



# Regret-based budgeted decision rules under severe uncertainty

Nawapon Nakharutai<sup>a</sup>, Sébastien Destercke<sup>b,\*</sup>, Matthias C.M. Troffaes<sup>c</sup>

<sup>a</sup> Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>b</sup> Université de Technologie de Compiègne, UMR CNRS 7253 Heudiasyc, Sorbonne Université, Compiègne, 60200, France

<sup>c</sup> Department of Mathematical Sciences, Durham University, Durham, DH13LE, United Kingdom

## ARTICLE INFO

MSC:  
0000  
1111

### Keywords:

Decision  
Regret  
Minimax  
Maximin  
Imprecise probability  
Optimization  
Numerical algorithm

## ABSTRACT

One way to make decisions under uncertainty is to select an optimal option from a possible range of options, by maximizing the expected utilities derived from a probability model. However, under severe uncertainty, identifying precise probabilities is hard. For this reason, imprecise probability models uncertainty through convex sets of probabilities, and considers decision rules that can return multiple options to reflect insufficient information. Many well-founded decision rules have been studied in the past, but none of those standard rules are able to control the number of returned alternatives. This can be a problem for large decision problems, due to the cognitive burden decision makers have to face when presented with a large number of alternatives. Our contribution proposes regret-based ideas to construct new decision rules which return a bounded number of options, where the limit on the number of options is set in advance by the decision maker as an expression of their cognitive limitation. We also study their consistency and numerical behaviour.

## 1. Introduction

Consider a decision problem where a decision maker wants to choose the best option from a set of available options, but where there are some uncertain variables that also influence the value of the decision. Such problems are routinely met in machine learning, where a classifier will typically return an uncertainty model over the possible classes (see [5, Ch. 10] and [22]); in structure design, where the quality of each design depends on multiple uncertain factors [12,4,6,23]; in agronomy where the weather and soil qualities are uncertain by nature [21]; etc. A classical way to model such a situation is to associate a reward with each option assumed to be expressed in numerical utility scale and that depends on the uncertain state of nature. In this way, we can view an uncertain reward, and thereby, each decision, as a bounded real-valued function on the set of states of nature. Such function is commonly called an act [2] or a gamble [26] in decision theory.

If uncertainty is described by precise probabilities for all events, a classical approach is to select the act that yields the highest expected utility [2]. However, such precise probabilities may not always be obtainable in a reliable way, particularly when we only have little information about the states of nature and their chance to happen. One way to handle this situation is to relax the need for a unique, precise probability and to consider imprecise probability theory, in which uncertainty is modelled by sets of probabilities [5]. This theory includes as special cases many commonly used uncertainty models, such as precise probabilities, possibility distributions,

\* Corresponding author.

E-mail address: [sebastien.destercke@hds.utc.fr](mailto:sebastien.destercke@hds.utc.fr) (S. Destercke).

<https://doi.org/10.1016/j.ins.2024.120361>

Received 2 November 2023; Received in revised form 26 January 2024; Accepted 21 February 2024

Available online 27 February 2024

0020-0255/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and belief functions. In this paper, we assume that our knowledge about the true state of nature is represented by a closed convex set of probability mass functions.

Using sets of probabilities then requires to generalize the classical expected utility criterion. This is usually done either by proposing decision rules that return a single<sup>1</sup> option as optimal such as  $\Gamma$ -maximin,  $\Gamma$ -maximax or Hurwicz, or decision rules that can return multiple incomparable options as optimal, such as interval dominance, maximality and E-admissibility [24]. Nevertheless, these criteria can return any subset of options, from a single optimal act to the whole set of acts [18]. This may be problematic, as it is conceivable that even if decision makers are keen to consider multiple options, they still may wish to limit the number of returned options, for example because of cognitive or monetary constraints. Such requirements are natural in many settings, such as set-valued classification [14,7] or recommendation problems [25].

However, such *budgeted* decision rule that limits the number of returned decisions still remains to be defined and explored within the imprecise probabilistic setting. This paper aims to propose and study such rules, using a regret criterion and expressing the rule either as a minimax or maximin problem. More precisely, we define the value of a given subset (of limited size) as the regret one would feel if an adversary was to pick an alternative from outside the retained subset. Note that we already studied the minimax version of this rule in a previous conference paper [19]. This paper provides extended proofs, examples and numerical simulation of this minimax criterion, and introduces the maximin version, which was not studied before.

The paper is organised as follows. In section 2, we present necessary notations and basic concepts for regret-based budgeted decision rules. We devote section 3 to review regret-based budgeted decision rules by recalling the minimax criterion from our previous study. In section 4, we introduce the new maximin criterion and study their properties and consistency with classical imprecise probability decision criteria and provide an algorithm for the maximin criterion. In section 5, we perform some computational experiments of these two regret-based budgeted decision rules. Section 6 provides two illustrative applications of the proposed rules. Finally, section 7 concludes the paper.

## 2. Preliminaries and definitions

We denote by  $\Omega$  a finite set of possible states of nature about which we are uncertain. We will consider decision problems where a subject can choose a finite number of acts from a set of acts  $A$ . For each act  $a \in A$ , if  $\omega \in \Omega$  turns out to be the true state of nature, then  $a(\omega)$  will represent the reward obtained by the subject (an end-user, a decision maker, ...). We assume that the subject can specify a utility for each reward. Therefore, an act  $a$  can be viewed as a real-valued function on  $\Omega$  representing an uncertain reward expressed on a utility scale.

In order to select an act, the subject can take their beliefs about the true state of nature into account. If their beliefs can be expressed through a probability mass function  $p$ , then they can simply select an act that maximizes their expectation  $\mathbb{E}_p(a) := \sum_{\omega \in \Omega} p(\omega)a(\omega)$ . However, when information is lacking, they might only be able to represent their knowledge about the unknown true value  $\omega$  via a closed convex set of probability mass functions on  $\Omega$ . Such a set  $\mathcal{P}$  of probability mass functions is called a credal set [11].

Let  $\mathcal{A}$  denote the set of all finite non-empty sets of acts. Mathematically, we can then define:

**Definition 1.** A *decision rule* is a function  $D : \mathcal{A} \rightarrow \mathcal{A}$  such that  $D(A) \subseteq A$  for every  $A \in \mathcal{A}$ .

In the case of a precise probability mass function, the classical decision rule is  $D(A) = \arg \max_{a \in A} \mathbb{E}_p(a)$ . Given a credal set  $\mathcal{P}$ , there are several decision rules that extend expected utility. Some can return a single act, e.g.  $\Gamma$ -maximin (or  $\Gamma$ -maximax) which maximizes the worst case (or minimal) expected utility of acts in  $A$ , and others a set of multiple acts, e.g. interval dominance and maximality [24]. In this study, we will compare our budgeted decision rules to maximality, which is induced by a partial ordering as follows: for any acts  $a$  and  $a'$ , we say that an act  $a$  dominates  $a'$ , denoted  $a \succ a'$ , whenever  $\underline{\mathbb{E}}(a - a') > 0$ , where  $\underline{\mathbb{E}}(f) := \min_{p \in \mathcal{P}} \mathbb{E}_p(f)$  is the lower bound of the expectation of a function  $f$  taken over all probabilities of  $\mathcal{P}$ ; observe that  $a - a'$  is the point-wise difference between two acts, and is therefore a real-valued function over  $\Omega$ . Note that  $\underline{\mathbb{E}}(a - a') > 0$  is equivalent to requiring that  $\mathbb{E}_p(a) > \mathbb{E}_p(a')$  for all  $p \in \mathcal{P}$ , and is therefore equivalent to a robust version of the classical expectation-based decision rule where  $a$  dominates  $a'$  if  $\mathbb{E}_p(a) > \mathbb{E}_p(a')$ .

The set of maximal acts in  $A$  with respect to  $\succ$  is then defined by:

$$D_M(A) := \{a \in A : \nexists a' \in A \text{ s.t. } a' \succ a\} \tag{1}$$

which contains all undominated acts in  $A$  with respect to  $\succ$ . In other words,  $a$  is a maximal element in  $A$  if and only if

$$\min_{a' \in A} \overline{\mathbb{E}}(a - a') \geq 0, \tag{2}$$

where  $\overline{\mathbb{E}}(f) := \max_{p \in \mathcal{P}} \mathbb{E}_p(f)$  is defined as the upper expectation operator, which is dual to the lower expectation since  $\underline{\mathbb{E}}(f) = -\overline{\mathbb{E}}(-f)$ . Note that maximality can return any non-empty subset of  $A$ , from a single act up to the whole set  $A$  [18]. If most or all elements of  $A$  are returned by maximality, this may not be helpful to a decision maker, especially when  $A$  is very large or when checking alternatives is costly.

<sup>1</sup> Or multiple indifferent options.

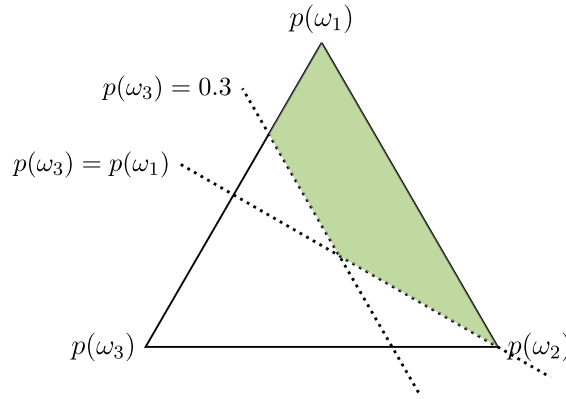


Fig. 1. Credal set of Example 1 in barycentric coordinates (in green).

Table 1  
Acts of Example 1.

	$\omega_1$	$\omega_2$	$\omega_3$
$a_1$	6	3	1
$a_2$	2	7	4
$a_3$	5	1	8
$a_4$	5	4	3
$a_5$	1	2	6

To address this issue, we introduce a so-called budgeted decision rule.

**Definition 2.** A decision rule  $D$  is said to be  $k$ -budgeted if  $|D(A)| \leq k$  for all  $A \in \mathcal{A}$ .

In the above,  $|D(A)|$  denotes the cardinality of  $D(A)$  (i.e. the number of elements of  $D(A)$ ). The parameter  $k$  is therefore an upper bound on the number of returned alternatives. In practice, its value can be settled accordingly to the application constraints, i.e., the number of alternatives that can be realistically inspected, for instance by a decision maker or by a second automatic processing step. It may also depend on the cardinality  $|\mathcal{A}|$ , e.g., passing on at most 10% of  $\mathcal{A}$  to the next processing step.

Next, we consider whether a  $k$ -budgeted decision rule preserves maximality. Specifically, we define consistency properties with respect to maximality as follows:

**Definition 3.** A decision rule  $D$  is said to be strongly consistent (with  $D_M$ ) if for all  $A \in \mathcal{A}$ , we have

$$D(A) \subseteq D_M(A). \tag{3}$$

**Definition 4.** A decision rule  $D$  is said to be weakly consistent (with  $D_M$ ) if for all  $A \in \mathcal{A}$

$$D(A) \cap D_M(A) \neq \emptyset. \tag{4}$$

Strong consistency ensures that a rule selects maximal acts only. Weak consistency ensures that it selects at least one maximal act, though it may also select non-maximal acts. Clearly, strong consistency implies weak consistency. One can easily adapt those definitions to other decision rules possibly returning multiple elements, but we will focus here on maximality.

