**Author: Tanishq Harit / tharit@stevens.edu / harit.tani@gmail.com**

# Introduction to Probability

## Probability Terminologies:

- **Random Experiment**: An experiment is called random experiment if it satisfies following conditions:
  - It has more than one possible outcome.
  - It is not possible to predict outcome in advance.
  - Example: Tossing a coin (possibilities - heads or tails)
- **Trial and Outcome**: Trial is single execution of a random experiment. Each trial produces an Outcome. In tossing a coin, when we toss the coin it is called trial and whether the coin is heads or tails is its outcome.
- **Sample Space**: Sample space of a random experiment is the set of all possible outcomes that can occur. Generally, one random experiment will have one **set** of sample space.
  - Again in tossing a coin example, we have sample space of {H,T} or {Heads, Tails}.
  - Another example is rolling a dice, which will have a sample space of {1, 2, 3, 4, 5, 6}.
- **Event**: Event is specific set of outcomes from a random experiment or process. Essentially, it's a subset of sample space. An event can include a single outcome, or it can include multiple outcomes. One random experiment can have multiple events.
  - In rolling a dice example, if we want to measure the probability of getting a number less than 3 then our event is "number less than 3" and this would be {1, 2}. As we can see, this is still subset of sample space {1, 2, 3, 4, 5, 6}.
  - We can have multiple events in rolling the dice example. Getting an odd number {1, 3, 5} or getting an even number {2, 4, 6}. Again, both of these events are subset of sample space {1, 2, 3, 4, 5, 6}.

## Examples:

Now, we will take some examples and for every example, we will understand their terminologies.

- Example 1 **Rolling a Dice**:
  - Random Experiment: Rolling the dice.
  - Trial: Rolling the die once.
  - Outcome: Suppose, when we rolled it, we got outcome of {3}. Maybe, if we roll again, we get outcome of {2} or {5} or any other number.
  - Sample Space: {1, 2, 3, 4, 5, 6}
  - Event: Could be anything, depends on the question or a person.
- Example 2 **Tossing a Coin Twice**: Tossing the coin once and then tossing it again.
  - Random Experiment: Tossing the coin twice.
  - Trial: Tossing the coin twice (once).
  - Outcome: Let's say, when first time we tossed it we got Heads and when we toss it for the second time, we get Tails. Som outcome is {Heads, Tails}.
  - Sample Space: {{H,H}, {H,T}, {T,H}, {T,T}}
  - Event: Again, event could be anything. Example: getting 2 heads then outcome would be {H,H} or getting at-least one head then outcome would be {{H,H}, {H,T}, {T,H}}

- Example 3 **Titanic Dataset Example**: There are 891 passengers, randomly picking out 1 passenger than calculating it's P-class.
  - Random Experiment: Randomly drawing out a passenger.
  - Trial: Randomly drawing out the passenger and finding it's p-class.
  - Outcome: Let's say, there are 3 p-class {1, 2, 3} then outcome can be {1} or {2} or maybe {3}.
  - Sample Space: Take any passenger, it's p-class would be in {1, 2, 3}. So, sample space is {1, 2, 3}.
  - Event: Again, it could be anything. Example: .A passenger is from p-class {3} or a passenger is not from p-class {2}.

