LECTURE FOUR

Bayesian Inference for Simple Linear Regression

$(x_1, y_1) \ldots (x_n, y_n)$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$$

likelihood: $\displaystyle\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right]$

$$\propto \sigma^{-n} \exp\Big[-\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2}_{S(\beta_0, \beta_1)}\Big]$$

$$= \sigma^{-n} \exp\left[\frac{-S(\beta_0, \beta_1)}{2\sigma^2}\right]$$

prior: $\beta_0, \beta_1, \log\sigma \overset{iid}{\sim} \text{Unif}(-C, C)$

(think of $C = \infty$)

$$f_{\beta_0, \beta_1, \sigma}(\beta_0, \beta_1, \sigma) = \frac{I\{-C < \beta_0 < C\}}{2C} \, \frac{I\{-C < \beta_1 < C\}}{2C} \, \frac{I\{e^{-C} < \sigma < e^{C}\}}{2C\sigma}$$

$\log\sigma \sim \text{Unif}(-C, C) \Rightarrow \sigma \sim$

$$\propto \frac{I\{-C < \beta_0, \beta_1, \log\sigma < C\}}{\sigma}$$

$$\underset{(\beta_0, \beta_1, \sigma)}{\text{posterior}} \propto \underset{(\beta_0, \beta_1, \sigma)}{\text{prior}} \times \text{likelihood}(\beta_0, \beta_1, \sigma)$$

$$\propto \frac{I\{-C < \beta_0, \beta_1, \log\sigma < C\}}{\sigma} \left(\frac{1}{\sigma}\right)^n \exp\left(-\frac{S(\beta_0, \beta_1)}{2\sigma^2}\right)$$

$$= I\{-c<\beta_0,\beta_1,\log\sigma<c\}\left(\frac{1}{\sigma}\right)^{n+1}\exp\left(-\frac{S(\beta_0,\beta_1)}{2\sigma^2}\right)$$

Integrate this to get marginal densities of the variables separately.

β_0, β_1 : main parameters $\qquad \sigma$: nuisance parameters

$$f_{\beta_0,\beta_1|data}(\beta_0,\beta_1) = \int_{-\infty}^{\infty} f_{\beta_0,\beta_1,\sigma|data}(\beta_0,\beta_1,\sigma)\,d\sigma \rightarrow$$ LAW OF TOTAL PROBABILITY

$$\propto I\{-c<\beta_0,\beta_1<c\}\int_{e^{-c}}^{e^{c}}\left(\frac{1}{\sigma}\right)^{n+1}\exp\left(-\frac{S(\beta_0,\beta_1)}{2\sigma^2}\right)d\sigma$$

$$\approx \underbrace{I\{-c<\beta_0,\beta_1<c\}\int_{0}^{\infty}\left(\frac{1}{\sigma}\right)^{n+1}\exp\left(-\frac{S(\beta_0,\beta_1)}{2\sigma^2}\right)d\sigma}$$

$$s = \frac{\sigma}{\sqrt{S(\beta_0,\beta_1)}}$$

$$= I\{-c<\beta_0,\beta_1<c\}\int_0^\infty\left(\frac{1}{s}\right)^{n+1}\left(S(\beta_0,\beta_1)\right)^{-\frac{n+1}{2}}\exp\left[-\frac{1}{2s^2}\right]ds\,\sqrt{S(\beta_0,\beta_1)}$$

$$= I\{-c<\beta_0,\beta_1<c\}\left(S(\beta_0,\beta_1)\right)^{-\frac{n}{2}}\int_0^\infty\left(\frac{1}{s}\right)^{n+1}\exp\left(-\frac{1}{2s^2}\right)ds$$

$$\propto I\{-c<\beta_0,\beta_1<c\}\left[\frac{1}{S(\beta_0,\beta_1)}\right]^{\frac{n}{2}}$$

$$f_{\beta_0,\beta_1 \mid \text{data}}(\beta_0,\beta_1) \propto I\{-c<\beta_0,\beta_1<c\}\left[\frac{1}{S(\beta_0,\beta_1)}\right]^{\frac{n}{2}}$$

Posterior mode: Least squares estimator: $\hat{\beta}_0, \hat{\beta}_1$

$$\propto I\{-c<\beta_0,\beta_1<c\}\left[\frac{S(\hat{\beta}_0,\hat{\beta}_1)}{S(\beta_0,\beta_1)}\right]^{\frac{n}{2}}$$

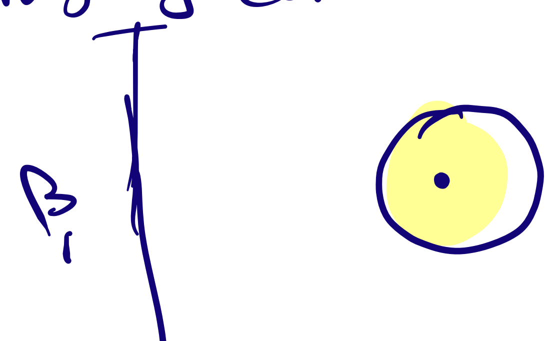
Claim: Posterior is concentrated around the least squares estimator $\hat{\beta}_0, \hat{\beta}_1$.

