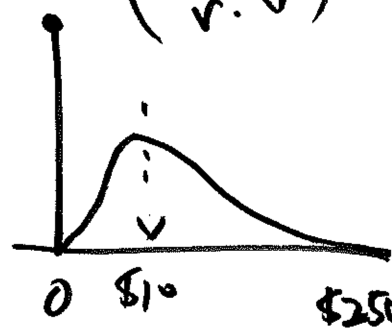
AMS 206 24 Jan 19

①

this time: frequentist inference (Neyman, Fisher)

next time: then Bayesian inference

(mixed r.v.)



monthly exp. of Ed.

want

$$P_F(|\underset{\uparrow \text{ my}}{\hat\theta} - \theta| \le \underset{\uparrow \text{ tolerance}}{T}) = \text{big}$$

$$\left(\text{Expected value of } \hat\theta\right) = EV_{\hat\theta} \text{ of } = \overset{\text{(AMS 131)}}{E_{IID}}(\overset{\text{r.v.}}{\underline{\hat\theta}}) = \theta$$

$$\left(\text{standard error of } \hat\theta\right) = SE_{\hat\theta} \text{ of } = SE_{IID}(\hat\theta) = \frac{\sigma}{\sqrt{n}}$$

$(i = 1, \dots, n)$

$Y_i \stackrel{IID}{\sim} E(Y_i) = \mu$, $SD(Y_i) = \sigma$

$V(Y_i) = \sigma^2 < \infty$

$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$

$$\boxed{E_{IID}(\bar{Y}_n) = \mu} \leftarrow \bar{Y}_n = \hat\theta \text{ is unbiased}$$

$$V(\bar{Y}_n) = \frac{\sigma^2}{n}$$

$$\sqrt{V(\bar{Y}_n)} = SE(\bar{Y}_n) \text{ for } \theta$$

Case Study 2 | frequentist inference (Neyman) ②

(imaginary) repeated-sampling data set

population

p^*

sample (2015–2018)

the observed AMI Medicare patients at JH

all possible $\hat{\theta}$

died?

$$\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix}$$

$N = ?$ $(>> n)$

mean $\theta = ?$

SD $\sigma = \sqrt{\theta(1-\theta)} = ?$

pop. PMF

$\boxed{1 = \text{yes} \\ 0 = \text{no}}$ died?

(actual) like IID

I_1 $Y_1 = 0$

I_2 $Y_2 = 1$

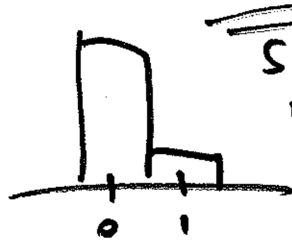
$\vdots$ y_i

I_n $Y_n = 0$

$n = 403$

mean $\hat{\theta}_0 = \frac{s}{n} = \frac{72}{403} \doteq 0.18$

hyp. IID



sample hist.

$$\begin{bmatrix} 0.18 \\ 0.19 \\ 0.185 \\ \vdots \end{bmatrix}$$

$M \to \infty$

$(M \to \infty)$

long run mean

Expected value of $\hat{\theta} = \theta$

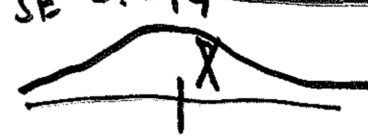
estimated SD

long run

estimated standard error of

$\hat{SE}(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \doteq 0.019$

$\hat{SE}$ 0.019



θ

long run hist.

$$\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix}$$ $n = 403$

$$\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix}$$ $n = 403$

mean $\hat{\theta} = ?$ (ex. 0.19)

mean $\hat{\theta} = 0.185$

$\left\{ \begin{array}{ll} (Y_i \mid \mu\, \sigma\, \mathcal{B}) \overset{IID}{\sim} & E(Y_i) = \mu \\ (i = 1, \ldots, n & SD(Y_i) = \sigma < \infty \\ \left(\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i\right) & \end{array} \right\}$

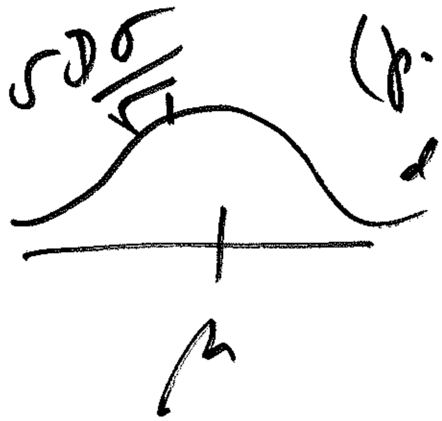
$$\rightarrow \quad \left(\frac{\bar{Y}_n - \mu}{\sigma / \sqrt{n}} \right) \xrightarrow{D} N(0, 1)$$

CLT ? ①

as long as n is large & σ is finite, R.S. dist. of $\bar{Y}_n$ under IID sampling will be close to $N\left(\mu, \frac{\sigma^2}{n}\right)$

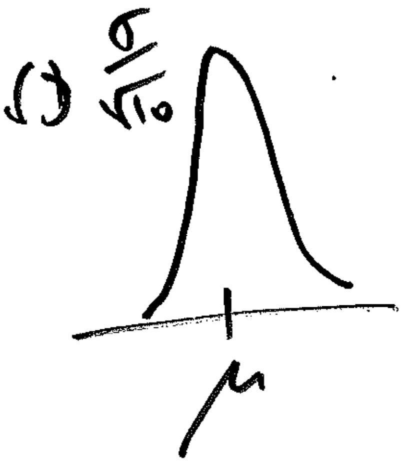
② the closer pop. dist. is to normal to begin with, smaller n needs to be to get normality of $\bar{Y}_n$

3 if pop dist is normal, $\bar{X}_n$ is normal $\forall n$, even $n=1$



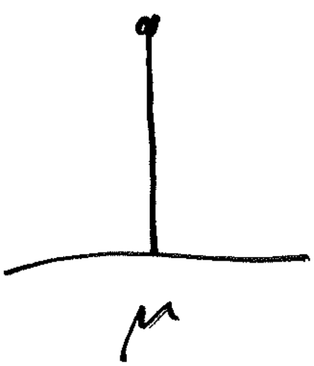
(p.p.)

dist. of $\bar{X}_1 = \frac{1}{1}\sum_{i=1}^{1} X_i$



dist of $\bar{X}_n$, $n = 10$

$\downarrow$ $n \to \infty$



inferential summary

(5)

(unknown) pop. quantity of main interest	θ = 30-day pop. mortality rate
(point) estimate of θ (in absence of external info.)	$\hat{\theta} = \frac{s}{n} \doteq 0.18$
give-or-take for $\hat{\theta}$ as est. of θ	$\leftarrow \hat{SE}(\hat{\theta}) \doteq 0.019$
95% interval for θ	$\hat{\theta} \pm 1.96\,\hat{SE}(\hat{\theta}) = (0.141, 0.216)$

$$SE_{IID}(\hat{\theta}) = \sqrt{\frac{\theta(1-\theta)}{n}} = \sqrt{V_{RS}(\hat{\theta})} = \sqrt{V_F(\hat{\theta})}$$

$$\hat{SE}_{IID}(\hat{\theta}) \triangleq \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = \sqrt{\frac{\left(\frac{72}{403}\right)\left(\frac{403-72}{403}\right)}{403}}$$

is defined to be

$$\doteq 0.019$$