Lec-III. Binomial Distribution

★ **Binomial Distribution :** A probability distribution that describes the nos of sucesses in a fixed number of independent bernoullis trial with two possible outcomes (often called "success" and "failure"), where the probability of sucess is constant for each trial.

- The binomial distribution ~~of~~ is characterized by two parameters :
  the nos of trial 'n' and probability of success 'p'

$$\boxed{P[X = \text{'r' success}] = {}^{n}C_{r}\, p^{r}\, q^{n-r}}$$

n : nos of trial, p : probability of success
q : prob. of failure, r : nos of success

✓ S/F

$$\text{Binomial} \xrightarrow[n,\ r,\ p]{} \boxed{B(n, p, r)} \longrightarrow P(X=r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

- Statistical averages

  - $E[X] = p$
  - $V[X] = pq$
  - $S.D = \sqrt{pq}$

$\xrightarrow[\text{trials}]{\text{for 'n'}}$

  - np
  - npq
  - $\sqrt{npq}$

eg. tossing a three coin

what is probability ?

$P(X = 1H)$, $P(X \geq 1)$, $P(X \leq 2H)$ = ?

exactly one head , atleast one Head , Atmost two head

single trial $\to P(H) = 1/2$

$\to P(T) = 1/2$

* $P[X = 1H] \to$ $r \to$ nos of success $\to 1$

$^nC_r\, p^r q^{n-r} \to {}^3C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{1-1} \Rightarrow \frac{3}{8}$

* $P[X \geq 1H]$ atleast 1 $\Rightarrow P(X=1) + P(X=2) + P(X=3)$

2
3
⋮
n

$P(X=2) \to {}^3C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{3-2} \Rightarrow \frac{3}{8}$

$P(X=3) \to {}^3C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{3-3} \Rightarrow \frac{1}{8}$

* 3 time tossing coin

0H
1H
2H
3H

"or else"

$P(X=0H) + P(X=1H) + P(X=2H) + P(X=3H) = 1$

$\to$ All probo $\Sigma = 1$

$P(X \geq 1) \Rightarrow 1 - P(X=0H)$

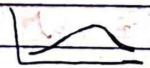
tossing 10 coin → perform this 1000 times.

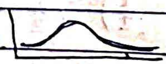
```
import numpy as np
import matplotlib.pyplot as plt

n=10  # nos of trials (tossing 10 coin)
p=0.5  # prob. of success
size = 1000  # nos of sample to generate

binomial_dist = np.random.binomial(n,p,size)

plt.hist(binomial_dist , density = True)

plt.show()
```

if → p = 0.8 # prob. of head high

similarly, if → p = 0.1
[histogram shift more foward left]

<u>outcome</u>: ① p = 0.5 → just like normal dist.

② p tend foward 1 →  shift rightward

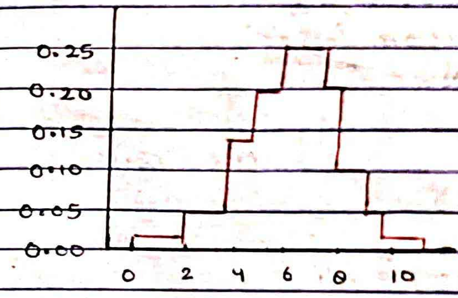
③ p backwards 1 →  shift leftward

1st expt → 10 coin → 5 head tosses

array([5, 5, 4, 4, 6, 5, - - - - -
- - - - - - - - - - -
5, 5, 5, 5, 6, 7])

