Convergence in probability 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 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:

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

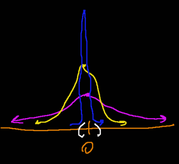
since probablities are in [0,1] if the rhs is 0, then the inequality becomes an equality. Thus, if we show:

$$\lim \frac{\sigma_n^2}{\varepsilon^2} = 0 \Rightarrow \boxed{X_n \xrightarrow{P} \mu}$$
$$\frac{1}{\varepsilon^2} \lim \sigma_n^2 = 0 \Rightarrow \lim \sigma_n^2 = 0$$

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

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

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



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

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

Let X_1, X_2, ... be iid with mean mu and variance sigsq both finite.

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

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

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

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

This is called weak because convergence in probability is actually a weak type of convergence. We don't have time to talk about "almost sure convergence".

The average converges to its mean with a "large number" of samples (n). It can't "escape"!

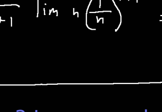
The third type of convergence is called "convergence in law" or "convergence in L^r norm" where $r \geq 1$. As before, we will only discuss convergence in law to a constant c. So:

$X_n \xrightarrow{L^r} c$ 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"

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

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



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

$= \lim n \int_0^{\frac{1}{n}} x^r dx = \lim n \left[\frac{x^{r+1}}{r+1}\right]_0^{\frac{1}{n}} = \frac{1}{r+1} \lim n \left(\frac{1}{n}\right)^{r+1} = \frac{1}{r+1} \lim \frac{1}{n^r} = 0.$

Which convergence is stronger? Law or probability?

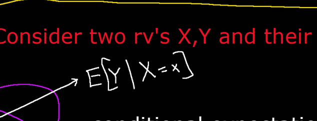
$X_n \xrightarrow{L^r} c \Rightarrow X_n \xrightarrow{P} c \quad \forall r \geq 1. \quad$ Pf:

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

$= \frac{1}{\varepsilon^r} \lim E[|X_n - c|^r] = 0$ ✓

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

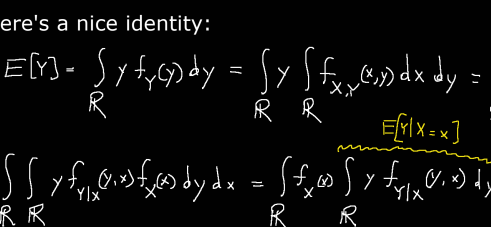
$X_n \sim \begin{cases} n^2 & \text{wp } \frac{1}{n} \\ 0 & \text{wp } 1 - \frac{1}{n} \end{cases}$



It's clear that $X_n \xrightarrow{P} 0$ But... $E[X^r] = \sum_{x\in\{0,n^2\}} x^r p(x)$

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

Law of Iterated Expectation. Consider two rv's X,Y and their joint density $f_{X,Y}(x,y)$.



conditional expectation function

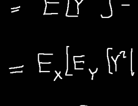
Here's a nice identity:

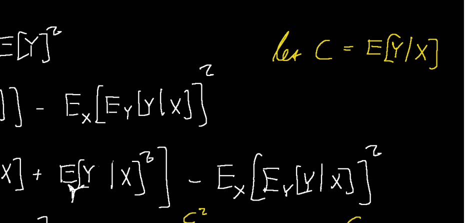
$E[Y] = \int_{\mathbb{R}} y f_Y(y)\, dy = \int_{\mathbb{R}} y \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy = \int_{\mathbb{R}} y \int_{\mathbb{R}} f_{Y|X}(y,x) f_X(x)\, dx\, dy$

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

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

Law of Total Variance





Here's a nice identity:

$Var[Y] = E[Y^2] - E[Y]^2 \qquad$ let $C = E[Y|X]$

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

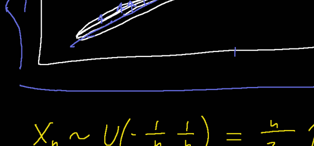
$= E_X\left[Var_Y[Y|X] + E_Y[Y|X]^2\right] - E_X[E_Y[Y|X]]^2$

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

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

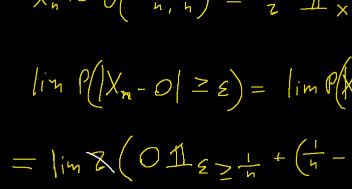
$Var_Y[Y] = \underbrace{E_X[Var_Y[Y|X]]}_{I} + \underbrace{Var_X[E_Y[Y|X]]}_{II}$

(I) Mean of the conditional variances. This is large when the spread around the CEF is high.
(II) Variance of the conditional means. This will be large when the CEF line is tilted in places of high X density.

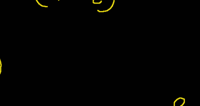


CEF $\Rightarrow$ II ≈ 0

I ≈ 0



CEF is variable $\Rightarrow$ II is large



$X_n \sim U\left(-\frac{1}{n}, \frac{1}{n}\right) = \frac{n}{2}\mathbb{1}_{x\in[-\frac{1}{n},\frac{1}{n}]}$ WTS $X_n \xrightarrow{P} 0$

$\lim P(|X_n - 0| \geq \varepsilon) = \lim P(|X_n| \geq \varepsilon) = \lim 2P(X_n \geq \varepsilon)$

$= \lim \cancel{2}\left(0\,\mathbb{1}_{\varepsilon \geq \frac{1}{n}} + \left(\frac{1}{n} - \varepsilon\right)\frac{n}{\cancel{2}}\mathbb{1}_{\varepsilon < \frac{1}{n}}\right)$

$= \lim_{n\to\infty} (1 - \varepsilon n)\mathbb{1}_{\varepsilon < \frac{1}{n}} = \lim_{n\to\infty}\left(1 - \cancel{\varepsilon}\frac{1}{\frac{\cancel{\varepsilon}}{2}}\right)\mathbb{1}_{\varepsilon < \frac{1}{\frac{1}{\varepsilon/2}}} = \lim (-1)\overbrace{\mathbb{1}_{\varepsilon < \frac{\varepsilon}{2}}}^{0} = 0$

Definition of limit as n-->infinity in calculus:

$\lim_{n\to\infty} a_n = a$ means $\forall \varepsilon_0 > 0 \ \exists n_0$ s.t $n \geq n_0 \ |a_n - a| < \varepsilon_0$

let $n_0 = \frac{1}{\frac{\varepsilon}{2}}$