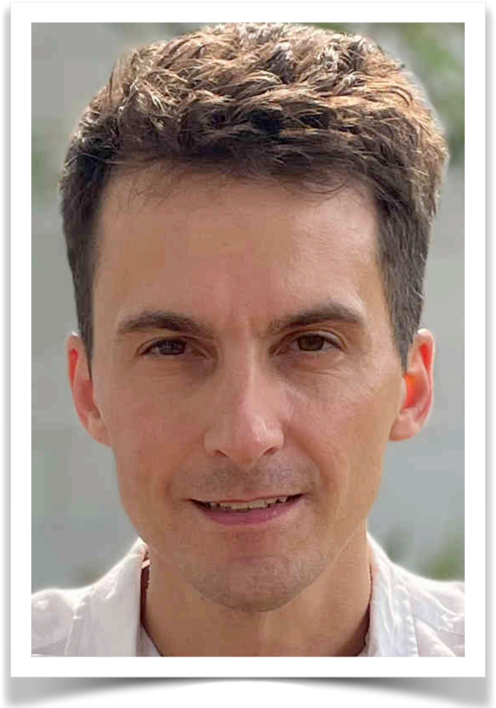
Welcome to the post-Bayesian seminar!

When: every 2 weeks @ Tuesdays either 9 AM (9:00) or 2 PM (14:00) GMT

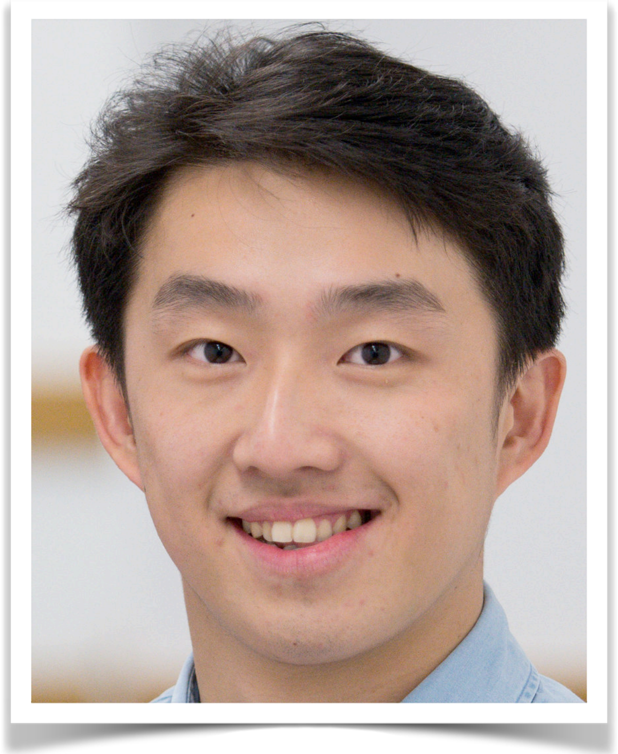
Structure:

Chapter 1: Generalised Bayes (11/02—22/04)
Chapter 2: Resampling & Martingale Posteriors (06/05—15/07)
Chapter 3: PAC-Bayes (after the summer break)

Organisers:



Prof. Pierre Alquier (ESSEC Singapore)



Dr. Edwin Fong (University of Hong Kong)

Matias Altamirano (UCL)

Yann McLatchie (UCL)

Welcome to the post-Bayesian seminar!

<u>Shameless Plug</u>:

Workshop @ UCL on post-Bayesian methods
15. /16. May 2025!!!

Matias Altamirano
(UCL)

Yann McLatchie
(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

Please share widely! :)

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

The Bayesian hangover:
updating beliefs about updating beliefs

Jeremias Knoblauch
Department of Statistical Science
University College London

11/02/25

Key Questions addressed in talk

Part I: What's the Bayesian hangover?
And why do we need this seminar?

Part II: What is the (post-Bayesian) aspirin?

Part III: What will Chapter 1 cover?

Power/Fractional/Cold Posterior

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

Gibbs/Quasi/Pseudo Posterior

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

Optimisation-centric Posterior

$$q_n^*(\theta) = \arg\min_{q \in \mathcal{Q}} \left\{ \mathscr{L}(q, x_{1:n}) \quad + \quad \mathrm{D}(q, \pi) \right\}$$

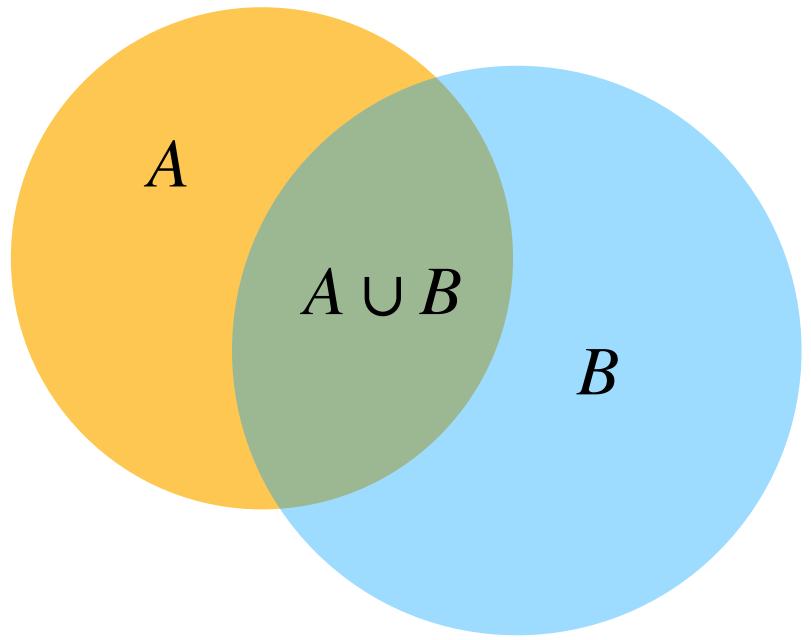
Part I: The Hangover

I

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

$$\boxed{P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}}$$

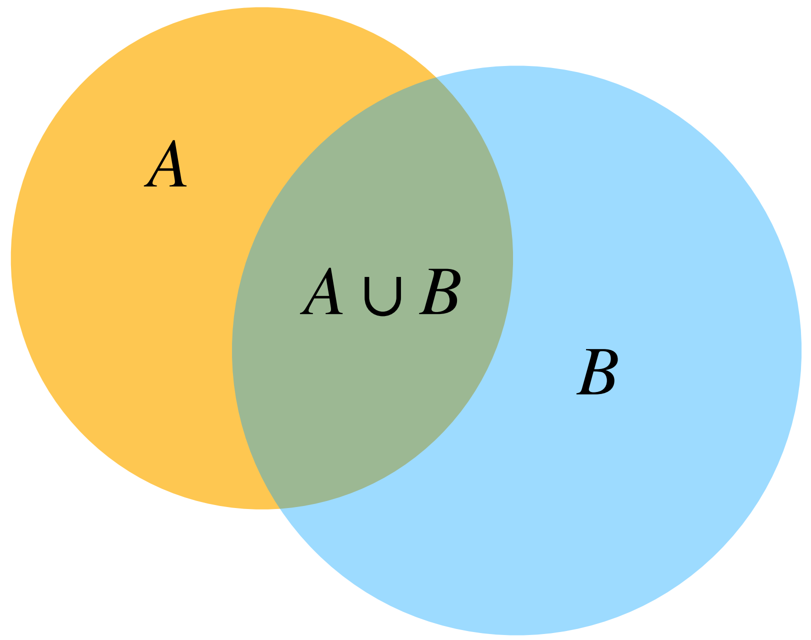


I

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$



Data model: $p(x_{1:n} \mid \theta)$

$x_{1:n} \in \mathscr{X}^n$

Prior probability: $\pi(\theta)$

$\theta \in \Theta$

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

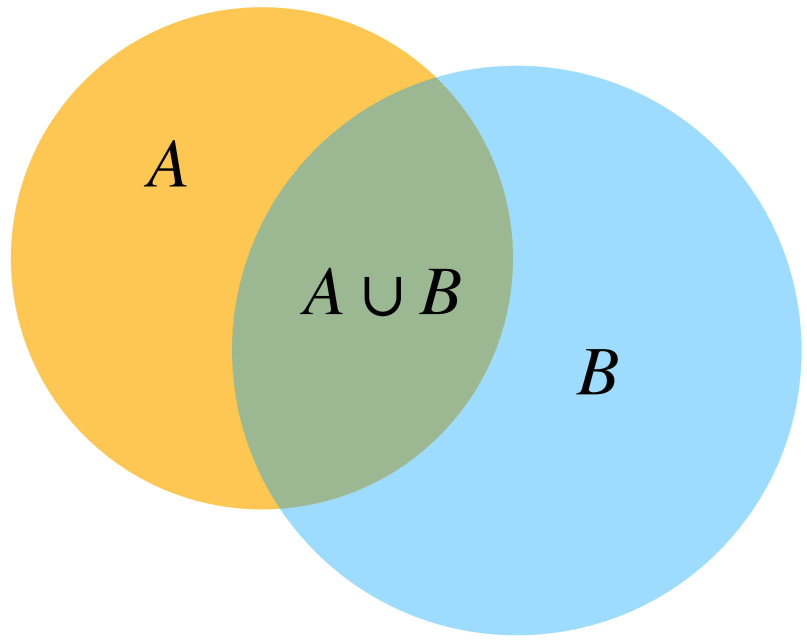
(Bayes) Posterior

I

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

$$\boxed{P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}}$$



Data model: $p(x_{1:n} \mid \theta)$

$x_{1:n} \in \mathcal{X}^n$

Prior probability: $\pi(\theta)$

$\theta \in \Theta$

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

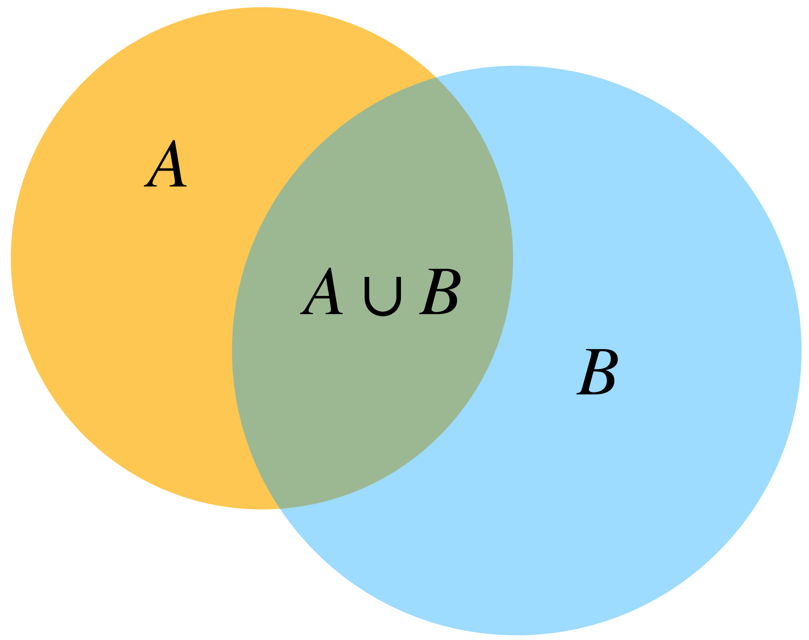
(Bayes) Posterior

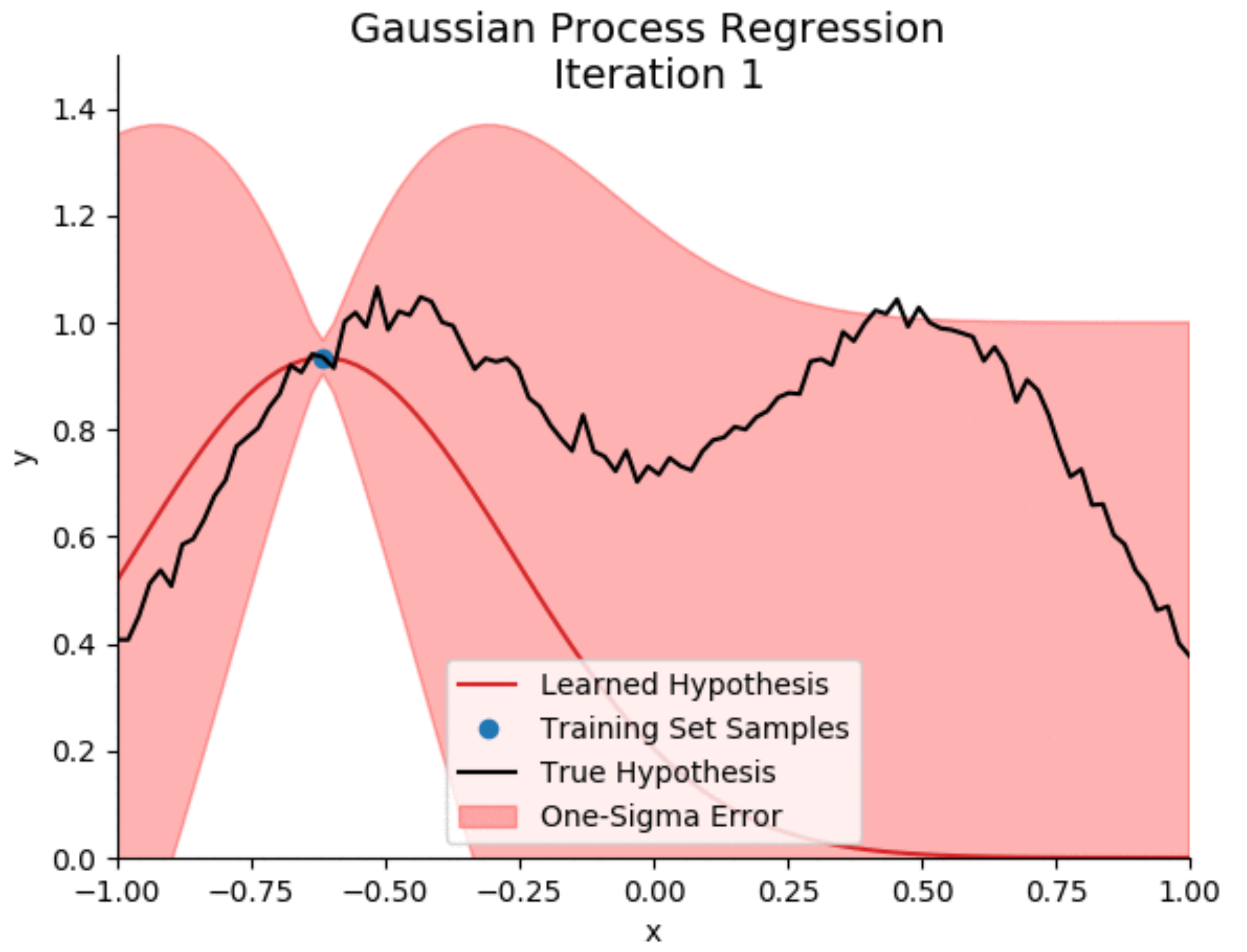
- (+) Averages models (instead of picking only one)
- (+) Quantifies uncertainty about θ via $\pi_n(\theta \mid x_{1:n})$
- (+) Inclusion of domain expertise via prior π

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

$$\boxed{P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}}$$





Data model: $p(x_{1:n} \mid \theta)$

$x_{1:n} \in \mathscr{X}^n$

Prior probability: $\pi(\theta)$

$\theta \in \Theta$

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

(Bayes) Posterior

- ⊕ Averages models (instead of picking only one)
- ⊕ Quantifies uncertainty about θ via $\pi_n(\theta \mid x_{1:n})$
- ⊕ Inclusion of domain expertise via prior π

Problematic Assumptions for Bayesian Analysis

I

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

Θ = Only relevant State of the world

Problematic Assumptions for Bayesian Analysis

I

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

Θ = Only relevant State of the world

(A2) $\boxed{\pi(\theta) = \text{ uncertainty about the true State of the world}}$

How rational decision-makers choose the prior

Problematic Assumptions for Bayesian Analysis

I

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

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

Θ = Only relevant State of the world

(A2) $\boxed{\pi(\theta) = \text{ uncertainty about the true State of the world}}$

How rational decision-makers choose the prior

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

Guarantees real-world relevance

Problematic Assumptions for Bayesian Analysis

I

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

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

$\Theta \;=\;$ Only relevant State of the world

FRAGILE

(A2) $\boxed{\pi(\theta) = \text{uncertainty about the true State of the world}}$

How rational decision-makers choose the prior

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

Guarantees real-world relevance

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with
research question

$x_{1:n}$

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

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge

$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$

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

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question	Statistical modelling & expert knowledge	Inference
$x_{1:n}$	$p(x_{1:n} \mid \theta), \pi(\theta)$	$\pi_n(\theta \mid x_{1:n})$

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

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question		Statistical modelling & expert knowledge		Inference
$x_{1:n}$	$\longrightarrow$	$p(x_{1:n} \mid \theta), \pi(\theta)$	$\longrightarrow$	$\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question	Statistical modelling & expert knowledge	Inference
$x_{1:n}$	$p(x_{1:n} \mid \theta),\ \pi(\theta)$	$\pi_n(\theta \mid x_{1:n})$

$$x_{1:n} \longrightarrow p(x_{1:n} \mid \theta),\ \pi(\theta) \longrightarrow \pi_n(\theta \mid x_{1:n})$$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

$$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$$

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question	Statistical modelling & expert knowledge	Inference
$x_{1:n}$	$p(x_{1:n} \mid \theta), \pi(\theta)$	$\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts

(A2) ✓

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question	Statistical modelling & expert knowledge	Inference
$x_{1:n}$	$p(x_{1:n} \mid \theta),\ \pi(\theta)$	$\pi_n(\theta \mid x_{1:n})$

$$x_{1:n} \longrightarrow p(x_{1:n} \mid \theta),\ \pi(\theta) \longrightarrow \pi_n(\theta \mid x_{1:n})$$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\underbrace{\log y_i}_{\text{willingness to pay}} = \sum_{j=1}^{J_1} p_j \underbrace{\log(x_{j,i})}_{\text{pollutants}} + c_0 + \sum_{j=J_1}^{J_2} c_j \underbrace{\log(x_{j,i})}_{\text{rooms, sqm, ...}} + \underbrace{\varepsilon_i}_{\text{measurement error}}$$

$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

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

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge → Inference

$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$ → $\pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model

$p(x_{1:n} \mid \theta), \pi(\theta)$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

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

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge → Inference

$x_{1:n} \longrightarrow p(x_{1:n} \mid \theta), \pi(\theta) \longrightarrow \pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model → Different data/ problems

$p(x_{1:n} \mid \theta), \pi(\theta) \longrightarrow x_{1:n}$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question		Statistical modelling & expert knowledge		Inference
$x_{1:n}$	$\longrightarrow$	$p(x_{1:n} \mid \theta), \pi(\theta)$	$\longrightarrow$	$\pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model		Different data/ problems		Inference
$p(x_{1:n} \mid \theta), \pi(\theta)$	$\longrightarrow$	$x_{1:n}$	$\longrightarrow$	$\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question		Statistical modelling & expert knowledge		Inference
●	⟶	●	⟶	●
$x_{1:n}$		$p(x_{1:n} \mid \theta), \pi(\theta)$		$\pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model		Different data/ problems		Inference
●	⟶	●	⟶	●
$p(x_{1:n} \mid \theta), \pi(\theta)$		$x_{1:n}$		$\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Pearce et al. (2020) [AISTATS]

Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question	Statistical modelling & expert knowledge	Inference
$x_{1:n}$	$p(x_{1:n} \mid \theta), \pi(\theta)$	$\pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model	Different data/ problems	Inference
$p(x_{1:n} \mid \theta), \pi(\theta)$	$x_{1:n}$	$\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

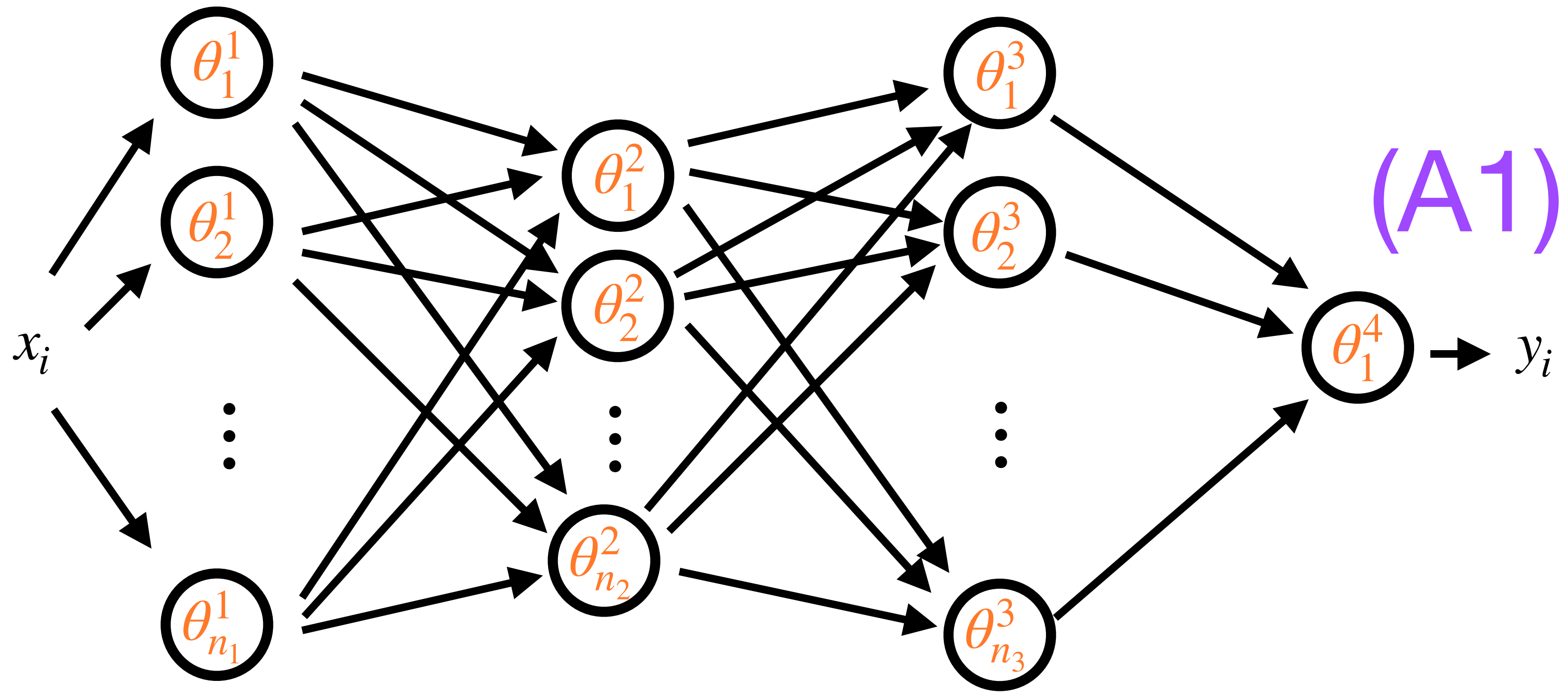
$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Pearce et al. (2020) [AISTATS]

Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?



(A1) ✓

F_θ : Input → Output

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge → Inference

$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$ → $\pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model → Different data/ problems → Inference

$p(x_{1:n} \mid \theta), \pi(\theta)$ → $x_{1:n}$ → $\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

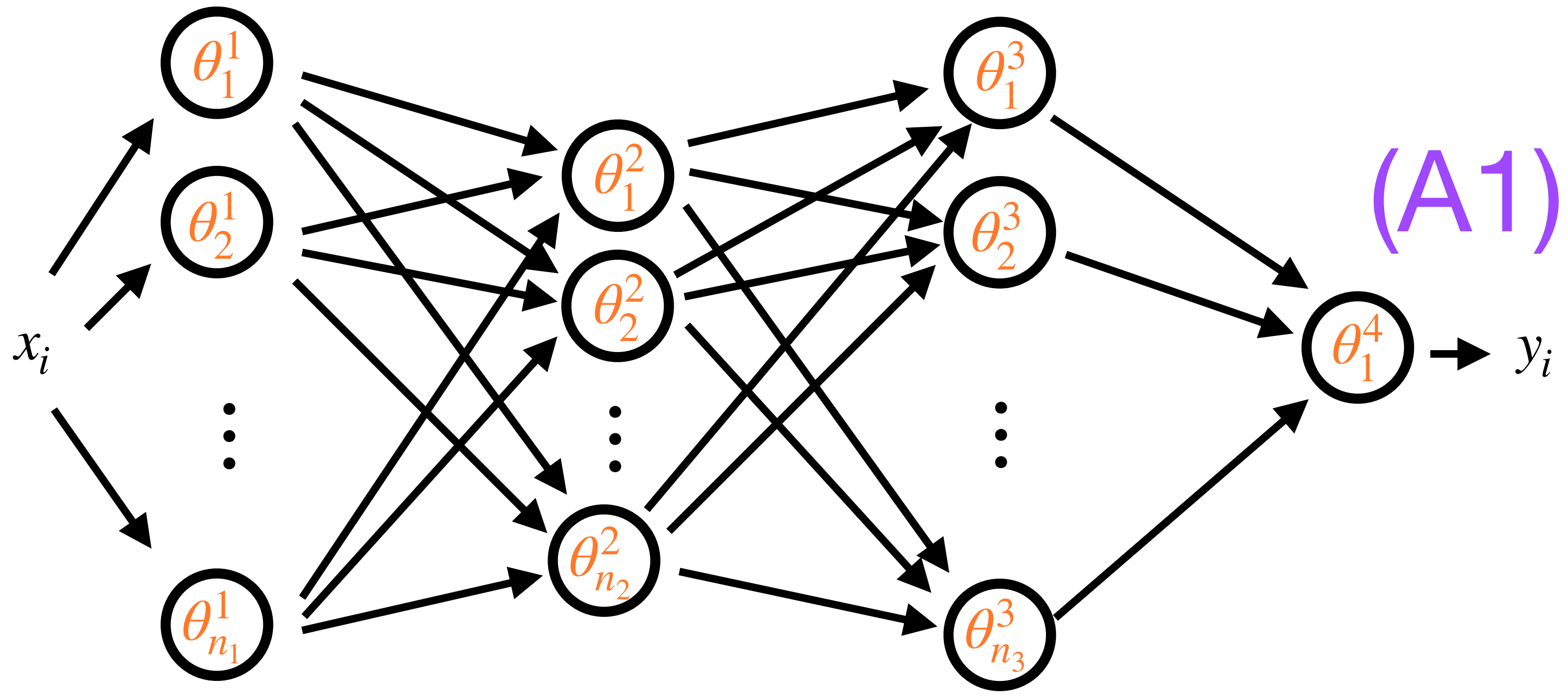
$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Pearce et al. (2020) [AISTATS]

Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?



(A1) ✓

F_θ : Input → Output

$\pi(\theta) \sim$ Normal

~~(A2)~~

Case Study: Bayesian ML & Boston Housing Data

I

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge → Inference

$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$ → $\pi_n(\theta \mid x_{1:n})$

Modern Bayesian ML

Flexible model → Different data/ problems → Inference

$p(x_{1:n} \mid \theta), \pi(\theta)$ → $x_{1:n}$ → $\pi_n(\theta \mid x_{1:n})$

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

willingness to pay

pollutants

rooms, sqm, ...

measurement error

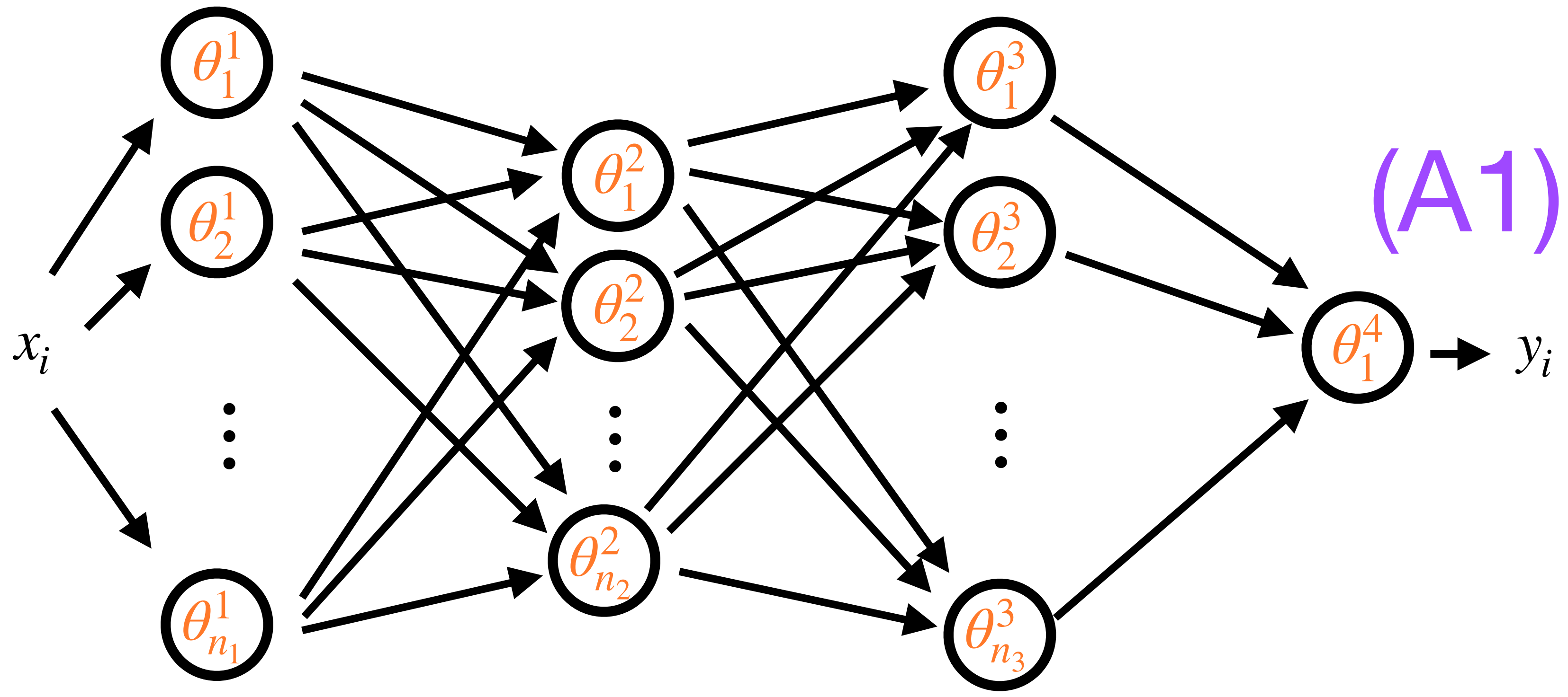
$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Pearce et al. (2020) [AISTATS]

Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?



(A1) ✓

F_θ : Input → Output

$\pi(\theta) \sim$ Normal ~~(A2)~~

$\pi_n(\theta \mid x_{1:n}) \longrightarrow$ coarse approximation ~~(A3)~~

Assumptions & Foundations

I

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge → Inference

$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$ → $\pi_n(\theta \mid x_{1:n})$

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

Modern Bayesian ML

Flexible model → Different data/ problems → Inference

$p(x_{1:n} \mid \theta), \pi(\theta)$ → $x_{1:n}$ → $\pi_n(\theta \mid x_{1:n})$

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

FRAGILE

Assumptions & Foundations

Traditional Bayesian analysis in science

Expert with research question → Statistical modelling & expert knowledge → Inference

$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$ → $\pi_n(\theta \mid x_{1:n})$

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Modern Bayesian ML

Flexible model → Different data/ problems → Inference

$p(x_{1:n} \mid \theta), \pi(\theta)$ → $x_{1:n}$ → $\pi_n(\theta \mid x_{1:n})$

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

FRAGILE

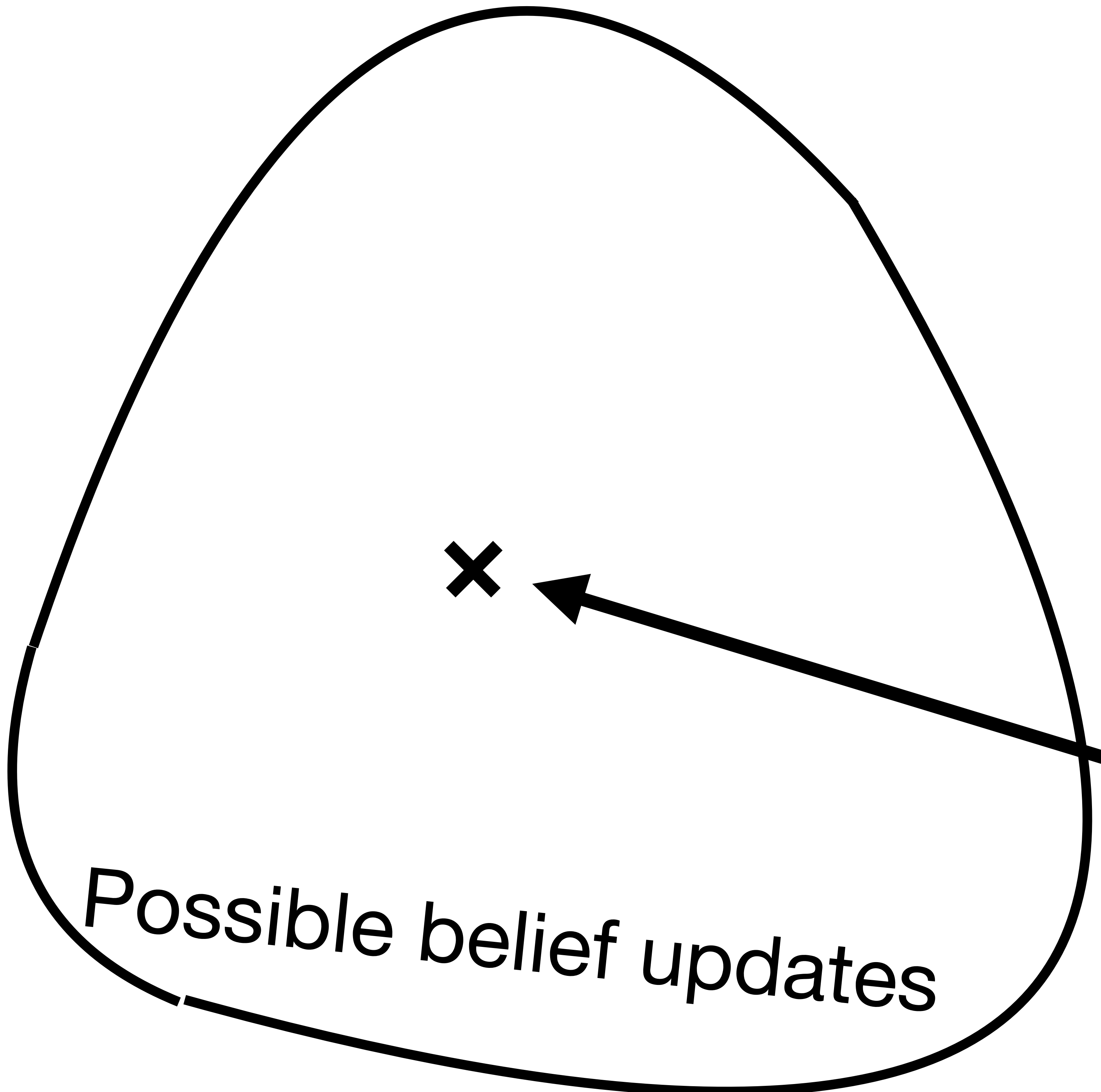
Post-Bayesian Approaches ask:

Can we keep benefits of Bayesianism without these assumptions???

This seminar is an attempt to organise ourselves under a common banner!!!

Part II: What is the (post-Bayesian) Aspirin?

