The Bayesian Infinitesimal Jackknife for Variance

Tamara Broderick
Associate Professor,
MIT

With: Ryan Giordano 

A motivating example: elections

A motivating example: elections

- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)

A motivating example: elections

- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals

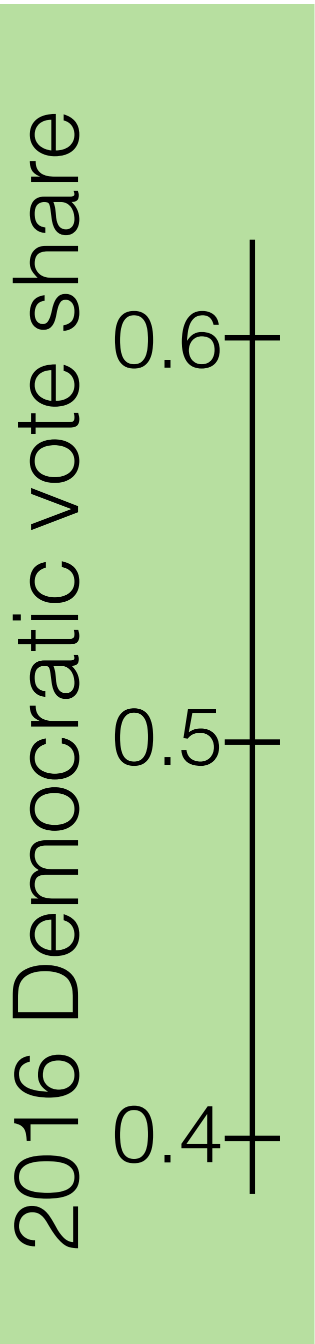
A motivating example: elections

- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day

A motivating example: elections

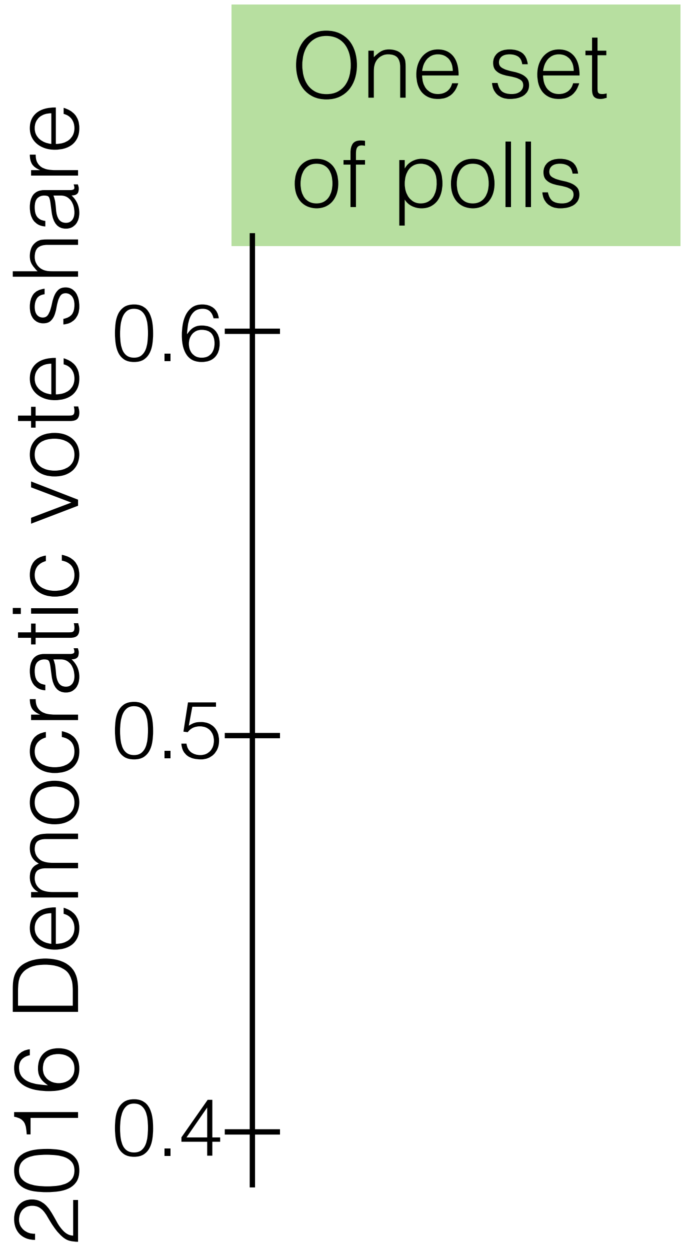
- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

A motivating example: elections



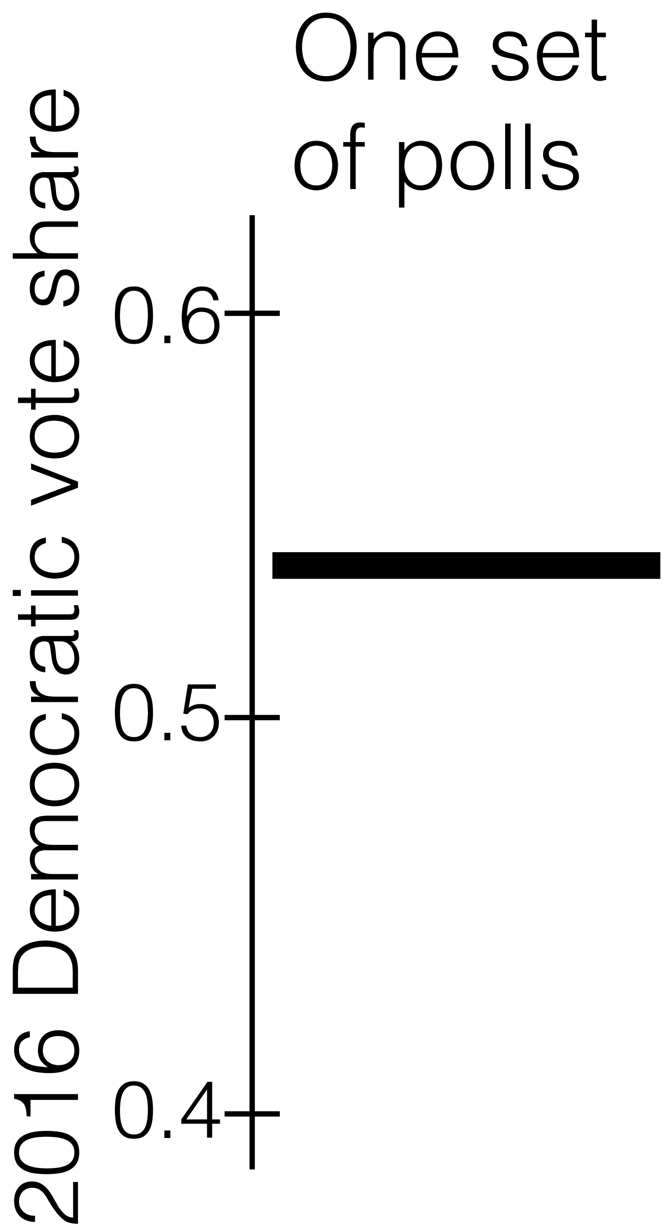
- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

A motivating example: elections



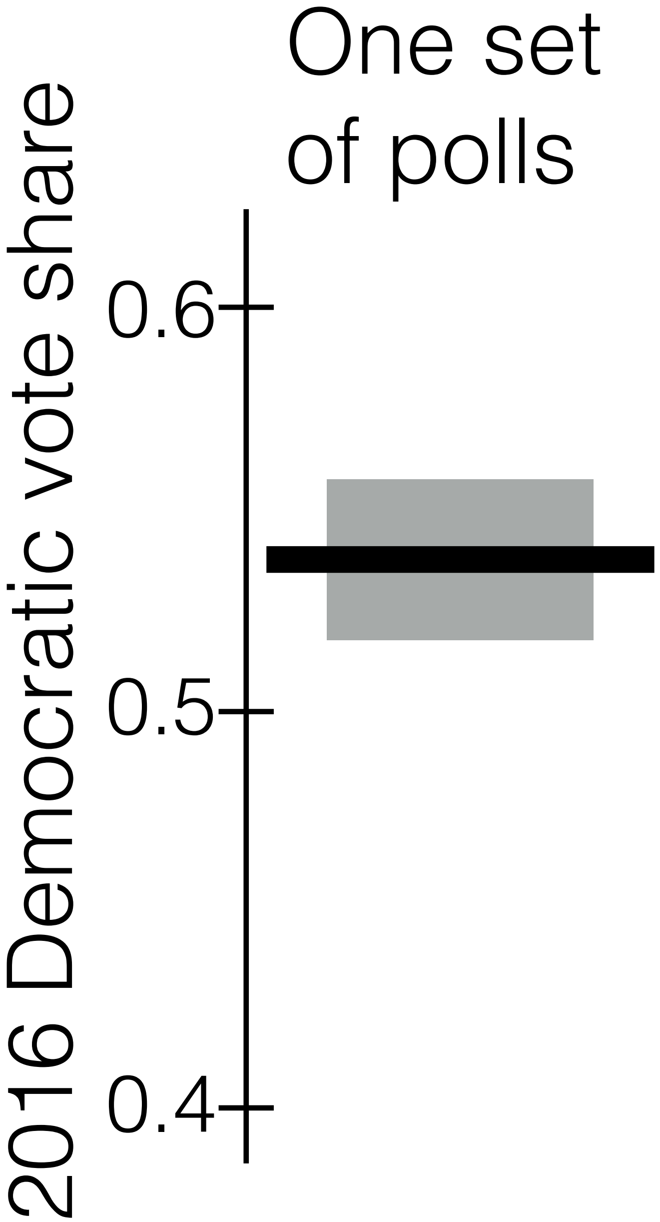
- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

A motivating example: elections



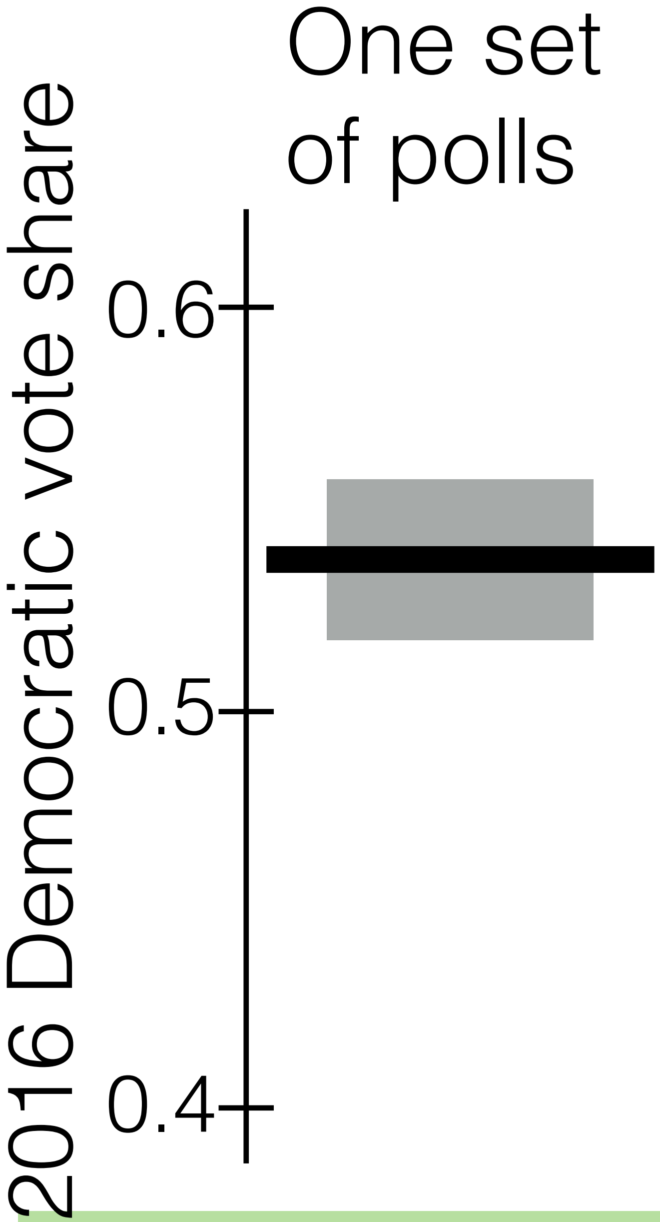
- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

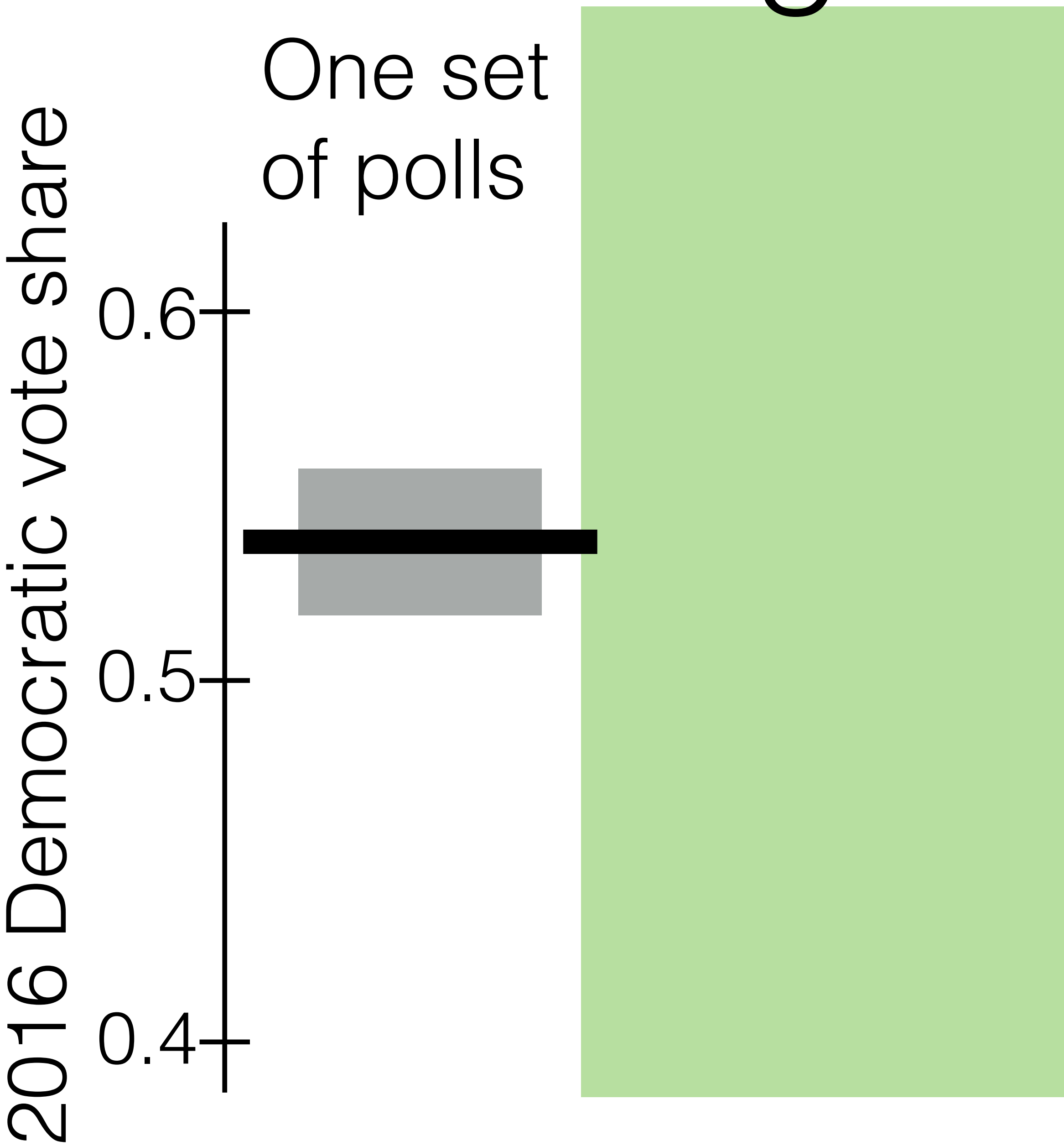
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data

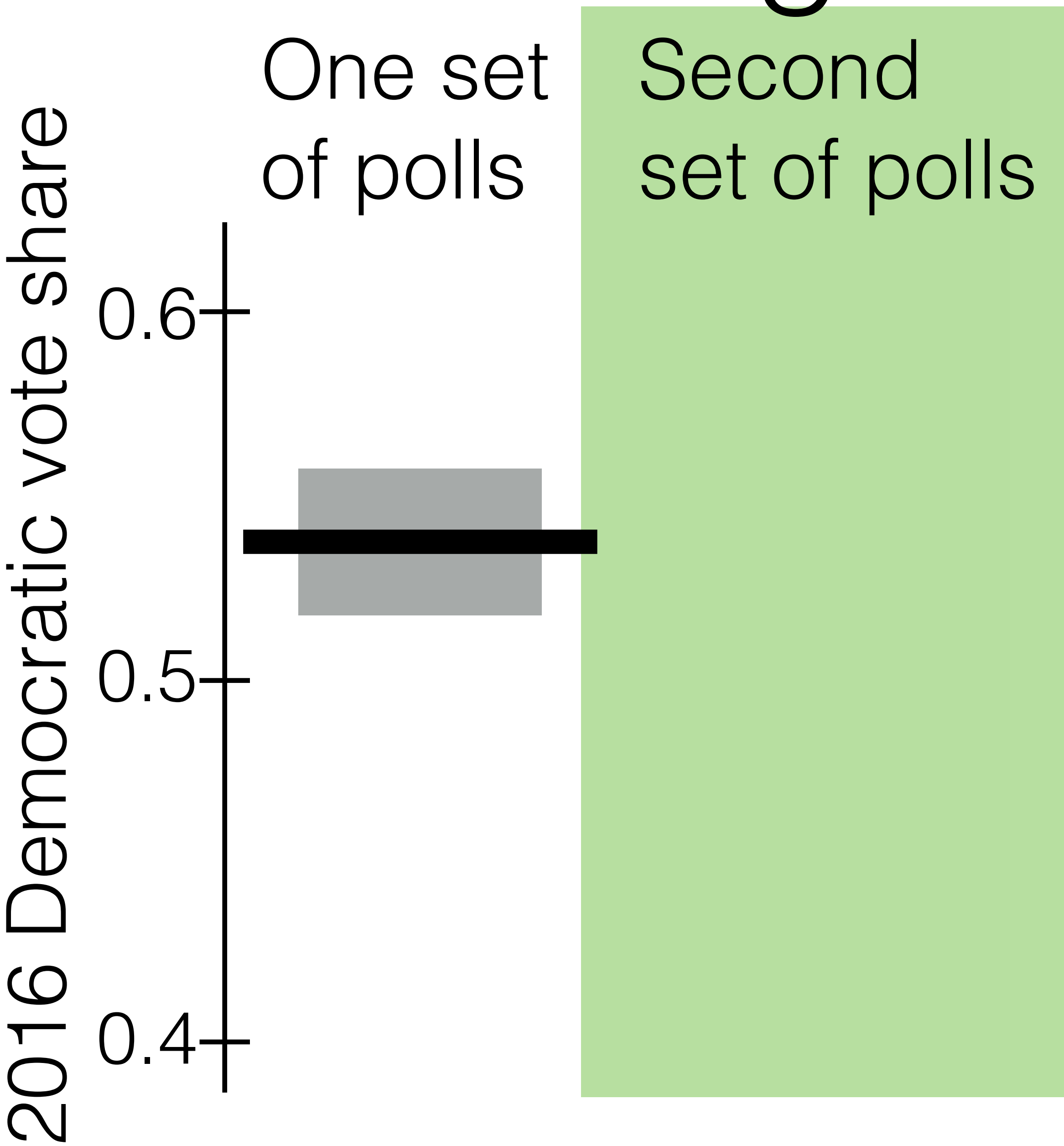
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data

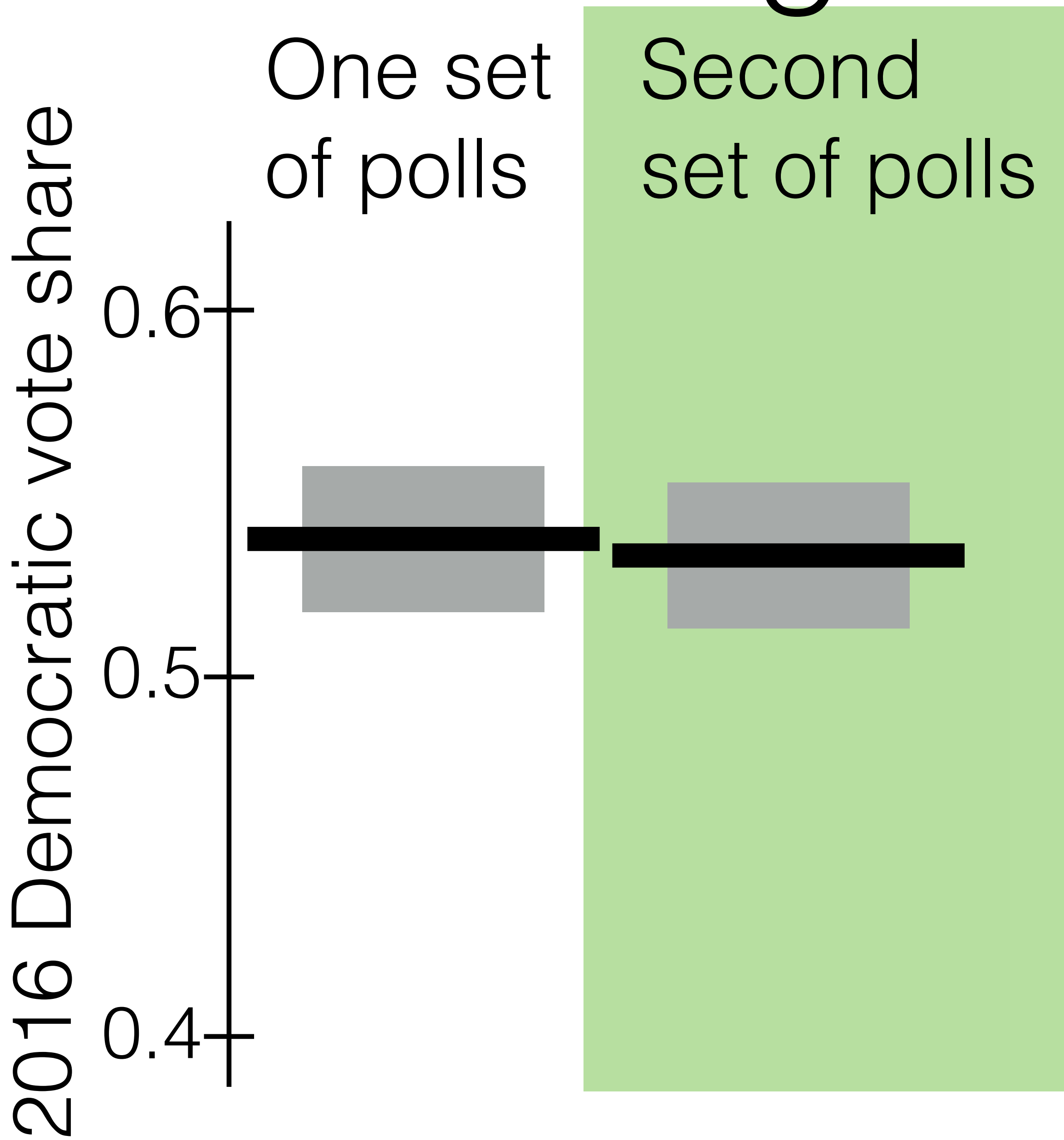
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data

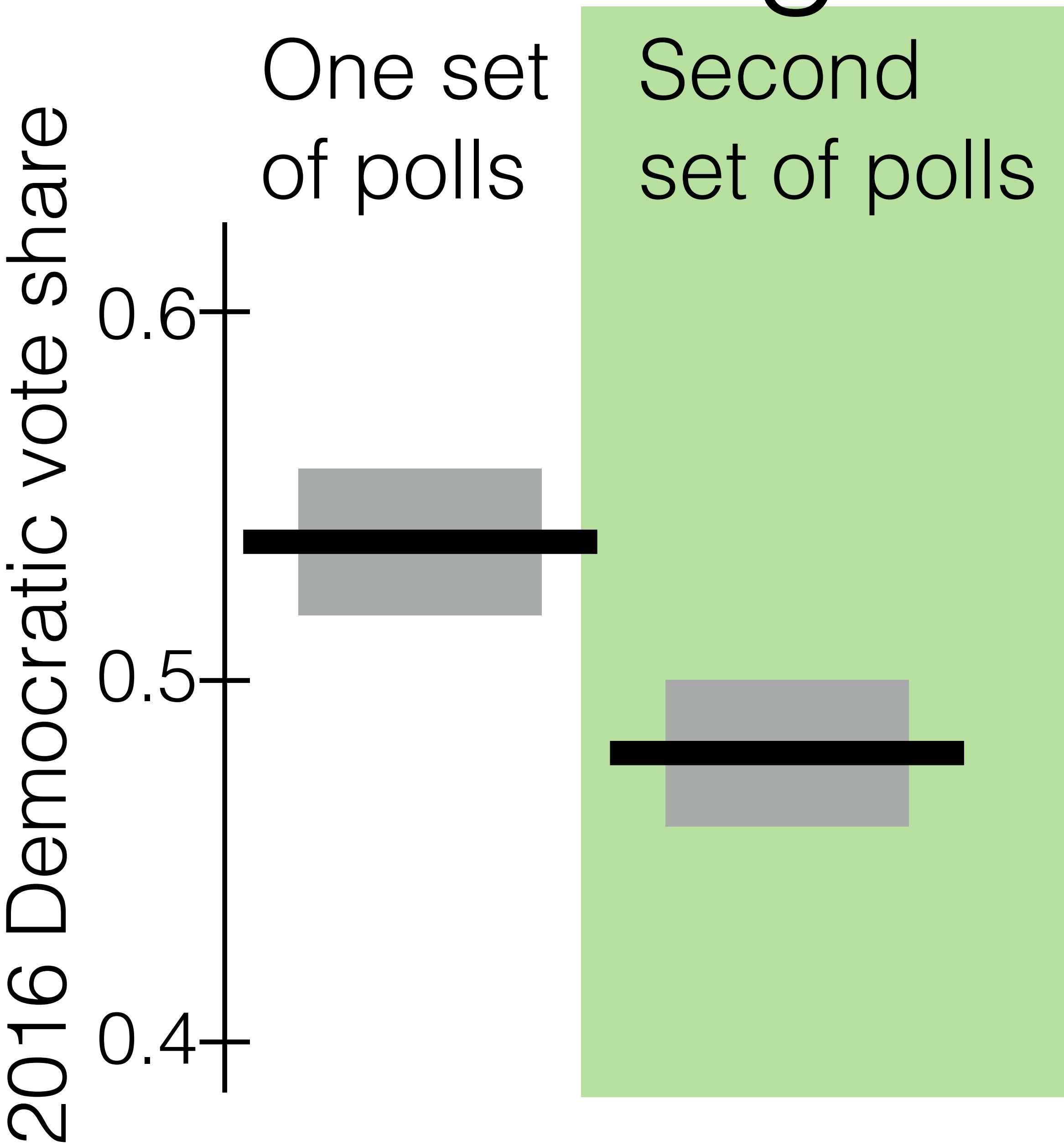
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data

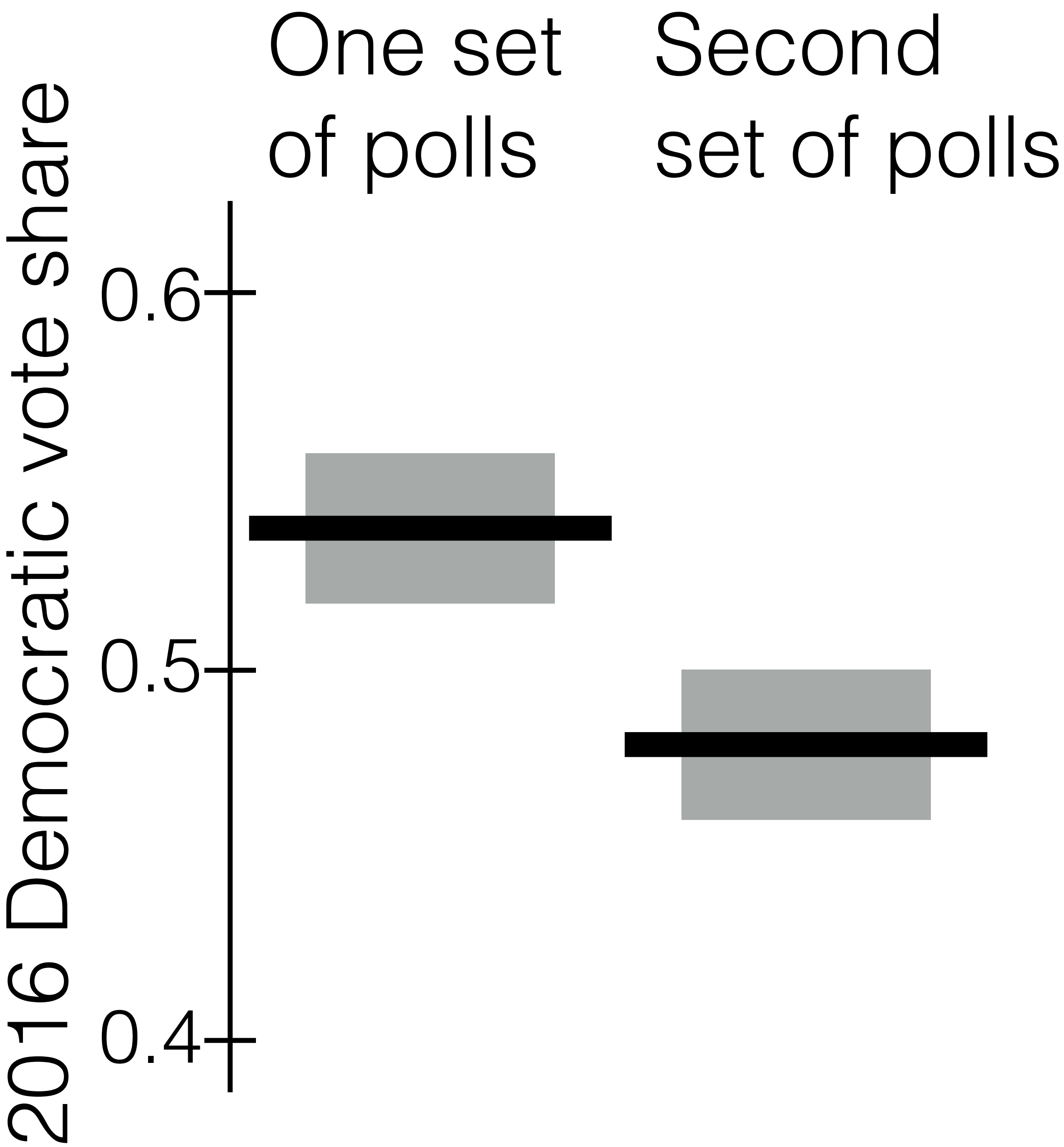
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data

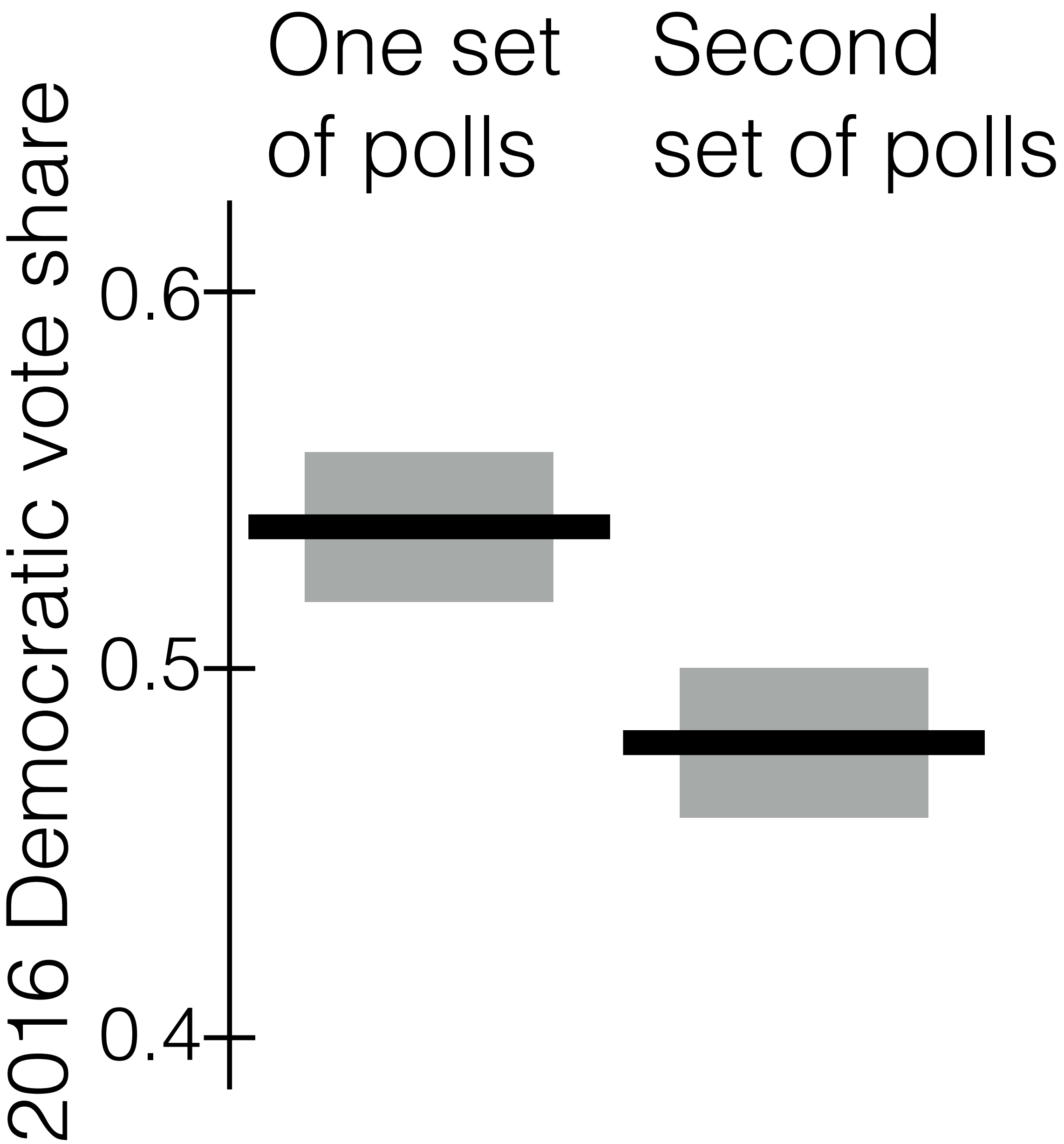
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data

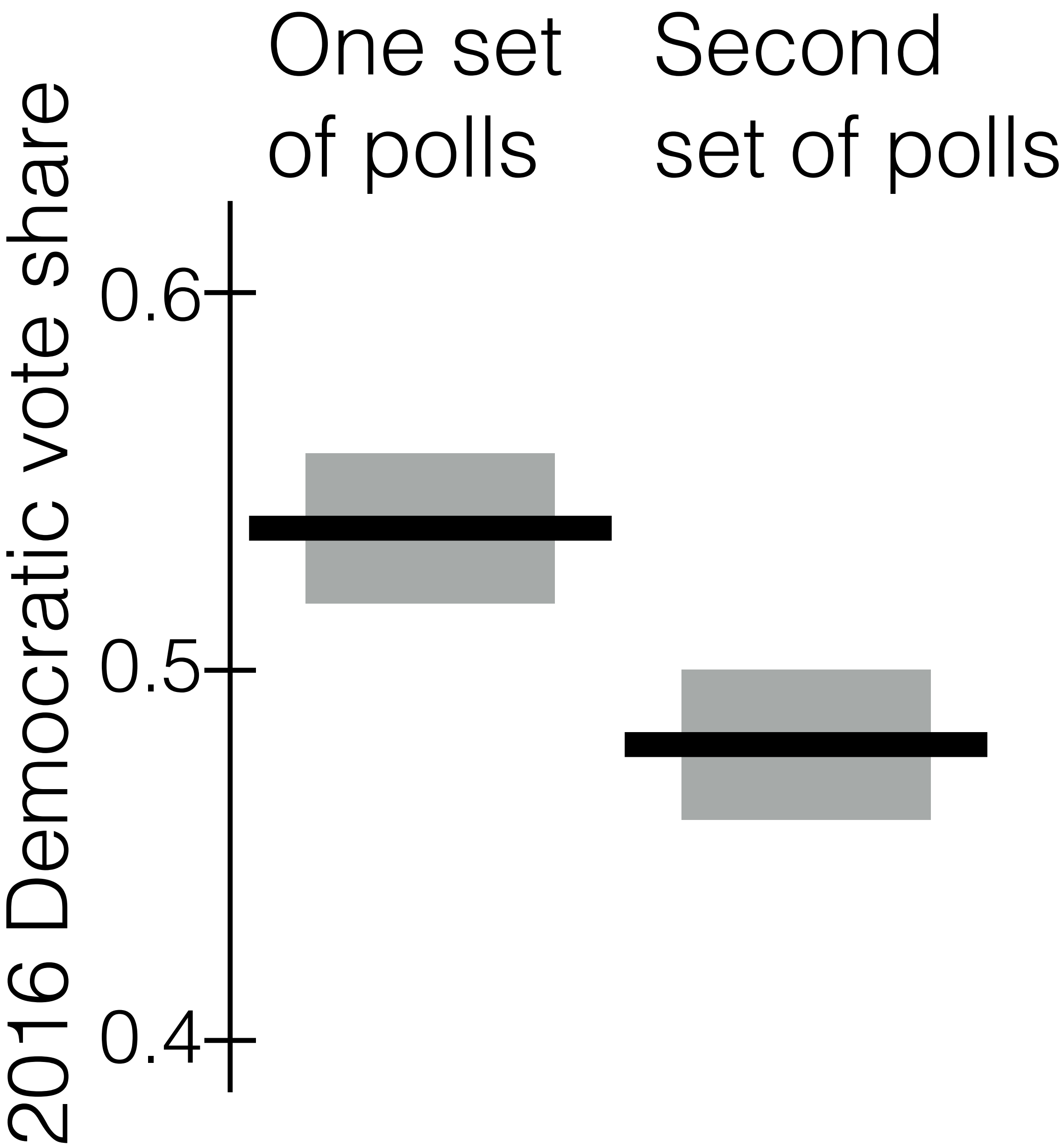
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data
  - If conclusions are very different, result is non-robust

