Lecture 2 370.03-02

$\hat{\theta} = \bar{X}$ is max. likelihood estimate.

V.V.

$\hat{\theta} = \bar{X}$ is known as the MAX likelihood estimator for $\mathcal{F}$ being Bernoulli. MLE's aren't the only estimators but they are the "best" in many situations for the following reasons:

1. $\hat{\theta} \xrightarrow{P} \theta$ meaning $E[\hat{\theta}] \to \theta$ as n gets large "Consistent" and $Var[\hat{\theta}] \to 0$ "
2. Asymptotically Normal $\dfrac{\hat{\theta} - \theta}{SE[\hat{\theta}]} \to N(0,1)$ as n gets large
3. $SE[\hat{\theta}]$ is smaller than other estimators (optimal).

$$\Rightarrow \bar{X} \approx N(\theta, \underbrace{SE[\theta]}_{\sqrt{Var[\bar{X}]}}{}^2) = N\left(\theta, \left(\sqrt{\theta(1-\theta)}\right)^2\right)$$

$X_1, \ldots, X_n \overset{iid}{\sim} Geom(\theta) := (1-\theta)^{x-1}\theta$

$$\mathcal{L}(\theta; x) = \prod_{i=1}^{n}(1-\theta)^{x_i - 1}\theta = \theta^n (1-\theta)^{\sum x_i - n} = \left(\frac{\theta}{1-\theta}\right)^n (1-\theta)^{n\bar{X}}$$

$$\ln \mathcal{L} = n\ln(\theta) - n\ln(1-\theta) + n\bar{x}\ln(1-\theta)$$

$$\frac{d}{d\theta}\ell = \frac{n}{\theta} + \frac{n}{1-\theta} - \frac{n\bar{x}}{1-\theta} = 0 \Rightarrow \frac{1-\theta}{\theta}n + n - n\bar{x} = 0$$

$$\Rightarrow \left(\frac{1}{\theta} - 1\right)n + n - n\bar{x}$$

$$\Rightarrow \frac{n}{\theta} \cancel{- n + n} - n\bar{x} = 0$$

$$\Rightarrow \frac{1}{\theta} = \bar{x} \Rightarrow \hat{\theta} = \frac{1}{\bar{X}}$$

- Not always $\bar{X}$
- Not done in 241

Can be shown $SE(\hat{\theta}) = \sqrt{\frac{(1-\theta)\theta^2}{n}} \approx \sqrt{\frac{(1-\hat{\theta})\hat{\theta}^2}{n}} = \hat{SE}(\hat{\theta})$

Need a course in math stats for this...

Pt est: just $\hat{\theta}_{MLE}$

Conf. Interval:

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

Interpretation

(1)

(2)

(3)

e.g. 0,0,1,0,1,0

$0.33 \pm 2\sqrt{0.33 \cdot 0.66} = [-0.60, 1.26]$

???? Why ???

$H_0: \theta = 0.5$
$H_a: \theta \neq 0.5$

$0.5 \pm 2 \cdot .05 = [0,1]$ ✓

Hyp Test

$H_0: \theta = \theta_0$ fair ← my theory

$H_a: \theta \neq \theta_0$ unfair

Under H_0

$\hat{\theta}_{MLE} \sim N(\theta_0, SE[\hat{\theta}_{MLE}])$

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

Interpret: Rej or Ret. Pval? $P(H_0)$? $P(H_a)$? $P(H_0)/P(H_a)$

How much does $\hat{\theta}$ vary from sample-sample?

$\hat{\theta} \sim N(\theta, SE(\hat{\theta}))$

based on large sample asymptotic result...

Hidden Assumption: θ is one fixed value that we know nothing about

Observe data, pick parametric model $\mathcal{F}$

- If goal to estimate θ, use $\hat{\theta}_{MLE}$
- " " " provide confidence use $\hat{SE}[\hat{\theta}_{MLE}]$

(need asymptotic theory for this)

- If goal is to test, use asymptotic normality and the fact that it's asymptotically normal

---

Disadvantages to pt estimation

(1) What if "degenerate" case $\langle 0,0,0 \rangle$ $\hat{\theta}=0$?

(2) No way to factor in prior knowledge about $\Theta_0 \subset \Theta$

ex: How old am I? 5? 80?

You kind of know the same thing about a coin.

Most likely it's fair

Disadvantages in confidence set creation

(3) $\hat{SE}[\hat{\theta}_{MLE}]$ is needed. Without n large, it's wrong as we've just seen... Asymptotic normality needed

(4) Interpretation of a CI is RETARDED.

