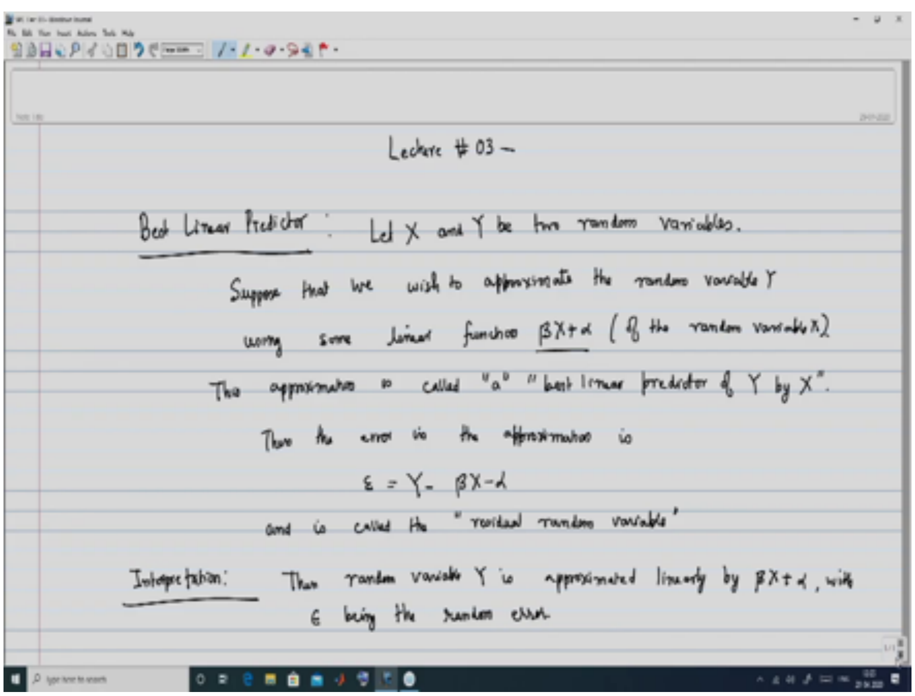
Mathematical Portfolio Theory

Prof. Siddhartha Pratim Chakrabarty
Department of Mathematics
Indian Institute of Technology Guwahati

Module 01: Basics of Probability Theory
Lecture 03: Linear regression, Binomial and normal distribution, Central limit theorem

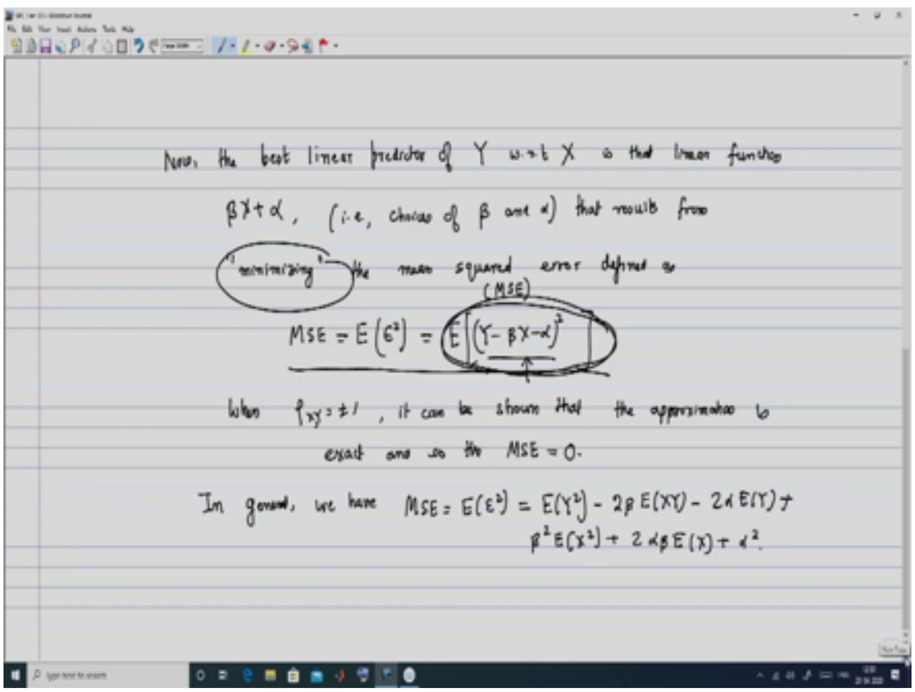
Hello viewers. Welcome to this third lecture for the course on Mathematical Portfolio Theory. You would recall that in the previous two lectures, we discussed in detail about probability space and in both discrete and in continuous time. And we looked at the definition, the properties and then in particularly in the last lecture we talked about the first two moments, namely, mean and variance, and then you talked about covariance, and correlation coefficient. So, in todays class, we will extend those concept to talk a little bit about what is known as the best linear predictor. The reason why we need to look at the best linear predictor is this will eventually be made use of in modern portfolio theory where we will basically talk about something called the single index model. This will be followed by a discussion on two distributions, one in discrete and one in continuous time which will be used extensively when you talk about asset pricing model in discrete and continuous time respectively. And we will look at some of the properties of those distribution.

