Essential of Data Science with R Software-1
Probability and Statistical Inference
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture No. 13
Axiomatic Definition of Probability

Hello friends, welcome to the course Essentials of Data Science with R Software-1 in which we are trying to understand the topics of probability theory and Statistical Inference. So, you can recall that in the last two lectures, I have given you a fair idea about that what is probability? What is this number indicating? Now I will try to pose here a question mark and you try to answer it yourself because you cannot answer it to me at the moment.

Suppose somebody says that person has repeated the experiment say 500 times. And one person says or the another person says that the person has repeated the experiment say 700 times. Now the question is which of them is going to give a real true probability of the event. And particularly in those situation where you really do not know how to compute the probability what is the exact value of the probability.

Now definitely according to the theory what we have proposed or what we have understood, we know that if the outcome is based on larger number of repetition that result is more dependable. But more dependable does it mean that this is the true value? Now how to solve this problem? We simply said as n goes to infinity this estimated value will be tending towards the true value. But how many number of times we have to repeat it? How to know these things?

Because if the process is very simple possibly smaller number will suffice to get a reasonable value. But if the process is complicated then possibly a larger number of repetitions might be needed. But how large is large and how small is small, that is the question we do not know the answer. So, the question is how to solve this problem? You know that in your say class if somebody has got 59.5 marks of out of 100 and somebody has got 60.1 marks out of 100. The question is to whom you will say that the person has achieved the first class?

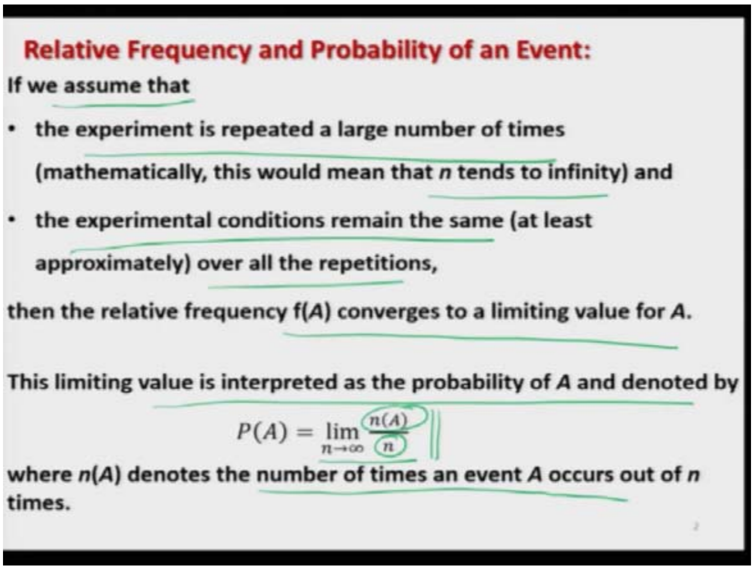
There will always be a confusion that okay the difference is very small. So, both are first class. But difference is small for me and that is a very qualitative decision. So, these qualitative decisions do not help us in real life. We have to make a rule to solve this type of problem. If your teacher or the college authorities make a rule that anybody who has got more than 60 percent of marks that student will be classified as first class and if the marks are less than 60 percent marks then the student will be classified as second class. Now after this there is no confusion. Because a rule has been given that you simply try to see.

If the marks are more than 60 percent the person has got first class otherwise second class as simple as that. So, how the problem is solved? Just by making a rule, making an axiom that before you go to the exam, this axiom is given to you that if the student gets more than 60 percent marks, the student will be classified as first class or if get smaller than or less than 60 percent marks, the student will be classified as second class, after that there is no confusion.

Now can we do the same thing in probability theory also? We have not made any rule that how large should be n so that one can say that the probability that you have computed is 100 percent correct. So why not to make this rule and try to define the probability in terms of those rules, then our problem is going to be solve. So, with this objective, some axioms were given to the probability. So, based on that, we have axiomatic definition of the probability.

