Discussion of "reIMagined: Probabilistic Approximations of Inferential Models"

Angelos Alexopoulos,
Department of Economics, Athens University of Economics and Business

O'Bayes

Athens, 11/06/2025

Motivation

When prior information is available, Bayesian inference offers major advantages:

- **Stability in small samples:** Prior knowledge can regularize inference, reducing variance and avoiding overfitting.
- **Probabilistic answers:** Posterior distributions allow direct computation of credible intervals and event probabilities (e.g., $\mathbb{P}(\theta > 0 \mid \text{data})$).
- **Coherence and interpretability:** Bayesian updating follows intuitive rules and aligns well with decision-theoretic principles.

Motivation

When prior information is available, Bayesian inference offers major advantages:

- **Stability in small samples:** Prior knowledge can regularize inference, reducing variance and avoiding overfitting.
- **Probabilistic answers:** Posterior distributions allow direct computation of credible intervals and event probabilities (e.g., $\mathbb{P}(\theta > 0 \mid \text{data})$).
- **Coherence and interpretability:** Bayesian updating follows intuitive rules and aligns well with decision-theoretic principles.

But prior knowledge is not always available, for example:

- **Predictions for new disease** (Covid-19)
- **New drug evaluation** (no clinical precedent)
- **First-time market entry** (no historical sales data)
- **Policy pilot studies** (no prior implementation)

Problems when prior is missing

- The Bayesian approach becomes **fragile** or arbitrary (e.g., sensitivity to flat or improper priors).
- Attempts to remain "non-informative" can lead to **paradoxes** or **misleading** results, especially in high dimensions.
- Objective priors are not always justifiable or practically meaningful.

Problems when prior is missing

- The Bayesian approach becomes **fragile** or arbitrary (e.g., sensitivity to flat or improper priors).
- Attempts to remain "non-informative" can lead to **paradoxes** or **misleading** results, especially in high dimensions.
- Objective priors are not always justifiable or practically meaningful.

Ryan's work offers alternative inferential frameworks that retain the **benefits** *of Bayesian reasoning—without requiring prior information.*

Author Background

Ryan is a pioneer in prior-free probabilistic inference.

Selected Publications:

- Martin, R. and Liu, C. (2013). *Inferential models: A framework for prior-free posterior probabilistic inference*. **Journal of the American Statistical Association**, 108(501), 301–313.
- Martin, R. and Liu, C. (2015). *Conditional inferential models: Combining information for prior-free probabilistic inference*. **Journal of the Royal Statistical Society: Series B**, 77(1), 195–217.
- Martin, R. and Liu, C. (2015). *Inferential Models: Reasoning with Uncertainty*. **Monograph**, Chapman & Hall/CRC Press.
- Martin, R. (2021). *On an inferential model construction using generalized associations*. **Journal of Statistical Planning and Inference**, 211, 80–94.
- Martin, R. (2025). *A new Monte Carlo method for valid prior-free possibilistic statistical inference*. Preprint available at `arXiv:2501.10585`.

Topic

- The paper proposes a **novel** framework for prior-free Bayesian inference.
- It constructs an **inner probabilistic approximation** to a valid inferential model (IM).

Key Contributions (1/2)

1. **Reimagined Prior-Free Bayesian Inference**
   - Starts from a data-driven *inferential model* (IM), framed as a **possibility** rather than **probability** measure with exact coverage.
   - Constructs a novel **inner probabilistic approximation**, which inherits many IM properties.

2. **Strong Reliability Guarantees**
   - Posterior-like distributions with exact frequentist coverage.
   - Avoids the pitfalls of default priors and the limitations described by the **false confidence theorem**; assignment of high confidence to false values with arbitrarily high probability.

Key Contributions (2/2)

3. **Agreement with Familiar Approaches (when appropriate)**
   - Recovers Bayes/fiducial solutions in classes of models (group-invariant ones).
   - Maintains connection to traditional inferential results like the Bernstein–von Mises theorem.
4. **Practical Relevance**
   - Introduces a Monte Carlo algorithm for computation.
   - Demonstrates valid and efficient inference in the Behrens–Fisher problem.

Critical Assessment: Theory

Strengths:

- Innovative and mathematically rigorous.
- Addresses longstanding issue of prior misspecification.
- Strong frequentist properties with Bayesian-style inference.

