ECI-114 Probabilistic Systems Analysis for Civil Engineers, 2024, Summer I
Department of Civil and Environmental Engineering
University of California Davis

Instructor: Yanlin Qi, ylqi@ucdavis.edu
Reader: Kanglong Wu, klowu@ucdavis.edu

Homework due: **July 30th, 2024**

Homework # 4: Estimators (100 pts)

The objective of this assignment is to familiarize yourself with the concepts discussed in class.
Remember to be as explicit as possible in your development of the solutions.

Note: Please be organized and clear. If we cannot understand your work, we will grade it accordingly. For the problems that you only provide a solution without the process we will mark down the grade.

For this assignment you need to:

1. 20 points. Suppose that a random variable X represents the output of a civil engineering process, and that X is uniformly distributed. The pdf of X is equal to 1 for any positive x smaller than or equal to 2, and it is 0 otherwise. If you take a random sample of 12 observations, what is the approximate probability distribution of $\overline{X} - 10$? (You need to find the mean and variance of this quantity and state your assumptions.)

   Hint:
   - **Determine the mean and variance of the uniform distribution X.**
   - **Use the Central Limit Theorem to approximate the distribution of the sample mean for 12 observations.**

   Denote the normal distribution with the mean μ and variance σ^2 as $\mathrm{N}(\mu, \sigma^2)$.

$$\mu = \mathrm{E(X)} = \frac{\mathrm{a+b}}{2} = \frac{0+2}{2} = 1$$
$$\sigma^2 = V(X) = \frac{(b-a)^2}{12} = \frac{(2-0)^2}{12} = \frac{1}{3}$$
$$\mu_{\bar{X}} = \mu = 1$$
$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{1}{3 \times 12} = \frac{1}{36}$$

   So by using CTL, $\overline{\mathrm{X}} \sim \mathrm{N}\left(\mu_{\overline{\mathrm{X}}}, \sigma_{\overline{\mathrm{X}}}^2\right)$ approximately, which is $\overline{\mathrm{X}} \sim \mathrm{N}\left(1, \frac{1}{3}\right)$.
   Approximately, $\mathrm{Y} = \bar{X} - 10 \sim N\left(1 - 10, \frac{1}{36}\right)$, or $\bar{X} - 10 \sim \mathrm{N}\left(-9, \frac{1}{36}\right)$

2. 10 points. The heights (in inches) of 15 students from the entire UC Davis student body are sampled. The standard deviation for height in the entire UC Davis student body is known to be 3.
   (a) What is the variance of the sample mean?

3. 20 points, 10 each. For two independent random variables with mean μ and variance σ^2, we have the following estimators of μ:

$$\hat{\Theta}_1 = \frac{X_1 + X_2}{2} \quad \text{and} \quad \hat{\Theta}_2 = \frac{X_1 + 2X_2}{4}$$

a) Find out if they are unbiased estimators of μ?

b) Find their variances, and determine which one is better in terms of the relative efficiency.

Hint:

- **To check if an estimator is unbiased, calculate its expected value and see if it equals <u>the population mean</u>.**
- **For the variances, use the properties of variances for <u>linear combinations</u> of independent random variables.**
- **Compare** $\mathrm{MSE}(\hat{\Theta}) = V(\hat{\theta}) + (\mathrm{bias})^2$ **since one is not an unbiased estimator.**

a)

$$E(\hat{\Theta}_1) = \mathrm{E}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}[\mathrm{E}(X_1) + \mathrm{E}(X_2)] = \frac{1}{2}[\mu + \mu] = \mu$$

∴ $\hat{\Theta}_1$ is unbiased estimator of μ

$$E(\hat{\Theta}_2) = \mathrm{E}\left(\frac{X_1 + 2X_2}{4}\right) = \frac{1}{4}[\mathrm{E}(X_1) + 2\mathrm{E}(X_2)] = \frac{1}{4}[\mu + 2\mu] = \frac{3}{4}\mu$$

∴ $\hat{\Theta}_2$ is not an unbiased estimator of μ. Bias $= E(\hat{\Theta}_2) - \mu = -1/4\,\mu$

b) **(Previous hint about this question is wrong. Award full marks for question 3.b)**

$$V(\hat{\Theta}_1) = \mathrm{V}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}[\mathrm{V}(X_1) + \mathrm{V}(X_2)] = \frac{1}{4}[\sigma^2 + \sigma^2] = \frac{\sigma^2}{2}$$

$$V(\hat{\Theta}_2) = \mathrm{V}\left(\frac{X_1 + 2X_2}{4}\right) = \frac{1}{16}[\mathrm{V}(X_1) + 2^2\mathrm{V}(X_2)] = \frac{1}{16}[\sigma^2 + 4\sigma^2] = \frac{5\sigma^2}{16}$$

The first estimator is unbiased, while the second one is not unbiased.

The mean squared error (MSE) included both of the variance of the estimator and the bias as $\mathrm{MSE}(\hat{\Theta}) = V(\hat{\Theta}) + (\mathrm{bias})^2$.

$$\mathrm{MSE}(\hat{\Theta}_1) = V(\hat{\Theta}_1) = \frac{\sigma^2}{2}$$

$$\mathrm{MSE}(\hat{\Theta}_2) = V(\hat{\Theta}_2) = \frac{5\sigma^2}{16} + \left(-\frac{1}{4}\mu\right)^2 = \frac{5\sigma^2}{16} + \frac{\mu^2}{16}$$

If $3\sigma^2 < \mu^2$,

$\frac{\mathrm{MSE}(\hat{\Theta}_1)}{\mathrm{MSE}(\hat{\Theta}_2)} < 1$, and the first estimator is better.

If $3\sigma^2 > \mu^2$,

$\frac{\mathrm{MSE}(\hat{\Theta}_1)}{\mathrm{MSE}(\hat{\Theta}_2)} > 1$, and the second estimator is better.

If $3\sigma^2 = \mu^2$,

$\frac{\mathrm{MSE}(\hat{\Theta}_1)}{\mathrm{MSE}(\hat{\Theta}_2)} = 1$, both estimators are equally good in terms of relative efficiency.

4. 30 points, 10 points each. Measurements of a critical dimension on a sample of automotive driveshafts are taken with the following results (in units of mm):

10.14	10.4	9.87	10.02	9.99	10.03
10.09	10.11	9.93	9.99	10	10.02

The intended μ is 10, but is known to drift from this mean during production. The measurement follows a normal distribution.

(a) Assume that the standard deviation for the production process is known as $\sigma = 0.10$ mm. Calculate the 95% confidence interval for the mean for the production process.

(b) Assume that the standard deviation for the production process is not known. Calculate the 95% confidence interval for the mean.

(c) Find β (the probability of a Type II error) for the case where a 95% confidence interval is used, the true mean 10.2, and σ is known.

Hint:

- **For known σ, use the Z-score to calculate the confidence interval.**
- **For unknown σ, use the t-distribution with the appropriate degrees of freedom.**
- **For the Type II error, calculate the probability of not rejecting the null hypothesis when the true mean is 10.2.**

(a)

Use: $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$,

$\bar{x}$ =(10.14+10.4+9.87+10.02+9.99+10.03+10.09+10.11+9.93+9.99+10+10.02)/12 = 10.049.

$10.049 - z_{0.025}\ 0.1/12^{1/2} \le \mu \le 10.049 + z_{0.025}\ 0.1/12^{1/2}$

$10.049-1.96(0.1)/3.464 \le \mu \le 10.049+1.96(0.1)/3.464$

$9.992 \le \mu \le 10.106$,

so the 95% C.I = [9.992,10.106].

(rounded answers, such as [9.99, 10.11] can be accepted.)

(b)

Here we need to calculate the S^2 before we can use the t-distribution to solve the problem.

$S^2 = (1212.024 - 12*(10.049)^2)/(11) = 0.0177$, $S = (0.0177)^{1/2} = 0.13297$

Use: $\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$, where n = 12.

$10.049 - t_{0.025,\,11}(0.13297)/(12^{1/2}) \le \mu \le 10.049 + t_{0.025,\,11}(0.13297)/(12^{1/2})$

$t_{0.025,\,11} = 2.201$

$9.9645 \le \mu \le 10.134$,

so the 95% CI = [9.9645, 10.134]

(rounded answers can be accepted.)

(c) (using the rounded CI from (a))

β = P(Type II) = P(Fail to Reject H_0 | H_0 is False) = P[$9.992 \le \bar{x} \le 10.106$]

$$\beta = \Pr\left[\frac{9.992 - 10.2}{\frac{0.1}{\sqrt{12}}} \le Z \le \frac{10.106 - 10.2}{\frac{0.1}{\sqrt{12}}}\right] = \Pr[-7.20533 \le Z \le -3.25626]$$
$$= \Pr[Z \le -3.25626] - \Pr[Z \le -7.20533] = 0.00039 - 0 = 0.04\%$$

5. 20 points. Suppose that X is a discrete uniform random variable, with pmf equal to ¼ for values of x = 1, 2, 3, and 4. Determine the probability that the sample mean is greater than 1 but less than 3, if you select a random sample of size n = 40.

Hint:

- **Calculate the mean and variance of the discrete uniform distribution.**
- **Use the Central Limit Theorem to approximate the distribution of the sample mean.**

- **Convert the range of interest into Z-scores and use the standard normal distribution to find the probability.**

X follows a discrete uniform distribution, so we can get the mean and variance of X.

$$\mu_X = \frac{a+b}{2} = \frac{(4+1)}{2} = 2.5$$

$$\sigma_X = \sqrt{\frac{(b-a+1)^2-1}{12}} = \sqrt{\frac{(4-1+1)^2-1}{12}} = \sqrt{\frac{15}{12}} = \sqrt{\frac{5}{4}}$$

For the sample mean, using central limit theorem (CTL) we know $\bar{X} \sim N\left(\mu_{\bar{X}}, \sigma_{\bar{X}}^2\right)$, where

$\mu_{\bar{X}} = 2.5, \quad \sigma_{\bar{X}} = \sigma/\sqrt{n} = \frac{\sqrt{5/4}}{\sqrt{40}} = \sqrt{\frac{1}{32}}.$

Standardize to the standard normal distribution by $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

$$P(1 < \bar{X} < 3) = P\left(\frac{1-2.5}{\sqrt{\frac{1}{32}}} < Z < \frac{3-2.5}{\sqrt{\frac{1}{32}}}\right) = \phi(2.828) - \phi(-8.485) = 0.9977 - 0 = 0.9977$$

Estimate the Z-score, and find the probabilities using the Normal table.