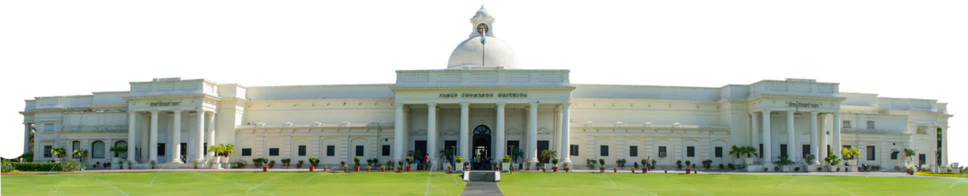
Data Mining for Business Intelligence (IBM 312)

Sumit Kumar Yadav

Department of Management Studies

January 18, 2023



Independent Random Variables

X	Y
x_1	y_1
x_2	y_2
⋮	⋮

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

- If X and Y are independent random variables, then having any information about X doesn't change anything in distribution of Y (and vice versa).
- Check that for independent random variables - E(XY) = E(X)E(Y).
- Is the reverse also true?

X \ Y	1	2	3
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
1	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

$$P(Y = 2) = \frac{3}{9}$$

$$P(X = 1) = \frac{2}{3}$$

$$\boxed{P(X = 1, Y = 2) = \frac{2}{9}}$$

Independent Random Variables

$$\left(\sum_i P(X=x_i)\, x_i\right)\left(\sum_j P(Y=y_j).\, y_j\right)$$

- If X and Y are independent random variables, then having any information about X doesn't change anything in distribution of Y (and vice versa)
- Check that for independent random variables - E(XY) = E(X)E(Y).
- Is the reverse also true?

$$\sum_i \sum_j \underbrace{P(X=x_i).P(Y=y_j)}\, x_i\, y_j$$

$$\sum_i \sum_j P(X=x_i, Y=y_j)\, \underline{x_i\, y_j} = E(XY)$$

Independent Random Variables

$E(X) = 1$

$E(Y) = \frac{1}{2}$

$E(XY) = \frac{1}{2}$

Expt: Tossing a coin twice

X = no. of heads in these 2 tosses

- If X and Y are independent random variables, then having any information about X doesn't change anything in distribution of Y (and vice versa)
- Check that for independent random variables - E(XY) = E(X)E(Y).
- Is the reverse also true?

$$Y = \begin{cases} 1, & \text{if both tosses are same} \\ 0, & \text{else} \end{cases}$$

	X	Y	XY
HH	2	1	2
HT	1	0	0
TH	1	0	0
TT	0	1	0

$P(X=2, Y=0) = ??$

Examples

"N"

- 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

Binomial Random Variable

$E(x) = p, \quad Var(X) = p(1-p)$

- Bernoulli Random Variable - Do an experiment once, probability of success = p. (X = 1, if success, 0 otherwise). Find E(X) and Var(X)
- Binomial Random Variable - Repeat independent Bernoulli trials *n* times. Y = total number of successes in these *n* trails.
- Find E(Y) and Var(Y)

Binomial Random Variable

$Y = X_1 + X_2 + \ldots + X_n$

$E(Y) = np \quad , \quad Var(Y) = np(1-p)$

- Bernoulli Random Variable - Do an experiment once, probability of success = p. (X = 1, if success, 0 otherwise). Find E(X) and Var(X)
- Binomial Random Variable - Repeat independent Bernoulli trials *n* times. Y = total number of successes in these *n* trails.
- Find E(Y) and Var(Y)

$(0, 1, 2, \ldots, n)$

$P(Y=r) = {}^nC_r\, p^r (1-p)^{n-r}$

Check that $\boxed{\sum_{r=0}^{n} r \cdot P(Y=r) = np}$

Binomial Random Variable

- Bernoulli Random Variable - Do an experiment once, probability of success = p. (X = 1, if success, 0 otherwise). Find E(X) and Var(X)
- Binomial Random Variable - Repeat independent Bernoulli trials *n* times. Y = total number of successes in these *n* trails.
- Find E(Y) and Var(Y)

Poisson Random Variable

$$\frac{e^{-\lambda}\lambda^{i}}{i!} = P(X=i)$$

- Let the number of events happening in a given period of time be X
- If X follows the following probability distribution, we say that X follows Poisson distribution
- $P(X=i) = \frac{e^{-\lambda}\lambda^{i}}{i!}$; i = 0,1,2,....
- Find E(X) and Var(X)

$$E(X) = e^{-\lambda}\sum_{i=0}^{\infty} \frac{i\,\lambda^{i}}{i!} = \lambda e^{-\lambda} \boxed{\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}}$$

$$= \lambda$$

Poisson Random Variable

$$Var(x) = E(x^2) - (E(x))^2$$
$$(\lambda^2 + \lambda) - \lambda^2 = \lambda$$

- Let the number of events happening in a given period of time be X
- If X follows the following probability distribution, we say that X follows Poisson distribution
- $P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$; i = 0,1,2,....
- Find E(X) and Var(X)

$$\boxed{\begin{aligned} E(x) &= \lambda \\ Var(x) &= \lambda \end{aligned}}$$

$$\sum_{i=0}^{\infty} i^2 \left(\frac{e^{-\lambda} \lambda^i}{i!}\right)$$

Random Sum of Random Numbers

$$C = \begin{cases} 0, & 0.1 \\ 1, & 0.8 \\ 2, & 0.1 \end{cases}$$

$E(C) = 1$

$Var(C) = 0.2$

- In a tea shop, the number of customers coming in a given day follows a Poisson distribution with parameter 500
- Each customer makes the purchase as per the following distibution - a.) No purchase with probability = 0.1, one cup of tea with probability = 0.8, two cups of tea with probability = 0.1
- Let X denote the number of tea cups sold in a day. What is E(X) and Var(X)

$E(X) = 500$

$Var(X) = 100$

$N \sim Poisson(500)$

$$\boxed{X = C_1 + C_2 + \dots + C_N}$$

Simulation of Random numbers in Python