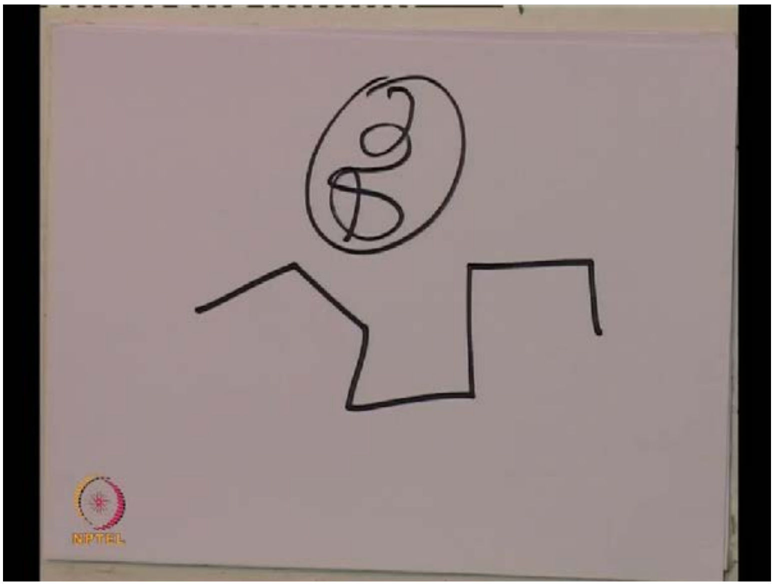
Dr. Ranjit Padinhateeri
Department of Biotechnology
Indian Institute Of Technology, Bombay
Lecture No. # 25

Size of a folded protein: Simplest model

Hello, welcome to this lecture on biomathematics. We have been discussing various ideas from statistics. We continue to discuss some more statistics and today we will discuss two separate issues; two small separate issues like one, something related to, so we have discussed about average standard deviation, etcetera. Now, we have today, first we will discuss the case of a flexible protein or a flexible polymer and if we have a flexible polymer of n monomers, what will be the average size of this protein or average folded size?

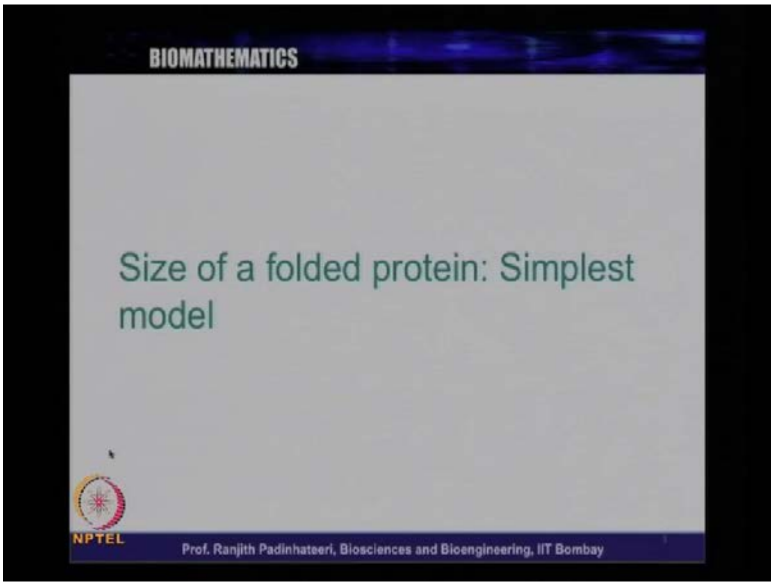
(Refer Slide Time: 01:13)



Like when, if we have a protein of n monomers N amino acids in the folded configuration it can be like anything. So, the size of this folded size, what will be the average size and what will be the standard deviation, etcetera, can be easily calculated from a simple model and this uses some ideas of statistics.

So, first we will discuss the size of a folded protein and then we will come to a slightly different distribution as per discussion, some more distribution. The main theme is statistics itself; so that is the bigger theme of today's lecture it is statistics as you see here.

(Refer Slide Time: 01:59)



BIOMATHEMATICS

Size of a folded protein: Simplest model

NPTEL

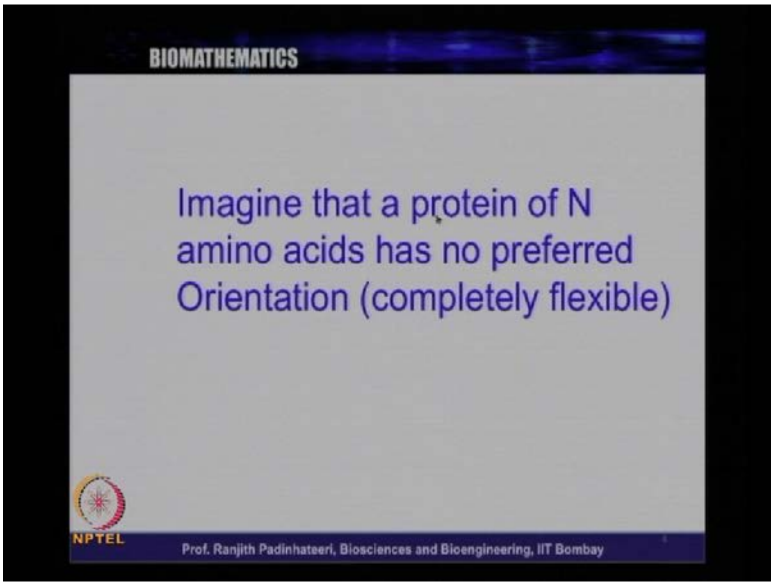
Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

Specifically, we will discuss first the part 1 of this lecture will be size of a folded protein as simplest model. What is a simplest model for a folded protein? Well, the real protein folding is a very complex problem and at this point we would not go into this but, the simplest model is just forget about all complexity and ask and tell that yeah, I have a protein of N amino acids or 5 amino acids or 10 amino acids or 40 amino acids and it takes and so you know that reality proteins like when they orient there is lot of constrained of angles and they orient in a particular angle and so on and so forth. But, the simplest model is to assume that protein amino acids or protein monomers can take any angle; they are flexible to take any configuration.

They are flexible to take this configuration, this configuration, this configuration, this configuration. So, any configuration it can take that is if we assume this, if we assume that a protein of N amino acids can take any orientation that is that makes the case that makes the problem very simple enough to analytically solve and get a formula for average size and standard deviation of the size and all that. So, we will see how we can do this; but, to make to do a calculation here, to learn a calculation by doing a simple example we will make an assumption which is kind of unrealistic. However, we will first we will have to do anything that is complex.

First, we have to learn how to do something very simple. If we do not even know how to do a simple calculation, we have no hope of doing something very complex. So, in reality, the size of a protein - the average size of a protein is the folded size of a protein. The size of the folded size of a (( )) is a little difficult to calculate; it is not very easy to calculate but, we will make an assumption and simplify the problem to a level where we can easily solve it and get an answer and many things that we can learn many things from this answer; so this is first.

(Refer Slide Time: 05:01)



BIOMATHEMATICS

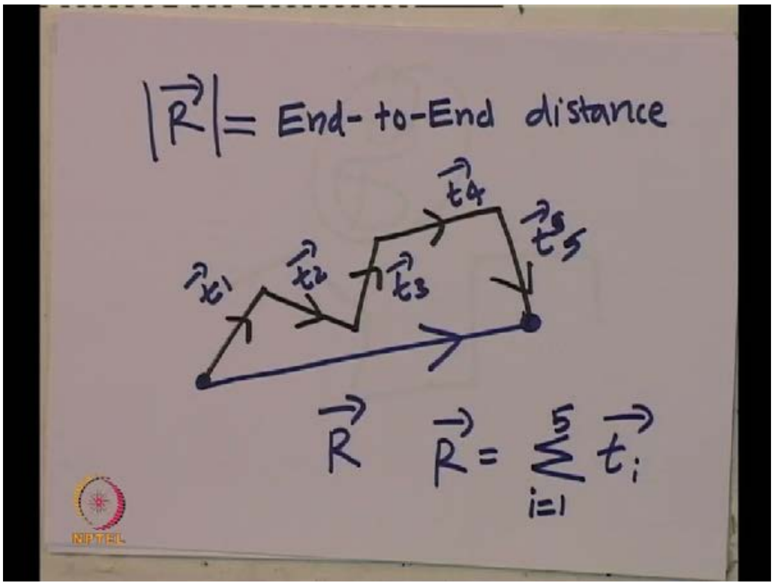
Imagine that a protein of N amino acids has no preferred Orientation (completely flexible)

NPTEL

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

Let us consider this particular case where the assumption is that, imagine that a protein of N amino acids has no preferred orientation; were completely flexible. This is true for proteins; true for any polymer actually. So, let we have a polymer and this polymer has n monomers; let us draw this here; so let us draw polymer of n monomers.

(Refer Slide Time: 05:23)

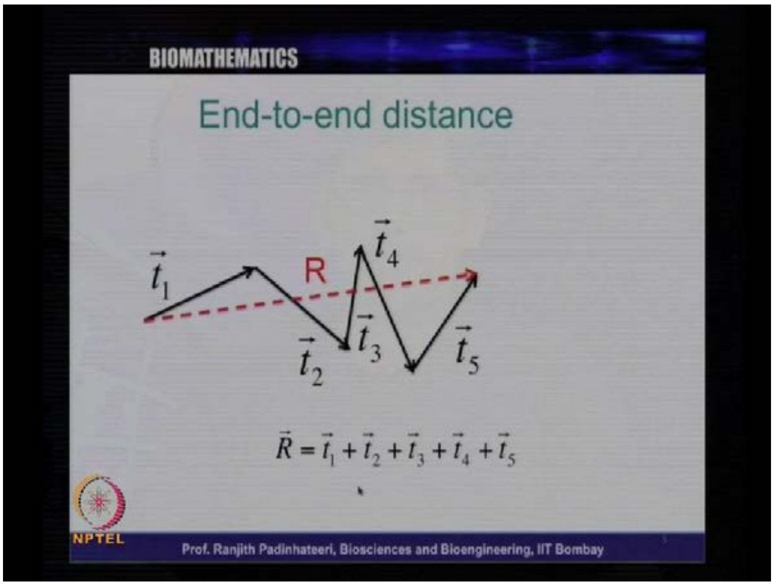


So this is the first monomer, second monomer, third monomer, fourth monomer, fifth monomer. Let us say imagine 5 monomer so they all oriented in some particular direction; first monomer, so there are 5 monomers. Now, how do we define the size of this protein?

There is some quantity called end-to-end distance, end-to-end distance or vector. We can this is one end and this is the other end; so we can define a vector which connects these 2 so from one end to the other end. What is this vector? We can call r which is the end-to-end vector or the mod of r is the end-to-end distance. So, the r is a vector and the mod of r we learn the, what do we mean by the mod of a vector? Basically, the distance; the distance is called end-to-end distance and the distance plus the direction which way it has to go; what is the direction from one end to the other end we have to go; in which direction to reach the other end these 2 information together form this end-to-end vector. Let each of this monomer have some orientation. Let, t 1, t 2, t 3, t 4 and t 5, these are the 5 vectors having 5 different orientation or any 5 random orientation then the end-to-end this vector r is nothing but, the sum of all these vectors.

