ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

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

Week 06a

This lecture

- New terminology, e.g., desirability
- De Finetti's theory of previsions (expectations)
- Imprecise version: lower and upper previsions
- Properties: coherence
- Examples

Introduction

- In probability theory, we usually start with P and, from there, build the notion of E
- Alternatively, could start with E and define P by a restriction to indicator functions, i.e., $\mathrm{P}(A) = \mathrm{E}(1_A)$
- The two are equivalent, so it's a matter of taste
- In imprecise probability, there are differences
- Benefits to generalizing the approach that starts with E...

Intro, cont.

- Benefits:
  - in the imprecise prob case, lower/upper probs don't uniquely determine lower/upper expectation
  - coherence conditions are more direct for expectations
- So it makes sense to build a general theory of imprecise probability based on lower/upper expectations
- All of the models we've discussed so far are special cases:
  - $(\underline{\mathrm{P}}, \overline{\mathrm{P}}) \to (\underline{\mathrm{E}}, \overline{\mathrm{E}})$ via Choquet integral
  - properties of $\underline{\mathrm{P}}$ lead to properties of the induced $\underline{\mathrm{E}}$
- I presented the other theories first because I think those are more accessible starting points...

Setup, notation, terminology

- $\mathbb{X}$ is a general space; we can take it to be finite
- $f : \mathbb{X} \to \mathbb{R}$ is called a *gamble*[1][2]
  - payoff is $f(X)$ utils, but $X \in \mathbb{X}$ is uncertain
  - e.g., $f \equiv 0$ means do nothing, no bet, no payoff
  - e.g., $f(x) = 1_A(x)$ means 1 util payoff if $x \in A$, 0 otherwise
- $\mathcal{K}$ is a collection of gambles
  - need not have specific structure, e.g., 1_A might not be in $\mathcal{K}$
  - I'll assume it's a *linear space*, i.e., closed under linear combos
- A gamble is *desirable* (to me) if I would accept it if offered
  - e.g., $\sup f < 0$ means sure loss, not desirable
  - e.g., $f(x) = 1_A(x) - \mu$ might be desirable to me or not, depends on my beliefs about A and magnitude of μ

---

[1] Walley's work takes gambles to be bounded functions, but the general case is addressed in Troffaes & De Cooman

[2] If $\mathbb{X}$ is finite, then gambles are just vectors

Setup, cont.

Axioms of desirability.

1. If $f \leq 0$ and $f \neq 0$, then not desirable
2. If $f \geq 0$ and $f \neq 0$, then desirable
3. If f is desirable and $\alpha > 0$, then αf is desirable
4. If f, g are desirable, then $f + g$ is desirable

- There's a non-empty set $\mathcal{D} \subset \mathcal{K}$ of desirable gambles, a cone
- The axioms *don't* determine which gambles are desirable, just like Kolmogorov's axioms don't determine the prob values
- Question: how to specify $\mathcal{D}$ so that I avoid incoherence?
- Two equivalent ways to proceed:
  - × formulate coherence etc. in terms of the set $\mathcal{D}$
  - ✓ let $\mathcal{D}$ be determined by (lower/upper) previsions and formulate coherence in terms of these

De Finetti's previsions

- De Finetti realized that there's no reason to restrict to bets on events; more convenient to focus on general gambles
- A *prevision*[3] is a map from gambles to fair prices
- That is, $\mathrm{P}(f)$ is the value μ such that

$$\alpha(f - \mu) \text{ is desirable for all } \alpha \in \mathbb{R}$$

- As we expect, De Finetti's conclusion is:
    - prevision is *coherent* if it corresponds to an expectation
    - that is, if[4] it's linear, $\mathrm{P}(f + g) = \mathrm{P}(f) + \mathrm{P}(g)$
- Requires pinning down a precise fair price for all gambles

[3] Here P stands for price/prevision, not probability

[4] also requires that $\mathrm{P}(f) \geq \inf_x f(x)$

Lower (and upper) previsions

- Since setting fair prices for all gambles may not be realistic, try relaxing by only requiring lower and upper bounds
- *Lower* and *upper previsions*, denoted by $\underline{\mathrm{P}}$ and $\overline{\mathrm{P}}$, are functionals mapping gambles in $\mathcal{K}$ to real numbers

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

- In other words:
  - $\underline{\mathrm{P}}(f)$ is my sup buying price for f
  - $\overline{\mathrm{P}}(f)$ is my inf selling price for f
- Technical note:
  - we don't specify whether, e.g., $f - \underline{\mathrm{P}}(f)$ is desirable,
  - only that $f - \{\underline{\mathrm{P}}(f) - \delta\}$ is desirable for all $\delta > 0$

Lower/upper previsions, cont.

