ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

Ryan Martin
North Carolina State University
www4.stat.ncsu.edu/~rmartin

Week 06b

This lecture

- Natural extension
- Conditional lower previsions
- Generalized Bayes rule
- ...

Introduction

- Last time we considered a collection of gambles $\mathcal{K}$
- Pair of (dual) functionals $(\underline{P}, \overline{P})$ on $\mathcal{K}$:

$$\underline{P}(f) = \sup\{\mu \in \mathbb{R} : f - \mu \text{ is desirable}\}$$
$$\overline{P}(f) = \inf\{\mu \in \mathbb{R} : \mu - f \text{ is desirable}\}$$

- "Desirability" means that I'd accept the gamble if offered:
  - I'd buy f from you for anything less than $\underline{P}(f)$
  - I'd sell f to you for anything more than $\overline{P}(f)$
- $\underline{P}$ is called the *lower prevision*
- Simple sufficient conditions on $\underline{P}$ to ensure coherence
  - ...
  - $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$

Intro, cont.

- Relatively easy to check coherence
- Holds for lower previsions induced by those imprecise probabilities we discussed before, e.g., belief functions
- Might want to modify $\underline{P}$ for some reason:
  - extend its domain from $\mathcal{K}$ to a larger $\mathcal{K}'$
  - or incorporate some newfound knowledge about X
- For these, there two such modifications:
  - natural extension
  - conditioning / generalized Bayes rule

Natural extension

> *Natural extension may be seen as the basic constructive step in statistical reasoning; it enables us to construct new previsions from old* —Peter Walley

- It won't look like "statistical reasoning" to us, at least not yet
- Similar to the *extension principle* from possibility theory
- Mathematical abstraction:
  - we have $\underline{P}$ defined on $\mathcal{K}$
  - "new" gambles are presented, larger domain $\mathcal{K}'$
  - how to *extend* $\underline{P}$ to $\mathcal{K}'$?
  - i.e., how to evaluate "$\underline{P}(h)$" for $h \in \mathcal{K}' \setminus \mathcal{K}$?

Natural extension, cont.

- Key points:
  - presumably $\underline{P}$ is coherent, so the extension should be too
  - extension should agree with $\underline{P}$ on $\mathcal{K}$
- Intuition:
  - let h be a particular gamble, i.e., one in $\mathcal{K}' \setminus \mathcal{K}$
  - μ a generic number
  - suppose there exists n, $\alpha_i \geq 0$, $f_i \in \mathcal{K}$, and $\delta > 0$ s.t.,

$$\inf_{x \in \mathbb{X}} \left[\{h(x) - \mu\} - \sum_{i=1}^{n} \alpha_i \{f_i(x) - \underline{P}(f_i) + \delta\} \right] \geq 0$$

  - summation term is desirable, so $h - \mu$ is desirable
  - so ought to be willing to pay at least μ for h
  - now find largest such $\mu = \mu(n, \alpha_i, f_i)$...

Natural extension, cont.

Definition.

The *natural extension* of $\underline{\mathrm{P}} : \mathcal{K} \to \mathbb{R}$ to $\mathcal{K}' \supseteq \mathcal{K}$ is

$$\underline{\mathrm{E}}(h) = \sup_{n, \alpha_i \geq 0, f_i \in \mathcal{K}} \inf_{x \in \mathbb{X}} \left[h(x) - \sum_{i=1}^{n} \alpha_i \{ f_i(x) - \underline{\mathrm{P}}(f_i) \} \right], \quad h \in \mathcal{K}'$$

Theorem.

Let $\underline{\mathrm{P}}$ be a lower prevision on $\mathcal{K}$ and $\underline{\mathrm{E}}$ its natural extension to $\mathcal{K}'$

1. $\underline{\mathrm{E}}$ is the smallest coherent lower prevision that dominates $\underline{\mathrm{P}}$ on $\mathcal{K}$
2. $\underline{\mathrm{E}}$ agrees with $\underline{\mathrm{P}}$ on $\mathcal{K}$ iff $\underline{\mathrm{P}}$ is coherent

Example, cont.[1]

- Experiment:
  - bag contains blue, green, and red balls
  - X is the color of a sampled ball, so $\mathbb{X} = \{B, G, R\}$
- Two gambles in $\mathcal{K}$:
  - $f(B) = 0$, $f(G) = 10$, $f(R) = 5$
  - $g(B) = 9$, $g(G) = 0$, $g(R) = 5$
- $\underline{P}(f) = 5$ and $\underline{P}(g) = 4$
- New gamble: $h(B) = 4$, $h(G) = 2$, $h(R) = 3$
- Lower prevision for h (from that on $\{f, g\}$)?

[1] Taken from Miranda & De Cooman, Ch. 2 of *Introduction to IP*

Example, cont.

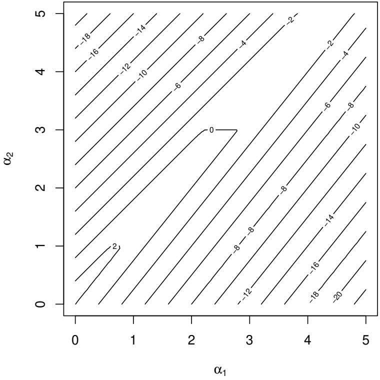
- Different cases:
  - if $x = B$, then diff $= 4 - \alpha_1(0-5) - \alpha_2(9-4)$
  - if $x = G$, then diff $= 2 - \alpha_1(10-5) - \alpha_2(0-4)$
  - if $x = R$, then diff $= 3 - \alpha_1(5-5) - \alpha_2(5-4)$
- "$\inf_x$" is the minimum of these three
- Induced lower prevision via natural extension is the max over (α_1, α_2) of this minimum, i.e.,

$$\underline{E}(h) = \sup_{\alpha_1, \alpha_2 \geq 0} \min\{4 + 5\alpha_1 - 5\alpha_2, 2 - 5\alpha_1 + 4\alpha_2, 3 - \alpha_2\}$$

