Probability

PAL WORKSHOP

OUR GOAL FOR TODAY

1. Laws of probabilities
2. Random Variables
3. Discrete and Continuous spaces
4. The Law of Total Probability
5. Measures of Central Tendency and Spread
6. Distributions

01

WHAT IS PROBABILITY?

- **Concerns the study on uncertainty (loosely speaking).**
- For certain types of events, we cannot predict the outcome with certainty in advance, e.g. tossing a coin or tossing a die.
- However we know the set of all possible outcomes for these events.
- We would like to use probability to measure the chance of something occurring in an experiment

PROBABILITY

H : What is it?

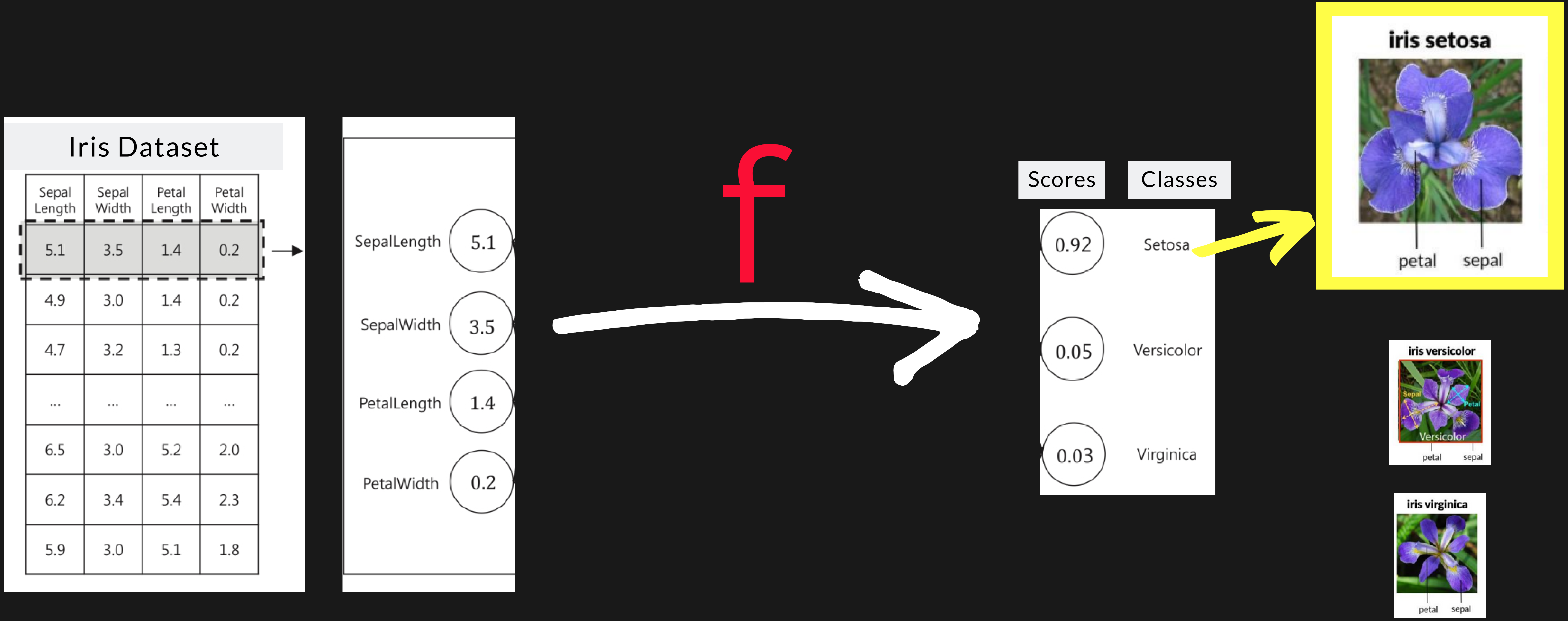
C : DOG

100% DOG

Not true!!

Then what is it?