(Refer Slide Time: 01:47)



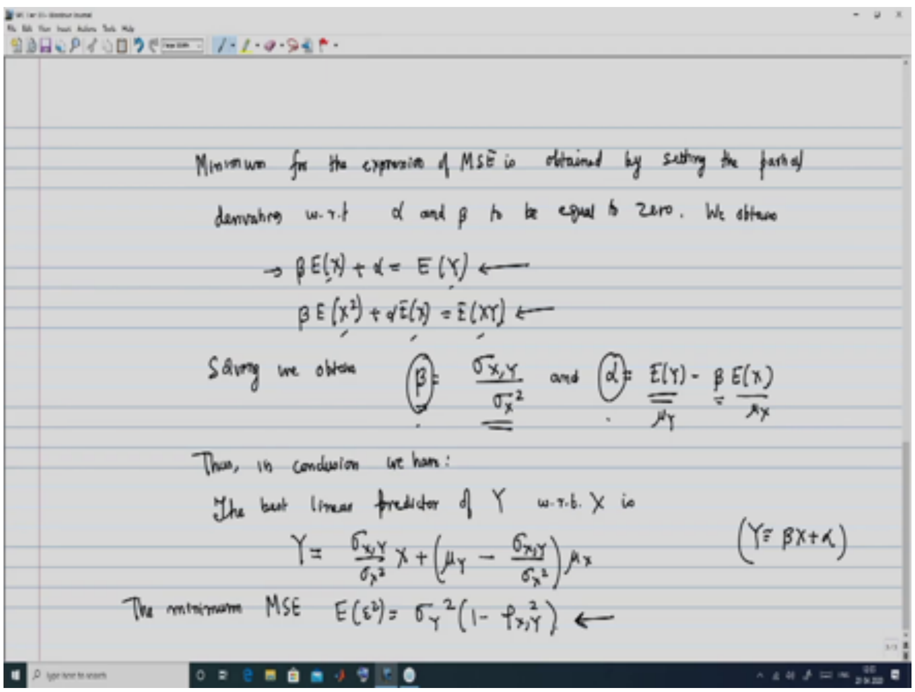
So, we start this lecture with the best linear predictor. So, this best linear predictor as the name itself suggest is about using some linear function to predict as a predictive tool. So, accordingly, I begin with, let X and Y be two random variables. Suppose that we wish to approximate the random variable Y using some linear function $\beta X + \alpha$ (of the random variable X). Now, this approximation and that means, this $\beta X + \alpha$ is called "a". So, this indicates that it is not necessarily the unique best linear predictor of Y by X. So, basically it is the best linear prediction of Y by X. So, once we have this random variable Y and which you want to approximate by the linear relation $\beta X + \alpha$, so clearly the set of random variables that are generated by this approximation $\beta X + \alpha$, where β and α have to be determined based on the information that we have about X and Y. And the prediction of $\beta X + \alpha$ results in a value of Y or other random variables are taking the values for Y, then between the actual random variable Y and the ones that are predicted by this linear approximation or the linear predictor $\beta X + \alpha$ there is going to be some difference. So, the next thing that we are going to look at is going to look at is the difference between these two or in particular what is going to be the error that happens in this prediction. So, accordingly let us define the error in terms of a variable epsilon. So, then the error in the approximation is the difference between Y and the predicted value of $\beta X + \alpha$. And I will denote this by ϵ, and this ϵ is called the residual random variable. So, let us just look at little bit about the interpretation of this of whatever I have discussed so far. So, the brief interpretation of this is that the random variable Y is approximately estimated or approximated linearly by $\beta X + \alpha$ with ϵ being the random error.

(Refer Slide Time: 06:18)



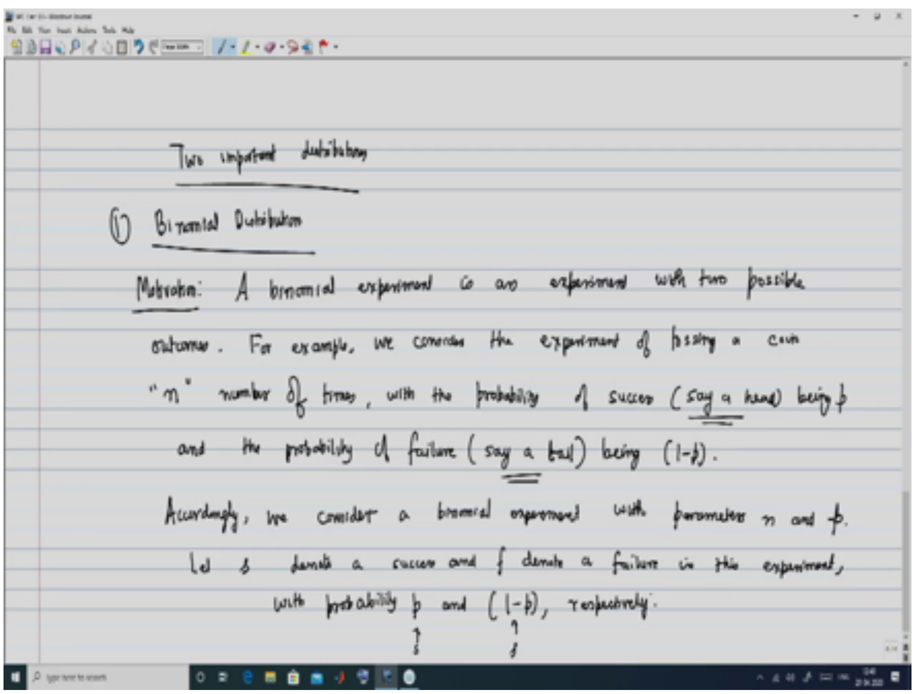
So, now the next thing that we look at is the best linear predictor. So, here basically the best linear predictor essentially means some sort of an optimized choice of alpha and beta. So, the best linear predictor of the random variable Y with respect to the random variable X is that linear function $\beta X + \alpha$. So, that linear function $\beta X + \alpha$ indicates the particular linear predictor with particular values of α and β. So, I will indicate this that is choices of β and α that results from minimizing the mean squared error, defined as, so, remember the error was ϵ which were the same as $Y - \beta X - \alpha$. So, we take the square of that. And as since these are this is a random variable, so we will essentially take the expectation of this and that is going to be the expectation of epsilon square and this is what we call as the mean squared error. So, mean squared error, I will just abbreviate this as MSE, okay. So, this two is not here. So, what I am doing, basically I am trying to what I am looking at here is I am looking at epsilon which is the error and I am squaring it. So, that is basically going to capture and penalize the larger deviations as far as from the actual value as compared to the linear predictor. And since, this is a random variable, obviously, we have to calculate is expectation, and then what I want to do is that we basically want this error that is a good linear predictor should be such that the difference between the predictor and the actual value Y should be as small as possible which is why we take this mean squared error and then we need to minimize it. So, that we basically get as possible closely or as best as possible the values of beta and alpha, okay. So, now, I want to begin with the observation that when a $\rho_{XY} = \pm 1$, you can show that, shown that the approximation is exact and the reason for this, and so, the MSE if it is an exact approximation, so obviously, the error is going to be equal to 0. But however, we need to look at the general case. So, accordingly, I will look at in general we have something else. So, in general, we have, what we are going to do is that, we are going to look at this MSE the expression of MSE and expand the right hand side of this that means, this term. So, $MSE = E(\epsilon^2) = E(Y^2)$. So, I am basically doing the squaring of the term inside and using the linearity property and scaling property of expectation. So, I get $E(Y^2) - 2\beta E(XY) - 2\alpha E(Y) + \beta^2 E(X^2) + 2\alpha\beta E(X) + \alpha^2$. And remember that our goal is basically to minimize the MSE.