**Example 1.** Consider the state space  $\Omega = \{\omega_1, \omega_2, \omega_3\}$  and the acts of Table 1. Suppose furthermore that the credal set is specified as

$$\mathcal{P} = \{p \in \mathbb{P} : p(\omega_3) \leq p(\omega_1), p(\omega_3) \leq 0.3\} \tag{5}$$

where  $\mathbb{P}$  denotes the set of all probability mass functions on  $\Omega$ . The corresponding credal set is displayed in Fig. 1.

Table 2 gives the values of  $\bar{\mathbb{E}}(a_j - a_i) = -\underline{\mathbb{E}}(a_i - a_j)$  where  $j \neq i$ . The set of maximal elements of  $A$  is  $D_M(A) = \{a_1, a_2, a_3, a_4\}$ , as only row 5 of Table 2 contains negative values. For instance, we have that  $a_5 - a_1 = (-5, -1, 5)$ , and the upper expected value  $\bar{\mathbb{E}}(a_5 - a_1) = -0.4$  is obtained by considering the distribution  $p(\omega_1) = 0.3, p(\omega_2) = 0.4, p(\omega_3) = 0.3$  (the extreme point of the credal set that is the intersection of the dotted lines in Fig. 1), which gives  $\bar{\mathbb{E}}(a_5 - a_1) = -5 \times 0.3 + (-1) \times 0.4 + 5 \times 0.3 = -0.4$ .

In the next section, we will construct  $k$ -budgeted decision rules based on the idea of regret. We introduce first the notion of regret we will consider in our decision rules, before proposing and studying its minimax and maximin versions to obtain recommendations with a limited budget.

**Table 2**  
Values of  $\bar{\mathbb{E}}(a_j - a_i)$ .

$j \backslash i$	1	2	3	4	5
1	-	4.0	2.0	1.0	5.0
2	4.0	-	6.0	3.0	5.0
3	1.4	3.3	-	1.5	4.0
4	1.0	3.0	3.0	-	4.0
5	-0.4	-0.1	1.0	-1.1	-

### 3. Minimax criterion

#### 3.1. Regret

Let  $a$  and  $a'$  be acts in  $A$ . Then  $\bar{\mathbb{E}}(a' - a) = \max_{p \in \mathcal{P}} \mathbb{E}_p(a' - a) = -\underline{\mathbb{E}}(a - a')$  is the maximal expected gain if we exchange  $a'$  for  $a$ , in other words, the worst possible expected loss, or regret, to us if we keep  $a$  instead of replacing  $a$  for  $a'$ . If  $a$  has a better expected utility than  $a'$  under all  $p \in \mathcal{P}$ , then this value will be negative. Consider an act  $a$  and any set of acts  $S'$  which does not contain  $a$ . The maximal loss, or worst regret, of not replacing  $a$  for any act  $a' \in S'$  is therefore

$$ML(a, S') := \max_{a' \in S'} \bar{\mathbb{E}}(a' - a), \tag{6}$$

where we take the maximum to be  $-\infty$  if  $S' = \emptyset$ . If we had to pick a single alternative from  $A$ , a natural choice would be to minimize our worst regret given by eq. (6) for  $S' = A \setminus \{a\}$ .

#### 3.2. Review minimax criterion

In this section, we recall the main results from [19], with some additional elements and illustrations here for the sake of completeness, and to enable comparison to the maximin approach that will be introduced later.

We now consider a set-valued criterion based on eq. (6). Consider any set  $S$  of acts such that  $\emptyset \neq S \subseteq A$ . In this first minimax scenario, we first have to choose an element  $a$  within the set  $S$ , and for each  $a \in S$ , the opponent can then pick  $a' \in S' := A \setminus S$  yielding the highest gain to them (the maximal loss to us),  $\max_{a' \in S'} \bar{\mathbb{E}}(a' - a)$ . The minimax regret of choosing a given set  $S$  is then [19, eq. (3)]:

$$mML(S, A) := \min_{a \in S} ML(a, S') = \min_{a \in S} \max_{a' \in S'} \bar{\mathbb{E}}(a' - a). \tag{7}$$

Note that  $mML(A, A) = -\infty$ , because for  $S = A$  we have  $S' = \emptyset$ .

Next, we recall some basic properties of  $mML(S, A)$ . First, eq. (7) is monotone with respect to set inclusion (the bigger the set we can select from, the lower our loss) [19, Lemma 1]:

**Lemma 1.** For any  $\emptyset \neq S \subseteq T \subseteq A \in \mathcal{A}$ , we have that  $mML(S, A) \geq mML(T, A)$ .

**Proof.** Recall  $S' := A \setminus S$  and  $T' := A \setminus T$ . Indeed,

$$mML(T, A) = \min_{a \in T} ML(a, T') \tag{8}$$

$$\leq \min_{a \in S} ML(a, T') \quad (\text{since } S \subseteq T) \tag{9}$$

$$= \min_{a \in S} \max_{a' \in T'} \bar{\mathbb{E}}(a' - a) \tag{10}$$

$$\leq \min_{a \in S} \max_{a' \in S'} \bar{\mathbb{E}}(a' - a) \quad (\text{since } T' \subseteq S') \tag{11}$$

$$= mML(S, A). \quad \square \tag{12}$$

Next, if  $mML(S, A)$  is negative, then  $S$  is a superset of the set of maximal elements [19, Theorem 1]:

**Theorem 1.** For  $\emptyset \neq S \subseteq A$ ,  $mML(S, A) < 0$  if and only if there is an  $a \in S$  such that for all  $a' \in S' := A \setminus S$  we have that  $a > a'$ .

**Proof.** By the definition, we have

$$mML(S, A) < 0 \iff \min_{a \in S} \max_{a' \in S'} \bar{\mathbb{E}}(a' - a) < 0 \tag{13}$$

$$\iff \exists a \in S, \max_{a' \in S'} \bar{\mathbb{E}}(a' - a) < 0 \tag{14}$$

$$\iff \exists a \in S, \forall a' \in S', \bar{\mathbb{E}}(a' - a) < 0 \tag{15}$$

$$\iff \exists a \in S, \forall a' \in S', \underline{\mathbb{E}}(a - a') > 0. \quad \square \tag{16}$$

**Corollary 1.** *If  $mML(S, A) < 0$  then  $D_M(A) \subseteq S$ .*

**Proof.** By Theorem 1, if  $mML(S, A) < 0$  then all  $a' \in S'$  are dominated by some  $a \in S$ , and therefore, no  $a' \in S'$  can be maximal. Consequently, it must be that  $S$  contains all maximal elements.  $\square$

Given the fact that a decision maker wants to choose  $S$  so as to minimize its regret, an optimal subset  $S \subseteq A$  with respect to the  $mML$  criterion can be defined as

$$S_k^*(A) := \arg \min_{\substack{\emptyset \neq S \subseteq A \\ |S| \leq k}} mML(S, A). \tag{17}$$

As the arg min may not be unique, but may contain sets of alternatives that are indifferent from a regret viewpoint, we assume one set is picked at random if need be. The minimum value is denoted by

$$mML_k^*(A) := mML(S_k^*(A), A). \tag{18}$$

By Lemma 1,  $S_k^*(A)$  will always be either  $A$  if  $|A| \leq k$ , or a set containing exactly  $k$  elements if  $|A| > k$ , and by Corollary 1,  $D_M(A) \subseteq S_k^*(A)$  if  $mML_k^*(A) < 0$ . In particular, we know that  $D_M(A) \subseteq S_k^*(A)$  if  $mML_k^*(A) < 0$  or  $|A| \leq k$ . Therefore, it makes sense to define an  $mML$  budgeted decision rule as follows:

$$D_{mML}^k(A) := \begin{cases} D_M(A) & \text{if } mML_k^*(A) < 0 \text{ or } |A| \leq k \\ S_k^*(A) & \text{otherwise.} \end{cases} \tag{19}$$

In [19], we provided an example to show that  $S_k^*(A)$  is not unique, and that we can have  $S_k^*(A) \not\subseteq S_{k+1}^*(A)$ . The latter fact shows in particular that the solution of Eq. (17) cannot always be obtained incrementally by gradually increasing  $k$ , i.e., that a greedy approach will generally be sub-optimal.

Fortunately, there is an efficient algorithm [19, Algorithm 1] to obtain  $S_k^*(A)$  and  $mML_k^*(A)$  without evaluating all possible sets  $S$  of size  $k$ , of which there are  $\binom{|A|}{k}$ . This algorithm is represented in Algorithm 1. Note that this algorithm is polynomial. Loop 2-4 requires computing a quadratic number of upper expectations (which can be done by linear programming), and loop 6-10 is a sorting procedure.

---

**Algorithm 1** Finding  $S_k^*(A)$ .

---

**Require:**  $A = \{a_1, a_2, \dots, a_n\}$ ,  $P$ ,  $k$

**Ensure:**  $S_k^*(A)$ ,  $mML_k^*(A)$

```

1: for  $i = 1 : n$  do
2:   for  $j = 1 : n, j \neq i$  do
3:     compute  $e_{ij} := \bar{\mathbb{E}}(a_j - a_i)$ 
4:   end for
5: end for
6: for  $i = 1 : n$  do
7:    $S[i] \leftarrow$  set such that  $\{e_{ij} : j \in S[i]\}$  are the  $k$  largest elements of  $\{e_{ij} : j \neq i\}$ 
8:    $M[i] \leftarrow \min_{j \in S[i]} e_{ij}$ 
9:    $J[i] \leftarrow \arg \min_{j \in S[i]} e_{ij}$ 
10: end for
11:  $i^* \leftarrow \arg \min_{i=1}^n M[i]$ 
12: return  $\{a_j : j \in \{i^*\} \cup S[i^*] \setminus \{J[i^*]\}\}$ ,  $M[i^*]$ 

```

---

The next example continues Example 1, illustrating the new decision rule and the algorithm to obtain  $S_k^*(A)$  [19, Example 1]. It also illustrates a case where  $S_k^*(A) \not\subseteq S_{k+1}^*(A)$ .

**Example 2.** Consider the credal set given by Example 1, then the optimal sets given by Eq. (17) and obtainable through Algorithm 1 are

- $S_1^*(A) = \{a_4\}$  with  $mML_1^*(A) = 3$ ,
- $S_2^*(A) = \{a_1, a_2\}$  with  $mML_2^*(A) = 1.4$ ,
- $S_3^*(A) = \{a_1, a_2, a_3\}$  or  $\{a_2, a_3, a_4\}$  with  $mML_3^*(A) = 1$  and
- $S_4^*(A) = \{a_1, a_2, a_3, a_4\}$  with  $mML_4^*(A) = -1.1$ .

Note that  $S_1^*(A) \not\subseteq S_2^*(A)$ . Let us detail how Algorithm 1 works to get  $S_2^*(A)$ . Line 7 gives  $S[1] = \{2, 3\}$  as the two largest elements of column 1 in Table 2 are in 4.0 (row 2) and 1.4 (row 3). For the other columns, we get  $S[2] = \{1, 3\}$ ,  $S[3] = \{2, 4\}$ ,  $S[4] = \{2, 3\}$ , and

$S[5] = \{1, 2\}$ . We then get  $M[1] = 1.4$ ,  $M[2] = 3.3$ ,  $M[3] = 3$ ,  $M[4] = 1.5$ , and  $M[5] = 5$  (of which the minimum is 1.4 at  $i^* = 1$ ), as well as  $J[1] = 3$ ,  $J[2] = 3$ ,  $J[3] = 4$ ,  $J[4] = 3$ , and  $J[5] = 1$ . From this, we obtain  $S_2^*(A) = \{1\} \cup \{2, 3\} \setminus \{3\}$  and the corresponding mML value 1.4. In the case of  $S_3^*(A)$ , the rule would return one of the two sets  $\{a_1, a_2, a_3\}$  or  $\{a_2, a_3, a_4\}$ , picking one at random.