So t 1 vector t 1 plus vector t 2 plus vector t 3 plus vector t 4 plus vector t 5 will give you this r vector. So, this is can be written as sum over I 1 to 5 t I. If you write so this is essentially what is shown here

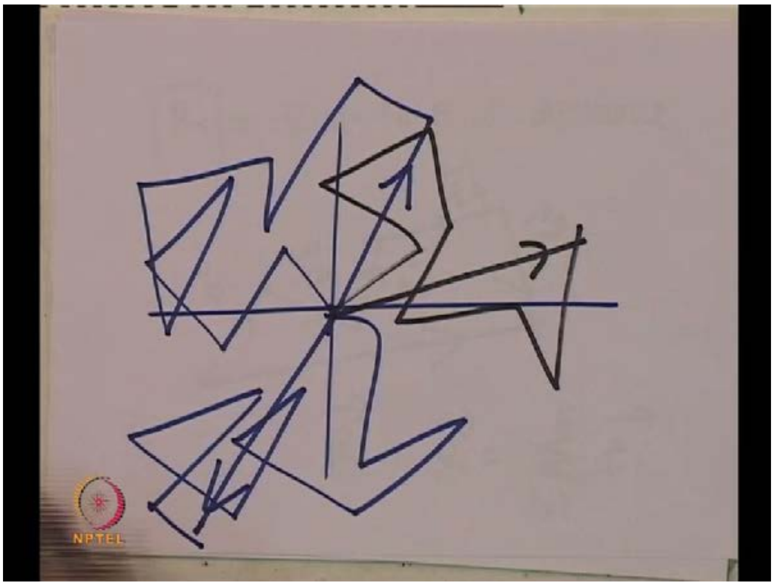
(Refer Slide Time: 07:52)



The end-to-end distance we have a vector t 1 t 2 t 3 t 4 and t 5 the end-to-end vector which is shown in red and represented by r can be obtained by summing t 1 vector plus t 2 vector plus t 3 vector plus t 4 vector plus t 5 vector. So, the sum of this 5 vector will give you this end-to-end vector r. So, this is 1 quantity r is a quantity that defines the size of a protein. In this particular configuration, the rough size can be estimated by this vector r so this is also a way you might have heard of something like there are various many other qualities that can define this folded size but, let us now define a quantity called end-to-end distance and the definition of this. Now, imagine that you are taking 1 million proteins the same protein, 1 million copies and you have put this 1 million copies in 1 million test tubes and at a particular moment you take a photo of all these proteins. Let us say you can exactly see or take a photo; imagine that you can take a photo of each of this configuration of this protein at a particular time.

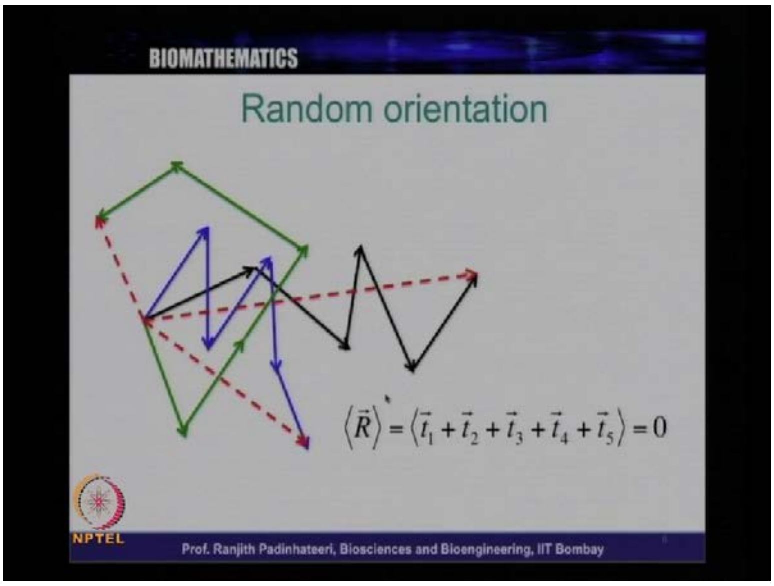
So, when you look at the watch and when the time is exactly 1 o clock in the afternoon you just take 1 million configuration photos. Imagine doing this so what do you expect to see? If you assume that they can take completely random configuration if you assume that each of this protein can take completely random configuration, what you expect to see is the following. What you expect? Let us draw what we expect to see when we take the first photograph.

(Refer Slide Time: 10:11)



Let us draw this x y axis; now, if you take keep statistics, so first photograph if you look it will have some shape like this with end-to-end distance like this. Next 1 will have some other configuration like this with end-to-end distance let us say, this with end-to-end distance this. Some other protein will have some other configuration completely random. The end-to-end distance can be something like this; so if you look at the end-to-end distance they might be oriented in completely random directions; some here one end-to-end distance here, one end-to-end distance here.

(Refer Slide Time: 11:07)



BIOMATHEMATICS

Random orientation

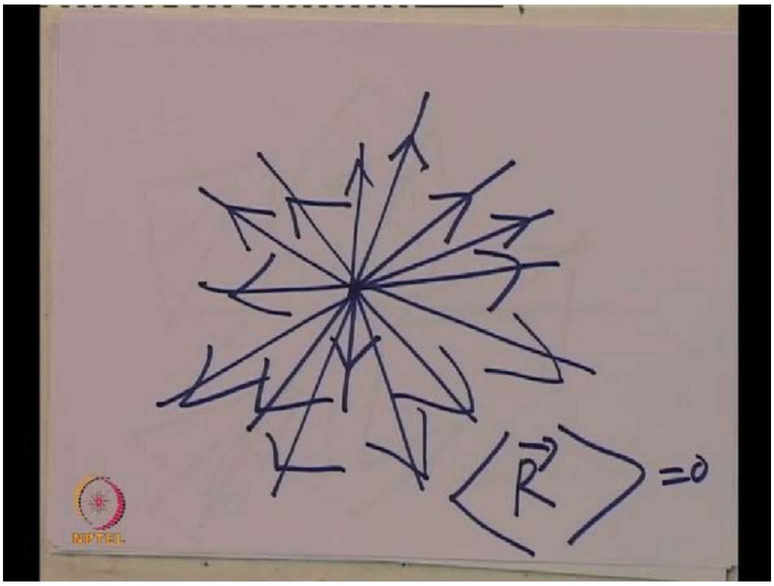
$$\langle \vec{R} \rangle = \langle \vec{t}_1 + \vec{t}_2 + \vec{t}_3 + \vec{t}_4 + \vec{t}_5 \rangle = 0$$

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

Have a look at this plot here; here, we have 3 different randomly oriented protein shown in 3 different colours. So, you can see black blue and green; black has this particular orientation and the end-to-end distance is this. Red here, the green has this orientation 2, 3, 4, 5; so, there are 5 vectors. They have an end-to-end distance this and the next 1 look at the blue one. They have some particular orientation and its end-to-end vector is this so if you just look at the red only if you look at the red only the end-to-end distance only end-to-end vector only and if you look at the end-to-end vector of 1 million configurations what do you expect to get one end-to-end vector like this another end-to-end vector like this another end-to-end vector like this another end-to-end vector like this some other end-to-end vector like this some other end-to-end vector like this some.

Other end-to-end vector like this so you will have end-to-end vector in completely random directions and random magnitudes so all sorts of directions all sorts of magnitudes so this are all of this are end-to-end vectors so you saw similar way you have 1 million or large number of end-to-end vectors now if you find average of all this vectors so average sum over all this and divided by the total number.

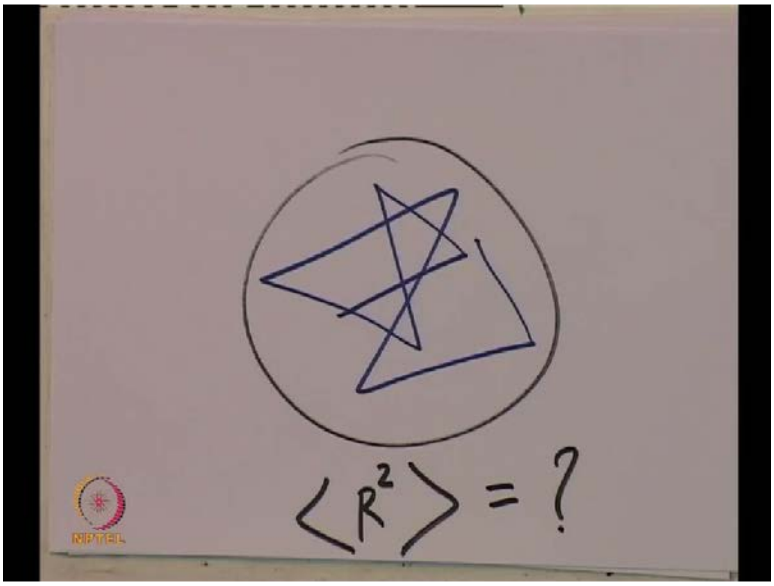
(Refer Slide Time: 12:03)



What do you expect to get? The average of all this will be 0 because this is 1 vector; exactly opposite direction there is another vector. So, they will be cancelling each other. So, if there is a vector in plus direction there has to be another vector in the minus direction because, completely random means it precisely means that they all will be uniformly distributed they all will be orienting in all sorts of directions. So, when you find the average of this there will be 0. If you do not trust this, you can go to a computer and try generating random configurations like this and calculate the end-to-end vector and calculate the average and see that they are indeed 0. If you go till large enough n there will be always 0. So, the average end-to-end distance 0; that is what is defined here; this r is defined as so the end-to-end distance r.

As you see here in the slide t 1 plus t 2 plus t 3 plus t 4 plus t 5, this average is equal to 0. So, this average means what does this angular bracket average mean? It means that you do many experiments in each experiment; you measure the end-to-end distance of a protein and you solve, average over this experiment and you get 0. So, well we know we can get that the average value 0 then next it does not have much meaning but, still even if its average is 0, if you look at this particular case, each of this protein has some size here.

