Extreme points of credal sets generated by 2-alternating capacities

Enrique Miranda *, Inés Couso and Pedro Gil

Department of Statistics and Operations Research

University of Oviedo

C/ Calvo Sotelo, s/n 33007 Oviedo, Spain

Abstract

The characterization of the extreme points constitutes a crucial issue in the investigation of convex sets of probabilities, not only from a purely theoretical point of view, but also as a tool in the management of imprecise information. In this respect, different authors have found an interesting relation between the extreme points of the class of probability measures dominated by a second order alternating Choquet capacity and the permutations of the elements in the referential. However, they have all restricted their work to the case of a finite referential space. In an infinite setting, some technical complications arise and they have to be carefully treated. In this paper, we extend the mentioned result to the more general case of separable metric spaces. Furthermore, we derive some interesting topological properties about the convex sets of probabilities here investigated. Finally, a closer look to the case of possibility measures is given: for them, we prove that the number of extreme points can be reduced even in the finite case.

Keywords. Upper probabilities, Choquet capacities, possibility measures, weak convergence, linear topological spaces, credal sets.

1 Introduction

The Theory of Imprecise Probabilities comprises different mathematical models such as upper and lower probabilities, upper and lower expectations or sets of probability measures. A special case of upper probabilities are those called *2-*

*Corresponding author. Tel: (+34) 985102958, Fax: (+34) 985103354. E-mail addresses: emiranda@correo.uniovi.es (E. Miranda), couso@pinon.ccu.uniovi.es (I. Couso), pedro@pinon.ccu.uniovi.es (P. Gil)

alternating capacities [5] or *submodular* set functions[11]. They play an important role as upper bounds of sets of probability measures (see, for instance [12]). This means that they are *coherent* [25] upper probabilities. A Choquet capacity does not need to be 2- alternating to satisfy the property of coherence. Nevertheless, this property leads to important mathematical simplifications, e.g. in the calculus of the formulae of upper expectations and conditional upper probabilities. Thus we can find detailed studies in the literature concerning this type of Choquet capacities.

On the other hand, several authors have studied the properties of convex sets of probability measures, or *credal sets* [15]. They are well suited to provide robustness in certain situations [25, 28]. Moreover, they have been successfully applied in different contexts, such as information theory [1] or classification [30]. In particular, Walley [25] has showed that a coherent upper probability μ contains the same information as the credal set $M(\mu)$, the set of probability measures dominated by μ. With a certain abuse of notation, we will refer to $M(\mu)$ as the credal set *generated* by μ.

The set of extreme points of the credal set generated by a 2-alternating capacity has been studied and characterized for the case of finite referential sets (see [4, 9, 23] for detailed discussions). In these works, the authors find a correspondence between the set of extreme points and the set of permutations of the elements in the referential set.

In this paper, we extend this useful result to the case of separable metric referential spaces. This is achieved after some intermediate work on the structure of the credal set; in particular, we are going to include it in a linear topological space with some properties. This profile can then be applied in a number or different contexts, such as random sets [9] or game theory [23].

The paper is organized as follows: in Section 2, we give a number of preliminary concepts from the theory of imprecise probabilities. In Section 3, we study the topological properties and the structure of the set of probabilities dominated by a 2-alternating capacity. The results are applied in Section 4, where we characterize the extreme points of this set. In Section 5, we give some particular results for possibility measures. Finally, in Section 6 some additional comments are given.

This paper is an updated and extended version of a paper ([18]) presented at AGOP'01, the International Summer School on Aggregation Operators and its Applications.

2 Preliminary concepts

Since they were first introduced by Choquet in his seminal work "Theory of capacities" ([5]), upper and lower probability models have been used as an alternative to the probability theory. They are based in the concept of *capacity*.

Definition 2.1. *[4] Given a referential set X, a* capacity *is a function μ defined on some class of subsets $\mathcal{A}$ of X satisfying the following properties:*

- $\mu(\emptyset) = 0,\ \mu(X) = 1.$
- $A \subset B \Rightarrow \mu(A) \leq \mu(B).$

A capacity is said to be *subadditive* when $\mu(A \cup B) \leq \mu(A) + \mu(B)\ \forall A, B \in \mathcal{A}$. Throughout this paper, we will work with subadditive capacities over a σ-algebra $\mathcal{A}$ of subsets of X. An analogous study could be done with superadditive capacities (those which reverse the inequality). A particular type of capacities will also be used:

Definition 2.2. *A capacity μ is called* outer continuous *if given a decreasing sequence $(A_n)_n \subset \mathcal{A}$, $\mu(\cap_n A_n) = \lim_n \mu(A_n)$.*

Capacities are in many cases too weak to deal with the uncertainty. Because of this fact, the concept of n-alternancy arises:

Definition 2.3. *[21] A subadditive capacity μ is* n-alternating *if $\mu(A_1 \cap \cdots \cap A_n) \leq \sum_{i=1}^n \mu(A_i) - \sum_i \sum_j \mu(A_i \cup A_j) + \sum_i \sum_j \sum_k \mu(A_i \cup A_j \cup A_k) - \cdots + (-1)^{n+1} \mu(\cup_{i=1}^n A_i)$ for all $A_1, \ldots, A_n \in \mathcal{A}$.*

A n-alternating capacity is also called Choquet capacity of order n. The dual condition for superadditive capacities is called *n-monotonicity*.

Choquet capacities of order 2 constitute a very powerful and manageable model when dealing with uncertainty: from the behavioural point of view, they provide some nice properties for conditioning or extending our information ([24, 25]). But they have also been applied in the context of robust statistics ([12]) or game theory ([23]).

When a capacity is n-alternating (resp., n-monotone) for every n, it is called ∞-alternating capacity (resp., ∞-monotone capacity). In the finite case, they are called plausibility and belief functions, respectively ([9, 21]), and they fulfill very interesting properties: for instance, they are characterized through a probability mass function m called Möbius inverse. Unfortunately, the Möbius inverse does not exist when we work with an infinite referential, and therefore the results cannot be immediately extended for that case.

An important particular case of ∞-alternating capacities are those called *possibility measures*. They play an important role in modeling linguistic information (see [8, 27, 29]).

Definition 2.4. *A* possibility measure *is a set function* $\mu : \mathcal{A} \to [0, 1]$ *satisfying that there is a function* $\pi : X \to [0, 1]$ *s.t.* $\mu(A) = \sup\{\pi(\omega) : \omega \in A\}$ $\forall A \in \mathcal{A}$. π *is called* possibility distribution *of* μ. *A possibility measure is called* normal *when* $\mu(X) = 1$.

We will assume throughout this paper that μ is normal when it is a possibility measure; otherwise, the set $M(\mu)$ should be empty.

Remark 2.1. *Possibility measures are in particular* maxitive *capacities, that is, they satisfy* $\mu(A \cup B) = \max\{\mu(A), \mu(B)\}$ $\forall A, B \in \mathcal{A}$. *These two concepts are equivalent in the case of a finite referential, but not in general. On the other hand, any maxitive measure is in particular an* ∞*-alternating capacity, as it is proven in [20].*

The next two sections are devoted to the study of the credal set generated by a 2-alternating capacity. In section 5, we will pay attention to the particular case of possibility measures.

3 The set $M(\mu)$

We are going to consider throughout this paper that (X, d) is a separable metric space, i.e., that it possesses a countable set $\mathcal{D} = \{x_n : n \in \mathbb{N}\}$ dense on $\tau(d)$. As a consequence, its induced topology $\tau(d)$ will have a countable base, that we will denote $\{B_n : n \in \mathbb{N}\}$, given by the sets $\{B(x_i, q_i) : q_i \in \mathbb{Q}, x_i \in \mathcal{D}\}$. Then, the Borel σ-algebra β_X induced by d is equal to $\sigma(B_1, \dots)$.

Let us denote by $M(\mu) = \{P \text{ probability} : P(A) \leq \mu(A) \ \forall A \in \beta_X\}$ the set of probability measures *dominated* by a given capacity μ. We are going to refer to this set as the *credal set generated by* μ.

In this section, we are going to take a closer look at this convex set; in particular, we are going to study some of its topological properties and we are going to give it an adequate structure. This is essential in order to be able to characterize the extreme points of $M(\mu)$, something that will be done in section 4. Walley ([25]), in the context of *linear previsions*, which correspond to finitely additive probabilities, has studied the credal set generated by a coherent upper probability. He has proven the applicability of some separation theorems under the weak-* topology, and has shown that in general the class of countably additive probabilities will not be weak-* compact. We will go deeply into this matter for our framework of countably additive probabilities dominated by 2-alternating capacities. In fact, throughout this section we are going to consider arbitrary capacities, not necessarily 2-alternating.

