ECON 6190 Section 5

Sept. 27, 2024

Yiwei Sun

1. [Hong 6.8] Establish the following recursion relations for sample means and sample variances. Let $\bar{X}_n$ and s_n^2 be the sample mean and sample variances based on random sample $\{X_1, X_2 \dots X_n\}$. Then suppose another observation, X_{n+1}, becomes available. Show:

   (a) $\bar{X}_{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}$.

   (b) $ns_{n+1}^2 = (n-1)s_n^2 + \frac{n}{n+1}(X_{n+1} - \bar{X}_n)^2$.

(a) Check solution

(b) Method 1 Start with LHS

$$n S_{n+1}^2 \overset{def}{=} \cancel{n}\left(\frac{1}{\cancel{n}}\sum_{i=1}^{n+1}(x_i - \bar{x}_{n+1})^2\right)$$

$$= \sum_{i=1}^{n+1}(\underbrace{x_i - \bar{x}_n}_{} + \underbrace{\bar{x}_n - \bar{x}_{n+1}}_{})^2$$

$$= \sum_{i=1}^{n+1}(x_i - \bar{x}_n)^2 + \underbrace{2\sum_{i=1}^{n+1}(x_i - \bar{x}_n)(\bar{x}_n - \bar{x}_{n+1})}_{} + \underbrace{\sum_{i=1}^{n+1}(\bar{x}_n - \bar{x}_{n+1})^2}_{}$$

only term indexed by i

not indexed by i

$$= (n+1)(\bar{x}_n - \bar{x}_{n+1})^2$$

$$= 2(\bar{x}_n - \bar{x}_{n+1})\underbrace{\sum_{i=1}^{n+1}(x_i - \bar{x}_n)}_{}$$

$$= \sum_{i=1}^{n+1} x_i - (n+1)\bar{x}_n$$

$$= (n+1)\left(\underbrace{\frac{1}{n+1}\sum_{i=1}^{n+1} x_i}_{\bar{x}_{n+1}} - \bar{x}_n\right)$$

$$= (n+1)(\bar{x}_{n+1} - \bar{x}_n)$$

$$= -2(n+1)(\bar{x}_n - \bar{x}_{n+1})^2$$

side note: on a high level, when doing +/- trick, you typically get forms like

$$(\square)^2 + \underbrace{2(\square)(\triangle)}_{= -2(\triangle)^2} + (\triangle)^2$$

$$= (\quad)^2 - (\quad)^2$$

$$= \sum_{i=1}^{n+1}(x_i - \bar{x}_n)^2 - (n+1)(\bar{x}_n - \bar{x}_{n+1})^2$$

$$= \underbrace{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}_{(n-1)S_n^2} + \underbrace{(x_{n+1} - \bar{x}_n)^2 - (n+1)(\bar{x}_n - \bar{x}_{n+1})^2}_{}$$

WTS: $\frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2$

Recall from (a): $\bar{x}_{n+1} = \frac{x_{n+1} + n\bar{x}_n}{n+1}$

$$\Rightarrow (\bar{x}_n - \bar{x}_{n+1})^2 = \left(\bar{x}_n - \frac{x_{n+1} + n\bar{x}_n}{n+1}\right)^2$$

$$= \left(\frac{x_{n+1}}{n+1} - \frac{(\cancel{n}+1-\cancel{n})\bar{x}_n}{n+1}\right)^2$$

$$= \left(\frac{1}{n+1}\right)^2(x_{n+1} - \bar{x}_n)^2$$

$$= (n-1)S_n^2 + (X_{n+1}-\bar{X}_n)^2 - \cancel{(n+1)}\left(\frac{1}{n+1}\right)^{\cancel{2}}(X_{n+1}-\bar{X}_n)^2$$

$$= (n-1)S_n^2 + \frac{n}{n+1}(X_{n+1}-\bar{X}_n)^2$$

Method #2 Start with RHS

$$(n-1)S_n^2 + \frac{n}{n+1}(X_{n+1}-\bar{X}_n)^2$$

DEF

$$= \cancel{(n-1)}\left(\frac{1}{\cancel{n-1}}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right) + \underbrace{\frac{n}{n+1}}_{= \frac{n+1}{n+1} - \frac{1}{n+1} = 1 - \frac{1}{n+1}}(X_{n+1}-\bar{X}_n)^2$$

$$= \underbrace{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 + (X_{n+1}-\bar{X}_n)^2}_{\text{combine}} - \frac{1}{n+1}(X_{n+1}-\bar{X}_n)^2$$

$$= \sum_{i=1}^{n+1}(X_i-\bar{X}_n)^2 - \frac{1}{n+1}(\underbrace{X_{n+1}}_{=(n+1)\bar{X}_{n+1} - n\bar{X}_n} - \bar{X}_n)^2$$

b/c $\cancel{(n+1)}\frac{1}{\cancel{n+1}}\sum_{i=1}^{n+1}X_i - \cancel{n}\frac{1}{\cancel{n}}\sum_{i=1}^{n}X_i = X_{n+1}$

