On the credal structure of consistent probabilities

Fabio Cuzzolin
Fabio.Cuzzolin@inrialpes.fr

INRIA Rhone-Alpes

Abstract. In this paper we introduce a novel, simpler form of the polytope of inner Bayesian approximations of a belief function, or "consistent probabilities". We prove that the set of vertices of this polytope is generated by all possible permutations of elements of the domain, mirroring a similar behavior of outer consonant approximations.

1 Introduction

Uncertainty description is a composite field in which different but related approaches compete to gain a wider audience in engineering [1] and business [2] applications. Belief [3], probability, and possibility [4] measures can all be used to represent uncertainty, even though some of them may be more fit to specific domains of applications. Their relation is then a popular object of study. More specifically, it is interesting in several situations to pose the problem of transforming one uncertainty measure into a measure of a different class. In the case of belief functions, this issue goes under the name of "Bayesian" or "consonant" approximation problem, according to whether we seek to approximate a belief measure with a probability or a possibility.
In particular we may request the approximating probabilities to be more informative of the original belief function. The "least commitment principle" [5] postulates indeed that, given a set of measures compatible with a number of constraints, the most appropriate one is the "least informative".
However, there are many ways of measuring the information content of a belief function [6–8]. If we adopt the classical ordering $b' \geq b \;\equiv\; b'(A) \geq b(A)$ $\forall A \subseteq \Theta$, we obtain the set of *inner Bayesian approximations* $\mathcal{P}[b]$ of b, i.e. the probabilities whose values dominates that of b on all events:

$$\mathcal{P}[b] = \big\{p \in \mathcal{P} : b(A) \leq p(A) \; \forall A \subseteq \Theta\big\}. \tag{1}$$

According to Equation (1) this "credal set" of probability distributions is in fact the set of probabilities which admit the original b.f. b as a lower bound. A powerful semantics to belief functions comes from the remark that such probabilities can be obtained by redistributing the basic belief or "mass" of each event $A \subseteq \Theta$ to the elements it contains. Belief functions can indeed be seen as constraints acting on the probability simplex, on which they define the credal set (1).

Geometrically, inner Bayesian approximations or *consistent probabilities* are known to form a *polytope* (the convex closure of a finite number of points) in the space of probability measures, whose center of mass coincides with the pignistic transformation [9, 10]. The credal semantics of belief functions is central in the "Transferable Belief Model" [11–14]. There belief is represented at credal level, while decision are made by recurring to the pignistic transformation. In robust Bayesian statistics, more in general, a large literature exists on the study of convex sets of probability distributions [15–17].

The goal of this paper is to prove a new result on the geometry of consistent probabilities, which greatly simplifies its classical expression.
We prove that the set of actual vertices of the polytope $\mathcal{P}[b]$ is indeed quite small, and determined by all possible permutations of elements of the domain:

$$\mathcal{P}[b] = Cl(p^\rho[b] \; \forall \rho)$$

where $p^\rho[b]$ is a probability determined by a permutation $\rho = \{x_{\rho(1)}, ..., x_{\rho(n)}\}$ of the singletons of Θ. This generates a beautiful symmetry with the (dual) case of *outer consonant approximations* [18, 19], i.e. the consonant b.f.s dominated by b: $\mathcal{O}[b] = \big\{co \in \mathcal{CO} : co(A) \leq b(A) \; \forall A \subseteq \Theta\big\}$ (here $\mathcal{CO}$ denotes the collection of all consonant b.f.s, i.e. belief functions whose focal elements are nested [3]).
As for each maximal chain of focal elements a vertex of the polytope of outer consonant approximations is also determined by a permutation of singletons, there exists a 1-1 correspondence between actual vertices of $\mathcal{P}[b]$ and $\mathcal{O}[b]$.

Paper outline. We recall in Section 2 the interpretation of belief functions as lower bounds to a convex set of probability distributions, called "consistent probabilities". In Section 3, the core of the paper, we prove that the actual vertices of the polytope of all consistent probabilities are each associated with a permutation of the elements of the domain, and discuss their uniqueness. These vertices are in 1-1 correspondence with the vertices of the region of outer consonant approximations induced by singleton permutations (Section 4).

2 Probabilities consistent with a belief function

Belief [3], probability, and possibility [4] theory are different but related descriptions of uncertainty, as (at least in the finite setting) both probabilities and possibilities are special cases of belief functions.
If we suppose that the ideal knowledge state is represented by a "true", but unknown probability measure (which we cannot estimate precisely because of imprecise measurements, missing data, etcetera) belief measures have in turn a natural interpretation as lower/upper bounds to this unknown true probability. Each belief function is then naturally associated with the set of probabilities which actually meet these constraints.

