Prob inequalities

Many experiments / prob. calculations are difficult to perform, which naturally leads us to inequalities.

Often we aspire to find some prob of events even without the knowledge of underlying distribution.

$X_1, \dots X_n$ (iid), with $E(X)=0$; $E(X^2)=1$

$\therefore \dfrac{X_1 + \dots X_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$

cdf of LHS converges to cdf of RHS.

$$F_{\frac{X_1+\dots X_n}{\sqrt{n}}}(x) \longrightarrow \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt = \Phi(x)$$

27/08

Central limit theorem (Uniform convergence result)

$X_1, \dots X_n$ are iid,

$$\frac{S_n}{\sqrt{n}} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0,1)$$

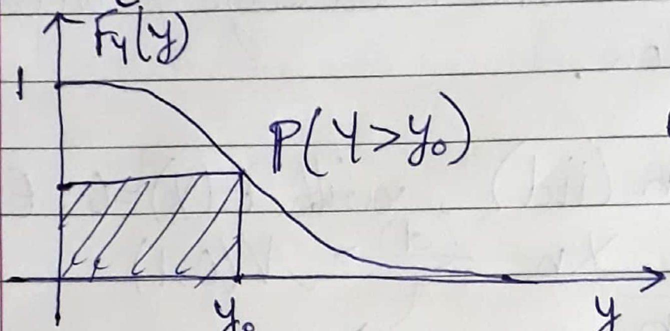
i.e. $\underbrace{F_{\frac{S_n}{\sqrt{n}}}(x)}_{P\left(\frac{S_n}{\sqrt{n}} \leq x\right)} \longrightarrow \Phi(x) \quad \forall x \in R$

↳ Prob. inequalities are the key to unravel these convergence results.

Markov inequality

Let $Y \geq 0$ be a r.v with $E(Y) < \infty$.
Then,
$$P(Y \geq y) \leq \frac{E(Y)}{y} \; ; \; y > 0$$

It is useful only when $E(y) < y$.



Recall,
$$E(Y) = \int_0^\infty \bar{F}_Y(y)\, dy$$
$$= \int_0^\infty P(Y>y)\, dy$$
$$= \int_0^\infty P(Y \geq y)\, dy$$

$$\therefore \underbrace{y_0 . P(Y \geq y_0)}_{\text{Area of shaded rect.}} \leq \int_0^\infty P(Y \geq y)\, dy = E(y)$$

Since $P(Y \geq y_0)$ is non-increasing

$$\therefore P(Y \geq y) \leq \frac{E(y)}{y}$$

That is, $P(Y \geq y)$ decreases at least as $\frac{1}{y}$ as $y \to \infty$ if $E(y) < \infty$

↳ One can show that if $E(Y) < \infty$, then
$$\lim_{y \to \infty} y . P(Y \geq y) = 0$$

Let X be a r.v. such that $E(X) = \mu$ & $E((X-\mu)^2) = \sigma_X^2$. Then,

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}, \quad \varepsilon > 0$$

↳ Let $Y = |X-\mu|^2$ & then apply the Markov inequality to get the result.

Both Markov & Chebyshev inequalities are achievable, ie. there exist r.vs for which the upper bound is achievable.

The Chernoff bound

Let X be a r.v. s.t. $g_X(r)$ is finite for $r \in (r_-(X), r_+(X))$.

Then Chernoff bound is obtained by setting $Y = e^{rX}$ in the Markov inequality.

$$\therefore P(e^{rX} \geq e^{rx}) \leq \frac{g_X(r)}{e^{rx}} \quad \forall r \in (r_-(X), r_+(X))$$

Case I

if $r > 0$, then

$$P(X \geq x) \leq \frac{g_X(r)}{e^{rx}}$$

Case II

if $r < 0$, then

$$P(X \leq x) \leq \frac{g_X(r)}{e^{rx}}$$

Since r is not present in LHS of these inequalities,

$\Rightarrow$ We minimise the RHS w.r.t r for $r>0$ & $r<0$ for the 2 cases

$$P(X \geq x) \leq \inf_{r \in (0, r_+(X))} \frac{g_X(r)}{e^{rx}} \quad \Bigg\| \quad P(X \leq x) \leq \inf_{r \in (r_-(X), 0)} \frac{g_X(r)}{e^{rx}}$$