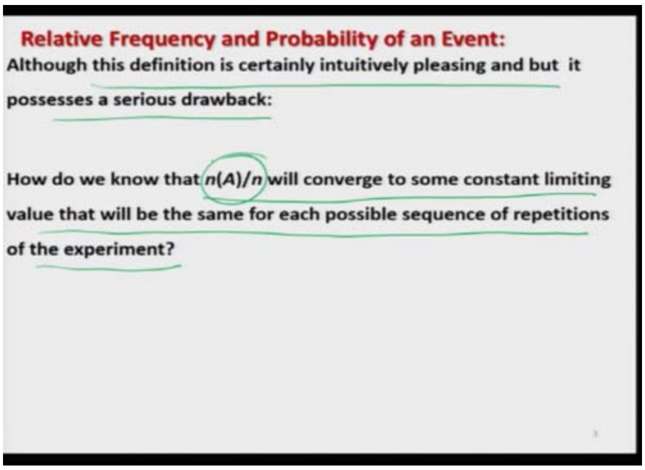
So, this type of concept and definitions will help you in taking the correct decision that the way you are trying to obtain the probability is correct or not. And if you try to solve many controversial issues. So in this lecture, I am going to introduce you with the concept of axiomatic definition of probability. So, let us begin our lecture and try to see what is this?

(Refer Slide Time: 5:48)



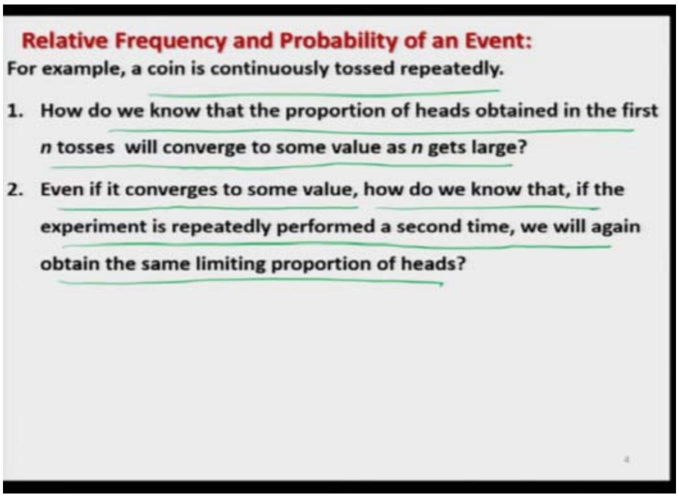
So, now we know that in the definition of probability where we are trying to observe the probability as the relative frequency, we assume that the experiment is repeated a large number of times. And mathematically this means that n is going to infinity that is the total number of times of repetition that is going to infinity and the experimental conditions remain the same, at least approximately over all the repetitions and then we say that the relative frequency f A converges to a limiting value for A and this limiting value is interpreted as the probability of A and it is indicated by these notations that is probability of A is equal to limit n tending to infinity n(A) upon n where this n(A) this is the number of times an event A occurs out of small n number of time. So, this is what have done in the earlier lectures.

(Refer Slide Time: 6:47)



Now I will try to move further. All thou this definition is certainly intuitive and pleasing but it possesses a serious drawback. How do we know that n(A) upon n will converge to some constant limiting value that will be the same for each possible sequence of repetitions of the experiment? What is this means if you try to see in the earlier lecture when took the example on the R Software? We are trying to repeat the experiment and we have tried to generate the say 100 observations, again and again. And every time you are getting a different value of the relative frequency. So, now, how are you going to take a call or how are you going to know that n(A) upon n is converging or not to some constant limiting value.

