Reliable statistical inference using consonant belief functions

Ryan Martin[1]
North Carolina State University
Researchers.One
www4.stat.ncsu.edu/~rgmarti3

7th BFAS School
University of Granada, Spain
October 20th, 2025

[1]Research partially supported by the NSF, SES–2051225

Introduction

> *Perhaps the most important unresolved problem in statistical inference is the use of Bayes theorem in the absence of prior information.* —Efron

- Dempster & Shafer are statisticians so, at some level, the foundations of statistics brought all of us here together
- Indeed, Dempster's work originally aimed to solve this "most important unresolved problem" in statistics
- This is the same problem that Fisher failed to solve
- My goal is to finally solve this problem
  - I don't work on belief functions specifically
  - but these ideas are necessary to solve my problem

Intro, cont.

- Rough timeline of statistical developments:
  - inverse probability $\rightsquigarrow$ "Bayesian statistics"
  - Fisher's likelihood, p-value, fiducial $\rightsquigarrow$ "Fisherian statistics"
  - Neyman rewrote Fisher $\rightsquigarrow$ "frequentist statistics"
  - Fisher's (& Dempster's) perspective was lost, Bayesianism and frequentism are in a complacency-driven stalemate

> *How can a discipline, central to science and to critical thinking, have two methodologies, two logics, two approaches that frequently give substantially different answers to the same problems?* —Fraser

Intro, cont.

- Forward progress on foundations requires new insights
- My basic claim: *probability theory is lacking, reliable inference & UQ requires different considerations*
- "Different considerations" = imprecise probability
- Note: imprecision is *necessary*, it's not a choice
- I'm proposing a general imprecise-probabilistic framework:
  - consonant belief structures, possibility theory
  - balances Bayesian and frequentist reliability
  - can do things Bs & Fs can't

This talk

- setup
- false confidence and implications for precise prob
- consonant beliefs/possibility = "correct" mode of imprecision
- possibilistic inferential model (IM) construction
- properties
  - validity, basic reliability
  - efficiency
  - credal set properties, computation, etc
- examples
- some "beyond the basics" stuff
- FAQs

Setup

- Data $Y \in \mathbb{Y}$, observed value y
- Model $\{\mathsf{P}_\theta : \theta \in \mathbb{T}\}$
- Uncertain true value is Θ, generic values denoted by θ
- For now, prior info about Θ is vacuous
- **Goal:** UQ/inference about Θ, given $Y = y$

- For example:
  - model $(Y \mid \Theta = \theta) \sim \mathsf{N}(\theta, 1)$
  - observation $Y = y$, e.g, $y = 7$
  - is the hypothesis "$\Theta > 8$" supported?

Setup, cont.

> *It would surely have been astonishing if all the complexities of such a subtle concept as probability in its application to scientific inference could be represented in terms of only three concepts—estimates, confidence intervals, and tests of hypotheses. Yet one would get the impression that this was possible from many textbooks purporting to expound the subject.* —Barnard

- *Inference isn't just estimation, testing, and confidence sets*
- Scientists don't really want, e.g., hypothesis tests!
- The UQ I'm after is richer & more informative

Setup, cont.

> *Statisticians want numerical measures of the degree to which data support hypotheses.* —Hacking

- UQ about Θ, given $Y = y$?
- Hacking says we want "numerical measures of..."
- Gut reaction:
  - "numerical measures of..." $\iff$ probability
  - i.e., return a "posterior" $(\Theta \mid Y = y) \sim \mathsf{Q}_y$
- However: *probability theory generally doesn't provide reliable UQ in the context statistical inference*

Why imprecision?

- *False confidence*[2] is a particular form of unreliability that <u>all</u> probabilistic UQ suffers from
- By "unreliability" I mean the following:
  - for some false hypotheses $H \not\ni \Theta$
  - random variable $Y \mapsto \mathsf{Q}_Y(H)$ tends to be large under P_Θ
- Define the *false confidence rate*

$$\text{FCR}_{\mathsf{Q}}(\alpha, H) = \sup_{\theta \notin H} \mathsf{P}_\theta\{\underbrace{\mathsf{Q}_Y(H) > 1 - \alpha}_{\text{confident } H \text{ is true}}\}$$

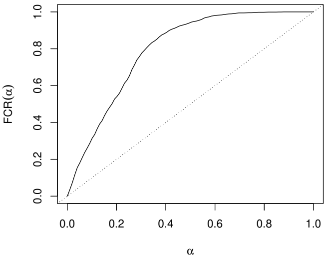
False confidence theorem.