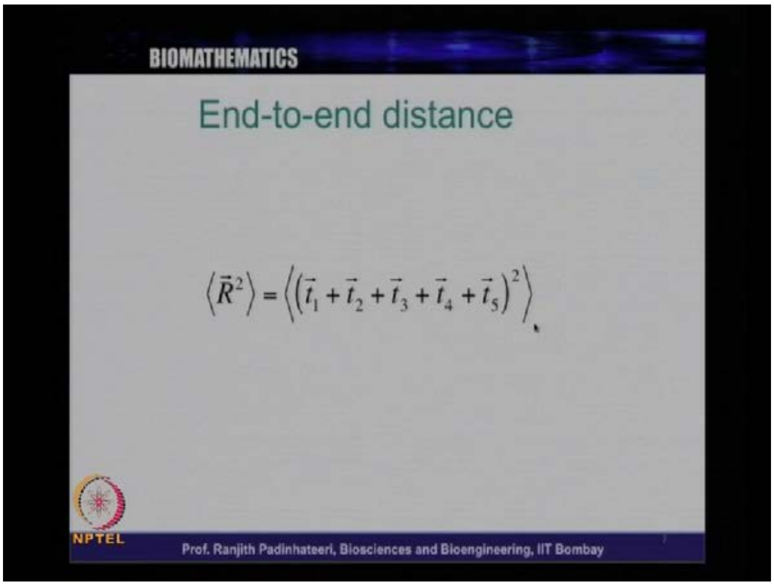
(Refer Slide Time: 15:03)



$\langle R^2 \rangle = ?$

If you look at this particular protein, there is there is some size, so we want to get some idea about what is this rough size like. You want to get some idea about this so the next quantity that is sensible is to calculate r square average. We saw that whenever r average is 0, it is sensible we always can calculate r square average and calculate the root mean square end-to-end distance. So, what is the r square average? That is the next; what is the r square average is the question; that is the next interesting question. So, how is r square average defined?

(Refer Slide Time: 15:50)


BIOMATHEMATICS

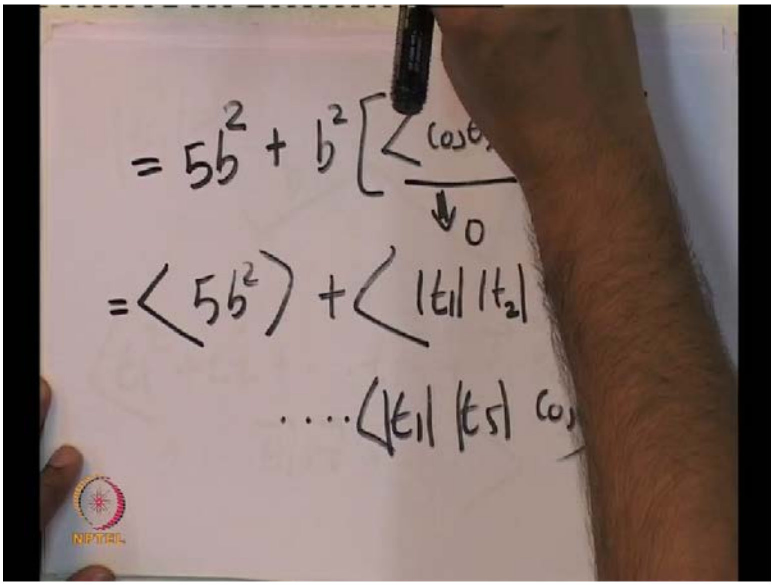
End-to-end distance

$$\langle \vec{R}^2 \rangle = \left\langle \left(\vec{t}_1 + \vec{t}_2 + \vec{t}_3 + \vec{t}_4 + \vec{t}_5 \right)^2 \right\rangle$$

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

So if you look at here the R square average is defined as t 1 plus t 2 plus t 3 plus t 4 plus t 5 whole square average. So what is t 1 plus t 2 plus t 3 plus t 4 whole r square?

(Refer Slide Time: 16:08)


$= 5b^2 + b^2 \Big[\langle \cos\theta$

$\downarrow 0$

$= \langle 5b^2 \rangle + \langle |t_1| \, |t_2|$

$\ldots\ldots \langle |t_1| \, |t_5| \cos$

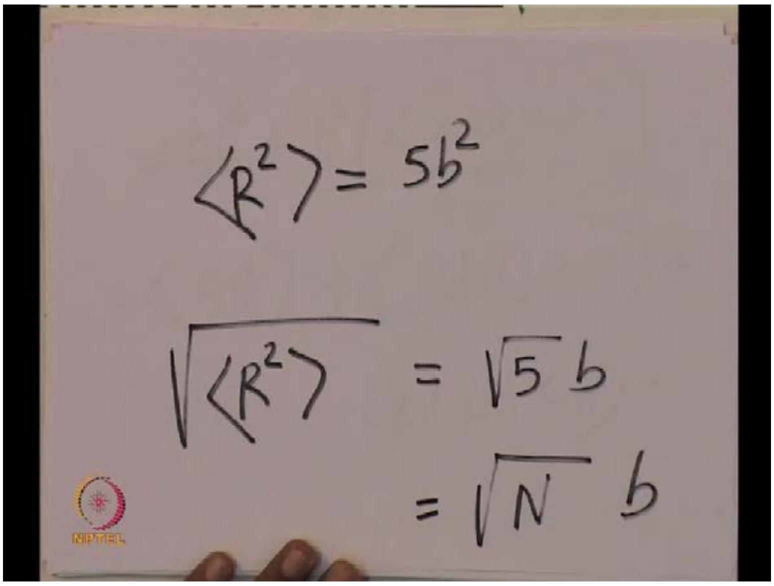
Average is defined as t 1 vector; there is vector here t 2 vector plus t 3 vector plus t 4 vector plus t 5 vector whole square and then average so, what is this thing, whole square? So, any vector square is r vector square is we know, that we saw that this is r dot r. So, basically you have this dot; you have basically t 1 vector plus t 2 vector plus t 3 vector plus t 4.Vector plus t 5 vector dot t 1 vector plus t 2 vector plus t 3 vector plus t 4 vector plus t 5 vector. What does this tell you? First is t 1 dot t 1 what is the answer of this so answer of this is first 1 is t 1 dot t 1 so that is t 1 square and then there is t 2 dot t 2 there is t t 2 square and there is t 3 dot t 3 t 4 dot t 4 t 5 dot t 5 t 5 square so t 1 square plus t 2 square plus t 3 square up to t 5 square then what do we have then we have t 1 dot t 2 so there is a terms like t 1 dot t 2.

There is t 1 dot t 3 dot dot there is t 1 dot t 5; so all these are plus dot dot dot all of this average. This is what we have to calculate; if you do this carefully what you get is there is t 1 square average t 2 square average t 5 up to t 5 square average plus all this product. We know what is t 1 square average so t 1 square average is we saw that t 1 square of any vector average will give you the average t 1 is the size of the first. So, we will have like 5 vectors; so t 1 is nothing but, the length of this.

So, T 1 square is nothing but, the length of this. If we define t 1 square as b square where b is the length of this monomer, if we define t 1 square as b square or in other words, if we define mod t 1 as b that is the length of each monomer is b t 1 square is b length of each monomer is same b if we if we if we imagine the length of each monomer each monomer as the same size and all of them are b so this is b square plus b square plus b square plus b square so you will get 5 b square from this.

So there are 1, 2, 3, 4, 5, there is what you get is that, so the answer is average; there is 5 b square plus there is t 1 dot t 2 so you know that t 1 dot t 2 is the mod of this and a dot b is cos theta. This is basically mod t 1 mod t 2 so mod t 1 mod t 2 cos theta average plus dot similarly, mod t 1 mod t 5 cos theta average plus dot dot dot this is the other terms there will be 5 b square now what do what do we get so t 1 dot t 1 t 2 is nothing but, b square so you can take b square out from everywhere and 5 b square average is 5 b square itself so 5 b square plus b square in to there is cos theta average plus dot dot many cos theta average now what is cos theta average cos theta average is nothing but 0 because, if you look at cos theta it can take, if you average cos theta, can take values from minus 1 to plus 1.

(Refer Slide Time: 21:29)



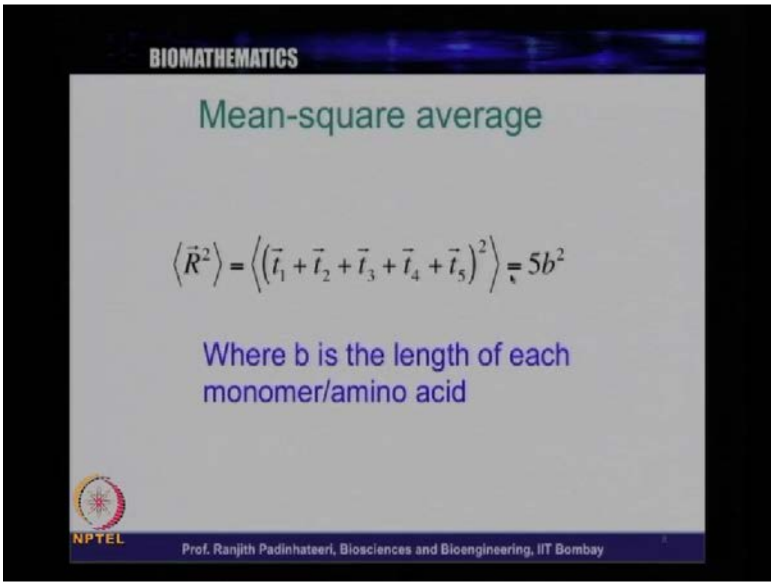
$$\langle R^2 \rangle = 5b^2$$

$$\sqrt{\langle R^2 \rangle} = \sqrt{5}\, b$$

$$= \sqrt{N}\, b$$

You sum over all this you will get 0 values from minus 1 to plus 1, 0 so the answer is, at the end of it the answer you will get r square average is 5 b square; so what do we get r square average is?

(Refer Slide Time: 21:42)



BIOMATHEMATICS

Mean-square average

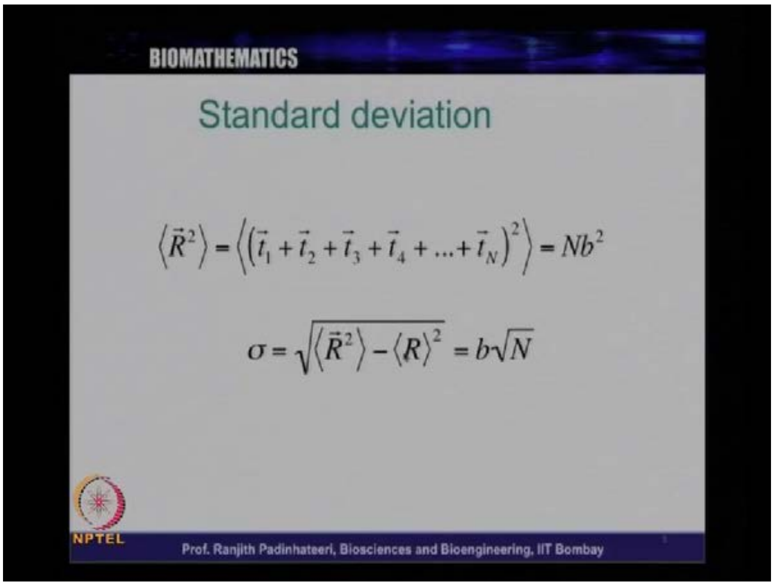
$$\langle \vec{R}^2 \rangle = \left\langle \left(\vec{t}_1 + \vec{t}_2 + \vec{t}_3 + \vec{t}_4 + \vec{t}_5 \right)^2 \right\rangle = 5b^2$$

Where b is the length of each monomer/amino acid

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

T 1 plus t 2 plus t 3 plus t 4 plus t 5 whole square average is 5 b square so the r m s end-to-end distance root mean square end-to-end distance is square root of 5 b. If we have 5 monomers the r square average is root of 5 b. Now, if you have n monomers you can extend this similar way to n monomers and you will see that the answer, if you have n monomers, this is root of n times b; so look at here.

