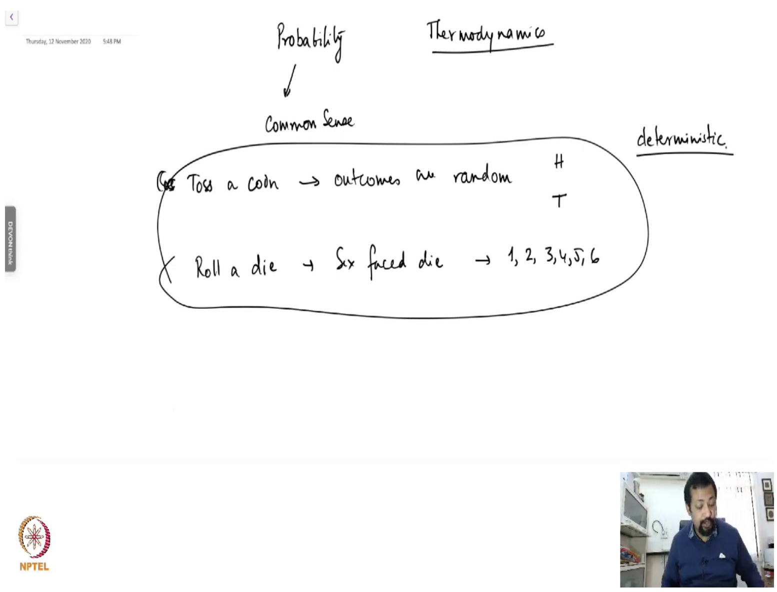
Statistical Mechanics
Prof. Dipanjan Chakraborty
Department of Physical Sciences
Indian Institute of Science Education and Research, Mohali

Lecture – 17
Introduction to Probability

(Refer Slide Time: 00:18)

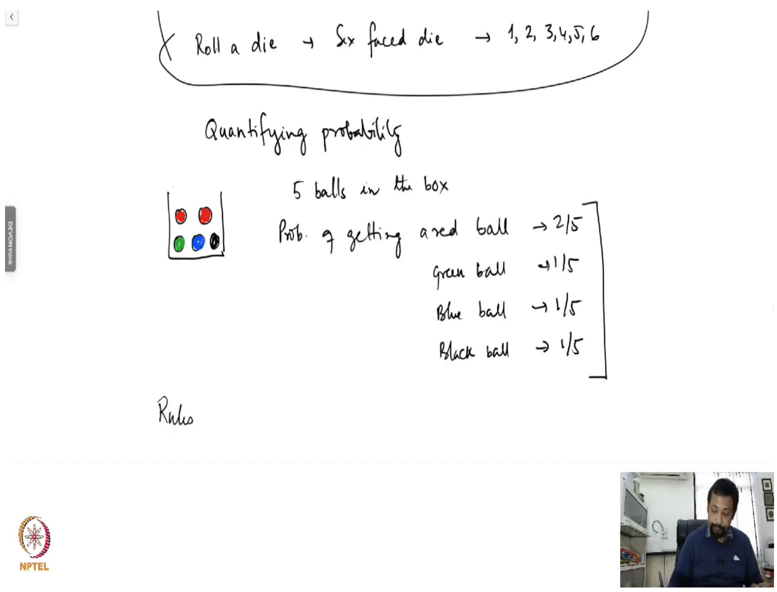


Welcome back. Today, we are going to look at Probability we have already looked at thermodynamics, we have looked at thermodynamics and now we want to bridge the gap with the Statistical Mechanics. So, now, probability is nothing, but common sense, we encountered this in plenty of occasions in our daily life.

For example, if you cross sorry if you toss a coin, then you know that the outcomes are random right. So, there is a certain probability of getting a head or getting a tail. Similarly, if you roll a die let us say it is a 6 face die then the number that comes up is also random so the possible outcomes are 1, 2, 3, 4, 5 and 6.

So, all of these tells you that nothing in nature is deterministic. So, this is completely opposite of what you have encountered in Newtonian mechanics where, given a particle given a single particle given the forces you know that the trajectory is completely deterministic and therefore, you can determine the trajectory at later times exactly. But here in these events that you see that the outcomes are random right.

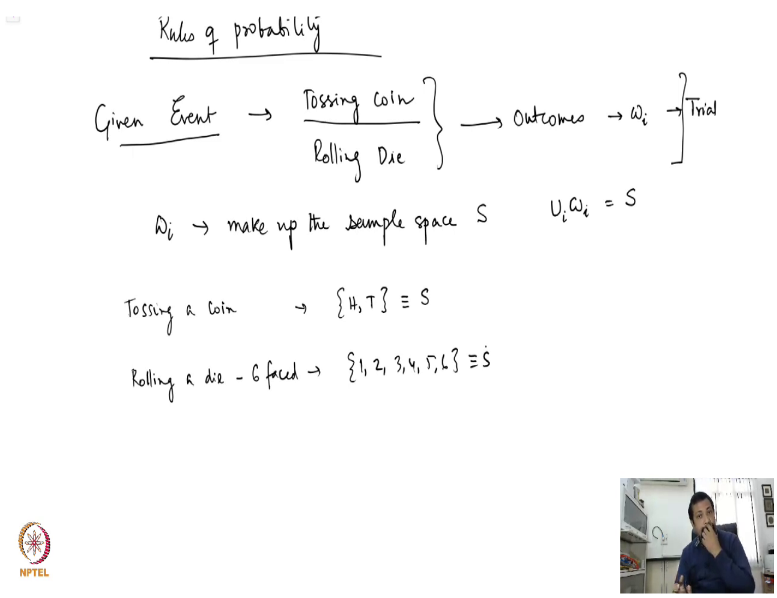
(Refer Slide Time: 01:46)



So, therefore, one has to figure out a way of quantifying probability right. So, let us say for example, if I have a box which has different balls colored balls in it right, some of them can be red. So, let us say 2 of them are red 1 of them is blue 1 of them is green and 1 of them is black.

