Using Bagged Posteriors for Robust Inference

Jonathan Huggins
Harvard University

Joint work with Jeff Miller

Bagged posterior corrects for model misspecification

[**H** & Miller 2019]

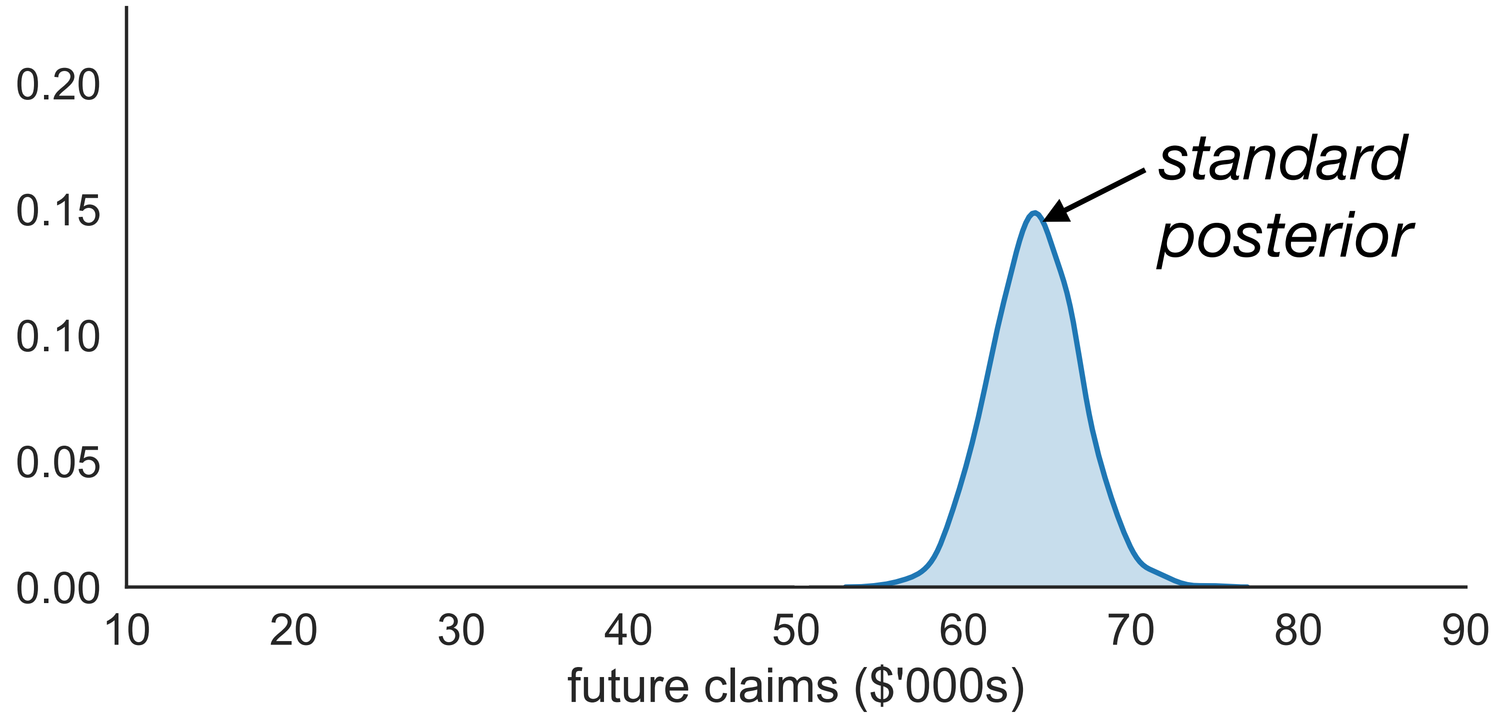
Bagged posterior corrects for model misspecification

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

[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

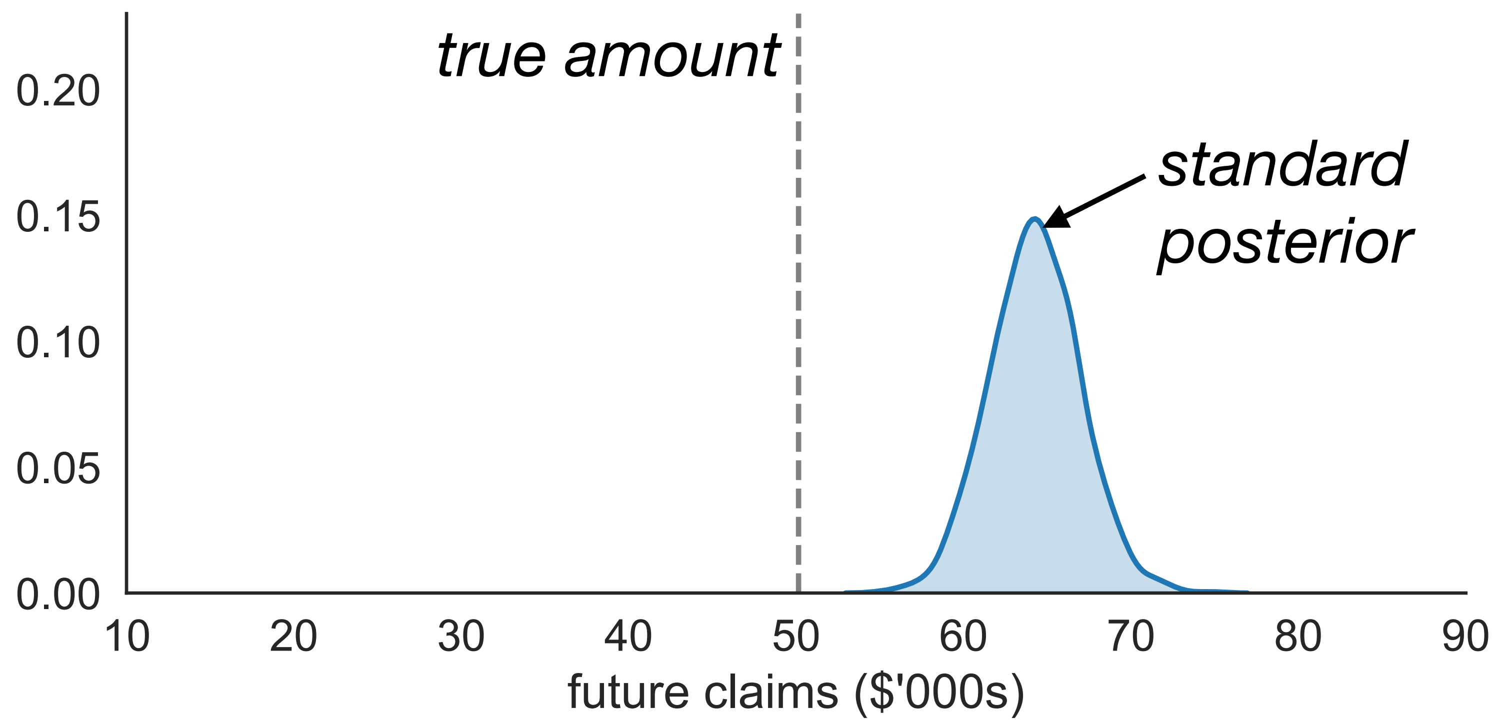
- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference with (non-trivial) model (data is 10 time series)



[H & Miller 2019]

Bagged posterior corrects for model misspecification

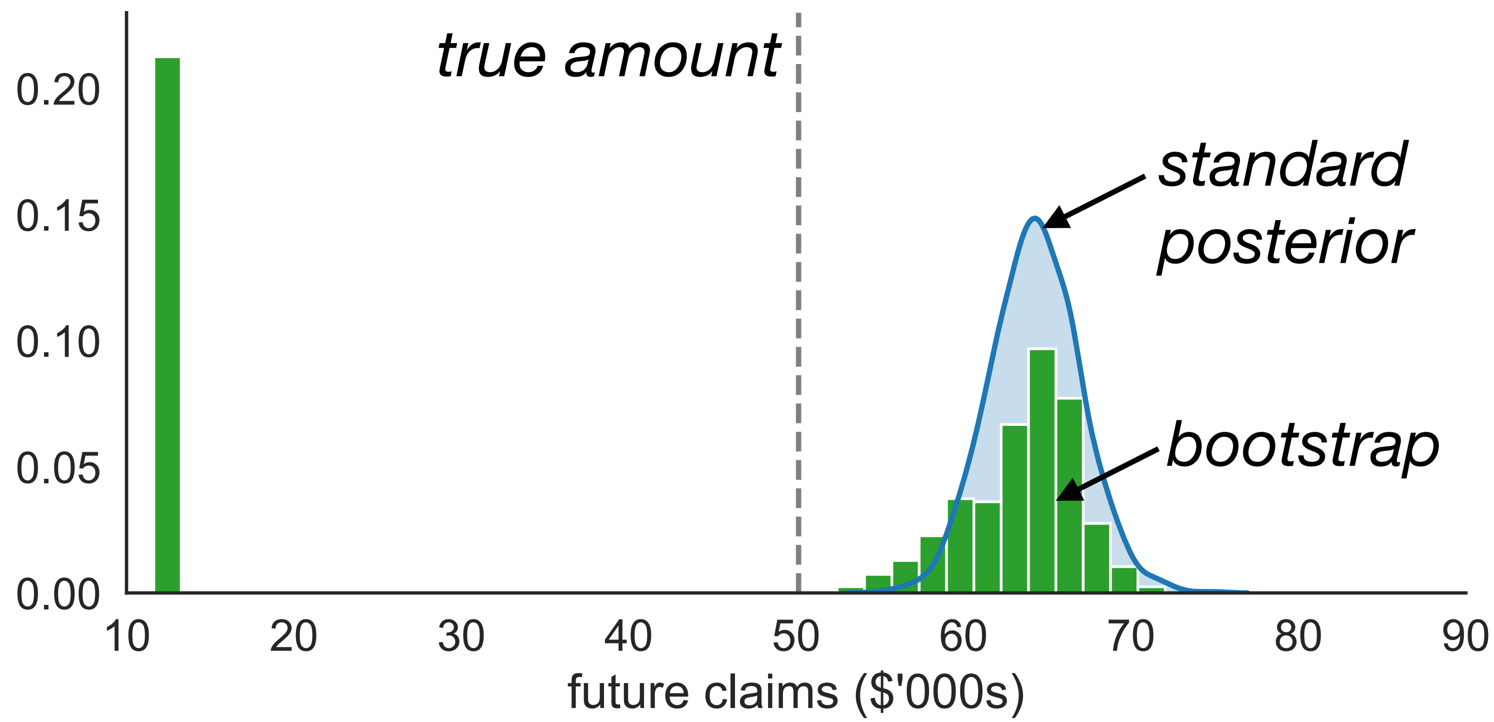
- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference with (non-trivial) model (data is 10 time series)
- **Problem:** uncertainty not well-calibrated because model is wrong



[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

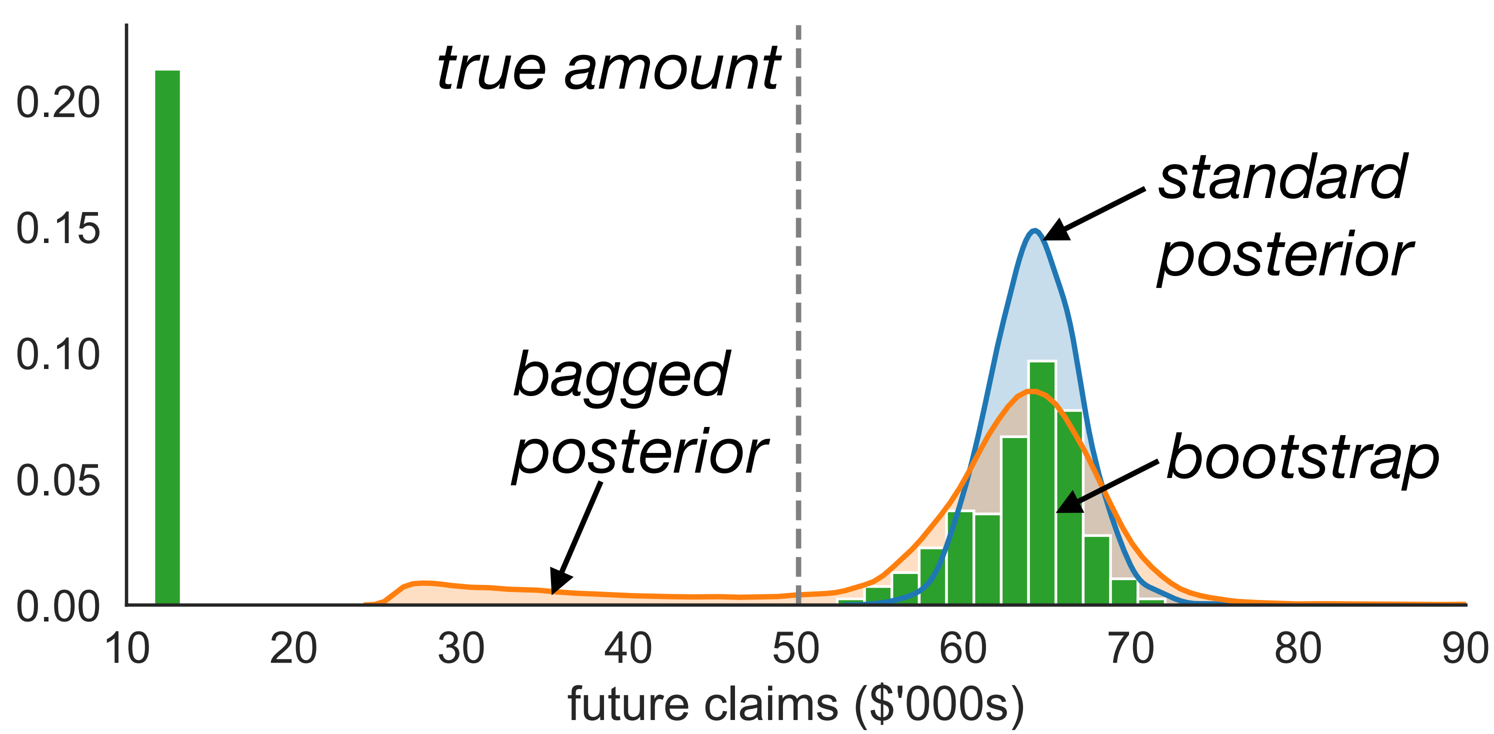
- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference with (non-trivial) model (data is 10 time series)
- **Problem:** uncertainty not well-calibrated because model is wrong
- **Alternative:** the bootstrap ⇒ too little data



[**H** & Miller 2019]

Bagged posterior corrects for model misspecification

- **Goal:** predict future insurance claims based on (real) historic data
- Try Bayesian inference with (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 theory and methodology
- BayesBag for model selection

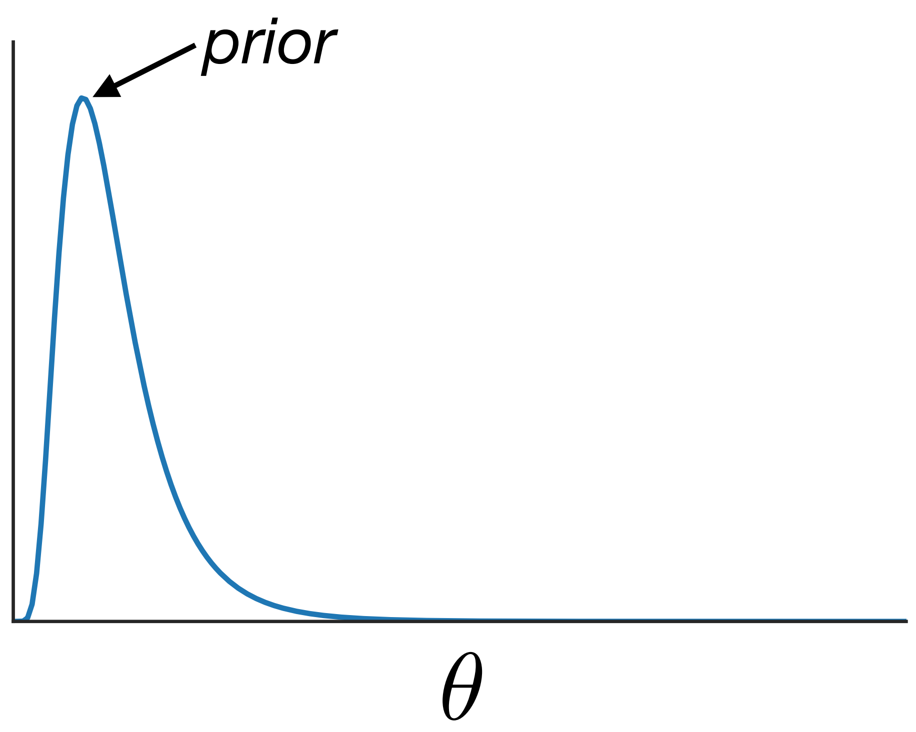
Bayesian inference

Bayesian inference

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

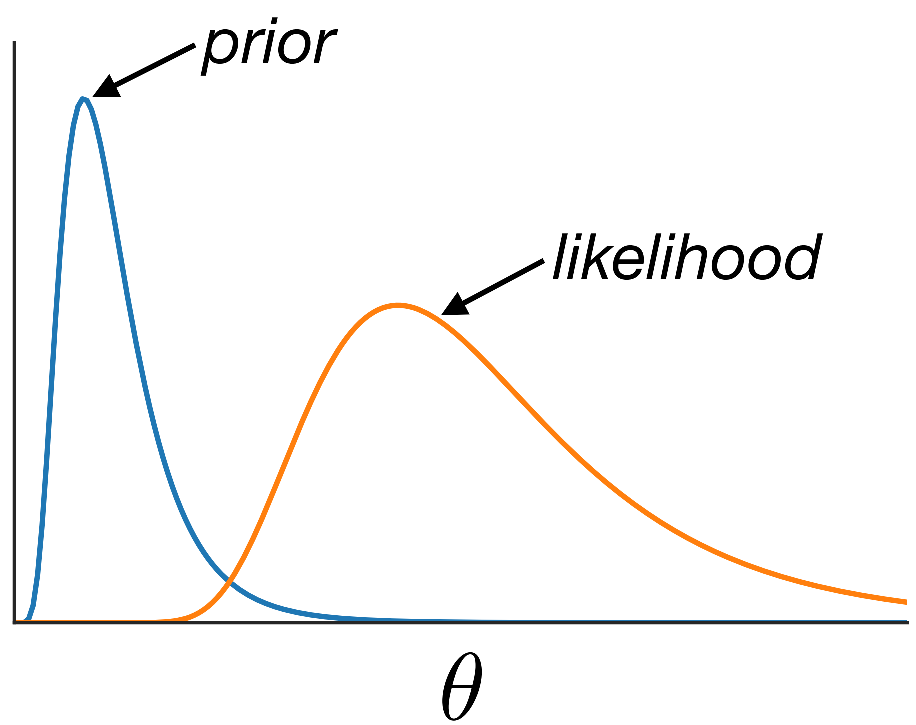
Bayesian inference

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



Bayesian inference

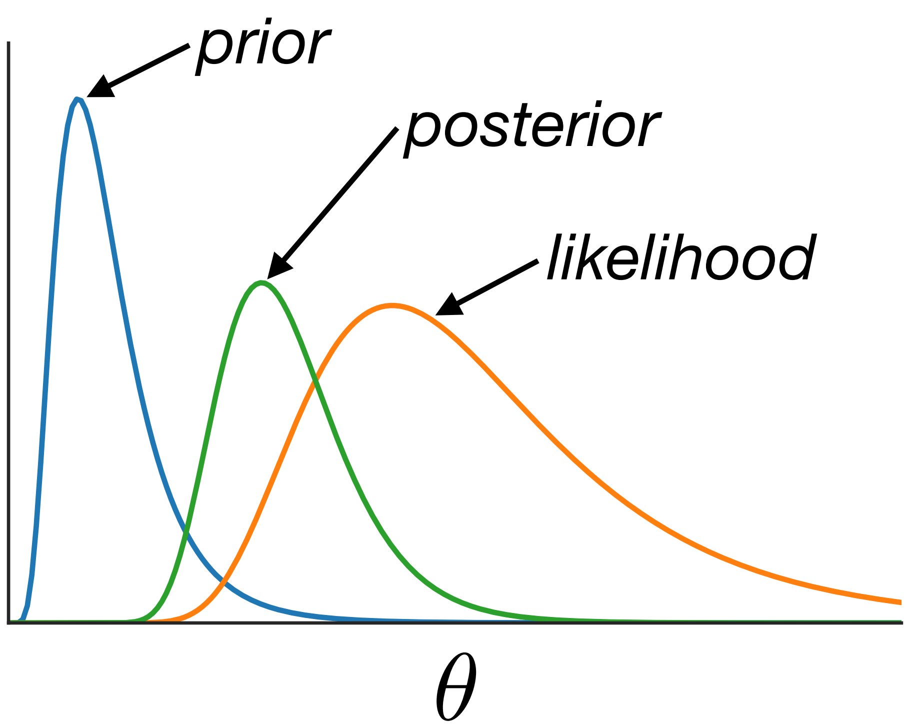
- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$



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**:

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

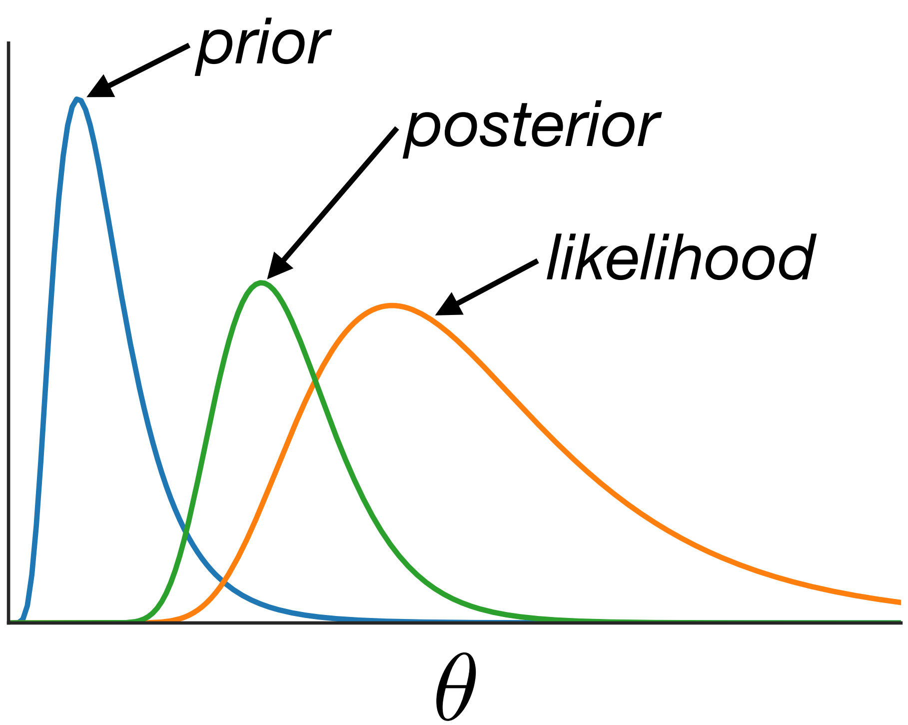


Bayesian inference

- **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

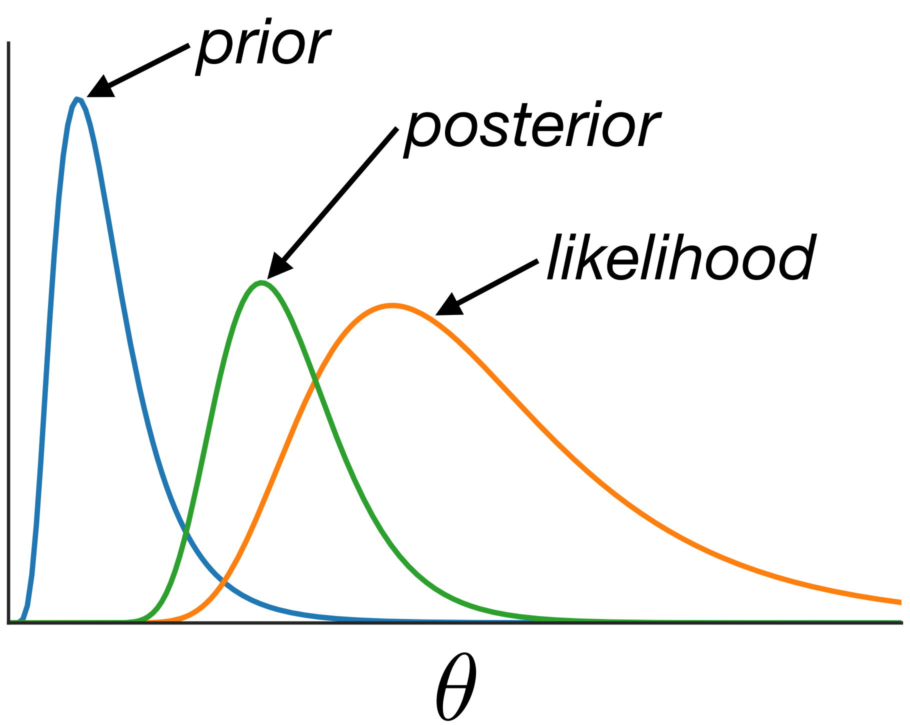


Bayesian inference

- **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 $p(Y \mid \theta_{\text{true}})$

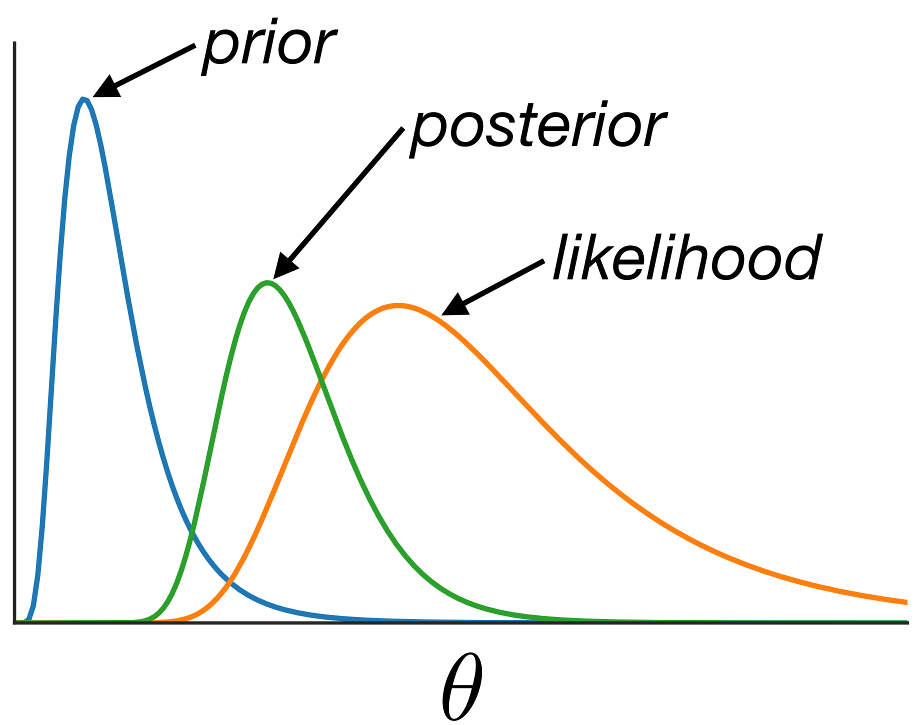


Bayesian inference

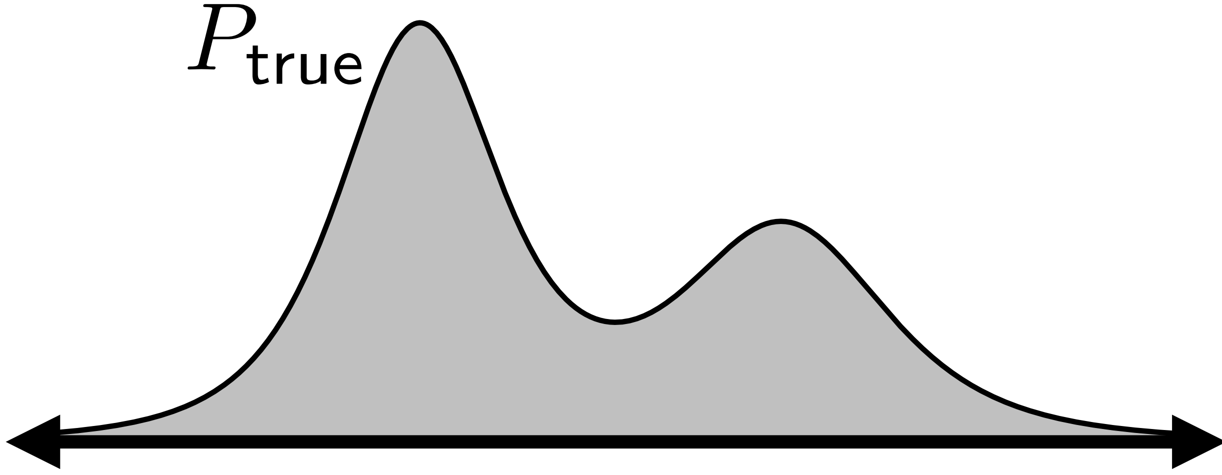
- **Goal:** learn about unobserved phenomenon (parameter) of interest θ **[e.g. future claims]**
- **Prior** beliefs $\pi_0(\theta)$ about the phenomenon
- Observe data Y via **model** $p(Y \mid \theta)$
- Combine prior & likelihood to form **posterior**:

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

- **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}



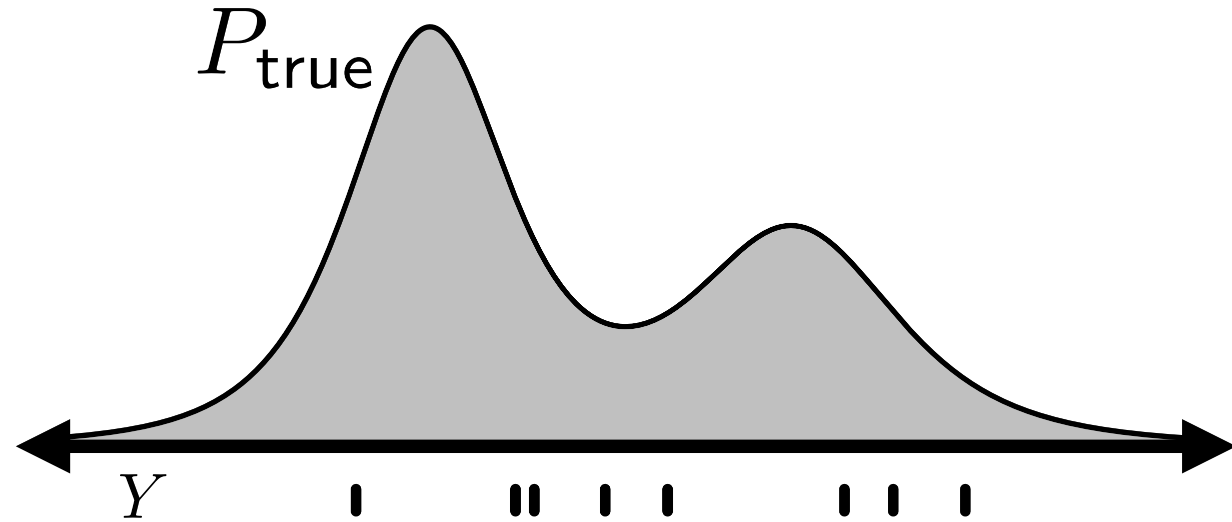
Bootstrapping



[Efron 1979]

