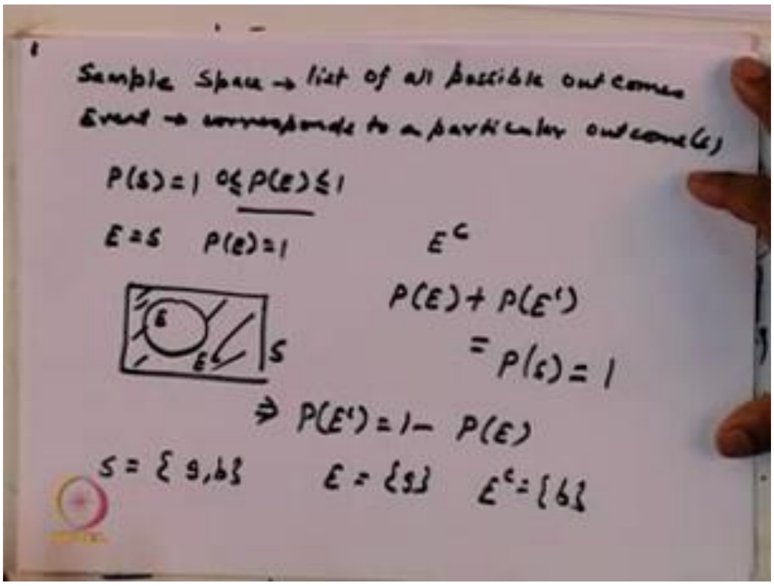
Introduction to Biostatistics

Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 17
Conditional probability

Hello and welcome to today's lecture. We had discussed about probability since last two classes, and that is what we will elaborate upon in today's lecture also. They will begin with a brief recap of what we had discussed in the last few lectures on probability.

(Refer Slide Time: 00:39)

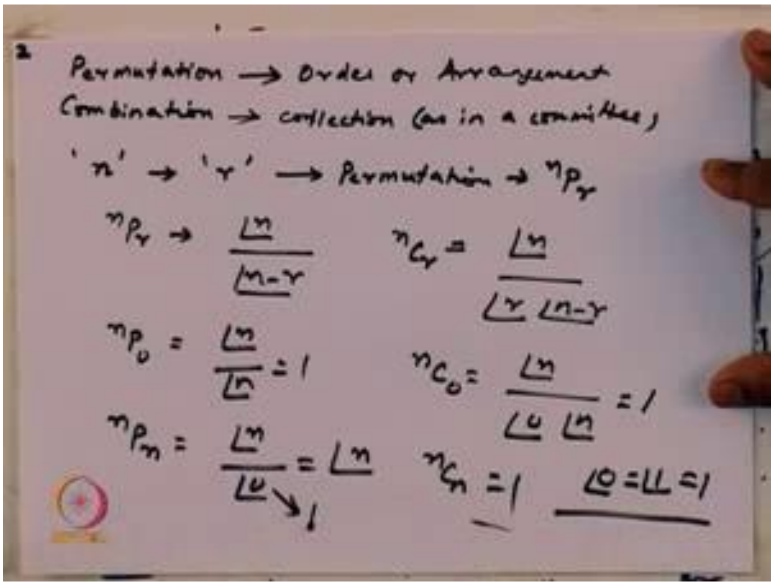


So, in probability we had introduced you to the concept of a sample, and a concept of A. So, sample space, which is a list of all possible events or outcomes, and an event which corresponds to a particular outcome or outcomes. So, in the language of probability, we define probability of sample. So, probability of a sample is equal to 1, and probability of an event E is bounded between 0 and 1. So, in the case where event E is equal to S, then probability of E is also equal to 1. So, you can represent events and samples using Venn diagrams.

So, imagine an event E, which is a subset of a sample. So, you would represent this information by drawing a circle E in a box, which corresponds to the complete sample. So, based on this then we could have. So, if this is event E then E. So, the complement of the event E is written as E C and E C is nothing but this entire area where E is not there.

So, this is E c. In other words you can write probability of E plus probability of E C, since together they constitute the sample. So, we have right P S is equal to 1 implying probability of E complement is equal to 1 minus probability of E. So, if you have the sample of a girl and a boy and you define E is equal to a girl, and then E complement has to be equal to boy. So, that E and E C put together will return you the sample space. So in order to calculate the probability of an event E what you have to do, is to do a good exercise of counting all the possibilities, and that brings us to permutation and combination respectively right, permutation and you have combination.