Belief measures. Belief functions are mathematical representations of the bodies of evidence we possess on a given decision or estimation problem Q. We assume that the possible answers to Q form a finite set $\Theta = \{x_1, ..., x_n\}$ called "frame of discernment". A *basic probability assignment* (b.p.a.) [3] over Θ is a

function $m : 2^\Theta \to [0,1]$ on its power set $2^\Theta = \{A \subseteq \Theta\}$ such that: 1. $m(\emptyset) = 0$; 2. $\sum_{A\subseteq\Theta} m(A) = 1$; 3. $m(A) \geq 0 \; \forall A \subseteq \Theta$. Subsets of Θ associated with non-zero values of m are called "focal elements".
The *belief function* (b.f.) $b : 2^\Theta \to [0,1]$ associated with a basic probability assignment m_b on Θ is defined as

$$b(A) = \sum_{B\subseteq A} m_b(B). \tag{2}$$

In the following we will denote by b_A the "dogmatic" b.f. which assigns unitary mass to a single event A: $m_b(A) = 1$, $m_b(B) = 0 \; \forall B \neq A$. We can then write each belief function b with b.p.a. $m_b(A)$ as [21]

$$b = \sum_{A\subseteq\Theta} m_b(A) b_A. \tag{3}$$

In the theory of evidence a probability is just a special belief function assigning non-zero masses to singletons only (*Bayesian* b.f.): $m_b(A) = 0 \; |A| > 1$. The *plausibility function* (pl.f.) $pl_b : 2^\Theta \to [0,1]$, $pl_b(A) \doteq 1 - b(A^c)$ measures instead the amount of evidence *not against* A.

Consistent probabilities or inner Bayesian approximations. Even though originally defined as set functions of the form (2) on the power set of a finite universe [3], belief functions have a natural interpretation as constraints on the "true", unknown probability distribution describing a state of belief.
Given a certain amount of evidence we are allowed to describe our belief on the outcome of Q in several possible ways: the classical option is to assume a probability distribution on Θ. However, as we may need to incorporate imprecise measurements and people's opinions in our knowledge state, or cope with missing or scarce information, a more sensible approach is to assume that we have no access to the "correct" probability distribution but that the available evidence provides us with some sort of constraint on this true distribution. Belief functions are mathematical descriptions of such a constraint.

According to this interpretation a f.e. A of mass $m_b(A)$ can be seen as the indication of the existence of a mass $m_b(A)$ "floating" inside A. The mass assigned to each event $A \subseteq \Theta$ can float freely among its elements $x \in A$. A probability distribution compatible with b emerges by redistributing the mass of each focal element to its singletons. This set of Bayesian b.f.s is said "consistent" with b.

Example. To illustrate the notion of probability consistent with a belief function let us consider a little toy example, namely a b.f. b on a frame of cardinality three $\Theta = \{x, y, z\}$ with focal elements (Figure 1-a)

$$m_b(\{x,y\}) = \frac{2}{3}, \qquad m_b(\{y,z\}) = \frac{1}{3}. \tag{4}$$

One way of obtaining a probability consistent with b is, for instance, to equally share the mass of $\{x, y\}$ among its elements x and y, while attributing the entire mass of $\{y, z\}$ to y (Figure 1-b). Or, we can assign all the mass of the focal

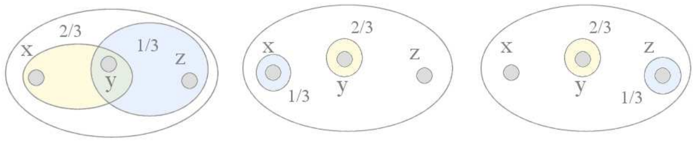

Fig. 1. a) A simple belief function in the ternary frame. b) and c) two admissible probabilities consistent with a).

element $\{x, y\}$, and give the mass of $\{y, z\}$ to z only, obtaining the Bayesian belief function of Figure 1-c.

Belief functions as lower bounds. An alternative vision of the set of probabilities consistent with a b.f. b relates to the fact that the set of all and only the (admissible) probabilities obtained as above by re-assigning the mass of each f.e. to its elements is (1) $\mathcal{P}[b] \doteq \big\{p \in \mathcal{P} : p(A) \geq b(A)\ \forall A \subseteq \Theta\big\}$ i.e. the distributions whose values dominate that of b on all events A.
A b.f. can then be interpreted as a "lower bound" to the set of probabilities it determines. For instance, the distribution of Figure 1-b) meets all the lower and upper bounds determined by the belief function (4). Indeed: $p(x) = 1/3 \geq b(x) = 0$, $p(\{x, y\}) = 1/3 + 2/3 = 1 \geq b(\{x, y\}) = m_b(\{x, y\}) = 2/3$, etcetera.

Consistent belief functions as inner Bayesian approximations. Finally, consistent probabilities can be seen as the set of all the probabilities *more committed* than the original b.f. b.
The "least commitment principle" [5] postulates that, given a set of b.p.a.s compatible with a number of constraints, the most appropriate mass functions is the "least informative". As pointed out by Denoeux [6], in some sense it plays a role similar to that of maximum entropy in probability theory. There are many ways of measuring the information content of a belief function. This is done in practice by defining a partial order in the space of belief functions [7, 8, 20]. If we adopt the order relation called *weak inclusion*

$$b \leq b' \equiv b(A) \leq b'(A)\ \ \forall A \subseteq \Theta, \tag{5}$$

according to which a belief function b' dominates another b.f. b if the belief values of b' are greater than those of b for all events $A \subseteq \Theta$, the consistent probabilities (1) are exactly the group of Bayesian belief functions more committed than b according to (5). They therefore assume the meaning of *inner Bayesian approximations* of the original belief function b.

