



Cross and Relative Entropies of Mass Functions Inspired by the Plausibility Entropy

Xinyang Deng^{1,*} and Wen Jiang¹

¹School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China

Abstract

Related concepts of entropy play a very important role in dealing with uncertainty in terms of Shannon's information theory. However, for uncertain information involving epistemic uncertainty, which is usually modelled by using Dempster-Shafer theory, the concepts of cross entropy and relative entropy are still not well defined currently. Facing this issue, by reviewing and importing existing related work, this study gives new definitions of cross entropy and relative entropy of mass functions, which are respectively named as cross plausibility entropy and relative plausibility entropy since they are both based on an uncertainty measure called plausibility entropy. The properties of cross and relative plausibility entropies are also given, which shows a strong connection with classical cross entropy and relative entropy in Shannon's information theory. An example of application regarding parameter estimation is provided to show the effectiveness and reasonability of the presented entropies, which has implemented the parameter estimation for a

generalized Bernoulli distribution with plausibility distribution observations.

Keywords: cross entropy, relative entropy, plausibility entropy, mass functions, dempster-shafer theory, uncertainty.

Nomenclature

Symbol	Meaning
Ω	Frame of discernment (FOD)
2^Ω	Power set of a FOD
m	Mass function
Bel	Belief function
Pl	Plausibility function
$Pl_{\rightarrow} P_m$	Plausibility transformation of a mass function
$U(m)$	Dezert's entropy of a mass function
$U(m_1, m_2)$	Dezert and Dambreville's cross entropy
$U(m_1 m_2)$	Dezert and Dambreville's relative entropy
$H_D(m)$	Deng's entropy of a mass function
$H_D(m_1, m_2)$	Gao et al.'s cross entropy
$H_{pignistic}(m)$	Pignistic entropy of a mass function
$H_{pignistic}(m_1, m_2)$	Cross pignistic entropy
$H_{Yager}(m)$	Yager entropy of a mass function
$H_{Yager}(m_1, m_2)$	Cross Yager entropy
$H_{Pl}(m)$	Plausibility entropy of a mass function
$H_{Pl}(m_1, m_2)$	Cross plausibility entropy
$H_{Pl}(m_1 m_2)$	Relative plausibility entropy
$H_S(P)$	Shannon's entropy of a probability distribution
$H_S(P_1, P_2)$	Shannon's cross entropy
$H_S(P_1 P_2)$	Shannon's relative entropy



Academic Editor:

Jun Liu

Submitted: 19 March 2025

Accepted: 23 July 2025

Published: 18 September 2025

Vol. 2, No. 3, 2025.

10.62762/CJIF.2025.592789

*Corresponding author:

✉ Xinyang Deng

xinyang.deng@nwpu.edu.cn

Citation

Deng, X., & Jiang, W. (2025). Cross and Relative Entropies of Mass Functions Inspired by the Plausibility Entropy. *Chinese Journal of Information Fusion*, 2(3), 212–222.



© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

1 Introduction

How to represent and measure the uncertainty is one of the central issues in information sciences. In general, the uncertainty can be briefly classified into random uncertainty and epistemic uncertainty [33]. Attributed to Shannon's innovative contributions [1], quantifying the randomness of uncertain information is solved through the mathematical framework of information theory. Dempster-Shafer theory [2, 3], also known as belief function theory, is a widely used framework to represent information with epistemic uncertainty, where a mathematical structure called mass functions is provided to simultaneously describe discord and non-specificity involved in the given information [4]. However, measuring the uncertainty of mass functions is not yet well solved currently, especially it still has not a consensus with respect to the definitions of entropy, cross entropy, and relative entropy of mass functions.

With respect to the entropy of mass functions, many researchers have paid attention on the problem [5–7]. Klir [8] has proposed generalized information theory (GIT) which aims to generalize Shannon's information theory for probabilities to various uncertainty theories including imprecise probabilities, fuzzy sets, belief functions, and so on. However, the proposed aggregated uncertainty (AU) in GIT for mass functions is of some shortcomings [9], especially some of the underlying axiomatic requirements of AU have been challenged [10]. Recent years, with the proposal of Deng's entropy [11], a new entropy measure for calculating the uncertainty of a mass function, the research of uncertainty measures in Dempster-Shafer theory has welcomed a "strong resurgence" [12]. Many novel entropy definitions of mass functions have been put forward [13–15]. For example, Jirousek and Shenoy [10] designed an entropy measure for mass functions by combining plausibility transformation and weighted Hartley entropy. Zhou and Deng [16] proposed a fractal-based belief entropy on the basis of Deng's entropy. In terms of belief intervals of single elements, Moral-Garcia and Abellan [17] developed an uncertainty measure of mass functions which is analogous to AU. Besides, Cui and Deng [18] presented a total uncertainty measure of mass functions based on plausibility function, which is called plausibility entropy. Facing existing uncertainty measures, Dezert and Tchamova [19] have raised the effectiveness problem of uncertainty measures, and provided four desiderata to check if an uncertainty measure is effective, and a new effective measure of uncertainty for mass functions has been proposed in

[20].

Cross entropy and relative entropy are another two important concepts according to Shannon's information theory. Although the research of entropy of mass functions is flourishing, cross entropy and relative entropy of mass functions, however, are relatively rare. This matter of fact is easy to understand, because cross and relative entropies are strongly connected with the concept of entropy, their forms are usually on the basis of the definition of entropy. Many existing entropy measures of mass functions are hard to yield corresponding cross and relative entropies. Very recently, Dezert and Dambreville [21] have provided definitions of cross entropy and relative entropy of mass functions in terms of Dezert's effective measure of uncertainty presented in [20]. In addition, Gao et al. [22] proposed a definition of cross entropy of mass functions based on Deng's entropy [11]. However, as will be analyzed in this paper, the existing definitions of cross entropy of mass functions are of some defects in underlying properties, new cross entropy, as well as corresponding relative entropy, of mass functions are required, which is the purpose of the study.

Specifically, inspired by the plausibility entropy, a total uncertainty measure of mass functions, presented in [18], new cross entropy and relative entropy of mass functions are given in this paper, which are named as cross plausibility entropy and relative plausibility entropy, respectively. The properties of cross and relative plausibility entropies are given, which shows a strong connection with classical cross entropy and relative entropy in Shannon's information theory. In addition, an illustrative example in parameter estimation is given to show the potential application of the presented entropies.

The remainder of the paper is organized as follows. Section 2 briefly introduces the basic knowledge of Dempster-Shafer theory. Related work regarding existing definitions of cross entropy of mass functions are reviewed in Section 3. Then, new cross and relative entropies of mass functions are given in Section 4. An example of application is provided in Section 5. Finally, Section 6 concludes the study.

