

Evaluation of Evidential Combination Operators

Alexander Karlsson

Infofusion/Informatics Research Center
University of Skövde, Sweden
alexander.karlsson@his.se

H. Joe Steinbauer

Infofusion/Informatics Research Center
University of Skövde, Sweden
joe.steinbauer@his.se

Abstract

We present an experiment for evaluating precise and imprecise evidential combination operators. The experiment design is based on the assumption that only limited statistical information is available in the form of multinomial observations. We evaluate three different evidential combination operators; one precise, the Bayesian combination operator, and two imprecise, the credal and Dempster's combination operator, for combining independent pieces of evidence regarding some discrete state space of interest. The evaluation is performed by using a score function that takes imprecision into account. The results show that the precise framework seems to perform equally well as the imprecise frameworks.

Keywords. Evidential combination, imprecise probability, credal sets.

1 Introduction

The problem of *combining independent pieces of evidence*, most often stemming from multiple sources of information, e.g., sensors, is common in many application scenarios [24]. Typically such applications involve one or several sensors where for each sensor a feature can be extracted and used for constructing an appropriate *evidence* with regard to the unknown state of interest. Even though the pieces of evidence might not be completely independent, in many application scenarios (e.g., [21]) where different sources are used, e.g., different type of sensors, it is reasonable to assume independence.

In order to investigate the question of how well different evidence combination operators, precise and imprecise, perform compared to each other, we design an experiment for an object recognition scenario. We restrict the comparison to three different evidential combination operators, Bayesian [1, 2, 20], credal [1, 2, 20], and Dempster's combination operator [26].

The latter is one of the most commonly used operator for combining pieces of evidence. The obvious difference between the operators is that the Bayesian one is precise, i.e., its operands are based on a single function, and the other two are imprecise, i.e., the operands are either a set of functions or can be cast to a set of functions.

Karlsson et al. [20] have previously empirically compared the performance of the Bayesian and credal combination operators. They found that the Bayesian combination operator performs better due to the fact that the credal counterpart could “overestimate” imprecision and in a sense become too “cautious”. However, in that evaluation imprecision was inherent in the state estimation problem, i.e., imprecise operands were sampled directly, without any particular statistical information, and the Bayesian operator was applied on the centroids of these operators while the credal counterpart was applied directly on the operands.

In contrast to the work by Karlsson et al. [20], we here design an experiment specifically aimed at evaluating the performance of combination operators when only a limited statistical amount of information is available and used in the precise and imprecise statistical models, namely Dirichlet models. This type of situation, i.e., when only limited information is available, is often one of the main motivations for using *imprecise probability* (including credal sets) [31, 33]. In addition, we also include Dempster's combination operator, i.e., another imprecise operator, in our evaluation. Since the imprecise operators have exponential worst case complexity in comparison to the precise one, we are specifically interested in comparing these two classes of operators.

The paper is organized as follows: in Section 2, we describe the different operators considered in the experiment. In Section 3, we elaborate on the design, performance, and result of the empirical evaluation, and lastly, in Section 4, we summarize the work pre-

sented and discuss the results of the evaluation as well as include ideas for future research.

2 Preliminaries

In this section, we present the different *combination operators* that we later will use in the empirical evaluation. One important aspect to note for all of these combination operators is that the pieces of evidences must satisfy different types of independence requirements which will discuss preceding each formal definition of the operators. Due to this requirement, a joint evidence is most often stronger if both operands constitute strong evidence for a certain state; this state gets reinforced in the combination. This should be put in contrast to other operators that often goes under the name *aggregation operators* [30] which are typically more “consensus-inspired” which means that the joint result is an agreement between the operands, e.g., if both operands are identical the joint result is equivalent to the operands.

2.1 Bayesian Combination

A Bayesian approach to combining independent pieces of evidence can be derived by modeling evidence as *likelihood functions* [1, 2, 20]. To realize this, assume that we have a random variable X taking values $x \in \Omega_X$. Furthermore, assume that we can obtain observations y_1, y_2 , from two different sources within the environment of interest and that these observations are informative about X in the sense that your *belief*, i.e., the *posterior probability* $p(X|y_1, y_2)$ could be affected. Now since:

$$p(X|y_1, y_2) = \frac{p(y_1, y_2|X)p(X)}{\sum_{x \in \Omega_X} p(y_1, y_2|x)p(x)}, \quad (1)$$

we see that the only way the observations can affect the belief $p(X|y_1, y_2)$ is through the *joint likelihood* $p(y_1, y_2|X)$. By assuming that the observations are *conditionally independent* given that we know the true state of X , we obtain:

$$p(y_1, y_2|X) = p(y_1|X)p(y_2|X) . \quad (2)$$

In terms of evidence, the above equation is a simple method for combining two independent pieces of evidence, i.e., likelihood functions, into a single *joint evidence*, i.e., a joint likelihood function. However, in order to avoid monotonically decreasing values of the joint evidence, we normalize after each combination to obtain a probability function. Such a normalization can be performed without loss of generality since it is the relative strength of the likelihoods that constitute the evidence structure and such relativeness is

preserved under normalization (see further Karlsson et al. (2011) [20]).

Definition 1. *The Bayesian combination operator Φ_B is defined as [1, 2, 20]:*

$$\Phi_B(\hat{p}(y_1|X), \hat{p}(y_2|X)) \triangleq \frac{\hat{p}(y_1|X)\hat{p}(y_2|X)}{\sum_{x \in \Omega_X} \hat{p}(y_1|x)\hat{p}(y_2|x)}, \quad (3)$$

where $\hat{p}(y_i|X)$, $i \in \{1, 2\}$, are normalized likelihood functions and where the joint evidence $\Phi_B(\hat{p}(y_1|X), \hat{p}(y_2|X))$ satisfies the *conditionally independence assumption* in Eq. (2). The operator is undefined when $\sum_{x \in \Omega_X} \hat{p}(y_1|x)\hat{p}(y_2|x) = 0$.

