

An Empirical Comparison of Bayesian and Credal Combination Operators

Alexander Karlsson
Informatics Research Centre
University of Skövde
Sweden
alexander.karlsson@his.se

Ronnie Johansson
Informatics Research Centre
University of Skövde
Sweden
ronnie.johansson@his.se

Sten F. Andler
Informatics Research Centre
University of Skövde
Sweden
sten.f.andler@his.se

Abstract – We are interested in whether or not representing and maintaining imprecision is beneficial when combining evidences from multiple sources. We perform two experiments that contain different levels of risk and where we measure the performance of the Bayesian and credal combination operators by using a simple score function that measures the informativeness of a reported decision set. We show that the Bayesian combination operator performed on centroids of operand credal sets outperforms the credal combination operator when no risk is involved in the decision problem. We also show that if a risk component is present in the decision problem, a simple cautious decision policy for the Bayesian combination operator can be constructed that outperforms the corresponding credal decision policy.

Keywords: Bayesian combination operator, credal combination operator, Imprecise probability

1 Introduction

Bayesian theory [3] is one of the most commonly utilized theories for managing uncertainty in *information fusion* [9, 13]. The theory relies on two main assumptions: (1) a *probability function* should be used for representing belief and (2) *Bayes' theorem* should be used for updating belief when a new observation has been made. The main criticism of Bayesian theory that can be found in the literature (e.g., [15]) is that the first assumption is unrealistically strong since one is forced to quantify belief precisely even if one only possesses scarce information about the reality of interest. For this reason, a family of alternative theories has been introduced, which usually goes under the name *imprecise probability* [16], where belief can be expressed imprecisely.

One common theory that belongs to the family of imprecise probability is *credal set theory* (cf [1,2,6,7,11,12]), also known as “theory of credal sets” [8] and “quasi-Bayesian theory” [5], where one utilizes a *closed convex set of probability functions*, denoted as a *credal set* [12], for representing belief. In credal set theory one is also allowed to express

evidence regarding some random variable imprecisely, i.e., instead of a single *likelihood function* as a representation for the evidence, one can adopt a closed convex set of such functions. When updating is performed in credal set theory, one applies Bayes' theorem “point-wise” on all possible combinations of functions from the prior credal set (representing prior belief) and the set of likelihood functions (representing evidence) and then applies the convex-hull operator in order to enforce convexity of the posterior credal set. An attractive feature of credal set theory is that it reduces to Bayesian theory if singleton sets are adopted, hence, it can be seen as a straightforward generalization of Bayesian theory to imprecise probability.

So far, empirical evaluations that in some sense compare “decision performance” of the Bayesian and credal methods have been scarce in the literature. Indeed, if it turns out that Bayesian theory performs equally well as credal set theory, one should prefer the former since it is less computational expensive. In this paper we are interested whether or not it can be beneficial, with respect to decision making, to utilize credal set theory instead of Bayesian theory for *combining evidences from multiple sources*. More specifically, we are interested in determine whether the *credal combination operator* outperforms the *Bayesian combination operator* with respect to a *score function* that measures decision performance.

In Sect. 2 we derive the Bayesian and credal combination operators. We also present measures for *conflict* and *imprecision*, which we later use when elaborating on the design of the experiments. In Sect. 3, we present two experiments; one where no risk component is present in the decision problem and one where such a component exists. We discuss the design and analyze the result of each experiment. In Sect. 4, we summarize the paper and present the main conclusions.

2 Preliminaries

We derive the Bayesian and credal combination operators and present measures for degree of conflict and imprecision. We also elaborate on how the credal combination operator can be computed.

2.1 Bayesian Combination Operator

Let X and Y_1, \dots, Y_n be discrete random variables with state spaces Ω_X and $\Omega_{Y_1}, \dots, \Omega_{Y_n}$, respectively. Assume that we have n sources and that source $i \in \{1, \dots, n\}$ has made observation $y_i \in \Omega_{Y_i}$ and reported a *likelihood function* $p(y_i|X)$ as a representation of the *evidence* provided by y_i regarding X . By assuming that the observations are *conditional independent given X*, we can construct the *joint evidence* (or joint likelihood):

$$p(y_1, \dots, y_n|X) = p(y_1|X) \dots p(y_n|X) \quad (1)$$

In principle we can use Eq. (1) as a Bayesian way of *combining the evidences*, however, this is not convenient when implemented in an operational system since the joint evidence monotonically decreases with the number of sources n . Let us therefore elaborate on how this problem can be solved. Let:

$$p_i(X) \triangleq \frac{p(y_i|X)}{\sum_{x \in \Omega_X} p(y_i|x)}, \quad (2)$$

i.e., $p_i(X)$ are probability functions (normalized likelihood functions). By using Bayes' theorem and the assumption of conditional independence, we obtain:

$$\begin{aligned} p(X|y_1, \dots, y_n) &= \frac{p(y_1, \dots, y_n|X)p(X)}{\sum_{x \in \Omega_X} p(y_1, \dots, y_n|x)p(x)} \\ &= \frac{p(y_1|X) \dots p(y_n|X)p(X)}{\sum_{x \in \Omega_X} p(y_1|x) \dots p(y_n|x)p(x)} \\ &= \frac{p_1(X) \dots p_n(X)p(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)p(x)} \\ &= \frac{p_1(X) \dots p_n(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)} p(X) \\ &= \frac{\sum_{x \in \Omega_X} \frac{p_1(x) \dots p_n(x)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)} p(x)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)} \end{aligned} \quad (3)$$