Post-Bayesian Inference



Bayes' Posterior (A1), (A2), (A3)

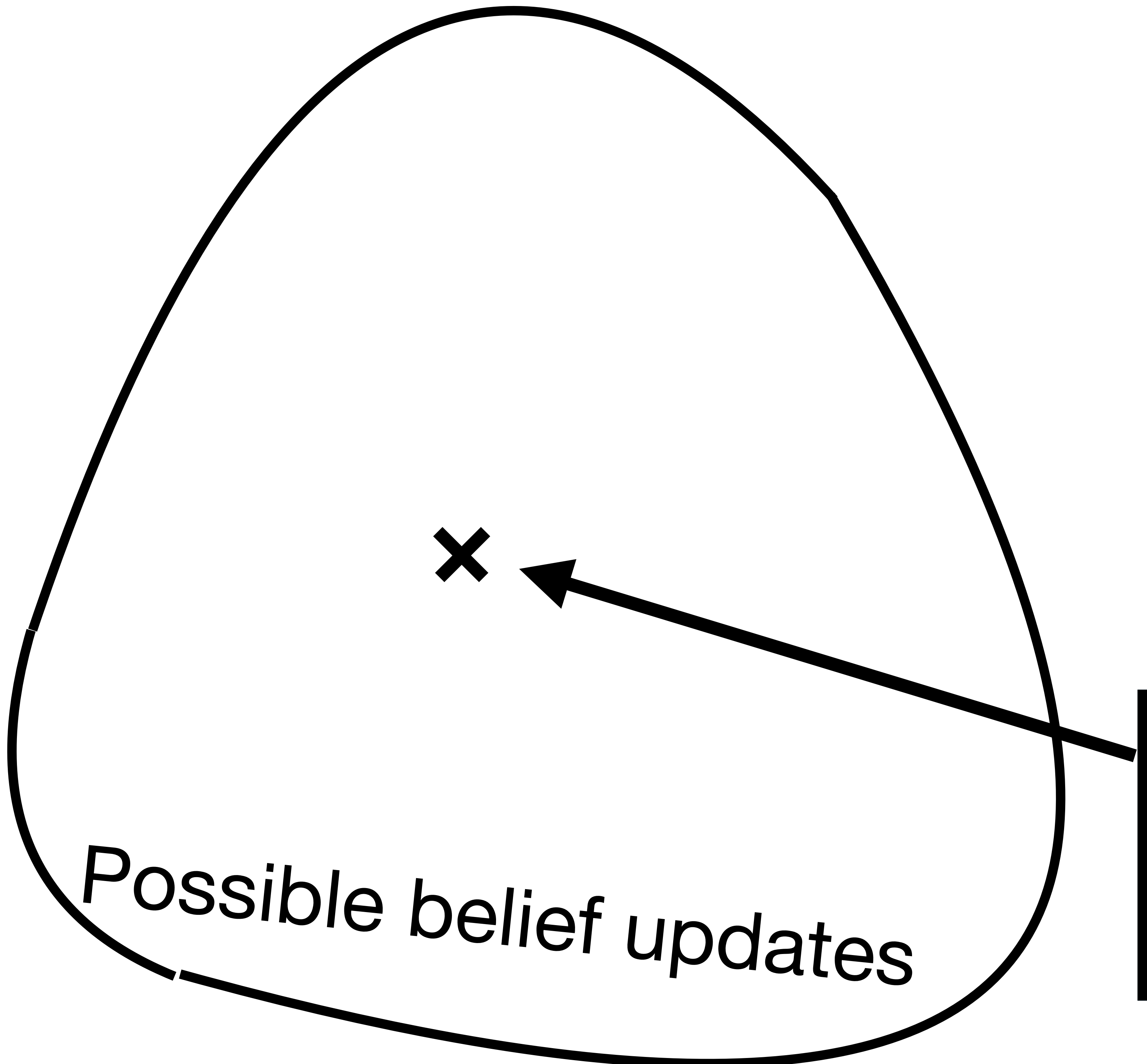
$$\pi_n(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta) \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta) \cdot \pi(\theta) d\theta}$$

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

Post-Bayesian Inference

II

$$\text{Belief updates} = \Big\{ \mathscr{B} : \underset{\text{space of priors}}{\mathscr{P}(\Theta)} \times \underset{\text{data space}}{\mathscr{X}^n} \times \underset{\text{space of hyperparameters}}{\mathscr{H}} \longrightarrow \underset{\text{space of posteriors}}{\mathscr{P}(\Theta)} \Big\}$$



Bayes' Posterior (A1), (A2), (A3)

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

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

Post-Bayesian Inference

space of priors

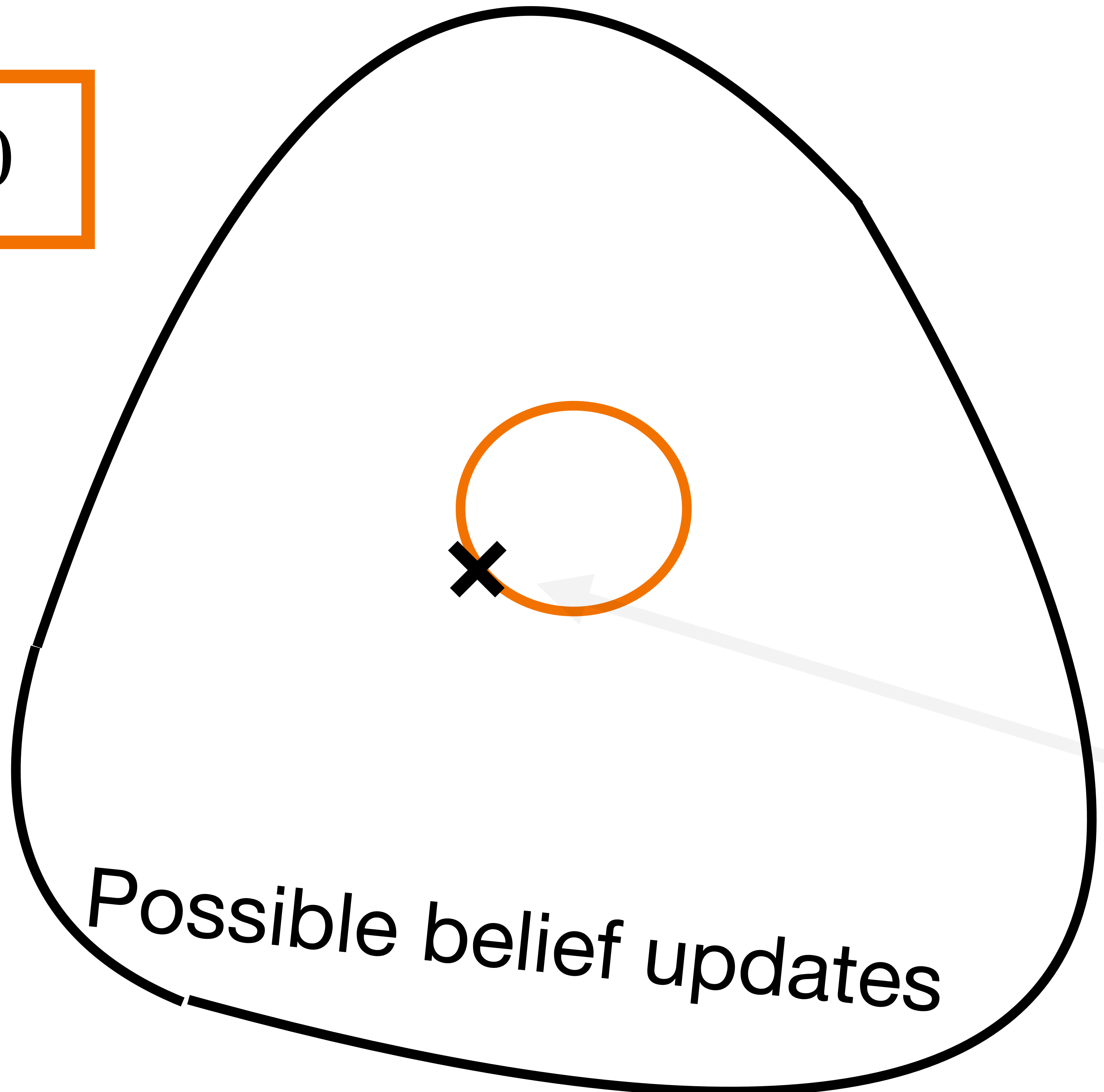
space of hyperparameters

$$\text{Belief updates} = \left\{ \mathscr{B} : \mathscr{P}(\Theta) \times \mathscr{X}^n \times \mathscr{H} \longrightarrow \mathscr{P}(\Theta) \right\}$$

data space

space of posteriors

$$p(x_{1:n} \mid \theta) \longrightarrow p(x_{1:n} \mid \theta)^{\lambda}, \ \lambda > 0$$



(A1), (A2), (A3)

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

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Post-Bayesian Inference

II

space of priors

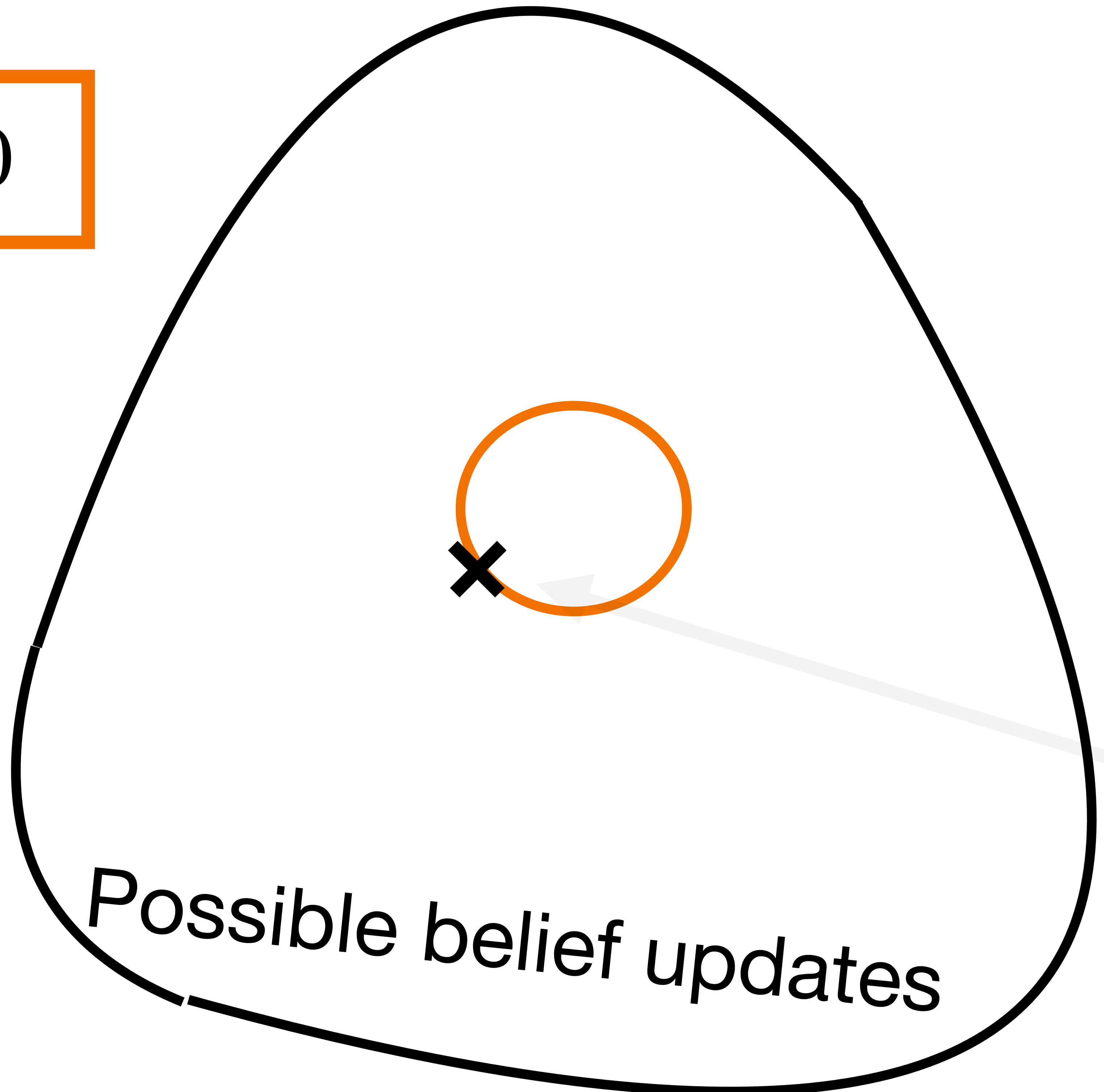
space of hyperparameters

$$\text{Belief updates} = \left\{ \mathscr{B} : \mathscr{P}(\Theta) \times \mathscr{X}^n \times \mathscr{H} \longrightarrow \mathscr{P}(\Theta) \right\}$$

data space

space of posteriors

$$\boxed{p(x_{1:n} \mid \theta) \longrightarrow p(x_{1:n} \mid \theta)^{\lambda}, \ \lambda > 0}$$



[See Grünwald (2011)]

Power/Fractional/
Cold Posterior

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

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

(A1), (A2), (A3)

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

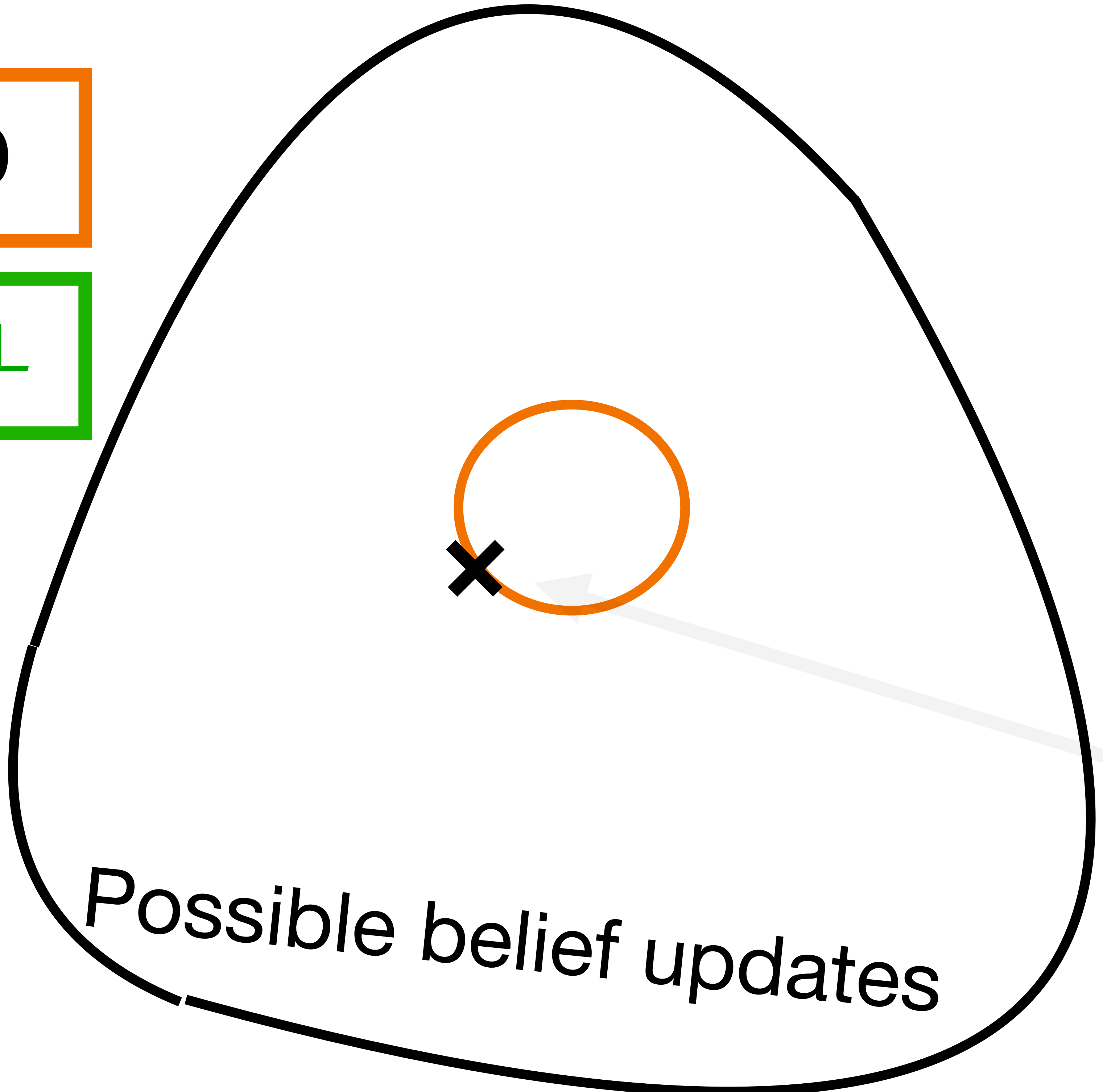
(A1) model well-specified
(A2) prior well-specified
(A3) computationally feasible

Post-Bayesian Inference

$$p(x_{1:n} \mid \theta) \longrightarrow p(x_{1:n} \mid \theta)^{\lambda}, \;\; \lambda > 0$$

$$p(x_{1:n} \mid \theta) \longrightarrow \exp\{-\mathsf{L}(x_{1:n}, p_\theta)\}, \;\; \text{loss } \mathsf{L}$$

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



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

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

(A1), (A2), (A3)

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

Post-Bayesian Inference

II

[See Bissiri, Holmes & Walker (2016)]

Gibbs/Generalised/
Pseudo Posterior

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

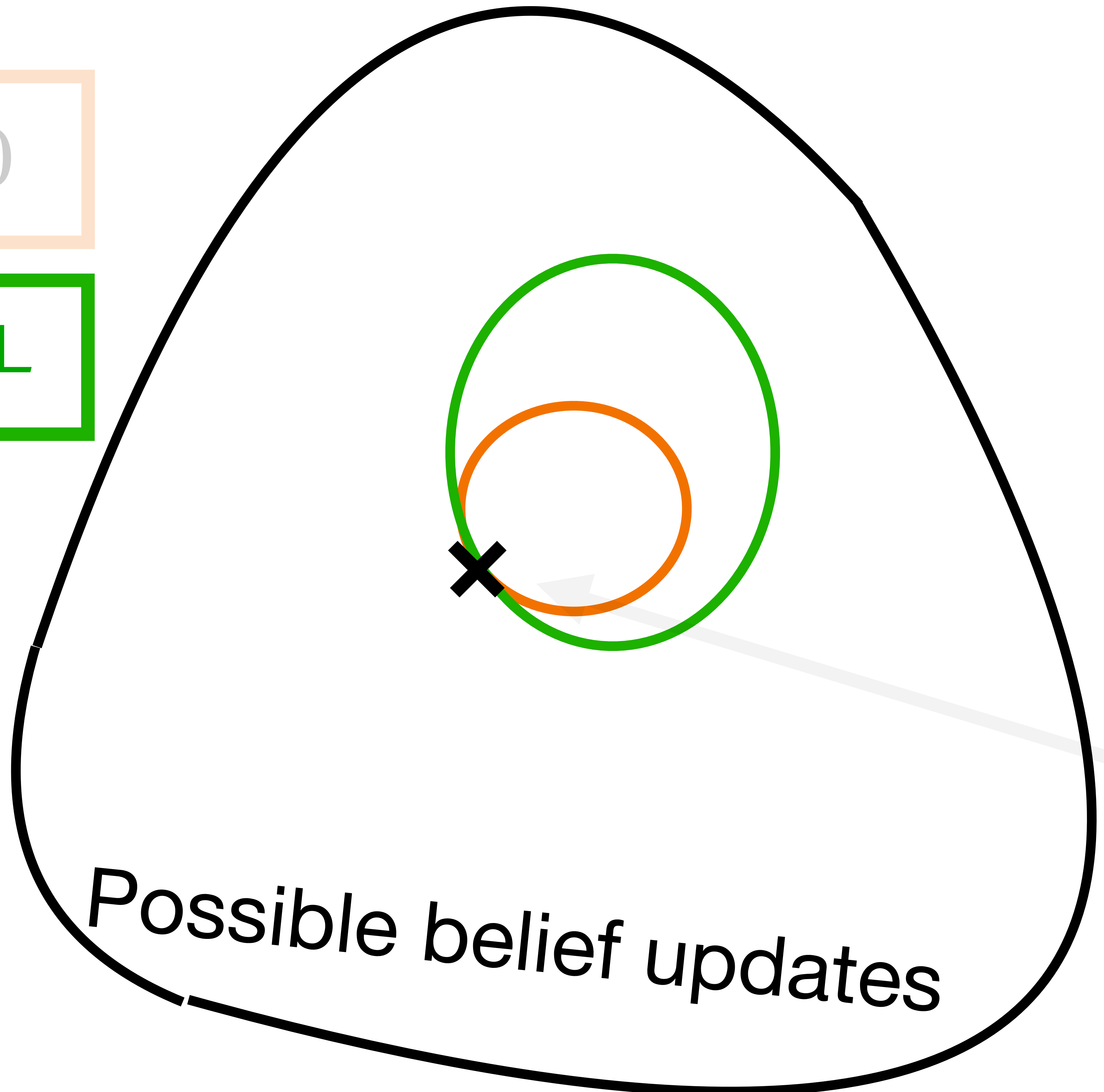
$$\boxed{\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathsf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathsf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}}$$

