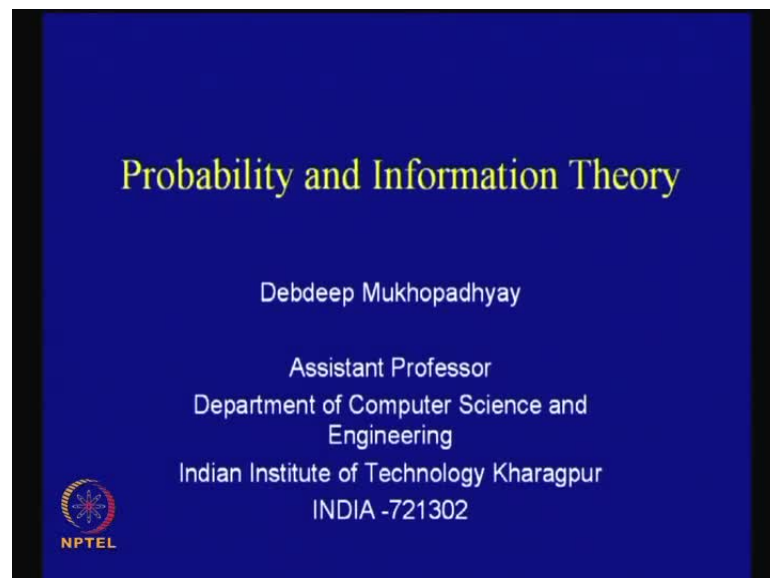


**Cryptography and Network Security**  
**Prof. D. Mukhopadhyay**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology, Kharagpur**

**Module No. # 01**  
**Lecture No. # 04**  
**Probability and information Theory**

(Refer Slide Time: 00:21)



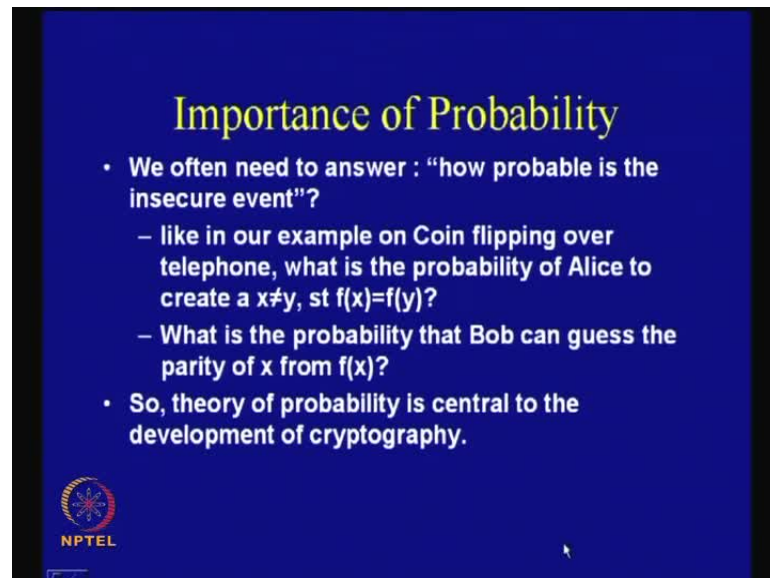
Welcome to today's class on probability and information theory. This is an extremely important field of study and a huge field. So, today we shall be trying to concentrate on some of the basic principles, which are necessary to understand the design and analysis of ciphers.

(Refer Slide Time: 00:35)



So, in today's talk, we shall be talking about the importance of probability in cryptography and then discuss about computational security, then follow it up with some discussions on binomial distributions and its applications, and very important birthday paradox, and then conclude with some concepts of entropy and information theory.

(Refer Slide Time: 00:54)



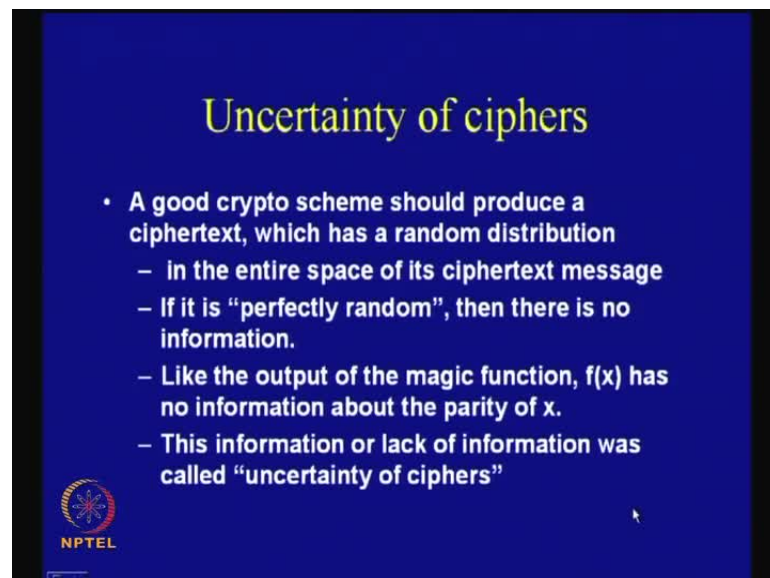
Now, as we have been discussing, if you remember in our first class that we are trying to answer the questions of this nature, like, how probable is the insecure event?

So, you remember that in our example on the coin flipping over telephone, the question was, what is the probability of Alice to create  $x$  which is not equal to  $y$ , such that,  $f(x)$  and  $f(y)$  are essentially the same? If you remember that the question was like, whether Alice is able to choose two different  $x$  and  $y$  values, such that, the outcome, which is denoted by  $f(x)$  and  $f(y)$  are the same?

The other question that we tried to address **was**, what is the probability that the Bob can guess the parity of  $x$ ? That was the important information and the question was, whether from the value of  $f(x)$ , Bob is able to extract out the information of the parity of  $x$ ?


Therefore, these types of questions will again and again appear, in the course of, when we are trying to design and also analyse the ciphers; therefore, the theory of probability is quite central to the development of this field.

(Refer Slide Time: 02:03)



**Uncertainty of ciphers**

- A good crypto scheme should produce a ciphertext, which has a random distribution
  - in the entire space of its ciphertext message
  - If it is “perfectly random”, then there is no information.
  - Like the output of the magic function,  $f(x)$  has no information about the parity of  $x$ .
  - This information or lack of information was called “uncertainty of ciphers”

 NPTEL

So, a good cryptosystem, as we discussed in the last to last classes, was that it should produce a cipher text, which has got a random distribution that means - we should look as much random as possible, to a distinguisher.

Therefore, I mean it should not be detectible, it should not be easily distinguished from a random distribution; so therefore, **which has to random** in the entire space of the cipher text message and the thing is that if it is perfectly random, let us leave it perfectly

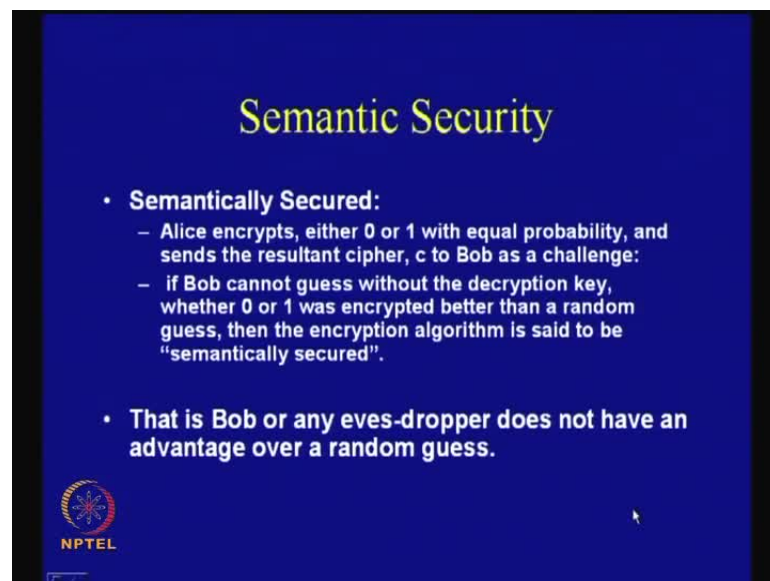
random; we will try to make this notion more and more concrete as we proceed. There should not be any information leakage.

So, lot of terms have been coined out here, like, information and also the notion of perfect randomness; we will try to make these things more concrete and more mathematical as we proceed.

Now, you just try to understand that intuitively; it means that when something is perfectly random, then essentially is not giving us any extra information; therefore, the idea is that it has got zero information content.


If you remember, we talked about the magic function  $f(x)$  and therefore, if the function  $f(x)$  does not leak any information or any other fact about the parity of  $x$ , then we say that it does not leak any information. Therefore, this information, or rather, the lack of information, is sometimes also referred to as uncertainty of ciphers. So, we will try to measure these terms, like, perfect randomness, information and also uncertainty.

(Refer Slide Time: 03:47)



**Semantic Security**

- **Semantically Secured:**
  - Alice encrypts, either 0 or 1 with equal probability, and sends the resultant cipher,  $c$  to Bob as a challenge:
  - if Bob cannot guess without the decryption key, whether 0 or 1 was encrypted better than a random guess, then the encryption algorithm is said to be "semantically secured".
- **That is Bob or any eves-dropper does not have an advantage over a random guess.**

 NPTEL

We rely heavily on the theory of probability and also develop theory of information using these probabilistic notions. So, another important concept, although we shall not be really going deep into this, is something - the concept of provable security.

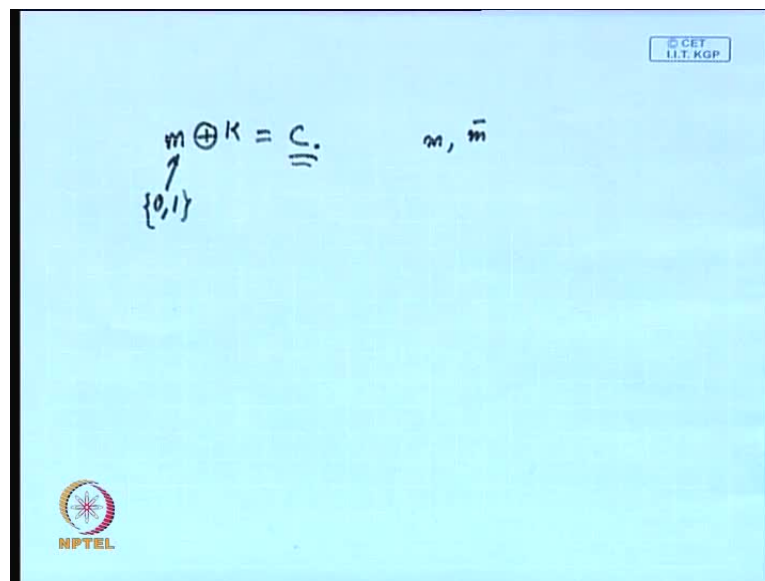
So, something which is called semantic security is very central, and it is defined as follows: remember, Alice and Bob, therefore, the idea is that Alice essentially encrypts

either 0 or 1 with equal probability; therefore, Bob knows either 0 is encrypted or 1 is encrypted and the probability of choosing a 0 and 1 is half.

Now, Alice encrypts these either 0 or 1 and sends the resultant cipher  $c$  to Bob as a challenge. Now, Bob has to guess from the challenge without the decryption key that whether 0 was encrypted or 1 was encrypted.

So, if Bob is not provided with a cipher text, then what will Bob does? Bob will simply guess. Now, when Bob is provided the cipher text, he should not be able to guess any better than the random guess; so, that is the idea. If he is not able to guess any better than the random guess, then we say that the encryption algorithm which Alice is using is semantically secured.