We previously gave some properties of the  $mML(S, A)$  function. Next are some properties of  $S_k^*(A)$  [19, Theorem 2]:

**Theorem 2.** For all  $k \geq 1$ ,

$$S_k^*(A) \cap D_M(A) \neq \emptyset. \tag{20}$$

**Proof.** For brevity, define  $S := S_k^*(A)$  and  $S' := A \setminus S_k^*(A)$ . If  $|A| \leq k$  then  $S = A$  by Lemma 1 and therefore  $S \cap D_M(A) = D_M(A) \neq \emptyset$ . If  $mML(S, A) < 0$  then the statement follows from Corollary 1. So, suppose that  $|A| > k$  and  $mML(S, A) \geq 0$ . Let

$$a_{i^*} := \arg \min_{a_i \in S} \max_{a_j \in S'} \bar{\mathbb{E}}(a_j - a_i), \tag{21}$$

$$a_{j^*} := \arg \max_{a_j \in S'} \bar{\mathbb{E}}(a_j - a_{i^*}). \tag{22}$$

Note that

$$0 \leq mML(S, A) = \min_{a_i \in S} \max_{a_j \in S'} \bar{\mathbb{E}}(a_j - a_i) = \bar{\mathbb{E}}(a_{j^*} - a_{i^*}), \tag{23}$$

and it follows that

$$\bar{\mathbb{E}}(a_{j^*} - a_{i^*}) \geq 0, \tag{24}$$

$$\forall a_j \in S', \bar{\mathbb{E}}(a_j - a_{i^*}) \leq \bar{\mathbb{E}}(a_{j^*} - a_{i^*}) \tag{25}$$

$$\forall a_i \in A, \forall j \in S[i], \bar{\mathbb{E}}(a_j - a_i) \geq \bar{\mathbb{E}}(a_{j^*} - a_{i^*}) \tag{26}$$

where  $S[i]$  is defined as in Algorithm 1. Equation (26) holds because, from the algorithm, we know that

$$M[i] = \min_{j \in S[i]} \bar{\mathbb{E}}(a_j - a_i) \geq M[i^*] = \bar{\mathbb{E}}(a_{j^*} - a_{i^*}). \tag{27}$$

We have now everything in place to show that  $a_{i^*}$  is maximal, i.e.  $\bar{\mathbb{E}}(a_{i^*} - a_\ell) \geq 0$  for all  $a_\ell \in A$ . Fix any  $a_\ell \in A$  and consider the set

$$B := \{a_m : \bar{\mathbb{E}}(a_m - a_\ell) \geq \bar{\mathbb{E}}(a_{j^*} - a_{i^*})\}. \tag{28}$$

This set has at least  $k$  elements by Equation (26) and the definition of  $S[i]$ . If  $a_{i^*} \in B$ , then we are done, by Equation (24). Otherwise,  $B$  must contain at least one element outside of  $S$  and thus in  $S'$ , since  $S$  has exactly  $k$  elements and  $a_{i^*} \in S$ . Choose  $a_m \in B \cap S'$ . Then

$$\begin{aligned} \bar{\mathbb{E}}(a_{i^*} - a_\ell) &\geq \bar{\mathbb{E}}(a_m - a_\ell) - \bar{\mathbb{E}}(a_m - a_{i^*}) \\ &= \underbrace{\bar{\mathbb{E}}(a_m - a_\ell) - \bar{\mathbb{E}}(a_{j^*} - a_{i^*})}_{\text{non-negative by eq. (28)}} + \underbrace{\bar{\mathbb{E}}(a_{j^*} - a_{i^*}) - \bar{\mathbb{E}}(a_m - a_{i^*})}_{\text{non-negative by eq. (25)}} \geq 0, \end{aligned}$$

and thus, in this case, the desired inequality also holds.  $\square$

From this theorem follows two simple corollaries [19, Corollaries 2 and 3]:

**Corollary 2.** For all  $k \geq 1$ ,  $S_k^*$  and  $D_{mML}^k$  are weakly consistent with  $D_M$ .

**Corollary 3.**  $S_1^*$  and  $D_{mML}^1$  are strongly consistent with  $D_M$ .

We conclude this section by noting that Algorithm 1 gives us an efficient way to compute the decision rule  $D_{mML}^k$ , which is weakly consistent with the usual maximality criterion by Theorem 2. However, we will see in the empirical experiment that the observed consistency is indeed kind of weak, and in any case weaker than for the maximin criterion, studied in the next section.

## 4. Maximin criterion

### 4.1. Definition

In the previous formulation of the problem, the minimax criterion is expressed as the fact that given a subset  $S$  to us, for each possible  $a \in S$  we select, the adversary is then free to choose  $a'$  among the remaining options  $S' = A \setminus S$  for which it will bring the highest possible gain to him (or the highest possible loss to us). It then makes sense to look for the set that minimizes this possible loss.

Conversely, given a subset  $S \subseteq A$ , we could consider that the adversary first picks any action within  $S' := A \setminus S$ , knowing that we are then free to choose any action within  $S$  to limit our loss as much as possible. This is expressed as a maximin criterion as follows:

$$MmL(S, A) := \max_{a' \in S'} \min_{a \in S} \bar{\mathbb{E}}(a' - a). \quad (29)$$

Note that  $MmL(A, A) = -\infty$  because for  $S = A$  we have  $S' = \emptyset$ , and as before we take the maximum over the empty set to be  $-\infty$ . For any  $S \subseteq A$ , the following relation holds:

$$\max_{a' \in S'} \min_{a \in S} \bar{\mathbb{E}}(a' - a) \leq \min_{a \in S} \max_{a' \in S'} \bar{\mathbb{E}}(a' - a). \quad (30)$$

Thus,  $MmL(S, A) \leq mM L(S, A)$ . We will show further that this inequality can be strict, but first we will explore some properties of  $MmL(S, A)$ .

### 4.2. Properties of $MmL(S, A)$

We can show that the maximin criterion satisfies properties that are very similar to those of the minimax criterion.

**Lemma 2.** For any  $\emptyset \neq S \subseteq T \subseteq A \in \mathcal{A}$ , we have that

$$MmL(S, A) \geq MmL(T, A).$$

**Proof.** Recall  $S' := A \setminus S$  and  $T' := A \setminus T$ . Then

$$MmL(T, A) = \max_{a' \in T'} \min_{a \in T} \bar{\mathbb{E}}(a' - a) \quad (31)$$

$$\leq \max_{a' \in T'} \min_{a \in S} \bar{\mathbb{E}}(a' - a) \quad (\text{since } S \subseteq T) \quad (32)$$

$$\leq \max_{a' \in S'} \min_{a \in S} \bar{\mathbb{E}}(a' - a) \quad (\text{since } T' \subseteq S') \quad (33)$$

$$= MmL(S, A). \quad \square \quad (34)$$

The next property shows that  $MmL(S, A)$  is negative if and only if every element in  $S'$  is dominated by some maximal element in  $S$ . This is similar but not quite the same as the corresponding property for the minimax criterion in Theorem 1: the quantifiers are swapped.

**Theorem 3.**  $MmL(S, A) < 0$  if and only if for all  $a' \in S' := A \setminus S$ , there is an  $a \in S$  such that  $a > a'$ .

**Proof.** By the definition, we have

$$\begin{aligned} MmL(S, A) < 0 &\iff \max_{a' \in S'} \min_{a \in S} \bar{\mathbb{E}}(a' - a) < 0 \\ &\iff \forall a' \in S', \exists a \in S, \bar{\mathbb{E}}(a' - a) < 0 \\ &\iff \forall a' \in S', \exists a \in S, \underline{\mathbb{E}}(a - a') > 0. \quad \square \end{aligned}$$

So, contrary to Theorem 1, the act  $a \in S$  dominating a given  $a' \in S'$  in Theorem 3 can be different for each  $a'$ . This has an important practical impact, as will be confirmed in further experiments. Theorem 3 also leads to the following corollary.

**Corollary 4.** If  $MmL(S, A) < 0$  then  $D_M(A) \subseteq S$ .

**Proof.** By Theorem 3, each  $a' \in S'$  is dominated by a maximal element in  $S$ . Therefore,  $S'$  contains no maximal elements, so consequently all maximal elements must be in  $S$ .  $\square$

**Table 3**  
Acts of Example 4.

	$\omega_1$	$\omega_2$	$\omega_3$
$a_1$	6	4	2
$a_2$	7	1	4
$a_3$	10	4	8
$a_4$	2	7	2
$a_5$	7	1	9
$a_6$	7	8	2

**Table 4**  
Values of  $\bar{E}(a_j - a_i)$ .

$j \backslash i$	1	2	3	4	5	6
1	-	3.0	0.0	4.0	3.0	-0.7
2	1.3	-	-3.0	5.0	0.0	0.6
3	4.6	3.3	-	8.0	3.0	3.9
4	3.0	6.0	3.0	-	6.0	-1.0
5	2.8	1.5	-1.8	5.6	-	2.1
6	4.0	7.0	4.0	5.0	7.0	-

Similar to the minimax criterion, we can define an optimal maximin subset of a given size as

$$S_k^+(A) := \arg \min_{\substack{\emptyset \neq S \subseteq A \\ |S| \leq k}} MmL(S, A) \tag{35}$$

and we denote the minimum value by

$$MmL_k^+(A) := MmL(S_k^+(A), A). \tag{36}$$

Note that  $S_k^+(A)$  is also not guaranteed to be unique. If this occurs, we assume one set is picked at random, as these are indifferent sets. This can be employed as a budgeted decision rule as follows:

$$D_{MmL}^k(A) := \begin{cases} D_M(A) & \text{if } MmL_k^+(A) < 0 \text{ or } |A| \leq k \\ S_k^+(A) & \text{otherwise.} \end{cases} \tag{37}$$

### 4.3. Computing $S_k^+(A)$

Let us now deal with the problem of computing  $S_k^+(A)$ . A first question that arises is then to know whether  $S_k^+(A)$  and  $S_k^*(A)$  do not coincide in general, in which case one could use Algorithm 1 to solve the problem in polynomial time. Let us first give the solution for our previous example.

**Example 3.** Consider the same situation as in Example 1. For this situation and various values of  $k$ , we have that

- $S_1^+(A) = \{a_4\}$  with  $MmL(S_1^+, A) = 3.0$ ,
- $S_2^+(A) = \{a_1, a_2\}$  with  $MmL(S_2^+, A) = 1.4$ ,
- $S_3^+(A) = \{a_1, a_2, a_3\}$  or  $\{a_2, a_3, a_4\}$  with  $MmL(S_3^+, A) = 1.0$  and
- $S_4^+(A) = \{a_1, a_2, a_3, a_4\}$  with  $MmL(S_4^+, A) = -1.1$ .