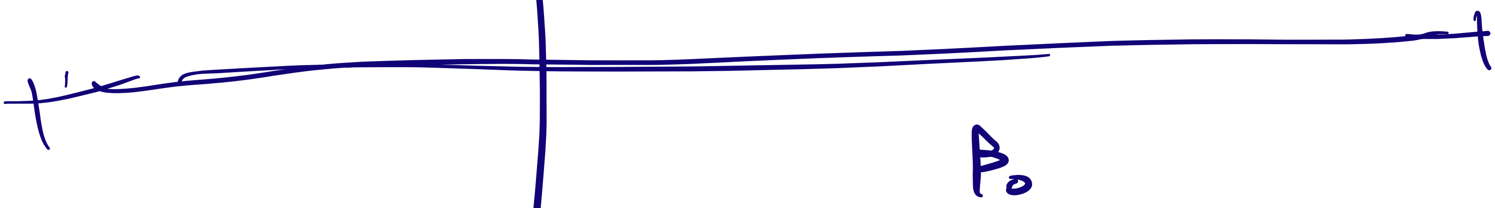
(a) If $S(\beta_0,\beta_1) = (1.1)\, S(\hat{\beta}_0,\hat{\beta}_1)$, then posterior at (β_0,β_1): $\left(\frac{1}{1.1}\right)^{\frac{n}{2}}$

(b) If $S(\beta_0,\beta_1) = (1.01)\, S(\hat{\beta}_0,\hat{\beta}_1)$

posterior: $\left(\frac{1}{1.01}\right)^{\frac{n}{2}}$

Conclude: Bayesian inference with $\beta_0,\beta_1,\log\sigma \sim \text{Unif}(-C,C)$ is also based on the least squares estimators $\hat{\beta}_0, \hat{\beta}_1$. Posterior typically is highly concentrated: $\left[\frac{S(\hat{\beta}_0,\hat{\beta}_1)}{S(\beta_0,\beta_1)}\right]^{\frac{n}{2}}$





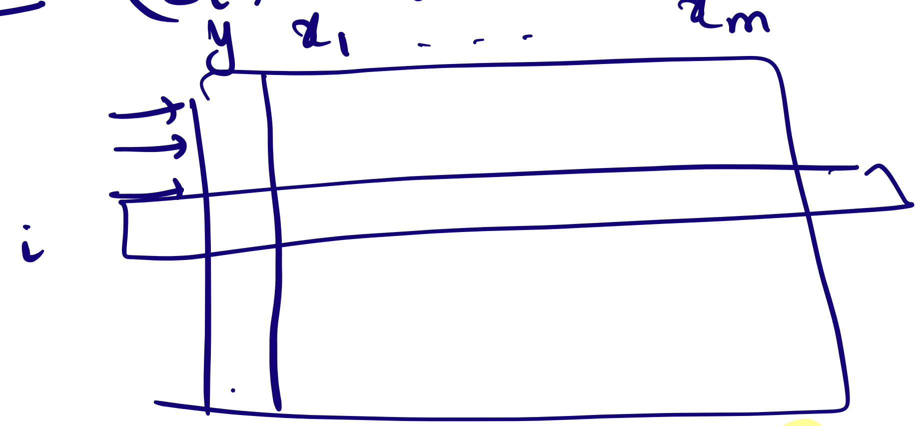
Posterior: $\propto \left[\dfrac{S(\hat{\beta}_0, \hat{\beta}_1)}{S(\beta_0, \beta_1)}\right]^{\frac{n}{2}} \Bigg\} \boxed{t\text{-density}}$

Multiple Linear Regression

$\underbrace{y}_{\text{response}} \qquad \underbrace{x_1, x_2, \ldots, x_m}_{\text{covariates}}$

Data: $(y_i, \; x_{i1} \; x_{i2} \; \ldots \; x_{im}), \; i = 1 \ldots n$

Tabular Data



Model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im} + \varepsilon_i$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

(a) Functions of Time as Covariates:

$x_1 = \text{time} \rightarrow x_{i1} = i$

$x_2 = (\text{time})^2 \rightarrow x_{i2} = i^2$

$x_3 = \cos(2\pi \times \text{time}) \rightarrow x_{i3} = \cos(2\pi \frac{i}{12})$

ⓑ AutoRegression:

$x_1 = y$ at the previous time $\rightarrow x_{i1} = y_{i-1}$

$x_2 = y$ two time points before $\rightarrow x_{i2} = y_{i-2}$

. . . .

$$S(\beta_0, \dots, \beta_m) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_m x_{im})^2$$

$\rightarrow$ minimize to get $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$ (least squares estimator)

