Feb-03-16 / Lecture-2 / 1

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\arg\max} \left[\ln \mathcal{L}(\theta; X) \right]^{n}$$

real F

presupposed

$$\hat{\theta}_{MLE} \neq \bar{X}$$

1. $\hat{\theta}_{MLE} \xrightarrow{P} \theta$
2. $\hat{\theta}_{MLE} \longrightarrow N(\theta, SE(\hat{\theta}_{MLE}))$
3. $SE[\hat{\theta}_{MLE}]$ is optimal

$X_1, \ldots, X_n \overset{i.i.d}{\sim} geom(\theta) = (1-\theta)^{x-1}\theta$

$$\mathcal{L}(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} (1-\theta)^{x_i - 1}\theta$$

$$= (1-\theta)^{\sum(x_i - 1)} \theta^n$$

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

$$\frac{d}{d\theta}(\mathcal{L}()) = \frac{n(\bar{X}-1)}{1-\theta} + \frac{n}{\theta} = 0$$

$n(\bar{X}-1)$

$n\bar{X} - n$

$\sum X_i - n$

$$\frac{1}{\theta} = \frac{\bar{x}-1}{1-\theta}$$

$$1-\theta = \theta(\bar{X}-1)$$

$$\hat{\theta}_{MLE} = \frac{1}{\bar{X}}$$

2

$$SE[\hat{\theta}_{MLE}] = \sqrt{\frac{(1-\theta)\theta^2}{n}}$$

$$\approx \sqrt{\frac{(1-\hat{\theta}_{MLE})\hat{\theta}_{MLE}^2}{n}}$$

Inference:

1. Point Estimate $= \hat{SE}[\hat{\theta}_{MLE}]$
2. Confidence Sets $[\hat{\theta}_{MLE} \pm Z_{\frac{\alpha}{2}} \hat{SE}(\hat{\theta}_{MLE})]$
3. Hypothesis Testing

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

Retain Region

$$[\theta_0 \pm Z_{\alpha/2} SE(\hat{\theta}_{MLE})]$$

$$* \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$X_1, \ldots, X_6 \overset{i.i.d}{\sim} Ber(\theta)$

$$\hat{\theta}_{MLE} = \bar{X}$$

$\langle 0, 0, 1, 0, 1, 0 \rangle$ $\quad \bar{x} = \frac{0+0+1+0+1+0}{6} = \frac{2}{6} = \frac{1}{3}$

Best guess $\rightarrow 0.33$

3

95% CI BIG

$$\left[0.33 \pm 2\sqrt{\frac{0.33 \times 0.66}{6 = n}}\right]$$

$$= [-0.051, 0.711]$$

$H_0: \theta = 0.5$

$H_1: \theta \neq 0.5$

$$\left[0.5 \pm 2\sqrt{\frac{0.5(0.5)}{6}}\right] = [0.092, 0.908]$$

- (1) Disadvantages to Point Estimation:

(1) What if "divergence" case $\langle 0,0,0 \rangle$

(2) No way to factor in prior knowledge (we have a lot of prior idea, eg: age...)

Disadvantages for Confidence Interval

(3) $\hat{SE}[\hat{\theta}_{MLE}]$. Without n large, wrong Asymptotic Normality

(4) Interpretation bad.
$P(\theta \in C.I) = 0.95$

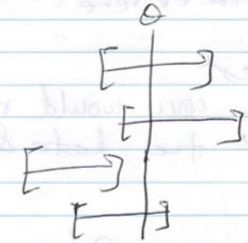
4

$\hat{\theta}_{MLE} \to$ R.V's

Fn

(1) Before experiment begin;
$P(\theta \in CI) = 0.95$

(2) $$\frac{\sum \mathbb{1}_{\theta \in CI_i}}{n} \longrightarrow 0.95$$



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

Disadvantages to testing

Asymptotic Normality?

