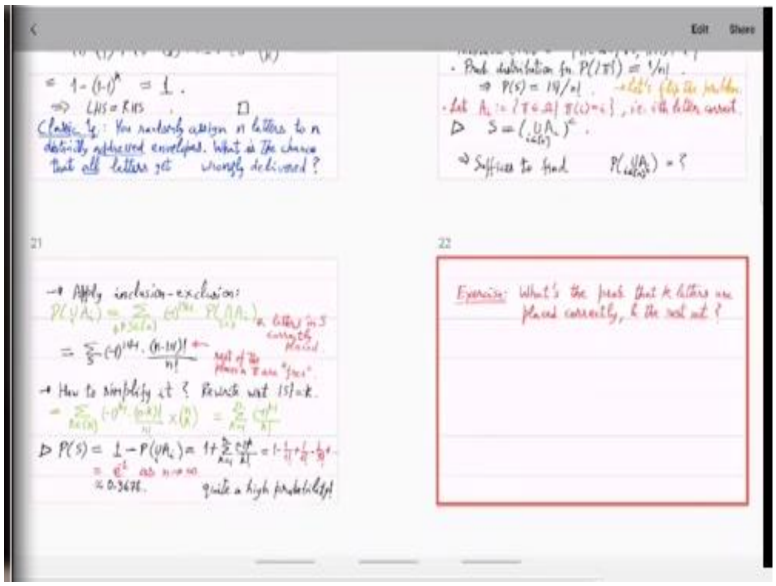
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science & Engineering
Indian Institute of Technology-Kanpur

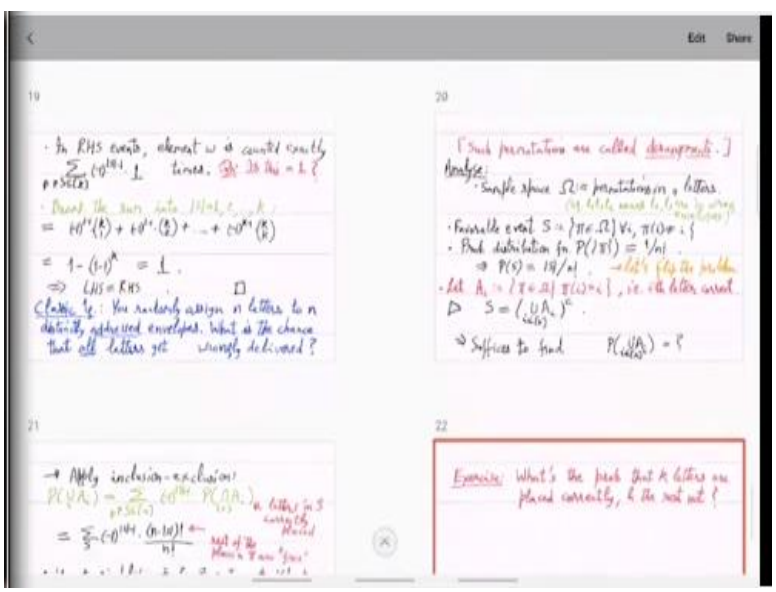
Lecture-05
Probability Over Infinite Space

(Refer Slide Time: 00:17)



Last time we did the basics of probability. And you saw the proof of inclusion, exclusion principle which is probability of union.

(Refer Slide Time: 00:28)

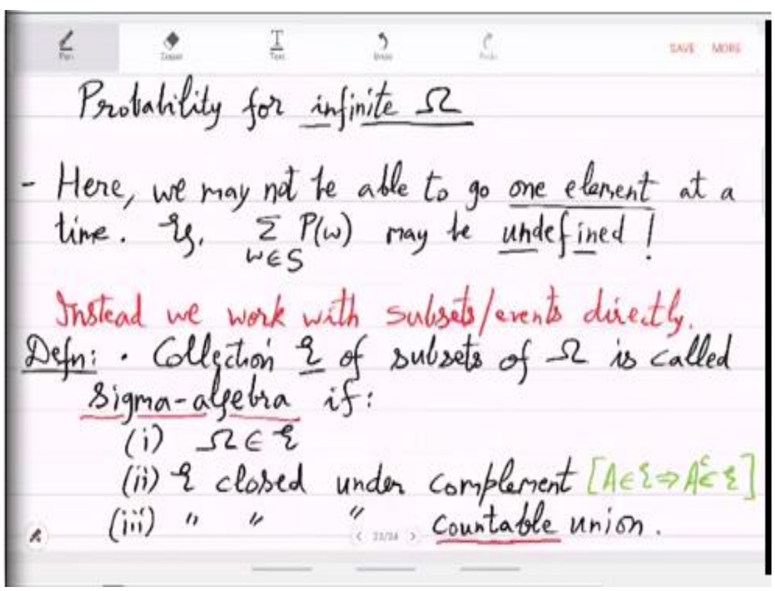


And then you saw this very nice classical example of counting derangements or the probability that all the letters will be wrongly delivered.

(Refer Slide Time: 00:38)

So, today we will start this one formalisation that we postponed which is what is the meaning of probability in an infinite space. So, let us start that.

(Refer Slide Time: 00:48)



So, probability for infinite omega, infinite sample space. So, the thing is here we may not be able to do a sum of probabilities because sum may not always be defined, if it is an infinite, sum of infinitely many things then you get into convergence issues. So, we may not be able to go one element at a time. And the reason is this sigma P omega, this way of doing things may not work, specially when you are doing it for a subset, this may be undefined, that is the problem.

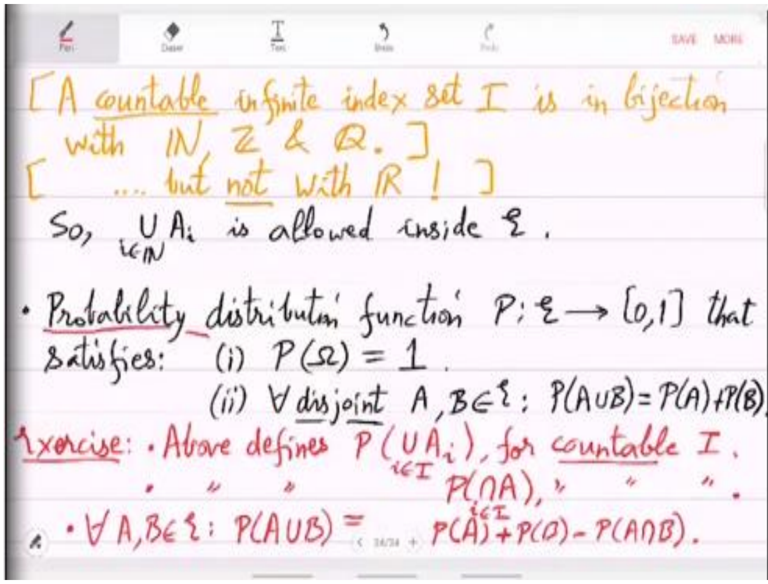
So, instead what we will do is, we will work at the level of subsets, we will define probability directly for subsets and not for elements. So, instead we work with subsets which is event directly. So, how do you do that? So, for that we have to define what kind of subsets are allowed? So, that is sigma algebra. So, collection E of events or subsets is called a sigma algebra, if it has nice closure properties.

So, since we are not working with elements but subsets, we still want the subsets to contain some information about the elements within. So, in particular you want to take intersection, union, complement those kinds of things. And then the probability map or operator should extend suitably. So, you want full space to be in E omega should be in E, you want closed under complement, which means what is closure?

Closure simply means that if A is in E then a compliment should also be in E, that is what closure means. And finally E should be closed under union, but remember that now you are talking about infinitely many subsets. So, what is union? How many subsets could you take the union of? So, we will want to work with good kinds of infinity which is countably many subsets, so countable union.

So, what is this word countable? Countable simply means that the number of subsets that you are taking union of, that number or that index set should be in bisection with or should not be bigger than let us say natural numbers or integers. So, it is like 1, 2, 3, 4, so first subset, second subset, third subset and so on, that kind of union you are allowed. But not something which is bigger than this which is not in bisection with 1, 2, 3, 4, natural numbers. So, let me mention that separately.

(Refer Slide Time: 05:33)



So, in case you do not know a countable infinite set is always in bisection with natural numbers and also in fact integers, and also with rationals. So, that you can take also as the definition of countably infinite, it is like natural numbers integers or rationals but not with reals. So, reals is not in bisection with natural numbers, real is something far more, it is actually like subsets of natural numbers, but we will not go too deep into that.

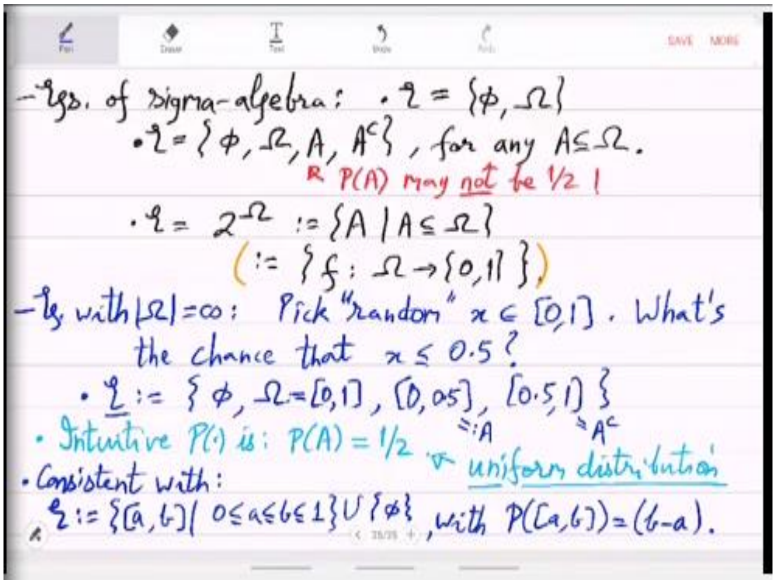
So, it suffices to say that you are allowed to take union, fed the index set is national numbers, that is the intuitive thing. So, union of subsets, natural numbers is allowed, inside E, so that is what is meant. So, such a collection is called a sigma algebra, and now once you have a sigma algebra you can define the probability map. So, next thing is probability distribution function P is a map from this sigma algebra to the real line reflect the segment 0 to 1.

Obviously which satisfies, so to define probability, remember. So, what it should satisfy is that probability of the full sample space should be 1. And since you want the probability to also extend to other subsets, what you would like is? It should respect somehow the sigma algebra operations in particular union. So, for all A, B that are disjoint, probability of union is the sum, and this is enough to extend P to the whole of sigma algebra.

You define it over disjoint and then this will help you together with the countable union to extend to any subset because E is a sigma algebra. So, that I leave as an exercise. This is the meaning of probability over infinite spaces. So, above defines probability of A i for any index set that is countable. And second is the same thing for intersection, so defines for union, countable union, countable intersection and other kinds of union also.

For example, you can recover the union formula by subtracting intersection. So, all these things could be deduced. So, these things should be intuitively clear, what we have done is just given you the formal definition of probability distribution over infinite sample space. The idea being moved from elements to subsets, do not talk about probability of an element. That in fact may not even be defined, but what is defined is probability of a subset of an event. So, what are the examples of sigma algebra?

(Refer Slide Time: 11:19)



These are the immediate examples. So, if you just take empty set and everything that is a sigma algebra, it is a trivial kind of a sigma algebra. It has omega it is closed under countable union and complement because for the 3 conditions. Then you can extend this; make it slightly bigger by adding a subset then you have to add it is complement. So, this you can do for any subset.

Now remember that this probability may not be half, like just A and it is complement being the subsets in your sigma algebra does not mean that they are equally probable, that it may also be one third, two third probability. So, that will depend on the problem definition, how you define

P? And there is another trivial example, which is just all subsets. You can also take all subsets of omega, so which is E such that A is a subset of omega, all of them.

And which is also kind of functions Boolean valued. So, you can look at the events as either subsets of the sample space or you can think of them as 0, 1 valued functions. But then the sample space will be that of functions as well. But this I should that is kind of a diversion, the important definition is the first one. Another example slightly more interesting example, infinite sample space example is suppose you pick a random number, pick.

So, physically you are picking a random number x between 0, 1. And you are interested in the probability or chance that it is smaller than 0.5, you are interested in this probability. So, what is the probability distribution function for this with the sample space? So, the way you can model it is, define your sigma algebra to be empty set, everything which is 0 to 1 and the subset of interest which is 0 to 0.5 and the complement of it.

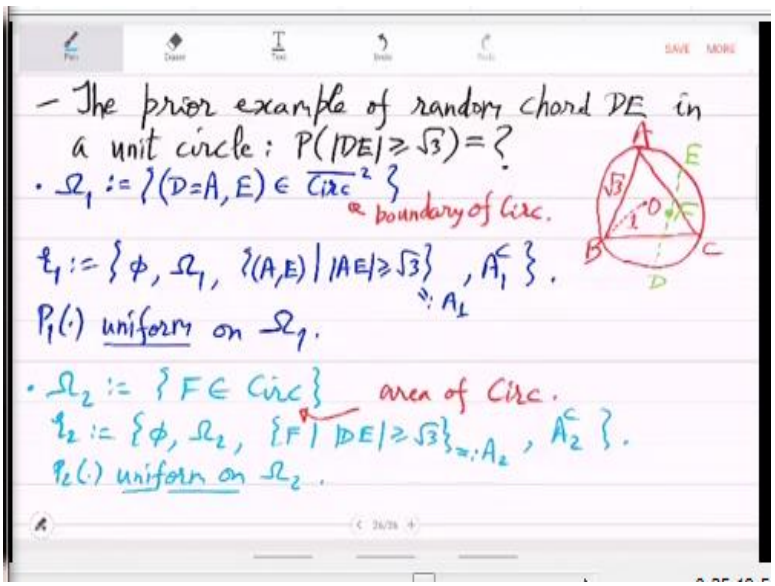
So, let us call this E and then this is A compliment. So, the sigma algebra only these 4 elements, everything nothing and subset of interest event of interest and it is complement these 4 things. And intuitively the probability of A should be attached half, because it is length is half of the sample space. So, intuitively you define like that. So, intuitive P is with sets probability of A to B half.

But you could have given this A any probability, that also will be a valid probability map. For example, you can assign probability of A to be two thirds or one thirds and then the other one will be two thirds that is also a valid probability map. But the intuitive one is this, and this is called uniform distribution. So, the uniform probability distribution map or function is the one that gives an interval probability value difference of the endpoints, the length of the interval.

So, we can in fact, then define E to be something like this. So, it is also consistent with, and E where you have interval a, b for every possible a b. And also you have to put obviously the empty interval with probability of the interval, the difference. So, this is how you can make sense of picking a random number x in the interval 0 to 1. And saying that the probability it is utmost 0.5 is half, 50% chance.

So, you make sense of it by saying that ok, my sigma algebra is this and my probability distribution function is this consistent with uniform distribution. So, I can mathematically say that the probability is half for getting 0 to 0.5. So, an intuitive thing can be explained mathematically like this, in fact this even models reality.

(Refer Slide Time: 18:40)



So, in the same spirit, we can go back and explain the random chord example that we did last week. So, the prior example for random chord DE in a unit circle, so recall that like picture was like this, you have an equilateral triangle, centred at O and the circumscribing circle is also drawn. We have assumed the radius to be 1 and hence the side is root square root 3, it is in trigonometry you can show that, and the random chord is this DE with centre F.

So, what is the chance? Question was what is the chance that this random DE chord has length more than square root 3? So, the probability that the chord length is at least square root 3, what is this? And in fact, since it is an infinite sample space, how do we make sense of this? That there are infinitely many possibilities of picking the chord, even picking the point D. So, we got 3 different answers by using 3 different experiments.

So, I will just quickly explain to you the formalisation of each of those 3. So, the first was you model your sample space as, take your D to be A and E to be random on the circumference. So, the border of the circle squared, so this is the boundary of the circle. So, on the boundary, you are thinking in terms of picking a random E, taking D could have been anything.

So, without loss of generality, we fixed it to be A, and then E is what is being picked from the boundary of the circle. So, these are all possible such chords. And, so this is the sample space and the relevant sigma algebra that you have to define is contains empty set, sample space. And those chords which are long, those are the favourable chords. So, let us call this set A 1 and A 1 compliment.

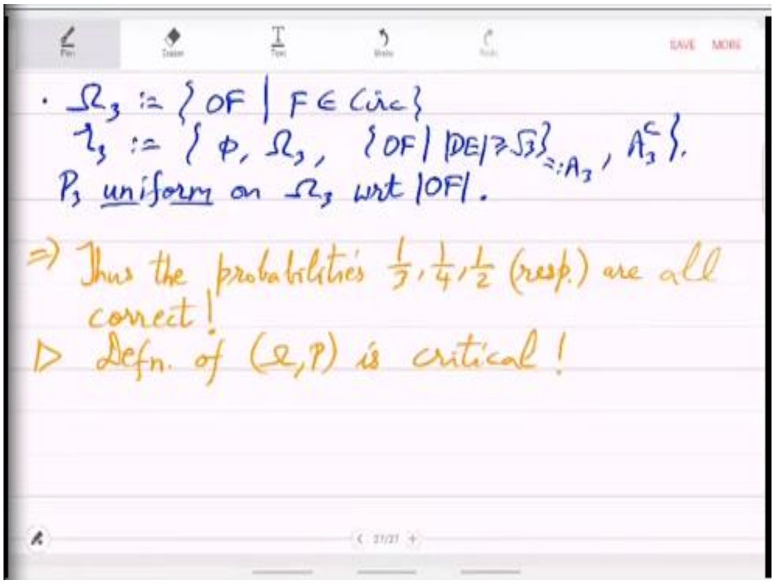
And P 1 you take to the uniform on omega 1, so uniform on omega 1 means that the only thing that is changing is E. So, you take it to be everywhere will fit the same chance, whatever that means, mathematically you will just define P 1 to be on an arc, what is the chance of point being on an arc? So, length of the arc over 2 pi, the circumference of the circle, that will be the probability.

And so this is enough to give you the probability of A 1 which I think you got one third. Second way to model this mathematically and rigorously is to take omega 2 to be this collection of F, so F on the circle. So, the circle is not just boundary but everything, so this is area of the circle. So, and then in this case the sigma algebra that we will pick is the empty set, everything, points F such that DE is long, so let us call this A 2 and it is complement.

And then the probability map will be uniform on omega 2. So, uniform on omega 2 means that this point F can be anywhere in the area of the circles. So, if you restricted to some zone, then the area of that zone over pi which is the area of the circle is the probability, that is what is meant. So, this I think you probability one fourth, probability of A 2 came out to be one fourth, again it is completely correct in this formalism.

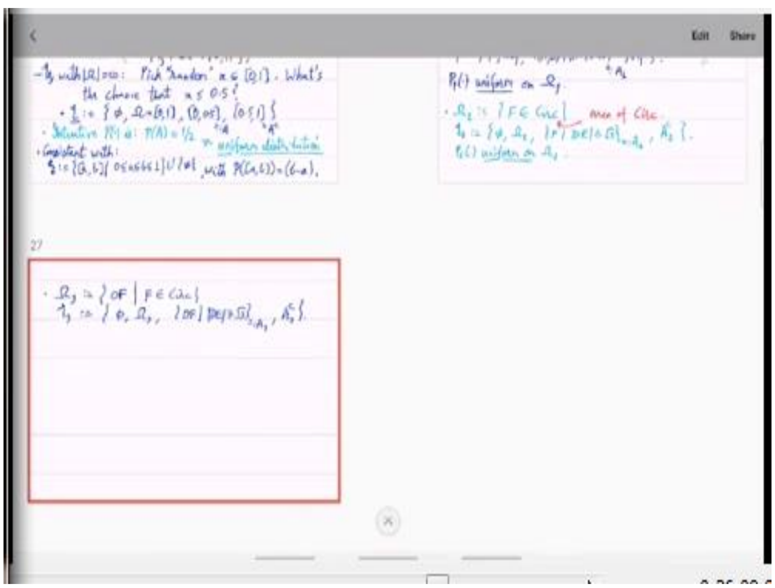
So, P 1 and P 2 are different probability distributions, that was the point, so both of them hold true. And then the third one was looking at the length of OF, that is the third kind of formalism.

(Refer Slide Time: 25:20)



So, you can take sample space to be the segment OF centre of the circle to centre of the chord, F everywhere in the circle. And the sigma algebra you can define as empty set, sample space, OF with DE long, let us call this A 3 and it is complement.

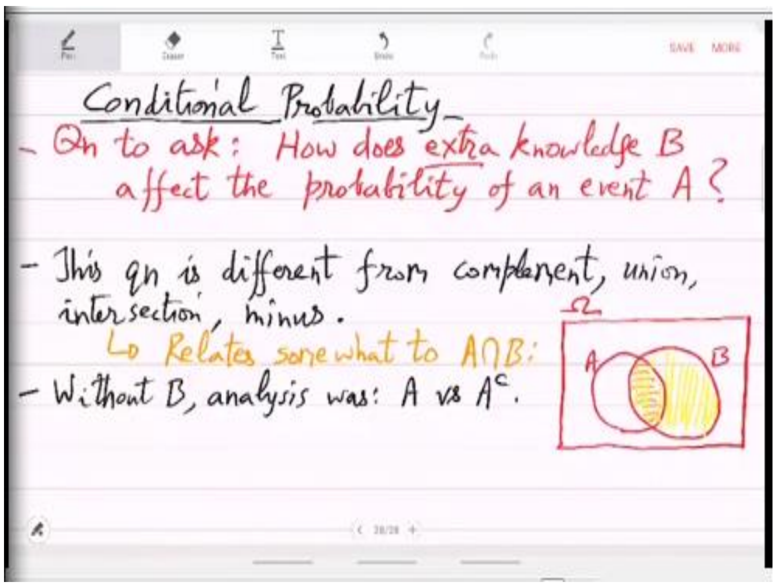
(Refer Slide Time: 26:08)



Similar to the definition before and of course P 3 is now uniform on omega 3 with respect to length. So, the segment you can take all lengths with kind of equal chance and also all directions with equal chance. So, it ultimately boils down to the length which the favourable length was half or less, that is you got probability half there. So, one third, one fourth and one by two, these were the 3 possibilities, they are all correct depends on which probability distribution you are looking at P 1, P 2 or P 3.

Thus the probability, respectively are all correct, and why are they all correct for the same event? Well, because it is not really the same event, it is happening in different probability distribution, so omega P is critical. So, definition of omega P is critical, that is the reason that is all you need to define probability over discrete space or over continuous space. So, next what we will do is a very important you can say operator on events or you can say type of probability, which is called conditional probability.

(Refer Slide Time: 28:14)



So, here the question you ask is how does extra knowledge help you? Extra knowledge B affect the probability of an event A. So, the question you are asking is what is the probability of event A? And in a different scenario, if you had prior information about some other event B, what is the probability of A? So, with the knowledge and without the knowledge of B, how does the probability of A change?

So, this relative probability or conditional probability is a very natural question and it has huge implications, in fact as you will see later huge practical implications as well. So, this is a question different from what we have studied till now. So, this question is different from complement, union, intersection minus etcetera that we saw. It is a different kind of operation that we are doing with A and B.

In all the other cases B was not certain, it was always either B complement or A union B or A intersection B or A minus B. But now we are seeing that certainly B has happened, and then A is yet to happen. But this is still relates somewhat intersection, why is that? Well, just consider the following picture. So, in this picture, this is the intersection, so B has already happened and what is the chance that A will happen?

So, B happening means that you are considering the shaded part, so in this yellow shaded part what is the chance of or what is the fraction of this orange shaded part? This is what you are interested in. So, it is somehow related to intersection but it is not intersection in the whole sample space but relative to B. And that can change everything, because then the question arises how big is B compared to omega?

If B is very small, then the conditional probability will should be very small, well, it can be it is not clear. Even if B is very small or B is very large, still you have to look at the intersection relatively, so it really depends on. So, without B the analysis was looking at A versus A compliment, with B it will be different.