Let us introduce the notation:

$$\Phi(p_1(X), \dots, p_n(X)) \triangleq \frac{p_1(X) \dots p_n(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)}, \quad (4)$$

From Eq. 3 we see that the joint evidence $p(y_1, \dots, y_n|X)$ has the same effect on the *posterior* $p(X|y_1, \dots, y_n)$, irrespective of the prior $p(X)$, as $\Phi(p_1(X), \dots, p_n(X))$, i.e., $p(y_1, \dots, y_n|X)$ and $\Phi(p_1(X), \dots, p_n(X))$ are equivalent evidences. Now it can easily be shown that:

$$\Phi(p_1(X), \dots, p_n(X)) = \Phi(\Phi(\dots \Phi(p_1(X), p_2(X)) \dots, p_{n-1}(X)), p_n(X)) \quad (5)$$

Hence, we can *recursively combine* the evidences in order to obtain the joint evidence. Note that the normalization in each combination in the recursion eliminates the problem of a monotonically decreasing joint evidence when n increases. We use the recursive form of Eq. (5) as our basis for the definition of a *Bayesian combination operator* denoted by Φ_B (i.e., we define the operator for two operands) [1, 2, 11]:

Definition 1. *The Bayesian combination operator is defined as:*

$$\Phi_B(p_1(X), p_2(X)) \triangleq \frac{p_1(X)p_2(X)}{\sum_{x \in \Omega_X} p_1(x)p_2(x)} \quad (6)$$

where $p_i(X)$, $i \in \{1, 2\}$, are conditionally independent evidences in the form of probability functions (normalized likelihood functions). The operator is undefined when $\sum_{x \in \Omega_X} p_1(x)p_2(x) = 0$.

Note that the operator is both associative and commutative. One important concept when combining evidences from multiple sources is *the degree of conflict* between the sources, measured on the evidences reported by the sources. Intuitively, such measure can be thought of as an "inverse similarity measure", i.e., the more similar the reported evidences are, the less conflict exists between the sources. We here simply use the *Euclidean distance* for our conflict measure. Before we define the measure, we need to define the meaning of *probability simplex*:

Definition 2. *The probability simplex $\mathcal{P}^*(X)$ for a discrete random variable X with state space Ω_X is defined as:*

$$\mathcal{P}^*(X) \triangleq \left\{ p(X) : (\forall x \in \Omega_X)(p(x) \in [0, 1]), \sum_{x \in \Omega_X} p(x) = 1 \right\}, \quad (7)$$

Let us now define a measure for degree of conflict by [11]:

Definition 3. *The degree of conflict between two evidences, in the form of probability functions $p_1(X)$ and $p_2(X)$ is defined as:*

$$\Gamma_B(p_1(X), p_2(X)) \triangleq \frac{\|p_1(X) - p_2(X)\|}{\sqrt{2}}, \quad (8)$$

where $\|\cdot\|$ is the Euclidean norm and where the denominator constitutes the diameter of the probability simplex, i.e.:

$$\max_{p_i(X) \in \mathcal{P}^*(X)} \left\{ \max_{p_j(X) \in \mathcal{P}^*(X)} \|p_i(X) - p_j(X)\| \right\} = \sqrt{2} \quad (9)$$

2.2 Credal Combination Operator

The credal combination operator can be derived by using *credal set theory* (cf [1, 2, 6, 7, 11, 12]) and a notion of independence known as *strong independence* [4]. Let $\mathcal{P}(X)$ denote a *prior credal set*, i.e., a closed convex set of probability

functions of the form $p(X)$; $\mathcal{P}(X|y)$ a *posterior credal set* of functions $p(X|y)$; and $\mathcal{P}(x|Y)$ denote a *closed convex set of likelihood functions* $p(x|Y)$. Let $\mathcal{E}(\mathcal{P}(X))$ denote the set of *extreme points* of $\mathcal{P}(X)$, i.e., points that belong to the set and cannot be expressed as a *convex combination* of other points in the set. We can now define strong independence as follows [4]:

Definition 4. *The discrete random variables X and Y are strongly independent iff all $p(X, Y) \in \mathcal{E}(\mathcal{P}(X, Y))$ can be expressed as $p(X, Y) = p(X)p(Y)$, where $p(X) \in \mathcal{P}(X)$ and $p(Y) \in \mathcal{P}(Y)$. Similarly, X and Y are strongly conditionally independent given Z iff all $p(X, Y|z) \in \mathcal{E}(\mathcal{P}(X, Y|z))$ can be expressed as $p(X, Y|z) = p(X|z)p(Y|z)$, $\forall z \in \Omega_Z$, where $p(X|z) \in \mathcal{P}(X|z)$ and $p(Y|z) \in \mathcal{P}(Y|z)$.*

