36-705 06/05/2020

YouTube 14/09/16

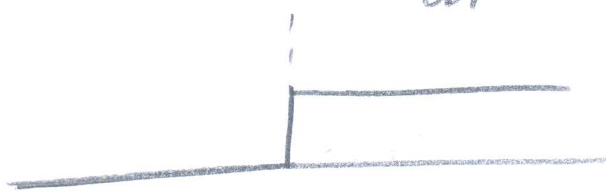
continue LN4 - convergence

Example 4

- sequence of r.v.s. $X_n \sim N(0, \frac{1}{n})$
- we proved

$X_n \overset{d}{\rightsquigarrow} \delta_0$

CDF



- $\delta_0(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$ (*) point-mass distn at 0

let's prove $X_n \xrightarrow{P} 0$ (i.e. converges in probability to 0 not r.v.) (a constant)

WTS

$P(|X_n| > \epsilon) \to 0$ as $n \to \infty$

W A1 (via Markov)

$P(|X_n| > \epsilon) \overset{(i)}{=} P(|X_n|^2 > \epsilon^2) \leq \frac{\mathbb{E}[X_n^2]}{\epsilon^2} = \frac{\frac{1}{n}}{\epsilon^2} \to 0$ special case (c)

Theorem 5: Following rel.

a.s. $\overset{(d)}{\Rightarrow}$ prob $\overset{(b)}{\Rightarrow}$ distn

prob $\overset{(a)}{\Uparrow}$ q.m.

a) $X_n \xrightarrow{qm} X \Rightarrow X_n \xrightarrow{P} X$

b) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$

c) If $X_n \xrightarrow{d} X$ and if $P(X=c)=1$ for some c, then $X_n \xrightarrow{P} X$

(convergence in distn to degenerate X)

d) $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$

- conduct proofs; examine counterexamples / failures of imp.

W A2: A little stumped by dir. of logic

a) $X_n \xrightarrow{q.m} X \Rightarrow X_n \xrightarrow{P} X$ Suppose $X_n \xrightarrow{q.m.} X$. Fix $\epsilon > 0$; via Markov and squaring

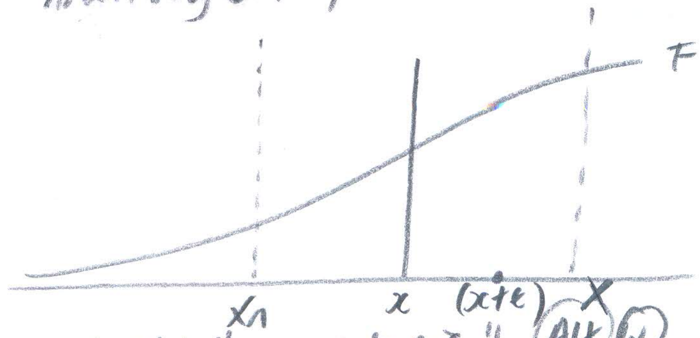
$P(|X_n - X| > \epsilon) = P(|X_n - X|^2 > \epsilon^2) \leq \frac{\mathbb{E}[|X_n - X|^2]}{\epsilon^2} \to 0$ as $n \to \infty$

hence $X_n \xrightarrow{P} X$

b) $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

- only need to show convergence in distn at any point (x) where it's (F) continuous

Arbitrary CDF of r.v. X



A4

HW: Typo : "continuous when he meant strictly monotonic" (last question)

- we have a sequence of r.v.s. X_n

$F_n(\cdot) = P(X_n \leq x)$

(OT) $P(A) = P(A \cap B) + P(A \cap B')$

- consider a point x and $(x+\epsilon)$

Then $P(X_n \leq x) = P(X_n \leq x, X \leq x+\epsilon) + P(X_n \leq x, X > x+\epsilon)$ (?) w A5 ✓

$\leq \underset{(i)}{P(X \leq x+\epsilon)} + \underset{(ii)}{P(|X_n - X| > \epsilon)}$

(i) $P(A \cap B) \leq P(B)$
(ii) (?)

$= F(x+\epsilon) + P(|X_n - X| > \epsilon)$

(ii) Drop lines at X_n and X on diagram

(Q): What does that mean about distance between X_n and X ?

(W): Trick: ~~into~~ Introduce new event so X_n and X both involved in calc.

