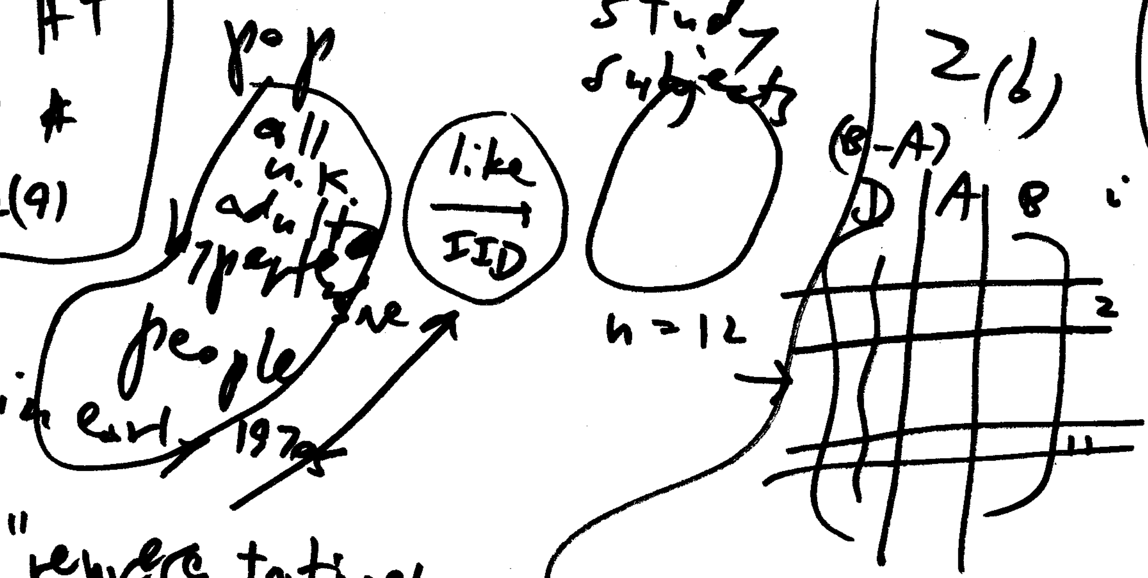
THT
3 #
2(9)

7

STAT 131
11 Jun 20

DD extra office 1.5-hour session (+)



"representatively chosen"

$$\bar{D}_n = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad D_i = B_i - A_i$$ ①

$$V_{RM}(\bar{D}_n) = V_{RM}\left(\frac{1}{n}\sum_{i=1}^{n} D_i\right)$$

$$= \frac{1}{n^2} V_{RM}\left(\sum_{i=1}^{n} D_i\right) \overset{\textcircled{I}}{=} \frac{1}{n^2}\sum_{i=1}^{n} V_{RM}(\underline{D_i})$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V_{RM}(\underline{B_i} + (-1)\underline{A_i})$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left[V(B_i) + V(-A_i) + 2\,C(B_i, -A_i)\right]$$

$$V(\underline{X} + \underline{Y}) = V(\underline{X}) + V(\underline{Y}) + 2\,C(\underline{X}, \underline{Y})$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \Big[\sigma_B^2 + (-1)^2 \sigma_A^2$$

(*) $$- 2\, \underline{C(\underline{B_i}, \underline{A_i})} \Big]$$

$$\updownarrow =$$

$$\rho_{AB} \doteq \frac{C(A_i, B_i)}{\underset{\overset{\uparrow}{\sigma_A}}{\boxed{SD(A_i)}} \cdot \underset{\overset{\uparrow}{\sigma_B}}{\boxed{SD(B_i)}}} \rightarrow \underline{C(A_i, B_i) = \rho_{AB}\, \sigma_A\, \sigma_B}$$

(*) $$= V_{RM}(\bar{D}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \left(\sigma_A^2 + \sigma_B^2 - 2 \rho_{AB}\, \sigma_A\, \sigma_B \right)$$

$$= \frac{n(\quad)}{n^2} = \frac{\sigma_A^2 + \sigma_B^2 - 2\rho\,\sigma_A\sigma_B}{n}$$

3(b)(i) $\text{Support}(X) = \{0, 1, 2, \ldots\}$

$$p_X(x \mid s, \theta) = \binom{s+x-1}{x}\theta^{s}(1-\theta)^{x} \cdot I_{\{0,1,2,\dots\}}(x)$$