By using this notion of independence, the *credal combination operator*, also known as the “the robust Bayesian combination operator” [1, 2, 11], can be derived as a straightforward generalization of the Bayesian combination operator:

Definition 5. *The credal combination operator is defined as:*

$$\Phi_C(\mathcal{P}_1(X), \mathcal{P}_2(X)) \triangleq \mathcal{CH}\left(\left\{\Phi_B(p_1(X), p_2(X)) : \begin{array}{l} p_1(X) \in \mathcal{P}_1(X), \\ p_2(X) \in \mathcal{P}_2(X) \end{array}\right\}\right), \quad (10)$$

where $\mathcal{P}_i(X)$, $i \in \{1, 2\}$, are strongly conditionally independent evidences in the form of credal sets (closed convex sets of normalized likelihood functions) and where \mathcal{CH} is the convex-hull operator. The Φ_C operator is undefined iff there exists $p_i(X) \in \mathcal{P}_i(X)$, $i \in \{1, 2\}$, such that Φ_B is undefined.

The operator is associative and commutative. Note that the operator reduces to the Bayesian combination operator for singleton sets. In this paper we only consider credal sets in the form of *polytopes*:

Definition 6. *A credal set $\mathcal{P}(X)$ is a polytope iff:*

$$\mathcal{P}(X) \triangleq \mathcal{CH}\left(\{p_1(X), \dots, p_n(X)\}\right), \quad (11)$$

where $\{p_1(X), \dots, p_n(X)\} \subset \mathcal{P}^*(X)$ is a finite set and where \mathcal{CH} is the convex-hull operator.

The following theorem provides a convenient way of computing the credal combination operator when the credal sets are polytopes (cf [1, 2, 11]):

Theorem 1.

$$\Phi_C(\mathcal{P}_1(X), \mathcal{P}_2(X)) \Leftrightarrow \Phi_C(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))) \quad (12)$$

Proof. See Karlsson et al (2009), Theorem 1 [11] \square

One important feature of a credal set is its *imprecision*. We use the following measure of *degree of imprecision* [11]:

Definition 7. *The degree of imprecision of a credal set $\mathcal{P}(X)$ is defined as:*

$$\mathcal{I}(\mathcal{P}(X)) \triangleq \frac{1}{|\Omega_X|} \sum_{x \in \Omega_X} \Delta(x, \mathcal{P}(X)) \quad (13)$$

where:

$$\Delta(x, \mathcal{P}(X)) \triangleq \max_{p \in \mathcal{P}(X)} p(x) - \min_{p \in \mathcal{P}(X)} p(x), \quad (14)$$

is *Walley's measure for degree of imprecision for a single event* [15].

Similar to the Bayesian conflict measure (Definition 3), we base a conflict measure for credal sets on the notion of similarity. Such similarity measure for closed convex sets exists under the name of *Hausdorff distance* [10]. Let us define the following conflict measure for credal sets [11]:

Definition 8. *The degree of conflict between two credal sets $\mathcal{P}_1(X)$ and $\mathcal{P}_2(X)$ is defined as:*

$$\Gamma_C(\mathcal{P}_1(X), \mathcal{P}_2(X)) \triangleq \frac{\mathcal{H}(\mathcal{P}_1(X), \mathcal{P}_2(X))}{\sqrt{2}}, \quad (15)$$

where the denominator constitutes the diameter of the probability simplex $\mathcal{P}^*(X)$ and where \mathcal{H} is the Hausdorff distance defined by [10]:

$$\mathcal{H}(\mathcal{P}_1(X), \mathcal{P}_2(X)) \triangleq \max \left\{ \vec{\mathcal{H}}(\mathcal{P}_1(X), \mathcal{P}_2(X)), \vec{\mathcal{H}}(\mathcal{P}_2(X), \mathcal{P}_1(X)) \right\}, \quad (16)$$

where $\vec{\mathcal{H}}$ is the *forward Hausdorff distance*:

$$\vec{\mathcal{H}}(\mathcal{P}_1(X), \mathcal{P}_2(X)) \triangleq \max_{p_1(X) \in \mathcal{P}_1(X)} \left\{ \min_{p_2(X) \in \mathcal{P}_2(X)} \|p_1(X) - p_2(X)\| \right\}, \quad (17)$$

where $\|\cdot\|$ is the *Euclidean norm*.

Note that the credal conflict measure reduces to the Bayesian conflict measure for singleton sets.

3 Experiments

We present two main experiments for combining evidences reported by a number of sources, where there exists some degree of conflict between them. In the first experiment there is no risk component whereas in the second experiment a risk component is included.