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

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

$$\boxed{p(x_{1:n} \mid \theta) \longrightarrow p(x_{1:n} \mid \theta)^\lambda, \ \lambda > 0}$$

$$\boxed{p(x_{1:n} \mid \theta) \longrightarrow \exp\{-\mathsf{L}(x_{1:n}, p_\theta)\}, \ \text{loss } \mathsf{L}}$$



(A1), (A2), (A3)

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

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

Post-Bayesian Inference

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

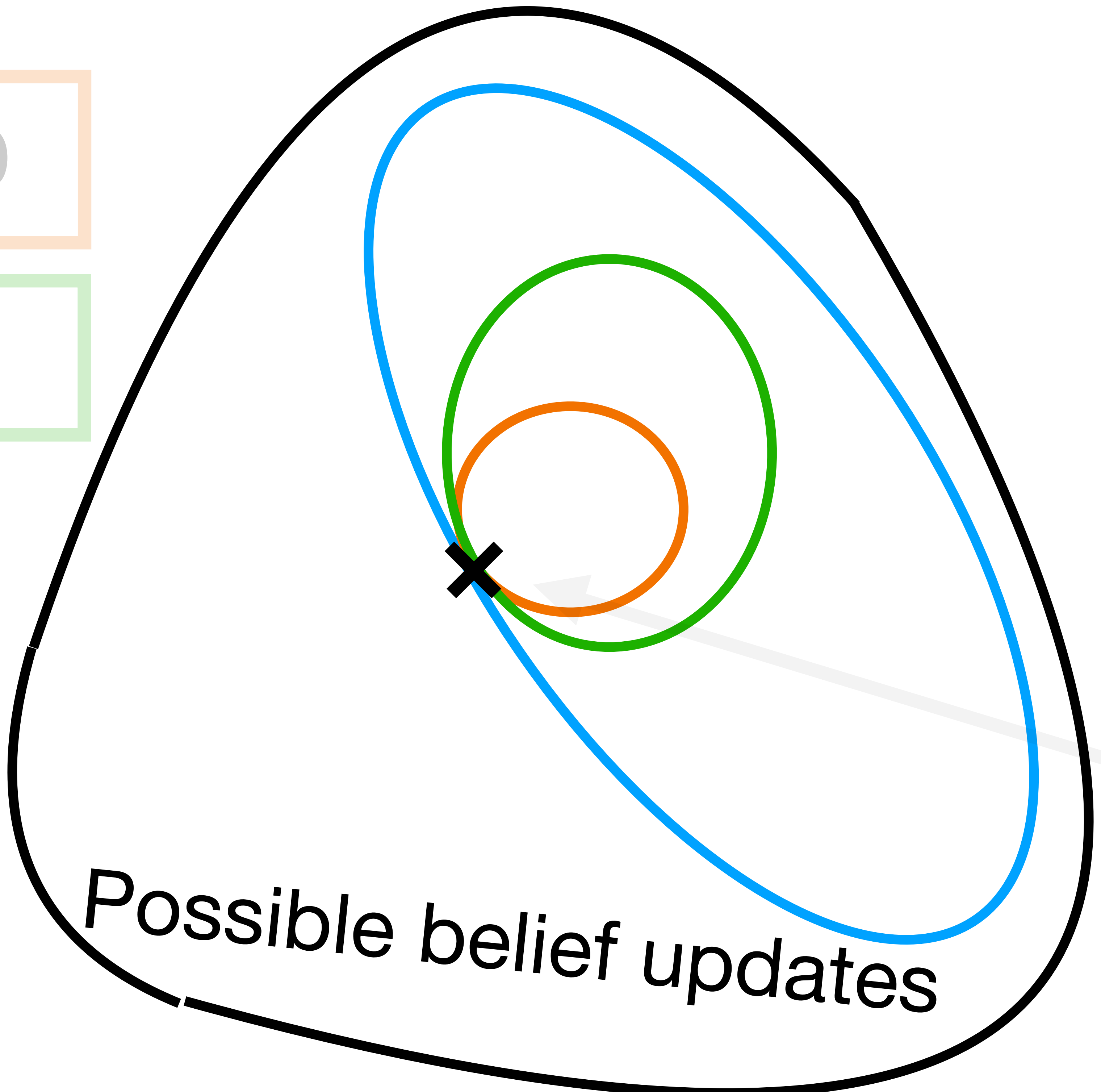
$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$

[See Knoblauch, Jewson, & Damoulas (2019/2022)]

$$p(x_{1:n} \mid \theta) \longrightarrow p(x_{1:n} \mid \theta)^{\lambda}, \;\; \lambda > 0$$

KL	$\longrightarrow$	D
$\mathcal{P}(\Theta)$	$\longrightarrow$	$\mathcal{Q}$

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



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

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

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

(A1), (A2), (A3)

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

Post-Bayesian Inference

II

Optimisation-centric posteriors / Generalised Variational Inference

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

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

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

Gibbs/Generalised/ Pseudo Posterior

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

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

Martingale posteriors & resampling-based approaches

[See Fong, Holmes, & walker (2023)]

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

For $i = 1,2,\dots$

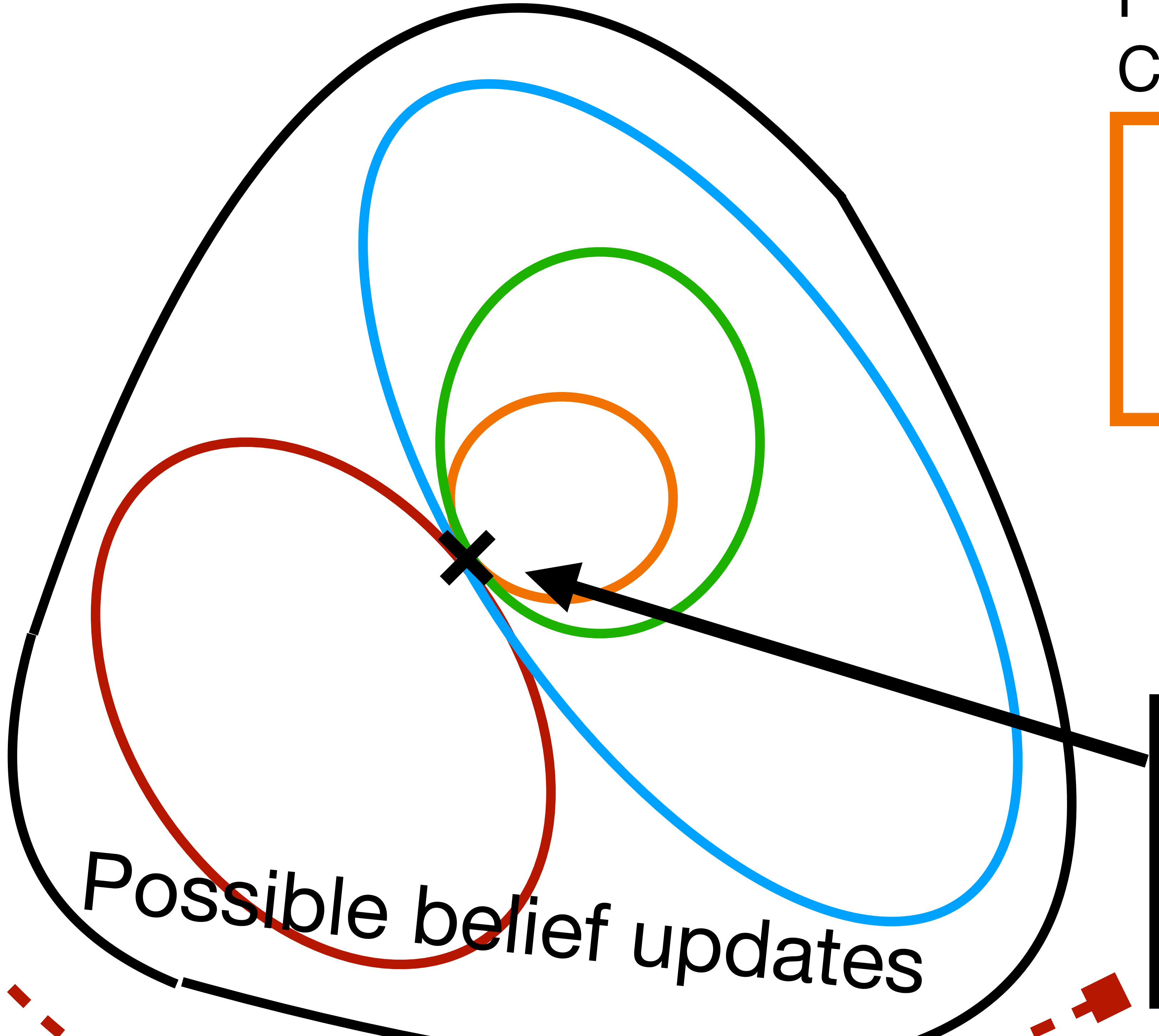
$$X_{n+i+1} \sim p(X_{n+i} \mid x_{1:n}, X_{n+1:n+i})$$

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

Power/Fractional/ Cold Posterior

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

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



Bayes' Posterior

(A1), (A2), (A3)

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

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

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

For $i = 1,2,\ldots$

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

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

Chapter 2

Resampling & Martingale Posteriors (06/05—15/07)

Dr. Edwin Fong (University of Hong Kong)

Optimisation-centric posteriors / Generalised Variational Inference ~~(A1)~~, ~~(A2)~~, ~~(A3)~~

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

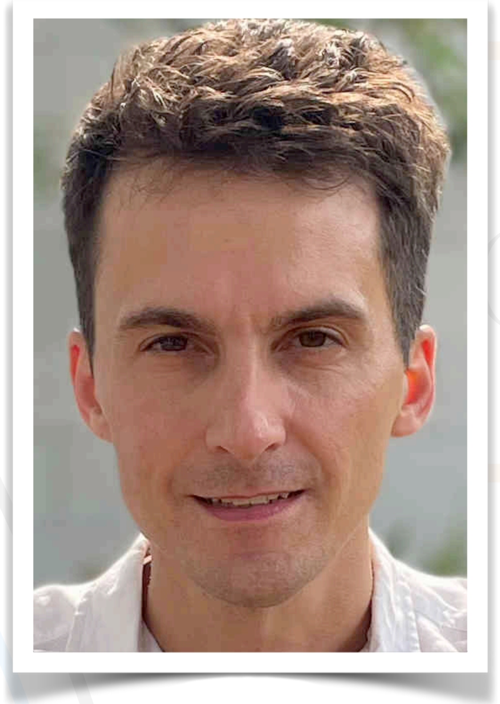
$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$

Gibbs/Generalised/ Pseudo Posterior ~~(A1)~~, (A2), (A3)

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

Chapter 3

PAC-Bayes (after summer break)



Prof. Pierre Alquier (ESSEC Singapore)

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)

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

Chapter 1

Generalised Bayes (11/02—22/04)

Structure of Chapter 1

Today: Overview of post-Bayesian ideas & generalised Bayes

25/02: Theoretical foundations
(Prof. David Frazier)

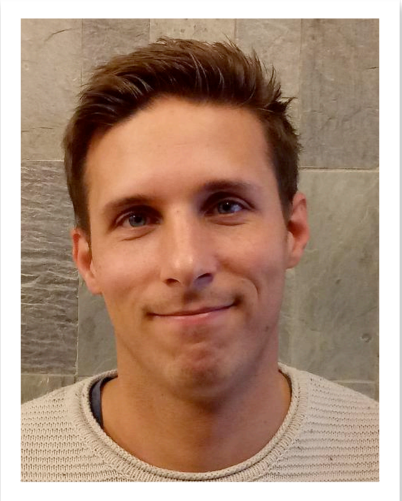
11/03: Learning rate selection & the power posterior
(Prof. Ryan Martin)

25/03: Prediction-centric approaches
(Prof. Chris Oates)

08/04: Coarsened Bayes & applications for biomedical problems
(Prof. David Dunson)

22/04: From generalised Bayes to Martingale Posteriors
(Prof. Chris Holmes)

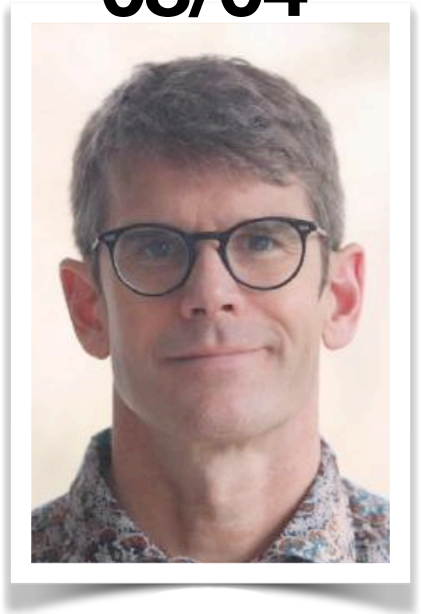
25/02



11/03

25/03

08/04



22/04

Structure of Chapter 1

Today: Overview of post-Bayesian ideas & generalised Bayes

25/02: Theoretical foundations
(Prof. David Frazier)

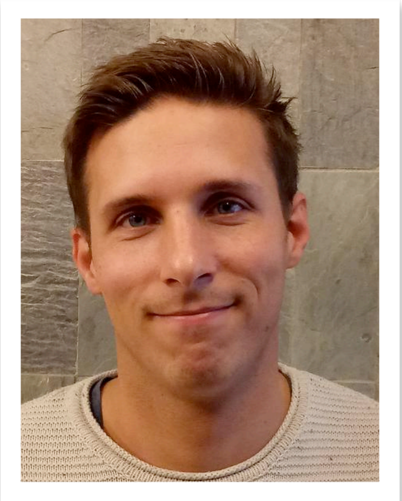
11/03: Learning rate selection & the power posterior
(Prof. Ryan Martin)

25/03: Prediction-centric approaches
(Prof. Chris Oates)

08/04: Coarsened Bayes & applications for biomedical problems
(Prof. David Dunson)

22/04: From generalised Bayes to Martingale Posteriors
(Prof. Chris Holmes)

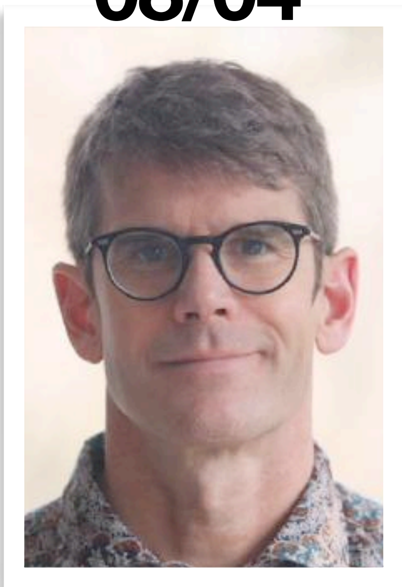
25/02



11/03

25/03

08/04



22/04

Part III: Basics of generalised Bayes

Power/Fractional/Cold Posteriors

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

Gibbs/Generalised/
Pseudo Posterior

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

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

Data-fitting

Prior regularisation

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

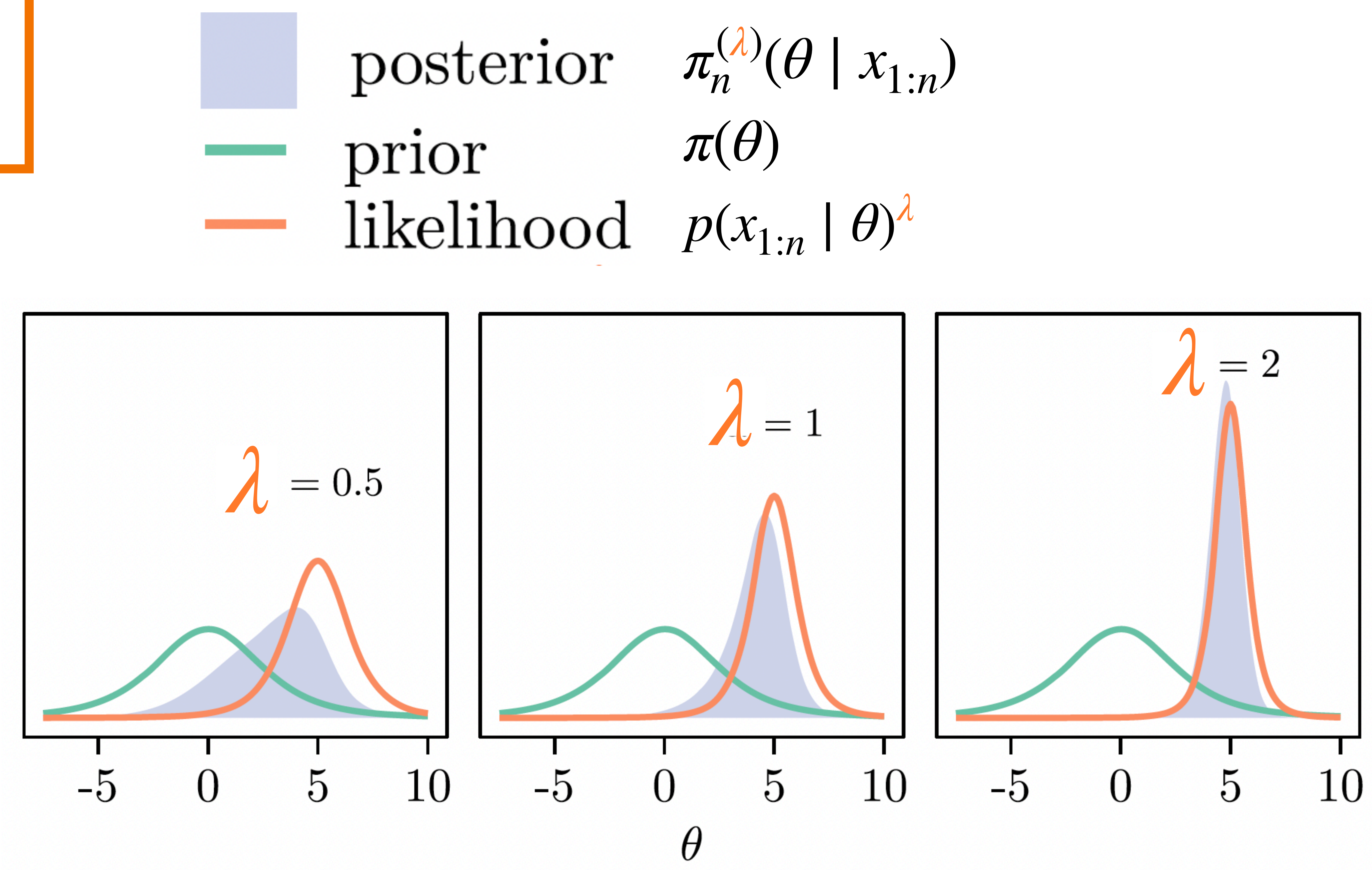
Q: What does it do?

