Social Networks
Prof. S. R. S. Iyengar
Department of Computer Science
Indian Institute of Technology, Ropar
Rich Get Richer Phenomenon

Lecture – 116
Why do Normal Distributions Appear?

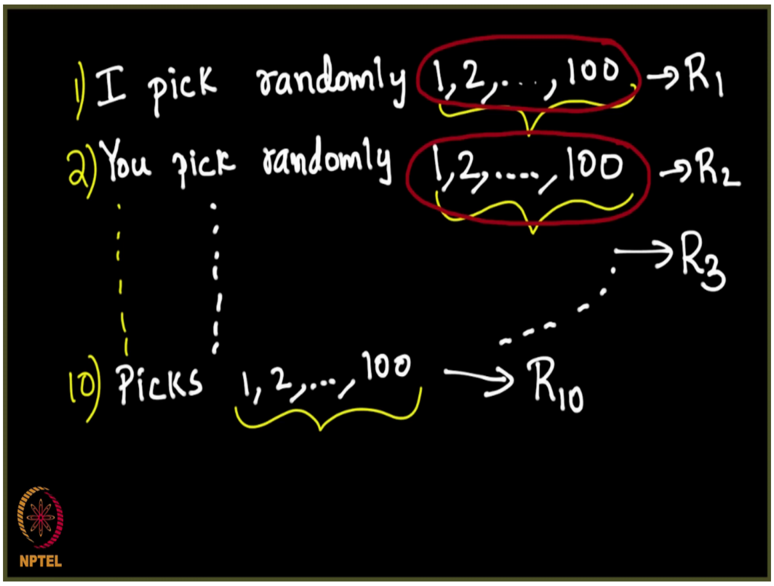
(Refer Slide Time: 00:11)

So, we saw two things. One was the normal distribution right. We saw basically two plots. So, far one is that I told you people there is this normal distribution; the inverted curve and then a drop like this ok. So, the definition the example heights of people in a in a town and this happens to be the degree distribution of nodes in a real world network you know real world; we see this ok.

Now, before asking why do we see a very different kind of a plot here? Why in most of the situations, we see this kind of a plot. Before asking this question, we would like to ask individually, why and how does this come about? This one same time why and how of this as well.

So, to begin with let me ask the why and how of this normal distribution. How does normal distribution come by ok? Let us take a look at a nice example. And try to motivate the reason behind why the normal distribution appears in nature ok. Here is an example you should invite you all to think a little about it ok. It is not very difficult to understand it is pretty straight forward ok.

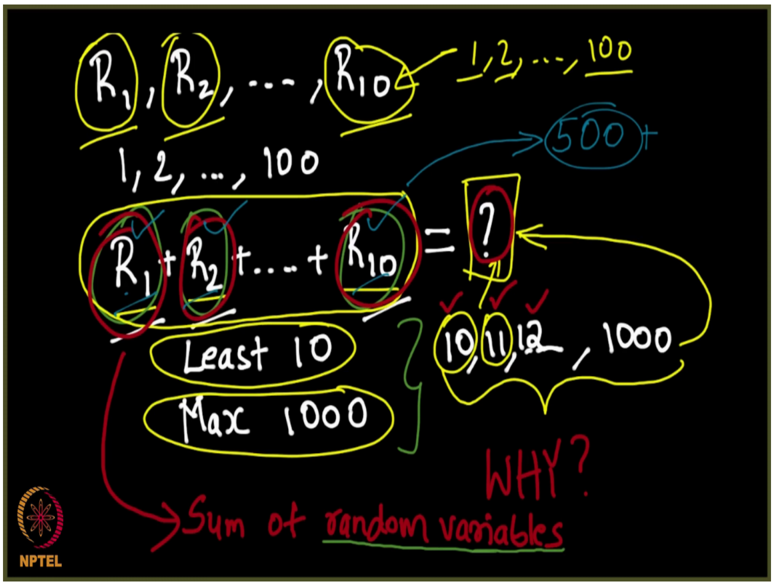
(Refer Slide Time: 02:01)



So, assume I pick a random number; randomly I pick randomly a number from 1 to 100. One of these numbers I pick some one number ok. And then you as another person, you also pick randomly some number from again from the same set 1 to 100. Some number you will pick randomly here all right good. Now see when I pick a number randomly here from 1 to 100, I can pick any number; it could be a 1 it could be a 2, it could be a 3, it could be a 50, 60, 73 89 100. Whatever I want when you pick as well you will you will have the liberty to pick any number.

Let us assume 10 such people pick. I am the first person, you are second person so on and so forth; some 10 people pick like this ok. Tenth person also picks a number from 1, 2, 100 right. So, what I, now what I am now going to do is let us say the number here that I pick let me call it R 1. The number you pick I call it R 2, the third person picks a number I am going to call it R 3 and so on. And the 10 person who picks the number that number is R 10 ok.

