 14 March 2024

Accepted: 22 March 2024

Published: 1 April 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Directed distances—particularly known as divergences— $D(\mathbf{Q}, \mathbf{P})$ between probability vectors (i.e., vectors of probability frequencies) $\mathbf{Q}, \mathbf{P}$ are widely used in statistics as well as in the adjacent research fields of information theory, artificial intelligence and machine learning. Prominent examples are, e.g., the Kullback–Leibler information distance/divergence (also known as relative entropy), the Hellinger distance, the Jensen–Shannon divergence and Pearson’s chi-square distance/divergence; those are special cases of the often-used wider class of the so-called Csiszár–Ali–Silvey–Morimoto (CASM) φ -divergences $D_\varphi(\mathbf{Q}, \mathbf{P})$ (cf. [1–3]). For some comprehensive overviews on CASM φ -divergences, the reader is referred to, e.g., the insightful books [4–9], the survey articles [10–13], and the references therein. It is well known that the optimization of such CASM φ -divergences plays an important role, in obtaining estimators (e.g., the omnipresent maximum likelihood estimation method can be equivalently seen as a minimum Kullback–Leibler information distance estimation method), as well as in quantifying the model adequacy in the course of a model-search (model-selection) procedure; for the latter, see, e.g., [14–17].

In the literature, one can also find a substantial amount of special cases of CASM φ -divergences $D_\varphi(\mathbf{Q}, \mathbf{P})$ between other prominent statistical quantities $\mathbf{Q}, \mathbf{P}$ (other than probability frequencies), see, e.g., [18] for a corresponding recent survey. In contrast, there also exist special cases of CASM φ -divergences $D_\varphi(B, A)$ between other basic objects B, A for the quantification of uncertain/imprecise/inexact/vague information such as *fuzzy sets* (cf. [19]) and *basic belief assignments* from Dempster–Shafer evidence theory (cf. [20,21]). Indeed, as far as the former is concerned, for instance, ref. [22] employ (a variant of) the Kullback–Leibler information distance between two fuzzy sets B and A (which they call

fuzzy expected information for discrimination in favor of B against A), ref. [23] investigate the Jensen–Shannon divergence between two intuitionistic fuzzy sets B and A (which they call *symmetric information measure between B and A*), whereas [24] deal with the Jensen–Shannon divergence between two extended representation type (i.e., hesitancy-degree including) Pythagorean fuzzy sets B and A . As far as CASM φ -divergences $D_\varphi(B, A)$ between basic belief assignments (BBAs) B, A is concerned, for instance, refs. [25,26] employ the Jensen–Shannon divergence for multi-sensor data fusion, whereas [27] use the Hellinger distance for characterizing the degree of conflict between BBAs.

In view of the above-mentioned illuminations, the main goals of this paper are as follows:

- (M1) to define—dissimilarity-quantifying— *generalized CASM φ -divergences* between fuzzy sets, between ν -rung orthopair fuzzy sets in the sense of [28] (including intuitionistic and Pythagorean fuzzy sets), between extended representation type ν -rung orthopair fuzzy sets, between those fuzzy set types and vectors, between (rescaled) basic belief assignments as well as between (rescaled) basic belief assignments and vectors;
- (M2) to present how one can tackle corresponding constrained minimization problems by appropriately applying our recently developed dimension-free *bare (pure) simulation method* of [29].

This agenda is achieved in the following way: in the next Section 2, we recall the basic definitions and properties of *generalized CASM φ -divergences* between vectors. The follow-up Section 3 explains the basic principles of our above-mentioned bare-simulation optimization method of [29] (where, for the sake of brevity, we focus on the minimal values and not on the corresponding minimizers). In Section 4, we achieve the main goals (M1) and (M2) for the above-mentioned types of fuzzy sets, whereas Section 5 concerns with (M1) and (M2) for (rescaled) basic belief assignments. The conclusions are discussed in the final Section 6.

2. Divergences between Vectors

As usual, a divergence is a function $D : \mathbb{R}^K \times \mathbb{R}^K \mapsto [0, \infty]$ with the following properties: $D(\mathbf{Q}, \mathbf{P}) \geq 0$ for K -dimensional vectors $\mathbf{Q}, \mathbf{P} \in \mathbb{R}^K$, and $D(\mathbf{Q}, \mathbf{P}) = 0$ iff $\mathbf{Q} = \mathbf{P}$. Since, in general, $D(\mathbf{Q}, \mathbf{P}) \neq D(\mathbf{P}, \mathbf{Q})$ and the triangle inequality is not satisfied, $D(\mathbf{Q}, \mathbf{P})$ can be interpreted as *directed distance* from $\mathbf{Q}$ to $\mathbf{P}$; accordingly, the divergences D can be connected to dissimilarity quantification and to geometric issues in various different ways, see, e.g., the detailed discussion in Section 1.5 of [18]. Typically, a divergence D is generated by some function φ . For the latter, here we fundamentally require the following:

- (G1) $\varphi :] - \infty, \infty[\rightarrow [0, \infty]$ is lower semicontinuous and convex;
- (G2) $\varphi(1) = 0$;
- (G3) the effective domain $\text{dom}(\varphi) := \{t \in \mathbb{R} : \varphi(t) < \infty\}$ has interior $\text{int}(\text{dom}(\varphi))$ of the form $\text{int}(\text{dom}(\varphi)) =]a, b[$ for some $-\infty \leq a < 1 < b \leq \infty$ (notice that (G3) follows from (G1), (G2) and the requirement that $\text{int}(\text{dom}(\varphi))$ is non-empty);
- (G4') φ is strictly convex in a neighborhood $]t_-^{sc}, t_+^{sc}[\subseteq]a, b[$ of one $(t_-^{sc} < 1 < t_+^{sc})$.

Furthermore, we set $\varphi(a) := \lim_{t \downarrow a} \varphi(t)$ and $\varphi(b) := \lim_{t \uparrow b} \varphi(t)$. The class of all functions φ with (G1), (G2), (G3) and (G4') will be denoted by $\tilde{Y}(]a, b[)$. For $\varphi \in \tilde{Y}(]a, b[)$, $\mathbf{P} := (p_1, \dots, p_K) \in \mathbb{R}_{>0}^K := \{\mathbf{R} := (r_1, \dots, r_K) \in \mathbb{R}^K : r_i > 0 \text{ for all } i = 1, \dots, K\}$ and $\mathbf{Q} := (q_1, \dots, q_K) \in \Omega \subset \mathbb{R}^K$ we define as directed distance the *generalized Csiszár–Ali–Silvey–Morimoto φ -divergence* (cf. [1–3,30,31]) —in short, the *generalized φ -divergence*—

$$D_\varphi(\mathbf{Q}, \mathbf{P}) := \sum_{k=1}^K p_k \cdot \varphi\left(\frac{q_k}{p_k}\right) \geq 0, \quad (1)$$

where for finiteness reasons one often even has $\mathbf{Q} \in \Omega \subset \mathbb{R}_{\geq 0}^K := \{\mathbf{R} := (r_1, \dots, r_K) \in \mathbb{R}^K : r_i \geq 0 \text{ for all } i = 1, \dots, K\}$; for a comprehensive technical treatment, see, e.g., [32] (for instance, if $]a, b[=]0, \infty[$ then one can include zeros in (1) by the “conventions” $p_k \cdot \varphi\left(\frac{0}{p_k}\right) =$

$p_k \cdot \lim_{t \downarrow 0} \varphi(t)$ for $p_k > 0$, $0 \cdot \varphi\left(\frac{q_k}{0}\right) = q_k \cdot \frac{0}{q_k} \cdot \varphi\left(\frac{q_k}{0}\right) = q_k \cdot \lim_{t \downarrow 0} t \cdot \varphi\left(\frac{1}{t}\right)$ for $q_k > 0$, and $0 \cdot \varphi\left(\frac{0}{0}\right) = 0$). Comprehensive overviews on these important (generalized) φ -divergences are given in e.g., [4–18], and the references therein.

3. Optimization of Generalized φ -Divergences via the Bare Simulation Solution Method

In the following we deal with

Problem 1. For pregiven $\varphi \in \tilde{Y}([a, b])$, positive-entries vector $\mathbf{P} := (p_1, \dots, p_K) \in \mathbb{R}_{>0}^K$ (or from some subset thereof), and subset $\Omega \subset \mathbb{R}^K$ with regularity properties

$$cl(\Omega) = cl(int(\Omega)), \quad int(\Omega) \neq \emptyset, \quad (2)$$

find

$$\inf_{\mathbf{Q} \in \Omega} D_\varphi(\mathbf{Q}, \mathbf{P}), \quad (3)$$

provided that

$$\inf_{\mathbf{Q} \in \Omega} D_\varphi(\mathbf{Q}, \mathbf{P}) < \infty. \quad (4)$$

Remark 1. (a) When Ω is not closed but merely satisfies (2), then the infimum in (3) may not be reached in Ω although it is finite; additionally, Ω is a closed set, then a minimizer $\mathbf{Q}^* \in \Omega$ exists. In the subset where Ω is a closed convex set and $int(\Omega) \neq \emptyset$, (2) is satisfied and the minimizer $\mathbf{Q}^* \in \Omega$ in (3) is attained and even unique. When Ω is open and satisfies (2), then the infimum in (3) exists, but is generally reached at some generalized projection of $\mathbf{P}$ on Ω (see [33] for the Kullback–Leibler divergence case of probability measures, which extends to any generalized φ -divergence in our framework). However, in this paper, we only deal with finding the infimum/minimum in (3) (rather than a corresponding minimizer).

(b) Our approach is predestined for non- or semiparametric models. For instance, (2) is valid for the appropriate tubular neighborhoods of parametric models or for more general non-parametric settings, such as, e.g., shape constraints.

(c) Without further mentioning, the regularity condition (2) is supposed to hold in the full topology.

According to our work [29], the above-mentioned Problem 1 can be solved by a new dimension-free precise bare simulation (BS) method to be explained in the following. We first suppose

Condition 1. With $M_{\mathbf{P}} := \sum_{i=1}^K p_i > 0$, the divergence generator φ in (3) is such that its multiple $\tilde{\varphi} := M_{\mathbf{P}} \cdot \varphi$ satisfies (G1) to (G4') (i.e., $\tilde{\varphi} \in \tilde{Y}([a, b])$) and additionally there holds the representation

$$\tilde{\varphi}(t) = \sup_{z \in \mathbb{R}} \left(z \cdot t - \log \int_{\mathbb{R}} e^{z \cdot y} d\tilde{\zeta}(y) \right), \quad t \in \mathbb{R}, \quad (5)$$

for some probability measure $\tilde{\zeta}$ on the real line $\mathbb{R}$ such that the function $z \mapsto \int_{\mathbb{R}} e^{z \cdot y} d\tilde{\zeta}(y)$ is finite on some open interval containing zero (notice that the latter implies that $\int_{\mathbb{R}} y d\tilde{\zeta}(y) = 1$ and that $\tilde{\zeta}$ has light tails).

