Probability

05.09.05

~~WE~~

Borel - Cantelli lemma

(i) $\sum_n P(A_n) < \infty \Rightarrow P(A_n \text{ i.o.}) = 0$.

(ii) $\langle A_n \rangle$ ~~are~~ indep, $\sum_n P(A_n) = \infty \Rightarrow$

$P(A_n \text{ i.o.}) = 1$.

$$\langle \omega : A_n \text{ i.o.} \rangle = \langle \omega : \sum_n I_{A_n} = \infty \rangle$$

Pf. (i) $$\sum_n P(A_n) = \lim_{N \uparrow \infty} \sum_{n=1}^{N} P(A_n)$$

$$= \lim_{N \uparrow \infty} \sum_{n=1}^{N} E[I_{A_n}]$$

$$= \lim_{N \uparrow \infty} E\left[\sum_{n=1}^{N} I_{A_n}\right] \underbrace{= E\left[\sum_{n=1}^{\infty} I_{A_n}\right]}_{\text{MCT}} <$$

$$\therefore \sum_{n=1}^{\infty} I_{A_n} < \infty \text{ a.s.}$$

(ii) ~~Need~~ Need to Show:

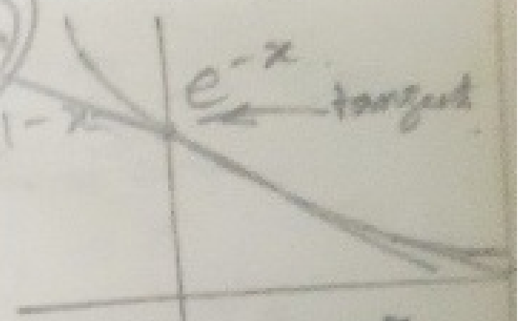
$$e^{-\sum_{n=1}^{\infty} I_{A_n}} = 0 \text{ a.s.}$$

i.e. $$\lim_{N \uparrow \infty} e^{-\sum_{n=1}^{N} I_{A_n}} = 0 \text{ a.s.}$$

$$E\left[e^{-\sum_{n=1}^{\infty} I_{A_n}}\right] = \lim_{N\uparrow\infty} E\left[e^{-\sum_{n=1}^{N} I_{A_n}}\right] \quad \text{— (MCT)}$$

$$= \lim_{N\uparrow\infty} \prod_{n=1}^{N} E\left[e^{-I_{A_n}}\right] = \lim_{N\uparrow\infty} \prod_{n=1}^{N} \left(1\cdot(1-P(A_n)) + e^{-1}\cdot P(A_n)\right)$$

$$\cancel{\lim_{N\uparrow\infty}} = \lim_{N\uparrow\infty} \prod_{n=1}^{N} \left(1-(1-e^{-1})P(A_n)\right)$$



$$\le \lim_{N\uparrow\infty} \prod_{n=1}^{N} e^{-(1-e^{-1})P(A_n)} \qquad \because 1-x \le e^{-x}$$

$$= \lim_{N\uparrow\infty} e^{-(1-e^{-1})\sum_{n=1}^{N} P(A_n)}$$

$$= 0 \qquad \left(\because \sum_n P(A) = \infty\right).$$

$A_n = A \ \forall n \quad 1 > P(A) > 0 \Rightarrow \sum P(A_n) = \infty$

$P(A_n \text{ i.o.}) = P(A) < 1$

$A_\alpha, \alpha \in I$ indep if every finite subfamily is indep.

pairwise indep if A_α, A_β indep for $\alpha \neq \beta$. independence $\Rightarrow$ pairwise indep
$\not\Leftarrow$

Ex. X, Y, i.i.d. $P(X=1) = P(X=-1) = \frac{1}{2}$
$Z = XY$, (X, Y, Z) pairwise indep.
but not dep.

$P(X=1, Z=1) = P(X=1, Y=1) = P(X=1)P(Y=1)$

$= \frac{1}{4}$

$= P(X=1)P(Z=1)$

$P(Z=1) = P(X=1, Y=1) + P(X=-1, Y=-1)$

$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(Z=-1) = \frac{1}{2}$.

$P(X=1, Y=1, Z=-1) = 0 \neq \frac{1}{8}$?

Z is a function of X & Y

Can't be independent.

$X_n \to X$ in probability, then $\exists \langle X_{n(k)} \rangle$ s.t. $X_{n(k)} \to X$ a.s.

Cor. $X_n \to X$ in q-th mean, $q \geq 1 \Rightarrow$

$\exists \langle X_{n(k)} \rangle$ s.t. $X_{n(k)} \to X$ a.s.

