

Problem Set 1

Question 1.

(a) Given an i.i.d. random sample $X_1, X_2, \dots, X_n$, we use the following formula for the *distribution variance*:

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

where $\bar{X}$ is the sample mean:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Show that this is a biased estimator for the distribution variance.

(b) To correct the bias in $\tilde{S}^2$, we multiply by $\frac{n}{n-1}$, yielding the unbiased sample variance estimator:

$$S^2 = \frac{n}{n-1} \tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Show that this estimator is unbiased for the distribution variance.

(c) Given the random sample (data set) drawn from a given distribution:

$$\{10, 15, 30, 45, 57, 52, 63, 72, 81, 93, 102, 105\},$$

estimate the distribution variance, using both estimators.

Question 2. The following 5 random number are generated from the uniform distribution $U[-1, 1]$:

$$\{0.1812, -0.6245, -0.0422, 0.2049, 0.3150\}.$$

The following 5 random numbers are generated from distribution $N(0, 1)$:

$$\{0.1772, -0.4802, 1.2499, 0.7698, -1.3827\}.$$

Let X be a random variable following uniform distribution $U[-1, 1]$, and $f(x) = \exp(-x^2/2)$. Estimate $E[f(X)]$ using the standard Monte Carlo method (using the given 5 data points). Use the importance sampling method with IS distribution $q(x) = N(0, 1)$ (using the given 5 data points).

Solutions

Question 1.

(a) We can express the sample variance $\tilde{S}^2$ as:

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Expanding this gives:

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$

Simplifying:

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

Since $\sum_{i=1}^n X_i = n\bar{X}$, the second term simplifies to:

$$-2\bar{X} \cdot n\bar{X} = -2n\bar{X}^2$$

Thus:

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

This simplifies to:

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Now, let's compute $E[\tilde{S}^2]$. Using the linearity of expectation:

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

We can split this into two parts:

$$E[\tilde{S}^2] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] - E[\bar{X}^2]$$

Since $E[X_i^2] = \sigma^2 + \mu^2$, we have:

$$E[\tilde{S}^2] = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - E[\bar{X}^2]$$

Simplifying:

$$E[\tilde{S}^2] = \sigma^2 + \mu^2 - E[\bar{X}^2]$$

To calculate $E[\bar{X}^2]$, note that:

$$E[\bar{X}^2] = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 = \frac{1}{n^2} \sum_{i=1}^n E[X_i^2] + \frac{1}{n^2} \sum_{i \neq j} E[X_i X_j]$$

Since $E[X_i X_j] = \mu^2$, we get:

$$E[\bar{X}^2] = \frac{1}{n^2} (n(\sigma^2 + \mu^2) + n(n-1)\mu^2)$$

Simplifying further:

$$E[\bar{X}^2] = \frac{1}{n}(\sigma^2 + \mu^2) + \frac{n-1}{n}\mu^2$$

Thus:

$$E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2$$

Finally, substituting this back into the expression for $E[\tilde{S}^2]$, we get:

$$E[\tilde{S}^2] = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) = \left(1 - \frac{1}{n}\right)\sigma^2$$

Thus, we have shown that:

$$E[\tilde{S}^2] = \frac{n-1}{n}\sigma^2$$

Therefore, $\tilde{S}^2$ is a biased estimator for σ^2 , underestimating the true population variance by a factor of $\frac{n-1}{n}$.

(b) To correct the bias in $\tilde{S}^2$, we multiply by $\frac{n}{n-1}$, yielding the unbiased sample variance estimator:

$$S^2 = \frac{n}{n-1}\tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

The expectation of S^2 is:

$$E[S^2] = \frac{n}{n-1}E[\tilde{S}^2] = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 = \sigma^2$$

Thus, S^2 is an unbiased estimator of the population variance σ^2 .

Question 2.

We are estimating $E[f(X)]$, where $f(x) = \exp\left(-\frac{x^2}{2}\right)$ and $X \sim U[-1, 1]$.

Monte Carlo:

The following 5 samples from the uniform distribution $U[-1, 1]$ are given:

$$\{0.1812, -0.6245, -0.0422, 0.2049, 0.3150\}$$

1. Evaluate the function $f(x)$ for each sample:

$$f(x) = \exp\left(-\frac{x^2}{2}\right)$$

- $f(0.1812) = \exp\left(-\frac{(0.1812)^2}{2}\right) \approx 0.9837$
- $f(-0.6245) = \exp\left(-\frac{(-0.6245)^2}{2}\right) \approx 0.8228$
- $f(-0.0422) = \exp\left(-\frac{(-0.0422)^2}{2}\right) \approx 0.9991$
- $f(0.2049) = \exp\left(-\frac{(0.2049)^2}{2}\right) \approx 0.9792$
- $f(0.3150) = \exp\left(-\frac{(0.3150)^2}{2}\right) \approx 0.9516$

Monte Carlo estimator:

$$E[f(X)] \approx \hat{f}_{MC} = \frac{1}{5} (0.9837 + 0.8228 + 0.9991 + 0.9792 + 0.9516) = \frac{4.7364}{5} = 0.9473$$

Importance Sampling:

The importance sampling distribution is $q(x) \sim N(0, 1)$. The following 5 random samples are provided from $N(0, 1)$:

$$\{0.1772, -0.4802, 1.2499, 0.7698, -1.3827\}$$

For each of the 5 samples x_i , we compute $f(x) = \exp\left(-\frac{x^2}{2}\right)$:

$$\begin{aligned} f(0.1772) &= \exp\left(-\frac{(0.1772)^2}{2}\right) \approx 0.9844 \\ f(-0.4802) &= \exp\left(-\frac{(-0.4802)^2}{2}\right) \approx 0.8948 \\ f(1.2499) &= \exp\left(-\frac{(1.2499)^2}{2}\right) \approx 0.5044 \\ f(0.7698) &= \exp\left(-\frac{(0.7698)^2}{2}\right) \approx 0.7379 \\ f(-1.3827) &= \exp\left(-\frac{(-1.3827)^2}{2}\right) \approx 0.3669 \end{aligned}$$

The weight for each sample is computed as:

$$w(x_i) = \frac{p(x_i)}{q(x_i)}.$$

Note that $p(x) = \frac{1}{2}$ for $x \in [-1, 1]$, and $p(x) = 0$ for $x \notin [-1, 1]$. $q(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.

For each sample:

$$\begin{aligned} w(0.1772) &= \frac{\frac{1}{2}}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(0.1772)^2}{2}\right)} \approx 1.2524 \\ w(-0.4802) &= \frac{\frac{1}{2}}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(-0.4802)^2}{2}\right)} \approx 1.1193 \\ w(1.2499) &= 0 \quad (\text{since } 1.2499 \notin [-1, 1]) \\ w(0.7698) &= \frac{\frac{1}{2}}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(0.7698)^2}{2}\right)} \approx 0.9227 \\ w(-1.3827) &= 0 \quad (\text{since } -1.3827 \notin [-1, 1]) \end{aligned}$$

The importance sampling estimate is given by:

$$\begin{aligned} E[f(X)] &\approx \frac{1}{n} \sum_{i=1}^n w(x_i) \cdot f(x_i) \\ &= \frac{1}{5} (1.2524 \times 0.9844 + 1.1193 \times 0.8948 + 0 \times 0.5044 + 0.9227 \times 0.7379 + 0 \times 0.3669) = 0.5830 \end{aligned} \tag{1}$$

Thus, the importance sampling estimate is 0.5830.