Module 17 self-assessment

Question 1

Consider $X_1, X_2, \dots, X_n$ iid random variables from a population with mean μ_X and variance σ_X^2. If we compute an estimator for the variance as

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

where $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$, what is the bias of this estimator?

Solution:

Out of the iid hint of the question we immediately have $\mathbb{E}[X_i] = \mu_X$ and $\mathrm{Var}(X_i) = \sigma_X^2$ for all X_i, $i = 1, \dots, n$. We also know that by virtue of the CLT $\bar{X}_n \sim \mathcal{N}\big(\mu_X, \frac{\sigma_X^2}{n}\big)$. From the definition of the bias

$$\mathrm{Bias}(S_n^2) = \mathbb{E}[S_n^2] - \sigma_X^2,$$

to compute the bias it suffices to compute $\mathbb{E}[S_n^2]$ and take away σ_X^2. Thus expanding the square of differences inside the expectation yields

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\Big] \\
&= \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2 - 2\bar{X}_n X_i + \bar{X}_n^2\Big] \\
&= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - 2\mathbb{E}\Big[\sum_{i=1}^{n} \bar{X}_n X_i\Big] + \mathbb{E}\Big[\sum_{i=1}^{n} \bar{X}_n^2\Big]\Big)
\end{aligned}$$

The above expression has $n+1$ random variables, the $X_1, \dots, X_n$ and the $\bar{X}_n$ (since different X_i will give a different n-sample mean). To simplify we can write X_n in terms of X_i by re-arranging its definition as $n\bar{X}_n = \sum_{i=1}^{n} X_i$. Also note that the terms $\bar{X}_n$ and $\bar{X}_n^2$ above do not change inside the sum, so we can pull them out as common factors.

$$
\begin{aligned}
\mathbb{E}[S_n^2] &= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - 2\mathbb{E}\Big[\sum_{i=1}^{n} \bar{X}_n X_i\Big] + \mathbb{E}\Big[\sum_{i=1}^{n} \bar{X}_n^2\Big]\Big) \\
&= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - 2\mathbb{E}\Big[\bar{X}_n \sum_{i=1}^{n} X_i\Big] + \mathbb{E}\Big[\bar{X}_n^2 \sum_{i=1}^{n} 1\Big]\Big) \\
&= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - 2\mathbb{E}\Big[\bar{X}_n \cdot n\bar{X}_n\Big] + \mathbb{E}\Big[n\bar{X}_n^2\Big]\Big) \\
&= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - 2\mathbb{E}\Big[n\bar{X}_n^2\Big] + \mathbb{E}\Big[n\bar{X}_n^2\Big]\Big) \\
&= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - \mathbb{E}\Big[n\bar{X}_n^2\Big]\Big) \\
&= \frac{1}{n}\Big(\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - n\mathbb{E}\Big[\bar{X}_n^2\Big]\Big)
\end{aligned}
$$

where the last equality is due to the linearity of the expectation function, e.g. $2\mathbb{E}[Z] - \mathbb{E}[Z] = \mathbb{E}[Z]$.

As $X_1, \ldots, X_n$ are independent, then the expectation of the sum is the sum of the expectation hence we can express the first term above as

$$
\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] = \sum_{i=1}^{n}\big(\mathrm{Var}(X_i) + \mathbb{E}[X_i]^2\big) = \sum_{i=1}^{n}\big(\sigma_X^2 + \mu_X^2\big) = n(\sigma_X^2 + \mu_X^2),
$$

where we have used the variance formula $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, (for $X = X_i$) and the information on the mean and variance of the X_i from the start of the solution. As we also know the mean and variance of $\bar{X}_n$ (via the CLT - also quoted at the start of the solution) the same formula for $X = \bar{X}_n$ gives

$$
\mathbb{E}[\bar{X}_n^2] = \mathrm{Var}(\bar{X}_n) + \mathbb{E}[\bar{X}_n]^2 = \frac{\sigma_X^2}{n} + \mu_X^2.
$$

Plugging into the derivation gives

$$
\begin{aligned}
\mathbb{E}[S_n^2] &= \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} X_i^2\Big] - \mathbb{E}\Big[\bar{X}_n^2\Big] \\
&= \frac{1}{n} n(\sigma_X^2 + \mu_X^2) - \frac{1}{n}\sigma_X^2 - \mu_X^2 \\
&= \big(1 - \frac{1}{n}\big)\sigma_X^2 \\
&= \frac{n-1}{n}\sigma_X^2.
\end{aligned}
$$

$$
\mathrm{Bias}(S_n^2) = \mathbb{E}[S_n^2] - \sigma_X^2 = \Big(\frac{n-1}{n} - 1\Big)\sigma_X^2 = -\frac{1}{n}\sigma_X^2.
$$

Question 2

Form and evaluate the likelihood function for the observations

$$x_1 = 2, \quad x_2 = 1, \quad x_3 = 3, \quad \text{and} \quad x_4 = 2,$$

if they are drawn from a Binomial distribution with $n = 3$ and some unknown probability of success p. Assume that the observations are iid random variables. Subsequently, compute the MLE estimator for p.

Solution:

Since the observations are iid the likelihood function, that is their joint distribution, is simply the product of their individual (marginal) distributions

$$L(x_1, x_2, x_3, x_4; p) = p_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4; p) = \prod_{i=1}^{4} p_{X_i}(x_i; p),$$

where

$$p_{X_i}(x_i; p) = \binom{3}{x_i} p^{x_i}(1-p)^{3-x_i},$$

from the definition of the Binomial with $n = 0$ and probability of success p. Multiplying we get

$$L(x_1, x_2, x_3, x_4; p) = \binom{3}{x_1}\binom{3}{x_2}\binom{3}{x_3}\binom{3}{x_4} p^{x_1+x_2+x_3+x_4}(1-p)^{12-(x_1+x_2+x_3+x_4)}.$$

Plugging in the values of $x_1, \ldots, x_4$ gives

$$L(2, 1, 3, 2; p) = 27p^8(1-p)^4.$$

To find $\hat{p}_{\text{MLE}}$ we find the negative log likelihood t

$$-\log L(p) = -\left(\log 27 + \log p^8 + \log(1-p)^4\right) = -\log 27 - 8\log p - 4\log(1-p)$$

and set its derivative to zero

$$\frac{\mathrm{d}}{\mathrm{d}p}(-\log L(p)) = -\frac{8}{p} + \frac{4}{1-p} = \frac{4p - 8 + 8p}{p(1-p)} \Longrightarrow p = \frac{2}{3},$$

since $0 < p < 1$.