Bayesian Linear Regression

Adam Smith, Spring 2017

# 1 Introduction

The standard linear regression model measures the relationship between a response variable $y$ and a set of predictor variables $x_1, \ldots, x_k$. In its simplest form, the model is specified as

$$y_i = x_i'\beta + \varepsilon_i \quad \text{for } i = 1, \ldots, n \tag{1}$$

where the set of errors $\{\varepsilon_i\}$ are assumed to be iid $N(0, \sigma^2)$ random variables. This parametric assumption on the error terms induces a distribution on $y$ given $x_i$. In particular, we have $y|x_i \sim N(x_i'\beta, \sigma^2)$. Collecting the predictor variables into a matrix $X$ allows us to rewrite the model using matrix notation:

$$y = X\beta + \varepsilon. \tag{2}$$

Here, $y$ is an $n$-dimensional vector of response variables, $X$ is an $n \times k$ design matrix, $\beta$ is a $k$-dimensional vector of regression coefficients, and $\varepsilon$ is an $n$-dimensional vector of errors assumed to have a multivariate $N(0, \Sigma)$ distribution. This implies that the conditional distribution of $y$ is $N(X\beta, \Sigma)$. We allow for a more general covariance structure of the errors by only requring $\Sigma$ to be an $n \times n$ positive definite matrix. To match the standard linear regression specification of (1), however, we can simply let $\Sigma = \sigma^2 I_n$.

# 2 Bayesian Linear Model

A Bayesian linear regression model reformulates the standard linear regression model in (2) for Bayesian inference. In any Bayesian model, it is necessary to specify a distribution for the data (likelihood) and a distribution for all model parameters (prior). Assuming the error vector is distributed $N(0, \sigma^2 I_n)$ yields a multivariate normal likelihood:

$$p(y|X, \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}. \tag{3}$$

Next, we must specify a prior distribution for $(\beta, \sigma^2)$. We choose conjugate priors to ensure a known form for the posterior distribution as well as analytic expressions for posterior moments. With both $\beta$ and $\sigma^2$ unknown, the conjugate prior is specified as

$$p(\beta, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2) \tag{4}$$

where

$$\beta|\sigma^2 \sim N(\bar{\beta}, \sigma^2 A^{-1})$$

and

$$\sigma^2 \sim \frac{\nu_0 s_0^2}{\chi^2_{\nu_0}}.$$

Hence, $\beta|\sigma^2$ is multivariate normal and $\sigma^2$ is inverse-gamma. The joint posterior then takes the form:

$$\begin{aligned}
p(\beta, \sigma^2 | y, X) & \qquad (5) \\
&\propto p(y|X, \beta, \sigma^2)\ p(\beta, \sigma^2) \\
&= p(y|X, \beta, \sigma^2)\ p(\beta|\sigma^2)\ p(\sigma^2) \\
&\propto (\sigma^2)^{-\frac{n}{2}}\ \exp\left\{\frac{-1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \\
&\qquad \times (\sigma^2)^{-\frac{k}{2}}\ \exp\left\{\frac{-1}{2\sigma^2}(\beta - \bar{\beta})'A(\beta - \bar{\beta})\right\} \\
&\qquad \times (\sigma^2)^{-(\frac{\nu_0}{2}+1)}\ \exp\left\{\frac{-\nu_0 s_0^2}{2\sigma^2}\right\}.
\end{aligned}$$

Now, the goal is to derive the marginal posterior distributions of $\beta|\sigma^2$ and $\sigma^2$. To do this, we simplify (5) by noticing that $p(y|X, \beta, \sigma^2)$ and $p(\beta|\sigma^2)$ both contain quadratic forms in $\beta$. The first step is to expand out the sum of the two quadratic forms.

$$\begin{aligned}
(y - X\beta)'(y - X\beta) + (\beta - \bar{\beta})'A(\beta - \bar{\beta}) & \qquad (6) \\
&= (y' - \beta'X')(y - X\beta) + (\beta' - \bar{\beta}')A(\beta - \bar{\beta}) \\
&= y'y - \underbrace{y'X\beta - \beta'X'y}_{y'X\beta = (y'X\beta)'} + \beta'X'X\beta + \beta'A\beta - \underbrace{\beta'A\bar{\beta} - \bar{\beta}'A\beta}_{\beta'A\bar{\beta} = (\beta'A\bar{\beta})'} + \bar{\beta}'A\bar{\beta} \\
&= \beta'X'X\beta + \beta'A\beta - 2\beta'X'y - 2\beta'A\bar{\beta} + y'y + \bar{\beta}'A\bar{\beta}
\end{aligned}$$

The last line uses the fact that $y'X\beta$ and $\beta'A\bar{\beta}$ are both scalars, so $y'X\beta = (y'X\beta)'$ and $\beta'A\bar{\beta} = (\beta'A\bar{\beta})'$. Moreover, we can write

$$X'y = (X'X)(X'X)^{-1}X'y = X'X\hat{\beta} \qquad (7)$$

so that (6) reduces to

$$\left[\beta'(X'X + A)\beta - \beta'(2X'X\hat{\beta} + 2A\bar{\beta})\right] + y'y + \bar{\beta}'A\bar{\beta}. \qquad (8)$$

We can further simplify the terms in $[\cdot]$ by completing the square in $\beta$.

