36-705   08/05/2020

YouTube 16/09/16

Convergence theory

WLLN: $\bar{X}_n \xrightarrow{P} \mu$ i.e. $\bar{X}_n - \mu = o_P(1)$

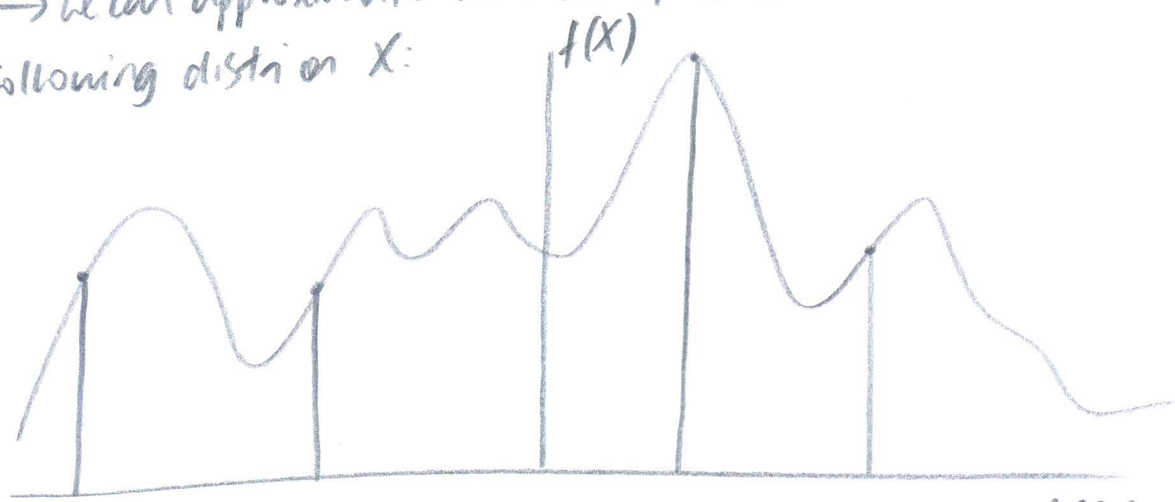
- IID case
- Many non-IID generalisations

Yields continuous mapping theorem:-

If g is continuous then
$$g(\bar{X}_n) \xrightarrow{d} g(\mu)$$

$X_1, \ldots, X_n \sim P$ $\mu = \mathbb{E}[X_i]$ $\sigma^2 = \mathbb{V}(X_i)$

- Examine distn of $\bar{X}_n$ - can be complicated $\rightarrow \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- CLT $\rightarrow$ we can approximate the distn of $\bar{X}_n$ with Normal distn
- E.g. following distn on X:



- No matter what the underlying distn P is; if we take n observations and evaluate $\bar{X}_n$ (itself an r.v.); the distn of $\bar{X}_n$ will be Normal.
- State theorem; give heuristic proof.

- Convergence in distribution normally uses standardisation.
- Convergence in probability to constant μ; may argue convergence in distribution to point mass in μ. (degenerate)
- Not very informative

LW: Find limiting distribution should be taken as find a converg. in distn to a non-degenerate distn (i.e. not point mass) (*)

correct scaling:

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{var(\bar{X}_n)}}$$

$$= \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

- standardise r.v. is a fundamental operation
  → subtract mean, divide by standard dev.
  - IW: intuitively $(\bar{X}_n - \mu) \xrightarrow{P} 0$
  - so it is converging to a degenerate distri
  - multiplying by $\sqrt{n}$ increases/enlarges fluctuations about the mean at just the right rate such that $\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma}$ converges to a well-defined distri.

Theorem(13) CLT:

- $X_1, \dots, X_n$ IID, $\mathbb{E}[X_i] = \mu$, $\mathbb{V}(X_i) = \sigma^2$ $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$; Then

$$Z_n \equiv \frac{\bar{X}_n - \mu}{\sqrt{var(\bar{X}_n)}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \quad \text{where } Z \sim N(0,1)$$

IW: intuitively, CDF of Z_n converges to CDF of $Z \sim N(0,1)$.

