Estimation of the variance

by Marco Taboga, PhD

Variance estimation is a statistical inference problem in which a sample is used to produce a point estimate of the variance of an unknown distribution.

The problem is typically solved by using the sample variance as an estimator of the population variance.

In this lecture, we present two examples, concerning:

1. IID samples from a normal distribution whose mean is known;
2. IID samples from a normal distribution whose mean is unknown.

For each of these two cases, we derive the expected value, the distribution and the asymptotic properties of the variance estimator.

Normal IID samples - Known mean

In this example of variance estimation we make assumptions that are similar to those we made in the mean estimation of normal IID samples.

The sample

The sample is made of n independent draws from a normal distribution.

Specifically, we observe the realizations of n independent random variables X_1, ..., X_n, all having

- known mean μ;
- unknown variance σ^2.

The estimator

We use the following estimator of variance:

$$\widehat{\sigma_n^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Expected value of the estimator

The expected value of the estimator $\widehat{\sigma_n^2}$ is equal to the true variance σ^2:

$$\mathrm{E}\left[\widehat{\sigma_n^2}\right] = \sigma^2$$

Proof

Therefore, the estimator $\widehat{\sigma_n^2}$ is unbiased.

Variance of the estimator

The variance of the estimator $\widehat{\sigma_n^2}$ is

$$\mathrm{Var}\left[\widehat{\sigma_n^2}\right] = \frac{2\sigma^4}{n}$$

Proof

Therefore, the variance of the estimator tends to zero as the sample size n tends to infinity.

Distribution of the estimator

The estimator $\widehat{\sigma_n^2}$ has a Gamma distribution with parameters n and σ^2.

Proof

Risk of the estimator

The mean squared error of the estimator is

$$\begin{aligned}
\mathrm{MSE}\left(\widehat{\sigma_n^2}\right) &= \mathrm{E}\left[\left\|\widehat{\sigma_n^2}-\sigma^2\right\|^2\right] \\
&= \mathrm{E}\left[\left|\widehat{\sigma_n^2}-\sigma^2\right|^2\right] && \text{(Euclidean norm in one dimension is equal to absolute value)} \\
&= \mathrm{E}\left[\left(\widehat{\sigma_n^2}-\sigma^2\right)^2\right] \\
&= \mathrm{Var}\left[\widehat{\sigma_n^2}\right] && \text{(by the definition of variance, because } \mathrm{E}\left[\widehat{\sigma_n^2}\right] = \sigma^2\text{)} \\
&= \frac{2\sigma^4}{n}
\end{aligned}$$

Consistency of the estimator

The estimator

$$\widehat{\sigma_n^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

can be viewed as the sample mean of a sequence $\{Y_n\}$ where the generic term of the sequence is

$$Y_n = (X_n-\mu)^2$$

Since the sequence $\{Y_n\}$ is an IID sequence with finite mean, it satisfies the conditions of Kolmogorov's Strong Law of Large Numbers.

Therefore, the sample mean of Y_n converges almost surely to the true mean $\mathrm{E}[Y_n]$:

$$\widehat{\sigma^2} \xrightarrow{a.s.} \mathrm{E}[Y_n] = \sigma^2$$

In other words, the estimator $\widehat{\sigma^2}$ is strongly consistent.

It is also weakly consistent, because almost sure convergence implies convergence in probability:

$$\operatorname*{plim}_{n\to\infty} \widehat{\sigma^2} = \sigma^2$$

Normal IID samples - Unknown mean

This example of variance estimation is similar to the previous one. The only difference is that we relax the assumption that the mean of the distribution is known.

The sample

The sample is made of independent draws from a normal distribution.

Specifically, we observe the realizations of n independent random variables X_1, ..., X_n, all having a normal distribution with:

- unknown mean μ;
- unknown variance σ^2.

The estimator

In this example also the mean of the distribution, being unknown, needs to be estimated.

