Applied Intelligence
https://doi.org/10.1007/s10489-022-04407-1

Belief entropy rate: a method to measure the uncertainty of interval-valued stochastic processes

Zhiyuan Wang[1,2] · Qianli Zhou[1] · Yong Deng[1,3,4]

Accepted: 13 December 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Entropy rate, as an effective tool in information theory, can measure the uncertainty of stochastic processes modeled by probability mass function. However, when the stochastic process to be measured cannot be accurately modeled, i.e., it is more sense to describe the phenomenon with an interval, the stochastic process needs a more general method to represent. In this paper, the interval-valued stochastic process is modeled with the basic belief assignment in an ordered frame of discernment and the corresponding belief entropy rate is proposed to measure its uncertainty. Two common stochastic processes are discussed. The first is the case of independent identically distributed stochastic processes, where the belief entropy rate is formally the same as the Shannon entropy rate. The second is Markov processes. We construct the evidential Markov chain and calculate its belief entropy rate. Compared with the Shannon entropy rate, the belief entropy rate is easier to implement the Markov chains. By validating in real dataset, the proposed method can better deal with interval information with stronger practicability. When encountering tiny disturbances, the variance of the Shannon entropy rate is more than 50 times the variance of the belief entropy rate, which reflects the stronger robustness of the belief entropy rate.

Keywords Belief entropy rate · Interval-valued stochastic processes · Dempster-Shafer theory · Evidential Markov chain · Robustness

1 Introduction

As an important method for modeling the stochastic process, entropy rate can effectively represent the uncertainty of the stochastic process. Shannon entropy [1] can measure the randomness of probability mass function and is used in dealing with sequence fusion [2] and measurements of chaos or noise [3]. Shannon entropy is generally used to measure the uncertainty of a random variable, but in the case of multivariable or stochastic processes, Shannon entropy faces the problem of non-convergence. Therefore, the Shannon entropy rate [4] is proposed to measure the uncertainty of stochastic processes. It takes a long-term view of a stochastic process and considers the connections between samples so more information is needed to calculate the entropy rate, and more comprehensive results are obtained. Shannon entropy rate is used in network security monitoring [5], behavior studies [6] and image processing [7]. Considering different usage occasions, many scholars have proposed different calculation methods of the entropy rate. Feutrill and Roughan [8] reviewed both parametric and non-parametric techniques to calculate the entropy rate for discrete or continuous-valued data. However, in practical application, the sample cannot be observed with complete precision. The sample can be displayed in an interval rather than a certain value, and the interval indicates that the real value of the sample fluctuates in this interval. It is difficult to calculate the Shannon entropy rate for such interval-valued stochastic processes. Although some methods based

✉ Yong Deng
dengentropy@uestc.edu.cn; prof.deng@hotmail.com

Zhiyuan Wang
wangzhiyuanuestc@hotmail.com

Qianli Zhou
zhou_qianli@hotmail.com

1 Institute of Fundamental and Frontier Science, UESTC, Chengdu, 610054, China

2 Yingcai Honors College, UESTC, Chengdu, 610054, China

3 School of Education, SNNU, Xi'an, 710062, China

4 Department of Management, Technology, and Economics, ETH Zurich, Zurich, 8093, Switzerland

Published online: 05 January 2023

on probability mass function using Shannon entropy have been proposed to deal with interval-valued data [9], the consequence of using entropy instead of entropy rate is to ensure that the samples are independent. For a non-independent multivariate interval-valued stochastic process, it cannot be accurately modeled by probability mass function. Therefore, there are few suitable methods to measure the uncertainty of interval-valued stochastic processes by taking into account both the interval information and the dependencies between variables.

Considering that interval-valued data contain more uncertainty than point-valued data, Dempster-Shafer theory (DST) [10, 11], as an effective means of dealing with uncertainty, is used to model the interval-valued data. For a frame of discernment (FOD), the beliefs can be assigned to both singletons and multi-element subsets, which makes the event space more flexible to better model uncertain information. Compared with probability mass function, the basic belief assignment (BPA) in DST can better model the interval-valued stochastic processes. The reason is that DST allows the existence of ambiguous information and the beliefs can be assigned to any elements in the power set. Since DST has the ability to deal with ambiguous information, it is widely used in decision making [12–15], threat assessment [16], information fusion [17–20] and classification [21–23]. DST can also be extended to complex number [24, 25] and the uncertain information of the complex-valued information can be measured by complex-valued Rényi entropy [26]. Based on the idea of permutation, the random permutation set is proposed [27, 28]. When dealing with multivariate dependencies, it is necessary to discuss the definition of conditional and joint distributions in DST. Some researchers put forward the idea of updating evidence based on Dempster's rule of combination. Fagin and Halpern [29] propose a method to make belief conditioning consistent with imprecise conditional probability calculus. Then the BPA is updated by a linear combination of original BPA and conditional BPA [30]. Other researchers, based on the generalization of Bayesian theory, consider conditional and joint distributions by assigning beliefs to Cartesian sets [31, 32]. However, these methods consider all elements in the power set, which is not suitable for a FOD consisting of ordered elements [33]. For the interval-valued data, due to the continuous range of variables, the ordered FOD is obtained. When generating BPA, the beliefs cannot be assigned to all elements of the power set. So, for ordered FOD in DST, new conditional distribution and joint distribution need to be defined.

Compared with the probability mass function, the incompletely mutually exclusive focal elements make the Shannon entropy unable to reasonably measure the uncertainty of BPA in DST. Therefore, the belief entropy is proposed to measure the uncertainty of BPA. Many scholars have proposed different expressions of the uncertainty of BPA. Hőhle [34] regards the belief function as an extension of the probability mass function, and proposes the entropy of BPA based on the form of Shannon entropy. Similarly, based on commonality function and plausibility function, Smets [35] and Yager [36] propose the information content and dissonance of BPA respectively. The information quality is expressed by Bouhamed et al. [37]. In general, belief entropy needs to consider two aspects of ambiguity, that is discord and non-specificity [38]. Some measurement methods considering both parts are proposed. Wang and Song [39] use the belief function and the plausibility function, regard the belief as an interval, and propose the SU measurement. Jiroušek and Shenoy [40] show discord and non-specificity explicitly, proposing the JS entropy in the form of sum of two parts. In addition, based on the idea of splitting, some scholars have also proposed different forms of belief entropy. According to the different split forms, Pal [41] divides the mass functions into singletons, and Deng [42] divides the mass functions into power sets. Zhou and Deng [43] based on the process of pignistic probability transformation (PPT), propose the Fractal-based Belief (FB) entropy. Furthermore, Time Fractal-based Belief (TFB) entropy [44] explains Deng entropy from the perspective of information volume, and interprets the Deng entropy as the 1-order information volume. There are also methods that completely get rid of the Shannon entropy form and measure uncertainty in terms of the distance between evidence [45]. However, the above methods also do not discuss a FOD consisting of ordered elements [33]. So, based on the idea of splitting, we propose a new belief entropy to model the uncertainty for ordered FOD.

In this paper, we model the interval-valued stochastic processes with BPA and propose the corresponding belief entropy rate. Through theoretical derivation and practical verification, the belief entropy rate can well measure the uncertainty of interval-valued stochastic processes. The contributions of the paper are summarized as follows: (1) A new conditional distribution and joint distribution is defined in DST. It applies not only to ordered FODs mainly used in this paper, but also can be updated to all general cases. (2) We define the belief entropy applicable to ordered FODs. According to conditional and joint distributions, we obtain conditional and joint belief entropies, and show that they obey the chain rule. (3) We calculate the belief entropy rate separately under the independent condition and the Markov model. Especially, the belief entropy rate of the sequence of independent identically distributed (i.i.d.) random variables is degenerated to Shannon entropy rate in form, which implies that belief entropy rate has properties similar to Shannon entropy rate but can be more adaptable in real situations. We construct the evidential Markov chain and find the belief entropy rate is more applicable to Markov processes and has stronger practicability and higher robustness than Shannon entropy rate when dealing with the real data. More refined results can be obtained by increasing the complexity of the model.

The rest of the paper is organized as follows. In Section 2, some preliminaries such as the entropy rate and DST are briefly introduced. In Section 3, propose the conditional and joint distributions and the belief entropy for ordered FOD. Then obtain the belief entropy rate separately under the independent condition and the Markov model. Some real world problems are proposed and the belief entropy rate can show its effectiveness in Section 4. Finally, conclude the paper in Section 5.

2 Preliminaries

2.1 Dempster-Shafer theory

As an extension of probability theory, DST [10, 11], also called belief function theory has the ability to deal with the uncertain information with lots of flexibility. It is defined on a power set, thus it can present more situations than probability theory. Modeling interval data with DST can better handle the uncertainty contained in the interval itself because the beliefs can be assigned to both singletons and multi-element subsets.

