WLLN

If $X_1, \dots, X_n$ are iid RVs w/ mean μ & finite variance, then

$$P\Big(\Big|\underbrace{\tfrac{1}{n}\sum_{i=1}^{n} X_i}_{\substack{\text{emperical mean} \\ \text{Def } M_n}} - \underbrace{\mu}_{\text{actual mean}}\Big| \geq \epsilon\Big) \to 0 \quad \text{as } n \to \infty$$

Proof: By Chebychev, $P\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$

The WLLN tells us that the emperical frequency estimate is a pretty good estimate of the actual mean

$\Rightarrow \boxed{\lim_{n\to\infty} P(|M_n - \mu| \geq \epsilon) = 0}$

Converging in probability:

$\forall \epsilon, \delta, \exists\, n(\epsilon, \delta)$ s.t. $P(|M_n - \mu| \geq \epsilon) < \delta \;\; \forall n > n(\epsilon, \delta)$

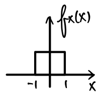
Remarks:

- ϵ captures "accuracy" level
- δ captures "confidence" level
- $n(\epsilon, \delta)$ captures "threshold" for a given accuracy & confidence level

We say $\boxed{M_n \xrightarrow{P} \mu}$ M_n converges to μ in probability

Formally, let $Y_1, \dots, Y_n$ be a sequence of RVs.
Then, we say the sequence converges to a number a in prob if
$\forall \epsilon > 0 \;\; \lim_{n\to\infty} P(|Y_n - a| > \epsilon) = 0.$

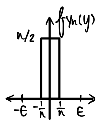
Ex 1: $X_1, \dots, X_n$ iid Unif$[-1, 1]$ RVs.



1/ $Y_n = \frac{X_n}{n}$ converge in prob?

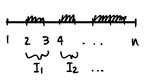
$Y_n \leq y \Rightarrow X_n \leq ny \Rightarrow F_{Y_n}(y) = F_X(ny)$

$f_{Y_n}(y) = n f_X(ny)$



$P(|Y_n - \underset{\text{limit}}{0}| > \epsilon) = 0$ if $\frac{1}{n} < \epsilon$ or $n > \frac{1}{\epsilon}$

Ex 2: Suppose time in discrete units and $Y_n = 1$,
there's some arrival at time n & $Y_n = 0$ else.
Def $I_k = \{2^k, 2^{k+1}, 2^{k+1} - 1\}$



Suppose exactly 1 arrival in each interval & it is equally likely in that interval.

$\Rightarrow P(Y_n = 1) = \frac{1}{2^k}$ if $n \in I_k$

$\lim_{n\to\infty} P(Y_n > 1) = \lim_{k\to\infty} \frac{1}{2^k} = 0$

↳ Y_n converges to 0 in prob.

Remark: Given any finite n, there are certain to be an infinite num of arrivals after n
$\Rightarrow$ this is the weakness of the weak law.

SLLN

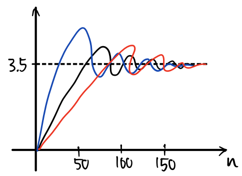
Let X_n be a seq of iid RVs w/ μ mean,

then $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$ as $n\to\infty$ w.p. 1.

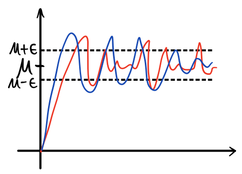
i.e. sample mean $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges to the exp value μ w.p. 1.

$\boxed{P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1}$ $\quad M_n \xrightarrow{a.s.} \mu$

Rolling 6-sided die → sample mean converges to 3.5



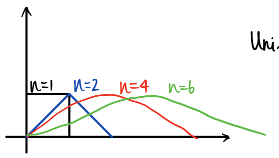
SLLN: Every realization converges to μ.



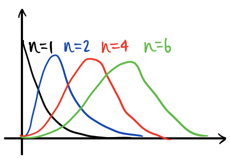
WLLN: May diverge out of the bound.

