The information volume of uncertain informaion: (1) Mass function

Yong Deng[a,b,c,*]

[a] *Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China*
[b] *School of Eduction, Shannxi Normal University, Xi'an, 710062, China*
[c] *Department of Biostatics, Medical Center, Vanderbilt University, Nashville, TN. 37235, USA*

Abstract

Given a probability distribution, its corresponding information volume is Shannon entropy. However, how to determine the information volume of a given mass function is still an open issue. Based on Deng entropy, the information volume of mass function is presented in this paper. Given a mass function, the corresponding information volume is larger than its uncertainty measured by Deng entropy. The so called Deng distribution is defined as the BPA condition of the maximum Deng entropy. The information volume of Deng distribution is called the maximum information volume, which is lager than the maximum Deng entropy. In addition, both the total uncertainty case and the Deng distribution have the same information volume, namely, the maximum information volume. Some numerical examples are illustrated to show the efficiency of the proposed information volume of mass function.

Keywords: information volume, mass function, Shannon entropy, Deng entropy, Deng distribution.

1. Introduction

In the past decades, plenty of theories have been developed for expressing and dealing with the uncertainty in the uncertain environment, for instance,

*Corresponding author
Email address: `dengentropy@uestc.edu.cn`; `prof.deng@hotmail.com` (Yong Deng)

probability theory [1], fuzzy set theory [2], Dempster-Shafer evidence theory [3, 4], rough sets [5], and D numbers [6].

Entropy function is very important in uncertainty modelling. Since firstly derived from thermodynamics, different kinds of entropy have been proposed, such as Shannon entropy [7], Tsallis entropy [8], nonadditive entropy [9]. Recently, a new entropy, called Deng entropy [10], is presented for measuring the uncertainty in evidence theory. Deng entropy is the generalization of Shannon entropy. Compared with traditional methods, Deng entropy is more reasonable, and it takes both discord and non-specificity into account.

Given a probability distribution, its corresponding information volume can be measured by Shannon entropy. However, how to determine the information volume of mass function in evidence theory is still an open issue. In this paper, an information volume of mass function based on Deng entropy is presented. The information volume of mass function is constructed with a new distribution, named as Deng distribution. If the mass function is degenerated into probability distribution, the proposed information volume is the same as Shannnon entropy.

The rest of this paper is organized as follows. In section **2**, some preliminaries are briefly reviewed. In section **3**, based on Deng entropy, the information volume of mass function is proposed. In section **4**, numerical examples are expounded to illustrated the proposed method and definition. In section **5**, we have a brief conclusion.

2. Preliminaries

Several preliminaries are briefly introduced in this section, including mass function, Deng entropy, the maximum Deng entropy, fuzzy sets and intuitionistic fuzzy sets.

2.1. Dempster-Shafer evidence theory

Dempster-Shafer evidence theory[3, 4] can be used to deal with uncertainty. Besides, evidence theory satisfies the weaker conditions than the probability theory, which provides it with the ability to express uncertain information directly. Some basic conceptions of evidence theory are given as follows:

Definition 2.1: *Frame of discernment and its power set*

Let Θ*, called the frame of discernment, denote an exhaustive nonempty set of hypotheses, where the elements are mutually exclusive. Let the set* Θ *have* N *elements, which can be expressed as:*

$$\Theta = \{\theta_1, \theta_2, \theta_3, \cdots, \theta_N\} \tag{1}$$

The power set of Θ*, denoted as* 2^Θ*, contains all possible subsets of* Θ *and has* 2^N *elements, and* 2^Θ *is represented by*

$$\begin{aligned} 2^\Theta &= \{A_1, A_2, A_3, \cdots, A_{2^N}\} \\ &= \{\ \emptyset, \{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}, \{\theta_1, \theta_2\}, \\ &\qquad \{\theta_1, \theta_3\}, \cdots, \{\theta_1, \theta_N\}, \cdots, \Theta\ \} \end{aligned} \tag{2}$$

where the element A_k *is called the focal element of* Θ*, if* A_k *is nonempty.*

Definition 2.2: *Mass function*

A mass function is also called Basic probability assignment (BPA), which map m *from* 2^Θ *to* $[0, 1]$*, and it is defined as follows:*

$$m : 2^\Theta \rightarrow [0, 1] \tag{3}$$

which is constrained by the following conditions:

$$\sum_{A \in 2^\Theta} m(A) = 1 \tag{4}$$

$$m(\emptyset) = 0 \tag{5}$$

2.2. Shannon entropy

In the field of classical probability theory, Shannon entropy [7] is often used to measure the uncertainty of a probability distribution. Consider a probability distribution P defined on the set $\Theta = \{H_1, H_2, H_3, \cdots, H_N\}$.