3 Vertices of consistent probabilities and permutations

The region of all consistent probabilities or inner Bayesian approximations of any given belief function b is a *polytope*, i.e. the convex closure of a finite number of probabilities in the probability simplex. Given an arbitrary belief function b with focal elements $E_1, ..., E_m$, we can define for each choice of m representatives $\{x_1, ..., x_m\}$, $x_i \in E_i \ \forall i$, the (*extremal*) probability measure

$$b_{x_1 \dots x_m} \doteq \sum_{i=1}^{m} m_b(E_i) b_{x_i}. \tag{6}$$

i.e. the Bayesian b.f. we obtain by assigning the mass of each focal element E_i to one of its elements $x_i \in E_i$. Recall what said above about the interpretation of focal elements as mass "floating" around in a subset of Θ.
For instance, consider again the belief function (4) of Figure 1-a. If we take as representative x_1 of $E_1 = \{x, y\}$ the element y and as representative x_2 of $E_2 = \{y, z\}$ the element z, we obtain the extremal probability

$$b_{y,z} : m_{b_{y,z}}(y) = m_b(\{x, y\}) = 2/3, \qquad m_{b_{y,z}}(z) = m_b(\{y, z\}) = 1/3$$

of Figure 1-c. Note that instead the probability of Figure 1-b, even though is consistent with b, cannot be obtained this way.

It is well known that $\mathcal{P}[b]$ is indeed the polytope formed by the convex closure of those extremal probabilities [21]:

$$\mathcal{P}[b] = Cl(b_{x_1 \dots x_m}, \{x_1, ..., x_m\} \in E_1 \times \cdots \times E_m). \tag{7}$$

That, though, does not imply that all the points (6) are actual vertices of the simplex $\mathcal{P}[b]$, as several of them may lie on some side of the polytope, i.e., be expressed as a convex combination of the others. In fact, as we show here, the actual vertices of $\mathcal{P}[b]$ can be found in a much smaller set of probabilities, each one associated with a different permutation of the elements of Θ.

Theorem 1. *The simplex $\mathcal{P}[b]$ of the probability measures consistent with a b.f. b is the polytope $\mathcal{P}[b] = Cl(p^\rho[b] \ \forall \rho)$, where ρ is any permutation $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$ of the singletons of Θ, and the vertex $p^\rho[b]$ is the Bayesian b.f. such that*

$$p^\rho[b](x_{\rho(i)}) = \sum_{A \ni x_\rho(i), A \not\ni x_\rho(j) \forall j < i} m_b(A). \tag{8}$$

Each probability function (8) attributes to each singletons $x = x_{\rho(i)}$ the mass of all focal elements of b which contains it, but does not contain the elements which precede x in the ordered list $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$ generated by the permutation ρ.

In the binary case $\Theta = \{x, y\}$, for instance, it is clear that there exist only two possible permutations of singletons: $\rho^1 : \{x, y\}$, $\rho^2 : \{y, x\}$. They correspond to the following probabilities (Figure 3):

$$p^{x,y}(x) = \sum_{A \supseteq \{x\}} m_b(A) = m_b(x) + m_b(\Theta), \ p^{x,y}(y) = \sum_{A \supseteq \{y\}, A \not\supset \{x\}} m_b(A) = m_b(y);$$
$$p^{y,x}(y) = \sum_{A \supseteq \{y\}} m_b(A) = m_b(y) + m_b(\Theta), \ p^{y,x}(x) = \sum_{A \supseteq \{x\}, A \not\supset \{y\}} m_b(A) = m_b(x).$$

Proof. We need to prove that:
1. each probability $p \in \mathcal{P}$ s.t. $p(A) \geq b(A)$ for all $A \subseteq \Theta$ can be put as a convex combination of the points (8): $p = \sum_\rho \alpha_\rho p^\rho[b]$ with $\sum_\rho \alpha_\rho = 1$, $\alpha_\rho \geq 0\ \forall \rho$;
2. vice-versa, each convex combination of the $p^\rho[b]$ satisfies $\sum_\rho \alpha_\rho p^\rho[b](A) \geq b(A)$ $\forall A \subseteq \Theta$.

Point **2** is easily proven after we notice that each probability (8) associated with a permutation ρ of elements of Θ is indeed consistent with b, i.e. $p^\rho[b](A) \geq b(A)\ \forall A$. Whatever ρ the mass of each of the subsets B of A is attributed by (8) to some element x of A: on the other side the mass of some other events $B \not\subset A$ is also given to elements of A, so that $p^\rho[b](A) = \sum_{x \in A} p^\rho[b](x) \geq \sum_{B \subseteq A} m_b(A) = b(A)$, i.e., $p^\rho[b]$ is consistent whatever the permutation ρ. Therefore

$$\sum_\rho \alpha_\rho p^\rho[b](A) \geq \sum_\rho \alpha_\rho b(A) = b(A) \sum_\rho \alpha_\rho = b(A).$$

Concerning point **1**, we recalled in Section 2 that $b'(A) \leq b(A)$ iff m_b is the result of a redistribution of the mass $m_{b'}(A)$ of each f.e. of b' to its subsets. In the case of inner Bayesian approximations the mass of each event has to be redistributed among its elements $x \in A$:

$$m_b(A) \mapsto \alpha_x^A m_b(A) \qquad \forall x \in A, \quad \sum_{x \in A} \alpha_x^A = 1. \tag{9}$$

