Contents

8
Bayesian regression

In this chapter we focus on a linear *DGP*

$$Y = X\beta + \varepsilon$$

uninformative priors for β, and known variance, $\sigma^2 = 1$. The posterior for β is Gaussian or normal with mean $b = \left(X^T X\right)^{-1} X^T y$ and variance $\sigma^2 \left(X^T X\right)^{-1}$

$$p\left(\beta \mid \sigma^2, Y, X\right) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\beta - b\right)^T \left(X^T X\right)\left(\beta - b\right)\right]$$

Next, we return to the (*ANOVA* and *ANCOVA*) examples from the projections chapter but apply Bayesian simulation.

8.1 Mean estimation

As discussed in the classical linear models chapter, we can estimate an unknown mean from a Gaussian distribution via regression where $X = \iota$ (a vector of ones). The posterior distribution is Gaussian with mean $\overline{Y}$ and variance $\frac{\sigma^2}{n}$. For an exchangeable sample, $Y = \{4, 6, 5\}$, we have $\overline{Y} = 5$ and variance $\frac{\sigma^2}{n} = 0.577$. The table below reports statistics from $1,000$

posterior draws.

statistic	$(\beta \mid \sigma^2, Y, X)$
mean	4.99
median	4.99
standard deviation	0.579
maximum	6.83
minimum	3.25
quantiles:	
0.01	3.62
0.025	3.91
0.05	4.04
0.10	4.26
0.25	4.57
0.75	5.39
0.9	5.78
0.95	5.92
0.975	6.06
0.99	6.23

Sample statistics for posterior draws of the mean
$DGP: Y = X\beta + \varepsilon, \quad X = \iota$

These results correspond well with, say, a 95% classical confidence interval $\{3.87, 6.13\}$ while the 95% Bayesian posterior interval is $\{3.91, 6.06\}$.

8.2 Different means (ANOVA)

8.2.1 ANOVA example 1

Suppose we have exchangeable outcome or response data (conditional on D) with two factor levels identified by D.

Y	D
4	0
6	0
5	0
11	1
9	1
10	1

We're interested in estimating the unknown means, or equivalently, the mean difference conditional on D (and one of the means). We view the former DGP as

$$\begin{aligned} Y &= X_1\beta^{(1)} + \varepsilon \\ &= \beta_0 (1 - D) + \beta_1 D + \varepsilon \end{aligned}$$

a no intercept regression where $X_1 = \begin{bmatrix} 1-D & D \end{bmatrix}$ and $\beta^{(1)} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, and the latter DGP as

$$\begin{aligned} Y &= X_2\beta^{(2)} + \varepsilon \\ &= \beta_0 + \beta_2 D + \varepsilon \end{aligned}$$

where $X_2 = \begin{bmatrix} \iota & D \end{bmatrix}$, $\beta^{(2)} = \begin{bmatrix} \beta_0 \\ \beta_2 \end{bmatrix}$,and $\beta_2 = \beta_1 - \beta_0$.

The posterior for $\beta^{(1)}$ is Gaussian with means

$$b^{(1)} = \begin{bmatrix} \left(\overline{Y} \mid D=0\right) & \left(\overline{Y} \mid D=1\right) \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$$

and variance $\sigma^2 \left(X_1^T X_1\right)^{-1} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$.

$$p\left(\beta^{(1)} \mid \sigma^2, Y, X_1\right) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\beta^{(1)} - b^{(1)}\right)^T X_1^T X_1 \left(\beta^{(1)} - b^{(1)}\right)\right]$$

While the posterior for $\beta^{(2)}$ is Gaussian with means

$$b^{(2)} = \begin{bmatrix} \left(\overline{Y} \mid D=0\right) & \left(\overline{Y} \mid D=1\right) - \left(\overline{Y} \mid D=0\right) \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

and variance $\sigma^2 \left(X_2^T X_2\right)^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$.

$$p\left(\beta^{(2)} \mid \sigma^2, Y, X_2\right) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\beta^{(2)} - b^{(2)}\right)^T X_2^T X_2 \left(\beta^{(2)} - b^{(2)}\right)\right]$$

The tables below report statistics from 1,000 posterior draws for each model.