A detailed discussion on the representability (5) can be found in [34]. By means of this cornerstone Condition 1, we construct in [29] a sequence $(\tilde{\xi}_n^{\mathbf{W}})_{n \in \mathbb{N}}$ of $\mathbb{R}^K$ -valued random variables/random vectors (on an auxiliary probability space $(\mathcal{X}, \mathcal{A}, \mathbb{P})$) as follows: for any $n \in \mathbb{N}$ and any $k \in \{1, \dots, K\}$, let $n_k := \lfloor n \cdot \tilde{p}_k \rfloor$ where $\lfloor x \rfloor$ denotes the integer part of x and $\tilde{p}_k := p_k / M_{\mathbf{P}} > 0$. Thus, one has $\lim_{n \rightarrow \infty} \frac{n_k}{n} = \tilde{p}_k$. Moreover, we assume that $n \in \mathbb{N}$ is large enough, namely, $n \geq \max_{k \in \{1, \dots, K\}} \frac{1}{\tilde{p}_k}$, and decompose the set $\{1, \dots, n\}$ of all integers from 1 to n into the following disjoint blocks: $I_1^{(n)} := \{1, \dots, n_1\}$, $I_2^{(n)} := \{n_1 + 1, \dots, n_1 + n_2\}$, and so on until the last block $I_K^{(n)} := \{\sum_{k=1}^{K-1} n_k + 1, \dots, n\}$, which, therefore, contains all

integers from $n_1 + \dots + n_{K-1} + 1$ to n . Clearly, $I_k^{(n)}$ has $n_k \geq 1$ elements (i.e., $\text{card}(I_k^{(n)}) = n_k$, where $\text{card}(A)$ denotes the number of elements in a set A) for all $k \in \{1, \dots, K-1\}$, and the last block $I_K^{(n)}$ has $n - \sum_{k=1}^{K-1} n_k \geq 1$ elements, which, anyhow, satisfies $\lim_{n \rightarrow \infty} \text{card}(I_K^{(n)})/n = \tilde{p}_K$. Furthermore, consider a vector $\tilde{\mathbf{W}} := (\tilde{W}_1, \dots, \tilde{W}_n)$ where the $\tilde{W}_i$'s are i.i.d. copies of the random variable $\tilde{W}$ whose distribution is associated with the multiple divergence-generator $\tilde{\varphi}$ through (5), in the sense that $\mathbb{P}[\tilde{W} \in \cdot] = \tilde{\mathbb{Q}}[\cdot]$. We group the $\tilde{W}_i$'s according to the above-mentioned blocks and sum them up blockwise, in order to build the following K -component random vector

$$\xi_n^{\tilde{\mathbf{W}}} := \left(\frac{1}{n} \sum_{i \in I_1^{(n)}} \tilde{W}_i, \dots, \frac{1}{n} \sum_{i \in I_K^{(n)}} \tilde{W}_i \right). \quad (6)$$

For such a context, in [29], we obtain the following solution of Problem 1:

Theorem 1. Under Condition 1, there holds the “bare-simulation (BS) minimizability”

$$\inf_{\mathbf{Q} \in \Omega} D_{\varphi}(\mathbf{Q}, \mathbf{P}) = - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\xi_n^{\tilde{\mathbf{W}}} \in \Omega]$$

for any $\Omega \subset \mathbb{R}^K$ with regularity properties (2) and finiteness property (4).

Theorem 1 provides our principle for the approximation of the solution of the deterministic optimization problem (3). Indeed, by replacing the involved limit by its finite counterpart, we deduce for given large n

$$\inf_{\mathbf{Q} \in \Omega} D_{\varphi}(\mathbf{Q}, \mathbf{P}) \approx - \frac{1}{n} \log \mathbb{P}[\xi_n^{\tilde{\mathbf{W}}} \in \Omega]; \quad (7)$$

it remains to estimate the probability on the right-hand side of (7). The latter can be performed either by a naive estimator of the frequency of those replications of $\xi_n^{\tilde{\mathbf{W}}}$ which hit Ω , or, more efficiently, by some improved estimator, see [29] for details, where we give concrete construction methods as well as numerous solved cases; for the latter, for the sake of brevity, we mention only two important special cases. The first one deals with the class of power-divergence generators $\varphi := \tilde{c} \cdot \varphi_{\gamma} : \mathbb{R} \mapsto [0, \infty]$ (with arbitrary multiplier $\tilde{c} \in]0, \infty[$) defined by

$$\tilde{c} \cdot \varphi_{\gamma}(t) := \begin{cases} \tilde{c} \cdot \frac{t^{\gamma} - \gamma \cdot t + \gamma - 1}{\gamma \cdot (\gamma - 1)}, & \text{if } \gamma \in]-\infty, 0[\text{ and } t \in]0, \infty[, \\ \tilde{c} \cdot (-\log t + t - 1), & \text{if } \gamma = 0 \text{ and } t \in]0, \infty[, \\ \tilde{c} \cdot \frac{t^{\gamma} - \gamma \cdot t + \gamma - 1}{\gamma \cdot (\gamma - 1)}, & \text{if } \gamma \in]0, 1[\text{ and } t \in [0, \infty[, \\ \tilde{c} \cdot (t \cdot \log t + 1 - t), & \text{if } \gamma = 1 \text{ and } t \in [0, \infty[, \\ \tilde{c} \cdot \frac{(t-1)^2}{2}, & \text{if } \gamma = 2 \text{ and } t \in]-\infty, \infty[, \\ \tilde{c} \cdot \left\{ \frac{t^{\gamma} - \gamma \cdot t + \gamma - 1}{\gamma \cdot (\gamma - 1)} \cdot \mathbf{1}_{]0, \infty[}(t) + \left(\frac{1}{\gamma} - \frac{t}{\gamma - 1} \right) \cdot \mathbf{1}_{]-\infty, 0]}(t) \right\}, & \text{if } \gamma \in]2, \infty[\text{ and } t \in]-\infty, \infty[, \\ \infty, & \text{else,} \end{cases} \quad (8)$$

which—by (1)—generate (the vector-valued form of) the generalized power divergences given by

$$D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{Q}, \mathbf{P}) := \begin{cases} \tilde{c} \cdot \left\{ \frac{\sum_{k=1}^K (q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma \cdot (\gamma-1)} - \frac{1}{\gamma-1} \cdot \sum_{k=1}^K q_k + \frac{1}{\gamma} \cdot \sum_{k=1}^K p_k \right\}, & \text{if } \gamma \in]-\infty, 0[, \mathbf{P} \in \mathbb{R}_{>0}^K \text{ and } \mathbf{Q} \in \mathbb{R}_{>0}^K, \\ \tilde{c} \cdot \left\{ \sum_{k=1}^K p_k \cdot \log\left(\frac{p_k}{q_k}\right) + \sum_{k=1}^K q_k - \sum_{k=1}^K p_k \right\}, & \text{if } \gamma = 0, \mathbf{P} \in \mathbb{R}_{>0}^K \text{ and } \mathbf{Q} \in \mathbb{R}_{>0}^K, \\ \tilde{c} \cdot \left\{ \frac{\sum_{k=1}^K (q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma \cdot (\gamma-1)} - \frac{1}{\gamma-1} \cdot \sum_{k=1}^K q_k + \frac{1}{\gamma} \cdot \sum_{k=1}^K p_k \right\}, & \text{if } \gamma \in]0, 1[, \mathbf{P} \in \mathbb{R}_{>0}^K \text{ and } \mathbf{Q} \in \mathbb{R}_{\geq 0}^K, \\ \tilde{c} \cdot \left\{ \sum_{k=1}^K q_k \cdot \log\left(\frac{q_k}{p_k}\right) - \sum_{k=1}^K q_k + \sum_{k=1}^K p_k \right\}, & \text{if } \gamma = 1, \mathbf{P} \in \mathbb{R}_{>0}^K \text{ and } \mathbf{Q} \in \mathbb{R}_{\geq 0}^K, \\ \tilde{c} \cdot \sum_{k=1}^K \frac{(q_k - p_k)^2}{2 \cdot p_k}, & \text{if } \gamma = 2, \mathbf{P} \in \mathbb{R}_{>0}^K \text{ and } \mathbf{Q} \in \mathbb{R}^K, \\ \tilde{c} \cdot \left\{ \sum_{k=1}^K \frac{(q_k)^\gamma \cdot (p_k)^{1-\gamma}}{\gamma \cdot (\gamma-1)} \cdot l_{[0, \infty[}(q_k) - \frac{1}{\gamma-1} \cdot \sum_{k=1}^K q_k + \frac{1}{\gamma} \cdot \sum_{k=1}^K p_k \right\}, & \text{if } \gamma \in]2, \infty[, \mathbf{P} \in \mathbb{R}_{>0}^K \text{ and } \mathbf{Q} \in \mathbb{R}^K, \\ \infty, & \text{else;} \end{cases} \quad (9)$$

for a corresponding detailed literature embedding (including applications and transformations), the reader is referred to [29]. For any fixed $M_{\mathbf{P}} := \sum_{i=1}^K p_i \in]0, \infty[$, Condition 1 is satisfied for $\varphi := \tilde{c} \cdot \varphi_\gamma$ for all $\tilde{c} \in]0, \infty[$ and all $\gamma \in \mathbb{R} \setminus]1, 2[$, and thus the BS-minimizability concerning Theorem 1 can be applied. Notice that the case $\gamma \in]1, 2[$ has to be left out for technical reasons. The corresponding crucial simulation distributions $\mathbb{P}[\tilde{W} \in \cdot] = \tilde{\zeta}[\cdot]$ (cf. (5)) are given by the following:

(DIS1) a tilted stable distribution on $[0, \infty[$ for the case $\gamma \in]-\infty, 0[$;

(DIS2) the “Gamma($M_{\mathbf{P}} \cdot \tilde{c}$, $M_{\mathbf{P}} \cdot \tilde{c}$) distribution” for $\gamma = 0$;

(DIS3) the “Compound Poisson($\frac{M_{\mathbf{P}} \cdot \tilde{c}}{\gamma}$) – Gamma($\frac{M_{\mathbf{P}} \cdot \tilde{c}}{1-\gamma}$, $\frac{\gamma}{1-\gamma}$) distribution” for $\gamma \in]0, 1[$;

(DIS4) the “ $\frac{1}{M_{\mathbf{P}} \cdot \tilde{c}}$ -fold of Poisson($M_{\mathbf{P}} \cdot \tilde{c}$) distribution” for $\gamma = 1$;

(DIS5) the “Normal($1, \frac{1}{M_{\mathbf{P}} \cdot \tilde{c}}$) distribution” for $\gamma = 2$;

(DIS6) a distorted stable distribution on $]-\infty, \infty[$ for $\gamma \in]2, \infty[$.

for details see our paper [29].

The second important special case to be mentioned here, deals with

$$\varphi_{snKL, \tilde{c}}(t) := \begin{cases} \tilde{c} \cdot \left[t \cdot \log t + (t+1) \cdot \log\left(\frac{2}{t+1}\right) \right] \in [0, \infty[, & \text{if } t \in]0, \infty[, \\ \tilde{c} \cdot \log 2, & \text{if } t = 0, \\ \infty, & \text{if } t \in]-\infty, 0[, \end{cases} \quad (10)$$

which leads to

$$D_{\varphi_{snKL, \tilde{c}}}(\mathbf{Q}, \mathbf{P}) = \tilde{c} \cdot \left\{ \sum_{k=1}^K q_k \cdot \log\left(\frac{2q_k}{q_k + p_k}\right) + \sum_{k=1}^K p_k \cdot \log\left(\frac{2p_k}{q_k + p_k}\right) \right\}, \quad \text{if } \mathbf{P} \in \mathbb{R}_{>0}^K, \mathbf{Q} \in \mathbb{R}_{\geq 0}^K. \quad (11)$$

For any fixed $M_{\mathbf{P}} := \sum_{i=1}^K p_i \in]0, \infty[$, Condition 1 is satisfied for $\varphi := \varphi_{snKL, \tilde{c}}$ for all $\tilde{c} \in]0, \infty[$, and thus the BS-minimizability concerning Theorem 1 can be applied. The corresponding crucial simulation distribution $\mathbb{P}[\tilde{W} \in \cdot] = \tilde{\zeta}[\cdot]$ (cf. (5)) is given by the “ $\frac{1}{M_{\mathbf{P}} \cdot \tilde{c}}$ -fold of NegativeBinomial($M_{\mathbf{P}} \cdot \tilde{c}$, $\frac{1}{2}$)” (cf. [29]). For the special subcase $\tilde{c} = 1$ we derive

$$D_{\varphi_{snKL, 1}}(\mathbf{Q}, \mathbf{P}) = D_{\varphi_1}(\mathbf{Q}, (\mathbf{Q} + \mathbf{P})/2) + D_{\varphi_1}(\mathbf{P}, (\mathbf{Q} + \mathbf{P})/2), \quad \text{if } \mathbf{P} \in \mathbb{R}_{>0}^K, \mathbf{Q} \in \mathbb{R}_{\geq 0}^K \text{ and } \sum_{i=1}^K p_i = \sum_{i=1}^K q_i, \quad (12)$$

which means that in such a situation the divergence (11) can be rewritten as a sum of two generalized Kullback–Leibler divergences (cf. (9)). For the important subset that

$\sum_{i=1}^K p_i = \sum_{i=1}^K q_i = 1$, and thus both $\mathbf{P} = \mathbb{P}$ as well as $\mathbf{Q} = \mathbb{Q}$ are probability vectors, the divergence $D_{\varphi_{snKL,1}}(\mathbf{Q}, \mathbb{P})$ in (12) is the well-known (cf. [35–41]) *Jensen–Shannon divergence* (being also called symmetrized and normalized Kullback–Leibler divergence, symmetrized and normalized relative entropy, and capacitor discrimination).

