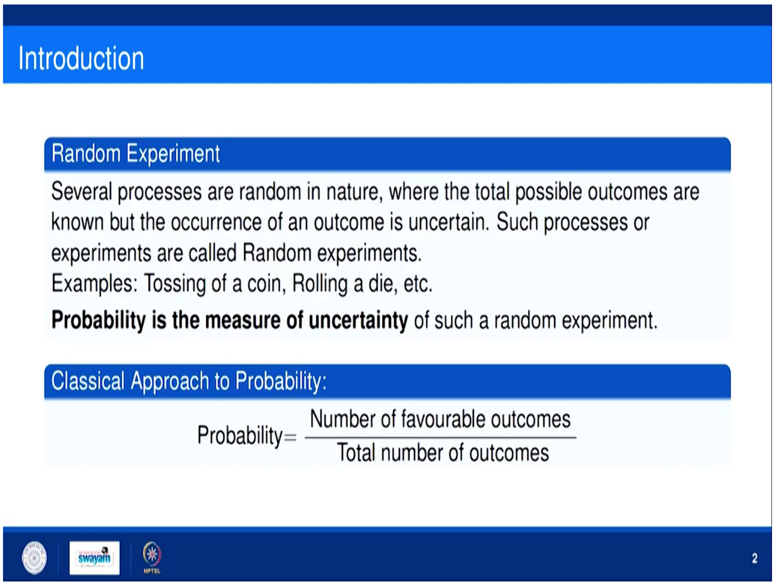
Essential Mathematics for Machine Learning
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture – 31
Review of Probability

Hello friends. Welcome to lecture series on Essential Mathematics for Machine Learning. Now, the few in few lectures we will see that what probability is, and how can how it is important in machine learning. So, first let a have; let us have a recap on Basic Probability.

(Refer Slide Time: 00:41)



So, what is a random experiment first? So, several processes are random in nature, where the total possible outcomes are known, but the occurrence of an outcome is uncertain, ok. Such processes or experiments are called random experiments.

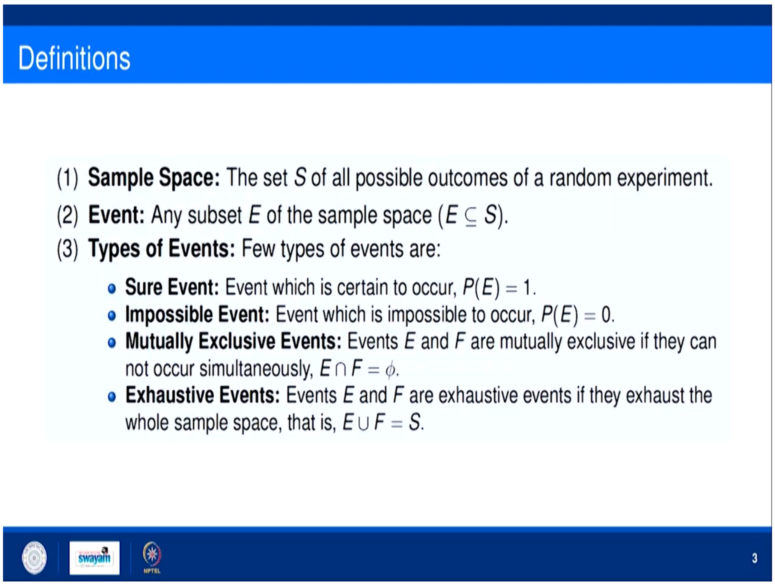
For example, you toss a coin, ok. If you toss a coin then either you will get head or tail, so the things are not certain, things are uncertain. So, that is basically a random experiment, ok. Now, you roll a dice, if you roll a die you will get either 1, 2, 3, 4, 5, or 6. So, we do not know that which number will come, so this is basically uncertain outcome is uncertain. So, it is also a random experiment.

Similarly, if you take a card from a pack of 52 cards, then that card may be spade, maybe club, may be heart or maybe diamond, ok, we do not know, we do not know which card it will be. So, that is basically a random experiment.

Now, for random experiment, how can we measure the uncertainty? So, the probability is basically the measure of uncertainty of such a random experiment. And classically, how can we define a probability? The simplest definition of probability is you simply take number of favourable outcomes in the numerator and divided by total number of outcomes.

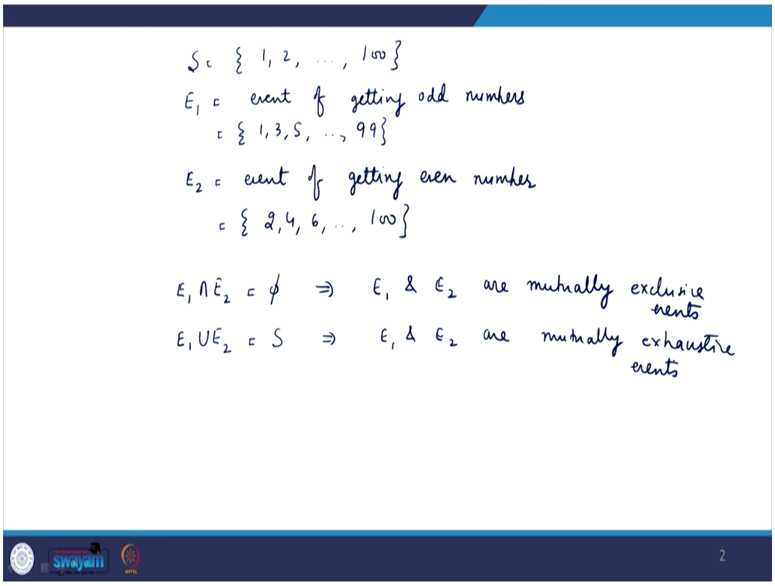
If you perform any random experiment, you take the total number of outcomes and what are the favourable outcomes for which you want to find out the probability that will come in the numerator divided by the total outcomes. That give the probability of that experiment.