(Refer Slide Time: 05:12)

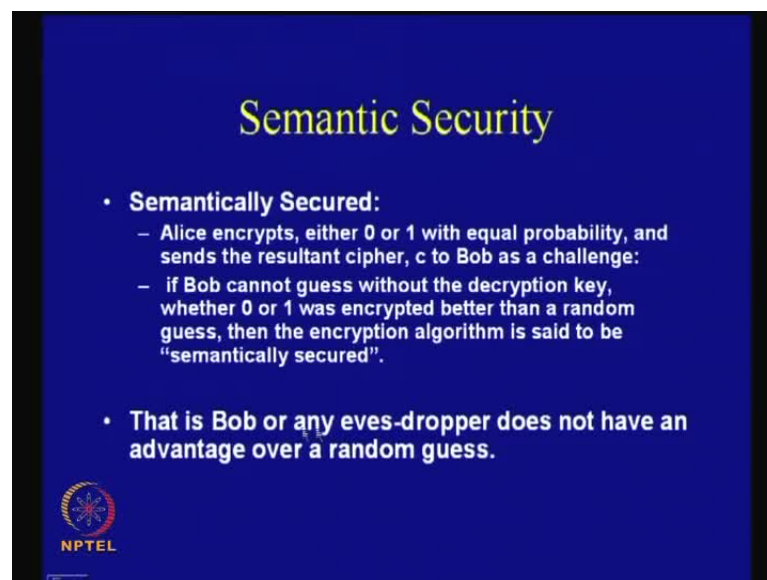


If we just consider a simple kind of stream cipher, for example, if there is a kind of message  $m$ , what Alice does is that it chooses randomly a key value and it creates a cipher text; therefore, this message can either be 0 or 1. So, the key will also be either 0 or 1 and it is randomly chosen.

Now, this particular cipher text has been provided to Bob and Bob from this cipher text value has to guess, whether 0 was encrypted or whether 1 was encrypted. This is an encryption algorithm, which Alice uses. Bob is either receiving  $m$  or it is receiving  $\bar{m}$ , because the key can either take 0 value or 1 value; we are just considering a very simple case.


So, Bob has to guess whether 0 has been encrypted or whether 1 has been encrypted and if Bob does not have any information of the cipher text, then Bob would have simply guessed. So, in this case also, when Bob is even provided with the cipher text, the probability of Bob being able to guess whether the message is 0 or 1 should even, then be close to half; so, that is the idea of semantic security.

(Refer Slide Time: 06:18)



**Semantic Security**

- **Semantically Secured:**
  - Alice encrypts, either 0 or 1 with equal probability, and sends the resultant cipher,  $c$  to Bob as a challenge:
  - if Bob cannot guess without the decryption key, whether 0 or 1 was encrypted better than a random guess, then the encryption algorithm is said to be "semantically secured".
- That is Bob or any eves-dropper does not have an advantage over a random guess.

  
NPTEL

So, Bob or any eves-dropper does not have any advantage over the random guess; so semantic security tries to encapsulate or capture this particular notion.

(Refer Slide Time: 06:32)



We have been discussing about something called message indistinguishability, that is, semantic security and message indistinguishability are the same notions; these notions say that the attacker is not able to distinguish between the encryptions of either a 0 or a 1.

Often we talk about something which is called computational security. We will see that there are two ways, you can model the attacker, you can either assume that the attacker is an unbounded adversary, that is, he is all powerful; it has got large access to large amount of resource and it can do large computations; so, that is the notion of an unbounded attacker.

So, you can make your security algorithms, which are protected against an unbounded attacker, you can try to do that. The other thing which you can do is that you can assume that in today's computational power or today's world, the maximum computations which an adversary can do – like, what is believed in today's thumb rule is that, if there is any algorithm which requires more than  $2^{80}$  computations, then it is termed as infeasible.

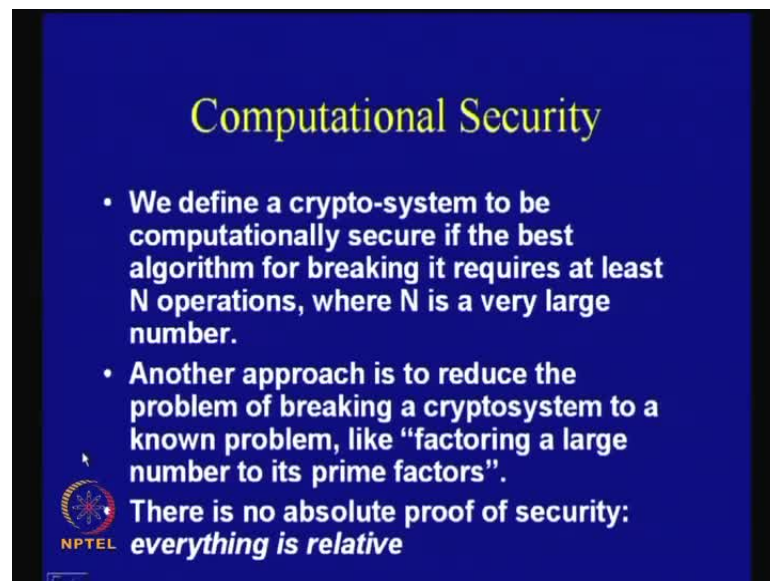
So, if they can limit or bound the attacker of today's world by, say,  $2^{80}$  computations, and if we can prove that the particular attack for an encryption algorithm requires more than  $2^{80}$  computations, then we are happy as a designer.

So, that is the idea of our computational security analysis, that is, I am not trying to really give you guarantees of security against an unbounded attacker, but we are trying to give you security against a bounded adversary - an adversary, which is bounded by computational power.

The other advantage that you arrive apart from the simplicity is, often you will find that when we are trying to give security guarantees against an unbounded adversary, you may end up in an encryption algorithm or a technique which is not practical.

But as we discussed, we also need practical security, that is, the cipher should be practical; so, in order to make it practical, it is often advantageous to assume that the adversary is actually not unbounded by this kind of bounded.

(Refer Slide Time: 08:51)



**Computational Security**

- We define a crypto-system to be computationally secure if the best algorithm for breaking it requires at least  $N$  operations, where  $N$  is a very large number.
- Another approach is to reduce the problem of breaking a cryptosystem to a known problem, like “factoring a large number to its prime factors”.

**There is no absolute proof of security:  
everything is relative**

NPTEL

So, that is the notion and importance of computational security analysis; we try to make this idea more clear as we proceed. Therefore, we define a crypto-system to be computationally secure if the best algorithm for breaking it requires at least  $N$  operations, where  $N$  is a very large number.

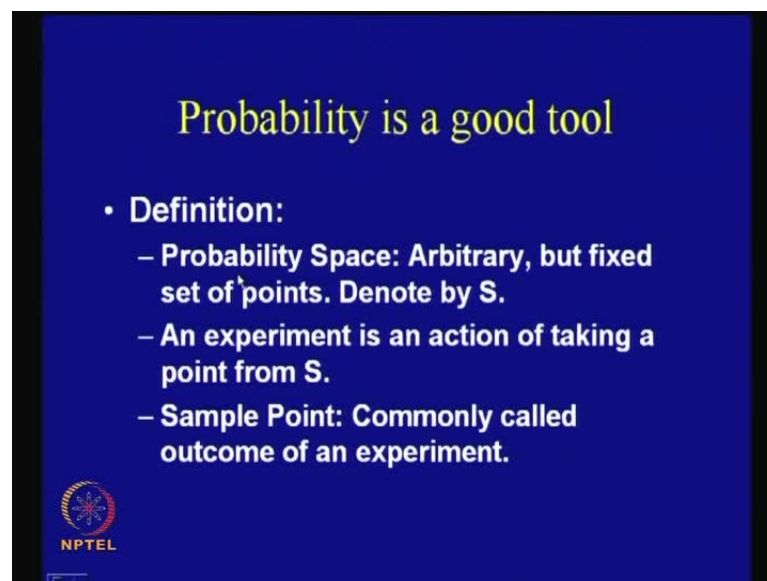
Another approach is to reduce the problem; therefore, this is another approach which is often taken to reduce the problem of breaking a cryptosystem to a known problem, like, “factoring a large number to its prime factors”.



It is often assumed that factoring a large number to its prime factor is a difficult problem; this is a quite hard problem. Although, we do not have real proofs for proving this fact, but the idea is that this particular problem has withstood large number of analysis and withstood a long period of time and it is believed that it is fairly hard.


So, the approach which is taken is that when a crypto-system is given, for example, if I just take the example of an asymmetric crypto-system called RSA, and somebody asks me to prove the security of this system, the approach which is adopted is, to reduce this problem of proving, or reduce this problem of breaking this RSA crypto-system to this known problem; that is, we do a reduction proof and show that if this problem, **the idea is that if this problem** - is a difficult problem, so is the problem of breaking this RSA crypto-system.

(Refer Slide Time: 10:17)



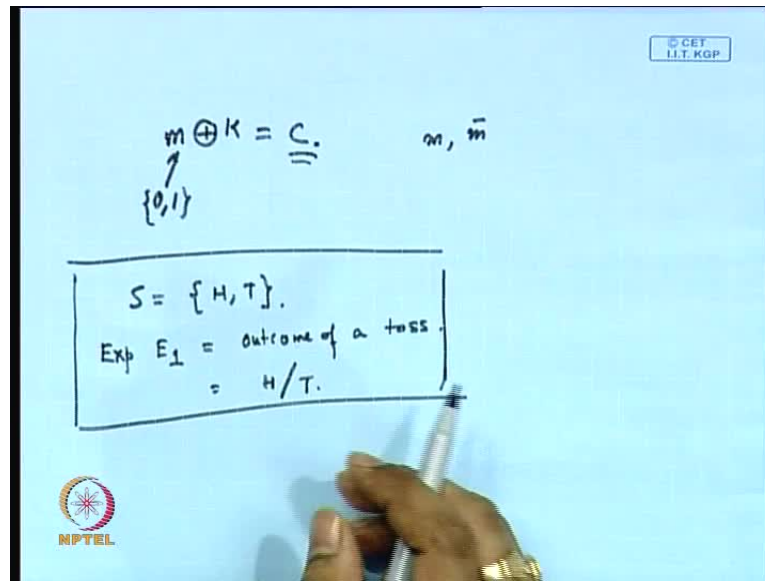
**Probability is a good tool**

- **Definition:**
  - **Probability Space: Arbitrary, but fixed set of points. Denote by S.**
  - **An experiment is an action of taking a point from S.**
  - **Sample Point: Commonly called outcome of an experiment.**

  
NPTEL

Thus these kinds of proofs are relative and they are not absolute and they are sometimes called, proof by reduction; so we will also see some examples of such kind of security proofs in our class also. Probability is a good tool and it is a tool which helps us to analyse the ciphers. So, let us try to make the concepts little bit more well-defined. So, there are some important definitions, one of them is probability space. The probability space is a fixed set of points which is arbitrary, that its kind is denoted by  $s$  often.

(Refer Slide Time: 10:49)



So, let us consider an example – suppose, there is an unbiased coin; so the unbiased coin can take two possible values. Therefore, we define its sample space to be the values like, head and tail; so, it can take either head value or it can take either tail value. Then, we define an experiment; so, the experiment is defined as follows - experiment  $e_1$  is nothing but the outcome of a toss. So, if we assume that this is an unbiased coin, then this experiment, what it does is that it chooses or samples out a point from the sample space.