N = # of trials needed to see $\textcircled{s}$ successes

↗ fixed known

ex. $s = 2$

FFFSFS $\left\{ \begin{array}{l} N = 6 \\ X = 4 \end{array} \right.$

X = # failures needed to see $\textcircled{s}$ successes

$$\boxed{\textcircled{N} = X + s}$$

~~$n = x + s$~~

$x = n - s$

$$p_N(n \mid s, \theta) = \binom{\cancel{s+}\, n \cancel{x} - 1}{n-s} \cdot \theta^{s} \cdot (1-\theta)^{n-s} \, I_{\{s, s+1, \dots\}}(n)$$

$$p_X(x \mid s, \theta) = \binom{s+x-1}{x}\theta^{s}(1-\theta)^{x} \cdot I_{\{0,1,2,\dots\}}(x)$$

$$\binom{5}{2} = \frac{5!}{2!\,3!} = \underline{\frac{5!}{3!\,2!}} = \binom{5}{3}$$

$$\binom{m}{k} = \binom{m}{m-k} \qquad m = 1, 2, \dots \quad k = 0, 1, \dots, m$$

$$\binom{\overset{m}{\boxed{n-1}}}{\underset{\underset{k}{\uparrow}}{\boxed{n-5}}} = \binom{n-1}{\underset{\underset{(n-1)-(n-5)}{\nearrow}}{m-k}} = \binom{n-1}{5-1} \checkmark$$

$$f_N(n|5,\theta) = \binom{n-1}{5-1}\theta^5(1-\theta)^{n-5} \cdot I_{\{5,5+1,\dots\}}(n)$$

neg. bin

$$\rightarrow f_S(s|n,\theta) = \binom{n}{s}\theta^s(1-\theta)^{n-s} \cdot I_{\{0,1,\dots,n\}}(s)$$

bin.

$$f_{X}(x|r,p) = \binom{r+x-1}{x} p^{r}(1-p)^{x} \cdot I_{\{0,1,\dots\}}(x)$$

DS def. of neg. bin.

who	count	prob.
us	s	θ
DS	r	p

	DS	us
DS: $E(X) =$	$\frac{r(1-p)}{p}$	$\frac{s(1-\theta)}{\theta}$
$V(X) =$	$\frac{r(1-p)}{p^2}$	$\frac{s(1-\theta)}{\theta^2}$

$$\boxed{N = X + s}$$

$$E(N) = E(X+\underline{\underline{s}}) = s + E(X) = s + \frac{s(1-\theta)}{\theta}$$

$$= \frac{s\theta + s - s\theta}{\theta} = \frac{s}{\theta}$$

$$V(N) = V(X+s)$$

$$= V(X) = \frac{s(1-\theta)}{\theta^2}$$

$\hat{\theta}_B = \frac{S}{n}$ ← r.v. : $E(\hat{\theta}_B) = \theta$, $V(\hat{\theta}_B) = \frac{\theta(1-\theta)}{n}$

const.

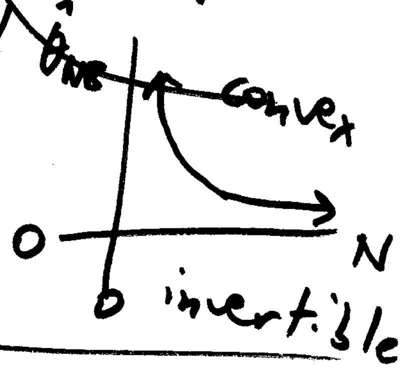
(easier) exact unbiasedness

$\hat{\theta}_{NB} = \frac{s}{N}$

(harder)

Delta method

$\hat{\theta}_{NB} = g(N) = \frac{s}{N}$



$E(\hat{\theta}_{NB}) = E\left(\frac{s}{N}\right)$

$\boxed{E[g(N)] \doteq} \; g[E(N)]$ ①

$\boxed{E(N) = \frac{s}{\theta}}$ ② $= \frac{s}{\frac{s}{\theta}} = \boxed{\theta}$

$E(\hat{\theta}_{NB}) \doteq \theta$

approximate unbiasedness

$V[g(N)] \doteq \left\{g'[E(N)]\right\}^2 \cdot V(N)$

$V(N) = \frac{s(1-\theta)}{\theta^2}$

$g(N) = \frac{s}{N} = s \cdot N^{-1}$

$g'(N) = -\frac{s}{N^2}$

$g'[E(N)] = \frac{-s}{[E(N)]^2} = \frac{-s}{(\frac{s}{\theta})^2} = -\frac{\theta^2}{s}$

so $V[g(N)] = V(\hat{\theta}_{NB})$

$$\doteq \left(\frac{-\theta^2}{s}\right)^2 \frac{s(1-\theta)}{\theta^2} = \frac{\theta^4 s(1-\theta)}{s^2 \theta^2} = \frac{\theta^2(1-\theta)}{s}$$