Let $\Theta = \{\theta_1, \theta_2, ..., \theta_M\}$ be a finite nonempty set of M mutually exclusive and collectively exhaustive events, and Θ is called the frame of discernment (FOD) [10]. Let 2^Θ be the power set of Θ , which contains all 2^M subsets of Θ. The BPA is defined as

$$m : 2^\Theta \to [0, 1].$$

It is a kind of mapping and satisfies

$$m(\emptyset) = 0, \quad \sum_{A_i \in 2^\Theta} m(A_i) = 1.$$

If $m(A_i) > 0$, then A_i is called the focal element. The mass $m(A_i)$ represents the support of the proposition A_i when more information cannot be obtained by refinement.

In DST, there are some methods to measure uncertainty of BPAs and belief entropy is one of them. Two types of belief entropy based on splitting ideas are introduced. First, Deng proposes Deng entropy [42] based on splitting, which considers all potential states under FOD. Deng entropy has wide usage scenarios and is applied in evidential reasoning [46]. The concept of Deng extropy based on Deng entropy was defined [47, 48] and used to pattern recognition. For a BPA m defined on a FOD Θ, Deng entropy is defined as

$$E_{\mathrm{D}} = -\sum_{A_i \in 2^\Theta} m(A_i) \log \frac{m(A_i)}{2^{|A_i|} - 1},$$

where A_i is a proposition in 2^Θ and $|A_i|$ is the cardinality of A_i.

Second, FB entropy uses a unit time transformation of PPT to recombine the beliefs and is defined as [43]

$$E_{\mathrm{FB}} = -\sum_{A_i \in 2^\Theta} m^{\mathrm{F}}(A_i) \log m^{\mathrm{F}}(A_i),$$

where

$$m^{\mathrm{F}}(A_i) = \sum_{A_i \subseteq A_j} \frac{m(A_j)}{2^{|A_j|} - 1}. \tag{1}$$

The newly generated belief assignment $m^{\mathrm{F}}(A_i)$ is called fractal-based basic probability assignment (FBPA). It is neither probability nor BPA, but has some properties of both [43]. As a modification of Deng entropy, FB entropy improves the counter-intuitive result of maximum Deng entropy [49].

Theorem 1 *(The maximum FB entropy) For a BPA m defined on a FOD $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$, the maximum FB entropy is* [43]

$$E_{\mathrm{FB}}^{\uparrow} = \log(2^M - 1),$$

if and only if

$$m(\Theta) = 1.$$

This result is intuitive and, in this case, the FBPA behaves as a uniform distribution, which is also consistent with the Shannon entropy.

2.2 Entropy rate

Shannon entropy can well measure the uncertainty of a probability mass function and is the basis of the Shannon entropy rate.

Definition 1 (Shannon entropy) For a discrete random variable X whose all possible values are $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$, and it follows a probability mass function P, the Shannon entropy [1] of variable X_p is

$$H(X_p) = -\sum_{i=1}^{M} P(\theta_i) \log P(\theta_i),$$

where $P(\theta_i)$ is the probability of $X_p = \theta_i$.

Shannon entropy measures the uncertainty of a random variable generally, but in the face of a multivariate stochastic process, the Shannon entropy will increase with the increase of the number of random variables, which leads to inconvenient measurement. In this case, entropy rate is proposed to effectively measure the uncertainty of stochastic processes.

Definition 2 (Entropy rate) For a stochastic process $X = \{X_1, X_2, \ldots, X_n\}$, its entropy rate $H(X)$ is defined as [4]

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n).$$

if the limit exists.

Especially, For the sequence of i.i.d. random variables $\{X_1, X_2, \ldots, X_n\}$, its entropy rate can be calculated as

$$\begin{aligned} H(X) &= \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \\ &= \lim_{n\to\infty} \frac{nH(X_1)}{n} = H(X_1). \end{aligned} \tag{2}$$

The entropy rate has another form of expression, let

$$H'(X) = \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, ..., X_1).$$

It can be proved that for a stationary stochastic process

$$H(X) = H'(X).$$

And for an irreducible-aperiodic Markov chain, entropy rate has a simpler expression, that is

$$\begin{aligned} H(X) = H'(X) &= \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, ..., X_1) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1}) = H(X_2|X_1). \end{aligned} \tag{3}$$

Let the stationary distribution of the Markov chain is μ, (3) can be expressed as

$$H(X) = -\sum_{ij} \mu_i P_{ij} \log P_{ij}. \tag{4}$$

2.3 Evidential Markov chain

Evidential Markov chain [50] is based on DST and serve as an extension of classical Markov chain. In this model, each random variable in the stochastic process no longer obeys a probability mass function, but a BPA. Therefore, the Markov transition belief matrix represents the transition belief from propositions to propositions. Evidential Markov chain has many applications such as simulation [51], prediction [50] and image classification [52]. For an evidential Markov process $X = \{X_1, X_2, \ldots, X_n\}$, if the BPAs of consecutive n samples can be obtained and they share the same FOD Θ, then the transition belief matrix P can be defined as [50]

$$P_{ij} = \frac{\sum_{t=1}^{n-1} (m_t(A_i) \cdot m_{t+1}(A_j))}{\sum_{A_k \in 2^\Theta} \sum_{t=1}^{n-1} (m_t(A_i) \cdot m_{t+1}(A_k))}, \quad A_i, A_j \in 2^\Theta. \tag{5}$$

P_{ij} is the simplified form of $P_{A_i A_j}$, representing the average belief of the transition of proposition A_i to proposition A_j. $m_t(A_i)$ represents the assigned belief to A_i when generating BPA according to X_t. It can be proved that the sum of each row of matrix P is 1, which satisfies the properties of Markov chains.

3 Belief entropy rate

Consider a real situation, we want to measure the entropy rate of a stochastic process, but when we sample in the stochastic process, each sample cannot be observed with complete precision. The sample is displayed in an interval which indicates that the real value of the sample fluctuates in this interval. Probability theory cannot accurately model this situation so DST is used to solve this problem. However, in this case, the range of random variables is continuous. Although the continuous value range can be transformed into a discrete form by division, the elements in FOD are still ordered. DST rarely considers the case of an ordered FOD [33]. Therefore, for ordered FOD, we propose the conditional and joint distributions and belief entropy, then use belief entropy rate to measure the uncertainty of interval-valued stochastic processes.

3.1 Conditional and joint distributions for ordered FOD

We take inspiration from the idea of splitting and propose the conditional and joint distributions for ordered FOD. For a FOD $\Theta = \{\theta_1, \theta_2, ..., \theta_M\}$ consisting of ordered elements, first the potential states of ordered FOD need to be defined. In order to be consistent with the power set in the classical FOD, the potential states of ordered FOD are simplified to ordered power set and noted as OP(Θ). The beliefs can only be assigned to the propositions of the ordered power set.

Definition 3 (Ordered power set) The ordered power set $OP(\Theta)$ contains all sets consisting of consecutive elements of Θ and is defined as [33]

$$OP(\Theta) = \{\{\theta_i, \theta_j\}^{ord}, 1 \le i \le j \le M\},$$

where

$$\{\theta_i, \theta_j\}^{ord} = \{\theta_i, \theta_{i+1}, ..., \theta_j\}.$$

Under the ordered power set, the BPA can be defined as a kind of mapping

$$m : OP(\Theta) \to [0, 1],$$

satisfying

$$\sum_{A_i \in OP(\Theta)} m(A_i) = 1.$$

According to the idea of splitting in (1), the ordered fractal-based basic probability assignment (OFBPA) is obtained by considering all potential states of an ordered FOD. It is expressed as

$$m^{\mathrm{OF}}(A_i) = \sum_{A_i \subseteq A_j} \frac{m(A_j)}{|A_j|(|A_j|+1)/2}. \tag{6}$$

how ?

$C_{|A_i|}^2 + |A_i|$

Compared with FBPA in (1), the belief is split into different numbers. The reason is that for ordered FOD, elements in the subset must exist consecutively. Therefore, for a focal element A_i, there are two kinds of subsets that can obtain the belief. The first is all single elements, with a total of $|A_i|$. The other is all continuous subsets containing at least two elements in A_i, with a total of $\binom{|A_i|}{2}$. So the number of all potential subsets is the sum of the two types and is $|A_i|(|A_i|+1)/2$.

The advantages of using OFBPAs are as follows. First, when the belief of a proposition in BPA changes, in OFBPA this variation is distributed between its subsets, which offsets this variation. OFBPA and BPA can be converted into each other, so they have the same information. Using OFBPA to deal with the uncertain information can make the model more robust. Second, since OFBPA divides the belief into subsets, more propositions gain belief. If the belief is assigned to Θ, then in OFBPA, all propositions obtain beliefs. Therefore, for the evidential Markov chain, at each moment, it is only necessary to assign the beliefs to a few focal elements, including Θ, to generate BPA, and then convert it into OFBPA, so that all propositions can obtain the belief. The Markov chains thus generated are irreducible and aperiodic, because all propositions are assigned the belief at each moment, and any two propositions are reachable, which is the basic condition for the entropy rate to be calculated. Therefore, the conditional distribution and joint distribution are proposed based on OFBPAs.

