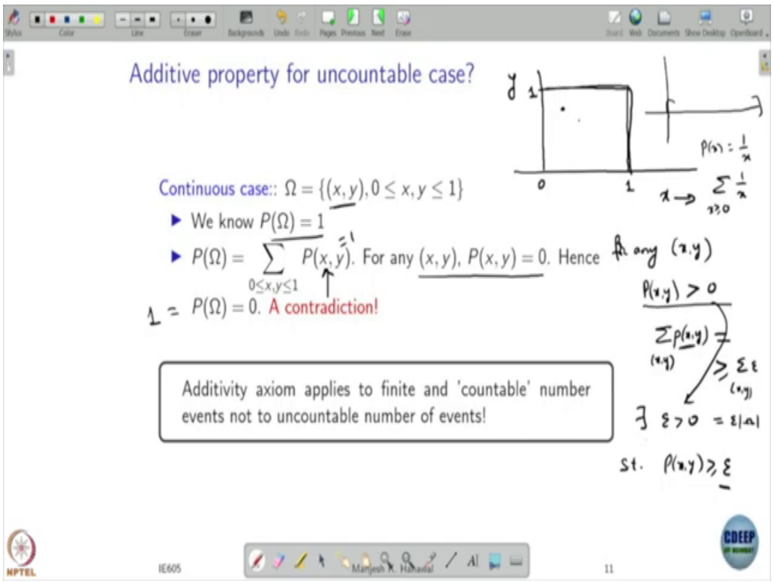
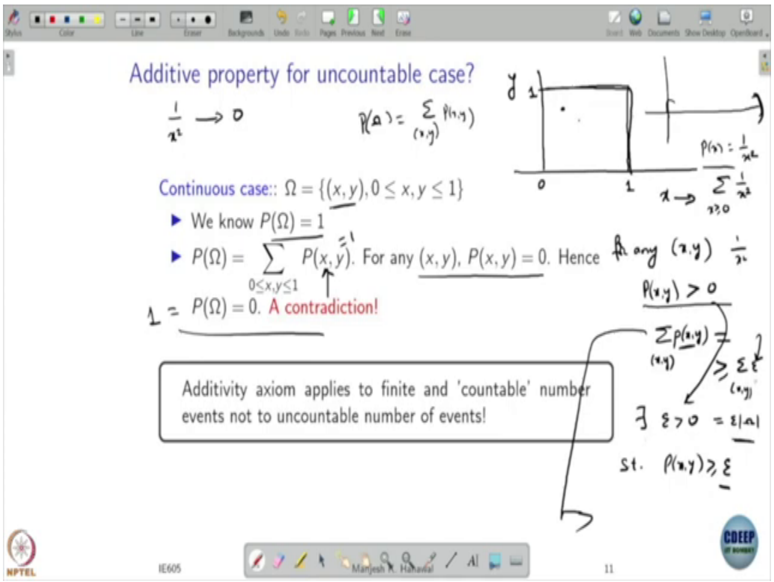
Engineering Statistics
Professor Manjesh Hanawal
Department Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Week 01- Lecture 03
Interpretation of Probability

You may actually end up uncountable case also let us look into an example.

(Refer Slide Time: 0:25)





Let us say this is my x axis and this is my y axis and I am interested in this region, 0 to 1, 0 to 1. So, just think of a geographical location which is like a square area, one-unit square area which is represented by x and y coordinates. Now, I am interested in this square area, this is my region and I am interested in let us say some particular point here and my omega is all x, y which are between 0 and 1 and also my second axiom said that probability of omega has to be 1. Then only it is a proper probability function otherwise it is not a probability function.

Now, let us look into this, suppose let us say in this region x and y for any x, y pair and that x, y you gave a strictly positive number, strictly positive number you assigned. Then if I am going to sum them all over all of them and everybody is strictly positive, remember and those who know real analysis, what does this mean? $P(x, y) \geq 0$, greater than 0 what does this mean? That means there exists an $\epsilon > 0$ such that $P(x, y) \geq \epsilon$. If this is the case what is this sum is going to be?

What summation of P of x, y or x, y is going to be? 1? Will it be 1? See how many x, y points are there?

Student is answering: Infinity.

Professor: Infinity and each point's probability is larger than epsilon, we are adding epsilon infinitely many times,

Student is answering: Summation (())(3:17).

