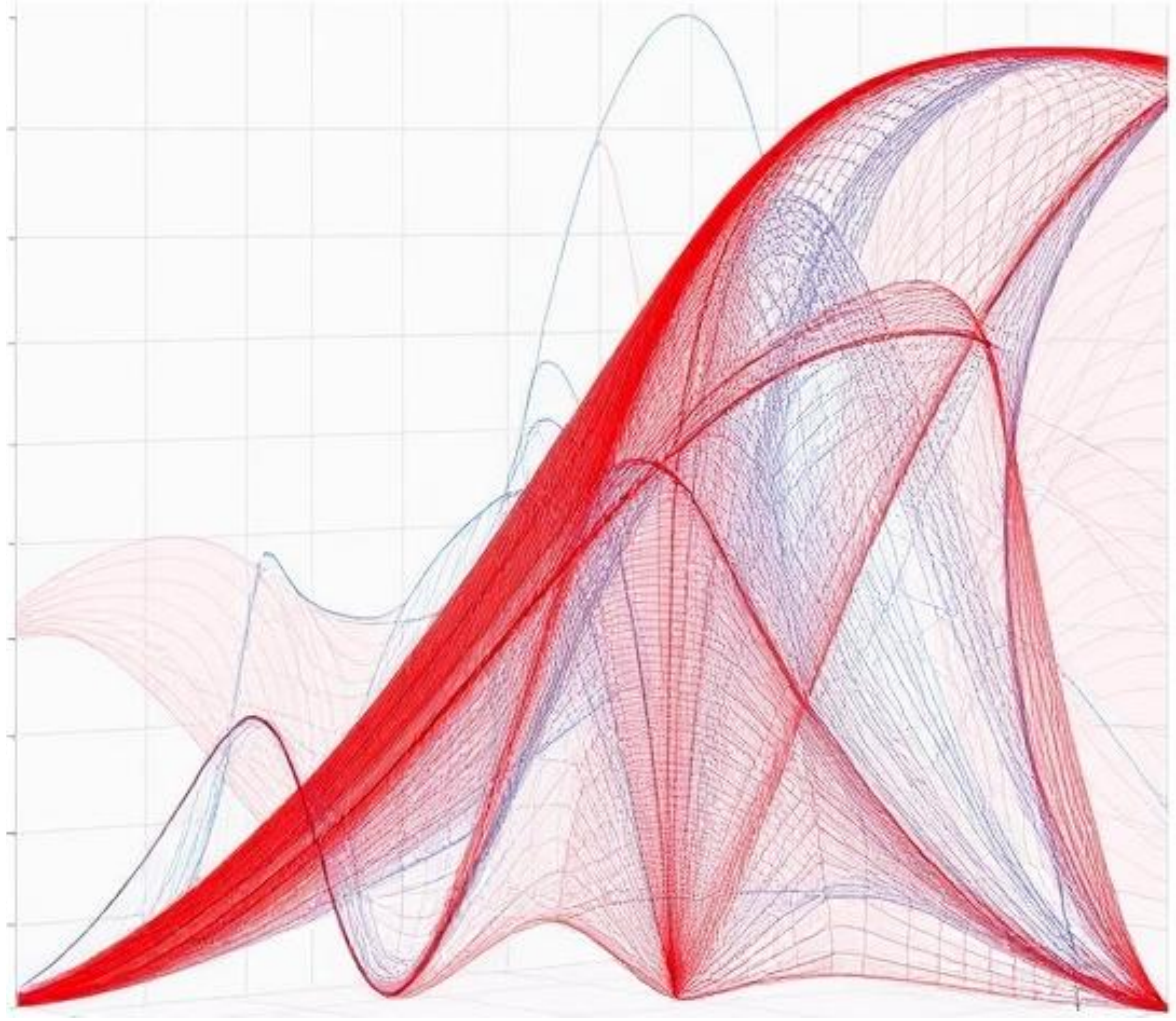
Post-Bayesian Beliefs



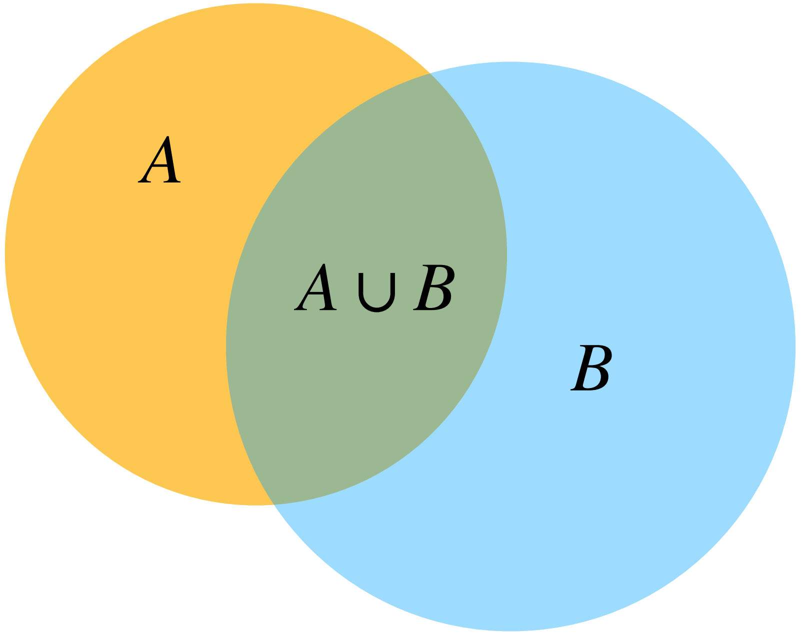
Jeremias Knoblauch; Associate Prof & EPSRC Fellow @ UCL Stats

24/03/26

Preamble: Bayesian Uncertainty

Bayes' Theorem: Inversion of conditionals

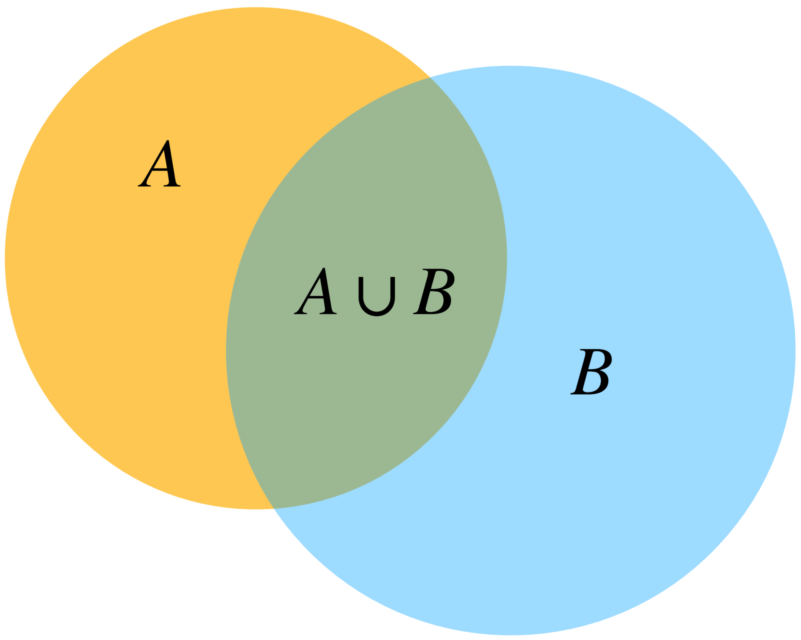
$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$



Preamble: Bayesian Uncertainty

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: $\pi(\theta)$

$\theta \in \Theta$

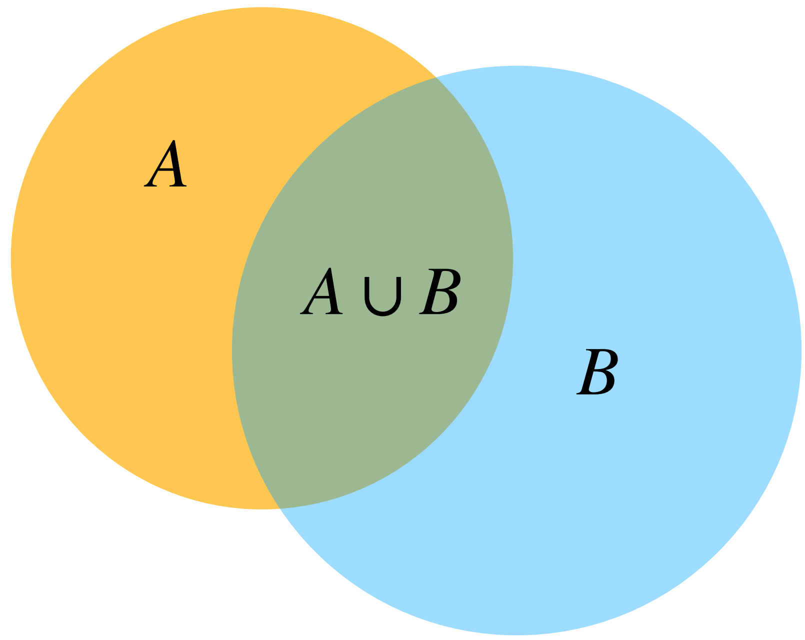
$$\boxed{\pi_n(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta) \cdot \pi(\theta)}{\int p(x_{1:n} \mid \theta) \cdot \pi(\theta) d\theta}}$$

(Bayes) Posterior

Preamble: Bayesian Uncertainty

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: $\pi(\theta)$

$\theta \in \Theta$

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

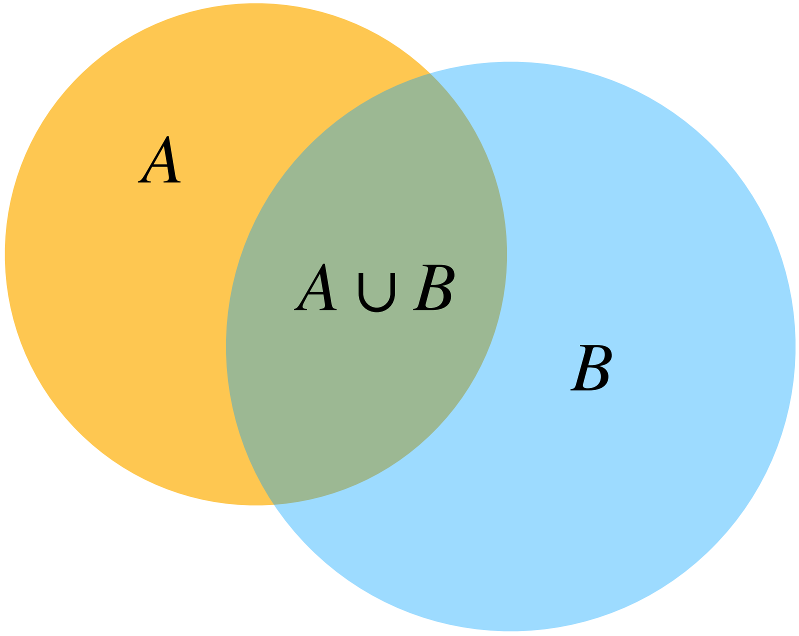
(Bayes) Posterior

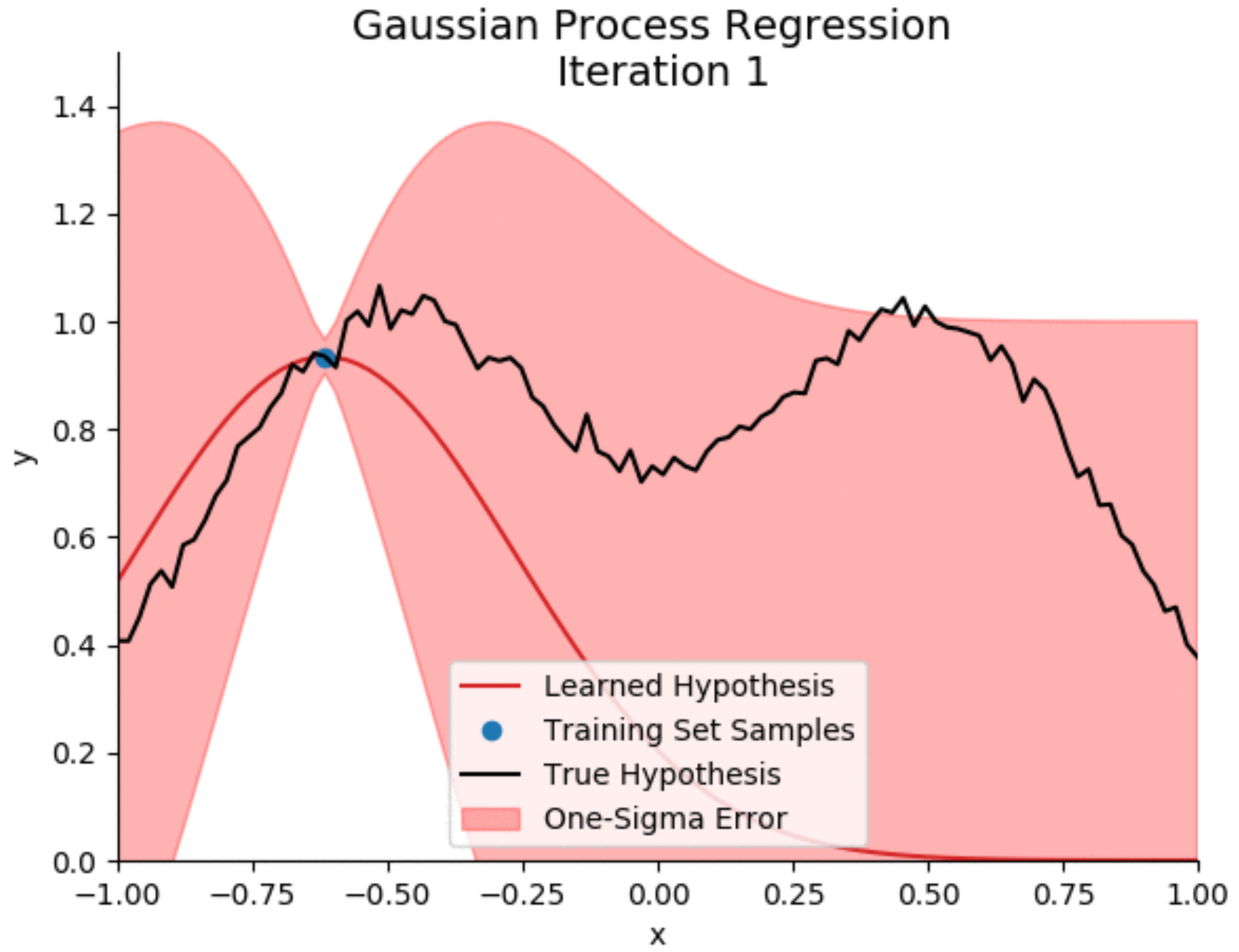
- ⊕ Averages models (instead of picking only one)
- ⊕ Quantifies uncertainty about θ via $\pi_n(\theta \mid x_{1:n})$
- ⊕ Inclusion of domain expertise via prior π

Preamble: Bayesian Uncertainty

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

Assumptions for Bayes Posteriors

(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

Assumptions for Bayes Posteriors

(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

Assumptions for Bayes Posteriors

(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

Assumptions for Bayes Posteriors

(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

FRAGILE

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

Guarantees real-world relevance

Case Study: Boston Housing Data

Traditional Bayesian analysis in science

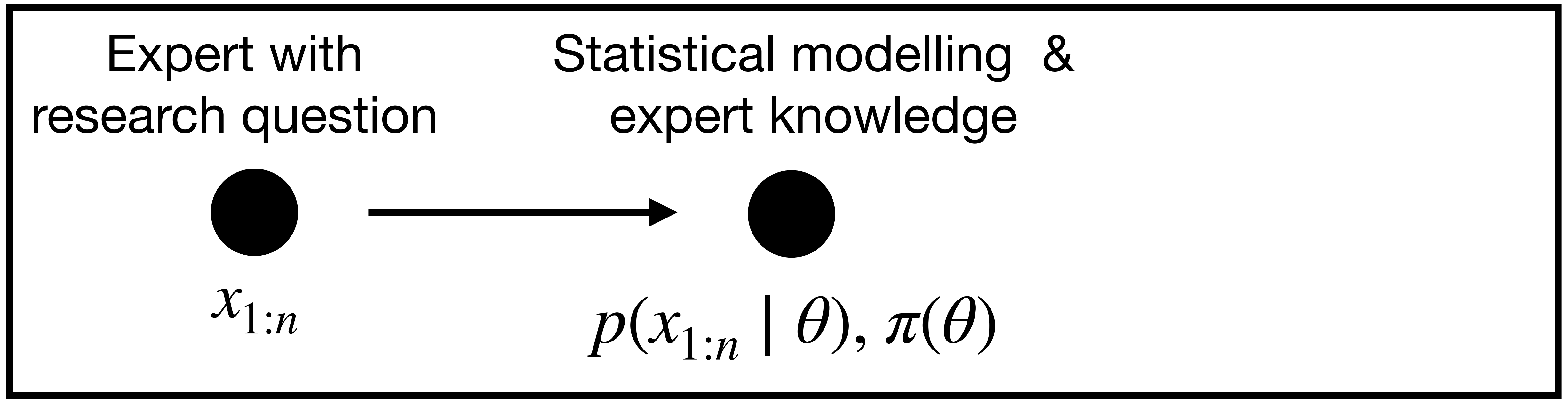
Expert with research question

$x_{1:n}$

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

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



Expert with research question

Statistical modelling & expert knowledge

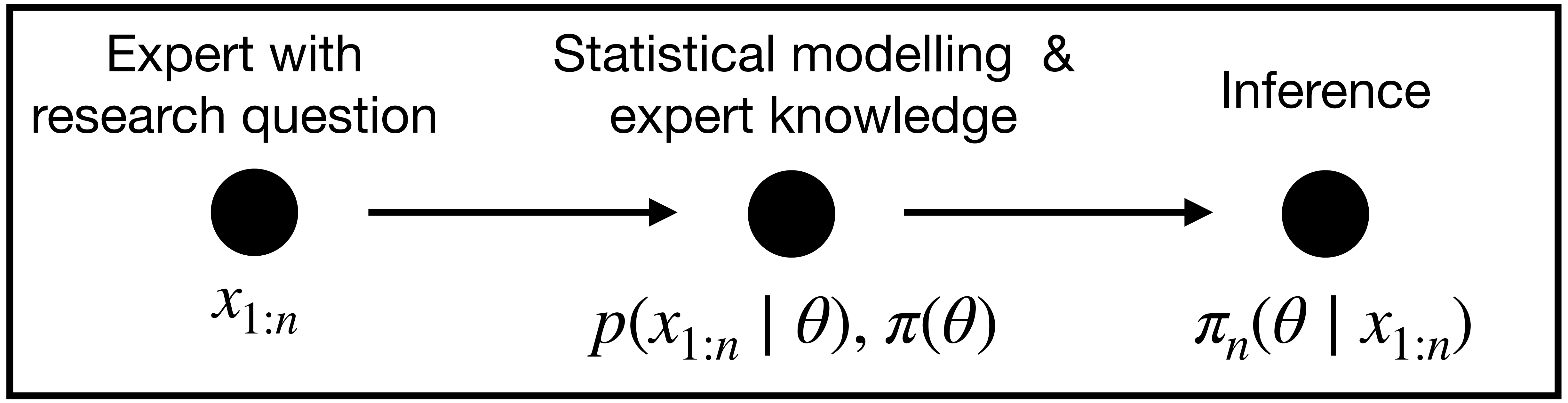
$x_{1:n}$

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

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

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



Expert with research question

Statistical modelling & expert knowledge

Inference

$x_{1:n}$

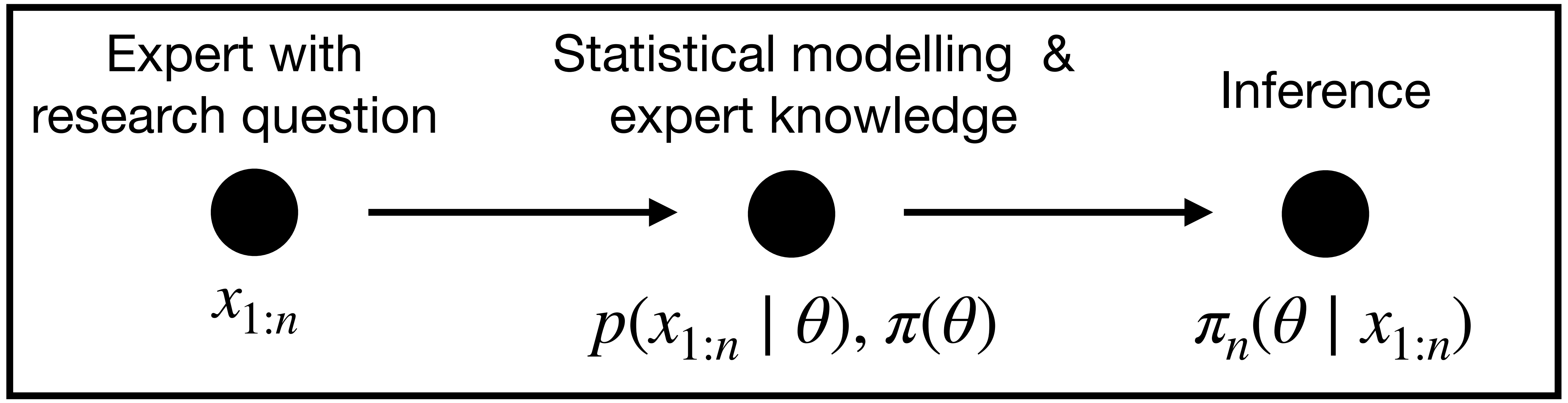
$p(x_{1:n} \mid \theta), \pi(\theta)$

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

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

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



Expert with research question

$x_{1:n}$

Statistical modelling & expert knowledge

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

Inference

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

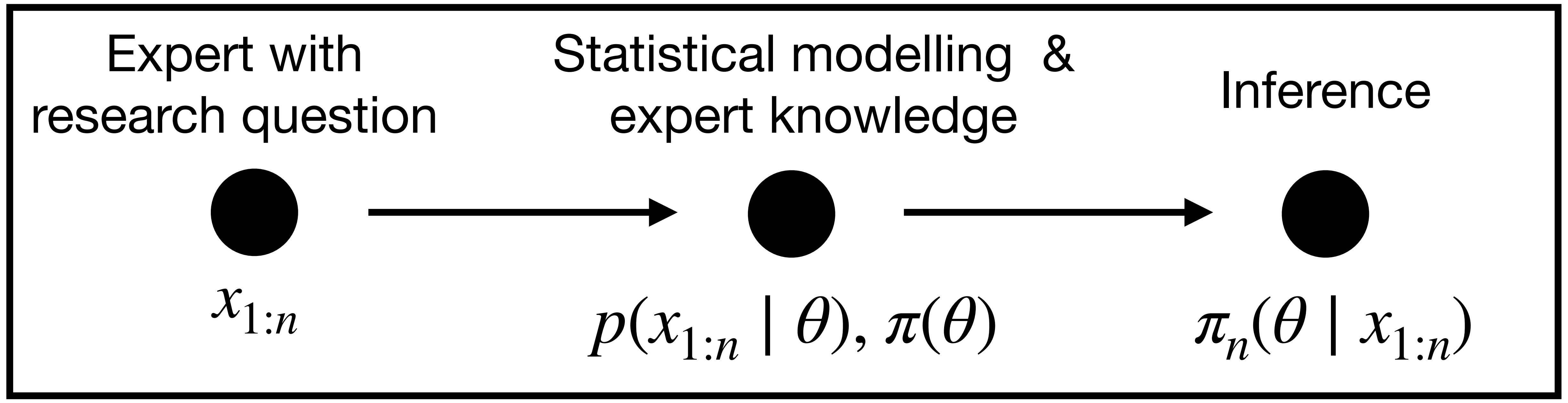
Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

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

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



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

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

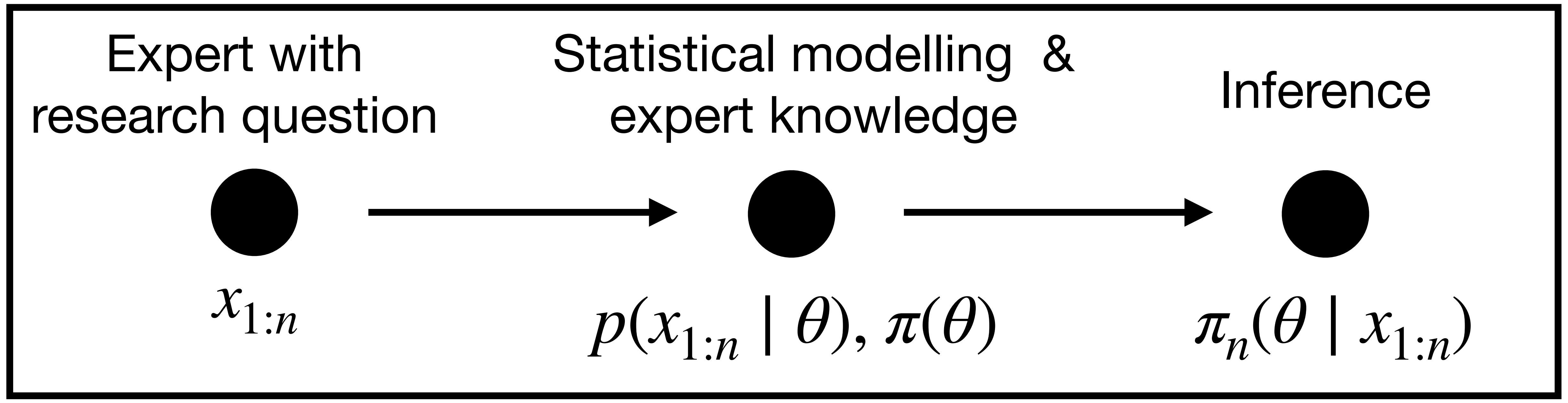
(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



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) ✓

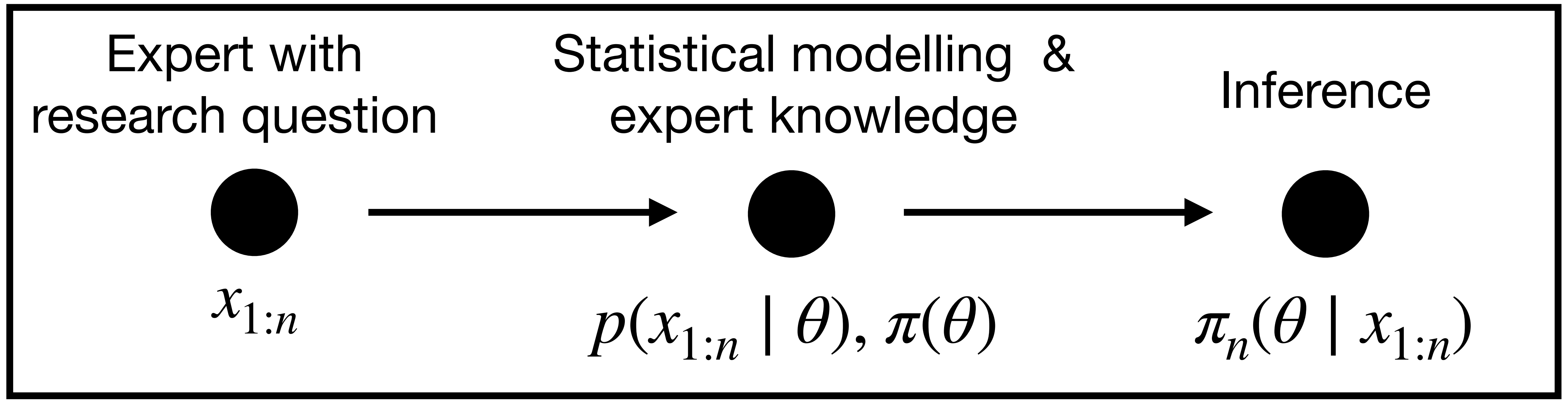
(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

