Decision making under severe uncertainty on a budget

Nawapon Nakharutai[1 [0000−0002−3145−713X]], Sébastien Destercke[2 [0000−0003−2026−468X]], and Matthias C. M. Troffaes[3 [0000−0002−1294−600X]]

[1] Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand. nawapon.nakharutai@cmu.ac.th
[2] UMR CNRS 7253 Heudiasyc, Sorbonne Université, Université de Technologie de Compiègne, France, sebastien.destercke@hds.utc.fr
[3] Durham University, Department of Mathematical Sciences, UK matthias.troffaes@durham.ac.uk

Abstract. Convex sets of probabilities are general models to describe and reason with uncertainty. Moreover, robust decision rules defined for them enable one to make cautious inferences by allowing sets of optimal actions to be returned, reflecting lack of information. One caveat of such rules, though, is that the number of returned actions is only bounded by the number of possibles actions, which can be huge, such as in combinatorial optimisation problems. For this reason, we propose and discuss new decision rules whose number of returned actions is bounded by a fixed value and study their consistency and numerical behaviour.

Keywords: Imprecise probabilities · Decision · Regret.

1 Introduction

Imprecise probability theories [1] provide very general tools to handle uncertainty, encompassing many existing uncertainty representations, including for instance classical probability, lower and upper previsions, sets of probability measures, choice functions, n-monotone capacities, and sets of desirable gambles. Imprecise probability theories often use convex sets of probability measures (or equivalent mathematical representations) as basic uncertainty models. They are used in practical applications that involve severe uncertainty, including for example wind-farm design [2] or machine learning [13].

Classical imprecise probability decision rules either deliver a single[4] optimal alternative as output, or a set of such alternatives whose size is unconstrained, apart from the trivial bounds that are 1 and the total number of alternatives. While such rules have been widely used and have strong theoretical properties, there are situations where one may want more than one decision, but still limit the number of proposed alternatives by, e.g., specifying an upper bound on the

[4] up to indifference

number of alternatives to return. This can arise for instance in situations where one has to account for natural human cognitive limits (a decision maker cannot inspect dozens of possible alternatives), or where inspecting more closely the different proposed alternatives represents a high monetary cost. Presenting set-valued recommendations or predictions with a limited budget is already treated in preference learning [14] and in standard machine-learning [4]. Hence, treating it in the setting of decisions under uncertainty appears as a natural next step.

We say that a decision rule is *budgeted* or *on a budget* if it limits the number of decisions it outputs. In Section 2, we present some basic ideas and notations about budgeted decision rules. We then propose an study two such rules in Sections 3 and 4, one based on the idea of minmax regret, the other on maximising diversity. We then discuss them in the light of numerical experiments as well as with respect to the previously proposed properties in Section 5.

2 Preliminaries and definitions

We start with a finite set $\mathcal{X} = \{x_1, \ldots, x_m\}$ of possible states of nature about which we are uncertain. We assume this is uncertainty is represented by a credal set $\mathcal{P}$, i.e. $\mathcal{P}$ is a closed convex set of probability mass functions on $\mathcal{X}$ and this set represents our knowledge about the unknown true value $x \in \mathcal{X}$.

An *act* $a\colon \mathcal{X} \to \mathbb{R}$ is a real-valued function on $\mathcal{X}$ that is interpreted as an uncertain reward, i.e. $a(x)$ represents the reward (in utiles) if $x \in \mathcal{X}$ is the true state of nature. We denote by $\mathcal{A}$ the set of all finite non-empty sets of acts. Each element of $\mathcal{A}$ represents a decision problem with a finite number of options. We set ourselves in the basic decision theoretic setting where we have to make a single decision [12], that is recommend once options from an element of $\mathcal{A}$, yet we consider that when information is lacking, we can return or recommend as a decision multiple options. This contrasts, for instance, with sequential problems where one should recommend a policy over multiple time-steps and for large state spaces [6].

A *decision rule* D is a function

$$D\colon \mathcal{A} \to \mathcal{A}.$$

satisfying $D(A) \subseteq A$ for every $A \in \mathcal{A}$. For example, maximising expected utility for a given $p \in \mathcal{P}$ is a decision rule. In case one allows for sets of possibly optimal decisions, maximality with respect to a credal set $\mathcal{P}$ is one of the most used rules. It works as follows: an act a maximally dominates a', denote $a \succ_{\mathcal{M}} a'$, if and only if $\underline{\mathbb{E}}(a - a') > 0$, where $\underline{\mathbb{E}}(f) \coloneqq \inf_{p \in \mathcal{P}} \mathbb{E}_p(f)$, $\mathbb{E}_p$ denoting the expectation operator with respect to f. The corresponding decision rule D_M then collects all maximal elements according to $\succ_{\mathcal{M}}$, i.e.,

$$D_M(A) \coloneqq \{a \in A\colon \nexists a' \in A \text{ s.t. } a' \succ_{\mathcal{M}} a\},$$

that is all the undominated elements in A with respect to $\succ_{\mathcal{M}}$. While this rule has strong theoretical appeal, it can deliver any subset of A, from a single decision to the whole set [10].

