Quiz

Let $Z_1, \dots, Z_7$ be a random sample from the standard normal distribution.
Let $W = Z_1^2 + \dots + Z_7^2$.

(a) Use CLT to estimate $\mathsf{P}(1.69 < W < 14.07)$.

(b) Find exact value of that probability.

Solution:

(a) The CLT states, that if $X_1, \dots, X_n$ are i.i.d. variables with $\mathsf{V}(X) = \sigma^2 > 0$ (where X is a variable with common distribution for X_i), then sum $S_n = X_1 + \ldots + X_n$ in standardized form Z_n:

$$Z_n = \frac{S_n - \mathsf{E}(S_n)}{\sigma_{S_n}} = \frac{S_n - n \cdot \mathsf{E}(X)}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal one:

$$Z_n \xrightarrow[n \to \infty]{d} Z \sim \mathcal{N}(0, 1).$$

In our case X_i are Z_i^2 with $n = 7$. We know that the sum $\sum_{i=1}^{n} Z_i^2$ is a random variable W with χ^2-distribution with number of degrees of freedom n. Moments of W are: $\mathsf{E}(W) = n$, $\mathsf{V}(W) = 2n$. Another approach is to use identity of Z_i^2, each of them is distributed as χ_1^2, which means that $\mathsf{E}(Z_i^2) = 1$ and $\mathsf{V}(Z_i^2) = 2$.

Anyhow, using standard normal distribution table and symmetry of Z, the required probability then is:

$$\begin{aligned}\mathsf{P}(1.69 < W < 14.07) &= \mathsf{P}\left(\frac{1.69 - 7 \cdot 1}{\sqrt{7} \cdot \sqrt{2}} < \frac{W - 7 \cdot 1}{\sqrt{7} \cdot \sqrt{2}} < \frac{14.07 - 7 \cdot 1}{\sqrt{7} \cdot \sqrt{2}}\right) \approx \\ &\approx \mathsf{P}\left(-\frac{5.31}{\sqrt{14}} < Z < \frac{7.07}{\sqrt{14}}\right) \approx \Phi(1.890) - \Phi(-1.418) = \\ &= \Phi(1.890) + \Phi(1.418) - 1 \approx 0.971 + 0.922 - 1 = \boxed{0.893}.\end{aligned}$$

(b) $W \sim \chi_7^2$. Using tables of χ^2-distribution:

$$\begin{aligned}\mathsf{P}(1.69 < W < 14.07) &= \mathsf{P}(W < 14.07) - \mathsf{P}(W \leq 1.69) = \\ &= \mathsf{P}(W > 1.69) - \mathsf{P}(W \geq 14.07) = 0.975 - 0.05 = \boxed{0.925}.\end{aligned}$$

Problem 1

Suppose $X_1, X_2, \ldots, X_{10}$ is a random sample taken from an $\mathcal{N}(12, 25)$-distributed population.

(a) Find the probability that the sample variance S^2 is between 20 and 30.

(b) Find the range for the middle 90% of the distribution of the sample variance.

Solution:

Fisher's lemma states that if $X_1, \ldots, X_n$ is a sample of i.i.d. random variables from $\mathcal{N}\left(\mu, \sigma^2\right)$ then

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

where $S^2 = \dfrac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)$ is an unbiased sample variance, and $\overline{X} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$ is a sample mean.

Multiplier $n-1$ instead of more obvious n in the denominator of S^2 definition is due to the fact that a sample variance is calculated from the vector of residuals $X_i - \overline{X}$:

$$\begin{pmatrix} X_1 - \overline{X} & X_2 - \overline{X} & \ldots & X_n - \overline{X} \end{pmatrix}.$$

While there are n independent observations in the sample, there are only $n-1$ independent residuals, as they sum to 0. Thus, there are $n-1$ degrees of freedom in S^2, and consecutively, in χ^2-distribution of S^2.

Also, if we tried to estimate variance with $\dfrac{1}{n} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)$ by the sample, which contains only one observation ($n = 1$), we would always get 0 for any population and any sample. In reality variance can't be estimated by the sample with one observation, because there's simply no variability. In the case of unbiased S^2 sample variance does not exist, which fully corresponds to the nature of variance.

(a) Applying Fisher's lemma to sample with $n = 10$ observations with $\mathcal{N}(15, 25)$ distribution:

$$\frac{9S^2}{25} \sim \chi^2_9.$$

Required probability then:

$$\mathsf{P}\left(20 < S^2 < 30\right) = \mathsf{P}\left(\frac{9}{25} \cdot 20 < \frac{9S^2}{25} < \frac{9}{25} \cdot 30\right) = \mathsf{P}\left(7.2 < \chi^2_9 < 10.8\right) \approx \boxed{0.327}.$$

Here we used simulations in Python directly, since χ^2_9-table does not provide probabilities for values 7.2 and 10.8 within good accuracy.

(b) Middle 90% of the distribution cut off left and right tails of 5% each. Thus, we need to find values s_1^2 and s_2^2 of sample variance, which correspond to percentiles $\chi^2_{9;\,0.95}$ and $\chi^2_{9;\,0.05}$, where the second index defines the probability of right tail. Illustration is given in the fig. 1.

