INTERNATIONAL JOURNAL OF GENERAL SYSTEMS
2020, VOL. 49, NO. 1, 112–141
https://doi.org/10.1080/03081079.2019.1658756

An expectation operator for belief functions in the Dempster–Shafer theory*

Prakash P. Shenoy

School of Business, University of Kansas, Lawrence, KS, USA

ABSTRACT

The main contribution of this paper is a new definition of expected value of belief functions in the Dempster–Shafer (D–S) theory of evidence. Our definition shares many of the properties of the expectation operator in probability theory. Also, for Bayesian belief functions, our definition provides the same expected value as the probabilistic expectation operator. A traditional method of computing expected of real-valued functions is to first transform a D–S belief function to a corresponding probability mass function, and then use the expectation operator for probability mass functions. Transforming a belief function to a probability function involves loss of information. Our expectation operator works directly with D–S belief functions. Another definition is using Choquet integration, which assumes belief functions are credal sets, i.e. convex sets of probability mass functions. Credal sets semantics are incompatible with Dempster's combination rule, the center-piece of the D–S theory. In general, our definition provides different expected values than, e.g. if we use probabilistic expectation using the pignistic transform or the plausibility transform of a belief function. Using our definition of expectation, we provide new definitions of variance, covariance, correlation, and other higher moments and describe their properties.

ARTICLE HISTORY
Received 5 December 2018
Accepted 19 July 2019

KEYWORDS
Dempster–Shafer theory of evidence; expected value; variance; covariance; correlation; higher moments; pignistic transformation; plausibility transformation

1. Introduction

The main goal of this paper is to propose an expectation operator for belief functions in the D–S theory of belief functions (Dempster 1967; Shafer 1976). The D–S theory of belief functions consists of representation of knowledge and evidence, and two operators for making inferences from the representations. Representations consist of basic probability assignments, belief functions, plausibility functions, commonality functions, credal sets of probability mass functions, etc. The two main operators are Dempster's combination rule for aggregating distinct knowledge and a marginalization rule for coarsening knowledge.

There are other theories of belief functions that use the same representations. For example, in the imprecise probability community, a belief function is regarded as a lower bound of a convex set of probability mass functions (PMFs) called a credal set. When we observe an event, the PMFs in the credal set are all conditioned on the observed event, and the lower

CONTACT Prakash P. Shenoy pshenoy@ku.edu
*Presented at the 11th Workshop on Uncertainty Processing (WUPES'18), Třeboň, Czech Republic, June 6–9, 2018.

bound of the updated PMFs again forms a belief function. Such an updating rule, called the Fagin–Halpern combination rule (Fagin and Halpern 1991), is different from Dempster's combination rule. In this paper, we are concerned only with the D–S theory of belief functions and not with the various other theories that use belief functions. Although a credal set representation of a belief function is mathematically equivalent to other representations, credal set semantics of belief functions are incompatible with Dempster's combination rule (Shafer 1981, 1990, 1992; Halpern and Fagin 1992), the center-piece of the D–S theory.

In probability theory, for discrete real-valued random variables characterized by a probability mass function (PMF), the expected value of X can be regarded as a weighted average of the states of X where the weights are the probabilities associated with the values. Our definition is similar. As we have probabilities associated with subsets of states, first we define the value of a subset as the weighted average of the states of the subset where the weights are the commonality values of the singleton states. Then the expected value of X is defined to be the weighted average of the values of the subsets where the weights are the commonality values of the subsets.

A traditional method of computing expectation of real-valued functions is to first transform a D–S belief function to a corresponding PMF, and then use the expectation operator for PMFs. Transforming a belief function to a probability function involves loss of information. Our expectation operator works directly with D–S belief functions.

In the decision-making with sets of probability distributions literature, there is a definition of expectation of a real-valued utility function called the Choquet integral (Choquet 1953; Gilboa and Schmeidler 1994). The Choquet integral consists of using the smallest value of a real-valued function over a subset as the value of the subset, and the expected value of the real-valued function is defined as the weighted average of the values of the subsets where the weights are the basic probability assignment values. Our definition of the expectation operator is different from the Choquet integral and is specifically designed for the D–S theory of belief functions. We believe that the Choquet integral is appropriate for the other theories of belief functions, and not for the D–S belief function theory.

In general, our definition provides different expected values than, e.g. if we use the pignistic transform or the plausibility transform, or if we use the Choquet integral. We use the expectation operator to define variance, covariance, correlation, and other higher moments and describe their properties. Most of the properties of these moments in probability theory are also satisfied by our definitions.

Our definition of expectation is motivated by a recent definition of entropy of D–S belief functions in Jiroušek and Shenoy (2018c). If we define $I(\mathsf{a}) = \log_2(1/Q_{m_X}(\mathsf{a}))$ as the information content of observing subset $\mathsf{a} \in 2^{\Omega_X}$ whose uncertainty is described by m_X, then similar to Shannon's definition of entropy of PMFs (Shannon 1948), we can define entropy of BPA m_X for X as an expected value of the function $I(X)$, i.e. $H(m_X) = E_{m_X}(I(X))$. This is what is proposed in Jiroušek and Shenoy (2018b, 2018c). This definition of entropy has many nice properties. In particular, it satisfies the compound distributions property: $H(m_X \oplus m_{Y|X}) = H(m_X) + H(m_{Y|X})$, where $m_{Y|X}$ is a BPA for (X, Y) obtained by $\oplus\{m_{x,Y} : x \in \Omega_X\}$, $m_{x,Y}$ is a BPA for (X, Y) obtained by conditional embedding of conditional BPA $m_{Y|x}$ for Y given $X = x$, and $\oplus$ is Dempster's combination rule. The compound distribution property of Shannon's entropy of a PMF is one of the most important properties that characterizes Shannon's definition (Shannon 1948). The definition of entropy of belief functions in Jiroušek and Shenoy (2018b, 2018c) is the only one of the many definitions proposed in the literature (see (Jiroušek and Shenoy 2018a) for a review) that satisfies the compound distributions property.

Our definition of expectation is using the commonality function representation of a belief function, which does not have easy semantics. Thus, we are unable to provide an intuitive justification of our definition. This is compensated for by showing that our definition satisfies many of the properties satisfied by the definition of expectation in probability theory, and the proofs of these properties are all simple and straightforward.

An outline of the remainder of the paper is as follows. In Section 2, we review the representations and operations of the D–S theory of belief functions. In Section 3, we provide our definition of the expected value of a real-valued random variable characterized by a commonality function. For a symbolic-valued random variable X, assuming we have a real-valued function g_X from the set of all non-empty subsets of the states of X, we also provide a definition of the expected value of g_X. Also, we show that our definition of expected value shares many of the properties of the probabilistic expected value, and we compare our definition with the probabilistic expectation using pignistic and plausibility transforms. Also, we compare our definition of expected value with the Choquet integral definition. In Section 4, we provide a new definition of variance and describe its properties. In Section 5, we provide a new definition of covariance and correlation and describe their properties. In Section 6, we provide a new definition of higher moments about the mean and about the origin and describe their relationship. Finally, in Section 7, we summarize and conclude.

2. Basic definitions in the D–S belief functions theory

In this section, we review the basic definitions in the D–S belief functions theory. Like the various uncertainty theories, D–S belief functions theory includes functional representations of uncertain knowledge, and operations for making inferences from such knowledge. Most of this material is taken from Jiroušek and Shenoy (2018b).

2.1. Representations of belief functions

Belief functions can be represented in five different ways: basic probability assignments, plausibility functions, belief functions, commonality functions, and credal sets. These are briefly discussed below.

Definition 2.1 (Basic Probability Assignment): Suppose X is a random variable with state space Ω_X. Let 2^{Ω_X} denote the set of all *non-empty* subsets of Ω_X. A basic probability assignment (BPA) m_X for X is a function $m_X : 2^{\Omega_X} \to [0, 1]$ such that

$$\sum_{\mathsf{a} \in 2^{\Omega_X}} m_X(\mathsf{a}) = 1.$$

The non-empty subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m_X(\mathsf{a}) > 0$ are called *focal* elements of m_X. An example of a BPA for X is the *vacuous* BPA for X, denoted by ι_X, such that $\iota_X(\Omega_X) = 1$. We say m_X is *deterministic* if m_X has a single focal element (with probability 1). Thus, the vacuous BPA for X is deterministic with focal element Ω_X. We say m_X is *consonant* if the focal elements of m are nested, i.e. if $F_1 \subset F_2 \subset \ldots \subset F_m$, where $\{F_1, \ldots, F_m\}$ denotes the set of all focal elements of m_X. Deterministic BPAs are trivially consonant. We say a BPA is *quasi-consonant* if the intersection of all focal elements is non-empty. Consonant BPA are quasi-consonant, but not vice-versa. Thus, a BPA with focal elements $\{x_1, x_2\}$ and $\{x_1, x_3\}$ is quasi-consonant, but not consonant. If all focal elements of m_X are singleton subsets of Ω_X, then we say m is *Bayesian*. In this case, m_X is equivalent to the PMF P_X for X such that $P_X(x) = m_X(\{x\})$ for each $x \in \Omega_X$.

Definition 2.2 (Plausibility Function): The information in a BPA m_X can be represented by a corresponding plausibility function Pl_{m_X} that is defined as follows:

$$Pl_{m_X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X}:\, \mathsf{b} \cap \mathsf{a} \neq \emptyset} m_X(\mathsf{b}) \quad \text{for all } \mathsf{a} \in 2^{\Omega_X}.$$

For an example, suppose $\Omega_X = \{x, \bar{x}\}$. Then, the plausibility function Pl_{ι_X} corresponding to BPA ι_X is given by $Pl_{\iota_X}(\{x\}) = 1$, $Pl_{\iota_X}(\{\bar{x}\}) = 1$, and $Pl_{\iota_X}(\Omega_X) = 1$.

Definition 2.3 (Belief Function): The information in a BPA m can also be represented by a corresponding belief function Bel_{m_X} that is defined as follows:

$$Bel_{m_X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X}:\, \mathsf{b} \subseteq \mathsf{a}} m_X(\mathsf{b}) \quad \text{for all } \mathsf{a} \in 2^{\Omega_X}.$$

For the example above with $\Omega_X = \{x, \bar{x}\}$, the belief function Bel_{ι_X} corresponding to BPA ι_X is given by $Bel_{\iota_X}(\{x\}) = 0$, $Bel_{\iota_X}(\{\bar{x}\}) = 0$, and $Bel_{\iota_X}(\Omega_X) = 1$.

Definition 2.4 (Commonality Function): The information in a BPA m_X can also be represented by a corresponding commonality function Q_{m_X} that is defined as follows:

$$Q_{m_X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X}:\, \mathsf{b} \supseteq \mathsf{a}} m_X(\mathsf{b}) \quad \text{for all } \mathsf{a} \in 2^{\Omega_X}.$$

For the example above with $\Omega_X = \{x, \bar{x}\}$, the commonality function Q_{ι_X} corresponding to BPA ι_X is given by $Q_{\iota_X}(\{x\}) = 1$, $Q_{\iota_X}(\{\bar{x}\}) = 1$, and $Q_{\iota_X}(\Omega_X) = 1$. If m_X is a Bayesian BPA for X, then Q_{m_X} is such that $Q_{m_X}(\mathsf{a}) = m_X(\mathsf{a})$ if $|\mathsf{a}| = 1$, and $Q_m(\mathsf{a}) = 0$ if $|\mathsf{a}| > 1$. Notice also that for singleton subsets $\mathsf{a} \in 2^{\Omega_X}$ ($|\mathsf{a}| = 1$), $Q_{m_X}(\mathsf{a}) = Pl_{m_X}(\mathsf{a})$. This is because for singleton subsets a, the set of all subsets that have non-empty intersection with a coincide with the set of all supersets of a. Q_{m_X} is a non-increasing function in the sense that if $\mathsf{b} \subseteq \mathsf{a}$, then $Q_{m_X}(\mathsf{b}) \geq Q_{m_X}(\mathsf{a})$. Finally, Q_{m_X} is a normalized function in the sense that:

$$\begin{aligned}
\sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} Q_{m_X}(\mathsf{a}) &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} \left(\sum_{\mathsf{b} \in 2^{\Omega_X}:\, \mathsf{b} \supseteq \mathsf{a}} m_X(\mathsf{b}) \right) \\
&= \sum_{\mathsf{b} \in 2^{\Omega_X}} m_X(\mathsf{b}) \left(\sum_{\mathsf{a} \in 2^{\Omega_X}:\, \mathsf{a} \subseteq \mathsf{b}} (-1)^{|\mathsf{a}|+1} \right) \\
&= \sum_{\mathsf{b} \in 2^{\Omega_X}} m_X(\mathsf{b}) = 1.
\end{aligned}$$

Finally, we will define a credal set representation of a belief function.

