Essentials of Data Science with R Software-1
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture 20
Independent Events

Hello friends. Welcome to the course Essentials of Data Science with R Software 1, in which we are trying to understand the topics of probability theory and statistical inference. So, we are still continuing with the topics of probability theory. Well, now, we have done different types of probabilities and now, in this short lecture today, I am going to introduce you with the concept of independence. What do you really mean by independence? That is a very simple word, two persons are said to be independent, if they are not dependent on each other.

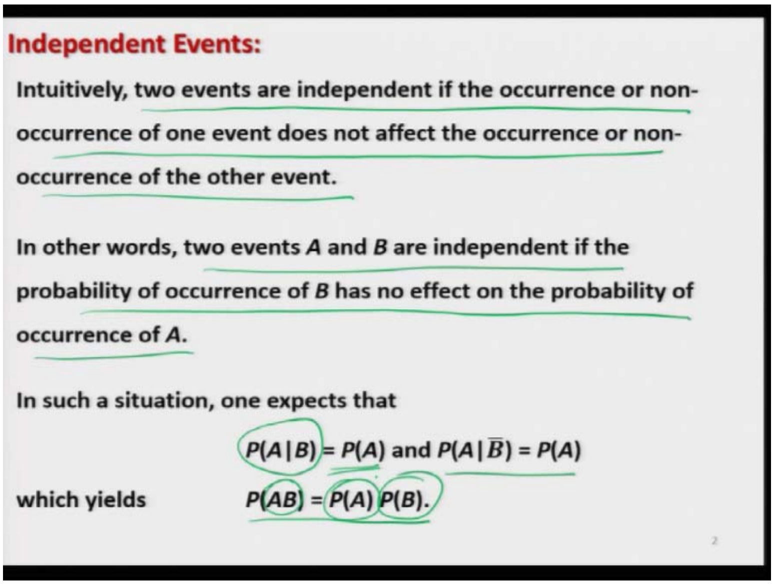
And in what sense they are independent? That the occurrence or non-occurrence of anything to that person does not affect me. But, on the other hand, if there is a friend or some person, suppose that person met with an accident. And suppose, if that affects me, means I get disturbed, and I feel unhappy about it. Do you think that I am independent of that man? Certainly not. That means, we have got some association, we have got some relationship.

Similarly, me and my parents are not independent. We do care for each other and we depend on each other. Similarly, in the case of statistics also, we have independence of events. Now, in case if you think about it, the thought process is very simple. If the occurrence of an event does not affect the occurrence of other events, we can simply say that both the events are independent of each other.

But definitely we are trying to do this concept in a mathematical way. So, we need some mathematical conditions through which we can make sure that the two events are independent or not. One thing I would like to clarify here that whenever we are trying to state the conditions for taking the independence of even there are various ways in which it can be done. At this moment, we have done only the probability theory.

So, in this lecture, I am going to talk about only the probability concept, means how using the concept of probability theory you can test, whether the two events are independent or not. And that will happen in a mathematical way. So, let us begin our lecture and try to understand this concept and try to use the, imply the concept of probability, to determine whether the two events are independent or not. Let us begin our lecture.

(Refer Slide Time: 03:14)



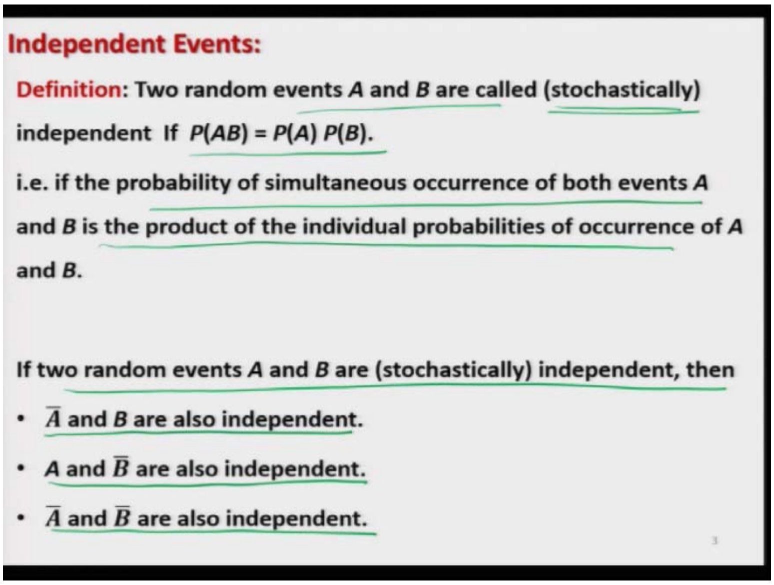
So, if you ask me clearly, two events are independent if the occurrence or non-occurrence of one event does not affect the occurrence or non-occurrence of the other event, no problem. Or in very simple words, I can say that two events A and B are independent if the probability of occurrence of B has no effect on the probability of occurrence of A.

