

Understanding Variance

Frame

January 23, 2025

1 Definition of Variance

Variance measures the dispersion of a set of data points around their mean value. It is a fundamental concept in statistics that quantifies the spread of the data.

1.1 Basic Definition

The variance of a random variable Z is denoted by σ^2 and defined as the expected value of the squared deviations from the mean:

$$\sigma^2 = E[(Z - \mu)^2]$$

where $\mu = E[Z]$ is the mean of the random variable Z .

1.2 Alternate Formulation

Variance can also be expressed in terms of the expectation of Z and the expectation of Z^2 :

$$\sigma^2 = E[Z^2] - E[Z]^2$$

This formulation highlights how variance measures the difference between the expectation of the square and the square of the expectation, emphasizing the spread of the distribution.

1.3 Discrete Case

In a discrete setting, where each outcome z has a probability p , the variance is calculated as:

$$\sigma^2 = \sum_{i=1}^n p_i (z_i - \mu)^2$$

This expression weights each squared deviation by its probability, reflecting the contribution of each deviation to the overall variance.

2 Sample Variance

When working with sample data, variance can be estimated from a finite set of data points.

2.1 Population vs. Sample Variance

The variance of a population is calculated with the divisor N (the number of observations), which assumes every data point is known:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

However, in sample variance, where data points are a sample from a larger population, the divisor is $N - 1$. This adjustment, known as Bessel's correction, provides an unbiased estimate of the population variance:

$$s^2 = \frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2$$

where \bar{x} is the sample mean.

2.2 Why Use $N - 1$ in Sample Variance?

Using $N - 1$ instead of N compensates for the fact that the sample mean is an imperfect estimate of the population mean. This adjustment helps correct the bias in the estimation of the population variance, making the sample variance an unbiased estimator.

3 Connecting Back to Probability Weights

In contexts where each observation x_i has a probability p_i , the variance formula adapts to use these probabilities as weights:

$$\sigma^2 = \sum_{i=1}^N p_i (x_i - \mu)^2$$

This approach generalizes the concept of variance to any distribution of probabilities, connecting it back to the fundamental principle of expected values.