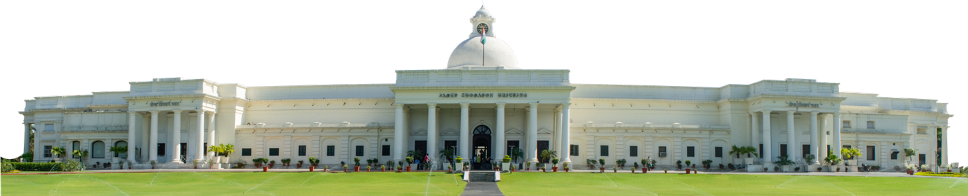
# Data Mining for Business Intelligence (IBM 312)

Sumit Kumar Yadav

Department of Management Studies

January 17, 2023



# Random Variable

$\{(1,1), (1,2), \ldots, (6,6)\}$

$2 \quad 3 \quad \ldots \quad 12$

- Definition - A mapping from sample space to real numbers
- Expectation, Variance, Correlation for Random Variables
- Examples - Throwing two dice, Sum of the two throws; Number of students who would come to 8am class

$$P(X=3) = ?? \quad \left(\frac{2}{36}\right)$$

# Distribution of a Random Variable

- Let $X$ be the random variable which can take values $x_1, x_2, ....$
- The function $P(X = x_i) = f(x_i)$ is called the distribution (probability distribution) of $X$
- Joint Distribution of Random Variables

# Basics about Random Variable

- A variable whose value depends on outcome of a random phenomenon
- A random variable is characterized by its distribution
- Sum of two or more different random variables is also a random variable, and thus will also have a distribution (we might not cover the mathematical tools required to find the distribution, but it is important to appreciate that it will have a distribution)
- Similarly, any other algebraic operation of two or more random variables also remain a random variable
- If $X$ and $Y$ are random variables, $X + Y, X - Y, XY, \frac{X}{Y}$ are all random variables

# Expectation, Variance and Covariance Definition

# Definition of Expectation

1, 2, 3, 4, 5, 6

$\boxed{3.5}$

$\left(\frac{1}{6}\right)$ $\left(\frac{1}{6}\right)$ - $\left(\frac{1}{6}\right)$

0.01 0.01 — 0.01 0.95

The expected value of a discrete random variable is

$$\mu = E(X) = \sum_x x\, p_X(x)$$

$$\frac{6+4+5+\_\_ + \cdots}{6000} = 3.5$$

# Variance of a random variable $X$

Let $E(X) = \mu$ (The Greek letter “mu").

$$Var(X) = E\left((X - \mu)^2\right)$$

# Definition of Covariance

$\boxed{E(XY) - \mu_X \mu_Y}$

Let $X$ and $Y$ be jointly distributed random variables with $E(X) = \mu_x$ and $E(Y) = \mu_y$. The *covariance* between $X$ and $Y$ is

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

- You could think of $Var(X) = E[(X - \mu_x)^2]$ as $Cov(X, X)$.

$\rightarrow E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y)$

$= E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y$

## Examples

$X_i = \begin{cases} 1, & \text{if } i^{th} \text{ student gets their phone} \\ 0, & \text{otherwise} \end{cases}$

"N" students

- Number of matching - I take your mobile phones and return the mobile phones randomly back. How many students get their own mobile phone back? (X = this random variable). Find E(X) and Var(X)
- Waiting time to get *r* unique objects - *N* different objects in a box. In each step, take out one object at random and keep it back. Repeat this until you get *r* unique objects. X = no. of trials required. Find E(X)
- Largest number in *n* drawings. A box contains balls numbered 1,2,...,N. Let *X* be the largest number drawn in *n* drawings, (done with replacement). Find E(X)

# Standard Distributions

- Discrete
  1. Binomial
  2. Poisson
  3. Geometric
- Continuous
  1. Normal
  2. Uniform
  3. Exponential

# Simulation of Random numbers in Python

# Learning Outcomes

- Standard discrete and Continuous Distributions
- Binomial, Poisson, Geometric, Normal, Uniform, Exponential
- Simulation for Random variables
- Sampling, Confidence Interval for mean and proportion