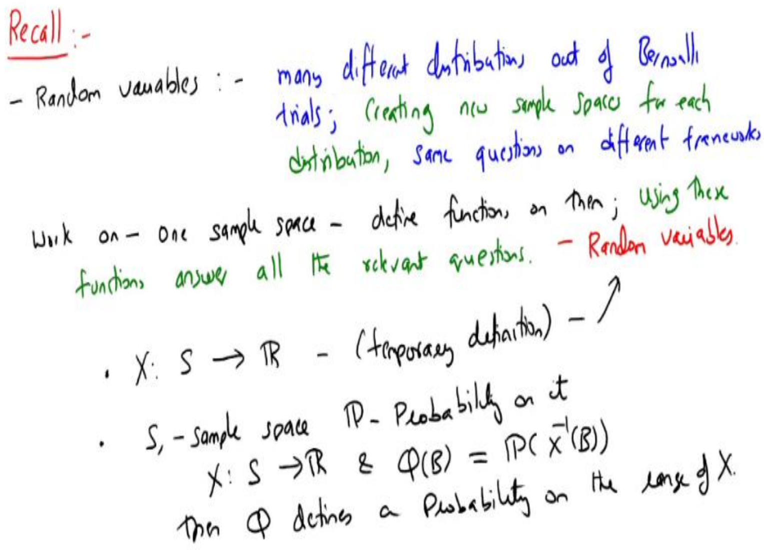
Introduction to Probability - With Examples Using R
Professor Siva Athreya
Theoretical Statistics and Mathematics Division
Indian Statistical Institute, Bangalore
Lecture 15
Discrete Random Variables - Part 02

(Refer Slide Time: 00:15)



So, we were doing random variables. And the reason we start random variable, is the following that that previously, we were doing many distributions, many different distributions. And they were all developed out of Bernoulli trials, out of Bernoulli trials. Then the key thing was, the key difficulty was we kept on creating new sample spaces. We kept on creating new sample spaces new sample spaces for each for each new distribution for each of them, each distribution. And once we did that, what happened was that, we were answering same questions, same questions on different frameworks. That was the, that was the process we were doing.

So, then we said no, no, we did not want to do this, we want to do, we want to define, we want to work on one sample space. So, then we want to work on one sample space. We want to work on, on one sample space, and then define functions on them, functions on them. And then using these functions, answer all the questions, using these functions answer all the questions, all the relevant questions. That is where we were that is where that is how we motivated the concept of need for random variables. That is how we started random variables.

There is various examples last time. So, the, the way we defined random variable was we said random variable is a function X from, so our first thing was random variable was the function X from from a sample space S to the real line, we will call any such function a random variable. It is the temporary definition. So, X is the function of random variable. That is what we called it as.

Then we showed another interesting fact that is if S is the sample space, P is a probability on it and X is a random variable from S to the real line. We showed the following fact you look at Q of an event B on the real line as the inverse image of events given by X, then Q defines a probability or not. Q defines a probability on the range of X and that is one thing we discussed last time.

(Refer Slide Time: 04:20)

functions answer

- $X: S \to \mathbb{R}$ - (temporary definition) -
- S, - sample space $\mathbb{P}$ - Probability on it

  $X: S \to \mathbb{R}$ & $Q(B) = \mathbb{P}(X^{-1}(B))$

  then Q defines a Probability on the range of X.
- $X: S \to T$ S - countable (finite) equipped with Probability $\mathbb{P}$ & T is countable subset of $\mathbb{R}$ then X is called as a Discrete Random variable.

$$f_X(t) = \mathbb{P}(X=t) \text{ - Probability mass function of } X.$$

$$\text{ie} \Rightarrow \mathbb{P}(X \in A) = \sum_{t \in A} f_X(t).$$

Then we came to the idea of what a discrete random variable was. So, we said our discrete random variable X is from S to T, where S is a countable space, countable or finite, equipped with probability, equipped with probability P and so is T, T is also countable, X is countable, T is, S is countable. T is countable subset of R of a real line then we called X as a X is called a discrete random variable.

