(r.v.s sequence ~ limiting r.v.
seq. of functions.

Defn (convergence in P) :- for any $\varepsilon > 0$, if $Z_n \xrightarrow{P} Z$, then

$$\lim_{n \to \infty} P(|Z_n - Z| > \varepsilon) = 0.$$

$\hookrightarrow$ $b_n = \frac{1}{n}$, $b_n \to 0$.

$\frac{1}{n} < \varepsilon$ eventually.

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

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

$\Rightarrow$ $|Z_n - Z| < \varepsilon$ eventually with probability 1.

* for any ε,

$$\lim_{n \to \infty} P\left(\bigcap_{m \geq n} |Z_m - Z| < \varepsilon\right) = 1.$$

$|Z_n - Z| < \varepsilon$ AND $|Z_{n+1} - Z| < \varepsilon$ AND $|Z_{n+2} - Z| < \varepsilon$.

* or for any $\varepsilon > 0$;

$$\lim_{n \to \infty} P\left(\bigcup_{m \geq n} |Z_m - Z| > \varepsilon\right) = 0.$$

(Here we have OR, not AND as above. If any 1 satisfies, breaks. should never satisfy.)

almost sure implies convergence in probability.
harder. simpler.

Page No.
Date

Defn (convergence in distribution) [weak convergence]

$Z_n \xrightarrow{d} Z$ if $F_{Z_n}(z) \longrightarrow F_Z(z)$ at all points where $F_Z(z)$ is continuous.

— x —

* - For conv. in probability → for a large n. satisfies.
  - for almost sure → for a large n, and every value of n beyond that.

* fact (The union bound)

Let $X_1, X_2, \ldots X_n$ be any r.v.s. let $(X_i \in g_i)$ be events.

Then, $P\left(\bigcup_{i=1}^{\infty} X_i \in g_i\right) \leq \sum_{i=1}^{\infty} P(X_i \in g_i)$ (3rd Axiom)

* For a.s. convergence, we want

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

→ A sufficient condn for almost s. conv. is for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} P\left(|Z_m - Z| > \varepsilon\right) = 0$$

* Dependencies of convergence

$\mathcal{L}^2$ (mean-sq) conv.

$\mathcal{L}$ (absolute) conv.

in $\mathcal{L}^p$ $\not\Leftrightarrow$ in a.s. (with prob 1)

subset inequality.

Markov ineq. on $|Z_n - Z|^p > \varepsilon$.

$\Downarrow$ $\Downarrow$

in P.

$\Downarrow$ } ⇝ hw question

in distribution. $\xrightarrow{d}$

(weak convergence)

consistent (consistency)

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

Example (counter examples of convergence dependencies)

i) $\xrightarrow{L^p} \not\Rightarrow \xrightarrow{a.s.}$ $(p \geq 1)$.

~~Let~~

~~Let $Z=0$. And $Z_n = 0$ wp $\left(1-\frac{1}{n}\right)$.~~

~~$Z_n = 1$ wp $\frac{1}{n}$.~~

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

~~Or $Z_n \xrightarrow{L^p} Z$.~~

Scrap the above ↑ ×.

Example :-

4	5	6	7
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1.
Z_4	Z_5	Z_6	Z_7.

$p = 1/4$.

1 2 3 4 5 6 7 8 9 10 14 15

let $Z_i = 1$, & $Z_j = 0$ for $2^k \leq i, j \leq 2^{k+1} - 1$ in equiprobable fashion. $i \neq j$.

If $2^k \leq i \leq 2^{k+1}$

then

$P(Z_i = 1) = \frac{1}{2^k}$.

Show that

$Z_n \xrightarrow{L^p} 0$

Since $E(|Z_n|^p) = P(Z_n = 1) = \frac{1}{2^k} \leq \frac{2}{n}$

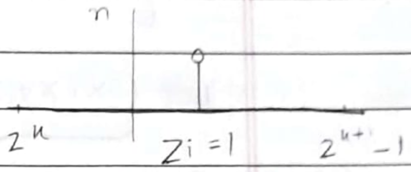
where $2^k \leq n < 2^{k+1}$

So, $\lim_{n\to\infty} E(|Z_n|^p) = 0$, or $Z_n \xrightarrow{L^p} 0$.