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

Q: What does it do?
A: Trades off prior vs data



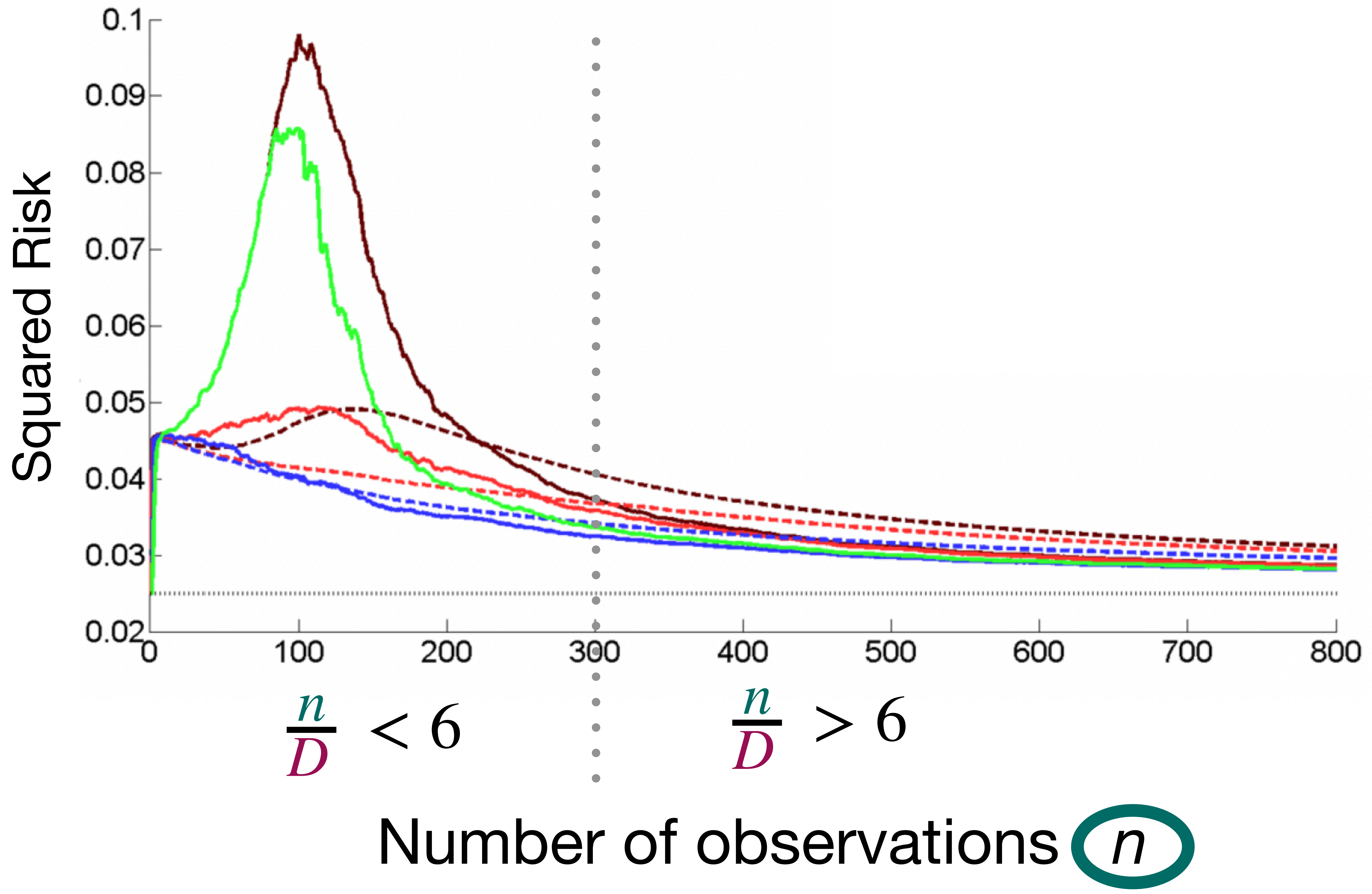
Picture from Kallionen, Paananen, Bürkner, & Vehtari (2023)

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

Q: Why do it do?



Regression model (misspecified):

$$p(y_i \mid \theta, x_i) = \mathcal{N}\left(y_i; \sum_{d=1}^{50} \theta_i \, x_{i,d} \, , \sigma^2\right)$$

D

Basics: power posteriors

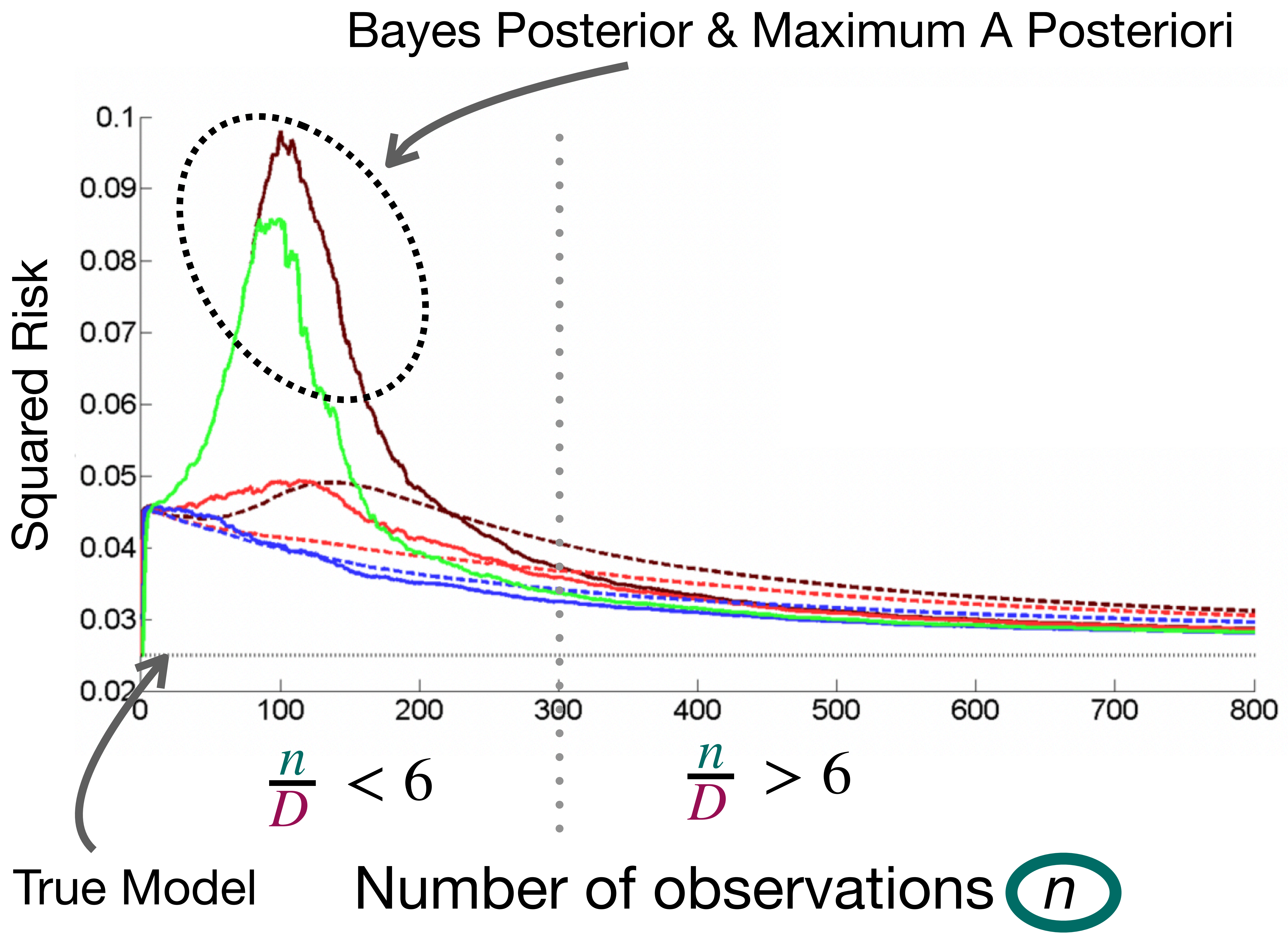
Power/Fractional/Cold Posteriors

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

Q: Why do it do?

Regression model (misspecified):

$$p(y_i \mid \theta, x_i) = \mathcal{N}\left(y_i; \sum_{d=1}^{\overset{D}{50}} \theta_i \, x_{i,d} \, , \sigma^2\right)$$



Basics: power posteriors

Power/Fractional/Cold Posteriors

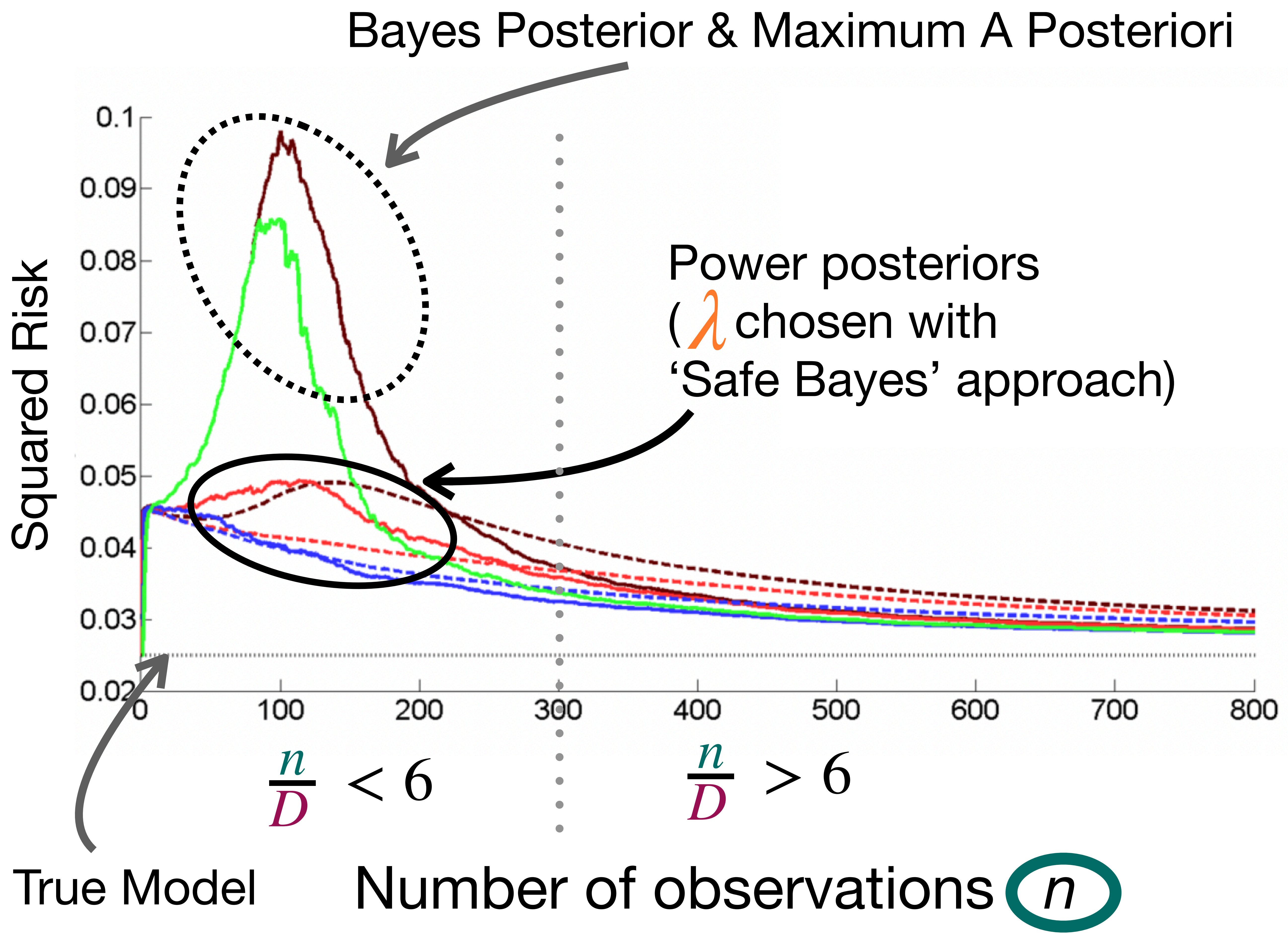
$$\pi_n^{(\lambda)}(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)^{\lambda} \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta)^{\lambda} \cdot \pi(\theta) d\theta}$$

Q: Why do it do?

Regression model (misspecified):

$$p(y_i \mid \theta, x_i) = \mathcal{N}\left(y_i; \sum_{d=1}^{\overset{D}{50}} \theta_i \, x_{i,d} \, , \sigma^2\right)$$

The **'Safe Bayes' effect** (see Grünwald, 2012)
(picture from Grünwald & van Ommen, 2017)



Basics: power posteriors

Power/Fractional/Cold Posteriors

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

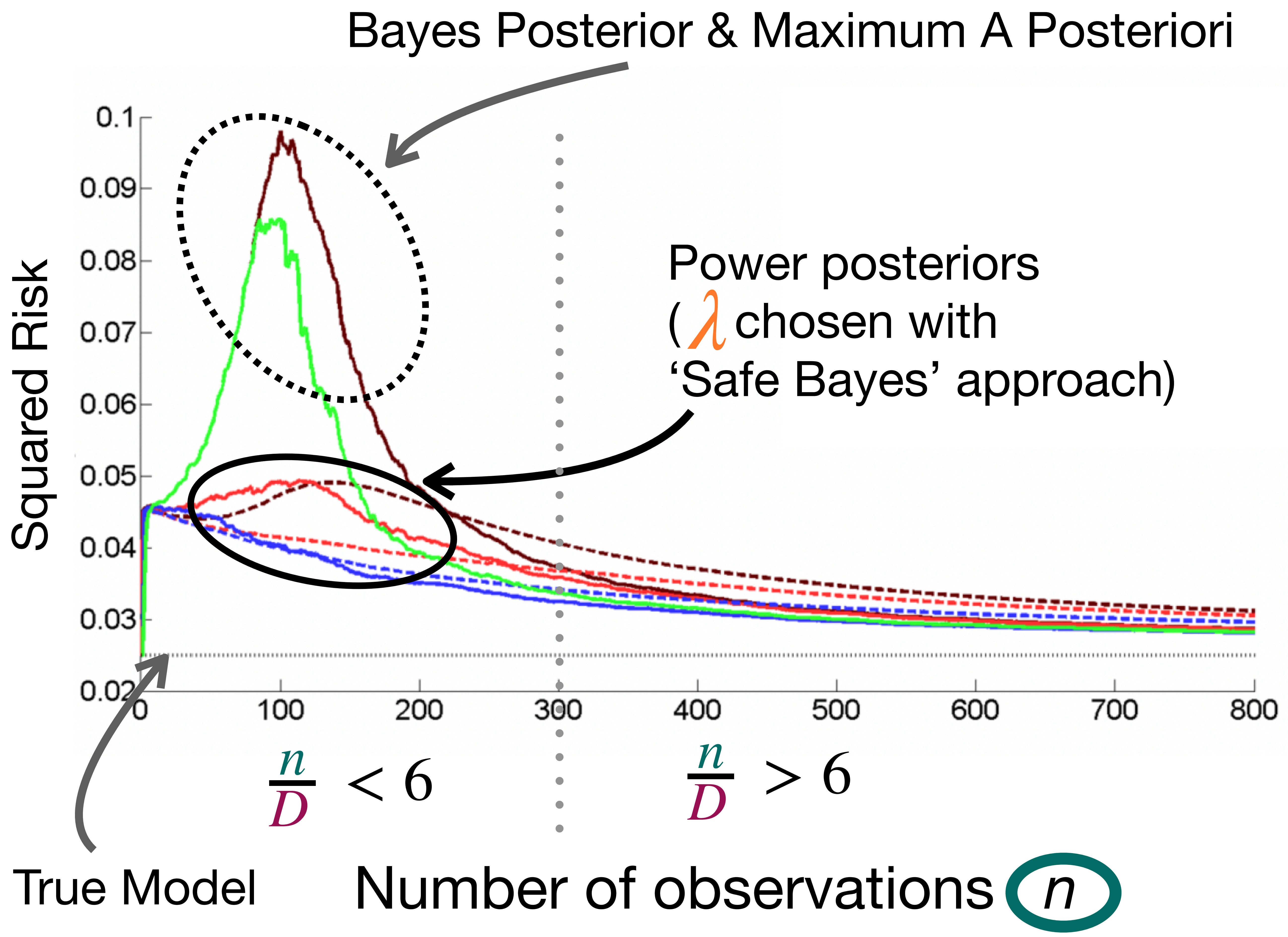
Q: Why do it do?
A: better risk properties/predictions

if $\frac{n}{D}$ is small

Regression model (misspecified):

$$p(y_i \mid \theta, x_i) = \mathcal{N}\left(y_i; \sum_{d=1}^{D = 50} \theta_i \, x_{i,d} \,, \sigma^2\right)$$

The **'Safe Bayes' effect** (see Grünwald, 2012)
(picture from Grünwald & van Ommen, 2017)



Basics: power posteriors

Power/Fractional/Cold Posteriors

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

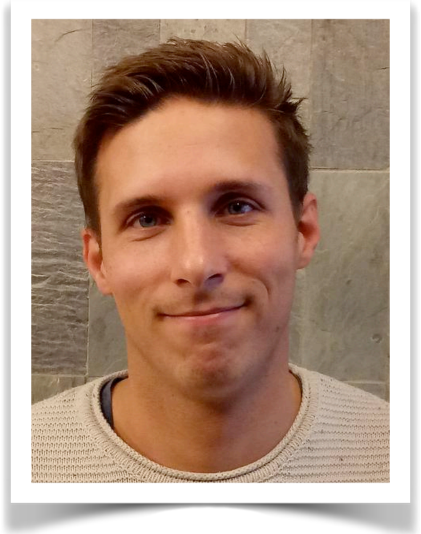
Q: What else have we discovered?

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

25/02



Q: What else have we discovered?

Power posterior & their variational approximations concentrate in situations where standard Bayes wouldn't

Bhattacharya, Pati, & Yang (2019)
Alquier & Ridgeway (2020)
Yang, Pati, & Bhattacharya (2020)

11/03

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

25/02

Q: What else have we discovered?

Power posterior & their variational approximations concentrate in situations where standard Bayes wouldn't

Bhattacharya, Pati, & Yang (2019)
Alquier & Ridgeway (2020)
Yang, Pati, & Bhattacharya (2020)

It is often surprisingly difficult to choose λ

Grünwald (2012)
Lyddon, Holmes, & Walker (2019)
Wu & Martin (2023)

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

25/02

Q: What else have we discovered?

Power posterior & their variational approximations concentrate in situations where standard Bayes wouldn't

Bhattacharya, Pati, & Yang (2019)
Alquier & Ridgeway (2020)
Yang, Pati, & Bhattacharya (2020)

11/03

It is often surprisingly difficult to choose λ

Grünwald (2012)
Lyddon, Holmes, & Walker (2019)
Wu & Martin (2023)

25/03

Predictive/robustness gains vanish provably & very quickly for even moderate $\frac{n}{D}$

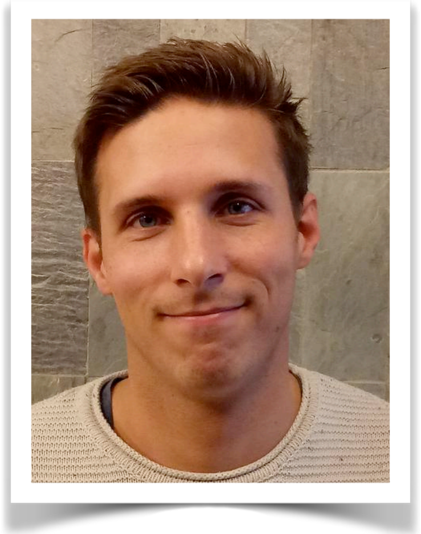
Medina, Olea, Rush, & Velez (2022)
McLatchie, Fong, Frazier, & Knoblauch (2024)

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

25/02



Q: What else have we discovered?

