

Bayesian Inference for a Mean

Jingchen (Monika) Hu

Vassar College

MATH 347 Bayesian Statistics

Outline

- 1 Example: Expenditures in the Consumer Expenditure Surveys
- 2 Prior and posterior distributions for mean and standard deviation
- 3 Bayesian inference for unknown mean μ (Lab 2)
- 4 Recap

Outline

- 1 Example: Expenditures in the Consumer Expenditure Surveys
- 2 Prior and posterior distributions for mean and standard deviation
- 3 Bayesian inference for unknown mean μ (Lab 2)
- 4 Recap

The Consumer Expenditure Surveys Data (CE)

- Conducted by the U.S. Census Bureau for the BLS.
- Contains data on expenditures, income, and tax statistics about consumer units (CU) across the country.
- Provides information on the buying habits of U.S. consumers.

The Consumer Expenditure Surveys Data (CE)

- Conducted by the U.S. Census Bureau for the BLS.
- Contains data on expenditures, income, and tax statistics about consumer units (CU) across the country.
- Provides information on the buying habits of U.S. consumers.
- The CE program releases data in two ways:
 - ▶ Tabular data (aggregated)
 - ▶ Micro-level data: public-use microdata PUMD (CU-level)

The Consumer Expenditure Surveys Data (CE)

- Conducted by the U.S. Census Bureau for the BLS.
- Contains data on expenditures, income, and tax statistics about consumer units (CU) across the country.
- Provides information on the buying habits of U.S. consumers.
- The CE program releases data in two ways:
 - ▶ Tabular data (aggregated)
 - ▶ Micro-level data: public-use microdata PUMD (CU-level)
- We work with PUMD micro-level data, with the continuous variable **TOTEXPPQ**: CU total expenditures last quarter.
- We work with Q1 2017 sample: $n = 6,208$.

The TOTEXPPQ variable

```
CEsample = read_csv("CEsample1.csv")
```

```
summary(CEsample$TotalExpLastQ)
```

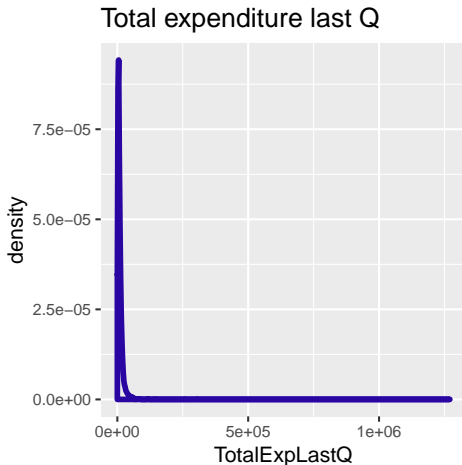
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      30    3522    6417    9513   11450  1270598
```

```
sd(CEsample$TotalExpLastQ)
```

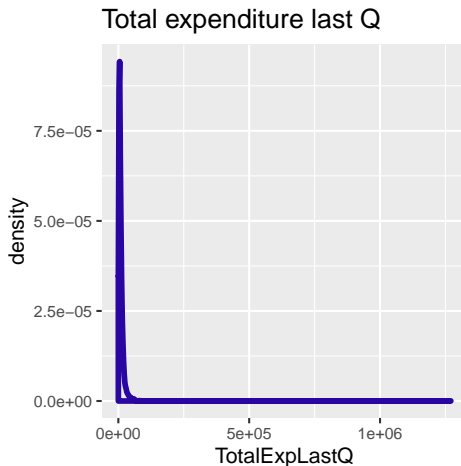
```
## [1] 19341.25
```

The TOTEXPPQ variable cont'd

```
ggplot(data = CEsample, aes(TotalExpLastQ)) +  
  geom_density(color = crcblue, size = 1) +  
  labs(title = "Total expenditure last Q") +  
  theme_grey(base_size = 8, base_family = "")
```



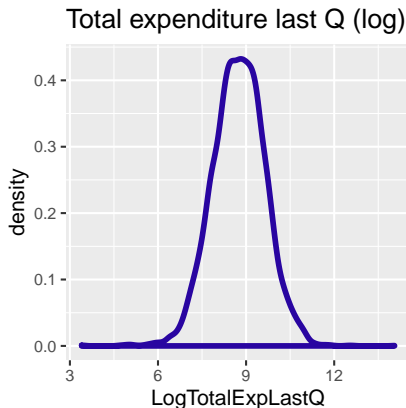
The TOTEXPPQ variable cont'd



- Very skewed to the right.
- Take log and transform it to the log scale.

Log transformation of the TOTEXPPQ variable

