36-705 01/07/2020

Lecture Notes 4 - Review

- Supplemented with Ch 5 (Wasserman)

(*) In calculus, a sequence of real nos x_n converges to a limit x if, for every $\epsilon > 0$, $|x_n - x| < \epsilon$ for all large n.

(*) Create subtlety for convergence in prob framework.

(*) Illustrating:-

- suppose $x_n = x$ $\forall n$.

- trivially, $\lim_{n\to\infty} x_n = x$

(*) suppose $X_1, X_2, \dots$ is a sequence of r.v.s. which are independent, each with $N \sim (0,1)$.

(*) we want to develop a rigorous framework within which to express the following kinds of intuitions (which cannot be expressed in the language of convergence only within calculus).

- i) X_n "converges" to $N(0,1)$ distri

  - $P(X_n = X) = 0$ $\forall n$ (no meaning)

(*) suppose $X_1, X_2, \dots$ where $X_i \sim N(0, 1/n)$

- i) intuitively, want to say that distn is concentrated around 0 for large n, so want to encode idea that X_n "converges" to 0.

  $P(X_n = 0) = 0$ $\forall n$.

(*) Illustration of convergence in distri.

- Recall X_n converges to X in distn: $X_n \xrightarrow{d} X$

  if $\lim_{n\to\infty} F_n(t) = F(t)$ $\forall t$ at which F is continuous

- For cdf F, which is a point mass at 0.

(Fig 5.1): - F_n converges to $F(t)$ $\forall t$ except $t=0$ (which is not a point of continuity)

(*) Types of convergence to a constant:-

- $X_n \xrightarrow{P} X$; $X_n \xrightarrow{P} c$ and $X_n \xrightarrow{d} X$; $X_n \xrightarrow{d} c$

- $X_n \xrightarrow{q.m.} X$; $X_n \xrightarrow{q.m.} c$ (almost sure left out).

(*) In all of these cases; convergence (in prob/dist/q.m.) to a constant can be viewed as a more general case of convergence (in ...) to an r.v.; but where $P(X=c)=1$.

(*) That is r.v. which sequence of r.v.s. is converging to is a point mass at the $X=c$. (i.e. degenerate)

(A1). $\dfrac{\mathbb{E}[X_n^2]}{\epsilon^2} = \dfrac{\frac{1}{n}}{\epsilon^2}$

- Note that, as $X_n \sim N(0,1)$ we have $\mathbb{E}[X_n]=0$ and $\mathbb{V}ar(X_n)=\frac{1}{n}$
- Also:- $\mathbb{V}(X_n)=\mathbb{E}[X_n^2]-(\mathbb{E}[X_n])^2$

$\Rightarrow \mathbb{E}[X_n^2]=\mathbb{V}(X_n)+\mathbb{E}[X_n]^2=\frac{1}{n}$.

(A2): Reviewed - fine.

- Fairly simple.
- Start with $P(|X_n-X|>\epsilon)$ (i.e. what we need to prove.)
- Manipulate this using standard algebra; apply Markov, invoke assumpti...