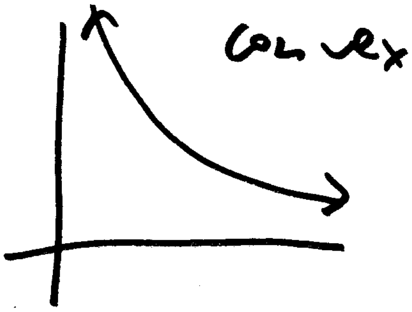
$$= \frac{\theta(1-\theta)}{\left(\frac{s}{\theta}\right)} = \frac{\theta(1-\theta)}{\checkmark\, E(N)}$$

$$\underline{\underline{SE}}(\hat{\theta}_{NB}) \doteq \sqrt{\frac{\theta(1-\theta)}{E(N)}}$$

$$V(\hat{\theta}_B) = \frac{\theta(1-\theta)}{n}$$

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

Jensen's ≠



if g is convex then

$$E[g(X)] \geq g[E(X)]$$

with equality only if $g(x) = a + bx$

$g(x) = x^2 \to$ convex

$V(X) = E(X^2) - (E(X))^2 \geq 0$ so $E(X^2) \geq (E(X))^2$

so here $E(\hat{\theta}_{NB}) = E[g(N)] > g[E(N)]$

$\boxed{g(N) = \frac{S}{N}}$ $\quad g[E(N)] = \frac{S}{E(N)} = \frac{S}{\frac{S}{\theta}} = \theta$, i.e.,

$\boxed{E(N) = \frac{S}{\theta}}$ $\quad \underline{\underline{E(\hat{\theta}_{NB})}} > \theta$, i.e.,

$\hat{\theta}_{NB}$ is biased on the high side

can show that

$$\boxed{E\left(\frac{S-1}{N-1}\right) = \theta}$$

$$\text{bias}(\hat{\theta}_{NB}) \triangleq E(\hat{\theta}_{NB}) - \theta > 0$$

so

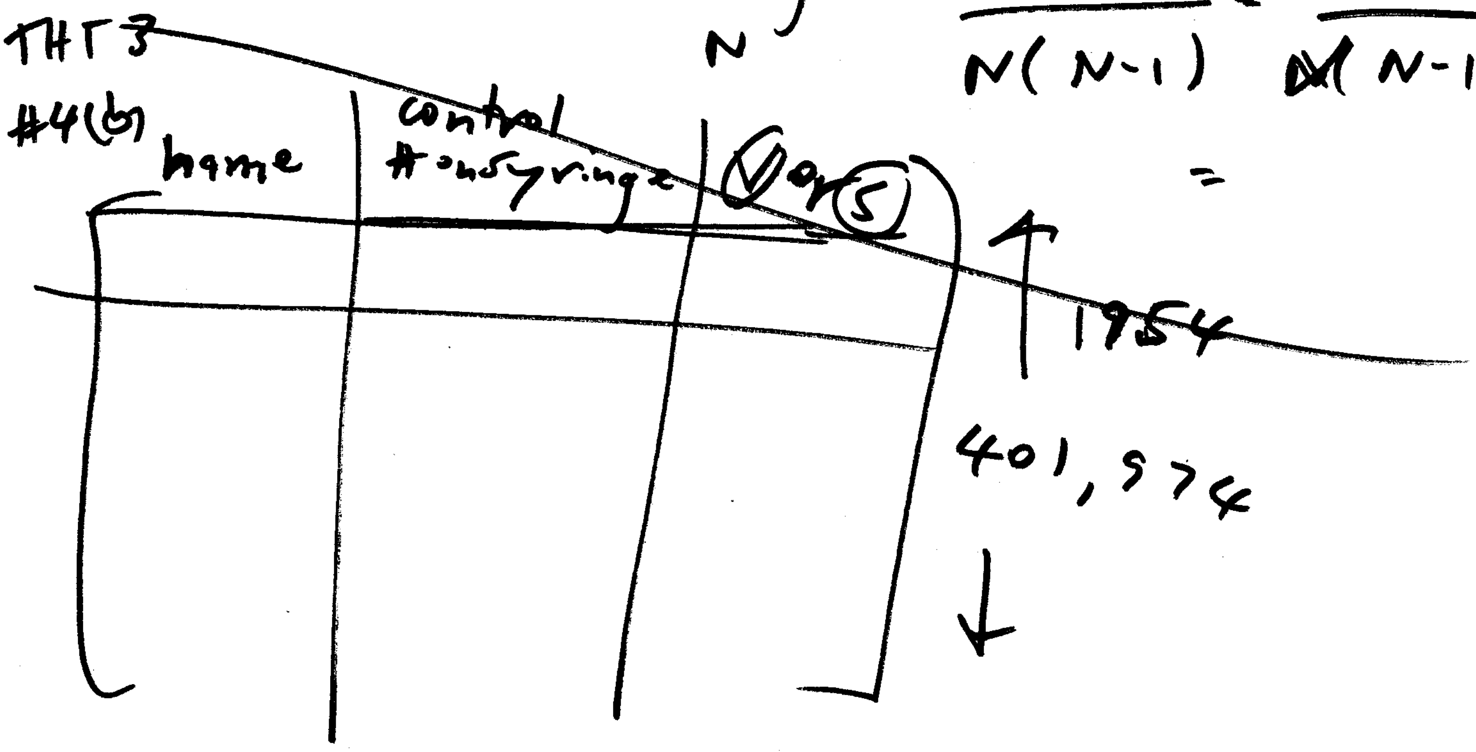
$$E(\hat{\theta}_{NB}) - \theta = E\left(\frac{S}{N}\right) - E\left(\frac{S-1}{N-1}\right)$$

