Using Bagged Posteriors for Robust Inference

Jonathan Huggins
Harvard University

Joint work with Jeff Miller

Bagged posterior corrects for model misspecification

[**H** & Miller 2019]

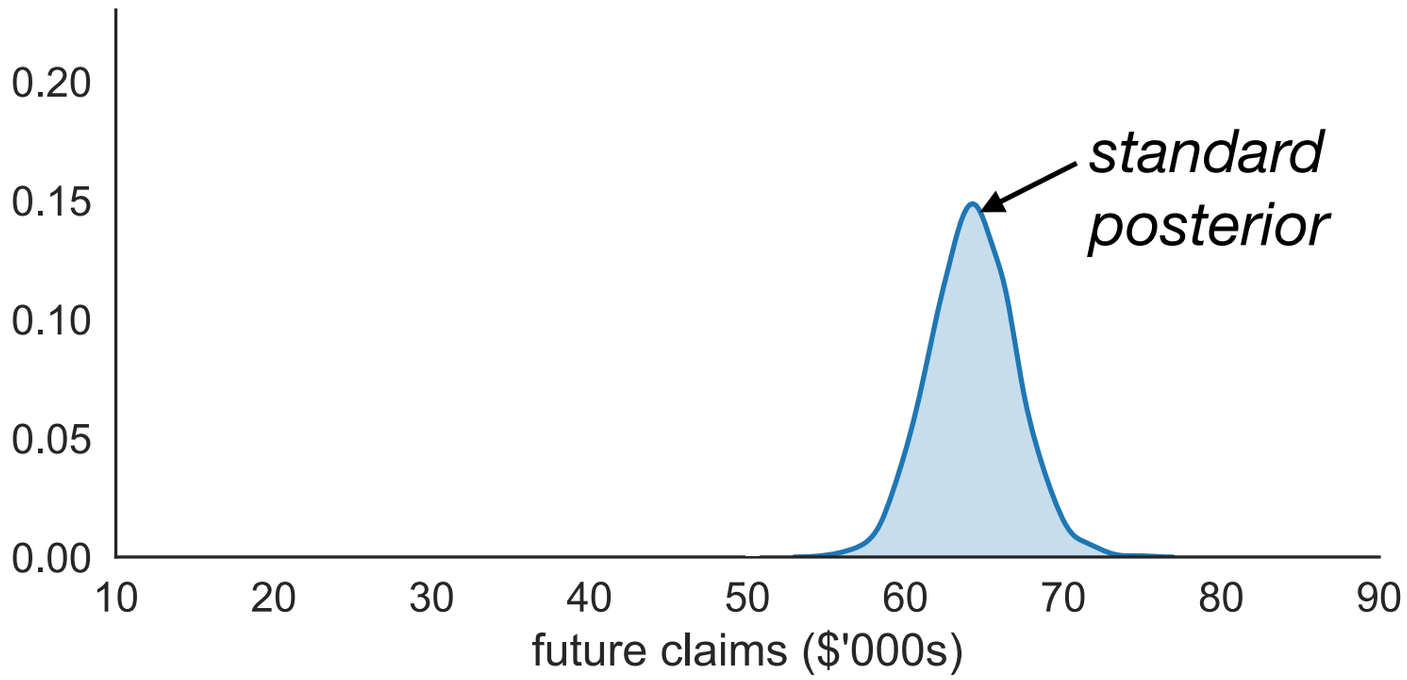
Bagged posterior corrects for model misspecification

- **Goal:** predict future insurance claims based on (real) historic data

[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

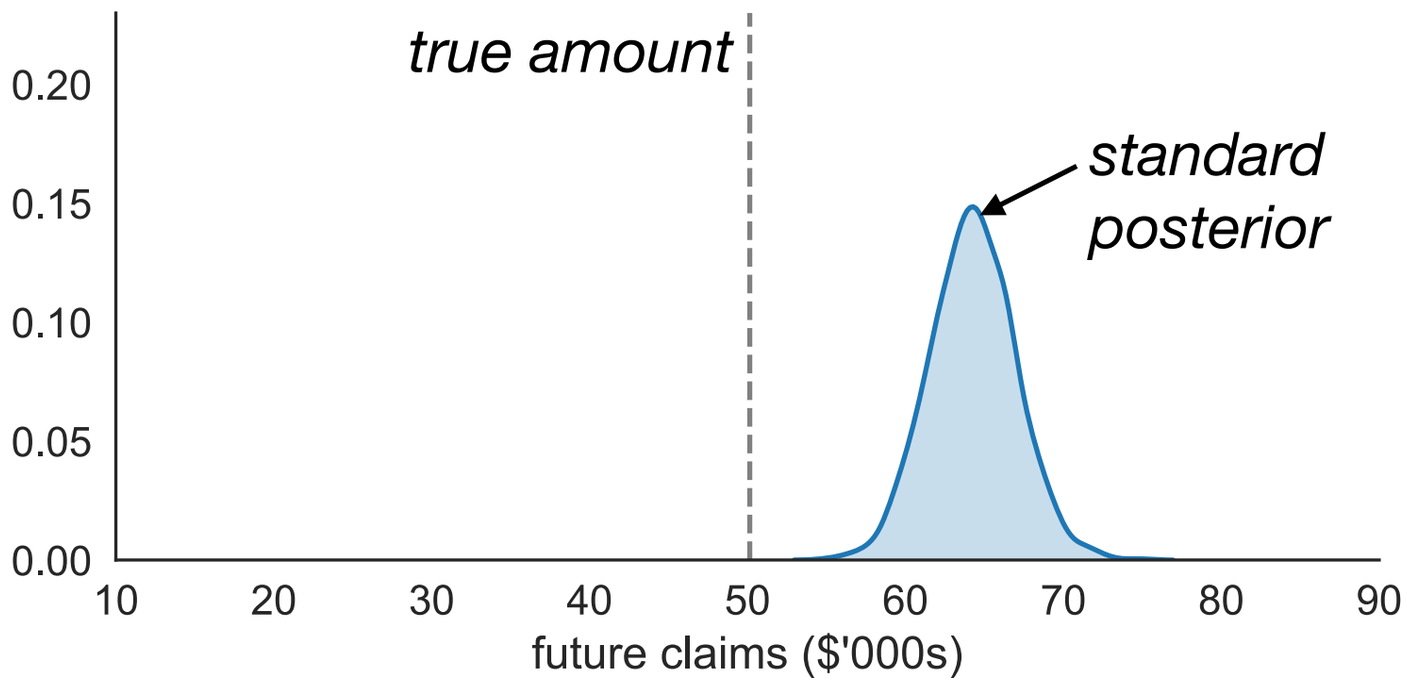
- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference (non-trivial) model (data is 10 time series)



[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

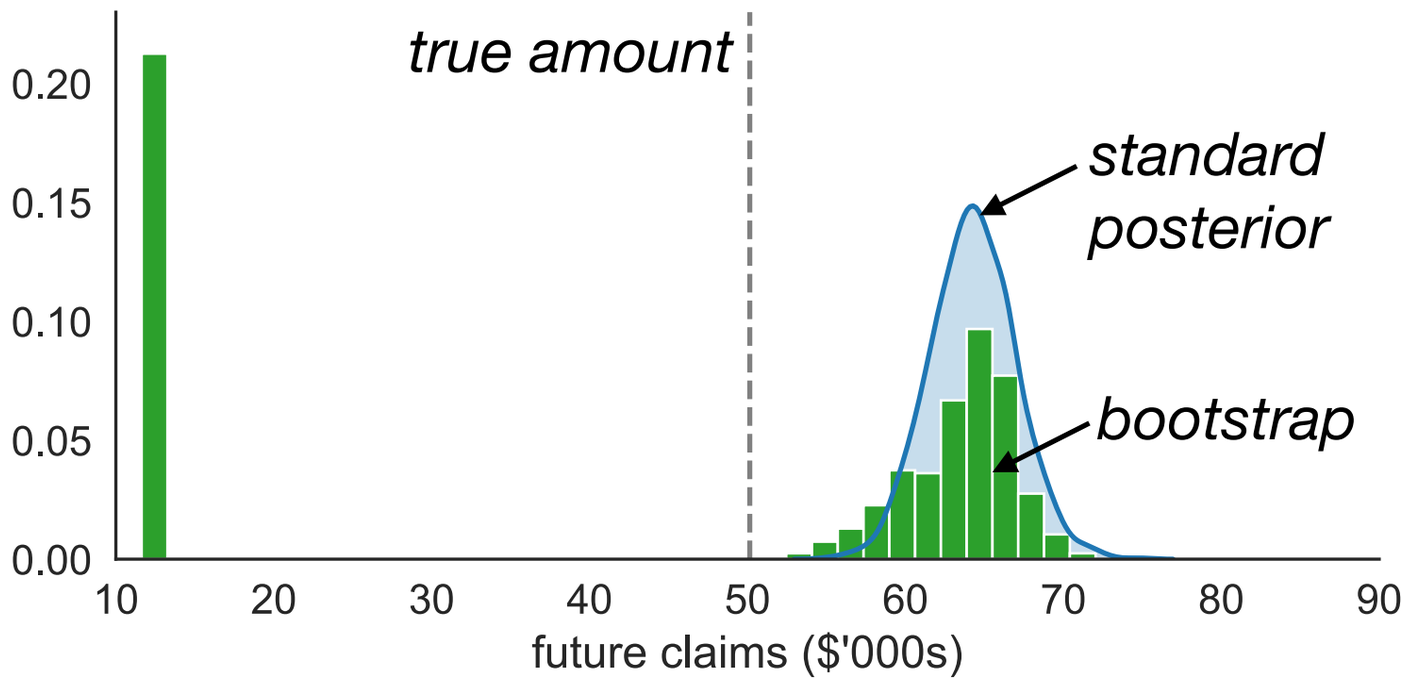
- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference (non-trivial) model (data is 10 time series)
- **Problem:** uncertainty not well-calibrated because model is wrong



[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

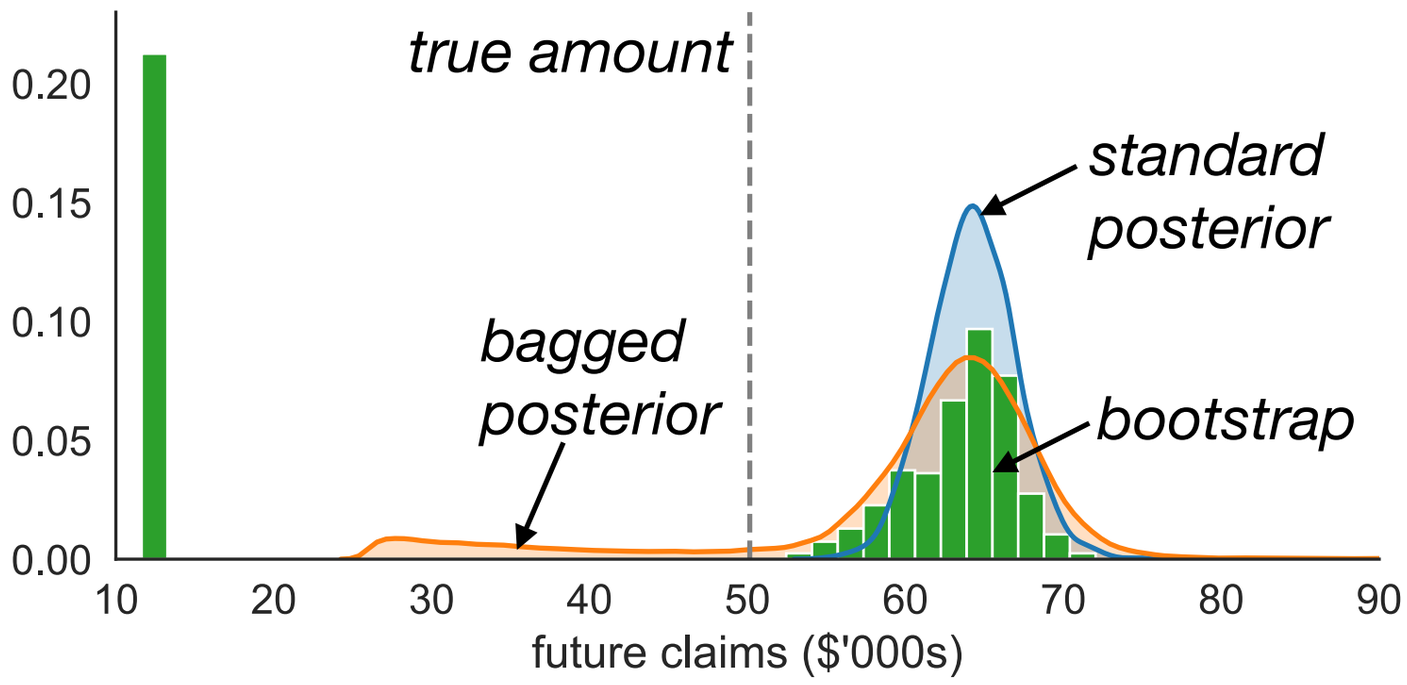
- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference (non-trivial) model (data is 10 time series)
- **Problem:** uncertainty not well-calibrated because model is wrong
- **Alternative:** the bootstrap ⇒ too little data



[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference (non-trivial) model (data is 10 time series)
- **Problem:** uncertainty not well-calibrated because model is wrong
- **Alternative:** the bootstrap ⇒ too little data
- **Solution:** use the *bagged posterior* (BayesBag)



[H & Miller 2019]

Agenda

- ➡ **BayesBag for parameter inference (and prediction)**
- ● BayesBag methodology
- ● BayesBag for model selection

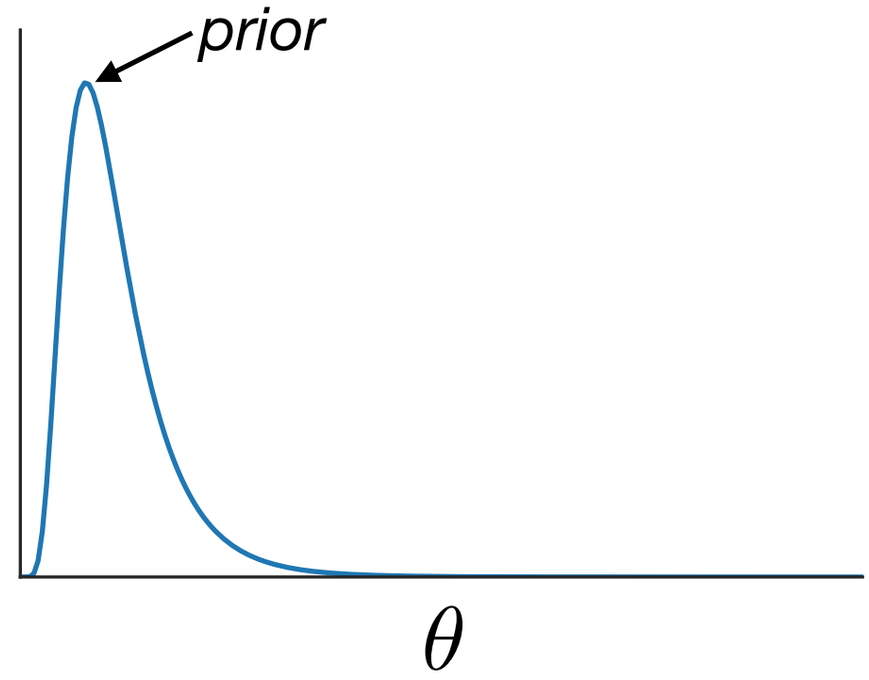
Bayesian inference

Bayesian inference

- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**

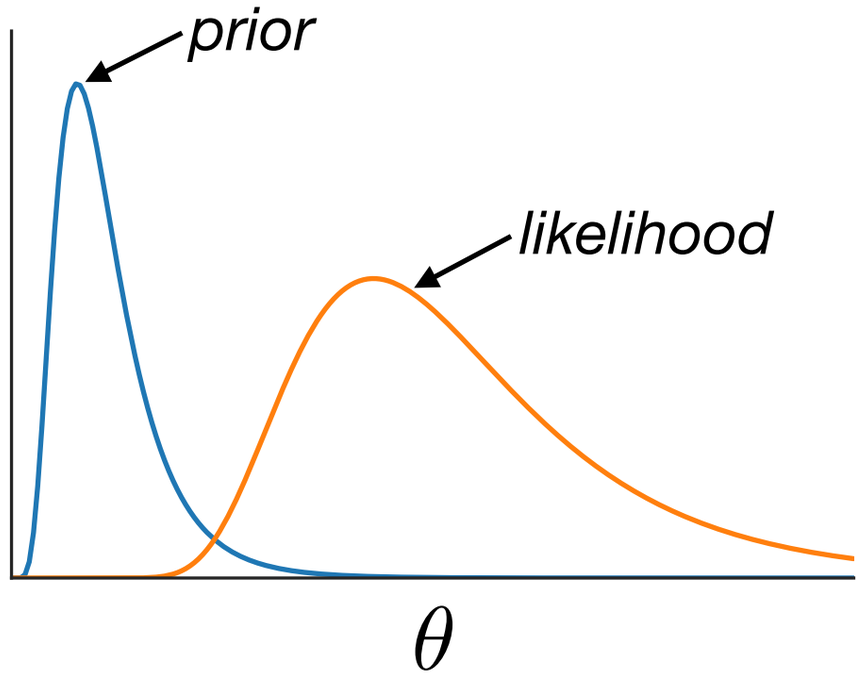
Bayesian inference

- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon



Bayesian inference

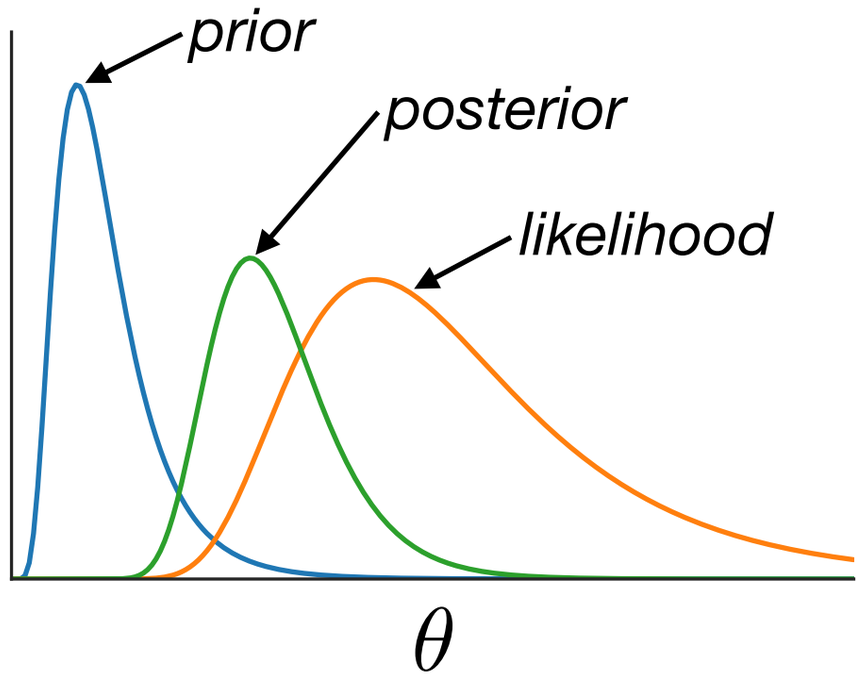
- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$



Bayesian inference

- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$
- Combine prior & likelihood to form **posterior**:

$$\pi(\theta \mid Y) \propto p(Y \mid \theta)\pi_0(\theta)$$

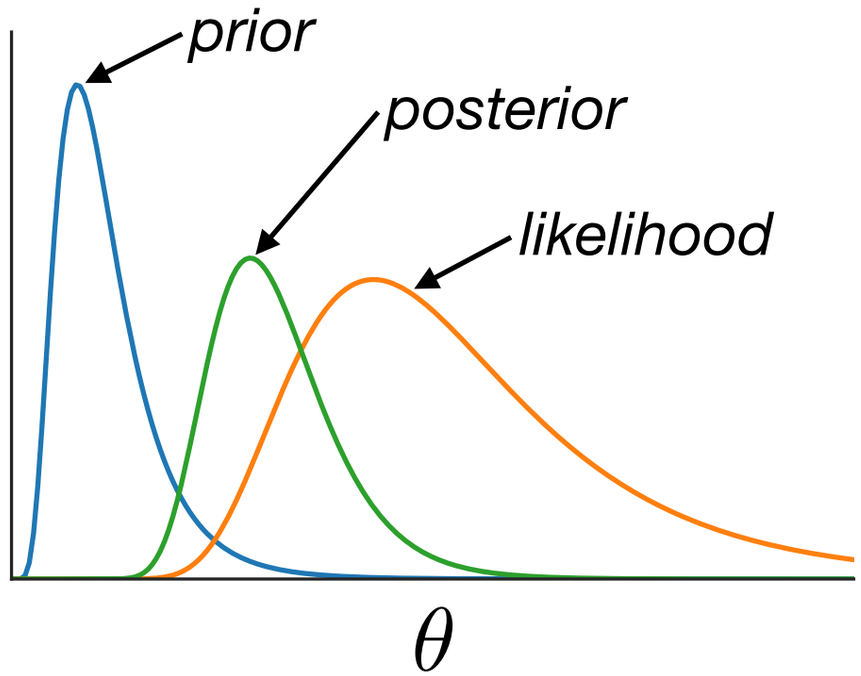


Bayesian inference

- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$
- Combine prior & likelihood to form **posterior**:

$$\pi(\theta \mid Y) \propto p(Y \mid \theta)\pi_0(\theta)$$

- **Benefits:** coherent belief updates, uncertainty quantification, flexible modeling, and more

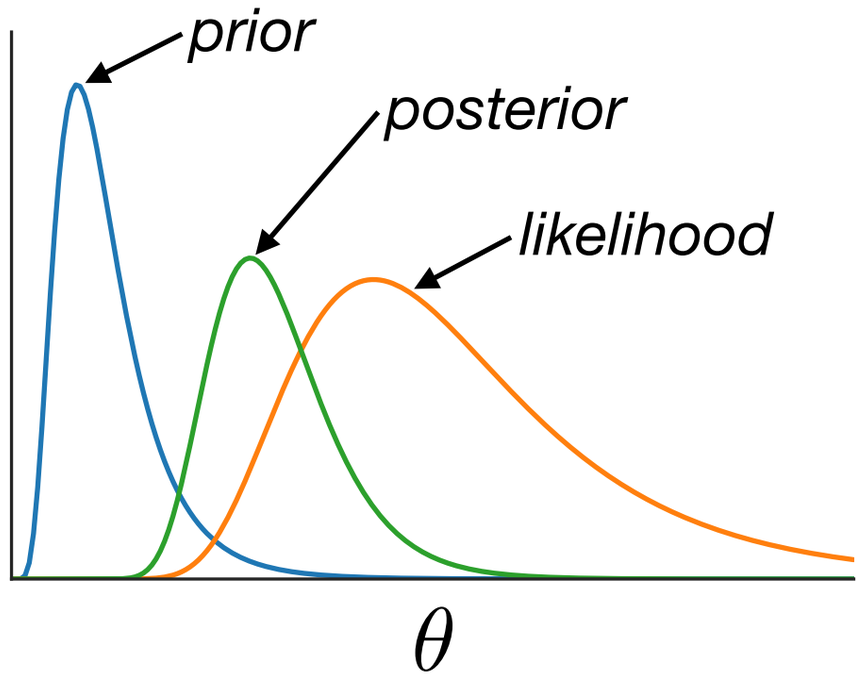


Bayesian inference

- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$
- Combine prior & likelihood to form **posterior**:

$$\pi(\theta \mid Y) \propto p(Y \mid \theta)\pi_0(\theta)$$

- **Benefits:** coherent belief updates, uncertainty quantification, flexible modeling, and more
- **Assumption #1:** measurement model correct: *observed* Y has distribution $p(Y \mid \theta_{\text{true}})$

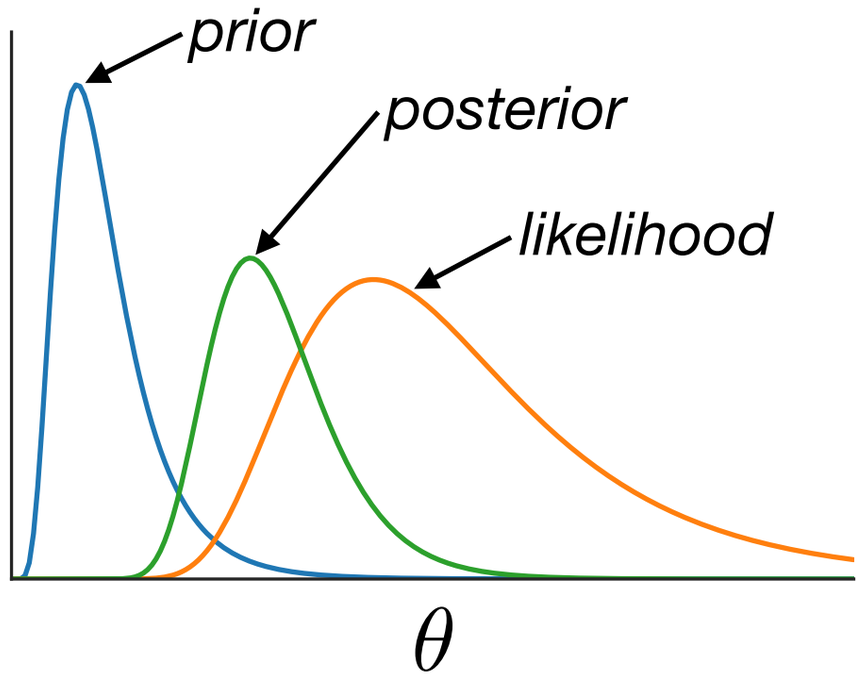


Bayesian inference

- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$
- Combine prior & likelihood to form **posterior**:

$$\pi(\theta \mid Y) \propto p(Y \mid \theta)\pi_0(\theta)$$

- **Benefits:** coherent belief updates, uncertainty quantification, flexible modeling, and more
- **Assumption #1:** measurement model correct: *observed* Y has distribution $p(Y \mid \theta_{\text{true}})$
- **Assumption #2:** Prior puts sufficient mass on true parameter θ_{true}