Definition 2.5: *Shannon entropy*

Shannon entropy $H_s(P)$ is defined as follows:

$$H_s(P) = \sum_{\theta \in \Theta} P(\theta) \log(\frac{1}{P(\theta)}). \tag{6}$$

where $\sum_{\theta \in \Theta} P(\theta) = 1$ and $P(\theta) \in [0, 1]$.

Usually, the base of logarithm is 2, and entropy has the unit of bit. It's not hard to find that $H_s(P)$ is on the scale $[0, \log N]$.

2.3. Deng entropy

In information theory, entropy can be used to measure the uncertainty of a system. Recently, a novel entropy, named as Deng entropy [10], is proposed to measure the uncertainty in evidence theory.

Definition 2.5: *Deng entropy*

Deng entropy is defined as:

$$H_{DE}(m) = -\sum_{A \in 2^\Theta} m(A) \log(\frac{m(A)}{2^{|A|} - 1}) \tag{7}$$

where $|A|$ is the cardinal of a certain focal element A.

Deng entropy is the generalization of Shannon entropy. When every focal element is singleton, Deng entropy degenerates into Shannon entropy.

Through a simple transformation, Eq.(7) can be rewritten as follows:

$$H_{DE}(m) = \sum_{A \in 2^\Theta} \log(2^{|A|} - 1) - \sum_{A \in 2^\Theta} m(A) \log m(A) \tag{8}$$

where $\sum_{A \in 2^\Theta} \log(2^{|A|} - 1)$ and $-\sum_{A \in 2^\Theta} m(A) \log m(A)$ are measurements of nonspecificity and discord, respectively. As a result, Deng entropy is a composite measurement of nonspecificity and discord, which means that it is a tool for measuring total uncertainty.

2.4. The maximum Deng entropy

Assume A is the focal element of a certain frame of discernment Θ and $m(A)$ is the BPA for A. According to [11], the analytic solution of the maximum Deng entropy and the conditions of BPA distribution is as follows:

Theorem 2.1: *The analytic solution of the maximum Deng Entropy and its BPA distribution*

If and only if $m(A) = \frac{(2^{|A|}-1)}{\sum_{A\in 2^\Theta}(2^{|A|}-1)}$, *Deng entropy reaches its maximum value, and the analytic solution of the maximum Deng entropy is*

$$H_{MDE}(m) = \log \sum_{A\in 2^\Theta} (2^{|A|} - 1) \tag{9}$$

3. Information volume of mass function

Given a probability distribution, the associated information volume can be measured by Shannon entropy. However, how to measure the information volume of a given mass function is still an open issue.

In this section, firstly, we define Deng distribution as the BPA condition of the maximum Deng entropy. Then, based on Deng entropy, the information volume of mass function is proposed.

3.1. The definition of Deng distribution

The maximum Deng entropy and the BPA condition of it have been analyzed in [11]. However, the terminology, the BPA condition of the maximum Deng entropy, is not convenient for discussing. As a result, we define Deng distribution as follows:

Definition 3.1: *Deng distribution*

Deng distribution is defined as

$$m_D(A) = \frac{(2^{|A|} - 1)}{\sum_{A \in 2^\Theta} (2^{|A|} - 1)} \tag{10}$$

which is the BPA distribution of the maximum Deng entropy. Namely, if and only if under this conditions, Deng entropy can reach its maximum value.

3.2. The definition of the information volume of mass function

Definition 3.2: *The definition of the information volume of mass function*

Let the frame of discernment be $\Theta = \{\theta_1, \theta_2, \theta_3, \cdots, \theta_N\}$. Use index i to denote the times of this loop, and use $m(A_i)$ to denote different mass function of different loops. Based on Deng entropy, the information volume of mass function can be calculated by following steps:

step 1: Input mass function $m(A_0)$.

step 2: Continuously separate the mass function of the element whose cardinal is larger than 1 until convergence. Concretely, repeat the loop from step 2-1 to step 2-3 until Deng entropy is convergent.

step 2-1: Focus on the element whose cardinal is larger than 1, namely, $|A_i| > 1$. And then, separate its mass function based on the proportion of Deng distribution:

$$m_D(A_i) = \frac{(2^{|A_i|} - 1)}{\sum_{A_i \in 2^\Theta} (2^{|A_i|} - 1)} \tag{11}$$