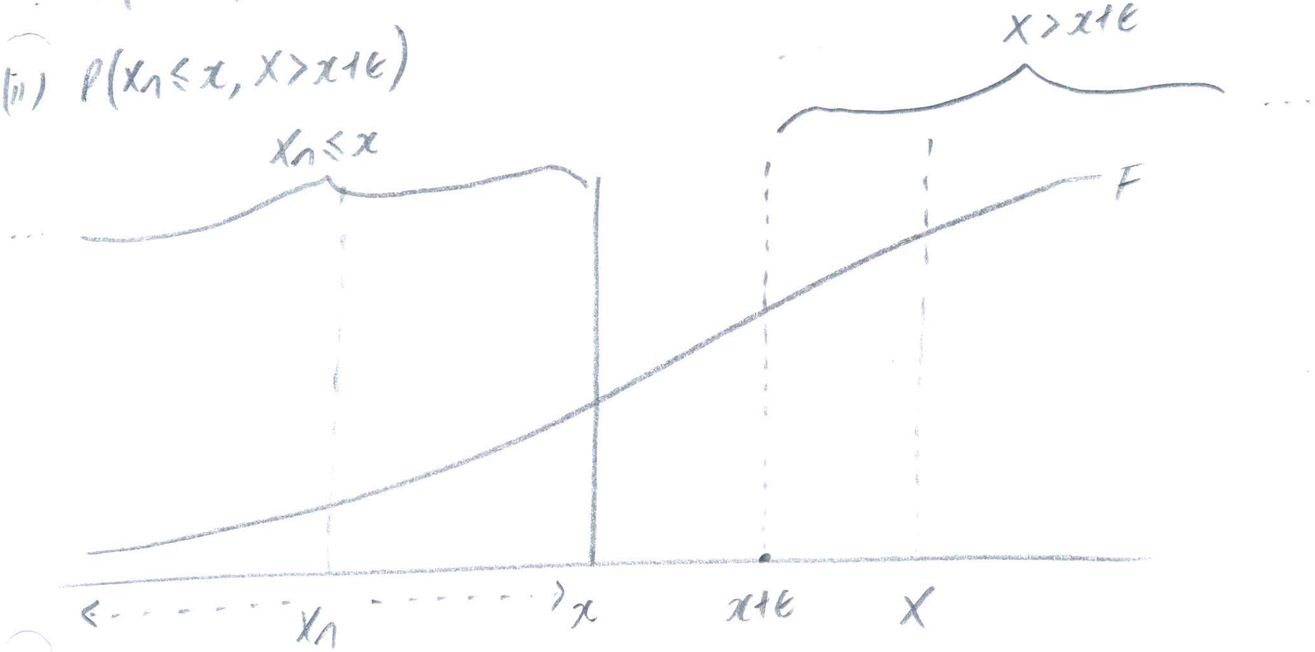
(A5): $P(X_n \leq x) = P(X_n \leq x, X \leq x+\epsilon) + P(X_n \leq x, X > x+\epsilon)$ as $P(A) = P(A \cap B) + P(A \cap B')$

(A6)

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

- (introducing event involving X)

- How do we get this?
- (i): $P(X_n \leq x, X \leq x+\epsilon)$
- defining event A to be $X_n \leq x$ and event B to be $X \leq x+\epsilon$:-
- note :- $P(A \cap B) \leq P(B)$.
- $\therefore$ $P(X_n \leq x, X \leq x+\epsilon) \leq P(X \leq x+\epsilon)$
- (ii) $P(X_n \leq x, X > x+\epsilon)$



- If $X_n \leq x$ and $X > x+\epsilon$, then their difference will be greater than ϵ.

i.e. $|X_n - X| > \epsilon$.

- Turn this into probability by defining events accordingly.

Hence; $P(X_n \leq x, X > x+\epsilon) = P(\{X_n \leq x\} \cap \{X > x+\epsilon\}) = P(|X_n - X| > \epsilon)$

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

(*) Doing the same for the lower bound on $F_n(x)$ (similar but diff.)

- we want a lower bound on $F_n(x)$.

- Again in terms of $F(x-\epsilon)$ and $P(|X_n - X| > \epsilon)$.
- start with event $\{X < x-\epsilon\}$:-

$$P(X \leq x-\epsilon) = P(X \leq x-\epsilon, X_n \leq x) + P(X \leq x-\epsilon, X_n > x)$$ ⊛

(via same logic)

⊛: Here we introduce event $B = \{X_n \leq x\}$
$B' = \{X_n > x\}$

- Note that, under similar arguments as before:-

i) $P(X \leq x-\epsilon, X_n \leq x) \leq P(X_n \leq x)$

ii) $P(X \leq x-\epsilon, X_n > x) = P(|X_n - X| > \epsilon)$

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

$F(x-\epsilon) - P(|X_n - X| > \epsilon) \leq F_n(x)$ yielding the required lower bound.

