this time: Markov, Chebyshev,

next time: WLLN, CLT

STAT 131
30 May 20

Catch-up lecture

(1)

$Z_1, Z_2, \ldots$ converges in probability to some constant b iff

for any $\epsilon > 0$ $P(|Z_n - b| < \epsilon) \uparrow 1$

fixed

as $n \uparrow$

Def $Z_n \xrightarrow{P} b \leftrightarrow (Z_n, n = 1, 2, \ldots)$ is consistent for b

is a <u>consistent</u> sequence of estimators of b

population

all U.S. ~~families~~ households in 2019

$\mathcal{P}$

~~family~~ household income (actual)

$\uparrow$ $N = 129$ M $\downarrow$

mean $\boxed{\mu = ?}$

SD $\sigma = ?$

0

PDF of Y_i (pop. hist.)

sample

the observed households

~~SRS~~ (like) $\boxed{IID}$

h. income (\$)

Y_1 y_1

$\vdots$

Y_n y_n

$n = \underline{\underline{1{,}052}}$

$(\bar{Y}_n)$ mean $\bar{y}_n = \$90{,}357$

SD $\underline{\underline{s}} = \$61{,}092$

hyp. IID

sample histogram

$n = 1{,}052$

$\vdots$

mean $\bar{Y}_n = ?$ (ex. \$88,550)

freq. stat. inf.

~~repeated~~ sim. like ~~data~~ ~~those~~ +

(2)

possible values of $\bar{Y}_n$

90,357

88,550

$\vdots$

$\uparrow$ $M \to \infty$ $\downarrow$

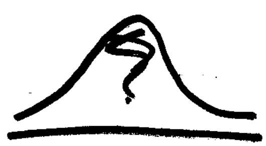
long run mean

$E_{IID}(\bar{Y}_n) = \mu$

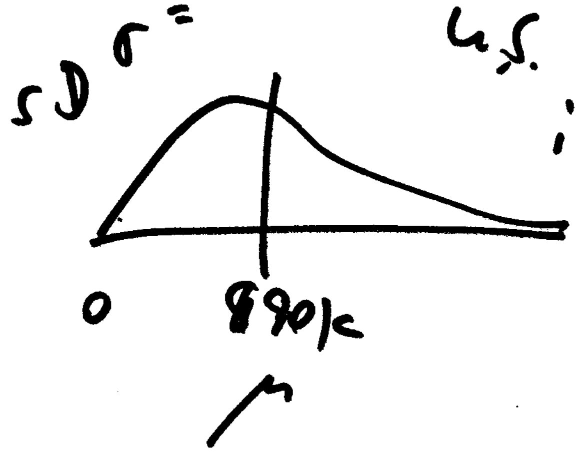
long run SD

$SD(\bar{Y}_n) = SE(\bar{Y}_n) = \sigma / \sqrt{n}$

long run hist.



to estimate how good a single $\bar{Y}_n$ is as a guess for μ, ironically we have to imagine getting a lot of $\bar{Y}_n$s



u.S. household income in 2019

CLT:

as long as $\sigma < \infty$, as $n \uparrow$ PDF of $\bar{Y}_n \to$ Normal

inferential summary

unknown pop. quantity of main interest	μ = pop. mean h. income in U.S. in 2019
estimate of μ	$\bar{y}$ = \$90,357
give or take for $\bar{y}$ or est. of μ	

on basis of this presumed-like-IID sample of size $n = 1,052$ from pop (P), we think that μ is around $\bar{y}$ = \$90k, give or take about ____ statistical prediction | If I choose one of the sampled households at random, I expect to see that their 2019 h. income is about $\bar{y} \doteq \$90k$, give or take about $s \doteq 861k$

RMSE $(\hat{y})$

How good is $\bar{Y}_n$ as estimate of μ?

want $|\bar{Y}_n - \mu|$ to be small

freq. μ = fixed unknown constant

$\bar{Y}_n$ is a r.v.

want $P_F(|\bar{Y}_n - \mu| \leq T)$ to be big

F ← repeated sampling

CLT Pop. SD has to be finite

$X_i \overset{IID}{\sim}$ $E(X_i) = \mu$, $SD(X_i) = \sigma < \infty$

$(i = 1, \dots, n)$ $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

part 1

For any PDF of X_i with $\sigma < \infty$, as $n \uparrow$ the PDF of $\bar{X}_n \to$ normal

---

part 2 The closer the (population) PDF of X_i is to normality to begin with, the ~~[illegible]~~ smaller n needs to be to get a good normal approximation to the PDF of $\bar{X}_n$

---

part 3 if $PDF(X_i) = N \to PDF(\bar{X}_n) = N$ even for $n = 1, 2, \dots$