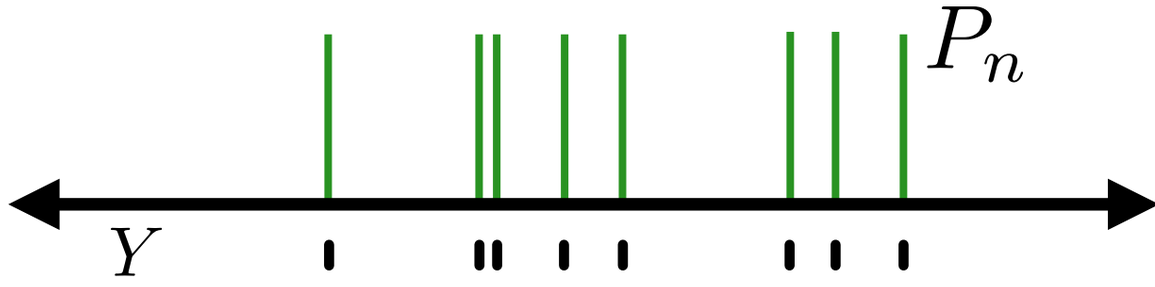
The bagged posterior

- Have data $Y = (Y_1, \ldots, Y_n)$

[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

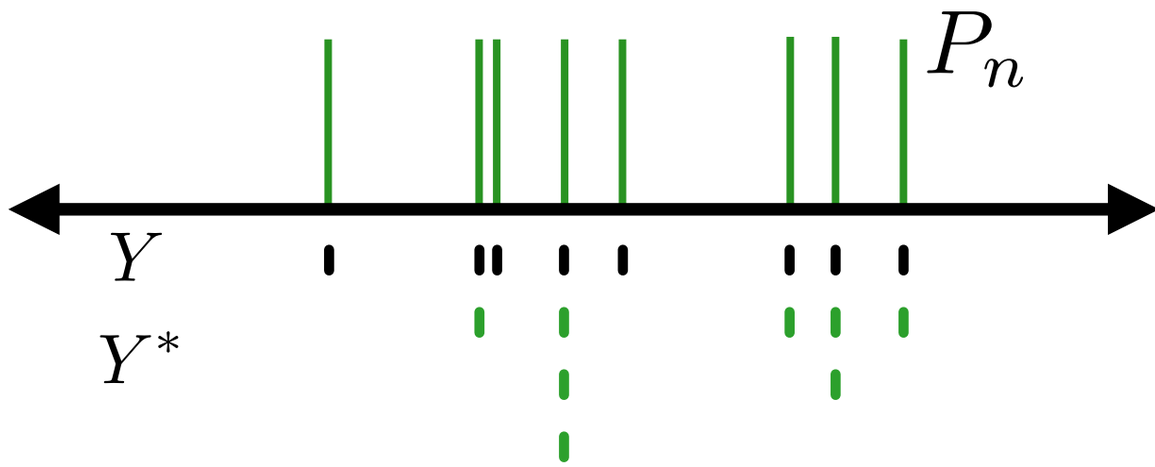
- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$



[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

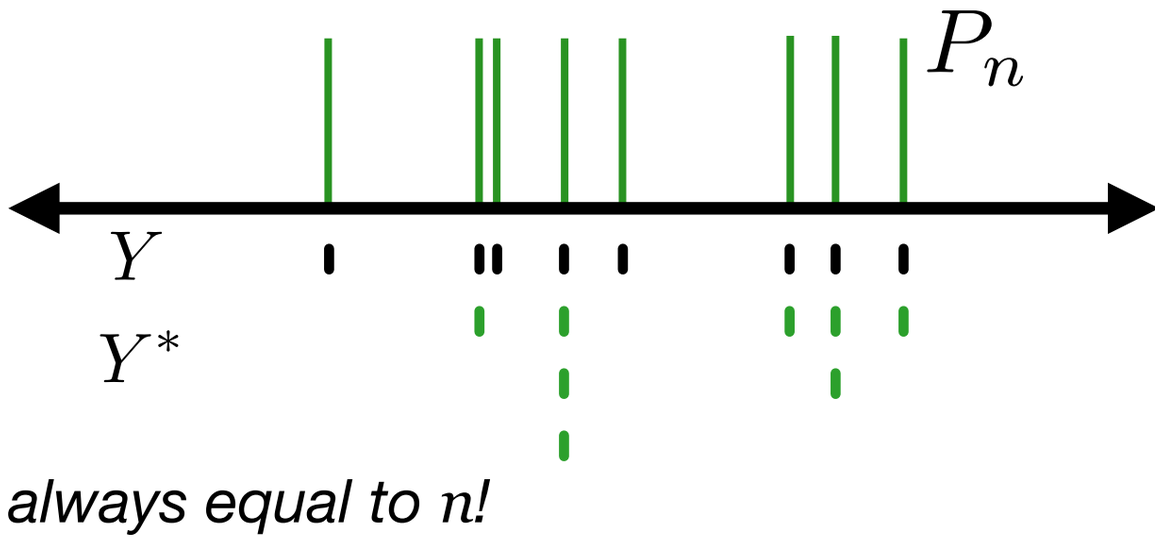
- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where Y_i^* i.i.d. $\sim P_n$



[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

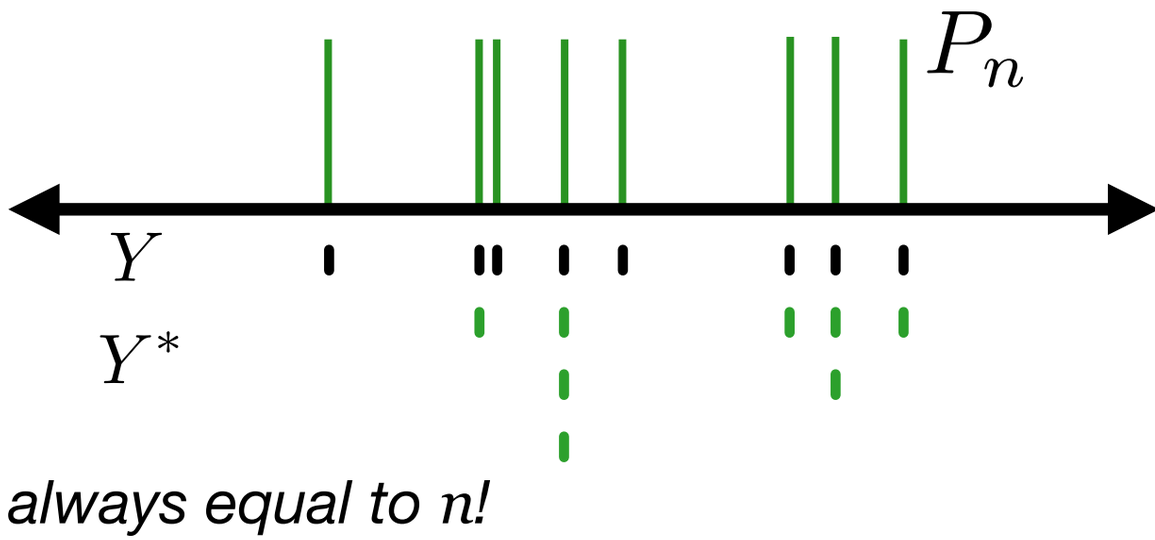
- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where Y_i^* i.i.d. $\sim P_n$



[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

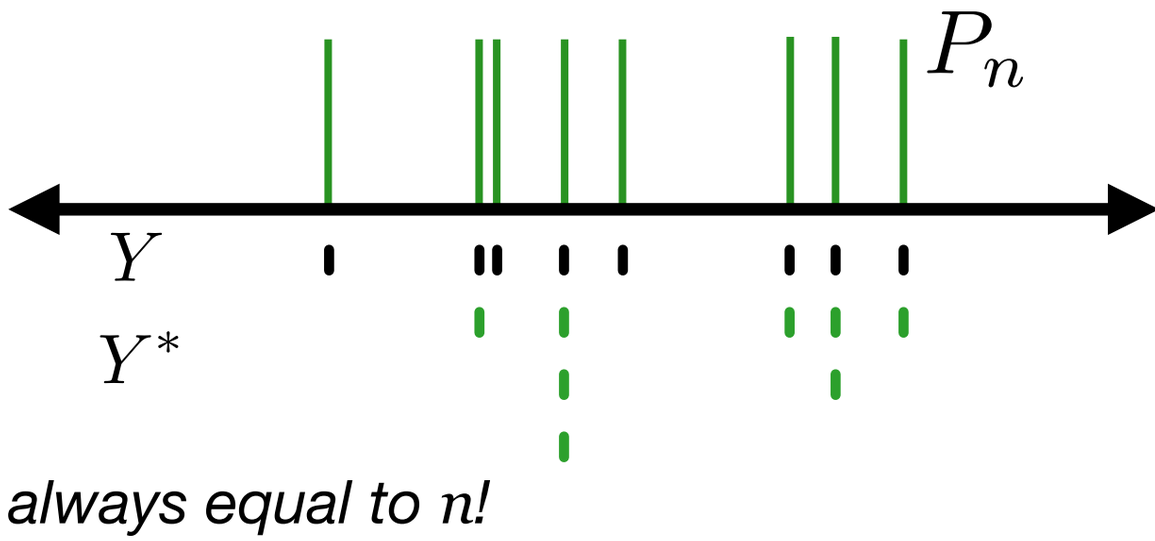
- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where $Y_i^* \text{ i.i.d.} \sim P_n$
- **Bagged posterior**
  $\pi^*(\theta \,|\, Y) = \mathbb{E}\{\pi(\theta \,|\, Y^*) \,|\, Y\}$



[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where Y_i^* i.i.d. $\sim P_n$
- **Bagged posterior**
  $\pi^*(\theta \,|\, Y) = \mathbb{E}\{\pi(\theta \,|\, Y^*) \,|\, Y\}$

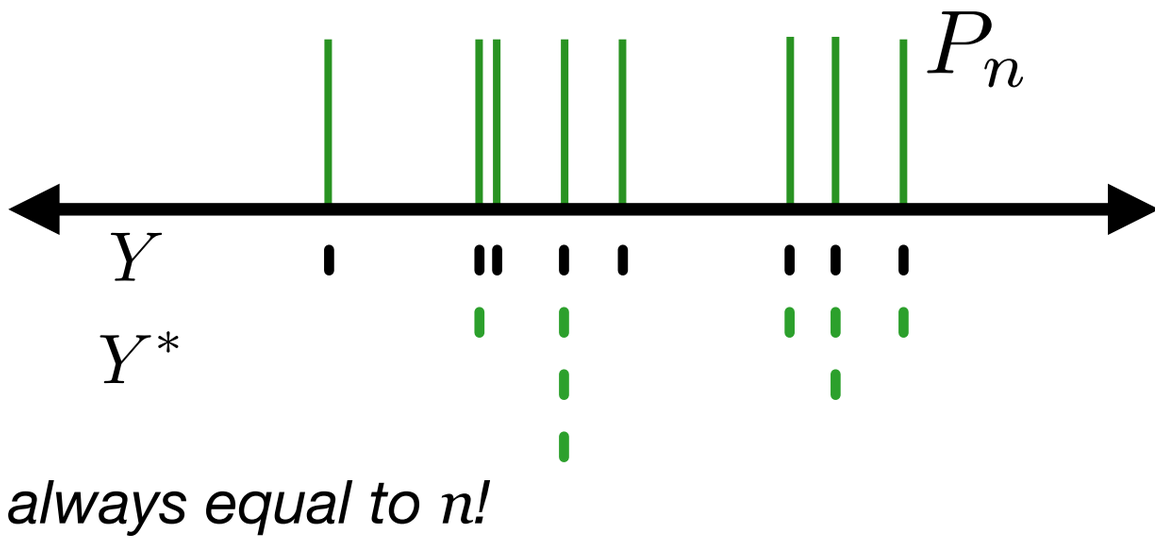


- We show: when $m = n$, **conservative** uncertainty even when model wrong

[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where Y_i^* i.i.d. $\sim P_n$

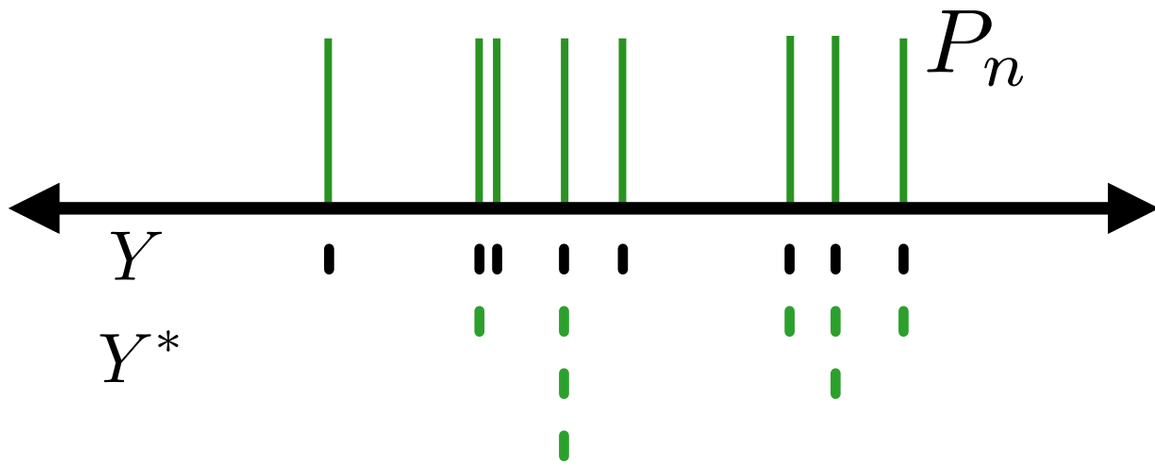


- **Bagged posterior** $\pi^*(\theta \,|\, Y) = \mathbb{E}\{\pi(\theta \,|\, Y^*) \,|\, Y\}$
- We show: when $m = n$, **conservative** uncertainty even when model wrong
- In practice, sample B bootstrap datasets: $\pi^*(\theta \,|\, Y) \approx \frac{1}{B} \sum_{b=1}^{B} \pi(\theta \,|\, Y_{(b)}^*)$

[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where Y_i^* i.i.d. $\sim P_n$

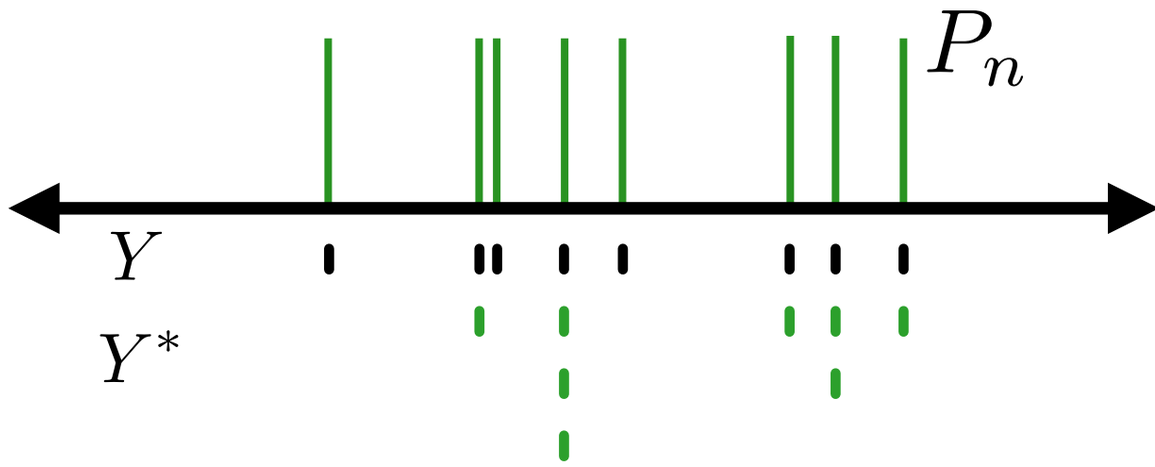


- **Bagged posterior** $\pi^*(\theta \,|\, Y) = \mathbb{E}\{\pi(\theta \,|\, Y^*) \,|\, Y\}$
- We show: when $m = n$, **conservative** uncertainty even when model wrong
- In practice, sample B bootstrap datasets: $\pi^*(\theta \,|\, Y) \approx \frac{1}{B} \sum_{b=1}^{B} \pi(\theta \,|\, Y_{(b)}^*)$
- Suffices to take $B = 50$ or 100

[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

The bagged posterior

- Have data $Y = (Y_1, \ldots, Y_n)$
- Empirical data distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$
- Bootstrap dataset $Y^* = (Y_1^*, \ldots, Y_m^*)$, where Y_i^* i.i.d. $\sim P_n$



- **Bagged posterior**
  $\pi^*(\theta \,|\, Y) = \mathbb{E}\{\pi(\theta \,|\, Y^*) \,|\, Y\}$
- We show: when $m = n$, **conservative** uncertainty even when model wrong
- In practice, sample B bootstrap datasets: $\pi^*(\theta \,|\, Y) \approx \frac{1}{B} \sum_{b=1}^{B} \pi(\theta \,|\, Y_{(b)}^*)$
- Suffices to take $B = 50$ or 100
- Benefits: **easy to use**, can **parallelize** across B

[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

Now entering asymptopia…

Now entering asymptopia…

- Assume $\theta \in \mathbb{R}^d$

Now entering asymptopia...

- Assume $\theta \in \mathbb{R}^d$
- Assume independent observations, so posterior is $\pi(\theta \mid Y) \propto \pi_0(\theta) \prod_{i=1}^n p_\theta(Y_i)$

Now entering asymptopia...

- Assume $\theta \in \mathbb{R}^d$
- Assume independent observations, so posterior is $\pi(\theta \mid Y) \propto \pi_0(\theta) \prod_{i=1}^n p_\theta(Y_i)$
- Assume data $Y_1, \ldots, Y_n$ i.i.d. $\sim P_{\text{true}}$

Now entering asymptopia...

- Assume $\theta \in \mathbb{R}^d$
- Assume independent observations, so posterior is $\pi(\theta \mid Y) \propto \pi_0(\theta) \prod_{i=1}^n p_\theta(Y_i)$
- Assume data $Y_1, \ldots, Y_n$ i.i.d. $\sim P_{\text{true}}$
- From now on, we consider the **large n limit**

Now entering asymptopia...

- Assume $\theta \in \mathbb{R}^d$
- Assume independent observations, so posterior is $\pi(\theta \mid Y) \propto \pi_0(\theta) \prod_{i=1}^n p_\theta(Y_i)$
- Assume data $Y_1, \ldots, Y_n$ i.i.d. $\sim P_{\text{true}}$
- From now on, we consider the **large n limit**
- We will be **ignoring the finite-sample benefits of Bayes**

Now entering asymptopia...

- Assume $\theta \in \mathbb{R}^d$
- Assume independent observations, so posterior is $\pi(\theta \mid Y) \propto \pi_0(\theta) \prod_{i=1}^n p_\theta(Y_i)$
- Assume data $Y_1, \ldots, Y_n$ i.i.d. $\sim P_{\text{true}}$
- From now on, we consider the **large n limit**
- We will be **ignoring the finite-sample benefits of Bayes**
- But, we still get many useful insights!

Uncertainty about the optimal parameter

Uncertainty about the optimal parameter

- Interested in quantifying uncertainty about **KL-optimal parameter**

$$\theta_{\mathsf{opt}} = \arg\max_{\theta} \mathbb{E}\{\log p_\theta(Y_1)\}$$

Uncertainty about the optimal parameter

- Interested in quantifying uncertainty about **KL-optimal parameter**

$$\theta_{\text{opt}} = \arg\max_{\theta} \mathbb{E}\{\log p_\theta(Y_1)\}$$

- Why? $p_{\theta_{\text{opt}}}$ best explains the data amongst the models $\{p_\theta : \theta \in \mathbb{R}^d\}$

Uncertainty about the optimal parameter

- Interested in quantifying uncertainty about **KL-optimal parameter**

$$\theta_{\mathsf{opt}} = \arg\max_{\theta} \mathbb{E}\{\log p_\theta(Y_1)\}$$

- Why? $p_{\theta_{\mathsf{opt}}}$ best explains the data amongst the models $\{p_\theta : \theta \in \mathbb{R}^d\}$
- When model correctly specified, $p_{\theta_{\mathsf{opt}}} = P_{\mathsf{true}}$

Uncertainty about the optimal parameter

- Interested in quantifying uncertainty about **KL-optimal parameter**

$$\theta_{\text{opt}} = \arg\max_{\theta} \mathbb{E}\{\log p_\theta(Y_1)\}$$

- Why? $p_{\theta_{\text{opt}}}$ best explains the data amongst the models $\{p_\theta : \theta \in \mathbb{R}^d\}$

- When model correctly specified, $p_{\theta_{\text{opt}}} = P_{\text{true}}$

- Define the maximum likelihood estimator

Uncertainty about the optimal parameter

- Interested in quantifying uncertainty about **KL-optimal parameter**

$$\theta_{\text{opt}} = \arg\max_{\theta} \mathbb{E}\{\log p_\theta(Y_1)\}$$

- Why? $p_{\theta_{\text{opt}}}$ best explains the data amongst the models $\{p_\theta : \theta \in \mathbb{R}^d\}$

- When model correctly specified, $p_{\theta_{\text{opt}}} = P_{\text{true}}$

- Define the maximum likelihood estimator

$$\hat{\theta}(Y) = \arg\max_{\theta} \prod_{i=1}^{n} p_\theta(Y_i)$$

Uncertainty about the optimal parameter

- Interested in quantifying uncertainty about **KL-optimal parameter**

$$\theta_{\text{opt}} = \arg\max_{\theta} \mathbb{E}\{\log p_\theta(Y_1)\}$$

- Why? $p_{\theta_{\text{opt}}}$ best explains the data amongst the models $\{p_\theta : \theta \in \mathbb{R}^d\}$

- When model correctly specified, $p_{\theta_{\text{opt}}} = P_{\text{true}}$

- Define the maximum likelihood estimator

$$\hat{\theta}(Y) = \arg\max_{\theta} \prod_{i=1}^{n} p_\theta(Y_i)$$

- Under mild conditions, $\hat{\theta}(Y) \to \theta_{\text{opt}}$ and $\pi(\cdot \,|\, Y) \overset{\mathcal{D}}{\Longrightarrow} \delta_{\theta_{\text{opt}}}$ as $n \to \infty$

The posterior can be over-confident when the model is wrong

The posterior can be over-confident when the model is wrong

[Kleijn & van der Vaart 2012] For $\vartheta \sim \pi(\cdot \,|\, Y)$,

$$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \qquad \Big(\text{that is, } n^{1/2}\{\vartheta - \hat{\theta}(Y)\} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma_{\mathsf{M}})\Big)$$

The posterior can be over-confident when the model is wrong

[Kleijn & van der Vaart 2012] For $\vartheta \sim \pi(\cdot \,|\, Y)$,

$$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \qquad \Big(\text{that is, } n^{1/2}\{\vartheta - \hat{\theta}(Y)\} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma_{\mathsf{M}})\Big)$$

"model" covariance

The posterior can be over-confident when the model is wrong

[Kleijn & van der Vaart 2012] For $\vartheta \sim \pi(\cdot \,|\, Y)$,

$$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \qquad \Big(\text{that is, } n^{1/2}\{\vartheta - \hat{\theta}(Y)\} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma_{\mathsf{M}})\Big)$$

"model" covariance

[Huber 1967, White 1982] Sampling distribution of the MLE satisfies

$$\hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n)$$

The posterior can be over-confident when the model is wrong

[Kleijn & van der Vaart 2012] For $\vartheta \sim \pi(\cdot \,|\, Y)$,

$$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \qquad \Big(\text{that is, } n^{1/2}\{\vartheta - \hat{\theta}(Y)\} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma_{\mathsf{M}})\Big)$$

"model" covariance

[Huber 1967, White 1982] Sampling distribution of the MLE satisfies

$$\hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n)$$

sandwich covariance

The posterior can be over-confident when the model is wrong

[Kleijn & van der Vaart 2012] For $\vartheta \sim \pi(\cdot \,|\, Y)$,

$$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \qquad \left(\text{that is, } n^{1/2}\{\vartheta - \hat{\theta}(Y)\} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma_{\mathsf{M}})\right)$$

"model" covariance

[Huber 1967, White 1982] Sampling distribution of the MLE satisfies

$$\hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n)$$

sandwich covariance

- If $P_{\mathsf{true}} = p_{\theta_{\mathsf{opt}}}$, then $\Sigma_{\mathsf{S}} = \Sigma_{\mathsf{M}}$ and posterior uncertainty about θ_{opt} correct

The posterior can be over-confident when the model is wrong

[Kleijn & van der Vaart 2012] For $\vartheta \sim \pi(\cdot \,|\, Y)$,

$$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \qquad \Big(\text{that is, } n^{1/2}\{\vartheta - \hat{\theta}(Y)\} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \Sigma_{\mathsf{M}})\Big)$$

