STAT J535: Chapter 4: The Bayesian Linear Regression Model

David B. Hitchcock
E-Mail: hitchcock@stat.sc.edu

Spring 2014

Setup of Linear Regression Model

- We now consider the **regression model** in which a response variable Y is related to one or more **explanatory** or **predictor** variables $X_1, X_2, \ldots, X_{k-1}$.
- For a random sample of n individuals, our model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{k-1} X_{i,k-1} + \epsilon_i, \; \epsilon_i \overset{\text{indep}}{\sim} N(0, \sigma^2)$$

Setup of Linear Regression Model

- This model can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,k-1} \\ 1 & X_{21} & \cdots & X_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{n,k-1} \end{bmatrix},$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix}$$

Likelihood for Linear Regression Model

- Based on this normal model, the likelihood is:

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}$$

- Note that the **least squares** estimates of $\boldsymbol{\beta}$ and σ^2 are:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})}{n-k}$$

Likelihood for Linear Regression Model

Then $L(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y})$

$$\propto \sigma^{-n} \exp\{-\tfrac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})\}$$

$$= \sigma^{-n} \exp\Big\{-\tfrac{1}{2\sigma^2}\Big(\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$
$$- 2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]'\mathbf{X}'\mathbf{y} + 2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]'\mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]\Big)\Big\}$$

Since $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\mathbf{b}}$,

$$= \sigma^{-n} \exp\Big\{-\tfrac{1}{2\sigma^2}\Big(\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$
$$- 2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}]'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}$$
$$+ 2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}]'\mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}]\Big)\Big\}$$

Likelihood for Linear Regression Model

$$= \sigma^{-n} \exp\Big\{-\tfrac{1}{2\sigma^2}\Big(\mathbf{y}'\mathbf{y} - 2\hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} + \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + 2\hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}$$

$$- \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} - 2\hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + 2\hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} - 2\beta'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + \beta'\mathbf{X}'\mathbf{X}\beta\Big)\Big\}$$

$$= \sigma^{-n} \exp\Big\{-\tfrac{1}{2\sigma^2}[(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) + \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}}$$

$$- 2\beta'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + \beta'\mathbf{X}'\mathbf{X}\beta]\Big\}$$

$$= \sigma^{-n} \exp\Big\{-\tfrac{1}{2\sigma^2}[\hat{\sigma}^2(n-k) + (\beta - \hat{\mathbf{b}})'\mathbf{X}'\mathbf{X}(\beta - \hat{\mathbf{b}})]\Big\}$$

Noninformative Priors for β and σ^2

Consider the independent vague priors

$$p(\boldsymbol{\beta}) \propto 1, \quad \boldsymbol{\beta} \in (-\infty, \infty)^k$$
$$\text{and } p(\sigma^2) = \frac{1}{\sigma}, \quad \sigma \in (0, \infty)$$

Then the joint posterior for $\boldsymbol{\beta}$ and σ^2 is:

$$\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) p(\boldsymbol{\beta}) p(\sigma^2)$$
$$\propto \sigma^{-n-1} \exp\{-\tfrac{1}{2\sigma^2}[\hat{\sigma}^2(n-k) + (\boldsymbol{\beta} - \hat{\mathbf{b}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\mathbf{b}})]\}$$

Noninformative Priors for $\boldsymbol{\beta}$ and σ^2

- Using the transformation $s = \sigma^{-2}$ with Jacobian $|J| = \frac{1}{2}s^{-3/2}$:

$$\begin{aligned}\pi(\boldsymbol{\beta}, s|\mathbf{X}, \mathbf{y}) &\propto (s^{-1/2})^{-n-1} \exp\Big\{-\tfrac{1}{2}s[\hat{\sigma}^2(n-k) \\ &\qquad + (\boldsymbol{\beta} - \hat{\mathbf{b}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\mathbf{b}})]\Big\}\Big(\tfrac{1}{2}s^{-3/2}\Big) \\ &\propto (s)^{\frac{n}{2}-1} \exp\{-\tfrac{1}{2}s[\hat{\sigma}^2(n-k) + (\boldsymbol{\beta} - \hat{\mathbf{b}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\mathbf{b}})]\}\end{aligned}$$

Noninformative Priors for $\boldsymbol{\beta}$ and σ^2

- To get the marginal posterior for $\boldsymbol{\beta}$, integrate out s:

So $\pi(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y})$

$$= \int_0^\infty (s)^{\frac{n}{2}-1} \exp\{-\tfrac{1}{2}[\hat{\sigma}^2(n-k) + (\boldsymbol{\beta} - \hat{\mathbf{b}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\mathbf{b}})]s\}\, ds$$

$$= \frac{\Gamma(\frac{n}{2})}{\frac{1}{2}[\hat{\sigma}^2(n-k) + (\boldsymbol{\beta} - \hat{\mathbf{b}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\mathbf{b}})]^{\frac{n}{2}}}$$

$$\propto [(n-k) + (\boldsymbol{\beta} - \hat{\mathbf{b}})'\hat{\sigma}^{-2}\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\mathbf{b}})]^{-\frac{n}{2}}$$

- This is the kernel of a multivariate t-distribution with $(n-k)$ degrees of freedom and covariance matrix

$$\frac{(n-k)\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}}{n-k-2}$$

Noninformative Priors for β and σ^2

- Now we integrate $\boldsymbol{\beta}$ out of the joint posterior to get the marginal posterior for σ^2:

$$
\begin{aligned}
\pi(\sigma^2|\mathbf{X},\mathbf{y}) &\propto (\sigma)^{-n-1} e^{-\frac{1}{2\sigma^2}\hat{\sigma}^2(n-k)} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\hat{\mathbf{b}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\hat{\mathbf{b}})}\, \mathrm{d}\boldsymbol{\beta} \\
&\propto (\sigma)^{-n-1} e^{-\frac{1}{2\sigma^2}\hat{\sigma}^2(n-k)} (2\pi\sigma^2)^{k/2} \\
&\propto (\sigma^2)^{-\frac{1}{2}(n-k-1)-1} e^{-\frac{\frac{1}{2}\hat{\sigma}^2(n-k)}{\sigma^2}}
\end{aligned}
$$

  which is clearly an $\mathrm{IG}(\frac{1}{2}(n-k-1), \frac{1}{2}\hat{\sigma}^2(n-k))$ posterior distribution.
- Example: Oxygen update data on course web page