Case Study: Boston Housing Data

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) ✓

$$\underset{\substack{\uparrow\\ \text{willingness to pay}}}{\log y_i} = \sum_{j=1}^{J_1} \overset{\text{parameters of interest}}{p_j} \log(\underset{\substack{\uparrow\\ \text{pollutants}}}{x_{j,i}}) + c_0 + \sum_{j=J_1}^{J_2} \overset{\text{incidental parameters}}{c_j} \log(\underset{\substack{\uparrow\\ \text{rooms, sqm, ...}}}{x_{j,i}}) + \underset{\substack{\uparrow\\ \text{measurement error}}}{\varepsilon_i}$$

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

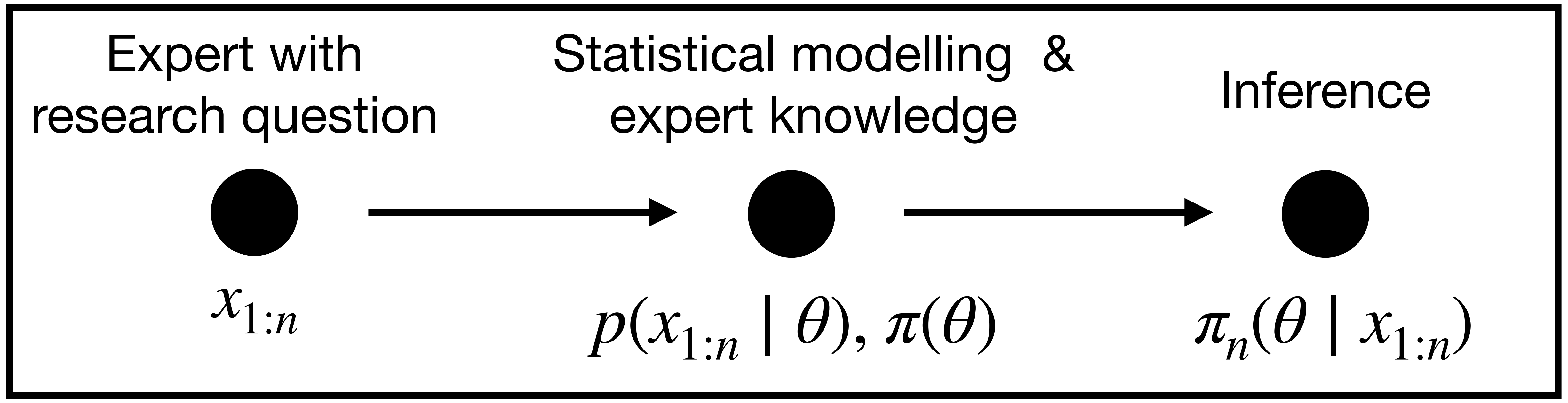
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

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

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

Case Study: Boston Housing Data

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

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

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

$$\underset{\substack{\uparrow \\ \text{willingness to pay}}}{\log y_i} = \sum_{j=1}^{J_1} \overset{\text{parameters of interest}}{p_j} \underset{\substack{\uparrow \\ \text{pollutants}}}{\log(x_{j,i})} + c_0 + \sum_{j=J_1}^{J_2} \overset{\text{incidental parameters}}{c_j} \underset{\substack{\uparrow \\ \text{rooms, sqm, ...}}}{\log(x_{j,i})} + \underset{\substack{\uparrow \\ \text{measurement error}}}{\varepsilon_i}$$

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

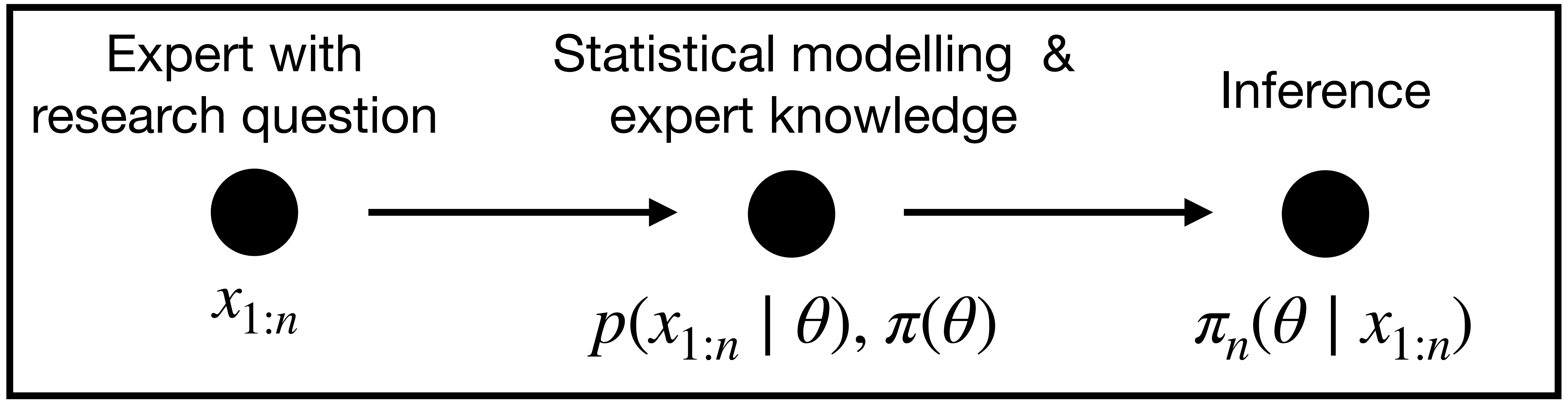
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

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

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

Case Study: Boston Housing Data

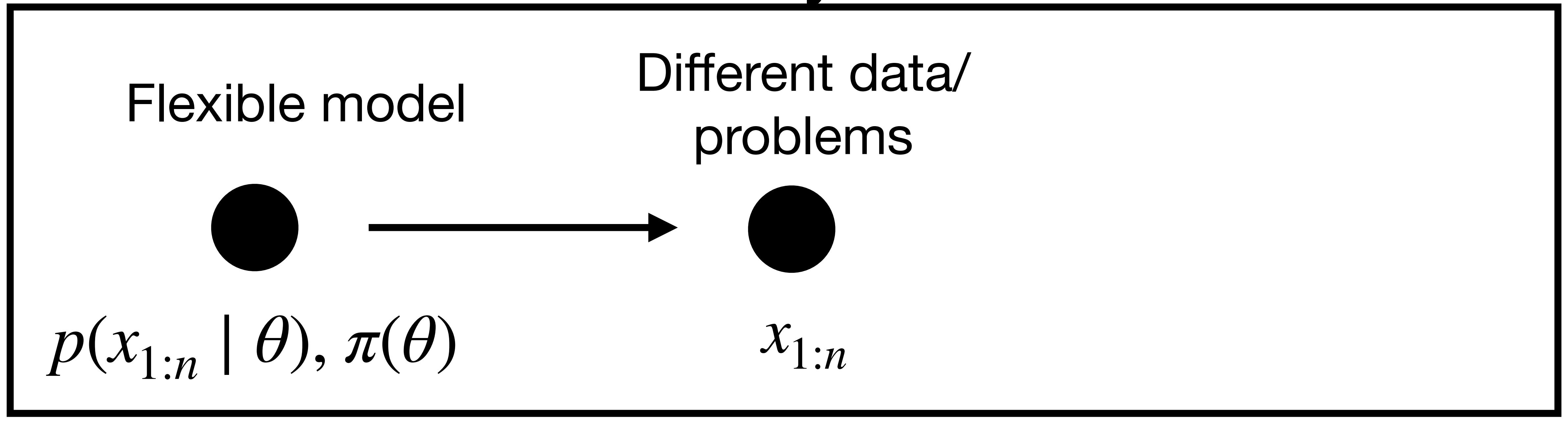
Traditional Bayesian analysis in science



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

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

Harrison & Rubinfeld (1978)

Research Question: influence of air pollution on house prices?

(A1) ✓

parameters of interest

incidental parameters

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

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

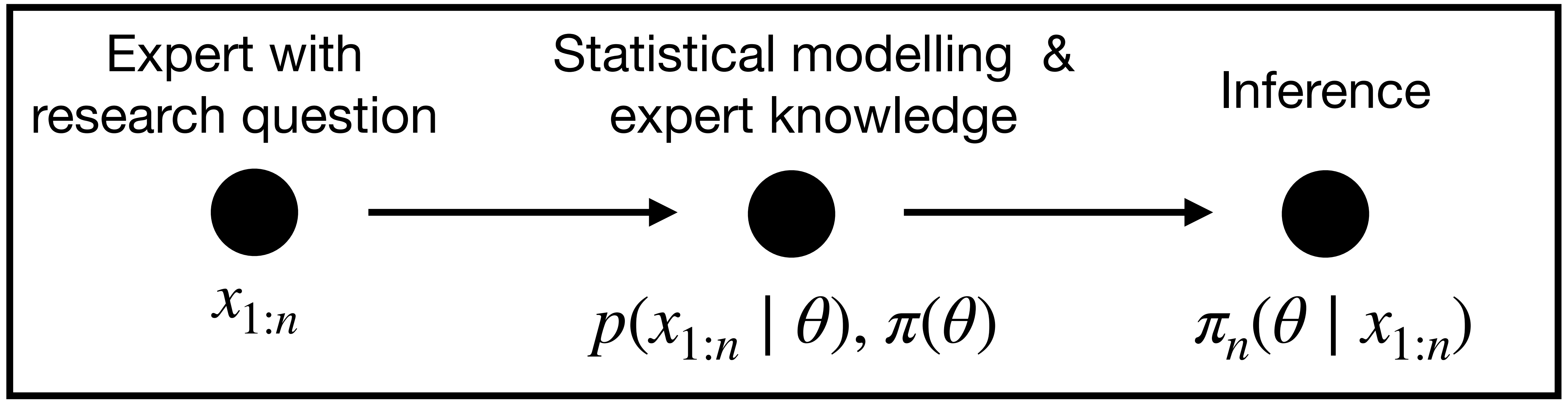
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

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

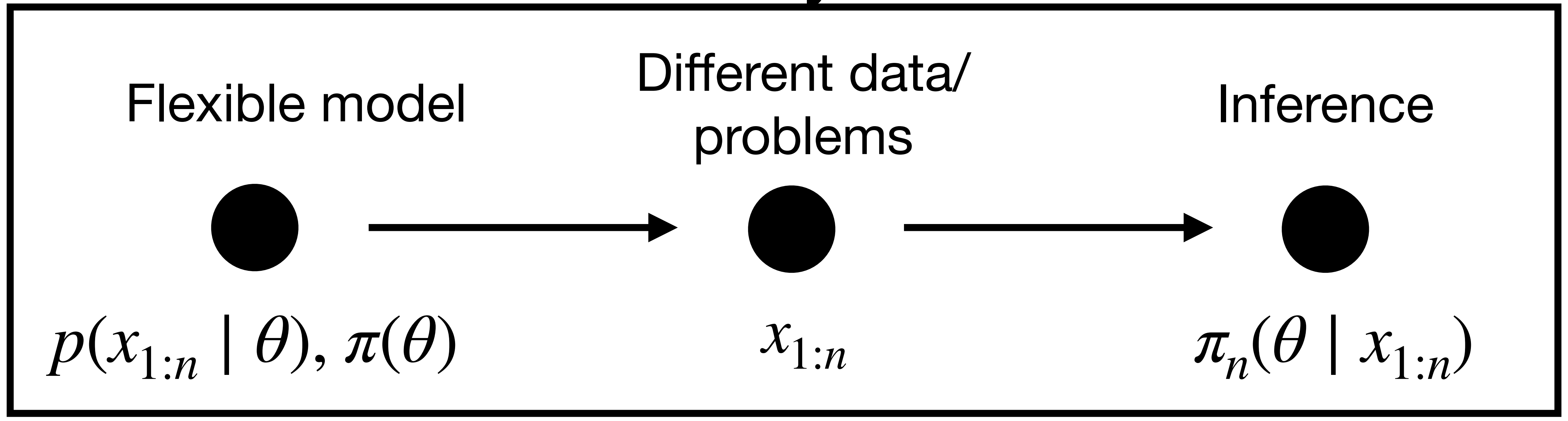
(A1) model well-specified
(A2) prior well-specified
(A3) computationally feasible

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



Modern Bayesian ML



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

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

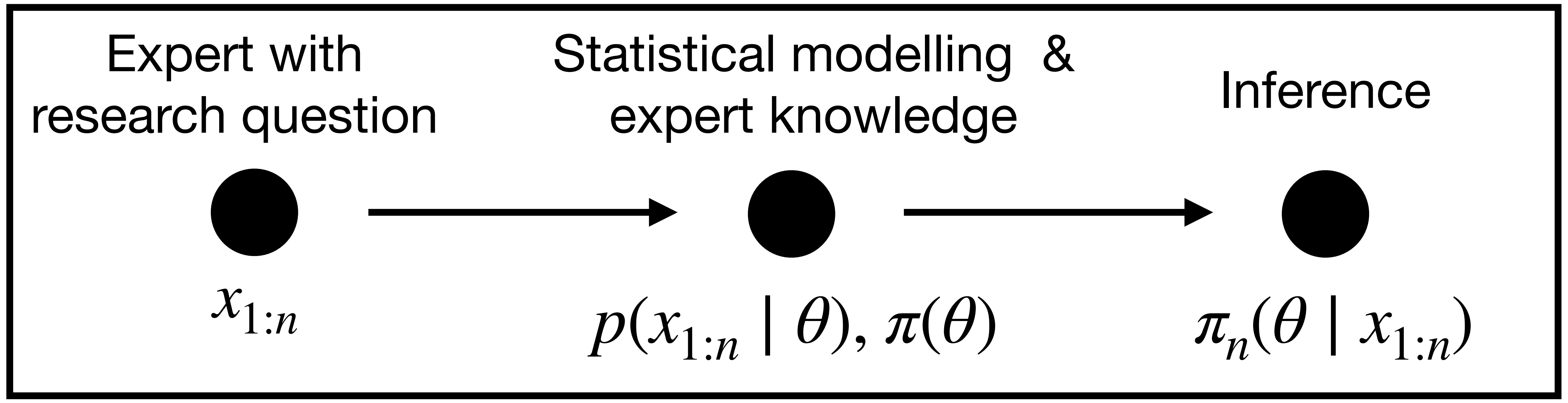
$\pi(\theta) \sim$ hand-crafted by experts (A2) ✓

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

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

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



Expert with research question

$x_{1:n}$

Statistical modelling & expert knowledge

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

Inference

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

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, ...

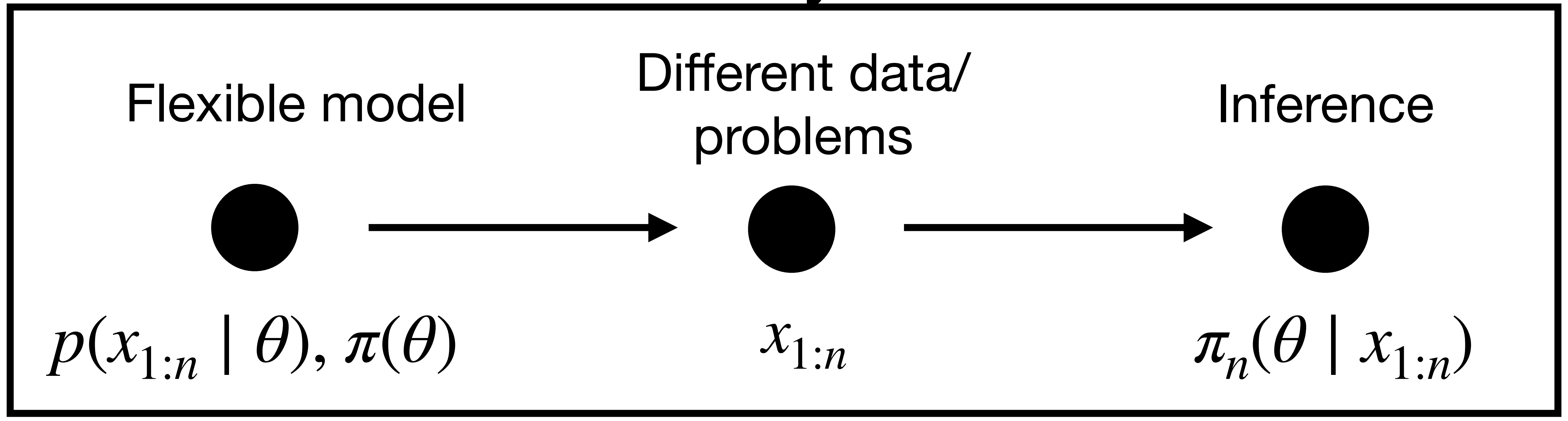
measurement error

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

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

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

Modern Bayesian ML



Flexible model

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

Different data/ problems

$x_{1:n}$

Inference

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

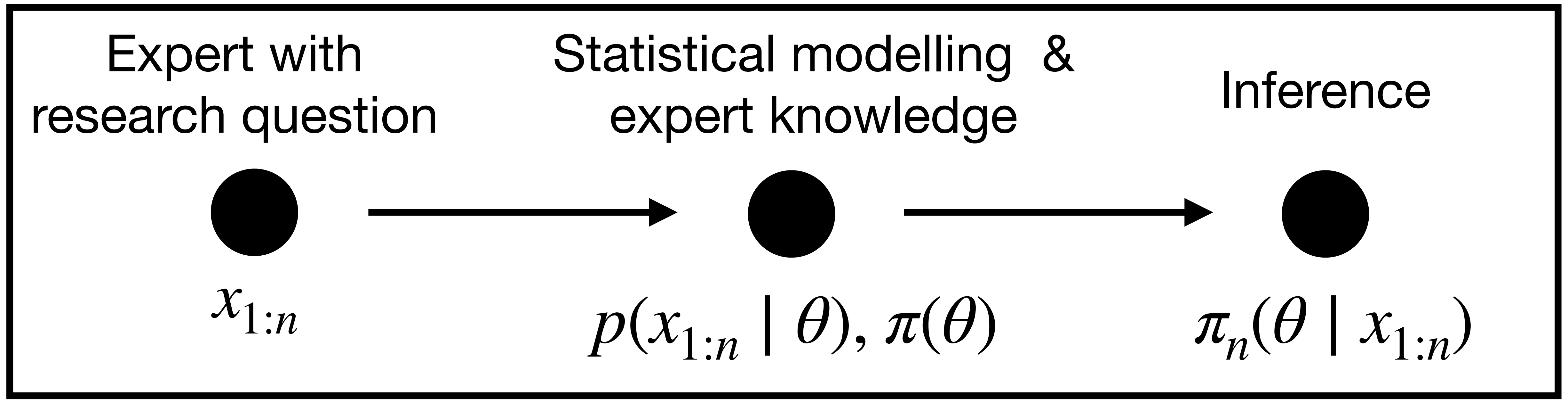
Pearce et al. (2020) [AISTATS]

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: Boston Housing Data

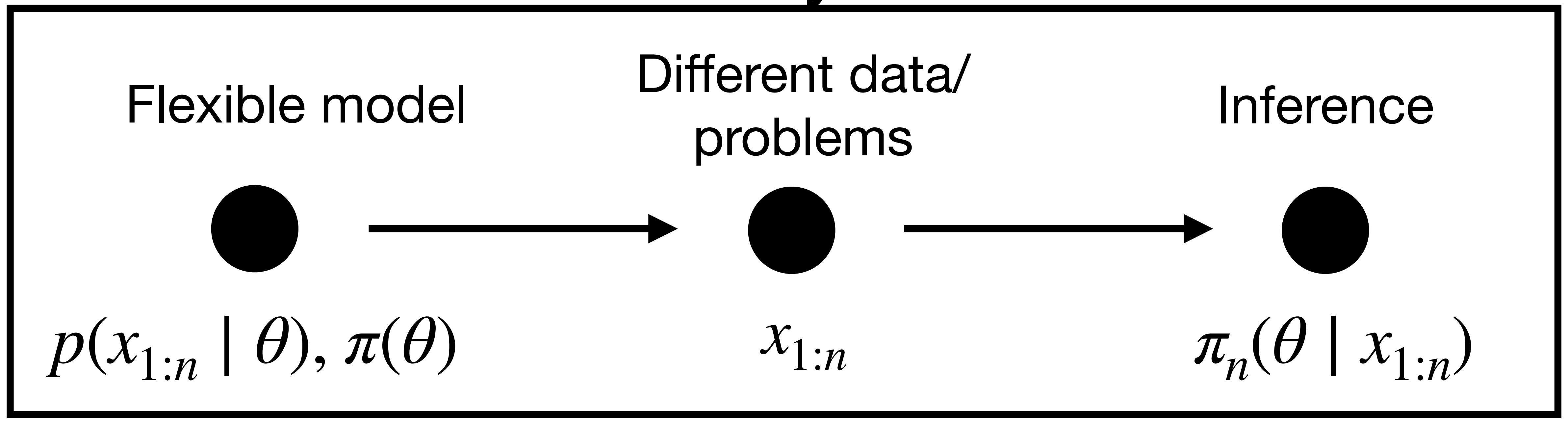
Traditional Bayesian analysis in science



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

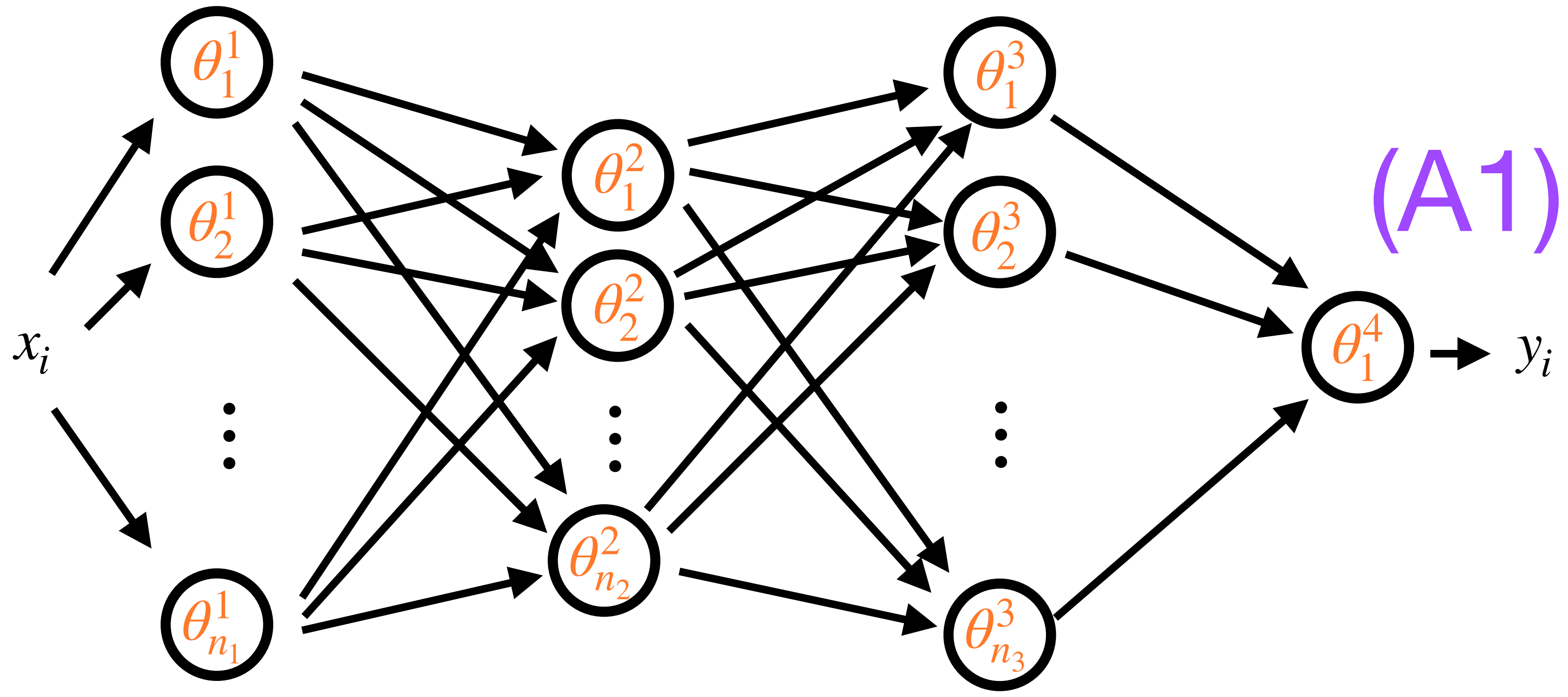
$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ 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?