Again, in the case of  $S_3^+(A)$ , the rule would return one of the two sets, picking one at random.

In the above example, we get the same results for  $S_k^+(A)$  and values  $MmL_k^+(A)$  as for the minimax criterion. This shows, among other things, that  $S_k^+(A) \not\subseteq S_{k+1}^+(A)$  and that a greedy approach applied to maximin will generally result in sub-optimal solutions. The next example however shows that such equalities do not hold in general, and experiments will later confirm that this is often the case.

**Example 4.** Suppose that we have the same space  $\Omega$ , the credal set specified by Equation (5) with the following set of acts, see Table 3.

According to values  $\bar{E}(a_j - a_i), \forall j \neq i$  in Table 4, we see that all maximal elements of  $A$  are  $D_M(A) = \{a_3, a_6\}$  (only lines 3 and 6 of Table 4 are non-negative for every column) and the optimal solution  $S_k^*(A)$  for each  $k$  are given as follows:

- $S_1^*(A) = \{a_6\}$  with  $mML_1^*(A) = 3.9$ ,
- $S_2^*(A) = \{a_3, a_6\}$  with  $mML_2^*(A) = 2.1$ ,



- $S_3^*(A) = \{a_3, a_4, a_6\}$  with  $mML_3^*(A) = 0$ ,
- $S_4^*(A) = \{a_1, a_3, a_4, a_6\}$  with  $mML_4^*(A) = -1.8$  and
- $S_5^*(A) = \{a_1, a_3, a_4, a_5, a_6\}$  with  $mML_5^*(A) = -3.0$ .

For  $S_k^+(A)$ , we have

- $S_1^+(A) = \{a_6\}$  with  $MmL_1^+(A) = 3.9$ ,
- $S_2^+(A) = \{a_3, a_6\}$  with  $MmL_2^+(A) = -0.7$ ,
- $S_3^+(A) = \{a_1, a_3, a_6\}$  with  $MmL_3^+(A) = -1.0$ ,
- $S_4^+(A) = \{a_1, a_3, a_4, a_6\}$  with  $MmL_4^+(A) = -1.8$  and
- $S_5^+(A) = \{a_1, a_3, a_4, a_5, a_6\}$  with  $MmL_5^+(A) = -3.0$ .

This example confirms that  $mML_k^*(A) \neq MmL_k^+(A)$  is possible, even when  $S_k^*(A) = S_k^+(A)$  (see  $k = 2$  and  $k = 3$  in Example 4), and that it is possible to have  $S_k^*(A) \neq S_k^+(A)$  (see  $k = 3$  in Example 4). This confirms that our previous algorithmic solution cannot be applied to find  $S_k^+(A)$  in general. Notable exceptions where the two criteria will always coincide are when  $k = 1$  and  $k \geq |A| - 1$ , as show the next two results.

**Theorem 4.** For  $k = 1$  and all  $k \geq |A| - 1$  we have that

$$S_k^*(A) = S_k^+(A) \text{ and } mML_k^*(A) = MmL_k^+(A). \tag{38}$$

**Proof.** For  $k \geq |A|$ , by Lemmas 1 and 2,

$$S_k^*(A) = S_k^+(A) = A \text{ and } mML_k^*(A) = MmL_k^+(A) = -\infty. \tag{39}$$

For  $k = 1$ , it holds because

$$\begin{aligned} S_1^*(A) &= \arg \min_{S \in \mathcal{A}_1} mML(S, A) \\ &= \arg \min_{a \in A} \max_{a' \in A \setminus \{a\}} \overline{\mathbb{E}}(a' - a) \\ &= \arg \min_{a \in A} \max_{a' \in A \setminus \{a\}} \min_{a'' \in \{a\}} \overline{\mathbb{E}}(a' - a'') \\ &= \arg \min_{S \in \mathcal{A}_1} MmL(S, A) \\ &= S_1^+(A) \end{aligned}$$

and note that the minima are achieved at the same values in each of the above steps so  $mML_1^*(A) = MmL_1^+(A)$ .

For  $k = |A| - 1$ , it holds because (with  $n := |A|$ )

$$\begin{aligned} S_{n-1}^*(A) &= \arg \min_{S \in \mathcal{A}_{n-1}} mML(S, A) \\ &= \arg \min_{S \in \mathcal{A}_{n-1}} \left( \min_{a \in S} \max_{a' \in A \setminus S} \overline{\mathbb{E}}(a' - a) \right) \end{aligned}$$

and now, writing  $S \in \mathcal{A}_{n-1}$  as  $S = A \setminus \{a'\}$  for  $a' \in A$ ,

$$\begin{aligned} &= \arg \min_{a' \in A} \left( \min_{a \in A \setminus \{a'\}} \max_{a'' \in A \setminus (A \setminus \{a'\})} \overline{\mathbb{E}}(a'' - a) \right) \\ &= \arg \min_{a' \in A} \left( \min_{a \in A \setminus \{a'\}} \overline{\mathbb{E}}(a' - a) \right), \end{aligned}$$

and

$$\begin{aligned} S_{n-1}^+(A) &= \arg \min_{S \in \mathcal{A}_{n-1}} MmL(S, A) \\ &= \arg \min_{S \in \mathcal{A}_{n-1}} \left( \max_{a' \in A \setminus S} \min_{a \in S} \overline{\mathbb{E}}(a' - a) \right) \end{aligned}$$

and again, writing  $S \in \mathcal{A}_{n-1}$  as  $S = A \setminus \{a'\}$  for  $a' \in A$ ,

$$\begin{aligned} &= \arg \min_{a' \in A} \left( \max_{a'' \in A \setminus (A \setminus \{a'\})} \min_{a \in A \setminus \{a'\}} \overline{\mathbb{E}}(a'' - a) \right) \\ &= \arg \min_{a' \in A} \left( \min_{a \in A \setminus \{a'\}} \overline{\mathbb{E}}(a' - a) \right), \end{aligned}$$

and again note that the minima are achieved at the same values in each of the above steps so  $mML_{n-1}^*(A) = MmL_{n-1}^+(A)$ .  $\square$

Let us now consider the problem of finding  $S_k^+(A)$  as well as the corresponding  $MmL_k^+(A)$ . In order to explain our algorithm, we first consider another question: given a value  $\alpha$ , can we find an  $S \subseteq A$  of size  $k$  such that  $MmL(S, A) \leq \alpha$ ? Finding  $S_k^+(A)$  then amounts to find the lowest  $\alpha$  for which the answer is yes. Denoting as before by  $e_{ij} := \overline{\mathbb{E}}(a_j - a_i)$  the upper expectation, let us first notice that  $MmL(S, A)$  can only take as value negative infinity (if  $S = A$ ), or one of the finite values  $e_{ij}$ .

Assume we choose  $\alpha = e_{i'j'}$ . We should then put within  $S$  elements  $a_i$  such that those elements can be responses to the adversary choices leading to losses  $e_{ij}$  lower than  $\alpha$ . The set  $C_\alpha[i]$  of adversary choices for which a given element  $a_i$  is an adequate response is given as

$$C_\alpha[i] := \{j : e_{ij} \leq \alpha\}. \quad (40)$$

Consider for instance Table 4, the value  $\alpha = -1.0$  and  $i = 3$ , then  $C_{-1.0}[3] = \{2, 5\}$  corresponding to the values  $-3$  ( $e_{32}$ ) and  $-1.8$  ( $e_{35}$ ). This means that if  $a_3 \in S$ , then it is a response to adversary choices  $a_2, a_5$  (if those are not in the set  $S$ ) leading to a value lower than  $\alpha = -1.0$ . Note that we can do this operation for any  $i$  and any  $\alpha$ . Given this, there is a subset  $S$  having  $k$  elements leading to  $MmL(S, A) \leq \alpha$  only if we can find  $k$  elements whose union of sets  $C_\alpha[i]$  includes all elements in  $A \setminus S$ , or formally if we can find a set  $S$  such that

$$\cup_{i \in S} (C_\alpha[i] \setminus S) = A \setminus S. \quad (41)$$

This equation means that for every act  $a_j$  in  $S' = A \setminus S$  that the adversary can pick (right hand-side of the equation), there is at least one element  $a_i$  in  $S$  (the union of the left-hand side) we can pick such that  $\overline{\mathbb{E}}(a_j - a_i) \leq \alpha$ , i.e. one for which  $j \in C_\alpha[i]$ .

**Lemma 3.** *Let  $\alpha \in \mathbb{R} \cup \{-\infty\}$ , and  $\emptyset \neq S \subseteq A$ . Then Equation (41) implies*

$$MmL(S, A) \leq \alpha. \quad (42)$$

**Proof.** For every  $a \in S$ , define

$$C'_\alpha(S, a) := \{a' \in S' : \overline{\mathbb{E}}(a' - a) \leq \alpha\}. \quad (43)$$

By Definition,

$$\forall a \in S, \forall a' \in C'_\alpha(S, a) : \overline{\mathbb{E}}(a' - a) \leq \alpha. \quad (44)$$

Consequently,

$$\forall a' \in \cup_{a \in S} C'_\alpha(S, a), \exists a \in S : \overline{\mathbb{E}}(a' - a) \leq \alpha. \quad (45)$$

Since  $\cup_{a \in S} C'_\alpha(S, a) = S'$  by assumption, equivalently,

$$\forall a' \in S', \exists a \in S : \overline{\mathbb{E}}(a' - a) \leq \alpha \quad (46)$$

or equivalently,

$$\max_{a' \in S'} \min_{a \in S} \overline{\mathbb{E}}(a' - a) \leq \alpha \quad (47)$$

which is what we had to show, by eq. (29).  $\square$

**Lemma 4.** *Let  $1 \leq k \leq |A|$ . If eq. (41) holds for some  $\alpha \in \mathbb{R} \cup \{-\infty\}$  and  $S \subseteq A$  with  $|S| = k$ , then  $MmL_k^+(A) \leq \alpha$ .*

**Proof.** Since  $|S| = k$ , by definition of  $MmL_k^+(A)$  we have

$$MmL_k^+(A) \leq MmL(S, A). \quad (48)$$

Now use Lemma 3.  $\square$

**Lemma 5.** *Let  $1 \leq k \leq |A|$ . Then eq. (41) holds for  $\alpha = MmL_k^+(A)$  and  $S = S_k^+(A)$ , where  $|S_k^+(A)| = k$ .*

**Proof.** That  $|S_k^+(A)| = k$  follows from Lemma 2 and eq. (35).

For brevity of notation, throughout the rest of the proof, we write  $S$  for  $S_k^+(A)$  and  $\alpha$  for  $MmL_k^+(A)$ . It now suffices to show that

$$\bigcup_{a \in S} C'_\alpha(S, a) = S'. \quad (49)$$

Since each  $C'_\alpha(S, a)$  is a subset of  $S'$ , equivalently, it suffices to show that

**Table 5**  
Representation of  $C_{-1.0}[i] = \{j : e_{ij} \leq -1.0\}$  in Example 5.

$i \backslash j$	1	2	3	4	5	6
1	-	3.0	0.0	4.0	3.0	-0.7
2	1.3	-	-3.0	5.0	0.0	0.6
3	4.6	3.3	-	8.0	3.0	3.9
4	3.0	6.0	3.0	-	6.0	-1.0
5	2.8	1.5	-1.8	5.6	-	2.1
6	4.0	7.0	4.0	5.0	7.0	-

$$\forall a' \in S', \exists a \in S : a' \in C'_\alpha(S, a). \tag{50}$$

Since  $C'_\alpha(S, a) = \{a' \in S' : \bar{\mathbb{E}}(a' - a) \leq \alpha\}$ , equivalently, we must show that

$$\forall a' \in S', \exists a \in S : \bar{\mathbb{E}}(a' - a) \leq \alpha \tag{51}$$

or equivalently,

$$\max_{a' \in S'} \min_{a \in S} \bar{\mathbb{E}}(a' - a) \leq \alpha \tag{52}$$

but this translates to  $MmL(S, A) \leq \alpha$ , and we know that  $MmL(S, A) = \alpha$  by choice of  $S$  and  $\alpha$ , so the condition is satisfied.  $\square$

These last two lemmas tell us that we need to find the lowest  $\alpha \in \mathbb{R} \cup \{-\infty\}$  for which there is a set  $S$  of size  $k$  such that eq. (41) is satisfied. While the search can be performed across all values in  $\mathbb{R} \cup \{-\infty\}$ , we will in practice (see Algorithm 2 below) restrict it to the finite set of distinct values  $e_{ij}$  (and  $-\infty$ ), as  $MmL(S, A)$  can only take one of these values.