Definition 2.5 (Credal Set): Suppose m_X is a BPA for X. Let $\mathcal{P}$ denote the set of all PMFs for X. Then, the credal set $\mathcal{C}_{m_X}$ corresponding to BPA m_X for X is defined as follows:

$$\mathcal{C}_{m_X} = \left\{ P \in \mathcal{P} : \sum_{x \in \mathsf{a}} P(x) \geq Bel_{m_X}(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}) \text{ for all } \mathsf{a} \in 2^{\Omega_X} \right\}$$

$\mathcal{C}_{m_X}$ is a convex set, i.e. if $P, P' \in \mathcal{C}_{m_X}$, then $\lambda\, P + (1-\lambda)\, P' \in \mathcal{C}_{m_X}$ for all $\lambda \in [0,1]$. If $m_X = \iota_X$, then $\mathcal{C}_{\iota_X} = \mathcal{P}$. If m_X is Bayesian, then $\mathcal{C}_{m_X} = \{P_X\}$, where P_X is the PMF corresponding to Bayesian BPA m_X.

All five representations – BPA, plausibility, belief, commonality, and credal set – have exactly the same information. Given any one, we can transform it to another (Shafer 1976). For example, given a commonality function Q_{m_X}, we can recover m_X from Q_{m_X} as follows Shafer (1976):

$$m_X(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \, \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q_{m_X}(\mathsf{b})$$

However, they have different semantics. Most importantly, credal set semantics are incompatible with Dempster's combination rule (to be described next) (Shafer 1981, 1990, 1992; Halpern and Fagin 1992).

2.2. *Basic operations in the D–S theory*

There are two main operations in the D–S theory – Dempster's combination rule and marginalization.

Dempster's Combination Rule In the D–S theory, we can combine two BPAs m_1 and m_2 representing distinct pieces of evidence by Dempster's rule (Dempster 1967) and obtain the BPA $m_1 \oplus m_2$, which represents the combined evidence. Dempster referred to this rule as the product-intersection rule, as the product of the BPA values are assigned to the intersection of the focal elements, followed by normalization. Normalization consists of discarding the probability assigned to $\emptyset$, and normalizing the remaining values so that they add to 1. In general, Dempster's rule of combination can be used to combine two BPAs for arbitrary sets of variables.

Let $\mathcal{X}$ denote a finite set of variables. The state space of $\mathcal{X}$ is $\times_{X \in \mathcal{X}} \Omega_X$. Thus, if $\mathcal{X} = \{X, Y\}$ then the state space of $\{X, Y\}$ is $\Omega_X \times \Omega_Y$.

Projection of states simply means dropping extra coordinates; for example, if (x, y) is a state of $\{X, Y\}$, then the projection of (x, y) to X, denoted by $(x, y)^{\downarrow X}$, is simply x, which is a state of X.

Projection of subsets of states is achieved by projecting every state in the subset. Suppose $\mathsf{b} \in 2^{\Omega_{\{X,Y\}}}$. Then $\mathsf{b}^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{b}\}$. Notice that $\mathsf{b}^{\downarrow X} \in 2^{\Omega_X}$.

Vacuous extension of a subset of states of $\mathcal{X}_1$ to a subset of states of $\mathcal{X}_2$, where $\mathcal{X}_2 \supseteq \mathcal{X}_1$, is a cylinder set extension, i.e. if $\mathsf{a} \in 2^{\mathcal{X}_1}$, then $\mathsf{a}^{\uparrow \mathcal{X}_2} = \mathsf{a} \times \Omega_{\mathcal{X}_2 \setminus \mathcal{X}_1}$. Thus, if $\mathsf{a} \in 2^{\Omega_X}$, then $\mathsf{a}^{\uparrow \{X,Y\}} = \mathsf{a} \times \Omega_Y$.

Definition 2.6 (Dempster's rule using BPAs): Suppose m_1 and m_2 are BPAs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Then $m_1 \oplus m_2$ is a BPA for $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$, say, given by

$$(m_1 \oplus m_2)(\mathsf{a}) = K^{-1} \sum_{\mathsf{b}_1,\mathsf{b}_2 \in 2^{\Omega_{\mathcal{X}}}:\, \mathsf{b}_1 \cap \mathsf{b}_2 = \mathsf{a}} m_1(\mathsf{b}_1^{\downarrow \mathcal{X}_1})\, m_2(\mathsf{b}_2^{\downarrow \mathcal{X}_2}),$$

for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$, where K is a normalization constant given by

$$K = 1 - \sum_{\mathsf{b}_1,\mathsf{b}_2 \in 2^{\Omega_{\mathcal{X}}}:\, \mathsf{b}_1 \cap \mathsf{b}_2 = \emptyset} m_1(\mathsf{b}_1^{\downarrow \mathcal{X}_1})\, m_2(\mathsf{b}_2^{\downarrow \mathcal{X}_2}). \tag{1}$$

The definition of Dempster's rule assumes that the normalization constant K is non-zero. If $K = 0$, then the two BPAs m_1 and m_2 are said to be in *total conflict* and cannot be combined. If $K = 1$, we say m_1 and m_2 are *non-conflicting*.

Dempster's rule can also be defined in terms of commonality functions (Shafer 1976).

Definition 2.7 (Dempster's rule using commonality functions): Suppose Q_{m_1} and Q_{m_2} are commonality functions corresponding to BPAs m_1 and m_2, respectively. The commonality function $Q_{m_1 \oplus m_2}$ corresponding to BPA $m_1 \oplus m_2$ is as follows:

$$Q_{m_1 \oplus m_2}(\mathsf{a}) = K^{-1} Q_{m_1}(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_{m_2}(\mathsf{a}^{\downarrow \mathcal{X}_2}),$$

for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$, where the normalization constant K is as follows:

$$K = \sum_{\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}} (-1)^{|\mathsf{a}|+1} Q_{m_1}(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_{m_2}(\mathsf{a}^{\downarrow \mathcal{X}_2}). \tag{2}$$

It is shown in Shafer (1976) that the normalization constant K in Equation (2) is exactly the same as in Equation (1). In terms of commonality functions, Dempster's rule is pointwise multiplication of commonality functions followed by normalization.

Marginalization Marginalization in D–S theory is addition of values of BPAs.

Definition 2.8 (Marginalization): Suppose m is a BPA for $\mathcal{X}$. Then, the marginal of m for $\mathcal{X}_1$, where $\mathcal{X}_1 \subset \mathcal{X}$, denoted by $m^{\downarrow \mathcal{X}_1}$, is a BPA for $\mathcal{X}_1$ such that for each $\mathsf{a} \in 2^{\Omega_{\mathcal{X}_1}}$,

$$m^{\downarrow \mathcal{X}_1}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{\mathcal{X}}}:\, \mathsf{b}^{\downarrow \mathcal{X}_1} = \mathsf{a}} m(\mathsf{b}).$$

Marginalization can also be described using commonality functions. Suppose m is a BPA for $\mathcal{X}$, and suppose $\mathcal{X}_1 \subset \mathcal{X}$. Then, for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}_1}}$,

$$Q_{m^{\downarrow \mathcal{X}_1}}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{\mathcal{X}}}:\mathsf{b}^{\downarrow \mathcal{X}_1} = \mathsf{a}} (-1)^{(|\mathsf{b}|-|\mathsf{a}|)}\, Q_m(\mathsf{b}).$$

2.3. Conditional belief functions

In probability theory, it is common to construct joint probability mass functions for a set of discrete variables by using conditional probability distributions. For example, we can construct joint PMF for (X, Y) by first assessing PMF P_X of X, and conditional PMFs $P_{Y|x}$ for each $x \in \Omega_X$ such that $P_X(x) > 0$. The pointwise multiplication of $P_{Y|x}$ for all $x \in \Omega_X$ is called a CPT, and denoted by $P_{Y|X}$. Then, $P_{X,Y} = P_X \otimes P_{Y|X}$. We can construct joint BPA for $\{X, Y\}$ in a similar manner.

Consider a BPA m_X for X such that $m_X(\{x\}) > 0$. Suppose that there is a BPA for Y expressing our belief about Y if we know that $X = x$, and denote it by $m_{Y|x}$. Notice that $m_{Y|x} : 2^{\Omega_Y} \to [0, 1]$ is such that $\sum_{\mathsf{b} \in 2^{\Omega_Y}} m_{Y|x}(\mathsf{b}) = 1$. We can embed this conditional BPA for Y into a conditional BPA for $\{X, Y\}$, which is denoted by $m_{x,Y}$, such that the following two conditions hold. First, $m_{x,Y}$ tells us nothing about X, i.e. $m_{x,Y}^{\downarrow X}(\Omega_X) = 1$. Second, if we combine $m_{x,Y}$ with the deterministic BPA $m_{X=x}$ for X such $m_{X=x}(\{x\}) = 1$ using Dempster's rule, and marginalize the result to Y we obtain $m_{Y|x}$, i.e. $(m_{x,Y} \oplus m_{X=x})^{\downarrow Y} = m_{Y|x}$. One way to obtain such an embedding is suggested by Smets (1978) (see also Shafer 1982), called *conditional embedding*, and it consists of taking each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of $m_{Y|x}$, and converting it to a corresponding focal element of $m_{x,Y}$ (with the same mass) as follows: $(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y)$. It is easy to confirm that this method of embedding satisfies the two conditions mentioned above.

Example 2.1 (Conditional embedding): Consider discrete variables X and Y, with $\Omega_X = \{x, \bar{x}\}$ and $\Omega_Y = \{y, \bar{y}\}$. Suppose that m_X is a BPA for X such that $m_X(x) > 0$ and $m_X(\bar{x}) > 0$. If we have a conditional BPA $m_{Y|x}$ for Y given $X = x$ as follows:

$$m_{Y|x}(y) = 0.8, \text{ and}$$
$$m_{Y|x}(\Omega_Y) = 0.2,$$

then its conditional embedding into BPA $m_{x,Y}$ for $\{X, Y\}$ is as follows:

$$m_{x,Y}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) = 0.8, \text{ and}$$
$$m_{x,Y}(\Omega_{\{X,Y\}}) = 0.2.$$

Similarly, if we have a conditional BPA $m_{Y|\bar{x}}$ for Y given $X = \bar{x}$ as follows:

$$m_{Y|\bar{x}}(\bar{y}) = 0.3, \text{ and}$$
$$m_{Y|\bar{x}}(\Omega_Y) = 0.7,$$

then its conditional embedding into BPA $m_{\bar{x},Y}$ for $\{X, Y\}$ is as follows:

$$m_{\bar{x},Y}(\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}) = 0.3, \text{ and}$$
$$m_{\bar{x},Y}(\Omega_{\{X,Y\}}) = 0.7.$$

Assuming we have these two conditional BPAs, and their corresponding embeddings, it is clear that the two BPA $m_{x,Y}$ and $m_{\bar{x},Y}$ are distinct and can be combined with Dempster's rule of combination, resulting in the conditional BPA $m_{Y|X} = m_{x,Y} \oplus m_{\bar{x},Y}$ for $\{X, Y\}$ as follows:

$$
\begin{aligned}
m_{Y|X}(\{(x, y), (\bar{x}, \bar{y})\}) &= 0.24, \\
m_{Y|X}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) &= 0.56, \\
m_{Y|X}(\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}) &= 0.06, \text{ and} \\
m_{Y|X}(\Omega_{\{X,Y\}}) &= 0.14.
\end{aligned}
$$

$m_{Y|X}$ is the belief function equivalent to CPT $P_{Y|X}$ in probability theory. ■

This completes our brief review of the D–S belief function theory. For further details, the reader is referred to Shafer (1976).

3. Expected value of D–S belief functions

In this section, we provide a new definition of expected value of belief functions in the D–S theory, and describe its properties.

In probability theory, the expected value of a PMF can be interpreted as a "central tendency", for example as a center of gravity if the probability masses are interpreted as weights on locations on the real line. It is neither a pessimistic value nor an optimistic value. Our goal is to define expectation of BPA functions that has the semantics of central tendency.

As in the probabilistic case, we will assume that Ω_X is a finite set of real numbers. In a PMF, we have probabilities assigned to each state $x \in \Omega_X$. In a BPA m_X for X and its equivalent representations, we have probabilities assigned to subsets of states $\mathsf{a} \in 2^{\Omega_X}$. Before we define expectation of X with respect to BPA m_X, we will define a real-valued value function $v_{m_X} : 2^{\Omega_X} \to \mathbb{R}$ for all subsets in 2^{Ω_X}. If $\mathsf{a} = \{x\}$ is a singleton subset, then it is natural to assume $v_m(\{x\}) = x$. Remember that the elements of Ω_X are real numbers. For non-singleton subsets $\mathsf{a} \in 2^{\Omega_X}$, it makes sense to define $v_{m_X}(\mathsf{a})$ such that the following inequality holds:

$$\min \mathsf{a} \leq v_{m_X}(\mathsf{a}) \leq \max \mathsf{a}. \tag{3}$$

One way to satisfy the inequality in Equation (3) is to define v_{m_X} as follows:

$$v_{m_X}(\mathsf{a}) = \frac{\sum_{x \in \mathsf{a}} x \cdot Q_{m_X}(\{x\})}{\sum_{x \in \mathsf{a}} Q_{m_X}(\{x\})}, \quad \text{for all } \mathsf{a} \in 2^{\Omega_X}. \tag{4}$$

In words, the value function $v_{m_X}(\mathsf{a})$ is the weighted average of all $x \in \mathsf{a}$, where the weights are the commonality numbers $Q_{m_X}(\{x\})$, which are also the plausibility values $Pl_{m_X}(\{x\})$. The rationale for this definition is similar to the rationale for the plausibility transformation of a BPA to a corresponding probability distribution (Cobb and Shenoy 2006), namely the plausibility transformation is the only transformation that is consistent with Dempster's rule in the following sense. Suppose m is a BPA for X, suppose Q_m denotes the commonality function corresponding to BPA m, and suppose P_{Pl_m} denotes a probability mass function for X obtained from BPA m as follows:

$$P_{Pl_m}(x) = \frac{Q_m(\{x\})}{\sum_{y \in \Omega_X} Q_m(\{y\})}$$

Then, P_{Pl_m} is the only probability transformation of m that satisfies the following property: Suppose m_1 and m_2 are two distinct BPA functions for X. Then,

$$P_{Pl_{m_1 \oplus m_2}} = P_{Pl_{m_1}} \otimes P_{Pl_{m_2}}.$$

Here, $\otimes$ denotes the combination rule for probability mass functions, namely pointwise multiplication followed by normalization.

