Introduction to regression models - Outline

- Linear regression
  - Classical regression
  - Default Bayesian regression
  - Conjugate subjective Bayesian regression
- Simulating from the posterior
  - Inference on functions of parameters

Linear Regression

Basic idea

- understand the relationship between response y and explanatory variables $x = (x_1, \ldots, x_k)$
- based on data from experimental units index by i.

If we assume

- linearity, independence, normality, and constant variance,

then we have

$$y_i \overset{ind}{\sim} N(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \sigma^2)$$

where $x_{i1} = 1$ if we want to include an intercept. In matrix notation, we have

$$y \sim N(X\beta, \sigma^2 I)$$

where $y = (y_1, \ldots, y_n)^T$, $\beta = (\beta_1, \ldots, \beta_k)^T$, and X is an $n \times k$ full-rank matrix with each row being $x_i = (x_{i1}, \ldots, x_{ik})$.

Classical regression

How do you find confidence intervals for β?

What is the MLE for β?

$$\hat{\beta} = \hat{\beta}_{MLE} = (X^T X)^{-1} X^T y$$

What is the sampling distribution for $\hat{\beta}$?

$$\hat{\beta} \sim t_{n-k}(\beta, s^2 (X^T X)^{-1})$$

where $s^2 = SSE/[n-k]$ and $SSE = (Y - X\hat{\beta})^T (Y - X\hat{\beta})$.

What is the sampling distribution for s^2?

$$\frac{[n-k]s^2}{\sigma^2} \sim \chi^2_{n-k}$$

Default Bayesian regression

Assume the standard noninformative prior

$$p(\beta, \sigma^2) \propto 1/\sigma^2$$

then the posterior is

$$\begin{aligned} p(\beta, \sigma^2|y) &= p(\beta|\sigma^2, y)p(\sigma^2|y) \\ \beta|\sigma^2, y &\sim N(\hat{\beta}, \sigma^2 V_\beta) \\ \sigma^2|y &\sim \text{Inv-}\chi^2(n-k, s^2) \\ \beta|y &\sim t_{n-k}(\hat{\beta}, s^2 V_\beta) \end{aligned}$$

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T y \\ V_\beta &= (X^T X)^{-1} \\ s^2 &= \tfrac{1}{n-k}(y - X\hat{\beta})^T (y - X\hat{\beta}) \end{aligned}$$

The posterior is proper if $n > k$ and $\text{rank}(X) = k$.

Comparison to classical regression

In classical regression, we have fixed, but unknown, true parameters β_0 and σ_0^2 and quantify our uncertainty about these parameters using the sampling distribution of the error variance and regression coefficients, i.e.

$$\frac{[n-k]s^2}{\sigma_0^2} \sim \chi^2_{n-k}$$

and

$$\hat{\beta} \sim t_{n-k}(\beta_0, s^2(X^T X)^{-1}).$$

In the default Bayesian regression, we still have the fixed, but unknown, true parameters, but quantify our uncertainty about these parameters using prior and posterior distributions, i.e.

$$\left.\frac{s^2[n-k]}{\sigma^2}\right| y \sim \chi^2(n-k)$$

and

$$\beta | y \sim t_{n-k}(\hat{\beta}, s^2(X^T X)^{-1}).$$

Fully conjugate subjective Bayesian inference

If we assume the following normal-gamma prior,

$$\beta|\sigma^2 \sim N(m_0, \sigma^2 C_0) \qquad \sigma^2 \sim \text{Inv-}\chi^2(v_0, s_0^2)$$

then the posterior is

$$\beta|\sigma^2, y \sim N(m_n, \sigma^2 C_n) \qquad \sigma^2|y \sim \text{Inv-}\chi^2(v_n, s_n^2)$$

with

$$\begin{aligned}
m_n &= m_0 + C_0 X^T (X C_0 X^T + I)^{-1}(y - X m_0) \\
C_n &= C_0 - C_0 X^T (X C_0 X^T + I)^{-1} X C_0 \\
v_n &= v_0 + n \\
v_n s_n^2 &= v_0 s_0^2 + (y - X m_0)^T (X C_0 X^T + I)^{-1}(y - X m_0)
\end{aligned}$$

Simulating from the posterior

Although the full posterior for β and σ^2 is available, the decomposition

$$p(\beta, \sigma^2 | y) = p(\beta | \sigma^2, y) p(\sigma^2 | y)$$

suggests an approach to simulating from the posterior via

1. $(\sigma^2)^{(j)} \sim \text{Inv-}\chi^2(n-k, s^2)$ and
2. $\beta^{(j)} \sim N(\hat{\beta}, (\sigma^2)^{(j)} V_\beta)$.