It is estimated with the sample mean $\bar{X}_n$:

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

We use the following estimators of variance:

1. the unadjusted sample variance:

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

2. the adjusted sample variance:

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

Expected value of the estimator

The expected value of the unadjusted sample variance is

$$\mathrm{E}[S_n^2] = \frac{n-1}{n}\sigma^2$$

Proof

Therefore, the unadjusted sample variance S_n^2 is a biased estimator of the true variance σ^2.

The adjusted sample variance s_n^2, on the contrary, is an unbiased estimator of variance:

$$\mathrm{E}[s_n^2] = \sigma^2$$

Proof

Thus, when also the mean μ is being estimated, we need to divide by $n-1$ rather than by n to obtain an unbiased estimator.

Intuitively, by considering squared deviations from the sample mean rather than squared deviations from the true mean, we are underestimating the true variability of the data.

In fact, the sum of squared deviations from the true mean is always larger than the sum of squared deviations from the sample mean.

Dividing by $n-1$ rather than by n exactly corrects this bias. The number $n-1$ by which we divide is called the **number of degrees of freedom** and it is equal to the number of sample points (n) minus the number of other parameters to be estimated (in our case 1, the true mean μ).

The factor by which we need to multiply the biased estimator S_n^2 to obtain the unbiased estimator s_n^2 is

$$\frac{n}{n-1}$$

This factor is known as **degrees of freedom adjustment**, which explains why S_n^2 is called unadjusted sample variance and s_n^2 is called adjusted sample variance.

Variance of the estimator

The variance of the unadjusted sample variance is

$$\mathrm{Var}\left[S_n^2\right] = \frac{n-1}{n}\frac{2\sigma^4}{n}$$

Proof

The variance of the adjusted sample variance is

$$\mathrm{Var}\left[s_n^2\right] = \frac{2\sigma^4}{n-1}$$

Proof

Therefore, both the variance of S_n^2 and the variance of s_n^2 converge to zero as the sample size n tends to infinity.

Note that the unadjusted sample variance S_n^2, despite being biased, has a smaller variance than the adjusted sample variance s_n^2, which is instead unbiased.

Distribution of the estimator

The unadjusted sample variance S_n^2 has a Gamma distribution with parameters $n-1$ and $\frac{(n-1)\sigma^2}{n}$.

Proof

The adjusted sample variance s_n^2 has a Gamma distribution with parameters $n-1$ and σ^2.

Proof

Risk of the estimator

The mean squared error of the unadjusted sample variance is

$$\text{MSE}\left(S_n^2\right) = \frac{2n-1}{n^2}\sigma^4$$

Proof

The mean squared error of the adjusted sample variance is

$$\text{MSE}\left(s_n^2\right) = \frac{2}{n-1}\sigma^4$$

Proof

Therefore the mean squared error of the unadjusted sample variance is always smaller than the mean squared error of the adjusted sample variance:

$$\begin{aligned}
\text{MSE}\left(S_n^2\right) &= \frac{2n-1}{n^2}\sigma^4 \\
&= \left(\frac{2n}{n^2} - \frac{1}{n^2}\right)\sigma^4 \\
&= \left(\frac{2}{n} - \frac{1}{n^2}\right)\sigma^4 \\
&< \frac{2}{n}\sigma^4 \\
&< \frac{2}{n-1}\sigma^4 = \text{MSE}\left(s_n^2\right)
\end{aligned}$$

Consistency of the estimator

Both the unadjusted and the adjusted sample variances are consistent estimators of the unknown variance σ^2.

Proof

Solved exercises

Below you can find some exercises with explained solutions.

Exercise 1

You observe three independent draws from a normal distribution having unknown mean μ and unknown variance σ^2. Their values are 50, 100 and 150.

Use these values to produce an unbiased estimate of the variance of the distribution.

Solution

Exercise 2

A machine (a laser rangefinder) is used to measure the distance between the machine itself and a given object.