Bootstrapping

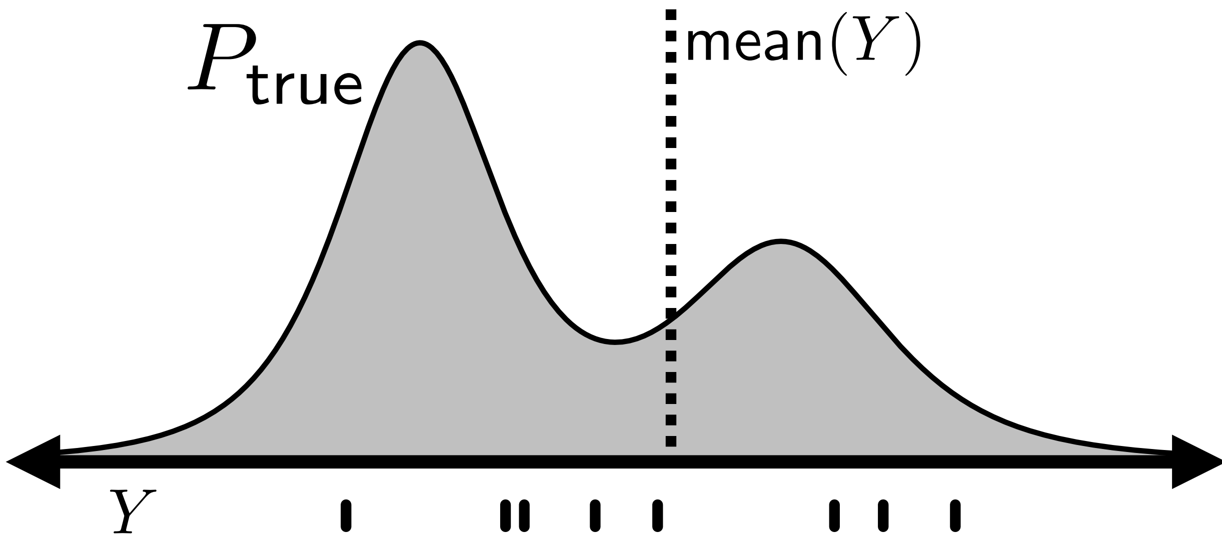
- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$



[Efron 1979]

Bootstrapping

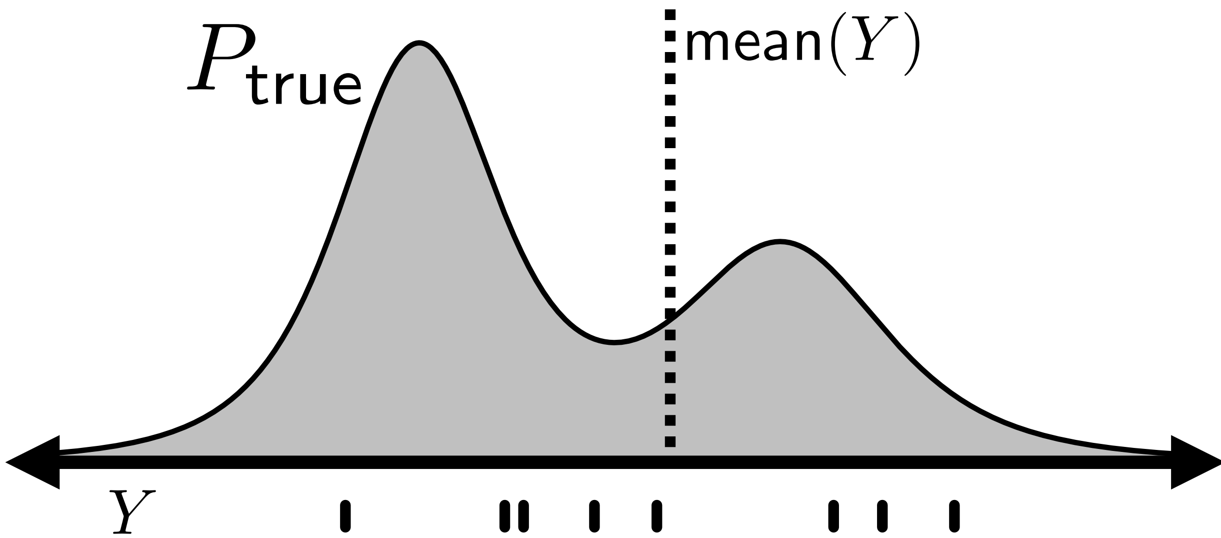
- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**



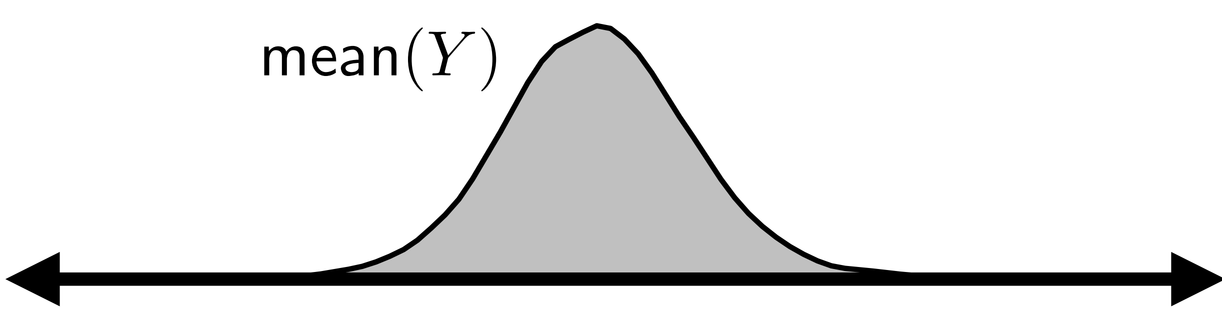
[Efron 1979]

Bootstrapping

- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty [e.g. distribution of mean**(Y) **under** P_{true}**]**



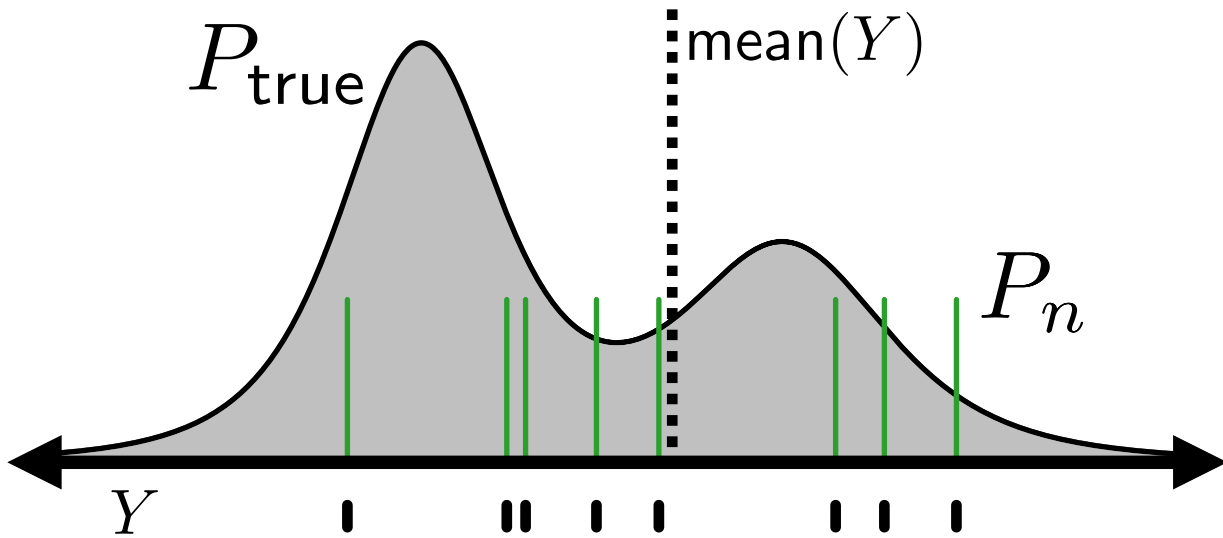
distribution of…



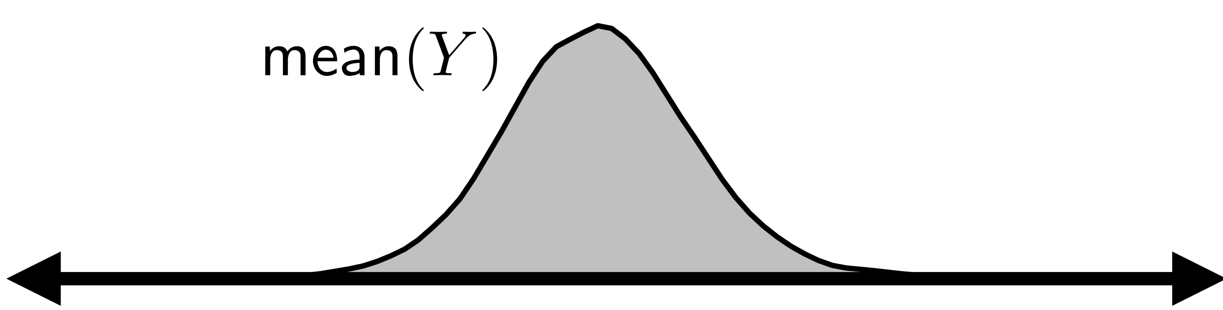
[Efron 1979]

Bootstrapping

- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty [e.g. distribution of mean(Y) under P_{true}]**
- **Bootstrap:** replace P_{true} with P_n

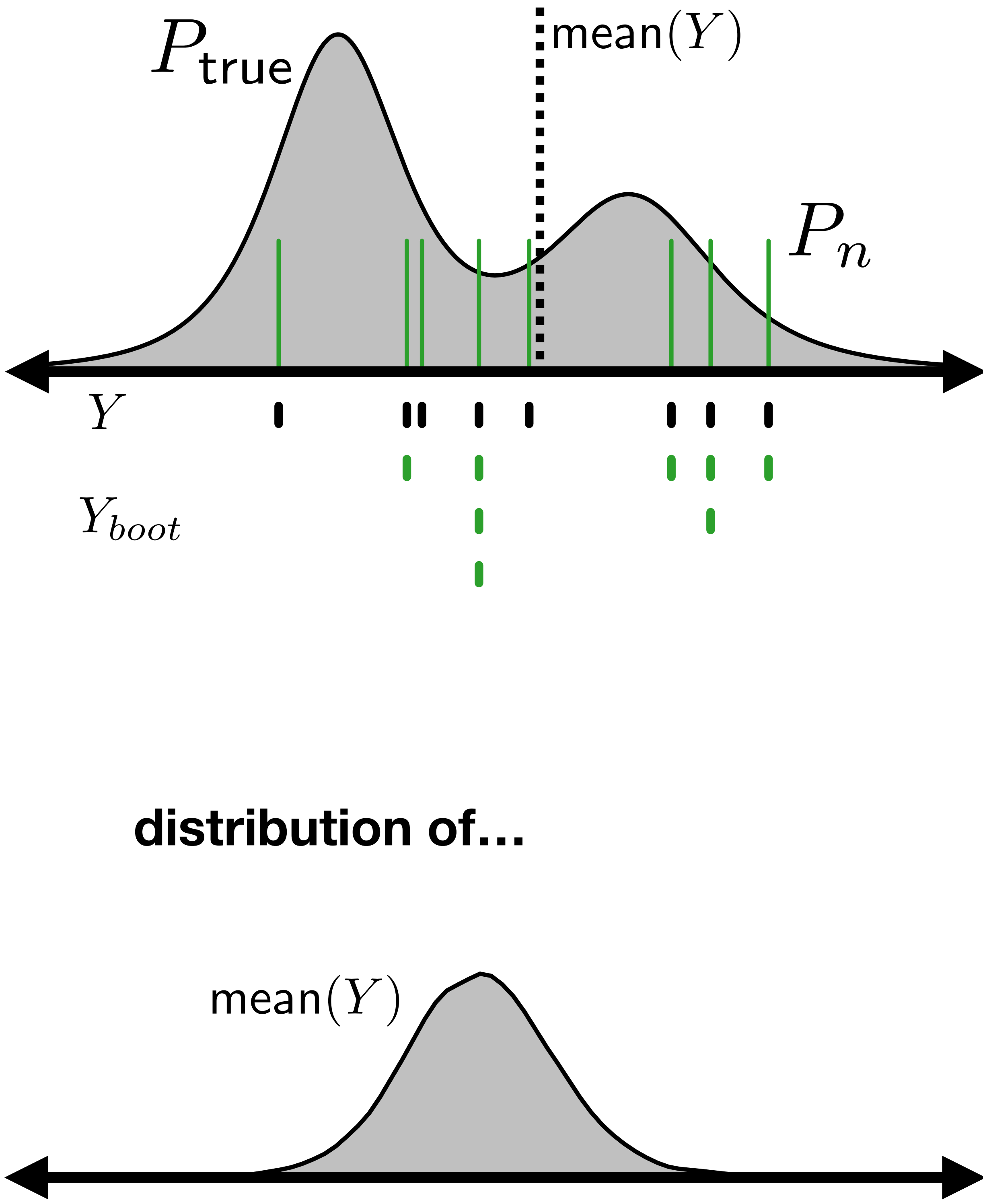


distribution of…



Bootstrapping

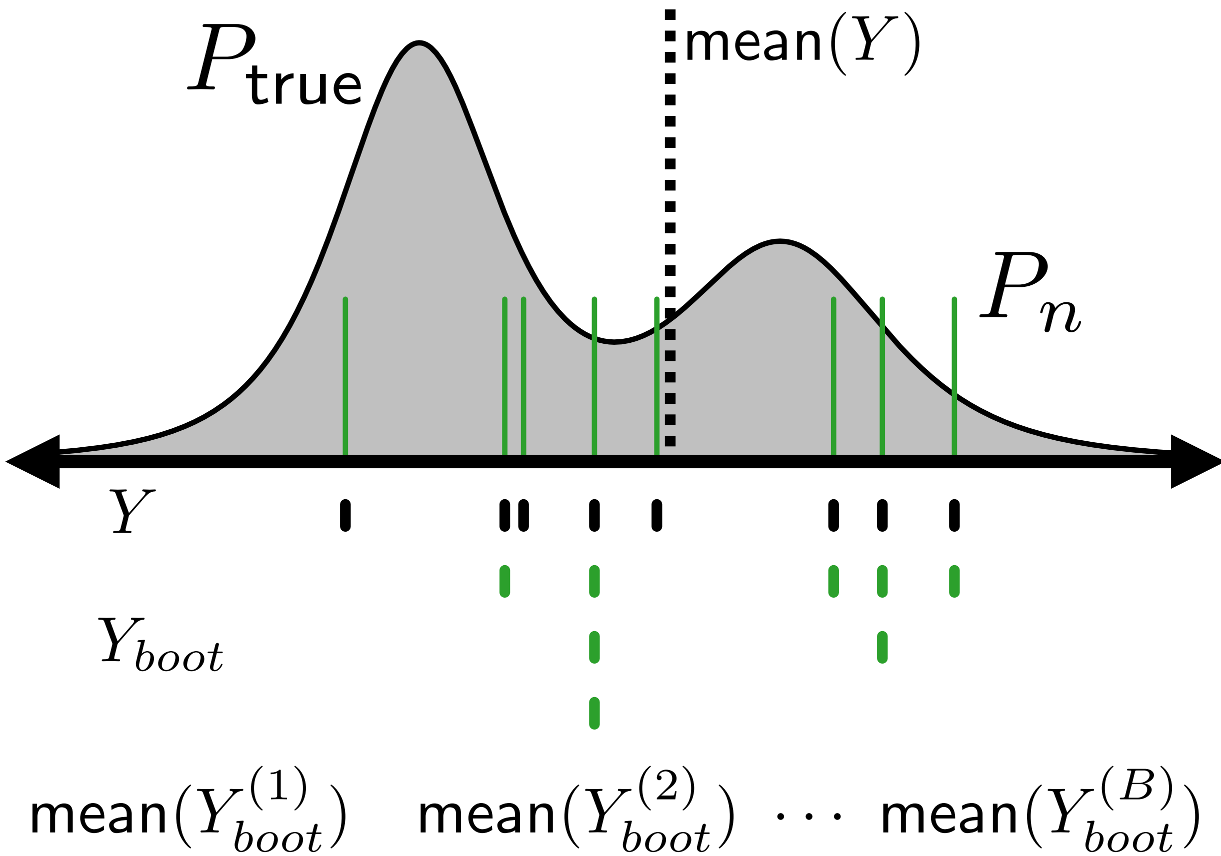
- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty [e.g. distribution of mean(Y) under P_{true}]**
- **Bootstrap:** replace P_{true} with P_n



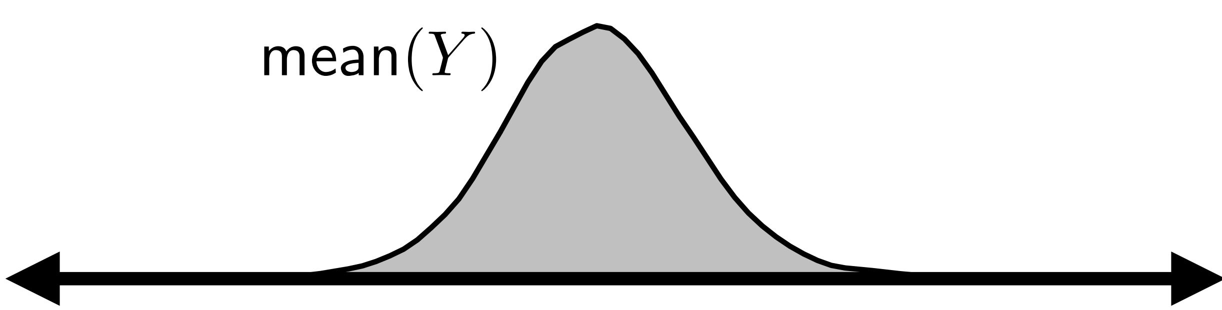
[Efron 1979]

Bootstrapping

- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty** **[e.g. distribution of mean(Y) under P_{true}]**
- **Bootstrap:** replace P_{true} with P_n
- Sample B bootstrap datasets to get empirical distribution **[e.g. mean(Y_{boot})]**

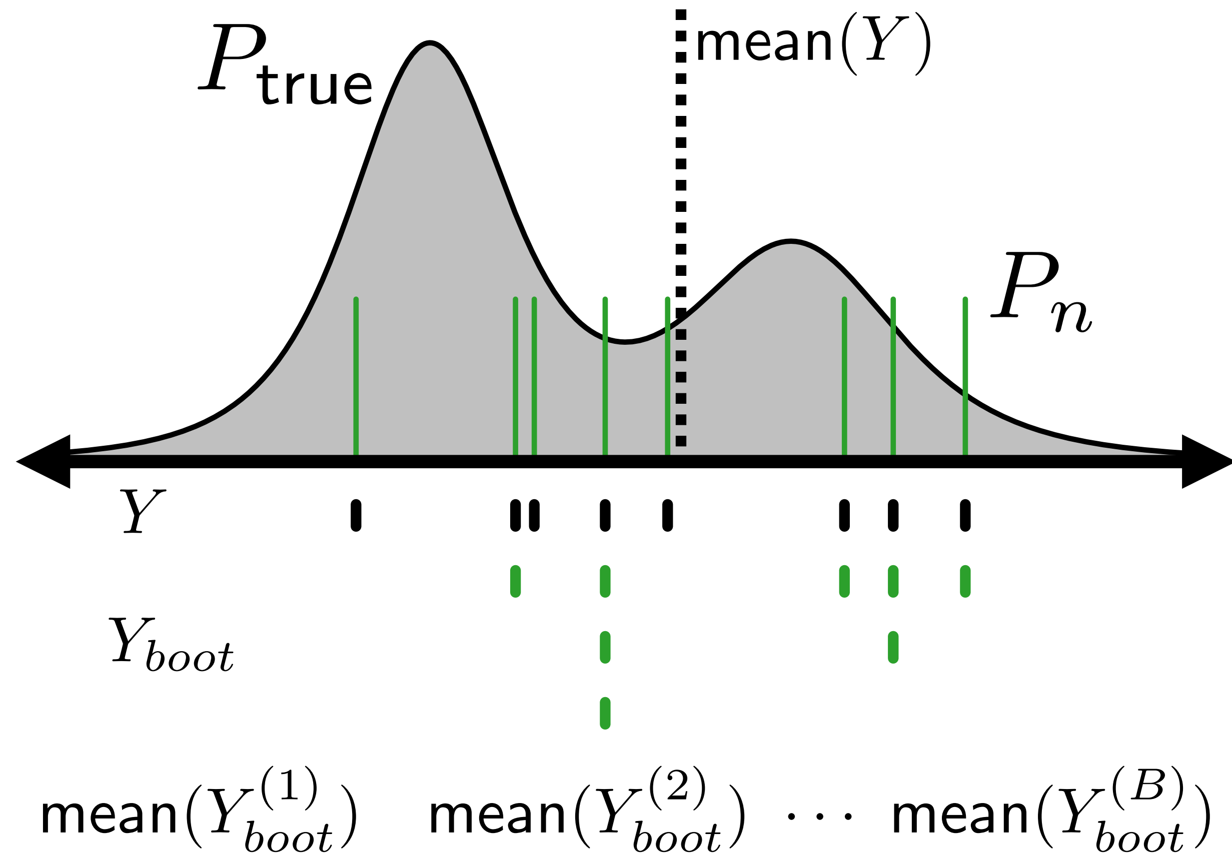


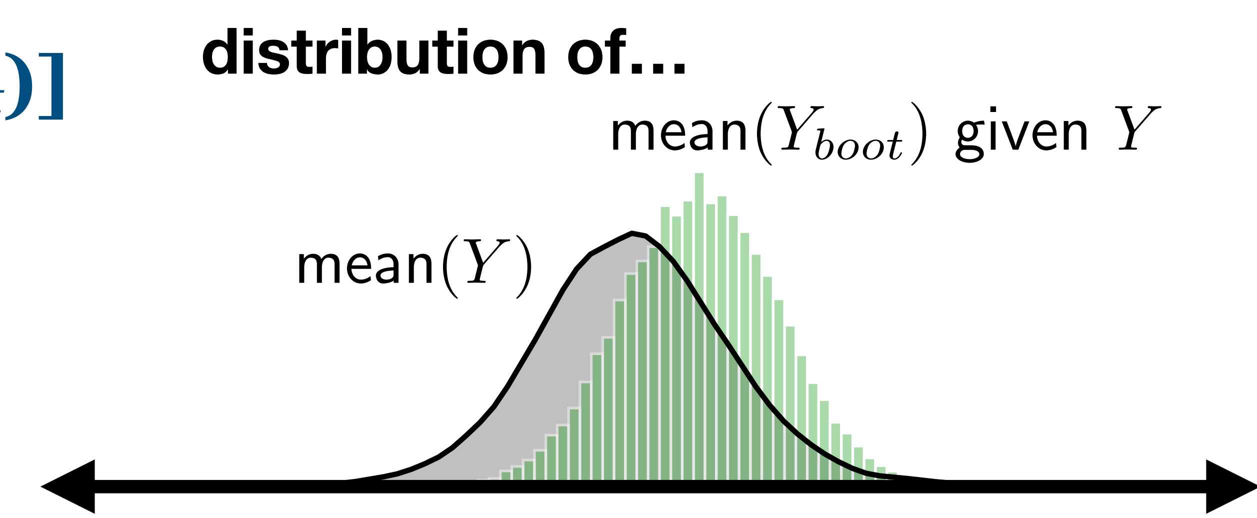
distribution of…



Bootstrapping

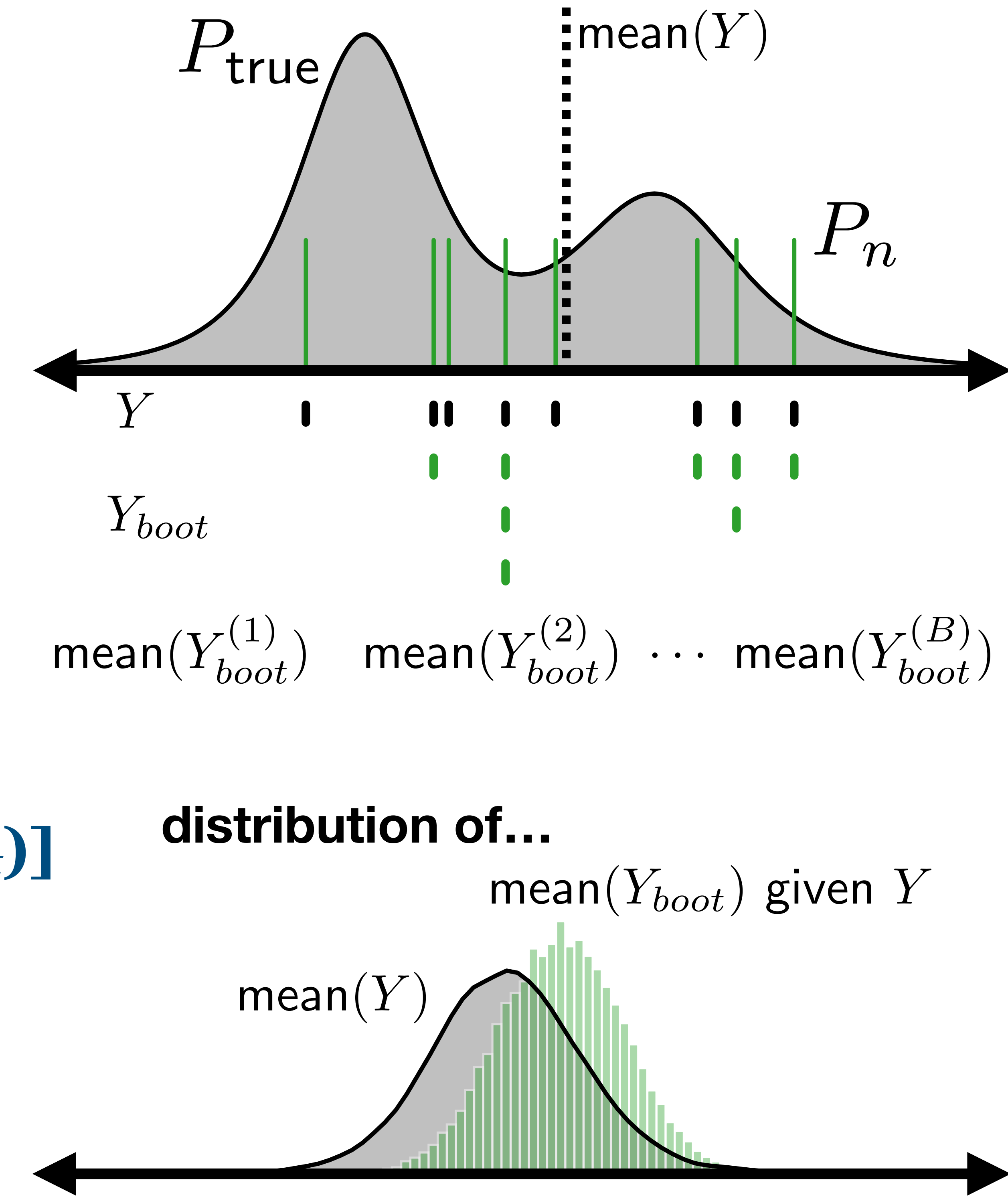
- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty** **[e.g. distribution of mean(Y) under P_{true}]**
- **Bootstrap:** replace P_{true} with P_n
- Sample B bootstrap datasets to get empirical distribution **[e.g. mean(Y_{boot})]**





Bootstrapping

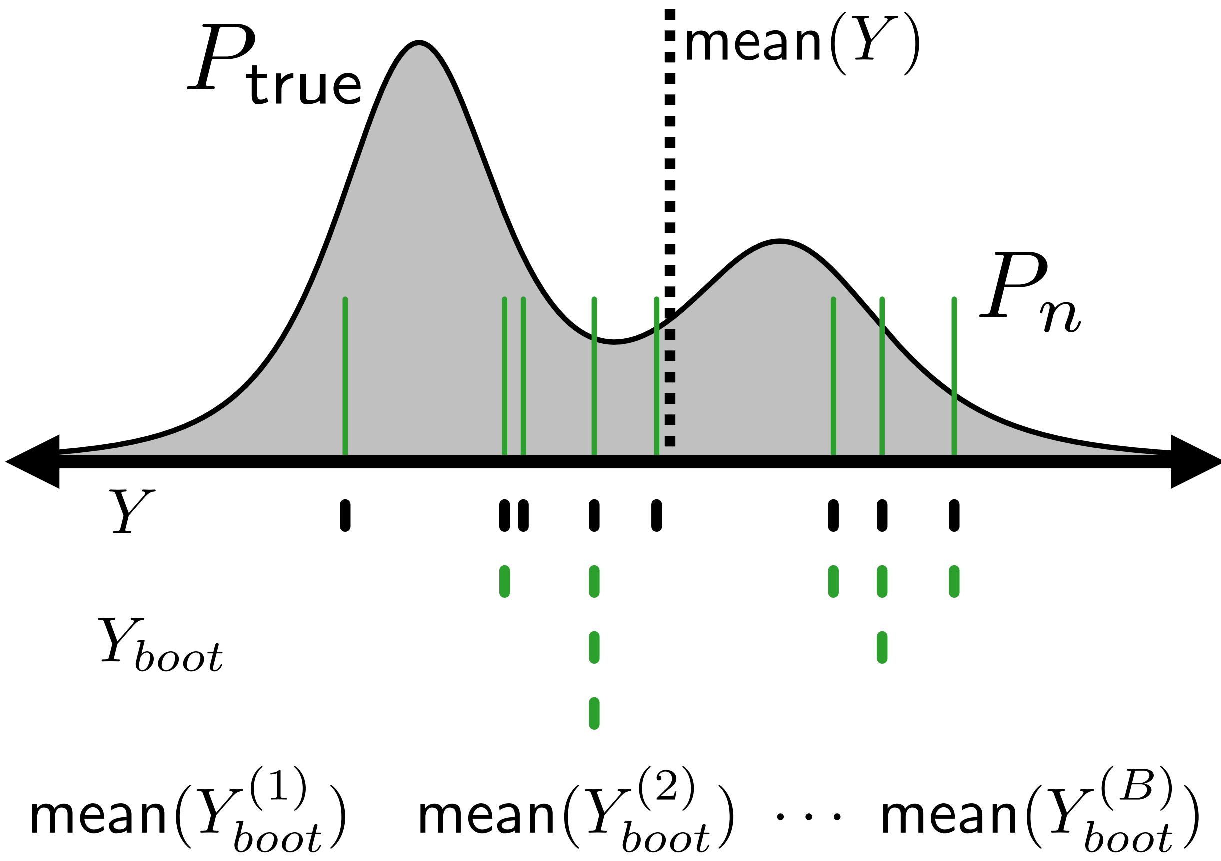
- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty [e.g. distribution of mean(Y) under P_{true}]**
- **Bootstrap:** replace P_{true} with P_n
- Sample B bootstrap datasets to get empirical distribution **[e.g. mean(Y_{boot})]**
- **Benefits:** no assumptions about P_{true}, easy to use, can parallelize across B



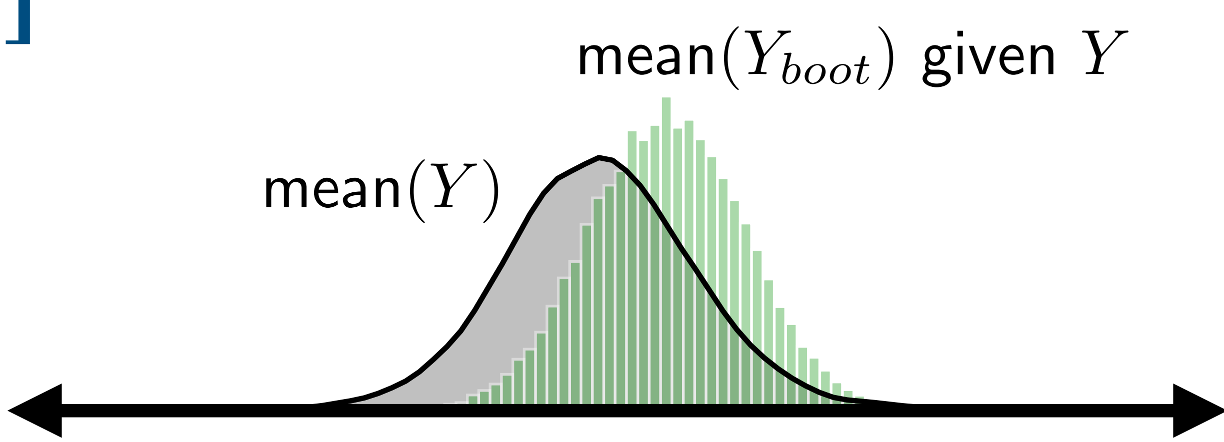
[Efron 1979]

Bootstrapping

- **Data** $Y = (Y_1, \ldots, Y_n)$, where $Y_i \sim P_{\text{true}}$
- Interested in parameter that best explains distribution **[e.g. mean of independent normal observations]**
- Want **sampling uncertainty** **[e.g. distribution of mean(Y) under P_{true}]**
- **Bootstrap:** replace P_{true} with P_n
- Sample B bootstrap datasets to get empirical distribution **[e.g. mean(Y_{boot})]**
- **Benefits:** no assumptions about P_{true}, easy to use, can parallelize across B
- **Challenges:** B large (1,000-100,000), finite-sample properties

