ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

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

Week 07a

This lecture

- Compare Dempster & generalized Bayes
- Two general phenomena:
  - dilation
  - contraction
- Paradoxical?
- Perspectives (mine and others)

Introduction

- We've seen different rules for updating imprecise probs:
  - Dempster's rule
  - Generalized Bayes rule
- There are others, these are just the ones we've discussed
- It's worth asking: *how do they compare?*
- This leads to consideration of
  - dilation
  - contraction
- Presentation is based on Gong & Meng (*Stat Sci* 2021)[1]

[1] https://ruobingong.github.io/files/GongMeng2021_StatSci.pdf

Intro, cont.

- Setup:
  - prior is a belief function, has a credal set
  - can apply both Dempster & generalized Bayes updates
- Issues *not* present in precise probability:
  - have to choose a rule
  - can have counter-intuitive/paradoxical results
- If imprecise prob can give counter-intuitive results, why not just stick with precise prob since I know it "works"?
- Two key points:
  - precise prob "works" because of strong assumptions
  - our intuition is based on experiences with how precise prob behaves under unrealistically-ideal circumstances

Setup

- Universe $\mathbb{X}$, finite in the examples here
  - $A \subseteq \mathbb{X}$ is a generic event
  - $\mathcal{B} = \{B_1, B_2, \ldots\}$ a partition
- *Generalized Bayes rule:*
  - $\mathscr{P}$ is a closed convex set of probabilities on $\mathbb{X}$
  - update $\mathscr{P}$ based on B

$$\overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) = \sup_{\mathrm{P} \in \mathscr{P}} \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \quad B \in \mathcal{B}$$

- *Dempster's rule:*
  - $\mathscr{P}$ is the credal set corresponding to prior belief $\underline{\mathrm{P}}$
  - update $\underline{\mathrm{P}}$ based on B

$$\overline{\mathrm{P}}_{\mathfrak{D}}(A \mid B) = \frac{\sup_{\mathrm{P} \in \mathscr{P}} \mathrm{P}(A \cap B)}{\sup_{\mathrm{P} \in \mathscr{P}} \mathrm{P}(B)}, \quad B \in \mathcal{B}$$

Example: boxer, wrestler, and coin[3]

- Two stage experiment:
  - fair coin is flipped, $X = 1(\text{heads})$
  - boxer and wrestler fight, $Y = 1(\text{boxer wins})$
- What's "$\mathrm{P}(Y = 1 \mid X = Y)$"?
- Challenges:
  - don't know marginal distribution of Y
  - don't know the dependence between X and Y
- We could make *assumptions*[2] and compute the answer
- Imprecise probability allows us to avoid assumptions, but not without creating different challenges

---

[2] *Manksi's Law of Decreasing Credibility:* The credibility of inference decreases with the strength of the assumptions maintained

[3] Originally from Gelman (*Amer Stat* 2006)

Example, cont.

- $\mathbb{X} = \{0,1\} \times \{0,1\}$, head-tail & win-loss pairs
- Incomplete/partial prior info about coin:
  - encode as a belief function
  - two focal elements, $A_0 = \{0\} \times \{0,1\}$ and $A_1 = \{1\} \times \{0,1\}$
  - $m(A_0) = m(A_1) = 0.5$
- "Prior ignorance" concerning the fight:

$$\underline{\mathrm{P}}(Y = y) = 0 \quad \text{and} \quad \overline{\mathrm{P}}(Y = y) = 1, \quad y \in \{0,1\}$$

- Let's see what happens when we update for "$X = Y$"...

Example, cont.