(Refer Slide Time: 22:24)



BIOMATHEMATICS

Standard deviation

$$\langle \vec{R}^2 \rangle = \left\langle \left(\vec{t}_1 + \vec{t}_2 + \vec{t}_3 + \vec{t}_4 + \ldots + \vec{t}_N \right)^2 \right\rangle = Nb^2$$

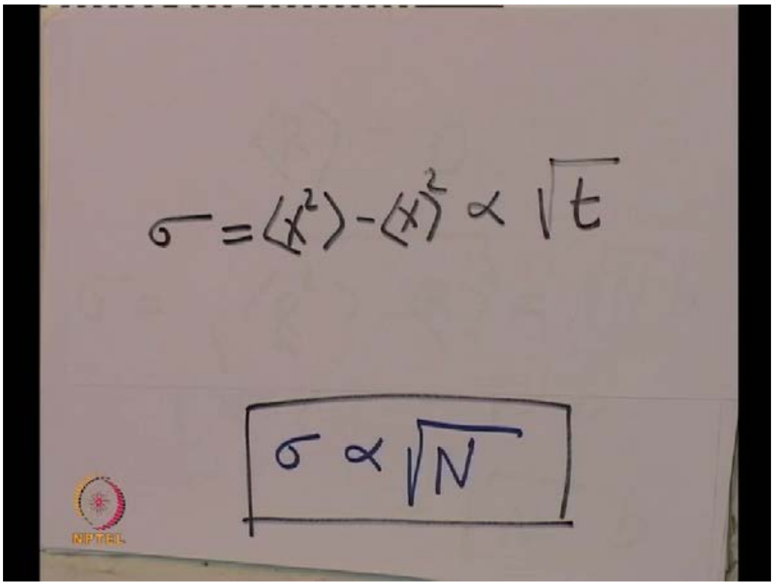
$$\sigma = \sqrt{\langle \vec{R}^2 \rangle - \langle R \rangle^2} = b\sqrt{N}$$

NPTEL

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

So, the r square average is n b square and you look at the standard deviation which is root of r square average minus r average square. You will (( )) find b times square root of n.

(Refer Slide Time: 22:45)



$$\sigma = \langle x^2 \rangle - \langle x \rangle^2 \propto \sqrt{t}$$

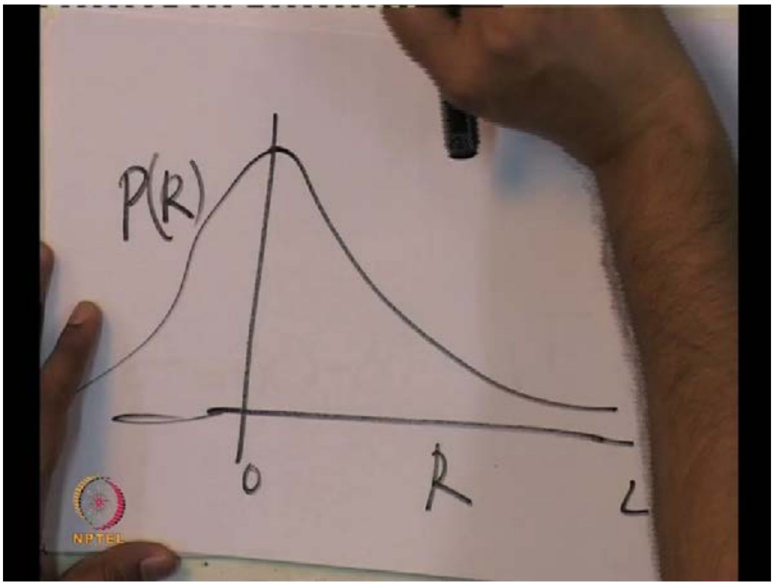
$$\boxed{\sigma \propto \sqrt{N}}$$

NPTEL

What did we find? We found that r average of a flexible protein or random protein that can take a random configuration will be 0 and the sigma r square average minus r average square; square root which is we define as sigma this will be square root of n times b. So, sigma is proportional to the square root of n; if there are n monomers the standard deviation of this end-to-end distance is proportional to the square root of n. so, this is an interesting, very interesting finding. This has some consequence; we had learnt in the case of diffusion previously that if you calculate sigma for diffusion sigma for diffusion which may be defined as x square average minus x average square, we will get this is proportional to square root of time. Here, we have similar way sigma is proportional to square root n. Here, it is time and here it is number of monomers; there is some relation between this and this. We will discuss this but, this is an interesting property that you should remember and this has some consequences. So, whenever you add, what did we essentially do here? We essentially did, what we did was we added some randomly oriented vectors. When we added randomly oriented vectors we got average 0 and the standard deviation going square root of n. Now, let us, we can extend just remember this argument. We will extend this to a different case in a different context in the coming class but, at this moment let us learn this for a protein and learn that the standard deviation sigma is proportional to the square root of the total number of monomers.

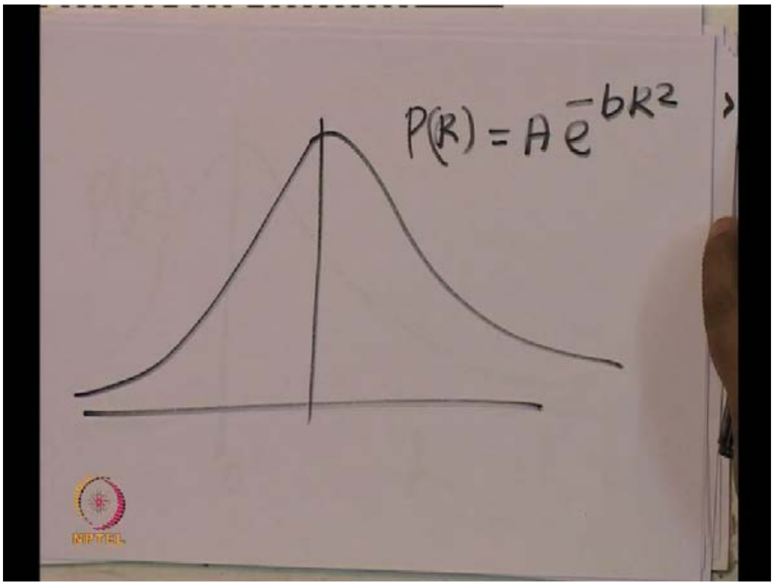
Now, the question we can ask is if we know the average and if we know the standard deviation, we know a lot of information about this but, what is the distribution of this? So, we briefly mentioned in the class - in one of the previous lectures that the distribution of the end-to-end distance actually turns out to be a Gaussian.

(Refer Slide Time: 25:47)



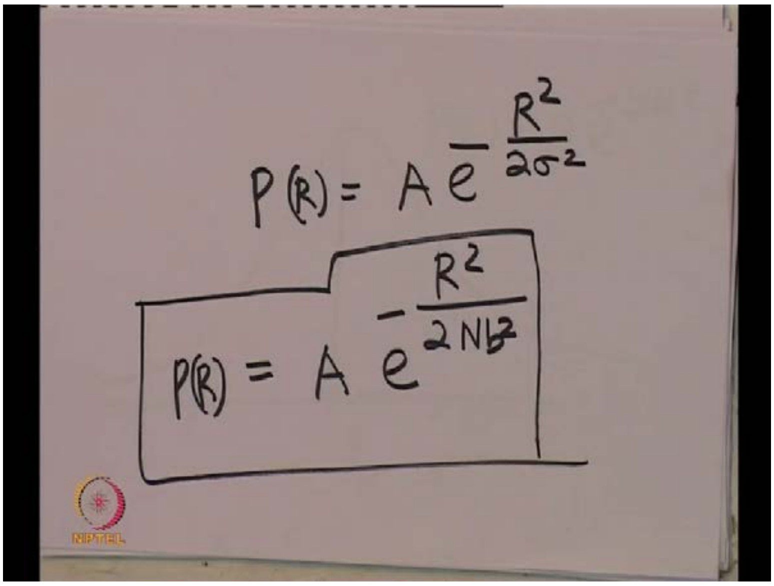
So, if you if you plot the distribution p of r verses r that we will get as a Gaussian with r if you plot r 0 to L. It will be a Gaussian style with p cut 0; so if you so this is this will this will have a shape of a Gaussian so if I draw it properly.

(Refer Slide Time: 26:18)



If I draw this, if you draw it properly, p of r has a e power minus b r square. So, it will have this Gaussian shape of a Gaussian and we know the standard deviation and average so if you know the standard deviation and if average we know that the Gaussian distribution will be if we know the standard deviation and average p of r has to be a e power minus r square.

(Refer Slide Time: 26:50)



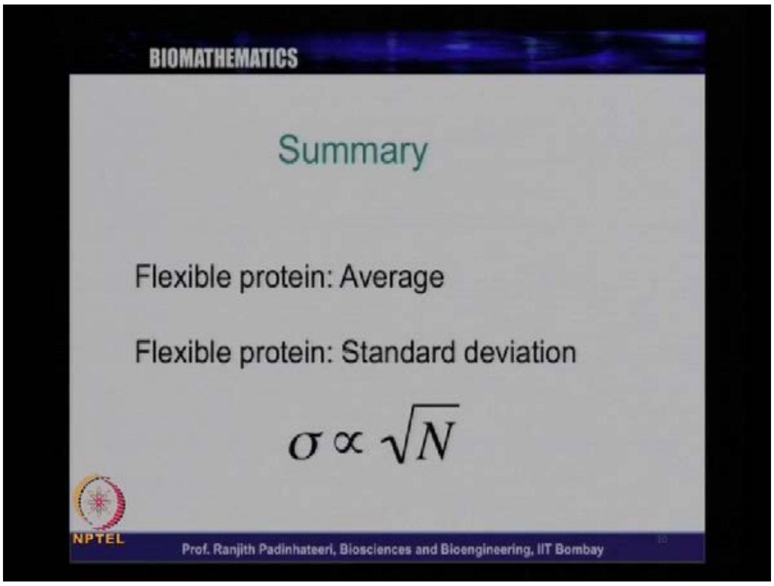
$$P(R) = A e^{-\frac{R^2}{2\sigma^2}}$$

$$\boxed{P(R) = A e^{-\frac{R^2}{2Nb^2}}}$$

By 2 sigma square that is what we learnt; so, what this would mean that the distribution is a e power minus r square by 2 sigma square sigma. We learnt as sigma square was n b square; so we found that sigma square is n b square.