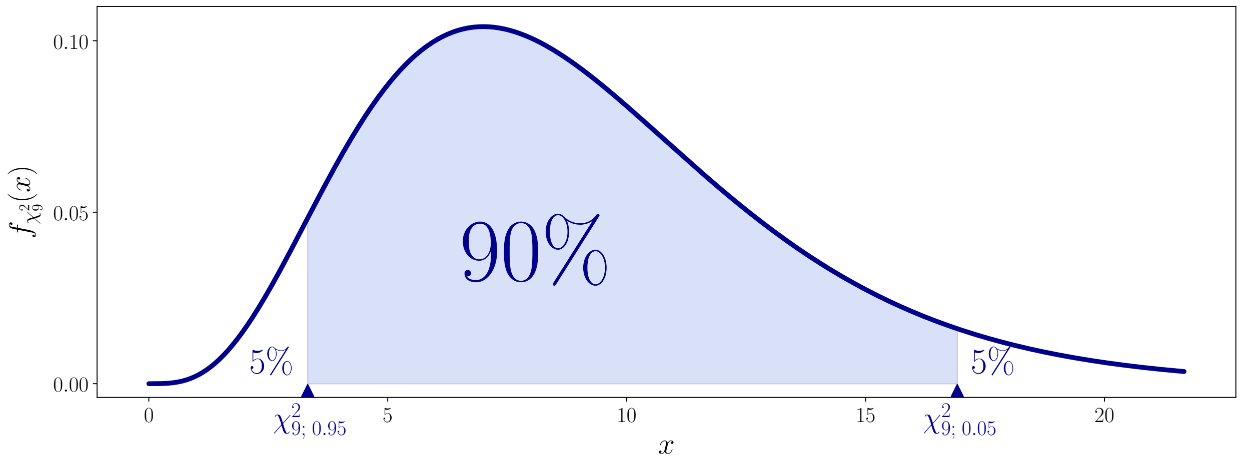

Figure 1: Middle 90% of χ^2_9-distribution.

Applying Fisher's lemma, the probability is:

$$\mathsf{P}\left(s_1^2 < S^2 < s_2^2\right) = \mathsf{P}\left(\frac{9s_1^2}{25} < \chi^2_9 < \frac{9s_2^2}{25}\right),$$

which means that

$$\begin{cases} \dfrac{9s_1^2}{25} = \chi^2_{9;\,0.95}, \\ \dfrac{9s_2^2}{25} = \chi^2_{9;\,0.05}, \end{cases} \implies \begin{cases} s_1^2 = \dfrac{25}{9}\cdot\chi^2_{9;\,0.95}, \\ s_2^2 = \dfrac{25}{9}\cdot\chi^2_{9;\,0.05}. \end{cases}$$

Using tables of χ^2-distribution:

$$\begin{cases} s_1^2 \approx \dfrac{25}{9}\cdot 3.33, \\ s_2^2 \approx \dfrac{25}{9}\cdot 16.92, \end{cases} \implies \boxed{\begin{cases} s_1^2 \approx 9.25, \\ s_2^2 \approx 47. \end{cases}}$$

Problem 2

An ordinary die is "fair" or "balanced" if each face has an equal chance of landing on top when the die is rolled. Thus the proportion of times a three is observed in a large number of tosses is expected to be close to 1/6. Suppose a die is rolled 240 times and shows three on top 36 times.

(a) Find the probability that a fair die would produce a proportion of 0.15 or less.

(b) Give an interpretation of the result in part (a). How strong is the evidence that the die is not fair?

(c) Suppose the sample proportion 0.15 came from rolling the die 2,400 times instead of only 240 times. Rework part (a) under these circumstances.

(d) Give an interpretation of the result in part (c). How strong is the evidence that the die is not fair?

Solution:

Let R be the number of successes in n trials. We know that probability of getting exactly R successes with probability of success p subjects to binomial distribution $\mathrm{Bin}(n, p)$. Probability p can be estimated with statistic $\widehat{P} = \dfrac{R}{n}$, which is called sample proportion.

Since we know expected value and variance of $R \sim \mathrm{Bin}(n, p)$: $\mathsf{E}(R) = np$ and $\mathsf{V}(R) = np(1-p)$, we can find those moments for $\widehat{P}$:

$$\mathsf{E}\left(\widehat{P}\right) = \mathsf{E}\left(\frac{R}{n}\right) = \frac{\mathsf{E}\left(R\right)}{n} = \frac{np}{n} = p,$$
$$\mathsf{V}\left(\widehat{P}\right) = \mathsf{V}\left(\frac{R}{n}\right) = \frac{\mathsf{V}\left(R\right)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

We could assume that $\widehat{P}$ has binomial distribution, since it is scaled binomial R. But $\widehat{P}$ is not binomial, at least because binomial distribution can take only integer values.

