Imprecise probability in frequentist inference

Ryan Martin[1,2]
North Carolina State University
Researchers.One

BFF6.5 – Virtual Workshop
https://researchers.one/conferences/bff65
February 5th, 2021

[1] Research partially supported by NSF DMS 1811802.
[2] www4.stat.ncsu.edu/~rgmarti3

Introduction

- Statistical inference is concerned with quantification of uncertainty about unknowns based on data, models, etc.
- "Uncertainty quantification" is usually associated with (ordinary) probability theory.
- BFFs seek out data-dependent probability distributions
  - (generalized) fiducial distributions
  - confidence distributions
  - (objective) Bayesian posterior distributions
  - .....
- Why bother constructing a full probability distribution?

- Answer: belief assignments "$A \mapsto \Pi_Y(A)$" are desirable.
- But the theory ignores the belief assignments, focusing on properties of *procedures* derived from them.
- My (frequentist) view:
  - Basic principle: *if* $\Pi_Y(A)$ *is sufficiently large, then infer* A
  - then $Y \mapsto \Pi_Y(\cdot)$ must be such that following the basic principle leads to "valid" inference
- What math structures in $\Pi_Y(\cdot)$ are consistent with validity?
- Take-away messages:
  - if $\Pi_Y(\cdot)$ is an ordinary probability, then it's not valid
  - if $\Pi_Y(\cdot)$ is a suitable *imprecise probability*, then it's valid

This talk

- Problem setup
- Inferential models (IMs) and validity
- Main results:
  - false confidence theorem
  - consonant plausibility can be valid
  - characterization via imprecise prob
- Open questions:
  - constructing valid IMs
  - false confidence phenomenon
  - relaxing the validity condition

Statistical inference problem

- Observable data: Y in a sample space $\mathbb{Y}$.
- Data are assumed to be relevant to the phenomenon under investigation, formalized through a *model*.
- Statistical model: a collection of probability distributions

$$\mathscr{P} = \{\mathsf{P}_{Y|\vartheta} : \vartheta \in \Theta\}.$$

- θ is the unknown true value; ϑ is a generic value.
- Assume there's no prior information available.
- Goal is uncertainty quantification...

Inferential models

Definition.

An *inferential model* (IM) is a mapping from $(y, \mathscr{P}, \ldots)$ to a data-dependent map $\underline{\Pi}_y : 2^\Theta \to [0,1]$ such that,

$$\underline{\Pi}_y(A) = \text{degree of belief in } A\text{, given } y\text{,...,} \quad A \subseteq \Theta.$$

- The "..." allows for other inputs, e.g., prior information.
- Could be non-additive:
  - Dual to $\underline{\Pi}_y$, $\overline{\Pi}_y(A) = 1 - \underline{\Pi}_y(A^c)$
  - $\overline{\Pi}_y(A)$ represents the plausibility of A, given y,...
- Examples include: Bayes, fiducial, DS, ...

IMs, cont.

- The IM definition is too vague; we need more constraints to pin down $\underline{\Pi}_Y$'s mathematical structure.
- *Basic principle:* if $\underline{\Pi}_Y(A)$ is large, infer A.
- We don't want $\underline{\Pi}_Y(A)$ to be large if A is false.
- Idea: require that $y \mapsto \underline{\Pi}_y(\cdot)$ satisfy

$$\{\theta \notin A \text{ and } Y \sim \mathsf{P}_{Y|\theta}\} \implies \underline{\Pi}_Y(A) \text{ tends to be small.}$$

- Consistent with the message in Reid & Cox:

  it is unacceptable if a procedure. . . of representing uncertain knowledge would, if used repeatedly, give systematically misleading conclusions.

Valid IMs

Definition.

An IM $(y, \mathscr{P}, \ldots) \mapsto \underline{\Pi}_y$ is *valid* if

$$\sup_{\theta \notin A} \mathsf{P}_{Y|\theta}\{\underline{\Pi}_Y(A) > 1-\alpha\} \leq \alpha, \quad \text{for all } A \subseteq \Theta,\ \alpha \in [0,1]$$

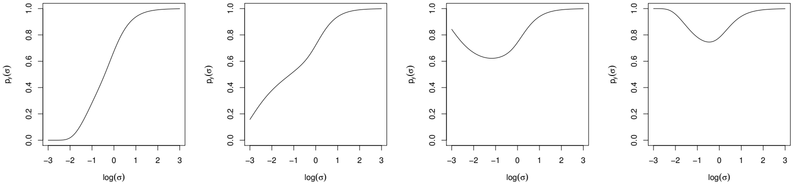
- Intuitively, validity controls the frequency at which the IM assigns relatively high beliefs to false assertions.
- There's an equivalent statement in terms of $\overline{\Pi}_y$.
- Theorem: *If belief assignments are valid, then procedures derived from them are good,* i.e.,
  - "reject $H_0 : \theta \in A$ if $\overline{\Pi}_y(A) \leq \alpha$" is a size α test,
  - the $100(1-\alpha)$% plausibility region $\{\vartheta : \overline{\Pi}_y(\{\vartheta\}) > \alpha\}$ has coverage probability $\geq 1-\alpha$.

