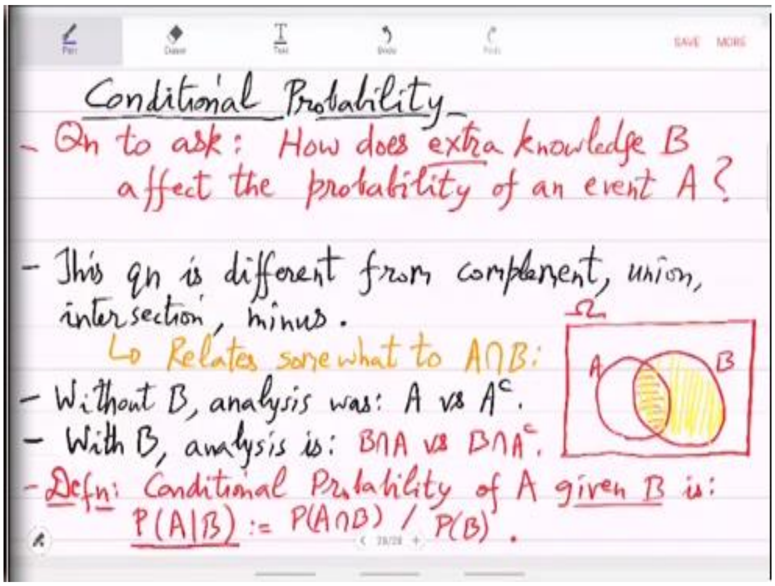
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science & Engineering
Indian Institute of Technology-Kanpur

Lecture-06
Conditional Probability, Partition formula

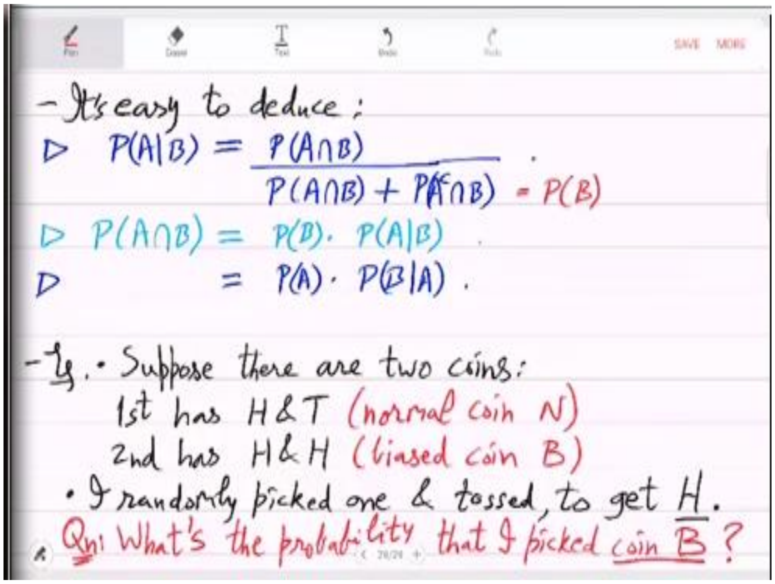
(Refer Slide Time: 00:14)



So, what happens with B? With the extra knowledge that B has happened, analysis is B intersection A versus B intersection A compliment, which is also equal to B - A, that is a difference. So, now if A happens, then you are actually saying that intersection B event, if A does not happen, then you are actually saying B - A, which is very different from the word without the knowledge of event B happening.

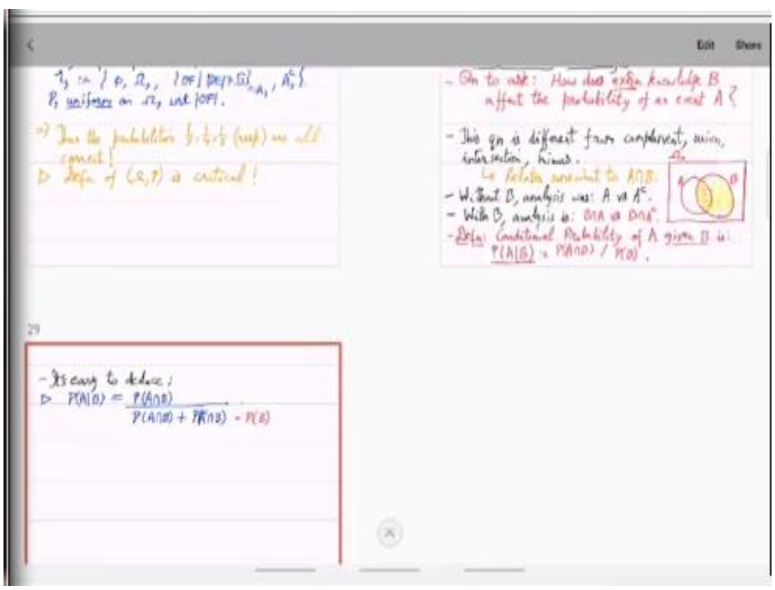
And to deal with this the most intuitive way formalise conditional probability is you defined conditional probability of A given B, P A given B, that is the symbol. This vertical bar is means given, so probability of A given B is defined as probability of the intersection happening divided by probability of B happening. So, take this as the definition of relative or conditional probability A relative to B, A given B. Then the intuition for this is clear from the Venn diagram. So, couple of things immediately follow from here.

(Refer Slide Time: 02:35)



It is easy to deduce that probability of A given B = A intersection B divided by A intersection B + A complement intersection B. This is because the denominator is equal to probability of B. So, you can as well write probability of B in terms of this first you see that B intersection A versus B intersection A complement, these are disjoint, it is a disjoint union. And then on that you use the probability map axiom or property which is PB is equal to sum of those two.

(Refer Slide Time: 03:43)



So, with that you can see that what we claimed in the picture in the Venn diagram, this is what is happening. So, it is the favourable over all in the definition. Other things that very easily follow, but are interesting is that P A intersection B is probability of B times probability of A given B, this is obvious from the definition. But when you write it this way, then it is telling you something about the intersection.

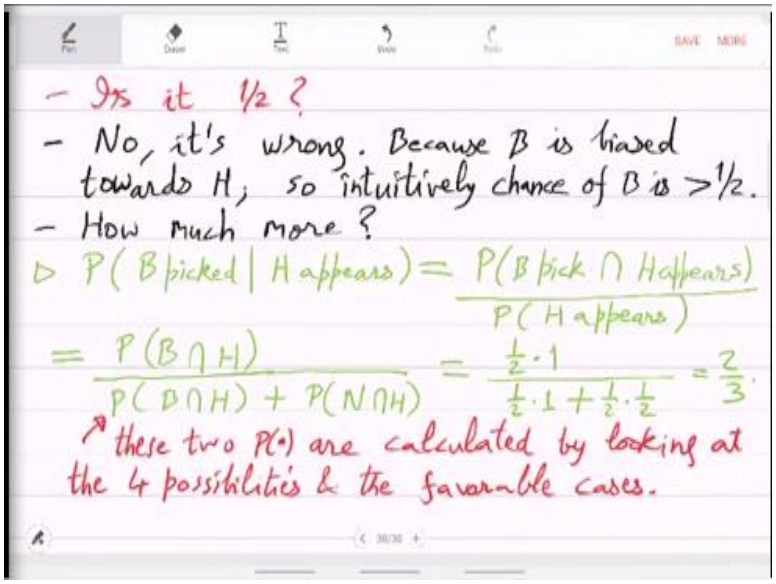
So, you have a new formula for intersection and what else? Yeah, so this is intersection is symmetric, so you also have probability of A times probability of B given A. So, you have this relationship between A given B, B given A. So, let us see a quick example of these formulas, an interesting example. So, suppose there are two coins, one is normal, so one has head and tail, so this we can call normal coin N.

And the other one let me use a slightly different notation. So, first has heads and tails, second has heads and heads, which means it is heavily biased, let us call it B. So, you have a unbiased normal coin and you have a biased coin, we are calling them N and B. You randomly pick one of them and toss, let me say I randomly picked one and tossed, to get heads. So, this is what I got at the end of the experiment.

Now it is a random experiment, so you can talk about probability, but probability of what? Because head I have already got, so there is no question about that. The question is which coin did I pick? Did I pick the normal coin or the biased coin? Probability of that. So, what is the probability that B was picked? So, what is the probability that I picked B? So, from the result of this random experiment being head what can you say about the coin that I pick which is secret from you, that being the biased coin?

Now intuitively vaguely speaking there are 2 coins, so the probability could have been half, the answer could have been half, is that correct?

(Refer Slide Time: 07:51)



No, this is completely wrong. Why is it wrong? Because the bias coin has more heads, it is biased towards head because B is biased towards, so intuitively chance of B is higher. This holds to reason. That the chance that I picked the bias coin is more once you know that head came out. But we are interested in measurement; we are interested in actually quantitative value of the probability, so how much more than half? That is the question.

So, let us do this calculation. How much more? So, let us calculate, so probability that B picked, given that head appears. Given that head appears the probability that biased coin was picked, this is by definition the intersection divided by the probability that H appears, which is now H can appear in 2 ways, either B was picked or B was not picked which is so it is equal to probability of B intersection H, I use the short form divided by probability that B was picked or B was not picked.

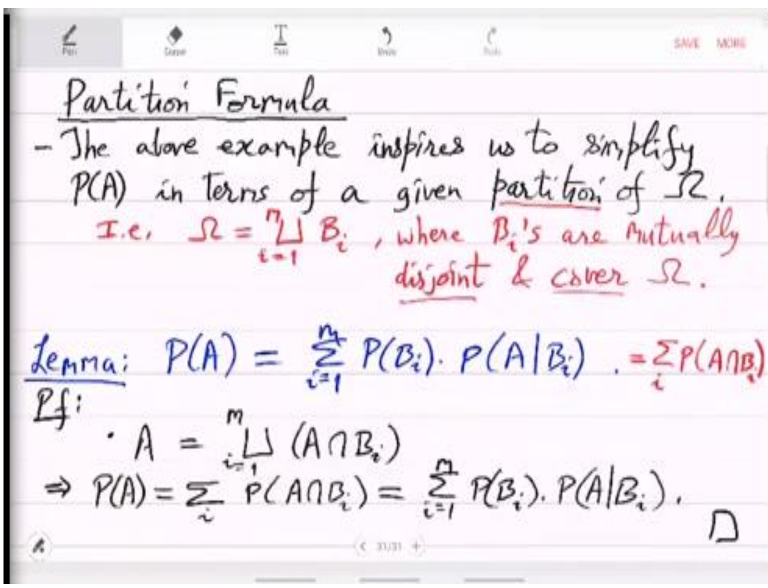
Now what is the probability that B and H both happen? So, B is picked by it is a uniformly random process, so either you will pick the bias coin or the unbiased coin, so probability is half but. If you have picked biased coin then probability of head is 1, same here. Now probability of B complement is what? B complement is essentially the normal coin, so let me put that.

So, normal coin is picked with probability half and in that case chance of head reduces, which is how much? So, this comes out to be 2 by 3, the way we calculated these probabilities, these are calculated by looking at the 4 possibilities and the favourable cases. So, there are 4 possibilities in this experiment, first you will pick normal or biased and then you will get head or tails.

And from these 4 possibilities, you can look at the favourable one and accordingly you get either half or 1 by 4. But then the amazing thing is that this ratio is what we consider and that gives you two thirds. So, once you know that head came out, it is significantly larger chance that the biased coin was picked. So, this is a very good example of conditional probability, it gives you a different view of probability.

And this is happening because you have extra information, without the extra information the probability was half, that you would be biased but now that probability jumps up. Let me do one more thing, so in this example we broke the denominator into two possibilities, B and B complement. So, let us formalise that because it is a useful trick, so that is called partition formula.

(Refer Slide Time: 13:47)



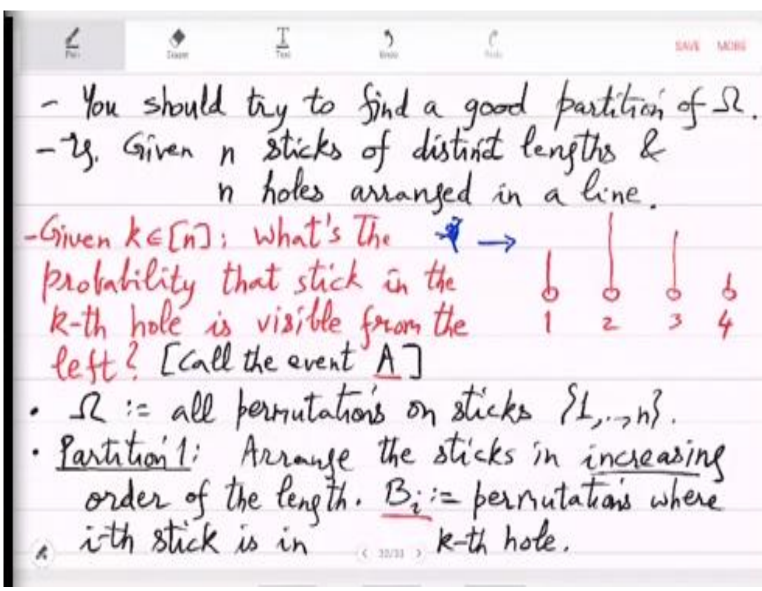
So, the above example inspires us to simplify probability of an event in terms of a given partition of omega. So, what is a partition? That is omega equal to disjoint union or union of disjoint subsets of omega called B i, where B i's are mutually disjoint. And as the equation says already they cover omega. So, in many situations you would have a understanding of omega in terms of simpler cases, which are these events B i's.

And you would want to break up A into these simpler cases. And then you will do those calculations and you will add up. So, what is the formula for adding up? So, probability of A is probability of B i times probability of A given B i. So, this is an immediate application of conditional probability. What is the proof? A is breaks up into these parts given by B I, so A intersection B i because well, B i covers and partitions omega.

So, obviously it also covers and partitions A like this, which means that probability of A by the axiom of P map is sigma probability of A intersection B i. Now probability of intersection by the definition of conditional probability we have seen is the following, so that is the proof. So, this is how you break up probability into sum of probabilities by using a partition. And this is also equal to probability of A intersection B I, maybe you just need that, so I am including it explicitly.

But the interesting part is that conditional probabilities involved finally. So, in applications or practical examples, you have to come up with a good partition. There will be many partitions possible you have to pick the one that gives you the best proof, simplest proof.

(Refer Slide Time: 18:02)



So, it is actually a design problem. So, you should try to find a good partitioning scheme, not every partition is good for your problem, for your application. So, let us see an example. Suppose you are given n sticks have distinct lengths, they are in distinct lengths, and you have to fix them in n holes, they are arranged in a line. So, you have these holes and you randomly arrange the sticks in these holes, one in each.

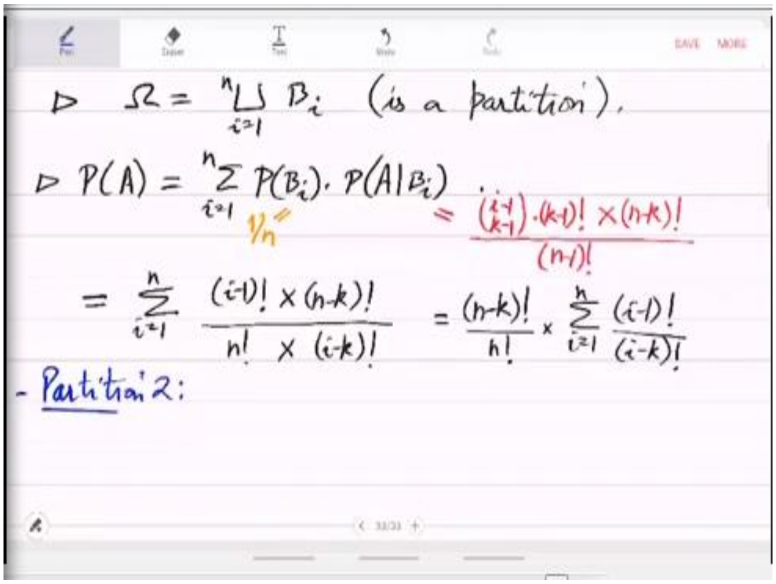
And somebody will look from let us say from the left. So, you are watching from the left in front of your eye or these holes, second one behind first and third one behind second and so on. And then sticks have been placed in these holes, so that is the scenario. It is a random arrangement of sticks, what is the probability we are interested in? So, we are interested in the following probability given K? What is the probability that the kth hole stick?

That the stick in the kth hole is visible, simple enough. So, you are interested in, for example in this picture there are 4 holes, 4 sticks, you randomly put the sticks these 4 sticks, there are 4 factorial ways, there are 24 ways. You can place them as you want and what is the probability? That let us say stick in the third hole is visible from the left, which means that, so that will be possible only first and second sticks are smaller than the third.

If any of those two is higher, then you cannot see the third stick. So, you will always be able to see the first stick but then the second one there is a condition, third one there is a stronger condition and so on. So, what is this probability? Now here I will show you the problem and the advantage of partitioning. So, let us define omega, omega is the permutations on n sticks, in fact all permutations.

You can arrange the sticks, so let me give you the first partition, partition 1. So, you arrange the sticks in increasing order. And define B i to be permutations where ith stick is in kth hole. So, with this again partitions the sample space omega, this is easy to see.

(Refer Slide Time: 23:22)



So, there are n sticks and now the permutations look at B 1, B 2, so the permutations which put 1 first stick or this stick 1 which is basically the smallest length in the Kth hole and permutations that put second smallest stick in kth hole, there obviously different permutations, so the sticks are different. So, B 1, B 2 are disjoint and hence all the B i's are mutually disjoint and they cover omega, why?

Well, because any permutation in omega has to put one of the sticks in the kth hole, kth hole cannot be kept empty because it is n on n, it is a bijection. So, this is a partition, so can we utilise it? So, let us call this event A that we are interested in which is only those permutations where kth whole stick is visible from the left. So, you get that probability of A is probability of B i times probability of A given B i. Now, what is probability of B i?

That is the ith stick goes in the kth hole, so it was a random process, ith stick at n option, so this is 1 by n. Now, what is the probability that kth hole is visible? The stick in kth hole is visible, when the ith stick went in the kth hole. So, remember that before i or overall there are only i - 1 sticks that are smaller than ith stick. So, those i - 1 those are the only ones which are allowed to go in holes number 1 to K - 1.

So, what you get is i - 1 choose k – 1, possibilities of these i - 1 going to first k - 1 holes. But you do not care about permutation, their permutations. So, you multiply it by k - 1 factorial times beyond the kth hole which is n - k factorial, that is the favourable cases. So, to compute the probability you have to count all the permutations. So, that is, is it n factorial? No, because you have already put ith in the kth hole, so it is the remaining n - 1 factorial.

So, what you get is this expression, so it is sum i to 1, i - 1 factorial times n - k factorial divided by n factorial times i - k factorial that is the simplification. Which is n - k factorial divided by n factorial times this expression, this is what you have to calculate. Maybe you will see a way to calculate this and simplify this but a priori it looks complicated. So, in the next class what we will do is, we will partition in a different way so that we get a much simpler calculation and a much simpler value directly. So, that we will do tomorrow.