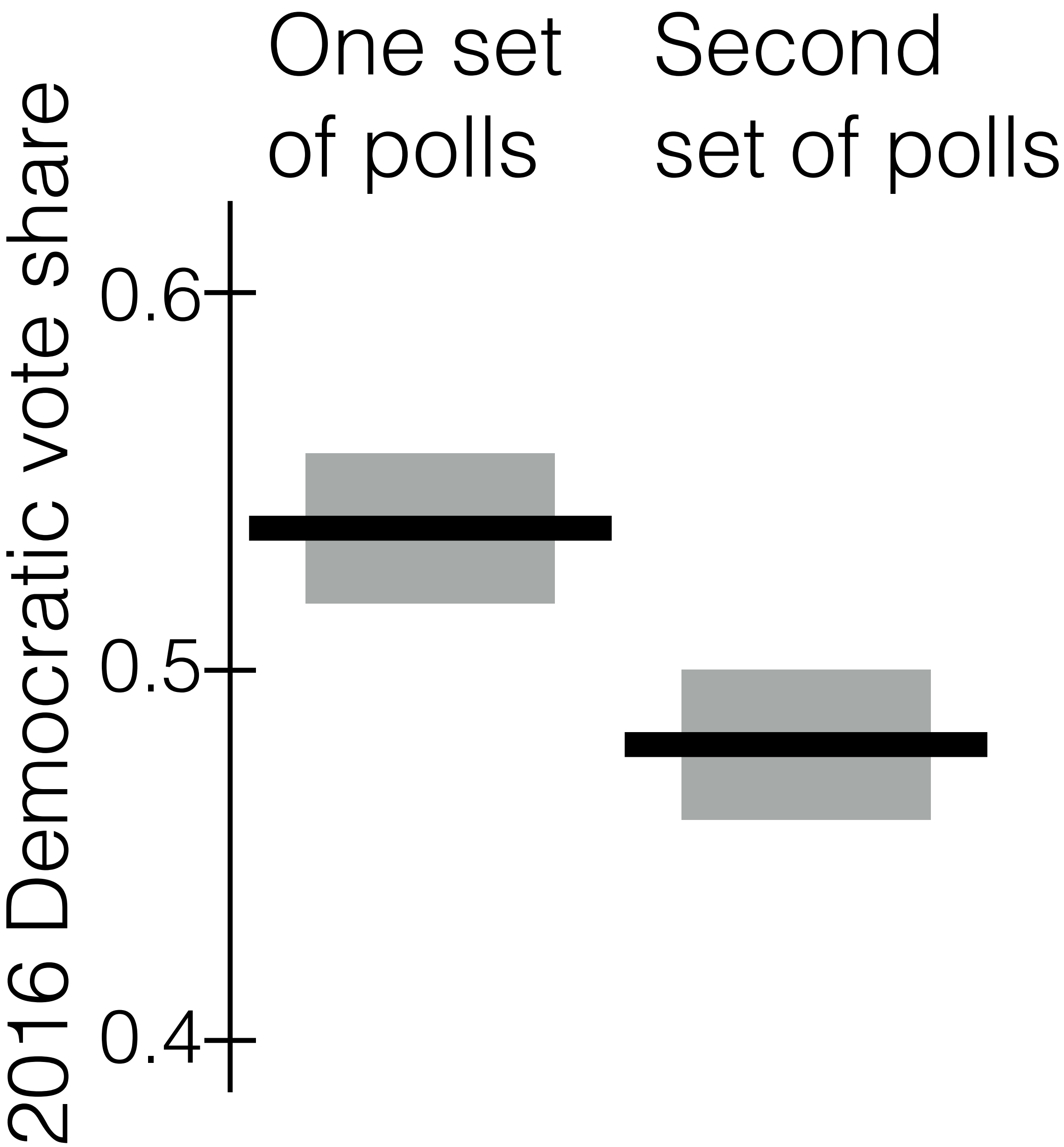
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data
  - If conclusions are very different, result is non-robust
- **How to check?**

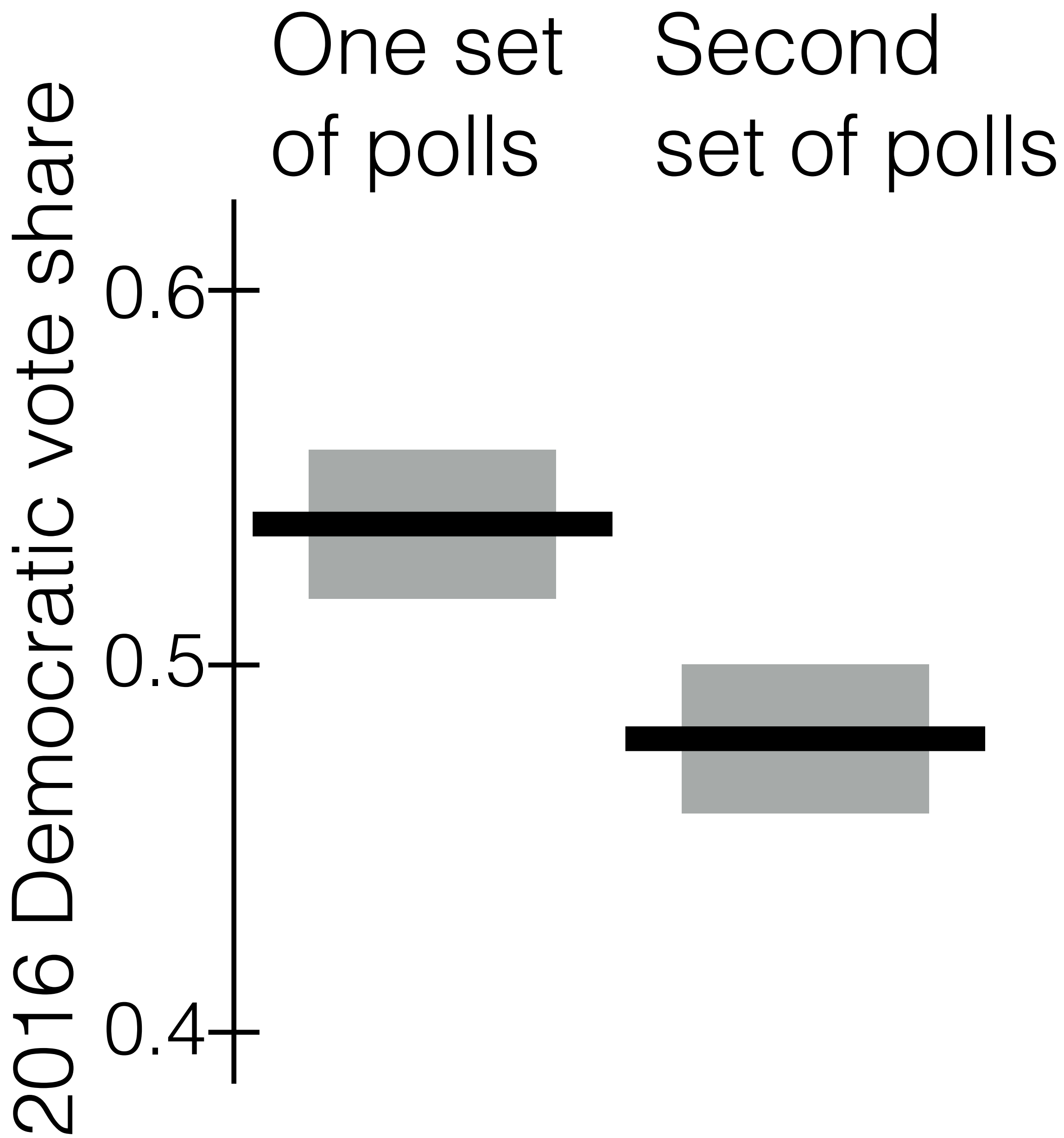
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data
  - If conclusions are very different, result is non-robust
- **How to check?** Bayesian uncertainty need not capture frequentist sampling uncertainty

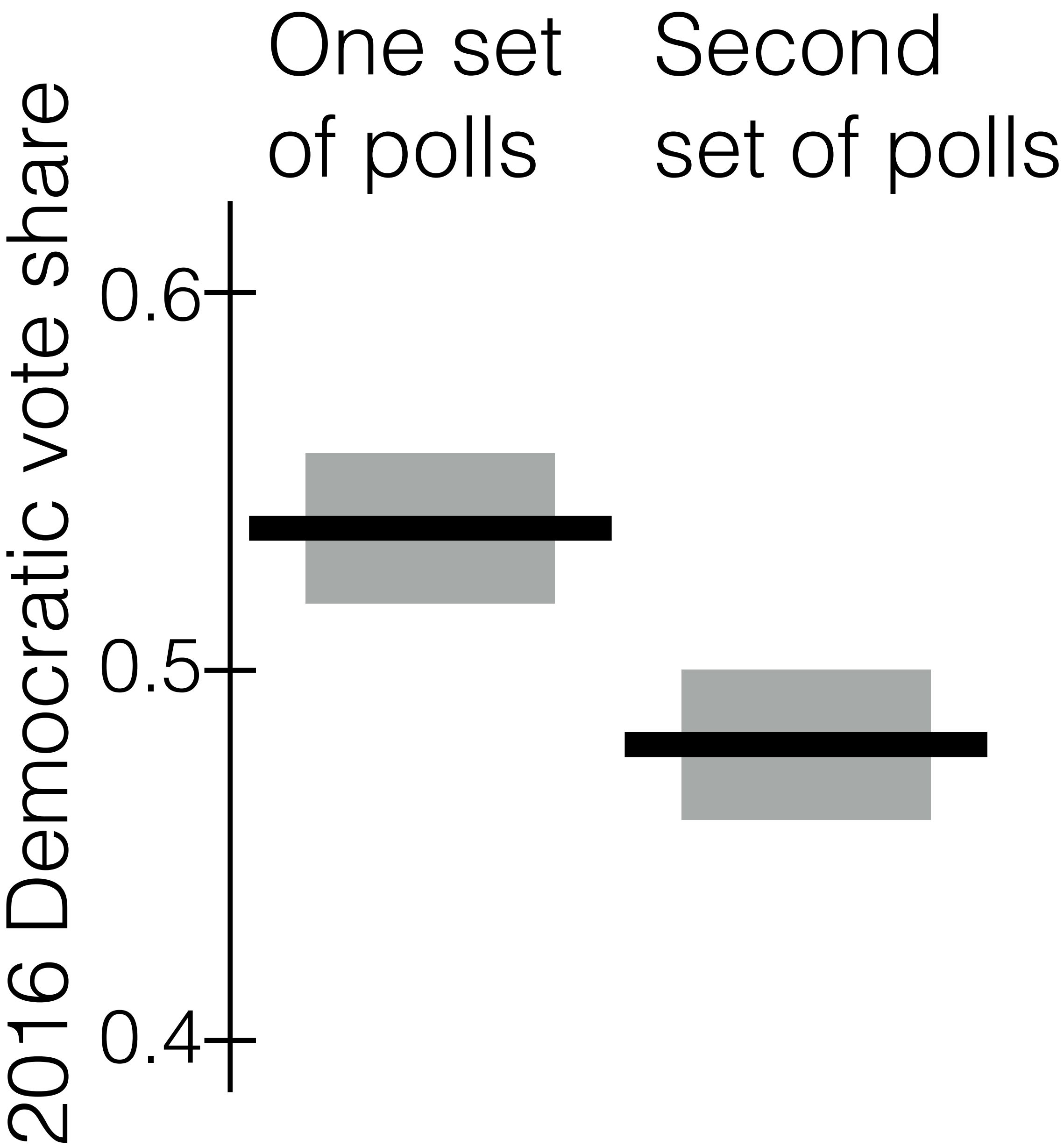
A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data
  - If conclusions are very different, result is non-robust
- **How to check?** Bayesian uncertainty need not capture frequentist sampling uncertainty; bootstrap too slow.

A motivating example: elections



- Lots of Bayesian analyses to predict US president election (The Economist, 538, etc)
- Polling data & fundamentals
- A parameter of interest: Democrat % on election day
- Report (MCMC-estimated) posterior mean & posterior standard deviation

- Suppose we re-sampled polls & got new data
  - If conclusions are very different, result is non-robust
- **How to check?** Bayesian uncertainty need not capture frequentist sampling uncertainty; bootstrap too slow. **We show infinitesimal jackknife is fast, accurate, automatic**

Roadmap

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)
- Idea: sandwich covariance (accuracy issues)

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)
- Idea: sandwich covariance (accuracy issues)
- Our proposal: use the infinitesimal jackknife (IJ)

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)
- Idea: sandwich covariance (accuracy issues)
- Our proposal: use the infinitesimal jackknife (IJ)
- What does the IJ estimator look like here?

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)
- Idea: sandwich covariance (accuracy issues)
- Our proposal: use the infinitesimal jackknife (IJ)
- What does the IJ estimator look like here?
- Some real data examples: elections and pilots

Roadmap

- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)
- Idea: sandwich covariance (accuracy issues)
- Our proposal: use the infinitesimal jackknife (IJ)
- What does the IJ estimator look like here?
- Some real data examples: elections and pilots
- Insights from theory

Roadmap

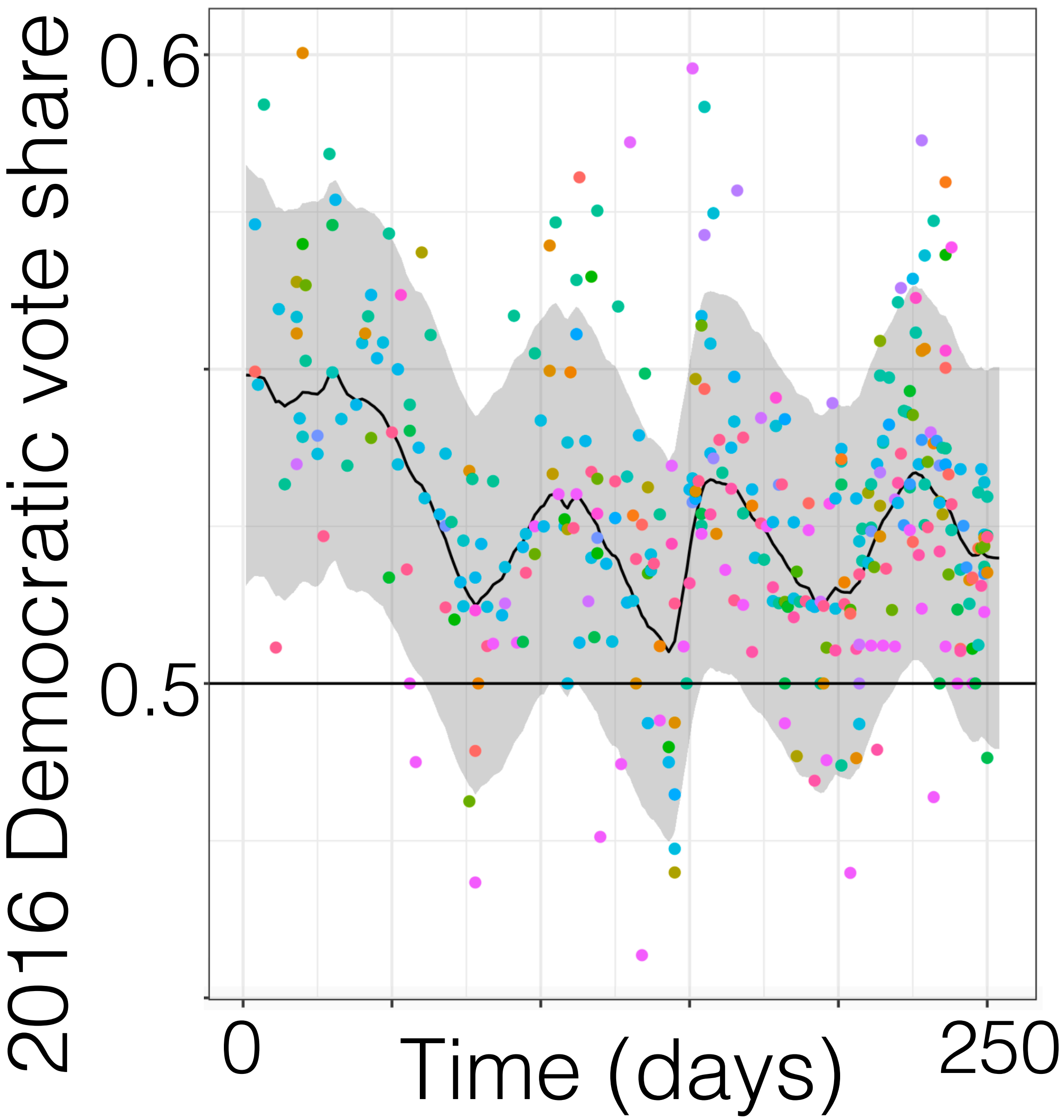
- Setup for frequentist uncertainty of a Bayesian estimator
- Idea: bootstrap (too slow)
- Idea: sandwich covariance (accuracy issues)
- Our proposal: use the infinitesimal jackknife (IJ)
- What does the IJ estimator look like here?
- Some real data examples: elections and pilots
- Insights from theory

General setup and notation

General setup and notation

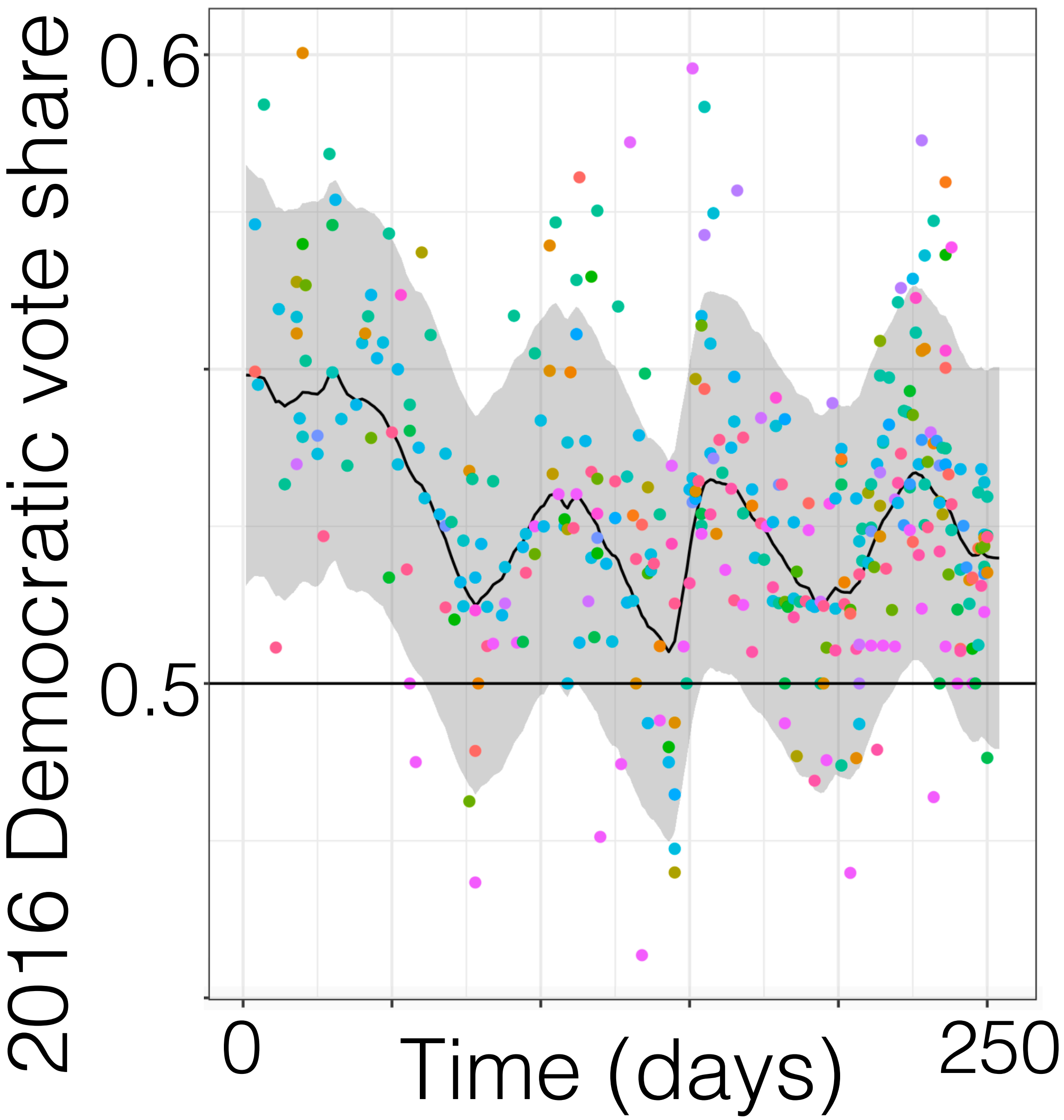
- Data: $X = (x_1, x_2, \ldots, x_N)$

General setup and notation



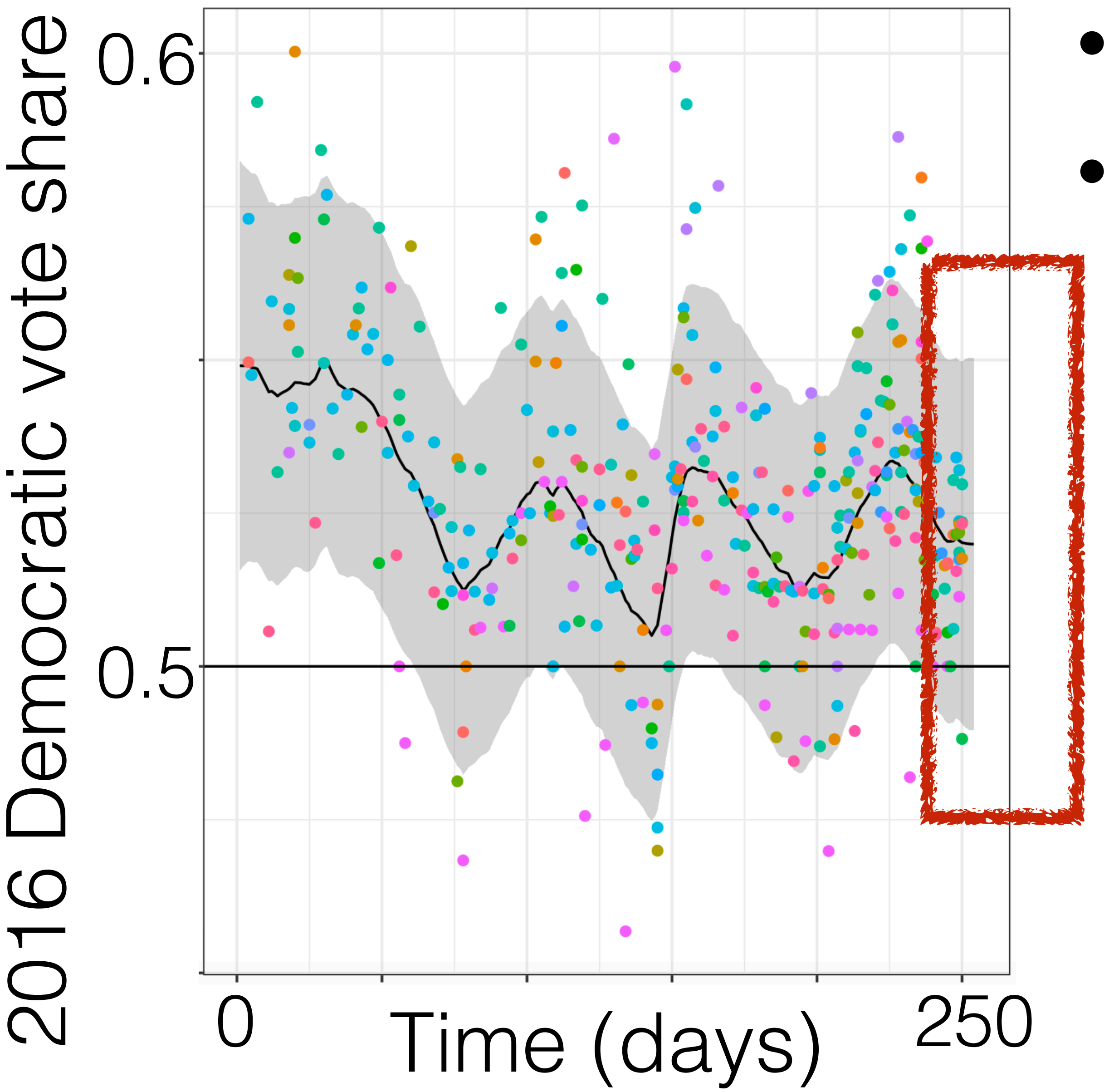
- Data: $X = (x_1, x_2, \ldots, x_N)$

General setup and notation



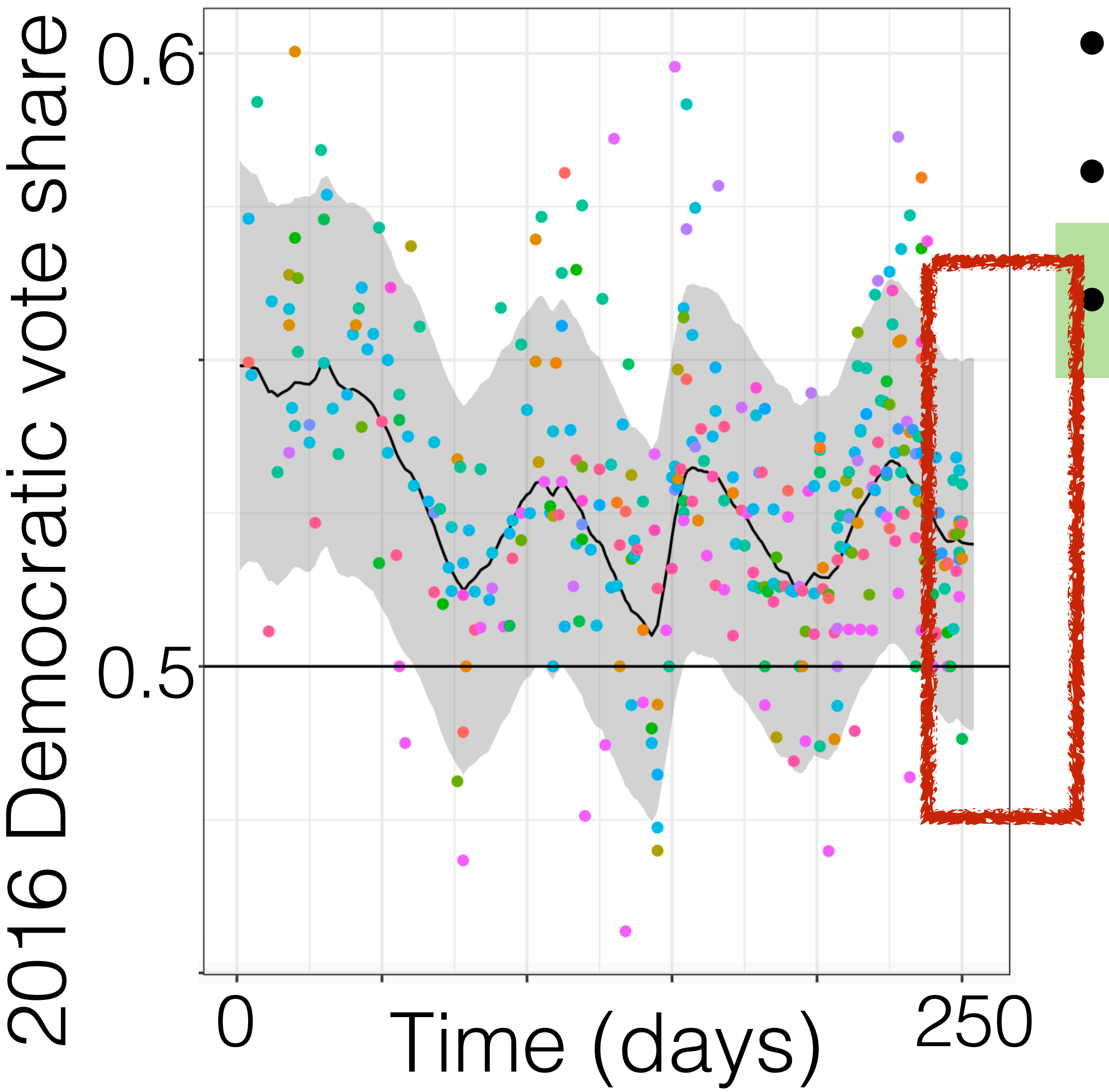
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ

General setup and notation



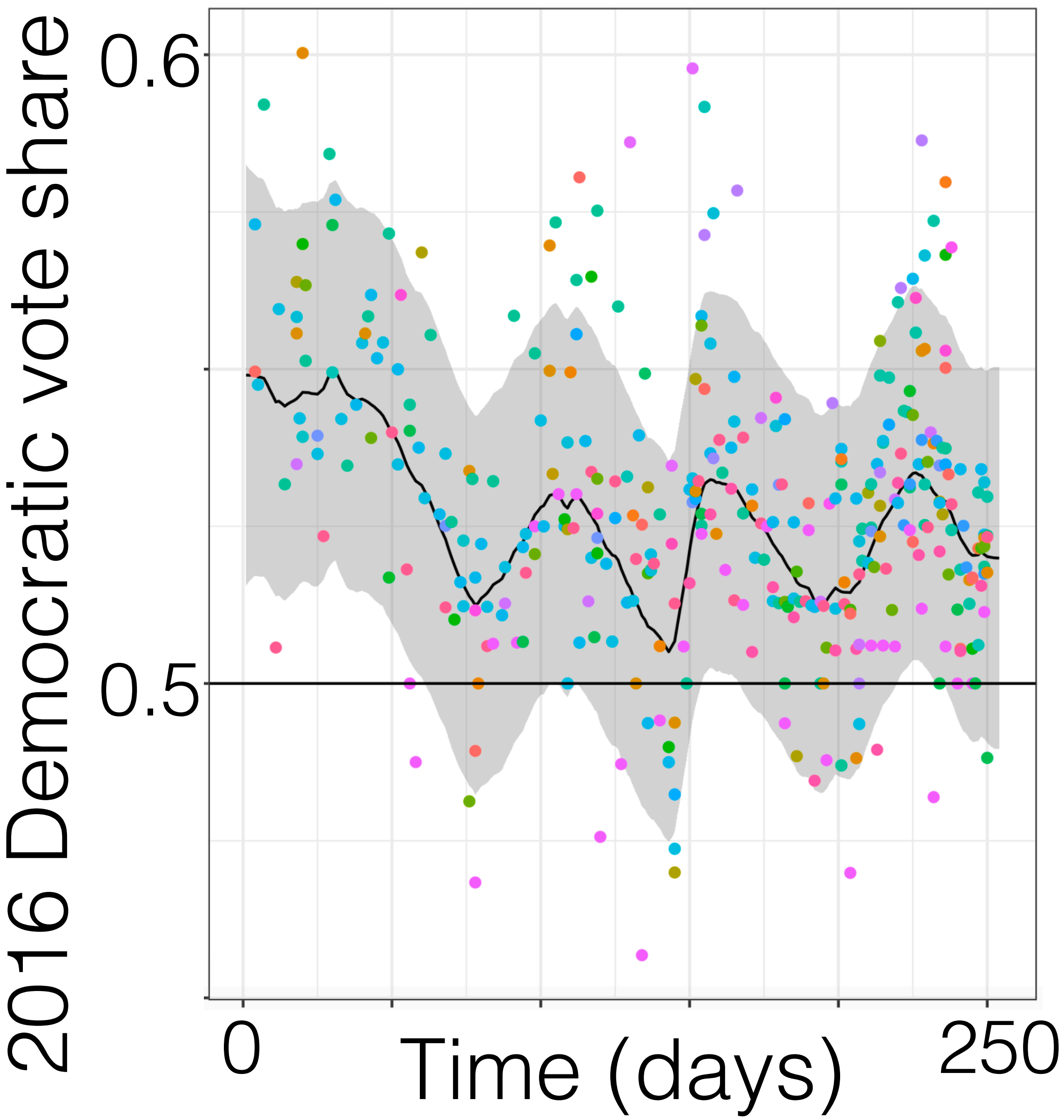
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ

General setup and notation



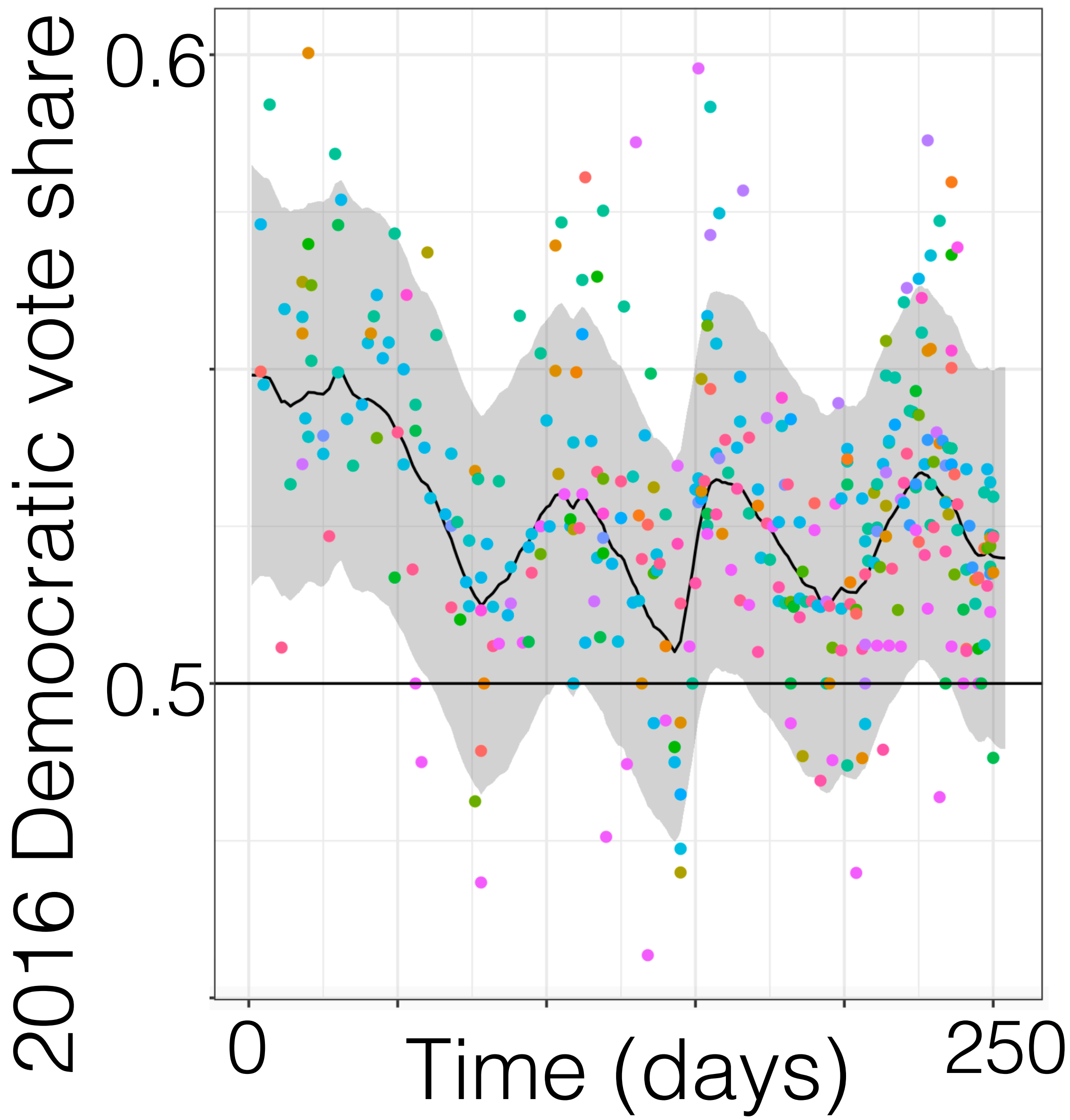
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$

General setup and notation



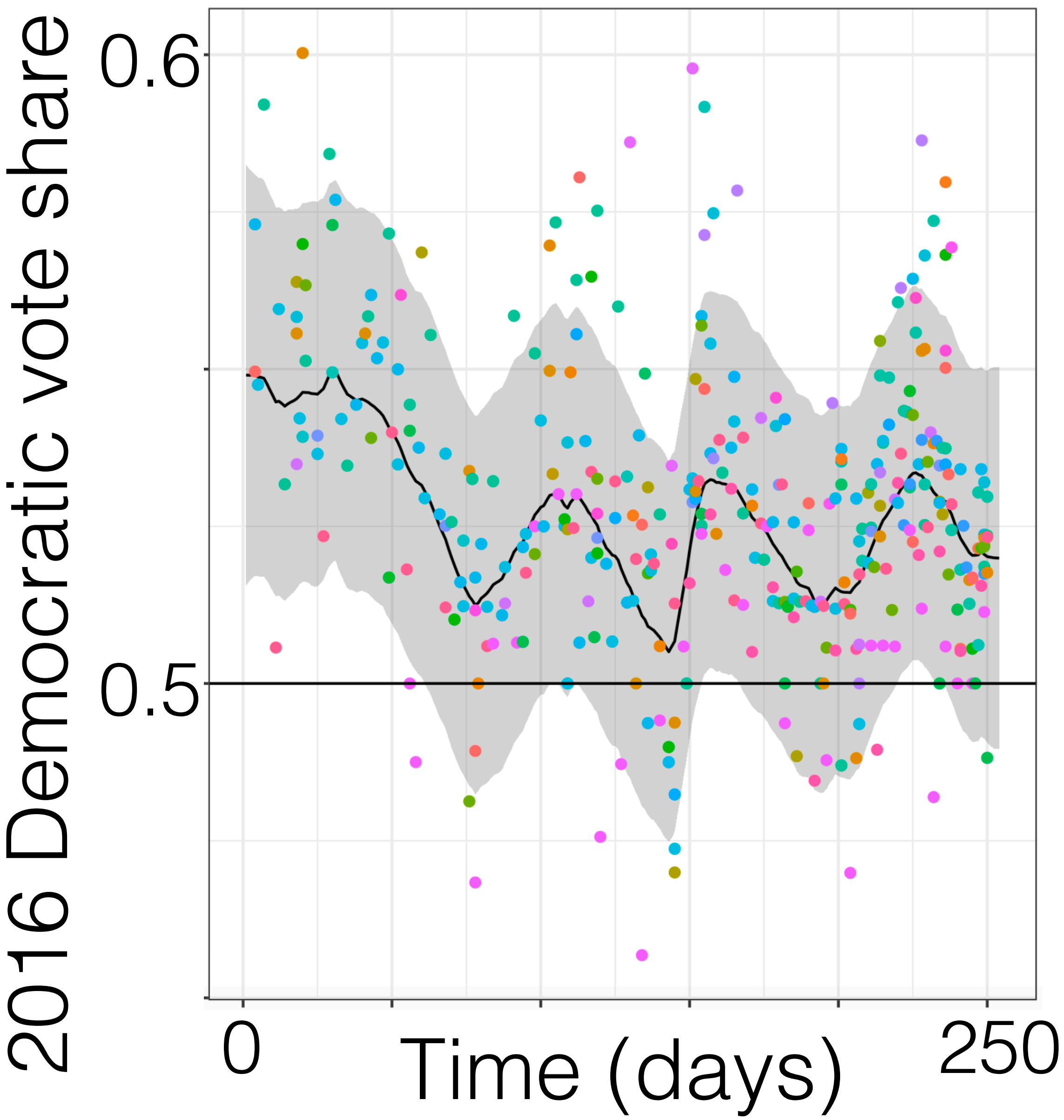
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:

General setup and notation



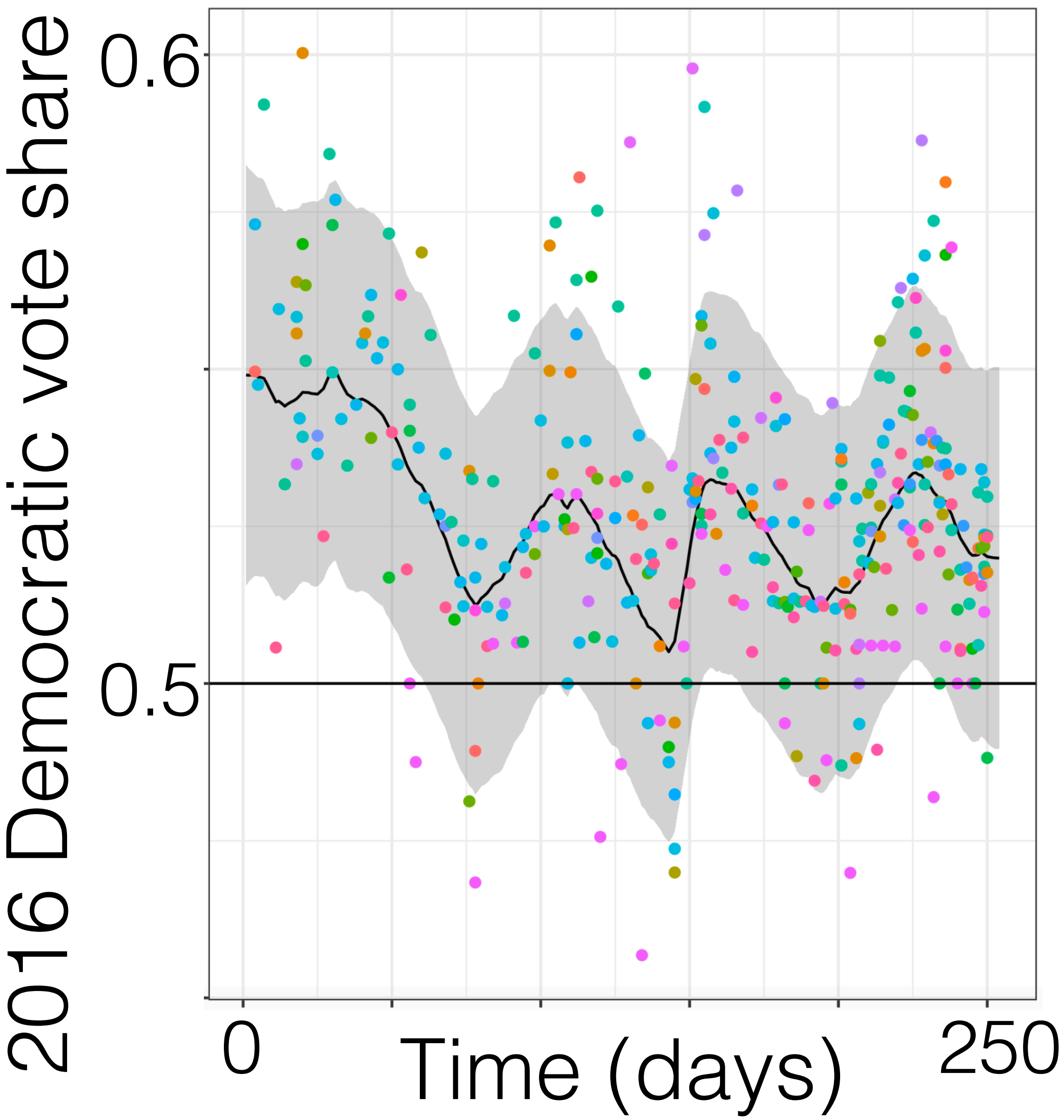
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model

General setup and notation



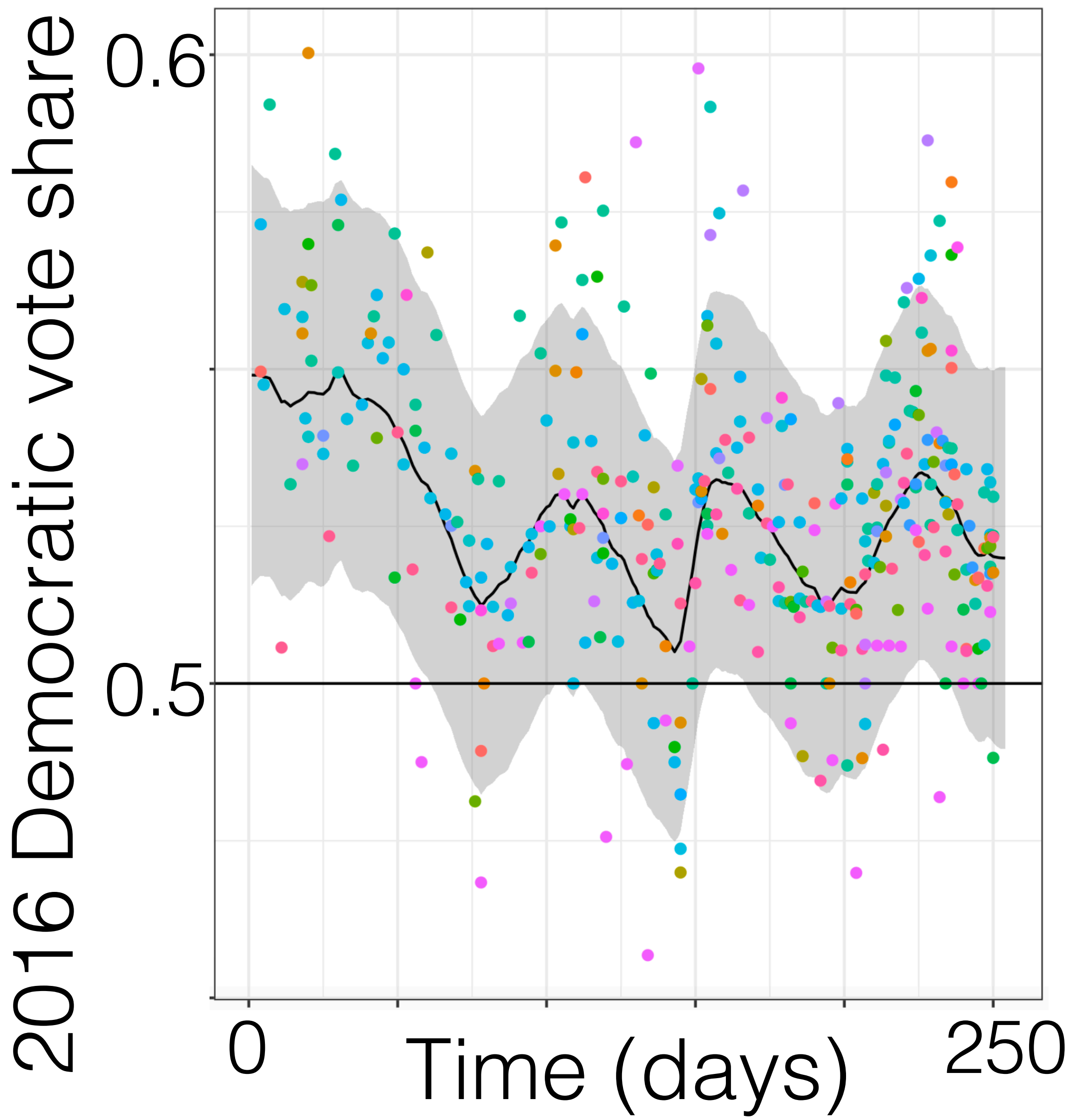
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta)$

General setup and notation



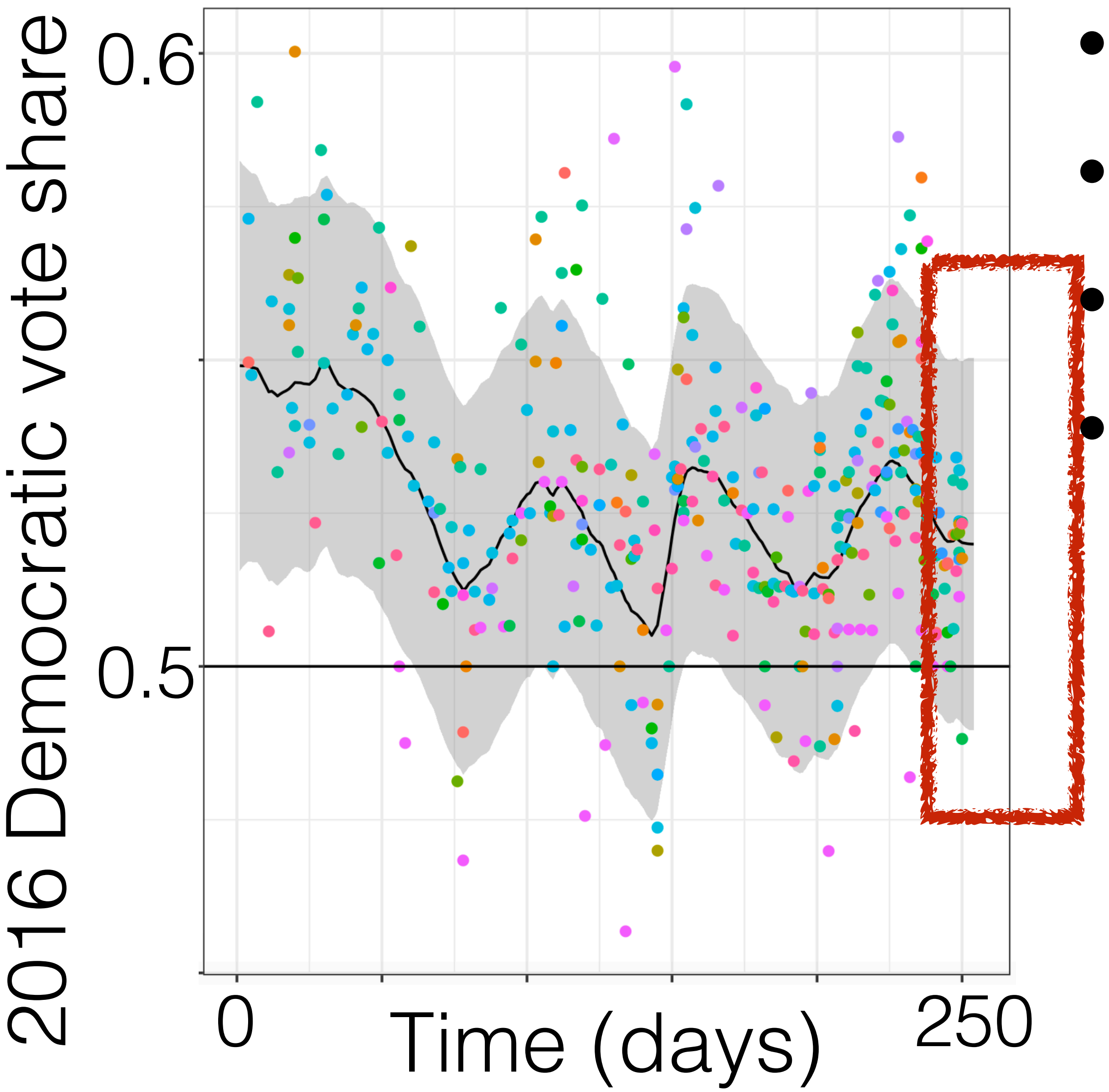
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta)$, $p(\theta)$

General setup and notation



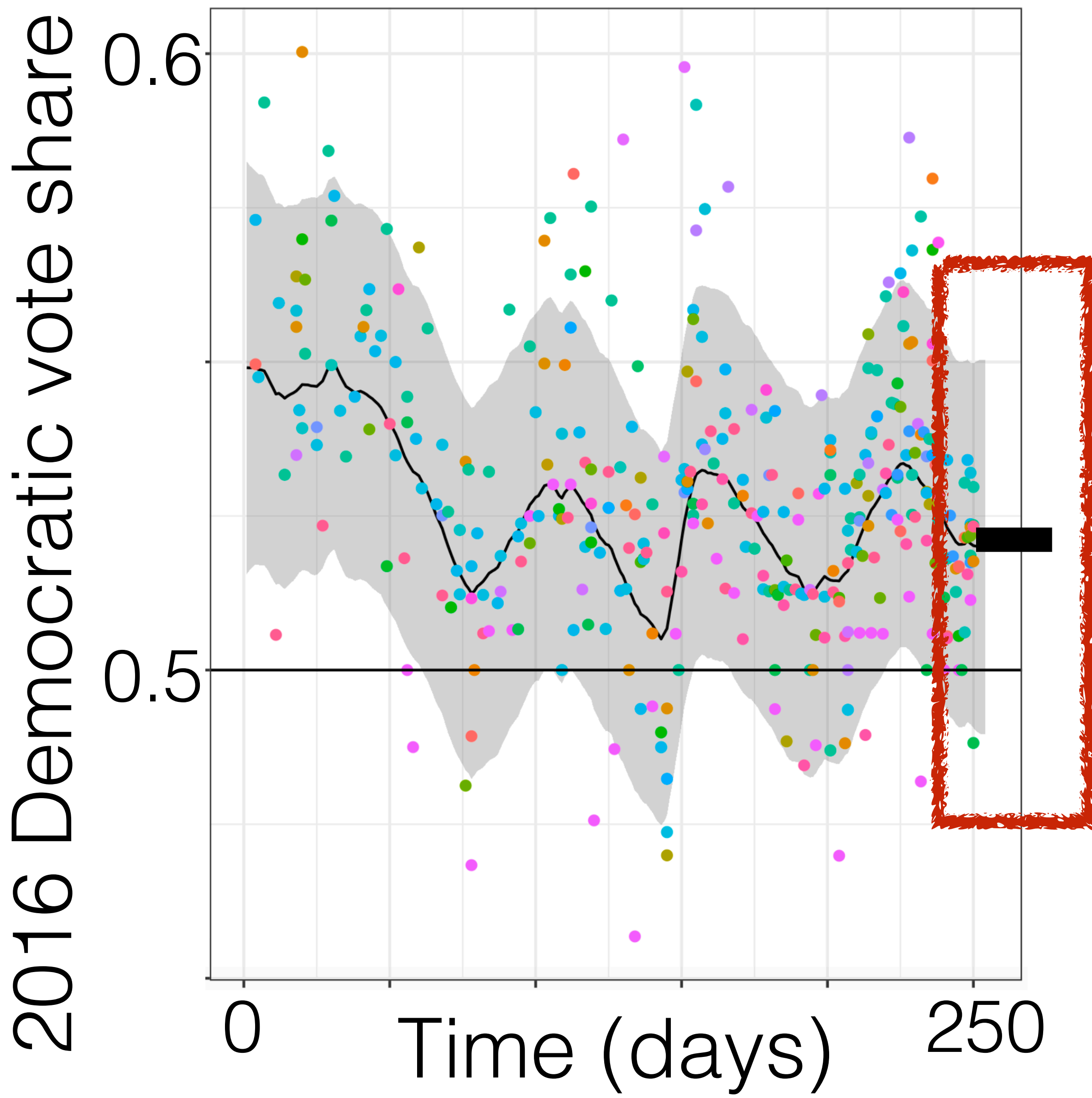
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$

General setup and notation



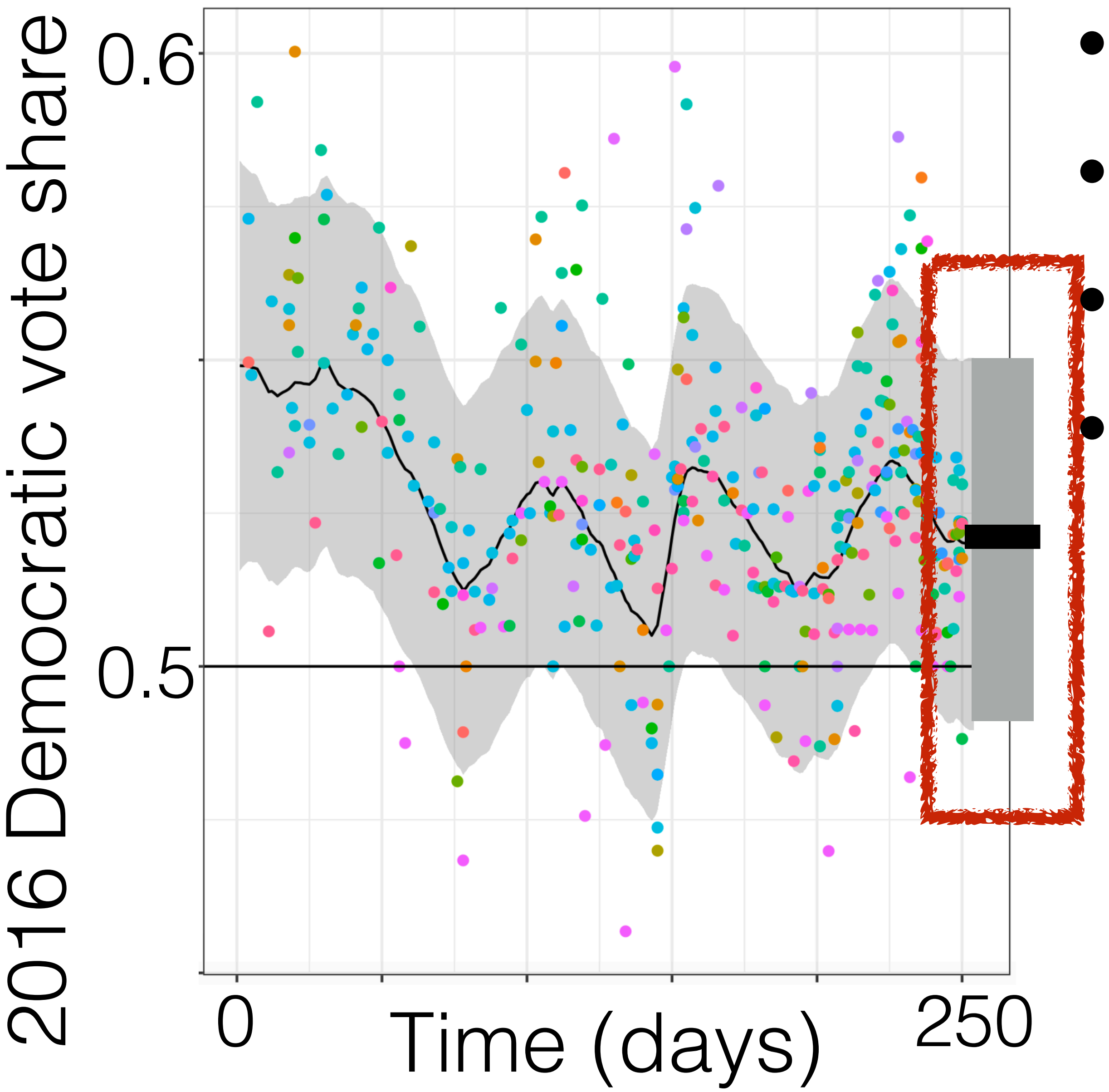
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$

General setup and notation



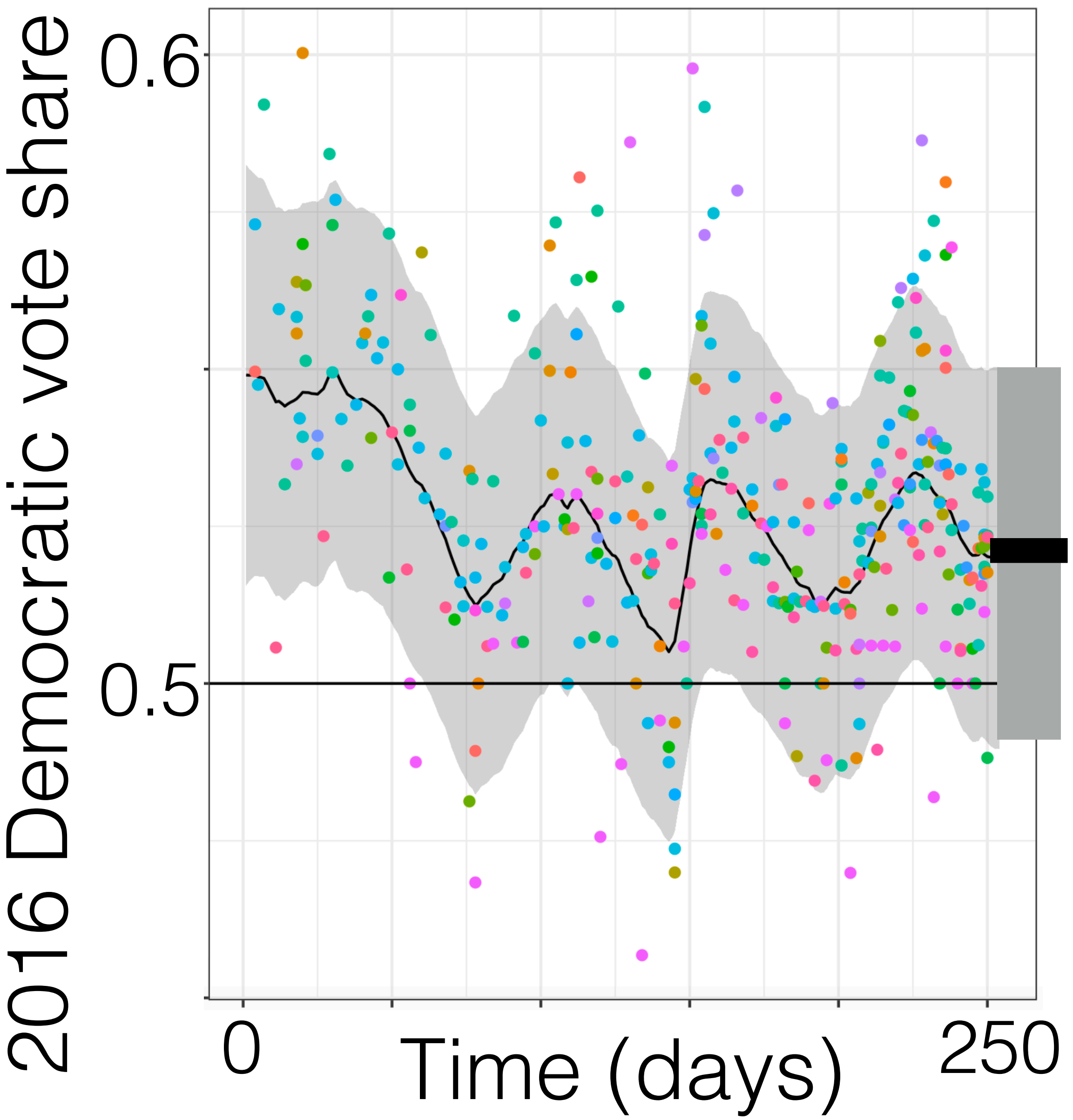
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta), \; p(\theta) \Rightarrow p(\theta|X)$

General setup and notation



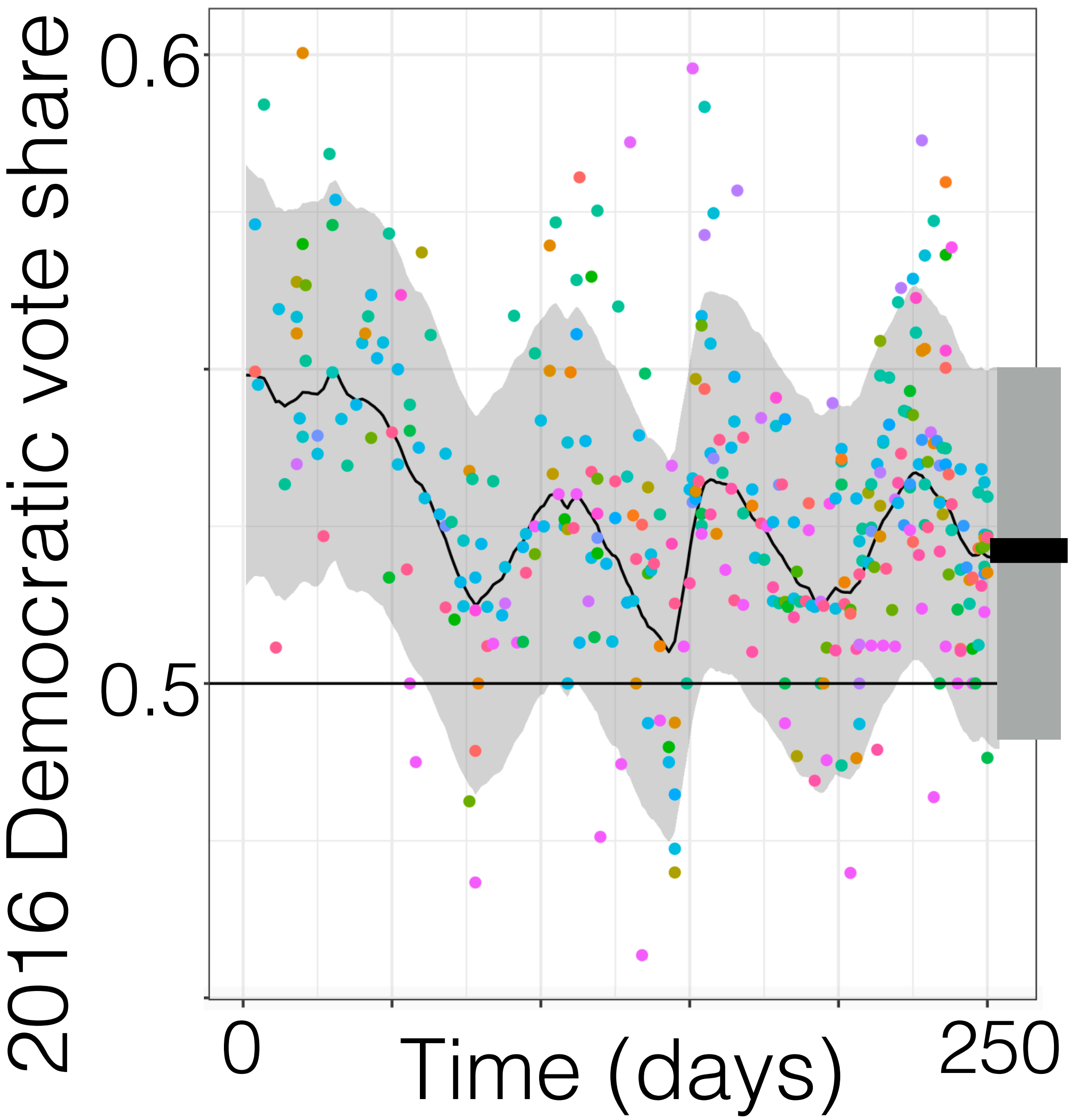
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$

General setup and notation



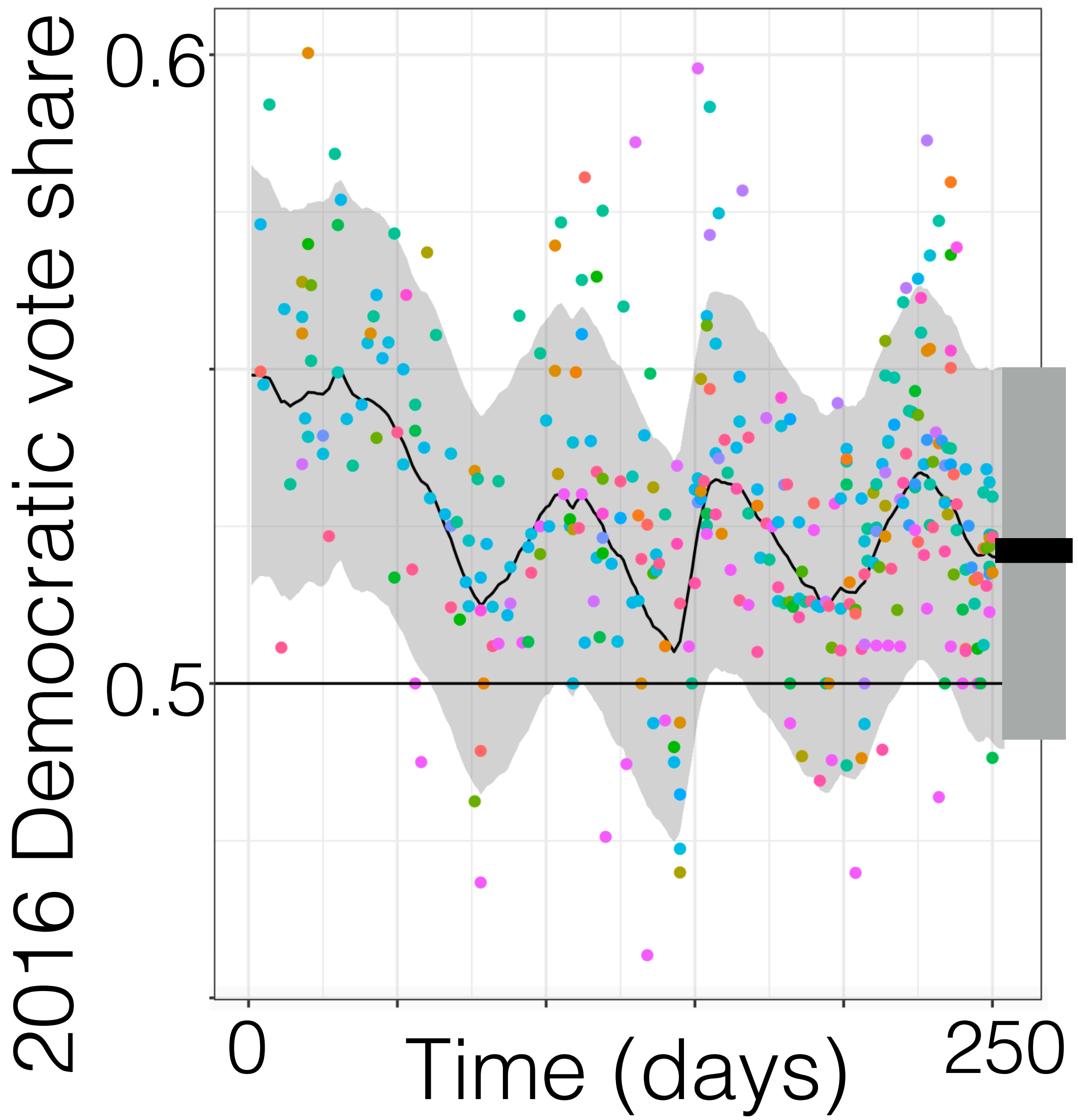
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$

General setup and notation



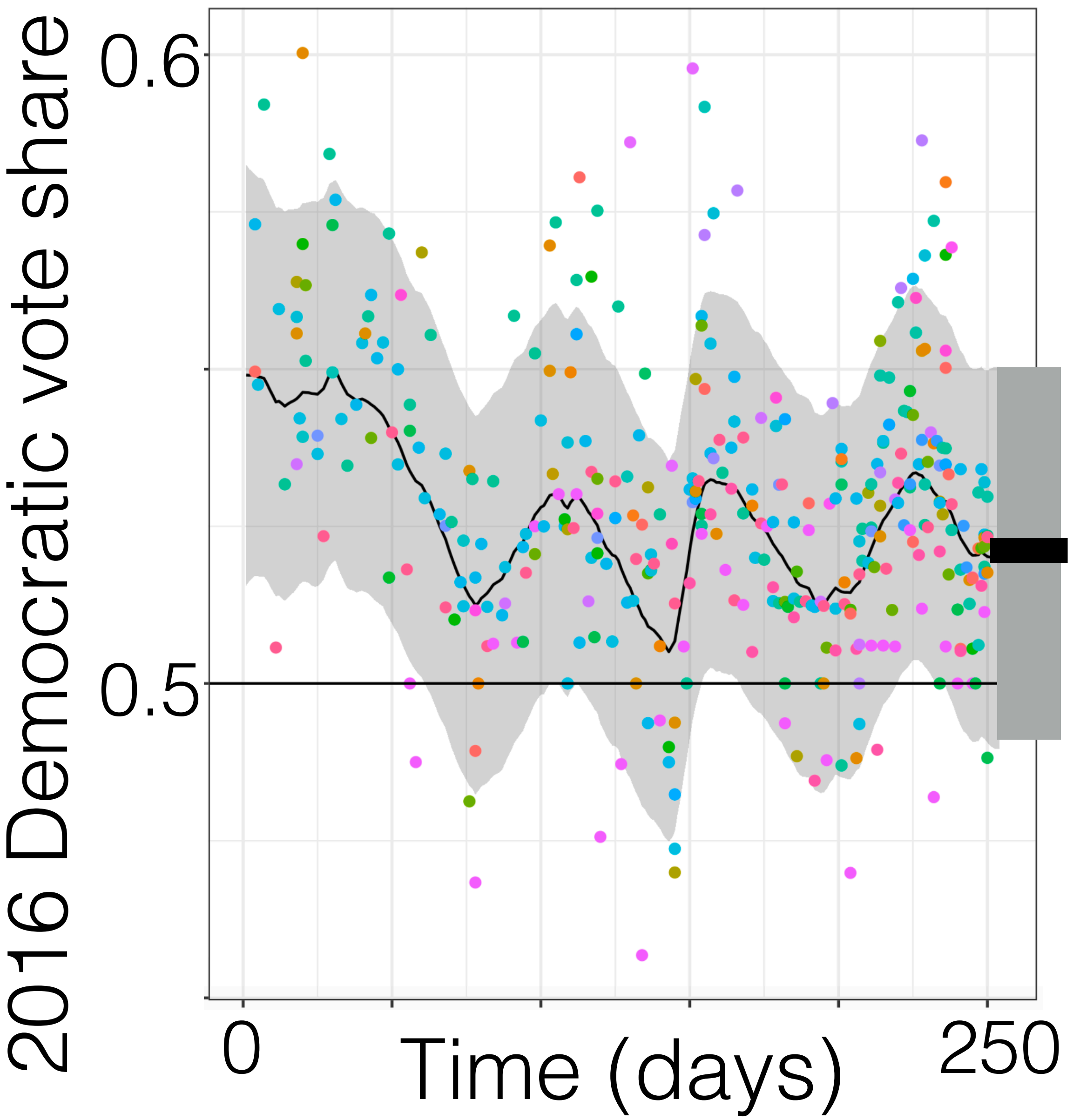
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$

General setup and notation



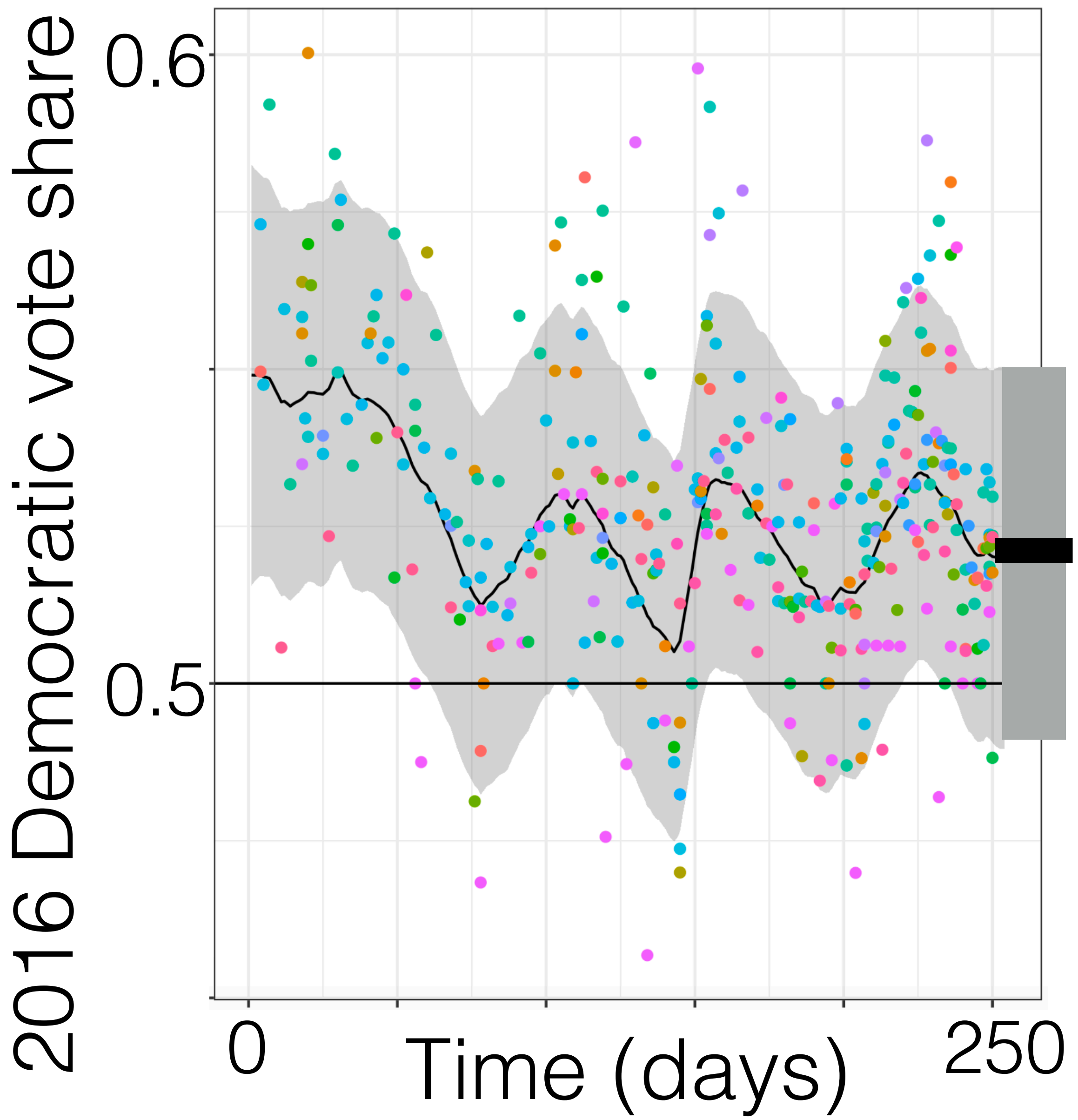
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$
  - Report MCMC approximations

General setup and notation



- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$
  - Report MCMC approximations
- Assumption: our model is misspecified, & $X \sim \mathbb{F}(X)$

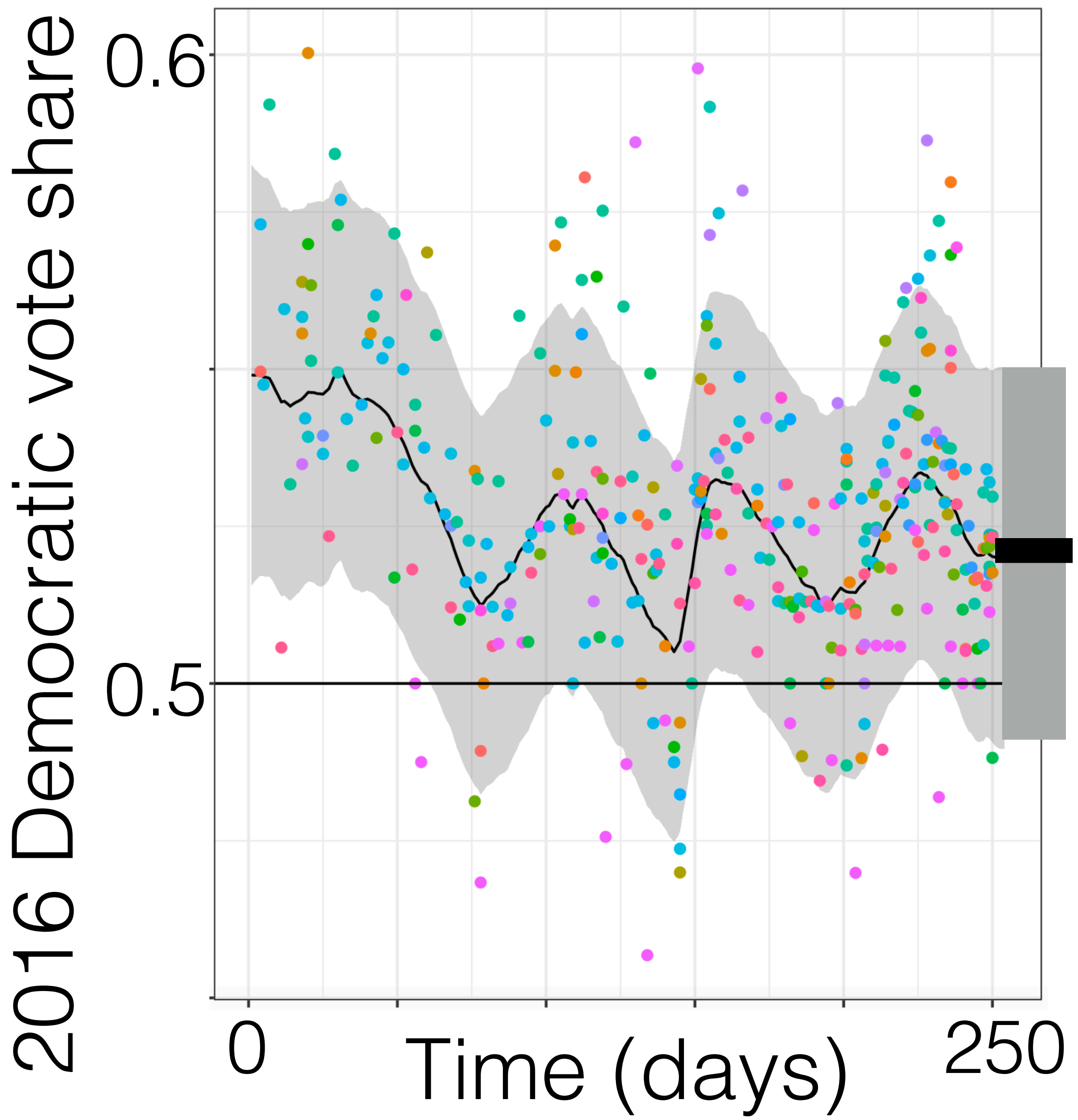
General setup and notation



- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$
  - Report MCMC approximations
- Assumption: our model is misspecified, & $X \sim \mathbb{F}(X)$

- Question: How much does $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)]$ change under different samples from the population?

