Second-Order Credal Combination of Evidence

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

David Sundgren
Department of Computer and Systems Sciences
Stockholm University, Sweden
dsn@dsv.su.se

Abstract

We utilize second-order probability distributions for modeling second-order information over imprecise evidence in the form of credal sets. We generalize the Dirichlet distribution to a shifted version, denoted the S-Dirichlet, which allows one to restrict the support of the distribution by lower bounds. Based on the S-Dirichlet distribution, we present a simple combination schema denoted as second-order credal combination (SOCC), which takes second-order probability into account. The combination schema is based on a set of particles, sampled from the operands, and a set of weights that are obtained through the S-Dirichlet distribution. We show by examples that the second-order probability distribution over the imprecise joint evidence can be remarkably concentrated and hence that the credal combination operator can significantly overestimate the imprecision.

Keywords. Second-order credal combination, imprecise probability, credal sets, second-order probability, combination, evidence

1 Introduction

One common way of representing belief imprecisely is by so called *credal sets* [19], i.e., a *closed and convex set of (discrete) probability distributions*. If one utilizes this belief structure in a *Bayesian context*, where a *prior* distribution is updated to a *posterior* by a *likelihood function*, one ends up with what is referred to as *credal set theory* [8, 18], similar to *robust Bayesian theory* [1, 2, 4, 14] but without the sensitivity interpretation. Credal set theory can be thought of as a straightforward generalization of Bayesian theory to imprecise probability since Bayes' theorem is applied point-wise on all elements in a prior credal set and a set of likelihood functions. In fact, Bayesian theory is the special case of credal set theory when all sets are singletons.

Karlsson et al. [18] have previously shown that credal set theory can yield posterior credal sets that are highly imprecise and that this extra imprecision can have a deteriorating effect on decision-making, even though there exists inherent imprecision in the decision situation. The main focus in this paper is to take one step further than only modeling imprecision by adding information in the form of *second-order probability* [9, 20, 28], thus qualifying the imprecision. This type of research was briefly initiated by Karlsson et al. [17] where the consequences of modeling second-order probability was explored by using the uniform distribution in the simple case where only two states were present in the state space. In that case, it was found that the second-order probability over posterior imprecision can be considerably skewed.

In order to model different second-order probability distributions, for any number of states, we will here generalize the *Dirichlet distribution* to allow for non-zero lower bounds, denoted the *shifted Dirichlet distribution* (S-Dirichlet). Compared to the Dirichlet distribution, the S-Dirichlet distribution has twice as many parameters since for each variable there is the usual Dirichlet parameter but also a lower bound that allows one to restrict the support of the distribution. The unique family of second-order distributions that factorizes into marginals, presented in Sundgren et al. [27], is a special case of the S-Dirichlet distribution with fixed Dirichlet parameters. In this case, the joint second-order distribution equals the normalized product of its own marginal distributions and do not model any dependencies between first-order probability other than the necessary requirement that the probabilities sum to one.

We will utilize the S-Dirichlet distribution for exploring the consequences of modeling second-order probability over *imprecise evidence*, in the form of credal sets, which we combine into a single imprecise joint evidence, also in the form of a credal set, while preserving second-order information. We introduce a simple particle based method for performing the computation and we denote such a schema by *second-order credal combination* (SOCC). The purpose is to utilize this schema in order to explore the concentration and placement of the resulting particle cloud. Our schema is based on considering likelihoods as first-order evidence and second-order probability in the form of the S-Dirichlet distribution, however, it should be noted that there exists previous approaches where likelihoods has been considered as a possibilistic second-order (hierarchical) model [5].

The paper is organized as follows: in Section 2, we formalize and give an overview of the problem of combining independent pieces of evidence from different sources. In Section 3, we generalize the Dirichlet distribution to the shifted version. In Section 4, we present a simple method for performing SOCC. Lastly, in Section 5, we summarize the paper and provide our conclusions.

2 Preliminaries

We here present some background material that the remaining paper relies on. We start by providing a general overview of the problem of combining independent pieces of evidence. Based on this overview, we then present how the combination problem can be tackled within the framework of credal sets [19], namely by using the *credal combination operator* also known as the *robust Bayesian combination operator* [1, 2].

2.1 Combination of Evidence

Combination of independent pieces of evidence from multiple sources, e.g., sensors, is a problem that has been extensively studied within many different variants of *evidence theory*, e.g., [10, 22, 24]. Common to all these theories is that evidence are represented imprecisely by so called *mass functions* which operate on the power set of some state space. Pieces of evidence are combined utilizing a so called *combination operator*, e.g., *Dempster's rule of combination* [22]. One important aspect to consider when performing such combination is that of independence. Loosely, this means that one piece of evidence should not be informative regarding the other piece (see further [23]).