For example, given a focal element $A_{i-1} = \{\theta_x, \theta_y\}$ and its mass function $m(A_{i-1})$, the separating proportion is that $\frac{1}{5} : \frac{1}{5} : \frac{3}{5}$. The ith times of separation divides $m(A_{i-1})$ and yields following new mass function: $m(X_i)$, $m(Y_i)$, $m(Z_i)$, where $X_i = \{\theta_x\}$, $Y_i = \{\theta_y\}$ and $Z_i = \{\theta_x, \theta_y\}$. In addition, they satisfy these equations:

$$m(X_i) + m(Y_i) + m(Z_i) = m(A_{i-1}) \tag{12}$$

$$m(X_i) : m(Y_i) : m(Z_i) = \frac{1}{5} : \frac{1}{5} : \frac{3}{5} \tag{13}$$

step 2-2: Based on Deng entropy, calculate the uncertainty of all the mass functions except for those who have been divided. The result is denoted as $H_i(m)$.

step 2-3: Calculate $\Delta_i = H_i(m) - H_{i-1}(m)$. When Δ_i satisfies following condition, jump out of this loop.

$$\Delta_i = H_i(m) - H_{i-1}(m) < \varepsilon \tag{14}$$

where ε is the allowable error.

step 3: Output $H_{IV-mass}(m) = H_i(m)$, which is the information volume of the mass function.

3.3. The maximum information volume of mass function

Theorem 3.1: *The maximum information volume of mass function*

If and only if the mass function is Deng distribution $m_D(A)$, the information volume achieve its maximum value, which is called the maximum information volume of mass function $H_{MIV-mass}(m)$.

4. Numerical examples and discussions

In this section, some examples are expounded to better understand the definition for the proposed information volume of mass function, and the discussion is followed after every example. In the following examples, the base of the logarithmic function is 2, and the allowable error is 0.001.

Example 4.1:

Consider the frame of discernment be $U = \{\theta_1, \theta_2\}$, $X = \{\theta_1\}$ and $Y = \{\theta_2\}$ be singletons. Let the mass function be $m_0(X) = m_0(Y) = \frac{1}{5}$ and $m_0(U) = \frac{3}{5}$.

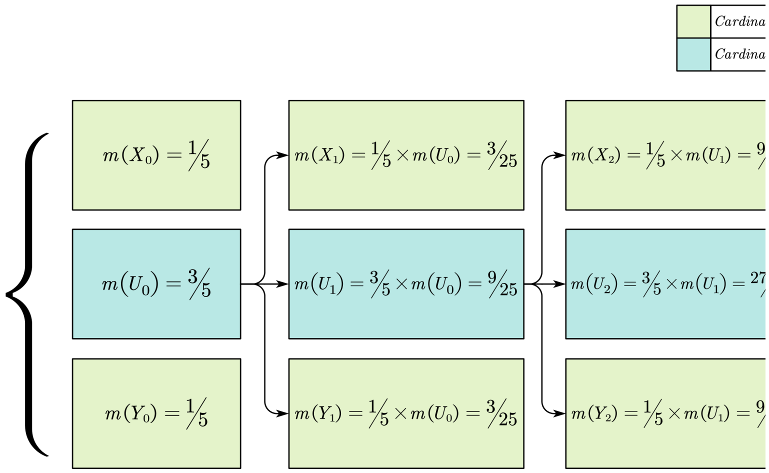

Figure 1: The procedure of from step 2-1 to step 2-3

The information volume of this mass function can be calculated by **Definition 3.2**, whose calculating procedure is illustrated in Figure 1. For the convenience of comprehension, the calculating procedure can be abstracted as a directed acyclic graphical model shown in Figure 2.

Then, the convergence procedure of $H_i(m)$ is listed in Table 1.

Table 1: The convergence procedure of $H_i(m)$

i	$H_i(m)$	i	$H_i(m)$
1	2.321928	8	3.396431
2	2.764107	9	3.408809
3	3.029415	10	3.416236
4	3.188600	11	3.420692
5	3.284110	12	3.423366
6	3.341417	13	3.424970
7	3.375801	14	3.425933

According to Table 1, when we continuously separate the BPA of the element whose cardinal is larger than 1, the Δ_i of Deng entropy becomes smaller and smaller. When $i = 14$, $H_i(m) - H_{i-1}(m) < 0.001$, which means that $H_i(m)$ finally converges to 3.425933. Hence the information volume of this mass function