As argued in the introduction, one could want to limit the number of returned decisions. To do so, we introduce the notion of a budgeted decision rule. If we now denote by $\mathcal{A}_k \coloneqq \{A \in \mathcal{A} \colon |A| \leq k\}$ the non-empty finite sets of acts with at most k alternatives, we define a *k-budgeted decision rule* D^k as a rule

$$D^k \colon \mathcal{A} \to \mathcal{A}_k,$$

meaning that $D^k(A)$ returns *at most* k alternatives. Once one accepts the need for such rules, it is natural to look which properties they should follow, as well as look how computable they are.

A property we may want is such rules to be partially consistent with well-known, and theoretically well-justified rules producing non-bounded recommendation sets (maximality, E-admissibility). We define two such consistency properties, a strong one and a weak one:

Definition 1. *A k-budgeted rule D^k is said to be* strongly consistent *with a rule D if for all $A \in \mathcal{A}$*

$$D^k(A) \subseteq D(A)$$

Definition 2. *A k-budgeted rule D^k is said to be* weakly consistent *with a rule D if for all $A \in \mathcal{A}$*

$$D^k(A) \cap D(A) \neq \emptyset$$

While strong consistency requires $D^k(A)$ to be a subset of $D(A)$, weak consistency merely requires $D^k(A)$ to contain some elements that are also in $D(A)$, while others may not belong to it. Note that a natural way to ensure strong consistency with a rule $D(A)$ is simply to first compute exactly $D(A)$, and then to apply $D^k(D(A))$. However, this requires first computing $D(A)$. This may be impossible in problems where $|A|$ is extremely large, such as in machine learning problems like multi-label ones [5] or in combinatorial optimisation problems [2,3].

Next, we will introduce two rules where with each $A \in \mathcal{A}$ a value is associated, either considered as a loss or as a utility. One can then see the problem of selecting k decisions as either picking the set within $\mathcal{A}$ that has at most k decisions and that either tries to minimize a loss or to maximize a utility. The first rule is based on the idea of regret, while the second is based on the idea of maximising the spread of selected alternatives in terms of a pseudometric.

3 Regret-based budgeted decision rule

3.1 Definition

For any two acts a and a', $\overline{\mathbb{E}}(a'-a) = \sup_{p \in \mathcal{P}} \mathbb{E}_p(a'-a) = -\underline{\mathbb{E}}(a-a')$ represents the maximal expected gain in exchanging a' for a, or alternatively the worst possible loss we would incur by keeping a instead of exchanging it for a'. It is negative only if a is better than a' in terms of expected utility under all $p \in \mathcal{P}$. The maximal loss of retaining only a from A is then

$$ML(\{a\}, A) \coloneqq \max_{a' \in A \setminus \{a\}} \overline{\mathbb{E}}(a'-a).$$

Note that this is negative if and only if a dominates (in the sense of $\succ_{\mathcal{M}}$) all other actions in A, and is therefore the unique optimum. Otherwise, it is positive and something we want to miminize, since it is a loss function. Therefore, if we have to pick exactly one action, we can define a minmax[5] loss of A as follows:

$$mML(A) \coloneqq \min_{a \in A} ML(\{a\}, A) \tag{1}$$

and the associated decision rule as

$$D^1_{mML}(A) \coloneqq \arg\min_{a \in A} ML(\{a\}, A). \tag{2}$$

This is a kind of minmax regret criterion. We can now turn it into a set-valued criterion. Consider a solution set $S \subseteq A$, then the maximal loss associated with this set of alternatives, considering that we can pick any alternative within S as our choice but that the opponent is then free to choose the worst adversary (we first pick $a \in A$, then the adversary picks the worst alternatives according to $ML(\{a\}, A)$), can be defined for any $\emptyset \neq S \subseteq A$ as

$$mML(S, A) \coloneqq \min_{a \in S} ML(\{a\}, A \setminus S) = \min_{a \in S} \max_{a' \in A \setminus S} \overline{\mathbb{E}}(a' - a). \tag{3}$$

We will use (3) to define a budgeted decision rule, yet before doing so we provide two properties of $mML(S, A)$.

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

Proof. Indeed,

$$\begin{aligned}
mML(S', A) &= \min_{a \in S'} ML(\{a\}, A \setminus S') \\
&\leq \min_{a \in S} ML(\{a\}, A \setminus S') && \text{(since } S \subseteq S'\text{)} \\
&= \min_{a \in S} \max_{a' \in A \setminus S'} \overline{\mathbb{E}}(a' - a) \\
&\leq \min_{a \in S} \max_{a' \in A \setminus S} \overline{\mathbb{E}}(a' - a) && \text{(since } S \subseteq S'\text{)} \\
&= mML(S, A)
\end{aligned}$$

Lemma 1 tells us that as $mML(S, A)$, any rule that tries to find a set S minimizing it will search for the biggest possible set. In particular, among the sets of $\mathcal{A}_k$, it will always pick a set of size k. The next property shows that $mML(S, A)$ is negative if only if every alternative outside of S is dominated (in the sense of maximality) by some alternative within S.

Theorem 1. *$mML(S, A) < 0$ if only if there is an $a \in S$ such that for all $a' \in A \setminus S$ we have that $a \succ_{\mathcal{M}} a'$.*

[5] as it minimizes a maximal loss that is $\overline{\mathbb{E}}(a' - a)$.