PROBABILITY

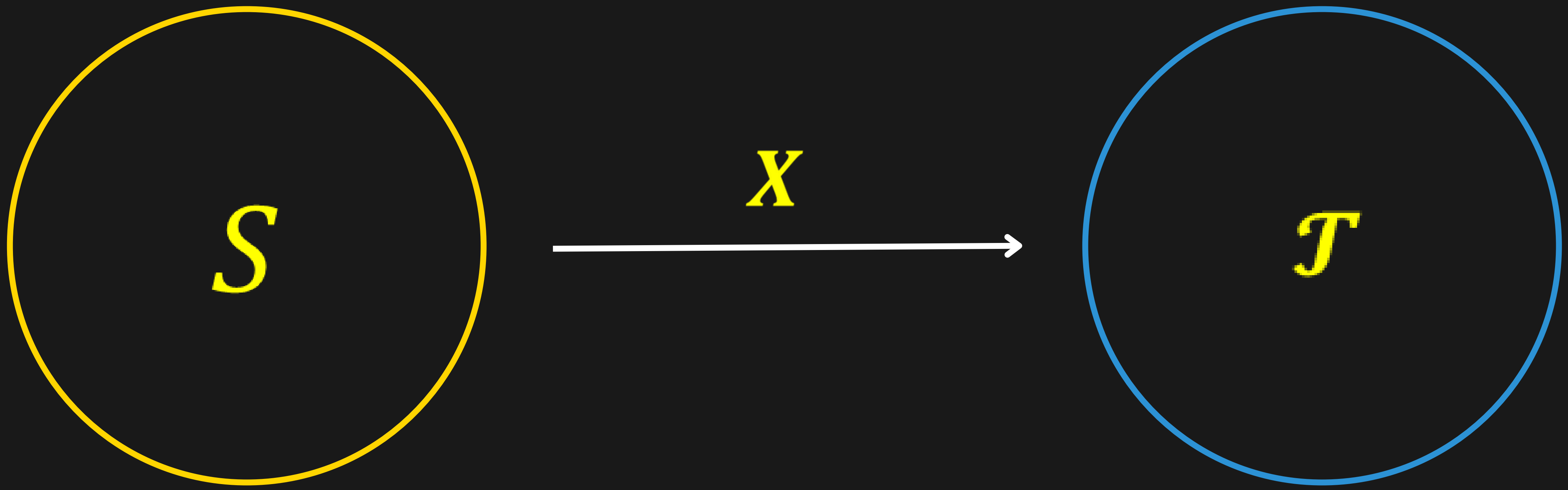


02

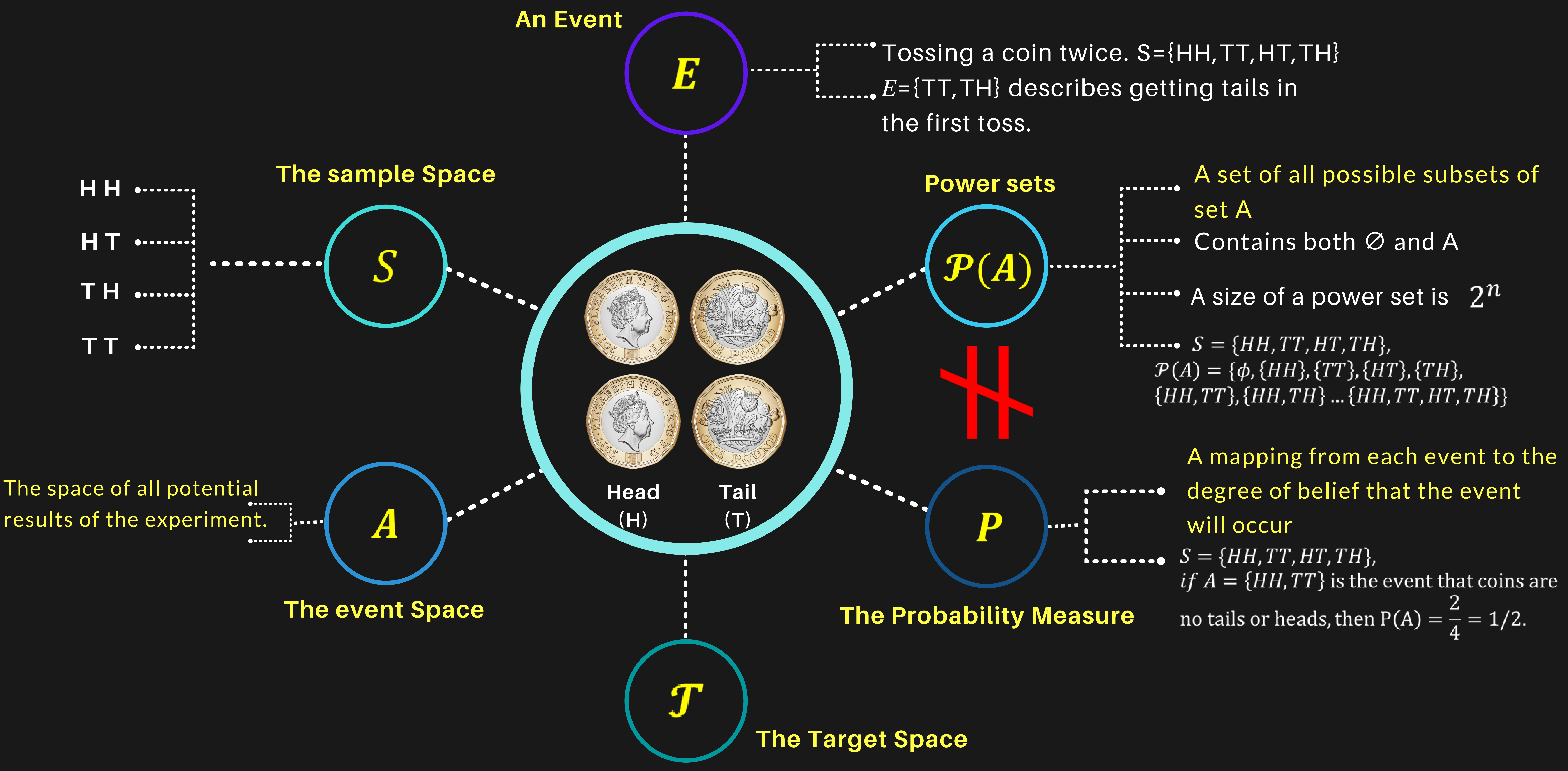
RANDOM VARIABLES

- The term itself is misleading as it is neither random nor is it a variable
- In fact it as a function

$$X: S \to \mathcal{T}$$

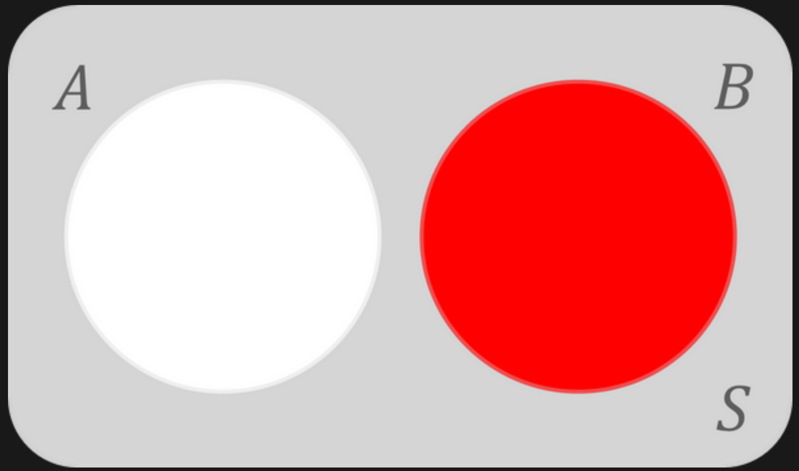


Let's consider the sample space of two successive coin tosses:



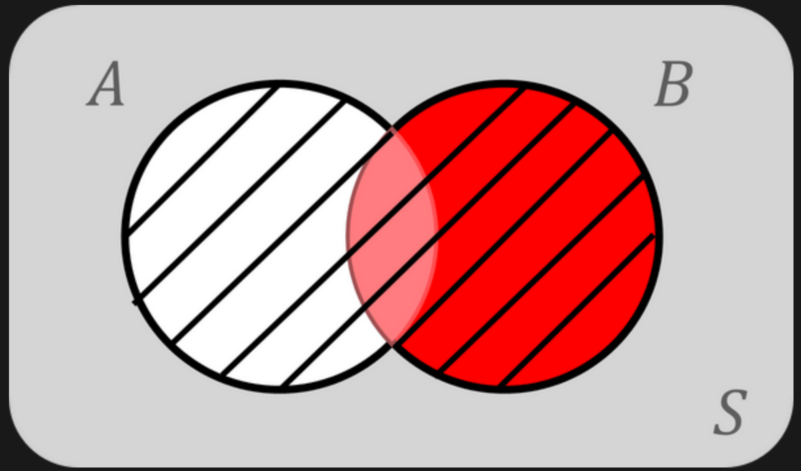
PROBABILITIES AS SETS

Mutually Exclusive events



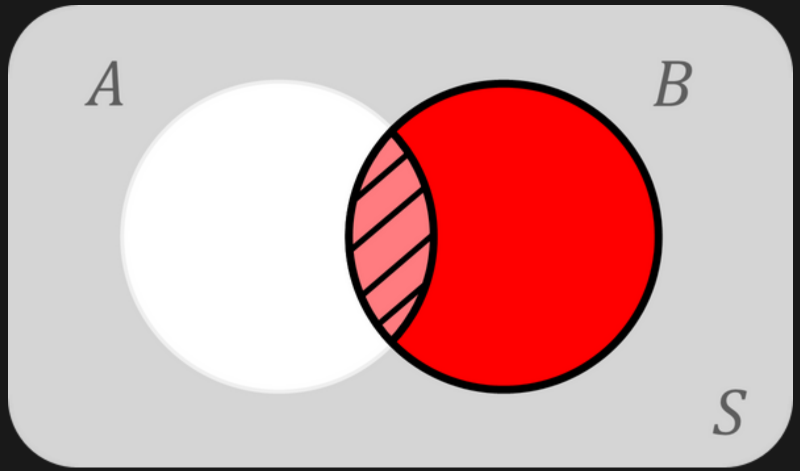
Probability of a union of two events

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

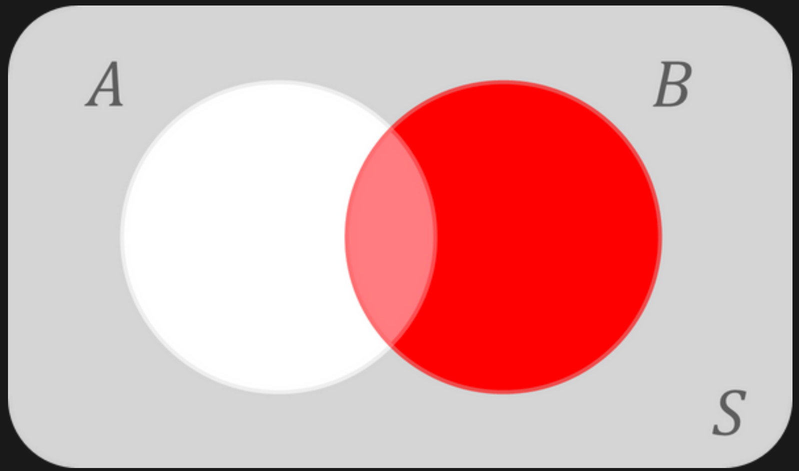


Probability of a conjunction of two events

$$P(A \cap B) = P(A) \times P(B|A)$$

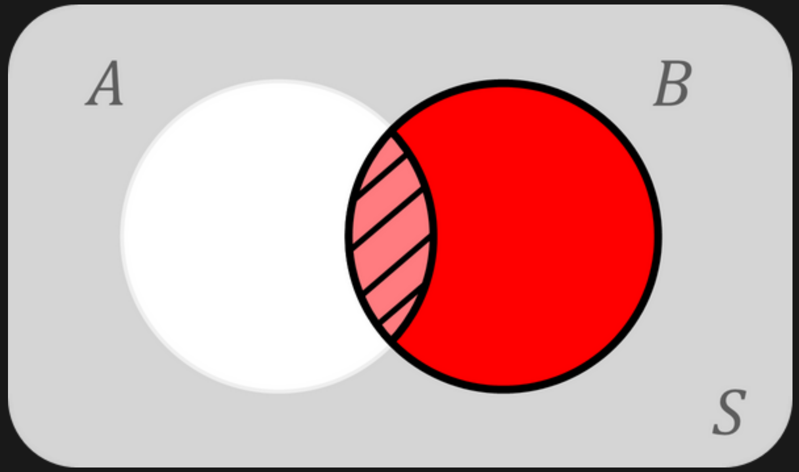


Non-mutually Exclusive events



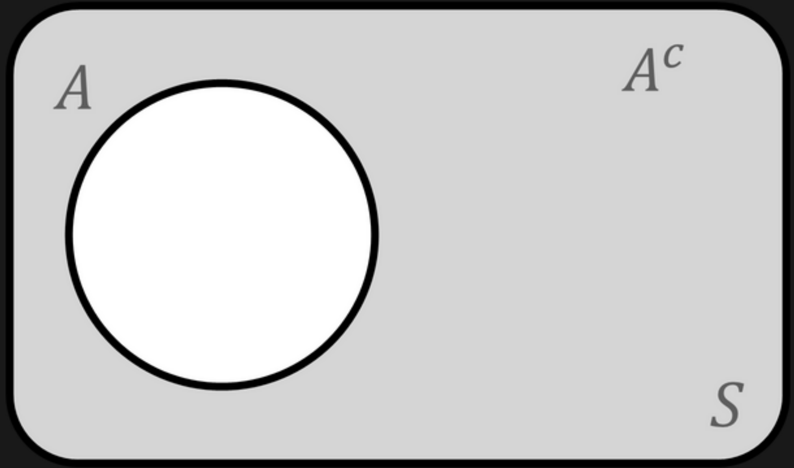
Conditional probability of one event given another