So, the probability distribution for the particular protein having random configuration; it turns out that it can be shown that it is this e power minus r square by 2 n b square. We will, we can especially show this if there is time towards the end of this course. If there is more time, we have once we cover our main part of the… we may exclusively show this in another lecture. But, just realise that this is the distribution function so this is an example of Gaussian distribution happening in biology. If you take up model for a protein or a polymer of flexible polymer r to be more precise, this can be single standard DNA. If you do experimentally look at single standard DNA as a flexible polymer you are likely to get this particular answer. Now we know these things; now we will look at to summarize this part, this is first part of this lecture and summarizes this part.

(Refer Slide Time: 28:40)



Summary

Flexible protein: Average

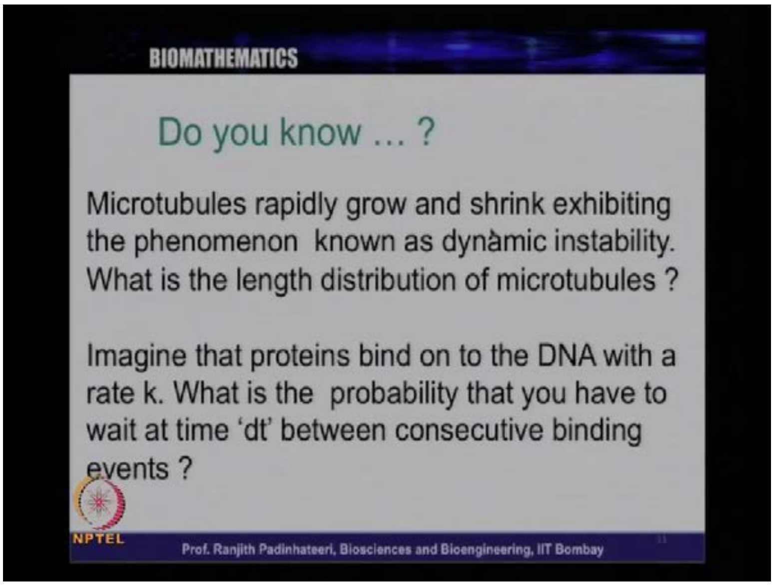
Flexible protein: Standard deviation

$$\sigma \propto \sqrt{N}$$

The summary of this part is that the flexible protein is if you took, we followed the average for a flexible protein, the average end-to-end distance, we found out the standard deviation of the end-to-end distance and we showed that the standard deviation is proportional to the square root of the number of monomers. So, this is the summary of the first part.

Now, we will switch gear to a different case; another, so we have seen this in general distribution functions; we discussed Gaussian distribution, normal distribution, we apply this idea of normal distribution to different cases, even to understand like many cases. Then, now we will try and understand another distribution function which might be very useful in biology. So, before doing that let us ask ourselves some question.

(Refer Slide Time: 29:49)



BIOMATHEMATICS

Do you know … ?

Microtubules rapidly grow and shrink exhibiting the phenomenon known as dynamic instability. What is the length distribution of microtubules ?

Imagine that proteins bind on to the DNA with a rate k. What is the probability that you have to wait at time 'dt' between consecutive binding events ?

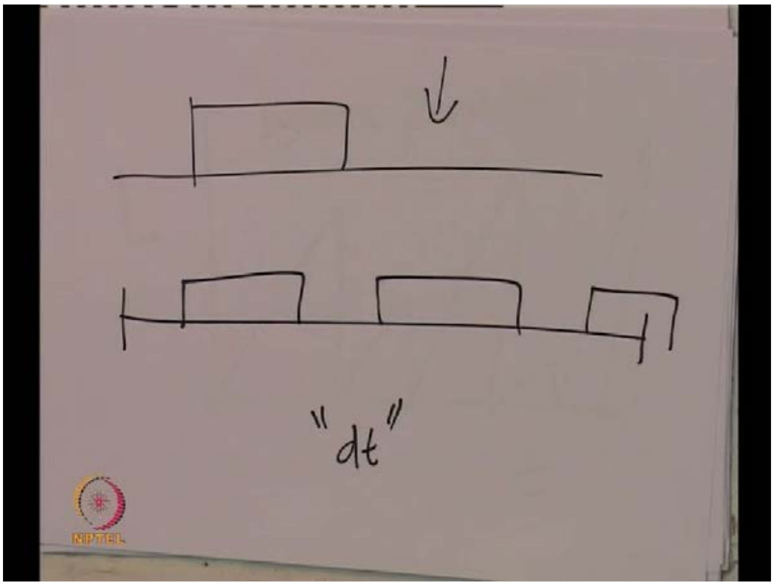
Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

The some question which I wanted to ask you is that, you know some… like something about microtubules; you know that microtubules rapidly grow and shrink exhibiting the phenomenon known as dynamic instability. You might have heard of dynamic instability otherwise, go and have a look at these text books, what is dynamic instability of microtubules. So, it turns out that dynamic instability is a phenomenon where microtubules rapidly grow and shrink. That is if you look at the length verses time for microtubules; what you would see is that they grow rapidly and shrink.

(Refer Slide Time: 30:25)

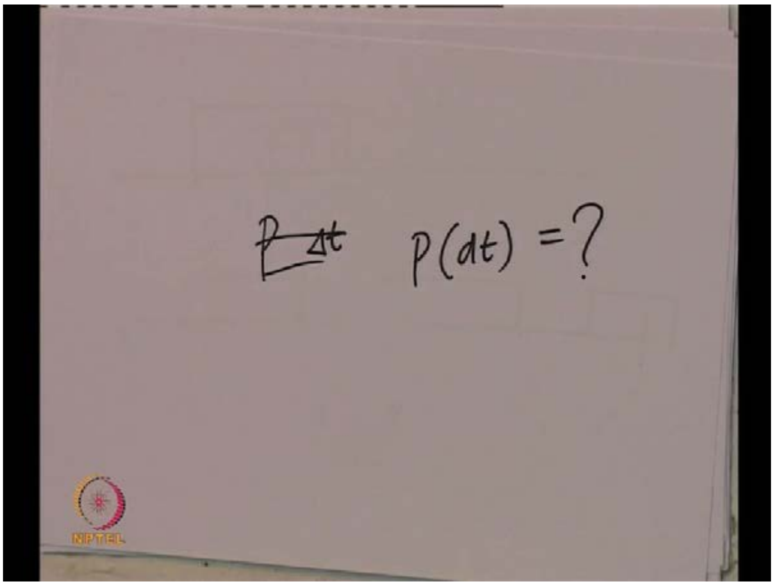
So, this is rapid growth and shrinkage; this kind of phenomenon is known as dynamic instability for microtubules. If you know length as a function of time, if you know the length as a function of time, we can always calculate the average length. From this, you can calculate the average length and we can also calculate the standard deviation. We can even calculate the distribution of length and did you know what the distribution of this length is? We will discuss that now. Also, imagine that proteins that bind on to the DNA and imagine that they rate by with a rate k, what is the probability that you have to wait at a time of d t between consecutive binding events? Imagine that you have a binding unbinding event happening and you have to, if we let us say there is a binding event happening with a rate k so consider this example of you know imagine that you have imagine proteins binding onto DNA so imagine that you have a (( )) DNA here so look at this paper and what you have is look at here.

(Refer Slide Time: 32:27)



Imagine DNA and you randomly let us say, proteins bind on to this DNA and you ask the question let us say 1 protein binds how long you have to wait for the next protein to bind so 1 protein comes and binds here. How long you have to wait for the next protein to come and bind? Let us say, d t is the time you have to wait for the next binding event to happen. In such a case typically, what will be this distribution of d t?

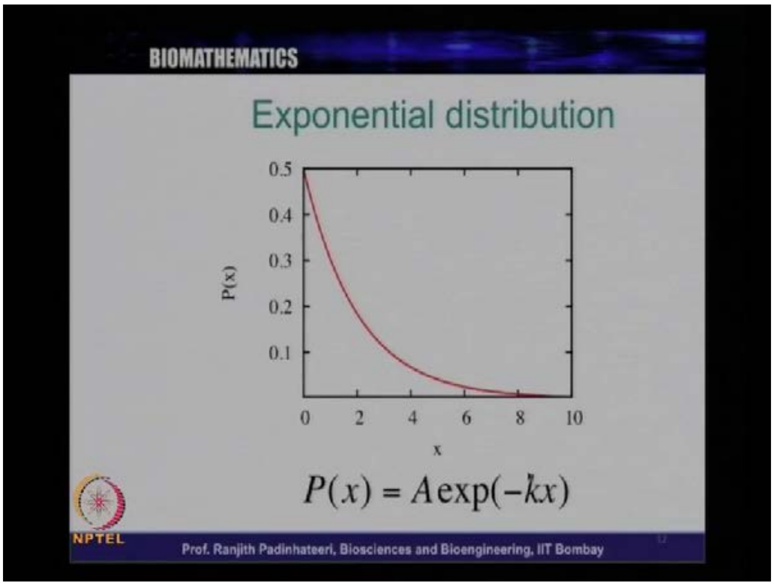
(Refer Slide Time: 33:03)



$P \Delta t \quad p(dt) = ?$

How long typically you all have to wait so p of d t? What is this? It turns out that the answer to this question is something called an exponential distribution.

(Refer Slide Time: 33:19)



BIOMATHEMATICS

Exponential distribution

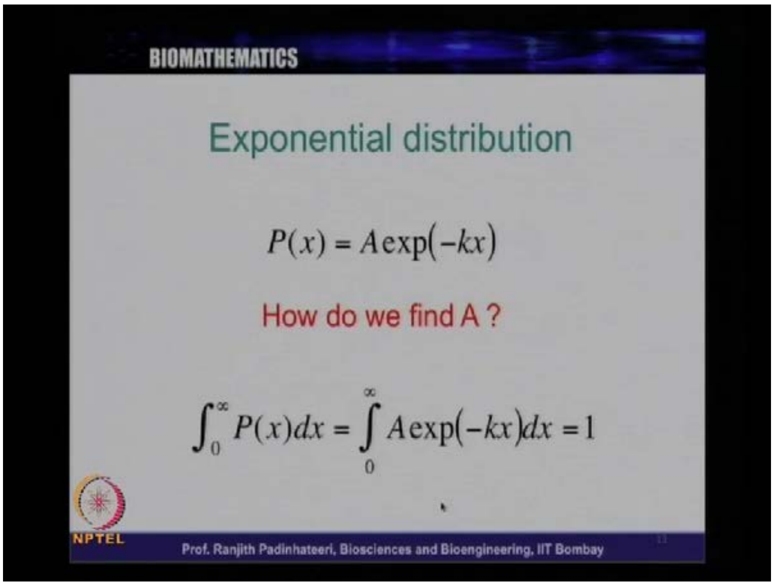
$$P(x) = A\exp(-kx)$$

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

Exponential distribution is having this particular form p of x is a e power minus k x which is a peak at 0 and it is coming down as a function of the parameter x. So, this distribution is called exponential distribution; length of microtubules during dynamic instability they will have exponential distribution that means small length we are likely to find small microtubules more than very long microtubules less likely to find very long microtubules.