3.1. Definition of expected value

Definition 3.1 (Expected Value of X): Suppose m_X is a BPA for X with a real-valued state space Ω_X, and suppose Q_{m_X} denotes the commonality function corresponding to m_X. Then the expected value of X with respect to m_X, denoted by $E_{m_X}(X)$, is defined as follows:

$$E_{m_X}(X) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} v_{m_X}(\mathsf{a})\, Q_{m_X}(\mathsf{a}), \tag{5}$$

where $v_{m_X}(\mathsf{a})$ is as defined in Equation (4).

Example 3.1 (Vacuous BPA): Suppose $\Omega_X = \{1, 2, 3\}$, and suppose the uncertainty of X is described by the vacuous BPA ι_X for X. The commonality function Q_{ι_X} is identically 1 for all $\mathsf{a} \in 2^{\Omega_X}$. $v_{\iota_X}(\mathsf{a}) = (\sum\{x : x \in \mathsf{a}\})/|\mathsf{a}|$, where $|\mathsf{a}|$ denotes cardinality of a (see Table 1, empty cells in column 2 have 0 values). Thus, $E_{\iota_X}(X) = 2$. $E_{\iota_X}(X)$ coincides with the expected value of the pignistic and plausibility transforms.

Example 3.2 (Consonant BPA): Suppose $\Omega_X = \{1, 2, 3\}$, and suppose the uncertainty of X is described by a consonant BPA m_X as shown in Table 2 (empty cells in column 2 have 0 values). For this example, $E_{m_X}(X) = 2.317$, which is different from the expected value of the pignistic transform: 2.495, and the expected value of the plausibility transform: 2.328.

Example 3.3 (Quasi-consonant BPA): Suppose $\Omega_X = \{1, 2, 3\}$, and suppose the uncertainty of X is described by a quasi-consonant BPA m_X as shown in Table 3 (empty cells in columns 2–3 have 0 values). For this example, $E_{m_X}(X) = 2.00$, which is different from the

Table 1. Expected value of a vacuous BPA.

$\mathsf{a} \in 2^{\Omega_X}$	$m_X(\mathsf{a})$	$Q_{m_X}(\mathsf{a})$	$v_{m_X}(\mathsf{a})$	$E_{m_X}(X)$	$E_{BetP_{m_X}}(X)$	$E_{Pl_{P_{m_X}}}(X)$
$\{1\}$		1	1	2	2	2
$\{2\}$		1	2			
$\{3\}$		1	3			
$\{1, 2\}$		1	1.5			
$\{1, 3\}$		1	2			
$\{2, 3\}$		1	2.5			
$\{1, 2, 3\}$	1	1	2			

Table 2. Expected value of a consonant BPA.

$a \in 2^{\Omega_X}$	$m_X(a)$	$Q_{m_X}(a)$	$v_{m_X}(a)$	$E_{m_X}(X)$	$E_{BetP_{m_X}}(X)$	$E_{PlP_{m_X}}(X)$
{1}		0.34	1.00	2.317	2.495	2.328
{2}		0.67	2.00			
{3}	0.33	1.00	3.00			
{1, 2}		0.34	1.66			
{1, 3}		0.34	2.49			
{2, 3}	0.33	0.67	2.60			
{1, 2, 3}	0.34	0.34	2.33			

Table 3. Expected value of a quasi-consonant BPA.

$a \in 2^{\Omega_X}$	$m_X(a)$	$Q_{m_X}(a)$	$v_{m_X}(a)$	$E_{m_X}(X)$	$E_{BetP_{m_X}}(X)$	$E_{PlP_{m_X}}(X)$
{1}		1.0	1.00	2.00	1.75	1.75
{2}		0.5	2.00			
{3}		0.5	3.00			
{1, 2}	0.5	0.5	1.33			
{1, 3}	0.5	0.5	1.67			
{2, 3}			2.50			
{1, 2, 3}			1.75			

Table 4. Expected value of a general BPA.

$a \in 2^{\Omega_X}$	$m_X(a)$	$Q_{m_X}(a)$	$v_{m_X}(a)$	$E_{m_X}(X)$	$E_{BetP_{m_X}}(X)$	$E_{PlP_{m_X}}(X)$
{1}	0.5	0.5	1.00	1.75	1.75	2.00
{2}		0.5	2.00			
{3}		0.5	3.00			
{1, 2}			1.50			
{1, 3}			2.00			
{2, 3}	0.5	0.50	2.50			
{1, 2, 3}			2.00			

expected value of the pignistic transform: 1.75, and the expected value of the plausibility transform: 1.75.

Example 3.4 (General BPA): Suppose $\Omega_X = \{1, 2, 3\}$, and suppose the uncertainty of X is described by a BPA m_X as shown in Table 4 (empty cells in columns 2–3 have 0 values). For this example, $E_{m_X}(X) = 1.75$, which is same as the expected value of the pignistic transform: 1.75, but different from the expected value of the plausibility transform: 2.00.

3.2. Definition of expected value of real-valued functions

Definition 3.2 (Expected value of a real-valued function of real-valued variable X): Suppose Q_{m_X} is a commonality function for X corresponding to BPA m_X for X, and $g_X : \mathbb{R} \to \mathbb{R}$ is a well-defined real-valued function of X, then we define expected value of g_X with respect to m_X, denoted by $E_{m_X}(g_X)$ as follows:

$$E_{m_X}(g_X) = \sum_{a \in 2^{\Omega_X}} (-1)^{|a|+1}\, g_X(v_{m_X}(a))\, Q_{m_X}(a)$$

The variable X in Definition 3.2 is assumed to have a real-valued frame Ω_X. If Ω_X consists of categorical values, it suffices to have a real-valued function $g_X : 2^{\Omega_X} \to \mathbb{R}$, and we define $E_{m_X}(g_X)$ as follows:

Definition 3.3 (Expected value of a real-valued function of symbolic-valued variable X): Suppose Q_{m_X} is a commonality function for X corresponding to BPA m_X for X, and suppose $g_X : 2^{\Omega_X} \to \mathbb{R}$ is a well-defined real-valued function of X. Then, we define expected value of g_X with respect to m_X, denoted by $E_{m_X}(g_X)$ as follows:

$$E_{m_X}(g_X) = \sum_{\mathsf{a}\in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\, g_X(\mathsf{a})\, Q_{m_X}(\mathsf{a})$$

Suppose we have a real-valued function $h_X : \Omega_X \to \mathbb{R}$, whose domain is Ω_X (compared to 2^{Ω_X} in g_X in Definition 3.3). Suppose our uncertainty of X is defined by BPA m_X. What is the expected value of h_X, $E_{m_X}(h_X)$? Jiroušek and Kratochvíl (2018) suggest first extending h_X to $\hat{h}_X : 2^{\Omega_X} \to \mathbb{R}$ as follows:

$$\hat{h}_X(\mathsf{a}) = \frac{\sum_{x\in\mathsf{a}} h_X(x)\, Q_{m_X}(\{x\})}{\sum_{x\in\mathsf{a}} Q_{m_X}(\{x\})}$$

Notice that $\hat{h}_X(\mathsf{a})$ satisfies $\min_{x\in\mathsf{a}} h(x) \le \hat{h}_X(\mathsf{a}) \le \max_{x\in\mathsf{a}} h(x)$. Then, we can define $E_{m_X}(h_X) = E_{m_X}(\hat{h}_X)$, where $E_{m_X}(\hat{h}_X)$ is defined as in Definition 3.3, i.e.

$$E_{m_X}(h_X) = \sum_{\mathsf{a}\in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\, \hat{h}_X(\mathsf{a})\, Q_{m_X}(\mathsf{a})$$

At this stage, it may be useful to define Choquet integral of a real-valued function that has been used to describe a decision-making theory for lotteries that are described by belief functions interpreted as credal sets of PMFs (Gilboa and Schmeidler 1989).

Definition 3.4: Suppose we have a real-valued function $h : \Omega_X \to \mathbb{R}$. The Choquet integral of h with respect to BPA m_X is defined as follows. First we extend h to $h_{\min} : 2^{\Omega_X} \to \mathbb{R}$ as follows:

$$h_{\min}(\mathsf{a}) = \min_{x\in\mathsf{a}} h(x) \tag{6}$$

Then, we can define the Choquet integral of h with respect to BPA m_X for X, denoted by $E^{\min}_{m_X}(h)$, as follows:

$$E^{\min}_{m_X}(h) = \sum_{\mathsf{a}\in 2^{\Omega}_X} h_{\min}(\mathsf{a})\, m_X(\mathsf{a}) \tag{7}$$

It has been shown (see e.g. Gilboa and Schmeidler 1994) that:

$$E^{\min}_{m_X}(h) = \min_{P\in\mathcal{C}_{m_X}} \sum_{x\in\Omega_X} h(x)\, P(x).$$

If we regard h_X as a utility function for the states of X, then Equation (6) represents a pessimistic or ambiguity-averse attitude. Also, the definition of $E^{\min}_{m_X}(h_X)$ in Equation (7) is appropriate for belief functions interpreted as credal sets of PMFs, which are not compatible with Dempster's combination rule. For a vacuous belief function on $\Omega_X = \{1, 2, 3\}$, the Choquet integral $E^{\min}_{m_X}(X) = 1$, the smallest value in $\Omega_X = \{1, 2, 3\}$. Thus, the Choquet integral is inconsistent with the central tendency semantics of probabilistic expectation.

Definition 3.2 can be generalized to a multivariate function.

Definition 3.5 (Expected value of a real-valued function of X and Y): Suppose $Q_{m_{X,Y}}$ is a commonality function corresponding to BPA $m_{X,Y}$ for (X, Y), suppose m_X and m_Y are marginals of $m_{X,Y}$ for X and Y, respectively, and suppose $g_{X,Y} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a well-defined real-valued function of X and Y. Then we define expected value of $g_{X,Y}$ with respect to $m_{X,Y}$, denoted by $E_{m_{X,Y}}(g_{X,Y})$ as follows:

$$E_{m_{X,Y}}(g_{X,Y}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, g_{X,Y}(v_{m_X}(\mathsf{a}^{\downarrow X}), v_{m_Y}(\mathsf{a}^{\downarrow Y}))\, Q_{m_{X,Y}}(\mathsf{a})$$

Similarly, we can extend Definition 3.3 to multivariate functions. We skip the details.

3.3. Properties of expected value

Some important properties of our definition in Equation (5) are as follows. Consider the situation in Definition 3.1.

(1) (*Consistency with probabilistic expectation*) If m_X is a Bayesian BPA for X, and P_X is the PMF for X corresponding to m_X, i.e. $P_X(x) = m_X(\{x\})$ for all $x \in \Omega_X$, then $E_{m_X}(X) = E_{P_X}(X)$.

Proof: As m_X is Bayesian, $Q_{m_X}(\mathsf{a}) = m_X(\mathsf{a})$ if $|\mathsf{a}| = 1$, and $Q_{m_X}(\mathsf{a}) = 0$ if $|\mathsf{a}| > 1$. Also, $v_{m_X}(\{x\}) = x$. Thus, $E_{m_X}(X)$ in Equation (5) reduces to probabilistic expectation. ■

(2) (*Expected value of a constant*) If X is a constant, i.e. $m_X(\{a\}) = 1$, where a is a real constant, then $E_{m_X}(X) = a$.

Proof: Notice that in this case, m is Bayesian, and as this property holds for the probabilistic case, it also holds for the D–S theory from the consistency with probabilistic expectation property. ■

(3) (*Expected value of a linear function of* X) Suppose $Y = g_X = a\,X + b$, then $E_{m_Y}(Y)$ can be computed as follows:

$$E_{m_Y}(Y) = E_{m_X}(g_X) = a\,E_{m_X}(X) + b. \tag{8}$$

In probability theory, this property is valid for any well-defined function of X. Our definition does not satisfy this property for any well-defined function (see Examples 3.5 and 3.6 that follow), but it is satisfied only for a linear function of X. This property allows us to compute the expected value of $Y = g_X$ without first computing its commonality function.