(*) From here take limits; use $X_n \xrightarrow{p} X$

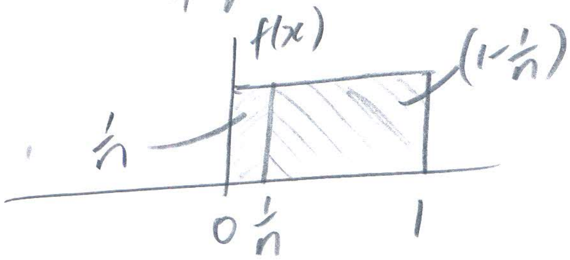
(*) showing reverse of theorem 5a does not hold:-

i.e. $X_n \xrightarrow{p} X \not\Rightarrow X_n \xrightarrow{q.m.} X$

construct example:- $X_n = \sqrt{n}\, \mathbb{I}(0 < U < \frac{1}{n})$ (sequence of r.v.s.)

- consider values of $X_n \in \{0, \sqrt{n}\}$
- $P(|X_n| > \epsilon) = P(0 < U < \frac{1}{n}) = \frac{1}{n} \to 0$ as $n \to \infty$ $\therefore X_n \xrightarrow{p} 0$

(?) Why is it the case that $\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2]$ (just a statement of quadratic mean)



- Logic of the table:-
- Note $U = \text{Unif}(0,1)$
- use this to create a sequence of r.v.s.
- $\mathbb{I}(0 < U < \frac{1}{n})$ can be thought of as. r.v.
- $P(\mathbb{I}(0 < U < \frac{1}{n})) = P(0 < U < \frac{1}{n}) = \frac{1}{n}$

- The logic of table:-

$X_n = \begin{cases} 0 \text{ when } u \notin (0, 1/n) \\ \sqrt{n} \text{ when } u \in (0, 1/n) \end{cases}$

As $u \notin (0, 1/n)$ with probability $(1 - \frac{1}{n})$

$u \in (0, 1/n)$ — " ——— $\frac{1}{n}$.

(*) so we use this to compute the quadratic mean of X_n assuming it does converge to 0, and show that it does not <u>converge</u>

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

(*) showing that 1 S b. does not hold

$X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{p} X$

- $X \sim N(0,1)$
- $X_n = -X$ for $n = 1, 2, 3, \ldots$
- Then $X_n \overset{d}{=} X$ (stronger sense in which $X_n \xrightarrow{d} X$)
- ~~compute~~ evaluate diff:- (substitute $X_n = -X$

$P(|X_n - X| > \epsilon) = P(|-X - X| > \epsilon) = P(|2X| > \epsilon) = P(2|X| > \epsilon)$

- As $X \sim N(0,1)$ is standard normal, it will not get <u>smaller</u>
- Hence $P(|X| > \epsilon/2) \not\to 0$

(*) Example 6 (A7)

$X_n \xrightarrow{p} b \not\Rightarrow \mathbb{E}[X_n] \longrightarrow b$

- Define $X_n = \begin{cases} 0 & \text{with } P(X_n = 0) = 1 - (1/n) \\ n^2 & \text{—"— } P(X_n = n^2) = 1/n \end{cases}$

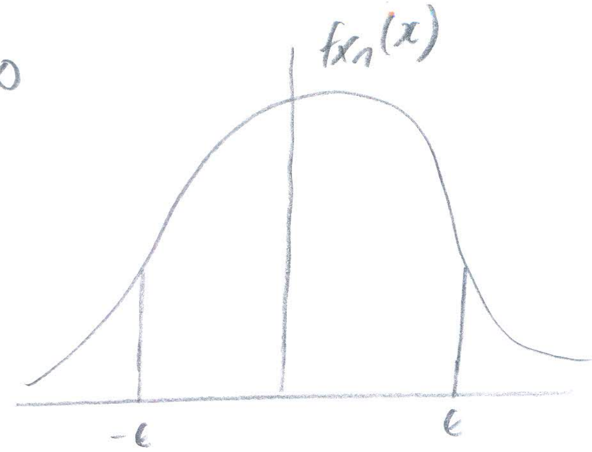
(?) why is $P(|X_n| < \epsilon) = P(X_n = 0) = (1 - 1/n)$?