$$= E\left(\frac{S}{N} - \frac{S-1}{N-1}\right) = E\left(\frac{\cancel{SN} - S - \cancel{SN} + N}{N(N-1)}\right)$$

$$\text{bias}(\hat\theta_{NB}) = E(\hat\theta_{NB}) - \theta \qquad \left(\frac{S}{N} \xrightarrow{P} \theta\right) \quad \therefore S \doteq N\theta$$

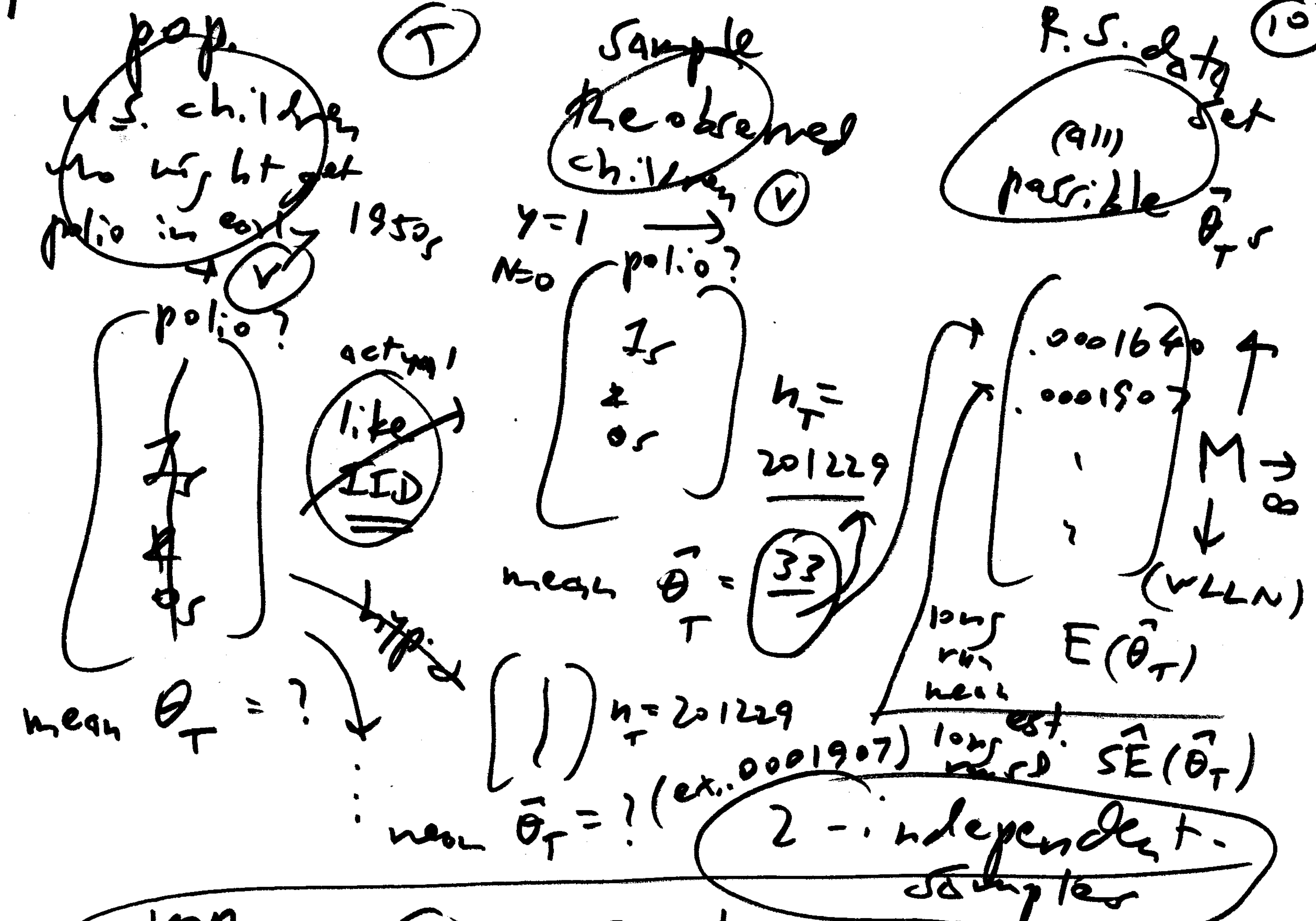
$$= E\left[\frac{N-S}{N(N-1)}\right] \; ; \; \frac{N-S}{N(N-1)} = \underline{\underline{O}}_P\left(\frac{1}{N}\right)$$