Proof: Suppose $a = 0$, then the result follows from the expected value of a constant property. Suppose $a \neq 0$. In this case, g_X is a 1-1 function. Therefore, $\Omega_Y = \{g_X(x) : x \in \Omega_X\}$. Thus, the values of the commonality function Q_{m_Y} for Y are the same as the corresponding values of the commonality function Q_m for X, i.e. $Q_{m_Y}(\mathsf{a}_Y) = Q_{m_X}(\mathsf{a})$, where $\mathsf{a}_Y \in 2^{\Omega_Y}$ is the subset that corresponds to subset a of Ω_X, i.e. $\mathsf{a}_Y = \{g_X(x) : x \in \mathsf{a}\}$. It suffices to show that $v_{m_Y}(\mathsf{a}_Y) = g(v_m(\mathsf{a}))$ for all $\mathsf{a} \in 2^{\Omega_X}$. Suppose $Y = g_X = aX + b$.

$$\begin{aligned}
v_{m_Y}(\mathsf{a}_Y) &= \frac{\sum_{y \in \mathsf{a}_Y} y \cdot Q_{m_Y}(\{y\})}{\sum_{y \in \mathsf{a}_Y} Q_{m_Y}(\{y\})} \\
&= \frac{\sum_{x \in \mathsf{a}} (ax + b) \cdot Q_{m_X}(\{x\})}{\sum_{x \in \mathsf{a}} Q_{m_X}(\{x\})} \\
&= a \frac{\sum_{x \in \mathsf{a}} x \cdot Q_{m_X}(\{x\})}{\sum_{x \in \mathsf{a}} Q_{m_X}(\{x\})} + b \\
&= a v_{m_X}(\mathsf{a}) + b \\
&= g_X(v_m(\mathsf{a})).
\end{aligned}$$

Thus, $E_{m_Y}(Y) = E_{m_X}(g_X)$. Next, using Definition 3.2,

$$\begin{aligned}
E_{m_X}(g_X) &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} (a v_{m_X}(\mathsf{a}) + b)\, Q_{m_X}(\mathsf{a}) \\
&= a \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} v_{m_X}(\mathsf{a})\, Q_{m_X}(\mathsf{a}) \\
&\quad + b \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} Q_{m_X}(\mathsf{a}) \\
&= a\, E_{m_X}(X) + b.
\end{aligned}$$

■

(4) (*Expected value of a function of* X and Y) The law of the unconscious statistician does not generalize to the multidimensional case, even for linear functions. Suppose X and Y are discrete random variables with state spaces Ω_X and Ω_Y, respectively, with joint BPA $m_{X,Y}$ for (X, Y), whose marginals for X and Y are m_X and m_Y, respectively. If $Z = g_{X,Y} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a *linear* function of (X, Y), then, in general,

$$E_{m_Z}(Z) \neq E_{m_{X,Y}}(g_{X,Y}),$$

where $E_{m_{X,Y}}(g_{X,Y})$ is as defined in Definition 3.5. A counter-example is given in Example 3.8.

(5) (*Expected value of a linear function of* X and Y) If $g_{X,Y} = a\,X + b\,Y + c$, where a, b, and c are real constants, and $m_{X,Y}$ is a joint BPA for (X, Y), then

$$E_{m_{X,Y}}(g_{X,Y}) = a\,E_{m_X}(X) + b\,E_{m_Y}(Y) + c$$

Proof: By Definition 3.5,

$$
\begin{aligned}
E_{m_{X,Y}}(g_{X,Y}) &= \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} \left(a\, v_{m_X}(\mathsf{a}^{\downarrow X}) + b\, v_{m_Y}(\mathsf{a}^{\downarrow Y}) + c\right) Q_{m_{X,Y}}(\mathsf{a}) \\
&= a \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} v_{m_X}(\mathsf{a}^{\downarrow X})\, Q_{m_{X,Y}}(\mathsf{a}) \\
&\quad + b \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} v_{m_Y}(\mathsf{a}^{\downarrow Y})\, Q_{m_{X,Y}}(\mathsf{a}) + c \\
&= a \sum_{\mathsf{a}_X\in 2^{\Omega_X}} (-1)^{|\mathsf{a}_X|+1}\, v_{m_X}(\mathsf{a}_X) \sum_{\mathsf{b}\in 2^{\Omega_{X,Y}}:\mathsf{b}^{\downarrow X}=\mathsf{a}_X} (-1)^{\mathsf{b}\setminus \mathsf{a}_X} Q_{m_{X,Y}}(\mathsf{b}) \\
&\quad + b \sum_{\mathsf{a}_Y\in 2^{\Omega_Y}} (-1)^{|\mathsf{a}_Y|+1}\, v_{m_Y}(\mathsf{a}_Y) \sum_{\mathsf{b}\in 2^{\Omega_{X,Y}}:\mathsf{b}^{\downarrow Y}=\mathsf{a}_Y} (-1)^{\mathsf{b}\setminus \mathsf{a}_Y} Q_{m_{X,Y}}(\mathsf{b}) + c \\
&= a \sum_{\mathsf{a}_X\in 2^{\Omega_X}} (-1)^{|\mathsf{a}_X|+1} v_{m_X}(\mathsf{a}_X) Q_{m_X}(\mathsf{a}_X) \\
&\quad + b \sum_{\mathsf{a}_Y\in 2^{\Omega_Y}} (-1)^{|\mathsf{a}_Y|+1} v_{m_Y}(\mathsf{a}_Y) Q_{m_Y}(\mathsf{a}_Y) + c \\
&= a\, E_{m_X}(X) + b\, E_{m_Y}(Y) + c
\end{aligned}
$$

■

(6) (*Bounds on expected value*) $\min \Omega_X \le E_{m_X}(X) \le \max \Omega_X$.

Proof: Suppose $|\Omega_X| = n$. The expected value $E_{m_X}(X)$ is a function of Ω_X and m_X. It can be regarded as a function $E_{m_X}(X) : \mathcal{R}^n \times [0,1]^{2^n} \to \mathcal{R}$. From the *expected value of a linear function of* X and Y property, it follows that $E_{m_X}(X)$ is a convex function. Any m_X for X can be considered as a convex combination of $\{m_{\mathsf{a}} : \mathsf{a} \in 2^{\Omega_X}\}$, where m_{a} is a deterministic BPA for X such that $m_{\mathsf{a}}(\mathsf{a}) = 1$, and where the weight associated with m_{a} is $m_X(\mathsf{a})$. Let $x_{\min}$ denote $\min \Omega_X$ and $x_{\max}$ denote $\max \Omega_X$. Notice that if $m_X = m_{\{x_{\min}\}}$, then $E_{m_X}(X) = x_{\min}$, and if $m_X = m_{\{x_{\max}\}}$, then $E_{m_X}(X) = x_{\max}$. For all other m_{a}, $\min \Omega_X \le E_{m_{\mathsf{a}}}(X) \le \max \Omega_X$. From Jensen's inequality (Rockafellar 1970), the result follows. ■

(7) (*Expected value of a product of independent random variables*) Suppose X and Y are random variables with joint BPA $m_{X,Y} = m_X \oplus m_Y$, where m_X and m_Y are BPAs for X and Y, respectively. Suppose $W = g_{X,Y} = X \cdot Y$. Then, $E_{m_{X,Y}}(g_{X,Y}) = E_{m_X}(X) \cdot E_{m_Y}(Y)$.

Proof: Notice that $Q_{m_{X,Y}}(\mathsf{a}) = Q_{m_X}(\mathsf{a}^{\downarrow X}) \cdot Q_{m_Y}(\mathsf{a}^{\downarrow Y})$ for all $\mathsf{a} \in 2^{\Omega_{X,Y}}$. Also, $g_{X,Y}(v_{m_X}(\mathsf{a}^{\downarrow X}), v_{m_Y}(\mathsf{a}^{\downarrow Y})) = v_{m_X}(\mathsf{a}^{\downarrow X}) \cdot v_{m_Y}(\mathsf{a}^{\downarrow Y})$ for all $\mathsf{a} \in 2^{\Omega_{X,Y}}$. Notice that for a given $\mathsf{a}_X \in 2^{\Omega_X}$ and a given $\mathsf{a}_Y \in 2^{\Omega_Y}$, $\sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}:\mathsf{a}^{\downarrow X}=\mathsf{a}_X,\, \mathsf{a}^{\downarrow Y}=\mathsf{a}_Y} (-1)^{|\mathsf{a}|+1} = (-1)^{|\mathsf{a}_X|+1}$

$(-1)^{|\mathsf{a}_Y|+1}$. Thus,

$$\begin{aligned}
E_{m_{X,Y}}(g_{X,Y}) &= \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} g_{X,Y}(v_{m_X}(\mathsf{a}^{\downarrow X}), v_{m_Y}(\mathsf{a}^{\downarrow Y}))\, Q_{m_{X,Y}}(\mathsf{a}) \\
&= \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} \left(v_{m_X}(\mathsf{a}^{\downarrow X}) \cdot v_{m_Y}(\mathsf{a}^{\downarrow Y})\right) \left(Q_{m_X}(\mathsf{a}^{\downarrow X}) \cdot Q_{m_Y}(\mathsf{a}^{\downarrow Y})\right) \\
&= \left(\sum_{\mathsf{a}_X\in 2^{\Omega_X}} v_{m_X}(\mathsf{a}_X)\, Q_{m_X}(\mathsf{a}_X)\right) \left(\sum_{\mathsf{a}_Y\in 2^{\Omega_Y}} v_{m_Y}(\mathsf{a}_Y)\, Q_{m_Y}(\mathsf{a}_Y)\right) \\
&\quad \times \left(\sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}:\mathsf{a}^{\downarrow X}=\mathsf{a}_X,\, \mathsf{a}^{\downarrow Y}=\mathsf{a}_Y} (-1)^{|\mathsf{a}|+1}\right) \\
&= \left(\sum_{\mathsf{a}^{\downarrow X}\in 2^{\Omega_X}} (-1)^{|\mathsf{a}^{\downarrow X}|+1}\, v_{m_X}(\mathsf{a}^{\downarrow X})\, Q_{m_X}(\mathsf{a}^{\downarrow X})\right) \\
&\quad \times \left(\sum_{\mathsf{a}^{\downarrow Y}\in 2^{\Omega_Y}} (-1)^{|\mathsf{a}^{\downarrow Y}|+1}\, v_{m_Y}(\mathsf{a}^{\downarrow Y})\, Q_{m_Y}(\mathsf{a}^{\downarrow Y})\right) \\
&= E_{m_X}(X) \cdot E_{m_Y}(Y).
\end{aligned}$$

■

Example 3.5 (Non 1-1 function): Consider a real-valued variable X with $\Omega_X = \{-1, 0, 1\}$, and suppose m_X is a BPA for X as shown in Table 5. Suppose $Y = g_X = X^2$. Notice that this function is not 1-1. Then, $\Omega_Y = \{1, 0\}$, and m_Y is as shown in Table 5. For this example, $E_{m_X}(g_X) = 1.188$, and $E_{m_Y}(Y) = 0.576$. Thus, Equation (8) does not hold. The Choquet integral $E^{\min}_{m_X}(g(X)) = 0.720$, and $E^{\min}_{m_Y}(Y) = 0.30$. Thus, $E^{\min}_{m_X}(g(X)) \neq E^{\min}_{m_Y}(Y)$.

Example 3.6 (Nonlinear 1-1 function): Consider a real-valued variable X with $\Omega_X = \{1, 2, 3\}$, and suppose m_X is a BPA for X as shown in Table 6. Suppose $Y = g_X = \log(X)$. Then, $\Omega_Y = \{\log(1), \log(2), \log(3)\} \approx \{0, 0.30, 0.48\}$, and m_Y is as shown in Table 6.

Table 5. Expected value of a function $Y = g_X = X^2$ that is not 1-1.

$\mathsf{a} \in 2^{\Omega_X}$	$m_X(\mathsf{a})$	$Q_{m_X}(\mathsf{a})$	$v_{m_X}(\mathsf{a})$	$E_{m_X}(X)$	$(v_{m_X}(\mathsf{a}))^2$	$E_{m_X}(g_X)$	$E^{\min}_{m_X}(g_X)$
$\{-1\}$	0.02	0.63	−1.00	0.059	1.00	1.188	0.720
$\{0\}$	0.05	0.70	0.00		0.00		
$\{1\}$	0.09	0.81	1.00		1.00		
$\{-1, 0\}$	0.12	0.42	−0.47		0.22		
$\{-1, 1\}$	0.19	0.49	0.13		0.02		
$\{0, 1\}$	0.23	0.53	0.54		0.29		
$\{-1, 0, 1\}$	0.30	0.30	0.08		0.01		
$\mathsf{b} \in 2^{\Omega_Y}$	$m_Y(\mathsf{b})$	$Q_{m_Y}(\mathsf{b})$	$v_{m_Y}(\mathsf{b})$	$E_{m_Y}(Y)$	$E^{\min}_{m_Y}(Y)$		
$\{1\}$	0.30	0.95	1.00	0.576	0.300		
$\{0\}$	0.05	0.70	0.00				
$\{1, 0\}$	0.65	0.65	0.58				

Table 6. Expected value of $Y = g_X = \log(X)$, a nonlinear 1-1 function.