Suppose that there are two variables X and Y and their marginal BPAs are m_1 and m_2. The conditional BPA $m_{1|2}$ is known. To obtain the joint OFBPA m_{12}^{OF}, a unit time PPT is performed on the marginal BPA and the conditional BPA respectively. Then we can obtain the OFBPAs m_2^{OF} and $m_{1|2}^{\mathrm{OF}}$ presented as follows.

$$m_2^{\mathrm{OF}}(B_j) = \sum_{B_j \subseteq B_k} \frac{m_2(B_k)}{|B_k|(|B_k|+1)/2},$$

$$m_{1|2}^{\mathrm{OF}}(A_i|B) = \sum_{A_i \subseteq A_k} \frac{m_{1|2}(A_k|B)}{|A_k|(|A_k|+1)/2}.$$

$m_{1|2}^{\mathrm{OF}}(A_i|B_j)$ represents the reassigned belief of X to be proposition A_i when Y is already proposition B_j. Then, combine two OFBPAs to obtain the joint OFBPA m_{12}^{OF}. Considering that whether in probability mass function or in BPA, the joint mass function is the product of the conditional mass function and the initial mass function, so a similar combination is also used in OFBPA and is presented as

$$m_{12}^{\mathrm{OF}}(A_i,B_j)=m_{1|2}^{\mathrm{OF}}(A_i|B_j)\cdot m_2^{\mathrm{OF}}(B_j),\quad \forall A_i, B_j \in OP(\Theta).$$

$m_{12}^{\mathrm{OF}}(A_i, B_j)$ represents how strongly the evidences support X and Y to be propositions A_i and B_j respectively aftering recombining the beliefs. In addition, independence is defined in this step. The necessary and sufficient condition for two random variables X, Y to be independent is

$$m_{12}^{\mathrm{OF}}(A_i, B_j) = m_1^{\mathrm{OF}}(A_i)\cdot m_2^{\mathrm{OF}}(B_j), \quad \forall A_i, B_j \in OP(\Theta). \tag{7}$$

The obtained conditional OFBPA $m_{1|2}^{\mathrm{OF}}$ and joint OFBPA m_{12}^{OF} are regarded as the conditional distribution and joint distribution respectively. The above operation is not only applicable to the ordered FOD, but can also be generalized to the normal case by changing the ordered FOD to classical and replace OFBPA with FBPA in (1).

3.2 Belief entropy for ordered BPA

A new belief entropy E_B is proposed based on the idea of splitting for the ordered BPA. It is defined as follows.

Definition 4 (Belief entropy) For a FOD $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$ consisting of ordered elements, there is a random variable X that obeys a BPA m on $OP(\Theta)$, the belief entropy of X can be represented as

$$E_B(X) = -\sum_{A_i \in OP(\Theta)} m^{\mathrm{OF}}(A_i) \log m^{\mathrm{OF}}(A_i), \tag{8}$$

where $m^{\mathrm{OF}}(A_i)$ is the OFBPA defined in (6).

Assume that two variables X and Y share the same ordered FOD $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$ and both of them comply with its own BPA $m_1(X)$ and $m_2(Y)$.

The conditional BPA between X and Y is $m_{1|2}(X|Y)$. Through the conditional and joint operations, the joint BPA $m_{12}(X, Y)$ is obtained. All definitions of OFBPA are similar to (6). Then the conditional and joint belief entropies are defined as follows.

Definition 5 (Joint belief entropy) Joint belief entropy measures the uncertainty of the joint system of X, Y and is represented as

$$E_{\mathrm{B}}(X, Y) = -\sum_{A_i, B_j \in OP(\Theta)} m_{12}^{\mathrm{OF}}(A_i, B_j) \log m_{12}^{\mathrm{OF}}(A_i, B_j). \tag{9}$$

Definition 6 (Conditional belief entropy) Conditional belief entropy measures the uncertainty of X given Y and is represented as

$$E_{\mathrm{B}}(X|Y) = -\sum_{A_i, B_j \in OP(\Theta)} m_{12}^{\mathrm{OF}}(A_i, B_j) \log m_{1|2}^{\mathrm{OF}}(A_i|B_j). \tag{10}$$

Theorem 2 *The conditional and joint belief entropies satisfy the chain rule:*

$$E_{\mathrm{B}}(X, Y) = E_{\mathrm{B}}(X|Y) + E_{\mathrm{B}}(Y).$$

Proof 1 of Theorem 2:

$$\begin{aligned}
E_{\mathrm{B}}(X, Y) &= -\sum_{A_i, B_j \in OP(\Theta)} m_{12}^{\mathrm{OF}}(A_i, B_j) \log m_{12}^{\mathrm{OF}}(A_i, B_j) \\
&= -\sum_{A_i, B_j \in OP(\Theta)} m_{12}^{\mathrm{OF}}(A_i, B_j) \log m_{1|2}^{\mathrm{OF}}(A_i|B_j) \\
&\quad -\sum_{A_i, B_j \in OP(\Theta)} m_{12}^{\mathrm{OF}}(A_i, B_j) \log m_{2}^{\mathrm{OF}}(B_j) \\
&= -\sum_{A_i, B_j \in OP(\Theta)} m_{12}^{\mathrm{OF}}(A_i, B_j) \log m_{1|2}^{\mathrm{OF}}(A_i|B_j) \\
&\quad -\sum_{B_j \in OP(\Theta)} m_{2}^{\mathrm{OF}}(B_j) \log m_{2}^{\mathrm{OF}}(B_j) \\
&= E_{\mathrm{B}}(X|Y) + E_{\mathrm{B}}(Y).
\end{aligned}$$

□

If X, Y meet the condition of independence in (7), it can be proved that

$$E_{\mathrm{B}}(X, Y) = E_{\mathrm{B}}(X) + E_{\mathrm{B}}(Y). \tag{11}$$

With the relationship in (11), the belief entropy rate of the i.i.d. stochastic process can be calculated. Assume a sequence of i.i.d. random variables $X = \{X_1, X_2, \ldots, X_n\}$, each X_i obey the same BPA m. Then the belief entropy rate $E_{\mathrm{B}}(X)$ is calculated as

$$\begin{aligned}
E_{\mathrm{B}}(X) &= \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{B}}(X_1, X_2, \ldots, X_n) \\
&= \lim_{n\to\infty} \frac{n E_{\mathrm{B}}(X_1)}{n} = E_{\mathrm{B}}(X_1).
\end{aligned}$$

The result is similar to (2), showing that this kind of expression can degenerate to Shannon entropy rate in form when the stochastic process is i.i.d. . However, in many cases, the stochastic process does not satisfy the i.i.d. condition. Next more general case will be discussed, that is the Markov process.

3.3 Construction of the evidential Markov chain

Assume that there is an interval-valued Markov chain $X = \{X_1, X_2, \ldots, X_n\}$, and a sample $x = \{x_1, x_2, \ldots, x_n\}$ is obtained. Each x_i is displayed in an interval $[l_i, r_i]$ and the range of X_i is $[L, R]$, which can be roughly estimated as

$$L = \min_{1\le i\le n} \{l_i\}, \quad R = \max_{1\le i\le n} \{r_i\}. \tag{12}$$

The first step is to define the FOD Θ. Consider that Θ is the range of X_i, so Θ is mapped to the interval $[L, R]$. By dividing $[L, R]$ into M intervals $\{\theta_1, \theta_2, \ldots, \theta_M\}$ of equal length to constitute the ordered FOD. In addition, every element in $OP(\Theta)$ can be mapped to an interval which is the part of the range of X_i.

After defining the FOD, BPA m_i is generated according to each x_i. Since each focal element A_j can be mapped to an interval $[l_{A_j}, r_{A_j}]$, a focal element A_j with the smallest length can be found, which contains $[l_i, r_i]$. We name the found A_j as $\lceil x_i \rceil$, then assign beliefs to $\lceil x_i \rceil$ and Θ according to the following rule:

(1) When $\lceil x_i \rceil = \Theta$,

$$m_i(\Theta) = 1. \tag{13}$$

(2) When $\lceil x_i \rceil \neq \Theta$,

$$\begin{aligned}
m_i(\Theta) &= \frac{r_i - l_i}{R - L}, \\
m_i(\lceil x_i \rceil) &= 1 - m_i(\Theta).
\end{aligned} \tag{14}$$

This rule shows that when $\lceil x_i \rceil = \Theta$, we cannot refine the belief due to the large uncertainty of x_i. It can only be define as a state of complete ignorance $m_i(\Theta) = 1$. When $\lceil x_i \rceil \neq \Theta$, we take the ratio of the observed interval length to the total length as uncertainty and assign it to Θ. The remaining belief is assigned to $\lceil x_i \rceil$, indicating a degree of elimination of the uncertainty. This ensures that as the observation interval becomes longer, the uncertainty

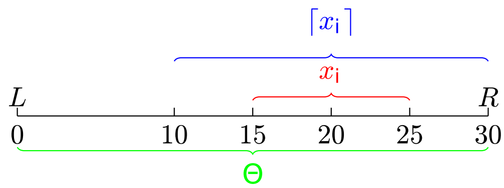


Fig. 1 Example of $x_i = [15, 25]$, $M = 3$

of the system becomes larger. Compared with the evidential Markov chain proposed by He and Jiang [50], although in their method, the problem of ordered FOD is taken into account, they only assign beliefs to subsets with one or two consecutive elements in FOD. In their method, The lengths of the intervals corresponding to all focal elements are similar, that is to say, for the ordered FOD, the uncertainty represented by $\{\theta_1, \theta_1\}^{ord}$ and $\{\theta_1, \theta_2\}^{ord}$ are similar, which is not intuitive. For generation rules in (13) and (14), the beliefs an be assigned to all elements in $OP(\Theta)$ and different focal elements are mapped into intervals of different lengths according to the uncertainty involved. He and Jiang [50] assign beliefs according to the similarity of the intervals, and each focal element needs to be compared with the sample, which is a complicated process. By directly assigning beliefs to $\lceil x_i \rceil$ and Θ, it is simpler and more intuitive to express the interval-valued data.

