Proof that the Variance Estimator using $\frac{1}{N}$ is Biased

Proof

Given a random sample $X_1, X_2, \ldots, X_N$ drawn from a population with mean μ and variance σ^2, we consider the biased variance estimator:

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu}_X)^2.$$

where the sample mean is given by:

$$\hat{\mu}_X = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Our goal is to compute $\mathbb{E}[\hat{\sigma}_X^2]$ and compare it to σ^2.

Step 1: Expanding the Squared Differences

Expanding the squared term:

$$(X_i - \hat{\mu}_X)^2 = X_i^2 - 2X_i\hat{\mu}_X + \hat{\mu}_X^2.$$

Summing over all i:

$$\sum_{i=1}^{N} (X_i - \hat{\mu}_X)^2 = \sum_{i=1}^{N} X_i^2 - 2\hat{\mu}_X \sum_{i=1}^{N} X_i + N\hat{\mu}_X^2.$$

Since $\sum_{i=1}^{N} X_i = N\hat{\mu}_X$, substitute and obtain:

$$\sum_{i=1}^{N} (X_i - \hat{\mu}_X)^2 = \sum_{i=1}^{N} X_i^2 - N\hat{\mu}_X^2.$$

Dividing by N to obtain $\hat{\sigma}_X^2$:

$$\hat{\sigma}_X^2 = \frac{1}{N} \left(\sum_{i=1}^{N} X_i^2 - N\hat{\mu}_X^2 \right).$$

Step 2: Taking Expectation

Taking expectation on both sides:

$$\mathbb{E}[\hat{\sigma}_X^2] = \frac{1}{N}\left(\sum_{i=1}^{N}\mathbb{E}[X_i^2] - N\mathbb{E}[\hat{\mu}_X^2]\right).$$

Since each X_i has mean μ and variance σ^2, we use:

$$\mathbb{E}[X_i^2] = \sigma^2 + \mu^2.$$

This is because $\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2$— a property we have proven earlier about the variance (see lecture 8).

Summing over all N:

$$\sum_{i=1}^{N}\mathbb{E}[X_i^2] = N(\sigma^2 + \mu^2).$$

For $\mathbb{E}[\hat{\mu}_X^2]$, we use:

$$\mathbb{E}[\hat{\mu}_X^2] = \mathrm{Var}(\hat{\mu}_X) + (\mathbb{E}[\hat{\mu}_X])^2 = \frac{\sigma^2}{N} + \mu^2.$$

Substituting these into our expectation:

$$\mathbb{E}[\hat{\sigma}_X^2] = \frac{1}{N}\left(N(\sigma^2 + \mu^2) - N\left(\frac{\sigma^2}{N} + \mu^2\right)\right).$$

Expanding:

$$\mathbb{E}[\hat{\sigma}_X^2] = \frac{1}{N}\left(N\sigma^2 + N\mu^2 - \sigma^2 - N\mu^2\right).$$

$$= \frac{1}{N}\left(N\sigma^2 - \sigma^2\right).$$

$$= \sigma^2\left(1 - \frac{1}{N}\right).$$

Step 3: Computing the Bias

Since the true variance is σ^2, the bias of $\hat{\sigma}_X^2$ is:

$$\mathbb{E}[\hat{\sigma}_X^2] - \sigma^2 = \sigma^2\left(1 - \frac{1}{N}\right) - \sigma^2.$$

$$= -\frac{\sigma^2}{N}.$$

Since this bias is negative, $\hat{\sigma}_X^2$ underestimates the true variance, proving that it is biased.

Conclusion

To obtain an unbiased estimator, we must instead use:

$$\bar{\sigma}_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2.$$

In this way:

$$\bar{\sigma}_X^2 = \frac{N}{N-1} \cdot \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2 = \frac{N}{N-1}\hat{\sigma}_X^2.$$

and therefore

$$\begin{aligned}
&\mathbb{E}(\bar{\sigma}_X^2) && (1)\\
=&\mathbb{E}(\frac{N}{N-1}\hat{\sigma}_X^2) && (2)\\
=&\frac{N}{N-1}\mathbb{E}(\hat{\sigma}_X^2) && (3)\\
=&\frac{N}{N-1}\sigma^2\left(1-\frac{1}{N}\right) && (4)\\
=&\sigma^2. && (5)
\end{aligned}$$

which is unbiased.

This correction compensates for the additional variability introduced by using $\hat{\mu}_X$ as an estimate of μ.