ST790 — Fall 2022

Imprecise-Probabilistic Foundations of Statistics

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

Week 10b

This lecture

- Wrap-up of new consonant IM developments
- Shift gears from inference to prediction
- Setup
- Model-based methods
- Non-model-based methods
  - nonparametric predictive inference
  - conformal prediction

IM recap

- Previous focus was on validity wrt vacuous priors
- This was achieved, roughly, by suitably enlarging the random sets in Dempster's framework
- Efficiency achieved by reducing the dimension of the "auxiliary variable" before carrying out the construction
- Other cool/novel developments...[1]
- So-called *generalized IMs* are a bit different:
    - dimension-reduction/efficiency built in[2]
    - a number of applications[3]
    - inspired some more recent developments (below)

---

[1] See the *conditional* and *marginal* IM papers

[2] `arXiv:1203.6665` and `arXiv:1511.06733`

[3] Cahoon and M., `arXiv:1910.00533` and `arXiv:1912.00037`

IM recap, cont.

- What's missing from these developments?
  - behavioral properties
  - incorporation of (partial) prior info
  - guidance about what construction is "best"
  - flexibility, computational & conceptual simplicity
  - ...
- New developments aim to fill these gaps
- In particular:
  - new results on (half-)coherence
  - construction and theory allows for partial prior info
  - no/fewer loose ends concerning *how* to do it
  - likelihood plays a central role
  - seems flexible, computationally doable

IM recap, cont.

- Technical novelties:
  - outer consonant approximation
  - imprecise-probability-to-possibility transform
- Other things I didn't tell you:
  - complete-prior case is interesting
  - applies to prediction & non- and semi-parametric cases
  - my focus on *consonant* structure is WLOG
- Next steps:
  - general computation, scaling up to higher dimensions
  - applications — literally everything is open!
  - model selection
  - decision theory
  - imprecision in the data/model
  - approximations (sometimes necessary...)

From inference to prediction

- In classical inference, there's a fixed unknown to learn
- Prediction is focused on a future observable
- Under a "probabilistic" framework, there's not much difference between the two problems
- Consider a simple case to start:
  - $(Y, \tilde{Y} \mid \Theta = \theta)$ iid from model $\mathsf{P}_{Y|\theta}$
  - goal is "inference" on $\tilde{Y}$, given $Y = y$
  - then Θ is a nuisance parameter, marginalize it out
- Lots of ways to handle this: Bayes, DS, gBayes, ...
- My IM construction can handle this too[4]

[4]Old-school IMs for prediction in `arXiv:1403.7589`

Prediction, cont.

- $(Y \mid \Theta = \theta) \sim \mathsf{N}(\theta, \sigma^2 n^{-1})$ and $(\tilde{Y} \mid \Theta = \theta) \sim \mathsf{N}(\theta, \sigma^2)$
- Bayesian approach:
  - flat prior for Θ
  - predictive distribution: $(\tilde{Y} \mid Y = y) \sim \mathsf{N}\big(y, \sigma^2(1 + n^{-1})\big)$
- My approach (with vacuous prior):
  - the “relative likelihood” is given by

$$\eta(y, \tilde{y}) = \frac{\sup_\theta p_\theta(y)\, p_\theta(\tilde{y})}{\sup_{x,\theta} p_\theta(y)\, p_\theta(x)} = \cdots = \exp\Big\{-\frac{(\tilde{y} - y)^2}{2\sigma^2(1 + n^{-1})}\Big\}$$

  - Since $\tilde{Y} - Y$ is a pivot, can easily get contour

$$\begin{aligned} \pi_y(\tilde{y}) &= \sup_\theta \mathsf{P}_{Y,\tilde{Y}|\theta}\{\eta(Y, \tilde{Y}) \le \eta(y, \tilde{y})\} \\ &= 1 - \texttt{pchisq}\Big(\frac{(\tilde{y} - y)^2}{\sigma^2(1 + n^{-1})}, \texttt{df} = 1\Big) \end{aligned}$$

Prediction, cont.

- Inference problems tend to be model-driven, i.e., the thing you want to infer is determined by the posited model
- Prediction problems don't inherently depend on a posited model, so there's a desire to be "model-free"
- This is tricky for model-based frameworks...
- There are some options:
  - nonparametric Bayes
  - my IM construction still works in principle
- Balancing validity and efficiency is a challenge
- Warrants considering a non-model-based perspective

Nonparametric predictive inference

- One idea is based on so-called *Hill's assumption*[5]
- Let $Y^n = (Y_1, \ldots, Y_n)$ be observable, Y_{n+1} to be predicted
- Assume $Y_1, \ldots, Y_n, Y_{n+1}$ are exchangeable[6]
- Then Hill's assumption states that

$$\underbrace{\text{“P}\{Y_{n+1} \in (y_{(i)}, y_{(i+1)})\}\text{”}}_{\text{posited predictive prob}} = \frac{1}{n+1}, \quad i = 0, \ldots, n$$

- A symmetry-based y^n-dependent prob statement about Y_{n+1}
- This idea forms the basis of the framework Frank Coolen calls *nonparametric predictive inference*[7]

---

[5] Hill (*JASA* 1968)

[6] i.e., joint distribution invariant to permutations of indices

[7] `https://npi-statistics.com/`

NPI, cont.

