Math 241 Lec 3 2/5/18

MLE's seem to solve all our problems!

We can do all goals of inference:

(1) Pt estimation: use $\hat{\theta}_{MLE}$

(2) Confidence set / interval,

use $CI_{\theta,1-\alpha} := \left[\hat{\theta}_{MLE} \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\right]$

we can do this because $\hat{\theta}_{MLE} \approx$ Normal due to thm in Math 633

(3) Hypothesis ~~tests~~: same thing!

What are the downsides??

① $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta)$ $X = \langle 0, 0, 0 \rangle$

$\Rightarrow \hat{\theta}_{MLE} = \bar{X}$ $\Rightarrow \hat{\theta}_{MLE} = 0$ no conf int, no hyp. testing!

② $\hat{SE} = 0$ $\Rightarrow$

③ What if you just know $\Theta \neq (0,1)$ but $\Theta = \Theta_0 < (0,1)$

prior knowledge... shouldn't that count for something?

How old am I? 5? No 80? No...

④ Frequentist interp. of conf int.

(A) over many repeats of experiment, 95% will cover θ

(B) before you begin... 95% chance cover θ

For any given interval $[0.27, 0.81]$ no interpretation applies!!

$\Rightarrow$ Frequentist interpretations are garbage.. you want $P(\theta \in CI) = 1 - \alpha$!

⑤ Hyp. Testing..

~~$H_0: \theta = \theta_0 = 0.5$ (fair)~~

~~$H_a: \theta \neq \theta_0 = 0.5$ (unfair)~~

Fail to reject Ho, Reject Ho

pval := $P\left(\text{seeing this evidence or more extreme} \;\middle|\; H_0 \text{ true}\right)$

You never get $P(H_a | X)$ or $P(H_0 | X)$

what you really want to know!

So.. very unsatisfying!!

Even in our case...

$\langle 0,0,1,0,1,0 \rangle \Rightarrow \bar{x} = \frac{2}{3}$

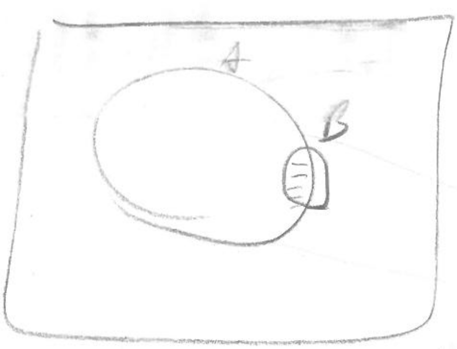
$$CI_{\theta,95\%} = \left[\hat{\theta} \pm z_{2.5\%}\sqrt{\hat{\theta}(1-\hat{\theta})}\right) = \left[0.33 \pm 2 \cdot \sqrt{0.33 \cdot 0.67}\right]$$
$$= \left[-0.60,\ 1.26\right] \text{ Why??}$$

asymptotic normality needed... doesn't kick in at $n = 6$.

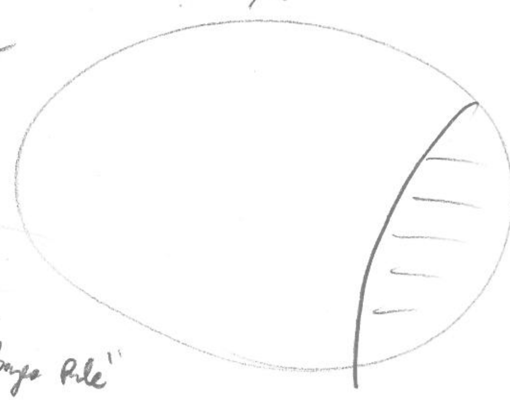
We will attempt to solve these problems... using Bayesian setup!

Recall

Ω

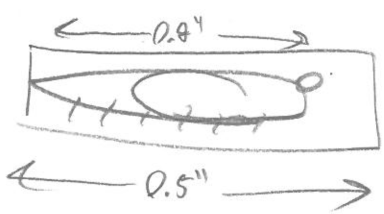


$P(A) = 0.2$ smoke
$P(B) = 0.06$ lung cancer
$P(A,B) = 0.036$



$$P(B|A) = c\,P(A,B) = P(A,B)\,\frac{P(\Omega)^{=1}}{P(A)} = \frac{P(A,B)}{P(A)}$$ "Bayes Rule"

is new universe



$$2 = \text{zoom} = \frac{1''}{0.5''} = \frac{\text{previous size}}{\text{new size}}$$

$\Rightarrow P(B,A) = P(B|A)P(A) = P(A|B)P(B)$ Bayes Rule as well

Also $P(A) = P(A,B) + P(A,B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$ law of total prob for 2 events

$\Rightarrow P(B|A) = \frac{P(A|B)P(B)}{}$ Bayes Thm.