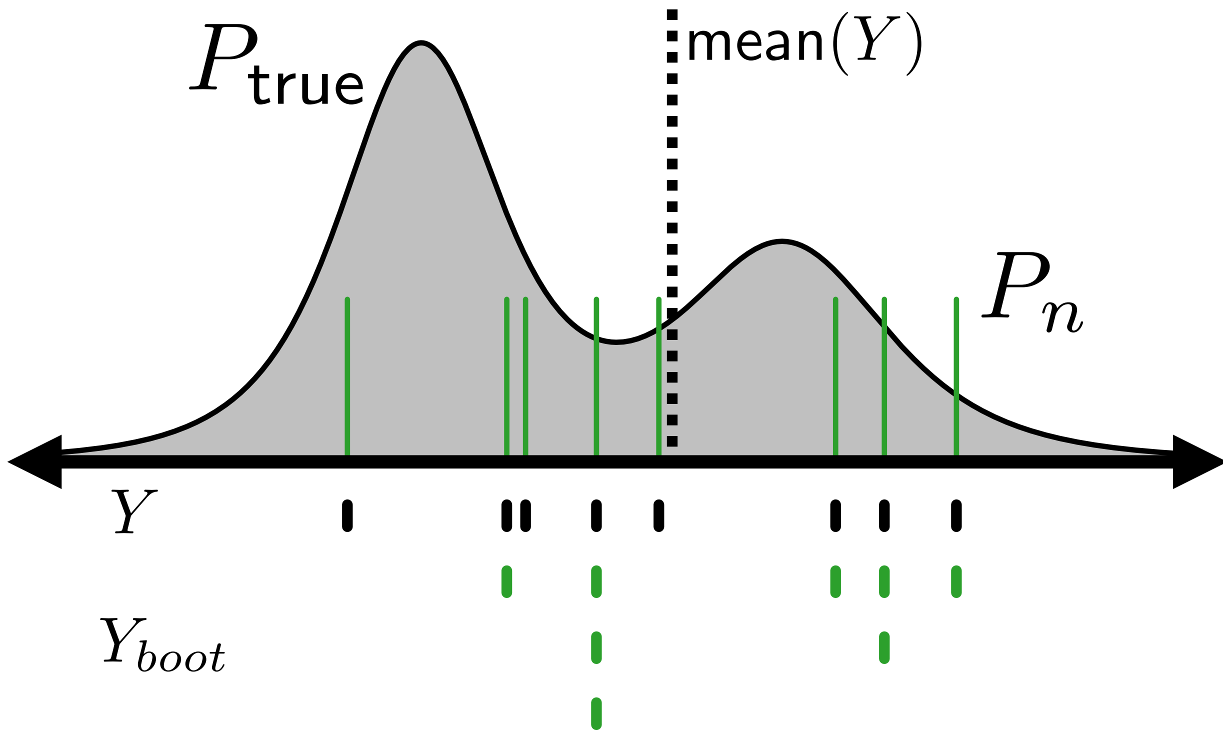


distribution of…



[Efron 1979]

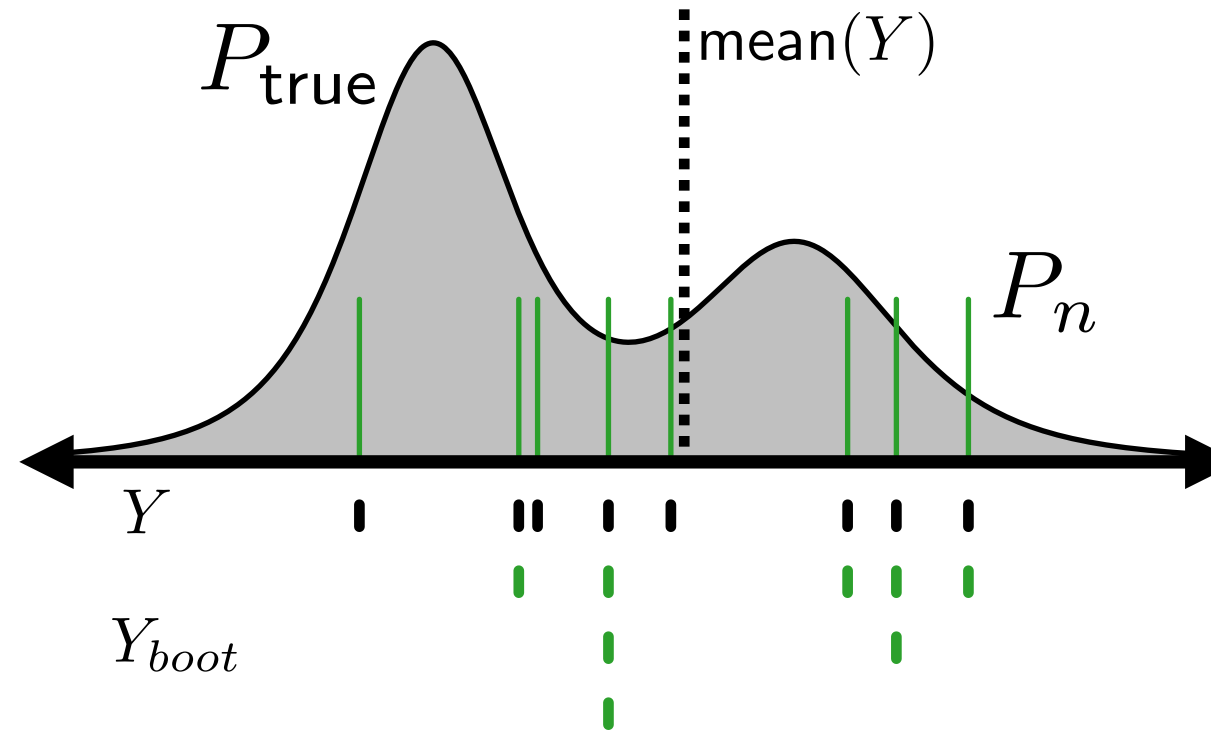
Bootstrapping Bayes



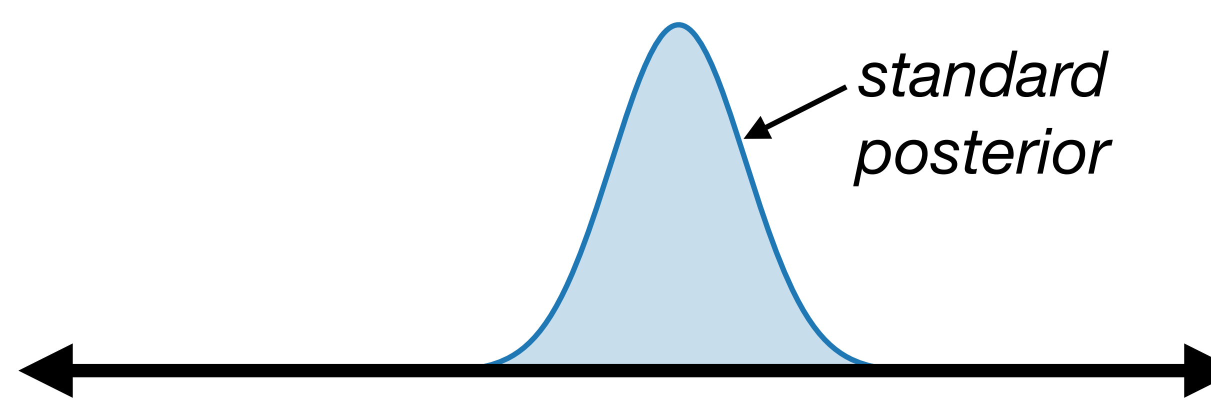
[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

Bootstrapping Bayes

- **Recall:** posterior given data Y is denoted $\pi(\theta \mid Y)$



uncertainty about true mean

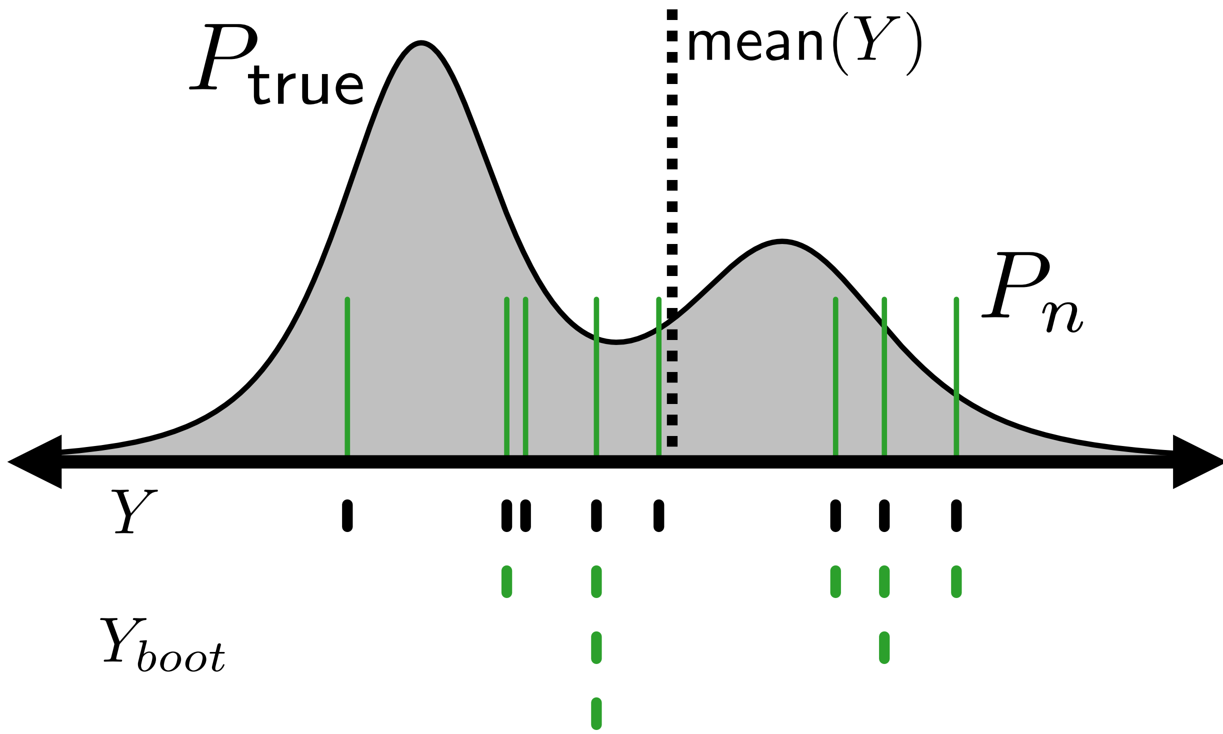


[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

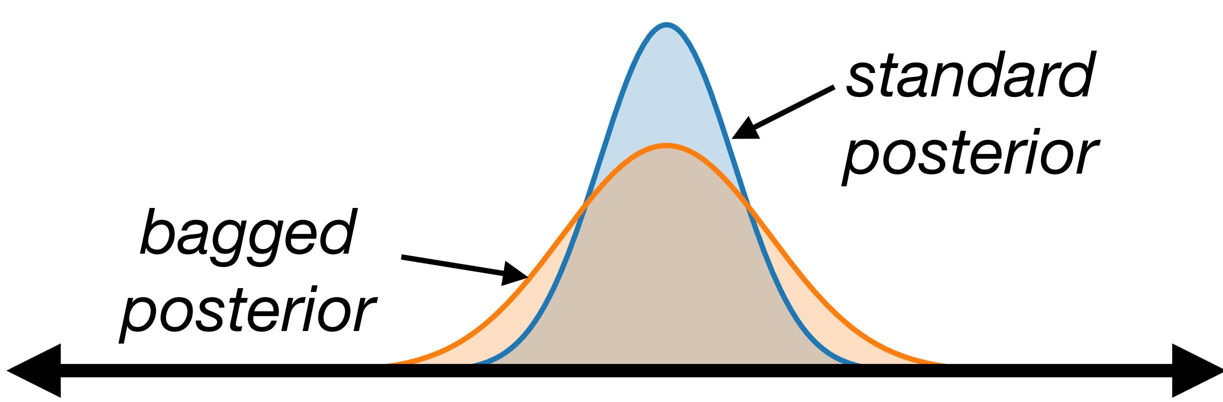
Bootstrapping Bayes

- **Recall:** posterior given data Y is denoted $\pi(\theta \mid Y)$
- **BayesBag method:** Sample B bootstrap datasets and average over posteriors

$$\pi_{BB}(\theta \mid Y) = \frac{1}{B}\sum_{b=1}^{B}\pi(\theta \mid Y_{boot}^{(b)})$$



uncertainty about true mean



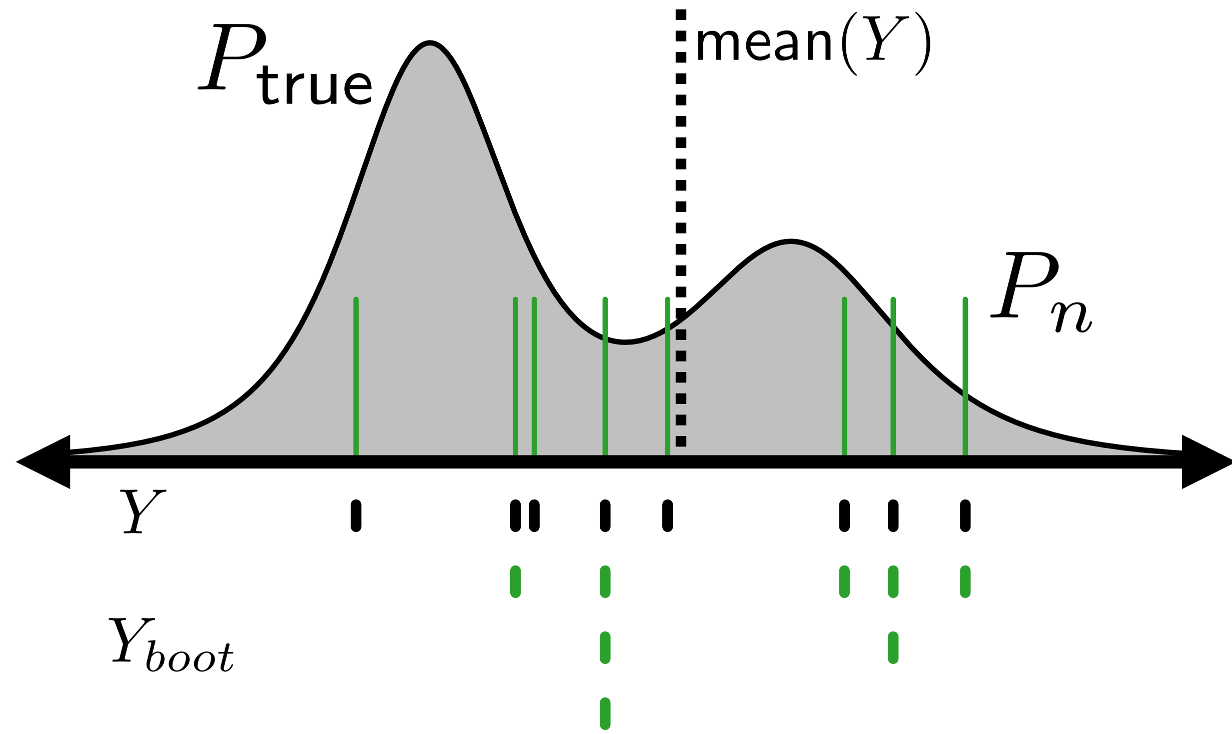
[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

Bootstrapping Bayes

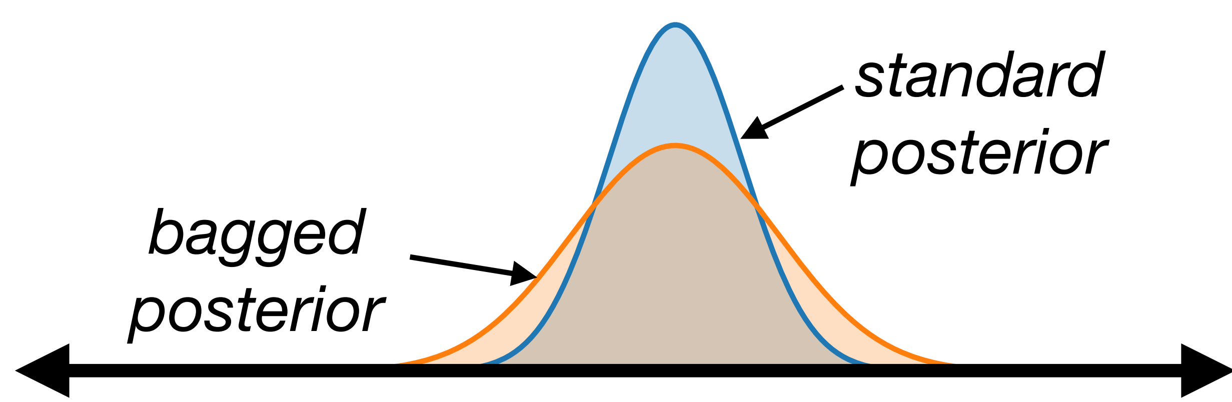
- **Recall:** posterior given data Y is denoted $\pi(\theta \mid Y)$
- **BayesBag method:** Sample B bootstrap datasets and average over posteriors

$$\pi_{BB}(\theta \mid Y) = \frac{1}{B} \sum_{b=1}^{B} \pi(\theta \mid Y_{boot}^{(b)})$$

- Same benefits as bootstrap: no correct model assumption, easy-to-use, can parallelize across B



uncertainty about true mean

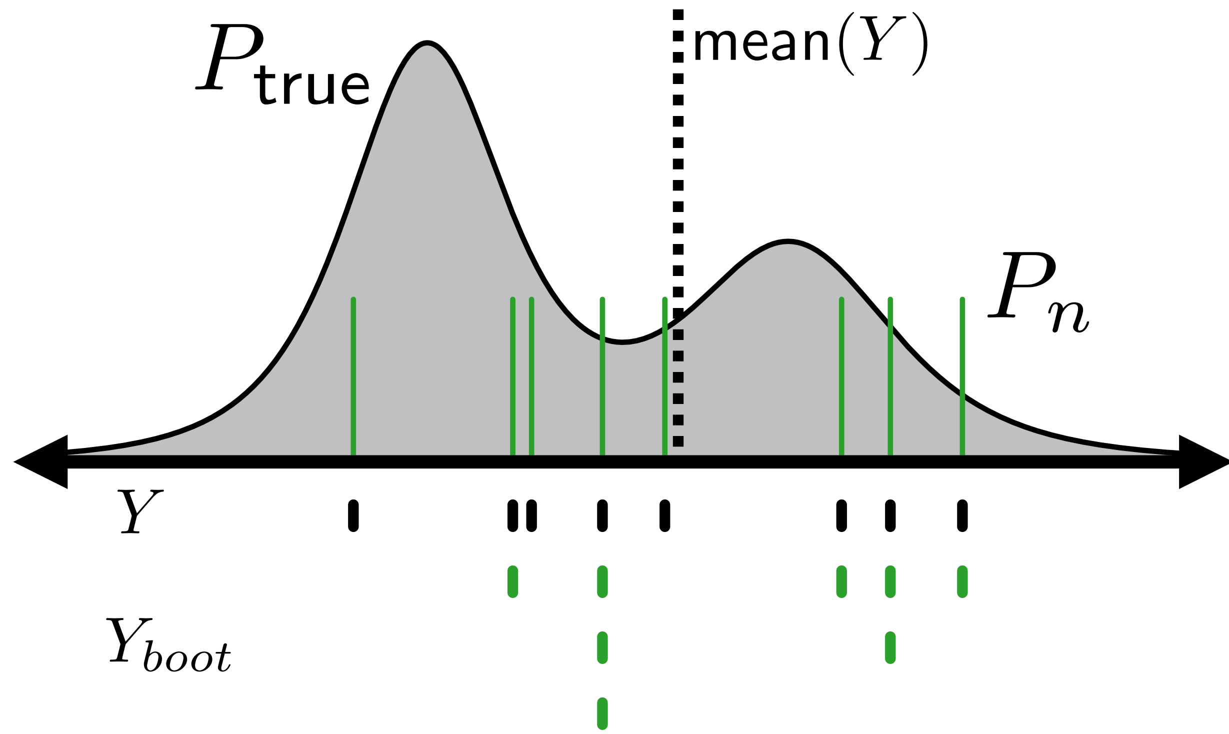


Bootstrapping Bayes

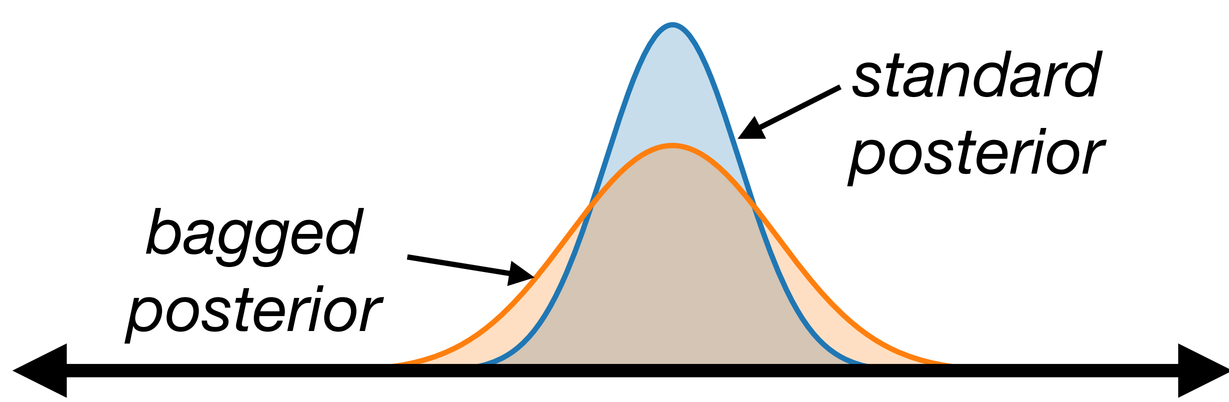
- **Recall:** posterior given data Y is denoted $\pi(\theta \mid Y)$
- **BayesBag method:** Sample B bootstrap datasets and average over posteriors

$$\pi_{BB}(\theta \mid Y) = \frac{1}{B} \sum_{b=1}^{B} \pi(\theta \mid Y_{boot}^{(b)})$$

- Same benefits as bootstrap: no correct model assumption, easy-to-use, can parallelize across B
- Suffices to take B = 50 or 100



uncertainty about true mean



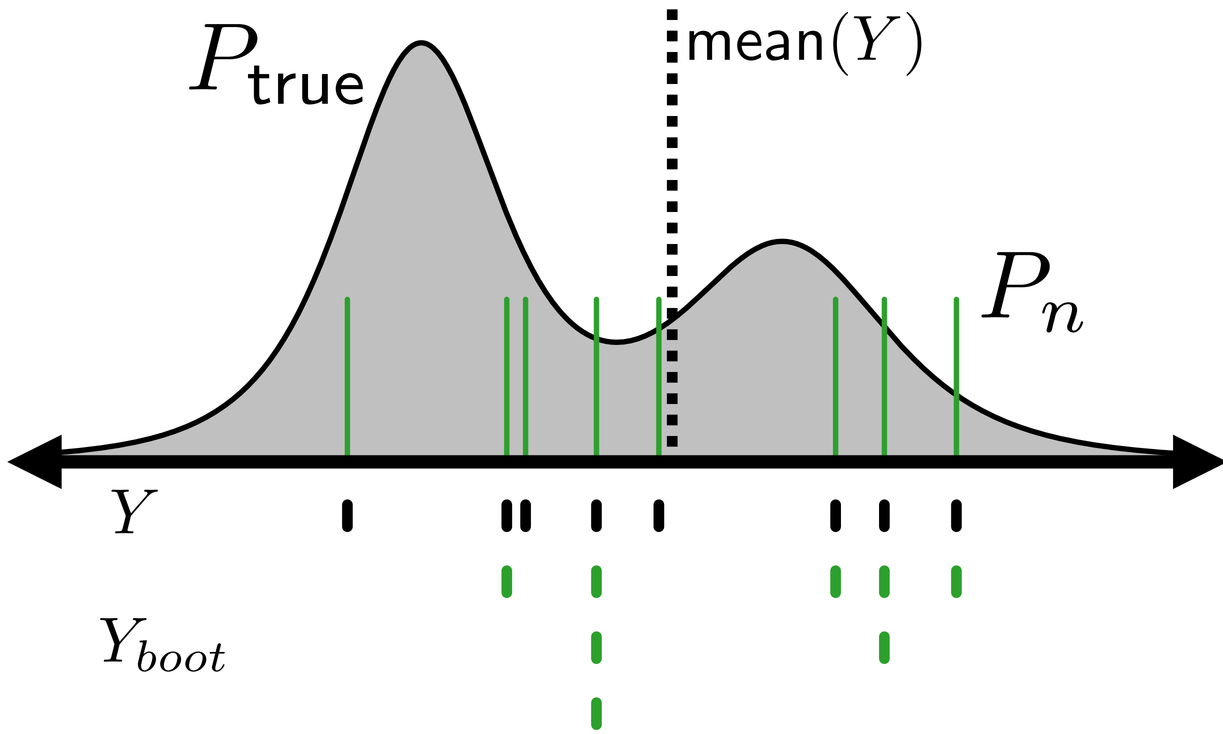
[Douady et al. 2003, Bühlmann 2014, **H** & Miller 2019]

Bootstrapping Bayes

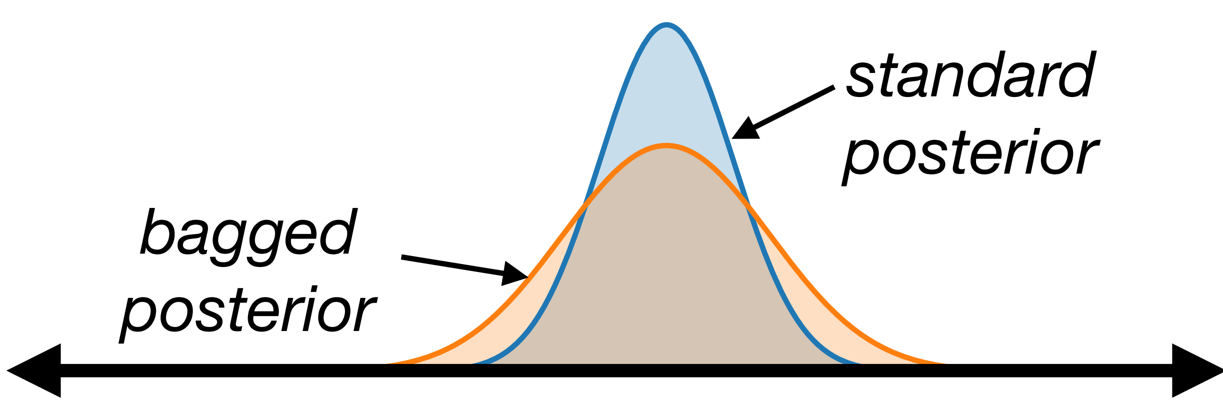
- **Recall:** posterior given data Y is denoted $\pi(\theta \mid Y)$
- **BayesBag method:** Sample B bootstrap datasets and average over posteriors

$$\pi_{BB}(\theta \mid Y) = \frac{1}{B}\sum_{b=1}^{B} \pi(\theta \mid Y_{boot}^{(b)})$$

- Same benefits as bootstrap: no correct model assumption, easy-to-use, can parallelize across B
- Suffices to take B = 50 or 100
- Finite-sample benefits of Bayes



uncertainty about true mean



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

Better parameter inference with BayesBag

Better parameter inference with BayesBag

- Assumed model: Gaussian linear regression with conjugate priors

Better parameter inference with BayesBag

- Assumed model: Gaussian linear regression with conjugate priors
- Data-generating distribution P_{true}: can be correct or misspecified

Better parameter inference with BayesBag

- Assumed model: Gaussian linear regression with conjugate priors
- Data-generating distribution P_{true}: can be correct or misspecified
- $\theta_{\text{opt}} =$ optimal parameter that is "closest" to P_{true}

[**H** & Miller 2019]

Better parameter inference with BayesBag

- Assumed model: Gaussian linear regression with conjugate priors
- Data-generating distribution P_{true}: can be correct or misspecified
- θ_{opt} = optimal parameter that is "closest" to P_{true}
- Performance metric is difference in log posterior density at θ_{opt}:

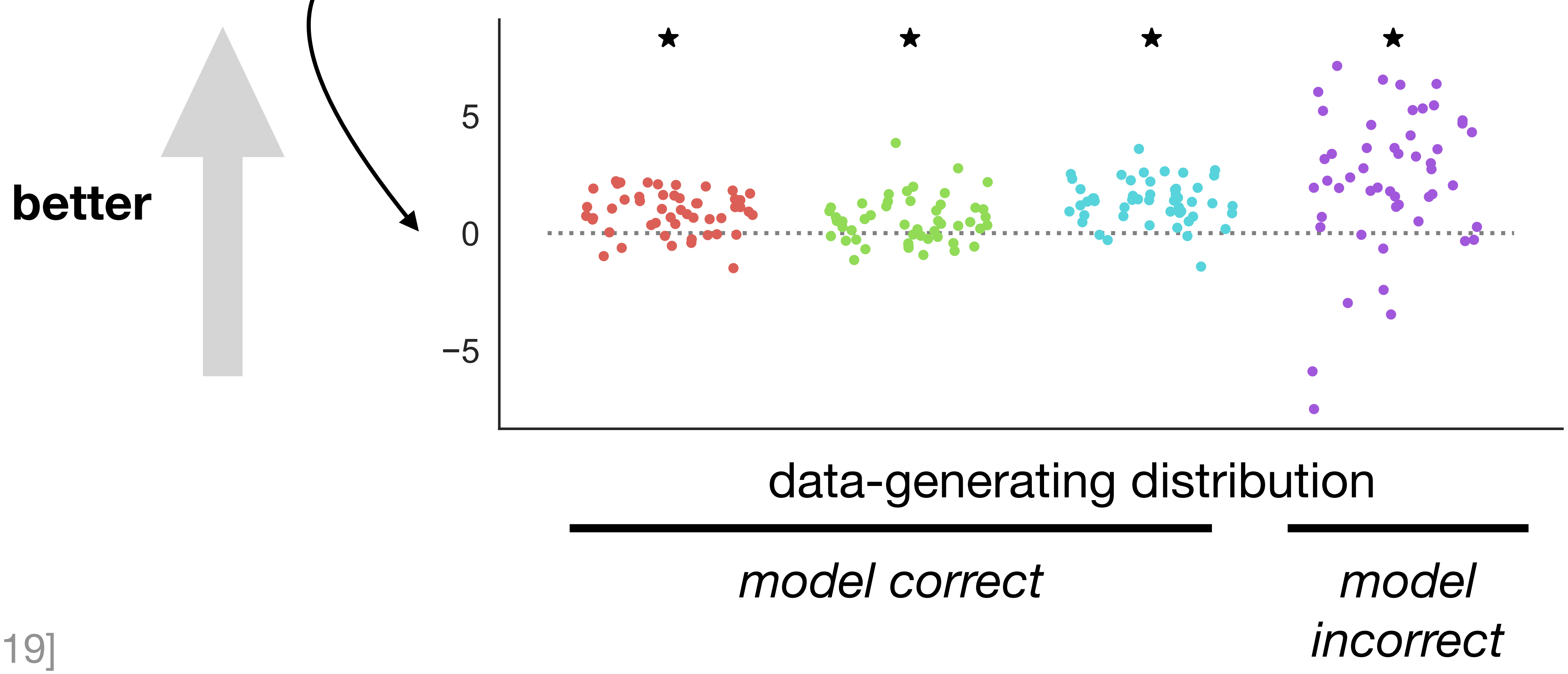
$$\log \pi_{BB}(\theta_{\text{opt}} \,|\, Y) - \log \pi(\theta_{\text{opt}} \,|\, Y)$$

[H & Miller 2019]

Better parameter inference with BayesBag

- Assumed model: Gaussian linear regression with conjugate priors
- Data-generating distribution P_{true}: can be correct or misspecified
- θ_{opt} = optimal parameter that is "closest" to P_{true}
- Performance metric is difference in log posterior density at θ_{opt}:

$$\log \pi_{BB}(\theta_{\text{opt}} \mid Y) - \log \pi(\theta_{\text{opt}} \mid Y)$$



[H & Miller 2019]

Agenda

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

BayesBag incorporates model- and sampling-based uncertainty

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$

point estimate

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \mid Y)$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \,|\, Y)$

Posterior variance: $\mathrm{Var}(\vartheta \,|\, Y)$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \,|\, Y)$