Therefore, for all p such that $b(A) \leq p(A)\ \forall A$ the mass $p(x)$ of each $x \in \Theta$ is

$$p(x) = \sum_{A \supseteq \{x\}} m_b(A) \alpha_x^A. \tag{10}$$

To prove (1) we then need to write (10) as a convex combination of the $p^\rho[b](x)$

$$p(x) = \sum_\rho \alpha_\rho p^\rho[b](x) = \sum_\rho \alpha_\rho \left(\sum_{A \ni x = x_{\rho(i)}, A \not\ni x_{\rho(j)} \forall j < i} m_b(A) \right)$$

where i is the position of the element x according to the permutation ρ.
For all $A \supseteq \{x\}$ there exists a permutation ρ such that the elements before x in $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$ fall outside A. Hence the above quantity reads as $\sum_{A \supseteq \{x\}} m_b(A) (\sum_{\rho : x_{\rho(j)} \notin A \forall j < i} \alpha_\rho)$, where again $x = x_{\rho(i)}$.
In summary we need to show that the system of equations

$$\left\{ \alpha_x^A = \sum_{\rho : x_{\rho(j)} \notin A \forall j < i, x = x_{\rho(i)}} \alpha_\rho \qquad \forall x \in \Theta, \quad \forall A \supseteq \{x\} \right. \tag{11}$$

has at least one solution $\{\alpha_\rho\}$ such that $\sum_\rho \alpha_\rho = 1$, $\alpha_\rho \geq 0\ \forall \rho$.

Parenthesis: proof in the ternary case. It is useful to first illustrate the existence of a convex solution to (11) in the simple but interesting case of a ternary frame $\Theta = \{x, y, z\}$. The possible permutations of singletons in this case are six: $\rho^1 = \{x, y, z\}$, $\rho^2 = \{x, z, y\}$, $\rho^3 = \{y, x, z\}$, $\rho^4 = \{y, z, x\}$, $\rho^5 = \{z, x, y\}$, $\rho^6 = \{z, y, x\}$. The system of equations (11) reads as

$$\begin{cases} \alpha_x^{\{x\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6}; \\ \alpha_x^{\{x,y\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + 0 \quad + 0 \quad + \alpha_{\rho^5} + 0; \\ \alpha_x^{\{x,z\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + 0 \quad + 0 \quad + 0; \\ \alpha_x^{\Theta} &= \alpha_{\rho^1} + \alpha_{\rho^2} + 0 \quad + 0 \quad + 0 \quad + 0; \\ \alpha_y^{\{y\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6}; \\ \alpha_y^{\{x,y\}} &= 0 \quad + 0 \quad + \alpha_{\rho^3} + \alpha_{\rho^4} + 0 \quad + \alpha_{\rho^6}; \\ \alpha_y^{\{y,z\}} &= \alpha_{\rho^1} + 0 \quad + \alpha_{\rho^3} + \alpha_{\rho^4} + 0 \quad + 0; \\ \alpha_y^{\Theta} &= 0 \quad + 0 \quad + \alpha_{\rho^3} + \alpha_{\rho^4} + 0 \quad + 0; \\ \alpha_z^{\{z\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6}; \\ \alpha_z^{\{x,z\}} &= 0 \quad + 0 \quad + 0 \quad + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6}; \\ \alpha_z^{\{y,z\}} &= 0 \quad + \alpha_{\rho^2} + 0 \quad + 0 \quad + \alpha_{\rho^5} + \alpha_{\rho^6}; \\ \alpha_z^{\Theta} &= 0 \quad + 0 \quad + 0 \quad + 0 \quad + \alpha_{\rho^5} + \alpha_{\rho^6}. \end{cases}$$

As by definition $\alpha_x^x = 1$ all equations associated with a singleton generate the normalization constraint $\alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6} = 1$. Also, many equations in the above system are actually linearly dependent, like for example equations 2 and 6 for $\alpha_x^{\{x,y\}}$ and $\alpha_y^{\{x,y\}}$.
This because by definition (9): $\sum_{x \in A} \alpha_x^A = 1$.
After eliminating the dependencies we get a reduced system

$$\begin{cases} \alpha_z^{\Theta} &= 0 \quad + 0 \quad + 0 \quad + 0 \quad + \alpha_{\rho^5} + \alpha_{\rho^6}; \\ \alpha_x^{\{x,y\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + 0 \quad + 0 \quad + \alpha_{\rho^5} + 0; \\ \alpha_y^{\{y,z\}} &= \alpha_{\rho^1} + 0 \quad + \alpha_{\rho^3} + \alpha_{\rho^4} + 0 \quad + 0; \\ \alpha_x^{\{x,z\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + 0 \quad + 0 \quad + 0; \\ \alpha_x^{\Theta} &= \alpha_{\rho^1} + \alpha_{\rho^2} + 0 \quad + 0 \quad + 0 \quad + 0; \\ \alpha_x^{\{x\}} &= \alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6}; \end{cases}$$

which clearly admits as solution

$$\alpha_{\rho^6} = \alpha_z^{\Theta}, \, \alpha_{\rho^5} = \alpha_x^{\{x,y\}}, \, \alpha_{\rho^4} = \alpha_y^{\{y,z\}}, \, \alpha_{\rho^3} = \alpha_x^{\{x,z\}}, \, \alpha_{\rho^2} = \alpha_x^{\Theta}, \, \alpha_{\rho^1} = \alpha_x^{\{x\}}$$

which represents a valid convex combination of the $p^\rho[b]$.

Of course there are many ways of obtaining a reduced system, and therefore many acceptable convex solutions to (11).

