extra notes
11 Jan

p(A and D)

		truth (A) HIV +	truth (not A) HIV −	
what ELISA says	(D) HIV +	95	198	293
	(not D) HIV −	5	9702	9707
		100	9900	10000

P(D)

p(A)

$$P(A \mid D) = \frac{95}{293} = .32$$

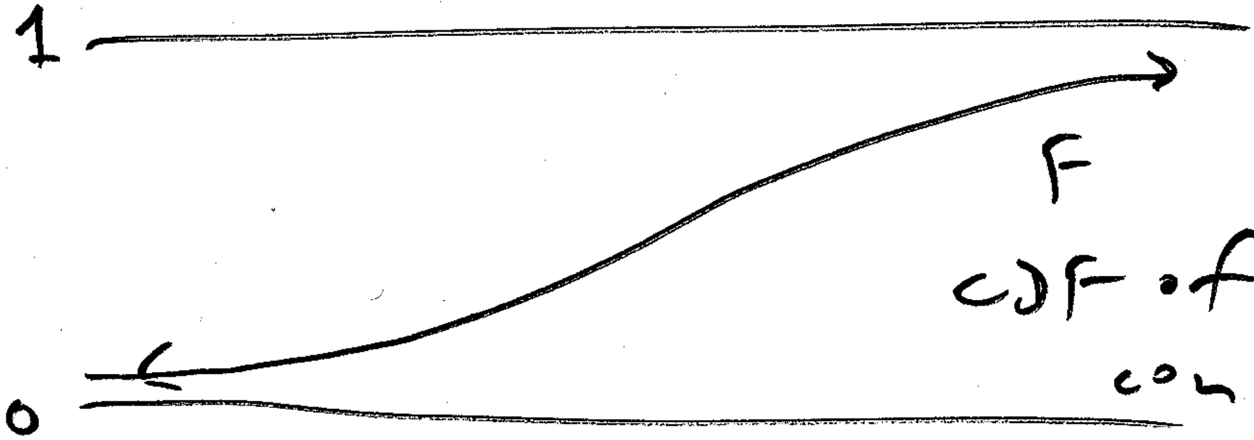
$P(A) = .01$ (prevalence)

$P(D \mid A) = .95$ (sensitivity)

$P(\text{not } D \mid \text{not } A) = .98$ (specificity)

$P(A \mid D) = ?$ expect to be big (close to 1)

$$= \frac{P(A)\,P(D \mid A)}{P(D)} = \frac{(.01)(.95)}{.0293} \doteq .32 \;!!$$



CDF of cont. r.v.

$$P(Y_1 = y_1 \overset{\text{and}}{,} Y_2 = y_2 \overset{\text{and}}{,} \dots, Y_n = y_n)$$

$$= p(y_1, \dots, y_n)$$

$$\overset{\text{indep}}{=} P(Y_1 = y_1) \cdot P(Y_2 = y_2) \cdots P(Y_n = y_n)$$

$$= \prod_{i=1}^{n} P(Y_i = y_i) = \prod_{i=1}^{n} p(y_i)$$

$$= \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}$$

$$= \theta^{y_1} (1-\theta)^{1-y_1} \, \theta^{y_2} (1-\theta)^{1-y_2} \cdots$$

$$\theta^{y_n} (1-\theta)^{1-y_n}$$

$$= \theta^{y_1 + \dots + y_n} (1-\theta)^{n - (y_1 + \dots y_n)}$$

$$p(y_1, \dots, y_n) = \theta^s (1-\theta)^{n-s}, \quad s = \sum_{i=1}^n y_i$$

$y = (y_1, \dots, y_n)$

$$p(y \mid \theta) = \theta^s (1-\theta)^{n-s}$$

joint sampling dist.

$$\ell(\theta \mid y) = c \, p(y \mid s)$$

likelihood function

(>0)

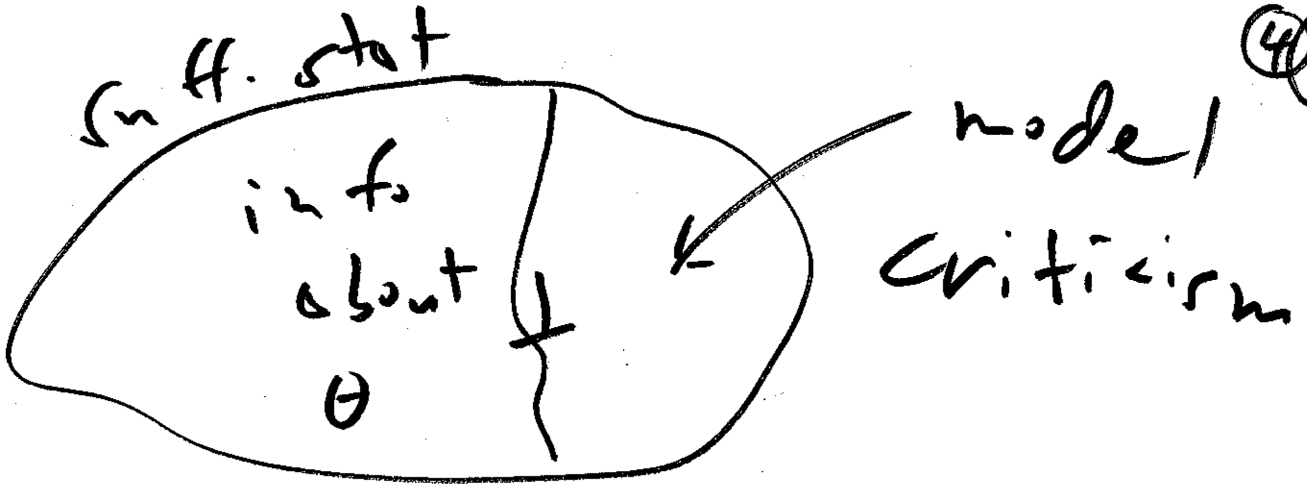
$$\ell(\theta \mid y) = \theta^s (1-\theta)^{n-s}, \quad s = \sum_{i=1}^n y_i$$