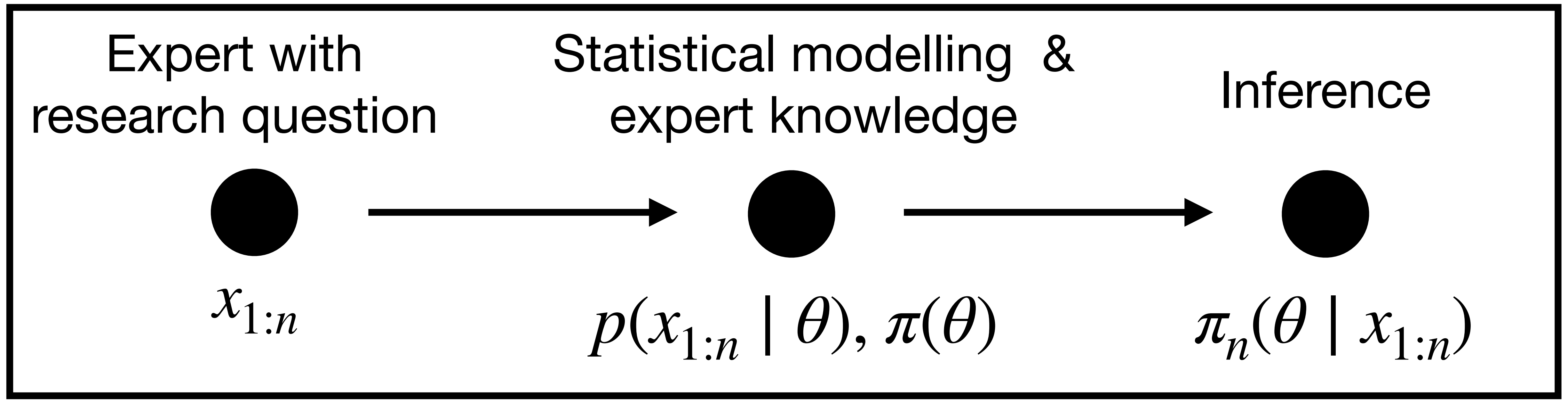


(A1) ✓

F_θ : Input → Output

Case Study: Boston Housing Data

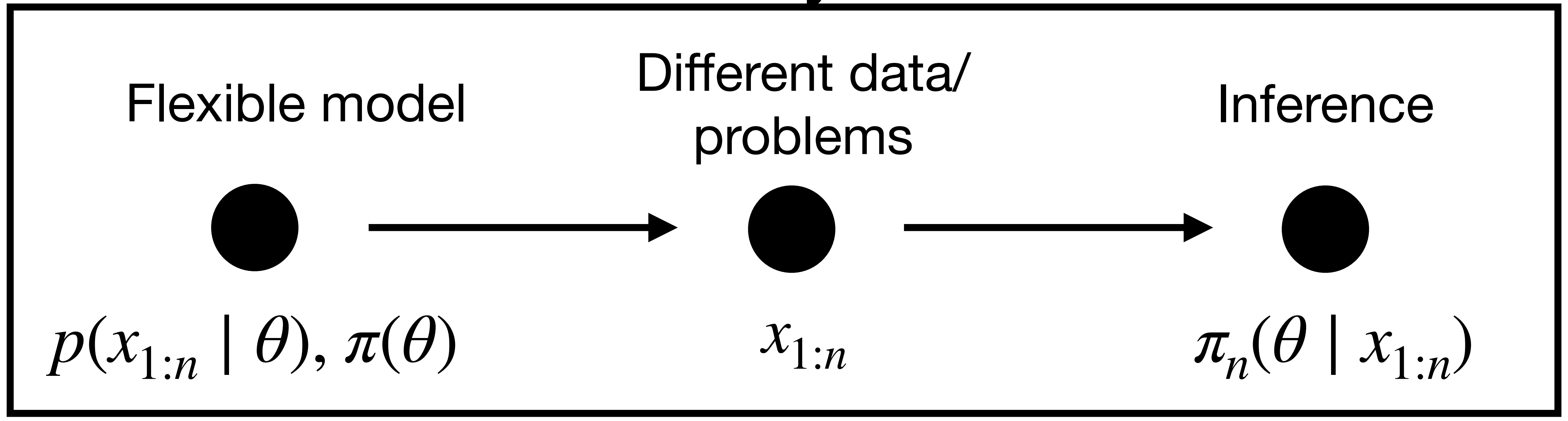
Traditional Bayesian analysis in science



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

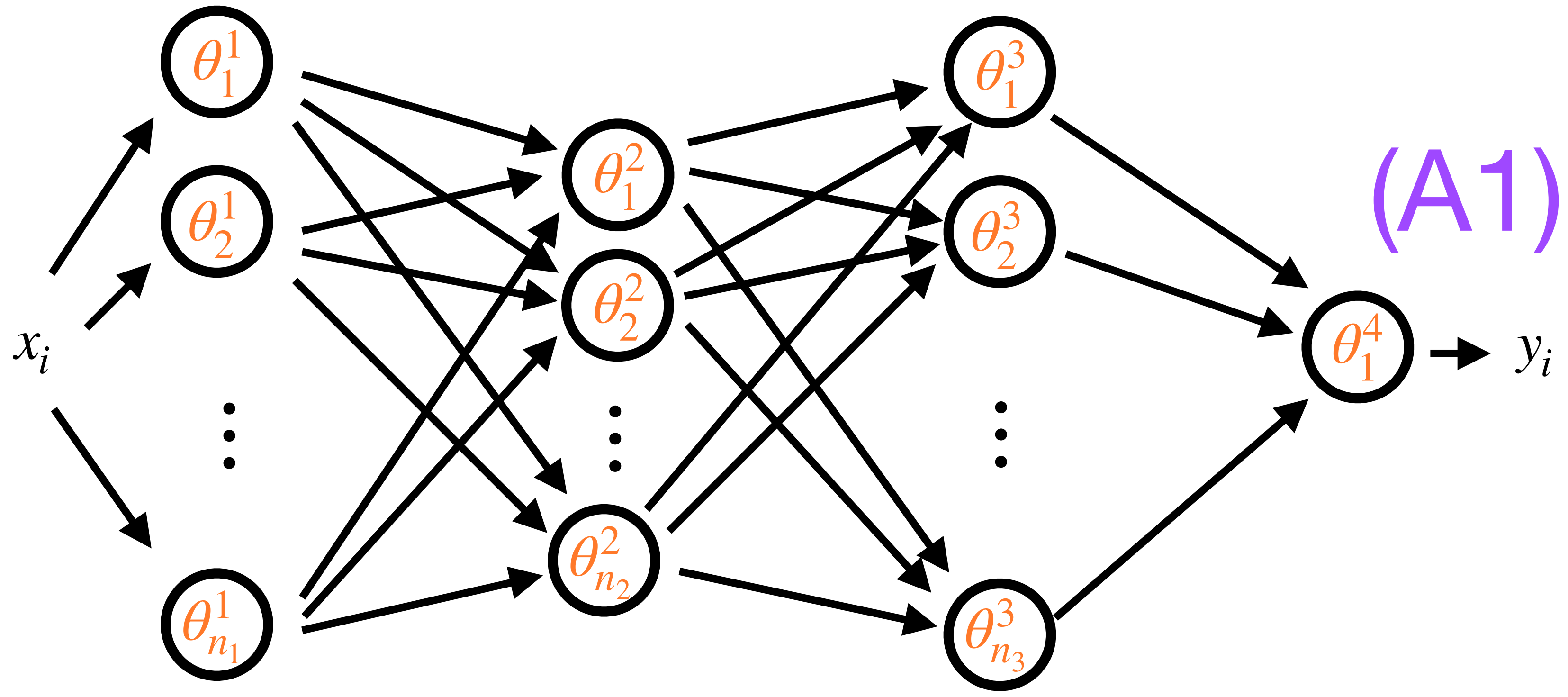
$\theta = (c_0, c_2, \ldots, c_{J_1}, p_1, p_2 \ldots p_{J_2})^\top$

$\pi(\theta) \sim$ 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) ✓

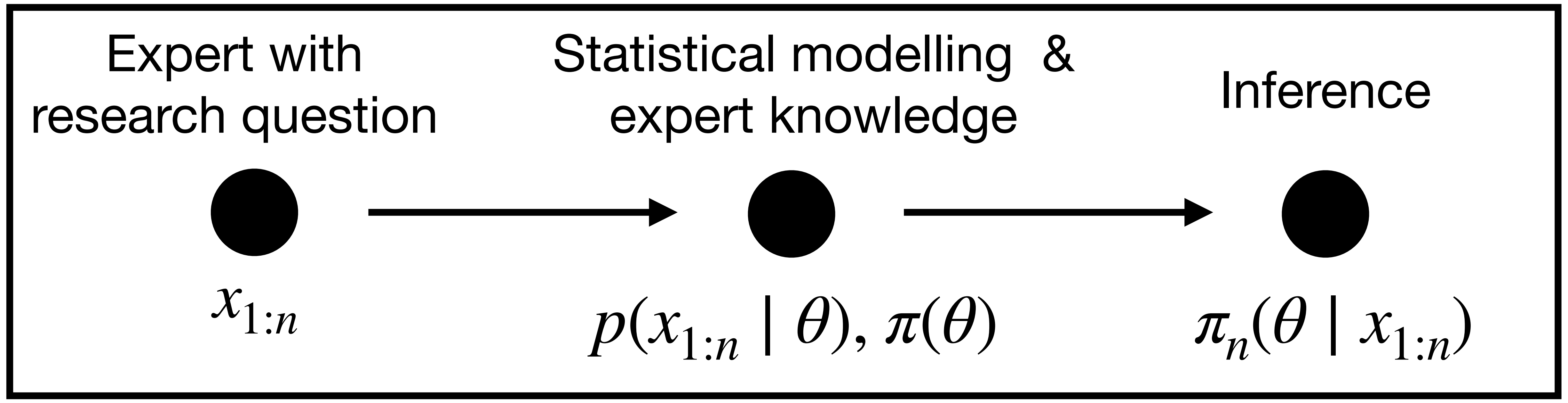
F_θ : Input → Output

$\pi(\theta) \sim$ Normal

~~(A2)~~

Case Study: Boston Housing Data

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, ...

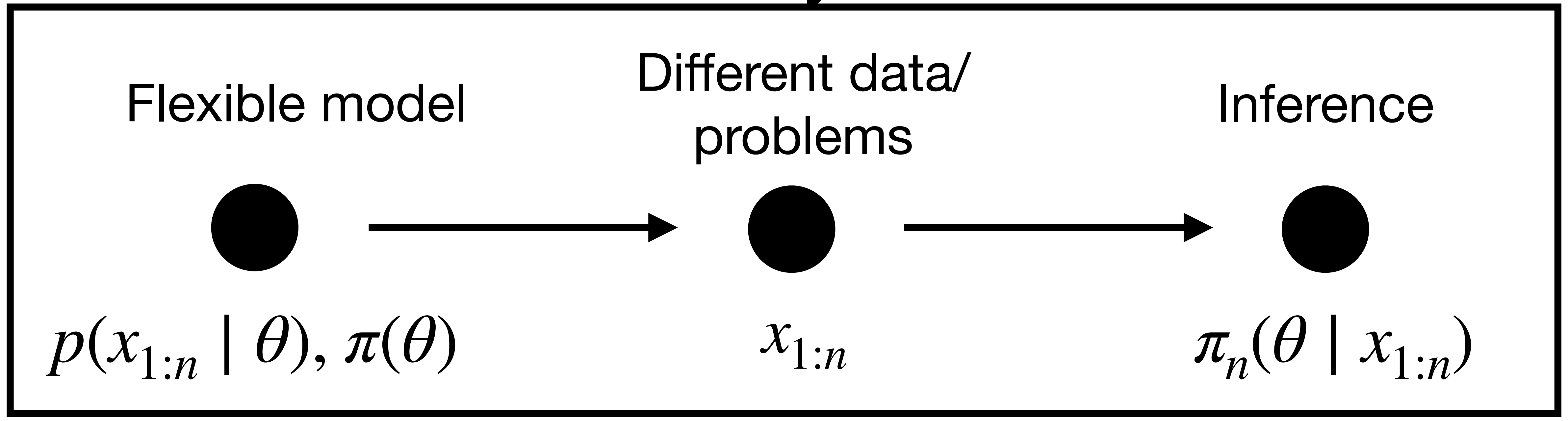
measurement error

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

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

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

Modern Bayesian ML

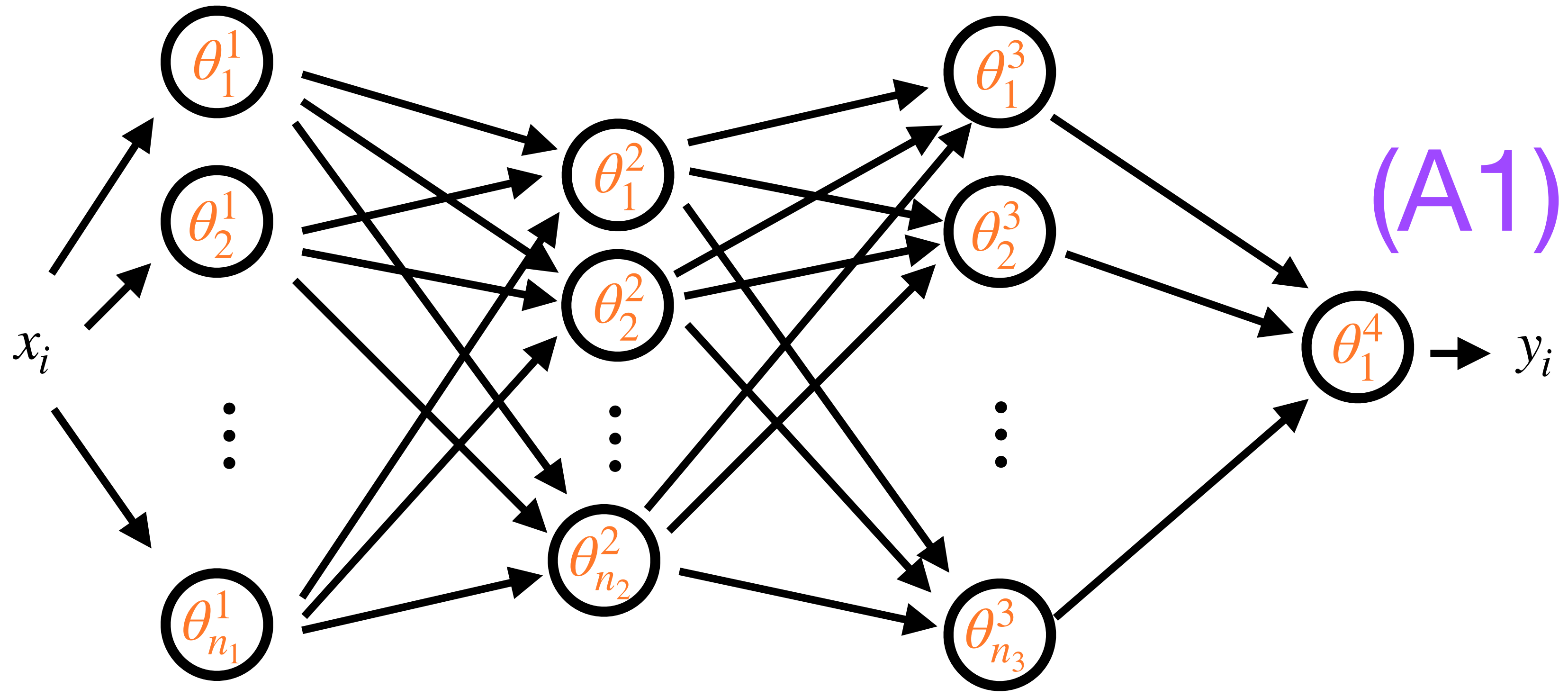


Flexible model → Different data/ problems → Inference

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

Pearce et al. (2020) [AISTATS]

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



(A1) ✓

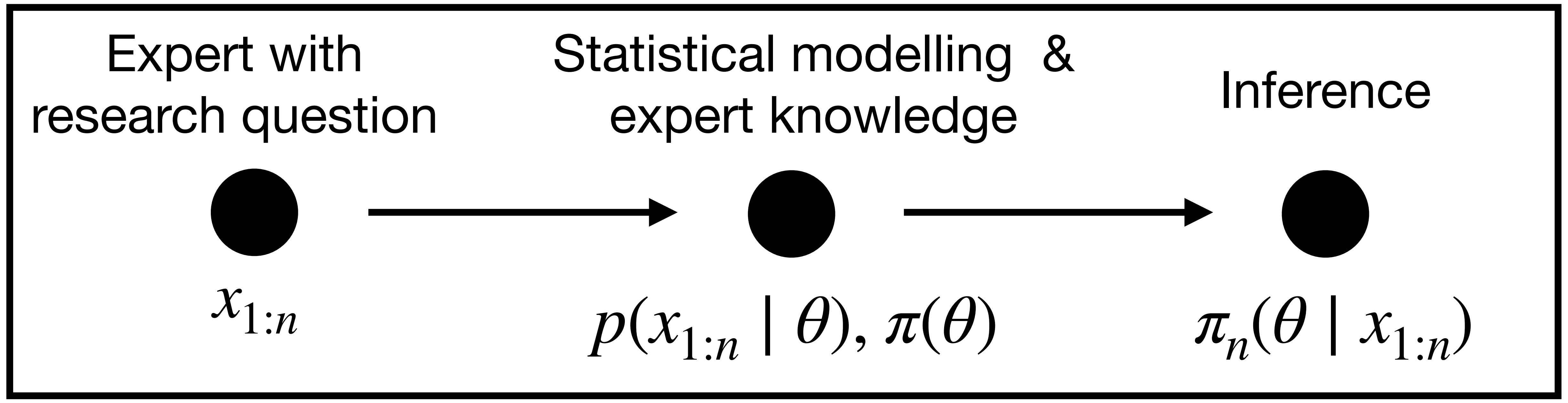
F_θ : Input → Output

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

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

Case Study: Boston Housing Data

Traditional Bayesian analysis in science



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

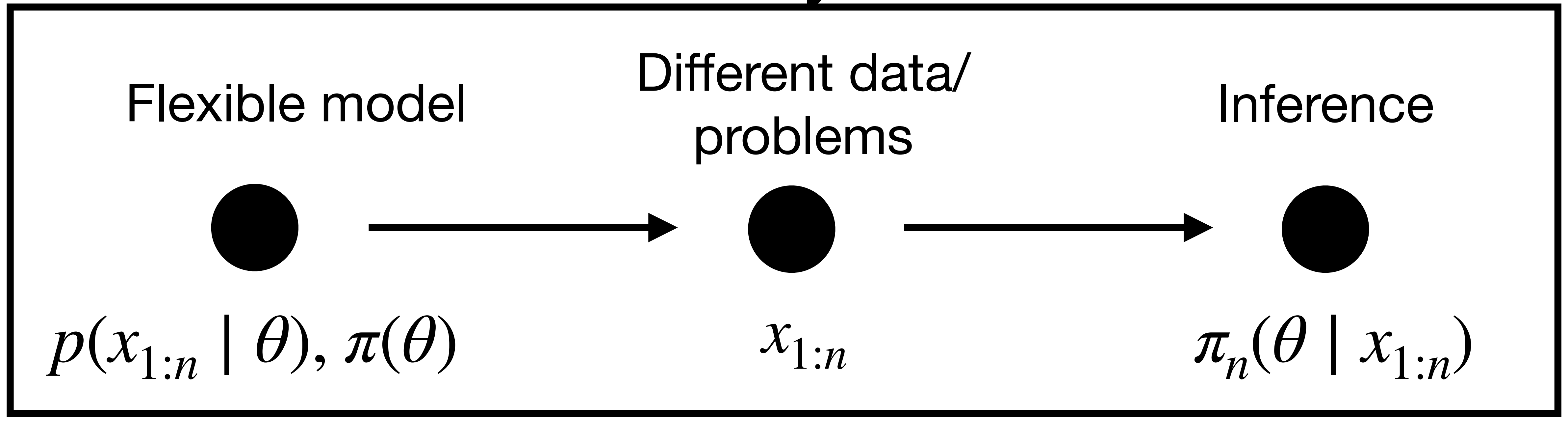
$x_{1:n}$ → $p(x_{1:n} \mid \theta), \pi(\theta)$ → $\pi_n(\theta \mid x_{1:n})$

(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

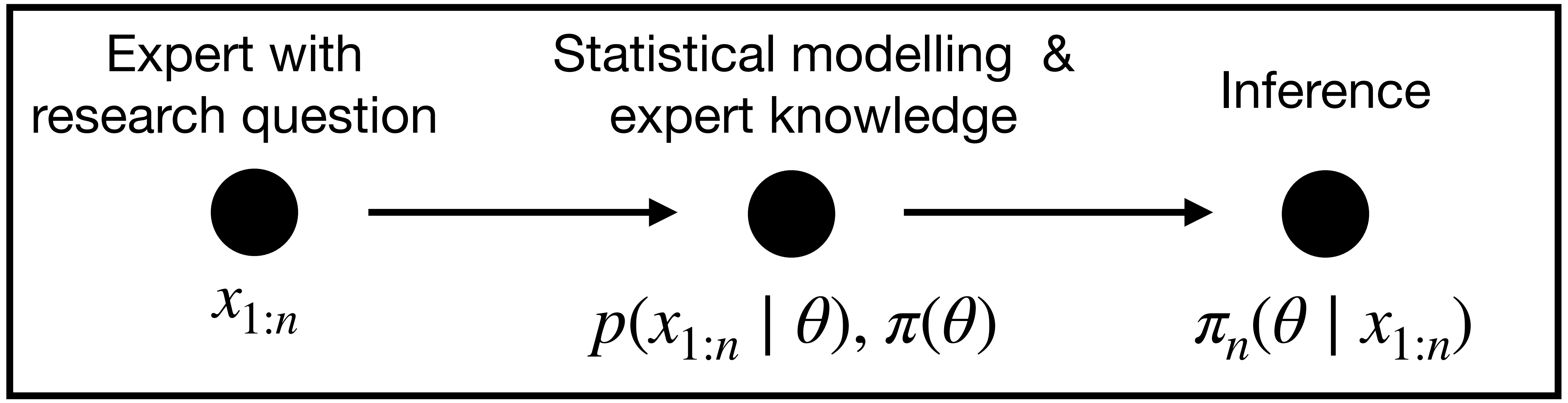
(A2) prior well-specified

(A3) computationally feasible

FRAGILE

Case Study: Boston Housing Data

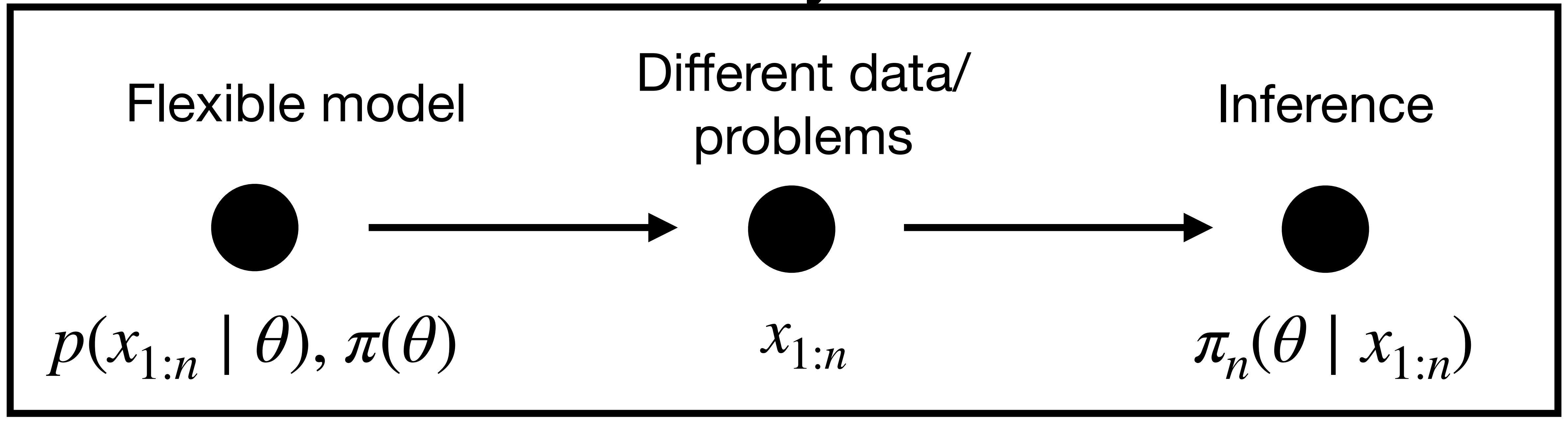
Traditional Bayesian analysis in science



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

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

(A1) model well-specified

(A2) prior well-specified

(A3) computationally feasible

(A1) model well-specified

(A2) prior well-specified

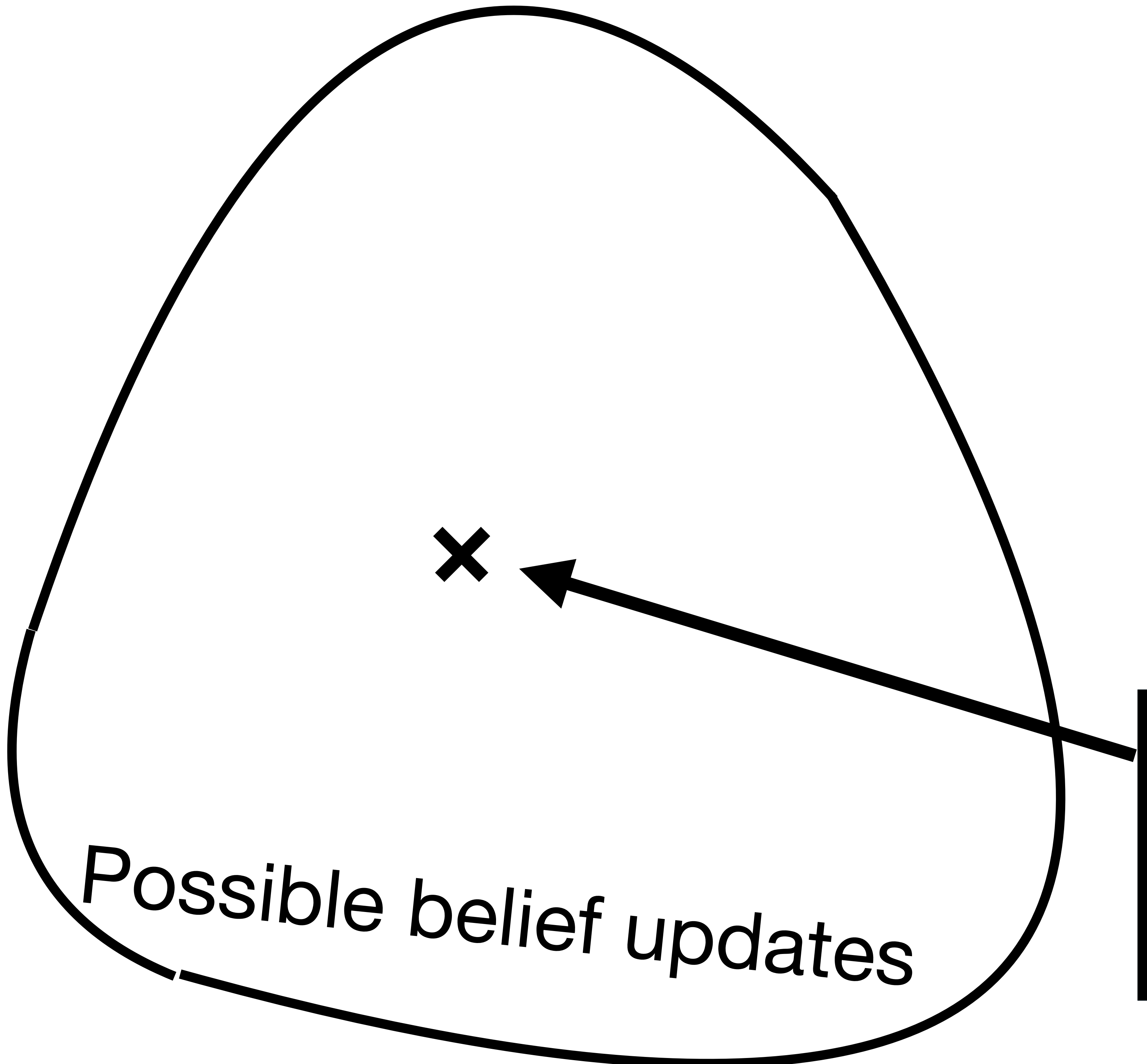
(A3) computationally feasible

FRAGILE

Post-Bayesian Approaches ask:

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

Post-Bayesian Beliefs



Possible belief updates

Bayes’ Posterior (A1), (A2), (A3)