# Types of Events:

- **Simple Event**: Event that consists of exactly one outcome. Also known as elementary event.
  - When rolling a die, getting a 6 is a simple but getting odd numbers is not a simple event.
- **Compound Event**: Consists of two or more simple events.
  - When rolling a die, event "rolling an odd number" is a compound event because it consists of 3 simple events: rolling a 1, rolling a 3 and rolling a 5.
- **Independent Event**: Two events are independent if the occurrence of one event does not affect the probability of the occurrence of the other event.
  - If we flip a coin and roll a die, the outcome of the coin flip does not affect the outcome of the die roll.
- **Dependent Events**: Events are dependent if the occurrence of one event does affect the probability of occurrence of the other event.
  - If we draw two cards from a deck of cards, the outcome of first draw affects the outcome of second draw as now there is 1 less card in the deck.
- **Mutually Exclusive Events**: Two events are mutually exclusive (or disjoint) if they cannot occur both at the same time.
  - When rolling a die, events "roll a 2" and "roll a 3" are mutually exclusive because a single roll of die cannot results in both.
- **Exhaustive Events**: A **set of events** is exhaustive if at-least one of the events must occur when the experiment is performed.
  - When we roll a die, the events "roll an even number" and "roll an odd number" are exhaustive because one or the other must occur.
- **Impossible Event and Certain Event**: As the name suggests.
  - Impossible: When rolling a die, event of getting a 7 is impossible.
  - Certain: Similarly, when we roll die, it is certain that value x is going to be in the range of **1 ¡= x ¡= 6**.

## What is Probability:

In simplest terms, probability is a measure of likelihood that a particular event will occur. It is a fundamental concept in statistics and is used to make predictions and informed decisions. Probability is usually expressed as a number between 0 and 1, inclusive:

- A probability of 0 means that an event will not happen.
- A probability of 1 means that an event will certainly happen.
- A probability of 0.5 means that an event will happen half the time.

## Empirical Probability vs Theoretical Probability:

Empirical Probability, also known as experimental probability, is a probability measure that is based in observed data, rather than theoretical assumptions. It is calculated as the ratio of the number of times a particular event occurs to the total number of trials.

- **Numerical 1:** Suppose that, in our 100 tosses, we get heads 55 times and tails 45 times. What is empirical probability of getting a head?

  **Solution:** Heads = 55, Trails = 100, So, P(H) = 55 / 100 = 0.55 = 55% ans.

- **Numerical 2:** Let's say you have a bag with 50 marbles. Out of these 50 marbles, 20 are red, 15 are blue, and 15 are green. You start to draw marbles one at a time, replacing the marble back into the bag after each draw. After 200 draws, you find that you've drawn a red marble 80 times, a blue marble 70 times, and a green marble 50 times. What is the empirical probability of getting a red marble?

  **Solution:** Easy, Trials = 200 times, Red marble = 80 times, So, P(R) = 80/200 = 0.4 = 40% ans.

Theoretical or classical probability is used when each outcome in sample space is equally likely to occur. If we denote an event of interests at event A, we can calculate theoretical probability of event A as:

$$\text{Theoretical Probability of event } A = \frac{\text{No. of favorable outcomes (outcomes in event } A)}{\text{Total number of outcomes in Sample Space}}$$

- **Numerical 1:** Consider a scenario of rolling 2 dice (at same time). What is the probability of getting a sum of both dice = 7?

  **Solution:** Sample Space: {{1,1}, {1,2}, {1,3}.... {6,6}} = 36, Event: {{1,6}, {6,1}, {2,5}, {5,2}, {3,4}, {4,3}}

  So, P(Sum of 7) = 6 / 36 = 1 / 6 = 0.167 = 16.67 % ans.

- **Numerical 2:** Consider a scenario of tossing 3 coins. Find probability of getting exactly 2 heads.

  **Solution:** Sample Space: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} = 8,

  Event = {HHT, HTH, THH} = 3, So, P(exactly 3 heads) = 3 / 8 = 0.375 = 37.5% ans.

**Note:** When large number of trials are conducted in an experiment (infinite trials), then empirical probability gets close to theoretical probability.

- **Example:** If we toss a coin 10 times, suppose we get heads 3 times, then Empirical probability = 0.3 but Theoretical probability = 0.5. Now, if we toss the coin 100 times, this time we may get heads 45 times, then EP = 0.45. Similarly, if we toss 1000 times, we may get 470 heads, then EP = 0.47. See, as we increase the number of trials we are getting closer to Theoretical Probability of 0.5.