When X is finite, the study of $M(\mu)$ benefits from a lot of simplifications: under the natural topology induced by the Euclidean distance, the set $M(\mu)$ is the closed convex hull of its set of extreme points. In the general case, we are going to apply the following corollary of Krein-Milman's theorem [13]:

Theorem 3.1. *Let $\mathcal{I}$ be a compact subset of a locally convex Hausdorff linear topological space $\mathcal{E}$, and suppose that the closed convex hull $\overline{Conv(\mathcal{I})}$ of $\mathcal{I}$ is compact. Then, all the extreme points of $\overline{Conv(\mathcal{I})}$ belong to $\mathcal{I}$.*

To be able to apply this theorem, we need to include the probability distributions on β_X in a linear topological space with some properties. This will be achieved at the end of the section, in theorems 3.5 and 3.7. We also intend to work with the topology of the weak convergence of probability distributions, as this topology is widely used in practice: Kelley and Namioka ([13]) use it as the natural topology in the context of linear topological spaces, and in the particular case of $(X, \beta_X) = (\mathbb{R}, \beta_{\mathbb{R}})$ it models the convergence of distribution functions. Other interesting topologies would be the *weak-** topology, the *strong* topology or the *Mackey* topology (see a complete study in [13]).

We are going to recall first the behaviour of the weak topology, for it will help to understand our construction of a linear topological space containing the class $\mathcal{P}$ of countably additive probability measures.

Definition 3.1. *Let (X, d) be a separable metric space, and consider $(P_n)_n, P$ probability measures defined on (X, β_X). $(P_n)_n$ is said to* converge weakly *to P when $\lim_n \int f dP_n = \int f dP$ for all bounded, continuous real function f on X. We will denote it $P_n \Rightarrow P$.*

We are only going to outline here some basic definitions necessary to follow the rest of the paper; see [2, 3] for a detailed study of the weak convergence and its implications. The following result ([2]) will be used later:

Proposition 3.2. *Let $\mathcal{U}$ be a subclass of β_X such that (i) it is closed under the formation of finite intersections and (ii) each open set in β_X is a finite or countable union of elements of $\mathcal{U}$. If $P_n(A) \to P(A)$ for every A in $\mathcal{U}$, then $P_n \Rightarrow P$.*

The weak topology is the topology on $\mathcal{P}$ taking as a neighborhood system of P the sets $\{Q : Q(G_i) > P(G_i) - \epsilon,\ i = 1, \ldots, k\}$, with G_i open and $\epsilon > 0$. Then, $P_n \Rightarrow P$ if and only if the sequence $(P_n)_n$ converges to P respect to the weak topology. Moreover, this topology is Hausdorff, so any pair of probability measures can be separated through disjoint open sets. Let us also introduce an auxiliary notion, to be used later:

Definition 3.2. *[2] A family Λ of probability measures on (X, β_X) is said to be* tight *if for every positive ϵ there exists a compact set K such that $P(K) > 1 - \epsilon$ for all P in Λ.*

It is clear that if X is a compact metric space (which is in particular separable), any family of probability measures will be tight.

Proposition 3.3. *[2] Consider Λ a tight family of probability measures on (X, β_X). Then, its closure respect to the weak topology is compact.*

Remark 3.1. *In particular, if (X, d) is a compact metric space, any closed family Ψ of probability measures on β_X is weakly compact: Ψ will be tight because X is compact, and then we apply the previous proposition.*

We are ready now to introduce the main result of this section. We are going to define a linear topological space $\mathcal{E}$ satisfying the conditions of theorem 3.1, having the class of probability measures on (X, β_X) as a subset and satisfying that the relative topology on this subset is the weak topology. The construction, that we will apply in this paper to the study of the extreme points of the credal set $M(\mu)$, could also be used in other contexts where the structure of a linear topological space is also essential; hence, its scope goes beyond the framework of this paper.

Let us denote by $\mathcal{P}$ the class of probability measures on (X, β_X), and consider the space $\mathcal{E} := \{\sum_{i=1}^{n} \lambda_i P_i : \lambda_i \in \mathbb{R}, P_i \in \mathcal{P}\}$. This is a linear space over the scalar field $\mathbb{R}$. Consider also the space $\mathcal{F}$ of bounded continuous real functions on (X, d). This is also a vectorial space over the scalar field $\mathbb{R}$.

Remark 3.2. *Note that any element $Q = \sum_{i=1}^{m} \lambda_i Q_i \in \mathcal{E}$ can be expressed in the form $Q = \alpha_1 Q_1' - \alpha_2 Q_2'$, with $\alpha_1, \alpha_2 \geq 0$, Q_1', Q_2' probabilities: if we consider $I_1 = \{i : \lambda_i > 0\}$, then $\sum_{i \in I_1} \lambda_i Q_i = (\sum_{i \in I_1} \lambda_i)(\sum_{i \in I_1} \frac{\lambda_i}{(\sum_{i \in I_1} \lambda_i)} Q_i) = (\sum_{i \in I_1} \lambda_i) Q_1'$. The measure Q_1' is a linear convex combination of probabilities, which is also a probability. Similarly, if we take $I_2 = \{i : \lambda_i < 0\}$, we can see that $\sum_{i \in I_2} \lambda_i Q_i = (\sum_{i \in I_2} \lambda_i) Q_2'$, with $Q_2' = (\sum_{i \in I_2} \frac{\lambda_i}{(\sum_{i \in I_2} \lambda_i)} Q_i)$ a probability.*

In order to provide $\mathcal{E}$ with a linear topological space structure, we are going to use the concept of *pairing* [13] of vectorial spaces. A pairing is an ordered pair of linear spaces over the same scalar field K, together with a bilinear functional $B : \mathcal{E} \times \mathcal{F} \to K$ on the product. We will denote $< x, f >= B(x, f)$.

Then, given a pairing $< \mathcal{E}, \mathcal{F} >$, we can define on $\mathcal{E}$ the $w(\mathcal{E}, \mathcal{F})$ topology as follows: consider the space $K^{\mathcal{F}} = \prod_{f \in \mathcal{F}} K$, with the product topology. We can establish a correspondence between the elements $x \in \mathcal{E}$ and the elements of $K^{\mathcal{F}}$ through the bilinear functional B. That is, we can define a map T such that $T(x)$ is the element $\prod_{f \in \mathcal{F}} < x, f >$ in $K^{\mathcal{F}}$. A subset U of $\mathcal{E}$ belongs to $w(\mathcal{E}, \mathcal{F})$ if and only if it is $U = T^{-1}(V)$ for some open set $V \subset K^{\mathcal{F}}$ in the product topology.

The $w(\mathcal{E}, \mathcal{F})$-topology satisfies the following property:

Theorem 3.4. *[13] Let $\mathcal{E}$ and $\mathcal{F}$ be paired linear spaces. Then:*

1. *The space $\mathcal{E}$ together with the topology $w(\mathcal{E},\mathcal{F})$ is a locally convex linear topological space, and it is Hausdorff if and only for every $0 \neq x \in \mathcal{E}$, there exists $f \in \mathcal{F}$ s.t. $< x, f > \neq 0$.*

2. *A sequence $\{x_n\}_n \subset \mathcal{E}$ converges to $x \in \mathcal{E}$ relative to the topology $w(\mathcal{E},\mathcal{F})$ if and only if $(< x_n, f >)_n$ converges to $< x, f >$ for each $f \in \mathcal{F}$.*

Let us apply these results in our context:

Theorem 3.5. *Consider the functional $B : \mathcal{E} \times \mathcal{F} \to \mathbb{R}$ by $B(\sum_{i=1}^n \lambda_i P_i, f) = \sum_{i=1}^n \lambda_i \int f dP_i$. Then,*

1. *B is well-defined.*

2. *B is a bilinear functional.*

3. *$\mathcal{E}$ with the $w(\mathcal{E},\mathcal{F})$ topology is a locally convex and Hausdorff linear topological space.*

4. *The relative topology on the subset $\mathcal{P}$ of probability measures coincides with the weak topology.*

Proof: We start by proving the first statement. Assume $Q = \sum_{i=1}^n \lambda_i Q_i = \sum_{i=1}^m \lambda'_i Q'_i$, and let us show that for any $f \in \mathcal{F}$, we have $\sum_{i=1}^n \lambda_i \int f dQ_i = \sum_{i=1}^m \lambda'_i \int f dQ'_i$. It is clear that it holds when $f = I_A$ for some $A \in \beta_X$: both terms are equal to $Q(A)$. Consequently, we also have the equality for simple β_X-measurable functions. Now, any $f \in \mathcal{F}$ can be uniformly approximated by a sequence $(f_k)_k$ of simple β_X-measurable functions, and then

$$\sum_{i=1}^n \lambda_i \int f dQ_i = \sum_{i=1}^n \lambda_i \lim_k \int f_k dQ_i = \lim_k \sum_{i=1}^n \lambda_i \int f_k dQ_i = \lim_k \sum_{i=1}^m \lambda'_i \int f_k dQ'_i = \sum_{i=1}^m \lambda'_i \lim_k \int f_k dQ'_i = \sum_{i=1}^m \lambda'_i \int f dQ'_i.$$

Therefore, B is well-defined.