(Refer Slide Time: 03:52)

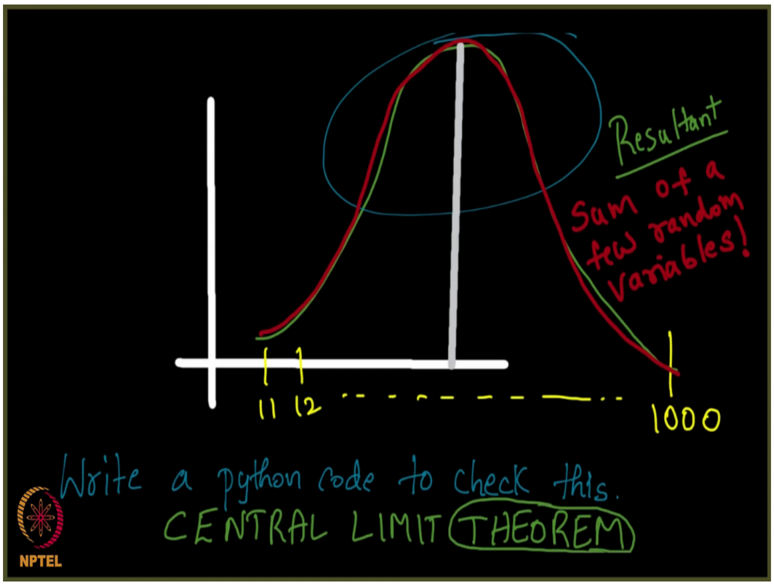


So, let me go to the next slide, let me write down. There are 10 random numbers each one between 1, 2, and 100. Each of these random numbers are from 1, 2, 100. Here is an important question while each of these random numbers can be anything from 1, to 100; just in case I added them look at this just in case I added all these, what will this be? See each one of these can be a number from 1, to100 which means the sum will be at the least it will be 1 plus 1 plus 1 10 times. Assume all them shows a 1, 1, 1 each then at least it will be 10 and a maximum of how much.

If each one of them picked 100 it will be 100, 100, 100 a 10 times. It will be 1000. So, this is the least and this is the maximum the minimum and maximum, but then, but then observe here that any number 10, 11, 12 up to 1000 can appear for what for the sum correct, but here is the climax. While whatever numbers you pick can be any number between 1, 2, and 100 when you take these ten numbers and add them, what you get as an answer is not uniformly at random from these numbers which means 10 and 11 do not appear uniformly at random here. While let us say in R 1 and 2 and 3 and what not up to 100 appears randomly. Each one of them is has the same chance of appearing here, but then the chances of 10 appearing here is less than the chances of 11 appearing here is less than the chances of let us say 12 appearing here, here in this place. And why is that?

If you can think; you will come to a conclusion that this is indeed once again a bell curve right.

(Refer Slide Time: 06:25)



It is going to be if I were to plot the values; let us say the least what 11 12 so, on up to it will go on and on and up to let us say 1000 is the maximum. You will observe that uniform. So, let me go back to the previous slide. So, each one of them each one of R 1 can be on an average let us say anything from 1 to 100 so on and so forth. So, this will be roughly; so, this will be like this the plot will look like this right. It is good to write a programming and then check.

Let me assign that to you the assignment to you all to write is to write, a piece of python code; write a python code to check this. You will be surprised to see that it is actually a very beautiful curve like this ok, to check this all right. So, why it is this happen that is very obvious. It is very very unlikely that you pick a 1 and a 1 and a 1 for all of them correct. It is equally unlikely that you pick a 100 and a 100 and a 100 for all of them correct.

So, what is more probable as a sum i if you if you have guessed it right it is roughly around 500, actually 550 but roughly around 500 plus right. That will go, that will be the most free; that is the peak here, rapped that the peak here is actually 550 and completely the peak here is going to be 550 that is the peak ok. Anyway you do not worry much about it. All you need to know is that you will see a normal distribution here. Now why is that? Why does it why what exactly is the moral of the story?

The moral of the story is the sum of random variables that you see here. The sum of random variables they are basically a bunch of a sum entities that are random. They are called the random variables; sum of random variables. And this is slightly it is the term that would not have really heard of well unless you have done a course in statistics, but just see random variables as simply random entities which can take any values all right. Now whenever, the resultant this; a resultant whenever the resultant is a sum of several random variables; whenever the resultant is a sum of a few random variables several few random entities random variables, then you will see a normal distribution like this.

You will see this distribution whenever the resultant is a sum of few random variables. This goes by the name this is actually a very well known theorem in statistics rather probability. It is called the of central limit theorem. So, I have stated in a very in a really easiest possible way, you do not worry much about the word theorem here. It is sort of scary for many of you people who do not like math much, but the all I am saying here is that the sum of random variables right.

This sum of random variables is this resultant and that resultant is generally a bell curve if its sum of few random variables. And that is how, this emerges. This always emerges when the resultant is a sum of random variables this. In fact, true in nature whenever you see an experiment. The sum total will basically the resultant that you are plotting will basically be a bunch of random events put together right ok.

So, let me go to the new slide and let me ask the next question that the next question is actually in the first slide itself. that I question is this. How does this emerge? What is the reason behind the emergence of this thing? We gave some reason for the left plot this one. So, what is the reason for this one is what we will be covering in the next lecture.