Let Q_Y be *any* data-dependent probability for Θ. For any (α, τ), there exists hypotheses $H \subset \mathbb{T}$ such that $\text{FCR}_{\mathsf{Q}}(\alpha, H) > \tau$.

[2] Balch, M., & Ferson 2019, *Proc Roy Soc A*, `arXiv:1706.08565`

Why imprecision?, cont.

- Simple linear regression: $(Y_i \mid x_i) \overset{\text{ind}}{\sim} \mathsf{N}(\beta_0 + \beta_1 x_i, \sigma^2)$
- Covariates are fixed, model parameter $\theta = (\beta_0, \beta_1, \sigma^2)$
- Q_y = Bayes posterior w/ diffuse normal–igamma prior
- Simulation setup:
  - $H = \{(\beta_0, \beta_1, \sigma^2) : -\beta_0/\beta_1 > -1\}$
  - data generated with $\Theta = (0.3, 0.1, 1)$, so H is false



Why imprecision?, cont.

- Two-player perspective on false confidence
  - statistician carrying out (probabilistic) UQ
  - scrutinizer who can bet against the statistician
- Condition "$\mathrm{FCR}_{\mathsf{Q}}(\alpha, H) > \alpha$" implies
  - there exists gambles that are acceptable to the statistician, i.e., non-negative Q_y-expected payoff, for each y
  - but are also unacceptable to the statistician, i.e., have negative payoff, on average over y for some θ
- Reveals a type of *incoherence* in probabilistic UQ...
- For details, see `arXiv:2312.14912`

Why imprecision?, cont.

- Examples of false confidence are easy to find
- But what is the cause of false confidence?
- Current best guess: caused by *non-linearity* in hypotheses

Home > Belief Functions: Theory and Applications > Conference paper

Which Statistical Hypotheses are Afflicted with False Confidence?

Conference paper | First Online: 20 August 2024

pp 140–149 | Cite this conference paper

Access provided by Consorcio de Bibliotecas Universitarias de Andalucía / CBUA

Why imprecision?, cont.

> *[Fisher] was the world's master of quantifying uncertainty, having developed many of the... procedures for doing so.* —Pearl

- The need for imprecision in UQ wasn't lost on Fisher
  - *[A p-value]... does not justify any exact probability statement*
  - *no exact probability statements can be based on [conf. sets]*
- "Fisher's biggest blunder" was just that he didn't find the right mathematical formulation
- Dempster's effort to fix Fisher is why we're here today
- But their original problem still isn't settled...

What imprecision?

- If not probability, then what? *Imprecise probability*
- Note: imprecise $\neq$ inaccurate
- Roughly, imprecise probability $=$ *sets of probabilities*
- Whether statisticians realize it or not, they're working with sets of (Y, Θ) joint distributions
  - e.g., the *confidence level* of a set estimator $C(Y)$ is

$$\sup_{\theta} \mathsf{P}_\theta\{C(Y) \not\ni \theta\} = \sup_{\text{all priors } \mathsf{Q}} \underbrace{\int \mathsf{P}_\theta\{C(Y) \not\ni \theta\} \, \mathsf{Q}(d\theta)}_{\text{linear functional of } \mathsf{Q}}$$

  - "$=$" because the supremum of a linear functional over a convex set is on the boundary (point masses)

What imprecision?, cont.

- If even frequentists have sets of joint distributions, then
  - imprecision in inference is clear/unavoidable
  - helps to explain the false confidence phenomenon
- Data-dependent imprecise probability for (reliable) UQ...

Definition — *inferential model* (IM).

- given data y, model, etc
- to every $H \subseteq \mathbb{T}$, assign a pair of numbers

$$\underline{\Pi}_y(H) \quad \text{and} \quad \overline{\Pi}_y(H)$$

- y-dependent measures of $\underline{\text{support for}}$ and $\overline{\text{plausibility of}}$ H

What imprecision?, cont.

- There's a wide range of probability alternatives
  - belief functions
  - consonant beliefs/possibility measures
  - credal sets
  - lower/upper previsions
  - ...
- All have advantages and disadvantages
- My focus: *consonant belief functions/possibility measures*
- I'll justify this later; for now just a quote...

> *Specific items of evidence can often be treated as consonant, and there is at least one general type of evidence that seems well adapted to such treatment. This is inferential evidence—the evidence for a cause that is provided by an effect.* —Shafer

What imprecision?, cont.