Note that when the denominator is zero, the sources mutually exclude all possibilities within the state space which is a contradiction to the assumption that the truth exists within this space (given the closed world assumption). From this viewpoint it is quite natural that the operator is undefined for such cases. One way of handling such situation is to perform discounting [15, 20].

2.2 Credal Combination

Credal combination is a straightforward generalization of the Bayesian combination operator to *imprecise probability* [33]. It relies on the notion of *credal sets* [23, 10, 11], i.e., *closed convex sets of probability functions*. Such sets can be conveniently represented by *extreme points* and therefore one uses credal sets in the form of *polytopes* since such a structure guarantees a finite number of such points. The combination schema was introduced as the *robust Bayesian combination operator* by Arnborg [1, 2], and further studied by Karlsson et al. [20] as the *credal combination operator*¹.

The main reason for considering *imprecision* in the form of credal sets is that it allows one to model problems when only scarce information is available regarding the environment of interest [31]. In such cases it can be considered to be more realistic to express, e.g., probabilities in terms of intervals instead of single probability values. Credal sets can also be thought of as being a result of performing sensitivity analysis as in robust Bayesian theory [17, 4].

In order to generalize the Bayesian combination operator in Def. 1 to a credal counterpart, we start by modeling evidence by credal sets of normalized likelihoods functions, denoted $\hat{\mathcal{P}}(y_1|X)$ and $\hat{\mathcal{P}}(y_2|X)$, instead of a single normalized likelihood function, where

¹We denote this operator as the credal combination operator for the simple reason that we do not want to impose any particular interpretation of the imprecision as “robust” imposes a sensitivity-analysis interpretation.

as previous X denotes a random variable for something unknown of interest in the environment and y_1 and y_2 are the observations. In order to model independent pieces of evidence we use a generalization of conditional independence denoted *strong independence* [9], which requires that all extreme points must factorize, i.e:

$$\hat{p}_e(y_1, y_2|X) = \hat{p}(y_1|X)\hat{p}(y_2|X), \quad (4)$$

$\forall \hat{p}_e(y_1, y_2|X) \in \mathcal{E}(\hat{\mathcal{P}}(y_1, y_2|X))$ where $\mathcal{E}(\cdot)$ denotes the set of extreme points and where $\hat{p}(y_i|X) \in \hat{\mathcal{P}}(y_i|X)$, $i \in \{1, 2\}$. By using this independence assumption the credal combination operator can be defined in terms of applying the Bayesian combination operator point-wise on all combinations of functions in the operand credal sets and as a last step applying the convex-hull operator $\mathcal{CH}(\cdot)$ in order to fulfill convexity of the joint evidence.

Definition 2. *The credal combination operator $\Phi_{\mathcal{C}}$ is defined as [1, 2, 20]:*

$$\begin{aligned} \Phi_{\mathcal{C}}(\hat{\mathcal{P}}(y_1|X), \hat{\mathcal{P}}(y_2|X)) \triangleq \\ \mathcal{CH}\left(\left\{\Phi_{\mathcal{B}}(\hat{p}(y_1|X), \hat{p}(y_2|X)) : \right. \right. \\ \left. \left. \hat{p}(y_i|X) \in \hat{\mathcal{P}}(y_i|X), i \in \{1, 2\}\right\}\right), \end{aligned} \quad (5)$$

where $\hat{\mathcal{P}}(y_i|X)$, $i \in \{1, 2\}$, are credal sets of normalized likelihoods functions and where the joint evidence $\Phi_{\mathcal{C}}(\hat{\mathcal{P}}(y_1|X), \hat{\mathcal{P}}(y_2|X))$ satisfies the conditional independence assumption in Eq. (4). The operator is undefined if $\Phi_{\mathcal{B}}(\hat{p}(y_1|X), \hat{p}(y_2|X))$ is undefined for any pair $\hat{p}(y_1|X) \in \hat{\mathcal{P}}(y_1|X)$, $\hat{p}(y_2|X) \in \hat{\mathcal{P}}(y_2|X)$.

Note that the credal combination operator inherits the property of being undefined for cases where the denominator is zero (see further the discussion after Def. 1) and that when only singleton sets are used the operator is equivalent to the Bayesian combination operator. For computation of $\Phi_{\mathcal{C}}$, one can restrict the application of the Bayesian combination operator to the extreme points of the operand credal sets [20, Theorem 2], i.e:

$$\begin{aligned} \Phi_{\mathcal{C}}(\hat{\mathcal{P}}(y_1|X), \hat{\mathcal{P}}(y_2|X)) = \\ \Phi_{\mathcal{C}}(\mathcal{E}(\hat{\mathcal{P}}(y_1|X)), \mathcal{E}(\hat{\mathcal{P}}(y_2|X))) . \end{aligned} \quad (6)$$

In order to measure the degree of imprecision of a credal set, we will utilize the following measure [20], which can be thought of as the average degree of imprecision for single events [31]:

$$\mathcal{I}(\mathcal{P}(X)) \triangleq \frac{1}{|\Omega_X|} \sum_{x \in \Omega_X} \left(\max_{p(X) \in \mathcal{P}(X)} p(x) - \min_{p(X) \in \mathcal{P}(X)} p(x) \right) \quad (7)$$

2.3 Dempster-Shafer Combination

Dempster-Shafer theory [12, 26], also known as *evidence theory*, is a variant of imprecise probability [33], where one models evidence imprecisely by so called *mass functions*. A mass function assigns mass to subsets $A \subseteq \Omega_X$. The idea is that this schema can be useful in cases where a source is only partly sure of the true value of X , e.g. for $\Omega_X = \{x_1, x_2, x_3\}$, a source might be able to exclude the alternative x_3 but not be able to specify more clearly whether the truth is x_1 or x_2 .