Let us prove now the second statement. There are two parts:

- Given $\gamma_1, \gamma_2 \in \mathbb{R}, Q_1 = \sum_{i=1}^n \lambda_i P_i, Q_2 = \sum_{j=1}^m \tau_j P'_j \in \mathcal{E}, f \in \mathcal{F}$, it is

$$B(\gamma_1 Q_1 + \gamma_2 Q_2, f) = B(\sum_{i=1}^n \gamma_1 \lambda_i P_i + \sum_{j=1}^m \gamma_2 \tau_j P'_j, f) =$$
$$\sum_{i=1}^n \gamma_1 \lambda_i \int f dP_i + \sum_{j=1}^m \gamma_2 \tau_j \int f dP'_j = \gamma_1 (\sum_{i=1}^n \lambda_i \int f dP_i) +$$
$$\gamma_2 (\sum_{j=1}^m \tau_j \int f dP'_j) = \gamma_1 B(Q_1, f) + \gamma_2 B(Q_2, f).$$

- Given $\gamma_1, \gamma_2 \in \mathbb{R}, Q = \sum_{i=1}^n \lambda_i Q_i \in \mathcal{E}, f_1, f_2 \in \mathcal{F}$, it is

$$B(Q, \gamma_1 f_1 + \gamma_2 f_2) = \sum_{i=1}^n \lambda_i \int (\gamma_1 f_1 + \gamma_2 f_2) dQ_i =$$
$$\sum_{i=1}^n \lambda_i \gamma_1 \int f_1 dQ_i + \sum_{i=1}^n \lambda_i \gamma_2 \int f_2 dQ_i = \gamma_1 B(Q, f_1) + \gamma_2 B(Q, f_2).$$

The first part of the third statement is a consequence of theorem 3.4. To see that $\mathcal{E}$ is Hausdorff, take $Q = \alpha_1 Q_1 - \alpha_2 Q_2 \neq 0$. There are two alternatives: either it is $\alpha_1 \neq \alpha_2$, and then $f = I_X$ satisfies $B(Q, f) = \alpha_1 - \alpha_2 \neq 0$; or either we have $\alpha_1 = \alpha_2$. In that case, it must be $Q_1 \neq Q_2$. But then there exists $f \in \mathcal{F}$ s.t. $\int f dQ_1 \neq \int f dQ_2 \Rightarrow \alpha_1 \int f dQ_1 - \alpha_2 \int f dQ_2 \neq 0 \Rightarrow B(Q, f) \neq 0$. Therefore, the pairing $< \mathcal{E}, \mathcal{F} >$ satisfies theorem 3.4 and $\mathcal{E}$ is Hausdorff.

We finally prove the fourth and last statement. Consider the subspace of $\mathcal{E}$ given by the class $\mathcal{P}$ of probability measures on (X, β_X), with the relative topology $w(\mathcal{E}, \mathcal{F}) \cap \mathcal{P}$. By theorem 3.4, a sequence $(P_n)_n$ converges to P respect to the relative topology if and only if for every $f \in \mathcal{F}$, $< P_n, f > \to < P, f >$. But this is equivalent to the convergence of $\int f dP_n$ to $\int f dP$ for every bounded continuous real-valued function on (X, d), that is, the weak convergence of P_n to P. This means that the relative topology on $\mathcal{P}$ coincides with the weak topology.■

Next, we are going to show that the set $\mathcal{P}$ of probability measures is a closed subset of $\mathcal{E}$ with the $w(\mathcal{E}, \mathcal{F})$ topology. First, we need to prove an auxiliary result.

Lemma 3.6. *For every $A \in \mathcal{Q}(B_1, \dots)$, there is decreasing sequence $(A_n)_n \subseteq \tau(d)$ s.t. $A = \cap_n A_n$.*

Proof: As $\mathcal{Q}(B_1, \dots) = \cup_m \mathcal{Q}(B_1, \dots, B_m)$, it suffices to prove the result for $A \in \mathcal{Q}(B_1, \dots, B_m)$ for some arbitrary m.

- Consider first $A = C_1 \cap \dots \cap C_m$, where $C_i \in \{B_i, B_i^c\}$ $\forall i = 1, \dots, m$. Recall that $B_i = B(x_i; q_i)$ for some $x_i \in X, q_i \in \mathbb{Q}$. Given a sequence $(r_n)_n \uparrow q_i$, with $r_n < q_i$ $\forall n$, we can easily check the equality $B_i^c = \cap_n \overline{B}(x_i; q_n)^c$. Then, each C_i is the limit of a decreasing sequence $(A_n^i)_n$ of open sets for every $i = 1, \dots, m$, and then so is A, taking the sequence $(A_n)_n$ given by $A_n = A_n^1 \cap \dots \cap A_n^m$.
- Given $B \in Q(B_1, \dots, B_m)$, it is a finite union of elements of the previous type; that is, $B = A_1 \cup \dots \cup A_l$, with $A_i = C_1^i \cap \dots C_m^i$ for all i, $C_j^i \in \{B_j, B_j^c\}$ for every i, j and $A_p \cap A_q = \emptyset$ $\forall p \neq q$. Each A_i is, by the previous point, the limit of a decreasing sequence $(A_n^i)_n$ of open sets; if we consider the sequence $(B_n)_n = (A_n^1 \cup \dots \cup A_n^l)_n$, we can easily check that it is decreasing and that its limit is $\cap_n B_n = \cap_n (A_n^1 \cup \dots \cup A_n^l) = (\cap_n A_n^1) \cup \dots \cup (\cap_n A_n^l) = A_1 \cup \dots \cup A_l = B$. ■

Equivalently, any element of the algebra $\mathcal{Q}(B_1, \dots)$ will also be the limit of an increasing sequence of closed sets. Now we are ready to check the closeness of $\mathcal{P}$ for the $w(\mathcal{E}, \mathcal{F})$-topology.

Theorem 3.7. *$\mathcal{P}$ is a closed subset of $\mathcal{E}$.*

Proof: Consider a sequence $(P_n)_n$ of probability measures converging to $Q = \sum_{i=1}^m \lambda_i Q_i \in \mathcal{E}$ with respect to the topology $w(\mathcal{E}, \mathcal{F})$. Any element of $\mathcal{E}$ is σ-additive and satisfies $Q(\emptyset) = 0$. Besides, the $w(\mathcal{E}, \mathcal{F})$-convergence implies that $\lim_n P_n(X) = 1 = Q(X)$, because $I_X \in \mathcal{F}$. It only remains to show that Q is monotone, that is, that $A \subseteq B \Rightarrow Q(A) \leq Q(B)$. Because of the additivity, it is equivalent to show that Q is non-negative.

The proof is made in three steps: first, we show that Q is non-negative in the closed sets; then, we prove it for the elements of $\mathcal{Q}(B_1, \dots)$; and finally we deduce that Q is non-negative on β_X.

- Let C be a closed subset of X. Then, by ([2, th.1.2]), for every $\delta > 0$

there exists $f_\delta \in F$ s.t. $f_\delta(x) = 1\ \forall x \in C, 0 \le f_\delta(x) \le 1$ if $d(x, C) \le \delta$, and $f_\delta(x) = 0$ if $d(x, C) > \delta$. These functions satisfy that $f_\delta(x) \to I_C(x)$ for every x when $\delta \to 0$.