Similarly, if you ask the question how long you have to wait it turns out that you would not have the probability that you have to wait very long is less and the probability that you have to wait short is more so the event will happen immediately as the large probability and even the event that it will happen after a long time as a very little probability that is what it precisely means so you know in such a case if you (( )) distribution what are the property of this distribution like average a what is the standard deviation and so on and so forth that is our interest to find out here discuss here so first question if you have this p of x is a e power minus k x so the first question is surely what is a.

(Refer Slide Time: 35:00)



BIOMATHEMATICS

Exponential distribution

$$P(x) = A\exp(-kx)$$

How do we find A ?

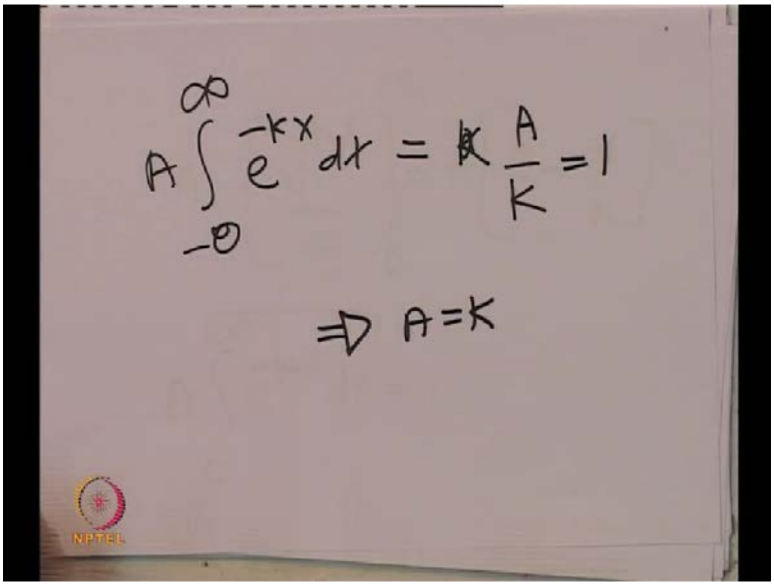
$$\int_0^\infty P(x)dx = \int_0^\infty A\exp(-kx)dx = 1$$

NPTEL

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

How do we find a? That is the question and like we did before we know that the integral of this the total probability has to be 1. So, the integral p of x d x is equal to a e power minus k x d x integral this has to be 1. What does that mean?

(Refer Slide Time: 35:33)



$$A\int_{-0}^{\infty} e^{-kx}\,dx = \cancel{k}\,\frac{A}{K} = 1$$

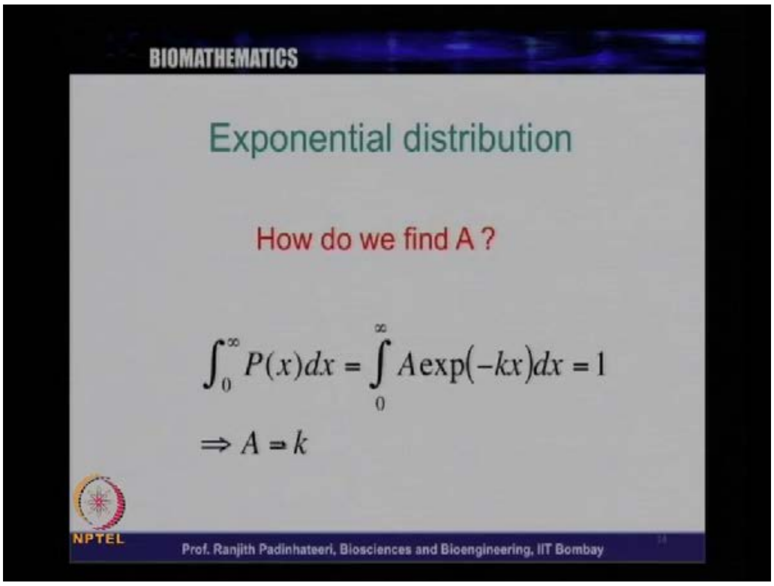
$$\Rightarrow A = K$$

If you have an integral so what we have we have? p of x is a e power minus k x and we want p of x d x equal to 1. That means, integral a e power minus k x d x is 1 this means that what does this mean, let us do it here this means.

This means that a is not a function of x; so you can take a outside so a integral e power minus k x d x equal to 1. What is integral e power minus k x d x? What is the limit the limit is 0 to infinity because, length of microtubule is typically either 0 or between 0 and infinity. It can go mi negative the time you have to wait typically cannot go negative. So, this is all this case is, this is the integral is from 0 to infinity. So, let us do this 0 to infinity; what is integral e power minus k x from 0 to infinity is so integral of e power minus k x is e power minus k x by minus k.

This is the integral of e power minus k x limit 0 to infinity with a constant a here what is this is a times if you apply this x is if you put x is equal to infinity e power minus infinity is 0 so this is 0 and minus when x equal to 0 this is minus 1 over k so the answer integral p of x d x turns out that a integral 0 to infinity e power minus k x d x is k, sorry, this is 1 over k. This answer is 1 over 0 minus K with an a here and this has to be 1; so this has to be 1 that implies that a is equal to k if a by k has to be 1 a has to be equal to k.

(Refer Slide Time: 38:20)



BIOMATHEMATICS

Exponential distribution

How do we find A ?

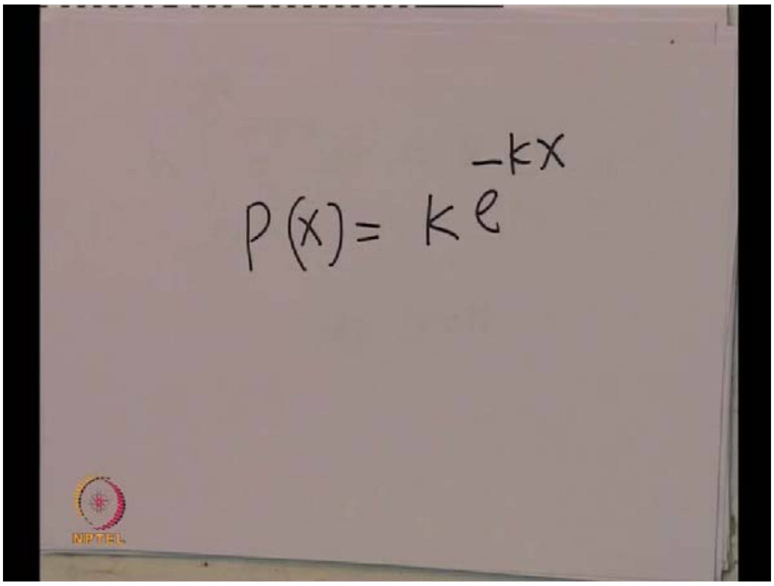
$$\int_0^\infty P(x)dx = \int_0^\infty A\exp(-kx)dx = 1$$

$$\Rightarrow A = k$$

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

This is what is shown here the way to find a is that you do this integral 0 to infinity p of x d x and equate that to 1 we will find that a is nothing but, k so our distribution function.

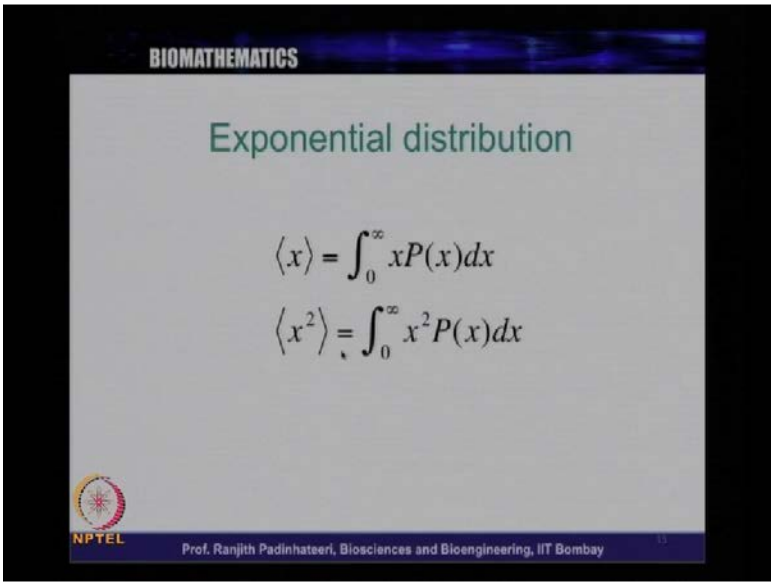
(Refer Slide Time: 38:39)



$$P(x) = k e^{-kx}$$

P of x is nothing but, k e power minus k x such that p of x g of x is 1. If this is the case, now we know the distribution function. How do we find the average? What is the average? Of this the average of this is so the average of this is defined in this per.

(Refer Slide Time: 39:01)



BIOMATHEMATICS

Exponential distribution

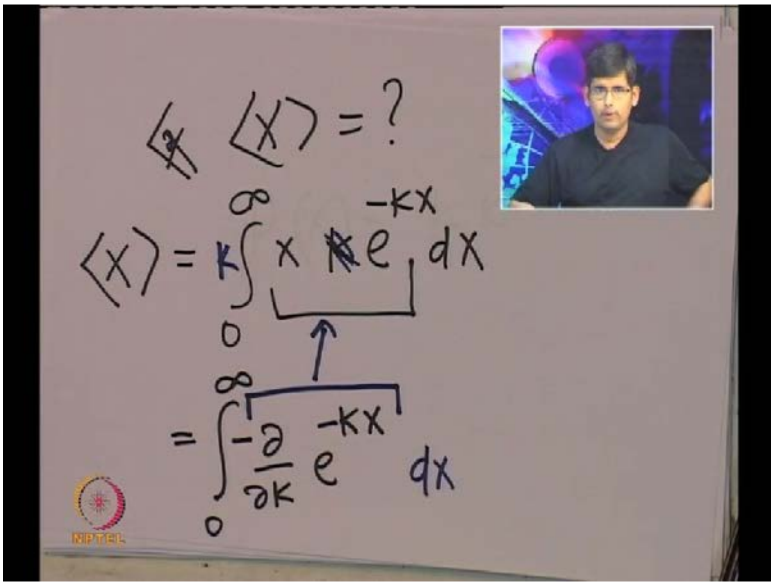
$$\langle x \rangle = \int_0^\infty x P(x) dx$$

$$\langle x^2 \rangle = \int_0^\infty x^2 P(x) dx$$

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

Typical fashion so the average for any distribution function is defined as 0 to infinity integral x p x d x. So, the integral here is 0 to infinity because the x can go only to 0 to infinity. In this case that we are interested in so x average is 0 to infinity x p of x d x x square average is 0 to infinity x square p of x d x so what is x average and what is x square average? Let us calculate this x average and x square average so let us first calculate x average.

