Bayesian Data Analysis, class 2b

Andrew Gelman

Chapter 2. Single-parameter models (part 2)

Discussion of homework due beginning of Class 2b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 2b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 2b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 2b

- Theory problem
- Computing problem
- Applied problem

Theory problem

- Show that if $y|\theta$ is exponentially distributed with rate θ, then the gamma prior distribution is conjugate ...
- Show that the equivalent prior specification for the mean, $\phi = 1/\theta$, is inverse-gamma. (That is, derive the latter density function.)
- The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

Theory problem

- Show that if $y|\theta$ is exponentially distributed with rate θ, then the gamma prior distribution is conjugate ...
- Show that the equivalent prior specification for the mean, $\phi = 1/\theta$, is inverse-gamma. (That is, derive the latter density function.)
- The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

Theory problem

- Show that if $y|\theta$ is exponentially distributed with rate θ, then the gamma prior distribution is conjugate ...
- Show that the equivalent prior specification for the mean, $\phi = 1/\theta$, is inverse-gamma. (That is, derive the latter density function.)
- The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

Theory problem

- Show that if $y|\theta$ is exponentially distributed with rate θ, then the gamma prior distribution is conjugate ...
- Show that the equivalent prior specification for the mean, $\phi = 1/\theta$, is inverse-gamma. (That is, derive the latter density function.)
- The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

Computing problem

- Your task is to estimate the percentatge of the (adult) population in each state (excluding Alaska and Hawaii) who label themselves as "very liberal," replicating the procedure that was used in Section 2.8 to estimate cancer rates ...
- This exercise has four challenges: first, manipulating the data in order to get the totals by state; second, replicating the calculations for estimating the parameters of the prior distribution; third, doing the Bayesian analysis by state; and fourth, making the graphs.

Computing problem

- Your task is to estimate the percentatge of the (adult) population in each state (excluding Alaska and Hawaii) who label themselves as "very liberal," replicating the procedure that was used in Section 2.8 to estimate cancer rates . . .
- This exercise has four challenges: first, manipulating the data in order to get the totals by state; second, replicating the calculations for estimating the parameters of the prior distribution; third, doing the Bayesian analysis by state; and fourth, making the graphs.

Computing problem

- Your task is to estimate the percentatge of the (adult) population in each state (excluding Alaska and Hawaii) who label themselves as "very liberal," replicating the procedure that was used in Section 2.8 to estimate cancer rates . . .
- This exercise has four challenges: first, manipulating the data in order to get the totals by state; second, replicating the calculations for estimating the parameters of the prior distribution; third, doing the Bayesian analysis by state; and fourth, making the graphs.

Applied problem

A (hypothetical) study is performed to estimate the effect of a simple training program on basketball free-throw shooting . . . Let θ be the average improvement in success probability. Give . . .

- A noninformative prior,
- A subjective prior based on your best knowledge, and
- A weakly informative prior.

Applied problem

A (hypothetical) study is performed to estimate the effect of a simple training program on basketball free-throw shooting ... Let θ be the average improvement in success probability. Give ...

- A noninformative prior,
- A subjective prior based on your best knowledge, and
- A weakly informative prior.

Applied problem

A (hypothetical) study is performed to estimate the effect of a simple training program on basketball free-throw shooting . . . Let θ be the average improvement in success probability. Give . . .

- A noninformative prior,
- A subjective prior based on your best knowledge, and
- A weakly informative prior.

Applied problem

A (hypothetical) study is performed to estimate the effect of a simple training program on basketball free-throw shooting . . . Let θ be the average improvement in success probability. Give . . .

- A noninformative prior,
- A subjective prior based on your best knowledge, and
- A weakly informative prior.

