Continuous Random Variables

Rosario Barone

Tor Vergata University of Rome

Statistical tools for decision making

Undergraduate Degree in Global Governance

A.Y. 2023/2024

Transition from Discrete to Continuous

When dealing with continuous quantities, we transition from discrete sums to integrals:

1. In discrete settings, we use summation notation, often denoted as $\sum$, to add up individual values.
2. For continuous variables, we use integration notation, often denoted as $\int$, to account for an uncountably infinite range of values.

- The essence remains the same - finding the total "amount" over a range.
- The key difference is that, in integration, we consider infinitesimally small pieces within the range.
- The transition is analogous to zooming in on ever-smaller intervals, making the sum continuous.

Integral as the Limit of a Sum

Definition

The integral of a function $f(x)$ over an interval $[a, b]$ is defined as the limit of a sum as the partition of the interval becomes increasingly fine. It is denoted by:

$$\int_a^b f(x)\,dx = \lim_{\substack{n\to\infty \\ \Delta x\to 0}} \sum_{i=1}^{n} f(x_i)\Delta x$$

- In this definition:
  - a and b define the interval of integration.
  - n represents the number of subdivisions in the partition.
  - x_i is a point within each subinterval.
  - Δx is the width of each subinterval, given by $\frac{b-a}{n}$.
- As n approaches infinity and the subintervals become infinitesimally small, the sum approaches the integral.
- This concept underlies the development of the Riemann integral, a fundamental tool in calculus.

Continuous Random Variables

Definition

A *continuous random variable* is a variable that can take any value in a continuous range of values. It is associated with a probability density function (PDF) rather than a probability mass function.

- Unlike discrete random variables, continuous random variables can take on an uncountably infinite number of values.
- Probability is defined over intervals rather than individual values.
- The total area under the PDF curve over its entire range is equal to 1.

Probability Density Function (PDF)

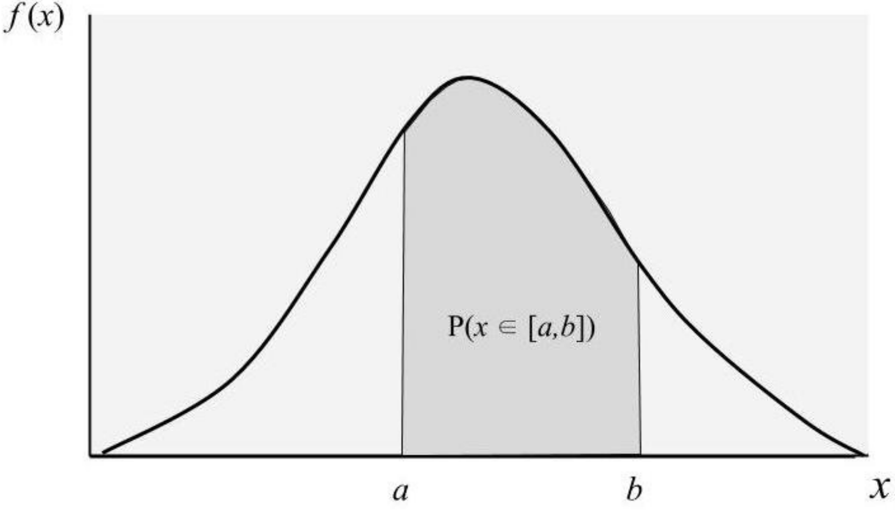
Definition

The *Probability Density Function (PDF)* of a continuous random variable X is a function denoted as $f_X(x)$, which describes the likelihood of X taking on a particular value or falling within a given interval.

- $f_X(x)$ is non-negative for all x; that is, $f_X(x) \geq 0$.
- The total area under the PDF curve over its entire range is equal to 1:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

- The PDF is used to compute probabilities associated with continuous random variables through integration over intervals.
- The PDF can be specific to the probability distribution of X, such as the Gaussian (normal) distribution or the uniform distribution, among others.


