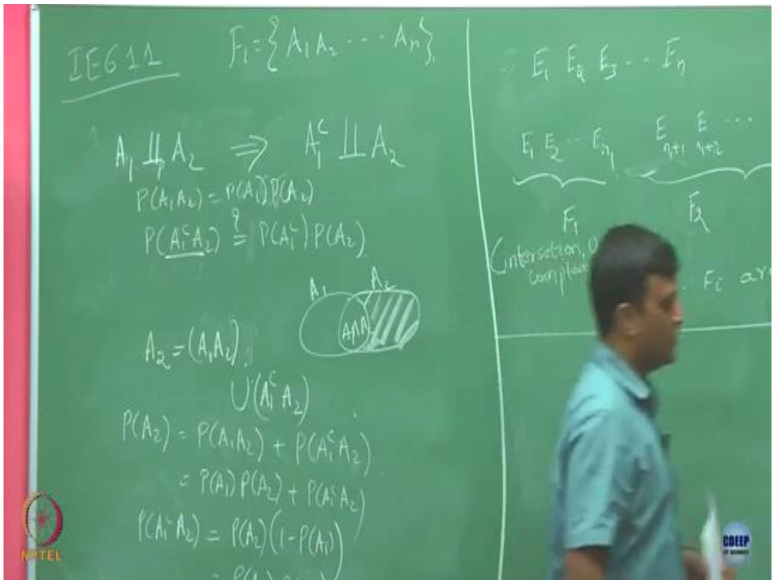
**Introduction to Stochastic Processes**
**Professor Manjesh Hanawal**
**Industrial Engineering & Operations Research**
**Indian Institute of Technology, Bombay**
**Lecture No. 04**
**Baye's Theorem and Introduction to Random Variables**

(Refer Time Slide: 01:04)



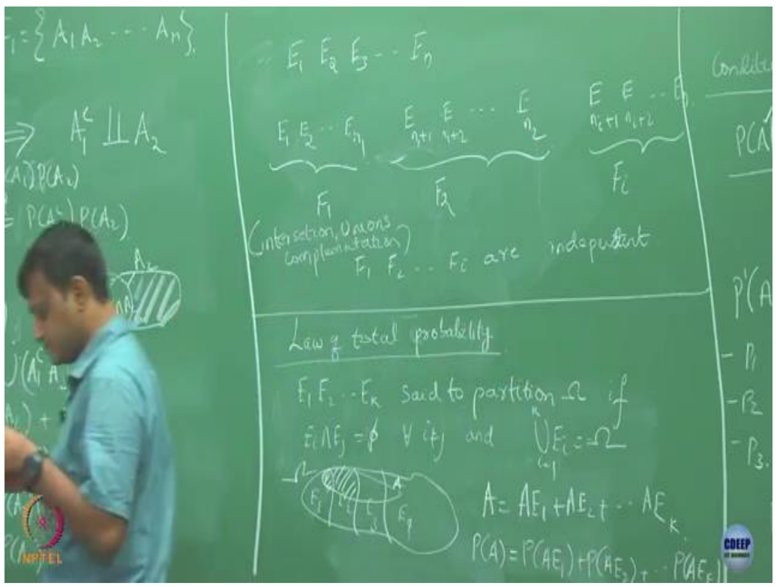
So, as of now I am simply defining these notions of independence, conditional probabilities and all, so, as we move on you will see that to certain results as we start analyzing the system, certain analysis become simpler if the independence conditions hold. So, as we move along we want to bit maybe model somewhat complex cases where it will involve many, many events.

But defining a probability for every joint so you have this script F, right? It will consist of so many events. Okay now, instead, now to define probabilities function, we need to apply given number to each of the elements in this. Whatever like I do not know how many events it (cont) let us say it contains n events. But suppose somehow this events turns out to be independent.

I only need to define probabilities for each of the elements in my sample space. For the others, I can just get multiply. For each if I take a set in this and let us say each of the elements in that are okay, if I pick a set, multiple sets and if they happen to be independent now if I have a probability for each of the sets, and they are happened to be independent, all I need to do is just multiply them. Otherwise, I need to define probability for each one of them. So, if there is an independent, my definition of probability becomes simpler.

All I need to define probabilities for the individual events from that I can derive for all the joint event. In that way. Independence brings lot of simplicity when I have to analyze a big chunk of events. For example, if you want to find the probability that that happened, this happened, that is going to happen, all these things. And if you know that they are independent. All you need to know is the probability of each of these terms. Okay, fine. So now moving on we are going to study this concept of total probability.

(Refer Slide Time: 03:04)



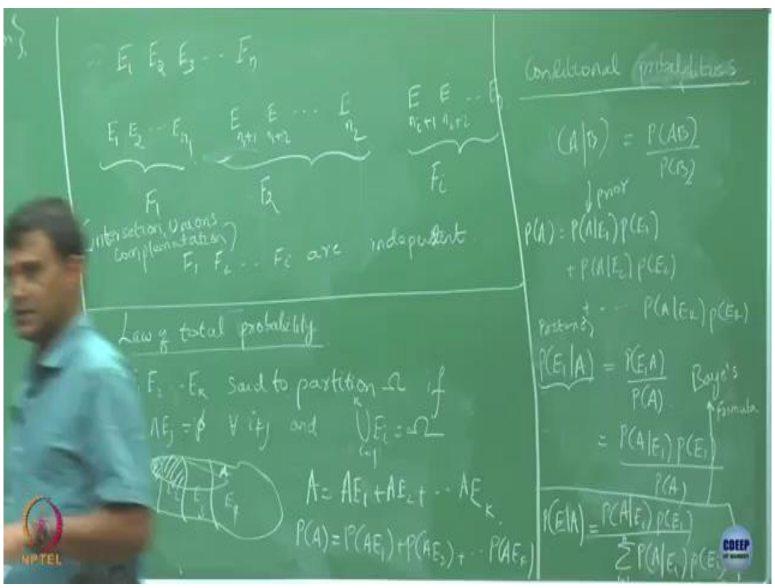

So, the basically the law of total Probability tells us that if I want to find probability of an event, we need to just look at its disjoint sets. Okay? And how we are going to find that disjoint set, by looking at some partition of my space and look at how much of these events falls in each of these partitions. So, let me make it more clear now. All of you understand what is a partition? Okay, so we are going to say E 1, E 2, E k said to partition omega if, and, so basically, let us say if you have an let us say this is your sample space omega, one possible partition could be like this simply E 1, E 2, E 3, E 4.

So, E 1, E 2, E 3, E 4 they are disjoint and they are such that their union covers my entire my sample space. Now, suppose let us say in this you have an event A, this is an event, which is a subset of my sample space. Now, if I want to find probability of this event, so do you think that if I know something about the probability of this E 1, E 2, E3 should help me? Because some portion of event F falls in E 4, some portion in E 3, some in E 2 and some in E 1.

So, if I know somehow calculate each of these portion, I should be able to compute the probability of the event A itself. So, that is what exactly the law of total probability says. So, if I have A, can I write A as A E 1 plus A E 2 plus A E k? Is this correct? So, for example, here A E 1 means this portion and A E 2 means this portion and this is true because my E 1, E 2 to all the way up to E k are disjoint, if they are not disjoint this is not necessarily correct. Now, are this A E 1, A E 2 themselves are disjoint?

Because E 1, E 2 are disjoint, it is necessary that A E 1, A E 2 all the way up to A E k themselves are disjoint. Now, can I apply if what is the probability of A is going to be, so probability of A is going to be probability of A E 1 plus probability of A E 2 all the way up to probability of A E k, why is that? And what property of probability I applied here? The third property, right? So, I have finitely many mutually exclusive events here, so their union should, the probability of their union is nothing but sum of their probabilities.

(Refer Slide Time: 8:32)




Now, I want to further use by definition of conditional probability here. So, can you tell me in terms of the conditional probability, how I can write this as?

Student is answering: (())(8:15).

Professor: Let us say probability of Ei is not equals to 0. Okay, so write it as a condition on the event, how we are going to write this? Okay, so let me write this for you guys. Now, think of A as my A, B as E 1. So, then I can write my this A as probability of A given E 1 into E 1 all the way up to probability of A given E k into.

Now, do anybody see any usefulness of this formula here? Fine, I have just manipulated my definition of conditional probability and express in this form. Do you think the way expressing it in this form should be of any use anywhere? Can anybody imagine why at all it is useful?

Student is answering: (())(9:43).

Professor: Yeah, it is not given. But suppose, assume that, you know, this probabilities of E 1, E 2, E 3, E 4, you have some information about them. So, you know that let us suppose you have omega, you know that the Omega can be partition like that you have further information on that and prior information. And you know, what is the probability of each of this partition?

In that way, do you think you can leverage that information to compute these probabilities? So, suppose, let us say I tell you these events E 1, E 2, E k partition. And I know this probability. Further, I know conditional probability of A on these partitions. These are partitions which I have defined at my convenience and whenever this condition that this partition that the person of A falling in each of this partition that is the probability of A given Ei, I know. I am, maybe let us say I can compute. Then using formula you are just, you can compute what is the probability of A. So, what is problem, what is this formula basically telling you?

Portion of A falling in E 1, portion of A falling in E 2, portion of A falling in E k, right? But before I take this but portion of a falling in E 1 then I have to also take the probability of E 1 happening itself, right. So, that is why, if I know what is the probability of each of these partitions happening and what is the likelihood that my event A happening condition that then if I going to take the sum of all these things that is going to give me the probability of A.

Okay, so, let us say, let us take a simple example. Okay, before I write an example let me give something called Baye's formula okay. So, here I have told you to compute P of A, you need to know that condition on E 14, what is the probability of A happening. But suppose, yeah this is what I have this probabilities E 1, E 2 I have defined a priori and also said you that I have somehow computed a probability of A given E, E 1, E 2 all this things, these are like prior probabilities that I have computed from this I got probability of A.

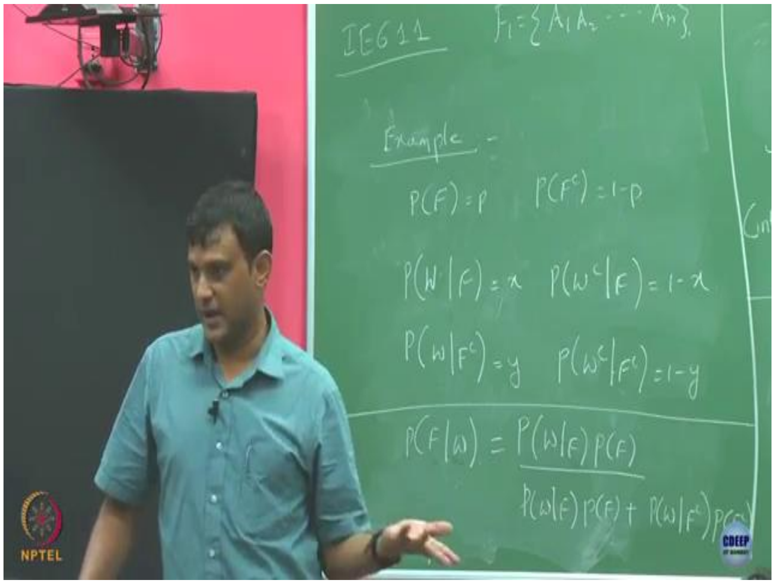
Suppose, now, I will say (prob), I ask the reverse question, event A has happened what is the probability that it might have been due to E 1? So, that is I want to ask E 1 given A, I am now asking basically posterior probability. So, these are my prior probabilities I have based on my partition, now, I think okay fine event A has happened that I have observed, now what is that it has this is due to E 1.

Now, can you express this in terms of this? So, let me call this as posterior probabilities and this as prior probabilities. Can you express my posteriors into the, in terms of the priors? How? So, this one I know already by definition of my conditional probability, this is probably of E 1 A divided by P of A. But this one I can further, again apply the definition of conditional probability to express it as P of A E 1 into P of E 1 and then P of A.

I have just applied the definition of my conditional probability. And now, this P of A, I am going to use my, this last total probability, which I have already derived, expressed in terms of my conditional probability. So, I am going to just write it as P of A 1 P of E 1 divided by summation of P of, right? And this is what we call it as Baye's formula. And this is indeed one of the important formulas we come across in analysis.

Let us try to see how, what is the usefulness of this formula. So what this formula is basically (())(15:58) is. The posterior probability based on my prior info probabilities, prior probabilities is something which I compute a priori, and how that is going to help me in computing my posterior probability.

(Refer Slide Time: 16:30)




So, let us see an example. Let us say you are working with a system and that that system has some critical component. Let us say it is your satellite that you want to launch and that satellites has some critical thing and you have built in NF redundancy in your system that if this critical component works fine everything should go normal, if this critical component fails, it is still possibility that your mission goes through even this go through. Now, the prior information you have try to compute is failure or success of this critical component.

Okay, there are any two possibilities, right? Failure or success of this critical component and given the system fails, what is the probability that your mission still complete successfully or not? Let us say you have that, you have computed this through your various lab experiments that you do extensively in your offline test moods. So, let us say you have that information.

So, when I say F this is the failure rate of my critical component I have. So, this is some very high pressure unit or whatever. Let us say, I test it in the lab environment, I say, I see that, when I put it into actual test mode with probability if P it fails and I have that information. And now, let us say I also have this suppose critical component fails. What is the probability that my mission still succeeds? So, let us call this or instead of that, let us say that critical component fails, my system still works.

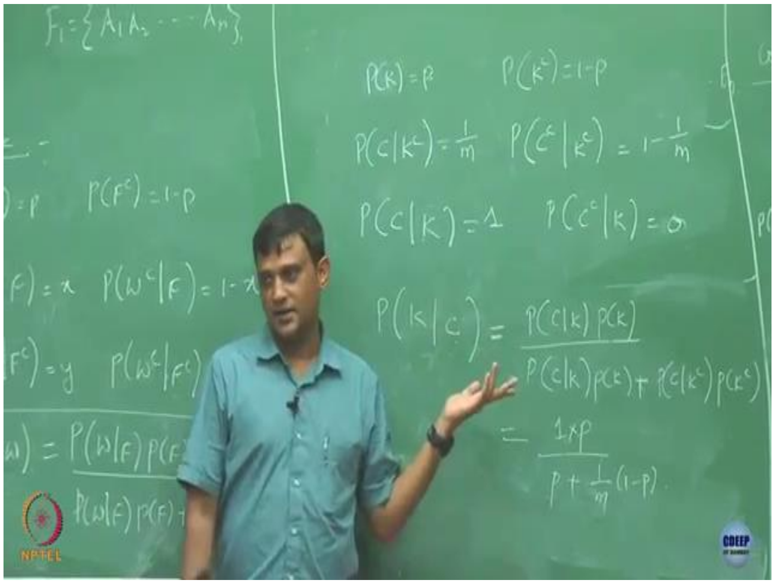
So that is what W I mean working condition. You also have completed it based on your lab environment like you based on your prior experiment. And let us call this probability y and let us call 1 minus x. And now, let us say you have probability that it works when your critical component fails. Let us say that is some y. And then probability that (())(19:25) does not fail is 1 minus y.

So, this is all you can compute right like before you actually launch your mission, you can do various experiments and try to compute and get these numbers. Now, based on this information, so these are all your prior probabilities that you could compute.

Now, based on that, let us say you want to compute the probability that you observed that your mission worked done, you are successful. But what was the probability that it was successful in spite of the fact that the critical component failed. So, this is the posterior probability maybe you want to compute like okay fine. You observe that your mission actually went through now, that is the condition what is that it went through in spite of that, you are critical component failed. Now, can you compute this probability based on your prior probabilities? What formula you are going to use? So, what is the Baye's theorem you are going to apply?

So, this I can write it as P of w by f into probability of f divided by P of w given f. Now, everything is expressed in terms of the things which you already computed beforehand. Do you know what is this probability of w given f? You know, what is P of f? Yes, I have already computed it. And I have already computed all these things. So, based on a prior profitability, I can go back and compute. Okay, what is the probability that my mission actually succeeds? Actually succeeded and in spite of there is like a failure in my critical component. Okay, let us say another simple example may be that you can relate a bit more.

(Refer Slide Time: 22:28)




So, let us say you have a question, for that question you either know the answer, or you do not know the answer. Okay, what is the probability? You know the answer with probability P and probability 1 minus p you do not know the answer. And what you guys do when you do not know the answer? You guess.

So, and when you guys guess, you guys know you are always not going to be lucky, right? So, you are going to be taking a risk. And that is why this probability comes. You need to see that what probability you are going to succeed and then let us say if you do not know the answer what is the probability that you end up? Guessing it correct. Let us say that probability is some 1 by m.

Okay, and then probability that you will not, when you do not know the answer, but your guess ended up in a wrong answer. Let us say that is going to be 1 by m. And then we can also compute, okay, what is when you know the answer. You are going to always make it correct, right you are going to get the correct one, let us say that probability is 1, and probability that c compliment is going to be 0.

Suppose, as an exam setter somebody setting multiple choice questions he derived these sets, he comes up with the statistics. And he wants to be fair, he want to make sure that he want to set a question paper in such a level that the guy who actually got the correct answer, he got it correct, because he knew the answer. It is not that he got it correct because he just guessed it smartly. He want to max, he wants to find out this probability.

So, that guy got the answer correctly, but he got it correct because he knew the answer. He want to compute this is my posterior probability I want to come, so if I want to compute based on this information, how I am going to do this? Can I do it at all? How? So, can you compute this? So, let us substitute these values.

So, if you know it correct, it is correct. Now, what is P of k? It is P and what is this? This is P plus what is C k complement 1 by m and what is this? This 1 minus p. So, let us try to understand like, how does this, suppose you increase P fix m, if you increase P this probability is going to increase or decrease? This function is an increasing function in P or decreasing function in P?

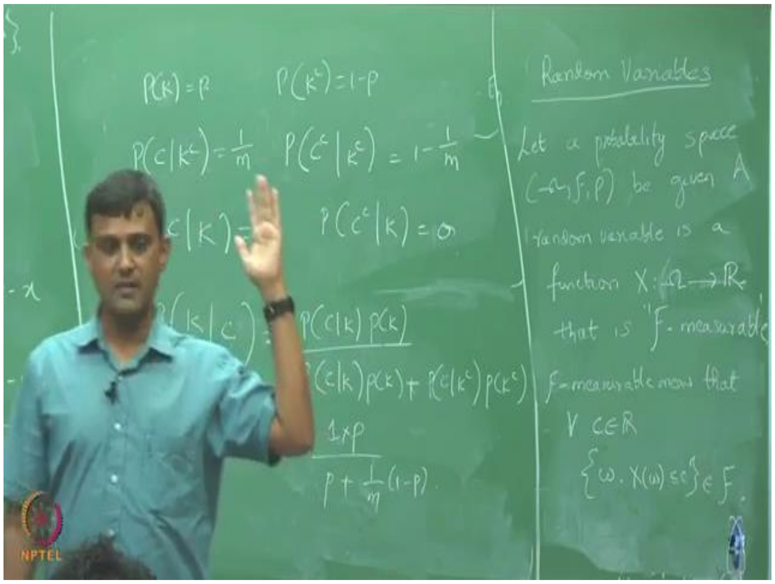
Student is answering: Increasing.

Professor: Increasing function. So it is good, right? Like if you know the answer, fine, the probability that you actually got the correct it is increasing. Now, let us say m here denotes the number of questions, the number of questions you have in your multiple choice. Now, suppose you want to be fair for your students, when I say fair, you want to make sure that if you got the correct you want the exam to be robust in the sense that if you got it correct, you got it correct because you knew it with high probability.

