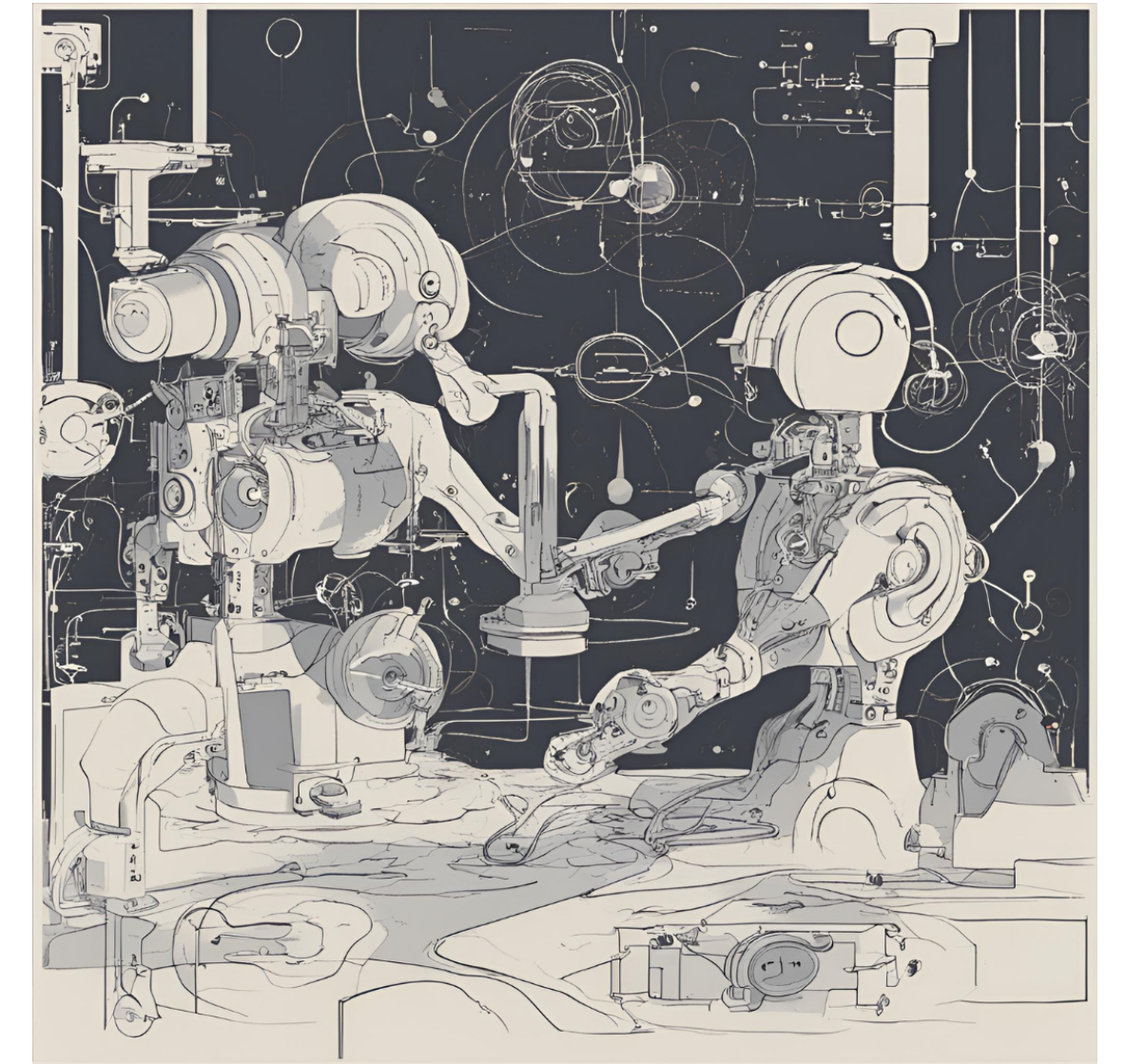


Post-Bayesian Inference

@ ISNET-11

Jeremias Knoblauch;
Associate Prof @ UCL Stats

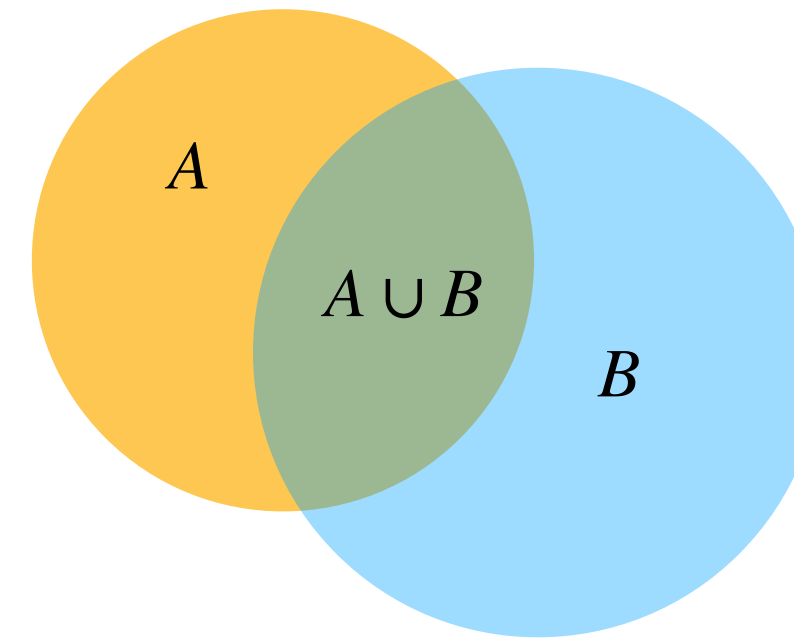


17/11/25

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

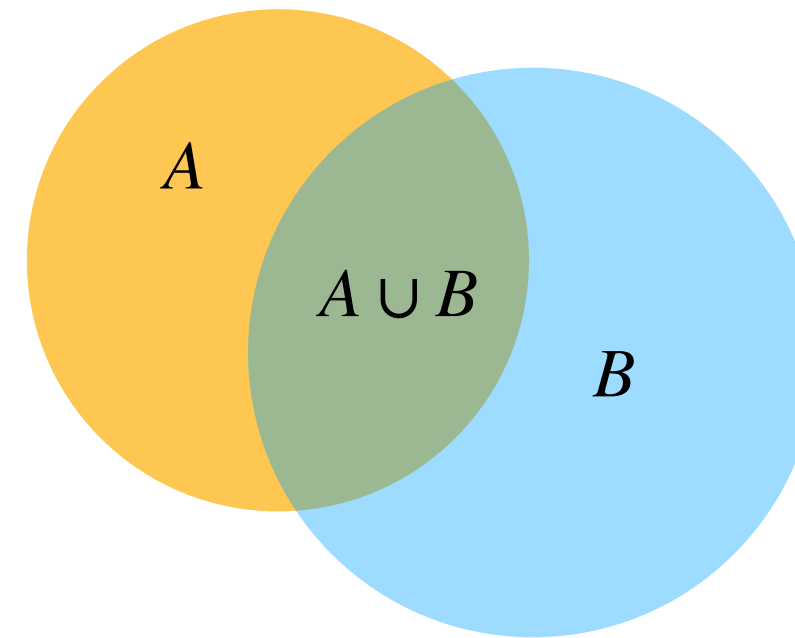
$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$



Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

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



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

$p(x_{1:n} | \theta)$

Prior probability: $\pi(\theta)$
 $\theta \in \Theta$

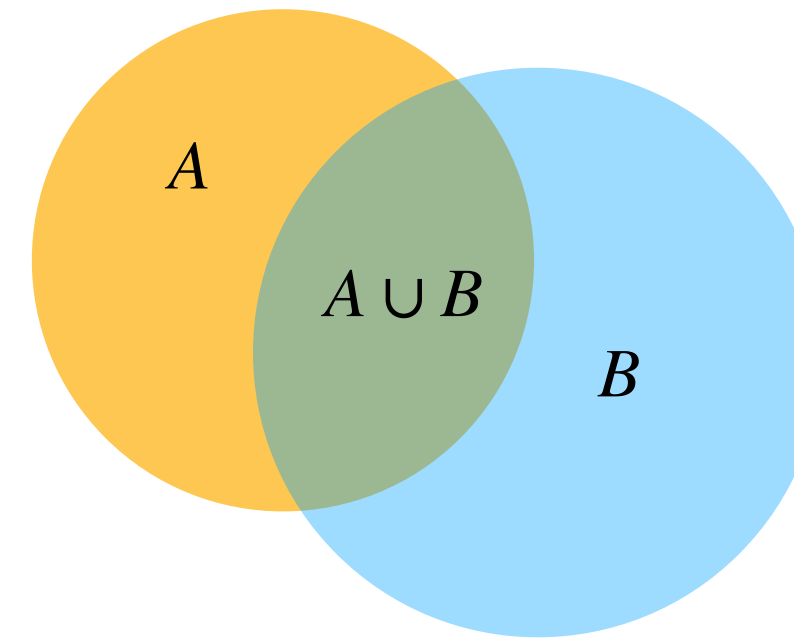
$$\pi_n(\theta | x_{1:n}) = \frac{p(x_{1:n} | \theta) \cdot \pi(\theta)}{\int p(x_{1:n} | \theta) \cdot \pi(\theta) d\theta}$$

(Bayes) Posterior

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

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



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

$p(x_{1:n} | \theta)$

Prior probability: $\pi(\theta)$
 $\theta \in \Theta$

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

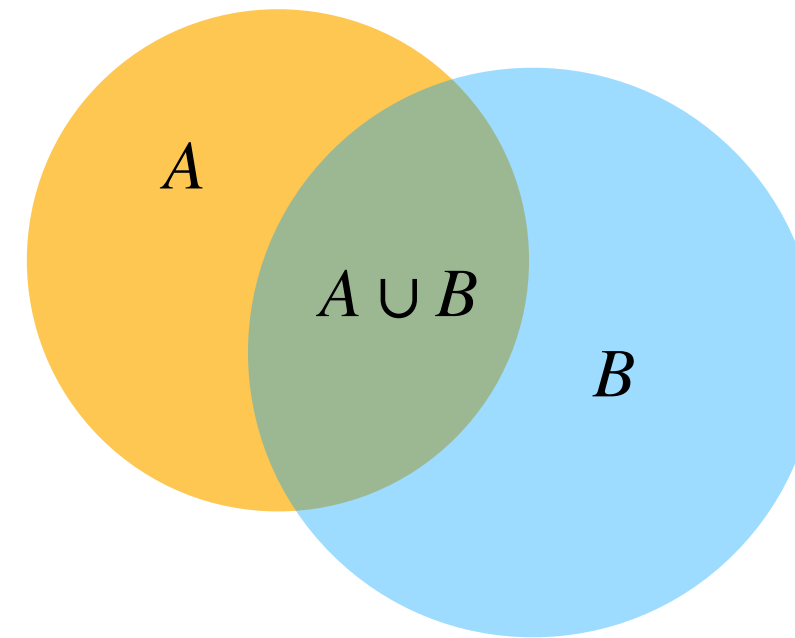
(Bayes) Posterior

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

Preamble: Bayesian Data Analysis

Bayes' Theorem: Inversion of conditionals

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



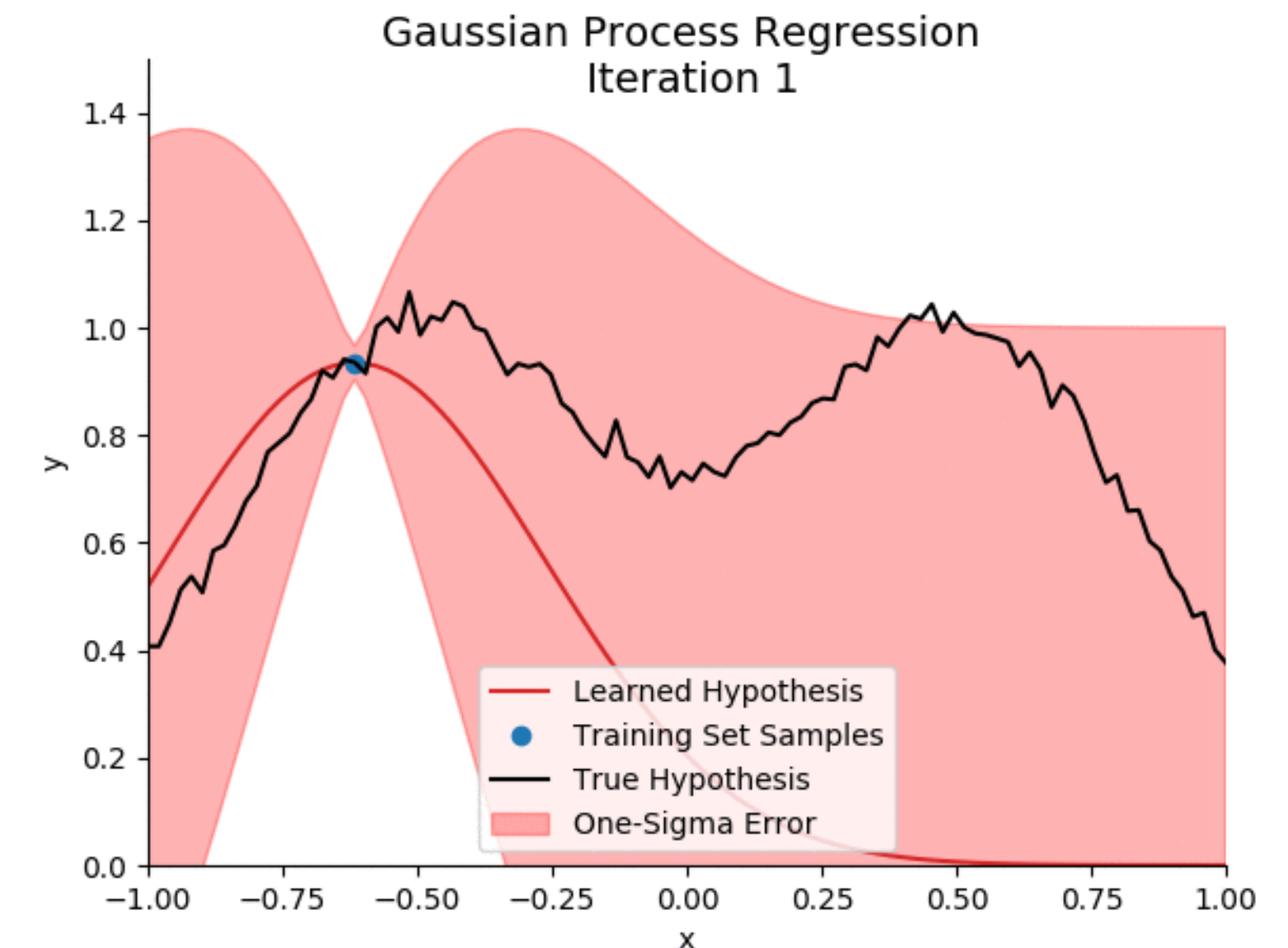
Data model:
 $x_{1:n} \in \mathcal{X}^n$

$p(x_{1:n} | \theta)$

Prior probability: $\pi(\theta)$
 $\theta \in \Theta$

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

(Bayes) Posterior



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

Problematic Assumptions for Bayesian Analysis ⓘ

(A1)

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

Θ = Only relevant State of the world

Problematic Assumptions for Bayesian Analysis ⓘ

(A1)

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

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

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 ①

- (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

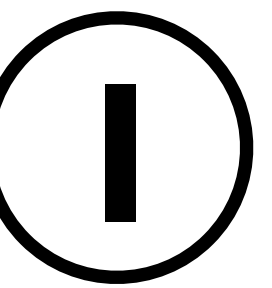
How rational decision-makers choose the prior

(A3)

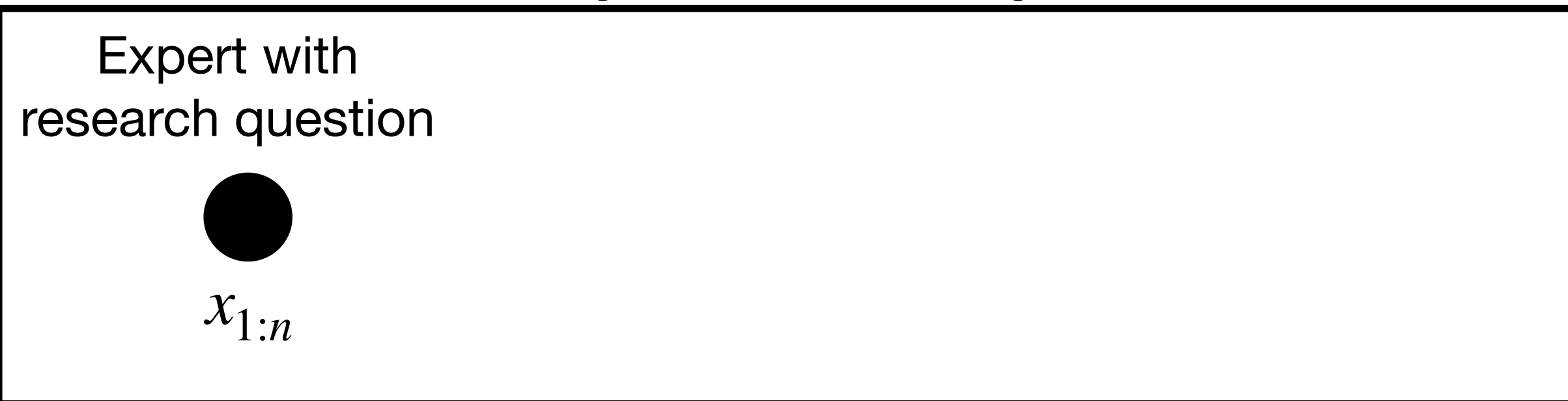
$\pi_n(\theta \mid x_{1:n})$ computable in practice

Guarantees real-world relevance

Case Study: Bayesian ML & Boston Housing Data

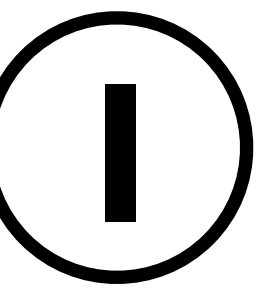


Traditional Bayesian analysis in science

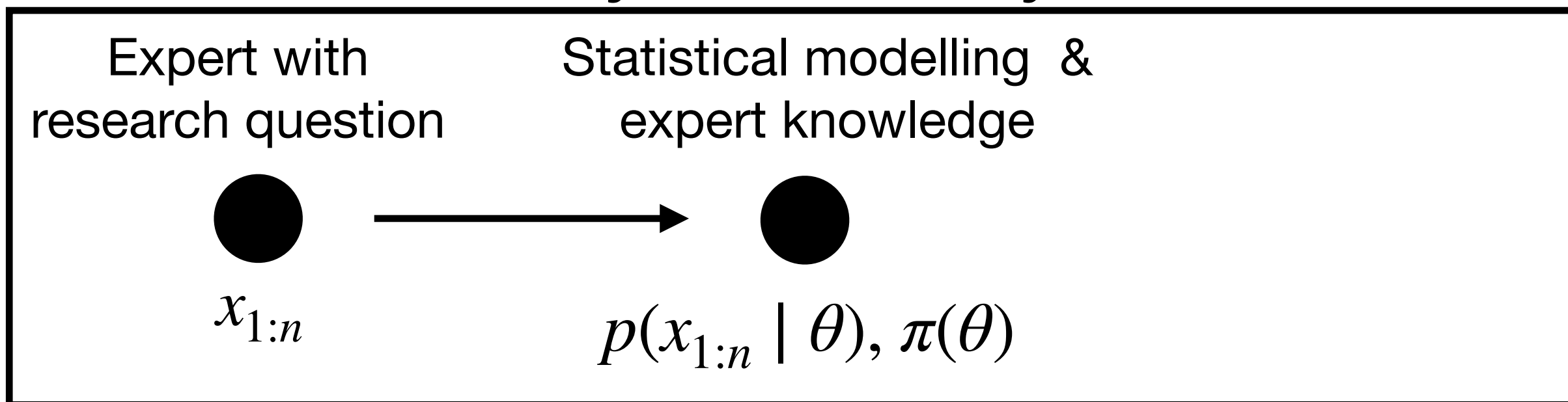


- (A1) model well-specified
- (A2) prior well-specified
- (A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

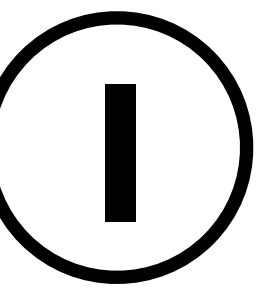


Traditional Bayesian analysis in science

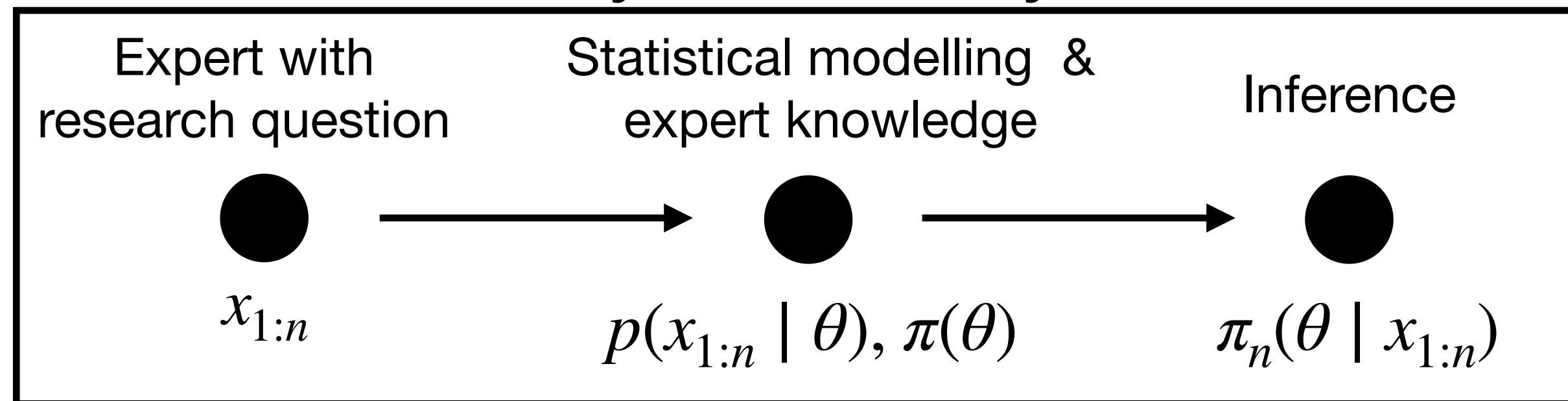


- (A1) model well-specified
- (A2) prior well-specified
- (A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data

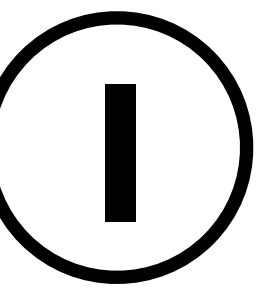


Traditional Bayesian analysis in science

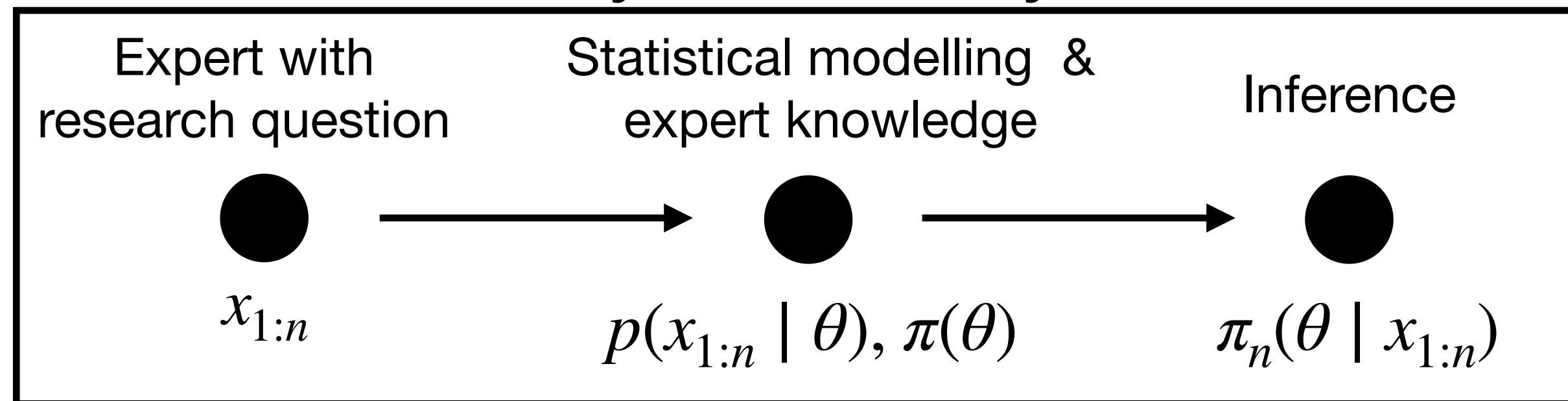


- (A1) model well-specified
- (A2) prior well-specified
- (A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science

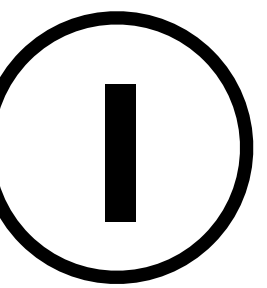


Harrison & Rubinfeld (1978)

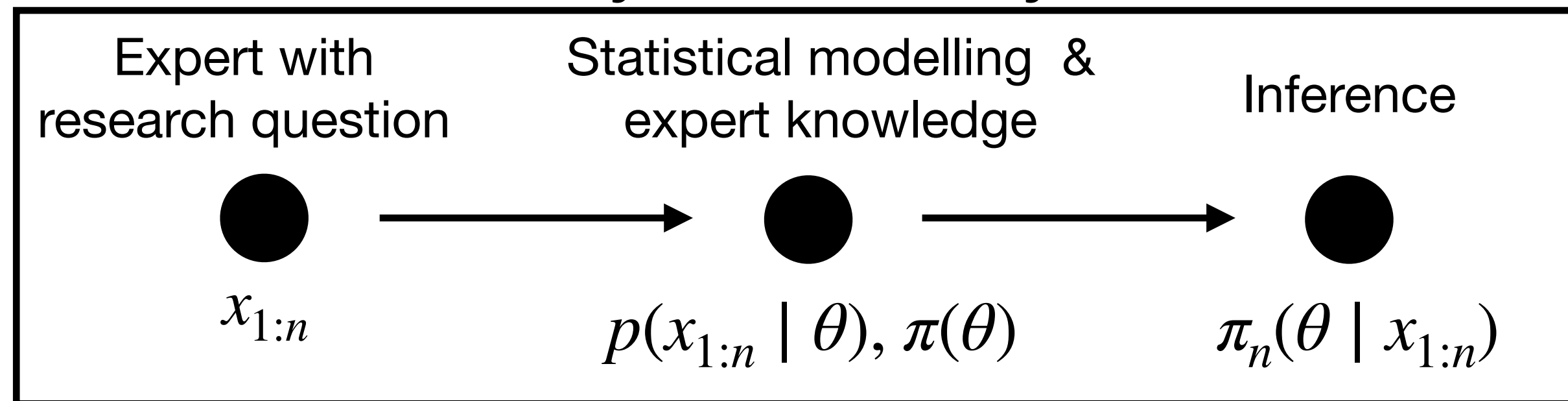
Research Question: influence of air pollution on house prices?

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

Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

↑ ↑ ↑ ↑

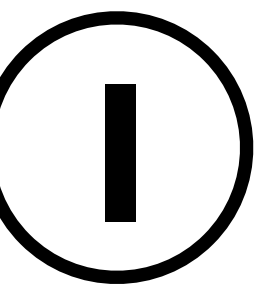
willingness to pay pollutants rooms, sqm, ... measurement error

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

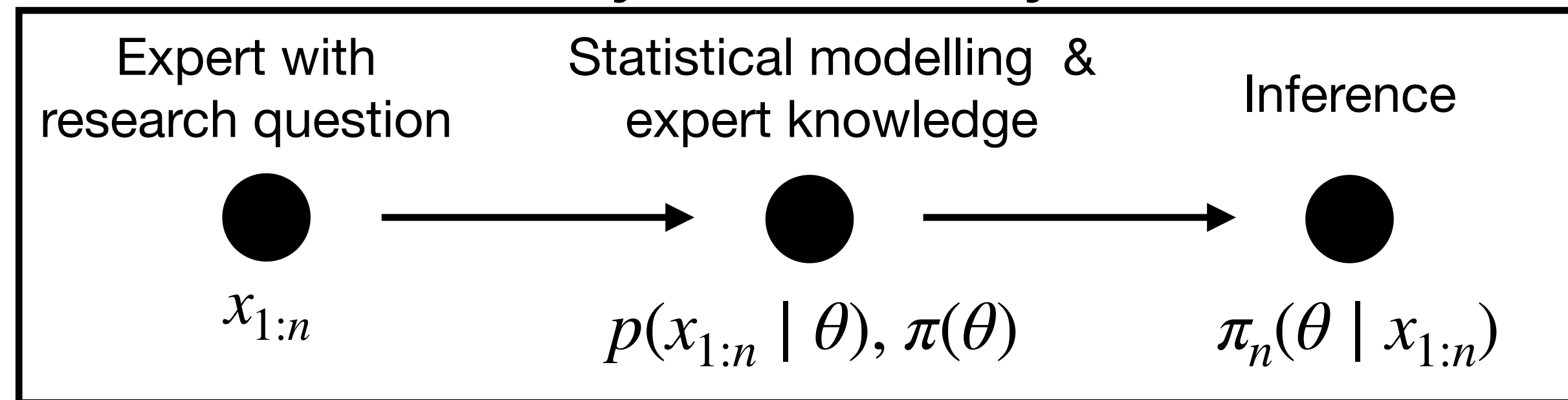
parameters of interest incidental parameters

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

Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

↑ ↑ ↑ ↑

willingness to pay pollutants rooms, sqm, ... measurement error

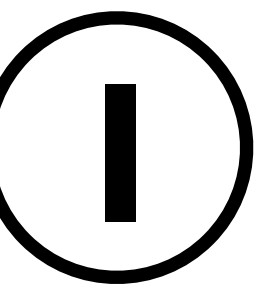
$\theta = (c_0, c_2, \dots, c_{J_1}, p_1, p_2, \dots, p_{J_2})^\top$

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

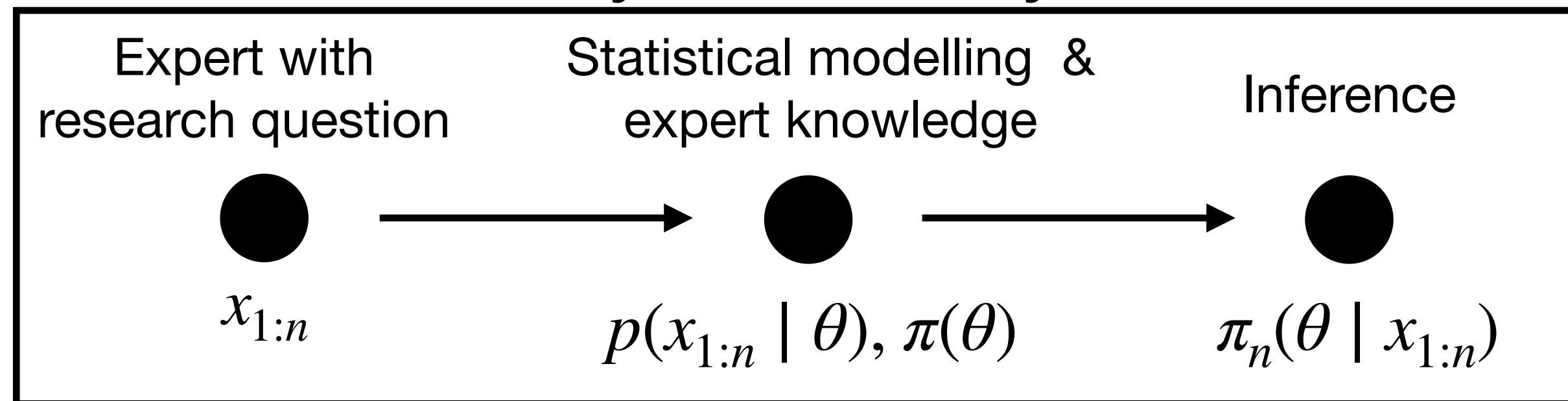
(A2) ✓

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

Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

↑ ↑ ↑ ↑

willingness to pay pollutants rooms, sqm, ... measurement error

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

parameters of interest incidental parameters

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

$\pi_n(\theta | x_{1:n}) \longrightarrow$ computed exactly

(A2)

(A3)

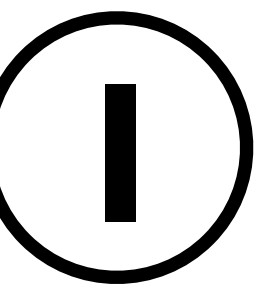


(A1) model well-specified

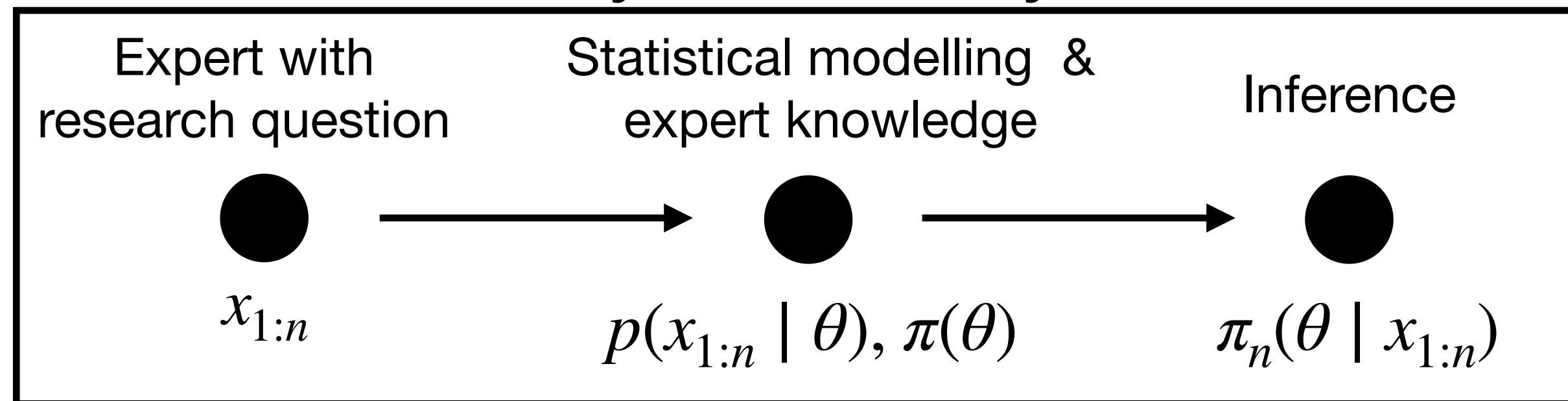
(A2) prior well-specified

(A3) computationally feasible

Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Modern Bayesian ML



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

$\uparrow$ willingness to pay $\uparrow$ pollutants $\uparrow$ rooms, sqm, ... $\uparrow$ measurement error

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

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

$\pi_n(\theta | x_{1:n}) \rightarrow$ computed exactly

(A2) ✓

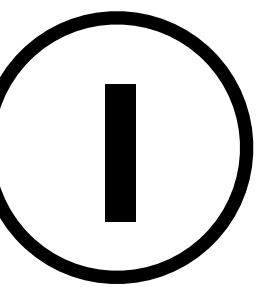
(A3) ✓

(A1) model well-specified

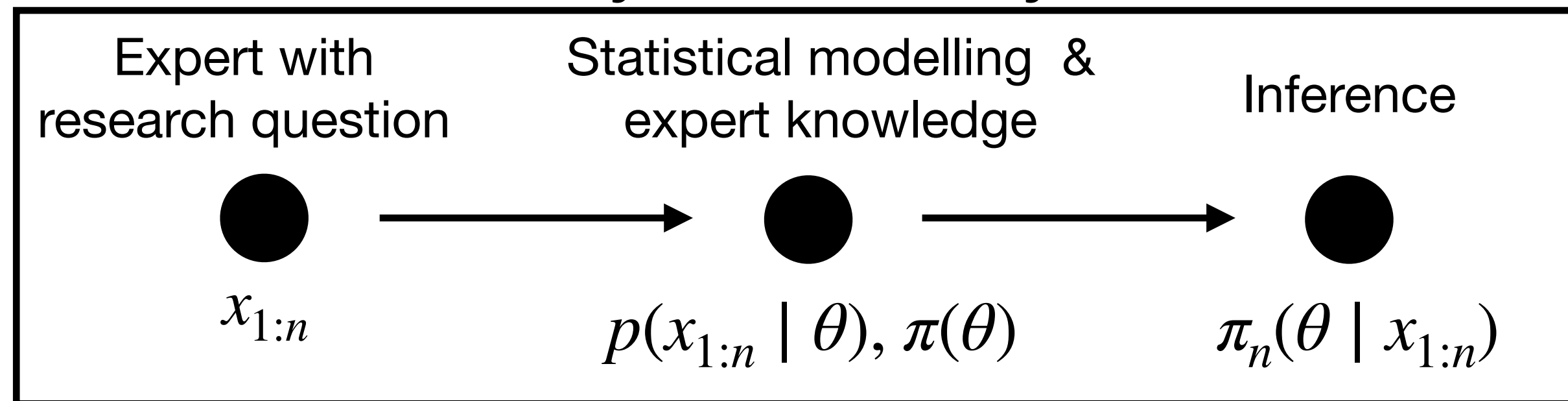
(A2) prior well-specified

(A3) computationally feasible

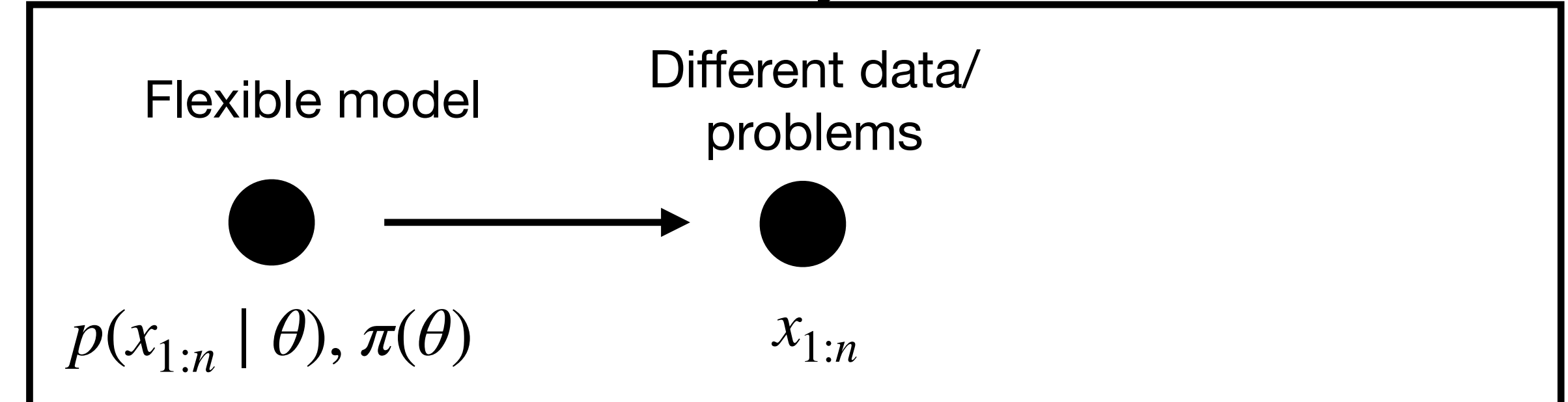
Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Modern Bayesian ML



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

$\uparrow$ willingness to pay $\uparrow$ pollutants $\uparrow$ rooms, sqm, ... $\uparrow$ measurement error

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

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

$\pi_n(\theta | x_{1:n}) \rightarrow$ computed exactly

(A2) ✓

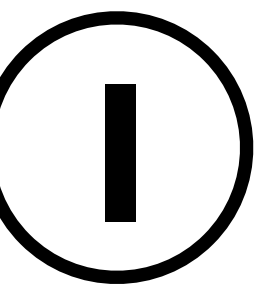
(A3) ✓

(A1) model well-specified

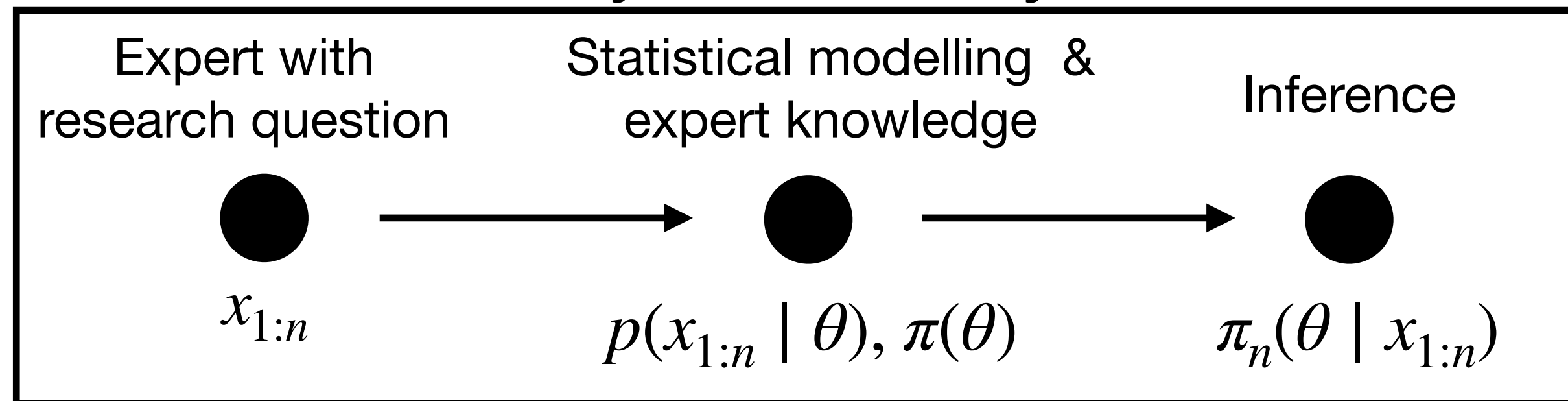
(A2) prior well-specified

(A3) computationally feasible

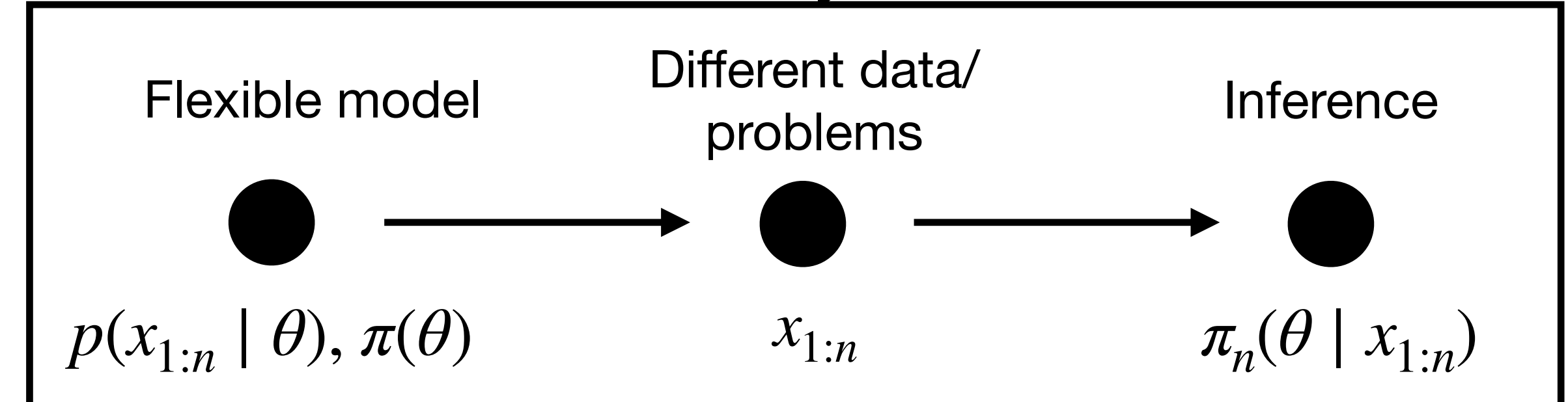
Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Modern Bayesian ML



