Conditional probability and Bayes Theorem

Conditional Probability

Conditional probability is one of the **types of probability** in **probability theory**, where the probability of one event is dependent on the other event already happened.

Simply put, conditional probability tells us the likelihood of the occurrence of an event based on the occurrence of some previous outcome.

When the probability of one event happening doesn't influence the probability of any other event, then events are called independent, otherwise dependent events.

Conditional Probability is defined as the probability of any event occurring when another event has already occurred. In other words, it calculates the probability of one event happening given that a certain condition is satisfied. It is represented as $P(A \mid B)$ which means the probability of A when B has already happened.

Applications

Healthcare and Diagnostics

Determining the probability of a patient having a specific disease given the results of diagnostic tests. Conditional probability is crucial in medical diagnoses and decision-making, helping healthcare professionals make informed decisions based on test results.

Conditional Probability

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

Where,

- **$P(A \cap B)$** represents the probability of both events A and B occurring simultaneously, and
- **$P(B)$** represents the probability of event B occurring.

How to Calculate Conditional Probability?

Step 1: Identify the Events. Let's call them Event A and Event B.
Step 2: Determine the Probability of Event A i.e., $P(A)$
Step 3: Determine the Probability of Event B i.e., $P(B)$
Step 4: Determine the Probability of Event A and B i.e., $P(A \cap B)$.
Step 5: Apply the **Conditional Probability Formula** and calculate the required probability.

Conditional Probability

For Example, let's consider the case of rolling two dice, sample space of this event is as follows:

{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

Now, consider an event A = getting 3 on the first die and B = getting a sum of 9.

Then the probability of getting 9 when on the first die it's already 3 is P(B | A), which can be calculated as follows:

All the cases for the first die as 3 are (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6).
In all of these cases, only one case has a sum of 9.
Thus, P(B | A) = 1/6.

In case, we have to find P(A | B),

All cases where the sum is 9 are (3, 6), (4, 5), (5, 4), and (6, 3).
In all of these cases, only one case has 3 on the first die i.e., (3, 6)
Thus, P(A | B) = 1/4.

Conditional Probability

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

Where,

- **P(A ∩ B)** represents the probability of both events A and B occurring simultaneously, and
- **P(B)** represents the probability of event B occurring.

P(A|B): Conditional probability of an event A given that the event B has already occurred.

P(A ∩ B): Joint probability that the two events A and B occur simultaneously.

P(B): Marginal probability of the event B.

Conditional Probability

Let us take an example that a disease is caused depending upon the mutation of two genes g1 and g2.

A data set was collected, let us say, for 300 patients with the disease and status of the mutation of the genes g1 and g2 were recorded.

		Gene g1 mutation	
		Yes	**No**
Gene g2 mutation	**Yes**	120	50
	No	45	85

P(g1 is mutated) = P(g1 = Yes) = 165/ 300 = 0.550
P(g1 is not mutated) = P(g1 = No) = 135 / 300 = 0.45
P(g2 is mutated) = P (g2 = Yes) = 170 / 300 = 0.567
P(g2 is not mutated) = P (g2 = No) = 130 / 300 = 0.433

These are marginal probabilities g1 and g2 either as mutated or not-mutated.

Conditional Probability

Now, let us calculate the probability of occurrence of both the genes i.e. the joint probability

P(g1 and g2 are mutated) = P(g1 = Yes and g2 = Yes) = 120 / 300 = 0.400

P(g1 is mutated but g2 is not mutated) = P(g1 = Yes and g2 = No) = 45 / 300 = 0.15

P(g1 is not mutated but g2 is mutated) = P(g1 = No and g2 = Yes) = 50 / 300 = 0.167

P(g1 and g2 are not mutated) = P(g1 = No and g2 = No) = 85 / 300 = 0.283

Conditional Probability

Now, let us try to the conditional probability i.e. the probability of an event given that the other event has occurred.

P(g1 is mutated / g2 is mutated) = P(g1 = Yes / g2 = Yes) = 120 / 170 = 0.705

P(g1 is mutated / g2 is not mutated) = P(g1 = Yes / g2 = No) = 45 / 130 = 0.346

P(g1 is not mutated / g2 is mutated) = P(g1 = No / g2 = Yes) = 50 / 170 = 0.294

P(g1 is not mutated / g2 are not mutated) = P(g1 = No / g2 = No) = 85 / 130 = 0.653

Conditional Probability

Now, let us see the relationship between the conditional probability, joint probability, and the marginal probability.