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

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

Post-Bayesian Beliefs

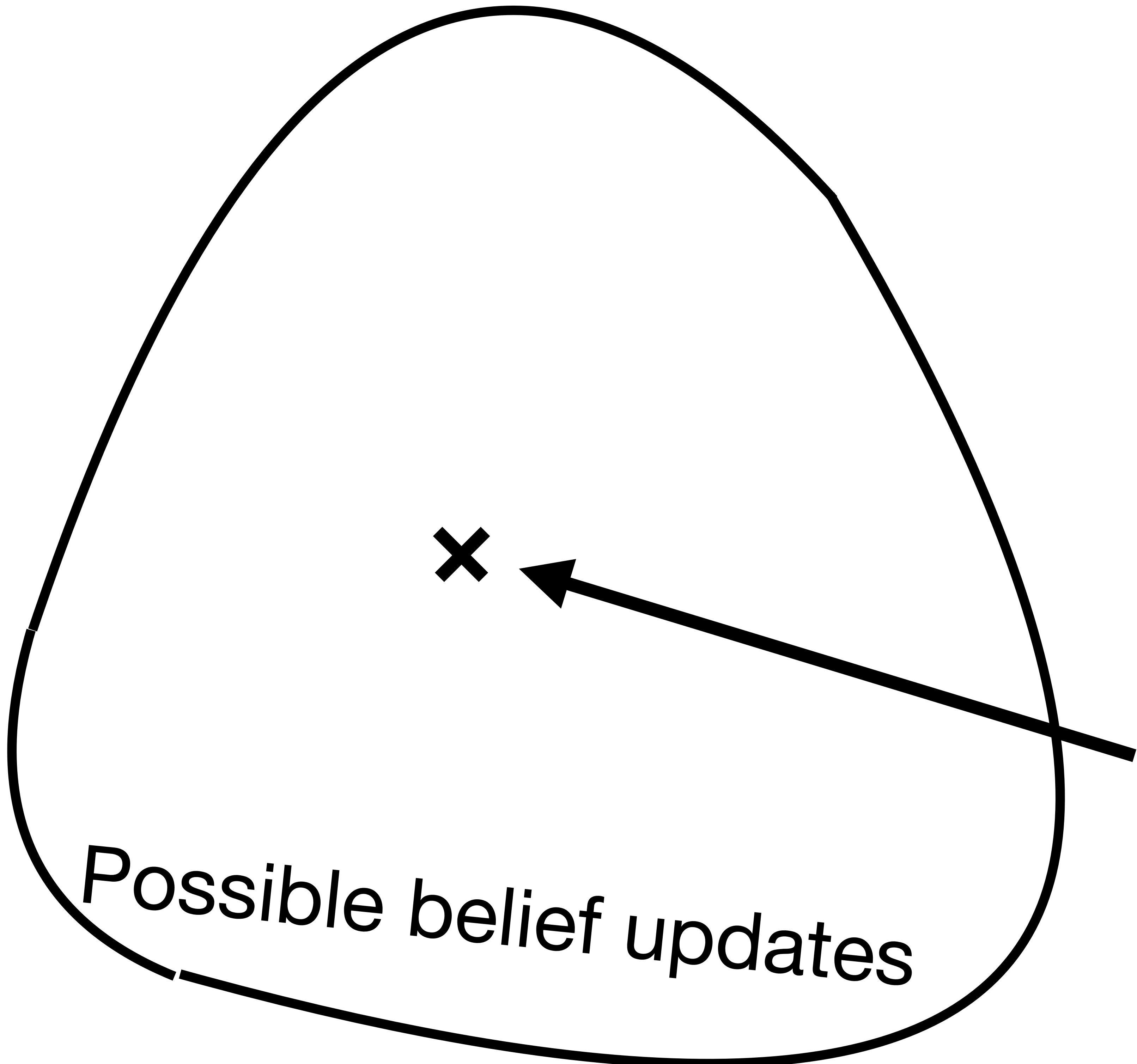
space of priors

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



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 Beliefs

space of priors

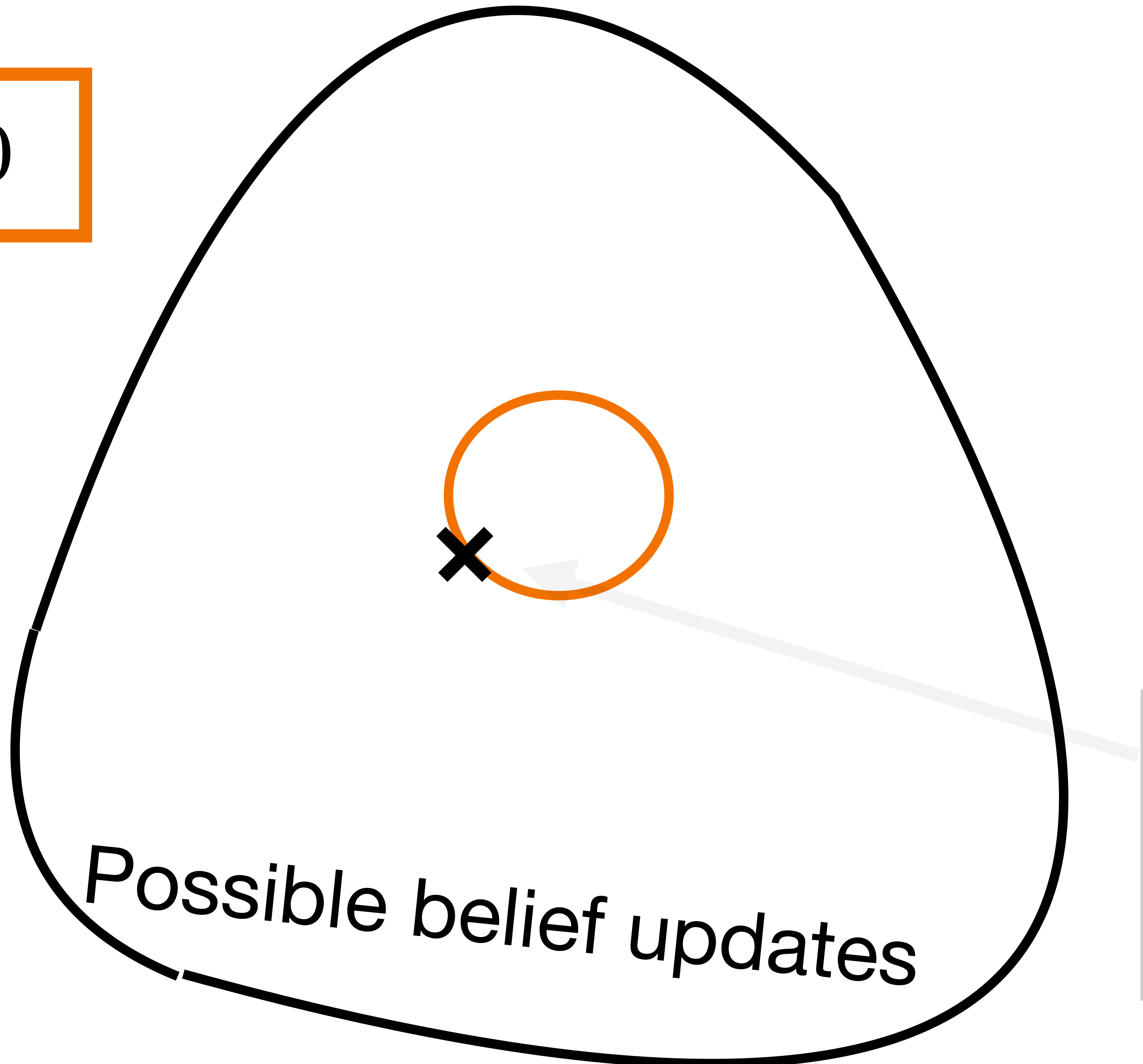
space of hyperparameters

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

data space

space of posteriors

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



(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 Beliefs

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)]

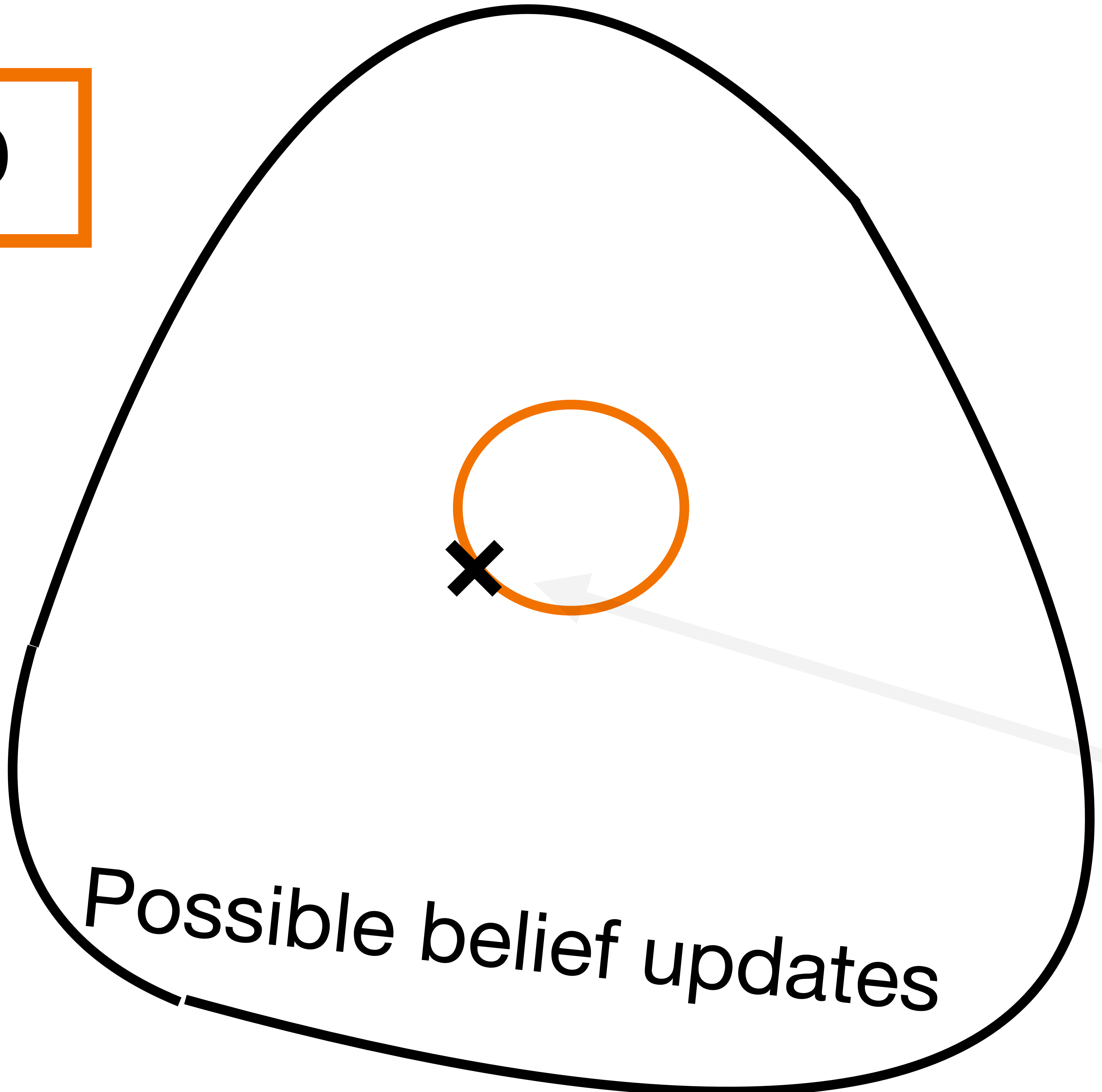
Power/Fractional/
Cold Posterior

~~(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)

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



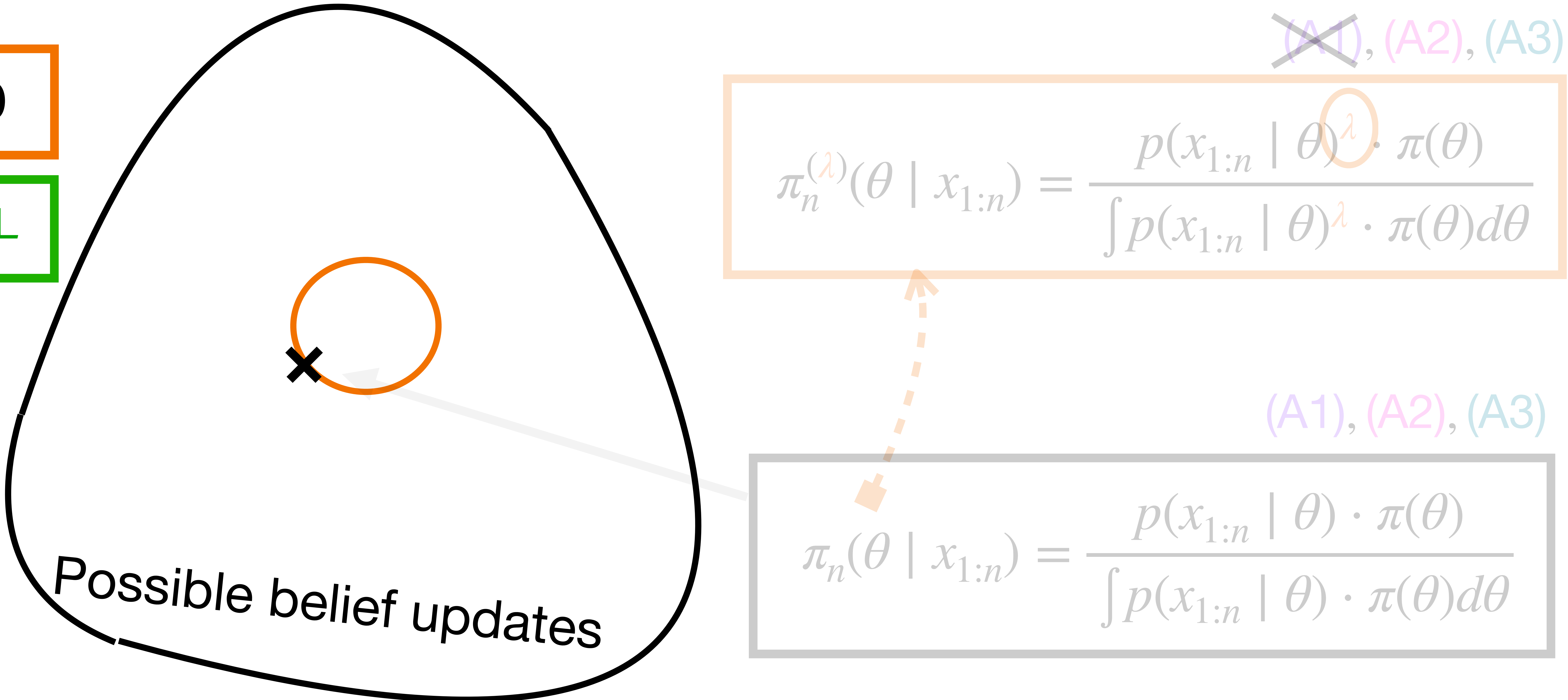
- (A1) model well-specified
- (A2) prior well-specified
- (A3) computationally feasible

Post-Bayesian Beliefs

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

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

(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 Beliefs

[See Bissiri, Holmes & Walker (2016)]

Gibbs/Generalised/
Pseudo Posterior

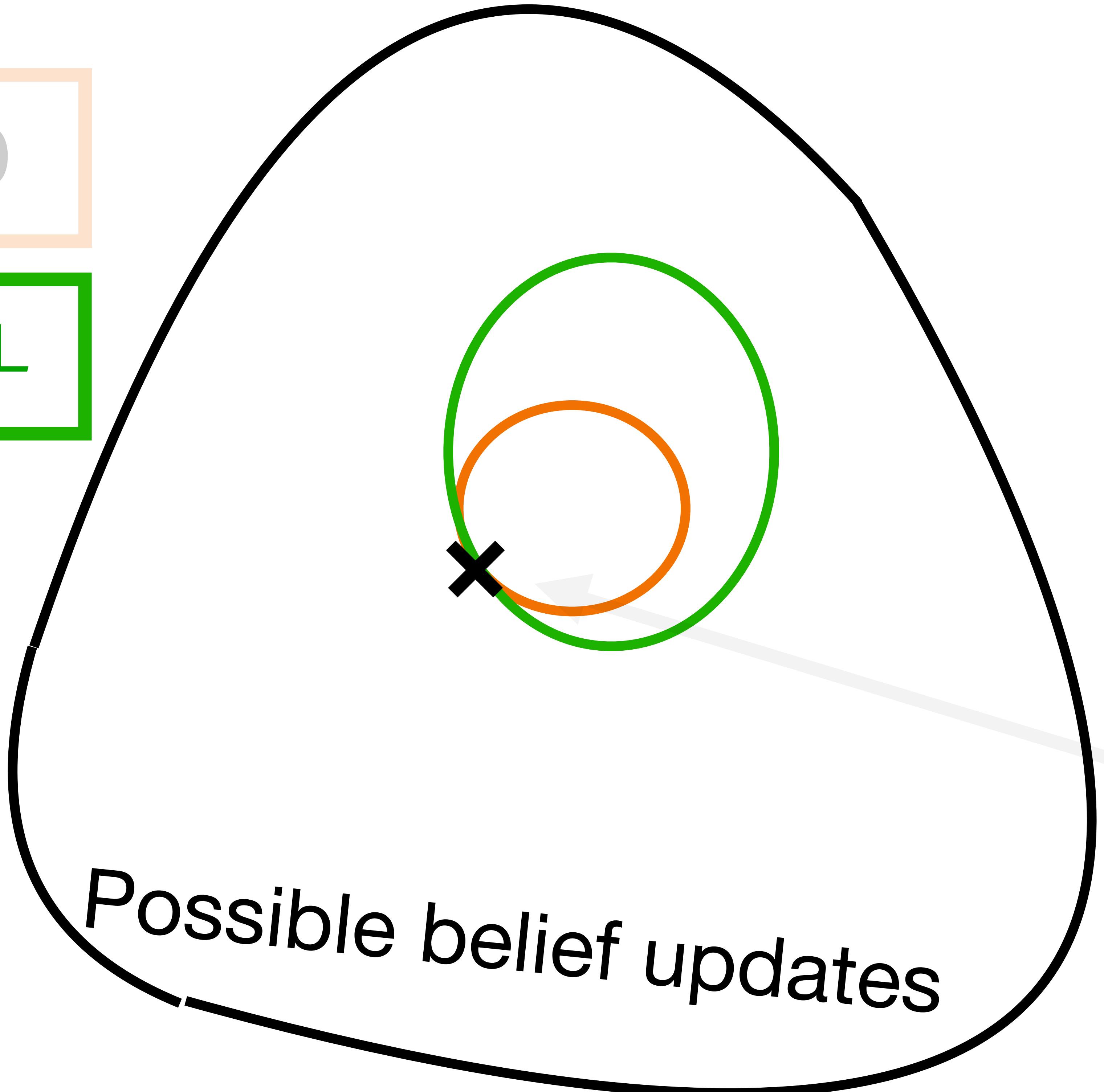
~~(A1)~~, (A2), (A3)

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

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

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

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

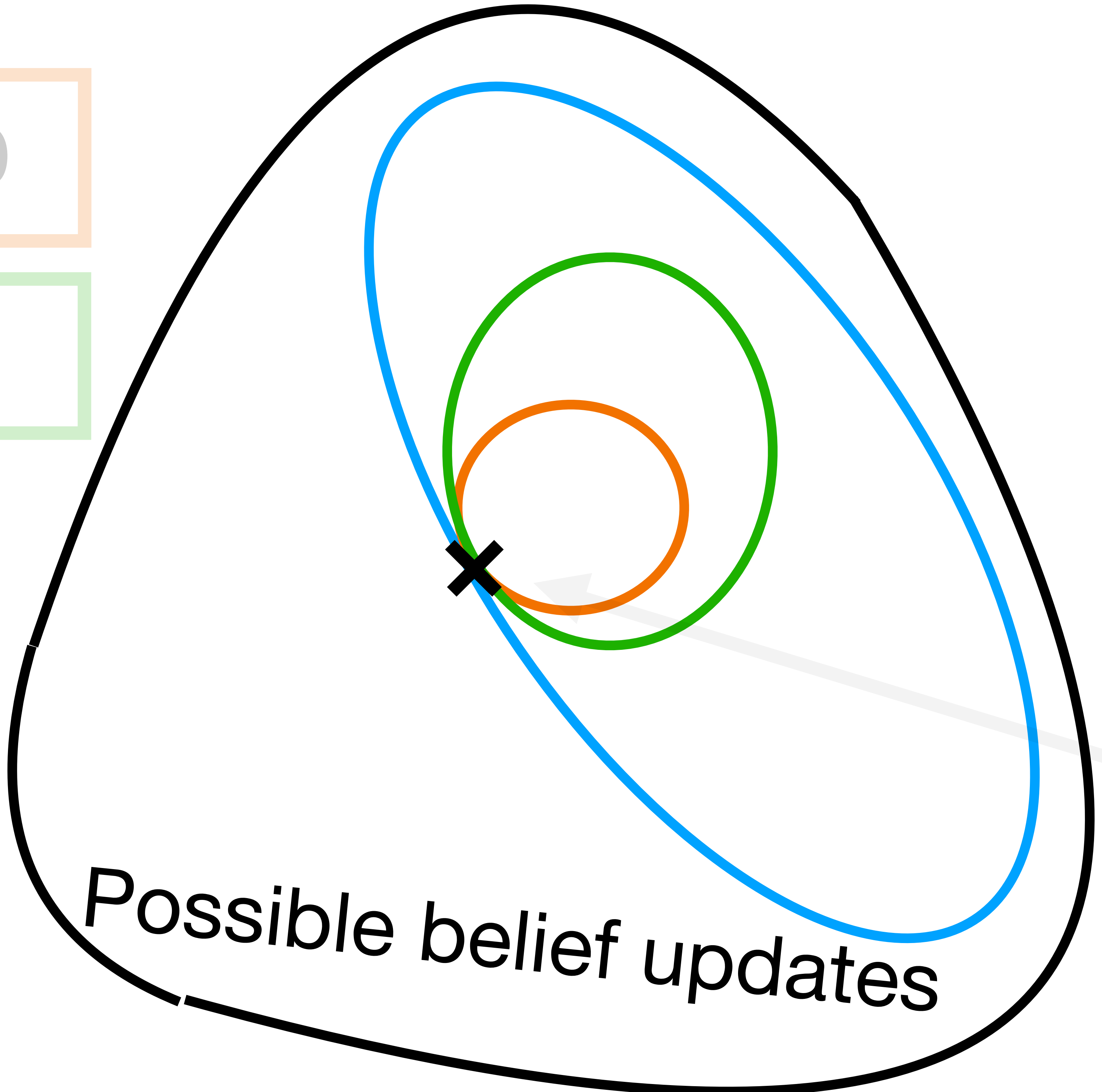
$$\mathcal{Q} \subseteq \mathcal{P}(\Theta)$$

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

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

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

(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 Beliefs

Optimisation-centric posteriors /
Generalised Variational Inference

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

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

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

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

Rest of talk: particular instantiation

Predictively Oriented (PrO) posterior

[See McLatchie, Cherieff-Abdelatiff, Frazier, & Knoblauch [2025]]

Bayes Posterior Predictives

$$P_{\pi_n}(x_{\text{test}}) = \int p_\theta(x_{\text{test}}) \underbrace{\mathrm{d}\pi_n(\theta \mid x_{1:n})}_{\text{Bayes posterior}}$$

Bayes posterior predictive

Bayes Posterior Predictives

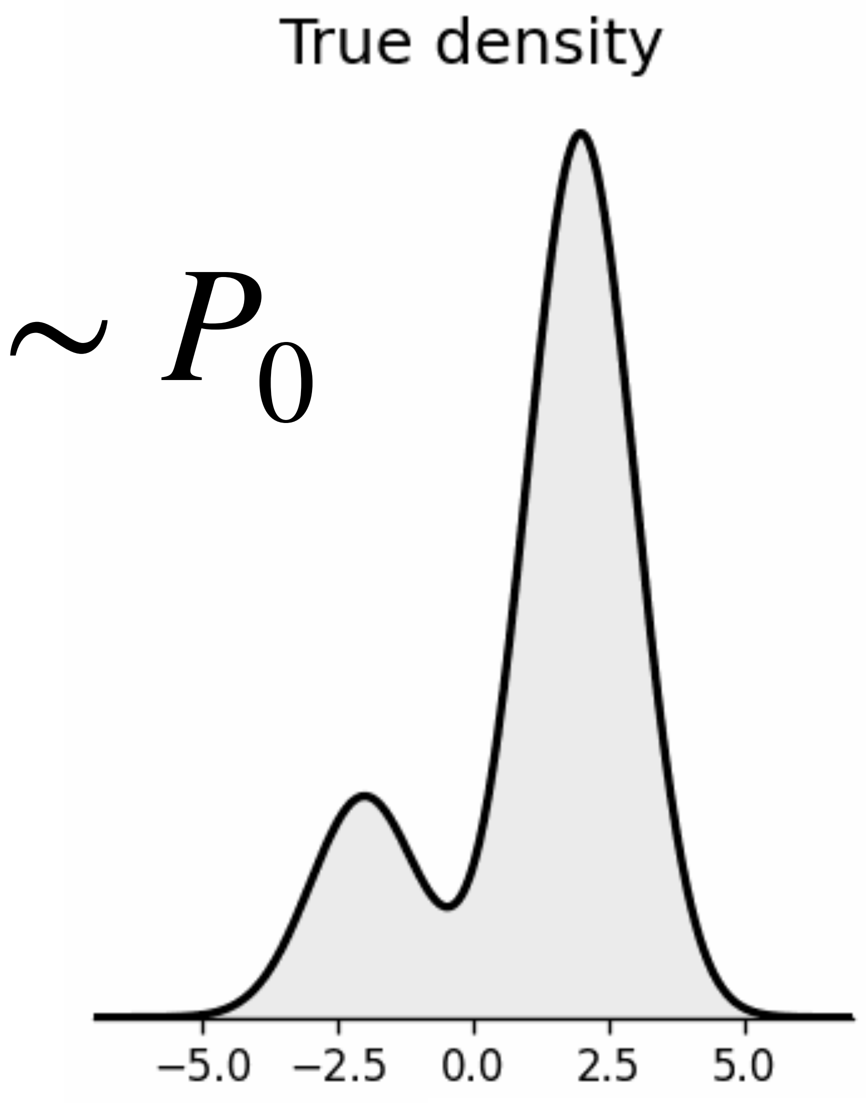
$$P_{\pi_n}(x_{\text{test}}) = \int p_\theta(x_{\text{test}})\, \underbrace{\mathrm{d}\pi_n(\theta \mid x_{1:n})}_{\text{Bayes posterior}}$$

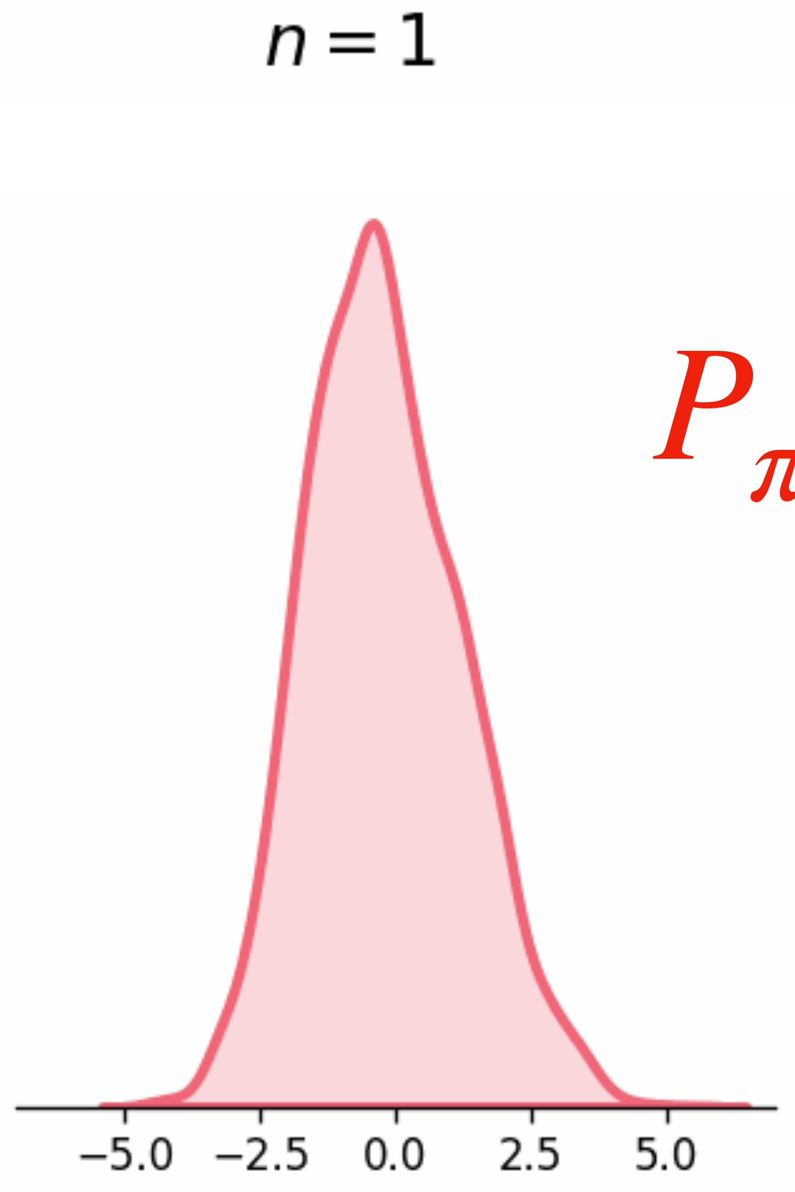
Bayes posterior predictive

Issue: Bayes posterior predictives fragile under model misspecification

(See e.g. Grünwald (2012) or Grünwald & van Ommen, 2017)