Consequently, $Z_n \xrightarrow{P} 0$

However, $Z_n \not\xrightarrow{a.s.} 0$ since for any $\varepsilon < 1$,

$$P\left(\bigcup_{m \geq n} |Z_m - 0| > \varepsilon \right) = 1.$$



Here, not converging a.s., as no matter what n_0 we choose, there will be some Z_is for which $P(\ ) \neq 0$. as for a.s, we need convergence for any $n > n_0$. Random sequence should not deviate again, even at a single point.

* in a.s. $\not\Rightarrow$ in $\mathcal{L}^p$

Let $Z_n = 0$ with prob $\left(1 - \frac{1}{n^2}\right)$

$= n$ with prob $1/n^2$.

Then $E(|Z_n|^2) = n^2 \cdot 1/n^2 = 1$.

$E(|Z_n|^3) = n^3 / n^2 = n$

$\therefore Z_n \not\xrightarrow{\mathcal{L}^p} 0$ for all $p \geq 2$.

However, for any $\varepsilon > 0$

$$P\left(\bigcup_{m \geq n} |Z_n| > \varepsilon \right) \leq \sum_{m=n}^{\infty} P(|Z_n| > \varepsilon).$$

$$= \sum_{m=n}^{\infty} \frac{1}{m^2} \longrightarrow 0 \text{ as } n \to \infty. \qquad \left(Z_n \xrightarrow{a.s.} 0\right)$$

— x —

If bounded, & converging in prob, normally converge in $\mathcal{L}^p$ as well

Also, from above, conv in $\mathbb{P} \not\Rightarrow$ in $\mathcal{L}^p$.

* **Example (data compression example)**

$L \times W \times 3 \times 8$

$1920 \times 1080 \times 3 \times 8 = 6\text{MB}.$

Let $\underbrace{(X_1 X_2 X_3 \cdots X_n)}_{\text{binary string.}}$ be a file.

Let $X_i \sim Ber(p)$, where p is known and $X_1, \ldots X_n$ are iid.

Let $p_X(1) = p(1) = p$. & $p_X(0) = p(0) = (1-p)$.

Consider the function. $\log_2 p(X)$

Then $\log_2 p(X) = \log_2(p)$; wp. p

$= \log_2(1-p)$; wp. $(1-p)$

By WLLN;

$$\frac{1}{n}\sum_{i=1}^{n} \log_2(p(X_i)) \approx E(\log_2 p(X))$$

$$= p\log_2 p + (1-p)\log_2(1-p)$$

$H_2(p)$.

$$= \underbrace{-H_2(p)}$$ Entropy of $X \sim Ber(p)$ [entropy always +ve

Z (degenerate r.v.) $\equiv C$ takes it as a $p(c) = 1$.

where, $H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$. (+ve)

By ~~WLLN~~ WLLN, for any $\varepsilon, \delta > 0$,

$$P\left(\left|\frac{\sum_{i=1}^{n} \log_2 p(X_i)}{n} + H_2(p)\right| > \varepsilon\right) \leq \delta.$$

If n is large enough.

I.e.,

$$-H_2(p) - \varepsilon \le \frac{1}{n}\sum_{i=1}^{n} \log_2 p(x_i) \le -H_2(p) + \varepsilon.$$

with prob. $> 1-\delta$.

$$\Rightarrow \quad -H_2(p) - \varepsilon \le \frac{1}{n}\log_2 p(x_1, \ldots, x_n) \le -H_2(p) + \varepsilon, \text{ w.p. } > 1-\delta$$

or, $$2^{-n(H_2(p)+\varepsilon)} \le p(x_1, \ldots, x_n) \le 2^{-n(H_2(p)-\varepsilon)}$$

with prob $> (1-\delta)$.