$f(x)$
$P(x \in [a,b])$
a
b
x

CDF for Continuous Random Variables

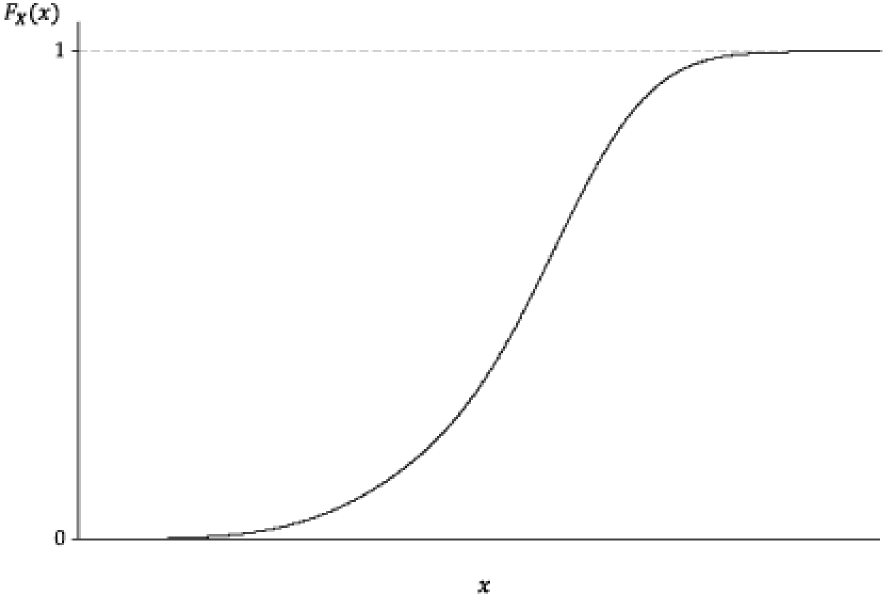
Definition

The *Cumulative Distribution Function (CDF)* of a continuous random variable X is a function denoted as $F_X(x)$, which gives the probability that X takes on a value less than or equal to x.

- The CDF provides a complete description of the probability distribution of X.
- Mathematically, the CDF is defined as:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

  where $f_X(x)$ is the Probability Density Function (PDF) of X.
- The CDF is a monotonically increasing function, and it ranges from 0 to 1 as x varies from $-\infty$ to ∞.
- The CDF is useful for calculating probabilities involving continuous random variables and for finding percentiles.


$F_X(x)$
1
0
x

Quantile Function

Relationship with CDF

The Cumulative Distribution Function (CDF), denoted as $F(x)$, gives the probability that a random variable is less than or equal to x. The quantile function and CDF are related as follows:

$$Q(p) = \inf\{x : F(x) \geq p\}$$

In other words, $Q(p)$ finds the smallest x for which $F(x)$ is greater than or equal to p.

Quantile Function

Quantile Function

The Quantile function takes a probability value p as input and returns the value in a distribution for which the cumulative probability is less than or equal to p. In mathematical terms:

$$Q(p) = x$$

- Q is the quantile function.
- p is the probability or percentile, where $0 \leq p \leq 1$.
- x is the corresponding value in the distribution.

The quantile function is commonly used to find percentiles. For example, Q(0.25) gives the 25th percentile, and Q(0.75) gives the 75th percentile.

Expected Value of a Continuous Random Variable

Definition

The *expected value* (or *mean*) of a continuous random variable X is a measure of the central tendency of its values. It is denoted as μ_X or $E(X)$ and is defined as:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

- $f_X(x)$ is the Probability Density Function (PDF) of the continuous random variable X.
- The integral is taken over the entire range of possible values for X.
- The expected value represents the "average" value that we would expect if we were to sample values of X many times.

Variance of a Continuous Random Variable

Definition

The *variance* of a continuous random variable X measures the spread or dispersion of its values around the expected value. It is denoted as σ_X^2 or $\mathrm{Var}(X)$ and is defined as:

$$\sigma_X^2 = \mathrm{Var}(X) = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f_X(x)\, dx$$

- $f_X(x)$ is the Probability Density Function (PDF) of the continuous random variable X.
- μ_X is the expected value of X.
- The variance quantifies how much the values of X deviate from their expected (average) value.
- A larger variance indicates greater variability or spread in the data.

