ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

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

Week 04a

This lecture

- Possibility theory recap
- (Imprecise-)probability-to-possibility transform
- Extension principle

Recap

- A function $\overline{\Pi} : 2^{\mathbb{X}} \to [0,1]$ is a *possibility measure* if
  - $\overline{\Pi}(\varnothing) = 0$
  - $\overline{\Pi}(\mathbb{X}) = 1$
  - it's maxitive,[1] i.e., $\overline{\Pi}(\bigcup_{n=1}^{\infty} A_n) = \sup_n \overline{\Pi}(A_n)$
- There exists a function $\pi : \mathbb{X} \to [0,1]$, called the *possibility contour*, such that $\sup_{x \in \mathbb{X}} \pi(x) = 1$ and

$$\overline{\Pi}(A) = \sup_{x \in A} \pi(x), \quad A \subseteq \mathbb{X}$$

- The dual, $\underline{\Pi}$, is a *necessity measure* and satisfies

$$\underline{\Pi}(A) = 1 - \overline{\Pi}(A^c) = 1 - \sup_{x \in A^c} \pi(x), \quad A \subseteq \mathbb{X}$$

[1] *Update:* For non-finite $\mathbb{X}$, I need to explicitly assume countable maxitivity in order to get the contour representation in 2nd bullet point

Recap, cont.

- $\overline{\Pi}$ determines a credal set $\mathscr{C}(\overline{\Pi}) = \{\mathsf{P} : \mathsf{P} \leq \overline{\Pi}\}$, containing all the probabilities *consistent* with it
- Characterization in terms of the sub-level sets of π
- Consequences:
  - $\mathscr{C}(\overline{\Pi}) \neq \varnothing \implies$ so no-sure-loss
  - $\overline{\Pi}(\cdot) = \sup_{\mathsf{P} \in \mathscr{C}(\overline{\Pi})} \mathsf{P}(\cdot) \implies$ coherent[2]

[2] *An aside:* There's a risk of incoherence, e.g., if we're not careful when "combining" distinct possibility measures about a common uncertain quantity, and we'll talk more about this later

Coherence[3]

Proof.

To show: *for each A, there exists* $\mathrm{P} \in \mathscr{C}(\overline{\Pi})$ *such that* $\mathrm{P}(A) = \overline{\Pi}(A)$.

Suppose $\overline{\Pi}(A) < 1$. Choose $(x_n^{\in}) \subset A$ and $(x_n^{\notin}) \subset A^c$ such that

$$\pi(x_n^{\in}) \to \overline{\Pi}(A) \quad \text{and} \quad \pi(x_n^{\notin}) \to \overline{\Pi}(A^c).$$

Define the sequences

$$\begin{aligned} g_n^{\in} &= \overline{\Pi}(\{x_1^{\in}, \dots, x_n^{\in}\}), \quad g_0 = 0 \\ g_n^{\notin} &= \overline{\Pi}(A \cup \{x_1^{\notin}, \dots, x_n^{\notin}\}), \quad g_0^{\notin} = \overline{\Pi}(A). \end{aligned}$$

Define a discrete probability P, with masses only on the above points, as

$$p(x_n^{\in}) = g_n^{\in} - g_{n-1}^{\in} \quad \text{and} \quad p(x_n^{\notin}) = g_n^{\notin} - g_{n-1}^{\notin}.$$

Then P is a probability, it's in $\mathscr{C}(\overline{\Pi})$, and $\mathrm{P}(A) = \overline{\Pi}(A)$....... $\square$

[3] From Hose's thesis, attributed to Fetz & Oberguggenberger (2004)

Probability-to-possibility transform

- Suppose we have a probability P on $\mathbb{X}$
- Goal: approximate[4] P by a possibility measure $\overline{\Pi} \geq \mathsf{P}$
- A complement to the question we considered before:
  - instead of asking which P are compatible with $\overline{\Pi}$
  - we're asking which $\mathscr{C}(\overline{\Pi})$ contains a given P?
- The *probability-to-possibility transform* says how to do this
- Of course, there's not a unique $\overline{\Pi} \geq \mathsf{P}$...