Challenges:

- Are inner approximations unique? If not, guidance on optimality, maybe by judging divergence from the IM contour?
- What are the limits of the approach in high dimensions?
- Theoretical guarantees for approximate sampling?

Critical Assessment: Practice

Strengths:

- Simple illustrations clarify the construction and intuition.
- Potentially more robust than default Bayes in practice.

Questions for Discussion:

- How do we scale this to complex models (e.g., hierarchical models)?
- Can one derive connections to empirical Bayes or variational Bayes?

Conclusion

Takeaway:

- The proposed approach offers a fresh perspective on prior-free inference with strong guarantees.
- Blends Bayesian reasoning with frequentist coverage – without reliance on arbitrary priors.

Promising Directions:

- Flexible, modular, and addresses a major gap in the literature.
- Worth exploring further for complex and real-world problems.
- Study theoretical properties of sampling algorithms.

Thank you!

Technical Appendix

Valid Possibilistic Inference and Confidence Sets

- Given data $X = x$, define the contour:

$$\pi_x(\theta) = \mathsf{P}_\theta\{R(X,\theta) \le R(x,\theta)\}$$

  where $R(x,\theta)$ is the relative likelihood.
- Induced **possibility measure**: $\overline{\Pi}_x(H) = \sup_{\theta \in H} \pi_x(\theta)$
- **Validity property:** $\sup_{\theta \in \mathbb{T}} \mathsf{P}_\theta\{\pi_X(\theta) \le \alpha\} \le \alpha$
- **Confidence Set**

  A set $C_\alpha(x) = \{\theta \in \mathbb{T} : \pi_x(\theta) \ge \alpha\}$ satisfies

$$\mathsf{P}_\theta\{C_\alpha(X) \not\ni \theta\} \le \alpha \quad \text{for all } \theta$$

- So $C_\alpha(x)$ is a $100(1-\alpha)\%$ confidence set for θ.
- IM's **inferential** weight: reject hypothesis H if $\overline{\Pi}_x(H) \le \alpha$ with the frequentist Type I error α.

Inner Probabilistic Approximations

- **Credal Set Characterization:**

$$\mathsf{Q}_x \in \mathscr{C}(\overline{\Pi}_x) \iff \mathsf{Q}_x\{C_\alpha(x)\} \geq 1-\alpha \quad \forall \alpha \in [0,1] \tag{1}$$

  - A credal set $\mathscr{C}(\cdot)$ contains probability measures that represent a decision maker's beliefs under uncertainty
  - Each Q_x gives at least $1-\alpha$ probability to the confidence region $C_\alpha(x)$.
  - These distributions are known as *confidence distributions*.

- **Mixture Characterization (Martin, 2025):**

$$\mathsf{Q}_x(\cdot) = \int_0^1 \mathsf{K}_x^\beta(\cdot)\,\mathsf{M}_x(d\beta) \tag{2}$$

where:

  - K_x^β is supported on $C_\beta(x)$.
  - M_x is a mixing measure stochastically larger than $\mathsf{Unif}(0,1)$.

Can we find $\mathsf{Q}_x^\star$ s.t (1) becomes equality for each $\alpha \in [0,1]$?

Inner probabilistic approximations

Yes, this is the **inner probabilistic approximation**

Inner probabilistic approximations

Yes, this is the **inner probabilistic approximation**

- A **Canonical Approximation** of the form

$$\mathsf{Q}_x^\star\{C_\alpha(x)\} = 1 - \alpha \quad \forall \alpha \in [0,1],$$

  is derived by making the following choices in (2)

  - Set $\mathsf{M}_x = \mathsf{Unif}(0,1)$ and support K_x^β on the boundary $\partial C_\beta(x)$.
  - A practical choice: $\mathsf{K}_x^\beta = \mathsf{Unif}(\partial C_\beta(x))$.

Sampling from the approximation

Monte Carlo sampling from a sort of ‘posterior distribution” reduces to sampling from $Q^{\star}_{x}$

Sampling from the approximation

Monte Carlo sampling from a sort of ‘posterior distribution”
reduces to sampling from $\mathsf{Q}_x^\star$

- **Sampling Algorithm:**
  1. Sample $A \sim \mathsf{Unif}(0, 1)$.
  2. Sample $\Theta \sim \mathsf{Unif}(\partial C_A(x))$.
- **Result:** Θ are samples from $\mathsf{Q}_x^\star$, a valid probabilistic approximation inside $\mathscr{C}(\overline{\Pi}_x)$.