statistic	β_0	β_1	$\beta_1 - \beta_0$
mean	4.99	10.00	5.00
median	4.99	10.00	4.99
standard deviation	0.597	0.575	0.851
maximum	6.87	11.63	7.75
minimum	2.92	8.35	2.45
quantiles:			
0.01	3.64	8.64	3.14
0.025	3.82	8.85	3.37
0.05	4.01	9.06	3.58
0.10	4.23	9.26	3.90
0.25	4.60	9.58	4.44
0.75	5.40	10.39	5.57
0.9	5.76	10.73	6.09
0.95	5.97	10.91	6.36
0.975	6.15	11.08	6.69
0.99	6.40	11.30	7.17
classical 95% interval:			
lower	3.87	8.87	3.40
upper	6.13	11.13	6.60

Sample statistics for posterior draws from *ANOVA* example 1
$DGP: Y = X_1\beta^{(1)} + \varepsilon, \quad X_1 = \left[\begin{array}{cc} (1-D) & D \end{array}\right]$

statistic	β_0	β_2	$\beta_0 + \beta_2$
mean	5.02	4.99	10.01
median	5.02	4.97	10.02
standard deviation	0.575	0.800	0.580
maximum	6.71	7.62	12.09
minimum	3.05	2.88	7.90
quantiles:			
0.01	3.65	3.10	8.76
0.025	3.89	3.43	8.88
0.05	4.05	3.66	9.02
0.10	4.29	3.91	9.22
0.25	4.62	4.49	9.62
0.75	5.41	5.51	10.38
0.9	5.73	6.02	10.75
0.95	5.93	6.30	10.96
0.975	6.14	6.54	11.18
0.99	6.36	6.79	11.33
classical 95% interval:			
lower	3.87	3.40	8.87
upper	6.13	6.60	11.13

Sample statistics for posterior draws from *ANOVA* example 1
$DGP : Y = X_2\beta^{(2)} + \varepsilon, \quad X_2 = \begin{bmatrix} \iota & D \end{bmatrix}$

As expected, for both models there is strong correspondence between classical confidence intervals and the posterior intervals (reported at the 95% level).

8.2.2 ANOVA example 2

Suppose we now have a second binary factor, W.

Y	D	W	$(D \times W)$
4	0	0	0
6	0	1	0
5	0	0	0
11	1	1	1
9	1	0	0
10	1	0	0

We're still interested in estimating the mean differences conditional on D and W. The DGP is

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= \beta_0 + \beta_1 D + \beta_2 W + \beta_3 (D \times W) + \varepsilon \end{aligned}$$

where $X = \begin{bmatrix} \iota & D & W & (D \times W) \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$. The posterior is Gaussian

$$p\left(\beta \mid \sigma^2, Y, X\right) \propto \exp\left[-\frac{1}{2\sigma^2}(\beta - b)^T X^T X (\beta - b)\right]$$

with mean

$$b = \left(X^T X\right)^{-1} X^T Y = \begin{bmatrix} 4.5 \\ 5 \\ 1.5 \\ 0 \end{bmatrix}$$

and variance $\sigma^2 \left(X^T X\right)^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 1 & \frac{1}{2} & -1 \\ -\frac{1}{2} & \frac{1}{2} & \frac{3}{2} & -\frac{3}{2} \\ \frac{1}{2} & -1 & -\frac{3}{2} & 3 \end{bmatrix}$.

The tables below reports statistics from 1,000 posterior draws.

statistic	β_0	β_1	β_2	β_3
mean	4.45	5.05	1.54	−0.03
median	4.47	5.01	1.50	−0.06
standard deviation	0.706	0.989	1.224	1.740
maximum	6.89	7.98	5.80	6.28
minimum	2.16	2.21	−2.27	−5.94
quantiles:				
0.01	2.74	2.76	−1.36	−4.05
0.025	3.06	3.07	−0.87	−3.36
0.05	3.31	3.47	−0.41	−2.83
0.10	3.53	3.82	−0.00	−2.31
0.25	3.94	4.39	0.69	−1.15
0.75	4.93	5.72	2.35	1.13
0.9	5.33	6.34	3.20	2.14
0.95	5.57	6.69	3.58	2.75
0.975	5.78	6.99	3.91	3.41
0.99	6.00	7.49	4.23	4.47
classical 95% interval:				
lower	3.11	3.04	−0.90	−3.39
upper	5.89	6.96	3.90	3.39

Sample statistics for posterior draws from *ANOVA* example 2
$DGP: Y = X\beta + \varepsilon, \quad X = \begin{bmatrix} \iota & D & W & (D \times W) \end{bmatrix}$

As expected, there is strong correspondence between the classical confidence intervals and the posterior intervals (reported at the 95% level).

8.2.3 ANOVA example 3

Now, suppose the second factor, W, is perturbed slightly.