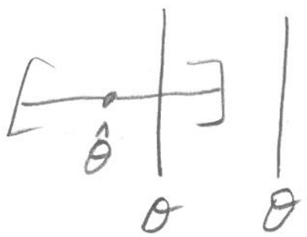
Disadvantages in hyp. testing

(3) Asymptotic normality needed... n low all bets off

(4) Can't ask what you really want... $P(H_0)$ e.g.

How did we get into this mess?

θ is considered one fixed #, unknowable.

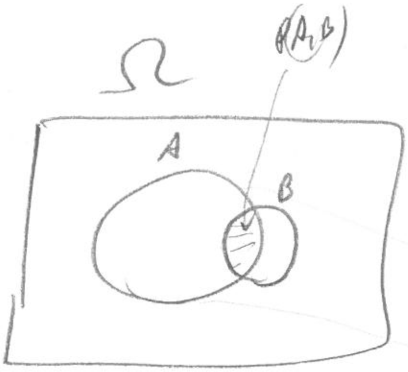


thus $P(\theta \in CI) = 0$ or 1

$H_0: \theta = 0.5$ $\quad P(H_0) = 0$ or 1 either $\theta = 0.5$ or it's not.

→ This makes lot of sense. But insistence upon it causes many problems...

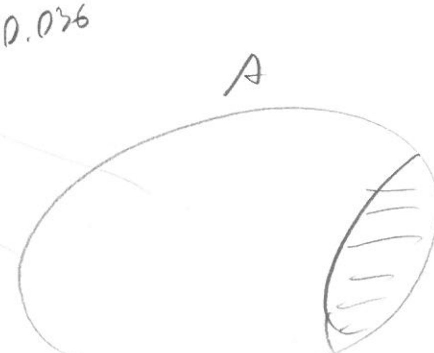
Recall...



$P(A) = 0.2$ smoking

$P(B) = 0.06$ l.c.

$P(A,B) = 0.036$



$P(\text{l.c.}|\text{smoking}) = P(B|A)$

$= P(A,B)$ in new universe

Which new universe?

Zoom $= 2 = \frac{1''}{0.5''} = 2$

Likewise here we begin with $P(\Omega)$ and zoom in to $P(A)$

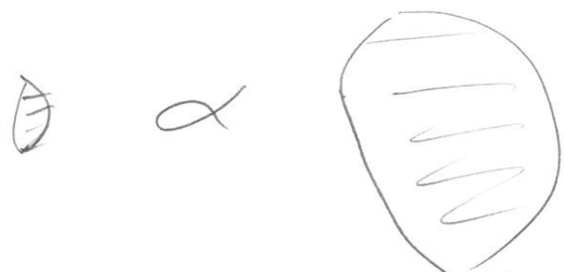
$\frac{P(\Omega)}{P(A)} = \frac{1}{P(A)}$ is the zoom factor $\Rightarrow$ $P(B|A) = P(B,A)\cdot\frac{1}{P(A)} = \frac{P(B,A)}{P(A)}$

Bayes Rule

proportional to:

same shape ... off by two diff. zoom factors.

$P(B,A) \quad \propto \quad P(B|A)$



$P(B,A) \qquad P(A|B)$

$P(B,A) = P(B|A)\,P(A) = P(A|B)\,P(B)$ also consistent Bayes Rule..

$\Rightarrow$ And $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$, $P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$ Bayes Rule again...

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

Law of Total Prob.

$\Rightarrow P(B|A) = \frac{P(A|B)\,P(B)}{\quad}$ Another Bayes Rule

What is our target of estimation here? prob l.c. in a certain insurance.
what's the data? smoking...

hissing in book...

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

P(B) ← prior on prob of l.c.

P(A) ← prior on the observed data

P(A|B) ← likelihood of the data

P(B|A) ← posterior prob of data

$$P(B) \longrightarrow P(B|A)$$ How? Multiply by $\frac{P(A|B)}{P(A)}$ ← dial it up or down

"Bayesian Conditionalism"

$P(A|B) < P(A) \Rightarrow$ posterior prob < prior prob

$P(A|B) > P(A) \Rightarrow$ " " > " "

More: ? .219 =

$$\underbrace{\frac{P(B|A)}{P(B^c|A)}}_{\text{posterior odds in favor of l.c.}} = \frac{P(A|B)}{P(A|B^c)} \underbrace{\frac{P(B)}{P(B^c)}}_{\text{prior odds in favor of l.c.}}$$

→ likelihood ratio
- called a Bayes Factor in favor of l.c. → Evaluate data and how to choose between two scenarios

3.640 positive / moderate evidence

.064

5:1 more likely not

16:1 more likely not

posterior odds / prior odds = Bayes Factor = $\frac{P(B|A)/P(B^c|A)}{P(B)/P(B^c)}$