Professor: Yeah, I mean, these are just all possible x, y's if you want to take it as an integration instead of summation, do an integration that is fine. But if you want to do integration of this P x, y here every point is greater than epsilon what is this value is going to be? Is it 1? Can this be 1? Can this be finite? Can the summation be finite if every point is greater than in, is going to be greater than epsilon here?

Let us do simple math further. So, for the time being do not worry about summation or integration, we can do. Now, if this is the case we know that this has to be greater than or equal to x, y epsilon, because every point here is greater than epsilon. And now you are adding this over all x, y, so now this value is nothing but epsilon times cardinality of omega. And what is the cardinality of omega?

That is going to be infinite, because there are how many points x, y are there which are between 0 and 1 grid, it is uncountably many actually. So, this becomes that if you are going to do like this this becomes actually infinity, this summation cannot be 1. So, as you see that if this guy, all this term has to be, if this guy has to be 1 that is it this value has to be finite many of the P(x, y) has to be 0, if they are not 0 then this you cannot guarantee this P of omega to be equals to 1.

So, this means, let us say now P(x, y) = 0 for many point, I mean take all of them to be 0, if that is the case if you are going to add P of x, y equals to 0 and add all of them, you see that right hand side is 0 but left hand side you want it to be 1. P of omega by axiom it should be 1, but now in this case if all P of x, y equals to 0 you will add infinitely many 0 is still 0, so, and you will end up with a contradiction.

So, this extension of additivity from countable to uncountable does not extend, just because of the fact that I mean many points in this space has to have a 0 likelihood value, they cannot all have positive likelihood and this is just for your reference do not worry much about that, I mean we are not going to use it, it is just to make it clear that the finite additivity, why it is extended to countable but why not to uncountably many terms, okay fine.

Student is questioning: what we said is there are infinitely many points and none of them is 0.

Professor: If none of them is 0…

Student: Yeah, if none of them is 0, the value cannot be finite, but. So, I can think of a simpler case where we take, instead of x and y we just take x, x tending from 0 to infinity and P(x) = 1/x. So, can we take the summation of that quantity that is the finite one.

Professor: P(x) = 1/x.

Student: None of the values are 0 over there as well and the values are infinite.

Professor: So, what is this example saying? So, suppose let us say P (x) is summation 1/ x is finite? Is it finite?

Student: Yeah.

Professor: Summation 1 by x, let us take it is to be integer only like here x is greater than or equal, instead of that just take x to be integers, is this finite? This is not.

Student: Let us take $1/x^2$.

Professor: Okay, let us take $1/x^2$, this is finite, you can make it a normal probability. Now, what we are saying here? Here in this interval for every point you have assigned a positive number, let us say this is $1/x$, I mean infinity, let us exclude the infinity part but open it interval infinity but everything else, so this is fine, every point has a mass. I am not talking about that case here, I am talking about the case where this axiom where popularity of omega has to be 1 whether that is going to be satisfied or not and probability of (())(8:45) is omega is nothing but the uncountably many additions over all x, y.

Student: So, in this case it could be 1.

Professor: This could be 1 here in this case.

Student: The values are uncountable, the x values 0 to infinity, those values are uncountable and each value has some finite mass.

Professor: So, finite mass here, yes.

Student: So, both the cases are same is the condition, the number of points are infinite and here also point is 0. So, over there it is coming to be 1. So how can you say that…

Professor: So, this has to hold for anything not for a particular example. I mean this, when we are talking about axiom this has to be generalized over all possible space.

Student: So, saying, in this example if we conclude that some of the lines have to be 0, this could be a wrong statement.

Professor: In this case countably many has to be 0.

Student: That is what I am asking, how we drive, how are we able to do that it has to be 0, it could be possible all these values have some mass but still…