Example, cont.

$$\underline{E}(h) = \sup_{\alpha_1, \alpha_2 \geq 0} \min\{4 + 5\alpha_1 - 5\alpha_2, 2 - 5\alpha_1 + 4\alpha_2, 3 - \alpha_2\}$$

- Supremum is near the origin
- Minimum comes from latter two terms
- Attained when two are equal
- $(\hat{\alpha}_1, \hat{\alpha}_2) = (0, 0.2)$
- $\underline{E}(h) = 2.8$



Conditional lower previsions

- There's a lot involved in this theory[2]
- I'm just going to wave my hands at a specific part

- Assessments about X are translated into a lower prevision $\underline{\mathrm{P}}$
- Then we learn that $X \in B$
- How do we update $\underline{\mathrm{P}}$ in light of this new info?
- This update is a *conditional lower prevision*

[2] Chapter 2.3 in *Intro to IP* barely touches all that's in Walley's book

Conditional lower previsions, cont.

- Let $\mathcal{B} = \{B_1, B_2, \dots\}$ be a partition of $\mathbb{X}$
- For any $B \in \mathcal{B}$, let

$$\underline{P}(f \mid B) = \text{sup buying price for } f \text{ \emph{after} learning } B \text{ occurred}$$

- LHS is just a *symbol* for the RHS
  - RHS is a judgment made by us/agent
  - question is how to make judgments in a "coherent" way
- What does "coherence" mean in this context?

- Notation: $\underline{P}(f \mid \mathcal{B}) = \sum_{B \in \mathcal{B}} 1_B \, \underline{P}(f \mid B)$

Conditional lower previsions, cont.

- In this context, there are various notions of coherence
- Collection $\{\underline{P}(\cdot \mid B) : B \in \mathcal{B}\}$ are called *separately coherent* if
  - each $\underline{P}(\cdot \mid B)$ is coherent like in previous lecture
  - $\underline{P}(1_B \mid B) = 1$ for each B
- Alternatively, let's consider the relationship between conditional and unconditional lower previsions
- The details of *joint coherence* are too technical for me to present here in a comprehensible way[3]
- This notion is important so I'm going to focus on a (overly?) simplified version...

[3] See Walley, Ch. 6.5, and Miranda & De Cooman, Sec. 2.3.3

Conditional lower previsions, cont.

- Let's say $\underline{P}(\cdot \mid \cdot)$ and $\underline{P}(\cdot)$ are *jointly coherent*[4] if

$$\inf_{B \in \mathcal{B}} \underline{P}(f \mid B) \leq \underline{P}(f), \quad f \text{ in domain } \mathcal{K}$$

- Similar condition for upper previsions using conjugacy
- Intuition:
  - suppose condition fails
  - then there exists f with $\underline{P}(f \mid B) > \underline{P}(f)$ for all B
  - so I'll pay strictly more for f after B is revealed than before, *no matter which B it is*
  - therefore, my original $\underline{P}(f)$ must be too low
- If original assessments are satisfactory, then the goal is to *define* so that joint coherence holds

[4] This is only half of the definition, see (C8) in Walley, Ch. 6.5.2

Generalized Bayes rule

- Recall that coherent lower previsions correspond to lower envelopes of (closed and convex) sets of previsions
- Intuition:
  - if we had a prevision to start, then we'd update in a coherent way by applying conditional probability/expectation
  - original $\underline{P}$ determines a set of previsions
  - just get conditional previsions for each one
  - then define conditional lower prevision as the lower envelope
- This intuition can be made formal
- Corresponds to the so-called *generalized Bayes rule*

Generalized Bayes rule, cont.

Theorem.

Suppose that $\underline{\mathrm{P}}(B) > 0$. Then the *generalized Bayes rule* is

$$\underline{\mathrm{P}}(f \mid B) = \inf\left\{\frac{\mathrm{P}(1_B\, f)}{\mathrm{P}(B)} : \mathrm{P} \in \mathscr{C}(\underline{\mathrm{P}})\right\}$$

- This is a consequence of (the full version of) joint coherence, provided that $\underline{\mathrm{P}}(B) > 0$
- That is, (full-blown) joint coherence determines the form of the conditional prevision in this case
- With more care, the "$\underline{\mathrm{P}}(B) > 0$" condition can be relaxed

Generalized Bayes rule, cont.

- There are immediate statistical implications
- Suppose $X = (Y, \Theta) \in \mathbb{Y} \times \mathbb{T}$, data–parameter pair
- Lower prevision for X might be based on
  - a (precise) model for Y, given $\Theta = \theta$
  - an imprecise prior prevision for Θ
- $\mathcal{B} = \{\{y\} \times \mathbb{T} : y \in \mathbb{Y}\}$ based on realizations of Y
- All we have to do is apply Bayes's rule to each prevision that's compatible with the specified $\underline{\mathrm{P}}$ for (Y, Θ)[5]
- Separate and joint coherence

- This is the context I'm working in now,[6] developing an *efficient alternative to generalized Bayes*

---

[5] See Ch. 7 of *Intro to IP*

[6] e.g., `http://arxiv.org/abs/2203.06703`

Next lecture

- Comparison of Dempster's & generalized Bayes rules
- Dilation & contraction, connection to sure-loss/incoherence
- Statistical perspectives
- ...

- Based largely on Gong & Meng (2021 *Stat Sci*)[7]

[7] `https://ruobingong.github.io/files/GongMeng2021_StatSci.pdf`