2 Basics of Dempster-Shafer theory

Dempster-Shafer theory [2, 3] has provided a well-defined framework to represent and deal with uncertainty information with epistemic uncertainty. In this theory, the set of possible answers to a given

question of interest is called as a frame of discernment (FOD), denoted as $\Omega = \{\theta_1, \theta_2, \dots, \theta_n\}$, in which the set is collectively exhaustive and all elements in a FOD are mutually exclusive. The power set of FOD Ω is represented by 2^Ω .

In order to represent the uncertain information involving epistemic uncertainty, mass functions, also known as basic probability assignments (BPAs), are defined in Dempster-Shafer theory. A mass function is a mapping from the power set of a FOD to interval $[0, 1]$, denoted as $m : 2^\Omega \rightarrow [0, 1]$, satisfying the following conditions

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in 2^\Omega} m(A) = 1 \quad (1)$$

where A is called as a focal element of m if $m(A) > 0$, and $m(A)$ measures the belief assigned exactly to A . In general, a probability distribution can be treated as a Bayesian mass function in Dempster-Shafer theory, where $\forall A$ with $|A| \geq 2$ there is $m(A) = 0$.

Belief function Bel and plausibility function Pl are two equivalent forms of mass functions, which respectively express the lower bound and upper bound of the support degree of a set A based on a given mass function m , $A \subseteq \Omega$. Given a mass function m , Bel and Pl are defined as follows

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

$$Pl(A) = 1 - Bel(\bar{A}) = \sum_{B \cap A \neq \emptyset} m(B) \quad (3)$$

where $\bar{A} = \Omega - A$. For each $A \subseteq \Omega$, there is $Bel(A) \leq Pl(A)$, and $[Bel(A), Pl(A)]$ is called the belief interval of A . In general, the wider the belief interval of a set A , the larger the uncertainty it contributes to the whole mass function. Therefore, in Dempster-Shafer theory, given a FOD Ω , the vacuous mass function m_γ , in which $m_\gamma(\Omega) = 1$, has the largest uncertainty.

3 Related work

In this section, two existing definitions of cross entropy of mass functions, proposed very recently, are reviewed.

3.1 Dezert and Dambreville's cross entropy

Dezert [20] has given an entropy definition to measure the uncertainty of a mass function m on a FOD Ω

$$U(m) = \sum_{A \in 2^\Omega} s(A) \quad (4)$$

with

$$s(A) = -m(A)(1 - u(A)) \ln(m(A)) + u(A)(1 - m(A)) \quad (5)$$

where $\ln(\cdot)$ represents the natural logarithm, and $u(A) = Pl(A) - Bel(A)$ for any $A \in 2^\Omega$.

On the basis of $U(m)$, Dezert and Dambreville [21] have further proposed a cross entropy of mass functions

$$U(m_1, m_2) = \sum_{A \in 2^\Omega} s_{1,2}(A) \quad (6)$$

with

$$s_{1,2}(A) = -m_1(A)(1 - u_1(A)) \ln(m_2(A)) + u_1(A)(1 - m_2(A)) \quad (7)$$

where $u_1(A) = Pl_{m_1}(A) - Bel_{m_1}(A)$ for any $A \in 2^\Omega$.

In addition, a relative entropy of mass functions was also proposed in [21] as below

$$U(m_1 || m_2) = \sum_{A \subseteq \Omega} m_1(A)(1 - u_1(A)) \ln \left(\frac{m_1(A)}{m_2(A)} \right) + \sum_{A \subseteq \Omega} u_1(A)(m_1(A) - m_2(A)) \quad (8)$$

It is proved in [21] that $U(m)$, $U(m_1, m_2)$, and $U(m_1 || m_2)$ can respectively degenerate into classical Shannon's entropy, cross entropy, and relative entropy (also known as Kullback-Leibler (KL) divergence), and they own a relation $U(m_1, m_2) = U(m_1 || m_2) + U(m_1)$.

3.2 Gao et al.'s cross entropy

Gao et al. [22] presented another cross entropy definition of mass functions as below

$$H_D(m_1, m_2) = - \sum_{A \subseteq \Omega} m_1(A) \log_2 \frac{m_2(A)}{2^{|A|} - 1} \quad (9)$$

in which the underlying entropy definition for mass functions is based on Deng's entropy [11] as follows

$$H_D(m) = - \sum_{A \subseteq \Omega} m(A) \log_2 \frac{m(A)}{2^{|A|} - 1} \quad (10)$$

Since the relative entropy inspired by Deng's entropy is not defined at present, quantitative relation between $H_D(m)$ and $H_D(m_1, m_2)$ is not established yet.

4 New cross and relative entropies of mass functions

4.1 Analysis of existing cross entropy definitions

In this subsection, these two cross entropy definitions of mass functions mentioned in the above section are analyzed simply.

At first, Gao et al.'s cross entropy $H_D(m_1, m_2)$ has been questioned in [21] because the underlying Deng's entropy $H_D(m)$ is non-effective. According to the four desiderata proposed in [19], "unicity of max value of MoU" (D4), i.e., $\text{MoU}(m_\gamma) > \text{MoU}(m)$ for any $m \neq m_\gamma$ in which m_γ is the vacuous mass function, is not satisfied by $H_D(m)$. In terms of Deng's entropy, given a FOD Ω , the vacuous mass function, i.e., $m(\Omega) = 1$, does not have the maximum uncertainty. Please refer to literature [21, 31, 32] for more details.

Secondly, based on a similar consideration, Dezert and Dambreville's cross entropy $U(m_1, m_2)$ may also not be recommended since its underlying entropy measure $U(m)$ violates the monotonicity [17] of a rational uncertainty measure in belief function theory. Specifically, the monotonicity means that $\text{Uncertainty}(m_1) \leq \text{Uncertainty}(m_2)$ holds if $\forall A \subseteq \Omega : [\text{Bel}_{m_1}(A), \text{Pl}_{m_1}(A)] \subseteq [\text{Bel}_{m_2}(A), \text{Pl}_{m_2}(A)]$ for arbitrary BPAs m_1, m_2 defined on a same FOD Ω . Literature [23] has first revealed the problem of $U(m)$, however the given counterexample in [23] is a little problem in the calculation of $U(m)$ (specifically, \log_2 is misused). In this paper, a real counterexample of $U(m)$ on the monotonicity is given as below.

Given a FOD $\Omega = \{a, b, c\}$, m_1 and m_2 are two mass functions defined on Ω , in which

$$m_1(a) = 0.05, \quad m_1(ac) = 0.05, \quad m_1(bc) = 0.9;$$

$$m_2(ab) = 0.05, \quad m_2(ac) = 0.05, \quad m_2(bc) = 0.9.$$

Obviously, there are $\text{Bel}_{m_1}(A) \geq \text{Bel}_{m_2}(A)$ and $\text{Pl}_{m_1}(A) \leq \text{Pl}_{m_2}(A)$ for any $A \subseteq \Omega$. Therefore, it clearly should be $\text{Uncertainty}(m_1) \leq \text{Uncertainty}(m_2)$ in terms of the monotonicity. However, by means of the uncertainty measure $U(m)$ expressed in *nats*, it obtains $U(m_1) = 3.9549$ and $U(m_2) = 3.9153$. Namely, there is $U(m_1) > U(m_2)$. Therefore, the monotonicity is violated by $U(m)$.