$X_i \sim P_0$





$P_{\pi_n}(x_{\text{test}})$

$$p_\theta(x) = \mathcal{N}(x; \theta_1, \theta_2)$$

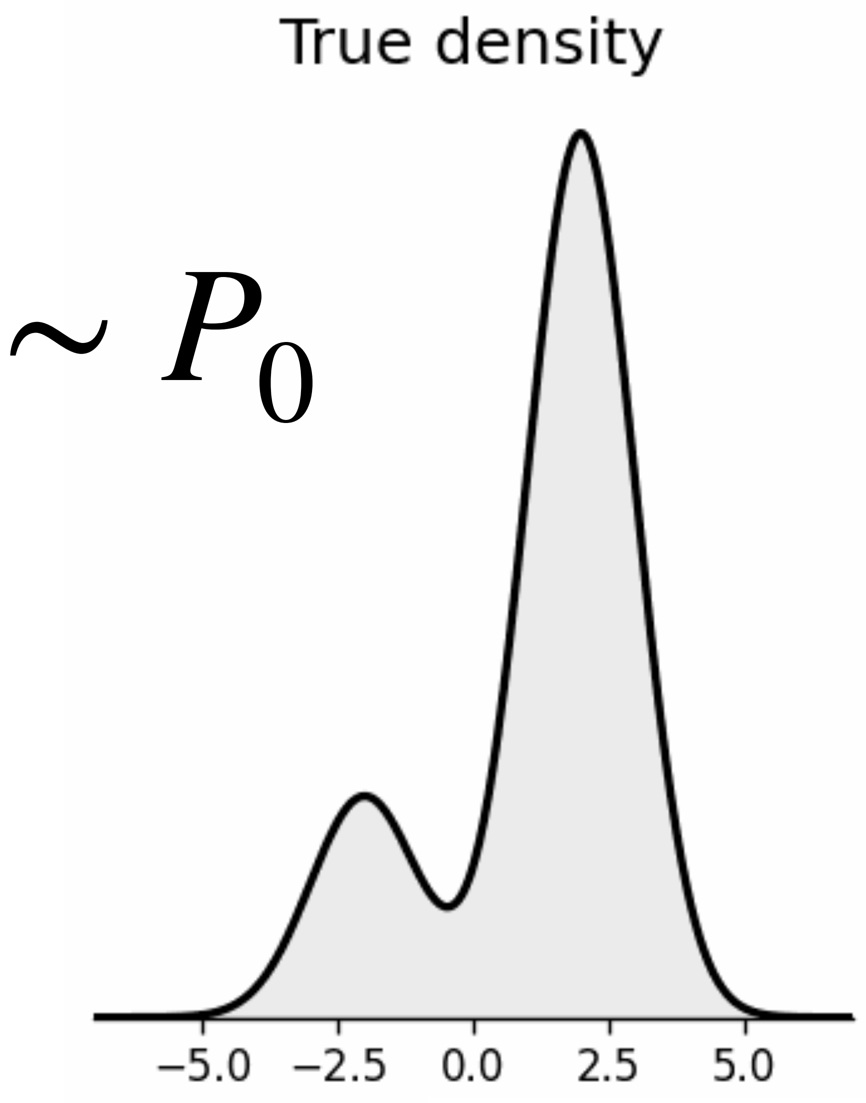
Bayes Posterior Predictives

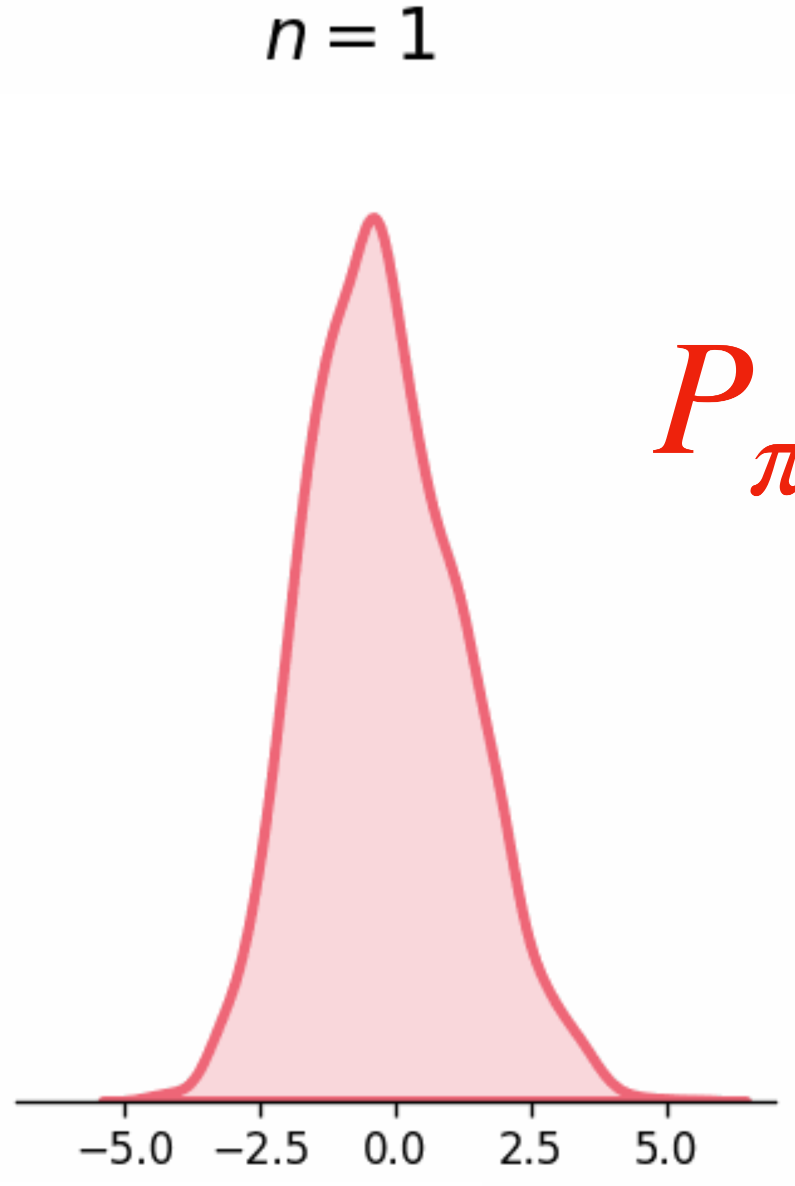
What we observe:

$$P_{\pi_n}(x_{\text{test}}) = \int p_\theta(x_{\text{test}})\, \mathrm{d}\pi_n(\theta \mid x_{1:n}) \xrightarrow[\text{(typically at rate } \sqrt{n})]{n \to \infty} p_{\hat{\theta}_n}(x_{\text{test}})$$

Plug-in predictive

$X_i \sim P_0$


True density


$n = 1$

$P_{\pi_n}(x_{\text{test}})$

$$p_\theta(x) = \mathcal{N}(x; \theta_1, \theta_2)$$

Bayes Posterior Predictives

Posterior Collapse

$$\boxed{\pi_n(\theta \mid x_{1:n}) \xrightarrow[\text{(typically at rate } \sqrt{n}\text{)}]{n \to \infty} \delta_{\hat{\theta}_n}}$$

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} - \log p_\theta(x_{1:n})$$

(= MLE)

$$P_{\pi_n}(x_{\text{test}}) = \int p_\theta(x_{\text{test}})\, \mathrm{d}\pi_n(\theta \mid x_{1:n}) \xrightarrow[\text{(typically at rate } \sqrt{n}\text{)}]{n \to \infty} p_{\hat{\theta}_n}(x_{\text{test}})$$

Plug-in predictive

McLatchie, Fong, Frazier, & K. (2025); *Biometrika* **(cf. Lemma 1)**

Bayes Posterior Predictives

Posterior Collapse

$$\boxed{\pi_n(\theta \mid x_{1:n}) \xrightarrow[\text{(typically at rate } \sqrt{n})]{n \to \infty} \delta_{\hat{\theta}_n}}$$

$\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} - \log p_\theta(x_{1:n})$

(= MLE)

$$P_{\pi_n}(x_{\text{test}}) = \int p_\theta(x_{\text{test}})\, \mathrm{d}\pi_n(\theta \mid x_{1:n}) \xrightarrow[\text{(typically at rate } \sqrt{n})]{n \to \infty} p_{\hat{\theta}_n}(x_{\text{test}})$$

Plug-in predictive

Note 1: makes sense if model correct! No uncertainty left if all data seen!

Note 2: makes less/little sense if model wrong! (Irreducible uncertainty!)

(Note 3: Same behaviour for power posteriors and Gibbs posteriors)

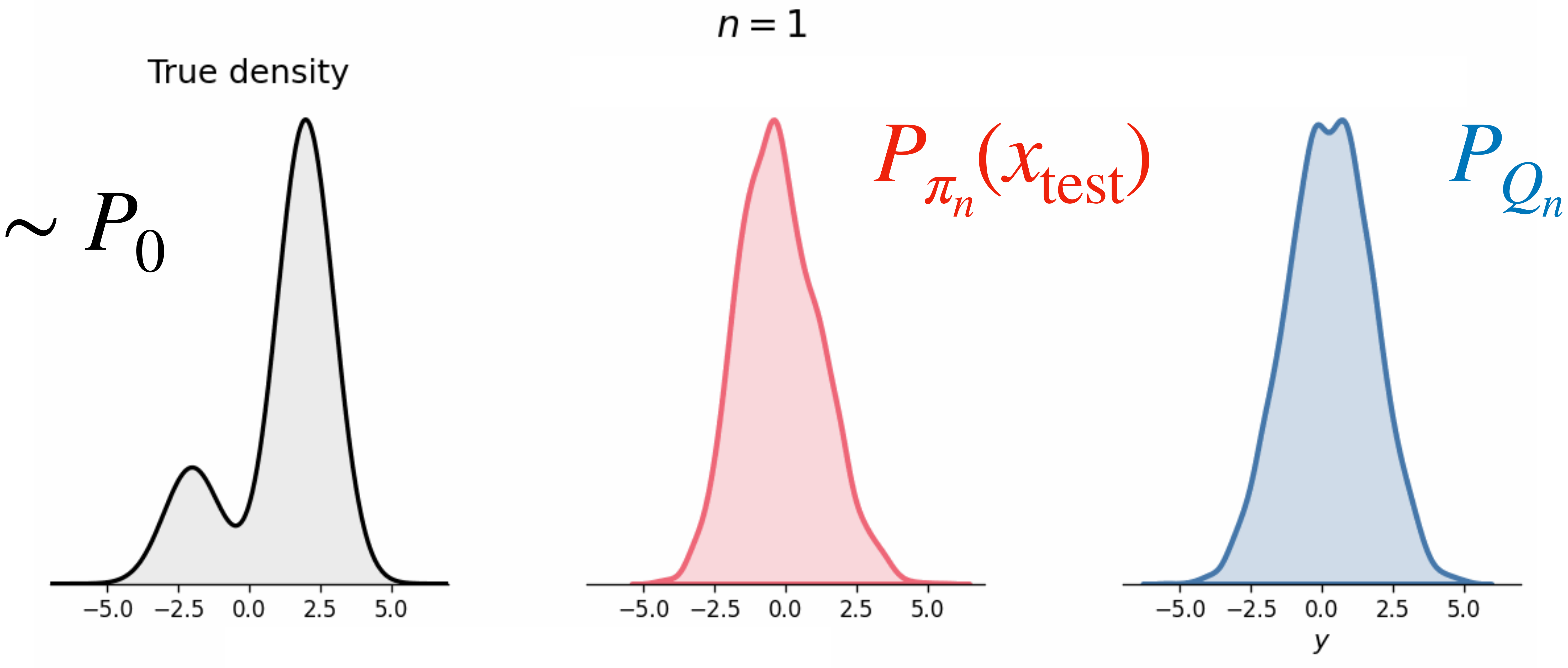
$\pi_n^{(\lambda)}(\theta \mid x_{1:n})$ $\pi_n^{\mathsf{L}}(\theta \mid x_{1:n})$

McLatchie, Fong, Frazier, & K. (2025); *Biometrika* **(cf. Lemma 1)**

Posterior predictives

$$p_\theta(x) = \mathcal{N}(x; \theta_1, \theta_2)$$

$X_i \sim P_0$



$P_{\pi_n}(x_{\text{test}})$

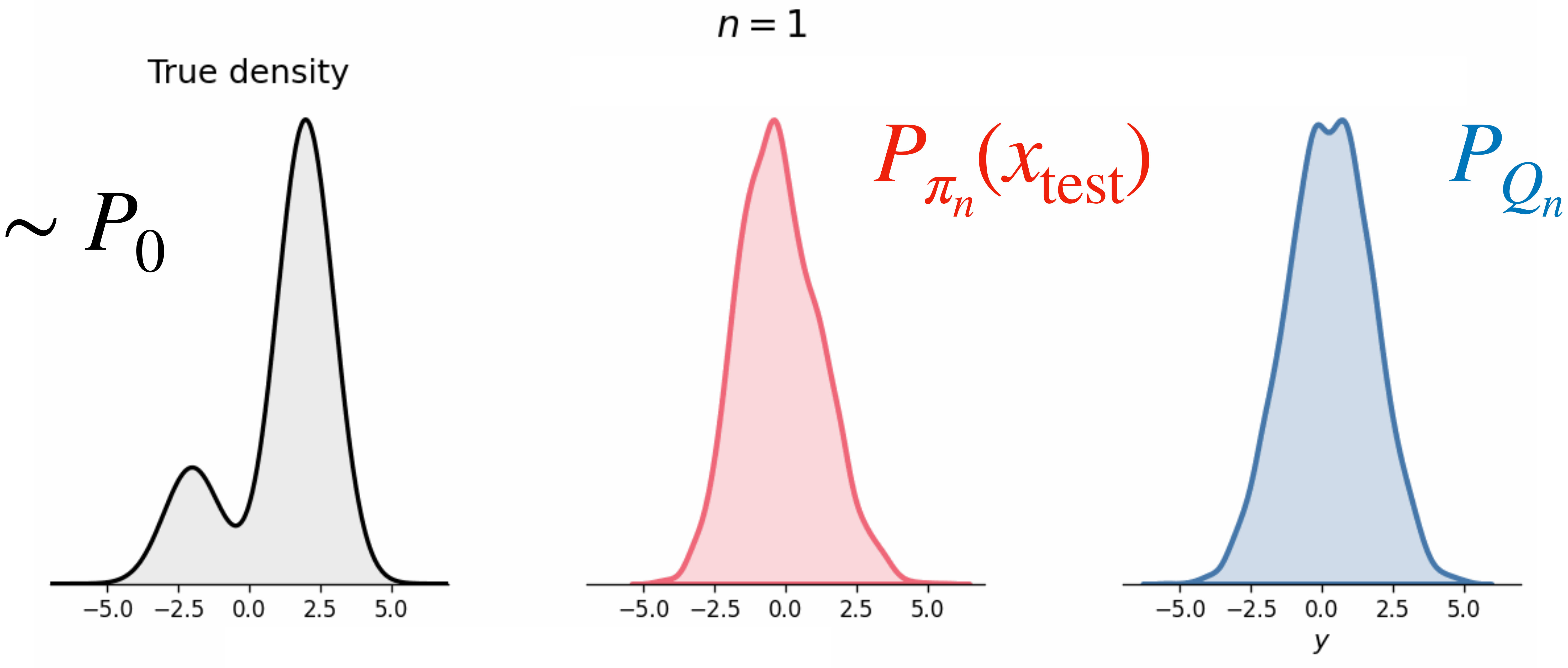
$P_{Q_n}(x_{\text{test}})$

Question: Can we construct better predictives without changing the model?

Posterior predictives

$$p_\theta(x) = \mathcal{N}(x; \theta_1, \theta_2)$$

$$X_i \sim P_0$$



$P_{\pi_n}(x_{\text{test}})$

$P_{Q_n}(x_{\text{test}})$

Question: Can we construct better predictives without changing the model?

$$P_{\boxed{Q_n}}(x_{\text{test}}) = \int p_\theta(x_{\text{test}})\, \underbrace{\mathrm{d}\boxed{Q_n(\theta \mid x_{1:n})}}_{\text{Predictively Oriented (PrO) posterior}}$$

PrO posterior predictive

Constructing Predictively Oriented (PrO) Posteriors

$$\arg\min_{Q\in\mathscr{P}(\Theta)} \left\{ \underbrace{\int -\log p_\theta(x_{1:n})\, \mathrm{d}Q(\theta)}_{\text{Data-fitting}} + \underbrace{\mathrm{KL}\big(Q,\pi\big)}_{\text{Prior regularisation}} \right\} = \pi_n(\theta \mid x_{1:n})$$

(= Bayes posterior)

Shen, K., Power, & Oates (2025); *AISTATS* + McLatchie*, Cherrief-Abdellatif*, Frazier*, K.* (2025)*; ArXiv preprint*

Constructing Predictively Oriented (PrO) Posteriors

$\mathrm{d}Q(\theta)$**-Averaged (log) loss of the model**

$$\arg\min_{Q\in\mathscr{P}(\Theta)}\left\{\boxed{\int -\log p_\theta(x_{1:n})\,\mathrm{d}Q(\theta)} + \mathrm{KL}\big(Q,\pi\big)\right\} \quad = \quad \pi_n(\theta \mid x_{1:n})$$

(= Bayes posterior)

Shen, K., Power, & Oates (2025); *AISTATS* + McLatchie*, Cherrief-Abdellatif*, Frazier*, K.* (2025)*; ArXiv preprint*

Constructing Predictively Oriented (PrO) Posteriors

$\mathrm{d}Q(\theta)$ **-Averaged (log) loss of the model**

$$\arg\min_{Q\in\mathscr{P}(\Theta)} \left\{ \boxed{\int -\log p_\theta(x_{1:n})\,\mathrm{d}Q(\theta)} + \mathrm{KL}(Q,\pi) \right\} = \pi_n(\theta \mid x_{1:n})$$

(= Bayes posterior)

(log) loss of the $\mathrm{d}Q(\theta)$ **-averaged model**

$$\arg\min_{Q\in\mathscr{P}(\Theta)} \left\{ \boxed{-\log\left(\underbrace{\int p_\theta(x_{1:n})\,\mathrm{d}Q(\theta)}_{=\ \textbf{predictive}\ P_Q(x_{\text{test}})}\right)} + \mathrm{KL}(Q,\pi) \right\} = Q_n(\theta)$$

(= PrO posterior)

Shen, K., Power, & Oates (2025); *AISTATS* + McLatchie*, Cherrief-Abdellatif*, Frazier*, K.* (2025)*; forthcoming*

Constructing Predictively Oriented (PrO) Posteriors

$\mathrm{d}Q(\theta)$**-Averaged (log) loss of the model**

$$\underset{Q\in\mathscr{P}(\Theta)}{\arg\ \min}\left\{\boxed{\int -\log p_\theta(x_{1:n})\,\mathrm{d}Q(\theta)} + \mathrm{KL}(Q,\pi)\right\} = \pi_n(\theta \mid x_{1:n})$$

(= Bayes posterior)

(log) loss of the $\mathrm{d}Q(\theta)$**-averaged model**

$$\underset{Q\in\mathscr{P}(\Theta)}{\arg\ \min}\left\{\boxed{-\log\left(\underbrace{\int p_\theta(x_{1:n})\,\mathrm{d}Q(\theta)}_{}\right)} + \mathrm{KL}(Q,\pi)\right\} = Q_n(\theta)$$

= predictive $P_Q(x_{\text{test}})$

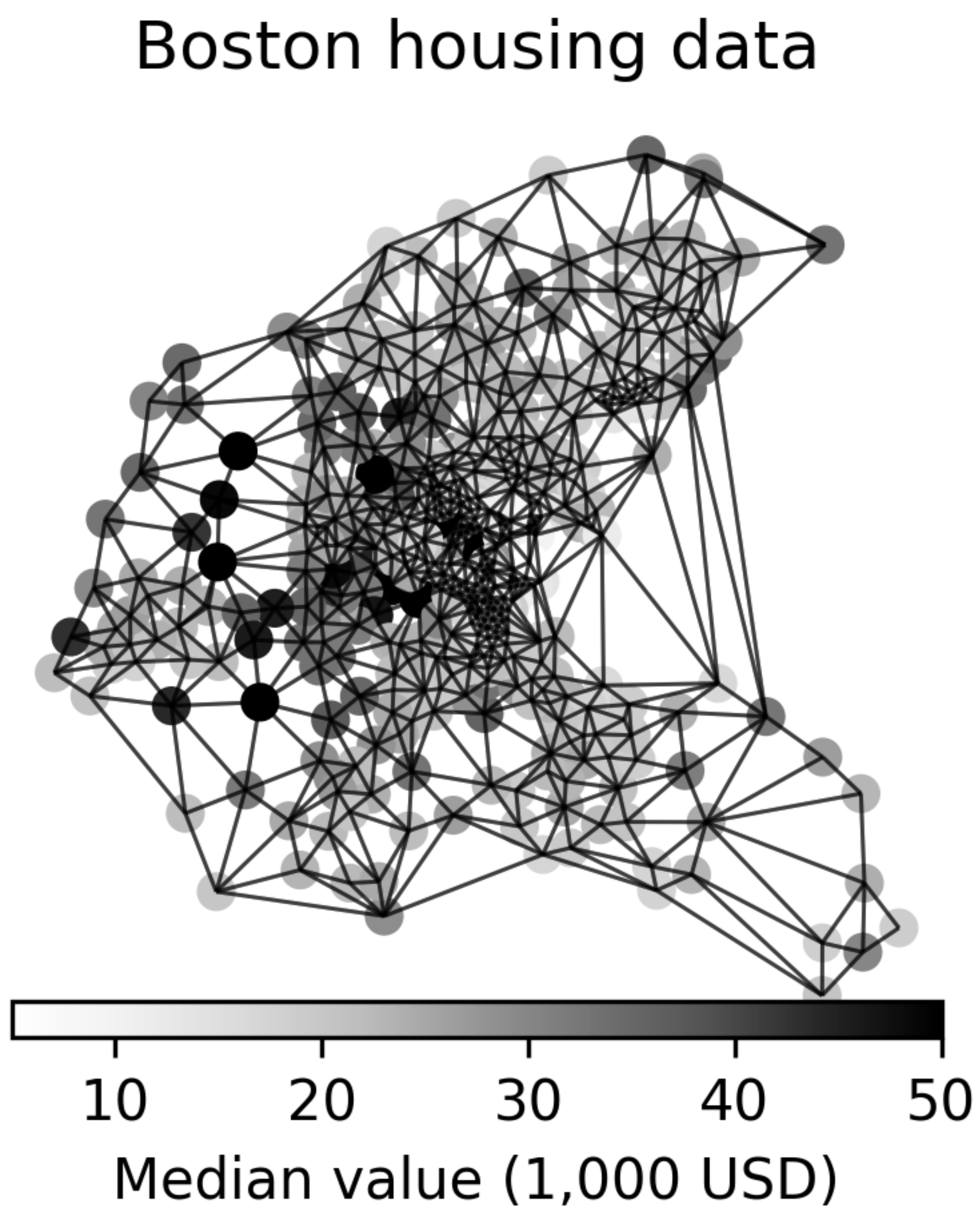
(= PrO posterior)

Note 1: Stated with log score here; but can use other scoring rules too

Note 2: Some questionable shortcuts taken here to present key idea

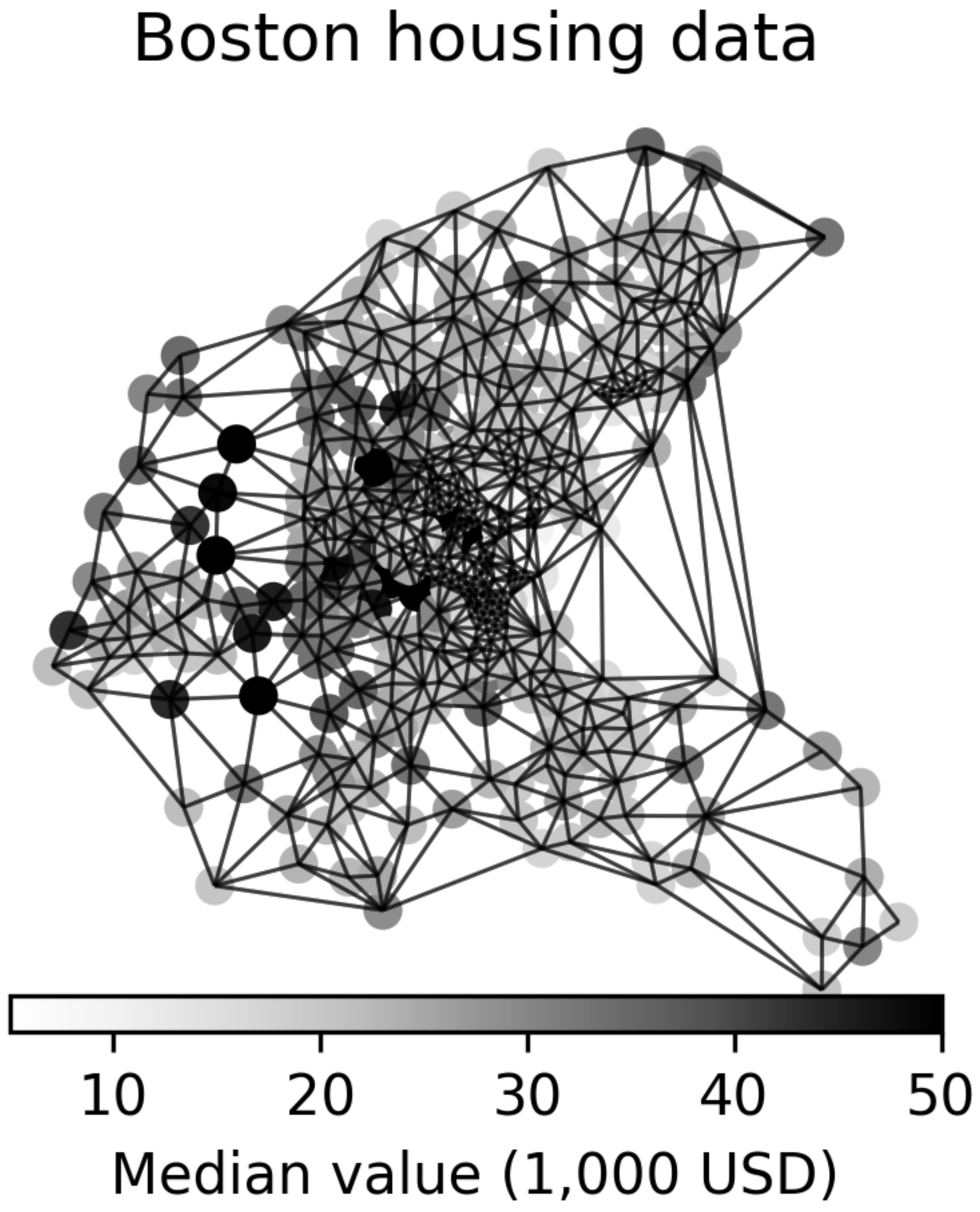
Shen, K., Power, & Oates (2025); *AISTATS* + McLatchie*, Cherrief-Abdellatif*, Frazier*, K.* (2025)*; forthcoming*

Building Intuition with the Boston Housing Data



Median house price for $n = 205$
tracts in Boston in the 1970/80s.

Building Intuition with the Boston Housing Data



Median house price for $n = 205$ tracts in Boston in the 1970/80s.