2. Single-parameter models (part 2)

- Equivalent sample size of prior distribution
- Basic single-parameter models (beyond normal location and binomial probability)
- One prior distribution applied to many different cases
- Noninformative priors
- Weakly informative priors

2. Single-parameter models (part 2)

- Equivalent sample size of prior distribution
- Basic single-parameter models (beyond normal location and binomial probability)
- One prior distribution applied to many different cases
- Noninformative priors
- Weakly informative priors

2. Single-parameter models (part 2)

- Equivalent sample size of prior distribution
- Basic single-parameter models (beyond normal location and binomial probability)
- One prior distribution applied to many different cases
- Noninformative priors
- Weakly informative priors

2. Single-parameter models (part 2)

- Equivalent sample size of prior distribution
- Basic single-parameter models (beyond normal location and binomial probability)
- One prior distribution applied to many different cases
- Noninformative priors
- Weakly informative priors

2. Single-parameter models (part 2)

- Equivalent sample size of prior distribution
- Basic single-parameter models (beyond normal location and binomial probability)
- One prior distribution applied to many different cases
- Noninformative priors
- Weakly informative priors

2. Single-parameter models (part 2)

- Equivalent sample size of prior distribution
- Basic single-parameter models (beyond normal location and binomial probability)
- One prior distribution applied to many different cases
- Noninformative priors
- Weakly informative priors

Another sex-ratio example

- The data
- A prior distribution
- Equivalent sample size

Another sex-ratio example

- The data
- A prior distribution
- Equivalent sample size

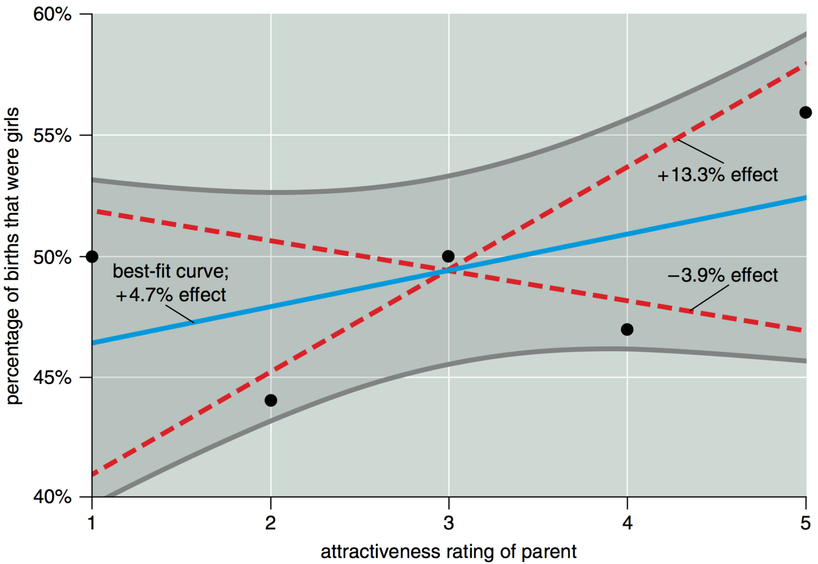
Another sex-ratio example

- The data
- A prior distribution
- Equivalent sample size

Another sex-ratio example

- The data
- A prior distribution
- Equivalent sample size

Beauty of parents and sex ratio of children



Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:

- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
    Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
    s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
    Equivalent info: $0.003 = 0.5\sqrt{6/n} \rightarrow n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3)+0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \rightarrow n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \ldots n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \ldots n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \ldots n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \ldots n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \ldots n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

Pr(girl), comparing beautiful to ugly parents

- Data: difference in Pr(girl) estimated from 3000 respondents
  - 0.08 ± 0.03 (selected comparison)
  - 0.047 ± 0.043 (linear regression)
- Prior distribution: $\mathrm{N}(0, 0.003^2)$
- Equivalent sample size:
  - Consider a survey with n parents
  - Compare sex ratio of prettiest $n/3$ to ugliest $n/3$
  - s.e. is $\sqrt{0.5^2/(n/3) + 0.5^2/(n/3)} = 0.5\sqrt{6/n}$
  - Equivalent info: $0.003 = 0.5\sqrt{6/n} \ldots n = 166{,}000$
- A study with $n = 166{,}000$ people would be weighted equally with the prior