a sufficient statistic

$n = 400$

$s = 72$

$\hat{\theta} = 0.18$



total info in $(y_1, \dots, y_n)$

$y_1, \dots, y_n \overset{iid}{\sim}$ dist mean μ, SD σ

$$E(\bar{y}) = \mu$$

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

$$SE(\bar{y}) = \frac{\sigma}{\sqrt{n}} = SE(\hat{\theta})$$

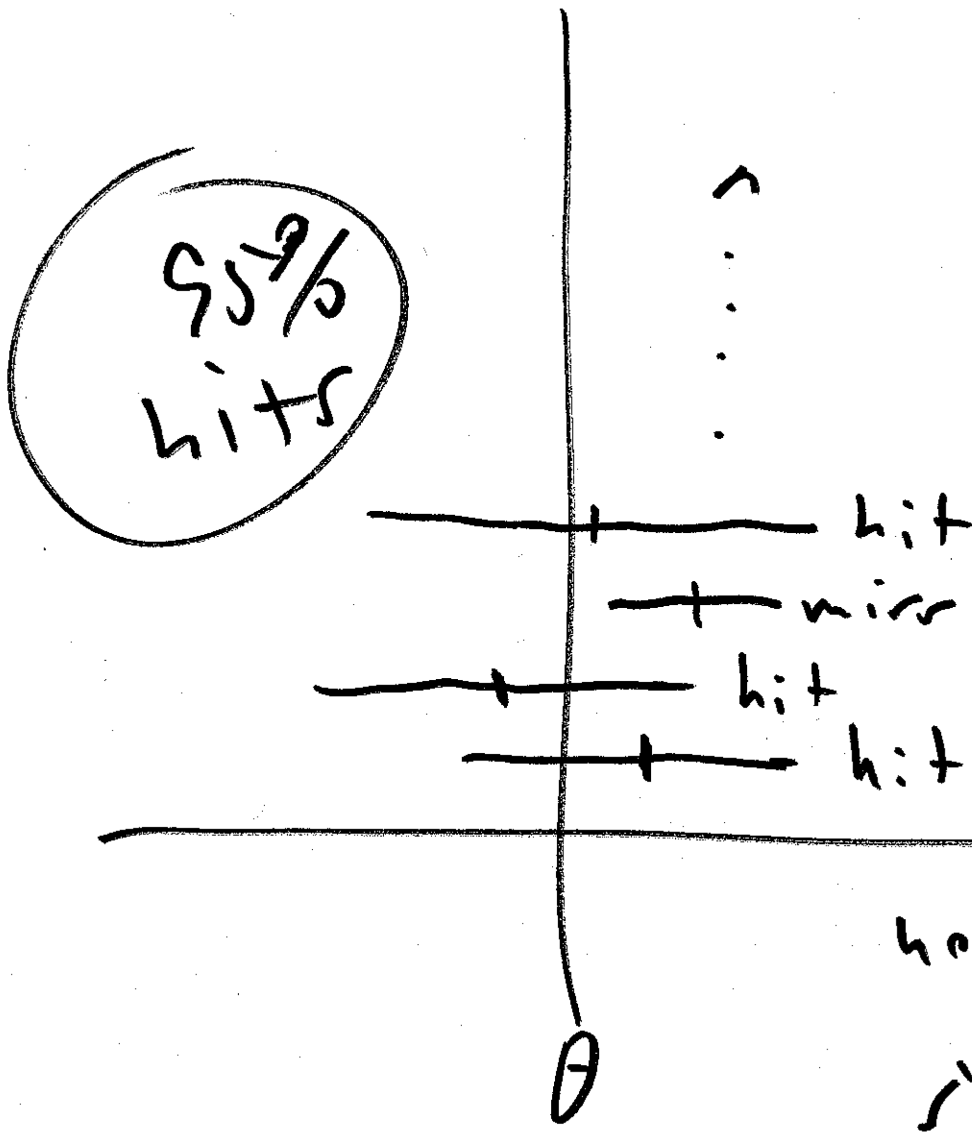
$= 1.9\%$
$= 0.019$

$$\hat{SE}(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = \sqrt{\frac{(.18)(.82)}{400}}$$

$$.18 \pm .04 = (.14, .22)$$

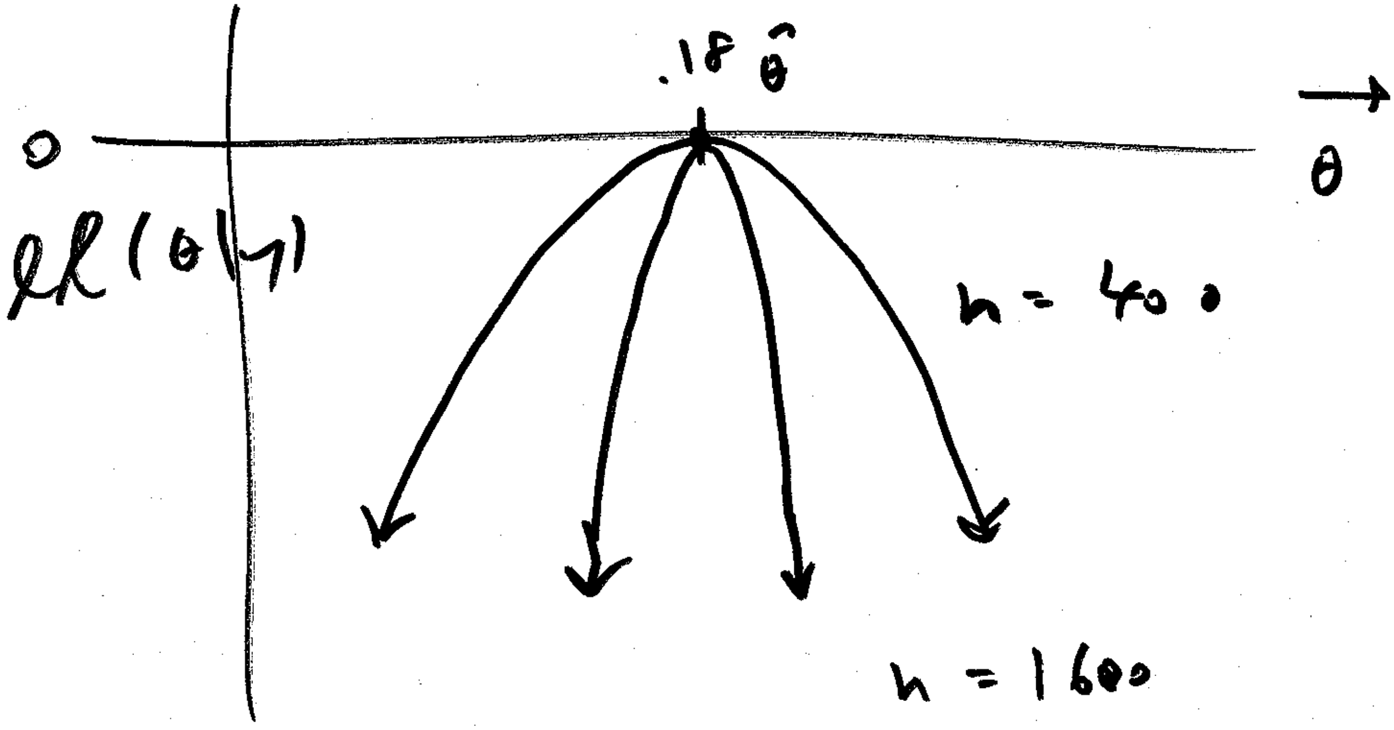
$$P_F(.14 < \theta < .22) = \cancel{\ }$$