Therefore, this result can either be a head or it can be a tail; so, it just chooses a point from this head and tail. Similarly, you can actually make this example a little bit more complicated and even for this simple example, I think we can actually understand the concepts quite well.


So, if I just make it little general, then it will look like that the probability space are arbitrary, but fixed set of points. For example, the head and tail, it could be more than that also and we denote that by  $S$ .

What the experiment does is that, it is an action of taking a point from  $S$ ; so, it just chooses a random point from the sample space, this sample point, which is commonly called outcome of an experiment.

(Refer Slide Time: 12:15)

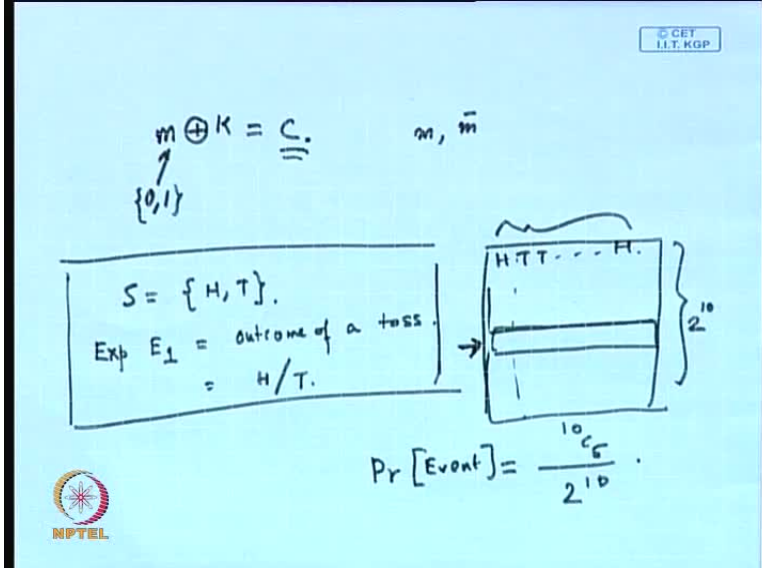
### Tossing an unbiased Coin

- Two possibilities of an experiment are Head or Tail
- An experiment is "toss the coin for 10 times"
- Event is 5 times head, 5 times tail.
- Probability of the event is:  $\frac{\binom{10}{5}}{2^{10}}$




So, you toss a coin, it is either head or a tail; so, either head or the tail is the sample point of the experiment. Let us try to make little bit more complicated, so you have got two possibilities, head or tail, but what you do is that you toss the coin for ten number of times.

(Refer Slide Time: 12:38)



The diagram shows the following components:

- Top left:  $m \oplus K = C$  with  $\{0,1\}$  below it.
- Top right:  $m, \bar{m}$
- Left box:  $S = \{H, T\}$ .  
Exp  $E_1 =$  outcome of a toss.  
 $= H/T$ .
- Right box: A sequence of outcomes  $H T T \dots H$  with a bracket on the right labeled  $2^{10}$ .
- Bottom right:  $Pr [Event] = \frac{{}^{10}C_5}{2^{10}}$ .

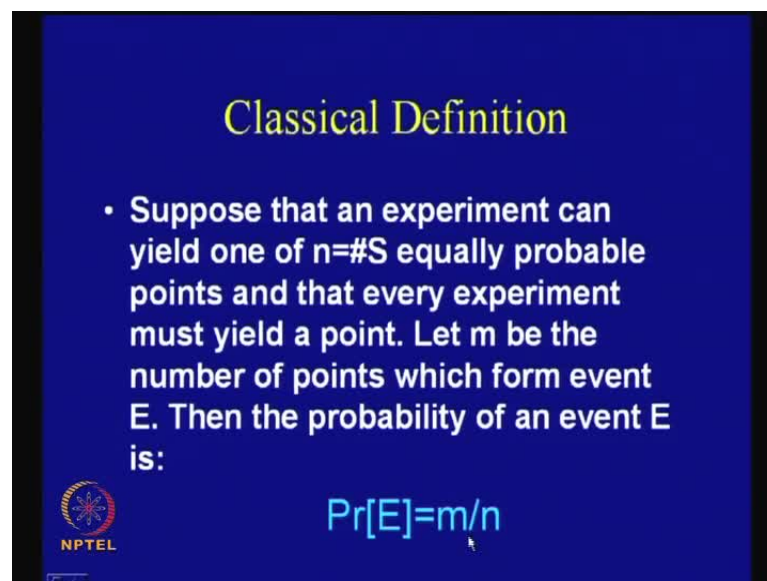


If you toss the coin for ten number of times, then there are several possible outcomes. So, if you toss the coin for ten number of times, then it could be like, it is a sequence of heads and tails - it can be head tail, tail, tail, and so on, tail head or something like that.

You denote the head by 0 or the tail by 1, you know that there are actually how many possible enumerations or possible outcomes? There are 2 power of 10 possible outcomes. Now, if I denote the particular kind of event as, saying that the end is five times head and five times tail, then it means, that from all the sample points, I am interested in the probability of one particular event.

So, this particular event, like this head and tail can occur actually in certain possible ways. So, we can actually compute the number of times, exactly five times head falls by simply computing 10 choose 5. Therefore, probability of this event, that is, the probability of the event that there are 5 heads will be nothing but 10 choose 5 divided by 2 power of 10; so, this way of computing the probability is something that we have quite seen.


(Refer Slide Time: 13:46)



**Classical Definition**

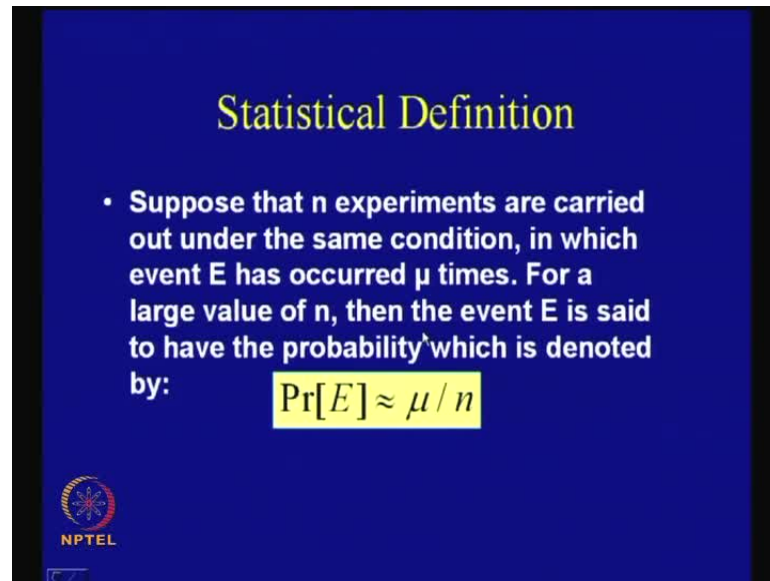
- Suppose that an experiment can yield one of  $n = \#S$  equally probable points and that every experiment must yield a point. Let  $m$  be the number of points which form event  $E$ . Then the probability of an event  $E$  is:

$Pr[E] = m/n$

 NPTEL

Therefore if I make it, or rather, define it, this is the classical definition of probability, that is, if there is an experiment, which yields one out of  $n$  possible equally probable points and that every experiment must yield a point of course, and let  $m$  be the number of points which form event  $E$ , then the probability of an event  $E$  is defined as the ratio of  $m$  and  $n$ .


(Refer Slide Time: 14:08)



**Statistical Definition**

- Suppose that  $n$  experiments are carried out under the same condition, in which event  $E$  has occurred  $\mu$  times. For a large value of  $n$ , then the event  $E$  is said to have the probability which is denoted by:

$$\Pr[E] \approx \mu / n$$

 NPTEL

There is a statistical definition also. Now, why do we require this statistical definition, if something we can easily understand? For example, if I tell you that there is an unbiased coin and if I tell you that the probability of a head is half, then that does not mean that if I toss the coin for 2 number of times, then one of them will be a head, it does not mean that. What it means is that if you keep on tossing the coin for a large number of times, then the half of the number of times you toss will be, at least, approximately half will be head.

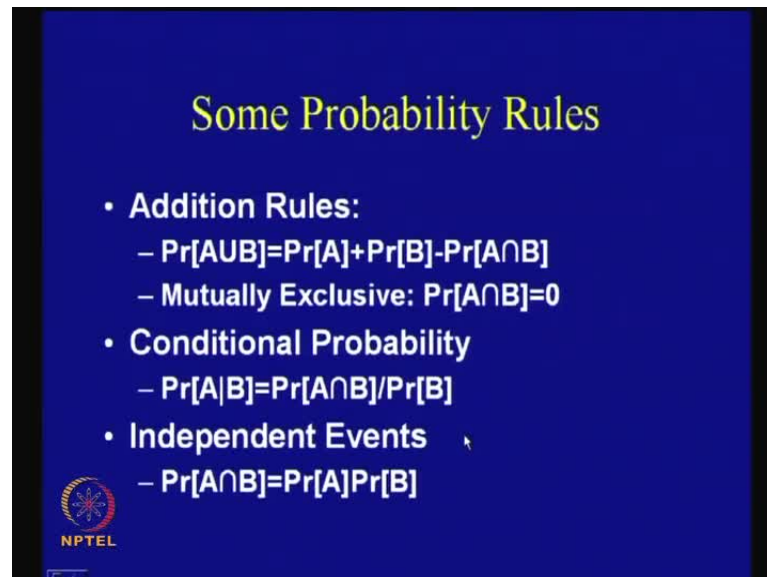
So, the notion of probability actually holds in reality, when you repeat the experiment for a large number of times; so, there is a notion of statistics involved in the way we are defining probability.

Suppose, there are  $n$  experiments and these are carried out under the same conditions in which the event  $E$  has occurred  $\mu$  times. So, for a large value of  $n$ , that means, this is important, that if I repeat the experiment for a large number of times, then the event  $E$  is said to have the probability which is denoted by probability of  $E$ , approximately equal to  $\mu$  by  $n$ .

So, what I do is that instead of computing the probability by finding out the outcomes and the number of possible ways these outcomes can come, what we can do is that we can keep on repeating the experiment.


We find that, if we actually repeat the experiment for a large number of times and out of the  $\mu$  number of times the particular event has occurred, then we can fairly approximate its probability by the ratio of  $\mu$  by  $n$ ; so, this is the statistical definition and often useful for analysis.