- There's a conjugacy relationship between $\underline{\mathrm{P}}$ and $\overline{\mathrm{P}}$

$$\overline{\mathrm{P}}(f) = \cdots = -\underline{\mathrm{P}}(-f), \quad f \in \mathcal{K}$$

- Like what we had for lower upper prob: $\overline{\mathrm{P}}(A) = 1 - \underline{\mathrm{P}}(A^c)$
- So it suffices to study properties of only one of the previsions in the pair — the literature focuses on $\underline{\mathrm{P}}$
- Speaking of lower/upper probabilities...
  - for sets A with $1_A \in K$, the lower prob can be defined in the natural way, i.e., $\underline{\mathrm{P}}(A) = \underline{\mathrm{P}}(1_A)$
  - $\underline{\mathrm{P}}$ can be *extended* to indicators if $\mathcal{K}$ isn't big enough

Example[5]

- Experiment:
  - bag contains blue, green, and red balls
  - X is the color of a sampled ball, so $\mathbb{X} = \{B, G, R\}$
- A desirable gamble: $f(B) = 0$, $f(G) = 10$, $f(R) = 5$
- I might be willing to pay some positive amount for f
- e.g., if I don't believe there are any B's, then $\underline{P}(f) = 5$
- It's in this sense that the lower/upper previsions I set correspond to degrees of belief

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

Example, cont.

- New gamble: $g(B) = 9$, $g(G) = 0$, $g(R) = 5$
- Suppose I set $\underline{P}(g) = 6$ — is this choice reasonable?
  - This specification, along with the desirability axioms, implies that I'm willing to pay up to \$11 for $f + g$
  - Absolute most I can win is \$10
  - Sure loss!
- So I'll set $\underline{P}(g) = 4$ to avoid sure loss
- But there are stricter notions of self-consistency we want

Coherence

- Like we saw before, $\underline{P}$ is *coherent* if

$$\sup_{x \in \mathbb{X}} \Big[\sum_{i=1}^{n} \{ f_i(x) - \underline{P}(f_i) \} - a \{ f_0(x) - \underline{P}(f_0) \} \Big] \geq 0, \quad \forall\ n,\ f_i,\ a \geq 0$$

- Intuition:
  - suppose the above condition fails
  - then there exists n, f_i, $a \geq 0$, and $\varepsilon, \delta > 0$ such that

$$\sum_{i=1}^{n} [f_i(x) - \{\underline{P}(f_i) - \varepsilon\}] \leq a[f_0(x) - \{\underline{P}(f_0) + \delta\}], \quad \forall\ x$$

  - LHS is desirable by definition, so RHS is too
  - then $f_0 - \{\underline{P}(f_0) + \delta\}$ is desirable
  - contradicts $\underline{P}(f_0)$ being the sup μ s.t. $f - \mu$ is desirable

Example, cont.

X	B	G	R
f	0	10	5
g	9	0	5
$f-5$	-5	5	0
$g-4$	5	-4	1
$6-g$	-3	6	1
$5-g$	-2	5	0

- Assessments: $\underline{P}(f)=5$, $\underline{P}(g)=4$, $\overline{P}(g)=??$
- None of the rows are strictly negative, so no sure loss
- Compare the two assessments of $\overline{P}(g)$
  - $f-5$ is desirable by assumption
  - $f-5 \leq 5-g$
  - so $5-g$ is desirable
  - then $\overline{P}(g)=6$ doesn't make sense

Coherence, cont.

Theorem.

