Paper Name: Basic Statistics

Paper Code: MS 204

Unit II

Conditional Probability

Conditional probabilities arise naturally in the investigation of experiments where an outcome of a trial may affect the outcomes of the subsequent trials.

We try to calculate the probability of the second event (event B) given that the first event (event A) has already happened. If the probability of the event changes when we take the first event into consideration, we can safely say that the probability of event B is dependent of the occurrence of event A.

Let’s think of cases where this happens:

- Drawing a second ace from a deck given we got the first ace
- Finding the probability of having a disease given you were tested positive
- Finding the probability of liking Harry Potter given we know the person likes fiction

And so on….

Here we can define, 2 events:

- Event A is the probability of the event we’re trying to calculate.
- Event B is the condition that we know or the event that has happened.

We can write the conditional probability as $P\left(\frac{A}{B}\right)$, the probability of the occurrence of event A given that B has already happened.

$$P\left(\frac{A}{B}\right) = \frac{P(\text{A and B})}{P(B)} = \frac{\text{Probability of the occurence of both A and B}}{\text{Probability of B}}$$

Let’s play a simple game of cards for you to understand this. Suppose you draw two cards from a deck and you win if you get a jack followed by an ace (without replacement). What is the probability of winning, given we know that you got a jack in the first turn?

Let event A be getting a jack in the first turn

Let event B be getting an ace in the second turn.

We need to find $P\left(\frac{B}{A}\right)$

P(A) = 4/52

P(B) = 4/51 {no replacement}

P(A and B) = 4/52*4/51= 0.006

$$P\left(\frac{B}{A}\right) = \frac{\text{P(A and B)}}{P(A)} = \frac{0.006}{0.077} = 0.078$$

Here we are determining the probabilities when we know some conditions instead of calculating random probabilities. Here we knew that he got a jack in the first turn.

Let’s take another example.

Suppose you have a jar containing 6 marbles – 3 black and 3 white. What is the probability of getting a black given the first one was black too.

P (A) = getting a black marble in the first turn

P (B) = getting a black marble in the second turn

P (A) = 3/6

P (B) = 2/5

P (A and B) = ½*2/5 = 1/5

$$P\left(\frac{B}{A}\right) = \frac{\text{P(A and B)}}{P(A)} = \frac{0.2}{0.5} = 0.4$$

Let us now consider a new example and implement in R.

Bayes Theorem

Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

The Bayes theorem describes the probability of an event based on the prior knowledge of the conditions that might be related to the event. If we know the conditional probability $P\left(\frac{A}{B}\right)$, we can use the bayes rule to find out the reverse probabilities $P\left(\frac{B}{A}\right)$.

How can we do that?

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

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

$$P\ (A \cap B) = P\left(\frac{A}{B}\right) * P\ (B) = P\left(\frac{B}{A}\right) * P\ (A)$$

$$P\left(\frac{B}{A}\right) = P\left(\frac{A}{B}\right) * \frac{P\ (B)}{P\ (A)}$$

Statement: Let $A_1, A_2, \ldots, A_n$ be a set of events associated with a sample space S, where all the events $A_1, A_2, \ldots, A_n$ have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes' theorem,

$$P\,(A_i/B) = \frac{P(B|Ai) * P(Ai)}{\sum(i=1\ to\ n)\ P(B|Ai) * P(Ai)}$$

Examples of Bayes' Theorem and Probability trees

Let's take the example of the cancer patients. The patients were tested thrice before the oncologist concluded that they had cancer. The general belief is that 1.48 out of a 1000 people have cancer in the US at that particular time when this test was conducted. The patients were tested over multiple tests. Three sets of test were done and the patient was only diagnosed with cancer if she tested positive in all three of them.

Let's examine the test in detail.

Sensitivity of the test (93%) – true positive Rate

Specificity of the test (99%) – true negative Rate

Let's first compute the probability of having cancer given that the patient tested positive in the first test.

P (has cancer | first test +)