P(g1 is mutated / g2 is mutated) = P(g1 = Yes / g2 = Yes) = 120 / 170 = 0.705

=> P(g1 = Yes / g2 = Yes) = 120 / (120 + 50)

$\Rightarrow$ P(g1 = Yes / g2 = Yes) = (120/300) / ((120 + 50)/300)

$\Rightarrow$ P(g1 = Yes / g2 = Yes) = P(g1 = Yes and g2 = Yes) / P(g2 = Yes)

$\Rightarrow$ P(g1 = Yes / g2 = Yes) = Joint Probability of g1 and g2 as mutated / Marginal Probability of g2 as mutated

This can be generalized as

P(g1 = Yes / g2 = No) = P(g1 = Yes and g2 = No) / P(g2 = No)

P(g1 = No / g2 = Yes) = P(g1 = No and g2 = Yes) / P(g2 = Yes)

P(g2 = Yes / g1 = Yes) = P(g1 = Yes and g2 = Yes) / P(g1 = Yes)

And so on.

Conditional Probability

Now, let us come back to the calculation of the marginal probabilities.

		Gene g1 mutation	
		Yes	**No**
Gene g2 mutation	**Yes**	120	50
	No	45	85

$$P(\text{g1 is mutated}) = P(\text{g1} = \text{Yes}) = 165/\ 300 = 0.550$$
$$\Rightarrow P(\text{g1} = \text{Yes}) = (120 + 45)\ /\ 300$$
$$\Rightarrow P(\text{g1} = \text{Yes}) = (120\ /\ 300) + (45\ /\ 300)$$
$$\Rightarrow P(\text{g1} = \text{Yes}) = P(\text{g1} = \text{Yes and g2} = \text{Yes}) + P(\text{g1} = \text{Yes and g2} = \text{No})$$

But, we know that

$$P(\text{g1} = \text{Yes}\ /\ \text{g2} = \text{Yes}) = P(\text{g1} = \text{Yes and g2} = \text{Yes})\ /\ P(\text{g2} = \text{Yes})$$

Conditional Probability

It implies that

$$P(g1 = Yes) = P(g1 = Yes/\ g2 = Yes) \times P(g2 = Yes) + P(g1 = Yes\ /\ g2 = No) \times P(g2 = No)$$

This can be generalized as

$$P\ (X) = P\ (X/Y)\ P(Y) + P(X/{\sim}Y)\ P\ ({\sim}Y)$$

where ~Y means “not Y”.

Conditional Probability

Conditional Probability of Independent Events

When two events are independent, those conditional probability is the same as the probability of the event individually i.e., P(A | B) is the same as P(A) as there is no effect of event B on the probability of event A. For independent events, A and B, the conditional probability of A and B with respect to each other is given as follows:

$$P(B|A) = P(B)$$
$$P(A|B) = P(A)$$

Conditional Probability

Properties of Conditional Property

Property 1: Let's consider an event A in any sample space S of an experiment.

$$P(S|A) = P(A|A) = 1$$

Property 2: For any two events A and B of a sample space S, and an event X such that $P(X) \neq 0$,

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

Property 3: The order of set or events is important in conditional probability, i.e.,

$$P(A|B) \neq P(B|A)$$

Property 4: The complement formula for probability only holds conditional probability if it is given in the context of the first argument in conditional probability i.e.,

$$P(\bar{A}|B)=1-P(A|B)$$

$$P(A|\bar{B}) \neq 1-P(A|B)$$

Property 5: For any two or three independent events, the intersection of events can be calculated using the following formula:

$$P(A \cap B) = P(A)\,P(B) \text{ and } P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

Bayes Theorem

Bayes Theorem

What is Bayes' Theorem?

It is used to determine the conditional probability of event A when event B has already happened. The general statement of Bayes' theorem is "The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the event of B, given A and the probability of A divided by the probability of event B." i.e.

$$P(A/B) = \frac{P(B|A)P(A)}{P(B)}$$

where, **P(A)** and **P(B)** are the probabilities of events A and B

P(A|B) is the probability of event A when event B happens

P(B|A) is the probability of event B when A happens

Bayes Theorem

Now let us use the formula for conditional probability of two random variable X and Y.

$$P(X / Y) = P(X \text{ and } Y) / P(Y)$$

$$P(Y / X) = P(X \text{ and } Y) / P(X)$$

By re-arranging these terms, we get

$$P(Y / X)\, P(X) = P(X / Y)\, P(Y)$$

$$\Rightarrow P(Y / X) = P(X / Y)\, P(Y) / P(X)$$

This is known as the Bayes Theorem.