The problem of combining evidence has not been equally well studied under a Bayesian perspective or within the framework of credal sets. However, Arnborg [1, 2] has explored the relationship between *robust Bayesian theory* [4, 14], which can be considered a sensitivity interpretation of credal sets, and evidence theory. He found that the results of these theories can even be disjoint. One key observation when considering the combination problem within a Bayesian or credal framework is that it is the *likelihoods* that constitute evidence. Let us further elaborate on this in the next section.

2.2 Credal Combination of Evidence

Since the *credal combination operator* [15, 18], introduced as the *robust Bayesian combination operator* by Arnborg [1, 2] (we deliberately avoid using this terminology since it imposes a sensitivity interpretation of the imprecision), is a direct generalization of its Bayesian counterpart, we start by elaborating on how the latter can be derived. The derivation is similar to Karlsson et al. [18], and is inspired by Arnborg [1, 2]. Assume that we have a random variable X for which we are uncertain about the true value. Let the state space for X be denoted by $\Omega_X \triangleq \bigcup_{i=1}^{n}\{x_i\}$ and that we can obtain observations $y_1 \in \Omega_{Y_1}$ and $y_2 \in \Omega_{Y_2}$. We can then use Bayes' theorem in order calculate the posterior distribution, or belief, regarding the true value of X given these observations:

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

From the above equation, we see that the observations y_1 and y_2 only affect the posterior through the joint likelihood $p(y_1, y_2|X)$, which hence constitutes the evidence based on the observations. Now by assuming conditional independence, we obtain:

$$p(y_1, y_2|X) = p(y_1|X)p(y_2|X), \tag{2}$$

i.e., one observation is not informative about the other given that we know the true state of X. The above equation is essentially all that we need in order to combine two pieces of evidence in the form of likelihood functions into a single joint evidence. However, in order to avoid a monotonically decreasing joint evidence, it is convenient to normalize the joint evidence to a probability function. By also normalizing the likelihoods, we have constructed an operator where both the operands and result are evidences in the form of probability functions. Note that these normalizations do not affect the resulting posterior distribution since it is only the relative strengths of likelihoods that determines the posterior (see further Karlsson et al. [18] for more detail). Based on this line of reasoning, we are now ready to define *the Bayesian combination operator* [1, 2, 15, 18]:

Definition 1. *The Bayesian combination operator* $\Phi_{\mathcal{B}}$ *is defined as*

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

where $\hat{p}(y_i|X)$*,* $i \in \{1, 2\}$*, are normalized likelihood functions satisfying conditional independence in the sense of Equation* (2)*. The operator is undefined iff* $\sum_{x \in \Omega_X} \hat{p}(y_1|x)\hat{p}(y_2|x) = 0$.

Let us continue by elaborating on how *the credal combination operator* can be derived based on Def. 1. Since we now move into the domain of imprecise probability, we are allowed to utilize *a closed and convex set of probability distributions*, i.e., a *credal set* [19]. Convexity enables one to perform computation by the sets' *extreme points* (see further Karlsson et al. [18, Theorem 2]). Now, instead of evidence in the form of a single normalized likelihood function, we have credal sets of such functions, denoted by $\hat{\mathcal{P}}(y_1|X)$ and $\hat{\mathcal{P}}(y_2|X)$, which we want to combine into a single joint evidence $\hat{\mathcal{P}}(y_1, y_2|X)$. In order to regard the evidences as independent, the *extreme points* needs to factorize, denoted *strong independence* [7], i.e., for each extreme point $\hat{p}_e(y_1, y_2|X)$ we have:

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

where $\hat{p}(y_i|X) \in \hat{\mathcal{P}}(y_i|X)$, $i \in \{1, 2\}$. The combination is then performed by the credal combination operator, which simply applies the Bayesian combination operator point-wise on each combination of functions from the operand sets and as last step one applies the convex-hull operator [1, 2, 15, 18]:

Definition 2. *The credal combination operator* $\Phi_{\mathcal{C}}$ *is defined as*

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

where $\hat{\mathcal{P}}(y_i|X)$*,* $i \in \{1, 2\}$*, are credal sets of normalized likelihood functions satisfying strong independence in the sense of Eq.* (4)*,* $\Phi_{\mathcal{B}}$ *is the Bayesian combination operator, and* $\mathcal{CH}$ *denotes the convex hull. The operator is undefined iff there exist a pair* $\hat{p}(y_1|X) \in \hat{\mathcal{P}}(y_1|X)$ *and* $\hat{p}(y_2|X) \in \hat{\mathcal{P}}(y_2|X)$ *for which* $\Phi_{\mathcal{B}}$ *is undefined.*

Note that when only singleton sets are used as operands, the credal combination operator is equivalent to the Bayesian counterpart.