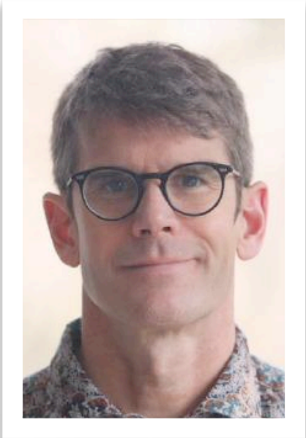
Power posterior & their variational approximations concentrate in situations where standard Bayes wouldn't

Bhattacharya, Pati, & Yang (2019)
Alquier & Ridgeway (2020)
Yang, Pati, & Bhattacharya (2020)

Power posterior $\approx$ 'Coarsened' posterior (conditioning on a neighbourhood of observed data)

Miller & Dunson (2019)

08/04



11/03

It is often surprisingly difficult to choose λ

Grünwald (2012)
Lyddon, Holmes, & Walker (2019)
Wu & Martin (2023)

25/03

Predictive/robustness gains vanish provably & very quickly for even moderate $\frac{n}{D}$

Medina, Olea, Rush, & Velez (2022)
McLatchie, Fong, Frazier, & Knoblauch (2024)

Basics: power posteriors

Power/Fractional/Cold Posteriors

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

Q: What else have we discovered?

Power posterior & their variational approximations concentrate in situations where standard Bayes wouldn't

Bhattacharya, Pati, & Yang (2019)
Alquier & Ridgeway (2020)
Yang, Pati, & Bhattacharya (2020)

Power posterior $\approx$ 'Coarsened' posterior (conditioning on a neighbourhood of observed data)

Miller & Dunson (2019)

It is often surprisingly difficult to choose λ

Grünwald (2012)
Lyddon, Holmes, & Walker (2019)
Wu & Martin (2023)

Predictive/robustness gains vanish provably & very quickly for even moderate $\frac{n}{D}$

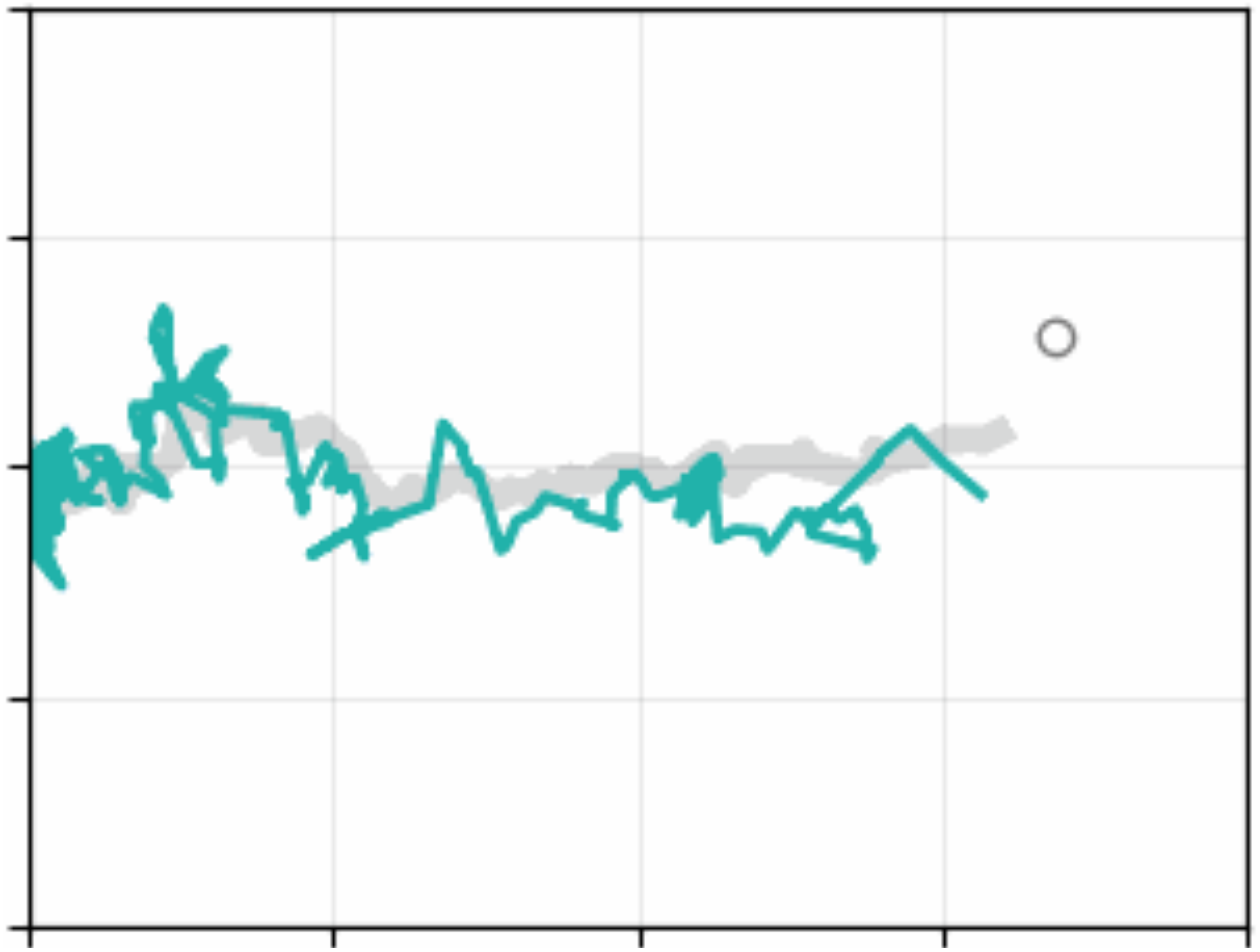
Medina, Olea, Rush, & Velez (2022)
McLatchie, Fong, Frazier, & Knoblauch (2024)

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

Kalman Filter Example: generalised / Gibbs posteriors

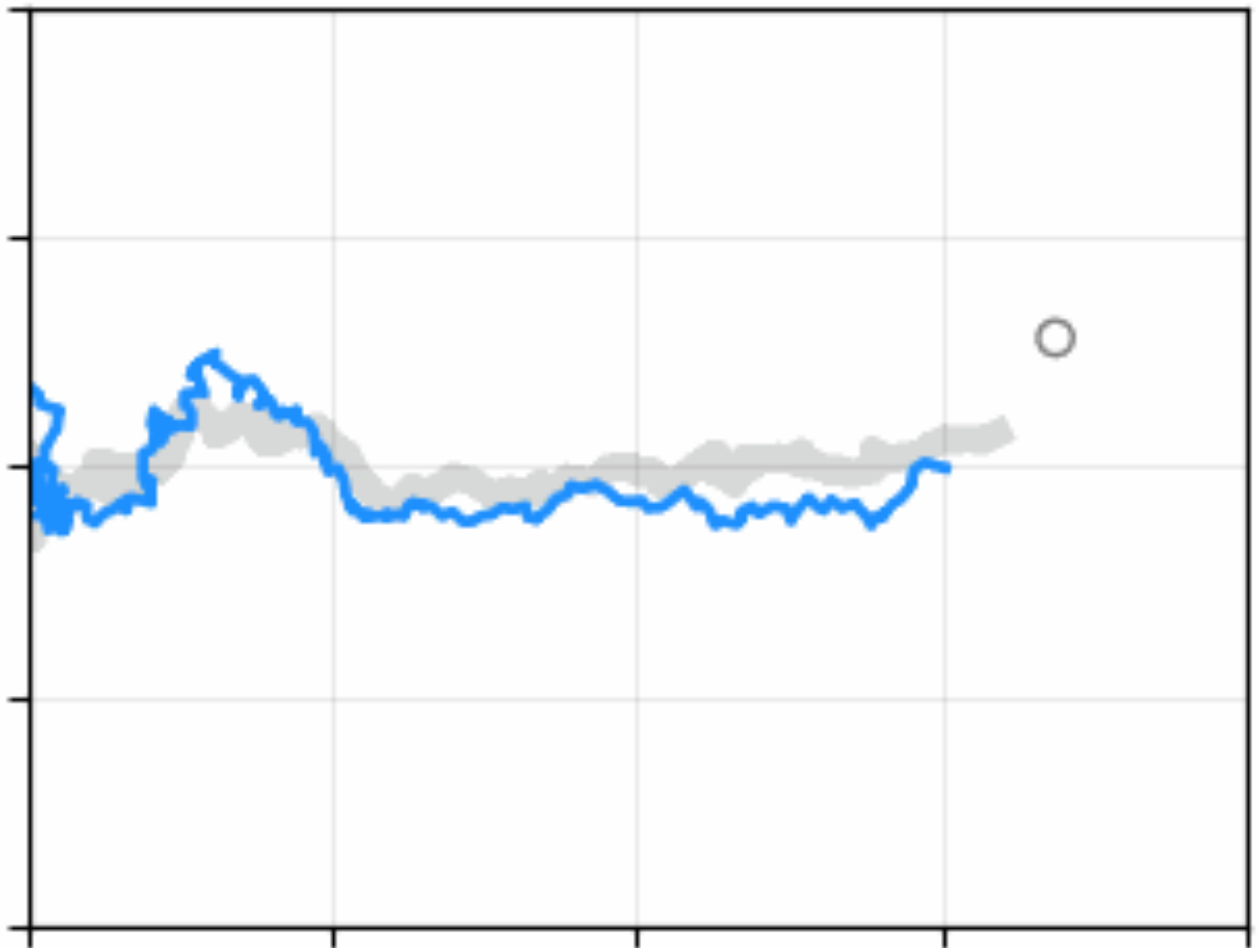
Bayes' Posterior

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



Gibbs/Generalised/Pseudo Posterior

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



Duran-Martin, Altamirano, Shestopaloff, Sanchez-Betancourt, Knoblauch, Briol, & Murphy (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

$\Big($ **NOT necessary for θ to come from a model p_θ** $\Big)$

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to think about this?

Perspective 1: 'General Bayes Updates'

(Conditional) independence: $p_\theta(x_{1:n}) = \prod_{i=1}^{n} p_\theta(x_i)$

$$\Downarrow$$

$$\pi_n(\theta \mid x_{1:n}) \propto \pi_{n-1}(\theta \mid x_{1:(n-1)}) \cdot p_\theta(x_n)$$

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) \propto \pi_n^{\mathsf{L}}(\theta \mid x_{1:(n-1)}) \cdot \exp\{-\lambda \cdot \ell(x_n, p_\theta)\}$$

$$\Uparrow$$

Summable Losses: $\mathsf{L}(x_{1:n}, p_\theta) = \sum_{i=1}^{n} \ell(x_i, p_\theta)$

e.g., Bissiri, Holmes, & Walker (2016)

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to think about this?

A: (1) General Bayes Updates

Perspective 1: 'General Bayes Updates'

(Conditional) independence: $p_\theta(x_{1:n}) = \prod_{i=1}^{n} p_\theta(x_i)$

$$\Downarrow$$

$$\pi_n(\theta \mid x_{1:n}) \propto \pi_{n-1}(\theta \mid x_{1:(n-1)}) \cdot p_\theta(x_n)$$

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) \propto \pi_n^{\mathsf{L}}(\theta \mid x_{1:(n-1)}) \cdot \exp\{-\lambda \cdot \ell(x_n, p_\theta)\}$$

$$\Uparrow$$

Summable Losses: $\mathsf{L}(x_{1:n}, p_\theta) = \sum_{i=1}^{n} \ell(x_i, p_\theta)$

e.g., Bissiri, Holmes, & Walker (2016)

Main restriction of this interpretation

Basics: generalised / Gibbs posteriors

III

Gibbs/Generalised/Pseudo Posterior

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

Q: How to think about this?

A: (1) General Bayes Updates

(2) Optimisation-centric view

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \underset{q \in \mathscr{P}(\Theta)}{\arg\min} \left\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda}\mathrm{KL}\left(q, \pi\right)}_{\text{Prior regularisation}} \right\};$$

All probability distributions over parameter space Θ

e.g., Knoblauch, Jewson, & Damoulas (2022)

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to think about this?

A: (1) General Bayes Updates
(2) Optimisation-centric view

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg \min_{q \in \mathscr{P}(\Theta)} \Big\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda} \mathrm{KL}\left(q, \pi\right)}_{\text{Prior regularisation}} \Big\};$$

All probability distributions over parameter space Θ

Need not be summable

e.g., Knoblauch, Jewson, & Damoulas (2022)

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to think about this?

A: (1) General Bayes Updates
(2) Optimisation-centric view

Perspective 2: 'Optimisation-centric'

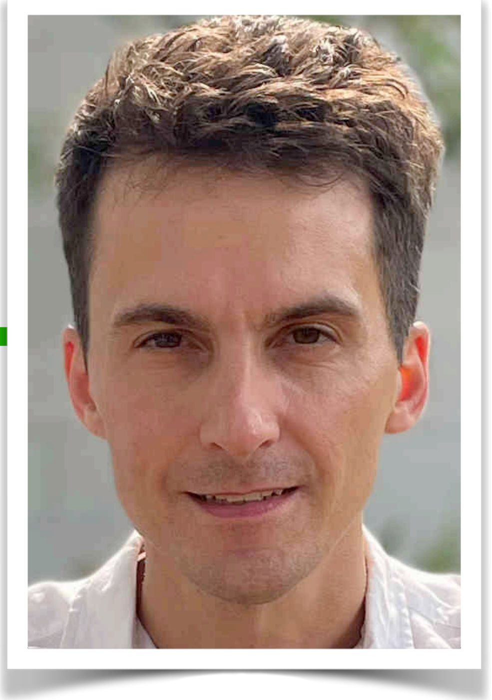
$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg \min_{q \in \mathscr{P}(\Theta)} \left\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda} \mathrm{KL}\left(q, \pi\right)}_{\text{Prior regularisation}} \right\};$$

All probability distributions over parameter space Θ

Role of PAC-Bayes:
what choices lead to what
generalisation guarantees?

Chapter 3

e.g., Knoblauch, Jewson, & Damoulas (2022)



Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to think about this?

A: (1) General Bayes Updates

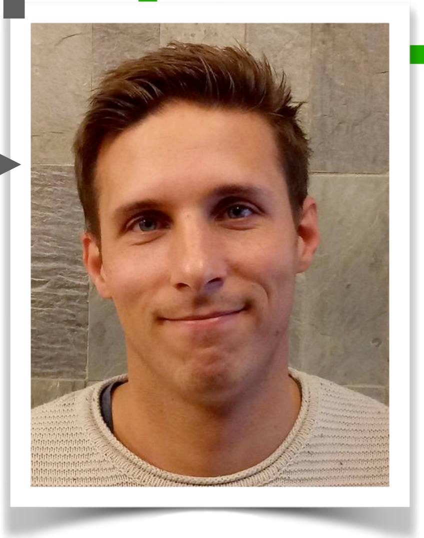
(2) Optimisation-centric view

NOT asked by PAC-Bayes:

When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

How should we design $\mathsf{L}(x_{1:n}, p_\theta)$?

What happens asymptotically?



Perspective 2: 'Optimisation-centric'

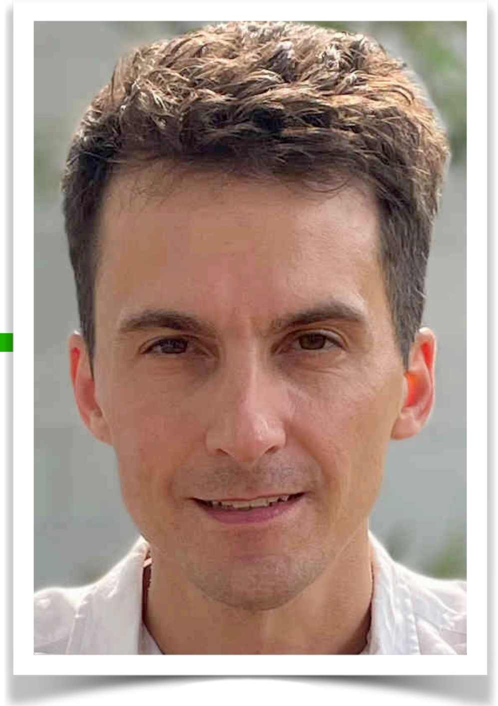
$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg\min_{q \in \mathcal{P}(\Theta)} \left\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda}\mathrm{KL}(q, \pi)}_{\text{Prior regularisation}} \right\};$$

All probability distributions over parameter space Θ

Role of PAC-Bayes: what choices lead to what generalisation guarantees?