"model" covariance

[Huber 1967, White 1982] Sampling distribution of the MLE satisfies

$$\hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n)$$

sandwich covariance

- If $P_{\mathsf{true}} = p_{\theta_{\mathsf{opt}}}$, then $\Sigma_{\mathsf{S}} = \Sigma_{\mathsf{M}}$ and posterior uncertainty about θ_{opt} correct
- If model misspecified, then in general $\Sigma_{\mathsf{S}} \neq \Sigma_{\mathsf{M}}$ and posterior uncertainty about θ_{opt} could be very wrong

Bootstrap uncertainty is well-calibrated

$$\text{MLE:} \qquad \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n)$$

$$\text{Posterior:} \qquad \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

Bootstrap uncertainty is well-calibrated

$$\text{MLE:} \qquad \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n)$$

$$\text{Posterior:} \qquad \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

[van der Vaart & Wellner 1996] Bootstrap distribution of the MLE satisfies

$$\hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)$$

Bootstrap uncertainty is well-calibrated

$$\begin{aligned} \text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\ \text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \end{aligned}$$

[van der Vaart & Wellner 1996] Bootstrap distribution of the MLE satisfies

$$\hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)$$

- With $m = n$, the bootstrap
  - quantifies **sampling variability**

Bootstrap uncertainty is well-calibrated

$$\begin{aligned} \text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\ \text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \end{aligned}$$

[van der Vaart & Wellner 1996] Bootstrap distribution of the MLE satisfies

$$\hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)$$

- With $m = n$, the bootstrap
  - quantifies **sampling variability**
  - provides correct frequentist confidence intervals

The bagged posterior is conservative

$$
\begin{aligned}
\text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\
\text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \\
\text{Bootstrap:} \quad & \hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)
\end{aligned}
$$

The bagged posterior is conservative

$$
\begin{aligned}
\text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\
\text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \\
\text{Bootstrap:} \quad & \hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)
\end{aligned}
$$

[H & Miller 2019] BayesBag includes **sampling and model-based uncertainty**: for $\vartheta^* \sim \pi^*(\cdot \,|\, Y)$,

$$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

The bagged posterior is conservative

$$\begin{aligned}
\text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\
\text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \\
\text{Bootstrap:} \quad & \hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)
\end{aligned}$$

[H & Miller 2019] BayesBag includes **sampling and model-based uncertainty**: for $\vartheta^* \sim \pi^*(\cdot \,|\, Y)$,

$$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- If model correct, $m = 2n$ yields **correct uncertainty**:

$$\mathrm{Cov}(\vartheta^* \,|\, Y) \approx \Sigma_{\mathsf{S}}/n$$

The bagged posterior is conservative

$$\begin{aligned}
\text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\
\text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \\
\text{Bootstrap:} \quad & \hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)
\end{aligned}$$

[H & Miller 2019] BayesBag includes **sampling and model-based uncertainty**: for $\vartheta^* \sim \pi^*(\cdot \,|\, Y)$,

$$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- If model correct, $m = 2n$ yields **correct uncertainty**:

$$\mathrm{Cov}(\vartheta^* \,|\, Y) \approx \Sigma_{\mathsf{S}}/n$$

- No matter what, $m = n$ ensures **conservative uncertainty**:

$$\mathrm{Cov}(\vartheta^* \,|\, Y) \geq \Sigma_{\mathsf{S}}/n$$

The bagged posterior is conservative

$$
\begin{aligned}
\text{MLE:} \quad & \hat{\theta}(Y) \approx \mathcal{N}(\theta_{\mathsf{opt}}, \Sigma_{\mathsf{S}}/n) \\
\text{Posterior:} \quad & \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n) \\
\text{Bootstrap:} \quad & \hat{\theta}(Y^*) \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{S}}/m)
\end{aligned}
$$

[H & Miller 2019] BayesBag includes **sampling and model-based uncertainty**: for $\vartheta^* \sim \pi^*(\cdot \,|\, Y)$,

$$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- If model correct, $m = 2n$ yields **correct uncertainty**:

$$\mathrm{Cov}(\vartheta^* \,|\, Y) \approx \Sigma_{\mathsf{S}}/n$$

- No matter what, $m = n$ ensures **conservative uncertainty**:

$$\mathrm{Cov}(\vartheta^* \,|\, Y) \geq \Sigma_{\mathsf{S}}/n$$

- Remember: we are ignoring finite-sample benefits of Bayes

Agenda

- BayesBag for parameter inference (and prediction)

➡ **BayesBag methodology**

- BayesBag for model selection

Selecting the optimal bootstrap dataset size

Posterior: $$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

BayesBag: $$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

[**H** & Miller 2019]

Selecting the optimal bootstrap dataset size

$$\text{Posterior:} \qquad \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

$$\text{BayesBag:} \qquad \vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- $\mathrm{Var}(\vartheta) =$ posterior variance $\approx \Sigma_{\mathsf{M}}/n$

[**H** & Miller 2019]

Selecting the optimal bootstrap dataset size

$$\text{Posterior:} \qquad \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

$$\text{BayesBag:} \qquad \vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- $\mathrm{Var}(\vartheta) =$ posterior variance $\approx \Sigma_{\mathsf{M}}/n$
- $\mathrm{Var}(\vartheta^*) =$ BayesBag variance $\approx (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/n$

[**H** & Miller 2019]

Selecting the optimal bootstrap dataset size

$$\text{Posterior:} \qquad \vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

$$\text{BayesBag:} \qquad \vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- $\mathrm{Var}(\vartheta) =$ posterior variance $\approx \Sigma_{\mathsf{M}}/n$
- $\mathrm{Var}(\vartheta^*) =$ BayesBag variance $\approx (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/n$

} *computable*

[**H** & Miller 2019]

Selecting the optimal bootstrap dataset size

Posterior: $$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

BayesBag: $$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- $\mathrm{Var}(\vartheta) =$ posterior variance $\approx \Sigma_{\mathsf{M}}/n$
- $\mathrm{Var}(\vartheta^*) =$ BayesBag variance $\approx (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/n$

} *computable*

- optimal bootstrap sample size m_{opt} estimate

$$\hat{m}_{\mathsf{opt}} = \frac{\mathrm{Var}(\vartheta^*)}{\mathrm{Var}(\vartheta^*) - \mathrm{Var}(\vartheta)} \times n$$

[**H** & Miller 2019]

Selecting the optimal bootstrap dataset size

Posterior: $$\vartheta \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), \Sigma_{\mathsf{M}}/n)$$

BayesBag: $$\vartheta^* \,|\, Y \approx \mathcal{N}(\hat{\theta}(Y), (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/m)$$

- $\mathrm{Var}(\vartheta) =$ posterior variance $\approx \Sigma_{\mathsf{M}}/n$
- $\mathrm{Var}(\vartheta^*) =$ BayesBag variance $\approx (\Sigma_{\mathsf{M}} + \Sigma_{\mathsf{S}})/n$

} *computable*

- optimal bootstrap sample size m_{opt} estimate

$$\hat{m}_{\mathsf{opt}} = \frac{\mathrm{Var}(\vartheta^*)}{\mathrm{Var}(\vartheta^*) - \mathrm{Var}(\vartheta)} \times n$$

$\rightarrow$ finite-sample version also available using a Gaussian model approximation

[**H** & Miller 2019]

BayesBag in practice

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$
   - e.g., use MCMC to get approximate samples $\theta_{(1)}, \ldots, \theta_{(T)}$ from $\pi(\cdot \,|\, Y)$

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$

   – e.g., use MCMC to get approximate samples $\theta_{(1)}, \ldots, \theta_{(T)}$ from $\pi(\cdot \,|\, Y)$

2) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = n$ using $B \approx 50$ bootstrap datasets

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$

- e.g., use MCMC to get approximate samples $\theta_{(1)}, \ldots, \theta_{(T)}$ from $\pi(\cdot \,|\, Y)$

2) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = n$ using $B \approx 50$ bootstrap datasets

- e.g., use MCMC to get approximate samples $\theta^*_{(b,1)}, \ldots, \theta^*_{(b,T)}$ from $\pi(\cdot \,|\, Y^*_{(b)})$ for $b = 1, \ldots, B$

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$

- e.g., use MCMC to get approximate samples $\theta_{(1)}, \ldots, \theta_{(T)}$ from $\pi(\cdot \,|\, Y)$

2) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = n$ using $B \approx 50$ bootstrap datasets

- e.g., use MCMC to get approximate samples $\theta^*_{(b,1)}, \ldots, \theta^*_{(b,T)}$ from $\pi(\cdot \,|\, Y^*_{(b)})$ for $b = 1, \ldots, B$

if Gaussian approximation to standard and bagged posteriors decent **then**

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$

- e.g., use MCMC to get approximate samples $\theta_{(1)}, \ldots, \theta_{(T)}$ from $\pi(\cdot \,|\, Y)$

2) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = n$ using $B \approx 50$ bootstrap datasets

- e.g., use MCMC to get approximate samples $\theta^*_{(b,1)}, \ldots, \theta^*_{(b,T)}$ from $\pi(\cdot \,|\, Y^*_{(b)})$ for $b = 1, \ldots, B$

if Gaussian approximation to standard and bagged posteriors decent **then**

3a) compute optimal bootstrap sample size estimate $\hat{m}_{\text{opt}}$

3b) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = \hat{m}_{\text{opt}}$ and output

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \,|\, Y)$

- e.g., use MCMC to get approximate samples $\theta_{(1)}, \ldots, \theta_{(T)}$ from $\pi(\cdot \,|\, Y)$

2) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = n$ using $B \approx 50$ bootstrap datasets

- e.g., use MCMC to get approximate samples $\theta^*_{(b,1)}, \ldots, \theta^*_{(b,T)}$ from $\pi(\cdot \,|\, Y^*_{(b)})$ for $b = 1, \ldots, B$

if Gaussian approximation to standard and bagged posteriors decent **then**

3a) compute optimal bootstrap sample size estimate $\hat{m}_{\text{opt}}$

3b) compute bagged posterior $\pi^*(\cdot \,|\, Y)$ with $m = \hat{m}_{\text{opt}}$ and output

else

4) output bagged posterior with $m = n$ computed in step 2

[**H** & Miller 2019]

Diagnosing model–data mismatch

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty
- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
    - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty

- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$

- Provides interpretable diagnostic for **model criticism**

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty

- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$

- Provides interpretable diagnostic for **model criticism**
  - $-1 < \mathcal{I} < 1$

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty

- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$

- Provides interpretable diagnostic for **model criticism**
  - $-1 < \mathcal{I} < 1$
  - $\mathcal{I} \approx 0$: no disagreement

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty

- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$

- Provides interpretable diagnostic for **model criticism**
  - $-1 < \mathcal{I} < 1$
  - $\mathcal{I} \approx 0$: no disagreement
  - $\mathcal{I} > 0$: posterior overconfident

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty

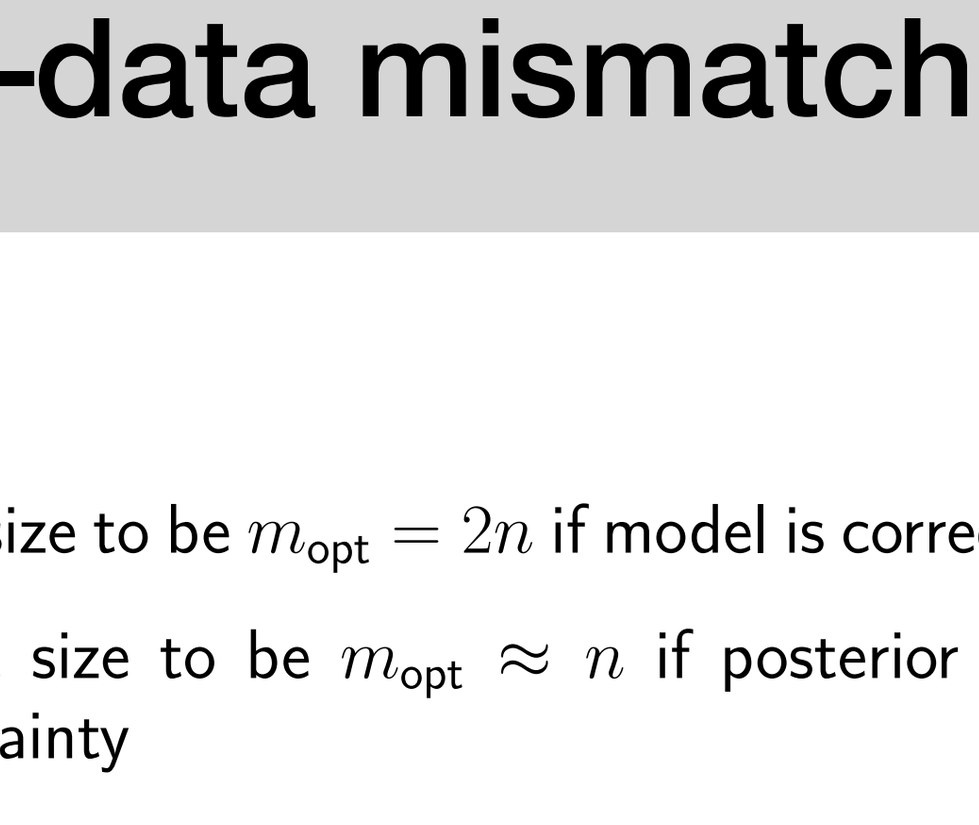
- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$

- Provides interpretable diagnostic for **model criticism**
  - $-1 < \mathcal{I} < 1$
  - $\mathcal{I} \approx 0$: no disagreement
  - $\mathcal{I} > 0$: posterior overconfident
  - $\mathcal{I} < 0$: posterior under-confident

[**H** & Miller 2019]

Diagnosing model–data mismatch

- Recall:
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} = 2n$ if model is correct
  - expect optimal bootstrap dataset size to be $m_{\text{opt}} \approx n$ if posterior is drastically underestimating uncertainty
- Define **model–data mismatch index** $\mathcal{I} = 2n/\hat{m}_{\text{opt}} - 1$
- Provides interpretable diagnostic for **model criticism**
  - $-1 < \mathcal{I} < 1$
  - $\mathcal{I} \approx 0$: no disagreement
  - $\mathcal{I} > 0$: posterior overconfident
  - $\mathcal{I} < 0$: posterior under-confident



[H & Miller 2019]

Simulations: linear regression

[**H** & Miller 2019]

Simulations: linear regression

- Assumed model: linear regression with a conjugate prior on $\theta = (\sigma^2, \beta)$

$$Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2) \qquad\qquad Y_i = (X_i, Z_i)$$

[**H** & Miller 2019]

Simulations: linear regression

- Assumed model: linear regression with a conjugate prior on $\theta = (\sigma^2, \beta)$

$$Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2) \qquad Y_i = (X_i, Z_i)$$

outcome

[**H** & Miller 2019]

Simulations: linear regression

- Assumed model: linear regression with a conjugate prior on $\theta = (\sigma^2, \beta)$

$$Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2) \qquad Y_i = (X_i, Z_i)$$

outcome *covariates*

[**H** & Miller 2019]

Simulations: linear regression

- Assumed model: linear regression with a conjugate prior on $\theta = (\sigma^2, \beta)$

$$Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2) \qquad Y_i = (X_i, Z_i)$$

Z_i: outcome; X_i: covariates

- True model P_{true} (that we simulate from):

$$Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1) \qquad X_i \sim G$$

[**H** & Miller 2019]

Simulations: linear regression

- Assumed model: linear regression with a conjugate prior on $\theta = (\sigma^2, \beta)$

$$Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2) \qquad Y_i = (X_i, Z_i)$$

Z_i: *outcome*; X_i: *covariates*

- True model P_{true} (that we simulate from):

$$Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1) \qquad X_i \sim G$$

- $n = 50$, $d = 10$, and $m = \hat{m}_{\text{opt}}$

[**H** & Miller 2019]

Simulations: linear regression

- Assumed model: linear regression with a conjugate prior on $\theta = (\sigma^2, \beta)$

$$Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2) \qquad Y_i = (X_i, Z_i)$$

outcome *covariates*

- True model P_{true} (that we simulate from):

$$Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1) \qquad X_i \sim G$$

- $n = 50$, $d = 10$, and $m = \hat{m}_{\text{opt}}$

- Important: when model misspecified, $\theta_{\text{opt}} \neq (1, \beta_{\text{gen}})$ (in general)

[**H** & Miller 2019]

Better parameter inference with BayesBag

Assumed model: $Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\mathsf{gen}}, 1), \qquad X_i \sim G$

[**H** & Miller 2019]

Better parameter inference with BayesBag

Assumed model: $Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\mathsf{gen}}, 1), \qquad X_i \sim G$

- Performance metric is difference in log posterior density at θ_{opt}:

$$\log \pi^*(\theta_{\mathsf{opt}} \,|\, Y) - \log \pi(\theta_{\mathsf{opt}} \,|\, Y)$$

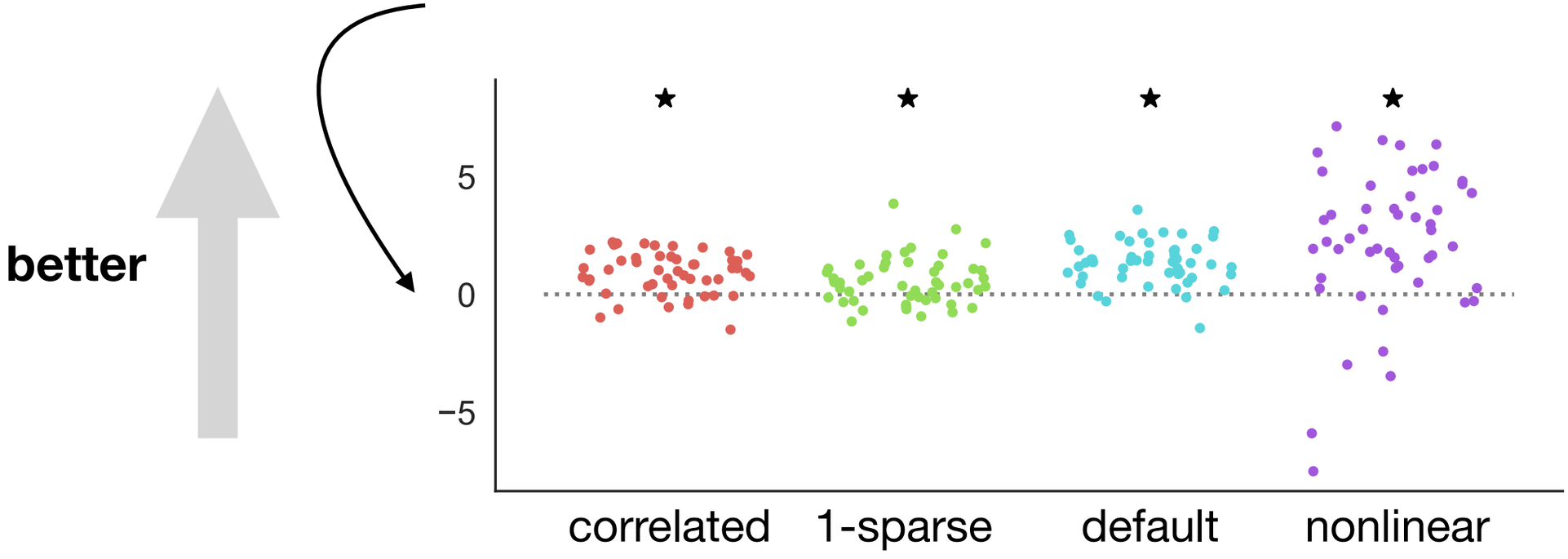
[**H** & Miller 2019]

Better parameter inference with BayesBag

$$\text{Assumed model:} \qquad Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$$
$$\text{True model:} \qquad Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$$

- Performance metric is difference in log posterior density at θ_{opt}:

$$\log \pi^*(\theta_{\text{opt}} \,|\, Y) - \log \pi(\theta_{\text{opt}} \,|\, Y)$$



[**H** & Miller 2019]

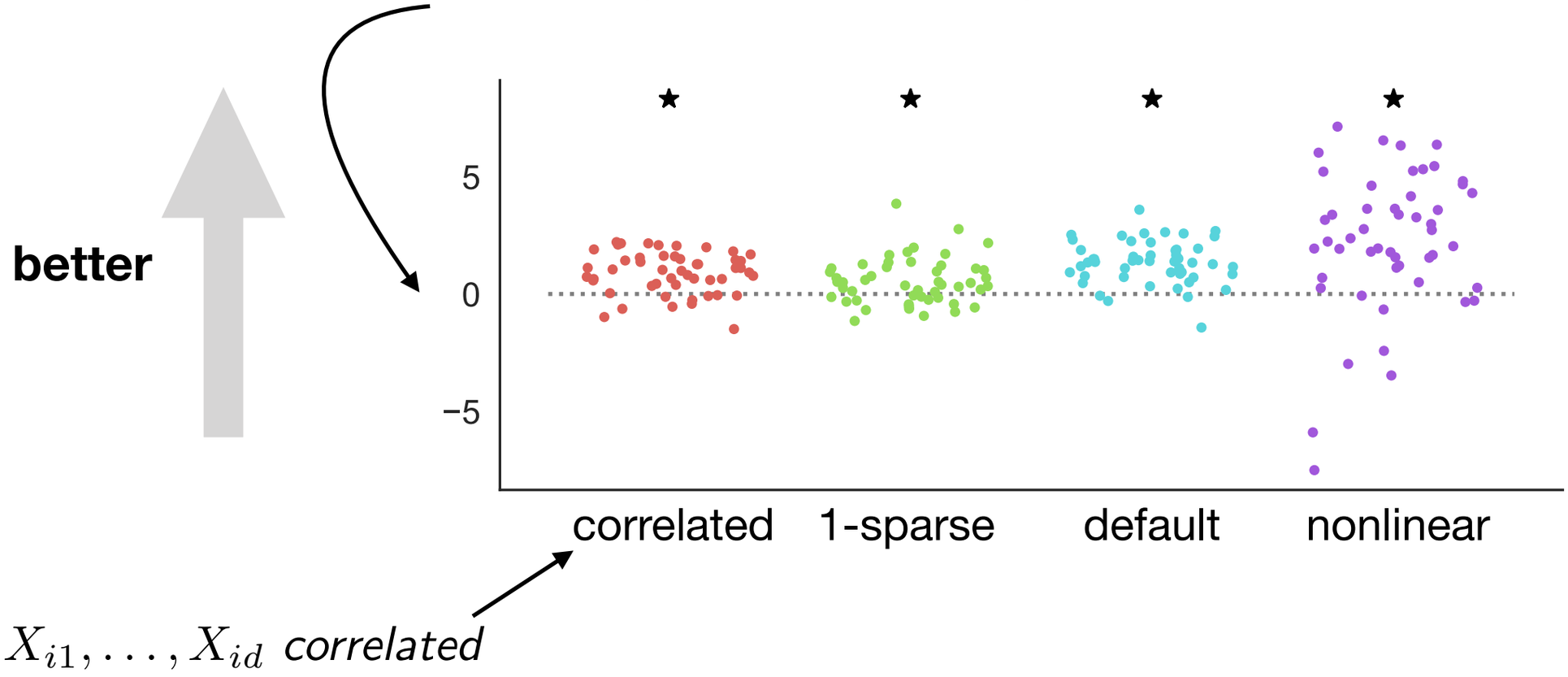
Better parameter inference with BayesBag

Assumed model: $Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$

- Performance metric is difference in log posterior density at θ_{opt}:

$$\log \pi^*(\theta_{\text{opt}} \,|\, Y) - \log \pi(\theta_{\text{opt}} \,|\, Y)$$



[**H** & Miller 2019]

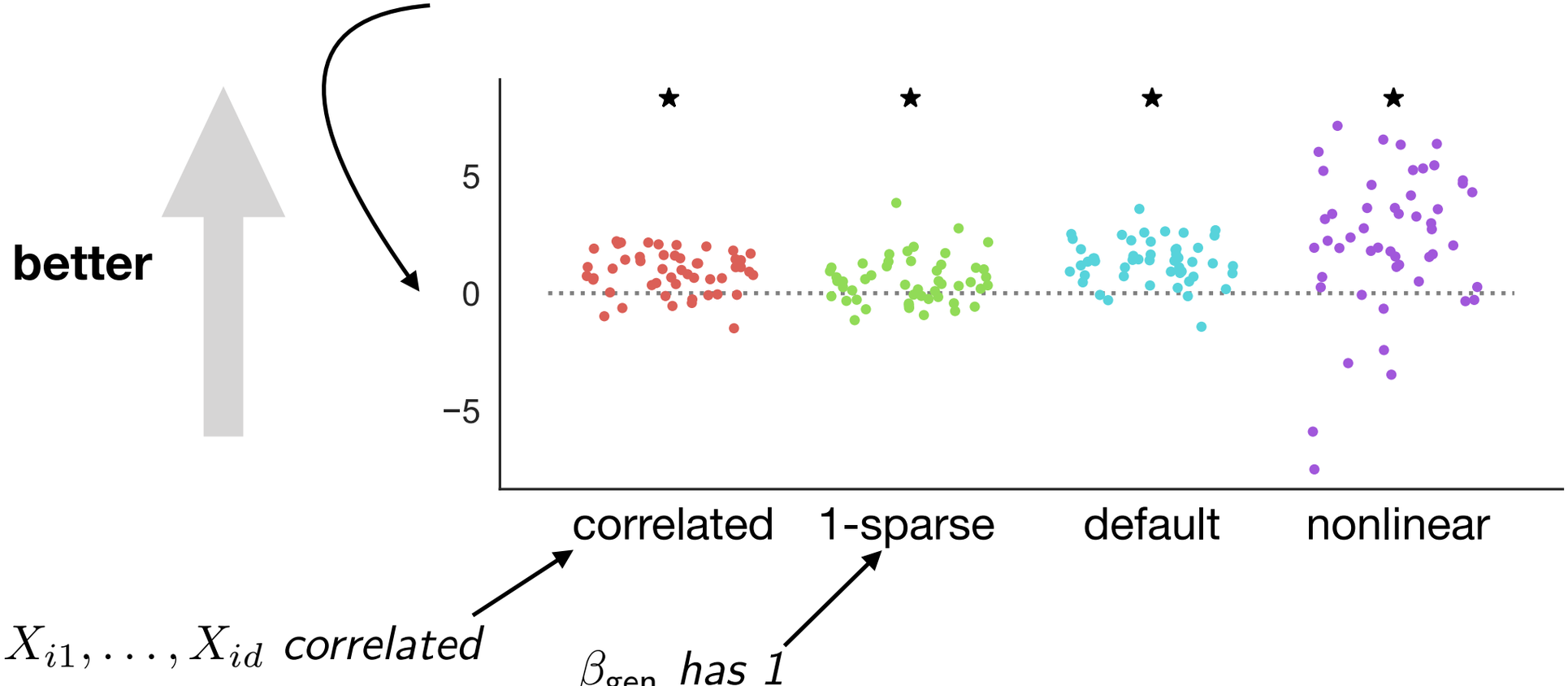
Better parameter inference with BayesBag

