Tutorial 2

YX

Feb. 9, 2022

1. Proof to Section 2.6 theorem:

   Suppose $X_1, \dots, X_n$ $i.i.d. \sim N(\mu, \sigma^2)$,$\bar{X} = \sum\limits_{i=1}^{n} X_i/n,\ S^2 = \sum\limits_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$, we have:

   1.$\bar{X} \sim N(\mu, \sigma^2/n)$

   2.$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

   3.$\bar{X} \perp\!\!\!\perp S^2$

   Proof:

   1. Proved in other courses.

   2. Let

   $$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

   we select elements for a_{ij} such that A is an orthogonal matrix (See Gram-Schmidt Process). We now have $\mathbf{AA^T} = \mathbf{I}$

   For the orthogonal transformation $Y = AX$,then

   $$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \sqrt{n}\bar{X}$$

   By properties of orthogonal transformation, we know that:

   $$Y_1^2 + \cdots + Y_n^2 = X_1^2 + \cdots + X_n^2$$

   Thus,

   $$\begin{aligned} (n-1)S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^{n} Y_i^2 - Y_1^2 \\ &= \sum_{i=2}^{n} Y_i^2 \end{aligned}$$

Now we prove $\{Y_i\}_{i=1}^n$ are independent of each other.

$(Y_1, Y_2, \dots, Y_n)^T = A(X_1, X_2, \dots, X_n)^T$. Hence, $(Y_1, Y_2, \dots, Y_n)$ are jointly distributed as a multivariate normal. We only need to prove that they are uncorrelated.

$$\begin{aligned} Cov(\mathbf{Y}) &= Cov(\mathbf{AX}) \\ &= \mathbf{A}Cov(X)\mathbf{A^T} \\ &= \mathbf{A}\sigma^2\mathbf{I}\mathbf{A^T} \\ &= \sigma^2\mathbf{I} \end{aligned}$$

Therefore, $Y_1, Y_2, \dots, Y_n$ are independent of each other and $Y_i \sim N(\mu_i, \sigma^2)$. Note that $\mathbf{A}$ is orthogonal:

$$\mu_i = \mu \sum_{k=1}^n a_{ik} = \sqrt{n}\mu \sum_{k=1}^n \frac{1}{\sqrt{n}} a_{ik} = 0$$

i.e. $Y_2, \dots, Y_n$ are i.i.d. from $N(0, \sigma^2)$. Hence, $Y_2/\sigma, \dots, Y_n/\sigma$ are i.i.d. from $N(0, 1)$.

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^n (Y_i/\sigma)^2 \sim \chi^2_{n-1}$$

Now we proved (2).

3.

From 2. S^2 is only related to $Y_2, \dots, Y_n$, $\bar{X}$ only depends on Y_1, and $Y_1, Y_2, \dots, Y_n$ are independent of each other. Then $\bar{X}$ is independent of S^2.

For the rest of the tutorial, please refer to the handwritten notes below

1. partial derivative $f(x,y) \leftarrow f_y(x)$.

$\frac{\partial}{\partial x} f(x,y) = \frac{d}{dx} f_y(x)$. e.g. $\underset{\substack{\uparrow \\ \text{constant}}}{\sin(xy)}$. $f_y(x) = \sin(\underset{\substack{\uparrow \\ y_0}}{y_0} \cdot x)$.

$\frac{\partial}{\partial x} f(x,y) = \frac{d}{dx} \sin(y_0 x)$.

$f(x,y) = x^2 y$
$$\begin{cases} \frac{\partial}{\partial x} f(x,y) = y \cdot 2x = 2xy \\ \frac{\partial}{\partial y} f(x,y) = x^2. \end{cases}$$

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) = \arg\max l(\underbrace{\beta_0, \beta_1, \sigma^2}_{\vec{\theta}}).$$

$$l(\beta_0, \beta_1, \sigma^2) = \ln(L(\beta_0, \beta_1, \sigma^2))$$
$$= \underbrace{-\frac{n}{2}\ln(2\pi)} - \underbrace{\frac{n}{2}\ln\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

