Lecture Notes 6: Non-Linear Regression

In this lecture note we will provide some more results related to linear regression model and they will be used later. Let us recall that a linear regression model in a matrix form can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1}$$

Here $\mathbf{Y}$ is a $n \times 1$ observation vector, $\mathbf{X}$ is a $n \times p$ design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ is a $n \times 1$ random vector. It is assumed for the time that

$$E(\epsilon_i) = 0 \quad \text{and} \quad \operatorname{Cov}(\epsilon_i, \epsilon_j) = \begin{cases} \sigma^2 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j. \end{cases} \tag{2}$$

We have seen that under the above assumption the LSEs of $\boldsymbol{\beta}$ can be obtained as follows,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}.$$

under the assumption that the design matrix $\mathbf{X}$ is of full rank. Now let us look at some of the properties of $\widehat{\boldsymbol{\beta}}$. First it is observed that under the error assumptions (2), the LSE of $\boldsymbol{\beta}$ is an unbiased estimator of $\boldsymbol{b}eta$, and it can be easily see as follows:

$$E(\widehat{\boldsymbol{\beta}}) = E((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top E\mathbf{Y} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

Now let us try to examine whether it will be a consistent estimator or not. Note that $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$ means

$$\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta},$$

i.e. $\widehat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ in probability. Alternatively, we can say that for any $\epsilon > 0$,

$$P(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}| > \epsilon) \to 0, \quad \text{as } n \ \to \ \infty$$

Now to check that let us look at the dispersion matrix of $\widehat{\boldsymbol{\beta}}$, and that is

$$\begin{aligned} D(\widehat{\boldsymbol{\beta}}) &= D(((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y})) = ((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top D(\mathbf{Y})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top D(\boldsymbol{\epsilon})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}. \end{aligned}$$

Suppose we make the following assumption of the design matrix $\mathbf{X}$ that

$$\frac{1}{n}(\mathbf{X}^\top\mathbf{X}) \;\rightarrow\; \mathbf{A} > 0, \tag{3}$$

here $\mathbf{A}$ is a positive definite matrix. Then,

$$D(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} = \frac{\sigma^2}{n}\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}.$$

Therefore,

$$\lim_{n\to\infty} D(\widehat{\boldsymbol{\beta}}) = \sigma^2\mathbf{A}^{-1}\lim_{n\to\infty}\frac{1}{n} = \mathbf{0}.$$

Hence, in this case

$$\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}.$$

Therefore, the condition (3) can be treated as a sufficient condition of consistency. But it cannot be taken as a necessary condition. Let us look at the following two examples, it should be clear.

Example 1: Consider the following linear regression model:

$$y(i) = A\cos(i) + B\sin(i) + \epsilon(i); \quad i = 1, \ldots, n. \tag{4}$$

Here the error $\epsilon(i)$ satisfies the error assumptions (2). Now in this case the design matrix $\mathbf{X}$ becomes

$$\mathbf{X} = \begin{bmatrix} \cos(1) & \sin(1) \\ \cos(2) & \sin(2) \\ \vdots & \vdots \\ \cos(n) & \sin(n) \end{bmatrix}.$$

Therefore,

$$\frac{1}{n}\mathbf{X}^\top\mathbf{X} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^n \cos^2(i) & \sum_{i=1}^n \cos(i)\sin(i) \\ \sum_{i=1}^n \cos(i)\sin(i) & \sum_{i=1}^n \sin^2(i) \end{bmatrix}$$

Now based on the following trigonometric results:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \cos^2(i) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^2 \sin^2(i) = \frac{1}{2}, \quad \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \cos(i)\sin(i) = 0,$$

it follows that

$$\frac{1}{n}\mathbf{X}^\top\mathbf{X} = \frac{1}{2}\,\mathbf{I} > 0.$$

Therefore, the LSEs of A and B of the model (4) are unbiased and consistent estimators of A and B, respectively.

Now consider the second example:

Example 2

Consider the following linear regression model:

$$y(i) = A + Bi + \epsilon(i); \quad i = 1, \ldots, n. \tag{5}$$

Here the error $\epsilon(i)$ satisfies the error assumptions (2). Now in this case the design matrix $\mathbf{X}$ becomes

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}.$$

Therefore,

$$\frac{1}{n}\mathbf{X}^\top\mathbf{X} = \frac{1}{n}\begin{bmatrix} n & \frac{n(n+1)}{2} \\ \frac{n(n+1)}{2} & \frac{n(n+1)(2n+1)}{6} \end{bmatrix}$$

Now it is clear that $\frac{1}{n}\mathbf{X}^\top\mathbf{X}$ does not converge to a positive definite matrix. Hence, it does not satisfy the sufficient condition. Verify whether LSEs are consistent or not?

Now let us discuss about the distribution of $\widehat{\boldsymbol{\beta}}$. It will help us in deriving the confidence set of $\boldsymbol{\beta}$ as well as answering the testing of hypothesis of problems also. Let us make a simplified assumptions of $\epsilon(i)$'s. Other than the conditions (2), it is further assumed that $\epsilon(i)$'s are i.i.d. normal random variables. Then, it is immediate $\boldsymbol{\beta}$ is a p-variate normal distribution with the mean vector $\boldsymbol{\beta}$ and the dispersion matrix $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$. Hence,

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N_P(\mathbf{0}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}).$$

Therefore,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N_p\left(\mathbf{0}, \sigma^2 \left(\frac{1}{n}\mathbf{X}^\top \mathbf{X}\right)^{-1}\right).$$

Now if the errors are not normally distributed, then if the errors follow (2), the condition (3) is satisfied, then

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N_p(\mathbf{0}, \sigma^2 \mathbf{A}^{-1}),$$

i.e. $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution to a p-variate normal distribution with the mean vector $\mathbf{0}$, and the dispersion matrix $\sigma^2\mathbf{A}^{-1}$.

We have already provided the confidence set for $\boldsymbol{\beta}$ when σ^2 is known and unknown also. Now construct a $100(1-\alpha)\%$ confidence set for A and B in case of Example 1, when (i) $\epsilon(i)$'s are i.i.d. N(0,1) and σ^2 is known, (ii) $\epsilon(i)$'s are i.i.d. N(0,1) and σ^2 is unknown, (iii) $\epsilon(i)$'s are i.i.d. random variables with mean 0, variance 1, and σ^2 is known, (iv) $\epsilon(i)$'s are i.i.d. random variables with mean 0, variance 1, and σ^2 is unknown.

Answer all the above questions in case of Example 2.