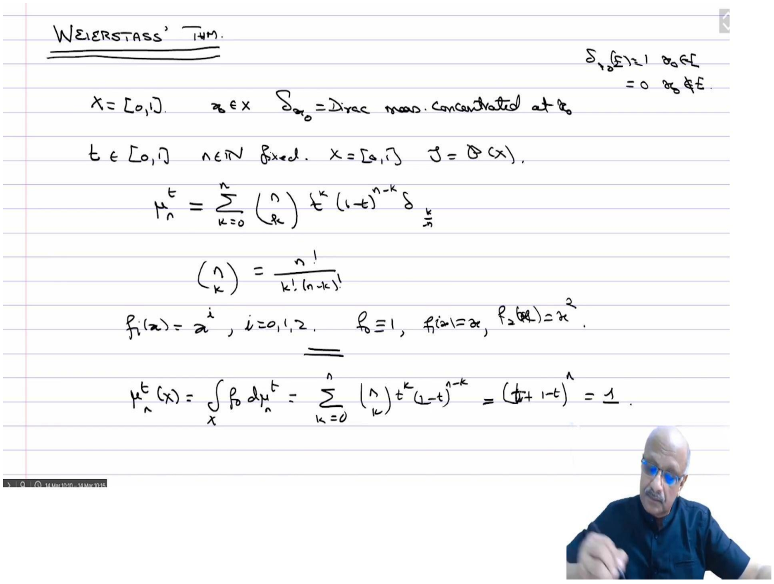
Measure and Integration

Professor S. Kesavan
Department of Mathematics
The Institute of Mathematical Sciences

Lecture No-39

So, now, we are going to see a really beautiful application of all that we have learned so far: integration and convergence and so on and so forth and through a very classical theorem from analysis which you have already seen shortly in a course on real analysis.

(Refer Slide Time: 0:17)



So, we will prove that Weierstrass's theorem: tells you that any continuous function on a bounded interval can be uniformly approximated by a sequence of polynomials. So, this is the Weierstrass approximation theorem. And so, let us try to prove this in a completely different way using the notions of integration which we have been seeing.

So, we work with the interval $X = [0, 1]$ and if $x_0 \in X$, we denote by δ_{x_0} = the Dirac measure concentrated at x_0, recall so, this means $\delta_{x_0}(E) = 1$ if $x_0 \in E$ and 0 if $x_0 \notin E$, so this is the Dirac measure and of course, sigma algebra is the entire power set every set is measurable.

So, now, we take $t \in [0, 1]$ and $n \in \mathbb{N}, X = [0, 1]$, $S = P(X)$ and then we define

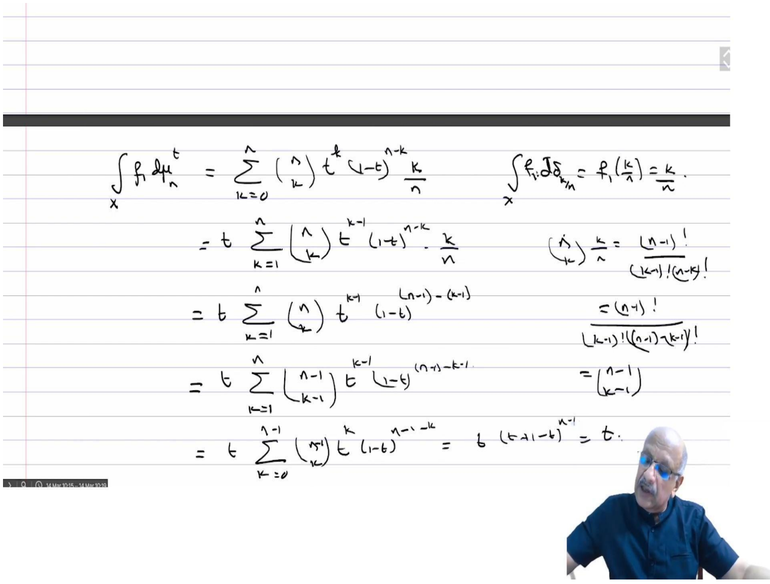
$$\mu_n^t = \sum_{k=0}^{n} {}^nC_k t^k (1 - t)^{n-k} \delta_{\frac{k}{n}} \; ; \qquad {}^nC_k = \frac{n!}{k!(n-k)!}.$$

So, let us so, we have $f_i(x) = x^i$, $i = 0, 1, 2$. $f_0 \equiv 1$, $f_1(x) = x$, $f_2(x) = x^2$.

So, we can compute all the integrals of these functions with respect to this measure. So, it is a good exercise in computing (())(4:02) so, what is $\mu_n^t(X)$? So,

$$\mu_n^t(X) = \int_X f_0 d\mu_n^t = \sum_{k=0}^{n} {}^nC_k t^k (1-t)^{n-k} = (t+1-t)^n = 1.$$

(Refer Slide Time: 4:53)



Now, let us compute the integral

$$\int_X f_1 d\mu_n^t = \sum_{k=0}^{n} {}^nC_k t^k (1-t)^{n-k} \frac{n}{k} \qquad ; \int_X f_1 d\delta_{\frac{k}{n}} = f_1\left(\frac{k}{n}\right) = \frac{k}{n}.$$

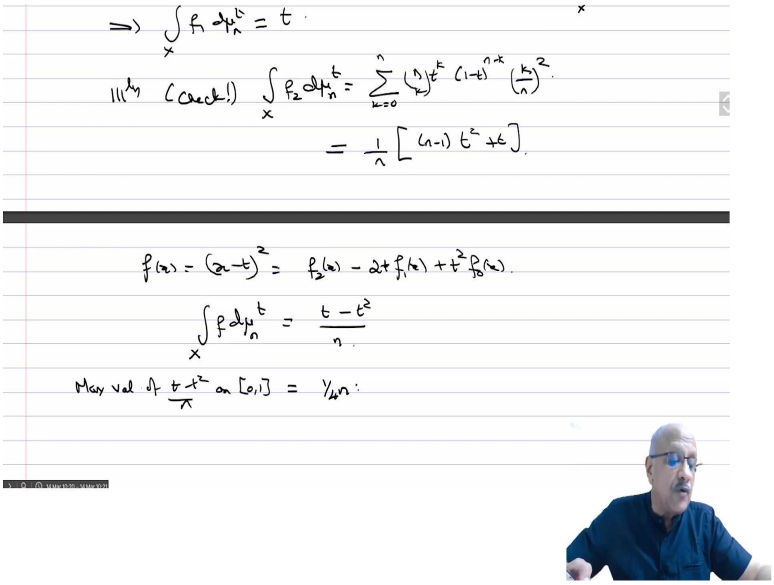
$$= t \sum_{k=1}^{n} {}^nC_k t^{k-1} (1-t)^{n-k} \frac{n}{k}$$

$$= t \sum_{k=1}^{n} {}^nC_k t^{k-1} (1-t)^{(n-1)-(k-1)}$$

$$= t \sum_{k=1}^{n} {}^{n-1}C_{k-1} t^{k-1} (1-t)^{(n-1)-(k-1)}$$

$$= t \sum_{k=0}^{n-1} {}^{n-1}C_{k-1} t^{k-1} (1-t)^{(n-1)-(k-1)} = t(t+1-t)^{n-1} = t.$$

(Refer Slide Time: 8:33)



So, we get $\int_X f_1 d\mu^t_n = t$.

So, similarly, you can make the calculation so, I will allow you to check this by just playing around with the binomial coefficients. So,

