

# 02-680 Module 17

## Essentials of Mathematics and Statistics

Aidan Jan

November 20, 2025

## Expectation and Moments

Given a probability distribution  $p(X)$  of a random variable  $X$ , how can we summarize the distribution with a single value? The **expectation** of a random variable is a number that attempts to capture the center of  $p(X)$  and can be interpreted as the long-run average of many independent samples from the given distribution.

### Mean of Discrete Random Variables

For a (discrete) random variable  $X$ , the **expected value** is:

$$\mathbb{E}[X] = \sum_{x \in \text{range}(X)} x \cdot p(X = x)$$

We sometimes also call this the **mean** or **first moment** of the variable.

$\text{range}(X)$  means the set of all possible values that the random variable  $X$  can take.

### Bernoulli Random Variable

Let  $X$  have the Bernoulli distribution with parameter  $p$ . That is, assume that  $X$  takes only the two values 0 and 1 with  $\Pr(X = 1) = p$ . Then, the mean of  $X$  is:

$$\mathbb{E}(X) = 0 \times (1 - p) + 1 \times p = p$$

### Expectation for a Continuous Distribution

The idea of computing a weighted average of the possible values can be generalized to continuous random variables by using integrals instead of sums. The distinction between bounded and unbounded random variables arises in the case for the same reasons.

For a continuous random variable, again we swap a sum for an integral:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Once again, the expectation is also called the **mean** or the **expected value**.

## Variance and Standard Deviation

In addition to the mean, we often want to know something about how ‘far’ the likely outcomes are from the mean. In the normal distribution, this is visually about how narrow the bell is.

To do this we define the **variance** of a random variable as

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

It represents the expected squared difference between the random variable and its mean.

## Alternative Method for Calculating Variance

For every random variable  $X$ ,

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Proof: Let  $\mathbb{E}(X) = \mu$ . Then,

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

## Standard Deviation

The standard deviation of a variable is the square root of the variance. It provides a measure of spread in the same units as  $X$ .

## Estimating Variance from Data

Variance is estimated as:

$$\text{Var}(X) \approx \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

where  $\bar{X} = \frac{1}{n} \sum X_i$

This is widely used in real-world statistics when true distribution is unknown.

## Properties of Expectation and Variance

### Summation

For some set  $X_1, X_2, \dots, X_n$  of independent random variables:

$$\begin{aligned}\mathbb{E} \left[ \sum_{i \in [n]} X_i \right] &= \sum_{i \in [n]} \mathbb{E}[X_i] \\ \text{Var} \left[ \sum_{i \in [n]} X_i \right] &= \sum_{i \in [n]} \text{Var}[X_i]\end{aligned}$$

### Linearity

For some random variable  $X$  and constraints  $a, b \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}[aX + b] &= a\mathbb{E}[X] + b \\ \text{Var}[aX + b] &= a^2 \text{Var}[X]\end{aligned}$$

# Conditional Expectation and Variance

**Discrete:**

$$\mathbb{E}[X|Y = y] = \sum_x (x \cdot p(X = x|y = y)) \text{Var}[X|Y = y] = \sum_x (x - \mathbb{E}[X|Y = y])^2 p(X = x|Y = y)$$

**Continuous:**

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x \cdot f(x|y) dx V[X|Y = y] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X|Y = y])^2 \cdot f(x|y) dx$$

# Covariance and Correlation

When we are interested in the joint distribution of two random variables, it is useful to have a summary of how much the two random variables **depend on each other**.

The covariance and correlation are attempts to measure that dependence, but they only capture a particular type of dependence, namely **linear dependence**.

**Concept:**

When we have two random variables  $X$  and  $Y$ , we often want to understand how they **vary together**.

## Covariance Definition:

Given two random variables  $X$  and  $Y$  with expectations  $\mu_X = \mathbb{E}[X]$  and  $\mu_Y = \mathbb{E}[Y]$  respectively, the **covariance** between the two is defined as

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

## Standard Deviations:

- $\sigma_X$  (standard deviation of  $X$ )
- $\sigma_Y$  (standard deviation of  $Y$ )

## Correlation Definition:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

You can think of correlation as a **normalized version of covariance**.