By theorem 3.4, we have that $\lim_n B(f_\delta, P_n) = \lim_n \int f_\delta dP_n = B(f_\delta, Q) = \sum_{i=1}^m \lambda_i \int f_\delta dQ_i$. If we make $\delta \to 0$ and apply the bounded convergence theorem, we obtain

$$\lim_{\delta\to 0} \sum_{i=1}^m \lambda_i \int f_\delta dQ_i = \sum_{i=1}^m \lambda_i \lim_{\delta\to 0} \int f_\delta dQ_i = \sum_{i=1}^m \lambda_i Q_i(C) = Q(C).$$

Moreover, $P_n(C) \le \int f_\delta dP_n\ \forall \delta > 0,\ n \in \mathbb{N}$; hence, $\limsup_n P_n(C) \le \limsup_n \int f_\delta dP_n = \lim_n \int f_\delta dP_n = \sum_{i=1}^m \lambda_i \int f_\delta dQ_i\ \forall \delta > 0$. Consequently, $\limsup_n P_n(C) \le \lim_{\delta\to 0} \sum_{i=1}^m \lambda_i \int f_\delta dQ_i = Q(C)$, and we deduce that $Q(C) \ge 0$.

- We have proven in lemma 3.6 that every element of $Q(B_1, \dots)$ is the limit of a decreasing sequence of open sets. Equivalently, it is the limit of an increasing sequence of closed sets, i.e., given $A \in \mathcal{Q}(B_1, \dots)$, it is $A = \cup_n C_n$, with $C_i \subseteq C_{i+1}$ closed $\forall i = 1, 2, \dots$. Then,

$$Q(A) = \sum_{i=1}^m \lambda_i Q_i(A) = \sum_{i=1}^m \lambda_i Q_i(\cup_n C_n) = \sum_{i=1}^n \lambda_i \lim_n Q_i(C_n) = \lim_n \sum_{i=1}^m \lambda_i Q_i(C_n) = \lim_n Q(C_n) \ge 0$$

  (the limit of a finite sum of convergent sequences is the sum of the limits).

  We see then that $Q \ge 0$ in the algebra $\mathcal{Q}(B_1, \dots)$.

- Define $\mathcal{H} = \{A \in \beta_X : Q(A) \ge 0\}$. We have proven the inclusion $\mathcal{Q}(B_1, \dots) \subseteq \mathcal{H}$. Let us also show that $\mathcal{H}$ is a monotone class: given $(A_n)_n \uparrow\subseteq \mathcal{H}$, it is $Q(\cup_n A_n) = \lim_n Q(A_n) \ge 0$, and the same applies to decreasing sequences. Therefore, $\mathcal{H}$ is a monotone class, whence $\beta_X = \mathcal{H}$.

We deduce that $Q(A) \ge 0\ \forall A \in \beta_X,\ Q(X) = 1,\ Q(\emptyset) = 0$, and that Q is σ-additive. Therefore, Q is a probability. ■

We conclude that a set of probability measures $A \subset \mathcal{P}$ will be closed respect to the weak topology if and only if it is closed respect to the $w(\mathcal{E}, \mathcal{F})$ topology on $\mathcal{E}$, and the closure of A will coincide on $\mathcal{P}$ and on $\mathcal{E}$.

Moreover, if $A \subset \mathcal{P}$ is compact respect to the weak topology, any family of closed subsets of A with the F.I.P. (finite intersection property) will have a (global) non-empty intersection. But the weak topology is Hausdorff, whence A is weakly closed ($\Leftrightarrow w(\mathcal{E}, \mathcal{F})$-closed) and the closed subsets of A will be the $w(\mathcal{E}, \mathcal{F})$- closed subsets. This means that any family of $w(\mathcal{E}, \mathcal{F})$-closed subsets of A with the F.I.P. will have a non-empty intersection, and hence A is $w(\mathcal{E}, \mathcal{F})$-compact.

This will allow us to restrict our work to the class $\mathcal{P}$ and to the well-known behaviour of the weak topology, and ultimately apply theorem 3.1, which characterizes the extreme points of convex compact sets on linear topological spaces.

Let us remark again that by this construction we are providing the probability measures on a separable metric space with a mathematical structure which is widely used in the field of functional analysis and that will allow us to look at some problems from probability theory with a different perspective.

Now, we are going to study under which conditions the class $M(\mu)$ is weakly closed and compact. Our results, combined with the previous remarks, will allow us to give a version of Krein-Milman's theorem applicable to our problem. We also want to stress that the following results are valid for arbitrary capacities, *not necessarily 2-alternating*. This property will only become important in our next section, when we characterize the set of extreme points. In particular, this means that our work in this section can be regarded as dual from Walley's study ([25]) for the weak-* topology.

Let us give a characterization of $M(\mu)$:

Proposition 3.8. *Consider (X, d) a separable space, and $\{B_n\}_n$ a countable base. Let P be a probability and μ a capacity defined on β_X. Suppose that $P(A) \leq \mu(A)$ for every A open. If μ is outer continuous, then $P \in M(\mu)$.*

Proof: Consider the set $\mathcal{H} = \{A \in \beta_X : P(A) \leq \mu(A)\}$. By hypothesis, $\mathcal{H}$ contains the open sets. From lemma 3.6, we can deduce that $\mathcal{H}$ contains the algebra $\mathcal{Q}(B_1, ...)$: given $B \in \mathcal{Q}(B_1, ...)$, there exists a sequence of open sets

$(A_n)_n \downarrow B$. Then, $P(B) = \lim_n P(A_n) \leq \lim_n \mu(A_n) = \mu(B)$, because μ is outer-continuous.

Let us check that $\mathcal{H}$ is also a monotone class:

- Take $(A_n)_n \uparrow \subset \mathcal{H}$. Then, $P(A_n) \leq \mu(A_n) \leq \mu(\cup_n A_n)\ \forall n \Rightarrow P(\cup_n A_n) = \sup_n P(A_n) \leq \mu(\cup_n A_n) \Rightarrow \cup_n A_n \in \mathcal{H}$.
- Consider now a sequence $(A_n)_n \downarrow \subset \mathcal{H}$. We have $P(A_n) \leq \mu(A_n)\ \forall n$; then, $P(\cap_n A_n) = \lim_n P(A_n) \leq \lim_n \mu(A_n) = \mu(\cap_n A_n)$, for the outer continuity. Hence, $\cap_n A_n \in \mathcal{H}$.

Therefore, $\mathcal{H}$ is a monotone class and it contains $\mathcal{Q}(B_1, ...)$, so $\mathcal{H} = \sigma(B_1, ...)$, and then $P \in M(\mu)$. ■

The thesis of this last proposition is not true in general when we drop the outer continuity condition:

Example 3.1. *Consider $\beta_{\mathbb{R}}$ the Borel σ-algebra on the real line, and let P denote the degenerate distribution on some irrational point, $x \in \mathbb{R} \setminus \mathbb{Q}$, i.e., $P(A) = 1,\ \forall A \supset \{x\}$, and $P(A) = 0$ otherwise. Consider the set function μ on $\beta_{\mathbb{R}}$ given by $\mu(\{x\}) = 0$, $\mu(A) = 1\ \forall \emptyset \neq A \neq \{x\}$.*

- *It is clear that μ is a capacity.*
- *It is not outer continuous: it suffices to take $A_n = (x - \frac{1}{n}, x + \frac{1}{n})$. Then, it will be $\mu(A_n) = 1\ \forall n$ and $\mu(\cap_n A_n) = \mu(\{x\}) = 0$.*

The set $\{x\}$ is not open on $\beta_{\mathbb{R}}$; hence, $\mu(A) = 1\ \forall A$ open, and it dominates P on the open sets trivially. However, it is $P(\{x\}) = 1 > \mu(\{x\})$, so $P \notin M(\mu)$.

Therefore, the outer continuity, which is trivially satisfied in the finite case, arises here as an interesting condition in order to characterize the dominance of a set of probability measures by a capacity [1].

In general, the class $M(\mu)$ is not going to be (weakly) closed. Let us see an example.

[1] Some definitions of capacity impose the outer continuity on closed or compact sets ([5, 16]). In this paper, we have avoided that to make our work as general as possible.