Finally, note that eq. (41) can be rewritten as follows:

$$\bigcup_{i \in S} (C_\alpha[i] \setminus S) = A \setminus S \tag{53}$$

$$\Leftrightarrow \left( \bigcup_{i \in S} C_\alpha[i] \right) \setminus S = A \setminus S \tag{54}$$

$$\Leftrightarrow \left( \bigcup_{i \in S} C_\alpha[i] \right) \cup S = A. \tag{55}$$

The next question is: what should be a range of  $\alpha$  that we are looking for? According to eq. (30), we know that for  $|A| = n$  and  $1 \leq k \leq |A|$ ,

$$MmL_k^+(A) \leq mM L_k^*(A), \tag{56}$$

where  $mM L_k^*(A)$  can be obtained by Algorithm 1, therefore,  $mM L_k^*(A)$  will be an upper bound of  $\alpha$ . For a lower bound, we notice that if we set  $\alpha$  as the  $(n - k)^{\text{th}}$  lowest value of all  $e_{ij}$ , then there are at least  $n - k$  values of  $e_{ij}$  that are lower or equal to  $\alpha$ . Thus, the corresponding set  $\cup_{i \in S} (C_\alpha[i] \setminus S)$  may contain  $n - k = |A \setminus S|$  elements for  $A \setminus S$ , and this is not true for any value lower than that.

Let us illustrate this notion on our previous example, before formalizing it into an algorithm.

**Example 5.** Consider our Example 4, together with the value  $\alpha = -1$  and  $k = 2$ . The obtained sets  $C_{-1.0}[i]$  are pictured in Table 5, with the  $\alpha$  value circled, and the values lower than it squared. We do have  $C_{-1.0}[3] = \{2, 5\}$ ,  $C_{-1.0}[6] = \{4\}$  and  $C_{-1.0}[i] = \emptyset$  for all other  $i$ 's. At best, we have for  $S = \{3, 6\}$  that  $C_{-1.0}[3] \cup C_{-1.0}[6] = \{2, 5, 4\}$  which is different from  $A \setminus S = \{1, 2, 5, 4\}$ . We therefore cannot find a set  $S$  of two elements such that  $MmL(S, A) \leq -1.0$ . Note that  $-1.0$  is not the 4<sup>th</sup> (or  $(n - k)^{\text{th}}$ ) lowest value of all  $e_{ij}$ , but it is the 3<sup>rd</sup> lowest value of all  $e_{ij}$ . Therefore, there are only three values lower or equal to  $-1.0$ , and the set  $\cup_{i \in S} (C_{-1.0}[i] \setminus S)$  cannot include, by definition, 4 elements for  $A \setminus S$  as we want.

The next value after  $-1.0$  in Table 4 is  $-0.7$ , which is the 4<sup>th</sup> lowest value of all  $e_{ij}$  for which there are 4 values lower or equal to  $-0.7$ . Therefore, the corresponding set  $\cup_{i \in S} (C_{-0.7}[i] \setminus S)$  will have a potential to be a set of 4 elements for  $A \setminus S$ . Specifically, the corresponding sets  $C_{-0.7}[i]$  are represented in Table 6 and are  $C_{-0.7}[3] = \{2, 5\}$ ,  $C_{-0.7}[6] = \{1, 4\}$  and  $C_{-1.0}[i] = \emptyset$  for all other  $i$ 's. This time, we do have for  $S = \{3, 6\}$  that  $C_{-0.7}[3] \cup C_{-0.7}[6] = \{1, 2, 5, 4\} = A \setminus S$ , meaning that we can find an  $S$  with  $MmL(S, A) \leq -0.7$ , and since this is the lowest value for which we can, we have  $S_2^+ = \{3, 6\}$ .

We are now ready to put these ideas into a corresponding algorithm. Algorithm 2 provides the iterative strategy to find  $S_k^+(A)$  and  $MmL_k^+(A)$ , by finding the lowest value for which the condition of Equation (41) is satisfied. Algorithm 3 simply checks that this condition can be satisfied.

**Table 6**  
Representation of  $C_{-0.7}[i] = \{j : e_{ij} \leq -0.7\}$  in Example 5.

$i \backslash j$	1	2	3	4	5	6
1	-	3.0	0.0	4.0	3.0	-0.7
2	1.3	-	-3.0	5.0	0.0	0.6
3	4.6	3.3	-	8.0	3.0	3.9
4	3.0	6.0	3.0	-	6.0	-1.0
5	2.8	1.5	-1.8	5.6	-	2.1
6	4.0	7.0	4.0	5.0	7.0	-

---

**Algorithm 2** Finding  $S_k^+(A)$ .

---

**Require:**  $A = \{a_1, a_2, \dots, a_n\}$ ,  $\mathcal{P}$ ,  $k$   
**Ensure:**  $MmL_k^+(A)$ ,  $S_k^+(A)$

- 1: **for**  $j = 1 : n$  **do**
- 2:     **for**  $i = 1 : n, i \neq j$  **do**
- 3:         compute  $e_{ij} := \bar{E}(a_j - a_i)$
- 4:     **end for**
- 5: **end for**
- 6:  $S_k^+(A) \leftarrow \emptyset$
- 7:  $m \leftarrow n - k$
- 8: **while**  $S_k^+(A) = \emptyset$  **do**
- 9:      $e' \leftarrow e_{i'j'}$  such that  $e_{i'j'}$  is the  $m^{\text{th}}$  lowest value of all  $e_{ij}$
- 10:     **for**  $i = 1 : n$  **do**
- 11:          $C[i] \leftarrow C_{e'}[i] := \{j : e_{ij} \leq e'\}$ .
- 12:     **end for**
- 13:      $S_k^+(A) \leftarrow$  result of Algorithm 3 with  $C[i]$  and  $e'$
- 14:      $m \leftarrow m + 1$
- 15: **end while**
- 16:  $MmL_k^+(A) \leftarrow e'$
- 17: **return**  $S_k^+(A)$ ,  $MmL_k^+(A)$

---



---

**Algorithm 3** Reachable value  $\alpha$ .

---

**Require:**  $C_\alpha[i]$  for  $i = 1, \dots, n, k$   
**Ensure:** A non-empty subset  $S$  if there is a solution,  $\emptyset$  if not

- 1:  $S \leftarrow \emptyset$
- 2: **for** every subset  $T \subseteq \{1, \dots, n\}$  with  $|T| = k$  **do**
- 3:     **if**  $\cup_{i \in T} C_\alpha[i] \setminus T = A \setminus T$  **then**
- 4:          $S \leftarrow T$  **break for**
- 5:     **end if**
- 6: **end for**
- 7: **return**  $S$

---

Looking at the algorithms, it is clear that Algorithm 3 has a combinatorial nature, and represents a bottleneck in our approach. The next result indicates that this part of the Algorithm is indeed a computational barrier.

**Theorem 5.** *Checking whether the condition of Equation (41) can be satisfied is NP-complete.*

**Proof.** To show this, we will show that it is equivalent to solving a dominating set problem, which is a known NP-complete problem. We can rephrase the problem we try to solve as having a set  $[n] = \{1, \dots, n\}$  of integer indices, to which are associated (possibly empty) subsets  $A_i \subseteq [n]$  such that  $i \notin A_i$ . For a given  $k \leq n$ , we want to solve the following optimisation problem:

Find a subset  $B \subseteq [n]$  of indices such that

$$|B| = k, \tag{57}$$

$$\cup_{i \in B} A_i = \{1, \dots, n\} \setminus B. \tag{58}$$

In the above reformulation, the subset  $B$  corresponds to  $S$ , and the subsets  $A_i$  to the subsets  $C_\alpha[i]$ . We can then consider the directed graph  $G = (V, E)$  where the vertices are  $V = \{1, \dots, n\}$ , and where there is an edge  $(i, j) \in E$  whenever  $j \in A_i$  ( $A_i$  are the out-neighbours of  $i$ ). Now, there is straightforward reduction of our problem to a dominating set problem, in the sense that taking  $i$  in the dominating set is equivalent to taking  $A_i$  in  $B \Rightarrow A_i$  “dominates”  $i$  and every index in  $A_i$ .

So, identifying whether there is a dominating set of size  $k$  in  $G$  is equivalent to identify whether there is a solution to our problem, hence the two problems have the same complexity.  $\square$

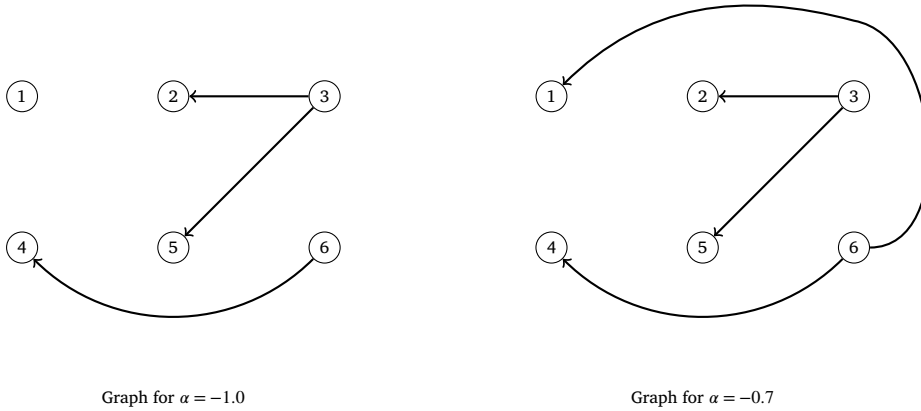


Fig. 2. Corresponding graphs of Example 5.

Fig. 2 illustrates the equivalent graphs of the sets used in Example 5. One can easily see that for  $\alpha = -1.0$ , there is no way to pick a cover for  $k = 2$ , i.e., two vertices  $i, j$  such that all other vertices except  $i, j$  are their direct neighbours (one would need to pick at least three vertices). In contrast, this is possible in the graph corresponding to  $\alpha = -0.7$ .