Bayes Theorem

Here, P(Y / X) is called Posterior Probability.

P(X / Y) is called likelihood of Y.

P(Y) is marginal probability of Y.

P(X) is marginal probability of X.

Bayes Theorem

Bayes Theorem Statement

Bayes' Theorem for n set of events is defined as,

Let E_1, E_2,…, E_n be a set of events associated with the sample space S, in which all the events E_1, E_2,…, E_n have a non-zero probability of occurrence. All the events E_1, E_2,…, E form a partition of S. Let A be an event from space S for which we have to find probability, then according to Bayes' theorem,

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{k=1}^{n} P(E_k)P(A|E_k)}$$

Bayes Theorem

Let us take an example of finding the probability of having a disease given that the test is positive.

The information given is that the prevalence of the disease (let us say) is 0.15%.

A rapid test kit has been prepared with
Sensitivity (true positive rate) of the test = 95%
Specificity (true negative rate) of the test =90%

What is the probability (chance) that a person has the disease given that he/she has already tested positive.

Bayes Theorem

Given the data, let us first try to identify the components of the Bayes Theorem.

Sensitivity = P(Test = +ve / Disease = +ve) = 0.95

Specificity = P(Test = -ve / Disease = -ve) = 0.90

Prevalence = P(Disease = +ve) = 0.0015

$$P(\text{Disease} = +ve \,/\, \text{Test} = +ve) = P(\text{Test} = +ve \,/\, \text{Disease} = +ve)\, P(\text{Disease} = +ve) \,/\, P(\text{Test} = +ve)$$
$$= 0.95 \times 0.0015 \,/\, P(\text{Test} = +ve)$$

$$P(\text{Test} = +ve) = P(\text{Test} = +ve \,/\, \text{Disease} = +ve)\, P(\text{Disease} = +ve) + P(\text{Test} = +ve \,/\, \text{Disease} = -ve)\, P(\text{Disease} = -ve)$$
$$= 0.95 \times 0.0015 + 0.10 \times 0.9985 = 0.101275$$

Bayes Theorem

Thus,

$$P(\text{Disease} = +ve \,/\, \text{Test} = +ve) = P(\text{Test} = +ve \,/\, \text{Disease} = +ve)\ P(\text{Disease} = +ve) \,/\, P(\text{Test} = +ve)$$
$$= 0.95 \times 0.0015 \,/\, P(\text{Test} = +ve)$$
$$= 0.95 \times 0.0015 \,/\, 0.101275$$
$$= 0.014$$

It implies that there is 1.4% (posterior probability) chance of having the disease.

Conditional Probability

Example 1:

Tossing a Coin: Let’s consider two events in tossing two coins be,
A: Getting a head on the first coin.
B: Getting a head on the second coin.

Sample space for tossing two coins is:
S = {HH, HT, TH, TT}

The conditional probability of getting a head on the second coin (B) given that we got a head on the first coin (A) is = $P(B|A)$

Since the coins are independent (one coin’s outcome does not affect the other), $P(B|A) = P(B) = 0.5$ (50%), which is the probability of getting a head on a single coin toss.

Conditional Probability

Example 2:

Drawing Cards: In a deck of 52 cards where two cards are being drawn, then let's consider the events be
A: Drawing a red card on the first draw, and
B: Drawing a red card on the second draw.

The conditional probability of drawing a red card on the second draw (B) given that we drew a red card on the first draw (A) is = $P(B|A)$

After drawing a red card on the first draw, there are 25 red cards and 51 cards remaining in the deck. So, $P(B|A) = 25/51 \approx 0.49$ (approximately 49%).

Conditional Probability

Example 3: A bag contains 5 red balls and 7 blue balls. Two balls are drawn without replacement. What is the probability that the second ball drawn is red, given that the first ball drawn was red?

Let the events be,
Event A: The first ball drawn is red.
Event B: The second ball drawn is red.

P(A) = 5/12

and P(B) = 4/11 (as first ball drawn is already red, thus only 4 red balls remain in the bag)

Therefore, the probability of the second ball drawn being red given that the first ball drawn was red is 4/11.

Conditional Probability

Example 4: A box contains 5 green balls and 3 yellow balls. Two balls are drawn without replacement. What is the probability that both balls are green?

