MTL108
Multivariate Random Variables and Independence

Rahul Singh

Bivariate and multiple random variables extend the univariate case to capture joint behaviors, enabling analysis of dependence (e.g., via correlation) and multivariate probabilities.

Definition 1 (Bivariate Discrete Random Variable)**.** A bivariate discrete random variable is a pair (X, Y) where X and Y are random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$. The joint behavior of X and Y is described by their joint cumulative distribution function (CDF):

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

The joint PMF $p_{X,Y}(x, y)$ provides the distribution over pairs of values, i.e.,

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y).$$

Marginal PMFs are obtained by summing over one variable,

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad p_Y(y) = \sum_x p_{X,Y}(x, y).$$

Example 1. Consider tossing three fair coins. Let X denote the number of heads and Y denote the number of tails. This can be visualized from the table below,

Outcomes	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	3	2	2	1	2	1	1	0
Y	0	1	1	2	1	2	2	3
Probability	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

The joint PMF $p_{X,Y}(x, y)$ is give by

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) = \begin{cases} 1/8, & \text{if } (x, y) \in \{(0, 3), (3, 0)\} \\ 3/8, & \text{if } (x, y) \in \{(1, 2), (2, 1)\} \\ 0, & \text{otherwise.} \end{cases}$$

The marginal PMF of X is given by

$$p_X(x) = \sum_y p_{X,Y}(x, y) = \begin{cases} 1/8, & \text{if } x \in \{0, 3\} \\ 3/8, & \text{if } x \in \{1, 2\} \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, the marginal PMF of Y is given by

$$p_Y(y) = \sum_x p_{X,Y}(x,y) = \begin{cases} 1/8, & \text{if } y \in \{0,3\} \\ 3/8, & \text{if } y \in \{1,2\} \\ 0, & \text{otherwise.} \end{cases}$$

Example 2. Consider rolling two fair dice. Let X and Y denote the outcomes on first and second die respectively. Then (X,Y) is a bivariate random variable; the table below provides a detailed view.

X\Y	**1**	**2**	**3**	**4**	**5**	**6**
1	(1,1) X=1, Y=1	(1,2) X=1, Y=2	(1,3) X=1, Y=3	(1,4) X=1, Y=4	(1,5) X=1, Y=5	(1,6) X=1, Y=6
2	(2,1) X=2, Y=1	(2,2) X=2, Y=2	(2,3) X=2, Y=3	(2,4) X=2, Y=4	(2,5) X=2, Y=5	(2,6) X=2, Y=6
3	(3,1) X=3, Y=1	(3,2) X=3, Y=2	(3,3) X=3, Y=3	(3,4) X=3, Y=4	(3,5) X=3, Y=5	(3,6) X=3, Y=6
4	(4,1) X=4, Y=1	(4,2) X=4, Y=2	(4,3) X=4, Y=3	(4,4) X=4, Y=4	(4,5) X=4, Y=5	(4,6) X=4, Y=6
5	(5,1) X=5, Y=1	(5,2) X=5, Y=2	(5,3) X=5, Y=3	(5,4) X=5, Y=4	(5,5) X=5, Y=5	(5,6) X=5, Y=6
6	(6,1) X=6, Y=1	(6,2) X=6, Y=2	(6,3) X=6, Y=3	(6,4) X=6, Y=4	(6,5) X=6, Y=5	(6,6) X=6, Y=6

The joint PMF of (X,Y) is

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) = \begin{cases} \frac{1}{36} & \text{if } x, y \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise.} \end{cases}$$

The marginal PMF of X

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y \mathbb{P}(X = x, Y = y) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the marginal PMF of Y

$$p_Y(y) = \sum_x p_{X,Y}(x,y) = \sum_x \mathbb{P}(X = x, Y = y) = \begin{cases} \frac{1}{6} & \text{if } y \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2 (Multiple Discrete Random Variables)**.** Multiple discrete random variables refer to a collection $(X_1, X_2, \ldots, X_n)$ of $n \geq 2$ random variables defined on the same probability space. Their joint behavior is described by the joint CDF:

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \mathbb{P}(X_1 \le x_1,\ldots,X_n \le x_n).$$

The joint PMF $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ is defined by

$$\mathbb{P}(X_1 = x_1,\ldots,X_n = x_n) = p_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

Marginal PMFs are obtained by summing over all but one variable, for example, marginal PMF of X_2, denoted by $p_{X_2}(x_2)$, is given by

$$\begin{aligned} p_{X_2}(x_2) &= \sum_{x_1}\sum_{x_3}\sum_{x_4}\cdots\sum_{x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \\ &= \sum_{x_1,x_3,x_4,\ldots,x_n} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n). \end{aligned}$$

Example 3. Consider rolling two fair dice. Let X and Y denote the outcomes on first and second die respectively, and define $Z = X + Y$. Then (X, Y, Z) is a trivariate random variable; the table below provides a detailed view.