Such a result clearly goes against trying to find the maximin approach rather than the minimax one, at least in terms of computability and with the algorithms we have presented. However, if  $k$  remains of low value (which we expect to be the case in most applications of a budgeted rule), the computational burden will remain limited. Also, we will see in the experiments that the maximin approach is very often (but not always) strongly consistent with  $D_M$  and the maximality decision rule in practice, which is not the case for the minimax one. Before that, we will show that the maximin rule is also always weakly consistent with  $D_M$

#### 4.4. Weak consistency of maximin criterion

Similarly to  $S_k^*$ , we will show that  $S_k^+$  is weakly consistent with maximality. Weak consistency for  $A$  such that  $MmL_k^+(A) < 0$  is ensured by Corollary 4, therefore we now look at the case  $MmL_k^+(A) \geq 0$ .

**Theorem 6.** For any  $S \subseteq A$  with cardinality  $k$ , there is a  $T \subseteq A$  with cardinality  $k$  such that  $T \cap D_M(A) \neq \emptyset$  and  $MmL(T, A) \leq MmL(S, A)$ .

**Proof.** If  $S \cap D_M(A) \neq \emptyset$  then we can take  $T = S$  and we are done. For the remainder of the proof, we can therefore assume that  $S \cap D_M(A) = \emptyset$ . It suffices to construct a subset  $T$  of  $A$  with cardinality  $k$  such that  $T \cap D_M(A) \neq \emptyset$  and  $MmL(T, A) \leq MmL(S, A)$ , establishing the desired result.

For brevity, define  $S' := A \setminus S$ . If  $S \cap D_M(A) = \emptyset$ , then  $MmL(S, A) \geq 0$  by Theorem 3. There are  $a_+ \in S$  and  $a'_+ \in S'$  such that

$$0 \leq MmL(S, A) = \overline{\mathbb{E}}(a'_+ - a_+). \tag{59}$$

Since  $S \cap D_M(A) = \emptyset$ , and  $a_+ \in S$ , we know that  $a_+$  is non-maximal in  $A$ . Therefore,  $a_+$  must be dominated by a maximal element of  $A$ , which is by assumption guaranteed to belong to  $S'$ . In other words, there is a  $b \in S' \cap D_M(A)$  such that

$$\overline{\mathbb{E}}(a_+ - b) < 0. \tag{60}$$

Define  $T := (S \setminus \{a_+\}) \cup \{b\}$ . Clearly,  $T$  has the same cardinality as  $S$ , since  $a_+ \in S$  and  $b \notin S$  by construction, and  $T \cap D_M(A) \neq \emptyset$  since  $b \in D_M(A)$ . Since  $MmL(S, A) = \overline{\mathbb{E}}(a'_+ - a_+)$ , also by construction, it suffices to show that

$$MmL(T, A) = \max_{a' \in T'} \min_{a \in T} \overline{\mathbb{E}}(a' - a) \leq \overline{\mathbb{E}}(a'_+ - a_+) \tag{61}$$

to finish our proof. In other words, we are left to prove that

$$\forall a' \in T', \exists a \in T : \overline{\mathbb{E}}(a' - a) \leq \overline{\mathbb{E}}(a'_+ - a_+). \tag{62}$$

Indeed, by the definition of  $MmL(S, A)$  we already know that

$$\forall a' \in S', \exists a \in S, \overline{\mathbb{E}}(a' - a) \leq \overline{\mathbb{E}}(a'_+ - a_+). \tag{63}$$

First note that we have  $\underline{\mathbb{E}}(b - a_+) > 0$ , due to Equation (60) and the duality relation  $\underline{\mathbb{E}}(f) = -\overline{\mathbb{E}}(-f)$ . From this we can deduce

$$\overline{\mathbb{E}}(b - a) \geq \underline{\mathbb{E}}(b - a_+) + \overline{\mathbb{E}}(a_+ - a) > \overline{\mathbb{E}}(a_+ - a) \tag{64}$$

so when  $a' = b$  in eq. (63), we can replace  $a'$  with  $a_+$  whilst still respecting the inequality. Therefore,

$$\forall a' \in T', \exists a \in S : \bar{E}(a' - a) \leq \bar{E}(a'_+ - a_+). \tag{65}$$

Now fix any  $a' \in T'$ . If the  $a \in S$  in the above condition is not equal to  $a_+$ , then  $a \in T$  and we are done. Otherwise, if  $a = a_+$ , note that again we can replace it with  $b$  whilst still respecting the inequality, because, since  $\bar{E}(a_+ - b) > 0$ ,

$$\bar{E}(a' - b) \leq \bar{E}(a' - a_+) + \bar{E}(a_+ - b) < \bar{E}(a' - a_+). \tag{66}$$

This concludes the proof.  $\square$

Theorem 6 tells us that whenever we have a set of  $k$  elements without any maximal ones, we can add a maximal ones and improve the solution by decreasing its  $MmL$  value. This is sufficient to prove that at least one set  $S_k^+(A)$  contains a maximal element. Consequently, we have the following corollaries from Theorem 6 and Corollary 4.

**Corollary 5.**  $S_k^+$  and  $D_{MmL}^k$  are weakly consistent with  $D_M$ .

**Corollary 6.**  $S_1^+$  and  $D_{MmL}^1$  are strongly consistent with  $D_M$ .

After investigating the properties of our two budgeted regret-based rules, as well as the computational methods to obtain them, we are ready to proceed to some experiments to test some of their behaviours.

### 5. Computational experimentation

In this section, we will perform some experiments to compare  $S_k^*$  and  $S_k^+$  with their greedy approximations  $S_k^{g*}$  and  $S_k^{g+}$ , obtained as follows: given a set of actions  $A$ , we compute  $S_1^{g*}(A) = S_1^*(A)$ . Then, for  $k \geq 2$ , we compute  $S_k^{g*}(A) = S_{k-1}^{g*}(A) \cup S_1^*(A \setminus S_{k-1}^{g*}(A))$  and determine  $mML(S_k^{g*}, A)$ . Similarly, we compute  $S_1^{g+}(A) = S_1^+(A)$  and determine  $MmL(S_k^{g+}, A)$ . Then, for  $k \geq 2$ , we can compute  $S_k^{g+}(A) = S_{k-1}^{g+}(A) \cup S_1^+(A \setminus S_{k-1}^{g+}(A))$  and determine  $mML(S_k^{g+}, A)$ . Recall that the greedy algorithms are not optimal and will usually not result in  $S_k^*$  and  $S_k^+$ , as already shown in Examples 2 and 4.

We aim to compare  $S_k^*$  and  $S_k^+$  with their greedy approach for consistency with respect to maximality. Basically, we want to find out how much  $S_k^*$ ,  $S_k^{g*}$ ,  $S_k^+$  and  $S_k^{g+}$  can capture maximal elements in the set  $D_M$ . This can measure the quality of the greedy approximations with our proposed algorithms to find  $S_k^*$  and  $S_k^+$ . Since  $S_1^*$  and  $S_1^+$  are weakly and strongly consistent with  $D_M$  by Corollaries 2 and 3 and Corollaries 5 and 6 respectively, we will not consider case  $k = 1$ .

We fix  $|A| = 20$ ,  $|\Omega| = 5$ ,  $|\mathcal{P}| = 20$ . We generate  $p \in \mathcal{P}$  by sampling a probability mass function  $p$  uniformly from the unit simplex as follows. For each  $\omega$ , we sample  $q(\omega)$  uniformly from  $(0,1)$  and then for each  $\omega$ , we assign  $p(\omega) := \frac{\ln q(\omega)}{\sum_{\omega'} \ln q(\omega')}$ . This ensures that generated distributions  $p(\omega_1), \dots, p(\omega_n)$  follow a Dirichlet  $\text{Dir}(1, \dots, 1)$  distribution which has uniform density over the unit simplex [8]. Note that the convex hull of  $\mathcal{P}$  may have less than twenty extreme points, as some generated probabilities may be convex combinations of the others. Next, we generate a set of elements  $A$  on  $\Omega$  for which  $|D_M(A)| = 6$  by using algorithm 6 in [20]. We will compute  $S_k^*$ ,  $S_k^{g*}$ ,  $S_k^+$  and  $S_k^{g+}$  for the cases  $k \in \{2, \dots, 6\}$ . To do so, for all  $a_i \neq a_j \in A$ , we compute  $\bar{E}(a_j - a_i)$  with respect to the credal set  $\mathcal{P}$  as an input to our algorithms and compute  $S_k^*$ ,  $S_k^{g*}$ ,  $S_k^+$  and  $S_k^{g+}$  with respect to  $A$ . Next, we check whether  $S_k^*$ ,  $S_k^{g*}$ ,  $S_k^+$  and  $S_k^{g+}$  are weakly consistent (having only one maximal element in the set) or strongly consistent (having all elements in the set being maximal) with  $D_M$  or not. We also compute the proportion of elements in  $S_k^*$ ,  $S_k^{g*}$ ,  $S_k^+$  and  $S_k^{g+}$  that are in  $D_M$ . We repeat this process 500 times and summarise the result in Table 7. The percentages of these sets that satisfy weakly and strongly consistent properties and the average percentages of elements in these sets which are in  $D_M$  are showed in the 3<sup>rd</sup>-5<sup>th</sup> columns of Table 7.

According to the results, we see that  $S_k^*$  and  $S_k^+$  and their greedy approach are weakly consistent with  $D_M$ . Interestingly, only  $S_k^+$  is likely to be strongly consistent while the rest of the sets are rarely strongly consistent with  $D_M$  (the numbers in the fourth column quickly drop for all sets as  $k$  increase, while they actually increase for  $S_k^+$ ). Moreover, the average percentages of maximal elements in  $S_k^*$  and  $S_k^+$  are higher than in their greedy approximations  $S_k^{g*}$  and  $S_k^{g+}$ .

By Theorem 4 and the procedure to construct  $S_k^{g*}$  and  $S_k^{g+}$ , we have  $S_k^{g*} = S_k^*$  for all  $k$ . Therefore, the percentages of  $S_k^{g*}$  and  $S_k^{g+}$  that are weakly and strongly consistent with  $D_M$  are equal. In addition, to see how close those greedy approximations are to  $S_k^*$  and  $S_k^+$ , we compare the optimal solutions with the greedy approach solutions. To do so, we record the number of  $S_k^* = S_k^{g*}$  and the number of  $S_k^+ = S_k^{g+}$  and present the averages percentages of these sets that satisfy these conditions in the column 6<sup>th</sup>. In the column 7<sup>th</sup>, we calculate the average of proportion of elements in  $S_k^{g*}$  that are in  $S_k^*$  and the proportion of elements in  $S_k^{g+}$  that are in  $S_k^+$ .

As we can employ  $mML$  and  $MmL$  as budgeted decision rules as in Equations (19) and (37), we want to find out how fast  $S_k^*$  and  $S_k^+$  will become supersets of  $D_M$  (or  $mML_k^*(A) < 0$  and  $MmL_k^+(A) < 0$ ) so that we can simply return  $D_M$  instead of  $S_k^*$  or  $S_k^+$ . To do so, we regenerate a set of elements  $A$  on  $\Omega$  such that  $|D_M| = m$  for  $m \in \{2, 5, 10\}$ . Next, we compute  $mML_k^*(A)$  and  $MmL_k^+(A)$  for  $k = m + i$ , where  $i \in \{0, 1, 2, 3\}$ . We repeat this process 100 times and present the result in Table 8. The average percentages of  $mML_k^*(A) < 0$  and  $MmL_k^+(A) < 0$  are presented in the 2<sup>nd</sup> and 3<sup>rd</sup> columns while the average percentages of  $mML_k^*(A) = MmL_k^+(A)$  are presented in the 4<sup>th</sup> column of Table 8. According to the result, we found that  $S_k^+$  becomes a superset of  $D_M$  much faster than

**Table 7**  
Percentages averages of  $S_k^*$ ,  $S_k^{g*}$ ,  $S_k^+$  and  $S_k^{g+}$  that satisfy different conditions.