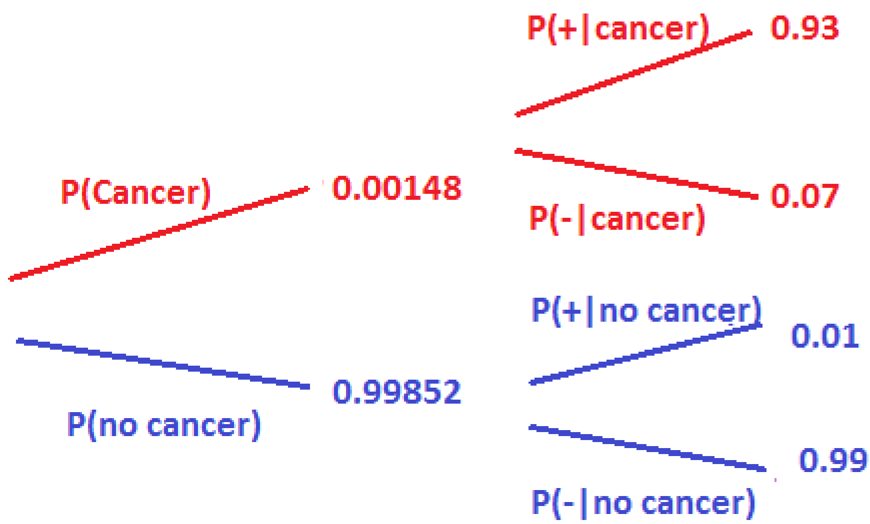
P (cancer) = 0.00148

Sensitivity can be denoted as P (+ | cancer) = 0.93

Specificity can be denoted as P (- | no cancer)

Since we do not have any other information, we believe that the patient is a randomly sampled individual. Hence our prior belief is that there is a 0.148% probability of the patient having cancer.

The complement is that there is a 100 – 0.148% chance that the patient does not have CANCER. Similarly we can draw the below tree to denote the probabilities.



Let's now try to calculate the probability of having cancer given that he tested positive on the first test i.e. P (cancer|+)

$$P\ (\text{cancer}|+) = \frac{P(cancer\ and+)}{P(+)}$$

P (cancer and +) = P (cancer) * P (+) = 0.00148*0.93

P (no cancer and +) = P (no cancer) * P(+) = 0.99852*0.01

To calculate the probability of testing positive, the person can have cancer and test positive or he may not have cancer and still test positive.

$$P\ (\text{CANCER}|+) = \frac{P(cancer\ and+)}{P(cancer\ and+) + P(no\ cancer\ and+)} = 0.12$$

This means that there is a 12% chance that the patient has cancer given he tested positive in the first test. This is known as the **posterior probability.**

Example 2: A factory has two Machines-I and II. Machine-I produces 60% of items and Machine-II produces 40% of the items of the total output. Further 2% of the items produced by Machine-I are defective whereas 4% produced by Machine-II are defective. If an item is drawn at random what is the probability that it is defective?

Probability of items produced by Machine 1

$$P(M_1) = 60/100$$

Probability of defective items produced by Machine 1

$$P(D/M_1) = 2/100$$

Probability of items produced by Machine 2

$$P(M_2) = 40/100$$

Probability of defective items produced by Machine 2

$$P(D/M_2) = 4/100$$

We need to find that if an item is drawn at random what is the probability that it is defective?

Randomly selected item will be defective either by machine 1 or machine 2.

$$P(D) = P(M_1) \cdot P(D/M_1) + P(M_2) \cdot P(D/M_2)$$

$$= (60/100)\ (2/100) + (40/100)\ (4/100)$$

$$= 120/10000 + 160/10000$$

$$= 280/10000$$

$$= 0.0280$$

Example 3: There are two identical urns containing respectively 6 black and 4 red balls, 2 black and 2 red balls. An urn is chosen at random and a ball is drawn from it. (i) find the probability that the ball is black (ii) if the ball is black, what is the probability that it is from the first urn?

Solution :

Total number of balls in Urn 1

$$= 6 \text{ black} + 4 \text{ red}$$

$$= 10 \text{ balls}$$

Probability of getting black ball from Urn 1

$$P(B/U_1) = 6/10$$

Total number of balls in Urn 1

$$= 2 \text{ black} + 2 \text{ red}$$

$$= 4 \text{ balls}$$

Probability of getting black ball from Urn 2

$$P(B/U_2) = 2/4$$

$P(U_1) = 1/2$ and $P(U_2) = 1/2$

(i) find the probability that the ball is black

$$P(B) = P(U_1) \cdot (P(B/U_1) + P(U_2) \cdot (P(B/U_2)$$

$$= (1/2) \cdot (6/10) + (1/2) \cdot (2/4)$$

$$= 3/10 + 1/4$$

$$= 11/20$$

(ii) if the ball is black what is the probability that it is from the first urn?

We need to find the probability that the selected black ball is from first urn.

$$= P(U_1) \cdot (P(B/U_1) / [P(U_1) \cdot (P(B/U_1) + P(U_2) \cdot (P(B/U_2)]$$

$$= (1/2) \cdot (6/10) / [(1/2) \cdot (6/10) + (1/2) \cdot (2/4)]$$

$$= (6/20) / (11/20)$$

$$= 6/11$$