Welcome to the post-Bayesian seminar!

When: every 2 weeks @ Tuesday either 9:00 or 14:00 GMT

Structure:
Chapter 1: Generalised Bayes (11/02–22/04)
Chapter 2: Predictive Bayes (06/05–15/07)
Chapter 3: PAC-Bayes (after the summer break)

Organisers:

Prof. Pierre Alquier (ESSEC Singapore)

Prof. Jeremias Knoblauch (UCL)

Yann McLatchie (UCL)

Matias Altamirano (UCL)

Welcome to the post-Bayesian seminar!

Important Links

At a glance/website: https://tinyurl.com/postBayesWebsite
Where to subscribe to mailing list: https://tinyurl.com/postBayesSubscribe
Where to subscribe to calendar: https://tinyurl.com/postBayesCalendar
Where to attend the seminars: https://tinyurl.com/postBayesZoom
Where recorded seminars are stored: https://tinyurl.com/postBayesYT
Where to register for the workshop: https://tinyurl.com/postBayesWorkshop

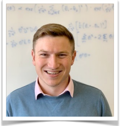
Please share widely! :)

Welcome to the post-Bayesian seminar!

Satellite Workshop @ BayesComp 2025:

Bayesian Computation and Inference with Misspecified Models

François-Xavier Briol
(UCL)



Jack Jewson
(Monash University)

Jeremias Knoblauch
(UCL)

https://postbayes.github.io/BayesMisspecificationSatellite/

Welcome to the post-Bayesian seminar!

Questions / Comments during talks

During talk:

- use Q/A function in zoom
- Other questions can be upvoted
- We will try to monitor questions and ask relevant ones in natural breaks

After talk:

- Raise your hand in zoom
- We will do our best to decide who gets to ask a question fairly
- We will do our best to resolve remaining questions in Q / A function

Problematic Assumptions for Bayesian Analysis

I

(A1)	model well-specified
(A2)	prior well-specified
(A3)	computationally feasible

(A1) $x_{1:n} \sim p(x_{1:n} \mid \theta^*)$ for some $\theta^* \in \Theta$

Θ = Only relevant State of the world

(A2) $\pi(\theta) =$ uncertainty about the true State of the world

How rational decision-makers choose the prior

(A3) $\pi_n(\theta \mid x_{1:n})$ computable in practice

Guarantees real-world relevance

Problematic Assumptions for Bayesian Analysis

I

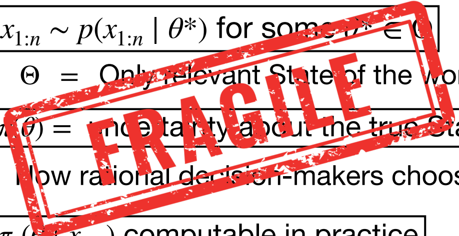
(A1) model well-specified
(A2) prior well-specified
(A3) computationally feasible

(A1) $x_{1:n} \sim p(x_{1:n} \mid \theta^*)$ for some $\theta^* \in \Theta$

Θ = Only relevant State of the world

(A2) $\pi(\theta)$ = uncertainty about the true State of the world

How rational decision-makers choose the prior



(A3) $\pi_n(\theta \mid x_{1:n})$ computable in practice

Guarantees real-world relevance

Optimisation-centric posteriors / Generalised Variational Inference

~~(A1)~~, ~~(A2)~~, ~~(A3)~~

$$q_n^*(\theta) = \arg\min_{q \in \mathcal{Q}} \Big\{ \underbrace{\mathcal{L}(q, x_{1:n})}_{\text{Data-fitting}} + \underbrace{\mathrm{D}(q, \pi)}_{\text{Prior regularisation}} \Big\};$$

$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$

Gibbs/Generalised/ Pseudo Posterior

~~(A1)~~, (A2), (A3)

$$\pi_n^{\mathrm{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathrm{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathrm{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Power/Fractional/ Cold Posterior

~~(A1)~~, (A2), (A3)

$$\pi_n^{(\lambda)}(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta) d\theta}$$

Prof. Jeremias Knoblauch (UCL)