$D^k$	$k$	w.c.	s.c.	$\frac{ D^k \cap D_M }{ D^k }$		
$S_k^*$	2	100%	100%	100%	$S_k^+ = S_k^{g*}$	$\frac{ S_k^{g*} \cap S_k^+ }{ S_k^{g*} }$
	3	100%	90.2%	96.7%		
	4	100%	73.8%	92.7%		
	5	100%	57.0%	89.8%		
	6	100%	44.0%	88.4%		
$S_k^{g*}$	2	100%	94.8%	97.4%	32.2%	64.3%
	3	100%	78.6%	92.5%	13.0%	61.4%
	4	100%	58.2%	87.8%	9.2%	65.9%
	5	100%	37.0%	82.9%	7.0%	69.4%
	6	100%	15.8%	77.9%	9.4%	73.6%
$S_k^+$	2	100%	99.2%	99.6%	$S_k^+ = S_k^{g+}$	$\frac{ S_k^{g+} \cap S_k^+ }{ S_k^{g+} }$
	3	100%	99.8%	99.9%		
	4	100%	100%	100%		
	5	100%	100%	100%		
	6	100%	100%	100%		
$S_k^{g+}$	2	100%	94.8%	97.4%	32.6%	63.0%
	3	100%	78.6%	92.5%	18.6%	65.8%
	4	100%	58.2%	87.8%	16.4%	71.1%
	5	100%	37.0%	82.9%	15.2%	74.9%
	6	100%	15.8%	77.9%	15.8%	77.9%

**Table 8**  
Percentages averages of  $S_k^*$  and  $S_k^+$  that satisfy different conditions.

$ D_M $	$k$	$mML_k^*(A) < 0$	$MmL_k^+(A) < 0$	$mML = MmL$
2	2	54%	100%	46%
	3	76%	100%	51%
	4	89%	100%	51%
	5	96%	100%	54%
5	5	33%	100%	23%
	6	64%	100%	25%
	7	85%	100%	31%
	8	96%	100%	34%
10	10	28%	100%	15%
	11	66%	100%	22%
	12	84%	100%	19%
	13	97%	100%	24%

$S_k^*$  as the average percentages of  $MmL_k^+(A) < 0$  are much higher than the average percentages of  $mML_k^*(A) < 0$ . If the cardinality of  $D_M$  is increasing, then the average percentages of  $mML_k^*(A) = MmL_k^+(A)$  tend to be decreasing.

All those numbers show that, in practice, the maximin approach has a quite stronger consistency with maximality, and its negativity can be used as a quite reliable signal that we have captured all the maximal elements. In contrast, the minimax rule shows a much weaker consistency, and will often contain non-maximal elements. As maximality rests on very strong theoretical foundations, our conclusion is that the maximin rule should be preferred whenever its computational burden remains affordable, and that one should resort to the minimax rule only when computational efficiency is a key issue.

## 6. Two illustrative use cases

In this section, we demonstrate how our method can be applied in practice. The first example is inspired from Jansen et al. [10], but adapted to provide more than 3 maximal acts, while the second one concerns a situation where we must predict binary vectors over a set of labels, which is the situation encountered in multi-label learning, a specific multi-task machine learning problem.

### 6.1. Financial investment example

We follow the financial application example in [10], where a subject wants to invest her money in some stocks. An act corresponds to investing in a specific stock. Suppose that a financial agent offers her ten different stocks, so we have  $A = \{a_1, \dots, a_{10}\}$ . There are five possible states of nature  $\Omega = \{\omega_1, \dots, \omega_5\}$  corresponding to different economic scenarios that are uncertain to the agent. The payoff for each stock under each possible scenarios is given in Table 9.

In addition, based on the decision maker's experience in the financial market, she specifies her credal set through probability bounds as follows (note these are slightly wider than the ones from Jansen et al. [10] to better demonstrate the benefits of our method):

**Table 9**  
Payoffs for the acts of the financial application example from [10].

	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$
$a_1$	37	25	23	73	91
$a_2$	50	67	2	44	94
$a_3$	60	4	96	1	83
$a_4$	16	24	31	26	100
$a_5$	3	86	76	85	11
$a_6$	12	49	66	56	14
$a_7$	39	10	92	88	57
$a_8$	62	52	80	71	42
$a_9$	90	8	74	70	38
$a_{10}$	63	68	36	69	9

**Table 10**  
Values of  $\bar{E}(a_j - a_i)$ .

$j \setminus i$	1	2	3	4	5	6	7	8	9	10
1	-	11.65	25.0	23.5	22.85	29.35	2.9	4.7	9.8	19.5
2	5.0	-	21.6	20.7	21.5	28.4	6.9	3.0	9.6	14.3
3	4.05	7.3	-	15.95	20.8	26.35	-3.0	-1.4	0.3	16.65
4	-7.4	-4.25	6.5	-	8.95	15.0	-7.4	-11.2	-3.0	4.35
5	16.2	22.0	33.1	34.8	-	19.8	2.6	2.6	10.6	12.2
6	-3.6	2.2	13.3	15.0	-5.55	-	-16.2	-17.2	-9.2	-3.1
7	16.15	26.3	30.5	37.75	23.8	30.4	-	8.7	10.6	26.95
8	19.0	23.45	32.3	34.95	23.1	28.7	8.0	-	8.6	18.5
9	18.9	26.95	30.3	39.25	27.0	32.85	5.8	4.6	-	20.25
10	10.0	11.6	27.2	27.2	8.5	19.5	2.7	-4.8	-0.5	-

**Table 11**  
 $S_k^*$  and  $S_k^+$  for different values of  $k$ .

$k$	$S_k^*(A)$	$mML_k^*(A)$	$S_k^+(A)$	$MmL_k^+(A)$
1	$a_7$	8	$a_7$	8
2	$a_7, a_8$	4.7	$a_7, a_8$	4.6
3	$a_1, a_7, a_8$	4.6	$a_7, a_8, a_9$	3
4	$a_2, a_7, a_8, a_9$	2.9	$a_2, a_7, a_8, a_9$	2.9
5	$a_1, a_2, a_7, a_8, a_9$	2.6	$a_1, a_2, a_7, a_8, a_9$	2.6
6	$a_1, a_2, a_5, a_7, a_8, a_9$	-1.4	$a_1, a_2, a_5, a_7, a_8, a_9$	-3.0

$$P = \{p \in \mathbb{P} : 0.1 \leq p(\omega_1) \leq 0.3, 0.05 \leq p(\omega_2) \leq 0.2, 0.1 \leq p(\omega_3) \leq 0.2, 0.2 \leq p(\omega_4) \leq 0.4, 0.1 \leq p(\omega_5) \leq 0.3\}. \tag{67}$$

From these bounds, the values of  $\bar{E}(a_j - a_i)$ , for all  $j \neq i$ , can be calculated by linear programming. They are provided in Table 10.

We see that while our information can discard some items as being non-maximal, we still have  $\{a_1, a_2, a_5, a_7, a_8, a_9\}$  for the set of maximal acts, which may be judged too high if these represent complex financial portfolios. Results of our algorithms are shown in Table 11. In this case, the two approaches only differ slightly, and are completely consistent with the notion of maximality, as all selected examples for  $k \leq 6$  are maximal.

### 6.2. Multi-label example

To demonstrate our approach in machine learning, we consider multi-label classification [13], a sub-category of multi-task learning where one first observes an input  $x$  and then has to predict binary vectors over a set of labels  $\mathcal{L} := \{\mathcal{L}_1, \dots, \mathcal{L}_n\}$  with  $\ell_i \in \{0, 1\}$ . A value  $\ell_i = 1$  usually means that the label is present in the instance  $x$ , while a zero means that the label is absent. There are many learning schemes to solve this problem, including imprecise ones [3,17,1], as well as many commonly used loss functions.

As is classically done in machine learning, our sets of acts and of states will coincide, acts being prediction of the possible ground truth when observing  $x$ . We will therefore have  $\Omega = \mathcal{A} = \{0, 1\}^n$ , with an exponentially increasing size of sets of alternatives as a function of the number of labels  $n$ . We will also consider here the utility version of the standard Hamming loss function, which means that if  $\hat{\mathcal{L}}$  is the predicted vector (i.e., the act, in our setting), and  $\mathcal{L}$  the true one (i.e., the state, in our setting), then

$$\hat{\mathcal{L}}(\mathcal{L}) = - \sum_{i=0}^n \mathbb{1}_{\hat{\mathcal{L}}_i \neq \mathcal{L}_i},$$

where  $\mathbb{1}_A$  denotes the indicator function of event  $A$ . Table 12 illustrates the obtained utility matrix when  $n = 3$ . As a model, we will consider in this example an imprecise version of the classical binary relevance scheme, where the probability mass of a given vector  $\ell$  is the product of label-wise probability masses, i.e.,



**Table 12**  
Example of multi-label utility based on Hamming distance for  $n = 3$ .

	[000]	[001]	[010]	[011]	[100]	[101]	[110]	[111]
[000]	0	-1	-1	-2	-1	-2	-2	-3
[001]	-1	0	-2	-1	-2	-1	-3	-2
[010]	-1	-2	0	-1	-2	-3	-1	-2
[011]	-2	-1	-1	0	-3	-2	-2	-1
[100]	-1	-2	-2	-3	0	-1	-1	-2
[101]	-2	-1	-3	-2	-1	0	-2	-1
[110]	-2	-3	-1	-2	-1	-2	0	-1
[111]	-3	-2	-2	-1	-2	-1	-1	0

**Table 13**  
Extreme probabilities of multi-label example.

	$p([000])$	$p([001])$	$p([010])$	$p([011])$	$p([100])$	$p([101])$	$p([110])$	$p([111])$
$p_1$	0.432	0.048	0.108	0.012	0.288	0.032	0.072	0.008
$p_2$	0.144	0.336	0.036	0.084	0.096	0.224	0.024	0.056
$p_3$	0.216	0.024	0.324	0.036	0.144	0.016	0.216	0.024
$p_4$	0.072	0.168	0.108	0.252	0.048	0.112	0.072	0.168
$p_5$	0.144	0.016	0.036	0.004	0.576	0.064	0.144	0.016
$p_6$	0.048	0.112	0.012	0.028	0.192	0.448	0.048	0.112
$p_7$	0.072	0.008	0.108	0.012	0.288	0.032	0.432	0.048
$p_8$	0.024	0.056	0.036	0.084	0.096	0.224	0.144	0.336

**Table 14**  
Upper pairwise expectations for the multi-label example.

	[000]	[001]	[010]	[011]	[100]	[101]	[110]	[111]
0.0	0.8	0.6	1.4	0.2	1.0	0.8	1.6	
0.4	0.0	1.0	0.6	0.6	0.2	1.2	0.8	
0.2	1.0	0.0	0.8	0.4	1.2	0.2	1.0	
0.6	0.2	0.4	0.0	0.8	0.4	0.6	0.2	
0.6	1.4	1.2	2.0	0.0	0.8	0.6	1.4	
1.0	0.6	1.6	1.2	0.4	0.0	1.0	0.6	
0.8	1.6	0.6	1.4	0.2	1.0	0.0	0.8	
1.2	0.8	1.0	0.6	0.6	0.2	0.4	0.0	

$$p(\mathcal{L}) = \prod_{i=1}^n p_i(\mathcal{L}_i) \tag{68}$$

Here, the  $p_i$ 's values are outputted by classifiers. We will consider here that instead of having precise label-wise estimates, we have imprecise ones given as follows:

$$p_1(\mathcal{L}_1 = 1) \in [0.4, 0.8], \quad p_2(\mathcal{L}_2 = 1) \in [0.2, 0.6], \quad p_3(\mathcal{L}_3 = 1) \in [0.1, 0.7].$$

By robustifying the product in Eq. (68), we get a set of extreme probabilities obtained by considering all the combinations of interval bounds. Those are summarised in Table 13.

