# Bayesian Data Analysis, class 3a

Andrew Gelman

Chapter 3: Introduction to multiparameter models (part 1)

# 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 $\theta$, 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 $\theta$. Suppose the prior distribution for $\theta$ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of $\theta$ 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 $\theta$, 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 $\theta$. Suppose the prior distribution for $\theta$ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of $\theta$ 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 $\theta$, 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 $\theta$. Suppose the prior distribution for $\theta$ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of $\theta$ 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 $\theta$, 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 $\theta$. Suppose the prior distribution for $\theta$ is a gamma distribution with coefficient of variation 0.5 ... If the coefficient of variation of the distribution of $\theta$ 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.

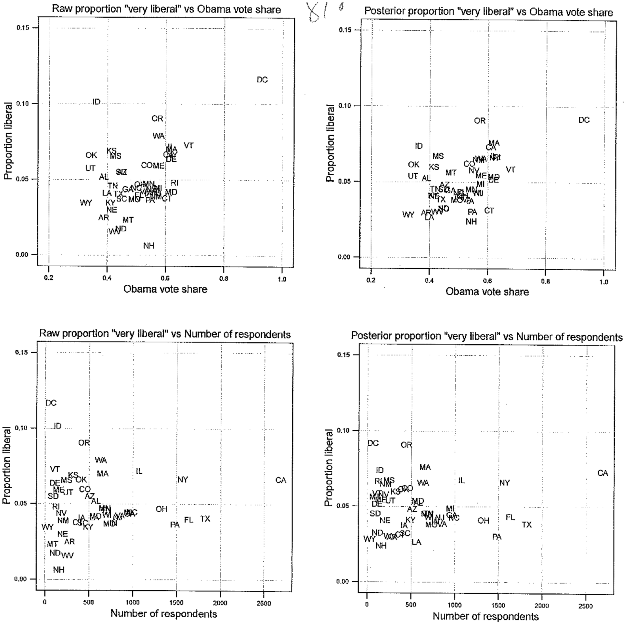
# Computing problem

- 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.

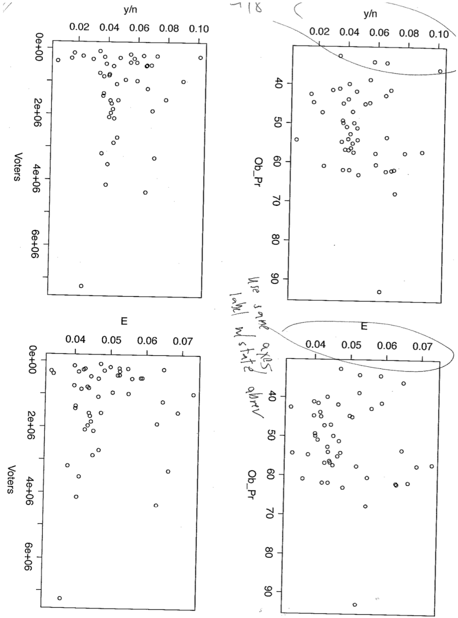
# Computing problem

- 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.

# Good graphs (but could be even better!)



# Bad graphs (but that's how you learn!)



# Applied problem

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

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

# One student's answer

- A noninformative prior:

  > "$U(-1, 1)$ *because the change in probabilities can't be less than* $-1$ *or more than 1"*

- A subjective prior based on your best knowledge:

  > "*Beta*$(2, 5)$*: most of the probability lies below* $\theta = 0.5$ *(mean = 0.29, mode = 0.20), and the probability of improving by more than 0.8 is essentially zero"*

- A weakly informative prior:

  > *"Assuming that taking 50 practice shots a day for a month does not worsen success probability, we can choose* $U(0, 1)$*"*

# One student's answer

- A noninformative prior:

  *"$U(-1, 1)$ because the change in probabilities can't be less than $-1$ or more than 1"*

- A subjective prior based on your best knowledge:

  *"$\text{Beta}(2, 5)$: most of the probability lies below $\theta = 0.5$ (mean = 0.29, mode = 0.20), and the probability of improving by more than 0.8 is essentially zero"*

- A weakly informative prior:

  *"Assuming that taking 50 practice shots a day for a month does not worsen success probability, we can choose $U(0, 1)$"*

# One student's answer

- A noninformative prior:

  *"$U(-1, 1)$ because the change in probabilities can't be less than $-1$ or more than 1"*

- A subjective prior based on your best knowledge:

  *"$\text{Beta}(2, 5)$: most of the probability lies below $\theta = 0.5$ (mean $= 0.29$, mode $= 0.20$), and the probability of improving by more than 0.8 is essentially zero"*

- A weakly informative prior:

  *"Assuming that taking 50 practice shots a day for a month does not worsen success probability, we can choose $U(0, 1)$"*

# One student's answer

- A noninformative prior:

  > *"$U(-1,1)$ because the change in probabilities can't be less than $-1$ or more than 1"*

- A subjective prior based on your best knowledge:

  > *"Beta$(2,5)$: most of the probability lies below $\theta = 0.5$ (mean = 0.29, mode = 0.20), and the probability of improving by more than 0.8 is essentially zero"*

- A weakly informative prior:

  > *"Assuming that taking 50 practice shots a day for a month does not worsen success probability, we can choose $U(0,1)$"*

# 3. Introduction to multiparameter models (part 1)

- Important algebra and geometry
- Basic computation
- Similarities and differences compared to classical statistics

# 3. Introduction to multiparameter models (part 1)

- Important algebra and geometry
- Basic computation
- Similarities and differences compared to classical statistics

# 3. Introduction to multiparameter models (part 1)

- Important algebra and geometry
- Basic computation
- Similarities and differences compared to classical statistics

# 3. Introduction to multiparameter models (part 1)

- Important algebra and geometry
- Basic computation
- Similarities and differences compared to classical statistics

# 3.1. Averaging over "nuisance parameters"

- Joint posterior distribution, $p(\theta_1, \theta_2|y)$
- Factorization, $p(\theta_1|y)p(\theta_2|\theta_1, y)$
- What is a nuisance parameter?

# 3.1. Averaging over "nuisance parameters"

- Joint posterior distribution, $p(\theta_1, \theta_2|y)$
- Factorization, $p(\theta_1|y)p(\theta_2|\theta_1, y)$
- What is a nuisance parameter?

# 3.1. Averaging over "nuisance parameters"

- Joint posterior distribution, $p(\theta_1, \theta_2 | y)$
- Factorization, $p(\theta_1 | y) p(\theta_2 | \theta_1, y)$
- What is a nuisance parameter?

# 3.1. Averaging over "nuisance parameters"

- Joint posterior distribution, $p(\theta_1, \theta_2|y)$
- Factorization, $p(\theta_1|y)p(\theta_2|\theta_1, y)$
- What is a nuisance parameter?

# Inference for the difference between two parameters

- Simple scenario of independence:
  - $y_1 \sim \text{dist}(\theta_1)$, prior $p(\theta_1)$
  - $y_2 \sim \text{dist}(\theta_2)$, prior $p(\theta_2)$
- Get 1000 posterior simulations for $\theta_1$ and 1000 posterior simulations for $\theta_2$
- Inference for $\theta_2 - \theta_1$ directly

# Inference for the difference between two parameters

- Simple scenario of independence:
  - $y_1 \sim \text{dist}(\theta_1)$, prior $p(\theta_1)$
  - $y_2 \sim \text{dist}(\theta_2)$, prior $p(\theta_2)$
- Get 1000 posterior simulations for $\theta_1$ and 1000 posterior simulations for $\theta_2$
- Inference for $\theta_2 - \theta_1$ directly

# Inference for the difference between two parameters

- Simple scenario of independence:
  - $y_1 \sim \text{dist}(\theta_1)$, prior $p(\theta_1)$
  - $y_2 \sim \text{dist}(\theta_2)$, prior $p(\theta_2)$
- Get 1000 posterior simulations for $\theta_1$ and 1000 posterior simulations for $\theta_2$
- Inference for $\theta_2 - \theta_1$ directly