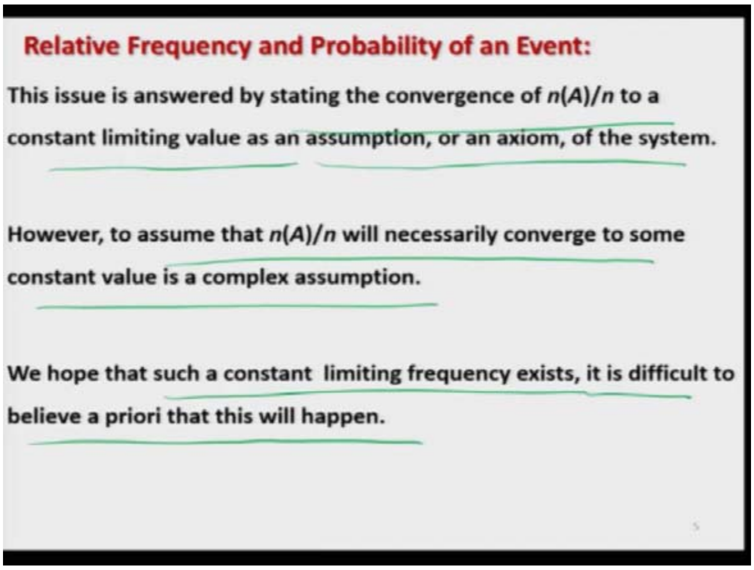
(Refer Slide Time: 7:43)



For example, in case if a coin is continuously tossed repeatedly, how do we know that the proportion of heads obtained in the first small n number of tosses will converge to some value as small n gets larger? And even if you try to find such a n and even if it converges to some value, how do we know that if the experiment is repeatedly perform a second time we will again obtain the same limiting proportion of the heads.

You have seen that when we were trying to increase the number of repetitions then the difference among the values of relative frequencies was decreasing, but when you are trying to repeat the experiment those values are not coming out to be exactly the same. You had not got the probability of observing say 1, 2, 3, 4, 5, or 6 in the roll of a die always as exactly 1 upon 6. So, as like 0.1666 and so on.

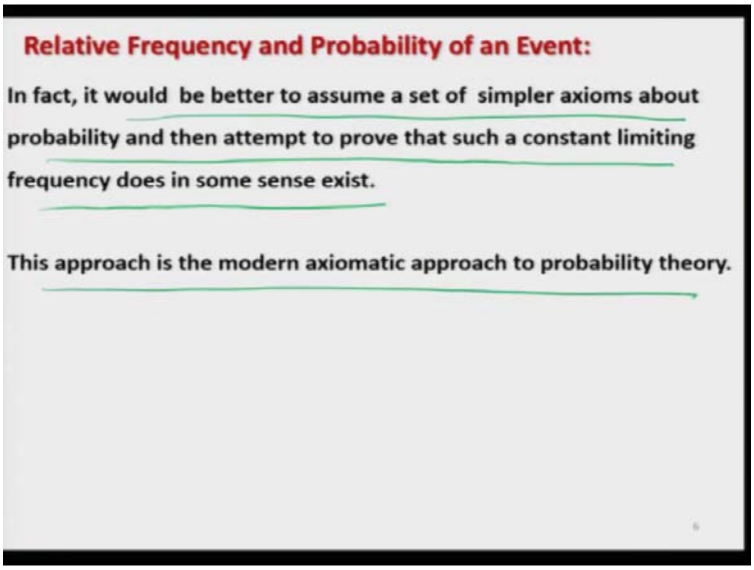
(Refer Slide Time: 8:47)



So, now, in order to answer this question, we have to do something more. And we need to state that the convergence of n(A) upon n to a constant limiting value as an assumption or an axiom of the system. We have to make some rule. All though even if you try to assume that n(A) upon n is getting larger but still to assume that n(A) upon n will necessarily converge to some constant value is not an essay job. It is a complex assumption.

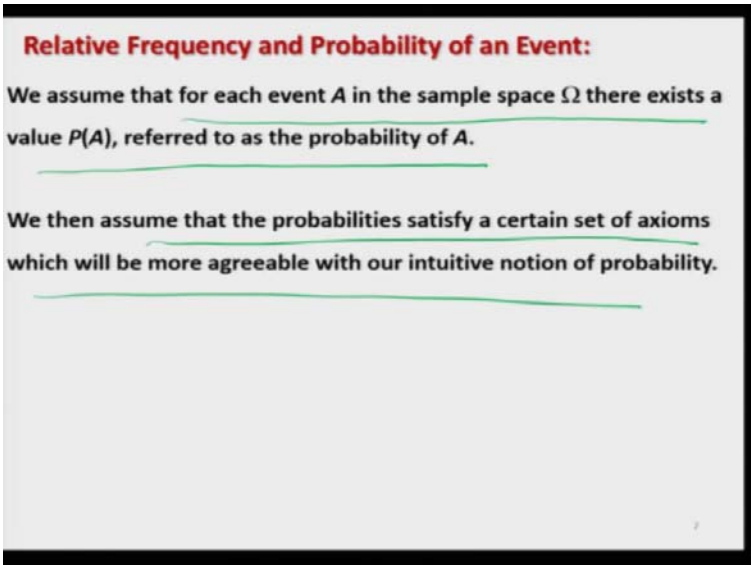
We hope when we try to compute such probabilities that such a constant limiting frequency or constant limiting relative frequency exists. But it is difficult to believe a priori that this will happen. How? Now you think about it that when we started conducting the experiment on the R console and we started generating 100 observations, 1000 observations, 10000 observations so on. We did not knew where to stop. 100, 500, 5000, 50000 what?

(Refer Slide Time: 10:10)



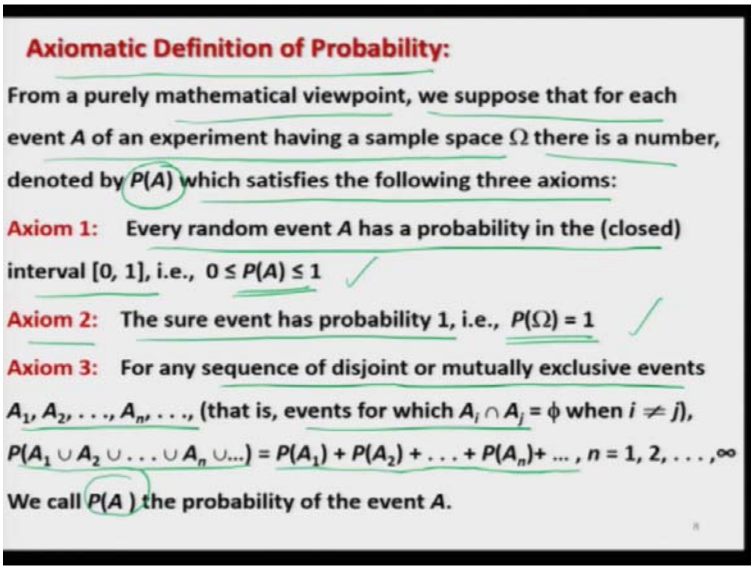
So, it would be a better option for us to assume a set of simpler axioms about probability and then attempt to prove that such a constant limited frequency does exist in some sense. And this approach is the modern axiomatic approach to the probability theory.

(Refer Slide Time: 10:36)



So, in this case what we try to do here that we assume that for each event A in the sample space Ω there exist a value of P(A) which is referred to as the probability of A. And then, we assume that the probabilities satisfy a certain set of axioms. Which will be more agreeable with our intuitive notion of probability. That is what we are doing up to now that we believe that as we are trying to increase the number of repetitions our probability is getting better and better. But the true probability that is unknown to us, but we are simply assuming that it is getting better and better.

(Refer Slide Time: 11:20)



So, why have to make some axioms and now we present here the axiomatic definitions of probability. Well, one thing you know that whenever we are trying to introduce something we had discussed in the beginning also, that when the concepts are based on the tools of mathematics then people do believe on them. So, that is the same approach we are going to follow here also. We are simply trying to solve this issues on the basis of some mathematical laws- Rules of mathematics.