Pf $P(|X_n - X| \geq \epsilon) \to 0$ as $n \uparrow \infty$, $\forall \epsilon > 0$

Let $\epsilon = \frac{1}{2^m}$ $P\left(|X_n - X| \geq \frac{1}{2^m}\right) \to 0$

as $n \to \infty$

$\exists\, n(m) > n(m-1)$ s.t.
$P(|X_{n(m)} - X| \geq \frac{1}{2^m}) < \frac{1}{2^m}$.

$m=1$ Pick $n(1)$ s.t.
$P(|X_{n(1)} - X| \geq \frac{1}{2}) < \frac{1}{2}$. [if $n(1)$ large enough then $P \to 0$]

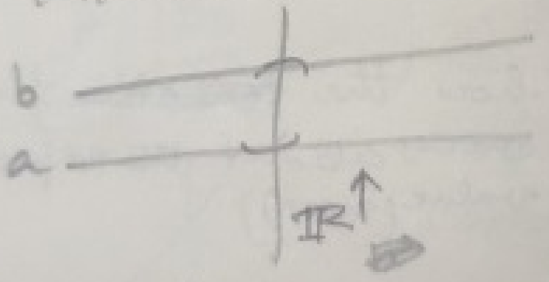
$$\sum_m P(|X_{n(m)} - X| \geq \tfrac{1}{2^m}) < \sum_m \tfrac{1}{2^m} < \infty$$

$\therefore P(|X_{n(m)} - X| \geq \frac{1}{2^m}$ i.o.$) = 0$.

$\therefore |X_{n(m)} - X| < \frac{1}{2^m}$ for m sufficiently large, a.s.

$\therefore X_{n(m)} \to X$ a.s.

X_n i.i.d., not constant. a.s.



$b > a$
s.t. $P(X_n \geq b) > 0$
$P(X_n < a) > 0$

$\Rightarrow$

$P(X_n$ converges$) = 0$.

$P(X_n \geq b$ i.o.$) = 1 = P(X_n \leq a$ i.o.$)$

Limit Theorems

$S_0 = 0,$

$S_n = \sum_{i=1}^{n} X_i,$

$\langle X_i \rangle$ i.i.d.

random walk. average $= 0$ here

---

(1) 'Typical 'average' behaviour

(2) 'fluctuations about the average'.

(3) 'upper & lower envelopes'.

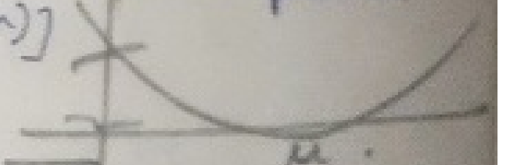
(4) 'delay of rare event probabilities'.

① $\frac{S_n}{n}$ Normalizations. (it will tell us that after normalization $\frac{1}{n}$ it converges

weak Laws of Large Numbers { WLLN → ~~tells~~ $\frac{S_n}{n}$ goes to expectation in prob.

SLLN → " " " a.s.

② $n \cdot \frac{S_n - E[X]}{\sqrt{n}}$ CLT. [how the ~~fluctuation~~ fluction varies about the expected value (mean)]



③ $\frac{S_n - nE[X_1]}{\sqrt{2n\log\log n}}$ LIL (lim sup converges to something) lim inf " "

④ $P\left(\frac{S_n}{n} \in A\right) \propto e^{-n \inf_A I}$ (telling about the dominant rate)

$$\lim_{N\to\infty} \frac{1}{n} \log \sum_{i=1}^{N} e^{-\lambda_i n} = -\min_i \lambda_i$$

$\lambda_i > 0$

Doob - Meyer

large deviations.

WLLN: $\langle X_i \rangle$ i.i.d., $E[X_i] = 0$. (in general, replace $X_i - E[X_i]$)

Let's assume $E[X_i^2] < \infty$.

Claim $\frac{S_n}{n} \to 0$ in Probability.

$\left[\frac{S_n}{n} \to E[X_i] \text{ in prob} \right]$

$$P\left(\left| \frac{S_n}{n} \right| \geq \epsilon \right) \leq \frac{E\left[S_n^2 / n^2 \right]}{\epsilon^2} = \frac{n E[X_i^2]}{n^2 \epsilon^2} \to 0.$$

Chebyshev.

$$E[S_n^2] = \sum_i E[X_i^2] + 2 \sum_{j<i} E[X_i X_j]$$

$\sum_i E[X_i^2] = n E[X_i^2]$

$E[X_i X_j] = E[X_i] E[X_j] = 0$

Remark: Pairwise independence enough
u.i. enough. (we can always chop off)

SLLN: $\langle X_i \rangle$ i.i.d.,
$E[X_i] = 0$, $E[X_i^2] < \infty$.

$$\frac{S_{n^2}}{n^2} \to 0 \text{ a.s.}$$

$$P\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \epsilon\right) \leq \frac{E[S_{n^2}^2]}{n^4 \epsilon^2} = \frac{n^2 K}{n^4 \epsilon^2} = \frac{K}{n^2 \epsilon^2}$$

[covariance terms drop out, summation of other terms remain Const K]

$$\sum_n P\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \epsilon\right) < \infty$$

$$\Rightarrow P\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \epsilon \text{ i.o.}\right) = 0$$

$\left|\frac{S_{n^2}}{n^2}\right| \leq \epsilon$ eventually, a.s. $\Rightarrow \frac{S_{n^2}}{n^2} \to 0$ a.s.