Based on the above analysis, new cross entropy definition of mass functions is required and it is exactly the purpose of the study.

4.2 Cross and relative plausibility entropies

In this paper, new cross and relative entropies of mass functions are presented, which is on the basis of an uncertainty measure called plausibility entropy.

The plausibility entropy was recently proposed in [18], which is defined as

$$H_{Pl}(m) = - \sum_{\theta_i \in \Omega} Pl(\theta_i) \log_2 \frac{Pl(\theta_i)}{\sum_{\theta_j \in \Omega} Pl(\theta_j)} \quad (11)$$

where m is a mass function defined on FOD $\Omega = \{\theta_1, \theta_2, \dots, \theta_n\}$. Alternatively, the plausibility entropy $H_{Pl}(m)$ can also be expressed in the form of Shannon's entropy

$$H_{Pl}(m) = \sum_{\theta_i \in \Omega} Pl(\theta_i) \times H_S(Pl_P_m) \quad (12)$$

where Pl_P_m is the plausibility transformation [24] of m , satisfying $Pl_P_m(\theta_i) = \frac{Pl(\theta_i)}{\sum_{\theta_j \in \Omega} Pl(\theta_j)}$, $\theta_i \in \Omega$, and $H_S(Pl_P_m) = - \sum_{\theta_i \in \Omega} Pl_P_m(\theta_i) \log_2 Pl_P_m(\theta_i)$ is Shannon's entropy of probability distribution Pl_P_m on Ω .

It can be proved that the plausibility entropy $H_{Pl}(m)$ has satisfied four desiderata given in [19] for an effective measure of uncertainty (MoU) including "zero min value of MoU" (D1), "increasing of MoU of vacuous BPA" (D2), "compatibility with Shannon's entropy" (D3), and "unicity of max value of MoU" (D4). In addition, many other desirable properties are also satisfied by $H_{Pl}(m)$, please refer to [18, 25, 26] for more information. Based on the well-defined plausibility entropy, new cross and relative entropies of mass functions, named as cross plausibility entropy and relative plausibility entropy respectively, are given.

Cross plausibility entropy. Given two mass functions m_1 and m_2 on a same FOD $\Omega = \{\theta_1, \theta_2, \dots, \theta_n\}$, a cross plausibility entropy, denoted as $H_{Pl}(m_1, m_2)$, is defined as

$$H_{Pl}(m_1, m_2) = - \sum_{\theta_i \in \Omega} Pl_{m_1}(\theta_i) \log_2 \frac{Pl_{m_2}(\theta_i)}{\sum_{\theta_j \in \Omega} Pl_{m_2}(\theta_j)} \quad (13)$$

Moreover, the cross plausibility entropy $H_{Pl}(m_1, m_2)$ can also be represented as

$$H_{Pl}(m_1, m_2) = \sum_{\theta_i \in \Omega} Pl_{m_1}(\theta_i) \times H_S(Pl_P_{m_1}, Pl_P_{m_2}) \quad (14)$$

where $Pl_{-P_{m_1}}$ and $Pl_{-P_{m_2}}$ are the plausibility transformations of m_1 and m_2 respectively, and $H_S(Pl_{-P_{m_1}}, Pl_{-P_{m_2}}) = - \sum_{\theta_i \in \Omega} Pl_{-P_{m_1}}(\theta_i) \log_2 Pl_{-P_{m_2}}(\theta_i)$ is the classical cross entropy between probability distributions $Pl_{-P_{m_1}}$ and $Pl_{-P_{m_2}}$.

In terms of Eq. (14), a series of properties satisfied by the cross plausibility entropy $H_{Pl}(m_1, m_2)$ can be obtained easily.

Proposition 1 $H_{Pl}(m_1, m_2) \geq H_{Pl}(m_1)$, the equality holds if and only if $Pl_{-P_{m_1}} = Pl_{-P_{m_2}}$.

Proposition 2 $H_{Pl}(m_1, m_2) \neq H_{Pl}(m_2, m_1)$, i.e., the cross plausibility entropy is not symmetric in general.

Proposition 3 If m_1 and m_2 are Bayesian mass functions, the cross plausibility entropy coincides with the classical cross entropy for probabilities, i.e., $H_{Pl}(m_1, m_2) = - \sum_{\theta_i \in \Omega} m_1(\theta_i) \log_2 m_2(\theta_i)$.

Having the above definition of cross plausibility entropy, the relative entropy of mass functions can be defined immediately in a similar means. We have noted that reference [27] provided a KL divergence as a straightforward derivative of the plausibility entropy, it is exactly the expected form of relative entropy, which is introduced as follows.

Relative plausibility entropy. Let m_1 and m_2 be two mass functions defined on a FOD $\Omega = \{\theta_1, \theta_2, \dots, \theta_n\}$, a relative plausibility entropy, denoted as $H_{Pl}(m_1||m_2)$, is defined as follows

$$H_{Pl}(m_1||m_2) = \sum_{\theta_i \in \Omega} Pl_{m_1}(\theta_i) \log_2 \frac{Pl_{m_1}(\theta_i) / \sum_{\theta_j \in \Omega} Pl_{m_1}(\theta_j)}{Pl_{m_2}(\theta_i) / \sum_{\theta_j \in \Omega} Pl_{m_2}(\theta_j)} \quad (15)$$

Similarly, the relative plausibility entropy $H_{Pl}(m_1||m_2)$ can also be simply represented as

$$H_{Pl}(m_1||m_2) = \sum_{\theta_i \in \Omega} Pl_{m_1}(\theta_i) \times H_S(Pl_{-P_{m_1}}||Pl_{-P_{m_2}}) \quad (16)$$

where $H_S(Pl_{-P_{m_1}}||Pl_{-P_{m_2}})$ is the classical cross entropy between probability distributions $Pl_{-P_{m_1}}$ and $Pl_{-P_{m_2}}$, i.e., $H_S(Pl_{-P_{m_1}}||Pl_{-P_{m_2}}) = \sum_{\theta_i \in \Omega} Pl_{-P_{m_1}}(\theta_i) \log_2 \frac{Pl_{-P_{m_1}}(\theta_i)}{Pl_{-P_{m_2}}(\theta_i)}$.

From Eq. (16), some properties of the relative plausibility entropy $H_{Pl}(m_1||m_2)$ are derived as below.

Proposition 4 $H_{Pl}(m_1||m_2) \geq 0$, the equality holds if and only if $Pl_{-P_{m_1}} = Pl_{-P_{m_2}}$.