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

↑ ↑ ↑ ↑

willingness to pay pollutants rooms, sqm, ... measurement error

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

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

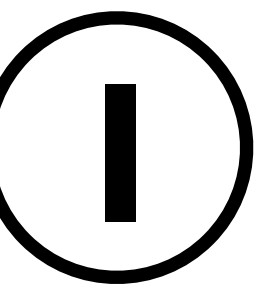
$\pi_n(\theta | x_{1:n}) \longrightarrow$ computed exactly

(A2) ✓

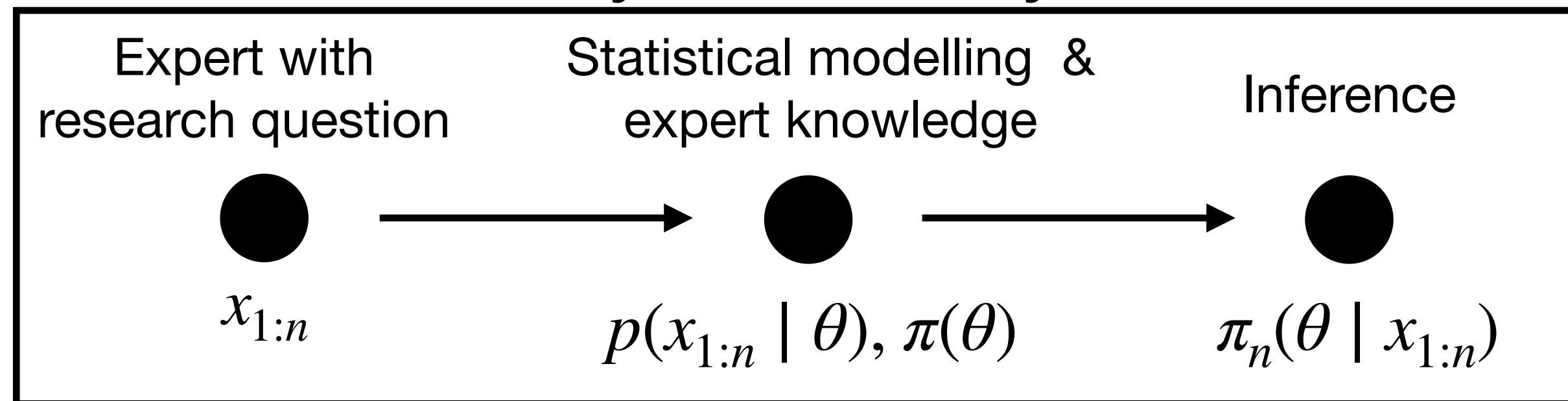
(A3) ✓

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

Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

↑ ↑ ↑ ↑

willingness to pay pollutants rooms, sqm, ... measurement error

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

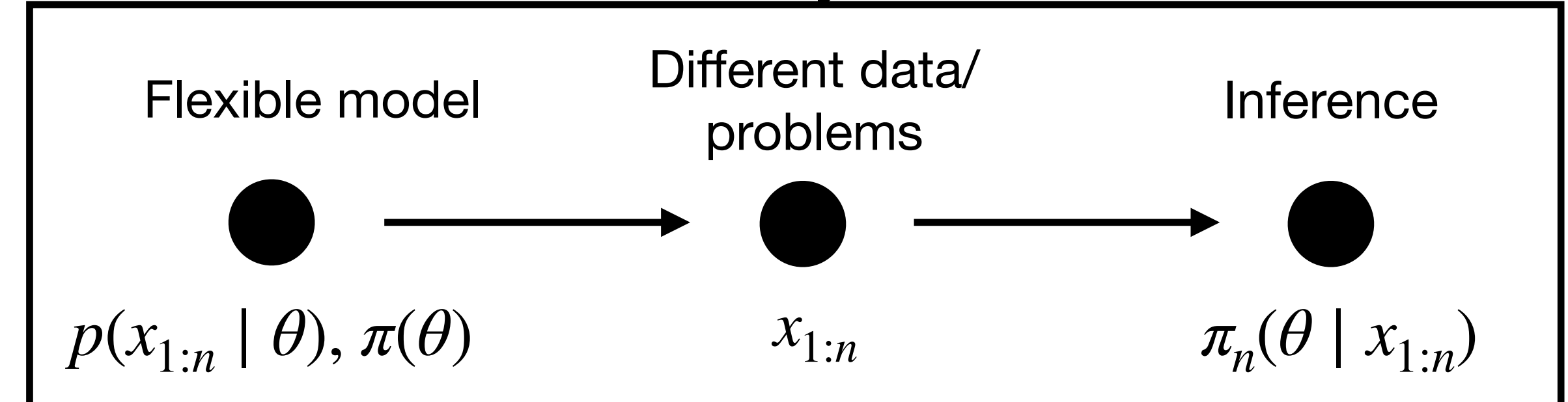
$\pi(\theta) \sim$ hand-crafted by experts

$\pi_n(\theta | x_{1:n}) \longrightarrow$ computed exactly

(A2) ✓

(A3) ✓

Modern Bayesian ML

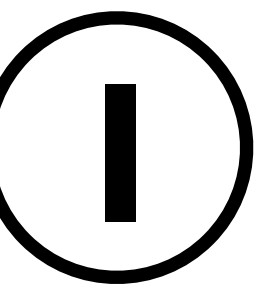


Pearce et al. (2020) [AISTATS]

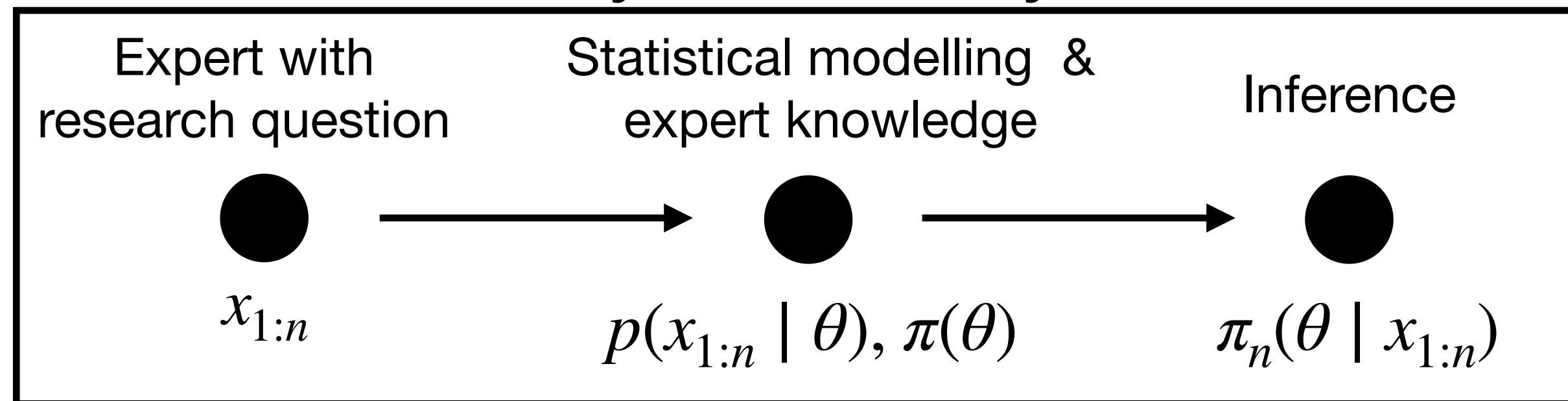
Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?

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

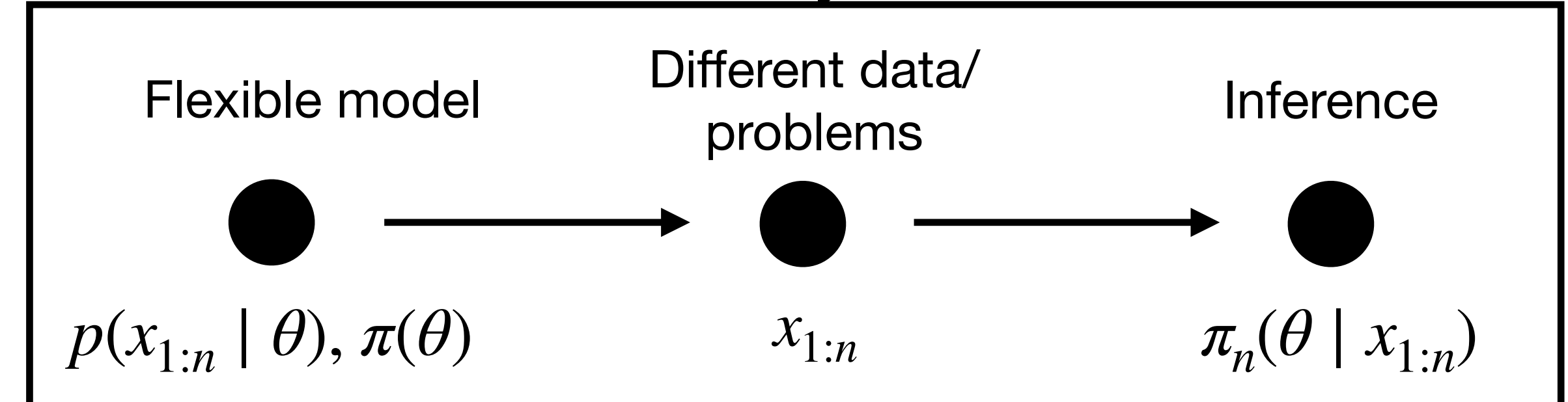
Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Modern Bayesian ML



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

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

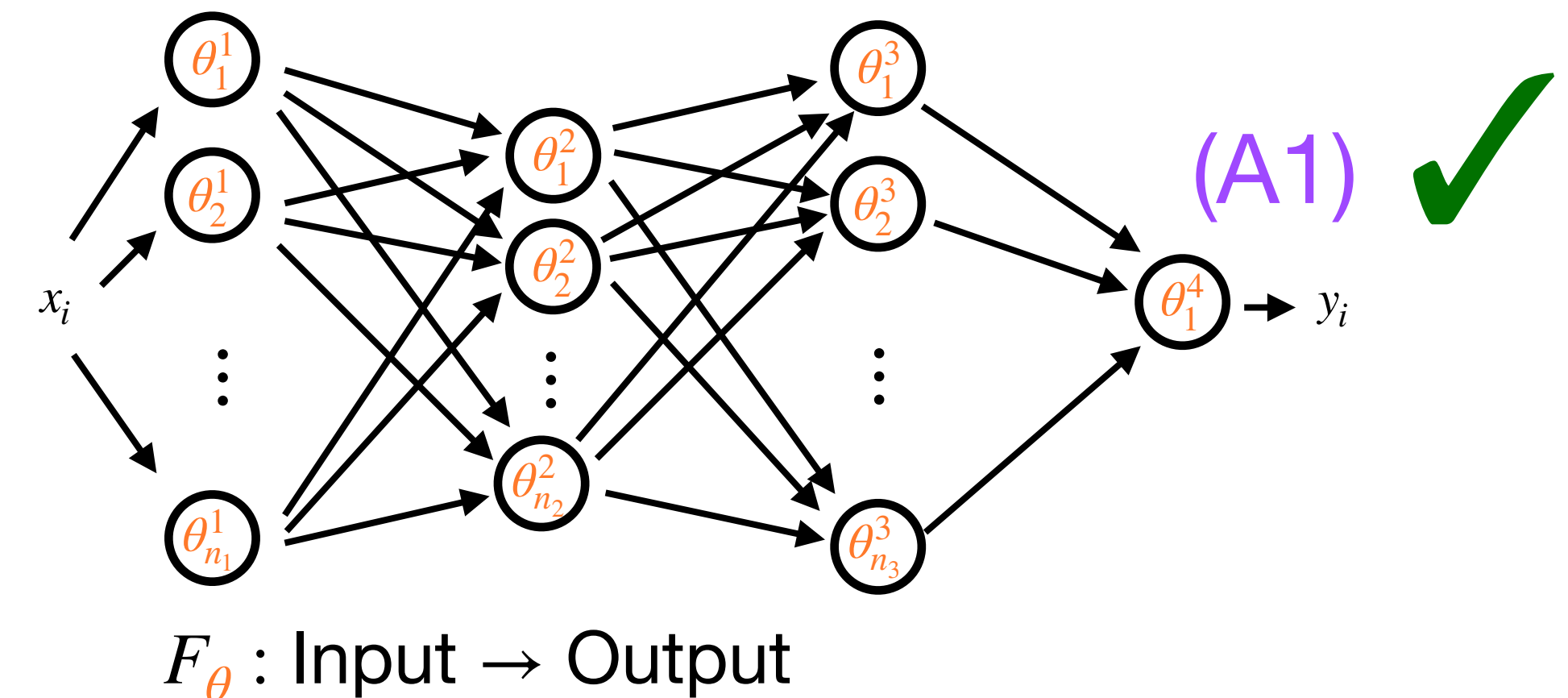
$\uparrow$ willingness to pay $\uparrow$ pollutants $\uparrow$ rooms, sqm, ... $\uparrow$ measurement error

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

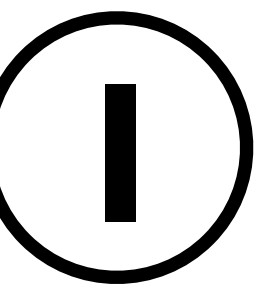
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓
 $\pi_n(\theta | x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Pearce et al. (2020) [AISTATS]

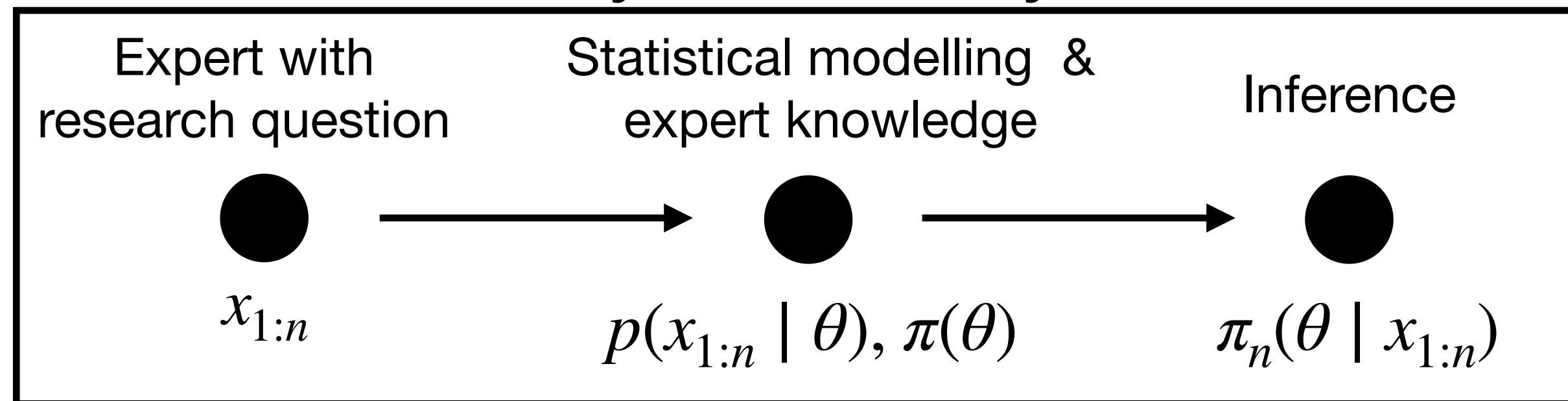
Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?



Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

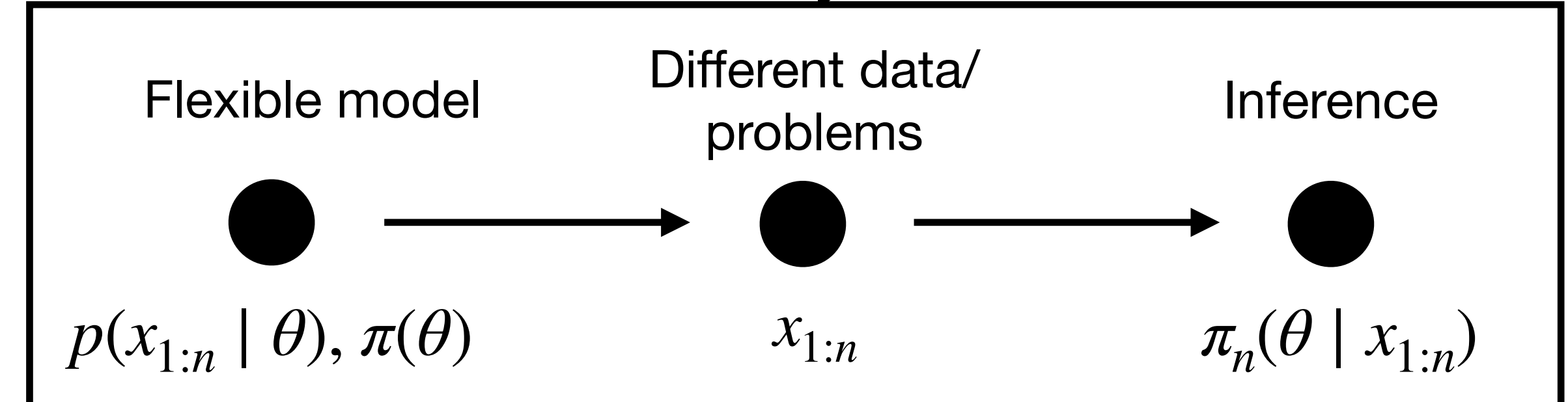
$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1+1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

$\uparrow$ willingness to pay $\uparrow$ pollutants $\uparrow$ rooms, sqm, ... $\uparrow$ measurement error

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

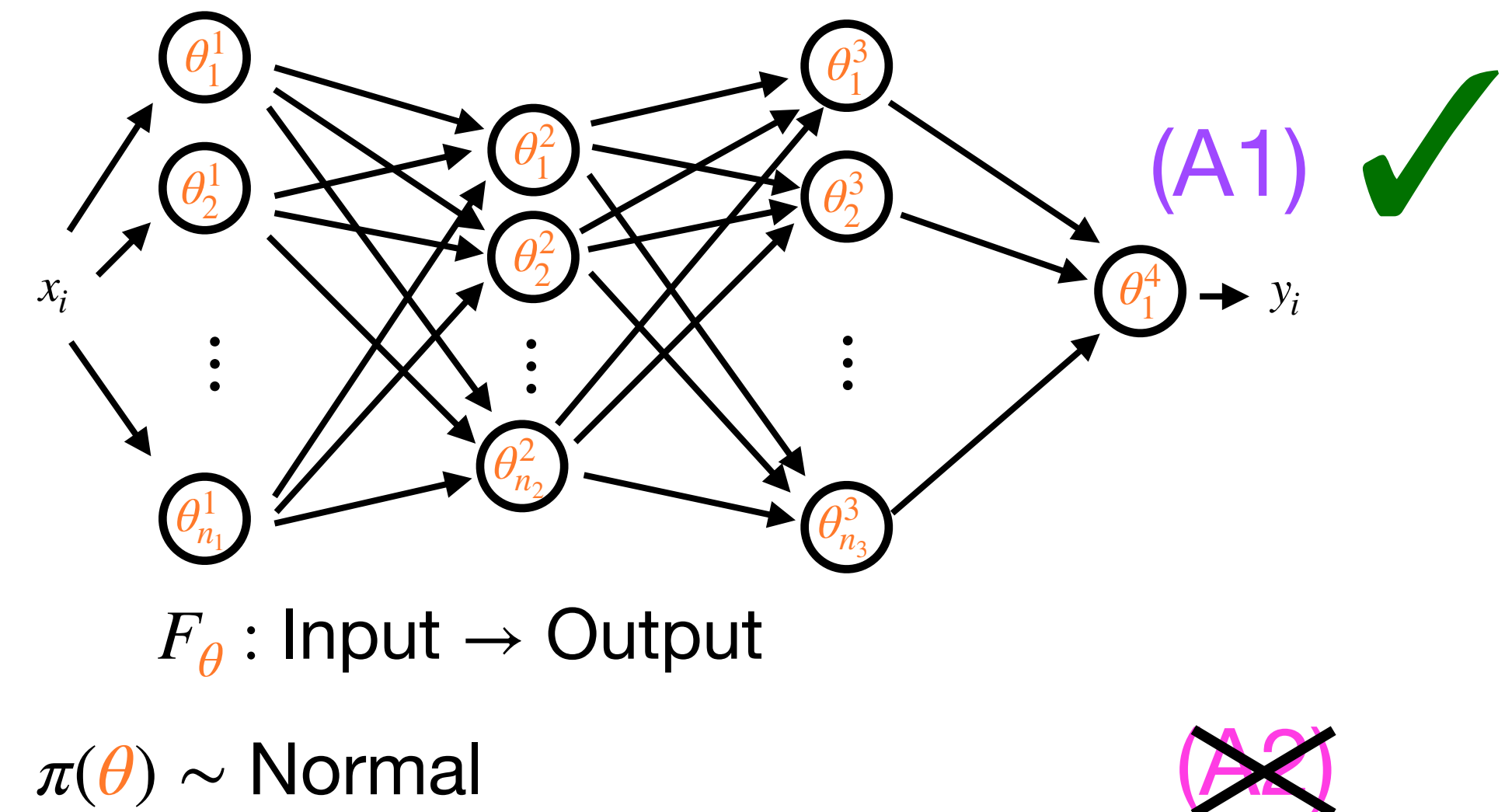
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓
 $\pi_n(\theta | x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Modern Bayesian ML

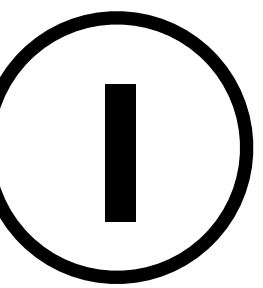


Pearce et al. (2020) [AISTATS]

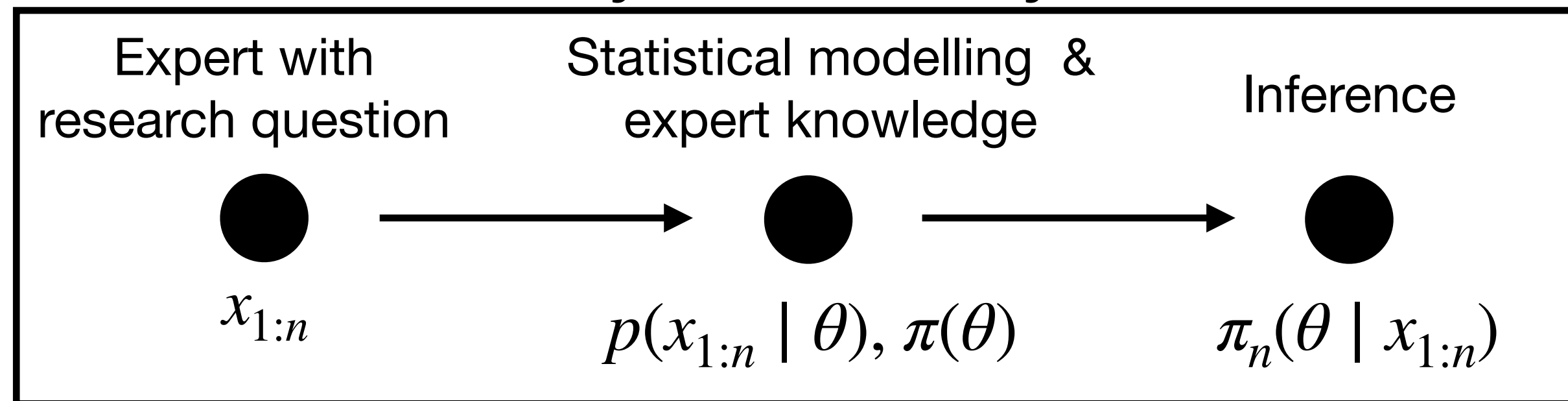
Research Question: Does my algorithm improve prediction on regression tasks like Boston UCI data?



Case Study: Bayesian ML & Boston Housing Data



Traditional Bayesian analysis in science



Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

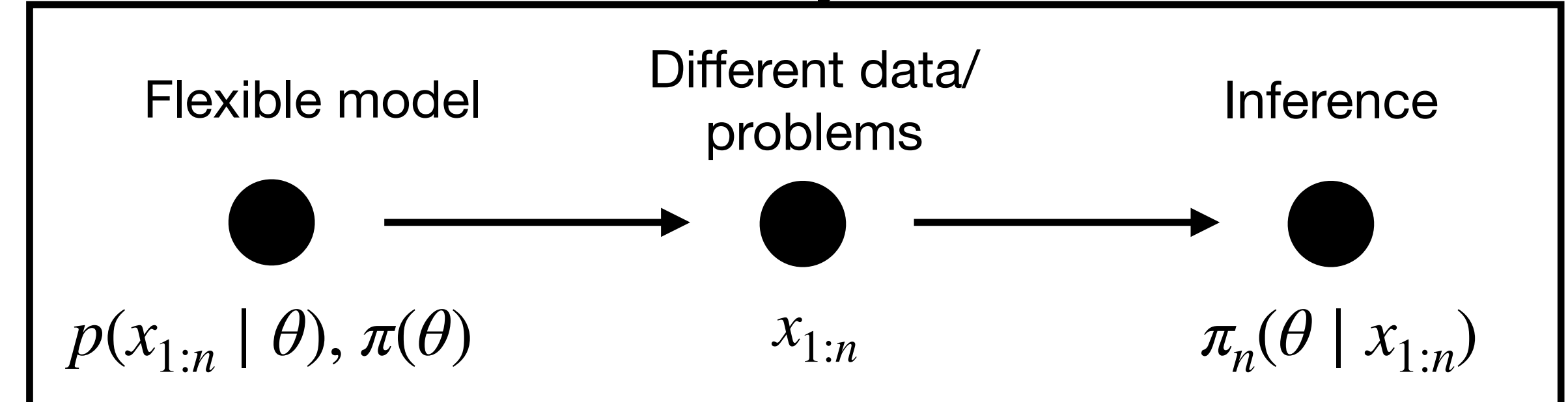
$$\log y_i = \sum_{j=1}^{J_1} p_j \log(x_{j,i}) + c_0 + \sum_{j=J_1+1}^{J_2} c_j \log(x_{j,i}) + \varepsilon_i$$

$\uparrow$ willingness to pay $\uparrow$ pollutants $\uparrow$ rooms, sqm, ... $\uparrow$ measurement error

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

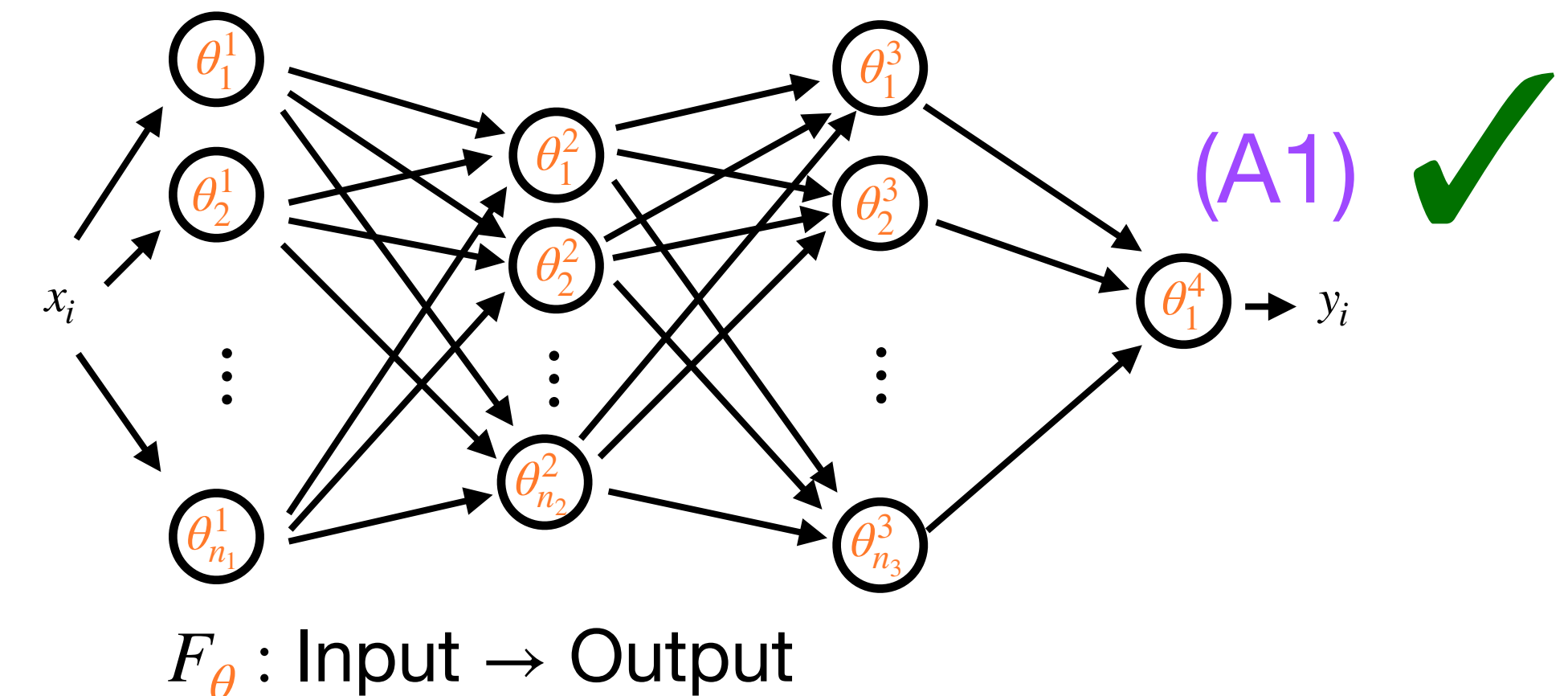
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓
 $\pi_n(\theta | x_{1:n}) \longrightarrow$ computed exactly (A3) ✓

Modern Bayesian ML



Pearce et al. (2020) [AISTATS]

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

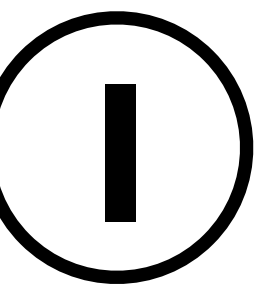


$\pi(\theta) \sim$ Normal

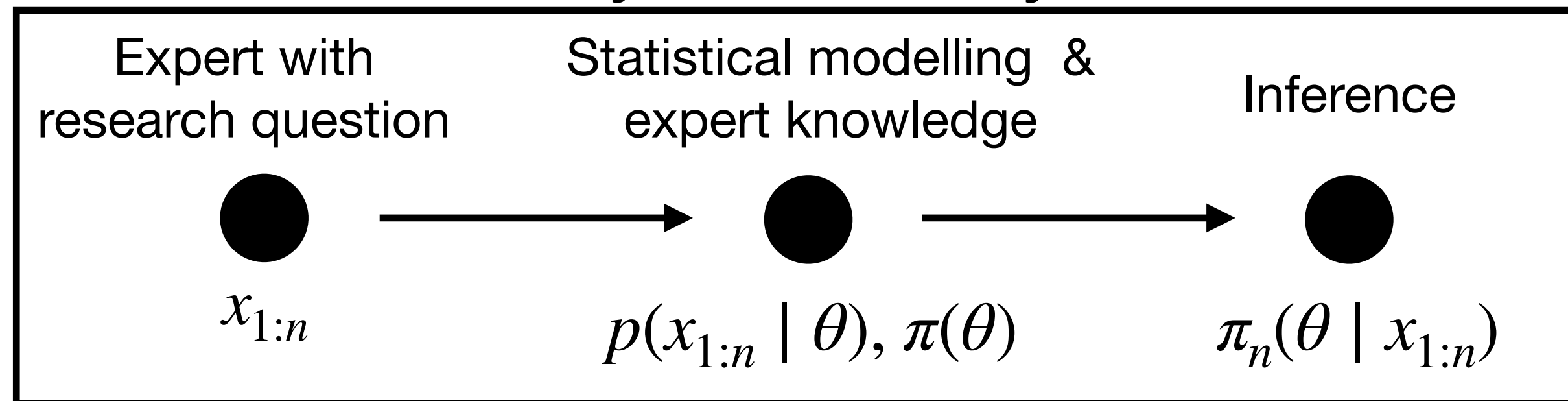
$\pi_n(\theta | x_{1:n}) \longrightarrow$ coarse approximation

(A2) ✗
(A3) ✗

Assumptions & Foundations

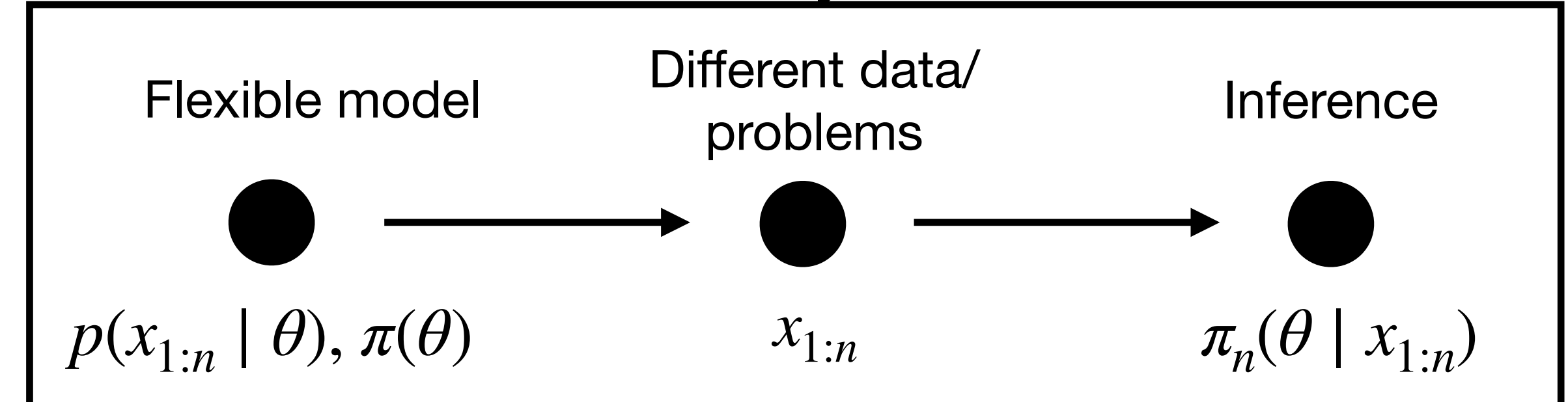


Traditional Bayesian analysis in science



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

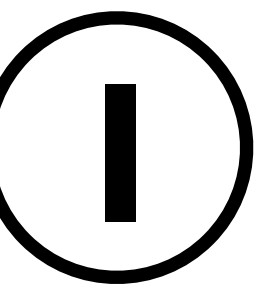
Modern Bayesian ML



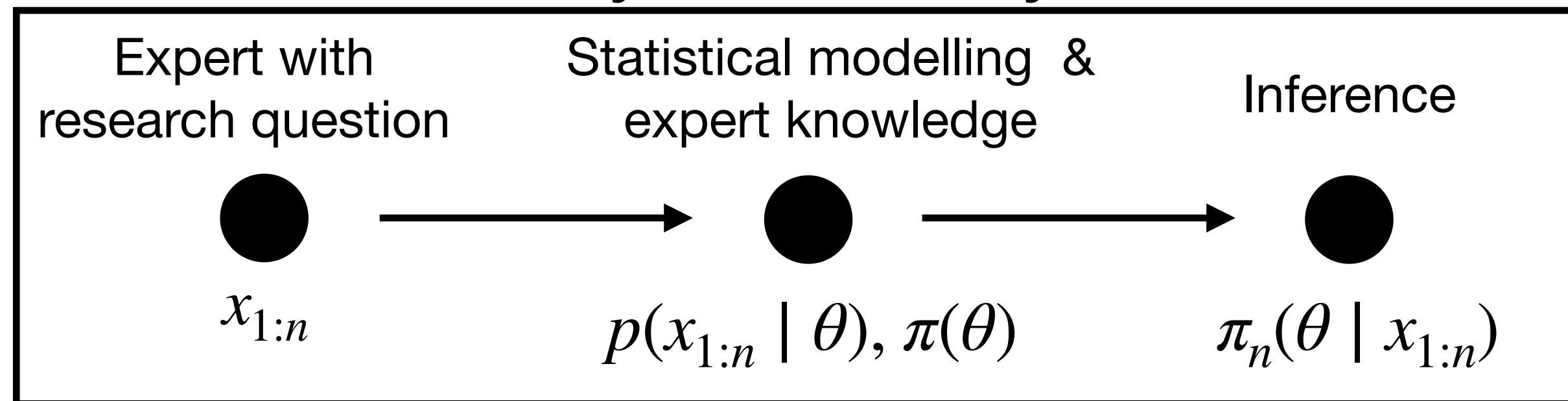
- (A1) model well-specified
- (A2) prior well-specified
- (A3) computationally feasible



Assumptions & Foundations

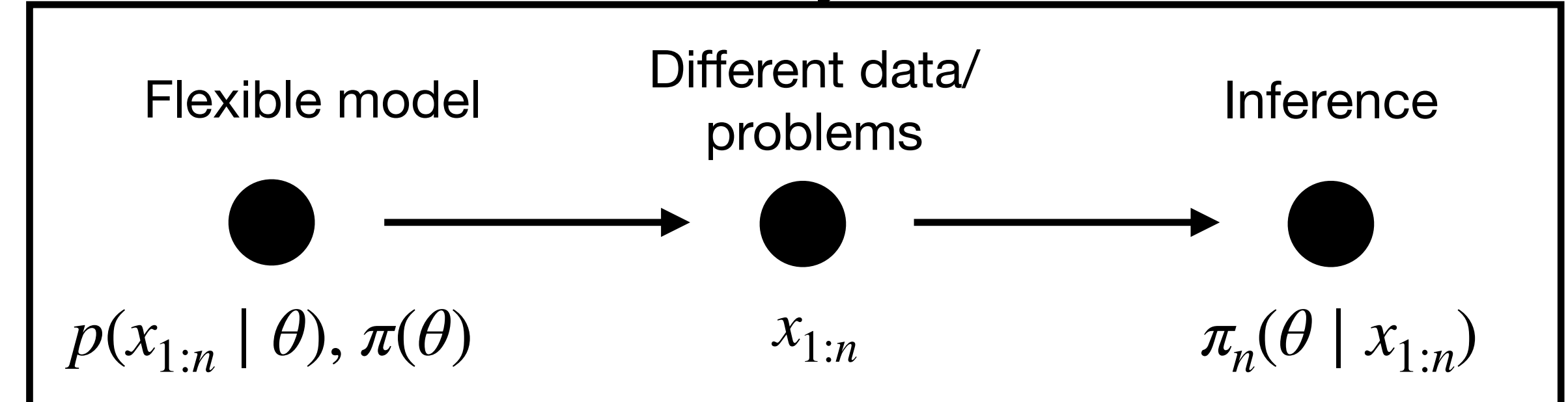


Traditional Bayesian analysis in science



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

Modern Bayesian ML



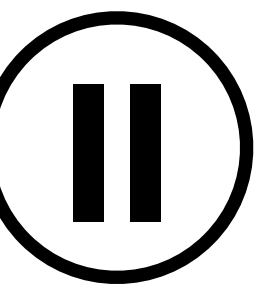
- (A1) model well-specified
- (A2) prior well-specified
- (A3) computationally feasible



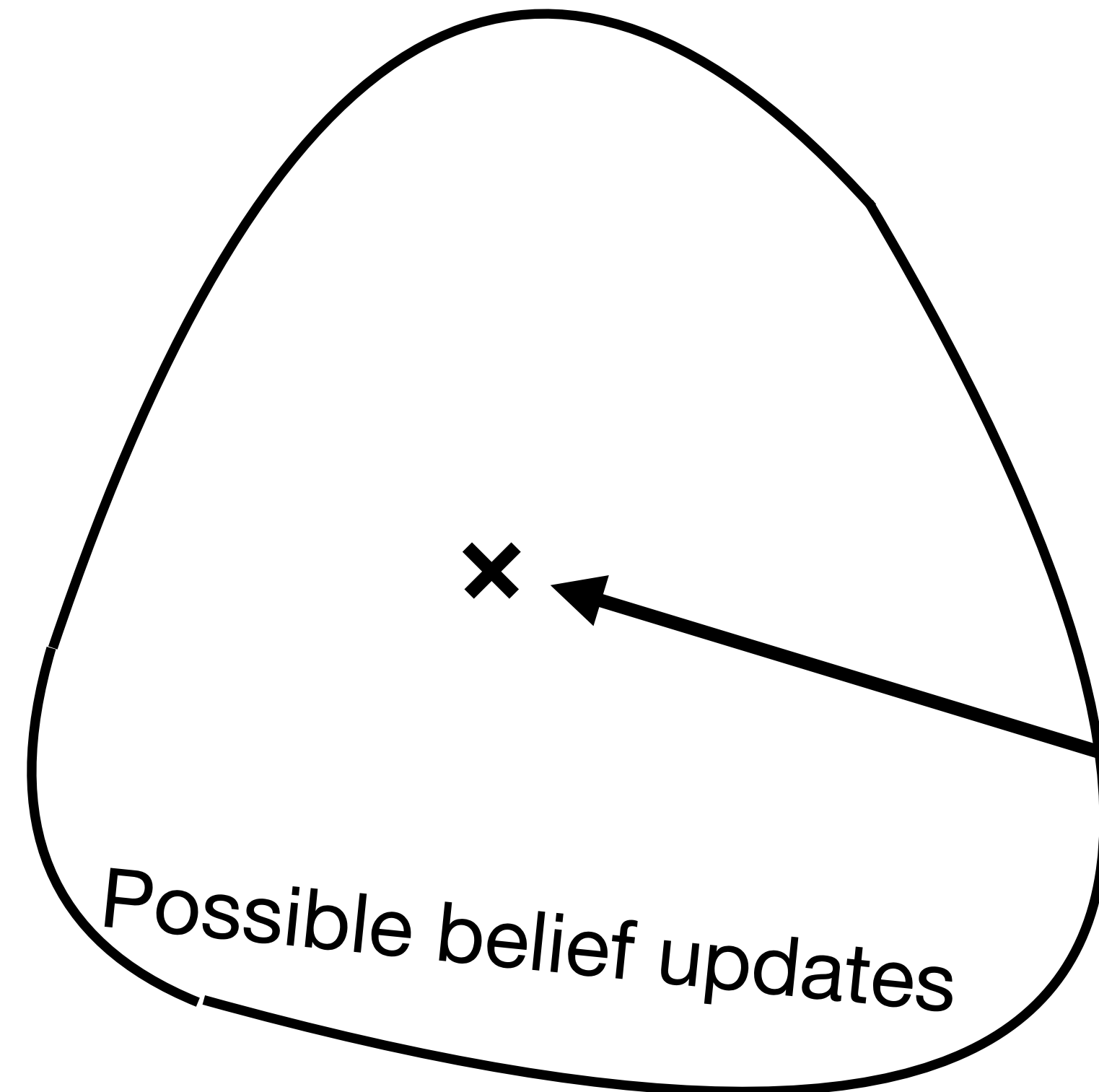
Post-Bayesian Approaches ask:

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

Post-Bayesian Inference



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



Possible belief updates

Bayes' Posterior

(A1), (A2), (A3)

