ST790 — Fall 2022

Imprecise-Probabilistic Foundations of Statistics

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

Week 09a

This lecture

- Generalized Bayes
- Properties — good & not-so-good
- Walley's formulation for vacuous priors
- IM efficiency considerations

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

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

  - validity $\implies$ "probabilistic reasoning" is simple & reliable
  - important to emphasize the *wrt...* part
- Last time: Dempster's construction of an IM
- Has some nice/interesting features, but not valid
- Need to keep looking for a valid IM

Intro, cont.

- Fortunately, other constructions are available
- One of those is *generalized Bayes*
- Walley's *Statistical Reasoning with Imprecise Probabilities* is a book about exactly this[1]
- Basic idea: Bayes with a bunch of priors simultaneously
- Statistically relevant points to discuss:
  - coherent
  - *valid*
  - tends to be *inefficient*

[1] Walley talks about *lower previsions*, I'll talk about lower/upper probs...

Generalized Bayes

- For simplicity, consider $(Y \mid \Theta = \theta) \sim \mathrm{P}_{Y|\theta}$, precise
- Then $\overline{\mathrm{P}}_{Y,\Theta}$ is determined by an imprecise prior for Θ
  - a credal set $\mathscr{P}_\Theta$ of precise priors, P_Θ
  - an upper envelope $\overline{\mathrm{P}}_\Theta$
- Update $\overline{\mathrm{P}}_\Theta \to \overline{\mathrm{P}}_{\Theta|y}$ based on $Y = y$
- Generalized Bayes rule[2] says to update as

$$\overline{\mathrm{P}}_{\Theta|y}(\cdot) = \sup_{\mathrm{P}_\Theta \in \mathscr{P}} \underbrace{\mathrm{P}_{\Theta|y}(\cdot)}_{\text{Bayes w/ prior } \mathrm{P}_\Theta}$$

- Roughly: *Bayes rule applied to a set of priors*

[2] e.g., Chapters 2 & 7 in *Intro to IP*

Generalized Bayes, cont.

- Computation??? More on this below...
- Imprecise/partial prior?
  - classically for robustness, e.g., contamination nbhd's
  - modern focus on marginalization and regularization
- Difference between "robust Bayes" and generalized Bayes is mainly in perspective:
  - robust Bayes *wants* to use a single posterior
  - generalized Bayes will use the set of posteriors
- Does the prior matter?
  - In classical problems, effect of prior washes out asymptotically
  - But partial priors for regularization won't — and we don't want them to!
- What about robustness to the likelihood...?

Validity

Theorem — generalized Bayes IM is valid.

[a]The generalized Bayes IM, $\overline{\Pi}_y = \overline{\mathsf{P}}_{\Theta|y}$, is valid wrt $\overline{\mathsf{P}}_{Y,\Theta}$, i.e.,

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

[a]Corollary 3 in `arXiv:2203.06703`

- Implies generalized Bayes is (coherent and) reliable
- Special case:
  - the ordinary Bayes IM is valid wrt $\mathsf{P}_{Y,\Theta}$
  - validity wrt the Bayes model isn't surprising at all
- The wrt... part is important

Computation

- Computation?
- It's the sup over priors that makes this challenging
- For finite $\mathbb{T}$, the sup is over a closed & convex subset of a finite-dim simplex, not impossible
- For other $\mathbb{T}$, it's an infinite-dim optimization problem :(
- However, depending on the structure of the prior credal set, some analytical results are available...

- Recall the *contamination neighborhood* example:

$$\mathscr{P}_\Theta = \{(1-\varepsilon)\mathsf{P}^\star + \varepsilon \mathsf{Q} : \text{ any prob } \mathsf{Q} \text{ on } \mathbb{T}\}$$

- Here, $\varepsilon \in [0,1]$ and $\mathsf{P}^\star$ are fixed
- Closed & convex, so we can take this as a prior credal set and apply generalized Bayes to get $(\underline{\Pi}_y, \overline{\Pi}_y)$
- This prior has enough structure to show,[3] e.g., that

$$\underline{\Pi}_y(A) = \frac{(1-\varepsilon)\int_A L_y(\theta)\,\mathsf{P}^\star(d\theta)}{(1-\varepsilon)\int_{\mathbb{T}} L_y(\theta)\,\mathsf{P}^\star(d\theta) + \varepsilon \sup_{A^c} L_y}$$

$$\overline{\Pi}_y(A) = \frac{(1-\varepsilon)\int_A L_y(\theta)\,\mathsf{P}^\star(d\theta) + \varepsilon \sup_A L_y}{(1-\varepsilon)\int_{\mathbb{T}} L_y(\theta)\,\mathsf{P}^\star(d\theta) + \varepsilon \sup_A L_y}$$

[3] Wasserman (*Annals* 1990), Example 5.2

Remarks

- Key properties of generalized Bayes:
  - coherent
  - valid
- Computation is non-trivial, but can be done when the prior credal set has enough structure
- But recall that generalized Bayes tends to *dilate*
  - in particular, prior vacuous $\implies$ posterior is vacuous
  - take $\varepsilon = 1$ in above example
- A sign of potential *inefficiency*...