$(\epsilon = \frac{1}{2}, \frac{1}{3}, \ldots)$

Enough to show $\frac{S_n}{K_n^2} \to 0$ where K is such that $\underline{K_n^2 \leq n \leq (K_n+1)^2}$.

This is because

$$\frac{S_n}{n} = \frac{S_n}{K_n^2} \cdot \frac{K_n^2}{n} \quad \& \quad \frac{K_n^2}{n} \to 1.$$

$$\because \left(\frac{K_n}{K_n+1}\right)^2 \leq \frac{K_n^2}{n} \leq 1$$

Enough to show:

$$\frac{S_n - S_{K_n^2}}{K_n^2} \to 0, \quad \because \underline{\frac{S_{K_n^2}}{K_n^2} \to 0}$$

$$P\left(\frac{|S_n - S_{K_n^2}|}{K_n^2} \geq \epsilon\right) \leq \; ?$$

$\boxed{\text{Kolmogorov one-sided inequality}}$

$$P\left(\max_{n^2 \leq k \leq (n+1)^2} |S_k - S_{n^2}| \geq \epsilon n^2\right)$$

$$\leq P\left(\sum_{j=1}^{2n+1} |X_{n^2+j}| \geq \epsilon n^2\right)$$

$$S_k - S_{n^2} = \sum_{m=n^2+1}^{k} X_m$$

$$|S_k - S_{n^2}| \le \sum_{m=n^2+1}^{k} |X_m|$$

$$\le \sum_{m=n^2+1}^{(n+1)^2} |X_m| .$$

$$\therefore \quad \max_k |S_k - S_{n^2}| \le \sum_{m=n^2+1}^{(n+1)^2} |X_m|$$

$$\therefore \quad P\left(\max_{n^2 \le k < (n+1)^2} |S_n - S_{n^2}| \ge \epsilon n^2 \right)$$

$$\le \frac{1}{\epsilon^2 n^4} E\left[\left(\sum_{m=n^2+1}^{(n+1)^2} |X_m| \right)^2 \right].$$

Lemma: (Cauchy-Schwartz-Bunikovski inequality)

$$|E[XY]| \le \sqrt{E[X^2]}\sqrt{E[Y^2]}$$

$\langle X, Y \rangle$

Inner Product Space.

$$\|X\| = \sqrt{\langle X, X \rangle}$$

$$|\langle X, Y \rangle| \le \|X\| \|Y\|$$

Pf $E[|X-\lambda Y|^2]$.

$= E[X^2] + \lambda^2 E[Y^2] - 2\lambda E[XY] \geq 0$.

$\therefore 2E[XY] \leq \frac{E[X^2]}{\lambda} + \lambda E[Y^2]$.

take $\lambda = \sqrt{\frac{E[X^2]}{E[Y^2]}}$

$\Rightarrow 2E[XY] \leq 2\sqrt{E[X^2]E[Y^2]}$.

$\Rightarrow E[XY] \leq \sqrt{E[X^2]E[Y^2]}$.

Similarly, $E[|X+\lambda Y|^2] \geq 0 \Rightarrow$

$E[XY] \leq \sqrt{E[X^2]E[Y^2]}$.

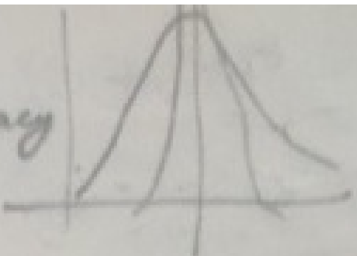
$\therefore E\left[\left(\sum_{i=n^2+1}^{(n+1)^2} |X_i|\right)^2\right]$

$= (2n+1) E[|X_i|^2] + \sum_{j \neq i} E[|X_i|] E[|X_j|]$

$\leq K(2n+1)$

$\therefore \sum_n P\left(\max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}| \geq \epsilon n^2\right) < \epsilon \frac{k}{\lambda^2} < \infty$

Machine learning



After n steps we are within ϵ of goal $\rightarrow$ Accuracy

with prob $1-\delta$ $\rightarrow$ confidence

If it does that then it's a good algorithm

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) < Ce^{-n(\ )}$$

Convergence in Law (Not a convergence of Random Variable at all)

$X_n \rightarrow X$ in law

if $E[f(X_n)] \rightarrow E[f(X)]$

$\forall$ bounded ~~function~~ continuous f.