28/08 **Sum of iid r.vs.**

Let $X_1, \dots X_n$ be iid r.vs with mean $E(X)$ & an mgf $g_X(r)$. Let $(r_-(x), r_+(x))$ be its ROC. We wish to find bounds on $P(S_n \geq na)$ or $P(S_n \leq na)$, where a is a constant & $S_n = (X_1 + \dots X_n)$

↳ By Chernoff bound,

$$P\left(e^{rS_n} \geq e^{rna}\right) \leq \frac{g_{S_n}(r)}{e^{rna}} = \frac{\left(g_X(r)\right)^n}{e^{rna}} = \left(\frac{g_X(r)}{e^{ra}}\right)^n$$

30/

Case I

$$P(S_n \geq na) \leq \inf_{r \geq 0} \left(\frac{g_X(r)}{e^{ra}}\right)^n$$

Case II

$$P(S_n \leq na) \leq \inf_{r \leq 0} \left(\frac{g_X(r)}{e^{ra}}\right)^n$$

Let $\underbrace{\gamma_X(r)}_{\text{semi-invariant mgf.}} = \log g_X(r)$. Then $\gamma_X(0) = 0$ & convex.

min. is well defined & unique

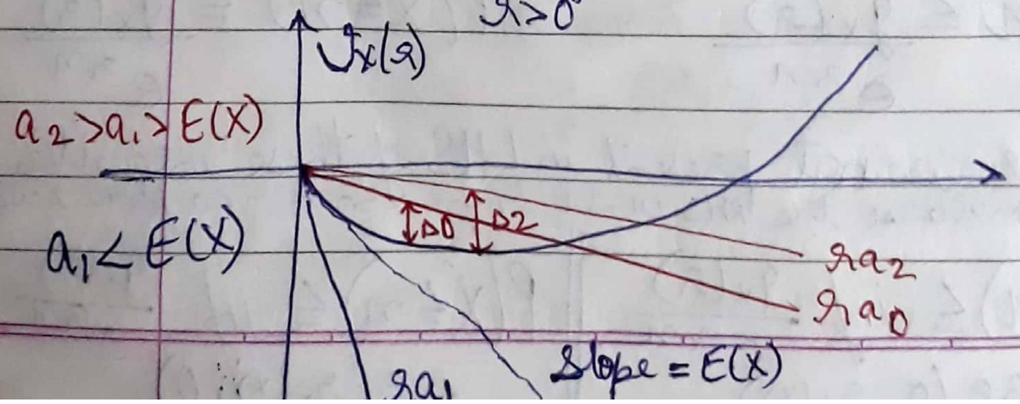
$\gamma_X'(0) = E(X)$

$\gamma_X''(r) \geq 0 \Rightarrow \gamma_X''(r)$ & $(\gamma_X''(r) - a)$ are both convex

Case I

$$P(S_n \geq na) \leq \inf_{r > 0} e^{n(\gamma_X(r) - ra)}$$

→ We want to make it as -ve as possible



Let $E(X) < 0$

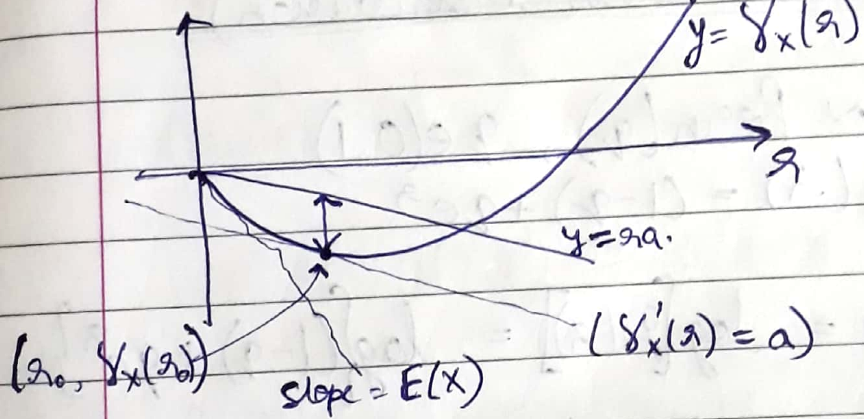
$\therefore P(S_n \geq na_1) \leq 1, \quad P(S_n \geq na_0) \leq e^{-n\Delta_0}$

The function $\gamma_X(r) - ra$ has a unique min. for $a > E(X)$ & $r > 0$; and the min. is strictly negative.

The minimum of $\gamma_X(r) - ra$ reduces further if a is increased (in the set $\{x : x > E(X)\}$)

0/08 $P(S_n \geq na) \leq \exp(n(\gamma_X(r) - ra))$

$\gamma_X(r) - ra$ is minimised at $\gamma_X'(r) = a$.



For $a < E(X)$, we can show that,
$$P(S_n \leq na) \leq \exp\{n(\gamma_X(r) - ra)\}$$
and the exponent $\gamma_X(r) - ra$ can be minimized for $r \in (r_-(X), 0)$

Let $\mu_X(a) = \inf_{r \in (r_-, r_+)} (\gamma_X(r) - ra)$

$\hookrightarrow$ 1 func^n only can be used, when $a < E(X)$, r_{min} lies in $r < 0$ & for $a > E(X)$, $r_{min} > 0$

Fact (Chernoff bound)

Let $X_1, \dots X_n$ be iid r.v. Let $S_n = X_1 + \dots X_n$ & $g_X(r)$ be the mgf of X.

Let $$M_X(a) := \inf_{r \in (r_-, r_+)} (\gamma_X(r) - ra)$$

Then, chernoff bound for sum of iid r.v says:-

for $a > E(X)$, $P(S_n \geq na) \leq \exp(-nM_X(a))$

& for $a < E(X)$, $P(S_n \leq na) \leq \exp(-nM_X(a))$

Example (Bernoulli Distribution)

Let $X \sim \text{Bern}(q)$, $q \in (0,1)$

$\therefore \; g_X(r) = (1-q) + qe^r$

$\therefore \; \gamma_X(r) = \log[g_X(r)] = \log\{(1-q) + qe^r\}$

3/09

$\therefore \; P(X_1 + \dots X_n \geq na) \leq \exp(nM_X(a))$

where $a > q \; (= E(X))$

$$\therefore M_X(a) = \inf_{r \in \mathbb{R}} \left[\log\left((1-q) + qe^r\right) - ra\right]$$

$\hookrightarrow \; \log\left(\dfrac{a(1-q)}{q(1-a)}\right) = r_0$

$$\therefore M_X(a) = -a \log \frac{a}{q} - (1-a) \log\left(\frac{1-a}{1-q}\right)$$

$\Rightarrow -D(a \| q)$, where $D(a\|q)$ is the K-L divergence (Kullback - Leibler) betn. the distrbn. $(a, 1-a)$ & $(q, 1-q)$. Unless $a = q$, $D(a\|q) > 0$

∴ We can conclude that,

$$P(S_n \geq na) \leq \exp\left(-n\left\{a\log\frac{a}{q} + (1-a)\log\left(\frac{1-a}{1-q}\right)\right\}\right)$$

if $a > q$, &

$$P(S_n \leq na) \leq \exp\left(-n\left\{a\log\frac{a}{q} + (1-a)\log\left(\frac{1-a}{1-q}\right)\right\}\right)$$

if $a < q$.

Jensen's inequality

Let $g: I \longrightarrow \mathbb{R}$ be a a convex function.
Assume that $P(X \in I) = 1$. Then,

$$E(g(X)) \geq g(E(X))$$

(I is interval on R)

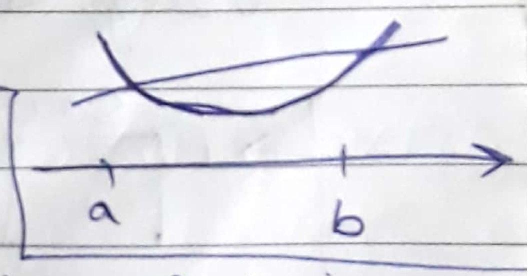
3/09

$g(x)$ is convex if $g''(x) \geq 0$

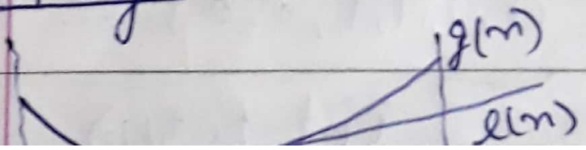
$g(x)$ is convex in I if for any $x_1, x_2 \in I$,

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2)$$

for all $\lambda \in [0,1]$



Proof:



Since $l(x)$ is a

By ~~convextity~~ convexity, $g(x) \geq l(x)$.

