ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

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

Week 15b

This lecture

- Course wrap-up
- Details:
  - what we covered
  - what we didn't cover
- Take-away messages
- The end

What was the plan again?

From Week 01a

Course outline, general

- My goal is achieve more breadth than depth, covering material from various sources — books and papers
- So, my presentation won't be 100% rigorous, but we will get into some technical details that require care
- Beyond technical details, there will be some philosophical issues that we have to deal with along the way too
- My coverage won't be comprehensive, there are interesting and relevant topics we won't cover
- Those skipped topics are fair game for the project!

24 / 25

Probability

- Interpretations of probability:[1]
  - frequentist
  - subjective — bets and de Finetti's coherence
- Shortcomings:
  - can't model (partial) ignorance
  - can't distinguish aleatory & epistemic uncertainty
  - false confidence phenomenon
- These shortcomings are all relevant for statistics & ML:
  - "All models are wrong..." (Box)
  - "[Bayes's rule] does not create real probabilities from hypothetical probabilities" (Fraser)
  - it's "unacceptable if a procedure... of representing uncertain knowledge would, if used repeatedly, give systematically misleading conclusions" (Reid & Cox)

[1] Lots of other perspectives, e.g., Keynes, Carnap, Popper, ...

Towards imprecision

> *As Bayes perceived, the concept of Mathematical Probability affords a means, in some cases, of expressing inferences from observational data, involving a degree of uncertainty, and of expressing them rigorously, in that the nature and extent of the uncertainty is specified with exactitude, yet it is by no means axiomatic that the appropriate inferences, though in all cases involving uncertainty, should always be rigorously expressible in terms of this same concept* —Ronald Fisher

My translation:

Probability can be used in some cases to quantify uncertainty precisely (i.e., "with exactitude"), but there are other cases wherein the appropriate inferences ought to be "rigorously" expressed in terms of *something else*

Towards imprecision, cont.

- What's the alternative that Fisher is eluding to?
- He didn't say — I think because he didn't know
- In retrospect, I think it's clear that what Fisher and many others have sought[2] can't be done with ordinary probability
- To me, a new perspective is needed
- So, I didn't get into imprecise prob because I wanted a more general theory — *I think we need this theory*[3]

[2] Aside from the "most important unresolved problem" quote in Week 01a, Efron once referred to this in a talk as the "Holy Grail" of statistics

[3] This is a 100 year-old problem, standard tools aren't gonna cut it

Imprecise probability

- Start with a more general set function, a *capacity*
- K-monotone, K-alternating, etc.
- First criticism:
  - what about de Finetti's result?
  - if not a precise, (finitely) additive prob, then *incoherent*
- We eventually confirmed that the following properties of a capacity are equivalent:
  - K-monotone for $K \geq 2$
  - corresponds to a closed & convex set of precise prob's
  - coherent, no sure loss, etc.
- So, we're on solid mathematical ground provided we work with K-monotone capacities, with $K \geq 2$
- Still leaves lots of degrees of freedom...

Different imprecise prob models

- Random sets
  - easiest entry point to this theory
  - mathematically elegant and rigorous
  - many other models are special cases
  - K-monotone with $K = \infty$
- Possibility measures
  - Shackle & Zadeh: UQ alternatives to probability
  - mathematically: *nested* random sets
  - *consonant*, determined by a contour function
  - Shafer: consonance makes sense for stat inference
  - extension principle
  - characterizations of the credal set, i.e.,

$$\mathsf{P} \in \mathscr{C}(\overline{\Pi}) \iff \mathsf{P}\{\pi(X) \leq \alpha\} \leq \alpha, \quad \alpha \in [0,1]$$

Different imprecise prob models, cont.

- Belief functions
  - K-monotone with $K = \infty$
  - equivalent to random sets in finite-$\mathbb{X}$ case
  - originated with Dempster, efforts to improve fiducial
  - part of Shafer's "mathematical theory of evidence"
  - characterization of credal set in terms of allocations
  - key feature is Dempster's rule of combination
- Lower/upper previsions
  - most general mathematical framework
  - e.g., *all* closed & convex sets of precise probabilities correspond to a lower prevision
  - behaviorally motivated, *coherence is king*
  - previsions for new gambles based on *natural extension*
  - *generalized Bayes rule*, updates preserve coherence

High-level comparison

- Possibility measures are the simplest, most constraints, but compatible with statistical reasoning
- Belief functions are more general than possibility, since not all *evidence* is of the statistical variety
- Theory of lower previsions insists on coherence:
  - generalized Bayes rule is strictly the only way to ensure coherence in updating beliefs
  - e.g., Dempster's update can be incoherent
  - generalized Bayes tends to be conservative
- Some other models we didn't discuss:
  - interval probabilities
  - p-boxes
  - desirable gambles
  - transferable belief model

