

Question 1(a): is this estimator unbiased?

Let $X_1, \dots, X_n$ be iid with

$$\mathbb{E}[X_i] = \mu, \quad 0 < \text{Var}(X_i) = \sigma^2 < \infty, \quad n \geq 2.$$

Consider the following *estimator of the population variance*:

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

Show that $\tilde{S}^2$ is biased and calculate its bias.

Try first 3 minutes

Expand the square and use $\sum_i X_i = n\bar{X}$.

A useful identity for any finite-variance variable

For $m = \mathbb{E}[Y]$,

$$\begin{aligned}\text{Var}(Y) &= \mathbb{E}[(Y - m)^2] \\ &= \mathbb{E}[Y^2] - 2m\mathbb{E}[Y] + m^2 \\ &= \mathbb{E}[Y^2] - m^2.\end{aligned}$$

Hence

$$\boxed{\mathbb{E}[Y^2] = \text{Var}(Y) + (\mathbb{E}[Y])^2.}$$

No Gaussian assumption

This is an algebraic identity. It applies to any random variable with finite second moment.

Solution 1(a): simplify the estimator

$$\begin{aligned}\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2.\end{aligned}$$

Divide by n :

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

We now need $\mathbb{E}[X_i^2]$ and $\mathbb{E}[\bar{X}^2]$.

Solution 1(a): take expectations

Identical distributions and independence give

$$\mathbb{E}[\bar{X}] = \mu, \quad \text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}.$$

Therefore

$$\mathbb{E}[X_i^2] = \sigma^2 + \mu^2, \quad \mathbb{E}[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2.$$

Substitute into the expanded estimator:

$$\begin{aligned} \mathbb{E}[\tilde{S}^2] &= \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2 \right) \\ &= \boxed{\frac{n-1}{n} \sigma^2}. \end{aligned}$$

What does downward bias mean?

$$\text{Bias}(\tilde{S}^2) = \mathbb{E}[\tilde{S}^2] - \sigma^2 = \boxed{-\frac{\sigma^2}{n}}.$$

- ▶ The estimator underestimates variance *on average over repeated samples*.
- ▶ An individual estimate can still exceed the true variance.
- ▶ The target σ^2 is fixed; $\tilde{S}^2$ is random before observing data.

Why the bias appears

The sample mean is fitted to the same data. Squared deviations around this fitted centre are smaller than around the true mean.

Question 1(b): correct the bias

Define

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

Show that S^2 is unbiased for σ^2 .

Pause 1 minute

Use the expectation found in part (a). No new expansion is needed.

Solution 1(b): the correction cancels the bias

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

Another useful identity is

$$\sum_i (X_i - \bar{X})^2 = \sum_i (X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

Taking expectations gives $n\sigma^2 - \sigma^2 = (n-1)\sigma^2$.

Be precise

S^2 is unbiased for the variance. Its square root S is generally *not* unbiased for the standard deviation.