(Refer Slide Time: 15:45)



**Some Probability Rules**

- **Addition Rules:**
  - $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$
  - Mutually Exclusive:  $\Pr[A \cap B] = 0$
- **Conditional Probability**
  - $\Pr[A|B] = \Pr[A \cap B] / \Pr[B]$
- **Independent Events**
  - $\Pr[A \cap B] = \Pr[A] \Pr[B]$

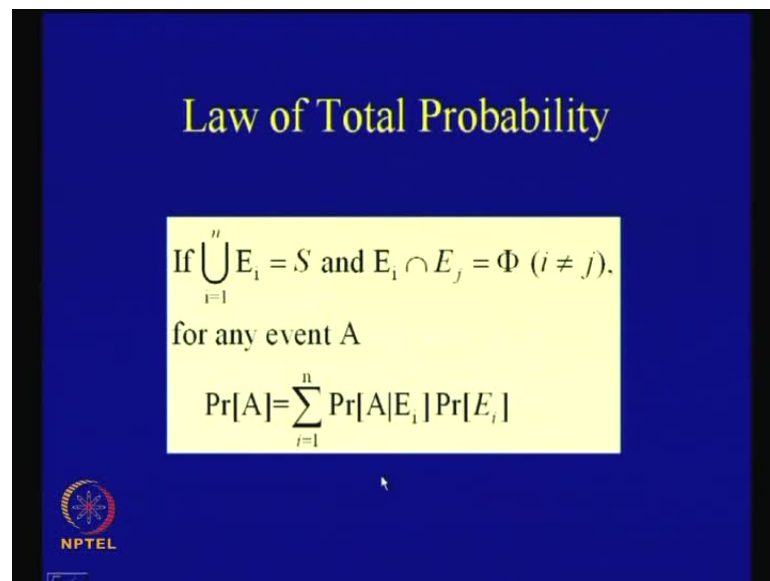
 NPTEL

Some of the very elementary probability rules are as follows, now, we know this, but this is a kind of recapitulation; that is, probability of A union B is equal to probability of A plus probability of B minus probability of A intersection B; if they are mutually exclusive, that we know, that probability of A intersection B works out to zero and therefore probability of A union B is equal to probability of A plus probability of B. Now, this particular rule often handy in giving us upper bounds of the probability of A union B, because we can say that probability of A union B will be lesser than equal to probability of A plus probability of B, because this probability is definitely greater than or equal to 0. So, we know the definition of conditional probability that probability of A, given B, so this probability means that we know the event B has occurred, what is the probability of A?

The way it is being computed is, probability of A intersection B divided by probability of B. Now, if A and B are independent, then what is the probability of A, given B? That is, it is same as probability of A because A does not really depend upon the outcome of B.

So, the probability of A given B becomes equal to probability of A when they are independent and therefore probability of A intersection B becomes equal to probability of A multiplied by probability of B. So, that is, if A and B are independent, then probability of A intersection B is nothing else, probability of A multiplied by probability of B.

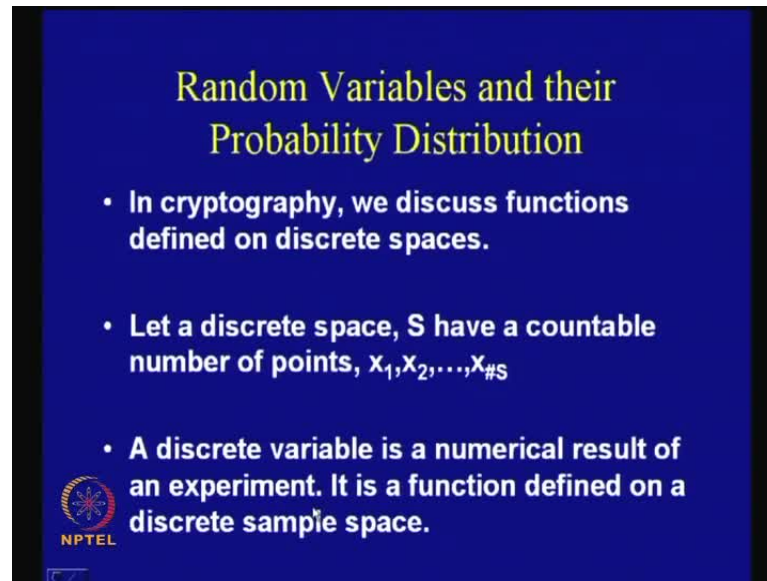
(Refer Slide Time: 17:17)



The slide features a dark blue background with a yellow rectangular box in the center containing text and a mathematical formula. The title 'Law of Total Probability' is written in yellow at the top. Below it, the conditions for the law are stated: 'If  $\bigcup_{i=1}^n E_i = S$  and  $E_i \cap E_j = \Phi$  ( $i \neq j$ ), for any event A'. The formula  $\Pr[A] = \sum_{i=1}^n \Pr[A|E_i] \Pr[E_i]$  is displayed in black. In the bottom left corner, there is a small circular logo with a star and the text 'NPTEL' below it.


Then there is a very important law, which is known as the law of total probability. If there are n events like E 1, E 2, and so on till E n and the union of them is a sample space S. If I know that they are mutually non-intersecting, that is, E i and E j cannot take place together; therefore, the intersection of E i and E j is equal to the null set. Then, for any event A, we can say probability of A is equal to, So what we do is, we multiply probability of E i, that is, the probabilities that the ith event has occurred with probability of A, given E i; that is, the probability of A. Given that the event E i has occurred, and then take a sum over all possible i values; so, i runs from 1 to n. So, this is also a very handy rule for doing our computations.

(Refer Slide Time: 18:13)



**Random Variables and their Probability Distribution**

- In cryptography, we discuss functions defined on discrete spaces.
- Let a discrete space,  $S$  have a countable number of points,  $x_1, x_2, \dots, x_{\#S}$
- A discrete variable is a numerical result of an experiment. It is a function defined on a discrete sample space.

 NPTEL

When we are talking about cryptography, we are not talking about a continuous probability space; we are actually talking about discrete space. Therefore, in this case, the total possible sample points can actually take some discrete possible outcomes; so it can run from  $x_1$  to say, till  $x_{\#S}$ ;  $\#S$  is nothing but the number of elements in the sample point.

What we do is, that for each of these sample points, we actually define a probability. So, first of all we define something which is called a random variable and then we say that this particular random variable can take such and such possible values; then, we try to assign a probability for this random variable. So, if there is a discrete space  $S$ , which has got a countable number of points like  $x_1, x_2$  so on till  $x_{\#S}$ , then a discrete variable is nothing but a numerical result of an experiment.

So, it is a function which is defined on a discrete sample space. For example, if I say that example of a discrete variable could be like, if there are some points like  $x_1, x_2$  and  $x_{\#S}$ , suppose, we define a function, like suppose the value of  $x_i$  and imagine that all of them are binary values. Suppose, there are four bit binary value, so there could be sixteen possible such numbers like, 000 00 00 to all 1s to all four 1s. We say that we are interested in finding out the probability, that is, the numbers are, for example, even; so, we know that what we will do is, from all these possible values we will try to find out




those binary values which end with zero. So, we define an experiment and therefore, this is a function which is defined on a discrete sample space.

(Refer Slide Time: 20:04)

**Random Variables and their Probability Distribution**

- Let  $S$  be a discrete probability space and  $X$  be a random variable (r.v).
- A discrete probability function of  $X$  is of type,  $S \rightarrow \mathbb{R}$  (set of reals), provided by a list of probability values:  
 $\Pr[X=x_i]=p_i$  ( $i=1,2,\dots,\#S$ ), st

$i) p_i \geq 0;$   
 $ii) \sum_{i=1}^{\#S} p_i = 1$



So, what we do is that now, let  $S$  be a discrete probability space and  $X$  be a random variable; so,  $X$  is the random variable, and we actually start assigning probability values for these random variables.

So, what are the possible values this random variable  $X$  can take? It can either take  $x_1$  value,  $x_2$  value, until  $x_{\#S}$ , and what we do is that for each of these possible values of the random variable, we assign a probability.

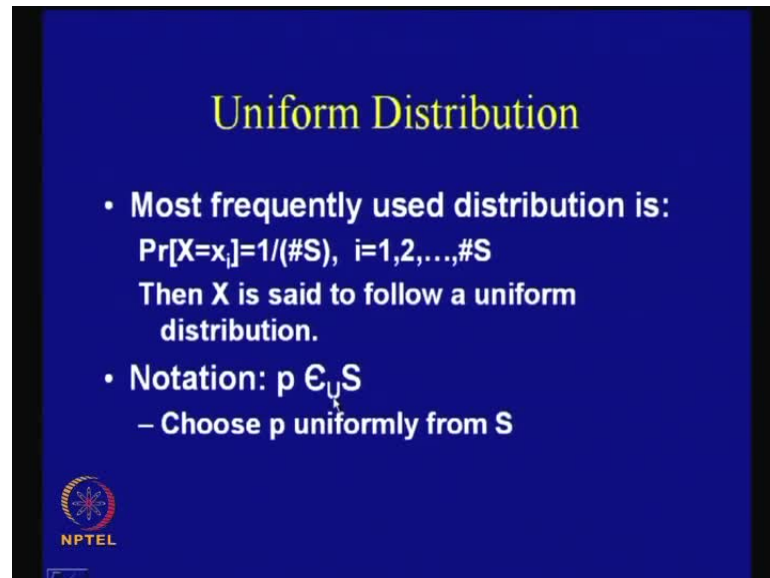
So, we say the random variable  $X$  can take the value of  $x_1$  with a probability of say,  $p_1$ , the random variable  $X$  can take the value of  $x_2$  with a probability of  $p_2$ . The random variable  $X$  can so on take the value of say  $x_{\#S}$  with a probability of  $p_{\#S}$ ; so, we can define the probabilities like this.

So, these probabilities will be discrete probabilities and it needs to satisfy two important properties - one of them is that each of this probability value should be greater than or equal to 0, that is, they should not be negative, and other thing is that the summation of all these probabilities should be equal to 1.

Therefore, these probability should satisfy these two properties and we can actually say that these probabilities are essentially a map from the sample space  $S$  to the set of real


numbers, because the probability values are nothing but the real numbers which lies between 0 and 1, both sides included; so, these are actually the individual probabilities and this should satisfy these two important properties.

(Refer Slide Time: 21:42)



**Uniform Distribution**

- **Most frequently used distribution is:**  
 $\Pr[X=x_i]=1/(\#S), i=1,2,\dots,\#S$   
Then X is said to follow a uniform distribution.
- **Notation:  $p \in_U S$**   
– Choose p uniformly from S

  
NPTEL

One of the very frequently used distribution is something which is called uniform distribution and we know that in this case, all the values like  $x_1$  till  $x_{\#S}$  are equally probable.

So, the random variable X can take  $x_1$  or  $x_2$  or  $x_3$  or  $x_{\#S}$  with the same probability and the probability is nothing but 1 divided by  $\#S$ . Then X is said to follow a uniform distribution and it is often denoted like this, that is, suppose, say that p choosing uniformly from s; therefore, it means that choose p uniformly from S. So, this is a very common notation which is useful.

(Refer Slide Time: 22:18)

**Binomial Distribution**

- Suppose an experiment has two possible outcomes, HEAD (success) or TAIL (failure)
- Repeated independent such experiments are called Bernoulli Trials
- $\Pr[H]=p, \Pr[T]=1-p$

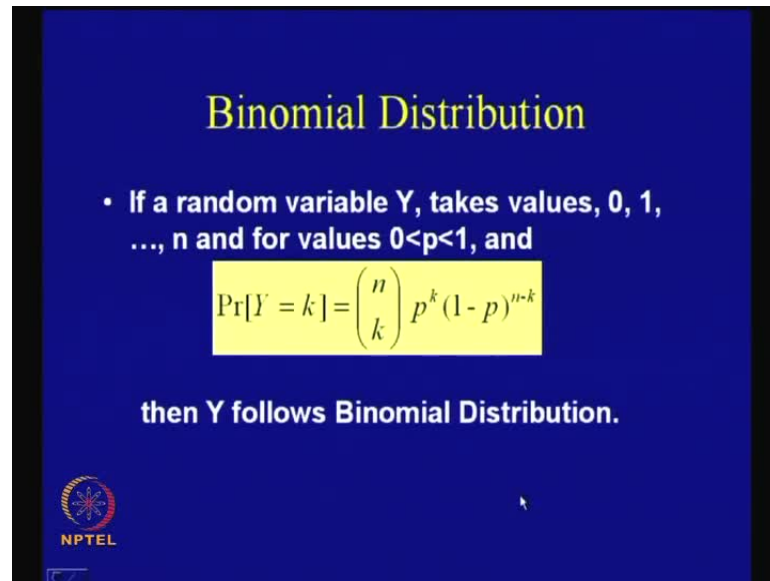
$\Pr[k \text{ "success" in } n \text{ trials}] = \binom{n}{k} p^k (1-p)^{n-k}$

NPTEL No of ways of choosing k points out of n

Then we actually define a very important distribution which is called a binomial distribution. It says that, suppose, there is an experiment and it has got two possible values or possible outcomes, like, it could be either a success or it could be either a failure or a HEAD or a TAIL, then we repeat the experiments independently. Such experiments and these are called Bernoulli Trials, and if I denote the probability of a HEAD to be  $p$  and the probability of a TAIL to be  $1 - p$  of course, because the HEAD and the TAIL together, if I add these two probabilities, should be equal to 1, what is the probability that there are  $k$  successes in  $n$  trials?

