ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

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

Week 01a

Today's lecture

- Boring administrative stuff
- High-level intro
  - what is imprecise probability?
  - why should I care about it?
  - seriously, what is imprecise probability?
- More detailed intro and motivation
  - some statistics background
  - Fisher and imprecision
- Course plan, objectives, etc.

Admin stuff

- Detailed syllabus is available on the course website
- Prerequisites/expectations:
  - ✓ probability/math-stat as in Casella & Berger
  - X measure-theoretic probability
  - X advanced statistical inference
  - X Bayesian inference
- No required textbook, but references are on the website
  - books (newer ones might have electronic copies)
  - journal/proceedings papers (open access if I can)
- I'll do some computing with R[1]
- For the registered students, grades are based on
  - occasional homework and "participation"
  - course project

[1] Available for free: `https://cran.r-project.org/`

Admin stuff, cont.

- Topic and objective are flexible
- *My vision:* Investigate how imprecise prob-related ideas could benefit efforts to solve a particular stat/ML problem[2]
- Basic structure:
  - students select a topic of interest
  - do some background reading, experimentation
  - summarize findings, new ideas, etc.
- Students work individually or in pairs
- Instructor can provide guidance along the way
- Could evolve into research papers or more...

[2] e.g., *causal inference* got considerable attention at ISIPTA'21

Admin stuff, cont.

- Admin stuff (mostly) clear?
- More details will be provided on the course website
- Questions?

Instructor

- Dr. Ryan Martin
- PhD, 2009, Statistics, Purdue
- NCSU Statistics since 2016
- co-author of *Inferential Models*
- co-founder of *Researchers.One*
- recent research efforts:
  - foundations of statistics
  - imprecise probability
  - .....

(Blog post about *Inferential Models* and how I got interested in imprecise probability: `https://www.sipta.org/blog/book-inferential-models/`)

Instructor, cont.

- I don't claim to know everything, I'm here to learn too
- Full disclosure: some of what I'll present in these lectures are things I'm developing in real time
- I think there are exciting opportunities for stat+IP, but the route is uncertain — that's what makes it fun!
- Feel free to ask questions, make suggestions, etc.

Introduction

> *So far as the laws of mathematics refer to reality, they are uncertain, and so far as they are certain, they do not refer to reality* —Albert Einstein

- *Imprecise probability* is a generalization of (*precise*) probability
- "There's more to uncertainty than probability"[3]
- Imprecise probability is meant to capture those aspects of uncertainty that (precise) probability doesn't
- Focus is on a higher-level uncertainty, i.e., *uncertainty about how to assign probabilities*
- Not a new idea...

[3]SIPTA mantra: `https://sipta.org/`

Intro, cont.

- This is exactly the situation statisticians face!
- We're aware of this — we love to say

  All models are wrong, but some are useful

- We don't take it seriously, however:
  - we focus on the convenient "but..." excuse
  - skipping over the challenges associated with accounting for this higher-level uncertainty
- Isn't it possible that ignoring some uncertainty is at least partially responsible for the *replication crisis*[4] in science?
- If so, then we should consider ways we might do better

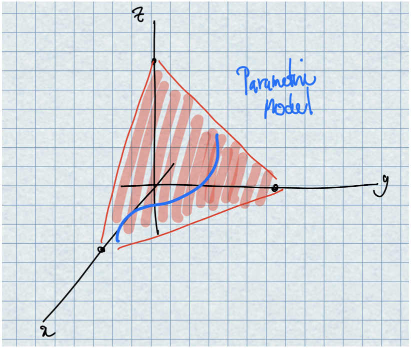
[4] https://en.wikipedia.org/wiki/Replication_crisis

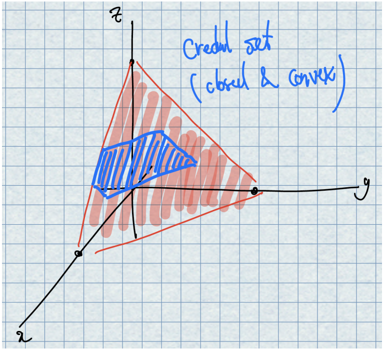
Intro, cont.