Here's a way to look at this...

What's the target of estimation here? Prob of l.c.

What's the data? Smoky ← Prob of data

$$P(B \mid A) = \frac{P(A|B)\, P(B)}{P(A)}$$

$P(B)$ ← prior prob

$P(A)$ ← prior predictive data.

param of interest   data   posterior prob.

$P(B)$ prior... why? Best guess at the outset!

$$P(B) \xrightarrow{A \text{ (data)}} P(B|A) \text{ (better estimate)}$$

Bayesian Conditionalism

~~$P(B|A) = P(B)\left(\frac{P(A|B)}{P(A)}\right)$~~

~~$P(A|B) < P(A) \Rightarrow$ posterior prob. lower than prior prob.~~

~~$P(A|B) > P(A) \Rightarrow$ posterior prob. higher than prior prob.~~

~~Is the data more likely under param than in general?~~

Recall $\text{Odds}(A) := \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}$ range? if $P(A) \in (0,1)$ non zero, not 1

$(0, \infty)$

Odds=5 "5:1" i.e. "5 times it will happen out of 6" (on avg).

$\text{Odds Against}(A) = \frac{1}{\text{Odds}(A)}$

$P(B|A) = P(A|B)\ P(B)$

$P(B^c|A) = P(A|B^c)\ P(B^c)$

$$\Rightarrow \underbrace{\frac{P(B|A)}{P(B^c|A)}}_{\text{posterior odds}} = \underbrace{\frac{P(A|B)}{P(A|B^c)}}_{\text{likelihood ratio}} \overbrace{\frac{P(B)}{P(B^c)}}^{\text{prior odds}}$$

$= 0.22 \qquad = 3.44 \qquad = 0.064$

16:1 odds against

data

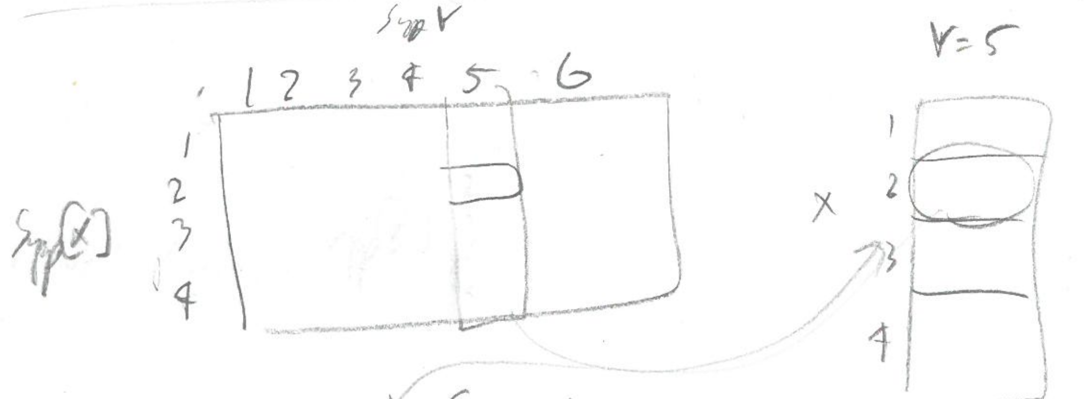
5:1 odds against

$P(A|B) = \frac{.036}{.06} = 0.6$

$P(A|B^c) = \frac{.164}{.94} = .174$

$P(A, B^c) = P(A) - P(A,B) = .2 - .036 = .164$

What is Marginalization?



$$P(X=2|Y=5) = \frac{P(X=2, Y=5)}{\underbrace{P(Y=5)}}$$

$P(X=1, Y=5) + P(X=2, Y=5) + P(X=3, Y=5) + P(X=4, Y=5)$

Bayes Rule for r.v. usually denoted

in shorthand.. $P(X|Y) = \frac{P(X,Y)}{P(Y)}$

or in longhand...

$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$

Marginalization Law of Total Prob

$\Rightarrow P(y) = \sum_{x \in Supp(X)} P(x,y)$ for PMFs or $p(y) = \sum_{x \in Supp(X)} p(y|x)\, p(x)$

$\Rightarrow f(y) = \int_{x \in Supp(X)} f(x,y)\, dx$ for PDFs or $f(y) = \int_{x \in Supp(X)} f(y|x)\, f(x)\, dx$

Back to the story:

$P(\theta|x) = \frac{P(x|\theta)\, P(\theta)}{P(x)}$ problems??