X\Y	**1**	**2**	**3**	**4**	**5**	**6**
1	(1,1) X=1, Y=1 Z=2	(1,2) X=1, Y=2 Z=3	(1,3) X=1, Y=3 Z=4	(1,4) X=1, Y=4 Z=5	(1,5) X=1, Y=5 Z=6	(1,6) X=1, Y=6 Z=7
2	(2,1) X=2, Y=1 Z=3	(2,2) X=2, Y=2 Z=4	(2,3) X=2, Y=3 Z=5	(2,4) X=2, Y=4 Z=6	(2,5) X=2, Y=5 Z=7	(2,6) X=2, Y=6 Z=8
3	(3,1) X=3, Y=1 Z=4	(3,2) X=3, Y=2 Z=5	(3,3) X=3, Y=3 Z=6	(3,4) X=3, Y=4 Z=7	(3,5) X=3, Y=5 Z=8	(3,6) X=3, Y=6 Z=9
4	(4,1) X=4, Y=1 Z=5	(4,2) X=4, Y=2 Z=6	(4,3) X=4, Y=3 Z=7	(4,4) X=4, Y=4 Z=8	(4,5) X=4, Y=5 Z=9	(4,6) X=4, Y=6 Z=10
5	(5,1) X=5, Y=1 Z=6	(5,2) X=5, Y=2 Z=7	(5,3) X=5, Y=3 Z=8	(5,4) X=5, Y=4 Z=9	(5,5) X=5, Y=5 Z=10	(5,6) X=5, Y=6 Z=11
6	(6,1) X=6, Y=1 Z=7	(6,2) X=6, Y=2 Z=8	(6,3) X=6, Y=3 Z=9	(6,4) X=6, Y=4 Z=10	(6,5) X=6, Y=5 Z=11	(6,6) X=6, Y=6 Z=12

The joint PMF of (X, Y, Z) is

$$p_{X,Y,Z}(x,y,z) = \mathbb{P}(X = x, Y = y, Z = z) = \begin{cases} \frac{1}{36} & \text{if } x, y \in \{1,2,3,4,5,6\} \text{ and } z = x + y \\ 0 & \text{otherwise.} \end{cases}$$

The marginal PMF of X

$$p_X(x) = \sum_{y,z} p_{X,Y,Z}(x,y,z) = \sum_{y,z} \mathbb{P}(X = x, Y = y, Z = z) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the marginal PMF of Y

$$p_Y(y) = \sum_{x,z} p_{X,Y,Z}(x,y,z) = \sum_{x,z} \mathbb{P}(X = x, Y = y, Z = z) = \begin{cases} \frac{1}{6} & \text{if } y \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise.} \end{cases}$$

Next, the marginal PMF of Z

$$p_Z(z) = \sum_{x,y} p_{X,Y,Z}(x,y,z) = \sum_{x,y} \mathbb{P}(X = x, Y = y, Z = z) = \begin{cases} \frac{1}{36} & \text{if } z \in \{2,12\} \\ \frac{2}{36} & \text{if } z \in \{3,11\} \\ \frac{3}{36} & \text{if } z \in \{4,10\} \\ \frac{4}{36} & \text{if } z \in \{5,9\} \\ \frac{5}{36} & \text{if } z \in \{6,8\} \\ \frac{6}{36} & \text{if } z = 7 \\ 0 & \text{otherwise.} \end{cases}$$

The expectation of a random variable is a fundamental concept in probability theory, extending naturally to functions of multiple random variables.

Definition 3 (Expectation of a Function of Multivariate Random Variables)**.** Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of n random variables defined on a probability space, with a joint probability distribution. Let $g(\mathbf{X}) = g(X_1, X_2, \ldots, X_n)$ be a real-valued function of these random variables. If $\mathbf{X}$ has a joint probability mass function (PMF) $p_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$, the expectation of $g(\mathbf{X})$, denoted $\mathbb{E}[g(\mathbf{X})]$, is defined as

$$\mathbb{E}[g(\mathbf{X})] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) p_{\mathbf{X}}(x_1, x_2, \ldots, x_n),$$

where the summation is over all possible values $(x_1, x_2, \ldots, x_n)$ in the support of $\mathbf{X}$, provided the sum converges absolutely.

Example 4. Consider two discrete random variables X and Y with joint PMF $\mathbb{P}(X = x, Y = y)$. For $g(X, Y) = X + Y$, the expectation is

$$\mathbb{E}[X + Y] = \sum_{x} \sum_{y} (x + y) \mathbb{P}(X = x, Y = y).$$

Theorem 1 (linearity property)**.** *For constants a and b,*

$$\mathbb{E}[ag(\mathbf{X}) + bh(\mathbf{X})] = a\mathbb{E}[g(\mathbf{X})] + b\mathbb{E}[h(\mathbf{X})],$$

assuming the expectations exist.