The assumption that S is countable can be relaxed but T has to countable for sure, then is an important concept we discussed was this probability mass function of X. So, we look at this function fX of t as the chance that X is equal to t. This is called the probability mass function. That is where we were at. And then, one advantages was that is this implies that if you want to understand the chance that X is an event A we just have to do the summation over t and A f sub X of t. So, any event concerning X can be gotten from f.

So, this is where we were at last time. So, the idea was random variables required so that we could work on one sample space, we need not work on many, many sample spaces. And a discrete random variable is one when a range is discrete, that is range is countable subset of real line, and has an important parameter, that is probability, mass, function of X what fX of t was probability X equal to t and chance that X takes in event A is the sum of this.

(Refer Slide Time: 07:15)

Then Ψ ...

- $X : S \to T$ S - countable (finite) equipped with Probability $\mathbb{P}$ & T is countable subset of $\mathbb{R}$ then X is called as a Discrete Random variable.

$$f_X(t) = \mathbb{P}(X=t) \text{ - Probability mass function of } X$$

$$\text{ie} \Rightarrow \mathbb{P}(X \in A) = \sum_{t \in A} f_X(t).$$

- $X: S \to T$, $Y: S \to T$ be discrete random variables.

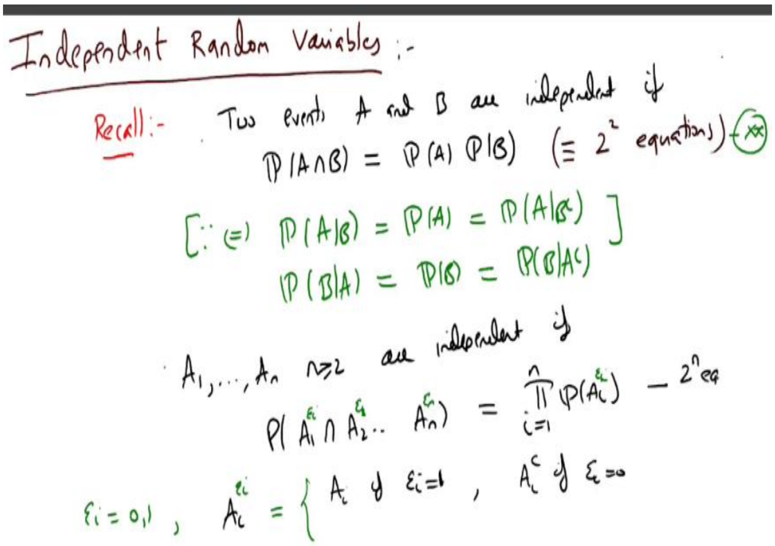
X and Y have equal distribution if

$$\underbrace{\mathbb{P}(X=t) = \mathbb{P}(Y=t)}_{\text{Same Probability mass function}} \quad \forall t \in T$$

And then we also said that if you have two random variables X and Y, so X is from S to T, Y is from S to T, and be that will just be discrete random variables. Then, we said that X and Y at equal distribution equal is defined this way, equal distribution, if the chance of X equal to t is the same as chance of Y equal to t, for all T. This is something that that one has to know. That is how you identify two random variables as being the same as, the distribution of them are the same.

So, the word distribution is signified that they have the same probability mass function. That means this is, they are the same probability mass function. They are, both have the same probability mass function. So, that is where we were at last time and I just began independence. So, let me just this time go back.

(Refer Slide Time: 09:13)



