Lecture 23 12/7/20

Math 621
Prof. Kapelner

Convergence in prob. to a constant: $X_n \xrightarrow{P} c$ means $\forall \varepsilon > 0 \lim_{n\to\infty} P(|X_n - c| \geq \varepsilon) = 0$.

* Thm - If $X_n$ has a finite variance $\forall n$ & $E[X_n] = \mu$, $\forall n$, then.....
First consider, Chebyshev's inequality:

$P(|X_n - \mu| \geq \varepsilon) \leq \frac{\sigma_n^2}{\varepsilon^2}$ now take limits w.r.t. n of both sides:

$\lim P(|X_n - \mu| \geq \varepsilon) \leq \lim \frac{\sigma_n^2}{\varepsilon^2}$ since probabilites 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. $X_n \sim U(-\frac{1}{n}, \frac{1}{n})$ W.T.S. $X_n \xrightarrow{P} 0$

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

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

e.g. $X_n \sim N(0, \frac{1}{n})$

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

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

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

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

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

the averge conv. to it's mean with a "large number" of samples (n).

Don't need a finite variance for this to be true.

this is called weak b/cuz conv. in prob. is actually a weak type of conv.

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

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

e.g. $X_n \sim U(0, \frac{1}{n})$

$\quad\quad\quad\quad ||$

$\quad\quad n \mathbb{1}_{x \in [0, \frac{1}{n}]}$

WTS $X_n \xrightarrow{L^r} 0$

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

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

$$= 0.$$

Which convergence is stronger? Law or probability?

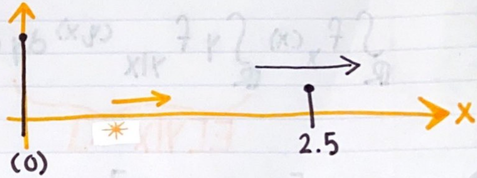
$$X_n \xrightarrow{L^r} c \Rightarrow X_n \xrightarrow{P} c \quad \forall r \geq 1$$

Proof: $\lim P(|X_n-c| \geq \varepsilon) = \lim P(|X_n-c|^r \geq \varepsilon^r) \leq \lim \frac{E[|X_n-c|^r]}{\varepsilon^r} = \frac{1}{\varepsilon^r} \lim E[|X_n-c|^r] = 0$ ✓

$X_n \xrightarrow{P} c \not\Rightarrow X_n \xrightarrow{L^r} c$ . $P_{X_n}(x)$

Counter example:

$$X_n \sim \begin{cases} n^2 & \text{w.p. } 1/n \\ 0 & \text{w.p. } 1-1/n \end{cases}$$



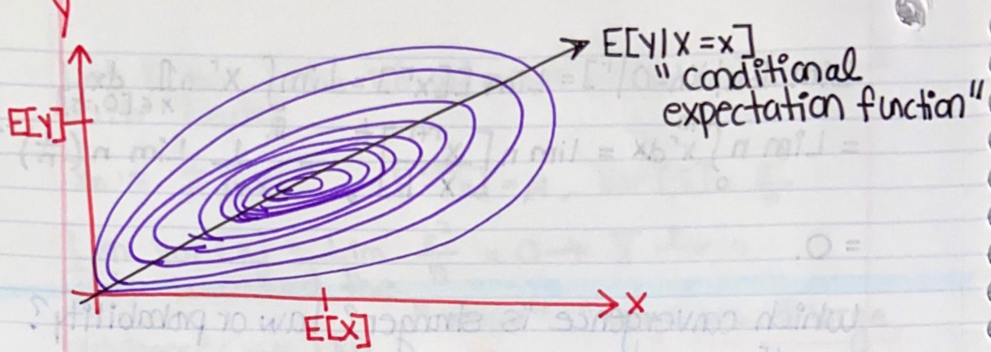

It is clear that

$X_n \xrightarrow{P} 0$ but.... $E[X^r] = \sum_{x\in\{0,n^2\}} x^r p(x)$

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

## 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:

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

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

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

$$= E_X\big[E_Y[Y|X=x]\big]$$

## Law of total Variance

Here's a nice identity: $Var[Y] = E[Y^2] - E[Y]^2$

$$= E_X[E_Y[Y^2|X]] - E_X[E_Y[Y|X]]^X = E_X[Var_Y[Y|X] +$$

$$E_Y[Y^2|X]^2] - E_X[E_Y[Y|X]]^2 = E_X[Var_Y[Y|X]] +$$

Let $c = E[Y|X]$

$$\underbrace{E_x[E_y[Y|X]^2]}_{c^2} - E_x[\underbrace{E[Y|X]}_{c}]^2 = E_x[Var_y[Y|X]] +$$

$$\underbrace{E_x[c^2] - E_x[c]^2}_{Var_x[c]}$$

$$Var_y[Y] = \underbrace{E_x[Var_y[Y|X]]}_{\mathrm{I}} + \underbrace{Var_x[E_y[Y|X]]}_{\mathrm{II}}$$

(I) Mean of conditional variances

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