```
CEsample$LogTotalExpLastQ <- log(CEsample$TotalExpLastQ)
ggplot(data = CEsample, aes(LogTotalExpLastQ)) +
  geom_density(color = crcblue, size = 1) +
  labs(title = "Total expenditure last Q (log)") +
  theme_grey(base_size = 8, base_family = "")
```



The Normal distribution

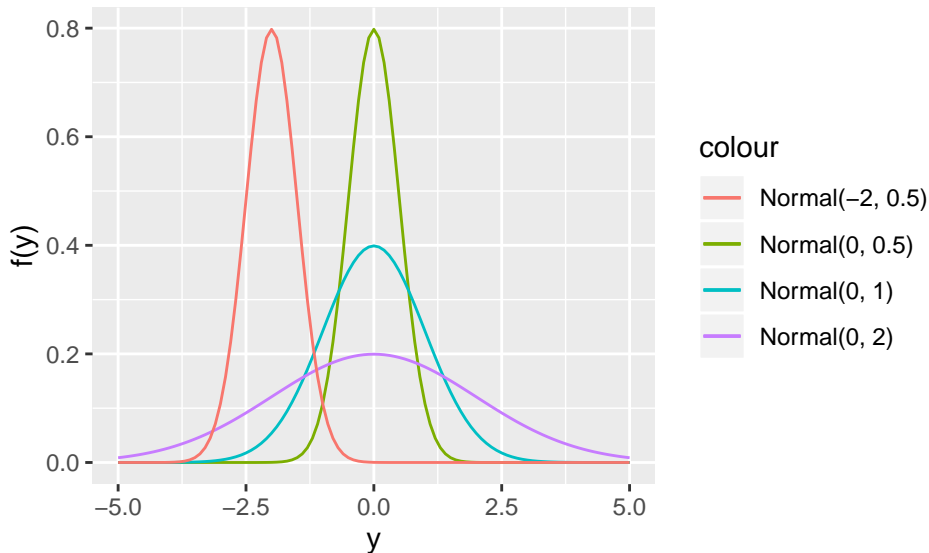
- The Normal distribution is a symmetric, bell-shaped distribution.
- It has two parameters: mean μ and standard deviation σ .

The Normal distribution

- The Normal distribution is a symmetric, bell-shaped distribution.
- It has two parameters: mean μ and standard deviation σ .
- The probability density function (pdf) of $\text{Normal}(\mu, \sigma)$ is:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right), -\infty < y < \infty.$$

The Normal distribution cont'd



i.i.d. Normals

- Suppose there are a sequence of n responses: Y_1, Y_2, \dots, Y_n .
- Further suppose each response **independently and identically** follows a Normal distribution:

$$Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- Then the joint probability density function (joint pdf) of y_1, \dots, y_n is:

$$f(y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y_i - \mu)^2}{2\sigma^2}\right), -\infty < y_i < \infty. \quad (1)$$

Outline

- 1 Example: Expenditures in the Consumer Expenditure Surveys
- 2 Prior and posterior distributions for mean and standard deviation
- 3 Bayesian inference for unknown mean μ (Lab 2)
- 4 Recap

Recap from proportion lecture

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of the proportion p before sampling (prior).
 - ▶ Step 2: take the sample and record the observed proportion (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about p given the information from the sample (posterior).

Recap from proportion lecture

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of the proportion p before sampling (prior).
 - ▶ Step 2: take the sample and record the observed proportion (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about p given the information from the sample (posterior).
- For Binomial data/likelihood, the Beta distributions are conjugate priors.
 - ▶ The prior distribution: $p \sim \text{Beta}(a, b)$
 - ▶ The sampling density: $Y \sim \text{Binomial}(n, p)$
 - ▶ The posterior distribution: $p \mid Y = y \sim \text{Beta}(a + y, b + n - y)$

Recap from proportion lecture

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of the proportion p before sampling (prior).
 - ▶ Step 2: take the sample and record the observed proportion (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about p given the information from the sample (posterior).
- For Binomial data/likelihood, the Beta distributions are conjugate priors.
 - ▶ The prior distribution: $p \sim \text{Beta}(a, b)$
 - ▶ The sampling density: $Y \sim \text{Binomial}(n, p)$
 - ▶ The posterior distribution: $p \mid Y = y \sim \text{Beta}(a + y, b + n - y)$
- What to do for a Normal model $Y_i \overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma)$?
 - ▶ Data model/sampling density is chosen: Normal.
 - ▶ What to do with two parameters μ and σ ?
 - ▶ How to specify priors? Conjugate priors exist?

Step 1: Prior distributions

- The data model/sampling density for n observations:

$$Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- There are two parameters μ and σ in the Normal model.
- Therefore, the likelihood is in terms of both unknown parameters:

$$f(y_1, \dots, y_n) = L(\mu, \sigma). \quad (2)$$

Step 1: Prior distributions

- The data model/sampling density for n observations:

$$Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- There are two parameters μ and σ in the Normal model.
- Therefore, the likelihood is in terms of both unknown parameters:

$$f(y_1, \dots, y_n) = L(\mu, \sigma). \quad (2)$$

- Need a joint prior distribution:

$$\pi(\mu, \sigma). \quad (3)$$

- Bayes' rule will help us derive a joint posterior:

$$\pi(\mu, \sigma \mid y_1, \dots, y_n). \quad (4)$$

If only mean μ is unknown

- Special case: μ is unknown, σ is known.
- There is only one parameter μ in $Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma)$.
- The Bayesian inference procedure simplifies to:
 - ▶ The data model for n observations with σ known:

$$Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- ▶ The likelihood is in terms of unknown parameter μ :

$$f(y_1, \dots, y_n) = L(\mu). \quad (5)$$

- ▶ Need a prior distribution for μ :

$$\pi(\mu \mid \sigma). \quad (6)$$

- ▶ Bayes' rule will help us derive a posterior for μ :

$$\pi(\mu \mid y_1, \dots, y_n, \sigma). \quad (7)$$

If only mean μ is unknown: Normal conjugate prior

- For this special case, Normal prior for μ is a conjugate prior:
 - ▶ The prior distribution:

$$\mu \mid \sigma \sim \text{Normal}(\mu_0, \sigma_0). \quad (8)$$

- ▶ The sampling density:

$$y_1, \dots, y_n \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma). \quad (9)$$

If only mean μ is unknown: Normal conjugate prior

- For this special case, Normal prior for μ is a conjugate prior:
 - The prior distribution:

$$\mu \mid \sigma \sim \text{Normal}(\mu_0, \sigma_0). \quad (8)$$

- The sampling density:

$$y_1, \dots, y_n \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma). \quad (9)$$

- The posterior distribution:

$$\mu \mid y_1, \dots, y_n, \phi \sim \text{Normal} \left(\frac{\phi_0 \mu_0 + n\phi \bar{y}}{\phi_0 + n\phi}, \sqrt{\frac{1}{\phi_0 + n\phi}} \right), \quad (10)$$

where $\phi = \frac{1}{\sigma^2}$ (and $\phi_0 = \frac{1}{\sigma_0^2}$), the precision. Since σ (and σ_0) is known, ϕ (and ϕ_0) is known too.

If only mean μ is unknown: Normal conjugate prior

- For this special case, Normal prior for μ is a conjugate prior:
 - The prior distribution:

$$\mu \mid \sigma \sim \text{Normal}(\mu_0, \sigma_0). \quad (8)$$

- The sampling density:

$$y_1, \dots, y_n \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma). \quad (9)$$

- The posterior distribution:

$$\mu \mid y_1, \dots, y_n, \phi \sim \text{Normal} \left(\frac{\phi_0 \mu_0 + n \phi \bar{y}}{\phi_0 + n \phi}, \sqrt{\frac{1}{\phi_0 + n \phi}} \right), \quad (10)$$

where $\phi = \frac{1}{\sigma^2}$ (and $\phi_0 = \frac{1}{\sigma_0^2}$), the precision. Since σ (and σ_0) is known, ϕ (and ϕ_0) is known too.

- We can then use the `rnorm()` R function to sample posterior draws of μ from Equation (10). **Known quantities:** $\phi_0, \mu_0, n, \bar{y}, \phi$

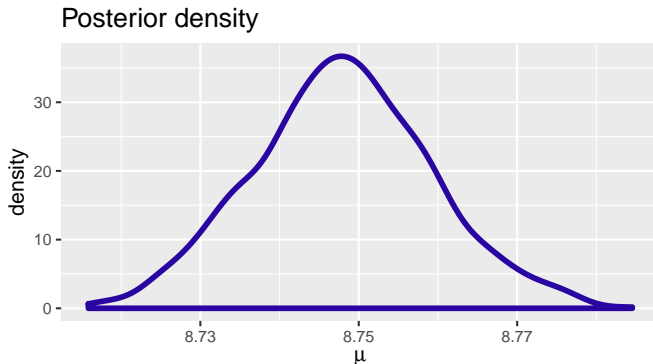
Simulate posterior draws of μ