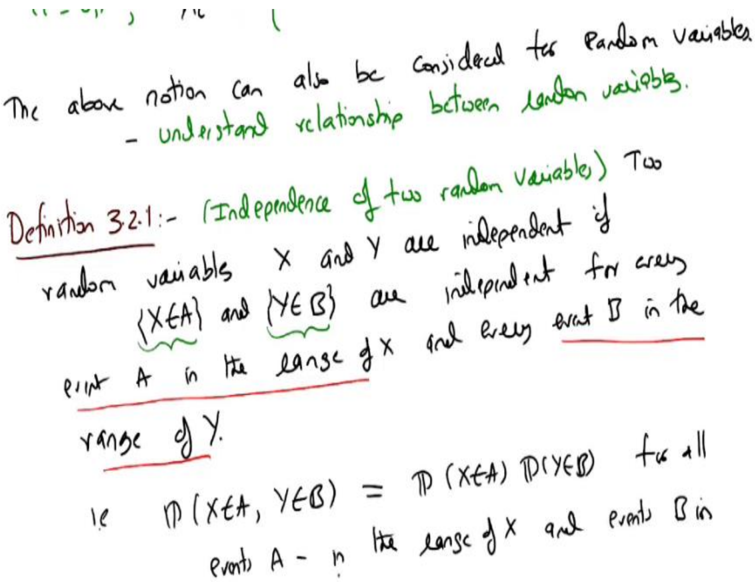
So, now in today's class, we will try and understand independent variables. So, independent random variables. So, independents are very used is a key concept in probability. It is the not the key defining feature problem. So, before we recall from before, recall from before, that we discussed the following idea, we said that two events A and B, A and B are independent, if the probability of A and B split up into probability of A time's probability.

And this is the definition of independence. And this kind of crucial idea and probability, because this gave you the following idea that this was kind of equivalent. It is kind of equivalent, this had actually, two to the two equations that is hidden inside it. And it gave, the observation we gave last time was because this actually was equivalent to the fact that the chance of A given B was the same as chance of A chance of A given B complement, give you two equations on either side, and also gave you the chance that B given A is same as probability of B is same as probability of B given A complement.

Just kind of a key idea in probability, that this equation that we have here, the start told you that one event does not affect the outcome probability of outcomes then other event. And to generalise this, we also said that, for any A1, A2, An, and Bn equal 2 are are independent. It is not enough to just look at just A1 and A2, and An was the product of i equal to 1 to n, your Ai is not the only equation to look at, you have to be a little bit more tricky to you have to put, you have to put an epsilon i at the top, epsilon 1, epsilon 2, epsilon n and put epsilon i where A super epsilon i was equal to epsilon i was say equal to 0 or 1 and A super epsilon i was either equal to A if epsilon i is equal to 0 or equal to 1 let us say and A compliment, if epsilon i was equal to 0.

So, you have to take all these combinations to show, this is a 2 to the n equations here, for independence of n events. This is what we discussed long time. So, now you have two random variables. So, you can understand they are functions of two random variables, they two sample spaces. And you know that they have the same law if the probability of X equals to t as same as probability of Y equal to t. Now, you can also try understand when are two random variables independent or dependent. So, the X affect the outcome of Y or Y affects the outcome of X.

(Refer Slide Time: 12:49)

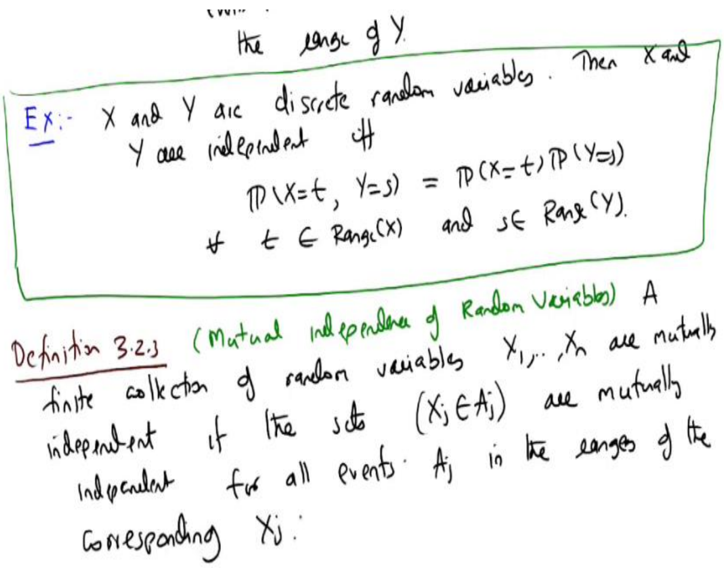


So, this can be generalised, but the above notion can also be considered for random variables. So, you may have several different random variables. So, this key to understanding relationship between another, so you understand the relationship between random variables. So, here is the definition. So, the first definition, I will write down as the following definition 3.2.1.