Example 3.2. *Consider the same capacity μ from example 3.1. Let $(Y_n)_n$ be a sequence of random variables, where Y_n is constant on $x+\frac{1}{n}$ for every n, and take Y a random variable constant on x. Then, the sequence $(Y_n)_n$ converges almost surely to Y, whence $(P_{Y_n})_n$ converges weakly to P_Y. However, P_{Y_n} belongs to $M(\mu)$ for every n, whereas P_Y does not. Hence, $M(\mu)$ is not closed. As any compact set in a Hausdorff space (such as $\mathcal{P}$ with the weak topology) is closed, $M(\mu)$ is not compact either.*

We also have the following result:

Proposition 3.9. *Let μ be an outer-continuous capacity on β_X. Then, $M(\mu)$ is closed. If in particular (X, d) is a compact metric space, then $M(\mu)$ is compact.*

Proof: Take a sequence of probability distributions $Q_n \Rightarrow Q$, with $Q_n \in M(\mu)$ $\forall n$. Then, by Portmanteau's theorem, we have that $\liminf_n Q_n(A) \geq Q(A)$ for every A open, which implies that $Q(A) \leq \mu(A)$ $\forall A$ open. Applying proposition 3.8, we conclude that $Q \in M(\mu)$. Therefore, this set is closed.

The second part of the result follows from the fact that when (X, d) is a compact metric space, any closed family of probability measures is compact, as we showed in remark 3.1. ∎

The credal set $M(\mu)$ is not closed in general without the outer continuity condition, as example 3.2 shows. When $M(\mu)$ is compact, we can apply theorem 3.1 to characterize its extreme points, because we have proven that a weakly compact set is also compact in the $w(\mathcal{E}, \mathcal{F})$ topology of the linear topological space $\mathcal{E}$. By proposition 3.9, we obtain:

Proposition 3.10. *Consider a compact metric space (X, d), and let μ be an outer-continuous capacity defined on β_X. If $\mathcal{I}$ is a weakly closed set of probability measures such that $M(\mu) = \overline{Conv(\mathcal{I})}$, then $\mathcal{I}$ contains the set of extreme points of $M(\mu)$.*

4 Extreme points of $M(\mu)$

Now, we are going to make use of the studies carried out in last section to characterize the extreme points of the set $M(\mu)$. We will find a family $\mathcal{I}$ of extreme points such that its convex hull $Conv(\mathcal{I})$ satisfies $Conv(\mathcal{I}) \subset M(\mu) \subset$

$\overline{Conv(\mathcal{I})}$. In particular, when $M(\mu)$ is closed, this family will satisfy $\overline{Conv(\mathcal{I})} = M(\mu)$.

We are first going to recall the results for a finite referential. Because $(X, \mathcal{P}(X))$ can be endowed with a separable metric space structure whenever X is finite and the continuity is trivial in that case, our results will generalize the ones we cite here.

Suppose that $|X| = n$, and take a permutation π, $\pi = (i_1, \ldots, i_n) \in S^n$. Let μ be a 2-alternating capacity on $(X, \mathcal{P}(X))$. Define a probability distribution P_π by $P_\pi(\{i_1\}) = \mu(\{i_1\}),\ P_\pi(\{i_j\}) = \mu(\{i_1, \ldots, i_j\}) - \mu(\{i_1, \ldots, i_{j-1}\})\ \forall j = 2, \ldots, n$. Then,

$P_\pi(\{i_1, \ldots, i_j\}) = \mu(\{i_1, \ldots, i_j\})\ \forall j = 1, \ldots, n.$

Proposition 4.1. *Consider μ a 2-alternating capacity on $(X, \mathcal{P}(X))$, where $|X| = n$. Then, the profile of $M(\mu)$ is given by the set $\{P_\pi : \pi \in S^n\}$.*

Therefore, $M(\mu)$ has at most n! extreme points.

This result was first established by Dempster ([9]) for ∞-alternating capacities, and alternative proofs for 2-monotone capacities have been made among others by Shapley ([23]), in the context of game theory, and Chateaneuf and Jaffray ([4]), using the Möbius inversion of a 2-alternating capacity.

Let us consider now (X, d) a separable metric space, and let us denote by $\{B_n\}_n$ a countable base of $\tau(d)$. Then, the Borel σ-algebra β_X is equal to $\sigma(B_1, \ldots)$. Take μ an outer-continuous 2-alternating capacity on β_X.

In order to define the extreme points of $M(\mu)$, it suffices to give their expression on the algebra $\mathcal{Q}(B_1, \ldots)$, and use then Carathéodory's extension [3]. Besides, from the proof of proposition 3.8 we see that given a probability measure P dominated by μ in the algebra $\mathcal{Q}(B_1, ...)$, its extension to β_X will also be dominated by μ, because μ is outer-continuous.

The key in our construction is the fact that the algebras $\mathcal{Q}(B_1, \ldots, B_n)$ are finite, and the restriction of μ to them can be seen as a 2-alternating capacity defined on a finite set.

Remark 4.1. *Consider the class $\mathcal{D}_n = \{C_1 \cap \cdots \cap C_n : C_i \in \{B_i, B_i^c\}\ \forall i = 1, \ldots, n\}$. It has at most 2^n different elements, which are disjoint; let us denote*

$\mathcal{D}_n = \{E_1^n, \ldots, E_{2^n}^n\}$. Then, every $A \in \mathcal{Q}(B_1, \ldots, B_n)$ is a finite (disjoint) union of elements of $\mathcal{D}_n$. Besides, $\mathcal{D}_{n+1} = \{A \cap B_{n+1}, A \cap B_{n+1}^c : A \in \mathcal{D}_n\}$.

We clearly observe that it suffices to have a probability mass function on $\mathcal{D}_n$ to be able to define to a probability measure on $\mathcal{Q}(B_1, \ldots, B_n)$ by additivity.

Consider an arbitrary (but fixed) $n \in \mathbb{N}$ and a permutation $\pi \in S^{2^n}$. For the sake of simplicity and without much loss of generality, let us denote it $\pi = (1, \ldots, 2^n)$. Let us define P_π^n on $\mathcal{D}_n$ by:

$P_\pi^n(E_1^n) = \mu(E_1^n)$

$P_\pi^n(E_k^n) = \mu(\cup_{j=1}^k E_j^n) - \mu(\cup_{j=1}^{k-1} E_j^n)\ \forall k = 2, \ldots, 2^n$.

We have

$P_\pi^n(\cup_{j=1}^k E_j^n) = \mu(\cup_{j=1}^k E_j^n)\ \forall k = 1, \ldots, 2^n$,

and by additivity we can define it on $\mathcal{Q}(B_1, \ldots, B_n)$.

Note that, if we denote by μ_n the restriction of μ to $\mathcal{Q}(B_1, \ldots, B_n)$, P_π^n is one of the extreme points of $M(\mu_n)$ by 4.1, and hence it belongs to $M(\mu_n)$. Thus, $P_\pi^n(A) \leq \mu(A)\ \forall A \in \mathcal{Q}(B_1, \ldots, B_n)$.

We are going to construct an extension of P_π^n to the algebra $\mathcal{Q}(B_1, \ldots) = \cup_m \mathcal{Q}(B_1, \ldots, B_m)$ satisfying that for every $m \geq n$, its restriction to the finite algebra $\mathcal{Q}(B_1, \ldots, B_m)$ is an extreme point of $M(\mu_m)$; this will be achieved associating its distribution to suitable permutations on $\mathcal{D}_m$ for every m, and repeating then the course of reasoning made above.

Let us consider, for each $m > n$, $\mathcal{D}_m$ in the following order defined recursively:

$\mathcal{D}_m := \{E_1^m, \ldots, E_{2^m}^m\} = \{E_1^{m-1} \cap B_m, E_1^{m-1} \cap B_m^c, \ldots, E_{2^{m-1}}^{m-1} \cap B_m, E_{2^{m-1}}^{m-1} \cap B_m^c\},\ \forall m > n$.

Let us now define P_π^m on $\mathcal{D}_m$ by

$P_\pi^m(E_1^m) = \mu(E_1^m)$

$P_\pi^m(E_k^m) = \mu(\cup_{j=1}^k E_j^m) - \mu(\cup_{j=1}^{k-1} E_j^m)\ \forall k = 2, \ldots, 2^m$.

Then we have

$P_\pi^m(\cup_{j=1}^k E_{i_j}^m) = \mu(\cup_{j=1}^k E_{i_j}^m)\ \forall k = 1, \ldots, 2^m$.

