STATISTICS FOR DATA SCIENCE

Probability

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Probability

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

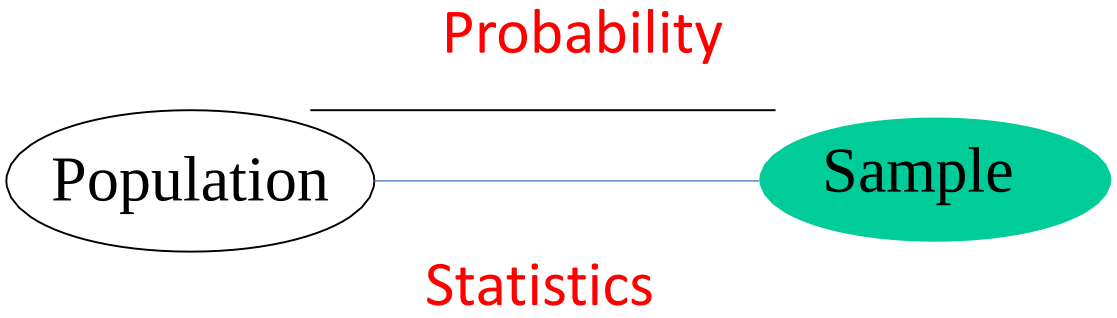
STATISTICS FOR DATA SCIENCE

Topics to be covered...

- Probability – Basics
- Random variables
- Types of Random variables

Why should we learn Probability?

- Nothing in life is certain. In everything we do, we gauge the chances of successful outcomes.
- A probability provides a quantitative description of the chances or likelihoods associated with various outcomes
- It provides a bridge between descriptive and inferential statistics.
- Probability theory – foundation for statistical inference



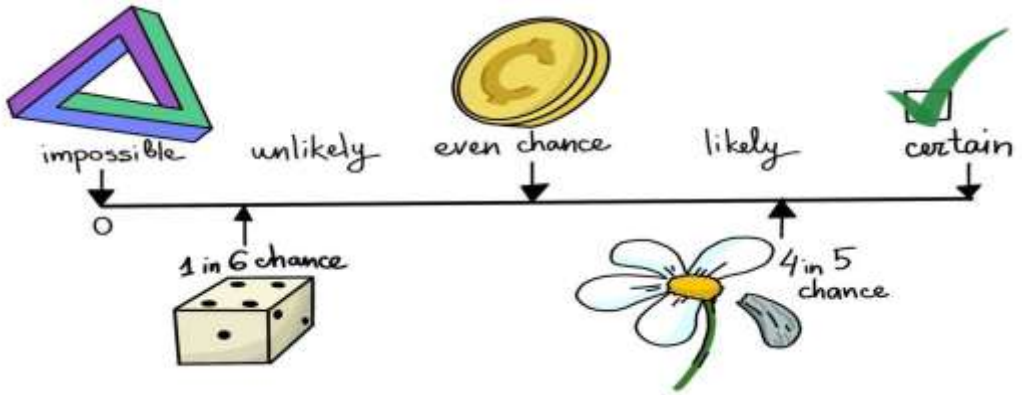
Probability:

Definition: Probability is a numerical index of the likelihood that a certain event will occur.

Examples:

What is the probability that Good Friday will come on a Monday?

What is the probability that Deepawali will come on a Friday?



Source: KDnuggests.com

Outcome: The **result of a single trial** of an **experiment**.

Ex : Getting **Head or Tail** is a result when you toss a coin.

Event : A collection of one or more outcomes of an experiment.

(or) A subset of a sample space is called an **event.**

Ex : Getting a sum as 6 when we roll dice.

A={(1,5),(2,4),(3,3),(4,2),(5,1)}

Sample Space: The set of **all possible outcomes** of an experiment is called the **sample space** for the experiment.

Ex : S={1,2,3,4,5,6}

Note: For any sample space, the **empty set ϕ** is an event, as is the **entire sample space**.

STATISTICS FOR DATA SCIENCE

What is Probability?

