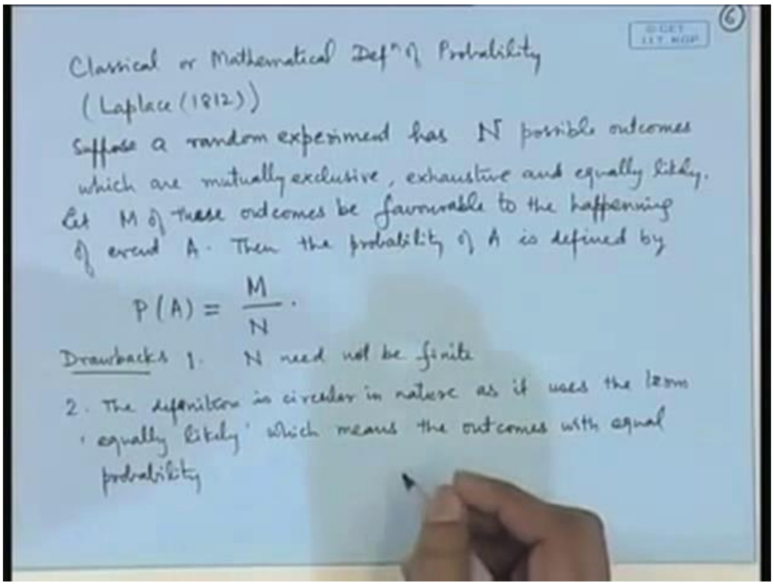
Probability and Statistics

Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 06
Definitions of Probability

Based on this we give a first or you can say a primary definition of probability it is called classical or mathematical definition of probability.

(Refer Slide Time: 00:33)



And this definition is due to Laplace, which are available on his critize on probability in 1812. Now this is based on certain conditions; I have already introduced the concept of several events so these all are subsets of the sample space and we make all event say A 1, A 2, A n to be equally likely now this is some terminology which is like circular in nature. Let me again explain.

So, I will say events A and B are equally likely; if A and B has the same chances of (Refer Time: 01:34), now till now we have not defined what is the chance, so this definition it is look circular, but anyway this is what has been used in the classical definition of probability. Suppose a random experiment has N possible outcomes, which are mutually exclusive exhaustive and equally likely. That means, we are looking at the elementary outcomes of the random experiment, which are color collected in the sample experiment; that means, the sample space as the total of N points, which is the finite number and naturally then when we are describing all of them they are suppose to be mutually exclusive and we have exhausted all the possibilities. So, I am making an assumption that they are equally likely.

Let M of these outcomes be favorable to the happening of event A, then the probability of A is defined by probability of A is equal to M by N. So, this definition was given by Laplace because the of the earlier statements which are that the probability theory has the origin in the given sub chance such as a tossing of a coin, throwing of a die, the numbers coming on a Rayleigh field etcetera. So, all of are driving up a card in a pack of cards.

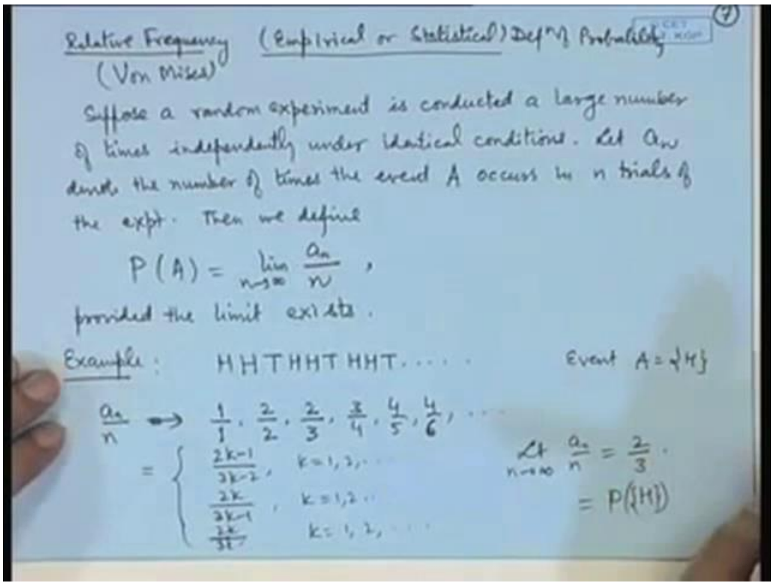
So, all of those experiments had a peculiar thing that they had a finite number of outcomes and assuming that the things game is fair for example, if you toss the coins, so you assume that it is a coin, if your die is thrown then you can see (Refer Time: 04:10) whether it is a pair die etcetera you could say a pack of card then you assume that it is a pack of well shuffled cards, 52 cards. So, these assumptions seem to be valid there, that is exhaustive mutually exclusive and equally likely, and therefore these definitions was given and this is the one which is used in the calculation of the probability in the classical examples.