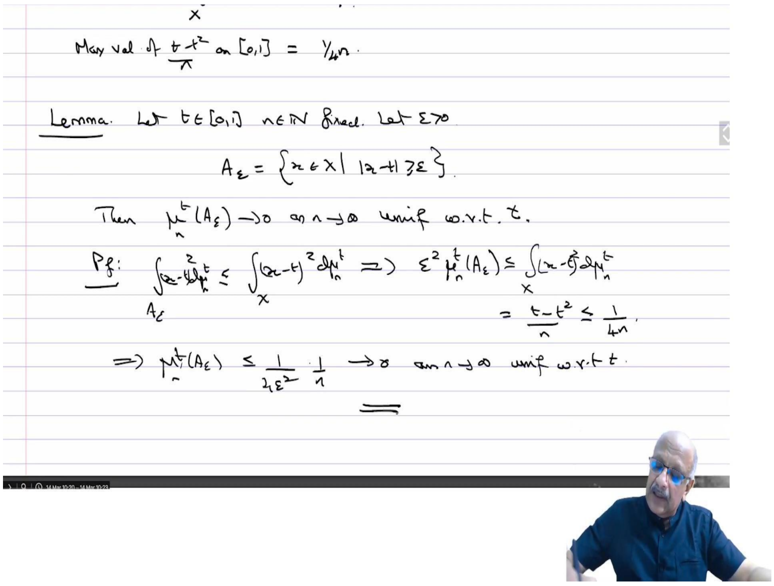
$$\int_X f_2 d\mu^t_n = \sum_{k=0}^{n} {}^nC_k t^k (1-t)^{n-k} \left(\frac{n}{k}\right)^2 = \frac{1}{n}\left[(n-1)t^2 + t\right].$$

So, now, if $f(x) = (x-t)^2 = f_2(x) - 2tf_1(x) + t^2 f_0(x)$ and therefore

$$\int_X f \; d\mu^t_n = \frac{t-t^2}{n}.$$

Now, the maximum value of $\frac{t-t^2}{n}$ on $[0, 1] = \frac{1}{4n}$. You can check.

(Refer Slide Time: 10:58)



Lemma: Let $t \in [0, 1]$, $n \in \mathbb{N}$ *fixed*. Let $\epsilon > 0$,

$$A_{\epsilon} = \{x \in X: |x - t| \geq \epsilon\}.$$

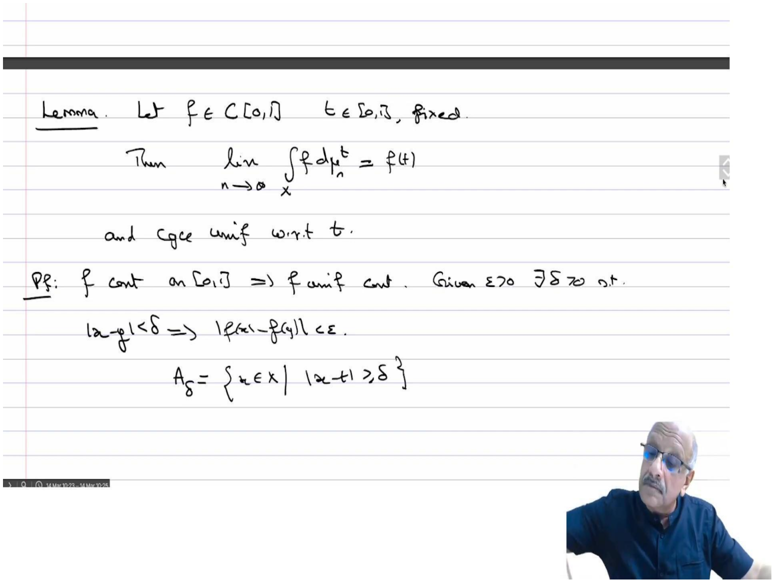
Then $\mu^{t}_{n}(A_{\epsilon}) \to 0$ as $n \to \infty$ uniformly w.r. to t.

proof: So, we are going to take $\int_{A_{\epsilon}} (x - t)^{2} d\mu^{t}_{n} \leq \int_{X} (x - t)^{2} d\mu^{t}_{n}$

$$\Rightarrow \epsilon^{2} \mu^{t}_{n}(A_{\epsilon}) \leq \int_{X} (x - t)^{2} d\mu^{t}_{n} = \frac{t-t^{2}}{n} \leq \frac{1}{4n}.$$

$$\Rightarrow \mu^{t}_{n}(A_{\epsilon}) \leq \frac{1}{4\epsilon^{2}} \frac{1}{n} \to 0 \ as \ n \to \infty \text{ uniformly w.r. to t.}$$

(Refer Slide Time: 13:12)