Posterior variance: $\mathrm{Var}(\vartheta \,|\, Y)$ ← *model-based uncertainty*

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \mid Y)$

Posterior variance: $\mathrm{Var}(\vartheta \mid Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \mid Y)$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \,|\, Y)$

Posterior variance: $\mathrm{Var}(\vartheta \,|\, Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \,|\, Y)$

BayesBag posterior variance:

$$\mathrm{Var}(\vartheta_{BB} \,|\, Y)$$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \mid Y)$

Posterior variance: $\mathrm{Var}(\vartheta \mid Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \mid Y)$

BayesBag posterior variance:

$$\mathrm{Var}(\vartheta_{BB} \mid Y) = \mathbb{E}\{\mathrm{Var}(\vartheta_{BB} \mid Y_{boot})\} + \mathrm{Var}\{\mathbb{E}(\vartheta_{BB} \mid Y_{boot})\}$$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \,|\, Y)$

Posterior variance: $\mathrm{Var}(\vartheta \,|\, Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \,|\, Y)$

BayesBag posterior variance:

$$\mathrm{Var}(\vartheta_{BB} \,|\, Y) = \underbrace{\mathbb{E}\{\mathrm{Var}(\vartheta_{BB} \,|\, Y_{boot})\}}_{\text{expected posterior variance}} + \mathrm{Var}\{\mathbb{E}(\vartheta_{BB} \,|\, Y_{boot})\}$$

[H & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \mid Y)$

Posterior variance: $\mathrm{Var}(\vartheta \mid Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \mid Y)$

BayesBag posterior variance:

$$\mathrm{Var}(\vartheta_{BB} \mid Y) = \mathbb{E}\{\mathrm{Var}(\vartheta_{BB} \mid Y_{boot})\} + \mathrm{Var}\{\mathbb{E}(\vartheta_{BB} \mid Y_{boot})\}$$

expected posterior variance

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

↖ *point estimate*

Sample from posterior: $\vartheta \sim \pi(\theta \mid Y)$

Posterior variance: $\mathrm{Var}(\vartheta \mid Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \mid Y)$

BayesBag posterior variance:

$$\mathrm{Var}(\vartheta_{BB} \mid Y) = \underbrace{\mathbb{E}\{\mathrm{Var}(\vartheta_{BB} \mid Y_{boot})\}}_{\text{expected posterior variance}} + \underbrace{\mathrm{Var}\{\mathbb{E}(\vartheta_{BB} \mid Y_{boot})\}}_{\text{variance of posterior mean}}$$

[H & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \,|\, Y)$

Posterior variance: $\mathrm{Var}(\vartheta \,|\, Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \,|\, Y)$

BayesBag posterior variance:

point estimate

$$\mathrm{Var}(\vartheta_{BB} \,|\, Y) = \underbrace{\mathbb{E}\{\mathrm{Var}(\vartheta_{BB} \,|\, Y_{boot})\}}_{\text{expected posterior variance}} + \underbrace{\mathrm{Var}\{\mathbb{E}(\vartheta_{BB} \,|\, Y_{boot})\}}_{\text{variance of posterior mean}}$$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

Bootstrap variance: $\mathrm{Var}\{\hat{\theta}(Y_{boot})\}$ ← *sampling uncertainty*

point estimate

Sample from posterior: $\vartheta \sim \pi(\theta \mid Y)$

Posterior variance: $\mathrm{Var}(\vartheta \mid Y)$ ← *model-based uncertainty*

Sample from BayesBag posterior: $\vartheta_{BB} \sim \pi_{BB}(\theta \mid Y)$

BayesBag posterior variance:

point estimate

$$\mathrm{Var}(\vartheta_{BB} \mid Y) = \underbrace{\mathbb{E}\{\mathrm{Var}(\vartheta_{BB} \mid Y_{boot})\}}_{\text{expected posterior variance}} + \underbrace{\mathrm{Var}\{\mathbb{E}(\vartheta_{BB} \mid Y_{boot})\}}_{\text{variance of posterior mean}}$$

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

- Summarizing the previous slide…

BayesBag incorporates model- and sampling-based uncertainty

- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

BayesBag incorporates model- and sampling-based uncertainty

- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*

[**H** & Miller 2019]

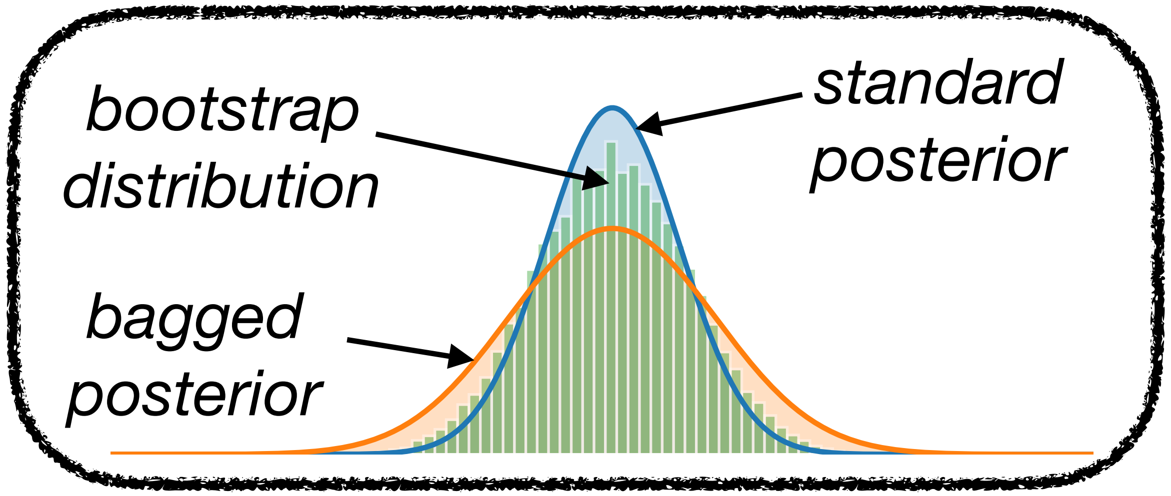
BayesBag incorporates model- and sampling-based uncertainty

- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*



[H & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

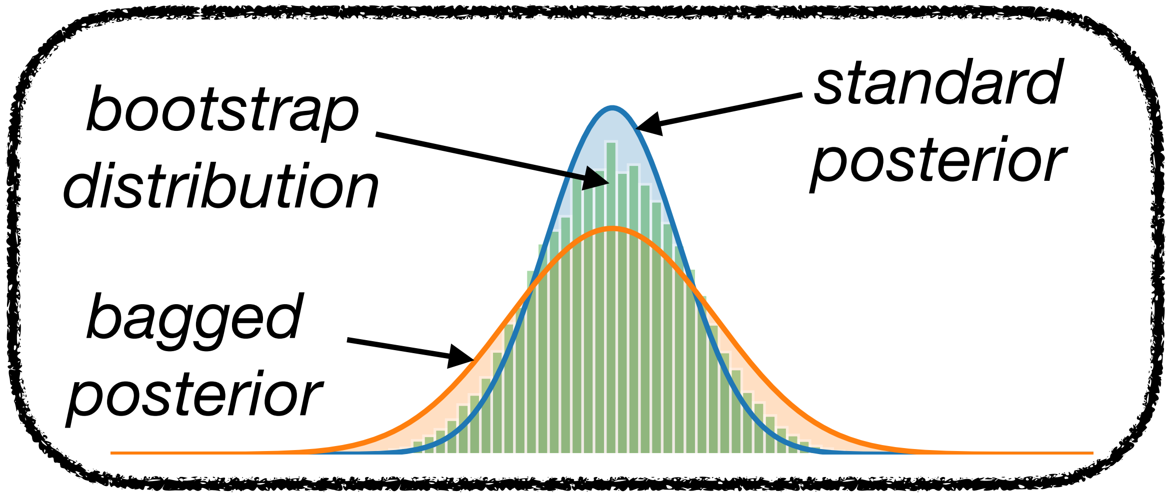
- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based + sampling-based uncertainty*

- Model correct: *model-based uncertainty = sampling-based uncertainty*



[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

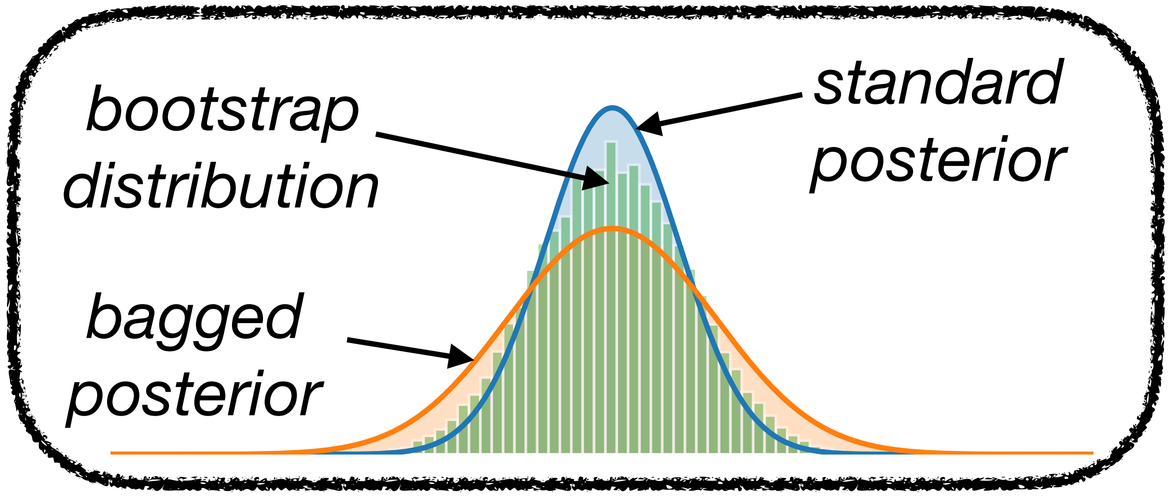
- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*

- Model correct: *model-based uncertainty* = *sampling-based uncertainty*
  - Posterior and bootstrap variances correct



BayesBag incorporates model- and sampling-based uncertainty

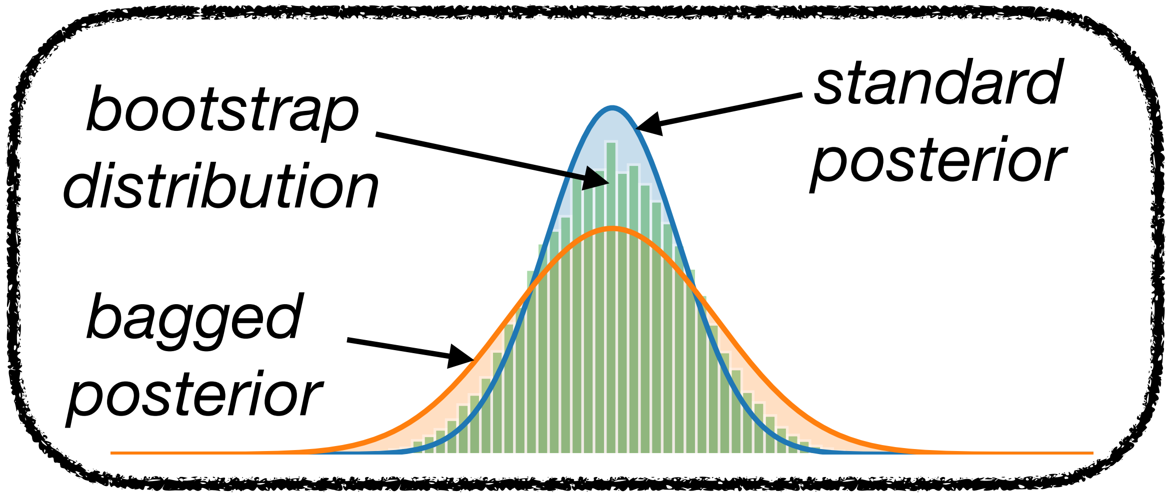
- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*

- Model correct: *model-based uncertainty* = *sampling-based uncertainty*
  - Posterior and bootstrap variances correct
  - BayesBag variance double-counts true uncertainty (conservative)



[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

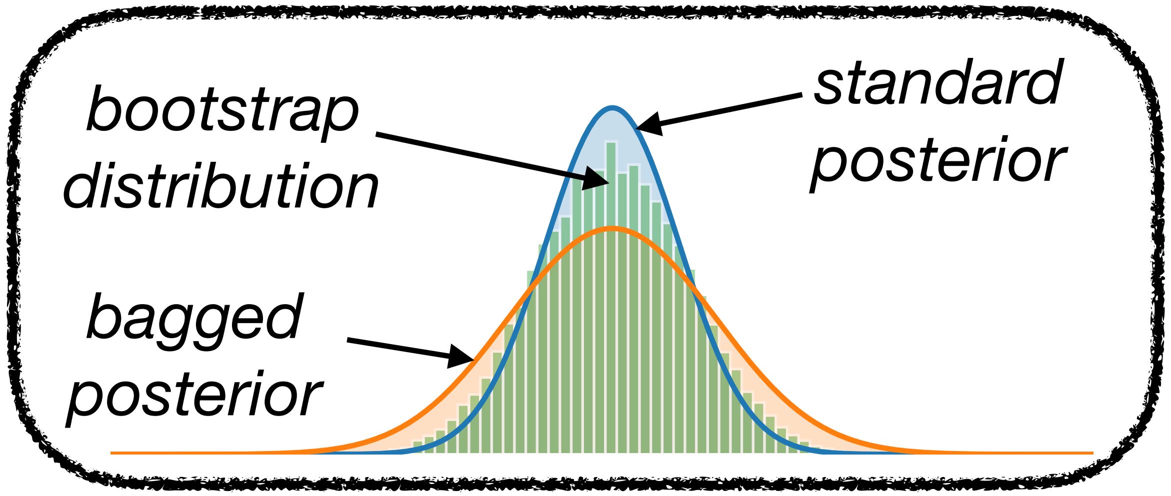
- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*

- Model correct: *model-based uncertainty* = *sampling-based uncertainty*
  - Posterior and bootstrap variances correct
  - BayesBag variance double-counts true uncertainty (conservative)
- Model incorrect: *model-based uncertainty* ≪ *sampling-based uncertainty*



[**H** & Miller 2019]

BayesBag incorporates model- and sampling-based uncertainty

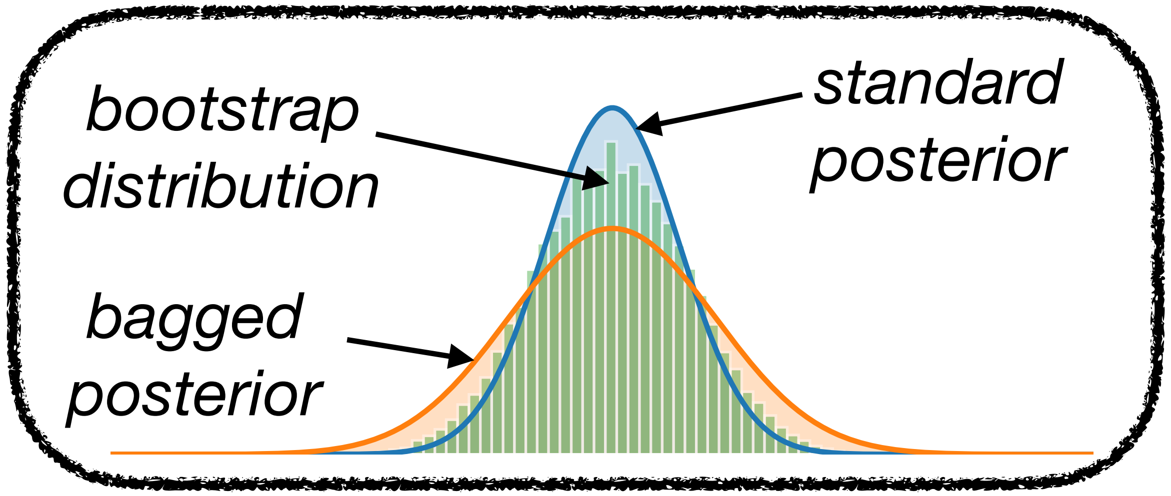
- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*

- Model correct: *model-based uncertainty* = *sampling-based uncertainty*
  - Posterior and bootstrap variances correct
  - BayesBag variance double-counts true uncertainty (conservative)
- Model incorrect: *model-based uncertainty* ≪ *sampling-based uncertainty*
  - Posterior variance far too small



[**H** & Miller 2019]

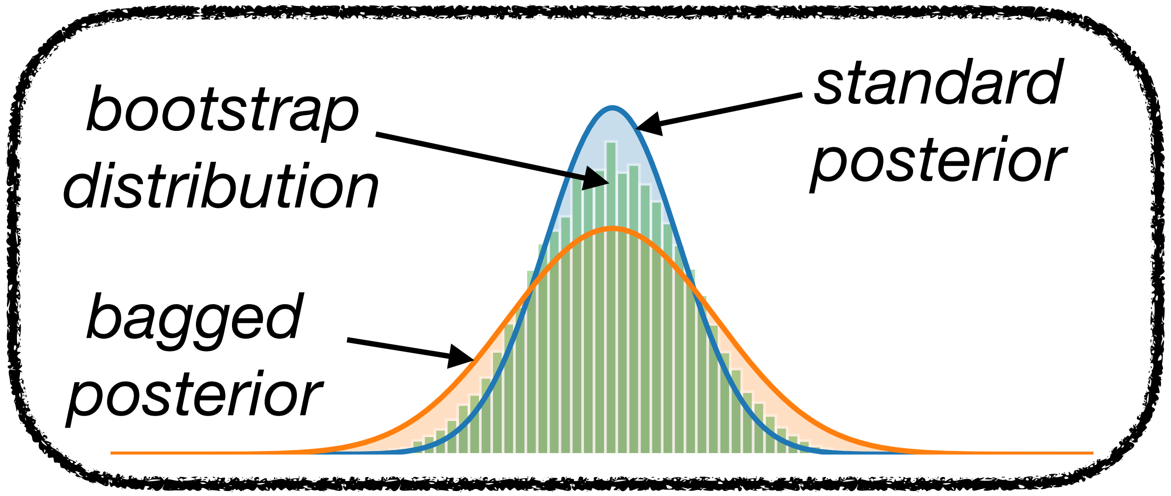
BayesBag incorporates model- and sampling-based uncertainty

- Summarizing the previous slide…

  Posterior variance = *model-based uncertainty*

  Bootstrap variance = *sampling-based uncertainty*

  BayesBag variance = *model-based* + *sampling-based uncertainty*
- Model correct: *model-based uncertainty* = *sampling-based uncertainty*
  - Posterior and bootstrap variances correct
  - BayesBag variance double-counts true uncertainty (conservative)
- Model incorrect: *model-based uncertainty* ≪ *sampling-based uncertainty*
  - Posterior variance far too small
  - BayesBag variance appropriately calibrated



[H & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

[**H** & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** $\mathcal{I}$ can diagnose when data disagrees with assumed model

[**H** & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
    - $-1 < \mathrm{I} < 1$

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
    - $-1 < \mathrm{I} < 1$
    - $\mathrm{I} \approx 0$: no disagreement

[**H** & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
  - $-1 < \mathrm{I} < 1$
  - $\mathrm{I} \approx 0$: no disagreement
  - $\mathrm{I} > 0$: posterior overconfident

[**H** & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
  - $-1 < \mathrm{I} < 1$
  - $\mathrm{I} \approx 0$: no disagreement
  - $\mathrm{I} > 0$: posterior overconfident
  - $\mathrm{I} < 0$: posterior under-confident

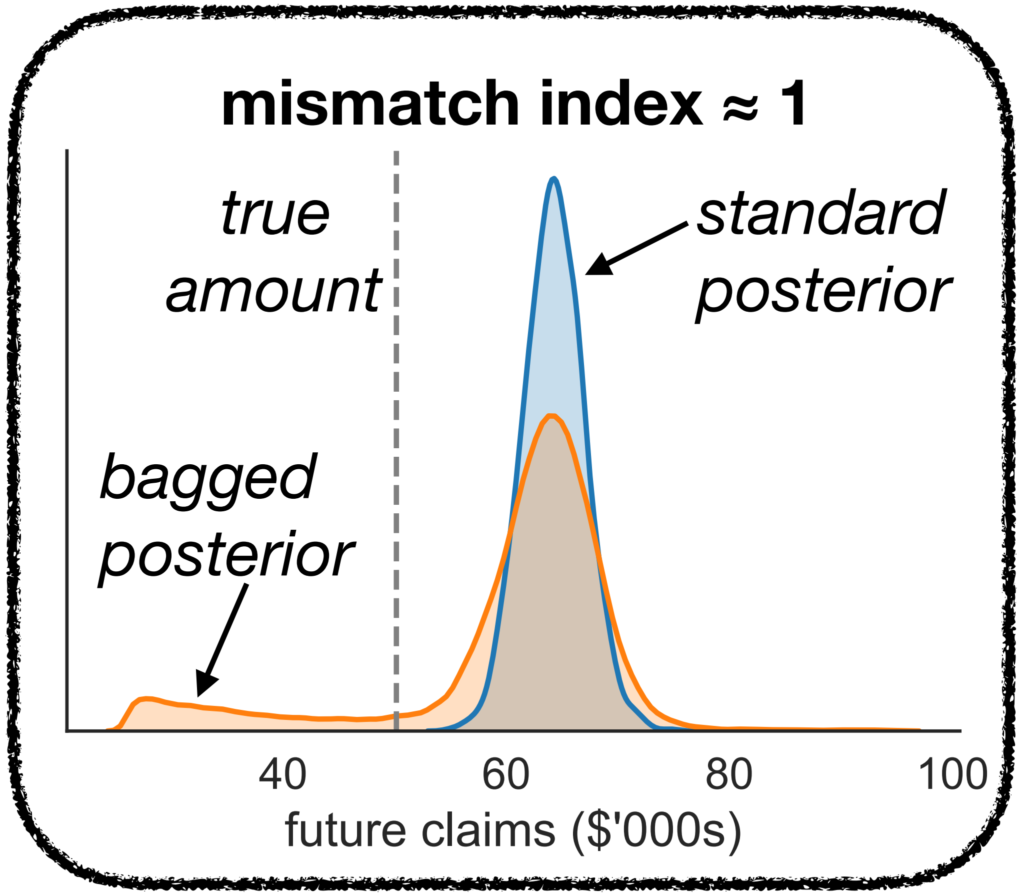
[H & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
  - $-1 < \mathrm{I} < 1$
  - $\mathrm{I} \approx 0$: no disagreement
  - $\mathrm{I} > 0$: posterior overconfident
  - $\mathrm{I} < 0$: posterior under-confident



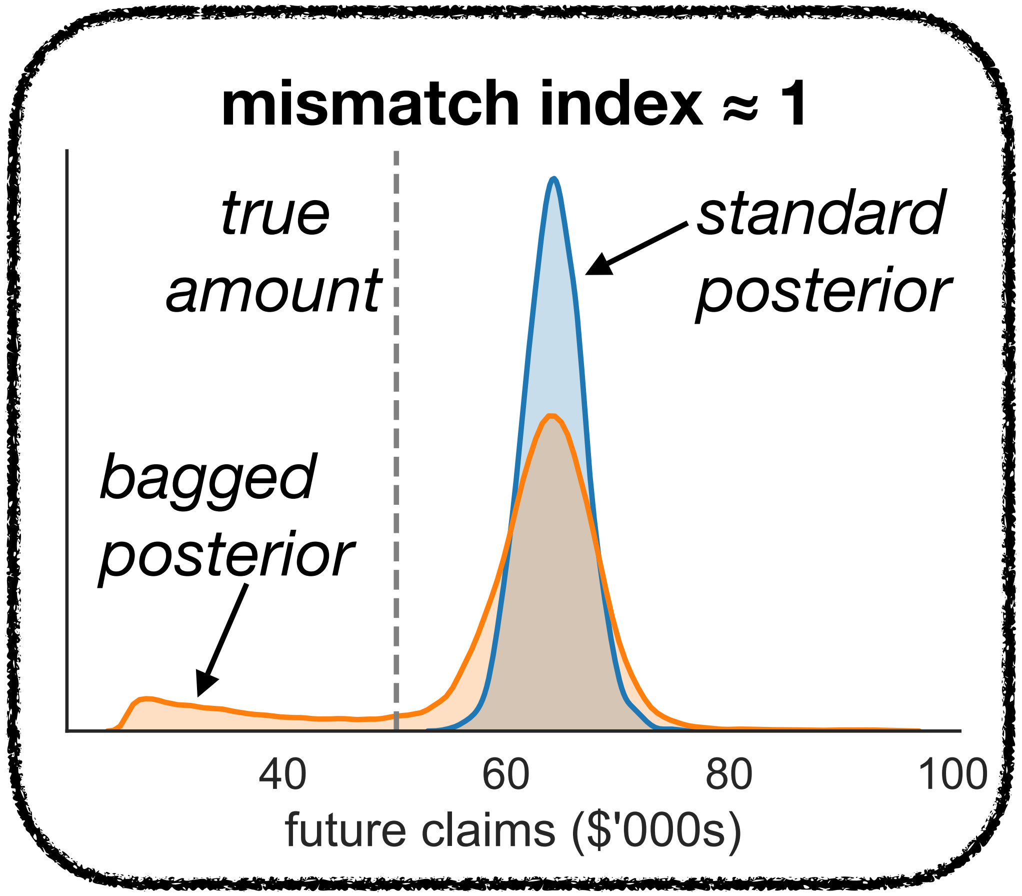
[H & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
  - $-1 < \mathrm{I} < 1$
  - $\mathrm{I} \approx 0$: no disagreement
  - $\mathrm{I} > 0$: posterior overconfident
  - $\mathrm{I} < 0$: posterior under-confident
- **Model criticism:** mismatch index indicates when model needs improvement



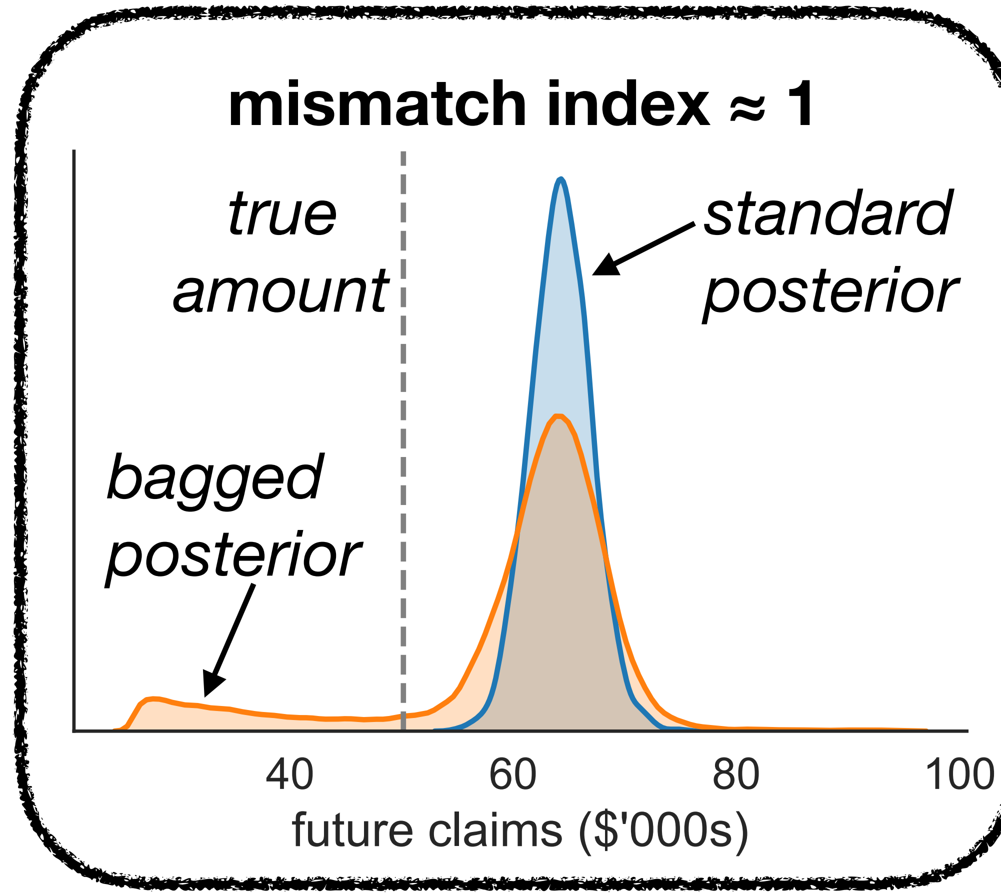
[H & Miller 2019]

Diagnosing model–data mismatch

Posterior variance = *model-based uncertainty*

BayesBag variance = *model-based* + *sampling-based uncertainty*

- **Mismatch index** I can diagnose when data disagrees with assumed model
  - $-1 < \mathrm{I} < 1$
  - $\mathrm{I} \approx 0$: no disagreement
  - $\mathrm{I} > 0$: posterior overconfident
  - $\mathrm{I} < 0$: posterior under-confident



- **Model criticism:** mismatch index indicates when model needs improvement
- Mismatch index can also detect problems with the prior

BayesBag in practice

[H & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \mid Y)$

BayesBag in practice

1) compute standard posterior $\pi(\cdot \mid Y)$

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

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \mid Y)$

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

2) compute bagged posterior $\pi_{BB}(\cdot \mid Y)$ using $B \approx 50$ bootstrap datasets

[H & Miller 2019]

BayesBag in practice

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

2) compute bagged posterior $\pi_{BB}(\cdot \mid Y)$ using $B \approx 50$ bootstrap datasets
   - e.g., use MCMC to get approximate samples $\theta^*_{(b,1)}, \ldots, \theta^*_{(b,T)}$ from $\pi(\cdot \mid Y^*_{(b)})$ 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 $\pi_{BB}(\cdot \,|\, Y)$ 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 $\pi(\cdot \mid Y)$

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

2) compute bagged posterior $\pi_{BB}(\cdot \mid Y)$ using $B \approx 50$ bootstrap datasets

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

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

3a) compute mismatch index $\mathbf{I}$

[**H** & Miller 2019]

BayesBag in practice

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

2) compute bagged posterior $\pi_{BB}(\cdot \mid Y)$ using $B \approx 50$ bootstrap datasets
   - e.g., use MCMC to get approximate samples $\theta^*_{(b,1)}, \dots, \theta^*_{(b,T)}$ from $\pi(\cdot \mid Y^*_{(b)})$ for $b = 1, \dots, B$

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

3a) compute mismatch index $\mathbf{I}$

3b) if $\mathbf{I} \gtrapprox .2$, consider refining the model and returning to step 1

[**H** & Miller 2019]

BayesBag in practice

1) compute standard posterior $\pi(\cdot \mid Y)$

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

2) compute bagged posterior $\pi_{BB}(\cdot \mid Y)$ using $B \approx 50$ bootstrap datasets

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

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

3a) compute mismatch index $\mathbf{I}$

3b) if $\mathbf{I} \gtrapprox .2$, consider refining the model and returning to step 1

4) output bagged posterior computed in step 2

[**H** & Miller 2019]

Agenda

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

➡ BayesBag for model selection

Bayesian model selection

Bayesian model selection

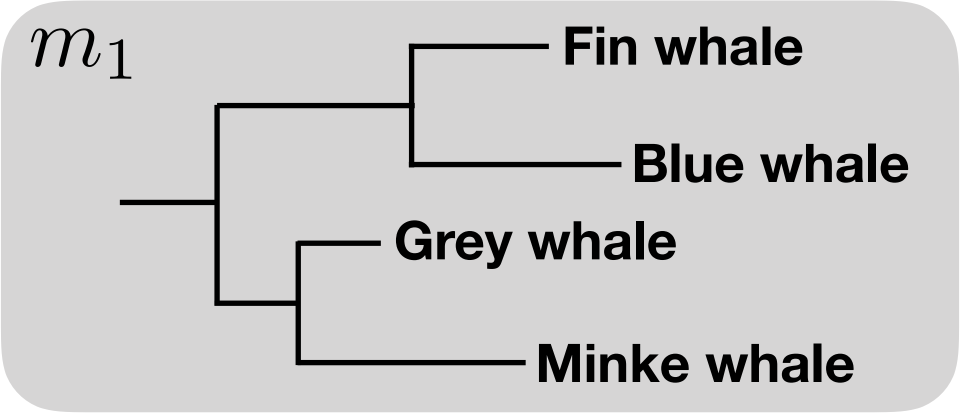
- **Goal:** based on data Y, select between a (finite or countable) set of models $M = \{m_1, m_2, \dots\}$