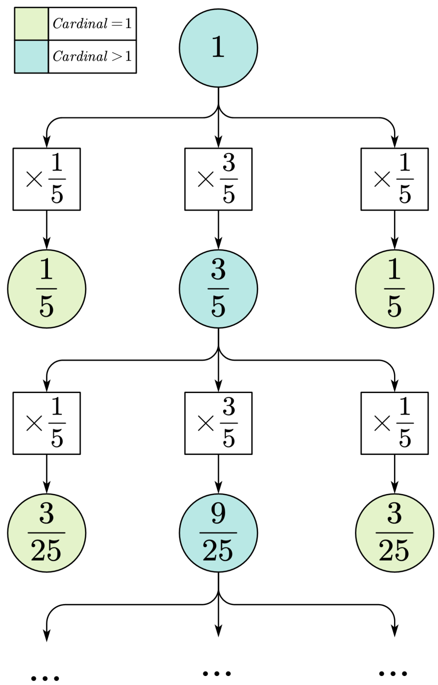

Figure 2: The directed acyclic graphical model

is $H_{IV-mass}(m) = 3.425933$.

Example 4.2:

Consider the focal element be $X = \{\theta_1\}$, $Y = \{\theta_2\}$ and $Z = \{\theta_3\}$. Let the mass function be $m_0(X) = m_0(Y) = m_0(Z) = \frac{1}{3}$.

Because there is no focal element whose cardinal is larger than 1, the step 2-1 can be skipped for all the times of the loop. Then, in step 2-2, use Deng entropy to calculate the uncertainty of this mass function:

$$H_i(m) = -\frac{1}{3}\log_2(\frac{1}{3}) - \frac{1}{3}\log_2(\frac{1}{3}) - \frac{1}{3}\log_2(\frac{1}{3}) = 1.585 \tag{15}$$

After going through the loop again, the new $H_i(m)$ is also 1.585 since step 2-1 is always skipped. As a result, we escape from the loop and get the information volume of this mass function $H_{IV-mass}(m) = 1.585$.

Actually, this form of mass function is the probability distribution $P_1 =$

$P_2 = P_3 = \frac{1}{3}$. Hence, when the mass function degenerates into the probability distribution, the value of $H_{IV-mass}(m)$ is identical to the Shannon entropy.

Example 4.3:

Consider the frame of discernment be $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Let the mass function be $m_0(\{\theta_1\}) = m_0(\{\theta_2\}) = m_0(\{\theta_3\}) = m_0(\{\theta_1, \theta_2\}) = m_0(\{\theta_1, \theta_3\}) = m_0(\{\theta_2, \theta_3\}) = m_0(\{\theta_1, \theta_2, \theta_3\}) = \frac{1}{7}$.

The information volume of this mass function can be calculated by **Definition 3.2**. The convergence procedure of $H_i(m)$ is listed in Table 2.

Table 2: The convergence procedure of $H_i(m)$

i	$H_i(m)$	i	$H_i(m)$
1	3.887675	9	5.178227
2	4.409314	10	5.187146
3	4.724509	11	5.192498
4	4.914440	12	5.195709
5	5.028700	13	5.197636
6	5.097366	14	5.198792
7	5.138606	15	5.199486
8	5.163366		

According to Table 2, when $i = 15$, $H_i(m) - H_{i-1}(m) < 0.001$, which means that $H_i(m)$ finally converges to 5.199486. Hence the information volume of this mass function is $H_{IV-mass}(m) = 5.199486$.

Example 4.4:

Consider the frame of discernment be $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Let the mass function be $m_0(\{\theta_1\}) = m_0(\{\theta_2\}) = m_0(\{\theta_3\}) = \frac{1}{19}$, $m_0(\{\theta_1, \theta_2\}) = m_0(\{\theta_1, \theta_3\}) = m_0(\{\theta_2, \theta_3\}) = \frac{3}{19}$, $m_0(\{\theta_1, \theta_2, \theta_3\}) = \frac{7}{19}$, which is Deng distribution when the cardinal of the frame of discernment is 3.

The information volume of this mass function can be calculated by **Definition 3.2**. The convergence procedure of $H_i(m)$ is listed in Table 3.

According to Table 3, when $i = 16$, $H_i(m) - H_{i-1}(m) < 0.001$, which means that $H_i(m)$ finally converges to 6.469009. Hence the information volume of Deng distribution is $H_{IV-mass}(m) = 6.469009$.

It should be noted that, the $H_{IV-mass}(m)$ in **Example 4.4** is 6.469009,

Table 3: The convergence procedure of $H_i(m)$

i	$H_i(m)$	i	$H_i(m)$
1	4.247928	9	6.432107
2	5.127754	10	6.447290
3	5.661354	11	6.456402
4	5.983615	12	6.461869
5	6.177746	13	6.465150
6	6.294510	14	6.467119
7	6.364674	15	6.468300
8	6.406810	16	6.469009

which is larger than $H_{IV-mass}(m) = 5.199486$ in **Example 4.3**. As a result, we can conclude that, under the same frame of discernment, the information volume of Deng distribution is larger than the information volume of other forms of mass function.