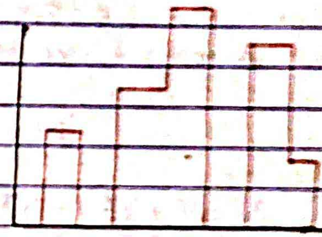
3rd expt → 10 coin → 4 head tossed

p = 0.5 → prob. Head आने का 50% है
That's why array contain 5 much capacity



→ ⓐ

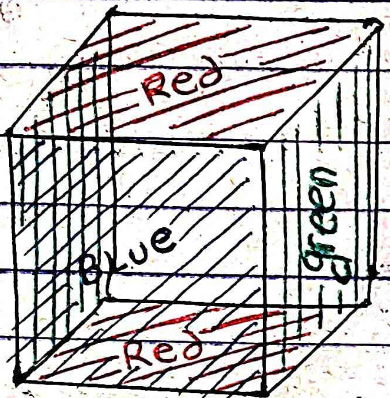
array([9, 5, 8, - - -
- - - - - - - -
8, 8, 8, 7, 6])



more biased toward 8 times head
[histogram shifts more foward right]

Applications
- binary classification problem
- hypothesis testing
- logistic regression
- A/B testing

Q). Consider a unbiased cubic die with opposite faces colored identically each face colored red, blue or green such that each color appear only two times on that die. If the die is thrown thrice, the prob. of obtaining red color on the top face of die at twice is —



front-back : green
top-bottom : red
up-down : blue

$n = 3$ (trials)   success = red

$P(\text{red}) = P(\text{success}) = 2/6$

$P(\text{not red}) = P(\text{failure}) = 4/6$

atleast twice $\rightarrow$ $P(X \geq 2)$

$\downarrow$

$P(X=2) + P(X=3)$

$${}^3C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{3-2} + {}^3C_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^{3-3}$$

$$\Rightarrow \frac{7}{27}$$

Q) The probability of obtaining atleast two 'SIX' in throwing a fair die 4 times —

$P(X \geq 2) \Rightarrow P(X=2) + P(X=3) + P(X=4)$

$n \Rightarrow 4$ trial $\quad p(\text{success}) = 1/6 \quad q = 5/6$

$${}^4C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{4-2} + {}^4C_3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^{4-3} + {}^4C_4 \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^{4-4}$$

Q) In an ~~experiment~~ examination of 10 multiple choice questions (one or more can be correct) out of 4 options. A student decided to mark the answer at random.
Find the prob. he gets exactly two questions corrected.

$\text{question} \to$ correct / not correct

(bernoulli's trial)

$n = 10$ ques $\to$ large
(binomial dist.)

$r = 2$ ques

$P(X=2) = ?$

4 option : $\boxed{A}$ $\boxed{B}$ $\boxed{C}$ $\boxed{D}$

case (1) $\rightarrow$ ${}^4C_1$ 4 में से 1 option सही है।

case (2) $\rightarrow$ ${}^4C_2$ 4 में से 2 option सही हैं

case (3) $\rightarrow$ ${}^4C_3$

case (4) $\rightarrow$ ${}^4C_4$

(+)

15 choices

p(one question correct)
$\hookrightarrow$ $1/15$

$q \rightarrow 14/15$

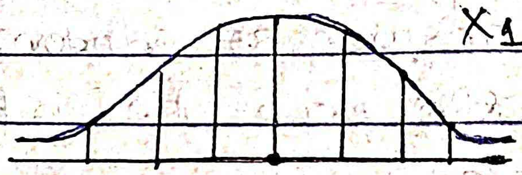
Q). X_1 and X_2 be two independent normal random variables with μ_1 and μ_2 and std. dev σ_1, σ_2 respectively. Consider, $Y = X_1 - X_2$; $\mu_1 = \mu_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 2$?

(a) Y is normally distributed with mean and variance

(b) Y has mean _ & var _ but not normally distributed

$\mu_1 = 1, \mu_2 = 1$

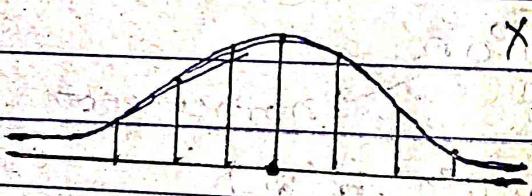
$\sigma_1 = 1, \sigma_2 = 2$



X_1

$\mu_Y = \mu_{X_1} - \mu_{X_2}$

$\Rightarrow 1 - 1 \Rightarrow 0$



X_2

$\sigma_Y^2 = \sigma_{X_1}^2 + (-\sigma_{X_2}^2)$

$\Rightarrow 1 + 4 \Rightarrow 5$

$Y = X_1 - X_2$

var = 5, mean = 0

Q). A die is rolled 180 times using gaussian random variable. find the prob. that face 4 will turn up atleast 35 times ?

p(success) = Face 4 ⇒ 1/6

p(failure) = Not Face 4 ⇒ 5/6.

n = nos of success = 35.

