A Decomposable Entropy of Belief Functions in the Dempster-Shafer Theory

Radim Jiroušek[1] and Prakash P. Shenoy[2]

[1] Faculty of Management, University of Economics, and
The Czech Academy of Sciences, Institute of Information Theory and Automation
Jindřichův Hradec and Prague, Czech Republic `radim@utia.cas.cz`
[2] School of Business, University of Kansas, Lawrence, KS 66045 USA
`pshenoy@ku.edu`

Abstract. We define entropy of belief functions in the Dempster-Shafer (D-S) theory that satisfies a compound distributions property that is analogous to the property that characterizes Shannon's definitions of entropy and conditional entropy for discrete probability distributions. None of the existing definitions of entropy for belief functions in the D-S theory satisfy such a compound distributions property. We describe some important properties of our definition.

1   Introduction

The main goal of this paper is to define entropy of belief functions in the Dempster-Shafer's theory [2] [4] that satisfies a compound distributions property analogous to the one that characterizes Shannon's definitions of entropy and conditional entropy for discrete probability distributions [6]. If $P_{X,Y}$ is a probability mass function (PMF) of (X, Y), and it is decomposed into PMF P_X for X, and conditional probability table $P_{Y|X}$ so that $P_{X,Y} = P_X \otimes P_{Y|X}$, then Shannon's definitions of entropy and conditional entropy satisfy $H_s(P_{X,Y}) = H_s(P_X) + H_s(P_{Y|X})$. Here, $\otimes$ denotes probabilistic combination, which is pointwise multiplication followed by normalization.

In this paper, we provide definitions of entropy and conditional entropy of belief functions, so that if $m_{X,Y}$ is a basic probability assignment (BPA) for (X, Y) that is constructed from a BPA m_X for X, and a conditional BPA $m_{Y|X}$ for Y given X, such that $m_{X,Y} = m_X \oplus m_{Y|X}$, where $\oplus$ is Dempster's combination rule, then our definitions satisfy $H(m_{X,Y}) = H(m_X) + H(m_{Y|X})$. This is the main contribution of this paper. Our definitions of entropy and conditional entropy have several nice properties similar to corresponding properties of Shannon's entropy. Here, we do not delve into philosophical discussions about what entropy means. Our exposition focusses exclusively on mathematical properties of entropy.

2 Shannon's Definition of Entropy

In this section, we briefly review Shannon's definition of entropy of PMFs of discrete random variables, and its properties. Most of the material in this section is taken from [6].

Definition 1. Suppose P_X is a PMF of discrete variable X. The *entropy* of P_X, denoted by $H_s(P_X)$, is defined as

$$H_s(P_X) = - \sum_{x \in \Omega_X} P_X(x) \log_2 \left(P_X(x) \right). \tag{1}$$

Suppose $P_{X,Y}$ is a joint PMF of (X, Y). Then, the *joint* entropy of $P_{X,Y}$ is as in Eq. (1), i.e.,

$$H_s(P_{X,Y}) = - \sum_{(x,y) \in \Omega_{X,Y}} P_{X,Y}(x,y) \log_2 \left(P_{X,Y}(x,y) \right).$$

Suppose $P_{X,Y}$ is a PMF of (X, Y) with P_X as its marginal PMF for X. Suppose we observe $X = a$ for some $a \in \Omega_X$ such that $P_X(a) > 0$. This observation is represented by the PMF $P_{X=a}$ for X such that $P_{X=a}(a) = 1$. Let $P_{Y|a} = (P_{X,Y} \otimes P_{X=a})^{\downarrow Y}$ denote the posterior PMF of Y, where $\otimes$ denotes pointwise multiplication followed by normalization, the combination rule in probability theory. The *posterior* entropy of $P_{Y|a}$ is as in Eq. (1), i.e., $H_s(P_{Y|a}) = -\sum_{y \in \Omega_Y} P_{Y|a}(y) \log_2(P_{Y|a}(y))$.