So, the advantage will be whenever we are trying to define a probability and that definitions or that value is going to satisfy these axioms, I will be confident, I will be sure that my value is correct. So, that is why it is very important to understand or to introduce the mathematical concept in such definitions. So, whenever we have a confusion. we try to take the help of mathematics.

For example, whether 59.5 marks out of 100 or 60.2 marks out of 100 is going to be classified as first class or second class, this all by a simple rule that if the marks are more than 60, the person is first class. And if not, then it is second class. Now you can argue whether the marks are 59.999 or so, from the mathematical point of view it is less than 60 that is all. And if the numbers or marks 60.000001 even, then the rules is very simple it is more than 60 so the candidate has got first class that is all. So, this is how we are trying to think and this is how I request all of you to think to solve the problems in decision sciences or data sciences.

So, let us try to understand the axiomatic definitions of this probability. So, as I said the axiomatic definitions of probability is the definition from the mathematical point of view and from purely mathematical view point, we suppose that for each event A of an experiment having a sample space Ω there is a number which is indicated by P(A). You can recall that the P(A) we had indicated as probability of an event A. But now at this moment we do not know what is P(A) but it is simply a function of event A.

So, this P(A) satisfies the following three axioms. The axiom number one says every random event A has a probability in the closed interval 0, to 1. That is P(A) lies between 0 and 1 where 0 and 1 are inclusive. The second axioms says that the sure event has probability 1 that is probability of Ω which is the sample space is equal to 1. And the axiom number 3 says that for any sequence of disjoint or mutually exclusive events this A_1, A_2, and … that is the event for which we have $A_i \cap A_j$ is equal to phi and i is not equal to j.

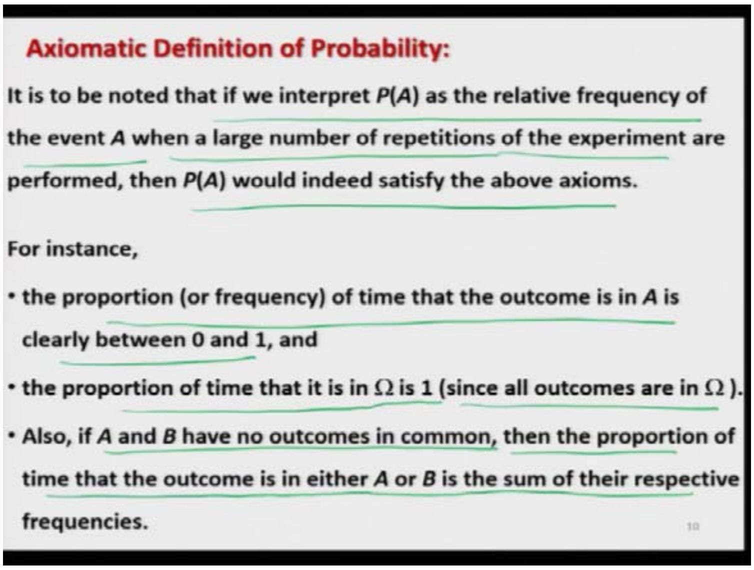
Do you remember that we had the concept of disjoint events in which we say that there are no common points between the two events A_i and A_j? So, in case if we have such a sequence of disjoint events then the probability of $A_1 \cup A_2 \cup \ldots \cup A_n$ that is probability of A_1 plus probability of A_2 plus probability of An …. And we call this value P(A) as the probability of the event A.

So, now I can say that here the axiom 1 is simply saying that the probability of any event will lie between 0 and 1. The probability of a sure event will always be equal to 1 and if we have the disjoint or mutually exclusive events, then the probability $A_1 \cup A_2 \cup \ldots \cup A_n$ can be express as the sum of their individual probabilities. So, this type of number will be called as the probability an event A.

