ST790 — Fall 2022

Imprecise-Probabilistic Foundations of Statistics

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

Week 09b

This lecture

- Towards a valid & efficient IM
- Possibility measures and *strong validity*
- Imprecise-probability-to-possibility transform
- ...

- Presentation will be based on ideas in a recent paper[1] and some new stuff I'm working on now

[1] http://arxiv.org/abs/2203.06703

Introduction

- Recall:
  - uncertain $(Y, \Theta) \in \mathbb{Y} \times \mathbb{T}$, imprecise prob $\overline{\mathsf{P}}_{Y,\Theta}$
  - data $Y = y$ is observed, goal is inference on Θ
  - an *inferential model* (IM) is a mapping $(y, \ldots) \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$, output is an imprecise prob on $\mathbb{T}$
  - an IM is *valid* (wrt $\overline{\mathsf{P}}_{Y,\Theta}$) if

$$\overline{\mathsf{P}}_{Y,\Theta}\{\overline{\Pi}_Y(A) \le \alpha, \, \Theta \in A\} \le \alpha, \quad (\alpha, A) \in [0,1] \times 2^{\mathbb{T}}$$

> *It is unacceptable if a procedure... of representing uncertain knowledge would, if used repeatedly, give systematically misleading conclusions* (Reid & Cox 2015)

Intro, cont.

- Dempster's IM isn't valid
- Generalized Bayes is valid but inefficient
- What else can we do?
- ...
- Options:
  - modify Dempster's approach to force validity
  - leverage connections between possibility theory & validity

First valid IMs

- Recall, Dempster's IM can incur sure-loss due to *contraction*, validity fails for the same reason
- Suggests Dempster's random sets, $\mathcal{T}_y(U)$, are "too small"
- How to enlarge them?
- Idea:[2] replace $U \sim \mathsf{P}_U$ with a random set $\mathcal{U}$,

$$\mathcal{T}_y(\mathcal{U}) = \bigcup_{u \in \mathcal{U}} \underbrace{\{\theta \in \mathbb{T} : y = a(\theta, u)\}}_{\mathcal{T}_y(u)}$$

- Union implies "$\mathcal{T}_y(\mathcal{U}) \supseteq \mathcal{T}_y(U)$" in a stochastic sense
- Need some conditions on $\mathcal{U}$...

[2] M. and Liu (2013, `arXiv:1206.4091`) and the book

First valid IM, cont.

Theorem (M. and Liu 2013).

Associate $Y = a(\Theta, U)$, with $U \sim \mathsf{P}_U$; vacuous prior case.

Define $f(u) = \mathsf{P}_{\mathcal{U}}(\mathcal{U} \ni u)$, a feature of $\mathcal{U} \sim \mathsf{P}_{\mathcal{U}}$. If

- $f(U) \geq_{\text{st}} \mathsf{Unif}(0,1)$ when $U \sim \mathsf{P}_U$,
- $\mathcal{T}_y(\mathcal{U})$ is non-empty for almost all y,

then the IM with $\overline{\Pi}_y(A) = \mathsf{P}_{\mathcal{U}}\{\mathcal{T}_y(\mathcal{U}) \cap A \neq \varnothing\}$ is valid.

- Condition on $\mathcal{U}$ is relatively mild, in particular, doesn't depend on special structure in the model
- Some recent[3] work replaced $\mathcal{U}$ with a possibility measure that dominates P_U

[3] https://researchers.one/articles/20.08.00004

Example

- $Y \sim \text{Bin}(n, \theta)$, vacuous prior
- $F_{n,\theta}(Y-1) \le 1 - U < F_{n,\theta}(U)$, with $U \sim \mathsf{P}_U = \text{Unif}(0,1)$
- How to choose $\mathcal{U}$ to satisfy above condition?
  - set[4] $h(u) = |u - \frac{1}{2}|$
  - and define $\mathcal{U} = \{u : h(u) \le h(U)\}$
- Then[5] the y-dependent random set for Θ is

$$
\begin{aligned}
\mathcal{T}_y(\mathcal{U}) &= \bigcup_{u \in \mathcal{U}} \{\theta : F_{n,\theta}(y-1) \le 1 - u < F_{n,\theta}(y)\} \\
&= \bigcup_{u \in \mathcal{U}} \{\theta : 1 - G_{y,n-y+1}(\theta) \le 1 - u < 1 - G_{y+1,n-y}(\theta)\} \\
&= \big(\underline{\vartheta}_y(U), \overline{\vartheta}_y(U)\big]
\end{aligned}
$$

[4] Same result holds with virtually any other h

[5] Let $G_{a,b}$ denote the Beta(a,b) CDF

Example, cont.