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

Post-Bayesian Inference

space of priors

space of hyperparameters

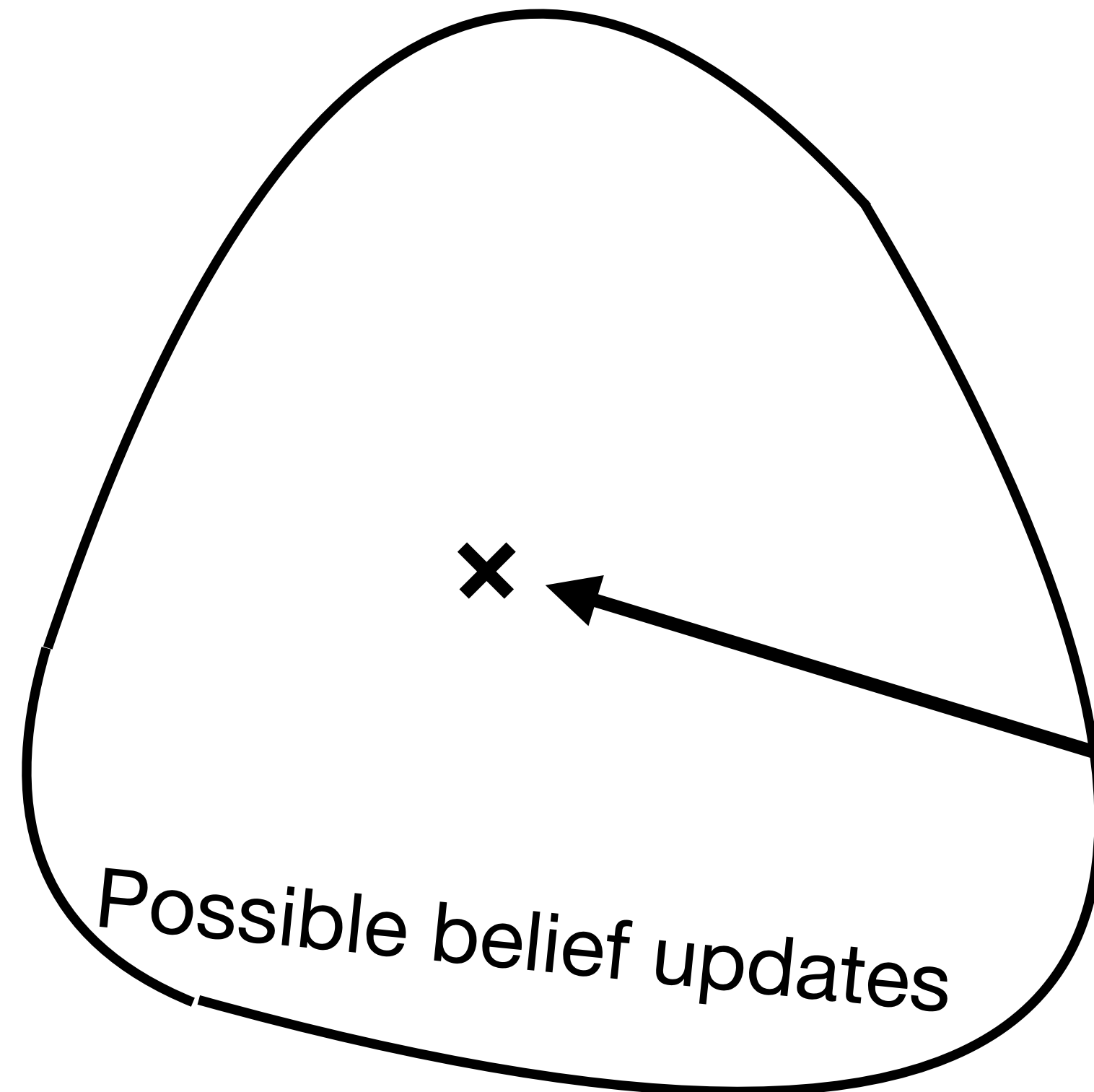
Belief updates = $\{ \mathcal{B} : \mathcal{P}(\Theta) \times \mathcal{X}^n \times \mathcal{H} \longrightarrow \mathcal{P}(\Theta) \}$

data space

space of posteriors

II

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



Bayes' Posterior

(A1), (A2), (A3)

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

Post-Bayesian Inference

space of priors

space of hyperparameters

Belief updates = $\{ \mathcal{B} : \mathcal{P}(\Theta) \times \mathcal{X}^n \times \mathcal{H} \longrightarrow \mathcal{P}(\Theta) \}$

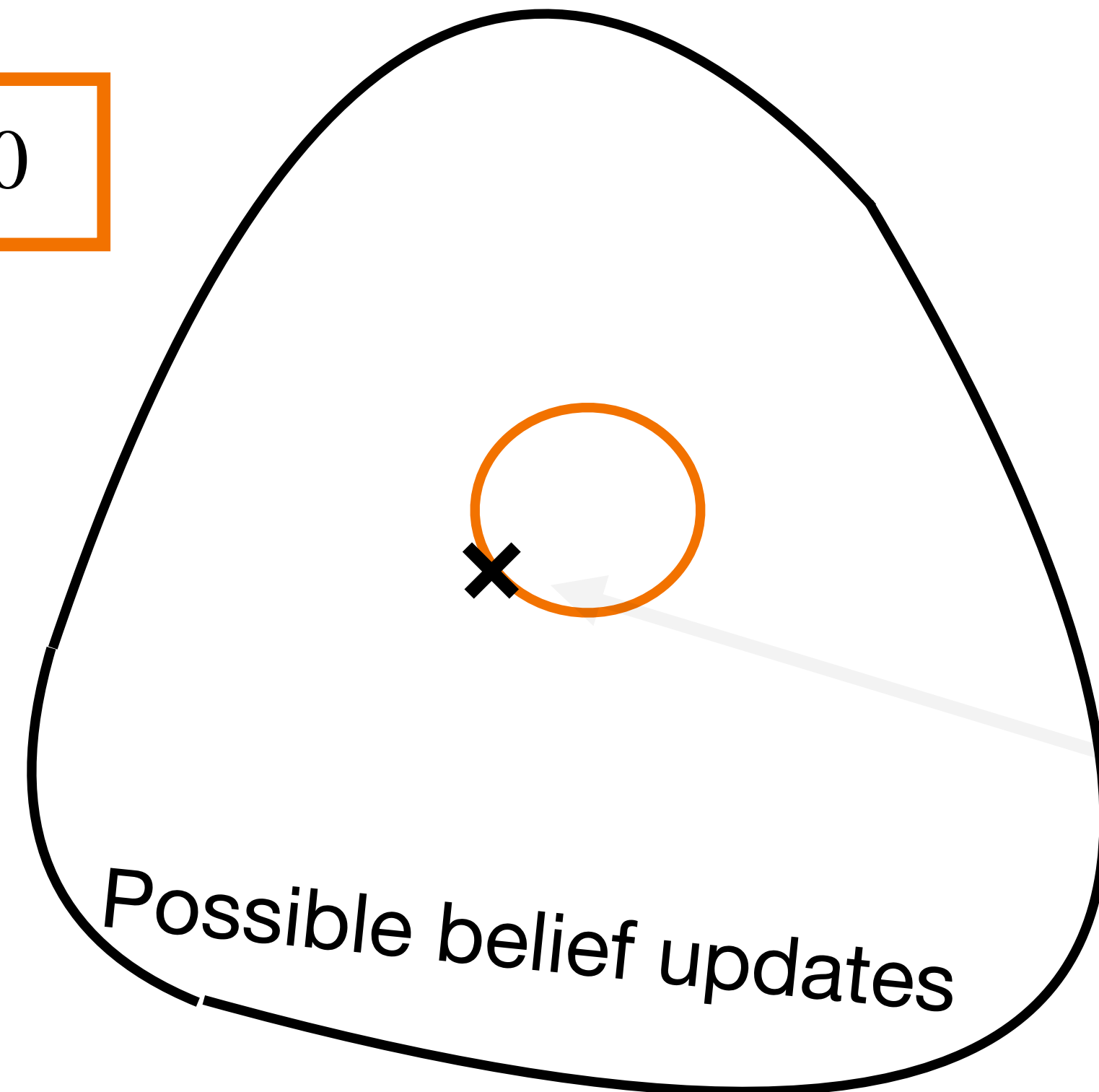
data space

space of posteriors

II

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

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



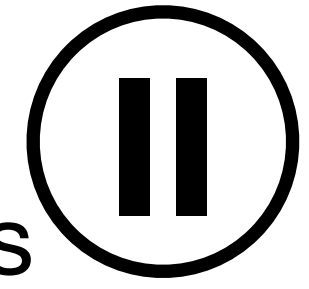
(A1), (A2), (A3)

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

Post-Bayesian Inference

space of priors

space of hyperparameters



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

space of posteriors

data space

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

[See Grünwald (2011)]

Power/Fractional/
Cold Posterior

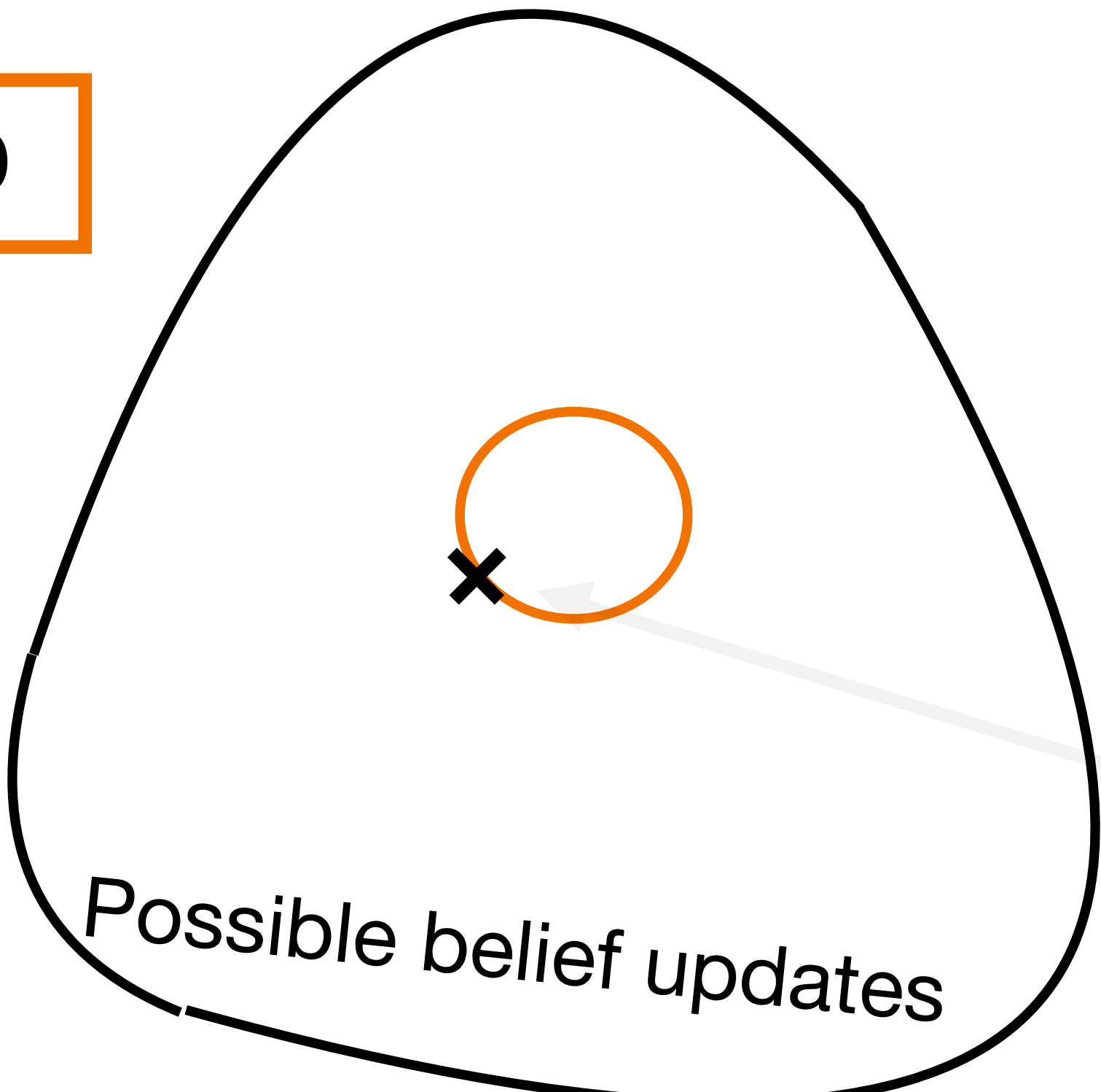
~~(A1)~~, (A2), (A3)

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

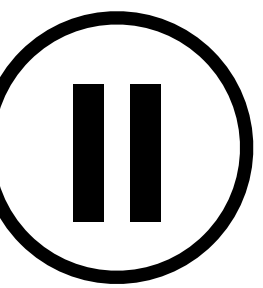
(A1), (A2), (A3)

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

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



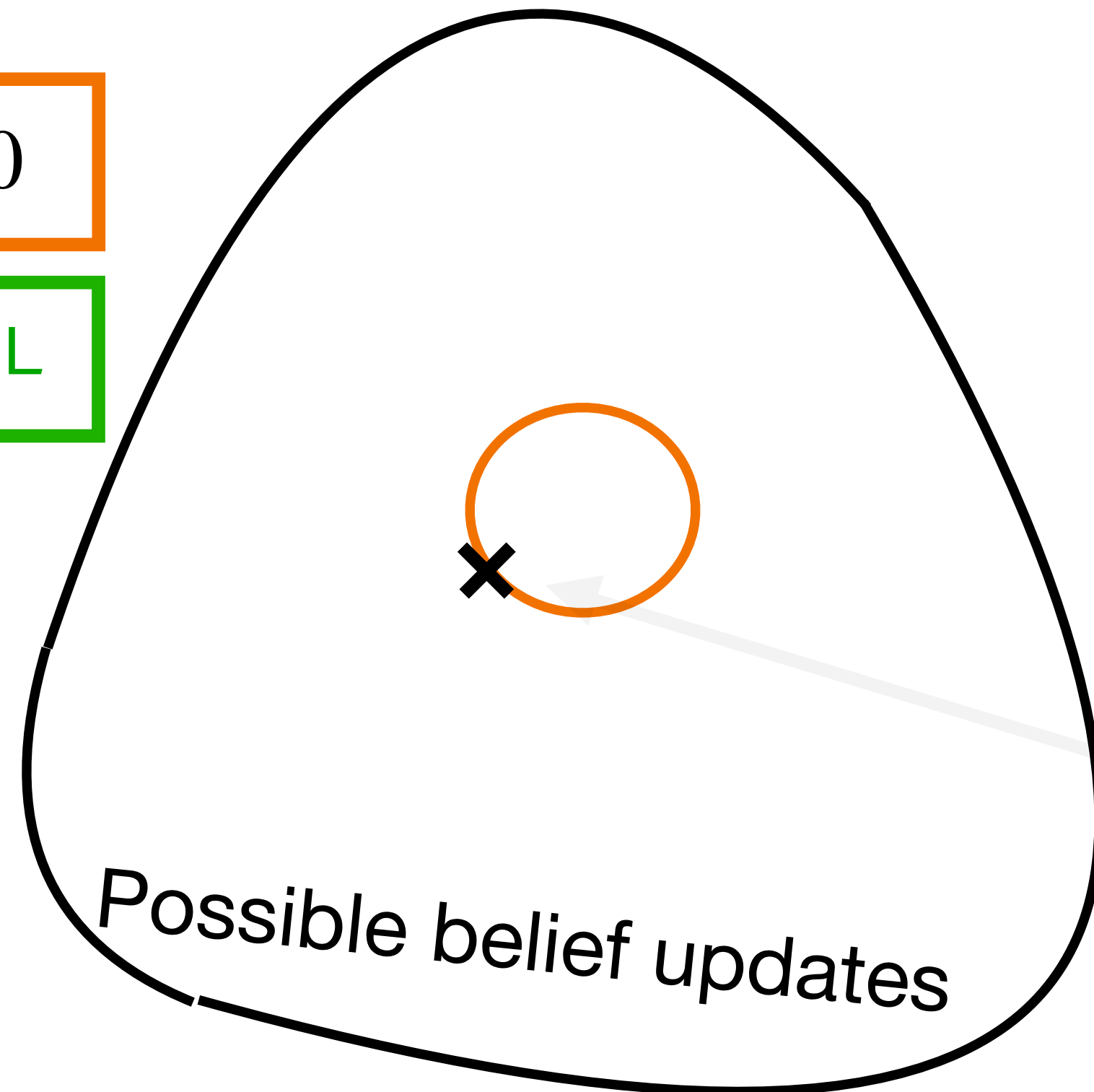
Post-Bayesian Inference



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

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

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



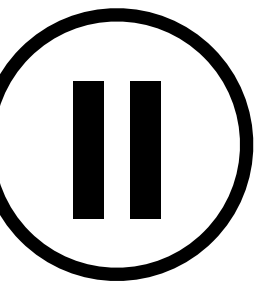
~~(A1)~~, (A2), (A3)

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

(A1), (A2), (A3)

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

Post-Bayesian Inference



[See Bissiri, Holmes & Walker (2016)]

Gibbs/Generalised/
Pseudo Posterior

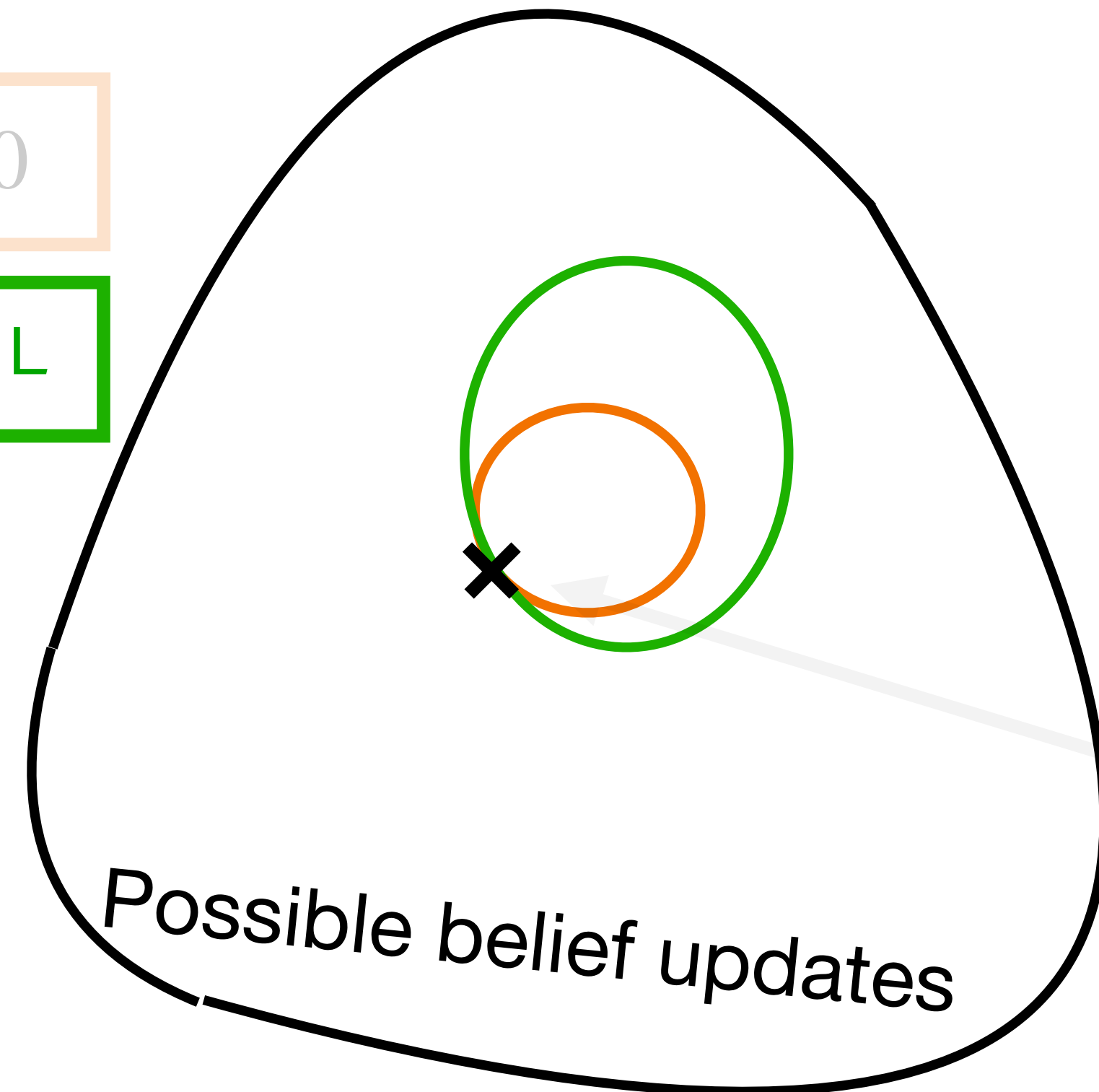
~~(A1)~~, (A2), (A3)

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

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

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

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



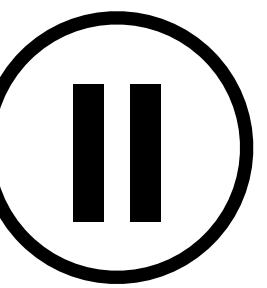
~~(A1)~~, (A2), (A3)

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

(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



Optimisation-centric posteriors /
Generalised Variational Inference

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

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

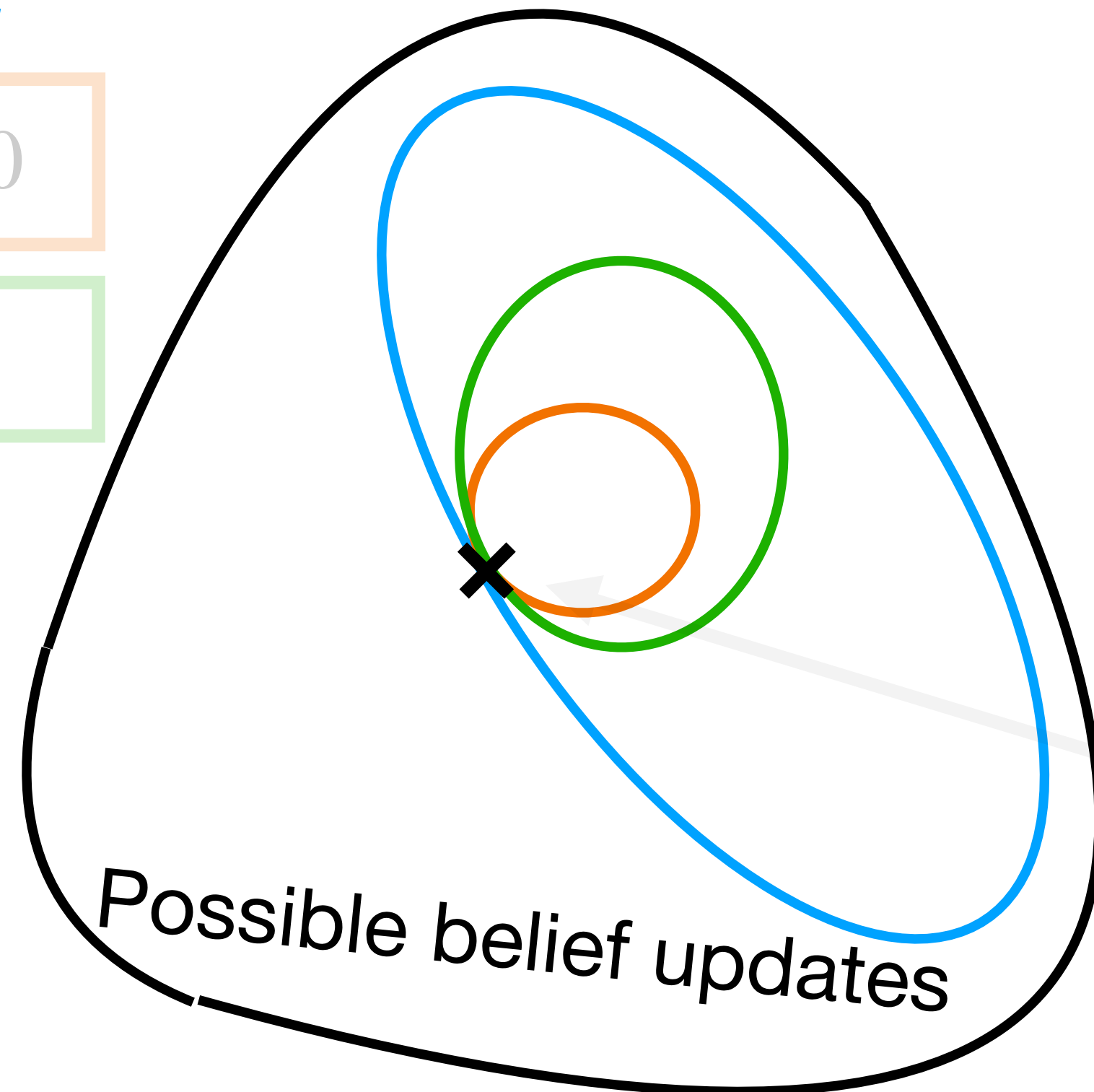
$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$

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

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

$$\begin{array}{ccc} \text{KL} & \longrightarrow & D \\ \mathcal{P}(\Theta) & \longrightarrow & \mathcal{Q} \end{array}$$

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



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

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

(A1), (A2), (A3)

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



Post-Bayesian Inference

Optimisation-centric posteriors /
Generalised Variational Inference

(A1), (A2), (A3)

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

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

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

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

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

$$\begin{array}{cc} \text{KL} & \longrightarrow & D \\ \mathcal{P}(\Theta) & \longrightarrow & \mathcal{Q} \end{array}$$

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

Gibbs/Generalised/
Pseudo Posterior

(A1), (A2), (A3)

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

Power/Fractional/
Cold Posterior

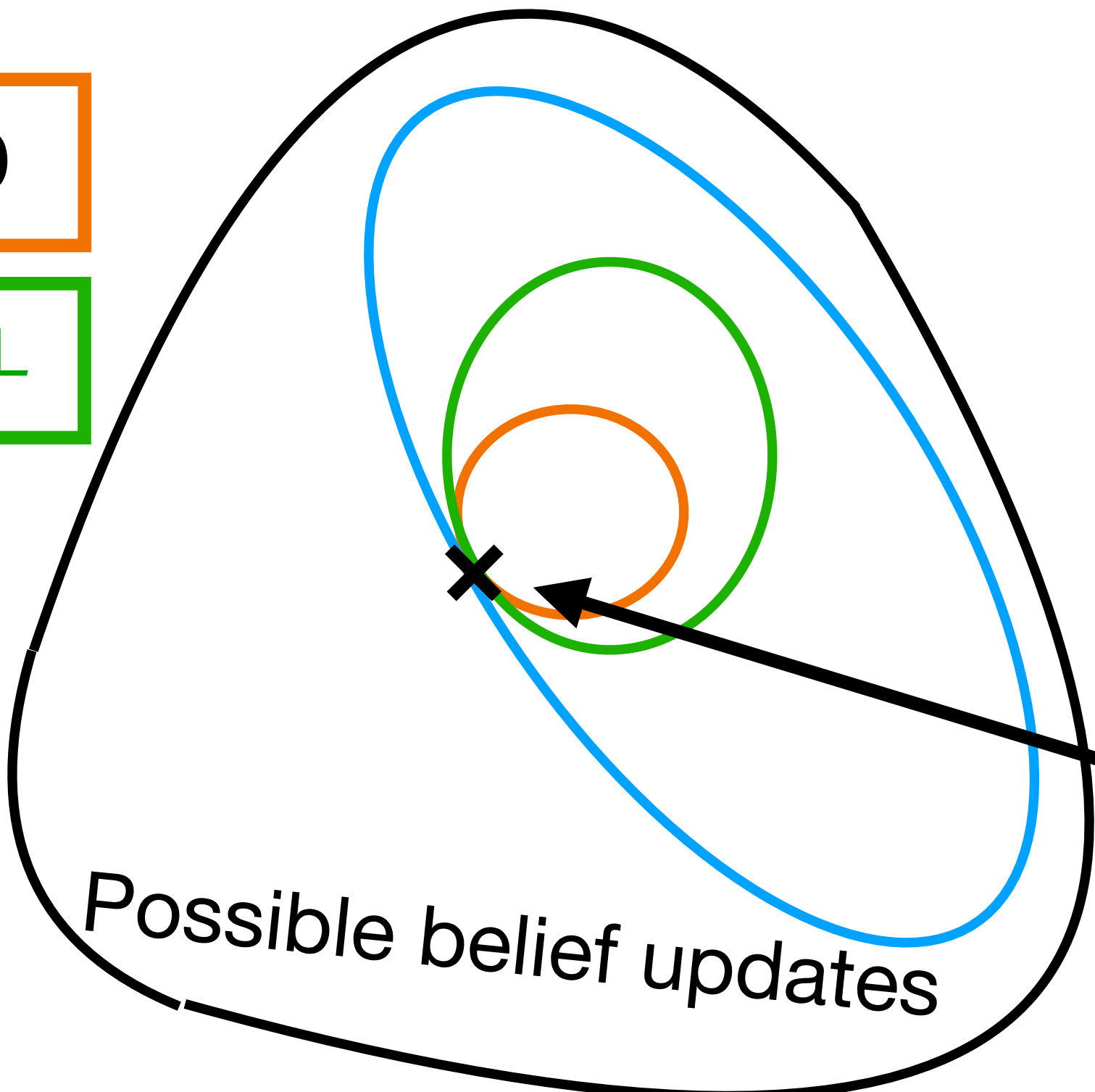
(A1), (A2), (A3)

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

Bayes' Posterior

(A1), (A2), (A3)

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



Post-Bayesian Inference

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

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

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

Predictively oriented Inference

VI

Gibbs/Generalised/
Pseudo Posterior

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

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

Power/Fractional/
Cold Posterior

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

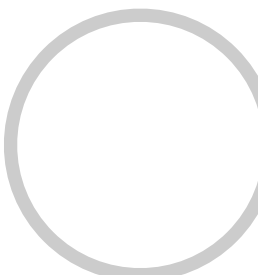
$$\pi_n^{(\lambda)}(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta) d\theta}$$

Parameter Inference

III

Post-Bayesian Inference

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



Gibbs/Generalised/
Pseudo Posterior

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

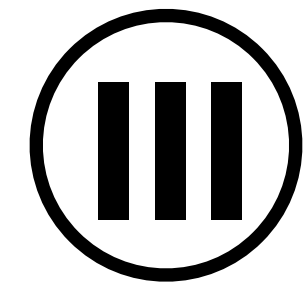
$$\pi_n^{\mathbf{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Power/Fractional/
Cold Posterior

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

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

Parameter Inference

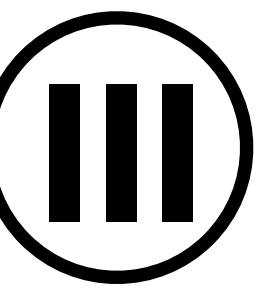


Power Posteriors

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?



Power Posteriors

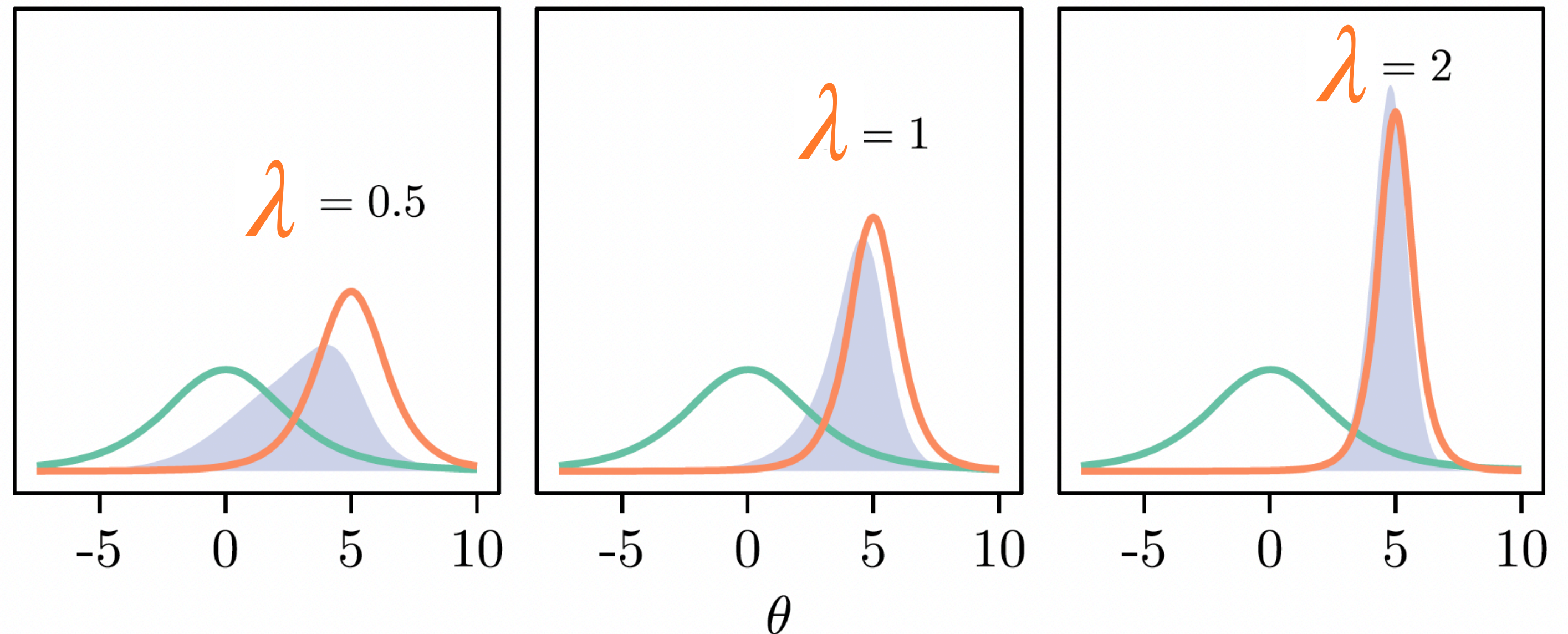
Power/Fractional/Cold Posteriors

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

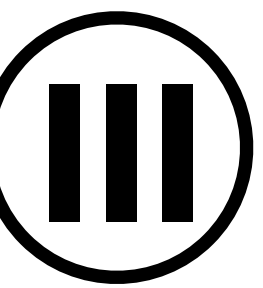
■ posterior $\pi_n^{(\lambda)}(\theta \mid x_{1:n})$
— prior $\pi(\theta)$
— likelihood $p(x_{1:n} \mid \theta)^\lambda$

Q: What does it do?

A: Trades off prior vs data



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



Power Posteriors

Power/Fractional/Cold Posteriors

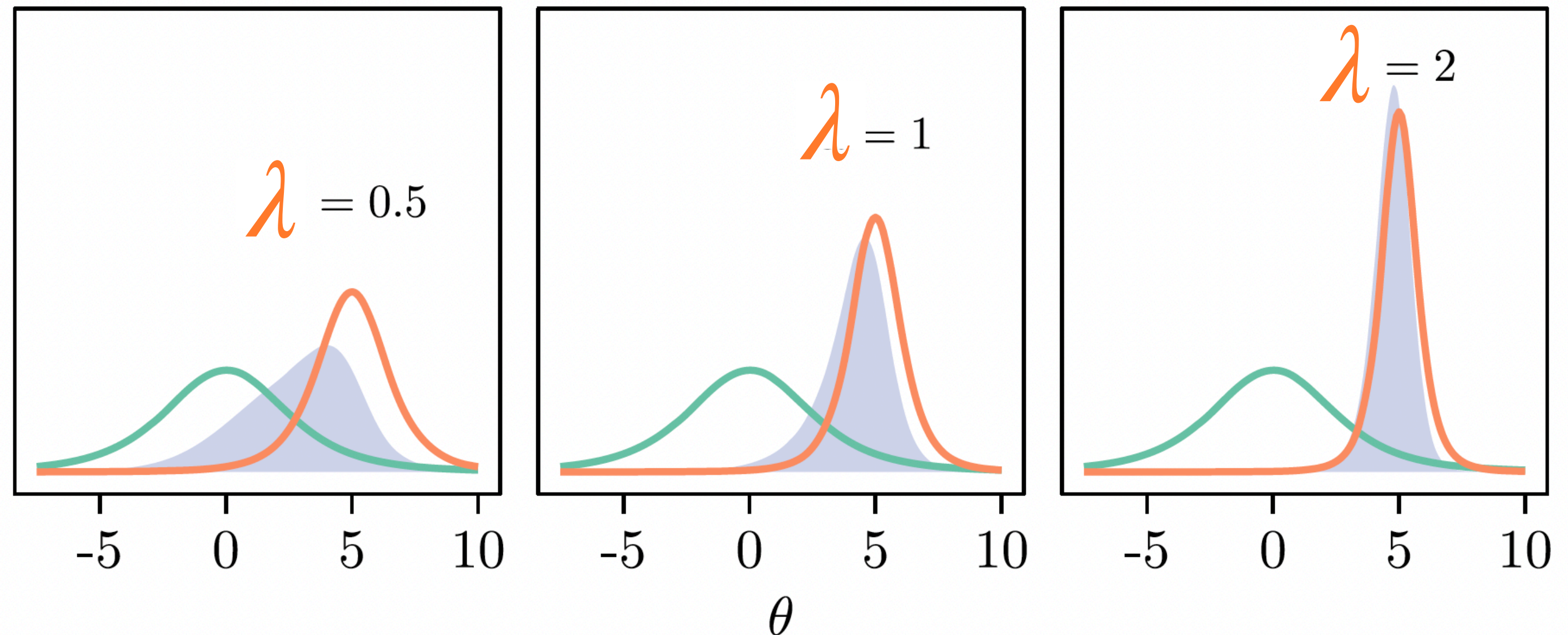
$$\pi_n^{(\lambda)}(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta) d\theta}$$

■ posterior $\pi_n^{(\lambda)}(\theta \mid x_{1:n})$
— prior $\pi(\theta)$
— likelihood $p(x_{1:n} \mid \theta)^\lambda$

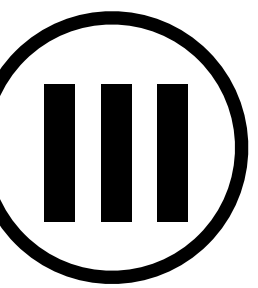
Q: What does it do?

