Math 341 Lec 2 2/2/17 $X_1, \dots X_6 \stackrel{iid}{\sim} \text{Bern}(\theta)$ saw $\langle 0,0,1,0,1,0 \rangle$ (S1)

Let's get $\hat{\theta}_{MLE}$ for this sample ... what do you think it is?

$$\ell(\theta; x) = \ln(\mathcal{L}(\theta; x)) = \ln\left(\prod_{i=1}^{6} \theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$f(x_1 \dots x_6; \theta) = \mathcal{L}(\theta; x_1 \dots x_6)$

$$= \sum_{i=1}^{6} \ln \theta^{x_i}(1-\theta)^{1-x_i}$$

$\prod \Rightarrow \sum$

prod $\Rightarrow$ sum

usually (HARD) usually (EASY)

$$= \sum_{i=1}^{6} x_i \ln\theta + (1-x_i)\ln(1-\theta)$$

$$= \ln\theta \sum x_i + \left(6 - \sum x_i\right)\ln(1-\theta)$$

Recall $\bar{x} = \frac{1}{n}\sum x_i \Rightarrow \sum x_i = n\bar{x}$

$$= \ln\theta\,(6\bar{x}) + (6 - 6\bar{x})\ln(1-\theta)$$

$$= 6\left(\bar{x}\ln(\theta) + (1-\bar{x})\ln(1-\theta)\right)$$

Now ... we maximize ...

use calculus!

$$0 \stackrel{set}{=} \frac{d}{d\theta}\left[\quad\right] = 6\left(\bar{x}\left(\frac{1}{\theta}\right) - (1-\bar{x})\frac{1}{1-\theta}\right)$$

$$0 = \bar{x}\frac{1}{\theta} - (1-\bar{x})\frac{1}{1-\theta}$$

$$0 = \bar{x}\left(\frac{1-\theta}{\theta}\right) - (1-\bar{x})$$

$$0 = \bar{x}(1-\theta) - (1-\bar{x})\theta = \bar{x} - \cancel{\bar{x}\theta} - \theta + \cancel{\bar{x}\theta} \Rightarrow \boxed{\theta = \bar{x} = \hat{\theta}_{MLE}} = \hat{p}$$

from Ps2#1

In our case $\bar{x} = \frac{2}{6} = \frac{1}{3}$

Our calc. above was regardless of $n=6$. It works for all n.

$\hat{\theta}_{MLE} = \bar{x}$, $\hat{\theta}_{MLE} = \bar{X}$ "estimator" ambiguous notation!

estimate

MLE's are not the only estimators but they have really nice properties

"Consistency"

1. $\hat{\theta}_{MLE} \xrightarrow{p} \theta$

By def: $\forall \varepsilon > 0 \quad \lim_{n \to \infty} P\left(|\hat{\theta}_{MLE} - \theta| \geq \varepsilon\right) = 0$

$\Rightarrow$ $\hat{\theta}_{MLE}$ becomes arbitrarily close to θ with high n

2. Asymptotic Normality

$$\hat{\theta}_{MLE} \xrightarrow{d} N\left(\theta, SE[\hat{\theta}_{MLE}]^2\right)$$

3. Efficiency $SE[\hat{\theta}_{MLE}]$ is the theoretically lowest std error for all consistent estimators (Cramer-Rao bound)

In the Bernoulli case $\hat{\theta}_{MLE} = \bar{X}$

$$\hat{\theta}_{MLE} = \bar{X} \approx N\left(\theta, \left(\sqrt{\frac{\theta(1-\theta)}{n}}\right)^2\right) \approx N\left(\bar{x}, \left(\underbrace{\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}}_{\hat{SE}[\hat{\theta}_{MLE}]}\right)^2\right)$$

$X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Bern}(\theta) \Rightarrow \hat{\theta}_{MLE} = \bar{X}$

$X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Geom}(\theta) := \underbrace{(1-\cancel{p})^x}_{\text{different parameterization than in M2\#1}} \theta$

Here, x is the # of failures before the first success.