Shannon [6] derives the expression for entropy of P_X axiomatically using four axioms as follows:

1. Axiom 1 (*Existence*): $H(P_X)$ exists.
2. Axiom 2 (*Continuity*): $H(P_X)$ should be a continuous function of P_X.
3. Axiom 3 (*Monotonicity*): If we have an equally likely PMF, then $H(P_X)$ should be a monotonically increasing function of $|\Omega_X|$.
4. Axiom 4 (*Compound distributions*): If a PMF is factored into two PMFs, then its entropy should be the sum of entropies of its factors, e.g., $P_{X,Y}(x,y) = P_X(x) P_{Y|x}(y)$, then $H(P_{X,Y}) = H(P_X) + \sum_{x \in \Omega_X} P_X(x) H(P_{Y|x})$.

Shannon [6] proves that the only function H_s that satisfies Axioms 1–4 is of the form $H_s(P_X) = -K \sum_{x \in \Omega_X} P_X(x) \log \left(P_X(x) \right)$, where K is a constant that depends on the choice of units of measurement.

Let $P_{Y|X} : \Omega_{X,Y} \to [0,1]$ be a function such that $P_{Y|X}(x,y) = P_{Y|x}(y)$ for all $(x,y) \in \Omega_{X,Y}$. As $P_{Y|x}(y)$ is only defined for $x \in \Omega_X$ such that $P_X(x) > 0$, we will assume that $P_{Y|X}$ is only defined for $x \in \Omega_X$ such that $P_X(x) > 0$. $P_{Y|X}$ is not a PMF, but can be considered as a collection of PMFs, and it is called a conditional probability table (CPT) in the Bayesian network literature. If we combine P_X and $P_{Y|X}$, we obtain $P_{X,Y}$, i.e., $P_{X,Y} = P_X \otimes P_{Y|X}$.

Definition 2. Suppose $P_{Y|X}$ is a CPT of Y given X for all $x \in \Omega_X$ such that $P_X(x) > 0$. Then the *conditional* entropy of $P_{Y|X}$ is defined as

$$H_s(P_{Y|X}) = \sum_{x \in \Omega_X} P_X(x) H_s(P_{Y|x}). \tag{2}$$

It follows from Axiom 4 that

$$H_s(P_{X,Y}) = H_s(P_X \otimes P_{Y|X}) = H_s(P_X) + H_s(P_{Y|X}). \tag{3}$$

3 Basic Definitions of the D-S Belief Functions Theory

In this section we review the basic definitions in the D-S belief functions theory, including functional representations of uncertain knowledge, and operations for making inferences from such knowledge.

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

BPAs. 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 BPA m for X is a function $m : 2^{\Omega_X} \rightarrow [0, 1]$ such that

$$\sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1. \tag{4}$$

The non-empty subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m(\mathsf{a}) > 0$ are called *focal* elements of m. We say m is *consonant* if the focal elements of m are nested, i.e., if $\mathsf{a}_1 \subset \ldots \subset \mathsf{a}_m$, where $\{\mathsf{a}_1, ..., \mathsf{a}_r\}$ denotes the set of all focal elements of m. We say m is *quasi-consonant* if the intersection of all focal elements of m is non-empty. A BPA that is consonant is also 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 are singleton subsets of Ω_X, then we say m is *Bayesian*. In this case, m is equivalent to the PMF P for X such that $P(x) = m(\{x\})$ for each $x \in \Omega_X$. If Ω_X is a focal element, then we say m is *non-dogmatic*, and *dogmatic* otherwise. Thus, a Bayesian BPA is dogmatic.

Commonality Functions. The information in a BPA m can also be represented by a corresponding commonality function Q_m that is defined as follows.

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}) \tag{5}$$

for all $\mathsf{a} \in 2^{\Omega_X}$. Q_m is a non-increasing function in the sense that if $\mathsf{b} \subseteq \mathsf{a}$, then $Q_m(\mathsf{b}) \geq Q_m(\mathsf{a})$. Finally, Q_m is a normalized function in the sense that

$$\sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X}} m(\mathsf{b}) = 1. \tag{6}$$

Thus, any non-increasing, non-negative function that satisfies Eq. (6) qualifies as a commonality function.