So, $E(g(X)) \geq E(l(X))$

$= l(E(X))$

$= g(E(X))$

Hence Proved.

↳ $l(x) = cx + d$, where (c, d) are some scalars.

So $l(X) = cX + d$

& $E(l(X)) = cE(X) + d = l(E(X))$

4/09

Using Jensen's inequality, one can show,

$$D(p\|q) := p \log \frac{p}{q} + (1-p) \log\left(\frac{1-p}{1-q}\right) \geq 0$$

↳ Let X & Y be 2 m-point discrete distributions, then

$$D(X\|Y) := \sum_{i=1}^{m} p_X(i) \log \frac{p_X(i)}{p_Y(i)}$$

KL divergence

Using Jensen's inequality, $D(X\|Y) \geq 0$

↳ Source → 0, 1, 1, 1, 0, 0, 1, ... $E(X) = p$

Bits

$\underbrace{0,1,1,1,0,0,1,\ldots}_{\text{iid Ber}(p)}$

p?

→ log liklihood

→ best fit

parameters → Model → stream of random bits

n-bits

Page No.

$\hat{p} = \frac{N_1}{n}$; p

$\mathbb{P}(N_1 \geq np) \leq \exp(-n.\overset{D(q\|p)}{0})$

$\mathbb{P}(N_1 \in (np - n\varepsilon, np + n\varepsilon))$

$= 1 - \mathbb{P}(N_1 \leq np - n\varepsilon) - \mathbb{P}(N_1 \geq np + n\varepsilon)$

$= 1 - \exp(-n D(p-\varepsilon \| p))$
$\quad - \exp(-n D(p+\varepsilon \| p))$

4/09 If X & Y are uncorrelated,

$E(XY) = E(X).E(Y)$

& $\underbrace{E((X-E(X))(Y-E(Y)))}_{\text{cov}(X,Y)} = 0$

Independence $\Longrightarrow$ Uncorrelated

↳ Let $X_1, \ldots X_n$ be iid r.v with mean 0 & variance σ^2

$\therefore \text{cov}(X_i, X_j) = \begin{cases} \sigma^2 & ; i = j \\ 0 & ; i \neq j \end{cases}$

or $\text{cov}(X_i, X_j) = \sigma^2 \delta[i-j]$

* $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2\,\text{cov}(X_1, X_2)$

$\therefore Var(X_1 + X_2) = \sigma_1^2 + \sigma_2^2 + 0 = 2\sigma^2$

* $Var(aX_1) = a^2 Var(X_1) = a^2\sigma^2$

* $Var(X_1 + X_2 + \dots X_n) = n\sigma^2$

$Var\left(\frac{X_1 + \dots X_n}{n}\right) = \frac{\sigma^2}{n}$

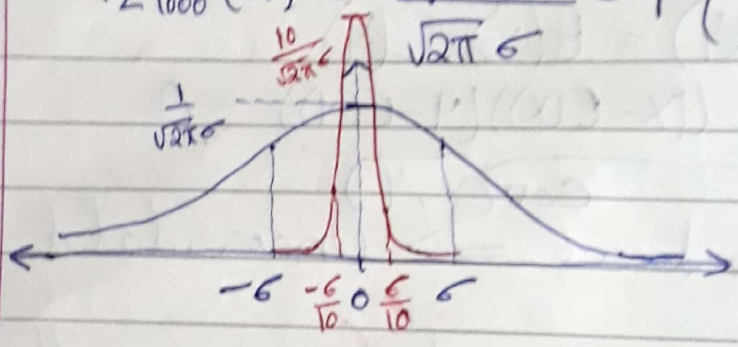
Let $S_n = X_1 + \dots X_n$, then $E(S_n) = 0$
& $Var(S_n) = n\sigma^2$

