Bayesian Linear Models

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

02/28/2024

Linear model

▶ For observations $y_1, \ldots, y_n$, the basic linear model is

$$y_i = x_{1i}\beta_1 + \ldots + x_{pi}\beta_p + \epsilon_i,$$

  ▶ $x_{1i}, \ldots, x_{pi}$ are predictors for the ith observation.

  ▶ ϵ_i are error terms.

▶ In matrix form:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

  ▶ $\mathbf{y} = (y_1, \ldots, y_n)$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$

  ▶ X is the matrix with entries $X_{ij} = x_{ij}$

Linear model

- ▶ Assume X is fixed (non-random)

- ▶ Assume errors are normal and iid with equal variance:

$$\epsilon \sim \text{Normal}(\mathbf{0}, \sigma^2 I).$$

- ▶ Standard frequentist estimates are

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y} \text{ and}$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-p} (\mathbf{y} - X\hat{\beta})^T (\mathbf{y} - X\hat{\beta}).$$

  - ▶ These estimates are unbiased, and can be motivated by least-squares.

- ▶ Under a Bayesian framework, we put a prior on β and σ^2.

- Consider uniform prior for β and Jeffreys prior for σ^2:

$$p(\beta, \sigma^2) \propto \frac{1}{\sigma^2}.$$

- The posterior for β, given σ^2, is

$$p(\beta \mid \mathbf{y}, \sigma^2) = \text{Normal}\left(\hat{\beta}, \sigma^2(X^T X)^{-1})\right)$$

- The marginal posterior of σ^2 is

$$p(\sigma^2 \mid \mathbf{y}) = IG\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right)$$

- Equivalently:

$$\sigma^2 \sim \frac{(n-p)s^2}{U} \quad \text{where } U \sim \chi^2_{(n-p)}.$$

Uninformative priors

▶ The marginal posterior for β_i is a non-central t-distribution:

$$\frac{\beta_i - \hat{\beta}_i}{s\sqrt{(X^TX)^{-1}_{ii}}} \sim t_{n-p}.$$

▶ For a new predictor vector $\mathbf{x}_{(n+1)}$, the posterior predictive for y_{n+1} is also a non-central t-distribution:

$$\frac{y_{n+1} - \mathbf{x}_{n+1}\hat{\beta}}{s\sqrt{1 + \mathbf{x}_{n+1}(X^TX)^{-1}\mathbf{x}_{n+1}}} \sim t_{n-p}.$$

▶ All given results for $p(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$ correspond to standard frequentist inference for linear regression!

Example: Body Fat

▶ The % body fat ($BF\%$) is measured for 100 adult males. [1]
  ▶ Using sophisticated and precise technique (water immersion)

▶ Also measure the following for each person:
  ▶ 1: *Age* (in years)
  ▶ 2: *Weight* (in pounds)
  ▶ 3: *Height* (in inches)
  ▶ Circumference of the *neck* (4), *chest* (5), *abdomen* (6), *ankle* (7), *bicep* (8), and *wrist* (9) in cm.

▶ Data available at
  http://www.lock5stat.com/datasets1e/BodyFat.csv

▶ Would like to predict $BF\%$ from the 9 additional measurements

---

[1]Johnson, R. "Fitting Percentage Body Fat to Simple Body Measurements," *Journal of Statistics Education*, 1996.

- Assume $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_{100})$ give $BF\%$ for subjects $1, \ldots, 100$
  - $\bar{\tilde{y}} = 18.6\%$
  - $s_{\tilde{y}} = 8.01\%$

- Let $X : 100 \times 9$ be the matrix of standardized predictors

$$X_{i,j} = \frac{\tilde{x}_{i,j} - \text{mean}(\tilde{\mathbf{x}}_{\cdot,j})}{\text{stdev}(\tilde{\mathbf{x}}_{\cdot,j})}$$