- we now have an upper bound ; do a lower bound as so:- (W) A6 : check ✓

$F(x-\epsilon) - P(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x+\epsilon) + P(|X_n - X| > \epsilon)$

- Take limit of every term as $n \to \infty$

• $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$ (via $X_n \xrightarrow{p} X$)

$\lim_{n \to \infty}$:-

$F(x-\epsilon) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x+\epsilon)$

HW: Can think of lim inf and lim sup as $\lim_{n\to\infty} F_n(x)$; technicality

(*)
Take limit as $\epsilon \to 0$

$$F(x) \le \lim_{n\to\infty} F_n(x) \le F(x)$$

- Hence $\lim_{n\to\infty} F_n(x) = F(x)$ · Analysis style proof

W: Pick an arbitrary ϵ - Trickier proof strategy
Take limit over n
Notice it holds for all ϵ
Take limit as $\epsilon \to 0$

- Above are 2 main relationships.
- More interesting is showing reverse implication hold.

Proof (c): Not covered yet
Proof (d): Omitted

Reverse of (a) does not hold
"convergence in probability" $\not\Rightarrow$ "convergence in quadratic mean"
i.e. $X_n \xrightarrow{p} X \not\Rightarrow X_n \xrightarrow{q.m} X$
Construct a counterexample: (reveals why quadratic mean convergence is stronger)

- Let $U \sim \text{Unif}(0,1)$ $U \sim \text{Unif}(0,1)$
- Construct a sequence of r.v.s defined by:-

$$X_n = \sqrt{n}\,\mathbb{I}(0<U<\tfrac{1}{n}) = \sqrt{n}\,\mathbb{I}_{(0,1/n)}(U)$$

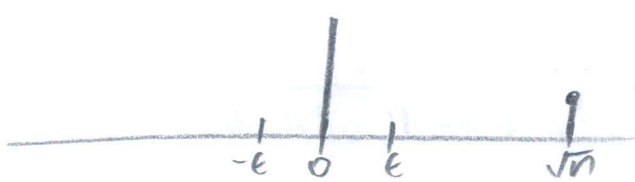
- This converges in probability, to 0 ; $X_n \xrightarrow{p} 0$

Q: What possible values can X_n take?

- X_n can only take 2 values - 0 and $\sqrt{n}$
- Only way for $X_n > \epsilon$ and $|X_n| > \epsilon$ is when $X_n \ne 0$

$$P(|X_n| > \epsilon) = P(0 < U < \tfrac{1}{n}) = \tfrac{1}{n} \to 0 \text{ as } n \to \infty$$ PMF

- Consider successive PMFs as $n \to \infty$



- This is converging in probability to 0 $X_n \xrightarrow{p} 0$
- consider neighbourhood of 0 i.e. interval around 0

(?) (W) (A7): narrative inset

- Intuitively, successive distributions 'squishing' to 0. (O1)

Q(?): $\mathbb{E}^2[\cdot]$ or $E[(\cdot)^2]$?

- quadratic mean

LW: But doesn't converge in q.m.

compute $\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2]$

X_n^2	X_n	$P(X_n)$	
0	0	$1-\frac{1}{n}$	
n	$\sqrt{n}$	$\frac{1}{n}$	

$\mathbb{E}[X_n^2] = n(\frac{1}{n}) = 1 \not\to 0$

LW: Reveals why convergence in prob. is a weaker statement than convergence in q.m.

LW: - probability mass concentrates near 0 as $n \to \infty$
- convergence in q.m is computing moments (2nd moments)
- moments are affected by tails
- Although very little probability mass in tail (see diagram), # it is far, and it is 'shooting to infinity'.
- It is 'shooting off' fast enough that when we compute 2nd moment $\mathbb{E}[X^2]$, it does not converge to 0.
- convergence in q.m is stronger statement because it is a statement about the tails of a distribution.
- convergence in probability - as long as probability mass near 0 approaches 1, how far out values for which there is a little probability mass does not matter.
- i.e. as long as an exceptional event has a small probability, it does not matter
- with 2nd moments in convergence in q.m; the value/realisation of the r.v. starts to matter
- Again about tails of distri.