For further examples the reader is referred to our paper [29]. In the latter, we also derive bare-simulation optimization versions for constraint sets Ω in—a strictly positive multiple $A > 0$ of—the probability simplex, to be explained in the following. First, we denote by $\mathbb{S}^K := \{\mathbf{Q} := (q_1, \dots, q_K) \in \mathbb{R}_{\geq 0}^K : \sum_{i=1}^K q_i = 1\}$ to be the simplex of probability vectors (probability simplex), and $\mathbb{S}_{>0}^K := \{\mathbf{Q} := (q_1, \dots, q_K) \in \mathbb{R}_{>0}^K : \sum_{i=1}^K q_i = 1\}$. For better emphasis, (as already performed above) for elements of these two sets we use the symbols $\mathbb{Q}, \mathbb{P}$ instead of $\mathbf{Q}, \mathbf{P}$, etc., but for their components we still use our notation q_k, p_k . Moreover, subsets of $\mathbb{S}^K$ or $\mathbb{S}_{>0}^K$ will be denoted by Ω instead of Ω , etc. As indicated above, in the following we deal with constraint sets of the form $A \cdot \Omega$ for some arbitrary $A \in]0, \infty[$, which automatically satisfy $\text{int}(A \cdot \Omega) = \emptyset$ in the *full topology* and thus the regularity condition (2) is violated (cf. Remark 1 (c)). Therefore, we need an adaption of the above-mentioned method. In more detail, we deal with

Problem 2. For pregiven $\varphi \in \tilde{Y}(]a, b[)$, positive-components vector $\mathbf{P} := (p_1, \dots, p_K) \in \mathbb{R}_{>0}^K$, and subset $A \cdot \Omega \subset A \cdot \mathbb{S}^K$ with regularity properties—in the relative topology (!!) —

$$\text{cl}(A \cdot \Omega) = \text{cl}(\text{int}(A \cdot \Omega)), \quad \text{int}(A \cdot \Omega) \neq \emptyset, \quad (13)$$

find

$$\inf_{\mathbf{Q} \in A \cdot \Omega} D_{\varphi}(\mathbf{Q}, \mathbf{P}), \quad (14)$$

provided that

$$\inf_{\mathbf{Q} \in A \cdot \Omega} D_{\varphi}(\mathbf{Q}, \mathbf{P}) < \infty$$

and that divergence generator φ additionally satisfies the Condition 1.

For the directed distance minimization Problem 2, we proceed (with the same notations) as above and construct the following K -component random vector (instead of $\xi_n^{\tilde{W}}$ in (6))

$$\xi_n^{w\tilde{W}} := \begin{cases} \left(\frac{\sum_{i \in I_1^{(n)}} \tilde{W}_i}{\sum_{k=1}^K \sum_{i \in I_k^{(n)}} \tilde{W}_i}, \dots, \frac{\sum_{i \in I_K^{(n)}} \tilde{W}_i}{\sum_{k=1}^K \sum_{i \in I_k^{(n)}} \tilde{W}_i} \right), & \text{if } \sum_{j=1}^n \tilde{W}_j \neq 0, \\ (\infty, \dots, \infty) =: \infty, & \text{if } \sum_{j=1}^n \tilde{W}_j = 0. \end{cases} \quad (15)$$

By construction, in case of $\sum_{j=1}^n \tilde{W}_j \neq 0$, the sum of the random K vector components of (15) are now automatically equal to one, but—as (depending on φ) the $\tilde{W}_i$'s may take both positive and negative values—these random components may be negative with a probability strictly greater than zero (respectively, non-negative with a probability strictly less than one). However, $\mathbb{P}[\xi_n^{w\tilde{W}} \in \mathbb{S}_{>0}^K] > 0$ since all the (identically distributed) random variables $\tilde{W}_i$ have and expectation of one (as a consequence of the assumed representability (5)); in case of $\mathbb{P}[\tilde{W}_1 > 0] = 1$, one has even $\mathbb{P}[\xi_n^{w\tilde{W}} \in \mathbb{S}_{>0}^K] = 1$. Summing up things, the probability $\mathbb{P}[\xi_n^{w\tilde{W}} \in \Omega]$ is strictly positive and finite at least for large n , whenever $\inf_{\mathbf{Q} \in A \cdot \Omega} D_{\varphi}(\mathbf{Q}, \mathbf{P})$ is finite.

As mentioned right after (9) above, the required representability (5) is satisfied for all (multiples of) the generators $\varphi(\cdot) := \tilde{c} \cdot \varphi_{\gamma}(\cdot)$ of (8) with $\tilde{c} \in]0, \infty[$ and $\gamma \in \mathbb{R} \setminus]1, 2[$ (cf. [29]). Within this context, for arbitrary constants $\check{A} > 0$ and $\check{c} > 0$ we define the auxiliary functions

$$H_{\gamma, \check{c}, \check{A}}(z) := \begin{cases} \frac{\check{c}}{\gamma \cdot (\gamma - 1)} \cdot \left\{ \check{A}^\gamma \cdot \left[1 - \frac{\gamma}{\check{c}} \cdot z \right]^{-(\gamma - 1)} - 1 - \gamma \cdot (\check{A} - 1) \right\}, & \text{if } \gamma \in]-\infty, 0[\cup]0, 1[\cup]2, \infty[\\ & \text{and } z \in \mathbb{R} \text{ such that } \gamma \cdot z \leq \check{c}, \\ z - \check{c} \cdot (1 - \check{A} + \log \check{A}), & \text{if } \gamma = 0 \text{ and } z \in \mathbb{R}, \\ \check{c} \cdot \left\{ 1 - \check{A} - \check{A} \cdot \left[\log \left(1 - \frac{z}{\check{c}} \right) - \log \check{A} \right] \right\}, & \text{if } \gamma = 1 \text{ and } z \in]-\infty, \check{c}[\end{cases}$$

we obtained (with slight rescaling) in [29] the following solution of Problem 2:

Theorem 2. Let $\mathbf{P} \in \mathbb{R}_{>0}^K$ with $M_{\mathbf{P}} := \sum_{i=1}^K p_i$, $\tilde{c} \in]0, \infty[$, $\gamma \in \mathbb{R} \setminus]1, 2[$ and $A \in]0, \infty[$ be arbitrary but fixed. Moreover, let $\check{c} := M_{\mathbf{P}} \cdot \tilde{c}$, $\check{A} := \frac{A}{M_{\mathbf{P}}}$ and $(\tilde{W}_i)_{i \in \mathbb{N}}$ be a family of independent and identically distributed $\mathbb{R}$ -valued random variables with probability distribution $\tilde{\mathbb{Q}}[\cdot] := \mathbb{P}[\tilde{W}_1 \in \cdot]$ being connected with the divergence generator $\varphi := \tilde{c} \cdot \varphi_\gamma(\cdot)$ via the representability (5) (cf. (DIS1)-(DIS6)). Then there holds the “bare-simulation (BS) minimizability”

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{Q}, \mathbf{P}) = H_{\gamma, \check{c}, \check{A}} \left(- \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{W}} \in \mathfrak{M}] \right) \quad (16)$$

for all sets $A \cdot \mathfrak{M} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (13) in the relative topology. Here, $\widetilde{\mathcal{M}}_\gamma := A \cdot \mathbb{S}_{>0}^K$ for $\gamma \in]-\infty, 0[$, respectively, $\widetilde{\mathcal{M}}_\gamma := A \cdot \mathbb{S}^K$ for $\gamma \in]0, 1[\cup]2, \infty[$.

Analogously to (7), Theorem 2 provides our principle for the approximation of the solution of the deterministic optimization problem (14). Indeed, by replacing the involved limit by its finite counterpart, we deduce for given large n

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{Q}, \mathbf{P}) \approx H_{\gamma, \check{c}, \check{A}} \left(- \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{W}} \in \mathfrak{M}] \right); \quad (17)$$

the probability in the latter can be estimated either by a naive estimator of the frequency of those replications of $\xi_n^{w\tilde{W}}$ which hit $\mathfrak{M}$, or more efficiently by some improved estimator; see [29] for details, where (for the case $A = 1$, with straightforward adaption to $A \neq 1$) we give concrete construction methods as well as numerous solved cases.

By means of the straightforward deterministic transformations, Theorem 2 carries straightforwardly over to the BS optimizability of, e.g., the following important quantities

$$\begin{aligned} R_\gamma(\mathbf{Q}, \mathbf{P}) &:= \frac{\log \left(\sum_{k=1}^K (q_k)^\gamma \cdot (p_k)^{1-\gamma} \right)}{\gamma \cdot (\gamma - 1)} \\ &= \frac{\log \left((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A + \gamma \cdot (\gamma - 1) \cdot D_{\varphi_\gamma}(\mathbf{Q}, \mathbf{P}) \right)}{\gamma \cdot (\gamma - 1)}, \\ &\quad \text{if } \gamma \in]-\infty, 0[\cup]0, 1[\cup]2, \infty[, \end{aligned} \quad (18)$$

(provided that all involved power divergences are finite), which are (by monotonicity and continuity) thus BS-minimizable on $\mathfrak{M} = A \cdot \mathfrak{M}$ for all $\gamma \in]-\infty, 0[\cup]0, 1[\cup]2, \infty[$:

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} R_\gamma(\mathbf{Q}, \mathbf{P}) = \lim_{n \rightarrow \infty} \frac{\log \left((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A + \gamma \cdot (\gamma - 1) \cdot H_{\gamma, \check{c}, \check{A}} \left(- \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{W}} \in \mathfrak{M}] \right) \right)}{\gamma \cdot (\gamma - 1)} \quad (19)$$

for all sets $A \cdot \mathfrak{M} \subset \widetilde{\mathcal{M}}_\gamma$ satisfying the regularity properties (13) in the relative topology; here, the simulation distribution $\mathbb{P}[\tilde{W} \in \cdot] = \tilde{\mathbb{Q}}[\cdot]$ (cf. (5)) is given by (DIS1), (DIS3), (DIS5) and (DIS6), respectively. The special subcase $A = 1$, $M_{\mathbf{P}} = 1$ in (18) (and thus, $\mathbf{Q}, \mathbf{P}$ are probability vectors $\mathbf{Q}, \mathbf{P}$) corresponds to the prominent Renyi divergences/distances [42] (in the scaling of, e.g., [4] and in probability vector form), see, e.g., [43] for a comprehensive

study of their properties. Notice that $R_\gamma(\mathbf{Q}, \mathbf{P})$ may become strictly negative, e.g., in the case that $(1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A \in]0, 1[$, however, in this case the variant

$$R_\gamma^{nn}(\mathbf{Q}, \mathbf{P}) := R_\gamma(\mathbf{Q}, \mathbf{P}) - \frac{\log((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A)}{\gamma \cdot (\gamma - 1)} \quad (20)$$

always stays non-negative and leads basically to the “same” optimization

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} R_\gamma^{nn}(\mathbf{Q}, \mathbf{P}) = \left(\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} R_\gamma(\mathbf{Q}, \mathbf{P}) \right) - \frac{\log((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A)}{\gamma \cdot (\gamma - 1)}. \quad (21)$$