- Think this is because we are assuming that $\epsilon > 0$

- If $\epsilon > 0$; then the only situation in which $X_n < \epsilon$ and hence $|X_n| < \epsilon$ is when $X_n = 0$.
- Hence $P(|X_n| < \epsilon) = P(X_n = 0) = 1 - (\frac{1}{n})$ ✓
- ~~No~~ Note that $P(|X_n| < \epsilon) \longrightarrow 1$ as $n \to \infty$

Note: $P(|X_n| \geq \epsilon) = 1 - P(|X_n| < \epsilon)$

$$\Rightarrow \lim_{n\to\infty} P(|X_n| > \epsilon) = 1 - \lim_{n\to\infty} P(|X_n| < \epsilon)$$



$$\Rightarrow \lim_{n\to\infty} P(|X_n| > \epsilon) = 0$$

$\Rightarrow P(|X_n| > \epsilon) \longrightarrow 0$ as $n \to \infty$ and $X_n \xrightarrow{P} 0$

- Evaluate:-

$$\mathbb{E}[X_n] = n^2(1/n) + 0(1 - \tfrac{1}{n}) = n$$

- As $n \to \infty$ $\mathbb{E}[X_n] \longrightarrow \infty$ and hence $\mathbb{E}[X_n] \not\longrightarrow 0$ as $n \to \infty$

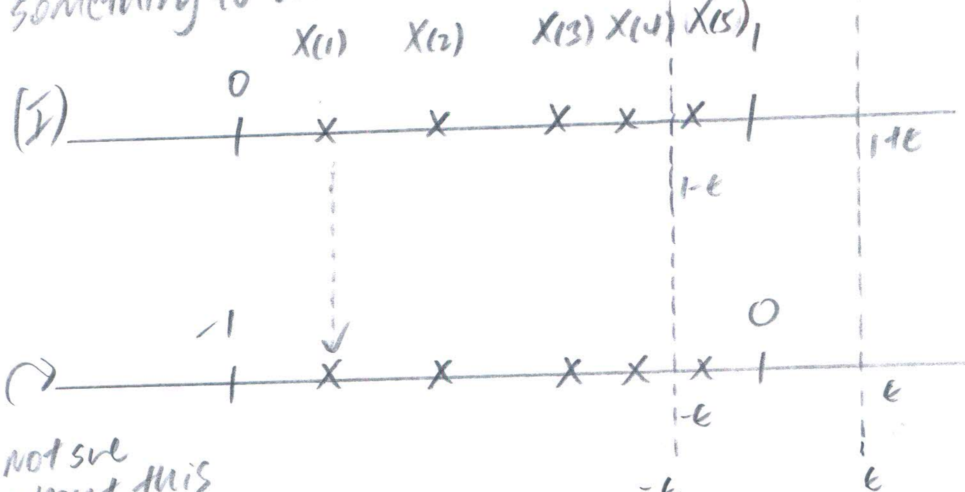
(*) use diagrams when considering tail probabilities (OTT)

(*) Example 7

0/5 1

- tighten up - get to bottom of confusion

- Logic of diagram is clearer on a 2nd pass.
- something to do with transformations of $\mathbb{R}$?



$X_i \sim \text{Unif}(0,1)$

$X_i - 1 \sim \text{Unif}(-1,0)$

- Have to consider $-\epsilon$ and $+\epsilon$ as considering $|X_{(n)} - 1|$

- Not sure about this comparison

(N): Got very confused with use of absolute values to encode 2 sets of inequalities at the same time.

