Bayesian Data Analysis, class 5b

Andrew Gelman

Chapter 6: Model checking (part 1)

Discussion of homework due beginning of Class 5b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 5b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 5b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 5b

- Theory problem
- Computing problem
- Applied problem

Theory problem

- $2J$ parameters $\theta_1, \ldots, \theta_{2J}$ clustered into two groups, with exactly half being drawn from a $N(1,1)$ distribution, and the other half being drawn from a $N(-1,1)$ distribution, but we have not observed which parameters come from which distribution
- Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?
- Show that this distribution cannot be written as a mixture of independent and identically distributed components.
- de Finetti's theorem

Theory problem

- $2J$ parameters $\theta_1, \ldots, \theta_{2J}$ clustered into two groups, with exactly half being drawn from a $N(1, 1)$ distribution, and the other half being drawn from a $N(-1, 1)$ distribution, but we have not observed which parameters come from which distribution
- Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?
- Show that this distribution cannot be written as a mixture of independent and identically distributed components.
- de Finetti's theorem

Theory problem

- $2J$ parameters $\theta_1, \ldots, \theta_{2J}$ clustered into two groups, with exactly half being drawn from a $N(1, 1)$ distribution, and the other half being drawn from a $N(-1, 1)$ distribution, but we have not observed which parameters come from which distribution
- Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?
- Show that this distribution cannot be written as a mixture of independent and identically distributed components.
- de Finetti's theorem

Theory problem

- $2J$ parameters $\theta_1, \ldots, \theta_{2J}$ clustered into two groups, with exactly half being drawn from a $N(1,1)$ distribution, and the other half being drawn from a $N(-1,1)$ distribution, but we have not observed which parameters come from which distribution
- Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?
- Show that this distribution cannot be written as a mixture of independent and identically distributed components.
- de Finetti's theorem

Theory problem

- $2J$ parameters $\theta_1, \ldots, \theta_{2J}$ clustered into two groups, with exactly half being drawn from a $N(1, 1)$ distribution, and the other half being drawn from a $N(-1, 1)$ distribution, but we have not observed which parameters come from which distribution
- Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?
- Show that this distribution cannot be written as a mixture of independent and identically distributed components.
- de Finetti's theorem

Computing problem

- Simulation of a stochastic process

Computing problem

- Simulation of a stochastic process

Applied problem

- Stochastic learning model with 2 or 3 parameters
- Weakly informative prior
- Simulate data from fake dogs

Applied problem

- Stochastic learning model with 2 or 3 parameters
- Weakly informative prior
- Simulate data from fake dogs

Applied problem

- Stochastic learning model with 2 or 3 parameters
- Weakly informative prior
- Simulate data from fake dogs

Applied problem

- Stochastic learning model with 2 or 3 parameters
- Weakly informative prior
- Simulate data from fake dogs

6. Model checking (part 1)

- Comparing estimates and predictions to substantive knowledge
- Comparing predictions to observed data
- Graphical and numerical tests
- P-values and u-values

- Comparing estimates and predictions to substantive knowledge
- Comparing predictions to observed data
- Graphical and numerical tests
- P-values and u-values

6. Model checking (part 1)

- Comparing estimates and predictions to substantive knowledge
- Comparing predictions to observed data
- Graphical and numerical tests
- P-values and u-values

6. Model checking (part 1)

- Comparing estimates and predictions to substantive knowledge
- Comparing predictions to observed data
- Graphical and numerical tests
- P-values and u-values

6. Model checking (part 1)

- Comparing estimates and predictions to substantive knowledge
- Comparing predictions to observed data
- Graphical and numerical tests
- P-values and u-values

6.1. The place of model checking in applied Bayesian statistics

- With great power comes great responsibility
- Sensitivity analysis
- All models are false
- Real Bayes vs. super-Bayes

6.1. The place of model checking in applied Bayesian statistics

- With great power comes great responsibility
- Sensitivity analysis
- All models are false
- Real Bayes vs. super-Bayes

6.1. The place of model checking in applied Bayesian statistics

- With great power comes great responsibility
- Sensitivity analysis
- All models are false
- Real Bayes vs. super-Bayes

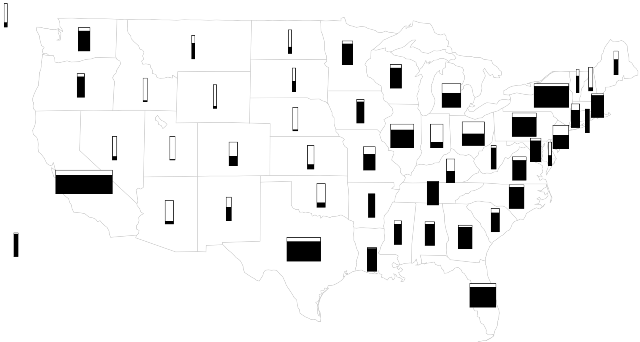
6.1. The place of model checking in applied Bayesian statistics