So, there are 5 balls in the box and you pick up 1 ball at a time. Then the probability that you pick up you will get a red ball is 2 by 5 probability, probability of getting a red ball is 2 out of 5. Similarly, probability of getting a green ball is 1 of 5 and probability of getting a blue ball is also 1 of 5 and probability of getting a black ball is again 1 of 5 right. This you are familiar with you have already encountered this you would have been many times right.

(Refer Slide Time: 03:20)

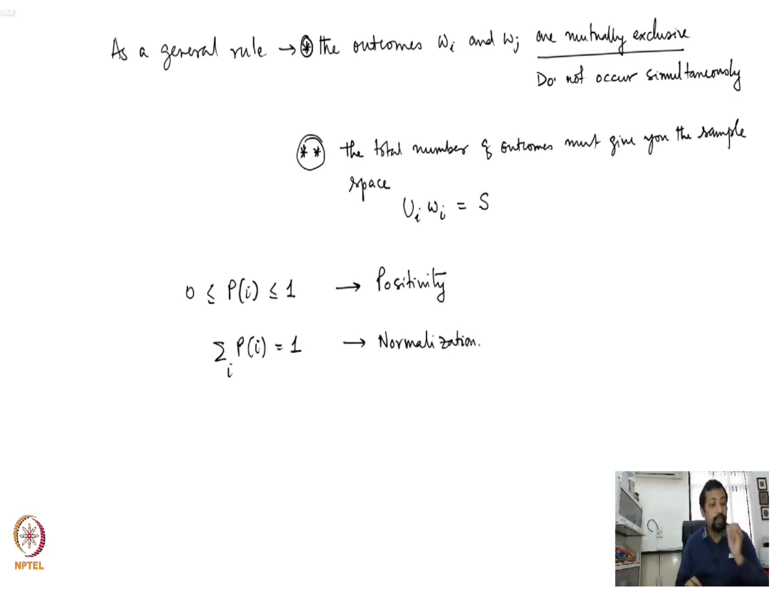


So, now we want to quantify this probability. So, what is called rules of probability? Let us drop it down. So, you see that for a event for a given event which could be like tossing a coin, rolling a die their outcomes, we will call these outcomes as omega i right. So, these are the elements that we normally talk about.

So, these outcomes make up what is called the sample space of the system. So, these outcomes omega i make up the sample space S so that we can write down omega i is equal to S. This event where you have tossed a coin is typically what is called a trial.

So, for tossing a coin I know the outcomes are either head or tail and therefore, this is the sample space for this event. For rolling a die let us be a little bit more specific which is a 6 faced die then the sample space are 1, 2, 3, 4, 5 and 6 and this is a sample space. So, all the outcomes taken together they form the sample space the only request. So, this choice of the sample space clearly depends on the event on the type of event that you are looking at and the type of probabilistic question you are asking right.

(Refer Slide Time: 05:38)

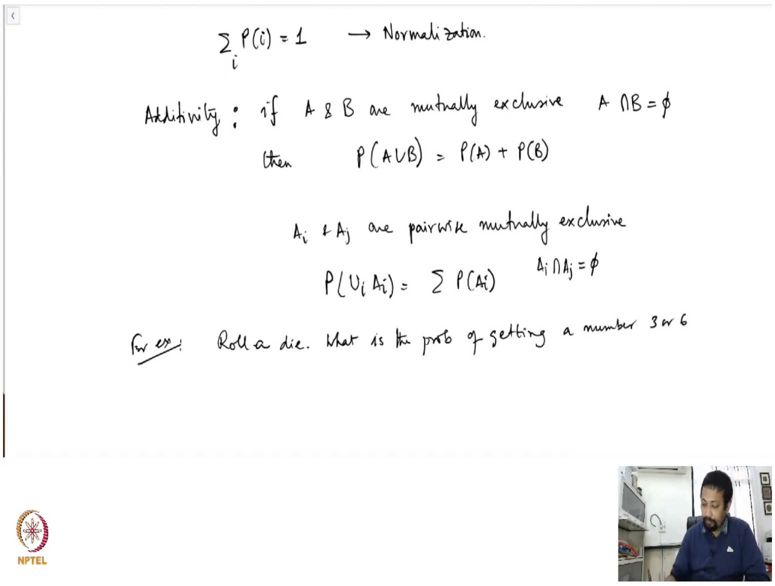


But as a general rule, as a general rule the outcomes omega i and omega j are mutually exclusive. What does this mean? This essentially means that they do not occur simultaneously right. This is my first, the second one that we want to also look at that the total number of outcomes must give you the sample space.

If you are missing something then you know that this is not the sample space which essentially means that the elements the set of elements should be complete. Which we wrote down in a earlier statement that omega i must give you the sample space.

So, given these two; one can write down that the probability you can assign this probability to this outcomes by saying that P of i must be greater than equal to 0 less than equal to 1. This is nothing, but common sense and this is what is called the positivity rule. The second quantity, the second rule that one should also define is sum of P of i must be 1.

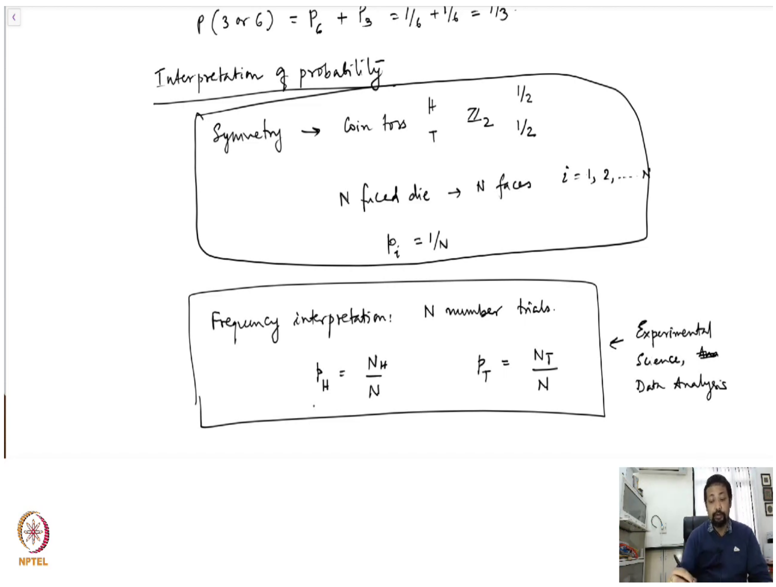
(Refer Slide Time: 07:50)



So, probability of all the outcomes must be unity and this is what is called normalization. Additivity if A and B are mutually exclusive, which means that A intersection B is equal to phi then, probability of A union B is P of A plus P of B. One can generalize this statement also by saying that if A i and A j two of these outcomes are pair wise mutually exclusive then P of this is sum over P of Ai.