Proof. By the definition, we have

$$\begin{aligned} mML(S,A)<0 &\iff \min_{a\in S} ML(\{a\},A\setminus S)<0 \\ &\iff \exists a\in S,\ ML(\{a\},A\setminus S)<0 \\ &\iff \exists a\in S,\ \max_{a'\in A\setminus S} \overline{\mathbb{E}}(a'-a)<0 \\ &\iff \exists a\in S,\ \forall a'\in A\setminus S,\ \overline{\mathbb{E}}(a'-a)<0 \\ &\iff \exists a\in S,\ \forall a'\in A\setminus S,\ \underline{\mathbb{E}}(a-a')>0 \end{aligned}$$

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

Proof. By Theorem 1, all $a'\in A\setminus S$ are dominated, hence $A\setminus S$ contains no maximal elements, so all maximal elements must be in S.

Let us now denote by $S_k^*(A)$ the optimal subset of size k w.r.t. mML criterion (that we want to minimize) within A, i.e.

$$S_k^*(A) \coloneqq \arg\min_{S\in\mathcal{A}_k} mML(S,A)$$

We can now define the mML budgeted decision rule as

$$D^k_{mML}(A) \coloneqq \begin{cases} D_M(A) & \text{if } mML(S_k^*(A),A)<0 \\ S_k^*(A) & \text{otherwise} \end{cases}$$

3.2 Example and computation

The next example illustrates some behaviour of $mML(S,A)$, as well as of $S_k^*(A)$.

Example 1. Let us consider the space $\mathcal{X}=\{x_1,x_2,x_3\}$ and the acts of Table 1. Suppose furthermore that the uncertainty on the states are specified by

$$\mathcal{P}=\{p\in\mathcal{P}\colon p(x_1)+2p(x_2)+3p(x_3)\le 2,\ p(x_3)\le 0.3\} \tag{4}$$

	x_1	x_2	x_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

Table 1. Acts of Example 1

$\overline{\mathbb{E}}(a_j-a_i)$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$j=1$	-	4.0	2.0	1.0	5.0
$j=2$	4.0	-	6.0	3.0	5.0
$j=3$	1.4	3.3	-	1.5	4.0
$j=4$	1.0	3.0	3.0	-	4.0
$j=5$	-0.4	-0.1	1.0	-1.1	-

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

Table 2 gives the values of $\overline{\mathbb{E}}(a_j-a_i), \forall j\neq i$. According to this table, we have $D_M(A)=\{a_1,a_2,a_3,a_4\}$ and

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

From Example 1, we have that $S_k^*(A) \nsubseteq S_{k+1}^*(A)$, therefore showing that a greedy approach iteratively picking the next best option to build $S_{k+1}(A)$ from $S_k(A)$ will not be optimal in general which is unfortunate, as such property might help constructing efficient computational algorithms that for instance iteratively increase k. Also, $S_3^*(A)$ is not unique. The result also agrees with theorem 1 that once $S_k(A)$ reaches the negative value of mML, then $S_k(A)$ is a superset of $D_M(A)$. Suppose we want to find $S_k^*(A)$ with respect to the mML criteria within a given set of acts A of size $n \geq k$. One may then wonder whether we have to verify all possible sets S of size k, of which there are $\binom{n}{k}$, to obtain the minimum of $mML(S, A)$. Fortunately, we can do it without checking all such sets. Let us consider the previous example and see how it can be obtained, with the formal procedure provided by Algorithm 1. Basically, it relies on the fact that the minimum of $mML(S, A)$ is reached at a specific alternative.

Example 2. From example 1, recall the values of $\overline{\mathbb{E}}(a_j - a_i)$ for all $a_j \neq a_i$ that are given in table 2. For $k = 1$, we notice that $S_1^*(A) = \{a_{i^*}\}$, where

$$i^* = \arg\min_{i=1}^{5} \max_{j=1, j\neq i}^{5} \overline{\mathbb{E}}(a_j - a_i)$$

which can be simply obtained by searching, for each i, the maximal value of $\overline{\mathbb{E}}(a_j - a_i)$ over $j \neq i$ and then finding the minimal value among maximal values that we have. In this case, $i^* = 4$ so $S_1^*(A) = \{a_4\}$

For $k = 2$, even though there are $\binom{5}{2} = 10$ sets of size 2 that we have to consider, we do not need to search all these sets. Let $a_{i^*} \in S_2^*(A)$ be such that

$$a_{i^*} = \arg \min_{a_i \in S_2^*(A)} \max_{a_j \in A \setminus S_2^*(A)} \overline{\mathbb{E}}(a_j - a_i).$$

We observe that a_{i^*} is obtained at the minimal value of the second highest value of $\overline{\mathbb{E}}(a_j - a_i)$ (circled values in table 3). This is because we are looking at the a_i that attains the minimal value of $\overline{\mathbb{E}}(a_j - a_i)$ for which a_j is not in the same set of a_i. In this case we have $a_{i^*} = a_1$ which can be obtained only if a_2 is also in $S_2^*(A)$. Otherwise, a_1 will not attain the minimum, as $\overline{\mathbb{E}}(a_2 - a_1) > 1.4$. Therefore, once we find the a_{i^*} that attains the minimal value of the second highest value of $\overline{\mathbb{E}}(a_j - a_i)$, the a_j such that $\overline{\mathbb{E}}(a_j - a_{i^*})$ is larger than the second highest values of a_{i^*} is also in $S_2^*(A)$ (boxed value in table 3).