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2+B\theta+C}$
- Posterior also has the form, $e^{A\theta^2+B\theta+C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.5. Estimating a normal mean with known variance

- Prior $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
- Data $\bar{y} \sim \mathrm{N}(\theta, \sigma^2/n)$
- Prior and likelihood both have the form, $e^{A\theta^2 + B\theta + C}$
- Posterior also has the form, $e^{A\theta^2 + B\theta + C}$
- Posterior is $\theta|y \sim \mathrm{N}\left(\frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}\right)$
- The posterior mean is a weighted average!
- Can derive by including all n factors in the likelihood at once or by updating one at a time

2.6. Other standard single-parameter models

- Important algebra
- Poisson with exposure
  - (Almost) never do $y_i \sim \text{Poisson}(\theta)$ for $i = 1, \ldots, n$
    Instead, do $y_i \sim \text{Poisson}(x_i\theta)$
    Rate θ, exposure x_i

2.6. Other standard single-parameter models

- Important algebra
- Poisson with exposure
  - (Almost) never do $y_i \sim \text{Poisson}(\theta)$ for $i = 1, \ldots, n$
  - Instead, do $y_i \sim \text{Poisson}(x_i\theta)$
  - Rate θ, exposure x_i

2.6. Other standard single-parameter models

- Important algebra
- Poisson with exposure
  - (Almost) never do $y_i \sim \text{Poisson}(\theta)$ for $i = 1, \ldots, n$
  - Instead, do $y_i \sim \text{Poisson}(x_i\theta)$
  - Rate θ, exposure x_i

2.6. Other standard single-parameter models

- Important algebra
- Poisson with exposure
  - (Almost) never do $y_i \sim \text{Poisson}(\theta)$ for $i = 1, \ldots, n$
  - Instead, do $y_i \sim \text{Poisson}(x_i\theta)$
  - Rate θ, exposure x_i

2.6. Other standard single-parameter models

- Important algebra
- Poisson with exposure
  - (Almost) never do $y_i \sim \text{Poisson}(\theta)$ for $i = 1, \ldots, n$
  - Instead, do $y_i \sim \text{Poisson}(x_i\theta)$
  - Rate θ, exposure x_i

2.6. Other standard single-parameter models

- Important algebra
- Poisson with exposure
  - (Almost) never do $y_i \sim \text{Poisson}(\theta)$ for $i = 1, \ldots, n$
  - Instead, do $y_i \sim \text{Poisson}(x_i\theta)$
  - Rate θ, exposure x_i

2.7. Example: informative prior distribution and multilevel structure for estimating cancer rates

- A puzzling pattern in a map
- Raw data y_j/n_j and Bayes estimate $\frac{\alpha+y_j}{\beta+n_j}$
- Estimating the prior distribution from data

2.7. Example: informative prior distribution and multilevel structure for estimating cancer rates

- A puzzling pattern in a map
- Raw data y_j/n_j and Bayes estimate $\frac{\alpha+y_j}{\beta+n_j}$
- Estimating the prior distribution from data

2.7. Example: informative prior distribution and multilevel structure for estimating cancer rates

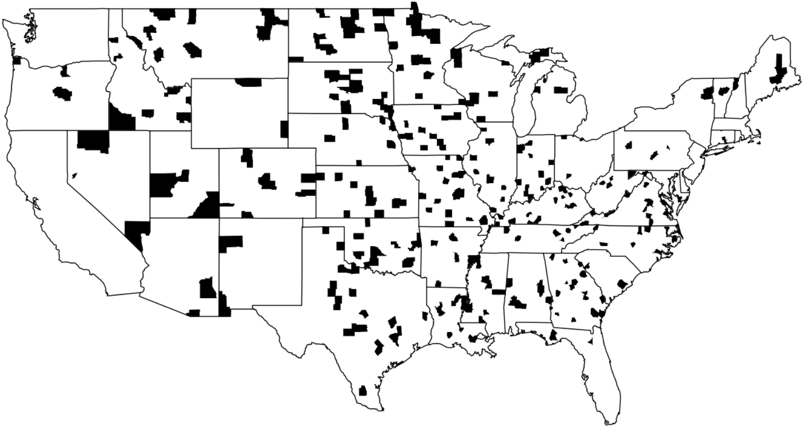
- A puzzling pattern in a map
- Raw data y_j/n_j and Bayes estimate $\frac{\alpha+y_j}{\beta+n_j}$
- Estimating the prior distribution from data