Then use a unit time PPT on the obtained BPA m_i, the OFBPA m_i^{OF} corresponding to x_i is obtained. Example 1 shows how to generate BPA and OFBPA from a certain x_i.

Example 1 In Fig. 1, assume the range of X_i is [0, 30] and is divided into 3 intervals {[0, 10], [10, 20], [20, 30]}, so $M = 3$ in this case. The observed x_i is [15, 25], so the shortest focal element A_j containing x_i is [10, 30]. Therfore, $\lceil x_i \rceil$ is $\{\theta_2, \theta_3\}^{ord}$. The generated BPA and OFBPA are shown in Table 1.

Considering that M is artificially specified, that is to say, the degree of division of the range can be controlled. If $M = 4$, the generated BPA and OFBPA are shown in Table 2.

Table 1 The generated BPA and OFBPA according to $x_i = [15, 25]$, $M = 3$

Proposition	Characterization range	BPA	OFBPA
$\{\theta_1, \theta_1\}^{ord}$	[0,10]	0	0.056
$\{\theta_2, \theta_2\}^{ord}$	[10,20]	0	0.278
$\{\theta_3, \theta_3\}^{ord}$	[20,30]	0	0.278
$\{\theta_1, \theta_2\}^{ord}$	[0,20]	0	0.056
$\{\theta_2, \theta_3\}^{ord}$	[10,30]	0.67	0.278
$\{\theta_1, \theta_3\}^{ord}$	[0,30]	0.33	0.056

Table 2 The generated BPA and OFBPA according to $x_i = [15, 25]$, $M = 4$

Proposition	Characterization range	BPA	OFBPA
$\{\theta_1, \theta_1\}^{ord}$	[0,7.5]	0	0.03
$\{\theta_1, \theta_2\}^{ord}$	[0,15]	0	0.03
$\{\theta_1, \theta_3\}^{ord}$	[0,22.5]	0	0.03
$\{\theta_1, \theta_4\}^{ord}$	[0,30]	0.33	0.03
$\{\theta_2, \theta_2\}^{ord}$	[7.5,15]	0	0.03
$\{\theta_2, \theta_3\}^{ord}$	[7.5,22.5]	0	0.03
$\{\theta_2, \theta_4\}^{ord}$	[7.5,30]	0	0.03
$\{\theta_3, \theta_3\}^{ord}$	[15,22.5]	0	0.26
$\{\theta_3, \theta_4\}^{ord}$	[15,30]	0.67	0.26
$\{\theta_4, \theta_4\}^{ord}$	[22.5,30]	0	0.26

From the OFBPA m_i^{OF} obtained by each x_i, the transition belief matrix P is built as follows.

$$P_{ij} = \frac{\sum_{t=1}^{n-1} (m_t^{\mathrm{OF}}(A_i) \cdot m_{t+1}^{\mathrm{OF}}(A_j))}{\sum_{A_k \in OP(\Theta)} \sum_{t=1}^{n-1} (m_t^{\mathrm{OF}}(A_i) \cdot m_{t+1}^{\mathrm{OF}}(A_k))}, \times A_i, A_j \in OP(\Theta). \tag{15}$$

P_{ij} is the simplified form of $P_{A_i A_j}$, representing the average belief of the transition of proposition A_i to proposition A_j and n is the observed length of the stochastic process. $m_t^{\mathrm{OF}}(A_i)$ denotes the generated OFBPA according to x_t. Let the stationary distribution of the evidential Markov chain is μ and μ is the solution of the following equation.

$$\mu_j = \sum_i \mu_i P_{ij}, \qquad A_i, A_j \in OP(\Theta). \tag{16}$$

μ_i is the simplified form of μ_{A_i}, representing the belief of proposition A_i after stabilization.

3.4 Belief entropy rate for Markov processes

Theorem 3 *Let $X = \{X_1, X_2, ..., X_n\}$ be the evidential Markov chain. P is the transition belief matrix and μ is the stationary distribution in (16), then the belief entropy rate $E_{\mathrm{B}}(X)$ can be calculated by*

$$E_{\mathrm{B}}(X) = -\sum_{A_i, A_j \in OP(\Theta)} \mu_i P_{ij} \log P_{ij}. \tag{17}$$

A_i and A_j are both focal elements which can be mapped to intervals.

Table 3 $\{x_i\}$ of length 4

i	1	2	3	4
x_i	[0,30]	[15,25]	[11,20]	[13,15]

Table 4 Transition belief matrix P

P_{ij}	$\{\theta_1,\theta_1\}^{ord}$	$\{\theta_1,\theta_2\}^{ord}$	$\{\theta_1,\theta_3\}^{ord}$	$\{\theta_2,\theta_2\}^{ord}$	$\{\theta_2,\theta_3\}^{ord}$	$\{\theta_3,\theta_3\}^{ord}$
$\{\theta_1,\theta_1\}^{ord}$	0.032	0.032	0.032	0.842	0.032	0.032
$\{\theta_1,\theta_2\}^{ord}$	0.032	0.032	0.032	0.842	0.032	0.032
$\{\theta_1,\theta_3\}^{ord}$	0.032	0.032	0.032	0.842	0.032	0.032
$\{\theta_2,\theta_2\}^{ord}$	0.022	0.022	0.022	0.892	0.022	0.022
$\{\theta_2,\theta_3\}^{ord}$	0.044	0.044	0.044	0.78	0.044	0.044
$\{\theta_3,\theta_3\}^{ord}$	0.044	0.044	0.044	0.78	0.044	0.044

Proof 2 of Theorem 3:

From (3) , in a stationary Markov chain

$$
\begin{aligned}
E_{\mathrm{B}}(X) &= E'_{\mathrm{B}}(X) = \lim_{n\to\infty} E_{\mathrm{B}}(X_n|X_{n-1}) \\
&= \lim_{n\to\infty} - \sum_{i,j\in OP(\Theta)} m_{n-1}(i) m_{n|n-1}(j|i) \\
&\quad \times \log m_{n|n-1}(j|i) \\
&= - \sum_{i,j\in OP(\Theta)} \mu_i P_{ij} \log P_{ij}.
\end{aligned}
$$

□

Example 2 shows the process of calculating the belief entropy rate of interval-valued Markov processes.

Example 2 There is an interval-valued Markov chain $X = \{X_i\}$ and the observed length is $n = 4$. Observed values $\{x_i\}$ are shown in Table 3.

The range of X_i is $[0, 30]$ by (12). Set $M = 3$ and use ordered FOD $\Theta = \{\theta_1, \theta_2, \theta_3\}$ as mentioned in Table 1, the generated OFBPAs in 4 periods can be calculated by the same way in Table 1. Then using (15) , the transition belief matrix P can be built and shown in Table 4.

Then according to (17) and using natural logarithm, the belief entropy rate of $\{X_i\}$ is

$$E_{\mathrm{B}}(X) = E'_{\mathrm{B}}(X) = 0.5458.$$

In Section 3, method to calculate the belief entropy rate is proposed. To get the entropy rate of interval-valued stochastic processes, belief entropy is proposed in (8) and the forms of joint belief entropy and conditional belief entropy are obtained in (9) and (10). For an i.i.d. random variable sequence, the belief entropy rate is consistent with the Shannon entropy rate in form. And for Markov chains, after generating OFBPAs through according rules in (13) and (14), the transition belief matrix can be obtained, so as to calculate the belief entropy rate. It is worth noting that the Markov chain obtained by this method is irreducible and aperiodic because under normal circumstances, Θ is assigned the belief, and then through a unit time PPT, all elements in $OP(\Theta)$ can obtain the belief, so that the two propositions can reach each other with a certain probability. Thus, the belief entropy rate can easily implement the Markov chains.