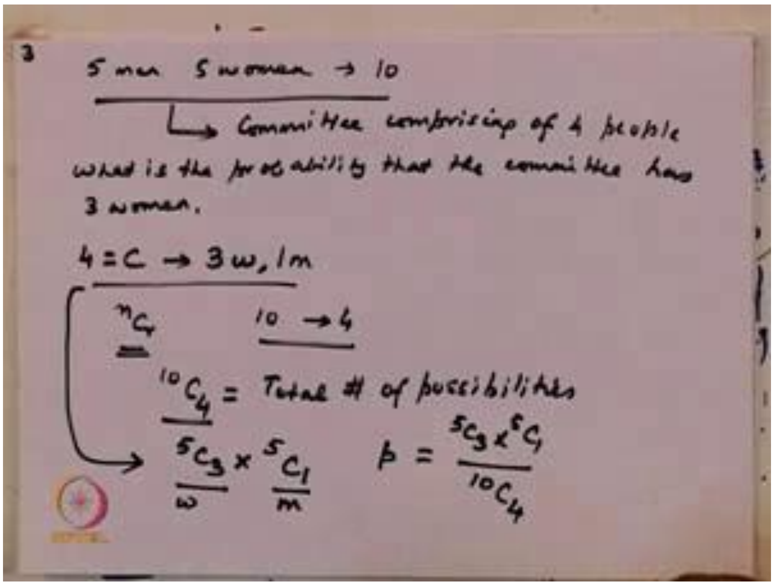
(Refer Slide Time: 03:16)



So, in when you talk about permutation you talk about order or arrangement, when you talk of combination you talk of a, you know you talk about a particular collection as in a committee. So, in case of a committee, the ordering does not matter. So, given n numbers you can arrange them in r different ways. So, if you are talking about arrangement then you are talking about permutation, and this is given by the expression n P r and n P r is defined by factorial n by factorial n minus r. So, n P 0 is equal to factorial n by factorial n is equal to 1, n P n is equal to factorial n by factorial 0 is equal to factorial n.

So, factorial n 0 is defined as 1. So, n objects can be placed, can be ordered in n different ways. And for expressing combinations you use the expression n C r, and this is equal to factorial n. So, you have factual divided by factorial n minus r and you have an additional factor of factorial r. So, n C 0 would give me factorial n by factorial 0 factorial n which is nothing but 1; similarly n C n will also be equal to 1. So, n P n gives me factorial n, but n C n gives me factor equal to 1, and you define factorial 0 equal to factorial 1 is equal to 1.

(Refer Slide Time: 05:42)

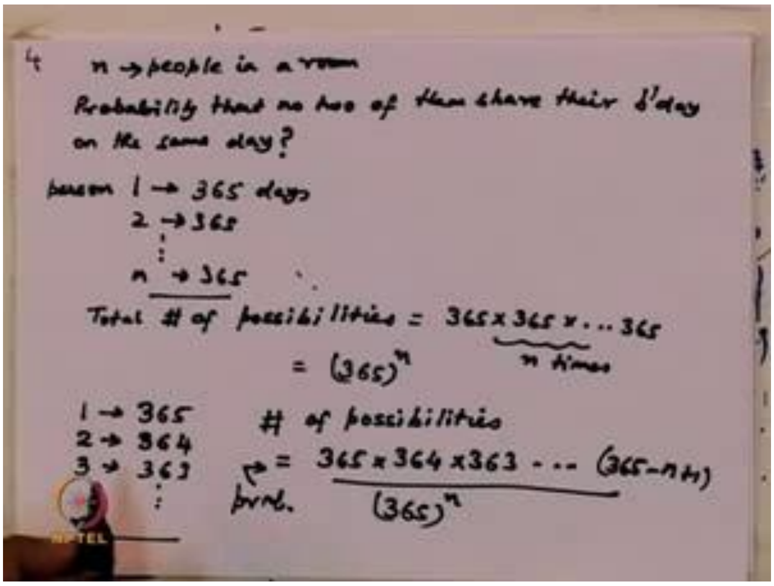


So, let us do a sample example. Imagine in a class you have 5 men and 5 women, from this group. So, in total this gives you a sample size of 10 people. From this group you want to constitute a committee, comprising of 4 people. So, question is, what is the chance or probability that the committee has 3 women. So, in total we have 5 men and 5 women, and you constitute a committee which is consisting of 3 women; that means,. So, this constitute of 4 people. So, if you have 3 women that could mean that you must have one man. How would you do this selection?