(Refer Slide Time: 02:40)



Now, we have few basic definitions in this first of all sample space. What do you mean by sample space? Sample space basically all possible outcomes of an experiment.

(Refer Slide Time: 02:53)



Suppose, you throw a die; so, if you throw a die, so what would be a sample space? Sample space here will be because if you throw a die we will get 1, 2, 3, 4, 5, or 6, so that will constitute a sample space. So, that is 1, 2, 3, 4, 5, or 6; if you So, if you toss a coin there are sample space will be either head or tail, so that would be a sample space of this experiment. In this way we can define sample space.

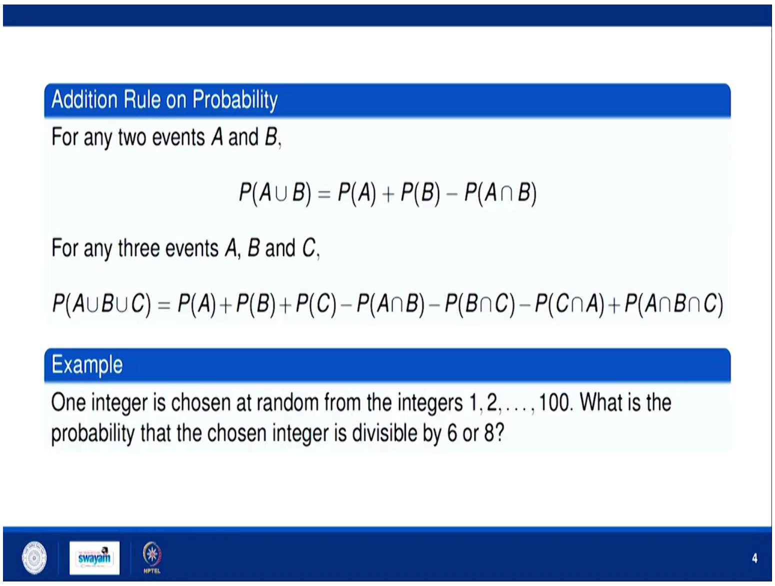
Now, what is an event? Any subset of the sample space is called an event, ok. So, basically denoted by E which a subset of S. Now, there are different types of events. The first one is sure event which will definitely happen. So, for such events probability is always 1, because that is always certain. Impossible events, events which is not possible to occur so, for such events probability is 0, ok.

Now, mutually exclusive events. So, if events E and F are set to be mutually exclusive, if the intersection of them is phi, if they cannot occur simultaneously. So, those events are called mutually exclusive events. Now, exhaustive events, events E and F are called exhaustive events if they exhaust the whole sample space that is union is S, ok.

Suppose, you take this example. Suppose sample space is numbers from 1 to 100, ok. Now, you take E event one which is basically event of getting odd numbers. So, of course, odd numbers will be what? 1, 3, 5 up to 99 from this sample space. Now, suppose E 2 is event of getting even number, so for getting even number it will be 2, 4, 6, 8 and so on up to 100.

Now, if you take these two events then clearly E 1 intersection E 2 is phi, ok. So, we can say that E 1 and E 2 are mutually exclusive events. So, this implies E 1 and E 2 are mutually exclusive events. Moreover, if you take E 1 union E 2, E 1 union E 2, so, this is nothing, but sample space S. So, we can say that E 1 and E 2 are also mutually exhaustive events. So, hence, for this example we can say that E 1 and E 2 are mutually exclusive and exhaustive events, ok.

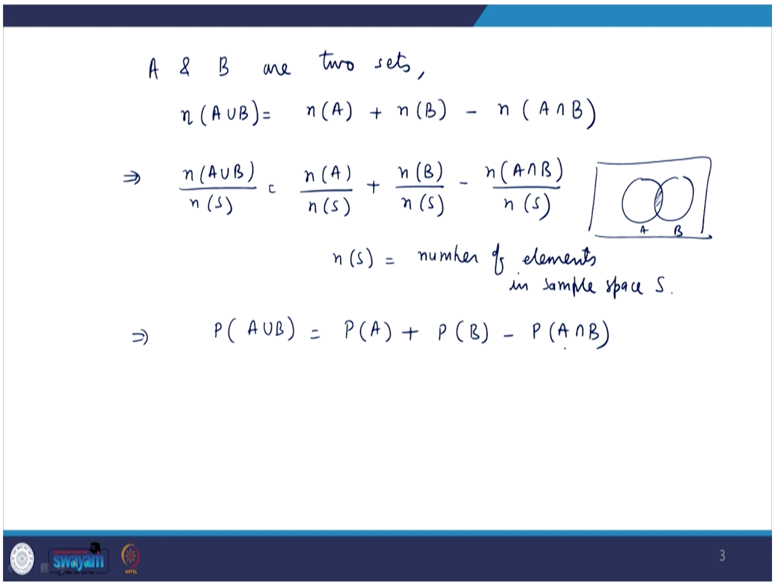
(Refer Slide Time: 06:28)



Now, come to the basic additional rule of probability. So, for any two events A and B we can define probability of A union B as probability of A plus probability of B minus probability of A intersection B. And for 3 sets, for 3 events A, B, and C, the rule the probability the probability of A union B union C will be equal to probability of A plus probability of B plus probability of C minus probability of A intersection B minus B intersection C minus C intersection A and plus A intersection B intersection C.