So, I repeat the experiments and the question is, what is the probability that among the  $n$  trials, there are actually  $k$  number of successes? We know, this actually works out to this, that is,  $n$  choose  $k$  which means that we choose the  $k$  success points and then for  $k$  success points, the probability will be actually  $p$  to raise to the power of  $k$ . Since, the other points are failure points, so they should be also multiplied by one minus  $p$  whole to the power of  $n$  minus  $k$ , because there are  $n$  minus  $k$  failure points; so, this gives us the number of ways of the probability that there are  $k$  success in  $n$  trials.

(Refer Slide Time: 23:36)




**Binomial Distribution**

- If a random variable  $Y$ , takes values, 0, 1, ...,  $n$  and for values  $0 < p < 1$ , and

$$\Pr[Y = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

then  $Y$  follows Binomial Distribution.

 NPTEL

Now, if a random variable  $Y$  takes values like 0 1 to  $n$ , so 0 1 to  $n$  means - if I say that this success, that is, the number of successes in  $n$  trials and I denote that by a random variable, then this random variable can take values from 0, that is, it can be that there are no success to  $n$ , that is, all of them are success. Therefore, the random variable  $Y$  can take values from 0 to  $n$  and for values, which lie between 0 to 1, that is, for  $p$ , the probability that  $Y$ , and we say that, the probability that  $Y$  is equal to  $k$  which means that there are  $k$  success is given by,  $n$  choose  $k$   $p$  to the power of  $k$  multiplied by 1 minus  $p$  whole to the power of  $n$  minus  $k$ .

If this random variable satisfies this probability distribution, then we say that  $Y$  follows binomial distribution and this is the very common and useful distribution.

(Refer Slide Time: 24:34)

The slide has a dark blue background. At the top, the title 'Law of large Numbers' is written in yellow. Below it are two bullet points in white text. The first bullet point says 'Repeat a trial for a large number of time (n → infinity) and note the number of success.' The second bullet point says 'After a point the number of success will remain constant and equal to np (often referred to as the Expected number of success) or the Expectation of the r.v.' Below the text is a yellow rectangular box containing the mathematical formula  $\lim_{n \rightarrow \infty} \Pr\left[\left|\frac{\xi_n}{n} - p\right| < \alpha\right] = 1$ . To the left of the formula is the NPTEL logo. To the right of the formula is a small green box with the text 'α: small fixed number' in red.

**Law of large Numbers**

- Repeat a trial for a large number of time ( $n \rightarrow \infty$ ) and note the number of success.
- After a point the number of success will remain constant and equal to  $np$  (often referred to as the Expected number of success) or the Expectation of the r.v.

$\lim_{n \rightarrow \infty} \Pr\left[\left|\frac{\xi_n}{n} - p\right| < \alpha\right] = 1$

α: small fixed number

Then we talk about something called law of large numbers very useful. So, it says that - if we repeat a trial for a large number of times, where  $n$  is suppose infinity and  $n$  tends to infinity and we note the number of successes, after a point the number of success will actually remain a constant and will be computed by something, which is called the expectation. The expectation is nothing but the number of times you are repeating the experiment multiplied by the probability of success.

The  $p$  is the probability of success; so this is often referred to the expectation of the random variable. As we stated, we have a limit  $n$  tends to infinity, the probability that  $\epsilon/n$  by  $n$  minus  $p$  is lesser than a very small number – small, but fixed number - and this probability is equal to one. Therefore, this is something which is called the law of large numbers, which says that if you repeat the experiment for a large number of times, then essentially you will find that the number of times you get the number of success will actually be found out by computing the expectation of the random variable.

So, the expectation of the random variable gives us an estimate about the number of times of success of an experiment will occur, if I repeat the experiment for a large number of times. Now, this particular law and the concept of binomial distribution have got a very important application in the field of cryptographic analysis.

(Refer Slide Time: 26:00)

## A useful result

Let  $\xi$  be an event in a probability space  $X$ , with  $\Pr[\xi]=p>0$ . Repeatedly, we perform the random experiment  $X$  independently. Let,  $G$  be the expected number of experiments of  $X$ , until  $\xi$  occurs the first time. Prove that:  $E(G)=\frac{1}{p}$



So, for that, let us consider the particular result. It says that let epsilon be an event in a probability space  $X$  with probability of epsilon being equal to  $p$  and  $p$  is greater than 0, and what we do is that repeatedly, we perform the random experiment  $X$  independently.

(Refer Slide Time: 26:25)

$\xi$   
 $\Pr[\xi] = p > 0$

$G$ : expected number of experiments of  $X$ , until  $\xi$  occurs the first time.  
 $E(G) = \frac{1}{p}$ .

$2^{-12}$   
 $128$

$\frac{1}{2^{-12}} = 2^{12}$   
 $2^{-128}, 2^{128}$

$\Pr[G=t] = (1-p)^{t-1} p$   
 $E(G) = \sum_{t=1}^{\infty} t (1-p)^{t-1} p$   
 $= -p \frac{d}{dp} \sum_{t=1}^{\infty} (1-p)^t$   
 $= -p \frac{d}{dp} \left( \frac{1-p}{p} \right) = \frac{1}{p}$

So, what we are saying here is, there is an event epsilon; so, epsilon is an event and there is the probability space  $X$ . So, what we know that the probability that epsilon occurs, that is, this particular event occurs, is actually given by  $p$ , where  $p$  is some non-zero value, it is greater than zero. So, what we then do is that we repeat this experiment again and again.

So, we repeat the experiment  $X$  once, we do the experiment  $X$  twice, and what we need is that we need to find out the expected number of experiments of  $x$  until  $\epsilon$  occurs the first time. So, what we need is that we need the expected number of the experiments of  $x$  until  $\epsilon$  occurs the first time. So, we are interested in a particular event; we know the probability of this event and we need to find out the number of times we would like to repeat this experiment, until we first get the success. The success is defined as the fact that this event,  $\epsilon$  occurs.

So, how do I compute that? What we say, if this  $G$  is a random variable, it says that this is the expected number of experiments of  $x$  until  $\epsilon$  occurs the first time, and what we will do is that we will try to give you a proof or develop a proof that this expectation  $E G$  is actually given by the reciprocal of  $p$ . Now, what is the impact of this result? Suppose, there is an attack, suppose, you develop an attack and you say that the probability that this attack works has a complexity of say,  $2$  to the power of minus  $n$ .

So, that means the fact, that you can actually find out that particular key is  $2$  to power of minus  $n$ , then these particular results gives us a kind of indication that, if I repeat the experiment, if I repeat the attack for  $1$  by  $2$  to the power of minus  $n$  number of times, that is,  $2$  to the power of  $n$  number of times, then I should get the attack working at least once.

That is, after  $2$  to the power of  $n$  operations, I should get the attack to work. So, therefore, if I say that for example, in US if I say, that they have got 128 bit security and if I tell you that the probability that you can actually make the attack work is  $2$  power of minus 128, then immediately you know that if I repeat the experiment for  $2$  to the power of 128 number of times, for which is very huge, you will get 1 success.

Therefore, this actually gives us a nice indication to find out, if there is a property which you can exploit for an attack, it gives you an estimate about the number of times you need to repeat the experiment to get a success once.

The proof of this is actually quite simple and follows from the binomial distribution notions and the idea is as follows: what you do is, you find out probability of  $G$  equal to the value at  $G$  is equal to  $t$ . You know that  $G$  equal to  $t$  means the success occurs at the  $t$ th number of times, that is, it occurs at  $t$ th instance, which means that the previous experiments till  $t$  minus 1 has been failures. Therefore, this probability can be computed

by  $1 - p$  raised to power of  $t - 1$  because they were all failures multiplied by  $p$ ; therefore, the expectation of  $g$ , by using the law of large numbers, will be equal to this probability multiplied by the number of times you are repeating the experiment.

So, what you are doing is that you are repeating this for  $t$  number of times and therefore, you multiply this by  $1 - p$  to the power of  $t - 1$  and multiply that by  $p$ . Now, you note that this particular  $t$  can be anything, it can go from  $t$  equal to 1, when the first time we are actually not going to prove this, even further, this is actually equal to minus  $p$  differential with respect to  $p$  of  $\sum_{t=1}^{\infty} (1 - p)^{t-1} p$ .

So, this will work out like this where  $t$  runs from 1 to infinity; you can check this that it actually gives you the same value; so, then that refers from simple differential calculus, if you differentiate this, you will get this.

So, this means that if I take this sigma, then this is nothing but  $\sum_{t=1}^{\infty} (1 - p)^{t-1} p$  and I would like to add this sigma. So, this  $1 - p$  to the power of 1 plus  $1 - p$  to the power of 2 plus  $1 - p$  to the power of 3, so until  $1 - p$  to the power of infinity.

So, you note that since this value is less than 1, this sigma actually converges and this summation actually converges and what we get is  $\sum_{t=1}^{\infty} (1 - p)^{t-1} p = 1 - (1 - p)^{\infty} = 1 - 0 = 1$ . So, therefore, from here we actually get an estimate about the number of times we have to repeat the experiment, so that we actually get this experiment to work.


Therefore, it is important to know this result, because this actually gives us an idea about if there is an experiments' success probability defined, then from there to get an estimate about the number of times we need to repeat the experiment to get that success or to get that event.



(Refer Slide Time: 32:55)


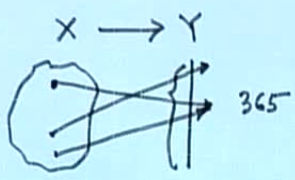
## The Birthday Paradox

- Consider a function,  $f: X \rightarrow Y$ , where  $Y$  is a set of  $n$  elements.
  - eg, consider this class of students form  $X$ . Let  $Y$  denote the birthday, say 15<sup>th</sup> September is the birthday of a person  $X$ .
  - thus,  $Y$  is the 365 days of a year (let us consider that no-body in the class was born on 29<sup>th</sup> February)



Then we come to the next important concept in today's class, which is called the birthday paradox. Birthday paradox is actually quite central to the idea of analysis of ciphers. So, consider a function  $f$ , which is a mapping from  $X$  to  $Y$ , where  $Y$  is a set of  $n$  elements. Consider this class of students form  $X$ , let  $Y$  denote the birthday, say, fifteenth September is a birthday of a person  $X$ .