As an extension on Moivre-Laplace theorem (or another particular case of the Central Limit Theorem), when n is large enough:

$$\frac{\widehat{P} - p}{\sqrt{p(1-p)/n}} \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}\left(0, 1\right).$$

Derivation is identical to that of Moivre-Laplace, but for averages. If $X_1, \ldots, X_n$ are n i.i.d. Bernoulli variables with probability of success p, then average $\widehat{P} = \dfrac{S_n}{n} = \dfrac{X_1 + \ldots + X_n}{n}$ in standardized form Z_n:

$$Z_n = \frac{\frac{S_n}{n} - \mathsf{E}\left(\frac{S_n}{n}\right)}{\sigma_{\frac{S_n}{n}}} = \frac{\frac{S_n}{n} - \mathsf{E}(X)}{\sigma_X/\sqrt{n}} = \frac{\frac{S_n}{n} - p}{\sqrt{p(1-p)}/\sqrt{n}} \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}(0,1).$$

The best way to define, when we can use normal approximation for $\widehat{P}$ is via its physical meaning. Since $\widehat{P}$ estimates probability p, its limits should be $[0,1]$. So, we want all possible outcomes $\widehat{p}$ of variable $\widehat{P}$ to lie within $[0,1]$. This is achieved if

$$\left[\mathsf{E}\left(\widehat{P}\right) - 3\sigma_{\widehat{P}},\ \mathsf{E}\left(\widehat{P}\right) + 3\sigma_{\widehat{P}}\right] = \left[p - 3\sqrt{\frac{p(1-p)}{n}},\ p + 3\sqrt{\frac{p(1-p)}{n}}\right] \subset [0,1]$$

by 3 sigma rule – 99.7% of probability lies inside 3σ range.

(a) Before we use normal approximation to calculate the probability, we should check that this approximation is viable. For $n = 240$ and $p = \frac{1}{6}$:

$$\left[\frac{1}{6} - 3\cdot\sqrt{\frac{\frac{1}{6}\cdot\frac{5}{6}}{240}},\ \frac{1}{6} + 3\cdot\sqrt{\frac{\frac{1}{6}\cdot\frac{5}{6}}{240}}\right] \approx [0.09,\ 0.24] \subset [0,1],$$

thus, we can use normal approximation. Using standardization and standard normal distribution table:

$$\mathsf{P}\left(\widehat{P} \le 0.15\right) = \mathsf{P}\left(Z \le \frac{0.15 - 1/6}{0.024}\right) \approx \Phi(-0.694) = 1 - \Phi(0.694) \approx 1 - 0.755 = \boxed{0.245}.$$

(b) We have 24.5% chance to get proportion $\widehat{p} = 0.15$ or less after 240 die rolls if the true success probability is $\frac{1}{6}$. Since 24.5% is sufficiently large, this event is probable and the evidence of die not being fair is $\boxed{\text{NOT strong enough}}$.

(c) Increasing number of rolls n, we decrease the variance of $\widehat{P}$, thus $[p - 3\sigma_{\widehat{P}},\ p + 3\sigma_{\widehat{P}}]$ for $n = 2400$ will be narrower than for $n = 240$ in vicinity of $\frac{1}{6}$ and definitely in range of $[0,1]$. New $\sigma_{\widehat{P}}$ is ≈ 0.0076:

$$\mathsf{P}\left(\widehat{P} \le 0.15\right) = \mathsf{P}\left(Z \le \frac{0.15 - 1/6}{0.0076}\right) \approx \Phi(-2.193) = 1 - \Phi(2.193) \approx 1 - 0.986 = \boxed{0.014}.$$

(d) Now it is only 1.4% to get proportion $\widehat{p} = 0.15$ or less. Such event is highly improbable and the evidence of die not being fair is $\boxed{\text{strong enough}}$.

Problem 3

Random variable X assumes values 0 and 1, each with probability $1/2$.

(a) Find population mean μ and variance σ^2.

(b) You have 9 independent observations of X: $\{X_1, \dots, X_9\}$. Consider the following estimators of the population mean μ:

i) $\widehat{\mu}_1 = 0.45$;

ii) $\widehat{\mu}_2 = X_1$;

iii) $\widehat{\mu}_3 = \overline{X}$;

iv) $\widehat{\mu}_4 = X_1 + \frac{1}{3}X_2$;

v) $\widehat{\mu}_5 = \frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3$.

Which of these estimators are unbiased? Calculate bias for each estimator. Which estimator is the most efficient?

(c) Which estimators from part (b) are consistent?

Solution:

(a) X is Bernoulli variable with probability of success $p = \frac{1}{2}$. For $X \sim \text{Bernoulli}(p)$ mean and variance are known: $\mathsf{E}(X) = p$, $\mathsf{V}(X) = p(1-p)$. Thus, in our case:

$$\mu = \mathsf{E}(X) = \boxed{\frac{1}{2}}, \qquad \sigma^2 = \mathsf{V}(X) = \frac{1}{2} \cdot \frac{1}{2} = \boxed{\frac{1}{4}}.$$

(b) The bias of an estimator is an average difference between estimator $\widehat{\mu}$ and true parameter value μ:

$$\text{Bias}(\widehat{\mu}) = \mathsf{E}(\widehat{\mu} - \mu) = \mathsf{E}(\widehat{\mu}) - \mu.$$

Clearly, estimator is considered to be unbiased if its bias is 0, or $\mathsf{E}(\widehat{\mu}) = \mu$.

