[redacted]

(1)

Final Exam

20.05.2022

(P1). Given a random sample $X_1 \dots X_n$ from Uniform$(1, \theta)$ where θ is unknown

a) Find method of moments estimator of θ. Is it biased?

Solution: Method of Moments sets mean $\bar{X}$ to First Moment.

$\mu = E(X) = \frac{1+\theta}{2}$ – first Moment.

$\bar{X} = \frac{1+\theta}{2} \Rightarrow \theta = 2\bar{X} - 1$

$E(X) = \int x \cdot f(x)\,dx = \int_1^\theta x \cdot \frac{1}{\theta-1}\,dx = \frac{1}{\theta-1}\left[\frac{x^2}{2}\right]_1^\theta =$

$= \frac{1}{\theta-1} \cdot \frac{\theta^2-1}{2} = \frac{\cancel{(\theta-1)}(\theta+1)}{\cancel{(\theta-1)} \cdot 2} = \frac{\theta+1}{2}$ first moment

$E(X^2) = \int x^2 \cdot f(x)\,dx = \int_1^\theta x^2 \cdot \frac{1}{\theta-1}\,dx =$

$= \frac{1}{\theta-1}\left[\frac{x^3}{3}\right]_1^\theta = \frac{1}{\theta-1} \cdot \frac{1}{3}\left[\theta^3 - 1\right] =$

$= \frac{1}{\cancel{\theta-1}} \cdot \frac{1}{3} \cdot \left[\cancel{(\theta-1)}(\theta^2+2\theta+1)\right] = \frac{\theta^2+2\theta+1}{3}$

Second moment

Is it biased?

Solution: We say the bias = 0 when estimate is unbiased and bias $\neq 0$ otherwise. The expected value and actual value in difference should give 0. Then $E(\hat{\theta}) - \theta = 0$ - unbiased

Bias: $B(\hat{\theta}) = E(\hat{\theta}) - \theta =$

$= E(2\bar{X} - 1) - \theta = 2E(\bar{X}) - 1 - \theta = \cancel{2} \cdot \frac{1+\theta}{\cancel{2}} - 1 - \theta =$

$= \cancel{1+\theta} \cancel{-1-\theta} = 0$ unbiased.

Why? Because $\bar{X} = \frac{1+\theta}{2}$

Why? Because $\bar{X} = \frac{1+\theta}{2} \Rightarrow \theta = 2\bar{X} - 1$

b) Find maximum likelihood estimator of θ?

Formula:

$$L(\theta) = \prod_{i=1}^{n} f(x_i)$$

We know that PDF for Uniform(a, b) is $\frac{1}{b-a}$

so,

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta-1} = \left(\frac{1}{\theta-1}\right)^n = \frac{1}{(\theta-1)^n} = (\theta-1)^{-n}$$

We can differentiate, but it is easier to take logarithm first.

log function is monotonic, so it will not affect the answer.

$$\ell(\theta) = \ln(L(\theta)) = -n \cdot \ln(\theta-1)$$

$$\frac{d\ell(\theta)}{d\theta} = -\frac{n}{\theta-1}$$

$$-\frac{n}{\theta-1} = 0 \qquad \theta = n \Rightarrow \text{Therefore,}$$

$(\theta-1)^{-n}$ is decreasing function but it is maximized at X_n

so $\theta = \max\{X_1 \dots X_n\}$

if X sample space is ~~sorted~~.

(P2) There is a biased coin and you want to test if it probability of heads is 0.6 You toss the coin 10000 time and observe 6152 heads

3

$H_0: p = 0.6$

$H_a: p \neq 0.6$

$n = 10000$

$p = \frac{6152}{10000} = 0,6152$ where $m = 6152$

It is our proportion.

let's find Z

$$Z = \frac{0.6152 - 0.6}{\sqrt{\frac{0.6 \cdot 0.4}{10000}}} = \frac{0.0152}{\sqrt{\frac{0.24}{10000}}} = \frac{0.0152}{0.0049} =$$

$= 3.1020$

Conclusion
$3.1 > 1.96 \Rightarrow$ reject H_0

$$p\text{-value} = P(Z > |z|) = 2 \cdot P(Z > 2.69) =$$

$$= 2 \cdot (1 - 0.9990) = 0.002$$

p-value $< \alpha$, where $\alpha = 0.05$, thus we do reject H_0

P3

Day	1	2	3	4
Sales	100	130	180	290

X_i	y_i	$X_i \cdot y_i$	X_i^2
1	100	100	1
2	130	260	4
3	180	540	9
4	290	1160	16
10	700	2060	30

Calculating means

$\overline{X} = \frac{10}{4} = 2.5$

$\overline{y} = \frac{700}{4} = 175$

$$\beta_1 = \frac{\sum (X_i - \overline{X})(y_i - \overline{y})}{\sum (X_i - \overline{X})^2} = \frac{\sum X_i \cdot y_i - n \cdot \overline{X} \cdot \overline{y}}{\sum X_i^2 - n \cdot (\overline{X})^2}.$$

$$\hat{\beta}_1 = \frac{2060 - 4 \cdot 2.5 \cdot 175}{30 - 4 \cdot (2.5)^2} = \frac{2060 - 1750}{30 - 25} =$$

$$= \frac{310}{5} = 62 \text{ - slope}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{X} \Rightarrow \hat{\beta}_0 = 175 - 62 \cdot 2.5 \Rightarrow$$

$$\hat{\beta}_0 = 175 - 155 = 20 \text{ - Interceptor}$$

The regression line model has formula:

$y_i = \beta_0 + \beta_1 \cdot X_i$, therefore $\Rightarrow$

$y_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i \Rightarrow$

$y_i = 20 + 62 \cdot X_i$

b) Estimate standard deviation of residuals (fluctuations).

Estimate std. deviation of residuals.

The predicted value or fitted value is $\hat{y}_i = \hat{F}(X_i)$, where $\hat{F}(X_i) \Rightarrow$ regression model.

The residual is $\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{F}(X_i)$

RSS - Residual sum of squares - how well fit of data.

$RSS = \sum_i \hat{\varepsilon}_i^2$

y_i	$\hat{y}_i$	$\hat{\varepsilon}_i = y_i - \hat{y}_i$	$\hat{\varepsilon}_i^2$
100	82	18	324
130	144	-14	196
180	206	-26	676
290	268	22	484
sum			1680

$Var = \frac{RSS}{n-2} = \left(\frac{1}{n-2}\right) \cdot \sum_i \hat{\varepsilon}_i^2 = \left(\frac{1}{n-2}\right) \cdot \sum_i (y_i - \hat{y}_i)^2$ - unbiased estimator.

$Var = \frac{1680}{n-2} = \frac{1680}{2} = 840$

Standard deviation is

$\sigma = \sqrt{Var} = \sqrt{840} = 28.98$

5

c) Find 75% confidence intervals for beta 0, beta 1 (intercept and slope)

$\alpha = 1-0.75 = 0.25$

Z-score for $Z_{\frac{\alpha}{2}}$ or

$Z_{\frac{0.25}{2}} = 1.150$

$\hat{\beta}_0 = 20$

$\hat{\beta}_1 = 62$

$S_x = \sqrt{\frac{\sum(X_i-\bar{X})^2}{n}}$

$\hat{SE}(\hat{\beta}_0) = \frac{\hat{6}}{S_x \sqrt{n}} \sqrt{\frac{\sum X_i^2}{n}}$

$\hat{SE}(\hat{\beta}_1) = \frac{\hat{6}}{S_x \sqrt{n}}$

First I compute S_x

$S_x = \sqrt{\frac{(1-2.5)^2+(2-2.5)^2+(3-2.5)^2+(4-2.5)^2}{4}} = \sqrt{\frac{5}{4}} =$

$= \sqrt{1.25} = 1.1180.$

$\hat{SE}(\hat{\beta}_0) = \frac{28.98}{1.1180 \cdot \sqrt{4}} \cdot \sqrt{\frac{30}{4}} = \frac{28.98}{2.236} \cdot \sqrt{\frac{30}{4}} =$

$= 12.96 \cdot 2.7386 = 35.49$

$\hat{SE}(\hat{\beta}_1) = \frac{28.98}{1.118 \cdot \sqrt{4}} = \frac{28.98}{2.236} = 12.96$

for β_0: $[20 \pm 1.15 \cdot 35.49]$

for β_1: $[62 \pm 1.15 \cdot 12.96]$

d) Predict on Day 5 and define 75% prediction interval.

$$y_i = 20 + 62 \cdot X_i$$

$$y_i = 20 + 62 \cdot 5 = 20 + 310 = 330$$

$$\rho = \hat{\sigma}\sqrt{1 + \frac{\sum(X_i - X_{new})^2}{n\sum(X_i - \bar{X})^2}} = 28.98\sqrt{1 + \frac{30}{4 \cdot 5}}$$

~~330 ± 1.150~~