So, the proof is quite easy. Suppose, you want to prove for two sets; similarly, we can go for 3 sets or higher or higher number of sets, say we want to show for two sets.

(Refer Slide Time: 07:24)



So, we know that if we are having two sets A and B, A and B are two sets then cardinality of A union B is nothing but cardinality of a plus cardinality of B minus cardinality of A intersection B, ok. This is very clear from the Venn diagram also. If you are having two events A and B, two sets A and B then A union B is nothing but number of elements in A, number of elements in B, and since intersection is coming two times, so minus A intersection B.

Now, you divide the entire expression by number of element in the sample space. Say number of element in a sample space is n S, ok. So, you can divide, so this implies n A union B divided by n S will be equal to n A upon n S plus n B upon n S minus n of A intersection B upon n S, where n S is nothing, but number of elements in sample space or cardinality of sample space; number of. So, it is basically a number of elements in sample space, sample space S.

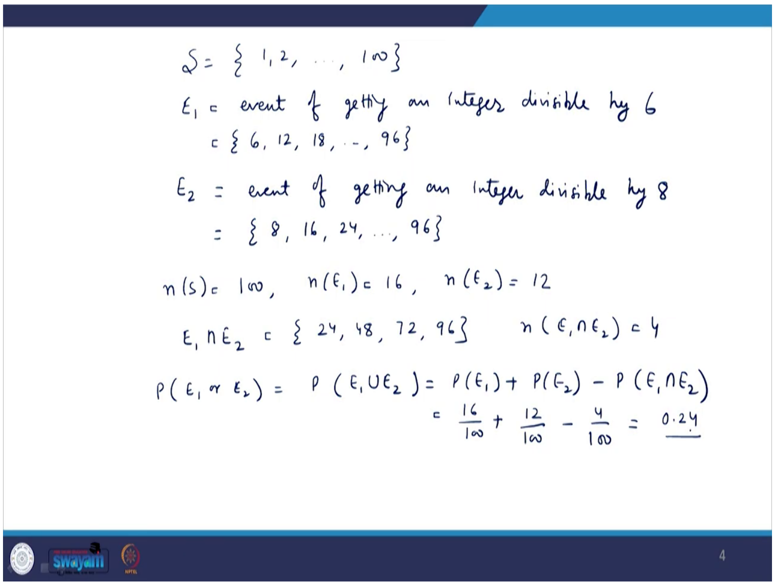
So, what this is? This is basically number of elements in A union B upon number of elements in the sample space. So, that will constitute probability of A union B. And similarly, this is number of elements in A that is possible out, that is favourable outcomes in A upon total outcomes, so that will constitute probability of A, probability of B, similarly, and similarly, this will be probability of A intersection B. So, this implies basically probability of A union B is equal to probability of A, probability of A plus probability of B minus probability of A intersection B. So, this give the additional addition rule of probability.

Now, similarly, if you want to go for proof for 3 set, similarly, by the Venn diagram we can easily prove. Now, let us discuss one problem based on this one integer is chosen at random from the integers 1 to 100. So, what your sample space here having? Sample space is all the integers from 1 to 100. So, what is the cardinality of the sample space? is 100.

Now, what is the probability that the chosen integer is divisible by 6 or 8? So, basically union means or if you are talking about A union B that means, A or B and if you are doing intersection that means, A and B. So, sometimes we can also say this expression as probability of A or B is equal to probability of A plus probability of B minus probability of A and B, ok.

So, now if you want to calculate probability of the chosen integer is divisible by 6 or 8, that means, we are interest to find out the union, or means union, ok. So, how can we calculate?

(Refer Slide Time: 11:00)



So, here sample space is what? Sample space is 1, 2 up to 100. So, let us suppose E 1 is the probability, E 1 is the event of getting an integer divisible by 6, ok. So, what will be the set E 1 then? See an events will be a 6, 12, 18 up to the last will be 96, because 96 is divisible by 6. Now, suppose E 2, E 2 will be event of getting an integer divisible by 8. So, that will be nothing, but 8, 16, 24 up to up to 96. It is a nearest integer nearest to 100 which is divisible by 8.

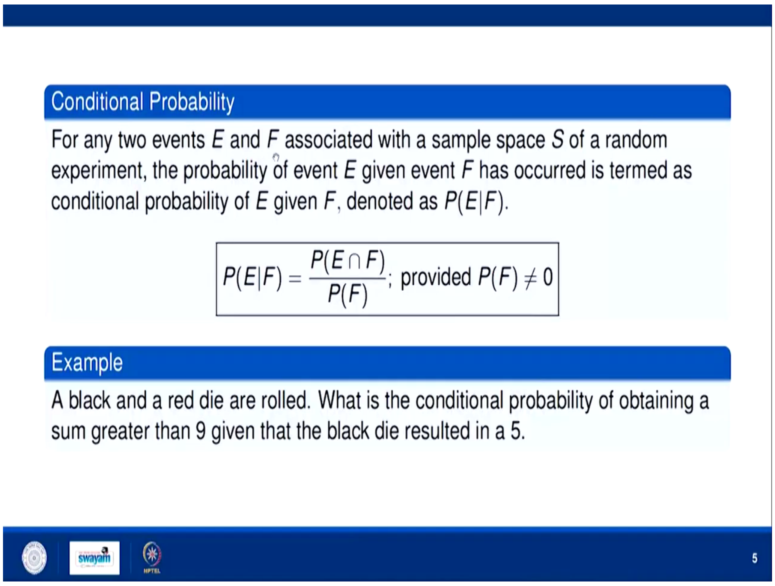
So, what will be the cardinality of S? It will be 100, cardinality of E 1, number of elements in E 1 these are simply 16, cardinality of E 2 that will be that will simply 12. Now, there are there are some elements which are common in both, like you are seeing here 96. Similarly, there are others also. So, what are they? They are simply the LCM of 6 and 8; 24; 48 like this.