Lemma: Let $f \in C[0, 1]$, $t \in [0, 1]$ fixed. Then

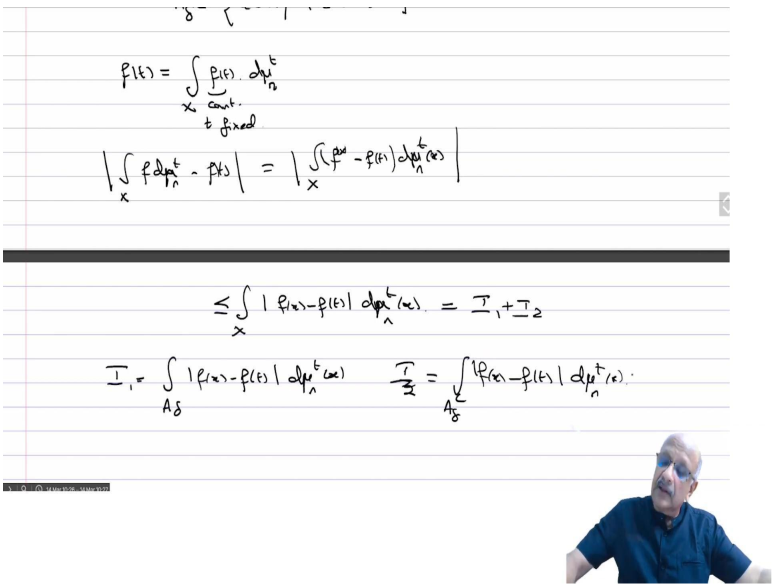
$$\lim_{n \to \infty} \int_X f d\mu_n^t = f(t)$$

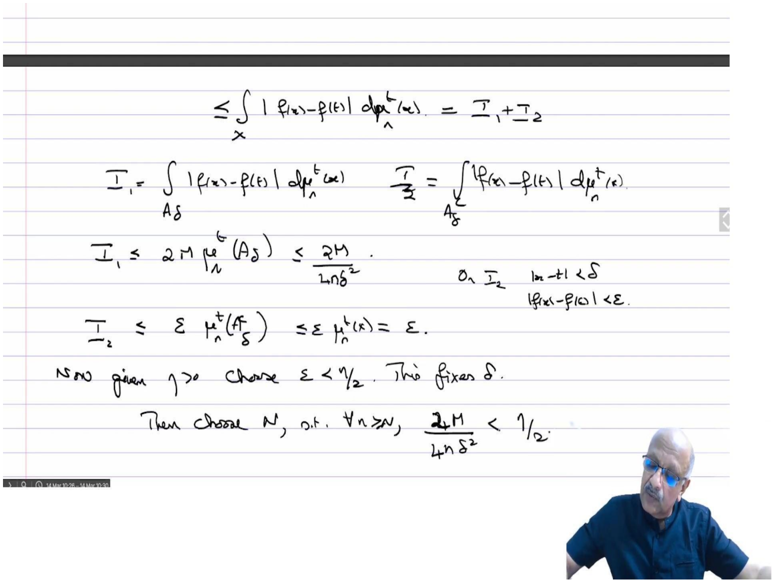
and convergence is uniform with respect to t.

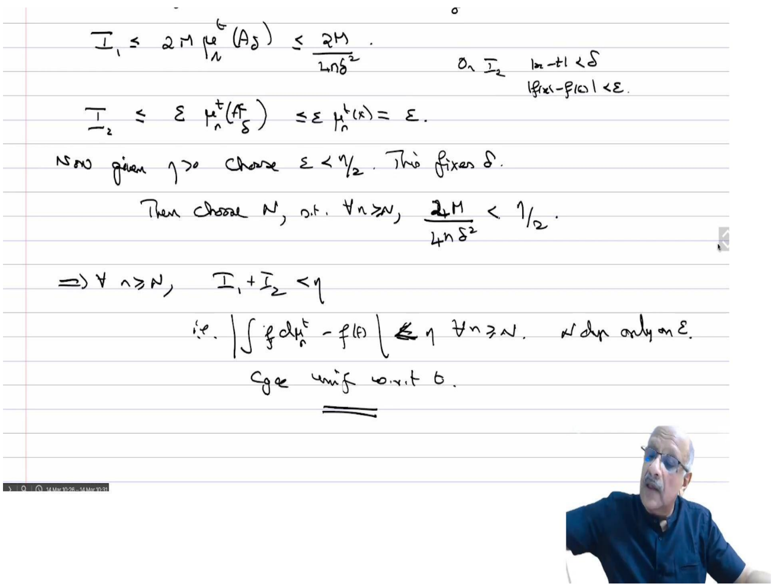
Proof: so f continuous on $[0, 1] \Rightarrow f$ is uniformly continuous. Given epsilon positive there exists delta positive such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon$. So, you take

$$A_\delta = \{x \in X: |x - t| \geq \delta\}.$$

(Refer Slide Time: 15:03)