One important type of credal set that we will use throughout the article is the *probability simplex*, i.e., the set of all probability distributions over a given state space, formally defined as:

Definition 3. *The set of all probability distributions* $\mathcal{P}^*(X)$*, i.e., the probability simplex, over a given state space* Ω_X *is defined as*

$$\mathcal{P}^*(X) \triangleq \left\{p(X) : p(x) \geq 0, \sum_{x \in \Omega_X} p(x) = 1\right\}. \quad (6)$$

Another important concept with respect to imprecise probability is the *degree of imprecision* of a credal set. When we refer to "imprecision" in this article we perform averaging of the imprecision for single states [29, 16]:

Definition 4. *The degree of imprecision* $\mathcal{I}(\hat{\mathcal{P}}(y|X))$ *of a credal set of normalized likelihood functions* $\hat{\mathcal{P}}(y|X)$ *is defined as:*

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

Please note that we only include the above definition to unambiguously declare the term imprecision. It will not be utilized for any computation in the paper.

3 Shifted Dirichlet Distributions

Probability values can be considered as random variables themselves and the corresponding distributions over such variables is referred to as a *second-order probability distribution* [9, 20, 28]. Any probability distribution that has support on the probability simplex (Def. 3), e.g. a Dirichlet distribution, can be seen as a second-order probability distribution.

The Dirichlet family of distributions can be generalized to have support on subsets of the probability simplex by using lower bounds l_i on the random variables P_i corresponding to first-order probabilities. Just as with related models such as possibility measures [30], belief functions [22], Choquet capacities of order 2 [6] and coherent upper and lower probabilities [25], lower bounds l_i of probabilities determine upper bounds by $1 - \sum_{j \neq i} l_i$. There are other possibilities for lower and upper bounds for the support of second-order probability distribution, e.g., it is possible to give lower and upper bounds for all but one of the first-order probabilities as in, e.g., Sundgren et al. (2009) [26], but for simplicity we give lower bounds l_i to all

$n = |\Omega_X|$ random variables P_i such that $\sum_{i=1}^n l_i \leq 1$ and $\sum_{i=1}^n P_i = 1$, $l_i \leq P_i \leq 1 - \sum_{j \neq i} l_i$.

A probability distribution whose support has been shifted needs renormalization to remain a probability distribution. Let us then look at the probability density function of the *Shifted Dirichlet* family that allows for non-zero lower bounds. If $\sum_{i=1}^n l_i \leq 1$ and $\sum_{i=1}^n P_i = 1$, $l_i \leq P_i \leq 1 - \sum_{j \neq i} l_i$ then the function:

$$f(\{P_i\}_{i=1}^n, \{\alpha_i\}_{i=1}^n, \{l_i\}_{i=1}^n) = \frac{\Gamma\left(\sum_{i=1}^n \alpha_i\right) \prod_{i=1}^n (P_i - l_i)^{\alpha_i - 1}}{\left(1 - \sum_{i=1}^n l_i\right)^{\sum_{i=1}^n \alpha_i - 1} \prod_{i=1}^n \Gamma(\alpha_i)} , \tag{8}$$

is the probability density function of a probability distribution, where P_i are random variables, α_i are the parameters of a proper Dirichlet distribution and l_i are parameters that determine lower bounds of the variables P_i (see the Appendix for a proof). Note that f is a function of P_i only ($\{\alpha_i\}_{i=1}^n$ and $\{l_i\}_{i=1}^n$ are parameters).

4 Second-Order Credal Combination

Now assume that two *agents*, $i \in \{1, 2\}$, for two different types of sensors, have extracted features y_i and that the agents based on this express imprecise independent evidence through lower bounds on normalized likelihoods $\{l_i^j \leq \hat{p}(y_i|x_j)\}_{j=1}^n$ where $\sum_{j=1}^n l_i^j \leq 1$. These lower bounds can then be utilized in order to construct evidence in the form of credal sets of normalized likelihoods by:

$$\hat{\mathcal{P}}(y_i|X) \triangleq \left\{ \hat{p}(y_i|X) : l_i^j \leq \hat{p}(y_i|x_j), \sum_{j=1}^n \hat{p}(y_i|x_j) = 1 \right\} . \tag{9}$$

In addition, the agents also express theirs beliefs over these imprecise operands by specifying alpha-values for the S-Dirichlet distribution, i.e., $\{\alpha_j^i\}_{j=1}^n$. The goal then is to construct a schema for combination that do not only takes evidence in the form of credal sets into account but also *second-order probability* in the form of S-Dirichlet distributions. We will denote such schema by *second-order credal combination* (SOCC).