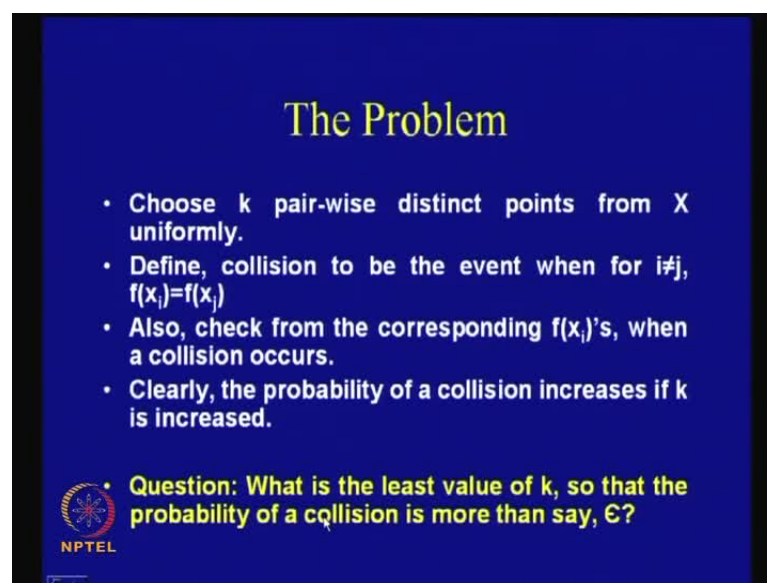
(Refer Slide Time: 32:24)



So, I mean, there are two things - one of them is  $X$  and the other one is  $Y$ . So, we are considering a mapping from  $X$  to  $Y$ . So, let us consider this class of students from  $X$ . Let

us consider that in this class there are  $X$  students, that is, there are  $X$  students and let us consider that their possible number of birthdays can be 365. So, there are 365 possible days of birthday. So, choose a person, say  $A$  and we know that  $A$  will be mapped to one of these days from this 365; if there is another person, then he will also or she will be mapped to another particular day among this 365. So, the question is, if you consider, for example, what we are essentially considering is that we are considering the fact, that there can be two persons or say, another person who are born on the same day; so, there is the collision of these two birthdays.

(Refer Slide Time: 34:21)



The slide has a dark blue background with yellow text. The title 'The Problem' is centered at the top. Below it are four bullet points. At the bottom left is the NPTEL logo, and at the bottom right is a question in yellow text.

### The Problem

- Choose  $k$  pair-wise distinct points from  $X$  uniformly.
- Define, collision to be the event when for  $i \neq j$ ,  $f(x_i) = f(x_j)$
- Also, check from the corresponding  $f(x_i)$ 's, when a collision occurs.
- Clearly, the probability of a collision increases if  $k$  is increased.

**Question: What is the least value of  $k$ , so that the probability of a collision is more than say,  $\epsilon$ ?**

NPTEL

So, we can say that the problem is like this, that is, we can abstract out this problem as this, that is, choose  $k$  pair-wise distinct points from  $X$  uniformly and define collision to be the event for  $i$  not equal to  $j$ , where  $f(x_i)$  is equal to  $f(x_j)$ . We check from the corresponding  $f(x_i)$ 's, when a collision occurs and clearly, the probability of a collision increases if  $k$  is increased. That is, if I choose large number of points, that is, if I choose large number of students from the class, then the probability of that they are born on the same day also increases.

The question now, is what is the least value of  $k$ ? We are actually interested in the least value of  $k$ , so that the probability of a collision is more than epsilon; so we would like to give a lower bound of the probability. So, first of all you understand that there are two

important things - one of them is this least value, and then we are actually giving the lower bound of the probability.

So, why are we choosing the least value of k? Because if I start increasing this k and if this k is quite large, then this probability will obviously increase and the probability of collision will obviously increase, but what we are interested is in finding out, what is the least value of this k so that this collision occurs?

(Refer Slide Time: 35:41)

**Let us compute for the class**

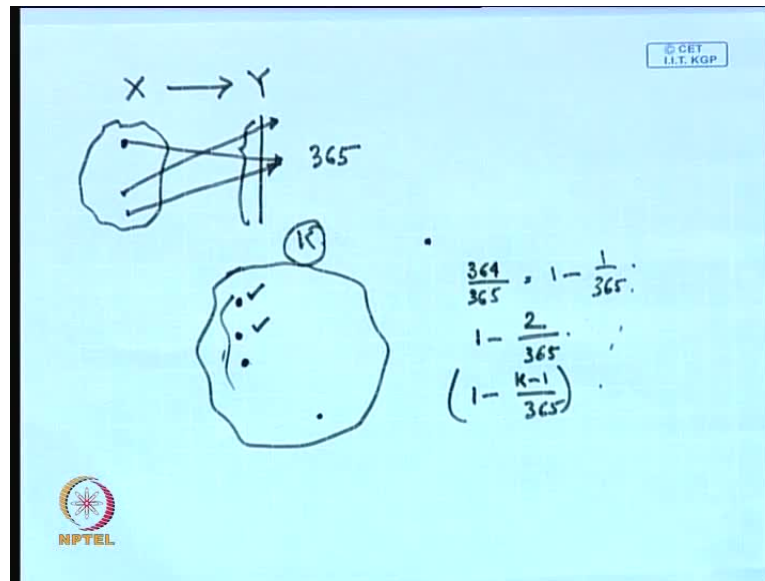
- **Probability of no collision in k persons in the class is:**  $(1 - \frac{1}{365})(1 - \frac{2}{365}) \dots (1 - \frac{k-1}{365}) = \prod_{i=1}^{k-1} (1 - \frac{i}{365})$
- **For a large n and a small x,**  $(1 + \frac{x}{n}) = e^{x/n}$
- **So, Pr of no collision is,**  $\prod_{i=1}^{k-1} (1 - \frac{i}{365}) \approx \prod_{i=1}^{k-1} e^{-i/365} = e^{-\frac{k(k-1)}{730}}$

NPTEL

Then we are actually trying to give a lower bound of this probability, that is, the probability should be at least this much, so we would like to give an answer to this question. So, in talking about in the birthdays' term, like we are actually interested in finding out, what is the number of students, which should be in that class, that is, what is the least number of students which should be in the class, so that the probability that two of them are born on the same day is say more than half.

So, I am interested in finding out, what is the least size of the class. So that there are two students, at least, who are born on the same day and a probability of this fact is more than or greater than or equal to half. So, we can actually compute this quite easily and we can do this as follows: like, the probability of no collisions, let us find that first in among this k persons.

(Refer Slide Time: 36:27)



So, we know that if there are  $k$  persons, that is, if there are  $k$  persons in the class and let us consider that this person is mapped to a particular date, that is, he is basically born on the same day. So, if we are considering the probability of no collisions, then it means that the second person should not be born on this date that means, from 365 days, he can be born on 364 days; so this is 364 divided by 365 or that is 1 minus 1 by 365. What about the third person? The third person cannot be born on this date or this date, therefore, his probability will be 1 minus 2 divided by 365.

Similarly, if I consider the last person, that is, like this person, then his probability will be 1 minus  $k$  minus 1 divided by 365 because there are  $k$  persons in this set. I am considering this least size of the set.

Therefore, this probability and all of them are independent events. So, the probability of no collision among these  $k$  persons can be found out by this, that is, I multiply 1 minus 1 by 365 with 1 minus 2 by 365 with 1 minus  $k$  minus 1 by 365 and so on; therefore, this is nothing but the product of these probabilities.

Now, for a large  $n$  and a small  $x$ , we have got this approximation, that is, one plus  $x$  by  $n$  is nothing but  $e$  to the power of  $x$  by  $n$ ; so for the large  $n$  and a small  $x$ , this holds. So, if I use this approximation, then this product of 1 minus  $i$  by 365, as each of these terms is substituted by  $e$  to the power of minus  $i$  divided by 365, then we try to find out the product of these terms. This actually works out to  $e$  to the power of minus  $k$  into  $k$  minus

1 divided by 730, because when we are doing a product of these terms, in the powers we are doing actually a sigma; so, if I do this sigma, then this sigma of minus i by 365 will work out as this.

(Refer Slide Time: 38:42)

$$\prod_{i=1}^{k-1} \left(1 - \frac{i}{365}\right) \approx \prod_{i=1}^{k-1} e^{-i/365}$$

$$= e^{-\sum_{i=1}^{k-1} \frac{i}{365}}$$

$$= e^{-\frac{1}{365} [1+2+\dots+(k-1)]}$$

$$= e^{-\frac{1}{365} \frac{k(k-1)}{2}}$$

$$= e^{-\frac{k(k-1)}{730}}$$

$$1 - e^{-\frac{k(k-1)}{730}} = 0.5$$

$$\left. \begin{aligned} e^{-\frac{k(k-1)}{730}} &= 0.5 \\ \frac{k(k-1)}{730} &= \ln(2) \\ k^2 - k &= 730 \ln(2) \\ k &= \sqrt{730 \ln(2)} \end{aligned} \right\}$$

See, what we are doing is this, that is, i equal to 1 to k minus 1 and we are multiplying 1 minus i by 365. So, by using the approximation is, i equal to 1 to k minus 1 into e to the power of minus i by 365.

Now, when we are doing this product, then this means that what we are doing is e to the power of minus i by 365 and then we are doing a summation here, so that means that it is e to the power of 1 by 365, if you take common, it is 1 plus 2 plus, so on till k minus 1. So, this is e to the power of 1 by 365 into k into k minus 1 by 2; there is a minus out. So, this minus will come out; so, minus 1 by 365 into k into k minus 1 by 2, so that is e to the power of minus k into k minus 1 divided by 730.

Suppose, we say this is the probability where there is no collision, so what is the probability that there is at least one collision? That is simple, that is 1 minus e to the power of minus k into k minus 1 divided by 730, and if I say this probability should be at least equal to 0.5, then we can actually get an estimate of this k by computing, or rather, equating this to 0.5. That means, you are trying to say, this probability should be greater than equal to 0.5, but in order to compute the value of k, let us equate this to be 0.5 and find out an estimate of k.

We can actually see this and I am not really going into this, that is, you can find by calculations like this, that is,  $1 - e^{-\frac{k(k-1)}{730}}$  is equal to 0.5 and therefore,  $\frac{k(k-1)}{730}$  will be equal to  $\ln 2$ ; so, you can actually do this type. Therefore, what you can do is that you can bring this  $e$  to the power of minus  $k(k-1)$  by 730 will be equal to 0.5 and then if you take a log on both sides, then it works out to  $k(k-1)$  by 730 is nothing but  $\ln 2$ . Therefore, you can get  $k^2 - k$  will be equal to  $730 \ln 2$ . Therefore, here if you neglect  $k$  with  $k^2$ , then  $k$  will be roughly equal to square root of  $730 \ln 2$  and that will be approximately 23; so, which means that if there is a random room of 23 people, then the probability that there are 2 persons with the same birthday is 0.5.