(Refer Slide Time: 39:46)



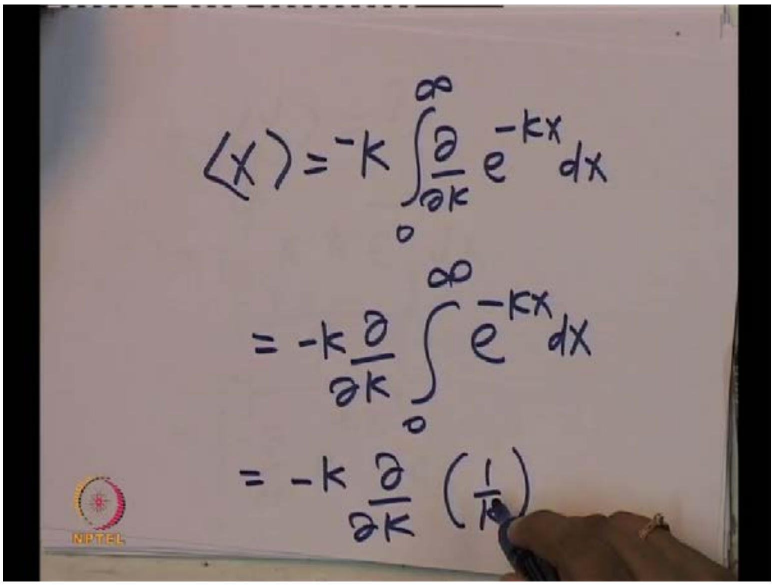
$\cancel{\langle x} \; \langle x \rangle = ?$

$$\langle x \rangle = k \int_0^\infty x \, \underbrace{\cancel{k} e^{-kx}} \, dx$$

$$= \int_0^\infty -\frac{\partial}{\partial k} e^{-kx} \, dx$$

X average can be defined as 0 to infinity x. The distribution function is k e power minus k x d x if you look at here we use a technique that x e power minus k x so this is 0 to infinity x e power minus k x can be defined as del by del k of e power minus k x with a minus sign so this is you look at this if you look at del by del k of e power minus k x with a minus sign this if you look at this if you look at del by del k power k minus x with a minus sign this is x e power minus k x d x you can do this yourself and convince yourself that del by del k will give you an m minus x and with a minus sign will be plus x and there is a d x here so there is a d x here so this is what essentially this integral is equal to this so x square average is this sorry this is not equal you can just take k out so this is what it is x e power minus this integral is over x so you can k take k out so this quantity e power minus k x is nothing but, minus del by del k of e power minus k x.

(Refer Slide Time: 41:42)



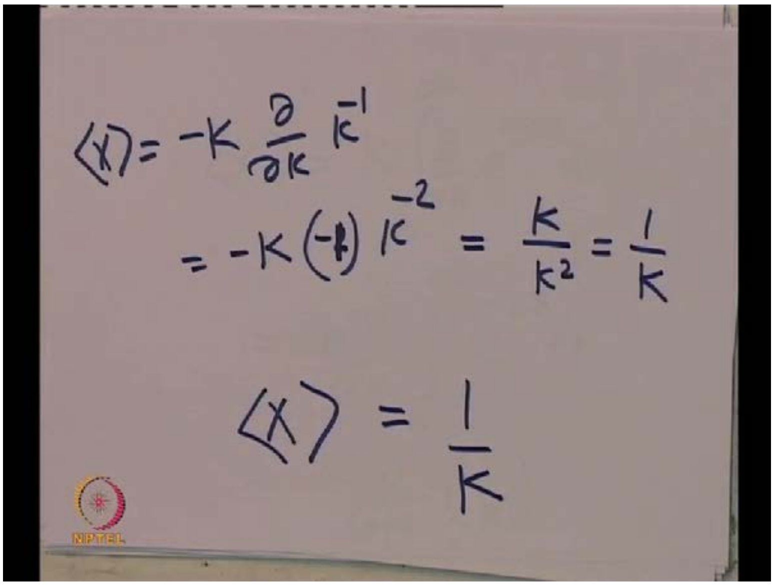
$$\langle x \rangle = -k \int_0^\infty \frac{\partial}{\partial k} e^{-kx}\, dx$$

$$= -k \frac{\partial}{\partial k} \int_0^\infty e^{-kx}\, dx$$

$$= -k \frac{\partial}{\partial k} \left(\frac{1}{k}\right)$$

So, this is you can convince yourself that this is true. If this is true x square average can be written as X square average can be written as we saw k with a minus sign 0 to infinity del by del k of e power minus k x d x this is minus k del by del k of 0 to infinity e power minus k x d x. We saw that 0 to infinity e power minus k x d x is 1 over k; so this is minus k del by del k of 1 over k so what is del by del k of 1 over k del by del k of 1 over k is this k so this is del if you look at it del by del k of 1 over k is del by so we have minus k times del by del k k power minus 1 which is so this is x average minus k times minus 1 times k power minus 2 so this is k by k square so the answer is 1 over k.

(Refer Slide Time: 42:37)



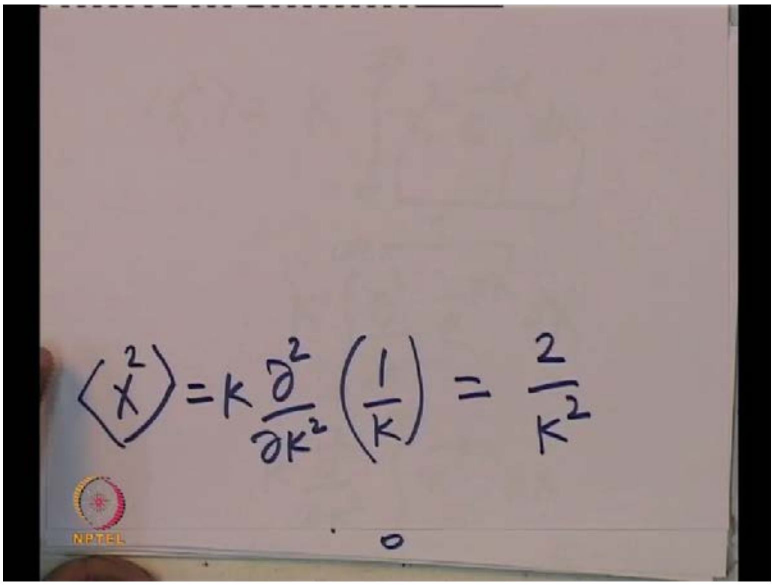
$$\langle x \rangle = -k \frac{\partial}{\partial k} k^{-1}$$

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

$$\langle x \rangle = \frac{1}{k}$$

So the x average turns out to be 1 over k so the x average the average x average for an exponential distribution is 1 over k now what is x square average.

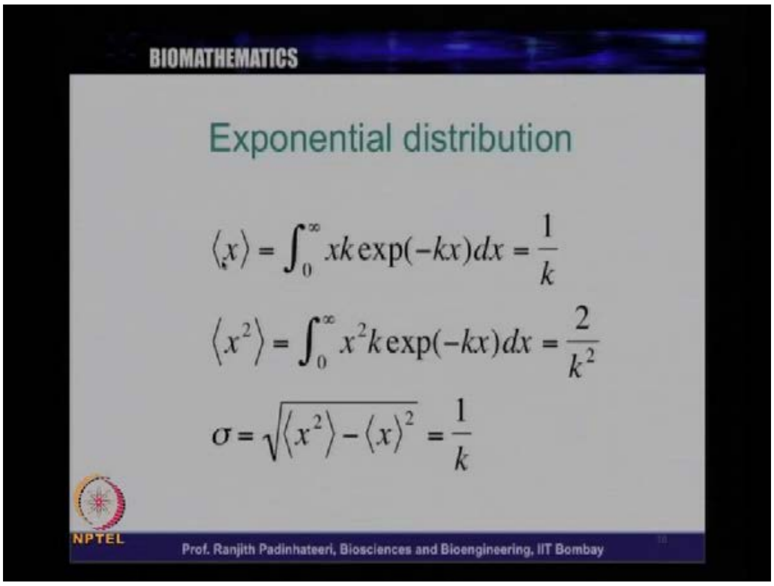
(Refer Slide Time: 43:21)



$$\langle x^2 \rangle = k \frac{\partial^2}{\partial k^2} \left(\frac{1}{k} \right) = \frac{2}{k^2}$$

As we typically define x square, average is defined as integral 0 to infinity x square p of x d x so here x square k e power minus k x d x integral 0 to infinity so you can take this k outside and write it as and write this thing as x square average can be written as k into integral 0 to infinity x square e power minus k x d x and the standard technique is that you can write this x square e power minus k x so this is k this is 0 to infinity. This part has del square by del k square e power minus k x d x if you do this and just look it is exactly this so this we saw we can see that this is this so just like we did before if we just you can convince yourself that del by del square and e power k x is x square e power minus k x and there is a d x sorry the d x now if you do this what you get is that k into del by del del square del k square 0 to infinity e power minus k x d x so you know that 0 to infinity e power minus k x d x is 1 over k so this is k del by del square k by del k square 1 over k and if you do this the answer is 2 over k square so this is what x square average is so x square average turns out to be 2 over k square and x average is 1 over k.

(Refer Slide Time: 45:50)



BIOMATHEMATICS

Exponential distribution

$$\langle x \rangle = \int_0^\infty x k \exp(-kx)\,dx = \frac{1}{k}$$

$$\langle x^2 \rangle = \int_0^\infty x^2 k \exp(-kx)\,dx = \frac{2}{k^2}$$

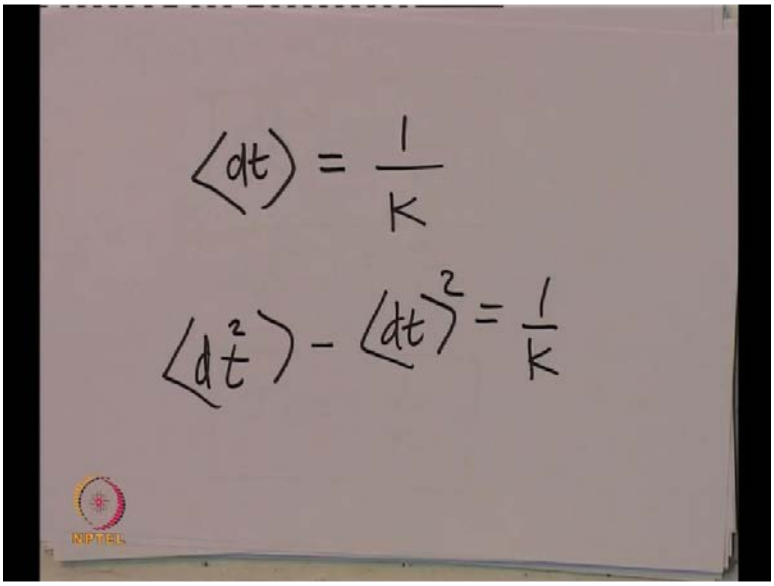
$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \frac{1}{k}$$

NPTEL

Prof. Ranjith Padinhateeri, Biosciences and Bioengineering, IIT Bombay

What do we get? We get x average is equal to 1 over k x square average is 2 over k square and now we can calculate the standard deviation. The standard deviation is sigma is equal to x square average minus x average square which is 1 over k so if you look at here what do we have?