Y	D	W	$(D \times W)$
4	0	0	0
6	0	0	0
5	0	1	0
11	1	1	1
9	1	0	0
10	1	0	0

We're still interested in estimating the mean differences conditional on D and W. The DGP is

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= \beta_0 + \beta_1 D + \beta_2 W + \beta_3 (D \times W) + \varepsilon \end{aligned}$$

where $X = \begin{bmatrix} \iota & D & W & (D \times W) \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$. The posterior is Gaussian

$$p\left(\beta \mid \sigma^2, Y, X\right) \propto \exp\left[-\frac{1}{2\sigma^2}(\beta - b)^T X^T X (\beta - b)\right]$$

with mean

$$b = \left(X^T X\right)^{-1} X^T Y = \begin{bmatrix} 5 \\ 4.5 \\ 0 \\ 1.5 \end{bmatrix}$$

and variance $\sigma^2 \left(X^T X\right)^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 1 & \frac{1}{2} & -1 \\ -\frac{1}{2} & \frac{1}{2} & \frac{3}{2} & -\frac{3}{2} \\ \frac{1}{2} & -1 & -\frac{3}{2} & 3 \end{bmatrix}$.

The tables below reports statistics from $1,000$ posterior draws.

statistic	β_0	β_1	β_2	β_3
mean	5.01	4.49	0.01	1.50
median	5.00	4.49	0.03	1.52
standard deviation	0.700	0.969	1.224	1.699
maximum	7.51	7.45	3.91	8.35
minimum	2.97	1.61	−5.23	−3.34
quantiles:				
0.01	3.46	2.28	−2.84	−2.36
0.025	3.73	2.67	−2.30	−1.82
0.05	3.87	2.93	−1.98	−1.18
0.10	4.08	3.25	−1.56	−0.62
0.25	4.50	3.79	−0.80	0.37
0.75	5.50	5.14	0.82	2.60
0.9	5.93	5.76	1.51	3.61
0.95	6.11	6.09	2.00	4.23
0.975	6.29	6.27	2.36	4.74
0.99	6.51	6.58	2.93	5.72
classical 95% interval:				
lower	3.61	2.54	−2.40	−1.89
upper	6.39	6.46	2.40	4.89

Sample statistics for posterior draws from *ANOVA* example 3
$DGP: Y = X\beta + \varepsilon, \quad X = \begin{bmatrix} \iota & D & W & (D \times W) \end{bmatrix}$

As expected, there is strong correspondence between the classical confidence intervals and the posterior intervals (reported at the 95% level).

8.2.4 ANOVA example 4

Now, suppose the second factor, W, is again perturbed.

Y	D	W	$(D \times W)$
4	0	0	0
6	0	1	0
5	0	1	0
11	1	0	0
9	1	0	0
10	1	1	1

We're again interested in estimating the mean differences conditional on D and W. The DGP is

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= \beta_0 + \beta_1 D + \beta_2 W + \beta_3 (D \times W) + \varepsilon \end{aligned}$$

where $X = \begin{bmatrix} \iota & D & W & (D \times W) \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$. The posterior is Gaussian

$$p\left(\beta \mid \sigma^2, Y, X\right) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\beta - b\right)^T X^T X \left(\beta - b\right)\right]$$

with mean

$$b = \left(X^T X\right)^{-1} X^T Y = \begin{bmatrix} 4 \\ 6 \\ 1.5 \\ -1.5 \end{bmatrix}$$

and variance $\sigma^2 \left(X^T X\right)^{-1} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1.5 & 1 & -1.5 \\ -1 & 1 & 1.5 & -1.5 \\ 1 & -1.5 & -1.5 & 3 \end{bmatrix}$.

The tables below reports statistics from 1,000 posterior draws.

statistic	β_0	β_1	β_2	β_3
mean	3.98	6.01	1.53	−1.46
median	3.95	6.01	1.56	−1.51
standard deviation	1.014	1.225	1.208	1.678
maximum	6.72	9.67	6.40	4.16
minimum	0.71	2.61	−1.78	−6.50
quantiles:				
0.01	1.69	3.28	−1.06	−5.24
0.025	2.02	3.69	−0.84	−4.73
0.05	2.37	3.94	−0.53	−4.26
0.10	2.70	4.34	−0.05	−3.60
0.25	3.27	5.19	0.69	−2.59
0.75	4.72	6.81	2.36	−0.27
0.9	5.28	7.58	3.09	0.73
0.95	5.64	8.08	3.45	1.31
0.975	5.94	8.35	3.81	1.75
0.99	6.36	8.79	4.21	2.20
classical 95% interval:				
lower	2.04	3.60	−0.90	−4.89
upper	5.96	8.40	3.90	1.89

Sample statistics for posterior draws from *ANOVA* example 4
$DGP : Y = X\beta + \varepsilon, \quad X = \begin{bmatrix} \iota & D & W & (D \times W) \end{bmatrix}$

As expected, there is strong correspondence between the classical confidence intervals and the posterior intervals (reported at the 95% level).