Proof. The proof relies on the linearity of the expectation. Let $\mathbf{X}$ have joint PMF $p_{\mathbf{X}}(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)$. The expectation of the linear combination is

$$\mathbb{E}[ag(\mathbf{X}) + bh(\mathbf{X})] = \sum_{\mathbf{x}} [ag(\mathbf{x}) + bh(\mathbf{x})] p_{\mathbf{X}}(\mathbf{x}),$$

where the sum is over all possible $\mathbf{x}$ in the support. By linearity of summation,

$$\mathbb{E}[ag(\mathbf{X}) + bh(\mathbf{X})] = a\sum_{\mathbf{x}} g(\mathbf{x})p_{\mathbf{X}}(\mathbf{x}) + b\sum_{\mathbf{x}} h(\mathbf{x})p_{\mathbf{X}}(\mathbf{x}) = a\mathbb{E}[g(\mathbf{X})] + b\mathbb{E}[h(\mathbf{X})].$$

□

Lemma 1. *If $g(\mathbf{X}) = X_i$ for some i, the expectation reduces to the marginal expectation $\mathbb{E}[X_i]$.*

Definition 4 (Continuous Bivariate Random Variable)**.** A bivariate random variable is a pair (X, Y) where X and Y are random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$. The joint behavior of X and Y is described by their joint cumulative distribution function (CDF):

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

The joint probability density function (PDF) $f_{X,Y}(x, y)$ (for continuous case) or probability mass function (PMF) $p_{X,Y}(x, y)$ (for discrete case) provides the distribution over pairs of values. Marginal distributions are obtained by integrating or summing over one variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

(for continuous), or $p_X(x) = \sum_y p_{X,Y}(x, y)$ (for discrete).

Example 5 (Bivariate Uniform Distribution)**.** Consider two random variables X and Y uniformly distributed over the unit square $[0, 1] \times [0, 1]$.
- Joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

- Marginal PDFs:

$$f_X(x) = \int_0^1 1\, dy = 1 \quad (0 \leq x \leq 1), \quad f_Y(y) = \int_0^1 1\, dx = 1 \quad (0 \leq y \leq 1).$$

- Probability $\mathbb{P}(X + Y \leq 1)$: This is the area of the triangle from $(0, 0)$ to $(1, 0)$ to $(0, 1)$, which is $\int_0^1 \int_0^{1-x} 1\, dy\, dx = \int_0^1 (1 - x)\, dx = [x - x^2/2]_0^1 = 1 - 1/2 = 1/2$.

Application: Modeling the coordinates of a randomly chosen point in a square region.

Multiple Random Variables

Definition 5 (Multiple Random Variables)**.** Multiple random variables refer to a collection $(X_1, X_2, \ldots, X_n)$ of $n \geq 2$ random variables defined on the same probability space. Their joint behavior is described by the joint CDF:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

For continuous variables, the joint PDF $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ satisfies:

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = \int_A f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, d\mathbf{x},$$

where A is a region in $\mathbb{R}^n$. Marginal distributions are obtained by integrating over all but one variable, for example, marginal PDF of X_2, denoted by $f_{X_2}(x_2)$, is given by

$$\begin{aligned}
f_{X_2}(x_2) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1\, dx_3\, dx_4 \cdots dx_n \\
&= \int_{x_1} \int_{x_3} \int_{x_4} \cdots \int_{x_n} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1\, dx_3\, dx_4 \cdots dx_n \\
&= \int_{x_1, x_3, x_4, \ldots, x_n} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1\, dx_3\, dx_4 \cdots dx_n \\
&= \int_{\mathbb{R}^{n-1}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1\, dx_3\, dx_4 \cdots dx_n.
\end{aligned}$$

Definition 6 (Expectation of a Function of Multivariate Random Variables)**.** Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of n random variables defined on a probability space, with a joint probability distribution. Let $g(\mathbf{X}) = g(X_1, X_2, \ldots, X_n)$ be a real-valued function of these random variables.

Discrete Case: If $\mathbf{X}$ has a joint probability mass function (PMF) $p_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$, the expectation of $g(\mathbf{X})$, denoted $\mathbb{E}[g(\mathbf{X})]$, is defined as

$$\mathbb{E}[g(\mathbf{X})] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) p_{\mathbf{X}}(x_1, x_2, \ldots, x_n),$$

where the summation is over all possible values $(x_1, x_2, \ldots, x_n)$ in the support of $\mathbf{X}$, provided the sum converges absolutely.

Continuous Case: If $\mathbf{X}$ has a joint probability density function (PDF) $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$, the expectation is

$$\mathbb{E}[g(\mathbf{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n,$$

where the multiple integral is taken over the support of $\mathbf{X}$, provided the integral converges absolutely.

In both cases, the expectation exists if the sum or integral of $|g(\mathbf{X})|$ with respect to the joint distribution is finite.

Theorem 2 (linearity property)**.** *For constants a and b,*

$$\mathbb{E}[ag(\mathbf{X}) + bh(\mathbf{X})] = a\mathbb{E}[g(\mathbf{X})] + b\mathbb{E}[h(\mathbf{X})],$$

assuming the expectations exist.