2.7. Example: informative prior distribution and multilevel structure for estimating cancer rates

- A puzzling pattern in a map
- Raw data y_j/n_j and Bayes estimate $\frac{\alpha+y_j}{\beta+n_j}$
- Estimating the prior distribution from data

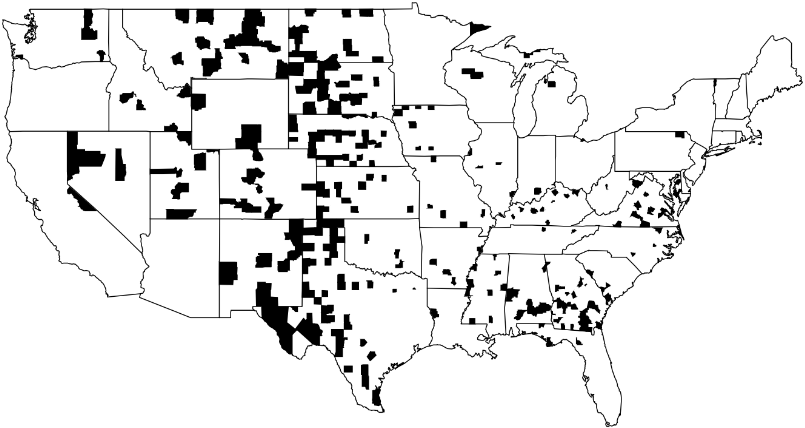
A puzzling pattern in a map

Highest kidney cancer death rates



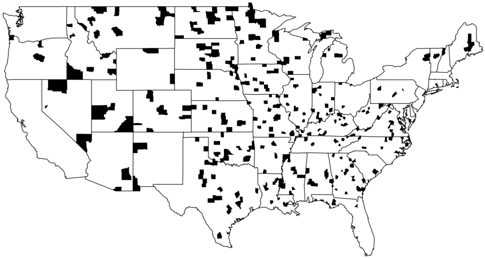
But wait!