Define: $$A_{n,\varepsilon} := \left\{ \vec{x} : \left| \frac{1}{n}\log p(x_1, \ldots, x_n) + H_2(p) \right| < \varepsilon \right\}.$$

typical set.

↳ Then $A^c_{n,\varepsilon}$ = atypical set. (Something which not happen normally)

$$H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

$A_{n,\varepsilon}$ is a collection of deterministic strings.

Now, WLLN says that

$$P\left((x_1, \ldots x_n) \in A_{n,\varepsilon}\right) > 1-\delta.$$

On the typical set,

$$2^{-n(H_2(p)+\varepsilon)} \le p(x_1, \ldots, x_n) \le 2^{-n(H_2(p)-\varepsilon)}$$

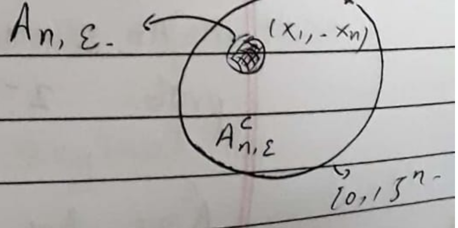
or $$p(x_1, \ldots, x_n) \approx 2^{-nH_2(p)} \quad \text{(ignoring } \varepsilon\text{)}.$$

for any $(x_1, \ldots x_n) \in A_{n,\varepsilon}$.

? ↳ So $|A_{n,\varepsilon}| \approx 2^{nH_2(p)}$ and $P((x_1, \ldots x_n) \in A^c_{n,\varepsilon}) \approx 0$.

~~for an~~ A set with $|A| = n$,

any probability $\to \frac{1}{n}$ of a particular element.



Page No.
Date

Given $(x_1, \dots x_n)$, check if $(x_1, \dots x_n) \in A_{n,\varepsilon}$.

If no, retain $(1, x_1, \dots, x_n)$ (1, 0 a flag which determines whether in typical set or not)
If yes, retain $(0, x_1 \dots x_n)$.

$|A_{n,\varepsilon}| \approx 2^{nH_2(p)}$ } reindex using $nH_2(p)$ bits. use lookup table to get a compressed representation for a typical bit string.

Expected length after compression is

$$\frac{\delta(n+1) + (1-\delta)(nH_2(p)+1)}{n}$$

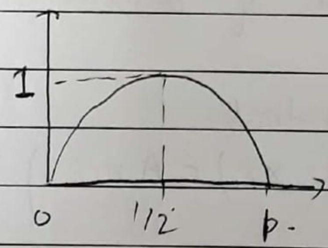
8-string
0000 0001
0000 0010
0000 0100
0000 1000
1000 0000

LUT
000
001
010
011
111

$\approx H_2(p)$

per bit or per kb compression.

$H_2(p)$



$\therefore$ farther away from 1/2, better compression.

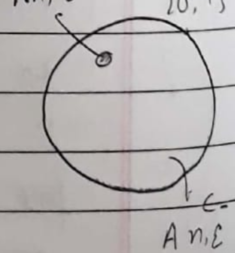
New day

Example. (WLLN and compression).

$(X_1, \dots, X_n)$; file.

X_i's are iid $Ber(p)$

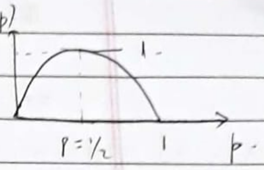
If n is large, then $\approx 2^{nH_2(p)}$ strings in $A_{n,\varepsilon}$ each nearly equiprobable with prob. $2^{-nH_2(p)}$.



$A^c_{n,\varepsilon}$ has negligible prob. collectively.

Expected length of compressed file $= (1 + nH_2(p))(1-\delta) + \delta(n+1)$.

