Math Revision Session

Statistics (6): Sampling Distribution of Sample Variance

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

April 10, 2026

1. Sampling Distribution of Sample Variance

1. Sampling Distribution of Sample Variance

Why Do We Need the Sample Variance?

In practice, the population variance σ^2 is usually unknown.

So we estimate it using the sample variance:

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

This quantity measures how spread out the sample observations are around the sample mean.

What is the Sampling Distribution of S^2?

The sample variance S^2 is itself a random variable.

So if we repeatedly draw samples, S^2 changes from sample to sample.

The **sampling distribution of** S^2 is the probability distribution of this random variable over repeated sampling.

Why Do We Divide by $n-1$?

The sample variance is defined as

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

Why $n-1$, not n?

- The sample mean $\bar{X}$ is estimated from the same data.
- Once $n-1$ deviations $X_i - \bar{X}$ are known, the last one is determined by

$$\sum_{i=1}^{n}(X_i - \bar{X}) = 0.$$

- So only $n-1$ independent pieces of information remain.

This is why we say the sample variance has $n-1$ degrees of freedom.

A Key Identity

A useful identity is

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

This tells us that the total variation around the sample mean is the total variation around the true mean, minus the part explained by estimating the mean.

Proof of the Identity

Start from

$$X_i - \bar{X} = (X_i - \mu) - (\bar{X} - \mu).$$

Then

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}\Big((X_i - \mu) - (\bar{X} - \mu)\Big)^2.$$

Expanding,

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + \sum_{i=1}^{n}(\bar{X} - \mu)^2.$$

Since

$$\sum_{i=1}^{n}(X_i - \mu) = n(\bar{X} - \mu),$$

we obtain

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

Unbiasedness of the Sample Variance

Using the identity from the previous slide, we can show that

$$\mathbb{E}[S^2] = \sigma^2.$$

So S^2 is an **unbiased estimator** of the population variance.

If we divided by n instead, then

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

which is biased downward.

Sampling Distribution in the Normal Case

If

$$X_1, \ldots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2),$$

then the sample variance has a very useful distribution:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

So in the normal case, the sample variance is closely related to the chi-squared distribution.

Why Is This Important?

This result is important because it leads directly to classical inference.

- It explains why the chi-squared distribution appears in variance estimation.
- Together with the sample mean, it leads to the t-distribution:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

  when the population is normal.

So the sampling distribution of S^2 is one of the key building blocks of confidence intervals and hypothesis tests.

Summary

- S^2 is the sample-based estimator of σ^2.
- We divide by $n-1$ because one degree of freedom is lost when $\bar{X}$ is estimated.
- S^2 is unbiased:
$$\mathbb{E}[S^2] = \sigma^2.$$
- If the population is normal, then
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$