Now, in such a situation if you try to bring in the concept of conditional probability, and if you try to compute the probability of A conditioned on B, and if this probability comes out to be the same as probability of A, that means, the occurrence of event A does not depend on the occurrence of event B. And even if you are trying to put a condition on the occurrence of event A in terms of given B, but that is not really going to affect it. And similar statement, we can also write that $P(A|\ \bar{B})$ is equal to P(A).

So, in this case, if you try to see here, this will yield P(A ∩ B) is equal to P(A) into P(B). So, now I can say that if the two events A and B are independent, what we can do? We can simply find out their joint probability of occurrence, that is P(A ∩ B).

We can find out the P(A)they are the occurrence of their individual events, P(B), and then we try to multiply together and if the P(AB) comes out to be the same as the P(A) into P(B), then we can say that the two events are independent, as simple as that.

(Refer Slide Time: 04:58)



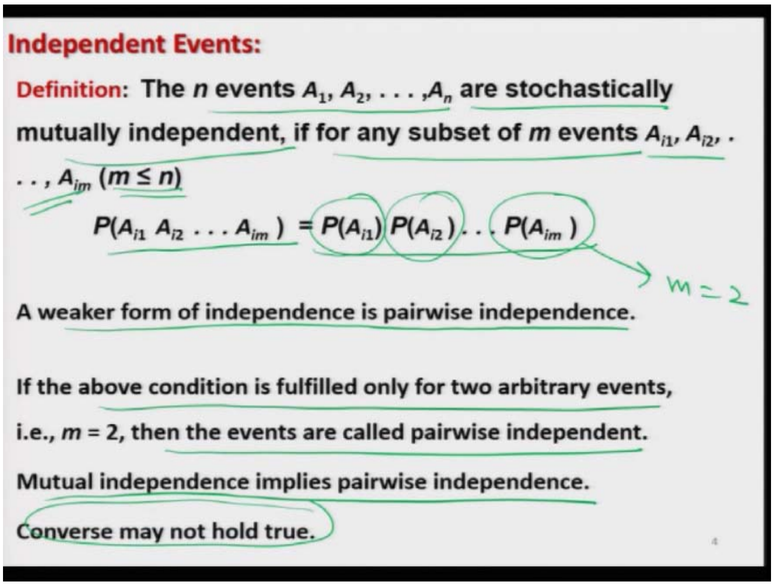
So, now I can formally define the independence of two random events A and B, or actually they are called actually as a stochastically independent or or statistically independent, like whatever we are saying here about the definition of independence, the independence is a statistical independence. But usually, many times you will see in the statistics, we do not use again and again the word statistical independence or say stochastically independent. Yes, stochastically means randomly means there is some randomness in the events A and B.

So, the two random events A and B are called independent or stochastically independent if P(A ∩ B) is the same as or can be expressed as P(A) into P(B). That is the probability of the simultaneous occurrence of both the events A and B is the product of their individual probabilities of occurrence of A as well as B.

Now, there is some more about this. There are some properties, which are related to the independence properties of two events A and B. And I am not giving you here the proof as I promised in the beginning of the lecture, but I will simply state it here and you can use it whenever you want it. But I am telling you, they are going to be very useful and you can take many conclusions directly without doing any algebra using these results.

So, now in case if we assume that the two random events A and B are stochastically independent then the first result says that the $\overline{A}$ and B they are also independent. A and $\overline{B}$ they are also independent as well as $\overline{A}$ and $\overline{B}$, they are also independent. So, if you are saying that there are two events A and B are independent. So, this result is related to the event and their complement, either one of the events is complement or both the events are complement. So, if you try to take any combination, they will also be independent.