3.1 Experiment A – No Risk

Let us start with a scenario that does not contain a risk component in the sense that there is no cost of reporting an erroneous state. Assume that we are interested in determining the state of a random variable X with a state space consisting of three possible states, i.e., $\Omega_X = \{x_1, x_2, x_3\}$ and that we base our decision regarding X on n sources that provide us with pieces of evidence regarding X in the form of strongly conditionally independent credal sets $\mathcal{P}_1(X), \dots, \mathcal{P}_n(X)$. Assume that the true state of X is x_2 . Obviously, if we have selected the sources well, a majority of these provide us with credal sets $\hat{\mathcal{P}}(X)$ that constitute evidence for the truth solely, i.e.:

$$\hat{p}(X) \in \hat{\mathcal{P}}(X) \Rightarrow x_2 = \arg \max_{x \in \Omega_X} \hat{p}(x) \quad (18)$$

Such type of credal set is completely contained in the region with vertical lines shown in Fig. 1. Let us assume that there

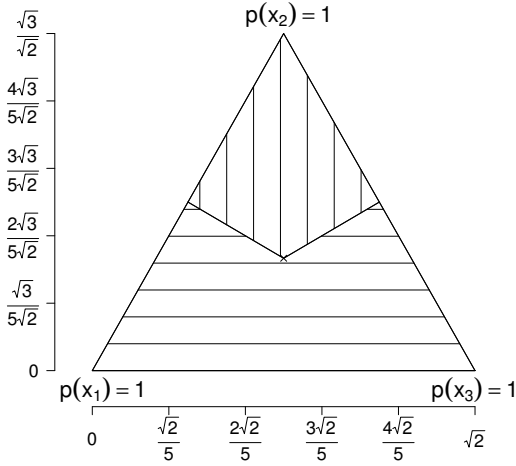


Figure 1: The probability simplex $\mathcal{P}^*(X)$ partitioned into evidence region (vertical lines) and counter-evidence region (horizontal line) with respect to the true state x_2 .

is a possibility for obtaining a counter evidence $\tilde{\mathcal{P}}(X)$ with respect to the truth from some of the sources, i.e.:

$$\tilde{p}(X) \in \tilde{\mathcal{P}}(X) \Rightarrow x_2 \neq \arg \max_{x \in \Omega_X} \tilde{p}(x) \quad (19)$$

The counter evidence $\tilde{\mathcal{P}}(X)$ is completely contained in the region with horizontal lines shown in Fig. 1. The imprecision of the credal evidence and counter evidence can be thought of as a *second-order uncertainty* regarding the strength of an evidence in the form of a probability function (i.e., a Bayesian evidence). Let us assume that the sources have no reason to favor any probability function in the credal evidence, i.e., the sources are *indifferent* regarding the probability functions.

Now, assume that we want to combine all the evidences obtained from the sources into a *joint evidence*. In the

Bayesian case, since we cannot apply the Φ_B operator to the operand credal sets, we need to select a single representative probability function from each operand to be utilized for combination. Since the sources are indifferent regarding the probability functions in the operand credal sets, we can assume an implicit uniform distribution over the sets. It is therefore reasonable to utilize the *expected value* of this distribution as a representative function, i.e., the *centroid distribution*. Consequently we obtain the following joint Bayesian evidence:

$$p_{1:n}(X) \triangleq \Phi_B(\Phi_B(\dots \Phi_B(\Upsilon(\mathcal{P}_1(X)), \Upsilon(\mathcal{P}_2(X))), \dots, \Upsilon(\mathcal{P}_{n-1}(X)), \Upsilon(\mathcal{P}_n(X))) \quad (20)$$

where the operator Υ is defined as:

$$\Upsilon(\mathcal{P}(X)) \triangleq E_{\text{Un}(\mathcal{P}(X))}[\mathcal{P}(X)], \quad (21)$$

where $\text{Un}(\mathcal{P}(X))$ denotes the uniform distribution over $\mathcal{P}(X)$ (i.e., $\Upsilon(\mathcal{P}(X))$ gives the centroid distribution of $\mathcal{P}(X)$). In the credal case, the joint evidence is straightforwardly obtained by utilizing the Φ_C operator:

$$\mathcal{P}_{1:n}(X) \triangleq \Phi_C(\Phi_C(\dots \Phi_C(\mathcal{P}_1(X), \mathcal{P}_2(X)), \dots, \mathcal{P}_{n-1}(X)), \mathcal{P}_n(X)) \quad (22)$$

Now, based on the joint Bayesian and credal evidences, we want to make a decision regarding the true state of the variable X . In the Bayesian case this is simply performed by reporting the most probable state(s):

$$\mathcal{D}_B(p_{1:n}(X)) \triangleq \{x_i \in \Omega_X : p_{1:n}(x_i) \geq p_{1:n}(x_j), x_j \in \Omega_X\} \quad (23)$$

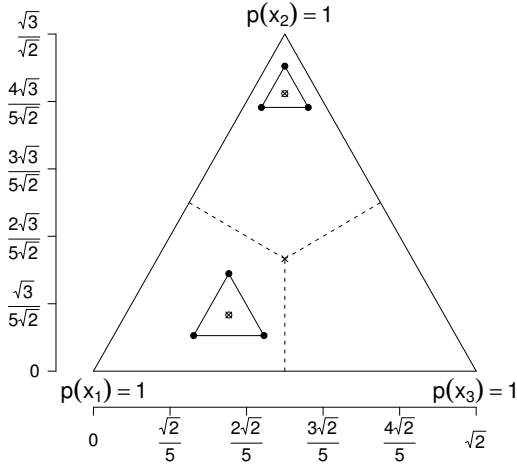
From the above equation, we see that the *Bayesian decision set* $\mathcal{D}_B(p_{1:n}(X))$ is singleton in a majority of the cases. In the credal case, however, it is quite likely, depending on the degree of imprecision reported by the sources, that the decision set is non-singleton:

$$\mathcal{D}_C(\mathcal{P}_{1:n}(X)) \triangleq \bigcup_{p_{1:n}(X) \in \mathcal{P}_{1:n}(X)} \mathcal{D}_B(p_{1:n}(X)) \quad (24)$$

