[Lec 23]

Convergence in probability to a constant $X_n \xrightarrow{P} c$ means

$$\forall \varepsilon > 0 \quad \lim P(|X_n - c| \geq \varepsilon) = 0.$$

Thm: If X_n has a finite variance for all n and $E[X_n] = \mu$ for all n, then...
First consider Chebyshev's Inequality:

$$P(|X_n - \mu| \geq \varepsilon) \leq \frac{\sigma_n^2}{\varepsilon^2}$$

now take limits wrt n of both sides:

$$\lim P(|X_n - \mu| \geq \varepsilon) \leq \lim \frac{\sigma_n^2}{\varepsilon^2}$$

Since probabilities are in $[0,1]$ if the rhs is 0, then the inequality becomes an equality. Thus, if we show:

$$\lim \frac{\sigma_n^2}{\varepsilon^2} = 0 \Rightarrow X_n \xrightarrow{P} \mu$$

$$\frac{1}{\varepsilon^2} \lim \sigma_n^2 = 0 \Rightarrow \lim \sigma_n^2 = 0$$

e.g.: $X_n \sim U(-\frac{1}{n}, \frac{1}{n})$ WTS $X_n \xrightarrow{P} 0$

$$\mu = E[X_n] = 0,\ \sigma_n^2 = \frac{(\frac{1}{n} - -\frac{1}{n})^2}{12} = \frac{4}{12n^2}$$

$$\lim \sigma_n^2 = \lim \frac{4}{12n^2} = 0 \Rightarrow X_n \xrightarrow{P} 0.$$

Another e.g.: $X_n \sim N(0, \frac{1}{n})$ WTS $X_n \xrightarrow{P} 0$

$$\mu = 0,\ \sigma_n^2 = \frac{1}{n} \quad \lim \sigma_n^2 = \lim \frac{1}{n} = 0 \Rightarrow X_n \xrightarrow{P} 0$$

Let $X_1, X_2, \ldots$ be iid w/ mean μ and variance σ^2 both finite.

$$\overline{X}_n := \frac{1}{n} \Sigma X_i \qquad E[\overline{X}_n] = \mu,\ Var[\overline{X}] = \frac{\sigma^2}{n}$$

$$\lim Var[\overline{X}] = \lim \frac{\sigma^2}{n} = 0 \Rightarrow \overline{X} \xrightarrow{P} 0.$$

This is called the "weak" "weak law of large numbers" (WLLN).

[illegible]e average converges to it's mean w/a "large #" of samples (n).

You actually don't need a finite variance for this to be true (see HW)

[illegible] in [illegible]

This is called weak b/c convergence in prob. is actually a weak type of convergence. We don't have time to talk [illegible] "[illegible] sure converges".

The third type of convergence is called "convergence in law" or "convergence in L^r norm" where $r \geq 1$. As before, we will only discuss convergence in law to a constant c. So:

$$X_n \xrightarrow{L^r} c \text{ means } \lim E[|X_n - c|^r] = 0.$$

E.g.: $X_n \xrightarrow{L^1} c$ means $\lim E[|X-c|] = 0$ "convergence in mean"

$X_n \xrightarrow{L^2} c$ means $\lim E[(X-c)^2] = 0$ "mean square convergence"

A good E.g.: $X_n \sim U(0, \frac{1}{n})$

$$\overset{\text{pdf}}{=} n \mathbb{1}_{x \in [0, \frac{1}{n}]}$$

WTS $X_n \xrightarrow{L^r} 0$ $\quad \lim E[|X_n - 0|^r] = \lim E[X^r]$

$$= \lim \int_{\mathbb{R}} x^r n \mathbb{1}_{x \in [0, \frac{1}{n}]} dx$$

$$= \lim n \int_0^{\frac{1}{n}} x^r dx = \lim n \left[\frac{x^{r+1}}{r+1}\right]_0^{\frac{1}{n}} = \frac{1}{r+1} \lim n \left(\frac{1}{n}\right)^{r+1}$$

$$= \frac{1}{r+1} \lim \frac{1}{n^r} \approx 0.$$

Which convergence is stronger? Law or probability?

$$X_n \xrightarrow{L^r} c \Rightarrow X_n \xrightarrow{P} c \quad \forall r \geq 1$$

Proof:

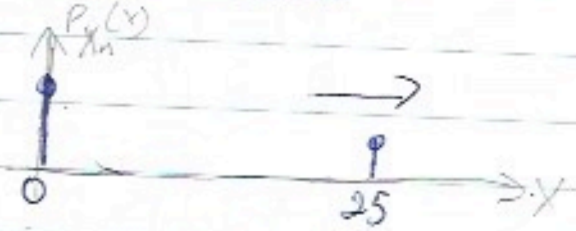
$$\lim P(|X_n - c| \geq \varepsilon) \overset{\text{raise to the r-th power}}{=} \lim P(|X_n - c|^r \geq \varepsilon^r) \leq$$

$$\leq \lim \frac{E[|X_n - c|^r]}{\varepsilon^r} = \frac{1}{\varepsilon^r} \lim E[|X_n - c|^r]$$

$$= 0$$

$$X_n \xrightarrow{P} c \not\Rightarrow X_n \xrightarrow{L^r} c$$ Counterexample:

$$X_n \sim \begin{cases} n^2 & \text{wp } \frac{1}{n} \\ 0 & \text{wp } 1 - \frac{1}{n} \end{cases}$$



It's clear that $X_1 \xrightarrow{P} 0$ But...