For the cases $\gamma = 1$ and $\gamma = 0$, important transformations are the *modified Kullback–Leibler information (modified relative entropy)*

$$-M_{\mathbf{P}} \leq I(\mathbf{Q}, \mathbf{P}) := \sum_{k=1}^K q_k \cdot \log\left(\frac{q_k}{p_k}\right) = D_{\varphi_1}(\mathbf{Q}, \mathbf{P}) + A - M_{\mathbf{P}} \quad (22)$$

and the *modified reverse Kullback–Leibler information (modified reverse relative entropy)*

$$M_{\mathbf{P}} - A \leq \tilde{I}(\mathbf{Q}, \mathbf{P}) := \sum_{k=1}^K p_k \cdot \log\left(\frac{p_k}{q_k}\right) = D_{\varphi_0}(\mathbf{Q}, \mathbf{P}) + M_{\mathbf{P}} - A;$$

notice that $I(\mathbf{Q}, \mathbf{P})$ can become negative if $A < M_{\mathbf{P}}$ and $\tilde{I}(\mathbf{Q}, \mathbf{P})$ can become negative if $A > M_{\mathbf{P}}$ (see [30] for counterexamples). Nevertheless, in general, we obtain immediately from $\gamma = 1$ in Theorem 2 that

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} I(\mathbf{Q}, \mathbf{P}) = \lim_{n \rightarrow \infty} \left(A - M_{\mathbf{P}} - \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{M}] \right) \quad (23)$$

for all sets $A \cdot \mathfrak{M} \subset \widetilde{\mathcal{M}}_1$ satisfying the regularity properties (13) in the relative topology; here, the simulation distribution $\mathbb{P}[\tilde{W} \in \cdot] = \tilde{\zeta}[\cdot]$ (cf. (5)) is given by (DIS4). Moreover, by employing $\gamma = 0$ in Theorem 2 we deduce

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} \tilde{I}(\mathbf{Q}, \mathbf{P}) = \lim_{n \rightarrow \infty} \left(M_{\mathbf{P}} - A - \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{M}] \right) \quad (24)$$

for all sets $A \cdot \mathfrak{M} \subset \widetilde{\mathcal{M}}_0$ satisfying the regularity properties (13) in the relative topology; here, the simulation distribution $\mathbb{P}[\tilde{W} \in \cdot] = \tilde{\zeta}[\cdot]$ (cf. (5)) is given by (DIS2).

Remark 2. By taking the special case $\mathbf{P} := \mathbb{P}^{unif} := (\frac{1}{K}, \dots, \frac{1}{K})$ to be the probability vector of frequencies of the uniform distribution on $\{1, \dots, K\}$ (and thus $M_{\mathbf{P}} = 1$) in the above formulas (16) to (24), we can deduce many bare-simulation solutions of constrained optimization of various different versions of entropies $\mathbf{Q} \mapsto \mathcal{E}(\mathbf{Q})$; for details, see Sections VIII and XII in [29].

From (19), (21), (23) and (24) we deduce the approximations (for large $n \in \mathbb{N}$)

$$\inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} R_\gamma(\mathbf{Q}, \mathbf{P}) \approx \frac{\log \left((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A + \gamma \cdot (\gamma - 1) \cdot H_{\gamma, \tilde{\zeta}, \tilde{A}} \left(-\frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{M}] \right) \right)}{\gamma \cdot (\gamma - 1)}, \quad (25)$$

$$\begin{aligned} \inf_{\mathbf{Q} \in A \cdot \mathfrak{M}} R_\gamma^{nn}(\mathbf{Q}, \mathbf{P}) &\approx \frac{\log \left((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A + \gamma \cdot (\gamma - 1) \cdot H_{\gamma, \tilde{\zeta}, \tilde{A}} \left(-\frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{M}] \right) \right)}{\gamma \cdot (\gamma - 1)} \\ &\quad - \frac{\log((1 - \gamma) \cdot M_{\mathbf{P}} + \gamma \cdot A)}{\gamma \cdot (\gamma - 1)}, \end{aligned} \quad (26)$$

$$\inf_{\mathbf{Q} \in \mathcal{A} \cdot \mathfrak{Q}} I(\mathbf{Q}, \mathbf{P}) \approx \left(A - M_{\mathbf{P}} - \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{Q}] \right), \quad (27)$$

$$\inf_{\mathbf{Q} \in \mathcal{A} \cdot \mathfrak{Q}} \tilde{I}(\mathbf{Q}, \mathbf{P}) \approx \left(M_{\mathbf{P}} - A - \frac{1}{n} \log \mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{Q}] \right), \quad (28)$$

where for the involved $\mathbb{P}[\xi_n^{w\tilde{\mathbf{W}}} \in \mathfrak{Q}]$ one can again employ either a *naive estimator* of the frequency of those replications of $\xi_n^{w\tilde{\mathbf{W}}}$ which hit $\mathfrak{Q}$, or an improved estimator, see [29] for details.

4. Minimization Problems with Fuzzy Sets

Our above-mentioned BS framework can be applied to the—imprecise/inexact/vague information describing —*fuzzy sets* (cf. [19]) and optimization problems on divergences between those. Indeed, let $\mathcal{Y} = \{d_1, \dots, d_K\}$ be a finite set (called the *universe (of discourse)*), $C \subset \mathcal{Y}$ and $M^C : \mathcal{Y} \mapsto [0, 1]$ be a corresponding *membership function*, where $M^C(d_k)$ represents the degree/grade of membership of the element d_k to the set C ; accordingly, the object $C^* := \{\langle x, M^C(x) \rangle | x \in \mathcal{Y}\}$ is called a *fuzzy set in $\mathcal{Y}$* (or fuzzy subset of $\mathcal{Y}$). Moreover, if $A \subset \mathcal{Y}$ and $B \subset \mathcal{Y}$ are two unequal sets, then the corresponding membership functions M^A and M^B should be unequal. Furthermore, we model the *vector of membership degrees to C* by $\mathbf{P}^C := (p_k^C)_{k=1, \dots, K} := (M^C(d_k))_{k=1, \dots, K}$, which satisfies the *key constraint* $0 \leq p_k^C \leq 1$ for all $k \in \{1, \dots, K\}$ and, consequently, the *aggregated key constraint* $0 \leq \sum_{k=1}^K p_k^C \leq K$ (as a side remark, $\sum_{k=1}^K M^C(d_k)$ is called *power of the fuzzy set C^**). For divergence generators φ in $\tilde{Y}([a, b])$ with (say) $0 \leq a < 1 < b$ and for two sets $A, B \subset \mathcal{Y}$ we can apply (1) to the corresponding membership functions and define the *generalized φ -divergence $D_\varphi(B^*, A^*)$ between the fuzzy sets B^* and A^** (on the same universe $\mathcal{Y}$) as (cf. [44])

$$D_\varphi(B^*, A^*) := D_\varphi(\mathbf{P}^B, \mathbf{P}^A) = \sum_{k=1}^K p_k^A \cdot \varphi\left(\frac{p_k^B}{p_k^A}\right) = \sum_{k=1}^K M^A(d_k) \cdot \varphi\left(\frac{M^B(d_k)}{M^A(d_k)}\right) \geq 0 \quad (29)$$

(depending on φ , zero degree values may have to be excluded for finiteness). For instance, we can take $\varphi(t) := \varphi_1(t) := t \cdot \log t + 1 - t \in [0, \infty]$ for $t \in [0, \infty]$ (cf. (8)) to end up with a generalized Kullback–Leibler divergence (generalized relative entropy) between B^* and A^* given by (cf. (9))

$$D_{\tilde{c}, \varphi_1}(B^*, A^*) = D_{\tilde{c}, \varphi_1}(\mathbf{P}^B, \mathbf{P}^A) = \tilde{c} \cdot \left\{ \sum_{k=1}^K M^B(d_k) \cdot \log\left(\frac{M^B(d_k)}{M^A(d_k)}\right) - \sum_{k=1}^K M^B(d_k) + \sum_{k=1}^K M^A(d_k) \right\} \geq 0, \\ \text{if } M^A(d_k) > 0 \text{ and } M^B(d_k) \geq 0 \text{ for all } k = 1, \dots, K;$$

this contrasts the choice $\varphi(t) := \check{\varphi}(t) := t \cdot \log t \in [-\frac{1}{e}, \infty]$ of [22]

$$D_{\check{\varphi}}(B^*, A^*) = I(\mathbf{P}^B, \mathbf{P}^A) = \tilde{c} \cdot \left\{ \sum_{k=1}^K M^B(d_k) \cdot \log\left(\frac{M^B(d_k)}{M^A(d_k)}\right) \right\}, \\ \text{if } M^A(d_k) > 0 \text{ and } M^B(d_k) \geq 0 \text{ for all } k = 1, \dots, K,$$

which they call *fuzzy expected information for discrimination in favor of B against A* , and which *may become negative* (cf. see the discussion after (22) and [30] in a more general context). Returning to the general case (29), as a special case of the above-mentioned BS concepts, we can tackle optimization problems of the type

$$\inf_{B^* \in \Omega^*} D_\varphi(B^*, A^*) := \inf_{\mathbf{P}^B \in \Omega} D_\varphi(\mathbf{P}^B, \mathbf{P}^A)$$

where Ω^* is a collection of fuzzy sets (on the same universe $\mathcal{U}$) whose membership degree vectors form the set Ω satisfying (2) and (4). Because of the inequality type key constraint

$$0 \leq M^B(d_k) \leq 1 \quad \text{for all } k \in \{1, \dots, K\}$$

which is incorporated into Ω , and which implies that Ω is contained in the K -dimensional unit hypercube and in particular $0 \leq \sum_{k=1}^K p_k^B \leq K$, Theorem 1 (and thus, (7)) applies correspondingly (e.g., to the generalized power divergences $D_{\tilde{c} \cdot \varphi_\gamma}(B^*, A^*) = D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}^B, \mathbf{P}^A)$ given by (9) and the generalized Jensen–Shannon divergence $D_{\varphi_{snKL, \tilde{c}}}(B^*, A^*) = D_{\varphi_{snKL, \tilde{c}}}(\mathbf{P}^B, \mathbf{P}^A)$ given by (11))—unless there is a more restrictive constraint that violates (2) such as, e.g., $\sum_{k=1}^K p_k^B = c$ with $c \leq K$ for which Theorem 2 (and hence, (17))—as well as its consequences (19), (21), (23), (24) (and thus, (25)–(28)) can be employed.