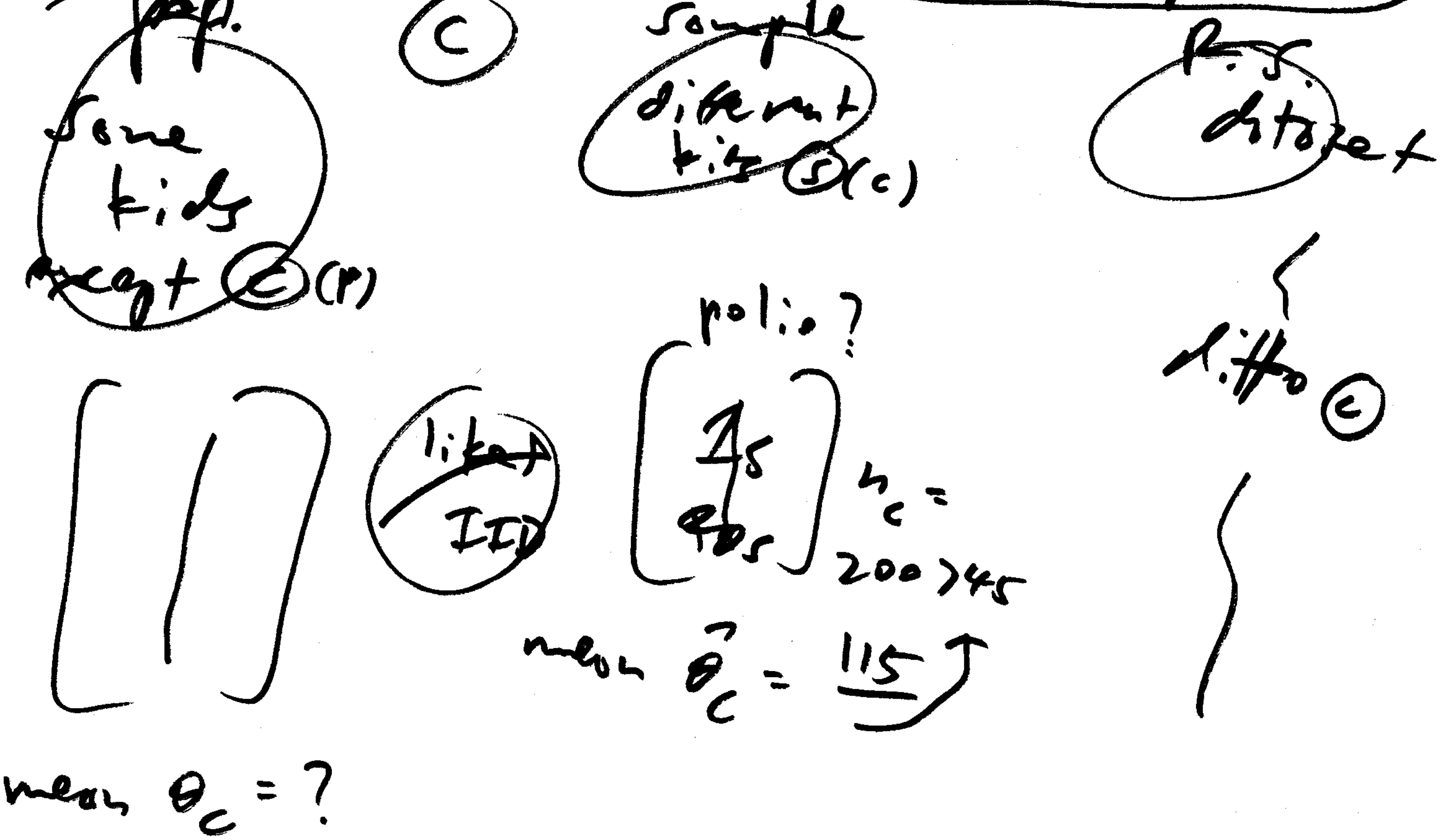
$$\text{for large } N \quad \frac{N-S}{N(N-1)} = \frac{N - N\theta}{N(N-1)} = \frac{\cancel{N}(1-\theta)}{\cancel{N}(N-1)} =$$