Let's calculate biases for given estimators:

i) $\text{Bias}(\widehat{\mu}_1) = \mathsf{E}(0.45) - \frac{1}{2} = \boxed{-0.05}$. Negatively biased (underestimates parameter).

ii) $\text{Bias}(\widehat{\mu}_2) = \mathsf{E}(X_1) - \frac{1}{2} = \frac{1}{2} - \frac{1}{2} = \boxed{0}$. Unbiased.

iii) $\text{Bias}(\widehat{\mu}_3) = \mathsf{E}(\overline{X}) - \frac{1}{2} = \frac{1}{2} - \frac{1}{2} = \boxed{0}$ $\left(\mathsf{E}(\overline{X}) = \mu\right)$. Unbiased.

iv) $\text{Bias}(\widehat{\mu}_4) = \mathsf{E}\left(X_1 + \frac{1}{3}X_2\right) - \frac{1}{2} = \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} - \frac{1}{2} = \boxed{\frac{1}{6}}$. Positively biased (overestimates parameter).

v) $\text{Bias}\left(\widehat{\mu}_5\right) = \mathsf{E}\left(\frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3\right) - \frac{1}{2} = \frac{2}{3}\cdot\frac{1}{2} + \frac{2}{3}\cdot\frac{1}{2} - \frac{1}{3}\cdot\frac{1}{2} - \frac{1}{2} = \boxed{0}$. Unbiased.

Let's compare efficiency of estimators in terms of Mean-Squared Errors (MSE). MSE of estimator $\widehat{\mu}$ is defined as

$$\text{MSE}\left(\widehat{\mu}\right) = \mathsf{E}\left(\left(\widehat{\mu} - \mu\right)^2\right),$$

which can be expanded via variance-mean identity $\mathsf{E}\left(Y^2\right) = \mathsf{E}(Y)^2 + \mathsf{V}(Y)$ as

$$\text{MSE}\left(\widehat{\mu}\right) = \mathsf{E}\left(\widehat{\mu} - \mu\right)^2 + \mathsf{V}\left(\widehat{\mu} - \mu\right) = \text{Bias}^2(\widehat{\mu}) + \mathsf{V}\left(\widehat{\mu}\right).$$

MSE of given estimators are following:

i) $\text{MSE}\left(\widehat{\mu}_1\right) = \text{Bias}^2(0.45) + \mathsf{V}\left(0.45\right) = (-0.05)^2 + 0 = \boxed{0.0025}$.

ii) $\text{MSE}\left(\widehat{\mu}_2\right) = \text{Bias}^2(X_1) + \mathsf{V}\left(X_1\right) = 0^2 + \frac{1}{4} = \boxed{\frac{1}{4}}$.

iii) $\text{MSE}\left(\widehat{\mu}_3\right) = \text{Bias}^2\left(\overline{X}\right) + \mathsf{V}\left(\overline{X}\right) = 0^2 + \frac{1}{9}\cdot\frac{1}{4} = \boxed{\frac{1}{36}}\ \left(\mathsf{V}\left(\overline{X}\right) = \frac{\sigma^2}{n}\right)$.

iv) $\text{MSE}\left(\widehat{\mu}_4\right) = \text{Bias}^2\left(X_1 + \frac{1}{3}X_2\right) + \mathsf{V}\left(X_1 + \frac{1}{3}X_2\right) = \left(\frac{1}{6}\right)^2 + \left(\frac{1}{4} + \left(\frac{1}{3}\right)^2\cdot\frac{1}{4}\right) =$
$\frac{1}{36} + \frac{5}{18} = \boxed{\frac{11}{36}}$.

v) $\text{MSE}\left(\widehat{\mu}_5\right) = \text{Bias}^2\left(\frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3\right) + \mathsf{V}\left(\frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3\right) =$
$0^2 + \left(\left(\frac{2}{3}\right)^2\cdot\frac{1}{4} + \left(\frac{2}{3}\right)^2\cdot\frac{1}{4} + \left(\frac{1}{3}\right)^2\cdot\frac{1}{4}\right) = \boxed{\frac{1}{4}}$.

Thus, we have $\text{MSE}\left(\widehat{\mu}_1\right) < \text{MSE}\left(\widehat{\mu}_3\right) < \text{MSE}\left(\widehat{\mu}_2\right) = \text{MSE}\left(\widehat{\mu}_5\right) < \text{MSE}\left(\widehat{\mu}_4\right)$, and estimator $\widehat{\mu}_1$ being best in terms of MSE.

We have to emphasize that with the increase of observations in sample up to $n = 100$, MSE of $\widehat{\mu}_3$ will catch up with MSE of $\widehat{\mu}_1$, and after greater increase the estimator $\widehat{\mu}_3$ will be more efficient than $\widehat{\mu}_1$ in terms of MSE.