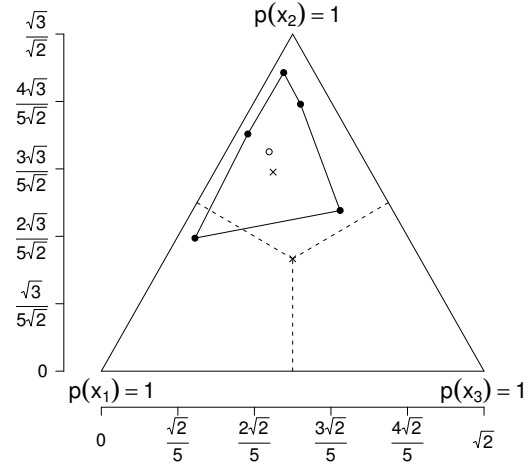
Note that unless all probability functions within $\mathcal{P}_{1:n}(X)$ agree on the most probable state, the decision set is non-singleton. Let us also introduce a credal method where the Φ_C operator is used for constructing the joint evidence but where the centroid distribution of the joint credal evidence is used for decision making in the same way as in the Bayesian case Eq. (23), i.e.:

$$\mathcal{D}_C^c(\mathcal{P}_{1:n}(X)) \triangleq \mathcal{D}_B(\Upsilon(\mathcal{P}_{1:n}(X))), \quad (25)$$

An example of using the Φ_B and Φ_C operators is seen in Fig. 2. In Fig. 2(a), we see that one of the sources has reported a quite strong evidence for the truth x_2 while the other source has reported a counter evidence to this state



(a) $\mathcal{P}_i(X)$ and $E_{\text{Un}(\mathcal{P}_i(X))}[\mathcal{P}_i(X)]$, $i \in \{1, 2\}$



(b) $\mathcal{P}_{1:2}(X)$ and $p_{1:2}(X)$

Figure 2: The example shows the probability simplex $\mathcal{P}^*(X)$ where one operand constitutes evidence for the true state x_2 and the other counter evidence to the true state (in this case evidence for x_1). The dashed lines show the decision regions for the state space. The extreme points of the credal operands $\mathcal{P}_i(X)$, $i \in \{1, 2\}$ and joint credal evidence $\mathcal{P}_{1:2}(X)$ are depicted by filled circles. The centroid of $\mathcal{P}^*(X)$ and $\mathcal{P}_{1:2}(X)$ is depicted by a cross. The joint Bayesian evidence $p_{1:2}(X)$ is depicted by an unfilled circle.

(an evidence for x_1). Figure 2(b) shows the results of the Bayesian and credal methods. We see that:

$$\begin{aligned} \mathcal{D}_{\mathcal{B}}(p_{1:2}(X)) &= \{x_2\} \\ \mathcal{D}_{\mathcal{C}}(\mathcal{P}_{1:2}(X)) &= \{x_1, x_2\} \\ \mathcal{D}_{\mathcal{C}}^c(\mathcal{P}_{1:2}(X)) &= \{x_2\} \end{aligned} \quad (26)$$

Note that the centroid of the joint credal evidence differs from the joint Bayesian evidence.

Now, obviously a decision set $\mathcal{D} \subseteq \Omega_X$ that contains two states where one of them is the true state, i.e., x_2 , should be less valued in comparison to a decision set that is singleton with the true state. Moreover, a decision set that is equal to the state space is clearly non-informative about X since we have already modeled the set of possibilities for X by Ω_X . Hence such decision set is not regarded to be of any value. Based on this reasoning we adopt the following *score function* for our experiment:

$$f_{\alpha}(\mathcal{D}) \triangleq \begin{cases} \frac{1}{|\mathcal{D}|}, & \text{if } x_2 \in \mathcal{D}, \mathcal{D} \neq \Omega_X \\ 0, & \text{if } \mathcal{D} = \Omega_X \\ -\alpha, & \text{Otherwise} \end{cases}, \quad (27)$$

where $\mathcal{D} \subseteq \Omega_X$. Note that α models a risk component. As we stated in the beginning of this section, we first explore the performance of the methods when no risk is involved in the decision problem, hence, we instantiate the score function with $\alpha = 0$, i.e., $f_0(\mathcal{D})$.