(Refer Slide Time: 10:42)



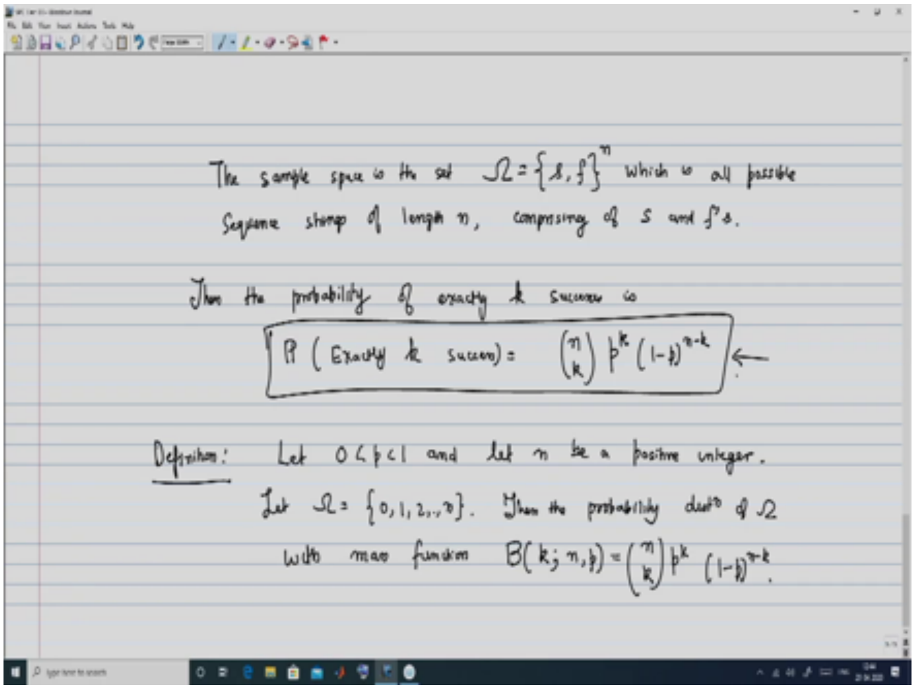
So, accordingly, we make the observation that the minimum for the expression of the mean squared error is obtained and remember that here the of the terms that I want to obtain are beta and alpha. So, it is obtained by setting the partial derivatives with respect to, since this optimization is done with respect to alpha and beta. So, with respect to alpha and beta to be equal to zero. So, once you essentially take the derivative of this you know term here with respect to α and β and set it equal to zero. So, we will basically get two equations. So, one of the equations will be, so we obtain accordingly we get $\beta E(X) + \alpha = E(Y)$. So, $\beta E(X) + \alpha = E(Y)$ and the second relation that you will get after differentiate with respect to β, this is going to be $\beta E(X^2) + \alpha E(X) = E(XY)$. So, what we get is basically we will get, we here we have one relation a linear equation in α and β being the unknown because we already know what is $E(X)$ and $E(Y)$. And here again, we have α and β unknowns because we know what is $E(X^2)$, $E(X)$ and $E(XY)$. So, these are known quantities. So, accordingly, what we do is that we will by solving we obtain. What do we obtain? We obtain that $\beta = \frac{\sigma_{XY}}{\sigma_X^2}$ and consequently from this from the first relation you can obtain that alpha from this relation here we get $\alpha = E(Y) - \beta E(X)$, where of course, you know I have already obtained what the β is going to be. Thus in conclusion of all this exercise we have the following. And we can make the following statement that the best linear predictor of Y with respect to X is given by Y is equal to remember our assumption was that Y will be approximated by $\beta X + \alpha$. So, what is β? I have calculate my $\beta = \frac{\sigma_{XY}}{\sigma_X^2} X + \alpha$. So, what is α? α is $E(Y)$. So, I will denote this by μ_Y and I will denote this by μ_X for consistency of notation. So, this is going to be $\mu_Y - \beta$ which is $\frac{\sigma_{XY}}{\sigma_X^2}\mu_X$. And so, if you use this particular values of α and β. So, if I use the value of α and β and we substitute this in our mean squared error, so in that case the minimum, that is, it is a mean square error for this particular α and β. So, the minimum mean square error is going to be $E(\epsilon^2)$, this is turns out to be after some calculation this turns out to be $\sigma_Y^2(1-\rho_{XY}^2)$. So, here you see you know it brings us back to the original statement. So, again this can be reconciled to our previous statement that when $\rho_{XY}=1$, the MSE is going to be equal to 0. So, this is consistent with this observation that we have here, okay. So, now, we next consider our two distributions, one in discrete and one in continuous time. So, for the discrete time, we will consider the binomial distribution and for the continuous time, we will consider the normal distribution. And the rationale for doing this in the context of this course is that, for the asset pricing models in discrete time, we will basically make use of something called a binomial model and for the continuous time we will essentially use something called the Black Scholes model framework, where the asset price will be modelled using a winner process driven mechanism, where essentially the winner process is some sort of is very closely related to and it satisfies conditions that are similar to the normal distribution. So, that part we will discuss in the next module. So, for this part we will talk about that some of the properties of the distribution and more we put an emphasis on the definition of the distribution. And in particular, we will just make a note of the first two moments of the distribution, namely, the mean and the variance.