If $\mathcal{K}$ is a linear space, then a lower prevision $\underline{P} : \mathcal{K} \to \mathbb{R}$ is *coherent* iff Properties 1–3 below hold:

1. $\underline{P}(f) \geq \inf_x f(x)$ for all $f \in \mathcal{K}$
2. $\underline{P}(\alpha f) = \alpha \underline{P}(f)$ for all $f \in \mathcal{K}$ and all $\alpha > 0$
3. $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$ for all $f, g \in \mathcal{K}$

- Relatively simple set of necessary & sufficient conditions
- Application:
  - Claim: lower prevision induced by a belief function is coherent
  - Recall that the Choquet integral is

$$\underline{P}(f) = \sum_A m(A) \left\{ \min_{x \in A} f(x) \right\}$$

Coherence, cont.

Theorem.

If $\underline{P} : \mathcal{K} \to \mathbb{R}$ is *coherent*, then, e.g.,

- $\underline{P}\{\alpha f + (1-\alpha)g\} \geq \alpha\underline{P}(f) + (1-\alpha)\underline{P}(g)$
- $|\underline{P}(f) - \underline{P}(g)| \leq \overline{P}(|f-g|)$
- $\underline{P}$ is continuous (wrt the topology of uniform convergence)

Theorem.

Suppose that $\underline{P}^\lambda$ is a coherent lower prevision on $\mathcal{K}$ for each $\lambda \in \Lambda$. Then the lower prevision $\underline{P}^{\inf}$, given by

$$\underline{P}^{\inf}(f) = \inf_{\lambda \in \Lambda} \underline{P}^\lambda(f), \quad f \in \mathcal{K},$$

is coherent too

Sets of previsions

- Recall from our discussion of coherent *lower probabilities*
  - coherent iff $\underline{P}(A) = \inf_{P \in \mathscr{C}(\underline{P})} P(A)$ for all A
  - so $\underline{P}$ and $\mathscr{C}(\underline{P})$ are equivalent
- Similar results for coherent *lower previsions*
- Define $\mathscr{C}(\underline{P}) = \{\text{previsions } P\colon\ P(f) \geq \underline{P}(f) \text{ for all } f \in \mathcal{K}\}$

Theorem.

- $\underline{P}$ is coherent iff it's the lower envelope of $\mathscr{C}(\underline{P})$
- If $\underline{P}$ is coherent, then $\mathscr{C}(\underline{P})$ is "compact" and convex
- If $\mathscr{P}$ is a "compact" and convex collection of previsions, then its lower envelop is a coherent lower prevision

Example, cont.

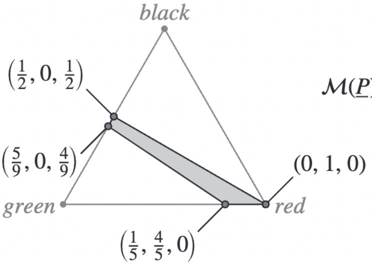
- Recall: $\underline{\mathrm{P}}(f) = 5$ and $\underline{\mathrm{P}}(g) = 4$
- This is a coherent lower prevision on $\mathcal{K} = \{f, g\}$[6]
- Previsions are determined by probability vectors, the simplex

$$\begin{aligned}\mathscr{C}(\underline{\mathrm{P}}) &= \{\mathrm{P} : \mathrm{P}(\cdot) \geq \underline{\mathrm{P}}(\cdot) \text{ on } \mathcal{K}\} \\ &= \{(p_{B}, p_{G}, p_{R}) : \mathrm{P}(f) \geq 5,\ \mathrm{P}(g) \geq 4\} \\ &= \text{solutions to the system } \begin{cases} 10p_{G} + 5p_{R} \geq 5 \\ 9p_{B} + 5p_{R} \geq 4 \end{cases}\end{aligned}$$

[6]This isn't a linear space, but that's not necessary

Example, cont.

$$\mathscr{C}(\underline{\mathrm{P}}) = \text{solutions to the system } \begin{cases} 10p_G + 5p_R \geq 5 \\ 9p_B + 5p_R \geq 4 \end{cases}$$



$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) : P(f) \geq 5 \text{ and } P(g) \geq 4\}.$$

Notation: Miranda & De Cooman's "black" is my B, their "$\mathcal{M}(\underline{P})$" is my $\mathscr{C}(\underline{\mathrm{P}})$

Next lecture

- Natural extension
- Conditional lower previsions
- Generalized Bayes rule