So, you can see here that this axiom is trying to state that the probabilities that the outcome of the experiment is contained in A is some number between 0 and 1. Well, 0 and 1 are inclusive. 1 is indicating the probability of sure event and 0 is indicating the probability of an impossible event. And similarly, the axiom 2 states that with probability 1 the outcome will be a member of the sample space Ω that is obvious that all the event what you are trying to define they have to be a or they need to belong to capital Ω the sample space.

And axiom 3 states that for any set of mutually exclusive or disjoint events the probability that at least one of these events occur is equal to the sum of their respective probabilities. And actually this axiom 3, what we have said this is the theorem of additivity of disjoint events. So, you will see in the books this name were there. So, I thought I should inform you.

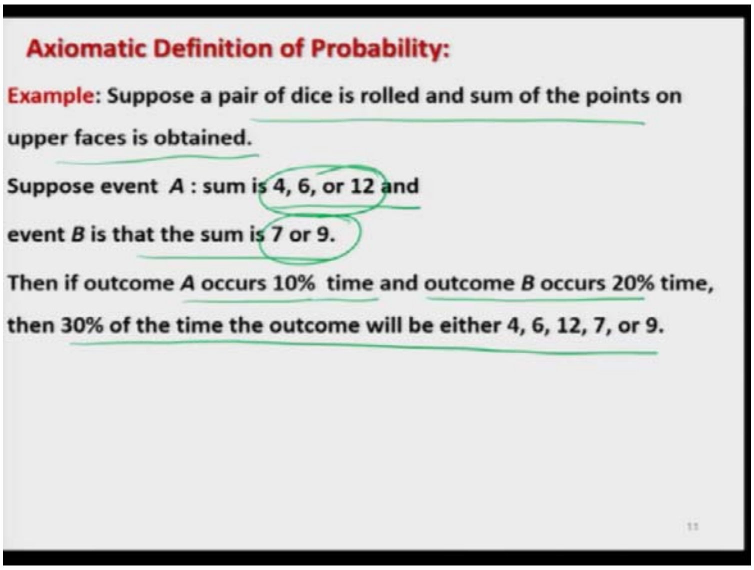
(Refer Slide Time: 16:51)



And we have to understand and we need to note here that if we interpret P(A) as the relative frequency of the event A. One and other total number of repetitions are becoming large in that experiment, then P(A) would indeed satisfy these three axioms. Axiom 1, axiom 2, and axiom 3. For example. the proportion well if you try to see the relative frequency what is that? This is only a proportion, the proportion of times that the outcome is in A is clearly between 0 and 1.

And the proportion of times that it is in Ω is always equal to 1 since all outcomes are in Ω that is our assumption and that is how we try to define the sample space Ω. And also if A and B have no outcomes in common, then the proportion of time that the outcomes is in either A or B is the sum of their respective frequencies. Frequencies means relative frequencies.

(Refer Slide Time: 17:56)



For example, if I try to take a the same example suppose a pair of dice is rolled and sum of the points on the upper faces they are obtained. Suppose, event A is defined as sum is 4, 6 or 12. And event B is that the sum is 7 or 9, then what is this mean? That this means that if outcome A occurs 10 percent time and outcome B occurs 20 percent time then 30 percent of the time the outcome will be either event A or event B or the point 4, 6, 12, 7, or 9. 4 6, 12 are coming from event A and 7 or 9 they are coming from event B. So, this is what we are trying to do.

So, now we come to an end to this lecture and you can see here now I have solved you problem that whenever you are trying to define the probability and if that number is going to satisfy these three axioms we will say that confidently that this is my probability. Well, these axioms are useful when we are trying to compute different types of probabilities of different types of events in real life.

And what are those things that you now can clearly understand that whatever the questions I ask you in the beginning of the lecture now they are satisfied. You have given the rules that if any number P(A) follow this three rules that can be call as the probability an event A. So, you try to revise this lecture, try to understand it, try to set inside your mind now what is this theory and what it is trying to indicate you, what it is trying to inform you that will help you? So, you try to practice it and I will see you in the next lecture with more topics till then goodbye.