$$P(A|B) = P(A \cap B)/P(B)$$



Probability of Non-event

Complement $P(A^c) = P(S) - P(A)$



HEAD(H) Random Variable X = 1

X(H) = 1

TAIL(T) Random Variable X = 0

X(T) = 0

$\mathcal{T}=\{0,1,2\}$

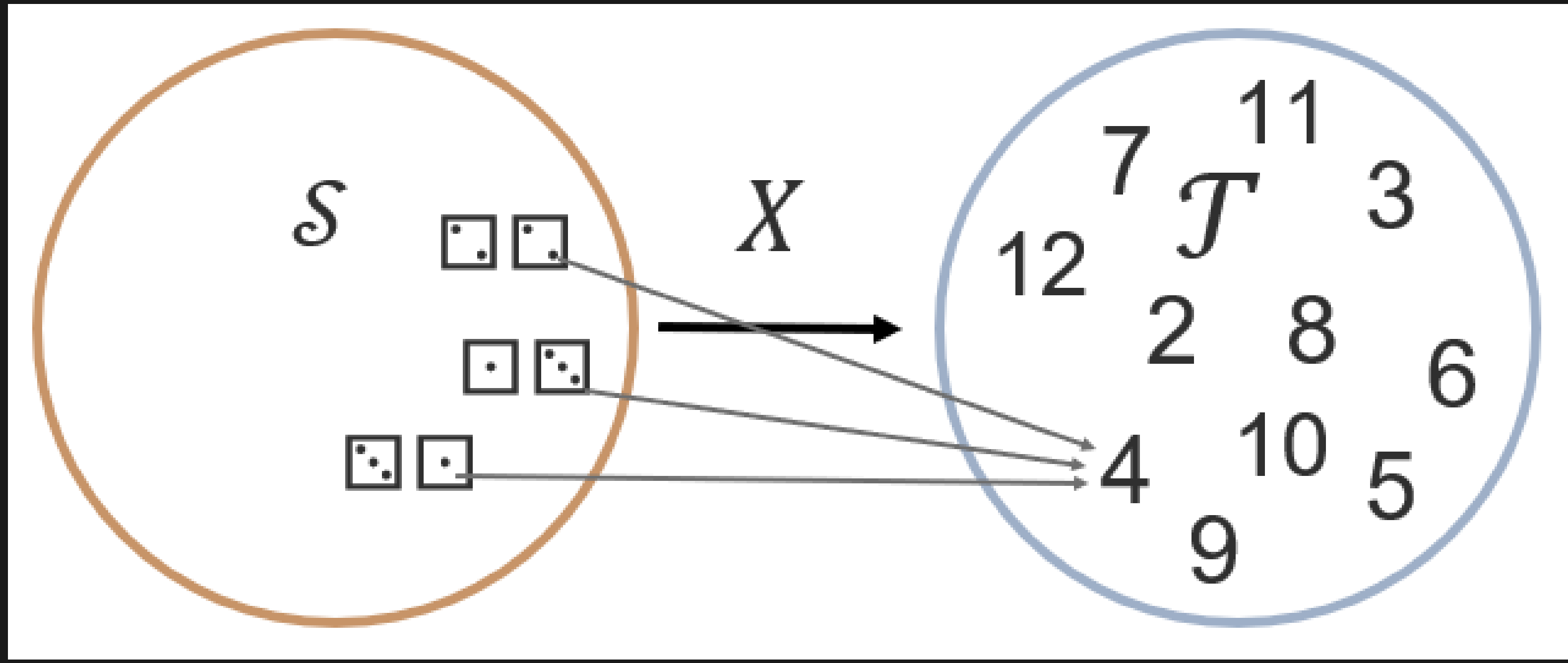
$$\mathrm{X}((H,H)) = 2$$
$$\mathrm{X}((H,T)) = 1$$
$$\mathrm{X}((T,H)) = 1$$
$$\mathrm{X}((T,T)) = 0$$

The Target Space $\mathcal{T}$

- Consider the set of all possible outcomes of throwing two six-sided dice.
- Then Ω can be represented as:

{(⚀, ⚀); (⚀, ⚁); (⚀, ⚂); (⚀, ⚃); (⚀, ⚄); (⚀, ⚅); (⚁, ⚀); (⚁, ⚁); (⚁, ⚂); (⚁, ⚃); (⚁, ⚄); (⚁, ⚅); (⚂, ⚀); (⚂, ⚁); (⚂, ⚂); (⚂, ⚃); (⚂, ⚄); (⚂, ⚅); (⚃, ⚀); (⚃, ⚁); (⚃, ⚂); (⚃, ⚃); (⚃, ⚄); (⚃, ⚅), (⚄, ⚀); (⚄, ⚁); (⚄, ⚂); (⚄, ⚃); (⚄, ⚄); (⚄, ⚅); (⚅, ⚀); (⚅, ⚁); (⚅, ⚂); (⚅, ⚃); (⚅, ⚄); (⚅, ⚅)}

- One possible random variable we can define is the *sum of the tosses.*
- In that case $\mathcal{T}$ will be a set of integers from 2 to 12.



PROBABILITY

(1) $0 \leq \mathrm{P(A)} \leq 1$

(2) $\mathrm{P(S)} = 1$

(3) $If\ A_1, A_2, \ldots are\ \boldsymbol{mutually\ exclusive\ events}, \mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

In Words	Notation_1	Notation_2
All Heads	n(A) =	P(A) =
All Tails	n(B) =	P(B) =
All coins	n(S) =	P(S) =
Intersection: Both head and head	$n(A \cap B) =$	$P(A \cap B) =$
Only Heads: Head coins that are not tail coins	$n(A) - n(A \cap B) =$	$P(A) - P(A \cap B) =$
Only Tails: Tail coins that are not head coins	$n(B) - n(A \cap B) =$	$P(B) - P(A \cap B) =$
Union: Head or Tail	$n(A \cup B) =$	$P(A \cup B) =$
Everything else	$n(A \cup B)' =$	$P(A \cup B)' =$

LET'S PRACTICE

$Q1.$ *Probability of getting an even number on rolling a dice once. what are smple space(S), Event(E) and probability?*

$Q2. If\ A\ \&\ B\ are\ two\ mutually\ exclusive\ events\ then\ P(A \cap B) = 0\ and\ P(A \cup B) = P(A) + P(B).$

$$A = \{Numbers\ greater\ than\ or\ equal\ to\ 4\ in\ a\ dice\ roll\} = \{4,5,6\}$$

$$B = \{Numbers\ lesser\ than\ or\ equal\ to\ 4\ in\ a\ dice\ roll\} = \{1,2,3,4\}$$

$Then, what\ is\ P(A \cup B)?$

DISCRETE & CONTINUOUS PROBABILITIES

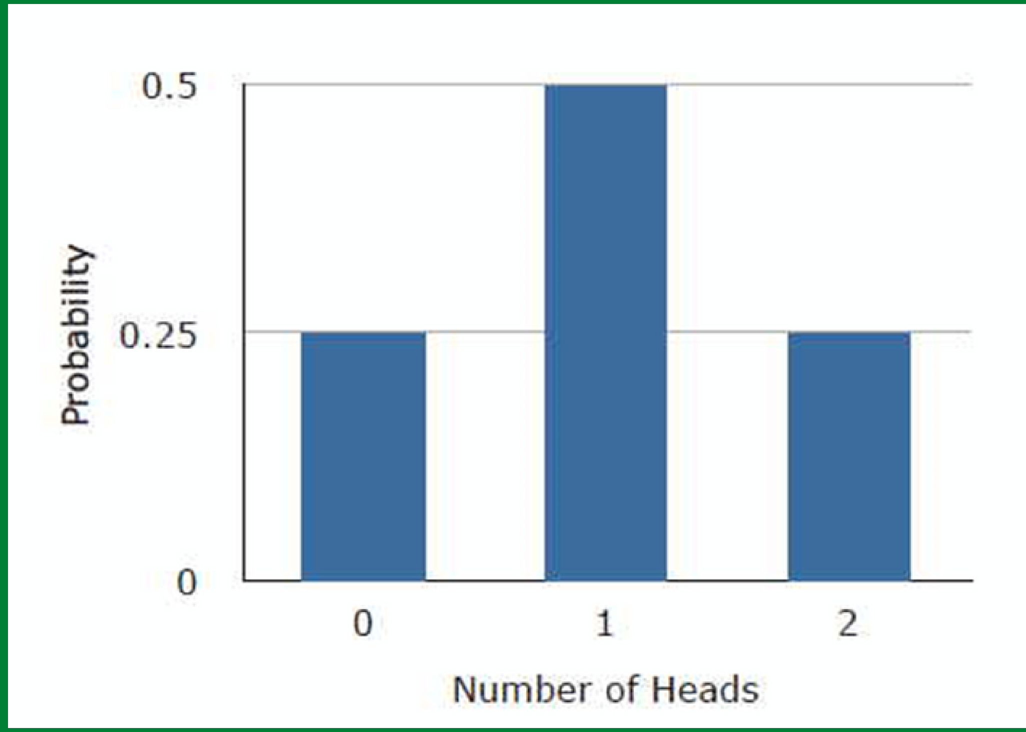
- **It is important to understand the difference between target space types.**
- Discrete random variables:
    - Variable can take on a ***discrete*** set of values.
    - Value can be obtained by counting.
- Continuous random variables:
    - Variable can take on a ***continuous*** set of values.
    - Value can be obtained by measuring.