Assumed model: $Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$

- Performance metric is difference in log posterior density at θ_{opt}:

$$\log \pi^*(\theta_{\text{opt}} \,|\, Y) - \log \pi(\theta_{\text{opt}} \,|\, Y)$$



[**H** & Miller 2019]

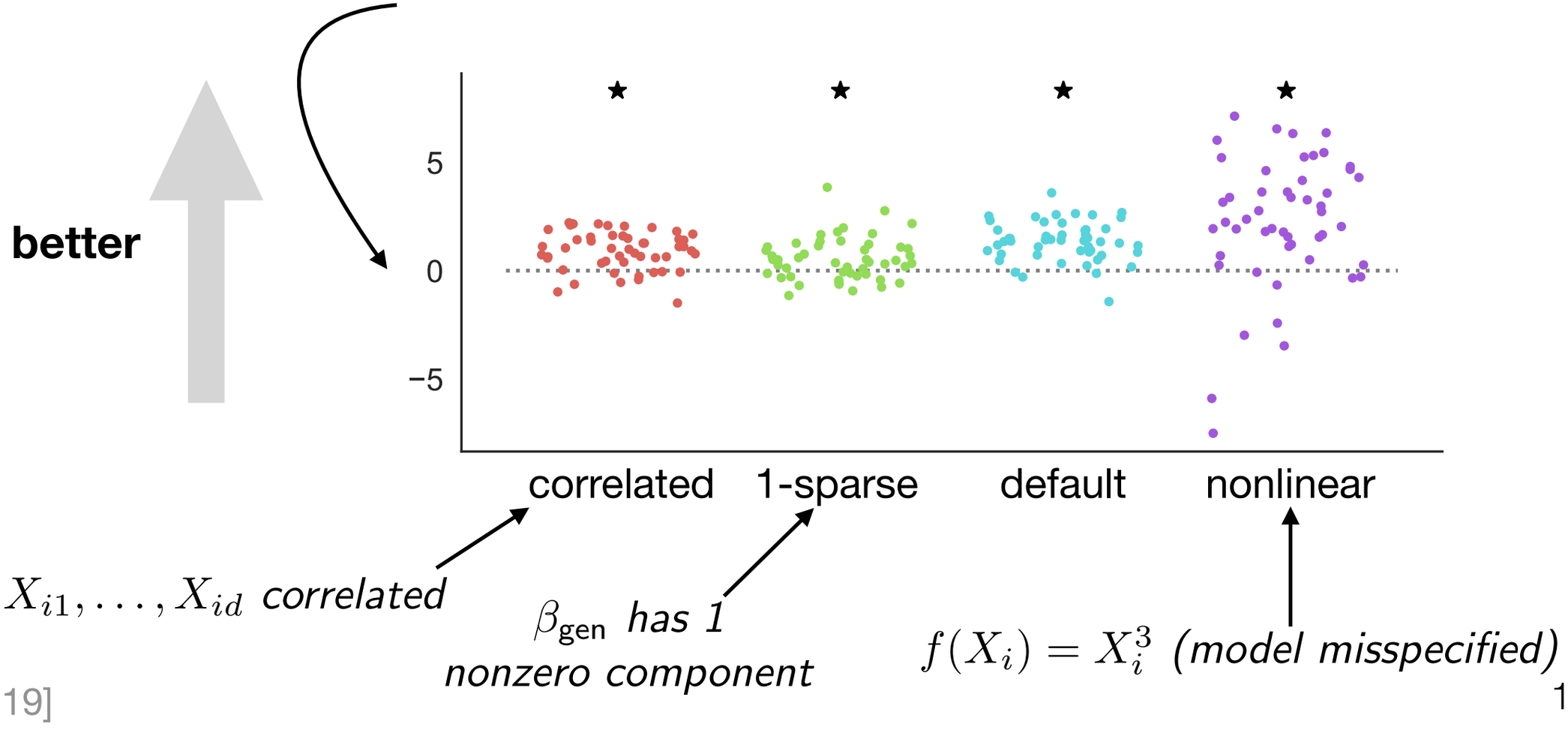
Better parameter inference with BayesBag

Assumed model: $Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$

- Performance metric is difference in log posterior density at θ_{opt}:

$$\log \pi^*(\theta_{\text{opt}} \,|\, Y) - \log \pi(\theta_{\text{opt}} \,|\, Y)$$



[**H** & Miller 2019]

Mismatch index detects problems with likelihood and prior

$$\text{Assumed model:} \qquad Z_i \,|\, X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$$

$$\text{True model:} \qquad Z_i \,|\, X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\mathsf{gen}}, 1), \qquad X_i \sim G$$

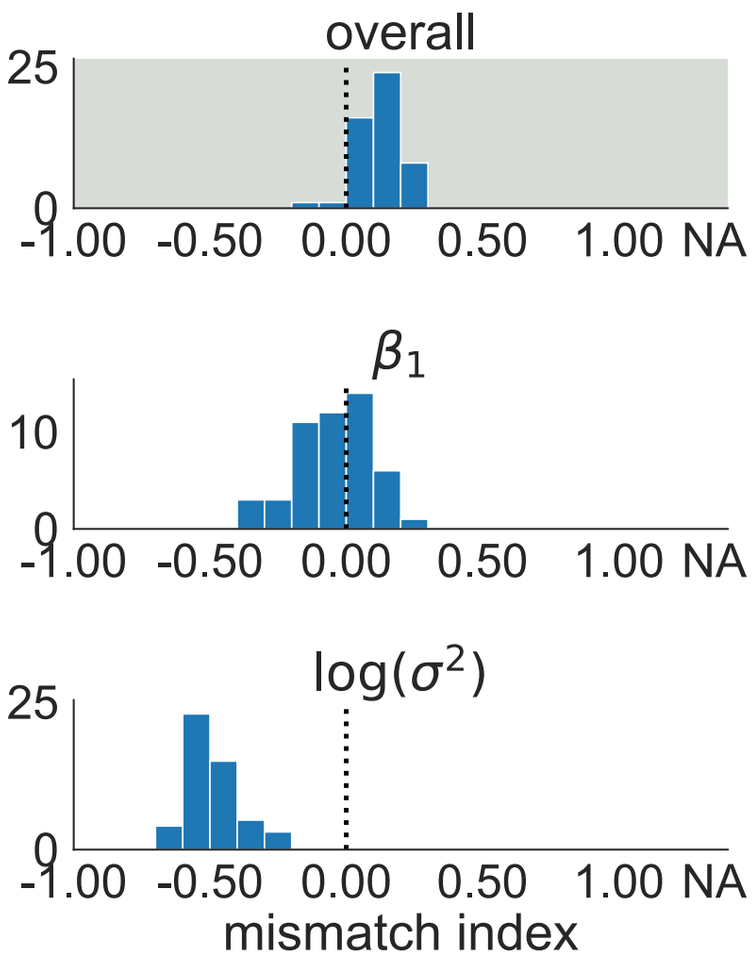
[**H** & Miller 2019]

Mismatch index detects problems with likelihood and prior

Assumed model: $Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\mathsf{gen}}, 1), \qquad X_i \sim G$

Assumed model correct



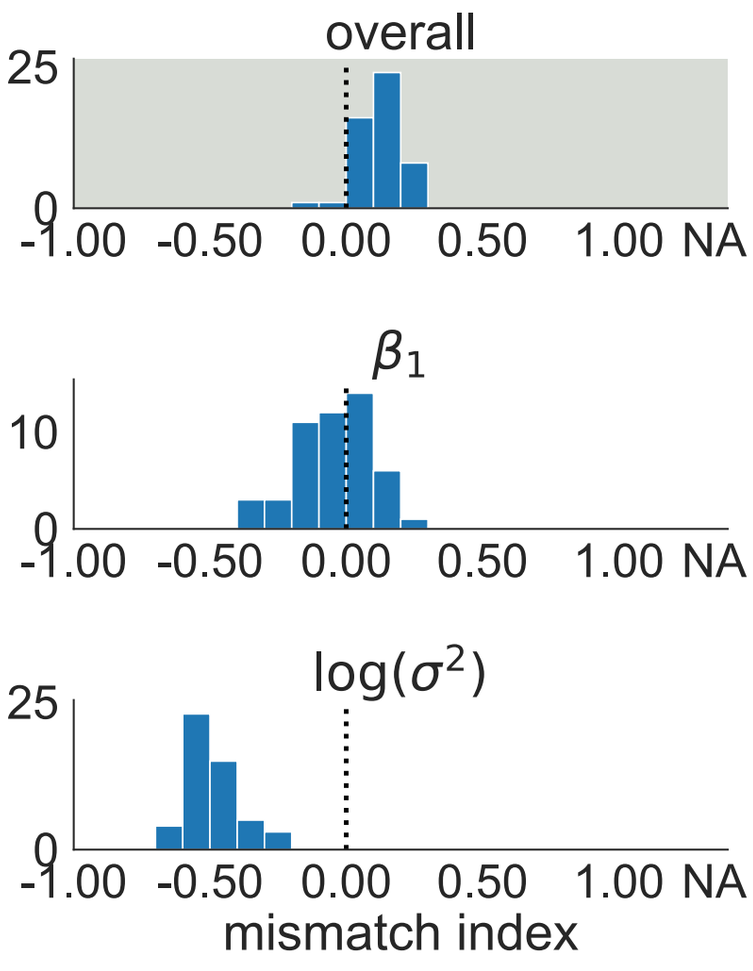
[**H** & Miller 2019]

Mismatch index detects problems with likelihood and prior

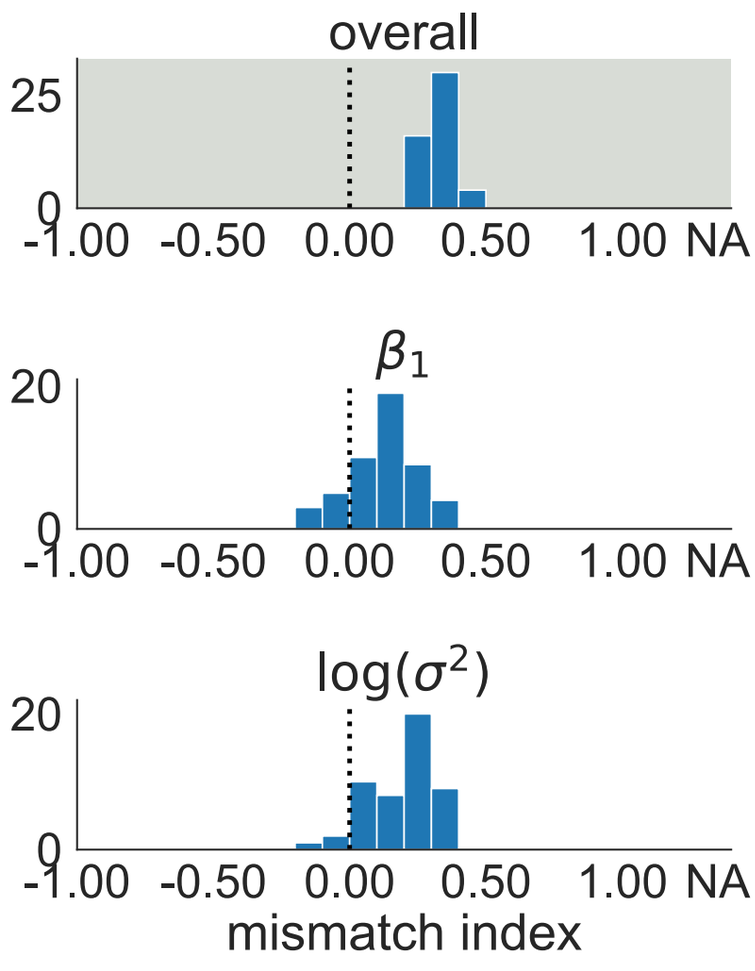
Assumed model: $Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$

Assumed model correct



Prior off (1-sparse)



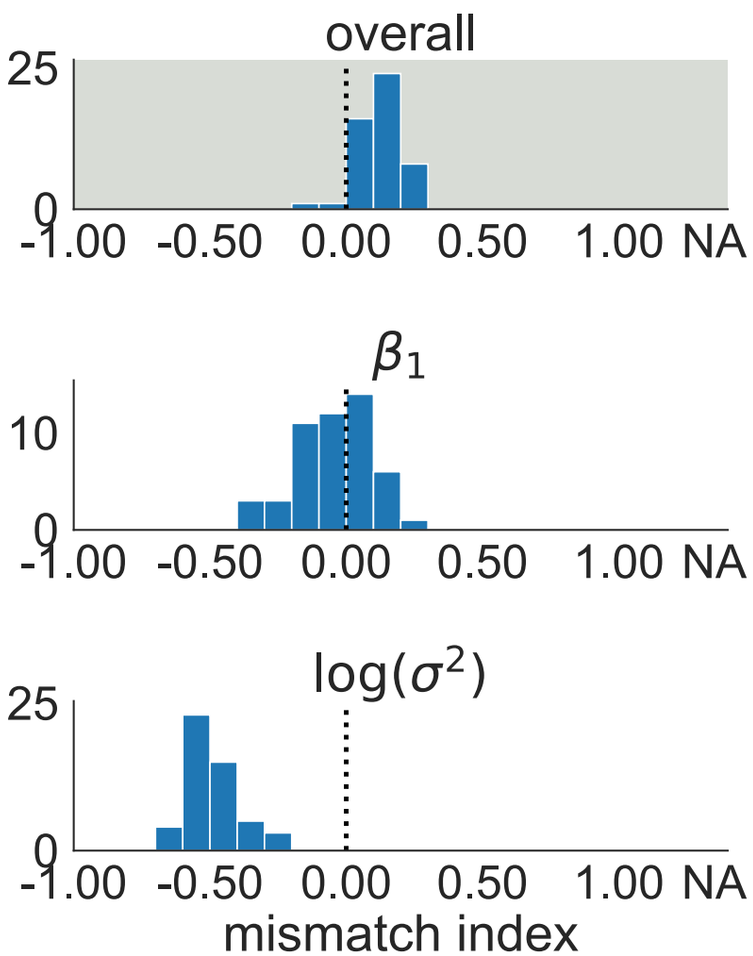
[**H** & Miller 2019]

Mismatch index detects problems with likelihood and prior

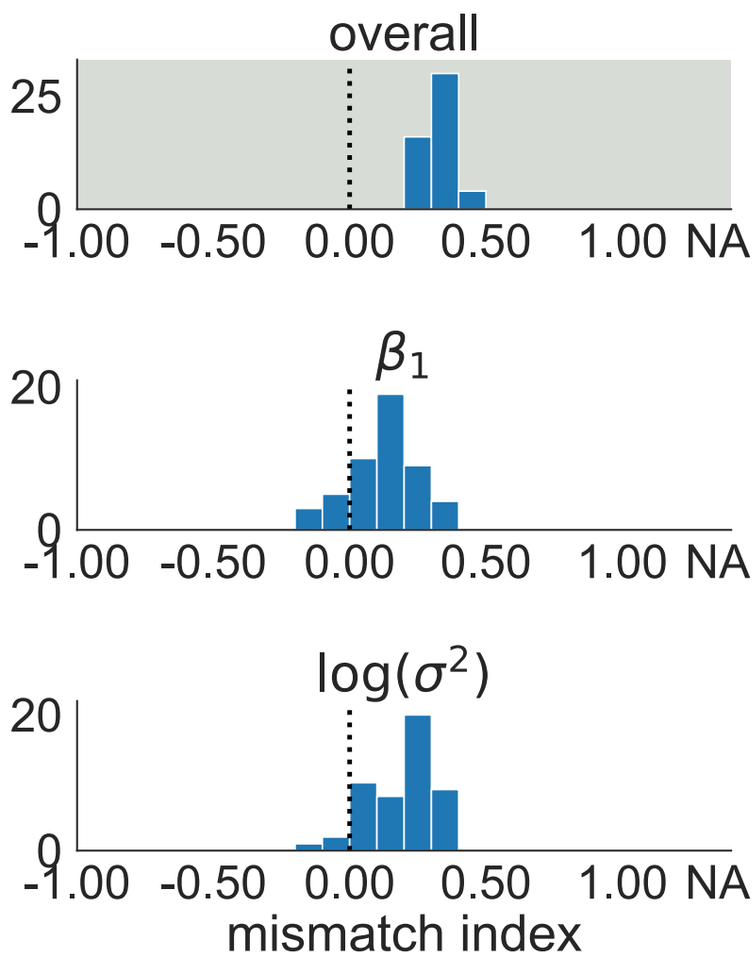
Assumed model: $Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$

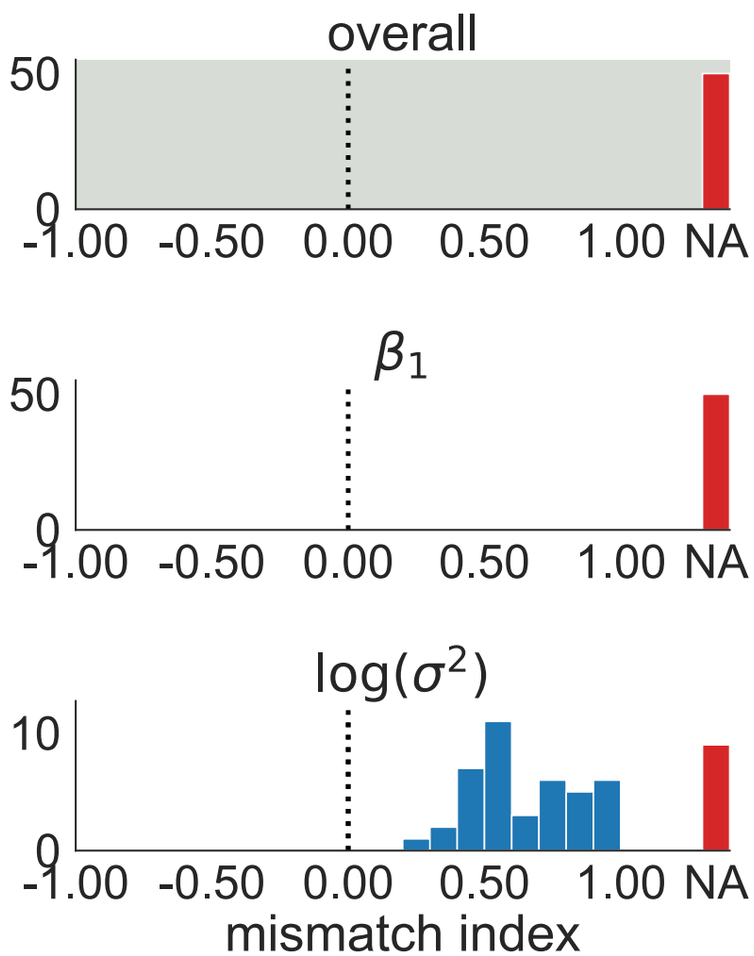
Assumed model correct



Prior off (1-sparse)



Likelihood wrong (nonlinear)



[**H** & Miller 2019]

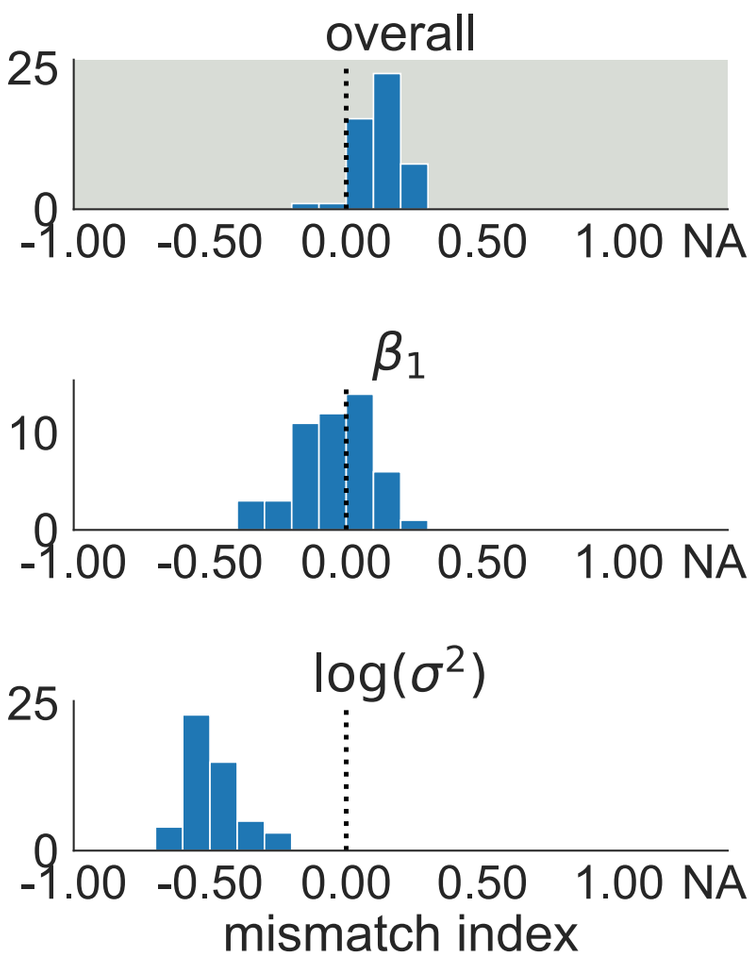
Mismatch index detects problems with likelihood and prior

something is very wrong

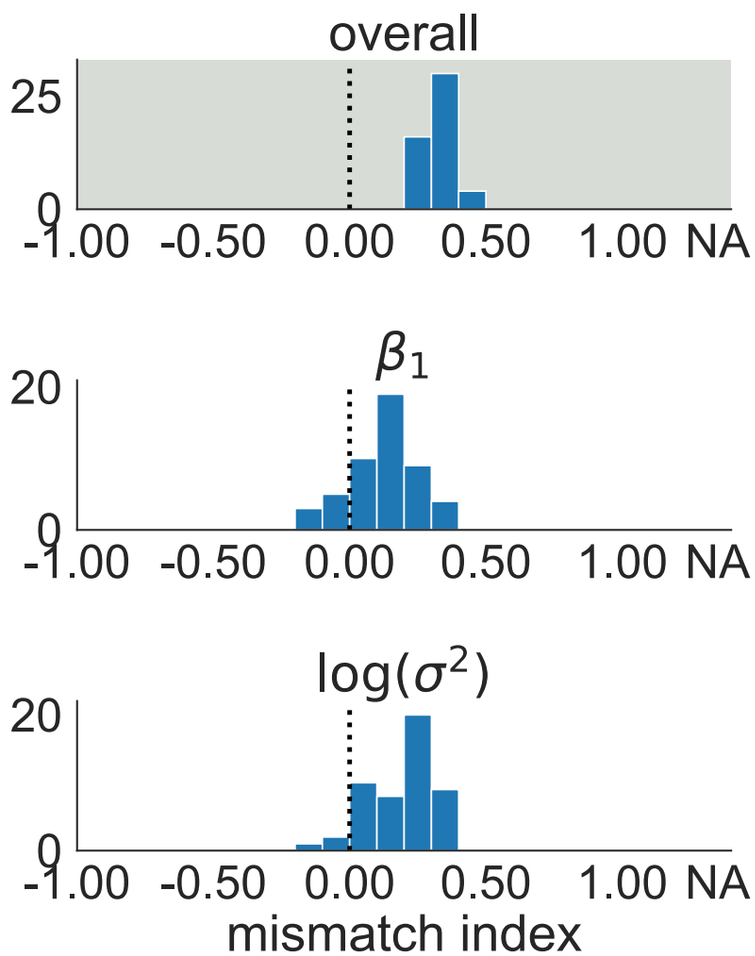
Assumed model: $Z_i \mid X_i, \beta, \sigma^2 \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$

True model: $Z_i \mid X_i \sim \mathcal{N}(f(X_i)^\top \beta_{\text{gen}}, 1), \qquad X_i \sim G$