4 Practical cases studies

4.1 A practical case about price volatility

4.1.1 Problem statement

As is known to all, prices of almost all commodities fluctuate over time and how to measure the volatility is an open issue. The price per day can be regarded as an interval-valued stochastic process $\{X_i\}$ because x_i per day may not be observed with complete precision. Some commodities such as crude oil, the prices fluctuate within a day and only the high and low prices can be obtained. When measuring volatility, not only the prices change between days, but also the prices change within a day needs to be considered. In the Appendix A, there are price data[1] of crude oil in February and March 2022 (20 consecutive trading days each month). The type of the oil is WTI crude oil futures. All prices are in US dollars per barrel. The price of WTI crude oil fluctuated greatly in February and March 2022 and the degree of price fluctuation can well reflect the uncertainty of the stochastic process. The price itself can also be seen as a Markov process, so it is suitable for study. The figure of transaction price per day is demonstrated in Fig. 2. Three methods are discussed and the belief entropy rate can show its efficiency in this case.

In Fig. 2, the part marked by the dotted line is the range of daily price fluctuations. Connect the midpoint of the interval to draw a line chart of price fluctuation.

4.1.2 Three methods for measuring volatility

Parkinson volatility Parkinson volatility [53] is a common method to measure price volatility. It considers the high and

[1] The data of oil price comes from https://cn.investing.com.

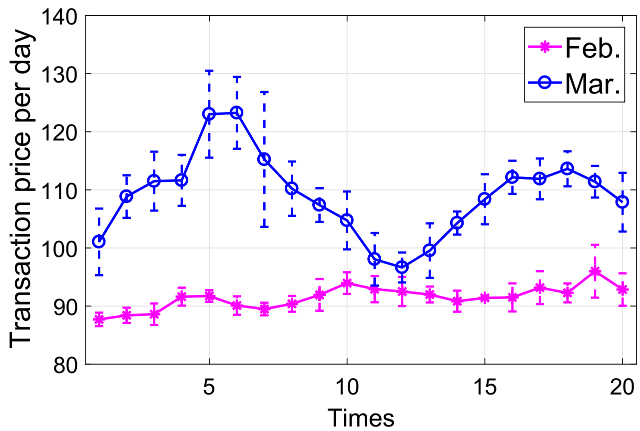


Fig. 2 Transaction price per day in Feb. and Mar

low price of the day, and can be calculated as

$$\sigma_{Parkinson} = \sqrt{\frac{1}{4Tln2}\sum_{t=1}^{T} ln\left(\frac{h_t}{l_t}\right)^2}.$$

T is the number of date, h_t is the high price on day t while l_t is the low price. Parkinson volatility just focus on the prices fluctuate within a day but ignore the prices change between days.

DTMC model to calculate Shannon Entropy rate Assume $X = \{X_1, X_2, \ldots, X_n\}$ is a stochastic process, and the Shannon entropy rate can be calculated by using Markov chain. Considering that price is a continuous variable, it needs to be turned into discrete by dividing and processed by the discrete time Markov chain (DTMC) model [50]. However, DTMC model can only process exact numbers, so x_i is calculated as the average of high and low price. The method to obtain the transition probability matrix P according to the known sequence $\{x_i\}$ by DTMC model is as follows. Firstly, it divides the range of X_i into M intervals as M states spaces to make sure that each x_i can be classified into only one state space. Θ consists of these basic state space. The range is roughly estimated as the upper and lower bounds of x_i. Secondly, N_{ij} is used to denote the number of transitions from the state i to state j. Then the transition probability matrix P can be calculated by

$$P_{ij} = \frac{N_{ij}}{\sum\limits_{k\in\Theta} N_{ik}}, \quad i, j \in \Theta.$$

If the Markov chain obtained by this method is irreducible and aperiodicity, then its stationary distribution μ can be obtained and Shannon entropy rate can be calculated as (4). Otherwise, the Shannon entropy rate cannot be calculated by this model, returning an error status.

Belief entropy rate The algorithm steps are as follows. Firstly, define FOD Θ and generate BPA for each day. The range of X_i can be obtained by calculating the minimum and maximum values of all numbers in (12), then choose M as the number of elements in FOD Θ. Divide the range into M intervals and each interval represents an element in Θ. After determining the focal elements in $OP(\Theta)$, the OFBPA per day is generated according to generation rules in (13) and (14). Then use (15) to gain transition belief matrix P. Finally use (17) to get the belief entropy rate.

4.1.3 Results of three methods

For the two methods of calculating entropy rate, M is both selected as 4 which means there are 4 basic states in DTMC model and 10 focal elements when calculating belief entropy rate. To make the result more explicit, both the entropy rate and relative entropy rate are used. The relative entropy rate is the entropy rate divided by the maximum. The entropy rate can be regarded the average of the entropy, so the maximum of entropy and entropy rate should be the same. For belief entropy, the maximum is 2.3026 when $m(\Theta) = 1$ with natural logarithm and for Shannon entropy, the maximum is the logarithm of the number of states, namely $ln4 = 1.3863$. When calculating the belief entropy rate, the upper and lower bounds of all data in February and March are considered for the range. While for Shannon entropy rate, since the data ranges in February and March are quite different, in order not to return to an error state, the Shannon entropy rate of February and March data are calculated separately, so the ranges are different for the two calculations. The result is shown in Table 5.

In Table 5, it is clear that all models have larger measurements for March, indicating that prices fluctuated more in March and this is intuitive by analyzing Fig. 2. But the three methods have significant differences. Parkinson volatility just focus on the price volatility in one day but lose information between days, so it is more a static description of data. The remaining two methods take into account the correlation between daily data because both of them is based on entropy rate. The comparison between Shannon entropy rate and belief entropy rate are discussed below and we just focus on the February data so for the two kinds of entropy

Table 5 The measured volatility by three methods

Volatility	February	March
Parkinson volatility	0.0272	0.0546
Shannon entropy rate	0.8085	0.8970
Relative Shannon entropy rate	0.5832	0.6470
Belief entropy rate	0.5651	1.7605
Relative belief entropy rate	0.2454	0.7646

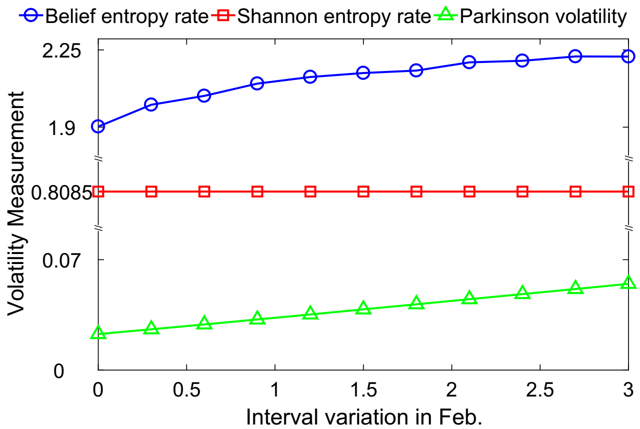


Fig. 3 Influence of interval length change on three methods

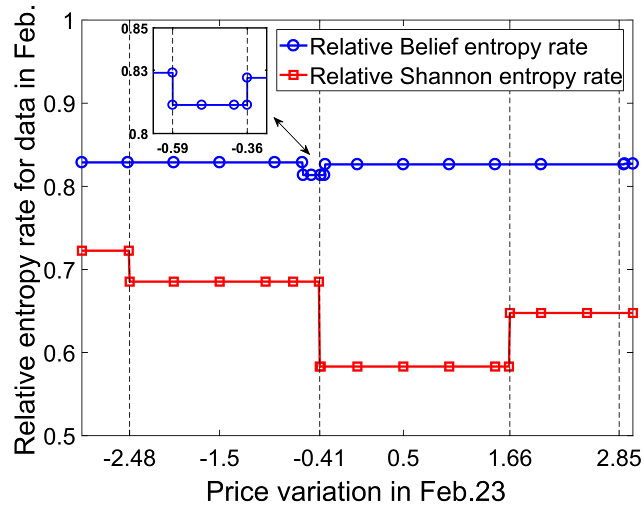


Fig. 4 Influence of modifying data of a certain day on entropy rate

rate, only February data is considered when calculating the data range. For the belief entropy rate, the result is 1.903 instead of 0.5651 in Table 5 if the February data are calculated separately.

Firstly, practicability. Belief entropy rate can better deal with interval data such as the price in this case. Shannon entropy rate can just deal with the accurate sequence and thus lose information within a day. Figure 3 demonstrates the ability of three methods to handle uncertain data when the interval length is continuously increased.

In Fig. 3, the abscissa represents the change of all February data which means the minimum value of the price fluctuation range of each day will decrease by a certain value and the maximum value will also increase by the same value to ensure that the average value per day remains unchanged. Since the Shannon entropy can only regard the interval as the average value, the Shannon entropy rate will not change. While, with the increase of fluctuation amplitude, belief entropy rate and Parkinson volatility show a monotonic increasing trend which symbolizes that the uncertainty of the stochastic process is gradually increasing. This shows that belief entropy rate has a strong ability to deal with interval-valued data and has better practicability.