Next, we describe two main operations for making inferences.

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 [2] and obtain the BPA $m_1 \oplus m_2$, which represents the combined evidence.

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 $\mathsf{a}^{\uparrow \{X, Y\}} = \mathsf{a} \times \Omega_Y$.

Suppose m_X is a BPA for X, and $\mathcal{X}$ is such that $X \in \mathcal{X}$. Then the vacuous extension of m to $\mathcal{X}$, denoted by $m_X^{\uparrow \mathcal{X}}$, is the BPA for $\mathcal{X}$ such that $m_X^{\uparrow \mathcal{X}}(\mathsf{a}^{\uparrow \mathcal{X}}) = m_X(\mathsf{a})$, for all $\mathsf{a} \in 2^{\Omega_X}$, i.e., all focal elements of $m_X^{\uparrow \mathcal{X}}$ are vacuous extensions of focal elements of m_X to $\mathcal{X}$, and they have the same corresponding values.

We will define Dempster's rule in terms of commonality functions [4]. Suppose m_1 and m_2 are BPAs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Suppose $Q_{m_1^{\uparrow \mathcal{X}}}$ and $Q_{m_2^{\uparrow \mathcal{X}}}$ are commonality functions corresponding to BPAs $m_1^{\uparrow \mathcal{X}}$ and $m_2^{\uparrow \mathcal{X}}$, respectively, where $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. The commonality function $Q_{m_1 \oplus m_2}$ corresponding to BPA $m_1 \oplus m_2$ is

$$Q_{m_1 \oplus m_2}(\mathsf{a}) = K^{-1} Q_{m_1^{\uparrow \mathcal{X}}}(\mathsf{a}) \, Q_{m_2^{\uparrow \mathcal{X}}}(\mathsf{a}), \tag{7}$$

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

$$K = \sum_{\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}} (-1)^{|\mathsf{a}|+1} Q_{m_1^{\uparrow \mathcal{X}}}(\mathsf{a}) \, Q_{m_2^{\uparrow \mathcal{X}}}(\mathsf{a}). \tag{8}$$

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*.

Marginalization 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 $\mathsf{a} \in 2^{\Omega_{\mathcal{X}_1}}$,

$$m^{\downarrow \mathcal{X}_1}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{\mathcal{X}} : \mathsf{b}^{\downarrow \mathcal{X}_1} = \mathsf{a}}} m(\mathsf{b}). \tag{9}$$

Conditional belief functions. 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 $\sum_{\mathsf{a} \in 2^{\Omega_Y}} m_{Y|x}(\mathsf{a}) = 1$. We can embed this BPA for Y into a conditional BPA for (X, Y), which is denoted by $m_{x,Y}$, such that the following four conditions hold. First, $m_{x,Y}$ tells us nothing about X, i.e., $m_{x,Y}^{\downarrow X}(\Omega_X) = 1$. Second, $m_{x,Y}$ tells us nothing about Y, i.e., $m_{x,Y}^{\downarrow Y}(\Omega_Y) = 1$. Third, if we combine $m_{x,Y}$ with the 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., $(m_{x,Y} \oplus m_{X=x})^{\downarrow Y} = m_{Y|x}$. Fourth, if we

combine $m_{x,Y}$ with the BPA $m_{X=\bar{x}}$ for X such $m_{X=\bar{x}}(\{\bar{x}\}) = 1$ using Dempster's rule, and marginalize the result to Y we obtain the vacuous BPA for Y, i.e., $(m_{x,Y} \oplus m_{X=\bar{x}})^{\downarrow Y}(\Omega_Y) = 1$. One way to obtain such an embedding is suggested by Smets [7] (see also [5]), 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 all four conditions mentioned above.

This completes our brief review of the D-S belief function theory. For further details, the reader is referred to [4].

4   A Decomposable Entropy for the D-S Theory

In this section, we provide a new definition of entropy of belief functions in the D-S theory, and describe its properties. This new definition is designed to satisfy a compound distributions property analogous to the compound distribution property that characterizes Shannon's entropy of PMFs.