A: Trades off prior vs data

⇒ **No robustness to model misspecification...**



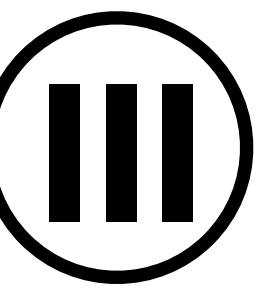
Picture from Kallionen, Paananen, Bürkner, & Vehtari (2023)



Gibbs Posteriors

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

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

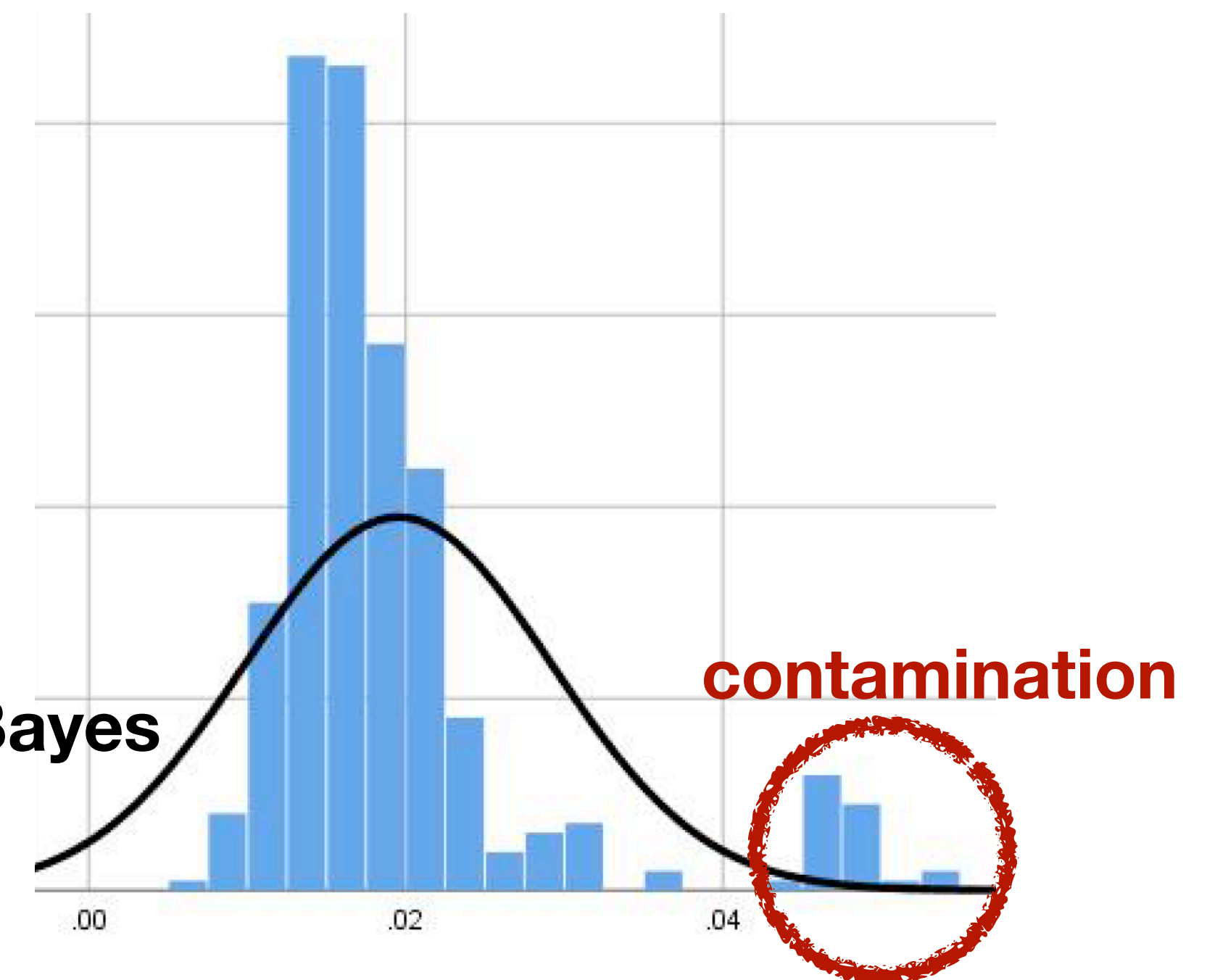


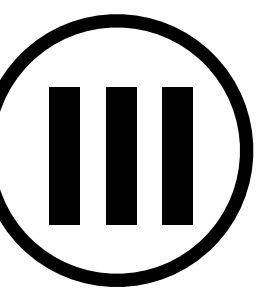
Gibbs Posteriors

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

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

Standard Bayes

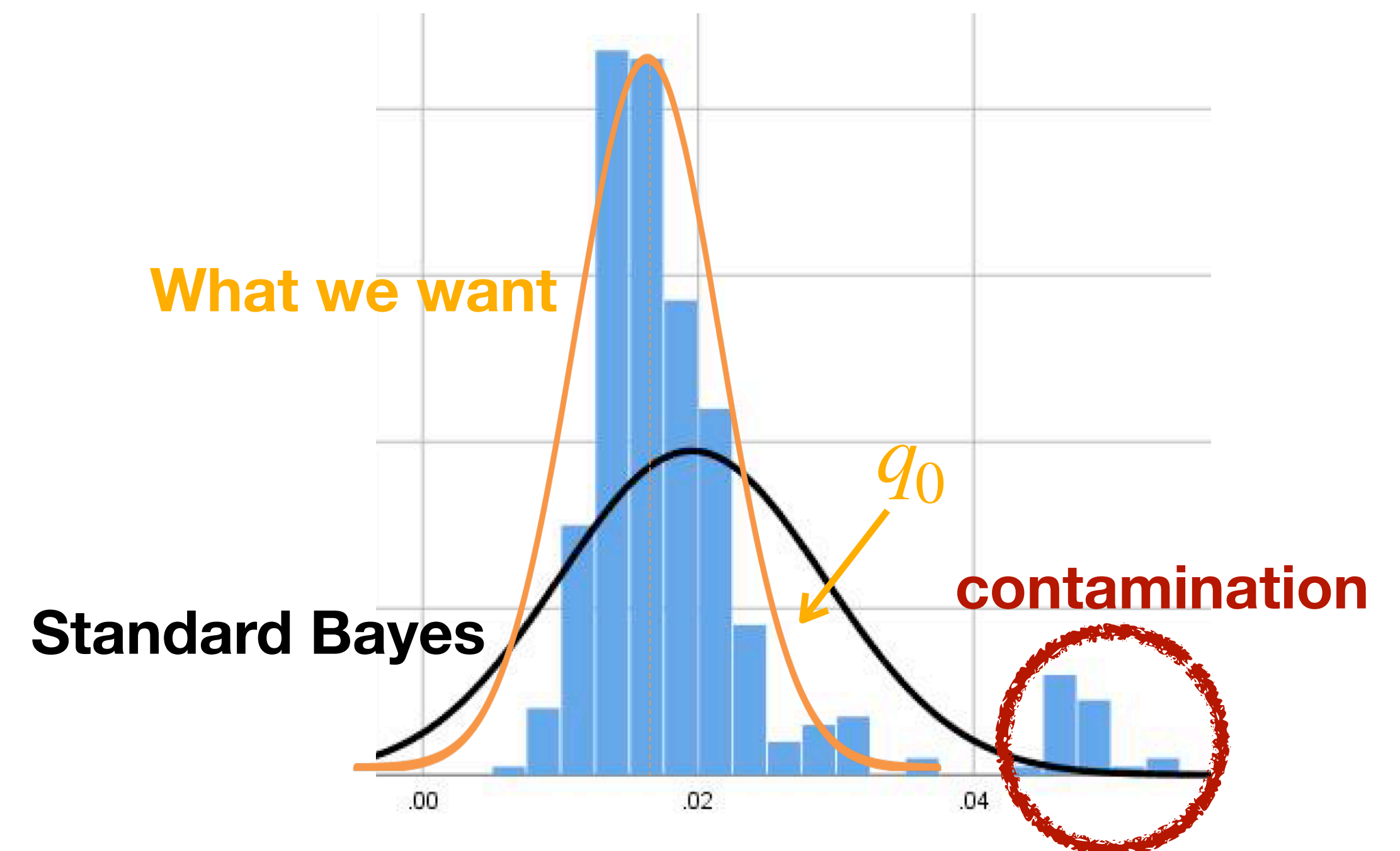


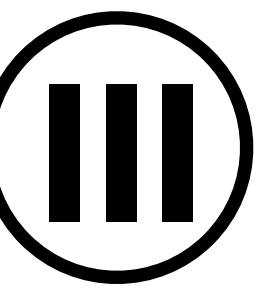


Gibbs Posteriors

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

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





Gibbs Posteriors

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

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

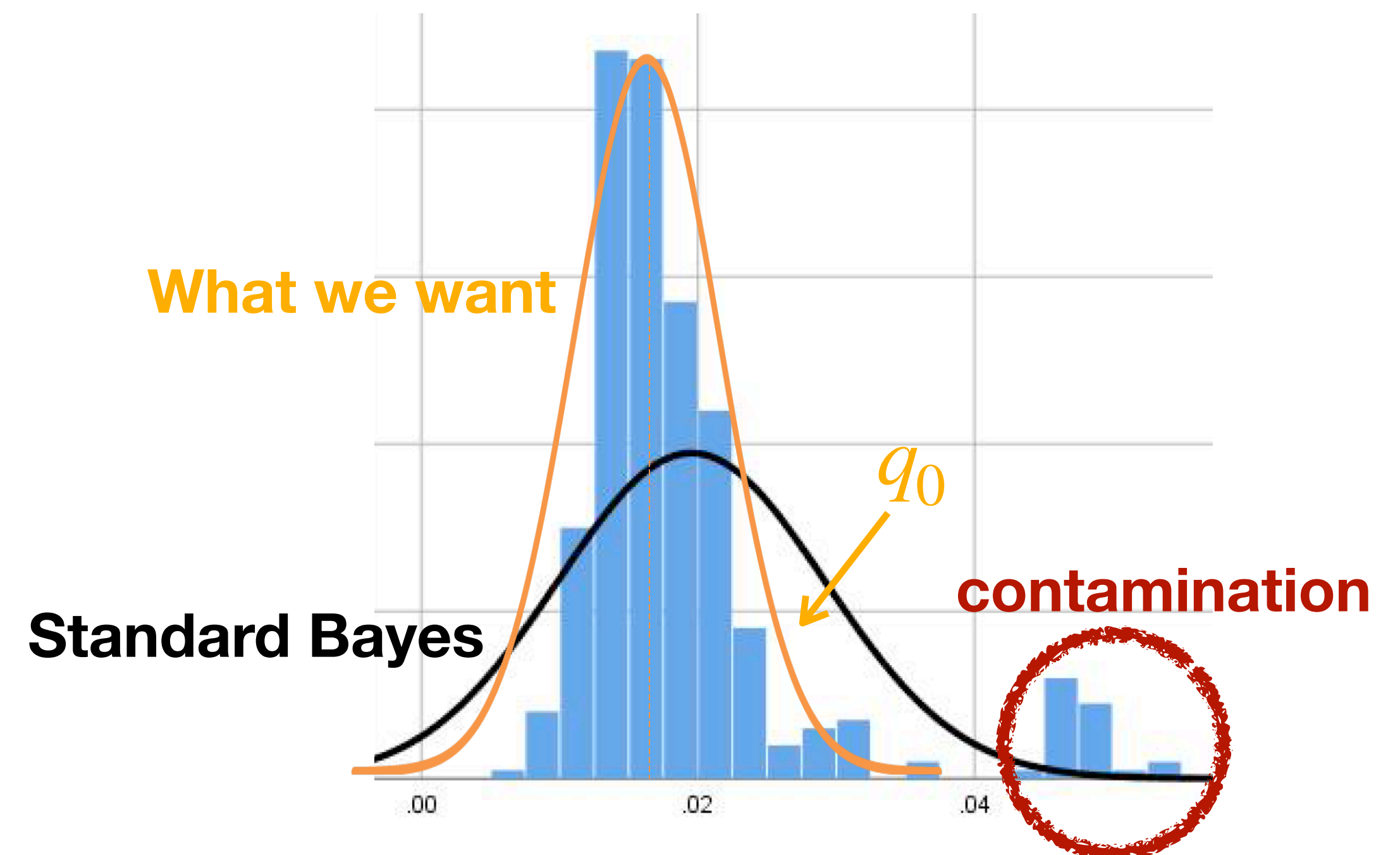
Setting: for some small $\varepsilon \geq 0$,

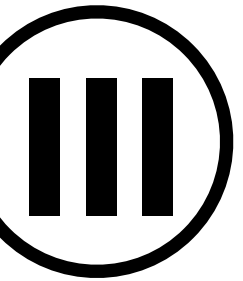
**Data-generating
probability distribution**

ε -**contamination
distribution**

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

Part of distribution our model captures





Gibbs Posteriors

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

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

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

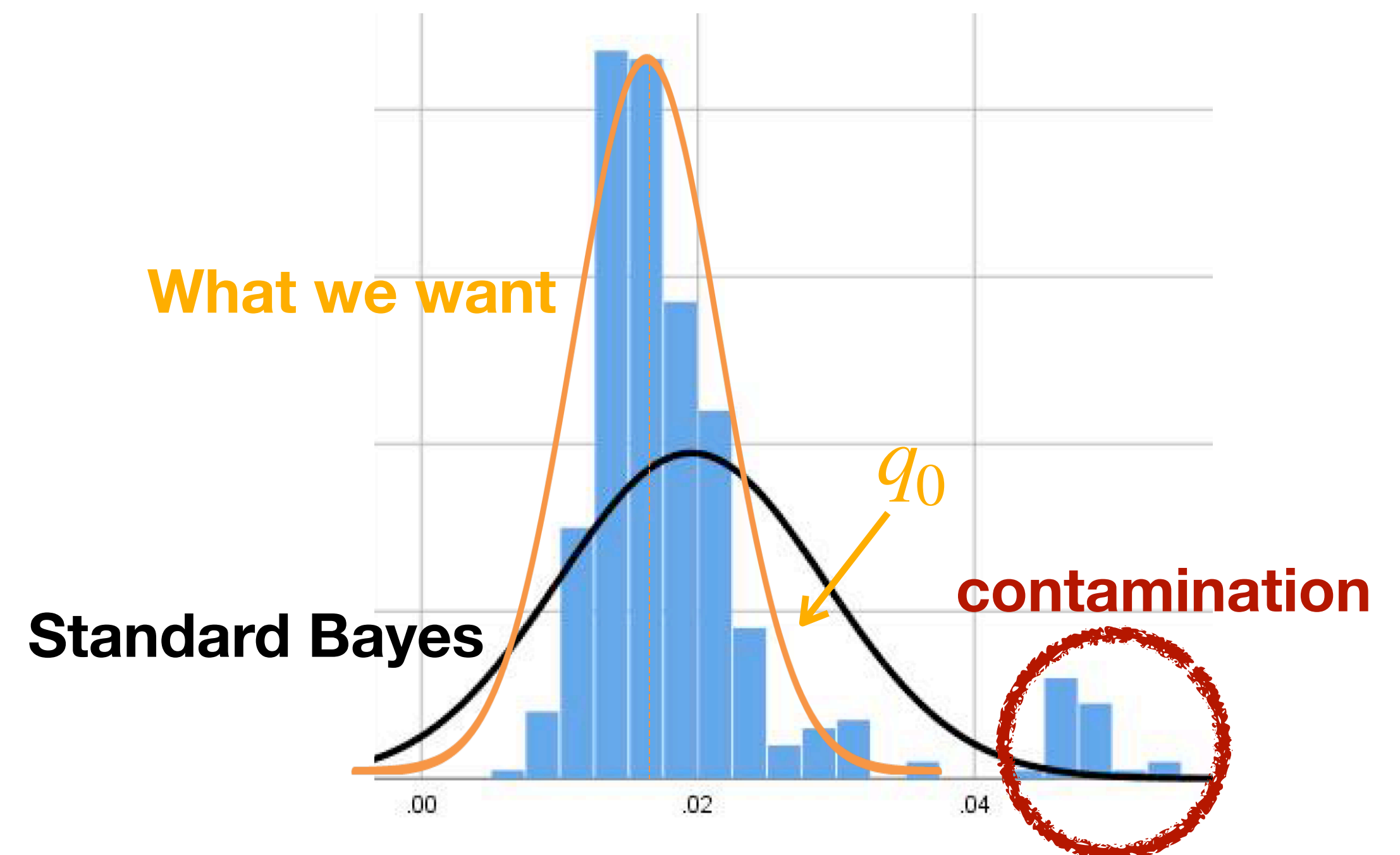
Setting: for some small $\varepsilon \geq 0$,

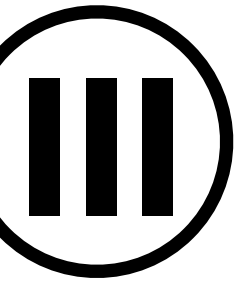
**Data-generating
probability distribution**

ε -**contamination
distribution**

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

Part of distribution our model captures





Gibbs Posteriors

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

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

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

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

**Data-generating
probability distribution**

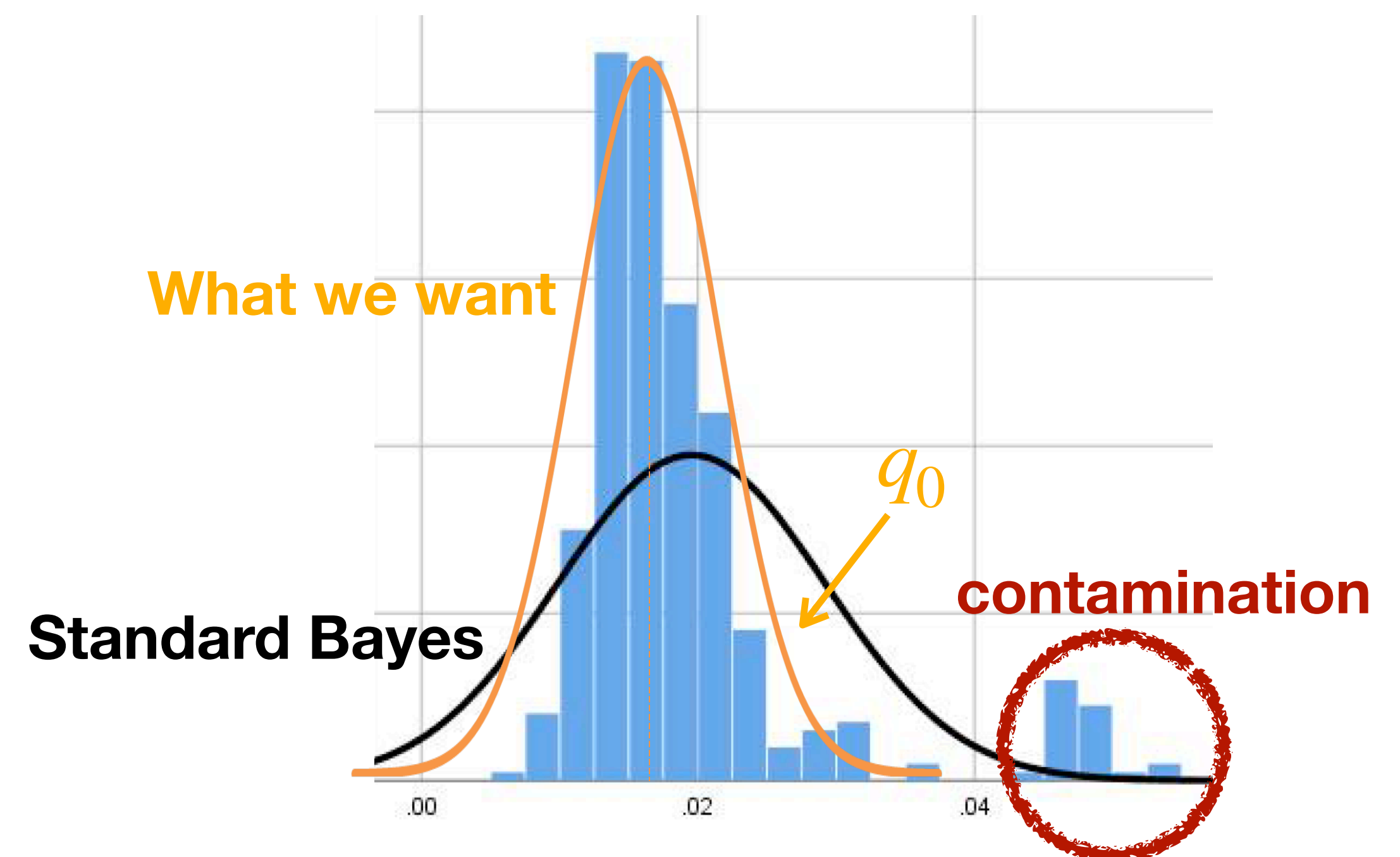
ε -**contamination
distribution**

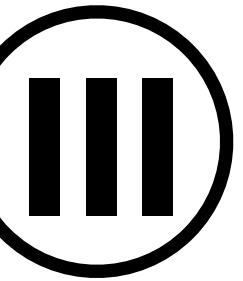
$$q_\varepsilon = (1 - \varepsilon) \cdot q_0 + \varepsilon \cdot c$$

Part of distribution our model captures

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

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





Gibbs Posteriors

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

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

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

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

**Data-generating
probability distribution**

ε -**contamination
distribution**

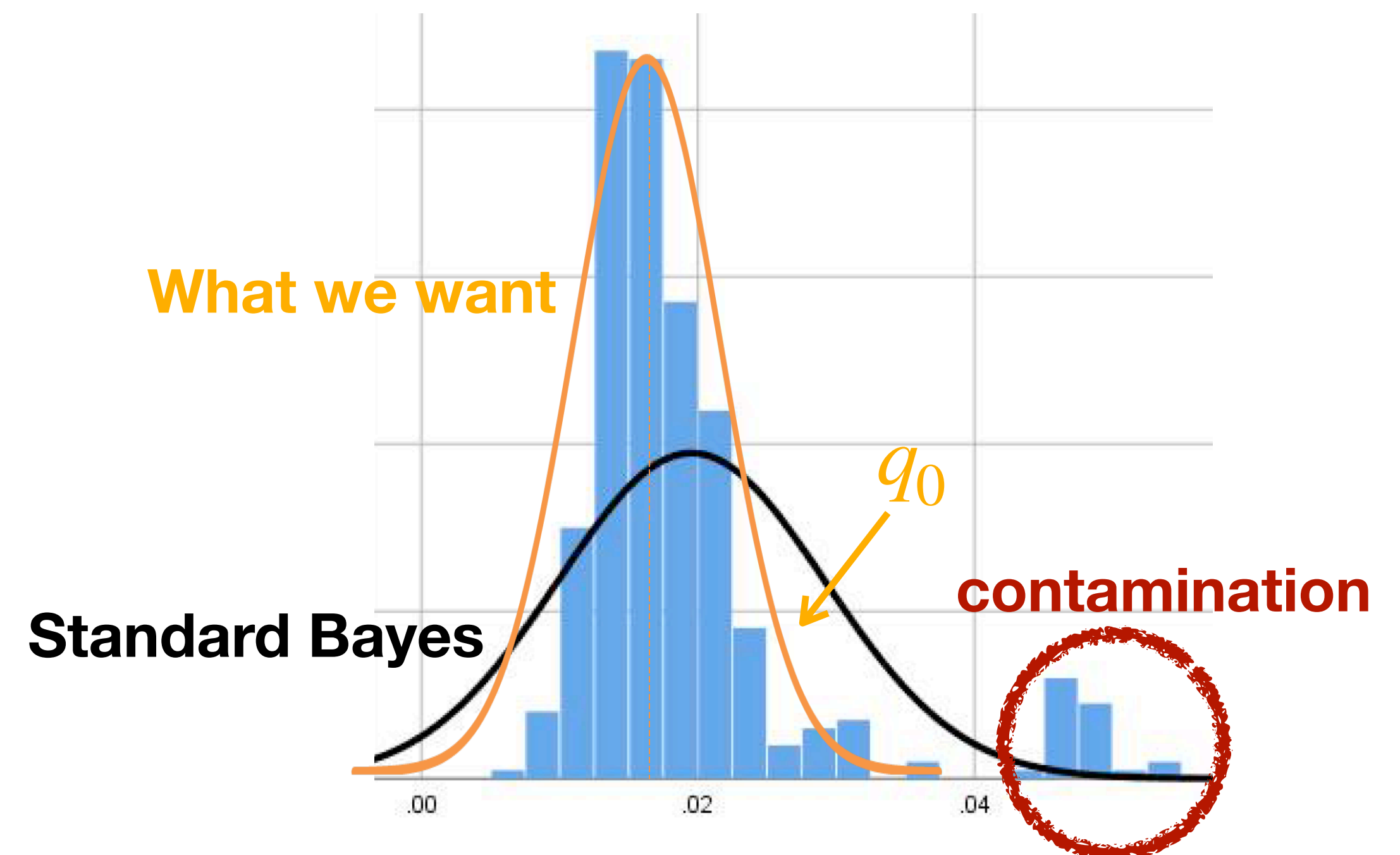
$$q_\varepsilon = (1 - \varepsilon) \cdot q_0 + \varepsilon \cdot c$$

Part of distribution our model captures

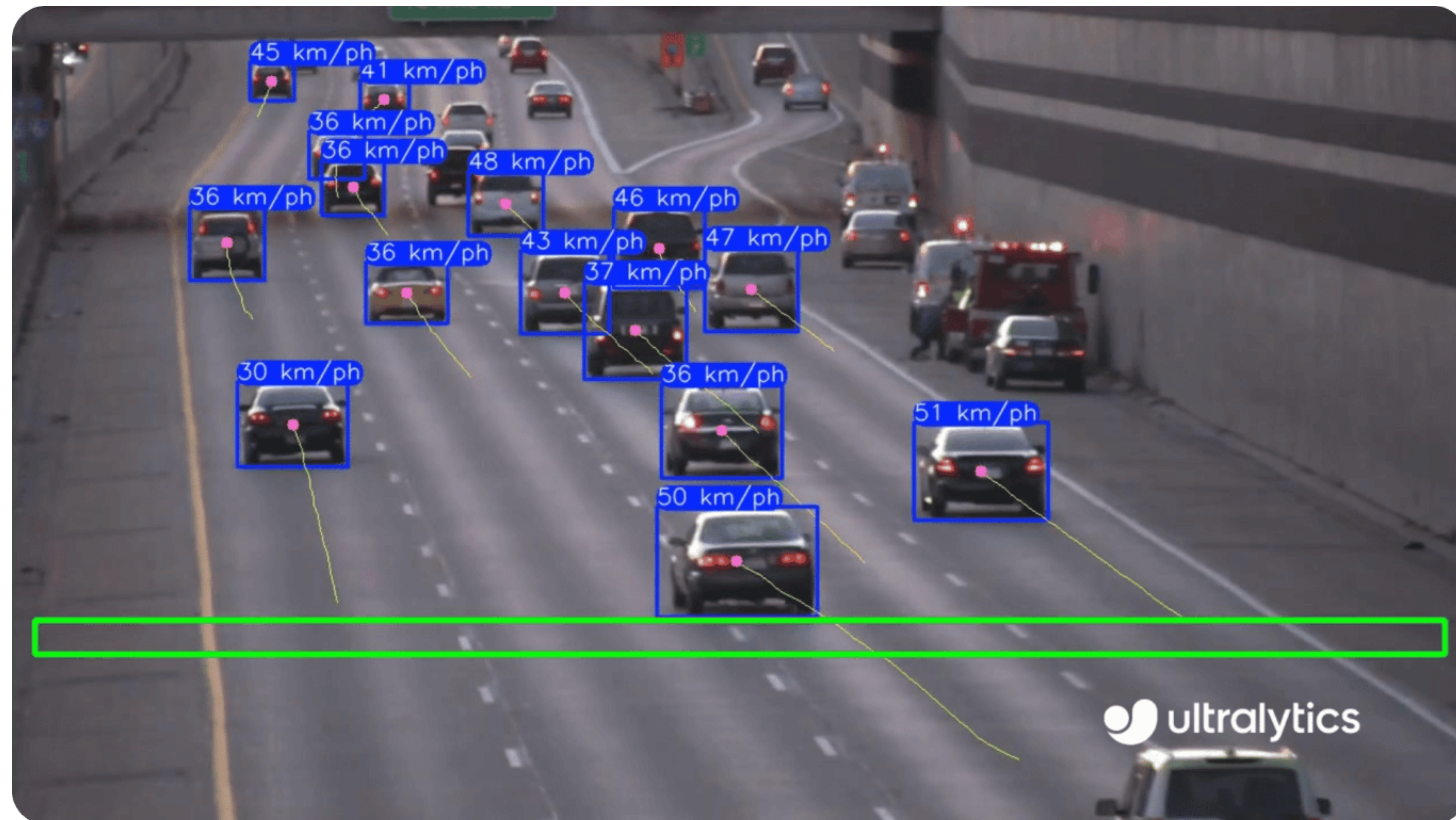
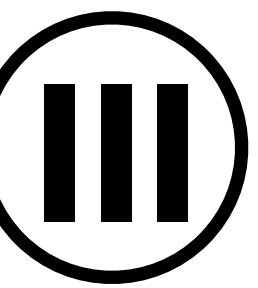
What we want:
$$\begin{cases} x_{1:n} \sim q_\varepsilon \longrightarrow \pi_n^{\mathbf{L}}(\theta \mid x_{1:n}) \\ z_{1:n} \sim q_0 \longrightarrow \pi_n^{\mathbf{L}}(\theta \mid z_{1:n}) \end{cases} \approx$$

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

Generally untrue for log likelihoods!

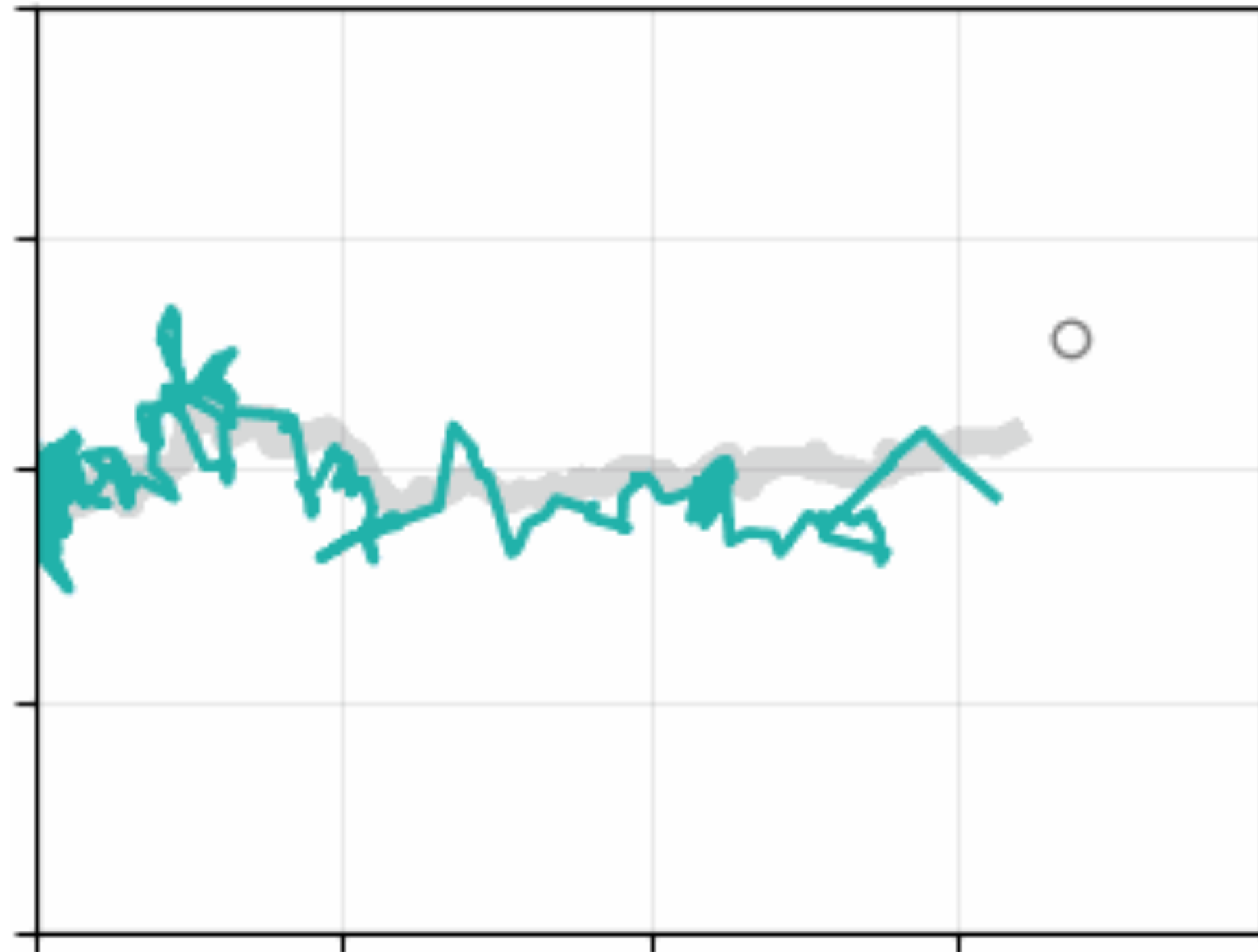


Gibbs Posteriors: Kalman Filtering/Tracking



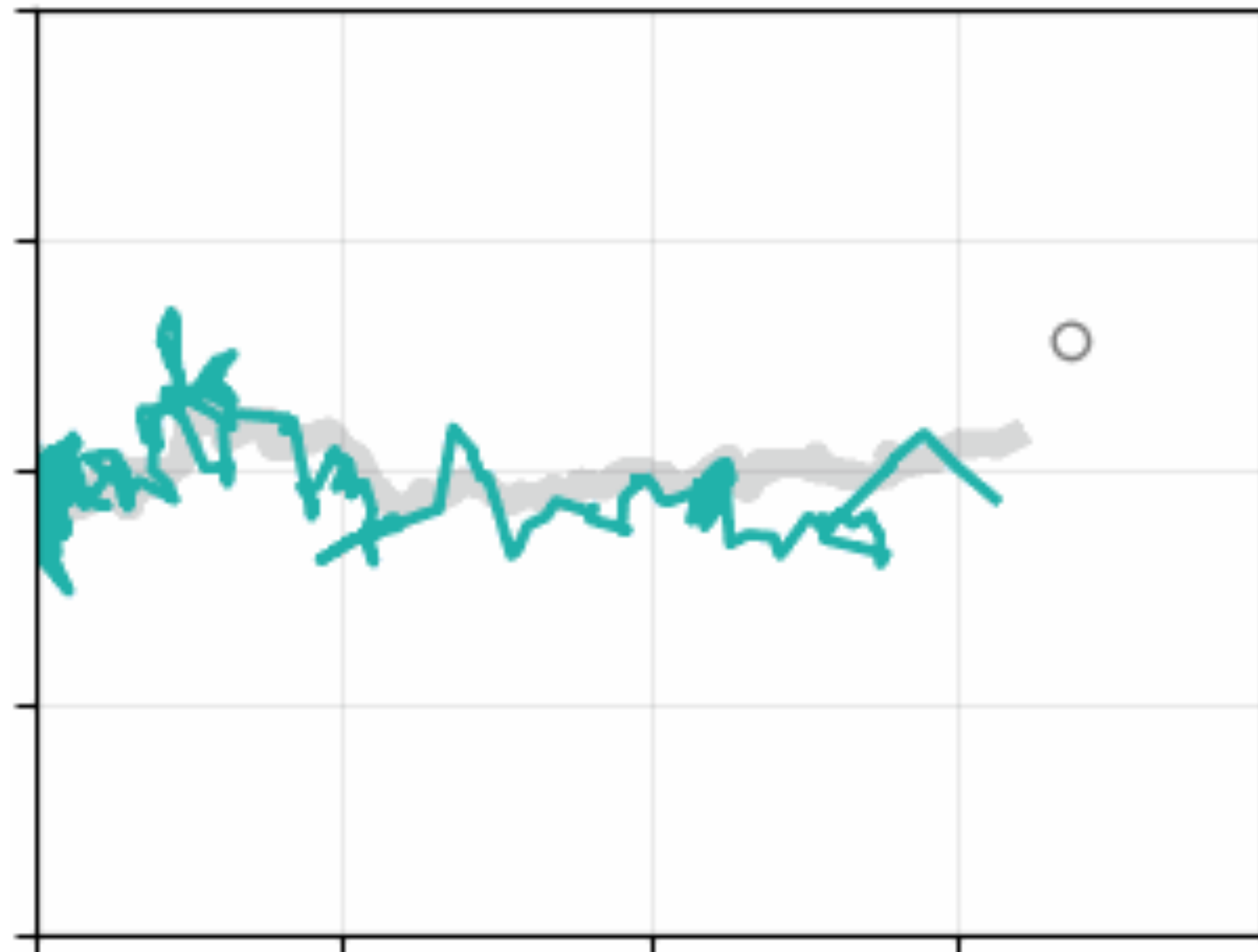
Gibbs Posteriors: Kalman Filtering/Tracking III

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

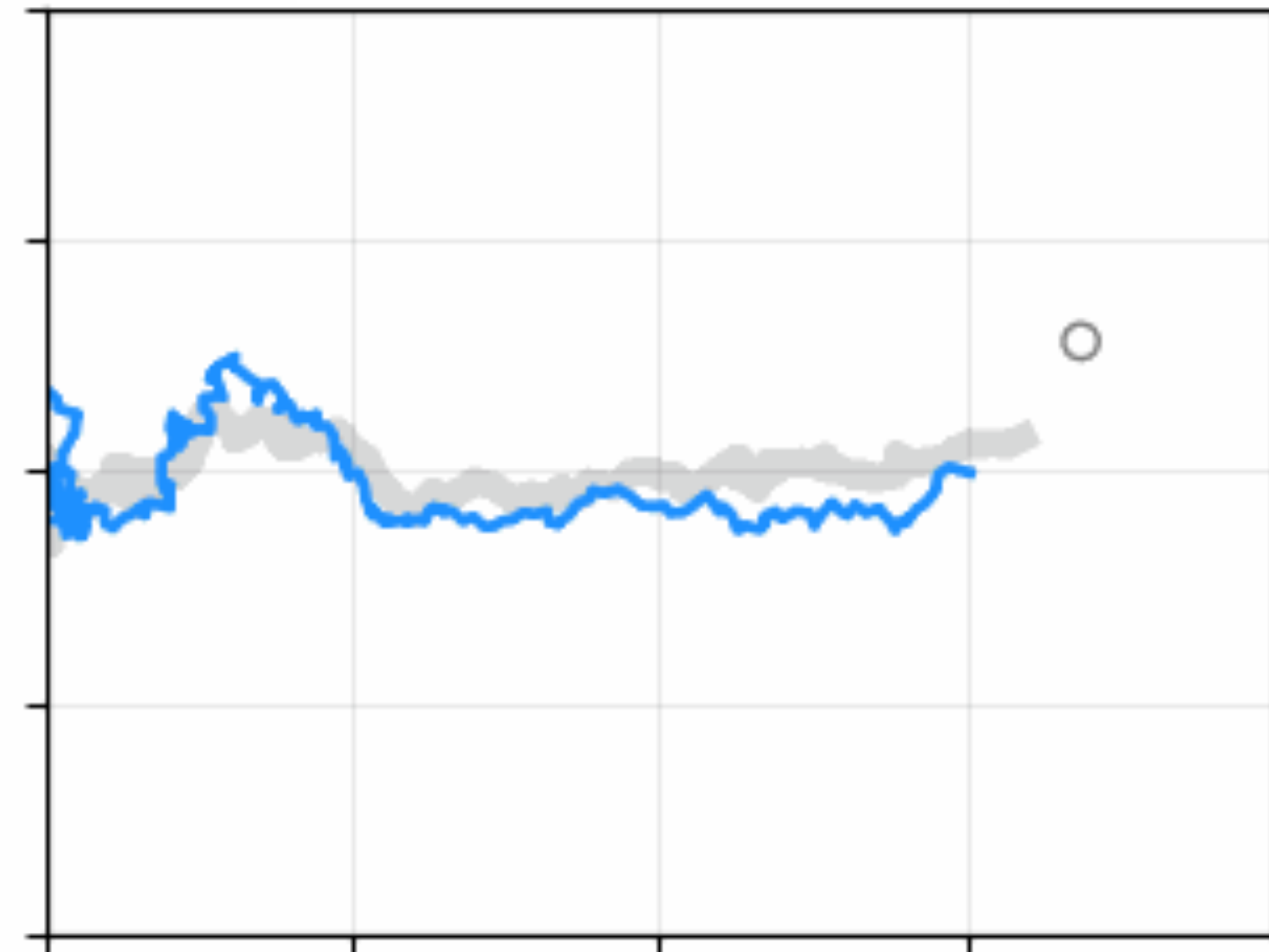


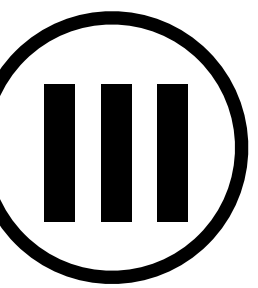
Gibbs Posteriors: Kalman Filtering/Tracking III

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



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



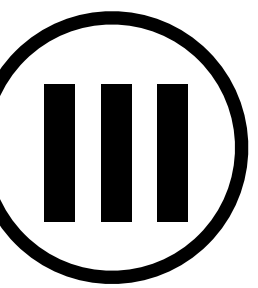


... why the Gibbs posterior approach?

Existing Approaches:

Outlier detection/removal before fitting

Extended Models (e.g. Student t Likelihood, Mixture models, ...)



... why the Gibbs posterior approach?

Existing Approaches:

Outlier detection/removal before fitting

× Only for very low-dimensional data

Extended Models (e.g. Student t Likelihood, Mixture models, ...)

... why the Gibbs posterior approach?

Existing Approaches:

Outlier detection/removal before fitting

✗ Only for very low-dimensional data

Extended Models (e.g. Student t Likelihood, Mixture models, ...)

✗ Only as good/robust as the new model

... why the Gibbs posterior approach?

Existing Approaches:

Outlier detection/removal before fitting

✗ Only for very low-dimensional data

Extended Models (e.g. Student t Likelihood, Mixture models, ...)

✗ Only as good/robust as the new model

✗ Posteriors do not have closed forms => slow

... why the Gibbs posterior approach?

Existing Approaches:

Outlier detection/removal before fitting

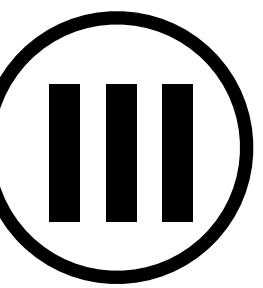
✗ Only for very low-dimensional data

Extended Models (e.g. Student t Likelihood, Mixture models, ...)

✗ Only as good/robust as the new model

✗ Posteriors do not have closed forms => slow

Gibbs posteriors: Robustness + closed forms!



... why the Gibbs posterior approach?

Gaussian Processes

Proposition 3.1. Suppose $f \sim \mathcal{GP}(m, k)$ and $\varepsilon \sim \mathcal{N}(0, I_n \sigma^2)$. Then, the RCGP posterior is

$$\begin{aligned} p^w(\mathbf{f}|\mathbf{y}, \mathbf{x}) &= \mathcal{N}(\mathbf{f}; \mu^R, \Sigma^R), \\ \mu^R &= \mathbf{m} + K (K + \sigma^2 J_{\mathbf{w}})^{-1} (\mathbf{y} - \mathbf{m}_{\mathbf{w}}), \\ \Sigma^R &= K (K + \sigma^2 J_{\mathbf{w}})^{-1} \sigma^2 J_{\mathbf{w}}, \end{aligned}$$

for $\mathbf{w} = (w(x_1, y_1), \dots, w(x_n, y_n))^T$, $\mathbf{m}_{\mathbf{w}} = \mathbf{m} + \sigma^2 \nabla_y \log(\mathbf{w}^2)$ and $J_{\mathbf{w}} = \text{diag}(\frac{\sigma^2}{2} \mathbf{w}^{-2})$. The RCGP's posterior predictive over $f_{\star} = f(x_{\star})$ at $x^{\star} \in \mathcal{X}$ is

$$\begin{aligned} p^w(f_{\star}|x_{\star}, \mathbf{x}, \mathbf{y}) &= \int_{\mathbb{R}^n} p(f_{\star}|x_{\star}, \mathbf{f}, \mathbf{x}, \mathbf{y}) p^w(\mathbf{f}|\mathbf{y}, \mathbf{x}) d\mathbf{f} \\ &= \mathcal{N}(f_{\star}; \mu_{\star}^R, \Sigma_{\star}^R), \\ \mu_{\star}^R &= m_{\star} + \mathbf{k}_{\star}^T (K + \sigma^2 J_{\mathbf{w}})^{-1} (\mathbf{y} - \mathbf{m}_{\mathbf{w}}), \\ \Sigma_{\star}^R &= k_{\star\star} - \mathbf{k}_{\star}^T (K + \sigma^2 J_{\mathbf{w}})^{-1} \mathbf{k}_{\star}. \end{aligned}$$

Altamirano, Briol, & K. (2024); ICML spotlight
Laplante, Altamirano, K., Duncan, & Briol (2025); ICML

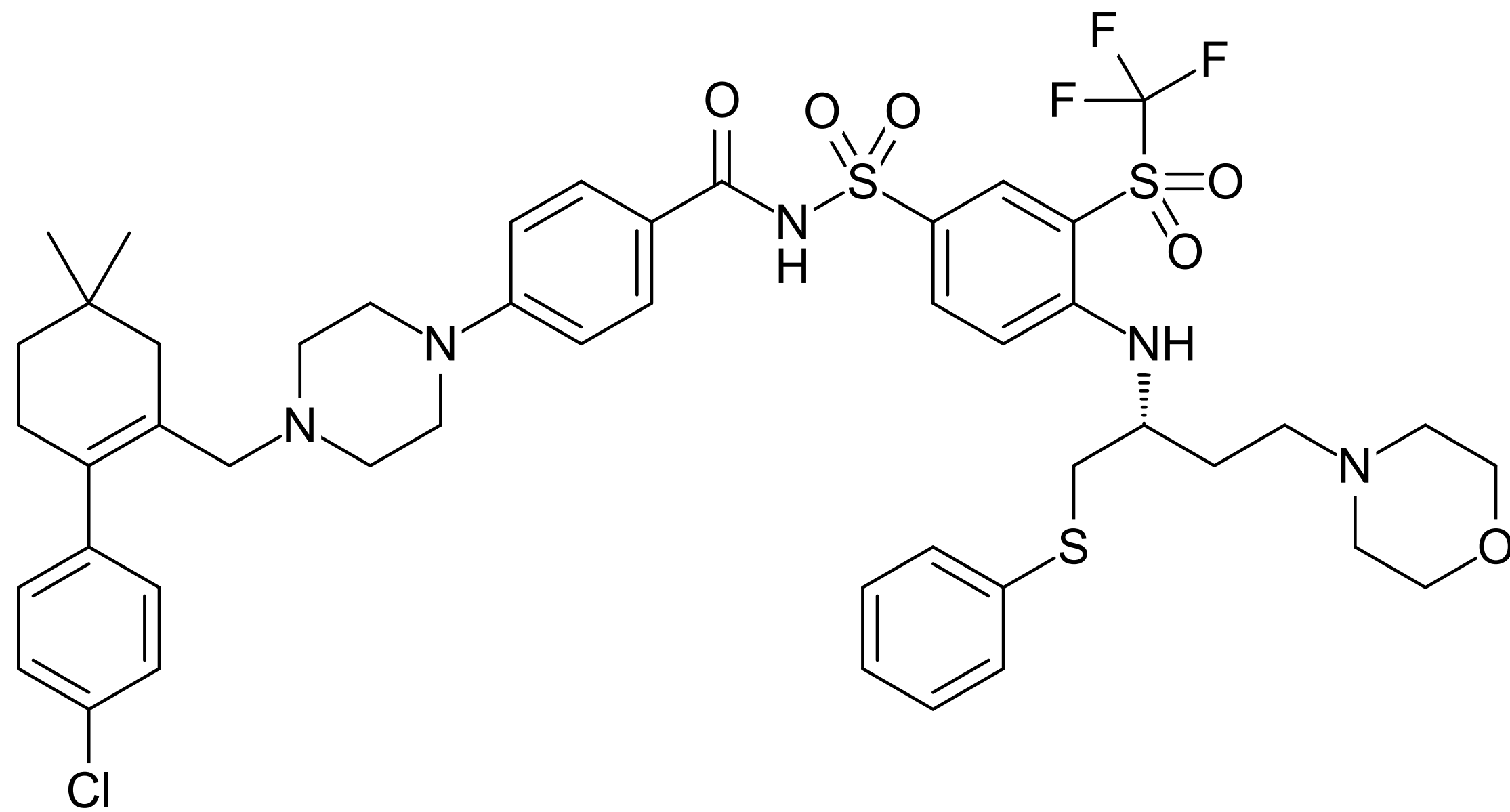


Posteriors do not have closed forms => slow

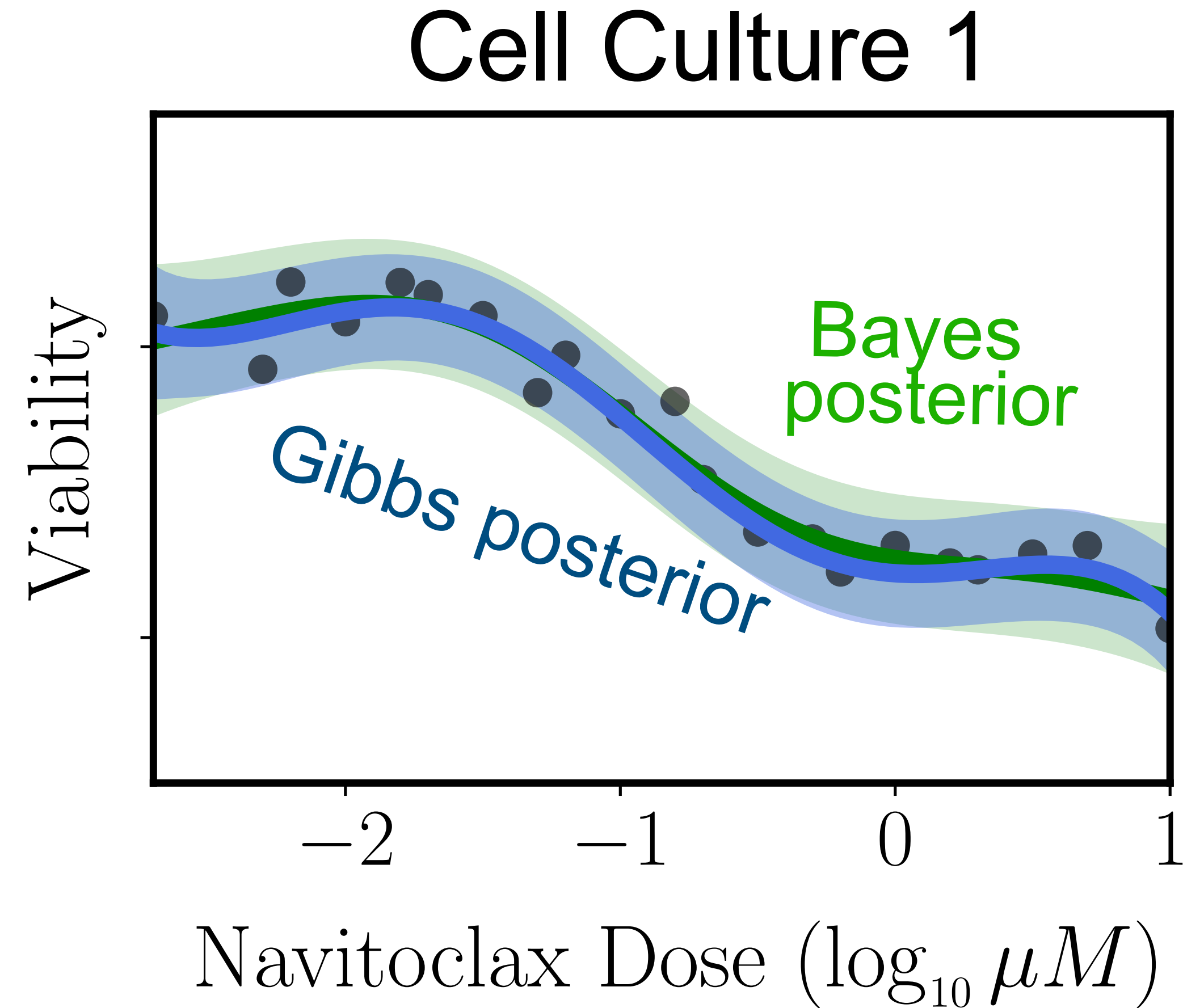
Gibbs posteriors: Robustness + closed forms!