Proof. We proved this result for the discrete case in Topic-8, so we need to prove it for the continuous case. Let $\mathbf{X}$ have joint PDF $f_{\mathbf{X}}(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)$. The expectation of the linear combination is

$$\mathbb{E}[ag(\mathbf{X}) + bh(\mathbf{X})] = \int_{x_1,\ldots,x_n} [ag(\mathbf{x}) + bh(\mathbf{x})] f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n,$$

where the integral is over the support of $\mathbf{X}$. By linearity of integration,

$$\begin{aligned}\mathbb{E}[ag(\mathbf{X}) + bh(\mathbf{X})] &= a \int_{x_1,\ldots,x_n} g(\mathbf{x}) f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \\ &\quad + b \int_{x_1,\ldots,x_n} h(\mathbf{x}) f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= a\mathbb{E}[g(\mathbf{X})] + b\mathbb{E}[h(\mathbf{X})].\end{aligned}$$

□

Lemma 2. *If $g(\mathbf{X}) = X_i$ for some i, the expectation reduces to the marginal expectation $\mathbb{E}[X_i]$.*

Example 6 (Trivariate Normal Distribution)**.** Consider three random variables (X, Y, Z) following a trivariate normal distribution with mean vector $\mu = (0, 0, 0)$ and a covariance matrix ensuring independence (e.g., diagonal with variances 1).

- **Joint PDF (for independent normals):**

$$f_{X,Y,Z}(x, y, z) = \frac{1}{(2\pi)^{3/2}} e^{-\frac{x^2+y^2+z^2}{2}}.$$

- **Marginal PDFs:** Each X, Y, Z is $N(0, 1)$, i.e., $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. - Probability $\mathbb{P}(X^2+Y^2+Z^2 \le 1)$: This is the volume of a unit ball in 3D, approximately 0.5236, computed via $\int_{x^2+y^2+z^2 \le 1} f_{X,Y,Z}(x, y, z)\, d\mathbf{x}$.

Application: Modeling the positions of particles in a 3D space with independent Gaussian noise.

Independent Random Variable

Independence is a key concept in probability theory that describes random variables whose outcomes do not influence each other. This property simplifies calculations, such as finding joint distributions or expectations of products.

Definition 7 (Independent Random Variables)**.** Two random variables X and Y are independent if the events $\{X \le x\}$ and $\{Y \le y\}$ are independent for all real numbers x and y. Equivalently, their joint cumulative distribution function (CDF) factors into the product of their marginal CDFs:

$$F_{X,Y}(x, y) = \mathbb{P}(X \le x, Y \le y) = F_X(x) F_Y(y),$$

for all $x, y \in \mathbb{R}$.

For discrete random variables with probability mass functions (PMFs) $p_X(x)$ and $p_Y(y)$, independence implies the joint PMF is the product:

$$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$

For continuous random variables with probability density functions (PDFs) $f_X(x)$ and $f_Y(y)$, independence implies the joint PDF is the product:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

More generally, a collection of random variables $X_1, X_2, \ldots, X_n$ is mutually independent if the joint CDF (or PMF/PDF) factors into the product of the individual marginals for any subset.

Example 7 (Independent Coin Flips)**.** Consider two independent fair coin flips, where X is 1 for heads on the first flip (0 for tails), and Y is 1 for heads on the second flip (0 for tails). Both X and Y are Bernoulli(0.5).
The joint PMF is:

$$p_{X,Y}(x,y) = \frac{1}{4}, \quad \text{for } (x,y) \in \{(0,0),(0,1),(1,0),(1,1)\}.$$

This factors as $p_X(x)p_Y(y) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$, confirming independence.
Application: Modeling independent binary outcomes, like success in separate trials.

Example 8 (Independent Dice Rolls)**.** Let X and Y be the outcomes of two independent fair six-sided dice rolls. Each is uniform on $\{1,2,3,4,5,6\}$.
The joint PMF is:

$$p_{X,Y}(x,y) = \frac{1}{36}, \quad x,y = 1,\ldots,6.$$

This is the product $\frac{1}{6} \times \frac{1}{6}$, so X and Y are independent.
Contrast: If $Z = X + Y$, then X and Z are dependent, as knowing X affects the distribution of Z.

Example 9 (Bivariate Uniform Distribution)**.** Consider two random variables X and Y uniformly distributed over the unit square $[0,1] \times [0,1]$.
- Joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

- Marginal PDFs:

$$f_X(x) = \int_0^1 1\,dy = 1 \quad (0 \le x \le 1), \quad f_Y(y) = \int_0^1 1\,dx = 1 \quad (0 \le y \le 1).$$

- Notice that

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x) \cdot f_Y(y) \text{ for all } x,y \in \mathbb{R} \\ &= \begin{cases} 1 & \text{if } 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, random variables X and Y are independent.

Theorem 3. *For real valued (nice) functions g and h, if X and Y are independent (discrete or continuous) random variables then*

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)],$$

provided the expectations exist.