Secondly, robustness. To verify the dynamic characteristics of each model, the data on February 23 (low price is 90.64 and the high price is 93.9) is constantly modified. By increasing or decreasing the high and low price of the data on Feb.23 at the same time, partial measurement results on the price volatility in February are shown in Table 6. Since the Parkinson volatility is almost unchanged, only the results of two methods based on entropy rate are shown in Fig. 4.

In Fig. 4, when the reduction of data on February 23 is between 0.41 and 0.42, Shannon entropy experienced an obvious mutation. This means Shannon entropy rate changes unevenly and a small disturbance may cause great changes, which is not convincing. For belief entropy rate, it has no such problem. the belief entropy rate will mutate but the change is not obvious. To numerically measure the robustness of the two methods, we calculated the variance of the data in Table 6 for the two methods. The variance is defined as

$$S^2 = \frac{\sum_{i=1}^{n}(H_i - \overline{H})^2}{n-1}.$$

H_i is the relative entropy rate, $\overline{H}$ is the mean of H_i and n is the amount of data. It can be calculated that the variance of the relative Shannon entropy rate is $S_H^2 = 2.89 * 10^{-3}$, while the variance of the relative belief entropy rate is $S_B^2 = 4.96 * 10^{-5}$. S_H^2 is more than 50 times S_B^2 which indicates that the robustness of the belief entropy rate is better when faced with disturbances.

Table 6 The volatility measurement with fluctuating data

Volatility	−0.7	−0.6	−0.59	−0.5	−0.42	−0.41	−0.36	−0.35	0	1
Parkinson volatility	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027
Shannon entropy rate	0.950	0.950	0.950	0.950	0.950	0.809	0.809	0.809	0.809	0.809
Relative Shannon entropy rate	0.685	0.685	0.685	0.685	0.685	0.583	0.583	0.583	0.583	0.583
Belief entropy rate	1.908	1.908	1.873	1.873	1.873	1.873	1.873	1.903	1.903	1.903
Relative belief entropy rate	0.829	0.829	0.814	0.814	0.814	0.814	0.814	0.826	0.826	0.826

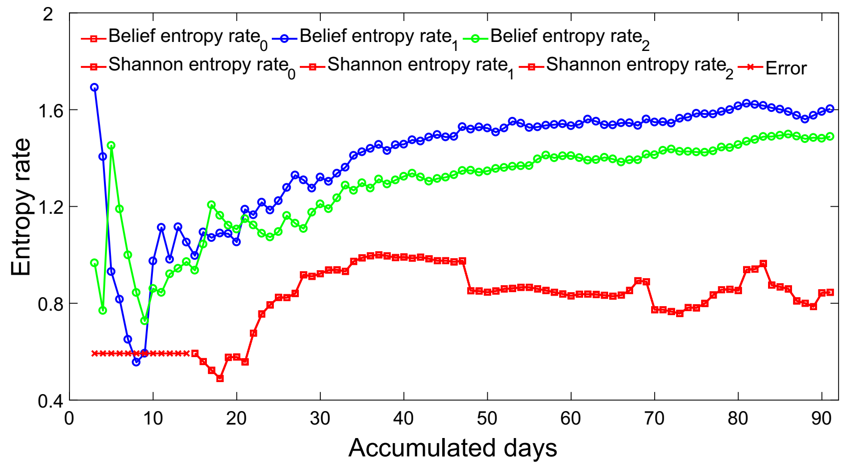


Fig. 5 Maximum temperature entropy rate simulation

The reason for the mutation of Shannon entropy rate is that it is difficult to process intervals, so for each interval, Shannon entropy rate can only perceive the midpoint of the interval, and this midpoint will be definitely divided into a certain basic state. While for belief entropy, the method of assigning beliefs is more flexible. It takes into account the uncertainty of the interval itself and this part of the uncertainty is assigned to Θ, so that even if a certain $\lceil x_i \rceil$ according to x_i changes due to the small disturbance, Θ is assigned the belief, then through a unit time PPT, all elements in $OP(\Theta)$ will obtain the belief. This method offsets part of the impact caused by the change of $\lceil x_i \rceil$. For Shannon entropy, there is no such counteracting strategy after encountering disturbances, which leads to a complete mutation of a certain x_i from one basic state to another basic state, and the calculation result is abruptly changed. So, the robustness of the belief entropy rate is better.

4.2 Weather parameter simulation experiment

Temperature can also be viewed as a common Markov process $X = \{X_1, X_2, \ldots, X_n\}$. We select the daily maximum temperature in Paris as experimental data.[2] The data range is from January 1 to April 1, 2022, a total of 91 days of maximum temperature data in Celsius (°C). The data of the highest temperature we have collected are all accurate values, but in practice, due to inaccurate observations and other disturbances, the data we finally obtain can be in the form of intervals. To simulate this perturbed situation, we randomly added perturbations to the exact collected data, but we ensured that the midpoint of the interval is still the exact collected value and the length of interval follows normal distribution. In Fig. 5, we model the entropy rate changes as the data increases, including the Shannon entropy rate and the belief entropy rate. We still use the DTMC model mentioned in 4.1.2 to calculate the Shannon entropy rate, use the method proposed in this paper to calculate the belief entropy rate, and let the number of basic states $M = 4$.

In Fig. 5, we add two perturbations, corresponding to Belief entropy rate$_1$ and Belief entropy rate$_2$. Belief entropy rate$_0$ shows the case where the belief entropy rate handles the point-valued data set, that is, the case without perturbation. When dealing with point-valued data, according to the BPA generation rules in (13) and (14), the beliefs will only be assigned to the single subsets. At this time, DST degenerates into probability theory, and the belief entropy rate also degenerates into Shannon entropy rate as shown in Fig. 5. For the two cases where the perturbation is added, the belief entropy rate produces different calculation results, while the Shannon entropy remains the same in all cases because the interval information is ignored. This again demonstrates the better practicality of the belief entropy rate. In addition, compare the belief entropy rate and Shannon entropy rate, it can also be found that when the amount of data is large enough, the Shannon entropy rate fluctuates more violently, while the belief entropy rate changes relatively smoothly, this is because Shannon entropy rate only considers the midpoint of the interval, and the fluctuation of the midpoint is large. The belief entropy rate considers the information of the interval and assigns the beliefs according to the interval and generates OFBPA based on the idea of splitting, so the fluctuation of the belief entropy rate is relatively stable.

At the beginning of the curve, the Shannon entropy rate and the degraded belief entropy rate go through a non-computable state. This is because the amount of data at the beginning is small, resulting in no data points in some state spaces. In this case, the generated Markov chain does not satisfy the aperiodic and irreducible conditions, there

[2]The data of maximum temperature comes from www.meteomanz.com.

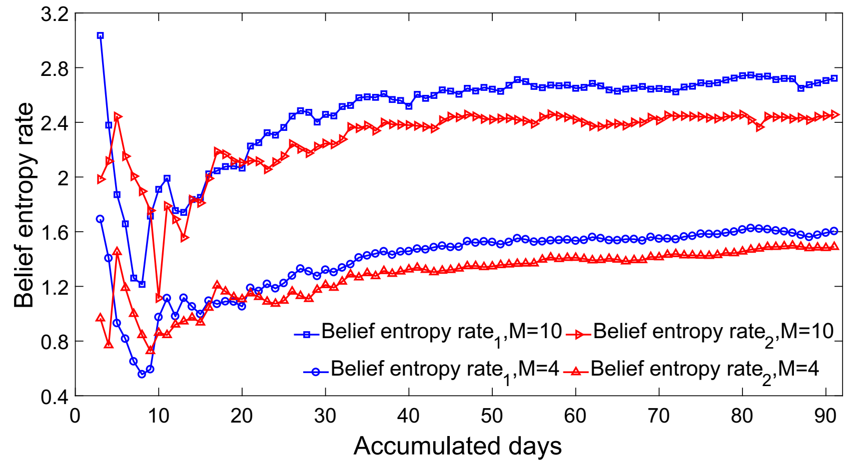


Fig. 6 The influence of M on the belief entropy rate

is no stationary distribution, so the Shannon entropy rate cannot be calculated. And for belief entropy rate, at the beginning of the curve, the fluctuation is very violent, which is also caused by the too small amount of data. As the amount of data gradually increases, the two curves Belief entropy rate$_1$ and Belief entropy rate$_2$ gradually converge. This shows that when dealing with interval-valued Markov processes, the Markov chain corresponding to the belief entropy rate is aperiodic and irreducible because through generation rules in (13) and (14), Θ is assigned the belief, then by splitting, all propositions are assigned beliefs and any two propositions are reachable. For the Shannon entropy rate, if the number of divided states is too large, that is, when M is large, or when the amount of data is small, there may be an error state. The reason behind it is that when calculating the Shannon entropy rate, each x_i is completely divided into a certain basic state, while the belief entropy rate will not divide x_i in such a fixed manner and all propositions are assigned the beliefs. Therefore, the belief entropy rate is easier to implement the Markov chains.