Possibilistic IM and Group Invariance

- Let $\mathcal{G}$ be a group of transformations on the sample space $\mathcal{X}$.
- The statistical model $\{\mathbb{P}_\theta : \theta \in \Theta\}$ is **group invariant** if:

$$\mathbb{P}_\theta(gX \in \cdot) = \mathbb{P}_{\bar{g}\theta}(X \in \cdot), \quad \forall g \in \mathcal{G}, \theta \in \Theta.$$

- Examples include location, scale and permutation models.
- In such models, the right Haar prior yields a **standard no-prior Bayes posterior**.
- In group invariant models, the inner probabilistic approximation of the possibilistic IM coincides with the Bayesian posterior under the right Haar prior (Martin, 2023).
- **Implication:** Possibilistic IMs recover fiducial/Bayes posteriors in invariant settings and explain properties like exact probability matching.

Asymptotic Normality and Efficiency

- Possibilistic IMs satisfy a Bernstein–von Mises-type result:

$$\pi_{x^n}(\theta) \approx 1 - F_D\big((\hat{\theta}_{x^n} - \theta)^\top J_{x^n}(\hat{\theta}_{x^n} - \theta)\big).$$

- This is the **Gaussian possibility contour** centered at MLE $\hat{\theta}_{x^n}$ with covariance $J_{x^n}^{-1}$.
- **Theorem:** Under classical regularity conditions, the possibilistic IM converges to a Gaussian possibility distribution.
- **Implication:** The inner probabilistic approximation of the IM converges to the usual asymptotic Bayes posterior $\mathcal{N}(\hat{\theta}_{x^n}, J_{x^n}^{-1})$.
- Matches the Cramér–Rao lower bound $\Rightarrow$ **Efficient**.

Naive Strategy for IM Computation

- **Naive Strategy:** Approximate IM contour by Monte Carlo:

$$\pi_x(\theta) \approx \frac{1}{M} \sum_{m=1}^{M} 1\{R(X_{m,\theta}, \theta) \leq R(x, \theta)\}$$

  - $X_{m,\theta} \sim \mathsf{P}_\theta$, independent samples.
  - Requires dense evaluation over $\mathbb{T}$—computationally intensive.

Efficient Strategy for Inner Probabilistic Approximations

- **Efficient Strategy:**
  - Sample from the inner probabilistic approximation $\mathsf{Q}_x^\star$ using elliptical confidence sets.
  - Construct Gaussian-shaped sets:

$$C_\alpha^\xi(x) = \left\{\theta : (\theta - \hat{\theta}_x)^\top J_x^\xi(\theta - \hat{\theta}_x) \leq \chi^2_{D,1-\alpha}\right\}$$

  - Embellished Fisher information: $J_x^\xi = E\,\mathrm{diag}(\xi^{-1})\Lambda\mathrm{diag}(\xi^{-1})E^\top$.
- **Implementation:**
  - Choose $\xi = \xi(x, \alpha)$ such that $C_\alpha(x) \subseteq C_\alpha^\xi(x)$.
  - Sample from $\mathsf{K}_x^\alpha = \mathrm{Unif}(\partial C_\alpha^{\xi(x,\alpha)}(x))$.
- **Result:** Practical, conservative implementation of $\mathsf{Q}_x^\star$ with theoretical guarantees.

Behrens–Fisher Problem

The Behrens–Fisher problem involves independent samples of size n_1 and n_2 from two distinct normal populations:

$$X_1 \sim \mathcal{N}(\Theta_{11}, \Theta_{12}^2), \quad X_2 \sim \mathcal{N}(\Theta_{21}, \Theta_{22}^2)$$

The goal is to perform marginal inference on the difference of the two means:

$$\Phi = m(\Theta) = \Theta_{21} - \Theta_{11}$$

The problem is straightforward when the variances are known or their ratio is known. However, the fully unknown variance case remains elusive, with various solutions proposed but no consensus on the "best" approach.