Proof. We proved this result for discrete case in Topic-8, so we need to prove for continuous case. Let X, Y have joint PDF $f_{X,Y}(x, y)$. The expectation is

$$\mathbb{E}[g(X)h(Y)] = \int_x \int_y g(x)h(y)\, f_{X,Y}(x, y)\, dx\, dy.$$

where the sum is over all possible x, y in the support. By independence assumption, $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Consequently, factoring out terms with x and Y, we have

$$\mathbb{E}[g(X)h(Y)] = \int_x \int_y g(x)h(y)\, f_X(x)f_Y(y)\, dx\, dy = \left(\int_x g(x)\, f_X(x)\, dx\right) \times \left(\int_y h(y)\, f_Y(y)\, dy\right).$$

Notice that

$$\left(\int_x g(x)\, f_X(x)\, dx\right) = \mathbb{E}[g(X)] \quad \text{and} \quad \left(\int_y h(y)\, f_Y(y)\, dy\right) = \mathbb{E}[h(Y)].$$

Therefore,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].$$

□

Theorem 4. *If X and Y are independent random variables, and define $M_Z(t) = \mathbb{E}(e^{tZ})$ then*

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Proof. Observe that

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX} \cdot e^{tY}).$$

Since e^{tX} is a function of X only and e^{tY} is a function of Y only, using independence and Theorem 2 above, we have

$$M_{X+Y}(t) = \mathbb{E}(e^{tX} \cdot e^{tY}) = \mathbb{E}(e^{tX}) \cdot \mathbb{E}(e^{tY}) = M_X(t) \cdot M_Y(t).$$

□

Covariance

Similar to variance we have a measure for together-variability of two random variables, known as covariance.

Definition 8 (Covariance)**.** Let X and Y be two random variables, then the covariance between them is denoted by $\text{Cov}(X, Y)$ and defined as

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

Theorem 5. *For any two random variables X and Y,*

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

provided the expectations $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$ exist.

Proof. We have from definition

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Expand the expression inside the expectation

$$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y].$$

Take the expectation of both sides, we have

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X\mathbb{E}[Y]] - \mathbb{E}[Y\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]], \text{ using linearity of expectation.}\end{aligned}$$

Now, $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are constants, so

$$\begin{aligned}\mathbb{E}[X\mathbb{E}[Y]] &= \mathbb{E}[X] \cdot \mathbb{E}[Y], \\ \mathbb{E}[Y\mathbb{E}[X]] &= \mathbb{E}[Y] \cdot \mathbb{E}[X] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \\ \text{and} \quad \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]] &= \mathbb{E}[X] \cdot \mathbb{E}[Y]\end{aligned}$$

Substitute these into the above equation, we get

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y].$$

Thus,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

□

Theorem 6. *If X and Y are independent, then $Cov(X,Y) = 0$.*

Proof. Using indepence of X and Y, and Theorem 2 we have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Therefore using Theorem 3, we have

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0.\end{aligned}$$

□

Theorem 7. *For any two random variables X and Y,*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

Additionally, if X and Y are independent random variables, then

$$Var(X+Y) = Var(X) + Var(Y).$$

Proof. Using definition,

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

So, using linearity of expectation, we have

$$\mathrm{Var}(X+Y) = \mathbb{E}[(X-\mathbb{E}(X))^2]+\mathbb{E}[(Y-\mathbb{E}(Y))^2]+2\mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))].$$

Now using definitions of variance and covariance, that is, $\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}(X))^2]$, $\mathrm{Var}(Y) = \mathbb{E}[(Y-\mathbb{E}(Y))^2]$ and $\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))]$, we have

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)+2\,\mathrm{Cov}(X,Y).$$

This proves the first part. Next, if X and Y are independent, from Theorem 4, we have $\mathrm{Cov}(X,Y) = 0$, substituting in above expression, we get

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y).$$

□

Remark 1. Covariance measures linear dependence; independence implies zero covariance, but not vice versa.

Theorem 8. *Let $X_1,\ldots,X_n$ be random variables defined on the same probability space. Then, for any constant $a \in \mathbb{R}$ and $i \in \{1,\ldots,n\}$, , $Var(aX_i) = a^2\,Var(X_i)$, and*

$$\begin{aligned}Var(X_1+X_2+\cdots+X_n) &= \sum_{i=1}^{n} Var(X_i) + \sum_{1\le i\ne j\le n} Cov(X_i,X_j) \\ &= \sum_{i=1}^{n} Var(X_i) + 2\sum_{1\le i<j\le n} Cov(X_i,X_j)\end{aligned}$$

Moreover, if $X_1,\ldots,X_n$ are independent then

$$Var(X_1+X_2+\cdots+X_n) = \sum_{i=1}^{n} Var(X_i).$$

Proof. By the definition of variance, $\mathrm{Var}(Y) = \mathbb{E}[Y^2]-(\mathbb{E}[Y])^2$. So, for any $i \in \{1,\ldots,n\}$,

$$\begin{aligned}\mathrm{Var}(aX_i) &= \mathbb{E}[(aX_i)^2]-(\mathbb{E}[aX_i])^2 \\ &= \mathbb{E}[a^2X_i^2]-(a\mathbb{E}[X_i])^2 \\ &= a^2\mathbb{E}[X_i^2]-a^2(\mathbb{E}[X_i])^2 \\ &= a^2\left(\mathbb{E}[X_i^2]-(\mathbb{E}[X_i])^2\right) \\ \Rightarrow \mathrm{Var}(aX_i) &= a^2\mathrm{Var}(X_i)\end{aligned}$$