Now we can define a measure P_π on $\mathcal{Q}(B_1, \ldots) = \cup_{m \geq n} \mathcal{Q}(B_1, \ldots, B_m)$: given $A \in \mathcal{Q}(B_1, \ldots)$, there exists some $m \geq n$ such that $A \in \mathcal{Q}(B_1, \ldots, B_m)$; we make $P_\pi(A) = P_\pi^m(A)$. This measure satisfies the following proposition:

Proposition 4.2. *P_π is a probability measure on the algebra $\mathcal{Q}(B_1, ...)$, and $P_\pi(A) \leq \mu(A)$ $\forall A \in \mathcal{Q}(B_1, \dots)$.*

Proof: Let us first remark that P_π is well-defined: this follows from the fact that P_π^m extends $P_\pi^{m'}$, for each pair of indices m, m' such that $m > m' \geq n$, because the order defined on $\mathcal{D}_m$ is compatible with the one defined on $\mathcal{D}_{m'}$.

We prove now that $P_\pi \leq \mu$: given $A \in \mathcal{Q}(B_1, \dots)$, there exists $m \in \mathbb{N}$ s.t. $A \in \mathcal{Q}(B_1, \dots, B_m)$. Then, $P_\pi(A) = P_\pi^m(A) \leq \mu(A)$, because P_π^m is an extreme point of $M(\mu_m)$ from proposition 4.1.

Finally, we prove that P_π is a probability. It is clear that P_π is non-negative, $P_\pi(\emptyset) = 0$, and $P_\pi(X) = 1$. The finite additivity holds because P_π^m is finitely additive for every m. Given a decreasing sequence $(A_k)_k \downarrow \emptyset$ on $\mathcal{Q}(B_1, ...)$, we have that $\lim_k P_\pi(A_k) \leq \lim_k \mu(A_k) = \mu(\emptyset) = 0$, for the outer continuity condition on μ. This implies ([3]) that P_π is countably additive. ■

For the sake of simplicity, let us also denote by P_π Carathéodory's extension of this measure to $\sigma(B_1, \dots)$. Consider $\mathcal{I} = \cup_n \{P_\pi : \pi \in S^{2^n}\}$ and $\mathcal{C} = Conv(\mathcal{I})$ its convex hull.

Proposition 4.3. *Let μ be an outer continuous 2-alternating capacity on the Borel σ-algebra β_X generated by a separable metric space (X, d). Then, every element of $\mathcal{I}$ is an extreme point of $M(\mu)$, that is, $\mathcal{I} \subseteq Ext(M(\mu))$.*

Proof: Consider a permutation $\pi \in S^{2^n}$, and denote by P_π its associated probability measure on β_X. Let us first check that P_π belongs to $M(\mu)$: by construction, we have that $P_\pi(A) \leq \mu(A)$ $\forall A \in \mathcal{Q}(B_1, \dots)$. Applying now the outer continuity of μ and proposition 3.8, we conclude that $P_\pi \in M(\mu)$.

Suppose now that there exist $P_1, P_2 \in M(\mu), \alpha \in (0, 1)$ such that $P_\pi = \alpha P_1 + (1 - \alpha) P_2$. By construction, given $m \geq n$, P_π is an extreme point of $M(\mu_m)$, and hence it is $P_1 = P_2 = P_\pi$ on $\mathcal{Q}(B_1, \dots, B_m)$ $\forall m \geq n$. Thus, $P_1 = P_2 = P_\pi$ in $\mathcal{Q}(B_1, \dots)$, and, by Carathéodory's theorem, $P_1 = P_2 = P_\pi$.■

We deduce that $\mathcal{C} \subset M(\mu)$. Now, we are going to prove the inclusion $M(\mu) \subset \overline{\mathcal{C}}$. For this, consider the function $G_n : M(\mu) \to M(\mu_n)$ such that $G_n(P)$ is the restriction of P to $\mathcal{Q}(B_1, \dots, B_n)$.

Lemma 4.4. *$G_n(M(\mu)) = G_n(\mathcal{C})$ for every n.*

Proof: It is clear that $G_n(\mathcal{C}) \subset G_n(M(\mu))$ and $G_n(M(\mu)) \subset M(\mu_n)$. All these sets are convex, and the extreme points of the latter are given, from proposition 4.1, by the probabilities associated to the permutations on $\mathcal{D}_n$. Take such a permutation $\pi \in S^{2^n}$, and denote P'_π the extreme point of $M(\mu_n)$ it induces. Then, there is an extension P_π to β_X belonging to $\mathcal{C}$; it is clear that $G_n(P_\pi) = P'_\pi \Rightarrow Ext(M(\mu_n)) \subset G_n(\mathcal{C})$. As we said, μ_n can be considered to be defined on a finite referencial, and then it is $M(\mu_n) = Conv(Ext(M(\mu_n)))$. $G_n(\mathcal{C})$ is a convex set; hence, $M(\mu_n) \subset G_n(\mathcal{C})$, and consequently we have the equality $G_n(\mathcal{C}) = G_n(M(\mu))$. ■

What we have just proven is that given an element of $M(\mu)$ and a finite n, there exists some distribution in $\mathcal{C}$ with the same values on $\mathcal{Q}(B_1, \ldots, B_n)$. We will use this to prove that every probability in $M(\mu)$ is the weak limit of some sequence in $\mathcal{C}$.

Theorem 4.5. $M(\mu) \subset \overline{\mathcal{C}}$.

Proof: Consider $P \in M(\mu)$. From the previous lemma, there exists $K_n \in \mathcal{C}$ such that P and K_n agree on $\mathcal{Q}(B_1, \ldots, B_n)$. Therefore, we have that $K_n(A) \to P(A)\ \forall A \in \mathcal{Q}(B_1, ...)$ (there exists some $n_o \in \mathbb{N}$ such that $A \in \mathcal{Q}(B_1, \ldots, B_{n_0})$, and consequently the sequence $(K_n(A))_{n \geq n_0}$ is constant on $P(A)$). The class $\mathcal{Q}(B_1, ...)$ satisfies the hypotheses of proposition 3.2: it is closed under finite intersections, and any open set will be a union of elements of the countable base $\{B_n\}_n$. We deduce that $K_n \Rightarrow P$. Therefore, $P \in \overline{\mathcal{C}}$. ■

We have obtained a class $\mathcal{I}$ of extreme points of $M(\mu)$ satisfying the relation $Conv(\mathcal{I}) \subset M(\mu) \subset \overline{Conv(\mathcal{I})}$. In the finite case, this implies that $\mathcal{I}$ contains all the extreme points of $M(\mu)$. Unfortunately, for a separable space there may be extreme points which are not in $\mathcal{I}$. We might think that at least the extreme points will be related to the extreme points of $M(\mu_n)$, in the sense that any extreme point will be an extension of extreme points of $M(\mu_n)\ \forall n \geq n_0$ for some n_0 (as it happens in our construction). This is not true, as we can see in the following example:

Example 4.1. *Let $x \in \mathbb{R} \setminus \mathbb{Q}$, and consider μ a capacity defined on $\beta_{[x-1,x+1]}$ by $\mu(A) = \min\{\lambda(A) + 0.5 I_A(x) + 0.5 I_A(x - 0.5), 1\}$, where λ is Lebesgue's* *measure. Let us define a probability measure P on $\beta_{[x-1,x+1]}$ by $P(\{x\}) = 0.5, P(\{x-0.5\}) = 0.5$.*

- *μ is 2-alternating, using the fact that the minimum between a measure and 1 is always 2-alternating: take $\mu = \min\{\mu', 1\}$, and consider $A, B \in \beta_X$. There are two possibilities. If either $\mu'(A)$ or $\mu'(B) \geq 1$, then $\mu(A \cup B) = 1$ and $\mu(A \cup B) + \mu(A \cap B) = 1 + \mu(A \cap B) \leq \mu(A) + \mu(B)$. Otherwise, it is $\mu'(A), \mu'(B) < 1$, and then $\mu(A \cup B) + \mu(A \cap B) \leq \mu'(A \cup B) + \mu'(A \cap B) = \mu'(A) + \mu'(B) = \mu(A) + \mu(B)$.*