Let events be:
Event A: The first ball drawn is green, and
Event B: The second ball drawn is green.

P(A) = 5/8

P(B) = 4/7 (as there are 4 green balls left out of 7)

Thus, the probability that both balls drawn are green is (5/8) × (4/7) = 20/56 = 5/14.

Conditional Probability

Example 5: In a bag, there are 8 red marbles, 4 blue marbles, and 3 green marbles. If one marble is randomly drawn, what is the probability that it is not blue?

Let the events be:
Event A: The marble drawn is not blue, and
Event B: The marble drawn is blue.

As A and B are complementary Events, we know
$P(A) = 1 – P(B)$
$\Rightarrow P(A) = 1 – 4/15$
$\Rightarrow P(A) = (15 – 4)/15$
$\Rightarrow P(A) = 11/15$

Thus, probability of drawing a marble out of bag which is not blue is 11/15.

Conditional Probability

Example 6: In a survey among a group of students, 70% play football, 60% play basketball, and 40% play both sports. If a student is chosen at random and it is known that the student plays basketball, what is the probability that the student also plays football?

Let’s assume there are 100 students in the survey.
Number of students who play football = $n(A) = 70$
Number of students who play basketball = $n(B) = 60$
Number of students who play both sports = $n(A \cap B) = 40$

To find the probability that a student plays football given that they play basketball, we use the **conditional probability formula**:

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

Substituting the values, we get:
$P(A|B) = 40 / 60 = 2/3$

Therefore, the probability that a randomly chosen student who plays basketball also plays football is 2/3.

Conditional Probability

Example 7: In a deck of 52 playing cards, 4 cards are drawn without replacement. What is the probability that all 4 cards are aces, given that the first card drawn is an ace?

Let the events be,
Event A: The first card drawn is an ace,
Event B: The second card drawn is an ace,
Event C: The third card drawn is an ace, and
Event D: The fourth card drawn is an ace.

P(A) = 4/52 (there are 4 ace out of 52)
P(B | A) = 3/51 (one is already drawn, thus 3 ace left)
P(C | A and B) = 2/50 (two is already drawn, thus 2 ace left)
P(D | A and B and C) = 1/49 (three is already drawn, thus 1 ace left)

To find the probability that all four cards are aces, we multiply the probabilities of the individual events.
P(A and B and C and D) = P(A) × P(B|A) × P(C|A and B) × P(D|A and B and C)
= (4/52) × (3/51) × (2/50) × (1/49)
= 1/270725

Therefore, the probability that all 4 cards drawn are aces, given that the first card drawn is an ace, is 1/270725**.**

Bayes Theorem

Example 8: A person has undertaken a job. The probabilities of completion of the job on time with and without rain are 0.44 and 0.95 respectively. If the probability that it will rain is 0.45, then determine the probability that the job will be completed on time.

Let E_1 be the event that the mining job will be completed on time and E_2 be the event that it rains. We have,
$P(A) = 0.45$,

$P(\text{no rain}) = P(B) = 1 - P(A) = 1 - 0.45 = 0.55$

By multiplication law of probability,
$P(E_1) = 0.44$
$P(E_2) = 0.95$

Since, events A and B form partitions of the sample space S, by total probability theorem, we have
$P(E) = P(A)\,P(E_1) + P(B)\,P(E_2)$
$= 0.45 \times 0.44 + 0.55 \times 0.95$
$= 0.198 + 0.5225 = 0.7205$

So, the probability that the job will be completed on time is 0.684.

Bayes Theorem

Example 9: There are three urns containing 3 white and 2 black balls; 2 white and 3 black balls; 1 black and 4 white balls respectively. There is an equal probability of each urn being chosen. One ball is equal probability chosen at random. what is the probability that a white ball is drawn?

Let E_1, E_2, and E_3 be the events of choosing the first, second, and third urn respectively. Then,
$P(E_1) = P(E_2) = P(E_3) = 1/3$

Let E be the event that a white ball is drawn. Then,
$P(E/E_1) = 3/5$, $P(E/E_2) = 2/5$, $P(E/E_3) = 4/5$

By theorem of total probability, we have
$P(E) = P(E/E_1) \, . \, P(E_1) + P(E/E_2) \, . \, P(E_2) + P(E/E_3) \, . \, P(E_3)$
$= (3/5 \times 1/3) + (2/5 \times 1/3) + (4/5 \times 1/3)$
$= 9/15 = 3/5$

Bayes Theorem

Example 10: A card from a pack of 52 cards is lost. From the remaining cards of the pack, two cards are drawn and are found to be both hearts. find the probability of the lost card being a heart.