$\frac{1}{n}$ Expected length $\approx H_2(p)$.



$$H_2(p) = -p\log p - (1-p)\log_2(1-p)$$

If $p = \frac{1}{2}$, the file is not compressible.

$$p_{X_1, X_2, X_3 \ldots X_n}(x_1, \ldots, x_n) = \left(\frac{1}{2}\right)^n = 2^{-n}$$ (0 & 1 equiprobable & all independent)

every string is equiprobable.

~~$p_{X_1, \ldots, X_n}(x_1, \ldots x_n)$~~

and $|\{0,1\}^n| = 2^n$. Entire set $\{0,1\}^n$ is typical. nothing atypical here.

— x —

Example: (cauchy distribution's expectation)

Recall that $X = X^+ - X^-$ and $E(X) = E(X^+) - E(X^-)$ exist if $E(X^+) < \infty$ and $E(X^-) < \infty$.

→ If X is Cauchy distributed then

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}; \quad x \in \mathbb{R}.$$

↳ $E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\,dx = 0$. However $E(X^+) = \infty$ and $E(X^-) = \infty$.

Let $X_1, \ldots X_n$ be iid Cauchy rvs with pdf $f_X(x)$.

↳ Then $\phi_X(t) = e^{-|t|}$, $t \in \mathbb{R}$.

→ From duality of the FT this result follows:-

$$y(t) \xrightarrow{\mathcal{F}} \tilde{y}(\omega) \iff \tilde{y}(t) \xrightarrow{\mathcal{F}} 2\pi y(-\omega)$$

$$f(x) = e^{-x}u(x) \xrightarrow{\phi} \frac{1}{1+jt}, \quad f(-x) = e^{x}u(-x) \xrightarrow{\phi} \frac{1}{1+jt}$$

$$\phi(t) = \int_{-\infty}^{\infty} e^{jtx} e^{-x} u(x)\, dx.$$

$$\Rightarrow \frac{1}{2}\left(\phi_1 + \phi_2\right) = \frac{1}{1+t^2}$$ ($\frac{1}{2}$ so prob. add up to 1)

The characteristic func. should have $\phi(0)=1$. Also $\int f(x)dx$. check these 2 while applying duality.

→ $\frac{X_1 + \cdots + X_n}{n}$ has a pdf

$$f_{\frac{X_1+X_2+\cdots X_n}{n}}(x) = \frac{1}{\pi}\frac{1}{(1+x^2)} ; x \in \mathbb{R}.$$

Since $\phi_{\frac{X_1+\cdots X_n}{n}}(t) = \phi_{X_1}\left(\frac{t}{n}\right)\phi_{X_2}\left(\frac{t}{n}\right) \cdots \phi_{X_n}\left(\frac{t}{n}\right)$

$$= e^{-|t|}, \quad t \in \mathbb{R}.$$

∴ averaging out doesn't change the distribution.
$\phi(t)$ & $f(x)$ are unique pairs having 1, can get the other deterministically.

If 2 signals have same FT, signals have to be the same.

→ if $E(X) < \infty$, WLLN tells us that

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} E(X)$$

→ However $f_{\frac{X_1+\cdots X_n}{n}}(x)$ does not show this concentration.

→ I.e. $E(X)=0$ is meaningless in the light of WLLN

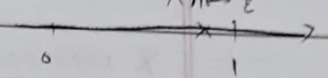
— x —

Example (convergence of $X_{n:n}$)

let $X_1, \dots X_n$ be iid Unif$(0,1)$
let $X_{n:n} = \max\{X_1, \dots X_n\}$; $X_{1:n} = \min\{X_1, \dots X_n\}$

and $X_{r:n}$ is the r^{th} largest in the set $\{X_1, \dots X_n\}$.
order statistics. we will show that

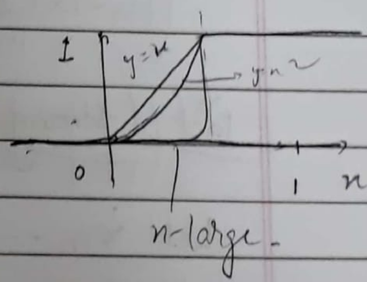
$X_{n:n} \xrightarrow{a.s.} 1$.



