

STA2002 - Homework 1

Xue Zhongkai 122090636

PROBLEM 1.

(a) By definition, we have

$$Mean = \frac{1.75 + 1.84 + 2.12 + 1.92 + 2.62 + 2.35 + 3.09 + 3.15 + 2.53 + 1.91 + 3.25 + 2.83}{12} = 2.447$$

$$Var = \frac{1}{12} \times \sum_{k=1}^{12} (x_i - 2.447)^2 = 0.298$$

$$Std = \sqrt{0.298} = 0.545$$

(b) We have the ordered list:

$$1.75 < 1.84 < 1.91 < 1.92 < 2.12 < 2.35 < 2.53 < 2.62 < 2.83 < 3.09 < 3.15 < 3.25$$

For Q_1 ,

$$h = 1 + (12 - 1) \times \frac{1}{4} = 3.75,$$

$$Q_1 = x_3 + (3.75 - 3)(x_4 - x_3) = 1.9175.$$

For Q_2 ,

$$Q_2 = Median = \frac{x_6 + x_7}{2} = 2.44.$$

For Q_3 ,

$$h = 1 + (12 - 1) \times \frac{3}{4} = 9.25$$

$$Q_3 = x_9 + (9.25 - 9)(x_{10} - x_9) = 2.895.$$

$$\text{As } IQR = Q_3 - Q_1 = 0.9775,$$

$$Q_1 - 1.5IQR = 0.45125 < x_1$$

$$Q_3 + 1.5IQR = 4.36125 > x_{12}$$

No outliers.

(See attached)

PROBLEM 2.

(See attached)

(Part III) Overall, these findings suggest that the type of smile or facial expression displayed by the accused student may have influenced the leniency judgments of the subjects. The results align with the smile-lenieny effect, indicating that smiling can indeed impact judgments of wrongdoing.

PROBLEM 3.

Denote the proportion of students owning a Mac at CUHKSZ as p_1 and at SUSTech as p_2 .

The sample proportion is P_1 and P_2 , respectively.

Given observations, i.e. $n_1 = 200$, $X_1 = 150$, $n_2 = 250$, $X_2 = 185$, we have

$$P_1 - P_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2} = \frac{150}{200} - \frac{185}{250} = 0.75 - 0.74 = 0.01$$

As n_1 and n_2 are sufficiently large (larger than 30), CLT could be applied, for which the sampling distribution of sample proportions will be approximately normal.

Used as an approximation,

$$\hat{\mu}(P_1 - P_2) = P_1 - P_2 = 0.01.$$

Using variance of binomial as an approximation where

$$Var(X) = np(1-p), Var\left(\frac{X}{n}\right) = \frac{p(1-p)}{n},$$

we have

$$\sigma = Var(P_1 - P_2) = Var(P_1) + Var(P_2) = \frac{0.75(1-0.75)}{200} + \frac{0.74(1-0.74)}{250} = 0.001707$$

As a result, it is a normal distribution with mean 0.01 and variance 0.001707.

PROBLEM 4.

(a) Use $\bar{X}_1 - \bar{X}_2$ as the unbiased estimator, as

$$E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2.$$

X_1 and X_2 are independent to each other s.t.

$$Std(X_1 - X_2) = \sqrt{Var(\bar{X}_1) + Var(\bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

As a result, the standard deviation is $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

(b) By definition, we have

$$Bias = E(\bar{X}_1^2 - \bar{X}_2^2) - (\mu_1^2 - \mu_2^2) = (\mu_1^2 + \frac{\sigma_1^2}{n_1}) - (\mu_2^2 + \frac{\sigma_2^2}{n_2}) - (\mu_1^2 - \mu_2^2) = \frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}$$

When n_1 and n_2 goes to ∞ ,

\bar{X}_1 and \bar{X}_2 will converge to μ_1 and μ_2 , and the remains will converge to 0.

That is why the bias converges to 0.

(c) By definition,

$$\begin{aligned} E[Sp^2] &= E\left[\frac{(n_1 - 1)S_1^2}{n_1 + n_2 - 2} + \frac{(n_2 - 1)S_2^2}{n_1 + n_2 - 2}\right] \\ &= \frac{(n_1 - 1)E(S_1^2)}{n_1 + n_2 - 2} + \frac{(n_2 - 1)E(S_2^2)}{n_1 + n_2 - 2} \\ &= \frac{(n_1 - 1)\sigma^2}{n_1 + n_2 - 2} + \frac{(n_2 - 1)\sigma^2}{n_1 + n_2 - 2} \\ &= \sigma^2 \end{aligned}$$

Therefore, S_p^2 is an unbiased estimator.

PROBLEM 5.

(a) For an exponential distribution with parametre λ , the population mean is $\frac{1}{\lambda}$.

By MoM, we have

$$\frac{1}{\lambda} = \bar{x}.$$

That is,

$$\hat{\lambda}_{mom} = \frac{n}{x_1 + x_2 + \dots + x_n}.$$

(b) By MLE, the likelihood function of the sample is

$$L(\theta) = f(x_1; \theta)f(x_2; \theta) \dots f(x_n; \theta) = \lambda^n \exp(-\lambda(x_1 + x_2 + \dots + x_n))$$

Take logarithm on both sides,

$$l(\theta) = n \ln(\lambda) - \lambda(x_1 + x_2 + \dots + x_n)$$

Take derivative w.r.t λ and make it equal to zero,

$$\begin{aligned}\frac{d}{d\lambda} \ln(L(\theta)) &= \frac{n}{\lambda} - (x_1 + x_2 + \cdots + x_n) = 0 \\ \frac{d^2}{d\lambda^2} &= -\frac{n}{\lambda^2} < 0.\end{aligned}$$

That is,

$$\hat{\lambda}_{mle} = \frac{n}{x_1 + x_2 + \cdots + x_n}.$$

(c) Given the sample,

$$\bar{x} = \frac{(3.8 + 3.24 + 1.4 + 1.22 + 4.5 + 4.6)}{6} = 3.1267$$

As a result,

$$\begin{aligned}\hat{\lambda}_{mom} &= \frac{1}{\bar{x}} = 0.3198, \\ \hat{\lambda}_{mle} &= \frac{1}{\bar{x}} = 0.3198\end{aligned}$$

(d) In exponential distribution, the mean of the population is $\frac{1}{\lambda}$.

By MLE, the mean of sample is $\lambda^n \exp(-\lambda(x_1 + x_2 + \cdots + x_n))$ as well.

When n goes to ∞ , the mean of sample is to converge to that of population.

As a result, it is an unbiased estimator of λ .

PROBLEM 6.

(a) For the distribution with pdf $(\theta + 1)x^\theta$, the population mean is

$$E(X) = \int_{\theta}^1 x(\theta + 1)x^\theta dx = \int_{\theta}^1 (\theta + 1)x^{\theta+1} dx = \frac{\theta + 1}{\theta + 2}$$

The sample mean is \bar{X} .

By MoM, we have

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

That is,

$$\theta_{mom} = \frac{2\bar{X} - 1}{1 - \bar{X}}.$$

(b)

By MLE,

$$L(\theta) = f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta) = (\theta + 1)^n (x_1 x_2 \cdots x_n)^\theta$$

Take the logarithm on both sides,

$$l(\theta) = n \ln(\theta + 1) + \theta \sum_{i=1}^n \ln x_i$$

Take the derivative,

$$\begin{aligned} \frac{d}{d\lambda} l(\theta) &= \frac{n}{\theta + 1} + \theta \sum_{i=1}^n \ln(x_i) = 0 \\ \frac{d^2}{d\theta^2} l(\theta) &= -\left(\frac{n}{\theta + 1}\right)^2 < 0 \end{aligned}$$

That is,

$$\hat{\theta}_{mle} = -1 - \frac{n}{\sum_{i=1}^n \ln(x_i)}$$

PROBLEM 7.

For a continuous uniform distribution over (a, b) ,

The first moment is

$$M_1 = \int_a^b x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

The sample mean is

$$\hat{\mu}_1 = E(X) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Set $\frac{a+b}{2} = \bar{X}$, then $b = 2\bar{X} - a$.

The second moment is

$$M_2 = \int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}$$

The sample square mean is

$$\hat{\mu}_2 = E(X^2) = \frac{1}{n} \sum_{i=1}^n x_i^2$$

Substituted a , we have

$$(2\hat{\mu}_1 - b)^2 + b^2 + (2\hat{\mu}_1 - b)b = 3\hat{\mu}_2$$

Since $a < E(X) < b$,

$$a = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \frac{1}{n} \sum_{i=1}^n x_i - \sqrt{3\left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right]} = \bar{X} - \sqrt{\frac{3(n-1)}{n} S^2}$$

$$b = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \frac{1}{n} \sum_{i=1}^n x_i + \sqrt{3\left[\frac{1}{n}x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right]} = \bar{X} + \sqrt{\frac{3(n-1)}{n}}S^2$$