Formally, a mass function is a mapping from the power set of the state space Ω_X , also known as the *frame of discernment*, to the interval $[0, 1]$:

$$m : 2^{\Omega_X} \rightarrow [0, 1] \quad (8)$$

$$m(\emptyset) = 0 \quad (9)$$

$$\sum_{A \subseteq \Omega_X} m(A) = 1 \quad (10)$$

Two additional functions that are often encountered when considering Dempster-Shafer theory are *belief* and *plausibility*, denoted $Bel(A)$ and $Pl(A)$, respectively, and defined by:

$$Bel(A) \triangleq \sum_{B \subseteq A} m(B) \quad (11)$$

$$Pl(A) \triangleq \sum_{B \cap A \neq \emptyset} m(B), \quad (12)$$

where $Bel(A)$ can be interpreted as the sum of all evidence that supports A and $Pl(A)$ as the sum of all evidence that does not contradict A . Belief and plausibility can also be regarded as a *lower and upper bound* for the probability of A , i.e:

$$Bel(A) \leq p(A) \leq Pl(A) . \quad (13)$$

The concept of independent pieces of evidence in Dempster-Shafer theory, also known as *distinct evidences*, is a bit problematic [27]. However, when the mass functions only operate on singleton sets, independent pieces of evidence can be defined in the same way as for the Bayesian combination operator, i.e., by using an assumption of conditional independence [27]. In the other cases, this assumption does not work, however, according to Smets [27], independent pieces of evidence can “in practice” be defined as:

$$m_{1,2}(A) = \begin{cases} m_1(B)m_2(C) & \text{if } A = B \times C \\ 0 & \text{Otherwise} \end{cases}, \quad (14)$$

where $B \subseteq \Omega_{X_1}$ and $C \subseteq \Omega_{X_2}$, i.e., ordinary stochastic independence.

Given two independent pieces of evidence m_1 and m_2 , e.g., in the sense of Eq. (14), we can combine them into a joint evidence $m_{1,2}$ utilizing *Dempster’s combination operator* [12].

Definition 3. *Dempster’s combination operator* $\Phi_{\mathcal{D}}$ is defined as [12, 26]:

$$\Phi_{\mathcal{D}}(A, m_1, m_2) \triangleq \frac{1}{1-k} \sum_{\substack{B \cap C = A \\ B, C \subseteq \Omega_X}} m_1(B)m_2(C), \quad (15)$$

where k is the conflict between evidence m_1 and m_2 , defined by:

$$k = \sum_{\substack{A \cap B = \emptyset \\ A, B \subseteq \Omega_X}} m_1(A)m_2(B). \quad (16)$$

The operator is undefined when $k = 1$.

Dempster’s combination operator is related to the Bayesian combination operator in the way that in case the mass is distributed only among singletons of Ω_X , the two operators produce the same results. Also note that similar to the Bayesian and credal combination operator, the operator is undefined in cases where the sources mutually exclude all possibilities of the state space.

Since a mass function imposes lower and upper bounds on a probability function, seen in Eq. (13), one can transform a mass function into a credal set. The question then arises if the mass function as a result of Dempster’s combination operator applied on two operands yields a mass function that when transformed to a credal set is equivalent to the result of first transforming the same operands to credal sets and then use the credal combination operator? Arnborg [1, 2] has shown that this is not the case, in fact the resulting credal sets can even be disjoint, hence the credal and Dempster’s combination operator are clearly different.

3 Empirical Evaluation

In this section we elaborate on the experiment design for evaluating the combination operators previously presented. We start by providing an overview of the application scenario where the combination takes place, and then move on to describe the design including assumptions, parameters, and score functions.

3.1 Overview

Assume that we want to implement an *object recognition* algorithm based on two different types of sensors: a camera and a microphone (we assume that the

objects of interest produce some form of sound). Naturally, using both sensors for performing the recognition should yield a better result than only using one. Since we utilize different sensors, that observes different features of the object, it is fair to make the assumption that the sensor readings yields independent pieces of evidence. As an example, if the object is positioned at an “unfamiliar” angle, yielding ambiguous output from an image analysis algorithm, this can be compensated for by the output from a pattern matching algorithm performed on the signal from the microphone. Also, if both sensor yields features that constitute evidence for one particular object, one would obtain an evidence that is reinforced towards that object.

Let the unknown object be denoted by X with a corresponding state space Ω_X . Assume that we use some technique to extract *discrete features* from each of the signals of the sensors. Let the features be denoted as y_1 and y_2 with corresponding feature spaces Ω_{Y_1} and Ω_{Y_2} . Furthermore, assume that we have performed a limited number of experiments where we have placed different objects at different positions in the range of the camera and microphone, and observed the extracted features. The goal then is to design an *agent*² \mathcal{A} that uses this limited set of information in order to construct evidence based on the observed features y_1 and y_2 and combine these pieces of evidence for the purpose of predicting the true object. In the remainder of this section, we present three agents based on the combination operators described in Section 2.

3.2 The Bayesian Agent – $\mathcal{A}_{\mathcal{B}}$

We here describe how an agent based on the Bayesian combination operator in Def. 1 can be used in order to decide on an object $x \in \Omega_X$ based on features from the sensor readings and previous mentioned limited statistical information. Since the features are extracted from different types of signals, it is fair to assume that y_1 and y_2 are conditionally independent given object X . By using a uniform (Bayes-Laplace) *Dirichlet model* (used in many scenarios, e.g., [7]), which amounts to calculating the expected value of a posterior Dirichlet density [16, 6], we can construct non-normalized evidence by [18]:

$$p(y_i|X) \triangleq \frac{\alpha_{y_i|X} + 1}{\sum_{y_i \in \Omega_{Y_i}} \alpha_{y_i|X} + |\Omega_{Y_i}|}, \quad (17)$$

where $i \in \{1, 2\}$ and where $\alpha_{y_i|X}$ denotes the number of times a specific feature y_i has been extracted given