# Inference for the difference between two parameters

- Simple scenario of independence:
  - $y_1 \sim \text{dist}(\theta_1)$, prior $p(\theta_1)$
  - $y_2 \sim \text{dist}(\theta_2)$, prior $p(\theta_2)$
- Get 1000 posterior simulations for $\theta_1$ and 1000 posterior simulations for $\theta_2$
- Inference for $\theta_2 - \theta_1$ directly

# Inference for the difference between two parameters

- Simple scenario of independence:
  - $y_1 \sim \text{dist}(\theta_1)$, prior $p(\theta_1)$
  - $y_2 \sim \text{dist}(\theta_2)$, prior $p(\theta_2)$
- Get 1000 posterior simulations for $\theta_1$ and 1000 posterior simulations for $\theta_2$
- Inference for $\theta_2 - \theta_1$ directly

# Inference for the difference between two parameters

- Simple scenario of independence:
  - $y_1 \sim \text{dist}(\theta_1)$, prior $p(\theta_1)$
  - $y_2 \sim \text{dist}(\theta_2)$, prior $p(\theta_2)$
- Get 1000 posterior simulations for $\theta_1$ and 1000 posterior simulations for $\theta_2$
- Inference for $\theta_2 - \theta_1$ directly

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
    Equivalent to $p(\mu, \log \sigma) \propto 1$
- Unnormalized joint posterior density:
  $p(\mu, \sigma^2|y) \propto (2\pi)^{-n/2}\sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2\right)$
- Integrate out $\mu$: $\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu|(\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log \sigma) \propto 1$
- Unnormalized joint posterior density:
  $p(\mu, \sigma^2|y) \propto (2\pi)^{-n/2}\sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu|(\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \dots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log\sigma) \propto 1$
- Unnormalized joint posterior density:
  $p(\mu, \sigma^2|y) \propto (2\pi)^{-n/2}\sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu|(\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log \sigma) \propto 1$
- Unnormalized joint posterior density:
  $p(\mu, \sigma^2 | y) \propto (2\pi)^{-n/2} \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2 | y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu | (\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log\sigma) \propto 1$
- Unnormalized joint posterior density:
  $p(\mu, \sigma^2|y) \propto (2\pi)^{-n/2}\sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu|(\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log\sigma) \propto 1$
- Unnormalized joint posterior density:
  $p(\mu, \sigma^2|y) \propto (2\pi)^{-n/2}\sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu|(\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log\sigma) \propto 1$
- Unnormalized joint posterior density: $p(\mu, \sigma^2|y) \propto (2\pi)^{-n/2}\sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu|(\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

## 3.2. Normal data with a noninformative prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Try $p(\mu, \sigma^2) \propto \sigma^{-2}$
  - Equivalent to $p(\mu, \sigma) \propto \sigma^{-1}$
  - Equivalent to $p(\mu, \log \sigma) \propto 1$
- Unnormalized joint posterior density: $p(\mu, \sigma^2 | y) \propto (2\pi)^{-n/2} \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$
- Integrate out $\mu$: $\sigma^2 | y \sim \text{Inv-}\chi^2(n-1, s^2)$
- Conditional on $\sigma^2$: $\mu | (\sigma^2, y) \sim \mathrm{N}(\bar{y}, \sigma^2/n)$

# Sufficient statistics

- Data distribution: $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Prior distribution $p(\mu, \sigma^2) \propto \sigma^{-2}$
- Joint posterior density depends only on $\bar{y}, s^2, n$
- $(\bar{y}, s^2, n)$ are *sufficient statistics*
- What happens if we change the prior distribution?

# Sufficient statistics

- Data distribution: $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Prior distribution $p(\mu, \sigma^2) \propto \sigma^{-2}$
- Joint posterior density depends only on $\bar{y}, s^2, n$
- $(y, s^2, n)$ are *sufficient statistics*
- What happens if we change the prior distribution?