So, $f(t) = \int_X f(t) d\mu_n^t$. Then

$$\left|\int_X f(t) d\mu_n^t - f(t)\right| = \left|\int_X (f(x) - f(t)) d\mu_n^t(x)\right| \le \int_X |f(x) - f(t)| d\mu_n^t(x) = I_1 + I_2.$$

So, $I_1 = \int_{A_\delta} |f(x) - f(t)| d\mu_n^t(x)$; $I_2 = \int_{A_\delta^c} |f(x) - f(t)| d\mu_n^t(x)$.

Now, let us assume that $|f(x)| \le M$, $\forall\, x \in [0, 1]$ because it is a continuous function and therefore, it is bounded. So,

$$I_1 \le 2M\mu_n^t A_\delta \le \frac{2M}{4n\delta^2}.$$

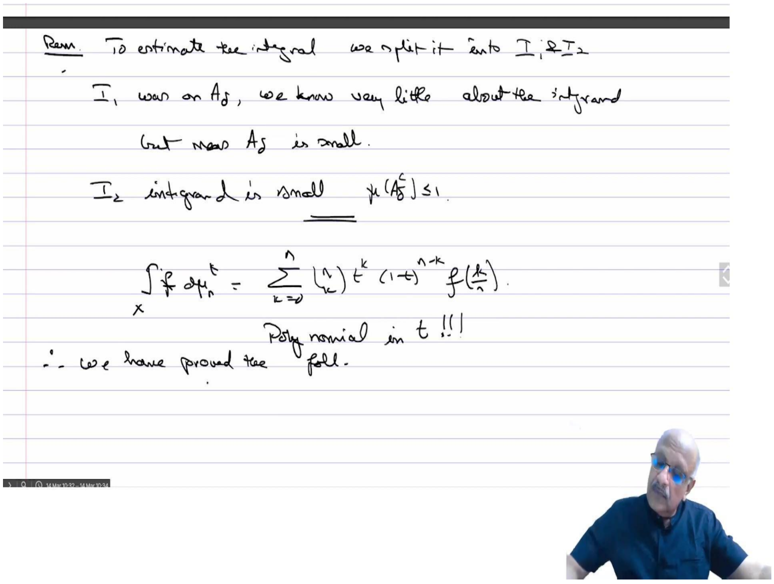
Now, $I_2 \le \epsilon\mu_n^t(A_\delta^c) \le \epsilon\mu_n^t(X) = \epsilon$.

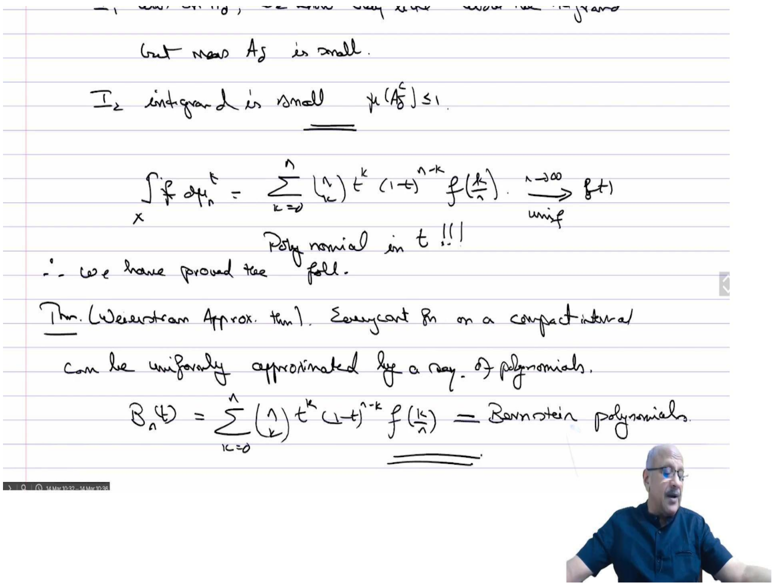
So, now, given $\eta > 0$, choose $\epsilon < \frac{\eta}{2}$, this fixes delta then choose capital N such that for all $n \ge N$ we have $\frac{2M}{4n\delta^2} < \frac{\eta}{2}$. Then for all $n \ge N$, you have $I_1 + I_2 < \eta$. So,