$\therefore Var\left(\frac{S_n}{n}\right) = \sigma^2/n$

Consider $N(0, \sigma^2/n) = Z_n$

$\therefore f_{Z_1}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-x^2}{2\sigma^2}\right)$

$f_{Z_{1000}}(x) = \frac{\sqrt{1000}}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{x^2 \cdot 1000}{2\sigma^2}\right)$



Weak law of large numbers (WLLN)

Let $X_1, \dots X_n$ be iid r.v with mean 0 & variance σ^2. Let $S_n = X_1 + \dots X_n$ & $\varepsilon > 0$ be a scalar,

Then, $P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$

Proof:- Since $Var\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$, by Chebychev inequality,

$$P\left(\left|\frac{S_n}{n} - \overset{E\left(\frac{S_n}{n}\right)}{0}\right| \geq \varepsilon\right) \leq \frac{Var\left(S_n/n\right)}{\varepsilon^2}$$

$$\frac{S_n}{n} \xrightarrow{P} 0, \text{ ie. } P\left(\left|\frac{S_n}{n} - 0\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty$$

$\Downarrow$

converges in probability

$\epsilon \to$ +ve,

$$0 \leq \lim_{n\to\infty} P\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon} = 0$$

By sandwich thm,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) = 0$$

$$0 \leq E\left(\left(\frac{S_n}{n} - \underbrace{E\left(\frac{S_n}{n}\right)}_{0}\right)^2\right) = Var\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$$

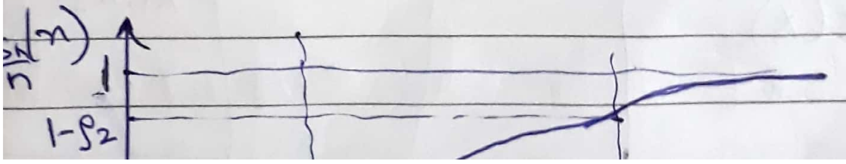
perfect square $\Rightarrow E(\ ) > 0$ since $\frac{S_n^2}{n^2} \geq 0$

Var. computⁿ.

$$\therefore \lim_{n\to\infty} E\left(\left(\frac{S_n}{n}\right)^2\right) = 0 \Rightarrow E\left(\frac{S_n}{n}\right) = 0$$

$\frac{S_n}{n} \to$ converges in mean square to zero.

ie. $\frac{S_n}{n}$ converges in mean-squared sense to $E\left(\frac{S_n}{n}\right) = 0$.



$n \to$ large

$\epsilon \to$ small, arbitary

Note that if $E(X) \neq 0$, then,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - E(X)\right| \geq \epsilon\right) = 0$$

& $$\lim_{n\to\infty} E\left(\left(\frac{S_n}{n} - E(X)\right)^2\right) = 0$$

$\hookrightarrow$ $\left|\frac{S_n}{n} - E(X)\right| \geq \epsilon$

$$\Rightarrow \left(\frac{S_n}{n} \geq E(X) + \epsilon\right) \cup \left(\frac{S_n}{n} \leq E(X) - \epsilon\right)$$

For any $\epsilon > 0$, as $n \to \infty$, $(\rho_1 + \rho_2) \to 0$

Convergence of R.V.

$$Z_n,\ n \in \mathbb{N} \xrightarrow{?} \underbrace{Z}_{\text{limiting r.v.}}$$

$g(t) = g(t+T)$;

$$\lim_{N\to\infty} \sum_{k=-N}^{N} a_k e^{\frac{j2\pi kt}{T}} \xrightarrow[\text{mean square}]{\text{point wise or}} g(t)$$

$$a_k = \left< g(t), e^{\frac{-j2\pi kt}{T}} \right>$$

Analysis	Emperical
X is iid. $E(X)$	$X_1, X_2 \ldots X_n$
model	data
$E(X)$	$\sum_{i=1}^{n} x_i / n$ empirical [illegible]

Def^n: Convergence in L^p

We say that $Z_n \xrightarrow{L^p} Z$ if.

$$\lim_{n\to\infty} E(|Z_n - Z|^p) = 0$$

→ Z_n, $n \in \mathbb{N}$ is a seqn. of r.v.s; Z is some fixed r.v.

ex: → $\frac{S_n}{n} \xrightarrow{L^2} E(X)$ in the WLLN setup.

