Journal of Control and Decision

ISSN: 2330-7706 (Print) 2330-7714 (Online) Journal homepage: www.tandfonline.com/journals/tjcd20

Make evidence theory probabilistic again

Dong-Ling Xu, Jian-Bo Yang & Ying-Ming Wang

To cite this article: Dong-Ling Xu, Jian-Bo Yang & Ying-Ming Wang (22 May 2025): Make evidence theory probabilistic again, Journal of Control and Decision, DOI: 10.1080/23307706.2025.2495805

To link to this article: https://doi.org/10.1080/23307706.2025.2495805

Published online: 22 May 2025.

JOURNAL OF CONTROL AND DECISION
https://doi.org/10.1080/23307706.2025.2495805

Make evidence theory probabilistic again

Dong-Ling Xu[a], Jian-Bo Yang[a] and Ying-Ming Wang[b]

[a]Alliance Manchester Business School, The University of Manchester, Manchester, UK; [b]Department of Management Science and Engineering, School of Economics and Management, Fuzhou University, Fuzhou, People's Republic of China

ABSTRACT
In this paper, we prove from a new angle that Dempster's rule is inherently probabilistic, extends Bayes' rule and reduces to Bayes' rule when precise probabilities are available, regardless of whether prior is uniform. We use examples to demonstrate this equivalence. Additionally, we explain that the Evidential Reasoning (ER) rule is also probabilistic and includes Bayes' and Dempster's rules as special cases. Furthermore, we address some criticisms of the behaviour of Dempster's rule from a probabilistic perspective and explain the rationality of the behaviour. We also identify instances where such critiques were misapplied. Finally, we clarify vital differences between belief degrees in belief functions and basic probabilities and highlight the critical differences between Shafer's discounting method and the ER rule. These differences make the latter probabilistic, while the former is not. Our motivation is to show that evidence theory has a probabilistic foundation and is possible to become probabilistic again.

KEYWORDS
Evidence theory; evidential reasoning; probabilistic reasoning; Bayesian inference; inference with imperfect data; inference under uncertainty

1. Introduction

Evidence theory was originated from Arthur P. Dempster's work on upper and lower probabilities (Dempster, 1967, 1968) and was later formalised by Glenn Shafer in his book entitled 'A Mathematical Theory of Evidence' (Shafer, 1976). It is a framework for reasoning under uncertainty. It can assign probability masses to subsets of possible outcomes, allowing the representation of partial ignorance. It is widely applied in decision-making, sensor fusion and diagnostics, where data from multiple sources may be imprecise or unreliable, such as fault detection, medical diagnosis and machine learning, for combining evidence under uncertainty, especially in risk assessment and AI-based reasoning.

Rules for evidence combination are the core of evidence theory. The first and the foundation of all rules in evidence theory is Dempster's rule. It generalises Bayesian inference by handling imprecise probabilities and unreliable data sources (Shafer, 1976). It is asserted that Dempster's rule is Bayesian when all probabilities are assigned to singleton hypotheses (Pearl, 1990).

Shortly after Dempster proposed the rule, Shafer formalised the concept of belief functions and used belief degrees and a discounting method to discount evidence before combining the evidence, which leads to a combination rule that is referred to as Shafer's discounting method. While some scholars have argued that belief functions diverge from probability theory, others highlighted the strong links between the two frameworks (Liu & Hong, 2000).

Lotfi Zadeh (1979) criticised Dempster's rule for producing counterintuitive results when combining highly conflicting evidence. His well-known example involved two medical experts assigning a 1% probability to meningitis and 99% to different diagnoses. Dempster's rule produced 100% certainty for meningitis, despite individual experts' low initial probability for meningitis. This critique sparked extensive debate, leading to the development of alternative combination rules and methods to address conflicting evidence (Haenni, 2005). As far as we know, all of those rules, excluding Dempster's rule, are non-probabilistic. So far, the only rules for evidence combination that are claimed to be probabilistic are Bayes' rule, Dempster's rule and the Evidential Reasoning rule (Yang and Xu, 2013), the latter two of which use Bayesian inference as their foundation.

In addition to Zadeh's critique, other scholars have also questioned the link between Dempster's rule and Bayes' rule, and the rationality of the behaviour of Dempster's rule, such as Dezert and Tchamova (2011), Dezert, Tchamova, Han, and Tacnet (2013) and Cheng et al. (1988). However, our research has demonstrated that Dempster's rule is probabilistic, serving as a natural extension to Bayes' rule, and when precise probabilities are available, Dempster's rule converges to Bayes' rule. In this paper, we will use examples to demonstrate that some of the critiques are not correct, Dempster's rule is probabilistic, and evidence theory has the foundation to go back to being probabilistic.

CONTACT Dong-Ling Xu Ling.Xu@Manchetser.ac.uk Alliance Manchester Business School, The University of Manchester, Manchester, M15 6PB, UK

The main contribution of this paper is to reinforce the probabilistic foundation of evidence theory by proving that Dempster's rule extends Bayes' rule and is equivalent to it whenever Bayes' rule is applicable. The paper also addresses the criticisms of Dempster's rule, explaining its rationality and identifying instances where such criticisms were misapplied. Additionally, it clarifies key differences between belief degrees in belief functions and basic probabilities, emphasising that only basic probabilities should be used for evidence combinations. Furthermore, it highlights the subtle but critical difference between Shafer's discounting method and the ER rule, which makes the ER rule probabilistic, whereas Shafer's discounting is not. We argue that the future development of evidence theory should follow a probabilistic direction, one that is principled, rigorous, and interpretable. Following this direction will ensure that its reasoning outcomes in big data analysis and explainable artificial intelligence are rational, reliable, interpretable and trustworthy.

2. Preliminaries

2.1. Common objective of Bayes inference and evidence theory: judging hypotheses

Both Bayesian inferences and evidence theory aim to address a fundamental practical problem: determining which hypothesis among a set of competing alternatives is true and to what extent, based on available evidence. This problem has a wide range of presence in fields such as multi-criteria decision-making, machine learning, artificial intelligence, large language models, and other decision support systems. Despite their shared purpose, they model the problem differently, with distinct theoretical foundations. This section focuses on demonstrating Bayesian inference and Dempster's evidence combination processes and exploring their relationships.

The practical goal of these methods is to make judgments about which hypothesis is true in a set of mutually exclusive and collectively exhaustive hypotheses, denoted as

$$\Theta = \{H_1, H_2, \ldots, H_N\}, \tag{1}$$

with $H_i \cap H_j = \emptyset$ for any $i, j \in \{1, \cdots, N\}$ and $i \neq j$ where $\emptyset$ is an empty set and Θ is referred to as a frame of discernment.

2.2. Assumptions in Bayesian inference

In Bayesian inference, probabilities are assumed to satisfy the following relationship:

$$p^t_{H_i} + p^f_{H_i} = 1, \quad \forall H_i \in \Theta, \tag{2}$$

where

- $p^t_{H_i}$: The probability that hypothesis H_i is true,
- $p^f_{H_i}$: The probability assigned to the negation (i.e. H_i being not true).

Bayesian inference is grounded in the use of Bayes' Rule to update probabilities given new evidence. This approach assumes precise probability values and a closed-world perspective, where the hypotheses completely describe the space of possibilities.

Another implied assumption is that all probabilities are regarded as correct and fully reliable. Later in Shafer's discounting method and the Evidential Reasoning rule, we will take into account the reliability of evidence explicitly.

2.3. Assumptions in evidence theory

Before we discuss the assumptions made in evidence theory, we need to introduce a few key concepts.

2.3.1. Power set (Shafer, 1976)

Evidence theory extends Bayesian approach to accommodate imprecise probabilities. That is, evidence theory allows probability to be allocated not only to single hypotheses but also to subsets of hypotheses.

For the frame of discernment defined by Equation (1), all possible subsets of Θ form a so-called power set of Θ, denoted as 2^Θ or $\mathrm{P}(\Theta)$, consisting of the following 2^N subsets of Θ, or

$$2^\Theta = \mathrm{P}(\Theta) = \{\emptyset, H_1, \cdots, H_N, \{H_1, H_2\}, \cdots, \times \{H_1, H_N\}, \cdots, \{H_1, \cdots, H_{N-1}\}, \Theta\} \tag{3}$$

In this paper, we let θ denote a subset of Θ, i.e. $\theta \subseteq \Theta$, or θ is any one of the elements in the power set shown in Equation (3), i.e. $\theta \in \mathrm{P}(\Theta)$. To differentiate it from a singleton hypothesis only, θ is referred to as a proposition in subsequent sections.

When probability is allowed to be allocated to subsets of hypotheses, it becomes essential to distinguish and track which portions of the probability are divisible among smaller elements of the subsets and which are not. Non-separable or indivisible probability is referred to as Basic Probability Mass or Basic Probability Assignment *(bpa)* by Dempster (1967, 1968, 2008), which is different from Shafer's belief degrees in a belief function that is separable and divisible. We will give an example to illustrate the important differences among these concepts in Section 2.3.5.

2.3.2. Basic probability mass in evidence theory

In evidence theory, basic probability, basic probability mass or basic probability assignment (bpa) are different names for the same concept. It is denoted by $m(\theta)$, representing the probability specifically assigned to a subset $\theta \subseteq \Theta$, where Θ is the frame of discernment.

It cannot be further divided among the individual elements in θ or any subset of θ. It is a function $m : 2^{\Theta} \to [0, 1]$, satisfying (Dempster 1967, 1968, 2008)

$$m(\emptyset) = 0, \quad 0 \le m(\theta) \le 1 \quad \text{and} \quad \sum_{\theta \subseteq \Theta} m(\theta) = 1 \tag{4}$$

It is important to note that in the evidential reasoning (ER) rule (Yang and Xu, 2013, 2025), basic probability $p(\theta)$and basic probability mass $m(\theta)$ are different, and a more detailed explanation is given in Section 2.3.6.