First part is proved. Next, let $S_n = X_1 + \cdots + X_n$. By the definition of variance:

$$\mathrm{Var}(S_n) = \mathbb{E}[S_n^2] - (\mathbb{E}[S_n])^2$$

Using the linearity of expectation, we have $\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i]$. For S_n^2, we write:

$$S_n^2 = \left(\sum_{i=1}^n X_i\right)\left(\sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n X_i X_j$$

Taking the expectation:

$$\mathbb{E}[S_n^2] = E\left[\sum_{i=1}^n \sum_{j=1}^n X_i X_j\right] = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i X_j]$$

Now we can express the variance of the sum:

$$\begin{aligned}
\mathrm{Var}(S_n) &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i X_j] - \left(\sum_{i=1}^n \mathbb{E}[X_i]\right)\left(\sum_{j=1}^n \mathbb{E}[X_j]\right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i X_j] - \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i]\mathbb{E}[X_j] \\
&= \sum_{i=1}^n \sum_{j=1}^n (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]) \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

The double summation can be split into terms where $i = j$ and $i \neq j$:

$$\mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Cov}(X_i, X_i) + \sum_{1 \le i \neq j \le n} \mathrm{Cov}(X_i, X_j)$$

Since $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$, the expression becomes:

$$\mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{1 \le i \neq j \le n} \mathrm{Cov}(X_i, X_j)$$

Due to the symmetry $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$, the second term can be rewritten:

$$\sum_{1 \le i \neq j \le n} \mathrm{Cov}(X_i, X_j) = 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j)$$

This gives the second form of the identity.

Finally, if $X_1, \ldots, X_n$ are independent, then for any $i \neq j$, $\mathrm{Cov}(X_i, X_j) = 0$. Substituting this into the general formula:

$$\mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} 0 = \sum_{i=1}^n \mathrm{Var}(X_i).$$

$\square$

Correlation

Definition 9 (Correlation)**.** Let X and Y be two random variables, then the correlation coefficient between them is denoted by $\mathrm{Corr}(X, Y)$ and defined as The correlation coefficient is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Theorem 9. $\mathrm{Corr}(X, Y)$ *ranges from -1 to 1.*

Correlation normalizes covariance to measure the strength and direction of a linear relationship.

Positive and Negative Correlation

- Positive Correlation ($\mathrm{Corr}(X, Y) > 0$): As X increases, Y tends to increase (e.g., height and weight).
- Negative Correlation ($\mathrm{Corr}(X, Y) < 0$): As X increases, Y tends to decrease (e.g., hours studied and errors on test).

Example 10 (Positive)**.** Height (X) and weight (Y) in adults: $\rho \approx 0.7$, positive.

Example 11 (Negative)**.** Price (X) and demand (Y) for a product: As price rises, demand falls, $\rho < 0$.

Remark 2. Zero correlation means no linear relationship, but nonlinear dependence may exist.

Example 12. Consider discrete random variables X and Y with their PMF

$$p_{X,Y}(x, y) = \begin{cases} 1/4, & \text{if} \quad x = 1, y = 1 \\ 1/4, & \text{if} \quad x = -1, y = 1 \\ 1/2, & \text{if} \quad x = 0, y = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Then, the marginal PMF of X is

$$p_X(x) = \sum_y p_{X,Y}(x, y) = \begin{cases} 1/4, & \text{if} \quad x = 1 \\ 1/4, & \text{if} \quad x = -1 \\ 1/2, & \text{if} \quad x = 0 \\ 0, & \text{otherwise.} \end{cases}$$

and the marginal PMF of Y is

$$p_Y(y) = \sum_x p_{X,Y}(x, y) = \begin{cases} 1/2, & \text{if} \quad y = 1 \\ 1/2, & \text{if} \quad y = 0 \\ 0, & \text{otherwise.} \end{cases}$$

So,

$$\mathbb{E}(XY) = 1 \cdot 1 \cdot \frac{1}{4} + (-1) \cdot 1 \cdot \frac{1}{4} + 0 \cdot 0 \cdot \frac{1}{2} = \frac{1}{4} - \frac{1}{4} = 0,$$

$$\mathbb{E}(X) = 1 \cdot \frac{1}{4} + (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} = \frac{1}{4} - \frac{1}{4} = 0,$$
$$\mathbb{E}(Y) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.$$

Thus,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 \quad \Rightarrow \mathrm{Corr}(X, Y) = 0.$$

However, $Y = X^2$ is the perfect quadratic relationship between X and Y.

Result: Verify that here X and Y are not independent. This means zero correlation (covariance) does not imply independence.