Families of Continuous Probability Distributions

Continuous probability distributions provide models for random variables that can take on an uncountable range of values.

1. Uniform Distribution
2. Exponential Distribution
3. Gamma Distribution
4. Normal Distribution (Gaussian)

Uniform Distribution

A *uniform distribution* is a continuous random variable that is equally likely to take any value within a specified interval. It has a constant probability density function (PDF) within the interval.

- $X \sim \text{Unif}(a, b)$
- The PDF of a uniform random variable X over the interval $[a, b]$ is given by:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

- Uniform random variables are commonly used in modeling situations where all values within an interval are equally likely.

Expected value and variance for Uniform Distribution

Moments of a Uniform Random Variable

A uniform random variable X over the interval $[a, b]$ has moments defined as:

$$\mu_X = E(X) = \frac{a+b}{2}$$

$$\sigma_X^2 = Var(X) = \frac{(b-a)^2}{12}$$

Derivation of the mean of a Uniform Distribution

Let $X \sim \mathrm{Unif}(a, b)$ with $a < b$.

$$\begin{aligned}
\mu_X &= \int_a^b x \cdot f_X(x)\, dx \\
&= \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^2}{2} - \frac{a^2}{2}\right] \\
&= \frac{b+a}{2}
\end{aligned}$$

Derivation of the mean of a Uniform Distribution

Let $X \sim \text{Unif}(a, b)$ with $a < b$.

$$\begin{aligned}
\mu_X &= \int_a^b x \cdot f_X(x)\, dx \\
&= \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\
&= \frac{1}{b-a} \left[\frac{b^2}{2} - \frac{a^2}{2} \right] \\
&= \frac{b+a}{2}
\end{aligned}$$

Derivation of the mean of a Uniform Distribution

Let $X \sim \text{Unif}(a, b)$ with $a < b$.

$$
\begin{aligned}
\mu_X &= \int_a^b x \cdot f_X(x)\, dx \\
&= \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^2}{2} - \frac{a^2}{2}\right] \\
&= \frac{b+a}{2}
\end{aligned}
$$

Derivation of the mean of a Uniform Distribution

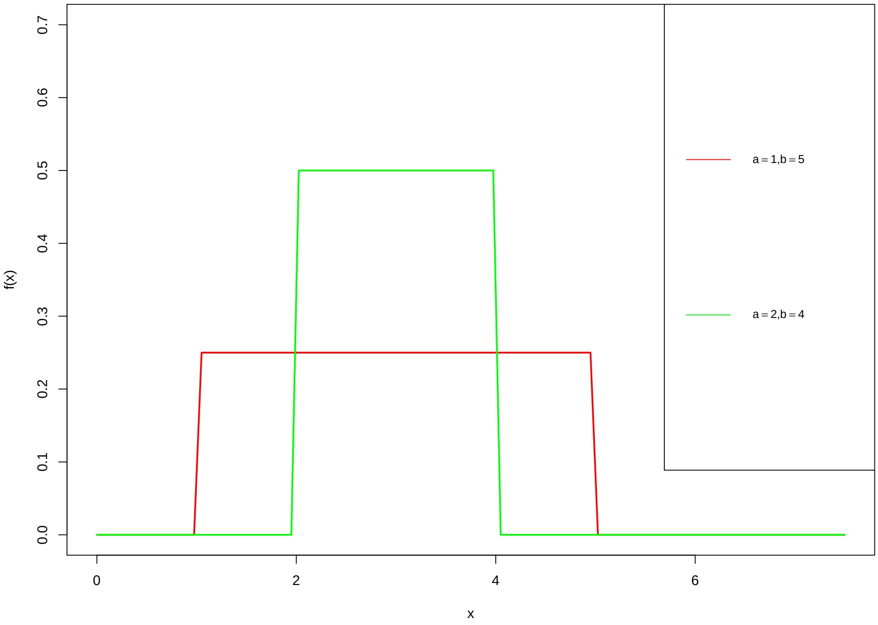
Let $X \sim \text{Unif}(a, b)$ with $a < b$.