We can use the same argument for $S_3^*(A)$, and in fact $S_k^*(A)$ for any k. Specifically, for each a_i, we look at the kth highest value of $\overline{\mathbb{E}}(a_j - a_i)$, find the minimum of these values say achieved at a_{i^*}, and then take all a_j such that $\overline{\mathbb{E}}(a_j - a_{i^*})$ is larger than this minimum. For instance, in this case, for $k = 3$, $a_{i^*} = a_1$ or a_4. Again, these a_{i^*} can attain the minimum of $S_3^*(A)$ if all a_j such that $\overline{\mathbb{E}}(a_j - a_{i^*})$ is larger than the third highest values of a_{i^*} are also in $S_3^*(A)$.

$\overline{\mathbb{E}}(a_j - a_i)$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$j=1$	-	4.0	2.0	1.0	5.0
$j=2$	[4.0]	-	6.0	3.0	(5.0)
$j=3$	(1.4)	(3.3)	-	(1.5)	4.0
$j=4$	1.0	3.0	(3.0)	-	4.0
$j=5$	-0.4	-0.1	1.0	-1.1	-

Table 3. Second highest values of $\overline{\mathbb{E}}(a_j - a_i)$ for each a_i (values in circle).

$\overline{\mathbb{E}}(a_j - a_i)$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$j=1$	-	4.0	(2.0)	(1.0)	5.0
$j=2$	[4.0]	-	6.0	[3.0]	5.0
$j=3$	[1.4]	3.3	-	[1.5]	(4.0)
$j=4$	(1.0)	(3.0)	3.0	-	4.0
$j=5$	-0.4	-0.1	1.0	-1.1	-

Table 4. Third highest values of $\overline{\mathbb{E}}(a_j - a_i)$ for each a_i (values in circle).

Thus, for selecting $a_{i^*} = a_1$, we have $S_3^*(A) = \{a_1, a_2, a_3\}$ and for selecting $a_{i^*} = a_4$, we have $S_3^*(A) = \{a_2, a_3, a_4\}$.

We now translate this argument into an algorithm that can find $S_k^*(A)$. Specifically, for each $a_i \in A$, we first compute $e_{ij} := \overline{\mathbb{E}}(a_j - a_i), \forall j \neq i$, and assign set $S[i]$ such that for all $j \in S[i]$, e_{ij} are the k largest elements of $\{e_{ij} : j \neq i\}$. Then, we calculate $M[i] = \min_{j \in S[i]} e_{ij}$ and $J[i] = \arg\min_{j \in S[i]} e_{ij}$. Next, we compute $i^* = \arg\min_{i=1}^{n} M[i]$. Finally, we have $S_k^*(A) = \{a_{i^*}\} \cup \{a_j : S[i^*] \setminus \{J[i^*]\}\}$. This process is summarised in algorithm 1.

The set $S[i]$ can be obtained through a partial sort, which can be much faster than a regular sort especially for small values of k. A partial sorting algorithm will normally also immediately give $M[i]$ (i.e. the value of the kth largest element). The algorithm can be easily adapted to perform a full sort instead of just finding the k largest elements. A full sorting can then be used to find $S_k^*(A)$ for all possible values of k simultaneously, with very little additional computational effort.

The process could be sped up further by presorting the acts a_j first by expectation with respect to $\mathbb{E}_p(a_j)$ for some $p \in \mathcal{P}$. A similar technique was shown to be very effective in the context of maximality [11].

Provided upper natural extensions $\overline{\mathbb{E}}(a_j - a_i)$ for all $j \neq i$ are all available, the (partial) sorting in algorithm 1 is much faster (a full sort of n elements typically takes $\mathcal{O}(n \log(n))$ comparisons, and we need to do this n times) than directly searching throughout all possibilities sets of choosing k elements from n, of which there are n^k. We therefore go from a polynomial number of operations

Algorithm 1 Finding $S_k^*(A)$

Input: $A = \{a_1, a_2, \dots, a_n\}$, $\mathcal{P}$, k
Output: $S_k^*(A)$, $mML(S_k^*(A), A)$
for $i = 1\colon n$ **do**
  for $j = 1\colon n$, $j \neq i$ **do**
    compute $e_{ij} \coloneqq \overline{\mathbb{E}}(a_j - a_i)$
  end for
end for
for $i = 1\colon n$ **do**
  $S[i] \leftarrow$ set such that $\{e_{ij}\colon j \in S[i]\}$ are the k largest elements of $\{e_{ij}\colon j \neq i\}$
  $M[i] \leftarrow \min_{j \in S[i]} e_{ij}$
  $J[i] \leftarrow \arg\min_{j \in S[i]} e_{ij}$
end for
$i^* \leftarrow \arg\min_{i=1}^{n} M[i]$
return $\{a_j\colon j \in \{i^*\} \cup S[i^*] \setminus \{J[i^*]\}\}$, $M[i^*]$

to an exponential one. However the task to evaluate $n(n-1)$ upper natural extensions $\overline{\mathbb{E}}(a_j - a_i)$ for all $j \neq i$ may still be challenging if n is quite high, which is particularly true for combinatorial problems, where adding an element (e.g., a node to a graph) may lead to an exponential increase of $|A|$.

One way to circumvent this needs is to be able to find $\max_{a_j \in A \setminus S_2^*} \overline{\mathbb{E}}(a_j - a_i)$ without a complete enumeration, something that is sometimes doable in structured problems [3]. In this case, one solution could either be to simply sample a reasonable number of alternatives from A, and use the fact that $\max_{a_j \in A \setminus S_2^*} \overline{\mathbb{E}}(a_j - a_i)$ can be evaluated without enumerating the whole set $\mathcal{A}$ to greedily add sampled alternatives to the set of k returned alternatives. This is why in further experiments, we include results of the greedy algorithm, denoted by $S_k^g(A)$, and compare them with the non-greedy, global optimal solution.