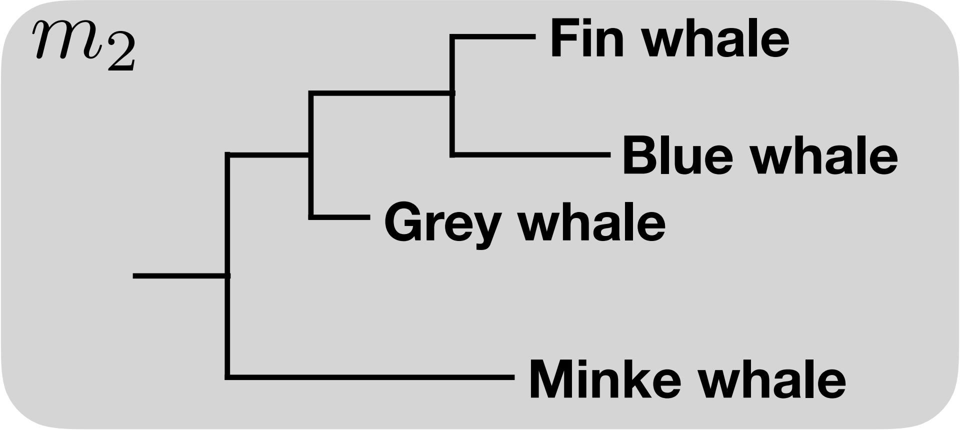
Bayesian model selection

- **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 $M = \{m_1, m_2, \ldots\}$
- **Example:** systematics
  - **Goal:** learn about evolutionary history of a set of species **[e.g. whales]**



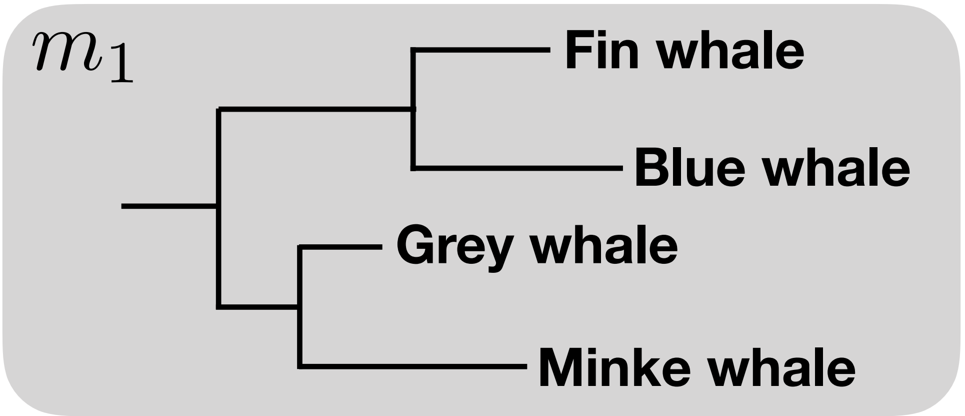


⋮

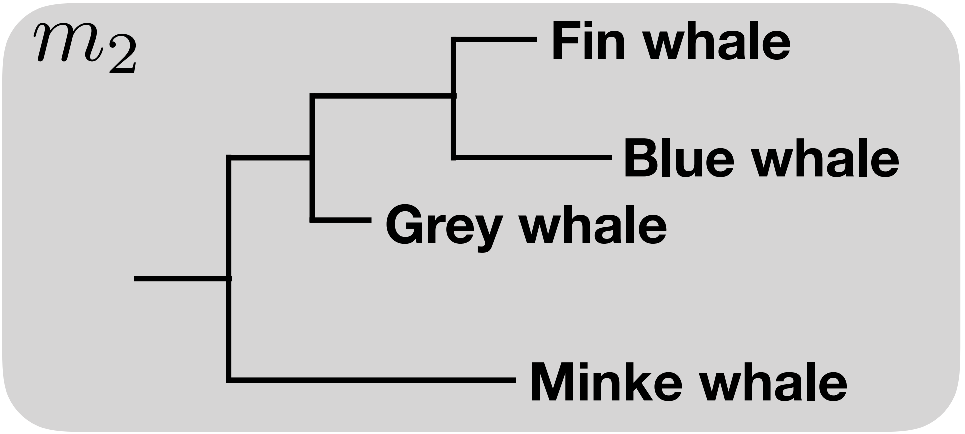
Bayesian model selection

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

$$\pi(m_1 \,|\, Y) = .8$$



$$\pi(m_2 \,|\, Y) = .1$$

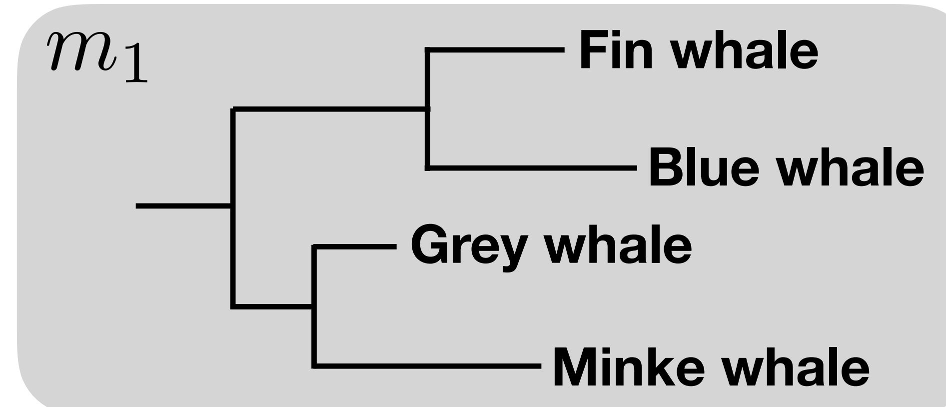


⋮

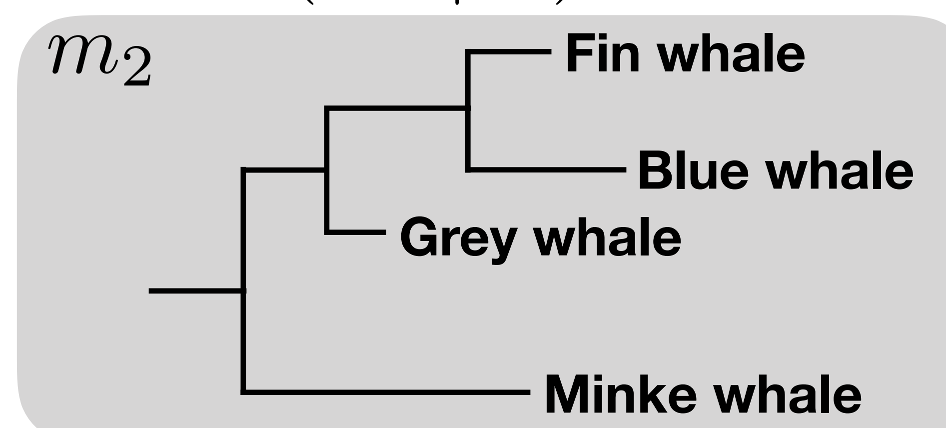
Bayesian model selection

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

$$\pi(m_1 \,|\, Y) = .8$$



$$\pi(m_2 \,|\, Y) = .1$$



$\vdots$

Illustration: two normal models

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

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 $P_{\text{true}} = \mathcal{N}(0, 1)$

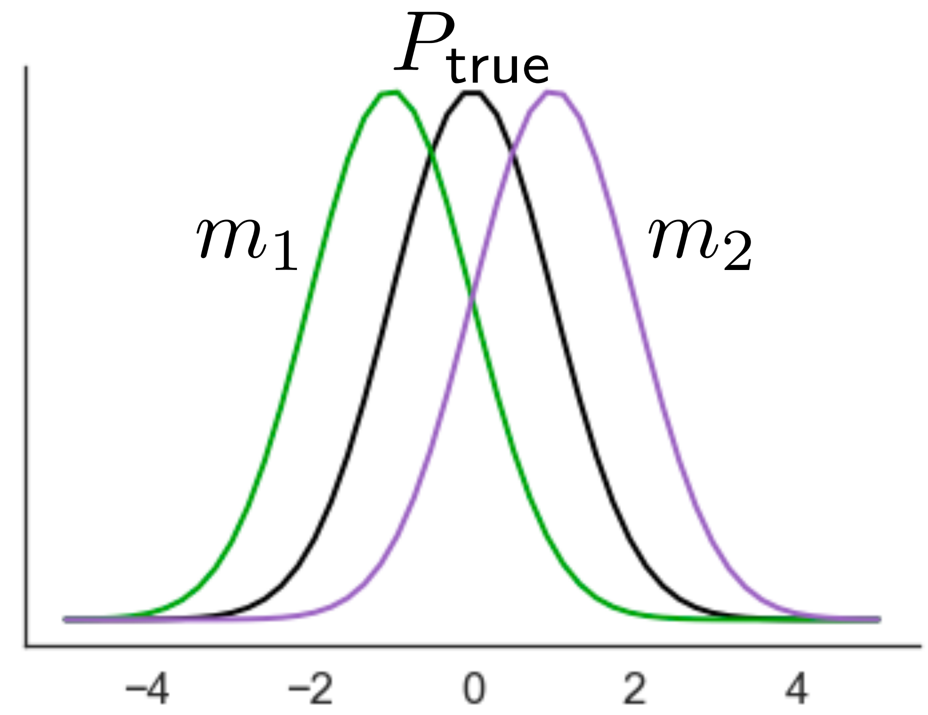
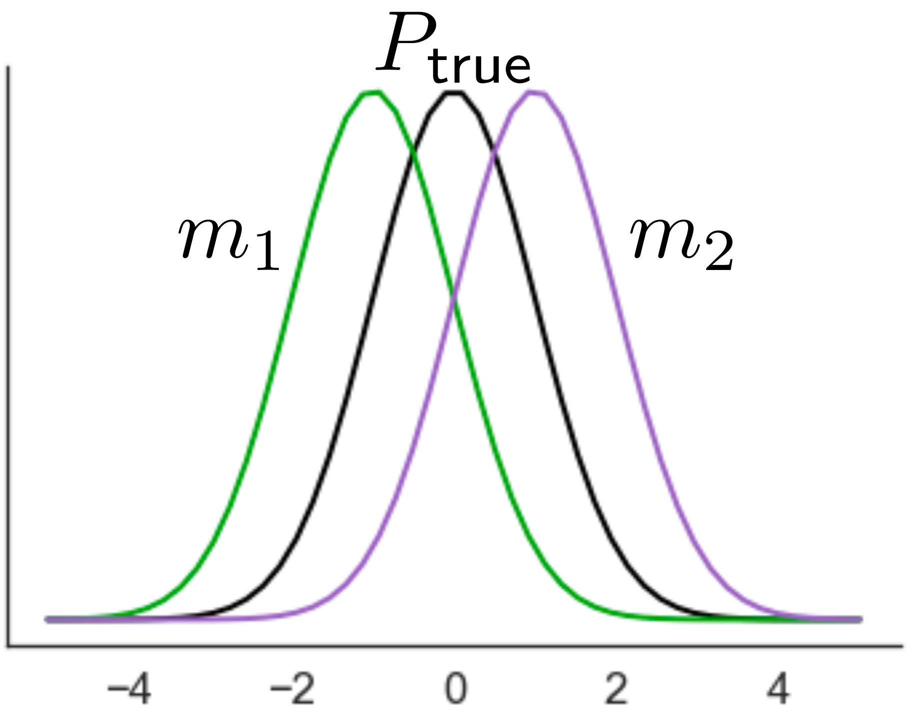


Illustration: two normal models

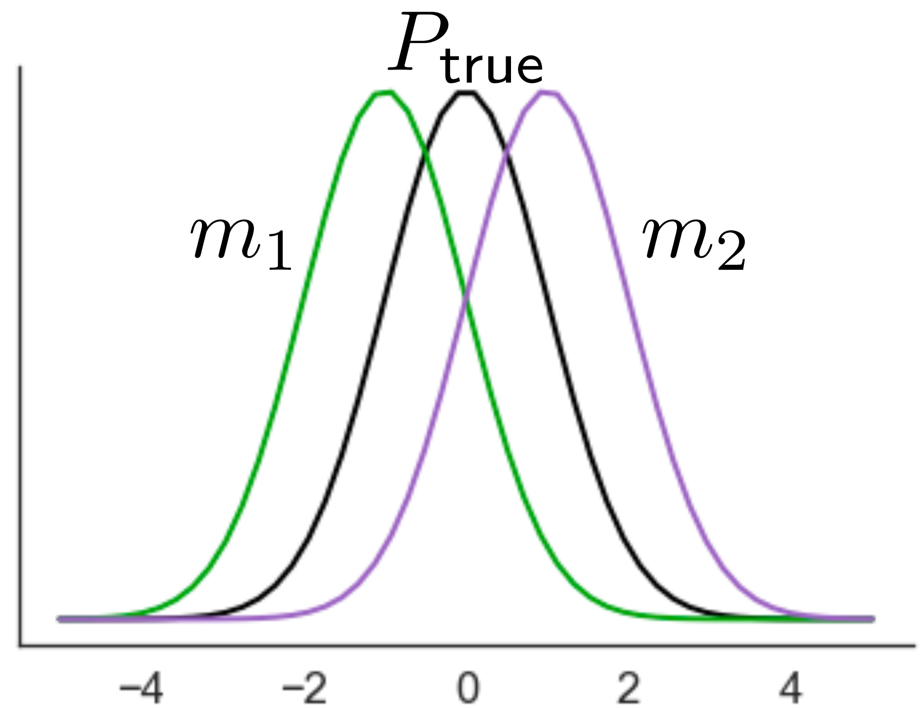
- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$
- Generate datasets $Y^{(1)}, Y^{(2)}, \ldots$ of size $n = 1000$, where $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



[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 $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



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

$$\pi_{BB}(m_1 \,|\, Y^{(1)}) = 0.82$$

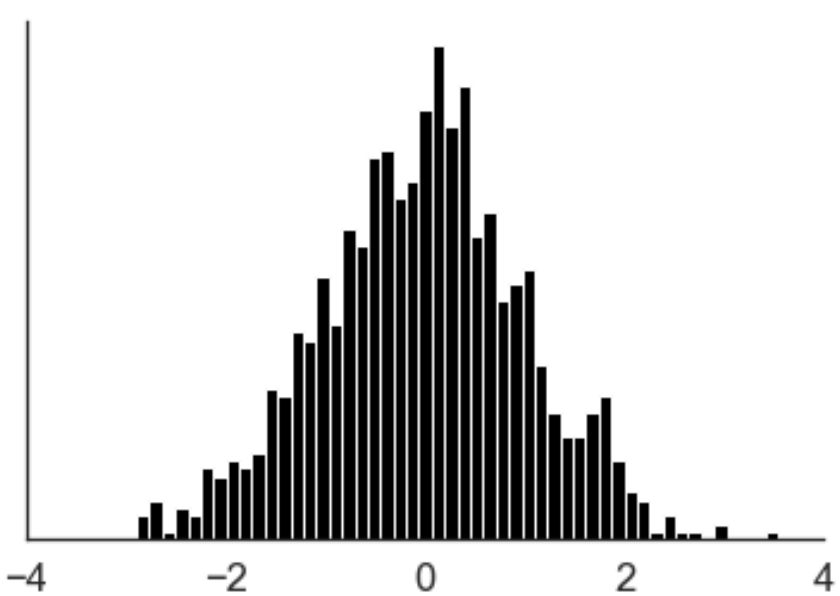
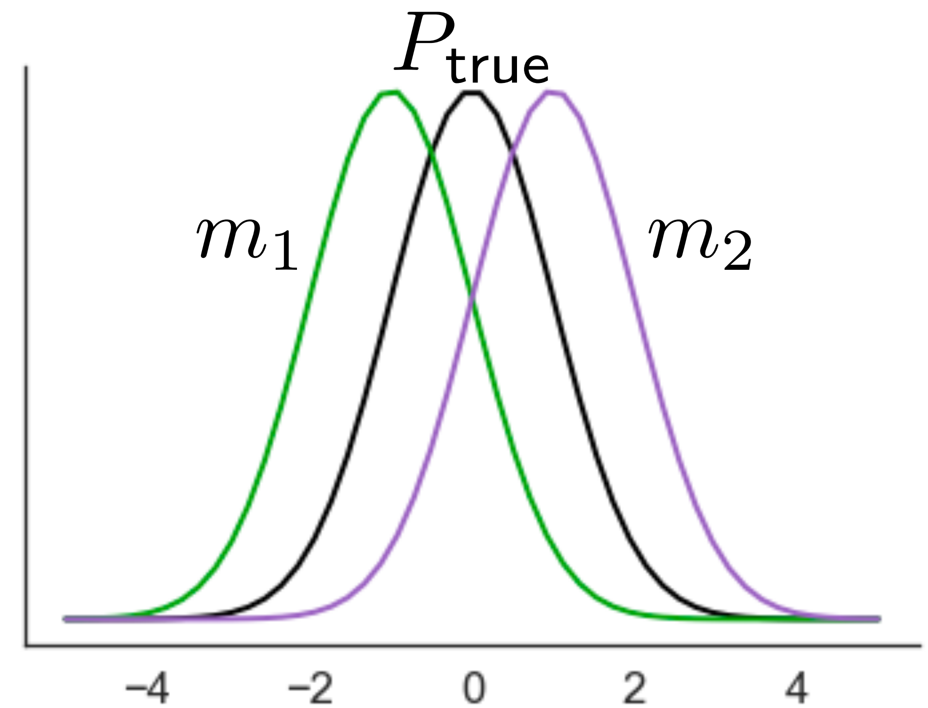
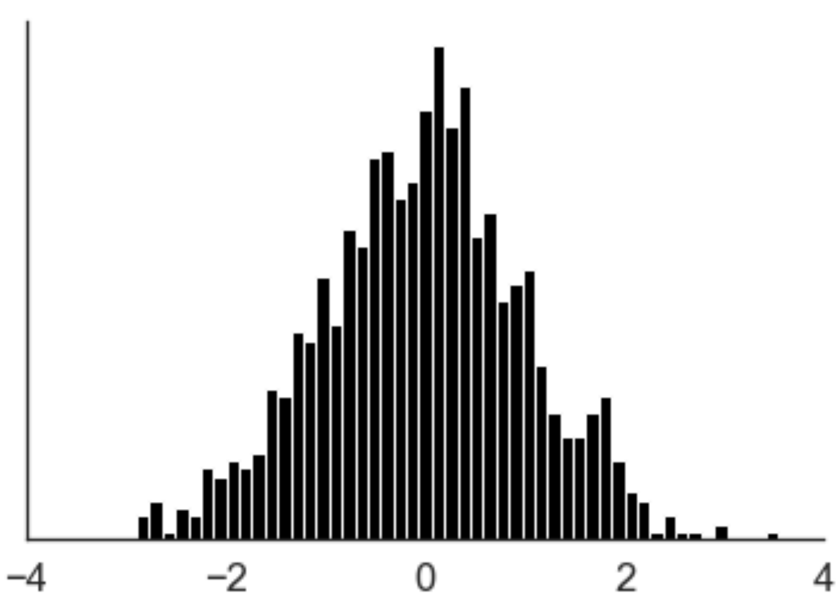


Illustration: two normal models

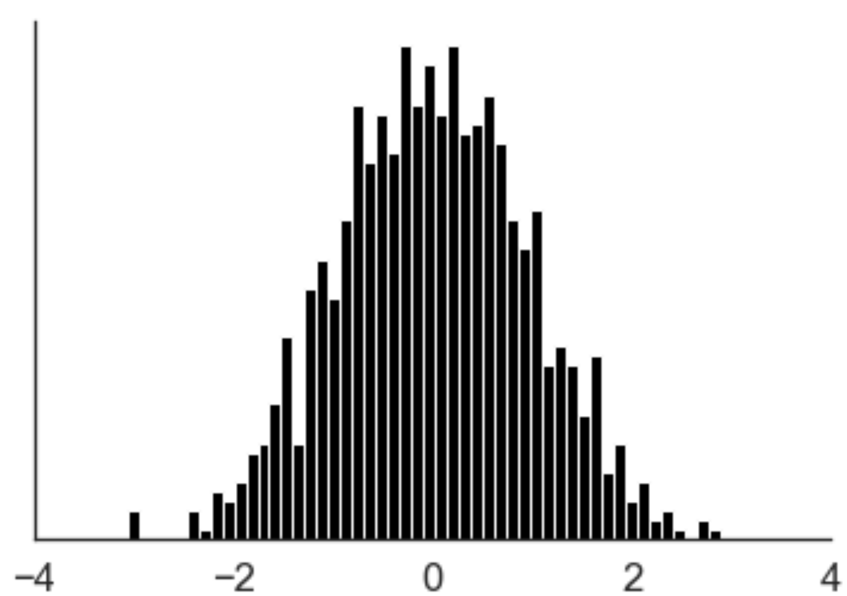
- Models are $m_1 = \mathcal{N}(-1, 1)$ and $m_2 = \mathcal{N}(1, 1)$
- True distribution is $P_{\text{true}} = \mathcal{N}(0, 1)$
- Generate datasets $Y^{(1)}, Y^{(2)}, \ldots$ of size $n = 1000$, where $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



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



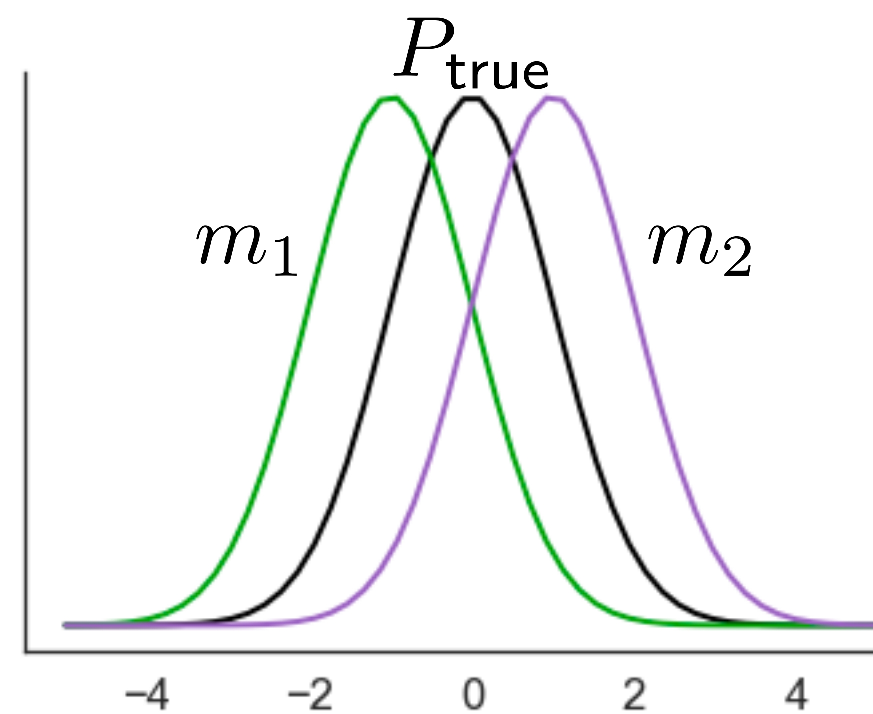
$$\pi(m_1 \mid Y^{(2)}) = 10^{-5}$$
$$\pi_{BB}(m_1 \mid Y^{(2)}) = 0.38$$



[H & Miller 2019]

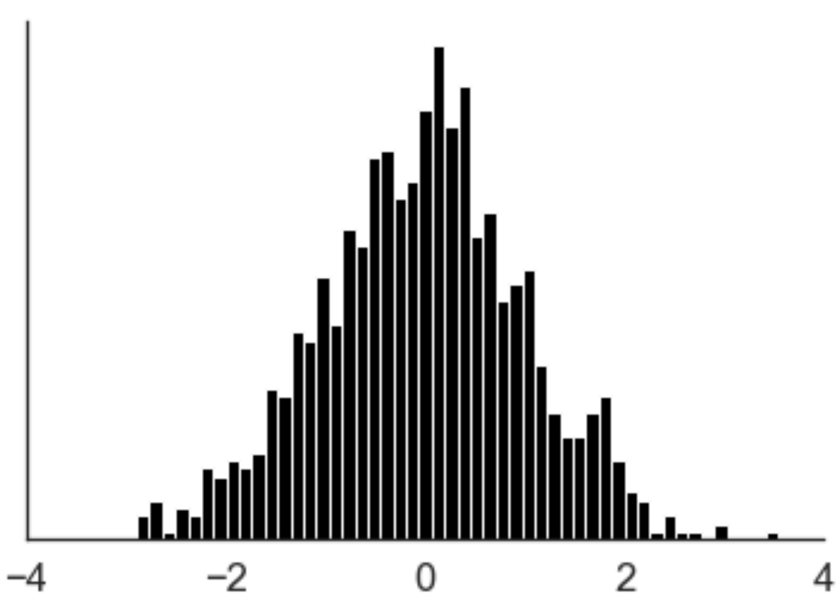
Illustration: two normal models

- 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 $Y_j^{(i)} \sim \mathcal{N}(0, 1)$.



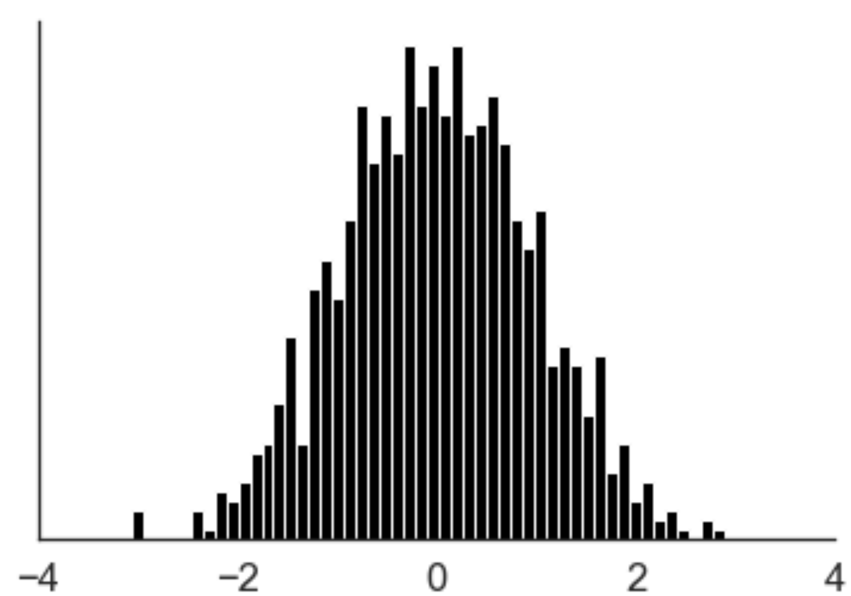
$\pi(m_1 \mid Y^{(1)}) = 1$

$\pi_{BB}(m_1 \mid Y^{(1)}) = 0.82$



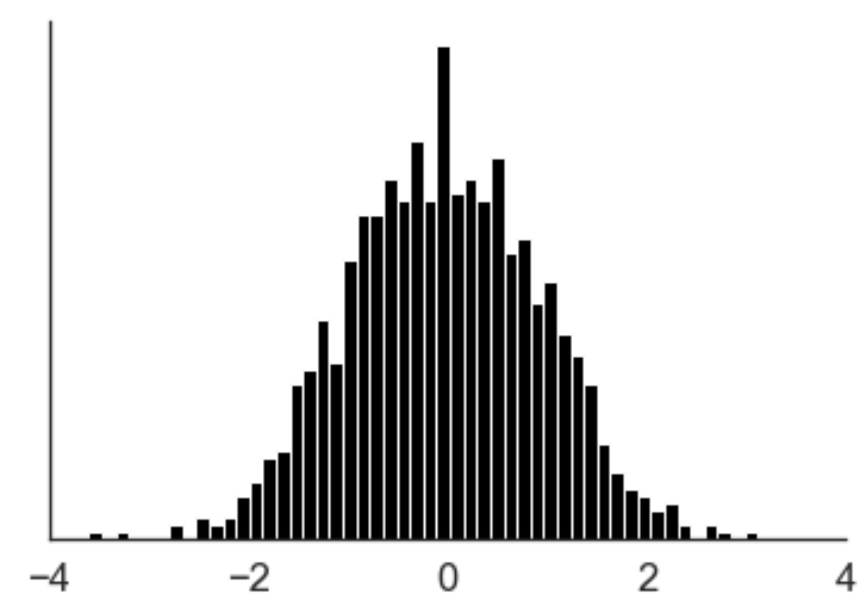
$\pi(m_1 \mid Y^{(2)}) = 10^{-5}$

$\pi_{BB}(m_1 \mid Y^{(2)}) = 0.38$



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

$\pi_{BB}(m_1 \mid Y^{(3)}) = 0.90$



[H & Miller 2019]

BayesBag stabilizes posterior probabilities of similar models

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

$\pi(m_1 \mid Y) = 0$ or 1 with equal probability

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

$\pi(m_1 \mid Y) = 0$ or 1 with equal probability

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

	$\pi(m_1 \mid \cdot)$
$Y^{(1)}$	0
$Y^{(2)}$	1
$Y^{(3)}$	0
$Y^{(4)}$	0
$Y^{(5)}$	1
$Y^{(6)}$	0
⋮	⋮

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

$\pi(m_1 \mid Y) = 0$ or 1 with equal probability

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

	$\pi(m_1 \mid \cdot)$
$Y^{(1)}$	0
$Y^{(2)}$	1
$Y^{(3)}$	0
$Y^{(4)}$	0
$Y^{(5)}$	1
$Y^{(6)}$	0
⋮	⋮

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

$\pi(m_1 \mid Y) = 0$ or 1 with equal probability

All posterior mass on a single, arbitrary model

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

	$\pi(m_1 \mid \cdot)$
$Y^{(1)}$	0
$Y^{(2)}$	1
$Y^{(3)}$	0
$Y^{(4)}$	0
$Y^{(5)}$	1
$Y^{(6)}$	0
⋮	⋮

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

   $\pi(m_1 \mid Y) = 0$ or 1 with equal probability

2. For the bagged posterior,

   $$\pi_{BB}(m_1 \mid Y) \sim \text{Uniform}[0,1]$$

All posterior mass on a single, arbitrary model

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

	$\pi(m_1 \mid \cdot)$	$\pi_{BB}(m_1 \mid \cdot)$
$Y^{(1)}$	0	0.03
$Y^{(2)}$	1	0.78
$Y^{(3)}$	0	0.75
$Y^{(4)}$	0	0.98
$Y^{(5)}$	1	0.95
$Y^{(6)}$	0	0.23
⋮	⋮	⋮

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

   $\pi(m_1 \mid Y) = 0$ or 1 with equal probability

2. For the bagged posterior,

   $$\pi_{BB}(m_1 \mid Y) \sim \text{Uniform}[0,1]$$

All posterior mass on a single, arbitrary model

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

BayesBag stabilizes posterior probabilities of similar models

- Assume two models m_1 and m_2 **[e.g. two possible trees]**
- If models explain the data-generating distribution equally well, **hope** equal posterior probability (with enough data): $\pi(m_1 \mid Y) = \pi(m_2 \mid Y) = 1/2$

	$\pi(m_1 \mid \cdot)$	$\pi_{BB}(m_1 \mid \cdot)$
$Y^{(1)}$	0	0.03
$Y^{(2)}$	1	0.78
$Y^{(3)}$	0	0.75
$Y^{(4)}$	0	0.98
$Y^{(5)}$	1	0.95
$Y^{(6)}$	0	0.23
⋮	⋮	⋮

However….