GP Gibbs posteriors: Cancer Treatment

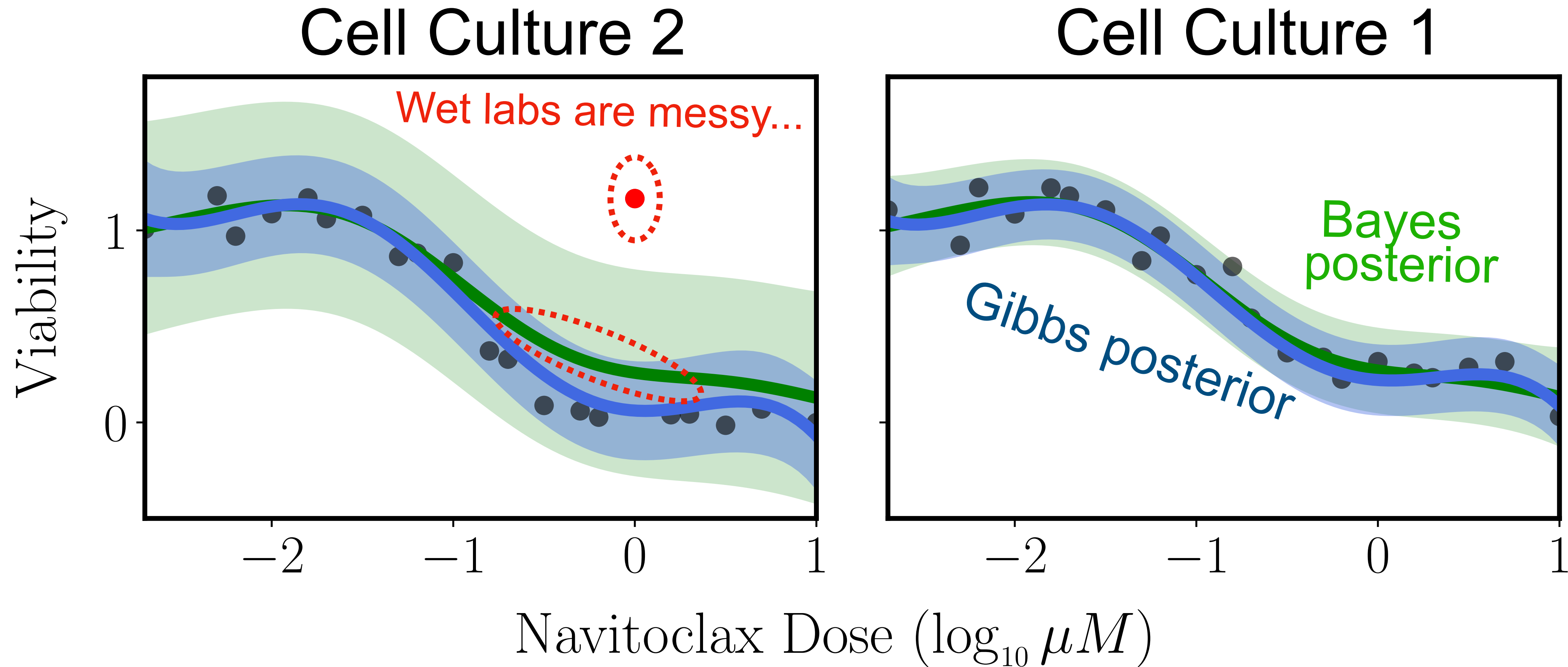
Navitoclax



GP Gibbs posteriors: Cancer Treatment



GP Gibbs posteriors: Cancer Treatment



... why the Gibbs posterior approach?



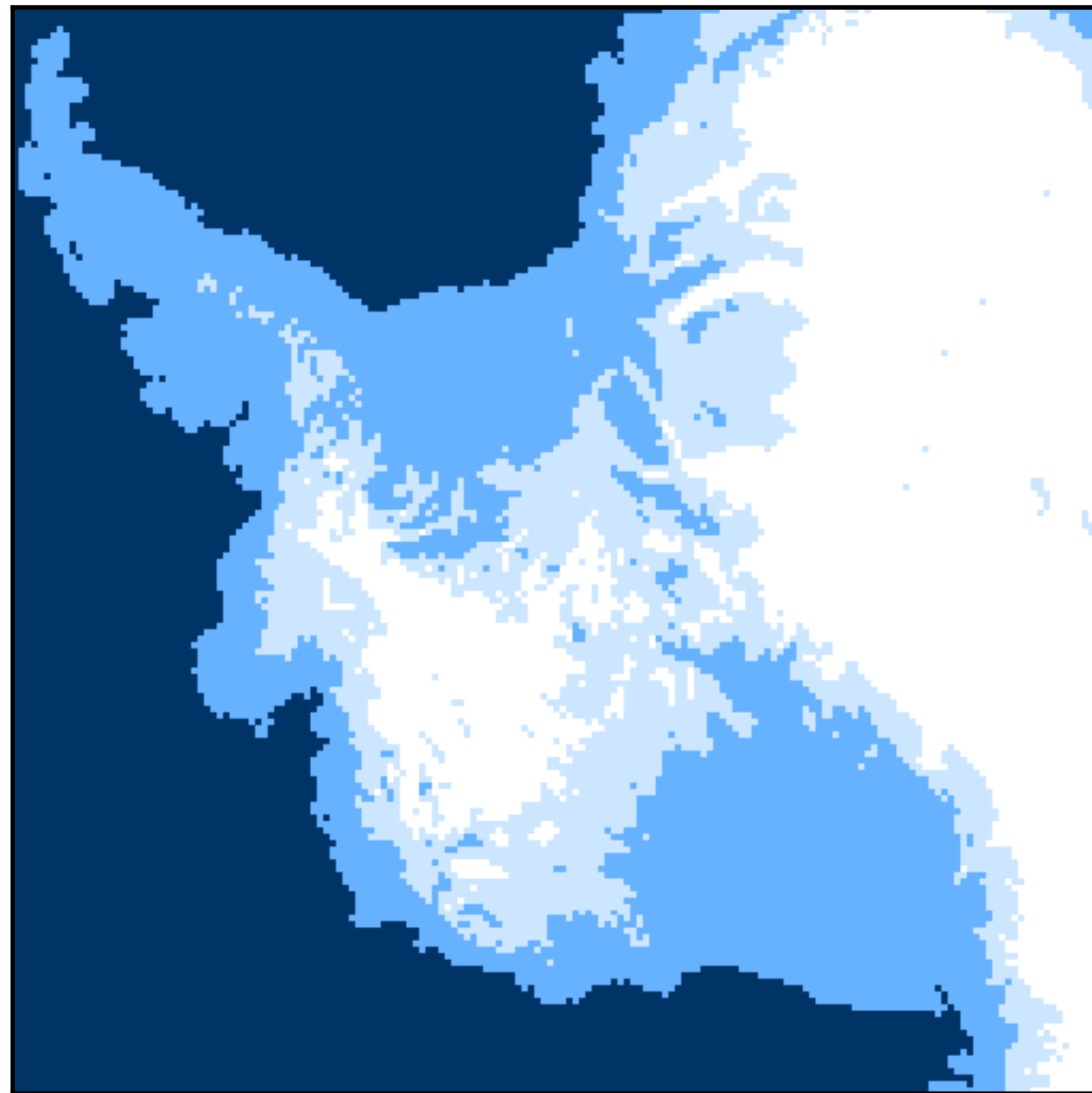
E.g. for **intractable** models like Potts/Ising Models



Posteriors do not have closed forms $\Rightarrow$ slow

Gibbs posteriors: ~~Robustness~~ + closed forms!

Gibbs posteriors: Ising/Potts Models

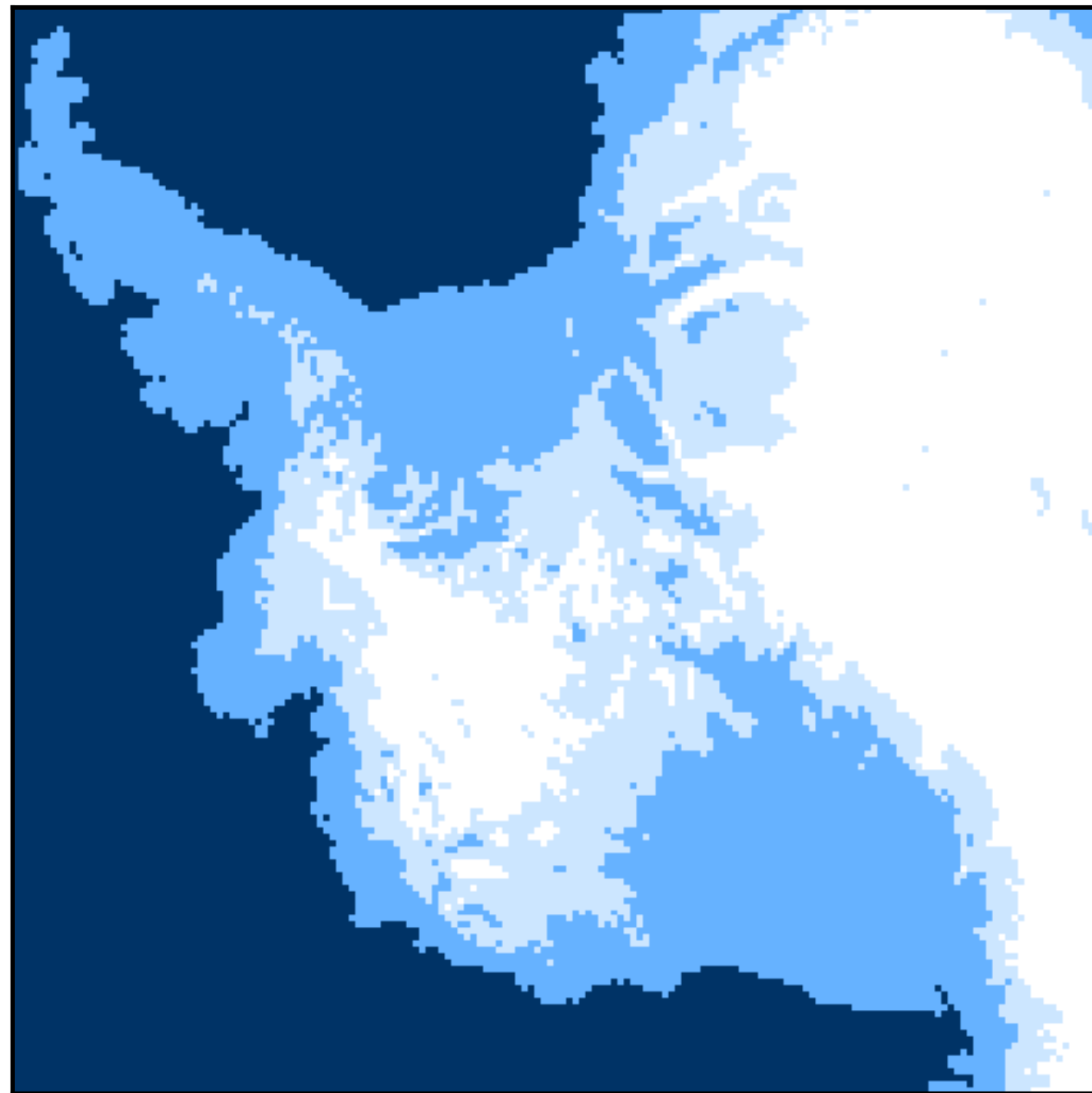


Potts Model
(for Ice Sheet thickness)
Grid size: 171 x 171

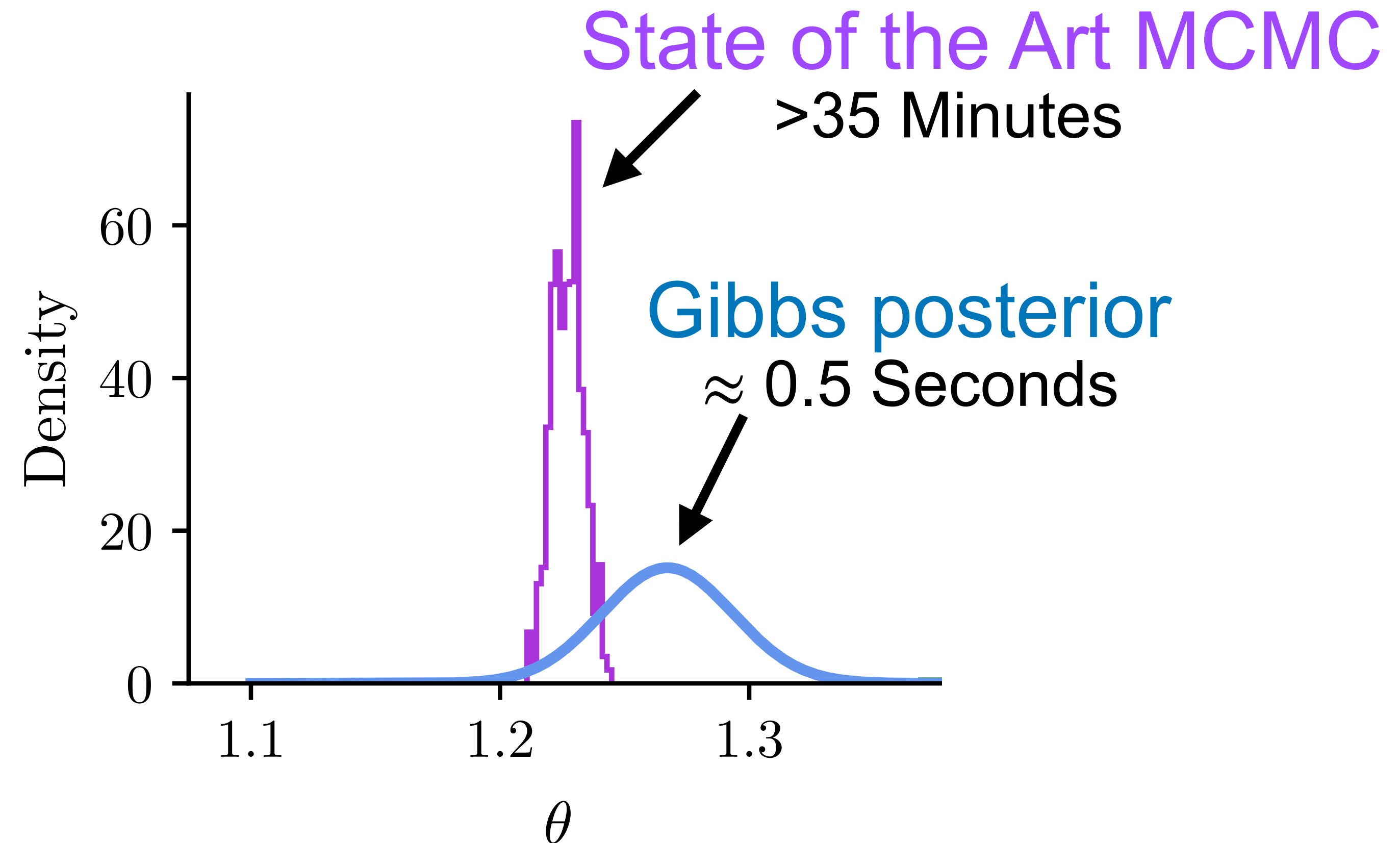
$$p_{\theta}(x) = \frac{1}{Z(\theta)} \exp \left\{ \theta \sum_{i \sim j} 1_{x_i = x_j} \right\}$$

intractable

Gibbs posteriors: Ising/Potts Models



Potts Model
(for Ice Sheet thickness)
Grid size: 171 x 171



Post-Bayesian Inference

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

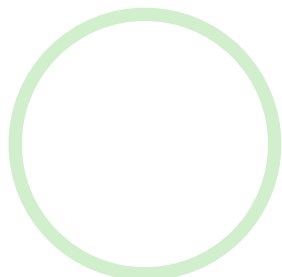
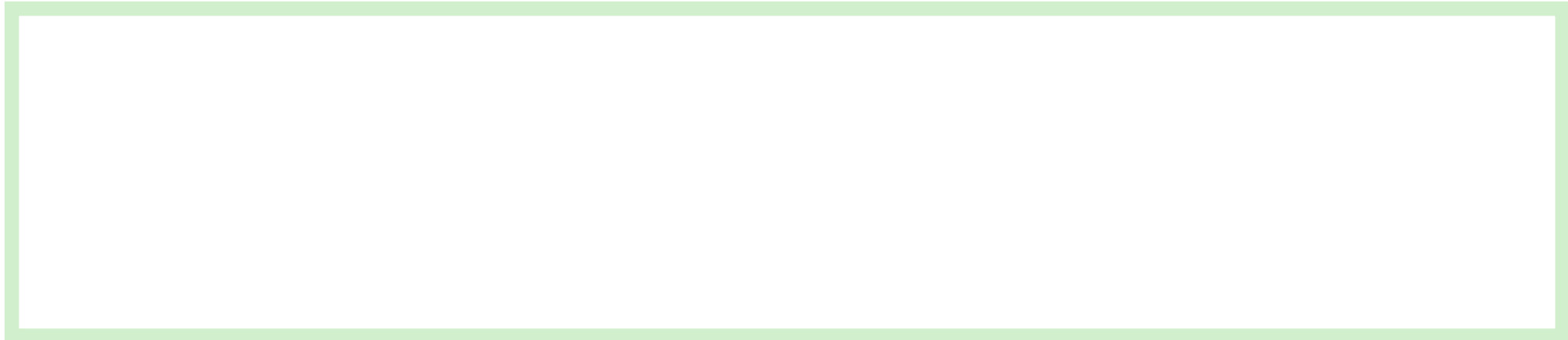
$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$

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

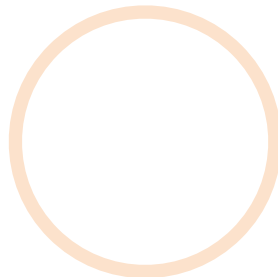
Predictively oriented Inference

VI

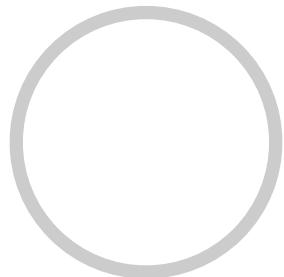
~~(A1)~~, (A2), (A3)



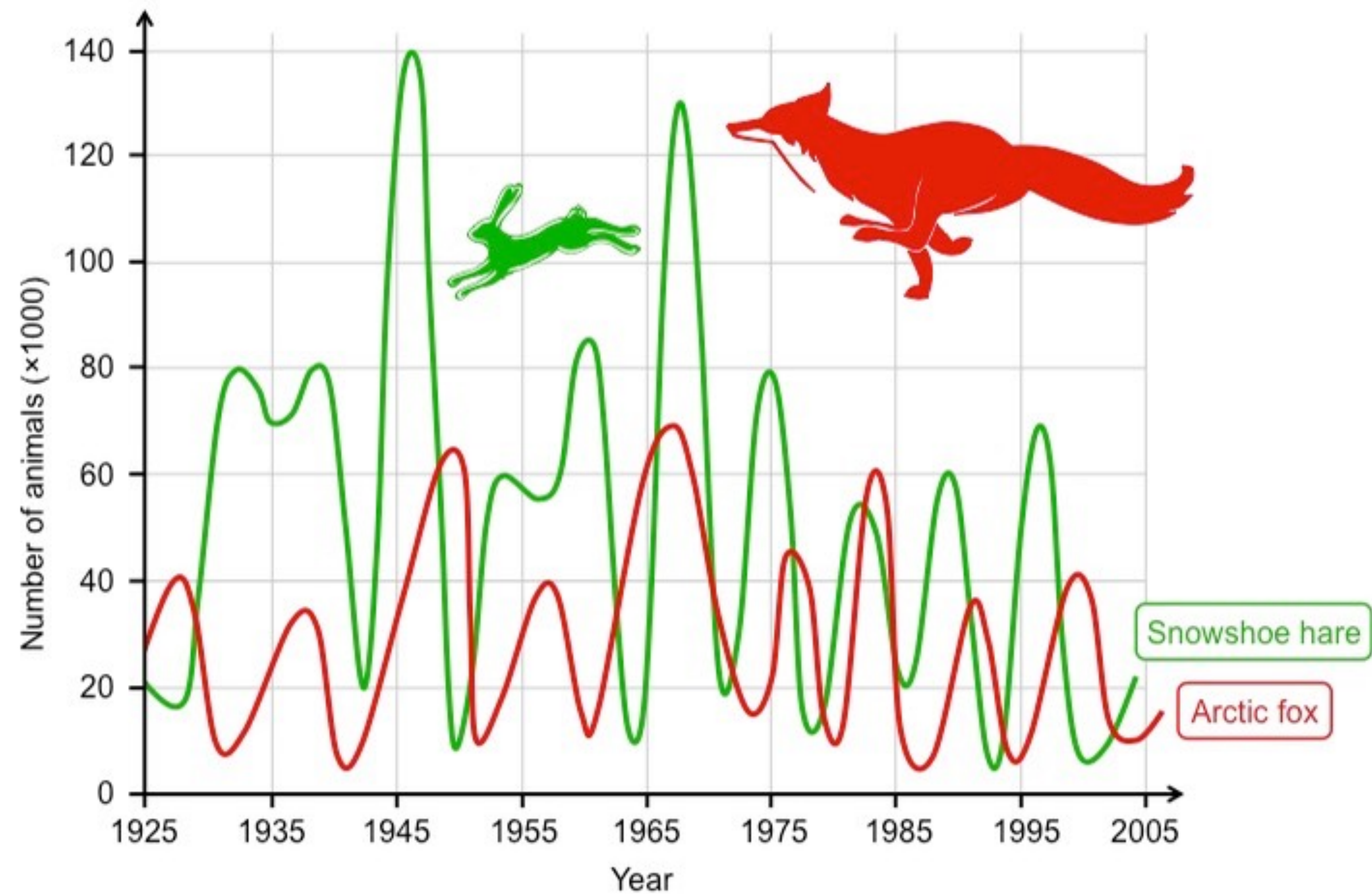
~~(A1)~~, (A2), (A3)



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



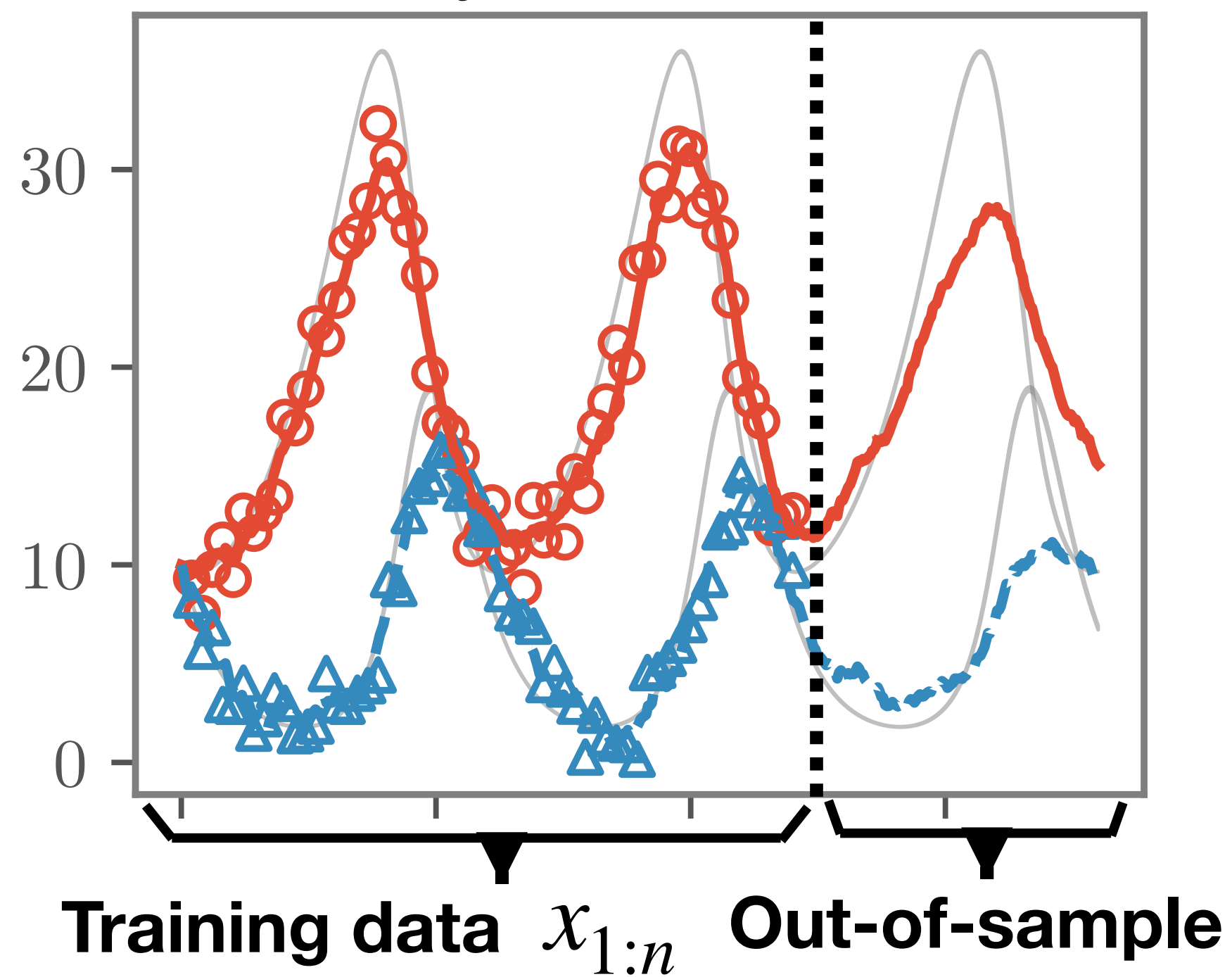
Predictively Oriented Posteriors: Lotka-Volterra Model



Predictively Oriented Posteriors: Lotka-Volterra Model

Bayes posterior

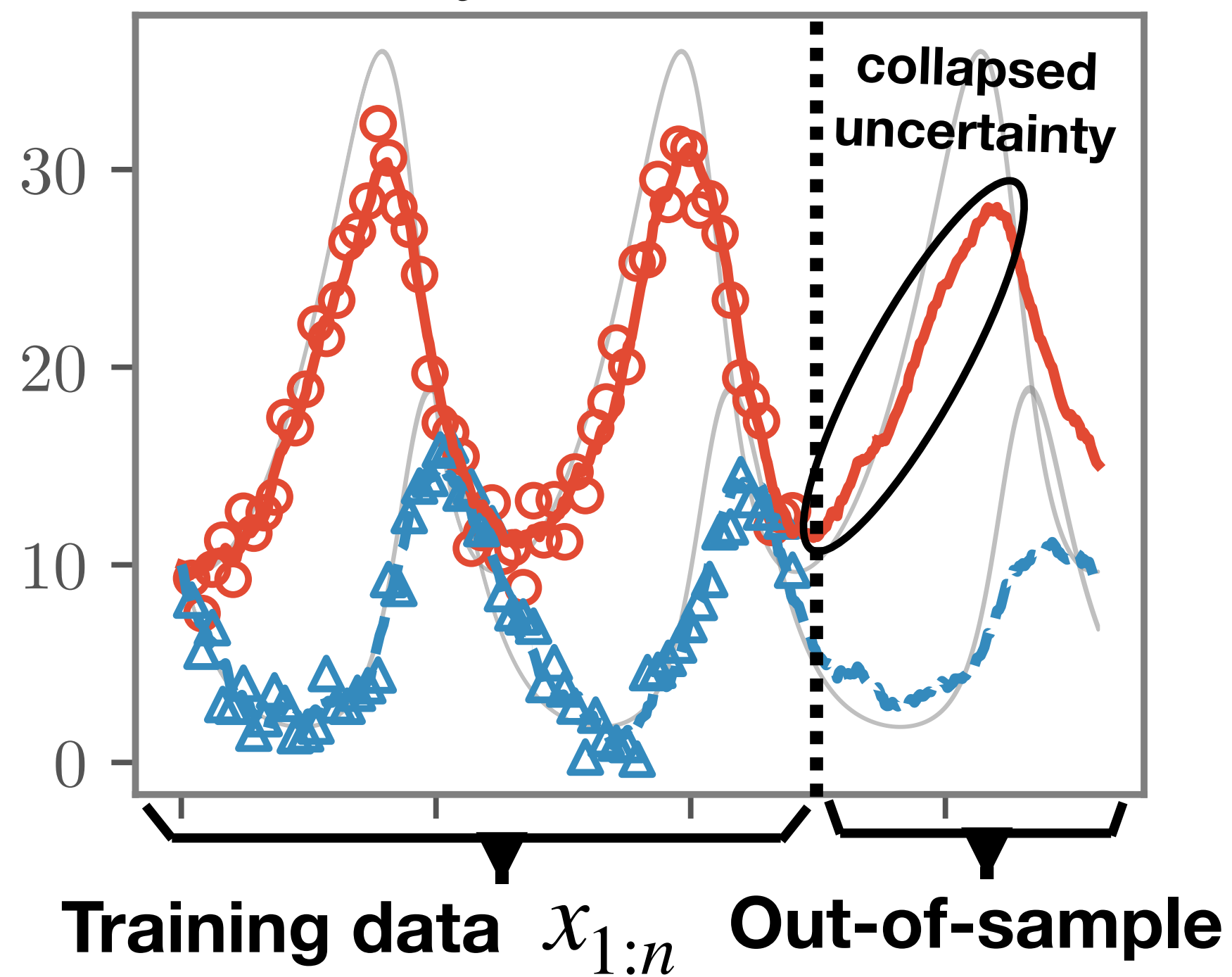
$$P_n(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) \pi_n(\theta \mid x_{1:n}) d\theta$$



Predictively Oriented Posteriors: Lotka-Volterra Model

Bayes posterior

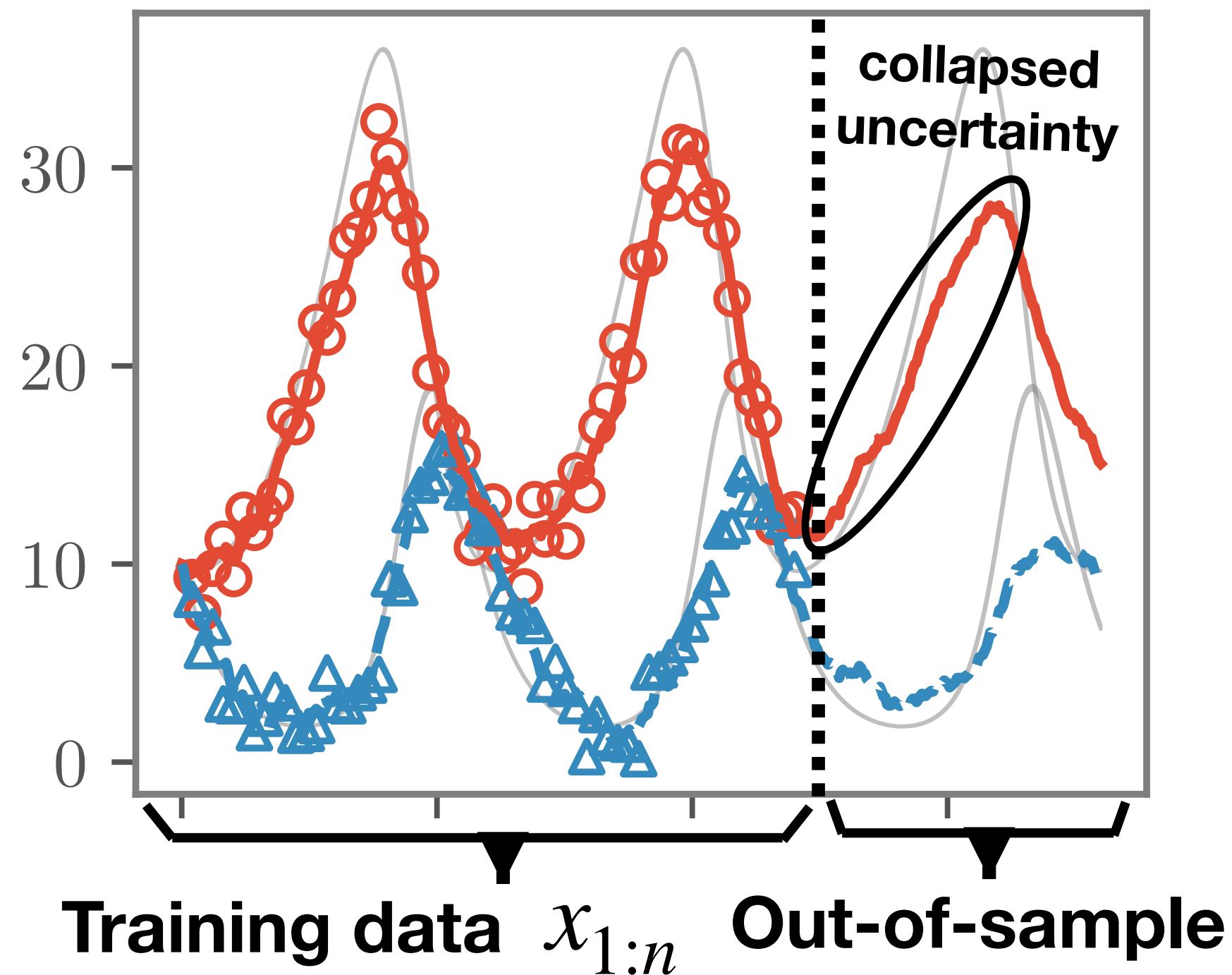
$$P_n(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) \pi_n(\theta \mid x_{1:n}) d\theta$$



Predictively Oriented Posteriors: Lotka-Volterra Model

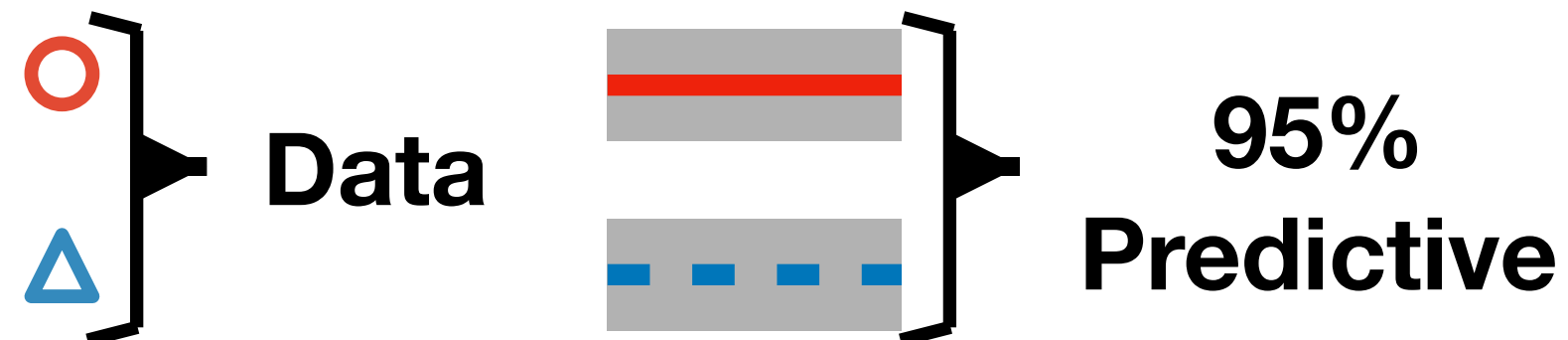
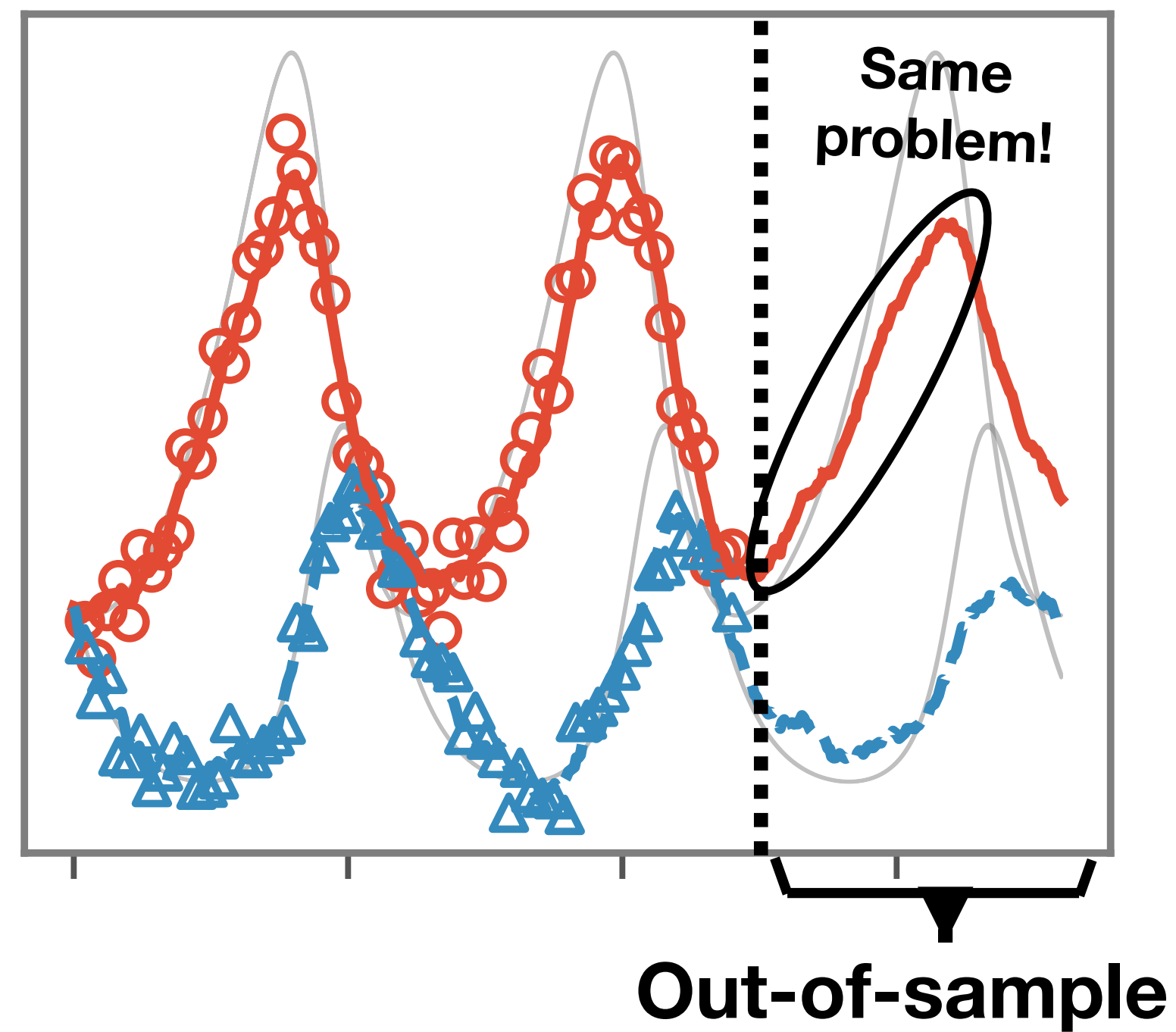
Bayes posterior

$$P_n(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) \pi_n(\theta \mid x_{1:n}) d\theta$$