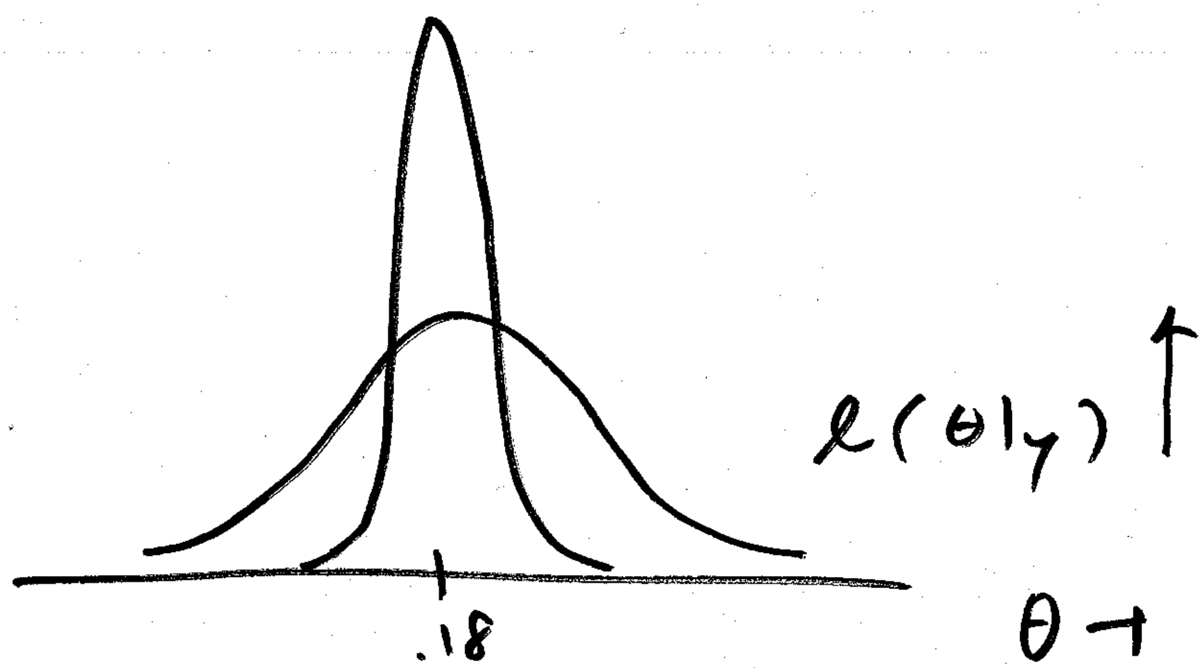
undefined



our confidence is in the process not in any single interval

6


.18 $\hat{\theta}$
0
θ
$\ell\ell(\theta|y)$
$n = 400$
$n = 1600$


$\ell(\theta|y)$
.18
θ

population (Neyman) sample repeated sampling 7

?

the observed patients

all possible $\hat{\theta}$

(1 = died, 0 = not)

died?

$N = ?$ (big)

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

mean $\theta = p = ?$

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

(actual) like IID

died?

$\begin{array}{c} Y_1 \\ \vdots \\ Y_n \end{array} \begin{pmatrix} 1s \\ \& \\ 0s \end{pmatrix}$ $n = 400$

mean $\hat{p} = \hat{\theta} = .18$

hyp. IID

$\begin{pmatrix} 1s \\ \& \\ 0s \end{pmatrix}$ $n = 400$

mean $\hat{\theta} = ?$ (ex. .20)

hyp IID

$\begin{pmatrix} \quad \end{pmatrix}$ $n = 400$

$\vdots$

mean $\hat{\theta} = ?$ (ex. .17)