The above-mentioned considerations can be extended to the recent concept of ν -rung orthopair fuzzy sets (cf. [28]) and divergences between those. Indeed, for $C \subset \mathcal{U}$, besides a membership function $M^C : \mathcal{U} \mapsto [0, 1]$ one additionally models a *nonmembership function* $N^C : \mathcal{U} \mapsto [0, 1]$, where $N^C(d_k)$ represents the degree/grade of nonmembership of the element d_k to the set C . Moreover, if $A \subset \mathcal{U}$ and $B \subset \mathcal{U}$ are unequal sets, then the corresponding nonmembership functions N^A and N^B should be unequal. For fixed $\nu \in [1, \infty]$, the *key constraint*

$$0 \leq (M^C(d_k))^\nu + (N^C(d_k))^\nu \leq 1 \quad \text{for all } k \in \{1, \dots, K\} \quad (30)$$

is required to be satisfied, too. Accordingly, the object $C^{**} := \{ \langle x, M^C(x), N^C(x) \rangle | x \in \mathcal{U} \}$ is called a ν -rung orthopair fuzzy set in $\mathcal{U}$ (or ... subset of $\mathcal{U}$). The object C^{**} is called *intuitionistic fuzzy set in $\mathcal{U}$* (cf. [45]) in case of $\nu = 1$, and *Pythagorean fuzzy set in $\mathcal{U}$* (cf. [46,47]) in the case of $\nu = 2$. For the choice $\nu = 1$ together with $N^C(x) := 1 - M^C(x)$, the object C^{**} can be regarded as an extended representation of the fuzzy set C^* in $\mathcal{U}$.

For any ν -rung orthopair fuzzy set C^{**} in $\mathcal{U}$, we model the corresponding *vector of concatenated membership and nonmembership degrees* to C by $\mathbf{P}^C := (p_k^C)_{k=1, \dots, 2K} := (M^C(d_1), \dots, M^C(d_K), N^C(d_1), \dots, N^C(d_K))$, which, due to (30), satisfies the *aggregated key constraint*

$$0 \leq \sum_{k=1}^{2K} (p_k^C)^\nu \leq K; \quad (31)$$

in other words, $\mathbf{P}^C$ lies (within the $2K$ -dimensional Euclidean space) in the intersection of the first/positive orthant with the ν -norm ball centered at the origin and with radius $K^{1/\nu}$. Analogously to (29), we can define the *generalized φ -divergence* $D_\varphi(B^{**}, A^{**})$ between the ν -rung orthopair fuzzy sets B^{**} and A^{**} (on the same universe $\mathcal{U}$) as (cf. [44])

$$\begin{aligned} D_\varphi(B^{**}, A^{**}) &:= D_\varphi(\mathbf{P}^B, \mathbf{P}^A) = \sum_{k=1}^{2K} p_k^A \cdot \varphi\left(\frac{p_k^B}{p_k^A}\right) \\ &= \sum_{k=1}^K \left\{ M^A(d_k) \cdot \varphi\left(\frac{M^B(d_k)}{M^A(d_k)}\right) + N^A(d_k) \cdot \varphi\left(\frac{N^B(d_k)}{N^A(d_k)}\right) \right\} \geq 0 \end{aligned} \quad (32)$$

respectively, as its variant (cf. [44])

$$\begin{aligned} D_\varphi^{var}(B^{**}, A^{**}) &:= D_\varphi((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu) := \sum_{k=1}^{2K} (p_k^A)^\nu \cdot \varphi\left(\frac{(p_k^B)^\nu}{(p_k^A)^\nu}\right) \\ &= \sum_{k=1}^K \left\{ (M^A(d_k))^\nu \cdot \varphi\left(\frac{(M^B(d_k))^\nu}{(M^A(d_k))^\nu}\right) + (N^A(d_k))^\nu \cdot \varphi\left(\frac{(N^B(d_k))^\nu}{(N^A(d_k))^\nu}\right) \right\} \geq 0. \end{aligned} \quad (33)$$

For the special choice $\varphi(t) := \varphi_{snKL,1}(t)$ (cf. (10)) and $\nu = 1$, the definition (32) leads to the generalized Jensen–Shannon divergence between B^{**} and A^{**} given by

$$\begin{aligned} D_{\varphi_{snKL,1}}(B^{**}, A^{**}) &:= D_{\varphi_{snKL,1}}(\mathbf{P}^B, \mathbf{P}^A) \\ &= \sum_{k=1}^K \left\{ M^B(d_k) \cdot \log \left(\frac{2M^B(d_k)}{M^B(d_k) + M^A(d_k)} \right) + M^A(d_k) \cdot \log \left(\frac{2M^A(d_k)}{M^B(d_k) + M^A(d_k)} \right) \right. \\ &\quad \left. + N^B(d_k) \cdot \log \left(\frac{2N^B(d_k)}{N^B(d_k) + N^A(d_k)} \right) + N^A(d_k) \cdot \log \left(\frac{2N^A(d_k)}{N^B(d_k) + N^A(d_k)} \right) \right\} \geq 0, \\ &\text{if } M^A(d_k) > 0, N^A(d_k) > 0, M^B(d_k) \geq 0, N^B(d_k) \geq 0, \text{ for all } k \in \{1, \dots, K\}; \end{aligned}$$

this coincides with the symmetric information measure between B^{**} and A^{**} of [23]. For the special choice $\varphi(t) := \varphi_1(t) := t \cdot \log t + 1 - t \in [0, \infty[$ with $t \in [0, \infty[$ (cf. (8)), $\nu = 1$, $N^A(x) := 1 - M^A(x)$, $N^B(x) := 1 - M^B(x)$ and thus (31) turning into $\sum_{k=1}^{2K} (p_k^A)^\nu = \sum_{k=1}^{2K} (p_k^B)^\nu = K$, one can straightforwardly see that the outcoming generalized Kullback–Leibler divergence (generalized relative entropy) between B^{**} and A^{**} given by

$$\begin{aligned} 0 \leq D_{\varphi_1}(B^{**}, A^{**}) &= D_{\varphi_1}(\mathbf{P}^B, \mathbf{P}^A) \\ &= \sum_{k=1}^K M^B(d_k) \cdot \log \left(\frac{M^B(d_k)}{M^A(d_k)} \right) - \sum_{k=1}^K M^B(d_k) + \sum_{k=1}^K M^A(d_k) + \sum_{k=1}^K N^B(d_k) \cdot \log \left(\frac{N^B(d_k)}{N^A(d_k)} \right) - \sum_{k=1}^K N^B(d_k) + \sum_{k=1}^K N^A(d_k) \\ &= \sum_{k=1}^K M^B(d_k) \cdot \log \left(\frac{M^B(d_k)}{M^A(d_k)} \right) + \sum_{k=1}^K N^B(d_k) \cdot \log \left(\frac{N^B(d_k)}{N^A(d_k)} \right), \end{aligned}$$

if $M^A(d_k) > 0, N^A(d_k) > 0$ and $M^B(d_k) \geq 0, N^B(d_k) \geq 0$ for all $k = 1, \dots, K$,

coincides with $D_{\check{\varphi}}(B^{**}, A^{**})$ where $\check{\varphi}(t) := t \cdot \log t$; the latter divergence was used, e.g., in [22] under the name *average fuzzy information for discrimination in favor of B against A*.

Returning to the general context, in terms of the divergences (32) and (33), we can tackle—as a special case of the above-mentioned BS concepts—optimization problems of the type

$$\begin{aligned} \inf_{B^{**} \in \Omega^{**}} D_{\varphi}(B^{**}, A^{**}) &:= \inf_{\mathbf{P}^B \in \Omega} D_{\varphi}(\mathbf{P}^B, \mathbf{P}^A) \quad \text{respectively} \\ \inf_{B^{**} \in \Omega^{**}} D_{\varphi}^{var}(B^{**}, A^{**}) &:= \inf_{\mathbf{P}^B \in \Omega} D_{\varphi}((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu), \end{aligned}$$

where Ω^{**} is a collection of ν -rung orthopair fuzzy sets whose concatenated membership–nonmembership degree vectors form the set Ω satisfying (2) and (4) as well as (30) for $C := B$. Because of the latter, Theorem 1 (and thus, (7)) applies correspondingly (e.g., to the generalized power divergences $D_{\tilde{c} \cdot \varphi_\gamma}(B^{**}, A^{**}) = D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}^B, \mathbf{P}^A)$ given by (9) and the generalized Jensen–Shannon divergence $D_{\varphi_{snKL,\tilde{c}}}(B^{**}, A^{**}) = D_{\varphi_{snKL,\tilde{c}}}(\mathbf{P}^B, \mathbf{P}^A)$ given by (11))—unless there is a more restrictive constraint that violates (2) such as, e.g., $\sum_{k=1}^{2K} p_k^B = c$ with $c \leq K$ for which Theorem 2 (and thus, (17))—as well as its consequences (19), (21), (23), (24) (and thus, (25)–(28)) can be employed; such a situation appears, e.g., in the case $\nu = 1$ together with $N^B(x) := 1 - M^B(x)$, which leads to $c = K$.

For the ν -rung orthopair fuzzy sets C^{**} in $\mathcal{U}$, we can also further “flexibilize” our divergences by additionally incorporating the *hesitancy degree* of the element d_k to C , which is defined as

$$H^C(d_k) := \left(1 - (M^C(d_k))^\nu - (N^C(d_k))^\nu \right)^{1/\nu} \in [0, 1]$$

(cf. [28]), and which implies the *key constraint*

$$(H^C(d_k))^\nu + (M^C(d_k))^\nu + (N^C(d_k))^\nu = 1 \quad \text{for all } k \in \{1, \dots, K\}. \quad (34)$$

Accordingly, the object $C^{***} := \{\langle x, M^C(x), N^C(x), H^C(x) \rangle | x \in \mathcal{X}\}$ can be regarded as an extended representation of the ν -rung orthopair fuzzy set C^{**} in $\mathcal{X}$. For C^{***} , we model the corresponding *vector of concatenated membership, nonmembership and hesitancy degrees* to C by

$$\mathbf{P}^C := (p_k^C)_{k=1, \dots, 3K} := (M^C(d_1), \dots, M^C(d_K), N^C(d_1), \dots, N^C(d_K), H^C(d_1), \dots, H^C(d_K))$$

which, due to (34), satisfies the *aggregated key constraint*

$$\sum_{k=1}^{3K} (p_k^C)^\nu = K;$$

in other words, $\mathbf{P}^C$ lies (within the $3K$ -dimensional Euclidean space) in the intersection of the first/positive orthant with the ν -norm sphere centered at the origin and with radius $K^{1/\nu}$. Analogously to (32) and (33), we can define the *generalized φ -divergence* $D_\varphi(B^{***}, A^{***})$ between the extended representation type ν -rung orthopair fuzzy sets B^{***} and A^{***} (on the same universe $\mathcal{X}$) as (cf. [44])

$$\begin{aligned} D_\varphi(B^{***}, A^{***}) &:= D_\varphi((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu) := \sum_{k=1}^{3K} (p_k^A)^\nu \cdot \varphi\left(\frac{(p_k^B)^\nu}{(p_k^A)^\nu}\right) \\ &= \sum_{k=1}^K \left\{ (M^A(d_k))^\nu \cdot \varphi\left(\frac{(M^B(d_k))^\nu}{(M^A(d_k))^\nu}\right) + (N^A(d_k))^\nu \cdot \varphi\left(\frac{(N^B(d_k))^\nu}{(N^A(d_k))^\nu}\right) \right. \\ &\quad \left. + (H^A(d_k))^\nu \cdot \varphi\left(\frac{(H^B(d_k))^\nu}{(H^A(d_k))^\nu}\right) \right\} \geq 0; \end{aligned} \quad (35)$$

for technical reasons, we do not deal with its variant

$$\begin{aligned} D_\varphi^{var}(B^{***}, A^{***}) &:= D_\varphi(\mathbf{P}^B, \mathbf{P}^A) = \sum_{k=1}^{3K} p_k^A \cdot \varphi\left(\frac{p_k^B}{p_k^A}\right) \\ &= \sum_{k=1}^K \left\{ M^A(d_k) \cdot \varphi\left(\frac{M^B(d_k)}{M^A(d_k)}\right) + N^A(d_k) \cdot \varphi\left(\frac{N^B(d_k)}{N^A(d_k)}\right) + H^A(d_k) \cdot \varphi\left(\frac{H^B(d_k)}{H^A(d_k)}\right) \right\} \geq 0. \end{aligned}$$

For instance, by taking the special choice $\nu = 2$ and $\varphi(t) := \varphi_{snKL,1}(t)$ (cf. (10)) in (35), we arrive at the *Jensen–Shannon divergence* between B^{***} and A^{***} of the form $D_{\varphi_{snKL,1}}(B^{***}, A^{***}) := D_{\varphi_{snKL,1}}(\mathbf{P}^B, \mathbf{P}^A)$, which—by the virtue of (35) and (11)—coincides with the (*squared*) *Pythagorean fuzzy set Jensen–Shannon divergence measure* between B^{***} and A^{***} of [24].