$p(x \geq 35) \longrightarrow P(x=35) + P(X=36) + \cdots + P(X=180)$

or

$1 - P(X=0) + P(X=1) + \cdots + P(X=35)$

$P(X \geq 35)$

↓ convert to Z

$$P\left(\frac{X-\mu}{\sigma} \geq \frac{35-\mu}{\sigma}\right)$$

binomial distribution,

$p = 1/6, \; q = 5/6, \; n = 180$

$E[X] = np \Rightarrow \mu = 180 \times \frac{1}{6} \rightarrow 30$

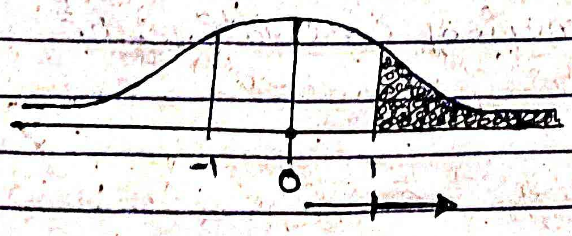
$Var.X = npq \Rightarrow 180 \times \frac{1}{6} \times \frac{5}{6} \rightarrow 25$

$Std \rightarrow \sqrt{25} \Rightarrow 5$

$$P\left(Z \geq \frac{35-30}{5}\right) \Rightarrow P(Z \geq 1)$$

$P(Z \geq 1) \Rightarrow 0.5 - 0.3416$

$\Rightarrow 0.1587$



A. Bernoulli's Trial :

Event
- Success
- Failure

Yes / No Success / failure Win / Loose

$\underbrace{\text{Yes / No} \quad \text{Success / failure} \quad \text{Win / Loose}}_{\text{independent events}}$

Sample	Success	Failure
x	1	0
$P(X=x)$	p	$1-p$

- prob. of success $= p$
- prob. of failure $= q$
- $P(S) + P(F)$ or $p + q$ $= 1$

$\underbrace{\text{mean} = ?,\ \text{var} = ?,\ \text{dev} = ?}_{\text{Statistical Average}}$

mean

$E[x] = 1 \times p + 0 \times (1-p)$

$E[x] = p$

variance

$E[X^2] - [E[X]]^2$

$[(1)^2 p - (0)^2 (1-p)] - p^2$

$\Rightarrow p - p^2 \Rightarrow p(1-p)$

$Var(x) = pq$

Standard deviation $= \sqrt{pq}$

for 'n' trials,

$n = 1$ S F

$n = 2$ S F S F

$n = 3$ = S SF S F S F S F

$\boxed{SSS}$ $\boxed{SSF}$ $\boxed{SFS}$ $\boxed{SFF}$ .... $\boxed{FFS}$ $\boxed{FFF}$

1

1 2 1

1 3 3 1

1 4 6 4

PASAL TRIANGLE

$P(S) = p$
$P(F) = q$
nos of trial $= n$
nos of succes $= ? = r$

$P(X = \text{Success})$

bernoullis trial

↓ n large

binomial distribution

prob. of success → 'r' times → $p \cdot p \cdot p \cdot \ldots$ 'r' times
$\Rightarrow p^r$

$n \to$ trial
$r \to$ success
failure $\to n-r$

prob of failure → $(n-r)$ times
$q \cdot q \cdot q \cdot \ldots (n-r)$ times
$\Rightarrow q^{n-r}$