Lowest kidney cancer death rates

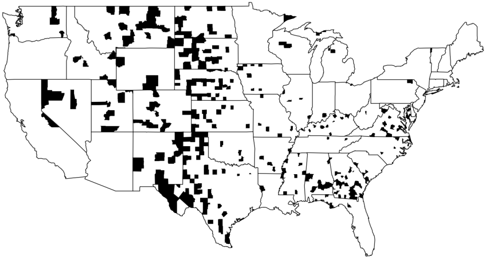


All at once

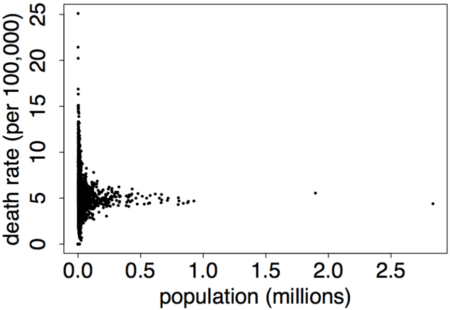
Highest kidney cancer death rates

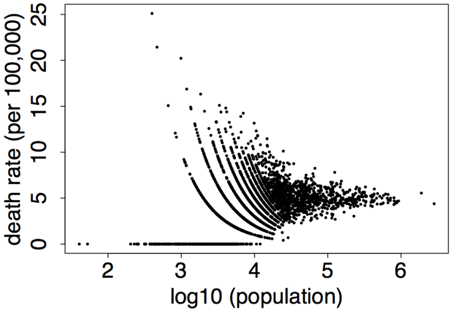


Lowest kidney cancer death rates

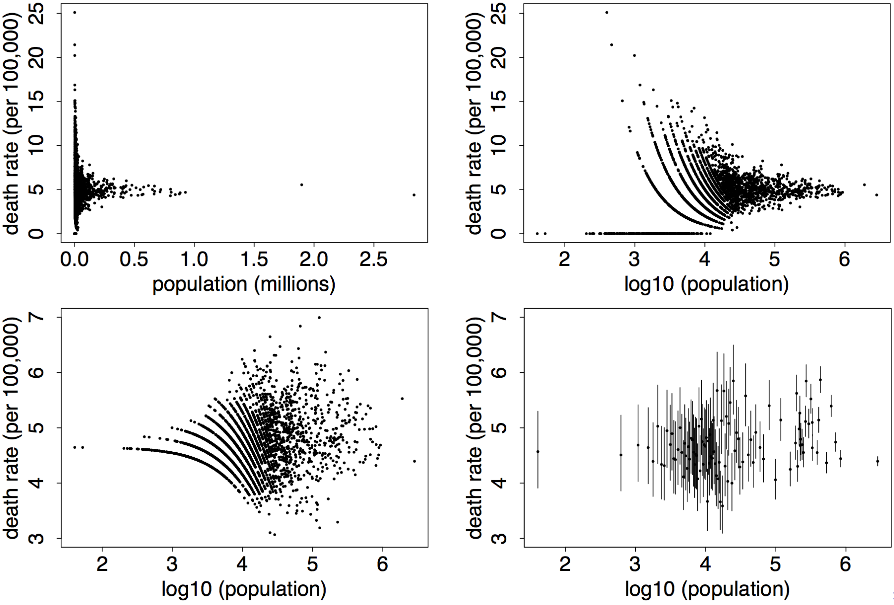


Cancer rates and sample size





Raw data and Bayes estimates



The parable of the two exams

- Option 1: 100 questions
- Option 2: 1 question, score is 0 or 100
- Which exam would you take to maximize the chance of getting the job?

The parable of the two exams

- Option 1: 100 questions
- Option 2: 1 question, score is 0 or 100
- Which exam would you take to maximize the chance of getting the job?

The parable of the two exams

- Option 1: 100 questions
- Option 2: 1 question, score is 0 or 100
- Which exam would you take to maximize the chance of getting the job?

The parable of the two exams

- Option 1: 100 questions
- Option 2: 1 question, score is 0 or 100
- Which exam would you take to maximize the chance of getting the job?

Estimating the prior distribution from data

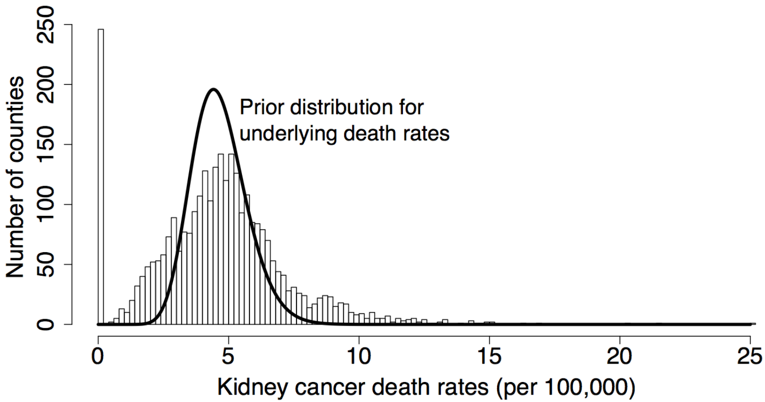


Figure 2.10 *Empirical distribution of the age-adjusted kidney cancer death rates,* $y_j/(10n_j)$*, for the 3071 counties in the U.S., along with the* $Gamma(20, 430000)$ *prior distribution for the underlying cancer rates* θ_j*.*

