Lecture 1

Mathematical Preliminaries

Ananya Iyengar

January 2026

St. Stephen's College

- **Random Variable**: A function that maps the outcome of a random experiment to a real number. Denoted by X, Y, etc. Foe e.g., number of times a head is observed in 5 flips of a coin.
- When the domain of a random variable is finite, it is discrete and when the domain is infinite, it is continuous.
- A **probability mass function** of a discrete RV X ia a function that returns probability $P(X = x)$ for each x in the domain of X.
- A **probability density function** of a continuous RV X operates the same as above, but instead of returning probabilities at a given point value of X, it returns probabilities within a range of realisations of X *(think about why!)*.
- Let E be the expectation operator, which is the **Expected Value** i.e. the weighted average of the RV X where the weights are according to the probabilities assigned to outcomes of X. Formally, $E(X) = \sum_x xf(x)$, where $f(\cdot)$ is the probability mass function of X. In case of a continuous RV, replace the summation by an integral. We denote the expected value by μ.
- The expectations operator has the following properties:

  1. $E(a) = a;\ a \in \Re$
  2. $E(aX) = aE(X);\ a \in \Re$ and X is a RV.
  3. $E(X + Y) = E(X) + E(Y)$ where X, Y are RVs.

4. $E(XY) = E(X)E(Y)$ iff X, Y are independent.

- The **Variance** of a RV is a weighted average of the squared deviation between a random variable and its expectation, where the weights are defined according to the probability distribution. We denote the variance by σ^2 (and the standard deviation by σ) where $\sigma^2 = E[(X - \mu)^2]$, where $\mu \equiv E(X)$.

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

- The variance operator is denoted by $V(\cdot)$ and has the following properties:

  1. $V(a) = 0;\ a \in \Re$
  2. $V(aX) = a^2V(X);\ a \in \Re$ and X is a RV.
  3. $V(X+Y) = V(X)+V(Y)+2Cov(X,Y)$, where X, Y are RVs. IF X and Y are independent, then $Cov(X,Y) = 0$.

- A **sample** is drawn from a population. Each unit of the sample is denoted by the subscript i. For example, if we collect data on heights (the RV, denoted by X) for 10 individuals, our data would be of the form $(X_1, X_2, ...X_i, ...X_{10})$.

- A **random sample** is **independently and identically distributed** (i.i.d). For example, successive tosses of a fair coin.

- Types of Economic Data:

  1. Cross-sectional: Many units (i) at a given point in time.
  2. Time Series: One unit across many time periods (t).
  3. Panel: Varies across i and t.