So, now as I briefly mentioned in case of a committee or a combination the order does not matter. So, you need to use the expression n C r. So, from a total of 10 people, from a total of temp 10 people you want to draw 4 people, in how many different ways can you draw 10 4 people from a group of 10 people, and the answer to that is 10 C 4. This is the total number of possibilities.

Now, so this is your total sample space. Now, but of these 4 you want to choose 3 women and 1 man. So, there are already 5 women in this group right. So, in how many ways can you choose 3 women and 1 man? So, it amounts to. So, these amounts to from 5 people you can choose 3 women, and for each of these combinations, from 4 men for 5 men you can choose 1 man. So, your probability is nothing but 5 C 3 into 5 C 1 by 10 C 4. You can do the exact calculation by put it plugging in the appropriate values and. So, this is your final answer. So, this is an example where you have made use of combinations, to de derive your final answer.

(Refer Slide Time: 08:51)



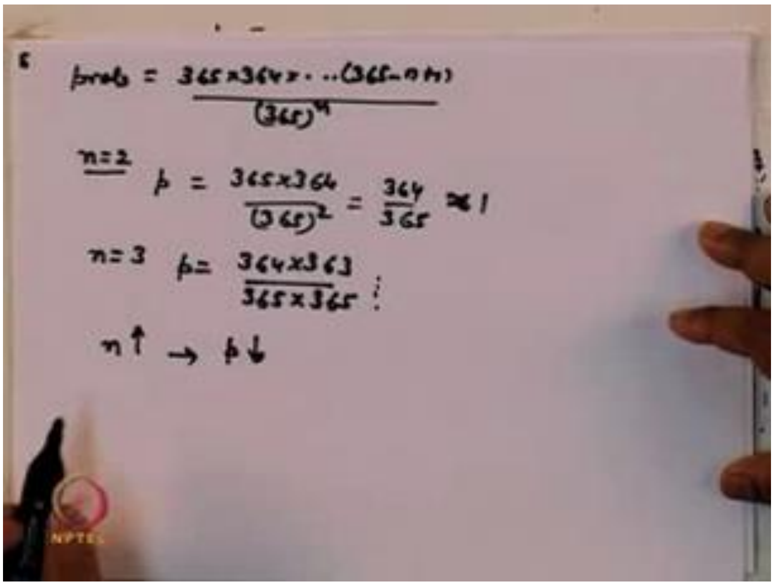
So, let us take another example. So, imagine you have a class in a in a in a room there are n people. So, you have n people in a room. So, what is the chance, what is the probability that no two of them share their birthday on the same day? So, you want to compute the chance of probability, that no two of these n people in the room, share their birthday on the same day. So, in order to address this, first we need to compute the total number of possibilities right. So, how many possibilities are there.

For person 1, for person 1, he or she can be born on any 1 of the 365 days. So, let us assume that the given year is not a leap year. So, otherwise you would have at 366 days. So, person 1 can be born on any one of those 365 days, same for person 2, and same for person n. So, what are the total number of possibilities? So, the total number of possibility is nothing but their product, since the birth of person 1 and person 2 is not constrained; that means that 1 and 2 can be born on any day. So, the total number of possibilities is equal to 365, for each birthday of person 1. There are 365 days possible for person 2 so on and so forth.

So, this thing is repeated n times. So, total number of possibilities is nothing but 365 whole to the power n. So, now, let us come to the question you want to calculate the probability that no two of them share their birthday on the same day. So, which means we can say, that if person 1 is born on any one of the 365 days, we do not know which day. Person 2 should be must be born on any one of the remaining 364 days. Person 3 must be born on any one of the remaining 363 days and like that you go. So, number of possibilities in this case is equal to 365 into 364 into 363.

So, this is for person n. So, your probability then becomes this. So, probability is this divided by 365 whole power n. So, as you can see, if you have only 2 people if they were n equal to 2 people.

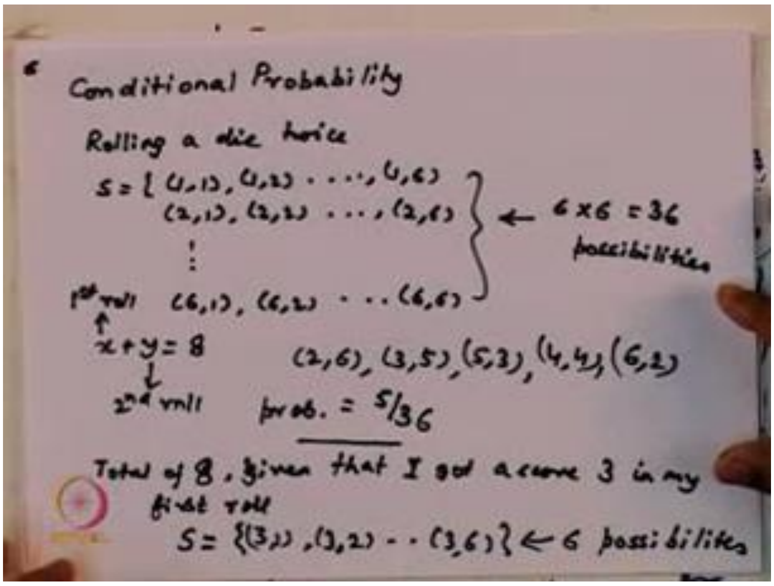
(Refer Slide Time: 12:30)



So, if I think of the probability. So, my probability is given by 365 into 364. So, if your n was 2 let us say, this probability would simply be 365 into 364 by 365 whole square, which is equal to 364 by 365, which is almost close to 1 right. So, because if there is n equal to 2 possibility, only 2 people are involved, then the probability is very high, because for each day except one day the second person can be born in any of the remaining days, but as you see if n equal to 3 this probability becomes 364 into 363 by 365 into 365.

So, with increasing n; so as n increases, this would lead to probability decreasing. So, we can compute what is the chance that of this entire probability being 50 percent, and it turns out to be nearly n equal to 23. So, with this I would like to conclude or stop complete my session on permutation and combination, and introduce a very important concept of probability called conditional probability.

(Refer Slide Time: 14:07)

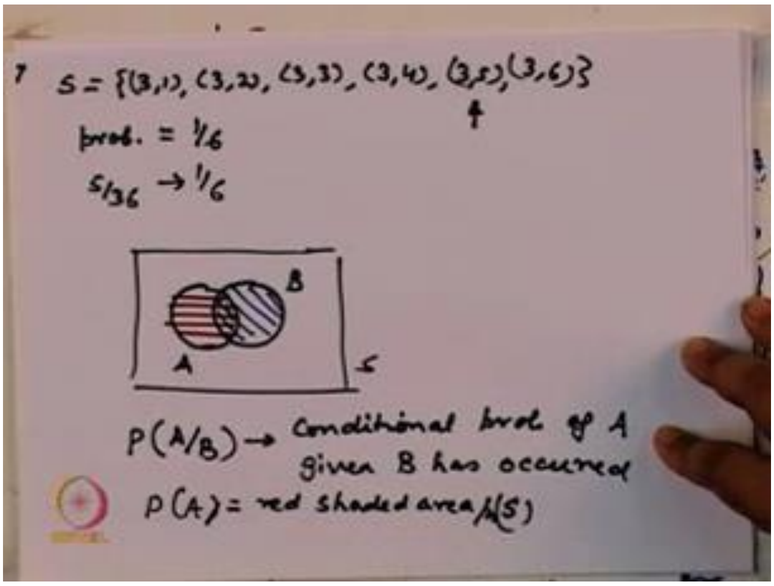


So, let us take a simple case. Imagine you are rolling a die twice. What are my possibilities? So, my sample space is given by all these combinations. So, you have a total of 6 into 6 equal to 36 possibilities. So, now, let us say I want to ask the question, what is the chance what is the probability, that the sum of the two sum of the scores on the 2 die equal is equal to 8. So, I want to ask the question what is the probability that we get x plus y, if x is the score associated with first roll of the die, and this is with the second roll.

What is the probability that x plus y is equal to 8. So, how will I find this out? I will find out all I will look for all those combinations, for which x plus y is equal to 8 and I get my answer as 2 comma 6 3 comma 5 5 comma 3 4 comma 4 and 6 comma 2. So, I can see that there are 5 possibilities of total of 8 so of these 5. So, my probability then becomes simply is equal to 5 by 36, that the sum is equal to a value of 8.

Now, let us say I want to find the same probability of the total score being 8, given that. So, I want to find the total score of 8, total of 8 given that, I got a score of 3 in my first roll. So, what you see is the sample space. So, the sample space now, instead of having thirty 6 different possibilities, you only have 6 possibilities, because I have said my first score is 3, my possibilities is simply 3 1 3 2, and this is 6 possibilities. So, of these if I were to ask. So, my sample space becomes 3 1 3 2 3 3 3 4 3 5 and 3 6.

(Refer Slide Time: 17:31)

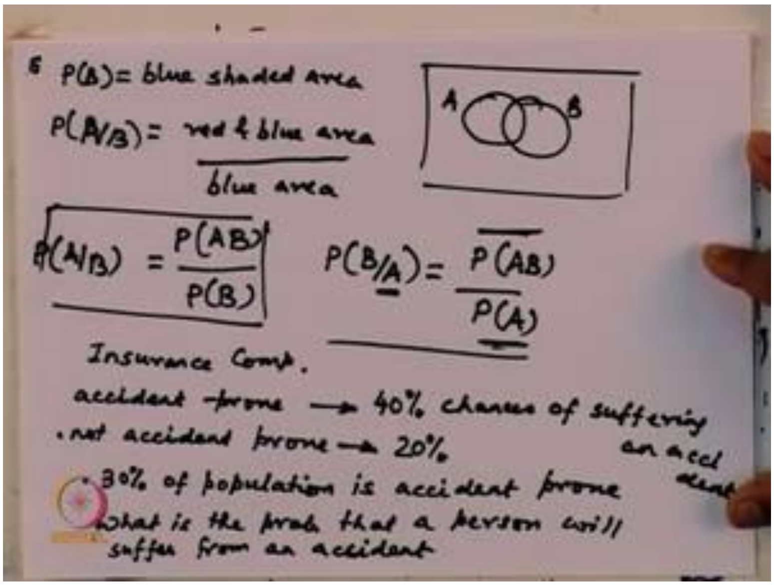


So, for which of them do I get a value of 8 which is nothing but 3 5. So, my now probability becomes, nothing but 1 out of 6 possibilities. So, my chance is have actually increased from the earlier value of 5 by 36, I have gone to 1 by 6.

So, it has changed. So, when you represent these particular phenomena in terms of a Venn diagram. So, imagine I have. So, this is a particular sample space, this is my event A, and this is an event b. So, there is some overlap between A and B, and the conditional probability of A given B has occurred is represented by P of A B. So, probability of A given B is the conditional probability of A, given that B has occurred. So, this is conditional probability of B has. So, if I see here my probability of A. So, my probability of A is, going to be given by this entire sample space, but now that I know that B has occurred.

So, B has occurred means this is, what is my sample space. So, of this only this common area, the common area is all that corresponds to A being occurring; so probability of A given B. So, probability of A is basically the red shaded area by the whole sample space n of S.

(Refer Slide Time: 20:20)