4(c) practsig? | statsig? | need prob./stat. inf. model

yes


pop.
T
U.S. children who might get polio in early 1950s
V
polio?
1s & 0s
actual
~~like IID~~
~~hyp.~~ d
mean θ_T = ?
sample
the observed children
V
Y=1
N=0
polio?
1s & 0s
n_T = 201229
mean $\hat{\theta}_T$ = 33
n_T = 201229
mean $\hat{\theta}_T$ = ? (ex. .0001907)
R.S. data set
(all) possible $\hat{\theta}_T$'s
.0001640
.0001907
M → ∞
(WLLN)
long run mean
E($\hat{\theta}_T$)
est. long run SD
$\hat{SE}(\hat{\theta}_T)$
2-independent samples


pop.
C
some kids except C (P)
~~like~~ IID
mean θ_C = ?
sample
different kids S (C)
polio?
1s 0s
n_C = 200745
mean $\hat{\theta}_C$ = 115
R.S. dataset
ditto C

inferential summary

pop	unknown pop qty. of main interest	$(\theta_C - \theta_T)$ = pop. mean diff. in polio incidence (C − T)
sample	estimate of $(\theta_C - \theta_T)$	$(\hat{\theta}_C - \hat{\theta}_T) = \begin{array}{r} 0.0005729 - \\ \underline{0.0001640} \\ 0.0004089 \end{array}$
R.S. & target	give or take for $(\hat{\theta}_C - \hat{\theta}_T)$ as est. of $(\theta_C - \theta_T)$	$\widehat{SE}(\hat{\theta}_C - \hat{\theta}_T) \doteq 0.000$[illegible]$6055$
	99.9% CI for $(\theta_C - \theta_T)$	(.00021, .00061)

$$E(\hat{\theta}_T) = \theta_T$$

$\underline{\underline{IID}}$

$$E_{IID}(\hat{\theta}_C) = \theta_C$$

$$E(\hat{\theta}_C - \hat{\theta}_T) = \theta_C - \theta_T$$

$$SE(\hat{\theta}_C - \hat{\theta}_T) = \sqrt{V_{IID}(\hat{\theta}_C - \hat{\theta}_T)}$$

$$V_{\underline{\underline{IID}}}(\hat{\theta}_C - \hat{\theta}_T) \overset{\textcircled{T}\text{ vs. }\textcircled{C}}{\overset{\textcircled{I}}{=}} V(\hat{\theta}_C) + V(-\hat{\theta}_T)$$

$$= V(\hat{\theta}_C) + V(\hat{\theta}_T)$$

$$V(\hat{\theta}_T) = V\left(\frac{S_T}{n_T}\right) = \frac{1}{n_T^2} \cdot \theta_T(1-\theta_T) \cdot n_T$$