Let the probability for the event that a source reports an evidence with respect to the truth (i.e., x_2) be denoted by

β . Note that if we sum the degree of conflict for both the Bayesian and credal conflict measures for all $n - 1$ combinations, i.e.:

$$\Gamma_{\bullet}^{1:n-1} \triangleq \sum_{i=1}^{n-1} \Gamma_{\bullet}^i \quad (28)$$

where Γ_{\bullet}^i denotes the conflict in the i^{th} combination, then we expected $\Gamma_{\bullet}^{1:n-1}$ to increase when β monotonically decreases in the interval $[0.5, 1]$, i.e., the total amount of conflict among the sources decreases. The experiment can now be defined by the following step-wise description:

1. Sample the number of sources $n \sim \text{Un}([5, 10])$
2. Sample the probability for obtaining an evidence for the true state $\beta \sim \text{Un}([0.7, 0.9])$
3. Sample evidences $\mathcal{P}_1(X), \dots, \mathcal{P}_n(X)$ where the probability of sampling an evidence $\tilde{\mathcal{P}}_i$ for the truth is β (see Eq. (18)). and $1 - \beta$ for a counter evidence $\tilde{\mathcal{P}}_i$ (see Eq. (19)), $i \in \{1, \dots, n\}$.
4. Calculate joint evidences $p_{1:n}(X)$ and $\mathcal{P}_{1:n}(X)$.
5. Calculate decision sets $\mathcal{D}_{\mathcal{B}}(p_{1:n}(X))$, $\mathcal{D}_{\mathcal{C}}(\mathcal{P}_{1:n}(X))$, and $\mathcal{D}_{\mathcal{C}}^c(\mathcal{P}_{1:n}(X))$
6. Calculate the score $f_{\alpha}(\cdot)$ for each decision set in the previous step.
7. Repeat $m = 10^5$ times

Table 1: Expected score $E[f_0(\cdot)]$, with 95% confidence intervals, for $\mathcal{D}_B(p_{1:n}(X))$, $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$ and $\mathcal{D}_C^c(\mathcal{P}_{1:n}(X))$.

Method	$E[f_0(\cdot)]$	$f_0(\cdot) > 0$ (%)		$f_0(\cdot) = 0$ (%)		
		$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 3$
$\mathcal{D}_B(p_{1:n}(X))$	0.93 ± 0.002	92.8	0.0	7.2	0.0	0.0
$\mathcal{D}_C(\mathcal{P}_{1:n}(X))$	0.85 ± 0.002	78.1	13.1	1.8	0.4	6.6
$\mathcal{D}_C^c(\mathcal{P}_{1:n}(X))$	0.92 ± 0.002	91.6	0.0	8.4	0.0	0.0

Remember that we have instantiated $\alpha = 0$ for this first experiment, i.e., there is no risk component involved. Let us elaborate somewhat on the implementation detail of the above description. In step three, we sample evidences by first deciding, utilizing β , if a specific source should report an evidence or a counter evidence for the truth. Then, when we know if it is an evidence or counter evidence that we should sample, we sample a centroid from the corresponding region (see Fig. 1), uniformly. Given the centroid, we sample imprecision by considering the distance from the centroid to the corner points of an equilateral triangle, under the condition that all corner points should reside in the same evidence region. Hence, the credal operand that we sample are all equilateral triangles (simplices) that are completely contained in the evidence or counter-evidence region with respect to the truth. Credal sets of this form can be obtained by interval constraints on marginal probabilities.

3.1.1 Results

The results of the experiment is seen in Table 1. We see that the expected score of the Bayesian method \mathcal{D}_B is clearly better than the credal method \mathcal{D}_C . This means that the credal method does not isolate the cases for which the Bayesian method performs poorly in an optimal way, since we would then have expected a higher score for the credal method. This is seen from the table since in 21.9% of the cases the credal method outputs a non-singleton set while the Bayesian method only outputs an erroneous state in 7.2% of the cases. The credal method outputs a decision set of no value (i.e., x_2 is not in the decision set or Ω_X is reported) in 8.8% of the cases. In fact, even if we would let the credal method obtain a reward of one in cases where two states is reported and one of them is the truth, the credal method would still perform worse than the Bayesian method ($78.1\% + 13.1\% = 91.2\%$ compared to 92.8%). Also note that the Bayesian method \mathcal{D}_B performs better than the credal centroid method \mathcal{D}_C^c , however, the difference is not as high compared to the former case.

3.2 Experiment B – Risk

One argument that one might have for using the credal method \mathcal{D}_C is that even though it cannot optimally isolate the cases where the Bayesian method \mathcal{D}_B performs poorly, it can still be an interesting choice when there exists a risk component in the decision problem, i.e., reporting an erroneous state is coupled with a large negative cost. Indeed, if we use the result from Table 1, we see that the Bayesian

method reports an erroneous state in 7.2% of the cases while the credal method only makes erroneous reports in $1.8\% + 0.4\% = 2.2\%$ of the cases. Hence, if we would have set $\alpha = 10$ in the score function in Eq. (27), we would have obtained an expected score $E[f_{10}(\mathcal{D}_B(p_{1:n}(X)))] \approx 0.21$ for the Bayesian method and $E[f_{10}(\mathcal{D}_C(\mathcal{P}_{1:n}(X)))] \approx 0.76$ for the credal correspondence. However, when a risk is incorporated in the decision problem, there clearly exist cases when using the Bayesian method for which one would not simply output the single state that maximizes the probability, e.g., whenever the joint Bayesian evidence $p_{1:n}(X)$ is close to the uniform distribution. Let us therefore modify \mathcal{D}_B to a *cautious Bayesian method* \mathcal{D}_B^δ in the following way:

$$\mathcal{D}_B^\delta(p_{1:n}(X)) \triangleq \{x \in \Omega_X : p_{1:n}(x) > \delta\}, \quad (29)$$

where $\delta \in [0, |\Omega_X|^{-1}]$. The method partitions the probability simplex into decision regions, seen in Fig. 3. Note that a high value of δ yields a less cautious method and the other way around. Note that when $\delta = 0$ we get $\mathcal{D}_B^0(p_{1:n}(X)) = \Omega_X$ for all joint evidences $p_{1:n}(X)$. Also note that when $\delta = |\Omega_X|^{-1}$ we still have decision regions that are non-singleton.