Applications

- I selfishly focused mostly on statistical inference
  - this is what I know best, most interesting to me
  - lots of potential impact for imprecise prob
  - even more on this below :)
- Some relatively high-level coverage of topics that fall (slightly) outside boundaries of statistical inference:
  - classification
  - regression
  - decision theory
- In each case, we discussed
  - classical solutions
  - their shortcomings
  - how imprecise prob can help to fill the gap
- Nowhere near exhaustive coverage

Regrets: things we didn't talk about

- Dependence notions in imprecise probability
- Game-theoretic probability
- Causal inference
- It would've been nice to dig deeper into:
  - mathematics of imprecise prob
  - stat/ML/econometrics applications
  - computational methods, e.g., for natural extension

Stat-related loose ends

- Statistics has two mainstream theories:
  - frequentist
  - Bayesian
- Clearly neither is satisfactory
- This is embarrassing for the field — gives the impression that we don't know what we're doing or that it's not serious
- Even worse, both theories miss almost all real problems
- A Bayesian's precise prior knowledge and a frequentist's prior ignorance are impractical extremes
- Despite the importance of statistics to the advancement of science, etc., still no urgency to resolve these issues?
- It's not enough to have a bunch of methods that "work"!

Stat-related loose ends, cont.

- Necessary conditions for a "resolution"
  - can't be framed within an existing theory
  - can't completely abandon the priorities of an existing theory
  - can't give completely different answers in standard examples
- The new inferential model (IM) construction I presented[4] in lecture satisfies these conditions
  - based on imprecise prob,[5] e.g., can handle partial prior info
  - guarantees strong validity and most of update-coherence
  - mostly agrees with freq/Bayes in respective contexts
- There are only necessary conditions available, so it's impossible to prove that a proposal like this is "right"
- Please: *tell me how my proposal is wrong!*

[4] Details here now: `https://arxiv.org/abs/2211.14567`

[5] Outer consonant approximation via IP-to-possibility transform

A few more details

- Important aspect of IM validity that I didn't stress enough
- Recall that (strong) validity *wrt the posited model* is

$$\overline{\mathsf{P}}_{Y,\Theta}\{\pi_Y(\Theta) \leq \alpha\} \leq \alpha, \quad \alpha \in [0,1]$$

- Model assumptions, in $\overline{\mathsf{P}}_{Y,\Theta}$, are built into target property
- Other approaches aim for properties *wrt vacuous model*
- Not always possible to achieve (something like) validity wrt vacuous model AND efficiency
- If I was willing to make model assumptions, *when I didn't have to*, then these should be reflected in the target property

A few more details, cont.

- Connection between new IM stuff and fiducial-like methods?
- I hinted at such a connection before, but no details
- Brand new result:
  - for a collection of models wherein fiducial argument applies[6]
  - the fiducial distribution is the "maximally diffuse" element in the IM's credal set
- The point:
  - valid IMs are always "right" (??)
  - in cases where fiducial-like things work, they correspond to the best precise-probabilistic approximation of the IM
- This sheds new light on
  - what Fisher was missing in his fiducial argument
  - IP and Efron's "most important unresolved problem"
  - false confidence phenomenon

[6] These are called *group invariant models*

Things I'm doing/thinking about

- Interpretation of IM output?
  - not frequentist
  - also not subjective/behavioral
- Model/structural uncertainty
- Decision theory
- Partial-prior elicitation
- Computation
- Applications
- ...

Take-away messages

- "Probability does not exist" (De Finetti)
- That is, probability only has meaning when you believe it
- You're not obligated to believe in any probability — imprecise probability is designed specifically for such cases
- Precision implies simplicity, but at what cost?
- Manski: *The credibility of inferences decreases with the strength of assumptions maintained*
- Even if you forget all the technical details we've covered, please don't forget about Manski!

Take-away messages, cont.

- You read lots of books and papers about stat/ML and none talk about imprecision, so maybe I'm just crazy
- But almost none of that will ever be used in real life[7]
- It's easy to insist on simplicity & precision when there are no real consequences, no *skin in the game*
- Credibility of inference matters a lot more than simplicity & precision when, e.g., lives are on the line
- Our foundations ought to assume that there are lives on the line and then relax as the applications allow

> *I'm talking about an extra type of integrity that is beyond not lying, but bending over backwards to show how you're maybe wrong, that you ought to have as a scientist.* —R. Feynman

[7] Not a criticism, just the truth; "publication" $\neq$ "real life"

The end

- Thank you for your...[8]
  - attention
  - patience
  - questions
- I learned a lot from this experience, hope you did too
- I'd be happy to talk more if you have questions, ideas, etc.
- Please share any feedback about the course with me

- Don't forget about IP-related societies and conferences:
  - SIPTA and BFAS
  - ISIPTA'23: `https://isipta23.sipta.org/`

[8] See the acknowledgment in `https://arxiv.org/abs/2211.14567`