(Refer Slide Time: 07:10)



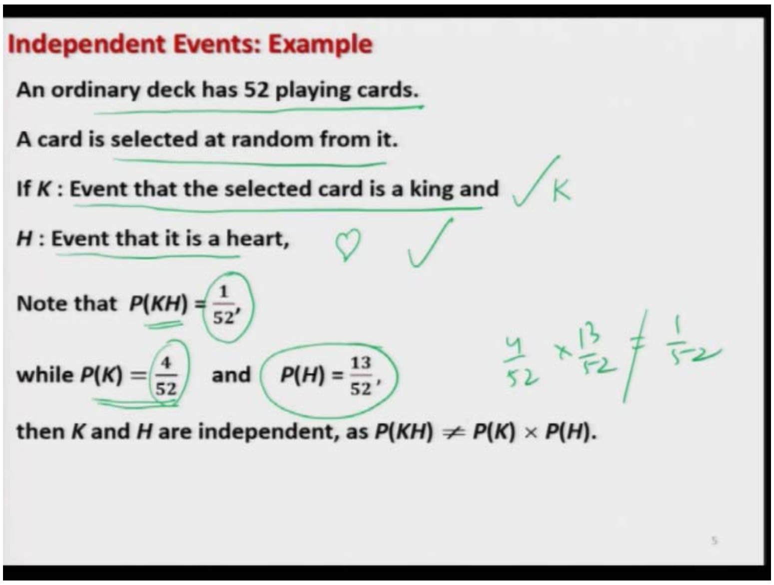
And now, this definition can also be extended to more than two events. Suppose, if I say there are n events A_1, A_2, …,A_n and they are stochastically mutually independent. When they are going to be independent? If for any subset of a small m events, when m is smaller than n or m is less than or equal to n.

And this subset is Ai1, Ai2,…, Aim means any subset from this A_1, A_2,…, A_m. For this subset the $P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{im})$ that is that can be expressed as the product of the individual probabilities of the events A_{i1}, A_{i2}, A_{im} that is $P(A_{i1})$ X $P(A_{i2})$ X $P(A_{i1})$.

And the and weaker form of this condition of this independent condition is pairwise independence. And we always said that, if this above condition is fulfilled only for two arbitrary event that is here, in this case, you are trying to take only here m equal to 2 then we call that the events are pairwise independent.

So, surely if there is a mutual independence, then this will imply the pairwise independence, but not the vice versa. The converse may not always hold true. So, you have to be very careful. So, that is the result which you always have to keep in mind whenever you are trying to deal with the independent events.

(Refer Slide Time: 08:59)



So, now, let me try to take an example here and try to show you here that how these things are working. Yeah, just for the sake of simplicity, it means I am just trying to consider here only one example, because I want to give you the idea that how are you really going to execute this concept of independence through the computation of probability and later on when we try to do some more concept, we will talk about this concept in more detail.

So, suppose, in the pack of a card or an ordinary deck of cards, there are 52 playing cards that we know there are 52 playing cards in a deck. And now, suppose you select one card at random. And suppose there are two events K and H and K is the event that the selected card is a King. And H is that event that it is a heart. You know that in cards there is a symbol K which we call as King and there is a symbol like this here is heart. So that we call those as heart.

So, now there are 13 hearts that we know and there are 4 kings. So, in case if you try to find out the probability of K i∩ H, then this is going to be 1/52. And similarly, if you try to find out the P(K), which is the probability of finding King that means, 4 out of 52. And similarly, if you want to find out the probability of here a heart, so there are 13 cards of this heart symbol. So, this probability of H will be 13/52.

Now, if you try to multiply probability of K and probability of H, which will be 4/52 into 13/52. Do you think that is it going to be 1/52? No. So, in this case, you can see that the product of their individual probabilities is not coming out to be the same as the probability of their joint occurrence. So, from this point of view, I can say that these two events K and H, they are not independent events. And this is what you have learned from the definition, which is using the probability theory.

So, now we come to an end this short lecture, as I said, but yeah, but this concept is very important. So, I will say now, why do not you take some examples from the book and try to practice this. And try to think about now that, how are you going to implement these concepts inside the software. If you have to compute a conditional probability, or if you want to compute a vision probability, or if you want to compute any other type of probability, how can you compute them through programming?

Now, I can give you a hint. When you want to compute the base probability differently that depends on simple probability and conditional probabilities. Simple probabilities you know how to count, for the conditional probability, how to find. Do not you think that we can use the logical operators? Well, that is my suggestion. But as a programmer, you are far much better and you can have different types of logics to compute the conditional probabilities.

So, in the next time, I am going to take up this issue, whatever possibilities we have considered, I will try to compute some of them through the software and I will try to give you some idea that how you can execute and compute these probabilities in the R software. So, it is important for you that before you go to the next lecture, all these concepts should be in your mind because if you do not understand these concepts, it will be difficult for you to understand that programming or the logic behind that programming. So, you try to revise these topics and I will see you in the next lecture. Till then, goodbye.