Now let us use the same simulation settings as in Experiment A (Sect. 3.1) but where we now introduce a risk component by setting $\alpha = 10$, yielding a score function f_{10} . We perform the simulation for at set of values of the parameter $\delta \in [0, |\Omega_X|^{-1}]$ to see if there exist parameters that causes the cautious Bayesian method to outperform the credal method.

3.2.1 Results

The result is shown in Fig. 4. The cautious Bayesian method outperforms the credal method when $\delta \in [0.005, 0.07]$. Let us explore the cautious Bayesian method at its peak performance, which approximately occur at $\delta = 0.02$. The result of this parameter value is seen in Table 2. We see that the cautious Bayesian method tends to output a non-singleton set more often than the credal correspondence. However, the Bayesian method only reports an erroneous state in $0.7\% + 0.5\% = 1.2\%$ of the cases compared to $1.8\% + 0.4\% = 2.2\%$ in the credal case and due to the high risk component this yields a better score for the Bayesian method. Note that since δ is quite low and the cautious Bayesian method only outputs Ω_X in approximately 7.1% of the cases, we can conclude that the joint Bayesian evidence $p_{1:n}(X)$ is close to the boundary of the probability

Table 2: Expected score $E[f_{10}(\cdot)]$, with 95% confidence intervals, for the methods $\mathcal{D}_B^{0.02}(p_{1:n}(X))$ and $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$.

Method	$E[f_{10}(\cdot)]$	$f_{10}(\cdot) \geq 0$ (%)			$f_{10}(\cdot) < 0$ (%)	
		$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 3$	$ \cdot = 1$	$ \cdot = 2$
$\mathcal{D}_B^{0.02}(p_{1:n}(X))$	0.69 ± 0.006	69.6	22.1	7.1	0.7	0.5
$\mathcal{D}_C(\mathcal{P}_{1:n}(X))$	0.62 ± 0.010	77.9	13.2	6.7	1.8	0.4

Table 3: The table shows the cautious Bayesian parameter δ in $\mathcal{D}_B^\delta(p_{1:n}(X))$ for different risks α . The intervals for δ depicts the region for which $\mathcal{D}_B^\delta(p_{1:n}(X))$ outperforms $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$ with respect to $E[f_\alpha(\cdot)]$ and where the 95% confidence intervals are non-overlapping for the methods.

α	0	2	4	6	8	10
δ	[0.05, 0.33]	[0.04, 0.33]	[0.03, 0.14]	[0.02, 0.10]	[0.01, 0.08]	[0.01, 0.07]

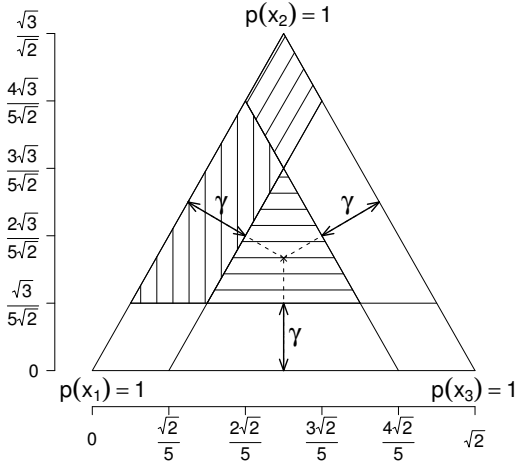


Figure 3: An example of the cautious Bayesian method in Eq. (29) where $\delta = 0.2$. The parameter δ imposes decision regions by planes that are parallel to the (proper) faces of the simplex with a distance $\gamma = \delta(\sqrt{3}/\sqrt{2})$. The horizontal lines depict the decision region for Ω_X , the vertical lines depicts $\{x_1, x_2\}$, and lastly the region with skewed lines depicts $\{x_2\}$.

simplex in a majority of the cases, which is quite natural since we have assumed that a majority ($\beta \in [0.7, 0.9]$) of the sources output evidence for the true state x_2 . Let us further study the sensitivity of the parameter δ by exploring how a parameter set for δ changes with respect to the risk for the case where the cautious Bayesian method outperforms the credal method (where the 95% confidence intervals are non-overlapping), seen in Table 3. From the table we see that when there is a low risk, $\alpha \in \{0, 2\}$, the parameter sets for δ where \mathcal{D}_B^δ outperforms \mathcal{D}_C are quite large. When the risk increases, $\alpha \in \{4, 6, 8, 10\}$, the parameter sets becomes considerably smaller. From the table we see that \mathcal{D}_B^δ with