(Refer Slide Time: 46:19)



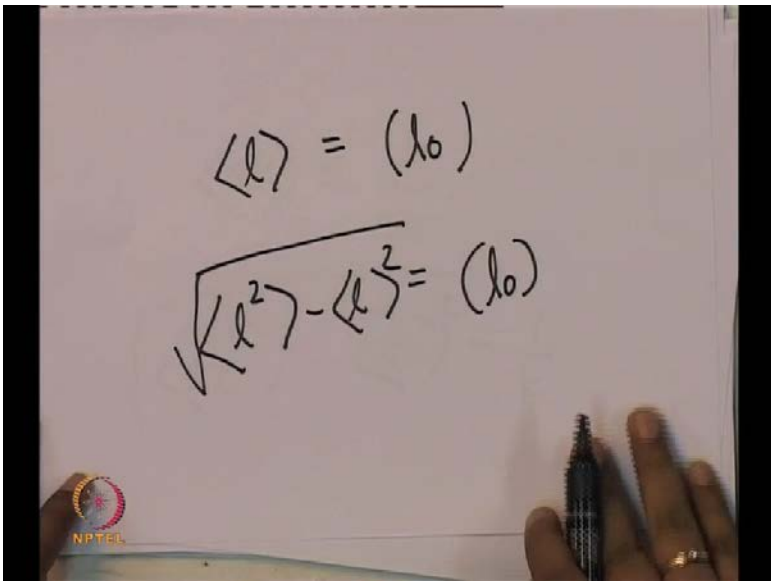
$$\langle dt \rangle = \frac{1}{k}$$

$$\langle dt^2 \rangle - \langle dt \rangle^2 = \frac{1}{k}$$

We have x average 1 over k x square average 2 over k square standard deviation 1 over k so what are we finding we find that the standard deviation and the average are the same for exponential distribution this is an very interesting property that it has average and standard deviation the same value.

This is an interesting thing to note or remember that exponential distribution of average the fluctuation the so if you if you if you look at the length of a microtubule now or if you look at the time that you have to wait d t the average time you have to wait will be the 1 over the rate or k so here you have a binding event happening with a rate k the average time you have to wait is 1 over k but, the deviation in this average that is d t square average minus d t average square is also 1 over k so this is the property of an exponential distribution and this is same for microtubule length.

(Refer Slide Time: 47:58)

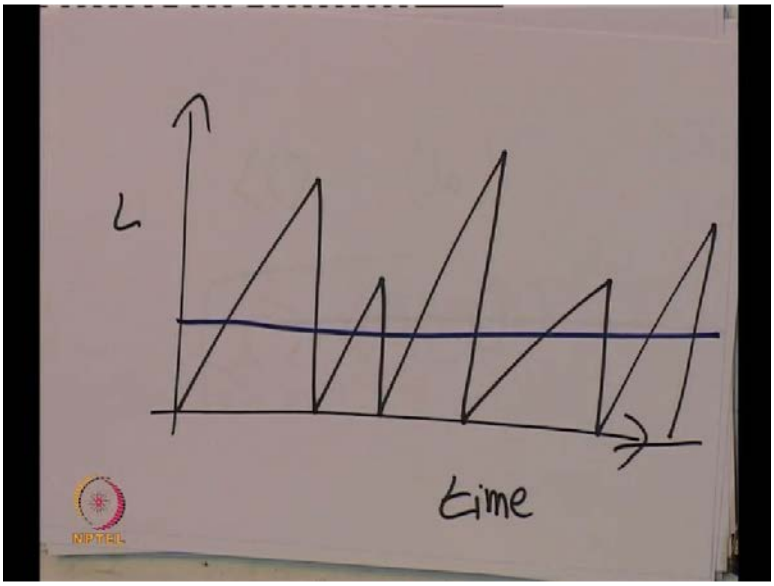


$$\langle l \rangle = (l_0)$$

$$\sqrt{\langle l^2 \rangle - \langle l \rangle^2} = (l_0)$$

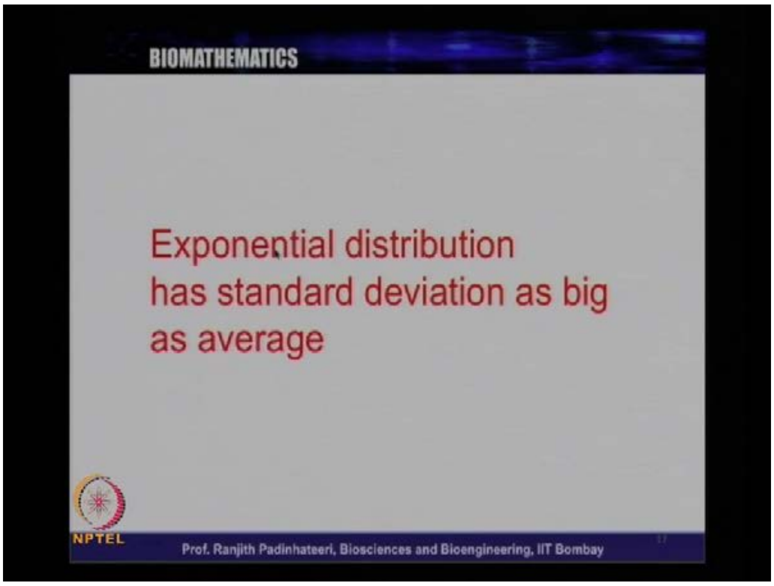
If the average length is some particular number and you will find that l square average minus l average square will also be the same number, so if this is some l 0 this will also be l 0, sorry, square root of this. Then, here it is square root of this; so, square root of this will be a number l 0 which is same as the average. What does that mean for a microtubule?

(Refer Slide Time: 48:37)



Look at length verses time for a microtubule. What you find like it grows, shrinks, grows, shrinks, grows, shrinks, grows; so this if you just follow this growth and shrinkage the average length. The average length you see and the standard deviation - the deviation from the average, they are roughly the same. So, the standard deviation is as big as the average. This is a property of exponential distribution that the standard deviation is as big as the average itself.

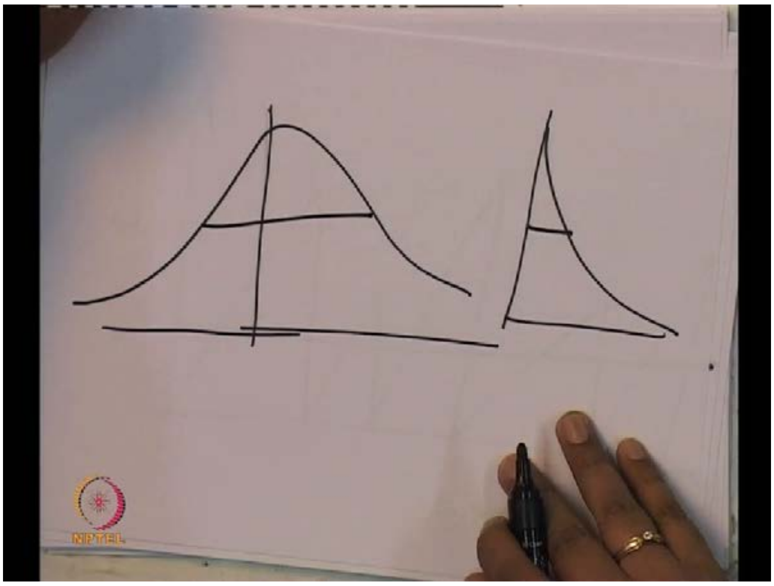
(Refer Slide Time: 49:32)



Exponential distribution has standard deviation as big as average

So, this is one thing to remember; that exponential distribution has a standard deviation as big as the average. So, dynamic instability have this property that the deviation - the length deviation is as big as the average itself so to summarize the problem in exponential distribution we learnt the average we learnt the standard deviation and we learnt that the standard deviation and average are the same. Now we can also as the similar question we asked for Gaussian distribution we asked for Gaussian distribution.

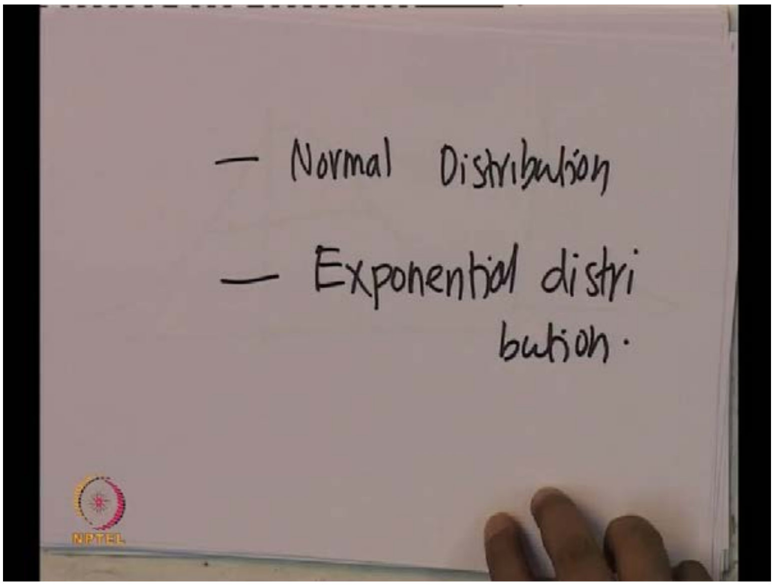
(Refer Slide Time: 50:07)



If you have a Gaussian distribution, what is the width at half maximum? Similarly, you can also ask the question what is the width at half maximum? So, do these questions yourself and find out the width at half maximum for each case, for standard exponential distribution as well as the Gaussian distribution. So, we learnt 2 kinds of distributions normal distribution exponential distribution.

These 2 distributions are very useful in biology. It appears in different, many different places. You will see that they come very often in biology and then you will have to use this, similar to know something about this, the property of this, you will have to always remember there average and standard deviation at least and the shape of distribution.

(Refer Slide Time: 50:44)



- Normal Distribution
- Exponential distribution.

So, with this idea, with these few things that we learned today, about exponential distribution normal distribution and property of flexible polymers, we will conclude today’s lecture and we will continue discuss on statistics in coming lectures; so, see you and bye for now.