Convergence in probability to a constant: $X_n \xrightarrow{p} c$. This means:
$\forall \varepsilon > 0 \quad \lim P(|X_n - c| \geq \varepsilon) = 0$.

Thm: If X_n has expectation mu for all n and variance sigsq_n which is finite for all n, then lim sigsq_n = 0 => $X_n \xrightarrow{p} \mu$. Proof:

Recall Chebyshev's Inequality:

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

now take limits of both sides:

$$\lim P(|X_n - \mu| \geq \varepsilon) \leq \lim \frac{\sigma_n^2}{\varepsilon^2} = \frac{1}{\varepsilon^2} \lim \sigma_n^2 = 0$$

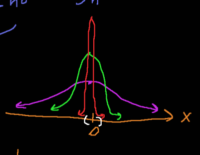
because probabilities are between 0 and 1, if you know the probability is <= 0. That means the probability is 0.

$$\Rightarrow \lim P(|X_n - \mu| \geq \varepsilon) = 0 \Rightarrow X_n \xrightarrow{p} \mu.$$

e.g. $X_n \sim U\left(-\frac{1}{n}, \frac{1}{n}\right)$ Prove $X_n \xrightarrow{p} 0$.

$$E[X_n] = 0 = \mu \ \forall n, \quad \sigma_n^2 = \frac{\left(\frac{1}{n} - -\frac{1}{n}\right)^2}{12} = \frac{4}{12n^2} = \frac{1}{3n^2}$$

$$\lim \sigma_n^2 = \lim \frac{1}{3n^2} = 0 \overset{\text{thm}}{\Longrightarrow} X_n \xrightarrow{p} 0 \checkmark$$



e.g. $X_n \sim N\left(0, \frac{1}{n}\right)$. Prove $X_n \xrightarrow{p} 0$.

$$E[X_n] = 0 = \mu \ \forall n, \quad \sigma_n^2 = \frac{1}{n}, \quad \lim \sigma_n^2 = \lim \frac{1}{n} = 0 \Rightarrow X_n \xrightarrow{p} 0 \checkmark$$

Let X_1, X_2, ... be iid with mean mu and variance sigsq < infinity.

$$\bar{X}_n = \frac{1}{n}\sum X_i, \quad E[\bar{X}_n] = \mu \ \forall n, \quad Var[\bar{X}_n] = \frac{\sigma^2}{n}$$

Prove $\bar{X}_n \xrightarrow{p} \mu$. $\lim Var\, \bar{X}_n = \lim \frac{\sigma^2}{n} = 0 \Rightarrow \bar{X}_n \xrightarrow{p} \mu$.

This is a very famous thm. It's called
the "weak" "weak law of large numbers" (WLLN).

thm.

because I assumed finite sigsq of the X_1, X_2,... rv's. You don't need it (see HW).

because convergence in probability is actually a weak type of convergence. It turns out you can prove "almost sure" convergence (but we won't discuss that).

As the "number" of samples increases, the average cannot "escape" from the mean.

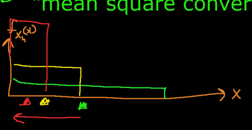
The last type of convergence we'll study is called "convergence in law" or "convergence in L^r norm" to a constant where r>=1 denoted: $X_n \xrightarrow{L^r} c$ which means by definition:

$$\lim E[|X_n - c|^r] = 0.$$

e.g. $r=1$ $\lim E[|X_n - c|] = 0$ "convergence in mean"

$r=2$ $\lim E[(X_n - c)^2] = 0$ "mean square convergence"

e.g. $X_n \sim U\left(0, \frac{1}{n}\right)$
$\quad \ddot{=} n \mathbb{1}_{x \in [0, \frac{1}{n}]}$



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

$$= \lim \int_{\mathbb{R}} x^r n \mathbb{1}_{x \in [0, \frac{1}{n}]} dx = \lim n \int_0^{\frac{1}{n}} x^r dx$$

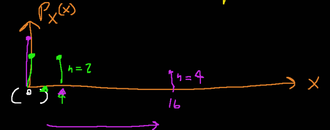
$$= \lim n \left[\frac{x^{r+1}}{r+1}\right]_0^{\frac{1}{n}} = \frac{1}{r+1} \lim n \frac{1}{n^{r+1}} = \frac{1}{r+1} \lim \frac{1}{n^r} = 0 \checkmark$$