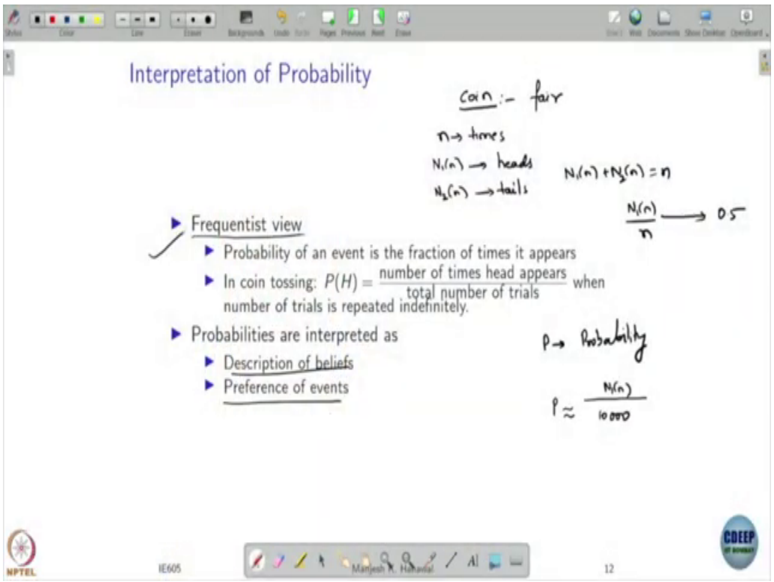
Professor: That is true that is possible, all of them can have some mass but now that is what I am saying like this is what we need to think about. Let us take if every point is positive then we are already saying that the summation cannot be anything finite it will be explored, this is what we are argued here.

Student: That is what I am not able to understand.

Professor: So, let us say in your case $1/x^2$, this is going to 0 as x tends to infinity and even those points which are tending to 0 are included in this. So, because of that you will not be able to find any epsilon positive and below which this value will not go, that epsilon here is arbitrary in this case, but where I mean what I am talking about is if everybody is positive with some fix epsilon this condition is violated, but the example you are giving is that epsilon itself is tending to 0.

So, we have to somehow be careful when we are dealing with uncountably many points, so that is why like I mean most of the arguments are going to make it for the finite case and wherever it is easy to extend it to the uncountable case will do but otherwise we will not get to that, for that is why like people have built nice sophisticated theories in terms of measures and all I mean those things will study in advanced classes.

(Refer Slide Time: 11:46)



Next, how to interpret probability. I talked about one thing like likelihood, like if you have a fair coin it is the likelihood of head happening is the same as tail, so I will assign equal values to them, that is based on your intuition, but is there any other interpretation of this probabilities? Other interpretations, so the one we talked about is based on description or preferences, there are other interpretations of something called frequentist view.

So, let me ask this question to understand this frequentist view, suppose let us say you have a coin, a fair coin and you are going to throw it n times, and out of this n times let us say you are going to see these many times heads and these many times tales and naturally N1(n) plus N2(n) has to be equals to small n, this is the total number of tosses you have made out of which N1 is.

Now, let us look into this, N1 by n of n. So, what you are doing in this? Out of n trials you are trying to compute what fraction of the time out of this n outcomes what fraction of them are corresponding to heads. Now, what do you expect this value to go if I increase n to larger and larger number? Tends to what? 0.5. Why is that?

Your understanding is if the coin is fair in the large number of trials it must have append equal number of, I mean it should its fraction should be almost half, so that is what like one could think probability as the fraction of the time it is going to happen when the experiment is done repeatedly and repeatedly means you continue till infinity.

So, the frequentist view is exactly like I can define let us say P is the probability of heads. How can I do this? You toss the coin again and again and again and again and see you throw it many times let us say 10000 times and see that out of that how many times it came head let us call that N1 of n we said and you take that as your probability.

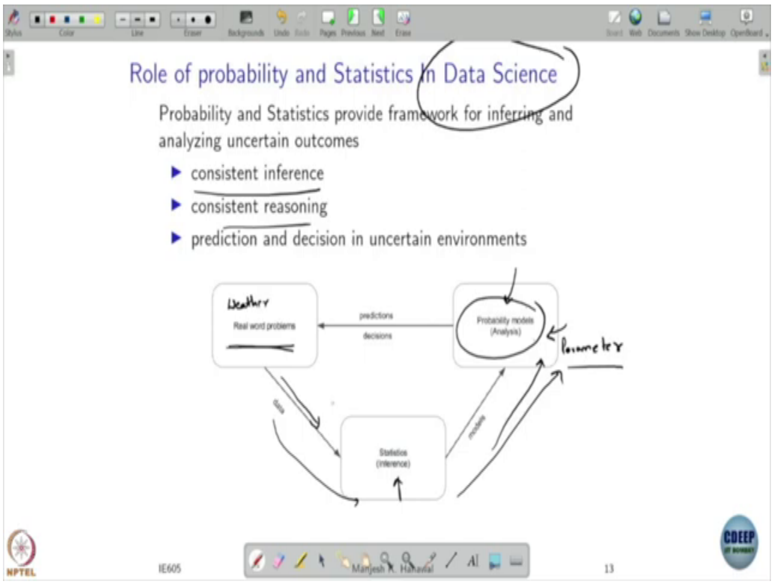
And this frequentist test is what is very handy let us say, suppose let us say in, we have all these drugs that have been come up which have been through various trials and they said that the efficacy of my drug is 86 percent or my efficacy of my drug is 95 percent. What does that mean how did they come up this number? It is not like they just put its likelihoods on that, they have to come up with these numbers. How did they come up with this?

