An Expectation Operator for Belief Functions in the Dempster-Shafer Theory

Prakash P. Shenoy
School of Business
University of Kansas
pshenoy@ku.edu

Abstract

The main contribution of this paper is a new definition of an expectation operator for 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 values 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. Our expectation operator works directly with D-S belief functions. 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.

1 Introduction

The main goal of this paper is to propose an expectation operator for belief functions in the D-S theory of evidence [2, 6].

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. Our expectation operator works directly with D-S belief functions. In general, our definition provides different expected values than, e.g., if we use the pignistic transform or the plausibility transform.

An outline of the remainder of the paper is as follows. In Section 2, we review the definition of expected value of a discrete real-valued random variable characterized by a PMF. Also, we review some of the main properties of the definition. In Section 3, we review the representations and operations of the D-S theory of belief functions. In Section 4, 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 from the set of all non-empty subsets of the states of X, we also provide a definition of the expected value of g. 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. Finally, in Section 5, we summarize and conclude.

2 Expected Values of Discrete Probability Distributions

In this section we briefly review the expectation operator for discrete random variables with finite state space whose behavior is described by probability mass functions.

2.1 Definition of probabilistic expectation

Suppose X is a discrete real-valued random variable with a finite state space Ω_X, and suppose $P_X : \Omega_X \to [0, 1]$ is a probability mass function (PMF) for X, i.e., $P_X(x) \geq 0$ for all $x \in \Omega_X$, and $\sum_{x \in \Omega_X} P_X(x) = 1$. Then the expected value of X with respect to P_X, denoted by $E_{P_X}(X)$, is defined as follows:

$$E_{P_X}(X) = \sum_{x \in \Omega_X} x \cdot P_X(x) \tag{1}$$

Notice that as X is real-valued, the definition in Eq. (1) is well defined. Also, as Ω_X is finite, $E_{P_X}(X)$ always exists.

2.2 Properties of probabilistic expectation

Consider the situation in the definition of probabilistic expectation. The expectation operator has the following properties.

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

2. (*Expected value of a function of X*) Suppose $Y = g_X : \Omega_X \to \mathbb{R}$ is a well-defined function of X, where $\mathbb{R}$ is the set of all real numbers. Then, $E_{P_Y}(Y)$ is as follows:

$$E_{P_Y}(Y) = \sum_{x \in \Omega_X} g_X(x) \cdot P_X(x) \tag{2}$$

For convenience, the right-hand-side of Eq. (2) is denoted by $E_{P_X}(g_X)$. This property is referred to as the *law of the unconscious statistician.* If $Y = g_X$, then Y is a random variable whose PMF P_Y is defined in terms of PMF P_X as follows:

$$P_Y(y) = \sum_{x \in \Omega_X : g_X(x) = y} P_X(x) \tag{3}$$

It follows from the definition of expected value that $E_{P_Y}(Y) = E_{P_X}(g_X)$. The result in Eq. (2) says that $E_{P_Y}(Y)$ can be computed directly from the PMF of X without computing the PMF of Y.

3. (*Expected value of a linear function of X*) Suppose $Y = g_X = aX + b$ where a and b are real constants. Then $E_{P_Y}(Y) = aE_{P_X}(X) + b$.

4. (*Expected value of a function of X and Y*) The law of the unconscious statistician generalizes to the multidimensional case. Suppose X and Y are discrete random variables with state spaces Ω_X and Ω_Y, respectively, with joint PMF $P_{X,Y}$, i.e., $P_{X,Y}(x, y) \geq 0$ for all $(x, y) \in \Omega_X \times \Omega_Y$, and $\sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} P_{X,Y}(x, y) = 1$. Then if $Z = g_{X,Y} : \Omega_X \times \Omega_Y \to \mathbb{R}$ is a well-defined function of (X, Y), then

$$E_{P_Z}(Z) = E_{P_{X,Y}}(g_{X,Y}) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g_{X,Y}(x, y) P_{X,Y}(x, y) \tag{4}$$

5. (*Expected value of a linear function of X and Y*) If $Z = g_{X,Y} = aX + bY + c$, where a, b, and c are real constants, then

$$E_{P_Z}(Z) = aE_{P_{X,Y}}(X) + bE_{P_{X,Y}}(Y) + c \tag{5}$$

3 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 [5].

3.1 Representations of belief functions

Belief functions can be represented in four different ways: basic probability assignments (BPA), belief functions, plausibility functions, and commonality functions. Here, we focus only on BPA and commonality functions.

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 $\iota_X(\Omega_X) = 1$. If all focal elements of m_X are singleton subsets of Ω_X, then we say m_X 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 $x \in \Omega_X$.

The information in a BPA m_X can also be represented by a corresponding commonality function Q_{m_X} that is defined as follows: $Q_{m_X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \supseteq \mathsf{a}} m_X(\mathsf{b})$ 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 $|\mathsf{a}| > 1$. 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:

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

Next, we describe the two main operations for making inferences.

3.2 Basic operations in the D-S theory

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

In the D-S theory, we can combine two BPAs m_1 and m_2 representing distinct pieces of evidence by Dempster's rule [2] 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 $\mathsf{b}^{\downarrow X} \in 2^{\Omega_X}$.

Dempster's rule can be defined in terms of commonality functions [6] as follows: Suppose m_1 and m_2 are BPAs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. 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 $m_1 \oplus m_2$ for $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$ is as follows:

$$Q_{m_1 \oplus m_2}(\mathsf{a}) = K^{-1} Q_{m_1}(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_{m_2}(\mathsf{a}^{\downarrow \mathcal{X}_2}), \tag{6}$$

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

$$K = \sum_{\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}} (-1)^{|\mathsf{a}|+1} Q_{m_1}(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_{m_2}(\mathsf{a}^{\downarrow \mathcal{X}_2}). \tag{7}$$

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. In terms of commonality functions, Dempster's rule is pointwise multiplication of commonality functions followed by normalization.

Marginalization in D-S theory is addition of values of BPAs. Suppose m is a BPA for $\mathcal{X}$. Then, the marginal of m for $\mathcal{X}_1$, where $\mathcal{X}_1 \subseteq \mathcal{X}$, denoted by $m^{\downarrow \mathcal{X}_1}$, is a BPA for $\mathcal{X}_1$ such that for each $\mathsf{a} \in 2^{\Omega_{\mathcal{X}_1}}$,

$$m^{\downarrow \mathcal{X}_1}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{\mathcal{X}}} :\, \mathsf{b}^{\downarrow \mathcal{X}_1} = \mathsf{a}} m(\mathsf{b}). \tag{8}$$

This completes a brief description of D-S theory of belief functions. For more details, see [6].

4 A New Definition of Expected Value for the D-S Theory

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

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 expected value 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 we can consider $v_m(\{x\}) = x$.

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:

$$\min \mathsf{a} \leq v_{m_X}(\mathsf{a}) \leq \max \mathsf{a} \tag{9}$$

One way to satisfy the inequality in Eq. (9) is as follows:

$$v_{m_X}(\mathsf{a}) = \frac{\sum_{x \in \mathsf{a}} x \cdot Q_{m_X}(\{x\})}{\sum_{x \in \mathsf{a}} Q_{m_X}(\{x\})} \quad \text{for all } \mathsf{a} \in 2^{\Omega_X} \tag{10}$$

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 $Q_{m_X}(\{x\})$.

4.1 Definition of expected value for D-S belief functions

Suppose m_X is a BPA for X with 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:

$$E_{m_X}(X) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} v_{m_X}(\mathsf{a}) \cdot Q_{m_X}(\mathsf{a}) \tag{11}$$

4.2 Properties of expected values of D-S belief functions

Some important properties of our definition in Eq. (11) are as follows. Consider the situation in the definition of expected value of D-S belief functions in Eq. (11).