²The use of an agent paradigm for describing the empirical evaluation was inspired by Aughenbaugh and Paredis [3].

an object X . The evidence can then be normalized:

$$\hat{p}(y_i|X) = \frac{p(y_i|X)}{\sum_{x \in \Omega_X} p(y_i|x)}, \quad (18)$$

and used as operands in the Bayesian combination operator in order to obtain a joint evidence $\hat{p}(y_1, y_2|X)$, i.e.:

$$\hat{p}(y_1, y_2|X) = \Phi_B(\hat{p}(y_1|X), \hat{p}(y_2|X)) . \quad (19)$$

Note that when we do not have any statistical information at all, $\hat{p}(y_1|X)$, $\hat{p}(y_2|X)$, and consequently $\hat{p}(y_1, y_2|X)$, are uniform. Finally, based on the joint evidence, the agent \mathcal{A}_B can define the most probable object(s) by:

$$\mathcal{A}_B \triangleq \mathcal{O}(\hat{p}(y_1, y_2|X)), \quad (20)$$

where $\mathcal{O}(\cdot)$ is defined as:

$$\mathcal{O}(p(X)) \triangleq \left\{ x \in \Omega_X : \right. \\ \left. (\forall x' \in \Omega_X) \left(p(x) \geq p(x') \right) \right\} . \quad (21)$$

where $p(X)$ is a probability function (remember that $\hat{p}(y_1, y_2|X)$ is a normalized likelihood function, i.e., a probability function).

3.3 The Credal Agent – \mathcal{A}_C

Consider the same setting but where one models the evidence by credal sets. Instead of utilizing the (precise) Dirichlet model, which was the case for the Bayesian agent, we utilize the corresponding imprecise model, denoted as the *imprecise Dirichlet model* [31, 32], where one calculates the expected value of a set of posterior Dirichlet densities. The difference between this model and the former is that one uses the imprecision, i.e., the “size” of a credal set, as a way of reflecting the amount of information that the evidence is based on. By utilizing the imprecise Dirichlet model, we can construct normalized evidence $\hat{\mathcal{P}}(y_i|X)$ by [18]:

$$\hat{\mathcal{P}}(y_i|X) \triangleq \left\{ \frac{p(y_i|X)}{\sum_{x \in \Omega_X} p(y_i|x)} : \right. \\ \left. (\forall x \in \Omega_X) \left(\frac{\alpha_{y_i|x}}{\sum_{y_i \in \Omega_{Y_i}} \alpha_{y_i|x} + \beta} \leq p(y_i|x) \right) \right. \\ \left. \leq \frac{\alpha_{y_i|x} + \beta}{\sum_{y_i \in \Omega_{Y_i}} \alpha_{y_i|x} + \beta} \right\}, \quad (22)$$

where $i \in \{1, 2\}$ and the parameter β determines how the imprecision of the set $\hat{\mathcal{P}}(y_i|X)$ is affected by the sample size. Note that when the sample size increases, the imprecision $\mathcal{I}(\hat{\mathcal{P}}(y_i|X))$ decreases since the lower and upper bounds for each $p(y_i|x)$ in Eq. (22) converge [31, 32]:

$$\lim_{\left(\sum_{y_i \in \Omega_{Y_i}} \alpha_{y_i|x} \right) \rightarrow \infty} \left(\frac{\alpha_{y_i|x} + \beta}{\sum_{y_i \in \Omega_{Y_i}} \alpha_{y_i|x} + \beta} - \frac{\alpha_{y_i|x}}{\sum_{y_i \in \Omega_{Y_i}} \alpha_{y_i|x} + \beta} \right) = 0, \quad (23)$$

i.e., imprecision is reflected by the sample size.

We can now utilize the credal combination operator in Def. 2 in order to obtain the joint evidence:

$$\hat{\mathcal{P}}(y_1, y_2|X) = \Phi_C(\hat{\mathcal{P}}(y_1|X), \hat{\mathcal{P}}(y_2|X)) . \quad (24)$$

Based on the joint evidence, agent \mathcal{A}_C can decide on the most probable object(s) in a similar way as in the Bayesian case:

$$\mathcal{A}_C \triangleq \bigcup_{\hat{p}(y_1, y_2|X) \in \hat{\mathcal{P}}(y_1, y_2|X)} \mathcal{O}(\hat{p}(y_1, y_2|X)), \quad (25)$$

where $\mathcal{O}(\cdot)$ is defined by Eq. (21). The intuition behind this set is that the agents includes all objects that are optimal for some probability function, i.e., there exists a probability function within the credal set that contains a probability that is highest for a given object within the set. In contrast to the Bayesian case, the above set is more likely to be non-singleton. This indicates that the agent does not possess enough information to distinguish between the objects within the set.

3.4 The Dempster-Shafer Agent – \mathcal{A}_D

In order to define an agent based on Dempster-Shafer theory, we first need to elaborate on how mass functions could be constructed based on the credal set obtained from the imprecise Dirichlet model in Eq. (22). A credal set is a more general structure in comparison to a mass function [33, 2]. Hence, transforming a credal set to a mass function, e.g. by [26, Theorem 2.2]:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B), \quad (26)$$

cannot in general be performed without some approximation [1, 2, 8]. One way, demonstrated by Campos et al. [8], is to approximate the credal set by certain types of intervals and then apply the algorithm proposed by Lemmer and Kyburg [22]. Another way,

suggested by Arnborg [2], is to utilize Eq. (26) and transfer mass upwards in the “set-lattice” to eliminate negative mass. Common to both these ways is that there is no unique minimal approximation that can be used for constructing the mass function. As a starting point for our evaluation, we here first consider a simple method for constructing a mass function based on the lower bounds of each single state $x \in \Omega_X$ of the credal set. Lower bounds on single states are often used in many models, e.g., [29], since they yield *simplices* which is the simplest type of imprecision.

