Statistical Techniques for Data Science

Probability

Introduction to Probability, Conditional Probability and Bayes Theorem

Dr. Subhabaha Pal
Manipal Global Academy of Data Science

Objective

After attending this session, you will be able to –

- **Explain what is Probability**
- **Describe what is meant by conditional probability**
- **Describe Bayes Theorem**

Different Scenarios

manipalglobal
Academy of Data Science

- We come across the following statements in day-to-day life
- Some statements are deterministic, i.e., you can directly determine the outcome using some mathematical expression
- Outcomes of certain statements can not be determined directly using any mathematical expression
- We call these as probabilistic statements

1. Initial Velocity, Acceleration and Time given. What will be final velocity after time t?

1. v = u + at,
   u = Initial velocity
   v = Final velocity
   a = Acceleration
   t = Time

Deterministic

2. Principal amount, ROI and time given. What will be simple interest?
I = PRT/100,
   I = Simple Interest
   P = Principal amount
   R = Rate of interest
   T = Duration of time

What is the difference between these statements?

Consider the example

3. Can it rain tomorrow?

Probabilistic

4. The price of gold on 11.11.2016 was Rs.3105.00 per gram. Will it remain the same on 18.11.2016?

Random Experiments

- We have learnt about deterministic and probabilistic statements in the last slide
- In order to understand what is probability, we should have concepts of experiment and random experiment
- Experiment is a planned process of data collection consisting of number of trials (replications) under similar identical conditions
- An experiment is termed as random experiment when the results are unpredictable
- Throwing a dice is a random experiment whereas experimenting with sugar and Sulphuric acid is not random experiment
- In the former you are not aware of the outcome whereas in the latter you are aware of the outcome

Trial and Outcome

- **Each time a random experiment is conducted, it is called a trial**
- **Result of a trial is called outcome**

Example – Suppose an IT engineer of a company is checking the laptops in store for distribution to the company's new employees. As those laptops were used by previous employees, those could be in perfect working condition or could be having some defects. Each laptop inspection by the IT engineer is a trial and getting the laptop in a perfect working condition or getting some defects in the laptop are the possible outcomes.

Exhaustive and Mutually Exclusive Outcomes

- **We have learnt about trials and outcomes in our previous slide**
- **All possible occurrence of outcomes in a trial together called exhaustive outcomes**
- **If a dice is thrown, the outcomes may be {1}, {2}, {3}, {4}, {5} and {6} and all these outcomes are exhaustive outcomes for the trial – throwing a dice – i.e., apart from these 6 outcomes, no other outcome can happen as result of the trial**
- **The outcomes which can not occur simultaneously are called mutually exclusive outcome**
- **In the dice throwing trial, the outcome {1} and {6} can not happen simultaneously and hence they are mutually exclusive outcomes**
- **In the same trial, the outcome {Appearance of odd number} and {6} can not also happen simultaneously and hence they are mutually exclusive outcomes**
- **However, the outcome {Appearance of even number} and {6} are not mutually exclusive**

Sample Space, Equally Likely and Independent Outcomes

manipalglobal
Academy of Data Science

- Exhaustive outcomes of an experiment together form the Sample Space of the experiment and it is denoted by Ω or S
- In the dice throw experiment, Ω is {1, 2, 3, 4, 5, 6}
- If every outcome of a random experiment has equal chance of being occurred, it is called equally likely outcome
- In an unbiased dice throw experiment, occurrences of all individual outcomes {1}, {2}, {3}, {4}, {5} and {6} are equally likely and hence they are called equally likely outcomes
- Outcomes of a random experiment are said to be independent of each other if occurrence of one outcome does not depend on the occurrence of another outcome
- {1} and {3} are independent outcomes whereas {1} and {Odd Number} are not independent

Event and Favorable Event

- An event is an outcome of an experiment usually denoted by capital letter
- The occurrence of {1} or {an even number} in dice throw is an event
- An event is said to be favourable to an outcome if the occurrence of the event leads to occurrence of the outcome
- The occurrence of the outcome {1} in dice throw is favourable for occurrence for the outcome {Odd Number} in dice throw
- Hence occurrences of {1}, {3} and {5} are favourable events for occurrence of the event {odd number}
- {1}, {2}, {3} are the elementary events whereas {even number} or {odd number} are not elementary events as these contain multiple elementary events