General solution. As like in the ternary case the normalization constraint is in fact trivially satisfied as from (11) it follows that when $A = \{x\}$, $x \in \Theta$

$$1 = \alpha_x^x = \sum_{\rho: x_{\rho(j)} \notin \{x\} \forall j < i, x = x_{\rho(i)}} \alpha_\rho = \sum_\rho \alpha_\rho$$

i.e. $\sum_\rho \alpha_\rho = 1$. Let us denote by $x_{|A|}$ any element representative of A. Due to the normalization constraint the system of equations (11) reduces to

$$\left\{ \alpha_x^A = \sum_{\rho: x_{\rho(j)} \notin A \forall j < i, x = x_{\rho(i)}} \alpha_\rho \qquad \forall A \subseteq \Theta, x \neq x_{|A|}. \right. \tag{12}$$

Again $\forall A$ s.t. $|A| = 1$ we get simply the normalization constraint.
To understand the structure of (12) consider some arbitrary ordering of the elements of Θ, $x_1, ..., x_n$. If we take as representative of any event A its last element according to this ordering, we can write (12) as

$$\left\{ \alpha^A_{x_k} = \sum_{\rho: x_{\rho(j)} \notin A \forall j < i, x_k = x_{\rho(i)}} \alpha_\rho, \quad \begin{array}{l} A \supseteq \{x_k\} \\ A \not\subset \{x_1, ..., x_k\} \end{array} \right. \tag{13}$$

each block associated with x_k, $k = 1, ..., n-1$. The number of equations in each block k for a frame Θ of size $|\Theta| = n$ is $|\{A \subseteq \Theta : A \supseteq \{x_k\}, A \not\subset \{x_1, ..., x_k\}\}| = |\{A \subseteq \Theta : A \supseteq \{x_k\}, A \cap \{x_1, ..., x_k\}^c \neq \emptyset\}| = 2^{i-1}(2^{n-i} - 1) = 2^{n-1} - 2^{i-1}$.
Now, all the equations of each block k involve (amongst others) the α_ρ related to permutations ρ which put x_k in the first position: $x_k = x_{\rho(1)}$, as it obviously has no predecessors so that there is no $j < i$ in the subscript of the sum in (12) or (13). The number of such permutations is clearly $(n-1)!$ (the number of possible orderings of the $n-1$ successors of x_k).

But for $n > 4$ we have that $(n-1)! \geq 2^{n-1} - 2^{i-1}$: Each block has less equations than the number of permutations associated with variables α_ρ which appear in all the equations of the block. Therefore we can assign the first term of each equation of the block to one of those variables: $\alpha^A_{x_k} = \alpha_\rho$ for some ρ which puts x_k in the first position, this for all $A : A \supseteq \{x_k\}, A \not\subset \{x_1, ..., x_k\}$ (all equations in the block).
Variables associated with the remaining permutations can be set to zero. This yields a convex solution to (12), and therefore to the original system (11).
If $n = 3$ we have seen that a solution also exists. For $n = 2$ the solution is trivial.
If $n = 4$ the condition $(n-1)! \geq 2^{n-1} - 2^{i-1}$ still holds for all blocks but the first one, for which the number of equations is $2^{n-1} - 2^{i-1} = 7$ while the number of variables in common is $(n-1)! = 6$. But it suffices to use the normalization constraint to replace the equation for Θ in the block x_1 (which is in excess) with an equation for Θ in the block x_3 and obtain an equivalent system of equations which meets the desired property. □

Uniqueness. We may wonder whether all the extremal points (8) generated by distinct permutations of singletons are guaranteed to be distinct. The answer is negative. Consider a belief function with b.p.a. $m_b(x) = 0.2$, $m_b(y) = 0.1$, $m_b(z) = 0.3$, $m_b(\{x, y\}) = 0.1$, $m_b(\{y, z\}) = 0.2$, $m_b(\Theta) = 0.1$ defined on a ternary frame $\Theta = \{x, y, z\}$. In this case there are six possible element permutations. Therefore by Theorem 1 $\mathcal{P}[b]$ has as vertices

$$\begin{array}{lll}
\rho^1 = \{x, y, z\} : p^{\rho^1}[b](x) = .4, & p^{\rho^1}[b](y) = .3, & p^{\rho^1}[b](z) = .3; \\
\rho^2 = \{x, z, y\} : p^{\rho^2}[b](x) = .4, & p^{\rho^2}[b](y) = .1, & p^{\rho^2}[b](z) = .5; \\
\rho^3 = \{y, x, z\} : p^{\rho^3}[b](x) = .2, & p^{\rho^3}[b](y) = .5, & p^{\rho^3}[b](z) = .3; \\
\rho^4 = \{y, z, x\} : p^{\rho^4}[b](x) = .2, & p^{\rho^4}[b](y) = .5, & p^{\rho^4}[b](z) = .3; \\
\rho^5 = \{z, x, y\} : p^{\rho^5}[b](x) = .3, & p^{\rho^5}[b](y) = .1, & p^{\rho^5}[b](z) = .6; \\
\rho^6 = \{z, y, x\} : p^{\rho^6}[b](x) = .2, & p^{\rho^6}[b](y) = .2, & p^{\rho^6}[b](z) = .6;
\end{array} \tag{14}$$

and we can notice that the permutations $\rho^3 = \{y, x, z\}$ and $\rho^4 = \{y, z, x\}$ yield the same function: $p^{\rho^3}[b] = p^{\rho^4}[b]$. According to the classical expression (7) of $\mathcal{P}[b]$, instead, there are many more (candidate) vertices (6): $\prod_{A \subseteq \Theta : m_b(A) \neq 0} |A|$. Many fall on the sides or the interior of $\mathcal{P}[b]$ (Figure 2).