(Refer Slide Time: 41:47)

**Let us compute for the class**

- Probability of a collision is:  $1 - e^{-\frac{k(k-1)}{730}}$
- Let this be  $\epsilon=0.5$
- Thus,

$$1 - e^{-\frac{k(k-1)}{730}} = 0.5$$

$$\therefore \frac{k(k-1)}{730} = \ln(2)$$

$$\therefore k^2 - k = 730 \ln(2)$$

$$\therefore k \approx \sqrt{730 \ln(2)} \approx 23$$

**Thus, in a random room of 23 people, the probability that there are two persons with the same birthday is 0.5 !!! Seems to be a paradox**

NPTEL

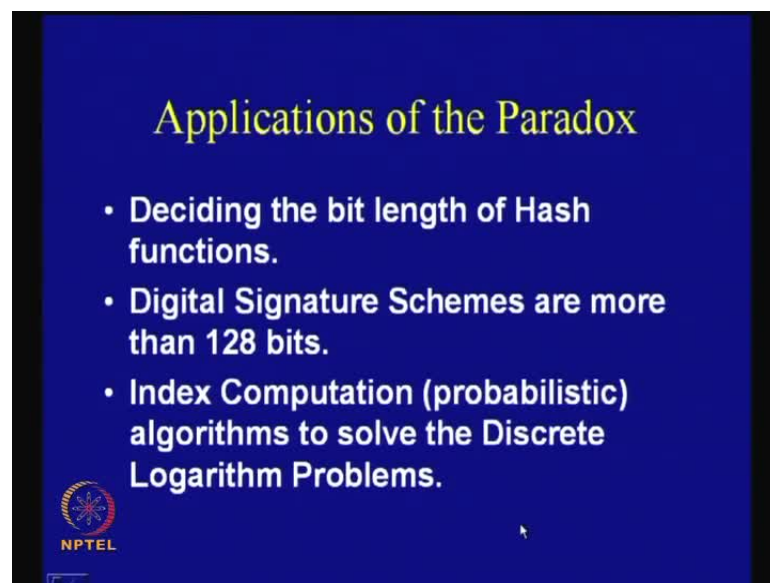
This actually seems to be a paradox, why? Because what if I asked you, what is the probability that there are two people who are born on the same date is actually very small; it is actually 1 by 365, so that is a very small number.

But we see that in a random room of 23 people, the probability that there are two persons which are born in the same birthday is actually quite high; it is actually 0.5 and we will see that if I increase this 23 people to larger and actually this probability shoots up quite fast.

Now, we have a very specific example, but you can work out the more general case, that is, instead of 365 we can have nearly 2 to the power of n possible outcomes and then try

to find out the estimate of  $k$ , but what we will find is that this  $k$  will roughly be proportional to the square root of total number of possible ways. So, which means that if you need to do a brute force search of  $2$  to the power of  $n$  possible values, then if you apply the birthday paradox, then this gives you an estimate that after  $2$  to the power of  $n$  by  $2$  random searches, there is a high probability that two of them will actually result in a collision.

(Refer Slide Time: 43:04)



So, this particular paradox or this particular analysis is used again and again to do the analysis of ciphers and develop with security proofs and other stuff. So, there are lot of applications deciding the bit length of the hash function, digital signature schemes are have to be kept more than 128 bits; it is used for doing cryptanalysis like index computation which are algorithms to solve something, which is called, the discrete logarithmic problems.



(Refer Slide Time: 43:22)

The slide features a dark blue background with yellow text. At the top, the title 'Cycle Finding Algorithms' is written in a yellow serif font. Below the title, there are three bullet points in white sans-serif font: '• Consider a function, F from S to itself', '• Starting from  $X_0$  in S generate a sequence by using  $X_{i+1}=F(X_i)$ ', and '• Goal is to find a collision,  $X_i=X_j$ '. At the bottom of the slide, there is a diagram showing a 'Tail' of a sequence leading into a circular 'Cycle'. The NPTEL logo is visible in the bottom left corner.

We actually try to find out a very interesting application of the birthday paradox; so it is something which is called cycle finding algorithms. Suppose, there is a linked list which is very large and I am interested in finding out a cycle in the linked list; so, this is one question that may appear. One way of doing that is, go through the entire linked list and this store a huge amount of data and once you see a repetition, you report a cycle, but can we do better than that?

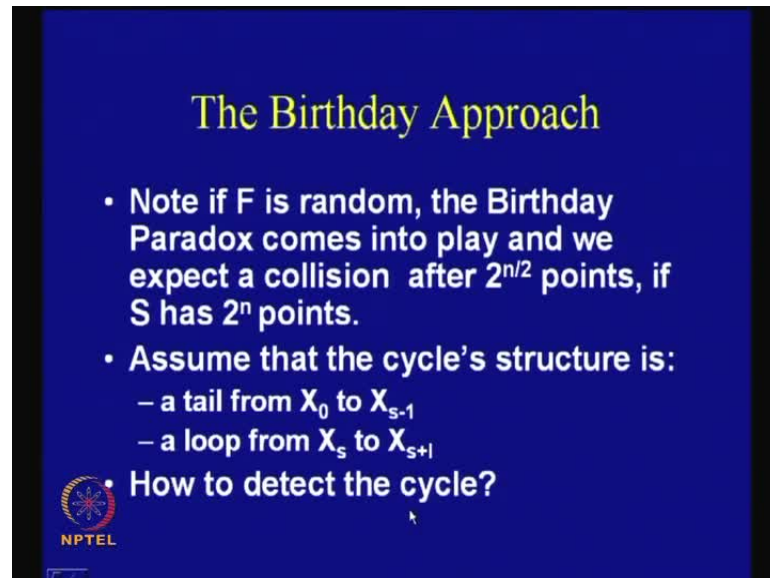
Because what the birthday paradox says is that if you choose the elements in these lists at random and if there are  $2^n$  possible elements, then after  $2^{n/2}$  possible searches or possible points, there is a high probability that two of them will actually lead to a collision.

Suppose, you actually keep on this arbitrarily choosing the points, then we will find that after some point it may happen that if you keep on computing, you will find there will be a resulting collision the moment you get a collision, you know that there is a cycle.

Consider a function  $F$  from  $S$  to itself, what you do is, you start from  $X$ , then what you that is you start from  $X_0$ , So,  $X_0$  could be here and then you generate a sequence by recursively as follows:  $X_{i+1}$  is equal to  $F(X_i)$  and you just keep on computing this value.



(Refer Slide Time: 44:58)



The slide has a dark blue background with yellow text for the title and white text for the bullet points. The NPTEL logo is in the bottom left corner.

### The Birthday Approach

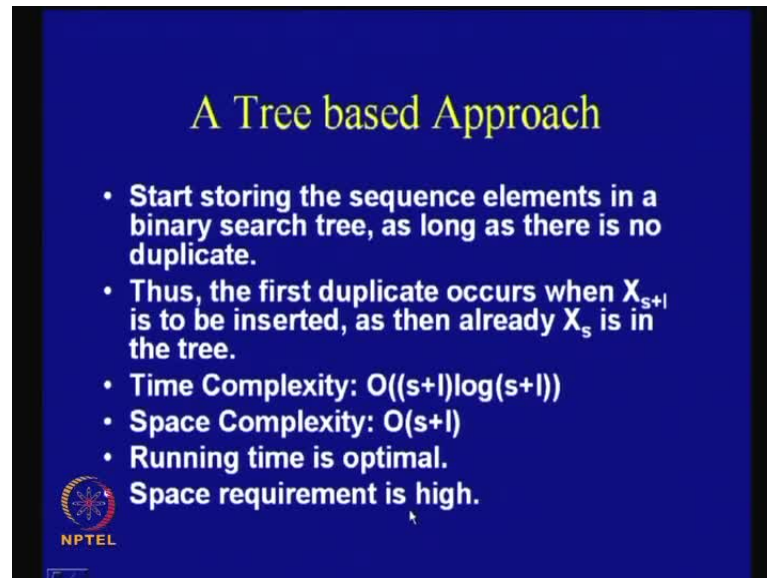
- Note if  $F$  is random, the Birthday Paradox comes into play and we expect a collision after  $2^{n/2}$  points, if  $S$  has  $2^n$  points.
- Assume that the cycle's structure is:
  - a tail from  $X_0$  to  $X_{s-1}$
  - a loop from  $X_s$  to  $X_{s+l}$
- How to detect the cycle?

NPTEL

The goal is to find a collision, such that  $X_i$  and  $X_j$  are same, that is,  $X_i$  and  $X_j$  are resulting in the same value. So, we can actually have a birthday approach, that is, note if  $f$  is random then the birthday paradox comes into play and we expect a collision after  $2$  to the power of  $n$  by  $2$  points, if  $S$  has got  $2$  to the power  $n$  points.


So, assume that the cycle structure is like this, that there is a tail from  $X_0$  to  $X_{s-1}$  and there is a loop from  $X_s$  to  $X_{s+l}$ , that is, from  $X_0$  to  $X_{s-1}$ , you come here - this is the tail, and then from the next one, that is, the  $X_s$  to  $X_{s+l}$ , there is a kind of cycle; so, that means this cycle has got a length of  $l$ .

(Refer Slide Time: 45:38)



**A Tree based Approach**

- Start storing the sequence elements in a binary search tree, as long as there is no duplicate.
- Thus, the first duplicate occurs when  $X_{s+1}$  is to be inserted, as then already  $X_s$  is in the tree.
- Time Complexity:  $O((s+1)\log(s+1))$
- Space Complexity:  $O(s+1)$
- Running time is optimal.
- Space requirement is high.

 NPTEL

Now, the question is how to detect this cycle? One way of doing this could be a tree based approach. So, what you do is that you start storing the sequence elements in a binary search tree as long as there is no duplicate; so, you keep on adding them to the tree and as long as there is no duplicate. Now, the first duplicate occurs when  $X_{s+1}$  is to be inserted because there is already  $X_s$  which is inserted into the tree and the moment you get to insert  $X_{s+1}$ , you know that there is a kind of duplication and therefore you report a collision.

Now, what is the time complexity here? We know that if we are using a binary search tree, then the complexity will be around  $O(s+1 \log(s+1))$ , but what about the space complexity? The space complexity is still  $O(s+1)$ , so that means that although the runtime is optimal because actually you cannot do better than this in terms of time, but the space requirement is quite high, that is, this could be an exponentially large; this could be quite a large value, so the question is, can I make the space requirement less.

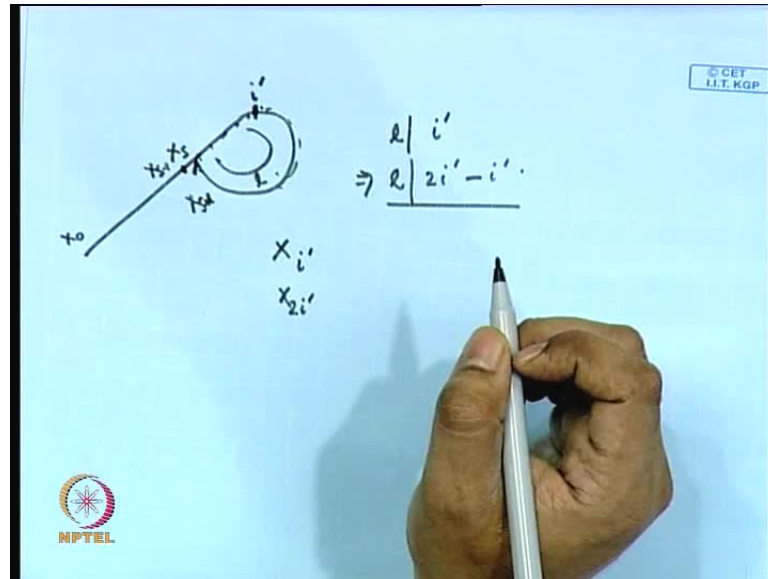
(Refer Slide Time: 46:39)

The slide features a dark blue background with yellow text. At the top, the title 'Floyd's Cycle Finding Algorithm' is displayed in a large, bold, yellow font. Below the title, two bullet points are listed: 'Define  $Y_0 = X_0$  and  $Y_{i+1} = F(F(Y_i))$ ' and 'Input initial sequence  $X_0$  and max iterations  $M$ '. A central yellow box contains the following pseudocode:   
`x = X0, y = X0  
for i from 1 to M do  
  x = F(x)  
  y = F(F(y))  
  if x == y  
    Output 'Collision between i and 2i'  
  exit  
  end if  
end for  
output Failed`  
In the bottom left corner, there is a circular logo with a star-like pattern and the text 'NPTEL' below it.

