Bayesian Thinking: Fundamentals, Regression and Multilevel Modeling

Jim Albert and Monika Hu

1/1/23

Webinar 1-2: Normal Models for Continuous Data

1. Example: Expenditures in the CE and normal distribution
2. Conjugate prior and posterior inferences for μ
3. Inferences for μ and σ
4. Wrap-up and additional material

Section 1

Example: Expenditures in the CE and normal distribution

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.
- 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 Total Expenditure variable

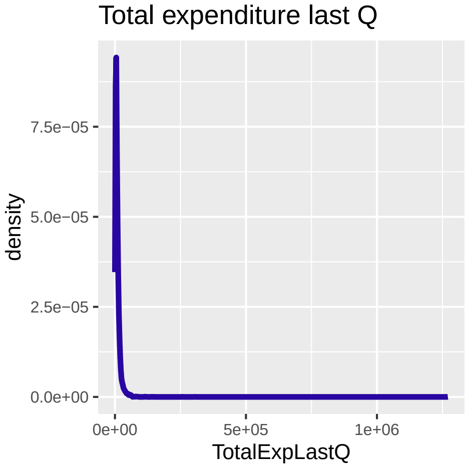
```
library(readr)
CEsample <- read_csv("CEsample.csv")

summary(CEsample$TotalExpLastQ)
```

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

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

```
[1] 19341.25
```

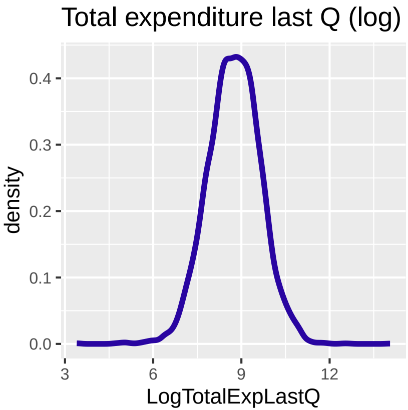
The Total Expenditure variable cont'd



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

Log transformation of the Total Expenditure variable

```
CEsample$LogTotalExpLastQ <- log(CEsample$TotalExpLastQ)
```

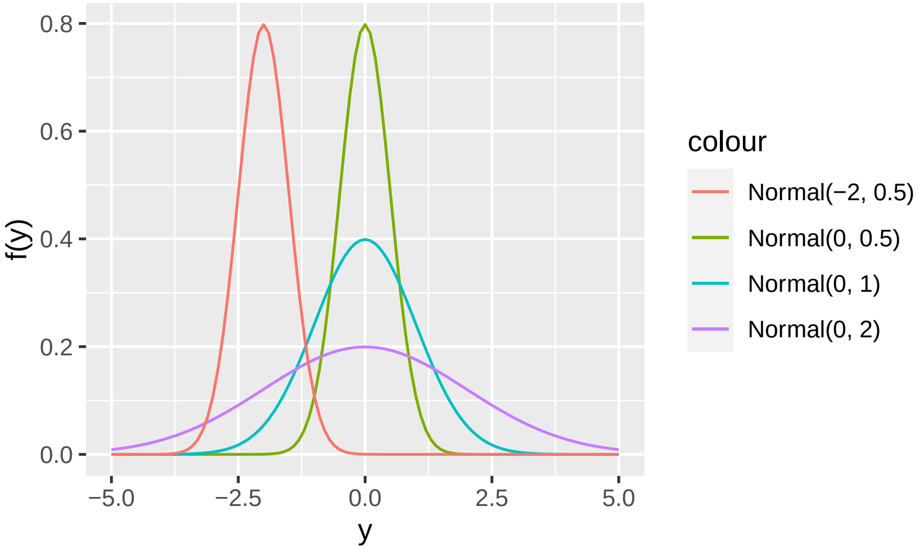


The normal distribution

- 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 $\mathrm{Normal}(\mu, \sigma)$ 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: $Y_1, Y_2, \cdots, Y_n$.
- Further suppose each response independently and identically follows a normal distribution:

$$Y_i \overset{i.i.d.}{\sim} \textrm{Normal}(\mu, \sigma).$$

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

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

Recap from beta-binomial

- Bayesian inference procedure:
  - The prior distribution: $p \sim \text{Beta}(\alpha, \beta)$
  - The sampling density: $Y \sim \text{Binomial}(N, p)$
  - The posterior distribution: $p \mid Y \sim \text{Beta}(\alpha + Y, \beta + N - Y)$

Recap from beta-binomial

- Bayesian inference procedure:
  - The prior distribution: $p \sim \text{Beta}(\alpha, \beta)$
  - The sampling density: $Y \sim \text{Binomial}(N, p)$
  - The posterior distribution: $p \mid Y \sim \text{Beta}(\alpha + Y, \beta + N - Y)$
- What to do for a normal model $Y_i \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma)$?
  - Data model/sampling density is chosen: normal.
  - What to do with two parameters μ and σ?
  - How to specify priors?

Section 2

Conjugate prior and posterior inferences for μ

Overview

- 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.
- Need a joint prior distribution (if both μ and σ are unknown):

$$g(\mu, \sigma).$$

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

$$g(\mu, \sigma \mid Y_1, \cdots, Y_n) \propto g(\mu, \sigma) f(Y_1, \cdots, Y_N \mid \mu, \sigma)$$

If only mean μ is unknown

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

$$Y_i \overset{i.i.d.}{\sim} \mathrm{Normal}(\mu, \sigma).$$

  - Need a prior distribution for μ:

$$g(\mu \mid \sigma).$$

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

$$g(\mu \mid Y_1, \cdots, Y_N, \sigma) \propto g(\mu \mid \sigma) f(Y_1, \cdots, Y_N \mid \mu, \sigma).$$

Normal conjugate prior

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

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

  - The sampling density:

$$Y_1, \cdots, Y_N \mid \mu, \sigma \overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

Normal conjugate prior

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

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

  - The sampling density:

$$Y_1, \cdots, Y_N \mid \mu, \sigma \overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

  - The posterior distribution:

$$\mu \mid Y_1, \cdots, Y_N, \phi \sim \text{Normal}\left(\frac{\phi_0\mu_0 + N\phi\bar{Y}}{\phi_0 + N\phi}, \sqrt{\frac{1}{\phi_0 + N\phi}}\right),$$

  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.

Example on log(Total Expenditure)

- Prior for μ is $\mu \sim \text{Normal}(5, 1)$, i.e. $\mu_0 = 5, \phi_0 = 1$
- Our log(Total Expenditure): $N = 6208$, $\bar{Y} = 8.75$
- Assume $\phi = 1.25$, i.e. $\sigma = \sqrt{1/1.25}$
- Use these quantities to obtain posterior for μ:

$$\mu \mid Y_1, \cdots, Y_N, \phi \sim \text{Normal}\left(\frac{\phi_0\mu_0 + N\phi\bar{Y}}{\phi_0 + N\phi}, \sqrt{\frac{1}{\phi_0 + N\phi}}\right).$$

Posterior for μ

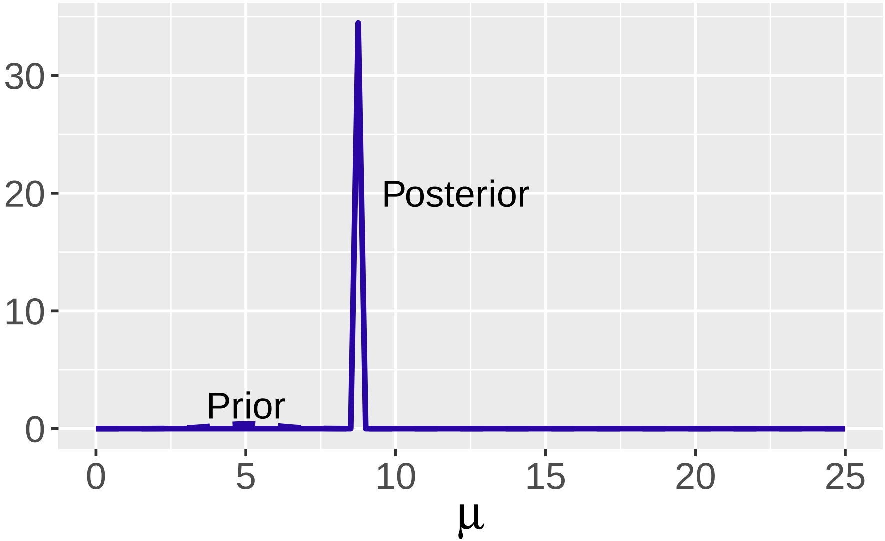
```
mu_0 <- 5
sigma_0 <- 1
phi_0 <- 1/sigma_0^2
ybar <- mean(CEsample$LogTotalExpLastQ)
phi <- 1.25
n <- dim(CEsample)[1]
mu_n <- (phi_0*mu_0+n*ybar*phi)/(phi_0+n*phi)
sd_n <- sqrt(1/(phi_0+n*phi))

mu_n
```

```
[1] 8.747753
```

```
sd_n
```

```
[1] 0.01135118
```

Posterior for μ cont'd



Bayesian inferences: hypothesis testing

- Suppose someone thinks the log(Total Expenditure) of CUs in the U.S. on average is at least \$8.5 (i.e. \$4914), is this statement reasonable?
- Exact solution:

```
1 - pnorm(8.5, mean = mu_n, sd = sd_n)
```

```
[1] 1
```

- Monte Carlo simulation solution:

```
set.seed(123)
S <- 1000
mu_post <- rnorm(S, mean = mu_n, sd = sd_n)
sum(mu_post >= 8.5) / S
```

```
[1] 1
```

Bayesian inferences: credible interval

- Bayesian credible interval: an interval contains the unknown parameter with a certain probability.
- What is a 95% credible interval for μ?
- Exact solution:

```
qnorm(c(0.025, 0.975), mean = mu_n, sd = sd_n)
```

```
[1] 8.725506 8.770001
```

- Monte Carlo simulation solution:

```
quantile(mu_post, c(0.025, 0.975))
```

```
    2.5%    97.5%
8.725714 8.770886
```

Bayesian inferences: prediction

- Suppose we are interested in predicting log(Total Expenditure) of another CU.
- The posterior predictive distribution is

$$f(Y^* \mid Y_1, \cdots, Y_N) = \int f(Y^* \mid \mu, \sigma) g(\mu \mid Y_1, \cdots, Y_N, \sigma) d\mu. \quad (1)$$

- The integration step in Equation (1) can be approximated through simulation.
  - Step 1: Sample a value of μ from its posterior distribution

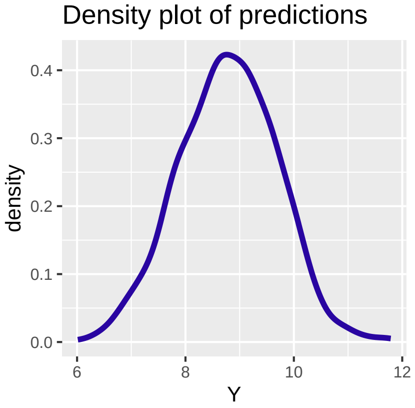
$$\mu \mid Y_1, \cdots, Y_N, \phi \sim \text{Normal}\left(\frac{\phi_0\mu_0 + N\phi\bar{Y}}{\phi_0 + N\phi}, \sqrt{\frac{1}{\phi_0 + N\phi}}\right).$$

  - Step 2: Sample a new observation Y^* from the sampling model

$$Y^* \sim \text{Normal}(\mu, \sigma)$$

Bayesian inferences: prediction cont'd

```
set.seed(123)
S <- 1000
mu_post <- rnorm(S, mean = mu_n, sd = sd_n)
y_pred <- rnorm(S, mean = mu_post, sd = sqrt(1 / phi))
```



Using Stan

- Write a Stan script defining the Bayesian model.

```
data {
  int<lower=0> N;  // number of observations
  real y[N];   // vector of continuous observations
}
parameters {
  real mu; // mean parameter
}
model {
  mu ~ normal(5, 1);  // prior
  for (i in 1:N) {
      y[i] ~ normal(mu, sqrt(1 / 1.25)); // observation model
  }
}
```