So, if you calculate E 1 intersection E 2, E 1 and E 2, so that will be nothing, but 24, 48, that will be 72 and 96, these 4 only. So, this the cardinality of E 1 intersection E 2 will be 4. So, you want to calculate probability of E 1 or E 2, that means, probability of E 1 union E 2.

And that is nothing, but by the addition rule that is nothing, but P of E 1 plus P of E 2 minus probability of E 1 intersection E 2 and that will be because cardinality of E 1 is 16. And sample space is 100, the so probability will be 16 upon 100 plus here it is 12 upon 100 and minus 4 upon 100. So, that will be simply 28 minus 424 that will be 0.24. So, that will be the probability of E 1 or E 2. So, in this way we can use addition rule on probability.

Now, come to conditional probability. So, what do you mean by conditional probability?

(Refer Slide Time: 14:05)

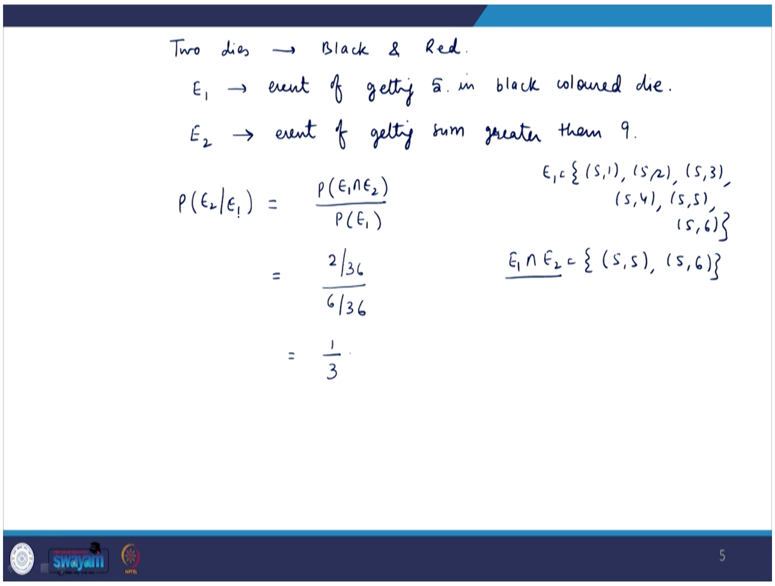


Now, for any two events E and F associated with a sample space S of a random experiment the probability of E given event F has occurred is termed as conditional probability of E given F.

So, we are finding, we are finding the probability of E when F has already happened. So, that is basically conditional probability, and that is mathematically given by this expression, that is nothing but probability of E intersection F divided by probability of F. Of course, probability of F should not equal to 0. So, that is how we can we can calculate the conditional probability of E given F. So, based on this let us discuss one example.

Now, a black and a red die are rolled. So, there are two die one is black coloured other is red coloured and they are rolled. What is a conditional probability of obtaining a sum greater than 9 given that the black die resulted in a 5? Ok. So, let us discuss this thing.

(Refer Slide Time: 15:16)



So, here we are having two dies, a black coloured and a red coloured, ok. Now, the question is obtaining a sum greater than 9 given that black die resulted is in a 5. So, black die resulted. So, let us suppose E 1 is your E, E 1 is the event of getting 5, ok, getting black die resulted in 5, getting 5 in black coloured die.

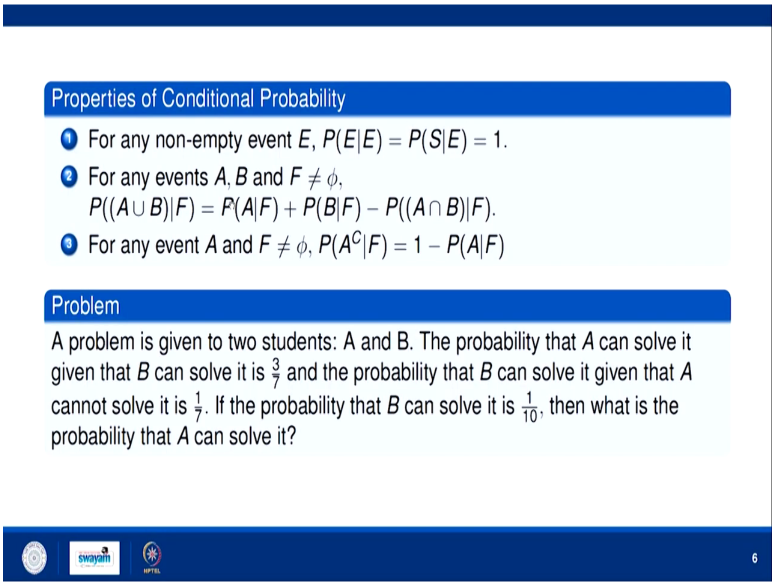
So, (Refer Time: 16:02) suppose E 2 is nothing but is want to find out sum greater than 9. So, event of getting sum greater than 9, ok. Now I want to calculate P of E 2 given E 1, ok; that means, probability of getting sum more than 9 when the when we are having 5 in a black coloured die. So, that is nothing, but probability of E 1 intersection E 2 upon probability of E 1 as per the formula of conditional probability.