So, main disadvantages or you can say drawbacks of these definition are that N need not be finite; for example, if you are considering the number of trails needed for the first success, then we do not know when we will start, suppose you are considering light of a bulb, suppose you are considering weight of birth of a child, then all are these outcomes the collection of outcomes those are either countably infinite or uncountably infinite set.

Then second is the more crucial thing that we are saying that the outcomes are equally likely. Now equally likely thing means that we are knowing that coin is sphere or the die is sphere or the packing well shuffled etcetera, but that is binding energy in inherent understanding of the definition of probability, whereas we are actually defining probability now, so it is a circular definition. The definition is circular in nature as it uses the term equally likely, which means outcomes with equal probability. Similarly in a given experiment we may not be able to express the outcomes as mutually exclusively outcomes. We may not be able to exhaust all the possibilities of the outcomes, because now total number of possibilities may be a really large to describe.

So, this definition though quite useful in the beginning of the development of the subject as it is limitations; later on in more important or you can say a more practical or a more applicable definition was developed which we call relative frequency definition of probability, which is based on actual conducting of the experiment.

(Refer Slide Time: 07:02)



So, we can also call it empirical because it is based on the actual observing of the outcomes or a statistical definition of probability. In the form of this definition is due to Vm Mises. Suppose a random experiment is conducted, a large number of times independently under identical conditions let a n denote the number of times the event A occurs in n trails of the experiment then we define the probability of the event A to be limit of a n by n as n tends to infinity provided the limit exists.

Now, let me give you an example of the actual application of these definition, let us consider say a trail of conducting a tossing of a coin and we want to find out the probability of it. So, this we are doing because we do not know whether the coin is sphere or not; in the classical definition we assume that the coin is sphere and then you try to find out probability of a (Refer Time: 09:46) etcetera.

But suppose I do not know what a coin is sphere, I actually want to find out the probability of it. Suppose an experiment of conducting of tossing of a coin results in T H followed by (Refer Time: 10:00) n. Now let us look at the sequence a n by n here. So, in the sequence a n by n here if you make it, in first trail you have a head and you are interested in the occurrence of head. So, a n by n is 1 by 1; in the second trail again there is a head and therefore, the ratio a n by n is 2 by 2.

So, you are looking at the event A is occurrence of H. So, if I look at the third trail then in third trail tail has occurred. So, the occurrence of head is 2 and the total number of trails is 3. Now if you just continuing this direction we can write the sequence a n by n like 1 by 1, 2 by 2, 2 by 3 then it becomes 3 by 4, 4 by 5, 4 by 6 etcetera and now in order to calculate the probability of head we need the limit of this particular sequence.

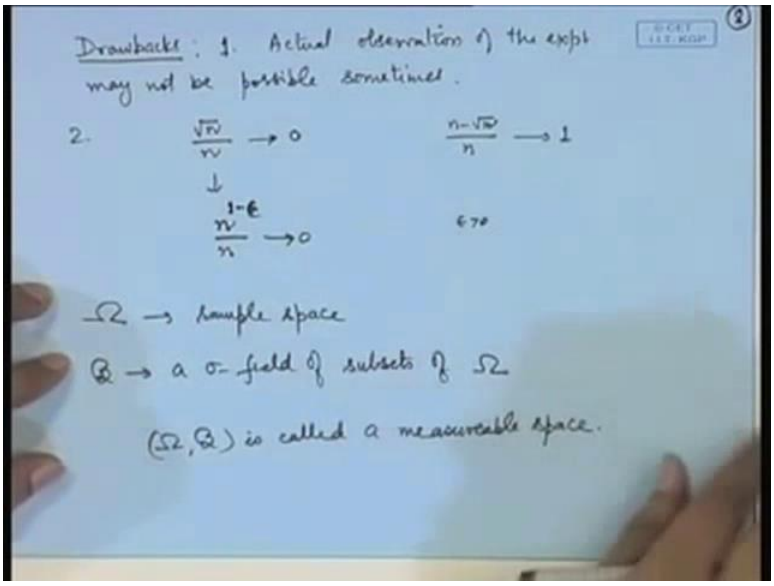
So, in order to do that let us write a proper mathematical expression for this, we can write it as say 2 k minus 1 by 3 k minus 2 for k equal to 1, 2 etcetera we may write it as 2 k by 3 k minus 1 for k equal to 1,2 etcetera.

If the number of the trails is of the form 3 k then it is 2 by 3, 2 k by 3 k for k is equal to one 2 and so on. So, in each of the cases you are able to describe the ratio a n by n and it is very obvious now that if I take the limit of this as a tends to infinity that is n tends to infinity, when the limit of this is 2 by 3 and therefore, probability of occurrence of head is 2 by 3; that means, this is a bias point in favor of head.

Now the relative frequency definition seems to be the one of the most reasonable definitions of probability, in the sense that it is based on the actual experience and that is what the subject probability should be all about in terms of a statistics; for example, if you are looking at the age at the death of a person, then it should be based on the experiment that is really how many people survive beyond the age 60, beyond the age 70; that means, it should be based on the correct age at death for the various persons.

If you want to find out the sex ratio in population, then we should look at how many children are born as male and female. So, it should be based on the actual data. If you want to say a something about a rain fall, then you should know in the past 10 years are in past 15 years, what is the pattern of the rainfall during monsoon season therefore, this relative frequency definition seems to be the most useful definition, for calculation of the probabilities; however, even these definition has certain drawbacks.

(Refer Slide Time: 13:30)



For example; we are making an assumption that the experiment is observed, but there may be certain experiments where we may not be able to look at the observations the observations may be in less in number or may be too complex too complicated, too costly for example, if you are looking at the failure rate of avalanched satellites then it is a not on experiment which can be conducted every day and if we are looking on a country like India then the overall launches of the satellite itself may be limited to a very small number and therefore, in order to find out the probability of a successful launch may be quite complicated.

Although we are not saying that this type of calculation of probability is impossible, but it is difficult certainly if we want to look at the full experiment; however, the probability can be calculated using certain rules of probability, by splitting the entire launch of satellite into various sections or various segments and then we combined the probabilities of those various things.

So, the actual conduct of the experiment are actual observation of the experiment may not be possible sometimes; also there are certain experiments which are destructive in nature for example, we want to look up how many of the match sticks kept in a match box are useful. So, conducting of an experiment mean that, we actually like a match stick and observe whether it is being good (Refer Time: 15:28) burn the holistic burns etcetera.

So, this kind of experiment is destructive in nature, because it will (Refer Time: 15:36) total destruction of the material itself and similarly there are various experiments conducted to test the extent of the materials, then it means then we apply a certain pressure on the material, the material which is used for making of certain mechanical things such as a car or an engine or a train line so, let us think of a material very important; how about the testing requires that we put some compressible force on that and observe it to break at a certain force and then we estimate that how much actual force will be required, if the actual force which is going to be applied on that material is less than that (Refer Time: 16:16) strength then we say that the material is all right so; however, such experiments are also costly in nature and sometimes it is escaped in all the and it is replaced by certain above process in a word that is strength.

Sometimes this probability relative frequency definition may give you a result, which may not be very intuitive for example, if I say that probability of an event is zero then, we must feel that the occurrence the number of occurrences in experiment for that particular event must be zero; however, real sense it is taken as the limit. So, we may have say a root N was the number of atoms in N trials and if I consider the limit as this when this is 0.

So, although the number of occurrences is not zero, but the probability of the event is zero; similarly we may have n minus root N by n, which will converge 1. So, every time the event does not occur; however, the probability is one which one indicates that this is the shear event; however, in the (Refer Time: 17:25) in the sense that it mean that the number of occurrences is negligible in comparison with the total number of trials.

Now, again this will lead to little bit of confusion for example, I may consider n to the power say 1 minus epsilon, divided by n; where epsilon may be a very small positive number, now here again this will go to 0 however, this will not be a negligible number. So, now, it is negligible in the sense that if I look at the order then in the terms of order of n is more than the order of n to the power 1 minus epsilon. So, (Refer Time: 17:59) justification can be given for the relative frequency definition.

Now, we consider a more rigorous definition of the probability; the first two definition which I have given they are based on the you can say basic they were based on the basic development of the subject of the probability itself for example, in the mathematical definition was developed as a consequence of the interest of some of the rambling houses to know probabilities of certain events and they contracted the mathematicians of that time and they looked at the entire thing has a finite set of outcomes, which may be equally likely and they of the a probability based on that.

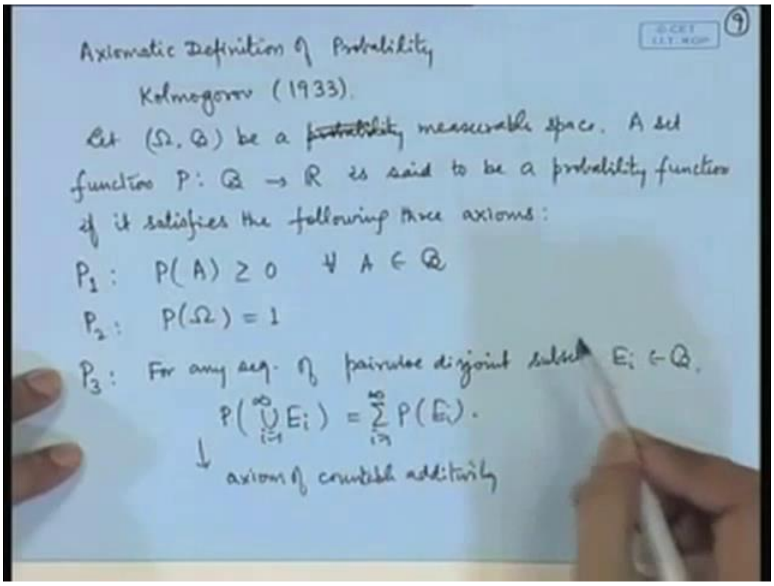
The relative frequency definitions are the statistical definition as a similarly based on the experience. So, when people were really looking on their statistical point of view certain events and they are not necessarily the events of the type where we have only head or tail or coin tossing etcetera, then they looked at that how many times the actually event occurs though in (Refer Time: 19:13) certain number of trails and we have seen that the definitions of drawbacks and therefore, they are not been eventually applicable; in order that vary the consistent or any (Refer Time: 19:26) applicable we need certain axioms.

So, in 1933 the Russian mathematician Kalmogorov he gave what is known as the axiomatic definition of probability and this is based on the set theoretic development which we gave on the very beginning that is on algebra source. So, now, we describe what is axiomatic definition of probability, what we have is that given a random experiment we have a sample space and now certain subsets of omega are the events, which may be of interest to us we may not be interested to consider all subsets of omega.

So, what we can consider is, we can consider a sigma field of subsets of omega. So, a sigma field subsets of omega will consist of certain events when they are unions the complementation, their intersections, their differences. In other words if we are considering certain events then all the manipulations of those events, which are of interest to the experimenter will be included in the set B and therefore, this definition of sigma field is useful in development of this definition or events the axiomatic definition.

So, given a random experiment we are considering a sample space and B is a sigma field of subsets of omega; that means, in the terminology of probability theory this is a set of events. So, omega B is called a measurable space and now we are interested to define the probability; so for every event which is included in B we must be able to define the probability function.

(Refer Slide Time: 21:37)



So, the axiomatic definition of probability it is due to Kalmogorov. Let omega B be a probability sorry be a measurable space. A set function P from B to R, R is a set of real numbers is said to be a probability function. If it satisfies the following three axioms; I will call them P 1 that is probability of A is greater than or equal to 0 for all A subset of B. Second axiom is that the probability of the sample space is equal to 1. The third axiom is that for any sequence of pair wise disjoint subsets E i 1 into B, probability of union E i is equal to sigma probability of E i, i is equal to 1 to infinity. This last axiom is known as the countable additivity axiom. The first axiom is known as the non negative (Refer Time: 24:09) axiom and the second axiom is known as the axima computeness.

So, we will continue from where we look at the probability properties of the probability function, which is given by the kolmogorova’s definition in the next class.

Thank you.