Error bounds for approximations of coherent lower previsions

Damjan Škulj

University of Ljubljana

WPMSIIP'16, Durham UK
6 September 2016

Contents

University *of Ljubljana*

Contents

Finite (imprecise) probability spaces

We study models with the following elements:

- sample space $\mathcal{X}$: a finite set with elements $x \in \mathcal{X}$;
- gamble: any map $f \colon \mathcal{X} \to \mathbb{R}$ or a vector in $\mathbb{R}^{\mathcal{X}}$;
- an arbitrary set of gambles $\mathcal{K}$;
- (precise) probability vector $p \in \mathbb{R}^{\mathcal{X}}$ satisfying $p(x) \geq 0 \, \forall x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} p(x) = 1$;
- linear prevision (expectation functional) $P \colon \mathcal{K} \to \mathbb{R}$ of the form $P(f) = \sum_{x \in \mathcal{X}} p(x) f(x) = p \cdot f$ where p is a precise probability vector;
- coherent lower prevision $\underline{P} \colon \mathcal{K} \to \mathbb{R}$ is a lower envelope of linear previsions.

Coherent lower previsions and lower expectation functionals

A coherent lower prevision $\underline{P}\colon \mathcal{K} \to \mathbb{R}$ can be expressed as a lower envelope of linear previsions

$$\underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P})} P(f),$$

where $\mathcal{M}(\underline{P})$ is the credal set of $\underline{P}$:

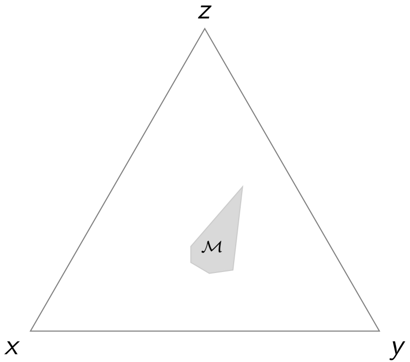
$$\mathcal{M}(\underline{P}) = \{P\colon P(f) \geq \underline{P}(f) \forall f \in \mathcal{K}\}.$$

A coherent lower prevision can be extended to a lower expectation functional $\underline{E}\colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}$, which is a coherent lower prevision defined everywhere in $\mathbb{R}^{\mathcal{X}}$.

Lower expectation functionals therefore form a family of coherent lower previsions.

Example

A credal set in probability simplex: The shaded points are the precise probabilities compatible with the corresponding coherent lower prevision.



The natural extension

Taking

$$\underline{E}(h) = \min_{P \in \mathcal{M}(\underline{P})} P(h) \ \forall h \in \mathbb{R}^{\mathcal{X}}.$$

gives the unique smallest (least committal) extension of the coherent lower prevision, called the natural extension.

If $\mathcal{K}$ is finite, the natural extension $\underline{E}(h)$ is calculated as a linear programming problem:

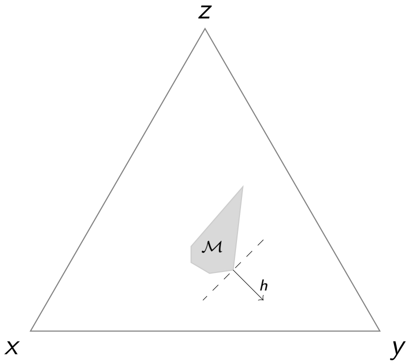
$$\text{minimize} P(h)$$

subject to

$$P(f) \geq \underline{P}(f) \ \forall f \in \mathcal{K}$$

Example

The value of the natural extension $\underline{E}(h)$ is a solution of a linear program.



Contents

Partially specified coherent lower prevision

- Let $\underline{P}$ be a coherent lower prevision on a set of gambles $\mathcal{H}$ (from now on $\mathcal{H} = \mathbb{R}^{\mathcal{X}}$).
- Sometimes we only know the values of $\underline{P}(f)\ \forall f \in \mathcal{K}$.
- Our best guess for $\underline{P}(h)$ is the value of its natural extension for h outside $\mathcal{K}$.

Problem

What is the maximal possible error that we make by taking the natural extension instead of the true value $\underline{P}(h)$?