Now, what will be E 1? Ok. Let us discuss E 1 first. E 1 will be, now you are throwing two dies and it is given that in a black coloured die it is 5 only. So, you are having 5 1, 5 2, 5 3, 5

4, 5 5, and 5 6. These are the various cases you are having for E 1. And for E 2, E 2 is nothing, but sum greater than 9, sum greater than 9 means only 5 5 and 5 6. So, there are two cases 5 5 and 5 6. This is E 1 intersection E 2 basically, ok.

So, probability of E 1 intersection E 2 will be only these two cases. So, it is 2 upon favourable cases are two total cases are 36, ok. And here for E 1 the favourable cases are 6; 1, 2, 3, 4, 5, 6 and total cases are 36 because you are throwing two dies, so total cases are 36, so which is 6 upon 36. So, total are so required probability is 1 upon 3, ok. So, in this way we can obtain conditional probability of getting sum more than 9 when you are having 5 in a black coloured die.

(Refer Slide Time: 18:14)



Now, what are the properties of conditional probability? So, the first properties which is very obvious property for any non-empty event E probability of E upon E. Now, now if you see the definition it is nothing, but E intersection E upon probability of E and E intersection E is itself. So, you can cancel. So, this will be nothing, but 1, ok.

And if you take; if you take probability of S, S when E has already happened; so, this is nothing, but S intersection E upon probability of E, S is the sample space, so S intersection E will be nothing, but E and probability of E upon probability of E is nothing but 1. So, that is quite easy result quite obvious result which we can obtain from the definition itself.

Now, for any two events A and B and F not equal to phi, not an empty set we are having another result which is probability of A union B, condition probability of this is equal to this. So, let us try to prove this result.

(Refer Slide Time: 19:24)

$$P((A \cup B)/F) = P(A/F) + P(B/F) - P((A \cap B)/F)$$

$$\underline{LHS} \quad P((A \cup B)/F) = \frac{P((A \cup B) \cap F)}{P(F)} = \frac{P((A \cap F) \cup (B \cap F))}{P(F)}$$

$$n\big((A \cap F) \cup (B \cap F)\big) = n(A \cap F) + n(B \cap F) - n\big((A \cap F) \cap (B \cap F)\big)$$
$$= n(A \cap F) + n(B \cap F) - n((A \cap B) \cap F)$$

$$\Rightarrow \frac{P((A \cap F) \cup (B \cap F))}{P(F)} = \frac{P(A \cap F)}{P(F)} + \frac{P(B \cap F)}{P(F)} - \frac{P((A \cap B) \cap F)}{P(F)}$$

$$\Rightarrow P((A \cup B)/F) = P(A/F) + P(B/F) - P((A \cap B)/F)$$

So, we have to show that probability of A union B upon F I mean conditional probability is equal to probability of A plus probability of conditional probability of B given F minus conditional probability of A intersection B given F.

So, let us try to prove this result. So, let us take the right left hand side. What is the left hand side? Probability of A union B given F, and by definition it is by definition it is probability of A union B intersection with F upon probability of F, ok. This is nothing, but A union F intersection with B union F upon P F sorry, it is A intersection F. So, so this is a distributive law basically. So, A intersection will come here and union will come here, ok.

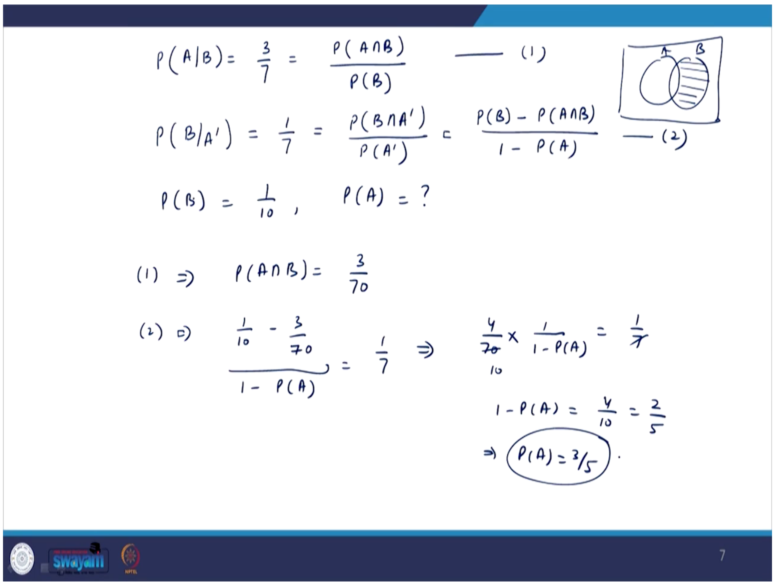
Now, if you if you take this set A intersection F union B intersection F, so let us suppose this is a set A, this is set B, then A union B can be defined as A the cardinality of this. The cardinality of this may be defined as cardinality of this plus cardinality of this minus cardinality of the intersection and this is cardinality of this plus cardinality of this minus cardinality of A intersection B intersection with F, ok.

Now, if you define the if you divide the entire expression by cardinality of S, so that we will get probability of this is A intersection F union with B intersection F and that will be equal to probability of A intersection F plus probability of B intersection F minus probability of A intersection B intersection with F, ok.

Now, you divide the entire expression by P F, probability of F. So, if you divide the entire expression by a probability of F now, so here you will get the left hand side which is which is probability of A union B given F and here is probability of A intersection F. It is basically what? It is a given F because what is probability of A given F? A intersection F upon probability of F. So, this is probability of this is probability of A given F plus probability of B given F minus probability of A intersection B given F. So, that is how we can obtain the proof of this result.

So, similarly, we can obtain this result also that for any event A and F is not equal to phi probability of A compliment, A complement means not A. So, not A given F is nothing but 1 minus P of A given F, ok. Now, let us come to this problem based on this. A problem is given to two students A and B, ok. The probability that a can solve it given B can solve is 3 3 by 7. So, the given thing is, so probability that A can solve given B is 3 upon 7 which is nothing, but P A intersection B upon P B. So, this is the first given to you.

(Refer Slide Time: 23:38)



And the probability that B can solve it given that A cannot solve it is 1 upon 7. So, that means, it is probability of B upon a compliment which is 1 by 7, so which is equal to probability of B intersection A complement upon P A complement which is equal to.

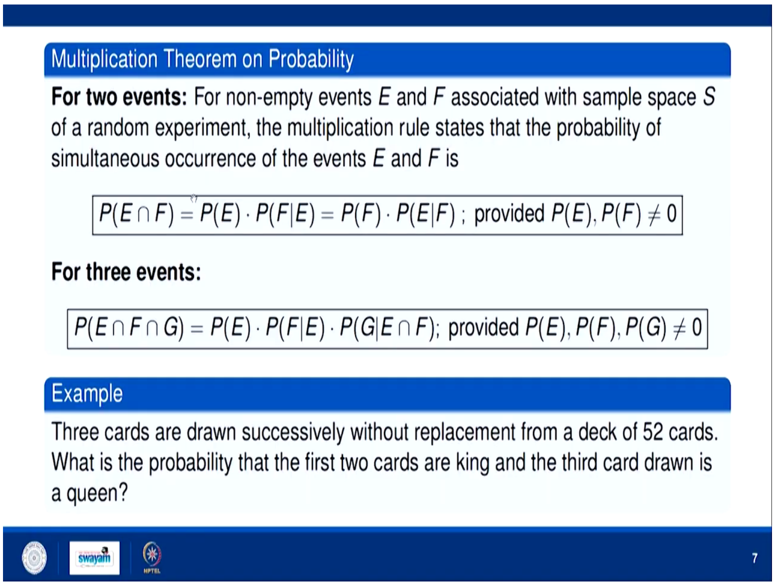
Now, you can see the Venn diagram. If you can see the Venn diagram then it is A, it is B. So, B intersection A complement, if you see the A compliment it is the region outside A and if you see with intersectional B, so only this region. So, that is nothing, but probability of B minus probability of A intersection B. And of course, probability of A complement is 1 minus probability of A. So, this and this is, so this is a second value, second expression given to you.

Now, if the probability that B can solve it is 1 upon 10, then what is the probability that B can solve, A can solve it? So, we have to find P probability of A, if probability of B is 1 upon 10. So, probability of B is 1 upon by 10 and we have to find probability of A.

So, from 1 what we obtain, from 1? So, you can substitute P B as 1 upon 10, so that will be nothing, but, so probability of A intersection B comes out to be 3 upon 70. Now, you can substitute 3 upon 70 over here and P B here, so you can compute P A. So, let us compute. So, 2 implies that is 1 by 10 minus 3 upon 10, 3 upon 70 upon 1 minus P A and that is equal to 1 upon 7, so this implies it is 4 upon 70 into 1 minus 1 minus P A which is equals to 1 upon 7.

So, basically this cancels with 10 times, so 1 minus P A will be 4 upon 10 that is 2 upon 5. So, this implies P A will be 3 upon 5. So, in this way we can compute this simple problem where we are having conditional probability of A with B, B given A compliment and P B, ok.

(Refer Slide Time: 26:31)



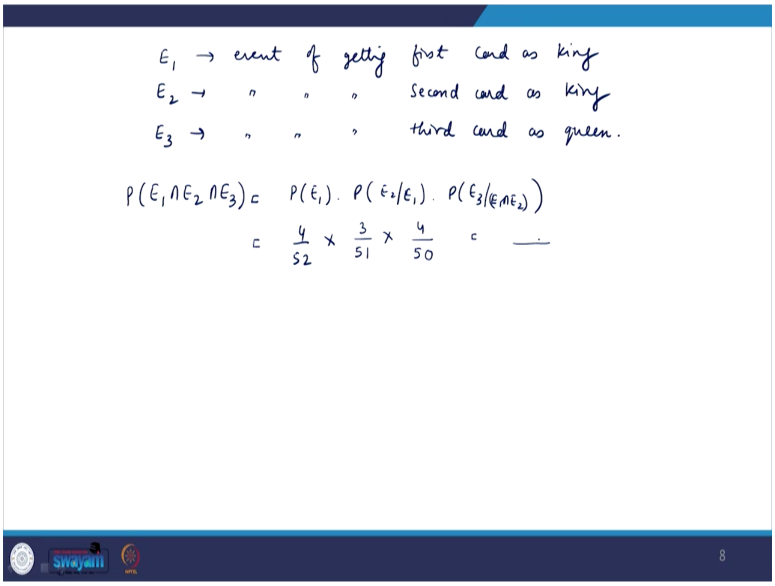
Now, we have multiplication theorem on probability which is states that, for non-empty events E and F associated with sample space S of a random experiment, the multiplication rule states that the probability of simultaneous occurrence of the events E and F is.

So, the simultaneous occurrence of E and F that is E intersection F, so that can be defined as probability of E multiplied by probability of F given E. So, very very easy expression, easy to understand because you see that, if what this mathematically what the expression is.

This is probability of F intersection E divided by probability of E which may which we can be cancels out and we will get back to probability of E intersection F. Or, we can further we can also write this expression as probability of F multiplied why probability of E given F. The same way we can also obtain this result.