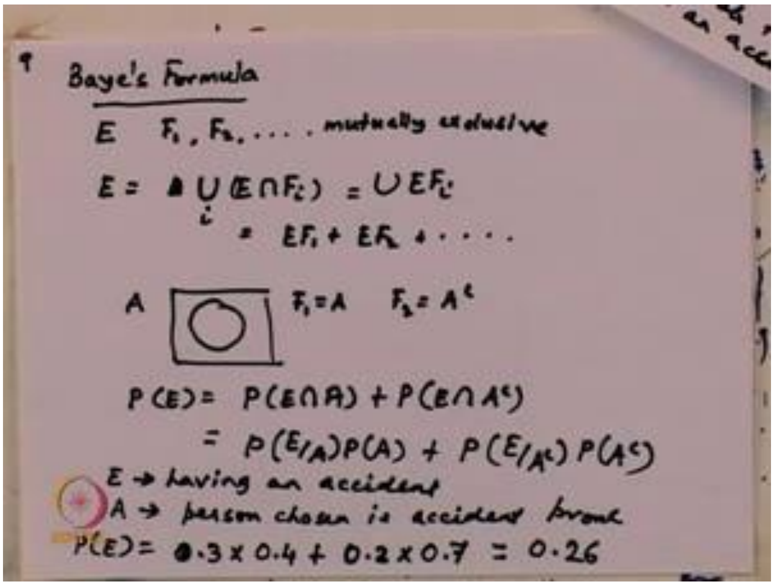
Probability of B; so I can write probability of B is equal to the blue shaded area, but probability of B given A. So, A given B is basically the red and blue area probability of A given B. Means only this common shaded area is where A can occur red and blue area divided by the blue area. So, I can see that the red and blue area, the red and blue area is nothing but (Refer Time: 21:01) is it is probability of A intersection B, this is nothing but probability of A intersection B, and the total B is simply probability of B.

So, this is a very important derivation. So, you see probability of A given B, is probability of A B by probability of B. Similarly. So, if this is my. So, I have A I have B I can also define probability of B given A. So, probability of B given A will similarly the probability of A B by probability of A. So, what you see from this definition is, what is the B given A whatever is you put given A, that probability absolute probability is appears in the denominator, and the probability of both events occurring together appears in the numerator. So, probability of A given B is P A B by P B and probability of B given A is P A B by P A.

So let us take a sample example; imagine you are running an insurance company and from your way of looking at people, you characterize them as accident prone or not accident prone. So, the person you call so you are in insurance company the person any person x. So, if you what you call as a person is labeled as accident prone, or not accident prone. What is your basis for characterizing people through these 2 terms, because the person who is accident prone has a chance, has a forty percent chance that he will actually suffer an accident in a given year. The person who is not accident prone has a chance of twenty percent. So, the person who is not accident prone has a lesser chance than the person who has an accident prone.

So, let us say there are thirty percent of the population is accident prone. So, this is forty percent chances of suffering an accident. So, and the person who is not accident prone has twenty percent chance. So, if you have thirty percent of the population is accident prone, what is the probability that a person will suffer from an accident. So, how do we do this? We begin, in order to be able to do this we need to have the following formalism which is Baye's formula. So, it brings us.

(Refer Slide Time: 24:43)



So, imagine if I define, you have an event E and a series of events F 1 F 2 so on and so forth, which are mutually exclusive. These are mutually exclusive. So, I can write down this formula E is equal to union of E intersection F I over all i which is nothing but E F 1 plus E F 2 plus dot dot dot. So, if I want to, just think of a single event a, if I want to think of a single event a right. So, what is my F 1 and F 2 if my F 1 is equal to A, then my F 2 is nothing but a complement.

So, based on that I can write, I can write my probability of event E is, equal to probability of E intersection A, plus probability of E intersection A complement. So, P E A E or can be written as probability of E given A into probability of A plus probability of given A complement into probability of A complement. So, you can break it down into conditional probabilities, and that is how you can address the earlier problem, that your answer your. Answer is nothing.

So, in this case you have to define what is your event A, and what is your event E, E event E corresponds to having an accident. So, and that is what we have been asked to compute. We want to compute what is the probability that the person will suffer from an accident. What is your event A. Event A is the person. So, that the person chosen is accident prone. So, I can then compute the total probability of E, as A probability of choosing accident prone person which is thirty percent or 0.3, into probability of having an accident if the person is accident prone which is forty percent.

Similar I can have one 2 more terms; probability is not he is not A. So, E given A compliment is 0.2 into A compliment is 0.7. So, this total will come to 0.26. So, you got an idea of how we can treat calculate a conditional probability, and you can calculate total probabilities, based on conditional probability.

In the next class we will again continue with conditional probability, and then we will go about discussing a random variables with that I.

Thank you for your attention.