Actually, the information volume of Deng distribution is called the maximum information volume $H_{MIV-mass}(m)$, which means that, Deng distribution has the largest information volume compare with other mass function.

Example 4.5:

Consider the frame of discernment be $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Let the mass function be $m_0(\Theta) = m_0(\{\theta_1, \theta_2, \theta_3\}) = 1$, which is called the total uncertainty case.

The information volume of the total uncertainty case can be calculated by **Definition 3.2**. The convergence procedure of $H_i(m)$ is listed in Table 4.

Table 4: The convergence procedure of $H_i(m)$

i	$H_i(m)$	i	$H_i(m)$
1	2.807355	10	6.432107
2	4.247928	11	6.447290
3	5.127754	12	6.456402
4	5.661354	13	6.461869
5	5.983615	14	6.465150
6	6.177746	15	6.467119
7	6.294510	16	6.468300
8	6.364674	17	6.469009
9	6.406810		

According to Table 4, when $i = 17$, $H_i(m) - H_{i-1}(m) < 0.001$, which means that $H_i(m)$ finally converges to 6.469009. Hence the information volume of the total uncertainty case is $H_{IV-mass}(m) = 6.469009$.

This example shows that, Deng distribution and the total uncertainty case has identical information volume. Since the information volume of Deng distribution is the maximum information volume, the total uncertainty case also has the maximum information volume. This point is consistent with the intuition.

5. Conclusion

In this paper, we define the information volume of a given mass function based on Deng entropy. In addition, the Deng distribution is presented with the case when Deng entropy achieves its maximum value.

Some concluding remarks can be shown as follows.

1) If the mass function degenerates as probability distribution, the information volume is the same as Shannon entropy.
2) Given a mass function, the corresponding information volume is larger than its uncertainty measured by Deng entropy.
3) If the mass function is Deng distribution, the corresponding information volume is called the maximum information volume, which is larger than the information volume of other forms of mass function.
4) The maximum information volume is lager than the maximum Deng entropy.
5) One interesting point is that Deng distribution and the total uncertainty case has the same information volume, which means that the total uncertainty case also has the maximum information volume. This point is coincide with the intuition.

Acknowledgment

The author greatly appreciates he China academician of the Academy of Engineering, Professor Shan Zhong and Professor You He, for their encouragement to do this research. The author greatly appreciates Professor Yugeng Xi in Shanghai Jiao Tong University to support this work. Research assistant, Jixiang Deng, and Ph.D student, Tao wen, discussed the idea and do a lot editorial works. The author greatly appreciates the continuous funding for the past nearly twenty years including National Natural Science Foundation of China, Grant Nos. 30400067, 60874105, 61174022, 61573290 and 61973332, Program for New Century Excellent Talents in University, Grant No. NCET-08-0345, Shanghai Rising-Star Program Grant No.09QA1402900, Chongqing Natural Science Foundation for distinguished scientist, Grant No. CSCT, 2010BA2003.

References

[1] P. Lee, Probability theory, Bulletin of the London Mathematical Society 12 (4) (1980) 318–319.

[2] L. A. Zadeh, Fuzzy sets, Information and control 8 (3) (1965) 338–353.

[3] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, The Annals of Mathematical Statistics 38 (2) (1967) 325–339. doi:10.1214/aoms/1177698950.

[4] G. Shafer, A mathematical theory of evidence, Vol. 1, Princeton university press Princeton, 1976.

[5] Z. Pawlak, Rough sets, International journal of computer & information sciences 11 (5) (1982) 341–356.

[6] B. Liu, Y. Deng, Risk evaluation in failure mode and effects analysis based on d numbers theory., International Journal of Computers, Communications & Control 14 (5).

[7] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (4) (1948) 379–423.

[8] C. Tsallis, Possible generalization of boltzmann-gibbs statistics, Journal of statistical physics 52 (1-2) (1988) 479–487.

[9] C. Tsallis, Nonadditive entropy: The concept and its use, The European Physical Journal A 40 (3) (2009) 257.

[10] Y. Deng, Deng entropy, Chaos, Solitons & Fractals 91 (2016) 549 – 553.

[11] B. Kang, Y. Deng, The maximum deng entropy, IEEE Access 7 (2019) 120758–120765.