- Where

$$\underline{\vartheta}_y(U) = G^{-1}_{y+1,n-y}(\tfrac{1}{2} - |U - \tfrac{1}{2}|)$$
$$\overline{\vartheta}_y(U) = G^{-1}_{y,n-y+1}(\tfrac{1}{2} + |U - \tfrac{1}{2}|)$$

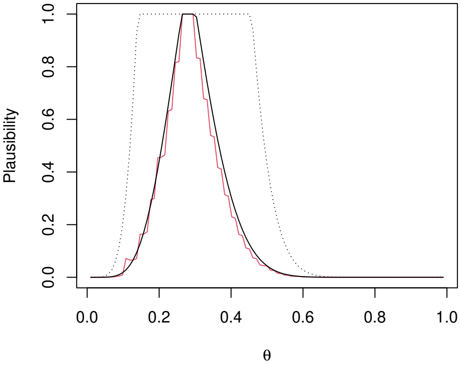
- IM's contour function

$$\begin{aligned}
\pi_y(\theta) &= \mathsf{P}_{\mathcal{U}}\{\mathcal{T}_y(\mathcal{U}) \ni \theta\} \\
&= 1 - \mathsf{P}_U\{\underline{\vartheta}_y(U) > \theta\} - \mathsf{P}_U\{\overline{\vartheta}_y(U) < \theta\} \\
&= \cdots \\
&= 1 - \max\{0, 2G_{y+1,n-1}(\theta) - 1\} - \max\{0, 1 - 2G_{y,n-y+1}(\theta)\}
\end{aligned}$$

- Plot on the next page...

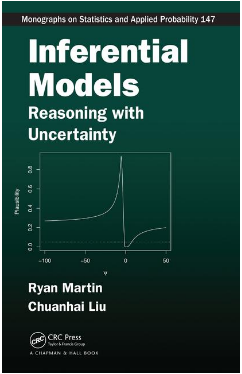
Example, cont.

- Data: $Y = 7$, $n = 25$
- Compare Walley (dotted), IM above, and new IM
- IM above is valid, more efficient than Walley, but slightly less efficient than the new IM



First valid IM, cont.

- Above result was the basis for the theory developed in the book
- I'm biased, but I think this is nice and interesting work
- Not perfect though:
  - requires "$Y = a(\Theta, U)$"
  - requires user to specify $\mathcal{U}$
  - too complicated
  - doesn't handle partial priors
- Need to modernize the idea



Strong validity

- Validity aims at making probabilistic reasoning reliable
- In the vacuous-prior settings, this condition was also sufficient for constructing IM-based confidence regions[6]
- With genuine partial priors, validity isn't strong enough

Definition.

Let $(y, \ldots) \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$ be an IM, and define the contour $\pi_y(\theta) = \overline{\Pi}_y(\{\theta\})$. This IM is *strongly valid* (wrt $\overline{\mathsf{P}}_{Y,\Theta}$) if

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

[6] e.g., Theorem 1 in `https://researchers.one/articles/21.01.00002`

Strong validity, cont.

- Similar to Walley (2002)'s *fundamental frequentist principle*
- Makes the construction of IM-based confidence regions easy

$$C_\alpha(y) = \{\theta \in \mathbb{T} : \pi_y(\theta) > \alpha\}, \quad \alpha \in [0,1]$$

- Implications (more details later):
  - some coherence-like properties
  - more-or-less forces the IM to be consonant
- Why stronger? Equivalent definition is

$$\overline{\mathsf{P}}_{Y,\Theta}\{\overline{\Pi}_Y(A) \leq \alpha, \underbrace{\text{for some } A \ni \Theta}_{\text{uniformity in } A}\} \leq \alpha, \quad \alpha \in [0,1]$$

- How to achieve strong validity?

New approach

- Two key ingredients:[7]
  - outer consonant approximations
  - imprecise-probability-to-possibility transform
- Basic general facts:
  - Let $\overline{\mathsf{P}}$ be an upper prob for X and define
    - $\pi(x) = \overline{\mathsf{P}}\{h(X) \leq h(x)\}$
    - $\overline{\Pi}(A) = \sup_{x \in A} \pi(x)$
  - then $\overline{\Pi}$ is an *outer consonant approximation* of $\overline{\mathsf{P}}$
  - that is, $\overline{\Pi}$ is consonant and $\mathscr{C}(\overline{\mathsf{P}}) \subseteq \mathscr{C}(\overline{\Pi})$
  - which implies $\overline{\mathsf{P}}\{\pi(X) \leq \alpha\} \leq \alpha$
- Use these ideas to construct a strongly valid IM

[7] See the Week 04a lecture material

New approach, cont.

- The explanation of *how* to get from the above ideas/facts to my specific construction is lengthy[8]
- I'll spare you that explanation and jump to the conclusion
- Easiest for the precise-model-partial-prior case

- Notation:
  - $p_\theta(y)$ is the density/mass function of the (precise) model
  - $q(\theta) = \overline{\mathsf{P}}_\Theta(\{\theta\})$ is the (imprecise) prior contour
  - the *relative likelihood* is

$$\eta(y, \theta) = \frac{p_\theta(y)\, q(\theta)}{\sup_{t \in \mathbb{T}} p_t(y)\, q(t)}, \quad \theta \in \mathbb{T}$$

  - kind of like Bayes's formula...