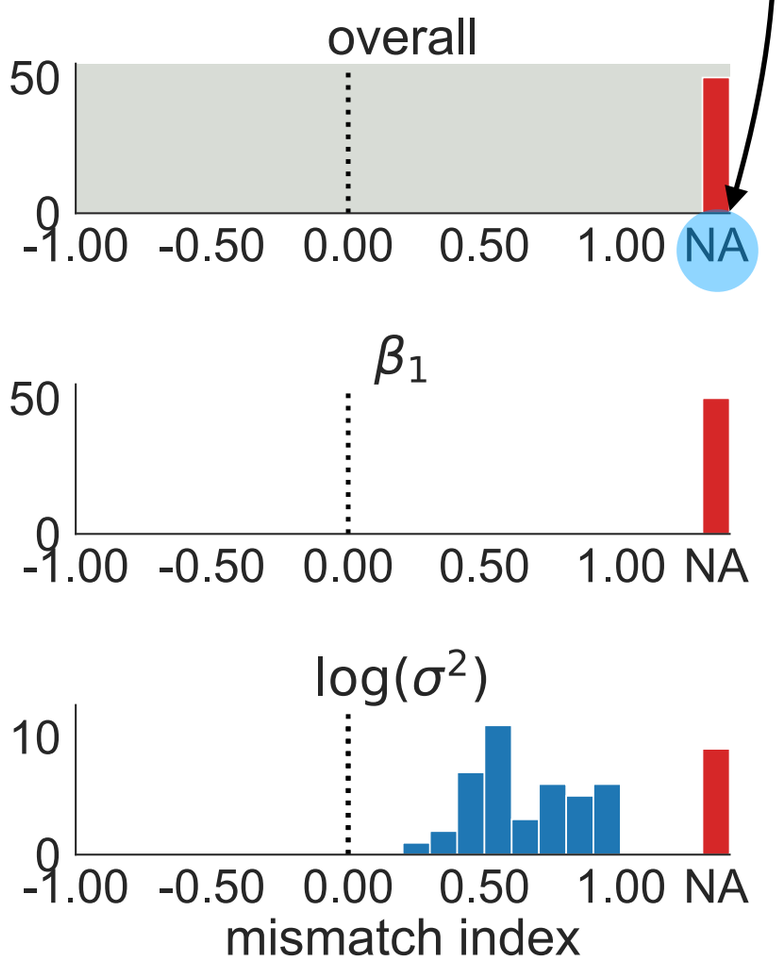
Assumed model correct



Prior off (1-sparse)



Likelihood wrong (nonlinear)



[H & Miller 2019]

Agenda

- BayesBag for parameter inference (and prediction)
- BayesBag methodology

➡ **BayesBag for model selection**

Bayesian model selection

Bayesian model selection

- **Goal:** based on data Y, select between a (finite or countable) set of models $M = \{m_1, m_2, ...\}$

Bayesian model selection

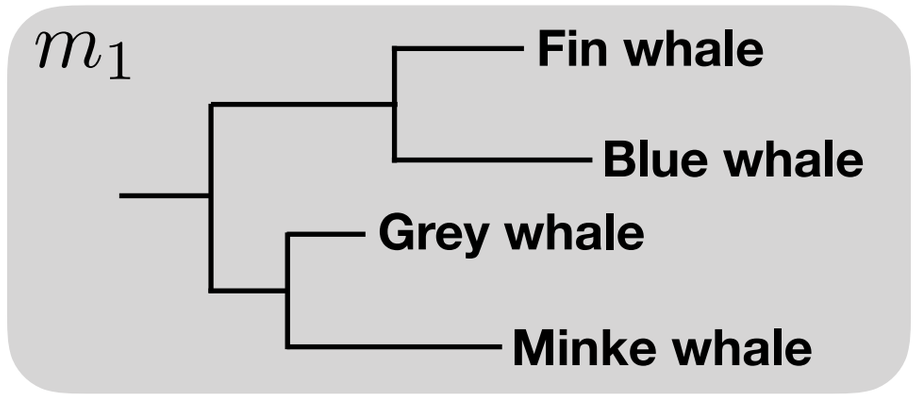
- **Goal:** based on data Y, select between a (finite or countable) set of models $M = \{m_1, m_2, ...\}$
- **Example:** systematics

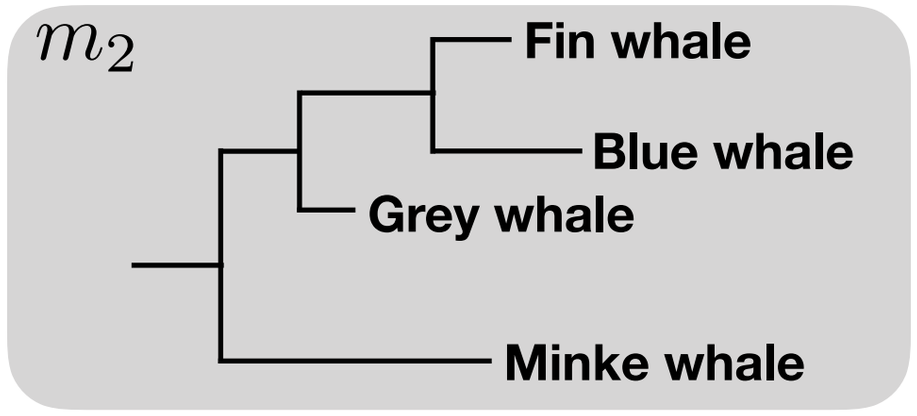
Bayesian model selection

- **Goal:** based on data Y, select between a (finite or countable) set of models $M = \{m_1, m_2, ...\}$

- **Example:** systematics

  - **Goal:** learn about evolutionary history of a set of species **[e.g. whales]**



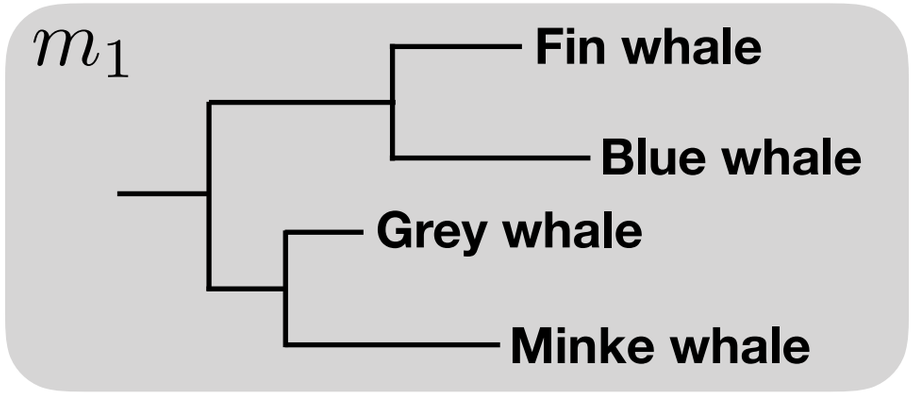


.
.
.

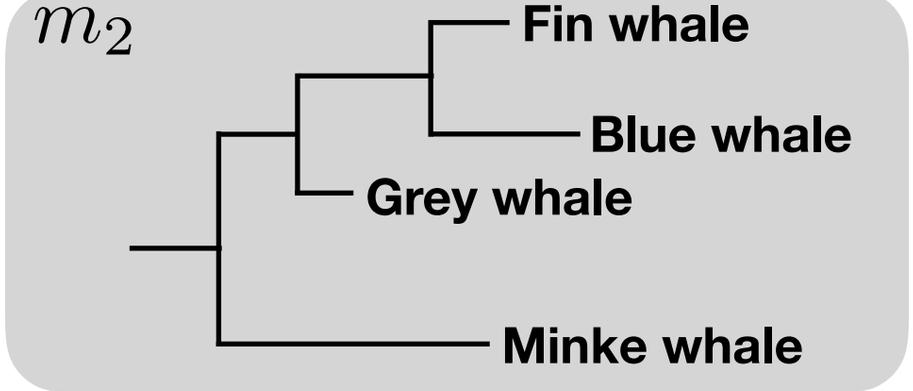
Bayesian model selection

- **Goal:** based on data Y, select between a (finite or countable) set of models $M = \{m_1, m_2, ...\}$
- **Example:** systematics
  - **Goal:** learn about evolutionary history of a set of species **[e.g. whales]**
  - **Approach:** infer which phylogenetic trees are consistent with observed species characteristics Y **[e.g. genetic data, physical features such as coloring]**

$\pi(m_1 \mid Y) = .8$



$\pi(m_2 \mid Y) = .1$

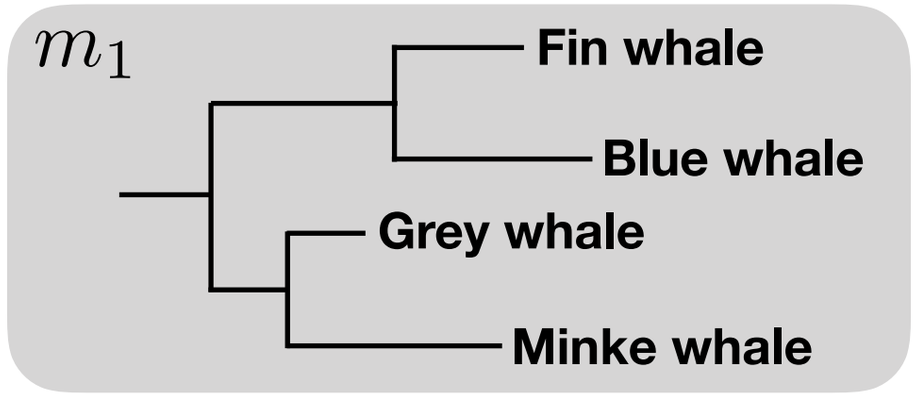


.
.
.

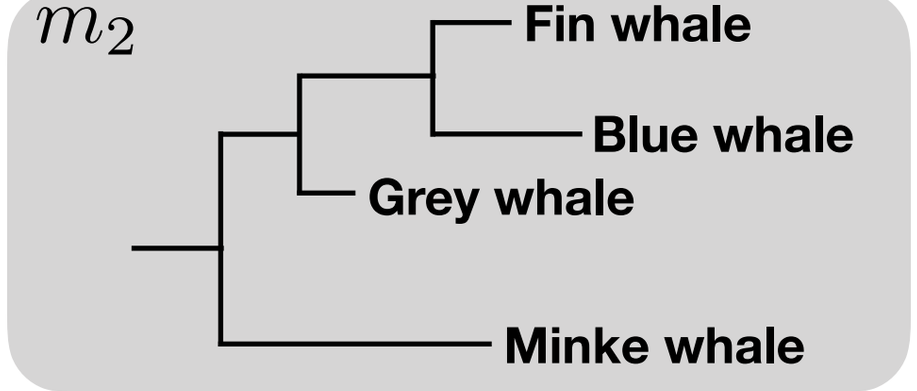
Bayesian model selection

- **Goal:** based on data Y, select between a (finite or countable) set of models $M = \{m_1, m_2, ...\}$
- **Example:** systematics
  - **Goal:** learn about evolutionary history of a set of species **[e.g. whales]**
  - **Approach:** infer which phylogenetic trees are consistent with observed species characteristics Y **[e.g. genetic data, physical features such as coloring]**
- **Problem:** Bayesian model selection still assumes some model in M is correct

$\pi(m_1 \mid Y) = .8$



$\pi(m_2 \mid Y) = .1$



.
.
.

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2

[Yang & Zhu 2018, **H** & Miller 2019]

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2
- Assume m_1 and m_2 explain the data-generating distribution equally well:
  $\mathbb{E}\{\log p(Y \mid m_1)\} = \mathbb{E}\{\log p(Y \mid m_2)\}$

[Yang & Zhu 2018, **H** & Miller 2019]

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2
- Assume m_1 and m_2 explain the data-generating distribution equally well:
  $\mathbb{E}\{\log p(Y \,|\, m_1)\} = \mathbb{E}\{\log p(Y \,|\, m_2)\}$
- Then, we **hope** models have equal posterior probability ($n \to \infty$):
  $\pi(m_1 \,|\, Y) = \pi(m_2 \,|\, Y) = 1/2$

[Yang & Zhu 2018, **H** & Miller 2019]

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2
- Assume m_1 and m_2 explain the data-generating distribution equally well:
  $\mathbb{E}\{\log p(Y \mid m_1)\} = \mathbb{E}\{\log p(Y \mid m_2)\}$
- Then, we **hope** models have equal posterior probability ($n \to \infty$):
  $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$
- However...

> **[Yang & Zhu 2018, H & Miller 2019]** $\pi(m_1 \mid Y) = 0$ or 1 with equal probability.

[Yang & Zhu 2018, **H** & Miller 2019]

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2
- Assume m_1 and m_2 explain the data-generating distribution equally well:
  $\mathbb{E}\{\log p(Y \mid m_1)\} = \mathbb{E}\{\log p(Y \mid m_2)\}$
- Then, we **hope** models have equal posterior probability ($n \to \infty$):
  $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$
- However...

> **[Yang & Zhu 2018, H & Miller 2019]** $\pi(m_1 \mid Y) = 0$ or 1 with equal probability.

all posterior mass on a single, arbitrary model

[Yang & Zhu 2018, **H** & Miller 2019]

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2
- Assume m_1 and m_2 explain the data-generating distribution equally well:
  $\mathbb{E}\{\log p(Y \mid m_1)\} = \mathbb{E}\{\log p(Y \mid m_2)\}$
- Then, we **hope** models have equal posterior probability ($n \to \infty$):
  $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$
- However...

[Yang & Zhu 2018, H & Miller 2019] $\pi(m_1 \mid Y) = 0$ or 1 with equal probability.

all posterior mass on a single, arbitrary model

[H & Miller 2019]
A) If $m = n$, then $\pi^*(m_1 \mid Y) \sim \text{Uniform}(0, 1)$
B) If $m/n \to 0$, then $\pi^*(m_1 \mid Y) \to 1/2$.

[Yang & Zhu 2018, H & Miller 2019]

BayesBag stabilizes Bayesian model selection

- Assume two models m_1 and m_2
- Assume m_1 and m_2 explain the data-generating distribution equally well:
  $\mathbb{E}\{\log p(Y \mid m_1)\} = \mathbb{E}\{\log p(Y \mid m_2)\}$
- Then, we **hope** models have equal posterior probability ($n \to \infty$):
  $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$
- However...

[Yang & Zhu 2018, H & Miller 2019] $\pi(m_1 \mid Y) = 0$ or 1 with equal probability.

all posterior mass on a single, arbitrary model

[H & Miller 2019]
A) If $m = n$, then $\pi^*(m_1 \mid Y) \sim \text{Uniform}(0, 1)$
B) If $m/n \to 0$, then $\pi^*(m_1 \mid Y) \to 1/2$.

bagged posterior mass more evenly distributed

Illustration: two normal models

- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$

[**H** & Miller 2019]

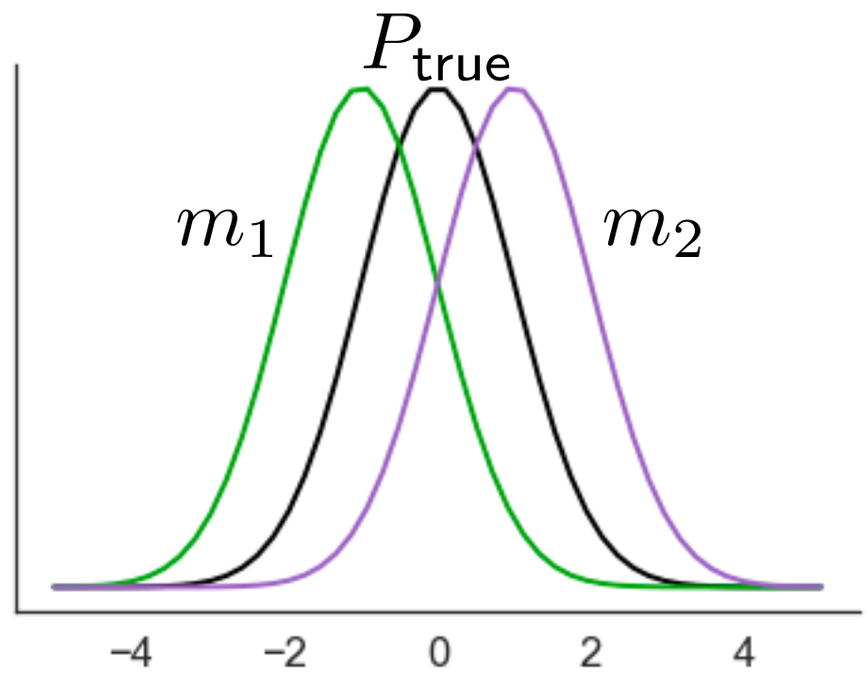
Illustration: two normal models

- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$

[**H** & Miller 2019]

Illustration: two normal models

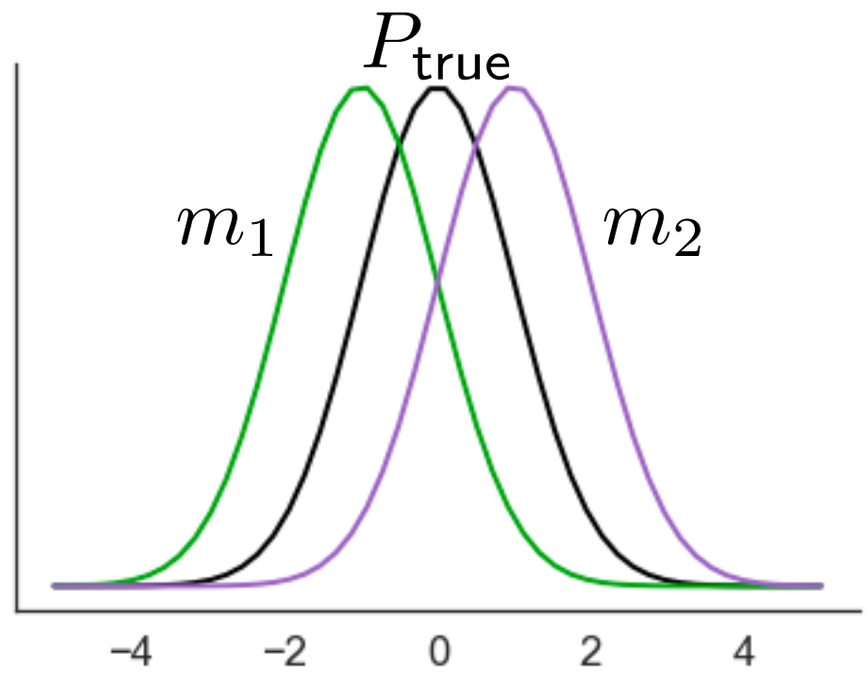
- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$



[**H** & Miller 2019]

Illustration: two normal models

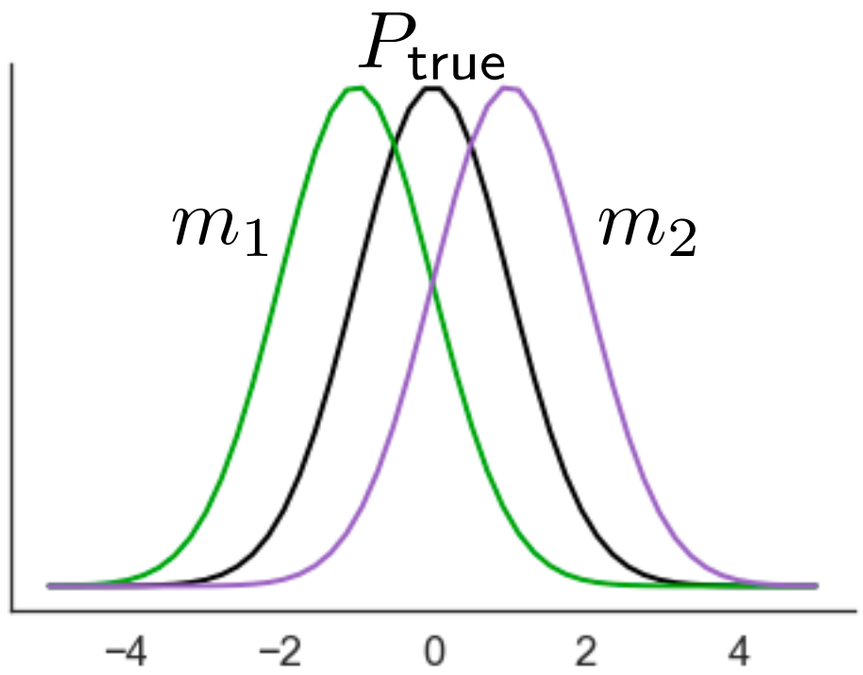
- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$
- Generate datasets $Y^{(1)}, Y^{(2)}, \ldots$ of size $n = 1000$, where $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



[**H** & Miller 2019]

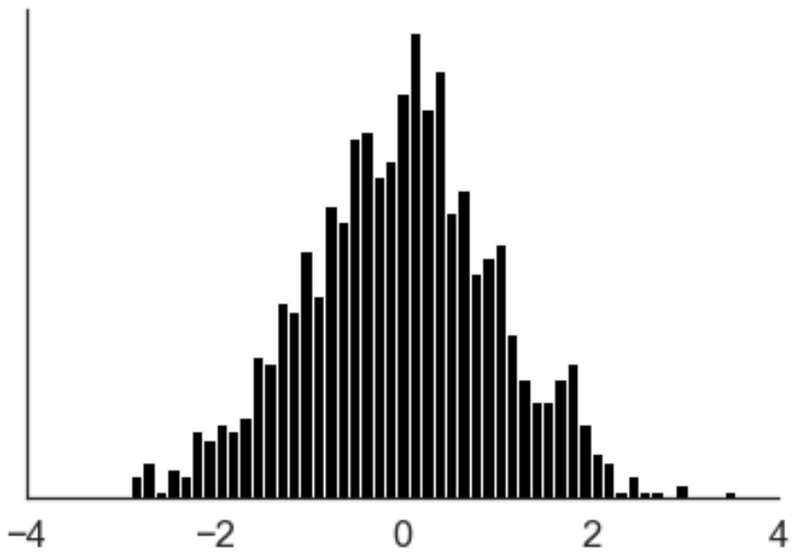
Illustration: two normal models

- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$
- Generate datasets $Y^{(1)}, Y^{(2)}, \ldots$ of size $n = 1000$, where $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



$$\pi(m_1 \mid Y^{(1)}) = 1$$

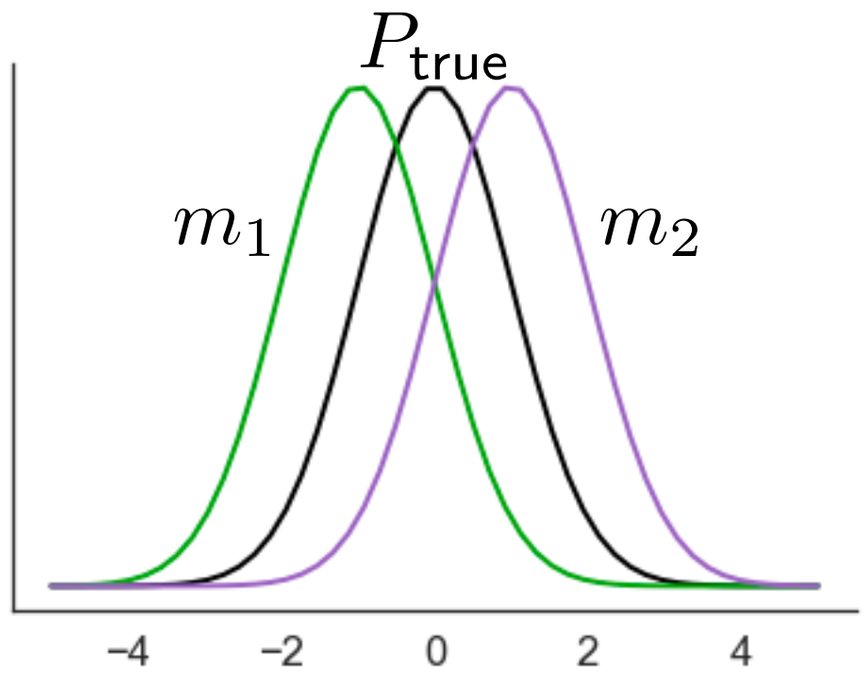
$$\pi^*(m_1 \mid Y^{(1)}) = 0.82$$



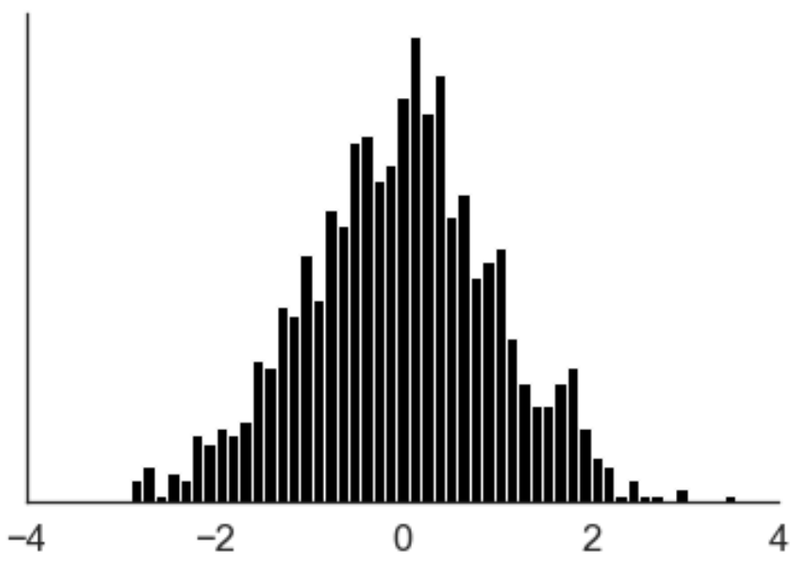
[**H** & Miller 2019]

