Bernoulli Distribution

$$P(X=x)=\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$$

The Bernoulli distribution is the **simplest discrete probability distribution**, representing a **single experiment** (or trial) that has only **two possible outcomes**:

Success (usually coded as 1)

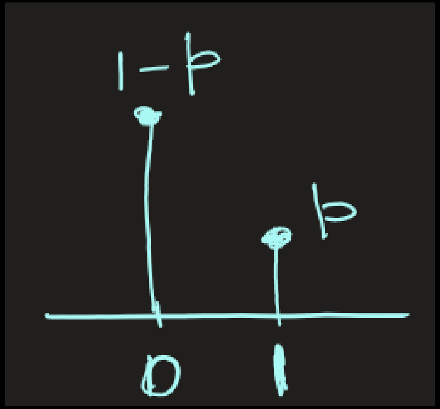
Failure (usually coded as 0)

The probability of success is denoted by p, and the probability of failure is $1 - p$.

The **Probability mass function (PMF):**

$$P(X=x)=\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$$

Here X = {0, 1}.



It is used in

- Coin toss: Head (1) / Tail (0)
- Passing a test: Pass (1) / Fail (0)
- Email: Spam (1) / Not Spam (0)
- Manufacturing: Defective (1) / Not Defective (0)

Bernoulli Distribution

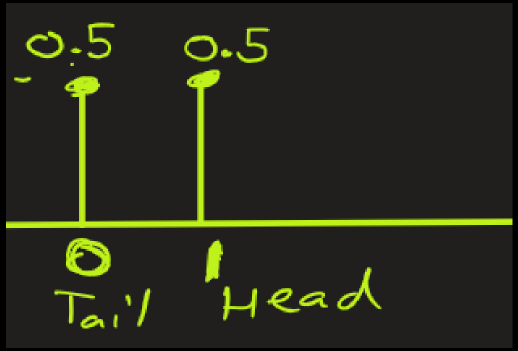
$$P(X=x)=\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$$

Example: Suppose you flip a fair coin once.
What is the probability mass function ?

Ans:
There are only 2 outcomes: Head and Tail.
Let us denote success as **h**ead (1) and failure as tail (0). Here X = {0, 1}.

Then for fair coin the PMF is

P (X=1)= 0.5 , probability of head
P (X=0)= 0.5 , probability of tail



Bernoulli Distribution

$$P(X=x)=\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$$

Example: A dataset containing emails received and whether it is spam is provided. Let X be a random variable representing spam email received on any given day. Assuming we can calculate the probability that an email is spam or not from this dataset, what is the probability mass function (PMF) of X ?

email	Is_spam
E1	Y
E2	Y
E3	Y
E4	N
E5	N
E6	Y
E7	N

Ans:

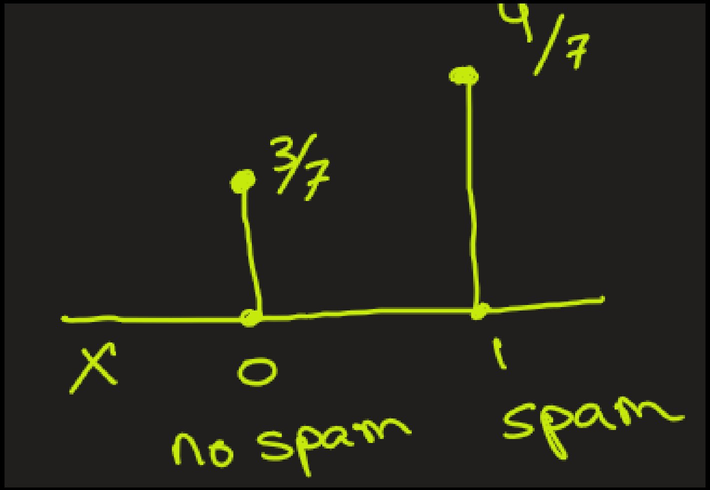
Here X is a random variable indicating whether new email is spam(1) or not spam(0)

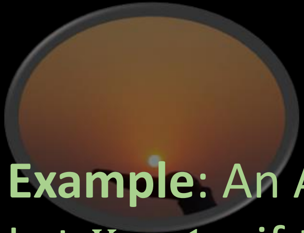
X = {0, 1} .

By looking at data, we see that PMF is

P(X=1) = 4/7 , (spam)

P(X = 0) = 3/7 , (not spam)





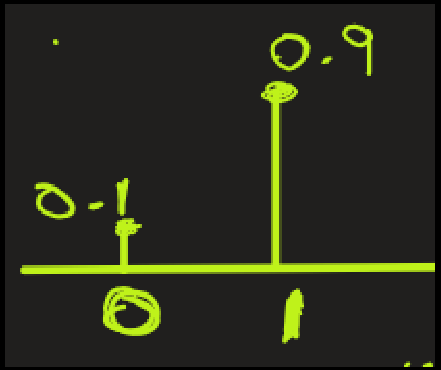
Bernoulli Distribution

$$P(X=x)=\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$$

Example: An AI system is asked to classify a single image as either **cat** or **not cat**.
Let $X = 1$ if the AI correctly classifies the image.
Let $X = 0$ if it misclassifies.
Suppose the probability that the AI correctly classifies the image is $p = 0.9$.
1) What is the PMF ?
2) If we repeat the classification process once more (independently), what is the probability that **both classifications are correct**?

Ans:
1) Here P(X=1) = 0.90 is the probability of correct classification.
P(X = 0) = 1 – 0.90 = 0.10 = 10 % , probability of misclassifying

So, PMF is
P(X = 1) = 0.90 (correct classification)
P(X = 0) = 0.10 (incorrect classification)



2) P(both classification correct) = P (X=1 and X=1)
= P(X=1) x P(X=2) (because classification processes are independent)
= 0.9 x 0.9
= 0.81 = 81%

STOP

Bernoulli Distribution

Key statistics:

Mean (Expected value):

$$\mu = E[X] = p$$

Variance:

$$\sigma^2 = Var(X) = p(1-p)$$

Standard deviation:

$$\sigma = \sqrt{p(1-p)}$$