Chapter 1

Generalised Bayes (11/02—22/04)

Martingale posteriors & resampling-based approaches ~~(A1)~~, ~~(A2)~~, ~~(A3)~~

For $i = 1,2,\dots$

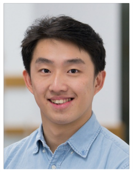
$$X_{n+i+1} \sim p(X_{n+i} \mid x_{1:n}, X_{n+1:n+i})$$

$$\theta^{\infty} = \mathrm{argmin}_{\theta\in\Theta} \mathsf{L}\left([x_{1:n}, X_{n+1:\infty}], \theta\right)$$

Chapter 2

Predictive Bayes

(06/05—15/07)



Dr. Edwin Fong

(HKU)

Optimisation-centric posteriors / Generalised Variational Inference

~~(A1)~~, ~~(A2)~~, ~~(A3)~~

$$q_n^*(\theta) = \arg\min_{q \in \mathcal{Q}} \left\{ \underbrace{\mathscr{L}(q, x_{1:n})}_{\text{Data-fitting}} + \underbrace{\mathrm{D}(q, \pi)}_{\text{Prior regularisation}} \right\};$$

$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$

Gibbs/Generalised/ Pseudo Posterior

~~(A1)~~, (A2), (A3)

$$\pi_n^{\mathrm{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathrm{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathrm{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Chapter 3

PAC-Bayes (after summer break)

Prof. Pierre Alquier (ESSEC Singapore)

Structure of Chapter 2

20/05: Theoretical foundations of predictive Bayes
(Prof. Sandra Fortini)

03/06: Predictive model selection and uncertainty
(Vik Shirvaikar)

01/07: Recursive methods for predictive Bayes
(Prof. Lorenzo Cappello)

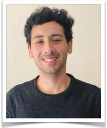
15/07: Applications of post-Bayesian methods
(Dr. Harita Dellaporta + Matias Altamirano)

20/05

03/06

01/07

05/07



Chapter 2: Predictive Bayes
Introduction and Overview

Edwin Fong

The University of Hong Kong

Post-Bayes Seminar 2025

1 Introduction
- Predictive Bayes in a nutshell
- History of predictive Bayes
- Predictive resampling
- A parametric example

2 The predictive Bayes framework

3 Bayesian bootstrap

4 Predictive models

5 Conclusions

Traditional Bayes in a Nutshell

The traditional Bayesian specifies the sampling density/likelihood $p(y \mid \theta)$ and the prior $\pi(\theta)$ as their model.

Given observations $y_{1:n}$, the Bayesian obtains:

1. The posterior for **parameter inference**:

$$\pi(\theta \mid y_{1:n}) \propto \prod_{i=1}^{n} p(y_i \mid \theta)\, \pi(\theta)$$

2. The marginal likelihood for **model selection**:

$$p(y_{1:n}) = \int \prod_{i=1}^{n} p(y_i \mid \theta)\, \pi(\theta)\, d\theta$$

3. The posterior predictive density for **prediction**:

$$p(y_{n+1} \mid y_{1:n}) = \int p(y_{n+1} \mid \theta)\, \pi(\theta \mid y_{1:n})\, d\theta.$$

Predictive Bayes in a Nutshell

The predictive Bayesian *reverses* the order, and directly specifies the predictive density $p(y_{n+1} \mid y_{1:n})$ as their model.

Given observations $y_{1:n}$, the predictive Bayesian obtains:

1. The posterior for **parameter inference**:

$$Y_{n+1:\infty} \sim \prod_{i=n+1}^{\infty} p\left(y_i \mid y_{1:i-1}\right), \quad \theta = \theta\left(Y_{n+1:\infty}\right)$$

2. The marginal likelihood for **model selection**:

$$p(y_{1:n}) = \prod_{i=1}^{n} p(y_i \mid y_{1:i-1})$$

3. The implicit likelihood $p(y \mid \theta)$ and prior $\pi(\theta)$ (sometimes)

Predictive Bayes in a Nutshell

Why is working with the predictive $p(y_{n+1} \mid y_{1:n})$ directly desirable?