To continue with the general context, as a particular application of the above-mentioned BS concepts, we can tackle the general optimization problems of the generalized power divergence type

$$\inf_{B^{***} \in \Omega^{***}} D_{\tilde{c} \cdot \varphi_\gamma}(B^{***}, A^{***}) := \inf_{\mathbf{P}^B \in \Omega} D_{\tilde{c} \cdot \varphi_\gamma}((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu), \quad \gamma \in \mathbb{R} \setminus]1, 2[, \quad \tilde{c} \in]0, \infty[, \quad (36)$$

where Ω^{***} is a collection of extended representation type ν -rung orthopair fuzzy sets whose concatenated membership–nonmembership–hesitancy degree vectors form the set Ω , satisfying (34) for $C := B$ (for each member) as well as the regularity properties (13) in the relative topology. Thus, for the minimization of (36) we can apply our Theorem 2 (and, consequently, (17)) by choosing there (with a slight abuse of notation) $A = M_P = K$. Of course, we can also apply our BS optimization method to the corresponding Renyi divergences $R_\gamma(B^{***}, A^{***}) := R_\gamma((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu)$ (via (19)), $R_\gamma^{nn}(B^{***}, A^{***}) := R_\gamma^{nn}((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu)$ (via (19) and (21)) as well as $I(B^{***}, A^{***}) := I((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu)$ (via (23)), $\tilde{I}(B^{***}, A^{***}) := \tilde{I}((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu)$ (via (24)), and employ the correspondingly applied approximations (25)–(28). For instance, by applying (20) with $A = M_P = K$ (with a slight

abuse of notation) we arrive at the non-negative γ -order Renyi divergence between ν -rung orthopair fuzzy sets given by

$$\begin{aligned} 0 \leq R_\gamma^{nn}(B^{***}, A^{***}) &:= R_\gamma^{nn}((\mathbf{P}^B)^\nu, (\mathbf{P}^A)^\nu) := \frac{1}{\gamma \cdot (\gamma - 1)} \cdot \left[\log \left(\sum_{k=1}^{3K} ((p_k^B)^\nu)^\gamma \cdot ((p_k^A)^\nu)^{1-\gamma} \right) - \log(K) \right] \\ &= \frac{1}{\gamma \cdot (\gamma - 1)} \cdot \left[\log \left(\sum_{k=1}^K \left\{ ((M^B(d_k))^\nu)^\gamma \cdot ((M^A(d_k))^\nu)^{1-\gamma} + ((N^B(d_k))^\nu)^\gamma \cdot ((N^A(d_k))^\nu)^{1-\gamma} \right. \right. \right. \\ &\quad \left. \left. \left. + ((H^B(d_k))^\nu)^\gamma \cdot ((H^A(d_k))^\nu)^{1-\gamma} \right\} \right) - \log(K) \right] \geq 0; \quad \gamma \in]-\infty, 0[\cup]0, 1[\cup [2, \infty[; \end{aligned} \quad (37)$$

depending on γ , zero degree values may have to be excluded for finiteness. As a side remark, let us mention that our divergence (37) contrasts to the recent (first) divergence of [48] who basically uses a different scaling, the product $\prod_{k=1}^K$ instead of the sum $\sum_{k=1}^K$, as well as $\frac{(p_k^A)^\nu + (p_k^B)^\nu}{2}$ instead of $(p_k^A)^\nu$. By appropriately applying (19) and (21), we can tackle with our BS method for $\gamma \in]-\infty, 0[\cup]0, 1[\cup [2, \infty[$ the minimization problem $\inf_{B^{***} \in \Omega^{***}} R_\gamma^{nn}(B^{***}, A^{***})$, where Ω^{***} is a collection of extended representation type ν -rung orthopair fuzzy sets B^{***} whose concatenated membership–nonmembership–hesitancy degree vectors form the set Ω satisfying (34) for $C := B$ (for each member) as well as the regularity properties (13) in the relative topology.

We can also apply our BS optimization method to “crossover cases” $D_\varphi(B^*, \mathbf{P}) := D_\varphi(\mathbf{P}^B, \mathbf{P})$ and $D_{\mathbf{P}, \varphi}(A^*) := D_\varphi(\mathbf{P}, \mathbf{P}^A)$ (instead of (29), $D_\varphi(B^{**}, \mathbf{P}) := D_\varphi(\mathbf{P}^B, \mathbf{P})$ and $D_\varphi(\mathbf{P}, A^{**}) := D_\varphi(\mathbf{P}, \mathbf{P}^A)$ (instead of (32)), $D_\varphi^{var}(B^{**}, \mathbf{P}) := D_\varphi((\mathbf{P}^B)^\nu, \mathbf{P})$ and $D_\varphi^{var}(\mathbf{P}, A^{**}) := D_\varphi(\mathbf{P}, (\mathbf{P}^A)^\nu)$ (instead of (33)), $D_\varphi(B^{***}, \mathbf{P}) := D_\varphi((\mathbf{P}^B)^\nu, \mathbf{P})$ and $D_\varphi(\mathbf{P}, A^{***}) := D_\varphi(\mathbf{P}, (\mathbf{P}^A)^\nu)$ (instead of (35)), $R_\gamma(B^{***}, \mathbf{P}) := R_\gamma((\mathbf{P}^B)^\nu, \mathbf{P})$ and $R_\gamma(\mathbf{P}, A^{***}) := R_\gamma(\mathbf{P}, (\mathbf{P}^A)^\nu)$, $R_\gamma^{nn}(B^{***}, \mathbf{P}) := R_\gamma^{nn}((\mathbf{P}^B)^\nu, \mathbf{P})$ and $R_\gamma^{nn}(\mathbf{P}, A^{***}) := R_\gamma^{nn}(\mathbf{P}, (\mathbf{P}^A)^\nu)$, $I(B^{***}, \mathbf{P}) := I((\mathbf{P}^B)^\nu, \mathbf{P})$ and $I(\mathbf{P}, A^{***}) := I(\mathbf{P}, (\mathbf{P}^A)^\nu)$, $\tilde{I}(B^{***}, \mathbf{P}) := \tilde{I}((\mathbf{P}^B)^\nu, \mathbf{P})$ and $\tilde{I}(\mathbf{P}, A^{***}) := \tilde{I}(\mathbf{P}, (\mathbf{P}^A)^\nu)$, $D_{\tilde{C} \cdot \varphi_\gamma}(B^{***}, \mathbf{P}) := D_{\tilde{C} \cdot \varphi_\gamma}((\mathbf{P}^B)^\nu, \mathbf{P})$ and $D_{\tilde{C} \cdot \varphi_\gamma}(\mathbf{P}, A^{***}) := D_{\tilde{C} \cdot \varphi_\gamma}(\mathbf{P}, (\mathbf{P}^A)^\nu)$, where $\mathbf{P} \in \mathbb{R}^{dim}$ (respectively, $\mathbb{R}_{\geq 0}^{dim}$, $\mathbb{R}_{> 0}^{dim}$, $A \cdot \mathbb{S}_{\geq 0}^{dim}$ or $A \cdot \mathbb{S}_{> 0}^{dim}$) is a general vector (not necessarily induced by a fuzzy set) having the same dimension dim (namely, K , $2K$ or $3K$) as the fuzzy set induced vector to be compared with. For instance, if we apply Remark 2 to $\mathbf{P} := \mathbb{P}^{unif} := (\frac{1}{3K}, \dots, \frac{1}{3K})$ and $\mathbf{Q} := (\mathbf{P}^B)^\nu$ and employ the corresponding straightforward application of the general results of Sections VIII and XII of [29], then we end up (via the appropriately applied Theorem 2) with the BS optimization results of $D_{\tilde{C} \cdot \varphi_\gamma}(B^{***}, \mathbb{P}^{unif})$, $R_\gamma(B^{***}, \mathbb{P}^{unif})$, $I(B^{***}, \mathbb{P}^{unif})$ and $\tilde{I}(B^{***}, \mathbb{P}^{unif})$, which can be deterministically transformed into the BS optimization results of various different versions of entropies $\mathcal{E}(B^{***}) := \mathcal{E}((\mathbf{P}^B)^\nu)$ of ν -rung orthopairs fuzzy sets B^{***} , where $\mathcal{E}(\cdot)$ is any entropy in Chapter VIII of [29].

As a final remark of this section, let us mention that we can carry over the above-mentioned definitions and optimizations to (classical, intuitionistic, Pythagorean and ν -rung orthopair) L -fuzzy sets, where the range of the membership functions, nonmembership functions and hesitancy functions is an appropriately chosen lattice L (rather than $L = [0, 1]$); for the sake of brevity, the details are omitted here.

5. Minimization Problems with Basic Belief Assignments

Our BS framework also covers—imprecise/inexact/vague information describing—*basic belief assignments* from Dempster–Shafer evidence theory (cf. [20,21]) and optimization problems on the divergences between those. Indeed, let $\mathcal{D} = \{d_1, \dots, d_K\}$ be a finite set (called the *frame of discernment*) of mutually exclusive and collectively exhaustive events d_k . The corresponding power set of $\mathcal{D}$ is denoted by $2^\mathcal{D}$ and has 2^K elements; we enumerate this by $2^\mathcal{D} := \{A_1, \dots, A_{2^K}\}$, where for convenience we set $A_1 := \emptyset$. A

mapping $M : 2^{\mathcal{Y}} \mapsto [0, 1]$ is called a *basic belief assignment (BBA)* (sometimes alternatively called basic probability assignment (BPA)) if it satisfies the two conditions

$$M(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in 2^{\mathcal{Y}}} M(A) = 1. \quad (38)$$

Here, the belief mass $M(A)$ reflects, e.g., the trust degree of evidence to proposition $A \in 2^{\mathcal{Y}}$. From this, one can build the belief function $Bel : 2^{\mathcal{Y}} \mapsto [0, 1]$ by $Bel(A) := \sum_{B: B \subseteq A} M(B)$ and the plausibility function $Pl : 2^{\mathcal{Y}} \mapsto [0, 1]$ by $Pl(A) := \sum_{B: B \cap A \neq \emptyset} M(B)$. Moreover, we model the 2^K -dimensional vector of (M -based) BBA values (vector of (M -based) belief masses) by $\mathbf{P}^M := (p_k^M)_{k=1, \dots, 2^K} := (M(A_k))_{k=1, \dots, 2^K}$, which satisfies the key constraint $0 \leq p_k^M \leq 1$ for all $k \in \{1, \dots, 2^K\}$ and, by virtue of (38), the aggregated key constraint $\sum_{k=1}^{2^K} p_k^M = 1$. Hence, $\mathbf{P}^M$ lies formally in the 2^K -dimensional simplex $\mathbb{S}^{2^K}$ (but generally not in the corresponding probability vector-describing $\mathbb{S}^K$).