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 Eq. (11) reduces to $E_{P_X}(X)$ in Eq. (1).

2. (*Expectation 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 function of X*) Suppose $Y = g_X : \Omega_X \to \mathbb{R}$ is a *linear* function, then $E_{m_Y}(Y)$ can be computed as follows:

$$E_{m_Y}(Y) = E_{m_X}(g_X) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} \, g_X(v_{m_X}(\mathsf{a})) \, Q_{m_X}(\mathsf{a}) \tag{12}$$

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 1 and 2 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: As g_X is linear, it 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., $Q_{m_Y}(\mathsf{a}_\mathsf{Y}) = Q_{m_X}(\mathsf{a})$, where $\mathsf{a}_\mathsf{Y} \in 2^{\Omega_Y}$ 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 $v_{m_Y}(\mathsf{a}_\mathsf{Y}) = g(v_m(\mathsf{a}))$ for all $\mathsf{a} \in 2^{\Omega_X}$. Suppose $Y = g_X = aX + b$.

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

This completes the proof.

4. (Expected value of a linear function of X) Suppose $Y = g_X = aX + b$ where a and b are real constants, and suppose m_X is a BPA for X. Then $E_{m_Y}(Y) = aE_{m_X}(X) + b$.

   Proof: From the *expected value of a function of* X property, it follows that that $E_{m_Y}(Y) = E_{m_X}(g_X) = E_{m_X}(aX + b)$. Thus,

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

5. (*Expected value of a function of* X *and* Y) The law of the unconscious statistician generalizes to the multidimensional case. 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). If $g_{X,Y} : \Omega_X \times \Omega_Y \to \mathbb{R}$ is a *linear* function of (X, Y), then

$$
E_m(g_{X,Y}) = \sum_{\mathsf{a} \in 2^{\Omega_X \times \Omega_Y}} (-1)^{|\mathsf{a}|+1} \, g_{X,Y}(v(\mathsf{a})) \, Q(\mathsf{a}) \tag{13}
$$

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

$\mathsf{a} \in 2^{\Omega_X}$	$m_X(\mathsf{a})$	$Q_{m_X}(\mathsf{a})$	$v_{m_X}(\mathsf{a})$	$E_{m_X}(X)$	$(v_{m_X}(\mathsf{a}))^2$	$E_{m_X}(g_X)$
$\{-1\}$	0.02	0.63	−1.00	0.059	1.00	1.188
$\{0\}$	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	
$\mathsf{b} \in 2^{\Omega_Y}$	$m_Y(\mathsf{b})$	$Q_{m_Y}(\mathsf{b})$	$v_{m_Y}(\mathsf{b})$	$E_{m_Y}(Y)$		
$\{1\}$	0.30	0.95	1.00	0.576		
$\{0\}$	0.05	0.70	0.00			
$\{1, 0\}$	0.65	0.65	0.58			

As in the case of *expected value of a function of* X property, this property holds only for the case where $g_{X,Y}$ is a linear function.

A proof of this property is similar to the proof of the *expected value of a function of* X property, and is therefore omitted.

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

$$E_{m_Z}(Z) = E_{m_{X,Y}}(aX + bY + c) = aE_{m_{X,Y}}(X) + bE_{m_{X,Y}}(Y) + c \tag{14}$$

A proof of this property is similar to the proof of the *expected value of a linear function of* X property, and is therefore omitted.