$a \in 2^{\Omega_X}$	$m_X(a)$	$Q_{m_X}(a)$	$v_{m_X}(a)$	$E_{m_X}(X)$	$\log(v_{m_X}(a))$	$E_{m_X}(g_X)$	$E^{\min}_{m_X}(g_X)$
{1}	0.02	0.63	1.00	2.059	0.00	0.241	0.127
{2}	0.05	0.70	2.00		0.30		
{3}	0.09	0.81	3.00		0.48		
{1, 2}	0.12	0.42	1.53		0.18		
{1, 3}	0.19	0.49	2.12		0.33		
{2, 3}	0.23	0.53	2.53		0.40		
{1, 2, 3}	0.30	0.30	2.08		0.32		
$a_Y \in 2^{\Omega_Y}$	$m_Y(a_Y)$	$Q_{m_Y}(a_Y)$	$v_{m_Y}(a_Y)$	$E_{m_Y}(Y)$	$E^{\min}_{m_Y}(Y)$		
{0}	0.02	0.63	0.00	0.273	0.127		
{0.30}	0.05	0.70	0.30				
{0.48}	0.09	0.81	0.48				
{0, 0.30}	0.12	0.42	0.16				
{0, 0.48}	0.19	0.49	0.27				
{0.30, 048}	0.23	0.53	0.40				
{0, 0.30, 0.48}	0.30	0.30	0.28				

As the function is 1-1, the values of m_Y are the same as the values of m_X. For this example, $E_{m_Y}(Y) = 0.273$, and $E_{m_X}(\log(X)) = 0.241$. Thus, Equation (8) does not hold. The Choquet integrals are as follows: $E^{\min}_{m_X}(g(X)) = 0.127$, and $E^{\min}_{m_Y}(Y) = 0.127$. Thus, $E^{\min}_{m_X}(g(X)) = E^{\min}_{m_Y}(Y)$. As $Y = g(X) = \log(X)$ is a 1-1 function, m_X and m_Y have the same values. Also $h_{\min}$ values are also the same.

Example 3.7 (Linear function of X): Consider a real-valued variable X with $\Omega_X = \{-1, 0, 1\}$, and suppose m_X is a BPA for X as shown in Table 7. Suppose $Y = g_X = 2\,X + 1$. Then, $\Omega_Y = \{-1, 1, 3\}$, and m_Y is as shown in Table 7. Notice that as a linear function is 1-1, the values of m_Y are the same as the corresponding values of m_X. Also, notice that as the function g_X is linear, $g(v_m(\mathsf{a})) = v_{m_Y}(\mathsf{a}_Y)$, where subset a_Y corresponds to subset a. For this example, $E_{m_Y}(Y) = 1.117$, and $E_{m_X}(g_X) = 1.117$. Thus, Equation (8) holds. Also notice that $E_{m_X}(g_X) = E_{m_X}(2\,X + 1) = 2\,E_{m_X}(X) + 1 = 2(0.059) + 1 = 1.117$. The Choquet integrals are as follows: $E^{\min}_{m_X}(X) = -0.54$ (not shown in Table 7, $E^{\min}_{m_X}(g(X)) =$

Table 7. Expected value of $Y = g_X = 2X + 1$, a linear 1-1 function.

$a \in 2^{\Omega_X}$	$m_X(a)$	$Q_{m_X}(a)$	$v_{m_X}(a)$	$E_{m_X}(X)$	$2\,v_{m_X}(a) + 1$	$E_{m_X}(g_X)$	$E^{\min}_{m_X}(g_X)$
{−1}	0.02	0.63	−1.00	0.059	−1.00	1.117	−0.080
{0}	0.05	0.70	0.00		1.00		
{1}	0.09	0.81	1.00		3.00		
{−1, 0}	0.12	0.42	−0.47		0.05		
{−1, 1}	0.19	0.49	0.12		1.24		
{0, 1}	0.23	0.53	0.53		2.07		
{−1, 0, 1}	0.30	0.30	0.08		1.16		
$a_Y \in 2^{\Omega_Y}$	$m_Y(a_Y)$	$Q_{m_Y}(a_Y)$	$v_{m_Y}(a_Y)$	$E_{m_Y}(Y)$	$E^{\min}_{m_Y}(Y)$		
{−1}	0.02	0.63	−1.00	1.117	−0.080		
{1}	0.05	0.70	1.00				
{3}	0.09	0.81	3.00				
{−1, 1}	0.12	0.42	0.05				
{−1, 3}	0.19	0.49	1.24				
{1, 3}	0.23	0.53	2.07				
{−1, 1, 3}	0.30	0.30	1.16				

-0.080, and $E^{\min}_{m_Y}(Y) = -0.080$. As $Y = 2X + 1$ is a 1-1 function, $E^{\min}_{m_X}(g(X)) = E^{\min}_{m_Y}(Y)$. Also, $E^{\min}_{m_X}(g(X)) = 2\,E^{\min}_{m_X}(X) + 1 = 2(-0.54) + 1 = -0.080$.

Example 3.8 (Linear Function of X and Y): Consider a real-valued variable X with $\Omega_X = \{1, 3\}$, and a real-valued variable Y with $\Omega_Y = \{2, 4\}$. Suppose $m_{X,Y}$ is a BPA for (X, Y), with marginal BPAs m_X for X, and m_Y for Y, as shown in Table 8. Suppose $Z = g_{X,Y} = X + 3Y + 5$. Then $E_{m_Z}(Z) = 16.044$ as shown in Table 9, and $E_{m_{X,Y}}(g_{X,Y}) = 16.046$ as shown in Table 8. Thus, $E_{m_Z}(Z) \neq E_{m_{X,Y}}(g_{X,Y})$. However, notice that $E_{m_{X,Y}}(g_{X,Y}) = E_{m_X}(X) + 3E_{m_Y}(Y) + 5$. This can be seen from Table 8 as $16.046 = 2.015 + 3(3.010) + 5$.

Table 8. Expected value of $g_{X,Y} = X + 3Y + 5$.

$a \in 2^{\Omega_{X,Y}}$	$m_{X,Y}(a)$	$Q_{m_{X,Y}}(a)$	$g_{X,Y}(v_{m_X}(a^{\downarrow X}), v_{m_Y}(a^{\downarrow Y}))$	$E_{m_{X,Y}}(g_{X,Y})$
$\{(1,2)\}$	0.003	0.636	12.000	16.046
$\{(1,4)\}$	0.009	0.673	18.000	
$\{(3,2)\}$	0.015	0.691	14.000	
$\{(3,4)\}$	0.021	0.703	20.000	
$\{(1,2),(1,4)\}$	0.034	0.385	15.030	
$\{(3,2),(3,4)\}$	0.040	0.440	17.030	
$\{(1,2),(3,2)\}$	0.052	0.416	13.016	
$\{(1,4),(3,4)\}$	0.058	0.446	19.016	
$\{(1,2),(3,4)\}$	0.070	0.453	16.046	
$\{(1,4),(3,2)\}$	0.089	0.459	16.046	
$\{(1,2),(1,4),(3,2)\}$	0.095	0.239	16.046	
$\{(1,2),(1,4),(3,4)\}$	0.113	0.257	16.046	
$\{(1,2),(3,2),(3,4)\}$	0.125	0.269	16.046	
$\{(1,2),(3,2),(3,4)\}$	0.131	0.275	16.046	
$\Omega_{X,Y}$	0.144	0.144	16.046	
$a_X \in 2^{\Omega_X}$	$m_X(a_X)$	$Q_{m_X}(a_X)$	$v_{m_X}(a_X)$	$E_{m_X}(X)$
$\{1\}$	0.046	0.924	1.000	2.016
$\{3\}$	0.076	0.954	3.000	
$\{1,3\}$	0.878	0.878	2.016	
$a_Y \in 2^{\Omega_Y}$	$m_Y(a_Y)$	$Q_{m_Y}(a_Y)$	$v_{m_Y}(a_Y)$	$E_{m_Y}(Y)$
$\{2\}$	0.070	0.911	2.000	3.010
$\{4\}$	0.089	0.930	4.000	
$\{2,4\}$	0.841	0.841	3.010	

Table 9. Expected value $E_{m_Z}(Z)$, where $Z = g_{X,Y} = X + 3Y + 5$, and $m_{X,Y}$ is as in Table 8.

$a \in 2^{\Omega_Z}$	$m_Z(a)$	$Q_{m_Z}(a)$	$v_{m_Z}(a)$	$E_{m_Z}(Z)$
$\{12\}$	0.003	0.636	12.000	16.044
$\{18\}$	0.009	0.673	18.000	
$\{14\}$	0.015	0.691	14.000	
$\{20\}$	0.021	0.703	20.000	
$\{12, 18\}$	0.034	0.385	15.084	
$\{14, 20\}$	0.040	0.440	17.026	
$\{12, 14\}$	0.052	0.416	13.041	
$\{18, 20\}$	0.058	0.446	19.022	
$\{12, 20\}$	0.070	0.453	16.201	
$\{18, 14\}$	0.089	0.459	15.973	
$\{12, 18, 14\}$	0.095	0.239	14.709	
$\{12, 18, 20\}$	0.113	0.257	16.802	
$\{12, 14, 20\}$	0.125	0.269	15.452	
$\{18, 14, 20\}$	0.131	0.275	17.343	
Ω_Z	0.144	0.144	16.086	

Table 10. Expected value $E_{m_{X,Y}}(Z)$, where $Z = g_{X,Y} = X \cdot Y$.

$a_X \in 2^{\Omega_X}$	$m_X(a_X)$	$Q_{m_X}(a_X)$	$v_{m_X}(a_X)$	$E_{m_X}(X)$
{1}	0.11	0.67	1.00	2.143
{3}	0.33	0.89	3.00	
{1, 3}	0.56	0.56	2.14	
$a_Y \in 2^{\Omega_Y}$	$m_Y(a_Y)$	$Q_{m_Y}(a_Y)$	$v_{m_Y}(a_Y)$	$E_{m_Y}(Y)$
{2}	0.23	0.65	2.00	3.085
{4}	0.35	0.77	4.00	
{2, 4}	0.42	0.42	3.08	
$a \in 2^{\Omega_{X,Y}}$	$m_{X,Y}(a)$	$Q_{m_{X,Y}}(a)$	$g_{X,Y}(v_{m_X}(a^{\downarrow X}), v_{m_Y}(a^{\downarrow Y}))$	$E_{m_{X,Y}}(g_{X,Y})$
{(1, 2)}	0.03	0.44	2.00	6.604
{(1, 4)}	0.04	0.52	4.00	
{(3, 2)}	0.08	0.58	6.00	
{(3, 4)}	0.12	0.69	12.00	
{(1, 2), (1, 4)}	0.05	0.28	3.08	
{(3, 2), (3, 4)}	0.14	0.37	9.25	
{(1, 2), (3, 2)}	0.13	0.36	4.28	
{(1, 4), (3, 4)}	0.20	0.43	8.56	
{(1, 2), (3, 4)}		0.24	6.60	
{(1, 4), (3, 2)}		0.24	6.60	
{(1, 2), (1, 4), (3, 2)}		0.24	6.60	
{(1, 2), (1, 4), (3, 4)}		0.24	6.60	
{(1, 2), (3, 2), (3, 4)}		0.24	6.60	
{(1, 2), (3, 2), (3, 4)}		0.24	6.60	
$\Omega_{X,Y}$	0.24	0.24	6.60	

Example 3.9 (Product Function): Consider a real-valued variable X with $\Omega_X = \{1, 3\}$, and suppose m_X is a BPA for X as shown in Table 10. Expected value of X is ≈ 2.14. Consider Y with $\Omega_Y = \{2, 4\}$, and suppose m_Y is as shown in Table 10. Expected value of Y is ≈ 3.09. Assuming X and Y are independent, $m_{X,Y} = m_X \oplus m_Y$, as shown in Table 10. Suppose $Z = g_{X,Y} = X \cdot Y$. Then, the computation of $E_{m_{X,Y}}(Z) \approx 6.62 = E_{m_X}(X) \cdot E_{m_Y}(Y)$. Details of the computation of $E_{m_{X,Y}}(Z)$ are shown in Table 10.

3.4. Comparison with expected values using pignistic and plausibility transformations

As we said earlier, a traditional method of computing expectations of random variables characterized by a D–S BPA is to first transform the BPA to a PMF, and then use the probabilistic expectation operator. There are several methods of transforming a BPA to a PMF. Here we focus on the pignistic (Smets 2002) and the plausibility (Cobb and Shenoy 2006) transforms.

As D–S theory is a generalization of probability theory, there is, in general, more information in a BPA m than in the corresponding transform of m to a PMF. Thus, by computing expectation of X whose uncertainty is described by BPA m by first transforming m to a pignistic PMF $BetP_m$ or to a plausibility PMF PlP_m, there may be loss of information.

In general, the expected value defined in this paper will yield different values than the probabilistic expectation using pignistic or plausibility transformation. Table 11 compares the expectation defined in this paper with probabilistic expectation using pignistic and plausibility transforms for the various BPAs described in Tables 1–7. Two observations. First, although the three definitions yield different answers, they are all in the same range

Table 11. A comparison of our expected value with probabilistic expectation using pignistic and plausibility transforms.

