False confidence, non-additive beliefs, and valid statistical inference[1][2]

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

6th African International Conference on Statistics
Arsi University, Ethiopia
May 28th, 2019

[1]Paper available at https://www.researchers.one/article/2019-02-1

[2]Thanks to friends and collaborators M. Balch, H. Crane, C. Liu,...

Introduction

- Statistics has developed a lot in the last 50+ years.
- But important fundamental questions about the role of probability in statistical inference remain unanswered.
- Currently we have two dominant schools of thought:
  - *frequentist*
  - *Bayesian*
- Lots of debate over the years about which one is "right."
- This hasn't really helped.
- I want to focus on what science needs from statistics which, in my assessment, is a level of *reliability*.

Intro, cont.

- Statistical problem:
  - Observable data is Y;
  - Model is $\mathscr{P} = \{\mathsf{P}_{Y|\theta} : \theta \in \Theta\}$ — taken as given today.
- Scientific questions correspond to hypotheses about θ.
- Goal is to quantify uncertainty about θ, given $Y = y$.
- Define an *inferential model*

$$(y, \mathscr{P}, \ldots) \mapsto b_y : 2^\Theta \to [0, 1],$$

  where $b_y(A)$ represents the data analyst's degrees of belief about a hypothesis $A \subseteq \Theta$ based on data y, model $\mathscr{P}$, etc.
- The inferential model could be lots of things:
  - a Bayesian posterior distribution;
  - a fiducial or confidence distribution;
  - ...

Intro, cont.

- Pointless if the inferential model isn't "reliable."
- Brad Efron said (roughly):

  > *the construction of reliable, prior-free, inferential models is the most important unresolved problem in statistics.*

- I'm going to tell you about my attempt at a solution.
- Key insight: *go non-additive!*
  - Familiar things are *additive*, i.e., b_y is a probability.
  - But additivity isn't necessary, might even be a constraint.
  - "There's more to uncertainty than probabilities"[3]
- Take-away messages:
  - additive beliefs are afflicted with false confidence (bad)
  - but good non-additive beliefs can avoid false confidence

[3] Theme of a conference I'm participating in this summer in Ghent on *imprecise probabilities*, http://www.isipta2019.ugent.be

This talk

- The price of additivity:
  - satellite collision example
  - false confidence theorem
- Going non-additive:
  - avoiding false confidence: the validity property
  - non-additive beliefs and random sets
  - construction of valid inferential models
- Satellite collision, revisited
- Conclusion

(Questions planted in blue)

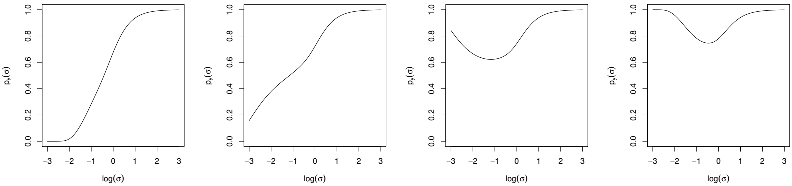
Satellite collision problem

- A satellite orbiting Earth could collide with another object.
- Potential mess, so navigators try to avoid collision.
- Data on position, velocity, etc is used, along with physical and statistical models, to compute a *collision probability*.
- If collision probability is low, then satellite is judged safe; otherwise, some action is taken.
- An unusual phenomenon has been observed in the satellite conjunction analysis literature:

noisier data makes non-collision probability large.

Satellite collision, cont.

- Simple illustration:
  - $\|y\|$ denotes measured distance between satellite and object.
  - True distance ≤ 1 implies collision.
  - Measurement error variance is σ^2.
  - $p_y(\sigma)$ = probability of non-collision, given y.
- When σ is large, $p_y(\sigma)$ is large, no matter what y is!
- *Potentially misleading conclusions!*



(a) $\|y\|^2 = 0.5$ (b) $\|y\|^2 = 0.9$ (c) $\|y\|^2 = 1.1$ (d) $\|y\|^2 = 1.5$

False confidence

- What's going on here?
- Apparently, data may not be sufficiently informative with respect to questions about collision/non-collision.
- Additivity forces the probability to go somewhere, happens that it goes to non-collision no matter what data says.
- i.e., even if satellite is on a direct collision course, if there's enough noise, then probability tells navigators that it's safe.
- *False confidence:*[4] a hypothesis tending to be assigned large probability even though data does not support it.
- Is this a general phenomenon...?