- With great power comes great responsibility
- Sensitivity analysis
- All models are false
- Real Bayes vs. super-Bayes

6.1. The place of model checking in applied Bayesian statistics

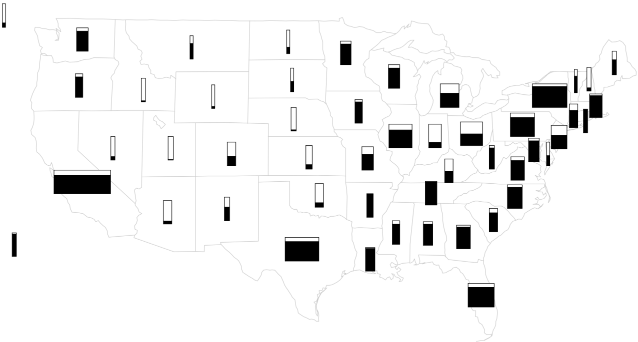
- With great power comes great responsibility
- Sensitivity analysis
- All models are false
- Real Bayes vs. super-Bayes

6.2. Do the inferences from the model make sense?



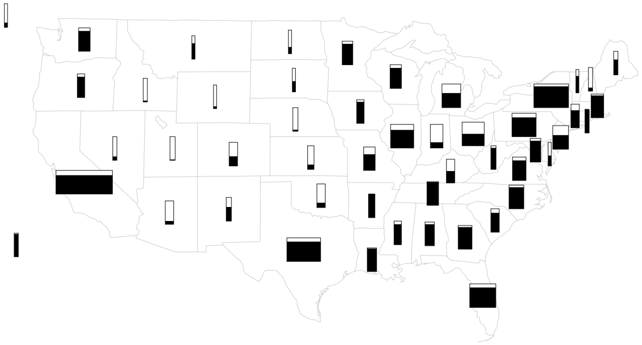
- Also, remember the 8-pound liver from class 1a
- We're talking about prior information!

6.2. Do the inferences from the model make sense?



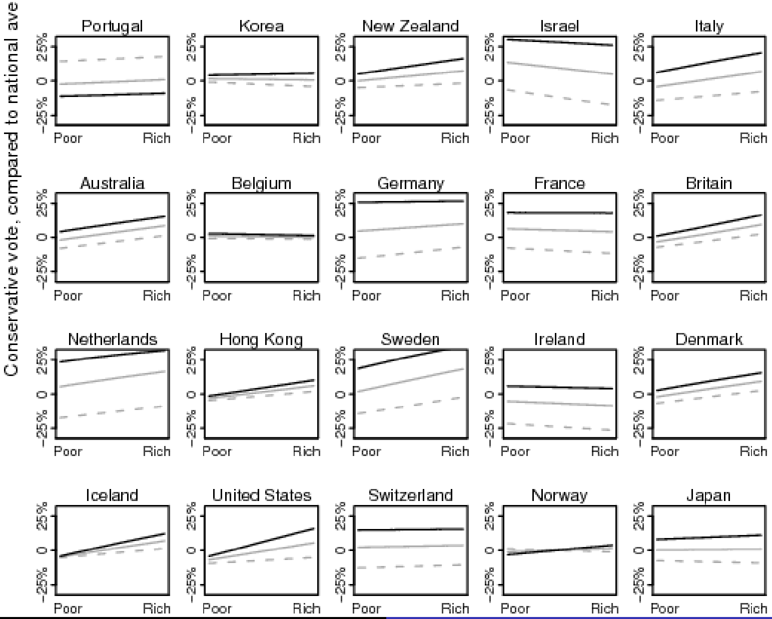
- Also, remember the 8-pound liver from class 1a
- We're talking about prior information!

6.2. Do the inferences from the model make sense?



- Also, remember the 8-pound liver from class 1a
- We're talking about prior information!

What's the matter with Portugal?