For example, so if A i and A j are pair wise mutually exclusive which means that if A i intersection A j is the null set phi then the additivity theorem can be generalized in this. For example, if I roll a die and then I ask the question that, what is the probability of getting a face or a number let us say 3 or 6?

(Refer Slide Time: 09:41)



That the probability of this is essentially P of 1 6 sorry, of getting a value 6 which we shall denote P 6 and a probability of getting a value P 3 which is 1 by 6 plus, 1 by 6. So, we will write down this as P of 3 or 6 is equal to 1 by 6 which is one-third right.

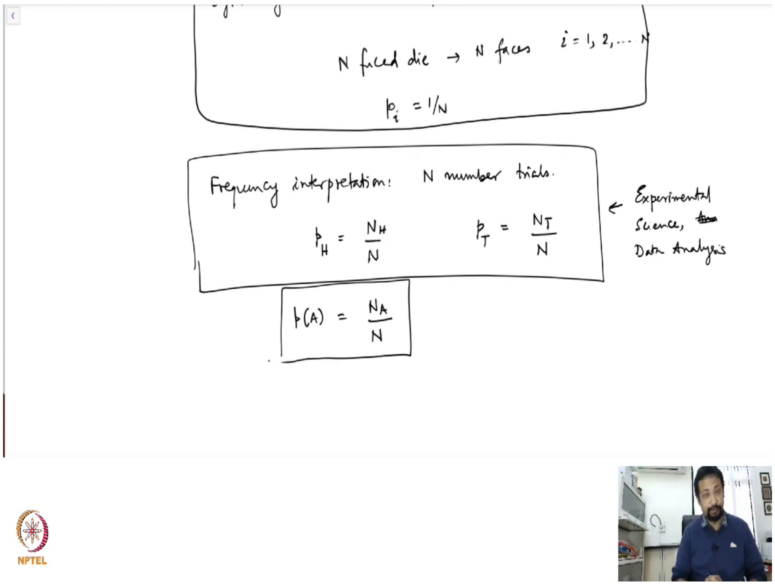
So, now, what we come to is essentially the interpretation of probability. So, whatever we have discussed is essentially we have quantified the rules of probability. So, now we want to look at interpretation of probability. And there are mainly two interpretations that is mostly encountered, one of them is based on symmetry and the example for this is a coin toss in which the possible outcomes are head and tail.

And therefore, here it is a 2 state system in the sense it has a 2 symmetry and each of this has a probability half and half. A similar argument is let us say an N faced die, in an N faced die there are N faces and if it s a fair die then all of these faces are equally likely to appear. So, the outcomes are i equal to 1, 2 all the way up to N. And this is also your sample space and therefore, probability p i of getting a face with number i is equal to 1 by N. So, this is clearly based on symmetry.

The other interpretation which is also often used is the frequency interpretation. Now, in the frequency interpretation what you have is essentially you have N number of trials. So, if you are tossing a coin you toss it N number of times then probability of getting a head is the number of times you get a head divided by the total number of trials that you have done.

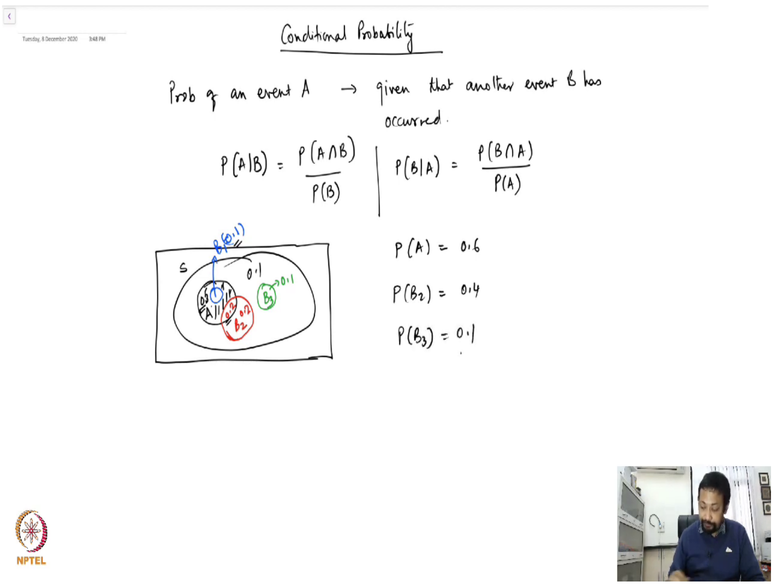
Similarly, probability of getting a tail is the number of tails divided by the total number of trials. So, this is typically what is called a frequency interpretation and this is what is called a symmetry interpretation. Most often in experimental science as well as in analysis in data analysis, this is the interpretation that you see the frequency interpretation is the interpretation that is most often.

(Refer Slide Time: 12:44)



So, for any in a in the frequency interpretation for any outcome A, the number of times that outcome appears divided by the total number of trials is the probability of this.

(Refer Slide Time: 12:57)



So, now we want to look at what is called the conditional probability. And the conditional probability as the name suggest is that you are asking a different question now conditional so, the question that you are asking now that, what is the probability of an event A given that another event B has occurred?