Theorem [**H** & Miller 2019]

Assume m_1 and m_2 are equally good.
Then in the large data limit,

1. For the standard posterior,

$\pi(m_1 \mid Y) = 0$ or 1 with equal probability

2. For the bagged posterior,

$\pi_{BB}(m_1 \mid Y) \sim \mathrm{Uniform}[0,1]$

All posterior mass on a single, arbitrary model

bagged posterior mass more evenly distributed

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

Reproducible phylogenetic inference

[**H** & Miller 2019]

Reproducible phylogenetic inference

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

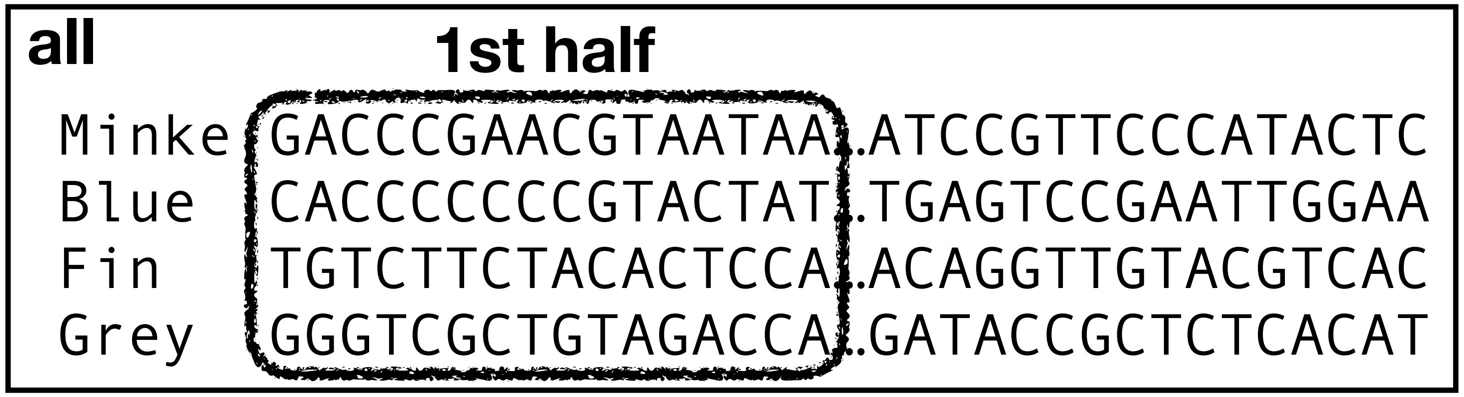
all

```
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



[**H** & Miller 2019]

Reproducible phylogenetic inference

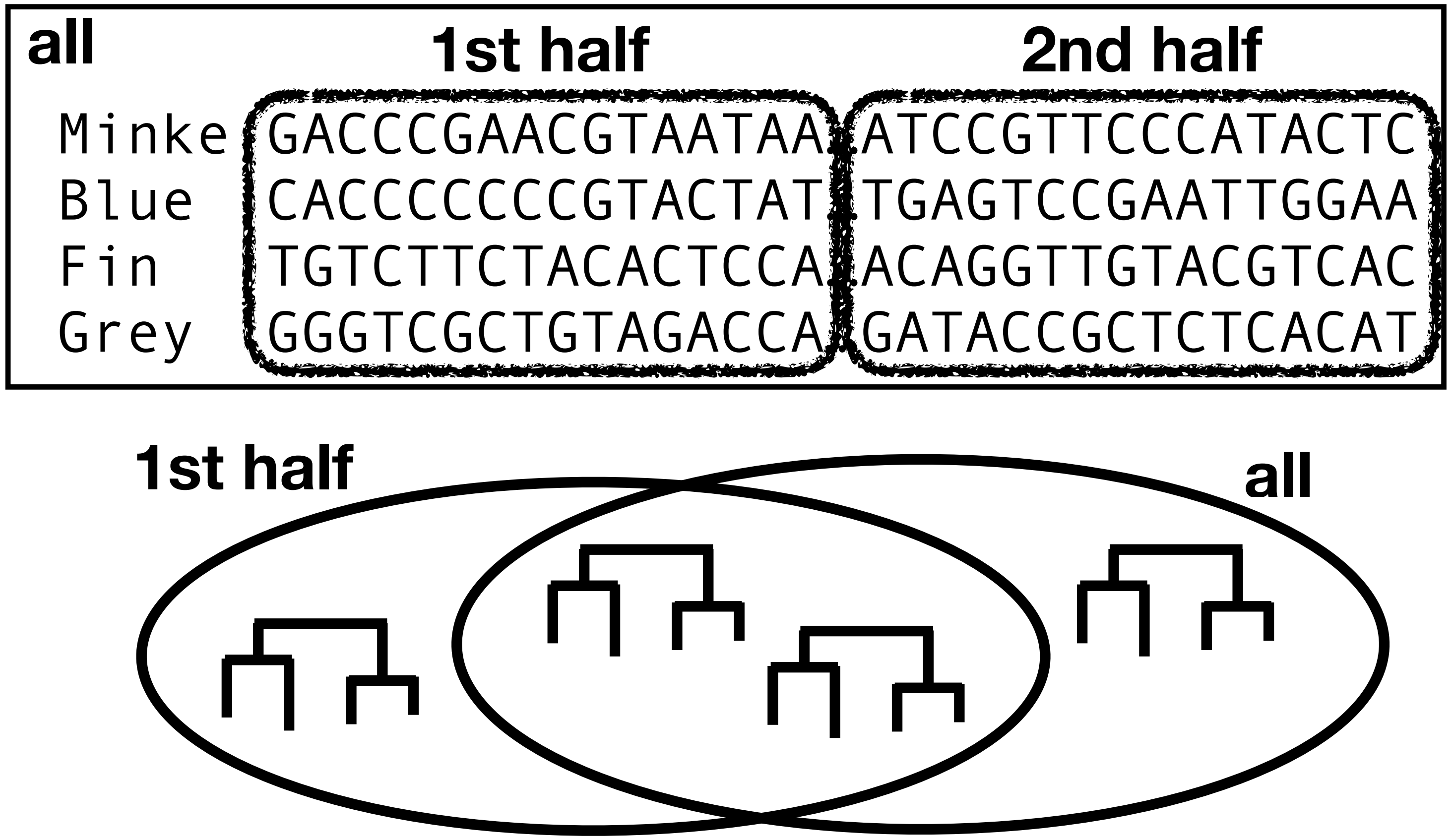
- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA

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

[H & Miller 2019]

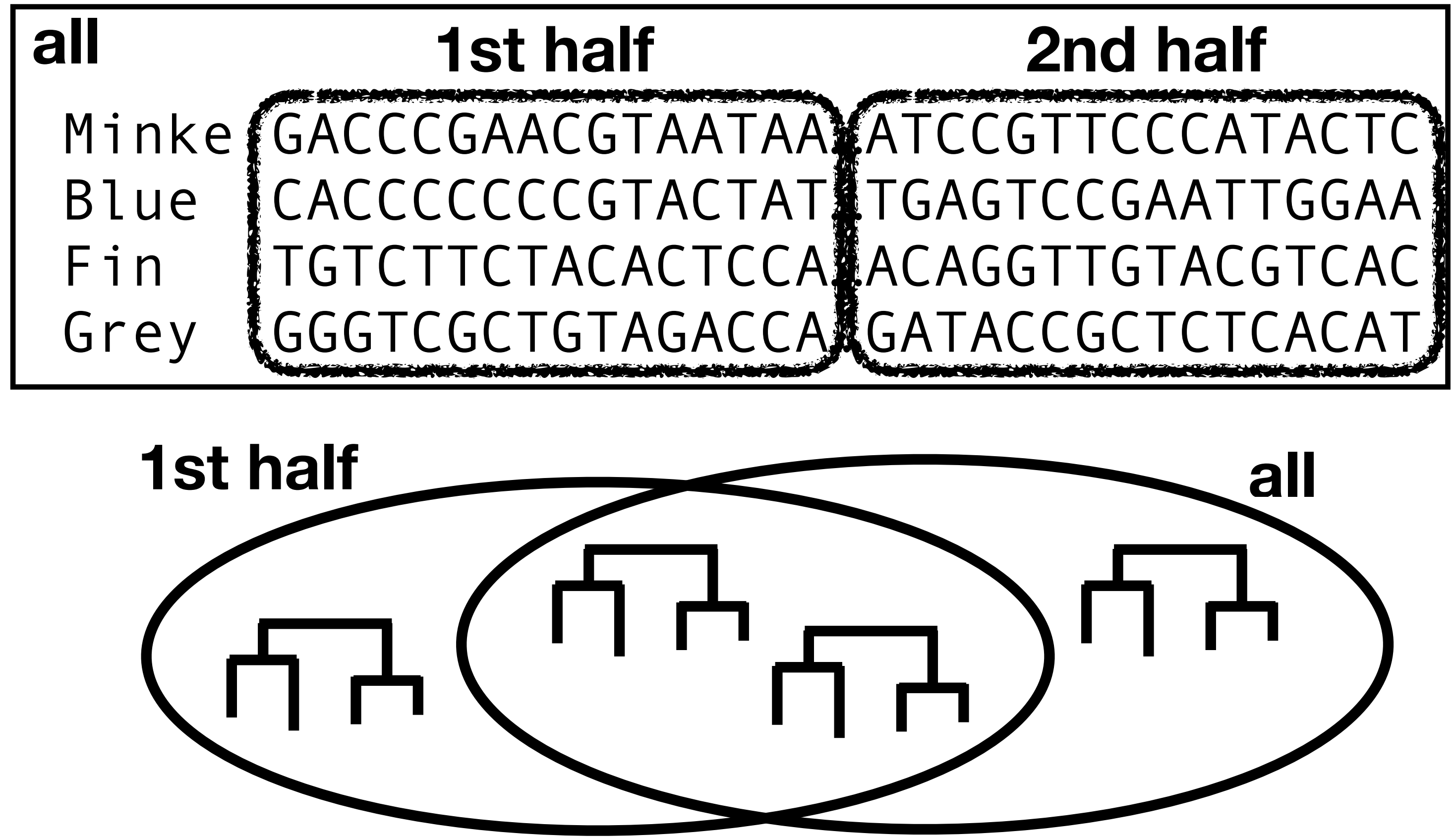
Reproducible phylogenetic inference

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



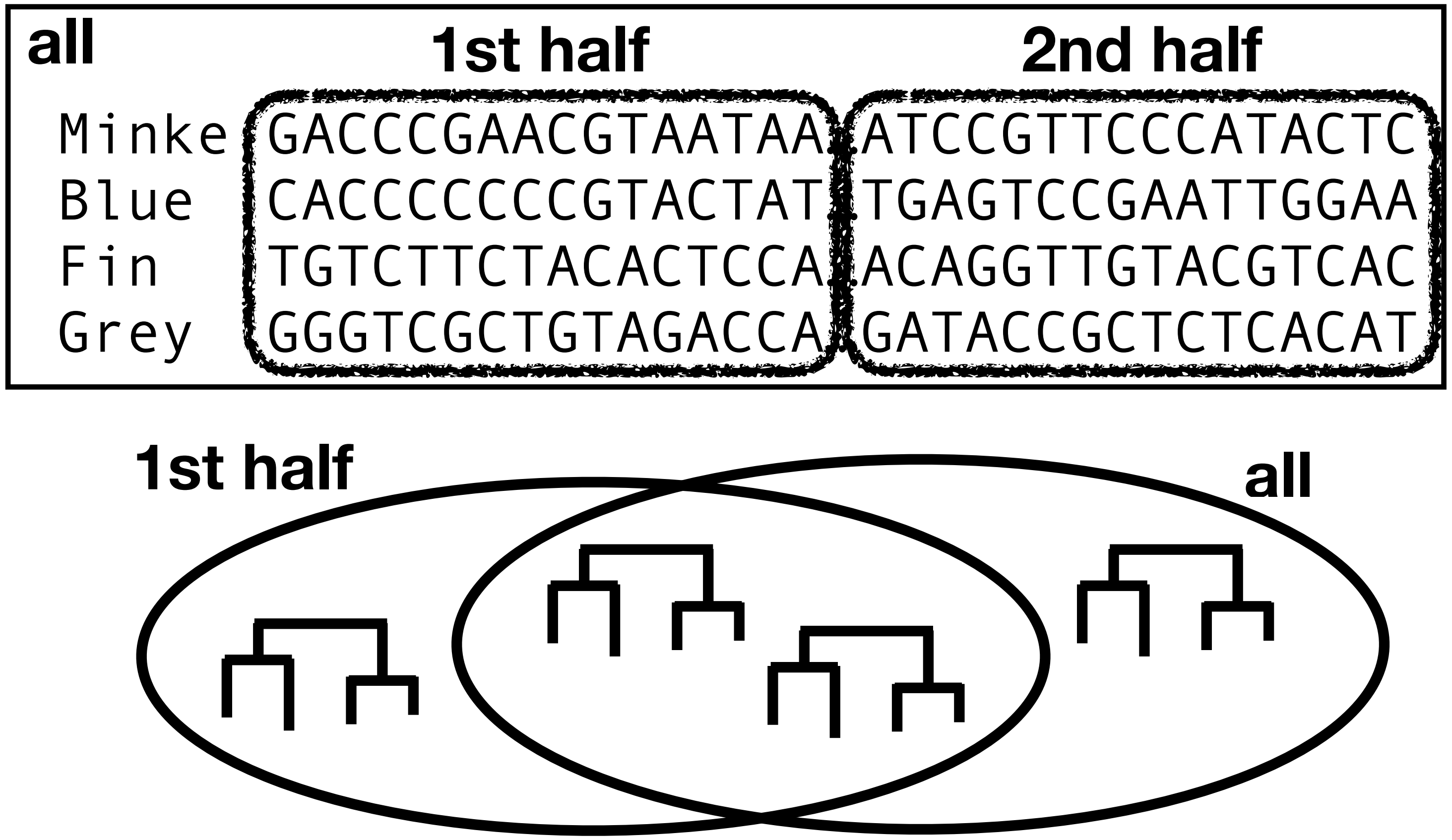
Reproducible phylogenetic inference

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



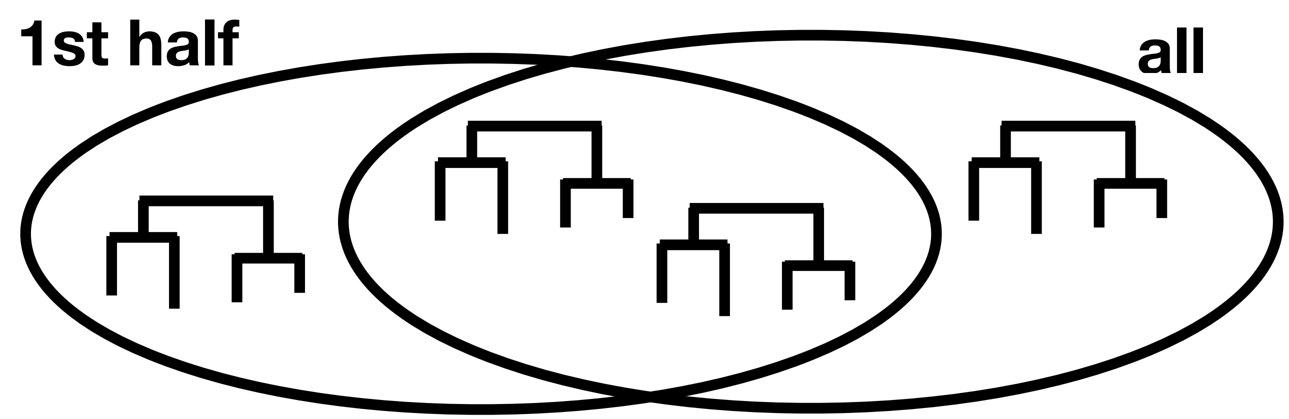
Reproducible phylogenetic inference

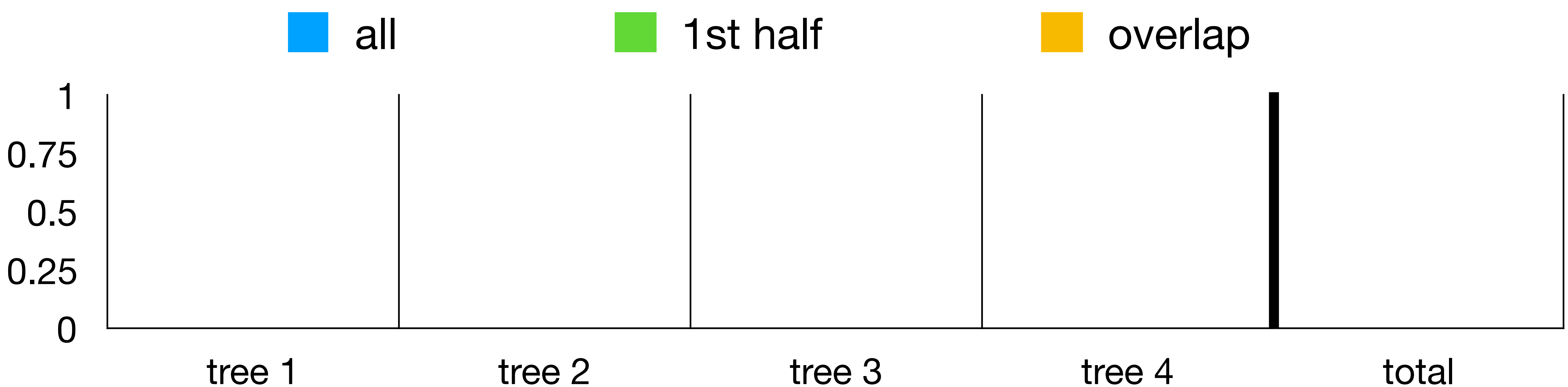
- **Goal:** infer phylogeny of 13 whale species from mitochondrial coding DNA
- Compute posterior tree probabilities based on **all**, **1st half**, and **2nd half**
- Compute overlap of 99% high probability regions



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

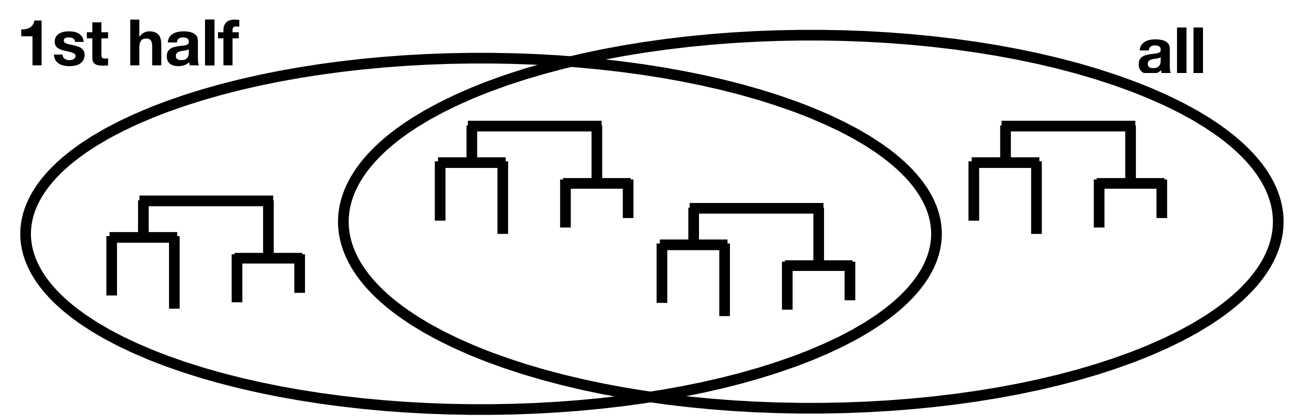


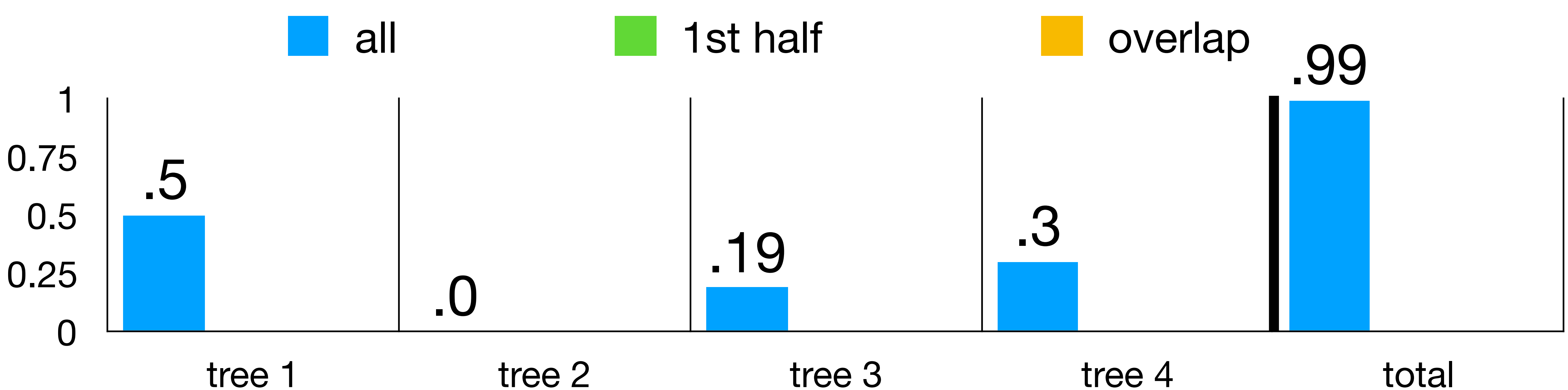


Reproducible phylogenetic inference

- **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

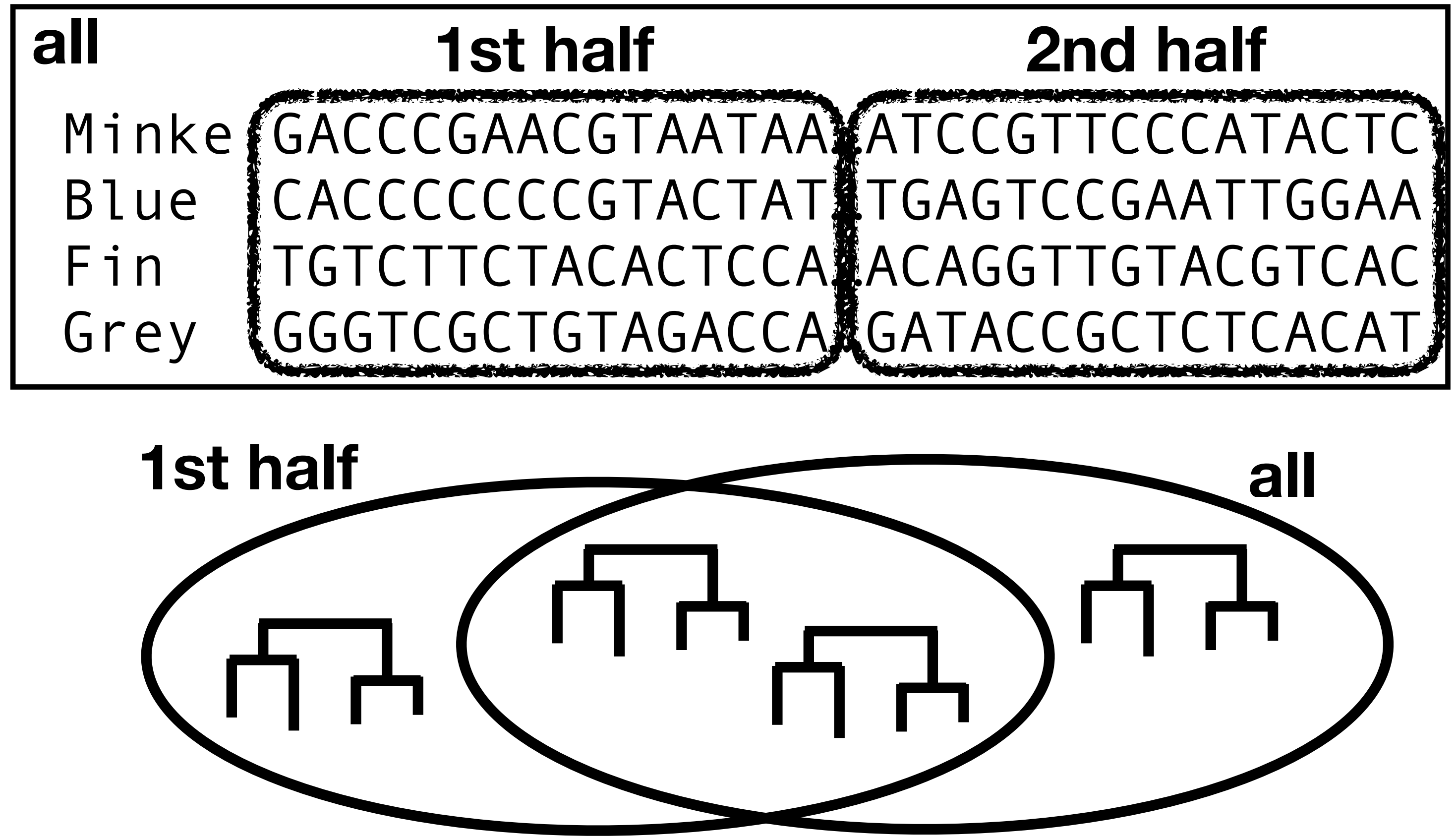


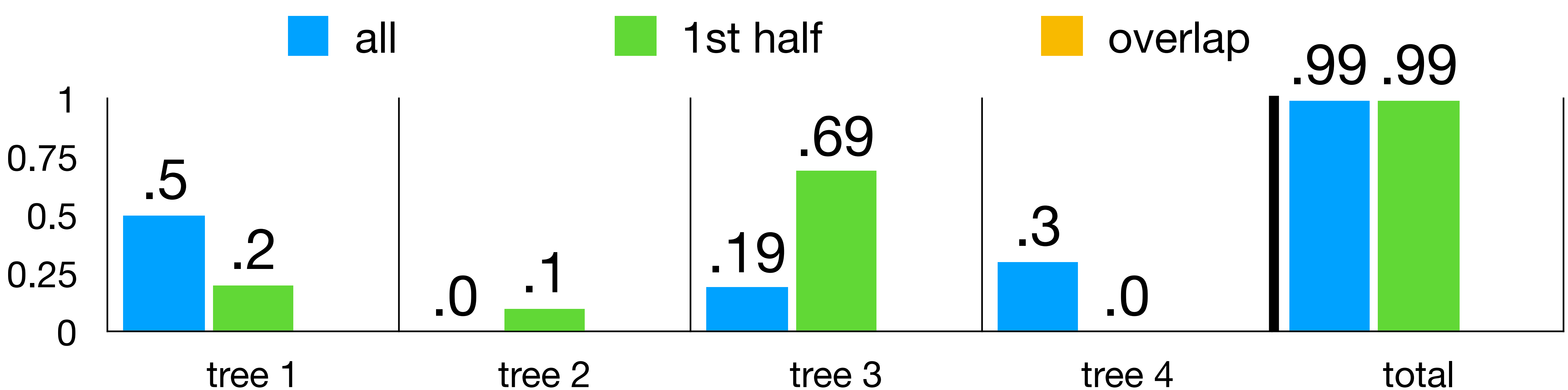


[H & Miller 2019]

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



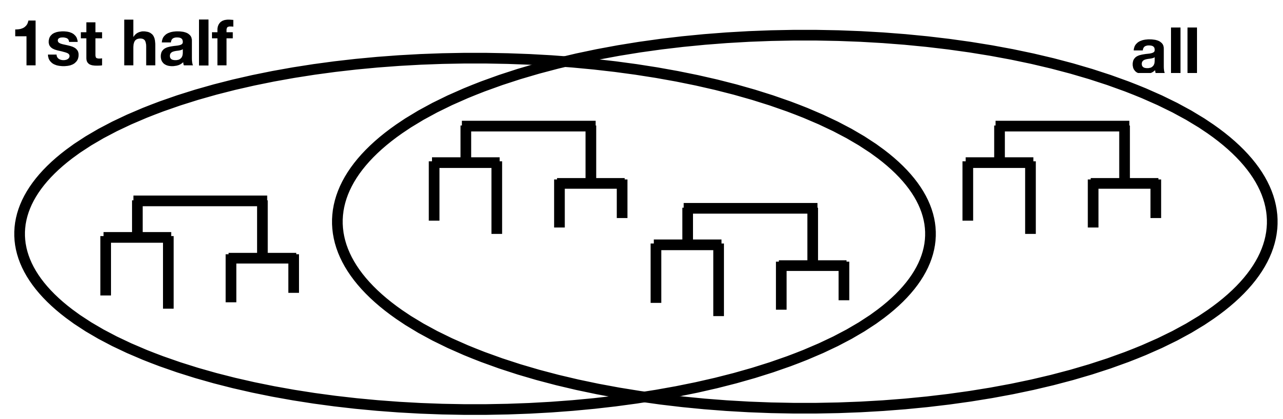


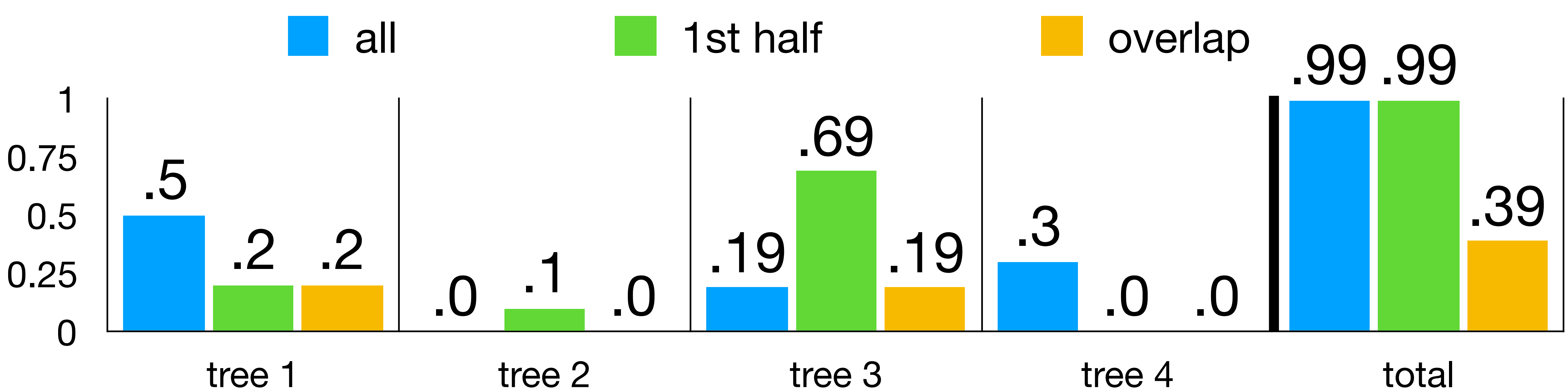
[**H** & Miller 2019]

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

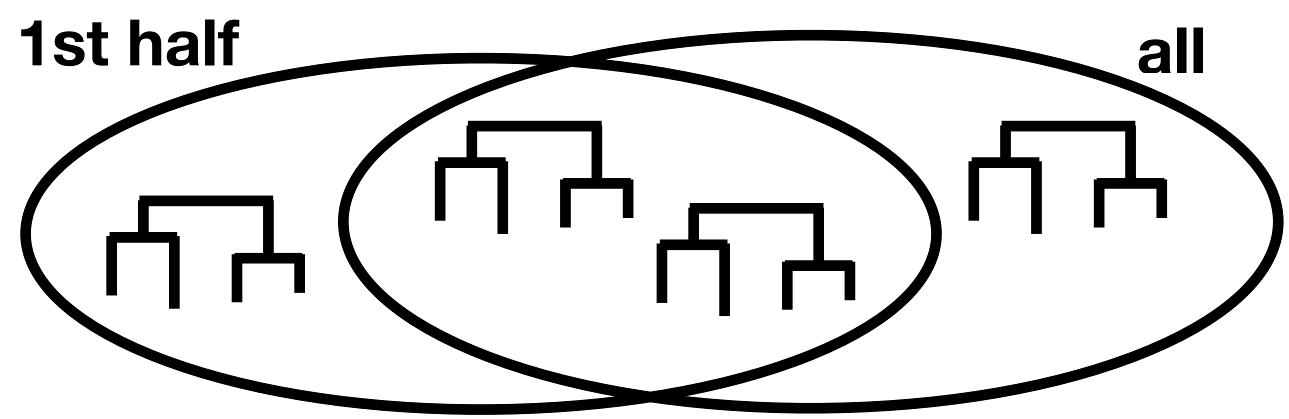


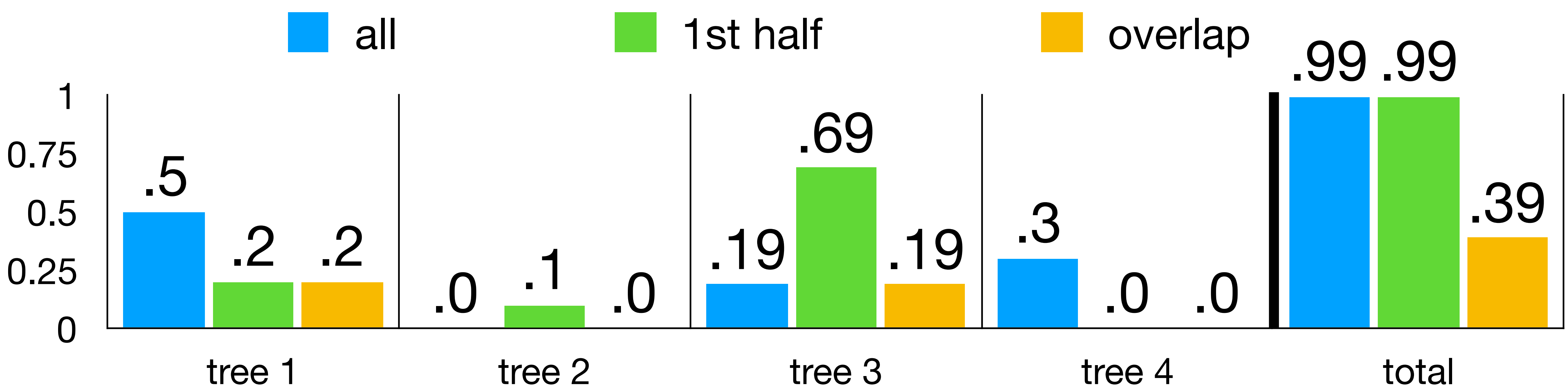


[H & Miller 2019]

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





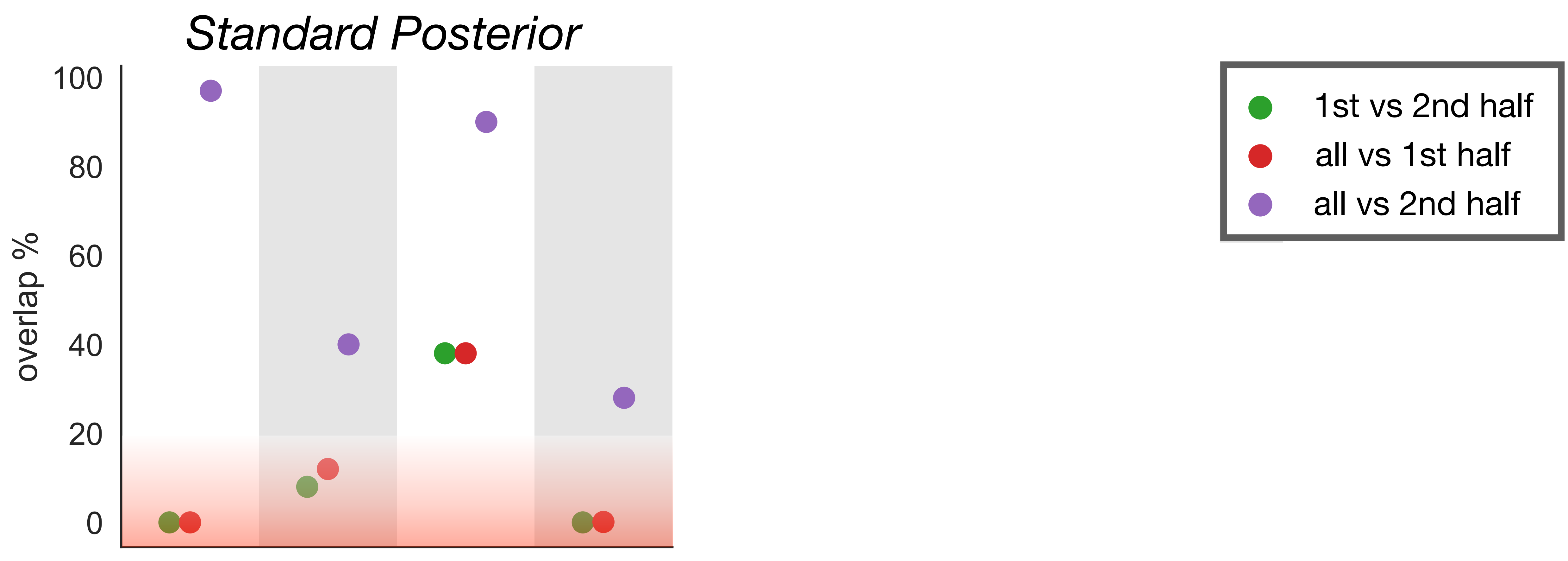
Stable, reproducible phylogenetic inference with BayesBag

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

[**H** & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

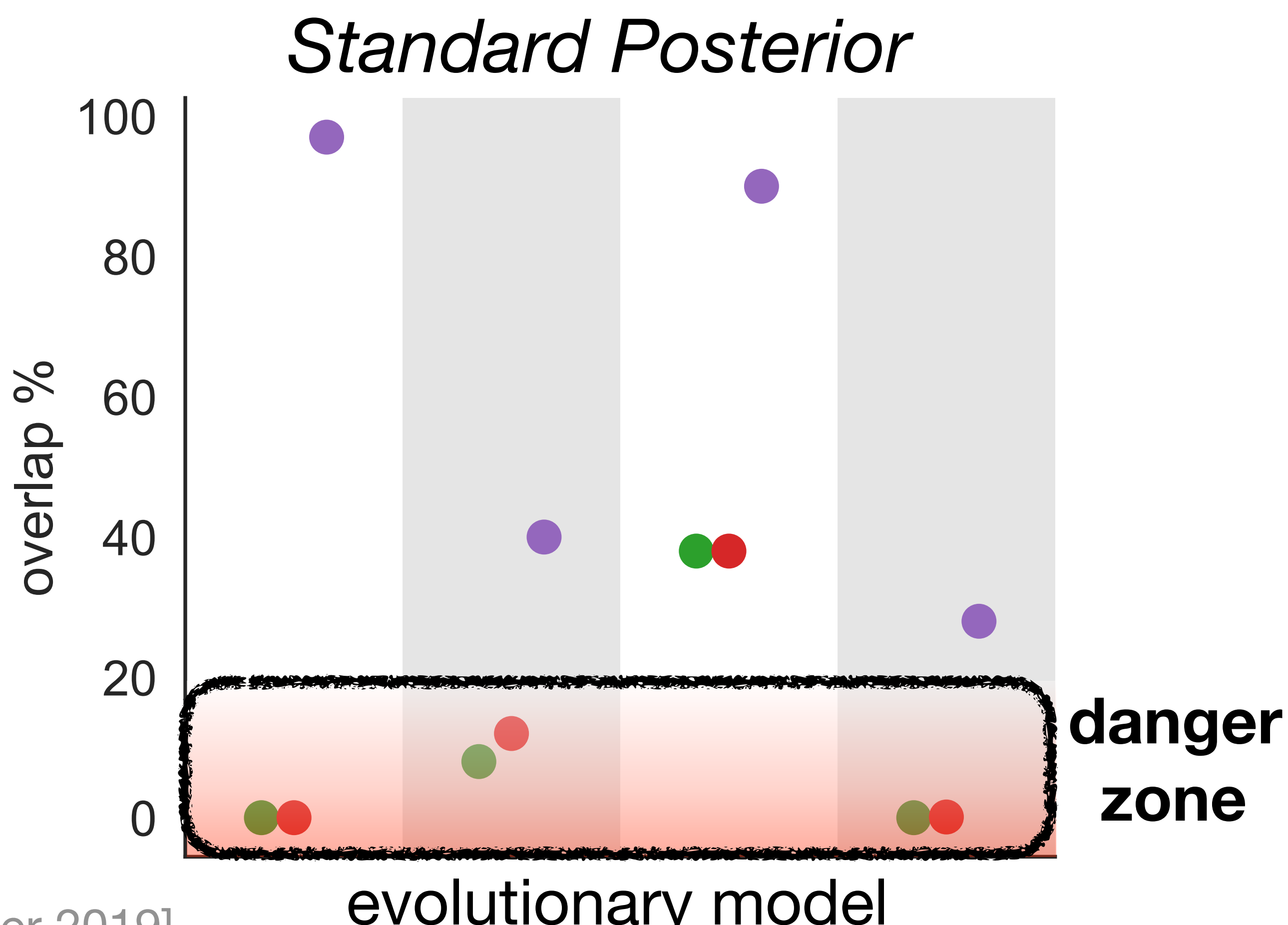
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap

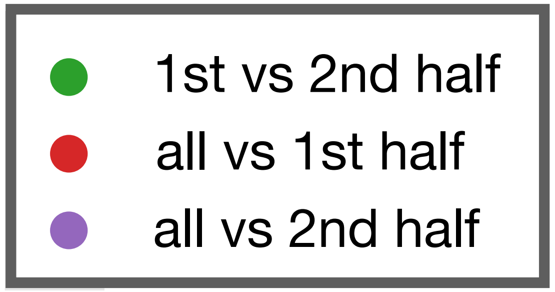


[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

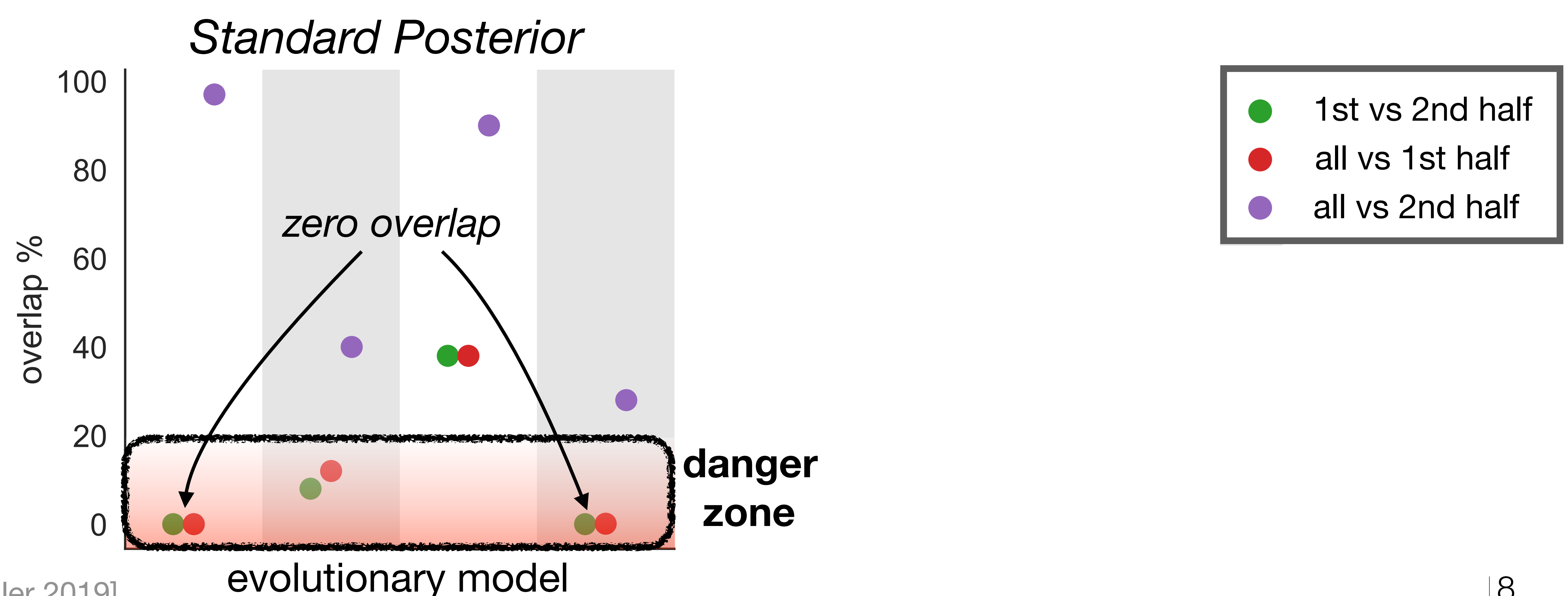
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap





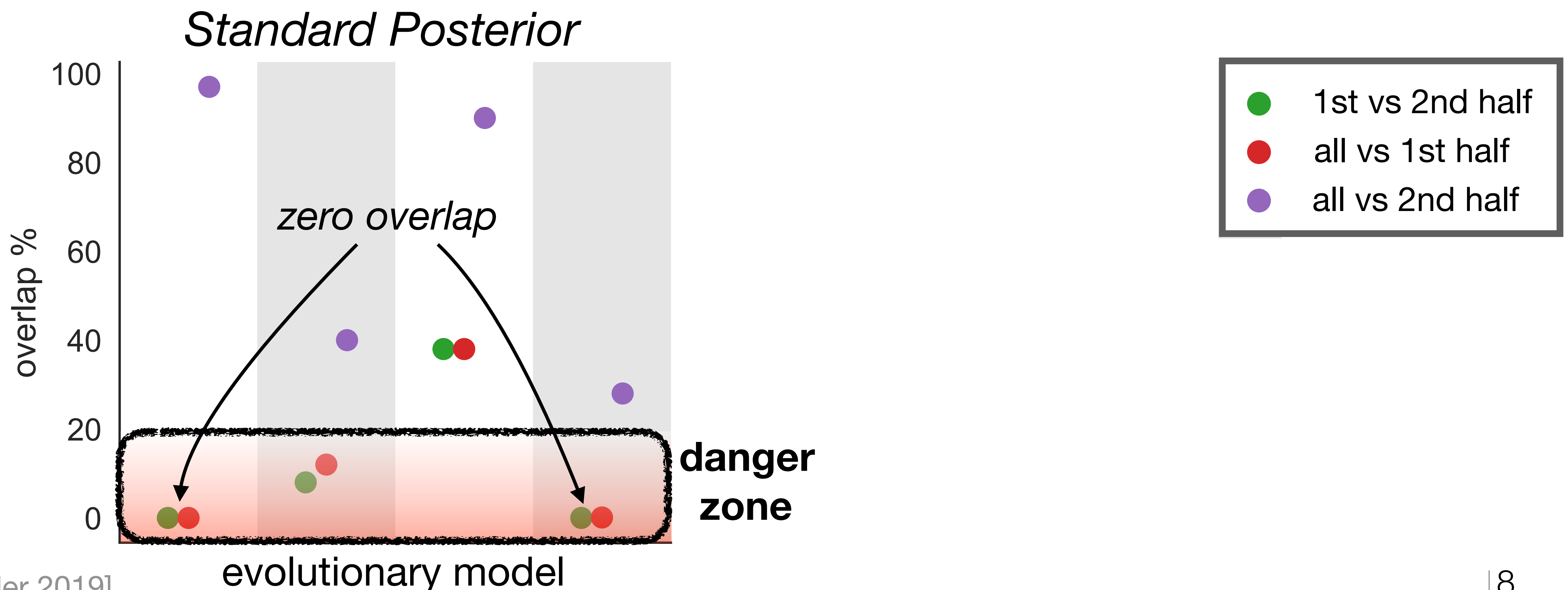
Stable, reproducible phylogenetic inference with BayesBag

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



Stable, reproducible phylogenetic inference with BayesBag

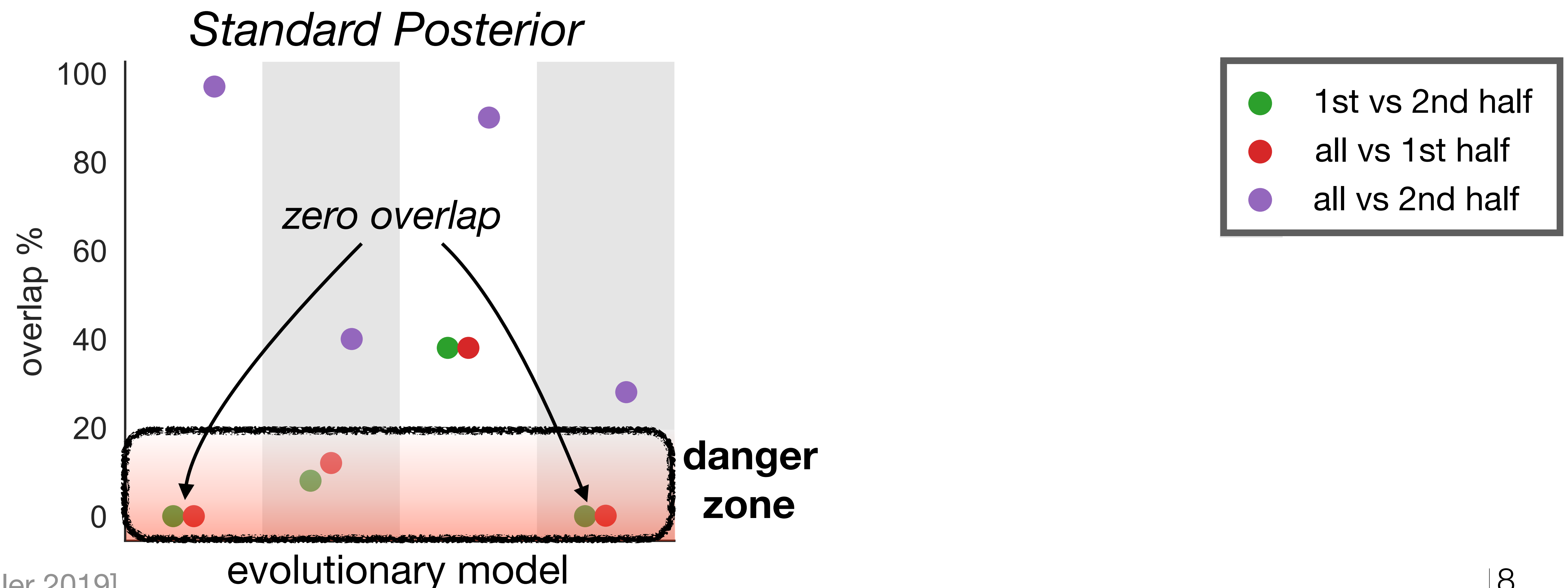
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]



[**H** & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

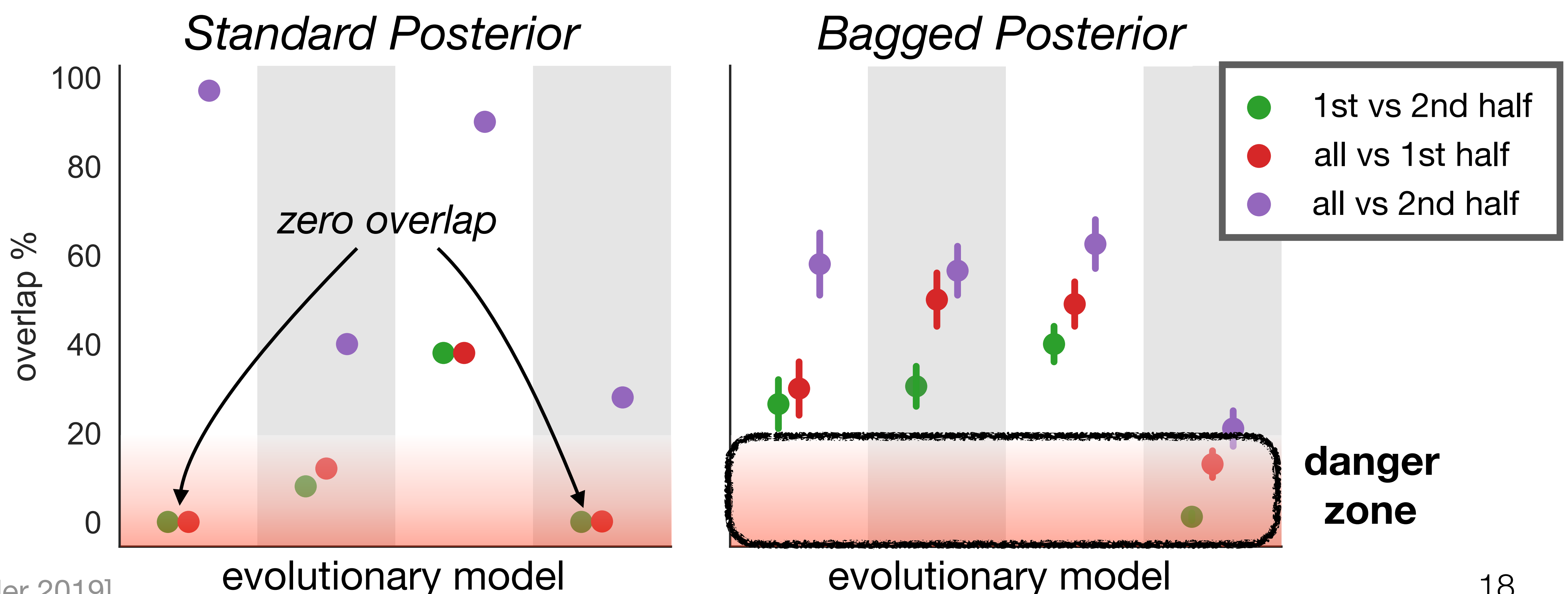
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]
- Same problem comparing evolutionary models with data fixed



[H & Miller 2019]

Stable, reproducible phylogenetic inference with BayesBag

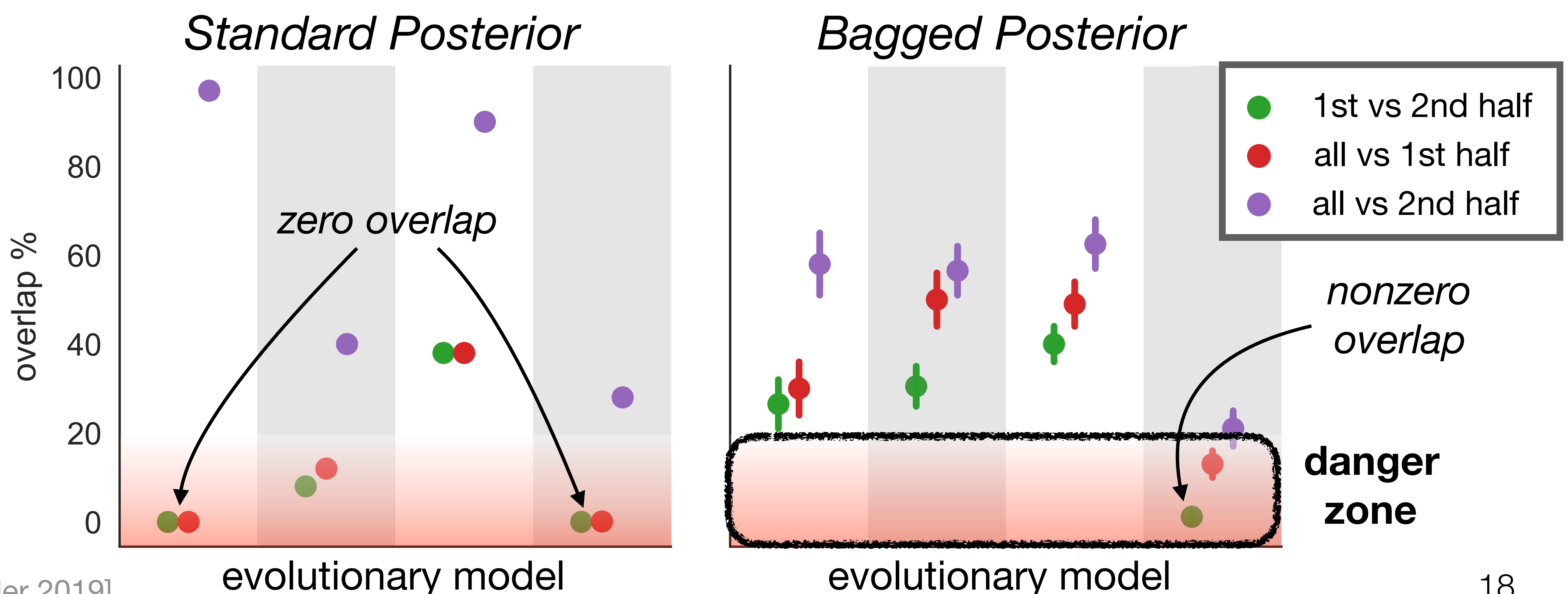
- **Goal:** infer phylogeny of 13 whale species