HW: convergence in probability is stronger than convergence in distribution. Which convergence is stronger? Law or probability? Law. Proof for $X_n \xrightarrow{L^r} c \Rightarrow X_n \xrightarrow{p} c$.

$$\lim P(|X_n - c| \geq \varepsilon) = \lim P(\underbrace{|X_n - c|^r}_{Y} \geq \underbrace{\varepsilon^r}_{a}) \overset{\text{Markov's}}{\leq} \lim \frac{E[|X_n - c|^r]}{\varepsilon^r} = 0 \checkmark$$

$X_n \xrightarrow{p} c \not\Rightarrow X_n \xrightarrow{L^1} c$. Counterexample:

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

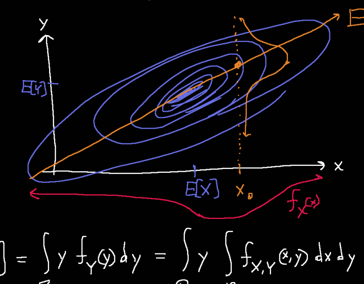


Clear that $X_n \xrightarrow{p} 0$.

$$\lim E[|X_n - 0|^1] = \lim E[X_n] = \lim \sum_{x \in \{0, n^2\}} x\, p_X(x) = \lim 0\cancel{\left(1 - \frac{1}{n}\right)}^{0} + n^2 \frac{1}{n}$$

$$= \lim n = \infty \neq 0 \Rightarrow X_n \not\xrightarrow{L^1} 0.$$

Law of Iterated Expectation. Consider two r.v. X, Y with jdf $f_{X,Y}(x,y)$



conditional expectation function (CEF).

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

now I'll switch the order of integration:

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

Law of Total Variance.

$$Var_Y[Y] = E_Y[Y^2] - E_Y[Y]^2$$

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

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

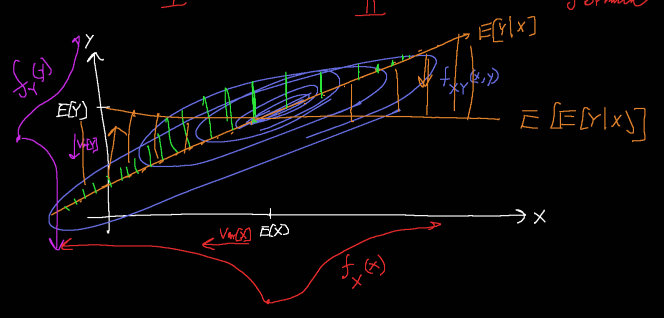
$$= E_X[Var_Y[Y|X]] + E_X[E_Y[Y|X]^2] - E_X[E_Y[Y|X]]^2$$

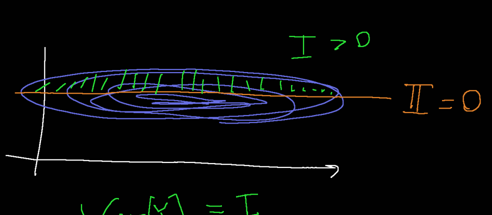
let $C = E_Y[Y|X]$

$$= E_X[Var_Y[Y|X]] + \underbrace{E_X[C^2] - E_X[C]^2}_{Var_X[C]}$$

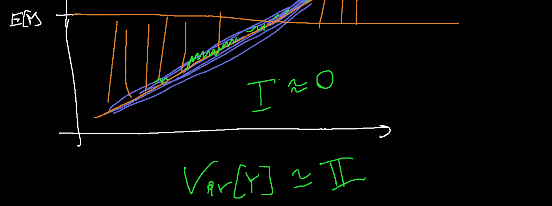
$$Var_Y[Y] = \underbrace{E_X[Var_Y[Y|X]]}_{\mathrm{I}} + \underbrace{Var_X[E_Y[Y|X]]}_{\mathrm{II}}$$

decomposition formula





$Var[Y] = \mathrm{I}$



$Var[Y] \approx \mathrm{II}$