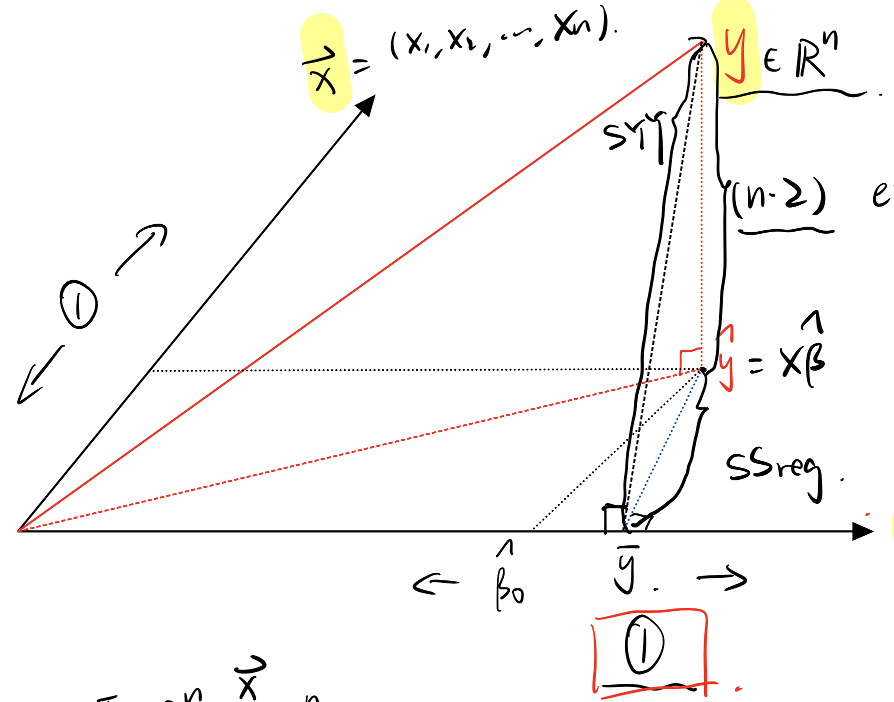
$$\begin{cases} \frac{\partial}{\partial \beta_0} l = 0 \leftarrow \\ \frac{\partial}{\partial \beta_1} l = 0 \quad ② \\ \frac{\partial}{\partial \sigma^2} l = 0 \quad ③ \end{cases}$$

$$\frac{\partial}{\partial \beta_0} l = -\frac{1}{\cancel{2}\sigma^2}\sum_{i=1}^{n}\cancel{2}(y_i - \beta_0 - \beta_1 x_i)\cdot(-1))$$
$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0.$$

2. Degree of Freedom. numbers of variables

rank. a linear space S $\text{rank}(S) = \dim(S) = k$.

$\{v_1, v_2, \cdots, v_k\} \subseteq S$. linearly independent. $\forall v \in S.\ \{\lambda_i\}_{i=1}^{k} \in \mathbb{R}.\ v = \sum_{i=1}^{k} \lambda_i \vec{v}_i$



n dof.

(n-2) error term.

dim-2 space.

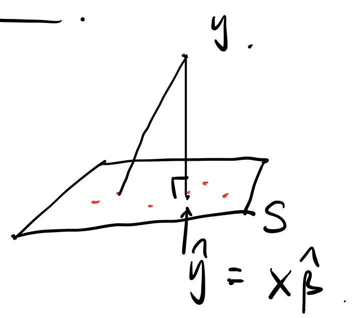
$\text{Span}\{J_n, \vec{x}\}$
$= \text{span}(X)$.

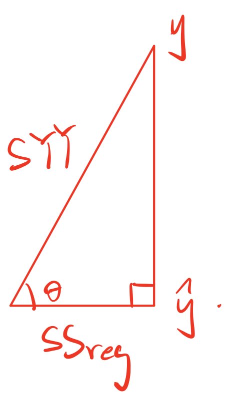
$\vec{J}_n = (1, 1, \cdots, 1)$.

$$R^2 = \frac{SS_{Reg}}{SYY}.$$

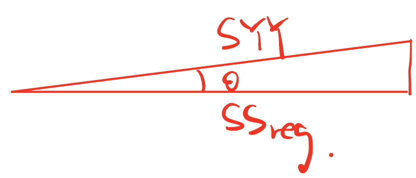
$$\begin{array}{ccccc} \in\mathbb{R}^n & & J_n\in\mathbb{R}^n \quad \vec{x}\in\mathbb{R}^n & & \\ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} & = & \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} & \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + & \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \end{array}$$

$$\min \sum_{i=1}^{n} e_i^2.$$





$R^2 = \cos(\theta)$.



$\cos\theta \to 1$.

$R^2 \to 1$.

$R^2 \to 1$.

$$R^2 = \frac{(SXY)^2}{SXX \cdot SYY} = \frac{SS_{reg}}{SYY} = r_{xy}^2.$$

$R^2 = r_{xy}^2$. $\quad r_{xy} = +/- \sqrt{R^2}$.

linear regression $y = \beta_0 + \beta_1 x$. $\boxed{y = \beta_0 + \beta_1 x^2}$. x^3 x^4.