$$= \sum_{i=1}^{n+1}(X_i-\bar{X}_n)^2 - \frac{1}{n+1}\left((n+1)\bar{X}_{n+1} - (n+1)\bar{X}_n\right)^2$$

$$= \sum_{i=1}^{n+1}(X_i-\bar{X}_n)^2 - \underbrace{(n+1)(\bar{X}_{n+1}-\bar{X}_n)^2}_{= \sum_{i=1}^{n+1}(\bar{X}_{n+1}-\bar{X}_n)^2}$$

constant, not indexed by i

$$= \sum_{i=1}^{n+1}(\underbrace{X_i-\bar{X}_n}_{a})^2 - (\underbrace{\bar{X}_{n+1}-\bar{X}_n}_{b})^2$$

Recall: $a^2-b^2 = (a+b)(a-b)$

$$= \sum_{i=1}^{n+1}(X_i-\bar{X}_n+\bar{X}_{n+1}-\bar{X}_n)(X_i-\cancel{\bar{X}_n}-\bar{X}_{n+1}+\cancel{\bar{X}_n})$$

$$= \sum_{i=1}^{n+1}\left(\underline{X_i-\bar{X}_{n+1}}+2\underline{(\bar{X}_{n+1}-\bar{X}_n)}\right)(X_i-\bar{X}_{n+1})$$

$$= \underbrace{\sum_{i=1}^{n+1}(X_i-\bar{X}_{n+1})^2}_{= n\left(\frac{1}{n}\sum_{i=1}^{n+1}(X_i-\bar{X}_{n+1})^2\right) = nS_{n+1}^2} + \underbrace{2\sum_{i=1}^{n+1}(\bar{X}_{n+1}-\bar{X}_n)(X_i-\bar{X}_{n+1})}_{\circledast}$$

WTS: $\circledast = 0$

Consider $\circledast$: $2\sum_{i=1}^{n+1}\left(\bar{X}_{n+1}X_i - (\bar{X}_{n+1})^2 - X_i\bar{X}_n + \bar{X}_n\bar{X}_{n+1}\right)$

$$= 2\left(\bar{X}_{n+1}\underbrace{\sum_{i=1}^{n+1}X_i}_{=(n+1)\bar{X}_{n+1}} - \cancel{(n+1)(\bar{X}_{n+1})^2} - \bar{X}_n\underbrace{\sum_{i=1}^{n+1}X_i}_{=(n+1)\bar{X}_{n+1}} + \cancel{(n+1)(\bar{X}_n\bar{X}_{n+1})}\right)$$

$$\cancel{(n+1)(\bar{X}_{n+1})^2} \qquad = -\cancel{(n+1)(\bar{X}_n\bar{X}_{n+1})}$$

$$=0$$

$\Rightarrow$ RHS $= nS_{n+1}^2$ $\blacksquare$

2. [Hong 6.6] Suppose $\mathbf{X}^n = (X_1, \ldots, X_n)$ is an iid $N(\mu, \sigma^2)$ random sample, $\mathbf{Y}^n = (Y_1, \ldots, Y_n)$ is an iid $N(\mu, \sigma^2)$ random sample, and the two random samples are mutually independent. Let $\bar{X}_n$ and $\bar{Y}_n$ be the sample means of the first and second random samples, respectively, and let s_X^2 and s_Y^2 be the sample variances of the first and second random samples respectively. Find:

(a) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{2\sigma^2/n}$;

Theorem If X_i, $i = 1, \ldots, n$ are iid $\mathcal{N}(\mu, \sigma^2)$, then

① $\bar{X}_n$ and S^2 are independent

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

$\boxed{\text{DEF}}$ Let $\{Z_1, Z_2, \ldots, Z_r\}$ be iid $\mathcal{N}(0,1)$, then $\sum_{i=1}^{r} Z_i^2 \sim \chi_r^2$

$\boxed{\text{DEF}}$ Let $Z \sim \mathcal{N}(0,1)$ and $Q \sim \chi_r^2$ be independent. Then $T = \frac{Z}{\sqrt{\frac{Q}{r}}} \sim t_r$

(a) (b) (c) see solution

(d) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{(s_X^2 + s_Y^2)/n}$;

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \sim \mathcal{N}(0,1)$$

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{S_X^2+S_Y^2}{n}}} = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \cdot \frac{\sqrt{\frac{2\sigma^2}{\cancel{n}}}}{\sqrt{\frac{S_X^2+S_Y^2}{\cancel{n}}}} = \frac{\overbrace{\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}}}^{\sim \mathcal{N}(0,1)}}{\underbrace{\sqrt{\frac{S_X^2+S_Y^2}{2\sigma^2}}}}$$

$$= \sqrt{\frac{(n-1)(S_X^2+S_Y^2)}{2(n-1)\sigma^2}} = \sqrt{\frac{(n-1)S_X^2}{2(n-1)\sigma^2} + \frac{(n-1)S_Y^2}{2(n-1)\sigma^2}}$$

(loose notation) $\sqrt{\frac{1}{2(n-1)}\underbrace{\chi^2_{n-1}}_{\sum_{i=1}^{n-1}Z_i^2} + \frac{1}{2(n-1)}\underbrace{\chi^2_{n-1}}_{\sum_{i=1}^{n-1}Z_i^2}} = \sqrt{\frac{1}{2(n-1)}\underbrace{\sum_{i=1}^{2(n-1)}Z_i^2}_{\chi^2_{2(n-1)}}}$

Since $\bar{X}_n$ and S_X^2 independent, $\bar{Y}_n$ and S_Y^2 are independent,

$$\left(\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{2\sigma^2}{n}}}\right) \text{ and } \left(\frac{S_X^2+S_Y^2}{2\sigma^2}\right) \text{ are also independent.}$$

$$\Rightarrow \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{S_X^2+S_Y^2}{n}}} \sim t_{2(n-1)}$$

(e) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{s_n^2/n}$, where s_n^2 is the sample variance of the difference sample $\mathbf{Z}^n = (Z_1, Z_2 \ldots Z_n)$, where $Z_i = X_i - Y_i$, $i = 1, 2 \ldots n$.

$$Z_i = X_i - Y_i, \quad Z \sim N(\mu-\mu, \sigma^2+\sigma^2+0)$$

$$\Rightarrow Z \sim N(0, 2\sigma^2)$$

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{S_n^2}{n}}} = \frac{\bar{Z}_n}{\sqrt{\frac{2\sigma^2}{n}}} \cdot \frac{\sqrt{\frac{2\sigma^2}{\cancel{n}}}}{\sqrt{\frac{S_n^2}{\cancel{n}}}} = \frac{\frac{\bar{Z}_n}{\sqrt{\frac{2\sigma^2}{n}}}}{\sqrt{\frac{S_n^2}{2\sigma^2}}}$$

$\frac{\bar{Z}_n}{\sqrt{\frac{2\sigma^2}{n}}} \sim N(0,1)$

$2\sigma^2$: σ_Z^2

$\frac{S_n^2}{2\sigma^2} \sim \frac{\chi^2_{n-1}}{n-1}$

Since $\bar{Z}_n$ and S_n^2 are independent,

$$\frac{\bar{X} - \bar{Y}_n}{\sqrt{\frac{S_n^2}{n}}} \sim t_{n-1}$$

4. [Final exam, 2022] Let $\{X_1, \ldots X_n\}$ be i.i.d with pdf $f(x \mid \theta) = e^{-(x-\theta)}\mathbf{1}\{x \geq \theta\}$. Show $Y = \min\{X_1, \ldots X_n\}$ is a sufficient statistic for θ **without** using the Factorization Theorem.

$$\mathbb{1}\{x \geq \theta\} = \begin{cases} 1, & x \geq \theta \\ 0, & o/w \end{cases}$$

$$f(x|\theta) = e^{-(x-\theta)}\mathbb{1}\{x \geq \theta\} \iff f(x|\theta) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & o/w \end{cases}$$

<u>Property of Indicator function:</u> $E[\mathbb{1}\{x \in A\}] = P(x \in A)$

WTS: $\frac{P(\mathbf{X}|\theta)}{q(t(\mathbf{X})|\theta)}$ is not a function of θ over the sample space

DATA

DATA

1) density of $\{x_1, \dots, x_n\}$.

$$P(x_1, \dots, x_n \mid \theta) \overset{iid}{=} \begin{cases} e^{-\sum_{i=1}^{n} x_i} e^{n\theta} , & \min\{x_1, \dots x_n\} \geq \theta. \\ 0 , & o/w \end{cases}$$

2) To find the density of Y, start with cdf of Y

$$\begin{aligned} P(Y \leq y) &= P(\min\{x_1, \dots, x_n\} \leq y) \\ &= 1 - P(\min\{x_1, \dots, x_n\} > y) \\ &= 1 - P(x_1 > y, x_2 > y, \dots x_n > y) \\ &= 1 - \prod_{i=1}^{n} P(x_i > y) \\ &= \begin{cases} 1 - e^{-n(y-\theta)} , & \text{for } y \geq \theta \\ 0 , & o/w \end{cases} \end{aligned}$$

For $y < \theta$, $P(x_i > y) = 1$

$\Rightarrow$ pdf of y $\quad q(y \mid \theta) = \begin{cases} n e^{-n(y-\theta)} , & \text{for } y \geq \theta \\ 0 , & o/w \end{cases}$

$$\Rightarrow \frac{P(x_1, \dots, x_n \mid \theta)}{q(y \mid \theta)} = \frac{e^{-\sum_{i=1}^{n} x_i}}{n e^{-n(\min\{x_1, \dots x_n\})}} \quad \text{for } \min(x_1, \dots x_n) \geq \theta.$$

↳ NOT A function of θ

$\Rightarrow$ $\min\{x_1, \dots, x_n\}$ is a s.s.