By using lower bounds we can obtain the simplex:

$$\hat{\mathcal{P}}^*(y_i|X) \triangleq \left\{ \hat{p}^*(y_i|X) : \begin{aligned} \hat{p}^*(y_i|x) &\geq \min_{\hat{p}^*(y_i|X) \in \hat{\mathcal{P}}^*(y_i|X)} \hat{p}(y_i|x), \\ \sum_{x \in \Omega_X} \hat{p}^*(y_i|x) &= 1, \quad x \in \Omega_X \end{aligned} \right\}. \quad (27)$$

The simplex $\hat{\mathcal{P}}^*(y_i|X)$ can then easily be transformed to a mass function using Eq. (26) on the belief function/lower probabilities defined by [1, 2]:

$$Bel(A) = \min_{\hat{p}^*(y_i|A) \in \hat{\mathcal{P}}^*(y_i|A)} \hat{p}^*(y_i|A), \quad (28)$$

where $A \subseteq \Omega_X$, which results in a mass function of the following form [8]:

$$\begin{aligned} m_i(x) &= \min_{\hat{p}^*(y_i|X) \in \hat{\mathcal{P}}^*(y_i|X)} \hat{p}^*(y_i|x) \\ m_i(\Omega_X) &= 1 - \sum_{x \in \Omega_X} \min_{\hat{p}^*(y_i|X) \in \hat{\mathcal{P}}^*(y_i|X)} \hat{p}^*(y_i|x), \end{aligned} \quad (29)$$

$\forall x \in \Omega_X$. Now, based on m_1 and m_2 , obtained by Eq. (29), we can perform the combination:

$$m_{1,2}(A) = \Phi_{\mathcal{D}}(A, m_1, m_2). \quad (30)$$

In order to be able to compare the results from the credal agents $\mathcal{A}_{\mathcal{C}}$ with the above mass function, we define a Dempster-Shafer agent which includes a transformation of $m_{1,2}$ back to a credal set by performing linear programming on the following set of constraints (cf. Eq. (13)):

$$\begin{aligned} \mathcal{P}_{1,2}(X) &\triangleq \{p_{1,2}(X) : Bel_{1,2}(A) \leq p_{1,2}(A) \\ &\leq Pl_{1,2}(A), A \subseteq \Omega_X\}, \end{aligned} \quad (31)$$

where $Bel_{1,2}$ is the belief, or lower probability, in Eq. (11) and $Pl_{1,2}$ is the plausibility, or upper probability, in Eq. (12), with respect to $m_{1,2}$. Now, based on the credal set $\mathcal{P}_{1,2}(X)$, the *Dempster-Shafer agent* $\mathcal{A}_{\mathcal{D}}$ can decide on objects according to:

$$\mathcal{A}_{\mathcal{D}} \triangleq \bigcup_{p_{1,2}(X) \in \mathcal{P}_{1,2}(X)} \mathcal{O}(p_{1,2}(X)), \quad (32)$$

where $\mathcal{O}(\cdot)$ is defined in Eq. (21).

3.5 Evaluation Schema

In order to evaluate the different agents, introduced in the previous sections, we consider a combination scenario where we have two sources $i \in \{1, 2\}$ that report evidences, based on features $y_i \in \Omega_{Y_i}$ where $\Omega_{Y_i} \triangleq \{f_{i,1}, \dots, f_{i,m}\}$, regarding a random variable $X \in \Omega_X$ where $\Omega_X = \{x_1, \dots, x_m\}$ (we will instantiate the parameter m later when we describe the experiment in more detail). Note that $|\Omega_{Y_i}| = |\Omega_X| = m$. Let us now assume that the true state is x_1 and that each agent has a limited set of multinomial observations from the two sources to base evidence upon. We will simulate the limited information stemming from the sources by drawing n samples, where n is a small number that we will instantiate later, from a multinomial distribution, i.e., we sample a vector:

$$\vec{\alpha}_{y_i|x}^n \triangleq [\alpha_{f_{i,1}|x} \dots \alpha_{f_{i,m}|x}] \quad (33)$$

$\forall x \in \Omega_X$ through:

$$\vec{\alpha}_{y_i|x}^n \sim \text{Mu}(\vec{\alpha}^n | p(f_{i,1}|x), \dots, p(f_{i,m}|x)), \quad (34)$$

where $\alpha_{f_{i,j}|x}$ denotes the number of times feature $f_{i,j}$ has been observed when the object is x and where $\text{Mu}(\cdot|\cdot)$ denotes the multinomial distribution with parameters $p(f_{i,j}|x)$, $j \in \{1, \dots, m\}$, i.e., the probability of observing a specific feature $f_{i,j}$ from source i given some object x . Note that

$$\sum_{j \in \{1, \dots, m\}} \alpha_{f_{i,j}|x} = n \quad (35)$$

$\forall x \in \Omega_X$ and $i \in \{1, 2\}$. The information contained in each sampled vector can then be used in the precise and imprecise Dirichlet models, Eqs. (17) – (18) and (22), by the agents in order to construct evidence. Since the imprecise agents are undefined in cases where the sources mutually excludes each other (see further the discussion following Def. 1 – 3), we simply omit such cases.

