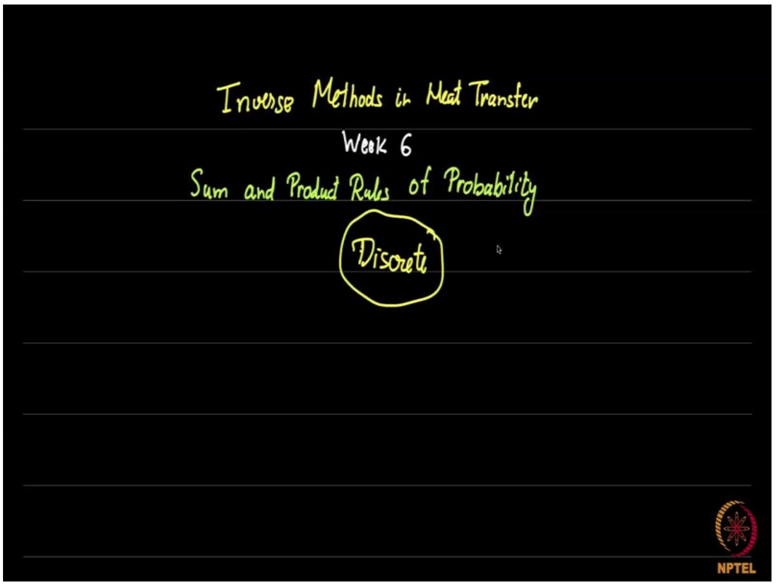
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology – Madras

Lecture – 31
Sum and Product Rules of Probability

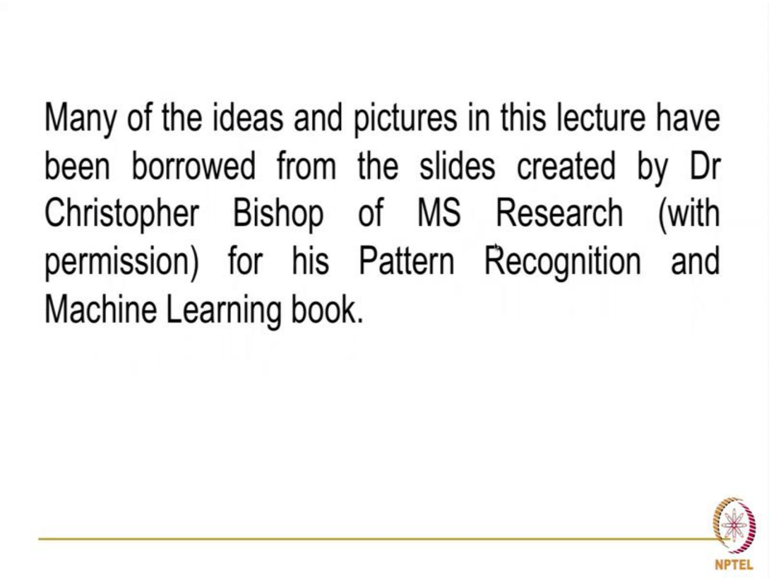
(Refer Slide Time: 00:21)



Welcome back, this is week six of inverse methods and heat transfer. In the last video we discussed why probability is important within inverse methods we also looked at definitions of sample space a random experiment and what a random variable is here we look at two major rules that govern probability which we call the sum and product rules. I am going to show those within the discrete setting here.

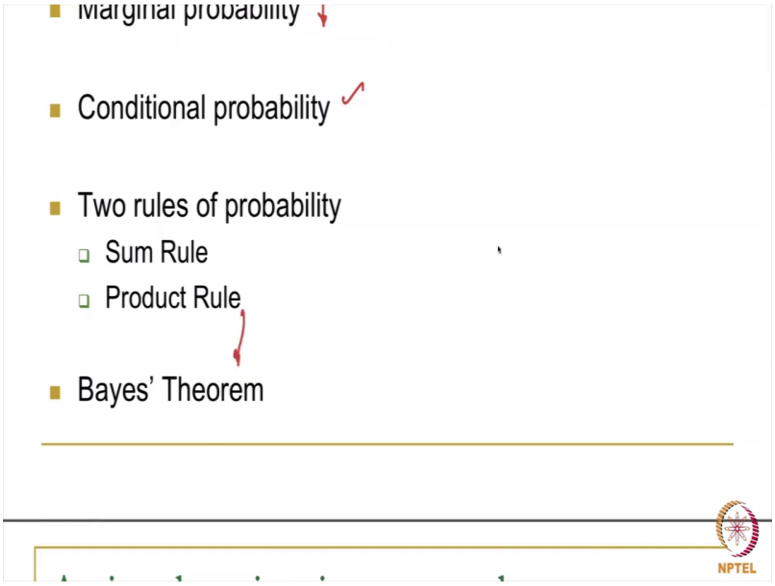
But these apply equally to the continuous setting which is where it is more important and we will discuss that further in the next week. It is often intuitive to work within the discrete setting and I will mostly work within that except for the end of this week but it is most useful to actually apply these within the continuous setting.

(Refer Slide Time: 01:13)



So, a quick acknowledgment many of the ideas that I have communicated here and the pictures also have been borrowed from the slides created by Dr Christopher Bishop's excellent book called Pattern Recognition Machine Learning, I had already told you about that earlier within this course. This book is available freely online, thanks to Microsoft research. you can take a look at that book the slides were borrowed some of the pictures from the slides were borrowed with his permission. So, this is just an acknowledgment.

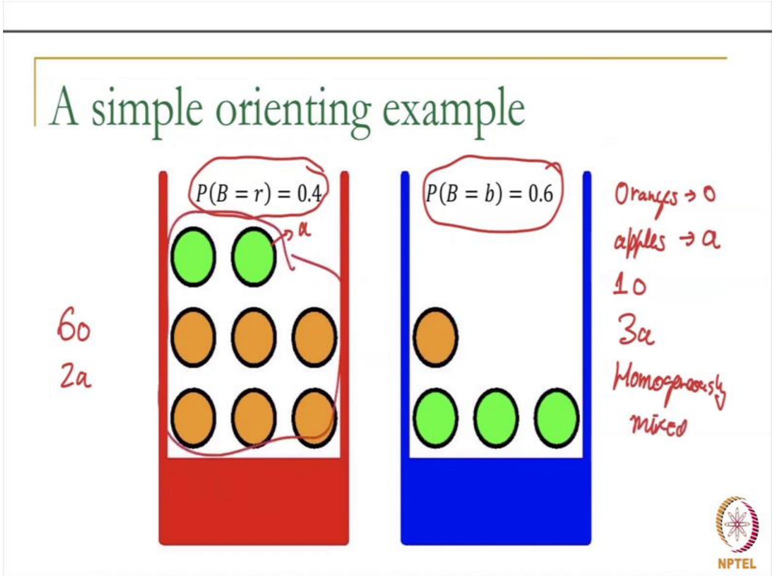
(Refer Slide Time: 01:50)



So, the ideas that we will be covering within this video are that of joint marginal and conditional probability. Now marginal probability will not come up too much with an inverse Methods at least the way I am teaching it in this course but joint probability and conditioning probability will come up. Marginal probability is sort of an intermediate idea that we will be using of course it is an independently important idea later on also with the probability theory or even generally with an inverse method.

But we will not be using it much. we will show two rules of probability that arise naturally the sum Rule and the product rule you will see that these are very intuitive rules and finally we will see that how product rule leads directly to basis theorem within the end of this video.

(Refer Slide Time: 02:34)

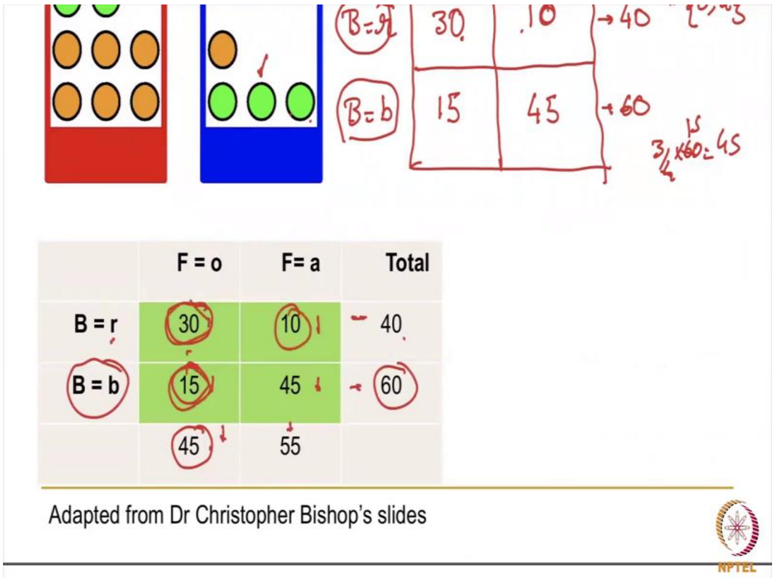


So, here is a simple example that we are looking at again like I said I borrowed it from Christopher Bishops book. Now like I said it is a discrete example and you might not see any relevance to inverse methods here, but I will explain how similar things occur even within the continuous setting shortly. but for now, just humor me and assume that this is an important problem. So, let us say you have two baskets and one of the baskets is red and another basket is blue.

Let us say within the red basket you have these six oranges and two apples and within the blue basket you have three apples and one orange. So, you just the orange corresponds to an orange and the green corresponds to an Apple. So, oranges I will denote as o and apples the green ones are they not as a. Now what happens is when you put your hand let us say you are not able to see which hand or which basket you are putting your hand in and let us say you pick the basket the red basket the probability of 0.4.

And you pick the blue basket with a probability of 0.6 and of course this has six oranges and two apples and this one has one orange and three apples. Both baskets are not equally likely and of course if you pick something like an orange within it is not equally likely that you will get an orange or an apple, but each fruit is picked up with the same process it is homogeneously mixed. For example, if I come to this basket and I pick an orange then the likelihood of picking an orange actually six by eight. It is not as if oranges are below and apples are above and like what is shown in the picture.

(Refer Slide Time: 04:48)



So, if we look at this picture here let us look at what the random variables are. Now there are two random variables. So, the random variable B is which basket I pick and the random variable F is which fruit I pick within that basket. So, you take your hand, you put it in a random basket, and you pick out a fruit and you land up with both a basket that you pick from as well as a fruit. So, there are multiple questions that you can ask at this point. for example, you can ask a question like which basket did I pick randomly.

If I just randomly touch something, I could ask which basket did I pick and you know that you will pick by the numbers given here the red basket with probability 0.4 and the blue basket with probability 0.6. Also, you could ask which basket did I pick, given that I picked an apple. Now this changes things. So, suppose you put your hand and you picked up an apple you might. Now want to think which basket are you more likely to have big.

Now notice that from this basket you are more likely to have picked an apple anyway because it has more apples and also this basket was more likely. So, now the answer no longer will be blue with the probability of 0.6, if I already know I picked an apple it is probably more likely than 0.6 that I picked the blue basket. So, that would be the example of a conditional probability case. So, let us come here and look at it sequentially before we come to this idea of joint marginal and conditional probability.

So, we let us make some trials. So, let us say we made 100 trials, that is somebody sat there and kept all picking up fruits, saw which fruit they picked which basket they picked and now made a table of how many they picked this should remind you of the histogram. Now we are dealing with two variables rather than one variable. In the previous video we primarily dealt with one variable. When you deal with two variables it becomes important to deal with what is known as joint probability.

Now let me take a brief moment here and talk about how this is relevant to inverse methods. So, remember once again, when we dealt with a slab example, we had let us say some six thermocouples. Now the question is we measure temperature here, let us say it came to 15 degrees became measured here and let us say it came to 14.7. So, we can treat this as two separate baskets. So, this could be the Red Basket and this would be the blue basket and we could also treat the temperature as picking an orange or an Apple.

So, remember this is now a continuous variable of course it is sample space is continuous whereas here in basket, you have a discrete sample space. So, you have basket equal to red or you have a basket equal to Blue these are the two samples. it is similarly the fruit has a sample space of Orange or Apple. Whereas here the sample space is infinite and you will have temperature variation which is infinite similarly sample space here is also infinite but you do have to discrete thermocouples you could either take this thermocouple or that thermocouple.

Now we come here and we ask what is the probability that the temperature in the first thermocouple was 15 and the temperature in the second thermocouple was 14.7. So, that is the sort of question that we can ask and that will be similar to the joint probability case that we will be discussing here. So, let us come back to this case. There are multiple ways of equating this as we will see in the next week to a joint probability case when we come to an actual inverse methods problem.

So, coming back to this let us say I made 100 trials just like I made 100 measurements I made 100 trials and I am going to make a Frequentist sort of assumption here and I am going to assume that things are going to fall exactly in proportion to their probabilities this will happen really speaking only for infinite trials, but for the sake of demonstration we are going to do this. So, we are going to say well the basket would either be red or it could either be blue similarly the fruit could either be an orange or it could be an apple.

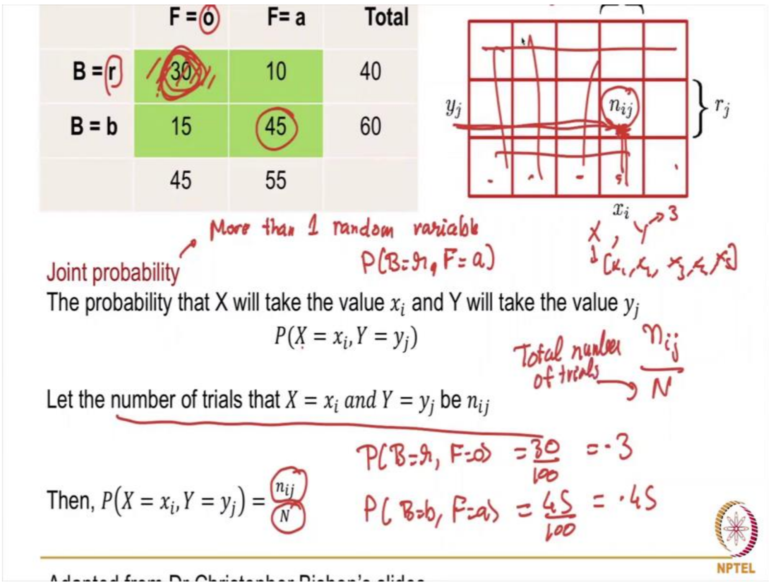
Now let us see the total number of cases that we come up with in 100 trials. So, once we make 100 trials in how many of those cases would we have landed up in the red basket. So, the net number of cases where we have landed up with the Red Basket is 40. Assuming it goes exactly according to proportion. Similarly, the net number of cases we would have landed up in the blue basket would be 60 because its probability of picking that is 0.6.

Now within this 40, what is the possibility that I got an orange, that is the basket was red and I picked an orange. Now I come here I see orange is 6 by 8 of the fraction of roots here. So, the probability of picking an orange is six by eight given that I have already picked up the Red Basket. So, 6 by 8 into 40 of the cases which comes through 30. So, in 30 cases I heard I have picked an orange and in the rest of the 10 cases I would have picked an apple.

Similarly, here once the basket is blue in three-fourths of the cases that is three-fourths of the 60 cases I would have picked an Apple. So, this will be 45 and in 15 of the remaining cases I would have picked up an orange. Now we can also sum it up in another way we can say that within the 100 trials in 45 of the cases I picked up an orange and then 55 of the cases I picked up an apple. How would that have happened in the 45 of the cases either I picked up the rest Red Basket, in which case 30 of those cases would have turned out to be orange 30 out of the 40.

Or I would have picked up the blue basket in 15 out of those 60 would have turned out to be oranges. So, we can sum this up in two different ways this is what will lead to what is known as the marginal probability and each individual one will lead to the Joint probability. So, let us see that in the next slide.

(Refer Slide Time: 12:24)



So, we can now use this box that we had to define this idea of joint probability. A joint probability is obviously for more than one random variable. So, you want to say what is the probability that event A occurs as well as segment B occurs and of course we can. Now talk about even C occurs D occurs etcetera. Once again going back to the slab case we can talk about six random variables and talk about the joint probability of all of these temperatures hitting the values that they did give our measurement.

So, we are actually trying to find out the probability of a specific event which is in turn made up of six events. In this case we have two possibilities for example or two random variables here, which is probability for example probability, that the basket was red and the fruit cause an orange. So, notice this this comma here is like an AND so, that is why the word joint occurs hearing. So, for example we would say notice the notation probability something first random variable b equal to red AND fruit equal to let us say apple.

So, x is the first random variable y is the second random variable. In our slab case this is now a continuous case. So, we could say something like probability that the temperature lies in the range 10 to 15 and the probability that the second temperature lies in the range 12 to 17 something of that sort. Now you might recall that obviously these temperatures are given specific values. So, where is this range coming from that is where you must remember our Sigma there is uncertainty in measurement.

You will make multiple measurements and you will say the range of measurements I have made with uncertainty is so much and that is where these ranges come from. For continuous variables these ranges are far more important than specific values which is where our uncertainties come in and automatically you will get our weighted linear regression when we come put continuous random variables in the next week.

Returning back to this example, let us generalize this this is of course the basket case with two variables and each of them had only two possibilities. You could have a more general case where you have two random variables x and y and x has five possible values. So, x has x_1, x_2, x_3, x_4, x_5, let us say it is a five-sided Dice and you are throwing and you could get the number 1, 2, 3, 4, 5 and similarly y has three possibilities in that case here we had a two cross two, here we would have a 5×3 5, because x has five possible values and y has three possible values, we are just generalizing this to a generic picture.

So, just like here, I was concentrating on some particular block, this would be the block which talks about in how many trials we had the red basket as well as the fruit as an orange. This will talk about in how many cases did I get the jth variable for y and ith variable for x. So, for example the number of trials, please notice here X equal to x_i and Y equal to y_j for example in the five-sided dice I threw four and the three-sided dice I threw two.

So, that would be an example of this case. of course, once again in our continuous case both these are infinite which is why we give ranges. So, both these are probability density functions and you have to be a little bit more careful while defining those we will come to that in the next week. Now it is easy to now talk about what is the probability that after my trial I pick up a fruit which was orange from the red basket.

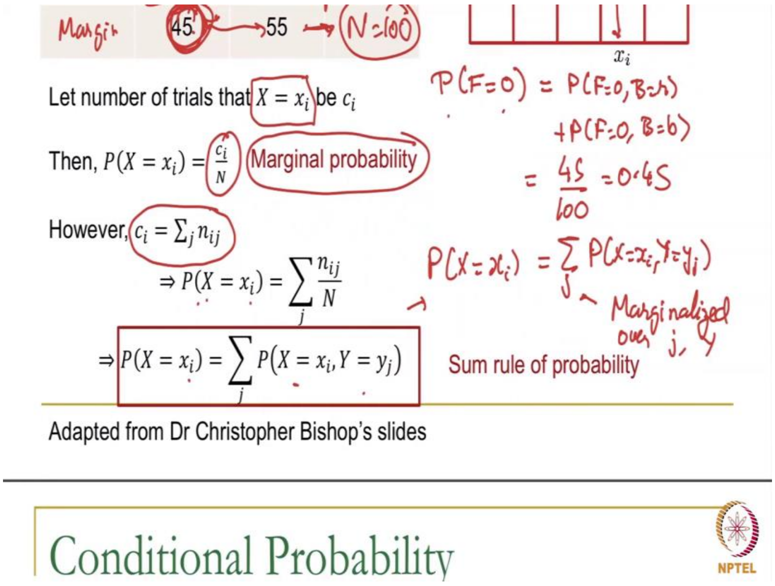
So, in that case all we need to do is to divide the number of cases we got this the joint number of trials divided by the total number of trials using the fundamental idea of probability. So, for example probability that the basket was red and fruit was orange. Now becomes 30 by 100 it is 0.3, similarly probability that the basket was blue and the fruit was an apple is 45 by 100. and the probability is 0.45 in general, it will be n_{ij} by N where N is the total number of trials.

So, that is what the joint probability is. Now you can imagine once again I will go back to our slab case you can. Now imagine what is the total number of cases or what is the joint probability of temperature being in some range here and being in some range here you have to. Now in imagine an infinite trial case where you are actually looking at all the physics of the problem and seeing where all in how many of these cases.

Now this will look because it is all infinite cases you have to know how to divide Infinities and we are not going to get into the limit process of that there is something called measure theory etcetera. We will just take it as if we took a lot of measurements maybe a million measurements which is what we will do next week. We will do something called Monte Carlo which actually takes a finite number of samples and assumes just like we did here.

That even though with 100 we took exact probability ranges or exact fractional ranges we will do the same thing there and there are several clever ways of doing that. So, please do keep that in mind I hope at least this example is reasonably clear what joint probability is.

(Refer Slide Time: 19:04)



So, once joint probability is there, we can also talk about something called the Sum Rule and the marginal probability. the sum rule is a very simple thing. So, the idea is this. So, we want the number of cases let us say X equal to x_i. So, for example I could ask the case in how many cases was the fruit and orange. So, that could be an example of X equal to x_i. So, that would be simply the marginal probability that is what is defined as the marginal probability I will explain very shortly why it is called marginal.

So, the way we do it is the number of cases where the fruit was an orange and we picked the red basket plus the number of cases where the fruit was orange and we picked the blue basket, this is simple you might actually say why am I struggling. So, much I might as well take this total case which I have summed up but many times we just have this table and this summary is not there I have put in here the sum for a specific reason.

So, we say 30 cases fruit was an orange, basket was red, 15 cases fruit was an orange, basket was blue put together we have 45 cases where the fruit was an orange. So, randomly if I look at all the cases where the fruit was an orange that would be 45 by the total number of Trials. So, the probability that a pick an orange is 0.45 out of either of the baskets and similarly the probability that I pick an apple from both these cases will be 0.55.

So, if I do 100 trials never look at the basket only look at the final fruit in 45 percent of the cases I will get an orange and 55 of the cases I am going to get an Apple. Now um one thing that is kind of strange here and it is not immediately obvious is if you look at the total number of fruits this is one place where probability does not always reasonably work. Total number of fruits, the number of fruits oranges is actually much higher than the number of apples.

And yet the probability of picking an apple is actually greater than the probability of picking an orange. So, this is one of these non-intuitive things or sometimes even counterintuitive things that happen with probability Theory, we don't have the time to go into that within this course but that is one of the major reasons why statistics can be confusing. So, once again you can see the final numbers the numbers speak for themselves when we work for numbers it is obvious.

So, you get 45 cases total oranges it is just that it is overwhelmed by the case where the basket was blue even though a greater number of oranges seven compared to four it is not as if you are getting uh 7 by 11 for orange it is actually 0.45 you have to be careful to split this. So, let us come to this probability which is called the marginal probability. Why is it called the marginal probability because this is a margin.

These numbers were written within margins when people drew these tables, historically this is what was done. you make the table then write this sum within the margin you would have seen the second Excel sheets also you know row some columns that is how you write it. So, if I write this number c_i is simply the summation of all these $n_{ij}'s$ within this column. This this summation is within j you can see this c_i is,

$$c_i = \sum_j n_{ij}$$

And

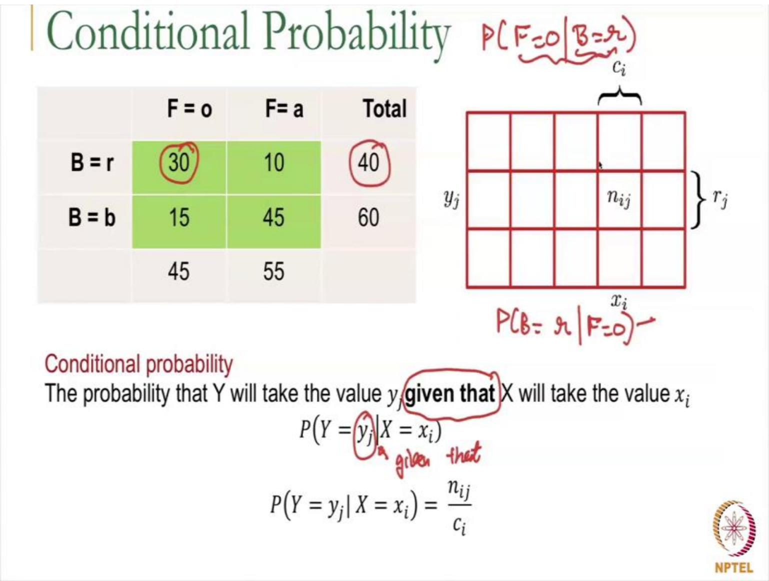
$$P(X = x_i) = \frac{c_i}{N}.$$

Whatever is the total sum this is what is all oranges that divided by total number of trials, whether you do this sum or this sum you will always end up with the same number of trials. So, this number divided by this number is what is called the marginal probability. it is a sum of the probability for one specific case. You could also do the marginal probability for basket being red.

But we already knew this I mean it is not a new quantity, we already started with this it is 40 divided by 100 which is of course 0.4. So, the more calculated probability is the marginal probability that the fruit is an orange. So, you can write this as in general this formula. this is what is known as the sum rule of probability. What it says is probability that I want one particular variable is the summation of all possible joint probabilities.

So, this is called marginalized it is an odd term especially given the social connotations. this is a marginalized over J marginalized over y something of that thought you would say. So, marginalizing over all possible baskets, I get the probability that the fruit is an orange as 0.45. So, all this means is I sum over all possible cases. Now X equal to x_i is given I know that I want an orange fruit but it could have been from a blue basket or from a red basket sum over those probabilities again we tend to do this intuitively but it becomes hard to see this when we come to multivariable or continuous cases that is why I am writing this out explicitly. So, this is what is known as the sum rule of probability and it is one of the key ideas or one of the key rules that govern probability Theory. we will also come to the probability the product rule of probability shortly.

(Refer Slide Time: 25:11)



The product rule of probability arises naturally out of this idea of conditional probability. So, let me give you a sort of colloquial example of conditional probability and then we will come to this specific example and then we will also write the product rule of probability. So, the colloquial idea of or the simple idea of conditional probabilities, I let me ask a question like if you walk on the road what is the probability that you will meet an Indian.

Now if all of you are watching in India your immediate answer would be well it is very high because the number of people of foreign origin in India is actually pretty small. But if you are watching this video abroad you will have a completely different answer because you have a condition automatically in mind. Now the trick about conditional probabilities there is always a condition there is always a prior that we have in our mind which we will use later of our bayes theorem.

So, suppose somebody is watching this video somewhere in Africa this number reduces if somebody is asking watching in West Indies this number slightly increases if depending on which country, they are within especially within the African continent this number can increase or decrease if you are in China the number again changes that is what is a conditional probability if I tell you, it could be anywhere in the world.

Now the probability that you meet an Indian then becomes approximately the population of Indians in the entire world depend divided by the population of the entire world. Now this is where uh conditional probability comes in the condition affects what the denominator is and what the context is and what the probability is. So, similarly here if I ask after 100 trials what is the probability that I picked an orange then you will say well you didn't give me any further information the probability is 45 by 100.

So, that would be simply probability that the fruit is an orange, but I give you something more specific I say what is the probability that the fruit is an orange given that the basket was red. Now your number immediately changes because your context changes. Now this becomes 30 by 100. So, this is the notation that we use given that some extra information is given. if nothing is given you assume that this simply says that a trial was made.

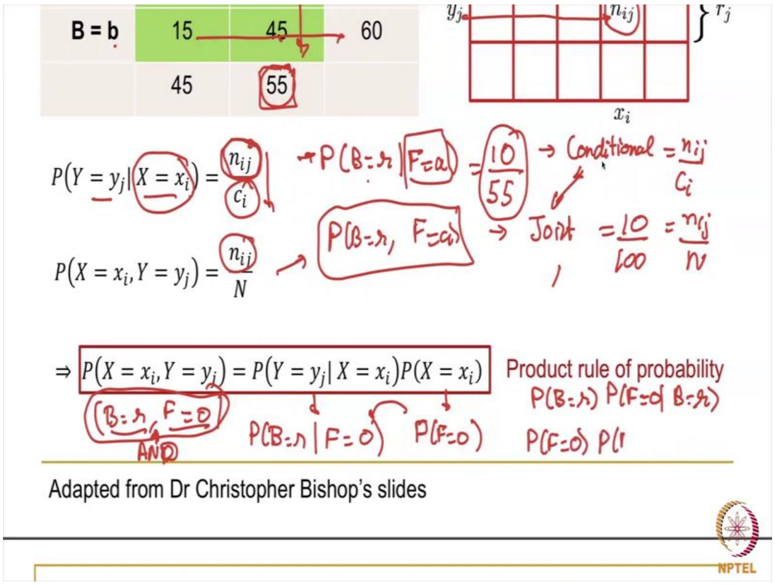
So, no context means a trial was made and I cannot give you any further information, but as I give you more and more and more information, you can give me more and more specific probabilities what is the probability that it will rain in India. Now if I give you a season and I give you a specific space or place then the conditional probability becomes more and more specific and more and more exact and more representative of what is actually happening.

So, more information changes probabilities, that is the basic idea behind conditional probability. So, we will use that with great power in the next week let us come back to this. So, a simple notation probability that Y is y_j given that. So, this bar rotation has chosen very carefully as you will see connects given that X is x_i. So, now this is easy once again we could decide something like this probability that the basket is red given that the fruit was an orange.

Now this is an inverse question to what I was asking before this was probability that fruit is an orange given that the basket was red, but here something opposite is given to you are told that you picked up an orange fruit. Now how likely is it that the basket that you picked was red. Now we know that without any information the basket being red this point four but given that you picked up an orange.

Now you can look at this and you can see that maybe that probability changes because if it was a red basket I would have more likely picked than orange than if it was a blue basket. So, this probability should hit weigh a little bit more heavily towards red with no information it was 0.4 with information, it should be a little bit more than that. So, now let us look at this in a more general case.

(Refer Slide Time: 29:57)



So, suppose I want the probability that this element in y was paid given that I picked xi, this becomes n_{ij} in this case it is in 10 divided by c_i by 55. let us make a concrete example probability. So, let us take this that the basket is red given that the fruit is an apple. Now how do we do this we say the basket is red and the fruit is an apple is 10 but I already know the fruit is an apple how many such cases are there only 55 cases out of the total trials.

Out of 55 trials in some cases in 10 of those cases I would have added up with a red basket and 45 of those cases I would have landed up with a blue basket. So, this becomes 10 divided by 55 not 10 divided by 100; 10 divided by 100 simply would be something else how many cases was basket red and the fruit red. So, fruit and apple so, this is the conditional whereas distinguish this from this case this is joined.

So, people often can get confused between these two. So, remember what is happening here. Here you made all these trials and out of those trials you want to find out how many of these were read as well as the fruit was an apple. Now this is a more restricted case this says amongst all the trials in a few of them the fruit was an apple but amongst those a few of those baskets were red which one was those and then you have this kind of case 10 divided by 55 which is why n ij is the total number of cases where both are proved.

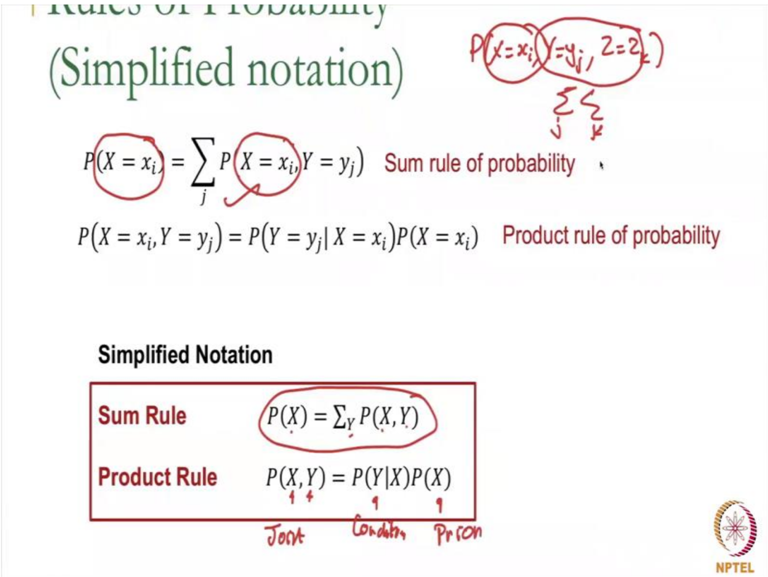
But c_i is the only number of cases where X was x_i as well as and I was n_{ij}. So, this is why you take n_{ij} divided by c_i here and similarly if you do it in this direction you will get n_{ij} sorry if you take the total number of Trials you will get n ij join probability you get n_{ij} divided by capital N this is n_{ij} divided by c_i . Now this becomes a little bit clearer when we talk about probabilities here numbers are more confusing whereas probabilities are clear.

So, let me ask this what is the probability that the basket was red and the fruit was an orange this is an AND condition this is what I will do is this. Let us assume that I first picked the basket and it was red. the probability that the basket was red. Now once I am given that I can say I already know the basket is red, what is the probability the basket is sorry the fruit is an orange. What is now the probability that the basket is red given that the fruit is orange.

If you multiply the two you will get the probability that the basket is red and the fruit is an orange. You can also flip this you can flip this in the following way you can say I want a red basket and an orange fruit. I can get it into this I first pick the basket then I say that the fruit was an orange and given that the basket was red or I can say that first I picked the orange and say what is the probability that basket was red given the fruit is an orange.

Now this is an easier calculation to make this is a more difficult calculation to make but switching between these two is how we get this idea of the base theorem which will come to shortly.

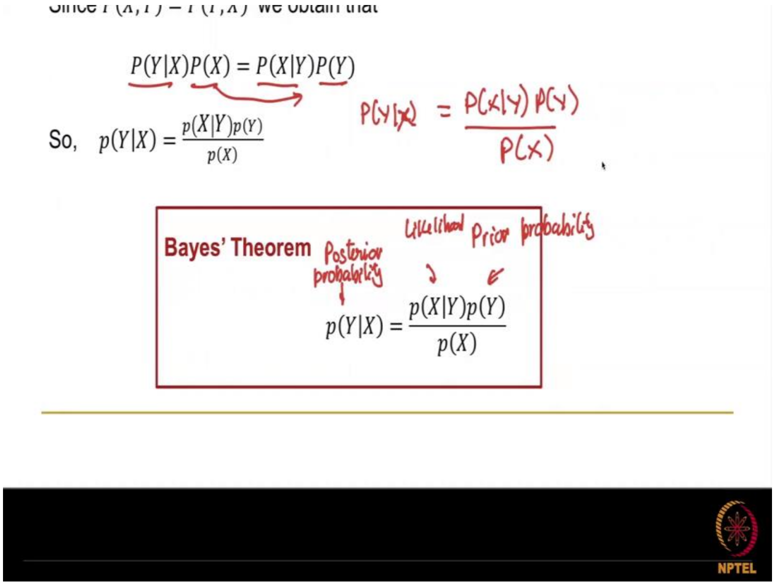
(Refer Slide Time: 34:21)



So, here are the rules of probability in a simplified notation remember original one simply says that if I want a single variable, I can sum over all the possibilities. If I say I have $P(X = x_i, Y = y_j, Z = z_k)$, then I will have to do a double Sigma over all possible values of j and all possible values of k and then I will get the possibility that $X = x_i$. So, here is a simplified notation really speaking you should use this notation but I am going to use this just for simplification probability of X, Y is summed over y gives me probability of X.

And the product rule is if you want both X and Y assume X is given and then say y given that X is given is the product rule. So, this is the conditional this is the joint and this is called the prior as we can see in the next slide.

(Refer Slide Time: 35:30)



So, here is a Bayes rule. bayes rule I am going to just define it sorry derived. it is a simple derivation assume the product rule. So, the product rule says that P of X and Y occurring the basket being red and the fruit being orange is probability that the fruit was orange given the basket was red multiplied by the probability that the basket was red. Now as I showed you in the previous slide.

You can flip this. you can ask the question what is the probability that the fruit is orange and the basket is red. now notice these two have to be the same because it is an AND so, it really does not make any difference whether I say basket is red and fruit is orange or fruit is orange and basket is red whereas conditional it does not commute

$$P(Y|X) \neq P(X|Y)$$

That is another easy way for you to see the difference between a joint probability and a conditional probability.

For example, a joint probability would be what is the probability that you are walking outside with an umbrella and it is raining, which is the same as the probability that you are that it is raining and you are walking outside with an umbrella. So, these two are exactly the same thing however if I flip the question and I ask what is the probability that it is raining given that you are carrying an umbrella.

So, let us say you stepped out of the house carried an umbrella what is the probability that it suddenly started raining this is different from saying what is the probability that it is raining sorry what is the probability that you are carrying an umbrella given that it is raining. Now that is much higher, you're about to step out it starts raining what's the probability that you're going to carry an umbrella.

So, notice that conditional probabilities do not commute, unless there are some special cases but these always commute with joint probabilities commute because they simply mean that all the events occur together. So, since these two are the same you can simply use these two equations let us say equation one and two equate them. So,

$$P(Y|X)P(X) = P(X|Y)P(Y)$$

So, now you can move one of these below it. So, let us say I move this below and I get,

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

This is what is known as base theorem. So, this is called the posterior probability. posterior means afterwards. This is usually called prior probability, which is called the likelihood. So, we will see this in greater detail in the next week.

But this is a powerful theorem even though it is coming through a very simple means. this becomes clearer the use of this becomes clearer with examples again you would have seen this in school but just for completion’s sake we will do a couple of simple examples in the next video. So, in this video we looked at the sum rule, the product rule and bayes theorem and we will see an application of bayes theorem in the next video. So, I will see you there thank you.