And that we denote as P A B, is essentially given by P A intersection B divided by P B right. If you want to ask a different question that is what is the other way around probability of P given that A has occurred the answer to that would be P B intersection A divided by P of A. Now, both this very simple expressions have a very nice interpretation.

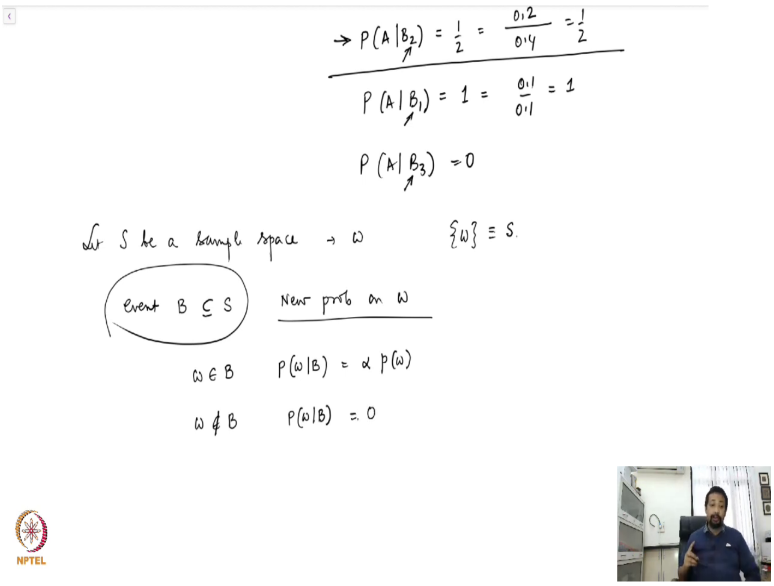
What you are doing, since it is being conditioned on an event A you are essentially restricting the whole sample space to the event of B and this becomes your new sample space. To illustrate this let us look at an Euler diagram, and let us say this is my sample space S, let us denote the events by different colored circles.

There is a green and then there is a red it is not drawn to scale right. So, let us now put level the events let us say this is A the blue one is B 1 the red one is B 2 and the green one is B 3 right. And now we need to put numbers to it.

So, let us this is 0.3 it does not this 0.3 does not include the area which is enclosed by the blue circle, but rather just the part over here this is the difference with the Venn diagram. This is 0.2, this is 0.2 right and then B 1 let us call this as 0.1 and B 3 also we will call it as 0.1.

The first thing to note is that I have what is the total number that it add? It should add up to unity so 0.3 plus 0.1 is 0.4, 0.6, 0.8, 0.9 and therefore, the rest of the sample space has 0.1. Now then you should note that probability of A is 0.3 plus 0.1 plus 0.2 which is 0.6. Probability of B 2 is 0.4, probability of B 3 is 0.1.

(Refer Slide Time: 16:02)



Now, let us say you want to ask the question probability of A given that B 2 has occurred right. Now, if you look at this very carefully then you see that half the time B 2 has happened A has happened. So, very quickly you can see that this answer is half and if I want to use the formula that I have given over here then A intersection B is 0.2 and the probability of B 2 is 0.4 which is half, which everything agrees very nicely.

A given that B 1 has happened, probability of a given that B 1 has happened is essentially if you look at this Euler diagram when you see every time B 1 has happened A 1 has happened. Unlike the first case when you looked at B 2.

Not always when B 2 has happened, A has happened only half the times B 2 has happened A has happened. And therefore, you see this is the first probability is half, but the second probability is 1 and the answer is also if you get back the same answer if you use the expression that is given over there.

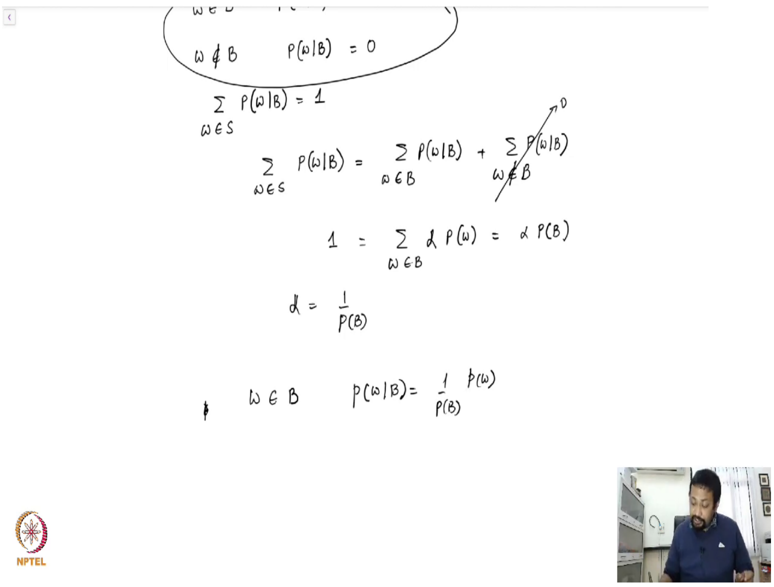
Probability of A given B 3 is 0 which is very true because probability of a intersection B 3 is 0. So, this essentially tells you that the two expressions you have it clearly tells you this illustration clearly tells you that you are restricting the sample space to the sample space of B, well in this case B 2 in this case B 1 and in this case B 2.

So, that these events B 1, B 2, B 3 they become your new sample space when you are asking for the conditional probability P A given B 1, P A given B 2, P A given B 3. Now, let us take a formal derivation try to see if we can formally get this ideas. So, let S be a sample space right and if this sample space has outcomes omega right.

So, that this forms the sample space, suppose, that now an event has occurred and this event B is a subset of S correct. Now, a new probability needs to be assigned on omega with this knowledge. So, we want to assign a new probability on omega, we want to assign a new probability of omega given that we know that event B which belongs to a subset S or a subset of S has happened.

So, if omega belongs to B then probability of omega given that B has happened is equal to let us say alpha times omega. Where, P omega was the unconditional probability need not condition it or anything and clearly if omega does not belong to B then P of omega given B has happened is equal to 0 right.

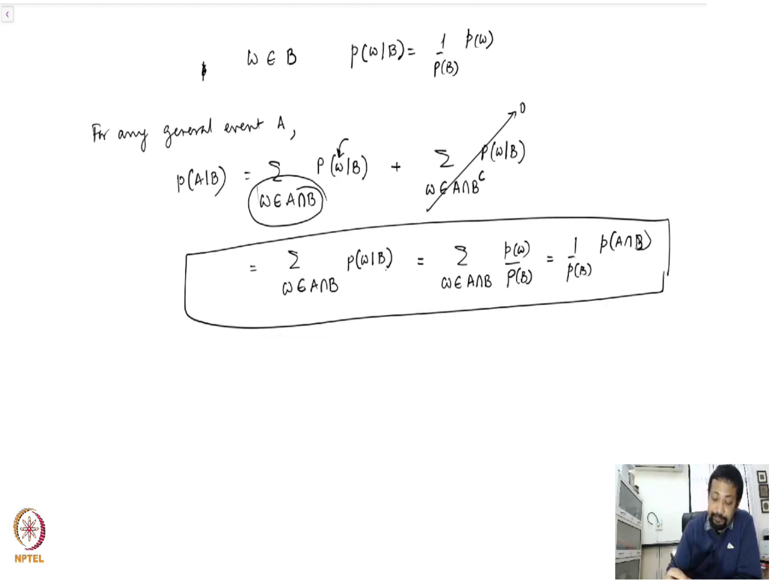
(Refer Slide Time: 19:34)



We further have the normalization condition that if this belongs to S probability that this is equal to 1 correct. So, I can now expand this so I can write down P omega belonging to S as this one as P omega belonging to B P omega B plus P omega belonging to B complement or P omega not belonging to B, since we have used them.

But by our conjecture with the hypothesis that we started off with this is 0. And therefore, I have omega belonging to B and this quantity is alpha times B of omega, since omega is now restricted to B therefore, this is P of B and the left hand side is 1. So, that alpha becomes very nicely P of B. And therefore, you see for any event omega that belongs to B you have reassigned the probabilities that P of omega given that B has happened is one of P of B times P of omega right.

(Refer Slide Time: 21:08)



Now, for any general event A let us say for any general event A, I want to write down P of A given B and that by definition is that omega belongs to A intersection B. So, P of omega given that this event belongs to this set of A intersection B plus omega belonging to A intersection B complement which is the rest of the sample space.

But this is 0 right so therefore, I have sum of a omega belonging to a intersection B P of omega given B, which is equal to sum of omega A into belonging omega belong sum of all the outcomes belonging to the set A intersection B P of omega divided by P of B and therefore, this is 1 over P of B P of A intersection B.

So, this is a very simplistic definition derivation of the expression that we did before. So, what we essentially again to summarize what we essentially did was we looked at the sample space and then we said look now an event has happened and I want to condition on this event.

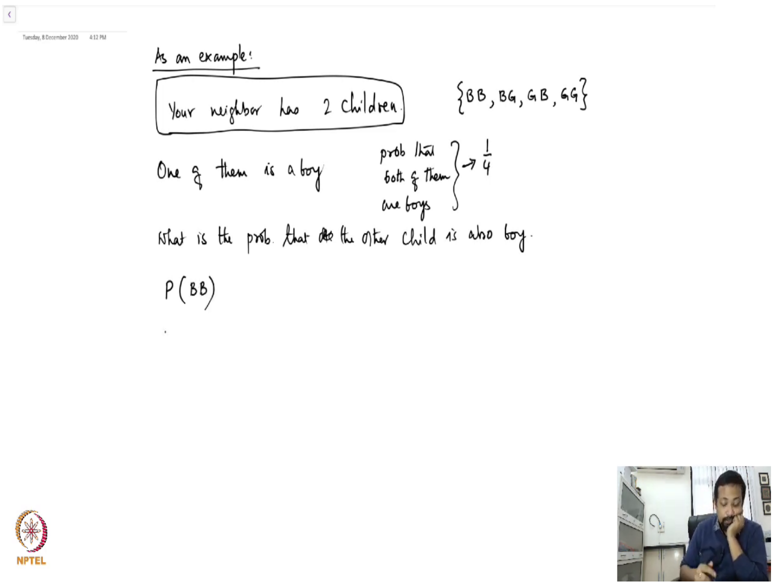
So, for that purpose I want to reassign the probabilities of the outcomes given that this event has happened.

And let us say that event is B. So, all we did was we looked at we said that if omega belongs to B then this must be a denormalized probability with alpha times P omega. If omega does not belong to p then that probability is 0.

So, then we use the normalization condition, if omega belongs to S P omega given B must be equal to 1 and once you have the normalization condition then you can expand this as either omega the outcome belongs to B or the outcome does not belong to B. If the outcome does not belong to B then the probability associated with has been is 0 right.

So, therefore, the left hand side is 1 and that gives you the normalization and then if you consider any event A for any general event A P A given B has happened is essentially P of omega. The probability of an outcome omega probability of an outcome omega given that B has happened such that omega belongs to A intersection B and A intersection B complement, but this probability is 0. Then this expression I can find to write down I can substitute for P omega given B and essentially you get this nice result.

(Refer Slide Time: 24:23)



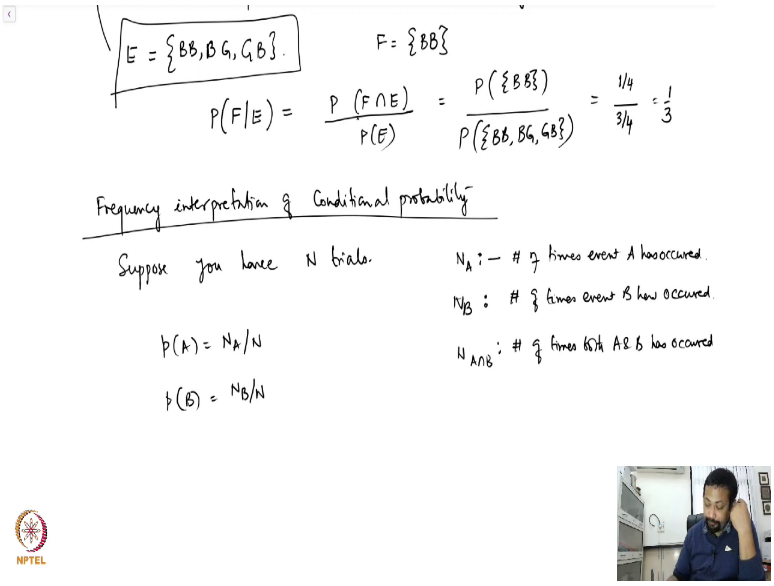
As an example, so of conditional probability let us take the following. So, you know that your neighbor let us say your neighbor has a son right they have 2 kids. So, let us say has 2 children right. Now, given just this information if I ask you that what is the probability that your neighbor the children of your neighbor are both sons he has 2 sons.

Then you will say that is very simple to answer let us look at the options that we have both of them are boys, one of them boy girl, girl boy and then girl girl. So, this is the sample space that you are looking at and you will immediately say that look if I do not have any other information, then the probability that both of them are boys is one-forth, probability that both of them are boys is one-forth equally likely to happen,

But, now I give you an additional information that one of them is a boy. So, now, you ask the second problem ask the question that, what is the probability that the other one is also a boy.

So, you ask the question, what is the probability that the other child is also a boy right? So, what you are looking for is essentially you are looking for probability of B B right.

(Refer Slide Time: 26:38)



So, in this whole when you would say that one of them is a boy then you see this sample space gets restricted to B B, B G, sorry there is no comma here and to G B right. So, let us call this as both of them are boys as F and this one as E, then the answer that I am looking for P of F given E. Because this part E is essentially is the answer to this question, that if one of them is a boy then your new sample space is just these 3 right, and which is E.

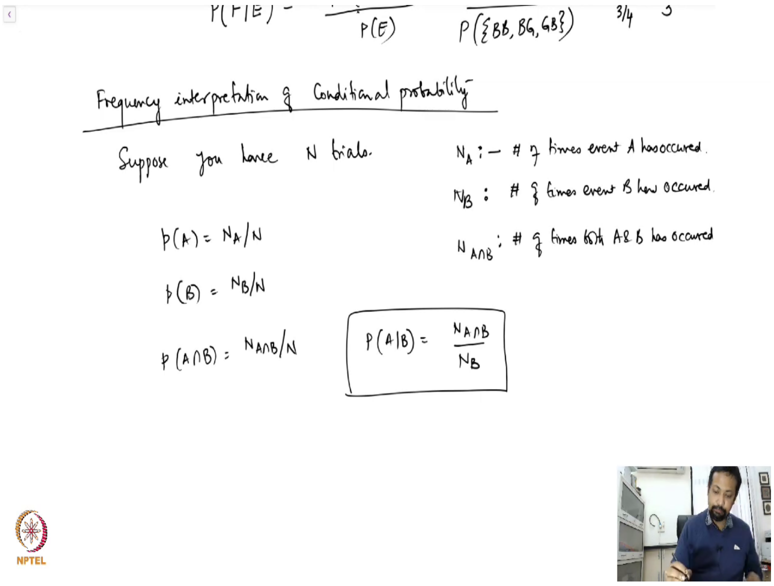
And the answer that you are looking for is what is the probability that the other child is also a boy is essentially that your neighbor has 2 boys is F let us denote them as F then the answer is P of F intersection E given P of E right.

So, that is probability of B B divided by probability of B B, B G and G B correct. The probability of B B is clearly one-forth which we have worked out and the other one is three-forth and therefore, this is one-third right. So, this is how you essentially use condition evaluate conditional probabilities. It becomes very very clear to you once you write down the questions and the once you write down the sample space that you are really interested in that the conditioning has forced you to do.

So, one can of course, write down a frequency interpretation pretation, of conditional probability and there suppose you have N trials right. And then essentially you count the number of times then event A has occurred and let us denote it as N A is number of times event A has occurred and N B is the number of times event B has occurred.

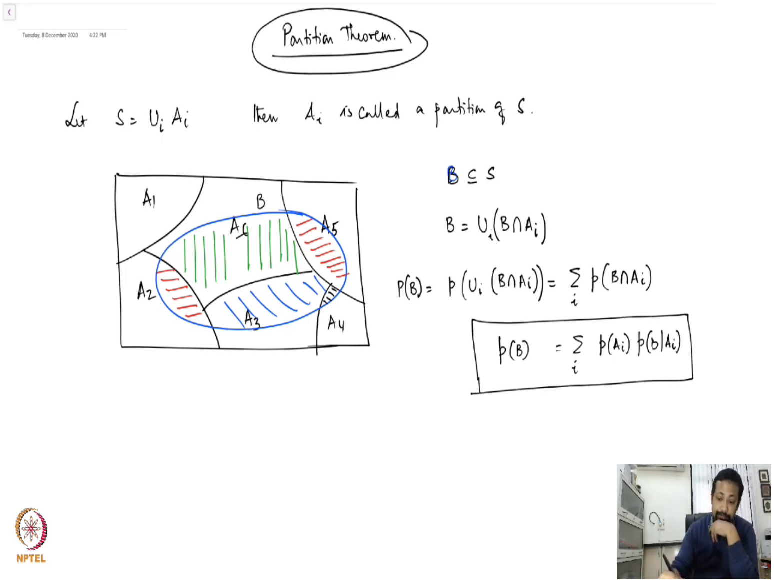
And one can also have the number of times when both A and B has occurred right. Then if N is large enough statistically speaking so then P of N A as we have P of A as we had worked out earlier the frequency interpretation P of B is N B over N.

(Refer Slide Time: 30:06)



And P of A intersection B is N of A intersection B by N, where this denotes essentially the number of times A and B has jointly occurred. So, the conditional probability in the frequency interpretation becomes N A intersection B divided by N B right.

(Refer Slide Time: 03:36)



So, we want to now look at what is called the partition theorem. So, although; so please realize that this material on probability is just a bridging between thermodynamics and statistical mechanics, we are not going to do anything in detail because I am sure you have learnt probability by now and several times.

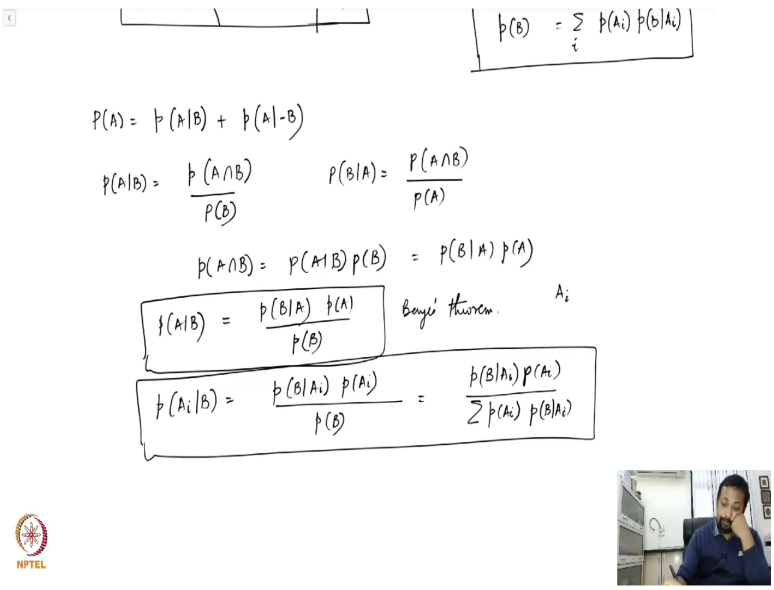
So, now, let us look at the partition theorem, what does the partition theorem tell you? That is let the sample space be written as right. Then essentially A i is called a partition of S, what does this mean? That means, I can have this is my whole of my sample space I can A right. So, this can be A 1 this can be A 2 this can be A 3, A 4, A 5 and this can be A 6.

So, these all these A 1, A 2, A 3, A 4, A 5 are events and these are subsets of S except that this whole thing essentially makes up the sample space S right. If I now consider B which is a subset of S and denote it over here like this way right, then you see there are several intersections right.

And, now with this so let us call this subset as B; I can write down B as union of right. So, then if I want to calculate the probability of P then essentially I will calculate indirectly the probability of this quantity right. And by the additivity theorem this is nothing, but sum over i probability of B A intersection A i right.

But then I know, that essentially this is going to be sum over i p of A i p of B given A i right. So, this tells me that the probability of B I can write down as in this particular form where, A i is at the partition of S.

(Refer Slide Time: 33:38)



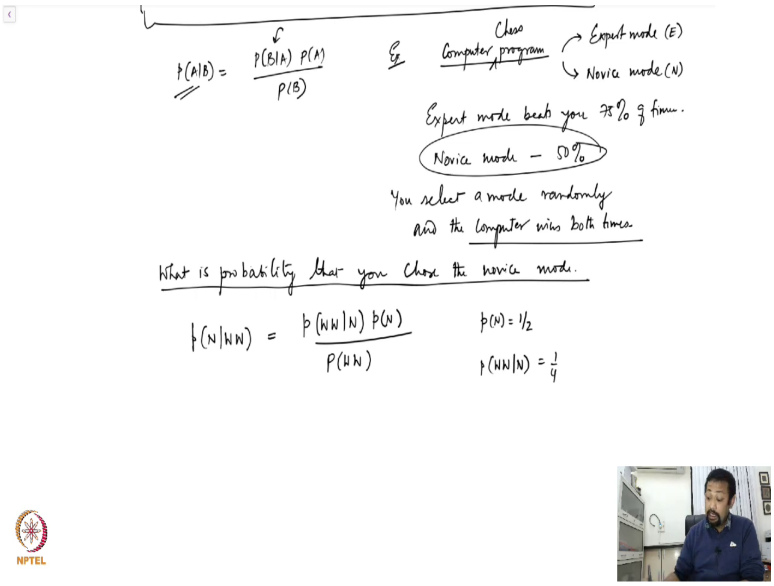
Now, I also know that probability of A is probability of A given B plus probability of A minus of B that is B has not occurred. P of A intersection B is sorry p of conditional probability of A given that P has occurred is p of A intersection B divided by P of B.

And as we had already written down P of B A given that A has happened, this P of A intersection B is divided by P of A. Which means it clearly that A intersection B is you can either write it down P of B which is also P of B given that A has happened times the probability of A right.

Therefore, you see I can write down this p of this conditional probability that probability of an event A given B has happened as probability of an event B given that A has happened times p A divided by p B right. Now, if there are multiple events this is just one event I am talking about it let us say this A there are multiple events A i and I want to know p of A i given B. Then p of A i given B is this just very simple expression p of B given A i times p of A i divided by p of B.

But p of B we just calculated right using the partition theorem we calculated this as sum over we use this expression p of A i times p of B A i. So, this is the generalization of the Bayes theorem this relation that you see over here is the Bayes theorem. Now, the advantage of this expression the Bayes theorem that we have seen is first of all not only it tells us how the probability is modified with presence of a new evidence.

(Refer Slide Time: 36:20)



But you see, in this expression what we have done is we have written down p A given B has occurred in terms of P B given A times P A divided by P B . So, we have kind of recast this calculation of this probability often the 3 quantities that you see on the right hand side of this expression is much easier to evaluate and is much easier to estimate also.

So, therefore, this gives us a very useful expression. As an example let us say I have a computer program right and this computer program has two modes. So, well this is a computer chess program and this has two modes one is an expert mode and the other one is a novice mode right.