(c) An estimator $\widehat{\theta}_n$ of the parameter θ is consistent if it converges to θ by probability:

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{\mathsf{P}} \theta,$$

or

$$\forall\varepsilon > 0: \qquad \mathsf{P}\left(\left|\widehat{\theta}_n - \theta\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

To find the sufficient condition of estimator consistency. Let's consider Markov's inequality, which has the following general formulation:

Markov's inequality. If X is a non-negative random variable and $\varepsilon > 0$, then

$$\mathsf{P}(X \geq \varepsilon) \leq \frac{\mathsf{E}(X)}{\varepsilon}.$$

Let's substitute X with $\left|\widehat{\theta}_n - \theta\right|$ and bring interior of probability to quadratic form:

$$\mathsf{P}\left(\left|\widehat{\theta}_n - \theta\right| \geq \varepsilon\right) = \mathsf{P}\left(\left(\widehat{\theta}_n - \theta\right)^2 \geq \varepsilon^2\right) \leq \frac{\mathsf{E}\left(\left(\widehat{\theta}_n - \theta\right)^2\right)}{\varepsilon^2} = \frac{\mathrm{MSE}\left(\widehat{\theta}_n\right)}{\varepsilon^2}.$$

The last inequality immediately implies the fact that

$$\mathrm{MSE}\left(\widehat{\theta}_n\right) \underset{n\to\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad \widehat{\theta}_n \xrightarrow[n\to\infty]{\mathsf{P}} \theta.$$

Thus, tendency of MSE to 0 is the sufficient condition of estimators consistency.

That last inequality is also sometimes called Chebyshev's inequality, since if $\mathsf{E}\left(\widehat{\theta}_n\right) = \theta$, Markov's inequality transforms into:

Chebyshev's inequality. If X is a random variable and with finite expected value and finite non-zero variance, then $\forall \varepsilon > 0$

$$\mathsf{P}\left(|X - \mathsf{E}(X)| > \varepsilon\right) \leq \frac{\mathsf{V}\left(X\right)}{\varepsilon^2}.$$

Overall, if inequality contains quadratic loss function, it's usually called Chebyshev's with a purpose not to confuse it with Markov's one, which uses absolute loss function.

The only consistent estimator in our problem is $\widehat{\mu}_3 = \overline{X}$, since

$$\lim_{n\to\infty} \mathrm{MSE}\left(\widehat{\mu}_3\right) = \lim_{n\to\infty} \mathsf{V}\left(\overline{X}\right) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0,$$

while variances of other estimators are constants due to limited set of observations X_i.

Problem 4

Let X_1, X_2, X_3 be a random sample from a population with mean μ and variance σ^2. Consider the following two estimators of variance σ^2:

(a) $\widehat{\sigma}_1^2 = c_1 (X_1 - X_2)^2$;

(b) $\widehat{\sigma}_2^2 = c_2 (X_1 - X_2)^2 + c_2 (X_1 - X_3)^2 + c_2 (X_2 - X_3)^2$.

Find constants c_1, c_2, such that $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are unbiased estimators of σ^2.

Solution:

For an estimator to be unbiased, its expected value should be equal to the estimated parameter: $\mathsf{E}(\widehat{\sigma}^2) = \sigma^2$.

(a) Let's find expected value of $\widehat{\sigma}_1^2$.

$$\begin{aligned}\mathsf{E}\left(\widehat{\sigma}_1^2\right) &= \mathsf{E}\left(c_1 (X_1 - X_2)^2\right) = c_1 \mathsf{E}\left(X_1^2 + X_2^2 - 2X_1X_2\right) = \\ &= c_1 \left(\mathsf{E}\left(X_1^2\right) + \mathsf{E}\left(X_2^2\right) - 2\,\mathsf{E}(X_1X_2)\right).\end{aligned}$$

From identity for variance and mean

$$\forall i \in \{1,2,3\}: \quad \mathsf{E}\left(X_i^2\right) = \mathsf{V}(X_i) + \mathsf{E}(X_i)^2 = \sigma^2 + \mu^2.$$

Due to independence of observations

$$\forall i \neq j \in \{1,2,3\}: \quad \mathsf{E}(X_iX_j) = \mathsf{E}(X_i)\,\mathsf{E}(X_j) = \mu^2.$$

Thus, expected value of $\widehat{\sigma}_1^2$:

$$\mathsf{E}\left(\widehat{\sigma}_1^2\right) = c_1\left(\sigma^2 + \mu^2 + \sigma^2 + \mu^2 - 2\mu^2\right) = 2c_1\sigma^2.$$

$\widehat{\sigma}_1^2$ will be unbiased if:

$$2c_1\sigma^2 = \sigma^2 \qquad \Longrightarrow \qquad c_1 = \boxed{\frac{1}{2}}.$$

(b) Based on derivations from previous part:

$$\forall i \neq j \in \{1,2,3\}: \quad \mathsf{E}\left((X_i - X_j)^2\right) = \mathsf{E}\left(X_i^2\right) + \mathsf{E}\left(X_j^2\right) - 2\,\mathsf{E}(X_iX_j) = 2\sigma^2.$$

Expected value of $\widehat{\sigma}_2^2$ then:

$$\mathsf{E}\left(\widehat{\sigma}_2^2\right) = c_2\mathsf{E}\left((X_1 - X_2)^2\right) + c_2\mathsf{E}\left((X_1 - X_3)^2\right) + c_2\mathsf{E}\left((X_2 - X_3)^2\right) = 6c_2\sigma^2.$$

$\widehat{\sigma}_2^2$ will be unbiased if:

$$6c_2\sigma^2 = \sigma^2 \qquad \Longrightarrow \qquad c_2 = \boxed{\frac{1}{6}}.$$

Problem 5

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(2\mu, 2\sigma^2)$. You have samples of size n and m from the two distributions: $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_m\}$. Consider the estimator $\widehat{\mu} = c_1\overline{X} + c_2\overline{Y}$.