Portmanteu theorem:

The following are equivalent:

(i) $E[f(X_n)] \rightarrow E[f(X)]$ for bdd. continuous f.

(ii) $E[f(X_n)] \rightarrow E[f(X)]$ for bdd. uniformly continuous f.

(iii) $\liminf_{n \to \infty} P(X_n \in G) \geq P(X \in G)$ $\forall$ open G.

(iv) $\limsup_{n \to \infty} P(X_n \in F) \leq P(X \in F)$ $\forall$ closed F

(v) $\lim_{n \to \infty} P(X_n \in A)$
$= P(X \in A)$ if
$P(X \in \partial A) = 0$.

(i) $\Rightarrow$ (ii) free

(ii) $\Rightarrow$ (iii) can find semi-cont.
$\langle f_m \rangle$ s.t. $f_m \uparrow I_G$

$\liminf_{n\to\infty} P(X_n \in G)$

$= \liminf_{n\to\infty} E[I(X_n \in G)]$

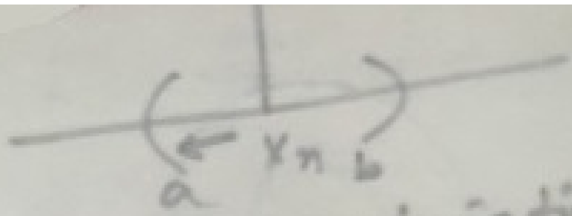
$\geq \lim_{n\to\infty} E[f_m(X_n)]$.

$= E[f_m(X)]$,

let $m \uparrow \infty$ RHS

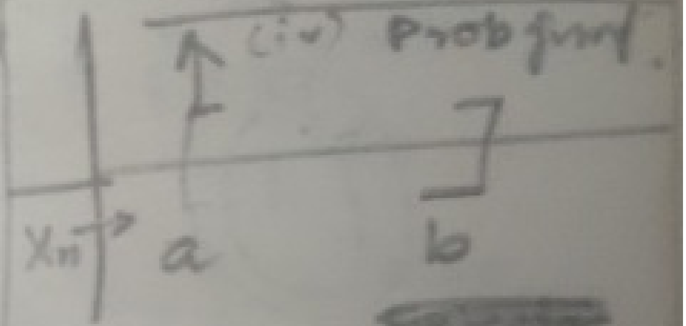
$\longrightarrow E[f(X)]$

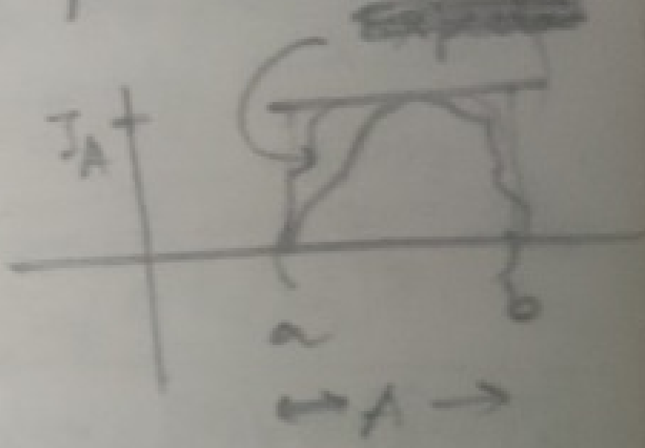
by MCT



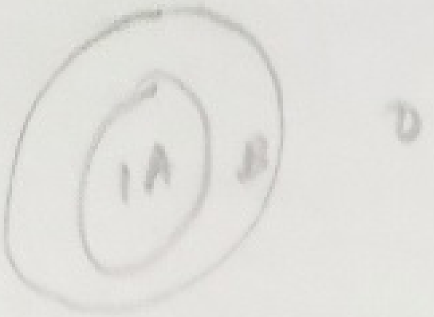
(iii) deterministic.
$\to$ concentrated at one pt.
Dirac at a, b
will be zero
(escape at bounds)

(iv) Prob fnd.





$\exists$ can always find
$\exists$ family of functions

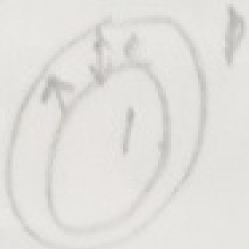


$$f(x) = \frac{d(x,B)}{d(X,A)+d(X,B)}$$

f can be approximated by the above function.

$$d(x,A) = \inf_{y \in A} \|x-y\|.$$

Open set



$\boxed{(iii) \Rightarrow (iv) \text{ take } F = G'}$