· Note:-

$P(|X_{(n)}-1|>\epsilon) = P(X_{(n)}-1>\epsilon) + P(-(X_{(n)}-1)>\epsilon)$ $\quad |x| = \begin{cases} x & \text{if } x>0 \\ -x & \text{if } x<0 \end{cases}$

$= P(X_{(n)}-1>\epsilon) + P(X_{(n)}-1<-\epsilon)$

$= \underbrace{P(X_{(n)}>1+\epsilon)}_{=0} + P(X_{(n)}<1-\epsilon)$

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

$= (1-\epsilon)^n \to 0$ as $n\to\infty$

· Here $X_{(n)} \xrightarrow{P} 1$.

(W)(Q151): I don't understand how it is he immediately writes down diagram (I). And how that fully relates to what is being done algebraically

---

(?): not sure how we have that $P(Y_n \leq t) \xrightarrow{n\to\infty} 1-e^{-t} \Rightarrow$

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

(*) note that $\xrightarrow{d}$ is concerned with CDFs

(*) not confuse the exponential distribution with the exponential family of distributions.

note:- exp distn:- $X \sim Exp(\beta)$ has:-

PDF: $f(x;\beta)\ \frac{1}{\beta} e^{-x/\beta} \quad x>0,\ \beta>0$

CDF: $F(x;\beta) = \begin{cases} 1-e^{-\beta x} & x \geq 0 \\ 0 & x<0 \end{cases}$

(*)(Q17): using above, we have shown that $\underset{=P(Y_n\leq t)}{F_n(t)} \xrightarrow{n\to\infty} F(t)$

where $F(t) = Exp(t;1)$

$\therefore\; Y_n = n(1-X_{(n)}) \xrightarrow{d} \exp(1)$

(*) need to know inequalities by **heart** (tail)

Gaussian:- $X \sim N(0,1)$

$P(|X|>\epsilon) \leq \dfrac{2e^{-\epsilon^2/2}}{\epsilon}$

$X_1,\dots,X_n \sim N(0,1)$ $\quad P(|\bar{X}_n|>\epsilon) \leq \dfrac{2}{\sqrt{n}\epsilon} e^{-n^2\epsilon/2} \leq e^{-\frac{n\epsilon^2}{2}}$ large n

Markov:- (1st moment) $X>0,\ \mathbb{E}[|X|]<\infty$

$\forall \epsilon>0$

$P(X>\epsilon) \leq \dfrac{\mathbb{E}[X]}{\epsilon}$

Chebyshev: (1st/2nd moments) $\mu = \mathbb{E}[X] \quad \sigma^2 = \mathrm{var}(X)$

$P(|X-\mu| \geq \epsilon) \leq \dfrac{\sigma^2}{\epsilon} \qquad \forall \epsilon>0$

Hoeffding: $X_1,\dots,X_n$ - IID r.v.s. $\quad a \leq X_i \leq b \quad \mathbb{E}[X_i]=\mu$

$\forall \epsilon>0$

$P(|\bar{Y}_n-\mu| \geq \epsilon) \leq 2e^{\frac{-2n\epsilon^2}{(b-a)^2}}$

(*) Theorem 11 - WLLN proof

WLLN: $X_1,\dots,X_n$ IID, $\bar{X}_n \xrightarrow{P} \mu \quad \mu = \mathbb{E}[X_i] \quad \bar{X}_n - \mu = o_p(1)$

- suppose $\sigma<\infty$
- uses Chebyshev and property of sample mean (you were getting confused)

· sample mean:- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \mathbb{E}[\bar{X}_n]=\mu \quad \mathrm{Var}(\bar{X}_n) = \dfrac{\sigma^2}{n}$

consider that which you want to show converges in prob.

· we apply Chebyshev to the sample mean $\bar{X}_n$ (an r.v.)