DISCRETE PROBABILITY

- The probability that a random variable $\boldsymbol{X}$ takes a particular value is $\boldsymbol{x} \in \boldsymbol{T}$ denoted as

$$\boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x})$$

- This expression is also called ***probability mass function.***



$$P(X = 0) = 0.25$$
$$P(X = 1) = 0.5$$
$$P(X = 2) = 0.25$$

DISCRETE PROBABILITY

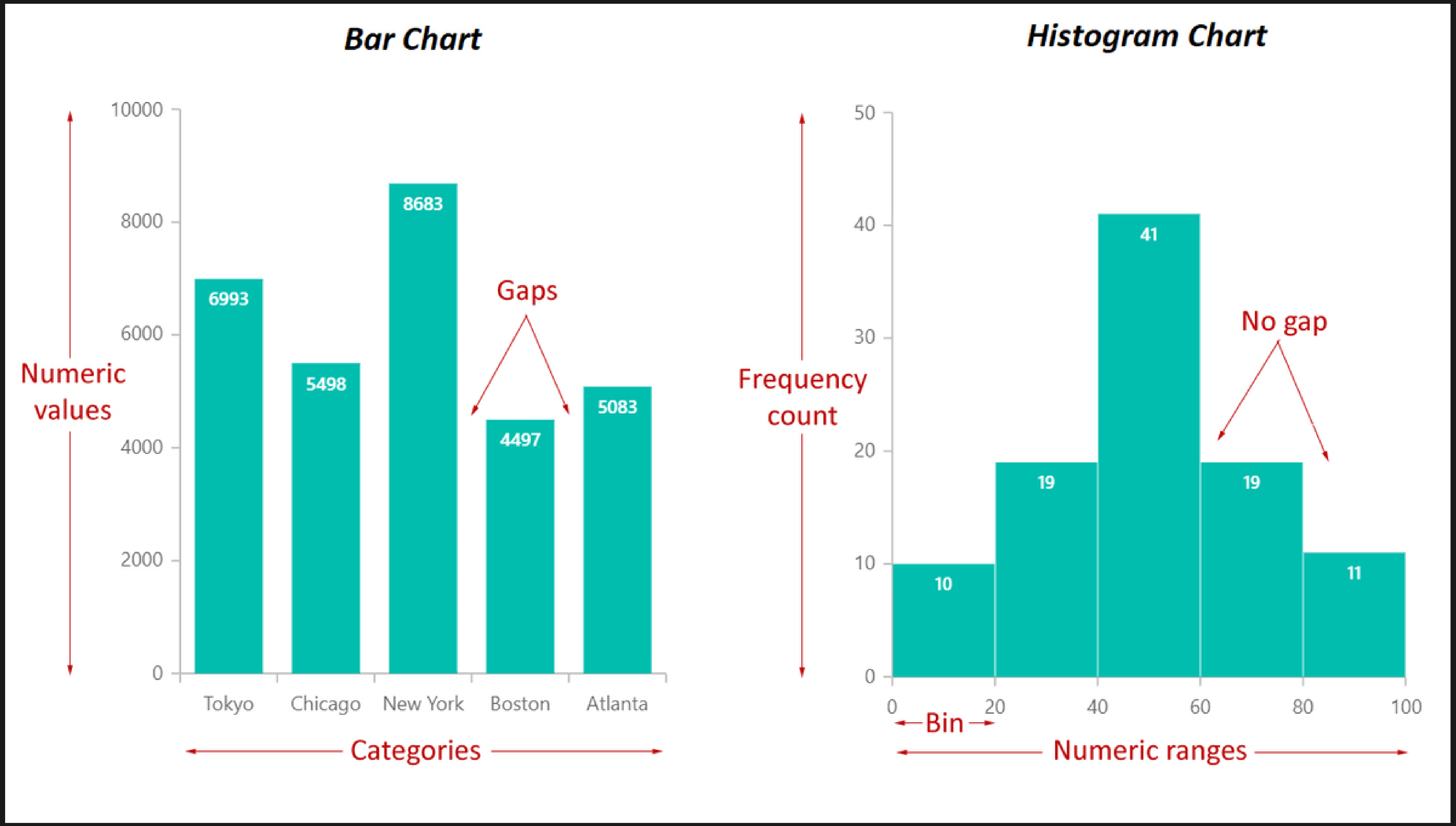
- When the target space is discrete we can imagine the probability distribution of multiple random variables as a multidimensional array of numbers

	x_1	x_2	x_3	x_4	x_5
y_3	0.1	0.07	0.06	0.03	0.1
y_2	0.12	0.09	0.02	0.01	0.05
y_1	0.18	0.01	0.11	0.02	0.03

- *Joint probability* is defined as $p(x, y) = P(X = x_i, Y = y_j)$
- *Marginal probability* $p(x)$ represents the probability that X takes the value x_i irrespective to the value of Y.
- *Conditional probability* $p(y|x)$ will only consider the value of Y for a particular value of X.

Wait!

What's difference between bar graph and histogram?