Proposition 5 $H_{Pl}(m_1||m_2) \neq H_{Pl}(m_2||m_1)$ in general, i.e., the cross plausibility entropy is not symmetric.

Proposition 6 For two Bayesian mass functions m_1 and m_2 , the cross plausibility entropy degenerates into classical relative entropy (or KL divergence) for probabilities, i.e., $H_{Pl}(m_1||m_2) = \sum_{\theta_i \in \Omega} m_1(\theta_i) \log_2 \frac{m_1(\theta_i)}{m_2(\theta_i)}$.

What's more, as same as the equality relation for probability distributions P_1 and P_2 in terms of Shannon's entropy, classical cross entropy and relative entropy, i.e., $H_S(P_1, P_2) = H_S(P_1||P_2) + H_S(P_1)$, the presented $H_{Pl}(m_1, m_2)$, $H_{Pl}(m_1||m_2)$, as well as plausibility entropy $H_{Pl}(m_1)$, of mass functions also meet the following equality relation

$$H_{Pl}(m_1, m_2) = H_{Pl}(m_1||m_2) + H_{Pl}(m_1) \quad (17)$$

5 An example of application

In this section, an illustrative example regarding parameter estimation is provided to show the potential application of proposed cross plausibility entropy $H_{Pl}(m_1, m_2)$. The example is originally from literature [28].

Assuming there are n patients randomly taken from a population which has a proportion θ to suffer from a disease, and each of them is represented by X_i to show if he/she has the disease (i.e., $X_i = 1$) or not (i.e., $X_i = 0$). Then, these random samples $\mathbf{X} = (X_1, X_2, \dots, X_n)$, which are independent and identically distributed (iid), can be viewed as the outcome of a Bernoulli variable. For realizations $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega_{\mathbf{X}} = \{0, 1\}^n$, the probability can be calculated by

$$p_{\mathbf{X}}(\mathbf{x}; \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \quad (18)$$

The task is to estimate the unknown parameter θ according to state descriptions $\mathbf{x} = (x_1, x_2, \dots, x_n)$. However, due to the uncertainty, these states are only partially known, and let m_i be the mass function concerning the state x_i associated with patient i . Table 1 gives a data set composed of $n = 6$ observations, in which the fourth one m_4 is uncertain and represented by $m_4(\{1\}) = \alpha$, $m_4(\{0\}) = \beta$, and $m_4(\{1, 0\}) = 1 - \alpha - \beta$, where $\alpha, \beta, 1 - \alpha - \beta \in [0, 1]$.

Literature [28] proposed an evidential expectation-maximization (E²M) algorithm to

Table 1. Data set for the example of parameter estimation.

Observation i	1	2	3	4	5	6
$m_i(\{1\})$	0.0	0.0	0.0	α	1.0	1.0
$m_i(\{0\})$	1.0	1.0	1.0	β	0.0	0.0
$m_i(\{1,0\})$	0.0	0.0	0.0	$1 - \alpha - \beta$	0.0	0.0

estimate the parameter θ , in terms of a maximum likelihood principle. Figure 1 gives the results of using E²M algorithm with respect to different α and β . It is found that, by using the E²M algorithm, the result of estimated θ is unchanged if the plausibility transformations of different m_4 caused by changed α and β are the same. For example, let m_4^1 and m_4^2 be $m_4^1(\{1\}) = 0.6$, $m_4^1(\{0\}) = 0.4$, $m_4^1(\{1,0\}) = 0$, and $m_4^2(\{1\}) = 0.4$, $m_4^2(\{0\}) = 0.1$, $m_4^2(\{1,0\}) = 0.5$, respectively. There are $Pl_{P_{m_4^1}}(\{1\}) = Pl_{P_{m_4^2}}(\{1\}) = 0.6$ and $Pl_{P_{m_4^1}}(\{0\}) = Pl_{P_{m_4^2}}(\{0\}) = 0.4$, i.e., $Pl_{P_{m_4^1}} = Pl_{P_{m_4^2}}$. Then, by using the E²M algorithm, parameter θ is estimated as $\theta^{1,*} = 0.4201$ associated with m_4^1 and $\theta^{2,*} = 0.4201$ associated with m_4^2 . Two observations with different uncertainty degrees, m_4^1 and m_4^2 , lead to the same estimation of θ .

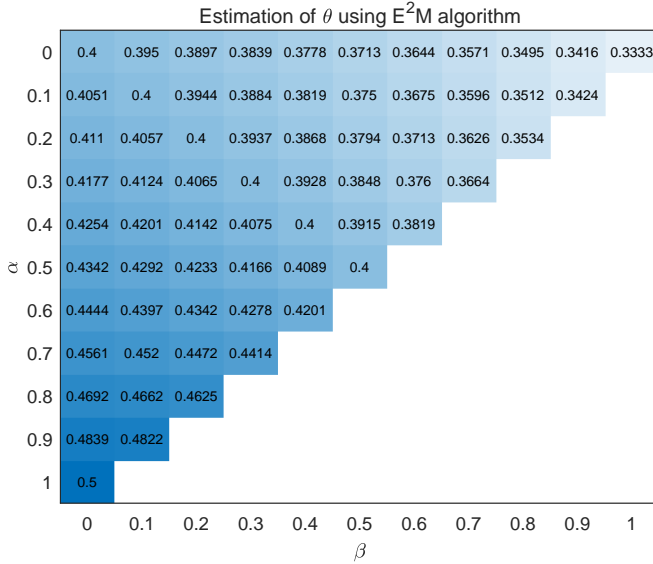


Figure 1. Estimation of θ with the use of E²M algorithm by considering different values of α and β .

Now, let us use a cross entropy-based method to solve the issue of estimating parameter θ , where θ is derived by minimizing a total cross entropy loss, which coincides with the maximum likelihood principle widely used in machine learning.

For the data set shown in Table 1, since there is uncertain observation m_4 involving epistemic

uncertainty, the proposed cross plausibility entropy is used to calculate the total cross entropy loss L_{Pl} . Let P be a distribution relying on parameter θ with $P(1) = \theta$ and $P(0) = 1 - \theta$, then

$$\begin{aligned}
 L_{Pl} &= \sum_{i=1}^6 H_{Pl}(m_i, P) \\
 &= -3\log_2(1 - \theta) - 2\log_2\theta \\
 &\quad - (1 - \beta)\log_2\theta - (1 - \alpha)\log_2(1 - \theta) \\
 &= -(4 - \alpha)\log_2(1 - \theta) - (3 - \beta)\log_2\theta
 \end{aligned}$$

By letting $\frac{\partial L_{Pl}}{\partial \theta} = 0$, we have

$$\theta_{Pl}^* = \frac{3 - \beta}{7 - \alpha - \beta}$$

Figure 2 shows the results of using the proposed cross plausibility entropy with the consideration of different α and β . Compared with the results of E²M algorithm, the estimated θ is changing with m_4 having different uncertainty degrees. For example, for the cases of $m_4 = m_4^1$ and $m_4 = m_4^2$, it is obtained that $\theta_{Pl}^{1,*} = 0.4333$ and $\theta_{Pl}^{2,*} = 0.4462$.