- *μ is outer continuous, making the same argument as in the previous point: given a decreasing sequence $(A_n)_n \subset \beta_{[x-1,x+1]}$, we have $\mu(\cap_n A_n) = \min\{\mu'(\cap_n A_n), 1\} = \min\{\lim_n \mu'(A_n), 1\}$. As the sequence $(P(A_n))_n$ is monotone and bounded, we can exchange the limit and the minimum; hence, $\min\{\lim_n \mu'(A_n), 1\} = \lim_n \min\{\mu'(A_n), 1\} = \lim_n \mu(A_n)$.*

- *$P(A) \leq \mu(A)$ trivially, because it is $\mu = \min\{1, P + \lambda\}$.*

- *P is an extreme point of $M(\mu)$: suppose the existence of $P_1, P_2 \in M(\mu)$ and $\alpha \in (0,1)$ s.t. $P = \alpha P_1 + (1-\alpha) P_2$; then, $P(\{x\}) = \mu(\{x\}) = 0.5 \Rightarrow P_1(\{x\}) = P_2(\{x\}) = 0.5 = \mu(\{x\})$, and similarly $P_1(\{x - 0.5\}) = P_2(\{x - 0.5\}) = P(\{x - 0.5\}) = 0.5$. Therefore, $P_1 = P_2 = P$.*

- *The spheres of rational center and radius form a base for the open sets on $\beta_{[x-1,x+1]}$. The algebra generated by these spheres is given by the finite unions of the intervals of rational extremes. Take B_1, B_2, two disjoint spheres separating x and $x - 0.5$; let us show that P is not an extreme point for any $\mathcal{Q}(B_1, ..., B_n)$ with $n \geq 2$. Take $\pi \in S^{2^n}$ a permutation, and consider P_π^n the extreme point it induces. Consider E_1^n, E_2^n the elements of $\mathcal{D}_n$ containing x and $x - 0.5$ respectively. It is $P(E_1^n) = 0.5 = P(E_2^n)$. If $P_\pi^n(E_1^n \cup E_2^n) = 1$, that is, when these are the first two elements considered by the permutation π on $\mathcal{D}_n$, we have $P_\pi^n(E_1^n) = \mu(E_1^n) > 0.5$ or $P_\pi^n(E_2^n) = \mu(E_2^n) > 0.5$, and it is $P \neq P_\pi^n$. Otherwise, it is $P_\pi^n(E_1^n \cup E_2^n) < 1$, and $P \neq P_\pi^n$.*

This example shows that in general $\mathcal{I}$ is not going to be the profile of $M(\mu)$. Using the results from the previous section, we can conclude the following theorem:

Theorem 4.6. *Consider μ an outer-continuous 2-alternating capacity defined on a compact metric space. Then, the set $\overline{\mathcal{I}}$ contains the extreme points of $M(\mu)$.*

Proof: It is clear that, since $M(\mu)$ is closed, $\overline{Conv(\overline{\mathcal{I}})} \subseteq M(\mu)$. But we have proven in theorem 4.5 that $M(\mu) \subseteq \overline{Conv(\mathcal{I})} \subseteq \overline{Conv(\overline{\mathcal{I}})}$. This implies the equality. Applying now proposition 3.1, we deduce that $\overline{\mathcal{I}}$ contains the extreme points of $M(\mu)$.■

We can see from example 4.1 that the set $\mathcal{I}$ is not closed in general: in that example, μ satisfied the hypotheses of this theorem, but we found an extreme point outside $\mathcal{I}$.

5 The case of possibility measures

Possibility measures constitute one of the most important branches of the theory of Imprecise Probabilities, and also one of the most widely used alternatives to probability measures. They were introduced by Zadeh [29] as a model for linguistic uncertainty, and they have been later related with Dempster-Shafer theory of evidence [14, 21], or with the behavioural interpretation of imprecise probabilities [25]. They have also been studied from the measure and integral-theoretic point of view [8], in parallel with probability theory, and particular attention has been devoted to the subject of conditional possibility measures and independence [6, 7, 17, 26]. All these works justify our interest in them and serve as a motivation for a study of the credal set generated by a (normal) possibility measure.

For this particular case, we are going to prove the existence of some simplifications in the set of extreme points that we have constructed in the previous section. In the finite case, the profile of $M(\mu)$ will have in general less than n! points, as we show in our next proposition:

Proposition 5.1. *If μ is a normal possibility measure on $(X, \mathcal{P}(X))$, with $|X| = n$, then $M(\mu)$ has at most 2^{n-1} extreme points.*

Proof: To simplify the notation, let us denote $X = \{1, \ldots, n\}$, and let us assume without loss of generality that $1 = \mu(\{1\}) \geq \mu(\{2\}) \geq \cdots \geq \mu(\{n\})$. We are going to prove our thesis using induction on n:

1. If $n = 2$, then $n! = 2 = 2^{n-1}$, and the extreme points are associated to the permutations $(2, 1)$ and $(1, 2)$.

2. If $n = k$ and the result holds for $n = k - 1$, let us consider a permutation $\pi \equiv (i_1, \ldots, i_k) \in S^k$:

   - If $i_1 < k$, then there exists some $h > 1$ s.t. $i_h = k$. Hence, $P_\pi(\{k\}) = \mu(\{i_1, \ldots, i_h\}) - \mu(\{i_1, \ldots, i_{h-1}\})$. Now, μ is a possibility measure, and $\mu(A \cup \{k\}) = \mu(A)\ \forall A \subset \{1, \ldots, k-1\} \Rightarrow P_\pi(\{k\}) = 0$. Thus, there exists $\pi' \in S^{k-1}$ such that $P_{\pi'}(\{i\}) = P_\pi(\{i\})\ \forall i : 1, \ldots, k-1$. Reciprocally, given $\pi' \in S^{k-1}$, we can consider $\pi = (i'_1, \ldots, i'_{k-1}, k)$, and we have that $P_\pi(\{i\}) = P_{\pi'}(\{i\})\ \forall i = 1, \ldots, k-1$. Therefore, $|\{P_\pi : \pi \in S^k, \pi(k) > 1\}| = |\{P_{\pi'} : \pi' \in S^{k-1}\}| \leq 2^{k-2}$.
   - If $i_1 = k$, then $P_\pi(\{k\}) = \mu(\{k\})$, and given $j = i_2$, it is $P_\pi(\{j\}) = \mu(\{j\}) - \mu(\{k\})$. Given the permutation $\pi' = (i_2, \ldots, i_k) \in S^{k-1}$, $P_{\pi'}$ satisfies the equalities $P_{\pi'}(\{j, k\}) = P_\pi(\{j\}) + P_\pi(\{k\})$, $P_{\pi'}(\{i_l\}) = P_\pi(\{i_l\})\ \forall l : 3, \ldots, k$. Conversely, given $\pi' \in S^{k-1}$, the probability P_π associated to the permutation $\pi = (k, i'_1, \ldots, i'_{k-1})$ satisfies the previous relation. Thus, $|\{P_\pi : \pi \in S^k, i_1 = k\}| = |\{P_{\pi'} : \pi' \in S^{k-1}\}| \leq 2^{k-2}$.

   Therefore, $|\{P_\pi : \pi \in S^k\}| \leq 2^{k-1}$. ■

This result does not hold in general for plausibility functions: we can construct a plausibility function μ on a finite space $(X, \mathcal{P}(X))$, with $|X| = n$, such that $M(\mu)$ has n! extreme points:

Example 5.1. *Take $X = \{1, \ldots, n\}$, and consider an arbitrary order on its $2^n - 2$ proper subsets. We denote them $\{A_1, \ldots, A_{2^n-2}\}$. Define the probability mass function m by $m(A_i) = \frac{1}{2^i}\ \forall i = 1, \ldots, 2^{n-2}$, and $m(X) = \frac{1}{2^{2^n-2}}$, and* *consider* μ *the plausibility function associated to this distribution through the formula* $\mu(A) = \sum_{B\cap A\neq\emptyset} m(B)$*. Then, given a permutation* $\pi = (i_1, \ldots, i_n) \in S^n$*, it is* $P_\pi(\{i_j\}) = \mu(\{i_1, \ldots, i_j\}) - \mu(\{i_1, \ldots, i_{j-1}\}) = \sum_{i_j\in B\subset\{i_j,\ldots,i_n\}} m(B)$*. The form of the distribution* m *makes* $P_\pi(i_j) \neq P_{\pi'}(i_j)$ *unless all the terms in their respective sums coincide, that is, unless they are considering the same sets* B *on the sum. Hence, the* $n!$ *permutations yield* $n!$ *different extreme points.*