  - $\tilde{X}_{i,j}$ is measurement j (unstandardized) for subject i

- The mean BF% for american adult men is 18.5%

- For $\mathbf{y} = \tilde{\mathbf{y}} - \mathbf{18.5}$ consider the model

$$\mathbf{y} = \beta X + \epsilon$$

Example: Body Fat

- ▶ Assume $\epsilon \sim \text{Normal}(\mathbf{0}, \sigma^2 I)$
- ▶ Use uninformative prior:

$$p(\beta, \sigma^2) = \frac{1}{\sigma^2}$$

- ▶ Recall $p(\beta_i \mid \mathbf{y})$ is a non-central t:

$$\frac{\beta_i - \hat{\beta}_i}{s\sqrt{(X^T X)_{ii}^{-1}}} \sim t_{91}.$$

where

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

and

$$s = \sqrt{\frac{1}{91} \|\mathbf{y} - X\hat{\beta})\|^2} = 4.11$$

Example: Body Fat

- Estimates and 95% credible intervals for $\beta_i's$:

Variable	$\hat{\beta}_i$	95% credible interval
Age	0.956	(-0.186, 2.099)
Weight	-2.458	(-7.397, 2.480)
Height	0.097	(-1.328, 1.523)
Neck	0.002	(-1.727, 1.732)
Chest	-1.181	(-3.889, 1.526)
Abdomen	10.597	(7.639, 13.554)
Ankle	0.304	(-1.137, 1.745)
Biceps	0.454	(-0.935, 1.844)
Wrist	-2.201	(-3.807, -0.596)

http://www.ericfrazerlock.com/More_on_Linear_
Models_Rcode1.r

Example: Body Fat

- Recall $p(\sigma^2 \mid \mathbf{y}) = IG\left(\frac{91}{2}, \frac{91\,s^2}{2}\right)$:
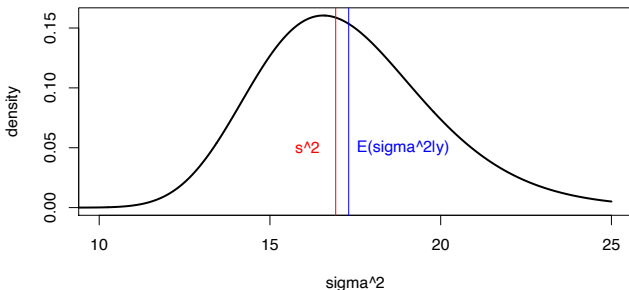
http://www.ericfrazerlock.com/More_on_Linear_Models_Rcode1.r

Variance estimate, uninformative priors

▶ Note for the uninformative prior $p(\mu, \sigma^2) = \frac{1}{\sigma^2}$,

$$E(\sigma^2 \mid \mathbf{y}) = \frac{s^2(n-p)}{n-p-2}$$

▶ However, the expected precision is

$$E(1/\sigma^2 \mid \mathbf{y}) = \frac{1}{s^2}$$

▶ s^2 still commonly used as point estimate for error variance.

Residuals

▶ Recall: defined Bayesian residual as

$$r_i' = y_i - E(Y_i \mid \mathbf{y}_{(i)})$$

where $\mathbf{y}_{(i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$

▶ For this context, the Bayesian residual is

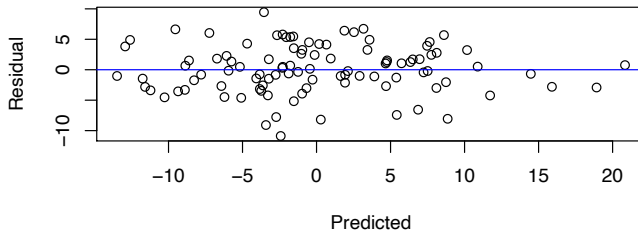$$r_i' = y_i - \mathbf{x}_i \hat{\beta}_{(i)}$$

where $\hat{\beta}_{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T \mathbf{y}_{(i)}$.