Valid IMs, cont.

- *Question:* What kind of mathematical structure in the belief assignments $A \mapsto \underline{\Pi}_y(A)$ is consistent with validity?
- Summary of the three main results:
  1. *if* $\Pi_Y = \underline{\Pi}_Y = \overline{\Pi}_Y$ *is a probability, then it's not valid*
  2. *if* $\overline{\Pi}_Y$ *is a consonant plausibility function, aka a possibility measure, then it can be valid*
  3. *procedures with good frequentist properties correspond to valid IMs that are possibility measures.*

Main result 1

- Satellite conjunction analysis:
  - Orbiting satellite could collide with another object.
  - To avoid this, analysts compute a *non-collision probability*.[3]
  - Satellite judged to be safe if non-collision probability is high.
  - *Noisier data makes non-collision probability large*[4]



(a) Close (b) Kinda close (c) Kinda far (d) Far

[3] Details in Balch, M., and Ferson (2019), `arXiv:1706.08565`.

[4] M. (2019), `researchers.one/articles/19.02.00001`

Main result 1, cont.

- *False confidence:* Hypotheses tending to be assigned high probability even if data don't support them.
- Probabilities are afflicted with false confidence, *not valid*.
- "Is [probability] just quick and dirty confidence?" (Fraser)

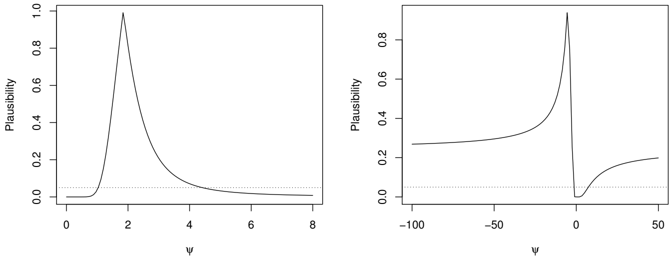
False confidence theorem (Balch, M., and Ferson 2019).

Let Π_Y be a probability on Θ, depending on data Y. Then, for any $\alpha \in (0,1)$ and any $\theta \in \Theta$, there exists $A \subset \Theta$ such that

$$A \not\ni \theta \quad \text{and} \quad \mathsf{P}_{Y|\theta}\{\Pi_Y(A) > 1 - \alpha\} > \alpha.$$

Main result 2

- Consonant plausibility function/possibility measure:
  - Plausibility contour, $\pi_y : \Theta \mapsto [0,1]$ with $\sup_\vartheta \pi_y(\vartheta) = 1$.
  - Consonant plausibility: $\overline{\Pi}_y(A) = \sup_{\vartheta \in A} \pi_y(\vartheta)$.
- Two examples of a plausibility contour function...



Main result 2, cont.

- Frequentist procedures are based on p-values:
  - $(y, \vartheta) \mapsto \pi_y(\vartheta)$, small values imply y and ϑ disagree
  - If $Y \sim \mathsf{P}_{Y|\theta}$, then $\pi_Y(\theta) \sim \mathsf{Unif}(0,1)$.
- Comparison:
  - P-value + additivity (= CD) $\to$ not valid.
  - P-value + consonance $\to$ valid.[5]

Theorem (M. 2021).

If the p-value π_y satisfies the conditions of a plausibility contour, then the IM with $\overline{\Pi}_y(A) = \sup_{\vartheta \in A} \pi_y(\vartheta)$ is valid.

[5] M. (2021), `researchers.one/articles/21.01.00002`

Main result 3

Theorem (M. 2021).

Let $\{C_\alpha : \alpha \in [0,1]\}$ be a family of confidence regions for $\phi = \phi(\theta)$ that satisfies the following properties:

- Coverage. $\inf_\theta \mathsf{P}_{Y|\theta}\{C_\alpha(Y) \ni \phi(\theta)\} \geq 1 - \alpha$ for all α;
- Nested. if $\alpha \leq \beta$, then $C_\beta \subseteq C_\alpha$;
- Compatible. ....