$$= \frac{\theta_T(1-\theta_T)}{n_T}$$

$$V(\hat{\theta}_C) = ?$$

$$\frac{\theta_C(1-\theta_C)}{n_C}$$

$$(S_T \mid \theta_T) \sim \text{Binomial}(n_T, \theta_T)$$

S_T = # (T) kids with polio

$(S_T = 33)$

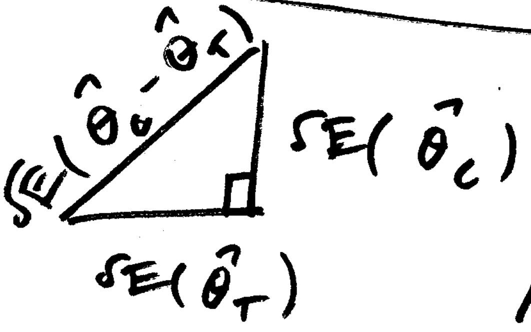
so $$V(\hat{\theta}_C - \hat{\theta}_T) = V(\hat{\theta}_C) + V(\hat{\theta}_T)$$

$$= \frac{\theta_C(1-\theta_C)}{n_C} + \frac{\theta_T(1-\theta_T)}{n_T}$$

and $$\hat{SD}(\hat{\theta}_C - \hat{\theta}_T) \doteq \hat{SE}(\hat{\theta}_C - \hat{\theta}_T)$$

$$= \sqrt{\left[\sqrt{\frac{\hat{\theta}_C(1-\hat{\theta}_C)}{n_C}}\right]^2 + \left[\sqrt{\frac{\hat{\theta}_T(1-\hat{\theta}_T)}{n_T}}\right]^2}$$

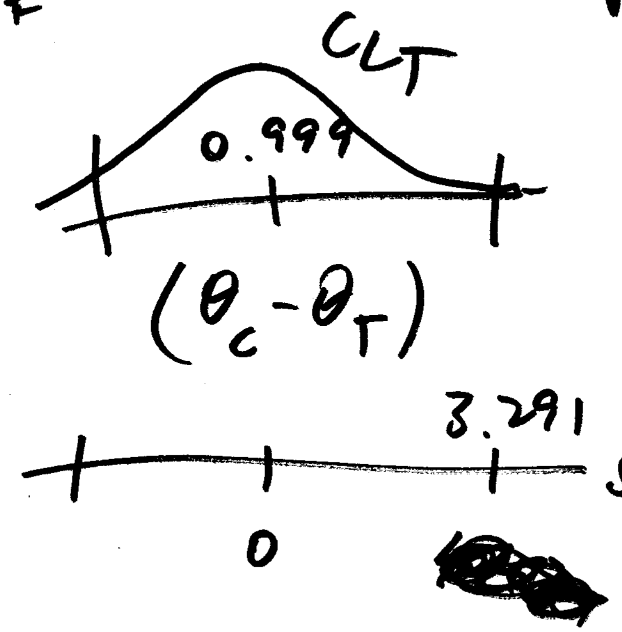
$$\hat{SE}(\hat{\theta}_T) = \sqrt{\frac{\hat{\theta}_T(1-\hat{\theta}_T)}{n_T}}$$



$$= .00006055$$

$\hat{SE}$ 0.00006

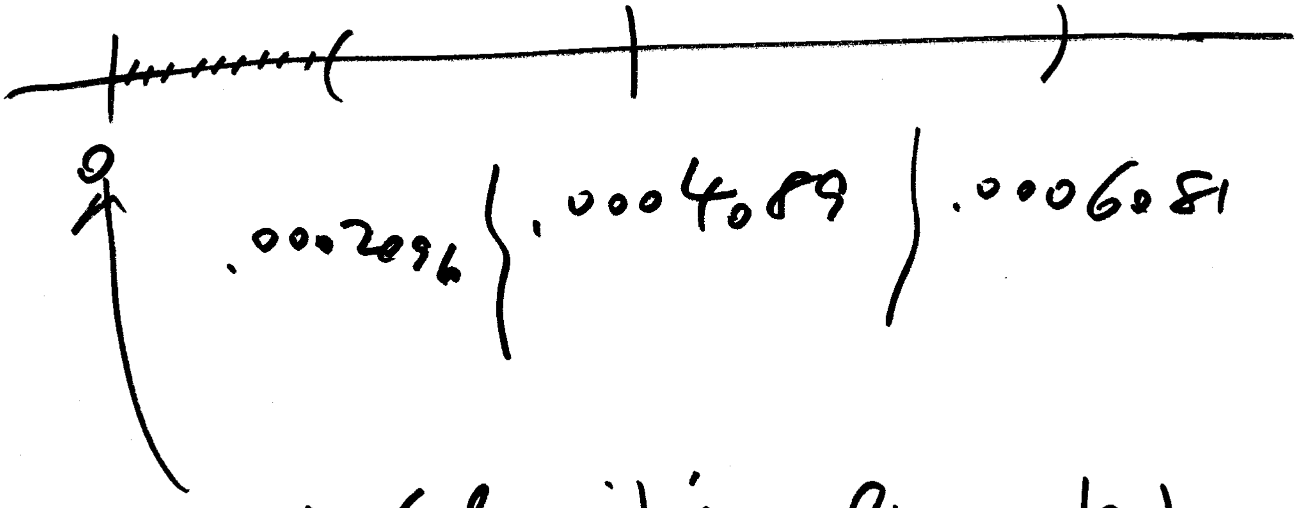
PDF/PMF of $(\hat{\theta}_C - \hat{\theta}_T)$