Section 3

Inferences for μ and σ

Overview

- The data model/sampling density for N observations:

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

- There are two parameters μ and σ in the normal model.
- Need a joint prior distribution (if both μ and σ are unknown):

$$g(\mu, \sigma).$$

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

$$g(\mu, \sigma \mid Y_1, \cdots, Y_n) \propto g(\mu, \sigma) f(Y_1, \cdots, Y_N \mid \mu, \sigma)$$

Priors for μ and for σ

- Suppose we keep the normal prior for μ:

$$\mu \sim \mathrm{Normal}(\mu_0, \sigma_0).$$

- Now let's also specify a prior for $\sigma > 0$:

$$\sigma \sim \mathrm{Cauchy}(\gamma_1, \gamma_2).$$

- We know that from Bayes' rule, we obtain our joint posterior:

$$g(\mu, \sigma \mid Y_1, \cdots, Y_n) \propto g(\mu)g(\sigma)f(Y_1, \cdots, Y_N \mid \mu, \sigma).$$

Markov chain Monte Carlo methods

- Our goal is to estimate the joint posterior:

$$g(\mu, \sigma \mid Y_1, \cdots, Y_n).$$

- As an approximation, we can iterate by sampling:
  - Sample μ at iteration i:

$$\mu^{(i)} \sim g(\mu \mid Y_1, \cdots, Y_N, \sigma^{(i-1)}).$$

  - Sample σ at iteration i:

$$\sigma^{(i)} \sim g(\sigma \mid Y_1, \cdots, Y_N, \mu^{(i)}).$$

- After convergence, $\{\mu^{(1)}, \cdots, \mu^{(S)}\}$ and $\{\sigma^{(1)}, \cdots, \sigma^{(S)}\}$ serve as approximations to the posterior distribution.

Using Stan

- Write a Stan script defining the Bayesian model.

```
data {
  int<lower=0> N;  // number of observations
  real y[N];   // vector of continuous observations
}
parameters {
  real mu; // mean parameter
  real<lower=0> sigma; // sd parameter
}
model {
  mu ~ normal(5, 1);  // prior for mu
  sigma ~ cauchy(0, 1); // prior for sigma
  for (i in 1:N) {
      y[i] ~ normal(mu, sigma); // observation model
  }
}
```

Run Stan using the rstan package

- Enter data by a list

```
n <- dim(CEsample)[1]
my_data <- list(N = n, y = CEsample$LogTotalExpLastQ)
```

- Inputs are Stan model file and the data list.

```
library(rstan)
fit_normal <- stan(file = "normal_2unknowns.stan",
                          data = my_data,
                          refresh = 0)
```

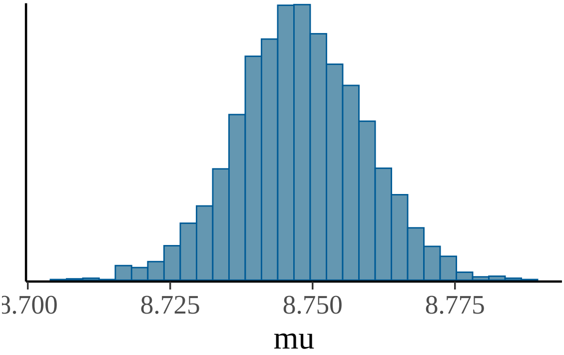
Extract the posterior draws

```r
draws <- as.data.frame(fit_normal)
head(draws)
```

```
        mu     sigma       lp__
1 8.731845 0.8921777 -2419.087
2 8.742834 0.8996197 -2418.376
3 8.750744 0.8951207 -2418.105
4 8.737180 0.8900023 -2418.647
5 8.753385 0.9037658 -2418.871
6 8.749248 0.8940300 -2418.074
```