```
mu_0 <- 5
sigma_0 <- 1
phi_0 <- 1/sigma_0^2
ybar <- mean(CESample$LogTotalExpLastQ)
phi <- 1.25
n <- dim(CESample)[1]
mu_n <- (phi_0*mu_0+n*ybar*phi)/(phi_0+n*phi)
sd_n <- sqrt(1/(phi_0+n*phi))

set.seed(123)
S <- 1000
mu_post <- rnorm(S, mean = mu_n, sd = sd_n)
df <- as.data.frame(mu_post)
```

Simulate posterior draws of μ cont'd

```
ggplot(data = df, aes(mu_post)) +  
  geom_density(color = "darkblue", size = 1) +  
  labs(title = "Posterior density") +  
  xlab(expression(mu)) +  
  theme_grey(base_size = 8, base_family = "")
```



If only standard deviation σ is unknown

- Special case: μ is known, σ is unknown.
- There is only one parameter μ in $Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma)$.
- The Bayesian inference procedure simplifies to:
 - ▶ The data model/sampling density for n observations with μ known:

$$Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- ▶ The likelihood is in terms of unknown parameter σ :

$$f(y_1, \dots, y_n) = L(\sigma). \quad (11)$$

- ▶ Need a prior distribution for σ :

$$\pi(\sigma \mid \mu). \quad (12)$$

- ▶ Bayes' rule will help us derive a posterior for σ :

$$\pi(\sigma \mid y_1, \dots, y_n, \mu). \quad (13)$$

If only standard deviation σ is unknown: Gamma conjugate prior for $1/\sigma^2$

- For this special case, Gamma prior for $1/\sigma^2$ is a conjugate prior:
 - ▶ The prior distribution:

$$1/\sigma^2 \mid \mu \sim \text{Gamma}(\alpha, \beta). \quad (14)$$

- ▶ The sampling density:

$$y_1, \dots, y_n \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma). \quad (15)$$

If only standard deviation σ is unknown: Gamma conjugate prior for $1/\sigma^2$

- For this special case, Gamma prior for $1/\sigma^2$ is a conjugate prior:
 - The prior distribution:

$$1/\sigma^2 \mid \mu \sim \text{Gamma}(\alpha, \beta). \quad (14)$$

- The sampling density:

$$y_1, \dots, y_n \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma). \quad (15)$$

- The posterior distribution:

$$1/\sigma^2 \mid y_1, \dots, y_n, \mu \sim \text{Gamma} \left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2 \right) \quad (16)$$

If only standard deviation σ is unknown: Gamma conjugate prior for $1/\sigma^2$

- For this special case, Gamma prior for $1/\sigma^2$ is a conjugate prior:
 - The prior distribution:

$$1/\sigma^2 \mid \mu \sim \text{Gamma}(\alpha, \beta). \quad (14)$$

- The sampling density:

$$y_1, \dots, y_n \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma). \quad (15)$$

- The posterior distribution:

$$1/\sigma^2 \mid y_1, \dots, y_n, \mu \sim \text{Gamma} \left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2 \right) \quad (16)$$

- We can then use `rgamma()` R function to sample posterior draws of σ from Equation (16). **Known quantities:** $\alpha, n, \beta, \{y_i\}, \mu$

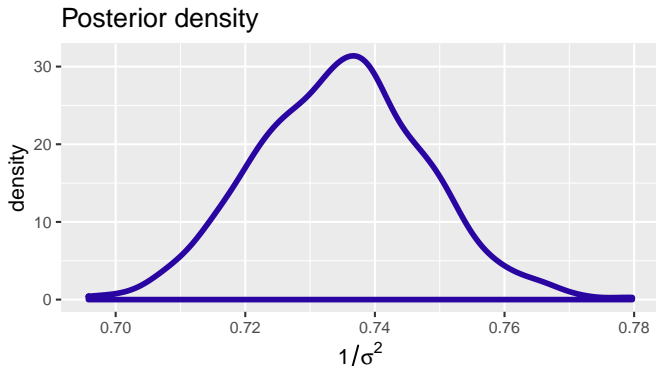
Simulate posterior draws of σ