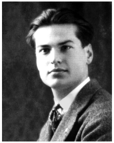
1. We have gotten pretty good at eliciting predictive distributions $p(y_{n+1} \mid y_{1:n})$, e.g. with machine learning.
2. Predictive statements can be validated as data y is actually observed, unlike probability statements on θ (like the prior).
3. Sometimes the computation required for posterior inference can be much more expedient with the predictive approach.

Some History on Predictive Bayes

Predictive Bayes has a long history, dating all the way back to de Finetti.

BRUNO DE FINETTI

Foresight: Its Logical Laws, Its Subjective Sources
(*1937*)



De Finetti's Representation Theorem: An infinitely exchangeable binary sequence $Y_i \in \{0, 1\}$ has the representation

$$p\left(Y_1, \ldots, Y_N\right) = \int_0^1 \left[\prod_{i=1}^{N} \theta^{Y_i} \left(1-\theta\right)^{1-Y_i}\right] \pi\left(\theta\right)\, d\theta$$

Some History on Predictive Bayes

Furthermore, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} Y_i = \Theta$$

where $\Theta \sim \pi$. De Finetti identified that the parameter Θ and the prior π can be written as a function of only binary exchangeable observables!

An insightful excerpt from [Bernardo and Smith, 2009, Chapter 4.9]:

$$\lim_{(n-m)\to\infty} P\left(\frac{y_{n-m}}{(n-m)} \leq \theta \,\middle|\, x_1, \ldots, x_m\right) = Q(\theta \mid x_1, \ldots, x_m).$$

Thus, ***a posterior distribution for a parameter is seen to be a limiting case of a posterior*** **(*****conditional*****)** ***predictive distribution for an observable.***

Some History on Predictive Bayes

20/05

Other well-known proponents of the predictive approach include Phil Dawid [Dawid, 1984], who coined the **pre-quential** approach, and Seymour Geisser [Geisser, 1993].

There has been a recent resurgence in interest of the predictive approach:

JOURNAL ARTICLE

Quasi-Bayes Properties of a Procedure for Sequential Learning in Mixture Models FREE

Sandra Fortini , Sonia Petrone ✉

Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 82, Issue 4, September 2020, Pages 1087–1114, https://doi.org/10.1111/rssb.12385

Published: 29 June 2020 **Article history** ▾

February 2021

A class of models for Bayesian predictive inference

Patrizia Berti, Emanuela Dreassi, Luca Pratelli, Pietro Rigo

Bernoulli 27(1): 702-726 (February 2021). DOI: 10.3150/20-BEJ1255

JOURNAL ARTICLE

Martingale posterior distributions

Edwin Fong , Chris Holmes ✉ , Stephen G Walker Author Notes

Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 85, Issue 5, November 2023, Pages 1357–1391, https://doi.org/10.1093/jrsssb/qkad005

Published: 02 February 2024 **Article history** ▾

Bayesian Uncertainty

Statistical uncertainty arises as only a finite sample $y_{1:n}$ is observed. Given the entire population $y_{1:\infty}$, the parameter $\theta_\infty = \theta(y_{1:\infty})$ would be known precisely.

We argue that the source of Bayesian uncertainty is precisely the unobserved remainder of the population, $y_{n+1:\infty}$.

Given $Y_{1:n} = y_{1:n}$, we view Bayesian posterior sampling as:

1. **Impute the population:** $Y_{n+1:\infty} \sim p(y_{n+1:\infty} \mid y_{1:n})$
2. **Compute the parameter:** $\theta_\infty = \theta(Y_{1:\infty})$

Doob's consistency theorem [Doob, 1949] shows that $\theta_\infty \sim \pi(\theta \mid y_{1:n})$. The Bayesian imputes what they need to know the parameter.

Posterior Sampling

Table 1: Observed Sample

Unit	A	B	C
1	10	5	8
$n = 2$	6	7	10
3	?	?	?
4	?	?	?
⋮	⋮	⋮	⋮
N	?	?	?