In order to achieve a computationally feasible schema for SOCC, we propose a simple method for approximating the second-order distribution over the joint evidence by simulation [12, 3]. Typically, these types of simulation utilize a set of so called *particles*[1], i.e., samples, and a set of corresponding *weights* of these particles. Such a representation has, as an example, previously been proposed as a model for *epistemic reliability* by Gärdenfors and Sahlin [11]. Now, we can obtain a set of particles, denoted $\{\hat{p}_j(y_i|X)\}_{j=1}^m$ where each $\hat{p}_j(y_i|X) \in \hat{\mathcal{P}}_j(y_i|X)$, and a set of weights, denoted $\{w_j^i\}_{j=1}^m$, by expanding a grid with a given precision over each operand. At each point in the grid we can compute the density value of the S-Dirichlet and then normalize with respect to all points in the grid [12, Chapter 11]. We can then use the grid as a basis for drawing m particles with replacement. Given these particles $\{\hat{p}_j(y_i|X)\}_{j=1}^n$, and the S-Dirichlet density, we can obtain the corresponding weights by:

$$w_j^i = \frac{f(\{\hat{p}_j(y_i|x_k)\}_{k=1}^n, \{\alpha_j^i\}_{j=1}^n, \{l_j^i\}_{j=1}^n)}{\sum_{j=1}^n f(\{\hat{p}_j(y_i|x_k)\}_{k=1}^n, \{\alpha_j^i\}_{j=1}^n, \{l_j^i\}_{j=1}^n)} . \tag{10}$$

where f is the S-Dirichlet density defined by Eq. (8). Since we now have particles and weights:

$$\Lambda_i \triangleq \{(\hat{p}_j(y_i|X), w_j^i)\}_{j=1}^m \tag{11}$$

from each operand $i \in \{1, 2\}$, we can compute an approximation of the second-order distribution over the joint evidence by combining pairs of particles:

$$\Lambda_{1,2} \triangleq \left\{ \left(\Phi_{\mathcal{B}}(\hat{p}_j(y_1|X), \hat{p}_j(y_2|X)), \frac{w_j^1 w_j^2}{\sum_{j=1}^m w_j^1 w_j^2} \right) \right\}_{j=1}^m . \tag{12}$$

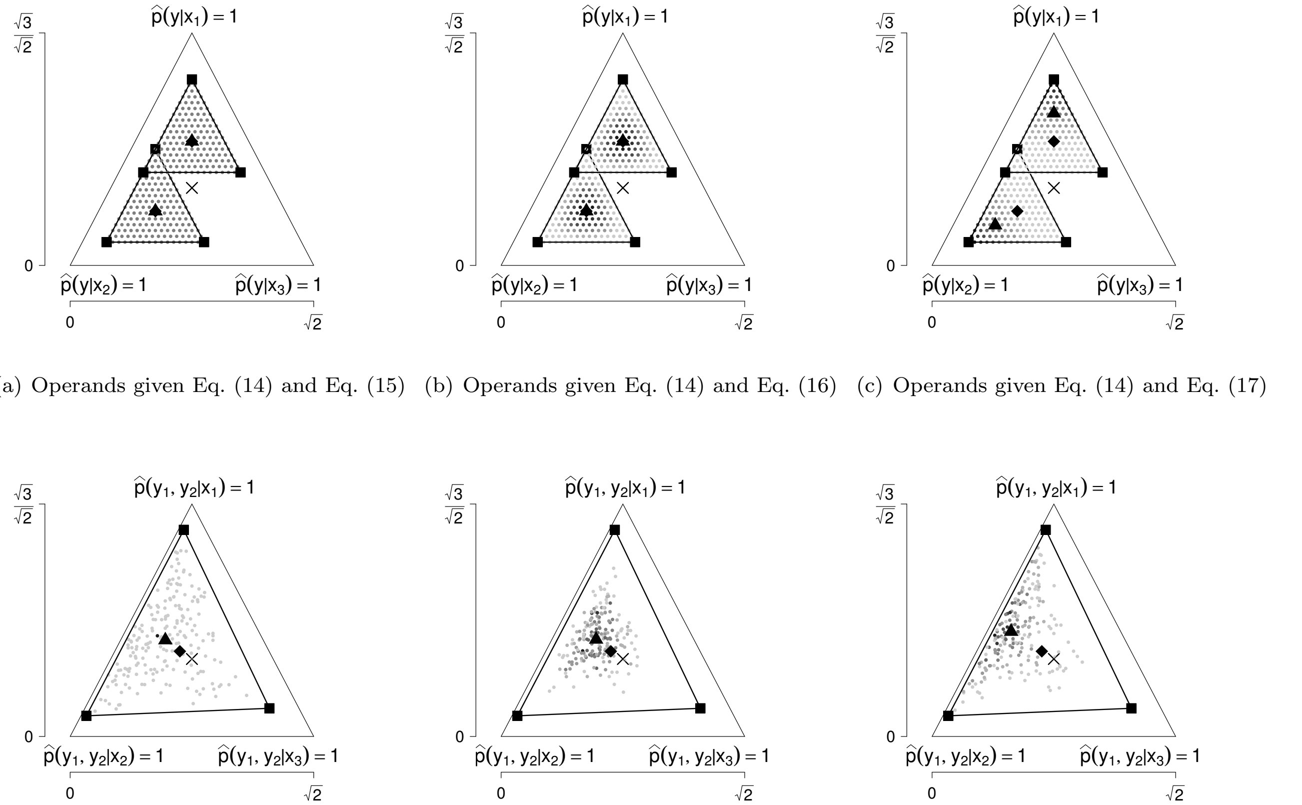
We can combine the above representation with a new operand by drawing a number of particles with replacement from $\Lambda_{1,2}$ where the sampling probability for each particle is given by its weight (similar to *resampling* in the case of particle filtering [3]). The difference is that since we already have a particle representation for $\Lambda_{1,2}$ we do not have to expand a grid for that operand.