To give an example, assume that $m = 3$ and $n = 5$ and that we have sampled the following:

$$\begin{bmatrix} \vec{\alpha}_{y_1|x_1}^5 \\ \vec{\alpha}_{y_1|x_2}^5 \\ \vec{\alpha}_{y_1|x_3}^5 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 1 \\ 1 & 0 & 4 \end{bmatrix} \quad (36)$$

$$\begin{bmatrix} \vec{\alpha}_{y_2|x_1}^5 \\ \vec{\alpha}_{y_2|x_2}^5 \\ \vec{\alpha}_{y_2|x_3}^5 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 5 & 0 \\ 0 & 1 & 4 \end{bmatrix}. \quad (37)$$

Further assume that an object $x \in \Omega_X$ have generated features $y_1 = f_{1,1}$ and $y_2 = f_{2,2}$. This would mean that we would utilize the first and second column of the matrices on the right hand side of Eqs. (36)

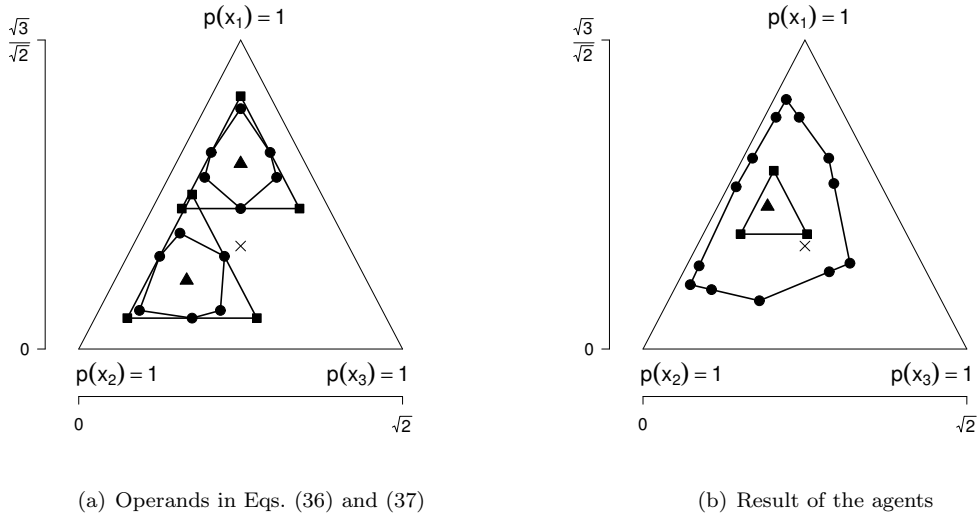


Figure 1: Precise and imprecise operands are shown in Fig. 1(a) and the result of applying agent \mathcal{A}_B , \mathcal{A}_C , and \mathcal{A}_D on the operands are shown in Fig. 1(b). The operands and results are denoted by triangles, circles, and squares for agent \mathcal{A}_B , \mathcal{A}_C , and \mathcal{A}_D , respectively.

and (37), correspondingly, for constructing the evidences based on the precise and imprecise Dirichlet models in Eqs. (17) – (18) and (22). The operands and result of applying the different agents, i.e., \mathcal{A}_B in Eq. (20), \mathcal{A}_C in Eq. (25), and \mathcal{A}_D in Eq. (32), on this data is shown in Fig. 1. Note in particular that the imprecision of the credal agent \mathcal{A}_C is considerably higher in comparison to the Dempster-Shafer agent \mathcal{A}_D . From the figure we also see that the result from the credal agent \mathcal{A}_C contains extreme points that could be removed without changing the shape of the set significantly, however, we omit such removal in this experiment. In a real-world application such a removal would be performed to reduce computational complexity.

In order to compare the performance of the agents with each other, we will use a score function that takes imprecision into account [19, 20]:

$$\Upsilon(\mathcal{A}) \triangleq \begin{cases} \frac{1}{|\mathcal{A}|} & \text{if } (x_1 \in \mathcal{A}) \wedge (\mathcal{A} \neq \Omega_X) \\ 0 & \text{otherwise} \end{cases}, \quad (38)$$

i.e., if the agent manage to minimize the imprecision and is able to return the true state, the agent obtains the highest possible reward of one. If the set \mathcal{A} contains two of the three states, where one of the states is the truth, i.e., x_1 , the agent gets half of this reward since the set is still informative due to exclusion of one erroneous state. The other two cases, i.e., the truth is not contained in the set and all the states are reported, are considered to be non-informative; the

latter due to that one already has modeled all possible states when the state space was designed.

Now, by simulating a large number of cases and apply the agents on each of these cases, we can obtain a good approximation of the expected score $E[\Upsilon(\mathcal{A})]$ of each agent. The experiment, including simulation parameters, is then defined by the following step-wise description:

1. For each source $i \in \{1, 2\}$ and $x \in \Omega_X$, draw γ according to:

$$\gamma \sim \text{Uniform}([0.7, 0.9]), \quad (39)$$

set:

$$p(f_{i,j}|x_k) \triangleq \begin{cases} \gamma & \text{when } k = j \\ \frac{1-\gamma}{m-1} & \text{otherwise} \end{cases} \quad (40)$$

and use these probabilities as multinomial parameters in Eq. (34) in order to draw vectors $\vec{\alpha}_{y_i|x}^n$. Note that given an object x_k , it is most likely that one observe the feature $f_{i,k}$ from source i .

2. Let us set $\beta = 2$ in Eq. (22) (this parameters is usually set to a value $1 \leq \beta \leq 2$, see further the discussion in [32, 5]) and sample new features to be used by the agents for predicting the true object by using the multinomial parameters in Eq. (40), i.e., sample $f_{i,j} \sim p(Y_i|x_1)$ (remember that x_1 is the true object) and apply the

agents \mathcal{A}_B in Eq. (20), \mathcal{A}_C in Eq. (25), and \mathcal{A}_D in Eq. (32), on the sampled vectors $\tilde{\alpha}_{y_i|X}^n$ from Step 1.

3. Evaluate the agents by $\Upsilon(\mathcal{A})$ in Eq. (38), and store the score and repeat from Step 1, 10^3 times.
4. Approximate the expected score $E[\Upsilon(\mathcal{A})]$ by using the stored scores from the previous step.

3.6 Results

The results are shown in Table 1, where the parameters m (dimension) and n (number of observations) have been varied. Taking the confidence interval into account, it seems that agent \mathcal{A}_B performs as well or better than agent \mathcal{A}_C and \mathcal{A}_D . One reason for this is that agent \mathcal{A}_C and \mathcal{A}_D tends to be too cautious in many cases, as can be seen from the number of cases where the complete state space is reported. The difference in performance seems to increase when the state space m increases while n is constant.