Example: coherent lower probabilities

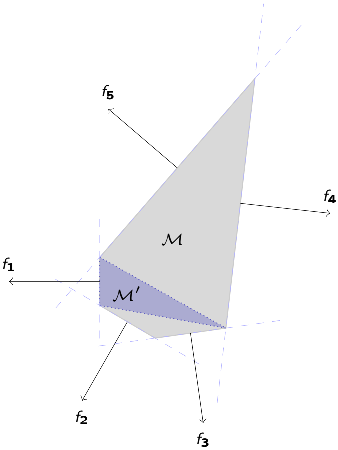
A popular model of imprecise probabilities are coherent lower probabilities: $\underline{P}(1_A)$ are given for every $A \subseteq \mathcal{X}$.

Coherent lower probabilities are also often used to approximate more general coherent lower previsions.

Example

Lower previsions $\underline{P}$ and $\underline{P}'$ with the credal sets $\mathcal{M}$ and $\mathcal{M}'$ respectively coincide on the set of gambles $\mathcal{K} = \{f_1, \ldots, f_5\}$.

(Note that $\underline{P}$ is the natural extension of $\underline{P}|_{\mathcal{K}}$.)



Contents

Formulations of the problem

Let $\underline{P}$ be a coherent lower prevision specified on a finite set of gambles $\mathcal{K}$.

- Let $\underline{P}_1$ and $\underline{P}_2$ be two extensions to $\mathbb{R}^{\mathcal{X}}$: what is the maximal possible distance between them?
- What is the maximal possible distance between an extension $\underline{P}$ and the natural extension $\underline{E}$?
- The distance denotes

$$d(\underline{P}_1, \underline{P}_2) = \max_{h \in \mathbb{R}^{\mathcal{X}}} \frac{|\underline{P}_1(h) - \underline{P}_2(h)|}{\|h\|},$$

where $\|\cdot\|$ is the Euclidean norm.

Credal set as a convex polyhedron

A credal set $\mathcal{M}$ of a coherent lower prevision $\underline{P}$ specified on a finite set $\mathcal{K}$ is a convex polyhedron:

- finite number of extreme points: linear previsions;
- finite number of faces: sets of the form $\mathcal{M}_f = \{P \in \mathcal{M} \colon P(f) = \underline{P}(f)\}$ for some gamble f.

Every extreme point is also a face.

Contents

Contents

Constraints for a credal set

The credal set $\mathcal{M}(\underline{P})$ contains vectors p satisfying the constraints:

$$p \in \mathbb{R}^{\mathcal{X}}$$
$$p \cdot 1_x \geq 0 \ \forall x \in \mathcal{X}$$
$$p \cdot 1_{\mathcal{X}} = \sum_{x \in \mathcal{X}} p(x) = 1$$

together with

$$p \cdot f \geq \underline{P}(f) \ \forall f \in \mathcal{K}.$$

Coherence requires that all inequalities in the last line are tight: for every $f \in \mathcal{K}$ there exists some p such that $p \cdot f = \underline{P}(f)$.

Transforming constraints

Each constraint of the form

$$p \cdot f \geq \underline{P}(f)$$

can be transformed into

$$p \cdot f' \geq 0$$

by taking $f' = f - \underline{P}(f)$.
(f' are thus marginally desirable gambles.)

From now on we assume that a credal set $\mathcal{M}$ is given by a finite set of tight constraints of the form:

$$p \cdot f_i \geq 0$$

and

$$p \cdot 1_{\mathcal{X}} = 1$$

Contents

University of Ljubljana

Normal cone

Let $\mathcal{M}$ be a credal set and E its boundary point.

The set

$$N_{\mathcal{M}}(E) = \{f : P(f) \geq \underline{P}(f)\}$$

is called the normal cone of $\mathcal{M}$ at point P.

The normal cone is the set of all gambles that reach minimal expectation at E.

Representation of the normal cone

For each boundary element E of $\mathcal{M}(\underline{P})$ there is a unique non-empty subset of constraints such that

$$E(f_i) = 0 \text{ for exactly} i \in I.$$

Every gamble $h \in N_{\mathcal{M}}(E)$ can be written in the form:

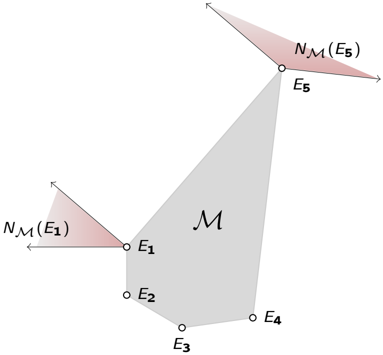
$$h = \sum_{i \in I} \alpha_i f_i + \beta 1_{\mathcal{X}}$$

for some $\alpha_i \geq 0$ and $\beta \in \mathbb{R}$.