- Probability is the way of expressing knowledge of belief that an event will occur on chance.
- The idea of probability is based on observation.
- Probability describes what happens over many, many trials.

Probability Formula:

$$P(A) = \frac{\text{Number of favorable outcomes to A}}{\text{Total number of outcomes}}$$

Text-Book-Chapter 2.1- Page.No:48

STATISTICS FOR DATA SCIENCE

Probability

Experiment	Outcomes	Event	Sample Space
Tossing a coin	Head, Tail	Getting a Head or Getting a Tail	S={Head, Tail}
Rolling a Die once	1, 2, 3, 4, 5 or 6	Getting an even number or Getting an odd number	S={1,2,3,4,5,6}
Taking a Test	A, B, C, D, E or F	Passing the Test	S={A,B,C,D,E,F}
Tossing a coin twice	(H,H),(H,T),(T,H),(T,T)	Getting at least one head Getting two tails	S={(H,H),(H,T), (T,H),(T,T)}
Rolling Dice		Dice add to 6	

Types of Probability

- **Subjective Probability**
  - Likelihood of a particular event estimated on the basis of individual intuition or some general information or some expert judgement.

Ex. : The probability that Yahoo will buy Google within the next 2 years is 0.1 or 10%.

The probability that RCB wins IPL next time is 0.8 or 80%.

- **Objective Probability**

The **relative possibility of occurrence** of an event defined as

P(A) = number of favorable outcomes/total number of outcomes
= m / n

Ex. : Let A= Getting 5 when you roll a die.
P(A) = 1/6

- **Event Relations**

The **intersection of two events**, A and B, is the event that both A and B occur when the experiment is performed and denoted by A∩B.

Ex.: Let A= An even number turns up

A = {2,4,6}

B= The number that turns up is divisible by 4

B={4}

A∩B = {4}

If two events A and B are **mutually exclusive**, then **P(A∩B)=0**.

Ex. : Drawing a king or an ace from a deck of cards.

The **complement of an event** A consists of all outcomes of the experiment that do not result in event A and written as A^c.

- **The Additive Rule for Unions**

For any two events A and B, the probability of their union, P(AUB) is

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

When two events A and B are mutually exclusive,

$$P(A \cup B) = P(A) + P(B) \quad \text{and} \quad P(A \cap B) = 0$$

- **Probability for Complement**

For any event A

$$P(A^c) = 1 - P(A)$$

- **Independent Events**

Two events A and B, said to be **independent** if the occurrence or non-occurrence of one of the events does not change the probability of the occurrence of the other event.

Ex. : Toss a fair coin twice.
Define A : head on first toss
B : head on second toss

P(A) does not change whether B happens or not.

- **Conditional Probability**

The probability that A occurs, given that event B has occurred is called the **conditional probability of A given B** and is defined as

$$P(A|B) = P(A \cap B) / P(B) \quad \text{if } P(B) \neq 0$$

- **Conditional Probability**

Example : In a group of 100 sports car buyers, 40 bought alarm systems, 30 purchased bucket seats, and 20 purchased an alarm system and bucket seats. If a car buyer chosen at random bought an alarm system, what is the probability they also bought bucket seats?

A : Bought bucket seats
B : Bought alarm systems

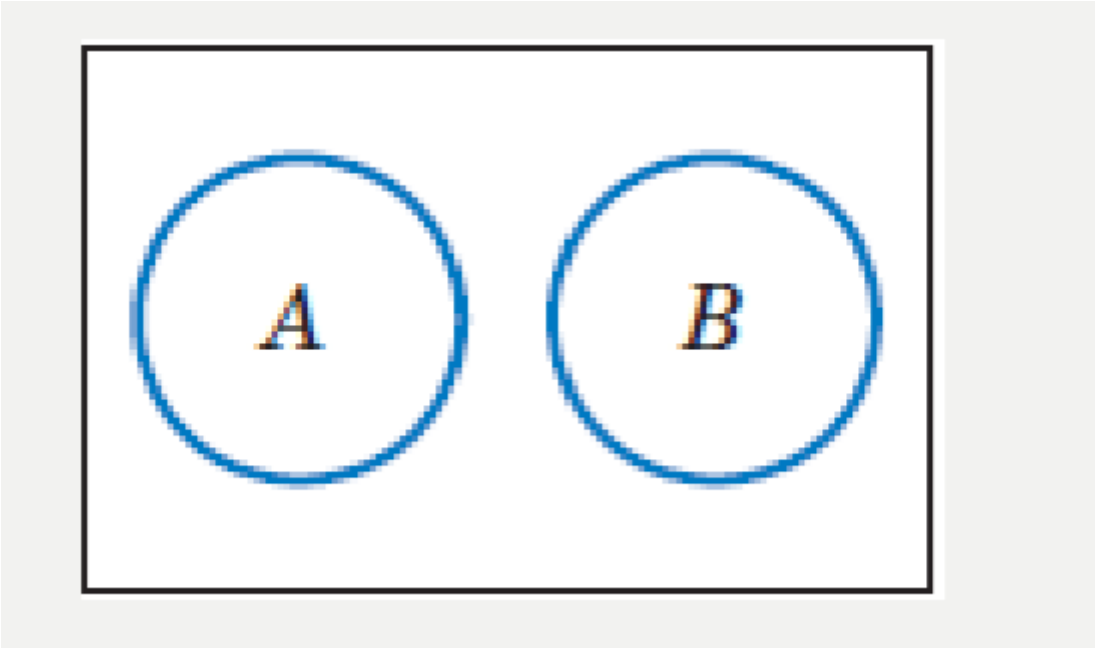
$P(A|B) = P(A|B) = P(A \cap B) / P(B) = 0.2 / 0.4 = 0.5 =$ **50%**

The **probability that a buyer bought bucket seats**, **given that** they **purchased an alarm system**, is **50%.**

STATISTICS FOR DATA SCIENCE

Mutually Exclusive Events

- The events A and B are said to be mutually exclusive if they have no outcomes in common.



Axioms of Probability

1. Let S be the sample space. Then P(S)=1
2. For any event A, $0 \leq P(A) \leq 1$.
3. If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

   More generally, if $A1, A2, \ldots$ are mutually exclusive events, then

$$P(A1 \cup A2 \cup \cdots) = P(A1) + P(A2) + \cdots .$$

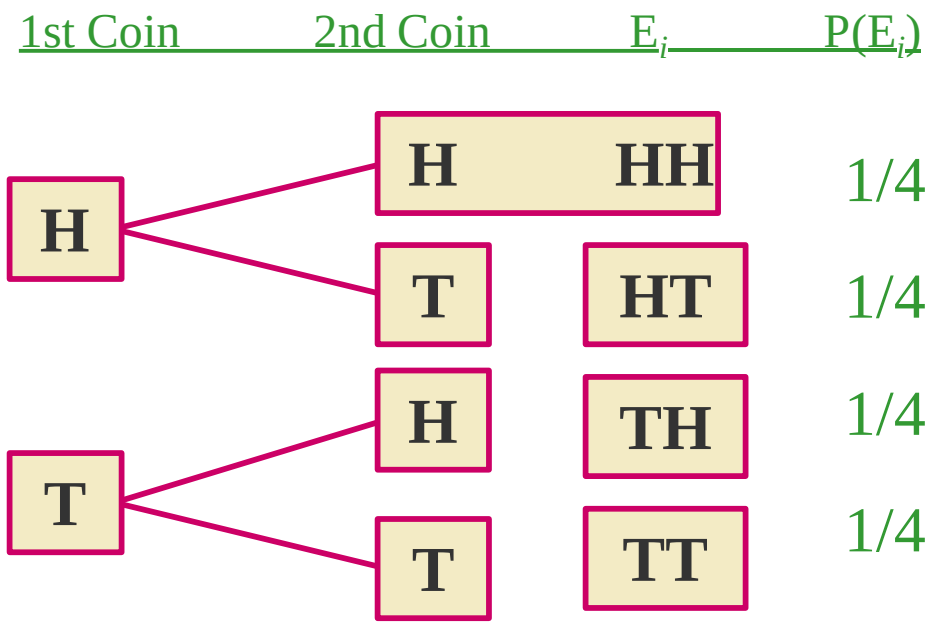
Axioms of Probability

- The probability that an event does not occur is 1 minus the probability that it does occur. (also called the complement of A)