$P(|\bar{X}_n-\mu|>\epsilon) = P(|\bar{X}_n-\mu|^2>\epsilon^2) \leq \dfrac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2}$ $\qquad \bar{X}_n \xrightarrow{P} \mathbb{E}[X]/\mu$

$\dfrac{\mathrm{var}(\bar{X}_n)}{\epsilon^2} = \dfrac{\sigma^2}{n\epsilon^2} \xrightarrow{n\to\infty} 0$; hence $P(|\bar{X}_n-\mu|>\epsilon) \xrightarrow{n\to\infty} 0$

(NT): Standardisation is a key point

- Allows us, when used appropriately, to find limiting distributions to non-degenerate distn.

- As $\bar{X}_n \xrightarrow{P} \mu$, intuitively expect $(\bar{X}_n - \mu) \xrightarrow{P} 0$ (i.e. to a degen distn)

- In terms of degenerate distn, this means that

$$(\bar{X}_n - \mu) \xrightarrow{P} X \quad \text{where} \quad P(X=0)=1$$

- So we use appropriate scaling to construct the sequence of r.v.s. $Z_1, Z_2, Z_3, \ldots$ from $\bar{X}_1, \bar{X}_2, \bar{X}_3 \ldots$

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{Var(\bar{X}_n)}} = \frac{(\bar{X}_n - \mu)}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

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

$$Z_n = \frac{(\bar{X}_n - \mu)}{\sqrt{Var(\bar{X}_n)}} = \frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma} \xrightarrow[n\to\infty]{d} Z \sim N(0,1)$$

(by collecting many samples)

(*) For $X_1, \ldots, X_n$, can construct sequence of sample means, rescale these. The sequence of CDFs of these normalised sample means converges to the CDF of a standard Normal $Z \sim N(0,1)$.

(*) That means as we get more and data, and only know mean and variance of these datapoints from an unknown distn P, we can approximate

(*) This is quite theoretical to remember; informally; $\bar{X}_n \,\hat{\sim}\, N(\mu, \sigma^2/n)$

(NT). Allows us to compute probabilities (see notes)

$$P(a < \bar{X}_n < b) \approx P\left(\frac{\sqrt{n}(a-\mu)}{\sigma} < Z < \frac{\sqrt{n}(b-\mu)}{\sigma}\right)$$ ↳ on sample mean (via rescaling)

(NT). We are approximating probability statements about the sample mean $\bar{X}_n$; not the ~~ex~~ underlying r.v. (!)

(*) Lemma 14:

- CDF and MGF completely characterise a distn

(A1): o/s (2): This proof (my notes) confuses me at the beginning

Notes: - assume X_i has $\mathbb{E}[X_i] = \mu = 0$ $Var(X_i) = \sigma^2 = 1$

$\Rightarrow \sigma = 1$

(*) Take sample mean of X_i: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and rescale to get a (sequence) of r.v.s.:-

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}(\bar{X}_n - 0)}{1} = \sqrt{n}\bar{X}_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$$

(A2): clarity on

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

- so if $a_n \to a$ then $\left(1 + \frac{a_n}{n}\right)^n \longrightarrow e^a$ as $n \to \infty$.

- Here, $a_n = \frac{t^2}{2} + \underbrace{\frac{t^2}{3!n^{1/2}}\psi'''(0) + \ldots}_{\to 0 \text{ as } n\to\infty}$

$$a = \frac{t^2}{2}$$

(*) T.15 Proof

- CLT with sampling variance (rather than true variance)

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

- Go over this - lots of steps

- Had difficulty understanding the similarities/differences between short and long version of proof.

(*) Key step is to decompose the sample variance normalised sample mean into a term involving the true variance.

Define:-

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right) \cdot \left(\frac{\sigma}{S_n}\right) = Z_n W_n$$

(*) Note this is a sequence of r.v.s. (as we consider $n \to \infty$)

(*) The key step:- (rest is details); and sense in which Slutsky's theorem is used.

i) $Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$ (via CLT)

ii) $W_n \xrightarrow{P} 1$; and hence $W_n \xrightarrow{d} 1$ as $\xrightarrow{P} \Rightarrow \xrightarrow{d}$

∴ use Slutsky:- $Z_n \xrightarrow{d} Z$, $W_n \xrightarrow{d} 1$ $\Rightarrow T_n = Z_n W_n \xrightarrow{d} 1 \cdot Z$

(*) Key part of extended proof is showing that $W_n \xrightarrow{P} 1$

- Main steps are:-

i) show $S_n^2 \xrightarrow{P} \sigma^2$ (by showing $R_n^2 \xrightarrow{P} \sigma^2$)

ii) show $\frac{S_n}{\sigma} \xrightarrow{P} 1$

iii) show $\frac{\sigma}{S_n} = W_n \xrightarrow{P} 1$

i) show $S_n^2 \xrightarrow{P} \sigma^2$

- $R_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ (i.e. biased sample variance)

(*) (N): Got very confused by following observation:-

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

- We will show this here:-

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

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_i^2 + \frac{1}{n}\sum_{i=1}^{n}(-2X_i\bar{X}_n + \bar{X}_n^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_i^2 + \frac{\bar{X}_n}{n}\sum_{i=1}^{n}(-2X_i + \bar{X}_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_i^2 + \frac{\bar{X}_n}{n}\left\{-2\sum_{i=1}^{n}X_i + \sum_{i=1}^{n}\bar{X}_n\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_i^2 + \frac{\bar{X}_n}{n}\left\{-2n\bar{X}_n + n\bar{X}_n\right\}$$

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

Hence $R_n^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2$

- set $Y_i = X_i^2$
- via WLLN:-

$$\frac{1}{n}\sum_{i=1}^{n}X_i^2 = \frac{1}{n}\sum_{i=1}^{n}Y_i \xrightarrow{P} \mathbb{E}[Y_i] = \mathbb{E}[X_i^2] = \mu^2 + \sigma^2$$

- via WLLN:-

$$\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} \mu$$

- As $g(t) = t^2$ is continuous, use <u>continuous mapping theorem</u>:-

$g()$ - continuous

$X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$

$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$

$$\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2 \xrightarrow{P} \mu^2$$

- using 1.8. (sum of conv. in prob.)

$$R_n^2 \xrightarrow{P} (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$$

Hence $R_n^2 \xrightarrow{P} \sigma^2$

- As $S_n^2 = \left(\frac{n}{n-1}\right) R_n^2$ (bias correction)
(unbiased sample var, biased s.v.)

As $R_n^2 \xrightarrow{P} \sigma^2$ and $\left(\frac{n}{n-1}\right) \longrightarrow 1$ (meaning $\frac{n}{n-1} \xrightarrow{P} 1$)

we have $S_n^2 \xrightarrow{P} \sigma^2$ (via prod. of conv in prob)

---

ii) show $\frac{S_n}{\sigma} \xrightarrow{P} 1$

$g(t) = \sqrt{t}$ is contin. (for $t \geq 0$); via CM,

$S_n^2 \xrightarrow{P} \sigma^2 \Rightarrow S_n \xrightarrow{P} \sigma$

$g(t) = \frac{t}{\sigma}$ is continuous; via C.M.

$\frac{S_n}{\sigma} \xrightarrow{P} 1$

---

iii) show $\frac{\sigma}{S_n} = W_n \xrightarrow{P} 1$

$g(t) = 1/t$ cont. for $t > 0$; via CM

$\frac{\sigma}{S_n} \xrightarrow{P} 1$.

last step (to link to above); $\frac{\sigma}{S_n} \xrightarrow{P} 1 \Rightarrow \frac{\sigma}{S_n} \xrightarrow{d} 1$.

$\boxed{E}$

1.16. Berry-Esseen

- Recall two ways of viewing CDFs
  i) Formal: convergence in distn (i.e. limiting sequence of CDFs) of rescaled/normalised sample mean approaches std. Normal.
  ii) informal: sample mean of a distn $\bar{X}_n$ is approximately Normal with mean μ, variance σ^2/n (i.e. in terms of mean, variance of underlying r.v.s.)

- Recall convergence in distn connot. of approximation of probabilistic statements about a sequence of rescaled sample means ~~can be~~ via prob. statements of std. Normal.
- Berry-Esseen requires 3rd moments to be finite; and 3rd moment
- A statement of maximum absolute diff. between true CDF of a rescaled sample mean and its Normal approx via CLT.
- Gives a sense of how fast the CLT Normal approx converges to true CDF.

5. Delta Method

- Your written notes mask the generality of this theorem.
- The delta method applies to an arbitrary sequence of random variables Y_n, that converge in distn to a standard Normal i.e. asymptotically normal.
- The rescaled, and normalised (sequence of) sample means, that is $\sqrt{n}(\bar{X}_n-\mu) \xrightarrow{d} N(0,\sigma^2)$ and $\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{d} N(0,1)$ are examples of sequences of r.v.s which are asymptotically normal (via CLT).
- Hence ~~wr~~ LW notes are at a greater level of generality
- And it looks like a key use case of this is to find the limiting distribution of a smooth function of an estimator.

(*) Example 19

$X_1,\dots,X_n$ IID, $\mathbb{E}[X_i]=\mu$, $Var[X_i]=\sigma^2$

$\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{d} N(0,1)$ $W_n = e^{\bar{X}_n}$

$g(s)=e^s$

$g'(s)=e^s$

- Note equivalence of:-

$$\sqrt{n}(W_n - e^{\mu}) \xrightarrow{d} N(0,\sigma^2(e^{\mu})^2)$$

$$\frac{\sqrt{n}(W_n-e^{\mu})}{\sigma|e^s|} = \frac{\sqrt{n}(W_n-e^{\mu})}{\sigma e^s} \xrightarrow{d} N(0,1)$$

✓ Q5: Get more fluency with how you move between these two reps.

→

- Notation to denote Normal:
(converg in distn) :-

- $Z_n \approx N(0,1)$
- $\bar{X}_n \approx N(\mu, \sigma^2/n)$
- $\bar{X}_n - \mu \approx N(0, \frac{\sigma^2}{n})$
- $\sqrt{n}(\bar{X}_n - \mu) \approx N(0, \sigma^2)$
- $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \approx N(0,1)$

(*) Multivariate Delta:-

For $\sqrt{n}(\underline{\bar{X}} - \underline{\mu}) \xrightarrow{d} N(0, \Sigma)$

- $\sqrt{n}(g(\underline{\bar{X}}) - g(\underline{\mu})) \xrightarrow{d} N(0, \nabla_{\mu}^T \Sigma \nabla_{\mu})$

$\nabla_{\mu} = \nabla_y g(y)\big|_{y=\mu}$

(*) - Note:- (from Casella & Berger)

- not emphasised by L.W.; but in notes, and you forgot.
- (underpins much of chapter and I did not appreciate it fully enough)
- in general, the various convergence results presented do not require an IID assumption on the sequence of r.v.s (often statistics) being considered.

E.g. A sequence of r.v.s. $X_1, X_2,$ converges in prob.

to an r.v. X if for every $\epsilon > 0$:-

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

- $X_1, X_2, \ldots$ and other definitions are typically NOT IID.
- For WLLN:-

$\bar{X}_n \xrightarrow{p} \mu \rightarrow P(|\bar{X}_n - \mu| \geq \epsilon) \xrightarrow{n\to\infty} 0$

- It is the sequence of statistics (sample means) $\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4$ that is converging in probability to the true pop mean of underlying IID r.v.
- However, sample mean is constructed from IID r.v.s.

(*) - Distribution of $\bar{X}_n$ (sampling distri / distri of sampling estimator of mean concentrates around true/pop. mean as n gets lage).

- The A-level definition of CLT (of which yours should be able to accomodate)
- When sample size is sufficiently lage, the distribution of the sample means is approximately Normal.