[8]Very much influenced by the fundamental principles described in Section 2.3.2.1 in Hose's PhD thesis

New approach, cont.

- Recall that $\overline{\mathsf{P}}_{Y,\Theta}$ is known
- Define a (consonant) IM with contour function[9]

$$\pi_y(\theta) = \overline{\mathsf{P}}_{Y,\Theta}\{\eta(Y,\Theta) \leq \eta(y,\theta)\}, \quad \theta \in \mathbb{T}$$

- Consonance means the upper prob is defined via optimization

$$\overline{\Pi}_y(A) = \sup_{\theta \in A} \pi_y(\theta), \quad A \subseteq 2^{\mathbb{T}}$$

- Can marginalize via extension principle...
- Outer consonant approximation $\implies$ strong validity

Theorem.

The IM above is strongly valid wrt $\overline{\mathsf{P}}_{Y,\Theta}$.

[9] just an imprecise-probability-to-possibility transform

New approach, cont.

- Generalized Bayes is valid but not strongly valid
- Other IMs are strongly valid, but tend to be inefficient[10]
- The only IM that achieves strong validity and efficiency?
- Why do I say "efficient"?
  - naive[11] construction in vacuous prior case gives contour

$$\pi_y(\theta) = \sup_{\vartheta} \mathsf{P}_{Y|\vartheta}\Big\{ \frac{p_\vartheta(Y)}{\sup_t p_t(Y)} \le \frac{p_\theta(y)}{\sup_t p_t(y)} \Big\}$$

  - relative likelihood is an approx pivot, so $\sup_\vartheta$ drops out
  - RHS $\approx$ the p-value for the optimal (?) likelihood ratio test

[10] e.g., Sec. 4.3 in `https://researchers.one/articles/21.05.00001`

[11] "Naive" because there are principles that can/should be applied to improve efficiency — more on this later

New approach, cont.

- What's important is how this deals with partial prior info
- Recall:

$$\pi_y(\theta) = \overline{\mathsf{P}}_{Y,\Theta}\{\eta(Y,\Theta) \le \eta(y,\theta)\}, \quad \theta \in \mathbb{T}$$

  with $\eta(y,\theta) = p_\theta(y)\, q(\theta) / \sup_t\{\cdots\}$
- Then partial prior info enters in two ways:
  - through the definition of η (depends on q)
  - through the model $\overline{\mathsf{P}}_{Y,\Theta}$
- This partial prior dependence creates opportunity for efficiency gain compared to vacuous-prior case
- Important: incorporating partial prior doesn't ruin validity!

Computation

- How to compute the $\overline{\mathsf{P}}_{Y,\Theta}$-probability that drives π_y?
- If $\overline{\mathsf{P}}_\Theta$ is consonant, then the Choquet integral simplifies:

$$\pi_y(\theta) = \int_0^1 \Big[\sup_{\vartheta : q(\vartheta) > \beta} \mathsf{P}_{Y|\vartheta}\{\eta(Y, \vartheta) \leq \eta(y, \theta)\} \Big] \, d\beta$$

- Suggests a simple/naive Monte Carlo strategy:

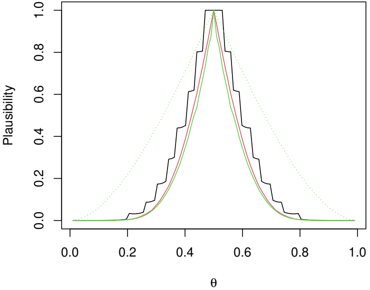
$$\pi_y(\theta) \approx \int_0^1 \max_{s : q(\vartheta_s) > \beta} \Big[\frac{1}{M} \sum_{m=1}^{M} 1\{\eta(Y_s^{(m)}, \vartheta_s) \leq \eta(y, \theta)\} \Big] \, d\beta$$

where

- $\vartheta_1, \ldots, \vartheta_S$ is a fixed, dense grid
- $Y_s^{(1)}, \ldots, Y_s^{(M)} \overset{\text{iid}}{\sim} \mathsf{P}_{Y|\vartheta_s}$, for $s = 1, \ldots, S$

Example

- $\mathsf{P}_{Y|\theta} = \mathsf{Bin}(n, \theta)$
- Vacuous, complete, and partial prior info[12]
- Partial prior info is *more* influential than complete



[12] There are dimension-reduction steps taken that I'm not explaining...

Summary

- New construction of a strongly valid IM
- Advantage is that it's based solely on the *posited model*
  - no choice of association
  - no choice of a random set ($\mathcal{U}$)
- Partial prior is like regularization
  - allows for efficiency gains
  - without sacrificing validity
- Wanna see how this works in high-dim cases
- Need to scale up the computations

Next lecture

- Coherence-like properties
- Efficiency considerations
  - basic dimension-reduction
  - marginalization
- ...