This also provides an approach to obtaining posteriors for any function $\gamma = f(\beta, \sigma^2)$ of the parameters via

$$\begin{aligned} p(\gamma|y) &= \int\int p(\gamma|\beta,\sigma^2,y)p(\beta|\sigma^2,y)p(\sigma^2|y)d\beta d\sigma^2 \\ &= \int\int p(\gamma|\beta,\sigma^2)p(\beta|\sigma^2,y)p(\sigma^2|y)d\beta d\sigma^2 \\ &= \int\int I(\gamma = f(\beta,\sigma^2))p(\beta|\sigma^2,y)p(\sigma^2|y)d\beta d\sigma^2 \end{aligned}$$

by adding the step

3. $\gamma^{(j)} = f(\beta^{(j)}, (\sigma^2)^{(j)})$.

Summary

- Model: $y \sim N(X\beta, \sigma^2 \mathrm{I})$
- Default Bayesian analysis corresponds exactly to classical regression analysis

$$p(\beta, \sigma^2) \propto 1/\sigma^2 \implies$$

$$\beta|\sigma^2, y \sim N(\hat{\beta}, \sigma^2 [X^T X]^{-1}), \sigma^2|y \sim \text{Inv-}\chi^2(n-k, s^2)$$

- Conjugate subjective Bayesian analysis:

$$\beta|\sigma^2 \sim N(m_0, \sigma^2 C_0), \sigma^2 \sim \text{Inv-}\chi^2(v_0, s_0^2) \implies$$

$$\beta|\sigma^2, y \sim N(m_n, \sigma^2 C_n), \sigma^2|y \sim \text{Inv-}\chi^2(v_n, s_n^2)$$

- Obtain functions of parameters and their uncertainty by simulating the parameters from their joint posterior, calculating the function, and taking posterior quantiles.

Computation

For numerical stability and efficiency, the QR decomposition can be used to calculate posterior quantities.

Definition

For an $n \times k$ matrix X, a QR decomposition is $X = QR$ for an $n \times k$ matrix Q with orthonormal columns and a $k \times k$ upper triangular matrix R.

The quantities of interest are

$$\begin{aligned} V_\beta &= (X^T X)^{-1} = ([QR]^T QR)^{-1} = (R^T Q^T QR)^{-1} = (R^T R)^{-1} \\ &= R^{-1}[R^T]^{-1} \end{aligned}$$

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T y = R^{-1}[R^T]^{-1} R^T Q^T y = R^{-1} Q^T y \\ R\hat{\beta} &= Q^T y \end{aligned}$$

The last equation is useful because R is upper triangular and therefore the system of linear equations can be solved without requiring the inverse of R.

Subjective Bayesian regression

Suppose

$$y \sim N(X\beta, \sigma^2 I)$$

and we use a prior for β of the form

$$\beta | \sigma^2 \sim N(b, \sigma^2 B)$$

A few special cases are

- $b = 0$
- $B = g(X^T X)^{-1}$

Ridge regression

Let

$$y = X\beta + e, \quad E[e] = 0, \quad Var[e] = \sigma^2 I$$

then ridge regression seeks to minimize

$$(y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$$

where λ is a penalty for $\beta^T\beta$ getting too large.

This minimization looks like -2 times the log posterior for a Bayesian regression analysis when using independent normal priors centered at zero with a common variance (τ^2) for β:

$$-2\sigma^2 \log p(\beta, \sigma | y) = C + (y - X\beta)^T(y - X\beta) + \frac{\sigma^2}{\tau^2}\beta^T\beta$$

where $\lambda = \sigma^2/\tau^2$. Thus the ridge regression estimate is equivalent to a MAP estimate when

$$y \sim N(X\beta, \sigma^2 I) \quad \beta \sim N(0, \tau^2 I).$$

Zellner's g-prior

Suppose

$$y \sim N(X\beta, \sigma^2 I)$$

and you use Zellner's g-prior

$$\beta \sim N(b_0, g\sigma^2(X'X)^{-1}).$$

The posterior is then

$$\begin{aligned} \beta|\sigma^2, y &\sim N\left(\tfrac{g}{g+1}\left(\tfrac{b_0}{g} + \hat{\beta}\right), \tfrac{\sigma^2 g}{g+1}(X'X)^{-1}\right) \\ \sigma^2|y &\sim \text{Inv-}\chi^2\left(n, \tfrac{1}{n}\left[(n-k)s^2 + \tfrac{1}{g+1}(\hat{\beta} - b_0)X'X(\hat{\beta} - b_0)\right]\right) \end{aligned}$$

with

$$\begin{aligned} E[\beta|y] &= \tfrac{1}{g+1} b_0 + \tfrac{g}{g+1}\hat{\beta} \\ E[\sigma^2|y] &= \frac{(n-k)s^2 + \frac{1}{g+1}(\hat{\beta}-b_0)X'X(\hat{\beta}-b_0)}{n-2} \end{aligned}$$

Setting g

In Zellner's g-prior,

$$\beta \sim N(b_0, g\sigma^2(X'X)^{-1}),$$

we need to determine how to set g.

Here are some thoughts:

- $g = 1$ puts equal weight to prior and likelihood
- $g = n$ means prior has the equivalent weight of 1 observation
- $g \to \infty$ recovers a uniform prior
- Put a prior on g and perform a fully Bayesian analysis.