Complement of an Event

- **If event A is a subset of sample space S, the complement of A contains the elements of S that are not members of A**
- **The symbol for the complement of an event A is A^C**
- **A die is rolled and let A be the event that the roll yields a number greater than 4, i.e. A = {5, 6}, then A^C will be (1, 2, 3, 4}**

Definition of Probability

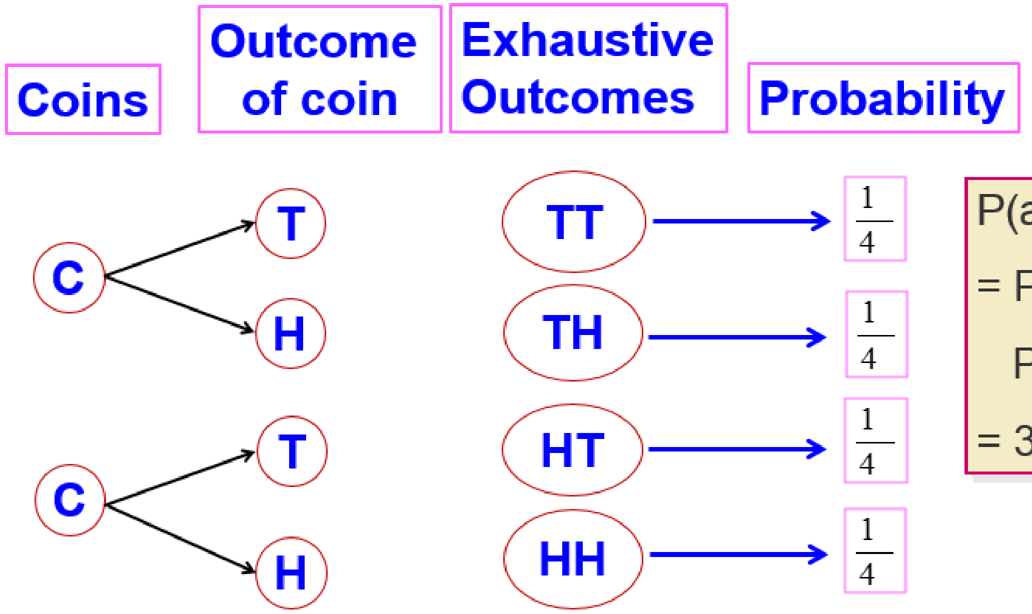
manipalglobal
Academy of Data Science

- **In a random experiment, out of 'n' exhaustive, mutually exclusive, equally likely, independent outcomes if 'm' of them are favourable to the occurrence of an event, say 'A', then the probability of an event 'A', denoted by P(A), is** $0 \leq \mathrm{P(A)} = \dfrac{\text{Events favourable to A}}{\text{Exhuastive outcomes}} = \dfrac{m}{n} \leq 1$ **.**
- **Suppose one dice is thrown and the possible outcomes from the dice throw will be {1}, {2}, {3}, {4}, {5} and {6}**
- **The probability of the event, appearance of {even number}, is 0.5 or 3/6 (because number of favourable outcomes for this event is 3 – {2}, {4} and {6} – total number of possible outcomes are 6 only**

Examples

manipalglobal
Academy of Data Science

- **Example 1 - A company has called for an interview for an executive position and 3 male and 2 female candidates have appeared for interview. What will be the probability of selection of a male candidate if all candidates are equally fit for the position? (It will be 3/5 or 0.6)**
- **Example 2 – A fair coin is tossed twice. What is the probability of getting at least one head?**



P(at least 1 head)
= P(TH) + P(HT)+ P(HH)
= 3/4

Examples

- **Example 3 – A bowl contains three balls – one red, one blue and one green. A child selects two balls at random. What is the probability that at lease one is a red ball?**

1st ball	2nd ball	Outcomes	Probability
R	B	RB	$\frac{1}{6}$
	G	RG	$\frac{1}{6}$
B	R	BR	$\frac{1}{6}$
	G	BG	$\frac{1}{6}$
G	B	GB	$\frac{1}{6}$
	R	GR	$\frac{1}{6}$

P(at least 1 red)

= P(RB) + P(BR)+ P(RG) + P(GR)

= 4/6

= 2/3

Types of Probability

- **Theoretical Probability – It is the likeliness of an event happening based on all possible outcomes. The probability of an event is P(event) = number of favourable outcomes divided by number of possible outcomes**

- **Example of Theoretical Probability – From the letters A, E, I, O, U, the theoretical probability of selecting the letter E is 1/5**

- **Relative Frequency Interpretation of Probability – Assume that an experiment can be repeated many times and assume that there are one or more outcomes that can result from each repetition. Then, the probability of a given outcome is the number of times that the outcome occurs divided by the total number of repetitions.**

$$\text{Probability of an event} = \frac{\text{How many times an event occurs}}{\text{How many trials}}$$

Types of Probability

- **Example of Relative Frequency Interpretation of Probability – A die is rolled 100 times and {3} occurs 12 times. Then, as per Relative Frequency interpretation, Probability of occurring 3 is 12/100.**
- **Personal or Subjective Probability – Subjective probability is an estimate that reflects a person's opinion, or best guess about whether an outcome will occur. These are values (between 0 and 1 or 0 and 100%) assigned by individuals based on how likely they think events are to occur.**
- **Example of Subjective Probability – The probability of a candidate winning in an election based on opinion poll is 60%**

Probability Laws - Additive

- **Addition Law for Two Events – If A and B are any two events with probabilities P(A) and P(B), then the occurrence of either A or B is P(A or B) = $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ where $P(A \cup B)$ is the probability of occurrence of either the event A or the event B or both, $P(A \cap B)$ is the probability of joint occurrence of A and B both**

Probability Laws – Mutually Exclusive Events

- **Addition Laws for 2 Events - If A and B are two mutually exclusive events with probabilities P(A) and P(B), then the occurrence of either A or B is $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$)**

- **Addition Laws for More than 2 Events - If $A_1, A_2, \ldots A_n$ are n mutually exclusive events with probabilities $P(A_1), P(A_2), \ldots, P(A_n)$, then the occurrence of either $A_1, A_2, \ldots, A_n$ is $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$**

Example – Additive Laws

manipalglobal
Academy of Data Science

Suppose that there were 120 students in the classroom, and that they could be classified as follows:

	Brown	Not Brown
Male	20	40
Female	30	30

A: brown hair
$P(A) = 50/120$
B: female
$P(B) = 60/120$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 50/120 + 60/120 - 30/120$$
$$= 80/120 = 2/3$$

Example – Additive Laws - Mutually Exclusive Case

When two events A and B are mutually exclusive,

$P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

	Brown	Not Brown
Male	20	40
Female	30	30

A: male with brown hair

$P(A) = 20/120$

B: female with brown hair

$P(B) = 30/120$

A and B are mutually exclusive, so that

$$P(A \cup B) = P(A) + P(B) = 20/120 + 30/120 = 50/120$$

Probability Laws – Multiplicative

manipalglobal
Academy of Data Science

- **Multiplicative Law for Two Events - If A and B are two independent events with probabilities P(A) and P(B), then the occurrence of A and B is P(A and B) = $P(A \cap B) = P(A) \times P(B)$**
- **Multiplicative Law for more than Two Events - If $A_1, A_2, \ldots A_n$ are n independent events with probabilities $P(A_1), P(A_2), \ldots, P(A_n)$, then the occurrence of A_1 and A_2 and …and A_n is**

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \times P(A_2) \times \ldots \times P(A_n)$$

Probability Rules – Mathematical Notations

PROBABILITY RULES

Rule 1. The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.

Rule 2. If S is the sample space in a probability model, then $P(S) = 1$.

Rule 3. Two events A and B are **disjoint** if they have no outcomes in common and so can never occur together. If A and B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B)$$

This is the **addition rule for disjoint events.**

Rule 4. For any event A,

$$P(A \text{ does not occur}) = 1 - P(A)$$

Conditional Probabilities

- In probability theory, the conditional probability is a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred
- If the event of interest is A and the event B is known or assumed to have occurred, “the conditional probability of A given B”, o “the probability of A under the condition B”, is generally written as P(A/B)
- Given two events A and B with P(B) > 0, the conditional probability of A given B is defined as the quotient of the probability of the joint events A and B, and the probability of B – $P(A/B) = P(A \cap B) / P(B)$ or $P(A \cap B) = P(A/B).P(B)$
- Also $P(A)P(B \mid A) = P(B)P(A \mid B)$

Conditional Probabilities – Examples

- **Example 1 - In a card game, suppose a player needs to draw two cards of the same suit in order to win. Of the 52 cards, there are 13 cards in each suit of same suit. Suppose the player draws a heart in the first attempt. Now the player wishes to draw a second heart. Since one heart has already been chosen, there are now 12 hearts remaining in a deck of 51 cards. So the conditional probability *P(Draw second heart|First card a heart)* = 12/51**

- **Example 2 - Suppose an individual applying to a college determines that he has an 80% chance of being accepted, and he knows that dormitory housing will only be provided for 60% of all of the accepted students. The chance of the student being accepted *and* receiving dormitory housing is defined by -**

  ***P(Accepted and Dormitory Housing)* = *P(Dormitory Housing|Accepted)P(Accepted)* =**

  (0.60)*(0.80) = 0.48.

Conditional Probabilities – Independent Events

- In case of independent events A and B, we know $P(A \cap B) = P(A) \times P(B)$

- We know from definition of conditional probability that –

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B)P(A \mid B)$$

- If A and B are independent, then conditional probability P(A/B) will become unconditional probability P(A), that is –

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Conditional Probabilities – Example

- **Example 3 - A manufacturer of airplane parts knows from the past experience that the probability is 0.80 that an order will be ready for shipment on time, and it is 0.72 that an order will be ready for shipment and will also be delivered on time. What is the probability that such an order will be delivered on time given that it was ready for shipment on time?**
- **Solution - Let R be the event that an order is ready for shipment on time and D be an event that it is delivered on time.**

$\therefore P(R) = 0.80$ **and** $P(R \cap D) = 0.72$; $\therefore P(D | R) = P(R \cap D)/P(R) = 0.72/0.80 = 0.90$

Thus, 90% of the shipments will be delivered on time provided they are shipped on time.

Bayes Theorem

Bayes' theorem is stated mathematically as the following equation –

P(A/B) = (P(B/A).P(A))/P(B)

Where A and B are events and P(B)≠0

- **P(A) and P(B) are the probabilities of observing A and B without regard to each other**
- **P(A/B) is the conditional probability of occurrence of A provided event B has already occurred**
- **P(B/A) is the conditional probability of occurrence of B provided event A has already occurred**

Note – Bayes Theorem is widely used for Naïve Bayes Classification Problem

Bayes Theorem - Example

- **Example 1 -** The following simple example shows how the Naïve Bayes classifier can be used for student performance prediction problem. Suppose we have the data of 10 students on the attribute previous academic performance and whether they have passed in the subject mathematics in the current term. The previous performance of the students are classified as – Less than 50% (LT 50), Between 50% to 80% (BT 50 to 80) and Greater than 80% (GT 80).

Table – Training Data for Naïve Bayes Classification

Student	1	2	3	4	5	6	7	8	9	10
Previous Academic Performance	LT 50	GT 80	BT 50 to 80	BT 50 to 80	BT 50 to 80	LT 50	GT 80	GT 80	BT 50 to 80	LT 50
Performance in Mathematics	PASS	PASS	PASS	PASS	PASS	FAIL	PASS	PASS	FAIL	FAIL

Bayes Theorem – Example (cont.)

As per the data, likelihood or probabilities are as follows -

P(LT 50) = 3/10 = 0.3

P(BT 50 to 80) = 0.4

P(GT 80) = 0.3

P(Pass) = 0.7

P(Fail) = 0.3

Now, we need to classify a student who has secured BT 50 to 80 and predict whether he will pass or not, i.e., it is needed to find out the following probabilities –

P(Pass/BT 50 to 80) and P(Fail/BT 50 to 80).

Now, P(Pass/BT 50 to 80) = P(BT 50 to 80/Pass).P(Pass)/P(BT 50 to 80)

And P(Fail/BT 50 to 80) = P(BT 50 to 80/Fail).P(Fail)/P(BT 50 to 80).

Now, P(BT 50 to 80/Pass) means likelihood that a student who has passed obtained between 50% to 80% in the previous examination.

Bayes Theorem – Example (cont.)

P(BT 50 to 80/Pass) = 3/7

P(BT 50 to 80/Fail) = 1/3

Hence, P(Pass/BT 50 to 80) = P(BT 50 to 80/Pass).P(Pass)/P(BT 50 to 80) = (3/7)*(0.7)/(0.4) = 0.75

P(Fail/BT 50 to 80) = P(BT 50 to 80/Fail).P(Fail)/P(BT 50 to 80) = (1/3)*(0.3)/(0.4) = 0.25

Hence, a student who has secured between 50% to 80% in previous examination has more chance (75% chance) to pass the mathematics subject and only 25% chance to fail.

Hence, if the performance of the student who has secured between 50% to 80% in the previous examination is to be predicted through Naïve Bayes algorithm based on the training dataset mentioned above, it will be predicted that the student will pass the examination.

manipalglobal
Academy of Data Science