$$
\begin{aligned}
\mu_X &= \int_a^b x \cdot f_X(x)\, dx \\
&= \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^2}{2} - \frac{a^2}{2}\right] \\
&= \frac{b+a}{2}
\end{aligned}
$$

Derivation of the mean of a Uniform Distribution

Let $X \sim \text{Unif}(a, b)$ with $a < b$.

$$\begin{aligned}
\mu_X &= \int_a^b x \cdot f_X(x)\, dx \\
&= \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a} \left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a} \left[\frac{b^2}{2} - \frac{a^2}{2}\right] \\
&= \frac{b+a}{2}
\end{aligned}$$



f(x)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
2
4
6
x
a = 1,b = 5
a = 2,b = 4

Exponential Distribution

An *exponential distribution* models the time between events in a Poisson process, where events occur at a constant rate and independently of the time since the last event.

- $X \sim \mathrm{Exp}(\lambda)$
- The probability density function (PDF) of an exponential random variable X with rate parameter λ is given by:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- Exponential random variables are commonly used in modeling waiting times, lifetimes, and reliability analysis.

Mean and Variance of the Exponential Distribution

Mean (Expected Value)

The mean (μ) of the exponential distribution is given by:

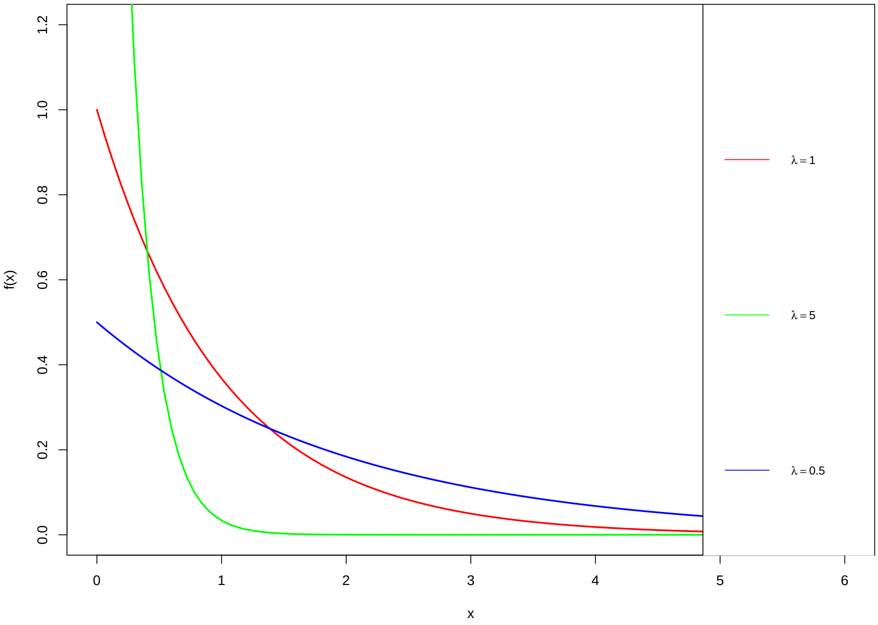
$$\mu_X = E(X) = \frac{1}{\lambda}$$

Variance

The variance (σ^2) of the exponential distribution is given by:

$$\sigma^2 = Var(X) = \frac{1}{\lambda^2}$$

- The exponential distribution is often used to model waiting times, lifetimes, and reliability.



f(x)
x
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
λ = 1
λ = 5
λ = 0.5

Gamma Distribution

The *Gamma random distribution* is a continuous probability distribution that represents the waiting time until a Poisson process with rate λ reaches a certain number of events.

- $X \sim \text{Gamma}(k, \lambda)$
- Parameters:
  - $k > 0$ (shape parameter)
  - $\lambda > 0$ (rate parameter)
- Probability Density Function (PDF):

$$f_X(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$$

  where Γ denotes the gamma function, a mathematical function that extends the concept of factorials to real and complex numbers.