6.3. Posterior predictive checking

- Compare observed data to replications simulated from the model
- Replication and p-values
- Several examples

6.3. Posterior predictive checking

- Compare observed data to replications simulated from the model
- Replication and *p*-values
- Several examples

6.3. Posterior predictive checking

- Compare observed data to replications simulated from the model
- Replication and p-values
- Several examples

6.3. Posterior predictive checking

- Compare observed data to replications simulated from the model
- Replication and p-values
- Several examples

Replications and p-values

- $\theta, y, y^{\mathrm{rep}}$
- Classical p-value is $p\text{-value}(y|\theta) = \Pr\left(T(y^{\mathrm{rep}}) \geq T(y) | \theta, y\right)$
- Bayesian posterior p-value is $\Pr\left(T(y^{\mathrm{rep}}) \geq T(y) | y\right) = \int p\text{-value}(y|\theta) p(\theta|y) d\theta$
- Compute using simulation

Replications and p-values

- $\theta, y, y^{\text{rep}}$
- Classical p-value is $p\text{-value}(y|\theta) = \Pr\left(T(y^{\text{rep}}) \geq T(y)|\theta, y\right)$
- Bayesian posterior p-value is $\Pr\left(T(y^{\text{rep}}) \geq T(y)|y\right) = \int p\text{-value}(y|\theta)p(\theta|y)d\theta$
- Compute using simulation

Replications and p-values

- $\theta, y, y^{\rm rep}$
- Classical p-value is $p\text{-value}(y|\theta) = \Pr\left(T(y^{\rm rep}) \geq T(y)|\theta, y\right)$
- Bayesian posterior p-value is $\Pr\left(T(y^{\rm rep}) \geq T(y)|y\right) = \int p\text{-value}(y|\theta)p(\theta|y)d\theta$
- Compute using simulation

Replications and p-values

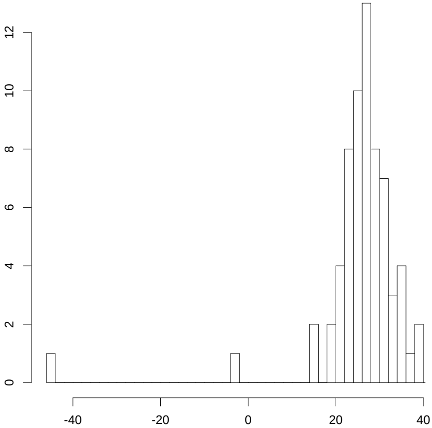
- $\theta, y, y^{\mathrm{rep}}$
- Classical p-value is $p\text{-value}(y|\theta) = \Pr\left(T(y^{\mathrm{rep}}) \geq T(y) | \theta, y\right)$
- Bayesian posterior p-value is
  $\Pr\left(T(y^{\mathrm{rep}}) \geq T(y) | y\right) = \int p\text{-value}(y|\theta) p(\theta|y) d\theta$
- Compute using simulation

Replications and p-values

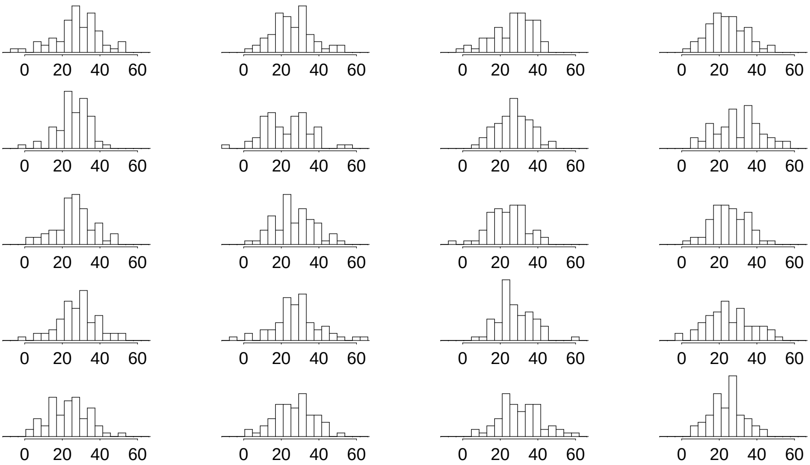
- $\theta, y, y^{\text{rep}}$
- Classical p-value is $p\text{-value}(y|\theta) = \Pr\left(T(y^{\text{rep}}) \geq T(y) | \theta, y\right)$
- Bayesian posterior p-value is $\Pr\left(T(y^{\text{rep}}) \geq T(y) | y\right) = \int p\text{-value}(y|\theta) p(\theta|y) d\theta$
- Compute using simulation

A dataset

A normal distribution is fit to the following data:

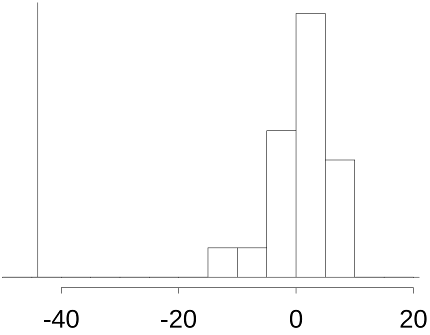


20 replications under the fitted model



Comparison using a numerical test statistic

$T(y) = \min_i y_i$:



Numerical posterior predictive checks

- Limitations of the likelihood principle
- Multiple comparisons
- Interpreting *p*-values