BPA m	$E_m(\cdot)$	$E_{BetP_m}(\cdot)$	$E_{P_{Pl_m}}(\cdot)$
m_X in Table 1 (vacuous)	2.00	2.00	2.00
m_X in Table 2 (consonant)	2.317	2.495	2.328
m_X in Table 3 (quasi-consonant)	2.00	1.75	1.75
m_X in Table 4 (general)	1.75	1.75	2.00
m_X in Table 5 (X^2)	0.059	0.125	0.084
m_Y in Table 5 ($Y = X^2$)	0.576	0.625	0.576
m_X in Table 6 ($\log(X)$)	2.059	2.125	2.084
m_Y in Table 6 ($Y = \log(X)$)	0.273	0.289	0.278
m_Y in Table 7 ($Y = 2X + 1$)	1.117	1.250	1.168

of values, $[\min \Omega_X, \max \Omega_X]$. Second, all three definitions satisfy the *expected value of a function of* X property. Thus, BPA m_Z in Table 6 can be obtained from BPA m_X in Table 5 using the transformation $Z = X + 2$. All three expected values satisfy the *expected value of a function of* X property. Also, BPA m_Y in Table 7 is obtained from BPA m_X in Table 5 using the transformation $Y = 2X + 1$. Again, all three expected values satisfy the *expected value of a linear function of* X property.

A question is whether there exists a transformation from BPA to a PMF that provides the same expected value as per our definition. We conjecture that such a transformation does not exist, and we do not have a proof.

3.5. Comparison with Choquet integrals

In this section, we compare our definition of expected value with the Choquet integral definition as given in Definition 3.4. Table 12 compares the expectation defined in this paper with Choquet integrals for the various BPAs described in Tables 1–7. In general, the expected value defined in this paper will yield different values than the Choquet integral. First, the expected value defined in this paper is for the D–S theory, whereas the Choquet integral is for the theory of belief functions interpreted as credal sets. Second, the expected value defined in this paper is a “central tendency”, whereas the Choquet integral is a pessimistic (lower bound) value. In this respect, we conjecture that $E^{\min}_{m_X}(X) \leq E_{m_X}(X)$, and we do not have a proof of this assertion. Some results in Coletti, Petturiti, and Vantaggi (2019) may be useful in proving this conjecture if it is true.

Table 12. A comparison of our expected value with Choquet integral.

BPA m	$E_m(\cdot)$	$E^{\min}_m(\cdot)$
m_X in Table 1 (vacuous)	2.00	1.00
m_X in Table 2 (consonant)	2.317	1.990
m_X in Table 3 (quasi-consonant)	2.00	1.00
m_X in Table 4 (general)	1.75	1.50
m_X in Table 5 (X^2)	1.188	0.720
m_Y in Table 5 ($Y = X^2$)	0.576	0.300
m_X in Table 6 ($\log(X)$)	2.059	0.127
m_Y in Table 6 ($Y = \log(X)$)	0.273	0.127
m_X in Table 7 ($2X + 1$)	1.117	−0.080
m_Y in Table 7 ($Y = 2X + 1$)	1.117	−0.080

For Bayesian belief functions, our expected value $E_{m_X}(X)$ will coincide with the Choquet integral for the function $h(X) = X$. In this case, the credal set consists of a single PMF corresponding to the Bayesian BPA.

The Choquet integral has the nice property that for 1–1 functions $Y = g(X)$, $E^{\min}_{m_X}(g(X)) = E^{\min}_{m_Y}(Y)$ regardless of whether the function $Y = g(X)$ is linear or not. Our definition satisfies this property for linear functions only.

4. Variance of D–S belief functions

In probability theory, the variance of a random variable X whose uncertainty is described by PMF P_X is defined as $V_{P_X}(X) = E_{P_X}((X - E_{P_X}(X))^2)$. Using the definition of expectation of a real-valued function in Definition 3.2, we can define variance of a belief function in a similar manner.

4.1. Definition of variance

Definition 4.1 (Variance): Suppose X is a random variable with real-valued state space Ω_X, and whose uncertainty is described by BPA m_X for X. Then, the variance of X with respect to BPA m_X, denoted by $V_{m_X}(X)$, is defined as follows:

$$V_{m_X}(X) = E_{m_X}((X - E_{m_X}(X))^2) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} (v_{m_X}(\mathsf{a}) - E_{m_X}(X))^2 \, Q_{m_X}(\mathsf{a})$$

To simplify notation, when it is clear which BPA is used to compute the expected value and variance, we will write $v(a)$ instead of $v_{m_X}(\mathsf{a})$, $E(X)$ in place of $E_{m_X}(X)$, and $V(X)$ instead of $V_{m_X}(X)$.

Example 4.1 (Vacuous BPA): Consider Example 3.1 where we have X with real-valued state space Ω_X and with vacuous BPA ι_X for X as shown in Table 1. Then, $V(X) = 1.5$. The details of the computation are shown in Table 13 (empty cells in columns 2 and 6 have 0 values).

4.2. Properties of variance

Consider the situation in Definition 4.1.

(1) (*Consistency with probabilistic variance*) Suppose m_X is Bayesian BPA for X. Let P_X denote the PMF corresponding to m_X, i.e. $P_X(x) = m_X(\{x\})$. Then, $V_{m_X}(X) = V_{P_X}(X)$.

Table 13. Variance of a vacuous BPA.

$\mathsf{a} \in 2^{\Omega_X}$	$m_X(\mathsf{a})$	$Q_{m_X}(\mathsf{a})$	$v(\mathsf{a})$	$E(X)$	$(v(\mathsf{a}) - E(X))^2$	$V(X)$
{1}		1	1	2	1	1.5
{2}		1	2			
{3}		1	3		1	
{1, 2}		1	1.5		0.25	
{1, 3}		1	2			
{2, 3}		1	2.5		0.25	
{1, 2, 3}	1	1	2			

Proof: As m_X is Bayesian, for singleton subsets $\{x\} \in 2^{\Omega_X}$, $Q_{m_X}(\{x\}) = m_X(\{x\}) = P_X(x)$, and $v_{m_X}(\{x\}) = x$. Also, from the *consistency with probabilistic expectation* property, $E_{m_X}(X) = E_{P_X}(X)$. Finally, for non-singleton subsets $\mathsf{a} \in 2^{\Omega_X}$, $Q_{m_X}(\mathsf{a}) = 0$. Thus, the result follows from the definition of $V_{m_X}(X)$. ■

(2) (*Computation of variance*) As in the probabilistic case, we can compute the variance as follows:

$$V_{m_X}(X) = E_{m_X}(X^2) - (E_{m_X}(X))^2$$

where $E_{m_X}(X^2)$ is as defined as follows:

$$E_{m_X}(X^2) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} (v_{m_X}(\mathsf{a}))^2 Q_{m_X}(\mathsf{a})$$

Proof:

$$\begin{aligned} V_{m_X}(X) &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} (v_{m_X}(\mathsf{a}) - E_{m_X}(X))^2 Q_{m_X}(\mathsf{a}) \\ &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} \left(v_{m_X}(\mathsf{a})^2 + (E_{m_X}(X))^2 - 2\, v_{m_X}(\mathsf{a})\, E_{m_X}(X)\right) Q_{m_X}(\mathsf{a}) \\ &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} (v_{m_X}(\mathsf{a}))^2\, Q_{m_X}(\mathsf{a}) - (E_{m_X}(X))^2 \\ &= E_{m_X}(X^2) - (E_{m_X}(X))^2 \end{aligned}$$

■

(3) (*Variance of a constant*) Suppose X is a constant, i.e. there exists $a \in \mathbb{R}$ such that $m_X(\{a\}) = 1$. Then, $V_{m_X}(X) = 0$.

Proof: As m_X is Bayesian, $V_{m_X}(X) = V_{P_X}(X) = 0$. ■

(4) (*Variance of a linear function of* X) If $Y = aX + b$, where a and b are real constants, then $V_{m_Y}(Y) = a^2 V_{m_X}(X)$.

Proof: As Y is a linear function of X, $v_{m_Y}(\mathsf{a}_Y) = a\, v_{m_X}(\mathsf{a}) + b$, $E_{m_Y}(Y) = aE_{m_X}(X) + b$, and $Q_{m_Y}(\mathsf{a}_Y) = Q_{m_X}(\mathsf{a})$, where $\mathsf{a}_Y = \{a\,x + b : x \in \mathsf{a}\}$.

$$\begin{aligned} V_{m_Y}(Y) &= \sum_{\mathsf{a}_Y \in 2^{\Omega_Y}} (-1)^{|\mathsf{a}_Y|+1} (v_{m_Y}(\mathsf{a}_Y) - E_{m_Y}(Y))^2 Q_{m_Y}(\mathsf{a}_Y) \\ &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} ((a\, v_{m_X}(\mathsf{a}) + b) - (a\, E_{m_X}(X) + b))^2 Q_{m_X}(\mathsf{a}) \end{aligned}$$

Table 14. Variance of $Y = 2X + 1$.

$a \in 2^{\Omega_X}$	$m_X(a)$	$Q_{m_X}(a)$	$v(a)$	$E(X)$	$(v(a))^2$	$E(X^2)$	$V(X)$
$\{-1\}$	0.02	0.63	−1	0.059	1.00	1.188	1.184
$\{0\}$	0.05	0.70					
$\{1\}$	0.09	0.81	1		1.00		
$\{-1, 0\}$	0.12	0.42	−0.47		0.22		
$\{-1, 1\}$	0.19	0.49	0.13		0.02		
$\{0, 1\}$	0.23	0.53	0.54		0.29		
$\{-1, 0, 1\}$	0.30	0.30	0.08		1.17		
$a_Y \in 2^{\Omega_Y}$	$m_Y(a_Y)$	$Q_{m_Y}(a_Y)$	$v(a_Y)$	$E(Y)$	$(v(a_Y))^2$	$E(Y^2)$	$V(Y)$
$\{-1\}$	0.02	0.63	−1.00	1.117	1.00	5.985	4.737
$\{1\}$	0.05	0.70	1.00		1.00		
$\{3\}$	0.09	0.81	3.00		9.00		
$\{-1, 1\}$	0.12	0.42	0.05		0.003		
$\{-1, 3\}$	0.19	0.49	1.25		1.56		
$\{1, 3\}$	0.23	0.53	2.07		4.30		
$\{-1, 1, 3\}$	0.30	0.30	1.17		1.36		

$$= a^2 \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} ((v_{m_X}(\mathsf{a}) - E_{m_X}(X))^2 \, Q_{m_X}(\mathsf{a})$$

$$= a^2 \, V_{m_X}(X) \qquad \blacksquare$$

(5) (*Non-negativity*) $V_{m_X}(X) \geq 0$.

Proof: $V_{m_X}(X) = E_{m_X}(g_X)$, where $g_X = (X - E_{m_X})^2$ and as the smallest value of g_X is 0, it follows from the *bounds on expected value* property of expected values that $V_{m_X}(X) \geq 0$. $\blacksquare$

Example 4.2 (Variance of linear function of *X*): Suppose $\Omega_X = \{-1, 0, 1\}$, m_X is as in Table 14, and suppose $Y = 2X + 1$. For this example, $E_{m_X}(X) \approx 0.059$, $E_{m_X}(X^2) \approx 1.188$, and $V_{m_X}(X) = E_{m_X}(X^2) - (E_{m_X}(X))^2 \approx 1.184$. Details are shown in the upper half of Table 14 (empty cells in columns 4 and 6 have 0 values). Also $E_{m_Y}(Y) \approx 1.117$, $E_{m_Y}(Y^2) \approx 5.985$, $V_{m_Y}(Y) = E_{m_Y}(Y^2) - (E_{m_Y}(Y))^2 \approx 4.737$. Details are shown in the lower half of Table 14. Notice that $V_{m_Y}(Y) = 4\, V_{m_X}(X)$.

5. Covariance and correlation for the D–S theory

Using the definition of expectation of a real-valued function in Definition 3.2, we can define covariance and correlation of a belief function in a manner similar to probability theory.

5.1. Definitions of covariance and correlation

Definition 5.1 (Covariance): Suppose $m_{X,Y}$ is a joint BPA for real-valued variables X and Y with marginals m_X and m_Y for X and Y, respectively. The covariance of X and Y with respect to $m_{X,Y}$, denoted by $C_{m_{X,Y}}(X, Y)$, is defined as follows:

$$C_{m_{X,Y}}(X, Y) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} \left(v_{m_X}(\mathsf{a}^{\downarrow X}) - E_{m_X}(X)\right) \left(v_{m_Y}(\mathsf{a}^{\downarrow Y}) - E_{m_Y}(Y)\right) Q_{m_{X,Y}}(\mathsf{a})$$

If we consider $g_{X,Y}(v_{m_X}(\mathsf{a}^{\downarrow X}), v_{m_Y}(\mathsf{a}^{\downarrow Y})) = (v_{m_X}(\mathsf{a}^{\downarrow X}) - E_{m_X}(X))(v_{m_Y}(\mathsf{a}^{\downarrow Y}) - E_{m_Y}(Y))$, then $C_{m_{X,Y}}(X,Y) = E_{m_{X,Y}}(g_{X,Y})$. $C_{m_{X,Y}}(X,Y)$ can be regarded as a measure of linear dependence between X and Y in joint BPA $m_{X,Y}$.