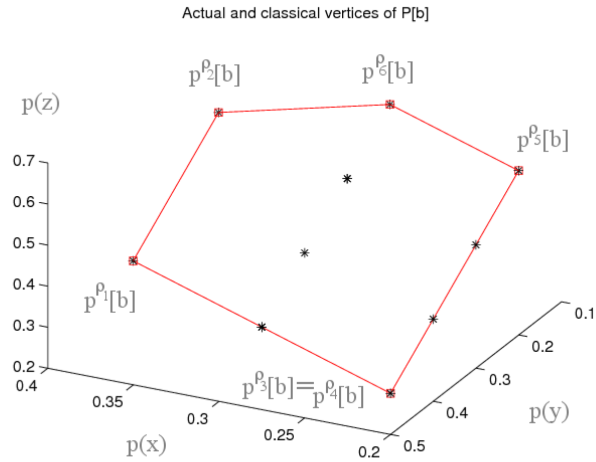

Fig. 2. The actual number of vertices (8), (14) of $\mathcal{P}[b]$ (red squares) is much smaller than the number of candidate points (6) of the classical expression. Here they are plotted as black stars for the belief function of the example. Some of them even fall inside the polytope.

4 Bayesian and consonant approximations: a symmetry

Besides including finite probabilities are a special case, belief measures also generalize finite *possibility* [22] measures, i.e. functions $Pos : 2^\Theta \rightarrow [0, 1]$ on Θ such that $Pos(\bigcup_i A_i) = \sup_i Pos(A_i)$ for any family of sets $\{A_i, i \in I\}$ (where I is an arbitrary set index).

More precisely, a b.f. is *consonant* (co.b.f.) when its focal elements $\{E_i, i = 1, ..., m\}$ are nested: $E_1 \subset E_2 \subset ... \subset E_m$. As a matter of fact it can be proven that [4, 23] the plausibility function pl_b associated with a belief function b on a domain Θ is a possibility measure iff b is consonant.

As possibility measures form a subclass of belief functions we can pose the problem of approximate a belief function with a possibility (or equivalently with a consonant b.f.) in perfect analogy to the case of Bayesian approximation. In particular, "outer consonant approximations" form a dual couple with inner Bayesian approximations or consistent probabilities.

Outer consonant approximations. We call *outer consonant approximations* of a belief function b [19] all the co.b.f.s which are *less committed* than the original belief function b: $\mathcal{O}[b] = \big\{co \in \mathcal{CO} : co(A) \leq b(A) \ \forall A \subseteq \Theta\big\}$. Here $\mathcal{CO}$ denotes the set of all consonant b.f.s.
According to the interpretation of the weak inclusion relation (5), $b' \leq b$ is equivalent to say that b' is obtained by letting the mass of each focal element A of b float to one or more events containing A: $B \supseteq A$.
If b' is also consonant, its focal elements have to form a chain $E_1 \subset \cdots \subset E_n$, $|E_i| = i$. An outer consonant approximation of b is then obtained by letting the mass of each f.e. be re-distributed to one or more elements of the chain.

Example. As an example, an outer consonant approximation of the belief function (4) of Figure 1 can be obtained by re-assigning the mass $2/3$ of $A = \{x, y\}$ one half ($1/3$) to $\{x, y\}$ itself and one half ($1/3$) to $\Theta \supset \{x, y\}$, and the mass of $A = \{y, z\}$ to $\Theta \supset \{y, z\}$ also. What we get is a consonant b.f. with focal elements $\{x, y\} \subset \Theta$ and b.p.a. $m'(\{x, y\}) = 2/3$, $m'(\Theta) = 1/3$.

Outer consonant approximations generated by permutations. In particular, with the purpose of finding outer approximations which are minimal with respect to the weak inclusion relation (5), Dubois and Prade [18] have introduced a family of outer consonant approximations obtained by considering all permutations ρ of the elements $\{x_1, ..., x_n\}$ of the frame of discernment Θ: $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$. A family of nested sets can be then built

$$\big\{S^\rho_1 = \{x_{\rho(1)}\}, S^\rho_2 = \{x_{\rho(1)}, x_{\rho(2)}\}, \cdots, S^\rho_n = \{x_{\rho(1)}, ..., x_{\rho(n)}\}\big\} \tag{15}$$

so that a new consonant belief function co^ρ can be defined with b.p.a.

$$m_{co^\rho}(S^\rho_j) = \sum_{i: E_i \subseteq S^\rho_j, E_i \not\subset S^\rho_{j-1}} m_b(E_i). \tag{16}$$

S^ρ_j concentrates all the mass of the f.e.s E_i of b included in S^ρ_j but not in S^ρ_{j-1}.

Example. Let us consider again the belief function (4). A possible permutation of the singletons of Θ is, for instance, $\rho = \{x_{\rho(1)}, x_{\rho(2)}, x_{\rho(3)}\} = \{y, z, x\}$. This permutation generates the following list of nested sets (15):

$$\big\{S^\rho_1 = \{x_{\rho(1)}\} = \{y\}, S^\rho_2 = \{x_{\rho(1)}, x_{\rho(2)}\} = \{y, z\}, \\ S^\rho_3 = \{x_{\rho(1)}, x_{\rho(2)}, x_{\rho(3)}\} = \{x, y, z\}\big\}.$$

By Equation (16) we assign to $S^\rho_1 = \{y\}$ the mass of all focal elements (4) of b included in $\{y\}$: there are none, so that $m_{co^\rho}(\{y\}) = 0$. To $S^\rho_2 = \{y, z\}$ we assign the mass of all f.e.s inside $\{y, z\}$ not contained in $\{y\}$, i.e. the mass $1/3$ of $\{y, z\}$ itself. Finally, $S^\rho_3 = \{x, y, z\}$ is assigned the mass of all f.e.s which are subsets of $\{x, y, z\}$, but not of $S^\rho_2 = \{y, z\}$, namely the mass $2/3$ of $\{x, y\}$.