In addition to the above described combination schema one can also perform credal combination, i.e:

$$\hat{\mathcal{P}}(y_1, y_2|X) \triangleq \Phi_{\mathcal{C}}(\hat{\mathcal{P}}(y_1|X), \hat{\mathcal{P}}(y_2|X))) . \tag{13}$$

By doing so, one preserve information about the extreme values, which could be valuable for a decision-maker in applications where there exists a strong risk component.

[1]The term "particle filtering" is frequently used in the tracking literature [3].



(a) Operands given Eq. (14) and Eq. (15) (b) Operands given Eq. (14) and Eq. (16) (c) Operands given Eq. (14) and Eq. (17)

(d) SOCC given Eq. (14) and Eq. (15) (e) SOCC given Eq. (14) and Eq. (16) (f) SOCC given Eq. (14) and Eq. (17)

Figure 1: The figures depict the probability simplex for three states where the upper figures show operands for Eq. (14) and Eqs. (15) – (17), and the lower figures show the result of performing SOCC. The intensity of grey shows the weights of the particle (identical particles have been merged by adding the weights), i.e., darker particles have more weights. Squares show extreme points of the credal sets, triangles show the expected value with respect to the particles, diamonds show the centroid of the credal sets, and the uniform distribution over Ω_X is indicated with a cross. The grid used for the operands has been obtained through probability vectors of rational numbers $(a, b, c)/40$, where $a+b+c = 40$, satisfying the lower bounds. In the figures, $m = 200$ particles have been sampled.

4.1 Examples

Let us now study SOCC through some examples where we use the S-Dirichlet distribution for expressing belief over imprecise operands and explore the appearance of the second-order distribution over the imprecise joint evidence. Assume that the two agents provide the following lower bounds on normalized likelihoods:

$$\begin{aligned} l_1 &= (0.1, 0.4, 0.1) \\ l_2 &= (0.4, 0.1, 0.1)\,, \end{aligned} \tag{14}$$

which then can be used in Eq. (9) for constructing $\hat{\mathcal{P}}(y_1|X)$ and $\hat{\mathcal{P}}(y_2|X)$, respectively. Given these lower bounds, we will explore the result of SOCC for three different, somewhat arbitrarily chosen although with some intuition, S-Dirichlet distributions:

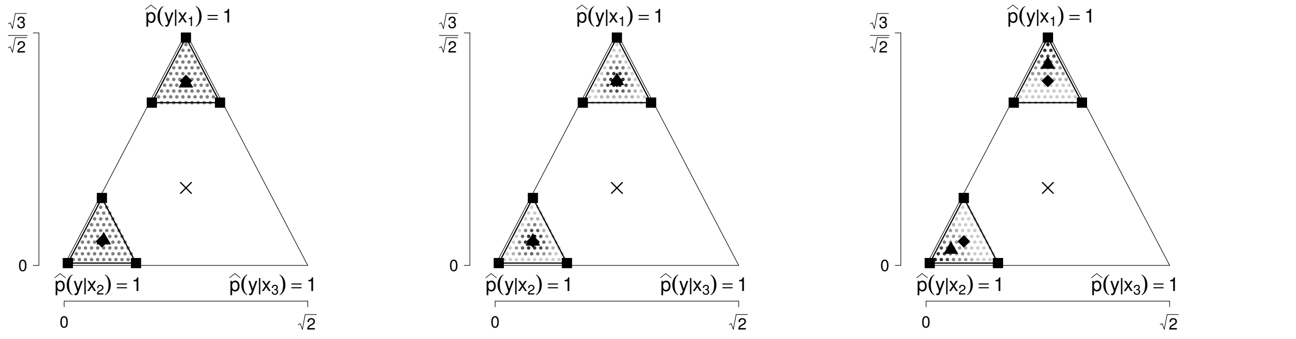
$$\alpha_1 = \alpha_2 = (1, 1, 1) \tag{15}$$