(Refer Slide Time: 16:28)



So, we now consider two important distributions. So, the first one as I said is going to be the binomial distribution. So, for the binomial distribution, I will just start off with a motivation and I will start off with a very simple set up. So, a binomial experiment, so the motivation is going to be a binomial experiment is an experiment with two possible outcomes. And let me explain this in detail. So, for example, we consider a particular such experiment of a tossing a coin. So, we consider the experiment of tossing a coin, n number of times with the probability of success, say, a head being p and naturally the probability of failure, since this is a two outcome situation and say this a head a tail being naturally $1-p$. Accordingly, we consider binomial experiment with parameters and p. So, a binomial experiment with parameters n and p basically means that you are repeating an experiment n number of times and for each time the probability of a success is going to be p and the probability of failure is going to be $1-p$. And a particular example of this is the coin tossing problem where the coin is being tossed n number of times. Also note that here we have talking that the probability of success, say, a head and the probability of failure is a tail. So, it is sort of not very rigid, it is just only for illustrative purposes that you are identifying the head to be a success and the tail to be a failure. You can actually also choose the other way around. So, there is no loss of generality in this particular observation, alright. So, now, we are talking about success and failure. So, accordingly this motivates us to use the alphabet s to denote a success and f denote a failure in this binomial experiment, right. So, s could have been a head and f could have been a tail in the context of the specific example we just looked at. And in the experiment with the probability p for f and $1-p$ for s and $1-p$ for f. So, with the probabilities p and $1-p$, respectively, alright.

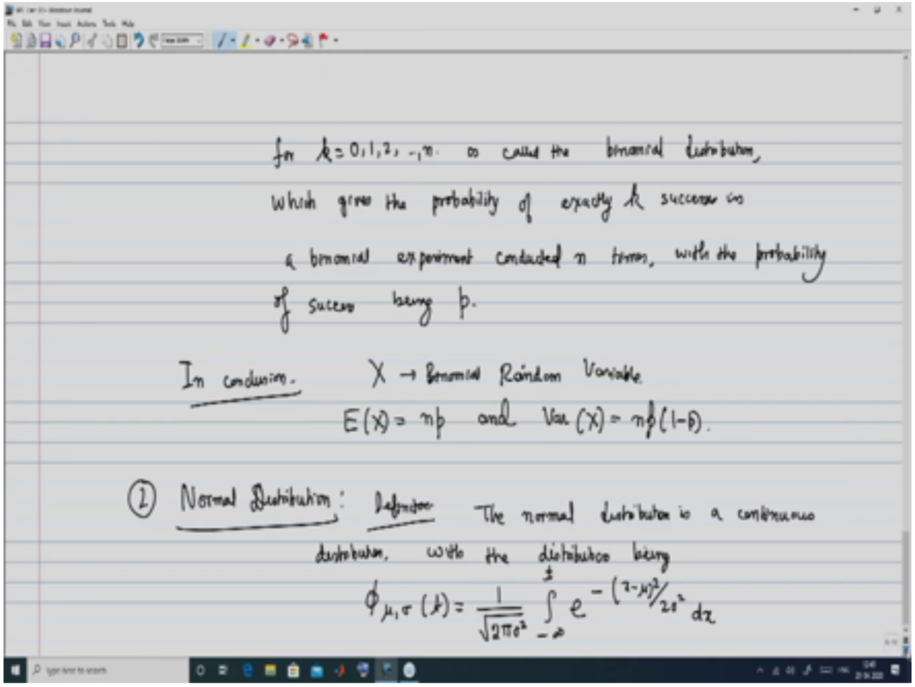
(Refer Slide Time: 20:47)



So, then we need to set the sample space. So, accordingly, the sample space is the set Ω and the sample space we will denote this by the notation $\{s, f\}^n$ and this notation means all possible sequence strings of length n comprising of s and f's. So, this means that this sample space omega is nothing, but a sequence of s and f, where the total number of such characters is going to be little n, okay. So, then the probability, so, the next thing we naturally look at is that if we are doing n number of experiments and we are curious to find out that what is going to be the possibility of k number of success. So, if you look at it in the context of the coin tossing problem, you are tossing the coin n number of times and you want to basically figure out what is going to be the probability that out of those n number of tosses you obtain k number of heads which he has been considered as the success. So, accordingly we will now define this probability. So, then the probability of exactly k successes is, I will just write this down elaborately, exactly k successes. What is this going to be? So, it is going to be p^k for k number of successes. So, this means that there has been $n-k$ number of failures with probability $1-p$ and these successes k number of success can happen in $\binom{n}{k}$ ways. So, this is going to be my probability of exactly k successes. So, now, I am in a position, once the motivation is done I am now in a position to start off with the definition. So, let $0 < p < 1$ and let n be a positive integer, then let $\Omega = \{0, 1, 2, \cdots, n\}$, then the probability distribution of Ω with mass function, remember we are using the term mass function because it is a discrete scenario will be given by $B(k; n, p)$, so that means, n number of trials with probability of success been p, this is going to be the same as $\binom{n}{k} p^k (1-p)^{n-k}$.

(Refer Slide Time: 24:20)

And this will hold for $k = 0, 1, 2, \cdots, n$. Remember, that the number of successes you could either have 0 success or 1 success, 2 success or a maximum possible n number of success which is the maximum possible number of experiments that you can actually do in this setup, okay. So, this is called the binomial distribution which the interpretation of this is that which gives the probability of exactly k successes in a binomial experiment conducted n times with the probability of success being p. So, in conclusion, if X



is a binomial random variable, so the $E(X) = np$ and $Var(X) = np(1-p)$, okay. So, now we are in a position to move to our second distribution which is the normal distribution. So, in this case, unlike the binomial distribution, in the case of a normal distribution we will not really motivate much except the point out that a lot of real life examples you can observe that there is a normal distribution that is being exhibited. So, for example, if you look at the distribution of marks secured by students in a class and you break them up into 10 intervals 0 to 10, 10 to 20, 20 to 30 and so on. And you look at the frequency for each of those intervals and you plot a histogram of that what you obtain is essentially it is a bell shaped curve. So, if we join the histograms by a smooth curve it turns out to be a bell shaped curve and this bell shaped curve is synonymous with what is known as the normal distribution. So, that is, one simple example where you actually see normal distribution real life. I will begin with the definition, right away. So, in the case of normal distribution, let me give the definition. So, the normal distribution is a continuous distribution with the distribution being; so, this is the cumulative distribution is this being. So, we have

$$\Phi_{\mu,\sigma}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

(Refer Slide Time: 28:39)

So, from here we can conclude, this means that is the density function or the probability density function for the normal distribution is; so, we will use this capital N to denote the density function. So,

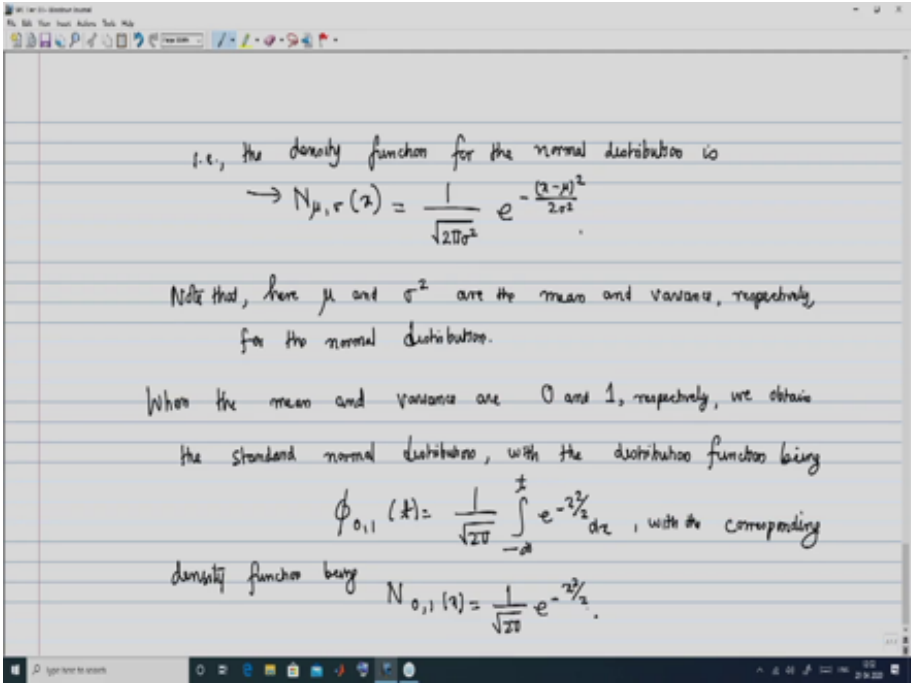
$$N_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

. Note that, here μ and σ^2 are the mean and variance respectively for the normal distribution. So, next we consider a particular case of this. So, when the mean and variance are 0 and 1, respectively, we obtain what is known as the standard normal distribution. So, naturally here it follows immediately from the general definition of the normal distribution. So, with the distribution function being

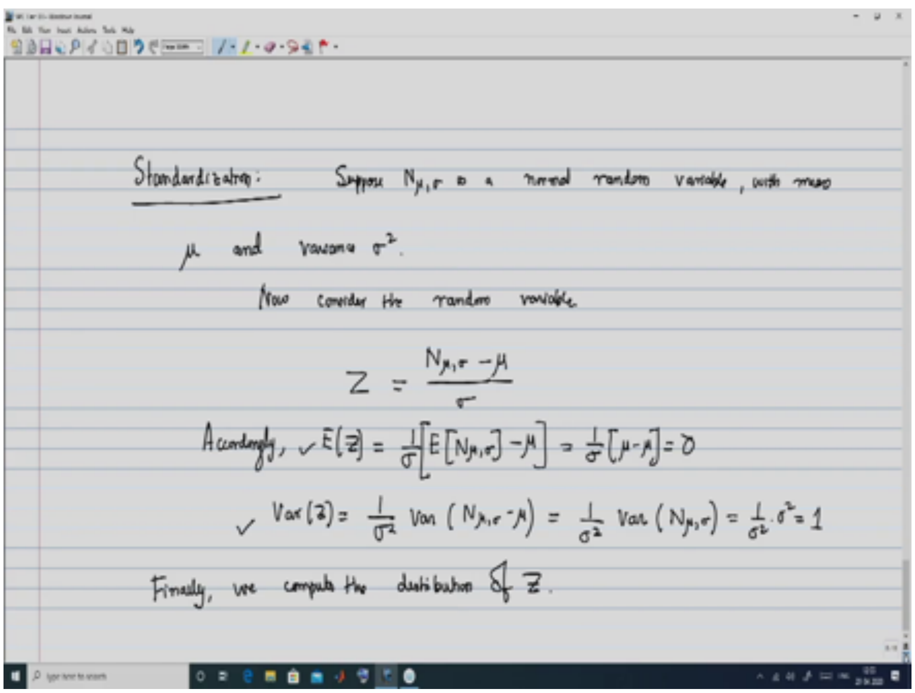
$$\Phi_{0,1}(t) = \frac{1}{\sqrt{2\sigma}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx.$$

And naturally with the corresponding density function being; so, we will use the notations similar to this. So, this is going to be

$$N_{0,1}(x) = \frac{1}{\sqrt{2\sigma}} e^{-\frac{x^2}{2}}.$$



(Refer Slide Time: 31:46)



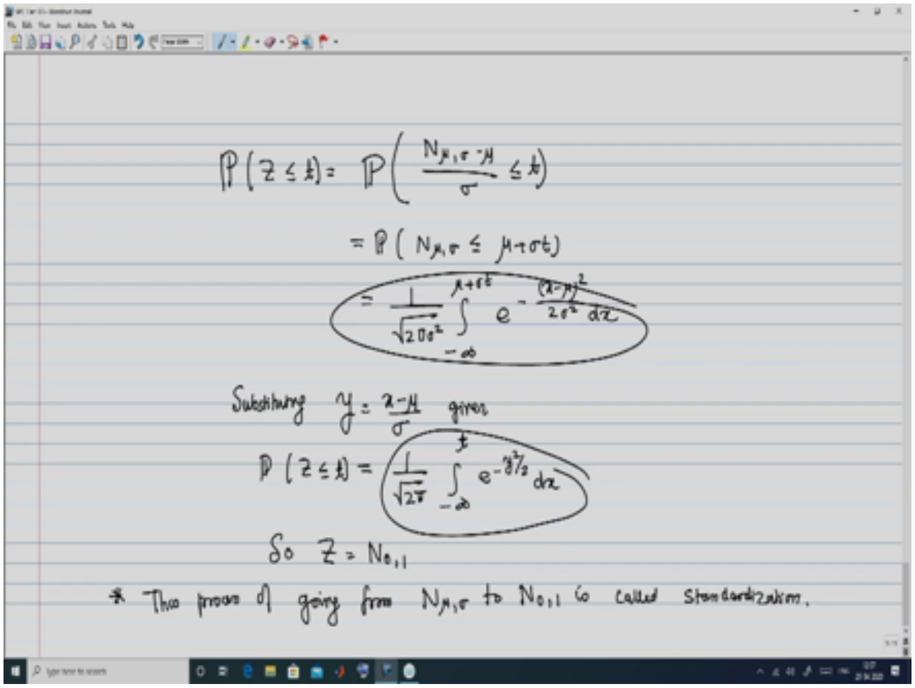
So, we now move on to an important fall out of the normal distribution which is known as standardization. So, let me motivate this in the following way. Suppose, $N_{\mu,\sigma}$ is a normal random variable obviously, with mean μ and variance σ^2. Now, we construct a new random variable. So, we consider the random variable. So, we take the random variable $\frac{N_{\mu,\sigma}-\mu}{\sigma}$ and we call this random variable as Z. Now, we look at the mean and variance of this Z. So, accordingly, expected value of Z, what is this going to be? This is going to be if I use the scaling property it is going to be

$$E(Z) = \frac{1}{\sigma}\left[E(N_{\mu,\sigma}) - \mu\right] = \frac{1}{\sigma}[\mu - \mu] = 0,$$

$$Var(Z) = \frac{1}{\sigma^2}Var(N_{\mu,\sigma} - \mu) = \frac{1}{\sigma^2}Var(N_{\mu,\sigma}) = \frac{1}{\sigma^2}.\sigma^2 = 1.$$

So, now, we know that the random variable Z has a mean of 0 and a variance of 1. So, the only thing that remains is to figure out making use of $N_{\mu,\sigma}$, figure out as to what exactly is the distribution of Z going to be. So, accordingly what you do, finally, we compute the distribution of Z.

(Refer Slide Time: 34:36)



So, how do you compute? The distribution will be given by probability of $Z \leq t$. What is this going to be? This is going to the probability and I replace Z with $\frac{N_{\mu,\sigma}-\mu}{\sigma} \leq t$. Now, this can be rewritten as probability of $N_{\mu,\sigma} \leq \sigma t + \mu$. Now, what is this going to be? I will make use of the distribution of $N_{\mu,\sigma}$. So, accordingly this becomes

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\mu+\sigma t} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

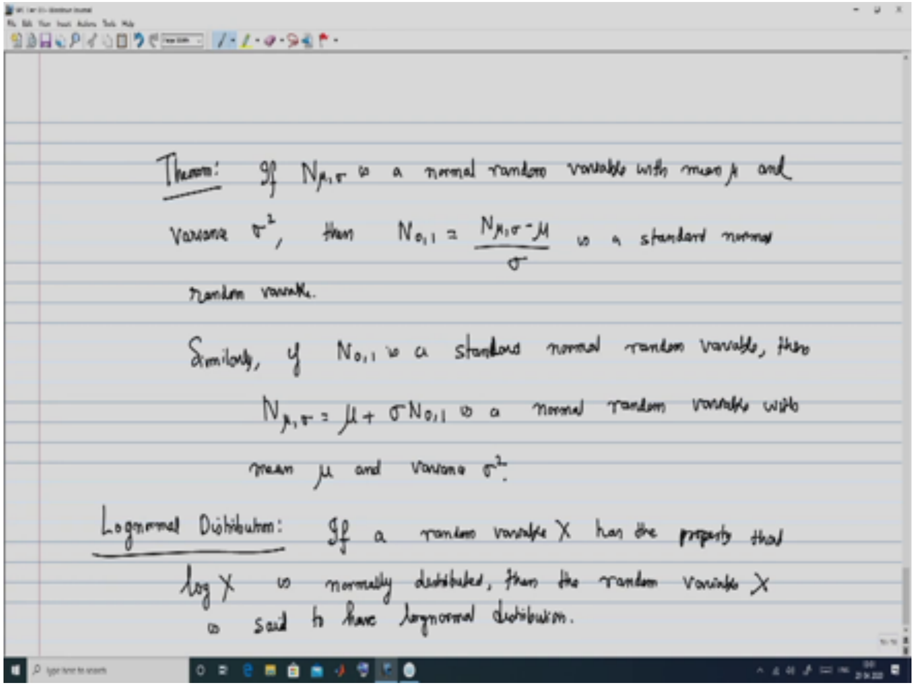
So, now what do we do? We now have this integral. So, we use the method of substitution. So, substituting $y = \frac{x-\mu}{\sigma}$, what will this give me? This gives me that

$$P(Z \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}} dy.$$

So, what is this? This is nothing, but the distribution for the standard normal variate. So, Z is nothing, but $N_{0,1}$. So, in conclusion this process of going from $N_{\mu,\sigma}$ to $N_{0,1}$ is what is called standardization.

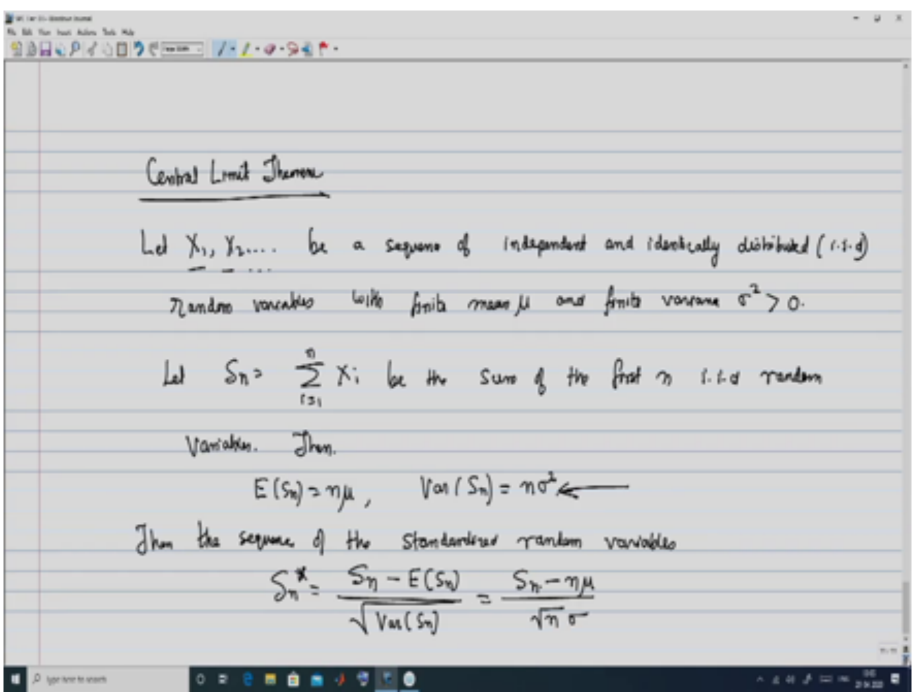
(Refer Slide Time: 36:48)