$$\left|\int_X f(t)d\mu_n^t - f(t)\right| < \eta,\ \forall\, n \ge N,\ N\ depends\ on\ \epsilon.$$

So, that proves the lemma.

(Refer Slide Time: 20:41)





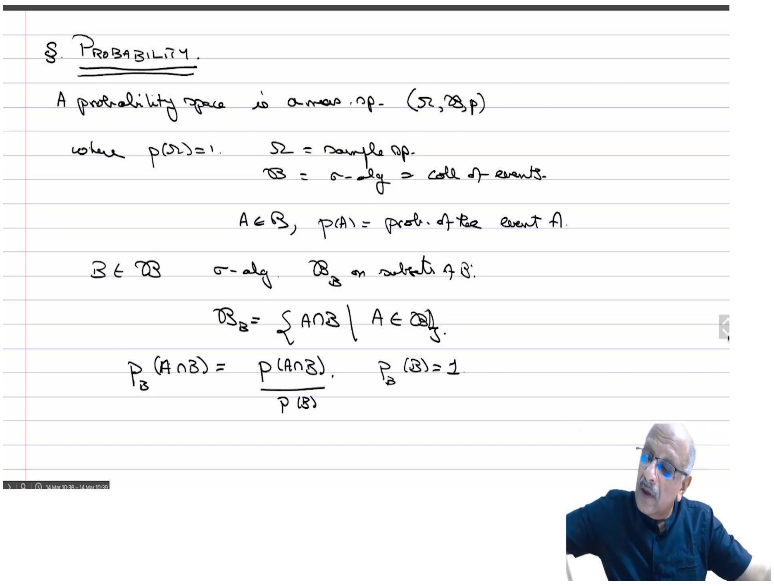
Remark: In this proof what we do so, we split so remark to estimate the integral we split it into integrals into I_1 and I_2, I_1 was on A_δ we know very little about the integrant but measure of A_δ is small. I_2 integrant is small and measures of $\mu(A_\delta^c) \leq 1$.

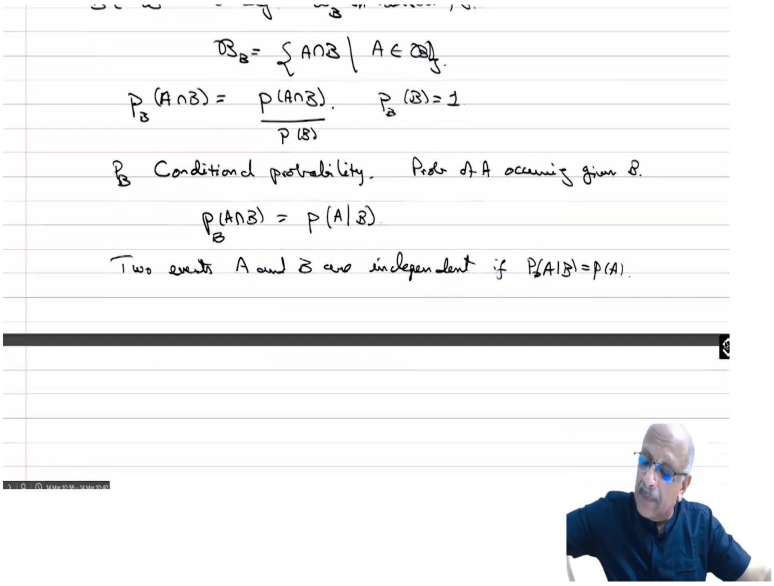
So, we split the integral on 1, we have information on the measure of the set on the other we have information on the integral. So, using these two inter complementary information we are able to estimate the integral this kind of divide and rule policy is very helpful it is a very useful technique to know. So, this is a technique which can come in useful whenever you want to estimate some intervals.

So, now $\int_X f(t) d\mu_n^t = \sum_{k=0}^{n} {}^nC_k t^k (1-t)^{n-k} f(\frac{k}{n}) \rightarrow f(t)\ uniformly\ as\ n \rightarrow \infty$.

Now, this polynomial Bn of t which is sigma k equals 0 to n, n ck t power k, 1 minus t power n minus k f of k by n these are called the Bernstein polynomials, so for each n, you have a polynomial. So, these are called the Bernstein partners. So, we have proved simultaneously two things, one is that you can approximate continuous function uniformly by a sequence of polynomials and we have also identified what those polynomials are and so, all this has come simply by discussing integration with respect to the Dirac measure suitably.

(Refer Slide Time: 25:25)





So, this is a nice application. So, now, finally, I want to talk a little about probability theory, so we will try to give a dictionary between measure theory and the theory of probability. So, the probability space is a measure space $(\Omega,\ B,\ P)$, $where\ P(\Omega) = 1$ and Ω is called the sample space, B the sigma algebra equals collection of events.

So, if $A \in B$, then $P(A) =$ the probability of event A. So, now, if $B \in B$ is an event, then you define a sigma algebra B_B on subsets of B:

$$B_B = \{A \cap B\colon A \in B\}.$$

We have already done this when defining the integral over subsets, so, BB is nothing but a set of all intersections A intersection B, A is an event, and you define a probability on B intersection A intersection B so BB of A intersection we already have done this before.

So, this is nothing but the usual probability of A intersection B it says, but we want to have this as a probability measure. So,

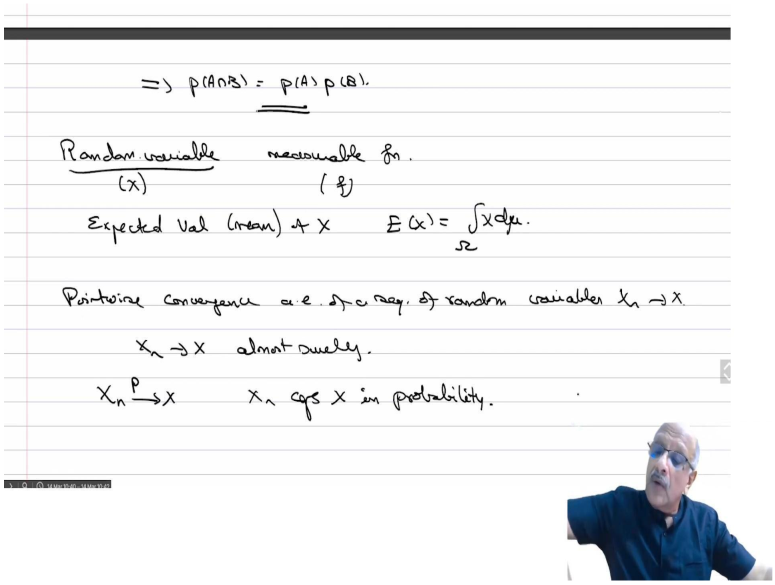
$$P_B(A \cap B) = \frac{P(A \cap B)}{P(B)};\ P_B(B) = 1.$$

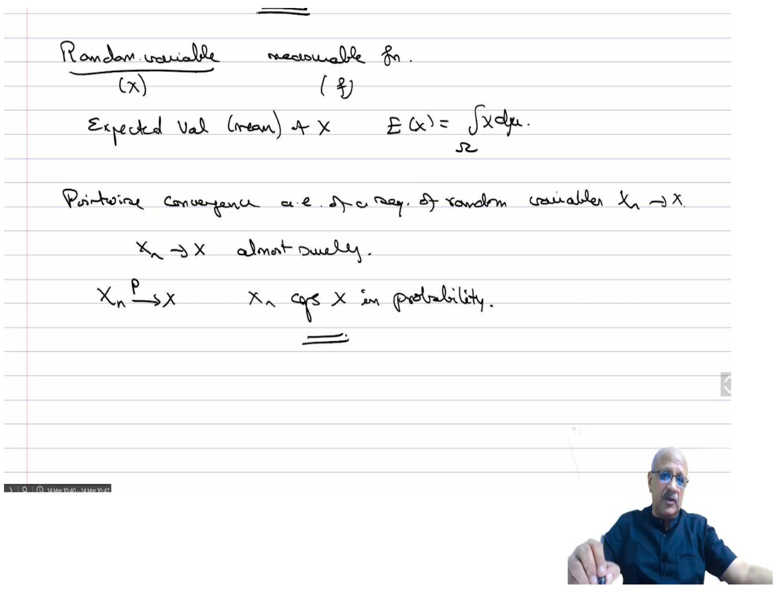
So, P_B is called the conditional probability. So, probability of A occurring given B given that B is happening, what is the probability of A also happening and that is precisely the probability measure the probability of A intersection B and then divided by the probability of B and you have this.

So, you have $P_B(A \cap B) = P(A|B)$.

Now, two events A and B are independent if $P(A|B) = P(A)$. so, it does not matter whether B happens or B does not happen, it does not affect the occurrence of A.

(Refer Slide Time: 29:14)



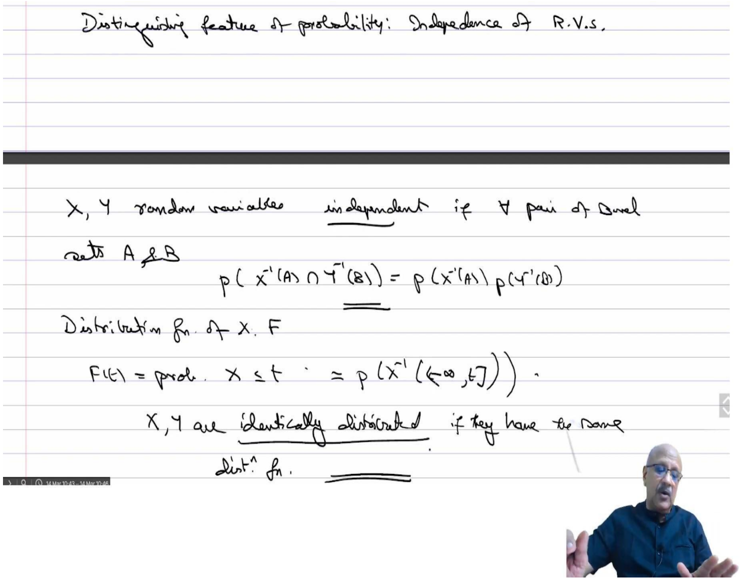


And therefore, this means that $P(A \cap B) = P(A)P(B)$. Now, the random variable (X) is nothing but a measurable function (f) only in probability theory we use the notation of here. So, instead we use value x, so X is a random variable, so, we use the symbol X for the set.

Now, the set is an omega that is a sample space and therefore, random variables are defined by this. So, the expected value or mean of X is $E(X) = \int_{\Omega} X d\mu$.

Point wise convergence a.e. of a sequence of random variables $X_n \to X$, then we say $X_n \to X$ almost surely. So, if $X_n \to X$ in measure then we say $X_n \to X$ in probability. So, it is probability really just measure theory no, because between there is something which is in probability which does not come in study in measure theory.

(Refer Slide Time: 31:22)



So, namely, the distinguishing feature of probability is independence of random variables. So, X, Y random variables independent if for every pair of Borel sets A and B, we have

$$P(X^{-1}(A) \cap Y^{-1}(B)) = P(X^{-1}(A))P(Y^{-1}(B)).$$

So, this is the thing that is whether the random variable X takes value in A is independent of the fact that the random variable of Y takes the value in B. So, this is what we call independence of random variables.

Similarly, distribution function of a random variable X so, this is equal to probability, so,

$$F(t) = \text{probability of } X \leq t = P(X^{-1}(-\infty, t]).$$

So, X and Y are identically distributed if they have the same distribution function so, sequences of independent identically distributed random variables is a very important study in the stochastic process. A stochastic process is nothing but a family of random variables.

So, it is your family so you study a family of measurable functions at the same time. So, there is a second index which will come in and therefore, you have so this is some kind of brief dictionary which tells you how to understand the probabilistic language in terms of measure theoretic language. So, now with this I will conclude this chapter on integration and we will do some exercises before we leave it altogether.