Example 1 (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 1. Suppose* $Y = g_X = X^2$*. Notice that* g_X *is not 1-1. Then,* $\Omega_Y = \{1, 0\}$*, and* m_Y *is as shown in Table 1. For this example,* $E_{m_Y}(Y) = 0.576$*, and* $E_{m_X}(g_X) = 1.188$*. Thus, Eq.* (12) *does not hold.*

Example 2 (Nonlinear 1-1 function) *Consider a real-valued variable* Z *with* $\Omega_X = \{1, 2, 3\}$*, and suppose* m_Z *is a BPA for* Z *as shown in Table 2. Suppose* $Y = g_Z = \log(Z)$*. Then,* $\Omega_Y = \{\log(1), \log(2), \log(3)\} \approx \{0, 0.30, 0.48\}$*, and* m_Y *is as shown in Table 2. As the function is 1-1, the values of* m_Y *are the same as the values of* m_Z*. For this example,* $E_{m_Y}(Y) = 0.273$*, and* $E_{m_Z}(\log(Z)) = 0.241$*. Thus, Eq.* (12) *does not hold.*

Table 2: Expected value of $Y = g_Z = \log(Z)$, a nonlinear 1-1 function

$\mathsf{a} \in 2^{\Omega_Z}$	$m_Z(\mathsf{a})$	$Q_{m_Z}(\mathsf{a})$	$v_{m_Z}(\mathsf{a})$	$E_{m_Z}(Z)$	$\log(v_{m_Z}(\mathsf{a}))$	$E_{m_Z}(g_Z)$
$\{1\}$	0.02	0.63	1.00	2.059	0.00	0.241
$\{2\}$	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	
$\mathsf{a}_Y \in 2^{\Omega_Y}$	$m_Y(\mathsf{a}_Y)$	$Q_{m_Y}(\mathsf{a}_Y)$	$v_{m_Y}(\mathsf{a}_Y)$	$E_{m_Y}(Y)$		
$\{0\}$	0.02	0.63	0.00	0.273		
$\{0.30\}$	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			

Example 3 (Linear 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 3. Suppose $Y = g_X = 2X + 1$. Then, $\Omega_Y = \{-1, 1, 3\}$, and m_Y is as shown in Table 3. 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, Eq. (12) holds. Also notice that $E_{m_X}(g_X) = E_{m_X}(2X + 1) = 2E_{m_X}(X) + 1 = 2(0.059) + 1 = 1.117$.*

4.3 A definition of expected value of a real-valued function of X

Suppose Q_{m_X} is a commonality function for X corresponding to BPA m_X for X, and Ω_X may not be real-valued, but $g_X : 2^{\Omega_X} \to \mathbb{R}$ 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:

$$E_{m_X}(g_X) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\, g_X(\mathsf{a})\, Q_{m_X}(\mathsf{a}) \tag{15}$$

The definition of the expected value of g_X with respect to m_X is similar to Eqs. (12) and (13). Such a definition may be useful in comparing preference for lotteries that are characterized by D-S belief functions similar to von Neumann-Morgenstern's utility theory for probabilistic lotteries [11].

Table 3: Expected value of $Y = 2X + 1$, a linear function

$\mathsf{a} \in 2^{\Omega_X}$	$m_X(\mathsf{a})$	$Q_{m_X}(\mathsf{a})$	$v_{m_X}(\mathsf{a})$	$E_{m_X}(X)$	$2\,v_{m_X}(\mathsf{a}))+1$	$E_{m_X}(g_X)$
$\{-1\}$	0.02	0.63	-1.00	0.059	-1.00	1.117
$\{0\}$	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	
$\mathsf{a}_Y \in 2^{\Omega_Y}$	$m_Y(\mathsf{a}_Y)$	$Q_{m_Y}(\mathsf{a}_Y)$	$v_{m_Y}(\mathsf{a}_Y)$	$E_{m_Y}(Y)$		
$\{-1\}$	0.02	0.63	-1.00	1.117		
$\{1\}$	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			

4.4 A comparison with expectation of pignistic and plausibility transforms

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 [9] and the plausibility [1] 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 Pl_P_m, there may be loss of information.