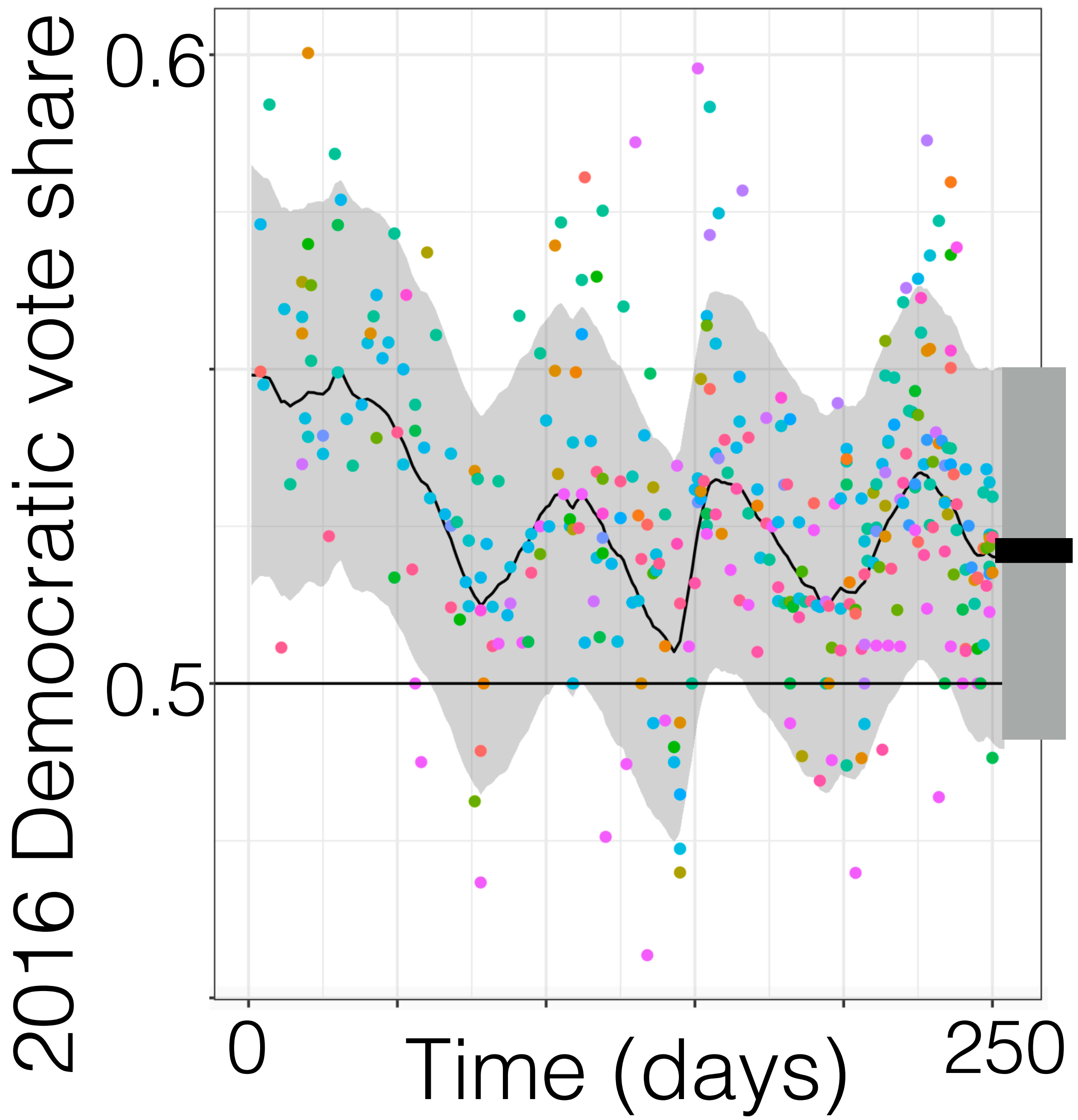
General setup and notation



- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\mathop{\mathbb{E}}_{p(\theta|X)}[g(\theta)],\ \mathop{\mathrm{Var}}_{p(\theta|X)}[g(\theta)]$
  - Report MCMC approximations
- Assumption: our model is misspecified, & $X \sim \mathbb{F}(X)$

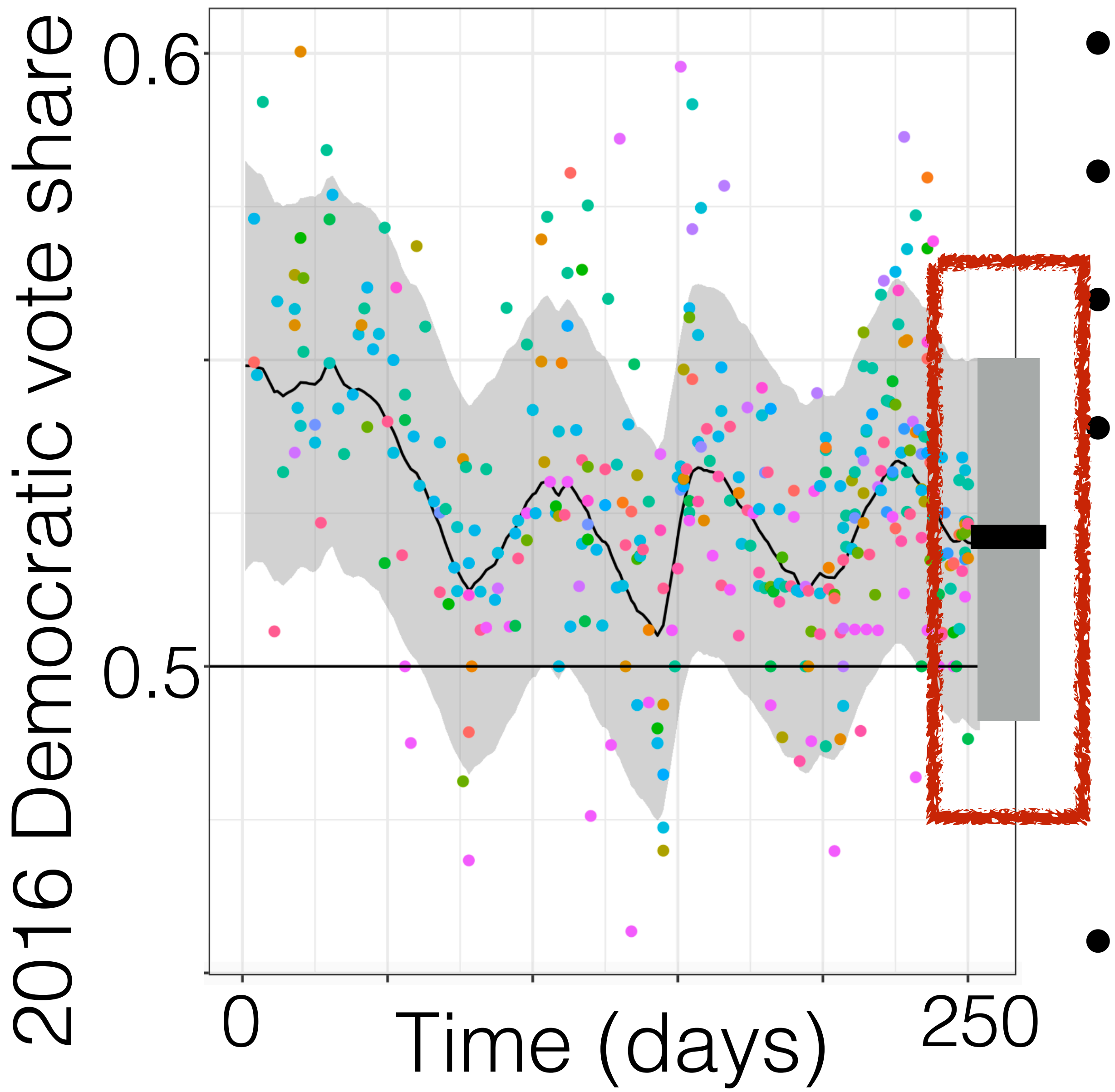
- Question: How much does $\mathop{\mathbb{E}}_{p(\theta|X)}[g(\theta)]$ change under different samples from the population?
  - One summary: $\mathop{\mathrm{Var}}_{\mathbb{F}(X)} \mathop{\mathbb{E}}_{p(\theta|X)}[g(\theta)]$

General setup and notation



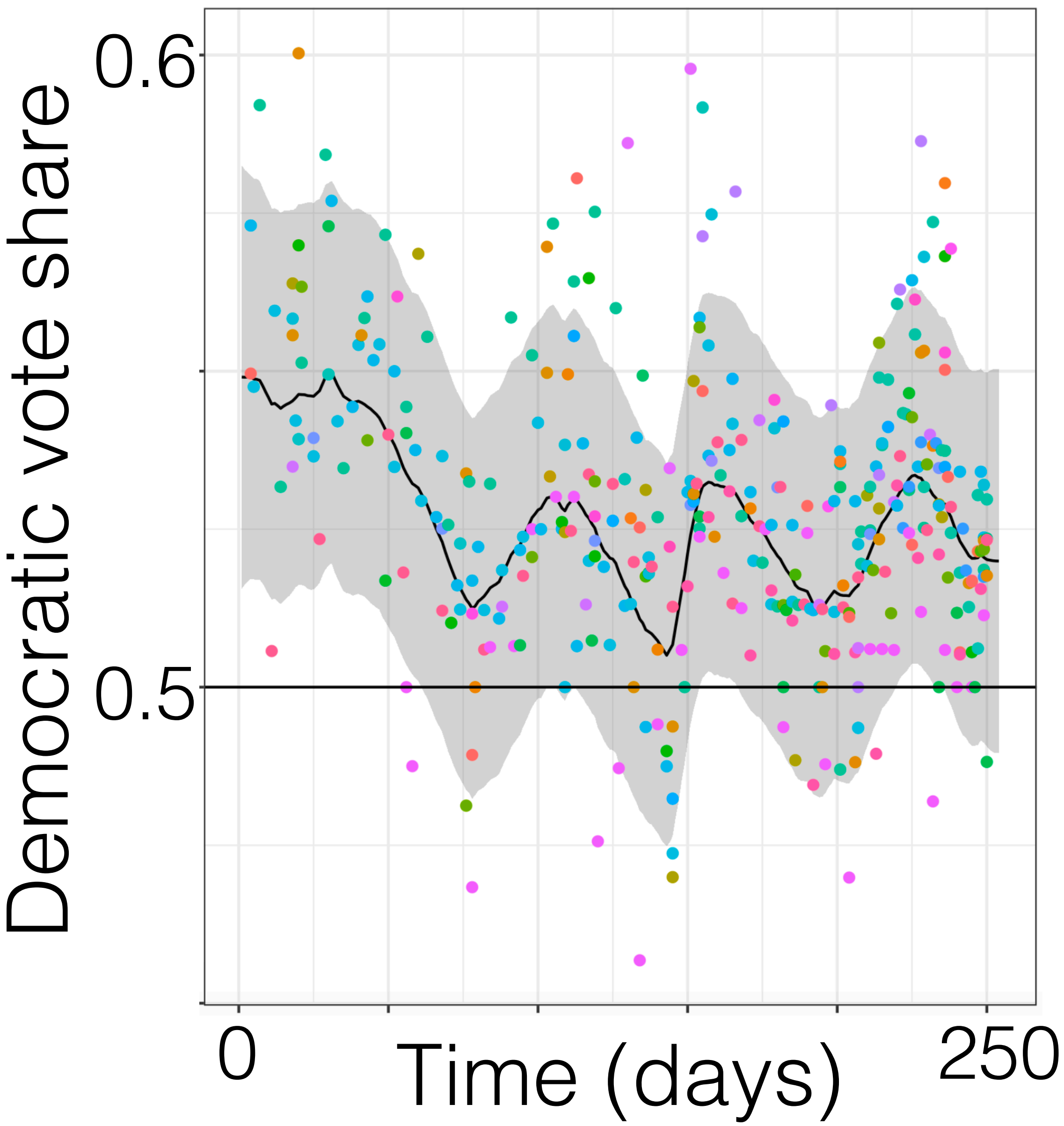
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$
  - Report MCMC approximations
- Assumption: our model is misspecified, & $X \sim \mathbb{F}(X)$
- Question: How much does $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)]$ change under different samples from the population?
  - One summary: $\underset{\mathbb{F}(X)}{\mathrm{Var}}\ \underset{p(\theta|X)}{\mathbb{E}}[g(\theta)]$
- If frequentist uncertainty exceeds $\underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$, say non-robust

General setup and notation



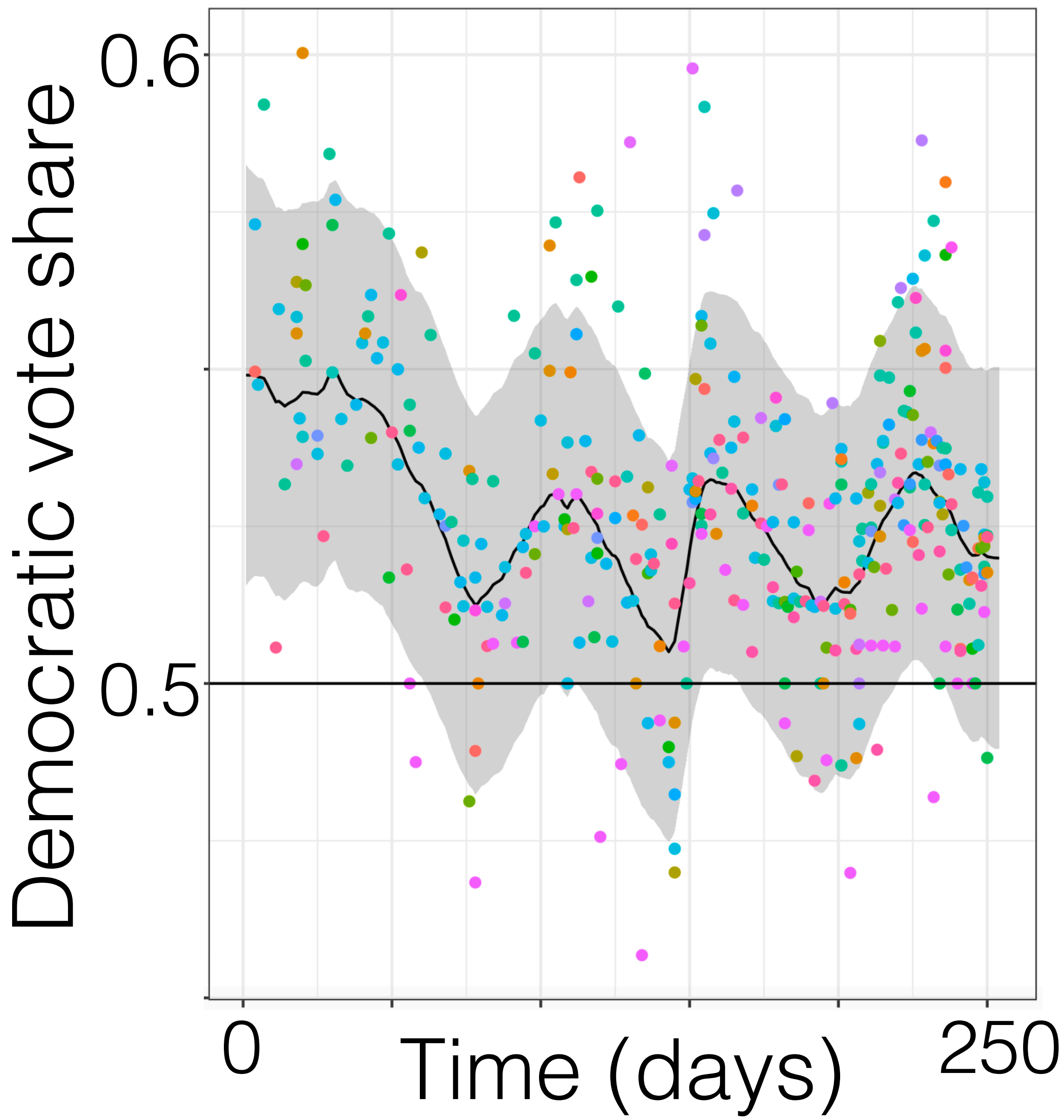
- Data: $X = (x_1, x_2, \ldots, x_N)$
- Parameters: θ
- Quantity of interest: $g(\theta)$
- A common Bayesian approach:
  - Model $p(X|\theta),\ p(\theta) \Rightarrow p(\theta|X)$
  - Want $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)],\ \underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$
  - Report MCMC approximations
- Assumption: our model is misspecified, & $X \sim \mathbb{F}(X)$

- Question: How much does $\underset{p(\theta|X)}{\mathbb{E}}[g(\theta)]$ change under different samples from the population?
  - One summary: $\underset{\mathbb{F}(X)}{\mathrm{Var}}\ \underset{p(\theta|X)}{\mathbb{E}}[g(\theta)]$
- If frequentist uncertainty exceeds $\underset{p(\theta|X)}{\mathrm{Var}}[g(\theta)]$, say non-robust



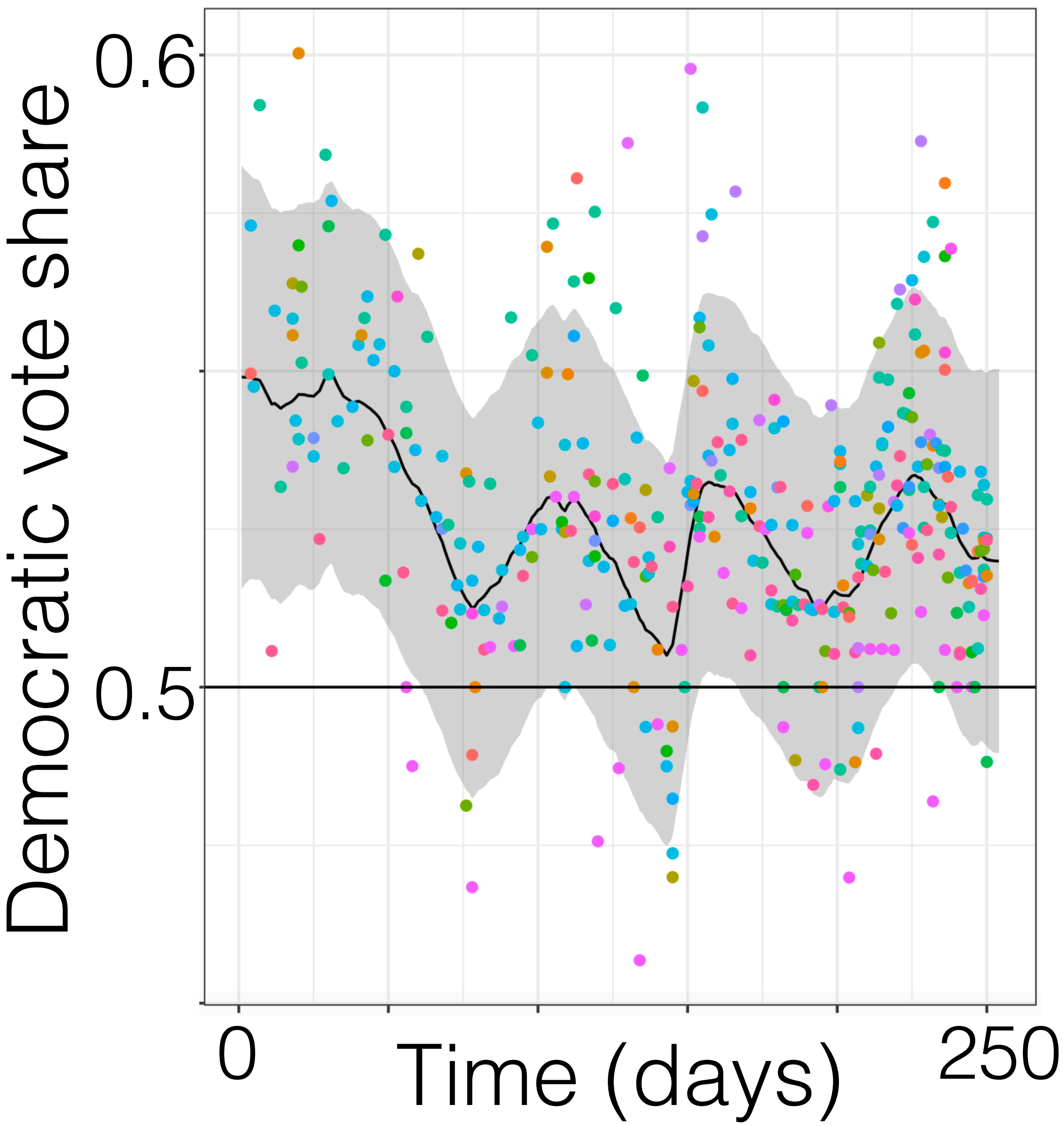
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

Idea: how about the bootstrap?



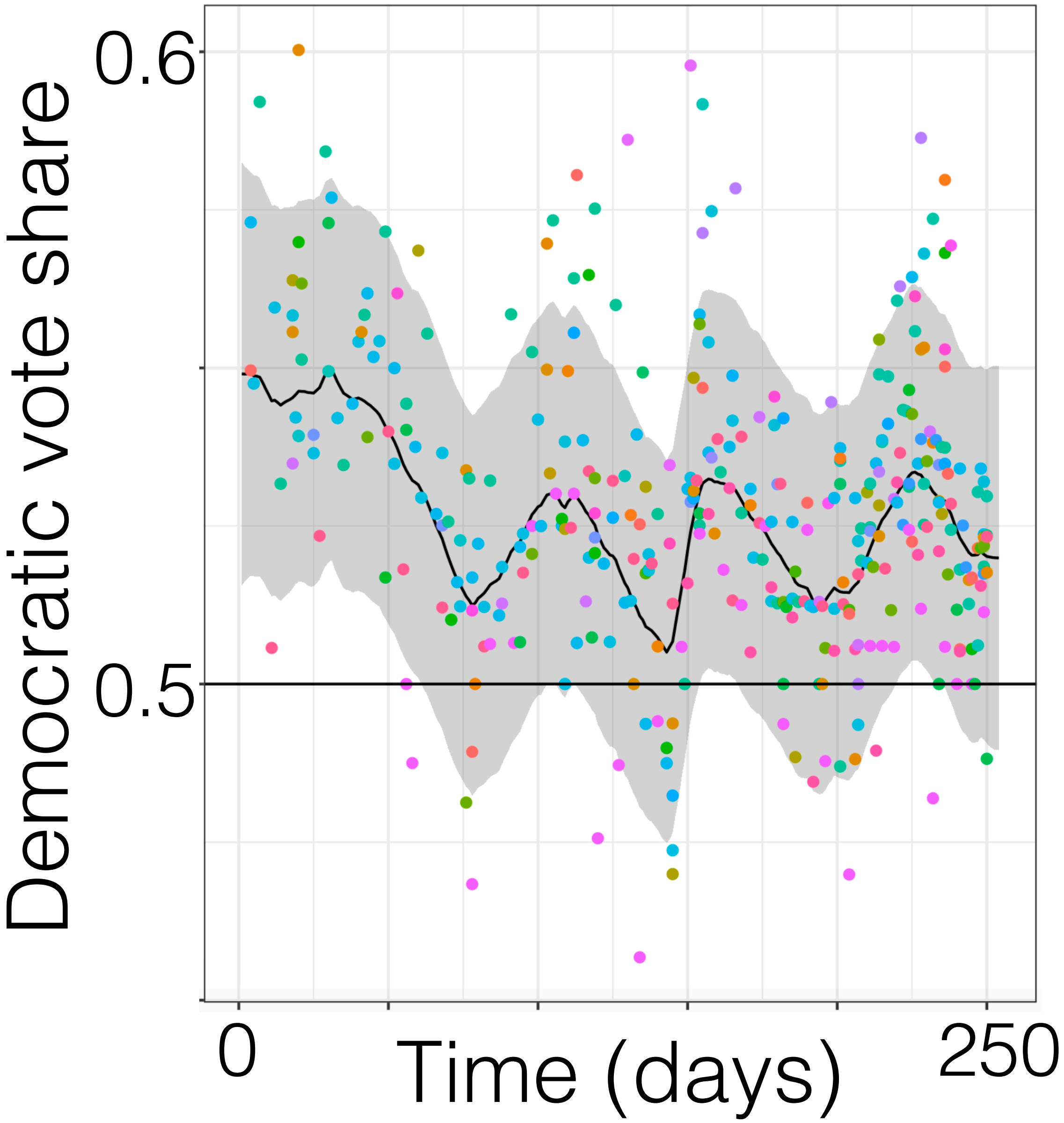
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

Idea: how about the bootstrap?



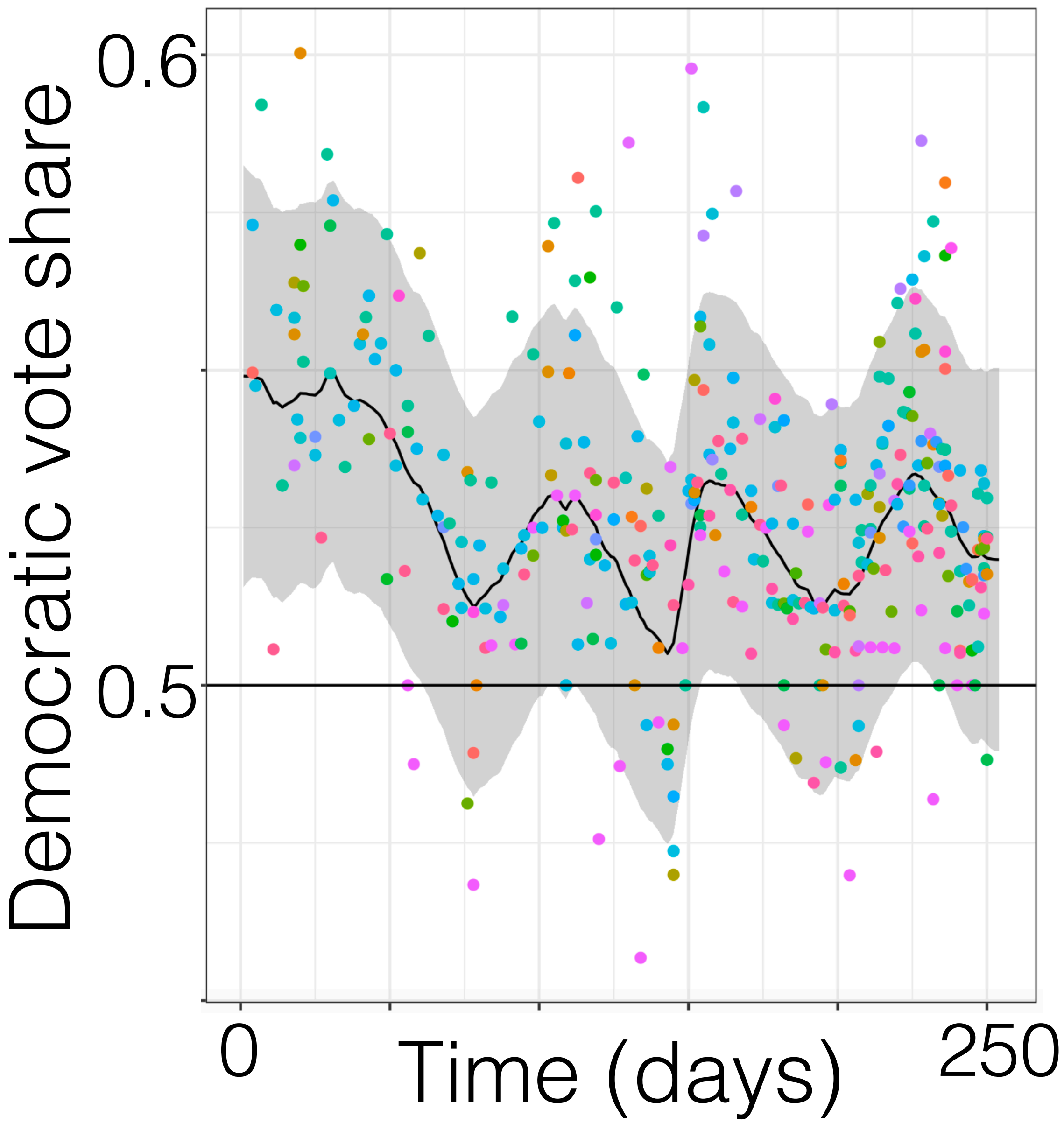
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019

Idea: how about the bootstrap?



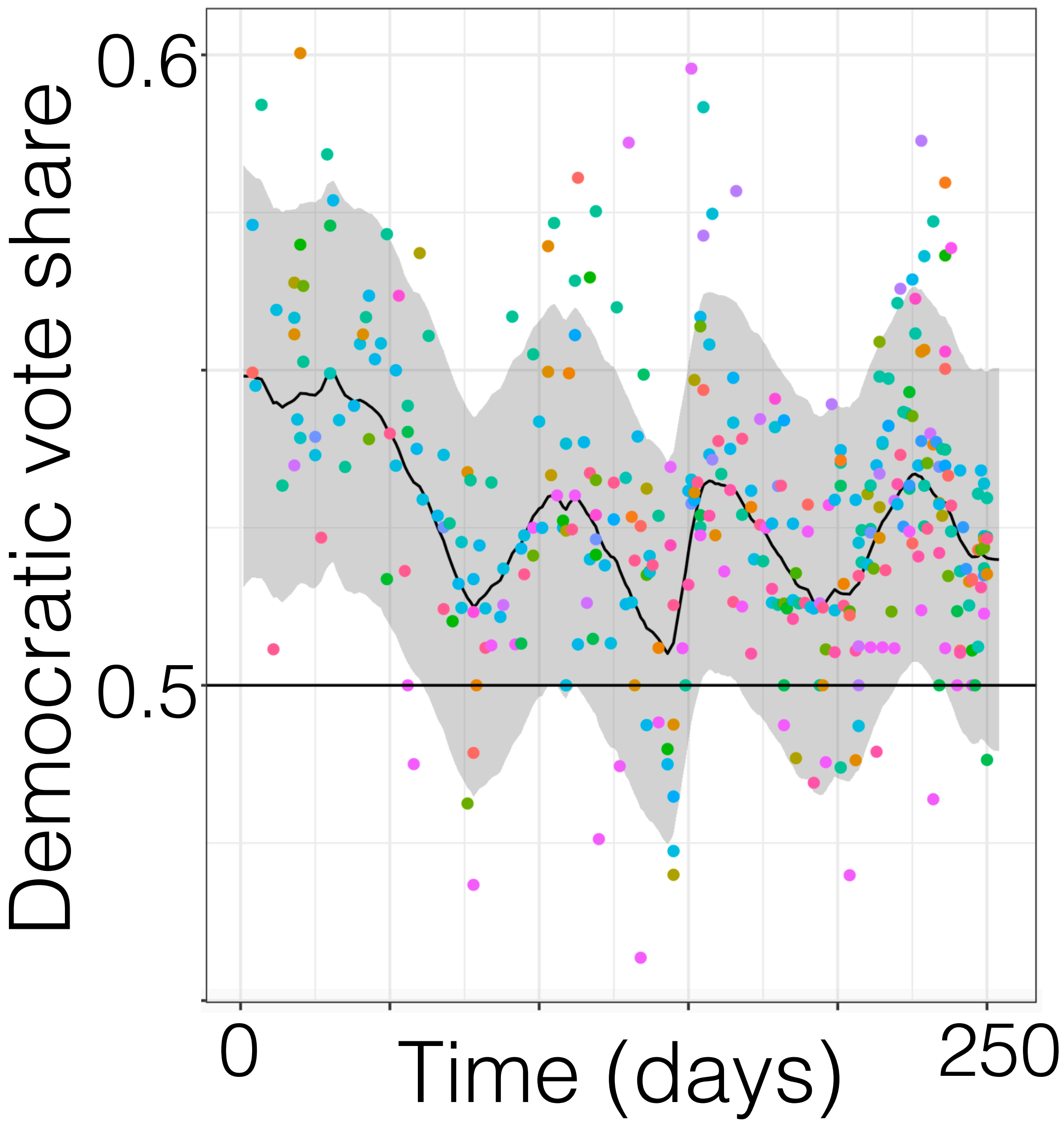
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:

Idea: how about the bootstrap?



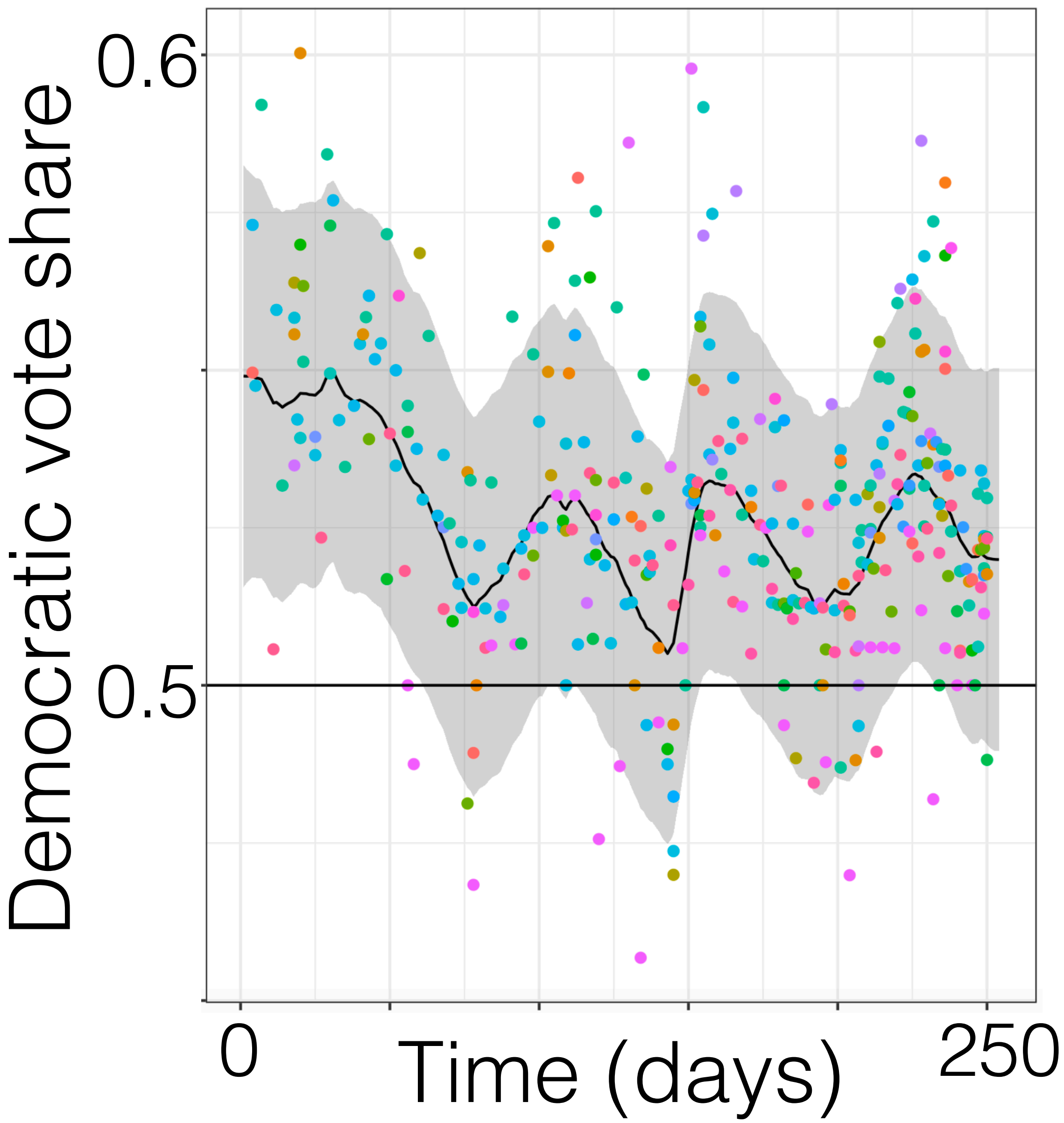
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)

Idea: how about the bootstrap?



- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X

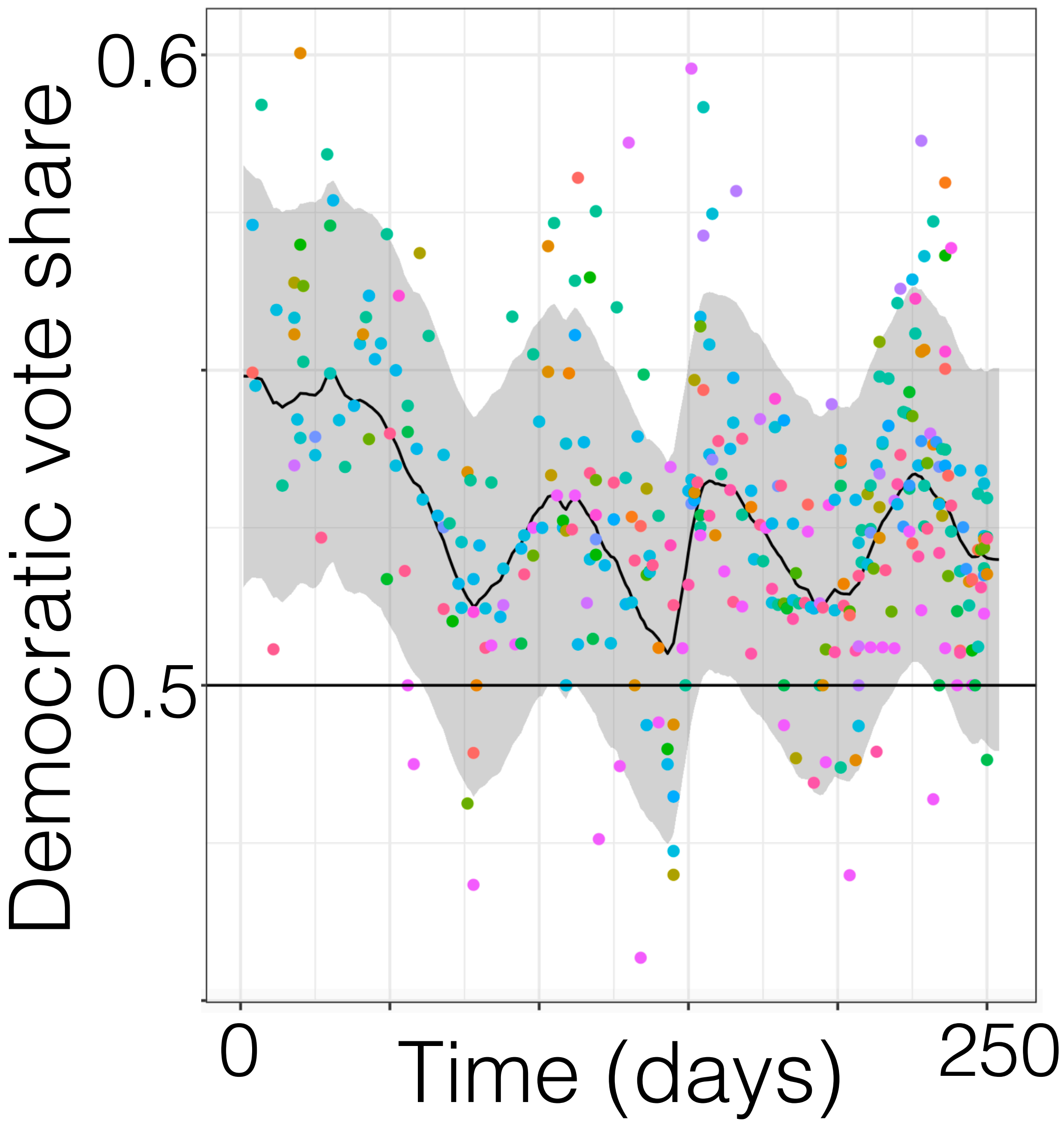
Idea: how about the bootstrap?



- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X

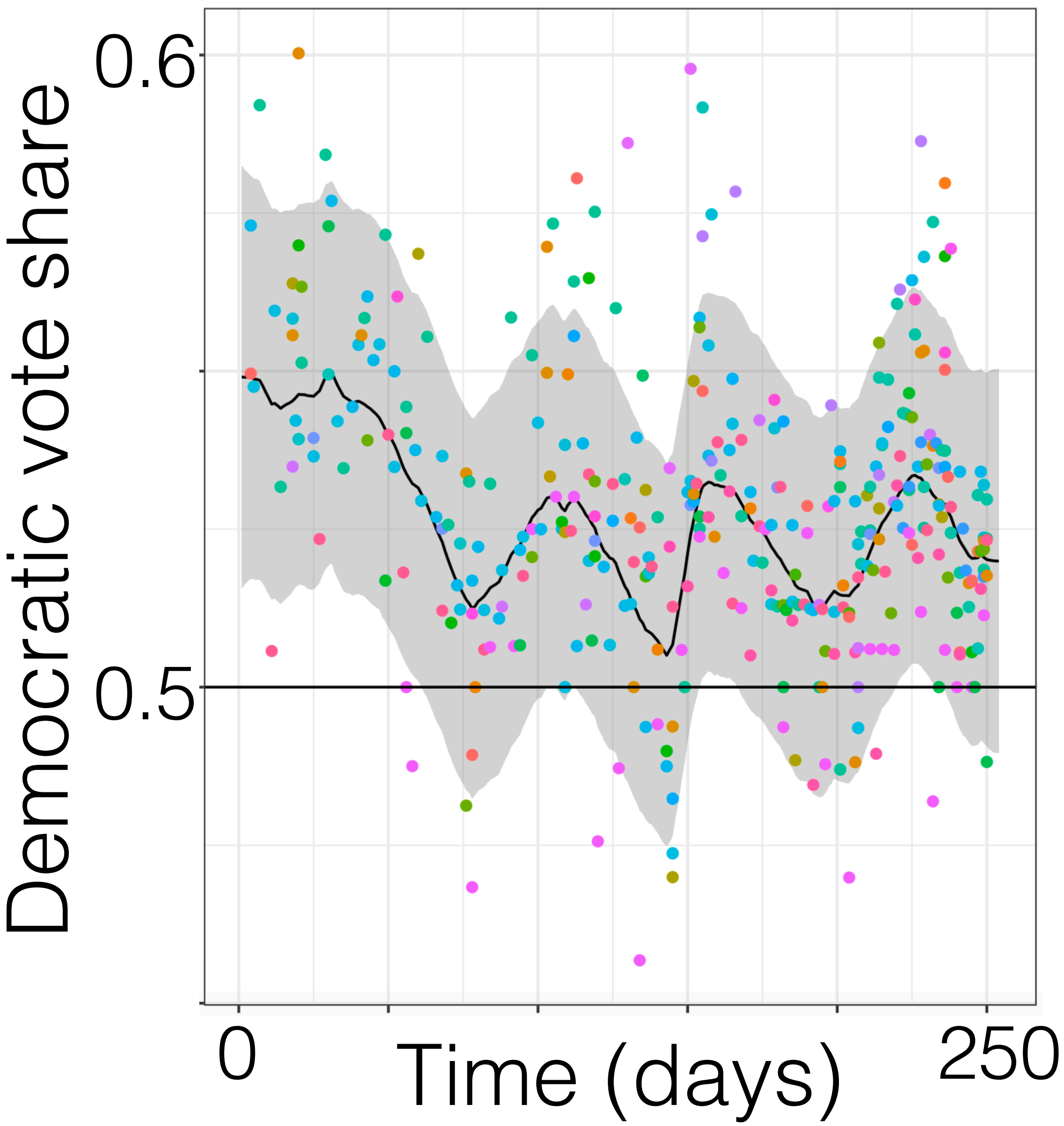
$$\underset{p(\theta|X_b^*)}{\mathbb{E}} [g(\theta)]$$

Idea: how about the bootstrap?



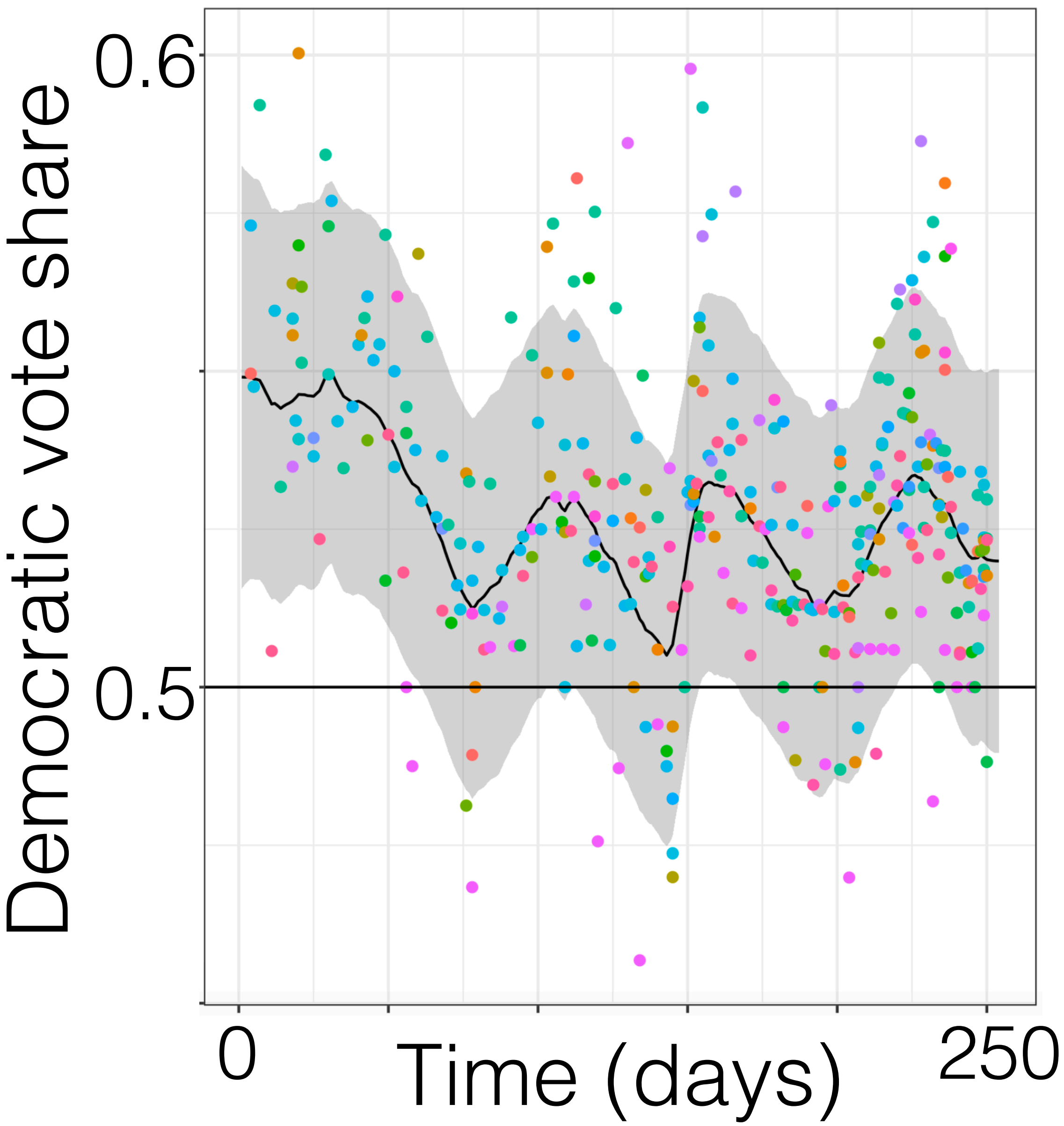
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X
    - Run MCMC to get $\underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [g(\theta)]$

Idea: how about the bootstrap?



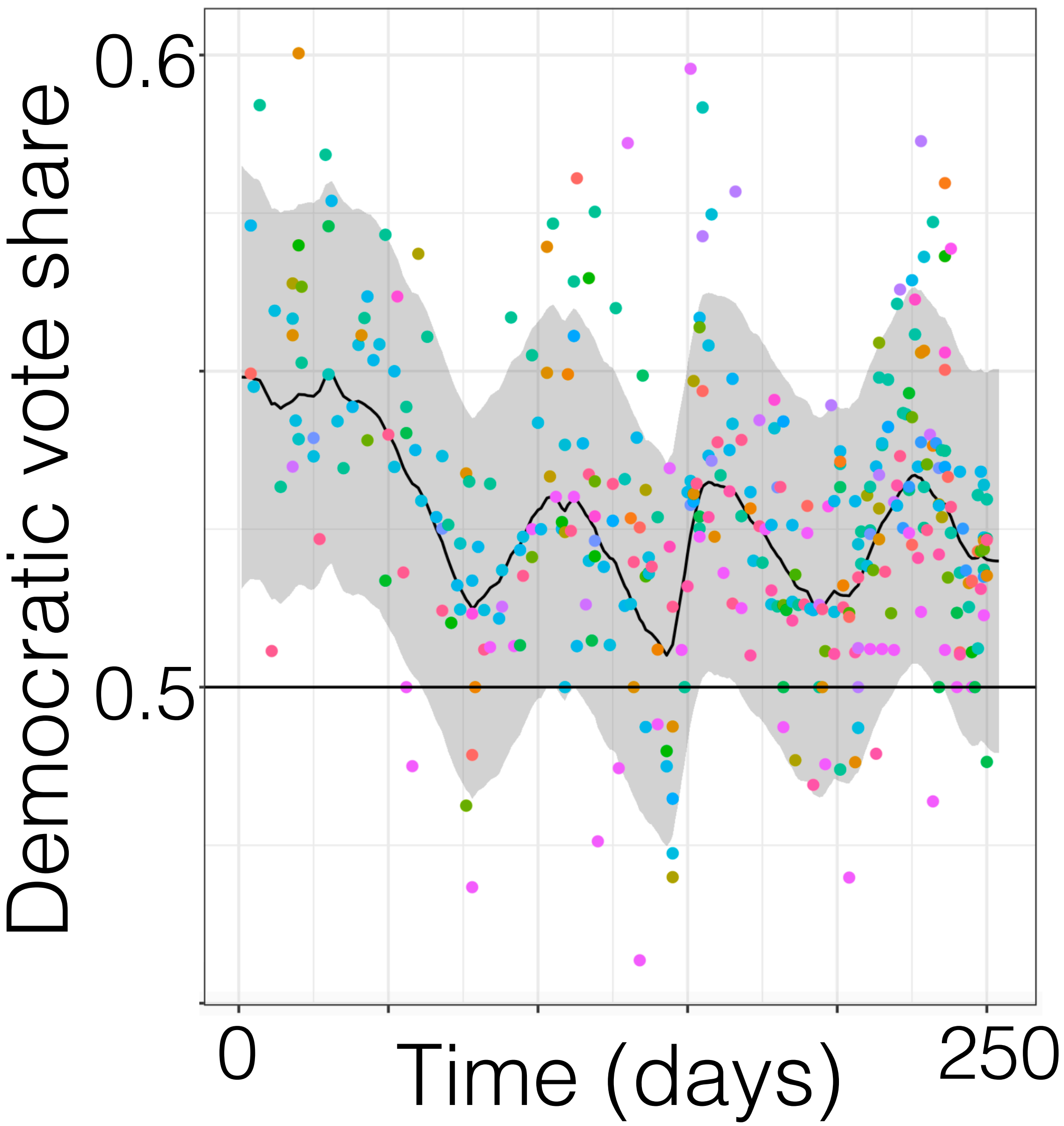
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X
    - Run MCMC to get $\underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [g(\theta)]$

Idea: how about the bootstrap?



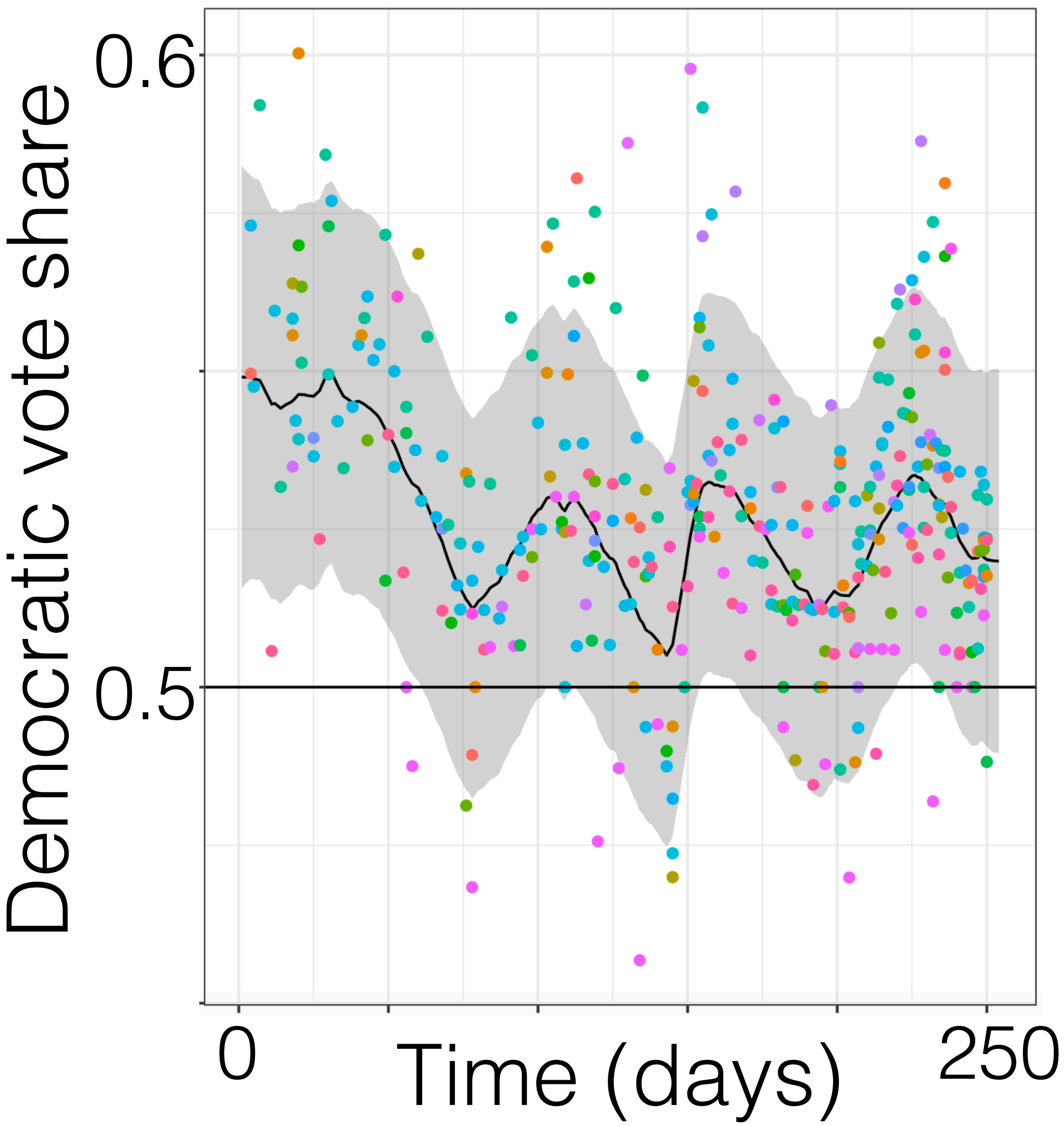
- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X
    - Run MCMC to get $\underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [g(\theta)]$
  - Report $\underset{b}{\widehat{\mathrm{Var}}} \; \underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [\theta]$

Idea: how about the bootstrap?



- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X
    - Run MCMC to get $\underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [g(\theta)]$
  - Report $\underset{b}{\widehat{\mathrm{Var}}} \underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [\theta]$
- Problem: computational cost

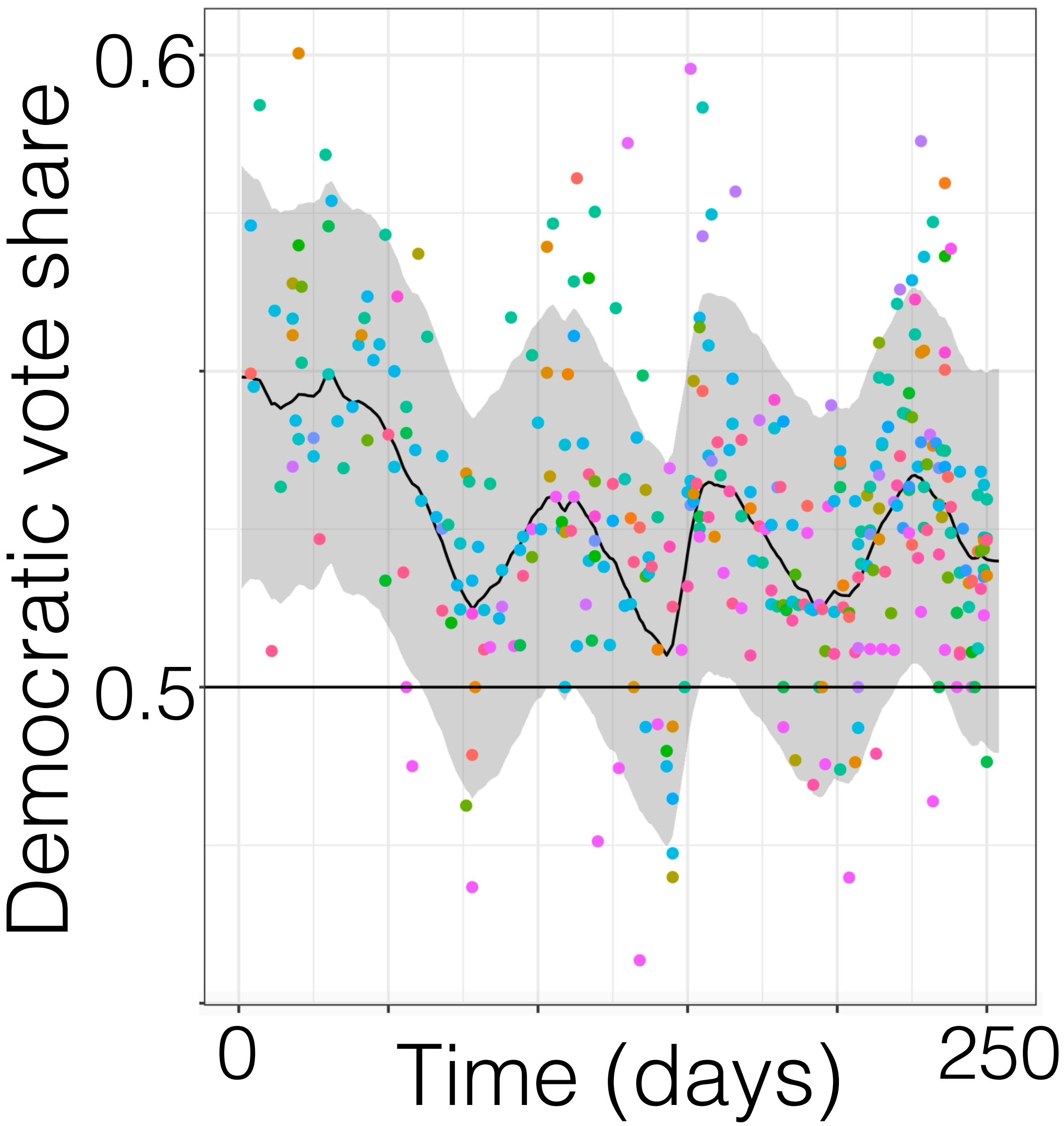
Idea: how about the bootstrap?



- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X
    - Run MCMC to get $\underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [g(\theta)]$
  - Report $\underset{b}{\widehat{\mathrm{Var}}} \; \underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [\theta]$

- Problem: computational cost
- In the election example, each MCMC run takes on average ~12.3 hours (Stan language)

Idea: how about the bootstrap?



- How to approx $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Check out Huggins & Miller 2019
- Algorithm:
  - For $b = 1, \ldots, B$ (e.g. B = 100)
    - Draw X_b^* with replacement from X
    - Run MCMC to get $\underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [g(\theta)]$
  - Report $\underset{b}{\widehat{\mathrm{Var}}} \underset{p(\theta|X_b^*)}{\widehat{\mathbb{E}}} [\theta]$

- Problem: computational cost
- In the election example, each MCMC run takes on average ~12.3 hours (Stan language)
- Bootstrap takes ~51 days to run

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
    - Approximate posterior mean with the MAP (no MCMC)
    - Then use the frequentist sandwich covariance

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
    - Approximate posterior mean with the MAP (no MCMC)
    - Then use the frequentist sandwich covariance
    - Use the delta method to account for *g*

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off, e.g. in (common) models with “global” and “local” parameters.

Idea: sandwich covariance

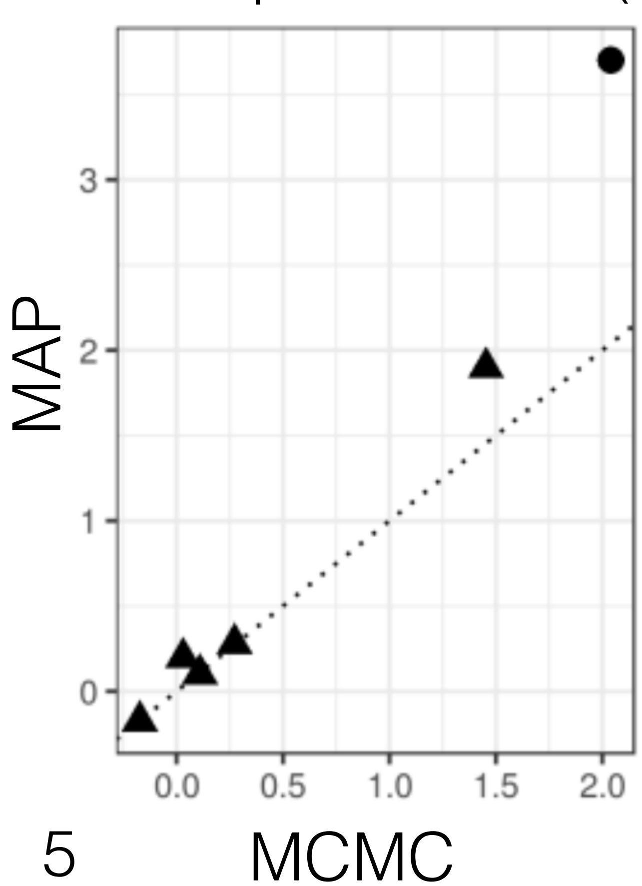
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off, e.g. in (common) models with "global" and "local" parameters. Example: logistic generalized linear mixed model for online ads (Criteo data)

[Giordano, Broderick, Jordan 2018]

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off, e.g. in (common) models with "global" and "local" parameters. Example: logistic generalized linear mixed model for online ads (Criteo data)

Global parameters (-τ)

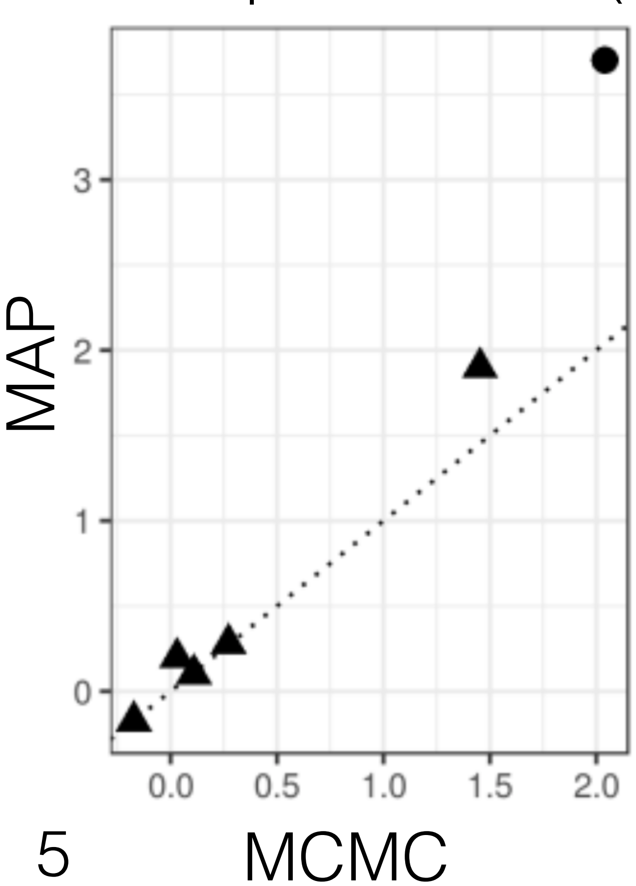


[Giordano, Broderick, Jordan 2018]

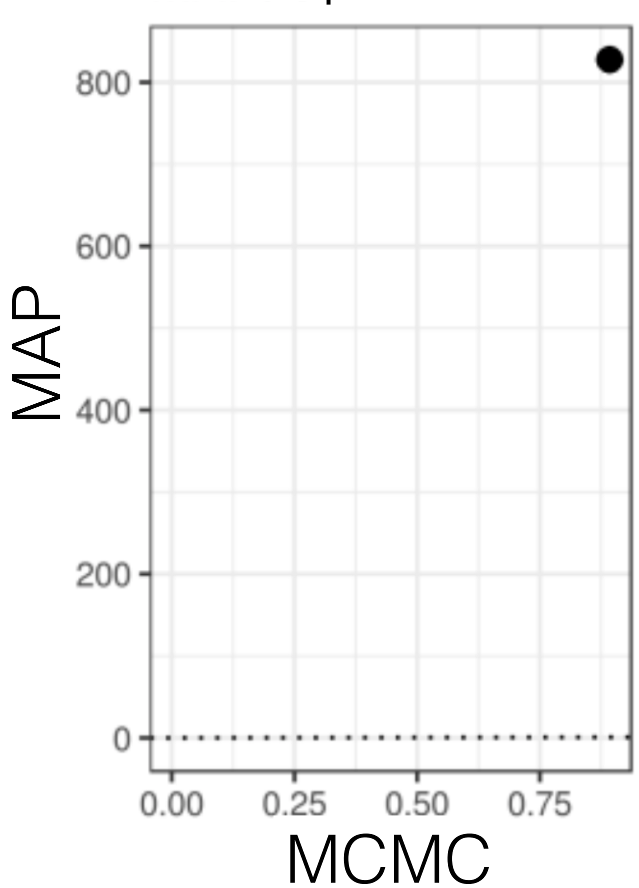
Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off, e.g. in (common) models with "global" and "local" parameters. Example: logistic generalized linear mixed model for online ads (Criteo data)

Global parameters (-τ)



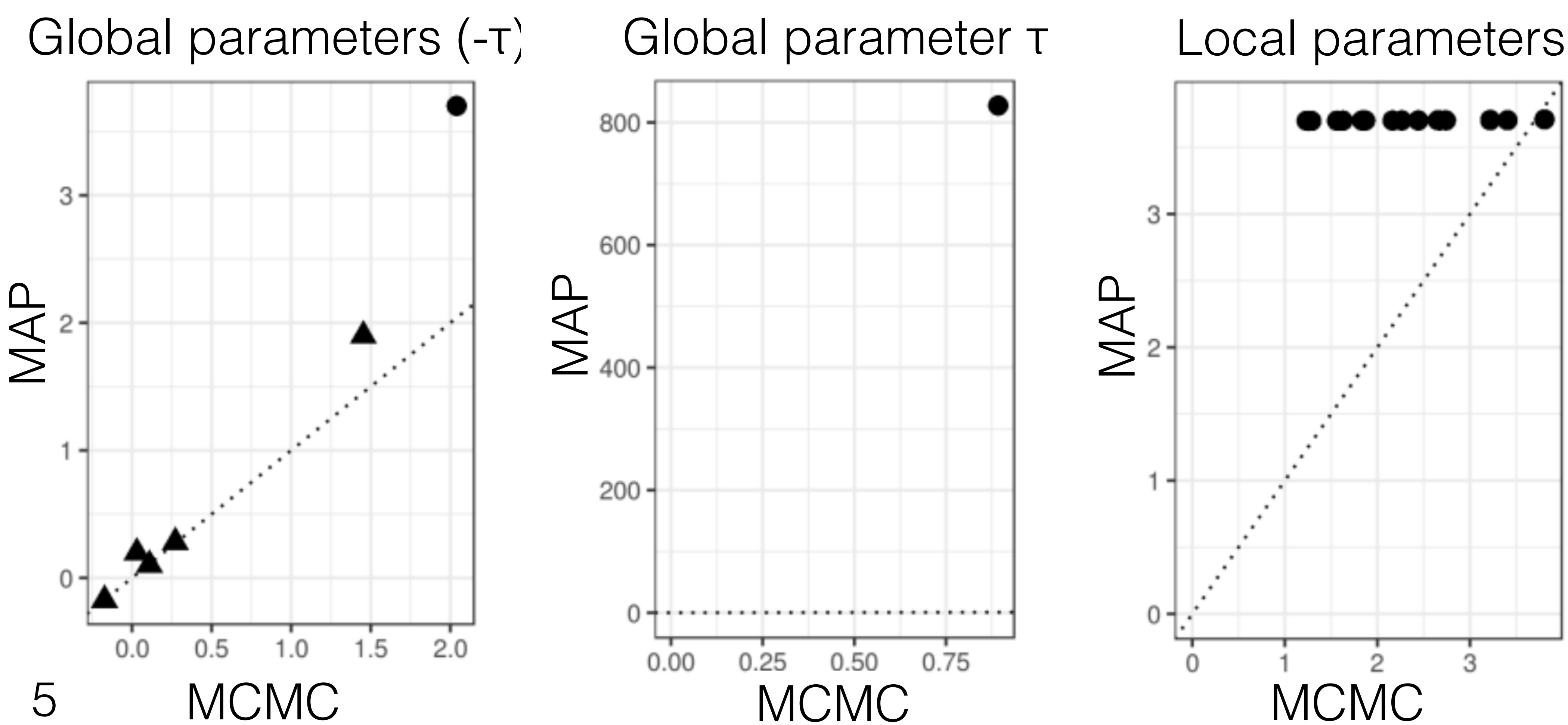
Global parameter τ



[Giordano, Broderick, Jordan 2018]

Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off, e.g. in (common) models with “global” and “local” parameters. Example: logistic generalized linear mixed model for online ads (Criteo data)

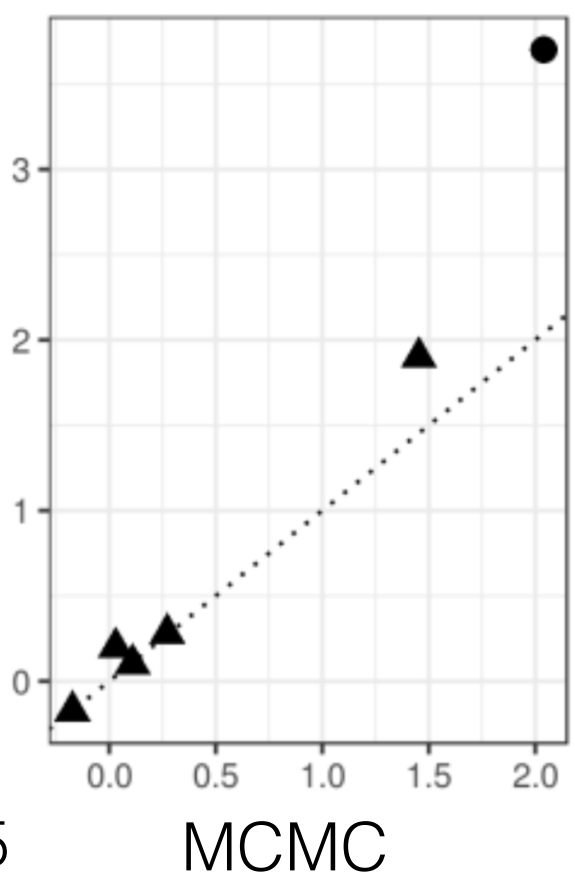


[Giordano, Broderick, Jordan 2018]

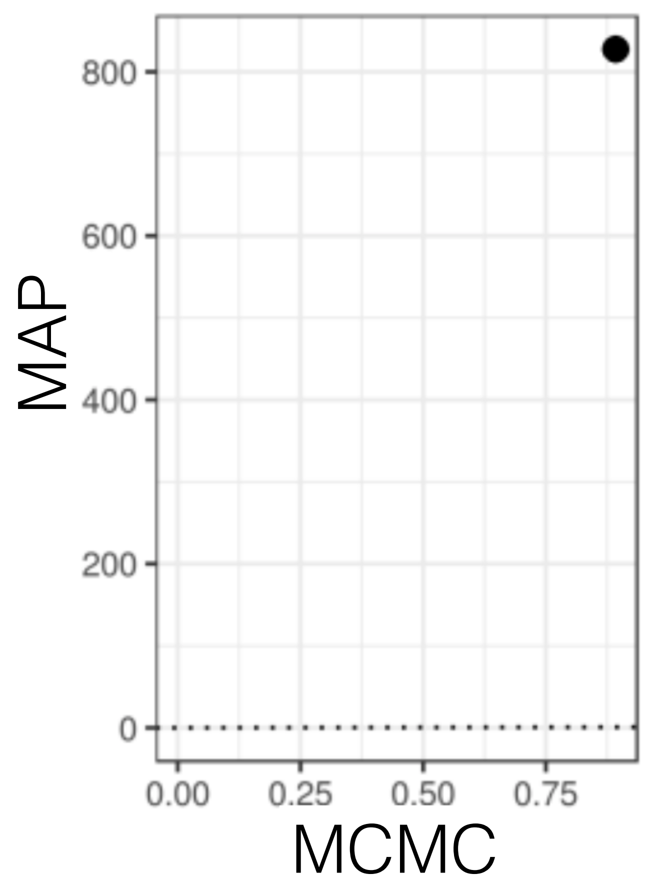
Idea: sandwich covariance

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
  - Approximate posterior mean with the MAP (no MCMC)
  - Then use the frequentist sandwich covariance
  - Use the delta method to account for *g*
- Problem: MAP can be way off, e.g. in (common) models with “global” and “local” parameters. Example: logistic generalized linear mixed model for online ads (Criteo data)

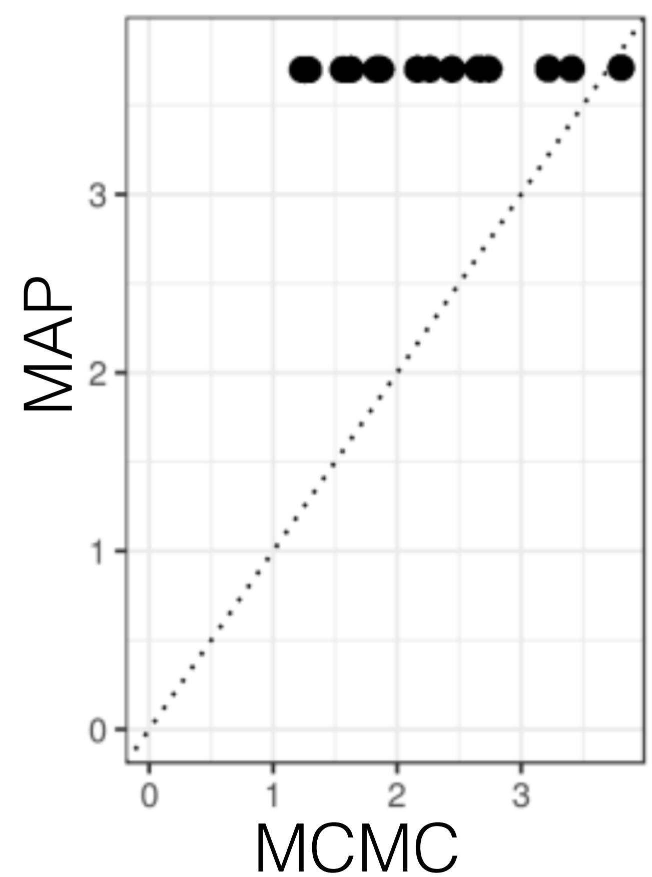
Global parameters (-τ)



Global parameter τ



Local parameters



- MAP: 12 s
- MCMC (5K samples): 5.85 h

[Giordano, Broderick, Jordan 2018]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*

Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013

Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available

Our idea: use the infinitesimal jackknife

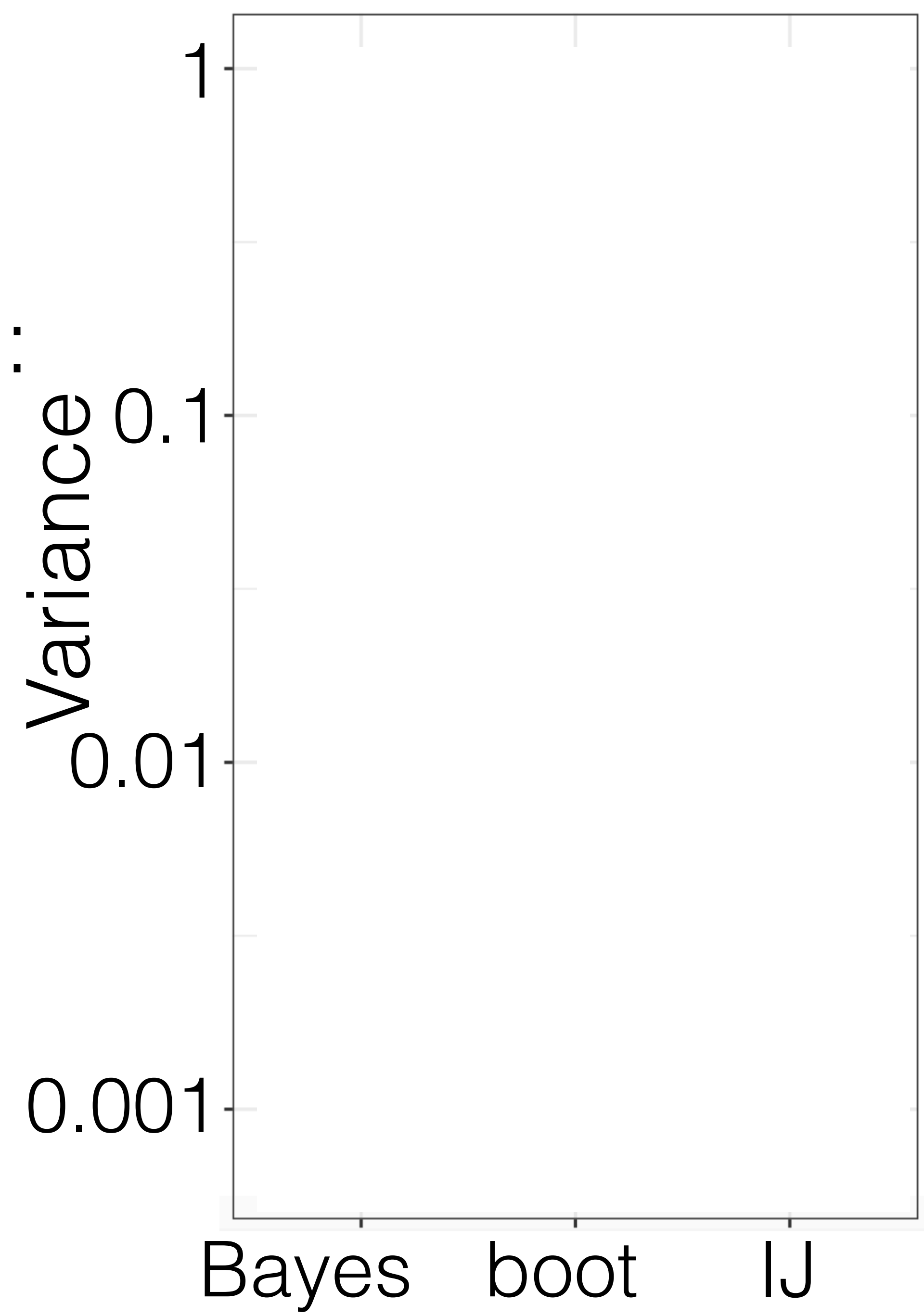
[Jaeckel 1972]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (*B*=100): ~51 days
  - IJ: seconds

Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

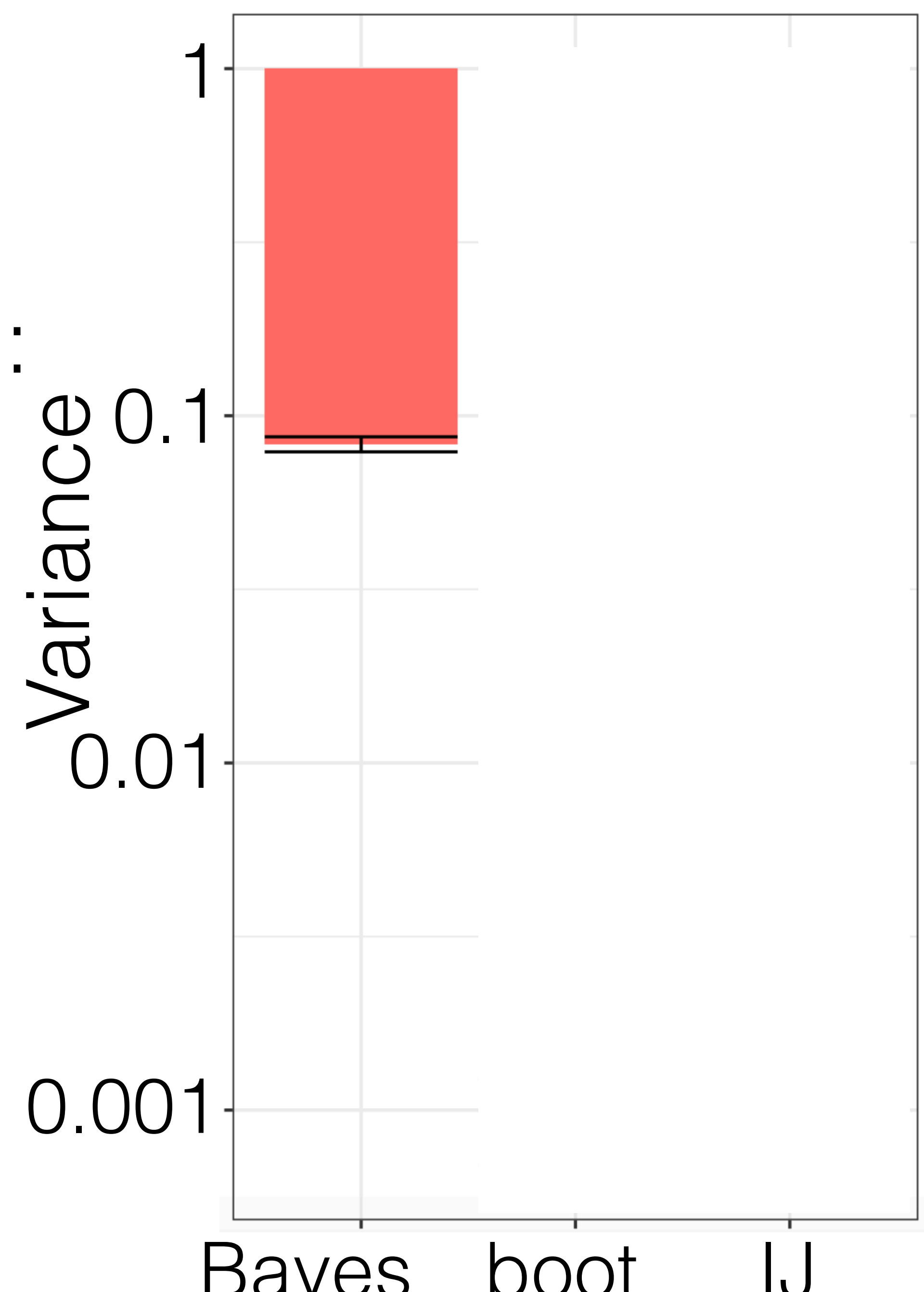
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (B=100): ~51 days
  - IJ: seconds



Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

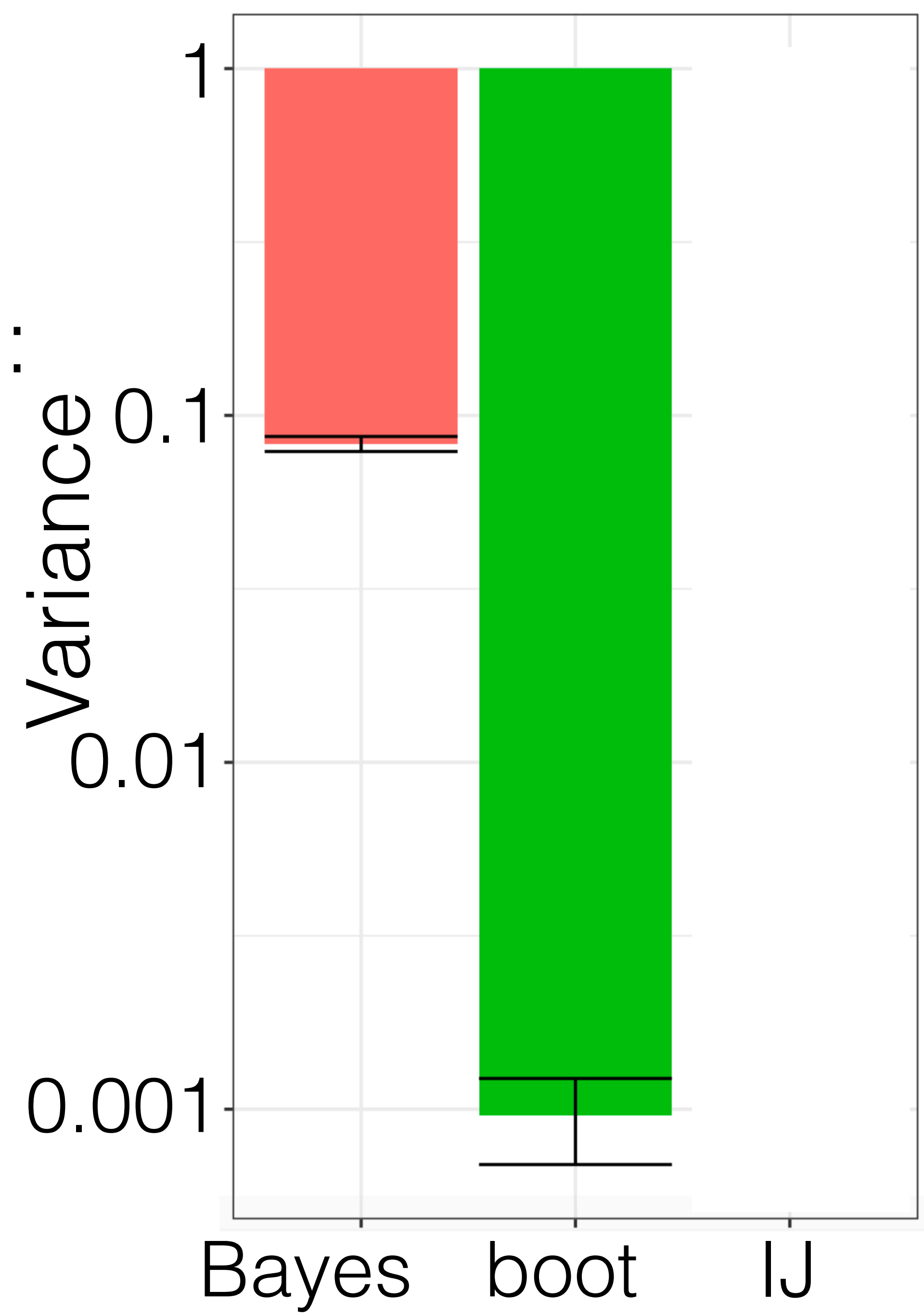
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (B=100): ~51 days
  - IJ: seconds



Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

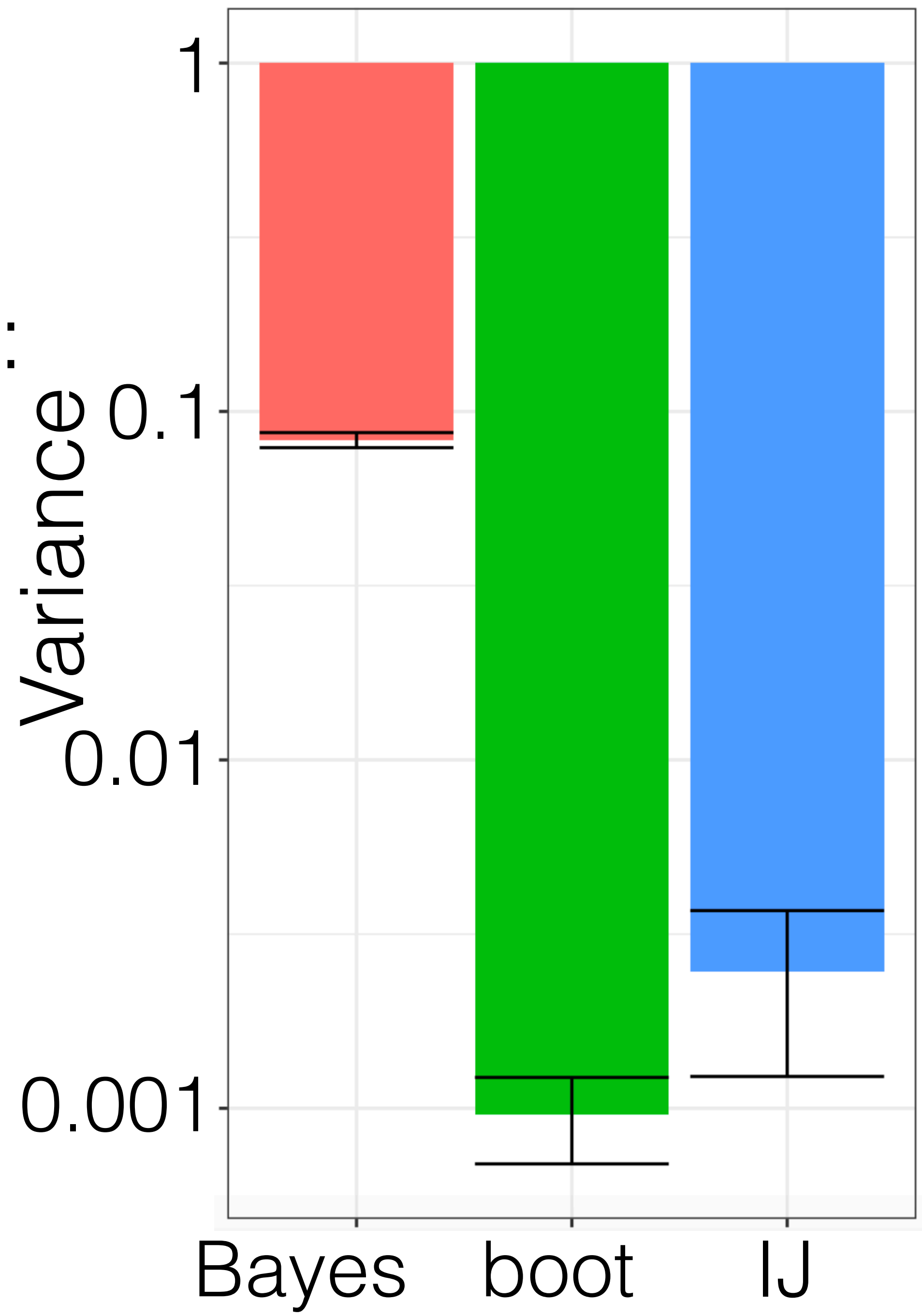
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (B=100): ~51 days
  - IJ: seconds



Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

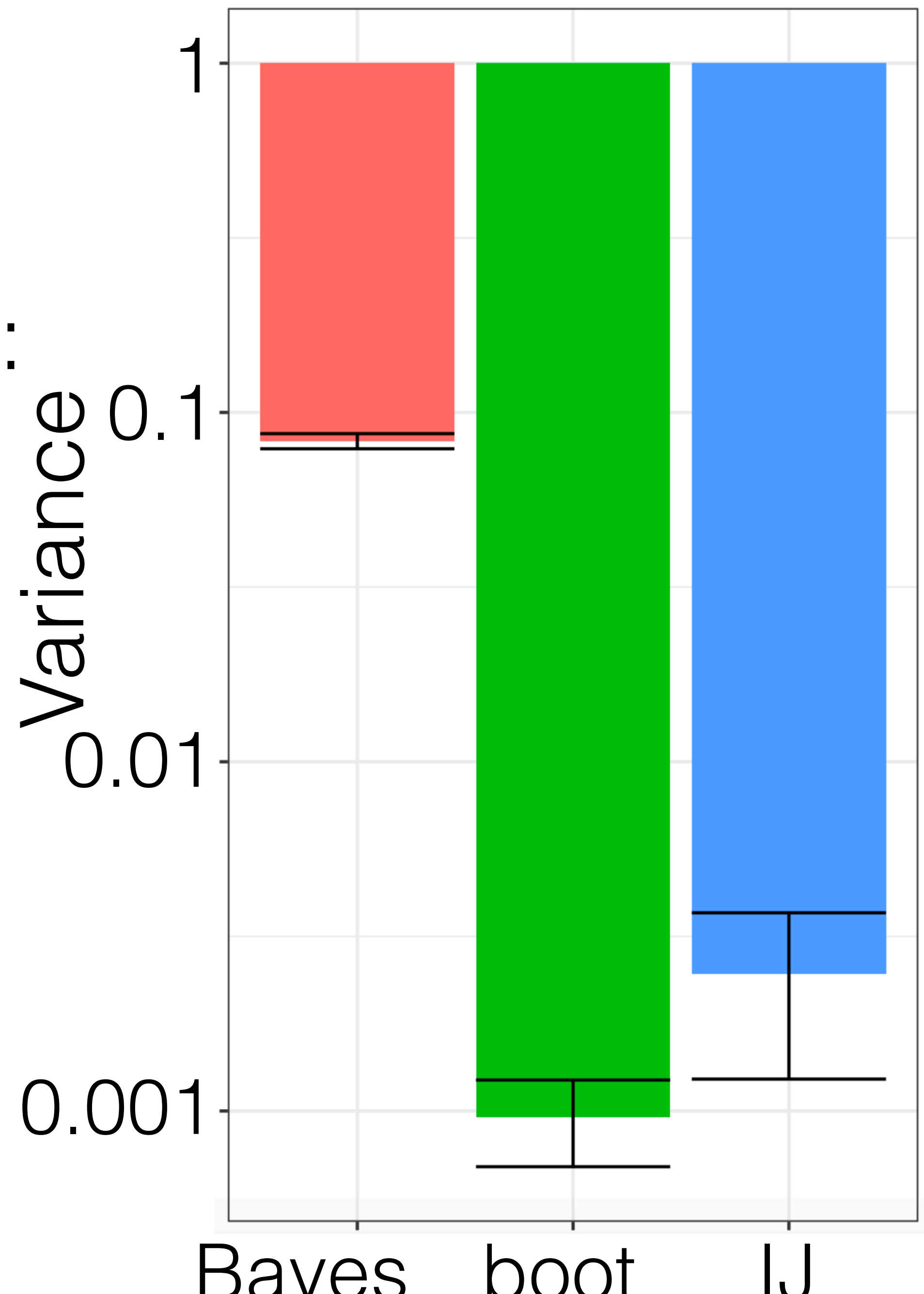
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (B=100): ~51 days
  - IJ: seconds



Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

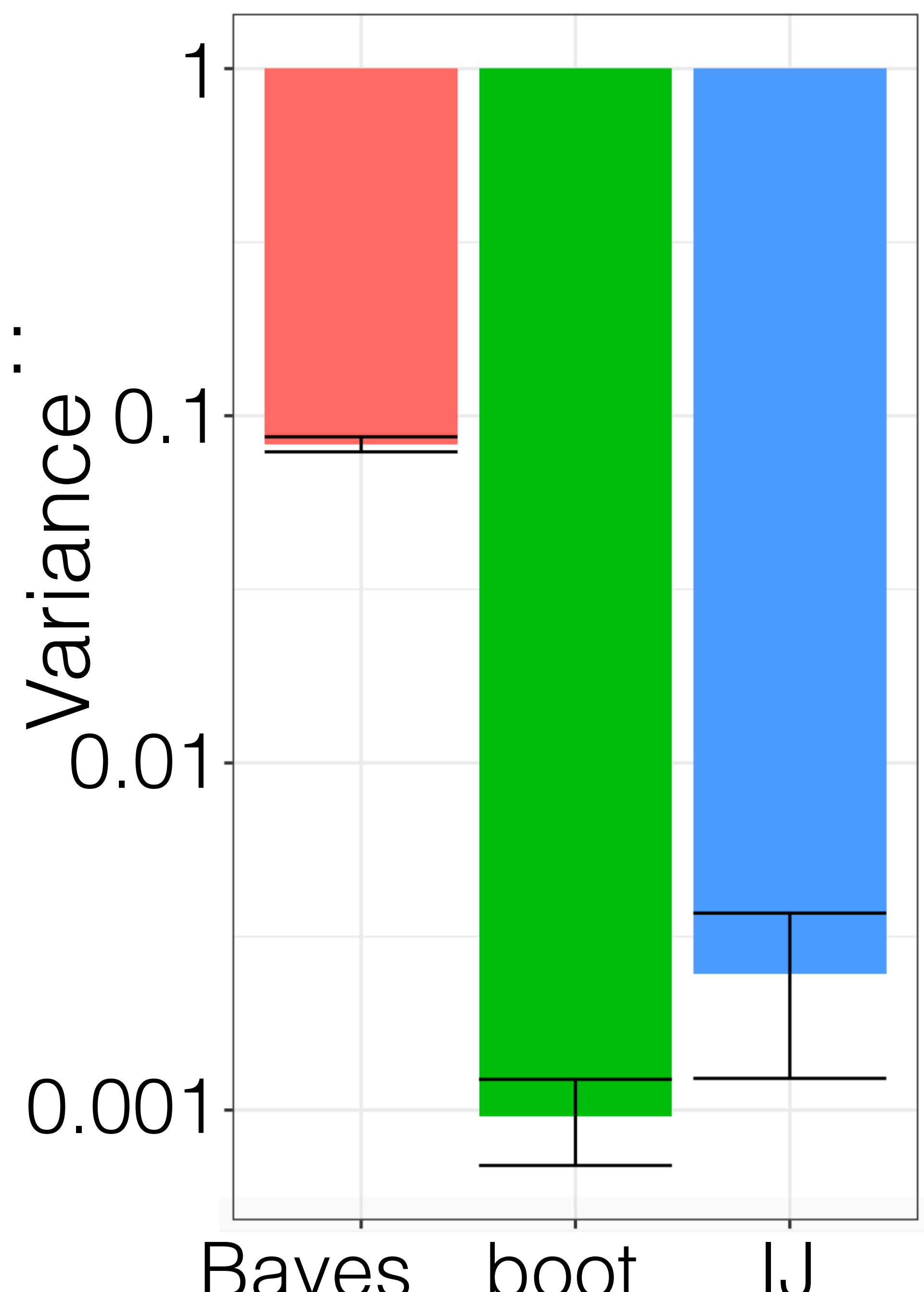
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
    - One MCMC run: ~12.3 hours
    - Bootstrap (B=100): ~51 days
    - IJ: seconds
- Qualitative conclusion: Bayesian uncertainty dominates frequentist



Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

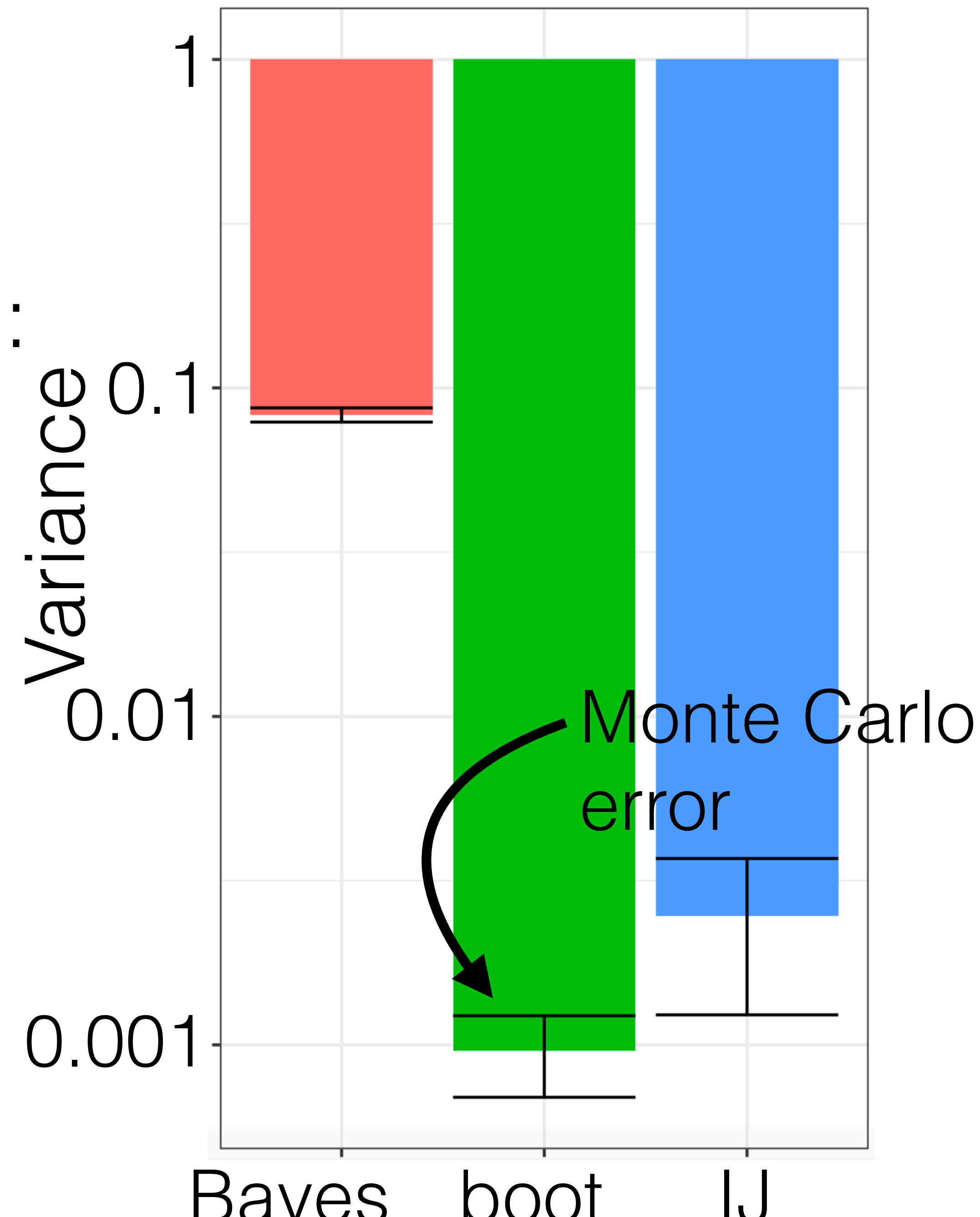
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (B=100): ~51 days
  - IJ: seconds
- Qualitative conclusion: Bayesian uncertainty dominates frequentist
- Later: we'll see a case where frequentist uncertainty dominates Bayesian



Our idea: use the infinitesimal jackknife

[Jaeckel 1972]

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Motivating example: election prediction, *The Economist*
- Gelman and Heidemanns 2020, building on Linzer 2013
- We examine the 2016 US presidential election: data & model publicly available
- Compute time (3000 MCMC draws):
  - One MCMC run: ~12.3 hours
  - Bootstrap (B=100): ~51 days
  - IJ: seconds
- Qualitative conclusion: Bayesian uncertainty dominates frequentist
- Later: we'll see a case where frequentist uncertainty dominates Bayesian



- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta)$

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

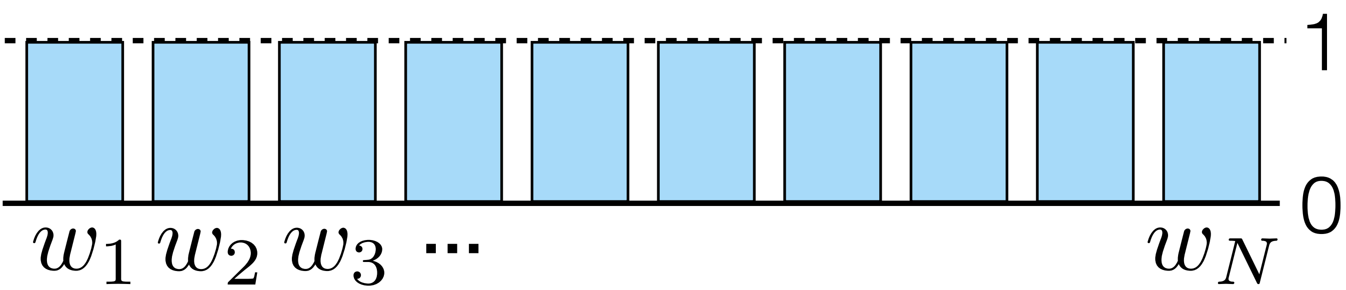
What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

What is the IJ estimate? Setup

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$
- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$

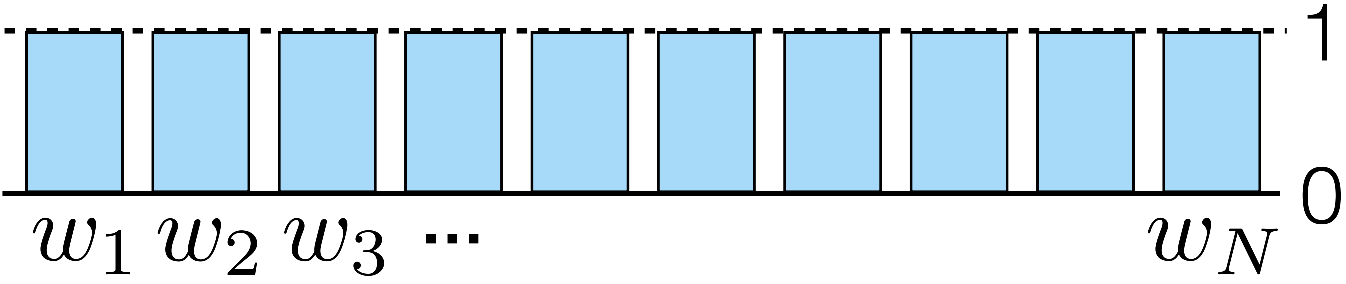


What is the IJ estimate? Setup

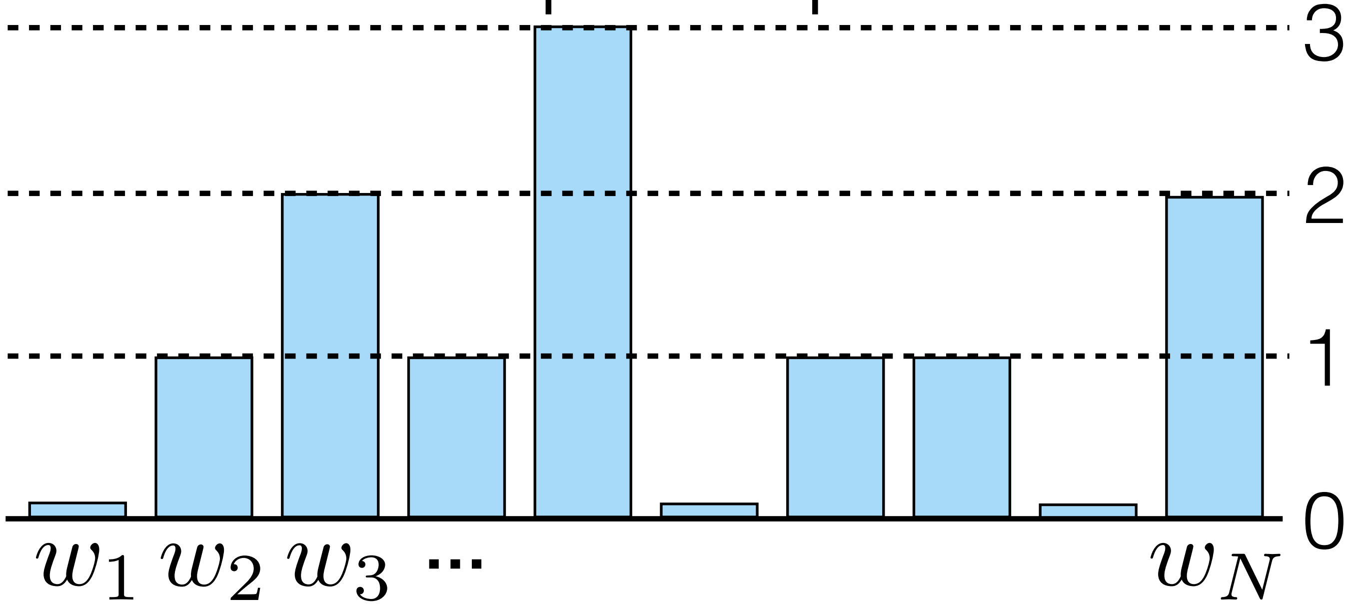
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:

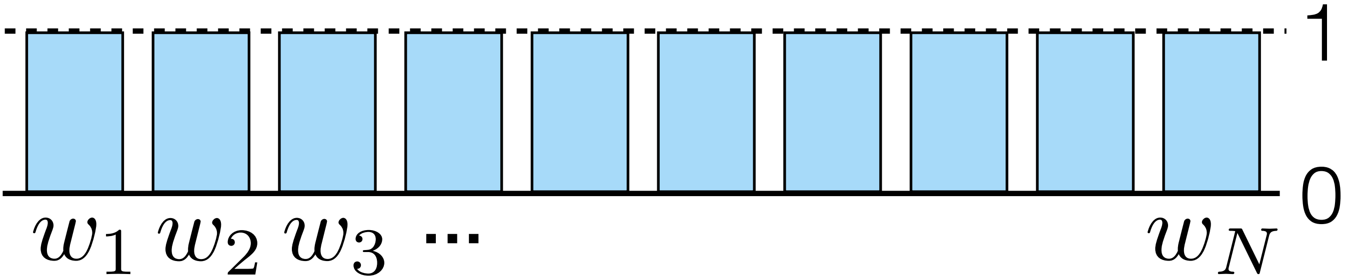


What is the IJ estimate? Setup

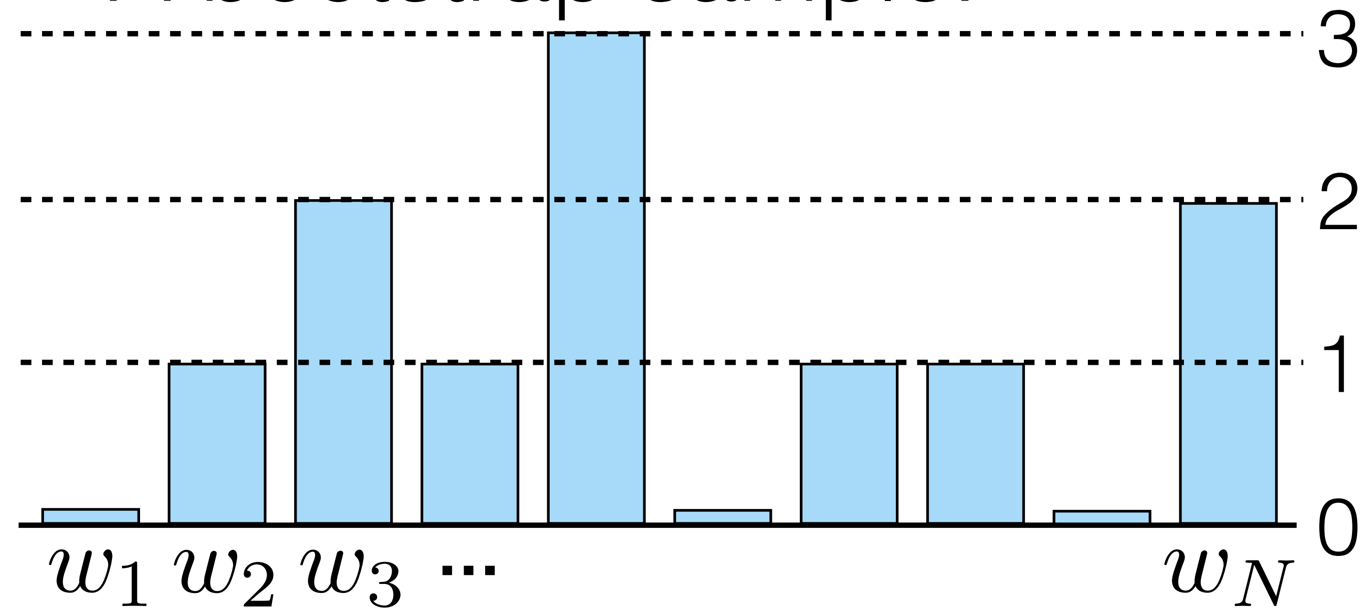
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



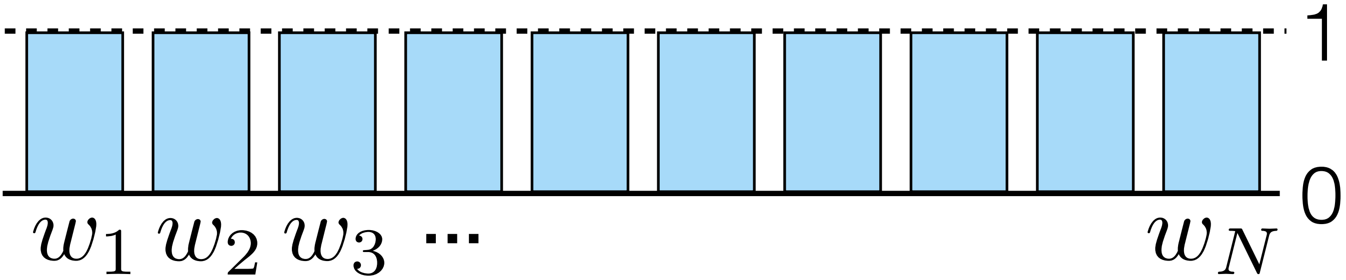
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

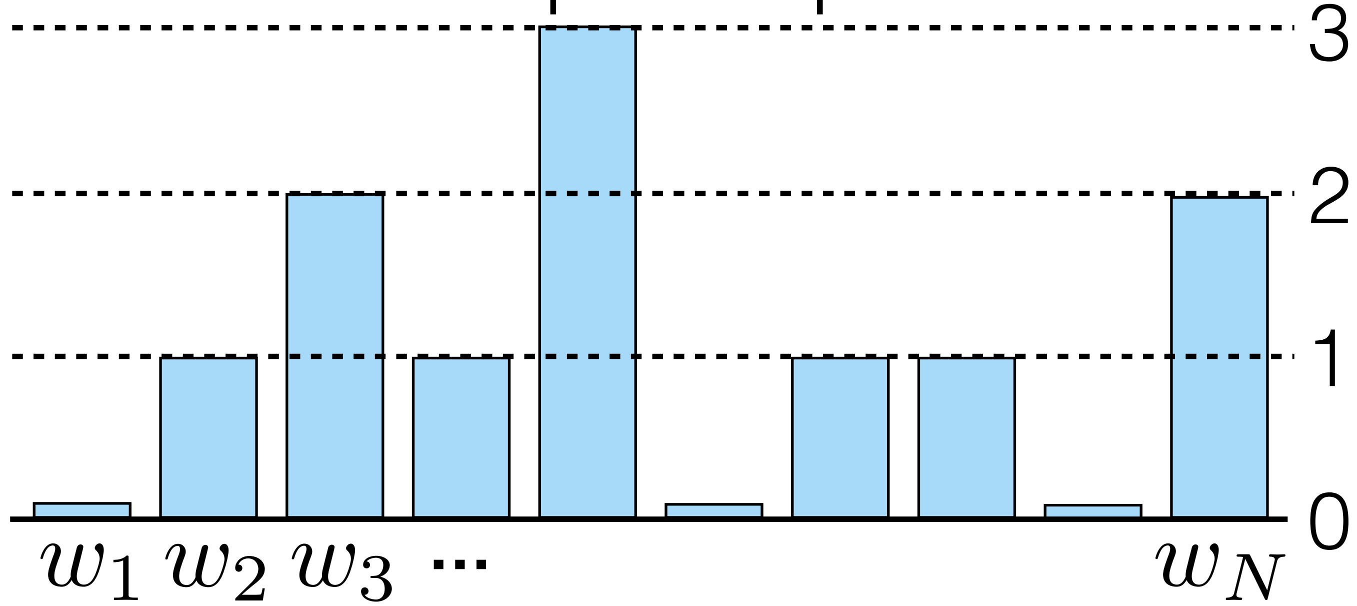
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



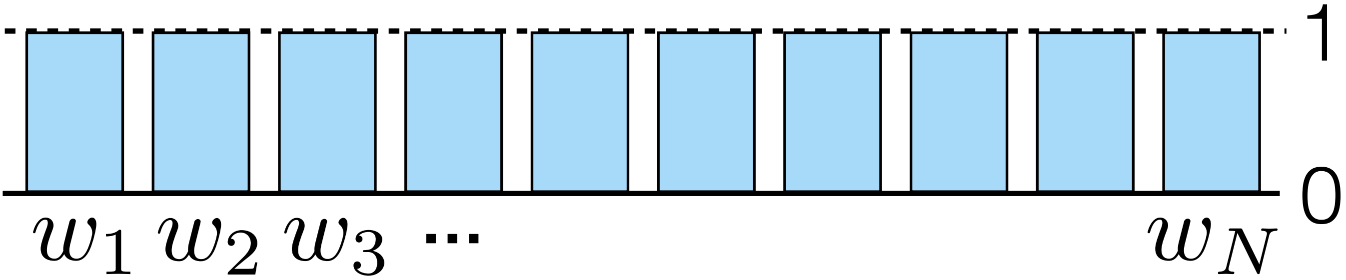
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

