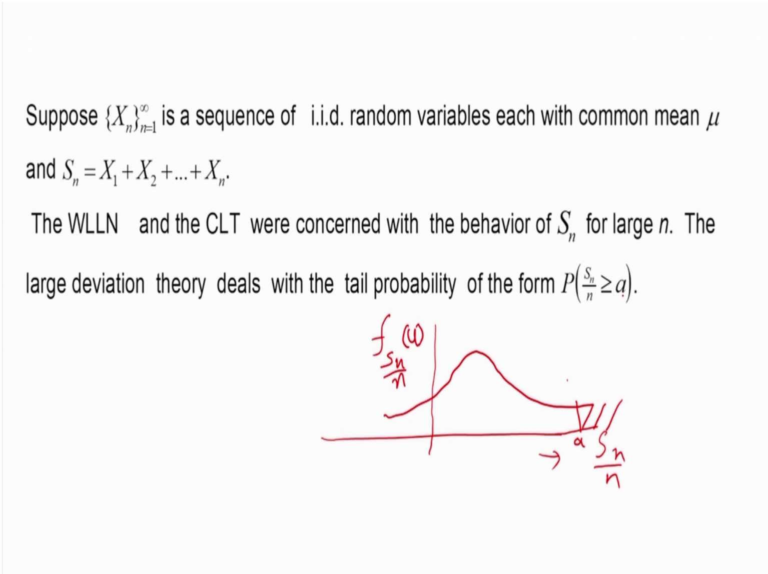
Advanced Topics in Probability and Random Processes
Prof. P. K. Bora
Department of Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture - 12
Large Deviation Theory

Large Deviation Theory, we discussed earlier about laws of large number then central limit theorem.

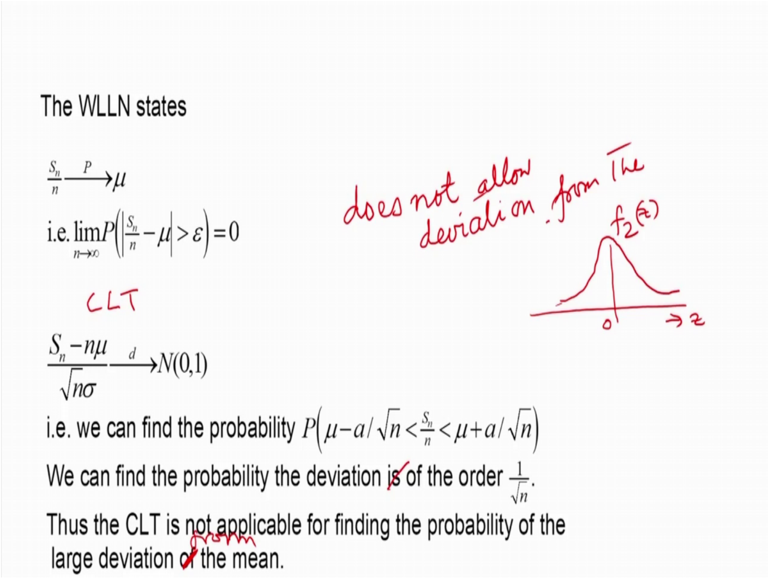
(Refer Slide Time: 00:32)



So, we will consider the same sequence of random variable again suppose X n is a sequence of i.i.d random variables is with common mean mu, and S n is summation of Xi S as defined earlier. Now, we know laws of large number and central limit theorem, they are concerned with the behaviour of S n for large n. This large deviation theory deals with the tail probability of the form probability S n by n greater than equal to a.

Suppose S n by n if I consider suppose, if it is a continuous case suppose this is some S n by n. And this is suppose f of S n by n the pdf we are considering the pdf may be this is a sum distribution like this. So, we are concerned with the tail probability suppose, this is a point a so, what is the probability that S n by n is a from a. So, S n by probability that S n by n is greater than equal to a; so, such tail probabilities are known as tail probabilities and they are very important for example, to determine the error or probability of some real event etcetera.

(Refer Slide Time: 02:13)



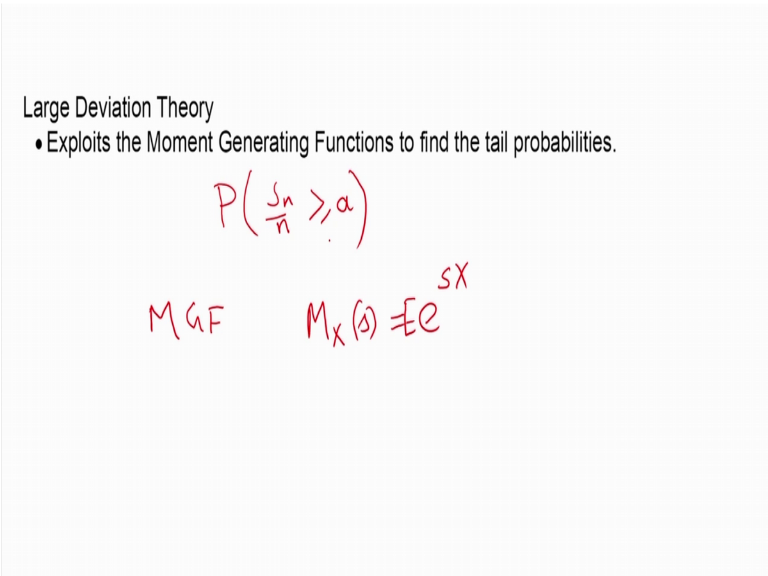
Now, let us go back to laws of large number and CLT how they are they can deal with deviation from the mean. As the name large deviation implies we have to deal with large deviation from the mean. Suppose S n by n that is the sample mean and according to weak law of large number, it converges in probability to mu, what does it mean that; the probability that S n by n, will deviate from mu by very small amount epsilon that probability is 0, as n tends to infinity that probability will become 0. Therefore, weak law of large number does not allow any deviation from the mean, does not allow deviation from the mean. So, any tail probability will be automatically 0 here.

Now, let us consider CLT, according to central limit theorem this is CLT; S n minus n mu that is the deviation divided by with n sigma, that is standardization in distribution it converges to normal 0 and normal distribution with mean 0 and variance 1. So, because the because of this now we can determine the probabilities of this form only. Because, it is normal distribution I know that normal distribution is a suppose if I it is a 0 mean normal distribution. So, this distribution standard normal distribution this is suppose z f z of z.

Now, it is a shorten distribution as normal distribution is and therefore, we can find out the probability of this form only, mu minus a by root n mu plus a by root n. So, that way because as n goes to infinity this term will becomes smaller n smaller therefore, it cannot deal with we can find probability deviation of the order of 1 by n ok. So, that way we can find out probability of this form and this deviation is very small, it cannot deal with if there are large deviation S n by n take some large deviation. This type of probability cannot deal with that is the limitation of central limit theorem.

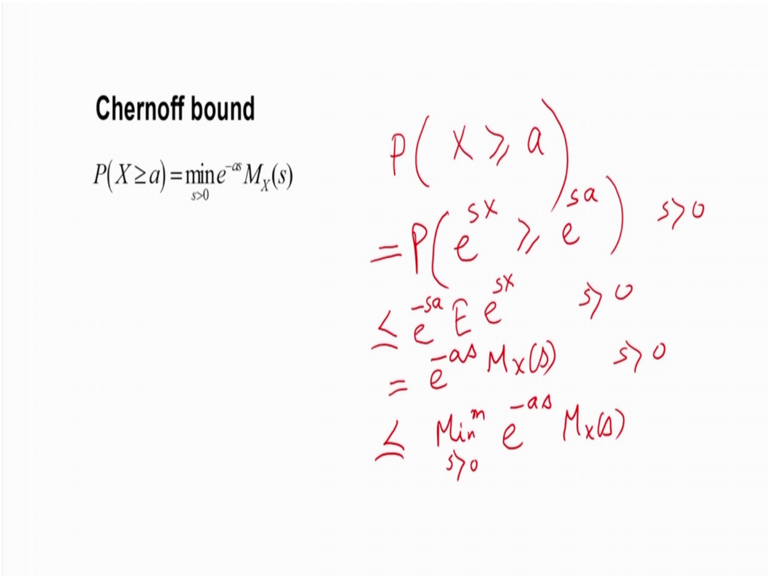
Thus, here it is not applicable for finding the probability of the large deviation of the mean from the mean. So, we see that weak law of large numbers does not allow any deviation, but central limit theorem allows deviation, but not very large deviation, the deviations are known as normal deviations.

(Refer Slide Time: 05:55)



So, now large deviation theory can deal with the probability of large deviation so, it basically exploits the moment generating functions to find out the these probabilities. That is, we want to find out suppose find out the probability that S n by n is greater than equal to a. So, we want to find out tail probability of this form and it exploits, large deviation theory exploits the moment generating function MGF, MGF for a random variable X. We know this is M X of s is equal to E of e, expected value of e to the power S X. So, this MGF is used to derive these probabilities ok.

(Refer Slide Time: 06:52)



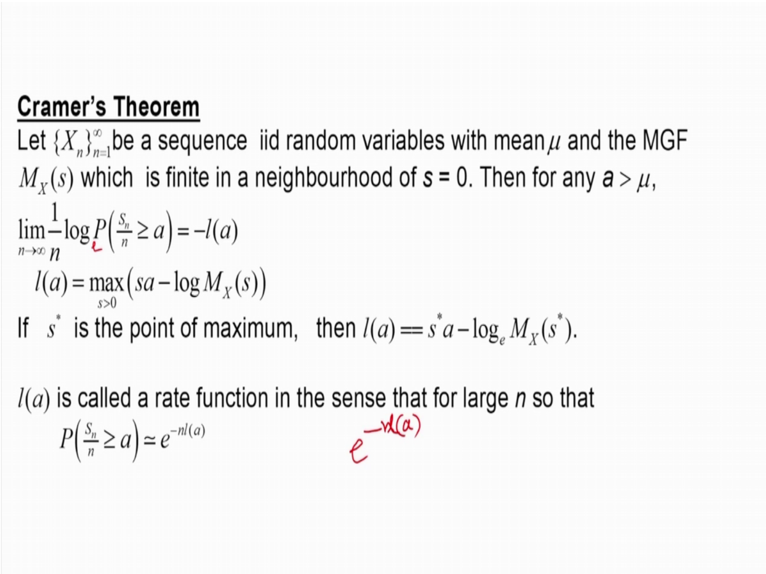
Let us see first we will go to Chernoff bound one very important result Chernoff bound that is probability that X is greater than equal to a. Suppose X is a random variable how to get the probability that X greater than equal to a. So, we have various inequalities like Chebyshev inequality, Markov inequality etcetera. Here also we can apply the Markov inequality to find out this Chernoff bound. You can consider like this; suppose we have to find out Chernoff probability that X is greater than equal to a.

Now, we will take the exponential of both sides. So, this is same as probability that e to the power S X is greater than equal to e to the power S a. If we have we have first taking the exponential with respect to S, S is a positive, S greater than 0. So, this we can write from here because the X is greater than equal to a. Now we are taking the exponential with respect to positive number. So, e to the power S X will be greater than e to the power S a.

Now, apply the Markov inequality so this will be less than equal to E of this is the positive quantity e to the power S X divided by e to the power S a that we can write as e to the power minus S a. So, this probability will be a less than equal to this is for S greater than 0 is n and this quantity is nothing but. So, this I can write as e to the power minus as into M X of s, but this is this quantity is less than equal to this and where s is greater than 0. So, if it is less than equal to this quantity for all s greater than 0, out of that there will be 1 minimum, suppose for all possible we consider the 1, which is the minimum of this expression then we will have a probability of s greater than a will be less than that minimum also.

Therefore, ultimately we can write that this is less than equal to minimum s greater than 0 e to the power minus as into M X of s. So, this is the Chernoff bound, so it is one of the tightest bounds. So, for probability of x being greater than equal to a is bounded by this. So, lower bound is this is the lower bound, sorry this is the upper bound, so this probability cannot achieve that. So, we will be in this Chernoff bound to find out the probability tail probability for S n by n. We will state one important theorem that is known as the Cramers theorem, that theorem the most fundamental results of large deviation theory that we will be discussing now.

(Refer Slide Time: 10:53)



We will consider the sequence of i.i.d random variable independent and identically distributed random variable with constant mean mu. And the MGF M X of s, which is finite in the neighbourhood of S is equal to 0. So, that is one important assumption that in the neighbourhood of s is equal to 0 M X of s is finite. So, we are considering a particular sequence of i.i.d random variables.

Then for a greater than mu, this a is greater than mu that is the mean of Xi mu. So, we are considering a number a greater than mu. So, for that the probability log of probability, this is log is base with respect to base e. Log of probability of S n by a so we are considering this limit, limit n tends to infinity 1 by n log of P of S n by n greater than equal to a. So, this term we are considering this is log of probability then we are dividing by n that is the limit is equal to minus l of a.

So, what is l of a? Now, this l of a is related with this moment generating function. So, this is equal to l of a is equal to maximum of s greater than equal to 0, s a minus log of M X of s. So and if s star is the point of maximum suppose l a is the expression it is a function of a, for which this s a minus log of M X of s attains the maximum value. Suppose s star is the point of maximum then l a we can write it is s star a because s is s star, now log of minus log of M X of s star. So, this is the rate function this function is called rate function.

Why it is called rate function; because it will determine the probability. Now it is a probability rate function how it can determine the probability, suppose this expression is there if I multiply by n then it will be l a into n. Therefore, and if we log is given so if we take the inverse log then probability of S n by n greater than equal to a will be approximately equal to e to the power minus m into rate function. So, this is the tail probability now. We can determine because this is a as n tends to infinity therefore, for large n we can approximate this probability by this result, that is probability that S n by n is greater than equal to a. That is approximately equal to e to the power minus there is a rate function l a is there that is the rate function multiplied by n so this is the probability.

What is l a now; l a is related with the log of the moment generating function, log of MGF this is also known as Cumulative Cumulant generating function. So, the difference between s a and log of M X of s. So, this is a line this is another function, so what is the maximum difference between them. So, that is l a and it will determine the tail probability this is what Cramer's theorem says.

(Refer Slide Time: 14:58)

Proof:

We can find an upper bound for $P\left(\left\{\frac{S_n}{n} > a\right\}\right)$ by applying Chernoff bound,

$$P\left(\left\{\frac{S_n}{n} > a\right\}\right) = P(S_n > na) \le \min_{s>0} e^{-ans} M_{S_n}(s) \qquad P(X > a) \le \min_{s>0} e^{-as} M_X(s)$$

$$M_{S_n}(s) = Ee^{sS_n} = E e^{s\sum_{i=1}^{n} X_i} = E\prod_{i=1}^{n} e^{sX_i} = \prod_{i=1}^{n} E e^{sX_i}$$

$$= \left(M_X(s)\right)^n$$

$$\therefore P\left(\frac{S_n}{n} > a\right) \le \min_{s>0} e^{-ans}\left(M_X(s)\right)^n$$

$$= \min_{s>0} e^{-ans + n\log_e M_X(s)} = e^{-nl(a)}$$

Now, this probability that S n by n is greater than equal to a Cramer's theorem says; what is the value of the probability as n as n tends to infinity. So, first of all we will find out, what is the upper bound for this probability. So, for this we can apply the Chernoff bound. So, probability of S n by n greater than a this is same as probability that S n is greater than n a.

So, if I apply the Chernoff bound, what we can write? That is less than equal to minimum over s greater than 0 of e to the power a and s because here now right hand side is n a. So, we are writing n a here into n this is the random variable M S n of s. This is because we know that probability of X greater than equal to a. Here we are writing greater than does not matter then here it will be less than equal to minimum over s greater than 0 of what e to the power a s into M X of s, that was the Chernoff bound that we have used here. So, here an concerned random variable is S n and then it is greater than n a.

Now since X i's are i.i.d so M S n of s received by definition E of e to the power s into S n. So, that is equal to now we can write this as equal to E of e to the power s into summation X i is equal to 1 to n. Now this we can write as a because it is a exponential. So, that way we can write E of product of i is equal to 1 to n, e to the power s Xi. Now we know that X i's are independent random variables therefore, e of s X i's that exponential of this independent random variable they will be also independent. So, because of that this e I can take inside so this will be for that term i is equal to 1 to n E of e to the power s Xi. So, this is nothing but normal generating function M of x i. So, this is nothing but M X i of s.

Since, i.i.d each of the random variable will have this a moment generating function therefore, this I can write as M S n of s is equal to moment generating function of X to the power n. So, this is the moment generating function of S n of the sum, when X i's are independent an identically distributed. Therefore, how to we can write now probability of S n by n greater than a, that is less than equal to minimum over s greater than 0, e to the power a n s because S n will be greater than equal to n a into. Now moment generating function of S n, S n is equal to M X of s to the power n so we can write like this.

So, now we have to find out the minimum of this quantity. If I write this also in terms of exponential then I can write as a n s plus n log M X of s. So, this minimum suppose this is now minimum of this expression will be obtained when this expression is maximum ok. So, if I consider suppose minimum of the entire exponential expression. So, the exponent whenever it is maximum that expression will be also maximum. So, that way we have to maximize the exponent part, so if that is the maximum the maximum is n into l a, then we can write that this probability that S n by n is greater than equal greater than a is bounded by it is less than equal to e to the power minus n into l a, what is l a here? Where l a is maximum s greater than 0 s a minus log of M X of s.

(Refer Slide Time: 20:08)

$$l(a) = \max_{s>0}\left(sa - \log_e M_X(s)\right)$$

$$= s^* a - \log_e M_X(s^*)$$

s^* is the value corresponding to the maximum and given by

$$\frac{d}{ds}\left(sa - \log M_X(s)\right)\Big]_{s=s^*} = 0.$$

Equivalently, $\frac{M_X(s^*)}{M_X(s^*)} = a$

Thus,

$$P\left(\frac{S_n}{n} > a\right) \le e^{-nl(a)}$$

$$\therefore \frac{1}{n}\log_e P\left(\frac{S_n}{n} > a\right) \le -l(a)$$

$P\left(\frac{S_n}{n} > a\right) \le e^{-n l(a)}$

$\frac{1}{n}\log_e P\left(\frac{S_n}{n} > a\right) \le -l(a)$

We have to show that $\lim_{n\to\infty} \frac{1}{n}\log_e P\left(\frac{S_n}{n} > a\right) \ge -(l(a)+\varepsilon)$, $\varepsilon > 0$

So, if s star is the maximum point then we can write it l a is equal to s star a minus log of M X of s star. Therefore, we have sum that that probability that S n by n is greater than a this is bounded by, e to the power minus n into l of a where l a is given by this. So, this is a consequence of Chernoff inequality only.

Now, how do I find s star? S star is the point where this expression s a minus log of M X of s is maximum therefore, if I take the derivative with respect to s that should be at s is equal to s star that should be is equal to 0. So, if I take the derivative which was (Refer Time: 21:21) to s this will become a, this will become 1 by M X of s into M X dash of s. So, from that equating to 0 we will get a equal to M X dash of s star divided by M X of s star.

So, now this l is related to the MGF by this relationship, a is equal to M X of s star divided M X dash of s star differentiation of M X of s, s star divided by value of M X of s of s star. So, therefore, what we get then probability that S n by n greater than a is less than equal to e to the power minus n into l a. Or if I take the logarithm so I take the logarithm first then it will be minus n into l a will come. So, that n if I keep at the denominator left hand side we will get this 1 by n log of probability S n by n greater than a, that will be always bounded by minus l a, where l a is the rate function given by this relationship ok.

So, that way we have a avail to find out an upper bound for this probability, tail probability upper bound is given by this relationship, but we have found out the upper bound. So, that is 1 by n log of base e, probability that S n by n is greater than a that is bounded by. So, we want to show that for large this is the upper bound will be we have to find the lower bound also for this expression so that we can find out the limit test. So, we have to show that for large m limit of limit as n tends to infinity 1 by n log of probability that S n by n is greater than a. So, this will be greater than some lower bound that is again given by l a plus epsilon here epsilon is arbitrarily small.

So, this result we have to establish this is a slightly complicated task. So, far we know that this with probability the logarithm of the tail probability is related to be rate function, but we have a upper bound for this expression which is bounded by the rate function.

Next we will show that this quantity as n tends to infinity it will be greater than minus l a plus epsilon. So, that in the limiting case we will get both lower bound and upper bound, if we consider then this will be equal to minus of l a that will establish.