So, if you want to do that, and m is your choice, that is the number of choices you can give, so you want to increase m or decrease m? What is happening with increasing m? So, if you increase m, this is also increasing.

So, that means you are trying to make your exam paper more robust. So, as you can see here, like fine, like I may not be able to model thing exactly, but this gives me in a some sense to model some things and get some interpretation out of it. Okay. And you will see a (())(28:14) of applications where this Baye’s formula comes kicks in. Some of them we are going to see it in the assignments.

(Refer Slide Time: 28:35)




So, now after this I want to move on to the definition of random variables. So, what is random variable? It says something random, some value which is random, so in many experiments our outcomes will be kind of descriptive. For example, when you toss a coin, the outcome is like I expressed in descriptively, like either heads or tails or when you go to a casino, your outcome is just you won or loss that win and loss may be associated with some numbers also, right? Like you won because you got more than this many points, or you lost because you did not get this many points. So, instead of making this outcomes very objective, maybe it is better, we are better off by assigning some numbers to this objective, we actually did that already in the case of dice.

In the case of dice, we threw it and we say, outcome is 1, 2, 3, 4 like that, right? Those are the different possible outcomes, but we actually assign the numbers there, and in many, many experiments that you do in life. It is not necessarily that the outcome is always numbers. The outcome is some description. Okay, this happened, this happened among all these ten possibilities, like if we are going to say weather, outcome of the weather, outcome of the weather could be sunny, rainy, cloudy, whatever different options are there, right?

Instead of explaining them in this descriptive way, maybe you want to assign numbers. Okay? I mean, just make it uniform, less than this much temperature, we are going to call it humid, less than, above this temperature, we are going to call it sunny or something of that sort.

So, what basically random variable will do is, it will try to quantize the outcomes in terms of number, so that it become more amenable to our analysis. Okay, so let me first give a formal definition of what I mean by a random variable. So, to define a random variable, I need to have this underlying probability space. I only need basically the first two components omega and script F you will see, why? And this random variable is basically a function from your sample space to (())(32:41).

When I say R this is (())(32:43). That means for every possible outcome in your sample space, it is going to assign a real number. But that is it not, that it is not just a random variable, if it is a function like this further, it has to be F measurable. Now, what I mean by F measurable? F measurable says that are all C in R so this is the definite meaning of F measurability. It means you take any C, and if I ask the question the number of outcomes, the set of outcomes that are going to take value, possibly less than R equals to C is going this is going to define a set that set has to be belong to event space.

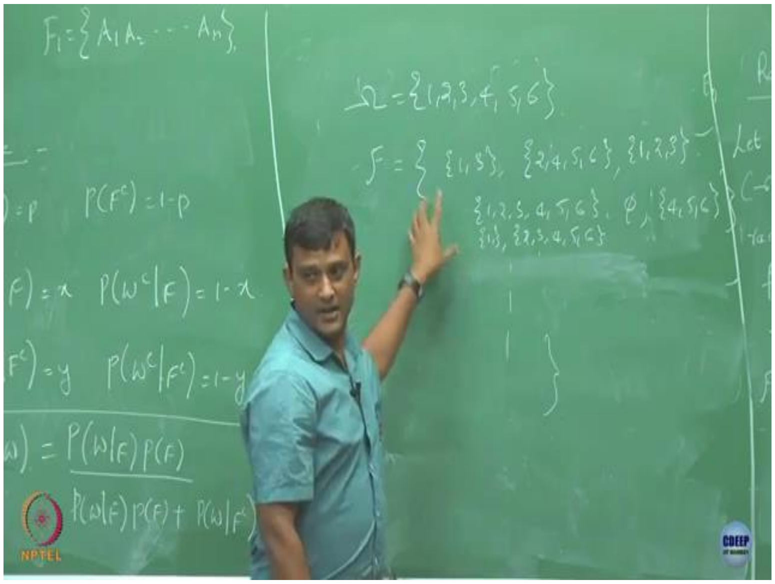
So, what it is basically asking for? I said X a random variable is basically function which is going to assign numbers to the outcomes. So, now I have mapped all my outcomes to numbers right. Now, in search experiment, I want to ask the question, what is the probability that my outcome of the experiment X value less than or equals to C? So, for example, in your case of dice throw, I could ask the question, what is the probability that outcome of the dice is less than or equal to 5?