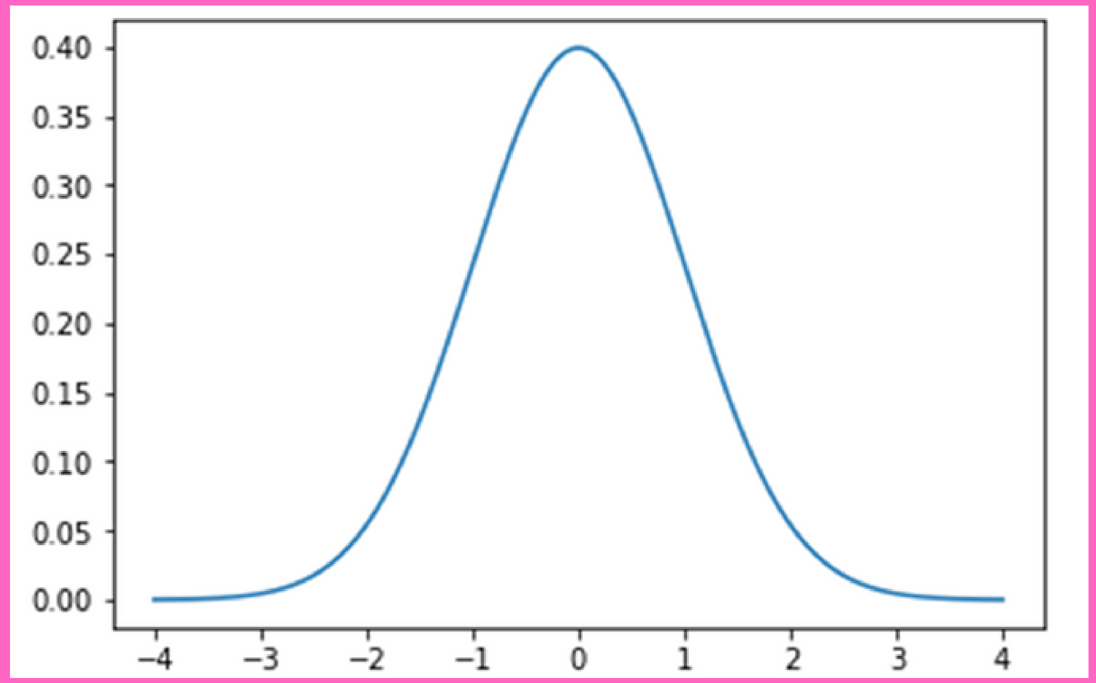
Comparison terms	Bar chart	Histogram chart
Usage	To compare different categories of data.	To display the frequency of occurrences.
Indicates	Discrete values.	Non-discrete values.
Data	Categorical data.	Quantitative data.
Rendering	Each data point is rendered as a separate bar.	The data points are grouped and rendered based on the bin value.
Space between bars	Can have space.	No space.
Reordering bars	Can be reordered.	Cannot reordered.
Axis label placement	Axis labels can be placed on or between the ticks.	Axis labels are placed on the ticks.
Required values	x and y.	Only y.

Reference: https://www.syncfusion.com/blogs/post/difference-between-bar-graph-and-histogram-chart.aspx

CONTINUOUS PROBABILITY

- Target spaces are intervals of the real line $\mathbb{R}$

- A ***probability density function*** is a function whose value at any given point in the sample space can be interpreted as providing a *relative likelihood* that the value of the random variable would be close to that sample.



04

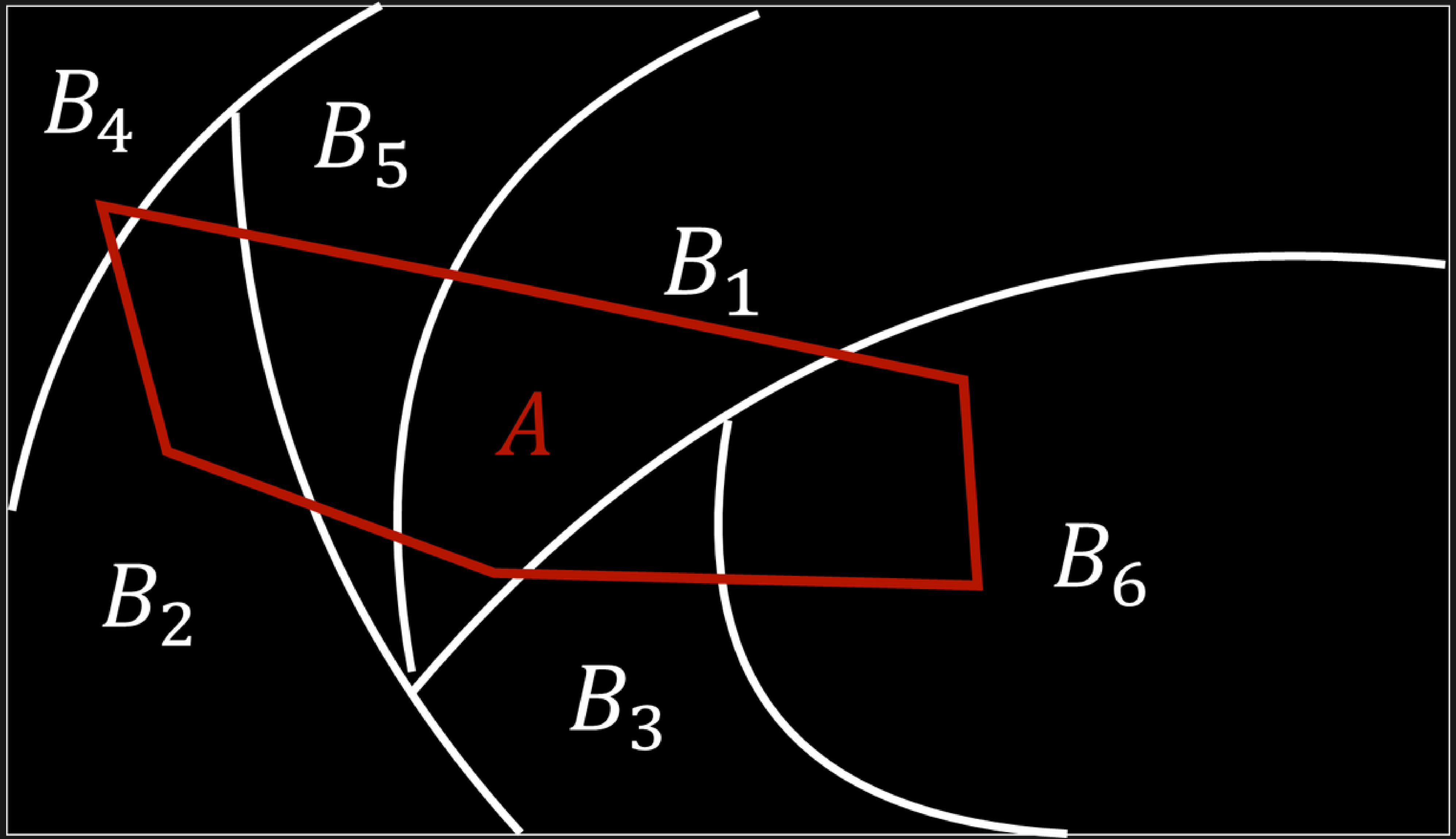
THE LAW OF TOTAL PROBABILITY

- Suppose $B_1, \ldots, B_n$ are mutually exclusive and collectively exhaustive events in a sample space. We can then sum/integrate over the set of states of variable B to get a marginal distribution of variable A.

$$P(A) = \sum_i^n P(A|B_i)P(B_i) = \sum_i^n P(A \cap B_i)$$

THE LAW OF TOTAL PROBABILITY

- Mutually exclusive - no overlap.
- Collectively exhaustive - cover the whole space.

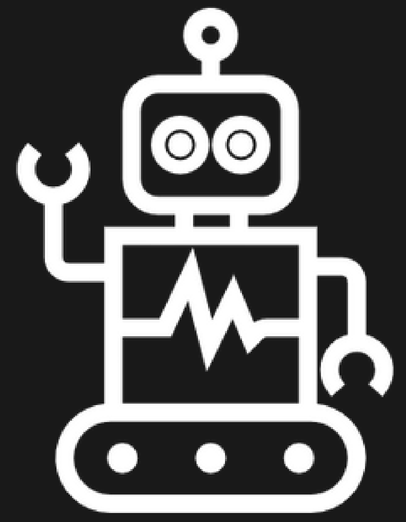


LET'S PRACTICE

Three robots are making parts at the *Sussex factory*. We know that:

A rate of Making parts

A rate of Deficiency of the parts, which the robot makes



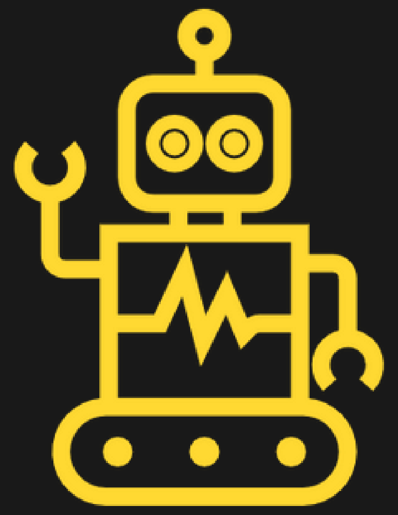
Robot 1

60%

$P(R_1) = 0.6$

7%

$P(D|R_1) = 0.07$



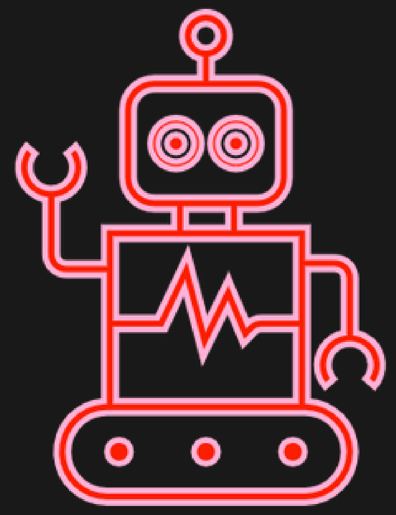
Robot 2

30%

$P(R_2) = 0.3$

15%

$P(D|R_2) = 0.15$



Robot 3

10%

$P(R_3) = 0.1$

30%

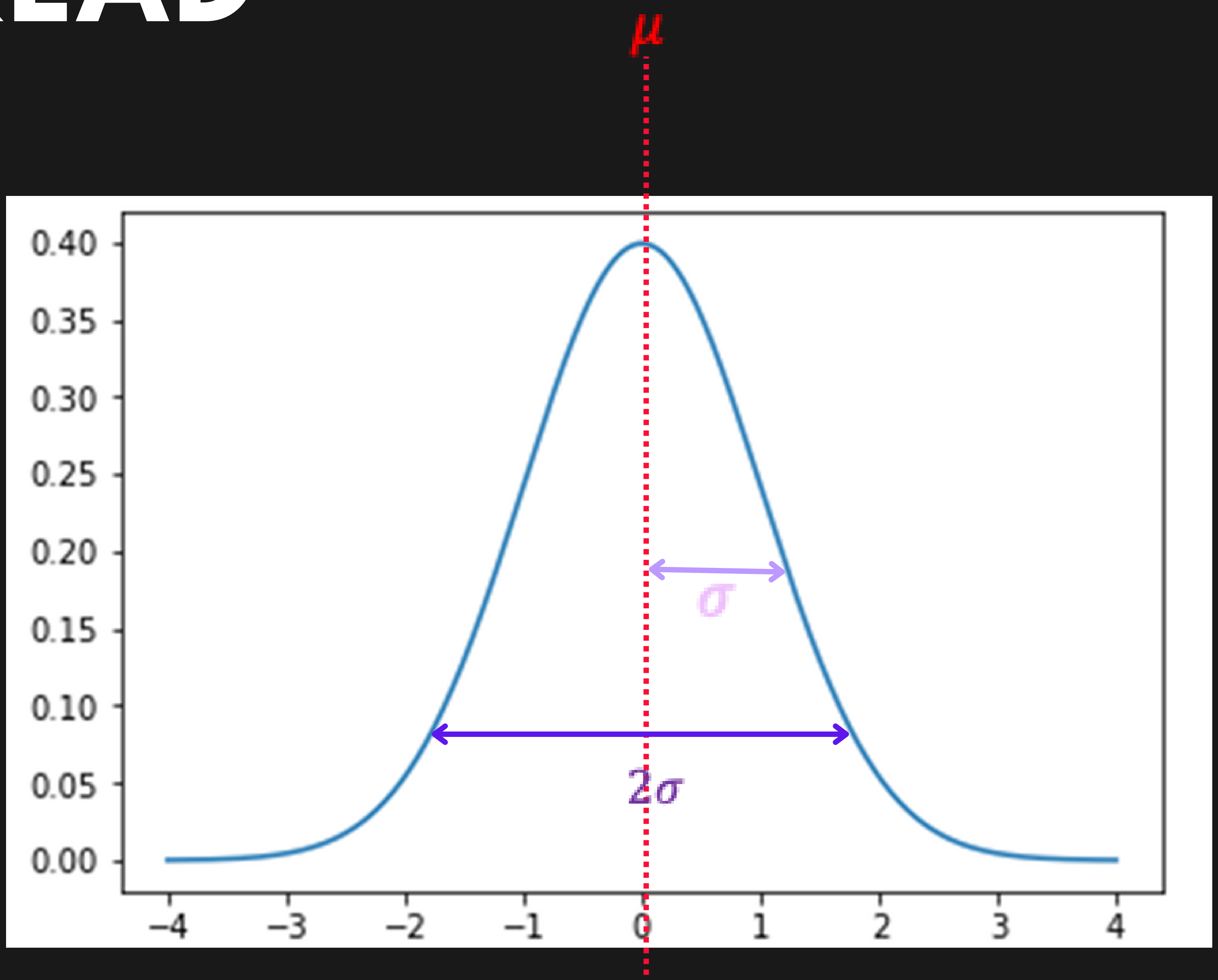
$P(D|R_3) = 0.3$

What is the probability of a randomly selected part being defective?

$$P(D) = ?$$

05

MEASURES OF CENTRAL TENDENCY AND SPREAD



EXPECTATION AND VARIANCE

- *Expected Value/Mean* gives the weighted average of all possible outcomes of the random variable. *Is not an expected outcome, but a theoretical mean!*

$$\mathbb{E}(X) = \sum xp(x)$$

- *Variance* represents the dispersion, i.e. how far a set of numbers is spread from the mean.

$$Var(X) = \mathbb{E}[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x)$$

- *Standard Deviation* is simply a square root of the variance

$$\sigma_X = \sqrt{Var(X)}$$

Covariance

Definition

Covariance of two univariate random variables X,Y∈ℝ is given by the expected product of their deviations from their respected means.