In addition, we also discussed the influence of the number of states M on the belief entropy rate. For the above two groups of perturbed data, two cases of $M = 4$ and $M = 10$ were considered respectively. The result is shown in Fig. 6.

The increase of M indicates that we increase the complexity of the model and can divide the value range space more carefully, and the beliefs can be assigned more accurately. Considering the $\lceil x_i \rceil$ corresponding to each x_i, that is, the focal element that completely contains x_i, when the interval endpoint of x_i changes within a certain range, $\lceil x_i \rceil$ cannot feel the change. For example, in Fig. 1 when l_i fluctuates between $[10, 20)$ or r_i fluctuates between $(20, 30]$, $\lceil x_i \rceil$ will not change. But as M increases, x_i can be locked in a more accurate focal element, thereby assigning the beliefs more precisely. Corresponding to Fig. 6, it is shown that when $M = 10$, the fluctuation of the entropy rate is more obvious and more refined result is obtained.

5 Conclusion

In this paper, we propose the OFBPA to model the interval-valued stochastic processes. To calculate the belief entropy rate, we discuss the case of i.i.d and Markov chains in detail. We propose a method for building the evidential Markov chain and calculate the belief entropy rate for it. Some experimental results show the practicability and robustness of the belief entropy rate, and it is also compatible with the Shannon entropy rate. The main contributions of the paper are briefly concluded as follows. First, the concepts about conditional and joint distributions are proposed based on the idea of splitting. The conditional and joint operations are suitable for dealing with dependency problems in the case of ordered FOD. Second, we propose the belief entropy and its joint and conditional forms. The definition of the belief entropy can help to measure the ambiguity of BPAs, and more generally, it provides a method for measuring the uncertainty of continuous random variables. The joint and conditional belief entropy can help to deal with dependence problems in DST. Third, the method to calculate the belief entropy rate of interval-valued stochastic processes is presented. For i.i.d. sequences, the belief entropy rate degenerates to Shannon entropy rate in form. For Markov processes, we discuss in detail how the evidential Markov chain is generated and calculate the belief entropy rate in this case. In the practical cases presented in Section 4, the belief entropy rate shows its strong practicability to deal with interval-valued data. In the face of disturbances, the variance of Shannon entropy rate is more than 50 times that of the belief entropy rate, reflecting the strong robustness of the belief entropy rate. In addition, the belief entropy rate is easier to implement the Markov chains compared with Shannon entropy rate and more refined results can be obtained by increasing the complexity of the model.

However, there are still some limitations in this model. First, the Markov process mentioned in this paper is only a first-order Markov process, and there is no discussion on the higher-order case. Other more complex stochastic processes, such as random walks, are also to be discussed. Second, the information of the interval can be further improved. In our model, we assume the data is uniformly distributed in the interval, but in practice the data can be normally distributed in the interval, or when we can obtain the accurate distribution of the data in this interval, a more accurate method is needed to calculate the belief entropy rate. Third, there is only one observer in our method. When there are multiple observers, we can obtain a more accurate measure of entropy rate by generating BPAs from the data returned by each observer and fusing them reasonably. Finally, in practical application, there are many stochastic processes where only interval-valued samples can be obtained. We can look for more application scenarios for the belief entropy rate and further explore its properties.

Appendix A: WTI crude oil price data in February and March 2022

Table 7 WTI daily price data in February 2022

Date	Low price	High price	Average price
2.1	86.55	88.87	87.71
2.2	87.10	89.72	88.41
2.3	86.75	90.45	88.60
2.4	90.07	93.17	91.62
2.7	90.73	92.73	91.73
2.8	88.51	91.68	90.09
2.9	88.41	90.58	89.50
2.10	89.03	91.74	90.38
2.11	89.19	94.66	91.92
2.14	92.09	95.82	93.95
2.15	90.66	95.17	92.91
2.16	90.00	95.01	92.50
2.17	90.62	93.36	91.99
2.18	89.03	92.66	90.84
2.20	91.08	91.74	91.41
2.21	89.08	93.91	91.50
2.22	90.35	96.00	93.17
2.23	90.64	93.90	92.27
2.24	91.45	100.54	96.00
2.25	90.06	95.64	92.85

Table 8 WTI daily price data in March 2022

Date	Low price	High price	Average price
3.1	95.32	106.78	101.05
3.2	105.18	112.51	108.84
3.3	106.43	116.57	111.50
3.4	107.25	116.02	111.63
3.7	115.54	130.50	123.02
3.8	117.07	129.44	123.25
3.9	103.63	126.84	115.23
3.10	105.53	114.88	110.20
3.11	104.48	110.29	107.39
3.14	99.76	109.72	104.74
3.15	93.53	102.58	98.06
3.16	94.07	99.22	96.64
3.17	94.85	104.24	99.54
3.18	102.30	106.28	104.29
3.21	104.08	112.69	108.38
3.22	109.30	115.01	112.16
3.23	108.38	115.40	111.89
3.24	110.61	116.64	113.62
3.25	108.68	114.12	111.40
3.28	102.83	112.93	107.88

Acknowledgements The work is partially supported by National Natural Science Foundation of China (Grant No. 61973332). Thanks to Mr. Jixiang Deng in the Information Fusion and Intelligent System Laboratory for his help and support.

Author Contributions All authors contributed to the study conception and design. All authors performed material preparation, data collection and analysis. Zhiyuan Wang wrote the first draft of the paper. All authors contributed to the revisions of the paper. All authors read and approved the final manuscript.

Funding The work was partially supported by National Natural Science Foundation of China (Grant No. 61973332), JSPS Invitational Fellowships for Research in Japan (Short-term).

Data Availability The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

Code Availability The code of the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate This article does not contain any studies with human participants or animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

Consent for Publication The participant has consented to the submission of the case report to the journal.

Conflict of Interests All the authors certify that there is no conflict of interest with any individual or organization for this work.