```
alpha <- 1
beta <- 1
mu <- 8
n <- dim(CESample)[1]
alpha_n <- alpha+n/2
beta_n <- beta+1/2*sum((CESample$LogTotalExpLastQ-mu)^2)

set.seed(123)
S <- 1000
invsigma2_post <- rgamma(S, shape=alpha_n, rate=beta_n)
df <- as.data.frame(invsigma2_post)
```

Simulate posterior draws of σ cont'd

```
ggplot(data = df, aes(invsigma2_post)) +
  geom_density(color = "darkblue", size = 1) +
  labs(title = "Posterior density") +
  xlab(expression(1/sigma^2)) +
  theme_grey(base_size = 8, base_family = "")
```



Outline

- 1 Example: Expenditures in the Consumer Expenditure Surveys
- 2 Prior and posterior distributions for mean and standard deviation
- 3 Bayesian inference for unknown mean μ (Lab 2)
- 4 Recap

Inference questions

- Bayesian hypothesis testing
 - ▶ exact solution: conjugacy
 - ★ `pbeta()` \rightarrow `pnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions

Inference questions

- Bayesian hypothesis testing
 - ▶ exact solution: conjugacy
 - ★ `pbeta()` \rightarrow `pnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions
- Bayesian credible interval
 - ▶ exact solution: conjugacy
 - ★ `qbeta()` \rightarrow `qnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions

Inference questions

- Bayesian hypothesis testing
 - ▶ exact solution: conjugacy
 - ★ `pbeta()` \rightarrow `pnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions
- Bayesian credible interval
 - ▶ exact solution: conjugacy
 - ★ `qbeta()` \rightarrow `qnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions
- Bayesian prediction
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions

Inference questions

- Bayesian hypothesis testing
 - ▶ exact solution: conjugacy
 - ★ `pbeta()` \rightarrow `pnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions
- Bayesian credible interval
 - ▶ exact solution: conjugacy
 - ★ `qbeta()` \rightarrow `qnorm()` R functions
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions
- Bayesian prediction
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions
- Posterior predictive checking
 - ▶ approximation by Monte Carlo simulation
 - ★ `rbeta()` \rightarrow `rnorm()` R functions

Outline

- 1 Example: Expenditures in the Consumer Expenditure Surveys
- 2 Prior and posterior distributions for mean and standard deviation
- 3 Bayesian inference for unknown mean μ (Lab 2)
- 4 **Recap**

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of mean μ and standard deviation σ (or precision ϕ) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about μ and σ (or precision ϕ) given the information from the sample (posterior).

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of mean μ and standard deviation σ (or precision ϕ) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about μ and σ (or precision ϕ) given the information from the sample (posterior).
- For Normal data/likelihood, Normal distributions are conjugate priors for μ , and Gamma distributions are conjugate priors for ϕ .

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of mean μ and standard deviation σ (or precision ϕ) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about μ and σ (or precision ϕ) given the information from the sample (posterior).
- For Normal data/likelihood, Normal distributions are conjugate priors for μ , and Gamma distributions are conjugate priors for ϕ .
- Bayesian inference
 - ▶ Bayesian hypothesis testing
 - ▶ Bayesian credible interval
 - ▶ Bayesian prediction
 - ▶ Posterior predictive checking

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of mean μ and standard deviation σ (or precision ϕ) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about μ and σ (or precision ϕ) given the information from the sample (posterior).
- For Normal data/likelihood, Normal distributions are conjugate priors for μ , and Gamma distributions are conjugate priors for ϕ .
- Bayesian inference
 - ▶ Bayesian hypothesis testing
 - ▶ Bayesian credible interval
 - ▶ Bayesian prediction
 - ▶ Posterior predictive checking
- What if we want to use a different prior for μ ? What if both μ and σ are unknown?