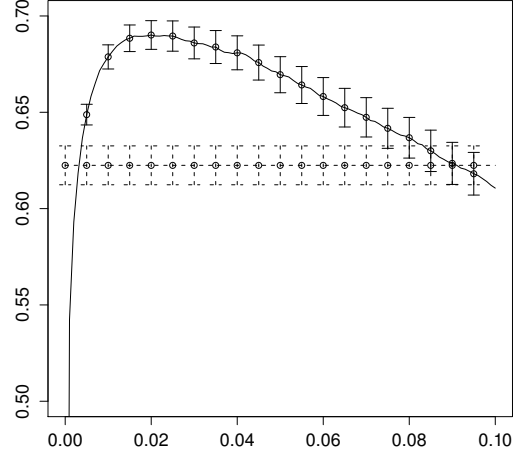


Figure 4: The solid line shows the cautious Bayesian method $\mathcal{D}_B^\delta(p_{1:n}(X))$ and the dashed line the credal method $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$. The x-axis depicts δ and the y-axis $E[f_{10}(\cdot)]$. Confidence intervals on the 95%-level are also shown.

$\delta = 0.05$ performs better than \mathcal{D}_C irrespective of the risk α .

4 Summary and Conclusions

We have presented two experiments that evaluate the Bayesian and credal combination operators for combining evidences, regarding a discrete state space, from multiple sources. In both experiments the sources report credal sets where the sources' second-order beliefs over the sets are uniformly distributed. For the Bayesian combination operator, we have utilized the expected value of the reported credal set, i.e., the centroid, for obtaining the joint evidence. We have evaluated the operators by using a simple score function that gives a reward corresponding to the informativeness of the joint evidence and a loss according to a specified risk. In the first experiment, we showed that in scenarios

where there exists no risk component, it is clearly beneficial to utilize the Bayesian combination operator instead of the credal correspondence. This is true even if one would maintain imprecision by using the credal combination operator and lastly utilize the centroid for decision making. However, this difference in performance was not as clear as in the previous case. Nevertheless, the latter results show that nothing is gained by maintaining imprecision and then using the centroid for constructing the decision set. By using the result from this experiment, we concluded that if a large risk component is present, the credal method is preferred due to a lower number of erroneous decision sets. However, we introduced a simple cautious Bayesian method, using a single parameter that partitions the probability simplex into regions corresponding to different decision sets, and we showed that such method can outperform the credal correspondence. One potential problem with the cautious Bayesian method is that one need to choose an appropriate parameter value. However, we showed that the parameter is not in particular sensitive since there existed values for which the method outperformed the credal method for a set of risk components.

In essence our results tells us that if there is no risk component in the scenario of interest, then one should use the Bayesian combination operator, even if the sources choose to report imprecision via credal sets. Furthermore, if a risk component does exists in the scenario, then one should use the cautious Bayesian method that we introduced. Hence for both cases it is sufficient to utilize a single probability function and the Bayesian combination operator for representing respectively combining evidences. From the perspective of computational complexity this is indeed positive results, considering that the number of extreme points of the joint credal evidence in worst case can grow exponentially with the number of combinations.

The question then is if there exists any cases at all where one might want to maintain imprecision with the credal combination operator? One possible such scenario could be when there is a human decision maker involved in the scenario, in particular when there exists a risk component. In such cases the decision maker might want to use the credal combination operator in order to maintain imprecision for the purpose of keeping track of worst-case scenarios with respect to the risk.

Acknowledgements

Computations have been performed with *R* [14]. This work was supported by the Information Fusion Research Program (University of Skövde, Sweden) in partnership with the Swedish Knowledge Foundation under grant 2003/0104 (URL: <http://www.infofusion.se>).

References

[1] S. Arnborg. Robust Bayesianism: Imprecise and paradoxical reasoning. In *Proceedings of the 7th International Conference on Information fusion*, 2004.

- [2] S. Arnborg. Robust Bayesianism: Relation to evidence theory. *Journal of Advances in Information Fusion*, 1(1):63–74, 2006.
- [3] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley and Sons, 2000.
- [4] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.
- [5] F. 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.
- [6] F. Cozman. A derivation of quasi-Bayesian theory. Technical report, Robotics Institute, Carnegie Mellon University, 1997.
- [7] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.
- [8] F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184, 2005.
- [9] S. Das. *High-Level Data Fusion*. Artech House, 2008.
- [10] A. Iripino and V. Tontodonato. Cluster reduced interval data using Hausdorff distance. *Computational Statistics*, 21:241–288, 2006.
- [11] A. Karlsson, R. Johansson, and S. F. Andler. On the behavior of the robust Bayesian combination operator and the significance of discounting. In *6th International Symposium on Imprecise Probability: Theories and Applications*, 2009.
- [12] I. Levi. *The enterprise of knowledge*. The MIT press, 1983.
- [13] M. E. Liggins, D. L. Hall, and J. Llinas, editors. *Multisensor Data Fusion*. CRC Press, 2009.
- [14] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.
- [15] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.
- [16] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.