Example: What is the probability of not rolling doubles?
P(not rolling doubles) = 1 – P(doubles)
= 1 – 6/36
= 30/36
= 83%

Example

1. Toss a fair coin twice. What is the probability of observing at least one head?



P(at least 1 head)

$= P(E_1) + P(E_2) + P(E_3)$

$= 1/4 + 1/4 + 1/4 = 3/4$

STATISTICS FOR DATA SCIENCE

Basic Concepts

- An **experiment** is the process by which an observation (or measurement) is obtained.
- The set of all possible outcomes of an experiment is called sample space for the experiment.
- A subset of sample space is called an event.

Example:

- **Experiment: Record an age**

  A: person is 30 yearsold

  B: person is older than65

Text-Book-Chapter 2.1- Page.No:48,49

Sample space-Example

Example 1:

$$P(\text{heads}) = \frac{1}{2} = 0.5$$

Tossing a coin.

The sample space is S = {H, T}. E = {H} is an event.

Example 2

Tossing a die.

The sample space is S = {1, 2, 3, 4, 5, 6}. E = {2, 4, 6} is an event, which can be described in words as ”the number is even”.

Source: https://faculty.math.illinois.edu/~kkirkpat/SampleSpace.pdf

STATISTICS FOR DATA SCIENCE

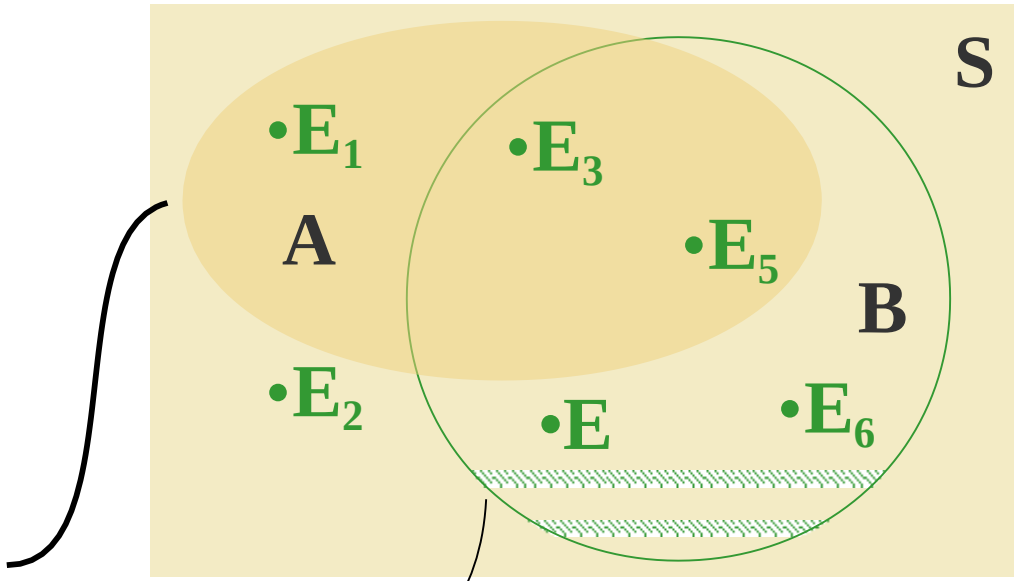
Basic Concepts

Event:

- **The die toss:**
  - A: an odd number
  - B: a number > 2

$A = \{E_1, E_3, E_5\}$

$B = \{E_3, E_4, E_5, E_6\}$



Source: www.slideshare.net

Do It Yourself !!!!

Example: In the process of tossing two coins.

- What is an experiment?
- What is an outcome?
- What is an event?
- What is a sample space?

Example: In rolling a six-sided die.

- What is an experiment?
- What is an outcome?
- What is an event?
- What is a sample space?

THANK YOU

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering