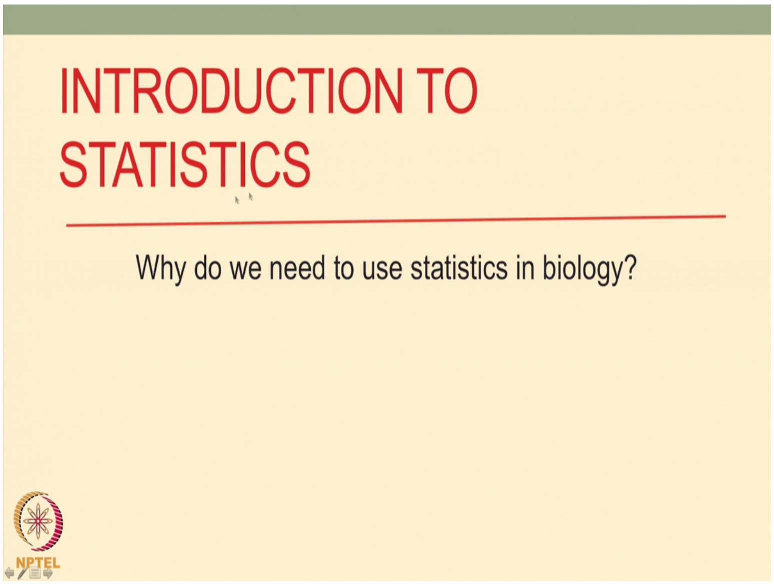
Introductory Mathematical Methods for Biologists
Prof. Ranjith Padinhateeri
Department of Biosciences & Bioengineering
Indian Institute of Technology, Bombay

Lecture – 35
Introduction to Statistics

Hi, welcome to this lecture on Mathematical Methods for Biologists; in this lecture we will start discussing about statistics.

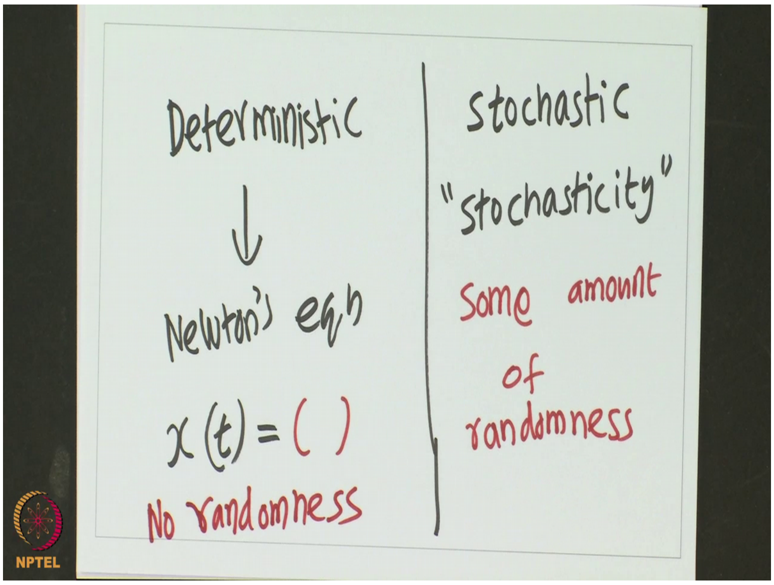
(Refer Slide Time: 00:28)



So, the title of this lecture is introduction to statistics and the question that we will answer in this lecture is; why do we need to use statistics in biology? This is the question and we will discuss some introductory things related to statistics.

So, let us start thinking about; why do we need to use statistics at all? So, there are two things that; two terms that you would want to know and these are called deterministic and stochastic.

(Refer Slide Time: 01:01)



So, these are deterministic and the other one is stochastic. You would often hear the word stochasticity somewhere you might hear this word stochasticity. So, let us quickly understand what does it mean?

So, in the world around us if we look at for example, if we ask a question; if I take a cricket ball or a football or a heavy ball, iron ball and drop it and ask the question; how long will it take for that ball to reach the ground? So, if I leave it from a 1 meter height and ask a question; how long it will take to reach the ground? 1 meter down and that answer can be obtained by solving the Newton's equations.

Something that you learn in school; like s is equal to ut plus half at square and things like that, you could solve various equations v is equal to u plus a t; these are the equations some set of equations that you learned in school, which are basically Newton's equations; equations of motion, we can solve them and obtain how long it would take.

And does not matter; if I do the experiment and somebody else repeat the experiment many times; the time would be exactly the same. Like if you drop a very heavy ball, the time can be precisely calculated and the answer will not change; under the same conditions if you repeat the experiment many many many times; essentially the answer will be the same.

On the other hand many things in biology this may not be the case. So, such things like Newton's equations if you solve the answer you get. For example, position as a function of time; this answer from Newton's equation is a deterministic outcome. You know this answer, there is only one answer and this how many r times you do; this experiment under a exactly same condition in a very precise condition? Once you fix the conditions, this answer the time or the distance; if I throw a ball and if I know the initial velocity and angle and all kinds of things needed to solve Newton's equations, I can precisely predict.

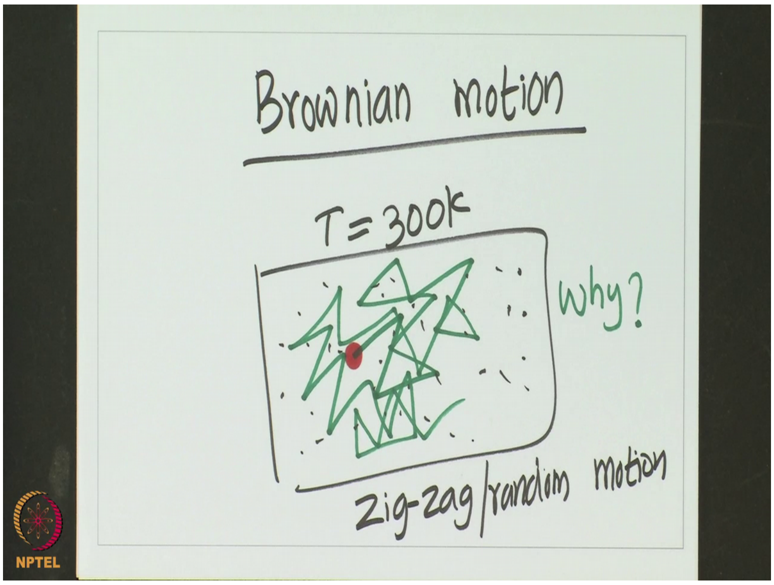
Similarly, like the motion of planets one could pretty precisely predict using equations of classical physics. So, those are very deterministic processes; however, things in biology is not really deterministic, they are stochastic. What does that mean? There is some amount of randomness in this process. So, there is some amount of randomness; how much is the randomness? So, there is some amount of; the processes with some amount of randomness is typically called stochastic processes. And this is a deterministic processes because there is no this is like determined, there is no randomness no randomness.

Example, as we said the motion of planets; the motion of a heavy ball under gravity or things like what one would describe using classical mechanics. Those kind of things are; many of those things are deterministic and; however, what happens in biology? What happens in a living system is slightly different; one important extra feature plays an important role in living system; that is temperature.

Temperature is not really affecting the motion of the planets as we would study; however, the temperature is an important factor for living systems; without the right temperature the systems cannot live. So, this temperature introduces some new features and through this temperature this randomness comes in.

So, let us first look at how randomness appear at the molecular level? So, many of you might have heard this name called Brownian motion.

(Refer Slide Time: 06:13)

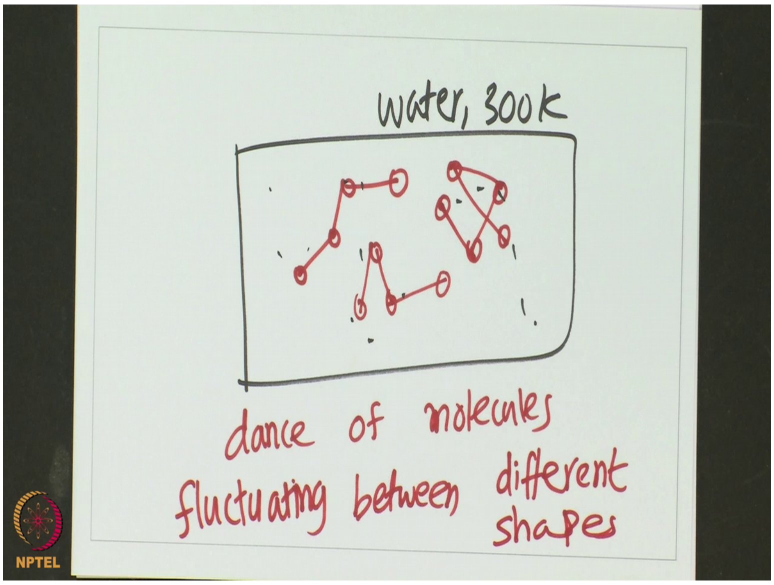


So, let me talk about this name which is a Brownian motion; what is Brownian motion? If I take water for example, so this is water at 300 Kelvin; a room temperature or 310 Kelvin, the typical temperature that you would want. And you put a protein molecules in this water, this protein molecules will randomly move around.

So, first the protein molecule will follow some random path like this; they will just keep moving they will not a small protein molecule if you put they would follow some random path. If assuming that the mass of this protein molecule is very small; so, if that is the case this protein molecules will do some random motion.

Why this random motion? Why? Similarly, if I take a extended protein of; protein molecules will have another kind of phenomena. Here we assume the protein molecule is a globule and folded nicely and it is like a dot; this is not the case, protein molecules are its polymers and if I think of a protein molecule in water.

(Refer Slide Time: 07:50)



So, there is water everywhere which I am not really showing. So, water 300 Kelvin if I put a protein molecule some particular kind of peptides or proteins, they can take various shapes or if it is a DNA. Or any biopolymer if you put in, they can take same protein molecules could take different shapes. So, this is one shape; sometime they will take some shape like this. So, they would fluctuate between different shapes; so, there will be a dance of molecules fluctuating between different shapes.

The amount of fluctuation will differ, but there can be some fluctuations; many proteins, if I take the protein; they would fluctuate between different shapes like this. If you think of this as monomers of a protein and if you imagine them in water; they could take various shape, this is one shape; this is some shape, they could fluctuate between different shapes as we observe; this is possible.

So, how is this happening? Why is this happening? And the answer to this is a phenomenon called a Brownian motion. So, let us come here; what happens in the Brownian motion is the following. The water molecule does a random motion because of the temperature. So, temperature is nothing, but the kinetic energy; so, the molecules of the medium that is the water molecules are doing a zigzag motion. So, they are doing a random zigzag; water molecules are doing a random motion and this random motion is basically the temperature.

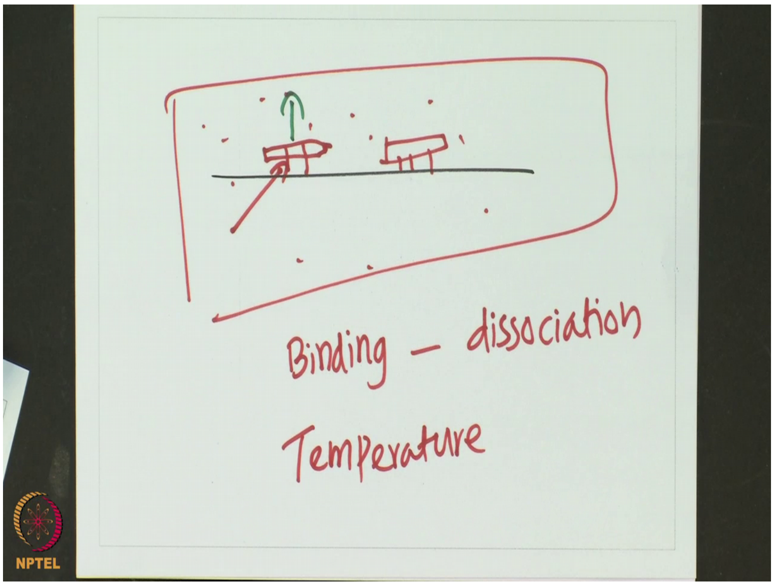
So, temperature is nothing, but the kinetic energy due to random motion; something that we learned in school that temperature is essentially is coming from the kinetic energy due to this random movement. So, this random movement of water molecules will; this is water molecules, which are randomly moving will go and hit the protein molecules and when they hit the protein molecules will be displaced.

And when it hits each amino acids each amino acids will be displaced randomly and that could lead to some kind of fluctuations of the shape itself. You take a single standard DNA or very long double standard DNA. And if you could look at the polymer itself; one could see they fluctuating; attaining different shapes and fluctuating, there will be shape fluctuations.

So, all of this is because at 300 Kelvin or at finite temperature or at higher temperatures, the water molecules are doing zigzag random motion. And this random motion of water molecules will affect the proteins or molecules in water itself. So, the inherent nature of water 300 Kelvin; the fact that temperature means random motion of molecules would lead to some random motion of protein molecules also; and this would lead to some amount of randomness in the system itself. And this can happen in many contexts for example, if you have a DNA and the protein is bound.

So, let us take this example of DNA and protein.

(Refer Slide Time: 12:00)

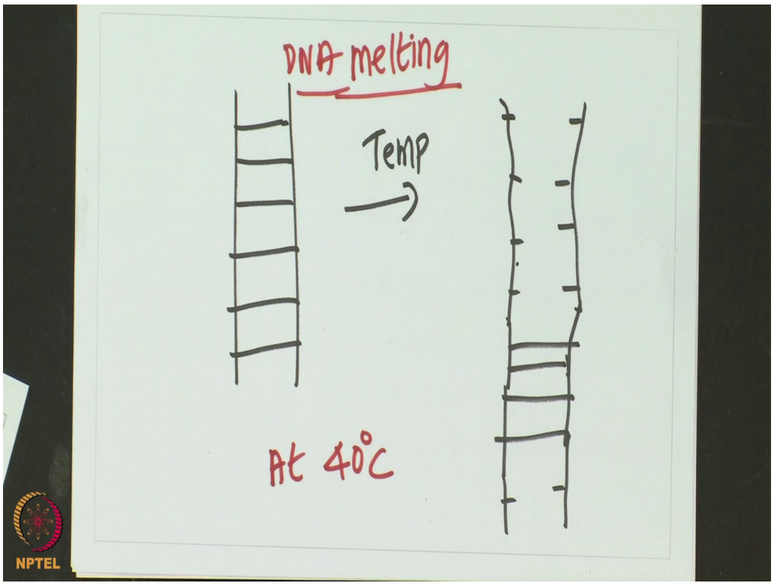


So, you have DNA on which proteins are bound; so, there is proteins are bound and this bonds can be broken when this is in water. So, water molecules are doing random zigzag motion and this water molecules can go and hit and break this bonds. And because of this breaking bonds, this protein molecules will come off.

So, there can be binding and dissociation; so, this binding dissociation process happens because of the temperature. At 0 temperature; if the temperature was very close to 0 Kelvin then all proteins would be bound; there will be no dissociation at all.

So, the binding dissociation of proteins essentially happen because the water molecules that doing zigzag motion and they are going and hitting, the protein molecules and the DNA. And because of that, this DNA protein directions break and they come apart. The melting of DNA is another phenomenon that you have heard, where you have a double standard DNA at due to because of the temperature the DNA strands break.

(Refer Slide Time: 13:32)



So, you have double stranded DNA; which is this hydrogen bond ATGC hydrogen bonds, at higher temperatures this will melt so; that means, this bonds will be break broken. And of course, there will be some bonds here all the bonds need not break an So, on so forth.

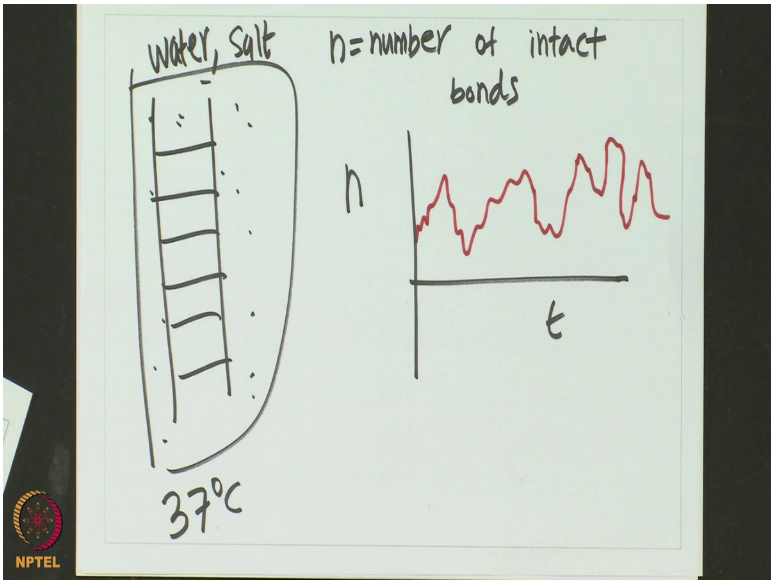
So, there will be some bonds will be broken at because of the finite temperature. So, as the temperature increases; there is a melting of DNA. So, this is called DNA melting sometime; basically the breaking of the hydrogen bonds. So, this also happens because the water molecules will go and hit this DNA molecules and because of that this bonds will be broken.

So, temperature is driving many random events at the molecular level because of this there will be lot of randomness. So, let us ask for example, a simple question of this DNA melting itself like. So, if I ask a question that at 40 degrees Celsius; if I have a DNA of let us say 10 base pair or 20 base pair; how many of this DNA bonds will be impacted?

And how many of the DNA bonds will be broken this is a question that we can ask and if I do this experiment at 40 degree Celsius for example, what I would see is that even at 37 degree Celsius; DNA would be occasionally this DNA double strand bonds will be the hydrogen bonds would be occasionally broken and they will reform.

So, at any time if I imagine that if you could take a snapshot of DNA; the double helix bonds, all the bonds would not be bound; some of them would be open even in vitro; let us forget in vitro; if you take a piece of DNA put it in vitro at 37 degree Celsius and look at the DNA bonds and you would see that the number of intact bonds would vary with time. So, if I plot the number of intact bonds with time; so, this is the experiment that we are thinking about.

(Refer Slide Time: 16:30)



So, we are thinking about taking a DNA at in with appropriate conditions; like water, ions and all that. So, there is salt and all that appropriate in vivo like conditions so, but only DNA and the temperature is 37 degree Celsius.

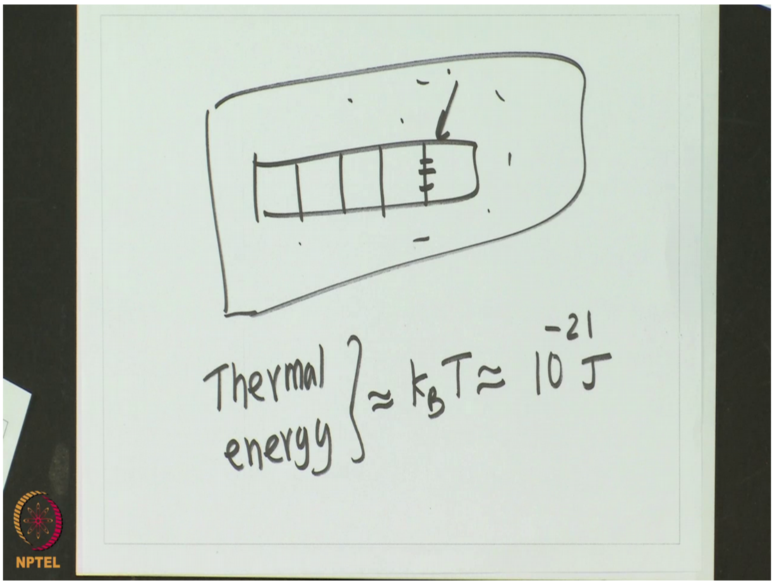
And then we will plot in n; which is as a function of t, where n is number of intact bonds. As suppose to the broken this is number is likely to fluctuate; so, this number would fluctuate, it could fluctuate like this. I do not know how would you fluctuate, but this will not be a constant number as the time goes; even though under given conditions this number of intact bonds could fluctuate. Therefore, this answer is not a constant; even though the conditions are exactly the same.

You could think of another question; that if you take a polymer which is a microtubule for example, or actin typically microtubule. And if you observe this in vitro, you would see that its length is increasing and decreasing. So, the length of microtubule is not a constant; it is increasing and decreasing all the time it is fluctuating. So, these are quantities where there is some lot of fluctuations, things are not static; things are dynamic and then under the same conditions, if I take different microtubules and look at.

So, if let us say you have 10 test tubes exactly the same conditions and you have microtubules and you look at 10 microtubules in each of the cell exactly same conditions, all conditions the same; at any time they could have different lengths. Their lengths would be 1 microtubule would have a different length another my test microtubule will be different lengths. So, the individual microtubule length at a given time; if we could take if we could take a photo of a microtubule in each of this test tubes, we would get different length

So, the length of a microtubule at a given time; I take a population of this microtubules will not be a constant value it will be a fluctuating value. And therefore, we need to think and introduce some quantity; which can deal with, this fluctuations, this variability; there will be a variability in length. So, if there is such variability at the molecular level which is the microtubule. So, this is fundamental reason; here is also this kind of Brownian motion.

(Refer Slide Time: 19:48)



In the case of polymers; What happens is that. So, you have a polymer microtubule or actin you have polymers whenever this is in water and water molecules which is doing a Brownian motion, will come and heat the polymer and they can break this bonds and therefore, it will lead to depolymerization.

So, the energy due to temperature is called a thermal energy. So, this thermal energy is the energy due to temperature; this KBT, the amount of thermal energy is KBT; if you do the math, you will add 300 Kelvin, this is appropriately 10 power minus 20 one joule. So, this 10 power minus 20 one joule thermal energy; if this energy is comparable to this bond energy or the other kind of energy of the hydro either; if I take this hydrogen bond energy in the DNA or even this bond energy in this polymers.

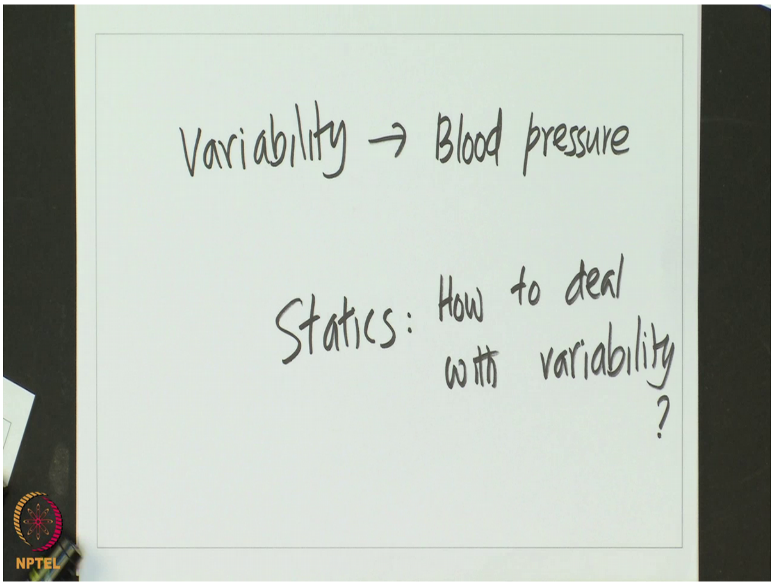
If those energy is comparable to this thermal energy which is the energy of the kick of the water molecule; then this bonds can break, there can be depolymerization, this will lead to some kind of randomness. So, this thermal energy is what is leading to this randomness and therefore, there is randomness in the molecular level.

And when there is randomness in the molecular level this could affect the gene expression there would be randomness in the gene expression level because proteins the DNA binding proteins would depend on the temperature. All the rate processes essentially the kind of there is a stochasticity and because this; the gene expression would be affected when the gene expression affects the cellular function itself would be affected.

Which means the shape of the cell various things the fate of the cell could itself get affected in principle by stochastic events and therefore, it is important to understand these stochastic events, the random events; the effect of randomness. So, we need a branch of (Refer Time: 22:03) mathematics which can deal with this kind of random events and therefore, that is why statistics is has emerged as a candidate to deal with this.

Now, in some other cases which is not really a molecular event for example, if we take people living in a city; even though the city has a same condition, broadly the same condition due to their background various other things, there is a lot of variability in the population. So, this is fluctuation due to thermal forces is one thing, variability arising from various sources. So, many of this variability can be finally, track down to this thermal temperature related Brownian motion.

(Refer Slide Time: 23:01)

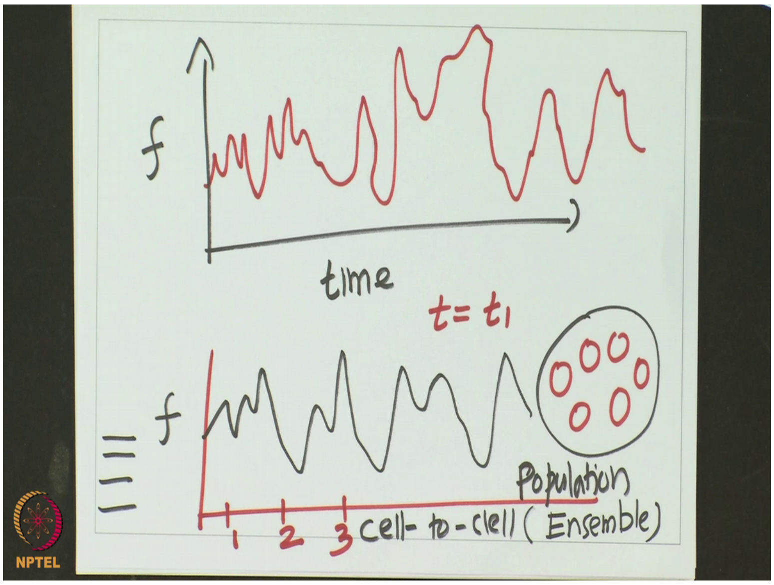


But there is lot of variability; if you look at blood pressure for example, something as a person doing livings life sciences. Something that would you would be interested for example, blood pressure; blood pressure even though there is a number that we would 1, 20, 80 or there is a typical number that doctors would suggest, there could be a lot of variability in the blood pressure itself. So, this is something that we would see it vary in nature itself temperature itself of course, can vary and many things in nature itself we will have lot of variability.

So, whenever there is a variability and which is there everywhere; to account for this, we would need to use a branch of mathematics and this is what we would use statistics for. How to deal with variability? This is something that we will learn and the statistics will teach us how to deal with variability? And how we could frame equations and how we could use some equations to describe some of those phenomenon. So, diffusion for example, is an example where some of this is applied, but we will learn few other things; so related to how to deal with this variability.

So, if we now think about a system where it is some quantity is varying for example, length is varying for a microtubule, something that we know. Number of bonds; DNA we saw and in many of the medically measurable quantities would be varying. So, very often all of this have the following characteristic that if I plot; this either as a function of time; so two things we will do variability.

(Refer Slide Time: 25:28)



So, this x axis could be time very often some quantity that you would measure some function that you would measure as a function of time; there would be a lot of variability this could be the case that there they could fluctuate.

So, this is one thing would come; so, very often at a particular time. So, let us say t is equal to like at a particular time t 1; you measure a particular quantity, which could be the length of a microtubule. In a population, so you have 100 cells; you measure the gene expression of a particular gene in 100 different cells or different different cells and there will be some variability in the gene expression itself.

So, there is a cell to cell variability or a population variability, variability across different people or constituents that we are thinking. So, this is at a particular time this could be the population or this could be, so this typically let us say a cell to cell. So, this would be cell number in the x axis; cell number or it could be persons number or this is call ensemble.

So, you have a collection of cells; so, you have a collection of cells. So, cell 1, cell 2, cell 3, cell 4. So, you have a collection of this is what is x axis; so cell 1, cell 2, cell 3,. So, you have different different cells; in each cell if you measure the amount of gene expressed or amount of protein expressed at a particular time; there also you could see some fluctuations maybe.

So, there could be a fluctuation which is in a population; so, this is also sometime I would call it population. So, if you take a population different individuals of the population, there will be individual cells or it could be individual whatever be the elements; it could be a need not even a cell it will be a microtubule.

If you have 10 microtubules each of their length under exactly same condition would could be fluctuating. So, this is an ensemble and then if you take population and if you look at population; each element of the population could have different length or different gene amount. And other thing is if you take same quantity as a function of time, you could see a variability.

So, these are the two kinds of variability that you would look as a function of time in the population. And in the coming lecture, we will learn how to deal with this variabilities and how we would use statistics to understand and make some sense out of these fluctuating quantities. So, to summarize the bottom line is the following many of the randomness arise from the Brownian motion of particles; proteins, starting from the water molecule itself is doing a zigzag motions Brownian motion. And from based on that the protein molecules will do zigzag motion, which is called the Brownian motion.

And thus will lead to many other phenomena and the molecular phenomena could lead to many other phenomena in the much larger scale. And there is lot variability in anything that we would measure in a real life; around us many things that we would measure around us would have a lot of variability. How to deal with this variability is what is understood; what will be learnt by learning statistics; that is what we will discuss in the next lecture. Bye.