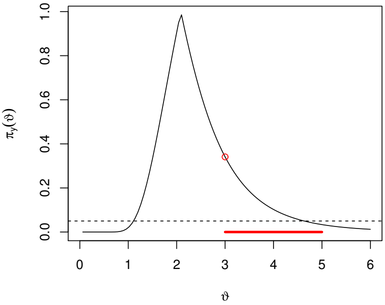
- *Possibility theory* is the simplest imprecise probability
- Equivalent to belief functions with nested focal elements
- Similar to probability theory
  - determined by a "density function"
  - different calculus — optimization replaces integration
- Specifically:
  - possibility contour $\theta \mapsto \pi_y(\theta)$ with $\sup_\theta \pi_y(\theta) = 1$
  - and then define

$$\overline{\Pi}_y(H) = \sup_{\theta \in H} \pi_y(\theta) \quad \text{and} \quad \underline{\Pi}_y(H) = 1 - \overline{\Pi}_y(H^c)$$

- *What makes the IM output meaningful is its properties*

What imprecision?, cont.

- Possibility contour plot
- Hypothesis $H = [3, 5]$
- $\overline{\Pi}_y(H)$ via optimization
- Level set defines interval of "sufficiently possible" θs



What imprecision?, cont.

- Possibility theory for statistics isn't my idea!
- A very natural choice of possibility contour/measure in stat applications (e.g., Shafer 1976, Ch. 11)
  - relative likelihood $R(y, \theta) = L_y(\theta) / \sup_\vartheta L_y(\vartheta)$
  - for fixed $Y = y$, set

$$\pi_y(\theta) = R(y, \theta), \quad \theta \in \mathbb{T}$$

  - and then $\overline{\Pi}_y(H) = \sup_{\theta \in H} \pi_y(\theta)$
- Further investigation into this proposal:
  - Wasserman (1990), *Canad. J. Stat.*
  - Denoeux (2014), *IJAR*
  - ...
- Shafer later (*JRSS-B*, 1982) rejected the idea in general
- I think it just doesn't go far enough...

Possibilistic IMs

- Special case of a more general construction[3]
- Workhorse: *imprecise-probability-to-possibility transform*[4]
- In the "no-prior" case, this takes a simple form:
  - same relative likelihood $R(y,\theta) = L_y(\theta) / \sup_\vartheta L_y(\vartheta)$
  - define a possibility contour

$$\pi_y(\theta) = \mathsf{P}_\theta\{R(Y,\theta) \leq R(y,\theta)\}, \quad \theta \in \mathbb{T}$$

  - then $\overline{\Pi}_y(H) = \sup_{\theta \in H} \pi_y(\theta)$, etc.
- Not unfamiliar...[5]
- Unexpected connections to Bayes/fiducial inference (later)

[3] M. `arXiv:2211.14567`

[4] e.g., Dubois et al (2004), *Rel. Comput.*; Hose's 2022 PhD thesis

[5] M. `arXiv:1203.6665` and M. `arXiv:1511.06733`

Possibilistic IMs, cont.

- A relatively simple example:
  - $Y = (Y_1, \ldots, Y_n)$ iid $\mathsf{P}_\theta = \mathsf{Exp}(\theta)$
  - likelihood $\theta \mapsto \theta^{-n} \exp(-n\bar{y}/\theta)$
  - relative likelihood $R(y, \theta) = (\theta/\bar{y})^n \exp(-n\bar{y}/\theta)$
- Compute the possibilistic IM contour:

$$\begin{aligned} \pi_y(\theta) &= \mathsf{P}_\theta\{R(Y, \theta) \leq R(y, \theta)\} \\ &= \mathsf{P}\{Z^{-n} e^{-Z} \leq (\theta/n\bar{y})^n e^{-n\bar{y}/\theta}\}, \quad \theta > 0 \end{aligned}$$

  where $Z \sim \mathsf{Gamma}(n, 1)$
- Plot from a few slides earlier is based on this

Possibilistic IMs, cont.

What makes the IM output meaningful is its properties. —M

It is unacceptable if a procedure... of representing uncertain knowledge would, if used repeatedly, give systematically misleading conclusions. —Reid & Cox

Strong validity theorem.

$$\sup_\theta \mathsf{P}_\theta\{\pi_Y(\theta) \leq \alpha\} \leq \alpha, \quad \alpha \in [0,1]$$

- Most basic property, familiar for the no-prior case
- Implies no false confidence, among other things

Possibilistic IMs, cont.

Corollary.