There exists a valid IM for θ, *a possibility measure*, whose derived marginal plausibility regions $C_\alpha^\star(y)$ for ϕ satisfy

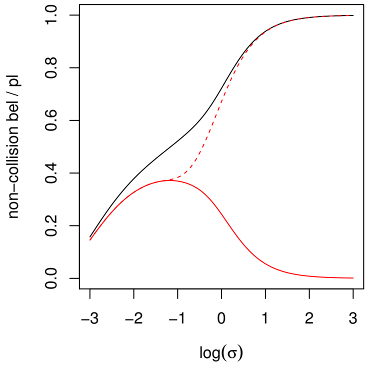
$$C_\alpha^\star(y) \subseteq C_\alpha(y) \quad \text{for all } (y, \alpha) \in \mathbb{Y} \times [0,1].$$

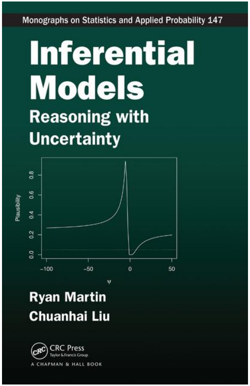
(There's an analogous result for tests.)

M. (2021), `researchers.one/articles/21.01.00002`

Open problem: valid IM construction

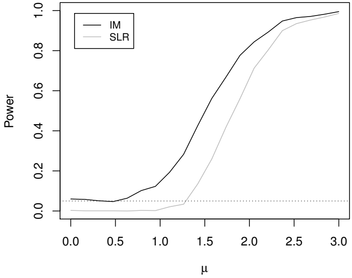
- Random sets on an auxiliary space.





Open problem: valid IM construction, cont.

- Possibility measures on an auxiliary space.[6]
- P-values + consonance
- New *IM algorithm*[7]
  - Proof of the above theorem is *constructive*
  - Defines an "algorithm" for constructing a valid/efficient IM.



[6] Liu and M. (2020), researchers.one/articles/20.08.00004
[7] M. (2021), researchers.one/articles/21.01.00002

Open problem: understanding false confidence

- Recall: *for any additive IM, there exists false assertions A that tend to be assigned high probability.*
- How serious of a problem is this? Can we characterize the assertions afflicted with false confidence?
- In my experience, the non-trivial examples occur when A is about some non-linear function $\phi = \phi(\theta)$.
- Marginalization challenges due to non-linearity:
  - Dawid et al (1973)
  - Gleser & Hwang (1987)
  - Fraser (2011), Fraser et al (2016)
- Do these issues relate to false confidence? If so, how?

Open problem: weakening validity

- Validity is a reasonable condition, but it is strong.
- Adjustments needed to incorporate (partial) prior information or structural assumptions, e.g., sparsity.
- New approach in prediction.[8]
  - $(Y, \tilde{Y}) \sim \mathsf{P}$; observe Y, predict $\tilde{Y}$.
  - "Probabilistic predictor" $\overline{\Pi}_Y$ is valid if

$$\underbrace{\mathsf{P}\{\overline{\Pi}_Y(A) \leq \alpha \, , \tilde{Y} \in A\}}_{\mathsf{E}[1\{\overline{\Pi}_Y(A) \leq \alpha\} \, \mathsf{P}(\tilde{Y} \in A \mid Y)]} \leq \alpha, \quad \text{all } (\alpha, A, \mathsf{P}).$$

  - There are some interesting consequences of this definition...
- How might this look in an inference problem?

[8] Cella and M. (2020), `researchers.one/articles/20.01.00010`

Open problem: weakening validity, cont.

Validity.

$$\sup_{\theta \notin A} \mathsf{P}_{Y|\theta}\{\underline{\Pi}_Y(A) > 1 - \alpha\} \leq \alpha, \quad \text{for all } A \in 2^{\Theta}, \; \alpha \in [0,1]$$

- Bound α away from 1 (Walley 2002)
- Restrict to a large but proper subset of assertions
- Replace supremum by a more general upper expectation

Validity – weaker version?

$$\overline{\mathsf{E}}_\theta \big[1_{\theta \notin A} \, \mathsf{P}_{Y|\theta}\{\underline{\Pi}_Y(A) > 1 - \alpha\} \big] \leq \alpha, \quad \text{for all } A \in \mathcal{A}, \; \alpha \in [0, \bar{\alpha}]$$

Conclusion

- Imprecise probabilities are closely tied to frequentist inference.
- Important for interpretation:
  - *p-values are the plausibility of the null hypothesis*
  - *confidence regions are collections of parameter values that are sufficiently plausible*
- Hopefully can capture the spectrum from no prior info to complete prior info, and a corresponding notion of validity.
- New and exciting territory.

Thanks for your attention!

`www4.stat.ncsu.edu/~rgmarti3`