3.3 Weak Consistency of S_k^* and D_{mML}^k

Let us now show that S_k^* (and consequently, also D_{mML}^k) is weakly consistent with maximality. Weak consistency for A that satisfy $mML(S_k^*(A), A) < 0$ follows from Corollary 1, so we are left to consider the case $mML(S_k^*(A), A) \geq 0$. The proof uses the observations we have made to obtain Algorithm 1.

Theorem 2. *For any k and A, if $mML(S_k^*(A), A) \geq 0$ then $S_k^*(A) \cap D_M(A) \neq \emptyset$.*

Proof. For brevity, define $S \coloneqq S_k^*(A)$ and $S' \coloneqq A \setminus S_k^*(A)$. Suppose that $mML(S, A) \geq 0$. Let

$$a_{i^*} \coloneqq \arg\min_{a_i \in S} \max_{a_j \in S'} \overline{\mathbb{E}}(a_j - a_i), \tag{5}$$

$$a_{j^*} \coloneqq \arg\max_{a_j \in S'} \overline{\mathbb{E}}(a_j - a_{i^*}). \tag{6}$$

Note that

$$0 \leq mML(S, A) = \min_{a_i \in S} \max_{a_j \in S'} \overline{\mathbb{E}}(a_j - a_i) = \overline{\mathbb{E}}(a_{j^*} - a_{i^*}), \tag{7}$$

and it follows that

$$\overline{\mathbb{E}}(a_{j^*} - a_{i^*}) \geq 0, \tag{8}$$

$$\forall a_j \in S',\, \overline{\mathbb{E}}(a_j - a_{i^*}) \leq \overline{\mathbb{E}}(a_{j^*} - a_{i^*}) \tag{9}$$

$$\forall a_i \in A,\ \forall j \in S[i],\, \overline{\mathbb{E}}(a_j - a_i) \geq \overline{\mathbb{E}}(a_{j^*} - a_{i^*}) \tag{10}$$

where $S[i]$ is defined as in the algorithm. Equation (10) holds because, from the algorithm, we know that

$$M[i] = \min_{j \in S[i]} \overline{\mathbb{E}}(a_j - a_i) \geq M[i^*] = \overline{\mathbb{E}}(a_{j^*} - a_{i^*}) \tag{11}$$

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

$$B \coloneqq \{a_m \colon \overline{\mathbb{E}}(a_m - a_\ell) \geq \overline{\mathbb{E}}(a_{j^*} - a_{i^*})\} \tag{12}$$

This set has at least k elements by eq. (10). If $a_{i^*} \in B$, then we are done, by eq. (8). 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

$$\begin{aligned}\overline{\mathbb{E}}(a_{i^*} - a_\ell) &\geq \overline{\mathbb{E}}(a_m - a_\ell) - \overline{\mathbb{E}}(a_m - a_{i^*}) \\ &= \underbrace{\overline{\mathbb{E}}(a_m - a_\ell) - \overline{\mathbb{E}}(a_{j^*} - a_{i^*})}_{\text{non-negative by eq. (12)}} + \underbrace{\overline{\mathbb{E}}(a_{j^*} - a_{i^*}) - \overline{\mathbb{E}}(a_m - a_{i^*})}_{\text{non-negative by eq. (9)}} \geq 0.\end{aligned}$$

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

We then have the following corollaries from Theorem 2 and Corollary 1.

Corollary 2. *S^*_k and D^k_{mML} are weakly consistent with D_M.*

Corollary 3. *S^*_1 and D^1_{mML} are strongly consistent with D_M.*

4 Metric-based budgeted decision rule

We now discuss an alternative to regret-based rules, by considering a metric argument according to which one selects alternatives that are the most dissimilar to one another. However, a naive application of this criterion may actually selects options without any considerations for their possible optimality, as will demonstrate our experiments in Section 5. For this reason, metric-based budgeted decision rule should only be used once good options have already been selected, typically by first applying a decision rule filtering sub-optimal options, e.g., by first applying D_M to $\mathcal{A}$.

4.1 Definition

We now consider a different angle, where we want to retained alternatives to cover as much as possible the space of all possible alternatives. The underlying idea is that retained alternatives should be as diverse as possible, so as to expose the decision maker to varied options. For this, we will try to maximise the distances between alternatives, given our knowledge represented by $\mathcal{P}$. The underlying idea is close to the one of space filling designs [8], where one tries to find samples that provide maximal coverage of a given space.

For easy of notation, for any act a, let $|a|$ be the act defined by $|a|(x) := |a(x)|$ for all $x \in \mathcal{X}$. For any pair of alternatives a and a', $\overline{\mathbb{E}}(|a - a'|)$ as a function of a and a' defines a pseudo-metric between alternatives. It is clearly non-negative and symmetric. Moreover, it satisfies the triangle inequality since $\overline{\mathbb{E}}(|a - a''|) \leq \overline{\mathbb{E}}(|a - a'|) + \overline{\mathbb{E}}(|a' - a''|)$ [15, §2.6.1]. Thus, $\overline{\mathbb{E}}(|a - a'|)$ is a pseudometric on the set of all acts.