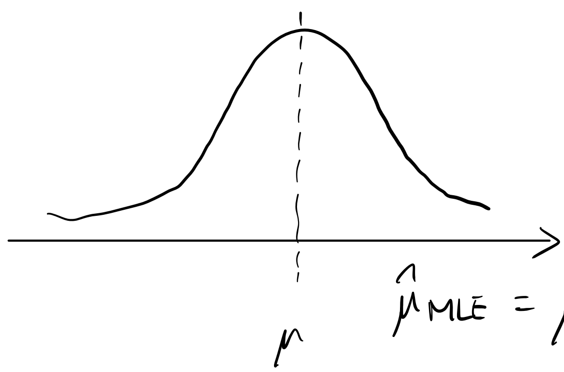
$z_i = x_i^2$ $y = \beta_0 + \beta_1 z$.

$(\bar{z_i}, \bar{y}) = (\overline{x^2}, \bar{y})$.

Section 2.3. MLE. $(\hat{\beta_0}, \hat{\beta_1})_{OLS} = (\hat{\beta_0}, \hat{\beta_1})_{MLE}$.

Assumption: $e_i \sim N(0, \sigma^2)$.

$\min E(\hat{\mu} - \mu)^2 \leftarrow \hat{\mu} = \bar{X}$. $\longleftrightarrow$ $\hat{\mu}_{MLE} = \bar{X}$.



$\hat{\mu}_{MLE} = \hat{\mu}_{OLS} = \bar{X}$.

T/F : $\varepsilon \sim$ distribution. $Y = X\beta + \varepsilon$ $\hat{\beta}_{OLS} = \hat{\beta}_{MLE}$ X.

normal

3. Theorem:

Suppose $X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

① $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$. ② $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ ③ $\overline{X} \perp\!\!\!\perp S^2$.

Proof: ① √. linear combination of $N(\cdot,\cdot)$ is still normal.

②

$$A \triangleq \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & & \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}$$ is an orthogonal matrix.

$\boxed{AA^T = I}$

(Gram-Schmidt Process). $A^{-1} = A^T$.

$AA^T = I$. $\rightarrow$ $i\begin{bmatrix} \text{———} \\ \\ \end{bmatrix}_A \begin{bmatrix} & \overset{j}{|} & \\ & | & \end{bmatrix}_{A^T}$.

$i \neq j$: $\text{——} \, | = 0$. $i = j$: $\text{——} \, | = 1$.

$Y = AX$. $\begin{pmatrix} Y_1 \\ \vdots \\ \vdots \\ Y_n \end{pmatrix} = A \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$.

$Y_1 = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} X_i = \sqrt{n} \cdot \sum_{i=1}^{n} X_i/n = \sqrt{n}\,\overline{X}$. $\Rightarrow Y_1^2 = n\overline{X}^2$. ①

$Y_1^2 + \cdots + Y_n^2 = Y^TY = (AX)^TAX = X^TA^TAX = X^TIX = X^TX$

$\Rightarrow$

$Y_1^2 + \cdots + Y_n^2 = X_1^2 + \cdots + X_n^2$. ②

$$
\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\
&= \sum_{i=1}^{n}X_i^2 - 2\bar{X}\cdot\sum_{i=1}^{n}X_i + n\bar{X}^2 \\
&= \sum_{i=1}^{n}X_i^2 - 2\bar{X}\cdot n\bar{X} + n\bar{X}^2 \\
&= \sum_{i=1}^{n}X_i^2 - n\bar{X}^2 .
\end{aligned}
$$

$$
= \sum_{i=1}^{n}Y_i^2 - Y_1^2 = \sum_{i=2}^{n}Y_i^2 = Y_2^2 + \cdots + Y_n^2
$$

$Y = AX$. A is orthogonal. $Y_1, \cdots, Y_n$ independent normal

$Y_i \sim N(\mu_i, \sigma^2)$.

$$
\mu_i = \sum_{i=1}^{n} a_{ik}\mu = \sqrt{n}\mu \sum_{i=1}^{n}\underline{\frac{1}{\sqrt{n}}a_{ik}} = 0 \leftarrow i \neq 1. \ (i=2,3,\cdots,n).
$$

$$
Y_2, \cdots, Y_n \sim N(0, \sigma^2). \Rightarrow \frac{Y_2}{\sigma}, \cdots, \frac{Y_n}{\sigma} \sim N(0,1).
$$

$$
\Rightarrow \quad (n-1)S^2/\sigma^2 = \sum_{i=2}^{n} Y_i^2/\sigma^2 = \sum_{i=2}^{n}\left(\frac{Y_i}{\sigma}\right)^2 \sim \chi^2_{n-1}
$$

③ $Y_1 = \sqrt{n}\bar{X} \Rightarrow$

$$
\left.\begin{aligned}
\bar{X} &= \frac{1}{\sqrt{n}}Y_1 . \leftarrow Y_1 .. \\
S^2 &= \underline{f(Y_2, \cdots, Y_n)}.
\end{aligned}\right\} \quad \bar{X} \perp\!\!\!\perp S^2 .
$$