$$\alpha_1 = \alpha_2 = (3, 3, 3) \tag{16}$$

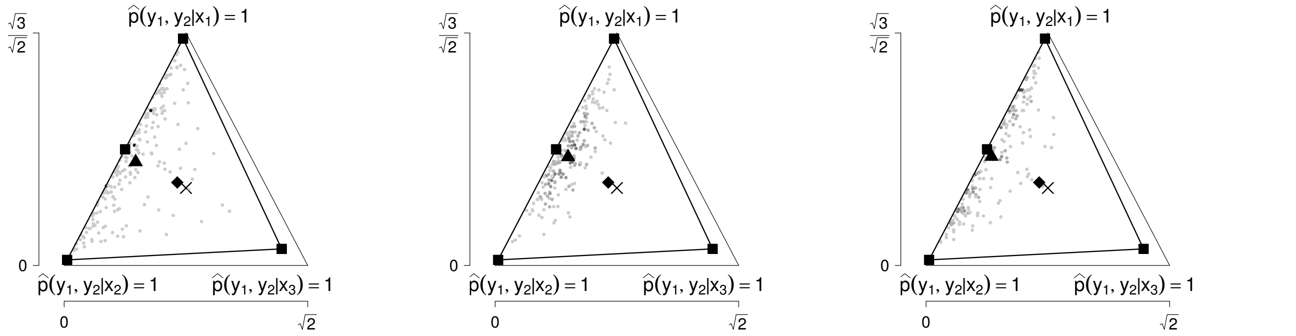
$$\alpha_1 = (1, 3, 1),\ \alpha_2 = (3, 1, 1)\,, \tag{17}$$

where Eq. (15) corresponds to the uniform (Bayes-Laplace) distribution; Eq. (16) is a case where the center of the credal sets is reinforced; and Eq. (17) is a case where the corner closest to some state has been reinforced. The result of applying SOCC on Eq. (14) and Eqs. (15) – (17) is shown in Fig. 1.

We see that the uniform distribution, i.e., Fig. 1(a) and 1(d), yields a particle cloud that is more scattered compared to the other S-Dirichlet distributions. Furthermore, we see that utilizing the S-Dirichlet



(a) Operands given Eq. (18) and Eq. (15) (b) Operands given Eq. (18) and Eq. (16) (c) Operands given Eq. (18) and Eq. (17)



(d) SOCC given Eq. (18) and Eq. (15) (e) SOCC given Eq. (18) and Eq. (16) (f) SOCC given Eq. (18) and Eq. (17)

Figure 2: The figures depict the probability simplex for three states where the upper figures show operands for Eq. (18) and Eqs. (15) – (17), and the lower figures shows the result of performing SOCC. The indicators and other settings are identical as in Fig. 1.

distribution that emphasizes the corners, defined by Eq. (17) and shown in Figs. 1(c) and 1(f), yields an expected value that has a lower probability for state x_3 than the others.

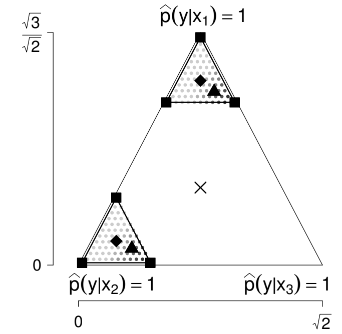
One key observation shown by Figs. 1(e) and 1(f), is that the particle cloud is fairly concentrated within the credal set, which in a sense means that the credal combination operator to some degree overestimates the imprecision. Such an overestimation is even more evident in the following example defined by:

$$\begin{aligned} l_1 &= (0.01, 0.7, 0.01) \\ l_2 &= (0.7, 0.01, 0.01)\,, \end{aligned} \tag{18}$$

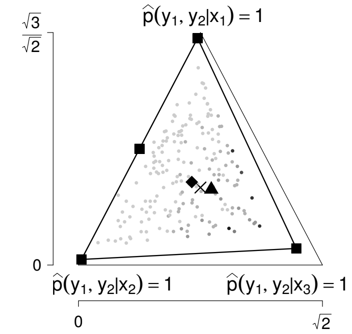
and shown in Fig. 2. If one would have ignored the particle cloud in this example and only base a decision upon the posterior imprecision, it is quite likely that the true state could be x_3 since the lower right extreme point is quite close to the lower right extreme point of the simplex. Such results are somewhat counter-intuitive when interpreting what the evidence from the agents actually express, i.e., evidence for x_1 and x_2, and both pieces constitute counter evidence against x_3 since the operands are positioned far away from the corner corresponding to x_3. However, when combining the two lower right extreme points of the operands, the states x_1 and x_2 are more or less eliminated by the agents, since both of these extreme points are close to the boundary of the simplex where the probability of the these states is close to zero, in contrast to the probability of state x_3, which is not close to any boundary in relative terms. Therefore these lower right extreme points of the operand credal sets gets reinforced to the lower right extreme point of the joint evidence. This case bares close resemblance to Zadeh's counter example [31] against Dempster-Shafer theory [22]. In that example, when combining evidence in the form of *mass functions*, one ends up with a result where all mass lies on the single state that the pieces of evidence only weakly indicated (see further Karlsson et al. [18]). In contrast, from the particle clouds and expected values, we see a clear concentration around the left boundary of the joint evidence; hence in agreement with the intuition that the true state is not likely to be x_3. If the mass instead are concentrated around the lower extreme points of the operands, e.g., by using an S-Dirichlet with the following parameters instead:

$$\alpha_1 = (1, 1, 3),\, \alpha_2 = (1, 1, 3)\,, \tag{19}$$

we obtain the results seen in Fig. 3. In contrast to the



(a) Operands given Eq. (18) and Eq. (19)



(b) SOCC given Eq. (18) and Eq. (19)

Figure 3: The figures depict the probability simplex for three states where the operands are defined by Eqs. (18) – (19). The indicators and other settings are identical as in Fig. 1.

former case, the mass is fairly uniformly distributed over the joint credal set.