Now, if for a 3 events E intersection F intersection G, the same expression can be written as probability of E into probability of F given E into probability of G given E and F. So, when you open this bracket and apply the definition of conditional probability we will get back to this result. So, to understand this let us discuss this example here. Three cards are drawn successively without replacement from a deck of 52 cards, ok. Then, what is the probability that the first two cards are king and a third card drawn is a queen?

(Refer Slide Time: 28:20)



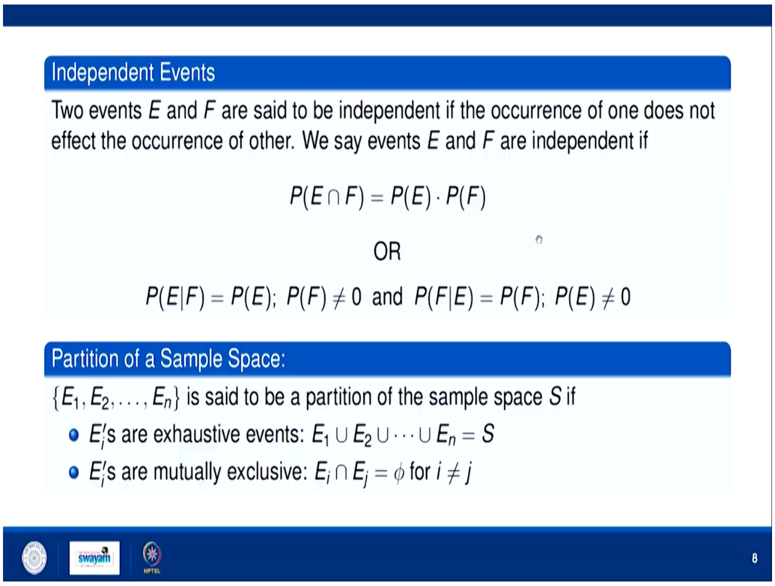
So, let us suppose E 1 is the probability E 1 is the event of getting first card as we want first card as king. So, first card as king. Suppose, E 2 is the event of getting second card also as king and E 3 is the event of getting this third card as, so third case here is basically queen. So, we are interested to find out P of E 1 intersection E 2 intersection E 3.

So, that by this multiplication theorem on probability we can write it as probability of E 1 into probability of E 2 when E 1 has already happened and probability of E 3 when E 1, E 2 both happened. So, what is probability of E 1? E 1 means the first card as king. So, first how many kings are there in the bunch of in a pack of 52 cards? 4. So, it is 4 upon 52.

Now, probability of E 2 given E 1; that means, the first card is already drawn as king and result is without replacement, so that means, if this has already been king what is the probability of getting second card as king. So, now, cards number of kings left are 3 and number of cards are 51. So, these are 3 upon 51.

Now, what is the probability that a third card is queen given that given that E 1 and E 2 both are king? So, now, we are having 4 queens of course, a number of cards are 50 now, because two cards are already been drawn. So, whatever the product will be that will be the required answer of this given problem, ok.

(Refer Slide Time: 30:35)

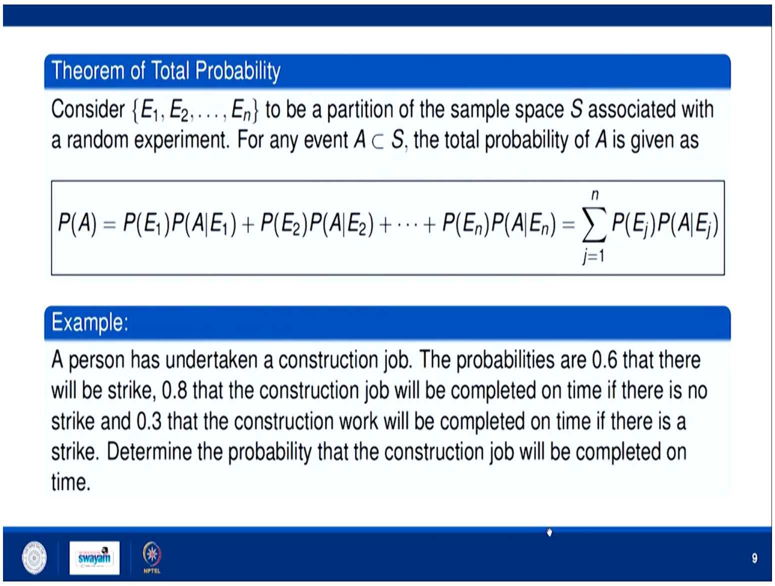


Now, independent events; so, two events E and F are said to be independent if the occurrence of one does not affect the occurrence of other, ok. And mathematically, the independent events are simply if you are finding E intersection F and that is nothing but P E into P F, ok.

And of course, if you are finding conditional probability of E given F, so here in this case because events are independent, so that we simply probability of E only and say. Similarly, if you are finding F given E probability of F given E that will be nothing, but probability of F only.