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
    - What are the assumptions here?
- Flat Gamma$(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat Gamma$(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat Gamma$(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat $\text{Gamma}(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat Gamma$(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat $\text{Gamma}(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat $\text{Gamma}(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Connecting cancer rate example to Bayesian themes

- Model $y_j \sim \text{Poisson}(10 n_j \theta_j)$
  - What are the assumptions here?
- Flat Gamma$(0, 0)$ prior gives posterior mean $\frac{y_j}{n_j}$
- Informative Gamma $(20, 430{,}000)$ prior gives posterior mean $\frac{20+y_j}{430{,}000+n_j}$
- Equivalent sample size (in each county) of 20, prior mean (in each county) of $\frac{20}{430{,}000} = 4.65 \times 10^{-5}$
- One prior, many different datasets
- Data-based prior distribution

Comparing Bayesian inferences to other estimates

- Do parameter estimates look reasonable?
- Do predictions look reasonable?
- Artifacts
- Cross-validation
- Bayesian inference as implicit cross-validation

Comparing Bayesian inferences to other estimates

- Do parameter estimates look reasonable?
- Do predictions look reasonable?
- Artifacts
- Cross-validation
- Bayesian inference as implicit cross-validation

Comparing Bayesian inferences to other estimates

- Do parameter estimates look reasonable?
- Do predictions look reasonable?
- Artifacts
- Cross-validation
- Bayesian inference as implicit cross-validation

Comparing Bayesian inferences to other estimates

- Do parameter estimates look reasonable?
- Do predictions look reasonable?
- Artifacts
- Cross-validation
- Bayesian inference as implicit cross-validation

Comparing Bayesian inferences to other estimates

- Do parameter estimates look reasonable?
- Do predictions look reasonable?
- Artifacts
- Cross-validation
- Bayesian inference as implicit cross-validation

Comparing Bayesian inferences to other estimates

- Do parameter estimates look reasonable?
- Do predictions look reasonable?
- Artifacts
- Cross-validation
- Bayesian inference as implicit cross-validation

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
    y_{j2} deaths out of $5n_j$ person-years
    Use $y_{j1}/5n_j$ to estimate θ_j
    Use $5n_j\theta_j$ to estimate y_{j2}

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
  - y_{j2} deaths out of $5n_j$ person-years
  - Use $y_{j1}/5n_j$ to estimate θ_j
  - Use $5n_j\theta_j$ to estimate y_{j2}

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
  - y_{j2} deaths out of $5n_j$ person-years
  - Use $y_{j1}/5n_j$ to estimate θ_j
  - Use $5n_j\theta_j$ to estimate y_{j2}

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
  - y_{j2} deaths out of $5n_j$ person-years
  - Use $y_{j1}/5n_j$ to estimate θ_j
  - Use $5n_j\theta_j$ to estimate y_{j2}

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
  - y_{j2} deaths out of $5n_j$ person-years
  - Use $y_{j1}/5n_j$ to estimate θ_j
  - Use $5n_j\theta_j$ to estimate y_{j2}

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
  - y_{j2} deaths out of $5n_j$ person-years
  - Use $y_{j1}/5n_j$ to estimate θ_j
  - Use $5n_j\theta_j$ to estimate y_{j2}

Cross-validation for the cancer-rate example

- For county j, observe y_j deaths out of $10n_j$ person-years
- Imagine predicting second half from first half:
  - y_{j1} deaths out of $5n_j$ person-years
  - y_{j2} deaths out of $5n_j$ person-years
  - Use $y_{j1}/5n_j$ to estimate θ_j
  - Use $5n_j\theta_j$ to estimate y_{j2}

2.8. Noninformative prior distributions

- Proper and improper prior distributions
- Unnormalized densities
- Uniform prior distributions on different scales
- Some examples

2.8. Noninformative prior distributions

- Proper and improper prior distributions
- Unnormalized densities
- Uniform prior distributions on different scales
- Some examples