Defn: Convergence in prob.

$Z_n \xrightarrow{P} Z$ if for any $\epsilon > 0$.

$$\lim_{n\to\infty} P(|Z_n - Z| \geq \epsilon) = 0$$

Defn: Convergence in almost sure sense

$Z_n \xrightarrow{a.s.} Z$ if $\mathbb{P}(\lim_{n\to\infty} Z_n = Z) = 1$

converges almost surely ($Z_n \to Z$ w.p 1)

diff^n from convergence in P

$$\mathbb{P}\Big(\underbrace{\omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)}\Big) = 0$$

$$\mathbb{P}(\lim_{n\to\infty} Z_n = Z) = 0$$

→ For $b_n \in \mathbb{R}$, we say that $b_n \to b$ for any $\epsilon > 0$, there exist an $n = n_0$ (b_n, ϵ) s.t. $|b_n - b| < \epsilon$ $\forall$ $n \geq n_0$

i.e. for any $\epsilon > 0$, $|b_n - b| < \epsilon$ eventually.

17/09 $\hookrightarrow b_n = 1/n$, $b_n \to 0$

$1/n < \epsilon$ eventually.

x

$\omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)$

$\therefore$ for any $\epsilon > 0$, $\{\omega : |Z_n(\omega) - Z(\omega)| < \epsilon$ eventually$\}$

$\hookrightarrow |Z_n - Z| < \epsilon$ eventually w.p. 1

$\therefore$ For any $\epsilon > 0$,

$$\underset{n\to\infty}{\text{limit}}\ \mathbb{P}\left(\bigcap_{m\geq n} |Z_m - Z| < \epsilon\right) = 1$$

or, for any $\epsilon > 0$,

$$\underset{n\to\infty}{\text{limit}}\ \mathbb{P}\left(\bigcup_{m\geq n} |Z_m - Z| > \epsilon\right) = 0$$

$\hookrightarrow \lim_{n\to\infty} \mathbb{P}(|Z_n - Z| > \epsilon) = 0$, $\boxed{Z_n \xrightarrow{P} Z}$

Defn: Convergence in distribution $\Rightarrow$ converges weakly

$$Z_n \xrightarrow{d} Z \text{ if } F_{Z_n}(z) \longrightarrow F_Z(z)$$

at all pts where $F_Z(z)$ is continuous

Fact: (The Union bound)

Let $X_1 \ldots X_n$ be any r.v. Let $(X_i \in \mathcal{S}_i)$ be events

Then,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} X_i \in \mathcal{S}_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(X_i \in \mathcal{S}_i)$$

For a.s. convergence, we want,

for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m\geq n} |Z_m - Z| > \epsilon \right) = 0$

↳ A sufficient condition for a.s conv. is

for any $\epsilon > 0$, $\lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(|Z_m - Z| > \epsilon) = 0$

Dependancies of convergence

$$\begin{tikzcd}
\text{in } L^p \arrow[dr, Rightarrow] \arrow[dr, bend left, "\text{markov ineq. on } |Z_n - Z|^p > \epsilon"] \arrow[dd, Leftrightarrow, "\times"'] & & \\
& \text{in } \mathbb{P} \arrow[r, Rightarrow] \arrow[r, bend right, "\text{H.W quest}^n"'] & \text{in distrib}^n. \\
\text{in a.S.} \arrow[ur, Rightarrow] \arrow[ur, bend right, "\text{Subset ineq.}"'] & &
\end{tikzcd}$$

$\Downarrow$

$$\mathbb{P}\left(\left\{ \bigcup_{m\geq n} |Z_m - Z| > \epsilon \right\}\right) \geq \mathbb{P}(|Z_n - Z| > \epsilon)$$

<u>ex:</u> i) $\xrightarrow{L^p} \not\Rightarrow \xrightarrow{a.s}$

Counter ex. $\Rightarrow$ of convergence dependancies

Let $Z = 0$ & $Z_n = 0$ w.p. $\left(1 - \frac{1}{n}\right)$

$= 1$ w.p. $\frac{1}{n}$

Then, $E(|Z_n - Z|^p) = \frac{1}{n} \longrightarrow 0$ as $n \to \infty$

or, $Z_n \xrightarrow{L^p} Z$