1. the test "reject $\Theta \in H$ iff $\overline{\Pi}_y(H) \le \alpha$" satisfies

$$\sup_{\theta \in H} \mathsf{P}_\theta\{\overline{\Pi}_Y(H) \le \alpha\} \le \alpha, \quad \alpha \in [0,1], \quad H \subseteq \mathbb{T}$$

2. the region $C_\alpha(y) = \{\theta : \pi_y(\theta) > \alpha\}$ satisfies

$$\sup_{\theta} \mathsf{P}_\theta\{C_\alpha(Y) \not\ni \theta\} \le \alpha, \quad \alpha \in [0,1] \tag{1}$$

- These are the classical frequentist properties investigated in math-stat textbooks
- Note: $C_\alpha'(y) = \{\theta : R(y,\theta) > \alpha\}$ doesn't satisfy (1)

Possibilistic IMs, cont.

- There's a sort of converse to the above corollary
- Suggests consonance is inherent in frequentism

Theorem (arXiv:2112.10904 or arXiv:2507.09007).

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

Coverage. $\sup_\theta \mathsf{P}_\theta\{C_\alpha(Y) \not\ni \phi(\theta)\} \leq \alpha$ for all α

Nested. if $\alpha \leq \beta$, then $C_\beta(y) \subseteq C_\alpha(y)$ for all y

There exists a valid possibilistic IM for Θ with contour π_y such that

$$\phi(\theta) \in C_\alpha(y) \iff \pi_y(\theta) > \alpha \quad \text{for all } (y, \alpha) \in \mathbb{Y} \times [0,1].$$

Possibilistic IMs, cont.

Theorem.

The possibilistic IM's error rate control is *uniform in H*, i.e.,

$$\sup_\theta \mathsf{P}_\theta\{\overline{\Pi}_Y(H) \leq \alpha \text{ for some } H \ni \theta\} \leq \alpha$$

- Inference goal isn't just "test one H and done"
- Often the goal is to *probe*
  - e.g., if null is rejected, find other H's that are supported
  - follow-ups depend on data, so H-wise control isn't enough
- Above result provides error rate control under probing[6]
- Implications of this aren't fully understood yet

[6] Cella and M. (2023), *IJAR*, `arXiv:2304.05740`

Possibilistic IMs, cont.

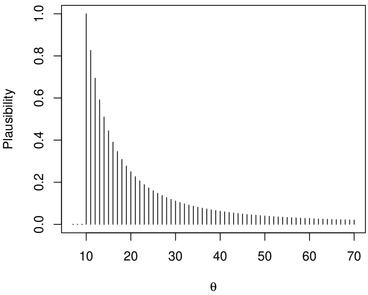
- Computation usually requires Monte Carlo
- For each θ on a grid
  - simulate independent $Y_\theta^{(m)} \sim \mathsf{P}_{Y|\theta}$, $m = 1, \ldots, M$
  - approximate the contour by

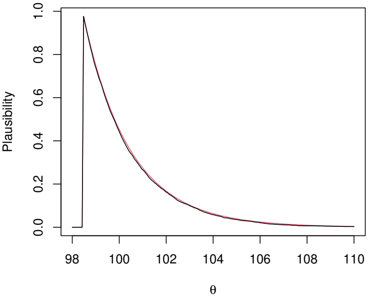
$$\pi_y(\theta) \approx \frac{1}{M} \sum_{m=1}^{M} 1\{R(Y_\theta^{(m)}, \theta) \le R(y, \theta)\}$$

- Simplifications are possible in specific examples
- Importance sampling can also be used to lower expenses
- More on computation later

Possibilistic IMs, cont.

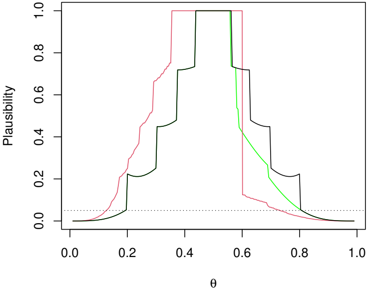
- Plots of $\theta \mapsto \pi_y(\theta)$ for two different uniform models
  - $P_\theta = \text{Unif}\{1, 2, \ldots, \theta\}$, $Y_{(n)} = 10$, $n = 2$
  - $P_\theta = \text{Unif}(\theta, \theta^2)$, $(Y_{(1)} = 281, Y_{(n)} = 9689)$, $n = 25$





Possibilistic IMs, cont.

- $\mathsf{P}_\theta = \mathsf{Bin}(n, \theta)$
- Plots of the contour $\theta \mapsto \pi_y(\theta)$ — focus on black curve[7]
- Contours: $n = 8$ (left) and $n = 16$ (right), $\hat{\theta} = 0.5$



[7]Others: partial prior IMs that assume "90% sure that $\Theta \leq 0.6$"

Possibilistic IMs, cont.

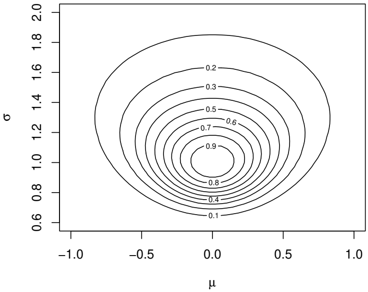
- Bivariate normal model:
  - means are 0, variances are 1
  - unknown correlation Θ
- Simulations to check validity & efficiency

			Θ				
n	Metric	Method	0.05	0.25	0.50	0.75	0.90
10	Coverage	IM	0.957	0.968	0.953	0.947	0.942
		$r^\star$	0.927	0.923	0.933	0.930	0.923
	Length	IM	1.004	0.979	0.888	0.619	0.242
		$r^\star$	0.947	0.919	0.818	0.540	0.238
25	Coverage	IM	0.949	0.960	0.957	0.950	0.941
		$r^\star$	0.938	0.943	0.948	0.944	0.942
	Length	IM	0.729	0.695	0.574	0.311	0.117
		$r^\star$	0.695	0.662	0.545	0.309	0.127

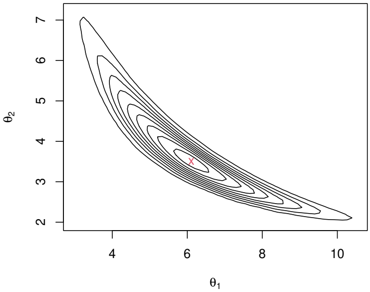
Table 1: Empirical coverage probabilities and expected lengths for the IM- and $r^\star$-based 95% confidence intervals in the bivariate normal correlation case described in Example 5.

Possibilistic IMs, cont.

- $\mathsf{P}_\theta = \mathsf{N}(\mu, \sigma^2)$
- Data: $n = 10$, $(\hat{\mu}, \hat{\sigma}) = (0, 1)$
- Map of the IM possibility contour



- $\mathsf{P}_\theta = \mathsf{Gamma}(\theta_1, \theta_2)$
- Simulated data, $n = 25$, $\theta_1 = 7$ and $\theta_2 = 3$
- Map of the IM possibility contour
- Computation via importance sampling



Possibilistic IMs, cont.

- Behrens–Fisher: difference of normal means, general variances
- New marginal IM solution[8]
- Strong validity is guaranteed, sim to check efficiency:
  - difficult unbalanced case, $(n_1, n_2) = (2, 20)$
  - $\mu_1 = 2$, $\mu_2 = 0$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$
  - compare coverage prob of 90% confidence intervals

Method	Coverage Prob
Hsu–Scheffe	0.9738
Jeffreys	0.9296
Ghosh & Kim	0.7873
Welch	0.8362
1st order	0.7399
Fraser et al	0.8617
IM	**0.9082**

[8] M. (2023), `arXiv:2309.13454`

Possibilistic IMs, cont.

- Validity concerns errors like the following:
  - true hypotheses H that I assign small $\overline{\Pi}_y(H)$
  - true Θ is outside my confidence set
- *Efficiency* concerns other kinds of errors:
  - false hypotheses H that I assign not-small $\overline{\Pi}_y(H)$
  - false θ values inside my confidence set
- Efficiency is more difficult to describe mathematically, it's different from specificity of possibility measures
- Roughly, efficiency to me means that π_Y *tends* to be more tightly concentrated as a function of Y
- Can look at large-sample concentration properties...

Possibilistic IMs, cont.

Theorem (possibilistic Bernstein–von Mises).

Simplest-to-state version: under the usual regularity conditions,

$$\sup_{z \in \text{compact}} \left| \pi_{Y^n}(\hat{\theta}_{Y^n} + J_{Y^n}{}^{-1/2} z) - \gamma(z) \right| \to 0 \quad \text{in } \mathsf{P}_\Theta\text{-probability},$$

where $\gamma(z) = \mathsf{P}(\|Z\|^2 > \|z\|^2)$, Gaussian possibility contour

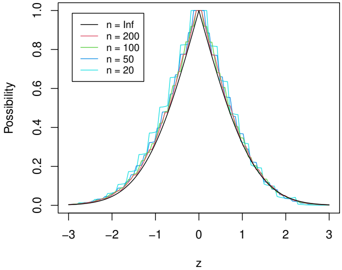
- Key points:[9]
  - IM contour looks like a Gaussian contour for large n
  - concentration determined by the observed Fisher information matrix $J_{y^n} \approx nI_\Theta$, hence efficient
- Case with nuisance parameters covered too
- Computational advances from this came soon after

[9] Details in M. and Williams (*IJAR* 2025), `arXiv:2412.15243`

- Binomial model from before
- Plot shows contour of the standardized parameter

$$Z = \frac{\sqrt{n}(\Theta - \hat{\theta}_y)}{\sqrt{\hat{\theta}_y(1 - \hat{\theta}_y)}}$$

- Looks Gaussian as n increases



Possibilistic IMs, cont.

- Imprecise prob corresponds to sets of probabilities
- This (non-empty) set is called the *credal set*

$$\mathscr{C}(\overline{\Pi}_y) = \{\mathsf{Q}_y \in \mathsf{probs}(\mathbb{T}) : \mathsf{Q}_y(\cdot) \leq \overline{\Pi}_y(\cdot)\}$$

- For possibility measures $\overline{\Pi}_y$, there's a characterization:[10]

$$\mathsf{Q}_y \in \mathscr{C}(\overline{\Pi}_y) \iff \mathsf{Q}_y(\underbrace{\{\theta : \pi_y(\theta) > \alpha\}}_{100(1-\alpha)\%\text{ pl region}}) \geq 1 - \alpha$$

- Elements of $\mathscr{C}(\overline{\Pi}_y)$ are confidence distributions — they assign $1 - \alpha$ probability to $100(1 - \alpha)\%$ confidence sets

[10] e.g., Destercke & Dubois, Ch. 4 of *Intro to IP*

Possibilistic IMs, cont.

Theorem (M., arXiv:2501.10585).

Each member Q_y of $\mathscr{C}(\overline{\Pi}_y)$ can be expressed as

$$\mathsf{Q}_y(\cdot) = \int_0^1 \mathsf{K}_y^\beta(\cdot)\, \mathsf{M}(d\beta)$$

- M is a probability stochastically no smaller than $\mathsf{Unif}(0,1)$
- K_y^β is a probability fully supported on $C_\beta(y)$ for each $\beta \in [0,1]$

- *Inner probabilistic approximation* $\mathsf{Q}_y^\star$ of possibility measure $\overline{\Pi}_y$ satisfies $\mathsf{Q}_y^\star\{\pi_y(\Theta) > \alpha\} = 1 - \alpha$
- Obtain $\mathsf{Q}_y^\star$ via above theorem by taking
  - $\mathsf{M} = \mathsf{Unif}(0,1)$
  - K_y^β fully supported on $\partial C_\beta(y)$ for each $\beta \in [0,1]$

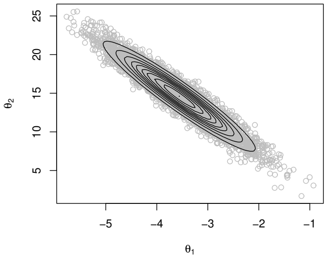
Possibilistic IMs, cont.

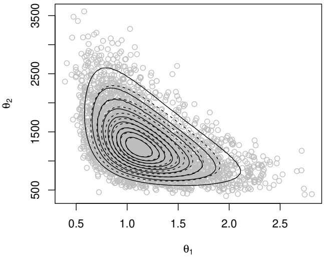
- Inner probabilistic approximation is interesting on its own[11]
- For example, in problems with group invariant structure, $\mathsf{Q}_y^\star$ is the Bayes posterior with right Haar prior
- My interest in it was primarily computational
- Could I use samples from $\mathsf{Q}_y^\star$ to simplify computation of the IM contour π_y in some way?
- Answer is *Yes*, but no time here for details
- The challenge is sampling on $\partial C_\alpha(y)$, but some strategies are given in `arXiv:2501.10585`

[11] M. `arXiv:2503.19748`

Possibilistic IMs, cont.

- Two non-trivial examples
  - logistic regression (left)
  - two-parameter Weibull with right censoring (right)
- Very expensive using the naive computational approach described earlier, but pretty easy with the new approach





Beyond the basics

- I've mostly talked about the "basic inference problem"
  - parametric model/likelihood
  - no nuisance parameters
- Most applications don't match this form!
- But the above framework is general, can be tweaked accordingly to address various practical challenges
- Simplest such "tweak" is to handle nuisance parameters
  - replace relative likelihood by relative *profile* likelihood
  - more efficient[12] than marginalizing the "joint IM"
- Other kinds of conditioning strategies...
- I'll give some details about one important case

[12] M. and Williams (*IJAR* 2025), `arXiv:2412.15243`

Beyond the basics, cont.

ELSEVIER
International Journal of Approximate Reasoning
Volume 150, November 2022, Pages 1-18

Valid inferential models for prediction in supervised learning problems ☆

Leonardo Cella, Ryan Martin

ELSEVIER
International Journal of Approximate Reasoning
Volume 151, December 2022, Pages 205-224

Direct and approximately valid probabilistic inference on a class of statistical functionals

Leonardo Cella [a], Ryan Martin [b]

Possibilistic IM prediction

- Prediction: use data Z^n to predict Z_{n+1}
- Minimal model assumptions, etc
- Predictive possibilistic IM constructed as before:

$$\pi_{z^n}(z_{n+1}) = \sup_{\mathsf{P} \in \mathscr{P}} \mathsf{P}\{R(Z^n, Z_{n+1}) \leq \underbrace{R(z^n, z_{n+1})}_{\text{ranking fn}}\}$$

- Satisfies a corresponding version of strong validity:[13]

$$\sup_{\mathsf{P} \in \mathscr{P}} \mathsf{P}\{\pi_{Z^n}(Z_{n+1}) \leq \alpha\} \leq \alpha, \quad \alpha \in [0, 1]$$

- *But this IM is a pain in the a** to compute!*
- Not necessarily...

[13] Thierry says this property is "too strong" :)

Possibilistic IM prediction, cont.

- Take R invariant to permutations of its first argument
- The set of values $\{z^n, z_{n+1}\}$ is a sufficient statistic
- Conditioning on a sufficient statistic removes dependence on the unspecified $\mathsf{P} \in \mathscr{P}$...[14]

$$\pi_{z^n}(z_{n+1}) = \sup_{\mathsf{P} \in \mathscr{P}} \mathsf{P}\big[R(Z^n, Z_{n+1}) \leq R(z^n, z_{n+1}) \mid \{z^n, z_{n+1}\}\big]$$

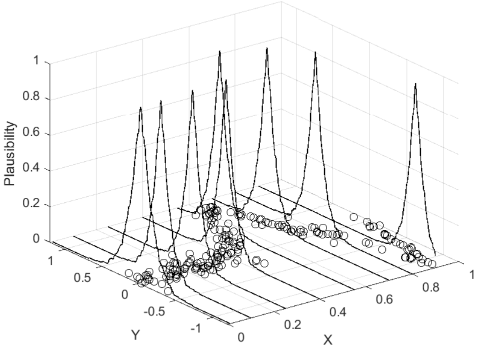
$$\text{sufficiency} \to = \frac{1}{(n+1)!} \sum_{\sigma} 1\{R(z^{\sigma(1:n)}, z_{\sigma(n+1)}) \leq R(z^n, z_{n+1})\}$$

$$\text{invariance} \to = \underbrace{\frac{1}{n+1} \sum_{i=1}^{n+1} 1\{R(z_{-i}^{n+1}, z_i) \leq R(z^n, z_{n+1})\}}_{\text{Vovk \& Shafer's conformal prediction}}$$

[14] Faulkenberry (*JASA* 1973), Hoff (*Bernoulli* 2023)

Possibilistic IM prediction, cont.

- For feature–label pairs $Z = (X, Y)$, IM produces possibility contour functions for Y at each $X = x$



FAQs

- There are a couple questions I'm often asked
  - why consonant structures?
  - you don't employ the full power of the DS calculus — why?
- Below I offer answers to these questions

FAQs: Why consonance?

- Consonant beliefs are very special, extra structure
- In particular, if $(\underline{\Pi}_y, \overline{\Pi}_y)$ is consonant, then

$$\underline{\Pi}_y(H) > 0 \implies \overline{\Pi}_y(H) = 1$$
$$\overline{\Pi}_y(H) < 1 \implies \underline{\Pi}_y(H) = 0$$

- Some might say a consonant IM is "too imprecise"
- So justification is needed for imposing such structure
- I have *two* reasons for this, see below

FAQs: Why consonance?, I

- In general, strong validity is needed to ensure that the IM induces confidence sets with nominal coverage probability
- If strong validity + efficiency are obligatory, then consonance is necessary

Proposition (M., `arXiv:2211.14567`).

For any strongly valid IM with upper prob $\overline{\Gamma}_y$, there exists a strongly valid consonant IM with upper prob $\overline{\Pi}_y$ such that

$$\overline{\Pi}_y(H) \leq \overline{\Gamma}_y(H) \quad \text{for all } H$$

FAQs: Why consonance?, II

- IM's output: *not bounds on a "true" precise prob for* Θ
- So, consonance $\not\Longrightarrow$ loose bounds on this "true" prob
- In fact, tighter bounds wouldn't add value to the IM
- Recall *Cournot's Principle*[15]
  - roughly, only the small & large probability values are meaningful in the real world
  - i.e., suggests what won't & will happen, respectively
- For us, Cournot says *inference* can be made only when

$$\underline{\Pi}_y(H) \text{ is large} \quad \text{or} \quad \overline{\Pi}_y(H) \text{ is small}$$

- That the complementary term is trivial doesn't matter

[15] Shafer & Vovk (2019), e.g., Ch. 10.2

FAQs: Why no DS rules?

- Note: I'm not following any formal rules (e.g., Dempster's)
- that is, $\underline{\Pi}_{(y_1,y_2)}(\cdot) \neq \underline{\Pi}_{y_1}(\cdot) \star \underline{\Pi}_{y_2}(\cdot)$ for any rule $\star$
- This is problematic to some...
- My perspective is different:
  - my top priority is validy + efficiency
  - enforcing a $\star$-identity like above is a constraint
  - can't be more efficient with constraint than without

FAQs: Why no DS rules?, cont.

- Of course, if necessary, I could use such a rule
  - either because I insist on it
  - or if the problem doesn't give me access to the full data (y_1, y_2)
  - e.g., if I only have access to π_{y_1} and π_{y_2}, then I can't directly construct $\pi_{(y_1,y_2)}$ as described above
- Natural idea: use a consonance-preserving rule,[16] e.g.,

$$(\pi_{y_1} \star \pi_{y_2})(\theta) = \frac{\pi_{y_1}(\theta) \circ \pi_{y_2}(\theta)}{\sup_\vartheta \pi_{y_1}(\vartheta) \circ \pi_{y_2}(\vartheta)}, \quad \circ = \text{t-norm}$$

- *Question:* Is this (strongly) valid?

[16] Dubois & Prade (*Comp Intell* 1988); M. & Syring, ISIPTA'19

Partial prior IMs

- There's a spectrum of problems indexed by prior info
- Key observations:
  - to reliably solve real problems, we can't ignore the spectrum
  - spectrum can't be described with ordinary probability
  - applying Bayes to a set of priors is inefficient (dilation)
- These considerations are essential for high-dim problems and cases involving model uncertainty
- Framework above generalizes, but no time to discuss details

Conclusion

> *Perhaps the most important unresolved problem in statistical inference is the use of Bayes theorem in the absence of prior information.* —Efron

- Statistical principles/foundations matter!
- Fisher had the right ideas but...
- Imprecision is imperative for reliable UQ
- New possibilistic IM framework:
  - likelihood-based
  - relatively simple (consonance)
  - achieves strong validity
  - can incorporate partial prior info

Conclusion, cont.

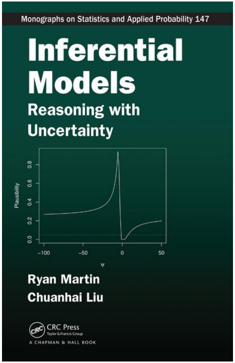
- Things I've more-or-less done but didn't discuss here:
  - formal decision-making
  - valid UQ about uncertain model (predictors in regression)
- Open methodological questions:
  - efficient computation?
  - discrete data problems?
  - partial prior elicitation?
- Open theoretical questions:
  - IM vs generalized Bayes contraction–dilation
  - models constructed with training data?
  - revisiting "impossibility theorems" & paradoxes?
- Open philosophical questions:
  - "what makes IM output meaningful is its properties"?
  - frequentist–Bayesian spectrum?
- Open applications: *all of them!*

References

- Learned a lot since my first book
- *n*-part series of working papers:

  Valid and efficient imprecise-probabilistic inference with partial priors

- (officially) $n = 3$ parts so far
  1. arXiv:2203.06703
  2. arXiv:2211.14567
  3. arXiv:2309.13454
- New review: arXiv:2507.09007
- Working towards a new book...



The end

https://wordpress-courses2223.wolfware.ncsu.edu/st-790-001-fall-2022/

- Papers, talks, etc? www4.stat.ncsu.edu/~rmartin/
- Question, etc? rgmarti3@ncsu.edu

Thanks for your attention!