$\beta_0, \beta_1, \dots, \beta_m, \log\sigma \overset{iid}{\sim} \text{Unif}(-C, C)$

posterior of $\beta_0, \dots, \beta_m \propto$ $\boxed{p = m+1}$ $\rightarrow$ $\left[\dfrac{S(\hat{\beta}_0, \dots, \hat{\beta}_m)}{S(\beta_0, \dots, \beta_m)}\right]^{\frac{n}{2}}$

Fact: This posterior is a multivariate t-density

Multivariate t-Density

$$\propto \left[\frac{1}{1 + \frac{1}{\nu}(x-\mu)^T \Sigma^{-1} (x-\mu)}\right]^{\frac{\nu+p}{2}}$$

$x = (x_1, \dots, x_p)$

p : dimension

ν : degrees of freedom

μ : location , Σ : scale or Covariance

$$\left[\left(\frac{S(\hat{\beta}_0, \ldots, \hat{\beta}_m)}{S(\beta_0, \ldots, \beta_m)}\right)^{\frac{n}{2}}\right.$$

Matrix Notation for Regression

Data: y_i, $x_{i1} \ldots x_{im}$ $\quad i = 1 \ldots, n$

$$\underset{\substack{\uparrow \\ \text{vector of response values}}}{y} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{n \times 1} \qquad \underset{n \times (m+1)}{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & & x_{nm} \end{bmatrix}$$

Values of covariate x_1

sm.OLS($\underline{y}$, $\underline{X}$).fit()

$$\underset{(m+1) \times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \qquad \underset{(m+1) \times 1}{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_m \end{pmatrix}$$

$$S(\beta_0, \ldots, \beta_m) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_m x_{im})^2$$

$$S(\beta) = \| \underbrace{y - \underset{n \times (m+1)}{X} \underset{(m+1) \times 1}{\beta}}_{n \times 1} \|^2$$

$$S(\beta) = \|y - X\beta\|^2$$

$$S(\beta) = (y - X\beta)^T (y - X\beta)$$
$$= y^T y - 2\beta^T X^T y + \beta^T X^T X \beta$$

Minimize to get a formula for least squares:

$$\nabla S(\beta) = \begin{pmatrix} \frac{\partial}{\partial \beta_0} S(\beta) \\ \frac{\partial}{\partial \beta_1} S(\beta) \\ \vdots \\ \frac{\partial}{\partial \beta_m} S(\beta) \end{pmatrix} = -2\nabla(\beta^T X^T y) + \nabla(\underbrace{\beta^T X^T X \beta})$$
$$= -2X^T y + 2X^T X\beta$$
$$= 0$$

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

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Facts about $S(\beta)$

① $\hat{\beta} = (X^T X)^{-1} X^T y$

② $S(\beta) = \|y - X\beta\|^2$
$$= \|y - X\hat{\beta} + X\hat{\beta} - X\beta\|^2$$
$$= (y - X\hat{\beta} + X\hat{\beta} - X\beta)^T (y - X\hat{\beta} + X\hat{\beta} - X\beta)$$
$$= \underbrace{(y - X\hat{\beta})^T (y - X\hat{\beta})}_{S(\hat{\beta})} + (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) + \underbrace{\text{Cross product term}}_{\text{Exercise: } = 0}$$

$$\boxed{S(\beta) = S(\hat{\beta}) + (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)}$$

Posterior β : $\left[\dfrac{S(\hat\beta)}{S(\beta)}\right]^{\frac{n}{2}}$

$$= \left\{\frac{S(\hat\beta)}{S(\hat\beta) + \underbrace{(\beta-\hat\beta)^T X^T X (\beta-\hat\beta)}}\right\}^{\frac{n}{2}}$$

t-density : $\left\{\dfrac{1}{1 + \frac{1}{\nu}(x-\mu)^T \Sigma^{-1}(x-\mu)}\right\}^{\frac{\nu+p}{2}}$

$$\left\{\frac{1}{1 + (\beta-\hat\beta)^T \frac{X^T X}{S(\hat\beta)} (\beta-\hat\beta)}\right\}^{\frac{n}{2}}$$

$p = m+1$, $\nu + p = n \Rightarrow \nu = n - p = n - m - 1$

$\mu = \hat\beta$: $\frac{1}{\nu}\Sigma^{-1} = \dfrac{X^T X}{S(\hat\beta)}$

$$\Rightarrow \Sigma = \frac{S(\hat\beta)}{\nu}(X^T X)^{-1}$$

$$\boxed{\Sigma = \frac{S(\hat\beta)}{n-m-1}(X^T X)^{-1}}$$

Posterior of $\beta_0, \dots, \beta_m$: $t_{m+1}\left(\underbrace{\hat{\beta}}_{\text{least squares}}, \frac{S(\hat{\beta})}{n-m-1}(X^T X)^{-1}, n-m-1\right)$

Simulate $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(N)}$

$S(\hat{\beta})$: Residual Sum of Squares

$$\sum \left(\underbrace{y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_m x_{im}}\right)^2$$

$y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_m x_{im}$: i^{th} residual