[4] Balch, M., and Ferson, `arXiv:1706.08565`

False confidence, cont.

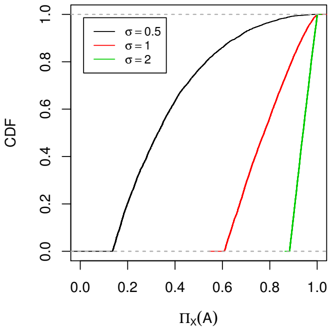
False confidence theorem.

Let Π_Y be any additive belief on Θ, depending on data Y. Then, for any α and any t, there exists $A \subset \Theta$ such that

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

- In words, there always exists a false hypothesis that tends to be assigned high posterior probability.
- If judgments about the plausibility of A are made based on the magnitude of $\Pi_Y(A)$, then there's risk of systematic error.
- Theorem doesn't say which hypotheses are afflicted (any clue?), only that they exist.
- What about asymptotics?

- Simplified version of the satellite example.
  - Let $A = \{\text{non-collision}\}$.
  - Then $\Pi_Y(A)$ as a random variable, with a CDF
  - Plot CDF when data are generated under a collision course.
- *False confidence:* $\Pi_Y(A)$ is almost always large!



Avoiding false confidence

- Nancy Reid and D. R. Cox write:

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

- In other words, unreliability is bad.
- Theorem says *every additive belief function* is afflicted with false confidence, at risk of being unreliable.
- To avoid this, b_y must be non-additive, e.g.,

$$b_y(A) + b_y(A^c) < 1.$$

- But not every non-additive belief avoids false confidence.
- So we need some additional restrictions...

Validity property

Definition.

An inferential model $(y, \mathscr{M}, \ldots) \mapsto b_y$ is *valid* if

$$\sup_{\theta \notin A} \mathsf{P}_{Y|\theta}\{b_Y(A) > 1 - \alpha\} \leq \alpha, \quad \forall\, A \subseteq \Theta, \quad \forall\, \alpha \in (0,1).$$

- Validity property implies that assigning high belief to a false hypothesis is a rare event — *no false confidence*.
- "For all $A \subseteq \Theta$" is a strong condition — why impose it?
- α on inside and outside calibrates the belief function values, i.e., so that I know what "large" and "small" means.
- Implies that procedures derived from a valid b_y have frequentist error rate control; details later.

Non-additive beliefs

- Simplest way to construct non-additive beliefs on a space $\mathbb{X}$ is via a *random set* $\mathcal{X} \sim \mathsf{P}_{\mathcal{X}}$ that takes values in $2^{\mathbb{X}}$.
- For a fixed set $A \subseteq \mathbb{X}$, a realization of $\mathcal{X}$ could be
  - fully contained in A,
  - fully contained in A^c,
  - or have non-empty intersection with both.
- Then the containment functional

$$b(A) = \mathsf{P}_{\mathcal{X}}(\mathcal{X} \subseteq A), \quad A \subseteq \mathbb{X},$$

is non-additive, in particular, $b(A) + b(A^c) \leq 1$.

Construction of valid non-additive beliefs

- Express the statistical model, $\mathsf{P}_{Y|\theta}$, as

$$Y = a(\theta, U), \quad U \sim \mathsf{P}_U, \quad \mathsf{P}_U \text{ known.}$$

- Intuition: *If U were observable*, then just solve for θ in terms of Y and U — done!
- Unfortunately, *U is not observable*...
- Its distribution is known, so we can "guess" its unobserved value with certain degree of reliability.
- This "guess" is based on a random set $\mathcal{S} \sim \mathsf{P}_{\mathcal{S}}$ in the U-space.
- There is theory to guide the choice of $\mathsf{P}_{\mathcal{S}}$.

Construction, cont.

- Given $\mathcal{S} \sim \mathsf{P}_{\mathcal{S}}$, push it forward to the θ space:

$$\Theta_y(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \{\vartheta : y = a(\vartheta, u)\}.$$

- The *belief function* is just the containment functional

$$b_y(A) = \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \subseteq A\}.$$

- Dual is the *plausibility function*

$$p_y(A) := 1 - b_y(A^c) = \mathsf{P}_{\mathcal{S}}\{\Theta_y(\mathcal{S}) \cap A \neq \varnothing\}.$$

- Non-additive, i.e., $b_y(A) \leq p_y(A)$ for all $A \subseteq \Theta$.

Validity theorem

Theorem.

With a "suitable choice of random set $\mathcal{S} \sim \mathsf{P}_{\mathcal{S}}$," the inferential model constructed above satisfies the validity condition.

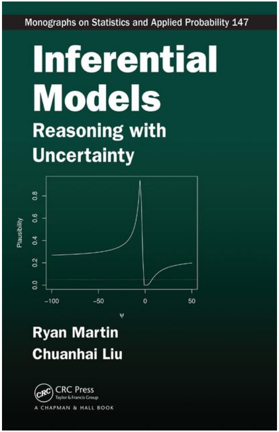
- Re: "suitable choice of $\mathcal{S}$."
  - Let $f(u) = \mathsf{P}_{\mathcal{S}}(\mathcal{S} \ni u)$.
  - Need $f(U) \geq_{\text{st}} \text{Unif}(0,1)$ when $U \sim \mathsf{P}_U$.
- This is actually easy to arrange...
- Consequences of validity:
  - "Reject $H_0 : \theta \in A$ if $p_y(A) \leq \alpha$" is a size α test.
  - The $100(1-\alpha)\%$ plausibility region

$$\{\vartheta : p_y(\{\vartheta\}) \geq \alpha\}$$

    is a $100(1-\alpha)\%$ confidence region.

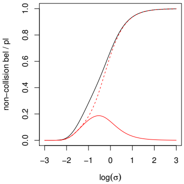
Commercial

If you're still looking for that perfect summer-time beach read, then your search is over!

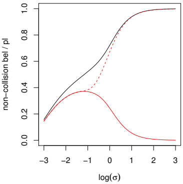


Satellite collision, revisited

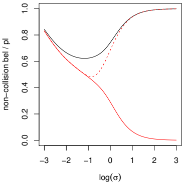
- Same illustration as before, but now showing
  - same non-collision probability (black)
  - non-collision belief (red) and plausibility (red, dashed).
- Notice belief and plausibility split when σ is large.
- What does the gap between them represent?



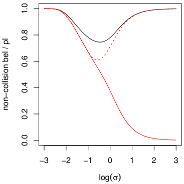
(e) $\|y\|^2 = 0.5$



(f) $\|y\|^2 = 0.9$

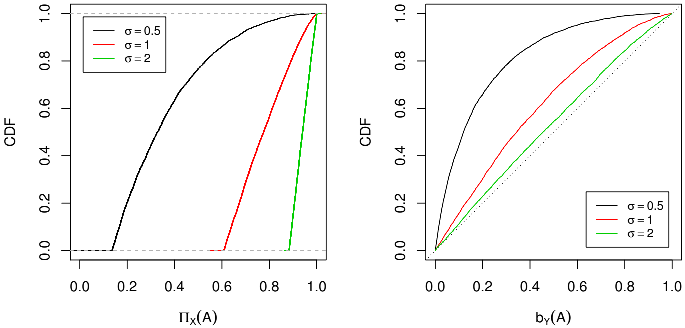


(g) $\|y\|^2 = 1.1$



(h) $\|y\|^2 = 1.5$

- By construction, there is no false confidence.
- Check by evaluating CDF of $b_Y(A)$ for $A = \{\text{non-collision}\}$.
- Same experiment as before.
  - Non-collision probabilities (left) tend to be near 1.
  - Non-collision belief (right) tends to be near 0 — *correct!*



Concluding remarks

- Additive beliefs put user at risk of false confidence.
- Avoid this risk with suitable non-additive beliefs.
- Close connections between non-additive beliefs and classical notions like p-values and confidence.[5]
- Many non-trivial details I didn't tell you...
- Lots of interesting questions that remain to be answered:
  - marginalization mysteries
  - efficiency and optimality
  - model uncertainty
  - computation
  - ...

[5] arXiv:1211.1547, arXiv:1606.02352, and arXiv:1707.00486

One more commercial

- The peer review system is broken in various ways.
- How so?
- Successful reform requires new ideas.
- Harry Crane and I developed a new open-access publication platform, featuring an *author-driven* peer review process.

RESEARCHERS.ONE

- For details, check us out at

www.researchers.one
www.twitter.com/@ResearchersOne

The end

Thank you!

rgmarti3@ncsu.edu
www4.stat.ncsu.edu/~rmartin

RES3ARCHERS.ONE