

# Sampling Variance vs. Variance Estimator

Xinyu Cindy Wang

January 10, 2025

## 1 Sampling Variance of an Estimator

In the case of the estimator of the Mean of the population  $\mu$ , i.e., the Sample Mean  $\bar{X}$ , **the sampling variance of the sample mean  $\bar{X}$**  is

$$\begin{aligned} V[\bar{X}] &= V\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] \\ &= \frac{1}{n^2}V[X_1 + X_2 + \dots + X_n] \\ &= \frac{1}{n^2}(V[X_1] + V[X_2] + \dots + V[X_n]) && \text{(independent of i.i.d., } COV(X_n, X_{n+1}) = 0) \\ &= \frac{1}{n^2}(V[X] + V[X] + \dots + V[X]) && \text{(identically distributed of i.i.d.)} \\ &= \frac{1}{n^2}nV[X] \\ &= \frac{V[X]}{n} \end{aligned}$$

(Aronow and Miller, 98).

## 2 Variance Estimator

**The population variance** is

$$V[X] = E[X^2] - E[X]^2$$

**The plug-in sample variance** is

$$\begin{aligned} \hat{V}[X]_{plug-in} &= \overline{X^2} - \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

The plug-in sample variance  $\hat{V}[X]_{plug-in}$  is a **biased sample variance**, i.e., a **biased estimator** of the population variance  $V[X]$  because

$$\begin{aligned}
E[\hat{V}[X]_{plug-in}] &= E[\overline{X^2} - \overline{X}^2] \\
&= E[\overline{X^2}] - E[\overline{X}^2] \\
&= E[X^2] - (E[X]^2 + \frac{V[X]}{n}) \\
&= (E[X^2] + E[X]^2) - \frac{V[X]}{n} \\
&= V[X] - \frac{V[X]}{n} \\
&= (1 - \frac{1}{n})V[X] \\
&= \frac{n-1}{n}V[X] \\
&\neq V[X]
\end{aligned}$$

But nevertheless,  $E[\hat{V}[X]_{plug-in}]$  converges to  $V[X]$ . Because  $\frac{n-1}{n} \rightarrow 1$  as  $n \rightarrow \infty$ .

Therefore, **the unbiased sample variance** is

$$\begin{aligned}
\hat{V}[X] &= \frac{n}{n-1} \hat{V}[X]_{plug-in} \\
&= \frac{n}{n-1} (\overline{X^2} - \overline{X}^2) \\
&= \frac{n}{n-1} \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right] = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2
\end{aligned}$$

Now, we have  $E[\hat{V}[X]] = V[X]$

(Aronow and Miller, 106-107).