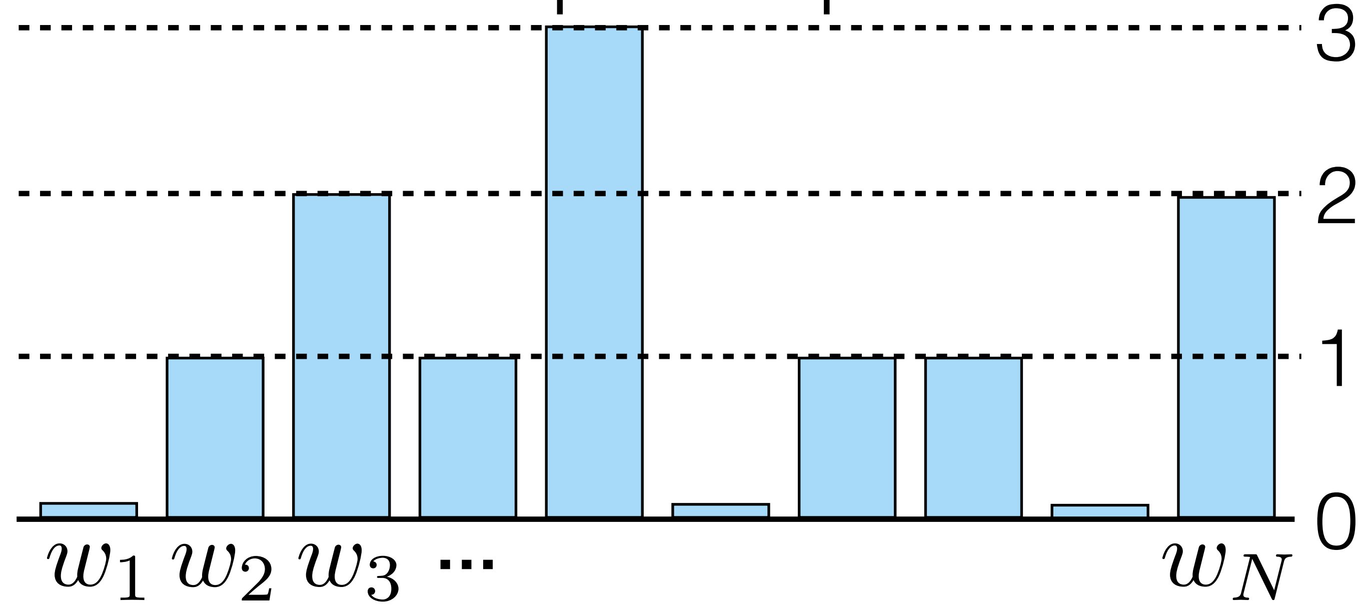
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



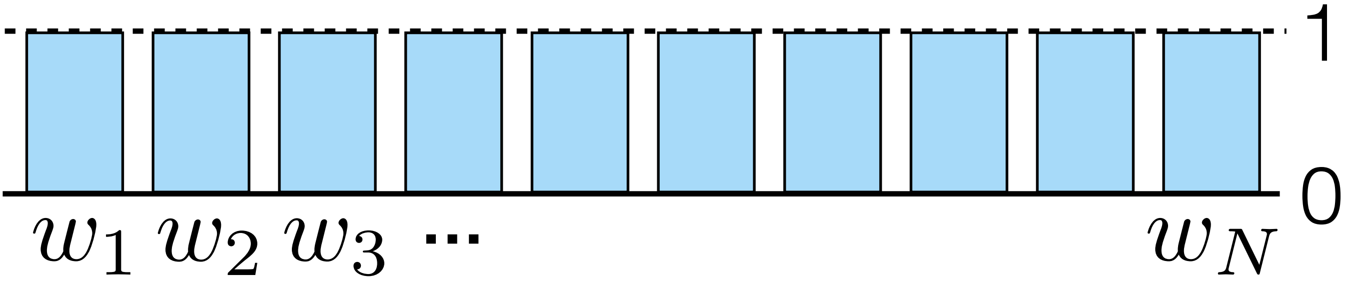
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

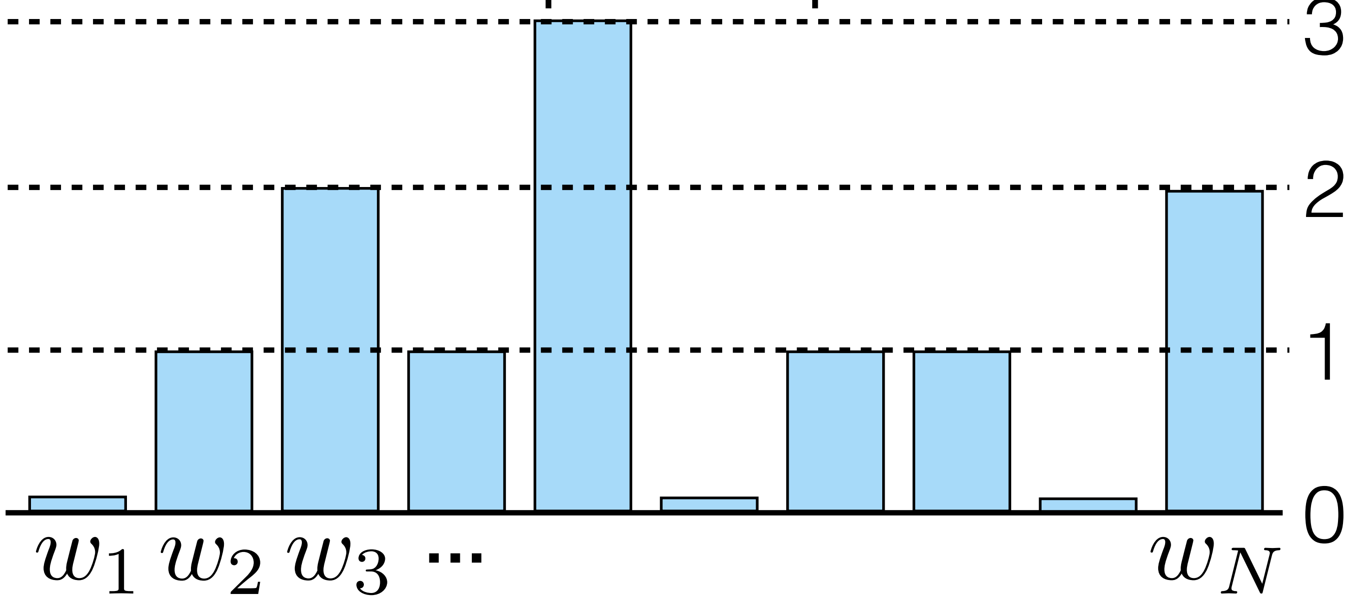
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



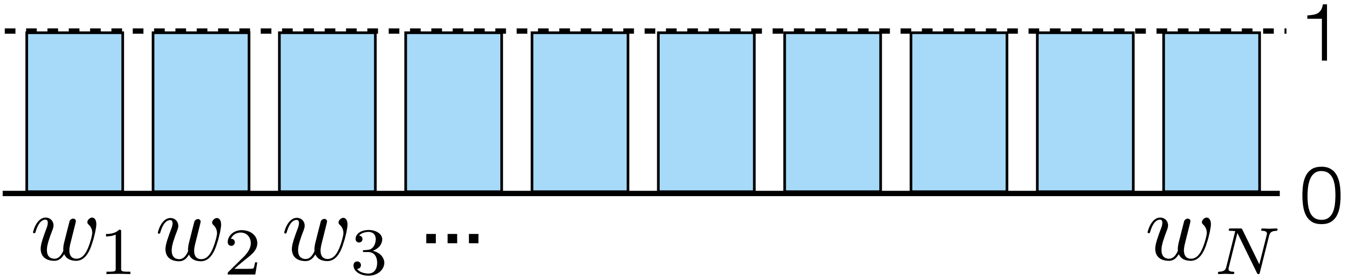
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

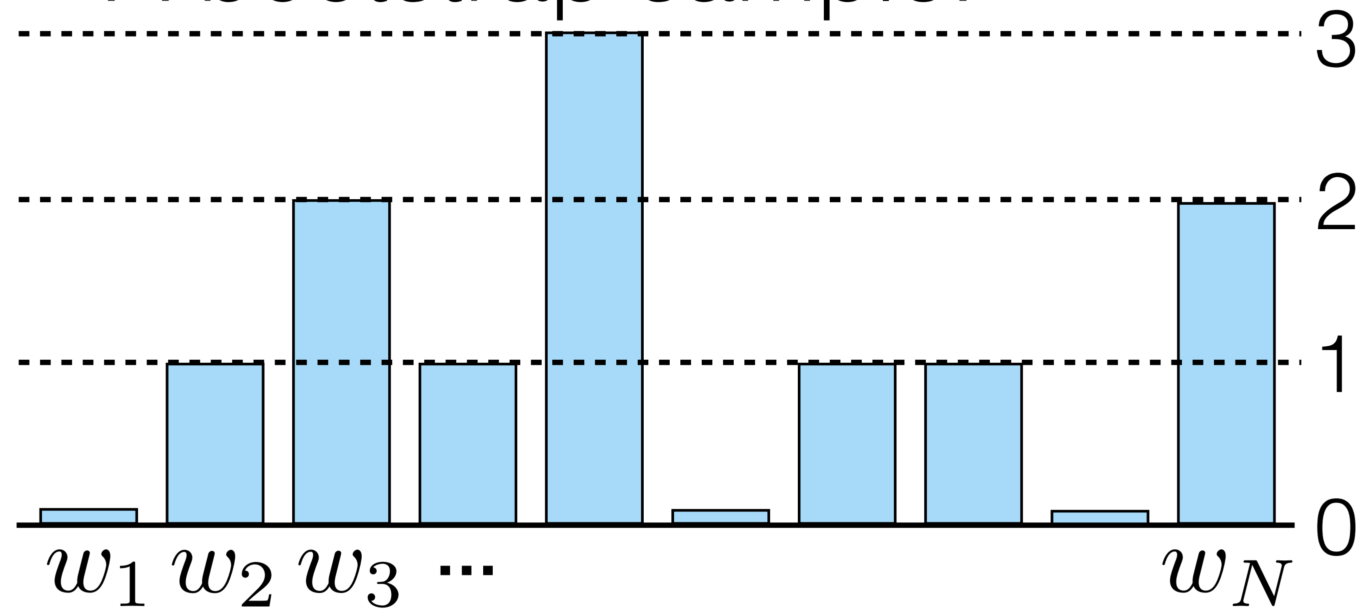
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



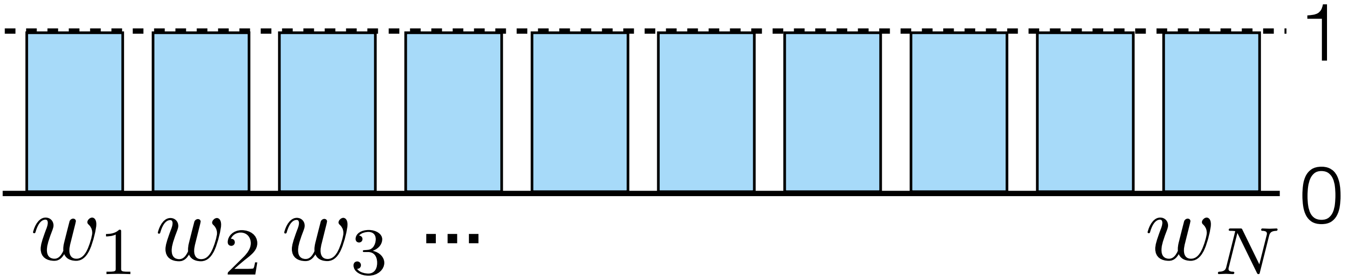
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

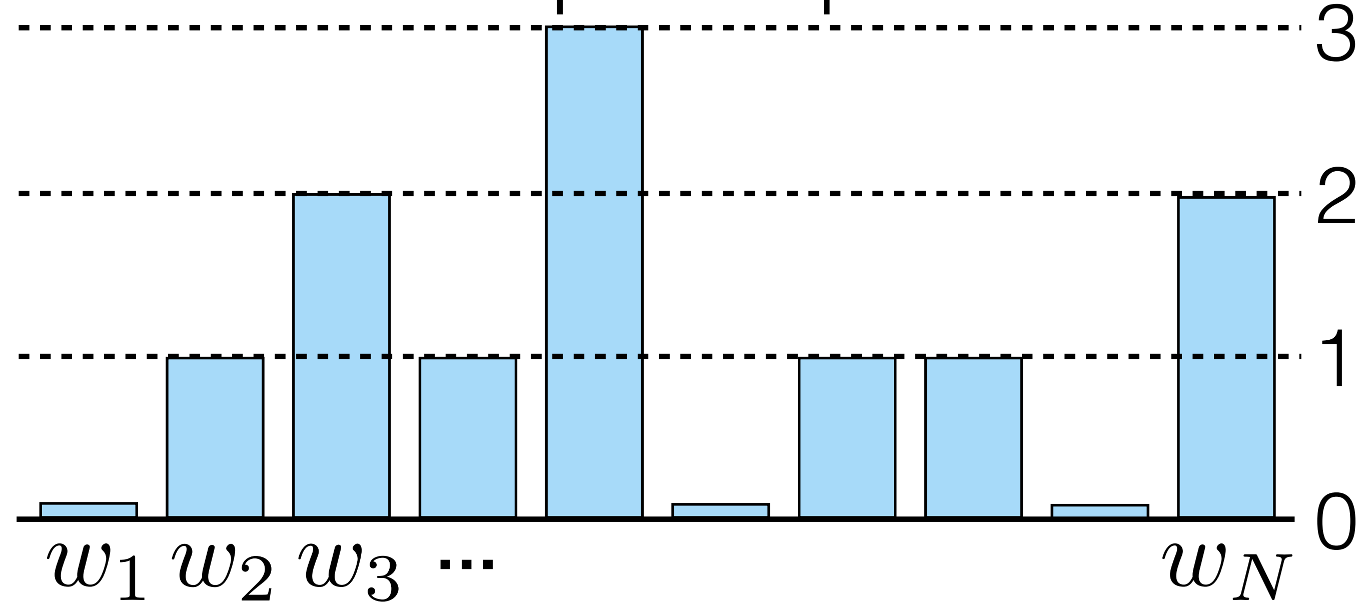
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



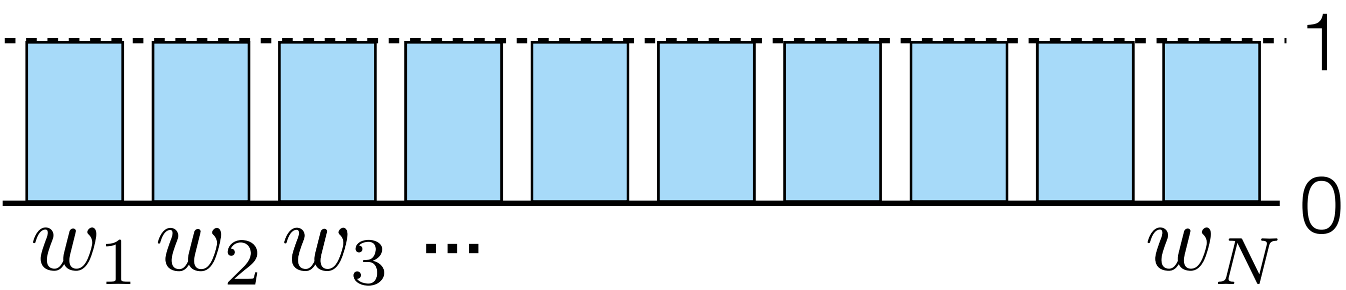
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

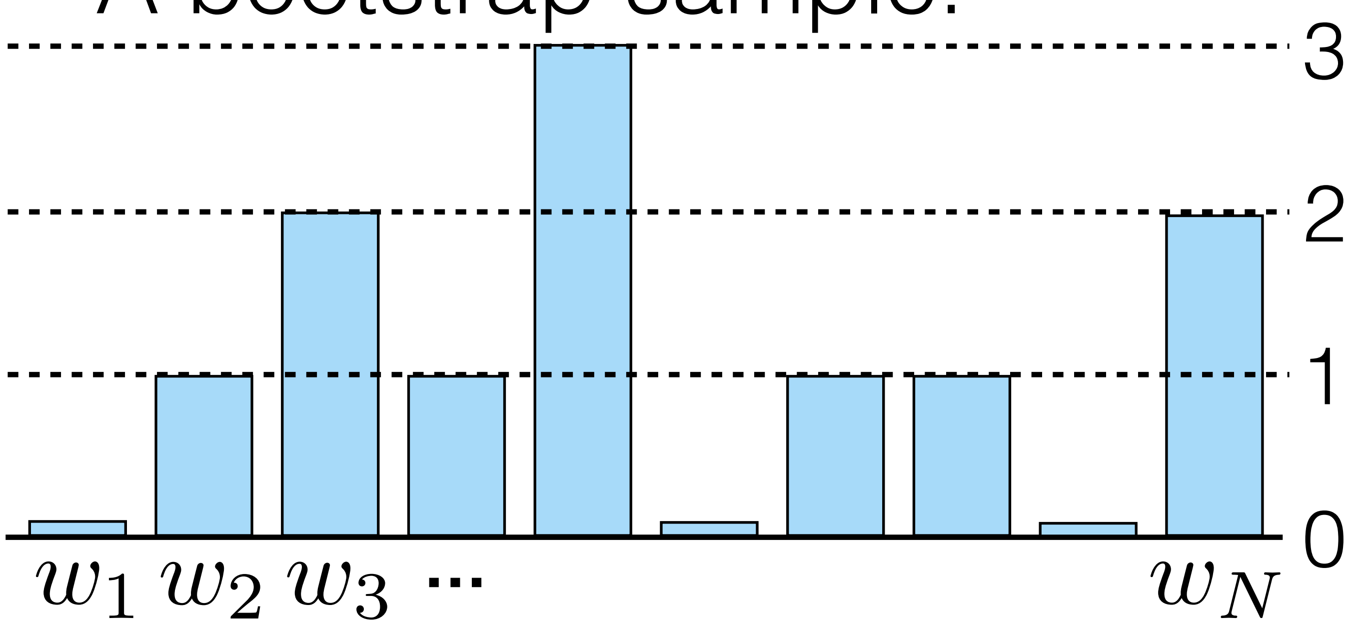
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



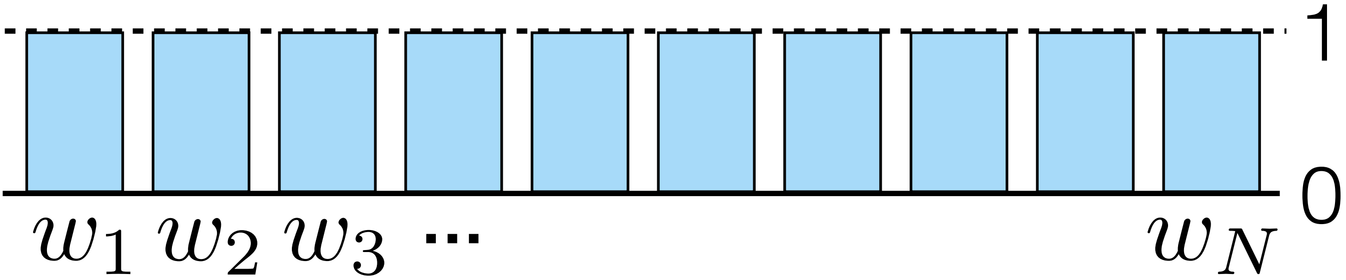
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

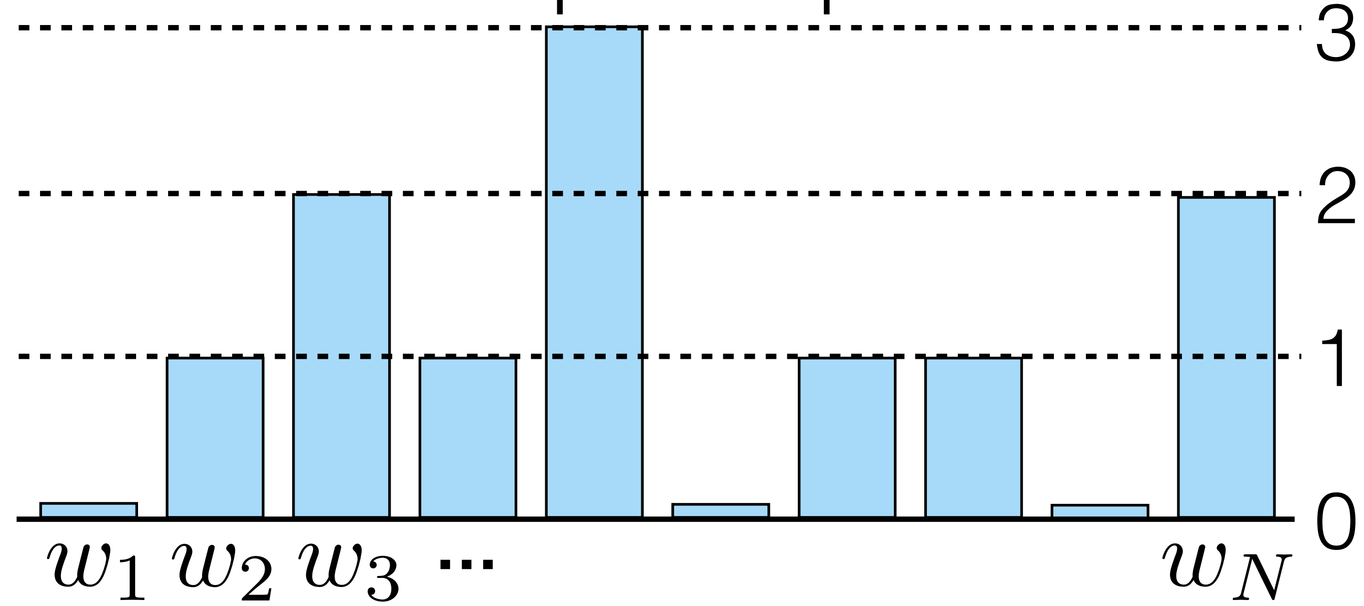
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



- Nonparametric bootstrap:

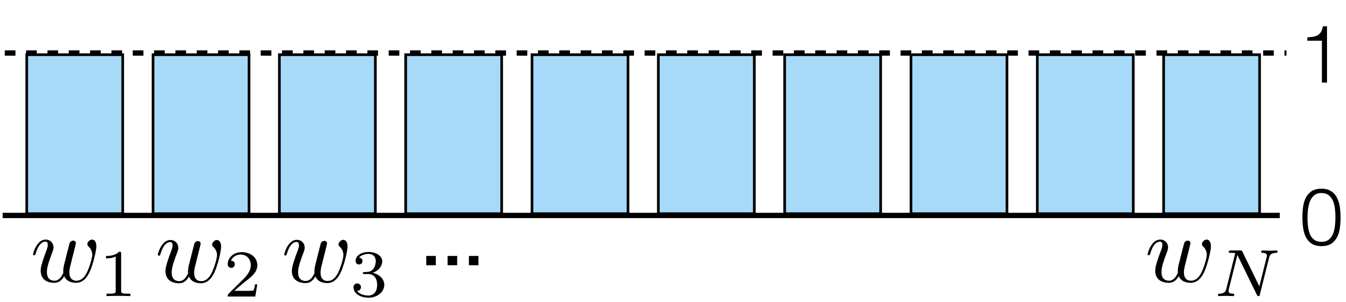
$$p(w) = \mathrm{Multinomial}(w; \tfrac{1}{N} 1_N, N)$$

What is the IJ estimate? Setup

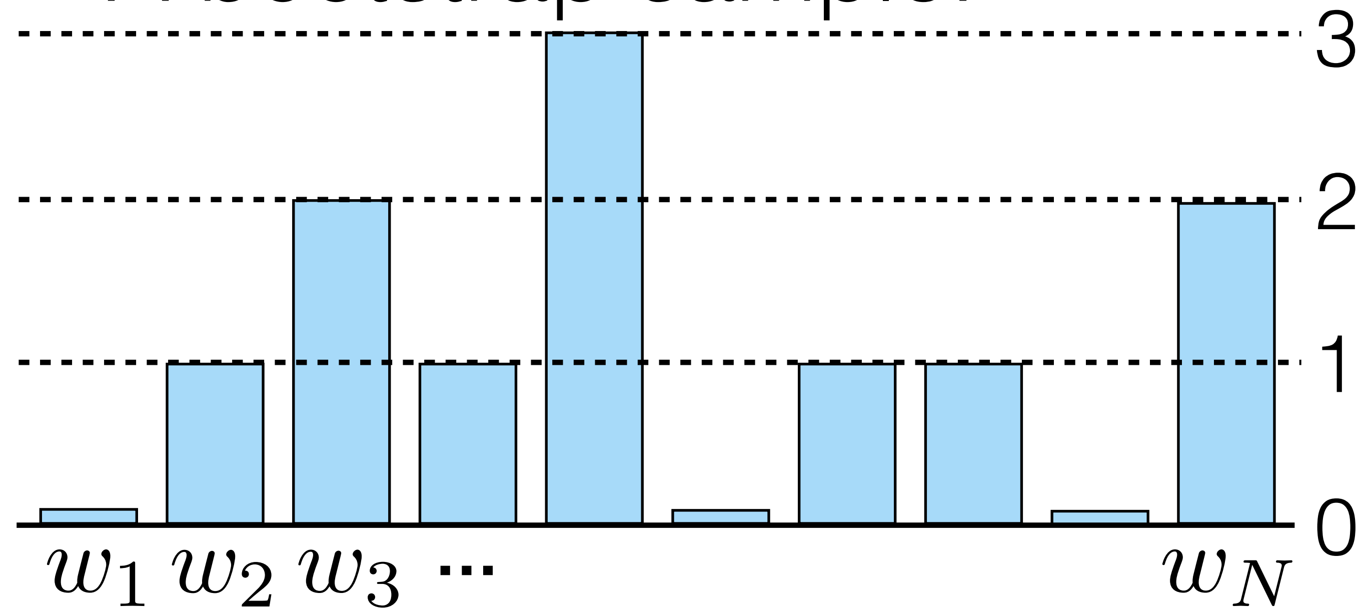
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



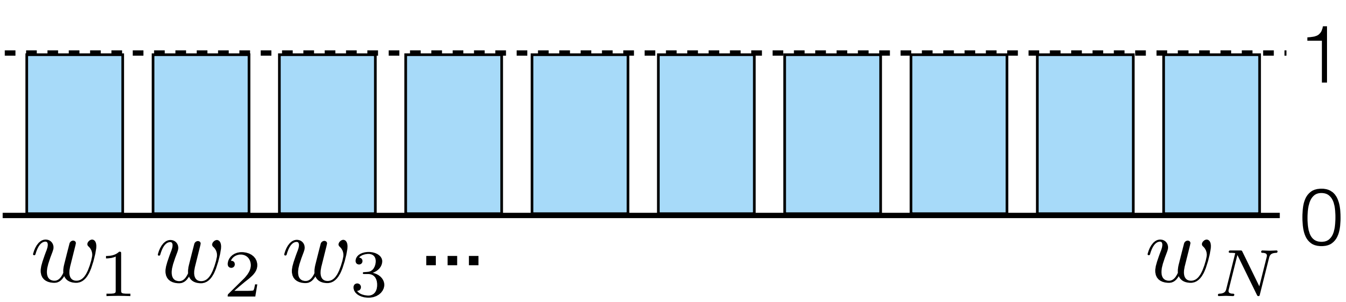
- Nonparametric bootstrap:
  - Step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Multinomial}(w; \frac{1}{N} 1_N, N)$

What is the IJ estimate? Setup

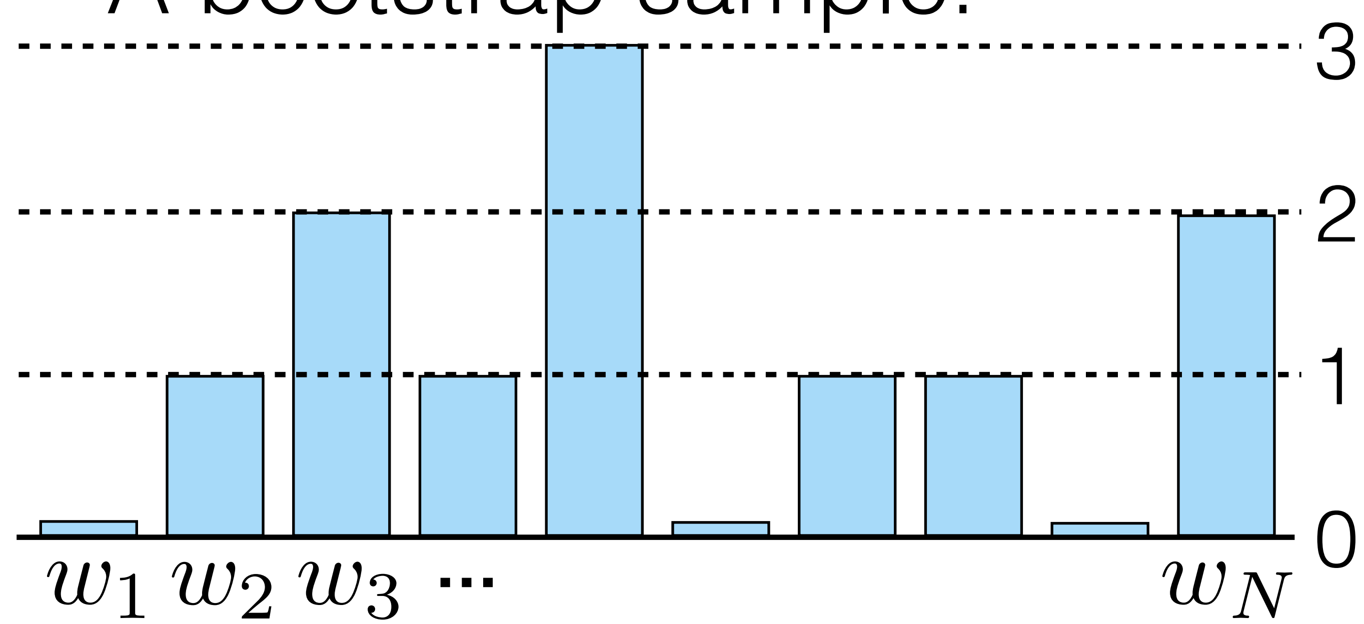
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



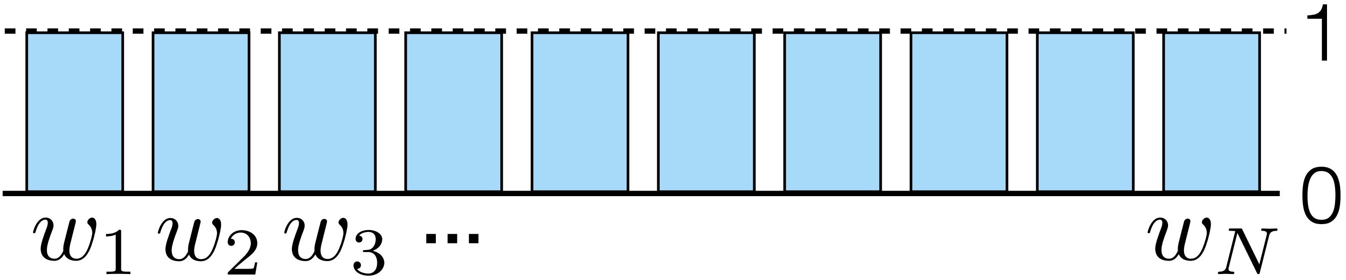
- Nonparametric bootstrap:
  - Step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Multinomial}(w; \frac{1}{N} 1_N, N)$

What is the IJ estimate? Setup

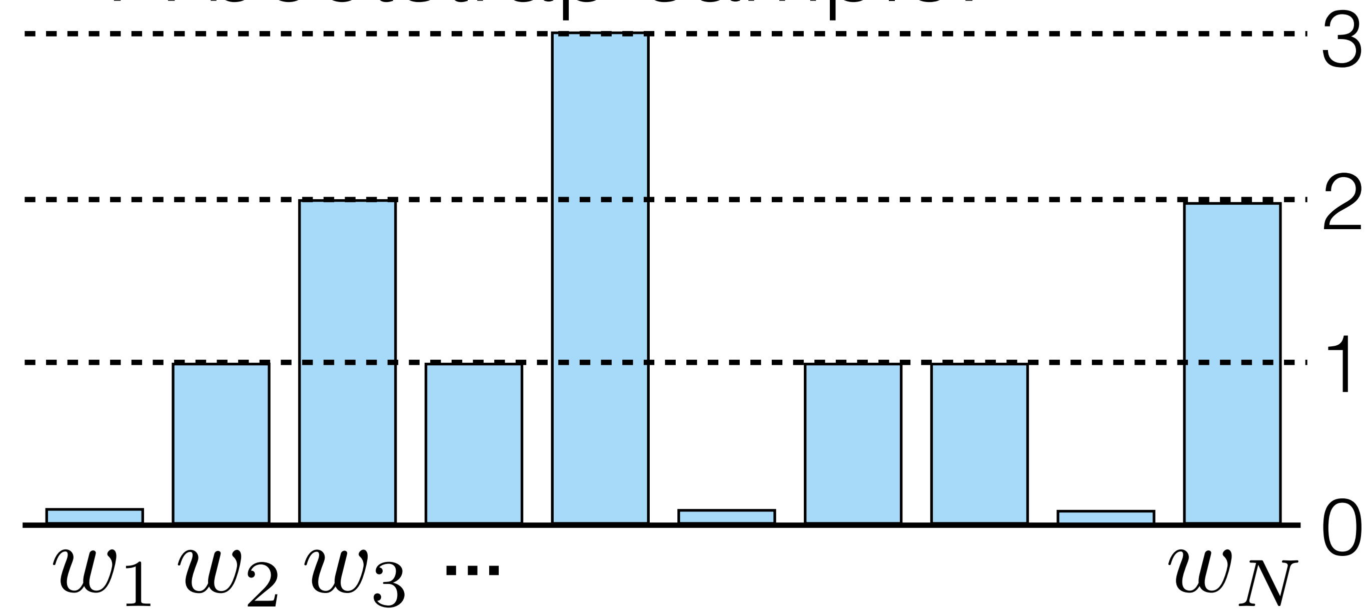
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



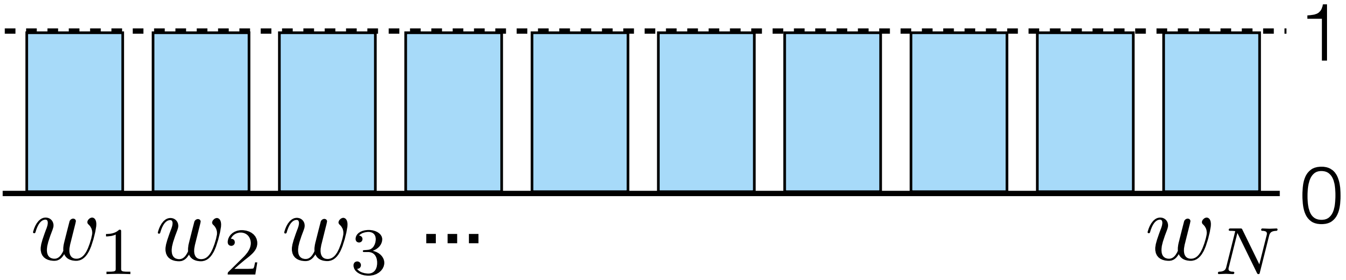
- Nonparametric bootstrap:
  - Step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Multinomial}(w; \frac{1}{N} 1_N, N)$

What is the IJ estimate? Setup

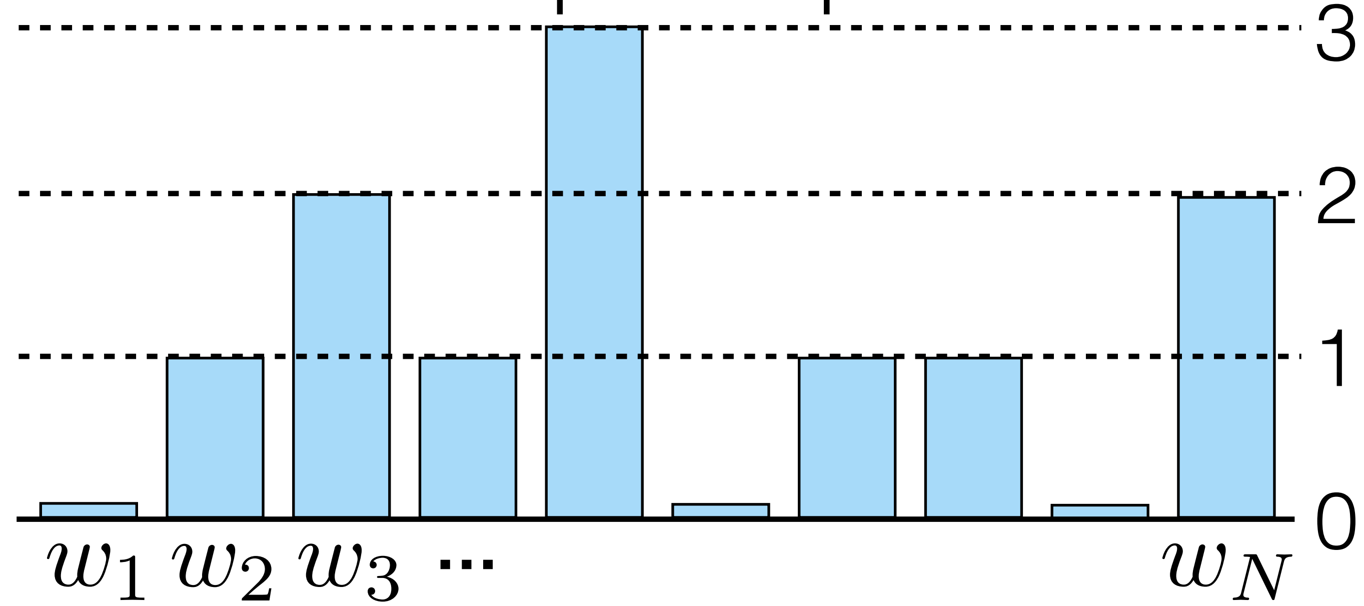
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



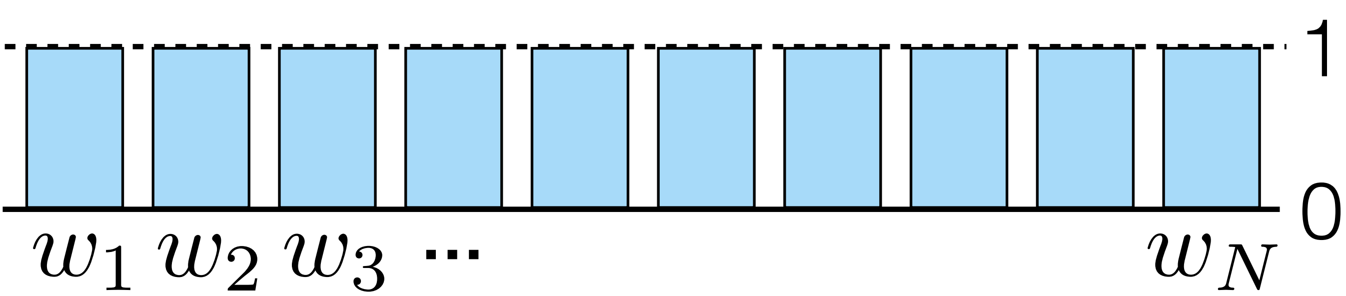
- Nonparametric bootstrap:
  - Step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)]$, $p(w) = \mathrm{Multinomial}(w; \frac{1}{N} 1_N, N)$
  - Step 2:

What is the IJ estimate? Setup

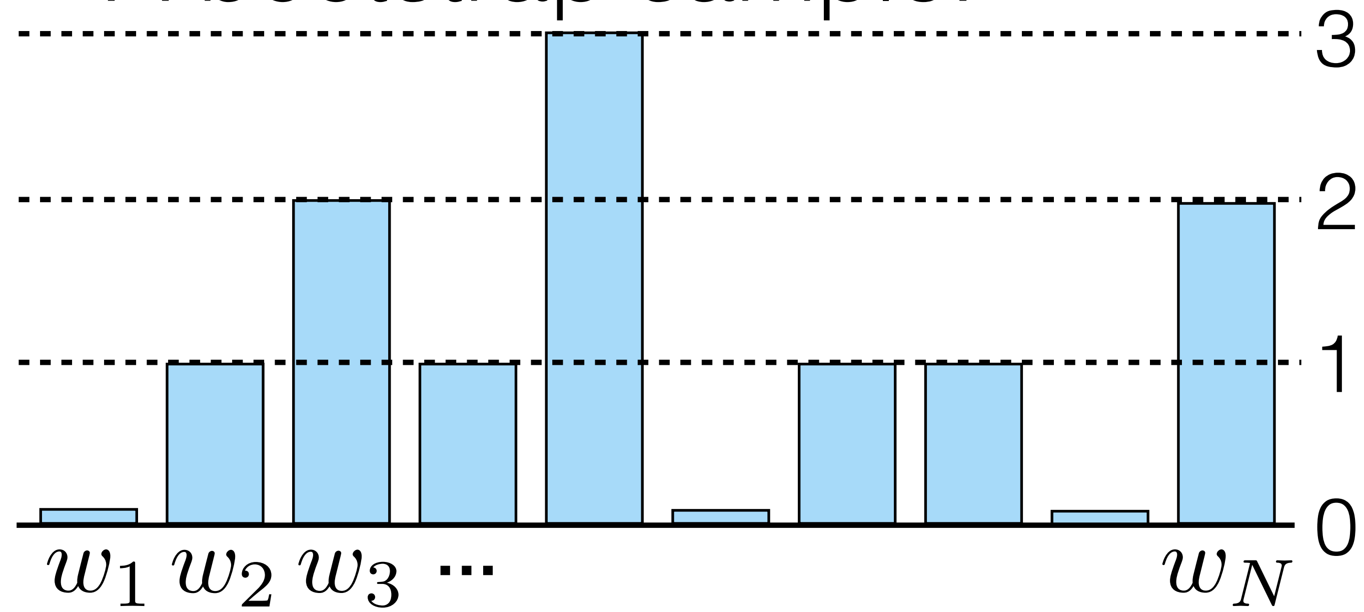
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