Model median price per tract as function of its neighbouring tracts:

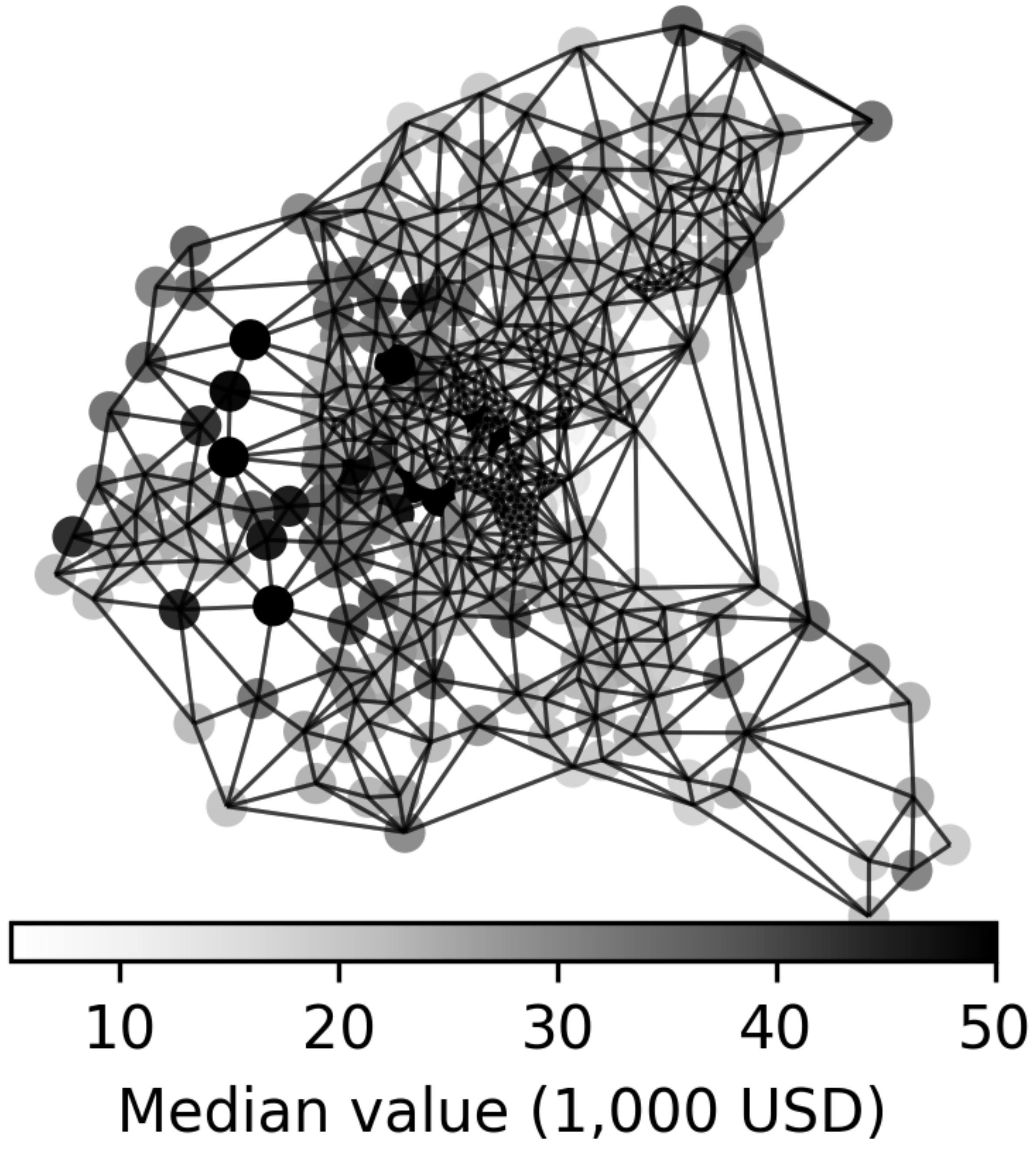
Median price in tract *i*

Influence of neighbouring tracts

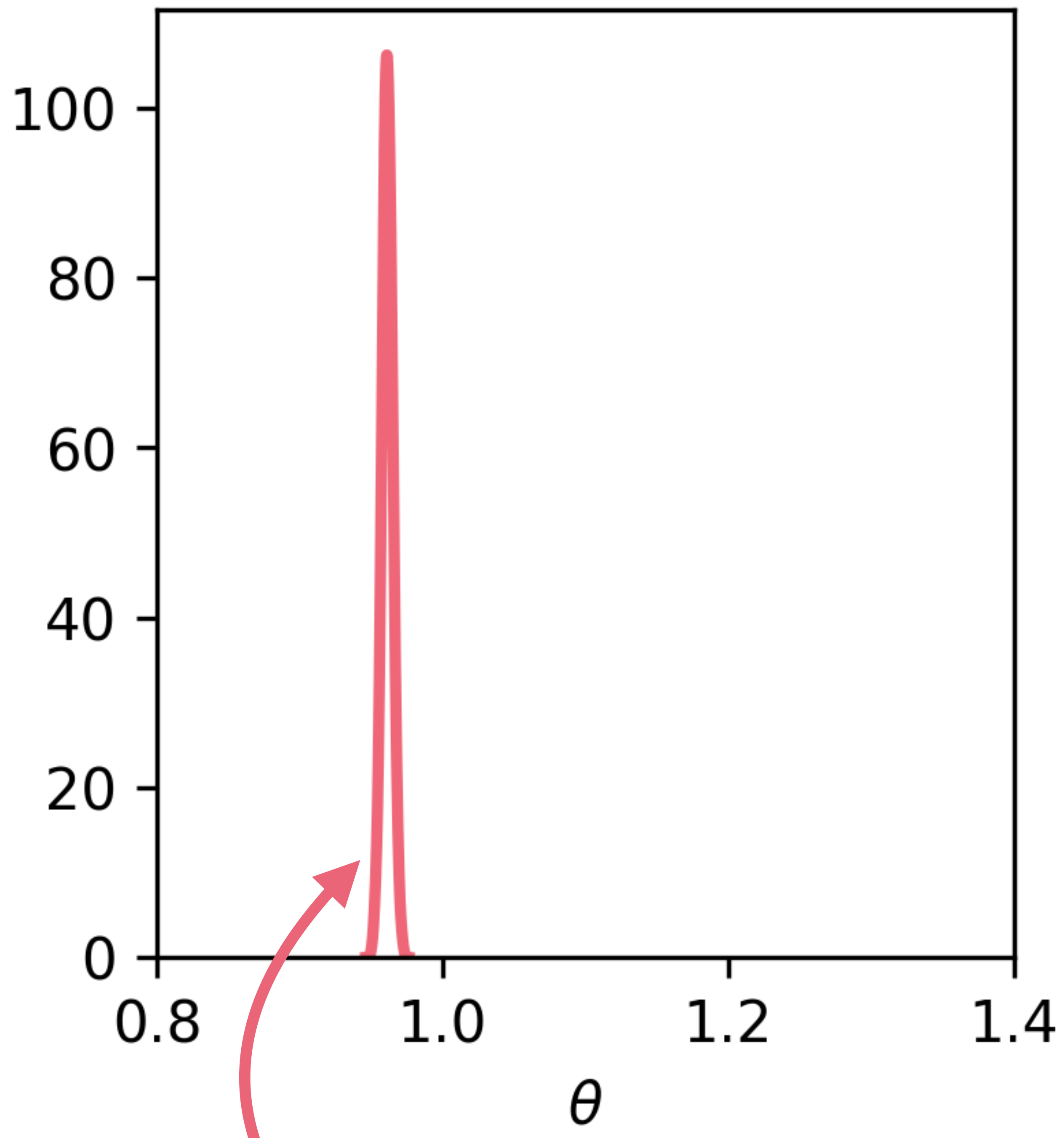
$$y_i \mid y_1, \ldots y_{i-1}, y_{i+1}, \ldots y_n \sim \mathrm{N}\left(\boxed{\theta}\sum_{j \neq i} \tilde{\omega}_{i,j} y_j, \sigma^2\right),$$

1 if tracts neighbouring
0 otherwise

Building Intuition with the Boston Housing Data



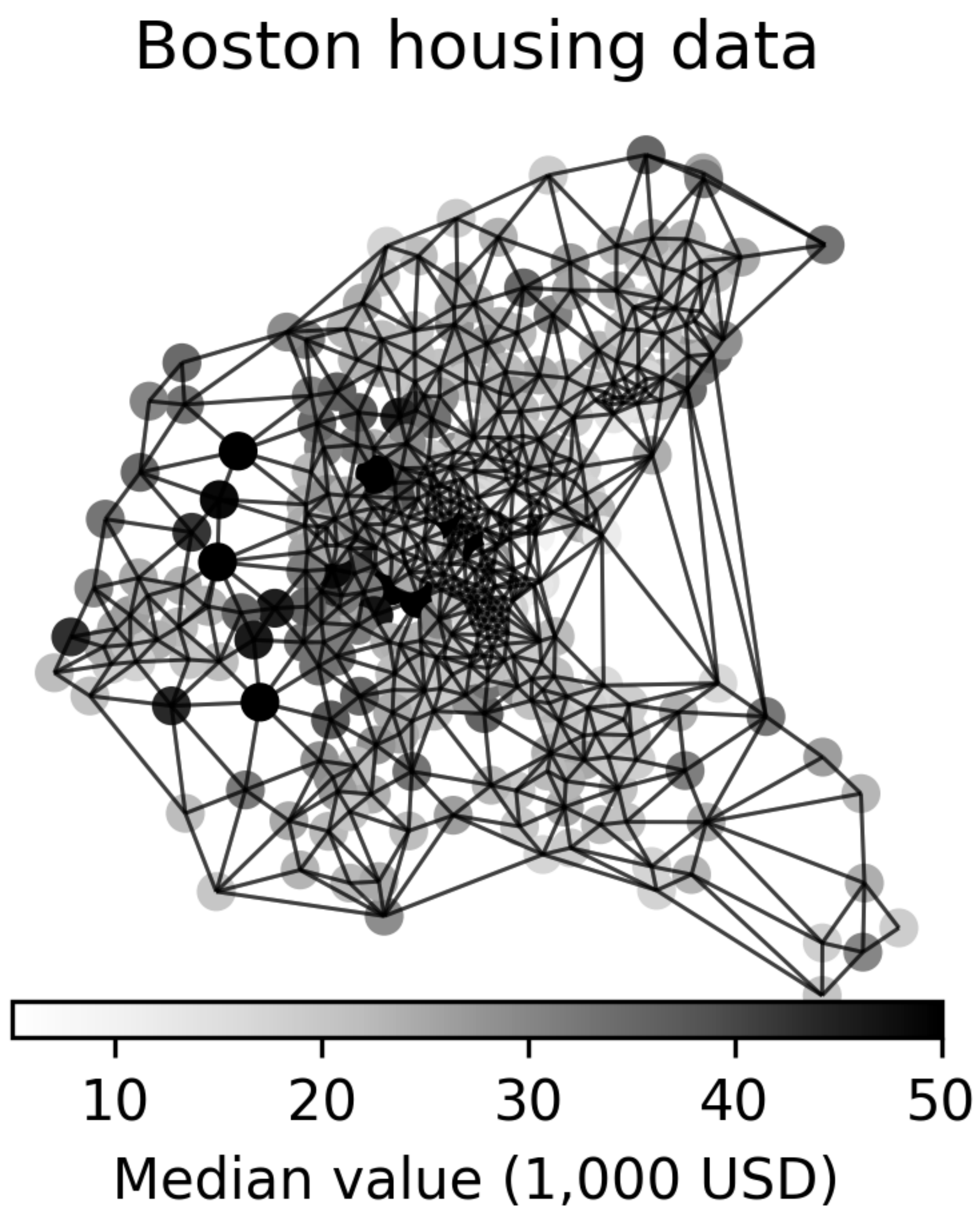
Boston housing data

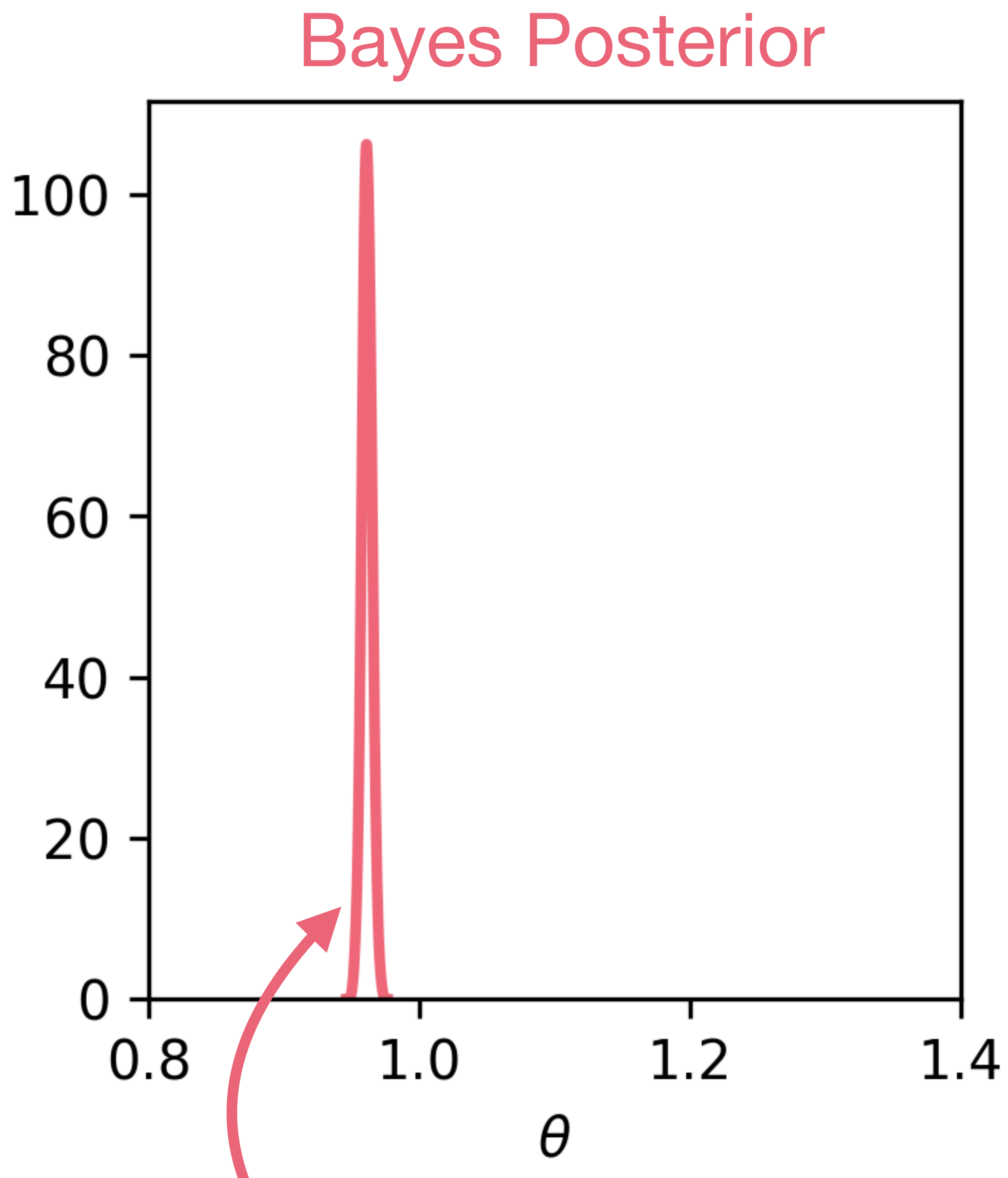


Bayes Posterior

Find the best single member of P_Θ

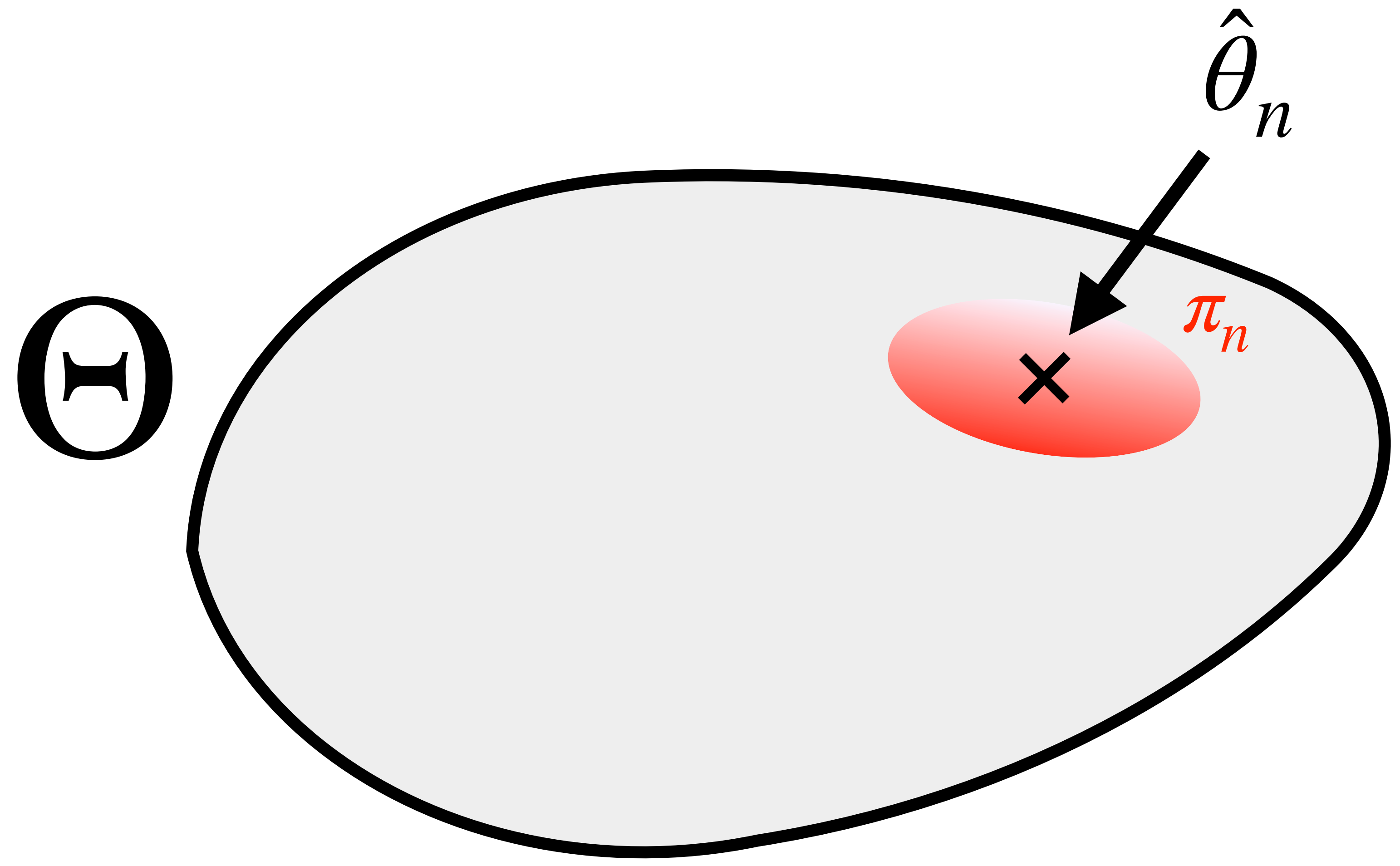
Building Intuition with the Boston Housing Data



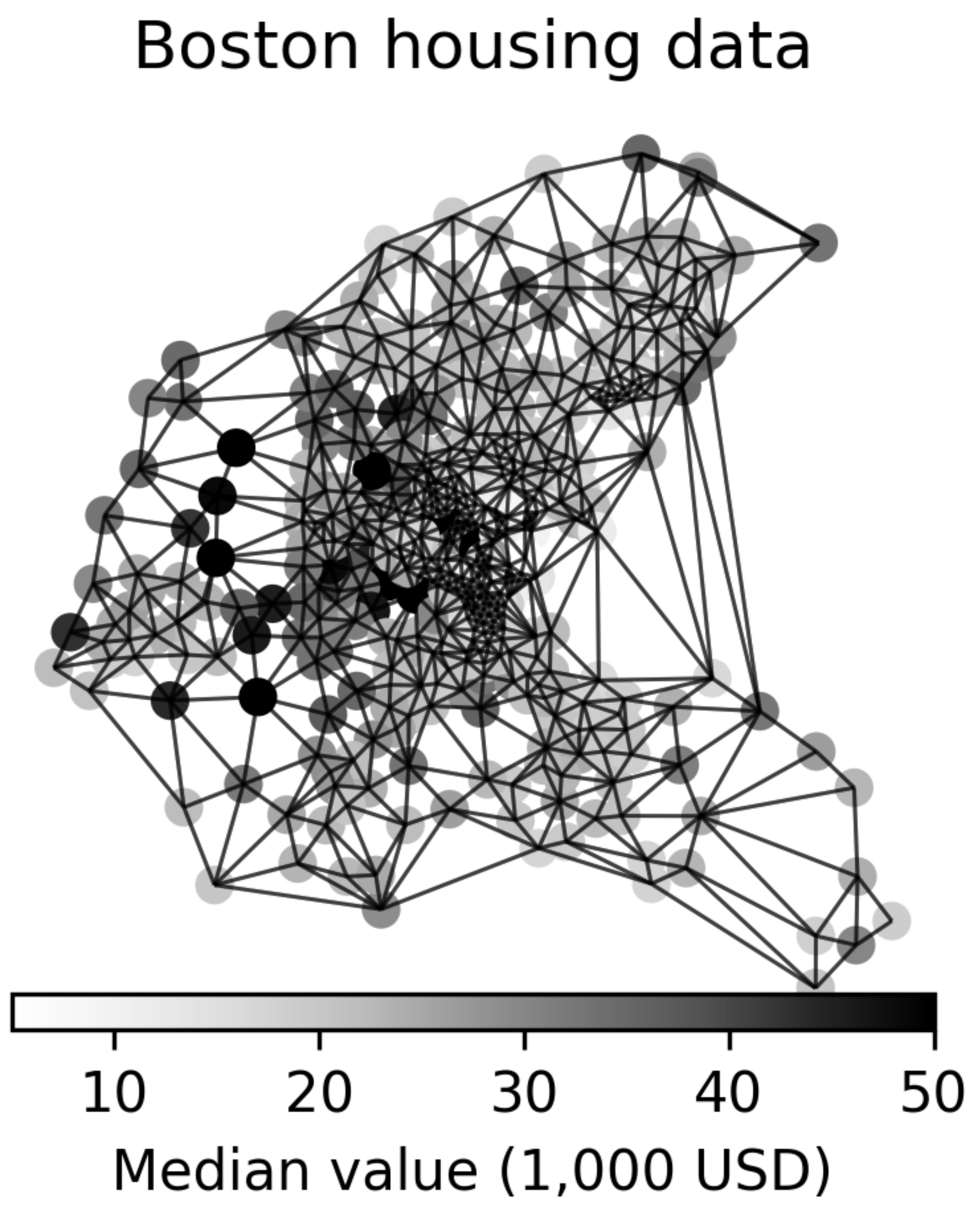


Find the best single member of P_Θ

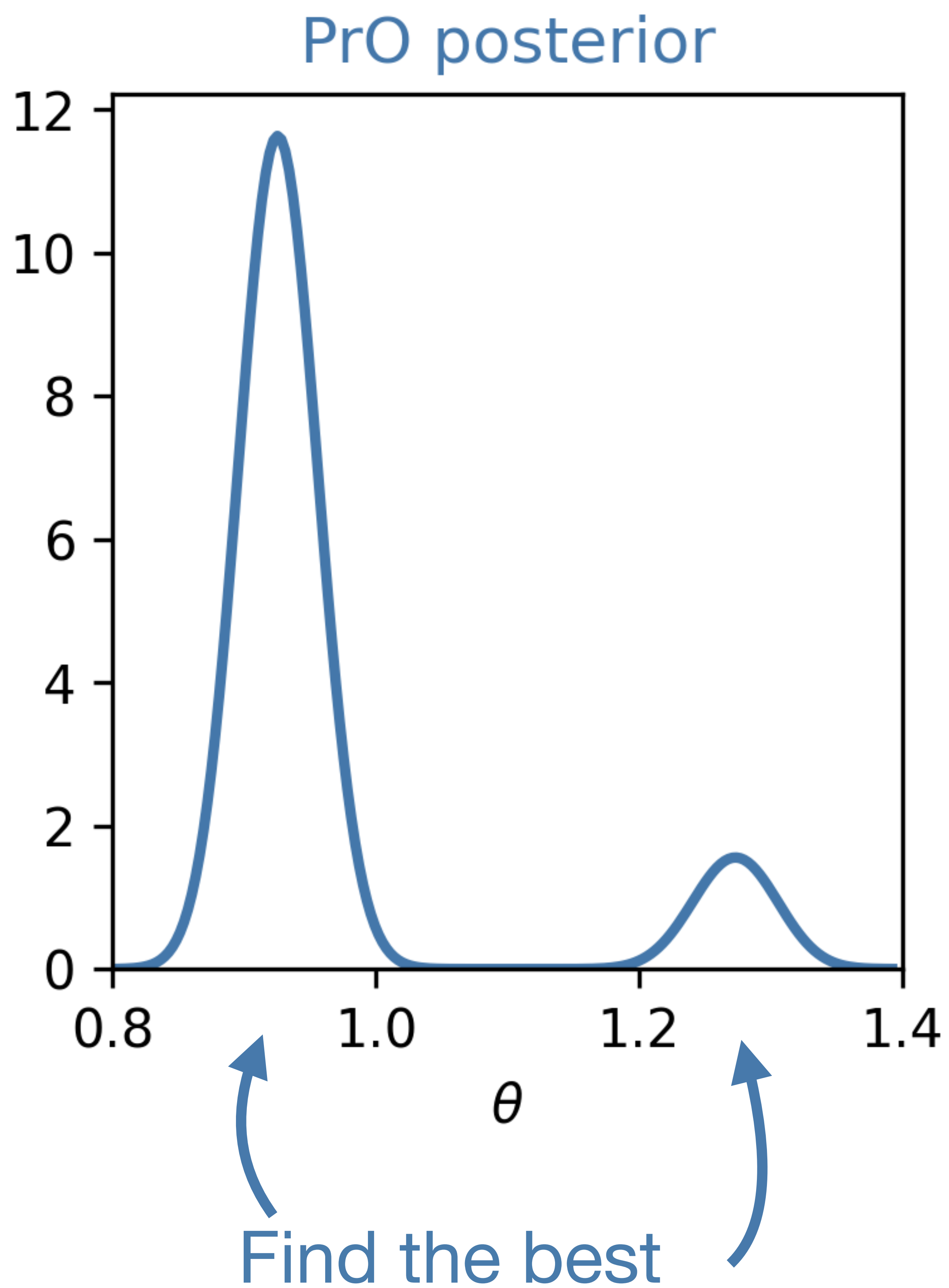
Bayes Posterior Collapses on Parameter Space