(a) For which c_1, c_2 the estimator is unbiased?

(b) For which c_1, c_2 the estimator is unbiased and most efficient?

Solution:

(a) Since X and Y are normally distributed, sample means, derived from their population, are also normal with known parameters:

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{and} \qquad \overline{Y} \sim \mathcal{N}\left(2\mu, \frac{2\sigma^2}{m}\right).$$

Let's find expected value of $\widehat{\mu}$.

$$\mathsf{E}(\widehat{\mu}) = \mathsf{E}\left(c_1\overline{X} + c_2\overline{Y}\right) = c_1\mathsf{E}\left(\overline{X}\right) + c_2\mathsf{E}\left(\overline{Y}\right) = c_1\mu + 2c_2\mu.$$

$\widehat{\mu}$ will be unbiased if:

$$c_1\mu + 2c_2\mu = \mu \qquad \Longrightarrow \qquad \boxed{c_1 + 2c_2 = 1}.$$

(b) We have to find such c_1 and c_2 that will minimize MSE of $\widehat{\mu}$, while being unbiased. The latter condition states that $c_1 + 2c_2 = 1$ should hold:

$$c_1^\star, c_2^\star = \arg\min_{c_1, c_2} \text{MSE}(\widehat{\mu}) \quad \text{s.t. } c_1 + 2c_2 = 1.$$

The MSE is:

$$\begin{aligned}\text{MSE}(\widehat{\mu}) &= \text{Bias}^2(\widehat{\mu}) + \mathsf{V}(\widehat{\mu}) = 0^2 + \mathsf{V}\left(c_1\overline{X} + c_2\overline{Y}\right) = c_1^2\,\mathsf{V}\left(\overline{X}\right) + c_2^2\,\mathsf{V}\left(\overline{Y}\right) = \\ &= c_1^2 \cdot \frac{\sigma^2}{n} + c_2^2 \cdot \frac{2\sigma^2}{m} = \sigma^2\left(\frac{c_1^2}{n} + \frac{2c_2^2}{m}\right).\end{aligned}$$

Minimization problem can be solved via Lagrange multipliers λ, assigned to constraint of unbiasedness. Lagrangian $\mathcal{L}$ will take a form:

$$\mathcal{L} = \sigma^2\left(\frac{c_1^2}{n} + \frac{2c_2^2}{m}\right) - \lambda(c_1 + 2c_2 - 1).$$

Necessary condition of the minimum:

$$
\begin{cases}
\dfrac{\partial \mathcal{L}}{\partial c_1} = \dfrac{2c_1^\star \sigma^2}{n} - \lambda^\star = 0, \\
\dfrac{\partial \mathcal{L}}{\partial c_2} = \dfrac{4c_2^\star \sigma^2}{n} - 2\lambda^\star = 0, \\
\dfrac{\partial \mathcal{L}}{\partial \lambda} = -(c_1^\star + 2c_2^\star - 1) = 0,
\end{cases}
\implies
\begin{cases}
c_1^\star = \dfrac{\lambda^\star n}{2\sigma^2}, \\
c_2^\star = \dfrac{\lambda^\star m}{2\sigma^2}, \\
\dfrac{\lambda^\star n}{2\sigma^2} + \dfrac{\lambda^\star m}{\sigma^2} = 0,
\end{cases}
\implies
\begin{cases}
c_1^\star = \dfrac{n}{n+2m}, \\
c_2^\star = \dfrac{m}{n+2m}, \\
\lambda^\star = \dfrac{2\sigma^2}{n+2m}.
\end{cases}
$$

We should check that $c_1^\star, c_2^\star$ give indeed a minimum. By sufficient condition of the minimum, let's see if the Hessian matrix $\mathcal{H}$ is positive-definite (all minors are positive):

$$
\mathcal{H} = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial c_1^2} & \dfrac{\partial^2 \mathcal{L}}{\partial c_1 \partial c_2} \\ \dfrac{\partial^2 \mathcal{L}}{\partial c_2 \partial c_1} & \dfrac{\partial^2 \mathcal{L}}{\partial c_2^2} \end{pmatrix} = \begin{pmatrix} \dfrac{2\sigma^2}{n} & 0 \\ 0 & \dfrac{4\sigma^2}{m} \end{pmatrix} \succ 0,
$$

since σ^2, n and m are always positive (except for degenerate cases of 0 sample size).

It means that

$$
\boxed{\begin{cases} c_1^\star = \dfrac{n}{n+2m}, \\ c_2^\star = \dfrac{m}{n+2m} \end{cases}}
$$

indeed produce the most effective unbiased $\widehat{\mu}$ in terms of MSE.

Problem 6

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a $\mathcal{U}(0, \theta)$ distribution, where θ is unknown. Define the estimator

$$\widehat{\Theta}_n = \max\{X_1, X_2, X_3, \ldots, X_n\}.$$

(a) Find the bias of $\widehat{\Theta}_n$.