Definition 3. Suppose m_X is a BPA for X with state space Ω_X, and suppose Q_{m_X} denotes the commonality function corresponding to m_X. Then the entropy of m_X, denoted by $H(m_X)$, is defined as follows:

$$H(m_X) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|} Q_{m_X}(\mathsf{a}) \log_2(Q_{m_X}(\mathsf{a})). \tag{10}$$

If $Q_{m_X}(\mathsf{a}) = 0$, we will follow the convention that $Q_{m_X}(\mathsf{a}) \log_2(Q_{m_X}(\mathsf{a})) = 0$ as $\lim_{\theta \to 0^+} \theta \log_2(\theta) = 0$.

This is a new definition of entropy that has not been proposed earlier in the literature. The closest definition is due to Smets [8], where $H(m)$ is defined as

$$H(m) = -\sum_{\mathsf{a} \in 2^{\Omega_X}} \log_2\left(Q_m(\mathsf{a})\right),$$

but only for non-dogmatic BPAs m. Our definition holds for all BPAs. Also, our sum is an alternating weighted sum, whose sign depends on the cardinality of non-empty subset a.

Suppose $m_{X,Y}$ is a joint BPA for (X,Y). Then the *joint* entropy of $m_{X,Y}$ is as in Eq. (10), i.e.,

$$H(m_{X,Y}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|} Q_{m_{X,Y}}(\mathsf{a}) \log_2(Q_{m_{X,Y}}(\mathsf{a})).$$

Suppose $m_{X,Y}$ is a BPA for (X,Y) with m_X as its marginal BPA for X. Suppose we observe $X = a$ for some $a \in \Omega_X$ such that $m_X(\{a\}) > 0$. This observation is represented by the BPA $m_{X=a}$ such that $m_{X=a}(\{a\}) = 1$. Let $m_{Y|a} = (m_{X,Y} \oplus m_{X=a})^{\downarrow Y}$ denote the posterior BPA for Y, and its *posterior* entropy is as in Eq. (10), i.e., $H(m_{Y|a}) = \sum_{\mathsf{a} \in 2^{\Omega_Y}} (-1)^{|\mathsf{a}|} Q_{m_{Y|a}}(\mathsf{a}) \log_2(Q_{m_{Y|a}}(\mathsf{a}))$.

The following theorem says vacuous extension of a BPA does not change its entropy.[3]

Theorem 1. *If m is a BPA for X with $\Omega_X = \{x, \bar{x}\}$, and m' is a vacuous extension of m to (X, Y), where $\Omega_Y = \{y, \bar{y}\}$, then $H(m') = H(m)$.*

Definition 4. Suppose m_X is a BPA for X such that $m_X(x) > 0$. Let $m_{x,Y}$ denote a BPA for (X, Y) representing a conditional BPA of Y given $X = x$. We define entropy of *conditional* BPA $m_{x,Y}$ as follows:

$$H(m_{x,Y}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|} Q_{m_X^{\uparrow\{X,Y\}}}(\mathsf{a}) Q_{m_{x,Y}}(\mathsf{a}) \log_2(Q_{m_{x,Y}}(\mathsf{a})). \qquad (11)$$

The definition in Eq. (11) is analogous to Eq. (2) for the probabilistic case. We have the following result about conditional entropy.

Theorem 2. *Suppose m_X is a BPA for X such that $\Omega_X = \{x, \bar{x}\}$ and $m_X(\{x\}) > 0$. Suppose Y is such that $\Omega_Y = \{y, \bar{y}\}$, and $m_{Y|x}$ is a BPA for Y given $X = x$. Let $m_{x,Y}$ denote a conditional BPA for (X, Y) obtained from $m_{Y|x}$ by conditional embedding. Then,*

$$H(m_{x,Y}) = m_X(\{x\}) H(m_{Y|x}). \qquad (12)$$

If $\Omega_X = \{x, \bar{x}\}$ and assuming $m_X(\bar{x}) > 0$, it follows from Eq. (11) that

$$H(m_{\bar{x},Y}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|} Q_{m_X^{\uparrow\{X,Y\}}}(\mathsf{a}) Q_{m_{\bar{x},Y}}(\mathsf{a}) \log_2(Q_{m_{\bar{x},Y}}(\mathsf{a})).$$