- Nonparametric bootstrap:
  - Step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Multinomial}(w; \frac{1}{N} 1_N, N)$
  - Step 2:

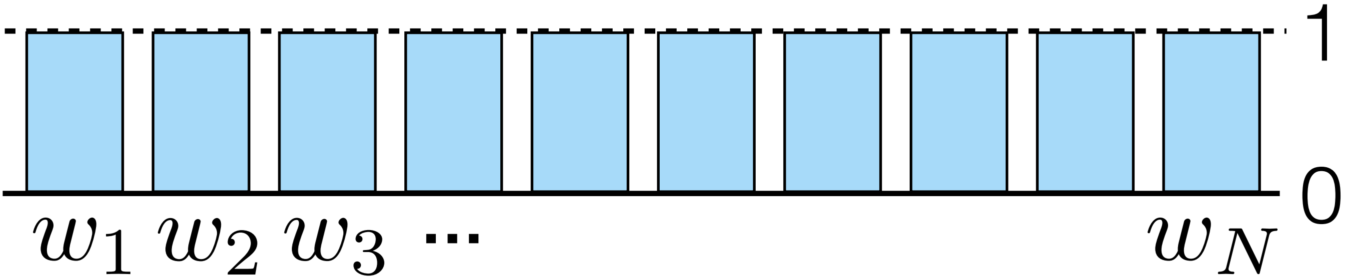
MC

What is the IJ estimate? Setup

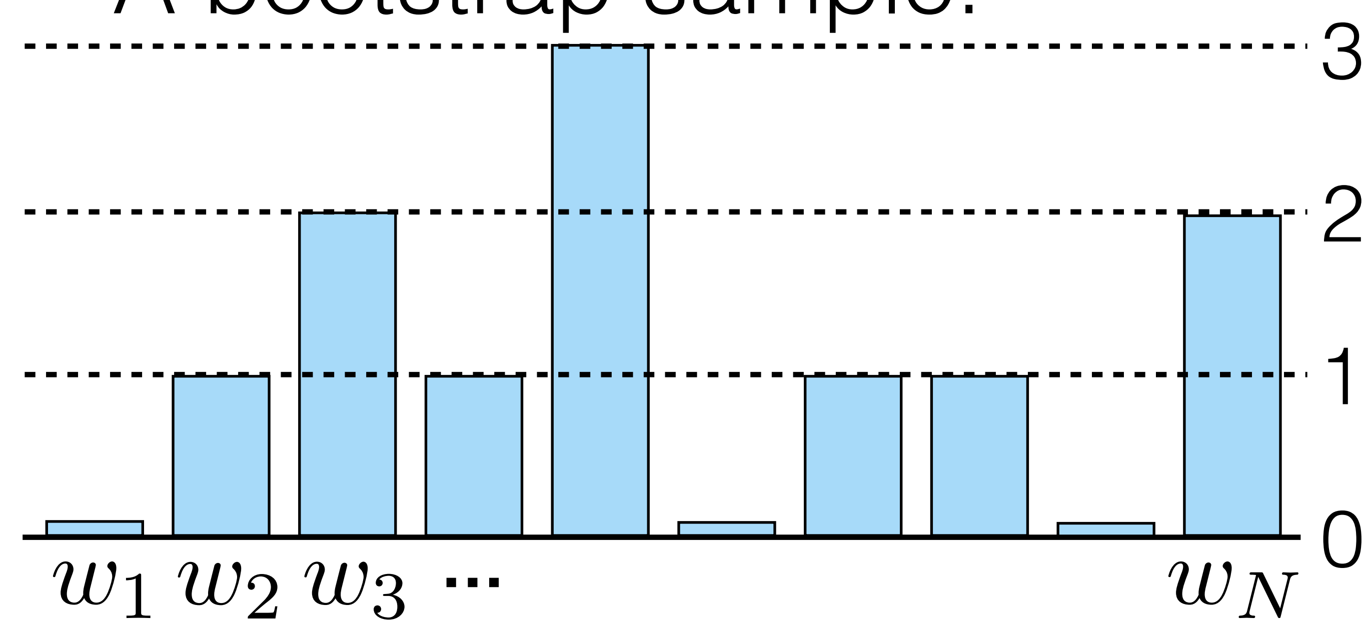
- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- With prior $p(\theta)$, yields exact posterior $p(\theta|X, w)$

- Original data:

$$w = 1_N = (1, \ldots, 1, 1, 1, \ldots 1)$$



- A bootstrap sample:



- Nonparametric bootstrap:
  - Step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Multinomial}(w; \frac{1}{N} 1_N, N)$
  - Step 2:

MC

MCMC

- And in practice:

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)]$$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)]$$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)]$$

exact

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \boxed{\phi^{\mathrm{lin}}(w)} := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \boxed{\phi^{\mathrm{lin}}(w)} := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^N w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^N (w_n - 1)\psi_n,$$

exact

linear approximation

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^N w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^N (w_n - 1)\psi_n,$$

exact

linear approximation

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left. \frac{\partial \phi(w)}{\partial w_n} \right|_{w=1_N}$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N}$

influence score

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left. \frac{\partial \phi(w)}{\partial w_n} \right|_{w=1_N} = \left. \underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\} \right|_{w=1_N}$

influence score

- Then:

$$\boxed{\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)]}$$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \boxed{\underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right)}$$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right) = \widehat{\mathrm{Var}}_n \, \psi_n$$

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \;\; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right) = \widehat{\mathrm{Var}}_n \; \psi_n$$

MCMC

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right) = \widehat{\mathrm{Var}}_n \; \psi_n$$

MCMC

- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)], \; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right) = \widehat{\mathrm{Var}}_n \; \psi_n$$

MCMC

- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\ p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right) = \widehat{\mathrm{Var}}_n\ \psi_n$$

MCMC

- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance
- IJ: no additional MCMC chains (vs. *B*-1 extra for bootstrap)

What is the IJ estimate?

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?

- Log likelihood $\log p(X|\theta, w) = \sum_{n=1}^{N} w_n \log p(x_n|\theta)$
- Bootstrap step 1: $\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)],\; p(w) = \mathrm{Mult}(w; \frac{1}{N} 1_N, N)$
- IJ idea: first-order Taylor expansion around $w = 1_N$:

$$\phi(w) := \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \phi^{\mathrm{lin}}(w) := \phi(1_N) + \sum_{n=1}^{N} (w_n - 1)\psi_n,$$

exact

linear approximation

where $\psi_n := \left.\frac{\partial \phi(w)}{\partial w_n}\right|_{w=1_N} = \left.\underset{p(\theta|X,w)}{\mathrm{Cov}} \{g(\theta), \log p(x_n|\theta)\}\right|_{w=1_N}$

influence score

- Then:

$$\underset{p(w)}{\mathrm{Var}} \; \underset{p(\theta|X,w)}{\mathbb{E}} [g(\theta)] \approx \underset{p(w)}{\mathrm{Var}} \left(\sum_{n=1}^{N} (w_n - 1)\psi_n \right) = \widehat{\mathrm{Var}}_n \; \psi_n$$

MCMC

- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance
- IJ: no additional MCMC chains (vs. *B*-1 extra for bootstrap)
- Fully automatable (given *g*, likelihood, and MCMC chain)

Related work

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance

Related work

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance
- The infinitesimal jackknife and influence functions are venerable ideas; they have been used to debias predictions, interpret black-box predictions, approximate cross-validation, etc. [Jaeckel 1972; Cook, Weisberg 1980; Belsley+ 1980 ch 2.1; Efron, 1982; Huh, Park 1990; Koh, Liang 2017; Giordano+ 2019a]

Related work

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \; \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance
- The infinitesimal jackknife and influence functions are venerable ideas; they have been used to debias predictions, interpret black-box predictions, approximate cross-validation, etc. [Jaeckel 1972; Cook, Weisberg 1980; Belsley+ 1980 ch 2.1; Efron, 1982; Huh, Park 1990; Koh, Liang 2017; Giordano+ 2019a]
- They have been used for local Bayesian robustness to changes in prior or likelihood. The covariance form is well known. [Diaconis, Freedman, 1986; Ruggeri, Wasserman, 1993; Gustafson, 1996; Giordano+ 2018; Mohamed+ 2020]. See also Billingsley, 2008, Thm 5.4 & Giordano+ 2023 rejoinder

Related work

- How to approximate $\underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)]$?
- IJ Algorithm: (1) Approximate *N* influence scores using one (original) MCMC chain; (2) Compute empirical variance
- The infinitesimal jackknife and influence functions are venerable ideas; they have been used to debias predictions, interpret black-box predictions, approximate cross-validation, etc. [Jaeckel 1972; Cook, Weisberg 1980; Belsley+ 1980 ch 2.1; Efron, 1982; Huh, Park 1990; Koh, Liang 2017; Giordano+ 2019a]
- They have been used for local Bayesian robustness to changes in prior or likelihood. The covariance form is well known. [Diaconis, Freedman, 1986; Ruggeri, Wasserman, 1993; Gustafson, 1996; Giordano+ 2018; Mohamed+ 2020]. See also Billingsley, 2008, Thm 5.4 & Giordano+ 2023 rejoinder
- To the best of our knowledge, no one has proposed or analyzed using the IJ/influence scores for assessing the frequentist variability of MCMC-estimated posterior means

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (*B*=200): 47 minutes, IJ: <1 second

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (*B*=200): 47 minutes, IJ: <1 second

$g(\theta) =$

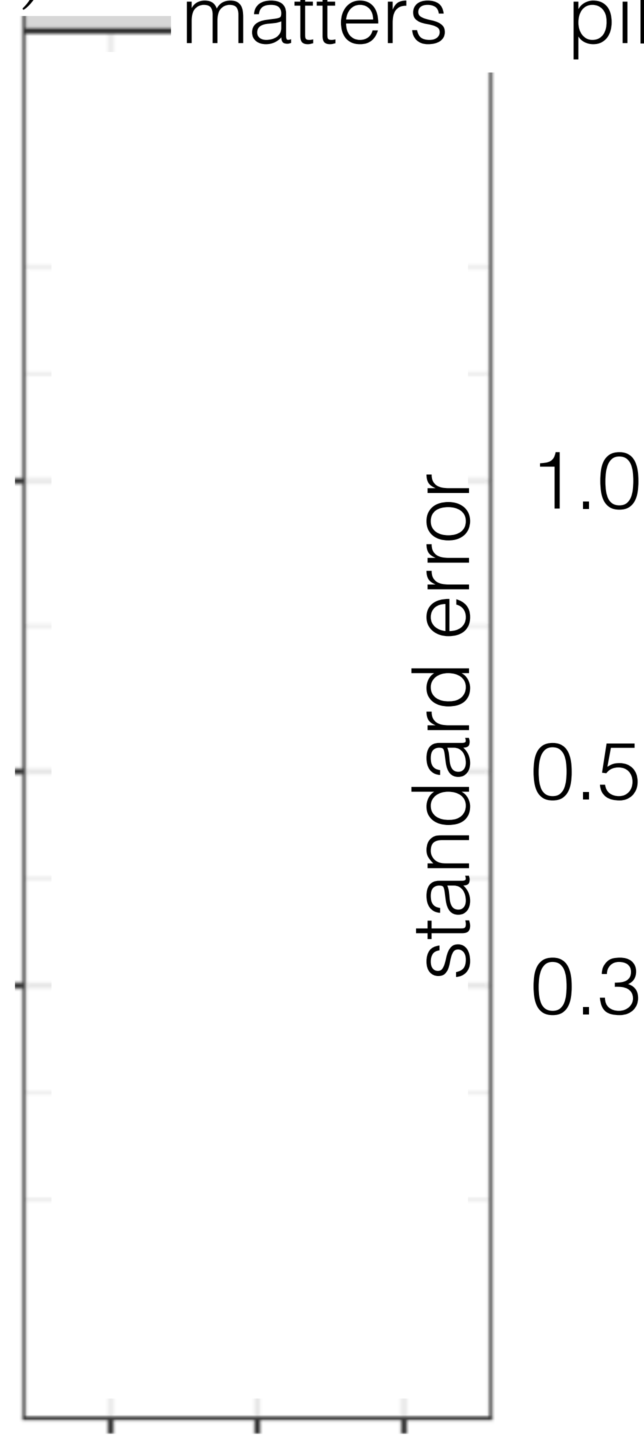
how much individual pilot skill matters	how much scenario matters relative to pilot skill	how much training matters relative to pilot skill

Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (B=200): 47 minutes, IJ: <1 second

$g(\theta) =$ how much individual pilot skill matters | how much scenario matters relative to pilot skill | how much training matters relative to pilot skill

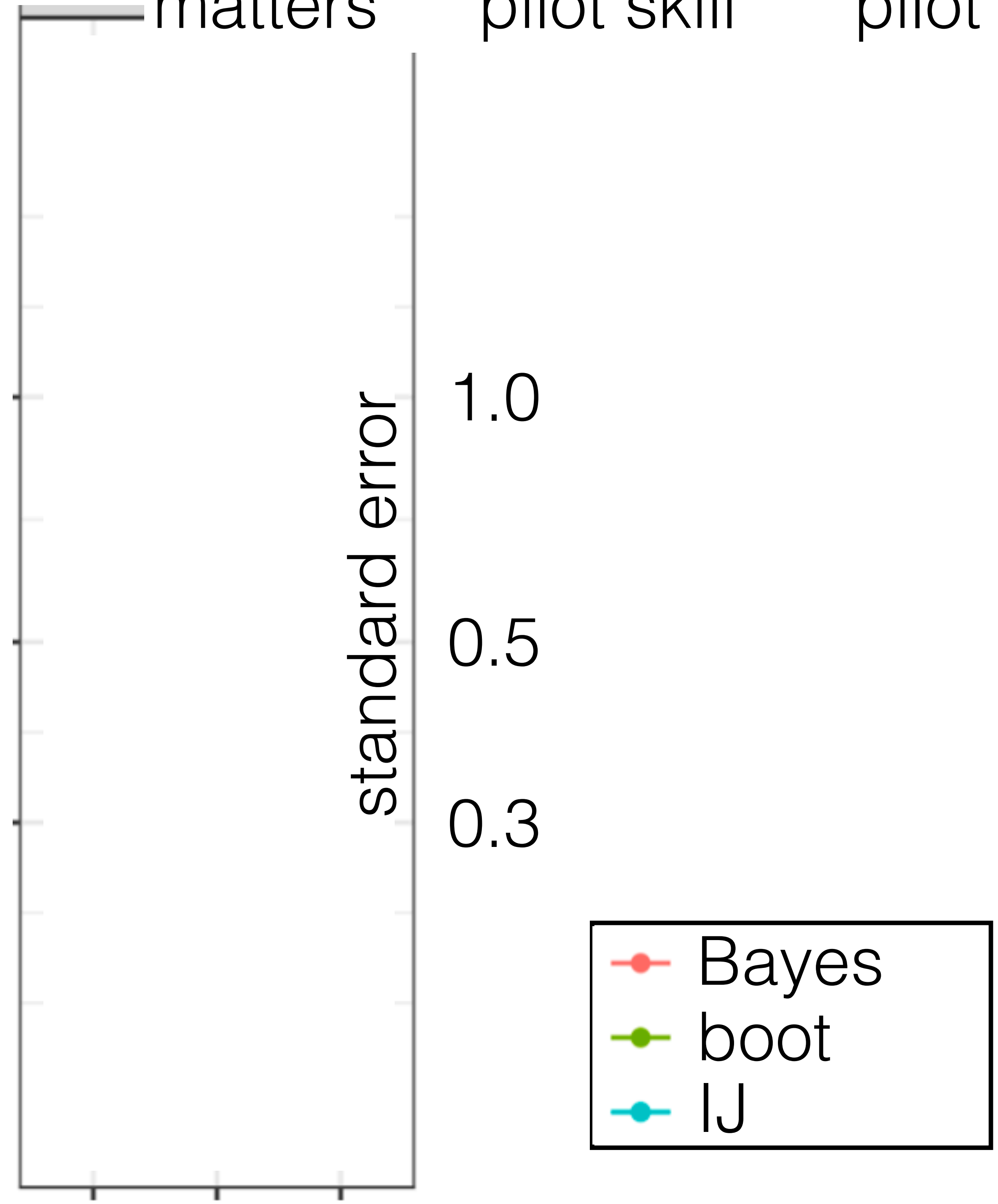


Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (B=200): 47 minutes, IJ: <1 second

$g(\theta) =$ how much individual pilot skill matters | how much scenario matters relative to pilot skill | how much training matters relative to pilot skill

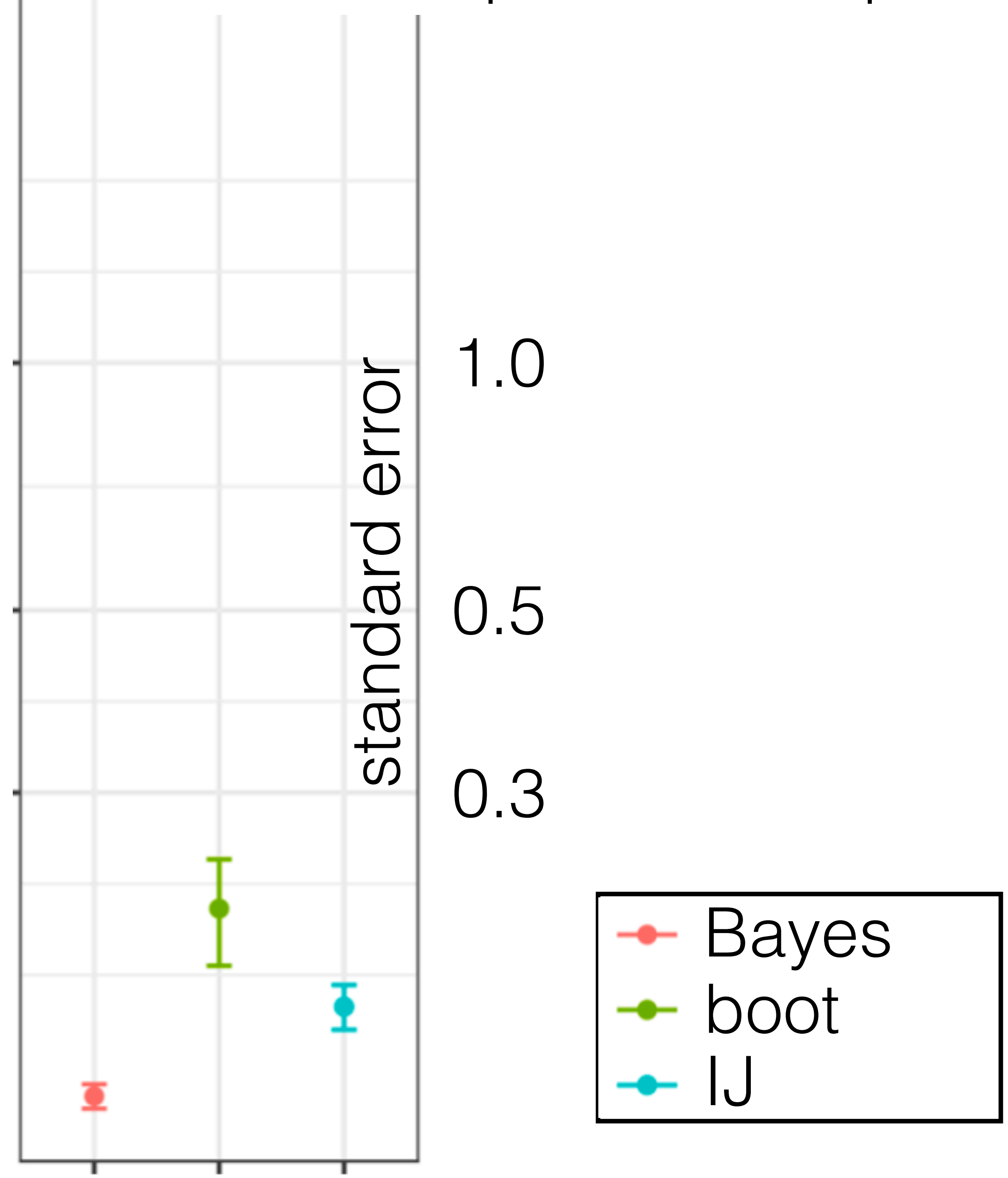


Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (B=200): 47 minutes, IJ: <1 second

$g(\theta) =$ how much individual pilot skill matters | how much scenario matters relative to pilot skill | how much training matters relative to pilot skill

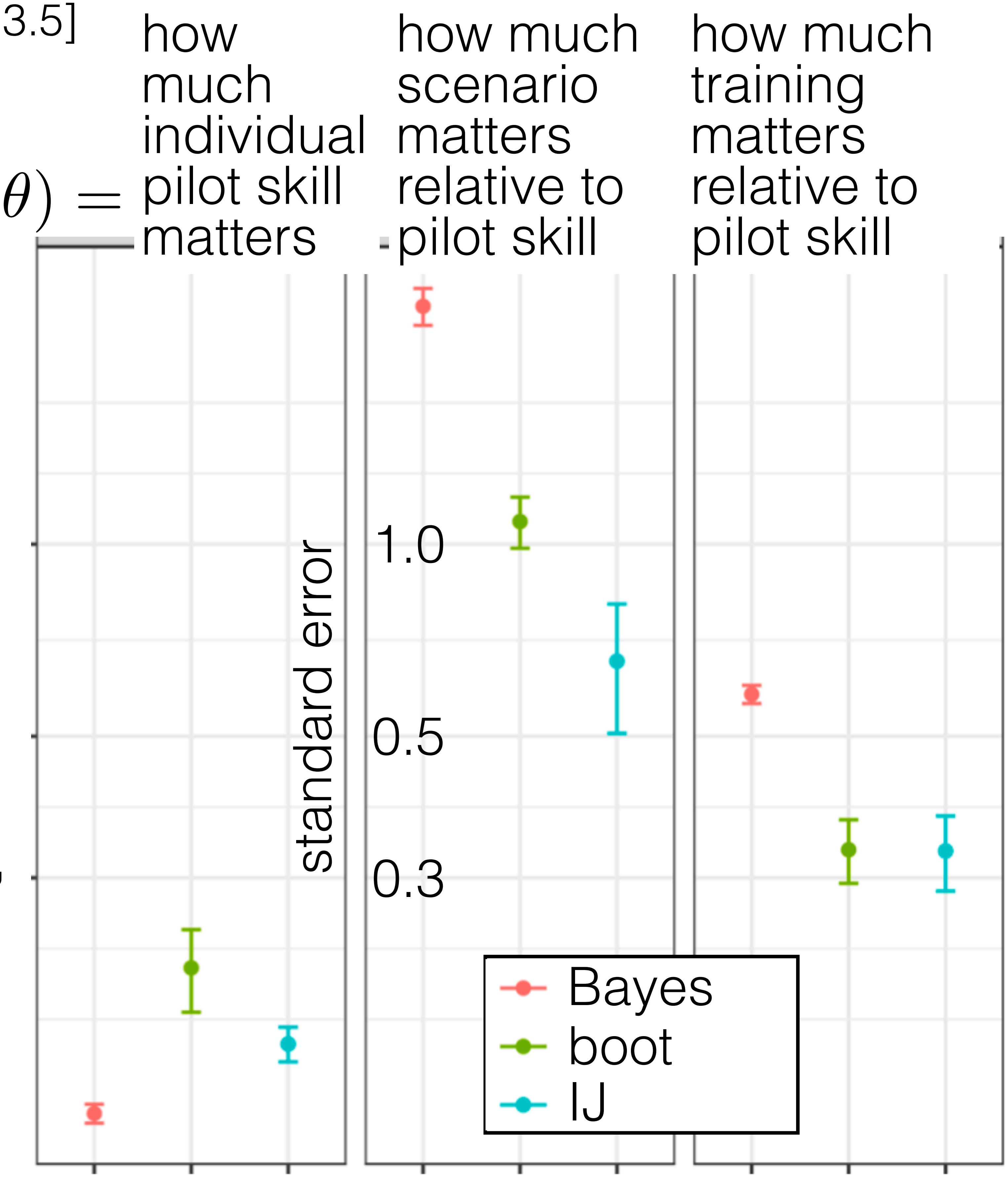


Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (B=200): 47 minutes, IJ: <1 second

$g(\theta) =$

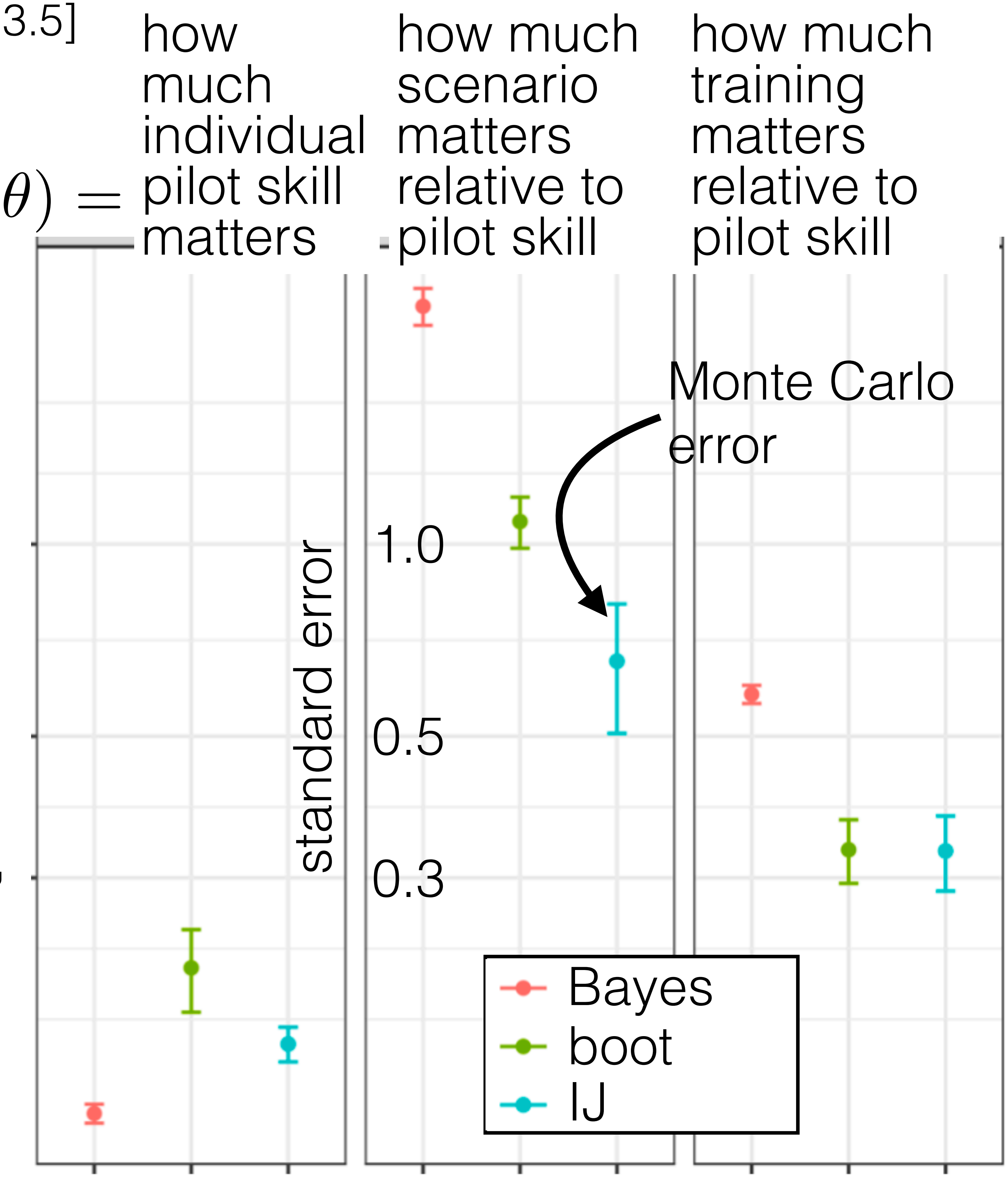


Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (B=200): 47 minutes, IJ: <1 second

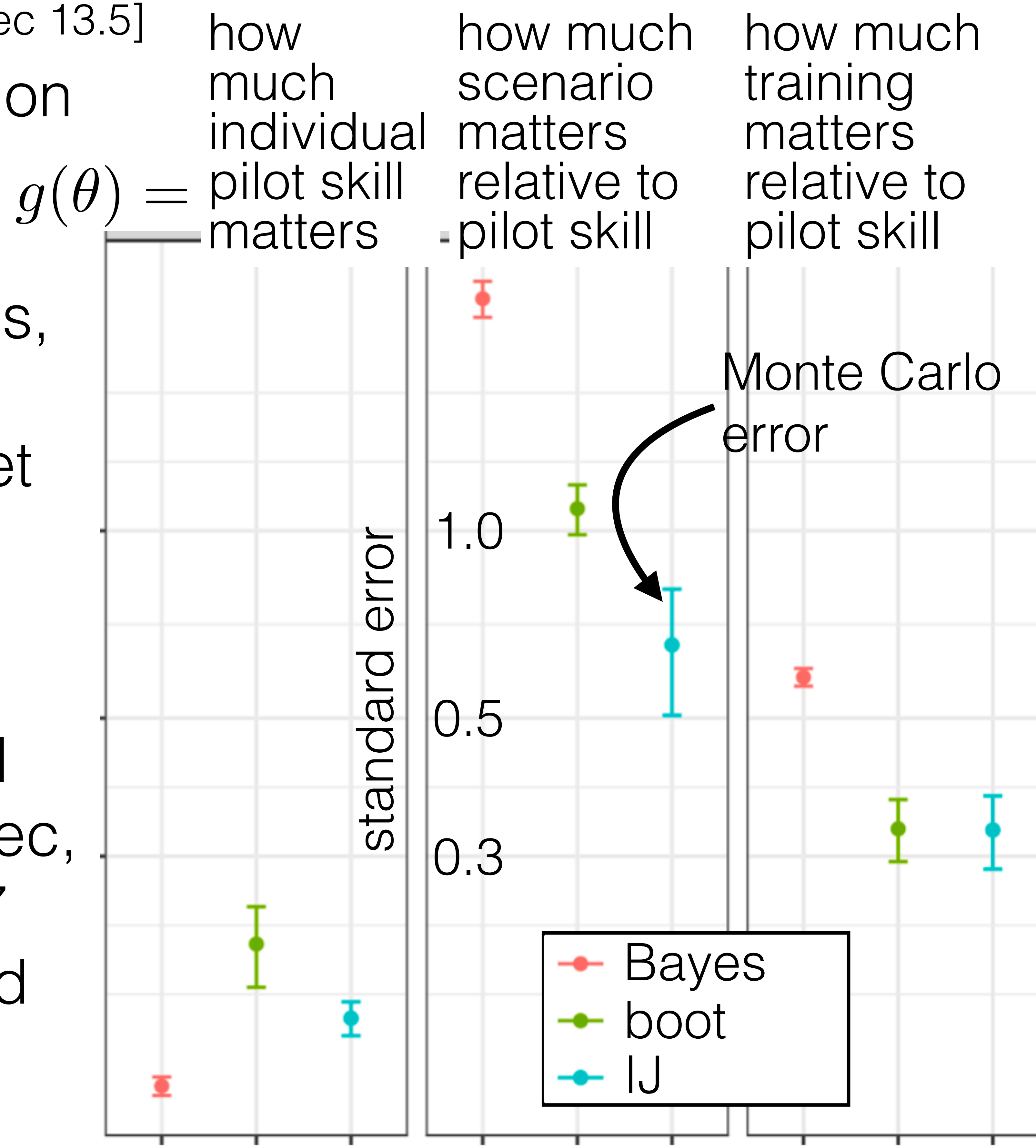
$g(\theta) =$



Another example: pilots analysis

[Gawron et al 03; Gelman, Hill 07 Sec 13.5]

- Performance of pilots on flight simulators
- 40 pilots; non-nested model with 8 scenarios, 5 training levels
- When we run lme4, get 0 for the treatment variance fit, but rstanarm MCMC diagnostics look good
- Original MCMC: 6.2 sec, bootstrap (B=200): 47 minutes, IJ: <1 second
- IJ & bootstrap share qualitative conclusion



Insights from theory

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \operatorname{Var}_{\mathbb{F}(X)} \mathbb{E}_{p(\theta|X)}[g(\theta)] \to V$

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \operatorname*{Var}_{\mathbb{F}(X)} \mathop{\mathbb{E}}_{p(\theta|X)}[g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \operatorname*{Var}_{p(\theta|X)}[g(\theta)] \to U$

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let V be such that: $N \operatorname*{Var}_{\mathbb{F}(X)} \mathop{\mathbb{E}}_{p(\theta|X)} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \operatorname*{Var}_{p(\theta|X)} [g(\theta)] \to U$
  - $U = V$ under correct specification, else generally different

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \operatorname*{Var}_{\mathbb{F}(X)} \mathop{\mathbb{E}}_{p(\theta|X)} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \operatorname*{Var}_{p(\theta|X)} [g(\theta)] \to U$
  - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \operatorname*{Var}_{\mathbb{F}(X)} \mathop{\mathbb{E}}_{p(\theta|X)} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \operatorname*{Var}_{p(\theta|X)} [g(\theta)] \to U$
  - *U* = *V* under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let V be such that: $N \underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \underset{p(\theta|X)}{\mathrm{Var}} [g(\theta)] \to U$
  - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$
- **OK, but we often have "global" and "local" parameters.**

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let V be such that: $N \underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \underset{p(\theta|X)}{\mathrm{Var}} [g(\theta)] \to U$
    - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$
- **OK, but we often have “global” and “local” parameters.**
    - Global: fixed dimension, local: dimension grows with N

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \underset{p(\theta|X)}{\mathrm{Var}} [g(\theta)] \to U$
  - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$
- **OK, but we often have "global" and "local" parameters.**
  - Global: fixed dimension, local: dimension grows with *N*
  - **We show**: Even when the global posterior concentrates, the IJ estimator is inconsistent

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \underset{p(\theta|X)}{\mathrm{Var}} [g(\theta)] \to U$
    - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$
- **OK, but we often have "global" and "local" parameters.**
    - Global: fixed dimension, local: dimension grows with *N*
    - **We show**: Even when the global posterior concentrates, the IJ estimator is inconsistent
    - Suggestive that bootstrap may behave same (not proven)

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \underset{p(\theta|X)}{\mathrm{Var}} [g(\theta)] \to U$
    - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$
- **OK, but we often have "global" and "local" parameters.**
    - Global: fixed dimension, local: dimension grows with *N*
    - **We show**: Even when the global posterior concentrates, the IJ estimator is inconsistent
    - Suggestive that bootstrap may behave same (not proven)
    - Still have reason to expect decent practical performance

Insights from theory

- **First, take *N* growing and parameter dimension fixed.**
- Take regularity assumptions & ignore posterior approx.
- Let *V* be such that: $N \underset{\mathbb{F}(X)}{\mathrm{Var}} \underset{p(\theta|X)}{\mathbb{E}} [g(\theta)] \to V$
- Classic Bayesian Central Limit Theorem: $N \underset{p(\theta|X)}{\mathrm{Var}} [g(\theta)] \to U$
  - $U = V$ under correct specification, else generally different
- Huggins and Miller 2019 show bootstrap estimator $\to V$
- **We show** IJ estimator $\to V$
- **OK, but we often have "global" and "local" parameters.**
  - Global: fixed dimension, local: dimension grows with *N*
  - **We show**: Even when the global posterior concentrates, the IJ estimator is inconsistent
  - Suggestive that bootstrap may behave same (not proven)
  - Still have reason to expect decent practical performance
  - See paper for simulation experiments

Conclusions

- We propose and analyze using the infinitesimal jackknife to estimate frequentist sampling uncertainty of a Bayesian estimator.
- We support our method's accuracy, speed, & automation (+ limitations) with simulated & real experiments and theory
- **Paper**: Giordano, Broderick. The Bayesian Infinitesimal Jackknife for Variance. In preparation.

`arxiv.org/abs/2305.06466`

Conclusions

- We propose and analyze using the infinitesimal jackknife to estimate frequentist sampling uncertainty of a Bayesian estimator.
- We support our method's accuracy, speed, & automation (+ limitations) with simulated & real experiments and theory
- **Paper**: Giordano, Broderick. The Bayesian Infinitesimal Jackknife for Variance. In preparation.
  `arxiv.org/abs/2305.06466`
- Robustness isn't a monolith.

Conclusions

- We propose and analyze using the infinitesimal jackknife to estimate frequentist sampling uncertainty of a Bayesian estimator.
- We support our method's accuracy, speed, & automation (+ limitations) with simulated & real experiments and theory
- **Paper**: Giordano, Broderick. The Bayesian Infinitesimal Jackknife for Variance. In preparation.

  `arxiv.org/abs/2305.06466`
- Robustness isn't a monolith.
  - Giordano et al. Covariances, robustness, and variational Bayes. *JMLR* 2018 & Giordano et al. Evaluating Sensitivity to the Stick-Breaking Prior in Bayesian Nonparametrics. *BA Discussion Paper*, 2023.

Conclusions

- We propose and analyze using the infinitesimal jackknife to estimate frequentist sampling uncertainty of a Bayesian estimator.
- We support our method's accuracy, speed, & automation (+ limitations) with simulated & real experiments and theory
- **Paper**: Giordano, Broderick. The Bayesian Infinitesimal Jackknife for Variance. In preparation.

  `arxiv.org/abs/2305.06466`

- Robustness isn't a monolith.
  - Giordano et al. Covariances, robustness, and variational Bayes. *JMLR* 2018 & Giordano et al. Evaluating Sensitivity to the Stick-Breaking Prior in Bayesian Nonparametrics. *BA Discussion Paper*, 2023. (Check out our rejoinder for an introduction to using influence scores for local robustness approximations.)

Conclusions

- We propose and analyze using the infinitesimal jackknife to estimate frequentist sampling uncertainty of a Bayesian estimator.
- We support our method's accuracy, speed, & automation (+ limitations) with simulated & real experiments and theory
- **Paper**: Giordano, Broderick. The Bayesian Infinitesimal Jackknife for Variance. In preparation.

  `arxiv.org/abs/2305.06466`

- Robustness isn't a monolith.
  - Giordano et al. Covariances, robustness, and variational Bayes. *JMLR* 2018 & Giordano et al. Evaluating Sensitivity to the Stick-Breaking Prior in Bayesian Nonparametrics. *BA Discussion Paper*, 2023. (Check out our rejoinder for an introduction to using influence scores for local robustness approximations.)
  - Broderick, Giordano, Meager. An Automatic Finite-Sample Robustness Metric: When Can Dropping a Little Data Make a Big Difference? arxiv.org/abs/2011.14999 (alphabetical)