CLT

Q/ What happens to $S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n$, X_i iid



Unif



Exp

Observe: Looks more like a normal as we take more copies of any distr

Var & mean of $S_n \to \infty$ as $n \to \infty$ (distr gets really flat & wide)

↳ need to normalize

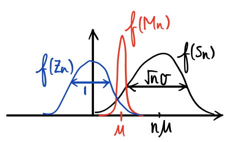
$\mathbb{E}[S_n] = n\mu$

$\mathrm{Var}(S_n) = n\sigma^2$

$M_n = \frac{S_n}{n}$

$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} \Rightarrow \mathbb{E}[Z_n] = 0$

$\mathrm{Var}(Z_n) = 1$



$\boxed{\text{CLT: } \lim_{n\to\infty} P(Z_n \le x) = \Phi(x) \;\; \forall x}$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}\,dz$ is the CDF of $N(0,1)$

Note: If there are a # of indep small factors, the aggregate of these factors will be approximately normally distributed.

⇒ models noise very well

Aside: If $Y \sim N(0,1) \Rightarrow M_Y(s) = \mathbb{E}[e^{sY}] = e^{s^2/2}$

$\log M_Y(s) = \frac{s^2}{2}$

Proof: Show that $M_Z(s)$ converges to $N(0,1)$.

$X_1, \dots, X_n$ are iid RVs w/ mean 0 & var 1.

$Z_n = \frac{\sum X_i}{\sqrt{n}}$

$$M_{Z_n}(s) = \mathbb{E}[e^{sZ_n}] = \mathbb{E}[e^{s\sum X_i/\sqrt{n}}]$$
$$= \mathbb{E}[e^{\frac{s}{\sqrt{n}}X_1} \cdot e^{\frac{s}{\sqrt{n}}X_2} \cdots e^{\frac{s}{\sqrt{n}}X_n}]$$
$$= \mathbb{E}[e^{\frac{s}{\sqrt{n}}X_i}]^n$$
$$\boxed{M_{Z_n}(s) = [M_X(\tfrac{s}{\sqrt{n}})]^n}$$

Recall: $M_X(0) = 1$, $M_X'(0) = 0$, $M_X''(0) = 1$

cumulant MGF

$$\lim_{n\to\infty} \log M_{Z_n}(s) = \lim_{n\to\infty} n \log M_X(\tfrac{s}{\sqrt{n}})$$
$$= \lim_{n\to\infty} \frac{\log M_X(\frac{s}{\sqrt{n}})}{1/n}$$

Let $y = \frac{1}{\sqrt{n}}$ $\quad = \lim_{y\to 0} \frac{\log M_X(sy)}{y^2}$

Use L'Hos $\quad = \lim_{y\to 0} \frac{s M_X'(sy)}{2y M_X(sy)}$

L'Hos $\quad = \lim_{y\to 0} \frac{s^2 M_X''(sy)}{2}$

$= \frac{s^2}{2}$

Ex: Polling

We ask n randomly chosen voters if they support candidate T $\Rightarrow X_i = \begin{cases} 1 & \text{if yes} \\ 0 & \text{if no} \end{cases}$ X_i's iid

Q/ We want 95% confidence that $|M_n - p| < \epsilon$ where p = true popularity of T,

M_n = emperical popularity of T $= \frac{1}{n}\sum X_i$

$X_i \sim B(p)$

$\mathbb{E}[X_i] = p$

$\mathrm{Var}(X_i) = p(1-p)$

Chebyshev: $P(|M_n - p| \geq a) \leq \frac{\mathrm{Var}(M_n)}{a^2} = \frac{p(1-p)}{n a^2}$
$\leq \frac{1}{4na^2}$

$a = 0.1$: $P(|M_n - p| \geq 0.1) \leq \frac{1}{4n(0.01)} = \frac{25}{n}$

≤ 0.05 (95% confidence)

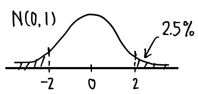
$\Rightarrow n \geq 500$

CLT: $\frac{M_n - \mathbb{E}[M_n]}{\sqrt{\mathrm{Var}(M_n)}} \xrightarrow{d} N(0,1)$

Want $P(|M_n - p| \geq 0.1) \leq 0.05$

$P(\underbrace{\frac{M_n - p}{\frac{1}{2}\sqrt{n}}}_{N(0,1)} \geq \underbrace{\frac{0.1}{\frac{1}{2}\sqrt{n}}}_{0.2\sqrt{n}}) \leq 0.05$

↳ now use Z table



$0.2\sqrt{n} \leq 2$

$\Rightarrow n \leq 100$