- For some evolutionary models, little to no overlap
- Bayesian model selection is **unstable** and **not reproducible** [Wilcox et al. 2002, Alfaro et al. 2003, Douady et al. 2003, …]
- Same problem comparing evolutionary models with data fixed
- Bagged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

Closing thoughts

- We show that BayesBag…
  1. has **provably good** statistical robustness properties
  2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)

Closing thoughts

- We show that BayesBag…
  1. has **provably good** statistical robustness properties
  2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
  3. is **easy to use** and **widely applicable**

Closing thoughts

- We show that BayesBag…
  1. has **provably good** statistical robustness properties
  2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
  3. is **easy to use** and **widely applicable**
  4. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods

Closing thoughts

- We show that BayesBag…
    1. has **provably good** statistical robustness properties
    2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
    3. is **easy to use** and **widely applicable**
    4. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work:

Closing thoughts

- We show that BayesBag…
    1. has **provably good** statistical robustness properties
    2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
    3. is **easy to use** and **widely applicable**
    4. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work:
    ➡ time series / other structured data

Closing thoughts

- We show that BayesBag…
  1. has **provably good** statistical robustness properties
  2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
  3. is **easy to use** and **widely applicable**
  4. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work:
  - ➡ time series / other structured data
  - ➡ speeding up computation

Closing thoughts

- We show that BayesBag…
  1. has **provably good** statistical robustness properties
  2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
  3. is **easy to use** and **widely applicable**
  4. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work:
  - ➡ time series / other structured data
  - ➡ speeding up computation
- On arXiv very soon (if you want a heads up: jhuggins@hsph.harvard.edu)

Closing thoughts

- We show that BayesBag…
  1. has **provably good** statistical robustness properties
  2. empirically, demonstrates **superior predictive performance** (compared to standard Bayes)
  3. is **easy to use** and **widely applicable**
  4. combines the **flexible modeling** features of Bayes with the **distributional robustness** of frequentist methods
- Future work:
  ➡ time series / other structured data
  ➡ speeding up computation
- On arXiv very soon (if you want a heads up: jhuggins@hsph.harvard.edu)

Thank you

Why can Bayesian model selection fail?

Why can Bayesian model selection fail?

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

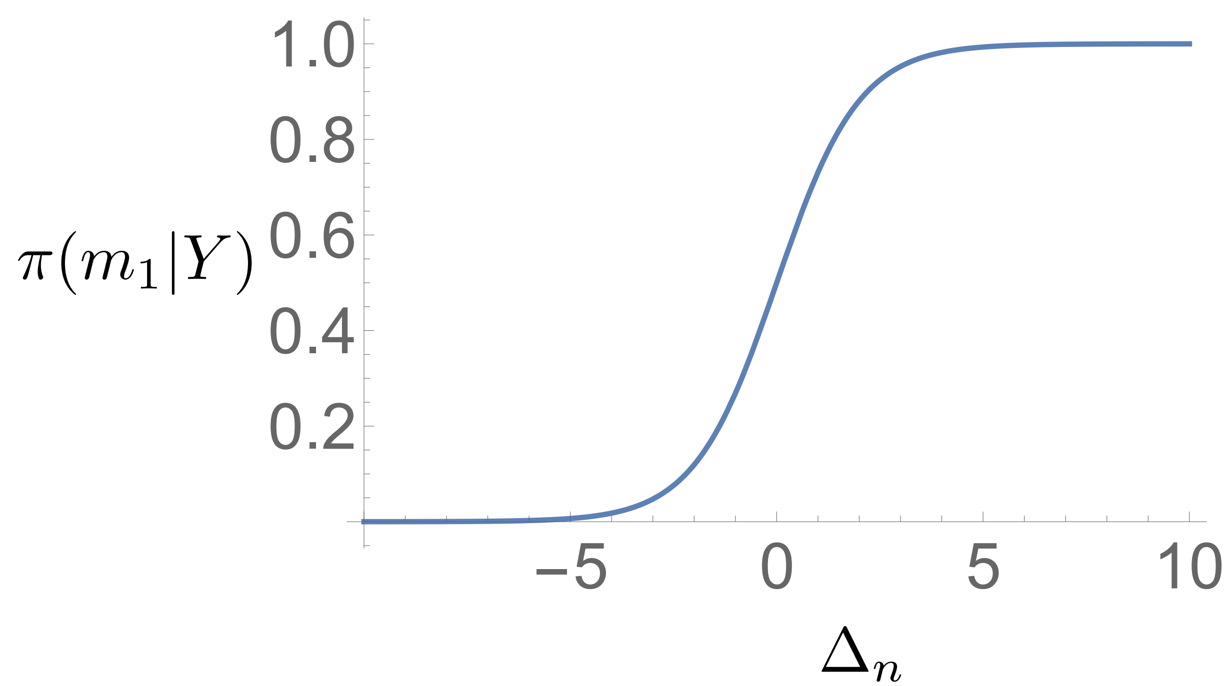
Why can Bayesian model selection fail?

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

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

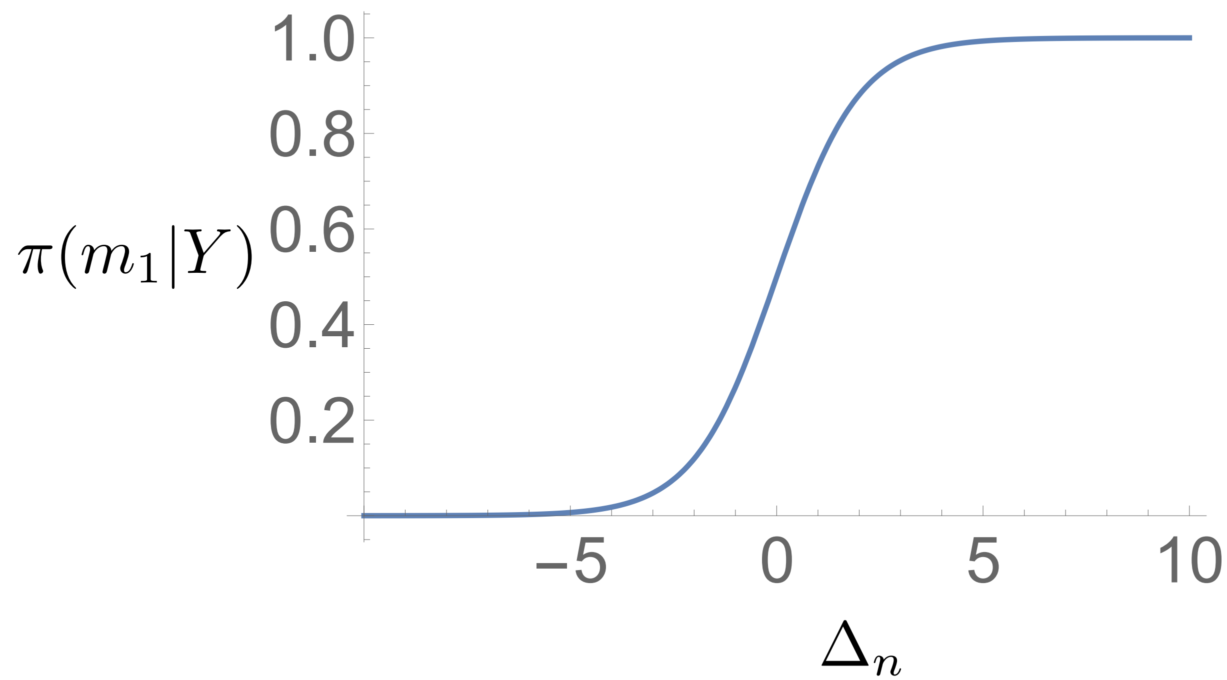


Why can Bayesian model selection fail?

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

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

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