A very interesting technique is being adopted for finding out the cycles. A very simple algorithm I will discuss here, is called Floyd's cycle finding algorithm and it works out as follows; we will see an application of this in context to factorization, when we discuss about factorization. So, you see that what we do is that we set that  $Y_0$  is equal to  $X_0$  and we compute another sequence  $Y_{i+1}$  as  $F$  of  $F$  of  $Y_i$ , so instead of applying once  $F$ , I am applying twice  $F$ .

So, the input initial sequence is  $X_0$  and we said the maximum interactions is  $M$ ; so, what we do is that we start  $x$  is equal to  $X_0$  and  $y$  is equal to  $X_0$ . So, we start at the same point and then for all these possible iterations that we have set as the maximum value, we compute  $x$  is equal to  $F(x)$  and we compute twice for  $y$ ; so, we apply  $F$  twice and we obtain the points. If we get a collision at some point, like if  $x$  and  $y$  is same, then we say that there is the collision between  $i$  and  $2i$ , because if you say that this  $x$  sequence that at point  $i$  and  $y$  sequence at point  $2i$  and **if you get**, so then you say that there is a collision, otherwise you say that there is a failure. If you see that actually you have got quite good chance of getting a collision because of the simple fact that this particular length, that is, the length of the cycle will divide  $2i - i$ ; it will actually divide this  $2i - i$ .

(Refer Slide Time: 48:19)



So, this is actually a very useful algorithm. It is often useful for doing analysis and one of the reasons is that for example, if you consider the cycle it is like this, it looks like a rho. If you start from  $X_0$  and you come to this, suppose this point is  $X_{s-1}$  and this is  $X_s$  and we keep on obtaining and then this is again  $X_{s+1}$ ; therefore the length of the cycle is  $l$ . There are some  $l$  successive points here and therefore, you know immediately that there must be from the  $l$  successive points, there must be one point here, call it  $i$  dash, which is divisible by  $l$ ; so, there are  $l$  successive terms, so there must be one term  $i$  dash which  $l$  divides. Therefore, if  $l$  divides  $i$  dash, then that means that  $l$  divides  $2i$  dash minus  $i$  dash, so that is  $i$  dash itself; that means, if you compute the series of  $X$  like this, that is, compute  $X_{i'}$  and if you compute  $X_{2i'}$ , then that means you will find out if you are basically computing these values, then you may not be able to find out the first time this particular collision occurs, but you will find out a successive point when the collision occurs.

Therefore that will give you this collision which will lead to the detection of the cycle, so we will see this in one of our future classes when we discuss about factorization.

(Refer Slide Time: 49:47)

**Measuring Information**

- $L = \{a_1, a_2, \dots, a_n\}$  : Language of  $n$  different symbols.
- Independent probabilities:  
 $\Pr[a_1], \Pr[a_2], \dots, \Pr[a_n]$
- Probabilities satisfy:  $\sum_{i=1}^n \Pr[a_i] = 1$

NPTEL

So, we conclude our talk with some notions of information measurement. Let us consider a language of  $n$  different symbols -  $a_1$  to  $a_n$ , and let us assign independent probabilities like - probability of  $a_1$ , probability of  $a_2$ , and so on, till probability of  $a_n$ .

(Refer Slide Time: 50:08)

**Entropy**

- Entropy of the source,  $S$ :  
$$H(S) = \sum_{i=1}^n \Pr[a_i] \log_2 \left( \frac{1}{\Pr[a_i]} \right)$$
- Number of bits required per source output


NPTEL

These probabilities must satisfy that sigma of these probabilities will be equal to 1; so, what is the entropy of the source  $S$ ? The entropy of source  $S$  is defined as  $H(S)$  is equal to sigma probability of  $a_i$  multiplied by logarithm of 1 by probability of  $a_i$  base 2.

(Refer Slide Time: 50:37)

**Properties of Entropy**

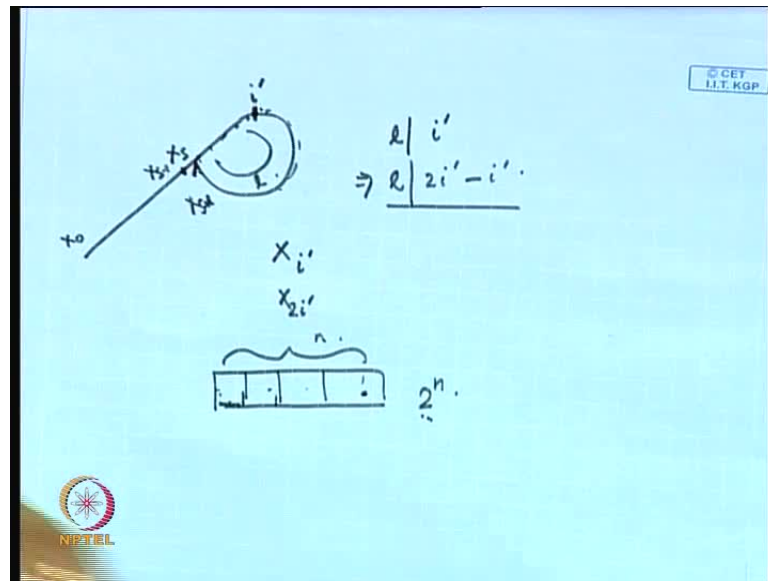
- If **S** outputs  $a_1$  with probability 1:  
 $H(S)=0$
- If **S** outputs  $n$  symbols with equal probability  $1/n$ , that is **S** is a source of a uniform distribution:  
$$H(S) = \frac{1}{n} \sum_{i=1}^n \log_2 n = \log_2 n$$
- **H(S)** can be thought as the amount of uncertainty or information in each output from **S**.

 NPTEL

So, this actually gives us the number of bits which are required per source output; therefore, this notion of entropy is often useful for computing the information content. Let us consider some examples, if **S** outputs a 1 with probability of 1, then  $H(S)$  will be equal to 0 because your probability is equal to one; therefore, if you plug in probability 1 here, then this logarithm computes to 0 and therefore, this  $H(S)$  is 0. But if **S** outputs  $n$  symbols with equal probability, that is, the probability is  $1/n$ , that is, **S** is the source of uniform distribution, then this  $H(S)$  will compute to  $\frac{1}{n} \sum_{i=1}^n \log_2 n$  and that is nothing but  $\log_2 n$ .

Therefore,  $H(S)$  can be thought of as the amount of uncertainty or information in each output from **S**; so consider, if you have got a binary sequence of values, like if there is a  $n$  bit length then, there are  $2^n$  possible values here and all these  $2^n$  possible values can be chosen. Therefore, the probability of a particular sequence is nothing but  $1/2^n$ .

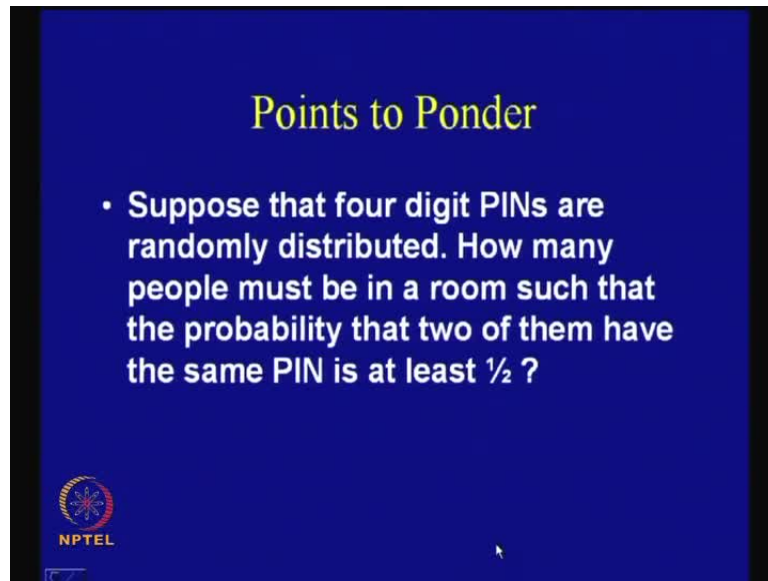
(Refer Slide Time: 51:40)



So, in that case the number of number of bits which are there, are in order to find out the value of  $X$  and are actually  $n$ ; therefore, you need to ascertain these  $n$  bits in order to find out the value.


So, that is essentially the amount of uncertainty or the amount of information you need to find out the value. If we think of the previous example, that is, if  $S$  outputs a particular  $A$  with the probability of 1 that means, you know that there is no information, that is, no uncertainty; therefore, the uncertainty is 0. In this case, the uncertainty was 0, but here the uncertainty is logarithm  $n$  base 2, that means, if I say, if  $S$  outputs  $n$  symbols and if  $n$  is equal to 2 to the power of capital  $N$ , which I denoted by  $n$  bits, then this logarithm will work out to be  $n$ . So, that means that there is an uncertainty of  $n$  bits, which is there in  $H$   $S$ . So, there is an uncertainty of capital  $N$  bits in the source  $S$ ; therefore this notion of entropy gives us an idea about uncertainty or information.  $S$

(Refer Slide Time: 52:52)



**Points to Ponder**

- Suppose that four digit PINs are randomly distributed. How many people must be in a room such that the probability that two of them have the same PIN is at least  $\frac{1}{2}$  ?

 NPTEL

Now, I will leave you with a question that is - suppose that there is a four digit PINs and which are randomly distributed. How many people must be in a room such that the probability that two of them have got the same pin is at least half?

(Refer Slide Time: 53:13)



**References**

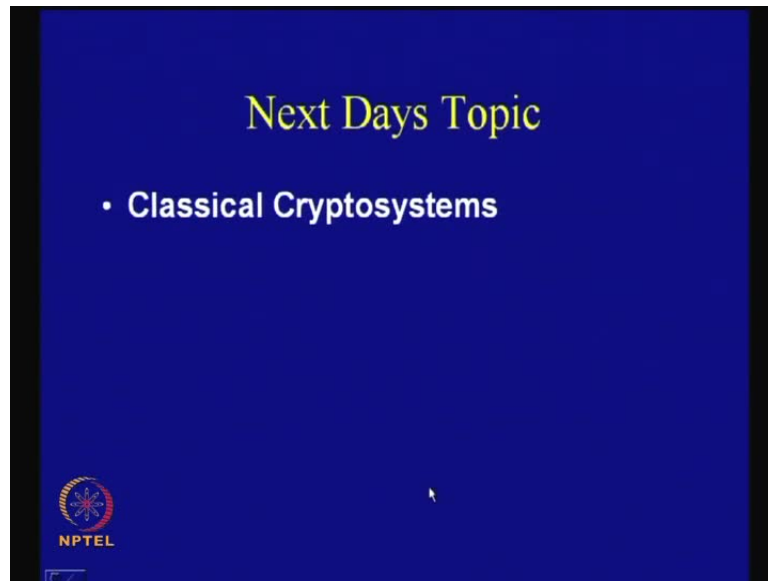
- W. Mao, "Modern Cryptography: Theory and Practice", Prentice Hall
- A. Joux, "Algorithmic Cryptanalysis", CRC
- Johannes A. Buchmann, "Introduction to Cryptography", Springer

 NPTEL



So, you can immediately understand that you should be applying birthday paradox to solve this problem. The references that are used are: Wenbo Mao's Modern Cryptography Theory and Practice and some of the things are taken from Algorithmic Cryptanalysis by Antoine Joux and by Buchmann, Introduction to Cryptography - it is a Springer book.

(Refer Slide Time: 53:29)



So, in next day's topic we will take up the topic of classical cryptosystems and discuss about some classical methods of doing cryptography.