2.8. Noninformative prior distributions

- Proper and improper prior distributions
- Unnormalized densities
- Uniform prior distributions on different scales
- Some examples

2.8. Noninformative prior distributions

- Proper and improper prior distributions
- Unnormalized densities
- Uniform prior distributions on different scales
- Some examples

2.8. Noninformative prior distributions

- Proper and improper prior distributions
- Unnormalized densities
- Uniform prior distributions on different scales
- Some examples

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
- Scale parameter $\sigma > 0$

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
- Scale parameter $\sigma > 0$

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
- Scale parameter $\sigma > 0$

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$
  - One possibility: $p(\sigma) \propto 1$ [improper]
  - Another possibility: $p(\log \sigma) \propto 1$ corresponds to $p(\sigma) \propto \sigma^{-1}$ [improper]

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$
  - One possibility: $p(\sigma) \propto 1$ [improper]
  - Another possibility: $p(\log \sigma) \propto 1$ corresponds to $p(\sigma) \propto \sigma^{-1}$ [improper]

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$
  - One possibility: $p(\sigma) \propto 1$ [improper]
  - Another possibility: $p(\log \sigma) \propto 1$ corresponds to $p(\sigma) \propto \sigma^{-1}$ [improper]

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$
  - One possibility: $p(\sigma) \propto 1$ [improper]
  - Another possibility: $p(\log \sigma) \propto 1$ corresponds to $p(\sigma) \propto \sigma^{-1}$ [improper]

Examples of noninformative priors

- Probability parameter $\theta \in (0, 1)$
  - One possibility: $p(\theta) = 1$ [proper]
  - Another possibility: $p(\text{logit}\,\theta) \propto 1$ corresponds to $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ [improper]
  - Example of political ideology
- Location parameter θ unconstrained
  - One possibility: $p(\theta) \propto 1$ [improper]
- Scale parameter $\sigma > 0$
  - One possibility: $p(\sigma) \propto 1$ [improper]
  - Another possibility: $p(\log \sigma) \propto 1$ corresponds to $p(\sigma) \propto \sigma^{-1}$ [improper]

2.9. Weakly informative prior distributions

- Some prior information constrains the inferences
- All prior distributions are weakly informative
- True prior and assumed prior
- Open questions

2.9. Weakly informative prior distributions

- Some prior information constrains the inferences
- All prior distributions are weakly informative
- True prior and assumed prior
- Open questions

2.9. Weakly informative prior distributions

- Some prior information constrains the inferences
- All prior distributions are weakly informative
- True prior and assumed prior
- Open questions

2.9. Weakly informative prior distributions

- Some prior information constrains the inferences
- All prior distributions are weakly informative
- True prior and assumed prior
- Open questions

2.9. Weakly informative prior distributions

- Some prior information constrains the inferences
- All prior distributions are weakly informative
- True prior and assumed prior
- Open questions

General theory for wips

- (Unknown) true prior, $p_{\mathrm{true}}(\theta) = N(\theta|\mu_0, \sigma_0^2)$
- Your subjective prior, $p_{\mathrm{subj}} = N(\theta|\mu_1, \sigma_1^2)$
- Weakly-informative prior, $p_{\mathrm{wip}} = N(\theta|\mu_1, (\kappa\sigma_1)^2)$, with $\kappa > 1$
- Tradeoffs if κ is too low or too high

General theory for wips

- (Unknown) true prior, $p_{\mathrm{true}}(\theta) = N(\theta|\mu_0, \sigma_0^2)$
- Your subjective prior, $p_{\mathrm{subj}} = N(\theta|\mu_1, \sigma_1^2)$
- Weakly-informative prior, $p_{\mathrm{wip}} = N(\theta|\mu_1, (\kappa\sigma_1)^2)$, with $\kappa > 1$
- Tradeoffs if κ is too low or too high