Binary case. In the binary case (Figure 3) a compelling symmetry emerges between $\mathcal{O}[b]$ and $\mathcal{P}[b]$. Given the definition of weak inclusion (5) it is straightforward to recognize that (as the abscissa measures the degree of belief $b(x)$ of x, and the ordinate the d.o.f. $b(y)$) the set of inner Bayesian approximations and the set of outer consonant approximations form respectively a segment $\mathcal{P}[b]$

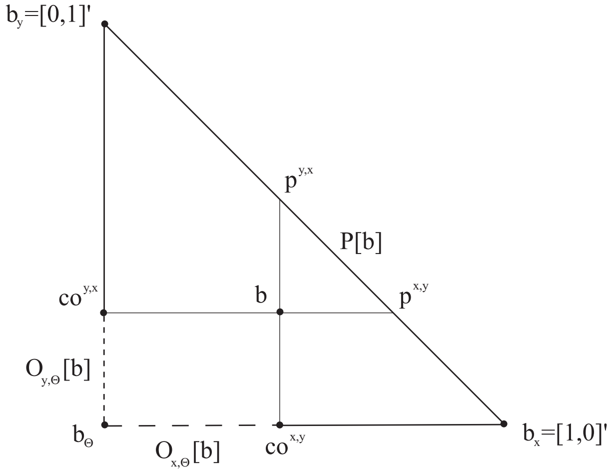

Fig. 3. Geometry of outer consonant approximations and inner Bayesian approximations in the binary case. Each b.f. is represented as a point b of coordinates $b(x), b(y)$. The set of all b.f.s on $\{x, y\}$ is the triangle in the figure. Probability measures lie on the line b_x, b_y. The sets of inner Bayesian $P[b]$ and outer consonant $O[b]$ approximations are highlighted.

delimited by a pair of probabilities, and the union of two segments $O_{x,\Theta}[b]$ and $O_{y,\Theta}[b]$.
From Figure 3 we can notice the existence of an apparent bijection between vertices of $\mathcal{O}[b]$ and $\mathcal{P}[b]$. The vertex $p^{y,x}$ of the interval of consistent probabilities has the same belief value on x as the vertex $co^{x,y}$ of outer consonant approximation associated with the opposite permutation $\{x, y\}$.
In fact we can prove that the outer consonant approximations (16) generated by permutations of singletons form a subset of the vertices of $\mathcal{O}[b]$ [24] even in the general case. Furthermore, the correspondence between vertices of $\mathcal{O}[b]$ and $\mathcal{P}[b]$ generated by permutations of singletons hold in the general case.

1-1 correspondence. Indeed, the family of outer consonant approximations (16) generated by permutations of singletons is linked by a very elegant geometric duality to the vertices (8) of the polytope of consistent probabilities $\mathcal{P}[b]$.
Let us go back to the example of the ternary frame $\Theta = \{x, y, z\}$. Given for instance the trivial permutation $\rho = \{x, y, z\}$ the vertex co^ρ has focal elements $\{x\}$, $\{x, y\}$, and $\{x, y, z\}$ and b.p.a.

$$\begin{aligned}
m_{co^\rho}(\{x\}) &= m_b(x); \\
m_{co^\rho}(\{x, y\}) &= \sum_{A \subseteq \{x,y\}, A \not\subset \{x\}} m_b(A) = m_b(\{y\}) + m_b(\{x, y\}); \\
m_{co^\rho}(\Theta) &= \sum_{A \subseteq \Theta, A \not\subset \{x,y\}} m_b(A) = m_b(\{z\}) + m_b(\{x, z\}) + m_b(\{y, z\}) + m_b(\Theta).
\end{aligned}$$

Consider instead the reverse permutation $\bar{\rho} = \{z, y, x\}$. The corresponding vertex $p^{\bar{\rho}}$ of $\mathcal{P}[b]$ has b.p.a.

$$\begin{aligned} p^{\bar{\rho}}(z) &= \sum_{A \supseteq \{z\}} m_b(A) = m_b(\{z\}) + m_b(\{x,z\}) + m_b(\{y,z\}) + m_b(\Theta); \\ p^{\bar{\rho}}(y) &= \sum_{A \supseteq \{y\}, A \not\supset \{z\}} m_b(A) = m_b(\{y\}) + m_b(\{x,y\}); \\ p^{\bar{\rho}}(x) &= \sum_{A \supseteq \{x\}, A \not\supset \{y\},\{z\}} m_b(A) = m_b(x) \end{aligned}$$

i.e. the b.p.a.s of co^{ρ} and $p^{\bar{\rho}}$ *coincide*. This is true in the general case.