Reverse of (b) does not hold:

"convergence in distribution" $\not\Rightarrow$ "convergence in probability"

i.e. $X_n \rightsquigarrow X$ / $X_n \xrightarrow{d} X$ $\not\Rightarrow$ $X_n \xrightarrow{p} X$

LW: convergence in distn is not about r.v.; it is about the CDF

(*) It's about what probability statements I would make about a random variable, not random variable themselves (*)

(!!)(*) Convergence in probability is a statement about random variables

(!!)(*) Convergence in distrib is about the probability statements you make about the random variables.

- Let $X \sim N(0,1)$
- Let $X_n = -X$ for $n = 1, 2, 3, \ldots$

→ - Generative process
  - Draw X from Normal
  - For every r.v. X_i in sequence; make it negative.
  - Straight sequence; highly corr.

- $X_n \stackrel{d}{=} X$ →
  - X_n converges to X in distn
  - They also have identical distn
  - As -ve of Normal(0,1) is Normal(0,1)
  - Equal in distn: does not mean r.v.s equal; but CDFs are equal.
  - Simulation of r.v.s → X_n and X may not be same
  - But any question about probabilities of X_n and X would get the same answer

- Hence a trivial, but strong sense in which:- $X_n \xrightarrow{d} X$

- Q: what is difference between X_n and X?

- $P(|X_n - X| > \epsilon) = P(|-X - X| > \epsilon)$

$= P(|2X| > \epsilon) = P(|X| > \epsilon/2) \not\to 0$

Hence $X_n \stackrel{p}{\not\to} X$.

LW: Think experimentally: $X = 6$ $X_n = -6$

Difference is large; r.v.s are not close

BUT if ask what's probability X_n or X is ... → same answers

- LW: Review diagram, proofs, counterexamples to build intuition

---

<u>Example 6</u> (LW)(17): review. ✓

- Examine in space time
- LW: (Q) If $X_n \xrightarrow{P} X$, does $\mathbb{E}[X_n] \to \mathbb{E}[X]$
  - In general, no
  - Recall, convergence in prob is about probability; RHS is a statement about <u>moments</u> (more <u>sensitive</u>)

LW: construct a convergent distn of an interesting sequence of r.v.s.

<u>Example 7</u>

- $X_1, \dots, X_n \sim$ uniform $(0,1)$
- $X_{(n)} = \max_i X_i$
- Examine convergence of $\max_i X_i$ / $X_{(n)}$ in <u>probabil.</u> and <u>distn</u>

- Notation: <u>order</u> of sequence

  $X_{(1)} \leq X_{(2)} \leq X_{(3)} \dots \leq X_{(n)}$;

  $\max_i X_{(i)} = X_{(n)}$

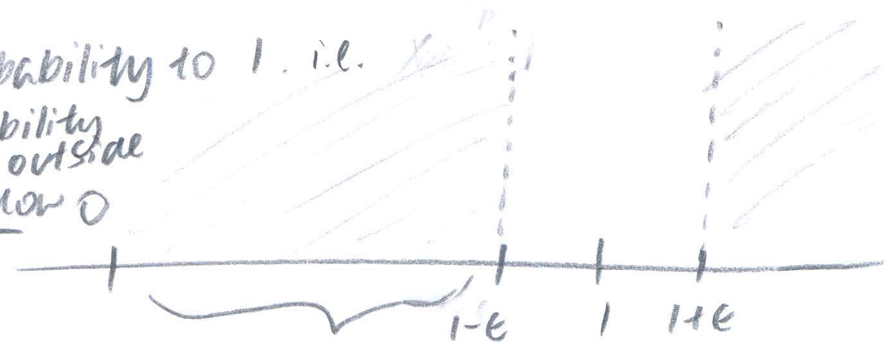
- <u>Diagram</u>:

- Claim: $X_{(n)}$ converges in probability to 1. i.e.

<u>$X_{(n)} \xrightarrow{P} 1$</u>

- $P(|X_{(n)} - 1| > \epsilon)$

probability here outside this window 0



- Hence

$$P(|X_{(n)} - 1| > \epsilon) = P(X_{(n)} < 1-\epsilon) = P(\text{all } X_i < 1-\epsilon) = \prod_i P(X_i < 1-\epsilon)$$

$$= (1-\epsilon)^n \to 0 \quad \text{as } n \to \infty$$

· convergence in distri:

LW: Good example; convergence in distri can be adapted to give useful, meaningful limit via rescaling the r.v.