Predictive Resampling $\Longrightarrow$

Table 2: Imputed Population 1

Unit	A	B	C
1	10	5	8
2	6	7	10
3	**4**	**20**	**12**
4	**12**	**12**	**18**
⋮	⋮	⋮	⋮
N	**19**	**15**	**12**

Predictive Resampling $\Longrightarrow$

Table 3: Imputed Population 2

Unit	A	B	C
1	10	5	8
2	6	7	10
3	**6**	**18**	**13**
4	**10**	**9**	**21**
⋮	⋮	⋮	⋮
N	**15**	**12**	**16**

Imputation with Predictive Resampling

Denoting $p_i = p(y_{i+1} \mid y_{1:i})$, we have the sequential imputation algorithm to draw from $p(y_{n+1:N} \mid y_{1:n}) = \prod_{i=n+1}^{N} p(y_i \mid y_{1:i-1})$:

Algorithm 1: Predictive Resampling (PR)

Compute p_n from the observed data $y_{1:n}$
$N > n$ is a large integer
for $j \leftarrow 1$ **to** B **do**
  for $i \leftarrow n+1$ **to** N **do**
    Sample $Y_i \sim p_{i-1}$
    Update $p_i \hookleftarrow \{p_{i-1}, Y_i\}$
  end
  Evaluate $\theta_N^{(j)} = \theta(Y_{1:N})$
end

Parametric Example

Example

Let $f_\theta(y) = \mathcal{N}(y \mid \theta, 1)$, with $\pi(\theta) = \mathcal{N}(\theta \mid 0, 1)$. The posterior density takes the form $\pi(\theta \mid y_{1:n}) = \mathcal{N}(\theta \mid \bar{\theta}_n, \bar{\sigma}_n^2)$ where

$$\bar{\theta}_n = \frac{\sum_{i=1}^n y_i}{n+1}, \quad \bar{\sigma}_n^2 = \frac{1}{n+1}.$$

The posterior predictive density is then

$$p(y_{n+1} \mid y_{1:n}) = \mathcal{N}(y_{n+1} \mid \bar{\theta}_n, 1 + \bar{\sigma}_n^2).$$

Predictive resampling (PR):

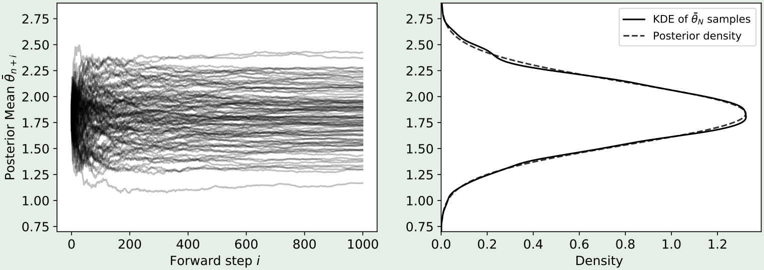
1. Draw $y_{n+1} \sim \mathcal{N}(y_{n+1} \mid \bar{\theta}_n, 1 + \bar{\sigma}_n^2)$
2. Update $\bar{\theta}_{n+1} = \sum_{i=1}^{n+1} y_i/(n+2), \quad \bar{\sigma}_{n+1}^2 = 1/(n+2)$
3. Repeat until N

Parametric Example

Example

We generated $y_{1:n} \overset{\text{iid}}{\sim} \mathcal{N}(y \mid \theta = 2, 1)$ for $n = 10$, giving $\bar{\theta}_n = 1.84$.

Doob's theorem: distribution of $\bar{\theta}_N$ is approximately $\pi(\theta \mid y_{1:n})$.



(a) (b)

Figure 1: (a) Sample paths of $\bar{\theta}_{n+i}$ through forward sampling; (b) Kernel density estimate of $\bar{\theta}_N$ samples (——) and analytical posterior density $\pi(\theta \mid y_{1:n})$ (- - -)

1 Introduction

2 The predictive Bayes framework
- Martingale posteriors
- Parameters
- Theoretical foundations
- Model evaluation

3 Bayesian bootstrap

4 Predictive models

5 Conclusions

Martingale Posterior: A Predictive Framework

Consider more general $\{p(y_i \mid y_{1:i-1})\}_{i=n+1,n+2,\dots}$ and $\theta_\infty = \theta(Y_{1:\infty})$. No need for likelihood/prior, but predictives must satisfy a *martingale* condition.