- Note that Hill's assumption doesn't identify a single probability distribution for Y_{n+1}, given y^n
- Instead, it just gives bounds

$$\underline{\Pi}_{y^n}(Y_{n+1} \in A) = \frac{1}{n+1} \sum_{i=0}^{n} 1\{(y_{(i)}, y_{(i+1)}) \subseteq A\}$$

$$\overline{\Pi}_{y^n}(Y_{n+1} \in A) = \frac{1}{n+1} \sum_{i=0}^{n} 1\{(y_{(i)}, y_{(i+1)}) \cap A \neq \varnothing\}$$

- These are posited/subjective probs...
- However:
  - prediction intervals: $[y_{(r)}, y_{(s)}]$ for $1 \leq r < s \leq n$
  - by exchangeability,

$$\inf_{\text{exchangeable P}} \mathsf{P}\{Y_{n+1} \in [Y_{(r)}, Y_{(s)}]\} \geq \tfrac{r-s}{n+1}$$

Conformal prediction

- Vladimir Vovk did this in 1990s,[8] recent interest in stat
- Exchangeable $Y_1, \ldots, Y_n$, goal is to predict Y_{n+1}
- Define a *non-conformity measure*, $M(B; z)$, which measures how representative z is of the data in bag B
- For example: $M(B; z) = |\text{median}(B) - z|$
- Needs to be symmetric in B, i.e., M does not depend on the order of the data in B

[8] Vovk, Gammerman, and Shafer (2010) book

Conformal prediction, cont.

- Fix a candidate value y of Y_{n+1}, and write $y_{n+1} = y$
- For $i = 1, \ldots, n, n+1$, evaluate

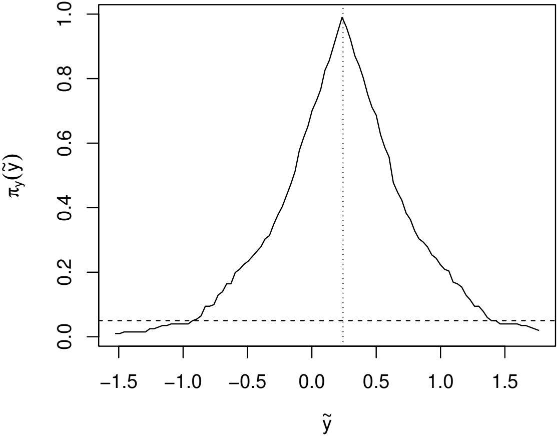
$$\mu_i(y) = M(\{y_1, \ldots, y_{n+1}\} \setminus \{y_i\}; y_i)$$

- Compute the "plausibility" of $Y_{n+1} = y$ as

$$\pi_{y^n}(y) = \frac{1}{n+1} \sum_{i=1}^{n+1} 1\{\mu_i(y) \geq \mu_{n+1}(y)\}$$

- Repeat this for "all" y values to get a $y \mapsto \pi_{y^n}(y)$ function

Conformal prediction, cont.



Conformal prediction, cont.

Theorem.

Define $C_\alpha(y^n) = \{y : \pi_{y^n}(y) > \alpha\}$. Then

$$\inf_{\text{exchangeable P}} \mathsf{P}\{C_\alpha(Y^n) \ni Y_{n+1}\} \geq 1 - \alpha, \quad \alpha \in [0,1]$$

- Coverage guarantee uniformly over all (exchangeable) models!
- Imprecise probability connection:[9]
  - π_{y^n} defines a genuine contour function
  - defines a possibility measure $\overline{\Pi}_{y^n}(A) = \sup_{y \in A} \pi_{y^n}(y)$
  - satisfies a prediction-version of *strong validity*

$$\overline{\mathsf{P}}_{Y^n, Y_{n+1}}\{\pi_{Y^n}(Y_{n+1}) \leq \alpha\} \leq \alpha, \quad \alpha \in [0,1]$$

[9] Cella and M., `https://researchers.one/articles/20.01.00010`

Discussion

- Prediction is a fundamental problem, some say might be more important than inference
- Can be handled via IMs and imprecise prob more generally
- More complex problems considered next time
- Specific IM-related remarks:
  - ML-style model-free problems are a challenge for model-based methods, including (original) IMs, Bayes, fiducial, ...
  - validity required over a very large class
  - need some clever tricks to accommodate this
  - "generalized IM" ideas led to model-free IM solutions[10]

[10] Cella and M., https://researchers.one/articles/21.12.00004

Next lecture

- Prediction with covariates
  - regression
  - classification
- Imprecise probabilistic methods
- ...