Chapter 3

25/02



e.g., Knoblauch, Jewson, & Damoulas (2022)

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: for some small $\varepsilon \geq 0$,

Data-generating probability distribution

ε-contamination distribution

$$q_\varepsilon = (1-\varepsilon) \cdot q_0 \; + \; \varepsilon \cdot c$$

Part of distribution our model captures

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

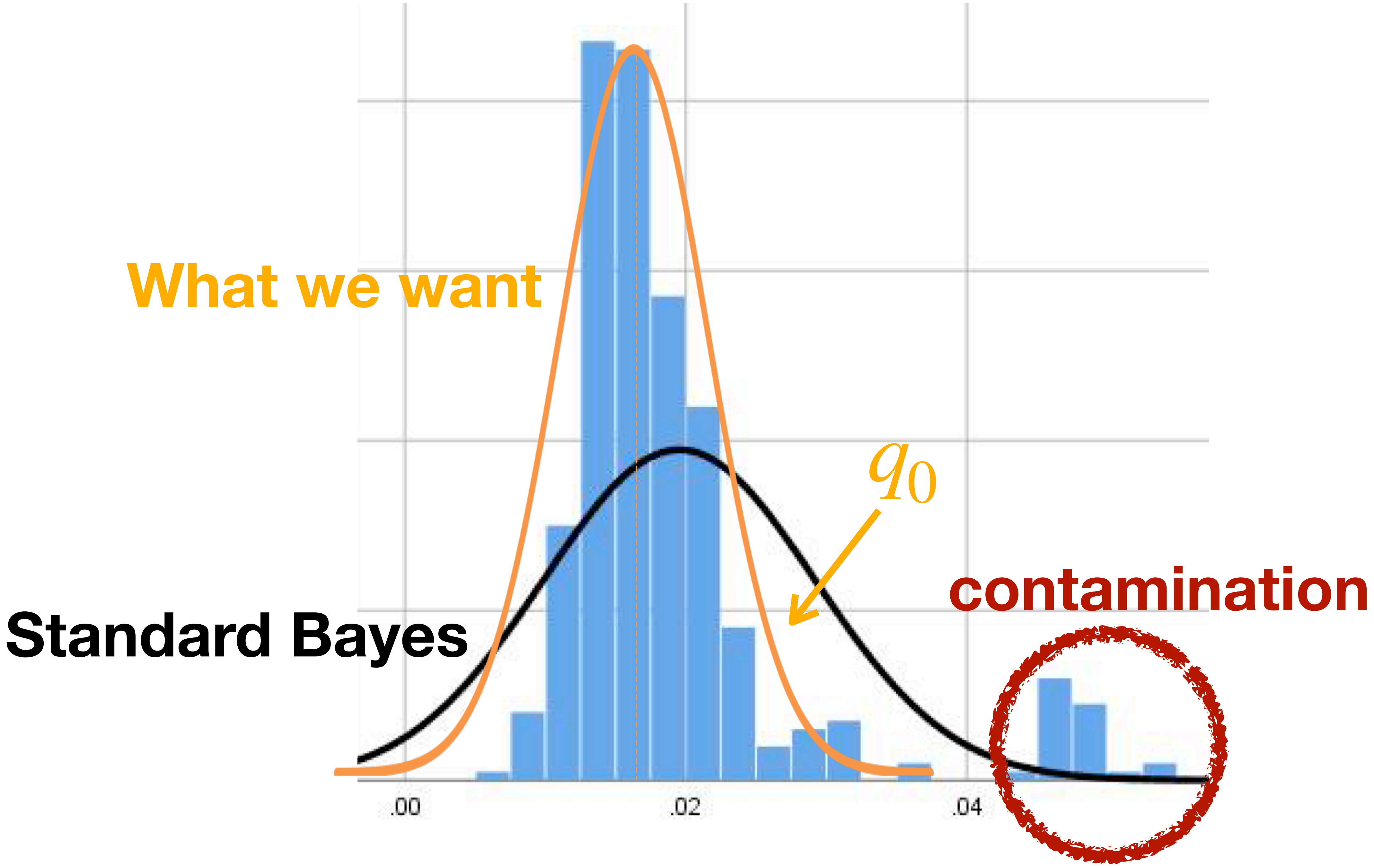
Setting: for some small $\varepsilon \geq 0$,

Data-generating probability distribution

ε **-contamination distribution**

$$q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$$

Part of distribution our model captures



Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

What we want:
$$\begin{cases} x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) \\ \qquad\qquad\qquad\qquad \approx \\ z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \end{cases}$$

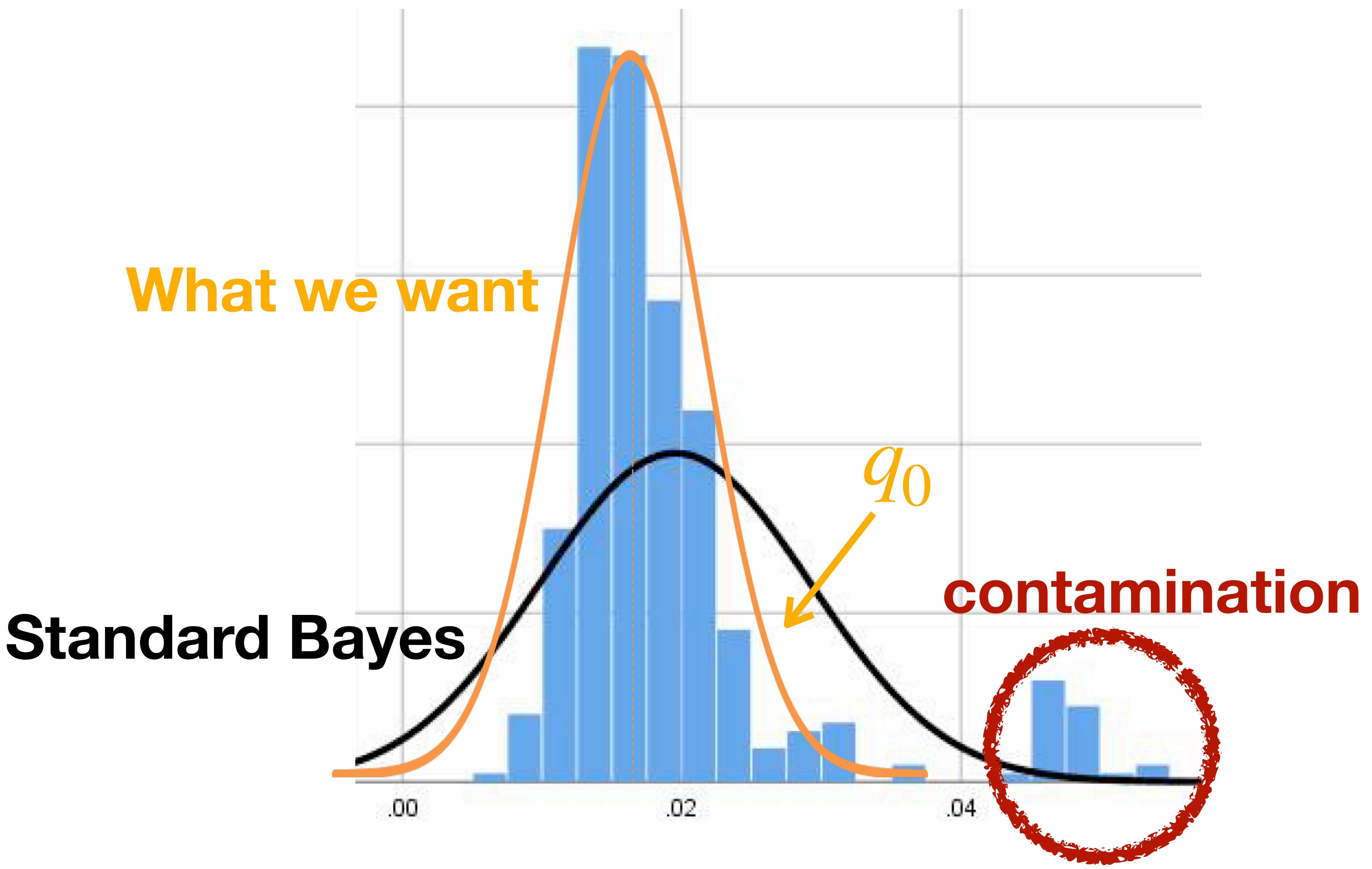
Setting: for some small $\varepsilon \geq 0$,

Data-generating probability distribution

ε-contamination distribution

$$q_\varepsilon = (1 - \varepsilon) \cdot q_0 + \varepsilon \cdot c$$

Part of distribution our model captures



Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: $q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$

$$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$$

$$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$$

Robustness: distance $\left\{\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})\right\} \leq$ constant$(c) \cdot \varepsilon$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: $q_\varepsilon = (1 - \varepsilon) \cdot q_0 + \varepsilon \cdot c$

$$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$$

$$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$$

Robustness: $$\sup_{c \in \mathcal{S}} \left\{ \text{distance} \left\{ \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: $q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$

$$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$$

$$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$$

Robustness: $\sup_{c \in \mathcal{S}} \left\{ \text{distance} \left\{ \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Key quantity: $\frac{1}{\varepsilon}\left[\mathsf{L}(p_\theta, x_{1:n}) - \mathsf{L}(p_\theta, z_{1:n})\right]$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: $q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$

$$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$$

$$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$$

Robustness: $$\sup_{c \in \mathcal{S}} \left\{ \text{distance}\left\{ \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Key quantity: $$\frac{\partial}{\partial \varepsilon} \mathsf{L}(p_\theta, x_{1:n}) \Big|_{\varepsilon=0} \approx \frac{1}{\varepsilon} \left[\mathsf{L}(p_\theta, x_{1:n}) - \mathsf{L}(p_\theta, z_{1:n}) \right]$$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: $q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$

$$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$$

$$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$$

Robustness: $\sup_{c \in \mathcal{S}} \left\{ \text{distance}\left\{ \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Loss robust!

Key quantity: $\boxed{\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \varepsilon} \mathsf{L}(p_\theta, x_{1:n}) \Big|_{\varepsilon=0} \right| < \infty}$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

Setting: $q_\varepsilon = (1 - \varepsilon) \cdot q_0 + \varepsilon \cdot c$

$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$

$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$

Theorem: $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ is robust over all $c \in \mathcal{S}$ if L is.

Robustness: $\sup_{c \in \mathcal{S}} \left\{ \text{distance} \left\{ \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Loss robust!

Key quantity: $\boxed{\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \varepsilon} \mathsf{L}(p_\theta, x_{1:n}) \Big|_{\varepsilon = 0} \right| < \infty}$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Setting: $q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$

$$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$$

$$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$$

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

A: Whenever $\mathsf{L}(x_{1:n}, p_\theta)$ is!

Theorem: $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ is robust over all $c \in \mathcal{S}$ if L is.

Robustness: $\sup_{c \in \mathcal{S}} \left\{ \text{distance}\left\{\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})\right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Loss robust!

Key quantity: $\boxed{\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \varepsilon} \mathsf{L}(p_\theta, x_{1:n}) \Big|_{\varepsilon=0} \right| < \infty}$

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Setting: $q_\varepsilon = (1-\varepsilon) \cdot q_0 + \varepsilon \cdot c$

$x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$

$z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})$

Q: When is $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ robust?

A: Whenever $\mathsf{L}(x_{1:n}, p_\theta)$ is!

Theorem: $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$ is robust over all $c \in \mathcal{S}$ if L is.

Robustness: $\sup_{c \in \mathcal{S}} \left\{ \text{distance}\left\{\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}), \pi_n^{\mathsf{L}}(\theta \mid z_{1:n})\right\} \right\} \leq \text{constant}(\mathcal{S}) \cdot \varepsilon$

$$= \sup_{\theta \in \Theta} \left| \pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) - \pi_n^{\mathsf{L}}(\theta \mid z_{1:n}) \right|$$

Key quantity: $\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \varepsilon} \mathsf{L}(p_\theta, x_{1:n}) \Big|_{\varepsilon=0} \right| < \infty$

Generally untrue for log likelihoods!

Ghosh & Basu (2015); AISM
Matsubara, **Knoblauch**, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & **Knoblauch** (2023); ICML
Altamirano, Briol, & **Knoblauch** (2024); ICML

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to design robust $\mathsf{L}(x_{1:n}, p_\theta)$ **?**

Hooker & Vidyashankar (2014); Test
Ghosh & Basu (2016); Statistica Sinica
Jewson, Smith, & Holmes (2018); Entropy
Knoblauch, Jewson, & Damoulas (2018); NeurIPS
Cherieff-Abdellatif & Alquier (2020); AABI
Matsubara, Knoblauch, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & Knoblauch (2023); ICML
Altamirano, Briol, & Knoblauch (2024); ICML
Matsubara, Knoblauch, Briol, & Oates (2023); JASA
Knoblauch*, Frazier*, & Drovandi (2024); preprint
Pacchiardi, Dhoo, & Dutta (2024); EJS
…

NOT robust to model misspecification

Standard Bayes

$$n \cdot \mathrm{KL}(q_\varepsilon, p(\cdot \mid \theta)) \overset{x_i \sim q_\varepsilon}{\approx} \mathsf{L}(x_{1:n}, p_\theta) = \sum_{i=1}^{n} -\log p(x_i \mid \theta)$$

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to design robust $\mathsf{L}(x_{1:n}, p_\theta)$?

Hooker & Vidyashankar (2014); Test
Ghosh & Basu (2016); Statistica Sinica
Jewson, Smith, & Holmes (2018); Entropy
Knoblauch, Jewson, & Damoulas (2018); NeurIPS
Cherieff-Abdellatif & Alquier (2020); AABI
Matsubara, Knoblauch, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & Knoblauch (2023); ICML
Altamirano, Briol, & Knoblauch (2024); ICML
Matsubara, Knoblauch, Briol, & Oates (2023); JASA
Knoblauch*, Frazier*, & Drovandi (2024); preprint
Pacchiardi, Dhoo, & Dutta (2024); EJS
…

NOT robust to model misspecification

$$n \cdot \mathrm{KL}(q_\varepsilon, p(\cdot \mid \theta)) \overset{x_i \sim q_\varepsilon}{\approx} \mathsf{L}(x_{1:n}, p_\theta) = \sum_{i=1}^{n} -\log p(x_i \mid \theta)$$

Standard Bayes

$$n \cdot \mathrm{D}(q_\varepsilon, p(\cdot \mid \theta))$$

Robust discrepancy

$$\boxed{\mathrm{D}(q_\varepsilon, p(\cdot \mid \theta)) \approx \mathrm{D}(q_0, p(\cdot \mid \theta))}$$

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Hooker & Vidyashankar (2014); Test
Ghosh & Basu (2016); Statistica Sinica
Jewson, Smith, & Holmes (2018); Entropy
Knoblauch, Jewson, & Damoulas (2018); NeurIPS
Cherieff-Abdellatif & Alquier (2020); AABI
Matsubara, Knoblauch, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & Knoblauch (2023); ICML
Altamirano, Briol, & Knoblauch (2024); ICML
Matsubara, Knoblauch, Briol, & Oates (2023); JASA
Knoblauch*, Frazier*, & Drovandi (2024); preprint
Pacchiardi, Dhoo, & Dutta (2024); EJS
…

Q: How to design robust $\mathsf{L}(x_{1:n}, p_\theta)$?
A: estimate robust divergence

NOT robust to model misspecification

$$n \cdot \mathrm{KL}(q_\varepsilon, p(\cdot \mid \theta)) \overset{x_i \sim q_\varepsilon}{\approx} \mathsf{L}(x_{1:n}, p_\theta) = \sum_{i=1}^{n} -\log p(x_i \mid \theta)$$

Standard Bayes

$$n \cdot \mathrm{D}(q_\varepsilon, p(\cdot \mid \theta)) \overset{x_i \sim q_\varepsilon}{\approx} \mathsf{L}(x_{1:n}, p_\theta)$$

Robust discrepancy

Robust loss

$$\boxed{\mathrm{D}(q_\varepsilon, p(\cdot \mid \theta)) \approx \mathrm{D}(q_0, p(\cdot \mid \theta))} \dashrightarrow \mathsf{L} \text{ is robust over all } c \in \mathcal{S}$$

Basics: generalised / Gibbs posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: How to design robust $\mathsf{L}(x_{1:n}, p_\theta)$?

A: estimate robust divergence

Hooker & Vidyashankar (2014); Test
Ghosh & Basu (2016); Statistica Sinica
Jewson, Smith, & Holmes (2018); Entropy
Knoblauch, Jewson, & Damoulas (2018); NeurIPS
Cherieff-Abdellatif & Alquier (2020); AABI
Matsubara, Knoblauch, Briol, & Oates (2022); JRSS-B
Altamirano, Briol, & Knoblauch (2023); ICML
Altamirano, Briol, & Knoblauch (2024); ICML
Matsubara, Knoblauch, Briol, & Oates (2023); JASA
Knoblauch*, Frazier*, & Drovandi (2024); preprint
Pacchiardi, Dhoo, & Dutta (2024); EJS
…

NOT robust to model misspecification

Standard Bayes

$$n \cdot \mathrm{KL}(q_\varepsilon, p(\cdot \mid \theta)) \overset{x_i \sim q_\varepsilon}{\approx} \mathsf{L}(x_{1:n}, p_\theta) = \sum_{i=1}^{n} -\log p(x_i \mid \theta)$$

$$n \cdot \mathrm{D}(q_\varepsilon, p(\cdot \mid \theta)) \overset{x_i \sim q_\varepsilon}{\approx} \mathsf{L}(x_{1:n}, p_\theta)$$

Robust discrepancy

Robust loss

$$\boxed{\mathrm{D}(q_\varepsilon, p(\cdot \mid \theta)) \approx \mathrm{D}(q_0, p(\cdot \mid \theta))} \longrightarrow \mathsf{L} \text{ is robust over all } c \in \mathcal{S}$$

Examples:

MMD
$\alpha/\beta/\gamma$-divergences
Stein discrepancies
…

Basics: optimisation-centric posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: Can we generalise this further?

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg \min_{q \in \mathscr{P}(\Theta)} \Big\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda}\mathrm{KL}(q, \pi)}_{\text{Prior regularisation}} \Big\}$$

All probability distributions over parameter space Θ

Variational Inference

Robustness to misspecified prior

$$q_n^*(\theta) = \arg \min_{q \in \mathscr{Q}} \Big\{ \mathscr{L}(q, x_{1:n}) + \frac{1}{\lambda} \mathrm{D}(q, \pi) \Big\}$$

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

Knoblauch, Jewson, & Damoulas (2022)

Basics: optimisation-centric posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: Can we generalise this further?
A: Yes! In many, many ways

Jankowiak, Pleiss, & Gardner (2020); ICML
Jankowiak, Pleiss, & Gardner (2020); UAI
Alquier (2020); ICML
Wild, Wu, & Sejdinovic (2022); NeurIPS
Javeed, Kouri, & Surowiec (2023); arXiv preprint
Wild, Ghalebikesabi, Sejdinovic, & Knoblauch (2023); NeurIPS
Shen, Knoblauch, Power, & Oates (2024); AISTATS
…

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg\min_{q \in \mathcal{P}(\Theta)} \Big\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda}\mathrm{KL}(q, \pi)}_{\text{Prior regularisation}} \Big\}$$

All probability distributions over parameter space Θ

Variational Inference

Robustness to misspecified prior

$$q_n^*(\theta) = \arg\min_{q \in \mathcal{Q}} \Big\{ \mathcal{L}(q, x_{1:n}) + \frac{1}{\lambda}\mathrm{D}(q, \pi) \Big\}$$

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

Knoblauch, Jewson, & Damoulas (2022)

Basics: optimisation-centric posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: Can we generalise this further?
A: Yes! In many, many ways

Jankowiak, Pleiss, & Gardner (2020); ICML
Jankowiak, Pleiss, & Gardner (2020); UAI
Alquier (2020); ICML
Wild, Wu, & Sejdinovic (2022); NeurIPS
Javeed, Kouri, & Surowiec (2023); arXiv preprint
Wild, Ghalebikesabi, Sejdinovic, & Knoblauch (2023); NeurIPS
Shen, Knoblauch, Power, & Oates (2024); AISTATS
…

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg \min_{q \in \mathscr{P}(\Theta)} \left\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda} \mathrm{KL}\left(q, \pi\right)}_{\text{Prior regularisation}} \right\}$$

All probability distributions over parameter space Θ

Variational Inference

25/03

Robustness to misspecified prior

$$q_n^*(\theta) = \arg \min_{q \in \mathscr{Q}} \left\{ \mathscr{L}(q, x_{1:n}) + \frac{1}{\lambda} \mathrm{D}\left(q, \pi\right) \right\}$$

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

e.g. $\mathscr{L}\left(\int p_\theta(\,\cdot\,)\, \mathrm{d}q(\theta)\ ,\ x_{1:n}\right)$

Knoblauch, Jewson, & Damoulas (2022)

Basics: optimisation-centric posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: Can we generalise this further?
A: Yes! In many, many ways

Jankowiak, Pleiss, & Gardner (2020); ICML
Jankowiak, Pleiss, & Gardner (2020); UAI
Alquier (2020); ICML
Wild, Wu, & Sejdinovic (2022); NeurIPS
Javeed, Kouri, & Surowiec (2023); arXiv preprint
Wild, Ghalebikesabi, Sejdinovic, & Knoblauch (2023); NeurIPS
Shen, Knoblauch, Power, & Oates (2024); AISTATS
...