Building Intuition with the Boston Housing Data



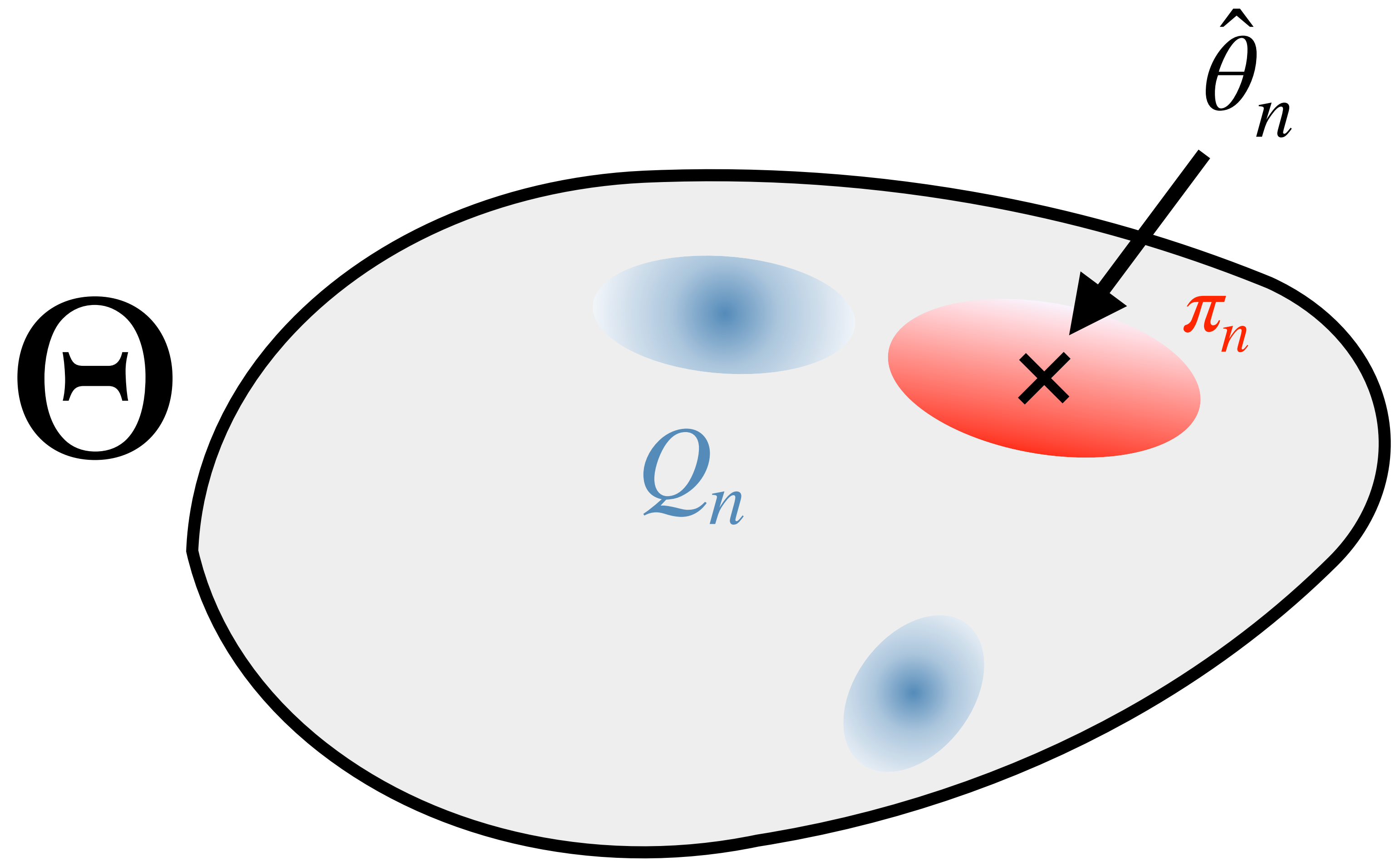
Boston housing data



PrO posterior

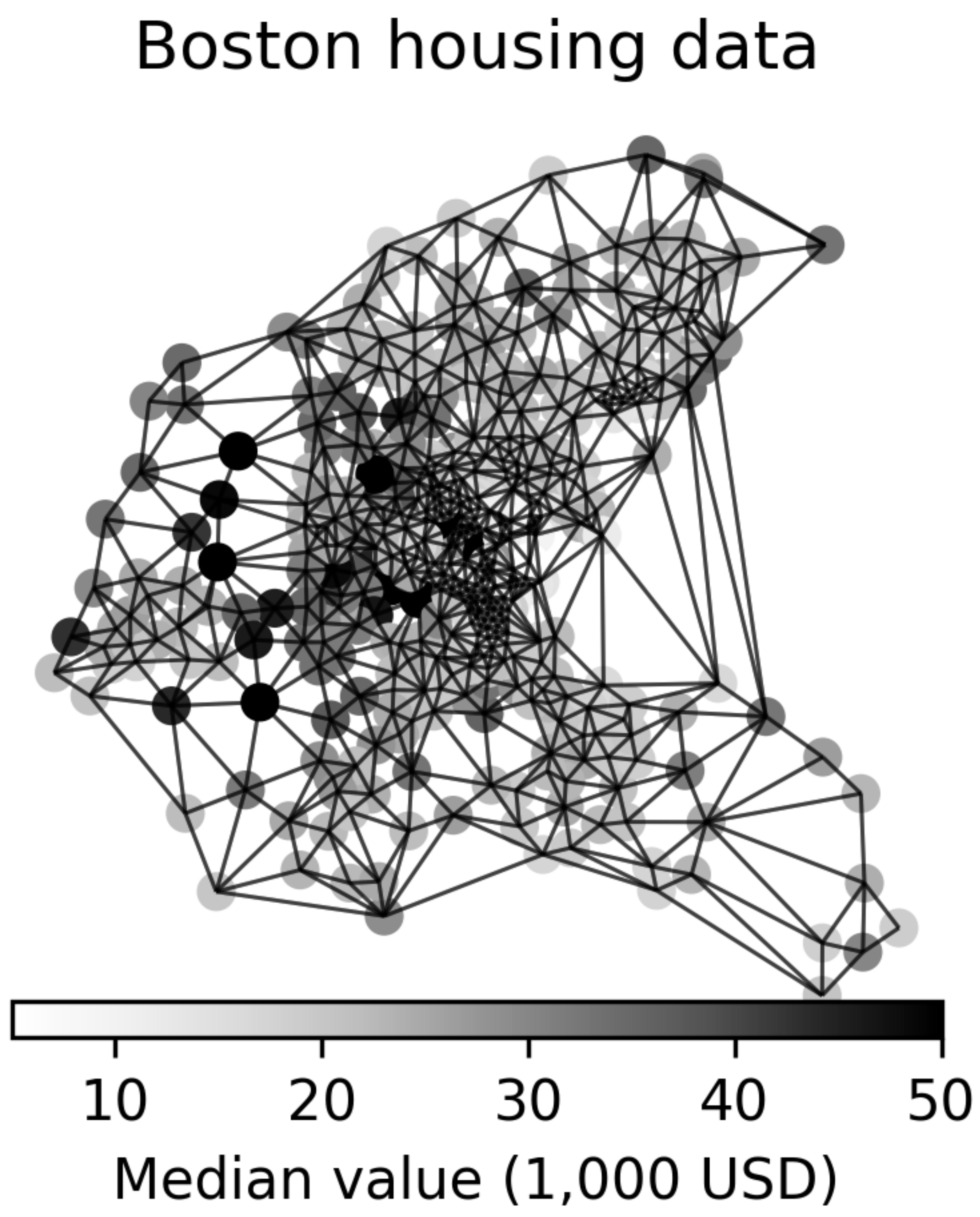
Find the best mixture over P_Θ

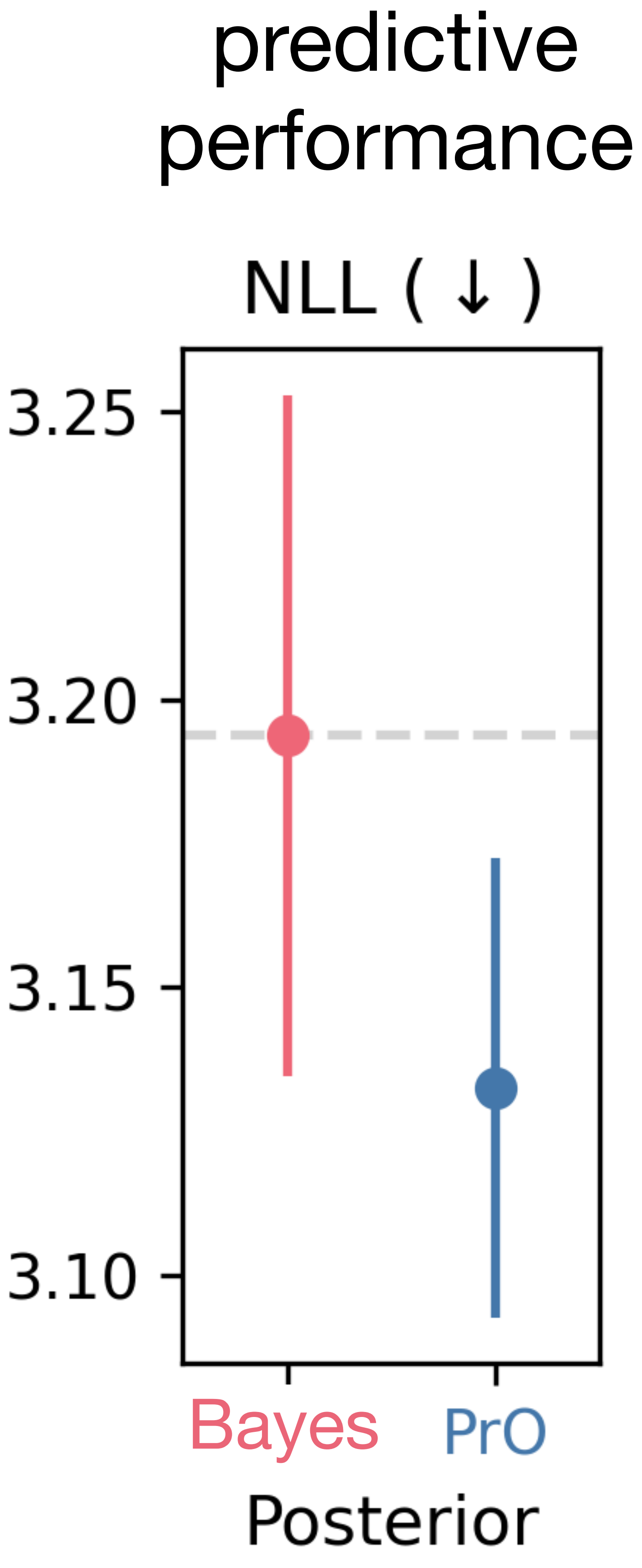
Bayes Posterior Collapses on Parameter Space



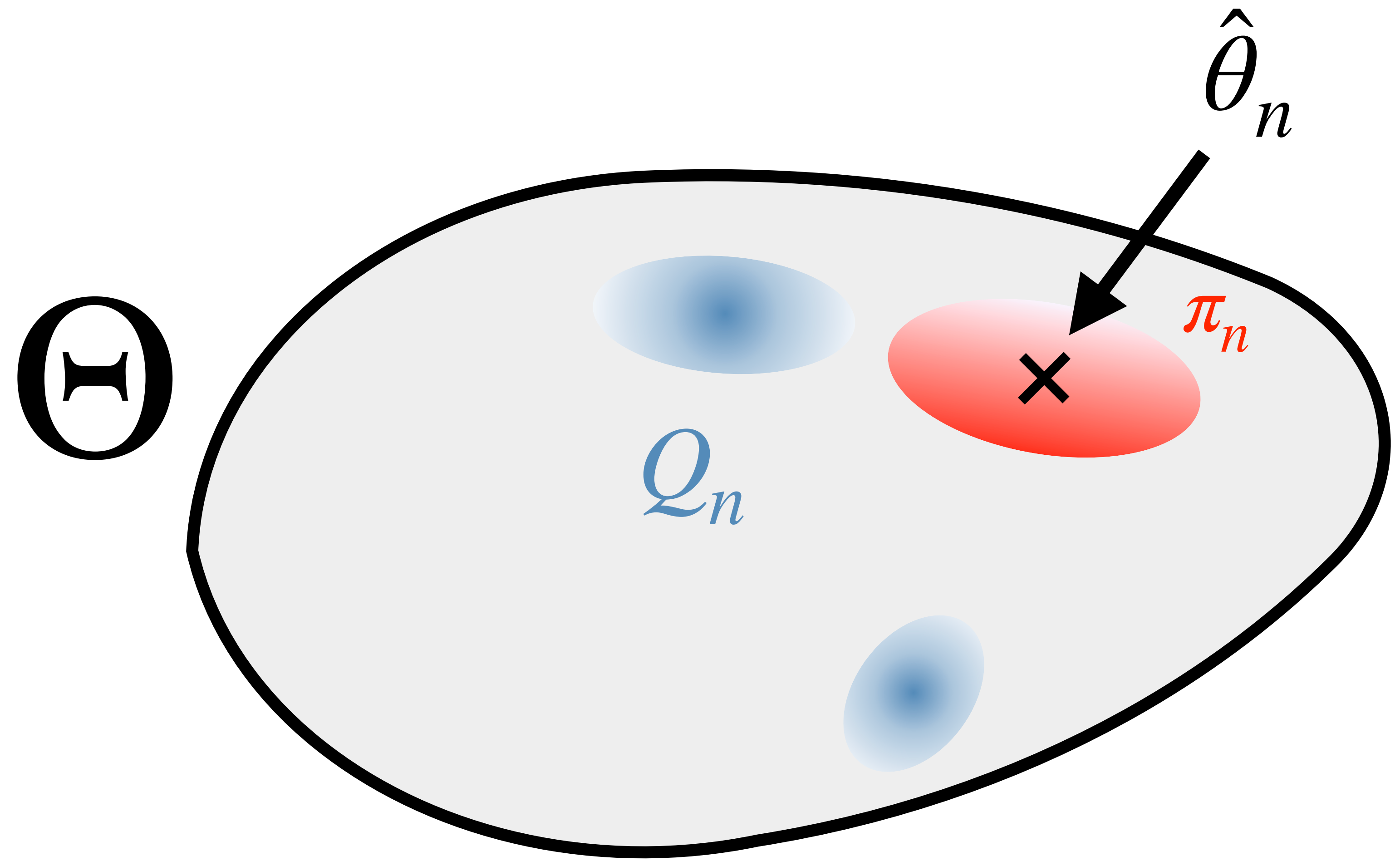
(PrO Posterior does not)

Building Intuition with the Boston Housing Data



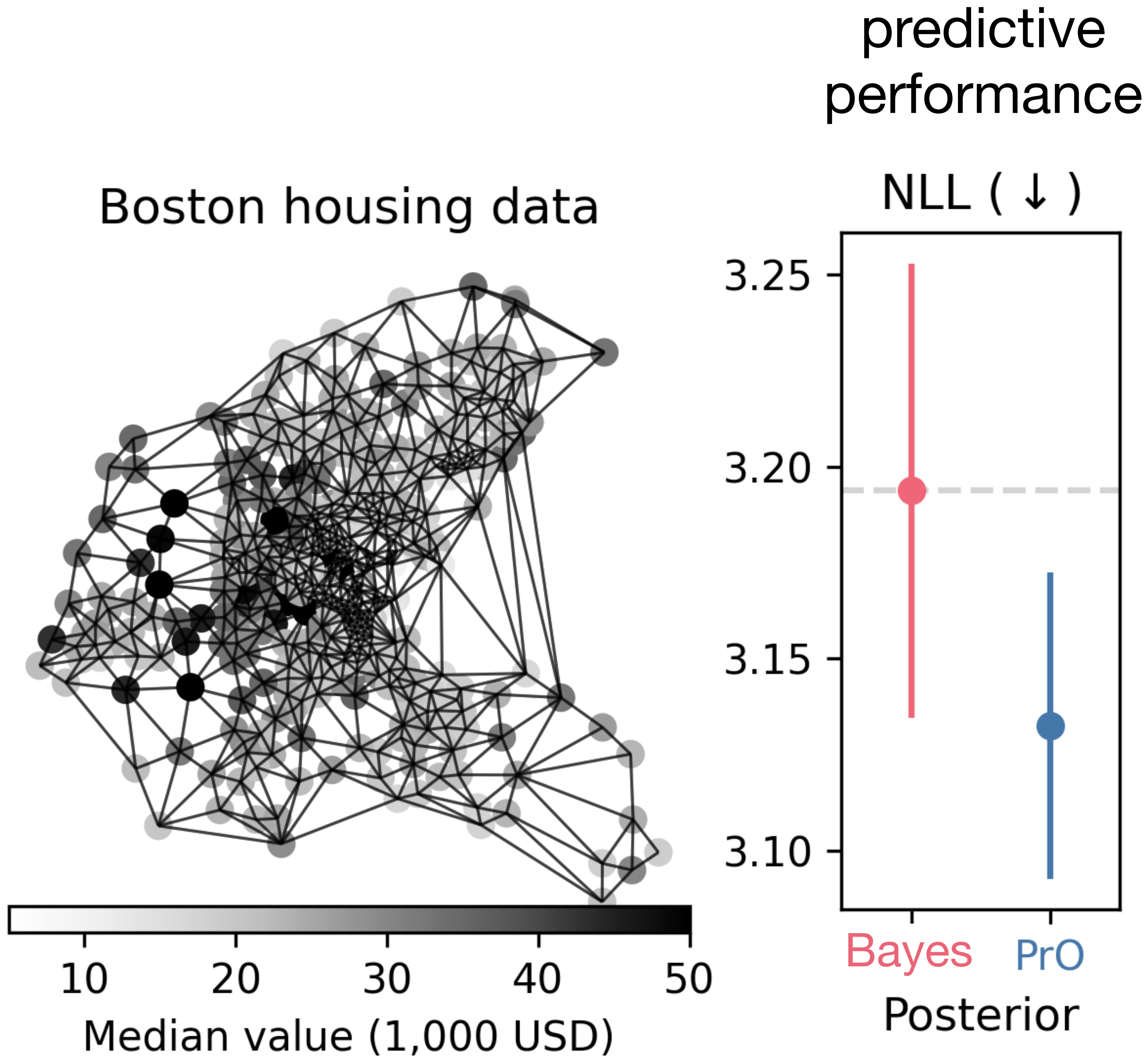


Bayes Posterior Collapses on Parameter Space

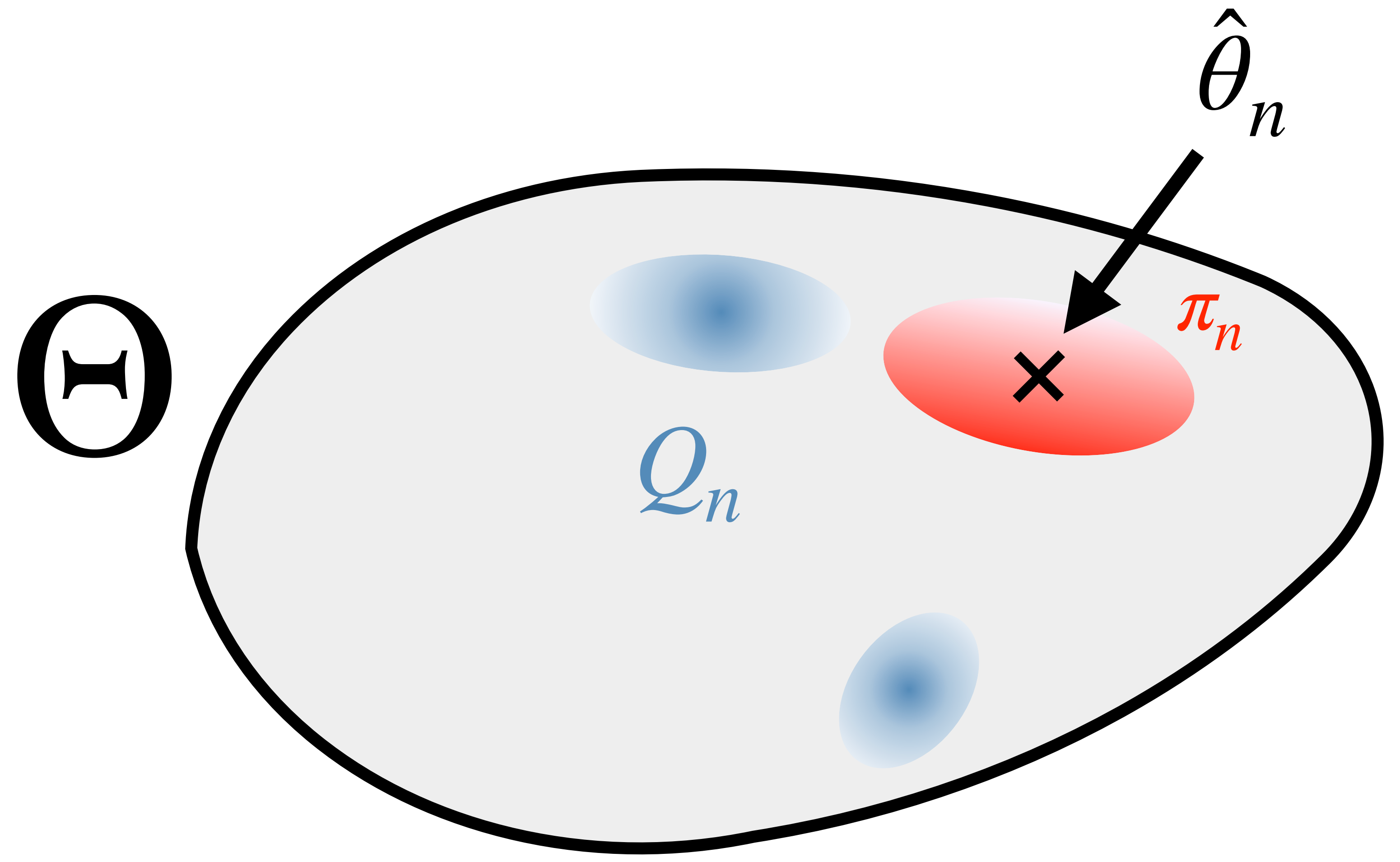


(PrO Posterior does not)

Building Intuition with the Boston Housing Data



Bayes Posterior Collapses on Parameter Space

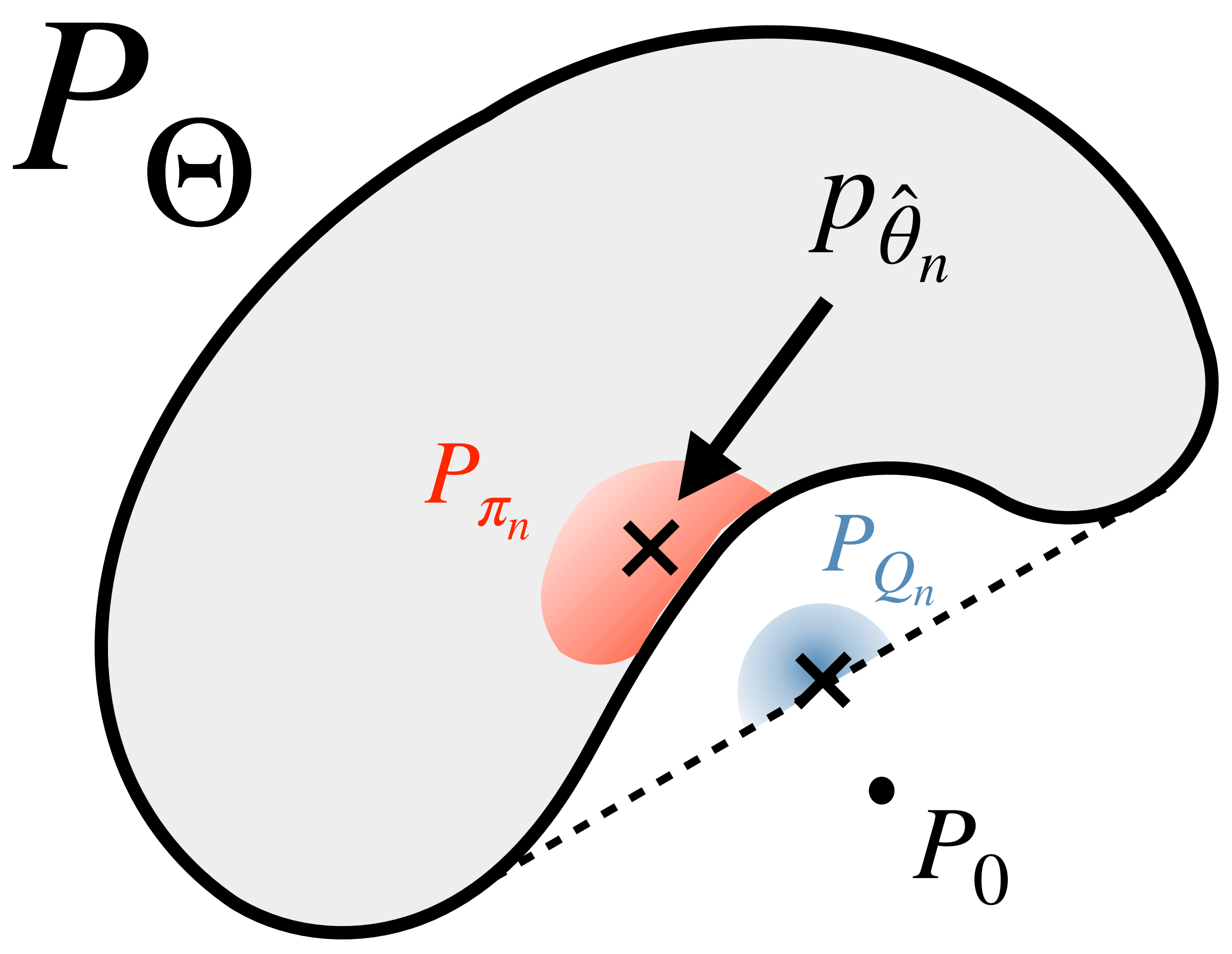


(PrO Posterior does not)

Question: So how do posteriors behave on predictive space?

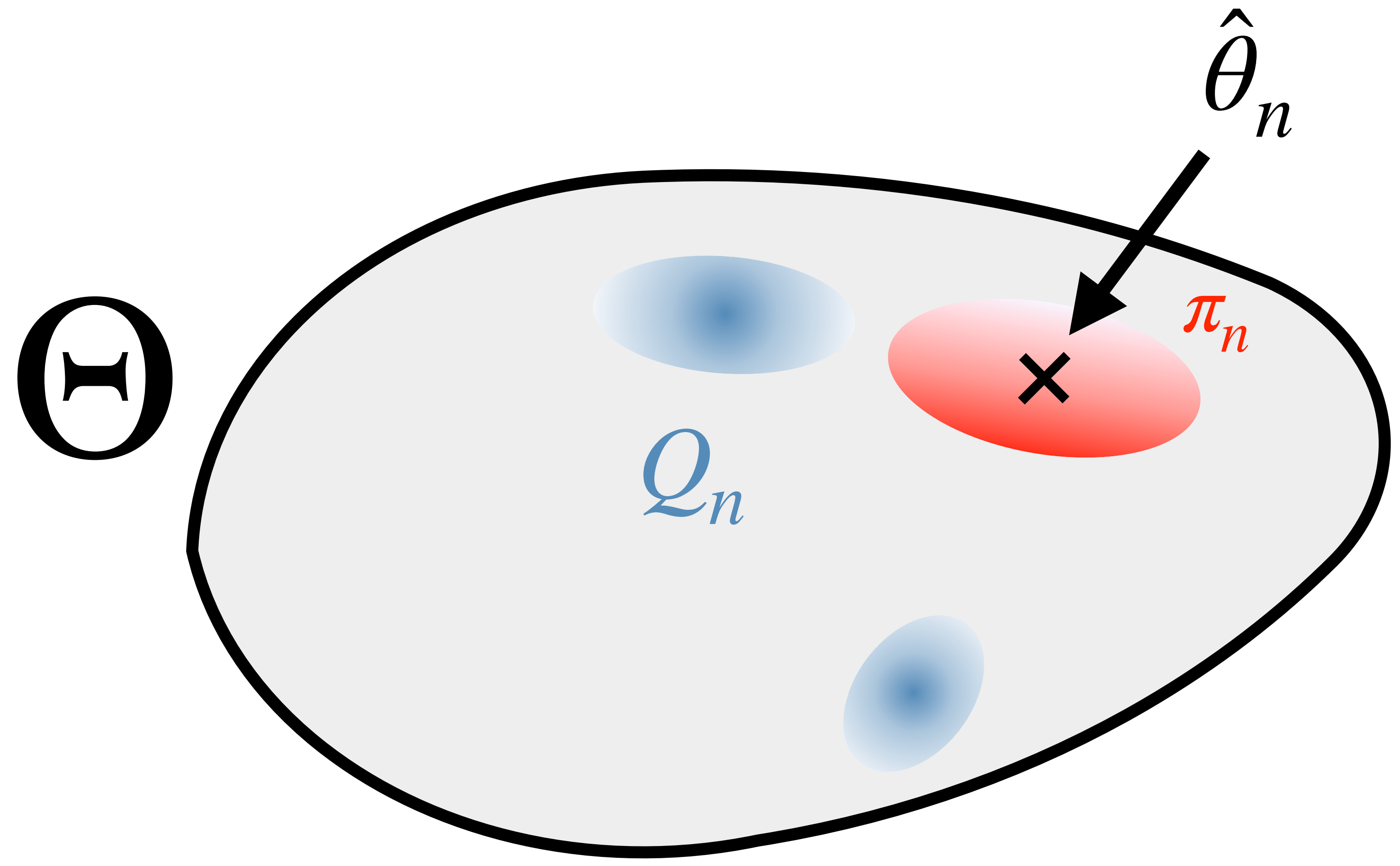
Building Intuition with the Boston Housing Data