For divergence generators φ in $\tilde{Y}([a, b])$ with (say) $0 \leq a < 1 < b$ and for two BBAs M_1, M_2 on the same frame of discernment $\mathcal{Y}$, we can apply (1) to the corresponding vectors of BBA values and define the generalized φ -divergence $D_\varphi(M_2, M_1)$ between the BBAs M_2 and M_1 (in short, *Belief- φ -divergence*) as

$$D_\varphi(M_2, M_1) := D_\varphi(\mathbf{P}^{M_2}, \mathbf{P}^{M_1}) = \sum_{k=1}^{2^K} p_k^{M_1} \cdot \varphi\left(\frac{p_k^{M_2}}{p_k^{M_1}}\right) = \sum_{k=1}^{2^K} M_1(A_k) \cdot \varphi\left(\frac{M_2(A_k)}{M_1(A_k)}\right) \geq 0. \quad (39)$$

This definition (39) of the Belief- φ -divergence was first given in our paper [44]; later on, [49] used formally the same definition under the different (but deficient) assumptions that φ is only convex and satisfies $\varphi(1) = 0$, which leads to a violation of the basic divergence properties (take, e.g., $\varphi(t) = 0$ for all t , which implies $D_\varphi(\mathbf{P}^{M_2}, \mathbf{P}^{M_1}) = 0$ even if $\mathbf{P}^{M_2} \neq \mathbf{P}^{M_1}$; such an effect is excluded in our setup due to our assumption (G4') being part of $\tilde{Y}([a, b])$); as a technical remark let us mention that (as already indicated in Section 2) depending on φ , zero belief masses may have to be excluded for finiteness in (39). For instance, we can take in (39) the special case $\varphi(t) := \varphi_{snKL,1}(t)$ (cf. (10)) to end up with the *Belief-Jensen-Shannon divergence* of [25,26] who applies this to multi-sensor data fusion. As another special case we can take $\varphi(t) := \varphi_{1/2}(t)$ (cf. (8)) to end up with the 4-times square of the *Hellinger distance of BBAs* of [27], who use this for characterizing the degree of conflict between BBAs. To continue with the general context, as a particular application of the above-mentioned BS concepts, we can tackle general optimization problems of the type

$$\inf_{M_2 \in \Omega^{BBA}} D_{\tilde{c} \cdot \varphi_\gamma}(M_2, M_1) := \inf_{\mathbf{P}^{M_2} \in \Omega} D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}^{M_2}, \mathbf{P}^{M_1}) \quad \gamma \in \mathbb{R} \setminus]1, 2[, \quad \tilde{c} \in]0, \infty[, \quad (40)$$

where Ω^{BBA} is a collection of BBAs whose vectors of BBA-values form the set $\Omega \in \mathbb{S}^{2^K}$ satisfying the regularity properties (13) in the relative topology; hence, we can apply our Theorem 2 (and thus, (17)) for the minimization of (40), by taking 2^K instead of K as well as (with a slight abuse of notation) $A = M_{\mathbf{P}^{M_1}} = 1$. Of course, we can also apply our BS optimization method to the corresponding Renyi divergences $R_\gamma(M_2, M_1) := R_\gamma(\mathbf{P}^{M_2}, \mathbf{P}^{M_1})$ (via (19)), $R_\gamma^{nn}(M_2, M_1) := R_\gamma^{nn}(\mathbf{P}^{M_2}, \mathbf{P}^{M_1})$ (via (19) and (21)) as well as $I(M_2, M_1) := I(\mathbf{P}^{M_2}, \mathbf{P}^{M_1})$ (via (23)), $\tilde{I}(M_2, M_1) := \tilde{I}(\mathbf{P}^{M_2}, \mathbf{P}^{M_1})$ (via (24)), and employ the correspondingly applied approximations (25)–(28). For instance, by applying (18) and (19) we arrive at

$$R_\gamma(M_2, M_1) := R_\gamma(\mathbf{P}^{M_2}, \mathbf{P}^{M_1}) = \frac{\log\left(\sum_{k=1}^{2^K} (M_2(A_k))^\gamma \cdot (M_1(A_k))^{1-\gamma}\right)}{\gamma \cdot (\gamma - 1)},$$

if $\gamma \in]-\infty, 0[\cup]0, 1[\cup]2, \infty[$,

and

$$\inf_{M_2 \in \Omega^{BBA}} R_\gamma(M_2, M_1) = \inf_{\mathbf{P}^{M_2} \in \Omega} R_\gamma(\mathbf{P}^{M_2}, \mathbf{P}^{M_1}) = \lim_{n \rightarrow \infty} \frac{\log \left(1 + \gamma \cdot (\gamma - 1) \cdot H_{\gamma, \tilde{c}, \tilde{A}} \left(-\frac{1}{n} \log \mathbb{P} \left[\tilde{\xi}_n^{w\tilde{\mathbf{W}}} \in \Omega \right] \right) \right)}{\gamma \cdot (\gamma - 1)}$$

where $\tilde{\xi}_n^{w\tilde{\mathbf{W}}}$ is chosen according to Theorem 2 with 2^K instead of K as well as $A = M_{\mathbf{P}^{M_1}} = 1$.

We can also apply our BS optimization method to the divergences between the rescaling of BBAs. For instance, let $\check{M}(A) := \frac{M(A)}{2^{|A|-1}}$ ($A \in 2^{\mathcal{X}}$) with the convention that $\frac{0}{0} := 0$, and denote the corresponding vector $\check{\mathbf{P}}^{\check{M}} := (p_k^{\check{M}})_{k=1, \dots, 2^K} := (\check{M}(A_k))_{k=1, \dots, 2^K}$. Accordingly, we define the *generalized φ -divergence* $D_\varphi(\check{M}_2, \check{M}_1)$ between the rescaled BBAs $\check{M}_2$ and $\check{M}_1$ (in short, *rescaled Belief- φ -divergence*) as (cf. [44])

$$\begin{aligned} D_\varphi(\check{M}_2, \check{M}_1) &:= D_\varphi(\mathbf{P}^{\check{M}_2}, \mathbf{P}^{\check{M}_1}) = \sum_{k=1}^{2^K} p_k^{\check{M}_1} \cdot \varphi \left(\frac{p_k^{\check{M}_2}}{p_k^{\check{M}_1}} \right) = \sum_{k=2}^{2^K} \check{M}_1(A_k) \cdot \varphi \left(\frac{\check{M}_2(A_k)}{\check{M}_1(A_k)} \right) \\ &= \sum_{k=2}^{2^K} \frac{M_1(A_k)}{2^{|A_k|-1}} \cdot \varphi \left(\frac{M_2(A_k)}{M_1(A_k)} \right) \geq 0, \end{aligned} \quad (41)$$

where for $A_1 := \emptyset$ we have used the convention that $0 \cdot \varphi(\frac{0}{0}) := 0$ (depending on φ , other zero rescaled belief masses may have to be excluded for finiteness). For the corresponding minimization problem

$$\inf_{\check{M}_2 \in \Omega^{resBBA}} D_\varphi(\check{M}_2, \check{M}_1) := \inf_{\mathbf{P}^{\check{M}_2} \in \Omega} D_\varphi(\mathbf{P}^{\check{M}_2}, \mathbf{P}^{\check{M}_1})$$

where Ω^{resBBA} is a collection of rescaled BBAs whose vectors of rescaled BBA values form the set Ω satisfying (2) and (4), we can apply Theorem 1 and (7) (with 2^K instead of K)—unless there is a more restrictive constraint that violates (2) such as, e.g., $\sum_{k=1}^{2^K} p_k^{\check{M}_2} = A$, for which Theorem 2 (and thus, (17)) can be employed.

We can also apply our BS optimization method to “crossover cases” $D_\varphi(M_2, \mathbf{P}) := D_\varphi(\mathbf{P}^{M_2}, \mathbf{P})$ and $D_\varphi(\mathbf{P}, M_1) := D_\varphi(\mathbf{P}, \mathbf{P}^{M_1})$ (instead of (39)), $D_\varphi(\check{M}_2, \mathbf{P}) := D_\varphi(\mathbf{P}^{\check{M}_2}, \mathbf{P})$ and $D_\varphi(\mathbf{P}, \check{M}_1) := D_\varphi(\mathbf{P}, \mathbf{P}^{\check{M}_1})$ (instead of (41)), $R_\gamma(M_2, \mathbf{P}) := R_\gamma(\mathbf{P}^{M_2}, \mathbf{P})$ and $R_\gamma(\mathbf{P}, M_1) := R_\gamma(\mathbf{P}, \mathbf{P}^{M_1})$, $R_\gamma^{nn}(M_2, \mathbf{P}) := R_\gamma^{nn}(\mathbf{P}^{M_2}, \mathbf{P})$ and $R_\gamma^{nn}(\mathbf{P}, M_1) := R_\gamma^{nn}(\mathbf{P}, \mathbf{P}^{M_1})$, $I(M_2, \mathbf{P}) := I(\mathbf{P}^{M_2}, \mathbf{P})$ and $I(\mathbf{P}, M_1) := I(\mathbf{P}, \mathbf{P}^{M_1})$, $\tilde{I}(M_2, \mathbf{P}) := \tilde{I}(\mathbf{P}^{M_2}, \mathbf{P})$ and $\tilde{I}(\mathbf{P}, M_1) := \tilde{I}(\mathbf{P}, \mathbf{P}^{M_1})$, $D_{\tilde{c} \cdot \varphi_\gamma}(M_2, \mathbf{P}) := D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}^{M_2}, \mathbf{P})$ and $D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}, M_1) := D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}, \mathbf{P}^{M_1})$, where $\mathbf{P} \in \mathbb{R}^{2^K}$ (respectively, $\mathbb{R}_{\geq 0}^{2^K}$, $\mathbb{R}_{> 0}^{2^K}$, $A \cdot \mathbb{S}_{\geq 0}^{2^K}$ or $A \cdot \mathbb{S}_{> 0}^{2^K}$) is a general vector (not necessarily induced by a (rescaled) BBA) having the same dimension (namely, 2^K) as the (rescaled) BBA-induced vector to be compared with. For instance, if we apply Remark 2 to $\mathbf{P} := \mathbb{P}^{unif} := (\frac{1}{2^K}, \dots, \frac{1}{2^K})$ and $\mathbf{Q} := \mathbf{P}^{M_2}$ and employ the corresponding straightforward application of the general results of Sections VIII and XII of [29], then we end up (via the appropriately applied Theorem 2) with the BS optimization results of $D_{\tilde{c} \cdot \varphi_\gamma}(M_2, \mathbb{P}^{unif})$, $R_\gamma(M_2, \mathbb{P}^{unif})$, $I(M_2, \mathbb{P}^{unif})$ and $\tilde{I}(M_2, \mathbb{P}^{unif})$, which can be deterministically transformed into the BS optimization results of various different versions of entropies $\mathcal{E}(M_2) := \mathcal{E}(\mathbf{P}^{M_2})$ of BBAs M_2 , where $\mathcal{E}(\cdot)$ is any entropy in Chapter VIII of [29].