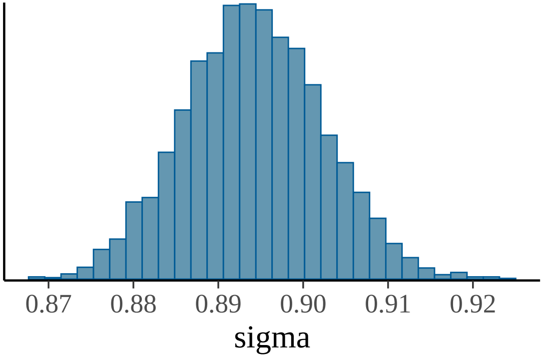
Histogram of the simulated draws of μ

```r
library(bayesplot)
mcmc_hist(draws, pars = 'mu')
```



Histogram of the simulated draws of Σ

```
mcmc_hist(draws, pars = 'sigma')
```



Posterior summaries

```
summary(fit_normal)
```

```
$summary
              mean      se_mean         sd          2.5%           25%
mu       8.7478510 0.0001936590 0.01131484     8.7251974     8.7402609
sigma    0.8941723 0.0001364165 0.00816012     0.8783017     0.8885314
lp__ -2419.0794746 0.0241412019 1.03262094 -2421.8243926 -2419.4754635
               50%          75%        97.5%    n_eff      Rhat
mu       8.7475162     8.755525     8.7703460 3413.673 1.0002326
sigma    0.8940308     0.899628     0.9103544 3578.155 0.9990457
lp__ -2418.7566891 -2418.357007 -2418.0890224 1829.633 1.0004411

$c_summary
, , chains = chain:1

         stats
parameter         mean          sd          2.5%           25%           5
    mu       8.7480701 0.011431011     8.7257254     8.7401911     8.74805
    sigma    0.8941528 0.007983593     0.8787577     0.8885802     0.89414
    lp__ -2419.0671886 0.984825323 -2421.5421235 -2419.4612982 -2418.77608
```

Bayesian inferences: hypothesis testing and credible interval

- Since exact posterior distribution is not available, our inferential methods are mainly Monte Carlo simulation.
- Hypothesis testing: μ at least \$8.5?

```
sum(draws$mu > 8.5) / dim(draws)[1]
```

```
[1] 1
```

- Credible interval: a 95% credible interval for σ?

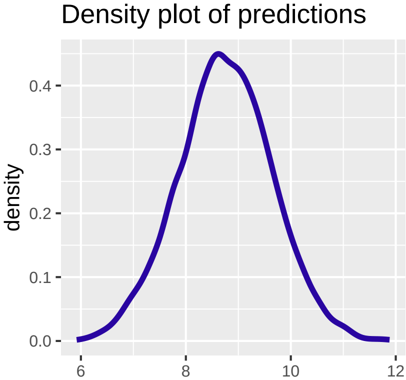
```
quantile(draws$sigma, c(0.025, 0.975))
```

```
     2.5%     97.5%
0.8783017 0.9103544
```

Bayesian inferences: prediction

$$Y^* \sim \text{Normal}(\mu, \sigma)$$

```
set.seed(123)
S <- dim(draws)[1]
y_pred2 <- rnorm(S, draws$mu, draws$sigma)
```



Section 4

Wrap-up and additional material

Wrap-up

- Bayesian inference procedure:
  - Step 1: express an opinion about the location of the parameters before sampling (prior).
  - Step 2: take the sample (data/likelihood).
  - Step 3: use Bayes' rule to sharpen and update the previous opinion about the parameters given the information from the sample (posterior).
- Bayesian inferences
  - Bayesian hypothesis testing
  - Bayesian credible interval
  - Bayesian prediction

Additional material: posterior predictive checks

- A way to check model fitting
- Sample S copies of predictions of the same sample size as the original data

```
set.seed(123)
S <- dim(draws)[1]
sim_ytilde <- function(j){
  rnorm(n, draws$mu, draws$sigma)
}
ytilde <- t(sapply(1:S, sim_ytilde))
```

Posterior predictive checks cont'd

- Use some statistics to check, e.g. the average

```
pred_ybar_sim <- apply(ytilde, 1, mean)
```