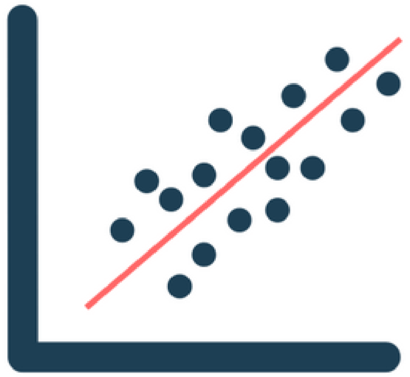
$$Cov(X,Y) = \mathbb{E}_{X,Y}[(x - \mathbb{E}_X[x])][(y - \mathbb{E}_Y[y])]$$

Correlation

Growth

- ***Correlation*** is the normalized form of Covariance.
- Is useful when we want to compare the covariances between different pairs of random variables.

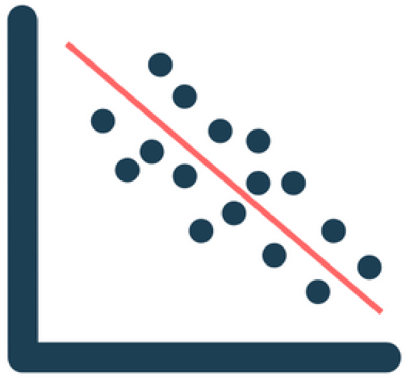
$$Corr[X,Y] = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \in [-1,1]$$



Positive correlation

As one variable increases so does the other variable.

$$\operatorname{cov}(X, Y) > 0$$



Negative correlation

As one variable increases the other variable decreases.

$$\operatorname{cov}(X, Y) < 0$$

No correlation

There is no relationship between the two variables.

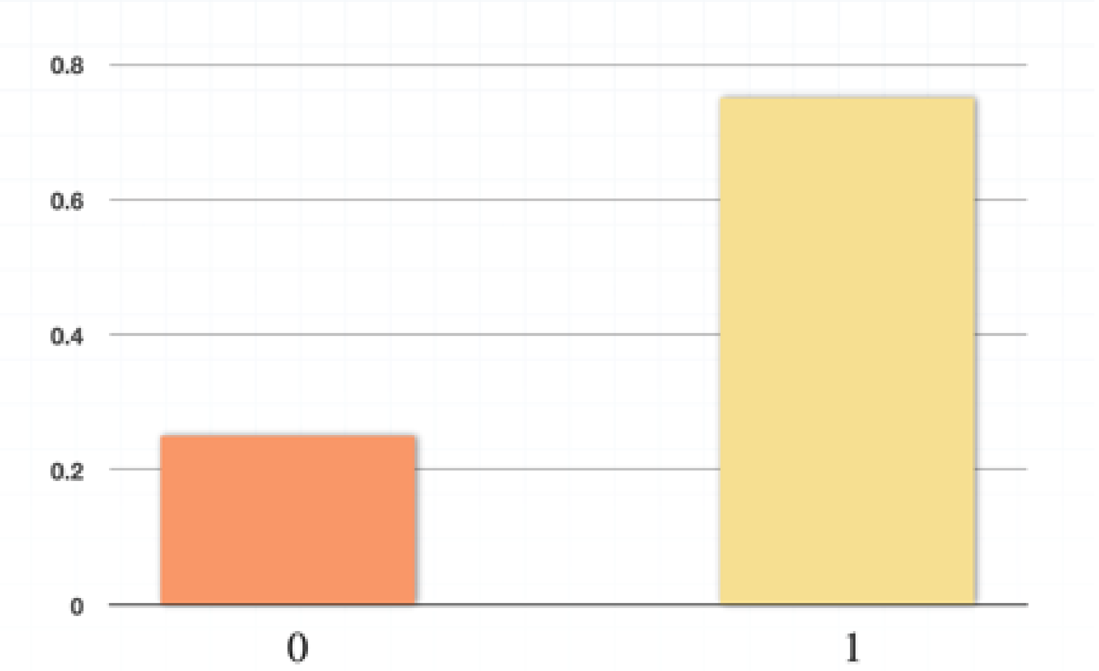
$$\operatorname{cov}(X, Y) \approx 0$$

Reference: https://www.internetgeography.net/scatter-graphs-in-geography/

DISTRIBUTIONS

Bernoulli distribution

- Models the set of possible outcomes for a *single* experiment.
- $X \in \{0,1\}$
- Parameter $\rho \in [0,1]$ reflects the probability of getting a 1.
- PMF: $f(x; p) = \rho^x (1-\rho)^{1-x}$
- $\mathbb{E}[X] = p$
- Example: tossing a biased coin.



DISTRIBUTIONS

Binomial distribution

- A generalization of Bernoulli for $\mathbb{N}$ random variables, i.e. $X \in \mathbb{N}$.
- Parameters $\rho \in [0,1], n \in \mathbb{N} = 0,1,2,3,\ldots$
- PMF: $f(x; p, n) = \binom{n}{k} \rho^x (1-\rho)^{n-x}$
- $\mathbb{E}[X] = n\rho$

06

DISTRIBUTIONS

Binomial distribution

- Let’s have a closer look at the PMF:

$$f(x; p, n) = \binom{n}{k} \rho^{x} (1-\rho)^{n-x}$$

- Combination

$$\binom{n}{k} = \frac{n!}{n!(n-k)!}$$

Where n is the total number of possible outcomes,

k is number of items you want to rearrange.

Useful materials

YouTube

Intelligent Systems Lab
@intelligentsystemslab907
4.7K subscribers

jbstatistics
@jbstatistics
182K subscribers

3Blue1Brown
@3blue1brown
4.96M subscribers

StatQuest with Josh Starmer
@statquest
876K subscribers

Textbook