(b) Find the MSE of $\widehat{\Theta}_n$.

(c) Is $\widehat{\Theta}_n$ a consistent estimator of θ?

Solution:

(a) The p.d.f. f_X of random variable X, sample of which was taken, and c.d.f. F_X are following:

$$f_X(x) = \frac{1}{\theta} \cdot I_{\{0 \le x \le \theta\}}, \qquad F_X(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x}{\theta}, & 0 \le x \le \theta, \\ 1, & x > \theta. \end{cases}$$

Since $\widehat{\Theta}_n$ is the maximal value among all X_i, we can make a conclusion, that when $\widehat{\Theta}_n$ is less than some chosen number x, all X_i are also less than x (simultaneously). C.d.f. $F_{\widehat{\Theta}_n}$ then:

$$F_{\widehat{\Theta}_n}(x) = \mathsf{P}\left(\widehat{\Theta}_n \le x\right) = \mathsf{P}\left(\bigcap_{i=1}^{n} X_i \le x\right) \overset{\text{ind}}{=} \prod_{i=1}^{n} \mathsf{P}\left(X_i \le x\right) = \prod_{i=1}^{n} F_{X_i}(x) =$$
$$= \prod_{i=1}^{n} F_X(x) = \begin{cases} 0^n, & x < 0, \\ \left(\dfrac{x}{\theta}\right)^n, & 0 \le x \le \theta, \\ 1^n, & x > \theta. \end{cases}$$

P.d.f. $f_{\widehat{\Theta}_n}$ is a derivative of $F_{\widehat{\Theta}_n}$:

$$f_{\widehat{\Theta}_n}(x) = \frac{d}{dx} F_{\widehat{\Theta}_n}(x) = \frac{n x^{n-1}}{\theta^n} \cdot I_{\{0 \le x \le \theta\}}.$$

Bias of estimator $\widehat{\Theta}_n$ is the expected difference beween estimator itself and estimated parameter θ:

$$\text{Bias}\left(\widehat{\Theta}_n\right) = \mathsf{E}\left(\widehat{\Theta}_n - \theta\right) = \mathsf{E}\left(\widehat{\Theta}_n\right) - \theta.$$

Expected value of $\widehat{\Theta}_n$ by definition:

$$\mathsf{E}\left(\widehat{\Theta}_n\right) = \int\limits_{-\infty}^{\infty} x f_{\widehat{\Theta}_n}(x)dx = \int\limits_{0}^{\theta} x \cdot \frac{nx^{n-1}}{\theta^n}dx = \frac{n}{\theta^n} \cdot \left.\frac{x^{n+1}}{n+1}\right|_0^\theta = \frac{n}{n+1}\theta.$$

Bias then:

$$\text{Bias}\left(\widehat{\Theta}_n\right) = \frac{n}{n+1}\theta - \theta = \boxed{-\frac{\theta}{n+1}}.$$

It means that $\widehat{\Theta}_n$ on average underestimates θ.

(b) MSE is a sum of variance and bias squared, thus we need to calculate $\mathsf{V}\left(\widehat{\Theta}_n\right)$.

$$\mathsf{E}\left(\widehat{\Theta}_n^2\right) = \int\limits_{-\infty}^{\infty} x^2 f_{\widehat{\Theta}_n}(x)dx = \int\limits_{0}^{\theta} x^2 \cdot \frac{nx^{n-1}}{\theta^n}dx = \frac{n}{\theta^n} \cdot \left.\frac{x^{n+2}}{n+2}\right|_0^\theta = \frac{n}{n+2}\theta^2.$$

$$\mathsf{V}\left(\widehat{\Theta}_n\right) = \mathsf{E}\left(\widehat{\Theta}_n^2\right) - \mathsf{E}\left(\widehat{\Theta}_n\right)^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

MSE then:

$$\text{MSE}\left(\widehat{\Theta}_n\right) = \text{Bias}^2\left(\widehat{\Theta}_n\right) + \mathsf{V}\left(\widehat{\Theta}_n\right) = \frac{\theta^2}{(n+1)^2} + \frac{n}{(n+2)(n+1)^2}\theta^2 =$$
$$= \boxed{\frac{2\theta^2}{(n+2)(n+1)}}.$$

(c) Since

$$\lim_{n\to\infty} \text{MSE}\left(\widehat{\Theta}_n\right) = \lim_{n\to\infty} \frac{2\theta^2}{(n+2)(n+1)} = 0,$$

estimator $\widehat{\Theta}_n$ is $\boxed{\text{consistent}}$.

Problem 7

When R successes occur in n trials, the sample proportion $\widehat{p} = R/n$ customarily is used as an estimator of the probability of success p. However, there are sometimes good reasons to use the estimator $p^* \equiv \dfrac{R+1}{n+2}$. Alternatively, p^* can be written as a linear combination of the familiar estimator $\widehat{p}$:

$$p^* = \frac{n\widehat{p}+1}{n+2} = \frac{n}{n+2} \cdot \widehat{p} + \frac{1}{n+2}.$$

(a) What is the MSE of $\widehat{p}$? Is it consistent?