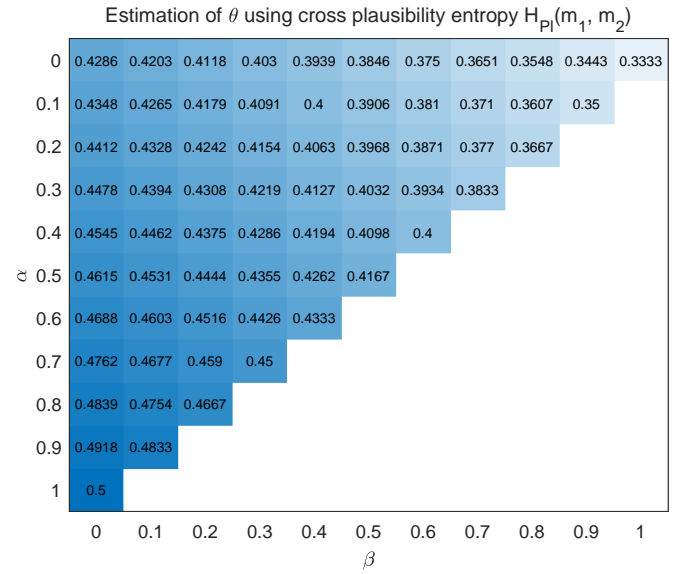


Figure 2. Estimation of θ with the use of proposed cross plausibility entropy by considering different values of α and β .

For the comparison, Dezert and Dambreville's cross entropy $U(m_1, m_2)$ is also used in the example to obtain the estimation of parameter θ . Let m_θ be the estimation of θ , in which $m_\theta(\{1\}) = \theta$, $m_\theta(\{0\}) = 1 - \theta - \varepsilon$, and $m_\theta(\{1,0\}) = \varepsilon$, where $\varepsilon \rightarrow 0$. Then, in terms of Dezert and Dambreville's cross entropy, the

total loss is

$$\begin{aligned}
 L_U &= \sum_{i=1}^6 U(m_i, m_\theta) \\
 &= -3 \ln(1 - \theta - \varepsilon) - 2 \ln \theta \\
 &\quad - \alpha(\alpha + \beta) \ln \theta + (1 - \alpha - \beta)(1 - \theta) \\
 &\quad - \beta(\alpha + \beta) \ln(1 - \theta - \varepsilon) - (1 - \alpha - \beta) \ln \varepsilon \\
 &\quad + (1 - \alpha - \beta)[1 - (1 - \theta - \varepsilon)]
 \end{aligned}$$

By means of $\frac{\partial L_U}{\partial \theta} = 0$, it obtains

$$\theta_U^* = \frac{2 + \alpha(\alpha + \beta)}{5 + (\alpha + \beta)^2} (1 - \varepsilon)$$

where $\varepsilon \rightarrow 0$. Figure 3 graphically shows the results obtained by using Dezert and Dambreville's cross entropy $U(m_1, m_2)$.

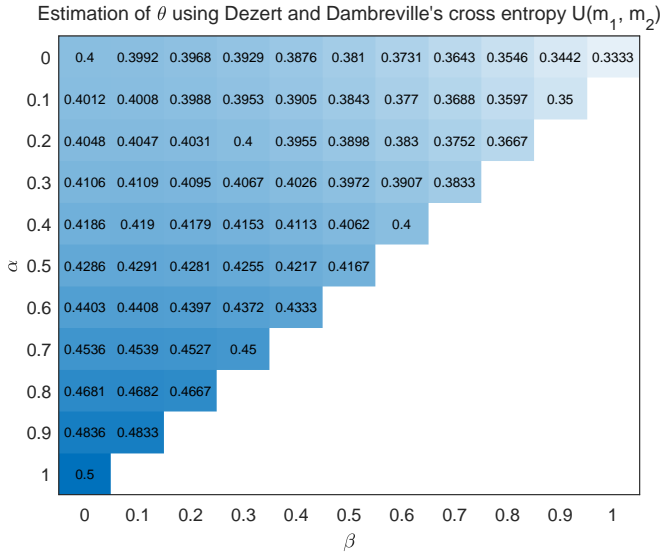


Figure 3. Estimation of θ with the use of Dezert and Dambreville's cross entropy by considering different values of α and β .

And the cross entropy $H_D(m_1, m_2)$ from Gao et al. [21] is also used in the example. Similarly, a total loss L_D is calculated as follows

$$\begin{aligned}
 L_D &= \sum_{i=1}^6 H_D(m_i, m_\theta) \\
 &= -3 \log_2(1 - \theta - \varepsilon) - 2 \log_2 \theta \\
 &\quad - \alpha \log_2 \theta - \beta \log_2(1 - \theta - \varepsilon) - (1 - \alpha - \beta) \log_2 \varepsilon
 \end{aligned}$$

Then, the estimation of θ is derived via $\frac{\partial L_D}{\partial \theta} = 0$ as below

$$\theta_D^* = \frac{2 + \alpha}{5 + \alpha + \beta} (1 - \varepsilon)$$

in which $\varepsilon \rightarrow 0$. The results with the use of cross entropy $H_D(m_1, m_2)$ are shown in Figure 4.

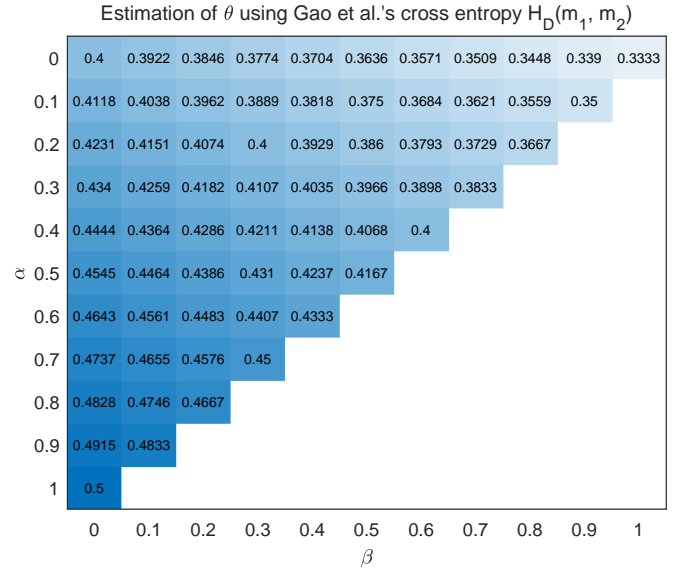


Figure 4. Estimation of θ with the use of cross entropy by considering different values of α and β .