So, here this is called independence of two random variable, you can write that maybe in green, so independence of two random variables. So here we say, two random variables X and Y, X and Y are independent, if X in A and Y in B are these two events, so you take these two events, events if the events X in A, and Y in B are independent for every event A in the range of X and every event B in the range of Y.

That is that is the definition of independence. So, you have to check if every if every any event in the range of X. If any event in the range in the range of Y and then you have to check whether these two events are independent. That is X in A and X in B are independent. So equivalently, you can write down to the following way as equivalently, that is to check probability of X in A and Y in B that should split up as probability of X in A probability of Y in B for all A events A in the range of X and events B in the range of Y. That is the ideal point, randoms. So, this one problem I will give you as a homework problem, we can take up as an exercise is the idea that that independence is equal to the following idea.

(Refer Slide Time: 16:54)



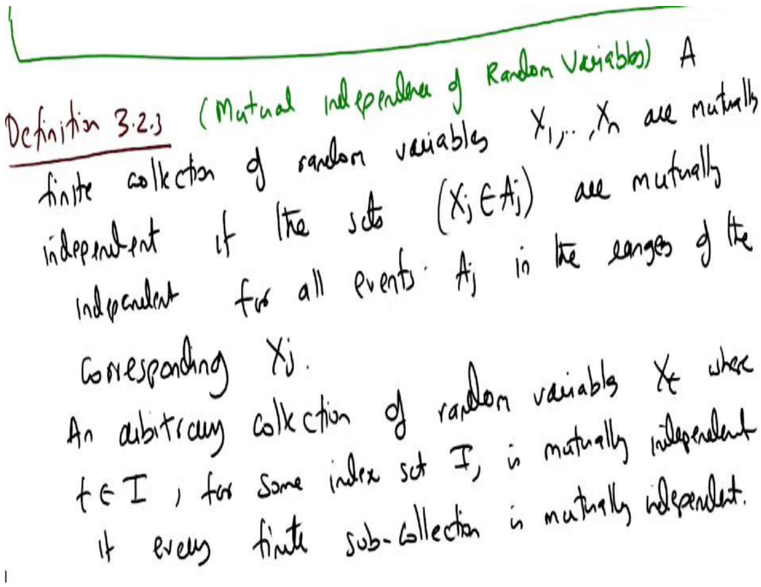
So, if X and Y are discrete random variables. So, X and Y are discrete random variables, then X and Y are independent if and only if, so you did not have to check for everybody, all events, it's enough to check for the following, you have to check probability X is equal to t and Y equal to S is the same as probability X equal to t and probability Y equal to S for all t in the range of X and S in the range of Y. That is enough to check.

This is some exercise that you can try and outline the homework assignment, but this you can check, it is not that hard to check. You did not have to check for all events A and B, you can just check it for just these particular events X equals t and Y equal to S, but t and S range over, range of X and range of Y respectively.

And then similarly, once you define this, I am going to slide down one more thing, you can define mutual independence of many random variables, as same way we defined for events define the same way. So, you can say this definition r 3.2.3. Think of it as mutually independent random variables, which are independence random variables.

You can think of this in the following way, you say a finite collection of random variables X1, X2, Xn are mutually independent if the sets Xj sets or event Xj and Xj are so these sets I always define by brackets respectively, are mutually independent for all Aj, all the events Aj, all events Aj all events A sub j in the ranges of ranges of the corresponding X of j.

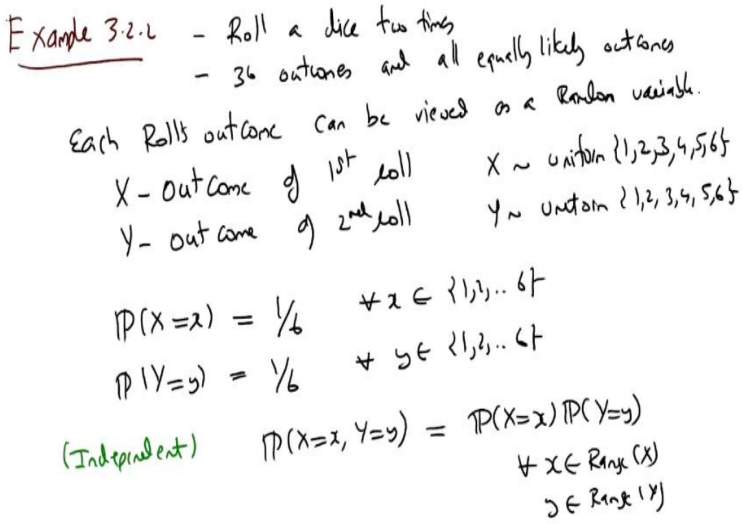
(Refer Slide Time: 21:18)



And similarly, like in the random variable case, j you can say for all the -- as the events case you can write again, an arbitrary collection random variables of random variables Xt, where t is in some set index set i for some index set i, is mutually independent if every finite sub-collection is mutually independent, that is one thing.

So, these are ideas that go with random variables quite easily. So, let us do a couple of examples. And then idea will be more clear. So, before we had this idea that we rolled the dice twice, and we had these 36 outcomes, so then we worked with it properly and said that to identify events of random variable and sorts of things.

(Refer Slide Time: 23:05)



So, let us do an example. Example let us say that is called 3.2.2, that is in our book. So, the idea is that, you roll a dice twice, so you roll a dice, so idea was to roll a dice two times. And that was the experiment. And then what we did was we viewed it as 36 outcomes. And then we and they were all equally likely. And handwriting is quite bad, let us change it properly.

So, you roll a dice twice. And then the way we set it up was 36 outcomes. And we designated all of them as equally likely. And then we try to understand various events. But then now we come about the following way, each roll each rolls outcome. So, now we can view each rolls outcome as a random variable. Each rolls outcome can be viewed as a random variable.

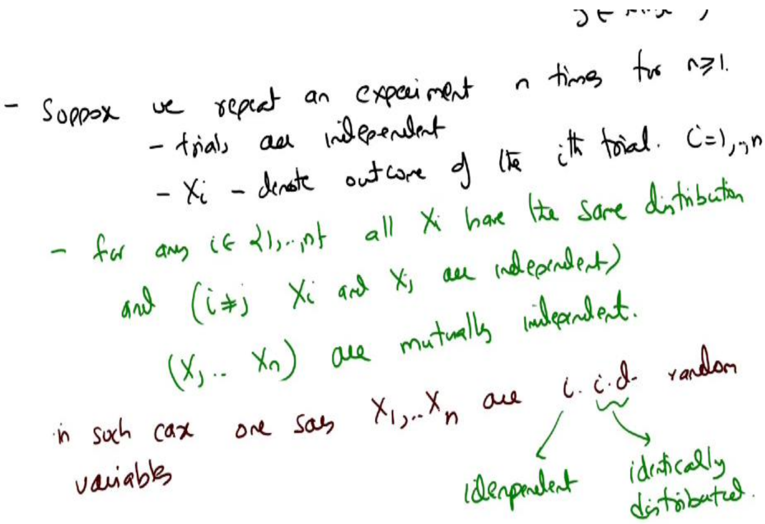
And then just think a little bit then you think of it as let us say X is the outcome of the first roll. And Y is the outcome of the second roll. Then from last class, you can easily check that X has the same distribution as uniform in the interval one up to one interval in the set 1, 2, 3, 4, 5, 6. And similarly, Y is also uniform 1, 2, 3, 4, 5, 6.

And you also know the following, you also know the chance of X equal to X is equal to 1/6th for all X in the range of X, which is just 1, 2, 3, 4, 5, 6. Similarly, more the chance that Y equal to Y is equal to one sixth for all Y in the range of X -- in the range of Y sorry, one up to six.

And more crucially, we also know the following fact, you also know that they are both independent. Because this calculation we have done before that the chance of let us say X equal to X and Y equal to Y, that is the outcome of the first roll was X and the outcome of second roll was Y, we already shown long time back that the same as probability X equal to X and Y equal to Y. This is true for all X in the range of X and all Y in the range of Y.

So, this is one simple example that you have seen of independent random variables. If you just focus on the first roll, the second roll, they are both independent events, independent out independent trials. And that forces X and Y to be independent. So, now this is sort of, let us say we repeat an experiment many, many times.

(Refer Slide Time: 27:09)



So, suppose above was only for two X -- two trials. So, suppose we repeat an experiment. Suppose we repeat an experiment n times for n bigger than or equal to 1, some n variable. And let us assume the trials are independent. Each trial of the experiment is independent.

And then, let us say and let us say let Xi denote the outcome of the outcome of the ith trial, for i equal to 1 up to n. So, from this there is there is two immediate consequences. So, one is, for any i, in let us say 1 up to n, all Xi have the same distribution. That is one thing. And i naught equal to j, Xi and Xj are independent. That is, that is one immediate consequence because each trial is independent.

Moreover, this is just a immediate by product. You can also say that X1, X2, Xn are mutually independent. That is the that is the consequence of having n, independent trials. So, there is notation for this in such a case, so in such a case, one refers to these exercise so in such in such cases, in such cases one says X1, X2, Xn are iid. So, the iid random variables.

So here the first i is for identical, sorry independent, first i is for independent and this id is for identical distributions, distributed identically distributed. So, this is what iid means, independent and identically distributed. That means the random variables are independent. And they have the same distribution. This is like a very, these are very interesting random variable, and the repeated trials can be, can be thought of as iid random variables. And this abstraction allows you to do many interesting calculations. And understand, let us do the following. Let us do a simple example.

(Refer Slide Time: 30:49)

- X_i - denote outcome of the ...
- for any $i \in \{1, .. n\}$ all X_i have the same distribution
  and ($i \neq j$ X_i and X_j are independent)
  $(X_1 .. X_n)$ are mutually independent.

In such case one says $X_1,..X_n$ are i.i.d. random variables

i. (independent) i.d. (identically distributed.)

Example 3.2.4: let $X_1,.. X_n$ be a sequence of i.i.d. random variable, such that $X_1 \sim$ Geometric (p), $0<p<1$.

So, let me do example 3.2.4. So, here what I do is, I say let X1, X2, Xn be geometric be random variables be iid random, let us say iid be a sequence of iid random variables, random variables such that each of X1 has the same law as geometric P, for some 0 less than P less than 1. So, but this means what, that means if the iid that means everybody has the same, that is all random variables are geometric random variables with parameter. So, suppose I want to understand what is the chance, let us ask you a question that is, so I am -- what does it mean I am repeating n trials, and I am waiting for the time to get the first success. That is what geometric is right?

(Refer Slide Time: 32:20)

(i.e. X_i - outcome of experiment i ; Experiment is :- observe the time taken for the first success)

What is the Probability all random variables are larger than j ? (for some $j \geq 1$)

i.e. Find $\mathbb{P}(X_1 \geq j, X_2 \geq j, \ldots, X_n \geq j) \equiv ?$

Step 1: Find $\mathbb{P}(X_1 \geq j)$

$X_1 \sim \text{Geometric}(p)$; $\mathbb{P}(X_1 = k) = p(1-p)^{k-1} \quad \forall k \geq 1.$

$$\therefore \mathbb{P}(X_1 \geq j) = \sum_{k=j}^{\infty} \mathbb{P}(X_1 = k)$$

So, let us just write this down the background, the background we have in our head that is Xi is the outcome of experiment i, experiment i and what is the experiment the experiment is you repeat trials all sets not use trails, you look you find out you observe the time taken for the first success.

That is what this is doing. So, I mean, each exercise is doing this following. So, what if I asked you the following question? What is the chance? What is the probability? So, let me write the question down properly? The question is what is the probability? That it took you at least j trials in each of the experiments? So, what is the probability that all random variables are larger than are larger than j, j let us say j is some number, j is some number its green j is some number.

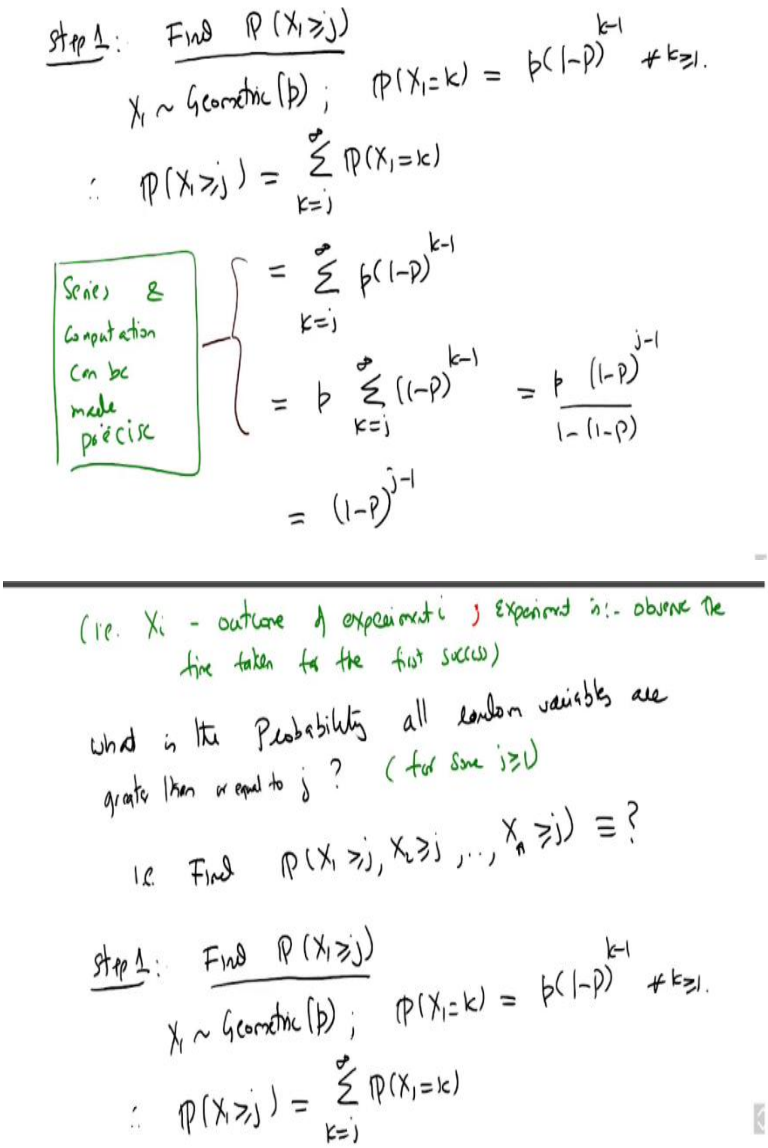
So, what does it mean what do you have to calculate? You have to calculate the following that is you want to find that is you want to find the chance that X1 is bigger than equal to j, X2 is bigger than equal to j and so on so forth all the way till Xn greater than or equal to j. This is what, is what you have to remember.

So, how do you solve such a problem? So, the first step is, step one is you find the chance at one random variables. First step one is to find the chance that X1 is greater than or equal to j. How do you do that for one random variable.

So, X1 is geometric P. Same law geometry P. So, I have been using tildes on, let me use the word tildes, X1 is the same law as geometric P. That means what that means probability X equals X1 equal to some letter K is the same as P times 1 minus P to the K minus 1 for K bigger than or equal to 1. So, the K minus 1 indicates that our K minus 1 failures and success of the Kth trial. So, probability mass function is given by this for all K.