- Support: $x \geq 0$

Mean and Variance of the Gamma Distribution

Mean (Expected Value)

The mean (expected value) of the Gamma distribution is given by:

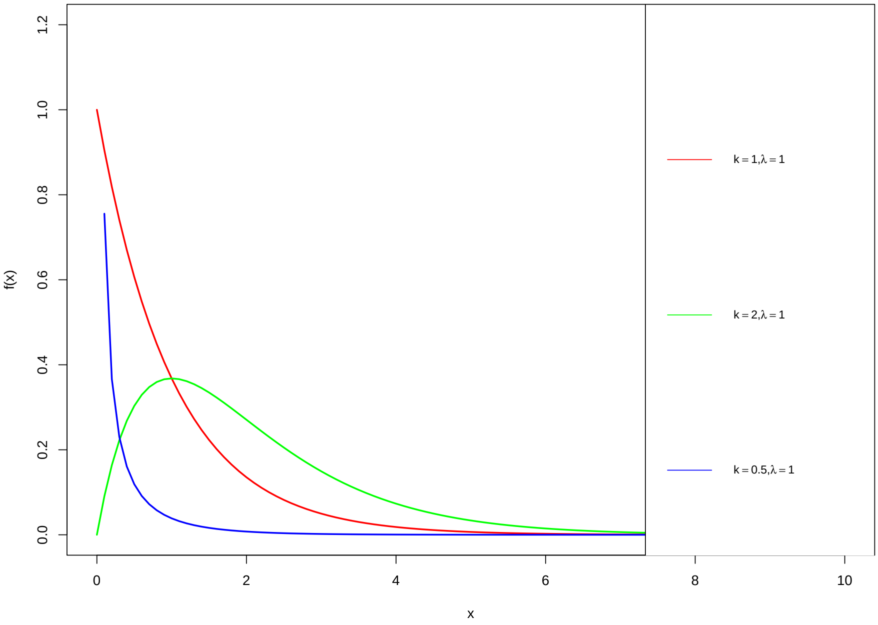
$$\mu_X = E(X) = \frac{k}{\lambda}$$

Variance

The variance of the Gamma distribution is given by:

$$\sigma_X^2 = Var(X) = \frac{k}{\lambda^2}$$

- Exponential distribution for $k = 1$
- Right-skewed for $k < 1$
- Left-skewed for $k > 1$



f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
2
4
6
8
10
k = 1,λ = 1
k = 2,λ = 1
k = 0.5,λ = 1

The Normal Distribution

The *Normal distribution*, also known as the Gaussian distribution, is a continuous probability distribution that is symmetric and bell-shaped. It is one of the most important probability distributions in statistics and is characterized by two parameters: the mean (μ) and the standard deviation (σ).

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- The probability density function of the Normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- It is symmetric with respect to the axis $x = \mu$
- It has a relative (and absolute) maximum at the point $x = \mu$.

The Normal Distribution

Mean (Expected Value)

The mean (expected value) of the Normal distribution is equal to its parameter μ:

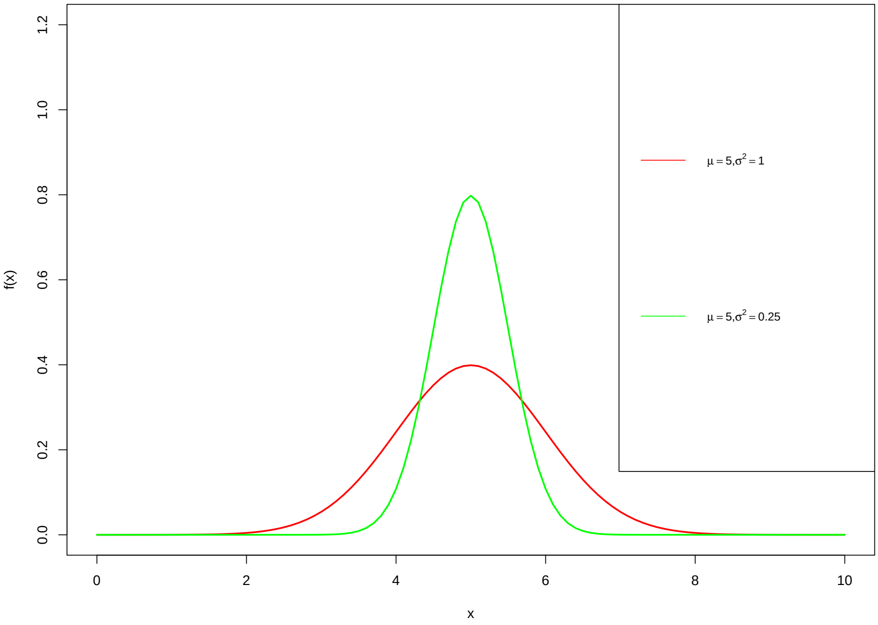
$$\mu_X = E(X) = \mu$$

Variance

The variance of the Normal distribution is equal to the square of its standard deviation σ:

$$\sigma_X^2 = Var(X) = \sigma^2$$

- Symmetric and bell-shaped.



f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
2
4
6
8
10
$\mu = 5, \sigma^2 = 1$
$\mu = 5, \sigma^2 = 0.25$

The importance of the Normal Distribution

- Many distributions encountered in the real world are not actually of the normal type, but are very well approximated by a Normal distribution. The Normal distribution is a model that adequately describes the distribution of numerous phenomena.
- Many distributions, even those far from the shape of the Normal distribution, may also be normalized through a variable transformation (e.g., $w = \log(x)$) as long as they are unimodal.
- The values produced by a measurement process are generally not normally distributed. When measuring the same object repeatedly, the instrument does not always produce the same value due to the so-called measurement error, which is the result of the sum of a large number of independent small factors that influence the process.
- The sum (or mean) of independent variables tends to follow a Normal distribution as n increases, regardless of the initial distributions (**Central Limit Theorem**).

Linear Combination of Random Variables

Consider two random variables, X and Y. A linear combination of these variables is given by:

$$Z = aX + bY$$

Where:

- Z is the linear combination of X and Y.
- a and b are constants.
- X and Y are random variables.

The expected value (mean) of the linear combination Z is calculated as:

$$E(Z) = a \cdot E(X) + b \cdot E(Y)$$

The variance of Z is calculated as:

$$Var(Z) = a^2 \cdot Var(X) + b^2 \cdot Var(Y) + 2ab \cdot Cov(X, Y)$$

Linear Combination of Random Variables

Consider two random variables, X and Y. A linear combination of these variables is given by:

$$Z = aX + bY$$

Where:

- Z is the linear combination of X and Y.
- a and b are constants.
- X and Y are random variables.

The expected value (mean) of the linear combination Z is calculated as:

$$E(Z) = a \cdot E(X) + b \cdot E(Y)$$

The variance of Z is calculated as:

$$Var(Z) = a^2 \cdot Var(X) + b^2 \cdot Var(Y) + 2ab \cdot Cov(X, Y)$$

Formal Definition

Definition

Let $X_1, X_2, \ldots, X_n$ be random variables, and let $a_1, a_2, \ldots, a_n$ be constants. The linear combination of these random variables is defined as:

$$Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

- Y is a new random variable.
- The a_i are constant coefficients ($a_i \in \mathbb{R}$).

Expectation of a Linear Combination

Expectation

The expectation (mean) of a linear combination $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ is calculated as:

$$E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$$

- The expectation of a linear combination is equal to the weighted sum of the expectations of the individual random variables.

Variance of a Linear Combination

Variance

The variance of a linear combination $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ is calculated as:

$$\mathrm{Var}(Y) = a_1^2\mathrm{Var}(X_1) + a_2^2\mathrm{Var}(X_2) + \ldots + a_n^2\mathrm{Var}(X_n) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

- The variance of a linear combination includes contributions from the variances of individual random variables and their pairwise covariances.