In addition, cross entropies inspired by two widely used entropies of mass functions, pignistic entropy [4, 29] and Yager entropy [30], are also considered for comparison. In terms of the formula of pignistic entropy $H_{\text{pignistic}}(m) = - \sum_{\theta_i \in \Omega} \text{BetP}_m(\theta_i) \log_2 \text{BetP}_m(\theta_i)$, where $\text{BetP}_m(\theta_i) = \sum_{A \in \mathcal{A}} \frac{m(A)}{|A|}$, cross pignistic entropy is naturally defined as $H_{\text{pignistic}}(m_1, m_2) = - \sum_{\theta_i \in \Omega} \text{BetP}_{m_1}(\theta_i) \log_2 \text{BetP}_{m_2}(\theta_i)$. Then, a total cross entropy loss can be obtained by

$$\begin{aligned}
 L_{\text{pignistic}} &= \sum_{i=1}^6 H_{\text{pignistic}}(m_i, P) \\
 &= -3 \log_2(1 - \theta) - 2 \log_2 \theta \\
 &\quad - \frac{1 + \alpha - \beta}{2} \log_2 \theta - \frac{1 - \alpha + \beta}{2} \log_2(1 - \theta) \\
 &= -\frac{5 + \alpha - \beta}{2} \log_2 \theta - \frac{7 - \alpha + \beta}{2} \log_2(1 - \theta)
 \end{aligned}$$

where $P(1) = \theta$ and $P(0) = 1 - \theta$. By letting $\frac{\partial L_{\text{pignistic}}}{\partial \theta} = 0$, we have

$$\theta_{\text{pignistic}}^* = \frac{5 + \alpha - \beta}{12}$$

whose graphical results with different values of α and β are given in Figure 5.

Similarly, according to the expression of Yager entropy $H_{\text{Yager}}(m) = - \sum_{A \in \Omega} m(A) \log_2 Pl_m(A)$, cross Yager entropy is defined as $H_{\text{Yager}}(m_1, m_2) = - \sum_{A \in \Omega} m_1(A) \log_2 Pl_{m_2}(A)$. Then, the corresponding

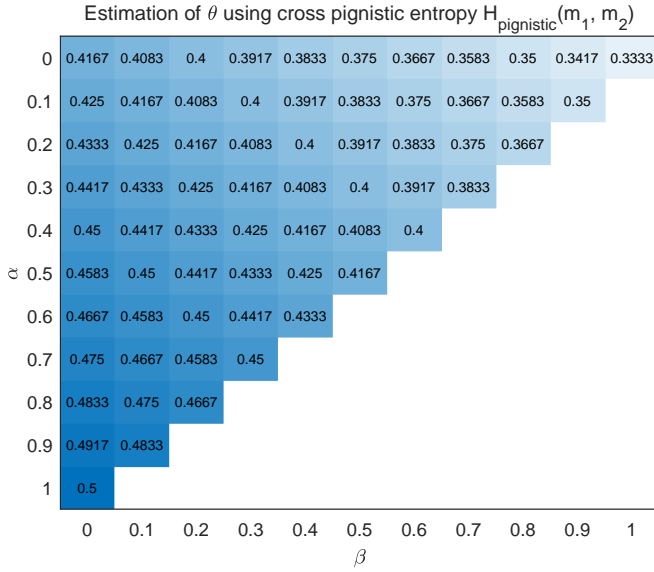


Figure 5. Estimation of θ with the use of cross pignistic entropy by considering different values of α and β .

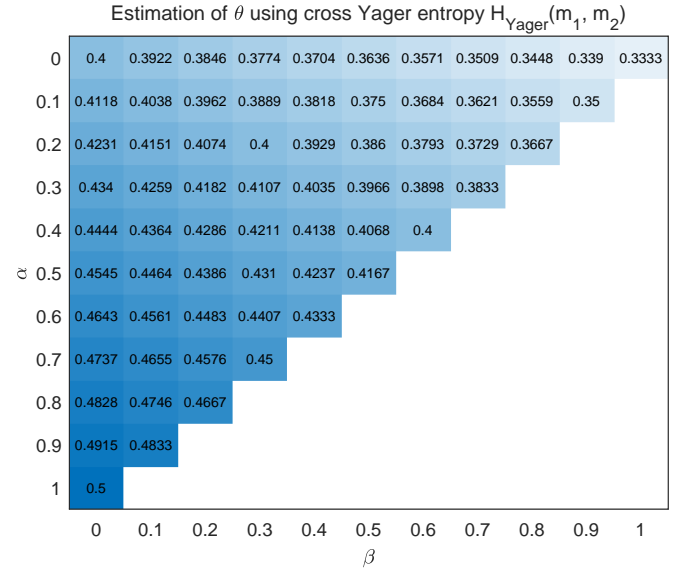


Figure 6. Estimation of θ with the use of cross Yager entropy by considering different values of α and β .

total cross entropy loss is

$$\begin{aligned}
 L_{\text{Yager}} &= \sum_{i=1}^6 H_{\text{Yager}}(m_i, P) \\
 &= -3\log_2(1 - \theta) - 2\log_2\theta \\
 &\quad - \alpha\log_2\theta - \beta\log_2(1 - \theta) - (1 - \alpha - \beta)\log_2 1 \\
 &= -(2 + \alpha)\log_2\theta - (3 + \beta)\log_2(1 - \theta)
 \end{aligned}$$

By calculating $\frac{\partial L_{\text{Yager}}}{\partial \theta} = 0$, it obtains

$$\theta_{\text{Yager}}^* = \frac{2 + \alpha}{5 + \alpha + \beta}$$

which is same with θ_D^* that is obtained by using cross entropy $H_D(m_1, m_2)$. Figure 6 shows the results of using cross Yager entropy by considering different α and β .

For the sake of further comparison of these methods, by letting $\alpha = \beta$, the uncertain observation m_4 becomes $m_4(\{1\}) = m_4(\{0\}) = \alpha$ and $m_4(\{1, 0\}) = 1 - 2\alpha$, where $\alpha \in [0, 0.5]$. If $\alpha = 0$, m_4 has the maximum uncertainty, and the uncertainty of m_4 is the least while $\alpha = 0.5$. And it is noted that there is not any preference in m_4 between states $\{1\}$ and $\{0\}$ because $Bel_{m_4}(\{1\}) = Bel_{m_4}(\{0\}) = \alpha$ and $Pl_{m_4}(\{1\}) = Pl_{m_4}(\{0\}) = 1 - \alpha$. Figure 7 shows the estimation results of θ by using different methods for the case of new m_4 in which $\alpha = \beta$.