8.3 Simple regression

Suppose we don't observe treatment, D, but rather observe only regressor, X_1, in combination with outcome, Y.

Y	X_1
4	-1
6	1
5	0
11	1
9	-1
10	0

For the perceived DGP

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

with $\varepsilon \sim N\left(0, \sigma^2 I\right)$ (σ^2 known) the posterior distribution for β is

$$p\left(\beta \mid Y, X\right) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\beta - b\right)^T X^T X \left(\beta - b\right)\right]$$

where $X = \left[\begin{array}{cc} \iota & X_1 \end{array}\right]$, $b = \left(X^T X\right)^{-1} X^T Y = \left[\begin{array}{c} 7.5 \\ 1 \end{array}\right]$, and $\sigma^2 \left(X^T X\right)^{-1} = \left[\begin{array}{cc} \frac{1}{6} & 0 \\ 0 & \frac{1}{4} \end{array}\right]$.

Sample statistics for $1,000$ posterior draws are tabulated below.

statistic	β_0	β_1
mean	7.48	0.98
median	7.49	1.01
standard deviation	0.386	0.482
maximum	8.67	2.39
minimum	6.18	−0.78
quantiles:		
0.01	6.55	−0.14
0.025	6.71	0.08
0.05	6.85	0.16
0.10	6.99	0.34
0.25	7.22	0.65
0.75	7.74	1.29
0.9	7.98	1.60
0.95	8.09	1.76
0.975	8.22	1.88
0.99	8.33	2.10
classical 95% interval:		
lower	6.20	0.02
upper	7.80	1.98

Sample statistics for posterior draws from *simple regression* example
$DGP: Y = X\beta + \varepsilon, \quad X = \begin{bmatrix} \iota & X_1 \end{bmatrix}$

Classical confidence intervals and Bayesian posterior intervals are similar, as expected.

8.4 ANCOVA example

Suppose in addition to the regressor, X_1, we observe treatment, D, in combination with outcome, Y.

Y	X_1	D
4	−1	0
6	1	0
5	0	0
11	1	1
9	−1	1
10	0	1

For the perceived *DGP*

$$Y = \beta_0 + \beta_1 D + \beta_2 X_1 + \beta_3 (D \times X_1) + \varepsilon$$

with $\varepsilon \sim N\left(0, \sigma^2 I\right)$ (σ^2 known) the posterior distribution for β is

$$p(\beta \mid Y, X) \propto \exp\left[-\frac{1}{2\sigma^2}(\beta - b)^T X^T X (\beta - b)\right]$$

where $X = \begin{bmatrix} \iota & D & X_1 & (D \times X_1) \end{bmatrix}$, $b = \left(X^T X\right)^{-1} X^T Y = \begin{bmatrix} 5 \\ 5 \\ 1 \\ 0 \end{bmatrix}$, and

$$\sigma^2 \left(X^T X\right)^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix}$$

Sample statistics for $1,000$ posterior draws are tabulated below.

statistic	β_0	β_1	β_2	β_3
mean	5.00	4.99	1.02	0.01
median	5.02	5.02	1.02	0.01
standard deviation	0.588	0.802	0.697	1.00
maximum	7.23	7.96	3.27	3.21
minimum	3.06	2.69	-1.33	-3.31
quantiles:				
0.01	3.62	3.15	-0.57	-2.34
0.025	3.74	3.38	-0.41	-2.03
0.05	4.02	3.65	-0.10	-1.66
0.10	4.28	3.98	0.14	-1.25
0.25	4.60	4.47	0.54	-0.66
0.75	5.34	5.53	1.48	0.65
0.9	5.73	5.96	1.91	1.25
0.95	5.97	6.30	2.16	1.68
0.975	6.20	6.50	2.37	2.07
0.99	6.49	6.93	2.72	2.36
classical 95% interval:				
lower	3.87	3.40	-0.39	-1.96
upper	6.13	6.60	2.39	1.96

Sample statistics for posterior draws from the *ANCOVA* example
$DGP : Y = X\beta + \varepsilon, \quad X = \begin{bmatrix} \iota & D & X_1 & (D \times X_1) \end{bmatrix}$

Classical confidence intervals and Bayesian posterior intervals are similar, as expected.