So now, therefore the chance that X1 is bigger than equal to j, that is going to be what now, or let us say K equal to j is, is going to be what? Is you have to you have to look at the chance that X is equal to K and you sum over all possibilities of K from j to infinity.

(Refer Slide Time: 36:15)

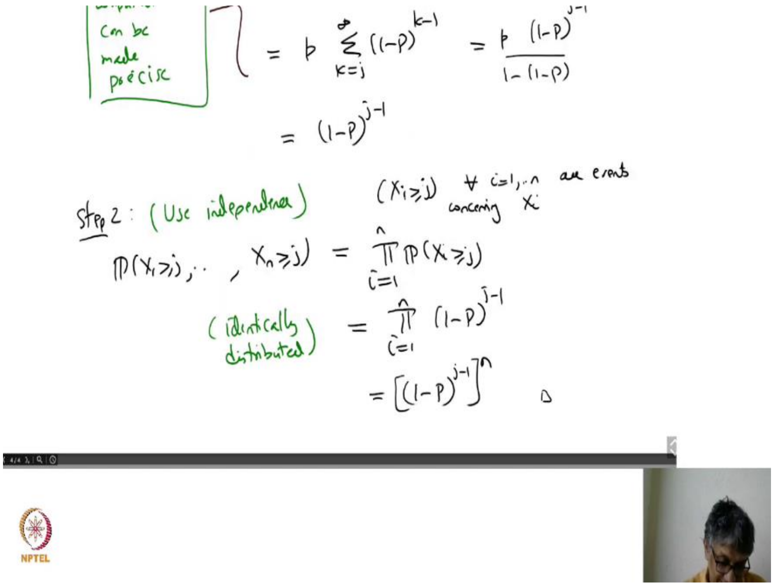


But then you already know the chance of X equal to K is this expression, so you plug that in, so K equal to j to infinity, P times 1 minus P to the K minus 1. And that is the same as P times sum from K equal to j to infinity of 1 minus P to K minus 1. This is a geometric series. Anyone can understand that those calculation as P times be the 1 minus are, P is 1 minus P. So, this K minus 1, 1 minus P to the j minus 1 divide by 1 minus 1 minus P.

And that if you do the calculation, you just get 1 minus P, j minus 1. You just get 1 minus P to j minus 1. This is what you get. Off course I have been a little bit loose in mathematical terms, you'll be a little careful when you do this computations do this computations, you have they are all limiting objects, so you should be a little careful. So, they are all a limiting object they are all series.

And this computation can be made precise, made precise. So, I have just manipulated series in an intuitive way. But you should be able to go back and do it, they are all series, so should think of them as limiting objects and do it for every partial sum, and then finish the computation, after I will not do these things. Since now we know that everybody is of this form. But what did we want to calculate? We want to calculate that chance that X1 was bigger than equal to j X2 was bigger than equal to j, and Xn. Maybe I just rephrase the question a little bit, or instead of writing larger than j I will make it precise; I will just write it as greater than or equal to j. This is what I am doing.

(Refer Slide Time: 38:28)



So now, I have to compute this question. I know what how each X, how X1 behaves. So now step 2, you have to use independence. So, using the problems, using the problems. So, what you do here, you know that the chance that X1 is bigger than equal to j, up to Xn is bigger than equal to j, you observe, what you observe? You observe that each of these events X1 is bigger than equal to j or let us say Xi bigger than equal to j, for all i equal to 1 to n, i events concerning Xi. So, by independence, you know that this is the same as the product from i equal to 1 to n, the chance that Xi is bigger than equal to j.

And this we know, now use the fact that they are the same distribution, use the fact that they are identically distributed. And that gives you the chance the product of i equal to 1 to n they are all the same, they are all the same as stuck, so the 1 minus P to the j minus 1, that is X1 has the same laws as Xi for any i.

And now you have, this doesn't depend on i. So, it just comes out as 1 minus P to j minus 1 to the whole power of n. So, that is the chance of, of all of them being bigger than, so this is one computation one into independent variables. So, chance that in n trials, we look at the number of successes to be the first time we get the time taken to calculate the first successes, at least j then this is the computation…