## 3 Completing the Square

The matrix version of completing the square is given by:

$$X'MX + X'n + p = (X-h)'M(X-h) + k \tag{9}$$

where $h = -\frac{1}{2}M^{-1}n$ and $k = p - \frac{1}{4}n'M^{-1}n$. Next, we plug in the matrices from (8) into the general form given above.

$$\begin{aligned} M &= X'X + A \\ n &= -2(X'X\hat{\beta} + A\bar{\beta}) \\ h &= (X'X + A)^{-1}(X'X\hat{\beta} + A\bar{\beta}) \\ k &= -(X'X\hat{\beta} + A\bar{\beta})'(X'X + A)^{-1}(X'X\hat{\beta} + A\bar{\beta}) \\ p &= 0 \end{aligned}$$

If we let $\tilde{\beta} = h = (X'X + A)^{-1}(X'X\hat{\beta} + A\bar{\beta})$, then the bracketed terms in (8) become

$$\begin{aligned} &\beta'(X'X + A)\beta - \beta'(2X'X\hat{\beta} + 2A\bar{\beta}) \\ &= (\beta - \tilde{\beta})'(X'X + A)(\beta - \tilde{\beta}) - (X'X\hat{\beta} + A\bar{\beta})'(X'X + A)^{-1}(X'X\hat{\beta} + A\bar{\beta}) \\ &= (\beta - \tilde{\beta})'(X'X + A)(\beta - \tilde{\beta}) - (X'X\hat{\beta} + A\bar{\beta})'\tilde{\beta} \end{aligned} \tag{10}$$

Now since $(X'X + A)^{-1}$ is symmetric and $I = \Big[(X'X + A)^{-1}(X'X + A)\Big]$, we can write the rightmost term above as

$$\begin{aligned} (X'X\hat{\beta} + A\bar{\beta})'\tilde{\beta} &= (X'X\hat{\beta} + A\bar{\beta})'\Big[(X'X + A)^{-1}(X'X + A)\Big]\tilde{\beta} \\ &= (X'X\hat{\beta} + A\bar{\beta})'\big((X'X + A)^{-1}\big)'(X'X + A)\tilde{\beta} \\ &= \Big[(X'X + A)^{-1}(X'X\hat{\beta} + A\bar{\beta})\Big]'(X'X + A)\tilde{\beta} \\ &= \tilde{\beta}'(X'X + A)\tilde{\beta}. \end{aligned} \tag{11}$$

Therefore, using the results of equations (8), (10), and (11), (6) simplifies to

$$\begin{aligned} (y - &X\beta)'(y - X\beta) + (\beta - \bar{\beta})'A(\beta - \bar{\beta}) \\ &= (\beta - \tilde{\beta})'(X'X + A)(\beta - \tilde{\beta}) + y'y + \bar{\beta}'A\bar{\beta} - \tilde{\beta}'(X'X + A)\tilde{\beta}. \end{aligned} \tag{12}$$

The joint posterior distribution is then

$$
\begin{aligned}
p(\beta, \sigma^2 | y, X) \propto (\sigma^2)^{-\frac{n}{2}} \ \exp\left\{ \frac{-1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right\} \\
\times (\sigma^2)^{-\frac{k}{2}} \ \exp\left\{ \frac{-1}{2\sigma^2} (\beta - \bar{\beta})' A (\beta - \bar{\beta}) \right\} \\
\times (\sigma^2)^{-(\frac{\nu_0}{2}+1)} \ \exp\left\{ \frac{-\nu_0 s_0^2}{2\sigma^2} \right\} \\
= (\sigma^2)^{-\frac{k}{2}} \ \exp\left\{ \frac{-1}{2\sigma^2} (\beta - \tilde{\beta})'(X'X + A)(\beta - \tilde{\beta}) \right\} \\
\times (\sigma^2)^{-(\frac{n+\nu_0}{2}+1)} \ \exp\left\{ \frac{-1}{2\sigma^2} \Big( \nu_0 s_0^2 + y'y + \bar{\beta}' A \bar{\beta} - \tilde{\beta}'(X'X + A)\tilde{\beta} \Big) \right\}.
\end{aligned}
\tag{13}
$$

But now we see that the joint posterior distribution factors into two parts: the conditional posterior of $\beta|\sigma^2$ and the marginal posterior of $\sigma^2$. Formally, we have

$$
\beta|\sigma^2, y, X \sim N\Big(\tilde{\beta}, \sigma^2 (X'X + A)^{-1}\Big) \tag{14}
$$

and

$$
\sigma^2 | y, X \sim IG\Big(a_n, b_n\Big), \tag{15}
$$

where

$$
\begin{aligned}
\tilde{\beta} &= (X'X + A)^{-1}(X'X\hat{\beta} + A\bar{\beta}) \\
a_n &= \frac{1}{2}(n + \nu_0) \\
b_n &= \nu_0 s_0^2 + \frac{1}{2}\Big( y'y + \bar{\beta}' A \bar{\beta} - \tilde{\beta}'(X'X + A)\tilde{\beta} \Big).
\end{aligned}
$$

# References


Rossi, P. E., G. M. Allenby, and R. McCulloch (2005), *Bayesian Statistics and Marketing*. New York: John Wiley and Sons.