$\text{Supp}[X] = \{0, 1, \ldots\} = \mathbb{N}_0$

$\Theta = (0, 1)$ param space

$$\mathcal{L}(\theta; x) = P(x; \theta) = \prod_{i=1}^{n} (1-\theta)^{x} \theta = \theta^{n} (1-\theta)^{\Sigma x_i}$$

$$\ell(\theta; x) = \ln(\;\;) = n \ln\theta + \Sigma x_i \ln(1-\theta)$$

$$\ell'(\theta; x) = \frac{n}{\theta} - \frac{\Sigma x_i}{1-\theta} \overset{set}{=} 0 \Rightarrow n(1-\theta) - \theta \Sigma x_i = 0$$

$$\Rightarrow n - n\theta = \theta \Sigma x_i$$

$$\Rightarrow n = \theta(\Sigma x_i + n)$$

$$\Rightarrow \hat{\theta}_{MLE} = \frac{n}{\Sigma x_i + n} = \frac{n}{\bar{x} n + n} = \frac{1}{\bar{x} + 1}$$

What would it be for old parameterization? $\frac{1}{\bar{X}}$ since $\bar{X}$ always gets one more trial

$SE[\hat{\theta}_{MLE}] = SE\left[\frac{1}{\bar{X}+1}\right]$ <... HARDER but totally possible

Inference Based on MLE's (frequentist method)

(1) Pt Estimate $\hat{\theta}_{MLE}$

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

(3) Hyp testing: $H_0 : \theta = \theta_0$

$H_a : \theta \neq \theta_0$

your theory

Retainment Region $:= \left[\theta_0 \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\big|_{\theta_0}\right]$

if $\hat{\theta}_{MLE} \in$ Retainment Region $\Rightarrow$ Fail to reject

Observe Data $\rightarrow$ Pick $\mathcal{F} \rightarrow$ Do Inference via MLE

What's wrong with this????

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

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

$\hat{SE} = 0$

(2) What if you just know $\Theta \neq (0,1)$ but $\Theta : \Theta_0 \subset (0,1)$

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

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

(3) 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 $P(\theta \in CI) = 1 - \alpha$!

(4) Hyp. Testing..

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

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

Fail to reject H_0,

Reject H_0

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...

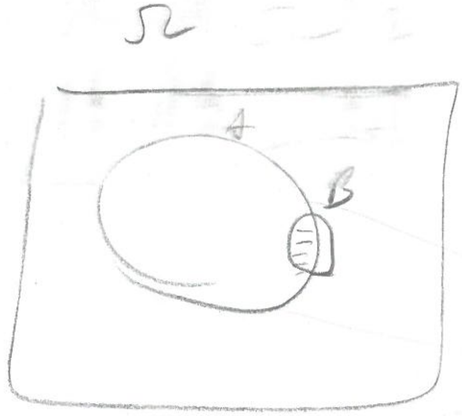
$\langle 0,0,1,0,1,0 \rangle \Rightarrow \bar{x} = \frac{2}{3}$

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

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

We will attempt to solve these problems...

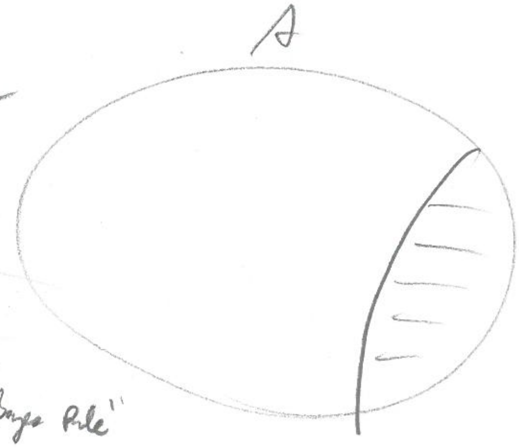
Recall



$P(A) = 0.2$ stroke

$P(B) = 0.06$ lung cancer

$P(A,B) = 0.036$



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

$$2 = \text{Zoom} = \frac{1''}{0.5''} = \frac{\text{previous size}}{\text{current size}}$$

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

Also $P(A) = P(A,B) + P(A,B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$

$\Rightarrow P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$ 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? Smoking

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

$P(A|B)$ ← prob of data

$P(B)$ ← prior prob

$P(A)$ ← prior prob of data

B ← param of interest

A ← data

$P(B|A)$ ← 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 $\text{Odds}(A) := \dfrac{P(A)}{P(A^c)} = \dfrac{P(A)}{1-P(A)}$ range? if $P(A) \in (0,1)$ non zero, non 1

$(0, \infty)$

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

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