Bayesian Statistics

Normal Linear Model

Nan Lin
Department of Mathematics
Washington University in St. Louis

Multiple linear regression

- Response: y
- Explanatory variables: $\boldsymbol{x} = (x_1, \dots, x_k)$
- Goal: find $y = f\,(\boldsymbol{x})$
- Data: $(\boldsymbol{x}_i, y_i), i = 1, \dots, n$
- Model: $y_i = f(\boldsymbol{x}_i) + \epsilon_i,$
  - Linear regression: $f(x_i) = x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + x_{ki}\beta_k$
- Typical assumptions for normal linear regression
  - $\epsilon_i \sim N(0, \sigma^2)$ i.i.d.
- $E(y|\boldsymbol{x}) = f(x)$
  - For linear regression, using matrix notations, we have $Y|X \sim N_n(X\boldsymbol{\beta}, \sigma^2 I)$

Frequentist inference

- Ordinary least square
  - $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T y$
  - $\hat{\sigma}^2 = s^2 = \frac{1}{n-k}\left(y - X\hat{\beta}\right)^T\left(y - X\hat{\beta}\right)$

Bayesian inference

- Noninformative Prior

$$f(\boldsymbol{\beta}, \sigma^2 | X) \propto \sigma^{-2}$$

- Posterior
  - $\boldsymbol{\beta} | \sigma^2, y \sim N(\widehat{\boldsymbol{\beta}}, V_\beta \sigma^2)$
    - $\widehat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T y$
    - $V_\beta = (X^T X)^{-1}$
  - $\frac{(n-k)s^2}{\sigma^2} | y \sim \chi^2_{n-k}$, i.e. $\sigma^2 | y \sim Inv - \chi^2\ (n - k, s^2)$
    - $s^2 = \frac{1}{n-k}\left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right)$
  - Marginal posterior of $\boldsymbol{\beta} | y$ is the multivariate t-distribution with $n - k$ degrees of freedom

Sampling from the posterior

- Calculate $\widehat{\boldsymbol{\beta}}$ and V_β from standard linear regression software
  - The following is a relatively efficient algorithm
    - Compute the QR factorization, $X = QR$, where Q is an $n \times k$ matrix of orthonormal columns and R is a $k \times k$ upper triangular matrix
    - Compute R^{-1}, which is an easy task since R is upper triangular. Then we obtain $V_\beta = (X^T X)^{-1} = R^{-1}(R^{-1})^T$.
    - Compute $\hat{\beta}$ by solving the linear system $R\hat{\beta} = Q^T y$.
- Simulate from $\sigma^2 | y$
- Simulate from $\boldsymbol{\beta} | \sigma^2, y$

Predictive distribution

- Given a new design matrix $\tilde{X}$, predict $\tilde{y}$
- Simulate from the predictive distribution
  - Draw from $\beta, \sigma^2|y$
  - Simulate from $N(\tilde{X}\beta, \sigma^2 I)$
- Analytical results
  - $E(\tilde{y}|\sigma^2, y) = \tilde{X}\hat{\beta}$
  - $Var(\tilde{y}|\sigma^2, y) = \left(I + \tilde{X}V_\beta\tilde{X}^T\right)\sigma^2$
  - $\tilde{y}|y$ is multivariate t with center $\tilde{X}\hat{\beta}$, squared scale matrix $\left(I + \tilde{X}V_\beta\tilde{X}^T\right)s^2$, and $n-k$ degrees of freedom