The cases shown in Figs. 2 – 3 demonstrates that second-order information could be valuable to a decision-maker when imprecision in decision problems are modelled.

5 Summary and Conclusions

We have generalized the Dirichlet distribution to the S-Dirichlet distribution, where the Dirichlet parameters can be used to model different second-order probability distributions over a restricted region defined by lower bounds. Based on the S-Dirichlet distribution, we presented a simple combination schema, denoted as *second-order credal combination* (SOCC), which takes second-order probability into account. The combination schema is based on a set of particles, sampled from the operands, and a set of weights that are obtained through the S-Dirichlet distribution. We then gave some example of SOCC utilizing different types of S-Dirichlet distributions. By the examples, we showed that the particle cloud over the joint evidence can be remarkably concentrated in comparison to the credal set obtained by credal combination.

One new feature that is enabled through SOCC is that it provides a grounded way of selecting a single probability function from the credal set to base one's decision upon; simply use the expected value with respect to the particle cloud. Such a schema is useful when a single decision is necessary, something that is common in many application scenarios, and is similar to what Smets and Kennes [24] has proposed in the *transferable belief model*, i.e., as long as a single decision does not have to be implemented, use mass functions, otherwise transform the mass function to a single probability function and use that for deciding on a single state. Utilizing the expected value of the particle cloud should be put in contrast to utilizing the centroid distribution, i.e., the expected value with respect to a uniform second-order distribution over the joint evidence. Since uniformity is in general not the case, as is seen in Figs. 2(d) – 2(f) (see also Karlsson et al. [17]), there is in principle no reason to utilize the centroid. Another alternative is utilizing the maximum entropy distribution [1, 2], representing a cautious choice, however, in applications without a risk component, the maximum entropy distribution is likely to be too cautious.

Given the examples where the particle clouds seems to be quite concentrated in comparison to the resulting credal sets, one legitimate question is whether or not it is reasonable to utilize the credal combination operator solely, i.e., without modeling second-order probability. Could it be so that it is always preferable to model second-order probability when imprecision appears in a decision problems? Perhaps the credal combination operator can be appropriate to utilize when the imprecise operands are a consequence of small perturbations of some precise evidence as is done in sensitivity analysis (robust Bayesian theory) [4, 14]. In such a setting it seems reasonable to only model imprecision, and not second-order probability, due to that only low degrees of imprecision in the operands are considered. For these cases one is likely to infer the same conclusions irrespective of any second-order probability since the perturbation of the operands is performed so that every point in the perturbed set is a reasonable precise evidence. Consequently, every perturbed point in the resulting joint evidence is a reasonable joint evidence that a decision maker should be willing to act upon, irrespective of the amount of density such a point possesses.

When the imprecise operands are not a consequence of sensitivity analysis, i.e., when the degree of imprecision of the operands could be considerably higher, then, as our results suggest, second-order probability is likely to be an important modeling tool that cannot be neglected without consequences. In our future research, we will therefore continue by exploring how one can perform different modeling tasks using second-order probability, i.e., how SOCC can be applied in an application scenario.

Acknowledgements

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), and Formas. R [21] and the package rcdd [13] have been used for computations in this paper.

Appendix

We here prove Eq. (8). Let us use the following shorthand notation ($n = |\Omega_X|$):

$$\gamma \triangleq \frac{\Gamma\left(\sum_{i=1}^{n} \alpha_i\right)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \ . \tag{20}$$

We need to show that:

$$\int_{\substack{\sum_{i=1}^{n} P_i = 1 \\ P_i \geq l_i}} \gamma \frac{\prod_{i=1}^{n} (P_i - l_i)^{\alpha_i - 1}}{\left(1 - \sum_{i=1}^{n} l_i\right)^{\sum_{i=1}^{n} \alpha_i - 1}} \, d\mathbf{P} = 1 \ . \tag{21}$$

Since a proper Dirichlet distribution has probability density function:

$$f(\{P_i\}_{i=1}^{n}, \{\alpha_i\}_{i=1}^{n}) = \gamma \prod_{i=1}^{n} P_i^{\alpha_i - 1} \ , \tag{22}$$

we know that:

$$\int_{\substack{\sum_{i=1}^{n} P_i = 1 \\ P_i \geq 0}} \gamma \prod_{i=1}^{n} P_i^{\alpha_i - 1} \, d\mathbf{P} = 1 \ . \tag{23}$$