We can define correlation for the D–S theory similar to the definition in probability.

Definition 5.2 (Correlation): Suppose $m_{X,Y}$ is a joint BPA for real-valued variables X and Y with marginals m_X and m_Y for X and Y, respectively. The *correlation* of X and Y with respect to joint BPA $m_{X,Y}$, denoted by $\rho_{m_{X,Y}}(X,Y)$, is defined as follows:

$$\rho_{m_{X,Y}}(X,Y) = \frac{C_{m_{X,Y}}(X,Y)}{\sqrt{V_{m_X}(X)}\sqrt{V_{m_Y}(Y)}}$$

5.2. Properties of covariance

Consider the situation in Definition 5.1.

(1) (*Equivalence to the probabilistic case*) If BPA $m_{X,Y}$ for (X,Y) is Bayesian, then $C_{m_{X,Y}}(X,Y) = C_{P_{X,Y}}(X,Y)$, where $P_{X,Y}$ corresponds to $m_{X,Y}$, i.e. $P_{X,Y}(x,y) = m_{X,Y}(\{(x,y)\})$ for all $(x,y) \in \Omega_{X,Y}$.

Proof: As $m_{X,Y}$ is Bayesian, for singleton subsets $\{(x,y)\} \in 2^{\Omega_{X,Y}}$, $Q_{m_{X,Y}}(\{(x,y)\}) = m_{X,Y}(\{(x,y)\}) = P_{X,Y}(x,y)$, $v_{m_X}(\{(x,y)^{\downarrow X}\}) = x$, and $v_{m_Y}(\{(x,y)^{\downarrow Y}\}) = y$. Also, from the *consistency with probabilistic expectation* property, $E_{m_X}(X) = E_{P_X}(X)$, and $E_{m_Y}(Y) = E_{P_Y}(Y)$. Finally, for non-singleton subsets $\mathsf{a} \in 2^{\Omega_X}$, $Q_{m_X}(\mathsf{a}) = 0$. Thus, the result follows from the definition of $C_{m_{X,Y}}(X,Y)$. ■

(2) ($C_{m_X}(X,X)$ *is* $V_{m_X}(X)$) If $m_{X,Y}$ is such that $m_{Y|x}(\{x\}) = 1$, i.e. $Y = X$, then $C_{m_{X,Y}}(X,Y) = V_{m_X}(X)$.

Proof: If $Y = X$, then the definition of $C_{m_{X,Y}}(X,Y) = C_{m_X}(X,X)$ reduces to the definition of $V_{m_X}(X)$. ■

(3) (*Computation of covariance*) As in the probabilistic case, we can compute the covariance as follows:

$$C_{m_{X,Y}}(X,Y) = E_{m_{X,Y}}(X \cdot Y) - \left(E_{m_X}(X)\right)\left(E_{m_Y}(Y)\right),$$

where $E_{m_{X,Y}}(X \cdot Y) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} v_{m_X}(\mathsf{a}^{\downarrow X})\, v_{m_Y}(\mathsf{a}^{\downarrow Y})\, Q_{m_{X,Y}}(\mathsf{a})$.

Proof:

$$
\begin{aligned}
C_{m_{X,Y}}(X,Y) &= \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} (v_{m_X}(\mathsf{a}^{\downarrow X}) - E_{m_X}(X))(v_{m_Y}(\mathsf{a}^{\downarrow Y}) \\
&\quad - E_{m_Y}(Y))\, Q_{m_{X,Y}}(\mathsf{a}) \\
&= \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, v_{m_X}(\mathsf{a}^{\downarrow X})\, v_{m_Y}(\mathsf{a}^{\downarrow Y})\, Q_{m_{X,Y}}(\mathsf{a}) \\
&\quad - E_{m_Y}(Y) \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, v_{m_X}(\mathsf{a}^{\downarrow X})\, Q_{m_{X,Y}}(\mathsf{a}) \\
&\quad - E_{m_X}(X) \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, v_{m_Y}(\mathsf{a}^{\downarrow Y})\, Q_{m_{X,Y}}(\mathsf{a}) + E_{m_X}(X)\, E_{m_Y}(Y)
\end{aligned} \tag{9}
$$

Next, we will show that $\sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, v_{m_X}(\mathsf{a}^{\downarrow X})\, Q_{m_{X,Y}}(\mathsf{a}) = E_{m_X}(X)$. This is because

$$
\begin{aligned}
\sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, v_{m_X}(\mathsf{a}^{\downarrow X})\, Q_{m_{X,Y}}(\mathsf{a}) &= \sum_{\mathsf{a}_X\in 2^{\Omega_X}} v_{m_X}(\mathsf{a}_X) \\
&\quad \times \sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}:\mathsf{a}^{\downarrow X}=\mathsf{a}_X} (-1)^{|\mathsf{a}|+1} Q_{m_{X,Y}}(\mathsf{a}) \\
&= \sum_{\mathsf{a}_X\in 2^{\Omega_X}} v_{m_X}(\mathsf{a}_X)\, (-1)^{|\mathsf{a}_X|+1}\, Q_{m_X}(\mathsf{a}_X) \\
&= E_{m_X}(X)
\end{aligned}
$$

Similarly we can show that $\sum_{\mathsf{a}\in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, v_{m_Y}(\mathsf{a}^{\downarrow Y})\, Q_{m_{X,Y}}(\mathsf{a}) = E_{m_Y}(Y)$. The result then follows from Equation (9). ■

(4) (*Independence*), If X and Y are independent[1], i.e. $m_{X,Y} = m_X \oplus m_Y$, then $C_{m_{X,Y}}(X,Y) = 0$.

Proof: In the *expected value of a product of independent variables* property, it follows that $E_{m_{X,Y}}(X \cdot Y) = E_{m_X}(X) \cdot E_{m_Y}(Y)$. From the *computation of covariance* property, it follows that $C_{m_{X,Y}}(X,Y) = 0$. ■

(5) (*Bilinear*) If m is a BPA for $\{X, Y, Z\}$, and a, b, and c are real constants, then $C_m(aX + bY + c, Z) = a\, C_m(X,Z) + b\, C_m(Y,Z)$.

Proof: It follows from the *computation of covariance* property that

$$
\begin{aligned}
C_m(aX + bY + c, Z) &= E_m((aX + bY + c)Z) - E_m(aX + bY + c)\, E_m(Z) \\
&= E_m(a\,X\,Z) + E_m(b\,Y\,Z) + E_m(c\,Z) - (a\,E_m(X)\,E_m(Z) \\
&\quad + b\,E_m(Y)\,E_m(Z) + c\,E_m(Z)
\end{aligned}
$$

$$= a(E_m(X\,Z) - E_m(X)\,E_m(Z)) + b(E_m(Y\,Z) - E_m(Y)\,E_m(Z))$$
$$= a\,C_m(X, Z) + b\,C_m(Y, Z) \qquad \blacksquare$$

(6) (*Bounds on covariance*) Suppose m is a BPA for $\{X, Y\}$. Then,

$$-\sqrt{V_m(X)}\,\sqrt{V_m(Y)} \leq C_m(X, Y) \leq \sqrt{V_m(X)}\,\sqrt{V_m(Y)}$$

Proof: This property follows from the corresponding property of correlation below. $\blacksquare$

5.3. Properties of correlation

Consider the definition of correlation in Definition 5.2.

- (*Equivalence to the probabilistic case*) If BPA $m_{X,Y}$ for (X, Y) is Bayesian, then $\rho_{m_{X,Y}}(X, Y) = \rho_{P_{X,Y}}(X, Y)$, where $P_{X,Y}$ corresponds to $m_{X,Y}$, i.e. $P_{X,Y}(x, y) = m_{X,Y}(\{(x, y)\})$ for all $(x, y) \in \Omega_{X,Y}$.

Proof: From the *equivalence to the probabilistic case* property of covariance, $C_{m_{X,Y}}(X, Y) = C_{P_{X,Y}}(X, Y)$, and from the *equivalence to the probabilistic case* property of variance, $V_{m_X}(X) = V_{P_X}(X)$ and $V_{m_Y}(Y) = V_{P_Y}(Y)$. Therefore, the result follows. $\blacksquare$

- (*Bounds on correlation*) Suppose m is a BPA for $\{X, Y\}$. Then,

$$-1 \leq \rho_m(X, Y) \leq 1$$

Proof: Let $X^* = (X - E_m(X))/\sqrt{V_m(X)}$, and let $Y^* = (Y - E_m(Y))/\sqrt{V_m(Y)}$. Then, it follows from properties of expected value and variance that $E_m(X^*) = E_m(Y^*) = 0$, and $V_m(X^*) = V_m(Y^*) = 1$. Also, it follows from the *bilinear* property of covariance that $C_m(X^*, Y^*) = \rho_m(X, Y)$. Next,

$$\begin{aligned} 0 \leq V_m(X^* \pm Y^*) &= C_m(X^* \pm Y^*, X^* \pm Y^*) \\ &= C_m(X^*, X^*) \pm 2\,C_m(X^*, Y^*) + C_m(Y^*, Y^*) \\ &= 1 \pm 2\,\rho_m(X, Y) + 1 \\ &= 2[1 \pm \rho_m(X, Y)] \end{aligned}$$

Thus, $1 \pm \rho_m(X, Y) \geq 0$ implies that $-1 \leq \rho_m(X, Y) \leq 1$. $\blacksquare$

Example 5.1 (Covariance and correlation): Suppose $\Omega_X = \{1, 3\}$, and suppose $\Omega_Y = \{2, 4\}$. Suppose m_X, $m_{Y|x=1}$, and $m_{Y|x=3}$ are as follows:

$$\begin{aligned} &m_X(\{1\}) = 0.56, && m_X(\{3\}) = 0.33, && m_X(\{1, 3\}) = 0.11, \\ &m_{Y|x=1}(\{2\}) = 0.42, && m_{Y|x=1}(\{4\}) = 0.23, && m_{Y|x=1}(\{2, 4\}) = 0.35, \\ &m_{Y|x=3}(\{2\}) = 0.29, && m_{Y|x=3}(\{4\}) = 0.39, && m_{Y|x=3}(\{2, 4\}) = 0.32. \end{aligned}$$

If we conditionally embed BPAs $m_{Y|x=1}$, and $m_{Y|x=3}$ to BPAs $m_{x=1,Y}$ and $m_{x=3,Y}$ for (X, Y), and combine these two BPAs with m_X using Dempster's rule, we obtain joint BPA $m_{X,Y}$ for

Table 15. Covariance and Correlation of (X, Y).

$\mathsf{a} \in 2^{\Omega_X}$	$m_X(\mathsf{a})$	$Q_{m_X}(\mathsf{a})$	$v_{m_X}(\mathsf{a})$	$E_{m_X}(X)$	$V_{m_X}(X)$
{1}	0.56	0.67	1.00	1.793	1.062
{3}	0.33	0.44	3.00		
{1, 3}	0.11	0.11	1.79		
$\mathsf{a} \in 2^{\Omega_{X,Y}}$	$m_{X,Y}(\mathsf{a})$	$Q_{m_{X,Y}}(\mathsf{a})$	$v_{m_X}(\mathsf{a}^{\downarrow X})$	$v_{m_Y}(\mathsf{a}^{\downarrow Y})$	$C_{m_{X,Y}}(X, Y)$
{(1, 2)}	0.24	0.52	1.00	2.00	0.153
{(1, 4)}	0.13	0.39	1.00	4.00	
{(3, 2)}	0.10	0.27	3.00	2.00	
{(3, 4)}	0.13	0.31	3.00	4.00	
{(1, 2), (1, 4)}	0.20	0.23	1.00	2.94	
{(1, 2), (3, 2)}	0.01	0.05	1.79	2.00	
{(1, 2), (3, 4)}	0.02	0.06	1.79	2.94	
{(1, 4), (3, 2)}	0.01	0.04	1.79	2.94	
{(1, 4), (3, 4)}	0.01	0.05	1.79	4.00	
{(3, 2), (3, 4)}	0.11	0.14	3.00	2.94	
{(1, 2), (1, 4), (3, 2)}	0.01	0.02	1.79	2.94	
{(1, 2), (1, 4), (3, 4)}	0.02	0.03	1.79	2.94	
{(1, 2), (3, 2), (3, 4)}	0.01	0.03	1.79	2.94	
{(1, 2), (3, 2), (3, 4)}	0.01	0.02	1.79	2.94	
$\Omega_{X,Y}$	0.01	0.01	1.79	2.94	
$\mathsf{a} \in 2^{\Omega_Y}$	$m_Y(\mathsf{a})$	$Q_{m_Y}(\mathsf{a})$	$v_{m_Y}(\mathsf{a})$	$E_{m_Y}(Y)$	$V_{m_Y}(Y)$
{2}	0.34	0.73	2.00	3.085	1.384
{4}	0.27	0.66	4.00		
{2, 4}	0.39	0.39	2.94		

(X, Y) as shown in Table 15. The marginal $m_{X,Y}^{\downarrow Y}$ is denoted by m_Y, which is also shown in Table 15. $C_{m_{X,Y}}(X, Y)$ is computed as $E_{m_X}(X\,Y) - E_{m_X}(X)\,E_{m_Y}(Y) = 0.153$. Correlation $\rho_{m_{X,Y}}(X, Y) = (C_{m_{X,Y}}(X, Y))/(\sqrt{V_{m_X}(X)}\,\sqrt{V_{m_Y}(Y)}) = 0.126$.