Theorem 2. *There exists a 1-1 correspondence between the vertices co^{ρ} (16) of the region of outer consonant approximations of b generated by a permutation ρ of the singletons, and the vertices $p^{\bar{\rho}}$ (8) of the polytope $\mathcal{P}[b]$ of probabilities consistent with b (all of which are associated with permutations of singletons), such that*

$$p^{\bar{\rho}}(x_{\bar{\rho}(i)}) = m_{co^{\rho}}(\{x_{\rho(1)}, ..., x_{\rho(n-i+1)}\}) \tag{17}$$

i.e. their b.p.a.s on $\{S^{\rho}_n, ..., S^{\rho}_1\}$, $\{x_{\bar{\rho}(1)}, ..., x_{\bar{\rho}(n)}\}$ respectively coincide.

Proof. It suffices to show that, as $\bar{\rho}(i) = \rho(n - i + 1)$,

$$\begin{aligned} p^{\bar{\rho}}(x_{\bar{\rho}(i)}) &= \sum_{A \ni x_{\bar{\rho}(i)}, A \not\ni x_{\bar{\rho}(j)} \forall j < i} m_b(A) = \sum_{A \ni x_{\rho(n-i+1)}, A \not\ni x_{\rho(j)} \forall j > n-i+1} m_b(A) \\ &= \sum_{\substack{A \subseteq \{x_{\rho(1)}, ..., x_{\rho(n-i+1)}\}, \\ A \not\subset \{x_{\rho(1)}, ..., x_{\rho(n-i)}\}}} m_b(A) = m_{co^{\rho}}(\{x_{\rho(1)}, ..., x_{\rho(n-i+1)}\}) = m_{co^{\rho}}(S^{\rho}_{n-i+1}). \end{aligned}$$

5 Conclusions

Belief functions possess a strong credal semantics in terms of convex sets of probability distributions or consistent probabilities, for whose values on all events belief values provide lower bounds. These probabilities can also be seen as more committed or "inner" Bayesian approximations of the original b.f.
In this paper we proved a more compact form of the polytope of consistent probabilities, as the latter has $n!$ (candidate) vertices each corresponding to a different permutation of the elements of the domain. This unveils an interesting link with the vertices of the polytopes of outer consonant approximations also generated by permutations of singletons both in terms of their analytical expression and their convex geometry.

References

1. Mercier, D., Denoeux, T., Masson, M.: Refined sensor tuning in the belief function framework using contextual discounting. In: IPMU. (2006)

2. Demotier, S., Schon, W., Denoeux, T.: Risk assessment based on weak information using belief functions: a case study in water treatment. IEEE Transactions on Systems, Man and Cybernetics, Part C **36(3)** (May 2006) 382– 396
3. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
4. Dubois, D., Prade, H.: Possibility theory. Plenum Press, New York (1988)
5. Smets, P.: Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning **9** (1993) 1–35
6. Denoeux, T.: A new justification of the unnormalized Dempsters rule of combination from the Least Commitment Principle. In: Proceedings of FLAIRS'08, Special Track on Uncertaint Reasoning. (2008)
7. Yager, R.R.: The entailment principle Dempster-Shafer granules. International Journal of Intelligent Systems **1** (1986) 247–262
8. Dubois, D., Prade, H.: A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. International Journal of General Systems **12** (1986) 193–226
9. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. Mathematical Social Sciences **17** (1989) 263–283
10. Ha, V., Haddawy, P.: Geometric foundations for interval-based probabilities. In Cohn, A.G., Schubert, L., Shapiro, S.C., eds.: KR'98: Principles of Knowledge Representation and Reasoning. Morgan Kaufmann (1998) 582–593
11. Smets, P.: The nature of the unnormalized beliefs encountered in the transferable belief model. In: Proceedings of the 8th Annual Conference on Uncertainty in Artificial Intelligence (UAI-92), San Mateo, CA, Morgan Kaufmann (1992) 292–29
12. Dubois, D., Grabisch, M., Prade, H., Smets, P.: Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: the assessment of the value of a candidate. Intern. J. Intell. Systems (2001)
13. Snow, P.: The vulnerability of the transferable belief model to dutch books. Artificial Intelligence **105** (1998) 345–354
14. Smets, P., Kennes, R.: The transferable belief model. AI **66**(2) (1994) 191–234
15. Melkonyan, T., Chambers, R.: Degree of imprecision: Geometric and algebraic approaches. International Journal of Approximate Reasoning (2006)
16. Cozman, F.G.: Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. Journal of Computational and Graphical Statistics **8(4)** (1999) 824–838
17. Seidenfeld, T., Wasserman, L.: Dilation for convex sets of probabilities. Annals of Statistics **21** (1993) 1139–1154
18. Baroni, P.: Extending consonant approximations to capacities. In: Proceedings of IPMU. (2004) 1127–1134
19. Dubois, D., Prade, H.: Consonant approximations of belief functions. International Journal of Approximate Reasoning **4** (1990) 419–449
20. Denoeux, T.: Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. Artificial Intelligence (2007)
21. Cuzzolin, F.: A geometric approach to the theory of evidence. IEEE Trans. Systems, Man, and Cybernetics C (in press) **38**(3) (May 2008)
22. Dubois, D., Prade, H., Smets, P.: New semantics for quantitative possibility theory. In: ISIPTA. (2001) 152–161
23. Joslyn, C.: Towards an empirical semantics of possibility through maximum uncertainty. In Lowen, R., Roubens, M., eds.: Proc. IFSA 1991. (1991) 86–89
24. Cuzzolin, F.: The geometry of consonant belief functions: Simplicial complexes of possibility measures. Fuzzy Sets and Systems (under review) (2007)