99.9% CI for $(\theta_C - \theta_T)$

$(\hat{\theta}_C - \hat{\theta}_T) \pm (3.291)\hat{SE}(\hat{\theta}_C - \hat{\theta}_T)$

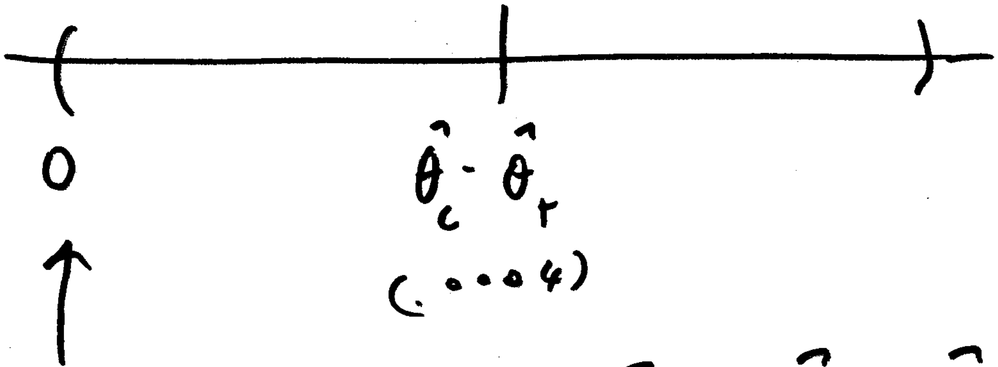
99.9% CI for $(\theta_C - \theta_T)$



null (devil's advocate) story (vaccine not effective)

0 is not in 99.9% CI, so dif. between $(\hat{\theta}_C - \hat{\theta}_T)$ and 0 is statsig

99.9% CI



$$(\hat\theta_c - \hat\theta_T) - \underline{3.291\,\hat{SE}(\hat\theta_c - \hat\theta_T)} = 0$$

$$n = \underset{\frac{n}{2}}{\underset{\downarrow}{n_c}} + \underset{\frac{n}{2}}{\underset{\downarrow}{n_T}}$$

$$(\hat\theta_c - \hat\theta_T) = 3.291\,\hat{SE}$$

$$\hat{SE}(\hat\theta_c - \hat\theta_+) = \frac{\hat\theta_c - \hat\theta_T}{3.291}$$

$$\parallel$$

$$\sqrt{\frac{2\cdot\hat\theta_c(1-\hat\theta_c)}{\frac{n}{\cancel{2}}} + \frac{2\cdot\hat\theta_T(1-\hat\theta_T)}{\frac{n}{\cancel{2}}}} =$$

$$\sqrt{\frac{2\left[\hat\theta_c(1-\hat\theta_c) + \hat\theta_T(1-\hat\theta_T)\right]}{n}} = \frac{\hat\theta_c - \hat\theta_T}{3.291}$$

$$\frac{2\left[\cdot\cdot\right]}{n} = \left(\frac{\vec{\theta}_c - \vec{\theta}_T}{3.291}\right)^2$$

$$n = \frac{3.291^2 \cdot 2 \cdot \left[\vec{\theta}_c(1-\vec{\theta}_c) + \vec{\theta}_T(1-\vec{\theta}_T)\right]}{\left(\vec{\theta}_c - \vec{\theta}_T\right)^2}$$