Note that, for all $p \in \mathcal{P}$,

$$|\mathbb{E}_p(a) - \mathbb{E}_p(a')| \leq \overline{\mathbb{E}}(|a - a'|)$$

Thus, $\overline{\mathbb{E}}(|a-a'|)$ is a measure of how different a and a' are with respect to expectation. Usually, maximizing dispersion only makes sense as a security criterion, i.e. in practice we want to apply it after we have already calculated the set of optimal decisions, to reduce the size of the optimal set whilst maximizing dispersion. Finding the pair of alternatives in A that are the most different according to this pseudometric comes down to find a pair of a_{i^*} and a_{j^*} such that

$$\overline{\mathbb{E}}(|a_{i^*} - a_{j^*}|) = \max_{a_i, a_j \in A,\ i<j} \overline{\mathbb{E}}(|a_i - a_j|)$$

A value function for a given set S could then be the sum of the pairwise distances, i.e.,

$$MS(S) = \sum_{a_i, a_j \in S,\ i<j} \overline{\mathbb{E}}(|a_i - a_j|)$$

that we would like to maximise, in order to select those alternatives that are far apart from each other. Consider a decision rule D returning k alternatives such that these k alternatives are spread over the set of acts.

$$D^k_{MS}(A) = \arg\max_{S \in \mathcal{A}_k} MS(S) \tag{13}$$

One can readily see that MS is an increasing function, meaning that if we restrict ourselves to sets of size k, the maximum of (13) will be reached for a set of size k. We can therefore restrict our attention to those. Note that this rule cannot handle the case $k = 1$, for which one can simply take S^*_1 as a solution.

4.2 Example and computation

Example 3. Consider a space $\mathcal{X} = \{x_1, x_2, x_3\}$ together with the actions provided in Table 5 and the credal set given by eq. (4). The values $\overline{\mathbb{E}}(|a_i - a_j|)$ for all $i < j$ are given in Table 6.

According to the result, we find that

	x_1	x_2	x_3
a_1	8	9	3
a_2	1	5	4
a_3	8	2	6
a_4	3	1	2
a_5	5	4	9

Table 5. Acts of Example 3

$\overline{\mathbb{E}}(\lvert a_i - a_j \rvert)$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$i = 1$	7.0	7.0	8.0	5.0
$i = 2$	-	7.0	4.0	4.3
$i = 3$	-	-	5.0	6.0
$i = 4$	-	-	-	3.9

Table 6. Pairwise distance for Example 3

- $D^2_{MS} = \{a_1, a_4\}$ with $MS(D^2_{MS}) = 8$,
- $D^3_{MS} = \{a_1, a_2, a_3\}$ with $MS(D^3_{MS}) = 21$ and
- $D^4_{MS} = \{a_1, a_2, a_3, a_4\}$ with with $MS(D^4_{MS}) = 38$.

We see that $D^2_{MS} \not\subseteq D^3_{MS}$. Therefore, $D^k_{MS} \not\subseteq D^{k+1}_{MS}$, from which we can conclude that a greedy algorithm will again not be optimal in general. However, unlike the regret-based approach, we were unable to find an efficient algorithm to directly find D^k_{MS}, and it is therefore quite relevant to study the quality of the answer provided by a greedy approach, summarised in algorithm 2. Note that for $k = 2$, $D^k_{MS} = D^k_{gMS}$. Also note that only a_1 and a_3 are maximal, so this example also verifies the behaviour that maximizing dispersion captures non-maximal options too. Despite seeing this result, we would like to see how much D^k_{MS} and its greedy approximation can capture the maximal elements or being consistency with maximality, which will be investigate in Section 5.

Algorithm 2 Greedy approximation of D^k_{MS}

Input: A, $\mathcal{P}$, k, where $k \geq 2$
Output: an approximate solution of $D^k_{MS}(A)$

$D^k_{gMS} \leftarrow \emptyset$
for $i = 1\colon n-1$ **do**
  for $j = 2\colon n$ **do**
    compute $\overline{\mathbb{E}}(|a_i - a_j|), \forall i < j$
  end for
end for
$D^k_{gMS} \leftarrow \arg\max_{a_i, a_j \in A} \overline{\mathbb{E}}(|a_i - a_j|), \forall i < j$
while $|D^k_{gMS}| < k$ **do**
  $a^* \leftarrow \arg\max_{a_j \in A \setminus D^k_{gMS}} \left(\sum_{a_i \in D^k_{gMS}} \overline{\mathbb{E}}(|a_i - a_j|) \right), \forall i < j$
  $D^k_{gMS} \leftarrow D^k_{gMS} \cup \{a^*\}$
end while
return D^k_{gMS} ▷ an approximate solution of D^k_{MS}

4.3 On some properties of D^k_{gMS}

When maximising a value function (here MS) under cardinality constraint (here k), submodularity is a property guaranteeing the quality of greedy approximation [9]. Recall that a set function f is submodular iff $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$ whenever $S \subseteq T$. Unfortunately, we can show that MS has the reverse, supermodularity property.