- Probabilistic statements about Z_n, we can approximate these with probability statements about Z. (*) formal state.

informally: $\bar{X}_n \approx N(\mu, \frac{\sigma^2}{n})$ (an intuitive way of thinking about this)

- recall $var(\bar{X}_n) = \frac{\sigma^2}{n}$

Example of usefulness:

- consider a sit. where we only have mean and variance of an unknown distri P.

$$P(a < \bar{X}_n < b) = P\left(\frac{\sqrt{n}(a-\mu)}{\sigma} < \underbrace{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}_{(ii)} < \frac{\sqrt{n}(b-\mu)}{\sigma}\right)$$

- (ii) $\xrightarrow{d} N(0,1)$

$$\approx P\left(\frac{\sqrt{n}(a-\mu)}{\sigma} < Z < \frac{\sqrt{n}(b-\mu)}{\sigma}\right)$$

Φ - std Normal CDF.

$$= \Phi\left(\frac{\sqrt{n}(b-\mu)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}(a-\mu)}{\sigma}\right)$$

- CLT usefulness; gone from something which potentially we did not know how to compute, to a 'formula'.
- A way of approximating probabilities, useful, ubiquitous in <u>statistics</u> (informally)

<u>CLT Proof</u>

$X_i \quad \mathbb{E}[X_i] = 0$ var[

LW: Define a mean 0, variance 1 r.v., $\mu = 0$, $\sigma = 1$, $\sigma^2 = 1$

- <u>can always</u> rescale:- $Y_i = \frac{X_i - \mu}{\sigma}$ if using different r.v. Y_i will have mean 0 and variance 1 by transf. X_i if X_i does not have $\mathbb{E}[X_i] = 0$, $\mathrm{Var}(X_i) = 1$

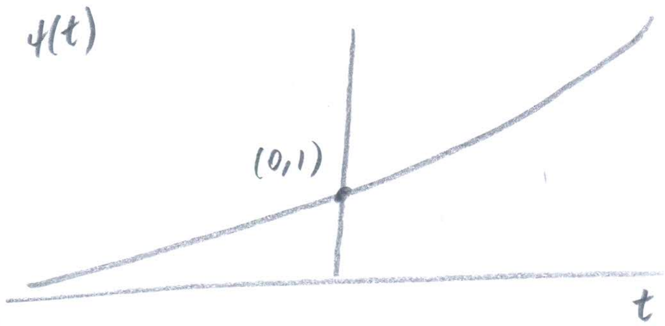
- Define (a sequence of r.v.s):-

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \sqrt{n}\,\bar{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

- Assume MGF of this Z_n is <u>well-defined</u> (made as part of a heuristic simpl.)

- Define MGFs for X_i and Z_n

- $\psi_i(t) = \mathbb{E}[e^{tX_i}]$ - same $\forall i$ as X_i IID.
- $\xi_n(t) = \mathbb{E}[e^{tZ_n}] = \mathbb{E}[e^{\frac{t}{\sqrt{n}}\sum_i X_i}] = \mathbb{E}[e^{\frac{t}{\sqrt{n}}X_1} \cdot e^{\frac{t}{\sqrt{n}}X_2} \cdot \ldots \cdot e^{\frac{t}{\sqrt{n}}X_n}]$
- As $X_1, \ldots, X_n$ are independent; <u>expectation of product</u> is <u>product of expect.</u>
- $\xi_n(t) = \prod_{i=1}^{n} \mathbb{E}[e^{\frac{t}{\sqrt{n}}X_i}] = \prod_{i=1}^{n} \psi_i\left(\frac{t}{\sqrt{n}}\right) = \left(\psi\left(\frac{t}{\sqrt{n}}\right)\right)^n$
- Consider MGF $\psi_X(t)$ (arbitr.)
- Consider Taylor exp. of $\psi_X(t)$ around 0.



$$\psi(t) = \underset{(i)}{\psi(0)} + \underset{(ii)}{t\psi'(0)} + \underset{(iii)}{\frac{t^2}{2!}\psi''(0)} + \frac{t^3}{3!}\psi'''(0) + \ldots$$

$$= 1 + 0 + \frac{t^2}{2} + \frac{t^3}{6}\psi'''(0) + \ldots$$

(i) $e^0 = 1$
(ii) $\psi'(0) = \mu = 0$
(iii) $\psi''(0) = \sigma^2 = 1$

Now:-

$$\xi_n(t) = \left[\psi\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

$$= \left[1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}}\psi'''(0) + \dots\right]^n$$

$$= \left[1 + \frac{\frac{t^2}{2} + \frac{t^2}{3!n^{1/2}}\psi'''(0) + \dots}{n}\right]^n \xrightarrow{n\to\infty} e^{\frac{t^2}{2}}$$

(A12)

(*) $\left(1+\frac{a_n}{n}\right)^n$ $a_n \to a$, $n\to\infty$

$\hookrightarrow e^a$

- We have shown MGF of $\xi_n(t)$ converges to MGF of a standard Normal.
- If MGF converges to another MGF; it implies convergence in distn $\times$ Lemma 14 ← not precise
- (1): similar to CDF specification; MGFs and CDFs go definitions specify whether distributions are equal
- Hence $Z_n \xrightarrow{d} N(0,1)$. - Not fully rigorous; but heuristic

→ and $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$

Lemma 14

- Let $Z_1, Z_2, \dots$ be a sequence of r.v.s.
- Let ψ_n be the mgf of Z_n
- Let Z be another r.v. and denote its mgf by ψ.
- If $\psi_n(t) \to \psi(t)$ $\forall t$ in some open interval around 0; then $Z_n \xrightarrow{d} Z$.