One interesting and a bit surprising effect that one can observe is that the performance of agent \mathcal{A}_D deteriorates considerably when the size of the state space increases when maintaining the same number of observations (i.e., five). The low performance is due, as can be seen from the table, that the agent tends to increase the fraction of times it reports the complete state space, e.g., when $m = 7$ and $n = 5$, it reports the complete state space in 90.7% of the cases. An explanation for such behavior is that when m increases under a constant n , the sum of the upper constraints in the imprecise Dirichlet model in Eq. (22) increases and this means that the lower bounds of $\hat{\mathcal{P}}(y_i|X)$ in Eq. (22) decreases due to the normalization. Also, since it is not likely that we have observed a feature $f_{i,j}$ where $j \neq 1$ (x_1 is the truth and according to Eqs. (39) – (40) it is most likely to observe feature $f_{i,1}$), the sum of the lower bounds is also likely to have decreased and then more mass has been allocated to the complete state space in Eq. (29). This increased mass is then distributed among the states when transforming back to a credal set, which means that the number of cases where a non-singleton set is reported has increased. In other words, when m increases under a given n , the degree of imprecision based on the mass functions increases, as is also seen from the table. Also note that when we increase the number of observations to $n = 20$ when $m = 7$ the degree of imprecision decreases again.

The performance of the credal agent \mathcal{A}_C does not seem to be equally sensitive when increasing m . In fact, the performance increases slightly for both agent \mathcal{A}_B and \mathcal{A}_C and for the latter agent the average degree of imprecision decreases. One explanation for such a

result could be that when the state space increases, the “noise probability” mass $1 - \gamma$ in Eq. (40) is distributed among more states, which could mean that it is less likely that a single erroneous state will be optimal for some probability function within the joint credal set due to noise.

It should also be noted that the credal combination operator can introduce substantial imprecision in the joint evidence, even though the operands are not that imprecise [20]. This can also be seen in Fig. 1, where both operands are less imprecise in comparison to the joint credal set. Such increase in imprecision of joint evidence mainly occur when the operands are in conflict with each other, i.e., when the operand credal sets are positioned at different positions within the simplex, especially when the operands are close to the boundary of the simplex.

We also observe that agent \mathcal{A}_B reports an erroneous set, i.e., when $\Upsilon(\mathcal{A}_B) = 0$ and $|\mathcal{A}_B| \neq \Omega_X$ in more cases compared to the other agents. This can be regarded as the usual trade-off between imprecision and precision, i.e., reducing erroneous output by increasing imprecision. It should be noted that the Bayesian agent \mathcal{A}_B has a quite crude way of reporting a decision set since the agent reports the most probable state also in cases where the difference of probability for this state in comparison to the other states is small. In a more refined Bayesian method one could use some form of thresholding (see further [20]). Nevertheless, even though such crude schema is utilized, \mathcal{A}_B performs well in comparison to the other agents.

4 Summary and Discussion

We have described an empirical experiment for evaluating and comparing the performance of different evidential combination operators. Besides comparing individual operators, our interest was also to compare precise and imprecise operators in general. The evaluation was restricted to the three operators, Bayesian combination (precise), credal combination (imprecise), and Dempster’s combination (imprecise). For each combination operator we implemented a corresponding agent. The evaluation was based on the precise and imprecise Dirichlet models and a limited number of multinomial observations. To measure the agents performance we used a score function that, based on the informativeness of the outcome, assigned a reward to the agent.

The results showed that the Bayesian agent seems to perform at least equally well as its imprecise counterparts. Since the imprecise frameworks are often motivated by their suitability to situations where only scarce information is available, i.e. the case in the ex-

Parameters	Agent	E [$\Upsilon(\cdot)$]	E [$\mathcal{I}(\cdot)$]	$\Upsilon(\cdot) > 0$ (%)		$\Upsilon(\cdot) = 0$ (%)		
				$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 1$	$ \cdot = 2$	$ \cdot = m$
$m = 3, n = 5$	\mathcal{A}_B	0.82 ± 0.03	0	80.5	2.3	17.0	0.1	0.1
	\mathcal{A}_C	0.75 ± 0.02	0.43 ± 0.01	61.5	11.4	2.6	1.3	23.2
	\mathcal{A}_D	0.80 ± 0.02	0.24 ± 0.00	69.3	16.5	6.0	1.1	7.1
$m = 5, n = 5$	\mathcal{A}_B	0.87 ± 0.02	0	86.3	1.2	11.4	0.9	0.0
	\mathcal{A}_C	0.81 ± 0.02	0.38 ± 0.01	73.7	5.1	0.7	0.1	16.0
	\mathcal{A}_D	0.64 ± 0.02	0.41 ± 0.00	49.4	12.0	0.5	0.0	29.4
$m = 7, n = 5$	\mathcal{A}_B	0.90 ± 0.02	0	89.1	1.8	8.9	0.1	0.0
	\mathcal{A}_C	0.82 ± 0.02	0.36 ± 0.01	77.0	3.8	0.9	0.4	12.8
	\mathcal{A}_D	0.17 ± 0.01	0.52 ± 0.00	1.1	2.8	0.0	0.0	90.7
$m = 7, n = 20$	\mathcal{A}_B	0.83 ± 0.02	0	82.0	1.3	16.5	0.1	0.0
	\mathcal{A}_C	0.79 ± 0.02	0.18 ± 0.01	66.4	21.3	1.2	2.1	7.1
	\mathcal{A}_D	0.80 ± 0.02	0.18 ± 0.00	64.9	29.3	0.6	2.9	0.2

Table 1: Results of the empirical evaluation in terms of expected scores $E[\Upsilon(\cdot)]$ and the degree of imprecision $E[\mathcal{I}(\cdot)]$, all with 95% confidence intervals. In addition, we also see how the cardinality of the reported set of each agent is distributed among some sets that contains the truth and not.

periment (limited multinomial information), this outcome appears to be unexpected.