Also, from Theorem 2, it follows that:

$$H(m_{\bar{x},Y}) = m_X(\{\bar{x}\}) H(m_{Y|\bar{x}}).$$

As the contexts in $m_{x,Y}$ and $m_{\bar{x},Y}$ are disjoint, and the beliefs of the contexts are described by the same BPA m_X such that $m_X(x) > 0$ and $m_X(\bar{x}) > 0$, we have the following result.

Theorem 3. *Suppose X and Y are such that $\Omega_X = \{x, \bar{x}\}$, and $\Omega_Y = \{y, \bar{y}\}$. Suppose that we have non-vacuous conditional BPAs $m_{Y|x}$, and $m_{Y|\bar{x}}$ for Y such that $m_X(\{x\}) > 0$, $m_X(\{\bar{x}\}) > 0$, and after conditional embedding, these are represented by conditional BPAs $m_{x,Y}$ and $m_{\bar{x},Y}$ for (X, Y). Then,*

$$H(m_{Y|X}) = H(m_{x,Y} \oplus m_{\bar{x},Y}) = H(m_{x,Y}) + H(m_{\bar{x},Y}). \qquad (13)$$

Notice that the result in Eq. (13) is analogous of the definition of conditional entropy in Eq. (2) in the probabilistic case.

---

[3] For lack of space, proofs of all theorems and properties are omitted, and can be found in a working paper that can be downloaded from `http://pshenoy.faculty.ku.edu/Papers/WP334.pdf`

Next, we state the main result of this paper.

Theorem 4. *Suppose X and Y are such that $\Omega_X = \{x, \bar{x}\}$, and $\Omega_Y = \{y, \bar{y}\}$. Suppose m_X is a BPA for X such that $m_X > 0$ and $m_X(\bar{x}) > 0$, and $m_{Y|X} = m_{x,Y} \oplus m_{\bar{x},Y}$ is a conditional BPA for Y given X. Let $m_{X,Y} = m_X \oplus m_{Y|X}$. Then,*

$$H(m_{X,Y}) = H(m_X) + H(m_{Y|X}). \tag{14}$$

Next, we show that a probability model for (X, Y) can be replicated exactly in the DS theory. Furthermore, our definition of entropy for all BPAs will coincide with Shannon's entropy of the corresponding probabilistic function.

Theorem 5. *Suppose X and Y are such that $\Omega_X = \{x, \bar{x}\}$, and $\Omega_Y = \{y, \bar{y}\}$. Suppose P_X is a PMF for X such that $P_X(x) > 0$, and $P_X(\bar{x}) > 0$, and $P_{Y|X}$ is a CPT for Y given X, i.e., $P_{Y|X}(x, y) = P_{Y|x}(y)$, where $P_{Y|x}$ is the conditional PMF for Y given $X = x$ for all $(x, y) \in \Omega_{X,Y}$. Let $P_{X,Y} = P_X \otimes P_{Y|X}$. Let m_X denote the Bayesian BPA corresponding to P_X, let $m_{Y|x}$ and $m_{Y|\bar{x}}$ denote the Bayesian BPAs for Y corresponding to PMFs $P_{Y|x}$ and $P_{Y|\bar{x}}$ for Y. Let $m_{x,Y}$ and $m_{\bar{x},Y}$ denote the conditional BPAs for (X, Y) obtained by conditional embedding of $m_{Y|x}$ and $m_{Y|\bar{x}}$. Let $m_{Y|X} = m_{x,Y} \oplus m_{\bar{x},Y}$, and let $m_{X,Y} = m_X \oplus m_{Y|X}$. Then, $m_{X,Y}$ is a Bayesian BPA for (X, Y) corresponding to PMF $P_{X,Y}$,*

$$H(m_{X,Y}) = H_s(P_{X,Y}), \tag{15}$$

$$H(m_X) = H_s(P_X), \text{ and} \tag{16}$$

$$H(m_{Y|X}) = H_s(P_{Y|X}). \tag{17}$$

Notice that $m_{x,Y}$, $m_{\bar{x},Y}$, and $m_{Y|X}$, are not Bayesian BPAs.