(b) What is the MSE of p^* ? Is it consistent?

(c) To decide which estimator is better, $\widehat{p}$ or p^*, does consistency help? What criterion would help?

(d) Tabulate the efficiency of p^* relative to $\widehat{p}$, for example when $n = 10$ and $p = 0, 0.1, 0.2, \ldots,$ $0.9, 1.0$.

(e) State some possible circumstances when you might prefer to use p^* instead of $\widehat{p}$ to estimate p.

Solution:

(a) MSE is a sum of variance and bias squared. Since

$$\mathsf{E}\left(\widehat{p}\right) = p \quad \text{and} \quad \mathsf{V}\left(\widehat{p}\right) = \frac{p(1-p)}{n},$$

Bias $(\widehat{p}) = \mathsf{E}\left(\widehat{p}\right) - p = 0$, and MSE:

$$\text{MSE}\left(\widehat{p}\right) = \text{Bias}^2(\widehat{p}) + \mathsf{V}\left(\widehat{p}\right) = 0^2 + \frac{p(1-p)}{n} = \boxed{\frac{p(1-p)}{n}}.$$

Since

$$\lim_{n\to\infty} \text{MSE}\left(\widehat{p}\right) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0,$$

estimator $\widehat{p}$ is $\boxed{\text{consistent}}$.

(b) Let's find $\mathsf{E}\left(p^*\right)$ and $\mathsf{V}\left(p^*\right)$:

$$\mathsf{E}\left(p^*\right) = \mathsf{E}\left(\frac{n}{n+2} \cdot \widehat{p} + \frac{1}{n+2}\right) = \frac{n}{n+2} \cdot \mathsf{E}\left(\widehat{p}\right) + \frac{1}{n+2} = \frac{np+1}{n+2},$$

$$\mathsf{V}\left(p^*\right) = \mathsf{V}\left(\frac{n}{n+2}\cdot\widehat{p}+\frac{1}{n+2}\right) = \left(\frac{n}{n+2}\right)^2\cdot\mathsf{V}\left(\widehat{p}\right) = \frac{np(1-p)}{(n+2)^2}.$$

MSE of p^* then:

$$\mathrm{MSE}\left(p^*\right) = \mathrm{Bias}^2(p^*)+\mathsf{V}\left(p^*\right) = \left(\frac{np+1}{n+2}-p\right)^2+\frac{np(1-p)}{(n+2)^2} = \boxed{\frac{(1-2p)^2+np(1-p)}{(n+2)^2}}.$$

Since

$$\lim_{n\to\infty}\mathrm{MSE}\left(p^*\right) = \lim_{n\to\infty}\frac{(1-2p)^2+np(1-p)}{(n+2)^2} = 0,$$

estimator p^* is $\boxed{\text{consistent}}$.

(c) Since both estimators $\widehat{p}$ and p^* are consistent, we can not decide based on this property which one is better. Relative efficiency (upon MSE) would help.

(d) We will not tabulate MSE, but rather plot them. Illustrations are given in fig. 2.

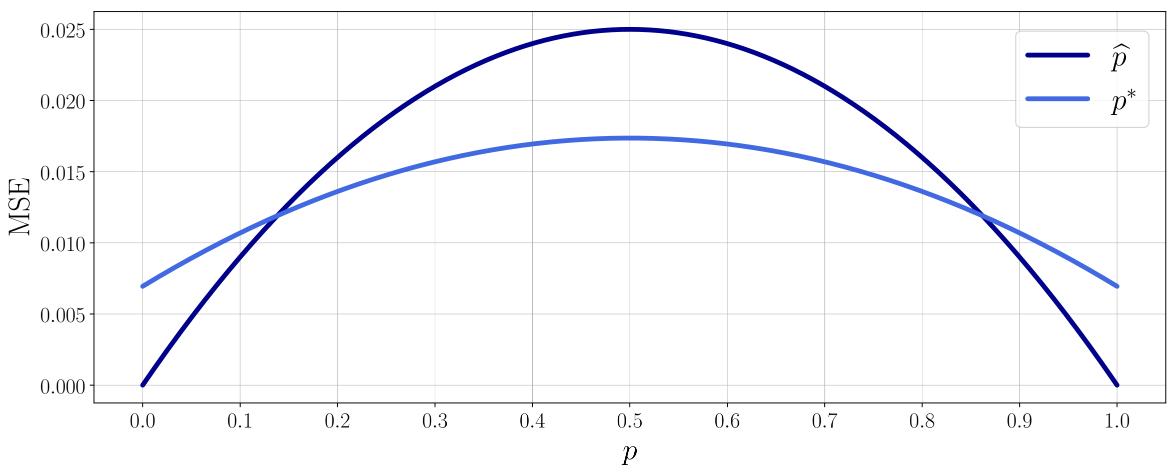

Figure 2: MSE of estimators $\widehat{p}$ and p^* for $n = 10$.

(e) We would prefer to use p^* instead of $\widehat{p}$ in the interval of probabilities, where $\mathrm{MSE}\left(p^*\right) < \mathrm{MSE}\left(\widehat{p}\right)$. For the case of $n = 10$ this interval is approximately $\boxed{p \in (0.14, 0.86)}$.