- Two interesting observations...[4]
- Inference about the boxer:
  - Dempster's rule *contracts*
    - $\rightarrow$ $\underline{\mathrm{P}}_{\mathfrak{D}}(Y = 1 \mid X = Y) = \overline{\mathrm{P}}_{\mathfrak{D}}(Y = 1 \mid X = Y) = 0.5$
    - $\rightarrow$ $\underline{\mathrm{P}}_{\mathfrak{D}}(Y = 1 \mid X \neq Y) = \overline{\mathrm{P}}_{\mathfrak{D}}(Y = 1 \mid X \neq Y) = 0.5$
  - vacuous prior turns precise
- Inference about the coin:
  - Generalized Bayes rule *dilates*
    - $\rightarrow$ $\underline{\mathrm{P}}_{\mathfrak{B}}(X = 1 \mid X = Y) = 0$, $\overline{\mathrm{P}}_{\mathfrak{B}}(X = 1 \mid X = Y) = 1$
    - $\rightarrow$ $\underline{\mathrm{P}}_{\mathfrak{B}}(X = 1 \mid X \neq Y) = 0$, $\overline{\mathrm{P}}_{\mathfrak{B}}(X = 1 \mid X \neq Y) = 1$
  - precise prior turns vacuous
- Both rules give counter-intuitive/paradoxical results
- *Dilation* and *contraction* are general phenomena

[4] Similar results in other examples analyzed in Gong & Meng

Characterizations

Theorem — gBayes can't contract.

If $\underline{\mathrm{P}}$ and $\overline{\mathrm{P}}$ are the lower and upper envelopes of $\mathscr{P}$, then

$$\inf_{B \in \mathcal{B}} \underline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) \leq \underline{\mathrm{P}}(A) \quad \text{and} \quad \sup_{B \in \mathcal{B}} \overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) \geq \overline{\mathrm{P}}(A)$$

- The above conclusion is the property I called "joint coherence" in the previous lecture
- So, no surprise that generalized Bayes has this property
- Proof below
- In particular, gBayes updates avoid sure loss
- *No such guarantee for Dempster's rule*

Characterizations, cont.

Proof:[5] **gBayes can't contract.**

Assume not, i.e., there exists A such that $\sup_{B \in \mathcal{B}} \overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) < \overline{\mathrm{P}}(A)$. Credal set is closed so there exists P^A such that $\overline{\mathrm{P}}(A) = \mathrm{P}^A(A)$. Then:

$$\begin{aligned}
\overline{\mathrm{P}}(A) &= \mathrm{P}^A(A) \\
&= \sum_{B \in \mathcal{B}} \mathrm{P}^A(A \mid B)\, \mathrm{P}^A(B) \\
&\leq \sum_{B \in \mathcal{B}} \underbrace{\overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B)}_{< \overline{\mathrm{P}}(A)}\, \mathrm{P}^A(B) \\
&< \overline{\mathrm{P}}(A) \quad \leftarrow \text{contradiction!}
\end{aligned}$$

□

[5] Also sheds light on the "overfitting" aspect of gBayes

Characterizations, cont.

Theorem — gBayes dilates more.

For any A, B

$$\underline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) \leq \underline{\mathrm{P}}_{\mathfrak{D}}(A \mid B) \leq \overline{\mathrm{P}}_{\mathfrak{D}}(A \mid B) \leq \overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B)$$

In other words, if $\mathscr{C}(\mathfrak{B})$ and $\mathscr{C}(\mathfrak{D})$ are the credal sets of conditional probs based on the two rules, then $\mathscr{C}(\mathfrak{D}) \subseteq \mathscr{C}(\mathfrak{B})$

- *Proof:* $\overline{\mathrm{P}}_{\mathfrak{D}}(A \mid B) = \frac{\sup \mathrm{P}(A \cap B)}{\sup \mathrm{P}(B)} \leq \sup \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B)$
- Avoiding incoherence requires more conservatism
- "more conservatism" $=$ more imprecision $=$ larger credal set
- Conservatism is fine, even good, but not too much

Characterizations, cont.

Theorem — gBayes can't sharpen prior ignorance.