Conditional Random Variables

In probability theory, a conditional random variable is a random variable whose probability distribution depends on the outcome of another random variable or event, extending the concept of conditional probability. If X and Y are random variables, $X|Y$ represents the distribution of X for each possible value of Y.

Discrete Conditional Random Variables

For discrete random variables X and Y with joint PMF $p_{X,Y}(x, y)$, the conditional PMF of X given $Y = y$ is defined as:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad \text{for } p_Y(y) > 0$$

where $p_Y(y)$ is the marginal PMF of Y.

Example 13 (Flipping three fair coins)**.** Consider the experiment of flipping a fair coin three times. The sample space Ω consists of 8 equally likely outcomes.

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Let X be a random variable representing the number of heads, and let Y be an indicator random variable such that $Y = 1$ if the first flip is heads, and $Y = 0$ otherwise. We want to find the conditional PMF of X given that $Y = 1$.

1. Marginal PMF

The event $Y = 1$ corresponds to the set of outcomes where the first flip is a head.

$$\{Y = 1\} = \{HHH, HHT, HTH, HTT\}$$

The marginal probability of this event is:

$$p_Y(1) = \mathbb{P}(Y = 1) = \frac{|\{Y = 1\}|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}$$

2. Joint PMF values for $Y = 1$

Next, we find the joint probabilities $p_{X,Y}(x, 1)$ for the possible values of X. Since the event $\{Y = 1\}$ has already occurred, X can only take values $\{1, 2, 3\}$.

- $p_{X,Y}(1, 1) = \mathbb{P}(X = 1 \text{ and } Y = 1)$: The outcome is HTT.

$$p_{X,Y}(1, 1) = \frac{1}{8}$$

- $p_{X,Y}(2, 1) = \mathbb{P}(X = 2 \text{ and } Y = 1)$: The outcomes are HHT and HTH.

$$p_{X,Y}(2, 1) = \frac{2}{8}$$

- $p_{X,Y}(3, 1) = \mathbb{P}(X = 3 \text{ and } Y = 1)$: The outcome is HHH.

$$p_{X,Y}(3, 1) = \frac{1}{8}$$

3. Conditional PMF

The conditional PMF of X given $Y = 1$ is defined as $p_{X|Y}(x|1) = \frac{p_{X,Y}(x,1)}{p_Y(1)}$. For each value of $x \in \{1, 2, 3\}$:

- For $x = 1$:

$$p_{X|Y}(1|1) = \frac{p_{X,Y}(1, 1)}{p_Y(1)} = \frac{1/8}{1/2} = \frac{1}{4}$$

- For $x = 2$:

$$p_{X|Y}(2|1) = \frac{p_{X,Y}(2, 1)}{p_Y(1)} = \frac{2/8}{1/2} = \frac{2}{4} = \frac{1}{2}$$

- For $x = 3$:

$$p_{X|Y}(3|1) = \frac{p_{X,Y}(3, 1)}{p_Y(1)} = \frac{1/8}{1/2} = \frac{1}{4}$$

Summary of the Conditional PMF

The conditional PMF can be summarized in a table:

x	1	2	3
$p_{X\|Y}(x\|1)$	1/4	1/2	1/4

Note that $\sum_x p_{X|Y}(x|1) = 1/4 + 1/2 + 1/4 = 1$, as required for a valid PMF.

Continuous Conditional Random Variables

For continuous random variables X and Y with joint PDF $f_{X,Y}(x, y)$, the conditional PDF of X given $Y = y$ is:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad \text{for } f_Y(y) > 0$$

where $f_Y(y)$ is the marginal PDF of Y.

Example 14 (Conditional distribution of uniform variables)**.** Given the joint PDF of continuous random variables X and Y:

$$f_{X,Y}(x, y) = \begin{cases} \frac{3}{2} & \text{for } x^2 \leq y \leq 1, \text{ and } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

We want to find the conditional PDF of Y given $X = x$.

1. Marginal PDF of X

The marginal PDF of X, denoted by $f_X(x)$, is found by integrating the joint PDF with respect to y over its full range. For a given x where $0 < x < 1$, the function is non-zero only for y between x^2 and 1.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \\ &= \int_{x^2}^{1} \frac{3}{2}\, dy \\ &= \frac{3}{2}[y]_{x^2}^{1} \\ &= \frac{3}{2}(1 - x^2) \end{aligned}$$

This marginal PDF is valid for $0 < x < 1$, and $f_X(x) = 0$ otherwise.

2. Conditional PDF of Y given $X = x$

The conditional PDF of Y given $X = x$, denoted by $f_{Y|X}(y|x)$, is given by the formula $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, for any x where $f_X(x) > 0$. Substituting the joint and marginal PDFs:

$$f_{Y|X}(y|x) = \frac{\frac{3}{2}}{\frac{3}{2}(1 - x^2)} = \frac{1}{1 - x^2}$$

This conditional PDF is defined for the range of y where the joint PDF is non-zero, which is $x^2 \leq y \leq 1$.

3. Conclusion

The result shows that for a fixed value of x, the conditional distribution of Y is uniform over the interval $[x^2, 1]$.