2.3.3. Belief degree in Shafer's belief function

To indicate the total support for $\theta \subseteq \Theta$, Shafer (1976) introduced the concept of belief function to complement Dempster's basic probability mass function. In belief function, belief degree in θ, denoted as $\mathrm{Bel}(\theta)$, represents the total basic probabilities committed to θ, including those committed to all its subsets $B \subseteq \theta$. It is computed by

$$\mathrm{Bel}(\theta) = \sum_{B \subseteq \theta} m(B) \tag{5}$$

The important difference between the basic probability mass assigned to θ and the belief degree of θ in a belief function is that the former cannot be divided among other smaller subsets in θ while the latter is the sum of the basic probability mass assigned to θ and its smaller subset B for any $B \subseteq \theta$.

Generally, in literature about probability theory, belief degrees, probability and basic probability mass usually mean the same thing – basic probability, or the indivisible type of probability. Indeed, they are the same thing when we consider only the traditional probabilities that are assigned to the singleton hypothesis only. To avoid confusion, Dempster used basic probability to compute a triplet $(p_\theta^t, p_\theta^f, p_\theta^u)$ for explaining the uncertainty of θ as discussed in the following subsection.

2.3.4. Assumptions in evidence theory: an extension to Bayesian inference

In evidence theory, the assumption in Bayesian Inference is relaxed to the following less restrictive condition (Dempster, 1967, 1968, 2008, 2015):

$$p_\theta^t + p_\theta^f \le 1, \tag{6}$$

or

$$p_\theta^t + p_\theta^f + p_\theta^u = 1, \tag{7}$$

where

- p_θ^t: Probability of θ being true,
- p_θ^f: Probability of θ being false (not true) and
- p_θ^u: Probability of θ being in unknown states.

Those three probabilities in Equation (7) are referred to as a triplet by Dempster. As p_θ^t is defined as the probability of θ being true, it represents the total basic probabilities committed to θ, including those committed to all its subsets $B \subseteq \theta$. This definition is the same as the definition of belief degree in θ in Shafer's belief function, i.e. $p_\theta^t = \mathrm{Bel}(\theta) = \sum_{B \subseteq \theta} m(B)$. It should be noted that p_θ^t is different from $m(\theta)$, the basic probability mass assigned to θ; $m(\theta)$ is the probability assigned exactly to θ and cannot be further divided to be assigned to any subsets of θ. There is always the relationship that $p_\theta^t \ge m(\theta)$ because p_θ^t contains $m(\theta)$ and all other basic probability masses assigned to all the subsets of θ.

2.3.5. Numerical example to illustrate the above basic concepts

The following numerical examples show some of the key concepts and the links among basic probability mass, belief degree $\mathrm{Bel}(\theta)$ and p_θ^t. Let $\Theta = \{H_1, H_2, H_3\}$, where H_1, H_2 and H_3 are three hypotheses. Suppose the basic probability masses are given by

$$\begin{aligned}
m(\{H_1\}) &= 0.1,\ m(\{H_2\}) = 0.2,\\
m(\{H_3\}) &= 0.2,\ m(\{H_1, H_2\}) = 0.1,\\
m(\{H_1, H_3\}) &= 0.2,\ m(\{H_2, H_3\}) = 0.1,\\
m(\{H_1, H_2, H_3\}) &= 0.1.
\end{aligned}$$

Power set $2^{\Theta} = \mathrm{P}(\Theta) = \{\emptyset, H_1, H_2, H_3, \{H_1, H_2\}, \{H_1, H_3\}, \{H_2, H_3\}, \{H_1, H_2, H_3\}\}$.

Belief degrees are computed using Equation (5) as follows:

$$\begin{aligned}
\mathrm{Bel}(\{H_1\}) &= m(\{H_1\}) = 0.1;\\
\mathrm{Bel}(\{H_2\}) &= m(\{H_2\}) = 0.2\\
\mathrm{Bel}(\{H_3\}) &= m(\{H_3\}) = 0.2\\
\mathrm{Bel}(\{H_1, H_2\}) &= m(\{H_1\}) + m(\{H_2\})\\
&\quad + m(\{H_1, H_2\})\\
&= 0.1 + 0.2 + 0.1 = 0.4\\
\mathrm{Bel}(\{H_1, H_3\}) &= m(\{H_1\}) + m(\{H_3\})\\
&\quad + m(\{H_1, H_3\})\\
&= 0.1 + 0.2 + 0.2 = 0.5\\
\mathrm{Bel}(\{H_2, H_3\}) &= m(\{H_2\}) + m(\{H_3\})\\
&\quad + m(\{H_2, H_3\})\\
&= 0.2 + 0.2 + 0.1 = 0.5\\
\mathrm{Bel}(\{H_1,\ H_2, H_3\}) &= m(\{H_1\}) + m(\{H_2\}) + m(\{H_3\})\\
&\quad + m(\{H_1, H_2\}) + m(\{H_1, H_3\})\\
&\quad + m(\{H_2, H_3\} + m(\{H_1, H_2, H_3\}))\\
&= 0.1 + 0.2 + 0.2 + 0.1\\
&\quad + 0.2 + 0.1 + 0.1 = 1
\end{aligned}$$

Dempster's triplet in Equation (7) are given as follows:

$$p^t_{H_1} = \mathrm{Bel}(\{H_1\}) = 0.1;$$

$$p^f_{H_1} = m(\{H_2\}) + m(\{H_3\}) + m(\{H_2, H_3\}) = 0.5;$$

$$p^u_{H_1} = 0.4.$$

$$p^t_{H_2} = \mathrm{Bel}(\{H_2\}) = 0.2;$$

$$p^f_{H_2} = m(\{H_1\}) + m(\{H_3\}) + m(\{H_1, H_3\}) = 0.5;$$

$$p^u_{H_2} = 0.3.$$

$$p^t_{H_3} = \mathrm{Bel}(\{H_3\}) = 0.2;$$

$$p^f_{H_3} = m(\{H_1\}) + m(\{H_2\}) + m(\{H_1, H_2\}) = 0.4;$$

$$p^u_{H_3} = 0.4.$$

$$p^t_{(H_1,H_2)} = \mathrm{Bel}(\{H_1, H_2\}) = 0.4;$$

$$p^f_{(H_1,H_2)} = m(\{H_3\}) = 0.2;$$

$$p^u_{(H_1,H_2)} = 0.4.$$

$$p^t_{(H_1,H_3)} = \mathrm{Bel}(\{H_1, H_3\}) = 0.5;$$

$$p^f_{(H_1,H_3)} = m(\{H_2\}) = 0.2;$$

$$p^u_{(H_1,H_3)} = 0.3.$$

$$p^t_{(H_2,H_3)} = \mathrm{Bel}(\{H_2, H_3\}) = 0.5;$$

$$p^f_{(H_2,H_3)} = m(\{H_1\}) = 0.1;$$

$$p^u_{(H_2,H_3)} = 0.4.$$

$$p^t_{(H_1,H_2,H_3)} = \mathrm{Bel}(\{H_1, H_2, H_3\}) = 1;$$

$$p^f_{(H_1,H_2,H_3)} = 0;$$

$$p^u_{(H_1,H_2,H_3)} = 0.$$

2.3.6. Roles of basic probability, basic probability mass and belief degree in evidence combination

In Bayesian inference, basic probability, basic probability mass and belief degree all reduce to the same concept, $p(H_i)$, the probability that hypothesis H_i is true. Equations (8) and (9) give such examples.

In Dempster's terminology, probability mass $m(\theta)$ and basic probability $p(\theta)$ are the same, or $p(\theta) = m(\theta)$. Equations (12), (13) and (14) provide such examples. Dempster has not used the term belief function in his papers but instead stated (Dempster, 2008) that

> The set function p^t_θ was called a belief function by Shafer, and DS theory itself is often called the theory of belief functions. The term belief function is unnecessarily formal, however, and has led to many misperceptions. In fact, p^t_θ is a mainline successor to ordinary textbook probability, principally designed to allow you to assign a non-zero probability to 'don't know'.

In the evidential reasoning (ER) rule, basic probability $p(\theta)$and basic probability mass $m(\theta)$ are different. The basic probability in the ER rule is the same as that in Dempster's rule. However, basic probability mass in the ER rule is equal to basic probability weighted by evidence weight, i.e. $m(\theta) = wp(\theta)$ where $0 \le w \le 1$ is evidence weight. In Dempster's rule, evidence weight is assumed to be one, or $w = 1$. This explains why basic probability mass $m(\theta)$ and basic probability $p(\theta)$are the same in Dempster's rule.

Shafer intended to extend belief degree to power-set and invented belief function. However, using belief functions to compute belief degrees can lead to a non-probabilistic inference process if the overlapping portions of probability masses are not properly separated. To make the inference process probabilistic, it is essential to use only basic probabilities masses directly, just as Dempster rightly asserted (Dempster, 2015):

> Among the four set functions that specify mass $m(\theta)$, p^t_θ, p^f_θ and p^u_θ, only the mass set function takes the familiar mathematical form of an ordinary probability distribution. When reduced to expressions in terms of masses, which is always possible, the operations and computations of the ECP (Extended Calculus of Probability) are abstractly equivalent to operations and computations with OCP (Ordinary Calculus of Probability) models.

In short, any probabilistic operation should only be conducted on the basic probability function or basic probability mass function.

2.4. Bayesian inference

2.4.1. Mathematical representation of evidence in Bayesian inference

Bayesian inference calculates the posterior probability of a hypothesis after an event is observed. It requires a known prior distribution of the hypotheses and the likelihood of observing the event given that each of the hypotheses is true. Using the language of evidence theory, in Bayesian inference, evidence is phrased as prior distribution (as the old knowledge, denoted as e_1) and likelihood function (as the new knowledge after an event, E, is observed).

A prior distribution, e_1, can be written as

$$e_1 = \{(H_1, p_{11}), (H_2, p_{12}), \ldots, (H_N, p_{1N})\}, \quad (8)$$

Or simply

$$e_1 = \{p(H_i), i = 1, \ldots, N\}, \quad (9)$$

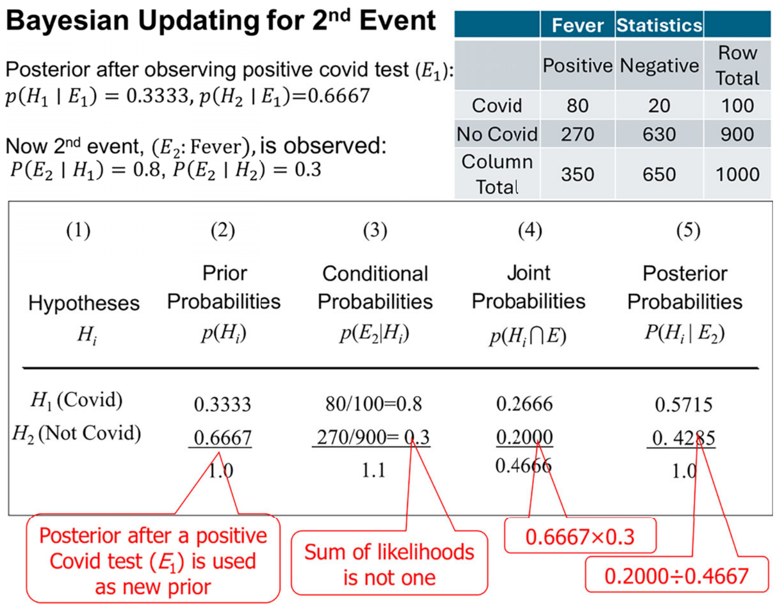

Figure 1. Bayesian updating after the observation of the second event.

where $p_{1i} = p(H_i)$ represents the probability or degree of belief assigned to hypothesis H_i, and $\sum_{i=1}^{N} p(H_i) = 1$.

Likelihood function is written as $p(E|H_i)$ and $p(E|\neg H_i)$, the likelihood of observing E if H_i is true and not true, respectively. In Bayesian inference, normally

$$p(E|H_i) + p(E|\neg H_i) \neq 1. \tag{10}$$

In medical test case studies, we are normally given the sensitivity and the specificity of a test, such as a rapid antigen Covid-19 test and PCR test. Such information is essentially likelihood function as shown in the example of Section 3.2. Other methods for obtaining the necessary information to conduct Bayesian inference are to collect data, as shown in Figure 1. From the data table in the figure, we can calculate prior and likelihoods for Bayesian inference.

2.4.2. Bayes' rule for evidence updating

In Bayesian inference, the probability in a hypothesis is updated when new evidence becomes available, using Bayes' Rule:

$$p(H_i|E) = \frac{p(E|H_i)p(H_i)}{p(E)}, \tag{11}$$

where

- $p(H_i|E)$: Posterior probability of hypothesis H_i given evidence E,
- $p(E|H_i)$: Likelihood of observing E if H_i is true,
- $p(H_i)$: Prior probability of H_i being true,
- $p(E) = \sum_{j=1}^{n} p(E|H_j)p(H_j)$: Total probability of observing E under all hypotheses.

Bayes' approach assumes precise prior probabilities and is widely used in machine learning and probabilistic inference.

2.5. Dempster's rule for evidence combination

2.5.1. Mathematical representation of evidence

In evidence theory, in general, the ith piece of evidence e_i can be profiled by a basic probability mass function, defined as follows

$$e_i = \left\{ (\theta, p_{\theta,i}), \forall \theta \subseteq \Theta, \sum_{\theta \subseteq \Theta} p_{\theta,i} = 1 \right\} \tag{12}$$

where $(\theta, p_{\theta,i})$ is an element of evidence e_i, representing that the evidence points to proposition θ to the degree of $p_{\theta,i}$. If $p_{\theta,i} > 0$, θ is referred to as a focal element of e_i. If $\theta = \Theta$, $p_{\theta,i}$ is referred to as a degree of global ignorance. If $\theta \neq \Theta$ and $\theta \neq H_i$, ($i = 1, \ldots, N$), i.e. when a focal element is not the frame of discernment or any singleton hypothesis, such elements θ are said to represent some local ignorance.

2.5.2. Dempster's rule of evidence combination

Dempster's Rule of Evidence Combination is used to combine evidence acquired from independent sources. If two mass functions m_1 and m_2 are provided over the same frame of discernment, their combination m is given by

$$m(\theta) = \frac{\sum_{B \cap C = \theta} m_1(B)m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)m_2(C)} \tag{13}$$

where

$$m_i(\theta) = p_{\theta,i} \tag{14}$$

is the probability mass assigned to θ indicating the degree of support from evidence e_i, and

$$K = 1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{15}$$

accounts for the total conflict in evidence because each term in $\sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ indicates a disagreement between the two pieces of evidence. K is also called a normalisation factor to make sure that $\sum_{\theta \subseteq \Theta} m(\theta) = 1$.

Dempster's rule allows basic probability (or basic probability mass) to be assigned to subsets of hypotheses, $\theta \subseteq \Theta$, representing global ignorance if $\theta = \Theta$ and local ignorance if $\theta \subset \Theta$ and θ is not any singleton hypothesis. Basic probability assigned to local or global ignorance is referred to as unknown probability.

3. Equivalence between Bayes' rule and Dempster's rule when there is no ignorance

When there is ignorance in the basic probability assignments, Bayes' rule cannot be applied for evidence updating, whereas Dempster's rule can. Therefore, Bayes' rule has a narrower scope of applicability than Dempster's rule. In this sense, we can see that Dempster's rule extends Bayes' rule and that the two rules are not equivalent when there is ignorance.

When θ is singleton hypothesis only, we have proved that Dempster's rule is equivalent to Bayes' rule (Yang and Xu, 2014, 2025; Yang et al., 2023) if and only if we use prior distribution as the very first piece of evidence and then use normalised likelihood as the new piece of evidence for every new observation. In the next section, we will prove this equivalence from a new perspective for situations where multiple events are observed, the precise prior distribution is known and the precise posterior probabilities of individual events are each given as well.

Dempster has always insisted that his rule of evidence combination is an extension of Bayes' rule. Pearl also asserted this claim (1990). Pearl's exact wording is as follows:

> Given two belief functions Bel_1 and Bel_2, their orthogonal sum $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$, also known as Dempster's rule of combination, is defined by their associated probability assignments:
>
> $$(m_1 \oplus m_2)(A) = K \sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)$$
>
> where K is a normalization constant:
>
> $$K^{-1} = 1 - \sum_{A_1 \cap A_2 = \emptyset} m_1(A_1) m_2(A_2)$$
>
> … …
>
> A belief function is called additive or Bayesian if each of its focal elements is a singleton, that is, an elementary event or a possible world. Bayesian belief $\mathrm{Bel}(A) = \mathrm{Pl}(A) = 1 - \mathrm{Bel}(\neg A)$. If Bel_1 is Bayesian, then $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ is also Bayesian, and Dempster's conditioning reduces to ordinary Bayesian conditioning. (Shafer [1976])

Although both Dempster and Pearl, and some other scholars, have asserted that Dempster's rule reduces to Bayes' rule when all probabilities are assigned to singleton hypotheses only, none of them has demonstrated the equivalence using examples or clarified the process of how the equivalence can be established, until 2023 when Yang et al. (2023) proved the necessary and sufficient condition for the two to be equivalent. That is, when Dempster's rule is applied, the prior distribution should be used as the first piece of evidence, and in subsequent evidence combinations, any other evidence to be combined should be acquired as normalised likelihoods with the prior no longer taken into account in evidence combination.

In Yang et al. (2023), the equivalence condition was established for situations where prior and likelihood probabilities are known. In the next subsection, we prove the equivalence for situations where the priors and posterior probabilities of individual events are known. The rationale behind the new proof is to show how to get rid of the effect of the prior which is embedded in each individual event's posterior distribution and why we need to do so. In fact, in Bayesian inference, prior distribution is used only once at the beginning of the inference process. In subsequent evidence updating, only likelihoods of new events are taken into account. Similarly, in Dempster's evidence combination, if we let the prior distribution be the first piece of evidence, and subsequent evidence is profiled as posterior probabilities, without getting rid of the effect of the prior from posterior probabilities, the prior's effect will be double counted every time we combine new evidence with the old one. The following equivalence theorem is established to show how to get rid of the effect of the prior in evidence combination.

3.1. Equivalence theorem and proof

Dempster's rule and Bayes' rule are equivalent in conditions where both rules are applicable. More precisely, suppose there are N hypotheses (or states) in the frame of discernment, $\Theta = \{H_1, H_2, \ldots, H_N\}$, and L events observed, $E_1, \cdots, E_L$, which occur independently of each other. The prior distribution of the hypotheses is given as follows:

$$e_0 = \{(H_1, p(H_1)), (H_2, p(H_2)), \ldots, (H_N, p(H_N))\},$$

and the individual posterior after the observation of each event, E_l, is given and denoted as $p_l(H_i|E_l)$. Then the joint *posterior probabilities, $p(H_i|E_1 \cap \cdots \cap E_L)$, of hypothesis $H_i(\ i = 1, \ldots, N)$ being true generated by applying Bayes' Theorem are the same as those generated by applying Dempster's rule if and only if the pieces of evidence to be combined are the prior distribution and the normalised likelihoods of each individual event.*

Proof: Let's denote

- $p(E_1 \cap \cdots \cap E_L|H_i)$ as the joint likelihood of observing all the L events $E_1 \cap \cdots \cap E_L$ given that hypothesis H_i is true
- $p_l(E_l|H_i)$ as the likelihood that each individual event, $E_l(l = 1, 2, \ldots, L)$, is observed given that H_i is true.

If the prior e_0 and the individual posterior $p_l(H_i|E_l)$ are given, then from Bayes' Theorem $p_l(H_i|E_l) = \frac{p_l(E_l|H_i)p_l(H_i)}{p_l(E_l)}$, we can compute the individual likelihood $p_l(E_l|H_i)$as follows:

$$p_l(E_l|H_i) = \frac{p_l(H_i|E_l)p_l(E_l)}{p_l(H_i)} \tag{16}$$

Here, $p_l(H_i)$ is a prior associated with the population from which event E_l is observed. If all events are observed from the same population, then the individual prior takes the common prior, i.e. $p_l(H_i) = p(H_i)$.

The term $p_l(E_l)$ is the total probability of observing the event E_l, which can be calculated from the formula $p_l(E_l) = \sum_{i=1}^{N} p_l(E_l|H_i)p_l(H_i)$. From the assumption that the L events $E_1, \cdots, E_L$ occur independently of each other, we have

$$p(E_1 \cap \cdots \cap E_L|H_i) = \prod_{l=1}^{L} p_l(E_l|H_i), \tag{17}$$

After taking into account all the L observations $E_1, \cdots, E_L$, the posterior probability of hypothesis H_i being true is generated as follows according to Bayes' Theorem:

$$\begin{aligned} p(H_i|E_1 \cap \cdots \cap E_L) &= \frac{p(E_1 \cap \cdots \cap E_L|H_i)p(H_i)}{p(E_1 \cap \cdots \cap E_L)} \\ &= \frac{p(H_i)\prod_{l=1}^{L} p_l(E_l|H_i)}{p(E_1 \cap \cdots \cap E_L)} \\ &= \frac{p(H_i)\prod_{l=1}^{L} p_l(E_l|H_i)}{\sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L} p_l(E_l|H_j)\right)} \end{aligned} \tag{18}$$

where $p(E_1 \cap \cdots \cap E_L) = \sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L} p_l(E_l|H_j)\right)$ because there must be $\sum_{i=1}^{N} p(H_i|E_1 \cap \cdots \cap E_L) = 1$.

Now let's substitute Equation (16) into Equation (18) so that we can make use of the known individual posteriors. We then have

$$\begin{aligned} p(H_i|E_1 \cap \cdots \cap E_L) &= \frac{p(H_i)\prod_{l=1}^{L} \frac{p_l(H_i|E_l)p_l(E_l)}{p_l(H_i)}}{p(E_1 \cap \cdots \cap E_L)} \\ &= \frac{p(H_i)\prod_{l=1}^{L} \frac{p_l(H_i|E_l)}{p_l(H_i)}}{\frac{p(E_1 \cap \cdots \cap E_L)}{\prod_{l=1}^{L} p_l(E_l)}} \\ &\qquad \forall i = 1, \cdots N, \end{aligned} \tag{19}$$

In Equation (19), we do not have all the information to compute $p(H_i|E_1 \cap \cdots \cap E_L)$, but we know there must be $\sum_{j=1}^{N} p(H_j|E_1 \cap \cdots \cap E_L) = 1$. Therefore, we have

$$\begin{aligned} &\sum_{j=1}^{N} \frac{p(H_j)\prod_{l=1}^{L} \frac{p_l(H_j|E_l)}{p_l(H_j)}}{\frac{p(E_1 \cap \cdots \cap E_L)}{\prod_{l=1}^{L} p_l(E_l)}} \\ &= \frac{1}{\frac{p(E_1 \cap \cdots \cap E_L)}{\prod_{l=1}^{L} p_l(E_l)}} \sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L} \frac{p_l(H_j|E_l)}{p_l(H_j)}\right) = 1 \end{aligned} \tag{20}$$

which means

$$\frac{p(E_1 \cap \cdots \cap E_L)}{\prod_{l=1}^{L} p_l(E_l)} = \sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L} \frac{p_l(H_j|E_l)}{p_l(H_j)}\right) \tag{21}$$

Now substituting Equations (21) to (19), we then have for any hypothesis $H_i, i = 1, \cdots, N$

$$p(H_i|E_1 \cap \cdots \cap E_L) = \frac{p(H_i)\prod_{l=1}^{L} \frac{p_l(H_i|E_l)}{p_l(H_i)}}{\sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L} \frac{p_l(H_j|E_l)}{p_l(H_j)}\right)} \tag{22}$$

As every element on the right-hand side of Equation (22) is known, we can use the equation to obtain the joint posterior for every H_i given common prior and the individual priors and posteriors for each event.

When $p_l(H_i) = p(H_i) \quad \forall i = 1, \cdots N,$ meaning every event is observed from the same population with the same prior, Equation (22) becomes

$$p(H_i|E_1 \cap \cdots \cap E_L) = \frac{p(H_i)\prod_{l=1}^{L} \frac{p_l(H_i|E_l)}{p(H_i)}}{\sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L} \frac{p_l(H_j|E_l)}{p(H_j)}\right)} \tag{23}$$

Equation (22) or (23) is the formula for computing the joint posterior when the common prior and the individual priors and posteriors are known.

Next, we show the equivalence between Bayesian updating and Dempster's combination. In this case, all probabilities are precisely known and assigned to individual hypotheses. Suppose the pieces of evidence to be combined are expressed as the prior distribution and the L normalised likelihoods for the L individual events,

$$e_0 = \{(H_1, p(H_1)), (H_2, p(H_2)), \ldots, (H_N, p(H_N))\},$$

$$e_1 = \left\{ \left(H_1, \frac{p_1(E_1|H_1)}{\sum_{k=1}^{N} p_1(E_1|H_k)} \right), \times \left(H_2, \frac{p_1(E_1|H_2)}{\sum_{k=1}^{N} p_1(E_1|H_k)} \right), \ldots, \times \left(H_N, \frac{p_1(E_1|H_N)}{\sum_{k=1}^{N} p_1(E_1|H_k)} \right) \right\},$$

$$\ldots$$

$$e_L = \left\{ \left(H_1, \frac{p_L(E_L|H_1)}{\sum_{k=1}^{N} p_L(E_L|H_k)} \right), \times \left(H_2, \frac{p_L(E_L|H_2)}{\sum_{k=1}^{N} p_L(E_L|H_k)} \right), \ldots, \times \left(H_N, \frac{p_L(E_L|H_N)}{\sum_{k=1}^{N} p_L(E_L|H_k)} \right) \right\}$$

Then by applying Dempster's rule to combine the $L+1$ pieces of evidence, we have

$$p(H_i|e_0 e_1 \cdots e_L) = \frac{p(H_i) \prod_{l=1}^{L} \frac{p_l(E_l|H_i)}{\sum_{k=1}^{N} p_l(E_l|H_k)}}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} \frac{p_l(E_l|H_j)}{\sum_{k=1}^{N} p_l(E_l|H_k)} \right)} \tag{24}$$

Therefore, to prove that Bayes updating is equivalent to Dempster's evidence combination, we need only to prove that the right-hand side of Equation (18) equals the right-hand side of Equation (24) or

$$\frac{p(H_i) \prod_{l=1}^{L} p_l(E_l|H_i)}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} p_l(E_l|H_j) \right)} = \frac{p(H_i) \prod_{l=1}^{L} \frac{p_l(E_l|H_i)}{\sum_{k=1}^{N} p_l(E_l|H_k)}}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} \frac{p_l(E_l|H_j)}{\sum_{k=1}^{N} p_l(E_l|H_k)} \right)} \tag{25}$$

Note that $\prod_{l=1}^{L} \sum_{k=1}^{N} p_l(E_l|H_k)$ is a term that does not change with respect to any particular event E_l or any particular hypothesis H_i. Suppose it is not zero so that it can be multiplied or divided to both the numerator and denominator of the right-hand side of Equation (18) without changing its value, leading to Equation (18) being equivalently re-written as follows:

$$p(H_i|E_1 \cap \cdots \cap E_L) = \frac{p(H_i) \prod_{l=1}^{L} p_l(E_l|H_i)}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} p_l(E_l|H_j) \right)}$$

$$\begin{aligned} p(H_i|E_1 \cap \cdots \cap E_L) &= \frac{p(H_i) \prod_{l=1}^{L} p_l(E_l|H_i)}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} p_l(E_l|H_j) \right)} \\ &= \frac{p(H_i) \prod_{l=1}^{L} p_l(E_l|H_i)}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} p_l(E_l|H_j) \right)} \times \frac{1/\prod_{l=1}^{L} \sum_{k=1}^{N} p_l(E_l|H_k)}{1/\prod_{l=1}^{L} \sum_{k=1}^{N} p_l(E_l|H_k)} \\ &= \frac{\frac{p(H_i) \prod_{l=1}^{L} p_l(E_l|H_i)}{\prod_{l=1}^{L} \sum_{k=1}^{N} p_l(E_l|H_k)}}{\sum_{j=1}^{N} \left(\frac{p(H_j) \prod_{l=1}^{L} p_l(E_l|H_j)}{\prod_{l=1}^{L} \sum_{k=1}^{N} p_l(E_l|H_k)} \right)} \\ &= \frac{p(H_i) \prod_{l=1}^{L} \frac{p_l(E_l|H_i)}{\sum_{k=1}^{N} p_l(E_l|H_k)}}{\sum_{j=1}^{N} \left(p(H_j) \prod_{l=1}^{L} \frac{p_l(E_l|H_j)}{\sum_{k=1}^{N} p_l(E_l|H_k)} \right)}. \end{aligned} \tag{26}$$

We can see that the first two terms of Equation (26) are the same as Equation (18) and the last term of Equation (26) is the same as the right-hand side of Equation (24). Therefore, we have shown that Equation (25) holds. End of Proof. ■

In the next two subsections, we will use examples to illustrate the equivalence of Bayesian inference and Dempster's evidence combination when Bayes' rule is applicable, assuming that prior distribution is not uniform, and prior and likelihoods of observed events are known.

3.2. Numerical examples of applying Bayes' rule

Example 3.1: Suppose in an area 10% of people are having Covid-19. A rapid test has 90% sensitivity and 80% specificity. If a person is tested positive, what is his probability of having Covid-19?

To apply Bayes' rule, we need to use Equation (11) to compute the posterior probability $p(H_1|E_1)$that a person has Covid-19 (H_1) given that the person's test result is positive (E_1).

Let

- H_1 be the hypothesis that the person has Covid-19, and
- H_2 the hypothesis that the person does not have Covid-19.
- Sensitivity is defined as the probability of a positive test result if the person has Covid-19, i.e. $p(E_1|H_1) =$ sensitivity $= 0.9$.
- Specificity is defined as the probability of a negative test if the person does not have Covid-19, i.e. specificity $= p(\text{Not } E_1|H_2)$. Given that the person does not have Covid-19, the probability of the person's test being positive is therefore $p(E_1|H_2) = 1 -$ Specificity $= 0.2$.
- Prior probabilities are given as:
  - ○ $p(H_1) = 0.1$ (The prior probability that the person has Covid-19).

○ $p(H_2) = 0.9$ (The prior probability that the person does not have Covid-19).

Substituting the above values into Equation (11), we have

$$p(E_1|H_1)p(H_1) = 0.9 \times 0.1 = 0.09.$$
$$p(E_1|H_2)p(H_2) = 0.2 \times 0.9 = 0.18.$$
$$p(E_1) = p(E_1|H_1)p(H_1) + p(E_1|H_2)p(H_2) = 0.09 + 0.18 = 0.27.$$

$$p(H_1|E_1) = \frac{0.09}{0.27} = 0.3333 \quad (27)$$
$$p(H_2|E_1) = \frac{0.18}{0.27} = 0.6667. \quad (28)$$

Interpretation of the results:

• Even though the test is 90% sensitive and 80% specific, the probability that the person has Covid-19 after being tested positive is only 33.33%.

• This is due to the low prior probability of having Covid-19 (10%) and the relatively high probability (20%) of a false positive rate due to the 80% specificity.

This example highlights the importance of considering prior probabilities and not relying solely on likelihoods (sensitivity and specificity).

When more events are observed, Equation (11) can be applied recursively by treating the most recently updated posterior probability as the prior probability for a new round of evidence updating. Figure 1 illustrates the process and results of updating when the 2nd event, fever is observed.

3.3. Numerical example of applying Dempster's rule of evidence combination

We use Example 1 above to illustrate the equivalence between Bayesian inference and Dempster evidence combination. Let e_1 be the prior distribution and e_2 the normalised likelihood, i.e.

$$\begin{aligned} e_1 &= \{(\theta, p_{\theta,1}), \forall\theta \subseteq \Theta\} \\ &= \{(H_1, p(H_1)), (H_2, p(H_2))\} \\ &= \{(H_1, 0.1), (H_2, 0.9)\} \text{or} \end{aligned}$$
$$m_1(H_1) = p(H_1) = 0.1 \quad \text{and}$$
$$m_1(H_2) = p(H_1) = 0.9$$

As likelihoods are $p(E_1|H_1) = 0.9$ and $p(E_1|H_2) = 0.2$, by normalising them, we get

$$p_{H_1,2} = \frac{p(E_1|H_1)}{p(E_1|H_1) + p(E_1|H_2)} = \frac{0.9}{0.9 + 0.2} = 0.8182 \quad (29)$$

$$p_{H_2,2} = \frac{p(E_1|H_2)}{p(E_1|H_1) + p(E_1|H_2)} = \frac{0.2}{0.9 + 0.2} = 0.1818 \quad (30)$$

$$\begin{aligned} e_2 &= \{(\theta, p_{\theta,2}), \forall\theta \subseteq \Theta\} = \{(H_1, p_{H_1,2}), (H_2, p_{H_2,2})\} \\ &= \{(H_1, 0.8182), (H_2, 0.1818)\} \end{aligned}$$

or

$$m_2(H_1) = p_{H_1,2} = 0.8182 \quad \text{and}$$
$$m_2(H_2) = p_{H_2,2} = 0.1818$$

Applying Dempster's rule in Equation (13), we get

$$\begin{aligned} m(H_1) &= \frac{\sum_{B\cap C=H_1} m_1(B)m_2(C)}{1 - \sum_{B\cap C=\emptyset} m_1(B)m_2(C)} \\ &= \frac{m_1(H_1)m_2(H_1)}{1 - m_1(H_1)m_2(H_2) - m_1(H_2)m_2(H_1)} \\ &= \frac{0.1 \times 0.8182}{1 - 0.1 \times 0.1818 - 0.9 \times 0.8182} = 0.3333 \end{aligned} \quad (31)$$

$$\begin{aligned} m(H_2) &= \frac{\sum_{B\cap C=H_2} m_1(B)m_2(C)}{1 - \sum_{B\cap C=\emptyset} m_1(B)m_2(C)} \\ &= \frac{m_1(H_2)m_2(H_2)}{1 - m_1(H_1)m_2(H_2) + m_1(H_2)m_2(H_1)} \\ &= \frac{0.9 \times 0.1818}{1 - 0.1 \times 0.1818 - 0.9 \times 0.8182} = 0.6667 \end{aligned} \quad (32)$$

3.4. Interpretation of the results

We can see that $m(H_1)$ and $m(H_2)$ in Equations (31) and (32) are equal to the posterior probabilities in Equations (27) and (28) obtained by applying Bayes' rule.

When a new event (E_2) is observed, the new evidence can be combined with the old one. For comparison purposes, suppose the second event observed is the same as what is shown in Figure 1. Equation (13) is then applied recursively by treating the previously combined evidence as the new 1st piece of evidence $m_1(\theta)$ and the normalised likelihood from the second event as the new second piece of evidence $m_2(\theta)$. Figure 2 illustrates the process and the results of combining the evidence from the second event that fever is observed.

Once again, we observe that the probabilities in the last column of Figure 1 match those in Figure 2, indicating that the final results obtained by applying Dempster's rule and Bayes' rule are identical. This equivalence is not coincidental. Yang et al. (2023) and Yang and Xu (2025) proved that Bayesian inference and Dempster's combination are strictly equivalent if and only if the prior distribution, regardless of its form, is treated as the initial piece of evidence and the normalised likelihoods for subsequent events as new evidence.

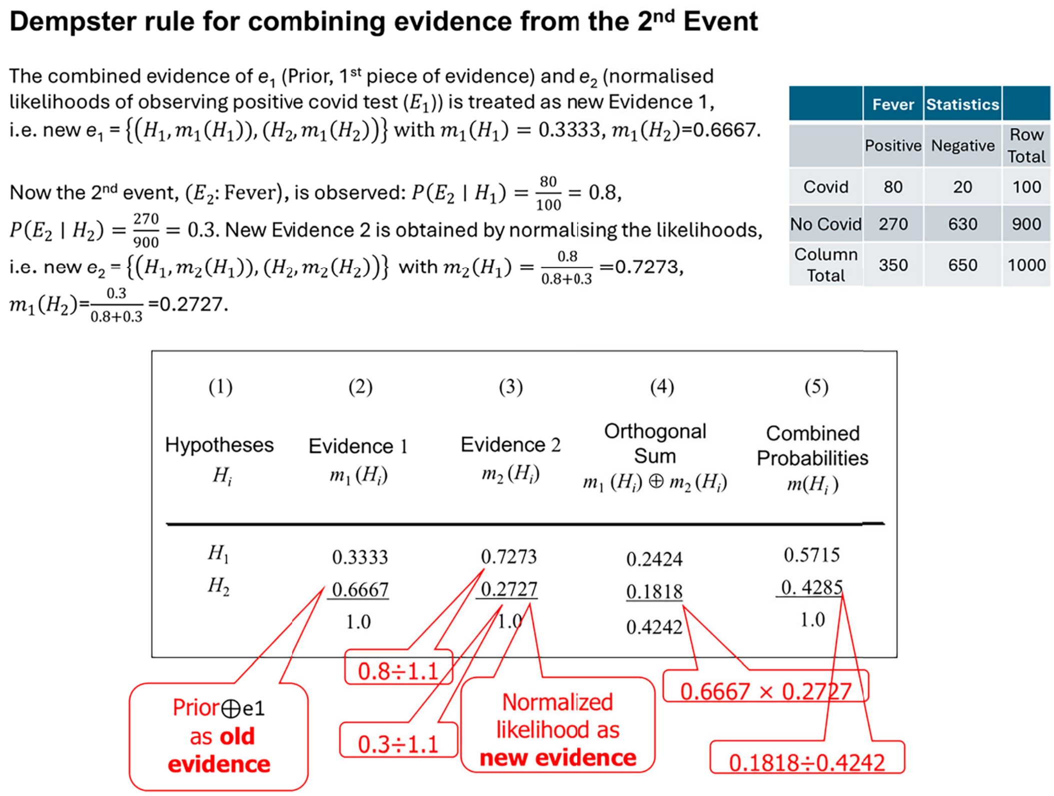

Figure 2. Combining new evidence from the second event using Dempster's rule.

4. Addressing critiques of Dempster's rule from a probabilistic perspective

There are quite a number of critiques against Dempster's rule. The most well-known is perhaps Zadeh's critique. He criticised Dempster's rule of combination for producing counterintuitive results, particularly in scenarios involving highly conflicting evidence. This issue has recently been completely clarified theoretically (Yang and Xu, 2025), and effective methods for addressing the issue have also been suggested in the past, for example using a robust yet practical perturbation analysis method (Yang and Xu, 2013) from a probabilistic perspective.

In this section, we will address a few observations about the behaviour of Dempster's rule and critical comments reported by Dezert, Tchamova, Han, and Tacnet (2013), Dezert and Tchamova (2011) and Cheng et al. (1988).

4.1. Addressing "Dempster's rule is equivalent to Bayes rule only when prior distribution is uniform"

It was asserted that Dempster's rule and Bayes' rule are not equivalent in general and are only equivalent when the prior distribution is uniform (Dezert et al., 2013).

From the proof of the Equivalence theorem, we can see that there is no need to assume a uniform prior. The reason that they made such a critique is that they suggested to use posterior distribution as evidence to be combined with Dempster's rule. If the posterior distributions of the L events are combined, the priors for these events will be used L times because each posterior already incorporates its prior. This results in the prior effect being accounted for L times. This is why they need a uniform (or non-informative) prior to establish the equivalence of the two rules.

Next, we use Example 1 in their paper (Dezert et al., 2013) to illustrate Bayesian updating (Equation (23)) and Dempster's combination (Equation (24)) and show the equivalence of the two without assuming a uniform prior. The example assumes that prior distribution is known and not uniform. The individual posterior distribution after each event is observed is also known. The task is to compute the joint posterior after three independent events are observed.

4.1.1. Bayesian updating

- Priors are given as follows:

$$p(H_1) = 0.2, \quad p(H_2) = 0.8$$

- Posteriors after individual events are given as follows:

  ○ After Event E_1 is observed: $p(H_1|E_1) = 0.1$, $p(H_2|E_1) = 0.9$
  ○ After Event E_2 is observed: $p(H_1|E_2) = 0.5$, $p(H_2|E_2) = 0.5$
  ○ After Event E_3 is observed: $p(H_1|E_3) = 0.6$, $p(H_2|E_3) = 0.4$

For this example, Equation (23) becomes

$$p(H_i \mid E_1 \cap E_2 \cap E_3) = \frac{p(H_i)\prod_{l=1}^{3}(p(H_i \mid E_l)/p(H_i))}{p(H_1)\prod_{l=1}^{3}(p(H_1|E_l)/p(H_1)) + p(H_2)\prod_{l=1}^{3}(p(H_2 \mid E_l)/p(H_2))}$$

- For E_1:

$$\frac{p(H_1|E_1)}{p(H_1)} = \frac{0.1}{0.2} = 0.5, \quad \frac{p(H_2|E_1)}{p(H_2)} = \frac{0.9}{0.8} = 1.125 \tag{33}$$

- For E_2:

$$\frac{p(H_1|E_2)}{p(H_1)} = \frac{0.5}{0.2} = 2.5,$$
$$\frac{p(H_2|E_2)}{p(H_2)} = \frac{0.5}{0.8} = 0.625 \tag{34}$$

- For E_3:

$$\frac{p(H_1|E_3)}{p(H_1)} = \frac{0.6}{0.2} = 3, \quad \frac{p(H_2|E_3)}{p(H_2)} = \frac{0.4}{0.8} = 0.5 \tag{35}$$

- *Numerator for H_1*:

$$p(H_1)\prod_{l=1}^{3}(p(H_1|E_l)/p(H_1)) = 0.2 \times 0.5 \times 2.5 \times 3 = 0.75 \tag{36}$$

- *Numerator for H_2*:

$$p(H_2)\prod_{l=1}^{3}(p(H_2|E_l)/p(H_2)) = 0.8 \times 1.125 \times 0.625 \times 0.5 = 0.28125 \tag{37}$$

- *The denominator is the sum of the numerators:*

$$\text{Denominator} = 0.75 + 0.28125 = 1.03125 \tag{38}$$

- Posterior probability for H_1:

$$p(H_1|E_1 \cap E_2 \cap E_3) = \frac{0.75}{1.03125} = 0.727 \tag{39}$$

- Posterior probability for H_2:

$$p(H_2|E_1 \cap E_2 \cap E_3) = \frac{0.28125}{1.03125} = 0.273 \tag{40}$$

4.1.2. Dempster's evidence combination

Using the same example, we illustrate Dempster's rule in Equation (24) as follows. Given the prior and the individual posteriors, likelihoods are computed by

$$p(E_l|H_i) = \frac{p(H_i|E_l)\cdot p(E_l)}{p(H_i)} \tag{41}$$

Let normalised likelihood be denoted as

$$\bar{p}(E_l H_i) = \frac{p_l(E_l|H_i)}{\sum_{k=1}^{N} p_l(E_l|H_k)} = \frac{p(H_i|E_l)\cdot p(E_l)/p(H_i)}{\sum_{k=1}^{N} p(H_k|E_l)\cdot p(E_l)/p(H_k)} \tag{42}$$

then Equation (24) can be re-written as follows:

$$p(H_i|m_1 m_2 \cdots m_L) = \frac{p(H_i)\prod_{l=1}^{L}\bar{p}(E_l|H_i)}{\sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L}\bar{p}(E_l|H_j)\right)} \tag{43}$$

- For E_1:

The normalised likelihoods are calculated by using Equations (42) and (33) for H_1, and (42) and (34) for H_2:

$$\bar{p}(E_1|H_1) = \frac{p(E_1|H_1)}{p(E_1|H_1)+p(E_1|H_2)} = \frac{p(H_1|E_1)\cdot p(E_1)/p(H_1)}{\sum_{k=1}^{N} p(H_k|E_1)\cdot p(E_1)/p(H_k)} = \frac{0.5p(E_1)}{0.5p(E_1)+1.125p(E_1)} = 0.3077$$

$$\bar{p}(E_1|H_2) = \frac{p(E_1|H_2)}{p(E_1|H_1)+p(E_1|H_2)} = \frac{1.125p(E_1)}{0.5p(E_1)+1.125p(E_1)} = 0.6923$$

- For E_2:

$$\bar{p}(E_2|H_1) = \frac{2.5p(E_2)}{2.5p(E_2)+0.625p(E_2)} = 0.8$$

$$\bar{p}(E_2|H_2) = \frac{0.625p(E_2)}{2.5p(E_2)+0.625p(E_2)} = 0.2$$

- For E_3:

$$\bar{p}(E_3|H_1) = \frac{3p(E_3)}{3p(E_3)+0.5p(E_3)} = 0.8571$$

$$\bar{p}(E_3|H_2) = \frac{0.5p(E_3)}{3p(E_3)+0.5p(E_3)} = 0.1429$$

- Combined probability for H_1:

$$p(H_1|m_1 m_2 \cdots m_L) = \frac{p(H_1)\prod_{l=1}^{L}\bar{p}(E_l|H_1)}{\sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L}\bar{p}(E_l|H_j)\right)} = \frac{0.2 \times 0.3077 \times 0.8 \times 0.8571}{0.042197 + 0.015829} = 0.7272 \tag{44}$$

- Combined probability for H_2:

$$p(H_2|m_1 m_2 \cdots m_L)$$

Table 1. Input basic belief assignments (BBAs) $m_1(\cdot)$ and $m_2(\cdot)$.

Focal elements	$m_1(\cdot)$	$m_2(\cdot)$
A	a_1	0
B	a_2	0
$A \cup B$	$1 - a_1 - a_2$	b_1
C	0	$1 - b_1 - b_2$
$A \cup B \cup C$	0	b_2

$$= \frac{p(H_2)\prod_{l=1}^{L}\bar{p}(E_l|H_2)}{\sum_{j=1}^{N}\left(p(H_j)\prod_{l=1}^{L}\bar{p}(E_l|H_j)\right)}$$

$$= \frac{0.8 \times 0.6923 \times 0.2 \times 0.1429}{0.042197 + 0.015829} = 0.2723 \quad (45)$$

We can see that both Bayesian updating and Dempster's combination have generated the same results.

4.2. Addressing 'Bayes rule is not associative or commutative'

The above example was used to claim that Bayes' rule is neither associative nor commutative (Dezert et al., 2013). However, no detailed explanations were provided to support the claim.

From the proof of the above Equivalence theorem and the calculation process for the example in Section 4.1, it is obvious that for multiple pieces of independent evidence, the left-hand side of Equation (25) shows that Bayes' rule is both associative and commutative, while the right-hand side demonstrates the same for Dempster's rule. From Equations (36), (37), (44) and (45), we can see that the outcomes of both Bayesian updating and Dempster's combination are proportional to the multiplications of the prior and the individual event's likelihoods, which are both associative and commutative. Those properties of Dempster's combination are also confirmed by Pearl (Pearl, 1988, p. 432). This means that the three pieces of evidence can be updated or combined in any order or grouped in any configuration, and the final results will remain the same in both Bayesian updating and Dempster's combination.

4.3. Addressing 'Dempster's rule is not sensitive to some evidence'

Dezert and Tchamova (2011) observed that Dempster's rule does not respond adequately to certain sources of evidence, even when the level of conflict between the sources is low. Table 1 presents the details of the example they used to illustrate this phenomenon.

By applying Dempster's rule, the two pieces of evidence represented by the two probability mass functions shown in Table 1 can be combined. The combined mass function $m_{12}(\cdot)$ is given by

- $m_{12}(A) = a_1(b_1 + b_2)$
- $m_{12}(B) = a_2(b_1 + b_2)$
- $m_{12}(A \cup B) = (1 - a_1 - a_2)(b_1 + b_2)$
- $K_{12} = m_{12}(\emptyset) = 1 - b_1 - b_2$

After normalisation by $1 - K_{12} = b_1 + b_2$, the combined probability masses are

- $m_{DS}(A) = \frac{m_{12}(A)}{1-K_{12}} = a_1$
- $m_{DS}(B) = \frac{m_{12}(B)}{1-K_{12}} = a_2$
- $m_{DS}(A \cup B) = \frac{m_{12}(A \cup B)}{1-K_{12}} = 1 - a_1 - a_2$

where $m_{DS}(\cdot)$are the combined and normalised probability masses, which are exactly the same as $m_1(\cdot)$, meaning that $m_2(\cdot)$ plays no role in the final result. While this is an interesting and correct observation, does it mean that Dempster's rule is not rational?

To answer this question, let's examine what $m_2(\cdot)$ represents.

First, although $m_2(\cdot)$ assigned some probability mass to hypothesis C, $m_1(\cdot)$ categorically denies it by assigning 0 probability to C and any subsets that contain C. Therefore, after combination, the probability mass assigned to hypothesis C is 0 regardless of what $m_2(\cdot)$ says about C. This is because Dempster's rule and Bayes' rule assume full reliability for any evidence. This is the limitation of both rules. This can be addressed by introducing a discounting factor, such as in the Evidential Reasoning rule (probabilistic) or Shafer's discounting method (non-probabilistic).

With C out of the picture, we can now focus on A and B. From Table 1, we can see that $m_2(\cdot)$ only assigns probability to subsets $A \cup B$ and $A \cup B \cup C$ (which is essentially $A \cup B$ as C was already excluded as explained in the previous paragraph) indifferently, nothing to singleton A or B. This means that $m_2(\cdot)$ is effectively telling $m_1(\cdot)$: 'I don't know how to divide the probability between A or B. I leave that decision to you'. The final result reflects exactly what $m_2(\cdot)$ intends. Therefore, we argue that Dempster's rule is rational under the assumption that any evidence is fully reliable and the outcome is what can be expected from Dempster's rule under the assumption.

4.4. Interpreting imprecise probabilities

Let's first look at the following example. Suppose

$$e_1 = ((H_1, 0.8), (H_2, 0.1), ((H_1, H_2), 0.1) \quad (46)$$

$$e_2 = ((H_1, 0.6), (H_2, 0.2), ((H_1, H_2), 0.2) \quad (47)$$

Dempster's combination of these pieces of evidence yields:

$$e_1 \oplus e_2 = (H_1, 0.8974), (H_2, 0.0769), \times ((H_1, H_2), 0.0256) \quad (48)$$

One may argue that the probability 0.8974 for H_1 and 0.0769 for H_2 in Equation (48) should be treated as the lower bound of $p(H_1)$ and $p(H_2)$, respectively, since the unknown probability 0.0256 could be allocated to H_1 or H_2 to make them larger. On the other hand, if all the unknown probability were assigned to only one of the singleton hypotheses, such as H_1, it would seem reasonable to assume that the sum:

$$0.8974 + 0.0256 = 0.9231 \tag{49}$$

becomes the upper bound of $p(H_1)$. Therefore, the upper and lower bound for $p(H_1)$ would be [0.8974, 0.9231]. This was Dempster's original interpretation of the unknown probability (Dempster, 1967, 1968).

However, suppose we now assign all the unknown probabilities, $[(H_1, H_2), 0.1]$ in e_1 and $[(H_1, H_2), 0.2]$ in e_2 to H_1 before combining them, so that the two pieces of evidence look like the ones shown in Equations (50) and (51). We can now apply either Dempster's rule or the Bayes rule to combine them because all probabilities are precisely known.

$$e_1 = ((H_1, 0.9), (H_2, 0.1), ((H_1, H_2), 0) \tag{50}$$

$$e_2 = ((H_1, 0.8), (H_2, 0.2), ((H_1, H_2), 0) \tag{51}$$

Applying either rule to combine them, we get:

$$e_1 \oplus e_2 = (H_1, 0.9730), (H_2, 0.0270), ((H_1, H_2), 0) \tag{52}$$

The combined probability for H_1 is $0.9730 > 0.9231$, which exceeds the upper bound in Equation (49). Conversely, the combined probability for H_2 in Equation (52), $p(H_2) = 0.0270$, is much lower than its lower bound of 0.0769 in Equation (48). This demonstrates that the probabilities in Equations (48) and (49) should not be interpreted as true upper and lower bounds.

Cheng et al. (1988) observed a similar phenomenon. They reported:

> We show that the width of the interval given by the Dempster rule is narrower than that of Bayes.

How to address this phenomenon? From the evidence combination process of Dempster's rule, we can see that the unknown probabilities in a piece of evidence must be respected as it is. It should not be divided into pieces to be reallocated by the evidence itself because this evidence alone cannot make such a reallocation decision but essentially allows other evidence to decide where to allocate the unknown probabilities in the combination process so that the unknown could be reduced by other evidence. This phenomenon agrees with our best practice that whenever possible we do not make unnecessary or unjustifiable assumptions but collect more information whenever there is a need to clarify any doubt.

As a matter of fact, Dempster (2008) also realised the phenomenon by stating that:

> My original adoption of the terms lower and upper probability invites confusion with other theories that use these same terms, and provokes debate concerning the implied existence or non-existence of unknown true probabilities lying between defined lower and upper bounds. No such existence is implied, since true probabilities nowhere appear in the theory.

It is in the same paper (Dempster, 2008) that Dempster made it crystal clear that:

> The mass $m(A)$ associated with each assertion $A \subset \Theta$ is an atom in the sense that it cannot be further broken down into pieces assigned to subsets of A.

5. Relationship among Bayes' rule, Dempster's rule and the evidential reasoning rule

Figure 3 (Xu and Yang, 2017) illustrates the relationship among Bayes' rule, Dempster's rule and the evidential reasoning (ER) rule. Bayes' rule is a special case of Dempster's rule as proved by Yang et al. (2023) and Yang and Xu (2014) and also demonstrated in this paper. Earlier, Yang and Xu (2013) proved that Dempster's rule is a special case of the ER rule. In other words, the ER rule extends Dempster's rule to cases where evidence is not fully reliable, and Dempster's rule extends Bayes' rule to cases where there is unknown probability. All of them are probabilistic.

5.1. Assumptions in Bayes' rule

$$r = 1, \; p(\theta) + p(\bar{\theta}) = 1$$

Here, r represents the reliability of evidence which is used for discounting $p(\theta)$. All probabilities are assigned to singleton hypotheses, i.e. θ can only be singleton hypotheses. In Bayes' rule, basic probability, basic probability mass, and belief degree all mean the same thing. There is no need to distinguish them.

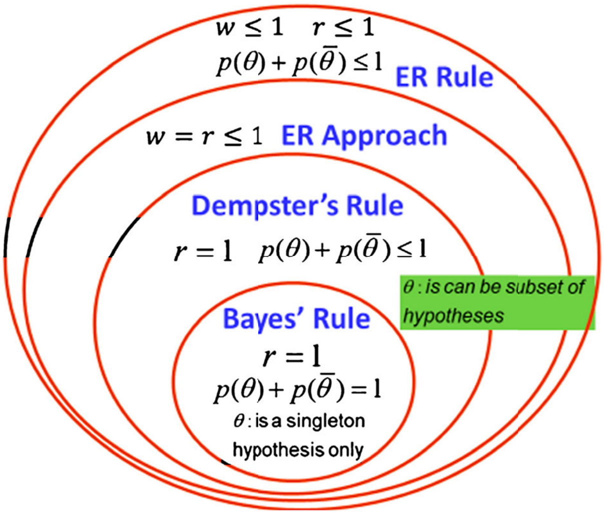

Figure 3. Relationship among Bayes' rule, Dempster's rule, and the evidential reasoning rule (Xu and Yang, 2017).

5.2. Assumptions in Dempster's rule

$$r = 1,\ p(\theta) + p(\bar{\theta}) \le 1$$

Dempster's rule allows basic probability (or basic probability mass) to be assigned to subsets of hypotheses, $\theta \subseteq \Theta$ and assumes that evidence is fully reliable ($r = 1$).

It is important to emphasise that Dempster's rule operates on basic probability mass only. He has made this point crystal clear that evidence combination processes should be carried out on basic probability mass (Dempster, 2015).

5.3. Relationship between Dempster's rule and Bayes' rule

When $r = 1$and $p(\theta) + p(\bar{\theta}) = 1$ for any θ, that is, the basic probability is only assigned to singleton hypotheses (θ), Dempster's rule reduces to Bayes' rule.

5.4. Assumptions in the ER Rule (Yang and Xu 2013)

$$w \le 1, \quad r \le 1, \quad p(\theta) + p(\bar{\theta}) \le 1$$

where w and r are weights and reliability representing the importance and the reliability of evidence, respectively. They can be equal or not equal. Weight is defined as conditional probability and reliability is also a probability defined as the ability of evidence to provide a correct conclusion (Yang and Xu, 2025). When they are used to weigh basic probabilities $p(\theta)$, the weighted basic probabilities $wp(\theta)$ are referred to as basic probability mass. We can see that in Bayes' rule and Dempster's rule, those two concepts (basic probability and basic probability mass) are the same because $r = w = 1$.

5.5. Relationship between the ER rule and Dempster's rule

When $w = r = 1$ (assuming full reliability), the ER rule becomes Dempster's rule.

5.6. Relationship between the Evidential Reasoning (ER) Rule and Bayes' rule

When $w = r = 1$ and $p(\theta) + p(\bar{\theta}) = 1$, that is, all θ are singleton hypotheses, the ER rule becomes Bayes' rule.

5.7. Relationship between the ER Rule and Shafer's discounting method

Shafer's discounting method is not shown in Figure 3 because it does not lead to probabilistic inference even though it uses Dempste'r's rule for combining discounted belief degrees. Shafer's discounting method does not conform to the likelihood principle (Birnbaum, 1962), as explained in Yang and Xu (2013). However, it is closely related to the ER rule in the following context. Shafer uses a discounting factor to take into account the reliability of evidence for discounting each piece of evidence but allocates the unreliable mass (or leftover belief) to the frame of discernment, Θ, as follows (Shafer, 1976, 1990):

$$m(\theta) = \begin{cases} \alpha \cdot p(\theta), & \theta \ne \Theta, \\ \alpha \cdot p(\theta) + (1 - \alpha), & \theta = \Theta, \\ 0, & \theta = \emptyset, \end{cases} \tag{53}$$

where $p(\theta)$is the basic probability to which a piece of evidence points to proposition θ, and α ($0 \le \alpha \le 1$) a factor to discount $p(\theta)$. After the evidence is discounted, then Dempster's rule is applied to combine the evidence.

Inspired by Shafer's discounting method, the ER rule also discounts evidence with a factor which is a conditional probability representing the importance of the evidence. The key difference between the ER rule and Shafer's discounting method is that Shafer allocates the unreliable portion of the evidence directly to the frame of discernment, while the ER rule buffers it in the power set of the frame of discernment as shown in Equation (54) in the anticipation that other evidence will redistribute it among the elements of the power set.

$$m(\theta) = \begin{cases} \alpha \cdot p(\theta), & \theta \subseteq \Theta, \\ (1 - \alpha), & \theta = P(\Theta), \\ 0, & \theta = \emptyset, \end{cases} \tag{54}$$

This difference is subtle but vital in making the ER rule probabilistic as explained in Yang and Xu (2013, 2025) and Yang et al. (2023). The necessity and rationale of buffering the unreliable portion of evidence to the power set are carefully argued and theoretically proven within the new MAKER (maximum likelihood evidential reasoning) framework (Yang and Xu, 2025).

6. Conclusion

In this paper, we have examined the relationships among Bayes' rule, Dempster's rule, Shafer's discounting method and the Evidential Reasoning (ER) rule for evidence combination. We have also addressed some critiques and observations of the behaviours of Dempster's rule from a probabilistic perspective,

Bayesian inference is perhaps an ideal engine for inference with perfect data (Yang and Xu, 2025) but is generally perceived as a tool for reasoning with 'small data' because it requires precise and complete probability information. In contrast, evidence theory relaxes this requirement and is therefore more suitable for 'big data' applications. If evidence theory is theoretically sound and interpretable, it can be applied more widely and give users greater confidence in the results it generates. This motivation underpins our work to investigate some of the critiques of Dempster's rule, with the intention of bringing the development of evidence theory back to its original probabilistic foundation.

We have tried to uncover a probabilistic rule for evidence combination that can take advantage of both Dempster's rule and Shafer's discounting method, which has led to the establishment of the ER rule. By building on Dempster's and Shafer's original insights that probabilities may be imprecise and evidence may be partially unreliable, the ER rule retains the theoretical rigour of Dempster's rule and the practical utility of Shafer's discounting method. This explains why the ER rule is not only probabilistic and rational but also practical.

Our work so far demonstrates that the goal of making evidence theory probabilistic again is achievable if

- Any evidence is acquired as normalised likelihood distribution except for the first piece of evidence which should be prior distribution if available,
- All elements of a powerset are treated with equal respect in discounting or conditioning, with the probability meaning of evidence kept intact,
- All probabilistic operations in evidence combination are carried out on basic probability functions instead of belief functions, and
- The combination of multiple pieces of evidence constitutes a process for generating their joint probabilities.

However, much work remains to fulfill this goal. This effort requires the collaboration of both the statistics and evidence theory communities to extend rigorous statistical theories, methods, and tools to handle imprecise probabilities and unreliable data sources.

Potential-related research directions include exploring the links between evidence theory and random set theory, since both generalise classical probability to handle uncertainty about sets rather than single elements; and investigating connections between evidence theory and quantum theory, as both frameworks deal with the probabilities of multiple hypotheses.

Responses to reviewers comments.docx

Acknowledgements

Special thanks to Professor Arthur Dempster, who provided invaluable advice to Yang and Xu both in person and via email. It is due to his advice that a principled approach is explored to establish the maximum likelihood evidential reasoning rule.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work is partially supported by the European Union's Horizon 2020 Research and Innovation Programme RISE under grant number: 823759 (ReMESH), the Natural Science Foundation of China under grant number: 72010107002, and InnoHK initiative, the Government of the HKSAR, and Hong Kong Laboratory for *AI*-Powered Financial Technologies.

Notes on contributors

Dong-Ling Xu is Chair Professor of Decision Science and Systems at Alliance Manchester Business School, University of Manchester, UK. With over 30 years of research experience, she has made significant contributions to data analysis, statistical inference, machine learning, decision support systems under uncertainty, system and process modelling, and statistical fault detection. Her work spans a variety of applications across multiple fields. Together with her collaborators, she developed several interactive web-based decision support systems and the Windows-based Intelligent Decision System (IDS) software. IDS implements the Evidential Reasoning approach, enabling multiple criteria decision analysis under diverse forms of uncertainty – including randomness, subjective judgments, vague information and missing data – without requiring the deletion or distortion of uncertain data. Xu has also designed statistical fault detection systems that have been adopted by major organisations such as General Motors, Tesco, the NHS, Ford, Shell, BP and CNOOC. These tools have addressed critical challenges in areas such as healthcare, finance, system safety and security, and quality management through organisational self-assessment. Today, the Evidential Reasoning approach and IDS software are utilised by practitioners and researchers in over 50 countries worldwide. Xu has authored more than 100 peer-reviewed journal articles, book chapters and books.

Jian-Bo Yang is Chair Professor of Decision and Systems Sciences at Alliance Manchester Business School, The University of Manchester, UK. Over the last three decades, He has conducted theoretical and methodological research in many areas, including the evidential reasoning (ER) theory; multiple criteria decision analysis under uncertainties; probabilistic inference and decision analysis using both data and judgments; multiple objective optimisation; intelligent system modelling; explainable artificial intelligence (XAI) and interpretable machine learning (IML); hybrid decision methodologies and technologies combining concepts and techniques from decision science, systems science, operational research and AI. His current applied research covers a range of applications driven by historical data, enabled by human knowledge and powered by ER, including modelling and decision support for professional services such as finance, insurance and healthcare; diagnosis and prognosis, design and operation decision analysis in healthcare, engineering and social systems; pattern identification and analysis of consumer behaviours; analysis of public sentiments and system risks (financial or non-financial); new product development; aggregated production management; system maintenance management; risk and security modelling and analysis; performance analysis and improvement of products, processes and organisations.

He is the principal investigator or co-investigator for over 70 research projects at a total value of over £23 m, funded by many organisations, including the UK Engineering and Physical Science Research Council (EPSRC), the UK Economic and Social Research Council (ESRC), Innovate UK, UK Department of Environment, Food and Rural Affairs (DEFRA), European Commission (EC), Natural Science Foundation of China (NSFC), Hong Kong Research Grant Council (HKRGC), US Government and industry. He has published 4 books, over 280 journal papers and book chapters, and a similar number of conference papers, with extensive citations in Web of Science and Google Scholar, and developed several software packages in optimisation and decision support with wide applications worldwide.

Yingming Wang received a B.S. degree in industrial electrical automation from Jiangsu University, Jiangsu, China, in 1984, an M.S. degree in system engineering from Huazhong University of Science and Technology in 1987 and the Ph.D. degree in automatic control theory and application from Southeast University in 1991. He worked as a postdoctoral fellow in the Department of Environmental Engineering, Tsinghua University from 1991 to 1993, an associate professor, professor and doctoral supervisor successively in the Department of Automation and the Department of Management Science, Xiamen University, from 1993 to 2001, a research fellow in Manchester Business School, the University of Manchester from 2002 to 2006, a senior research fellow in the Department of Systems Engineering and Engineering Management, City University of Hong Kong from 2007 to 2008. He joined the School of Economics and Management, Fuzhou University in 2009. Yingming Wang has received the following honours and rewards: a distinguished professor of 'Minjiang Scholar' of Fujian Province and the 'National Science Fund for Distinguished Young Scholars' both in 2009, a distinguished professor of 'Yangtze River Scholar' and the outstanding returnees from overseas studies in Fujian Province both in 2012 and the 'High Cited Chinese Researchers' from 2014 to 2024. He has published over 230 SCI and 72 SSCI-indexed journal papers. His research interests include Decision Theory and Methods, Data Envelopment Analysis, Rule Base Reasoning and Quality Function Deployment.

References

Birnbaum, A. (1962). On the foundations of statistical inference. *Journal of the American Statistical Association*, *57*(298), 269–326.

Cheng, Y., Kashyap, R. L., Erickson, G. J., & Smith, C. R. (1988). Comparison of Bayesian and Dempster's rules in evidence combination. In *Maximum-entropy and Bayesian methods in science and engineering: Fundamental theories of physics* (pp. 427–433). Springer. https://doi.org/10.1007/978-94-010-9054-4_27

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

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society – Series B (Methodological)*, *30*(2), 205–247.

Dempster, A. P. (2008). The Dempster–Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, *48*(2), 365–377. https://doi.org/10.1016/j.ijar.2007.03.004

Dempster, A. P. (2015, July 30–31). *Why the econometrics of finance should be recast in Sempster–Shafer terms*. The Second GOSS Private Equity Forum and Ninth RMI Annual Risk Management Conference. https://goss.com.hk/en/GOSS_2ndForum.php

Dezert, J., & Tchamova, A. (2011, April 4–8). *On the behavior of Dempster's rule of combination*. Presented at the Spring School on Belief Functions Theory and Applications (BFTA). http://hal.archives-ouvertes.fr/hal-00577983/

Dezert, J., Tchamova, A., Han, D., & Tacnet, J. M. (2013). Why Dempster's rule doesn't behave as Bayes' rule with informative priors. In *Proceedings of the IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1–5). IEEE. https://www.onera.fr/sites/default/files/297/INISTA2013-p80.pdf

Haenni, R. (2005). Shedding new light on Zadeh's criticism of Dempster's rule of combination. In *Proceedings of the 7th International Conference on Information Fusion (FUSION)* (pp. 2–6). Philadelphia, PA, USA: IEEE Xplore. https://doi.org/10.1109/ICIF.2005.1591951

Liu, W., & Hong, J. (2000). Reinvestigating Dempster's idea on evidence combination. *Knowledge and Information Systems*, *2*(2), 223–241.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Pearl, J. (1990). Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning*, *4*(5–6), 363–389.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Shafer, G. (1990). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, *4*, 323–362.

Xu, D. L., & Yang, J. B. (2017, June 20–21). ER rule for Bayes inference with imperfect data. In *Network plus: Industrial systems in the Digital Age Conference 2017*.

Yang, J. B., & Xu, D. L. (2013). Evidential reasoning rule for evidence combination. *Artificial Intelligence*, *205*, 1–29.

Yang, J. B., & Xu, D. L. (2014). A study on generalising Bayesian inference to evidential reasoning. In F. Cuzzolin (Ed.), *Belief function: Theory and applications, lecture notes in artificial intelligence* (pp. 180–189). Springer.

Yang, J. B., & Xu, D. L. (2025). Maximum likelihood evidential reasoning. *Artificial Intelligence*, *340*, 1–61. https://doi.org/10.1016/j.artint.2025.104289

Yang, J. B., Xu, D. L., Xu, X. B., & Fu, C. (2023). Likelihood analysis of imperfect data and Supplementary materials. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, *53*(8), 5046–5057.

Zadeh, L. A. (1979). A theory of approximate reasoning. *Machine Intelligence*, *9*, 149–194.