Step 1: Predictive resampling

- Sequentially draw $Y_{n+1} \sim p(y_{n+1} \mid y_{1:n})$, $Y_{n+2} \sim p(y_{n+2} \mid y_{1:n+1})$, ... until we have $Y_{1:\infty}$, noting $Y_{1:n} = y_{1:n}$ is fixed.

Step 2: Recover parameter of interest

- Compute parameter of interest:

$$\theta_\infty = \theta(Y_{1:\infty}),$$

  e.g. the mean or median of $Y_{1:\infty}$.

We call the distribution of θ_∞ the *martingale posterior*.

Parameter of Interest

For $y_{1:n} \stackrel{\text{iid}}{\sim} F_0$, a more general parameter of interest is defined:

$$\theta(F_0) = \arg\min_{\theta} \int \ell(\theta, y) dF_0(y)$$

- For example, $\ell(\theta, y) = |y - \theta|$ gives the median and $(y - \theta)^2$ gives the mean.
- For model fitting, let $\ell(\theta, y) = -\log f_\theta(y)$, where f_θ is the likelihood or pseudolikelihood

Martingale Posterior Distributions

To summarize:

1. Impute $Y_{n+1:\infty}$ from joint predictive:

$$p(y_{n+1:\infty} \mid y_{1:n}) = \prod_{i=n+1}^{\infty} p(y_i \mid y_{1:i-1})$$

2. Compute $\theta_\infty = \theta(Y_{1:\infty})$

Choice of predictive $p(y_{n+1} \mid y_{1:n})$:

- Posterior predictive: Bayesian posterior
- Empirical distribution: Bayesian bootstrap
- General predictive distribution: *martingale posterior*
  - Usually required to satisfy a martingale condition
  - We will cover a few parametric/nonparametric examples later in this talk!

Theoretical Foundations

20/05

Our sequence of predictive distributions $p(y_i \mid y_{1:i-1})$ should not be chosen arbitrarily. What conditions should the predictive distribution satisfy?

A *martingale condition** is required in [Berti et al., 2020], [Fortini and Petrone, 2020] and [Fong et al., 2023].

An excellent review of the theoretical foundations can be found in [Fortini and Petrone, 2025].

2025

Exchangeability, Prediction and Predictive Modeling in Bayesian Statistics

Sandra Fortini, Sonia Petrone

Author Affiliations +

Statist. Sci. 40(1): 40-67 (2025). DOI: 10.1214/24-STS965

*Known as conditionally identically distributed (c.i.d.).

Model Evaluation

03/06

There are many choices for the predictive $p(y_i \mid y_{1:i-1})$ which may be a good fit to the observations $y_{1:n}$.

How do we evaluate and choose between predictive models? How do we obtain model uncertainty?

JOURNAL ARTICLE

Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach

A. P. Dawid

Journal of the Royal Statistical Society. Series A (General), Vol. 147, No. 2, The 150th Anniversary of the Royal Statistical Society (1984), pp. 278-292 (15 pages)

JOURNAL ARTICLE

Strictly Proper Scoring Rules, Prediction, and Estimation

Tilmann Gneiting, Adrian E. Raftery

Journal of the American Statistical Association, Vol. 102, No. 477 (Mar., 2007), pp. 359-378 (20 pages)

1 Introduction

2 The predictive Bayes framework

3 Bayesian bootstrap
- The empirical predictive
- Bayes vs frequentism

4 Predictive models

5 Conclusions

The Empirical Predictive

Let us elicit the simplest nonparametric predictive, the empirical distribution:

$$p_n(y_{n+1}) = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}.$$

Predictive resampling involves repeating:

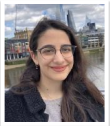
1. Resample $y_{n+1} \sim \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$
2. Update $p_{n+1}(\cdot) = \frac{1}{n+1} \sum_{i=1}^{n+1} \delta_{y_i}$

The drawn $y_{n+1:\infty}$ will be repeats of $y_{1:n}$, i.e. a Pólya urn giving

$$F_\infty := \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i} = \sum_{i=1}^{n} w_i \delta_{y_i}$$

with $w_{1:n} \sim \text{Dirichlet}(1, \dots, 1)$.