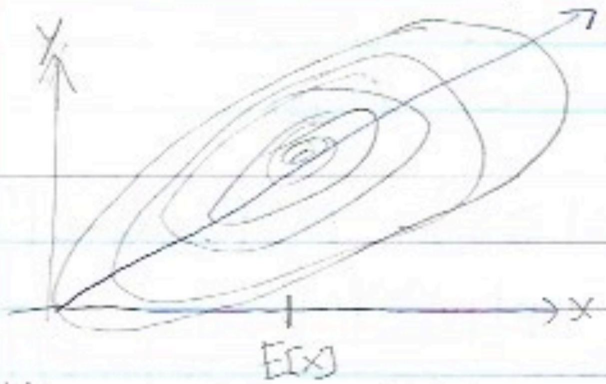
$$E[X^r] = \sum_{x \in \{0, n^2\}} x^r p(x) \quad r=1$$

$$= 0^r\left(1 - \frac{1}{n}\right) + (n^2)^r \frac{1}{n} = n^{2r-1} = n$$

Complete the Unit

Law of Iterated Expectation

Consider two r.v's X, Y and their joint density $f_{X,Y}(x,y)$



Here's a nice identity:

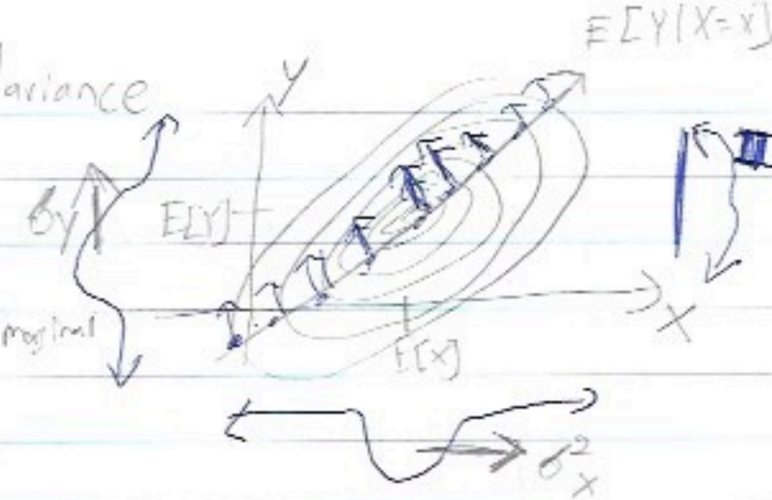
$$E[Y] = \int_{\mathbb{R}} y\, f_Y(y)\, dy = \int_{\mathbb{R}} y \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy =$$

$$\int_{\mathbb{R}} y \int_{\mathbb{R}} f_{Y|X}(y,x)\, f_X(x)\, dx\, dy = \int_{\mathbb{R}}\int_{\mathbb{R}} y\, f_{Y|X}(y,x)\, f_X(x)\, dy\, dx$$

$$= \int_{\mathbb{R}} f_X(x) \underbrace{\int_{\mathbb{R}} y\, f_{Y|X}(y,x)\, dy}_{E[Y|X=x]}\, dx = \int_{\mathbb{R}} \underbrace{E[Y|X=x]\, f_X(x)}_{g(x)}\, dx$$

$$= E_X\left[E_Y[Y|X=x]\right]$$

Law of Total Variance



Here's a nice identity:

$$\mathrm{Var}[Y] = E[Y^2] - E[Y]^2$$

$$= E_X\left[E_Y[Y^2|X]\right] - E_X\left[E_Y[Y|X]\right]^2$$

$$= E_X\left[\mathrm{Var}_Y[Y|X] + E_Y[Y|X]^2\right] - E_X\left[E_Y[Y|X]\right]^2$$

$$= E_X\left[\mathrm{Var}_Y[Y|X]\right] + E_X\left[\overbrace{E_Y[Y|X]}^{C}{}^2\right] - E_X\left[\overbrace{E[Y|X]}^{C}\right]^2$$

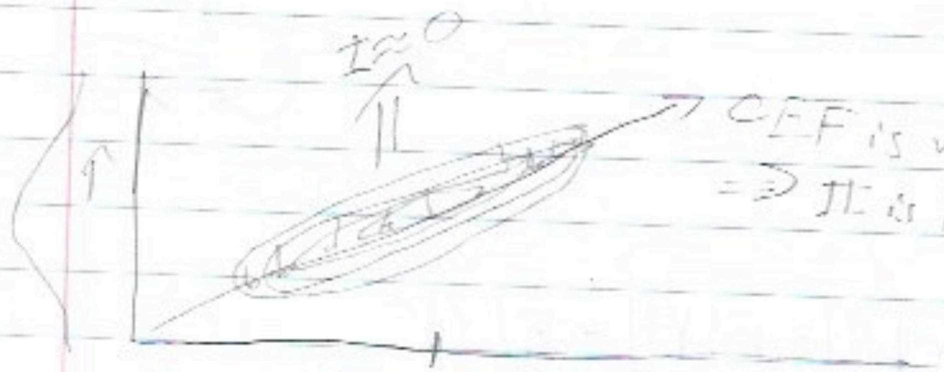
$$= E_X\left[\mathrm{Var}_Y[Y|X]\right] + \underbrace{E_X[C^2] - E_X[C]^2}_{\mathrm{Var}_X[C]}$$

$$\mathrm{Var}_Y[Y] = \underbrace{E_X\left[\mathrm{Var}_Y[Y|X]\right]}_{\mathrm{I}} + \underbrace{\mathrm{Var}_X\left[E_Y\left[E_Y[Y|X]\right]\right]}_{\mathrm{II}}$$

(I) Mean of the conditional variances. This is large when

(II) Variance of the conditional means. This will be large when the CEF line is tilted in place of high X density.

CFF $\Rightarrow \Pi \approx 0$



CFF is variable
$\Rightarrow \Pi$ is large