In general, the expected value defined in this paper may yield different values than the probabilistic expectation using pignistic or plausibility transformation. Table 4 compares the expectation defined in this paper with probabilisitic expectation using pignistic and plausibility transforms for the various BPAs described in Tables 1, 2, and 3. Two observations. First, although the three definitions yield different answers, they are all approximately of the same order of magnitude. Second, all three definitions satisfy the *expected value of a linear function of* X property. Thus, BPA m_Z in Table 2 can be obtained from BPA m_X in Table 1 using the transformation $Z = X + 2$. All three expected values satisfy the *expected value of a linear function of* X property. Also, BPA m_Y in Table 3 is obtained from BPA m_X in Table 1 using the transformation $Y = 2X + 1$. Again, all three expected values satisfy the *expected value of a linear function of* X property.

Table 4: A comparison of our expected value with probabilistic expectation using pignistic and plausibility transforms

BPA m	$E_m(\cdot)$	$E_{BetP_m}(\cdot)$	$E_{Pl\_P_m}(\cdot)$
m_X in Table 1	0.059	0.125	0.084
m_Y in Table 1	0.576	0.625	0.576
m_Z in Table 2	2.059	2.125	2.084
m_Y in Table 2	0.273	0.289	0.278
m_Y in Table 3	1.117	1.250	1.168

5 Summary and Conclusions

We propose a new definition of expected value for real-valued random variables whose uncertainty is described by D-S belief functions. Also, if we have a random variable with 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 Christiaan Huygens [3] in the context of the problem of points posed by Chevalier de Méré to Blaise Pascal.

The expectation operator can be used to define variance, covariance, correlation, and higher moments of D-S belief functions [8].

If we define $I(\mathsf{a}) = \log_2(\frac{1}{Q_{m_X}(\mathsf{a})})$ 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 [7], we define entropy of BPA m_X for X as an expected value of the function $I(\mathsf{a})$, i.e., $H(m_X) = E_{m_X}(I(\mathsf{a}))$. This is what is proposed in [5]. 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 conditional BPA for Y given X obtained by $\oplus\{m_{Y|x} : x \in \Omega_X\}$, and $m_{Y|x}$ is a conditional BPA for Y given $X = x$.

There are several decision theories for lotteries whose uncertainty is described by D-S belief functions theory. The most prominent ones are by Jean-Yves Jaffray [4]/Thomas Strat [10], and Philippe Smets [9]. The proposal by Jaffray/Strat is to first reduce a D-S belief function to an upper and lower PMFs, and then define an expected value that is a convex combination of the upper and lower probabilistic expectation. This proposal is justified in [4] by some axioms similar to the axioms proposed by John von Neumann and Oskar Morgenstern [11] for probabilistic lotteries. The proposal by Smets is to transform a D-S belief function to a corresponding PMF called the pignistic transform, and then use von Neumann-Morgenstern's expected utility theory. Our definition of expected value can be used in a decision theory for D-S theory without transforming belief functions to PMFs. This remains to be done.

Acknowledgments

This paper is a short version of [8], which 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 [5]. Thank you, Radim! Thanks also to Suzanna Emelio for a careful proofreading of this paper.

References

[1] B. R. Cobb and P. P. Shenoy. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3):314–340, 2006.

[2] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[3] C. Huygens. *De ratiociniis in ludo aleæ*. 1657 (English translation, published in 1714).

[4] J.-Y. Jaffray. Linear utility theory for belief functons. *Operations Research Letters*, 8(2):107–112, 1989.

[5] R. Jiroušek and P. P. Shenoy. A decomposable entropy of belief functions in the Dempster-Shafer theory. Working Paper 334, University of Kansas School of Business, Lawrence, KS 66045 USA, 2018.

[6] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[7] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[8] P. P. Shenoy. An expectation operator for belief functions in the Dempster-Shafer theory. Working Paper 335, University of Kansas School of Business, Lawrence, KS 66045, 2018.

[9] P. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. P. Srivastava and T. J. Mock, editors, *Belief Functions in Business Decisions*, pages 316–332. Physica-Verlag, Heidelberg, 2002.

[10] T. M. Strat. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4(5–6):391–417, 1990.

[11] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Priceton, NJ, 1953.