Okay, in that case, this is exactly so that is set of all outcomes that are going to take value less than or equals to request to 5 and that event, whatever it is it has to belong to this script F and why is this? If I belongs script F then I know I can assign probabilities to that, right? Because my probability space is such that I am going to assign probabilities to only to the events that belong to event space. So that is why and why I am allowing, and notice that this condition need to be satisfied for every C, you may ask any question at this point.

What is that outcome of my experiment is going to be less than 10? You may ask what is the outcome of my experiment is going to be less than 10.003, whatever the number you are going to ask, for that. I want to be able to assign a probability and I could do that if that event belongs to my event space. That is why for any C I want this to be belong to my script F.

So, basically, what we are saying is a random variable is a function that maps my outcomes to real numbers. And it is such that for every possible real number, the set of outcomes that takes value less than that real number should be an event that lives in my event space.

(Refer Slide Time: 36:40)




So, let us look an example what is a random variable and what is not as a variable. So, let us take the simple case of our dice problem. In my dice throw problem what are the outcomes? Outcomes are, in this outcome, I am also, let us say I am also going to define my event space on this, take, let us take one special case. So, let us take I have 1, 3 and let us say I have 2, 4, 5, 6 and I have 1, 2, 3, 4 and I have phi. I have defined this to be my event space and let me first define a random variable on this, like this, which is going to assign to every outcome a real number.

In this case, what I am going to do, I am going to assign omega to be simply omega for all omega in some outcomes are 1, 2, 3, 4, 5, 6 I am going to define a random variable x, which is just going to give the same numbers as the outcome. In this case, it is (())(38:03) example because outcomes already in terms of the numbers, but it is not necessary that the outcomes need to be in terms of numbers. For example, head tail I already talked about. Okay, fine. I have just defined a map from omega to R.

Now, before I call it a random variable on this experiment, it has to be F measurable. Is it F measurable? Yes? Sure? So, how you are going to check F measurability here? I have given you the definition of F measurability here.

Student is answering: (())(38:56).

Professor: Why?

Student is answering: (())(39:03).

Professor: Say, let us apply this, let us take some C, C belong to R let me and this should be true for every C so I have a freedom to choose C, so let me choose C to be 3. Now, and what is that, in that case what is this set is going to be? 1, 2, 3 and now I had an event which consists of outcomes 1, 2, 3, does that belong to my of F here? No, right. So, this is not a random variable on this F.

But suppose, let us expand this. Let me include 1, 2, 3 and if I include 1, 2, 3 let me also include its complement 4, 5, 6. Now, for C equals to 3, this satisfies. Is it now still an F measurable? Okay, let us take as C equals to 1, C equals to 1 outcome is just 1, 1 does not belong to F. So, let me include that also. Now, is it F measurable? No, for it violates? It violates (())(40:33). So, let me, now let us include all possible values. How many subsets are there?

Student is answering: (())(40:41)

Professor: Let us say all subsets are there, all possible subsets are there. Now is it F measurable? Yes? So you take possible C, you will end up with one of this one and that belongs to F. So, it is not necessary that every function that we are going to define on your sample space is a random variable, it has to be F measurable that is because I can the very purpose of me defining a random variable, assigning numbers to the outcome is that I want to measure events.

And this is exactly doing that. And if this measurability condition fails, that random variable may be the way I applied defined my random variable is not appropriate one. Okay, fine. So, let us stop here then. You will see, so as you see, like, maybe you need to just digest this fact that any function on sample space need not be a random variable. It has to be appropriately satisfying my measurability conditions, okay.