All is left to do now is to compute the pairwise matrix of upper expectations, which is summarised in Table 14. From this table it is clear that all alternatives are maximal, hence using a robust decision rule such as maximality is not helpful at all to select an alternative. In general, we would expect the number of maximal vectors to be quite large but not equal to  $\Omega$ . Also note that due to the simplicity of the model and the highly structured form of the utility matrix in Table 12, upper expectations only take a limited number of values. This would however not be the case in actual applications, where an analyst would probably consider more complex models, as well as utility or loss functions that would be different for different label mistakes.

Assuming that we want to return only two vectors, let us first consider the minimax criterion. Applying Algorithm 1 to Table 14, we get that the minimal second largest element is obtained for column [100] (the only one for which it is below 1), with  $S_2^* = \{[100], [011]\}$  and  $mML(S_2^*) = 0.6$ . Considering now the maximin criterion, if we fix  $\alpha = 0.4$  (from the  $mML$  value, we already know it is equal or lower than 0.6), we get  $C_{0.4}([100]) = \{[000], [010], [101], [110]\}$  and  $C_{0.4}([101]) = \{[001], [011], [111]\}$ , that do provide a dominating set, and one can check  $S_2^+ = \{[100], [101]\}$  and  $MmL(S_2^+) = 0.4$ . Despite the high structure of the problem, the two approaches deliver distinct, unique solutions. They also seem to adopt different strategies, the minimax recommending very diverse, complementary vectors, while the maximin sends back two similar vectors, that differ only by the label that is the most uncertain ( $\mathcal{L}_3$ ).

### 7. Discussion and conclusion

In this study, we have studied  $k$ -budgeted regret-based decision rules that return an optimal subset of size  $k$  with respect to some value function. To do so, we recalled minimax criteria which minimizes the maximal gain to the adversary on a given set of

alternatives and proposed a new regret-based decision rule called maximin criteria which swaps the order of selecting alternatives by the decision maker and the adversary. We also provided algorithms for both criteria and discussed their properties with respect to maximality. Note that our framework can be extended straightforwardly to continuous spaces as long as one can compute upper expectations over it, while extending it to an infinite set of acts would be trickier.

From the experimental perspective, we compared the minimax and maximin criteria and their greedy approximation on generated sets. Overall, both algorithms perform better than their greedy approximation. We also observed that the maximin criteria can capture all maximal elements faster than the minimax criteria and are more consistent with the maximality decision rule. However, the computational complexity of our proposed algorithm for maximin criteria is much harder than for the minimax criteria, as we showed that the former requires to solve an NP-complete problem, while the latter can be solved in polynomial time. This drawback is however of limited importance for small values of  $k$ . So, one should clearly prefer the maximin approach as long as it is computationally affordable, and otherwise take the minimax approach.

Note that whilst we are not aware of other works considering the specific problem of delivering at most  $k$  alternatives when modelling uncertainty as a credal set, some works provide methods and ideas that could easily be leveraged to do so. A first strand of works in this direction would consist in considering nested models where the imprecision is controlled by some parameter, allowing one to go from a precise probability mass function (and having a unique optimal action) to the full credal set. Jansen et al. [9] consider such a parametric model for imprecisely defined utility functions (that we assume here well-defined), further discussed by Miranda et al. [16] in combination with credal set approximation. Those same authors also discuss the notion of centroid for credal sets [15], and it would be easy to go from this to the idea of a nested sequence. A second proposal is to associate an evaluation to each possible maximal and/or E-admissible (a concept we did not consider here) alternatives [10], to rank-order them according to this evaluation and to take the top- $k$  alternatives. Such approaches also appear legitimate to solve the issue considered in this paper, with a different philosophy. In particular, they would consider alternatives individually, rather than offering an evaluation for a whole set of alternatives. One risk is then that potential interactions between alternatives would be ignored, and that similar alternatives could be selected in the  $k$  retained ones (or, in the case of E-admissibility degrees, that none of a collection of similar alternatives would be retained, while it would probably be desirable to retain at least one of those). On the other hand, one advantage over our approach is that such approaches makes it easier to be coherent with existing decision rules, and some of them can be solved efficiently. Finally, we also think that regret formulations are well-known and appealing to many researchers, which is also an argument to develop such methods in addition to others. One of our future endeavour would be to compare those various approaches, both from theoretical and practical perspectives, yet this would require to discuss desirable properties of credal budgeted decision rules, something that is out of the scope of the current paper.

In other future work, we can look at a more practical perspective, for example, we may apply these proposed budgeted rules to actual decision problems such as machine learning with structured outputs or system design, where the decision maker is limited by human cognitive limits or where inspecting more closely the different proposed options can lead to a high monetary cost.

### CRedit authorship contribution statement

**Nawapon Nakharutai:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Sébastien Destercke:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Matthias C.M. Troffaes:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nawapon Nakharutai reports financial support was provided by National Research Council of Thailand. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgement

This project is funded by National Research Council of Thailand (NRCT). This research was supported by Chiang Mai University, as well as by the fondation partenariale de l'université de technologie de Compiègne, through the safe AI chair. NN would like to thank Assoc. Prof. Manad Khomkong for his support. We also thank Tom Davot for discussions about the complexity proof.

### References

- [1] Y.C.C. Alarcón, S. Destercke, Multi-label chaining with imprecise probabilities, in: *European Conference on Symbolic and Quantitative Approaches with Uncertainty*, Springer, 2021, pp. 413–426.
- [2] F.J. Anscombe, R.J. Aumann, A definition of subjective probability, *Ann. Math. Stat.* 34 (1963) 199–205.
- [3] A. Antonucci, G. Corani, The multilabel naive credal classifier, *Int. J. Approx. Reason.* 83 (2017) 320–336.
- [4] J.M. Aughenbaugh, C.J.J. Paredis, The value of using imprecise probabilities in engineering design, *J. Mech. Des.* 128 (2005) 969–979.
- [5] T. Augustin, F.P.A. Coolen, G. de Cooman, M.C.M. Troffaes, *Introduction to Imprecise Probabilities*, John Wiley & Sons, 2014.

- [6] H. Bains, A. Madariaga, M.C.M. Troffaes, B. Kazemtabrizi, An economic model for offshore transmission asset planning under severe uncertainty, *Renew. Energy* 160 (2020) 1174–1184, <https://doi.org/10.1016/j.renene.2020.05.160>.
- [7] C. Denis, M. Hebiri, Confidence sets with expected sizes for multiclass classification, *J. Mach. Learn. Res.* 18 (2017) 3571–3598.
- [8] S. Feng, *The Poisson-Dirichlet Distribution and Related Topics*, Springer Berlin Heidelberg, 2010.
- [9] C. Jansen, G. Schollmeyer, T. Augustin, Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences, *Int. J. Approx. Reason.* 98 (2018) 112–131.
- [10] C. Jansen, G. Schollmeyer, T. Augustin, Quantifying degrees of e-admissibility in decision making with imprecise probabilities, in: *Reflections on the Foundations of Probability and Statistics: Essays in Honor of Teddy Seidenfeld*, Springer, 2022, pp. 319–346.
- [11] I. Levi, *The Enterprise of Knowledge*, MIT Press, London, 1980.
- [12] P. Limbourg, R. Savic, J. Petersen, H. Kochs, Fault tree analysis in an early design stage using the Dempster-Shafer theory of evidence, in: *Risk, Reliability and Societal Safety*, 2007, pp. 713–722.
- [13] W. Liu, H. Wang, X. Shen, I.W. Tsang, The emerging trends of multi-label learning, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (2021) 7955–7974.
- [14] T. Lorieul, A. Joly, D. Shasha, Classification under ambiguity: when is average-k better than top-k?, preprint, arXiv:2112.08851, 2021.
- [15] E. Miranda, I. Montes, Centroids of the core of exact capacities: a comparative study, *Ann. Oper. Res.* 321 (2023) 409–449.
- [16] E. Miranda, I. Montes, A. Presa, Inner approximations of coherent lower probabilities and their application to decision making problems, *Ann. Oper. Res.* (2023) 1–39.
- [17] S. Moral-García, S. Destercke, Partial calibrated multi-label ranking, in: *International Conference on Soft Methods in Probability and Statistics*, Springer, 2022, pp. 287–294.
- [18] N. Nakharutai, Algorithms for generating sets of gambles for decision making with lower previsions, in: V.N. Huynh, T. Entani, C. Jeenanunta, M. Inuiguchi, P. Yenradee (Eds.), *Integrated Uncertainty in Knowledge Modelling and Decision Making*, Springer International Publishing, Cham, 2020, pp. 62–71.
- [19] N. Nakharutai, S. Destercke, M.C.M. Troffaes, Decision making under severe uncertainty on a budget, in: F. Dupin de Saint-Cyr, M. Öztürk-Escoffier, N. Potyka (Eds.), *15th International Conference on Scalable Uncertainty Management (SUM 2022)*, Springer International Publishing, 2022, pp. 186–201.
- [20] N. Nakharutai, M.C.M. Troffaes, C.C.S. Caiado, Improving and benchmarking of algorithms for decision making with lower previsions, *Int. J. Approx. Reason.* 113 (2019) 91–105, <https://doi.org/10.1016/j.ijar.2019.06.008>, arXiv:1906.12215.
- [21] L. Paton, M.C.M. Troffaes, N. Boatman, M. Hussein, A. Hart, A robust Bayesian analysis of the impact of policy decisions on crop rotations, in: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur (Eds.), *ISIPTA'15: Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications, ARACNE, Pescara, Italy, 2015*, pp. 217–226, <http://www.sipta.org/isipta15/data/paper/17.pdf>.
- [22] B. Quost, S. Destercke, Classification by pairwise coupling of imprecise probabilities, *Pattern Recognit.* 77 (2018) 412–425.
- [23] A.H. Schumann, D. Nijssen, *Application of Scenarios and Multi-Criteria Decision Making Tools in Flood Polder Planning*, Springer, Netherlands, Dordrecht, 2011, pp. 249–275.
- [24] M.C.M. Troffaes, Decision making under uncertainty using imprecise probabilities, *Int. J. Approx. Reason.* 45 (2007) 17–29, <https://doi.org/10.1016/j.ijar.2006.06.001>, arXiv:1807.03705.
- [25] P. Viappiani, C. Boutilier, On the equivalence of optimal recommendation sets and myopically optimal query sets, *Artif. Intell.* 286 (2020) 103328.
- [26] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York, 1991.