The Bayesian Bootstrap

15/07

The key connection: the martingale posterior with the empirical distribution is equivalent to the Bayesian bootstrap [Rubin, 1981].

1. Predictive resampling:

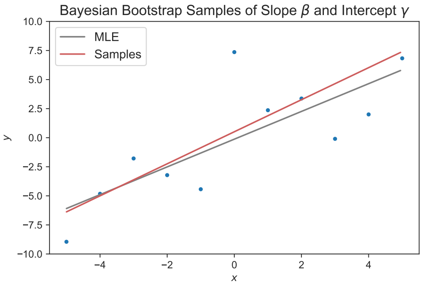
$$w_{1:n} \sim \text{Dirichlet}(1, \ldots, 1), \quad F_\infty = \sum_{i=1}^{n} w_i \delta_{y_i}$$

2. Compute parameter:

$$\theta_\infty = \theta\left(F_\infty\right)$$

- Good properties under model misspecification
- Computationally fast and parallelizable compared to MCMC, can handle multimodality

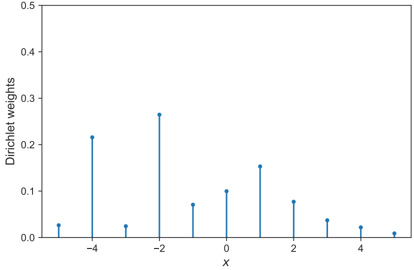
Bayesian Bootstrap for the Linear Model

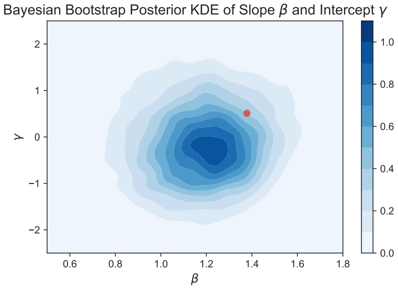


For a simple linear model

$$\ell(\beta, \gamma, y, x) = (y - \{\beta x + \gamma\})^2$$

sample $\left(\beta^{(j)}, \gamma^{(j)}\right)$ from Bayesian bootstrap.





Bayesian and Frequentist Uncertainty

Frequentist bootstrap:

1. Draw $[Y^*_{1:n} \mid y_{1:n}] \overset{\text{iid}}{\sim} \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$, giving us a random dataset $Y^*_{1:n}$
2. Compute $\theta(Y^*_{1:n})$

Bayesian bootstrap:

1. Draw $[Y_{n+1:\infty} \mid y_{1:n}] \sim P(\cdot \mid y_{1:n})$ from *joint* predictive of the empirical distribution, giving us a random complete dataset $Y_{1:\infty}$
2. Compute $\theta(Y_{1:\infty})$

- Bayesians consider uncertainty in $Y_{n+1:\infty}$ and estimand θ_0; frequentists consider uncertainty in $Y_{1:n}$ and estimator $\hat{\theta}$.
- Both methods only specify the empirical distribution, and resample.
- No need for a prior distribution to define posterior.

1 Introduction

2 The predictive Bayes framework

3 Bayesian bootstrap

4 Predictive models
- Parametric predictives
- Nonparametric predictives

5 Conclusions

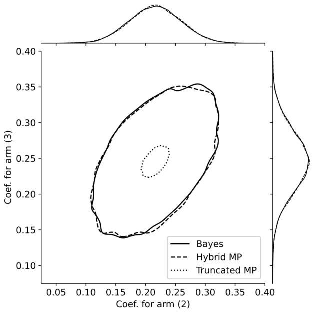
Parametric Predictives

We can utilize a plug-in parametric predictive density, where our recursive update is based on stochastic gradient descent [Holmes and Walker, 2023]

$$
\begin{aligned}
y_{i+1} &\sim p(y_{i+1} \mid y_{1:i}) = p_{\theta_i}(y_{i+1}) \\
\theta_{i+1} &= \theta_i + (i+1)^{-1} \nabla_\theta \log p_{\theta_i}(y_{i+1})
\end{aligned}
$$