[4] The same approximation strategy applies when P is replaced by a general imprecise probability $\overline{\mathsf{P}}$, and we'll consider this later

Prob-to-poss transform, cont.

- Let $f : \mathbb{X} \to \mathbb{R}$ be a (measurable) function
- For the given P, define a contour function

$$\pi_f(x) = \mathsf{P}\{f(X) \leq f(x)\}, \quad x \in \mathbb{X}$$

- Define the corresponding possibility measure

$$\begin{aligned} \overline{\Pi}_f(A) &= \sup_{x \in A} \pi_f(x) \\ &= \mathsf{P}\Big\{f(X) \leq \sup_{x \in A} f(x)\Big\}, \quad A \subseteq \mathbb{X} \end{aligned}$$

Theorem.

For any f as above, $\overline{\Pi}_f \geq \mathsf{P}$ or, in other words, $\mathscr{C}(\overline{\Pi}_f) \ni \mathsf{P}$

Prob-to-poss transform, cont.

Proof.

At least two different proofs:

- For any $A \subseteq \mathbb{X}$, it's clear that

  $$E := \{y : f(y) \leq \sup_A f\} \supseteq A$$

  So, $\mathrm{P}(A) \leq \mathrm{P}(E) = \overline{\Pi}(A)$, hence $\mathrm{P} \leq \overline{\Pi}_f$
- Define the sub-level set of π_f:

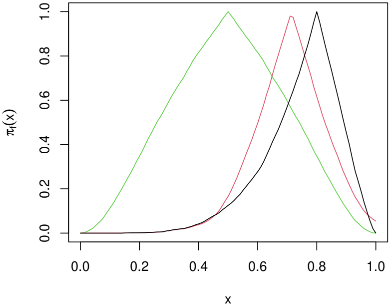
  $$S_\alpha^c = \{x : \mathrm{P}[f(X) \leq f(x)] \leq \alpha\}$$

  The right-hand side consists of those x such that $f(x)$ is no larger than the lower-α quantile of $f(X)$. That event has probability $\leq \alpha$, so $\mathrm{P}(S_\alpha^c) \leq \alpha$ and $\mathrm{P} \in \mathscr{C}(\overline{\Pi}_f)$

□

Example

- P = Beta(a, b) with $a = 5$ and $b = 2$
- Induced π based on three different f
  - $f(x) = x(1 - x)$
  - $f(x) = -|x - \frac{a}{a+b}|$
  - $f(x) = \text{dbeta}(x, a, b)$



Example, cont.

- FYI, I'm doing this via Monte Carlo
- R code I used is below

```r
# Probability-to-possibility transform

a <- 5; b <- 2
m <- a / (a + b)
N <- 5000
X <- rbeta(N, a, b)
f1 <- function(x) x * (1 - x)
f2 <- function(x) -abs(x - m)
f3 <- function(x) dbeta(x, a, b)
p1 <- function(x) sapply(x, function(y) mean(f1(X) <= f1(y)))
p2 <- function(x) sapply(x, function(y) mean(f2(X) <= f2(y)))
p3 <- function(x) sapply(x, function(y) mean(f3(X) <= f3(y)))
curve(p1, xlim=c(0,1), ylab=expression(pi[f](x)), col=3)
curve(p2, add=TRUE, col=2)
curve(p3, add=TRUE)
```

Prob-to-poss transform, cont.

- Natural question: *which f is best?*
- The so-called *specificity principle* suggests that a "best" $f^\star$, if it exists, would be such that

$$\pi_{f^\star}(x) \leq \pi_f(x) \quad \text{for all } x \text{ and all } f$$

- Too few constraints to solve this...
- Suppose we have a *subjective plausibility order*
  - determined by a function r,
  - i.e., x is no less plausible than y iff $r(x) \leq r(y)$
- Then the *most specific* approximation of P by possibility measure (relative to the order r) is with $f^\star = r$.

Prob-to-poss transform, cont.