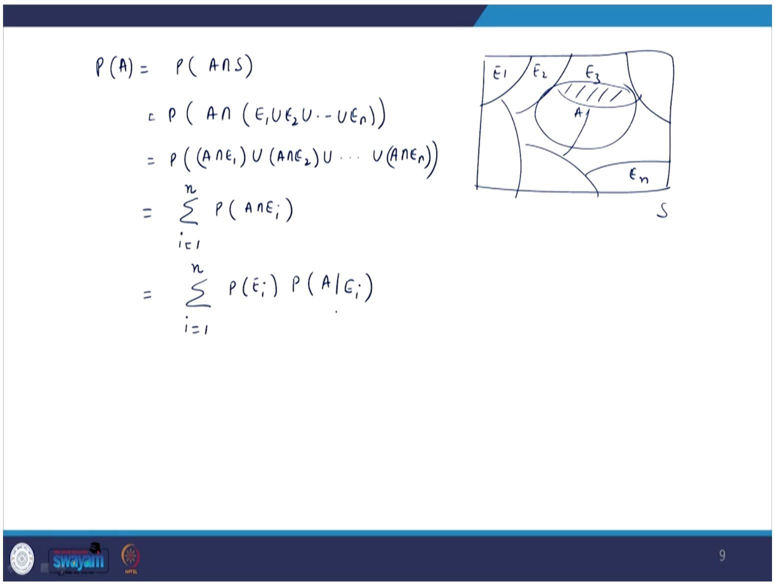
Now, the partition, the partition is simply; suppose you are having a sample space S and you are having E 1, E 2 up to E n. So, these are called partition of a sample space S if the events are exhaustive; that means, the union will constitute the entire sample space S and they are also mutually exclusive, ok, two events are mutually exclusive, any two. So, this is called partition.

(Refer Slide Time: 31:44)



Now, theorem on total theorem of total probability; what is its state let us see. Now, consider E 1, E 2 up to E n be a partition of the sample space S associated with a random experiment. Now, for any event A subset of S the total probability of A is given as given by this expression. So, let us let us discuss the proof of this so.

(Refer Slide Time: 32:12)



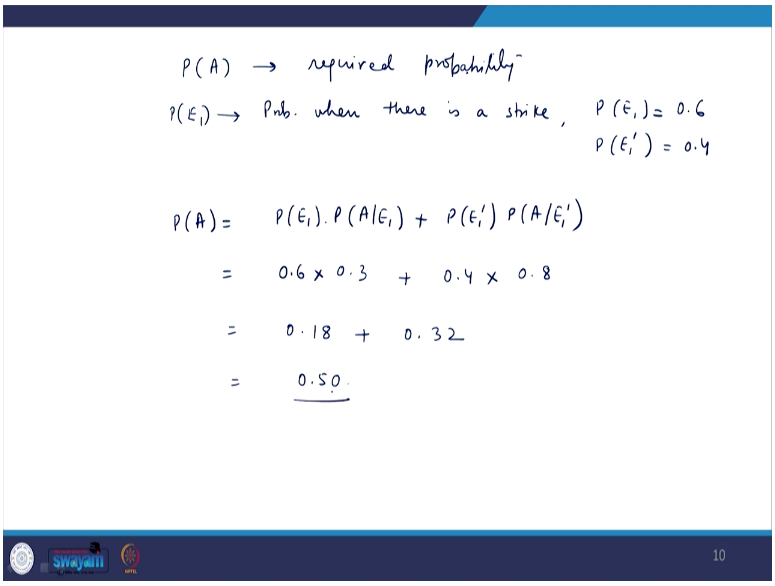
Now, you are having a sample space S, ok. And you are having various partition of this, various events which are partition of this. Let us suppose it is E 1, it is E 2, it is E 3 and so on, suppose it is E n. So, these are partition. Why are we are calling as partition? Because the union of these events we will constitute S and they are mutually they are mutually disjoint also, any two, if you take any two events they are disjoint intersection is 5. You take a set A here, any set A, any event A.

So, what would be A? Probability of A you want to calculate. Probability of A will be nothing, but probability of A intersection with S. Now, probability of A intersection with S. S is what? S is u E 1 union E 2 union and so on up to E n. So, it can be written as probability of A intersection with E 1 union A intersection with E 2 union and so on union with A intersection with E n because A will come somewhat here, A will come here also. So, there are different events from which we have to taken the intersection of A, so that will constitute P A, probability of A.

Now, since intersection of because since e E i's are mutually exclusive, I mean intersection of any two E i's is phi. So, we can write this as simply summation P of A intersection E i, i from 1 to n. And by the by the multiplication rule this can be written as summation i from 1 to n, it is P of E i multiplied by P A given E i. So, that is basically total probability, ok. So, that is basically total probability.

Now, we here is one example based on this, one problem based on this. Let us quickly discuss this. A person has undertaken a construction job. The probabilities are 0.6 that there will be strike, 0.8 that the construction job will be completed on time if there is no strike and 0.3 that the construction work will be completed on time if there is a strike. Determine the probability construction job will be completed on time.

(Refer Slide Time: 34:45)



P A is the required probability, required probability. That means, that means the required work; that means, a construction work will be complete it on time. So, let us suppose P of E 1, probability of probability when there is a strike. So, clearly P of E 1 will be, what is given to us? It is 0.6.

So, of course, P of E complement, that means, when there is no strike that will be 0.4. So, I want to compute P A and by the total probability it is will be P of E 1, we multiplied by P A given E 1 plus P of E 1 complement multiplied by A, when E 1 given E 1 complement, ok.

Now, P of E 1 is what? is 0.6 into. Now, probability of probability of the construction job is completed given the given a strike. See, A is a event of construction job should be completed, given when there is a strike, so when there is a strike when there is a strike the probability of construction job will be completed as 0.8, ok.

So, sorry it will be construction job will be completed on time if there is a strike is 0.3. So, it is 0.3 here, ok, plus it is 0.4 into and here it is 0.8. As given here the construction job will be completed on time if there is no strike, ok. So, this is 0.18, this is 0.32, so the resultant is 0.50. So, that will be the required answer. So, in this way the required answer will be the construction job, the probability that the construction job will be completed on time will be half, ok.

So, in this in this way in this lecture, we have seen that what is a what are the basic concepts of probability. We have seen that how we can define probability, how what is the conditional probability, total probability, multiplication rule, etcetera. In the next lecture, we will see some more concepts on probability.

Thank you.