- The plug-in predictive $p_{\theta_i}(y_{i+1})$ replaces the posterior predictive
- As the score function has mean zero under the model, our parameter is a martingale
- Allows for prior-free parametric posteriors without MCMC

Parametric Predictives



Student-t regression example :
Bayes: 2 min for MCMC
Parametric MP: 0.03 sec for PR [Fong and Yiu, 2024a]

- We can extend this to a semiparametric predictive (work in progress):

$$p(y_{i+1} \mid y_{1:i}) = \frac{c}{c+i}\, p_{\theta_i}(y_{i+1}) + \frac{1}{c+i} \sum_{j=1}^{i} \delta_{y_j}$$

Nonparametric Recursive Updates

01/07

The Bayesian bootstrap, while intuitive, only returns a discrete F_∞. We want a more general recipe to elicit a predictive distribution that:

- Satisfies the martingale property
- Has a continuous density and is nonparametric
- Utilizes recursion for computational ease

For $p_i(y) = p(y \mid y_{1:i})$, consider the recursive update

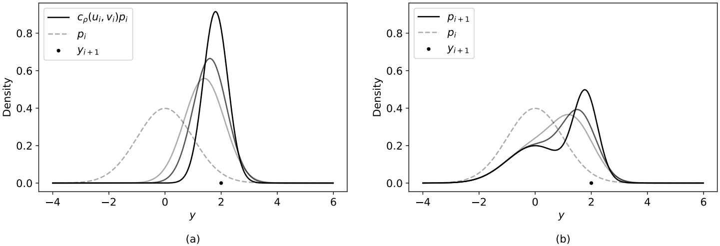
$$\{p_i(y), y_{i+1}\} \to p_{i+1}(y).$$

One can look to Bayesian nonparametric mixture models for inspiration, e.g. [Newton and Raftery, 1994, Hahn et al., 2018].

Nonparametric Recursive Updates: Copulas

As an example, an online Bayesian kernel density estimate can be constructed using copulas:

$$p_{i+1}(y) = \frac{i}{i+1}\, p_i(y) + \frac{1}{i+1} \underbrace{k_i(y, y_{i+1})}_{\text{Copula kernel}}$$



Can be extended to conditional density estimate $p_i(y \mid x)$ for regression!

Nonparametric Recursive Updates: Copulas

Predictive resampling can be very expedient compared to traditional MCMC.

Copula (GPU): 0.5 seconds for p_n, 2 seconds for PR
DPMM (CPU): 25 seconds for Gibbs sampling

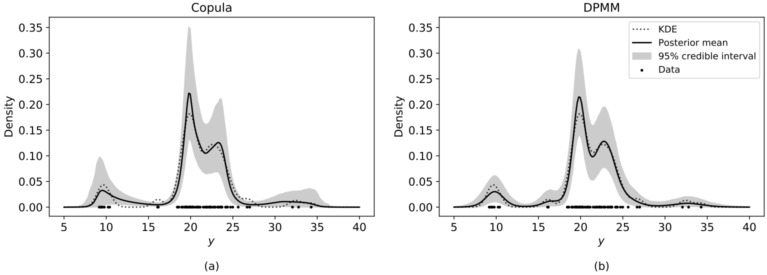


Figure 2: Martingale posterior for the density

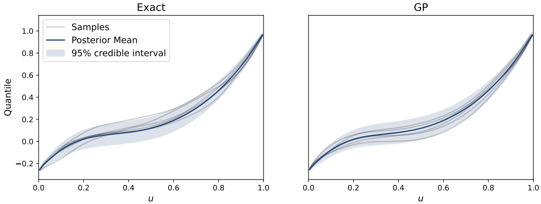
Nonparametric Recursive Updates: Quantiles

Nonparametric quantile function estimate acts as a *generative* predictive:

$$Y_{n+1} \sim Q_n(V) \quad \text{for } V \sim \mathcal{U}(0,1)$$

where $Q_{i+1}(u) = g_i(Q_i(u), Y_{i+1})$ is a recursive update.