Computation:

$\mathcal{Q} =$ parametric $\implies$ generalised VI

$\mathcal{Q} = \mathscr{P}_2(\Theta) \implies$ Wasserstein Gradient Flow

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg\min_{q \in \mathscr{P}(\Theta)} \Big\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda} \mathrm{KL}(q, \pi)}_{\text{Prior regularisation}} \Big\}$$

All probability distributions over parameter space Θ

Variational Inference

25/03

Robustness to misspecified prior

$$q_n^*(\theta) = \arg\min_{q \in \mathcal{Q}} \Big\{ \mathscr{L}(q, x_{1:n}) + \frac{1}{\lambda} \mathrm{D}(q, \pi) \Big\}$$

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

e.g. $\mathscr{L}\left(\int p_\theta(\,\cdot\,)\, \mathrm{d}q(\theta)\, ,\, x_{1:n} \right)$

Knoblauch, Jewson, & Damoulas (2022)

Basics: optimisation-centric posteriors

Gibbs/Generalised/Pseudo Posterior

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

Q: Can we generalise this further?
A: Yes! In many, many ways

Jankowiak, Pleiss, & Gardner (2020); ICML
Jankowiak, Pleiss, & Gardner (2020); UAI
Alquier (2020); ICML
Wild, Wu, & Sejdinovic (2022); NeurIPS
Javeed, Kouri, & Surowiec (2023); arXiv preprint
Wild, Ghalebikesabi, Sejdinovic, & Knoblauch (2023); NeurIPS
Shen, Knoblauch, Power, & Oates (2024); AISTATS
...

Computation:

$\mathcal{Q} =$ parametric $\Longrightarrow$ generalised VI

$\mathcal{Q} = \mathcal{P}_2(\Theta)$ $\Longrightarrow$ Wasserstein Gradient Flow

First Sampler of its kind! + 'Morality Tale'

Perspective 2: 'Optimisation-centric'

$$\pi_n^{\mathsf{L}}(\theta \mid x_{1:n}) = \arg\min_{q \in \mathcal{P}(\Theta)} \Big\{ \underbrace{\mathbb{E}_{\theta \sim q}\left[\mathsf{L}(x_{1:n}, p_\theta)\right]}_{\text{Data-fitting}} + \underbrace{\frac{1}{\lambda}\mathrm{KL}(q, \pi)}_{\text{Prior regularisation}} \Big\}$$

All probability distributions over parameter space Θ

Variational Inference

25/03

Robustness to misspecified prior

$$q_n^*(\theta) = \arg\min_{q \in \mathcal{Q}} \Big\{ \mathcal{L}(q, x_{1:n}) + \frac{1}{\lambda}\mathrm{D}(q, \pi) \Big\}$$

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

e.g. $\mathcal{L}\left(\int p_\theta(\,\cdot\,)\, dq(\theta)\;, x_{1:n}\right)$

Knoblauch, Jewson, & Damoulas (2022)

Morality Tale: Why Post-Bayesian thinking is needed

$$q_n^*(\theta) \quad = \arg\min_{q \in \mathcal{Q}} \left\{ \mathcal{L}(q, x_{1:n}) + \frac{1}{\lambda} \mathrm{D}\big(q, \pi\big) \right\}$$

Objective: $q \mapsto \mathbb{E}_{\theta \sim q}\left[-\log p(x_{1:n} \mid \theta)\right] + \frac{1}{\lambda}\mathrm{KL}(q, \pi)$

Target: **Cold Posterior** $(\lambda \gg 1)$

/ Bayes Posterior $(\lambda = 1)$

Wild, Ghalebikesabi, Sejdinovic, & **Knoblauch** (2023); NeurIPS Oral

Morality Tale: Why Post-Bayesian thinking is needed

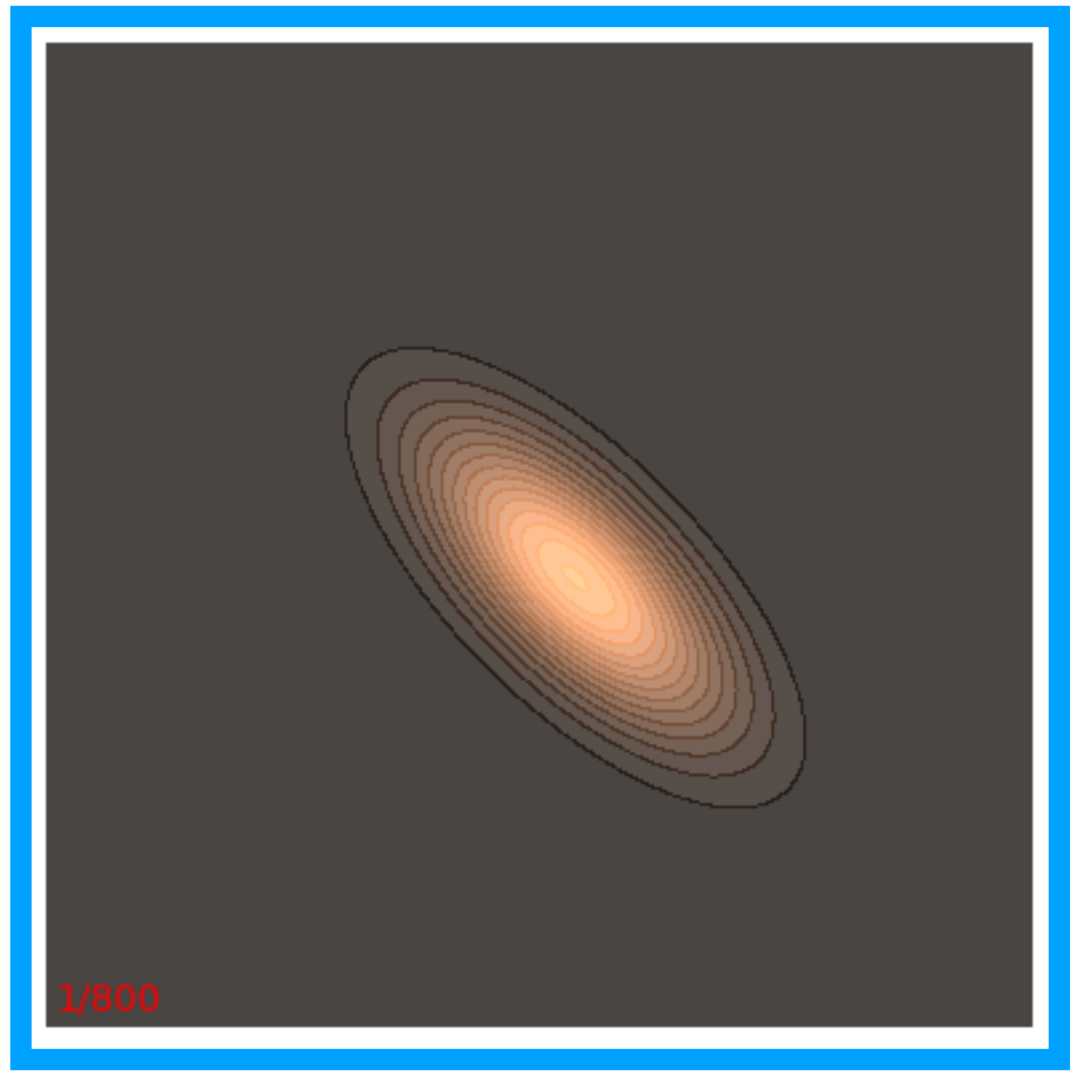
$$q_n^*(\theta) \quad = \arg\min_{q \in \mathcal{Q}} \left\{ \mathcal{L}(q, x_{1:n}) + \frac{1}{\lambda} \mathrm{D}\left(q, \pi\right) \right\}$$

Objective: $q \mapsto \mathbb{E}_{\theta \sim q}\left[-\log p(x_{1:n} \mid \theta)\right] + \frac{1}{\lambda}\mathrm{KL}(q, \pi)$

Target: **Cold Posterior** $(\lambda \gg 1)$

/ Bayes Posterior $(\lambda = 1)$

Wasserstein Gradient Flow = Langevin Diffusion



Converges to well-defined density

$$q_n^*(\theta) = \pi_n^{(\lambda)}(\theta \mid x_{1:n})$$

Wild, Ghalebikesabi, Sejdinovic, & **Knoblauch** (2023); NeurIPS Oral

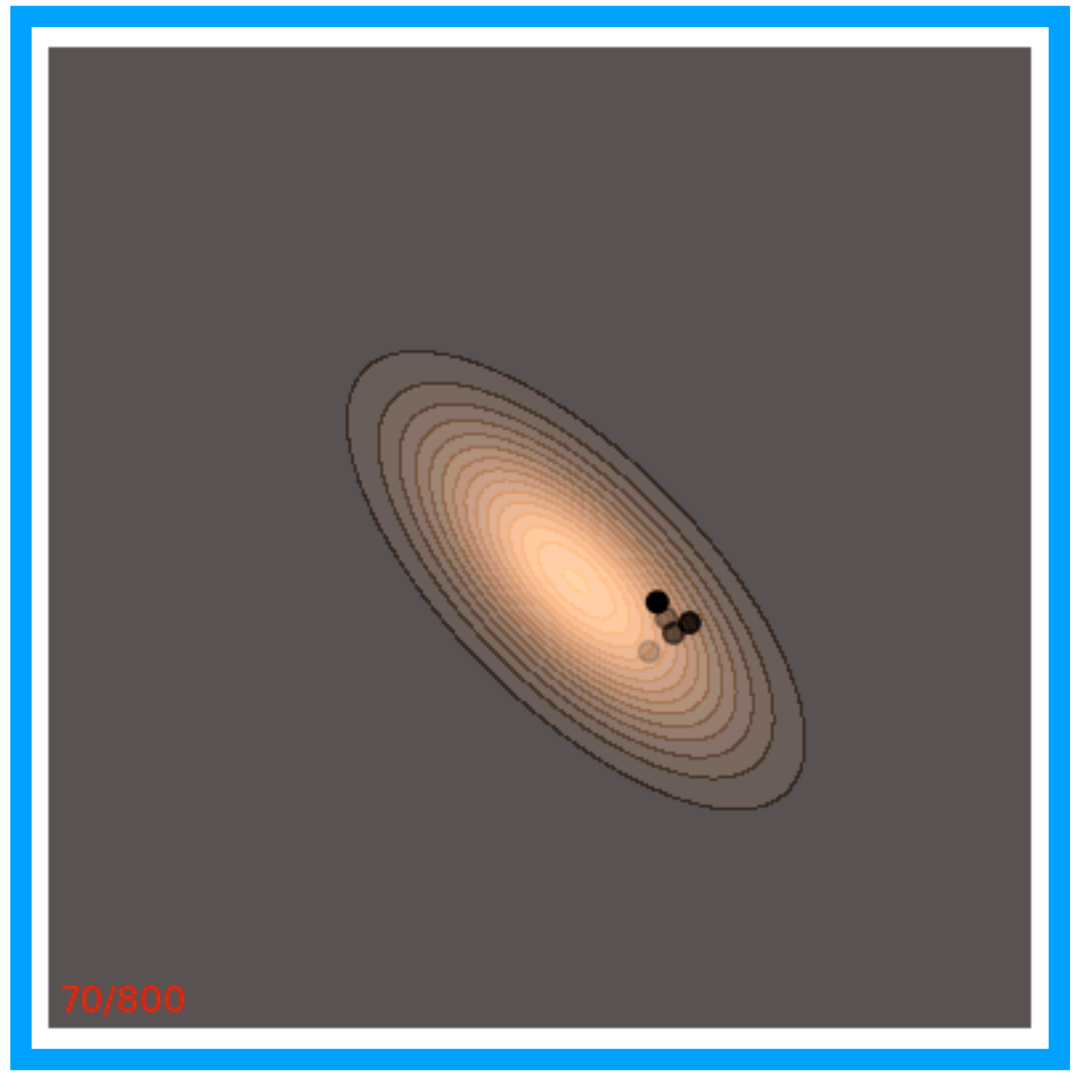
Morality Tale: Why Post-Bayesian thinking is needed

$$\boxed{q_n^*(\theta) \quad = \underset{q \in \mathcal{Q}}{\arg\min} \left\{ \mathcal{L}(q, x_{1:n}) + \frac{1}{\lambda} \mathrm{D}(q, \pi) \right\}}$$

Objective: $q \mapsto \mathbb{E}_{\theta \sim q}\left[-\log p(x_{1:n} \mid \theta)\right] + \frac{1}{\lambda}\mathrm{KL}(q, \pi) \xrightarrow{\lambda \to \infty} q \mapsto \mathbb{E}_{\theta \sim q}\left[-\log p(x_{1:n} \mid \theta)\right]$

Target: **Cold Posterior** $(\lambda \gg 1)$ **/ Bayes Posterior** $(\lambda = 1)$ — **Deep Ensemble (DE)** $(\lambda \to \infty)$

Wasserstein Gradient Flow = Langevin Diffusion

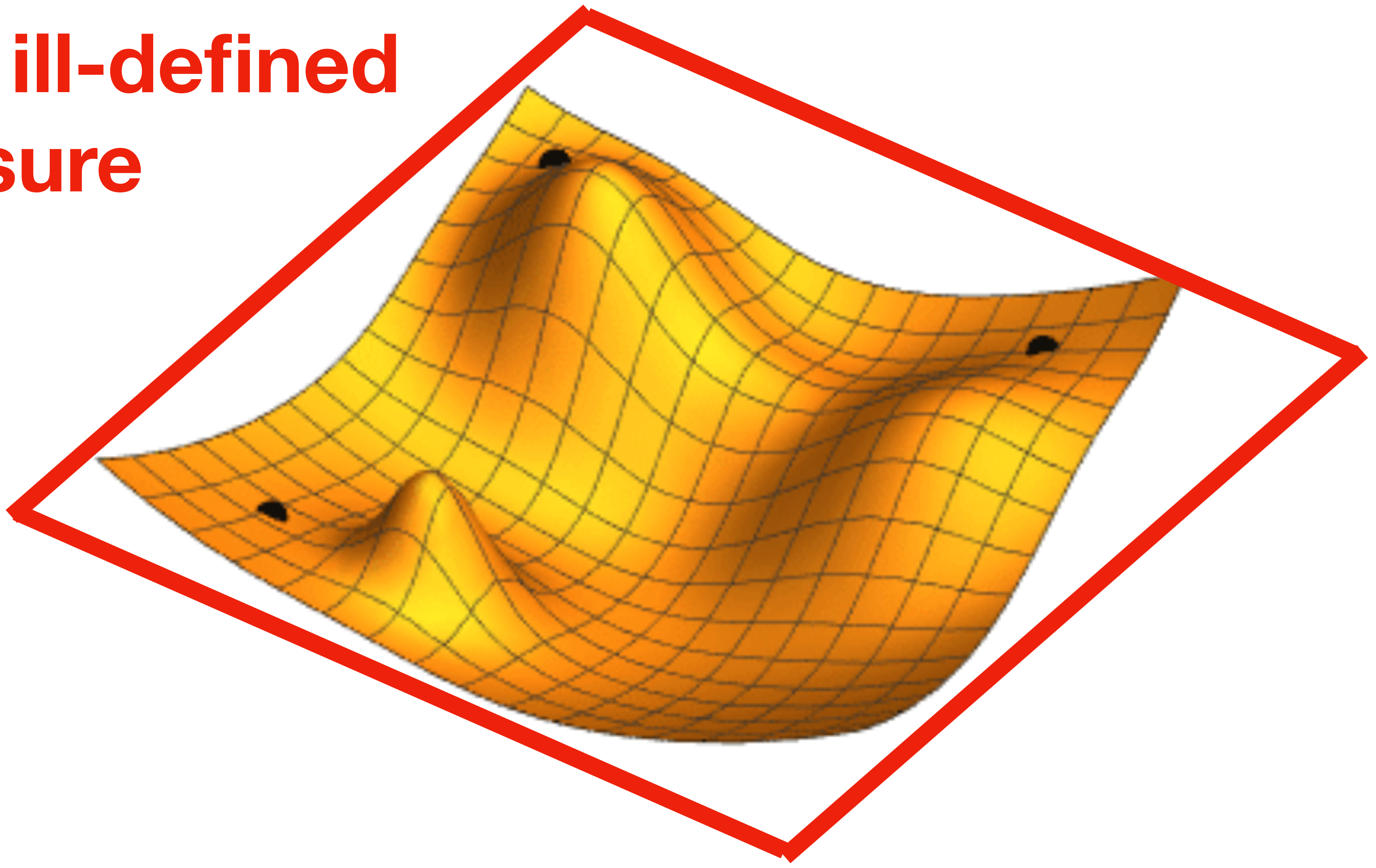


Converges to well-defined density

$$q_n^*(\theta) = \pi_n^{(\lambda)}(\theta \mid x_{1:n})$$

Wasserstein Gradient Flow = DE

Converges to ill-defined discrete measure



Wild, Ghalebikesabi, Sejdinovic, & **Knoblauch** (2023); NeurIPS Oral

Morality Tale: Why Post-Bayesian thinking is needed

Claim: ‘Deep Ensembles = Bayesian Inference’

[…] Deep ensembles (Lakshminarayanan et al., 2017) are not a competing approach to Bayesian inference, but […] a compelling mechanism for Bayesian marginalization.

Published 2020 @ NeurIPS
(cited ≈ 800 times according to Google scholar)

Morality Tale: Why Post-Bayesian thinking is needed

Claim: ‘Deep Ensembles = Bayesian Inference’

[…] Deep ensembles (Lakshminarayanan et al., 2017) are not a competing approach to Bayesian inference, but […] a compelling mechanism for Bayesian marginalization.

Published 2020 @ NeurIPS
(cited ≈ 800 times according to Google scholar)

Unfortunately, this is not correct.

Morality Tale: Why Post-Bayesian thinking is needed

III

I. In practice, orthodox Bayesianism has already been abandoned
(Bayes posterior: prior regulariser, densities ;
Deep Ensembles: no prior regulariser, discrete measures)

Morality Tale: Why Post-Bayesian thinking is needed

III

I. In practice, orthodox Bayesianism has already been abandoned
(Bayes posterior: prior regulariser, densities ;
Deep Ensembles: no prior regulariser, discrete measures)

II. But practicioners often don't realise this / pay attention to the ramifications, which in turns leads to incorrect claims and conclusions.
('Deep Ensembles are Bayesian')

Morality Tale: Why Post-Bayesian thinking is needed

III

I. In practice, orthodox Bayesianism has already been abandoned
(Bayes posterior: prior regulariser, densities ;
Deep Ensembles: no prior regulariser, discrete measures)

II. But practicioners often don't realise this / pay attention to the ramifications, which in turns leads to incorrect claims and conclusions.
('Deep Ensembles are Bayesian')

III. It's on us to help them! We need to:
1. acknowledge that post-Bayesian methods are already in use; and
2. develop clear formalisms and design principles for them.

Ideas either adapt, or die.