Gibbs posterior

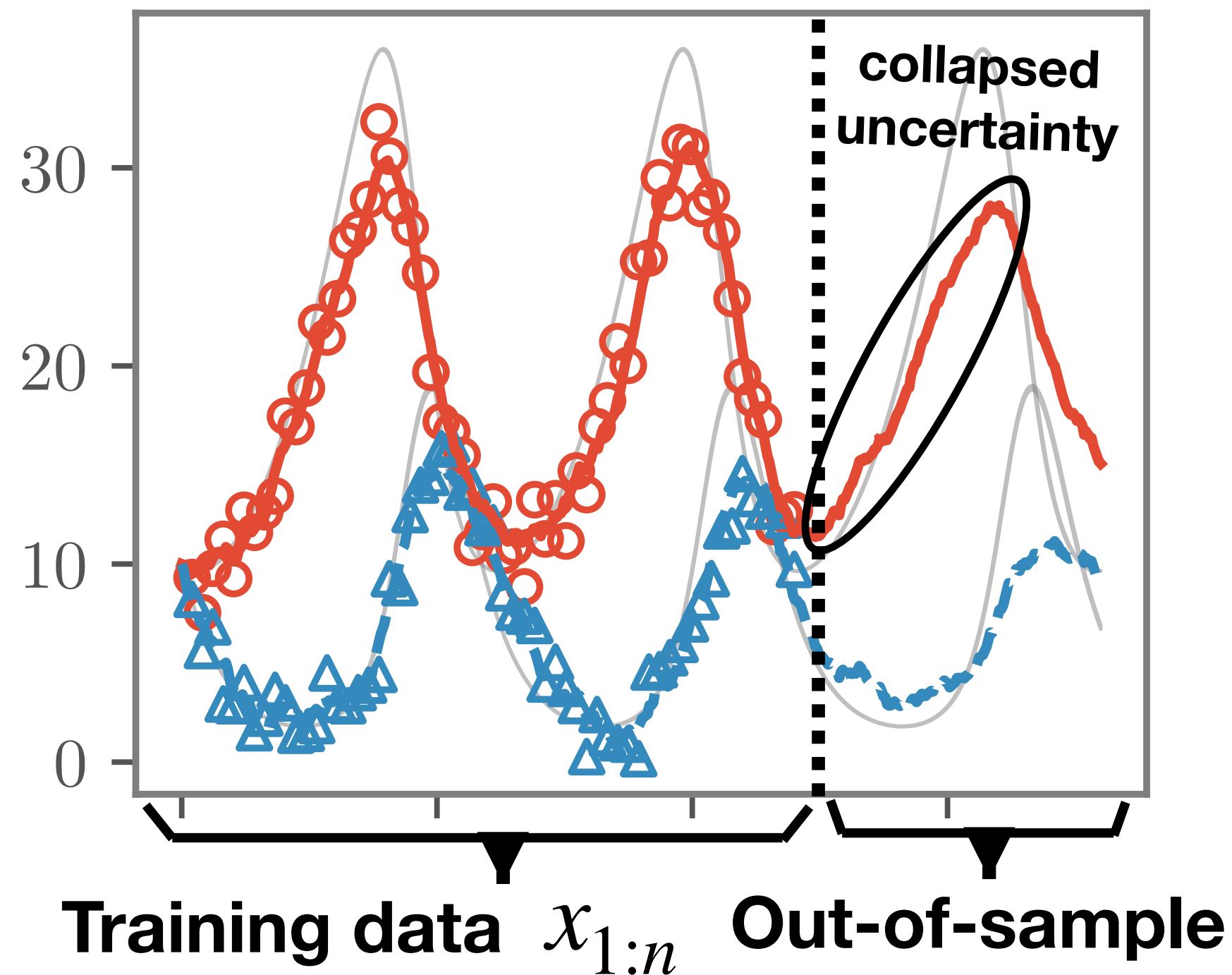
$$P_{\pi_n^L}(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) \pi_n^L(\theta \mid x_{1:n}) d\theta$$



Predictively Oriented Posteriors: Lotka-Volterra Model

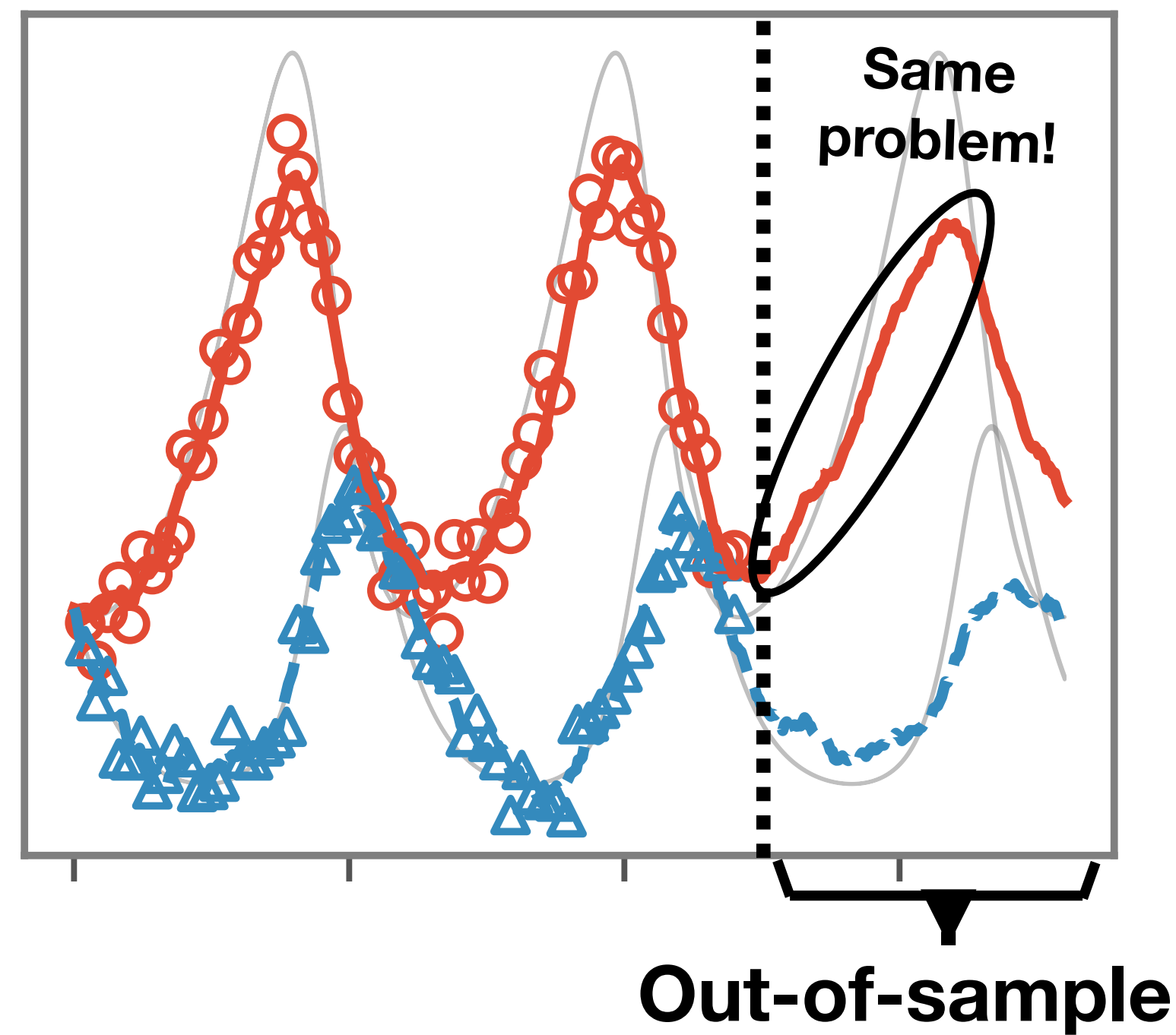
Bayes posterior

$$P_n(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) \pi_n(\theta \mid x_{1:n}) d\theta$$



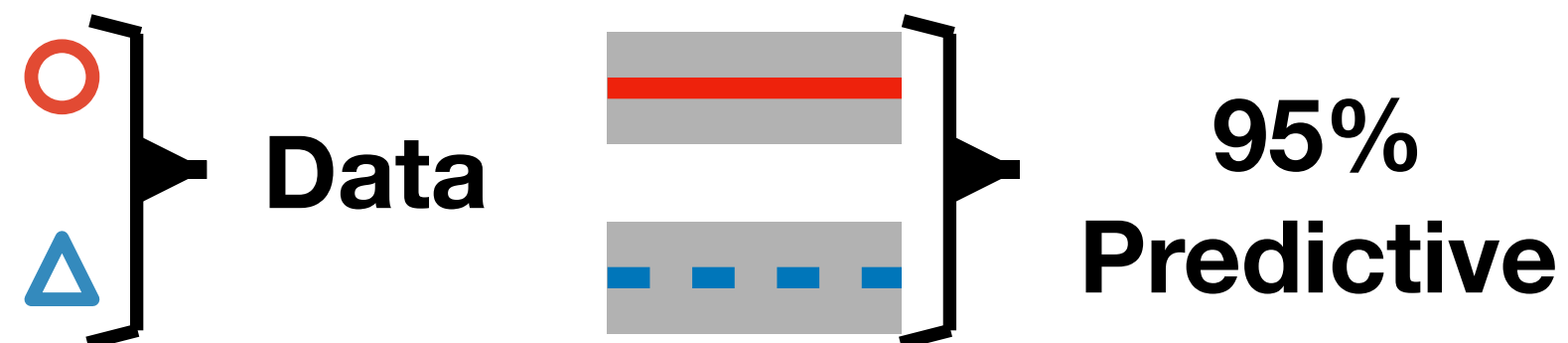
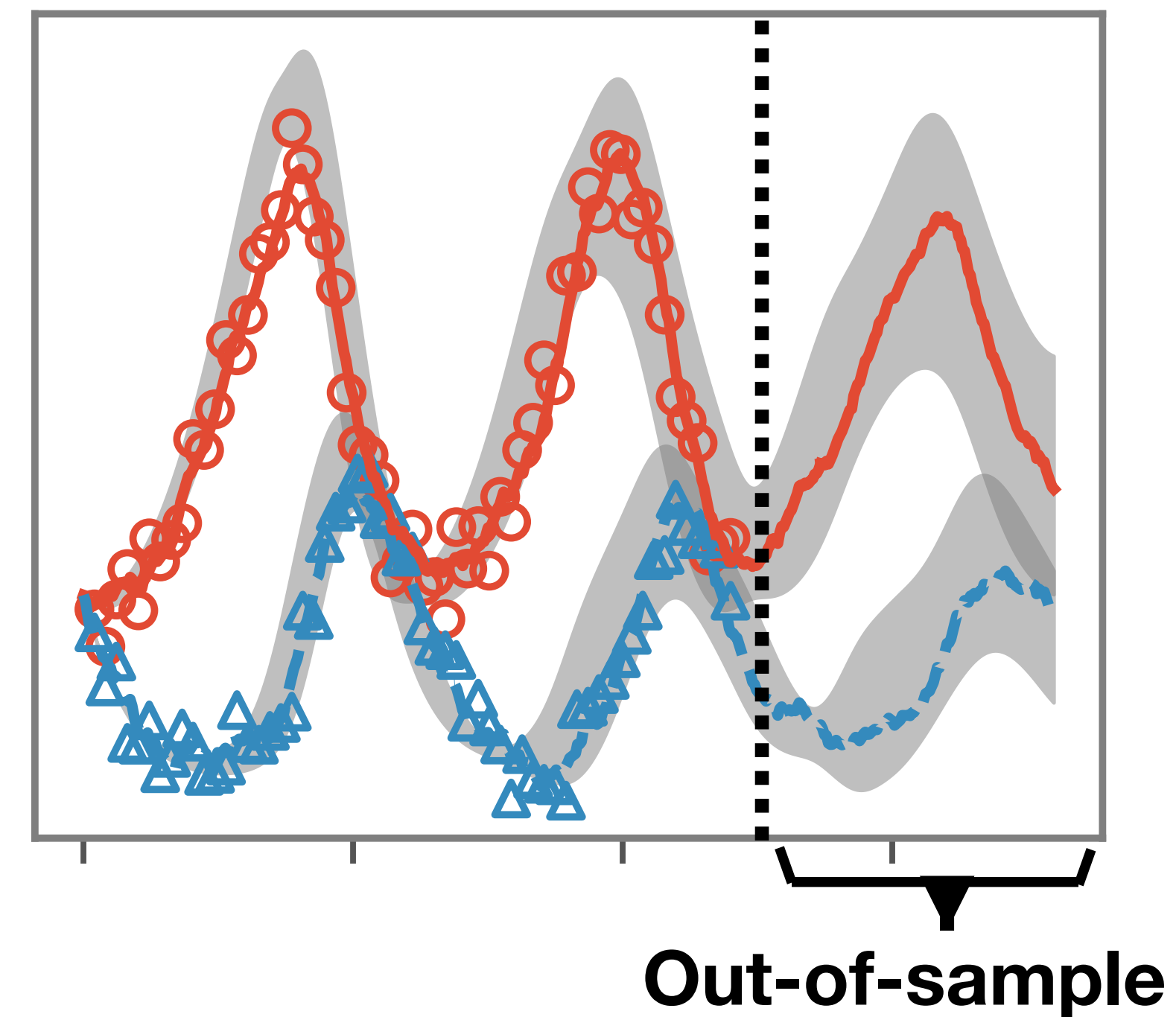
Gibbs posterior

$$P_{\pi_n^L}(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) \pi_n^L(\theta \mid x_{1:n}) d\theta$$



Prediction-centric posterior

$$P_{q_n}(x_{\text{test}}) = \int p_{\theta}(x_{\text{test}}) q_n(\theta) d\theta$$



Constructing Predictively Oriented (PrO) Posteriors

Gibbs Posterior based on L

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

Masegosa (2020); NeurIPS

Jankowiak, Pleiss, & Gardner (2020); ICML; (2020); (UAI)

Morningstar, Alemi, & Dillon (2022); AISTATS

Wild, Ghalebikesabi, Sejdinovic, & K. (2024); NeurIPS (oral)

Shen, K., Power, & Oates (2025); AISTATS

Constructing Predictively Oriented (PrO) Posteriors

Target best parameter

Gibbs Posterior based on L

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int L(x_{1:n}, p_\theta) dq(\theta) + \text{KL}(q, \pi) \right\} = \pi_n^L(\theta \mid x_{1:n}) = \frac{\exp\{-L(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-L(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Masegosa (2020); NeurIPS

Jankowiak, Pleiss, & Gardner (2020); ICML; (2020); (UAI)

Morningstar, Alemi, & Dillon (2022); AISTATS

Wild, Ghalebikesabi, Sejdinovic, & K. (2024); NeurIPS (oral)

Shen, K., Power, & Oates (2025); AISTATS

Constructing Predictively Oriented (PrO) Posteriors

Target best parameter

Gibbs Posterior based on L

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int L(x_{1:n}, p_\theta) dq(\theta) + \text{KL}(q, \pi) \right\} = \pi_n^L(\theta \mid x_{1:n}) = \frac{\exp\{-L(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-L(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Target best predictive

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ L\left(x_{1:n}, \int p_\theta dq(\theta)\right) + \text{KL}(q, \pi) \right\}$$

Masegosa (2020); NeurIPS
 Jankowiak, Pleiss, & Gardner (2020); ICML; (2020); (UAI)
 Morningstar, Alemi, & Dillon (2022); AISTATS
 Wild, Ghalebikesabi, Sejdinovic, & K. (2024); NeurIPS (oral)
 Shen, K., Power, & Oates (2025); AISTATS

Constructing Predictively Oriented (PrO) Posteriors

Target best parameter

Gibbs Posterior based on L

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int L(x_{1:n}, p_\theta) dq(\theta) + \text{KL}(q, \pi) \right\} = \pi_n^L(\theta | x_{1:n}) = \frac{\exp\{-L(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-L(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Target best predictive

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ L\left(x_{1:n}, \underbrace{\int p_\theta dq(\theta)}_{\text{Implied predictive for } q}\right) + \text{KL}(q, \pi) \right\}$$

Implied predictive for q

Constructing Predictively Oriented (PrO) Posteriors

Target best parameter

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int \mathbf{L}(x_{1:n}, p_\theta) dq(\theta) + \text{KL}(q, \pi) \right\} = \pi_n^{\mathbf{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Gibbs Posterior based on $\mathbf{L}$

Target best predictive

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \mathbf{L}\left(x_{1:n}, \underbrace{\int p_\theta dq(\theta)}_{\text{Implied predictive for } q}\right) + \text{KL}(q, \pi) \right\} = q_n(\theta)$$

Implied predictive for q

Predictively oriented posterior based on $\mathbf{L}$

**Targets predictive directly.
... and doesn't collapse!**

Constructing Predictively Oriented (PrO) Posteriors

Target best parameter

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int \mathbf{L}(x_{1:n}, p_\theta) dq(\theta) + \text{KL}(q, \pi) \right\} = \pi_n^{\mathbf{L}}(\theta | x_{1:n}) = \frac{\exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta}$$

Gibbs Posterior based on $\mathbf{L}$

Target best predictive

$$\arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \mathbf{L}\left(x_{1:n}, \underbrace{\int p_\theta dq(\theta)}_{\text{Implied predictive for } q}\right) + \text{KL}(q, \pi) \right\} = q_n(\theta)$$

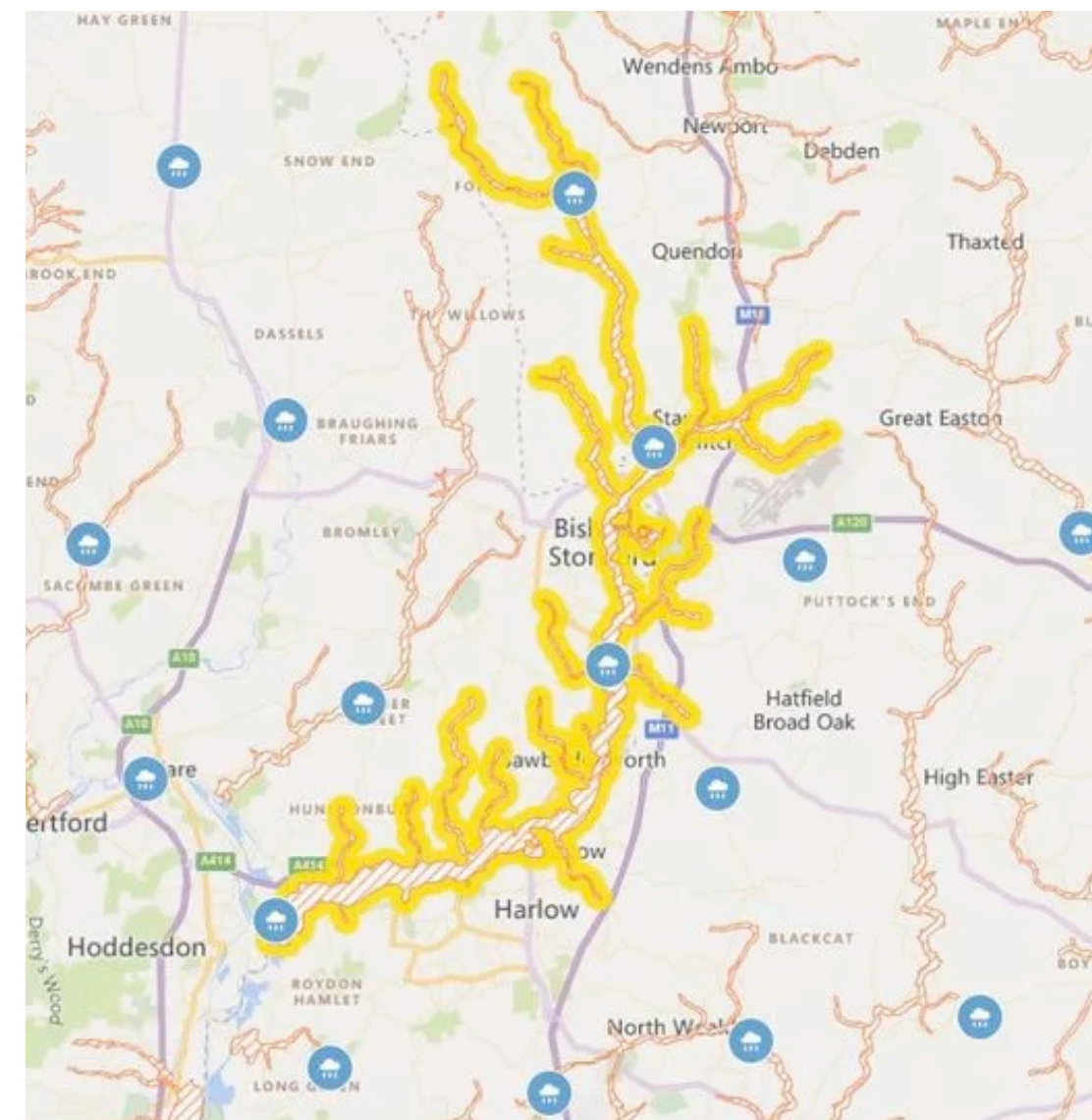
Implied predictive for q

Predictively oriented posterior based on $\mathbf{L}$

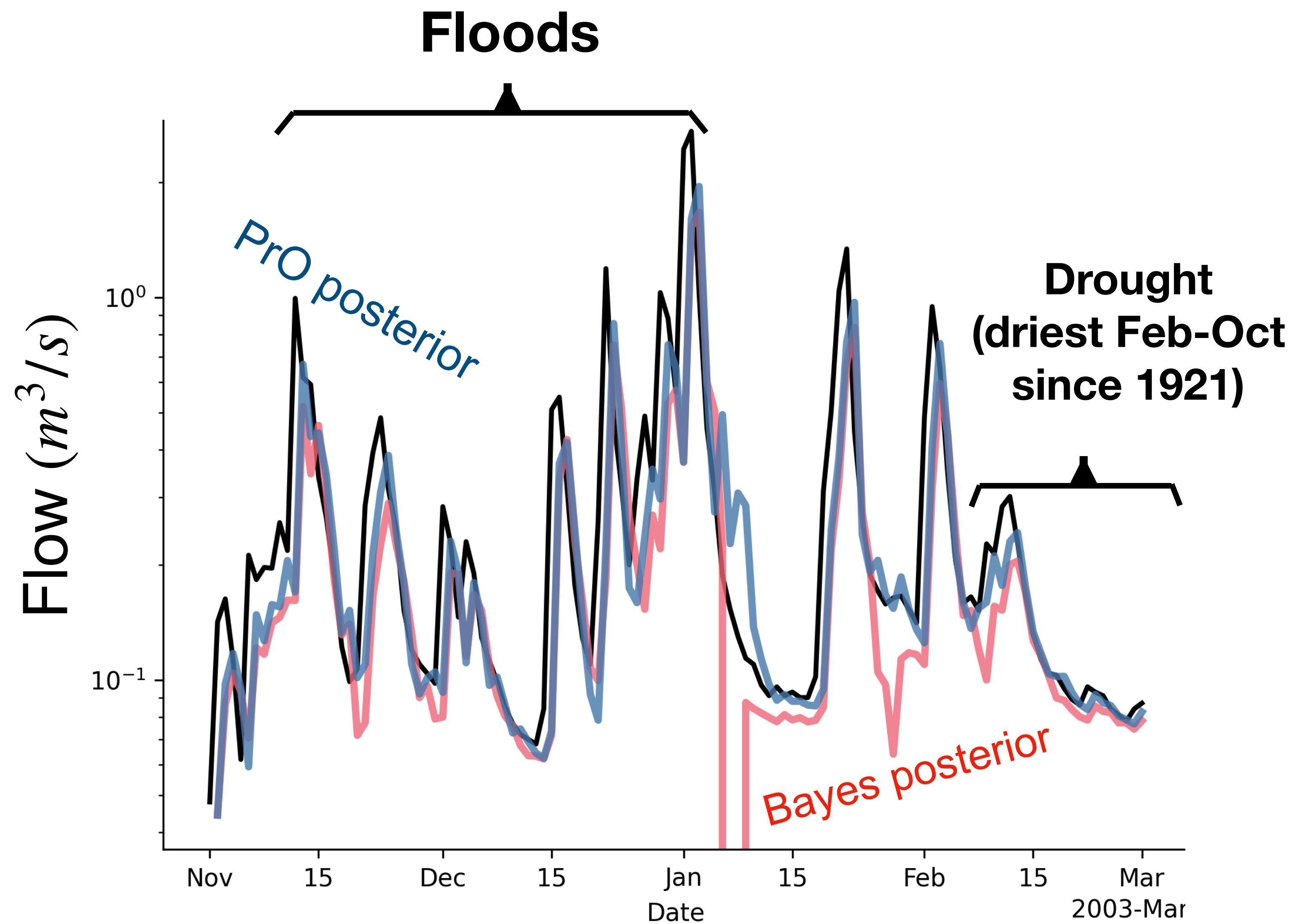
**Targets predictive directly.
... and doesn't collapse!**

Theorem: $q_n(\theta)$ is predictively superior to $\pi_n^{\mathbf{L}}(\theta | x_{1:n})$

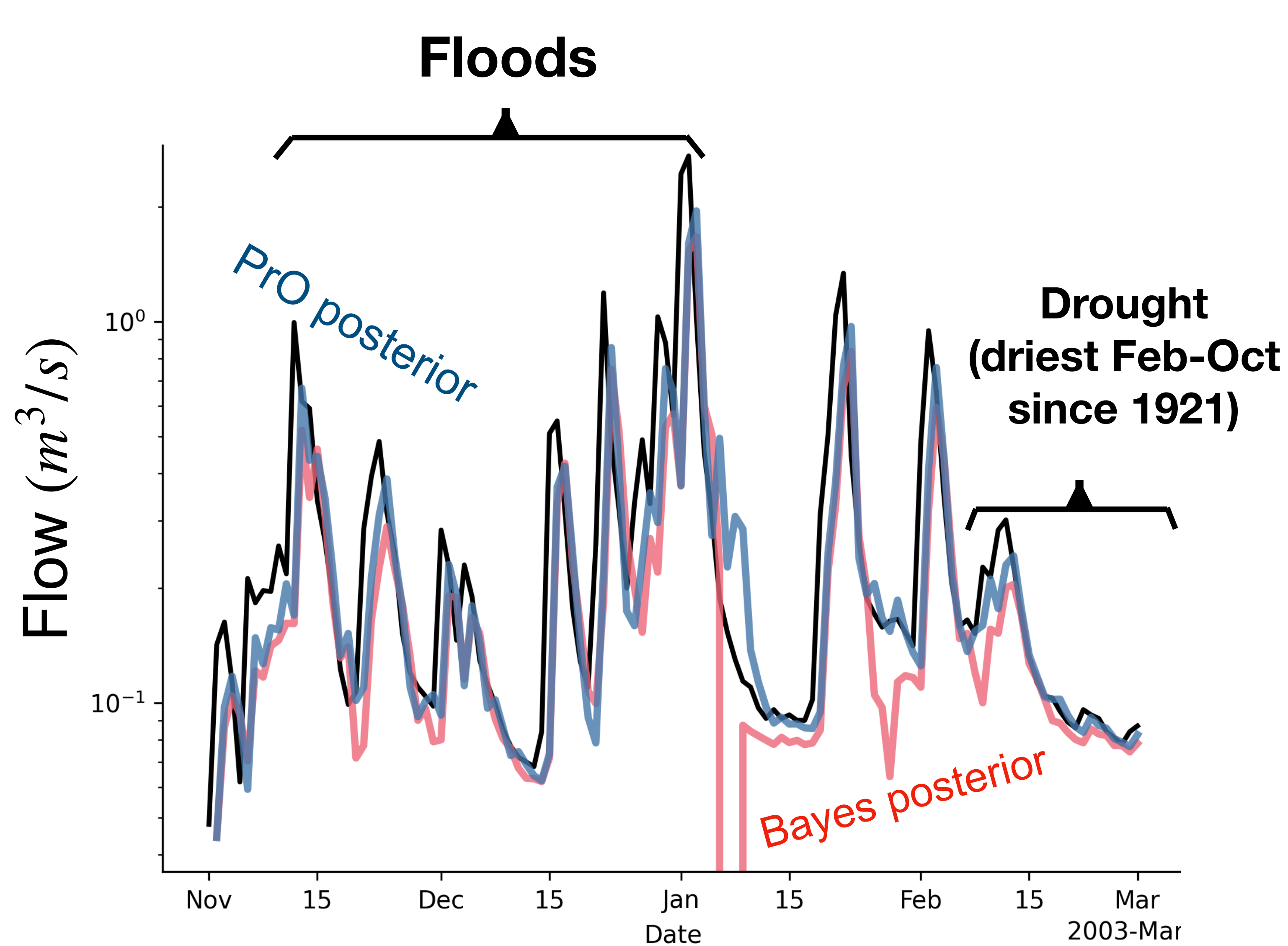
PrO Posteriors: Stansted Brook / River Flow



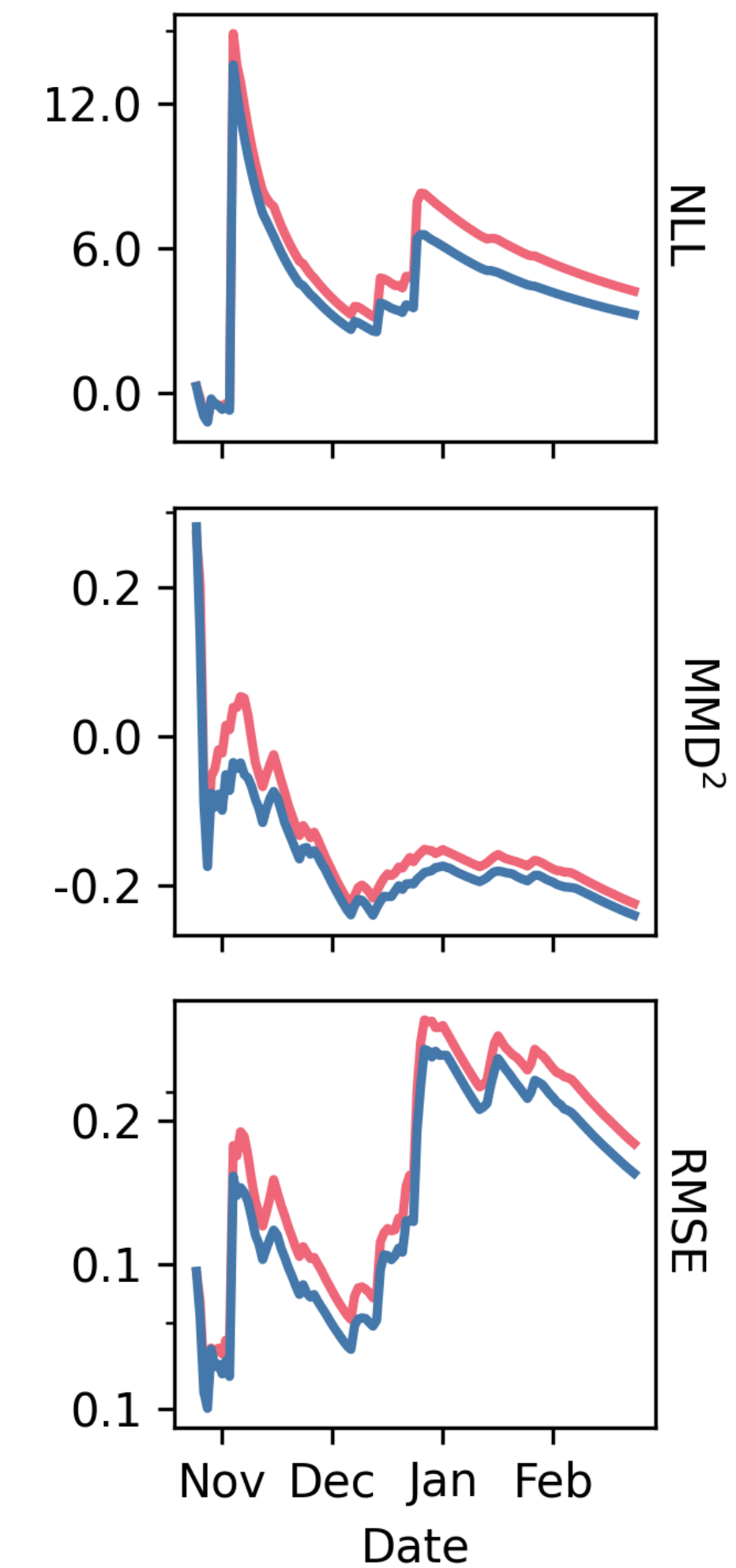
PrO Posteriors: Stansted Brook / River Flow



PrO Posteriors: Stansted Brook / River Flow

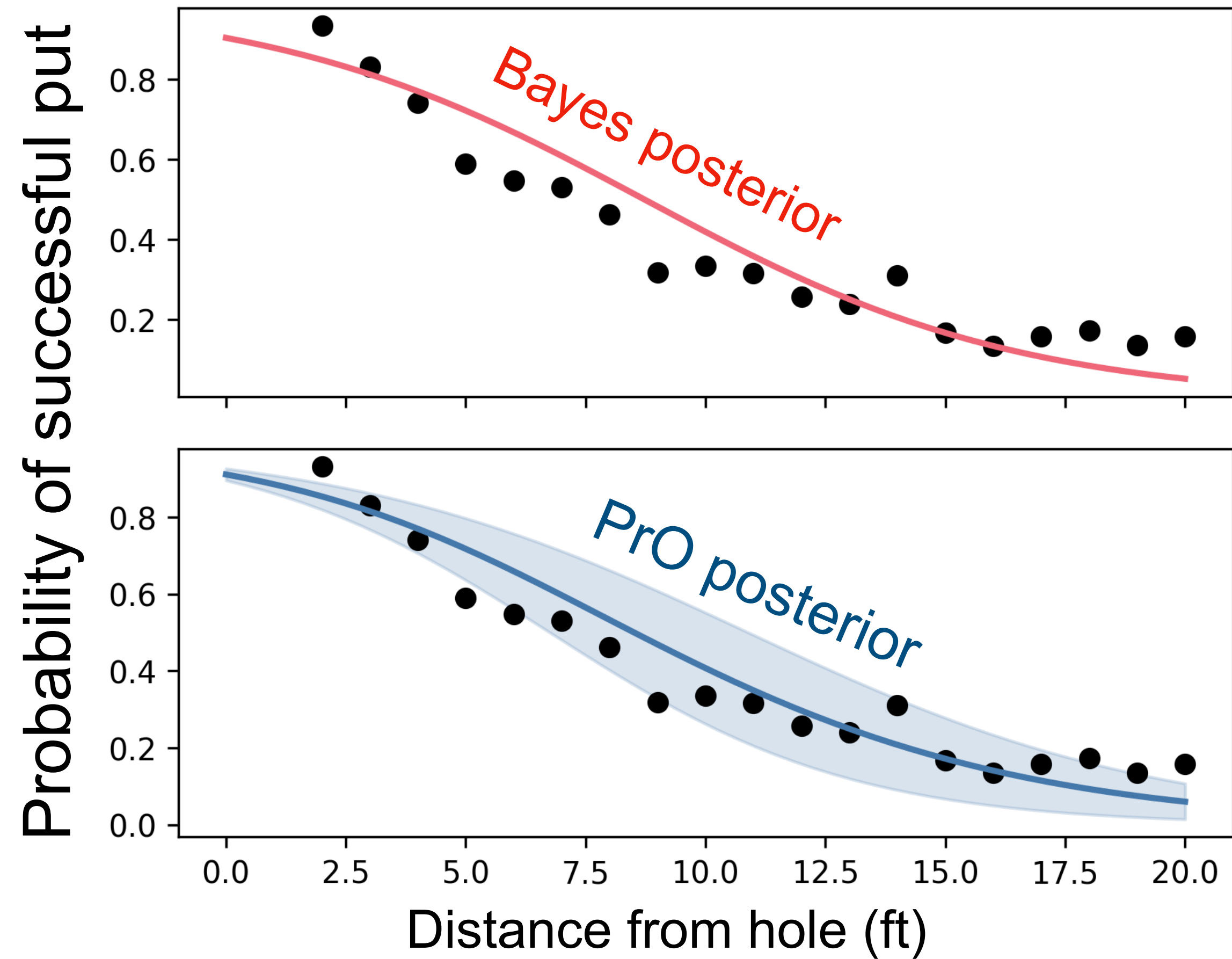


**1-step ahead
forecast errors**



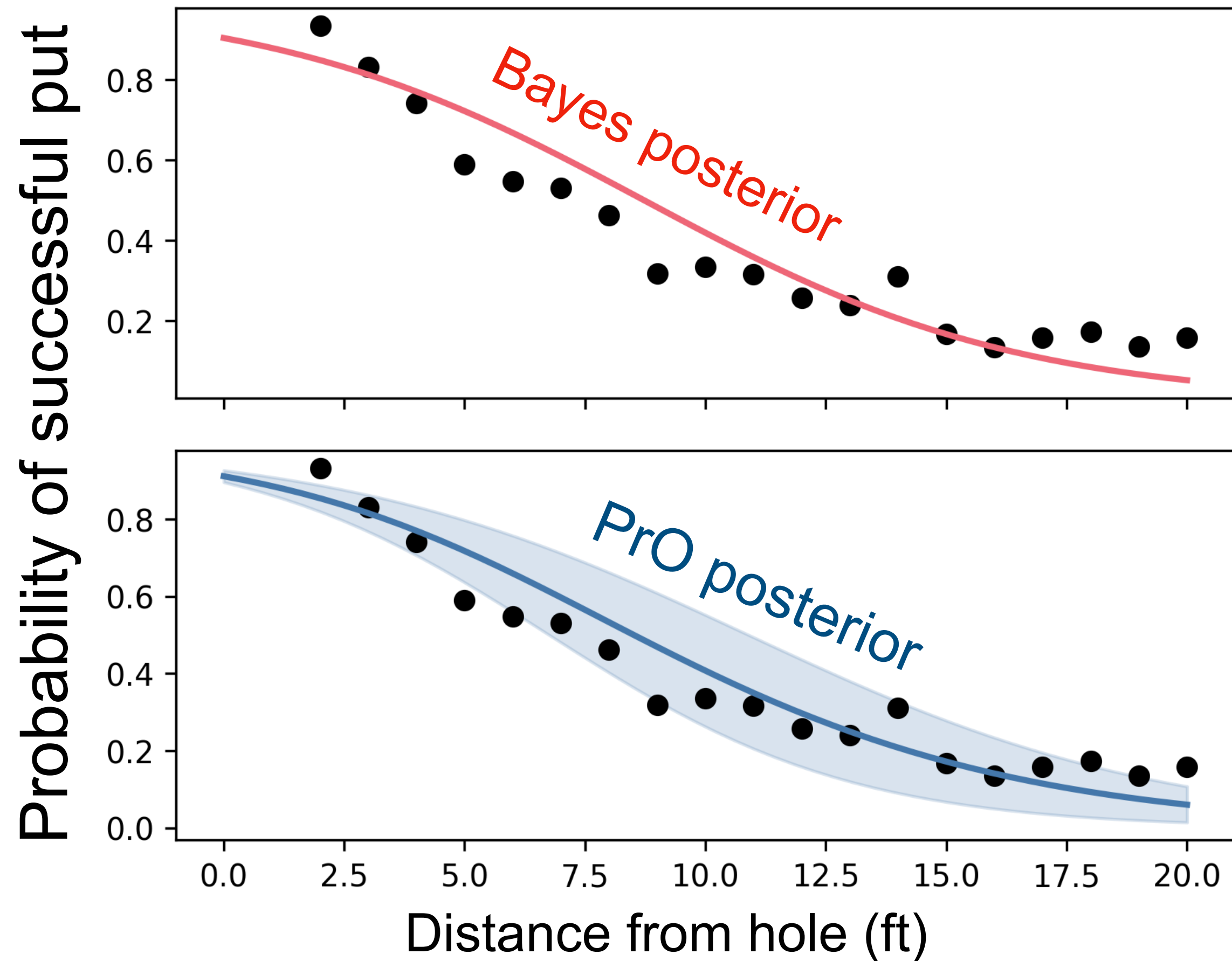
PrO Posteriors: Why/how does it work?

Posterior Predictive

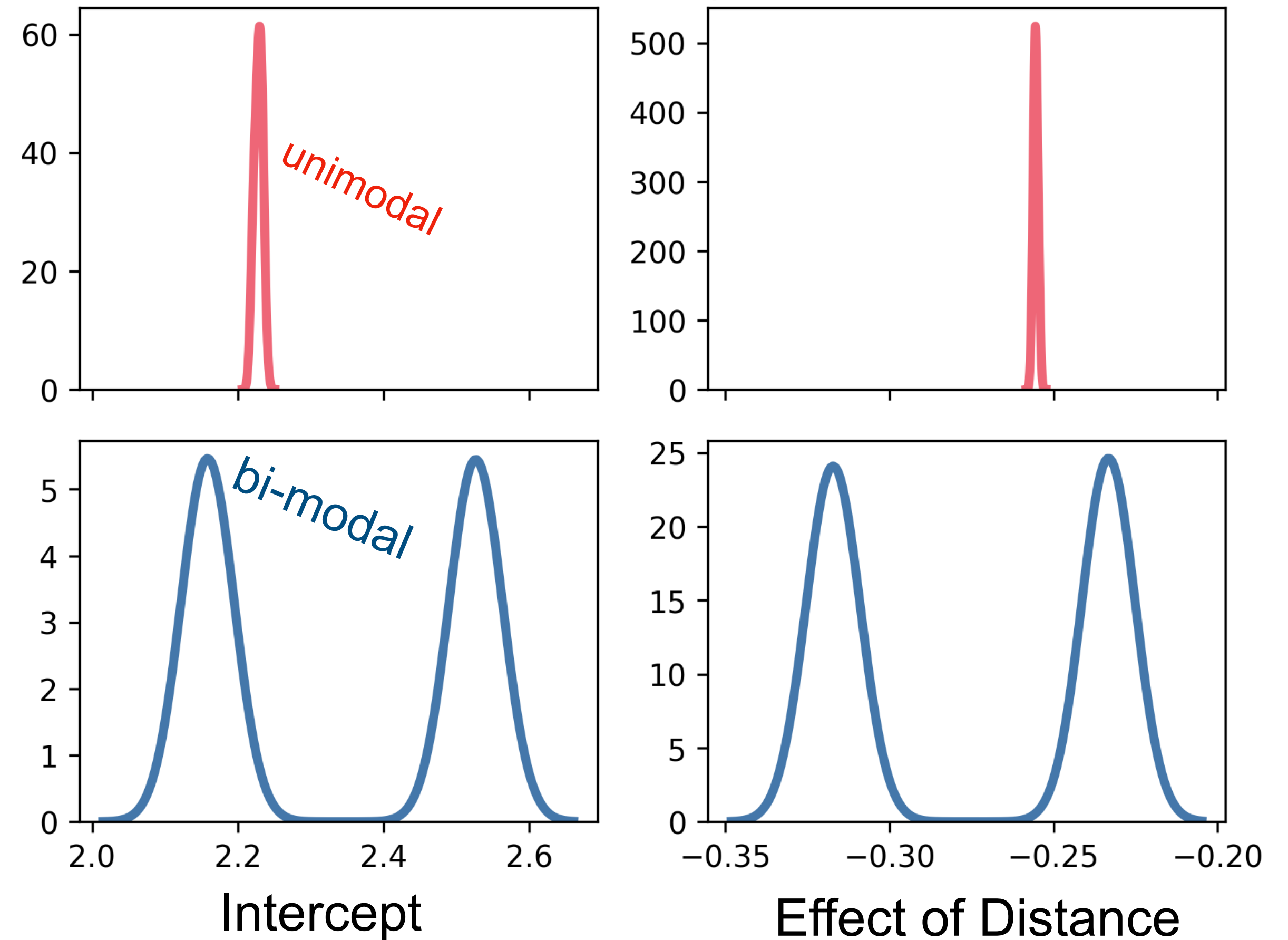


PrO Posteriors: Why/how does it work?