- Experimental; pick any strange distn
- Drawn n observations; compute $\bar{X}_n$ and compute a histogram of $\bar{X}_n$
- Increase sample size

CLT: As $n \to \infty$; it says CDF gets close to Normal; but how close to Normal?

Theorem 16 (Berry-Esseen)

- Assumed 1st and 2nd moments finite (x)
- But now also: 3rd moment :- $\mu_3 = \mathbb{E}[|X-\mu|^3] < \infty$
- Consider :-

$$F_n(z) = P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z\right)$$

Then

$$\sup_z \left| F_n(z) - \Phi(z) \right| \leq \frac{33}{4} \frac{\mu_3}{\sigma^3 \sqrt{n}}$$

LW: 1) The difference the true CDF $F_n(z)$ and approximation $\Phi(z)$ approaches 0 at around rate $\frac{1}{\sqrt{n}}$.

2) Statistically, can't do much better / faster convergence than $\frac{1}{\sqrt{n}}$. (not exponential however)

3) constant probably not optimal

4) Large 3rd moment is, worse bound is.

5) Gives a sense for how fast normal approx 'converges' to true

6) Around $n \simeq 30$ observ. and normal becomes good approx. of sampling est. distn.

LW: Generally don't know σ^2 (true variance); but have sampling variance s_n^2 (unbiased) or standard dev.

- $\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n} = \underset{(i)}{\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma}} \cdot \underset{(ii)}{\frac{\sigma}{S_n}}$ (i) $\quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$

(i) multiply/divide by σ

- Note:- $\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{d} Z$ and S_n^2 is an estimator of σ^2
(we will show $S_n \xrightarrow{p} \sigma^2$)

LW: Q: Does $S_n^2 \xrightarrow{p} \sigma^2 \Rightarrow S_n \xrightarrow{p} \sigma$

A: Yes - continuous mapping theorem (as $g(x)=\sqrt{x}$; continuous)

Q: Does $S_n \xrightarrow{p} \sigma \Rightarrow \frac{\sigma}{S_n} \xrightarrow{p} \frac{\sigma}{\sigma} = 1$

A: Yes - over $\mathbb{R}^+$; $1/x$ is continuous (same cont. mapping)

Q: Does $\left(\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma}\right) \xrightarrow{d} Z$ and $\frac{\sigma}{S_n} \xrightarrow{p} 1$ both imply that

$\left(\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n}\right) \xrightarrow{d} Z$?

- Yes, via Slutsky's theorem (?)

(?) - start (W) (A): Slutsky's theorem does not explicitly specify that a product of an r.v. (distr. conv.) and another r.v. (prob convergence) yields distr conv.

- Fill in logic here:

$\frac{\sigma}{S_n} \xrightarrow{p} 1 \Rightarrow \frac{\sigma}{S_n} \xrightarrow{d} 1$; $\left.\begin{array}{l} \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{d} Z \\ \frac{\sigma}{S_n} \xrightarrow{d} 1 \end{array}\right\} \Rightarrow \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \cdot \frac{\sigma}{S_n} \rightarrow 1(Z)$

Q: How do we know $S_n^2 \xrightarrow{p} \sigma^2$?

- Start with $R_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}X_i - \bar{X}_n^2$

- what does $\frac{1}{m}\sum_{m=1}^{m} X_i$ converge to (sample average)
- via WLLN; $\frac{1}{m}\sum_{m=1}^{m} X_i \xrightarrow{P} \mathbb{E}[X_m^2] = (\mu^2+\sigma^2)$
- $\bar{X}_m \xrightarrow{P} \mu$ (WLLN)
- $\bar{X}_m^2 \xrightarrow{P} \mu^2$ (WLLN + cont. mapping)
- Hence $R_m^2 \xrightarrow{P} (\mu^2+\sigma^2) - \mu^2 = \sigma^2$ (via pres. of conv under sum.)
- we want:-

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{n}{n-1} \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}_{R_n^2}$$

- switching indexing letters $m \to n$

- $R_n^2 \xrightarrow{P} \sigma^2$ $n/(n-1) \to 1$; we have $S_n^2 \xrightarrow{P} \sigma^2$

- 2 versions of CLT:-

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1) \quad \text{and} \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$$

- Next lesson: Multivariate CLT
  Delta method

- HW: Tests