Ömer Oktay Gültekin
21901413
Mat 230 – 5
CS

1. a) Let $X_1, X_2, \cdots X_n$ be a random sample from a distribution with mean μ and variance σ^2. Show that sample variance S^2 is an unbiased estimator of σ^2. Recall that sample variance is

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left((x_i - \mu)^2 - (\bar{x} - \mu)\right)^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i - \mu)^2 - 2\sum_{i=1}^{n}(x_i - \mu)(\bar{x} - \mu)\right.$$

$$\left. + \sum_{i=1}^{n}(\bar{x} - \mu)^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}E(x_i - \mu)^2 - nE(\bar{x} - \mu)^2\right]$$

Using the definitions of $\sigma^2 = E(x_i - \mu)^2$, $Var(\bar{x}) = E(\bar{x} - \mu)^2 = \frac{\sigma^2}{n}$ (Proofs are in the lecture notes, I assumed they can be used without explicit proof) $\Rightarrow E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right]$

$= \frac{1}{n-1}\left[n\sigma^2 - \sigma^2\right] = \frac{1}{n-1}\cdot(n-1)\sigma^2 = \sigma^2 \Rightarrow$ $\boxed{S^2 \text{ is an UE of } \sigma^2.}$

b) Let $X_1, X_2, \cdots X_n$ and $Y_1, Y_2, \cdots Y_m$ denote random samples from independent populations with common unknown variance σ^2. An estimator for σ^2 is given by .

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2},$$

where S_1^2 and S_2^2 are sample variances of the first and second samples. Show that S_p^2 is an unbiased estimator of σ^2.

$$E(S_p^2) = E\left(\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}\right) = \frac{(n-1)E(S_1^2) + (m-1)E(S_2^2)}{n+m-2}$$

using the results from part a, $E(S_1^2) = \sigma^2$ and $E(S_2^2) = \sigma^2$ where σ^2 is the common unknown variance $\Rightarrow E(S_p^2) = \frac{(n-1)\sigma^2 + (m-1)\sigma^2}{n+m-2}$

$= \frac{\sigma^2(n-1+m-1)}{n+m-2} = \sigma^2 \Rightarrow$ $\boxed{S_p^2 \text{ is an UE of } \sigma^2.}$

2. Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are each unbiased estimators of θ, with $V(\hat{\theta}_1) = \sigma_1^2$ and $V(\hat{\theta}_2) = \sigma_2^2$. A new unbiased estimator (verify this) for θ can be formed by

$$\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2$$

$(0 \le a \le 1)$. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, how should a be chosen so as to minimize $Var(\hat{\theta}_3)$?

If $\hat{\theta}_1$ and $\hat{\theta}_2$ is an UE of $\theta \Leftrightarrow E(\hat{\theta}_1) = \theta$ and $E(\hat{\theta}_2) = \theta$

$\hat{\theta}_3$ is an UE of $\theta \Leftrightarrow E(\hat{\theta}_3) = \theta$.

$$E(\hat{\theta}_3) \overset{?}{=} \theta \Rightarrow E(\hat{\theta}_3) = E(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) = a\theta + (1-a)\theta = \theta \Rightarrow E(\hat{\theta}_3) = \theta. \text{ (verified)} //$$

$$V(\hat{\theta}_3) = V(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) \xRightarrow{\hat{\theta}_1 \text{ and } \hat{\theta}_2 \text{ are ind.}} V(\hat{\theta}_3) = a^2 V(\hat{\theta}_1) + (1-a)^2 V(\hat{\theta}_2) = a^2(\sigma_1)^2 + (1-a)^2(\sigma_2)^2$$

$$= a^2(\sigma_1)^2 + (1-2a+a^2)(\sigma_2)^2 = a^2(\sigma_1)^2 + a^2(\sigma_2)^2 - 2a(\sigma_2^2) + (\sigma_2)^2$$

To minimize $Var(\hat{\theta}_3)$, we need to find the critical point with respect to independent variable, a.

$$\Rightarrow 0 = \frac{\partial}{\partial a} var(\hat{\theta}_3) = 2a\sigma_1^2 + 2a\sigma_2^2 - 2\sigma_2^2 \Rightarrow$$

$$a\sigma_1^2 + a\sigma_2^2 = \sigma_2^2 \Rightarrow \boxed{a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}$$

3. An elevator is used in a factory to carry product boxes upstairs. These boxes have a mean of 60 kg and a variance of 121. Each time elevator is filled by 50 boxes The elevator can lift up safely if the total weight no more than 3100 kg. What is the probability that elevator lifts safely in a specific loading?

X_i: weight of i-th box. $\sigma^2 = 121 \Rightarrow \sigma = 11$

$$P\left(\sum_{i=1}^{50} X_i < 3100\right) = P\left(\bar{X} < \frac{3100}{50}\right) = P(\bar{X} < 62) = ?$$

since $n = 50 > 30$, sample mean $\bar{X}$ is approximately normally distributed with mean μ and variance $\frac{\sigma^2}{n}$, then $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is app. standart normal

by the CLT $\Rightarrow P(\bar{X} < 62) \sim P\left(Z < \frac{62-60}{11/\sqrt{50}}\right)$

$= P(Z < 1.2856) = F(1.2856) \approx \boxed{0.9015} \Rightarrow$

Elevator lifts safely in a specific loading by app. 90.15 %.

4. Let $X_1, X_2, \ldots, X_n$ be a random sample of a random variable X, which has Gamma distribution with parameters $\alpha = 3$ and β.

a) Find maximum likelihood estimator (MLE) of β.

b) The weekly downtime X (in hours) for a certain industrial machine has Gamma distribution with parameters $\alpha = 3$ and β. The loss, in dollars, to the industrial operation as a result of this downtime is given by

$$L = 30X + 2X^2$$

Find MLE of expected lost.

pdf of Gamma dist., $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$ $x>0$

where $\Gamma(\alpha) = (\alpha-1)!$, $\alpha = 3$. $\Rightarrow f(x) = \frac{1}{2.\beta^3} \cdot x^2 e^{-\frac{x}{\beta}}$ $x>0$

$X \sim$ gamma $(\alpha = 3, \beta)$

a) Likelihood function, $L_X(\beta) = \prod_{i=1}^{n} \frac{1}{2\beta^3} x_i^2 e^{-\frac{x_i}{\beta}}$.

Then logarithmic likelihood function, $K_X(\beta) = \ln L_X(\beta)$

$\Rightarrow K_X(\beta) = \ln\left[\prod_{i=1}^{n} \frac{1}{2\beta^3} x_i^2 e^{-\frac{x_i}{\beta}}\right] = \sum_{i=1}^{n} \ln\left(\frac{1}{2\beta^3} x_i^2 e^{-\frac{x_i}{\beta}}\right)$

$= \sum_{i=1}^{n}\left(\ln\frac{1}{2\beta^3} + \ln x_i^2 - \frac{x_i}{\beta}\right) = n\ln\frac{1}{2\beta^3} + 2\sum_{i=1}^{n}\ln x_i - \frac{\sum x_i}{\beta}$

$= n(\ln 1 - 3\ln 2\beta) + 2\sum_{i=1}^{n}\ln x_i - \frac{\sum x_i}{\beta} = -3n\ln 2\beta + 2\sum_{i=1}^{n}\ln x_i - \frac{\sum x_i}{\beta}$

To find MLE of β $\Rightarrow$ $\frac{d K_X(\beta)}{d\beta} = 0 \Rightarrow -\frac{3n}{\beta} + \frac{\sum_{i=1}^{n} x_i}{\beta^2} = 0$

$\Rightarrow \frac{3n}{\beta} = \frac{\sum_{i=1}^{n} x_i}{\beta^2} \Rightarrow 3n = \frac{\sum_{i=1}^{n} x_i}{\beta} \Rightarrow$ MLE of β, $\boxed{\hat{\beta} = \frac{\sum x_i}{3n} = \frac{\bar{x}}{3}}$

b) By the invariant property of MLE's,

MLE of $E(L) = E(30X + 2X^2)$

$= 30E(X) + 2E(X^2)$

$X \sim$ Gamma dist $(\alpha = 3, \beta)$

For gamma dist. $E(x) = \alpha\beta = 3\beta$ and $Var(x) = \alpha\beta^2 = 3\beta^2$

$E(x^2) = Var(x) + [E(x)]^2 \Rightarrow E(x^2) = 3\beta^2 + 9\beta^2 = 12\beta^2$

$\Rightarrow E(L) = 30.(3\beta) + 2.(12\beta^2) = 90\beta + 24\beta^2$

MLE of $E(L)$ is $E(\hat{L}) = 24(\hat{\beta})^2 + 90\hat{\beta}$

$$= 24 \cdot \frac{\bar{X}^2}{9} + 90\frac{\bar{X}}{3} = \boxed{\frac{8\bar{X}^2}{3} + 30\bar{X} = E(\hat{L})}$$

5. Let $X_1, X_2, \ldots X_n$ be a random sample from a distribution that is uniform on $(0, 2\theta)$, where θ is unknown parameter of the distribution.

a) Is $\bar{X}$ an unbiased estimator of θ ? Why ?

b) Construct a 98% large-sample confidence interval for θ.

a) $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and X_i's are independent identically distributed rv, $i = 1 \ldots n \Rightarrow E(X_i) = \mu \Rightarrow E(\bar{X}) = \frac{n\mu}{n} = \mu$.

Dist. is uniform on $(0, 2\theta) \Rightarrow E(\bar{X}) = \mu = \theta \Rightarrow \bar{x}$ is an UE of θ.

b) $\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}$ is a confidence interval

$1 - \alpha = 0.98 \Rightarrow \alpha = 0.02 \Rightarrow \frac{\alpha}{2} = 0.01 \Rightarrow z_{0.01} = 2.326$

Dist. is uniform $\Rightarrow Var(X) = \frac{(2\theta - 0)^2}{12}$, $\sigma = \sqrt{Var(x)}$

$\Rightarrow \sigma = \sqrt{\frac{4\theta^2}{12}} = \frac{\theta}{\sqrt{3}}$

$\Rightarrow E = \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} = \frac{\theta}{\sqrt{3}\sqrt{n}} \cdot (2.326) \approx \frac{1.34 \cdot \theta}{\sqrt{n}}$

Dist. is uniform $\Rightarrow \bar{x} = \mu = \theta$

$\Rightarrow$ 98% large-sample confidence interval for θ is

$$\left(\theta - \frac{1.34\theta}{\sqrt{n}}, \theta + \frac{1.34\theta}{\sqrt{n}}\right) = \boxed{\left(\frac{(\sqrt{n} - 1.34)\theta}{\sqrt{n}}, \frac{(\sqrt{n} + 1.34)\theta}{\sqrt{n}}\right)}$$

6. To estimate the average time it takes to assemble a certain computer component, the engineer at an electronics firm timed 40 technicians in the performance of this task, getting a mean of 12.73 minutes and standard deviation of 2.06 minutes. Estimate the true average time it takes to assembly the computer component with a 98% confidence interval and interpret it.

$n = 40$, $\bar{x} = 12.73$, $s = 2.06$, $1-\alpha = 0.98 \Rightarrow \frac{\alpha}{2} = 0.01$

Assuming normality of assembly times, 98% CI for the true average of assembly times

$$\bar{x} \pm \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}} = (12.73) \pm \frac{2.06}{\sqrt{40}} \cdot z_{0.01}$$

$$= (12.73) \pm \frac{2.06}{\sqrt{40}} \cdot (2.326) \approx (12.73) \pm 0.7576$$

or $(11.9724, 13.4876)$

$\boxed{\text{We are 98\% confident that assembly time of a certain computer computer will approximately be in the range between (11.9724, 13.4876) minutes.}}$

7. An interval estimate will be done for the average gain in a circuit on a semiconductor device. Assume that gain is normally distributed with a standard deviation of 25. How large must sample size be if the length of the 99% CI is to be 30?

σ is known $\Rightarrow z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ and $100(1-\alpha)\%$ CI is $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ where $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is maximum sampling error.

Length of the 99% CI $= 30 \Rightarrow 2\, z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 30$

$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005 \Rightarrow z_{0.005} = 2.576$

$2 . z_{0.005} \frac{25}{\sqrt{n}} = 30 \Rightarrow \dfrac{(2.576) . 25}{\sqrt{n}} = 15 \Rightarrow n = \left(\dfrac{(2.576) . 25}{15}\right)^2 = 18.69$

$\Rightarrow$ Sample size should be at most 18 to have 99% CI with length 30.