Posterior Predictive

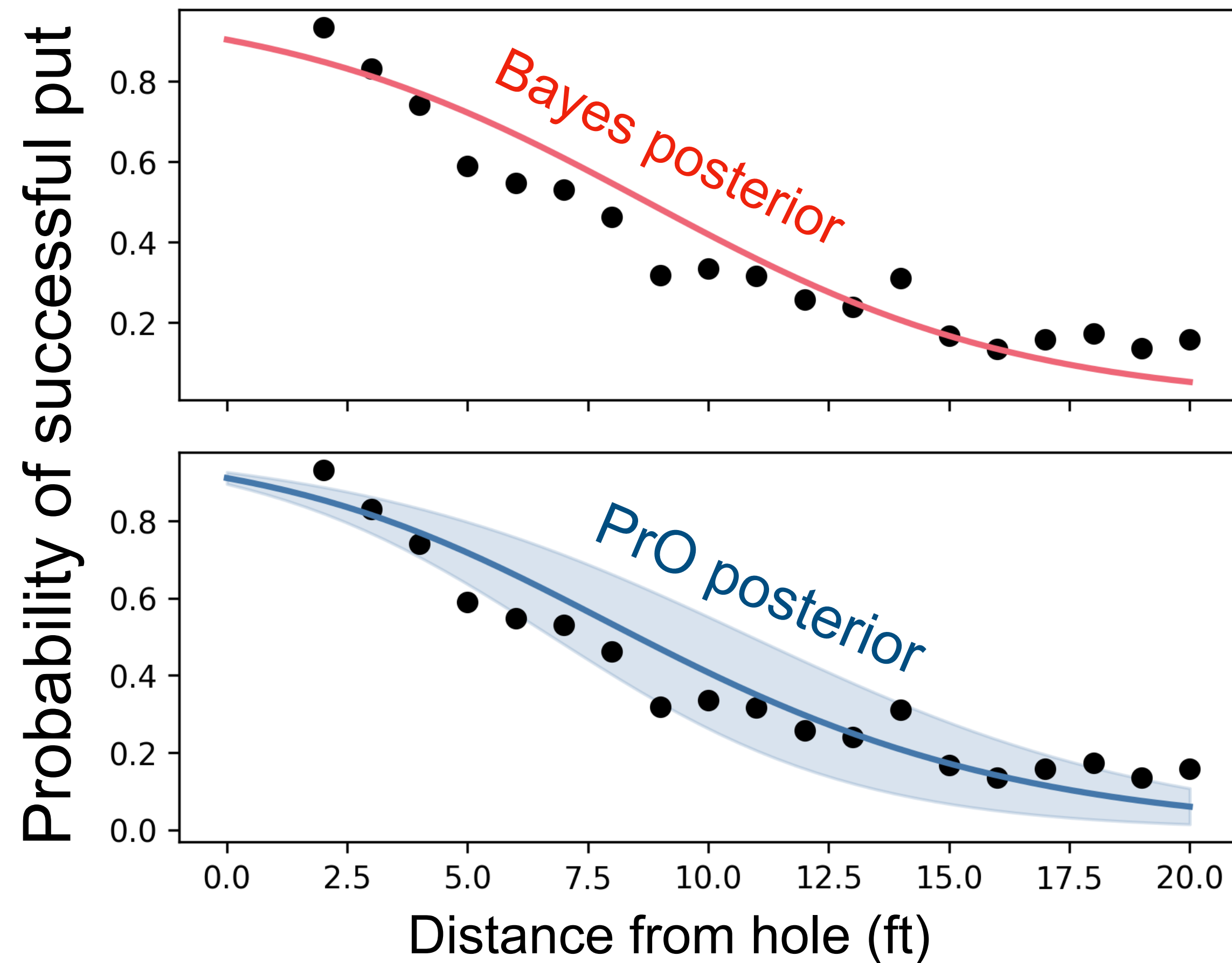


Corresponding Parameter Posterior

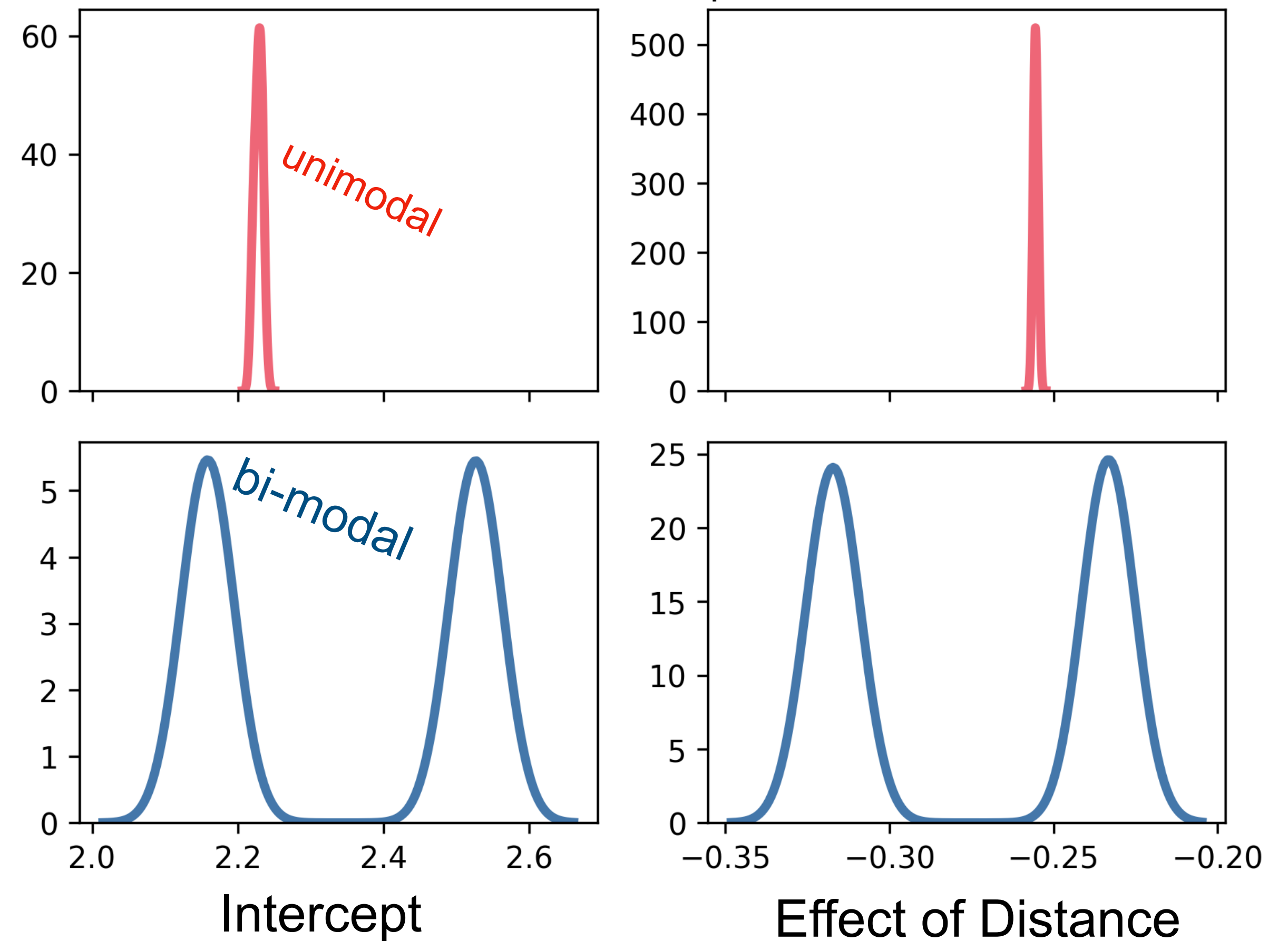


PrO Posteriors: Why/how does it work?

Posterior Predictive



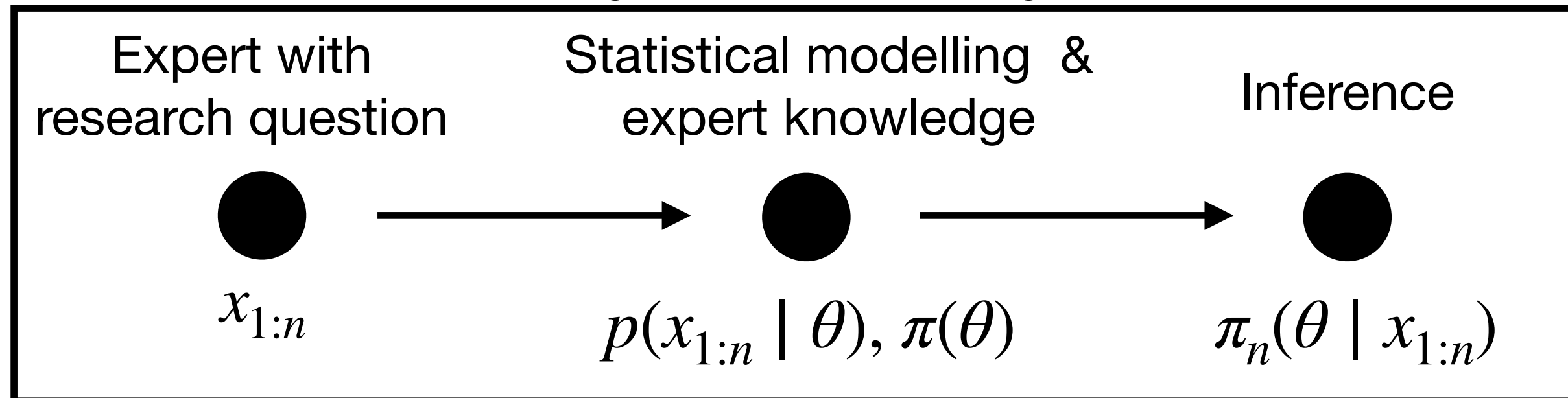
Corresponding Parameter Posterior



⇒ PrO posterior identifies 2 groups of golfers!
(those good at long puts & those not good at it)

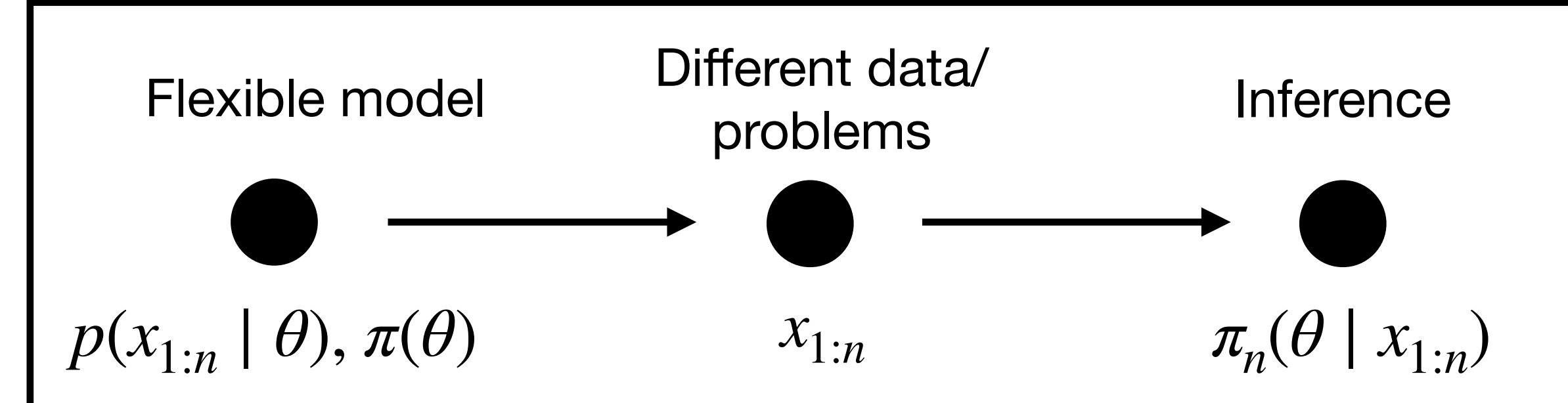
Summary: The issue

Traditional Bayesian analysis in science



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

Modern Bayesian ML



- (A1) model well-specified
 - (A2) prior well-specified
 - (A3) computationally feasible
- A large red stamp with the word "FRAGILE" is overlaid on the list, indicating that these assumptions are fragile in the context of modern Bayesian ML.

Summary: The solutions

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

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

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

Predictively oriented Inference

VI

Gibbs/Generalised/
Pseudo Posterior

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

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

Power/Fractional/
Cold Posterior

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

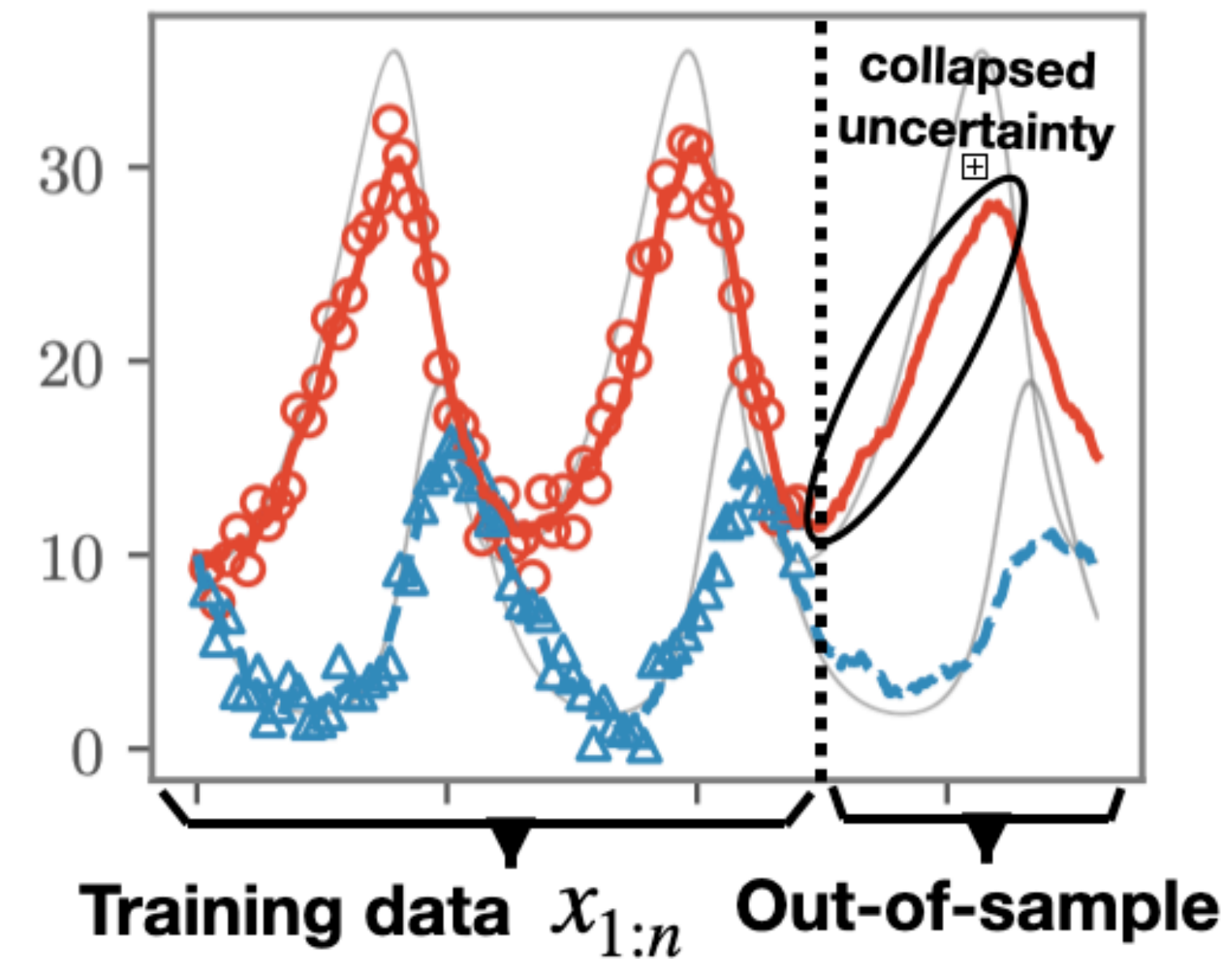
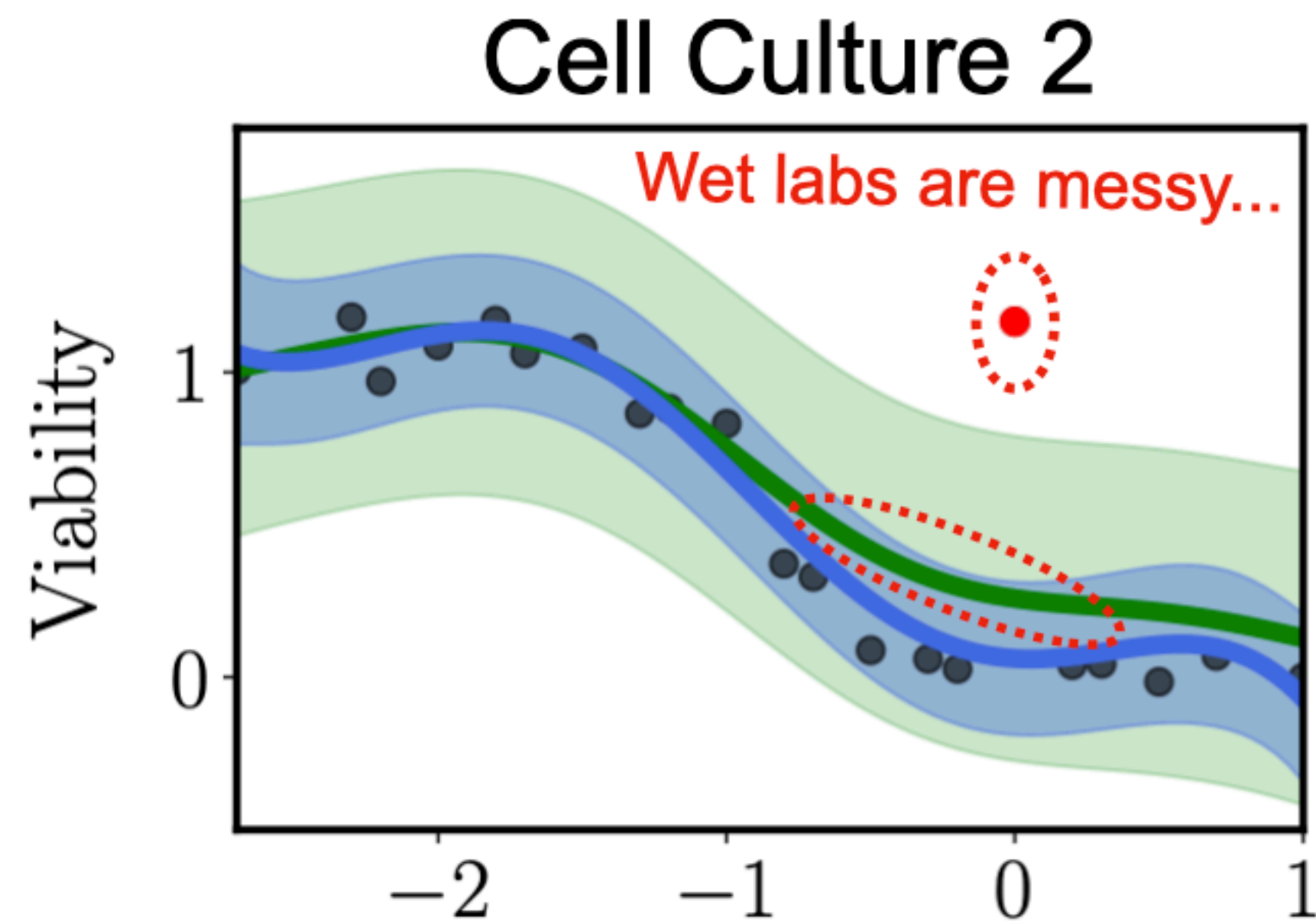
$$\pi_n^{(\lambda)}(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta) d\theta}$$

Parameter Inference

III

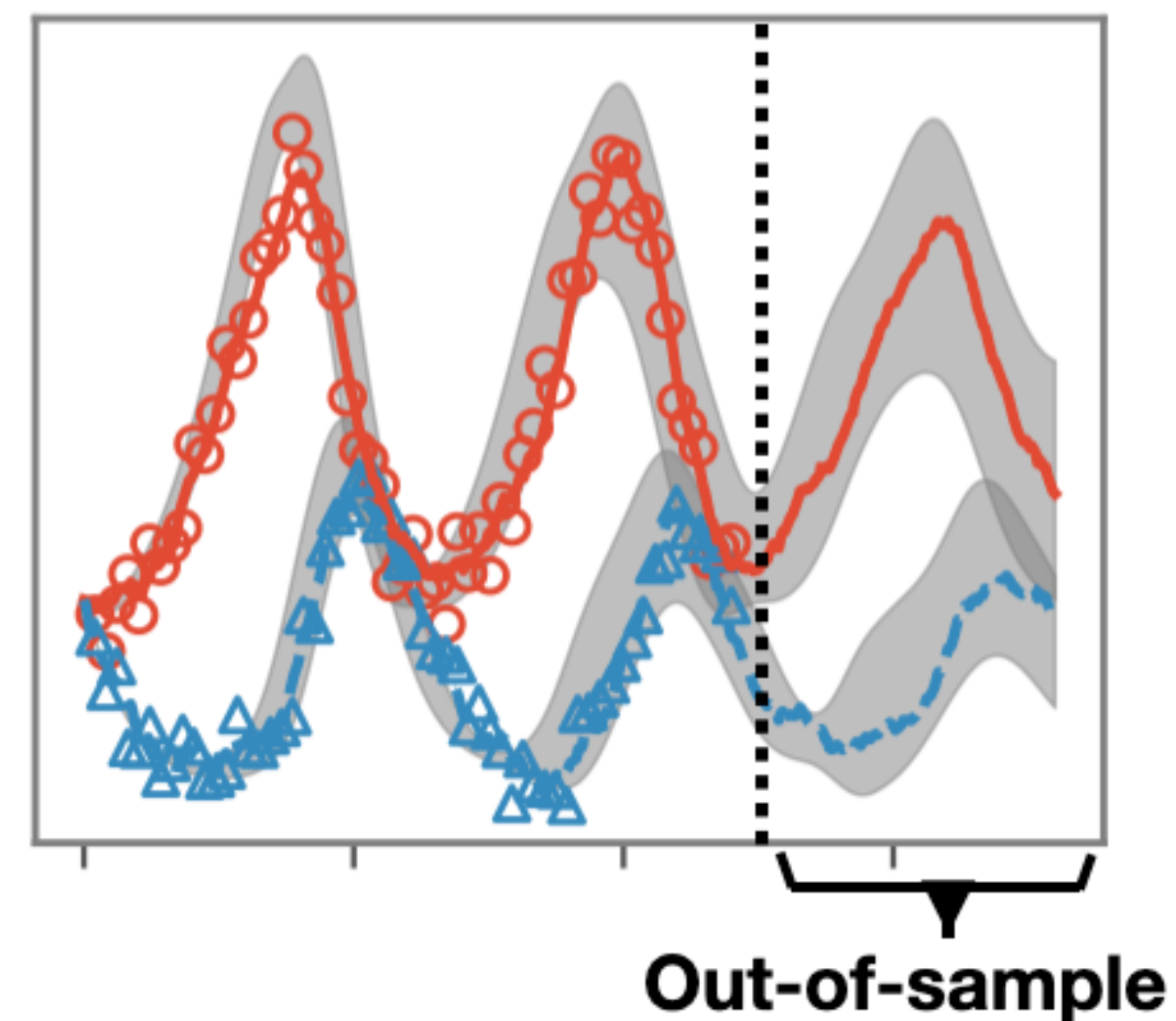
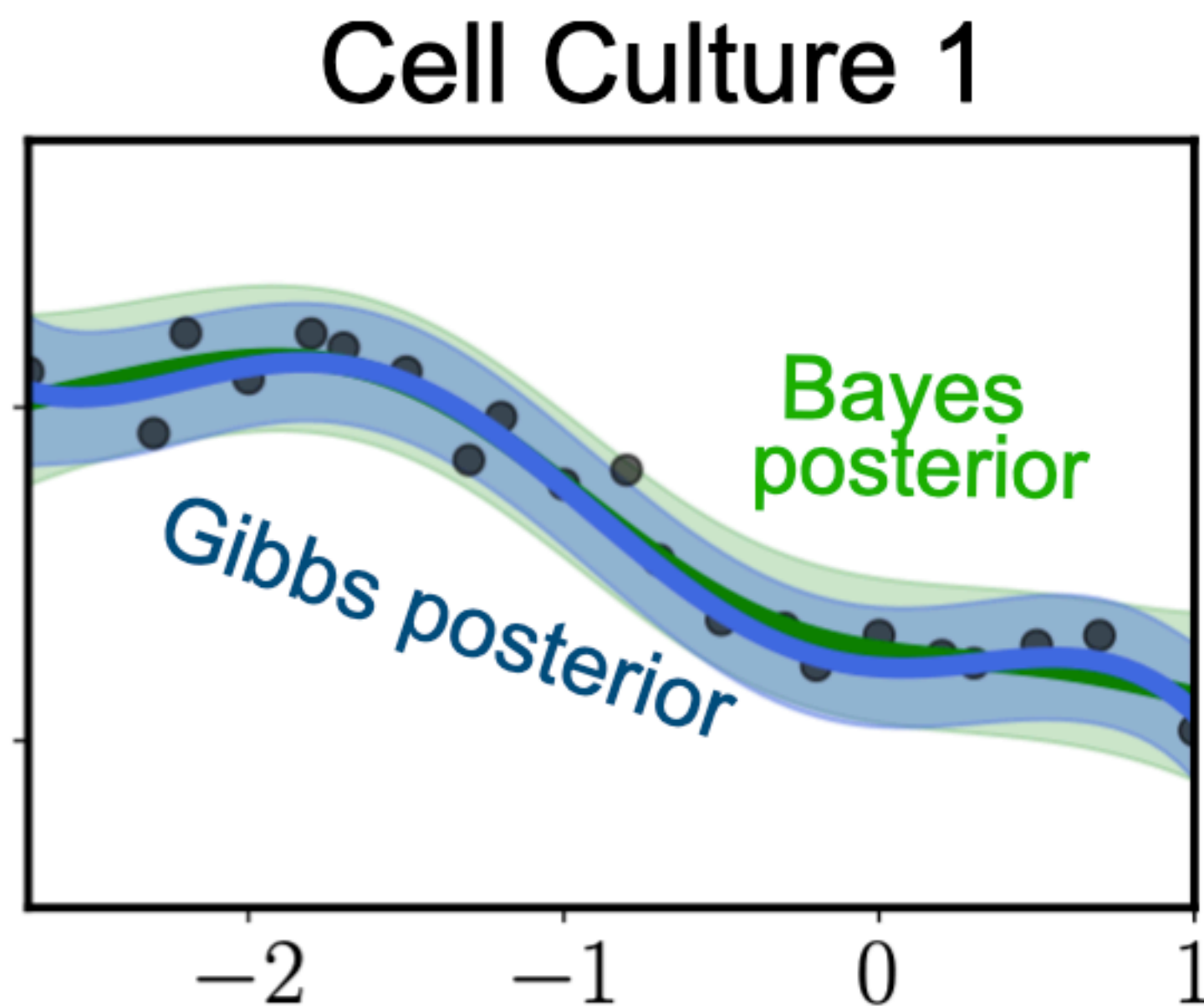
Conclusion

- I. Orthodox Bayes is inappropriate for many modern problems



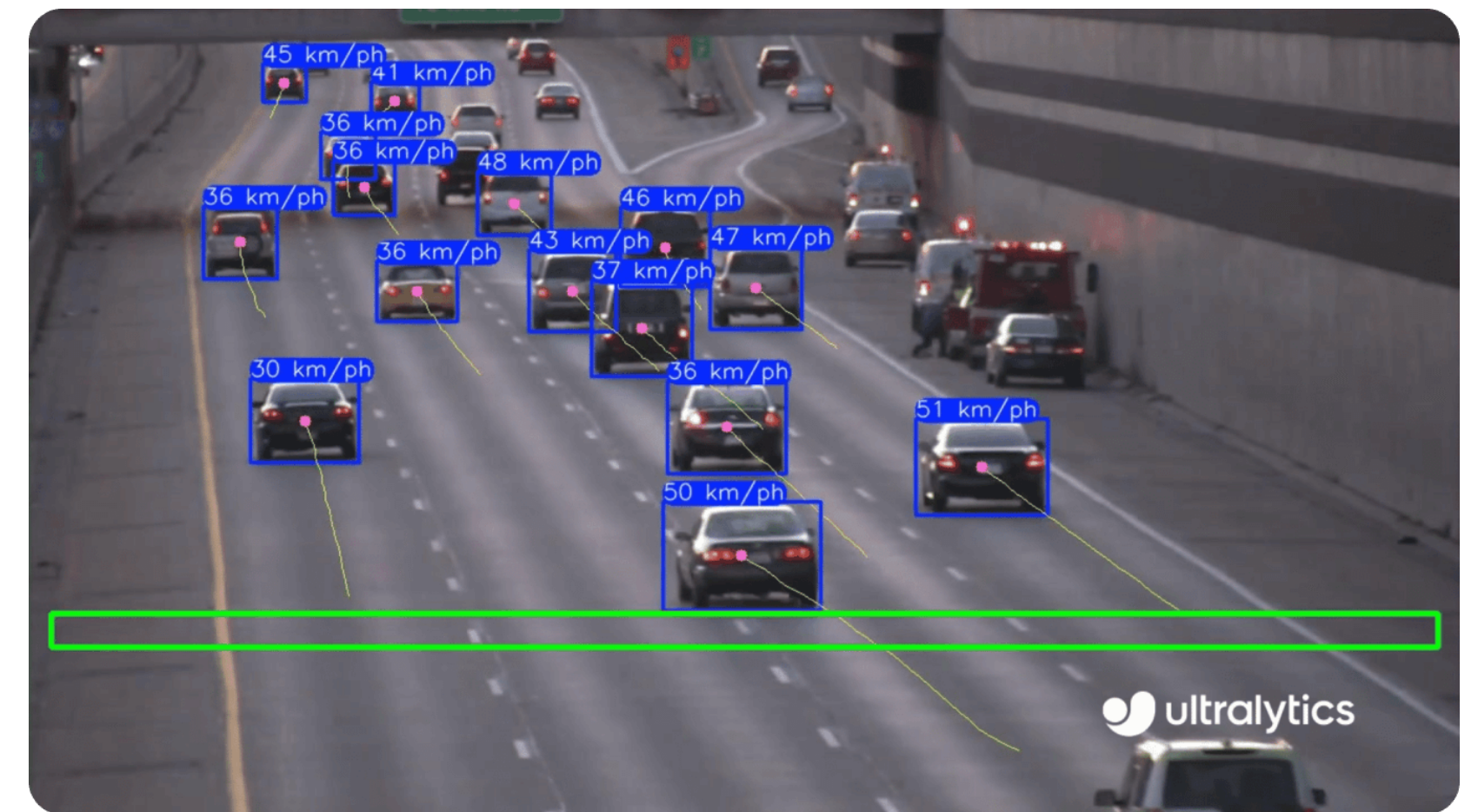
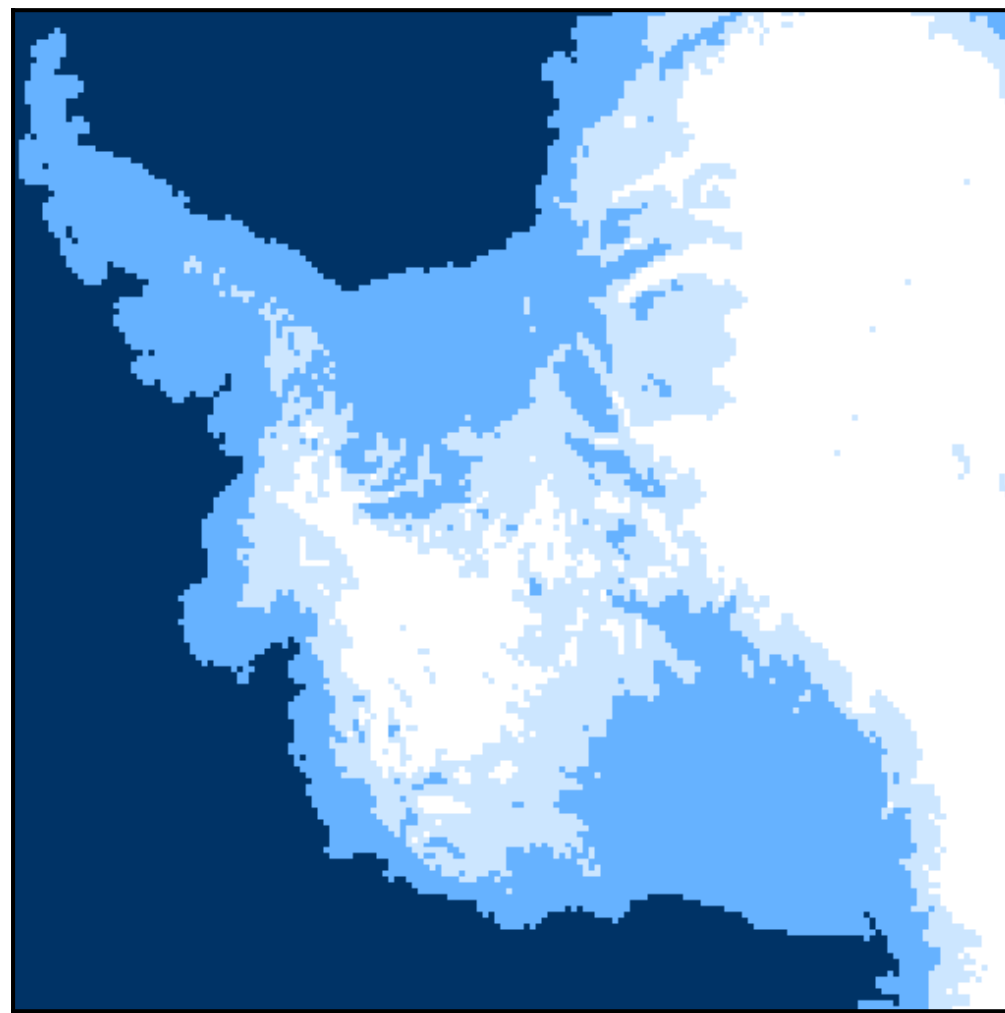
Conclusion

- I. Orthodox Bayes is inappropriate for many modern problems
- II. But we have new ways of fixing this!



Conclusion

- I. Orthodox Bayes is inappropriate for many modern problems
- II. But we have new ways of fixing this!
- III. The theory/methodology has been worked out! It's application time!



Ways to engage: Post-Bayes seminar

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

At a glance/website:

Where to subscribe to mailing list:

Where to subscribe to calendar:

Where to attend the seminars:

Where recorded seminars are stored:

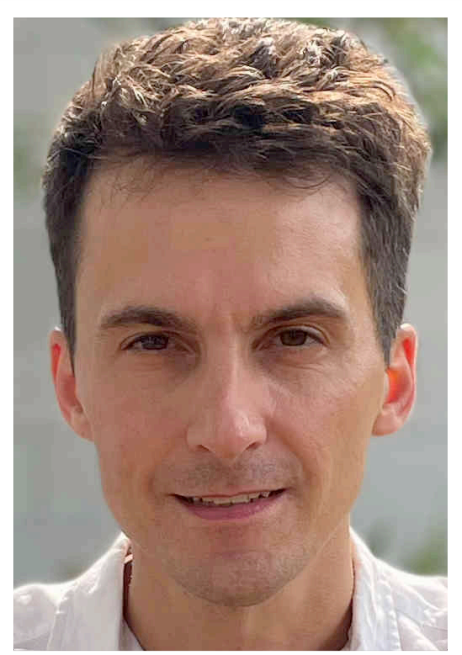
<https://tinyurl.com/postBayesWebsite>

<https://tinyurl.com/postBayesSubscribe>

<https://tinyurl.com/postBayesCalendar>

<https://tinyurl.com/postBayesZoom>

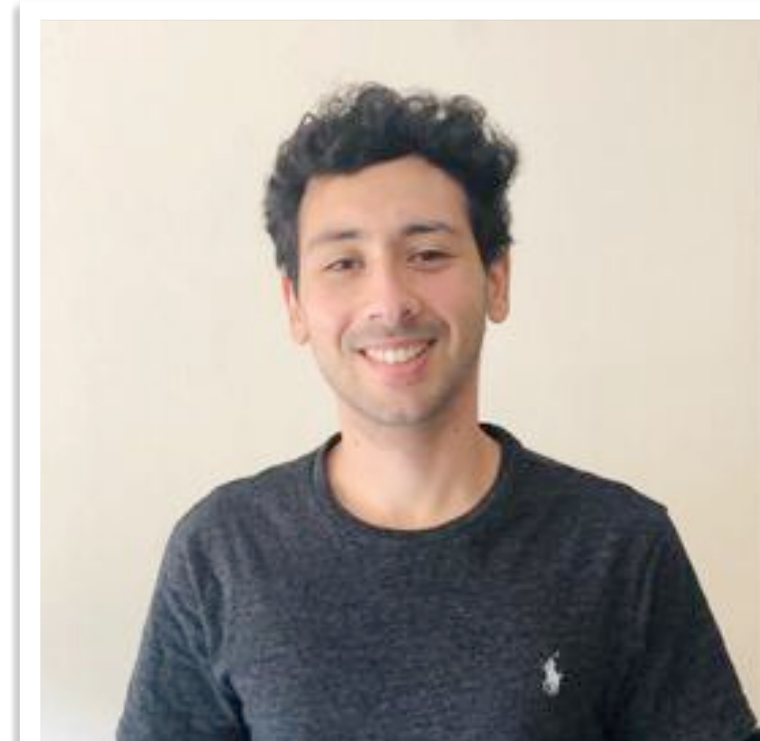
<https://tinyurl.com/postBayesYT>



Prof. Pierre Alquier
(ESSEC Singapore)



Dr. Edwin Fong
(University of
Hong Kong)



Matias Altamirano
(UCL)



Yann McLatchie
(UCL)

Relevant References for today's talk

Power Posteriors:

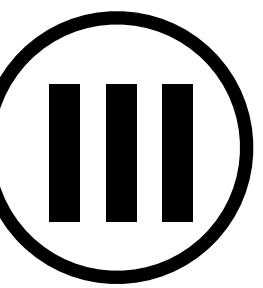
McLatchie, Fong, Frazier, & **K.** (2025); Biometrika
Grunwald & van Ommen (2017); Bayesian Analysis
Grunwald (2012); ALT

Gibbs Posteriors:

Rooijakkers, Ronneberg, Briol, **K.**, & Altamirano (2025); ArXiv preprint 2510.26401
Laplante, Altamirano, **K.**, Duncan, & Briol (2025); ArXiv preprint
Laplante, Altamirano, **K.**, Duncan, & Briol (2025); ICML
Ezzerg & **K.** (2025); ArXiv preprint
Altamirano, Briol, & **K.** (2023); AISTATS
Altamirano, Briol, & **K.** (2024); ICML
Duran-Martin, Altamirano, Shestopaloff, Sanchez-Betancourt, **K.**, Briol, & Murphy (2024); ICML
Matsubara, **K.**, Briol, & Oates (2023); JASA
Matsubara, **K.**, Briol, & Oates (2022); JRSS-B
Husain & **K.** (2022); ALT

PrO Posteriors:

McLatchie, Cherrief-Abdellatif, Frazier, & **K.** (2025); ArXiv preprint 2510.01915
Shen, **K.**, Power, & Oates (2025); AISTATS
Wild, Ghalebikesabi, Sejdinovic, & **K.** (2024); NeurIPS (oral)
K., Jewson, & Damoulas (2022); JMLR
Masegosa (2020); NeurIPS
Jankowiak, Pleiss, & Gardner (2020); ICML; (2020); (UAI)
Morningstar, Alemi, & Dilon (2022); AISTATS



GP Gibbs posteriors: Twitter Flash Crash

AP Twitter hack causes panic on Wall Street and sends Dow plunging

Market recovers after hackers tweeted from the official AP feed that two explosions had hit the White House

Heidi Moore in
New York and
Dan Roberts in
Washington

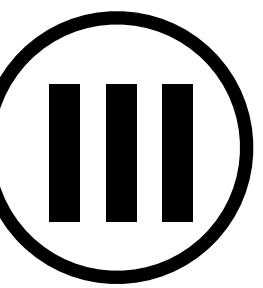
Tue 23 Apr 2013 20:41
BST

 **Share**

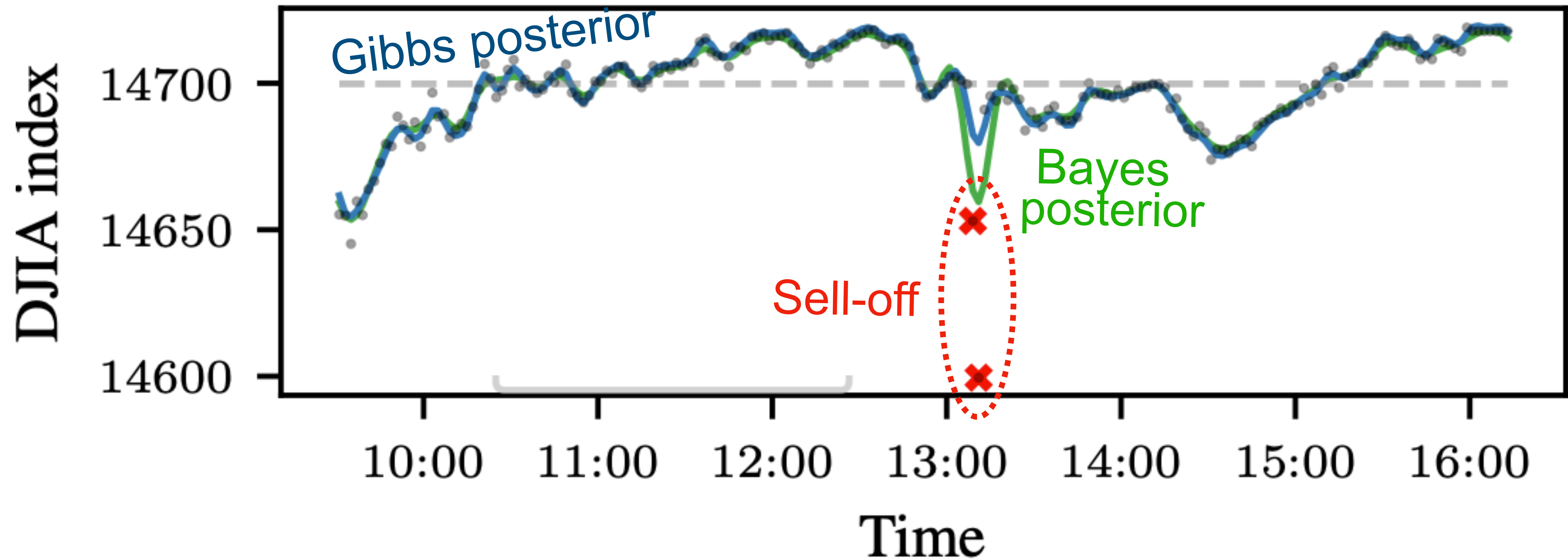
 **61**



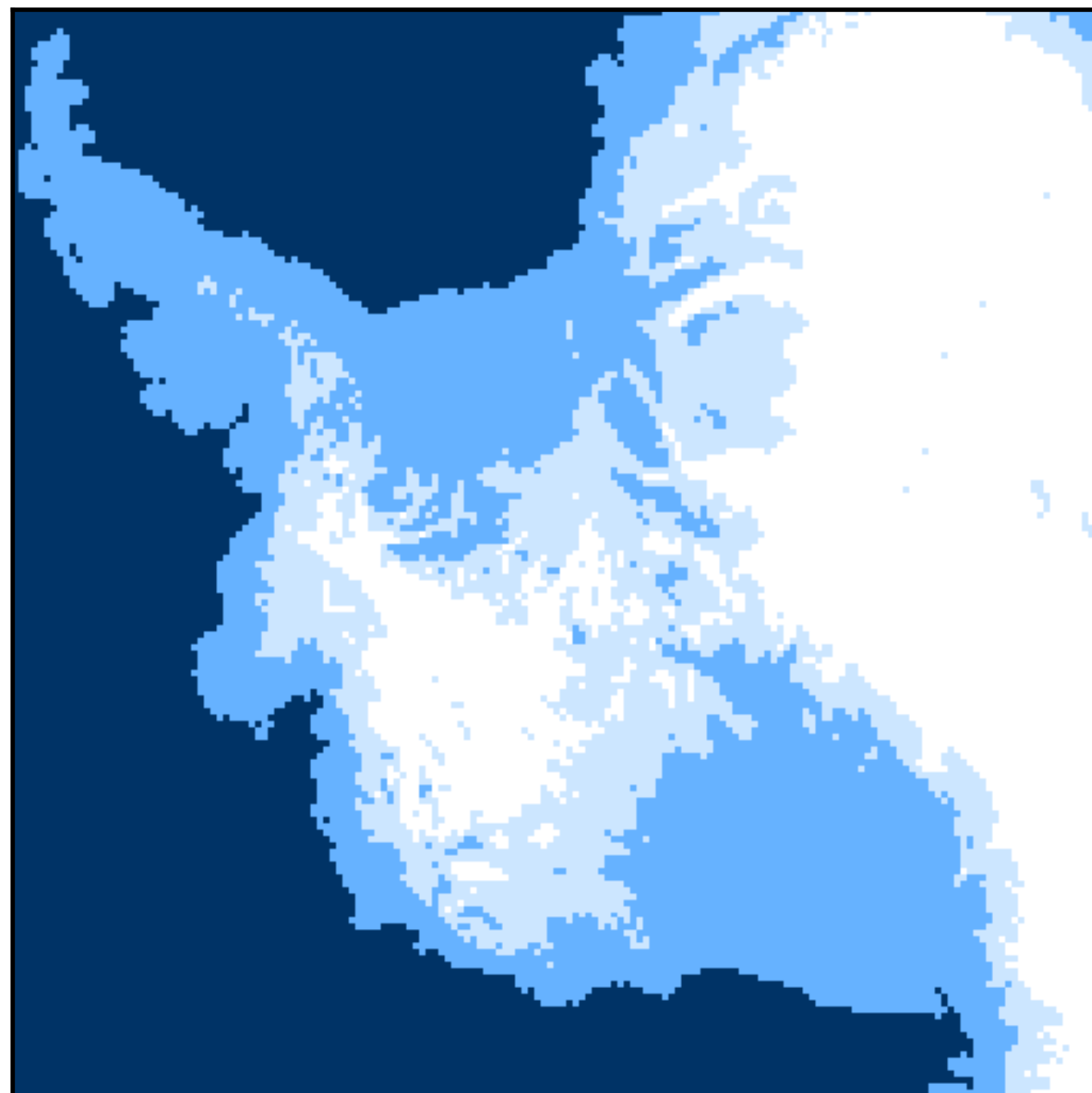
 The panic, however brief, demonstrates how tightly intertwined Wall Street has become with Twitter. Photograph: Spencer Platt/Getty Images