Let us replace P_i with $P_i - l_i$ and restrict the support from $\sum_{i=1}^{n} P_i = 1, P_i \geq 0$ to $\sum_{i=1}^{n} P_i = 1, l_i \leq P_i \leq 1 - \sum_{j \neq i} l_i$. Then, through the variable change:

$$Y_i = \frac{P_i - l_i}{1 - \sum_{i=1}^{n} l_i} \ , \tag{24}$$

where $i \in \{1, \ldots, n\}$, we find that:

$$\begin{aligned}
&\int_{\substack{\sum_{i=1}^{n} P_i = 1 \\ P_i \geq l_i}} \gamma \prod_{i=1}^{n} (P_i - l_i)^{\alpha_i - 1} \, d\mathbf{P} = \\
&\int_{\substack{\sum_{i=1}^{n} Y_i = 1 \\ Y_i \geq 0}} \gamma \prod_{i=1}^{n} \left(Y_i \left(1 - \sum_{i=1}^{n} l_i\right)\right)^{\alpha_i - 1} \left|\frac{\partial \mathbf{P}}{\partial \mathbf{Y}}\right| \, d\mathbf{Y} = \\
&\int_{\substack{\sum_{i=1}^{n} Y_i = 1 \\ Y_i \geq 0}} \gamma \prod_{i=1}^{n} \left(Y_i \left(1 - \sum_{i=1}^{n} l_i\right)\right)^{\alpha_i - 1} \\
&\qquad \left(1 - \sum_{i=1}^{n} l_i\right)^{n-1} d\mathbf{Y} = \\
&\int_{\substack{\sum_{i=1}^{n} Y_i = 1 \\ Y_i \geq 0}} \gamma \prod_{i=1}^{n} Y_i^{\alpha_i - 1} \left(1 - \sum_{i=1}^{n} l_i\right)^{\sum_{i=1}^{n} \alpha_i - n} \\
&\qquad \left(1 - \sum_{i=1}^{n} l_i\right)^{n-1} d\mathbf{Y} = \\
&\left(1 - \sum_{i=1}^{n} l_i\right)^{\sum_{i=1}^{n} \alpha_i - 1} \int_{\substack{\sum_{i=1}^{n} Y_i = 1 \\ Y_i \geq 0}} \gamma \prod_{i=1}^{n} Y_i^{\alpha_i - 1} \, d\mathbf{Y} = \\
&\left(1 - \sum_{i=1}^{n} l_i\right)^{\sum_{i=1}^{n} \alpha_i - 1} \ .
\end{aligned} \tag{25}$$

Therefore:

$$\frac{1}{\left(1 - \sum_{i=1}^{n} l_i\right)^{\sum_{i=1}^{n} \alpha_i - 1}} \tag{26}$$

is the normalization factor required for compensating the restricted support.

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] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174–188, 2002.

[4] James O. Berger. An overview of robust Bayesian analysis. *Test*, 3:5–124, 1994.

[5] Marco Cattaneo. A generalization of credal networks. In Thomas Augustin, Frank P. A. Coolen, Serafín Moral, and Matthias C. M. Troffaes, editors, *ISIPTA '09, Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 79–88. SIPTA, 2009.

[6] Gustave Choquet. Theory of capacities. *Annales de lInstitut Fourier*, 5:131–295, 1954.

[7] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.

[8] Fabio G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[9] Love Ekenberg and Johan Thorbiörnson. Second-order decision analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, No 1*, 9(1):13–38, 2001.

[10] Dale Fixsen and Ronald P. S. Mahler. The modified Dempster-Shafer approach to classification. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27:96–104, 1997.

[11] Peter Gärdenfors and Nils-Eric Sahlin. Unreliable probabilities, risk taking, and decision making. *Synthese*, 53:361–386, 1982.

[12] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 2004.

[13] Charles J. Geyer, Glen D. Meeden, and incorporates code from cddlib written by Komei Fukuda. *rcdd: rcdd (Computational Geometry)*, 2012. R package version 1.1-7.

[14] David Rios Insua and Fabrizio Ruggeri, editors. *Robust Bayesian Analysis*. Springer, 2000.

[15] 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.

[16] Alexander Karlsson, Ronnie Johansson, and Sten 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.

[17] 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.

[18] 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.

[19] Isaac Levi. *The enterprise of knowledge*. The MIT press, 1983.

[20] Robert F. Nau. Uncertainty aversion with second-order utilities and probabilities. *Management Science*, 52(1):136–145, 2006.

[21] 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.

[22] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[23] Philippe Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8:387–412, 2007.

[24] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[25] Cedric A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B, xxiii*, pages 1–25, 1961.

[26] David Sundgren, Mats Danielson, and Love Ekenberg. Warp effects on calculating interval probabilities. *International Journal of Approximate Reasoning*, 50(9):1360–1368, 2009.

[27] 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.

[28] Lev V. Utkin and Thomas Augustin. Decision making with imprecise second-order probabilities. In *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pages 547–561, 2003.

[29] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.

[30] Lotfi A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, pages 3–28, 1978.

[31] Lotfi A. Zadeh. Review of books: A mathematical theory of evidence. *AI Magazine*, 5:81–83, 1984.