▶ The standard (non-Bayesian) definition of residual is
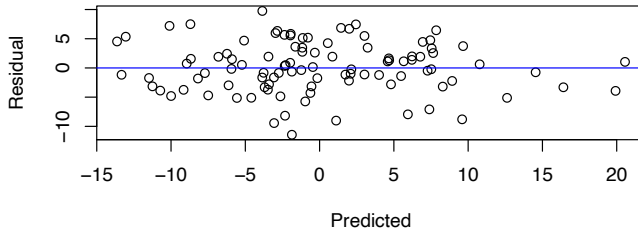
$$r_i = y_i - \mathbf{x}_i \hat{\beta}$$
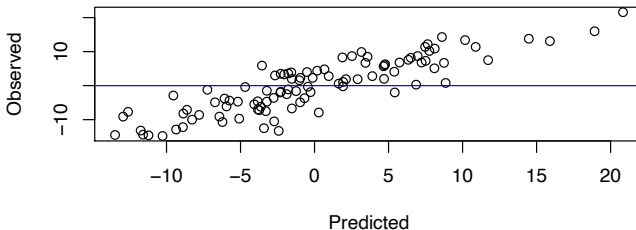
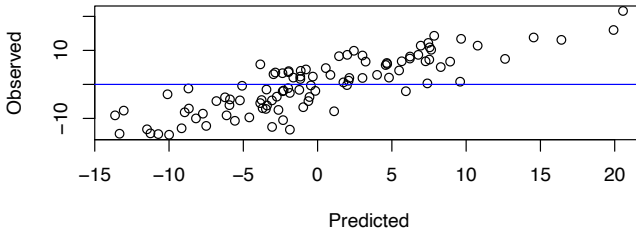# Example: Body Fat

Standard residuals

Bayesian residuals

Example: Body Fat

Predicted vs observed (standard)

Predicted vs observed (Bayesian)

Normal-inverse-gamma prior

- Consider independent normal priors for the $\beta_i's$:

$$\beta \mid \sigma^2 \sim \text{Normal}(0, \sigma^2 T)$$

  where $T_{ij} = \tau_i^2$ if $i = j$, 0 otherwise.

- And an inverse-gamma prior for σ^2:

$$\sigma^2 \sim IG(a, b).$$

- The full prior is

$$p(\beta, \sigma^2) = IG(\sigma^2 \mid a, b) \prod_{i=1}^{p} \text{Normal}(\beta_i \mid 0, \sigma^2 \tau_i^2)$$

- The posterior for β, given σ^2, is

$$p(\beta \mid \mathbf{y}, \sigma^2) = \text{Normal}\left(\tilde{\beta}, \sigma^2 V_\beta\right)$$

where $\tilde{\beta} = (X^T X + T^{-1})^{-1}(X^T \mathbf{y})$

and $V_\beta = (X^T X + T^{-1})^{-1}$

Normal-inverse-gamma prior

- The estimate $\tilde{\beta}$ solves a penalized least squares criterion:

$$\tilde{\beta} = \underset{\beta}{\mathrm{argmin}} ||\mathbf{y} - XB||^2 + \sum_{i=1}^{p} \beta_i^2 / \tau_i^2$$

- Shrinks unbiased estimate $\hat{\beta}$ toward **0**.

▶ The marginal posterior for σ^2 is

$$p(\sigma^2 \mid \mathbf{y}) = IG\left(a_n, b_n\right)$$

where $a_n = a + \frac{n}{2}$ and $b_n = b + \frac{1}{2}[\mathbf{y}^T\mathbf{y} - \tilde{\beta}^T V_\beta^{-1} \tilde{\beta}]$

▶ The marginal posterior for β is a multivariate t-distribution

$$\frac{\beta_i - \tilde{\beta}_i}{\sqrt{\frac{b_n}{a_n}(V_\beta)_{ii}}} \sim t_{2a+n}.$$

▶ For a new predictor vector $\mathbf{x}_{n+1}$, the posterior predictive for y_{n+1} given σ^2 is

$$y_{n+1} \mid \sigma^2, \mathbf{y} \sim \text{Normal}(\mathbf{x}_{n+1}\tilde{\beta}, \sigma^2(1 + \mathbf{x}_{n+1} V_\beta \mathbf{x}_{n+1}^T))$$

▶ The full posterior predictive distribution is a non-central t:

$$\frac{y_{n+1} - \mathbf{x}_{n+1}\tilde{\beta}}{\sqrt{\frac{b_n}{a_n} \left(1 + \mathbf{x}_{n+1} V_\beta \mathbf{x}_{n+1}^T\right)}} \sim t_{2a+n}.$$

Extensions

▶ There are many other versions of the Bayesian linear model.

▶ E.g.: Could use non-trivial mean and covariance for β:

$$\beta \sim \text{Normal}(\mu_\beta, T)$$

▶ E.g.: Could relax iid assumption for $y_i's$, model general covariance:

$$\mathbf{y} \sim \text{Normal}(X\beta, \Sigma)$$

requires a prior for Σ.

▶ For more details and derivations see
http://www.ericfrazerlock.com/LM_GoryDetails.pdf
and Carlin & Louis 4.1.1