If n is large, the max value better be close to 1.

Since $F_X(x) = x$; $0 \le x \le 1$.
$= 0$, $x < 0$
$= 1$, $x \ge 1$.

So $F_{X_{n:n}}(x) = x^n$; $0 \le x \le 1$.
$= 0$; $x \le 0$
$= 1$; $x \ge 1$.



We want to show that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\bigcup_{m \ge n} |X_{m:m} - 1| > \varepsilon \right) = 0.$$

$P(|X_{n:n} - 1| > \varepsilon)$
$= P(X_{n:n} < 1-\varepsilon)$
$= (1-\varepsilon)^n$. ($\to 0$ as $n\to\infty$)

i.e., $X_{n:n} \xrightarrow{P} 1$.

For a.s. convergence, note that,

For a.s. convergence note that.

$$0 \le P\left(\bigcup_{m\ge n} |X_{m:m}-1|>\varepsilon\right)$$

$$\le \sum_{m\ge n} P(|X_{m:m}-1|>\varepsilon) = \sum (1-\varepsilon)^m$$

as $n\to\infty$.

↳ So, $X_{n:n} \xrightarrow{a.s.} 1$

$X_{1000:1000}$.

$X_{10000:10000}$.

— x —

$$1-X_{n+1:n+1} \le 1-X_{n:n}.$$

and $X_{n:n}\xrightarrow{P}1$, it follows that $X_{n:n}\xrightarrow{a.s.}1$

→ or $\left\{\bigcup_{m\ge n}|X_{m:m}-1|>\varepsilon\right\} = \left\{|X_{n:n}-1|>\varepsilon\right\}$.

— x — as any way it shrinks, the smallest one covers everything.

Fact (Strong law of large numbers)

Let $X_1, \dots, X_n$ be iid rv with mean $E(X)$. Then,

$$\frac{X_1+\dots+X_n}{n} \xrightarrow{a.s.} E(X)$$

— x —

Fact (Strong law of large numbers (SLLN))

Let $X_1, \dots X_n$ be iid r.v. with mean $E(X)$. Then

$$\frac{X_1+\dots+X_n}{n} \xrightarrow{a.s.} E(X)$$

— x — Here, we assumed nothing abt variance. Also, in WLLN, it was conv. in prob, Here a.s.

Fact: (central limit Theorem).

let $X_1, \dots, X_n$ be iid rvs such that $E(X) = \mu$ and $Var(X) = \sigma^2$.

let $Y_n := \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n}\,\sigma}$. Then

$$Y_n \xrightarrow{d} N(0,1).$$

— x —

Proof (sketch):

Tool: Characteristic functions.

Fact: ( ) i) If $Z_n \xrightarrow{d} Z$, then $\phi_{Z_n}(t) \longrightarrow \phi_Z(t)$. for all $t \in \mathbb{R}$.

ii) If $\phi_{Z_n}(t) \longrightarrow \phi_Z(t)$ and $\phi_Z(t)$ is continuous at $t=0$, then $Z_n \xrightarrow{d} Z$ (levy continuity theorem).

let $\phi_X(t)$ be the charact. func. of $X_1, \dots, X_n$.

Then. $\phi_{Y_n}(t) = E\left(e^{jtY_n}\right)$

$$= E\left(e^{\left(jt\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu)\right)}\right)$$

$$= \prod_{i=1}^{n} \phi_{X-\mu}\left(\frac{t}{\sqrt{n}\,\sigma}\right) = \left[\phi_{X-\mu}\left(\frac{t}{\sqrt{n}\,\sigma}\right)\right]^n.$$

Observe that

$$\phi_{X-\mu}(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

$$\boxed{\phi_X(t) = E\left(e^{jtX}\right) = E\left(1 + jtX + j^2\frac{t^2X^2}{2!} + j^3\frac{t^3X^3}{3!} + \dots\right)}$$

$$\underbrace{\lim_{t \to 0} \frac{o(t^2)}{t^2} = 0}_{\text{defn of.}}$$, as $E(X-\mu) = 0$, no linear term

So, $\phi_{V_n}(t) = \left(1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$

$\longrightarrow \underbrace{e^{-t^2/2}}$ as $n \to \infty$.

So. $V_n \xrightarrow{d} N(0,1)$, $\phi_{N(0,1)}(t)$.

Let $c_n \in \mathbb{C}$ and $c_n \to c$.

Then, $\left(1 + \frac{c_n}{n}\right)^n \to e^c$, in $\mathbb{C}$

$V_n = \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} (X_i - \mu)$; $\phi_{V_n}(t) \longrightarrow e^{-t^2/2}$.

for every $t \in \mathbb{R}$.

So $V_n \xrightarrow{d} Z$ where $Z \sim N(0,1)$.

$F_{V_n}(x) \longrightarrow F_Z(x)$ at all pts $x \in \mathbb{R}$ where $F_Z(x)$ is continuous.

i.e. $F_{V_n}(x) \longrightarrow F_Z(x)$ for all $x \in \mathbb{R}$

(since the latter is continuous)

Example (denoising & quantization)

noise there is avged out if exposure time is increased

ISO speed → determines the exposure time req. to get a neutral image.

$c \longrightarrow \oplus \rightarrow \boxed{\mathbb{1}(\cdot)} \longrightarrow Y.$

(constant) to learn, estimate; $\oplus$ ← W (noise); $\mathbb{1}(\cdot)$: one-bit quantizer (measurement).

In this signal acquisition model,

$$Y = \mathbb{1}(c + W \geq 0) = \begin{cases} 1 & \text{if } c + W \geq 0 \\ 0 & \text{if } c + W < 0 \end{cases}.$$

$Y = [1 + \mathrm{sgn}(c+W)]/2.$ (Observe)

↳ If W is unknown, c can't be inferred (accurately) from Y.

↳ Consider 'n' indep. readings

$Y_i = 1\ (c + W_i \geq 0)$

where $W_1, \ldots, W_n$ are indep. and identically distribution.

↳ Observe that Y_i are also iid ~~and~~ (Think why??)

$Y \sim \mathrm{Ber}(\underbrace{1 - F_W(-c)}_{P(W+c \geq 0)})$.

If W is a continuous r.v.

$(1 - F_W(-c))$ is invertible (requires $F_W(u)$) in c whenever

since the function is monotonic

$f_W(-c) > 0$.

Let W be a sym...

i.e. ...

$$Y = [1 + \text{sgn}(c+w)]/2.$$ (Observe)

↳ If w is unknown, c can't be inferred (accurately) from Y.

↳ Consider 'n' indep. readings
$$Y_i = 1\ (c + w_i \geq 0)$$
where $w_1, \ldots, w_n$ are indep. and identically distribution.

↳ Observe that Y_i are also iid ~~and~~ (Think why?)
$$Y \sim \text{Ber}(\underbrace{1 - F_w(-c)}_{P(w+c \geq 0)})$$
If w is a continuous rv.

$(1 - F_w(-c))$ is invertible (requires $F_w(u)$) in c since the function is monotonic whenever $f_w(-c) > 0$.

Let w be a symmetric rv.,
i.e. $f_w(x) = f_w(-x)$.

Then $Y \sim \text{Ber}(F_w(c))$.

→ let $S_n = Y_1 + \ldots + Y_n$. Then
$$\frac{S_n}{n} \xrightarrow{\mathcal{L}^2} E(Y) = F_w(c).$$ (something with weak)

and $\text{var}\left(\frac{S_n}{n}\right) = \frac{1}{n}\text{var}(Y) = \frac{1}{n}F_w(c)(1 - F_w(c)) \leq \frac{1}{n}\cdot\frac{1}{4}$

→ Let $\hat{c} = F_w^{-1}\left(\frac{S_n}{n}\right)$

where $F_w(u)$ is known.