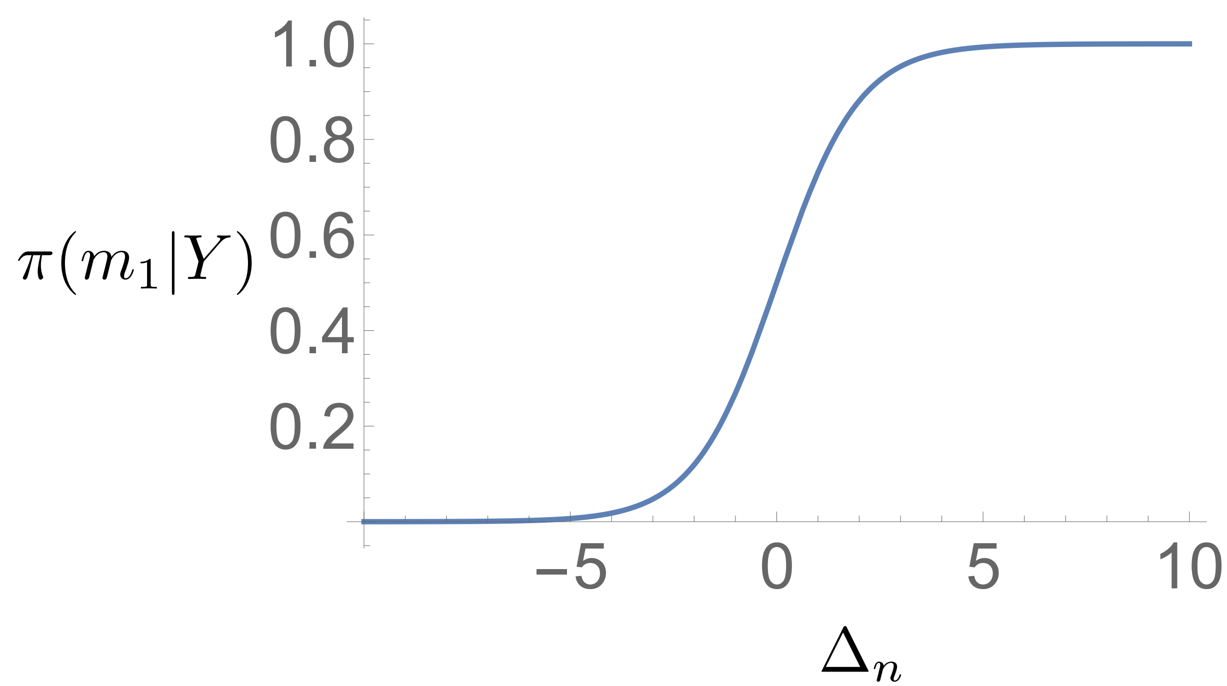
Why can Bayesian model selection fail?

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

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

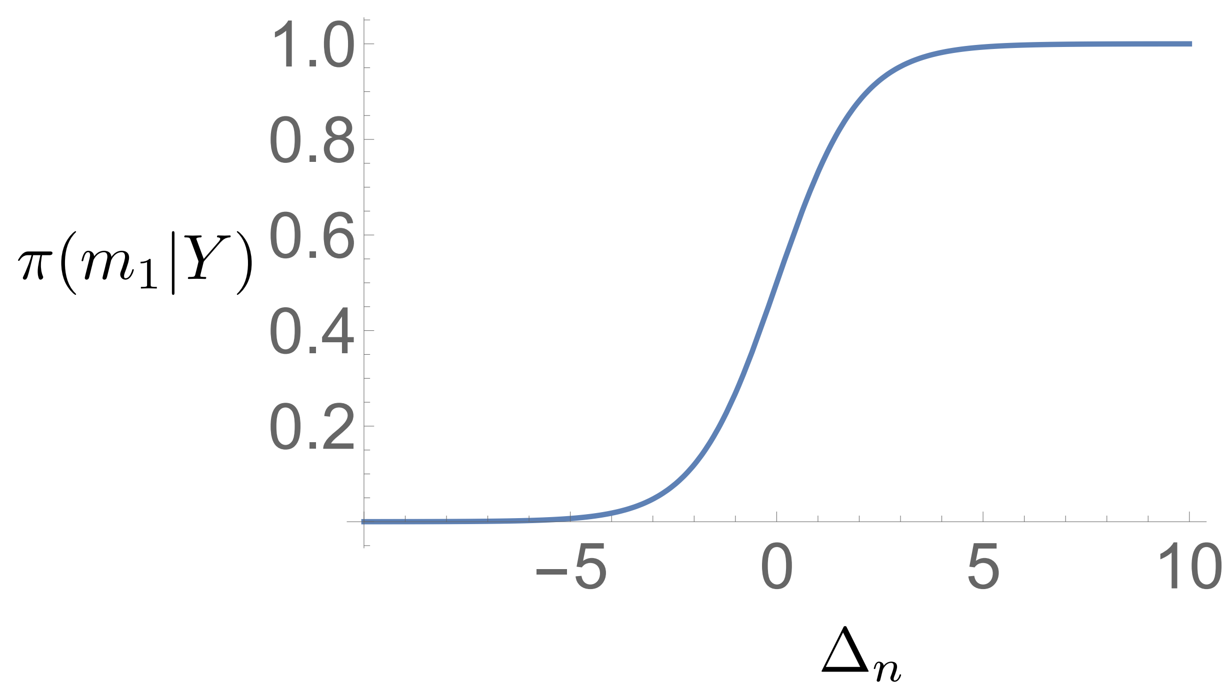
- By assumption, $\mathbb{E}[\delta_i] = 0$
  but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$

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



Why can Bayesian model selection fail?

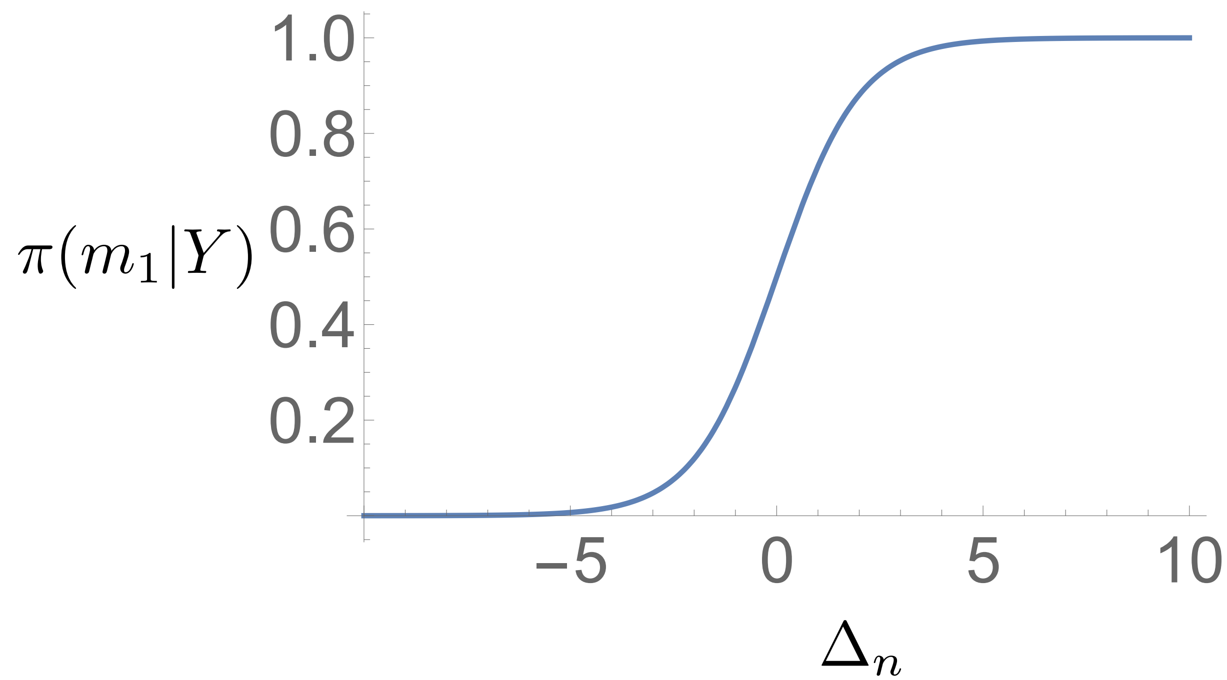
- Let $\Delta_n \triangleq \sum_{i=1}^{n} \underbrace{\log p(Y_i \mid m_1) - \log p(Y_i \mid m_2)}_{\delta_i}$
- Then $\pi(m_1 \mid Y) = (1 + \exp\{-\Delta_n\})^{-1}$
- By assumption, $\mathbb{E}[\delta_i] = 0$ but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$
- Hence, Δ_n is a random walk with $\mathbb{E}[\Delta_n^2] = \sigma^2 n$
- 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 $\pi(m_1 \,|\, Y) = (1 + \exp\{-\Delta_n\})^{-1}$



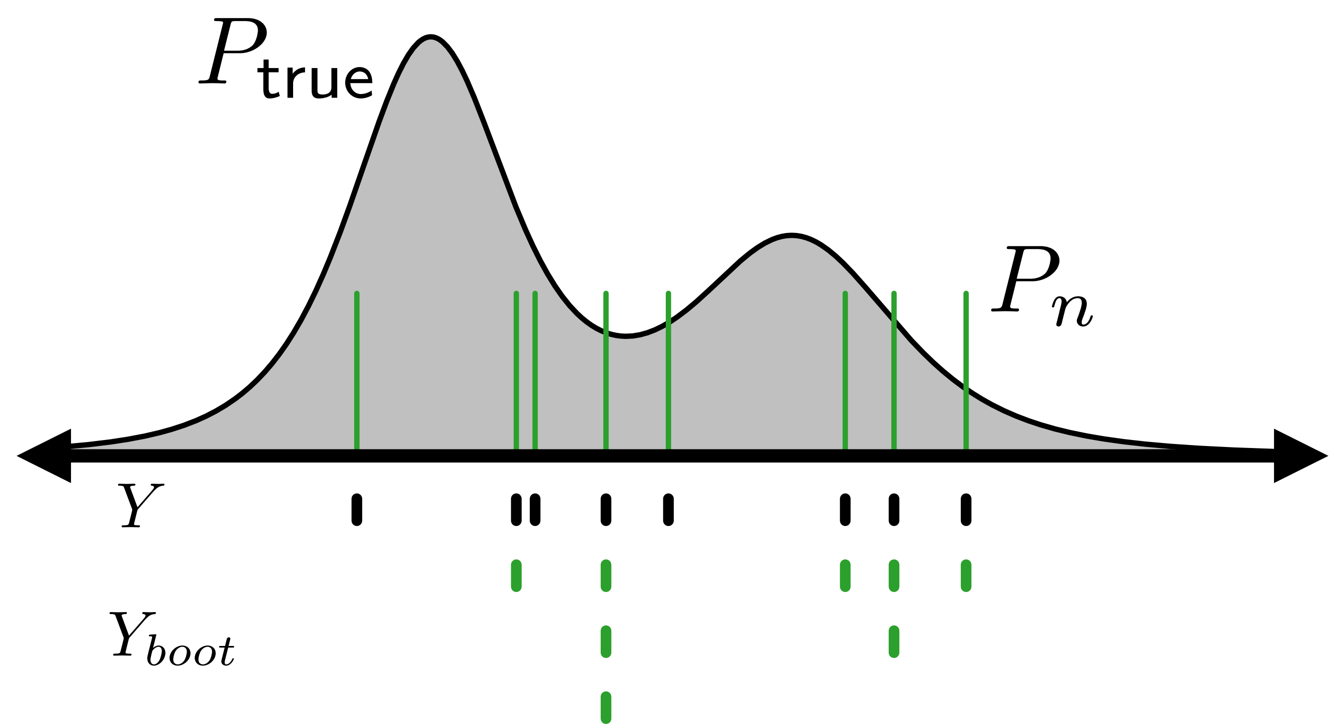
- By assumption, $\mathbb{E}[\delta_i] = 0$ but $\sigma^2 = \mathrm{Var}(\delta_i) > 0$

- 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

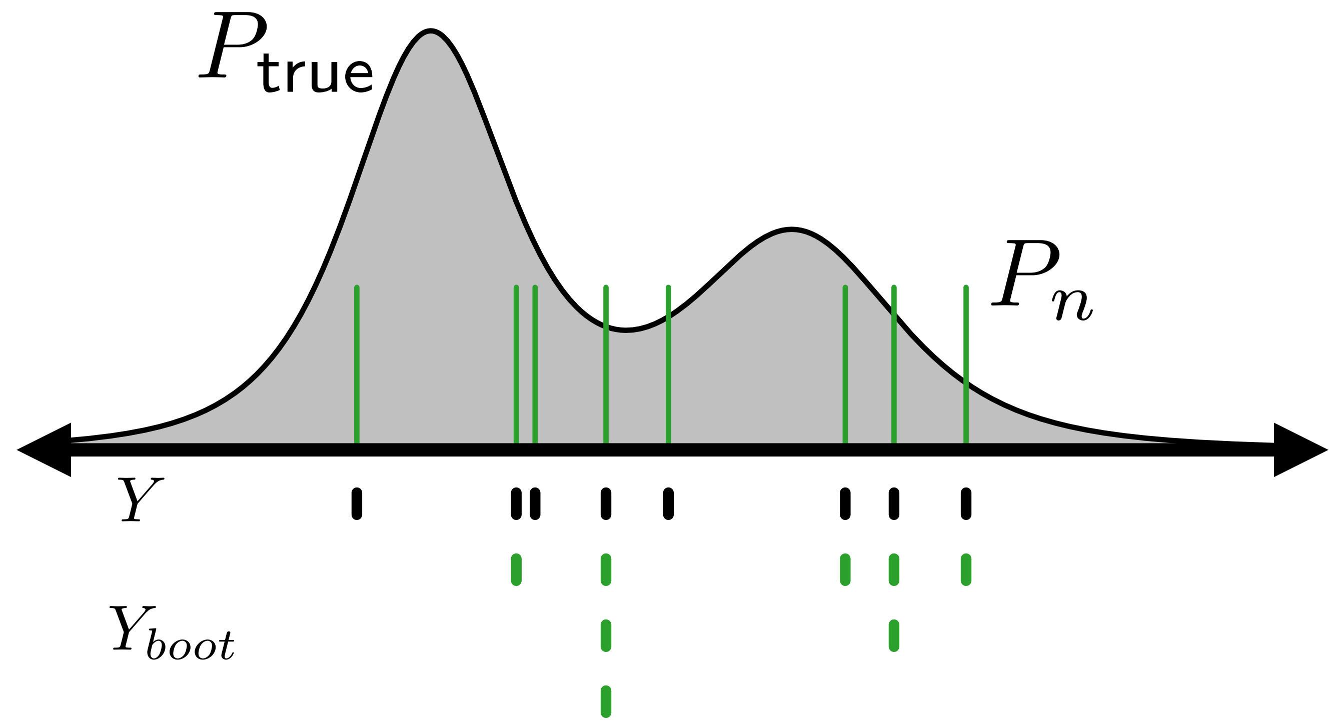
Bootstrap aggregating (bagging)



[Breiman 1995, Bühlmann & Yu 2002]

Bootstrap aggregating (bagging)

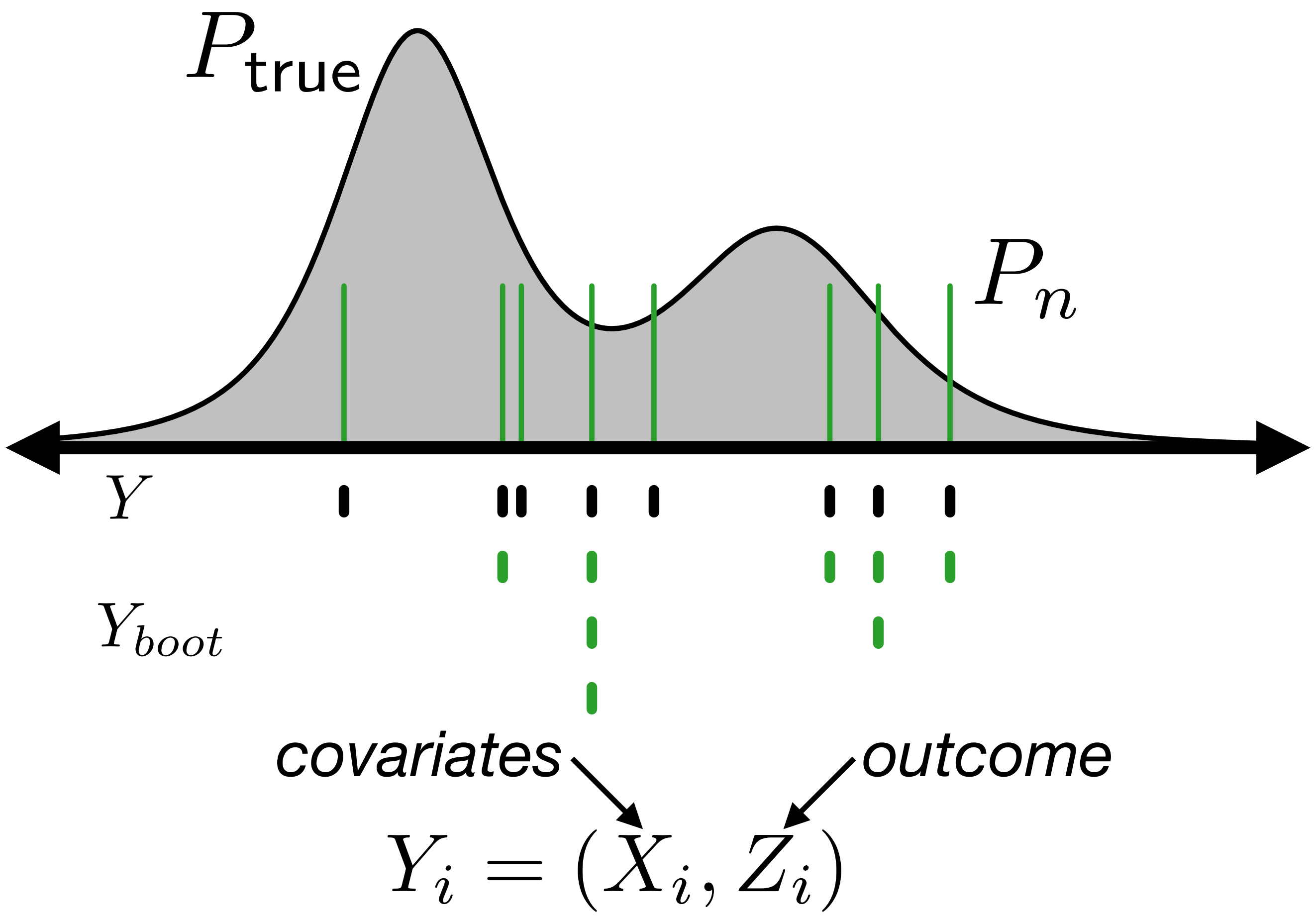
- **Have:** samples $Y = (Y_1, \ldots, Y_n)$



[Breiman 1995, Bühlmann & Yu 2002]

Bootstrap aggregating (bagging)

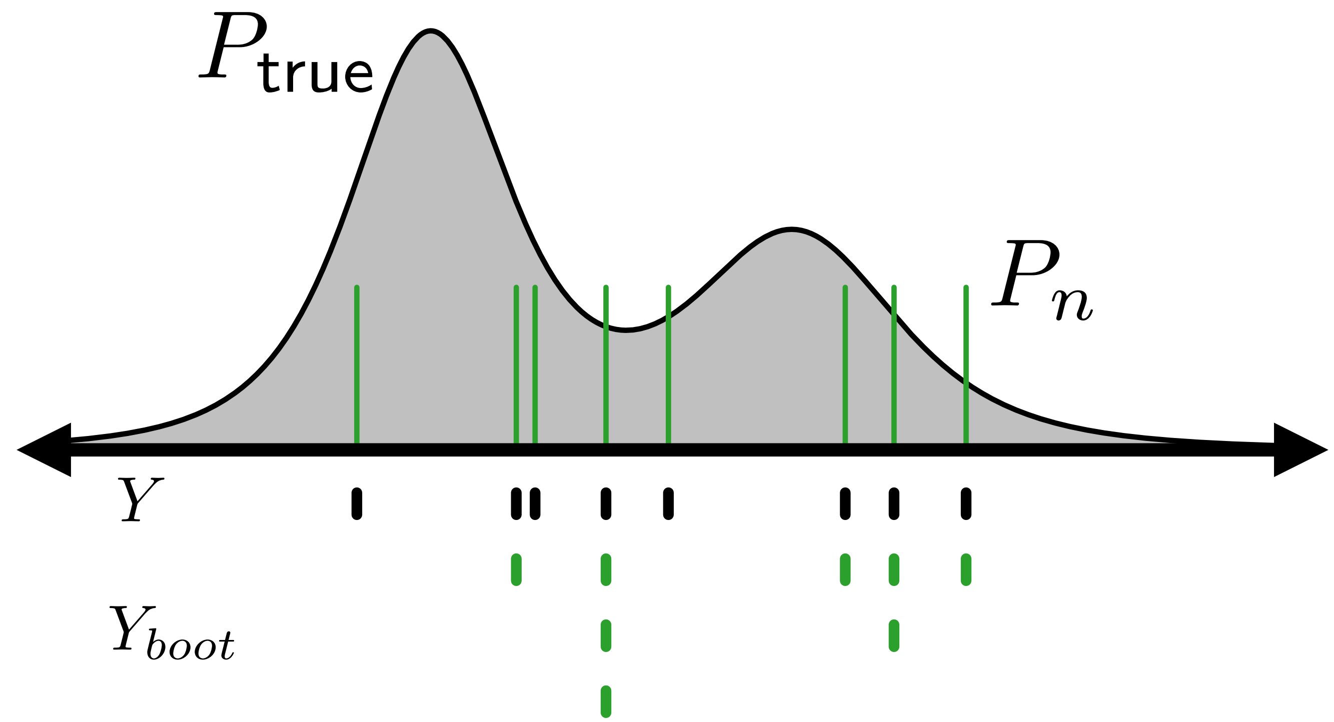
- **Have:** samples $Y = (Y_1, \ldots, Y_n)$
- **Goal:** predict future outcome based on Y **[i.e. regression]**



[Breiman 1995, Bühlmann & Yu 2002]

Bootstrap aggregating (bagging)

- **Have:** samples $Y = (Y_1, \ldots, Y_n)$
- **Goal:** predict future outcome based on Y **[i.e. regression]**



covariates → *outcome*

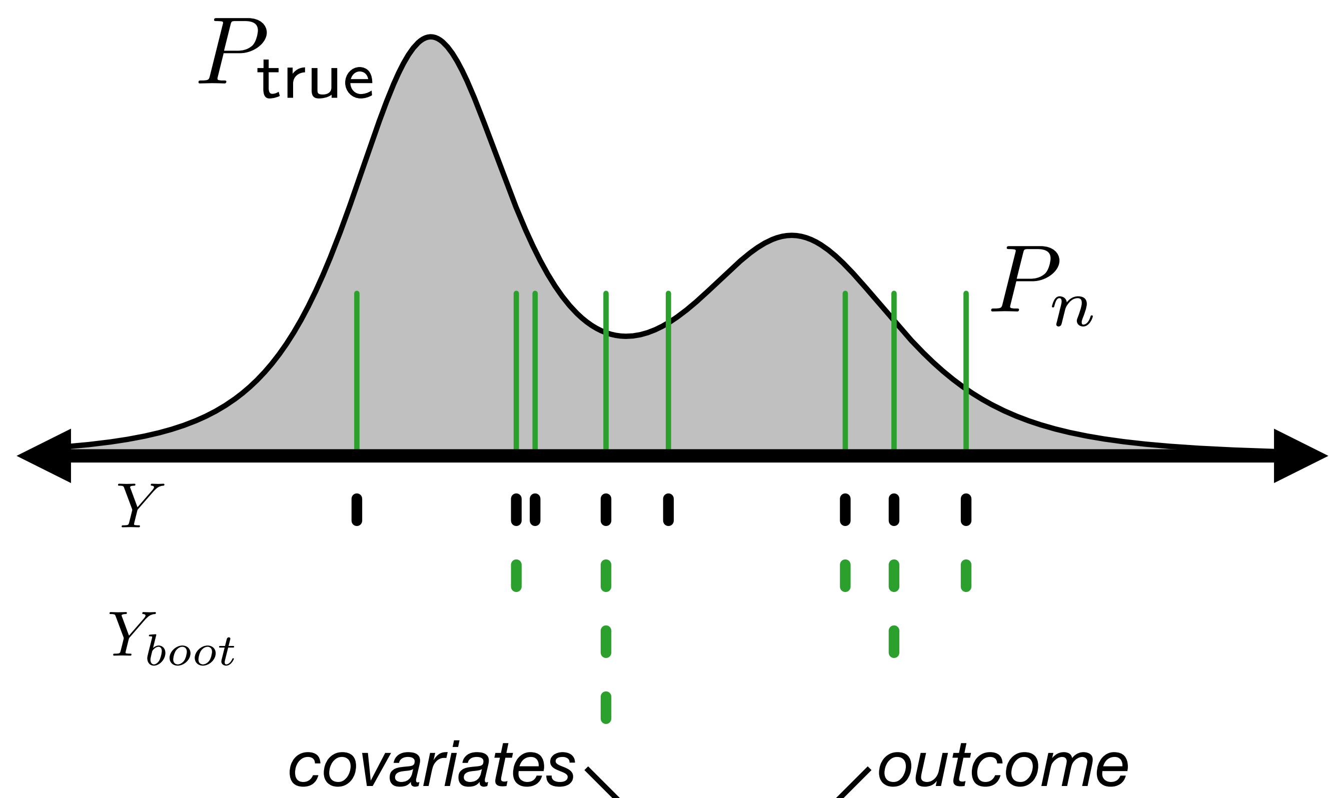
$$Y_i = (X_i, Z_i)$$

$$Z_{pred} = f(X_{new}; Y)$$

[Breiman 1995, Bühlmann & Yu 2002]

Bootstrap aggregating (bagging)

- **Have:** samples $Y = (Y_1, \ldots, Y_n)$
- **Goal:** predict future outcome based on Y **[i.e. regression]**
- **Problem:** prediction algorithm is unstable **[e.g. regression trees]**



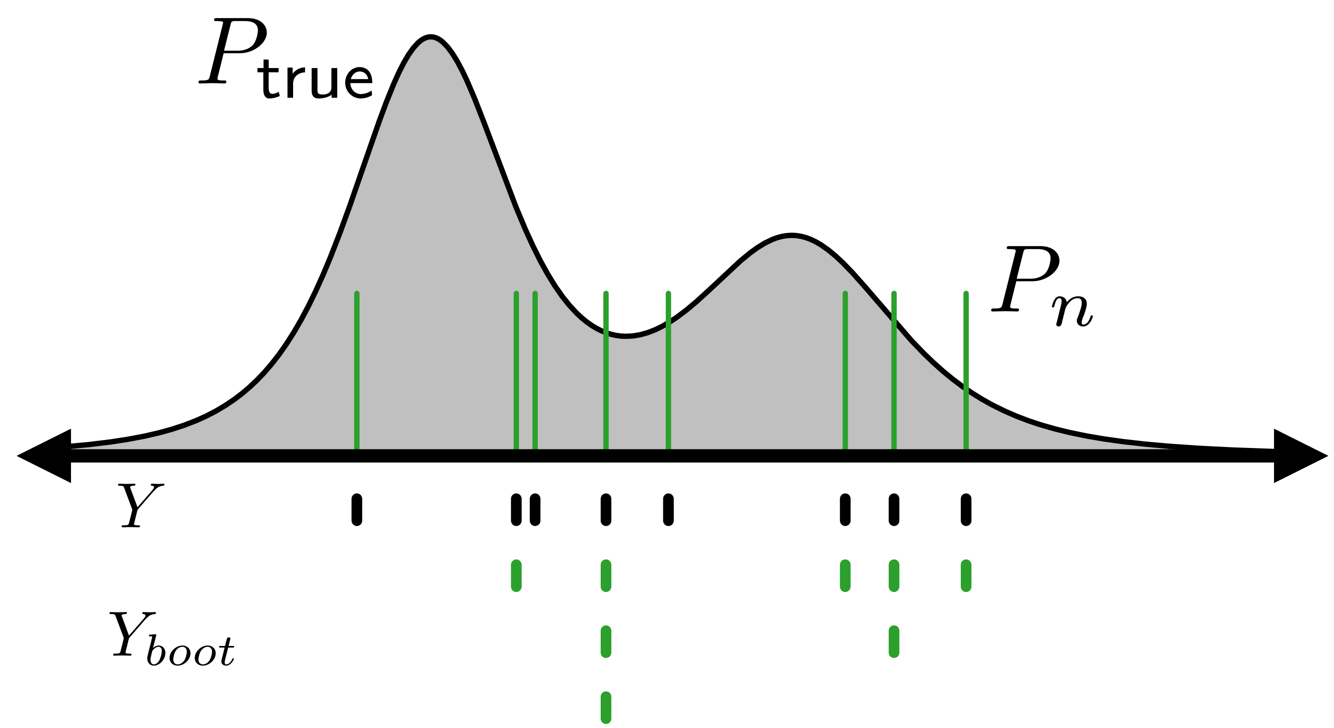
$$Y_i = (X_i, Z_i)$$

$$Z_{pred} = f(X_{new}; Y)$$

[Breiman 1995, Bühlmann & Yu 2002]

Bootstrap aggregating (bagging)

- **Have:** samples $Y = (Y_1, \ldots, Y_n)$
- **Goal:** predict future outcome based on Y **[i.e. regression]**
- **Problem:** prediction algorithm is unstable **[e.g. regression trees]**
- **Bagging:** stabilize predictions by aggregating (averaging) over predictions based on bootstrapped datasets



covariates *outcome*

$$Y_i = (X_i, Z_i)$$

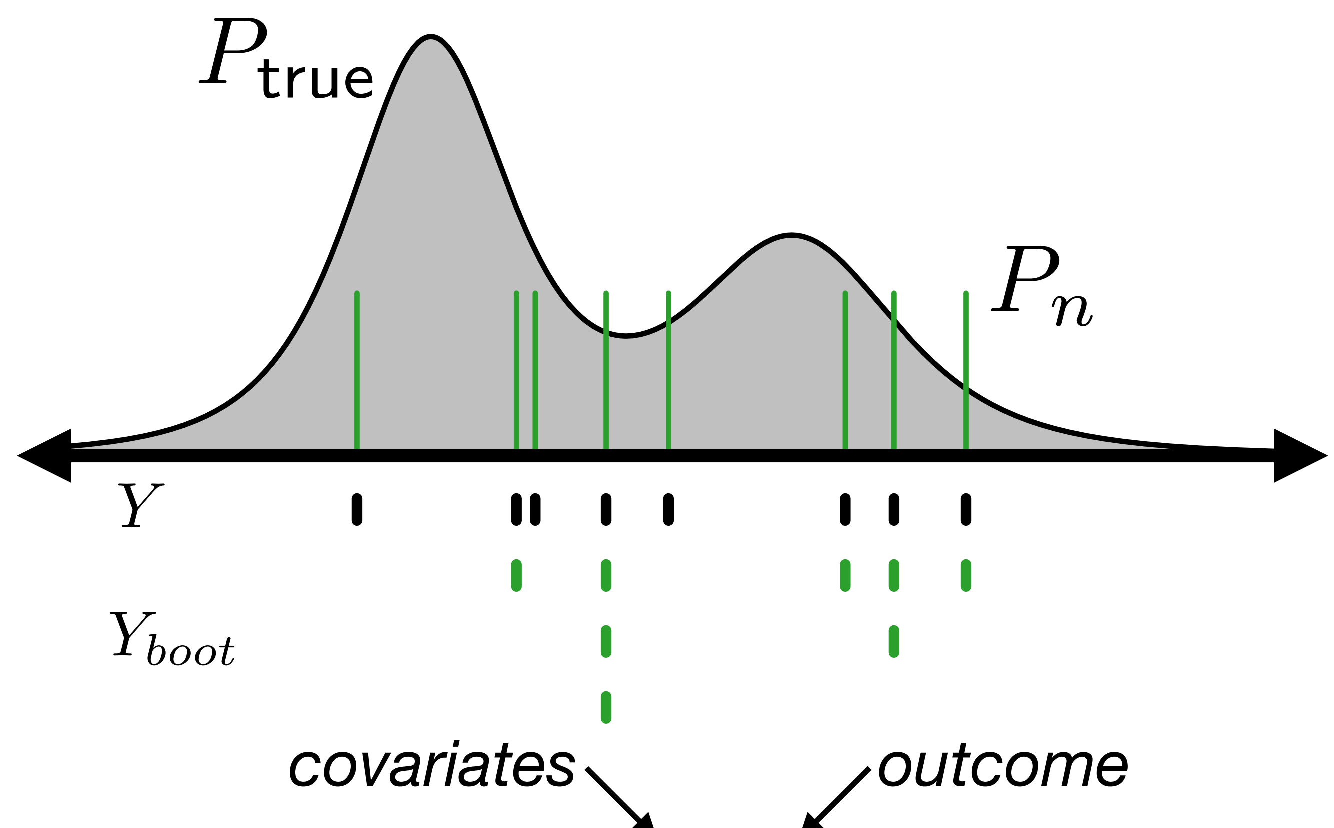
$$Z_{pred} = f(X_{new}; Y)$$

$$Z_{pred}^{bag} = \frac{1}{B}\sum_{b=1}^{B} f(X_{new}; Y_{boot}^{(b)})$$

[Breiman 1995, Bühlmann & Yu 2002]

Bootstrap aggregating (bagging)

- **Have:** samples $Y = (Y_1, \ldots, Y_n)$
- **Goal:** predict future outcome based on Y **[i.e. regression]**
- **Problem:** prediction algorithm is unstable **[e.g. regression trees]**
- **Bagging:** stabilize predictions by aggregating (averaging) over predictions based on bootstrapped datasets
- Like bagging, BayesBag seems to work well with B = 50 or 100



$$Y_i = (X_i, Z_i)$$

$$Z_{pred} = f(X_{new}; Y)$$

$$Z_{pred}^{bag} = \frac{1}{B}\sum_{b=1}^{B} f(X_{new}; Y_{boot}^{(b)})$$

$$\pi_{BB}(\theta \mid Y) = \frac{1}{B}\sum_{b=1}^{B} \pi(\theta \mid Y_{boot}^{(b)})$$

[Breiman 1995, Bühlmann & Yu 2002]