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x^2} & \text{for } x^2 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The length of this interval is $(1 - x^2)$, and the height of the uniform density is $\frac{1}{\text{length}}$, which confirms that it is a valid uniform distribution.

Conditional Expectation

The conditional expectation of a random variable is its expected value with respect to its conditional distribution.

Definition 10 (Conditional Expectation)**.** For discrete and continuous cases, conditional expectation are defind as:

- **Discrete:** $E[X|Y = y] = \sum_x x \cdot p_{X|Y}(x|y)$
- **Continuous:** $E[X|Y = y] = \int_x x \cdot f_{X|Y}(x|y)dx$

$E[X|Y]$ is a random variable that is a function of Y.

Theorem 10. *The Law of Total Expectation states* $E[X] = E[E[X|Y]]$.

Example 15 (Three coins example)**.** Let X be the number of heads in three fair coin flips, and Y be an indicator random variable such that $Y = 1$ if the first flip is a head, and $Y = 0$ otherwise. We want to find the conditional expectation $\mathbb{E}[X|Y = 1]$.
First, we use the conditional PMF $p_{X|Y}(x|1)$ derived from the example. The formula for the conditional expectation of a discrete random variable is:

$$\mathbb{E}[X|Y = 1] = \sum_x x \cdot p_{X|Y}(x|1)$$

Substituting the values of the conditional PMF for $x \in \{1, 2, 3\}$:

$$\begin{aligned} \mathbb{E}[X|Y = 1] &= (1) \cdot p_{X|Y}(1|1) + (2) \cdot p_{X|Y}(2|1) + (3) \cdot p_{X|Y}(3|1) \\ &= (1) \cdot \frac{1}{4} + (2) \cdot \frac{1}{2} + (3) \cdot \frac{1}{4} \\ &= \frac{1}{4} + 1 + \frac{3}{4} \\ &= \frac{4}{4} + 1 \\ &= 2 \end{aligned}$$

The calculation using the conditional PMF yields an expected value of 2.

Definition 11 (Identically distributed RVs)**.** Let random variables X and Y have CDFs $F_X(\cdot)$ and $F_Y(\cdot)$, respectively. The Rvs X and Y are identically distributed (or are equal in distribution) is denoted by

$$X \stackrel{d}{=} Y,$$

and defined as

$$F_X(a) = F_X(a) \quad \text{for all} \quad a \in \mathbb{R}.$$

For discrete RVs, equality of PMFs gives equality of distribution, and for continuous RVs, equality of PDFs gives equality of distribution.

The random variables $X_1, X_2, \ldots, X_n$ are identically distributed, can be written as,

$$X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \ldots \stackrel{d}{=} X_n.$$

Example 16. For example, if they all follow a normal distribution with mean 0 and variance σ^2, i.e.,

$$X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, \sigma^2),$$

then

$$X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \ldots \stackrel{d}{=} X_n.$$

IID Random Variables

Definition 12 (IID Random Variables)**.** Independent and Identically Distributed (IID) random variables are a sequence $X_1, X_2, \ldots$ that are independent and each has the same distribution.

IID is common in sampling, e.g., multiple independent trials from the same distribution. Let $X_1, X_2, \ldots, X_n$ be a sequence of n random variables that are independent and identically distributed (IID). This means:

- **Independent:** The outcome of any single variable does not influence the outcome of the others.
- **Identically Distributed:** All the variables are drawn from the same probability distribution. Consequently, they share the same mean (μ) and variance (σ^2).

This immediately follows:

- $\mathbb{E}[X_i] = \mu$ for all $i = 1, \ldots, n$.
- $\text{Var}(X_i) = \sigma^2$ for all $i = 1, \ldots, n$.

Example 17. Repeated fair die rolls: Each X_i uniform on {1,2,3,4,5,6}, independent.

Sample mean

The sample average, or sample mean, of these variables is denoted as $\bar{X}$ and

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Theorem 11. *Let $X_1, \ldots, X_n$ are IID random variables with $\mathbb{E}(X_1) = \mu$ and Var$(X_i) = \sigma^2$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\mathbb{E}[\bar{X}] = \mu$.*

Proof. This is derived using the linearity of expectation, which states that the expectation of a sum is the sum of expectations.

$$\begin{aligned}
\mathbb{E}[\bar{X}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu \\
&= \frac{1}{n}(n\mu) \\
\Rightarrow \mathbb{E}[\bar{X}] &= \mu
\end{aligned}$$

□

Theorem 12. *Let $X_1, \ldots, X_n$ are IID random variables with $\mathbb{E}(X_1) = \mu$ and $Var(X_i) = \sigma^2$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then,* $\mathrm{Var}(\bar{X}) = \sigma^2/n$.

Remark 3. This property demonstrates that as the sample size increases, the spread of the sample mean decreases.

Proof. We use Theorem 8 to find the variance of $\bar{X}$. Observe that

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) \\
\Rightarrow \mathrm{Var}(\bar{X}) &= \frac{\sigma^2}{n}.
\end{aligned}$$

□

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer