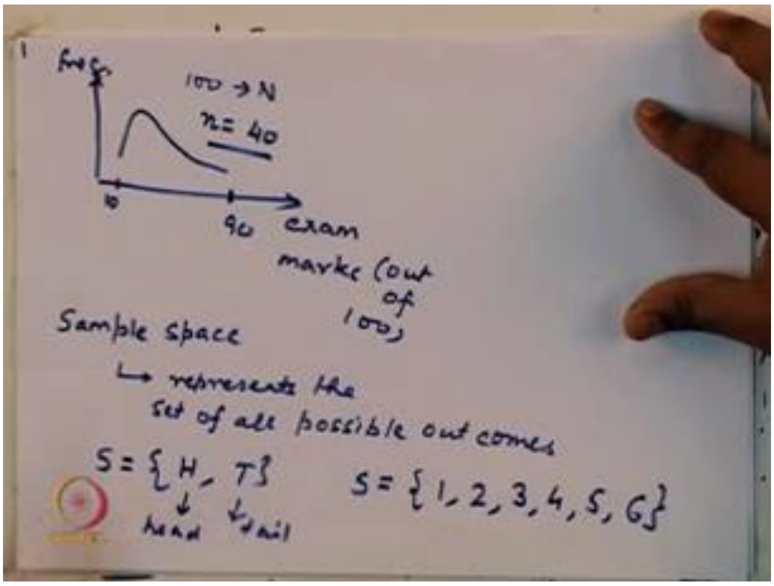
Introduction to Biostatistics

Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 15
Concept of Probability: introduction and basics

Hello and welcome to today’s lecture. In the last few lectures we had discussed how to collect data, how to represent data, and how to come up with useful statistical matrix of quantifying a data. And we also had a brief session on how to fit data depending on the nature of the curve? So, we have completed the first section of statistics, which is descriptive statistics and it brings us to our second section which is interferential statistics that is given a sample that you have collected you want to predict something about the whole population and that brings us to discussing one of the most important tools of statistics which is probability. So, if we think of why probability might be important.

(Refer Slide Time: 01:09)



Let us say you have collected the distribution of scores. So, this is exam marks and these are your frequency and let us just say this is how your distribution looks like. And let us say this is 10 and out of 100 and the maximum is 90. So, this is let us just say that you have collected this statistics not from the whole class, but you have asked this from the students who were present in a given day after the exam was over. So, this is a sample.

So, in a let us assume that out of 100 total students. So, N is 100 you have sample N is equal to 40 and this is the distribution that you get. So, based on this you can ask questions like, what is the chance that someone has scored 100 versus 30 versus 20 so on and so forth?

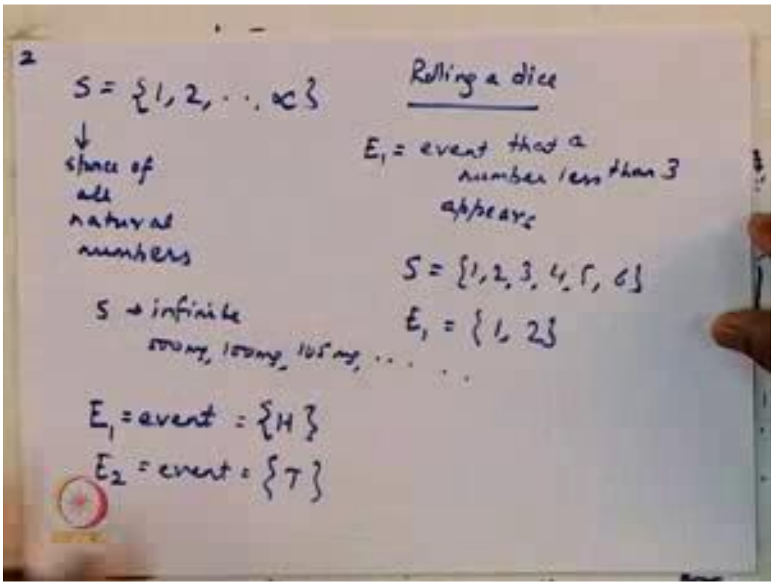
So, you know take let us take another example of that of a coin toss. You want to understand whether a coin is biased or unbiased. By unbiased I mean that when you flip a coin on an average you would expect to get at a tail or a head almost in equal proportion; that means, the chance or the probability of a head occurring should be nearly equal to the probability of a tale occurring which is into half, but if you were to do that experiment only thrice; that means, you flip the coin three times you know or four times right. Is there a chance that you will always get two out of two and you increase the sample size?

If you then do it for ten times let us say. So, if you do it only four times, it is a possible that you might have three heads in a tail or the other way round, but if you do it ten times right and let us say all ten times you are getting a head and not a tail at under any conditions. You are reasonably sure that the coin is not unbiased, but biased. So, probability is the tool which helps us to come up with computing chances or saying something about the population given a small sample.

So, when you come to probability we begin by definition of few terms; one is what is my sample or sample space? So, sample, so when you toss a coin let us say there are two possibilities or two outcomes possible either you have a head or you have a tail. So, the sample space is. So, it basically represents the set of all possible outcomes. So, in case of tossing a coin my sample space is typically written as S is nothing but a head appearing, H is short form for a head or a tail appearing, T is short form for a tail. So, now, imagine you roll a dice. So, our dice has six sides. So, when you roll a dice it is possible that you will get a 1, you can get a 2, and you can get a 3, a 4, a 5, or 6. So, there are in a sense six outcomes.

So, you will write the sample space for rolling a die as the number 1 appearing; the number 2 appearing, number 3 appearing, the number 4 appearing, the number 5 appearing and the number 6 appearing. So, this is your sample space. So, this is the entire sample space.

(Refer Slide Time: 05:24)

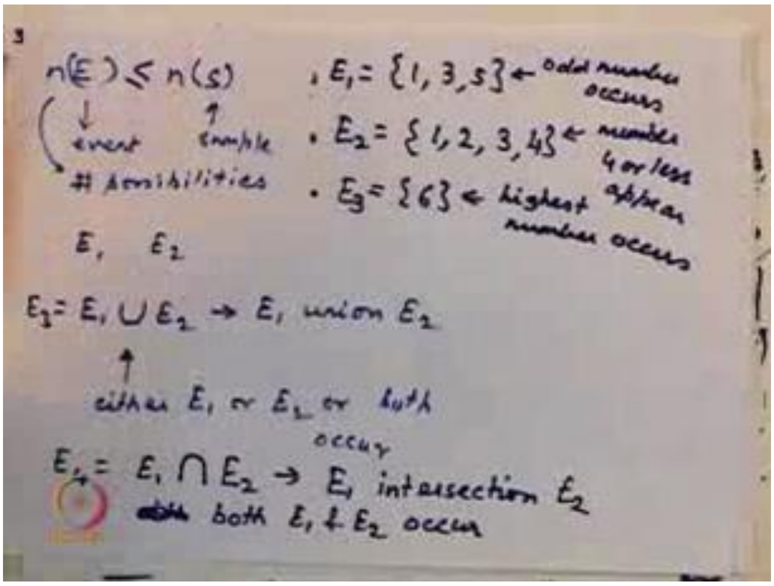


Now, let us say I say sample space S represents all natural numbers. So, my sample space all natural numbers are positive numbers greater than 0. So, my sample space will represent 1, 2, dot dot dot infinity. In other words, there are infinite possible outcomes which dictate. So, this is sample space is space of all natural numbers or let us say you can think of the dosage, you can administer to a patient with a certain drug. In that case also S has infinite possibilities right. You can give 500 milligrams, 100 milligrams; 105 milligrams so on and so forth. So, there are really infinite possibilities here.

So, a sample space can be infinite or can be finite as in case of a rolling of a die or tossing a coin. So, S is the list of all possible outcomes. Now let us say in case of a coin toss problem, you can define something of its interest to you is when a head appears. So, you can call E, E is short form for an event, which is associated with one particular outcome. So, in that case my event can be simply H or. So, I can write this as E 1 and I can write this another event E 2 as a tail appearing.

Now, let us go to the example of rolling a dice. So, I get, depending on the type of event that I want to define let us say, so I have a rolling a dice my say event E 1 is the event that a number less than 3 appears. So, in that case, so for rolling of a dice my sample space S is 1, 2, 3, 4, 5, and 6. My event E 1 is a number that less than 3 appear. So, I only have event there are two possibilities either you have a 1 or a 2 appearing. So, what you see? So, always event the number of possibilities in generally an event can be less or equal to the number of possibilities in case of a sample.

(Refer Slide Time: 08:21)

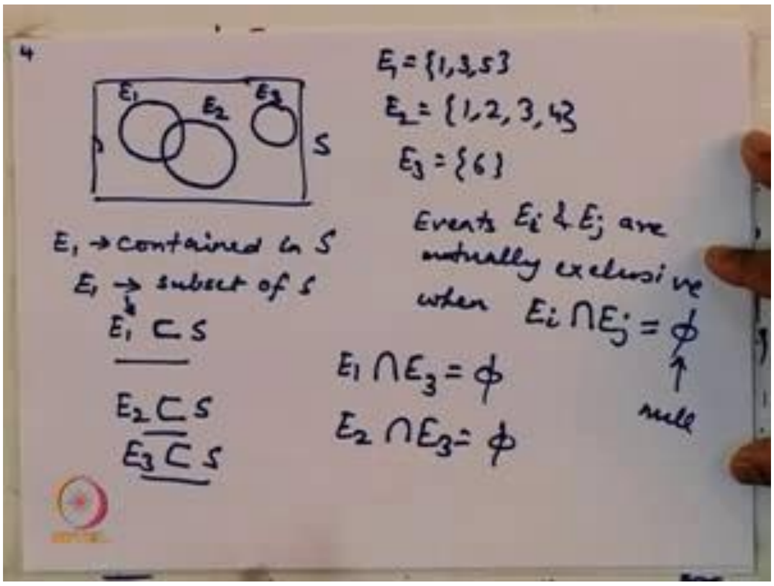


So, when I have. So, in the general case for any event E, the number of possibilities of E is less or equal to the number of possibilities in the sample space S. So, this is my sample and this is my event. So, n is a number of possibilities, now given two events let us say E 1 and E 2, I can think of a event which is represented by E 1 union E 2. So, this would be read as E 1 union E 2. So, this is interpreted is either E 1 or E 2 or both occur.

So, this is a union event let us say E 3 I have defined as the union of E 1 and E 2. It is either E 1 or E 2 or both occur. So, when you or I can have another event E 4, which is written as E 1, E 2. So, this as read as E 1 intersection E 2, which means either E 1 no. So, this means both E 1 and E 2 occur, so the possibility that both E 1 and E 2 are occurring. So, in the case of rolling of a dice, let us say if I define my E 1 as the set of all odd numbers appearing. In that case it will be 1, 3, 5 and E 2 is the set of all numbers less than 4; 1, 2, 3, 4.

So, basically numbers 4 or less occur and this is odd number occurs. So, you see there are few entries which are common to E 1 and E 2, like 1 and 3. There are some other entries which are present in E 1 like 5, which does not appear in E 2 and vice versa. Similarly the number 2 and 4 appear in E 2, but not in E 1. So that means, so it gives you a clue that depending on how you define an event you can have multiple possibilities and there might be overlap as in this case or let us see if I define the number and event E 3 is the highest number appears. So, which is nothing but 6. So, in this case there is nothing in common between E 2 and E 3 or E 1 and E 3. So, one of the easy ways of representing these phenomena is by using Venn diagrams.

(Refer Slide Time: 11:50)



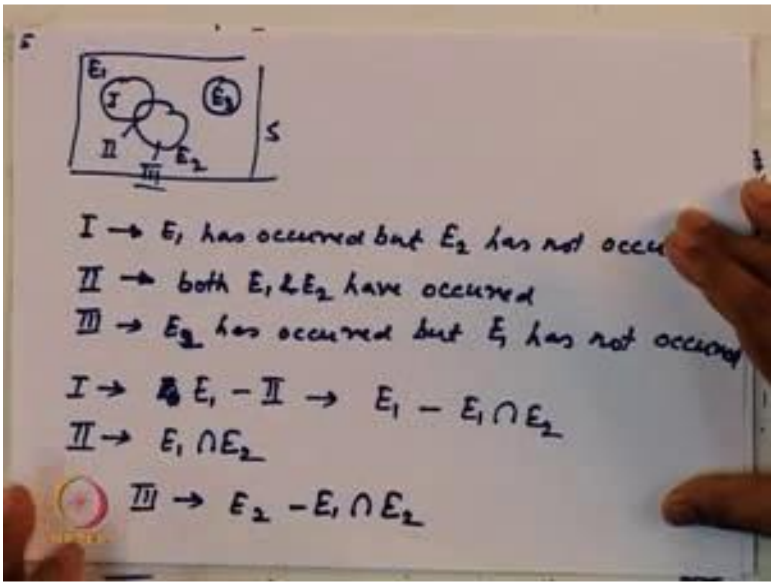
So, you can generally represent the whole sample space by a rectangular block. So, this is my S, I can represent an event E 1 like this by a circle. So, this conveys the message that E 1 is a subset of S.

So E 1, when E 1 is fully contained in S. So, we write that E 1 is a subset of S. So, I can also write this particular expression. So, this amounts to writing this term mathematically. So, I can represent an E 1, event E 1 I can represent another event E 2 and I can represent another event E 3 here. So, what were my previous events in the case of rolling a die? E 1 is 1, 3, 5 occurring, E 2 is 1, 2, 3 or 4 occurring and E 3 is 6 occurring. So, note the way I have drawn, all of these events E 1, E 2 and E 3 I represented as circles within this rectangle which means all these events that I have drawn fulfill this particular criteria, all are contained within the sample space.

Another important mission, so the way I have drawn E 1 and E 2 I have drawn with them some overlap between the events E 1 and E 2, but when I have drawn E 3 I have drawn it outside of E 1 and E 2. This gives you the idea of so what I can say is when E 3 occurs? So, the when number 6 occurs? I can say clearly that neither event E 1 or nor event E 2 have occurred. So, this is the case when I can say that events E 1 or E i and E j are mutually exclusive, when E i intersection E j is equal to phi; this is null. There is nothing in common shared between the events E 3 and E 2 or E 3 and E r.

So, I can write E 1 intersection E 3, E 3 is equal to null and E 2 intersection E 3 is equal to null. So, there is nothing in common between E 1 and E 3 or E 2 and E 3.

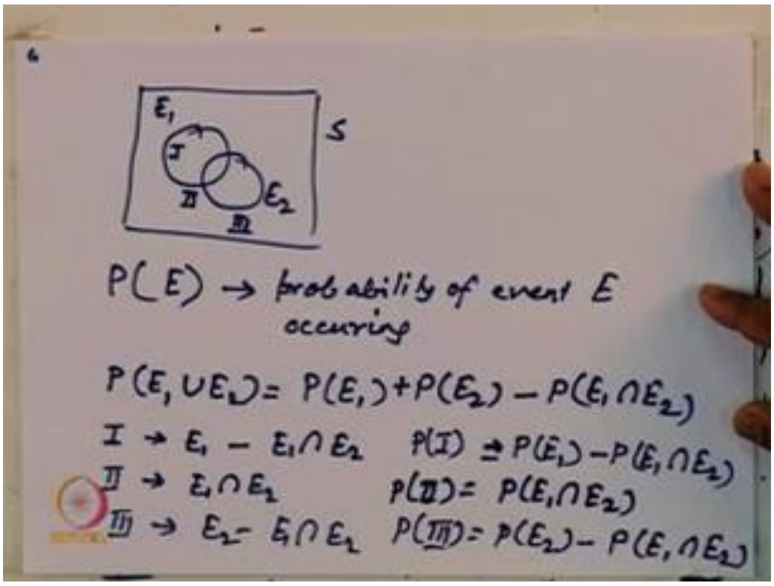
(Refer Slide Time: 14:56)



So, now let us again go back to the sample space. You have E 1 here, E 2 here and E 3 here and this is your whole sample space. So, let me divide this into 10 say, three segments I divide them into three segments labeled as 1, 2 and 3. So, what is the zone 1 implicates? Zone 1, E 1 has occurred, but E 3 has not occurred. The event 2 is you both E 1 and E 2 have occurred and similarly the event 3 like 1, E 3 has E 2 has occurred, but E 1 has not occurred.

So, using the concepts of union or intersection, I can write 1 is E 1. So, it is E 1 is the whole event. So, 1 is simply E 1 minus the area 2, let us find out what is the area two? Area two is nothing but E 1 intersection E 2 and. So, area 1 becomes E 1 minus E 1 intersection E 2. Similarly, the area 3 becomes E 2 minus E 1 intersection E 2. So, I can write down in the most general case.

(Refer Slide Time: 17:12)

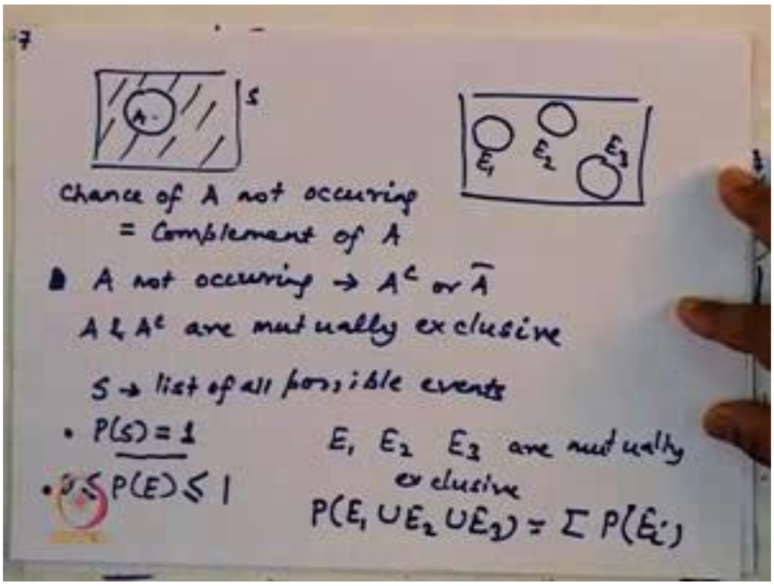


In the most general case if I draw two events, but there is overlap. Let us say this is my E 1, this is my event E 2.

So, probability of E is generally is this is basic this designates the probability of event E occurring. So, probability of E 1 union E 2 is given by the expression probability of E 1 plus probability of E 2 minus probability of E 1 intersection E 2. So, let us see how you can get this expression? We can use the earlier step right. So, we again break it down as before into three segments; 1, 2, 3. Your area 1 is E 1 minus E 1 intersection E 2. So, 2 is E 1 intersection E 2 and 3 is E 2 minus E 1 intersection E 2. So, if I calculate the probabilities and add them up, so I can get probability of event 1 is equal to probability of E 1 minus probability of E 1 intersection E 2.

Probability of 2 is equal to probability of E 1 intersection E 2 and probability of 3 is probability of E 2 minus probability of E 1 intersection E 2. So, if I add this three up, you have two negatives and one of them which is a positive, one of them will cancel each other you will have P E 1 plus P E 2 minus P E 1 intersection E 2. So, this is how you can you know compute the probabilities of composite events depending on how the events are defined, 6 (Refer Time: 19:38) write 7.

(Refer Slide Time: 19:44)



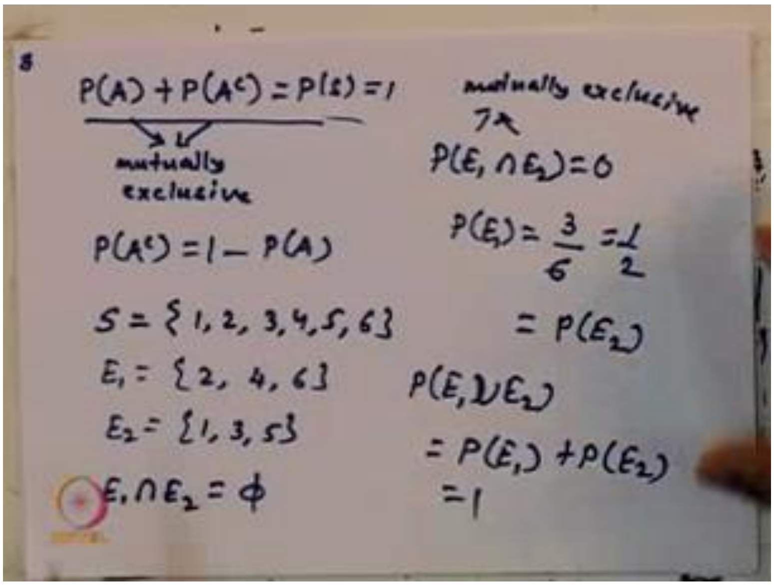
Now, if you have an event A let us say which is here this is my sample space. So, what is the chance of A not occurring? So, this is called the complement of A and this event, so event of A not occurring is denoted by A complement or A bar as the case may be. So, the probability of A not occurring is this area under the curve. All these possibilities are when A does not occur, this is when A occurs. So, I can write. So, I can clearly see. So, it brings us to the concepts of axioms of probability. So, when you have these two events. So, of course, A and A complement are mutually exclusive.

So, logically speaking, so what is the chance of probability of a samples occurring? So, sample is the list of all possible events. So, clearly, so I can say probability of S, it brings us to the chance that probability of S is equal to 1. If S is a list of all possibly outcomes then chance that S will occur is 100 percent. So, that is why you write the expression probability of S is equal to 1. In the generic case, so then the probability of an event E must fulfill. One particular criteria: first of all which I can say, 0 less is equal to probability of an event E less equal to 1.

So, in the case when your event is just the sample itself then your probability of the event is equal to 1 that is the highest bound. If you define an event where it does not occur under any case then 0 is the probability. So, in the case of when S is a list sample size of all natural numbers and you ask, what is the chance that the number minus 5 will occur? So, because minus 5 is not there in the sample at all conditions. So, you will have that probability to be 0, but in this case. So, this case which we do, probability of A is dividing this area by the total area you will get a fraction and hence this criteria is going to be solved. So, one axiom is probability of the sample space is equal to 1, the second axiom is probability of S is basically any event E is less equal to 1 and is also greater equal to 0.

So, I can write for events which are mutually exclusive. So, let us say E 1, E 2, E 3 are mutually exclusive. So, in this case, how will I represent E 1, E 2, E 3 in this, if they are mutually exclusive? So, this is my event E 1, this is my event E 2, this is my event E 3. So, for these events which are mutually exclusive I can write down the expression, what is the chance of probability of E 1 union E 2 union E 3? It is nothing but the summation of probability of the event E i, because there is no overlap. So, as we discussed you have this particular expression. So, when E 1 and E 2 are mutually exclusive then the area of the zone 2 reduces to 0. So, the probability of E 1 intersection E 2 is 0. So, as a consequence you will have E 1 union E 2 is equal to probability of E 1 plus probability of E 2. So, in case for case of exclusive events, I can write this particular expression.

(Refer Slide Time: 24:37)



So, next thing I can write for A and A complement, I can then write probability of A plus probability of A complement which are mutually exclusive. So, these events are mutually exclusive, but together they span the entire sample. So, this is nothing but the probability of the sample space itself is equal to 1. So, I can write probability of a complement is equal to 1 minus probability of A.

So, goes the few formula that we have derived let us take a specific example; the rolling of a dice. So, my sample space is 1, 2, 3, 4, 5, 6, let us say my event E 1 is the set of all even numbers. So, that will have 2, 4, 6 the set of E 2 is a set of all odd numbers which is 1, 3, 5 and so these are two events. So, I can clearly see that E 1 intersection E 2 is equal to phi there is no number which is common in these two events. So, I can write probability of E 1 intersection E 2 is equal to 0 because these are mutually exclusive. So, they are mutually exclusive, probability of the event E 1 occurring, and so probability of event E 1 occurring. So, there are three possibilities which is three and the total number of possibilities is six. So, probability of E 1 occurring is going to be half and this is same as the probability of the event E 2.

So, probability of E 1 plus E 2 union E 2 because these are mutually exclusive, is simply P of E 1 plus P of E 2 is equal to 1. So, we have in a sense confirmed that this expression that probability of A and probability of A complement is equal to 1. With that I stop our lecture today. So, what we have briefly discussed about the need for probability and some of the basic axioms of probability and how you can represent use Venn diagrams for describing events visually. So, you can understand what is what.

With that I thank you for today’s lecture and we look forward to our next lecture.

Thank you.