- Technically, how to describe higher-level uncertainty?
- It's not just a hierarchical model!
  - marginalization still gives a probability
  - e.g., coin with a uniform prior $\iff$ coin is fair
- Consider a simple universe with three states $\{x_1, x_2, x_3\}$
- Uncertain variable X takes one of these three values
- All possible probability models for X corresponds to the *probability simplex*

$$\mathscr{P} = \{(p_1, p_2, p_3) : p_k \geq 0 \text{ and } p_1 + p_2 + p_3 = 1\}$$

(lousy drawings of a) 3-dim probability simplex





Intro, cont.

- Points in $\mathscr{P}$ correspond to precise probability distn's for X
- e.g., $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ corresponds to "equally likely" outcomes
- Subsets of $\mathscr{P}$:
  - curve in $\mathscr{P}$ is a *parametric model*
  - closed and convex subsets of $\mathscr{P}$ are *credal sets*
- If all I can do is pin down a subset $\mathscr{P}_0$ of $\mathscr{P}$, then:
  - can't give a *precise* answer to $\mathsf{P}(X \in A) = ??$
  - my answer must be *imprecise* because I don't know which $\mathsf{P} \in \mathscr{P}_0$ to use to evaluate the probability
- Note: *imprecise* $\neq$ *inaccurate*

> *Whereof one cannot speak, thereon one must remain silent*
> —Ludwig Wittgenstein

Intro, cont.

- If all I can justify is that the distribution of X is in $\mathscr{P}_0$, then the best I can do is report, for each A,

$$\underline{\mathrm{P}}(A) = \inf_{\mathrm{P} \in \mathscr{P}_0} \mathrm{P}(A) \quad \text{and} \quad \overline{\mathrm{P}}(A) = \sup_{\mathrm{P} \in \mathscr{P}_0} \mathrm{P}(A)$$

- Note $\underline{\mathrm{P}}$ and $\overline{\mathrm{P}}$ are *not* probabilities, e.g., for $A \cap B = \varnothing$

$$\begin{aligned} \overline{\mathrm{P}}(A \cup B) &= \sup_{\mathrm{P} \in \mathscr{P}_0} \{\mathrm{P}(A) + \mathrm{P}(B)\} \\ &\leq \overline{\mathrm{P}}(A) + \overline{\mathrm{P}}(B) \quad \longleftarrow \text{sub-additive!} \end{aligned}$$

- The "bar" notation indicates lower and upper probabilities
- *Imprecise probability theory* focuses on the interpretation and properties/calculi of lower & upper-probabilities

Imprecise prob & statistics

- Not as unfamiliar/challenging as you might think!
- I'll demonstrate that classical statistical theory has a number of oft-unstated connections to imprecise prob
- Statistical problem setup:
  - observable data X in $\mathbb{X} = \mathbb{R}$ or $\mathbb{R}^q$
  - statistical model $X \sim \mathsf{P}_\theta$, indexed by $\theta \in \mathbb{T}$
  - true θ exists but is unknown[5]
- Goal is, roughly, to learn about the unknown θ based on the observation $X = x$

[5] Later we'll be more clear about what "unknown" means

- A familiar object is a *p-value*
- For now, focus on the simple normal mean case, where $X \sim \mathsf{P}_\theta = \mathsf{N}(\theta, \sigma^2)$, and $\sigma > 0$ is known
- Then the p-value function[6] based on $X = x$ is given by

$$\pi_x(\vartheta) = 2\{1 - \Phi(\sigma^{-1}|x - \vartheta|)\}, \quad \vartheta \in \mathbb{R},$$

  where Φ is the $\mathsf{N}(0,1)$ distribution function
- Textbooks stress that the p-value is not a probability for θ, but don't say what it is — maybe an imprecise probability?

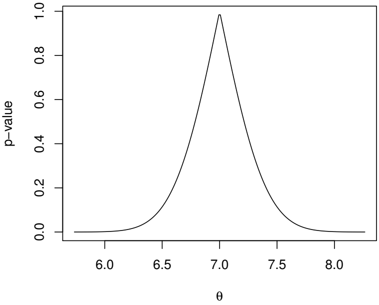
[6] This is for testing $H_0 : \theta = \vartheta$ against a two-sided alternative, but I don't explicitly refer to the alternative

Imprecise prob & stat, cont.

- P-value isn't a prob density, so can't integrate
- Textbooks recommend getting general p-values by optimization:

$$\overline{\Pi}_x(A) = \sup_{\vartheta \in A} \pi_x(\vartheta)$$

- Weird... why optimization?
- Clearly, $\overline{\Pi}_x$ isn't a probability, so what is it??