References

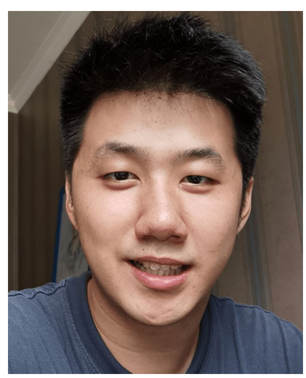
1. Shannon CE (2001) A mathematical theory of communication. ACM SIGMOBILE Mob Comput Commun Rev 5(1):3–55
2. Cheong KH, Saakian DB, Zadourian R (2017) Allison mixture and the two-envelope problem. Phys Rev E 96(6):062303
3. Wang C, Tan ZX, Ye Y, Wang L, Cheong KH, Xie N-G (2017) A rumor spreading model based on information entropy. Sci Rep 7(1):1–14
4. Cover TM (1999) Elements of information theory. John Wiley & Sons
5. Zhou L, Sood K, Xiang Y (2019) ERM: an accurate approach to detect DDoS attacks using entropy rate measurement. IEEE Commun Lett 23(10):1700–1703. https://doi.org/10.1109/LCOMM.2019.2931832. Conference Name: IEEE Communications Letters
6. Vegetabile BG, Stout-Oswald SA, Davis EP, Baram TZ, Stern HS (2019) Estimating the entropy rate of finite Markov chains with application to behavior studies. J Educ Behav Stat 44(3):282–308. https://doi.org/10.3102/1076998618822540. Accessed 15 Mar 2022
7. Wang W, Xiang D, Ban Y, Zhang J, Wan J (2017) Superpixel segmentation of polarimetric SAR images based on integrated distance measure and entropy rate method. IEEE J Sel Top Appl Earth Obs Remote Sens 10(9):4045–4058. https://doi.org/10.1109/JSTARS.2017.2708418. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
8. Feutrill A, Roughan M (2021) A review of Shannon and differential entropy rate estimation. Entropy 23(8):1046. https://doi.org/10.3390/e23081046. Accessed 14 Mar 2022. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute
9. Hu C, Hu ZH (2020) On statistics, probability, and entropy of interval-valued datasets. In: International conference on information processing and management of uncertainty in knowledge-based systems, Springer, pp 407–421
10. Dempster AP (2008) Upper and lower probabilities induced by a multivalued mapping. In: Classic works of the dempster-shafer theory of belief functions, Springer, pp 57–72
11. Shafer G (1976) A mathematical theory of evidence. vol 42 Princeton university press
12. Zhou M, Chen Y-W, Liu X-B, Cheng B-Y, Yang J-B (2020) Weight assignment method for multiple attribute decision making with dissimilarity and conflict of belief distributions. Comput Ind Eng 106648:147
13. Yager RR, Alajlan N (2015) Dempster-Shafer belief structures for decision making under uncertainty. Knowl-Based Syst 80:58–66. https://doi.org/10.1016/j.knosys.2014.12.031. Accessed 28 Jun 2022
14. Liu P, Li Y, Zhang X, Pedrycz W (2022) A multiattribute group decision-making method with probabilistic linguistic information based on an adaptive consensus reaching model and evidential reasoning. IEEE Trans Cybernet PP:1–15
15. Fei L, Wang Y (2022) An optimization model for rescuer assignments under an uncertain environment by using dempster-shafer theory. Knowl-Based Syst 255:109680
16. Chen X, Deng Y (2022) An evidential software risk evaluation model. Mathematics 10(13):10–339010132325
17. Huang L-Q, Liu Z-G, Dezert J (2021) Cross-domain pattern classification with distribution adaptation based on evidence theory. IEEE Trans Cybern :1–14
18. Bossé É, Solaiman B (2018) Fusion of information and analytics: a discussion on potential methods to cope with uncertainty in complex environments (big data and iot). Int J Digit Signals Smart Syst 2(4):279–316
19. Yang J-B, Xu D-L (2013) Evidential reasoning rule for evidence combination. Artif Intell 205:1–29
20. Deng X, Jiang W (2022) A framework for the fusion of non-exclusive and incomplete information on the basis of D number theory. Appl Intell :1010071048902203960
21. Liu Z, Zhang X, Niu J, Dezert J (2021) Combination of classifiers with different frames of discernment based on belief functions. IEEE Trans Fuzzy Syst 29(7):1764–1774. https://doi.org/10.1109/TFUZZ.2020.2985332
22. Liu Z-G, Qiu G-H, Wang S-Y, Li T-C, Pan Q (2021) A new belief-based bidirectional transfer classification method. IEEE Trans Cybernet
23. Song M, Sun C, Cai D, Hong S, Li H (2022) Classifying vaguely labeled data based on evidential fusion. Inf Sci 583:159–173
24. Xiao F, Pedrycz W (2022) Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification. IEEE Trans Pattern Anal Mach Intell :10–110920223167045
25. Pan L, Deng Y (2022) A new complex evidence theory. Inf Sci 608:251–261
26. Pan L, Deng Y (2022) Complex-valued rényi entropy. Commun Stat-Theory Methods :1–12
27. Deng Y (2022) Random permutation set. Int J Comput Commun Control 17(1):4542–0. https://doi.org/10.15837/ijccc.2022.1.4542
28. Deng J, Deng Y (2022) Maximum entropy of random permutation set. Soft Comput 26(21):11265–11275
29. Fagin R, Halpern JY (1990) A new approach to updating beliefs. In: Proceedings of the sixth annual conference on uncertainty in artificial intelligence, pp 347–374
30. Kulasekere EC, Premaratne K, Dewasurendra DA, Shyu M-L, Bauer PH (2004) Conditioning and updating evidence. Int J Approx Reason 36(1):75–108. https://doi.org/10.1016/j.ijar.2003.10.001. Accessed 31 Oct 2022
31. Su X, Li L, Qian H, Mahadevan S, Deng Y (2019) A new rule to combine dependent bodies of evidence. Soft Comput 23(20):9793–9799. https://doi.org/10.1007/s00500-019-03804-y. Accessed 11 May 2022
32. Dezert J, Tchamova A, Han D (2018) Total belief theorem and conditional belief functions. Int J Intell Syst 33(12):2314–2340
33. Martin A, Noûs C (2022) Belief functions on ordered frames of discernment. arXiv:2210.04535. http://arxiv.org/abs/2210.04535 Accessed 31 Oct 2022
34. U H (1982) Entropy with respect to Plausibility Measures. In: Proc. of 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982
35. Smets P (1983) Information content of an evidence. Int J Man-Mach Stud 19(1):33–43. https://doi.org/10.1016/S0020-7373(83)80040-6. Accessed 31 Oct 2022
36. Yager RR (2008) Entropy and specificity in a mathematical theory of evidence. In: Classic works of the dempster-shafer theory of belief functions, Springer, pp 291–310
37. Bouhamed SA, Kallel IK, Yager RR, Bossé É, Solaiman B (2020) An intelligent quality-based approach to fusing multi-source possibilistic information. Inf Fusion 55:68–90

38. Jousselme A-L, Liu C, Grenier D, Bosse E (2006) Measuring ambiguity in the evidence theory. IEEE Trans Syst Man Cybernet Part A Syst Hum 36(5):890–903. https://doi.org/10.1109/TSMCA.2005.853483. Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans
39. Wang X, Song Y (2018) Uncertainty measure in evidence theory with its applications. Appl Intell 48(7):1672–1688. https://doi.org/10.1007/s10489-017-1024-y. Accessed 08 Nov 2022
40. Jiroušek R, Shenoy PP (2018) A new definition of entropy of belief functions in the Dempster–Shafer theory. Int J Approx Reason 92:49–65. https://doi.org/10.1016/j.ijar.2017.10.010. Accessed 08 Nov 2022
41. Pal NR, Bezdek JC, Hemasinha R (1993) Uncertainty measures for evidential reasoning II: a new measure of total uncertainty. Int J Approx Reason 8(1):1–16. https://doi.org/10.1016/S0888-613X(05)80003-9. Accessed 31 Oct 2022
42. Deng Y (2020) Uncertainty measure in evidence theory. Sci China Inf Sci 63(11):210201
43. Zhou Q, Deng Y (2022) Fractal-based belief entropy. Inf Sci 587:265–282. https://doi.org/10.1016/j.ins.2021.12.032
44. Zhou Q, Deng Y (2022) Higher order information volume of mass function. Inf Sci 586:501–513. https://doi.org/10.1016/j.ins.2021.12.005
45. Yang Y, Han D (2016) A new distance-based total uncertainty measure in the theory of belief functions. Knowl-Based Syst 94:114–123. https://doi.org/10.1016/j.knosys.2015.11.014. Accessed 31 Oct 2022
46. Liao H, Ren Z, Fang R (2020) A deng-entropy-based evidential reasoning approach for multi-expert multi-criterion decision-making with uncertainty. Int J Comput Intell Syst 13:1281–1294. https://doi.org/10.2991/ijcis.d.200814.001
47. Buono F, Longobardi M (2020) A dual measure of uncertainty: the deng extropy. Entropy 22(5)
48. Balakrishnan N, Buono F, Longobardi M (2022) A unified formulation of entropy and its application. Phys A Stat Mech Appl 127214. https://doi.org/10.1016/j.physa.2022.127214
49. Abellán J (2017) Analyzing properties of deng entropy in the theory of evidence. Chaos, Solitons Fractals 95:195–199
50. He Z, Jiang W (2018) A new belief markov chain model and its application in inventory prediction. Int J Prod Res 56(8):2800–2817
51. Soubaras H (2010) On Evidential Markov Chains. In: Bouchon-Meunier B, Magdalena L, Ojeda-Aciego M, Verdegay J-L, Yager RR (eds) Foundations of Reasoning Under Uncertainty. Studies in Fuzziness and Soft Computing, Springer, pp 247–264. https://doi.org/10.1007/978-3-642-10728-3_13. Accessed 13 Jul 2022
52. Fouque L, Appriou A, Pieczynski W (2000) An evidential markovian model for data fusion and unsupervised image classification. In: Proceedings of the third international conference on information fusion, vol 1. IEEE, pp 4–25
53. Parkinson M (1980) The extreme value method for estimating the variance of the rate of return. J Bus 53(1):61–65. Publisher: University of Chicago Press. Accessed 2022-06-29

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Zhiyuan Wang is a Junior student at the Yingcai Honors college at the University of Electronic Science and Technology of China, Chengdu, China. He follows Prof. Deng to do research on evidence theory and he is interested in machine learning and uncertainty analysis.



Qianli Zhou is pursing his Ph.D degree in University of Electronic Science and Technology of China. He is visiting student in University of Alberta. His research interests is uncertain information processing on belief structure. He has published more than 10 papers on SCI-indexed journals, such as IEEE Trans. on SMC-A, Information Sciences, and Fractals, etc.

Yong Deng received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2003. From 2005 to 2011, he was an Associate Professor in the Department of Instrument Science and Technology, Shanghai Jiao Tong University. From 2010, he was a Professor in the School of Computer and Information Science, Southwest University, Chongqing, China. From 2012, he was a Visiting Professor in Vanderbilt University, Nashville, TN, USA. From 2016, he was also a Professor in School of Electronic and Information Engineering, Xi an Jiaotong University, Xian, China. From 2017 to now, he is the full professor of Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, China. He is the panel member of information division of NSFC. He is the JSPS Invitational Fellow in Japan Advanced Institute of Science and Technology, Ishikawa, Japan. Professor Deng has published more than 100 papers in referred journals. His research interests include uncertainty, quantum computation and complex system modelling. He presents generalized evidence theory, D numbers, Deng entropy, information volume of mass function and Random Permutation Set. He served as the PC member of many conferences such as International Conference on Belief Functions. He served as many editorial board members such as editorial board member of Entropy, Applied Intelligence, guest editor of IEEE trans. on fuzzy systems, International Journal of Approximate Reasoning, Sustainability, associate editor of Journal of Organizational and End User Computing. He served as a reviewer for more than 40 SCI-indexed journals. Professor Deng has received numerous honors and awards, including the Elsevier Highly Cited Scientist in China in 2014-2021, Clarivate Analytics Highly cited scholar in 2021.