6. Higher moments

In this section, we define higher moments of D–S belief functions about the mean and about the origin.

6.1. Definitions of higher moments

Definition 6.1: Suppose m_X is a BPA for X. The rth moment of X about the mean $\mu = E_{m_X}(X)$, denoted by $E_{m_X}[(X-\mu)^r]$, is defined as follows:

$$E_{m_X}[(X-\mu)^r] = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\,(v_{m_X}(\mathsf{a}) - \mu)^r\,Q_{m_X}(\mathsf{a}),$$

for all $\mathsf{a} \in 2^{\Omega_X}$.

Similarly, we can define higher moments about the origin as follows.

Definition 6.2: Suppose m_X is a BPA for X. The rth moment of X about the origin, denoted by $E_{m_X}(X^r)$, is defined as follows:

$$E_{m_X}(X^r) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\, v_{m_X}(\mathsf{a})^r\, Q_{m_X}(\mathsf{a}),$$

for all $\mathsf{a} \in 2^{\Omega_X}$.

6.2. *Property of higher moments*

(1) (*Computation of higher moments about the origin*) The rth moment about the mean can be computed using rth moment about the origin as follows:

$$E_{m_X}[(X-\mu)^r] = \sum_{j=0}^{r} \binom{r}{j} E_{m_X}(X^j)\,(-\mu)^{r-j}$$

Proof:

$$\begin{aligned} E_{m_X}[(X-\mu)^r] &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\, (v_{m_X}(\mathsf{a}) - \mu)^r\, Q_{m_X}(\mathsf{a}) \\ &= \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} \sum_{j=0}^{r} \binom{r}{j}\, v_{m_X}(\mathsf{a})^j\, (-\mu)^{r-j}\, Q_{m_X}(\mathsf{a}) \\ &= \sum_{j=0}^{r} \binom{r}{j}\, (-\mu)^{r-j} \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\, v_{m_X}(\mathsf{a})^j\, Q_{m_X}(\mathsf{a}) \\ &= \sum_{j=0}^{r} \binom{r}{j}\, (-\mu)^{r-j}\, E_{m_X}(X^j). \end{aligned}$$

■

7. Summary and conclusions

We propose a new definition of expected value for variables whose uncertainty is described by D–S belief functions defined on real-valued frames of discernment. Also, if we have a symbolic frame of discernment, but a real-valued function defined on the set of all non-empty subsets of the frame, then we propose a new definition of expectation of the function in a similar manner.

Our new definition satisfies many of the properties satisfied by the probabilistic expectation operator, which was first proposed by Huygens (1657) in the context of the problem of points posed by Chevalier de Méré to Blaise Pascal. For Bayesian BPAs, our definition of expected value coincides with the expected value of the corresponding PMF.

Analogous to the probabilistic case, we define variance of a BPA m_X for X as $E_{m_X}((X - E_{m_X}(X))^2)$. We show that our definition of variance of a BPA has many of the properties of the variance in the probabilistic case. For Bayesian BPA, our definition of variance coincides with the variance of the corresponding PMF.

Analogous to the probabilistic case, we define covariance of X and Y with respect to a joint BPA $m_{X,Y}$ for (X, Y) as $E_{m_{X,Y}}((X - E_{m_X}(X))(Y - E_{m_Y}(Y)))$. We show that our definition of covariance of a BPA has many of the properties of the covariance in the probabilistic case. For Bayesian BPA, our definition of covariance coincides with the covariance of the corresponding PMF. Also, we define correlation of X and Y with respect to a joint BPA $m_{X,Y}$ for (X, Y) as $C_{m_{X,Y}}(X, Y)/(\sqrt{V_{m_X}(X)}\sqrt{V_{m_Y}(Y)})$. Analogous to the probabilistic case, we define higher moments of X (with respect to m_X) about the mean and about the origin.

If we define $I(\mathsf{a}) = \log_2(1/(Q_{m_X}(\mathsf{a})))$ as the information content of observing subset $\mathsf{a} \in 2^{\Omega_X}$ whose uncertainty is described by m_X, then similar to Shannon's definition of entropy of PMFs (Shannon 1948), we can define entropy of BPA m_X for X as an expected value of the function $I(X)$, i.e. $H(m_X) = E_{m_X}(I(X))$. This is what is proposed in Jiroušek and Shenoy (2018b, 2018c). This definition of entropy has many nice properties. In particular, it satisfies the compound distributions property: $H(m_X \oplus m_{Y|X}) = H(m_X) + H(m_{Y|X})$, where $m_{Y|X}$ is a BPA for (X, Y) obtained by $\oplus\{m_{x,Y} : x \in \Omega_X\}$, and $m_{x,Y}$ is a BPA for (X, Y) obtained by conditional embedding of conditional BPA $m_{Y|x}$ for Y given $X = x$. The compound distribution property of Shannon's entropy of a PMF is one of the most important properties that characterizes Shannon's definition (Shannon 1948). The definition in Jiroušek and Shenoy (2018b, 2018c) is the only one that satisfies the compound distributions property.

One motivation in defining an expectation operator was to define an axiomatic linear utility theory for the D–S belief function lotteries along the lines of von Neumann–Morgenstern's (vN–M's) axiomatic theory for probabilistic lotteries (von Neumann and Morgenstern 1947). There are utility theories for D–S lotteries proposed by Jaffray (1989) and Smets (2002). Jaffray's linear utility theory does not use Dempster's combination rule. Instead, it is based on a mixture set of BPAs, which is not Dempster's combination rule, although Dempster's combination can be used to describe a mixture operation. Smets' decision theory is based on first transforming a belief function to a PMF and then using vN–M's theory. Unfortunately, Smets's theory is unable to explain ambiguity aversion as demonstrated by Ellsberg's paradox (Ellsberg 1961). Some preliminary results on this topic can be found in Denoeux and Shenoy (2019).

Note

1. The definition of independence here is based on factorization semantics, see, e.g. Shenoy (1994).

Acknowledgments

This work is inspired by recent work on entropy of belief functions in the Dempster-Shafer theory of evidence in collaboration with Radim Jiroušek from The Czech Academy of Sciences (Jiroušek and Shenoy 2018b). Thank you, Radim! Thanks also to Suzanna Emelio for a careful proofreading of this paper. A short version of this paper appeared as Shenoy (2018). Comments and suggestions for improvement of the paper are welcome.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by Ronald G. Harper Professorship at the University of Kansas (BUSHARPRO).

Notes on contributor

Prakash P. Shenoy is the Ronald G. Harper Distinguished Professor of Artificial Intelligence in Business, University of Kansas at Lawrence. He received a B.Tech. in Mechanical Engineering from the Indian Institute of Technology, Bombay, India, in 1973, and an M.S. and a Ph.D. in Operations Research/Industrial Engineering from Cornell University in 1975 and 1977, respectively. His research interests are in the areas of artificial intelligence and decision sciences. He is the inventor of valuation-based systems, an abstract framework for knowledge representation and inference that includes Bayesian probabilities, Dempster–Shafer belief functions, Spohn's kappa calculus, Zadeh's possibility theory, propositional logic, optimization, solving systems of equations, database retrieval, and other domains. He is also the co-author (with G. Shafer) of the so-called Shenoy–Shafer architecture for finding marginals of joint distributions using local computation. He has published many articles on management of uncertainty in expert systems, decision analysis, and the mathematical theory of games. His articles have appeared in journals such as Operations Research, Management Science, International Journal of Game Theory, Artificial Intelligence, and International Journal of Approximate Reasoning. http://pshenoy.faculty.ku.edu.

ORCID

Prakash P. Shenoy http://orcid.org/0000-0002-8425-896X

References

Choquet, G. 1953. "Theory of Capacities." *Annales de l'Institut Fourier* 5: 131–295.
Cobb, B. R., and P. P. Shenoy. 2006. "On the Plausibility Transformation Method for Translating Belief Function Models to Probability Models." *International Journal of Approximate Reasoning* 41 (3): 314–330.
Coletti, G., D. Petturiti, and B. Vantaggi. 2019. "Models for Pessimistic Or Optimistic Decisions Under Different Uncertain Scenarios." *International Journal of Approximate Reasoning* 105 (2): 305–326.
Dempster, A. P. 1967. "Upper and Lower Probabilities Induced by a Multivalued Mapping." *The Annals of Mathematical Statistics* 38 (2): 325–339.
Denoeux, T., and P. P. Shenoy. 2019. "An Interval-Valued Utility Theory for Decision Making with Dempster-Shafer Belief Functions." Working Paper 336, University of Kansas School of Business, Lawrence, KS.
Ellsberg, D. 1961. "Risk, Ambiguity and the Savage Axioms." *The Quarterly Journal of Economics* 75 (2): 643–669.
Fagin, R., and J. Y. Halpern. 1991. "A New Approach to Updating Beliefs." In *Uncertainty in Artificial Intelligence* 6, edited by P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, 347–374. Amsterdam: Elsevier Science.
Gilboa, I., and D. Schmeidler. 1989. "Maxmin Expected Utility with a Non-unique Prior." *Journal of Mathematical Economics* 18 (2): 141–153.
Gilboa, I., and D. Schmeidler. 1994. "Additive Representations of Non-additive Measures and the Choquet Integral." *Annals of Operations Research* 52: 43–65.
Halpern, J. Y., and R. Fagin. 1992. "Two Views of Belief: Belief As Generalized Probability and Belief As Evidence." *Artificial Intelligence* 54 (3): 275–317.
Huygens, C. 1657. *De ratiociniis in ludo aleæ* [English translation, published in 1714].

Jaffray, J.-Y. 1989. "Linear Utility Theory for Belief Functons." *Operations Research Letters* 8 (2): 107–112.
Jiroušek, R., and V. Kratochvíl. 2018. "Comparison of Shenoy's Expectation Operator with Probabilistic Transforms and Perez' Barycenter." In *Proceedings of the 11th Workshop on Uncertainty Processing*, edited by V. Kratochvíl and J. Vejnarová, 87–95. Prague.
Jiroušek, R., and P. P. Shenoy. 2018a. "A New Definition of Entropy of Belief Functions in the Dempster-Shafer Theory." *International Journal of Approximate Reasoning* 92 (1): 49–65.
Jiroušek, R., and P. P. Shenoy. 2018b. "A Decomposable Entropy of Belief Functions in the Dempster-Shafer Theory." Working Paper 334, University of Kansas School of Business, Lawrence, KS.
Jiroušek, R., and P. P. Shenoy. 2018c. "A Decomposable Entropy of Belief Functions in the Dempster-Shafer Theory." In *Belief Functions: Theory and Applications, 5th International Conference, BELIEF 2018*, vol. 11069 of Lecture Notes in Artificial Intelligence, edited by S. Destercke, T. Denoeux, F. Cuzzolin, and A. Martin, 146–152. Switzerland AG: Springer Nature.
Rockafellar, R. T. 1970. *Convex Analysis*. Princeton, NJ: Princeton University Press.
Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.
Shafer, G. 1981. "Constructive Probability." *Synthese* 48 (1): 1–60.
Shafer, G. 1982. "Belief Functions and Parametric Models." *Journal of the Royal Statistical Society, Series B* 44 (3): 322–352.
Shafer, G. 1990. "Perspectives on the Theory and Practice of Belief Functions." *International Journal of Approximate Reasoning* 4 (5–6): 323–362.
Shafer, G. 1992. "Rejoinders to Comments on "Perspectives on the Theory and Practice of Belief Functions"." *International Journal of Approximate Reasoning* 6 (3): 445–480.
Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27: 379–423, 623–656.
Shenoy, P. P. 1994. "Conditional Independence in Valuation-based Systems." *International Journal of Approximate Reasoning* 10 (3): 203–234.
Shenoy, P. P. 2018. "An Expectation Operator for Belief Functions in the Dempster-Shafer theory." In *Proceedings of the 11th Workshop on Uncertainty Processing (WUPES'18), Třeboň, Czech Republic*, edited by V. Kratochvíl and J. Vejnarová, 165–176. Praha, Czech Republic: MatfyzPress.
Smets, P. 1978. "Un Modele Mathematico-Statistique Simulant Le Processus Du Diagnostic Medical." PhD thesis, Free University of Brussels.
Smets, P. 2002. "Decision Making in a Context Where Uncertainty is Represented by Belief Functions." In *Belief Functions in Business Decisions*, edited by R. P. Srivastava and T. J. Mock, 316–332. Heidelberg: Physica-Verlag.
von Neumann, J., and O. Morgenstern. 1947. *Theory of Games and Economic Behavior*. 2nd ed. Princeton, NJ: Princeton University Press.