$\begin{pmatrix} .18 \\ .20 \\ .17 \\ \vdots \\ \vdots \end{pmatrix}$

$M \to \infty$

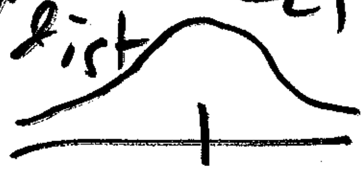
long run mean $E(\hat{\theta}) = \theta$

long run SD $SE(\hat{\theta}) = \sqrt{V(\hat{\theta})} = 0.019$

$\hat{SE}$ 0.019

long run dist

CLT



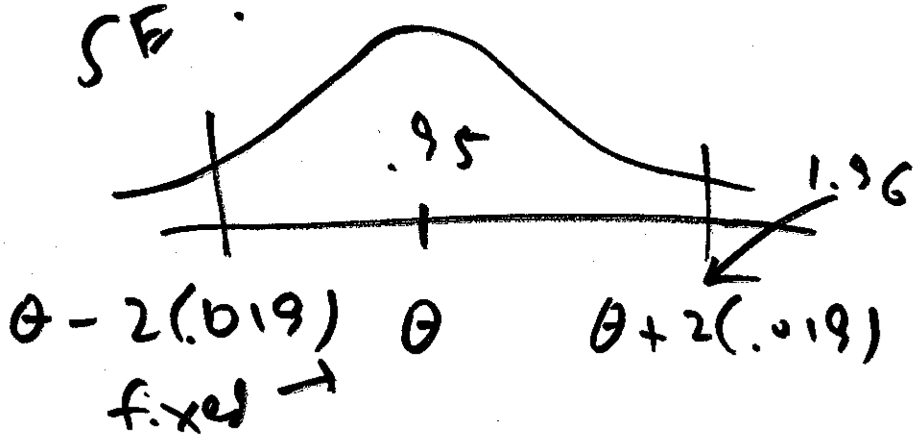
θ

sampling dist of $\hat{\theta}$

$\widehat{SE}$ .019

density of $\hat{\theta}$ in repeated sampling * RV

(8)



fixed ← rv

$$P_F(\theta - .038 < \hat{\theta} < \theta + .038) \doteq .95$$

$$\downarrow$$

$$P_F(\hat{\theta} - .038 < \theta < \hat{\theta} + .038) \doteq .95$$

$\hat{\theta} \pm 1.96\,\widehat{SE}(\hat{\theta})$ is

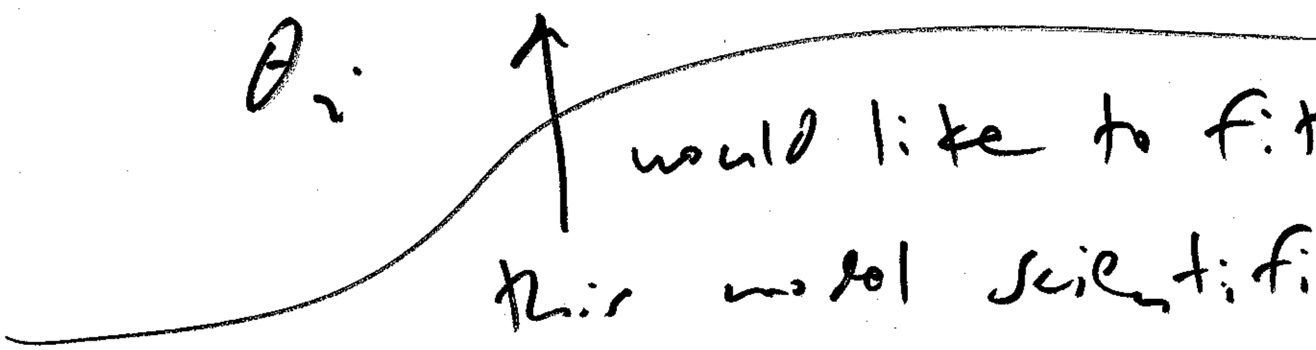
a 95% confidence interval for θ

$$\{\underbrace{Y_1, \ldots, Y_n}_{\text{sample}}, \ldots\} = P$$

$$(Y_i \mid \theta_i) \overset{\text{indep}}{\sim} \text{Bernoulli}(\theta_i)$$

$$i = 1, \ldots, n$$

θ_i



would like to fit this model scientifically

(because all patients are different with respect to covariates (predictors) (features) like age & other illnesses)

but can't: n observations $Y_1, \ldots, Y_n$

too many parameters $\rightarrow$ n parameters $\theta_1, \ldots, \theta_n$