To give a more explicit example concerning the preceding paragraph, our BS method of Theorem 2 and (17) (with 2^K instead of K as well as $A = M_{\mathbf{P}^{M_1}} = 1$) applies to the crossover case

$$\inf_{M \in \Omega^{BBA}} D_{\tilde{c} \cdot \varphi_\gamma}(M, \mathbf{P}) := \inf_{\mathbf{P}^M \in \Omega} D_{\tilde{c} \cdot \varphi_\gamma}(\mathbf{P}^M, \mathbf{P}) \quad \gamma \in \mathbb{R} \setminus]1, 2[, \quad \tilde{c} \in]0, \infty[,$$

where $\mathbf{P}^M$ is a vector of M -based BBA values and Ω^{BBA} is as above, and $\mathbf{P}$ is a vector whose sum of components may not necessarily be one. For instance, for the special choice $\gamma = 1$, i.e., $\varphi(t) := \varphi_1(t) := t \cdot \log t + 1 - t \in [0, \infty[$ (cf. (8)), $\mathbf{P}^M := (p_k^M)_{k=1, \dots, 2^K} := (M(A_k))_{k=1, \dots, 2^K}$, $\mathbf{P} := (p_k)_{k=1, \dots, 2^K}$ with $p_k := 2^{|A_k|} - 1$ employing the cardinality $|A_k|$ of A_k , and the usual convention $0 \cdot \log(\frac{0}{0}) := 0$, we end up with (cf. (9))

$$D_{\varphi_1}(M, \mathbf{P}) = D_{\varphi_1}(\mathbf{P}^M, \mathbf{P}) = \sum_{k=2}^{2^K} M(A_k) \cdot \log\left(\frac{M(A_k)}{2^{|A_k|} - 1}\right) - 1 + \sum_{k=2}^{2^K} (2^{|A_k|} - 1) =: -\mathcal{E}^{DE}(M) - 1 + \sum_{k=2}^{2^K} (2^{|A_k|} - 1)$$

where $\mathcal{E}^{DE}(M) := -\sum_{k=2}^{2^K} M(A_k) \cdot \log\left(\frac{M(A_k)}{2^{|A_k|} - 1}\right) \geq 0$ is nothing but (a multiple of) Deng's entropy of the BBA M (cf. [50], see also, e.g., [51]). Of course, we can also deal with the optimization of corresponding Renyi divergences $\inf_{M \in \Omega^{BBA}} R_\gamma(M, \mathbf{P}) = \inf_{\mathbf{P}^M \in \Omega} R_\gamma(\mathbf{P}^M, \mathbf{P})$ by applying (18) and (19). For the "reverse-argument" crossover case $\inf_{M \in \Omega^{BBA}} D_\varphi(\mathbf{P}, M) := \inf_{\mathbf{P} \in \Omega} D_\varphi(\mathbf{P}, \mathbf{P}^M)$ one can apply Theorem 1 (and thus, (7))—unless there is a more restrictive constraint, which violates (2) such as, e.g., $\sum_{k=1}^{2^K} p_k = A$, for which Theorem 2 (and hence, (17))—as well as its consequences (19), (21), (23), (24) (and thus, (25)–(28)) can be employed.

6. Conclusions

In this paper, we have—in terms of *generalized φ -divergences*—quantified the dissimilarity between fuzzy sets and some of their subsets such as ν -rung orthopair fuzzy sets and extended representation type ν -rung orthopair fuzzy sets; crossover-cases of *generalized φ -divergences* between fuzzy set types and general vectors have been treated, too. Moreover, we have presented how to tackle corresponding constrained minimization problems by appropriately applying our recently developed dimension-free *bare (pure) simulation method* of [29]. Afterwards, an analogous program is carried out by defining *generalized φ -divergences* between (rescaled) basic belief assignments as well as between (rescaled) basic belief assignments and vectors, and tackling corresponding constrained minimization problems.

As an outlook for future work, it is interesting to study ways of efficiently simulating the corresponding minimizers (in addition to the here-addressed minima). This ongoing research will appear in a future paper.

Author Contributions: Both authors have equally contributed to all aspects of this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to thank the reviewers for their helpful suggestions concerning a more comfortable readability of the paper. W. Stummer is grateful to the Sorbonne Université Paris for its multiple partial financial support and especially to the LPSM for its multiple great hospitality. M. Broniatowski thanks very much the FAU Erlangen-Nürnberg for its partial financial support and hospitality.

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

References

1. Csizsár, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci.* **1963**, *A-8*, 85–108.
2. Ali, M.S.; Silvey, D. A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. B* **1966**, *28*, 131–140. [[CrossRef](#)]
3. Morimoto, T. Markov processes and the H-theorem. *J. Phys. Soc. Jpn.* **1963**, *18*, 328–331. [[CrossRef](#)]
4. Liese, F.; Vajda, I. *Convex Statistical Distances*; Teubner: Leipzig, Germany, 1987.
5. Read, T.R.C.; Cressie, N.A.C. *Goodness-of-Fit Statistics for Discrete Multivariate Data*; Springer: New York, NY, USA, 1988.
6. Vajda, I. *Theory of Statistical Inference and Information*; Kluwer: Dordrecht, The Netherlands, 1989.

7. Csiszár, I.; Shields, P.C. *Information Theory and Statistics: A Tutorial*; Now Publishers: Hanover, NH, USA, 2004.
8. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.
9. Liese, F.; Miescke, K.J. *Statistical Decision Theory: Estimation, Testing, and Selection*; Springer: New York, NY, USA, 2008.
10. Liese, F.; Vajda, I. On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory* **2006**, *52*, 4394–4412. [[CrossRef](#)]
11. Vajda, I.; van der Meulen, E.C. Goodness-of-fit criteria based on observations quantized by hypothetical and empirical percentiles. In *Handbook of Fitting Statistical Distributions with R*; Karian, Z.A., Dudewicz, E.J., Eds.; CRC: Heidelberg, Germany, 2010; pp. 917–994.
12. Reid, M.D.; Williamson, R.C. Information, divergence and risk for binary experiments. *J. Mach. Learn. Res.* **2011**, *12*, 731–817.
13. Basseville, M. Divergence measures for statistical data processing—An annotated bibliography. *Signal Process.* **2013**, *93*, 621–633. [[CrossRef](#)]
14. Lindsay, B.G. Statistical distances as loss functions in assessing model adequacy. In *The Nature of Scientific Evidence*; Taper, M.P., Lele, S.R., Eds.; The University of Chicago Press: Chicago, IL, USA, 2004; pp. 439–487.
15. Lindsay, B.G.; Markatou, M.; Ray, S.; Yang, K.; Chen, S.-C. Quadratic distances on probabilities: A unified foundation. *Ann. Statist.* **2008**, *36*, 983–1006. [[CrossRef](#)]
16. Markatou, M.; Sofikitou, E. Non-quadratic distances in model assessment. *Entropy* **2018**, *20*, 464. . [[CrossRef](#)] [[PubMed](#)]
17. Markatou, M.; Chen, Y.; Statistical distances and the construction of evidence functions for model adequacy. *Front. Ecol. Evol.* **2019**, *7*, 447. . [[CrossRef](#)]
18. Broniatowski, M.; Stummer, W. A unifying framework for some directed distances in statistics. In *Geometry and Statistics*; Nielsen, F., Rao, A.S.R.S., Rao, C.R., Eds.; Handbook of Statistics; Academic Press: Cambridge MA, USA, 2022; Volume 46, pp. 145–223.
19. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [[CrossRef](#)]
20. Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [[CrossRef](#)]
21. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
22. Bhandari, D.; Pal, N.R. Some new information measures for fuzzy sets. *Inf. Sci.* **1993**, *67*, 209–228. [[CrossRef](#)]
23. Vlachos, I.K.; Sergiadis, G.D. Intuitionistic fuzzy information—Applications to pattern recognition. *Pattern Recogn. Lett.* **2007**, *28*, pp. 197–206. [[CrossRef](#)]
24. Xiao, F.; Ding, W. Divergence measure of Pythagorean fuzzy sets and its application in medical diagnosis. *Appl. Soft. Comput. J.* **2019**, *79*, 254–267. [[CrossRef](#)]
25. Xiao, F. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Inf. Fusion* **2019**, *46*, 23–32. [[CrossRef](#)]
26. Xiao, F. A new divergence measure for belief functions in D-S evidence theory for multisensor data fusion. *Inf. Sci.* **2020**, *514*, 462–483. [[CrossRef](#)]
27. Li, J.; Xie, B.; Jin, Y.; Hu, Z.; Zhou, L. Weighted conflict evidence combination method based on hellinger distance and the belief entropy. *IEEE Access* **2020**, *8*, 225507–225521. [[CrossRef](#)]
28. Yager, R.R. Generalized orthopair fuzzy sets. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1222–1230. [[CrossRef](#)]
29. Broniatowski, M.; Stummer, W. A precise bare simulation approach to the minimization of some distances. I. Foundations. *IEEE Trans. Inf. Theory* **2023**, *69*, 3062–3120. [[CrossRef](#)]
30. Stummer, W.; Vajda, I. On divergences of finite measures and their applicability in statistics and information theory. *Statistics* **2010**, *44*, 169–187. [[CrossRef](#)]
31. Broniatowski, M.; Keziou, A. Minimization of ϕ -divergences on sets of signed measures. *Stud. Scient. Math. Hungar.* **2006**, *43*, 403–442.
32. Broniatowski, M.; Stummer, W. Some universal insights on divergences for statistics, machine learning and artificial intelligence. In *Geometric Structures of Information*; Nielsen, F., Ed.; Springer Nature: Cham, Switzerland, 2019; pp. 149–211.
33. Csiszár, I. Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* **1984**, *12*, 768–793. [[CrossRef](#)]
34. Broniatowski, M.; Stummer, W. On a cornerstone condition of bare-simulation distance/divergence optimization. In *Geometric Science of Information GSI 2023*; Nielsen, F., Barbaresco, F., Eds.; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2023; Volume 14071, pp. 105–116.
35. Burbea, C.; Rao, C.R. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 489–495. [[CrossRef](#)]
36. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [[CrossRef](#)]
37. Pardo, M.C.; Vajda, I. About distances of discrete distributions satisfying the data processing theorem of information theory. *IEEE Trans. Inf. Theory* **1997**, *43*, 1288–1293. [[CrossRef](#)]
38. Topsøe, F. Some inequalities for information divergence and related measures of discrimination. *IEEE Trans. Inf. Theory* **2000**, *46*, 1602–1609. [[CrossRef](#)]
39. Endres, D.M.; Schindelin, J.E. A new metric for probability distributions. *IEEE Trans. Inf. Theory* **1993**, *40*, 1858–1860. [[CrossRef](#)]
40. Vajda, I. On metric divergences of probability measures. *Kybernetika* **2009**, *45*, 885–900.
41. Sason, I. Tight bounds for symmetric divergence measures and a new inequality relating f -divergences. In *Proceedings of the 2015 IEEE Information Theory Workshop (ITW)*, Jerusalem, Israel, 26 April–1 May 2015; 5p.

42. Renyi, A. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; Neyman, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
43. van Erven, T.; Harremoës, P. Renyi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [[CrossRef](#)]
44. Broniatowski, M.; Stummer, W. A precise bare simulation approach to the minimization of some distances. *Foundations. arXiv* **2021**, arXiv:2107.01693v1; Correction in *arXiv* **2022**, arXiv:2107.01693v2.
45. Atanassov, K.T. Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **1986**, *20*, 87–96. [[CrossRef](#)]
46. Yager, R.R. Pythagorean fuzzy subsets. In *Proceedings of the 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, Edmonton, AB, Canada, 24–28 June 2013; pp. 57–61.
47. Yager, R.R.; Abbasov, A.M. Pythagorean membership grades, complex numbers, and decision making. *Int. J. Intell. Syst.* **2013**, *28*, 436–452. [[CrossRef](#)]
48. Verma, R. Multiple attribute group decision-making based on order- α divergence and entropy measures under q-rung orthopair fuzzy environment. *Int. J. Intell. Syst.* **2020**, *35*, 718–750. [[CrossRef](#)]
49. Huang, J.; Song, X.; Xiao, F.; Cao, Z.; Lin, C.-T. Belief f -divergence for EEG complexity evaluation. *Inf. Sci.* **2023**, *643*, 119189. [[CrossRef](#)]
50. Deng, Y. Deng entropy. *Chaos Solitons Fract.* **2016**, *91*, 549–553. [[CrossRef](#)]
51. Kang, B.; Deng, Y. The maximum Deng entropy. *IEEE Access* **2020**, *7*, 120758–120765. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.