Now, the expert mode, so we will say that this is the expert mode this is the novice mode the expert mode the, the expert mode beats you 75 percent of times right. And the novice mode will the novice mode you have a 50 percent chance of winning. So, the what you do now is let us say you close your eyes and you select a mode you do not know which mode you have chosen right.

You randomly choose one of the modes and the computer wins both times. So, you select a mode randomly and the computer wins both times right. Now, you ask the question that, what is the probability that you chose the novice mode? So, I want to ask the question that what is the probability that you chose the novice mode?

See if all this information was not given which means all this information means that you the computer wins both times, if this information was not given to you then you would say that the probability of choosing a novice mode and winning twice is 1 4th right. So, sorry is 50 percent. So, probability of choosing the novice mode and winning is 50 percent right.

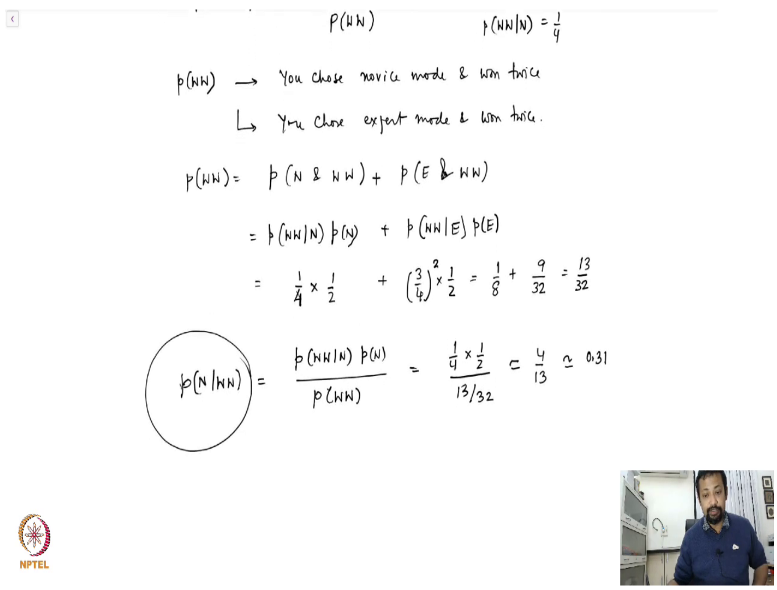
So, but now essentially what is given asked what is being asked here is what is the probability that you chose the novice mode given that you have the computer has won twice. So, what is the probability that you seek novice mode given W W right.

So, W denotes a win, but this I can clearly write down W W N I am reversing the question now, times p of N divided by P of W W right. By the Bayes theorem that we have written down I am just reversing this expression. So, I want to know, what is the probability that I have chosen the novice mode given that the computer has won twice?

And I want to rewrite this I want to evaluate this by choosing this expression where it is expressed in terms of a different thing. Probability that you have of winning twice given that you have chosen the novice mode right.

So, now, probability of choosing a novice mode is half because there are only either you choose an obvious mode or you choose the expert mode and you have closed your eyes and you have selected one option right. And therefore, probability of winning given that you have chosen the novice mode is one-forth that is very clear right.

(Refer Slide Time: 40:44)



Now, I have to figure out what is the probability of winning twice? And here essentially there are two possibilities one is you choose novice mode and won twice you choose expert mode and won twice correct.

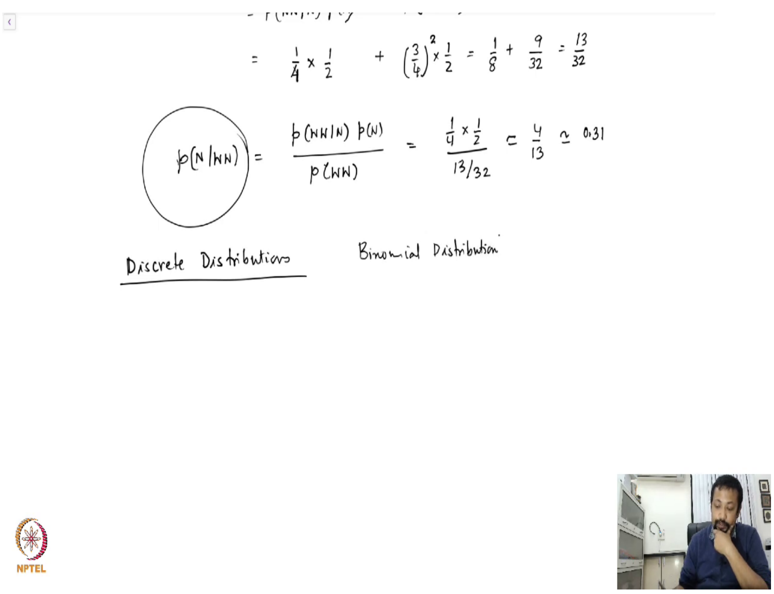
So, therefore, I can write down P W W as probability that you have chosen the novice mode and won twice plus the probability that you have chosen the expert mode and won twice right. So, probability of choosing the novice mode and winning twice is probability of N times WW N ok. So, let us write down the same thing in a different way so that it is much easier to.

So, probability of winning given novice mode times probability of novice mode plus probability of winning given expert mode times probability of expert mode. This we have already calculated to be one-forth, now this is creating a slight problem and this is half correct. And this I know is three-forth whole square into half and if I add everything let us say this is going to be one by 8 plus 9 by 32 which is 13 by 32.

So, now I have all the information. So, now I can calculate p of N WW as p WW N p N divided by p of W W which is p W W N is one-forth half and this is 13 by 32. So, this is going to be 4 over 13 which is approximately 0.31. So, now, you see the answer is that there is a 30 percent choice of you choosing the novice mode.

If it has reduced from 50 percent to 30 percent so the probability of choosing the novice mode was 50 percent right good. So, now, that we have the bare bone structures of probability theory ready we want to look at some this examples of probability distributions.

(Refer Slide Time: 43:47)



So, we will first consider discrete distributions since whatever we have been doing so far have discrete outcomes. So, the classic distribution that we will encounter in physics is what is called the binomial distribution.