Illustration: two normal models

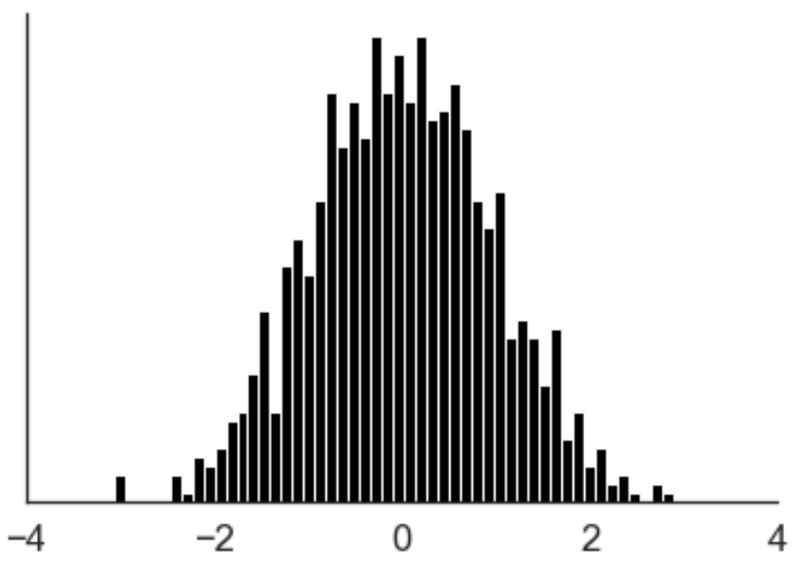
- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$
- Generate datasets $Y^{(1)}, Y^{(2)}, \ldots$ of size $n = 1000$, where $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



$$\pi(m_1 \mid Y^{(1)}) = 1$$
$$\pi^*(m_1 \mid Y^{(1)}) = 0.82$$



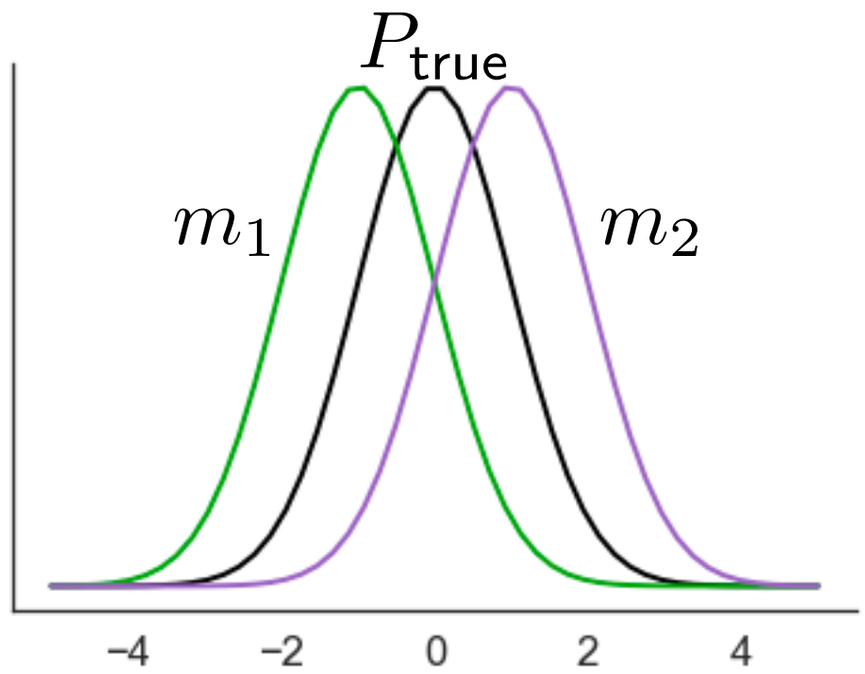
$$\pi(m_1 \mid Y^{(2)}) = 10^{-5}$$
$$\pi^*(m_1 \mid Y^{(2)}) = 0.38$$



[**H** & Miller 2019]

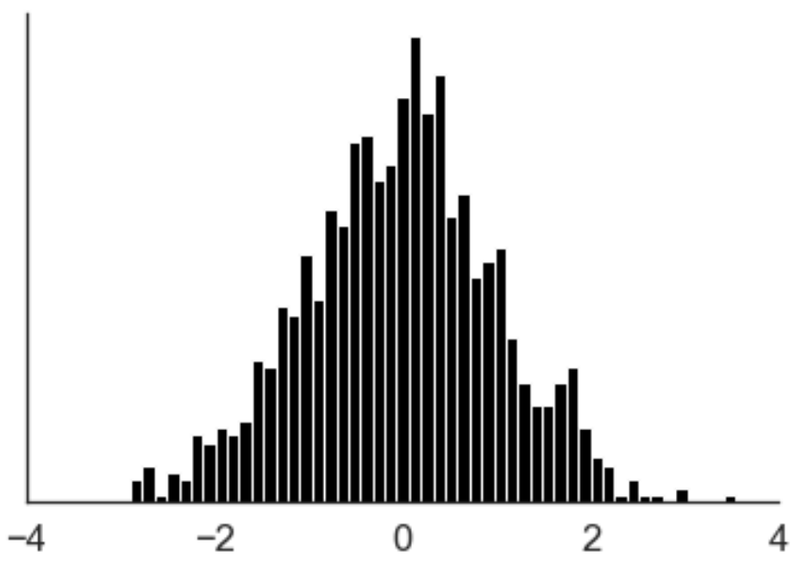
Illustration: two normal models

- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$
- Generate datasets $Y^{(1)}, Y^{(2)}, \ldots$ of size $n = 1000$, where $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



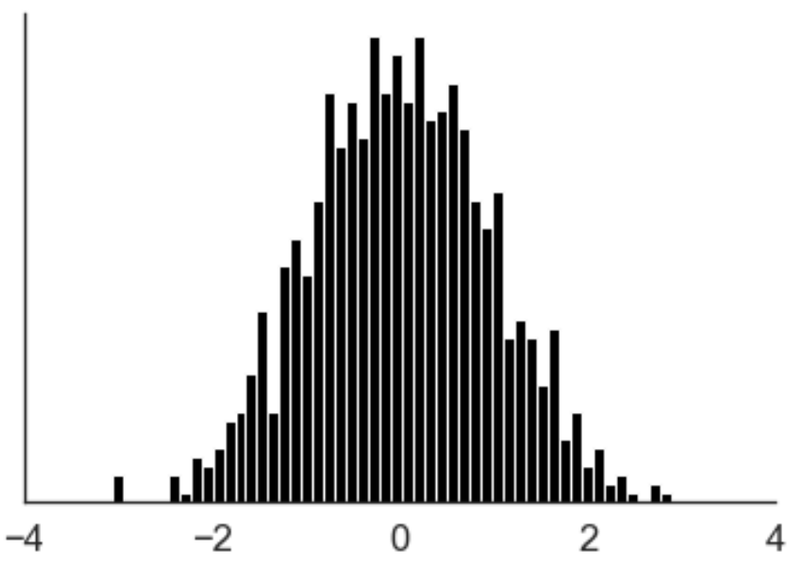
$\pi(m_1 \mid Y^{(1)}) = 1$

$\pi^*(m_1 \mid Y^{(1)}) = 0.82$



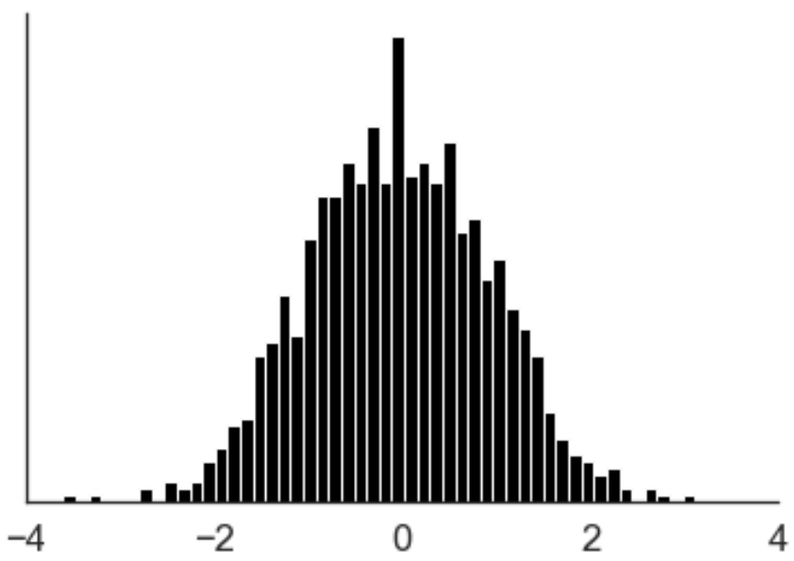
$\pi(m_1 \mid Y^{(2)}) = 10^{-5}$

$\pi^*(m_1 \mid Y^{(2)}) = 0.38$



$\pi(m_1 \mid Y^{(3)}) = 1$

$\pi^*(m_1 \mid Y^{(3)}) = 0.90$



[**H** & Miller 2019]

Reproducible phylogenetic inference

[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA

all

```
Minke GACCCGAACGTAATAA…ATCCGTTCCCATACTC
Blue  CACCCCCCCGTACTAT…TGAGTCCGAATTGGAA
Fin   TGTCTTCTACACTCCA…ACAGGTTGTACGTCAC
Grey  GGGTCGCTGTAGACCA…GATACCGCTCTCACAT
```

[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA

all **1st half**

```
Minke  GACCCGAACGTAATAA.ATCCGTTCCCATACTC
Blue   CACCCCCCCGTACTAT.TGAGTCCGAATTGGAA
Fin    TGTCTTCTACACTCCA.ACAGGTTGTACGTCAC
Grey   GGGTCGCTGTAGACCA.GATACCGCTCTCACAT
```

[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA

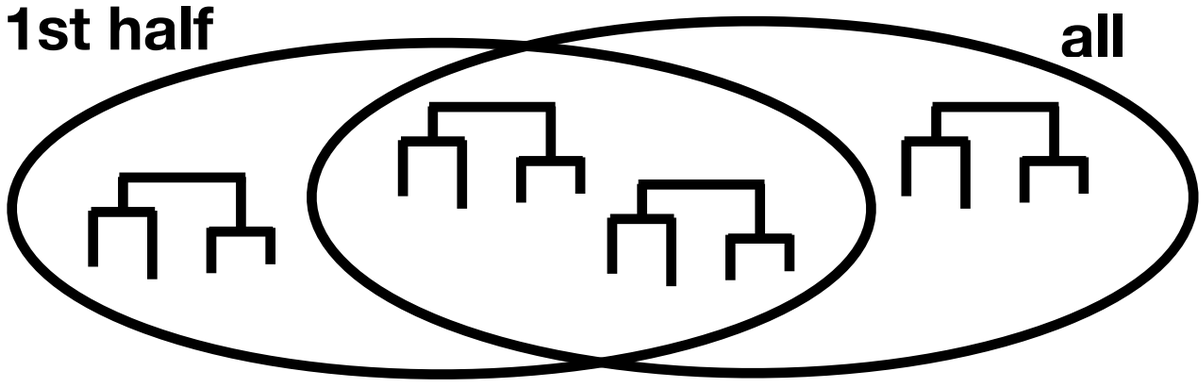
all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT

[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT

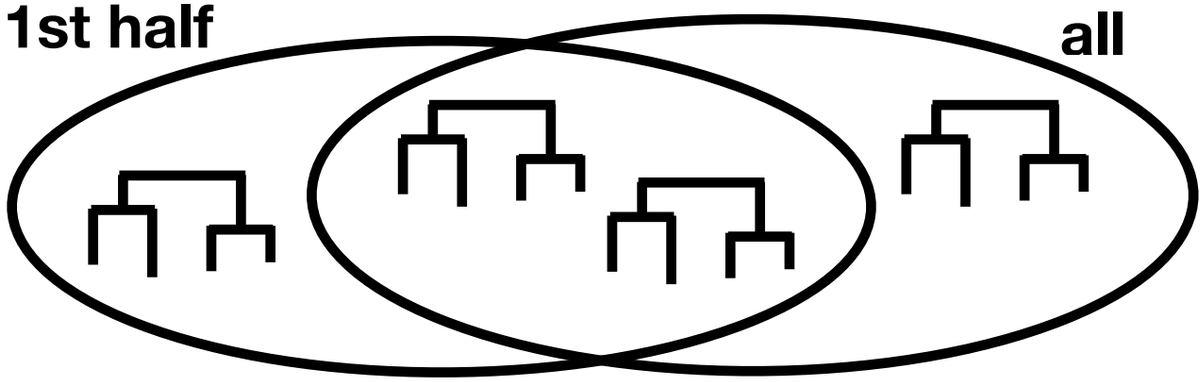


[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT

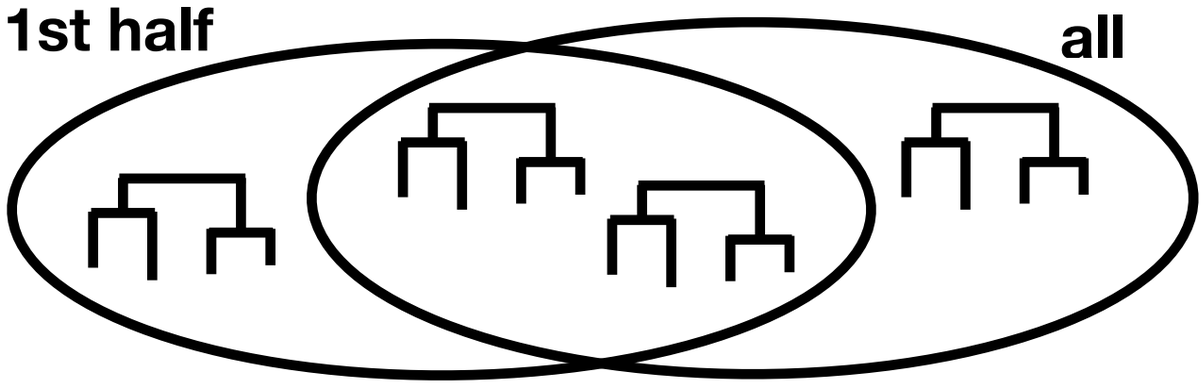


[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT

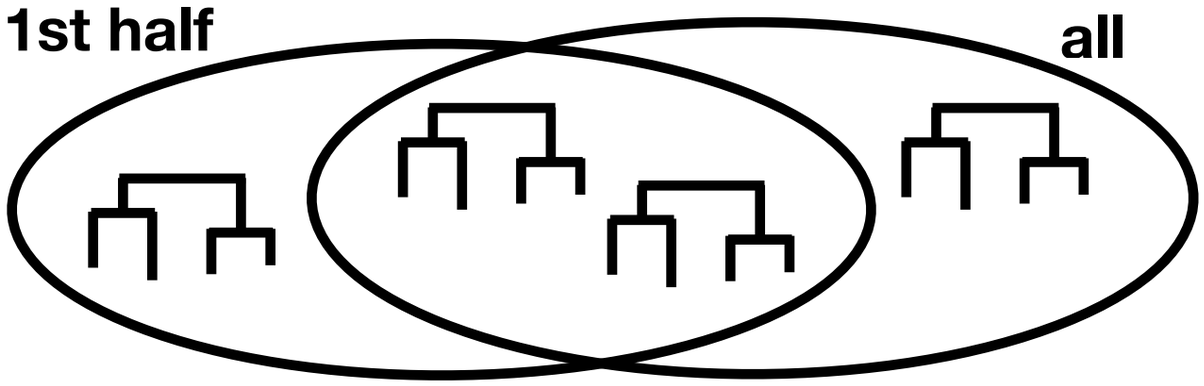


[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT

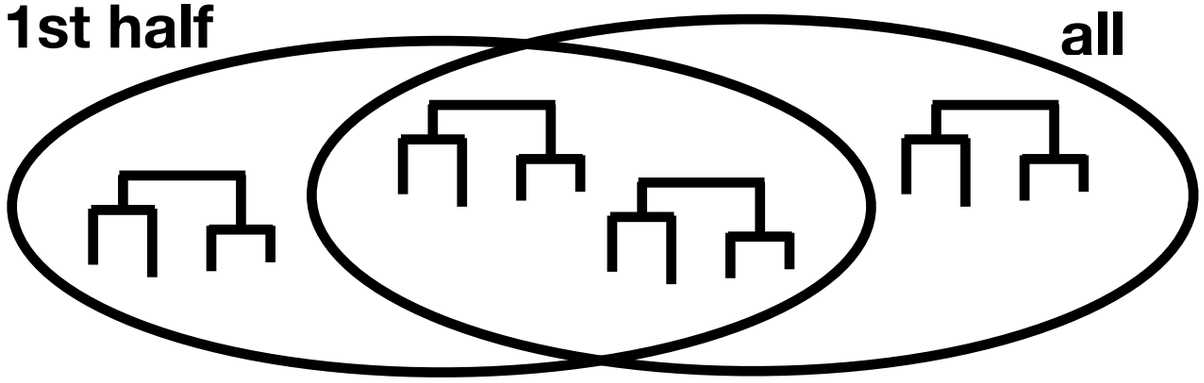


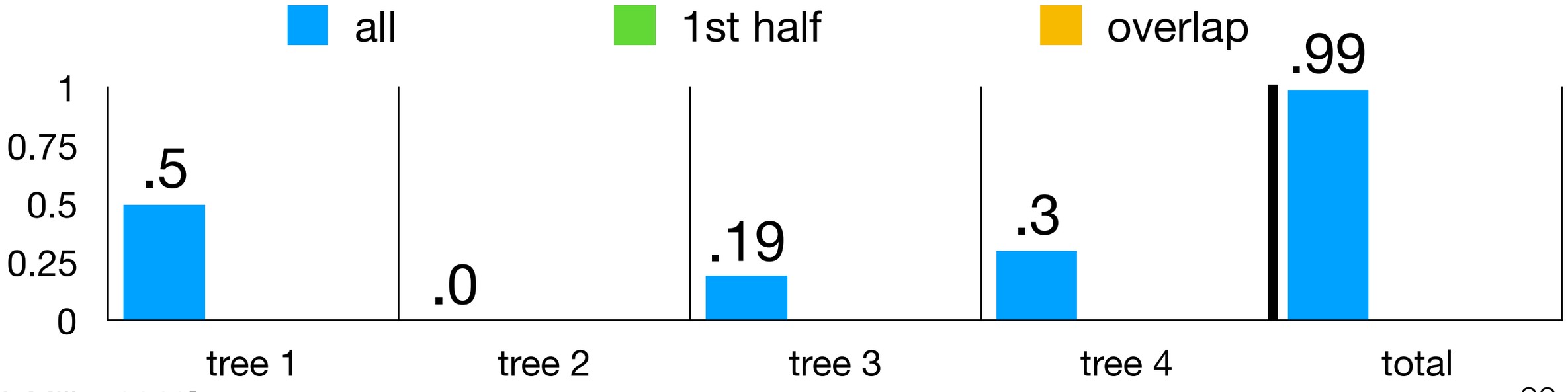
[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT



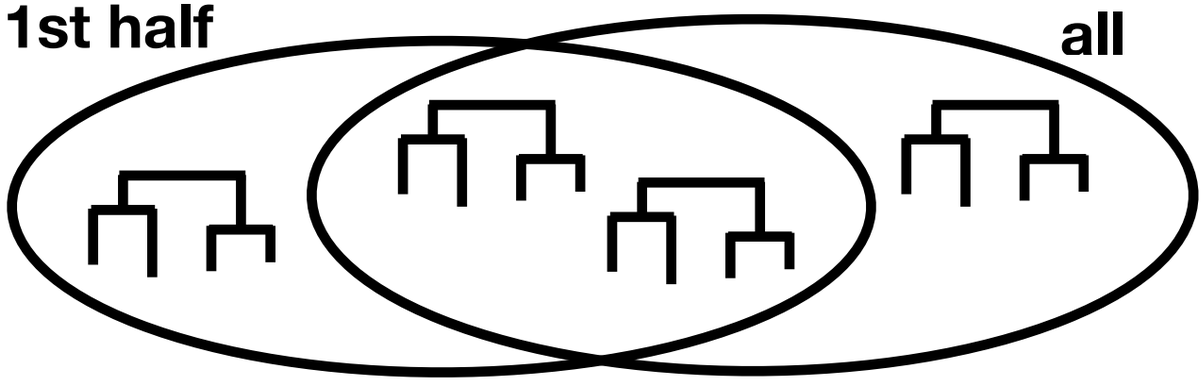


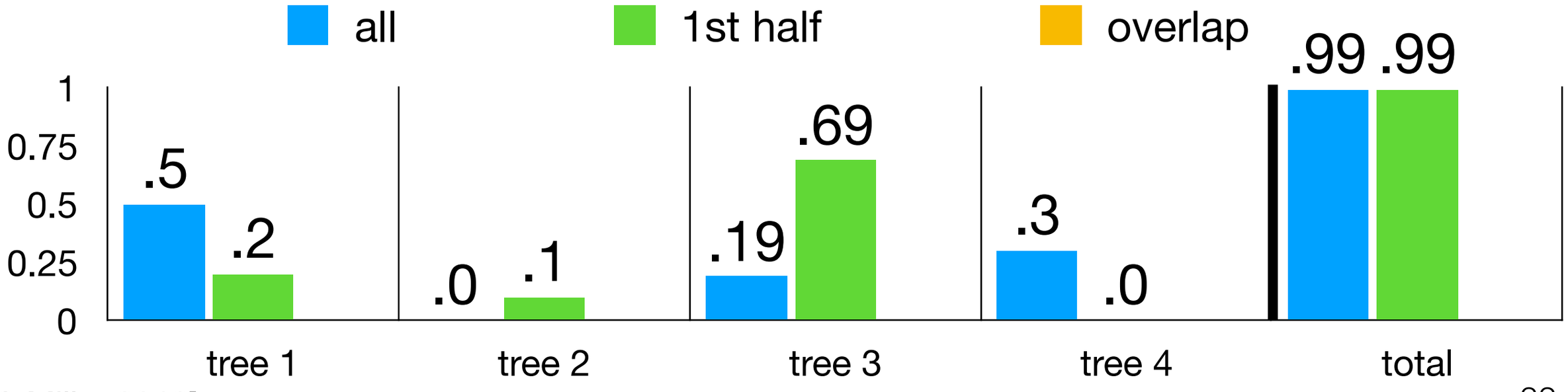
[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT



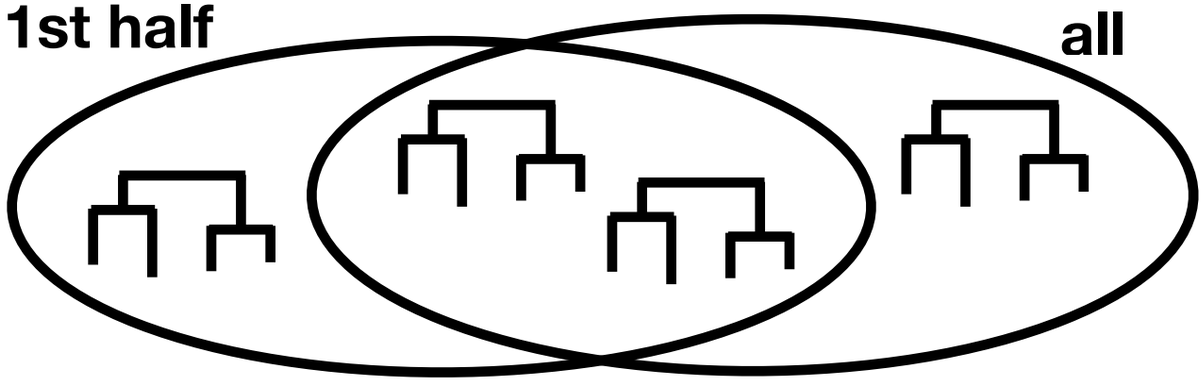


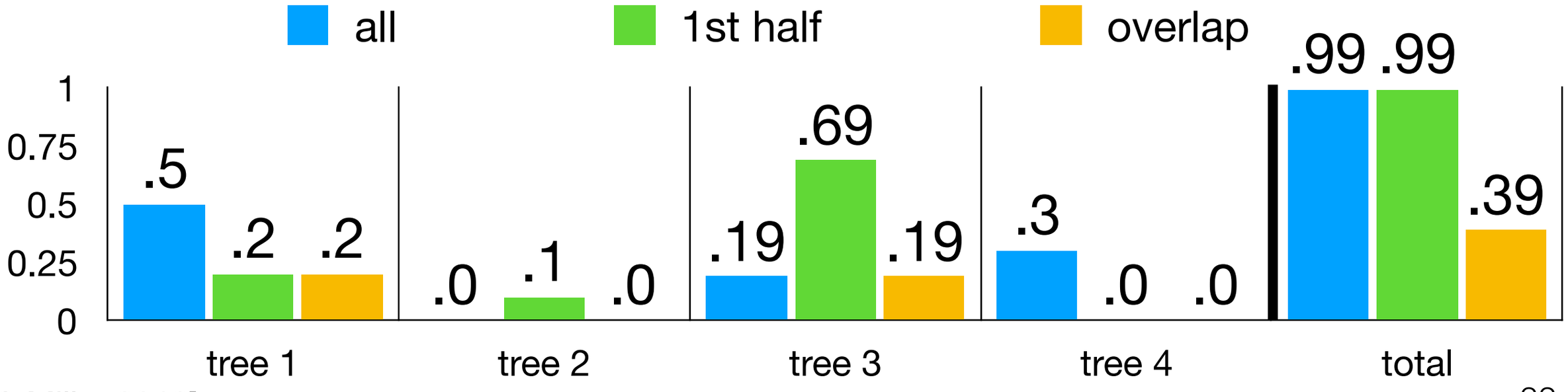
[**H** & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT



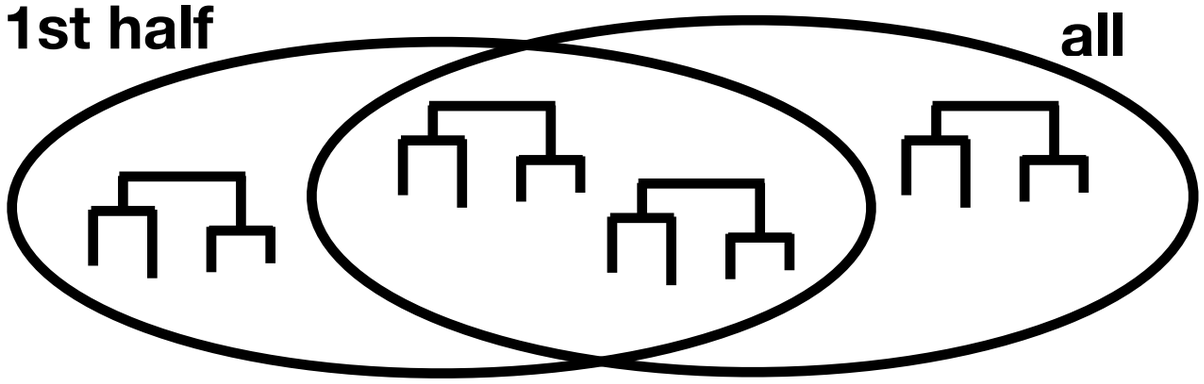


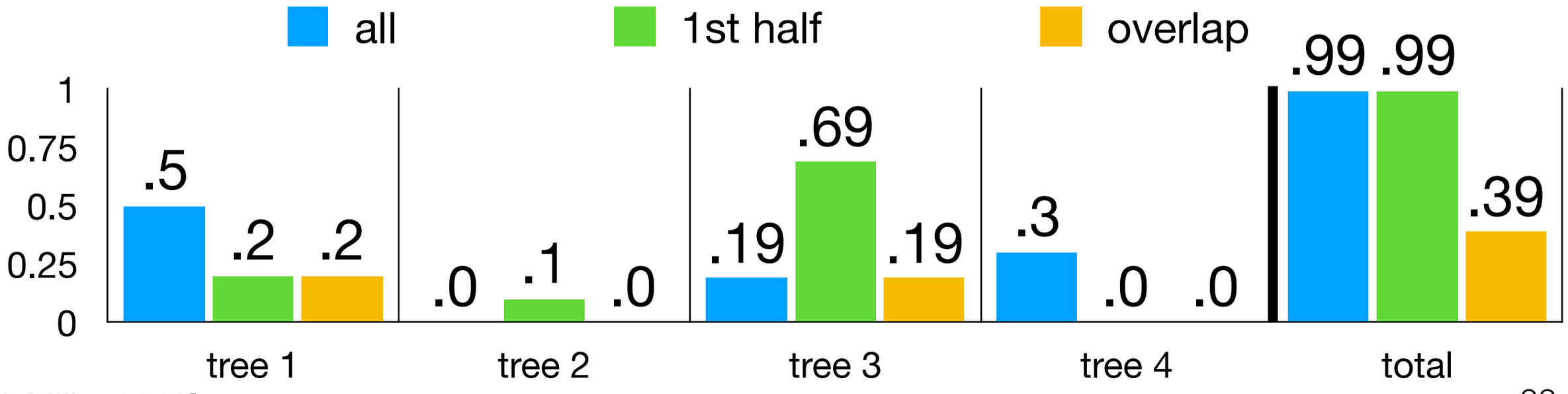
[H & Miller 2019]

Reproducible phylogenetic inference

- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions
- 0% overlap = contradiction

all	1st half	2nd half
Minke	GACCCGAACGTAATAA	ATCCGTTCCCATACTC
Blue	CACCCCCCCGTACTAT	TGAGTCCGAATTGGAA
Fin	TGTCTTCTACACTCCA	ACAGGTTGTACGTCAC
Grey	GGGTCGCTGTAGACCA	GATACCGCTCTCACAT





[H & Miller 2019]

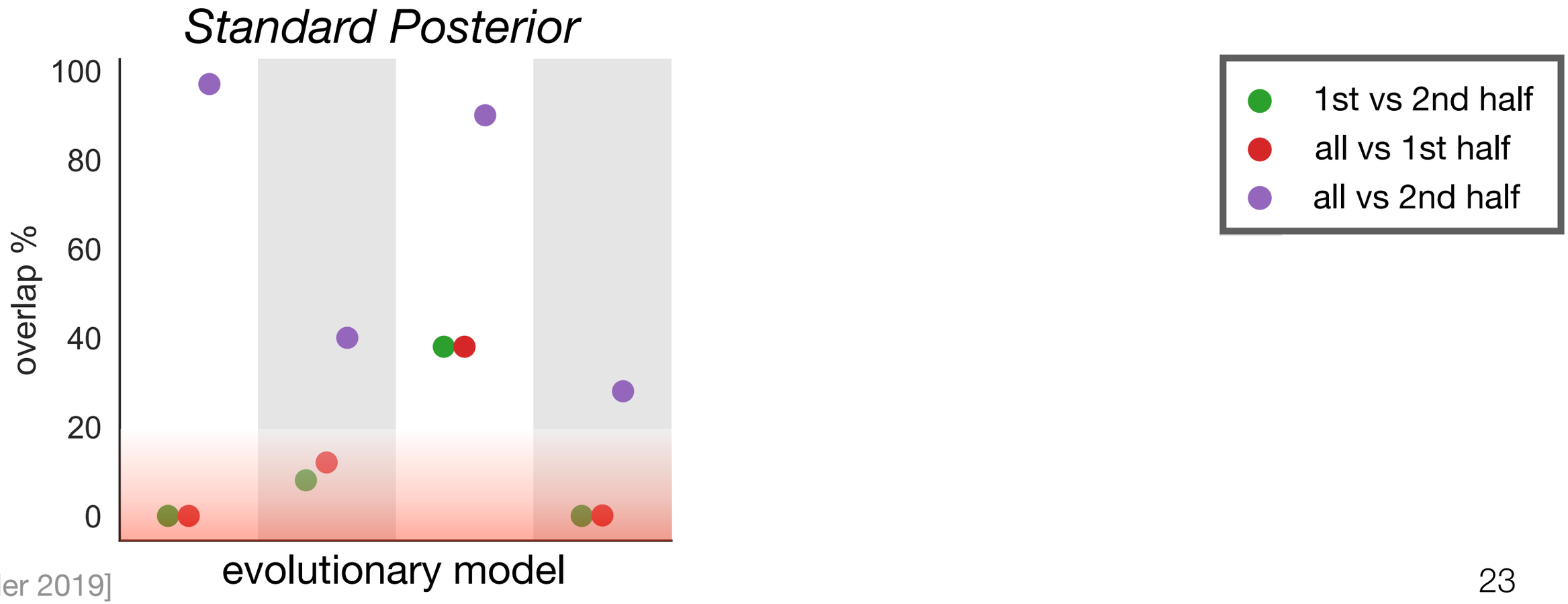
Stable, reproducible phylogenetic inference with BayesBag

- **Goal:** infer phylogeny of 13 whale species

[**H** & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

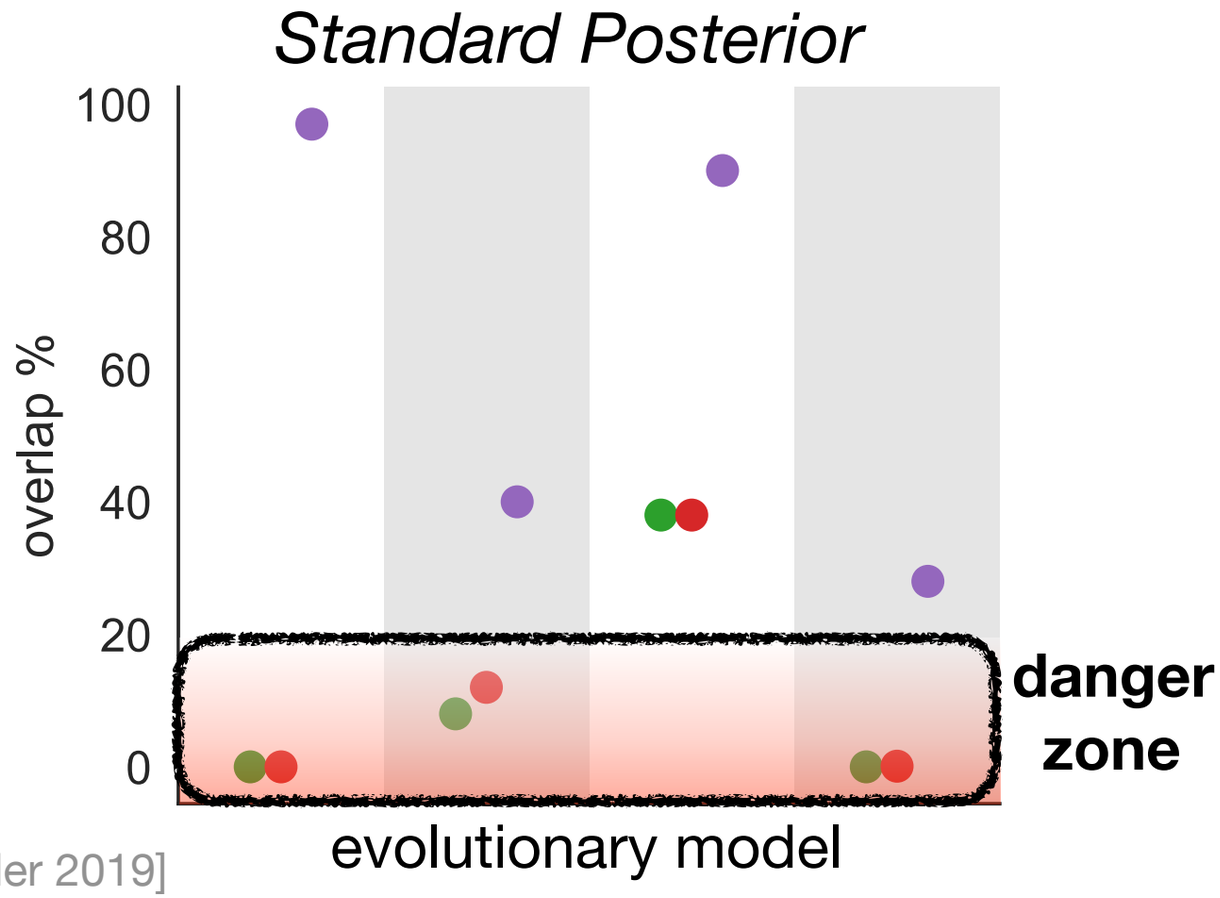
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap



[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

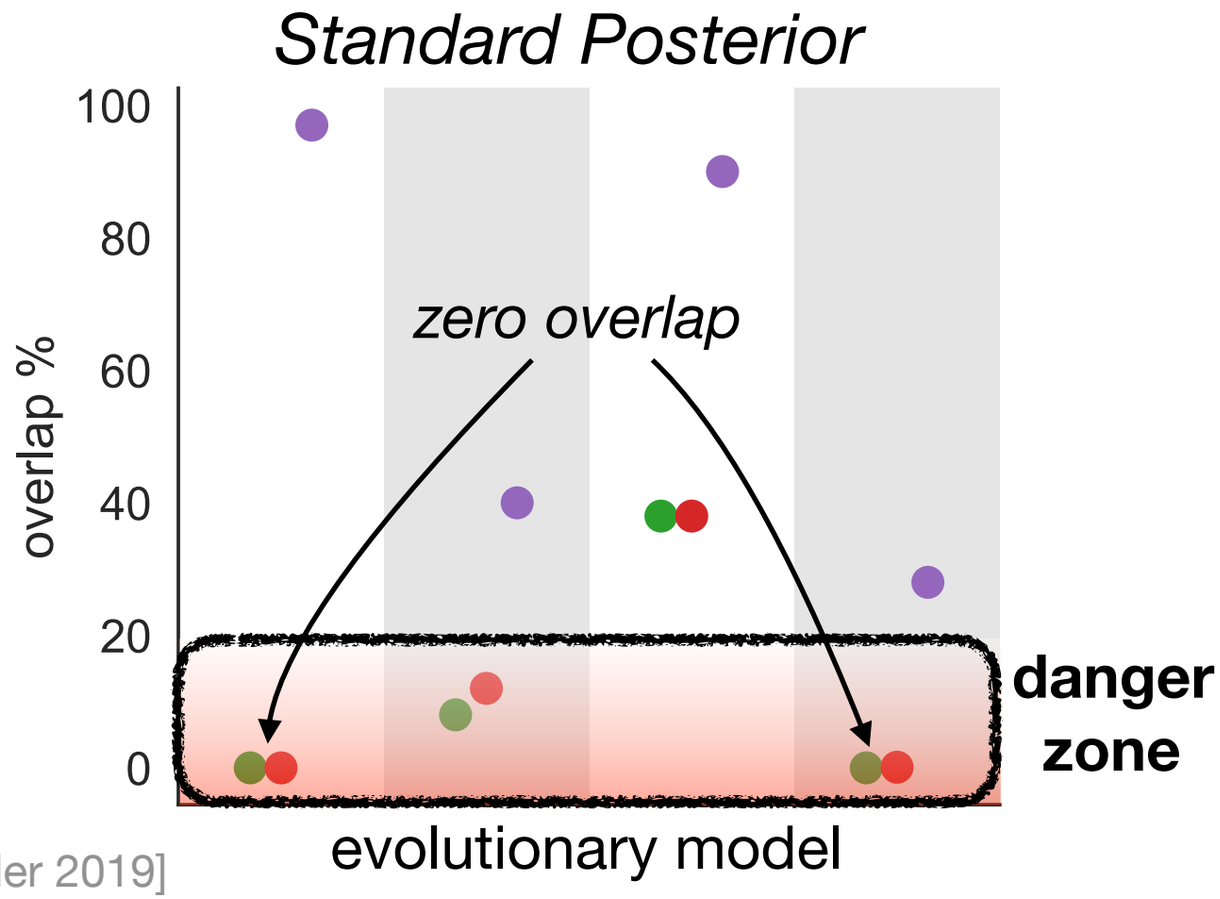
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap

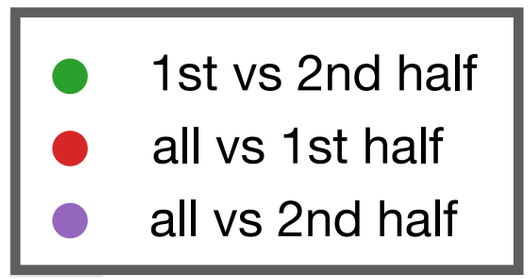


[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap

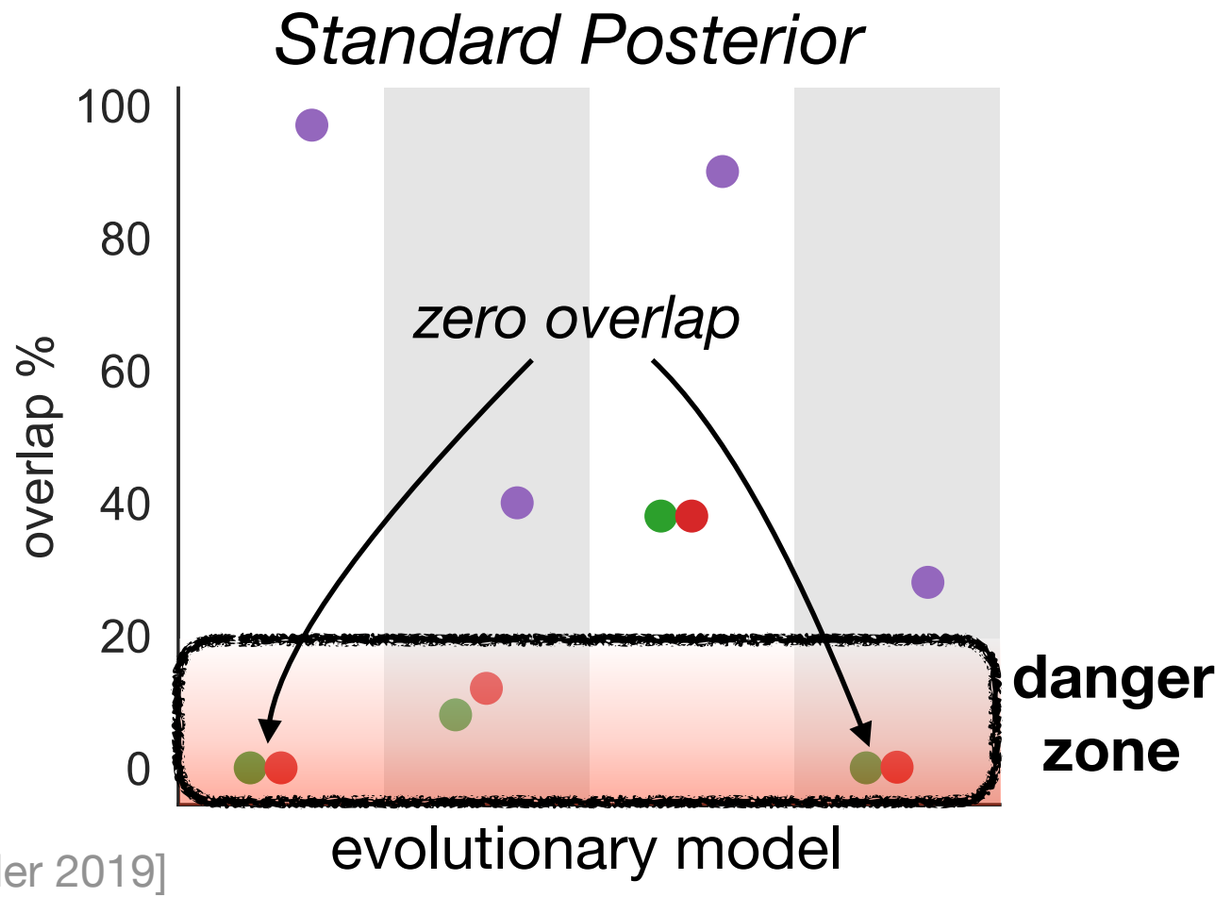




[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

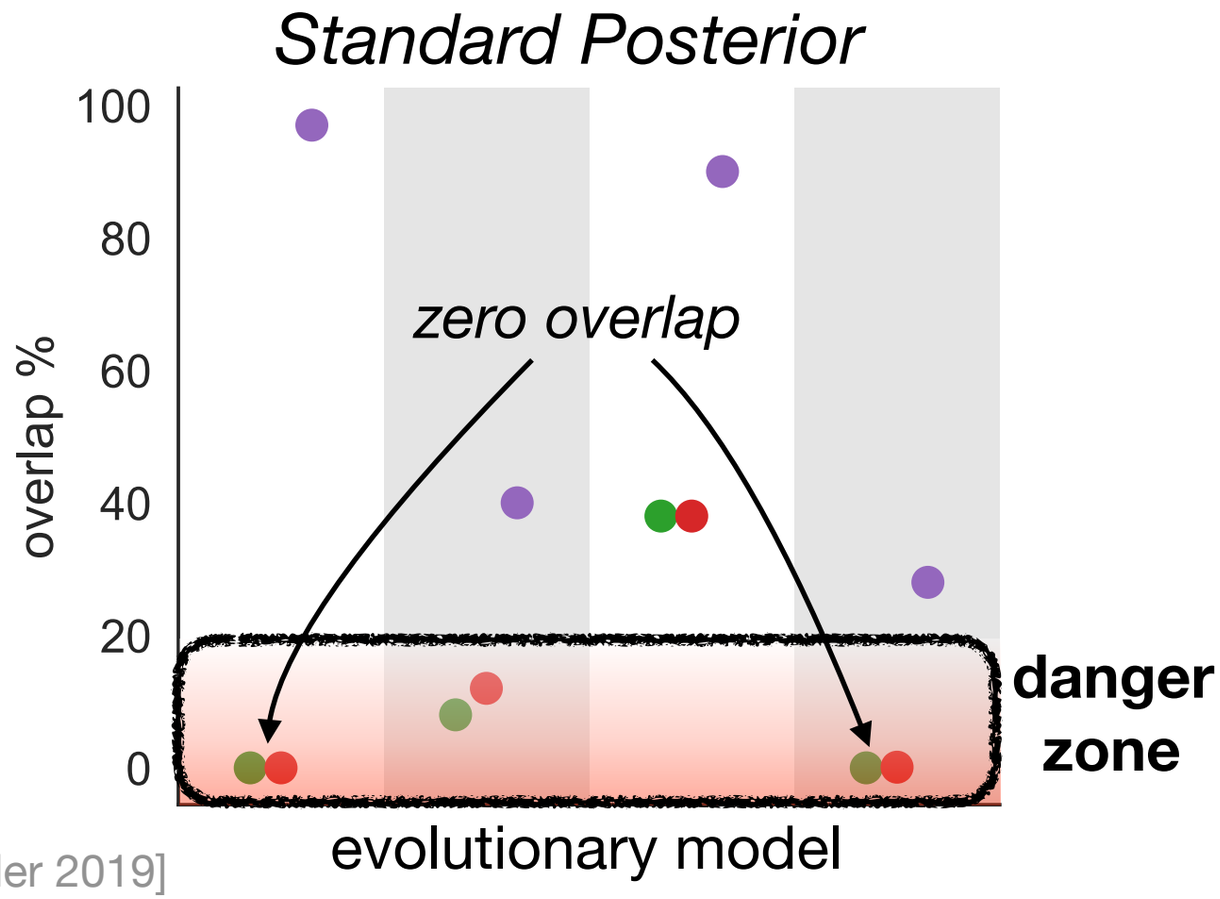
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]



[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

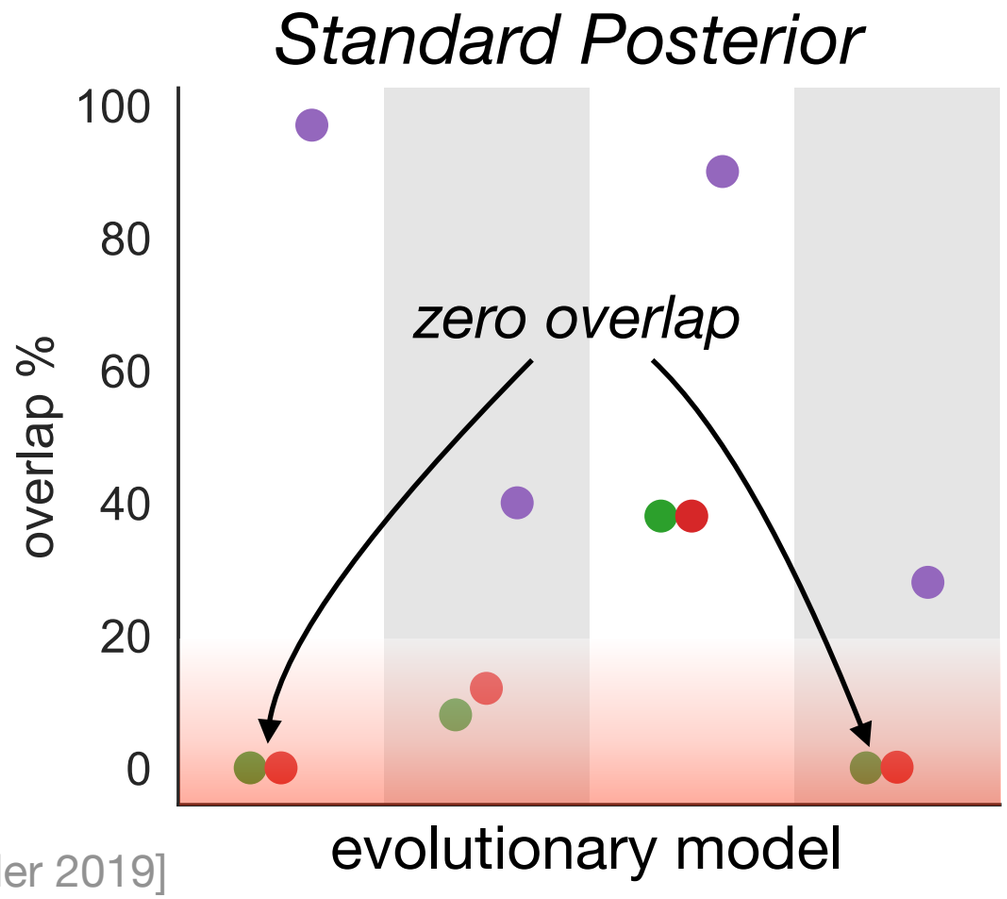
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]
- Same problem comparing evolutionary models with data fixed