$0.33 \in [0.092, 0.908] \Rightarrow$ Retain H_0

$H_0 \neq$ Accept H_0

~~0.99~~

$0.99 \notin [\quad] \Rightarrow$ Reject H_0 / Accept H_1

$P_{val} = P(\text{one data or more extreme} \mid H_0 \text{ true})$

smallest
$= \alpha$ you would reject at the data & H_0

Can't

$P(H_0; X)$ or $P(H_1; X)$

posterior prob of L.C

$$P(B \mid A) = \frac{\overset{f}{P(A \mid B)}\ \overset{\text{prior on prob of L.C}}{P(B)}}{\underset{\text{prior of your data}}{P(A)}} \quad \text{With Perspective}$$

B: target of estimation

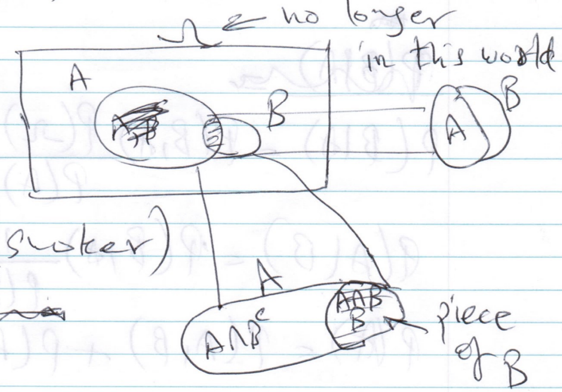
A = data

6% chance of getting lung cancer regardless of smoking

$P(A) = 0.2$ ← smoking

$P(B) = 0.06$ ← lung cancer

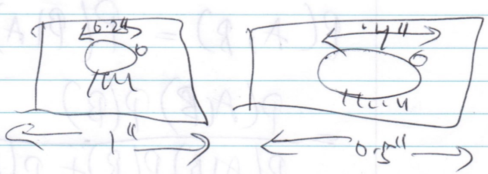
$P(A, B) = 0.036$



$P(\text{l.c} \mid \text{smoker})$

~~$P(B \mid A)$~~

$P(B \mid A) \propto P(B, A)$



$\text{Zoom} = \frac{.4''}{.2''} = 2 = \frac{1''}{0.5''}$

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

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

$\underline{P(B|A)}$

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

$$P(A|B) = P(B,A)\frac{1}{P(B)}$$

$$P(A) = P(A,B) + P(A,B^c) \leftarrow \text{law of Total Prob.}$$

$$\frac{P(A|B)P(B)}{P(A,B) + P(A,B^c)}$$

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

$$= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

$$\underset{\text{prior}}{P(B)} \xrightarrow{\text{data}} \underset{\text{posterior}}{P(B|A)}$$

know nothing of some data

after knowing data

Bayesian Conditionalism

Before that 6% chance, after knowing he is a smoker it goes up to 18%.

$$P(B|A) > P(B) \text{ if } \frac{P(A|B)}{P(A)} > 1$$

$$P(B|A) \leq P(B) \text{ if } \frac{P(A|B)}{P(A)} < 1$$

$$\text{Odds}(A) = \frac{P(A)}{P(A^c)}$$

$$\text{(Against) odds}(A) = \frac{P(A^c)}{P(A)} \quad \left\{\text{game odds}\right\}$$

likelihood ratio

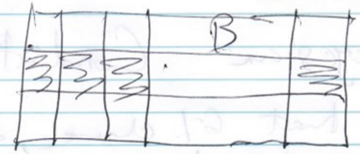
$$\text{Odds}(A) = \underbrace{\frac{P(B|A)}{P(B^c|A)}}_{\text{"posterior odds"}} = \underbrace{\frac{P(A|B)}{P(A|B^c)}}_{3.648} \underbrace{\frac{P(B)}{P(B^c)}}_{\text{"prior odds"}} \quad \left\{\text{Bayes Factor}\right\}$$

"posterior odds" 0.219 — 3.648 Greaters the no. — "prior odds" 0.067

5 : 1

$$= \frac{P(B|A) / P(B^c|A)}{P(B) / P(B^c)}$$

$A_1, \ldots, A_n$ mutually exclusive, collectively exhaustive



$$P(B) = P(B, A_1) + P(B, A_2) + \cdots + P(B, A_n)$$
$$= P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)$$

Bayes Theorem

$$P(A_\ell|B) = \frac{P(B|A_\ell)P(A_\ell)}{\underbrace{\sum_{i=1}^{n} P(B|A_i)P(A_i)}_{P(B)}}$$