Lemma 2. *If $S_1 \subseteq S_2$, then*

$$MS(S_2 \cup \{a_k\}) - MS(S_2) \geq MS(S_1 \cup \{a_k\}) - MS(S_1) \tag{14}$$

Proof. Without loss of generality, let's k be an index such that $k > j$ for all $a_j \in S_2$.

$$(MS(S_2 \cup \{a_k\}) - MS(S_2)) - (MS(S_1 \cup \{a_k\}) - MS(S_1)) =$$
$$\sum_{a_i \in S_2} \overline{\mathbb{E}}(|a_i - a_k|) - \sum_{a_i \in S_1} \overline{\mathbb{E}}(|a_i - a_k|) = \sum_{a_i \in S_2 \setminus S_1} \overline{\mathbb{E}}(|a_i - a_k|) \geq 0.$$

Since MS is positive, monotone increasing but supermodular, there is no guarantee on polynomial-time constant approximating algorithm for maximizing MS with respect to a maximum cardinality of size k. Nevertheless, in the next section, we will perform an experiment to see how close of the outcomes of the greedy algorithm to optimal solutions.

5 First experimentation

In this section, we will perform some first experiments to compare S_k^*, D_{MS}^k and their greedy approximations S_k^g and D_{gMS}^k. Each set will be checked for consistency with respect to maximality. Specifically, we would like to find out how much S_k^*, S_k^g, D_{MS}^k and D_{gMS}^k can capture maximal alternatives in the set D_M. In addition, we will measure the quality of the greedy approximations. Note that we do not consider case $k = 1$ since by Corollaries 2 and 3, S_1^* is weekly and strongly consistent with D_M while the size of D_{MS}^k requires $k \geq 2$.

We fix $|A| = 20$, $|\mathcal{X}| = 5$ and $k \in \{2, \dots, 6\}$. Throughout the experiment, we consider the credal set $\mathcal{P}$ that satisfies the following condition:

$$p(x_1) + p(x_2) + p(x_3) + p(x_4) + p(x_5) = 1$$
$$3p(x_1) + 2p(x_2) + p(x_3) + p(x_4) + p(x_5) \leq 1$$
$$p(x_1) \leq 0.3, 0.1 \leq p(x_2), 0.2 \leq p(x_3) \leq 0.4$$

We generate a set of alternatives A on $\mathcal{X}$ as follows. For each x_j, we sample $a_i(x_j)$ uniformly from $(0, 1)$. Then, we compute $\overline{\mathbb{E}}(a_j - a_i)$ for all $a_i, a_j \in A$ with respect to the credal set $\mathcal{P}$. Next, for each A, we find D_M and check whether $|D_M| > k$ or not. If $|D_M| \leq k$, then we regenerate A since we are not interested in this case due to S_k^* simply returning D_M. Otherwise, we compute S_k^*, S_k^g, D_{MS}^k and D_{gMS}^k with respect to A. Next, we verify whether S_k^*, S_k^g, D_{MS}^k and D_{gMS}^k are weakly consistent or strongly consistent with D_M or not. As being weakly (having only one maximal element in the set) and strongly (having all elements in the set being maximal) consistent are two extreme situations, we also calculate the proportion of alternatives in S^*_k, S^g_k, D^k_{MS} and D^k_{gMS} that are in D_M. The process was repeated 500 times. The percentages of these sets that satisfy the properties and the average percentages of elements in these sets that are in D_M are presented in the 3$^{\text{rd}}$-5$^{\text{th}}$ columns of table 7. As expected from the result, unlike S^*_k and S^g_k, D^k_{MS} and D^k_{gMS} are not guaranteed to be weakly consistency and rarely strongly consistency with D_M. In addition, the average percentages of maximal elements in D^k_{MS} and D^k_{gMS} are much smaller than in S^*_k and S^g_k. This result confirms our earlier comment that applying D^k_{MS} and D^k_{gMS} to obtain sets of size k makes little sense if we do not start with an optimal set, and that maximizing dispersion should only serve as a refinement of another rule.

To see how S^g_k is close to S^*_k and D^k_{gMS} is close to D^k_{MS}, we also compare the optimal solution with their greedy approximations. Specifically, for each iteration, we count how many time $S^g_k = S^*_k$ and $D^k_{gMS} = D^k_{MS}$. The percentages of these sets that satisfy this condition is presented in the 6$^{\text{th}}$ column. We also calculate the proportion of elements in S^g_k that are in S^*_k and the proportion of elements in D^k_{gMS} that are in D^k_{MS} and present the average of the proportions in the 7$^{\text{th}}$ column. Finally, we calculate $mML(S^g_k)/mML(S^*_k)$ and $MS(D^k_{gMS})/MS(D^k_{MS})$ and present the averages of these ratios in the last column of table 7. While the greedy approximation of S^*_k quickly degrade as k increases, this is not the case for D^k_{MS}, with the greedy set often being pretty close in terms of quality to the optimal set. This is rather good news, as we do not have an efficient algorithm at our disposal to compute D^k_{MS}.

6 Discussion and conclusion

In this study, we have introduced k-budgeted decision rules that return an optimal subset of size k according to some value function. We have adopted two different views: one where we consider a regret-based argument, and the other where we want to have alternatives that are well-dispersed in the space of alternatives. This second approach is very close in spirit to some recent work bearing on E-admissibility (another well-known decision rule) [7] as well as to space-filling designs.

Concerning future work, we could look at other possibilities in each direction drafted in this paper. For instance, one could look for an alternative to eq. (3) where the maximisation is done before the minimisation (i.e., the opponent chooses the alternative within $A \setminus S$ before we pick our alternative within S). Similarly, one could replace the sum with a minimum in eq. (13). Finally, as we hinted already in the paper, it would be interesting to look at situations where the alternatives are too numerous to be explicitly listed/treated, and where even estimating $\overline{\mathbb{E}}(a_i - a_j)$ for every pair would be computationally prohibitive.