Numerical posterior predictive checks

- Limitations of the likelihood principle
- Multiple comparisons
- Interpreting p-values

Numerical posterior predictive checks

- Limitations of the likelihood principle
- Multiple comparisons
- Interpreting p-values

Numerical posterior predictive checks

- Limitations of the likelihood principle
- Multiple comparisons
- Interpreting p-values

Connections to classical testing

- p-values and u-values
- The purpose of a model check
- Going beyond pivotal test statistics
- Two examples

Connections to classical testing

- p-values and u-values
- The purpose of a model check
- Going beyond pivotal test statistics
- Two examples

Connections to classical testing

- p-values and u-values
- The purpose of a model check
- Going beyond pivotal test statistics
- Two examples

Connections to classical testing

- p-values and u-values
- The purpose of a model check
- Going beyond pivotal test statistics
- Two examples

Connections to classical testing

- p-values and u-values
- The purpose of a model check
- Going beyond pivotal test statistics
- Two examples

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:

- If model is correct, test will essentially never "reject"

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
- If model is correct, test will essentially never "reject"

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
- If model is correct, test will essentially never "reject"

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta | y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}} | y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta|y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}}|y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(-\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta | y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}} | y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(-\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta | y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}} | y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(-\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"
  - p-value will always be close to 0.5

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta | y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}} | y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(-\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"
  - p-value will always be close to 0.5

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta | y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}} | y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(-\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"
  - p-value will always be close to 0.5

Example where the posterior predictive p-value is stuck near 0.5 and this makes sense

- Data $y \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) = y$ (the sample mean)
- To compute posterior predictive check:
  - Posterior distribution: $\theta | y \sim \mathrm{N}(0.9999y, .9999)$
  - Posterior predictive distribution: $y^{\mathrm{rep}} | y \sim \mathrm{N}(0.9999y, 1.9999)$
  - Compare to data:

$$p\text{-value} = \Phi\left(-\frac{y - \mathrm{E}(y^{\mathrm{rep}}|y)}{\mathrm{sd}(y^{\mathrm{rep}}|y)}\right) = \Phi(-y/14{,}000)$$

- If model is correct, test will essentially never "reject"
  - p-value will always be close to 0.5

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
  - $T(y)$ is a random variable (with 99% of the observations missing)
  - $\Pr(T(y^{\rm rep}) > T(y)) \approx 0.5$ for just about any $y_1, \ldots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
  - $T(y)$ is a random variable (with 99% of the observations missing)
  - $\Pr(T(y^{\mathrm{rep}}) > T(y)) \approx 0.5$ for just about any $y_1, \ldots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
  - $T(y)$ is a random variable (with 99% of the observations missing)
  - $\Pr(T(y^{\mathrm{rep}}) > T(y)) \approx 0.5$ for just about any $y_1, \ldots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
  - $T(y)$ is a random variable (with 99% of the observations missing)
  - $\Pr(T(y^{\mathrm{rep}}) > T(y)) \approx 0.5$ for just about any $y_1, \ldots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \dots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \dots, y_{10}$
  - $T(y)$ is a random variable (with 99% of the observations missing)
  - $\Pr(T(y^{\mathrm{rep}}) > T(y)) \approx 0.5$ for just about any $y_1, \dots, y_{10}$
- How to understand this result?

Example where the posterior predictive p-value is stuck near 0.5 and this seems bad

- Data $y_1, \ldots, y_{1000} \sim \mathrm{N}(\theta, 1)$
- Prior $\theta \sim \mathrm{N}(0, A^2)$ with $A = 100$
- Test statistic $T(y) =$ sample skewness of the 1000 observations
- Now suppose that all you've observed is $y_1, \ldots, y_{10}$
  - $T(y)$ is a random variable (with 99% of the observations missing)
  - $\Pr(T(y^{\mathrm{rep}}) > T(y)) \approx 0.5$ for just about any $y_1, \ldots, y_{10}$
- How to understand this result?

Homework due beginning of class 6b

- All assignments are at `http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf`
  - Theory problem: Bayes factors
  - Computing problem: Stan
  - Applied problem: Poststratification to go from sample to population

Homework due beginning of class 6b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Bayes factors
  - Computing problem: Stan
  - Applied problem: Poststratification to go from sample to population

Homework due beginning of class 6b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Bayes factors
  - Computing problem: Stan
  - Applied problem: Poststratification to go from sample to population

Homework due beginning of class 6b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Bayes factors
  - Computing problem: Stan
  - Applied problem: Poststratification to go from sample to population

Homework due beginning of class 6b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Bayes factors
  - Computing problem: Stan
  - Applied problem: Poststratification to go from sample to population