When measuring the distance to an object located 10 meters apart, measurement errors committed by the machine are normally and independently distributed and are on average equal to zero.

The variance of the measurement errors is less than 1 squared centimeter, but its exact value is unknown and needs to be estimated.

To estimate it, we repeatedly take the same measurement and we compute the sample variance of the measurement errors (which we are also able to compute because we know the true distance).

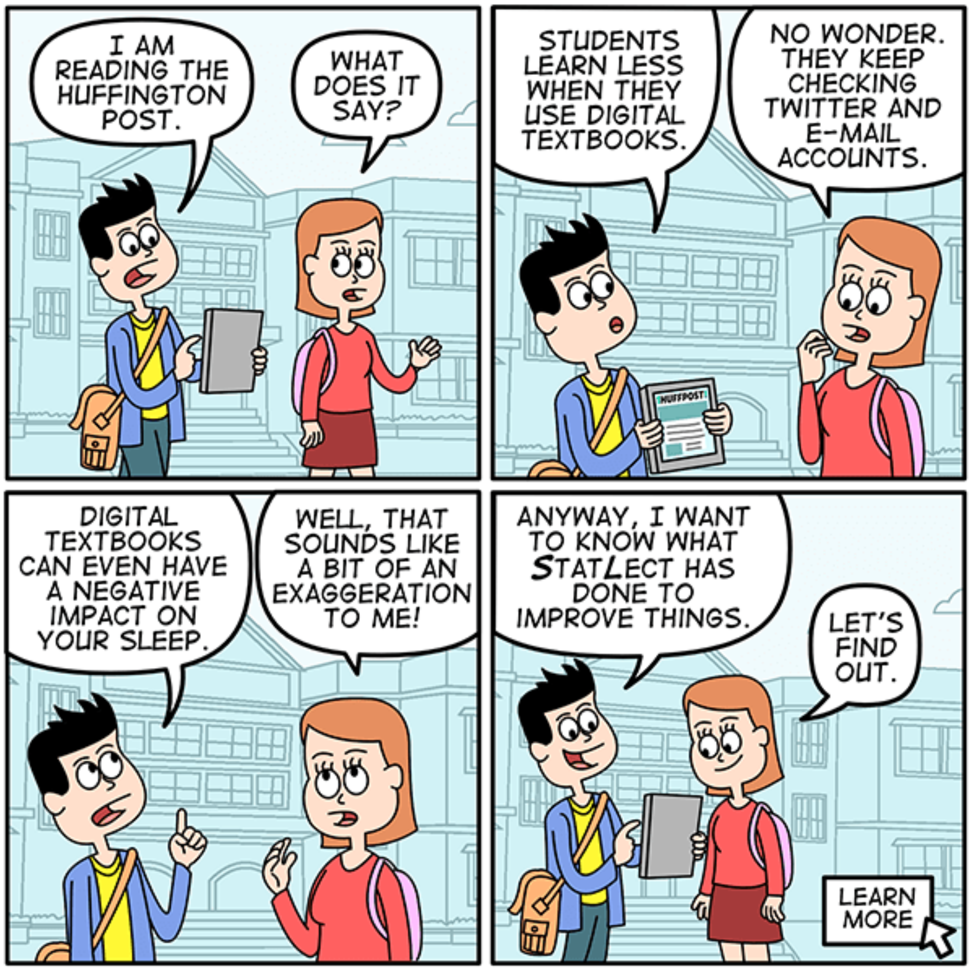
How many measurements do we need to take to obtain an estimator of variance having a standard deviation less than 0.1 squared centimeters?

Solution

How to cite

Please cite as:

Taboga, Marco (2021). "Estimation of the variance", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/fundamentals-of-statistics/variance-estimation.



The books

Most of the learning materials found on this website are now available in a traditional textbook format.

Share

To enhance your privacy,

we removed the social buttons,

but **don't forget to share**.