$\pi_x(\cdot)$ for $x = 7$ and $\sigma = 1$

Imprecise prob & stat, cont.

- $\overline{\Pi}_x$ takes the mathematical form of a *possibility measure*, a special type of imprecise probability
- In particular, optimization is to possibility theory what integration is to probability theory
- That theory interprets the upper probability $\overline{\Pi}_x(A)$ as a measure of how *plausible* hypothesis A is, given inputs $(x, \ldots)$
- Exactly how we interpret and use p-values in practice!
- All the misunderstandings about p-values disappear once we understand what they really are[7]
- More about this later...

[7] Student: "I wanted to let you know that your discussion of plausibility in ST503 made it easier for me to explain what a p-value is when I was interviewing. Thank you for the new perspective on the subject!"

Imprecise prob & stat, cont.

- Let's push this a bit further
- If $\overline{\Pi}_x$ is an imprecise probability, then it has a collection of probabilities that are *consistent* with it
- Define the associated credal set

$$\mathscr{C}(\overline{\Pi}_x) = \{\Pi_x : \Pi_x(A) \leq \overline{\Pi}_x(A) \text{ for all } A\}$$

- Define the $100(1-\alpha)\%$ confidence interval

$$C_\alpha(x) = \{\vartheta : \pi_x(\vartheta) > \alpha\}, \quad \alpha \in [0,1]$$

Imprecise prob & stat, cont.

- It can be shown that

$$\Pi_x\{C_\alpha(x)\} \geq 1 - \alpha \quad \forall\, \Pi_x \in \mathscr{C}(\overline{\Pi}_x)$$

- That is, every probability distribution in $\mathscr{C}(\overline{\Pi}_x)$ assigns at least probability $1 - \alpha$ to the $100(1 - \alpha)\%$ confidence int
- A reasonable *probabilistic approximation* to $\overline{\Pi}_x$ might be the minimally-specific/maximally diffuse element in $\mathscr{C}(\overline{\Pi}_x)$
- It can be shown that this minimally specific $\Pi_x^\star$ is

$$\Pi_x^\star = \mathsf{N}(x, \sigma^2)$$

which is Fisher's fiducial distribution (also flat-prior Bayes)

Imprecise prob & stat, cont.

- Fisher's fiducial argument was really clever, but there were technical issues he couldn't overcome
- Commonly (and unfairly) labeled *Fisher's biggest blunder*
- Unfair? Much of what statisticians consider "fundamental" can be traced back to Fisher's fiducial efforts, e.g.,
  - *confidence intervals* (Neyman)
  - *shrinkage estimation* (Stein)
- Also inspired the work of Art Dempster and Glenn Shafer, early leaders in the imprecise prob developments

> *There is still more to be learned from Fisher, even when he seems clearly wrong, than from any other contributor to statistical thinking* —Phil Dawid

Imprecise prob & stat, cont.

- Ironically, what Fisher lacked was imprecise prob theory
- There are hints in Fisher's writing, e.g.,
  - → *Valid tests of significance at all levels may exist without the possibility of deducing by an accurate argument, a probability distribution for the unknown parameter*
  - → *It is evidently easier for the practitioner of natural science to recognize the difference between knowing and not knowing than this seems to be for the more abstract mathematician*
- We have what's needed to "fix" Fisher's theory, and more...[8]

> *Perhaps the most important unresolved problem in statistical inference is the use of Bayes theorem in the absence of prior information* —Brad Efron

[8] e.g., `https://researchers.one/articles/21.01.00002`

Course outline, part 1

- Introduction to some basic imprecise prob models:
  - random sets
  - possibility measures
  - belief functions
  - lower previsions
- Go back through the list above and compare based on some specific criteria, e.g.,
  - conditioning/updating
  - combining information
  - ...
- Some more advanced/specific details, e.g., approximating one kind of imprecise prob by a simpler one

Course outline, part 2

- Statistical inference:
  - possibility measures (RM)
  - belief functions (Dempster, Shafer, Denoeux)
  - others (e.g., Walley)
  - comparison
- Methods for specific problems:
  - clustering
  - classification
  - prediction
  - decision-making
- Applications.....

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!

Next lecture

- Review of (precise) probability theory
- Shortcomings and the need for more flexibility
- Mathematical formulation of imprecise probability
- ...