Special cases

- In statistics, generalized Bayes methods typically fall under the umbrella of "robust Bayes"
- Two special cases developed by Peter Walley:
  - *imprecise Dirichlet model* (IDM)
  - a no-name approach for vacuous priors
- The IDM is for multinomial models, a generalization of the conjugate Dirichlet priors
- Lengthy summary of IDM in Ch. 7 of *Intro to IP*
- I'll focus here on Walley's other idea

GBayes for vacuous priors

- Generalized Bayes proper is uselss for vacuous priors
- Walley (*JSPI* 2002) proposed an ingenious workaround
- Take the same $(\varepsilon, \mathsf{P}^\star)$-contamination neighborhood prior
- Let $\underline{\Pi}_y^\varepsilon$ and $\overline{\Pi}_y^\varepsilon$ denote the gBayes IM output above

Theorem (Walley 2002).

Fix $\bar\alpha \in (0,1)$. Then $y \mapsto (\underline{\Pi}_y^\varepsilon, \overline{\Pi}_y^\varepsilon)$ satisfies:

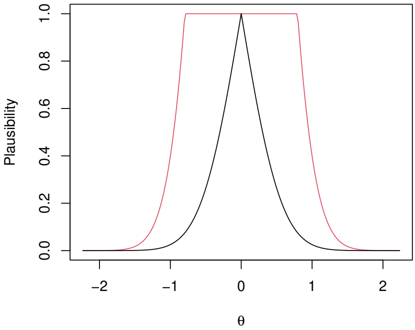
- coherence
- likelihood principle
- if $\varepsilon \geq (2-\bar\alpha)^{-1}$, then it's "valid" (wrt vacuous prior) on a restricted range $\alpha \in [0, \bar\alpha]$
- and more...

gBayes for vacuous priors, cont.

- Really impressive result!
- (Walley) A Bayesian & frequentist "reconciliation"
  - coherence follows from this being a gBayes solution
  - likelihood principle too (more later)
  - the "and more" says gBayes tests & confidence sets control error rates at levels $\alpha \in [0, \bar{\alpha}]$
- Doesn't get validity exactly, but that's not deal-breaker
  - taking $\bar{\alpha} = 0.5$, say, covers the usual significance levels
  - in that case, take $\varepsilon \geq 2/3$
- Visualization via plausibility contour in Walley's Eq. (3.3)
- Seems to do everything we want...

Examples

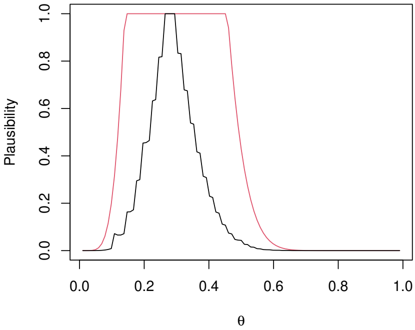
- Normal mean illustration: $(Y \mid \Theta = \theta) \sim \mathsf{N}(\theta, \sigma^2 n^{-1})$
- Contours for two (valid) IM solutions:
  - Walley's with $\varepsilon = \frac{10}{19}$, $\mathsf{P}^\star = \mathsf{N}(0, 3\sigma^2)$
  - mine I'll tell you about soon



$(y = 0,\ n = 5,\ \sigma = 1)$

Examples, cont.

- Binomial illustration: $(Y \mid \Theta = \theta) \sim \mathrm{Bin}(n, \theta)$
- Contours for two (valid) IM solutions:
  - Walley's with $\varepsilon = \frac{10}{19}$, $\mathsf{P}^\star = \mathrm{Unif}(0, 1)$
  - mine I'll tell you about soon



$(y = 7,\ n = 25)$

Remarks

- There's a price Walley pays for his "reconciliation"
- Significant[4] loss of efficiency compared to...
- Walley justifies this as a trade-off for satisfying the likelihood principle, e.g., stopping-rule invariance
- That's fine/nice, but I didn't ask for that
- To me, that's not worth sacrificing (a lot of) efficiency
- Fortunately, we can do better

[4] Even in the *rate*, e.g., $(\log n)^k n^{-1/2}$ compared to usual $n^{-1/2}$

Next 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[5] and some new stuff I'm working on now

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