In the general case of a separable metric space, we can also make some simplifications on our set $\mathcal{I}$. This is because the possibility distribution will induce an order on the countable family $\{B_n\}_n$. Take μ an outer continuous possibility distribution on β_X. Then, the restriction of μ to $\mathcal{Q}(B_1, \ldots, B_n)$ is also a possibility measure. From the previous lemma, we have that $|\{P_\pi : \pi \in S^{2^n}\}| \leq 2^{2^n-1}$

We can then select from S^{2^n} a permutation π associated to each of the extreme points, so that we obtain Π_n a subset of S^{2^n} with $\{P_\pi : \pi \in \Pi_n\} = \{P_\pi : \pi \in S^{2^n}\}$, and $P_\pi \neq P_{\pi'}$ $\forall \pi \neq \pi' \in \Pi_n$. This leads to the following proposition:

Proposition 5.2. *Let* μ *be an outer continuous possibility measure defined on the Borel* σ*-algebra* β_X *associated to a separable metric space* (X, d)*. Then,* $\mathcal{J} = \{P_\pi : \pi \in \Pi_n \text{ for some } n\}$ *satisfies* $Conv(\mathcal{J}) \subset M(\mu) \subset \overline{Conv(\mathcal{J})}$*.*

As a consequence, following the course of reasoning made in the previous section, we obtain:

Corollary 5.3. *Under the hypotheses of the previous proposition, if* (X, d) *is in particular a compact metric space, then* $\overline{\mathcal{J}}$ *contains the extreme points of* $M(\mu)$*.*

6 Conclusions and further work

In this paper, we have continued the studies made by Dempster [10], Shafer[22] or Walley [25] about upper and lower probability models in an infinite setting. In particular, we have investigated the extreme points of convex sets of probabilities. As far as we know, most of the work in this subject has been restricted to the finite case [4, 9, 23]. In these works, the problem has been solved for sets of probabilities dominated by second order alternating Choquet capacities.

These type of sets have interesting properties in connection with robust statistics [12] and the theory of imprecise probabilities [25]. The particular study of the finite case benefits from different technical facilities such as the existence of the Möbius inversion and the structure of Euclidean space on the set of probability measures. However, in the more general case considered here, we have been forced to include the class of probability measures in a linear topological space; we have done it recapturing the weak topology on $\mathcal{P}$. We have showed that this topology has some undesirable properties: in particular, we have given an example where the credal set $M(\mu)$ is not closed under this structure, let alone compact. Nevertheless, we have seen that under the additional condition of outer continuity, this set is closed. This condition is essential also in order to characterize the dominance over a probability measure, and becomes trivial on a finite referencial.

All this work on the functional structure of $M(\mu)$, though motivated by the subsequent study of the extreme points, is also useful on its own: many trivial characteristics of $M(\mu)$ in the finite case, such as its compactness, do not hold in general. Then, it becomes interesting to see whether they are implied by some extra conditions on the structure of the referential space.

It has been proven that in the finite case, the extreme points of $M(\mu)$ are related to the set of permutations of the referential set. Perhaps the shortest proof of this fact is in [4]. We have obtained here what could be called a generating set of extreme points for a general separable metric space. The set we obtain is in correspondence with the extreme points of the finite algebras. Two remarks must be done here: first, this set is not in general the profile of $M(\mu)$, so different generating sets could be found; and second, not all the extreme points of $M(\mu)$ will be extensions of extreme points on the algebras $\mathcal{Q}(B_1, ..., B_n)$.

It should also be remarked that for a possibility measure, and even in the finite case, we have succeeded in lessen the maximal number of extreme points, something that, as far as we know, had not been noted previously in the literature.

Let us point out some possible ways to continue our study: different topologies in the set of probabilities can be studied; in particular, we have already noted in section 3 that Walley [25] has obtained good results on the set of linear previsions dominated by an upper prevision using the weak-* topology [2]. On the other hand, we think that the construction made in this paper can be useful for handling different problems concerning credal sets.

We conclude outlining some connections with other theories: In [23], the results for the finite case were used for determining the Shapley value of a convex cooperative game. Specifically, it was shown that this value was the center of gravity of the extreme points of $M(\mu)$. We intend to study whether our results help to determine the Shapley value for some cases of non-atomic convex games.

Concerning random sets, the upper probability P^* they induce [9] is always 2-alternating ([19]), although it is not outer-continuous in general. We think that other hypothesis, such as the approximation of the upper probability by the distributions of the measurable selections, will allow us to use the work in this paper to determine the extreme points of $M(P^*)$.

Acknowledgements

This research has been supported in part by DGE grant number PB97-1286.

References

[1] Abellán, J. and Moral, S., A non-specificity measure for convex sets of probability distributions. *Int. J. of Unc., Fuz. and Know.-Based Systems*, 8:357–367, 2000.

[2] Billingsley, P., *Convergence of probability measures.* Wiley, New York, 1968.

[3] Billingsley, P., *Probability and measure.* Wiley, New York, 1986.

[4] Chateauneuf, A., and Jaffray, J.-Y., Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Math. Soc. Sciences*, 17:263–283, 1989.

[5] Choquet, G., Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] Even though his results would apply to *finitely* additive probabilities.

[6] de Campos, L.M., and Huete, J.F., Independence concepts in possibility theory: Part I. *Fuzzy Sets and Systems*, 103:127–152, 1999.

[7] de Campos, L.M., and Huete, J.F., Independence concepts in possibility theory: Part II. *Fuzzy Sets and Systems*, 103:487–505, 1999.

[8] de Cooman, G., Possibility theory I-III. *Int. J. of General Systems*, 25:291–371, 1997.

[9] Dempster, A.P., Upper and lower probabilities induced by a multivalued mapping. *Ann. of Math. Stat.*, 38:325–339, 1967.

[10] Dempster, A.P., Upper and lower probabilities generated by a random closed interval. *Ann. of Math. Stat.*, 39:957–966, 1968.

[11] Denneberg, D., *Non-additive measure and integral.* Kluwer Academic Publishers, 1994.

[12] Huber, P.J., *Robust Statistics.* Wiley, New York, 1981.

[13] Kelley, J.L., and Namioka, I., *Linear Topological Spaces.* Springer Verlag, New York, 1963.

[14] Klir, G. J., and Parviz, B., Probability-possibility transformations: a comparison. *Int. J. of General Systems*, 21:291–310, 1992.

[15] Levi, I., *The enterprise of knowledge.* MIT Press, 1980.

[16] Matheron, G., *Random sets and integral geometry.* Wiley, New York, 1975.

[17] Miranda, E., and de Cooman, G., Epistemic independence in numerical possibility theory. *Int. J. of Appr. Reasoning*, 32:23–42, 2003.

[18] Miranda, E., Couso, I., and Gil., P., On the probabilities dominated by a 2-alternating capacity on a separable metric space. *Proceedings of the International Summer School on Aggregation Operators and Their Applications*, Oviedo, Spain, 87–93, 2001.

[19] Nguyen, H. T., On random sets and belief functions. *J. of Math. Anal. and Appl.*, 63:531–542, 1978.

[20] Nguyen, H. T., Nguyen, N. T., and Wang, T., On capacity functionals in interval probabilities. *Int. J. of Unc., Fuz. and Know.-Based Systems*, 5:359–377, 1997.

[21] Shafer, G., *A mathematical theory of evidence*. Princeton University Press, New Jersey, 1972.

[22] Shafer, G., Allocations of probability. *Ann. of Prob.*, 7:827–839, 1979.

[23] Shapley, L. S., Cores of convex games. *Int. J. of Game Theory*, 1:11–26, 1971.

[24] Walley, P., Coherent lower (and upper) probabilities. *Statistics Research Report, University of Warwick (Coventry)*, 22, 1981.

[25] Walley, P., *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

[26] Walley, P., and de Cooman, G., Coherence of rules for defining conditional possibility. *Int. J. of Appr. Reasoning*, 21:63–107, 1999.

[27] Walley, P., and de Cooman, G., A behavioral model for linguistic uncertainty. *Inf. Sci.*, 134:1–37, 2001.

[28] Walley, P., and Fine, T. L., Towards a frequentist theory of upper and lower probability. *Ann. of Stat.*, 10:741–761, 1982.

[29] Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[30] Zaffalon, M., The naive credal classifier. *J. of Stat. Plan. and Inference*, 105:5–21, 2002.