Let E_1, E_2, $E_{3,}$ and E_4 be the events of losing a card of hearts, clubs, spades, and diamonds respectively.
Then $P(E_1) = P(E_2) = P(E_3) = P(E_4) = 13/52 = 1/4$.
Let E be the event of drawing 2 hearts from the remaining 51 cards. Then,
$P(E|E_1)$ = probability of drawing 2 hearts, given that a card of hearts is missing
$= {}^{12}C_2 / {}^{51}C_2 = (12 \times 11)/2! \times 2!/(51 \times 50) = 22/425$
$P(E|E_2)$ = probability of drawing 2 clubs ,given that a card of clubs is missing
$= {}^{13}C_2 / {}^{51}C_2 = (13 \times 12)/2! \times 2!/(51 \times 50) = 26/425$
$P(E|E_3)$ = probability of drawing 2 spades ,given that a card of hearts is missing
$= {}^{13}C_2 / {}^{51}C_2 = 26/425$
$P(E|E_4)$ = probability of drawing 2 diamonds ,given that a card of diamonds is missing
$= {}^{13}C_2 / {}^{51}C_2 = 26/425$
Therefore,
$P(E_1|E)$ = probability of the lost card is being a heart, given the 2 hearts are drawn from the remaining 51 cards
$= P(E_1) . P(E|E_1)/P(E_1) . P(E|E_1) + P(E_2) . P(E|E_2) + P(E_3) . P(E|E_3) + P(E_4) . P(E|E_4)$
$= (1/4 \times 22/425) / \{(1/4 \times 22/425) + (1/4 \times 26/425) + (1/4 \times 26/425) + (1/4 \times 26/425)\}$
$= 22/100 = 0.22$
Hence, The required probability is 0.22.

Bayes Theorem

Example 11: Suppose 15 men out of 300 men and 25 women out of 1000 are good orators. An orator is chosen at random. Find the probability that a male person is selected. Assume that there are equal numbers of men and women.

Let there be 1000 men and 1000 women.

Let E_1 and E_2 be the events of choosing a man and a woman respectively. Then,
$P(E_1) = 1000/2000 = 1/2$, and $P(E_2) = 1000/2000 = 1/2$

Let E be the event of choosing an orator. Then,
$P(E|E_1) = 50/1000 = 1/20$, and $P(E|E_2) = 25/1000 = 1/40$

Probability of selecting a male person ,given that the person selected is a good orator
$P(E_1/E) = P(E|E_1) \times P(E_1)/ P(E|E_1) \times P(E_1) + P(E|E_2) \times P(E_2)$
$= (1/2 \times 1/20) /\{(1/2 \times 1/20) + (1/2 \times 1/40)\}$
$= 2/3$

Hence the required probability is 2/3.

Bayes Theorem

Example 12: A man is known to speak the lies 1 out of 4 times. He throws a die and reports that it is a six. Find the probability that is actually a six.

In a throw of a die, let
E_1 = event of getting a six,
E_2 = event of not getting a six and
E = event that the man reports that it is a six.
Then, $P(E_1) = 1/6$, and $P(E_2) = (1 – 1/6) = 5/6$
$P(E|E_1)$ = probability that the man reports that six occurs when six has actually occurred
= probability that the man speaks the truth
$= 3/4$
$P(E|E_2)$ = probability that the man reports that six occurs when six has not actually occurred
= probability that the man does not speak the truth
$= (1 – 3/4) = 1/4$
Probability of getting a six ,given that the man reports it to be six
$P(E_1|E) = P(E|E_1) \times P(E_1)/P(E|E_1) \times P(E_1) + P(E|E_2) \times P(E_2)$ [by Bayes' theorem]
$= (3/4 \times 1/6)/\{(3/4 \times 1/6) + (1/4 \times 5/6)\}$
$= (1/8 \times 3) = 3/8$
Hence the probability required is 3/8.

Bayes Theorem

Exercise:
A rapid test kit has been prepared with
Sensitivity (true positive rate) of the test = 95%
Specificity (true negative rate) of the test =90%

P(Disease = +ve)	P(Disease = +ve / Test = +ve)
0.0010	
0.0020	
0.0025	
0.0030	
0.0035	
0.0040	
0.0045	
0.0050	

Calculate the probabilities and plot them at the Y-axis.

Thank You