This gives the *quantile martingale posterior* [Fong and Yiu, 2024b].



- Montonicity is guaranteed due to the imputation step
- Theory relies on function-valued martingales to show posterior support, consistency, contraction rate, etc. No longer c.i.d.!

1 Introduction

2 The predictive Bayes framework

3 Bayesian bootstrap

4 Predictive models

5 Conclusions

Conclusions

Foundations:

- Bayesian inference is about *imputing* $Y_{n+1:\infty}$ with $p(y_{n+1:\infty} \mid y_{1:n})$, which induces uncertainty on $\theta(Y_{1:\infty})$
- Bootstrap interpretation is insightful: Bayesian uncertainty arises from $Y_{n+1:\infty}$, whereas frequentist arises from $Y_{1:n}$

Methodology:

- The predictive Bayesian approach involves specifying the predictive distribution directly as the statistical model.
- We can generalize Bayes to the *martingale posterior* by considering other predictive distributions - no need for likelihood nor prior.

Conclusions

Strengths:

- (Almost) exact posterior sampling can be carried out without MCMC, which offers potentially large computational speed-ups.
- It is possible to carry out Bayesian inference without explicitly specifying a prior $\pi(\theta)$.
- Can be robust to model misspecification (but not always).

Weaknesses:

- Martingale condition is restrictive: can we relax it and incoporate machine learning?
- Incoporating structure into the predictive (e.g. hierarchy, dependence) is difficult without going through the likelihood-prior machinery.
- Theoretical properties are harder to show for predictive Bayesian methods.

Thank you!

References I

[Bernardo and Smith, 2009] Bernardo, J. and Smith, A. (2009).
Bayesian Theory.
Wiley Series in Probability and Statistics. Wiley.

[Berti et al., 2020] Berti, P., Dreassi, E., Pratelli, L., and Rigo, P. (2020).
A class of models for Bayesian predictive inference.
Bernoulli, 27(1):702–726.

[Dawid, 1984] Dawid, A. P. (1984).
Present position and potential developments: Some personal views statistical theory the prequential approach.
Journal of the Royal Statistical Society: Series A (General), 147(2):278–290.

[Doob, 1949] Doob, J. L. (1949).
Application of the theory of martingales.
Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 Juin–3 Juillet 1948), Paris CNRS, 23–27.

[Fong et al., 2023] Fong, E., Holmes, C., and Walker, S. G. (2023).
Martingale posterior distributions.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 85(5):1357–1391.

References II

[Fong and Yiu, 2024a] Fong, E. and Yiu, A. (2024a).
Asymptotics for parametric martingale posteriors.
arXiv preprint arXiv:2410.17692.

[Fong and Yiu, 2024b] Fong, E. and Yiu, A. (2024b).
Bayesian quantile estimation and regression with martingale posteriors.
arXiv preprint arXiv:2406.03358.

[Fortini and Petrone, 2020] Fortini, S. and Petrone, S. (2020).
Quasi-Bayes properties of a procedure for sequential learning in mixture models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4):1087–1114.

[Fortini and Petrone, 2025] Fortini, S. and Petrone, S. (2025).
Exchangeability, prediction and predictive modeling in bayesian statistics.
Statistical Science, 40(1):40–67.

[Geisser, 1993] Geisser, S. (1993).
Predictive inference, volume 55.
CRC press.

[Hahn et al., 2018] Hahn, P. R., Martin, R., and Walker, S. G. (2018).
On recursive Bayesian predictive distributions.
Journal of the American Statistical Association, 113(523):1085–1093.

References III

[Holmes and Walker, 2023] Holmes, C. C. and Walker, S. G. (2023).
Statistical inference with exchangeability and martingales.
Philosophical Transactions of the Royal Society A, 381(2247):20220143.

[Newton and Raftery, 1994] Newton, M. and Raftery, A. (1994).
Approximate Bayesian inference by the weighted likelihood bootstrap.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 56:3 – 48.

[Rubin, 1981] Rubin, D. B. (1981).
The Bayesian bootstrap.
The Annals of Statistics, 9(1):130–134.