In theory, if we only have observations 1,2,3,5,6, in terms of the maximum likelihood principle, the estimation of θ should be $\theta^* = 2/5 = 0.4$. With the consideration of observation 4, i.e., unbiased m_4 : (i) When $\alpha = 0$, the estimation of θ should lie in the

interval $[2/6, 3/6]$ (whose midpoint is $2.5/6$) since the existence of epistemic uncertainty in observation 4 with $m_4(\{1, 0\}) = 1$; (ii) When $\alpha = 0.5$, the estimation of θ should be $2.5/6$ since in this case the example becomes a mixture model for a two-dimensional Bernoulli distribution with $m_4(\{1\}) = m_4(\{0\}) = 0.5$ in which there is only random uncertainty; (iii) In the process of increasing α 's value from 0 to 0.5, the value of θ should be changing monotonically, because for the unbiased m_4 the only change is its uncertainty degree which is decreasing monotonically; (iv) Therefore, there is a path for the estimation of θ from start point $\theta|_{\alpha=0} \in [2/6, 3/6]$ to end point $\theta|_{\alpha=0.5} = 2.5/6$.

From Figure 7, the result of E²M algorithm is not reasonable since the estimated θ is 0.4 when $\alpha = 0.5$. Cross pignistic entropy $H_{\text{pignistic}}(m_1, m_2)$ produces insensitive result for the change of observation m_4 . The proposed cross plausibility entropy $H_{\text{Pl}}(m_1, m_2)$ gives that the value of θ is declining monotonically with the rise of α , while Dezert and Dambreville's cross entropy $U(m_1, m_2)$, cross entropy $H_D(m_1, m_2)$, and cross Yager entropy $H_{\text{Yager}}(m_1, m_2)$ present opposite trend of change. The difference between result of $H_{\text{Pl}}(m_1, m_2)$ and those of $U(m_1, m_2)$, $H_D(m_1, m_2)$, and $H_{\text{Yager}}(m_1, m_2)$, is caused by the underlying logic of different entropy measures. The plausibility entropy is based on plausibility function and it tends to get the maximum uncertainty degree that a mass function could possibly have, therefore the cross plausibility entropy $H_{\text{Pl}}(m_1, m_2)$ gives a relatively big estimation of θ . In contrast, it seems that $U(m_1, m_2)$, $H_D(m_1, m_2)$ and $H_{\text{Yager}}(m_1, m_2)$ are to obtain relatively small

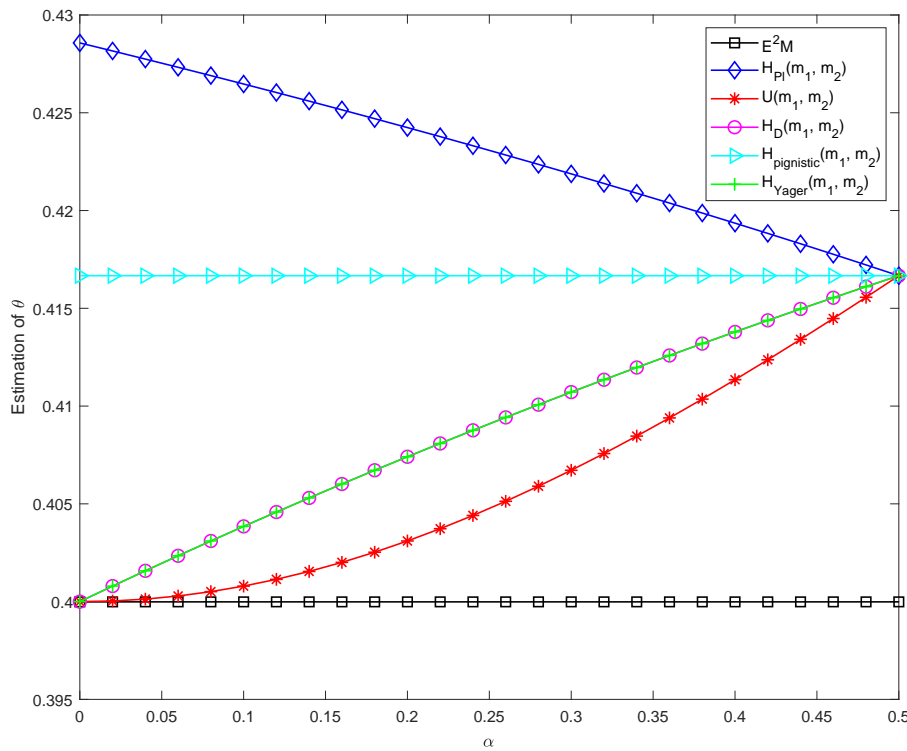


Figure 7. Estimation results of θ generated by using different methods where observation 4 is set as $m_4(\{1\}) = m_4(\{0\}) = \alpha$, $m_4(\{1, 0\}) = 1 - 2\alpha$.

Table 2. Generalized Bernoulli distribution with plausibility distribution observations.

Observation i	1	2	3	4	5	6
$Pl_i(\{1\})$	0.0	0.0	0.0	$1 - \beta$	1.0	1.0
$Pl_i(\{0\})$	1.0	1.0	1.0	$1 - \alpha$	0.0	0.0

estimation of θ .

Compared with $U(m_1, m_2)$, $H_D(m_1, m_2)$ and $H_{Yager}(m_1, m_2)$, the proposed $H_{Pl}(m_1, m_2)$ is more recommended in theory because, at first, the underlying entropy definitions of the formers are of some defects as analyzed in Section 4.1 and related references [5, 19], and the underlying mechanism of using cross plausibility entropy $H_{Pl}(m_1, m_2)$ to obtain the estimation of θ is more clear. In this example, with the use of cross plausibility entropy $H_{Pl}(m_1, m_2)$, the classical Bernoulli distribution based on probabilities is generalized to a new Bernoulli distribution with plausibility distribution observations as shown in Table 2. According to Table 2, the estimation of parameter θ can be obtained immediately as $\theta_{Pl}^* = \frac{3-\beta}{7-\alpha-\beta}$. Moreover, this Bernoulli distribution with plausibility distribution observations can be easily extended to the case of multi-dimensional Bernoulli distribution.

6 Conclusion

In this paper, new definitions of cross entropy and relative entropy of mass functions have been given on the basis of a recently presented total uncertainty measure of mass functions called plausibility entropy. And, properties of the cross plausibility entropy and relative plausibility entropy have been presented in the study. In addition, an illustrative example of application has been provided to show the effectiveness of the presented entropies compared with other methods and entropy definitions of mass functions. The presented cross plausibility entropy and relative plausibility entropy of mass functions can be used in the scenarios of multi-source information fusion based on Dempster-Shafer evidence theory for target recognition, fault diagnosis, and so on.

In the future study, on one hand, more theoretical analysis about the presented cross and relative

plausibility entropies will be conducted; on the other hand, practical applications with the use of cross and relative plausibility entropies for multi-sensor information fusion, decision support systems, intelligent diagnosis, and so forth, will be further considered.