Let A be s.t. $\underline{\mathrm{P}}(A) = 0$ and $\overline{\mathrm{P}}(A) = 1$. If B has $\underline{\mathrm{P}}(B) > 0$, then

$$\underline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) = 0 \quad \text{and} \quad \overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) = 1$$

- *Proof:* Quick sketch...
  - $\overline{\mathrm{P}}(A) = 1 \implies \underline{\mathrm{P}}(A^c) = 0 \implies \underline{\mathrm{P}}(A^c \cap B) = 0$ for all B
  - if $\underline{\mathrm{P}}(B) > 0$, then $\exists\, \mathrm{P} \in \mathscr{C}(\underline{\mathrm{P}})$ with $\mathrm{P}(A^c \mid B) = 0$
  - then $\underline{\mathrm{P}}_{\mathfrak{B}}(A^c \mid B) = 0$, hence $\overline{\mathrm{P}}_{\mathfrak{B}}(A \mid B) = 1$
- This kind of makes sense, ignorance is strong
- But hard to believe that we can't learn from observations
- Suggests gBayes might be "too conservative"

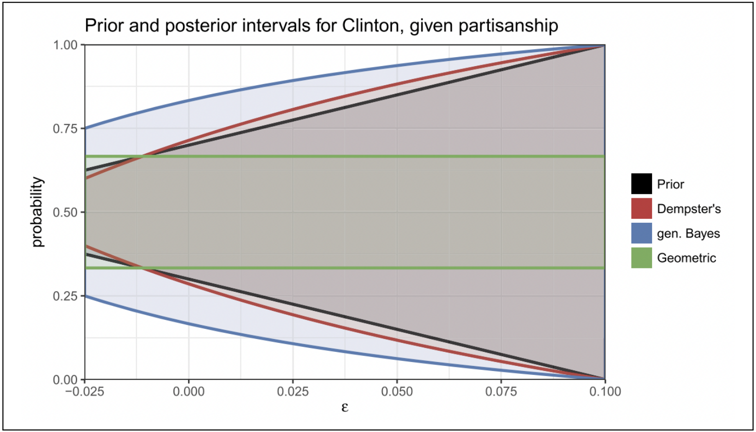
Election example

- Hypothetical *Trump vs. Clinton* election example
- Pre-election poll asks the following questions:
  - Q1 Do you intend to vote for Clinton or Trump?
  - Q2 Do you identify as Democrat or Republican?
- $\mathbb{X} = \{\mathsf{C}, \mathsf{T}\} \times \{\mathsf{Dem}, \mathsf{Rep}\}$
- Allow for non-response, i.e., set-valued responses
- Hypothetical data interpreted as a belief function
- Parameter ε controls "coarseness"

Q1	C	T	C	T	C	T	—	—	—
Q2	D	D	R	R	—	—	D	R	—
m	$\cdots\cdots$ $0.1 - \varepsilon$ $\cdots\cdots$								$0.2 + 8\varepsilon$

Election example, cont.

Plot shows $\underline{P}_\star(C \mid Q2)$ and $\overline{P}_\star(C \mid Q2)$ as a function of ε



Perspectives

- Gong & Meng: "the situation in the world of imprecise probability is more confusing and clearer at the same time"
  - it's right for our models to acknowledge what we don't know
  - imprecise prob gives us the flexibility to do so
  - but do we know how to operate the machine?
- Discussants from the IP community[6][7] argue that things are better understood than Gong & Meng suggest
- Roughly, Walley's theory of lower previsions settles it all
- ...except potential dilation

[6] Greg Wheeler and Thomas Augustin & Georg Schollmeyer

[7] Shafer's comments have a different focus, but are very insightful

Perspectives, cont.

- My take:
  - in stat/ML, data y is used to learn about θ
  - inference based on map $(y, \ldots) \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$
- Bayes, fiducial, Dempster, gBayes are examples
- *What properties do we want the map to satisfy?*
  - coherence
  - ...
  - validity
  - efficiency
- Latter two are related to reliability, important because now scientific inference is outsourced (e.g., R packages)
- If existing conditioning/updating rules fall short in terms of reliability, then let's come up with something new

Next lecture

- Today is the last lecture about imprecise probability
- Transitioning into "applications"
- Next time: some specifics in *statistical inference*
  - Dempster's approach
  - generalized Bayes (Walley and others)
  - .....
  - the "something new" I mentioned above