- A bit frustrating: we're left with the choice of ordering r
- Where does this come from?
- Good default approach, when P has density p:
  - Recall: probable $\implies$ possible
  - Makes sense for $\pi_f(x) > \pi_f(y)$ iff $p(x) > p(y)$
  - So, take $f^\star = r = p$
- This is what's shown in the black line in above plot
- Reminiscent of some ideas in statistics...

Statistics example

- $X \sim \mathsf{N}_q(\theta, I)$
- Default posterior: $\theta \sim \mathsf{N}(x, I)$
- Possibilistic approximation of the posterior

$$\pi_f(\vartheta) = \mathsf{P}_{\theta|x}\{f(\theta) \leq f(\vartheta)\}$$

- Above strategy suggests taking $f^\star = p_x$, post density, so

$$\begin{aligned} \pi_{f^\star}(\vartheta) &= \mathsf{P}_{\theta|x}\{p_x(\theta) \leq p_x(\vartheta)\} \\ &= \cdots \\ &= \mathsf{P}\{\mathsf{ChiSq}(q) \geq \|\vartheta - x\|^2\} \\ &= \text{p-value of the LR test of } H_0 : \theta = \vartheta \end{aligned}$$

Extension

- This is a simple but important idea
- Suppose $\overline{\Pi}_X$ is a possibility measure on $\mathbb{X}$, contour π_X
- This quantifies uncertainty about an "uncertain variable" X
- Suppose X is related to Y in $\mathbb{Y}$ via $c(X, Y) = 0$
- How to quantify uncertainty about Y?
- *Extension principle* says

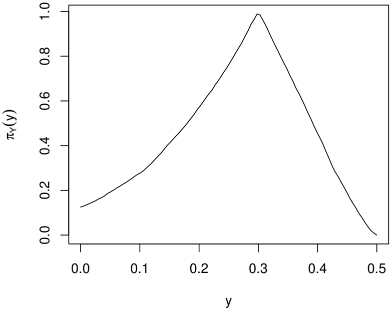
$$\overline{\Pi}_Y(B) = \sup_{y \in B} \pi_Y(y), \quad B \subseteq \mathbb{Y}$$

where

$$\pi_Y(y) = \sup_{x : c(x,y)=0} \pi_X(x), \quad y \in \mathbb{Y}$$

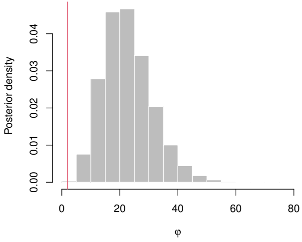
Example, cont.

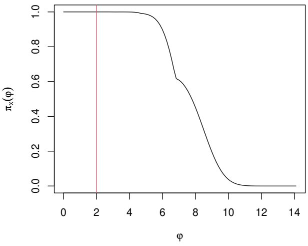
- X is the uncertain quantity, π_X from before
- e.g., take $Y = |X - \frac{1}{2}|$
- To evaluate $\pi_Y(y)$ for a given y
  - there's a pair $x = \frac{1}{2} \pm y$
  - maximize $\pi_X(x)$ over these two values



Statistics example, cont.

- Stein's example: $X \sim \mathsf{N}_q(\theta, I)$, interest in $\phi = \|\theta\|^2$
- Two solutions:
  - marginal Bayes (left), non-central chi-square
  - marginal possibility (right), via extension principle
- $q = 10$, *very naive* computation[5] of $\pi_x(\varphi)$
- Bayes misses true ϕ, other doesn't





[5]Semi-reliable: $\pi_x(\phi) \approx 0.99$

Remarks

- Interpretation of a possibility measure?
- At least two options:
  - interpret $\overline{\Pi}$ directly
  - interpret $\overline{\Pi}$ indirectly via $\mathscr{C}(\overline{\Pi})$
- Shackle had in mind the former
- Latter might be more natural, e.g.,
  - if $\overline{\Pi}(A)$ is small
  - then $\mathrm{P}(A)$ is small too for all compatible P's
- I don't know which is better/easier/etc.

Next lecture

- Belief functions
- Relations to random sets and possibility measures
- Properties
- ...