Data Availability Statement

Data will be made available on request.

Funding

The work was partially supported by the National Natural Science Foundation of China under Grant 62173272.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

- [1] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. [CrossRef]
- [2] Dempster, A. P. (2008). Upper and lower probabilities induced by a multivalued mapping. In *Classic works of the Dempster-Shafer theory of belief functions* (pp. 57–72). Berlin, Heidelberg: Springer Berlin Heidelberg. [CrossRef]
- [3] Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.
- [4] Jousselme, A. L., Liu, C., Grenier, D., & Bosse, E. (2006). Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36(5), 890–903. [CrossRef]
- [5] Abellan, J., & Bosse, E. (2018). Drawbacks of Uncertainty Measures Based on the Pignistic Transformation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(3), 382–388. [CrossRef]
- [6] Deng, Y. (2020). Uncertainty measure in evidence theory. *Science China Information Sciences*, 63(11), 210201. [CrossRef]
- [7] Abellan, J., & Bosse, E. (2017). Critique of recent uncertainty measures developed under the evidence theory and belief intervals. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(3), 1186–1192. [CrossRef]
- [8] Klir, G. J. (2006). Uncertainty and information: foundations of generalized information theory. *Kybernetes*, 35(7/8), 1297–1299. [CrossRef]
- [9] Abellan, J., & Masegosa, A. (2008). Requirements for total uncertainty measures in Dempster-Shafer theory of evidence. *International Journal of General Systems*, 37(6), 733–747. [CrossRef]
- [10] Jirousek, R., & Shenoy, P. P. (2018). A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92, 49–65. [CrossRef]
- [11] Deng, Y. (2016). Deng entropy. *Chaos, Solitons & Fractals*, 91, 549–553. [CrossRef]
- [12] Urbani, M., Gasparini, G., & Brunelli, M. (2023). A numerical comparative study of uncertainty measures in the Dempster-Shafer evidence theory. *Information Sciences*, 639, 119027. [CrossRef]
- [13] Zhou, M., Zhu, S. S., Chen, Y. W., Wu, J., & Herrera-Viedma, E. (2021). A generalized belief entropy with nonspecificity and structural conflict. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(9), 5532–5545. [CrossRef]
- [14] Kavya, R., Jabez, C., & Subhrakanta, P. (2023). A new belief interval-based total uncertainty measure for Dempster-Shafer theory. *Information Sciences*, 642, 119150. [CrossRef]
- [15] Deng, Z., & Wang, J. (2021). Measuring total uncertainty in evidence theory. *International Journal of Intelligent Systems*, 36(4), 1721–1745. [CrossRef]
- [16] Zhou, Q., & Deng, Y. (2022). Fractal-based belief entropy. *Information Sciences*, 587, 265–282. [CrossRef]
- [17] Moral-Garcia, S., & Abellan, J. (2021). Required mathematical properties and behaviors of uncertainty measures on belief intervals. *International Journal of Intelligent Systems*, 36(8). [CrossRef]
- [18] Cui, Y., & Deng, X. (2023). Plausibility entropy: a new total uncertainty measure in evidence theory based on plausibility function. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(6), 3833–3844. [CrossRef]
- [19] Dezert, J., & Tchamova, A. (2022). On the effectiveness of measures of uncertainty of basic belief assignments. *Information & Security*, 52, 9–36. [CrossRef]
- [20] Dezert, J. (2022, July). An effective measure of uncertainty of basic belief assignments. In *2022 25th International Conference on Information Fusion (FUSION)* (pp. 1–8). IEEE. [CrossRef]
- [21] Dezert, J., & Dambreville, F. (2023, June). Cross-entropy and relative entropy of basic belief assignments. In *2023 26th International Conference on Information Fusion (FUSION)* (pp. 1–8). IEEE. [CrossRef]
- [22] Gao, X., Pan, L., & Deng, Y. (2022). Cross entropy of mass function and its application in similarity measure. *Applied Intelligence*, 52(8), 8337–8350.

- [CrossRef]
- [23] Zhou, Q., Pedrycz, W., Liang, Y., & Deng, Y. (2023). Information Granule-based Uncertainty Measure of Fuzzy Evidential Distribution. *IEEE Transactions on Fuzzy Systems*, 31(12), 4385–4396. [CrossRef]
- [24] Cobb, B. R., & Shenoy, P. P. (2006). On the plausibility transformation method for translating belief function models to probability models. *International journal of approximate reasoning*, 41(3), 314–330. [CrossRef]
- [25] Deng, X., Xue, S., Jiang, W., & Zhang, X. (2024). Plausibility extropy: the complementary dual of plausibility entropy. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 54(11), 6936–6947. [CrossRef]
- [26] Deng, X., & Jiang, W. (2025). Upper bounds of uncertainty for Dempster combination rule-based evidence fusion systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 55(1), 817–828. [CrossRef]
- [27] Mao, K., Wang, Y., Zhou, W., Ye, J., & Fang, B. (2025). Evaluation of belief entropies: from the perspective of evidential neural network. *Artificial Intelligence Review*, 58(5), 133. [CrossRef]
- [28] Denoeux, T. (2013). Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on Knowledge and Data Engineering*, 25(1), 119–130. [CrossRef]
- [29] Xiao, F. (2019). Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Information Fusion*, 46, 23–32. [CrossRef]
- [30] Yager, R. R. (1983). Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9, 249–260. [CrossRef]
- [31] Abellan, J. (2017). Analyzing properties of Deng entropy in the theory of evidence. *Chaos, Solitons & Fractals*, 95, 195–199. [CrossRef]
- [32] Moral-Garcia, S., & Abellan, J. (2020). Critique of modified Deng entropies under the evidence theory. *Chaos, Solitons & Fractals*, 140, 110112. [CrossRef]
- [33] Hullermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110(3), 457–506. [CrossRef]



processing. (Email: xinyang.deng@nwpu.edu.cn)



with uncertainty. (Email: jiangwen@nwpu.edu.cn)

Xinyang Deng received the bachelor, master, and Ph.D. degrees from Southwest University, Chongqing, China, in 2010, 2013, and 2016, respectively. Since 2017, he has been an associate professor with the School of Electronics and Information, Northwestern Polytechnical University. His main research interests include multi-source information fusion, Dempster-Shafer evidence theory, and uncertain information modeling and processing. (Email: xinyang.deng@nwpu.edu.cn)

Wen Jiang received the bachelor and master degrees from Information Engineering University, Zhengzhou, China, in 1994 and 1997, and received her Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2009. She is currently a professor with the School of Electronics and Information, Northwestern Polytechnical University. Her main research interests include multi-source information fusion and artificial intelligence with uncertainty. (Email: jiangwen@nwpu.edu.cn)