However, it might also be the case that the particular type of imprecision considered in this experiment does not do the imprecise operators justice. In particular we need to be careful to judge the Dempster-Shafer agent based on this particular design of experiment, using the imprecise Dirichlet model and the given score function, since the agent was quite sensitive to the size of the limited statistical data in relation to the number of dimensions (which in principle could be reasonable). Also, in our experiment the mass function for the agent was derived by first approximating the credal set using a simplex and then transforming it into a mass function, which is a rather simple approximation. A more refined approximation might influence the performance of the Dempster-Shafer agent (see further the discussion in Section 3.4).

Note that the results are only valid in applications where the score function in Eq. (38) is accepted. It could still be useful in certain circumstances to report the entire state space, especially when you have the option to gather more information and thereby reduce the cardinality of the result. Depending on the application one needs to design the score function accordingly.

In our future research, we will explore different ways of obtaining the mass functions, and evaluate the imprecise operators found in this paper and also other variants, e.g., [13], with alternative forms of imprecision.

One interesting way ahead is to simulate mass functions directly, which then can be transformed into a credal set, instead of constructing them from credal sets as was done in the experiment. In that case one could, e.g., use the *pignistic transformation* [28] on the mass functions in order to obtain operands for the Bayesian combination operator.

The overall conclusion that we infer from the results is that the Bayesian framework could still be suitable in applications where only limited statistical data is available. Taking into account that the Bayesian framework has considerably lower computational complexity, this framework might even be the best choice for such type of applications.

Acknowledgements

We would like to thank the anonymous reviewers of this paper for their comments. This work was supported by the Information Fusion Research Program (University of Skövde, Sweden), in partnership with the Swedish Knowledge Foundation under grant 2010-0320 (URL: <http://www.infofusion.se>, UMIF project). R [25] and the package rcd [14] have been used for the computations in this paper.

References

- [1] Stefan Arnborg. Robust Bayesianism: Imprecise and paradoxical reasoning. In *Proceedings of the*

- 7th International Conference on Information fusion*, pages 407–414, 2004.
- [2] Stefan Arnborg. Robust Bayesianism: Relation to evidence theory. *Journal of Advances in Information Fusion*, 1(1):63–74, 2006.
- [3] Jason Matthew Aughenbaugh and Christiaan J. J. Paredis. The value of using imprecise probabilities in engineering design. *Journal of Mechanical Design*, 128:969–979, 2006.
- [4] James O. Berger. An overview of robust Bayesian analysis. *Test*, 3:5–124, 1994.
- [5] Jean-Marc Bernard. An introduction to the imprecise dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39:123 – 150, 2005.
- [6] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. John Wiley and Sons, 2000.
- [7] Henrik Boström. Estimating class probabilities in random forests. In *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*, pages 211–216, 2007.
- [8] Luis M. De Campos, Juan F. Huete, and Serafin Moral. Probability intervals: A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 02(02):167–196, 1994.
- [9] Inés Couso, Serafin Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.
- [10] Fabio G. Cozman. *Decision Making Based on Convex Sets of Probability Distributions: Quasi-Bayesian Networks and Outdoor Visual Position Estimation*. PhD thesis, The Robotics Institute, Carnegie Mellon University, 1997.
- [11] Fabio G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.
- [12] Arthur P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society*, pages 205–247, 1969.
- [13] Mihai Cristian Florea, Anne-Laure Jousselme, Éloi Bossé, and Dominic Grenier. Robust combination rules for evidence theory. *Information Fusion*, 10(2):183 – 197, 2009.
- [14] Charles J. Geyer, Glen D. Meeden, and incorporates code from cddlib written by Komei Fukuda. *reddl: reddl (Computational Geometry)*, 2012. R package version 1.1-7.
- [15] Rolf Haenni. Shedding new light on zadeh’s criticism of dempster’s rule of comb. In *7th International conference on Information Fusion (FUSION)*, pages 879–884, 2005.
- [16] David Heckerman. Bayesian networks for data mining. *Data Mining and Knowledge Discovery*, 1:79–119, 1997.
- [17] David Rios Insua and Fabrizio Ruggeri, editors. *Robust Bayesian Analysis*. Springer, 2000.
- [18] Alexander Karlsson. *Evaluating Credal Set Theory as a Belief Framework in High-Level Information Fusion for Automated Decision-Making*. PhD thesis, Örebro University, School of Science and Technology, 2010.
- [19] Alexander Karlsson, Ronnie Johansson, and Sten F. Andler. An empirical comparison of Bayesian and credal set theory for discrete state estimation. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), Communications in Computer and Information Science, Volume 80, ISSN 1865-0937*, 2010.
- [20] Alexander Karlsson, Ronnie Johansson, and Sten F. Andler. Characterization and empirical evaluation of bayesian and credal combination operators. *Journal of Advances in Information Fusion*, 6:150–166, 2011.
- [21] Max Krüger and Nane Kratzke. Monitoring of reliability in Bayesian identification. In *12th International Conference on Information Fusion*, 2009.
- [22] John F. Lemmer and Henry E. Kyburg. Conditions for the existence of belief functions corresponding to intervals of belief. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, 1991.
- [23] Isaac Levi. *The enterprise of knowledge*. The MIT press, 1983.
- [24] Martin E. Liggins, David L. Hall, and James Llinas, editors. *Multisensor Data Fusion, Second Edition*. CRC Press, 2009.
- [25] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
- [26] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

- [27] Philippe Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8:387–412, 2007.
- [28] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.
- [29] David Sundgren, Love Ekenberg, and Mats Danielson. Shifted dirichlet distributions as second-order probability distributions that factors into marginals. In *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 405–410, 2009.
- [30] Vicenç Torra and Yasuo Narukawa. *Modeling Decisions*. Springer, 2007.
- [31] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.
- [32] Peter Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:3–57, 1996.
- [33] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.