5   Other Properties

Some further properties of our definition in Eq. (10) are as follows.

   Non-negativity. Suppose m is a BPA for X and suppose $|\Omega_X| = 2$. Then, $H(m) \geq 0$. For $|\Omega_X| > 2$, $H(m)$ does *not* satisfy the non-negativity property.

Example 1. Consider a BPA m for X with $\Omega_X = \{a, b, c\}$ such that $m(\{a, b\}) = m(\{a, c\}) = m(\{b, c\}) = \frac{1}{3}$. Then Q_m is as follows: $Q_m(\{\mathsf{a}\}) = Q_m(\{\mathsf{b}\}) = Q_m(\{\mathsf{c}\}) = \frac{2}{3}$, $Q_m(\{a, b\}) = Q_m(\{a, c\}) = Q_m(\{b, c\}) = \frac{1}{3}$, and $Q_m(\{a, b, c\}) = 0$. Then, $H(m) = -3 \cdot \frac{2}{3} \log_2(\frac{2}{3}) + 3 \cdot \frac{1}{3} \log_2(\frac{1}{3}) = \log_2(\frac{3}{4}) \approx -0.415$.   □

We conjecture that $H(m) \geq \log_2(\frac{n}{2(n-1)})$, where $n = |\Omega_X|$. This is based on a BPA where each of $\binom{n}{2}$ doubleton subsets has a mass of $1/\binom{n}{2}$. If the conjecture is true, $H(m)$ would be on the scale from $[\log_2(\frac{n}{2(n-1)}), \log_2(n)]$. $\lim_{n \to \infty} \log_2(\frac{n}{2(n-1)}) = -1$. Lack of non-negativity is not a serious drawback. Shannon's definition of entropy for continuous random variables characterized by probability density functions can be negative [6].

Quasi-consonant. Suppose m is a BPA for X. If m is quasi-consonant, then $H(m) = 0$. As consonant BPAs are also quasi-consonant, $H(m) = 0$ for consonant BPAs. This property suggests that $H(m)$ is a measure of "dissonance" in m.

Maximum entropy. Suppose m is a BPA for X with state space Ω_X. Then, $H(m) \leq \log_2(|\Omega_X|)$, with equality if and only if $m = m_u$, where m_u is the Bayesian equiprobable BPA for X. This is similar to the corresponding property of Shannon's definition for PMFs.

6 Summary & Conclusion

The most important property of our definition of entropy is the compound distributions property. Such a property is not satisfied by any of the past definitions of entropy reviewed in [3], nor by the definition proposed there. The compound distributions property is fundamental to Shannon's definition of entropy as it constitutes the main property that characterizes Shannon's definition.

We should also note that the compound distributions property only applies to belief functions that are constructed from marginals and conditional belief functions. Given an arbitrary joint belief function, it is not always possible to factor it into marginals and conditionals that produce the given joint. Thus, our new definition is of particular interest for the class of joint belief functions that do factor into marginals and conditionals. In particular, it applies to graphical belief functions that are constructed from directed acyclic graphs models, also known as Bayesian networks, but whose potentials are described by belief functions [1].

This is work in progress. Although we have stated Theorems 1–5 for the case where X and Y are binary, we believe these theorems hold more generally, and are currently in the process of proving them.

References

1. R. G. Almond. *Graphical Belief Modeling*. Chapman & Hall, London, UK, 1995.
2. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.
3. R. Jiroušek and P. P. Shenoy. A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92(1):49–65, 2018.
4. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
5. G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B*, 44(3):322–352, 1982.
6. C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.
7. P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical*. PhD thesis, Free University of Brussels, 1978.
8. P. Smets. Information content of an evidence. *International Journal of Man-Machine Studies*, 19:33–43, 1983.