We will call the gambles f_i for $i \in I$ the positive basis of $N_{\mathcal{M}}(E)$.

Example: normal cones

Normal cones $N_{\mathcal{M}}(E_i)$ at the extreme points are the positive hulls of the normal vectors of adjacent faces.



Contents

Distance between extreme points

Let E be an extreme point of a credal set $\mathcal{M}$ and P another linear prevision in $\mathcal{M}$.

We will need to find the maximal possible distance

$$d_E(E, P) = \max_{h \in N_{\mathcal{M}}(E)} \frac{|P(h) - E(h)|}{\|h\|}.$$

The above distance is called the normed distance of P from E.

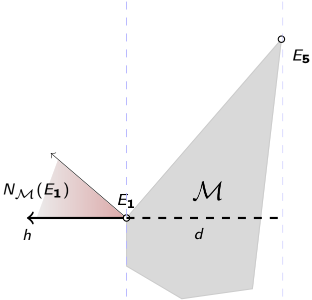
The reason for only considering elements of the normal cone is that in expression $\underline{P}(h)$ only those gambles will reach the minimal value in E.

Example: normal cones

The normed distance between E_1 and E_5 is the maximal distance on the normal cone $N_{\mathcal{M}}(E_1)$ which is reached in h.

Note that the normed distance is smaller than the Euclidean distance between the extreme points.

In higher dimensions the choice of the maximizing gamble is not so easy.



The minimal norm gambles

Given an element h from the normal cone $N_{\mathcal{M}}(E)$, we can write

$$h = \sum_{i \in I} \alpha_i f_i + \beta 1_{\mathcal{X}}$$

and $|P(h) - E(h)| = |P(h + \beta' 1_{\mathcal{X}}) - E(h + \beta' 1_{\mathcal{X}})|$.

To maximize the required norm, we must find β' where $\|h + \beta' 1_{\mathcal{X}}\|$ has the minimal norm. For this purpose, we apply an additional transformation to f_i, by subtracting a constant, to ensure that

$$f_i' \cdot 1_{\mathcal{X}} = 0$$

and then take

$$h = \sum_{i \in I} \alpha_i f_i'.$$

Setting up the problem

Every vector $(\alpha_i)_{i\in I}$ represents a minimal norm element $h \in N_{\mathcal{M}}(E)$ as

$$h = \sum_{i\in I} \alpha_i f_i$$

where we may assume that $f_i \cdot 1\mathcal{X} = 0$.
Recall that P and E are themselves vectors, and therefore we can write:

$$P(h) - E(h) = (P - E) \cdot h = D \cdot h$$

We can also decompose

$$f_i = \lambda_i D + u_i.$$

We thus obtain vectors $\underline{\alpha} = (\alpha_i)_{i\in I}$ and $\underline{\lambda} = (\lambda_i)_{i\in I}$ and a matrix U whose rows are u_i.

We have:

$$h = (\underline{\alpha} \cdot \underline{\lambda})D + \underline{\alpha}U$$
$$\|h\|^2 = \|D\|^2 \underline{\alpha\lambda}\underline{\lambda}^t \underline{\alpha}^t + \underline{\alpha}UU^t\underline{\alpha}^t$$
$$P(h) - E(h) = D \cdot (\underline{\alpha} \cdot \underline{\lambda})D = (\underline{\alpha} \cdot \underline{\lambda})\|D\|^2.$$

Further denote $\Pi = \|D\|^2 \underline{\lambda\lambda}^t + UU^t$, which is a symmetric positive semi-definite matrix.

Thus we would like to minimize the expression

$$\frac{(\underline{\alpha} \cdot \underline{\lambda})\|D\|^2}{\sqrt{\underline{\alpha}\Pi\underline{\alpha}^t}}$$

with respect to $\underline{\alpha}$.

Quadratic programming formulation

Since we may always multiply vector $\underline{\alpha}$ by a positive constant, we can always ensure the numerator in

$$\frac{(\underline{\alpha} \cdot \underline{\lambda})\|D\|^2}{\sqrt{\underline{\alpha}\Pi\underline{\alpha}^t}}$$

to be equal 1.

In this case, we can maximize the above expression by minimizing the norm:

$$\underline{\alpha}\Pi\underline{\alpha}^t$$

subject to

$$\begin{aligned} (\underline{\alpha} \cdot \underline{\lambda})\|D\|^2 &= 1 \\ \underline{\alpha} &\geq 0 \end{aligned}$$

University *of Ljubljana*

Contents

Coherent lower previsions coinciding on a set of gambles

Let two coherent lower previsions coincide on a set of gambles $\mathcal{K}$.

Without loss of generality we will assume that one of them is the natural extension $\underline{E}$ of $\underline{P}|_{\mathcal{K}}$.

Coherence implies that $\underline{P}(f) = \underline{E}(f)$ for every $f \in \mathcal{K}$.

Let $\mathcal{M}$ and $\mathcal{C}$ be the credal sets of $\underline{E}$ and $\underline{P}$.

Coherence implies

$\mathcal{C} \cap \mathcal{M}_f \neq \emptyset$ for every $f \in \mathcal{K}$, where $\mathcal{M}_f$ is a face of $\mathcal{M}$.

Maximal error on a gamble

Take any gamble h.

Question

What is the maximal possible distance $(\underline{P}(h) - \underline{E}(h))/\|h\|$ (notice that this is always non-negative).

We easily notice that

- There is some extreme point E of $\mathcal{M}$ so that $\underline{E}(h) = E(h)$;
- Also the minimum of $P(h) - E(h)$ over $P \in \mathcal{C}$ is reached in an extreme point $P \in \mathcal{C}$;
- h belongs to the normal cone $N_{\mathcal{M}}(E)$.

The maximal distance is clearly related to the normed distance described before.

Some additional observations

- The maximal distance $P(h) - E(h)$ is less than $\max_{P \in \mathcal{M}_f} P(h) - E(h)$ for any face $\mathcal{M}_f$.
- In fact, there always exists a face $\mathcal{M}_f$ so that the maximal possible value of $\max_{P \in \mathcal{C}} P(h) - E(h)$ over all possible credal sets $\mathcal{C}$ is equal to $\max_{P \in \mathcal{M}_f} P(h) - E(h)$.
- $\max_{P \in \mathcal{M}_f} P(h) - E(h)$ is an estimate of the maximal possible distance.

Contents

Practical estimation of the maximal possible distance

The above observations suggest that the following steps will provide the exact maximal possible distance between the natural extension of $\underline{P}|_{\mathcal{K}}$ and any other extension.

Step 1: Find extreme points of $\mathcal{M}$

There exist efficient algorithms for finding extreme points of convex polyhedra. Unfortunately, they are computationally expensive.

Step 2: For every extreme point find the maximal possible normed distance

This step requires finding the face with the minimal value of $\max_{P \in \mathcal{M}_f} P(h) - E(h)$. If only an estimate is required, we may only consider the faces that are closest to E. In most cases this is sufficient.

Calculating the normed distance

The normed distance is calculated by solving the quadratic programming problem described before. This is by far slowest step. Therefore we can measure the performance of the algorithm by the number of calls to this routine.

Results

The algorithm was tested on random generated lower coherent lower probabilities. Below are average results.

dimension	extreme points	distances calculated
3	5,9	11,8
4	23,6	124,2
5	101,2	1697,2
6	592,3	31179,7
7	2744,7	586728,0

Contents

University *of Ljubljana*

Computational complexity

The high computational complexity presents an obstacle to applying the method in high dimensional cases.

Possible solutions

- Quick approximations of the maximal distance.
- Optimization of the existing algorithm.

Related problems

- How to select a gamble h for which to compute $\underline{P}(h)$ so that the maximal error for extending $\underline{P}|_{\mathcal{K}\cup\{h\}}$ is as small as possible?
- How to select k such gambles?
- Suppose you have $\underline{P}$ and $\underline{P}'$ given on finite sets $\mathcal{K}$ and $\mathcal{K}'$:
  - How much can two coherent extensions differ?
  - How much do their natural extensions differ?