Normal linear model

.

Standard normal distribution, pdf

$$f(w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}w^2\right)$$

$$W \sim N(0,1)$$

Normal distribution with mean μ and variance σ^2

$$N\left(\mu, \sigma^2\right)$$

$$\mu + \sigma W$$

Random sampling

$$(Y_i, X_i) \sim F, \text{ i.i.d}$$

Terminology: independent and identically distributed – i.i.d.

Conditional distribution of Y_i given X_i

$$Y_i | X_i = x_i \sim N\left(x_i'\beta, \sigma^2\right)$$

$$y_i = x_i'\beta + u_i$$

Note: Polynomial regressions fit this framework as well:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

$$y = \beta_0 + \beta_1 x_{[1]} + \beta_2 x_{[2]} + u$$

Discussion: The purpose of confidence intervals vs. loss (MSE, cross-entropy). Generalization across datasets, experiments, or settings.

Here the prime means transpose

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_K \end{pmatrix}$$

$$x_i = \begin{pmatrix} x_{i,1} \\ x_{i,2} \\ \\ x_{i,K} \end{pmatrix}$$

Prediction error

$$U_i = Y_i - x_i'\beta$$

$$V_i = \frac{U_i}{\sigma}$$

$$V_i | X_i = x_i \sim N(0, 1)$$

$$Y_i = x_i'\beta + \sigma V_i$$

.

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \\ \\ Y_n \end{pmatrix}, \quad V = \begin{pmatrix} V_1 \\ V_2 \\ \\ \\ V_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1' \\ x_2' \\ \\ \\ x_n' \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} \\ & \\ & \\ & \\ x_{n,1} & \end{pmatrix}$$

$$Y = x\beta + \sigma V$$

Canonical form

$$Y^* = \begin{pmatrix} s \\ 0 \end{pmatrix} \beta + \sigma V^*$$

$$Y^* = \begin{pmatrix} Y^*_{(1)} \\ Y^*_{(2)} \end{pmatrix}$$

$$Y^* = \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix} \beta + \sigma V^*$$

$$V^* = \begin{pmatrix} V^*_{(1)} \\ V^*_{(2)} \end{pmatrix}$$

$$V^* \sim N(0, I_n)$$

The components of V^* are i.i.d. standard normal:

$$V_i^* \sim N(0, 1)$$

$V^*_{(1)}$ is $K \times 1$

$V^*_{(2)}$ is $(n-K) \times 1$

$$\begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ 0 & s_{22} & s_{23} \\ 0 & 0 & s_{33} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$s = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ 0 & s_{22} & s_{23} \\ 0 & 0 & s_{33} \end{pmatrix}$$

$$0_{(n-K)\times K} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Upper triangular matrix s of dimension $K \times K$

Orthogonal matrix q of dimension $n \times n$

$$Y^* = q'Y$$

QR decomposition:

If x is $n \times K$, then there is an orthogonal matrix q of dimension $n \times n$ and an upper triangular matrix r of dimension $n \times K$ such that

$$x = qr$$

For $n > K$

$$r = \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix}$$

For $K = 3$, for example, the matrix s is:

$$s = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ 0 & s_{22} & s_{23} \\ 0 & 0 & s_{33} \end{pmatrix}$$

Note: An orthogonal matrix q satisfies $qq' = 1$ and $q'q = 1$.

We will define

$$Y^* = q'Y$$

$$V^* = q'V$$

We can now derive the canonical form of the normal linear model:

$$Y = x\beta + \sigma V$$

$$q'Y = q'x\beta + \sigma q'V$$

$$Y^* = q'x\beta + \sigma V^*$$

Distribution of V

$$V \sim N(0, I_n),$$

where I_n is the $n \times n$ identity matrix. The distribution of V^* is the same

$$V^* \sim N(0, I_n)$$

$$V_i^* \sim N(0, 1)$$

By the QR decomposition theorem:

$$x = q \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix}$$

$$q'x = q'q \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix}$$

$$q'x = I_n \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix}$$

$$q'x = \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix}$$

Utilizing this fact gives us

$$Y^* = q'x\beta + \sigma V^*$$

$$Y^* = \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix} \beta + \sigma V^*$$

$$Y^*_{(1)} = s\beta + \sigma V^*_{(1)}$$

$$Y^*_{(2)} = \sigma V^*_{(2)}$$

Estimator for β

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \\ b_K \end{pmatrix}$$

$$b = \arg\min_{\tilde{b}} ||Y - x\tilde{b}||^2$$

$$MSE = \frac{1}{N}||Y - xb||^2$$

$$||Y - xb||^2 = ||q'\left(Y - xb\right)||^2 = ||q'Y - q'xb||^2$$

$$= ||Y^* - \begin{pmatrix} s \\ 0_{(n-K)\times K} \end{pmatrix} b||^2$$

$$= ||Y^*_{(1)} - sb||^2 + ||Y^*_{(2)} - 0_{(n-K)\times K}b||^2$$

$$= ||Y^*_{(1)} - sb||^2 + ||Y^*_{(2)}||^2$$

If we choose

$$b = s^{-1} Y^*_{(1)}$$

then $||Y^*_{(1)} - sb||^2 = 0$ and

$$||Y - xb||^2 = ||Y^*_{(2)}||^2$$

Sum of squared residuals

$$SSR = ||Y - xb||^2 = ||Y^*_{(2)}||^2 = ||\sigma V^*_{(2)}||^2 = \sigma^2 ||V^*_{(2)}||^2 = \sigma^2 \sum_{i=K+1}^{n} V^*_i$$

The regression coefficient vector b is independent of the SSR.

Student t-distribution, defined using a normal distribution and a chi-squared distribution.