From a more practical perspective, it would be useful to do more complete and varied experiments, even if those we conducted already allowed us to highlight several aspects of our proposals. In addition, it would be interesting to apply those rules to actual problems such as uncertainty elicitation or system design,

D^k	k	w.c.	s.c.	$\frac{\lvert D^k \cap D_M \rvert}{\lvert D^k \rvert}$			
	2	100%	100%	100%			
	3	100%	92.6%	97.5%			
S_k^*	4	100%	81.0%	97.5%	$S_k^* = S_k^g$	$\frac{\lvert S_k^g \cap S_k^* \rvert}{\lvert S_k^g \rvert}$	$\frac{mML(S_k^g)}{mML(S_k^*)}$
	5	100%	68.4%	94.9%			
	6	100%	55.8%	91.1%			
	2	100%	90.2%	95.1%	24.8%	60.9%	0.841
	3	100%	74.2%	90.1%	5.2%	52.7%	0.721
S_k^g	4	100%	60.8%	88.1%	1.8%	52.9%	0.651
	5	100%	44.8%	84.8%	0.4%	53.2%	0.569
	6	100%	33.2%	83.0%	0.4%	56.3%	0.509
	2	86.4%	15.4%	50.9%			
	3	96.6%	6.0%	51.9%			
D_{MS}^k	4	99.8%	1.8%	54.2%	$D_{MS}^k = D_{gMS}^k$	$\frac{\lvert D_{gMS}^k \cap D_{MS}^k \rvert}{\lvert D_{gMS}^k \rvert}$	$\frac{MS(D_{gMS}^k)}{MS(D_{MS}^k)}$
	5	100%	0.2%	52.4%			
	6	100%	0.4%	55.0%			
	2	86.4%	15.4%	50.9%	100%	100%	1.000
	3	95.2%	5.0%	51.5%	52.4%	71.9%	0.986
D_{gMS}^k	4	99.8%	1.8%	53.7%	44.4%	75.1%	0.987
	5	99.8%	0.6%	52.4%	41.0%	79.9%	0.991
	6	99.8%	1.0%	52.2%	36.4%	82.1%	0.993

Table 7. Percentages and averages of S_k^*, S_k^g, D_{MS}^k and D_{gMS}^k that satisfy different conditions.

where the decision maker can only scrutinise and analyse a limited number of options.

Acknowledgements Work by NN was supported by CMU Junior Research Fellowship Program, and the SAFE AI chair funded by the Fondation UTC pour l'innovation. This work was also supported by ANR grant PreServe (ANR-18-CE23-0008).

References

1. Augustin, T., Coolen, F.P., de Cooman, G., Troffaes, M.C.: Introduction to imprecise probabilities. John Wiley & Sons (2014)
2. Bains, H., Madariaga, A., Troffaes, M.C., Kazemtabrizi, B.: An economic model for offshore transmission asset planning under severe uncertainty. Renewable Energy **160**, 1174–1184 (2020)
3. Benabbou, N., Perny, P.: Interactive resolution of multiobjective combinatorial optimization problems by incremental elicitation of criteria weights. EURO journal on decision processes **6**(3), 283–319 (2018)
4. Chzhen, E., Denis, C., Hebiri, M., Lorieul, T.: Set-valued classification–overview via a unified framework. arXiv preprint arXiv:2102.12318 (2021)

5. Destercke, S.: Multilabel predictions with sets of probabilities: the hamming and ranking loss cases. Pattern Recognition **48**(11), 3757–3765 (2015)
6. Huntley, N., Troffaes, M.: Normal form backward induction for decision trees with coherent lower previsions. Annals of Operations Research **195**(1), 111–134 (2012)
7. Jansen, C., Georg, S., Thomas, A.: Quantifying degrees of e-admissibility in decicion making with imprecise probabilities (2018)
8. Joseph, V.R.: Space-filling designs for computer experiments: A review. Quality Engineering **28**(1), 28–35 (2016)
9. Krause, A., Golovin, D.: Submodular function maximization. Tractability **3**, 71–104 (2014)
10. Nakharutai, N.: Algorithms for generating sets of gambles for decision making with lower previsions. In: Huynh, V.N., Entani, T., Jeenanunta, C., Inuiguchi, M., Yenradee, P. (eds.) Integrated Uncertainty in Knowledge Modelling and Decision Making. pp. 62–71. Springer International Publishing, Cham (2020)
11. Nakharutai, N., Troffaes, M.C.M., Caiado, C.C.S.: Improving and benchmarking of algorithms for decision making with lower previsions. International Journal of Approximate Reasoning **113**, 91–105 (Oct 2019). `https://doi.org/10.1016/j.ijar.2019.06.008`
12. Troffaes, M.: Decision making under uncertainty using imprecise probabilities. Int. J. of Approximate Reasoning **45**, 17–29 (2007)
13. Utkin, L.V.: An imprecise extension of svm-based machine learning models. Neurocomputing **331**, 18–32 (2019)
14. Viappiani, P., Boutilier, C.: On the equivalence of optimal recommendation sets and myopically optimal query sets. Artificial Intelligence **286**, 103328 (2020)
15. Walley, P.: Statistical reasoning with imprecise Probabilities. Chapman and Hall, New York (1991)