Student: Observe data and put it in inside…

Professor: So, like they would have asked many volunteers like when you are go for a drug I mean evaluation you will ask volunteers or pay them to undergo trials, you collect a good number of people, you give this drug and say that on how many this drug is affective and then you take that on how many it is affected divided by the total number of people you administer this drug that can be taken as your probability of that effectiveness of your drug are curing that.

And that is why it like whenever you have data mostly you are going to go with this frequentist view and when you are actually not dealing with it, like you have to just simply model some probabilities, some probabilistic phenomena and you have just a belief about some probabilities then maybe you just to go and put these beliefs and try to see how things are.

(Refer Slide Time: 17:02)



Now, we are going to deal with three things as I said there is an underlying probability model, we are going to observe that probability model through the, through which the data is generated and we want to see that can we imitate that underlying probability model so that we can also generate the data in the same way that guy is generating.

If we could do that then basically we have understood that system, we are basically able to like see what potential it could have done and infer and that is why the role of the probability, statistics and data comes into picture and that is what is now called it as a data science, the interplay between these two. So, what happens in a, how they are used?

So, probability and statistics these two things kind of provide us a framework to understand the underlying phenomena and then make what we call it as consistent inference and also consistent reasoning and predict and make. So, how is that? Let us talk about some real-world problems, let us say this could be like the weather and in the weather, I am interested in predicting whether it is going to rain tomorrow or not, I am interested in that.

So, if I know how the weather affects, like how the rain comes based on what factors everything then I kind of myself build a probability model and then predict whether the rain is going to come based on that. But the weather is very complex, like I cannot simply understand everything based on which some weather phenomena happens.

So, what we actually do is, this is real world weather but we try to model with according to some probability model, this is our model, so this is reality, this is our thing model is ours, the reality is the real world problem. So, now what we are trying to do is we are trying to understand this real-world problem using some probability model and when we try to build some probability model I will say that this will come with some parameters, probability whenever we are going to do models, the models will come with some associated parameters.

Now, just assume that real world problem is also another probability but it is going to use some particular parameters and I do not know those parameters if I know those parameters maybe my real-world model is as good as, my probability model is as good as the real-world model, it is just like maybe I do not know those particular parameters that is using.

But then what is good is I can get the data from this real word problem, like whether I observe whether it is going to rain tomorrow or not and I have like that I have been the, I have recorded whether it rained or not in the past 10 years on each of these days, so I have this data.

I can use this data, try to do the inference. What could be the possible parameters that are going to explain this data and how to effectively identify and estimate those parameters? All these things, these statistics are going to tell me how to do that. And when this use those parameters you go back and use them in your probability model and see that whatever probability model this is your build model is going to tell like whether it is going to rain or not.

Then you actually observe what happened and then you know that whether what you predicted and what data set they are the same or not, if same good if not you note that you made a mistake, your probability model is not good enough it needs to be improved. So, you take that and you use again further statistics method to improve the parameters and go back and improve your probability model.

You try to continuously try to do this so that your probability model tries to give as good as predictions where your predictions are. You are making fewer and fewer errors in your prediction that means you have tried to understand the real world problems and your probability model is trying to capture that well.

So, this is the interplay between all the things and I mean if you just think of the forecasting weather forecast this is exactly happening and people have built, could have been collecting this data by putting sensors at various places and based on that they estimate the parameters and use that parameters to use in certain probability models based on that they predict whether it is going towards the weather is going to be and then they will see that what actually happened, get a new data and for that they will try to improve the parameters and discontinues.

So, that is why it is very important that we understand probability models and see how to use the data to make interference and get a good parameter for them and improve this probability model so that they capture the real-world problems well. I think we have already exceeded time, we will stop here.