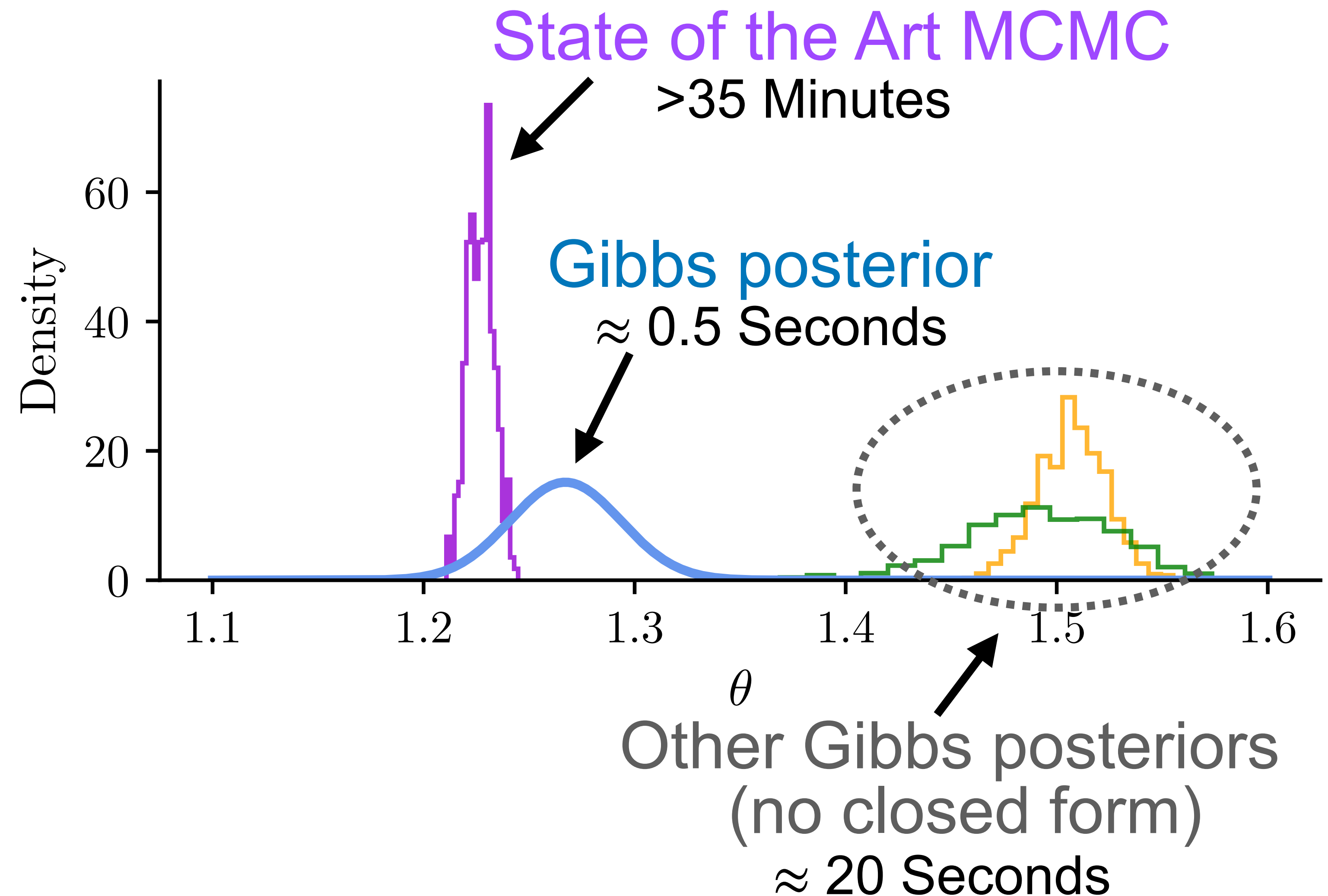
GP Gibbs posteriors: Twitter Flash Crash

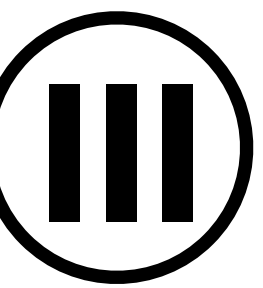


Gibbs posteriors: Ising/Potts Models

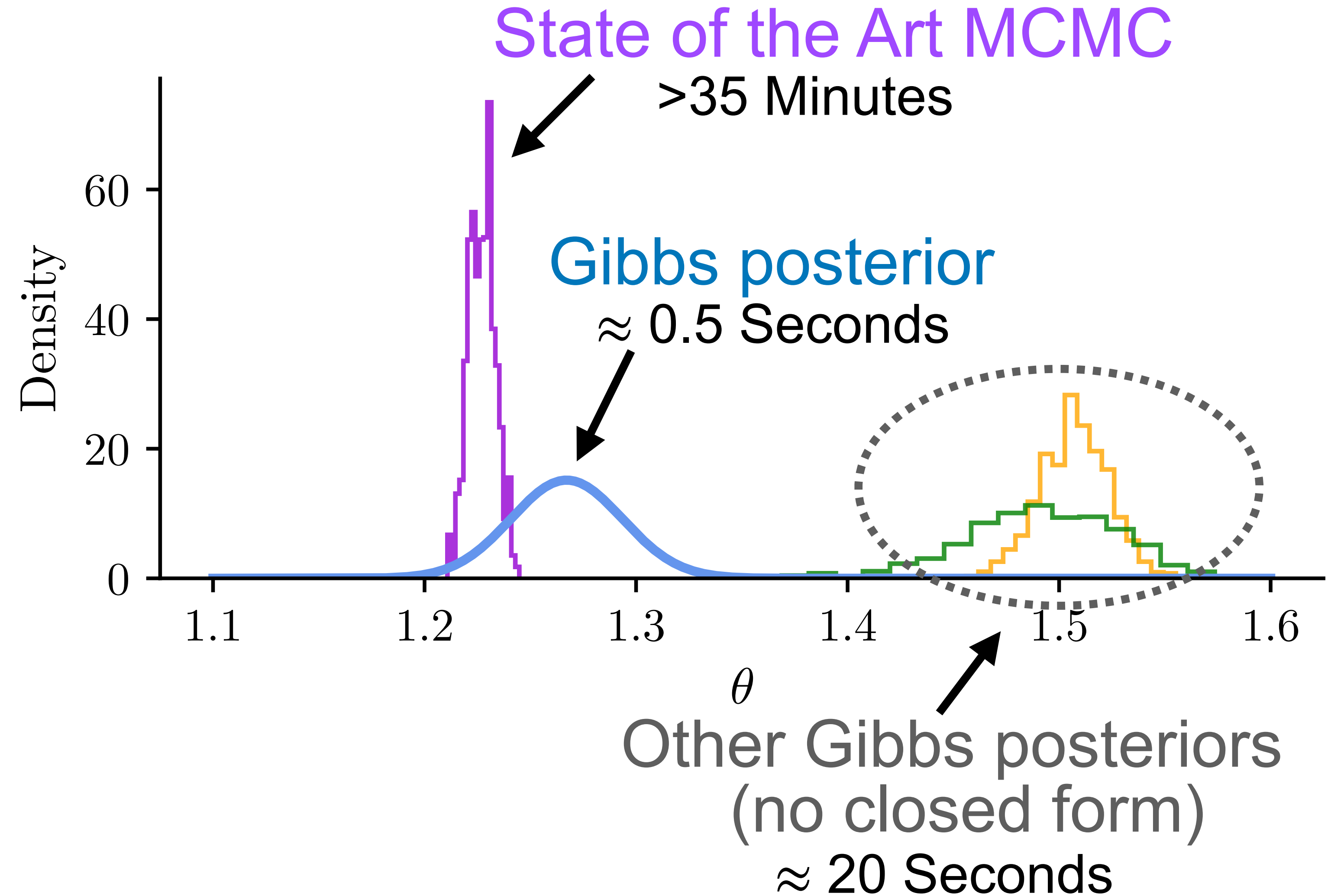
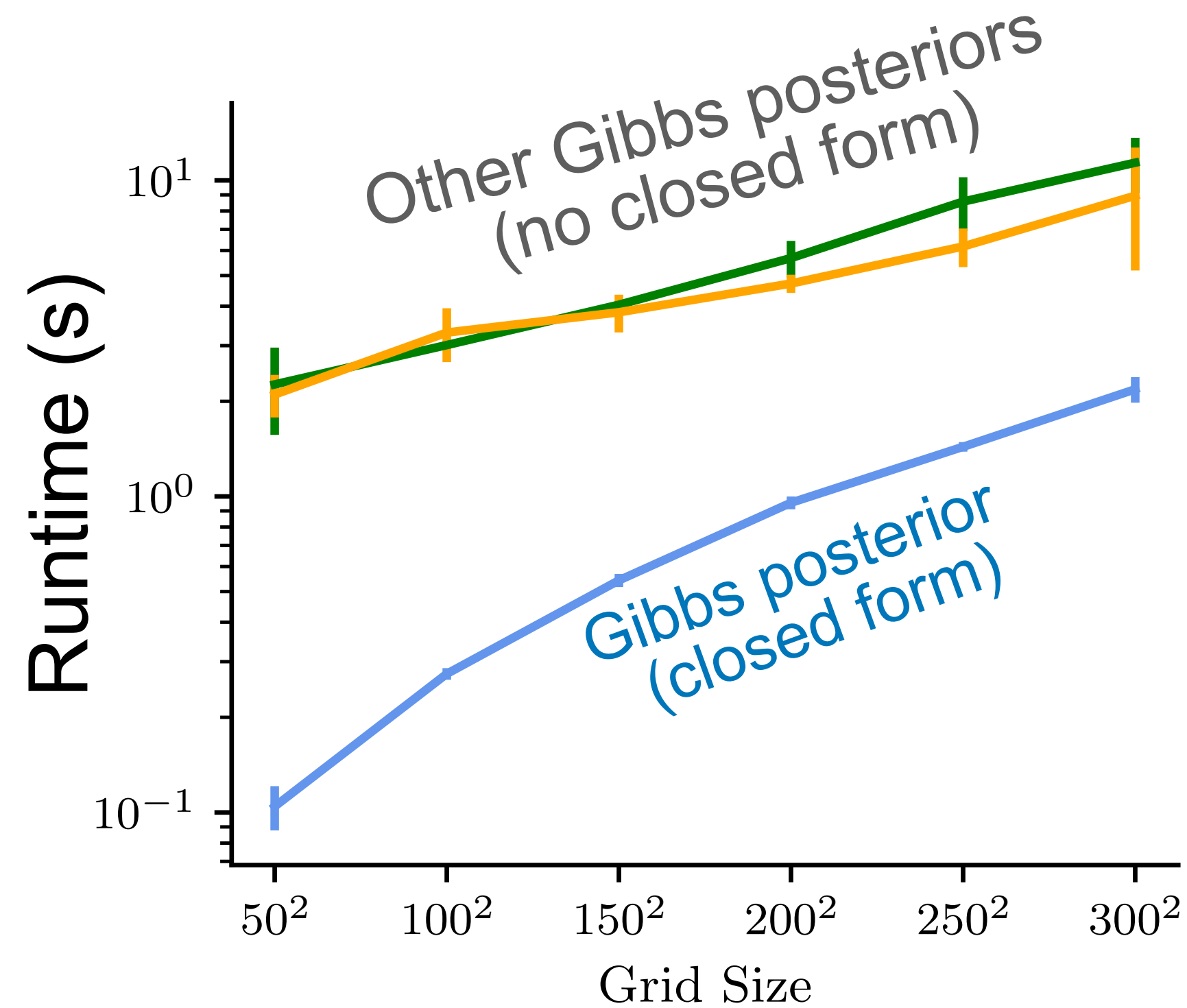


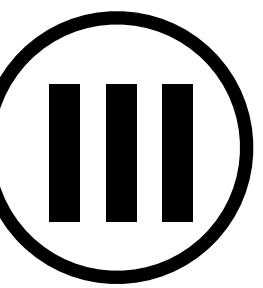
Potts Model
(for Ice Sheet thickness)
Grid size: 171 x 171



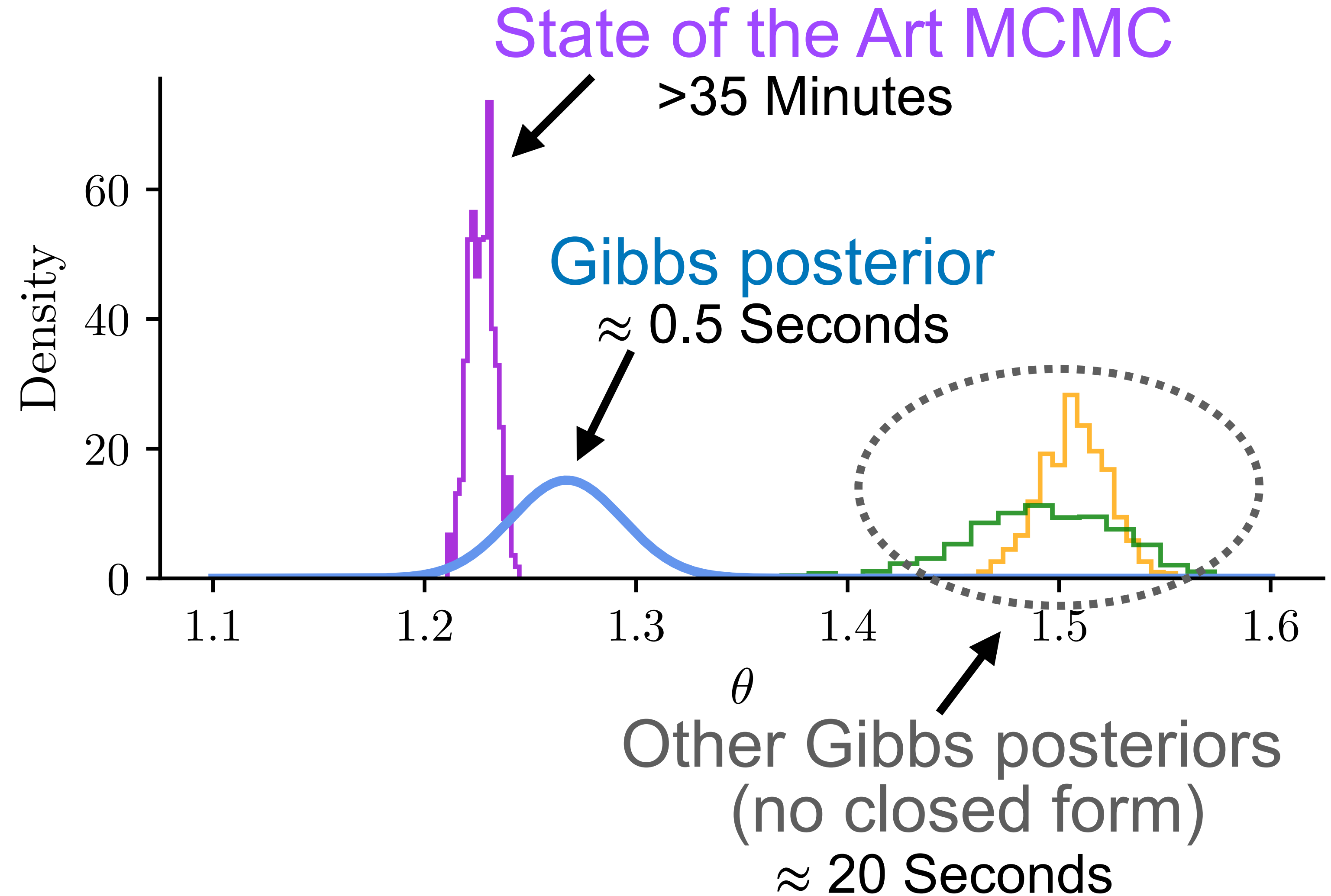
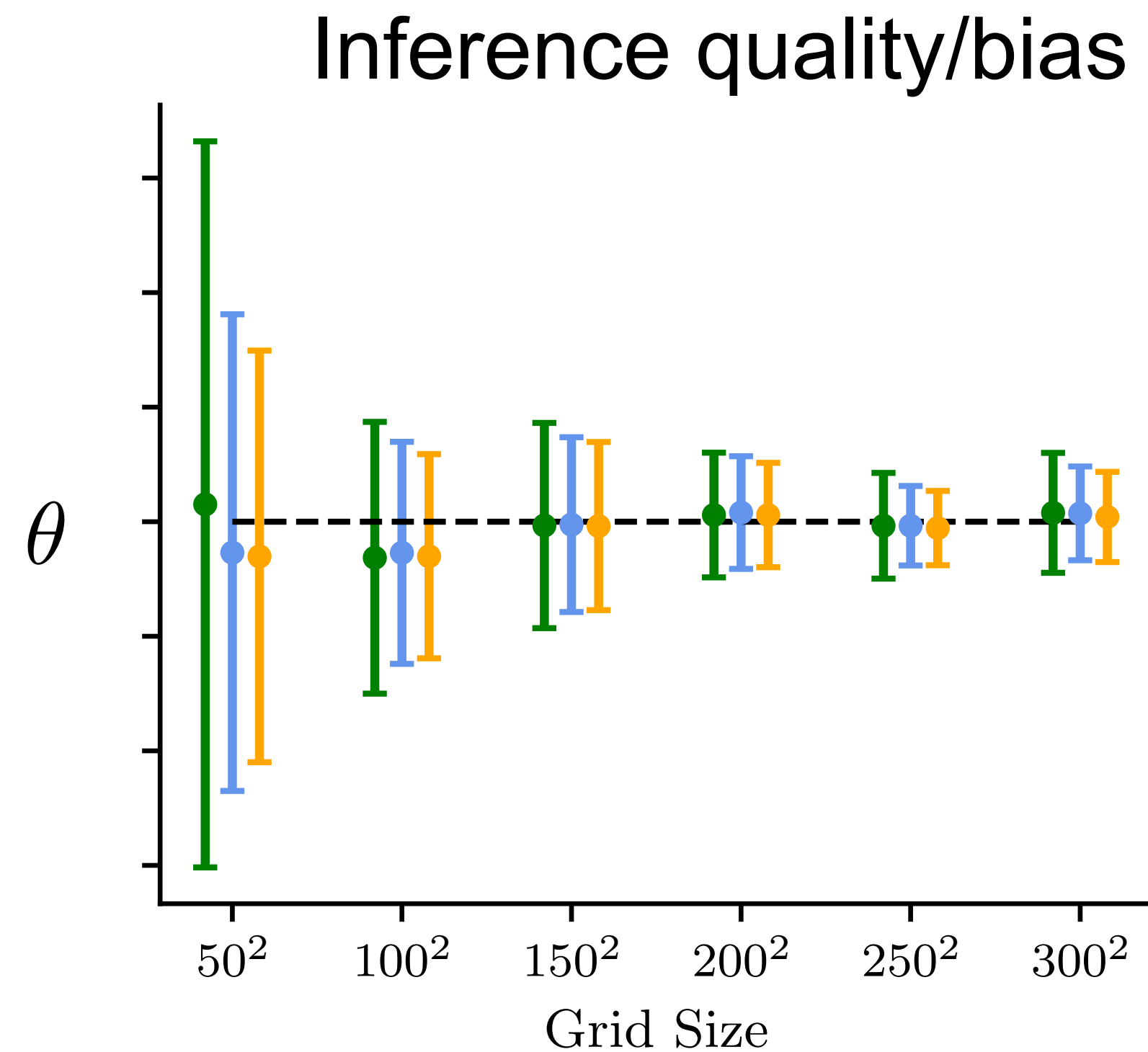


Gibbs posteriors: Ising/Potts Models





Gibbs posteriors: Ising/Potts Models



Optimisation-centric posteriors

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

$$\pi_n^{\mathbf{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta} = \arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int \mathbf{L}(x_{1:n}, p_\theta) q(\theta) d\theta + \text{KL}(q, \pi) \right\}$$

Optimisation-centric posteriors

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

$$\pi_n^{\mathbf{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta} = \arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int \mathbf{L}(x_{1:n}, p_\theta) q(\theta) d\theta + \text{KL}(q, \pi) \right\}$$

~~(A1)~~ by using
robust loss $\mathbf{L}$

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

Optimisation-centric posteriors

Optimisation-centric posteriors

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

$$\pi_n^{\mathcal{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathcal{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathcal{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta} = \arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int \mathcal{L}(x_{1:n}, p_\theta) q(\theta) d\theta + \text{KL}(q, \pi) \right\}$$

$\downarrow$ ~~(A1)~~ by using robust loss $\mathcal{L}$ $\downarrow$ ~~(A2)~~ by using robust regulariser $\mathcal{D}$

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

Optimisation-centric posteriors

Optimisation-centric posteriors

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

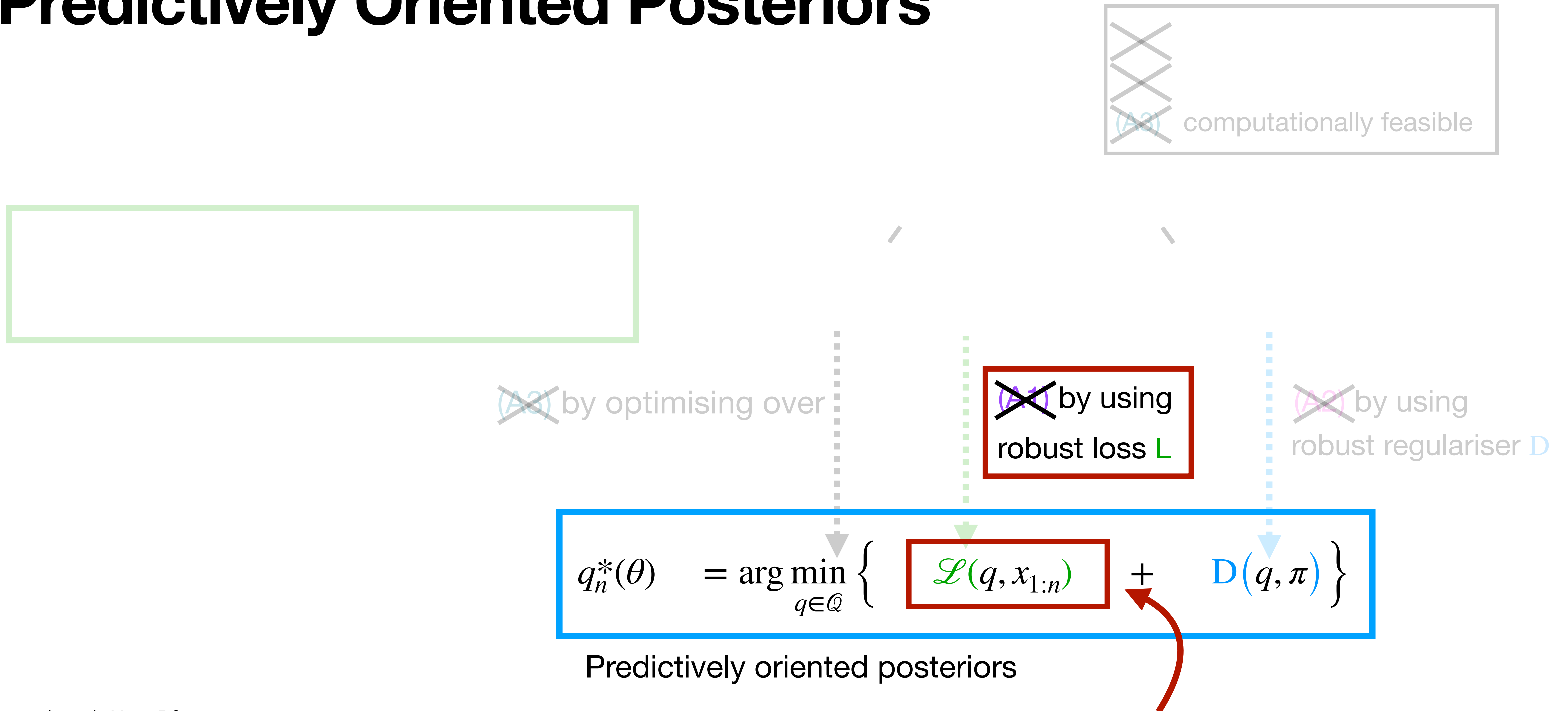
$$\pi_n^{\mathbf{L}}(\theta \mid x_{1:n}) = \frac{\exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta)}{\int \exp\{-\mathbf{L}(x_{1:n}, p_\theta)\} \cdot \pi(\theta) d\theta} = \arg \min_{q \in \mathcal{P}(\Theta)} \left\{ \int \mathbf{L}(x_{1:n}, p_\theta) q(\theta) d\theta + \text{KL}(q, \pi) \right\}$$

~~(A3)~~ by optimising over a set $\mathcal{Q} \subseteq \mathcal{P}(\Theta)$
~~(A1)~~ by using robust loss $\mathbf{L}$
~~(A2)~~ by using robust regulariser $\mathbf{D}$

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

Optimisation-centric posteriors

Predictively Oriented Posteriors



Power Posteriors

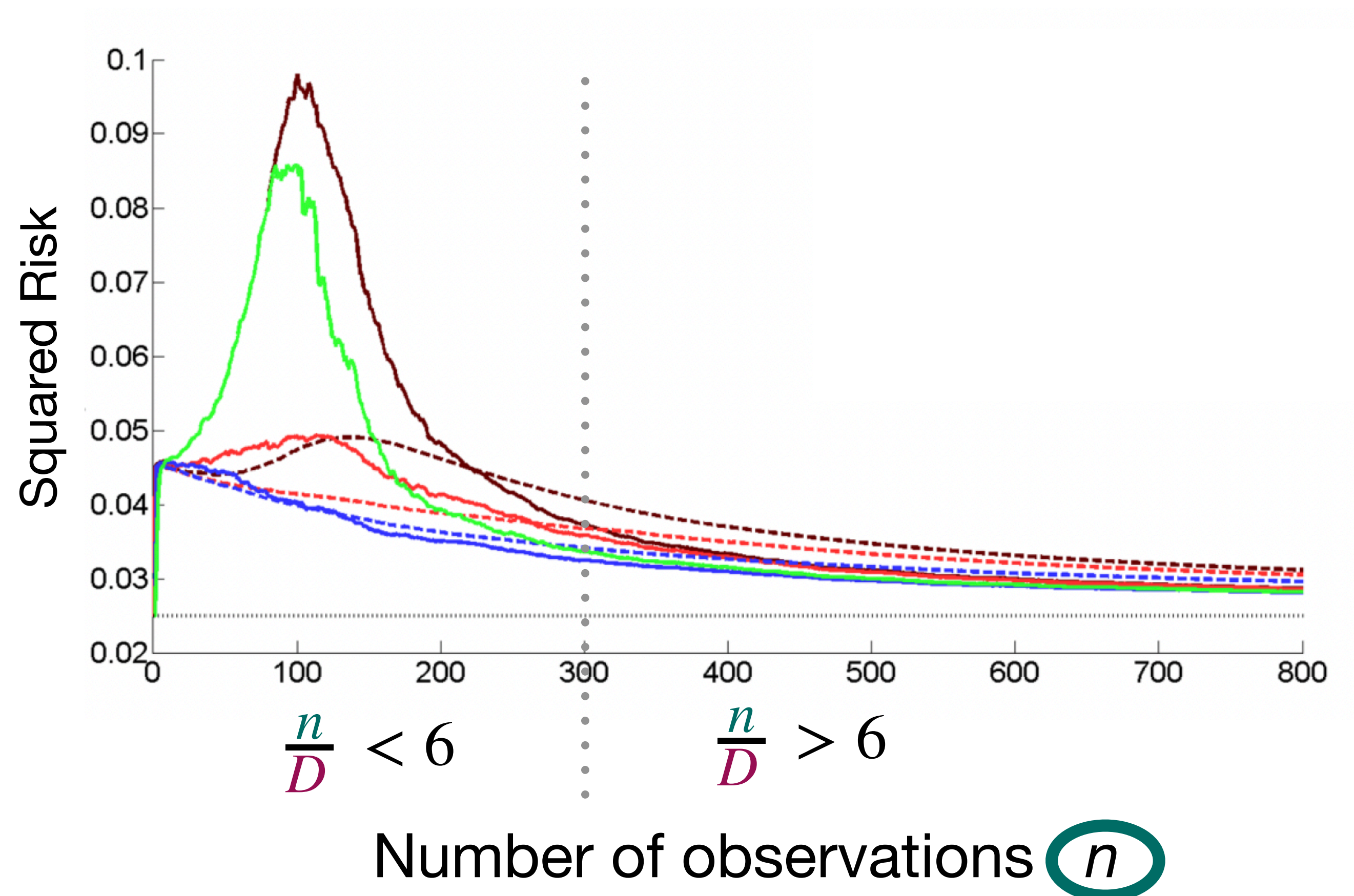
Power/Fractional/Cold Posteriors

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

Q: Why do it do?

Regression model (misspecified):

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



Power Posteriors

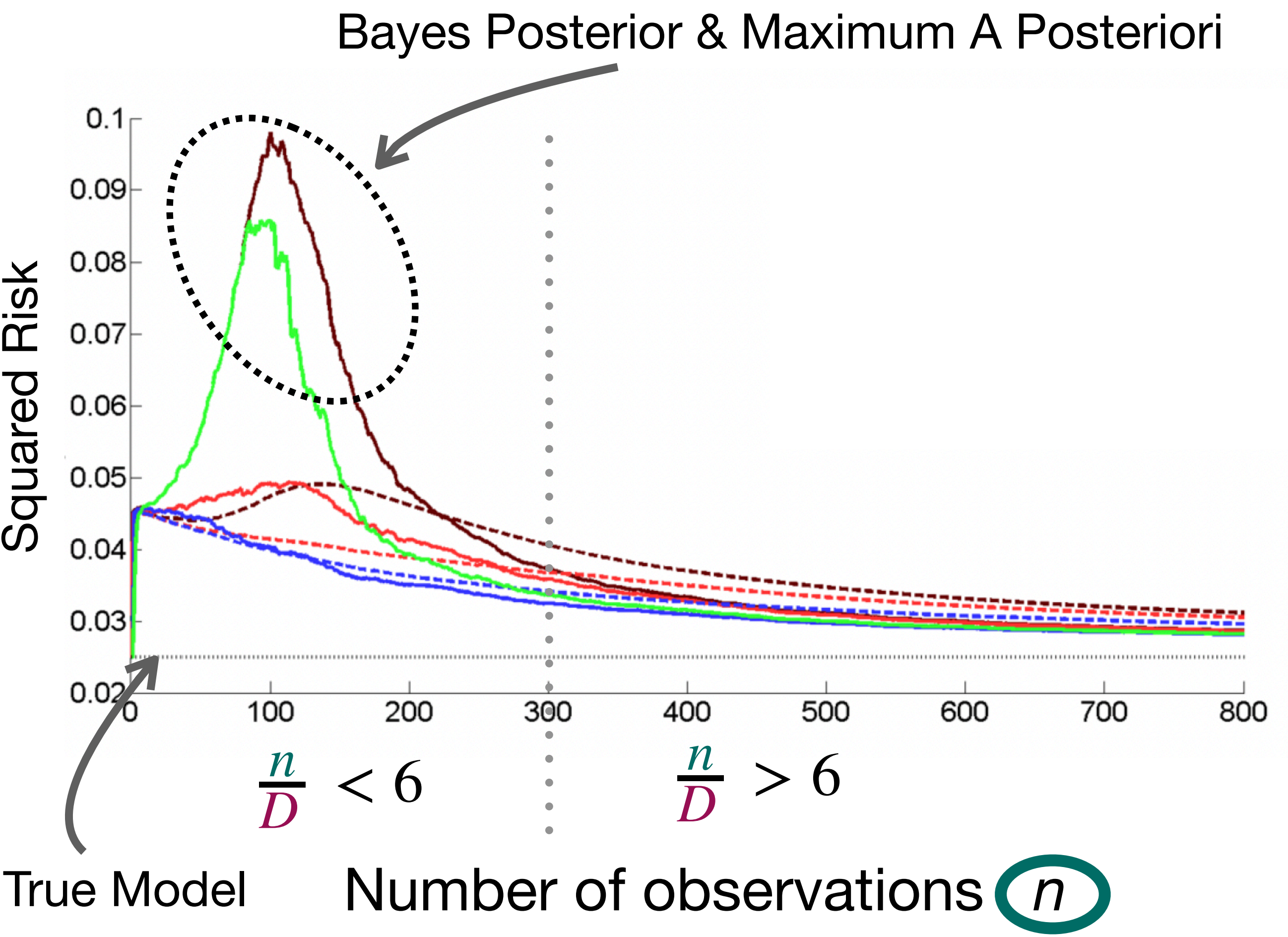
Power/Fractional/Cold Posteriors

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

Q: Why do it do?

Regression model (misspecified):

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



Power Posteriors

Power/Fractional/Cold Posteriors

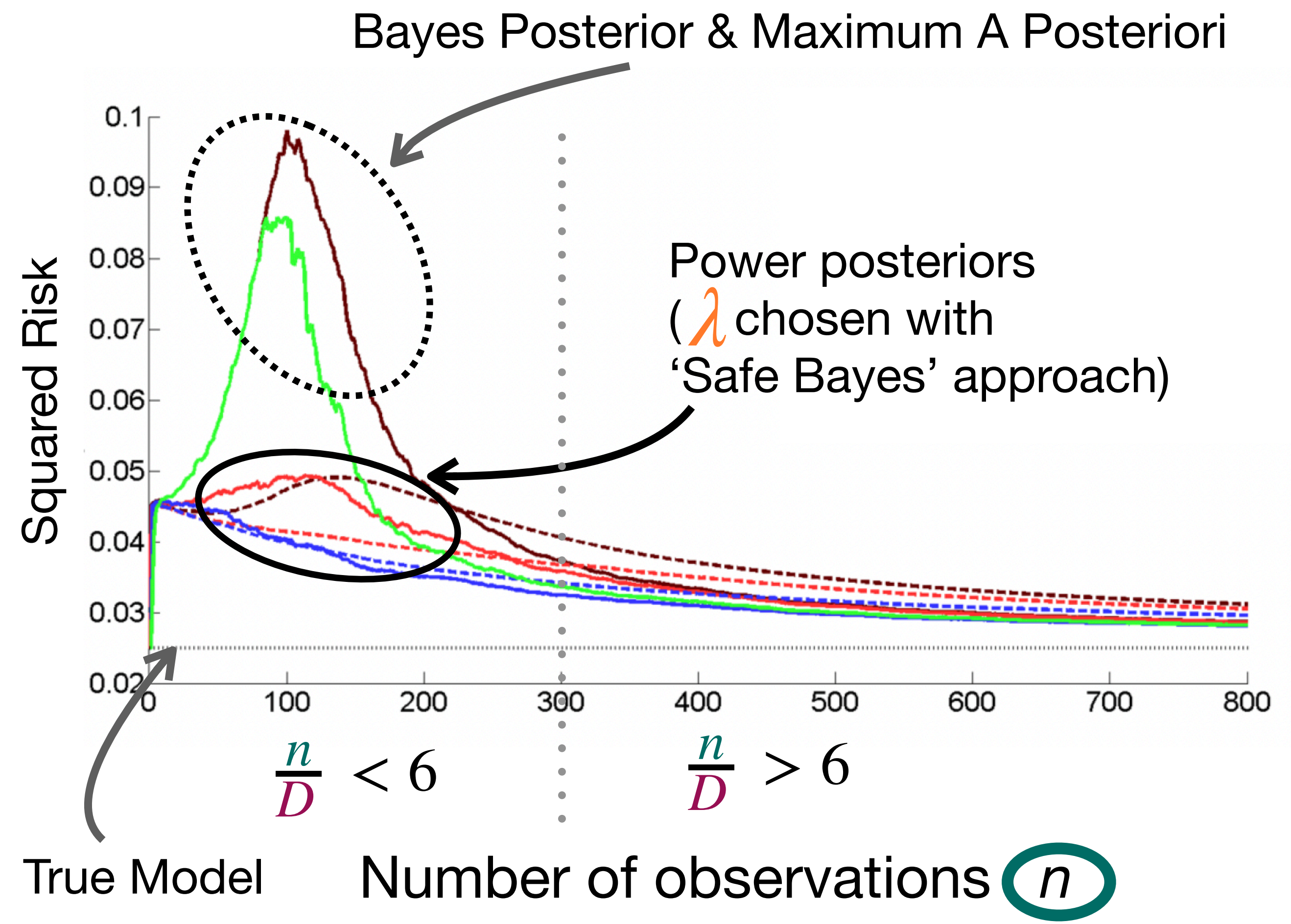
$$\pi_n^{(\lambda)}(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta)^\lambda \cdot \pi(\theta) d\theta}$$

Q: Why do it do?

Regression model (misspecified):

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

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



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?

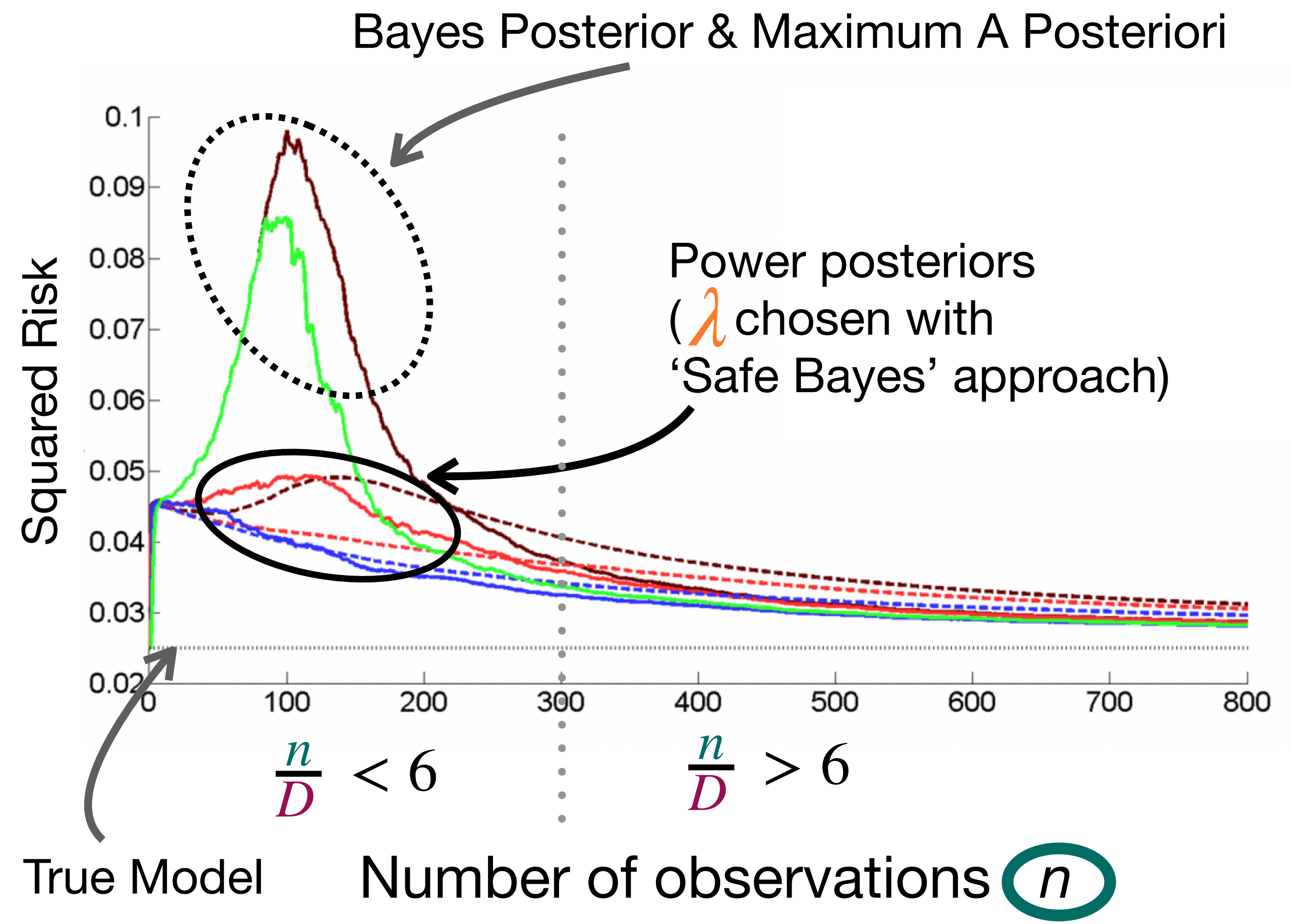
A: better risk properties/predictions

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

Regression model (misspecified):

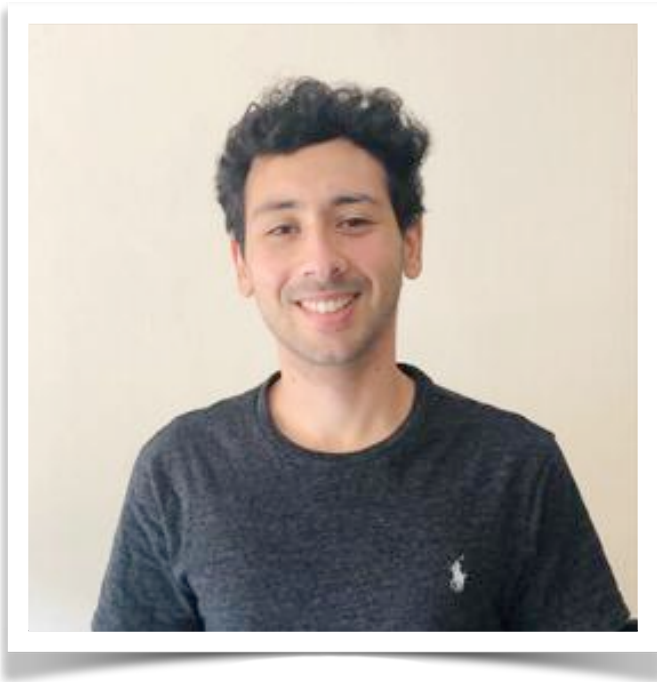
$$p(y_i \mid \theta, x_i) = \mathcal{N}\left(y_i; \sum_{d=1}^{50} \theta_d x_{i,d}, \sigma^2\right)$$

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



Ways to engage: Post-Bayes workshop

Held @ UCL 15. /16. May 2025;
>120 people in attendance!



**Matias
Altamirano
(UCL)**



**Yann
McLatchie
(UCL)**