- Define $Y_n = n(1 - X_{(n)})$; note multiplication by n.

(*) To find where Y_n is converging to in distribution, examine CDF.

$$P(Y_n \leq t) = P(n(1 - X_{(n)}) \leq t)$$

$$= P\left(X_{(n)} \geq 1 - \left(\tfrac{t}{n}\right)\right)$$

$$= 1 - P\left(X_{(n)} \leq 1 - \left(\tfrac{t}{n}\right)\right)$$

$$= 1 - \prod_i P\left(X_i \leq 1 - \left(\tfrac{t}{n}\right)\right)$$

$$= 1 - \left(1 - \tfrac{t}{n}\right)^n \xrightarrow{n \to \infty} 1 - e^{-t} \quad (*)$$

LW AQ: Fill in: ↓

(i) standard.

(ii) standard

(iii) use earlie logic

LW: No rescaling, we get CDF convergence to point-mass distri (degenerate)
rescaling → we get convergence to non-degenerate distri. (more interesting)

(*) CDF of exponential distri.

$$n(1 - X_{(n)}) \xrightarrow{d} Exp(1)$$

LW: useful when we cannot compute prob directly e.g. (complicated distri)

$$P(a < X_{(n)} < b) = P(1 - b < 1 - X_{(n)} < 1 - a)$$

$$= P(n(1-b) < n(1 - X_{(n)}) < n(1-a))$$

$$\approx P(n(1-b) < V < n(1-a)) \quad \text{(for large } n\text{)}$$

where V is an exponential distri r.v.

LW: shows why convergence in distri is powerful; start with a probability statement about an r.v., can use conveg in distri, recontextualise, substitute what it is converging to; maybe easier.

Q: How did you know what transf. to apply?

→ LLN/CLT results

LW: convergence in distn often involves rescaling

Preservation of convergence (Theorem 8) (under transf.)

(a) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ then $X_n + Y_n \xrightarrow{p} X + Y$

(b) If $X_n \xrightarrow{qm} X$ and $Y_n \xrightarrow{q.m} Y$ then $X_n + Y_n \xrightarrow{q.m} X + Y$

c) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ then $X_n Y_n \xrightarrow{p} XY$

• W A10: Summarise

- LW: Sums preserve convergence; except for in distn; where one has to be 'constant'

- Different for convergence in distn.

$X_n \xrightarrow{d} Y$ and $Y_n \xrightarrow{d} Y$ $\not\Rightarrow$ $X_n + Y_n \xrightarrow{d} X + Y$

BUT

Theorem 9 (Slutsky's theorem)

- If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ then $X_n + Y_n \xrightarrow{d} X + c$
- If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ then $X_n Y_n \xrightarrow{d} cX$

Theorem 10 (continuous mapping theorem)

(a) If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$

(b) If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$

LW: A lot of machinery; let's use it

- Main goal: LLN/CLT

- LLN - main theorem related to convergence in probability
- CLT - main —"— —"— convergence in distribution.

3 Law of large nos

Lw: There exist extensions of these to NON-IID settings

- let $X_1, X_2, \ldots,$ be an IID sample
- let $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = Var[X_i]$
- let sample mean: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\mathbb{E}[\bar{X}_n] = \mu$ $Var(\bar{X}) = \frac{\sigma^2}{n}$

Theorem 11 (WLLN)

- If $X_1, \ldots, X_n$ IID then $\bar{X}_n \xrightarrow{P} \mu$
- (means that $\bar{X}_n - \mu = o_p(1)$)

- Theorem is true even if variance does not exist
- Proof is for when variance exists.

(i)

W (A11): check you understand steps ✓

Proof - suppose that $\sigma < \infty$
- via chebyshev:-

$$P(|\bar{X}_n - \mu| > \epsilon) \underset{(i)}{=} P(|\bar{X}_n - \mu|^2 > \epsilon^2) \underset{(ii)}{\leqslant} \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

Intepretation: Distribution of $\bar{X}_n$ (sampling distn / distn of sampling estimator of mean) concentrates around true/population mean as n gets large.

Theorem 12 (SLLN)

- $X_1, \ldots, X_n$ IID with $\mathbb{E}[X_i] = \mu$
- Then $\bar{X}_n \xrightarrow{a.s.} \mu$

(*) These are the crowning achievements of probability

Next lecture: CLT → more subtle