General theory for wips

- (Unknown) true prior, $p_{\mathrm{true}}(\theta) = N(\theta|\mu_0, \sigma_0^2)$
- Your subjective prior, $p_{\mathrm{subj}} = N(\theta|\mu_1, \sigma_1^2)$
- Weakly-informative prior, $p_{\mathrm{wip}} = N(\theta|\mu_1, (\kappa\sigma_1)^2)$, with $\kappa > 1$
- Tradeoffs if κ is too low or too high

General theory for wips

- (Unknown) true prior, $p_{\rm true}(\theta) = N(\theta|\mu_0, \sigma_0^2)$
- Your subjective prior, $p_{\rm subj} = N(\theta|\mu_1, \sigma_1^2)$
- Weakly-informative prior, $p_{\rm wip} = N(\theta|\mu_1, (\kappa\sigma_1)^2)$, with $\kappa > 1$
- Tradeoffs if κ is too low or too high

General theory for wips

- (Unknown) true prior, $p_{\rm true}(\theta) = N(\theta|\mu_0, \sigma_0^2)$
- Your subjective prior, $p_{\rm subj} = N(\theta|\mu_1, \sigma_1^2)$
- Weakly-informative prior, $p_{\rm wip} = N(\theta|\mu_1, (\kappa\sigma_1)^2)$, with $\kappa > 1$
- Tradeoffs if κ is too low or too high

Ways to add prior information

- Weakly informative priors
- Fully informative priors
- Prior information that groups are similar to each other
- Prior information on predictive quantitites

Ways to add prior information

- Weakly informative priors
- Fully informative priors
- Prior information that groups are similar to each other
- Prior information on predictive quantitites

Ways to add prior information

- Weakly informative priors
- Fully informative priors
- Prior information that groups are similar to each other
- Prior information on predictive quantitites

Ways to add prior information

- Weakly informative priors
- Fully informative priors
- Prior information that groups are similar to each other
- Prior information on predictive quantitites

Ways to add prior information

- Weakly informative priors
- Fully informative priors
- Prior information that groups are similar to each other
- Prior information on predictive quantitites

Summary of Chapter 2

- Algebra of basic models
- Prior distribution as equivalent data
- Posterior as weighted average of prior and data
- Noninformative priors
- Weakly informative priors

Summary of Chapter 2

- Algebra of basic models
- Prior distribution as equivalent data
- Posterior as weighted average of prior and data
- Noninformative priors
- Weakly informative priors

Summary of Chapter 2

- Algebra of basic models
- Prior distribution as equivalent data
- Posterior as weighted average of prior and data
- Noninformative priors
- Weakly informative priors

Summary of Chapter 2

- Algebra of basic models
- Prior distribution as equivalent data
- Posterior as weighted average of prior and data
- Noninformative priors
- Weakly informative priors

Summary of Chapter 2

- Algebra of basic models
- Prior distribution as equivalent data
- Posterior as weighted average of prior and data
- Noninformative priors
- Weakly informative priors

Summary of Chapter 2

- Algebra of basic models
- Prior distribution as equivalent data
- Posterior as weighted average of prior and data
- Noninformative priors
- Weakly informative priors

Homework due beginning of class 3b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: proving that a posterior distribution is proper
  - Computing problem: computing posterior density on a grid
  - Applied problem: health effects of electromagnetic fields

Homework due beginning of class 3b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: proving that a posterior distribution is proper
  - Computing problem: computing posterior density on a grid
  - Applied problem: health effects of electromagnetic fields

Homework due beginning of class 3b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: proving that a posterior distribution is proper
  - Computing problem: computing posterior density on a grid
  - Applied problem: health effects of electromagnetic fields

Homework due beginning of class 3b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: proving that a posterior distribution is proper
  - Computing problem: computing posterior density on a grid
  - Applied problem: health effects of electromagnetic fields

Homework due beginning of class 3b

- All assignments are at `http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf`
  - Theory problem: proving that a posterior distribution is proper
  - Computing problem: computing posterior density on a grid
  - Applied problem: health effects of electromagnetic fields