# Sufficient statistics

- Data distribution: $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Prior distribution $p(\mu, \sigma^2) \propto \sigma^{-2}$
- Joint posterior density depends only on $\bar{y}, s^2, n$
- $(y, s^2, n)$ are *sufficient statistics*
- What happens if we change the prior distribution?

# Sufficient statistics

- Data distribution: $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Prior distribution $p(\mu, \sigma^2) \propto \sigma^{-2}$
- Joint posterior density depends only on $\bar{y}, s^2, n$
- $(y, s^2, n)$ are *sufficient statistics*
- What happens if we change the prior distribution?

# Sufficient statistics

- Data distribution: $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Prior distribution $p(\mu, \sigma^2) \propto \sigma^{-2}$
- Joint posterior density depends only on $\bar{y}, s^2, n$
- $(y, s^2, n)$ are *sufficient statistics*
- What happens if we change the prior distribution?

# Sufficient statistics

- Data distribution: $y_1, \dots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Prior distribution $p(\mu, \sigma^2) \propto \sigma^{-2}$
- Joint posterior density depends only on $\bar{y}, s^2, n$
- $(y, s^2, n)$ are *sufficient statistics*
- What happens if we change the prior distribution?

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
    $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
  - $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
  - $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
  - $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
  - $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
  - $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# 3.3. Normal data with a conjugate prior distribution

- $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$
- Conjugate family of priors:
  - $\sigma^2 \sim \text{Inv-}\chi^2$
  - $\mu|\sigma^2 \sim$ Normal with variance proportional to $\sigma^2$
- Posterior has same form
- Interpreting prior as equivalent data

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
    Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
    Two location parameters, $\mu_1, \mu_2$
    Baseline and treatment effect, $\mu, \theta$
    Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
    Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# Joint prior distributions

- Dependent prior on $\mu, \sigma$: does this make sense?
- Examples of multi-parameter models
  - Normal model: location $\mu$, scale $\sigma$
  - Student-$t$ model: d.f. $\nu$, location $\mu$, scale $\sigma$
  - Two location parameters, $\mu_1, \mu_2$
  - Baseline and treatment effect, $\mu, \theta$
  - Regression coefficients: $\beta_0, \beta_1, \beta_2, \ldots$
  - Main effects and interactions

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
- Conjugate prior as equivalent data:
- Consider some examples

## 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
- Conjugate prior as equivalent data:
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
- Conjugate prior as equivalent data:
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(\frac{1}{k}, \ldots, \frac{1}{k})$?
- Conjugate prior as equivalent data:
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$
- Consider some examples

# 3.4. Multinomial model for categorical data

- $y_1, \ldots, y_k \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$
- Unknown probabilities $\theta_1, \ldots, \theta_k$
- Independent prior distributions?
- Some noninformative priors:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(1, \ldots, 1)$ (uniform on the $\theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(0, \ldots, 0)$ (uniform on the $\log \theta_j$'s)?
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$?
- Conjugate prior as equivalent data:
  - $\theta_1, \ldots, \theta_k \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$
- Consider some examples

# 3.5. Multivariate normal model with known variance

- Important algebra

- Important algebra

# 3.6. Multivariate normal with unknown mean and variance

- Research on classes of prior distributions
- Parameterizing using covariances or correlations or eigenvalues
- Research on visualization

# 3.6. Multivariate normal with unknown mean and variance

- Research on classes of prior distributions
- Parameterizing using covariances or correlations or eigenvalues
- Research on visualization

# 3.6. Multivariate normal with unknown mean and variance

- Research on classes of prior distributions
- Parameterizing using covariances or correlations or eigenvalues
- Research on visualization

# 3.6. Multivariate normal with unknown mean and variance

- Research on classes of prior distributions
- Parameterizing using covariances or correlations or eigenvalues
- Research on visualization

# VIsualizing the inverse-Wishart in 4 dimensions