So, the preceding narrative will now be summed up as a very simple and obvious theorem, which I state as follows. So, if $N_{\mu,\sigma}$ is a normal random variable with mean μ and variance σ^2, then $N_{0,1}$ is equal to $N_{\mu,\sigma}$. So, I am just doing the standardization, is a standard normal random variable. On the other hand, in a similar manner if $N_{0,1}$ is a standard normal random variable, then $N_{\mu,\sigma}$ is equal to $\mu + \sigma N_{0,1}$ is a normal random variable obviously, with mean μ and variance σ^2, okay. So, just to wind up this discussion on normal distribution. I will just briefly mention what is the log normal distribution this is something that you are going to revisit when you talk about the asset pricing. So, I will just introduce the definition. So, if a random variable X has the property that $\log X$ and here $\log X$, it means with respect to base e is normally distributed, then the random variable X is said to have log normal distribution. So, this means that if you have a random variable X and you take the log of the random variable and those values are distributed normally, then the original random variable before it to the log is said to qualify of what is known as a



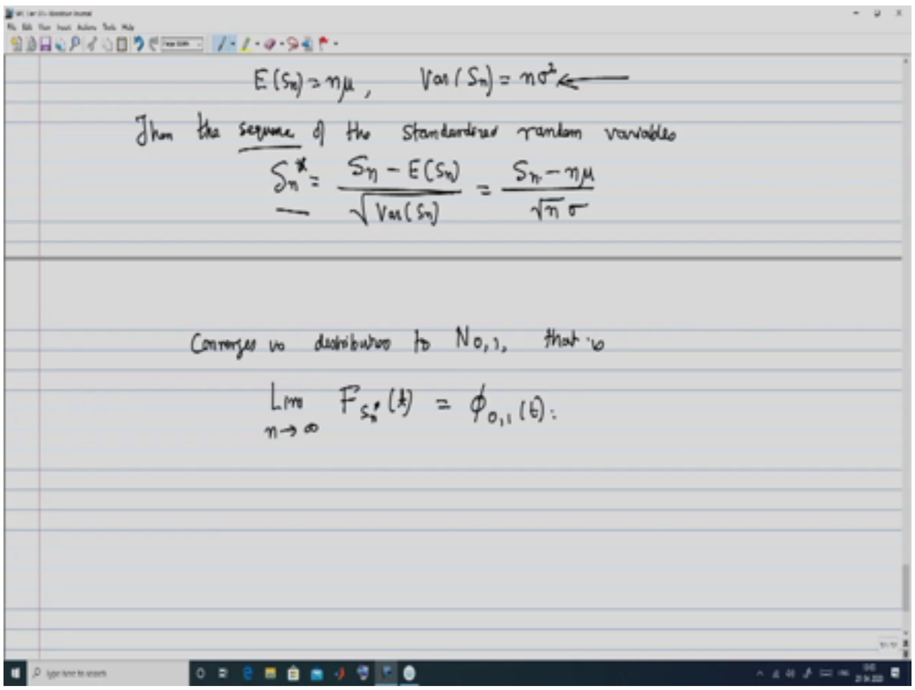
log normal distribution, okay. So, we conclude today's class with just one more topic and that is a very important result which is known as the central limit theorem. So, the central limit theorem plays a very key role in statistics where you basically look at the sequence of independent and identically distributed random variables and how its behaviour is related to a standard normal random variate.

(Refer Slide Time: 40:35)



So, accordingly we start off now with the Central Limit Theorem. The theorem states is the following, let X_1, X_2, this be a sequence of independent that means, they are independent of each other and identically distributed. This means that they follow the same distribution. And we abbreviate this as i i d. These are i i d random variables with finite mean μ and finite variance $\sigma^2 > 0$. So, this means your X_1, X_2 and so on there they are all independent of each other, first thing. The second thing that they are identically distributed and since they are identically distributed, so obviously, they will have the first two moments. So, those moment are μ, that is, the mean is μ and the variance is going to be σ^2 where as discussed before the μ has to be finite and σ^2 obviously is positive. So, now I defined S_n as the sum of the first nth this random variables. So, S_n be sum of X_i, $i = 1, \cdots, n$, this i let this be the sum of the first n i i d random variables. Then, obviously, $E(S_n)$ is going to be $n\mu$ using the additive property of expectation and using the property of variance of sum of independent random variables we get variance of S_n is $n\sigma^2$. So, the sequence, so I can say that then the sequence of the standardized random variables. So, here I basically I am using the standardized concept similar to the one you have done for normal distribution. So, I will call this as S_n^* to indicate that this is standardized. This is going to be the original random variable S_n minus the mean that is expected value of S_n divided by the standard deviation that is the square root of variance of S_n. What is this going to be? This using the observation here this is going to be $\frac{S_n - n\mu}{\sqrt{n}\sigma}$.

(Refer Slide Time: 43:57)



So, this sequence S_n^*, this converges in distribution to $N_{0,1}$ that is more explicitly this means that the limit of the distribution $F_{S_n^*}(t)$ as your n tends to ∞ this turns out to be $\Phi_{0,1}(t)$, which is the standard normal distribution. So, this brings us to the end of todays lecture. Just a brief recap of what we have done. We extended upon our observations whatever we have done as far as covariance and correlation coefficients is concerned. And of course, proceeding to that we had the properties of mean and variance, to make use of the concept of best linear predictor, and the best linear predictor serve the purpose of approximating one random variable by a linear function of another random variable. The next thing we did was that we looked at two important distributions from the context of the subsequent topics to be taught in this course, namely, the binomial and the normal distribution, and we look at couple of the properties. And then, we concluded our discussion today with a very important theorem which is the central limit theorem which looks as the sequence of sums of random variables which are independent and identically distributed and how it converges in distribution to the distribution of a standard normal random variate.

Thank you for watching.