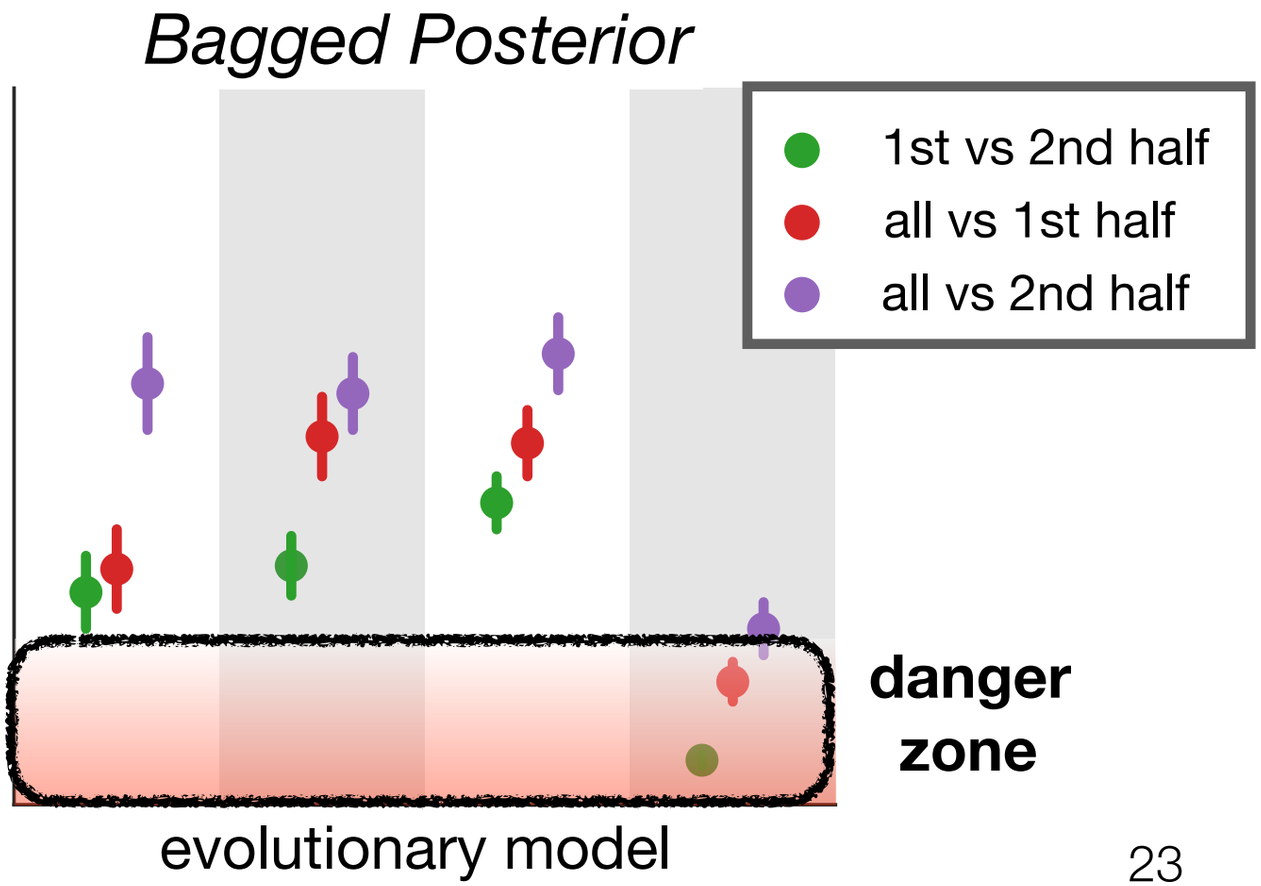


[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]
- Same problem comparing evolutionary models with data fixed
- Bagged posterior model probabilities more stable and reproducible

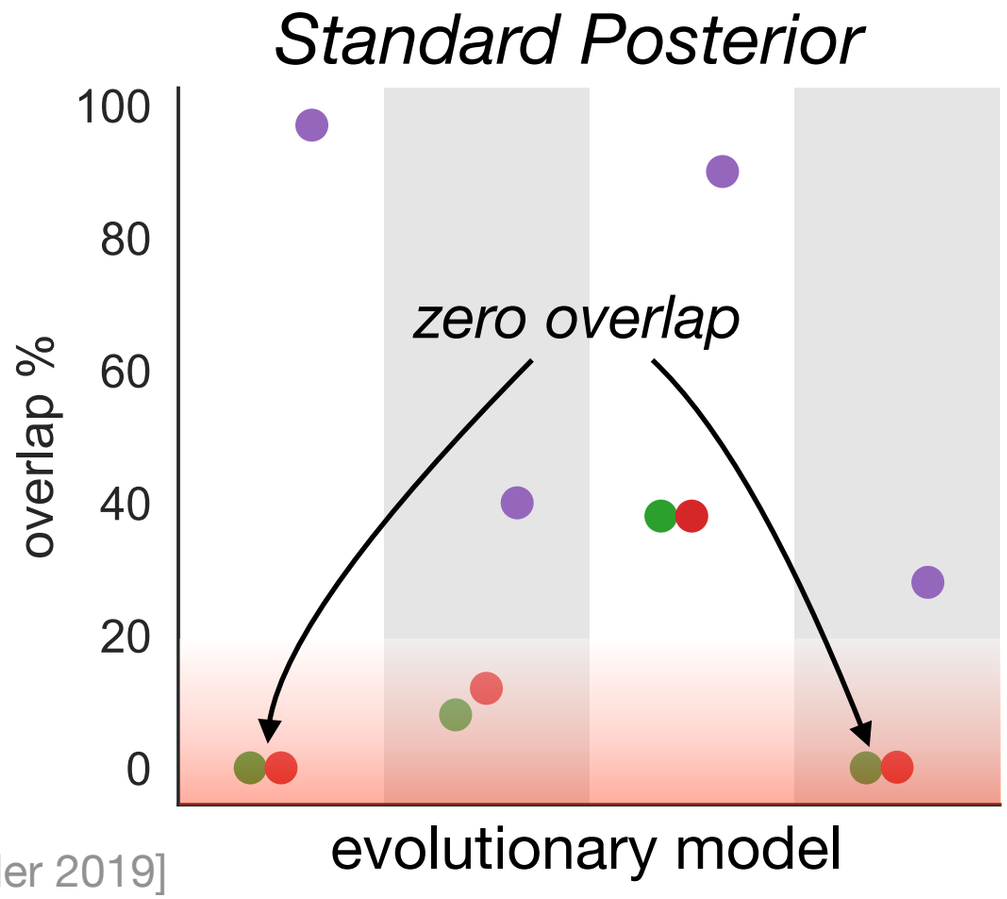


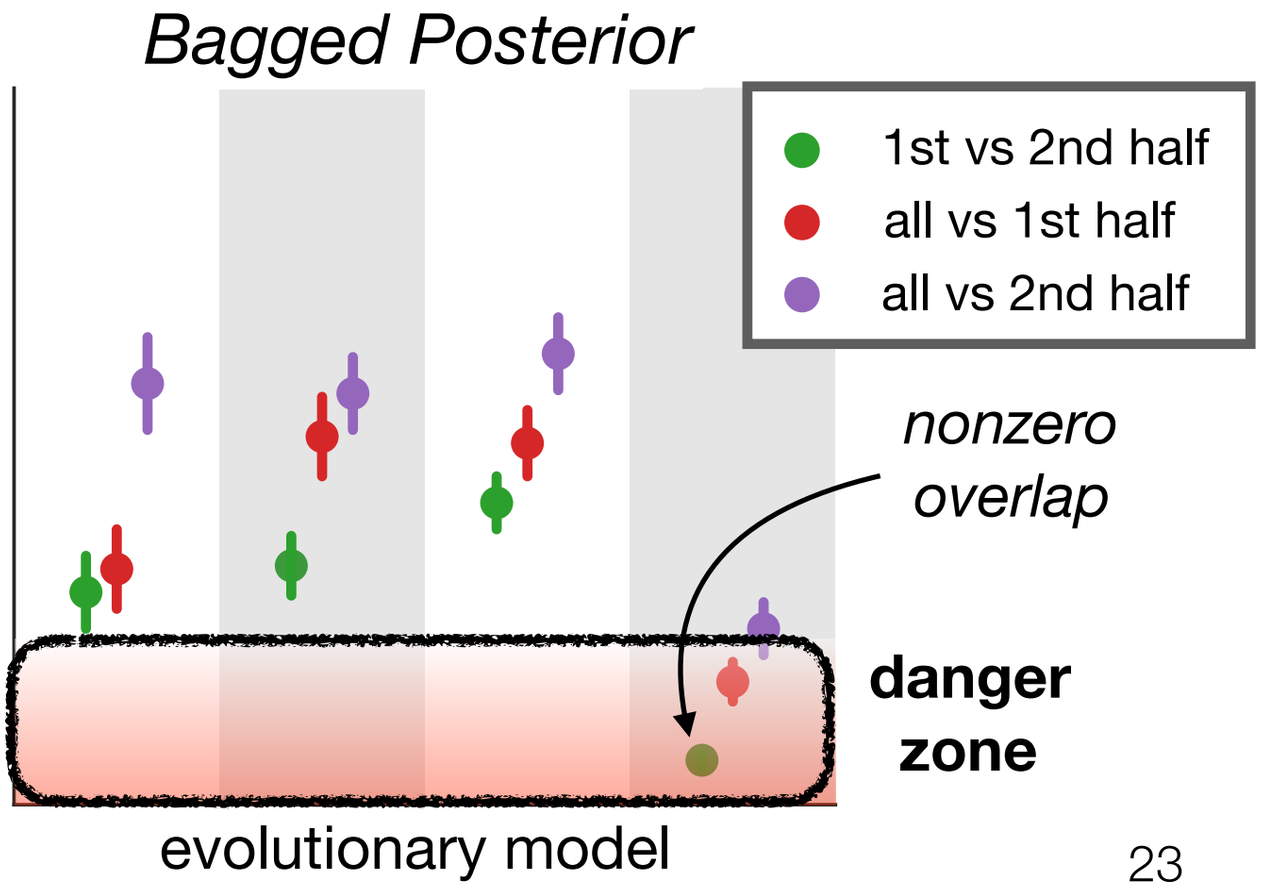


[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]
- Same problem comparing evolutionary models with data fixed
- Bagged posterior model probabilities more stable and reproducible





[H & Miller 2019]

Closing thoughts

Closing thoughts

- We show that BayesBag…

Closing thoughts

- We show that BayesBag…

    1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)

Closing thoughts

- We show that BayesBag…
    1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)
    2. more **stable** model probabilities (model selection)

Closing thoughts

- We show that BayesBag…

  1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)

  2. more **stable** model probabilities (model selection)

  3. demonstrates **excellent empirical performance**

Closing thoughts

- We show that BayesBag…

  1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)

  2. more **stable** model probabilities (model selection)

  3. demonstrates **excellent empirical performance**

  4. is **easy to use** and **widely applicable**

Closing thoughts

- We show that BayesBag…

  1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)

  2. more **stable** model probabilities (model selection)

  3. demonstrates **excellent empirical performance**

  4. is **easy to use** and **widely applicable**

  5. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods

Closing thoughts

- We show that BayesBag…
    1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)
    2. more **stable** model probabilities (model selection)
    3. demonstrates **excellent empirical performance**
    4. is **easy to use** and **widely applicable**
    5. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work: finite-sample properties of BayesBag and mismatch index, computation, non-independent data (e.g. time series & spatial models)

Closing thoughts

- We show that BayesBag…
    1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)
    2. more **stable** model probabilities (model selection)
    3. demonstrates **excellent empirical performance**
    4. is **easy to use** and **widely applicable**
    5. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work: finite-sample properties of BayesBag and mismatch index, computation, non-independent data (e.g. time series & spatial models)
- **Conclusion:** you should give BayesBag a try!

Closing thoughts

- We show that BayesBag…
    1. provides **better-calibrated uncertainty** that is conservative by default (parameter inference)
    2. more **stable** model probabilities (model selection)
    3. demonstrates **excellent empirical performance**
    4. is **easy to use** and **widely applicable**
    5. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work: finite-sample properties of BayesBag and mismatch index, computation, non-independent data (e.g. time series & spatial models)
- **Conclusion:** you should give BayesBag a try!
- On arXiv soon (if you want a heads up: jhuggins@hsph.harvard.edu)

Why can Bayesian model selection fail?

Why can Bayesian model selection fail?

- Let $\Delta_n \triangleq \sum_{i=1}^{n} \log p(Y_i \,|\, m_1) - \log p(Y_i \,|\, m_2)$

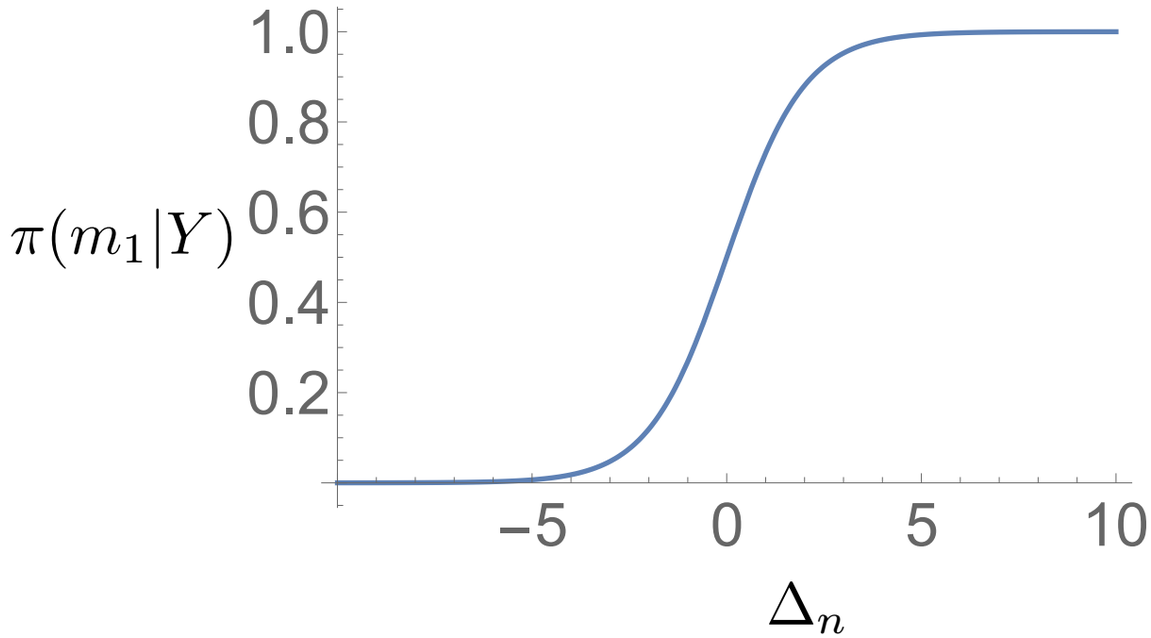
Why can Bayesian model selection fail?

- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \mid m_1) - \log p(Y_i \mid m_2)}_{\delta_i}$

Why can Bayesian model selection fail?

- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \,|\, m_1) - \log p(Y_i \,|\, m_2)}_{\delta_i}$

- Then $\pi(m_1 \,|\, Y) = (1 + \exp\{-\Delta_n\})^{-1}$

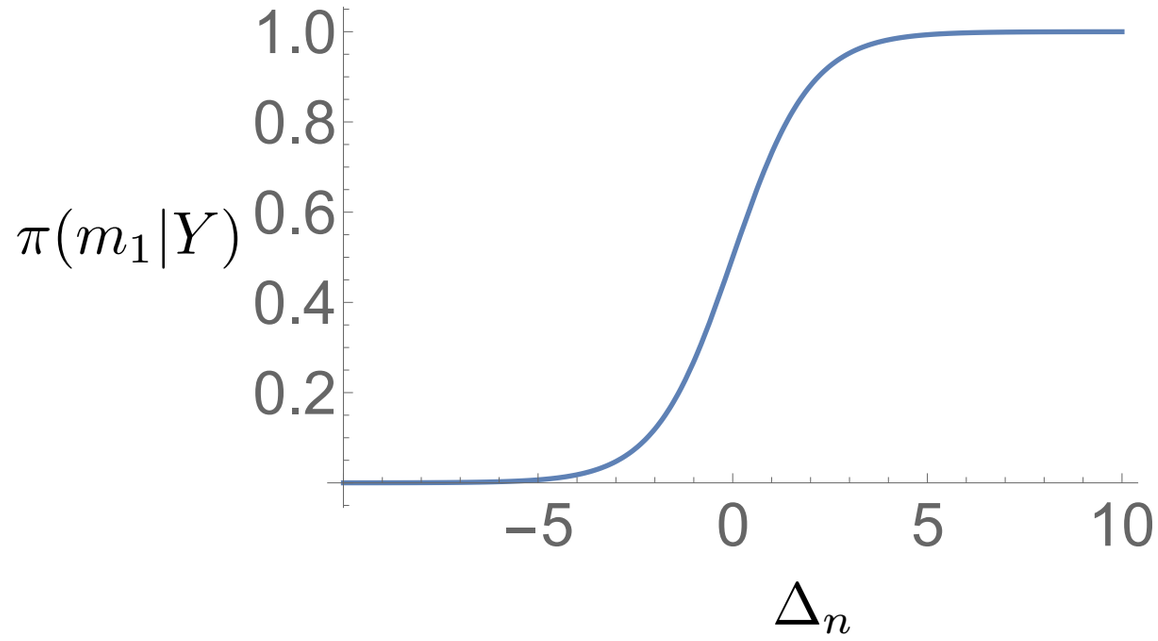


Why can Bayesian model selection fail?

- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \,|\, m_1) - \log p(Y_i \,|\, m_2)}_{\delta_i}$

- Then $\pi(m_1 \,|\, Y) = (1 + \exp\{-\Delta_n\})^{-1}$

- By assumption, $\mathbb{E}[\delta_i] = 0$ but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$



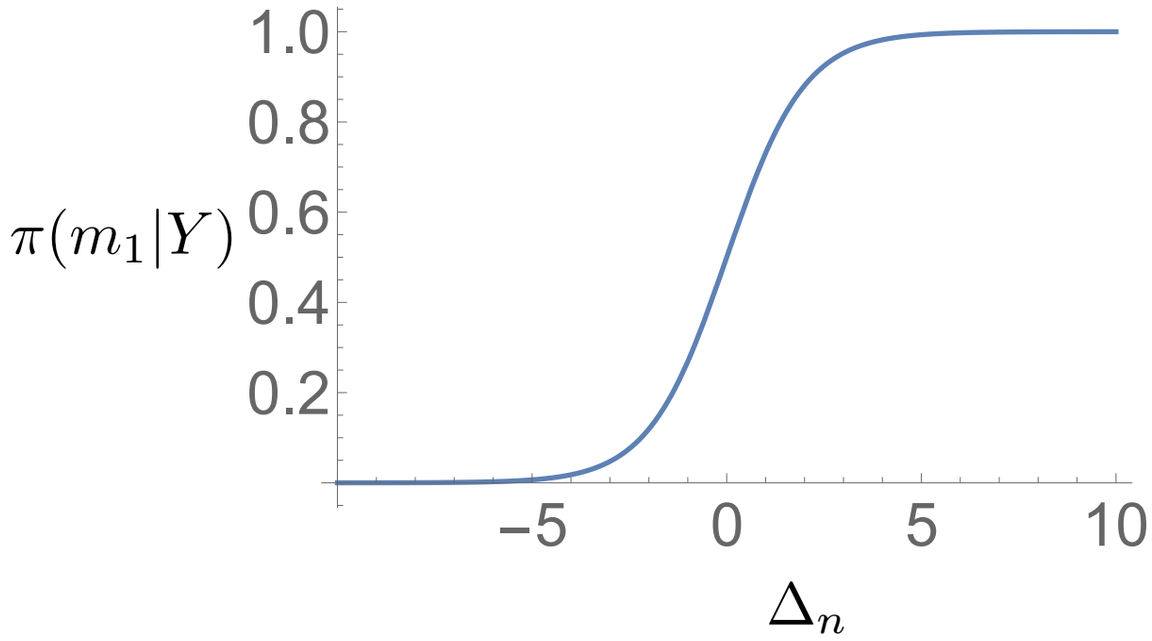
Why can Bayesian model selection fail?

- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \,|\, m_1) - \log p(Y_i \,|\, m_2)}_{\delta_i}$

- Then $\pi(m_1 \,|\, Y) = (1 + \exp\{-\Delta_n\})^{-1}$

- By assumption, $\mathbb{E}[\delta_i] = 0$ but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$

- Hence, Δ_n is a random walk with $\mathbb{E}[\Delta_n^2] = \sigma^2 n$



Why can Bayesian model selection fail?

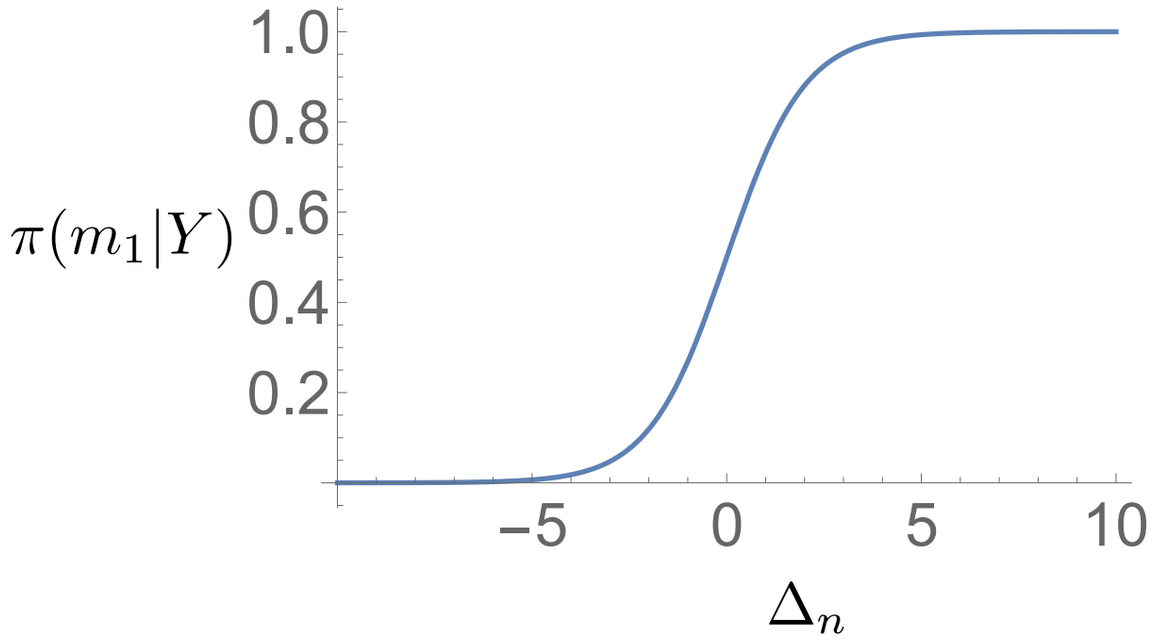
- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \,|\, m_1) - \log p(Y_i \,|\, m_2)}_{\delta_i}$

- Then $\pi(m_1 \,|\, Y) = (1 + \exp\{-\Delta_n\})^{-1}$

- By assumption, $\mathbb{E}[\delta_i] = 0$ but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$

- Hence, Δ_n is a random walk with $\mathbb{E}[\Delta_n^2] = \sigma^2 n$

- In other words, with very high probability, $|\Delta_n| = \Theta(n^{1/2})$



Why can Bayesian model selection fail?

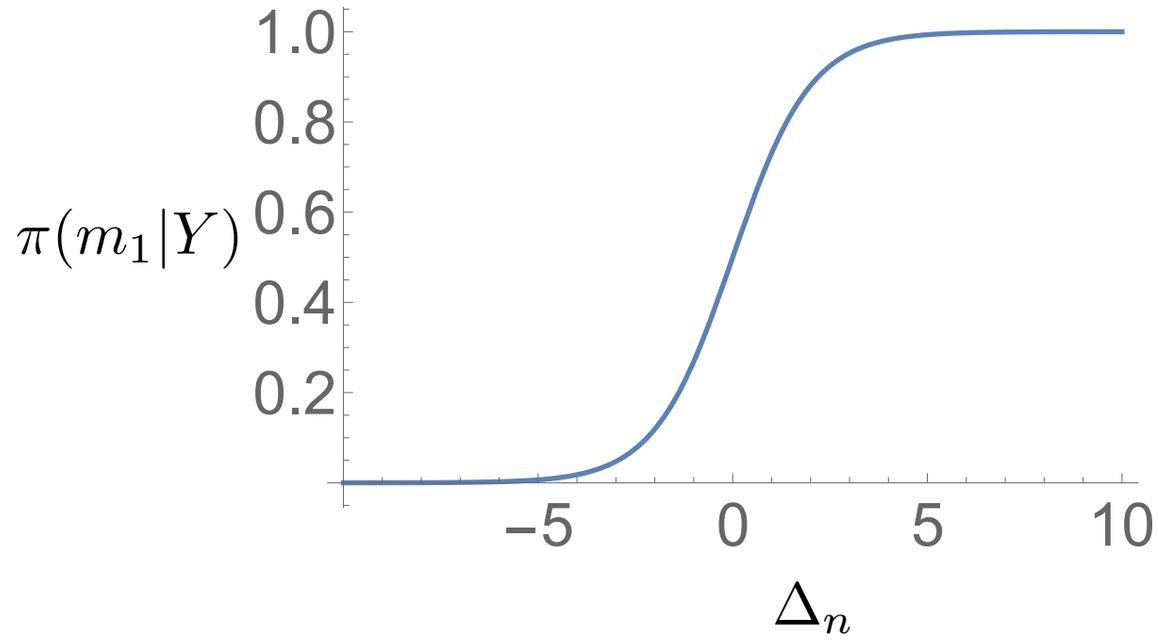
- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \,|\, m_1) - \log p(Y_i \,|\, m_2)}_{\delta_i}$

- Then $\pi(m_1 \,|\, Y) = (1 + \exp\{-\Delta_n\})^{-1}$

- By assumption, $\mathbb{E}[\delta_i] = 0$ but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$

- Hence, Δ_n is a random walk with $\mathbb{E}[\Delta_n^2] = \sigma^2 n$

- In other words, with very high probability, $|\Delta_n| = \Theta(n^{1/2})$



- Therefore, there is overwhelming evidence of order $n^{1/2}$ for either m_1 or m_2