PrO Posterior Collapses on Predictive Space



Convex Hull of models p_θ

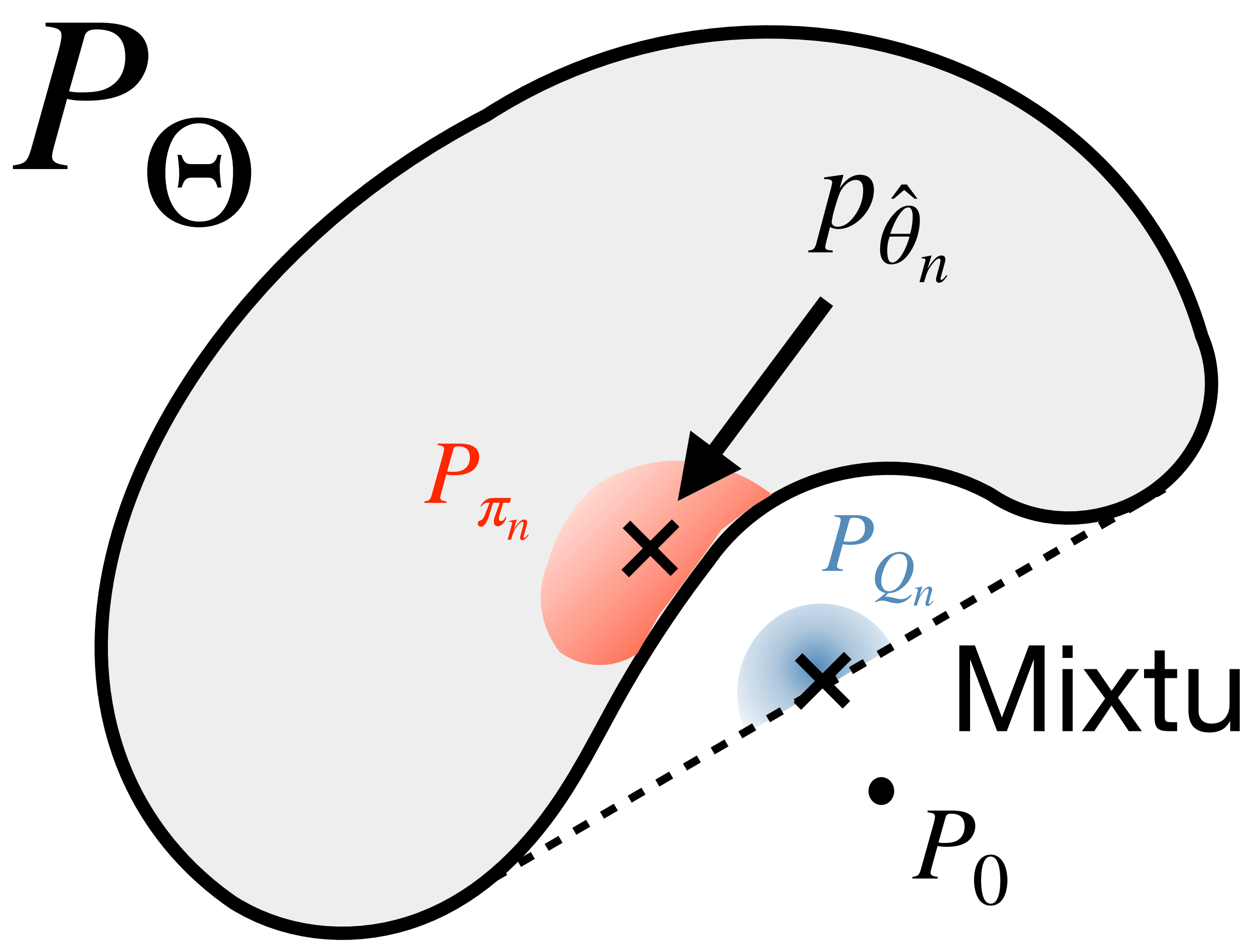
Bayes Posterior Collapses on Parameter Space



(PrO Posterior does not)

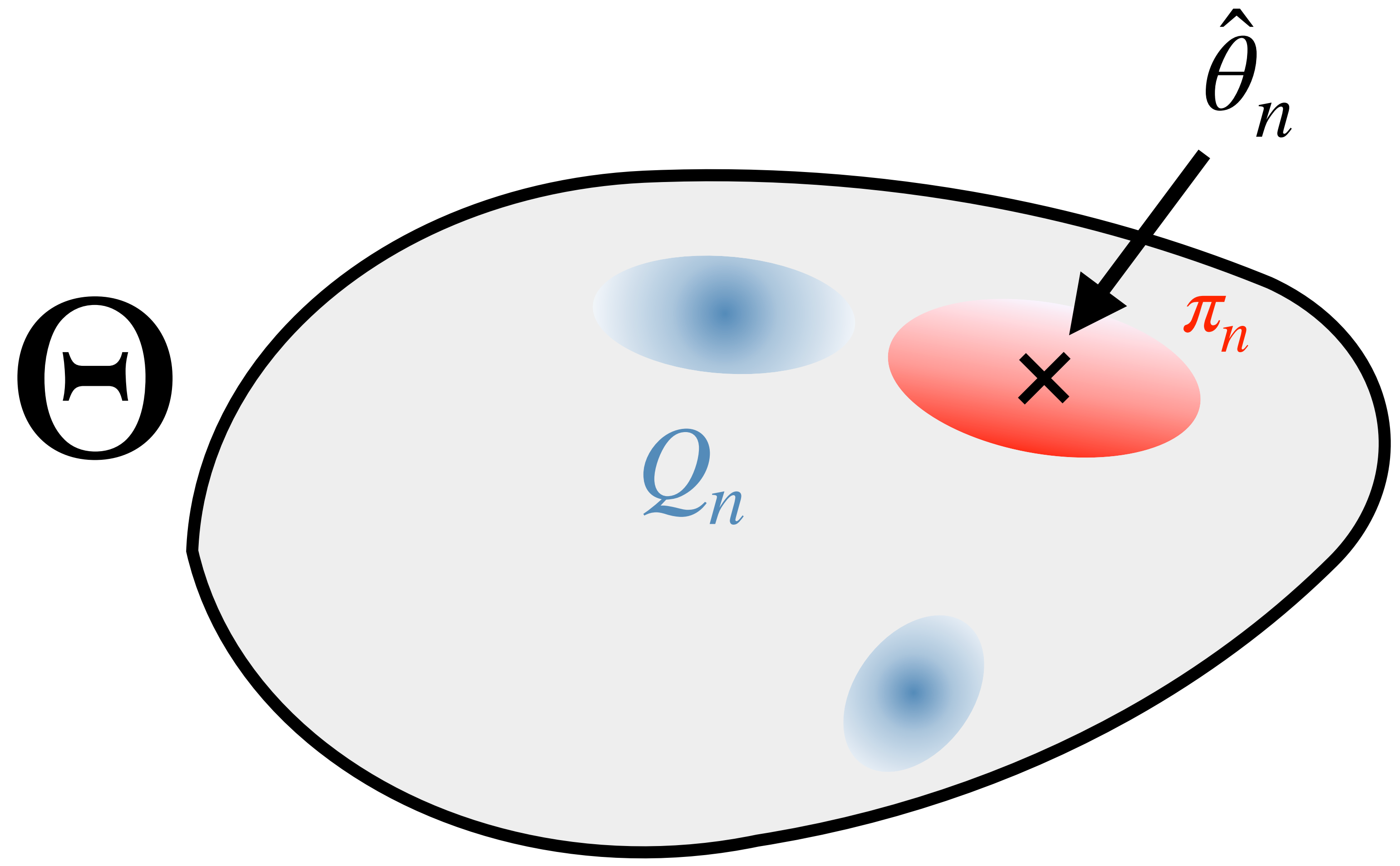
Building Intuition with the Boston Housing Data

PrO Posterior Collapses on Predictive Space



Mixture 'closest' to true process

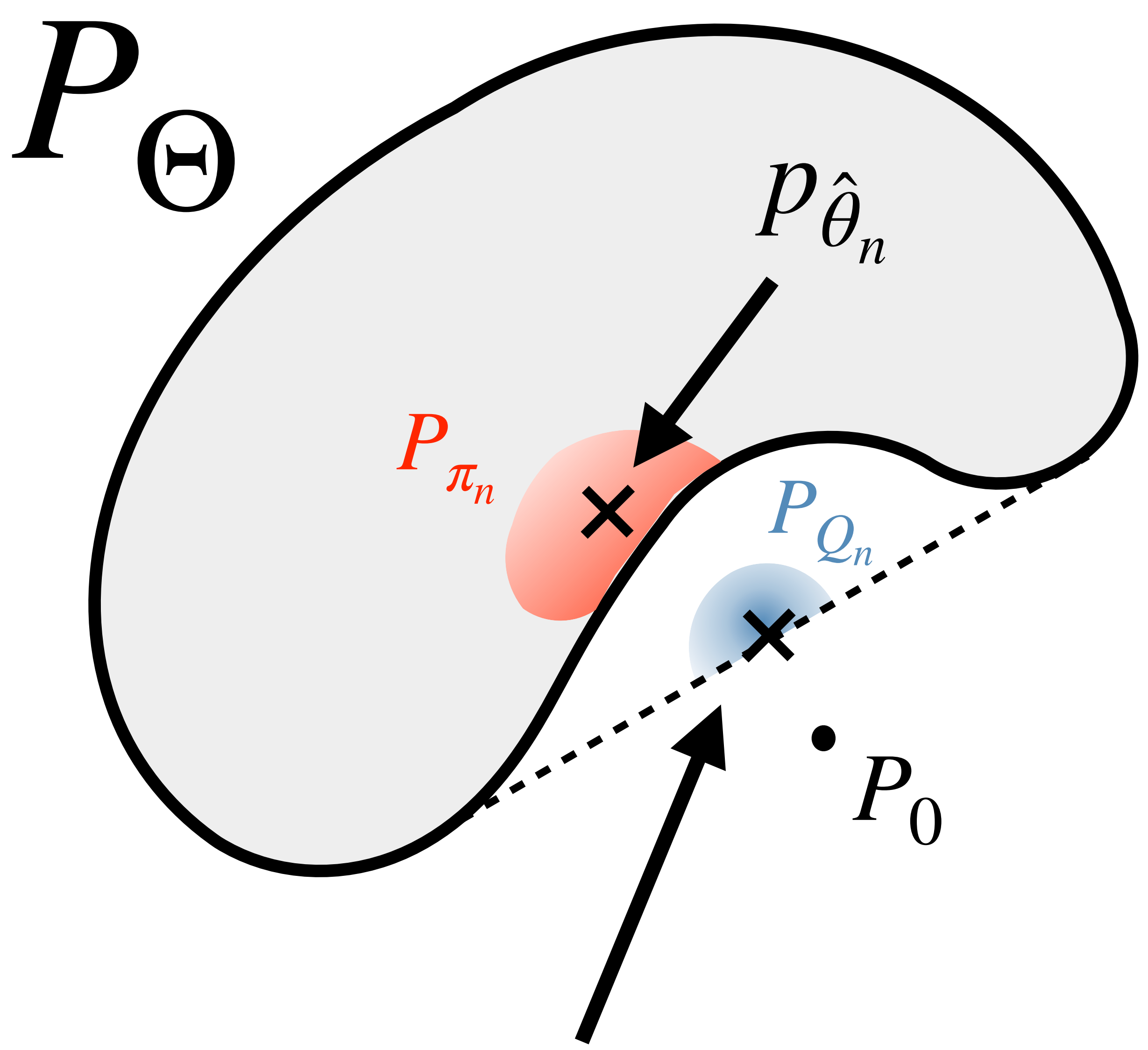
Bayes Posterior Collapses on Parameter Space



(PrO Posterior does not)

Building Intuition with the Boston Housing Data

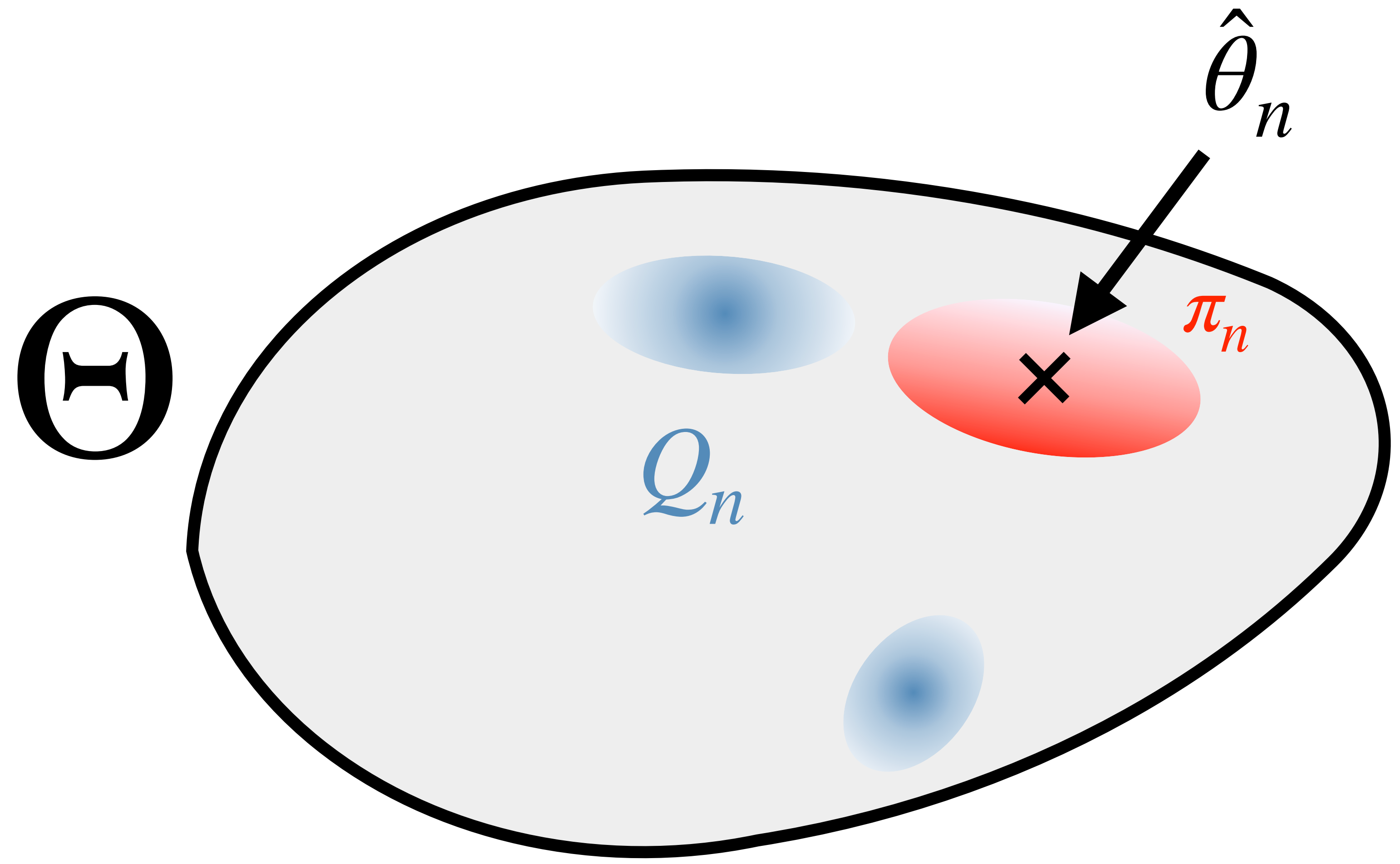
PrO Posterior Collapses on Predictive Space



$$\hat{Q} \in \mathrm{argmin}_{Q \in \mathscr{P}(\Theta)} -\log\left(\int p_\theta(x_{1:n}) \, \mathrm{d}Q(\theta) \right)$$

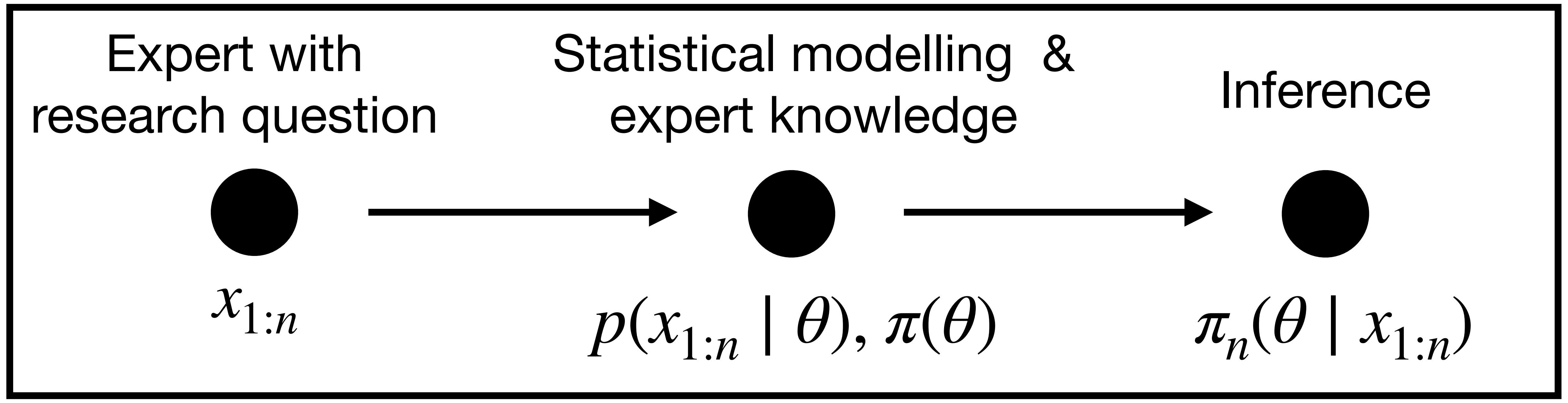
(Nonparametric MLE)

Bayes Posterior Collapses on Parameter Space



(PrO Posterior does not)

Summary: Post-Bayesian Beliefs & PrO posteriors



Expert with research question

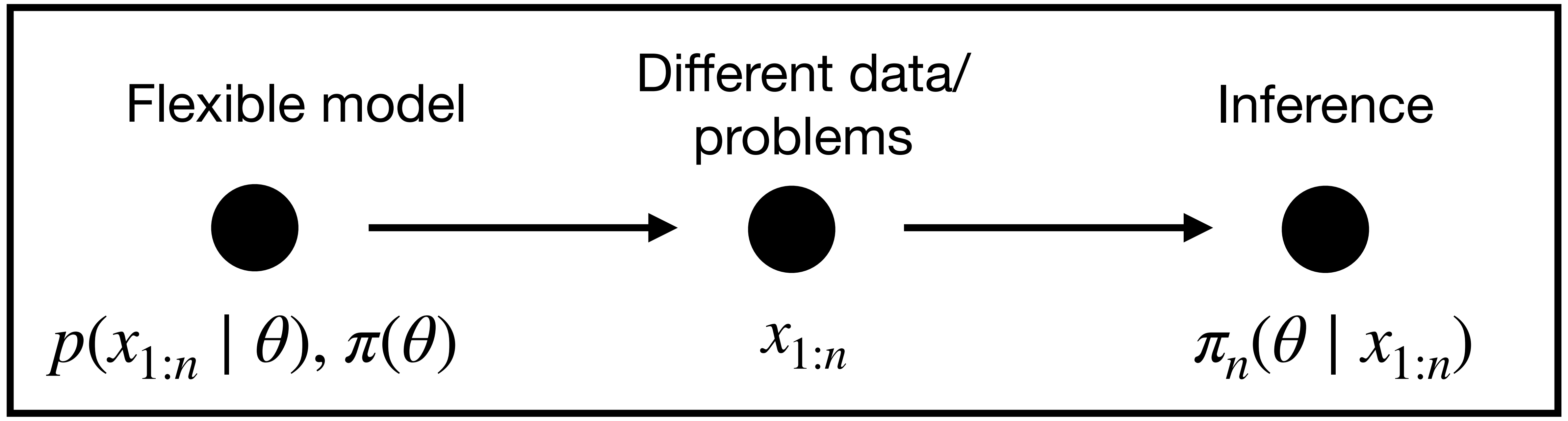
$x_{1:n}$

Statistical modelling & expert knowledge

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

Inference

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



Flexible model

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

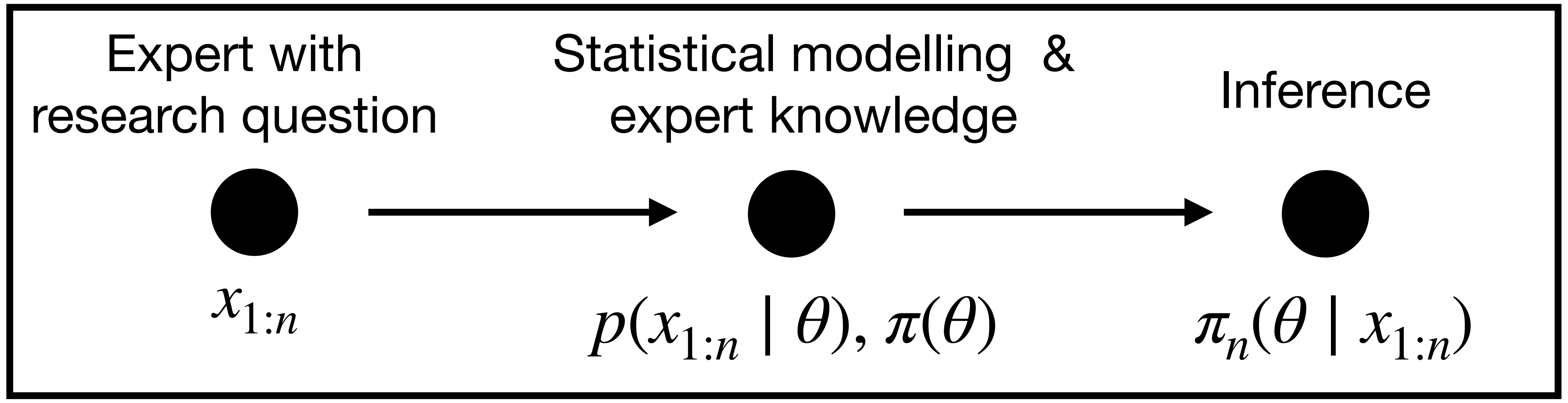
Different data/ problems

$x_{1:n}$

Inference

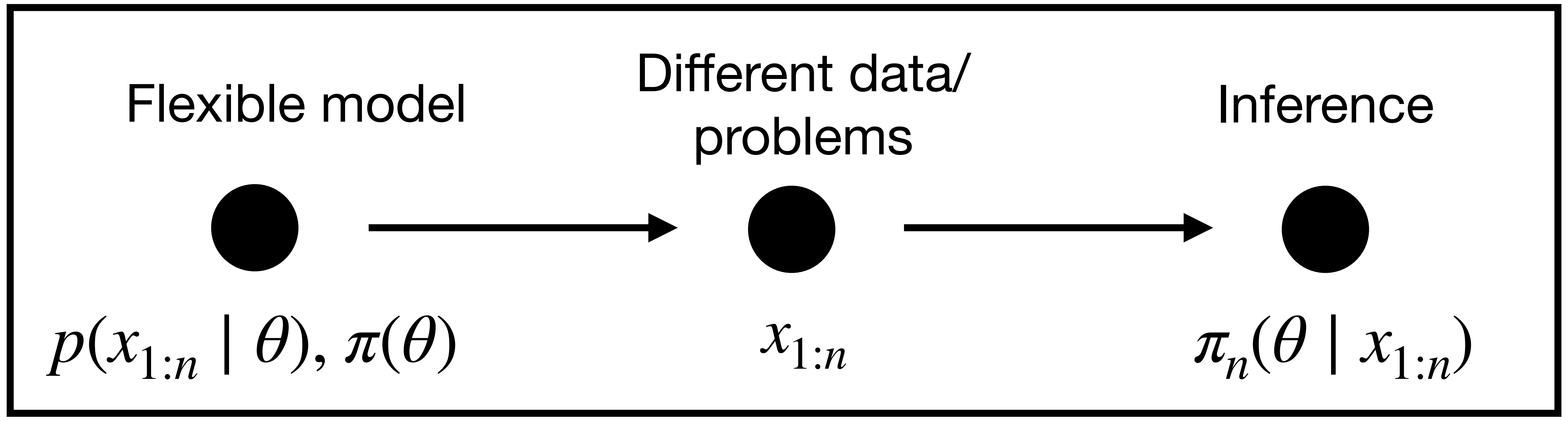
$\pi_n(\theta \mid x_{1:n})$

Summary: Post-Bayesian Beliefs & PrO posteriors



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

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



Flexible model → Different data/ problems → Inference

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

Breimann (2001): ‘The Two Cultures’

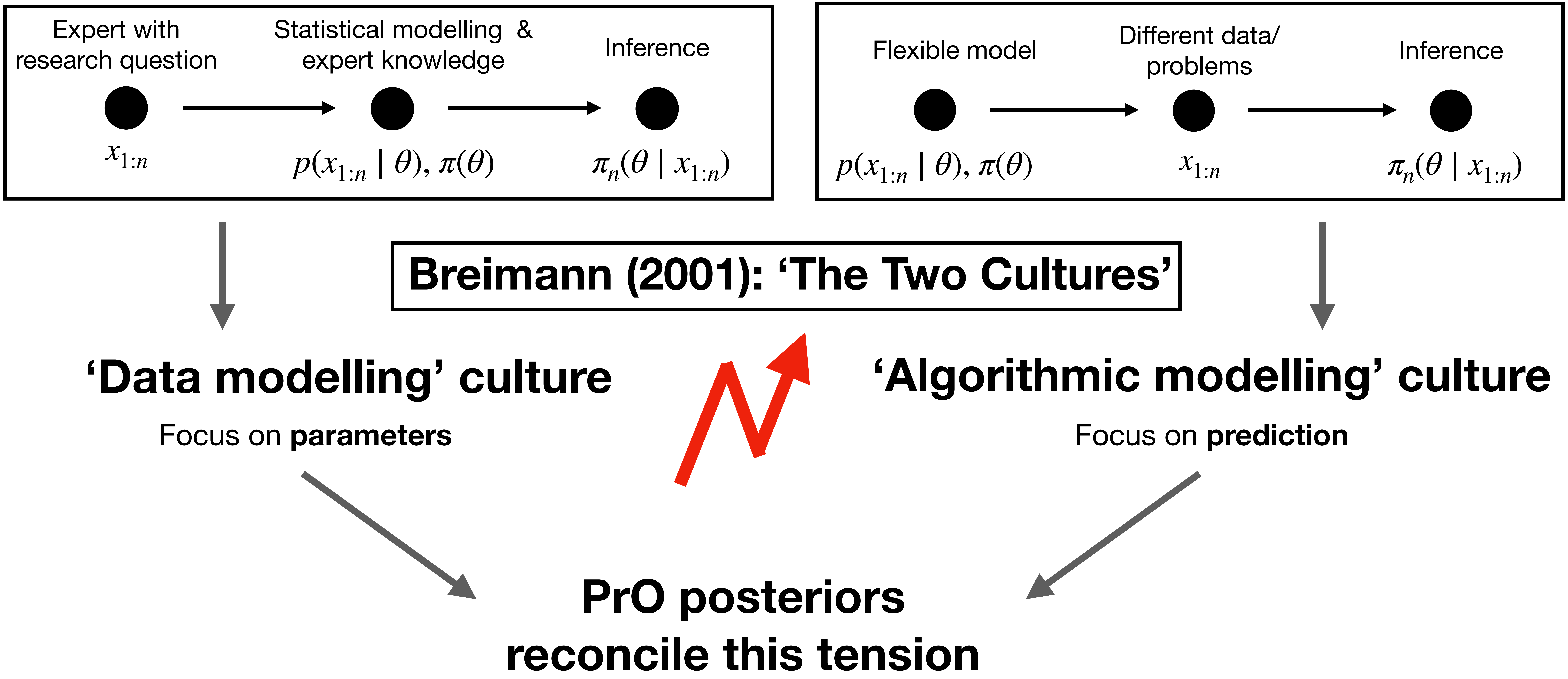
‘Data modelling’ culture

Focus on **parameters**

‘Algorithmic modelling’ culture

Focus on **prediction**

Summary: Post-Bayesian Beliefs & PrO posteriors



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

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

Flexible model → Different data/ problems → Inference

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

Breimann (2001): 'The Two Cultures'

'Data modelling' culture

Focus on **parameters**

'Algorithmic modelling' culture

Focus on **prediction**

PrO posteriors reconcile this tension

Getting Involved / Learning More

Mailing List:

https://tinyurl.com/postBayesSubscribe

Next Workshop:

AISTATS 2026: OPTIMAL (Optimisation and Post-Bayesian Inference in ML)

2 part tutorial-type overview (Cambridge's Newton Institute):

Part 1: https://www.youtube.com/watch?v=pNqg_So9brY
Part 2: https://www.youtube.com/watch?v=IwujAuVpeaw

Various lectures on selected topics in post-Bayes:

Youtube Channel https://www.youtube.com/@postbayes