Lec 3 Math 341 2/5/19

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\text{argmax}} \{ \ell(\theta; x) \}$$

MLE's allow for the 3 goals of inference

(I) Pt. Estimation. best guess is $\hat{\theta}_{MLE}$

(II) Confidence Set. $CI_{\theta, 1-\alpha} := \left[\hat{\theta}_{MLE} \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\big|_{\theta = \hat{\theta}_{MLE}} \right]$

(III) Hypothesis Test. $H_0: \theta = \theta_0$, Retention Region$_\alpha = \left[\theta_0 \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\big|_{\theta = \theta_0} \right]$

Trouble in Practice examples

(1) $\mathcal{F}$ = iid Bernoulli(θ) $\Rightarrow \hat{\theta}_{MLE} = \bar{x}$, $x = \langle 0, 0, 0 \rangle$

$\hat{\theta}_{MLE} = 0$. This is a bad idea. This means you're abs. sure the realization of the 1 is impossible

$CI_{\theta, 1-\alpha} = \{0\}$. You have no confidence in any other values!

$RR_\alpha = \{\theta_0\}$. Every null is rejected!

$\Rightarrow \hat{\theta}_{MLE}$ is bad for low n and/or very small / large true θ.

(2) What if you knew the Θ was restricted e.g. $\theta \notin [0.1, 0.2]$.. you wouldn't be able to use this information at all.

③ Consider the frequentist interpretation of a CI eg $CI_{\theta,95\%} = [0.37, 0.43]$

means ...

(a) If you repeat this experiment many times 95% of the CIs will capture the true θ.

(b) Before you begin, there is a 95% chance θ will be in the CI.

Very weird! $P(\theta \in [0.37, 0.43]) = 0$ or 1 since θ is fixed!

numbers!!

Any CI has no interpretation. What you want: $P(\theta \in [0.37, 0.43]) = 95\%$.

④ Frequentist Hypothesis Tests end in rejection of H_0 or retainment of H_0.

If you insist upon measuring a "strength" of rejection,

$$\text{pval} := P(\text{seeing } \hat{\theta}_{MLE} \text{ or more extreme} \mid H_0 \text{ true})$$

HOKEY!!

$$\neq P(H_0 \mid X)$$ which is really what you want

⑤ $\mathcal{F}$ = iid Bern.

$$X = \langle 0, 1, 0 \rangle \Rightarrow \hat{\theta}_{MLE} = 1/3$$

$$CI_{\theta,95\%} = \left[\frac{1}{3} \pm 1.96 \sqrt{\frac{\frac{1}{3} \cdot \frac{2}{3}}{3}} \right] = [-0.20, 0.87]$$

The width of the CI shouldn't bother you. Why? $n=3$... no precision

but CI $\not\subset$ Ⓗ! This is bad!

Why did this happen???

Asymptotic normality didn't kick in yet. When does it? In general, very difficult to know!

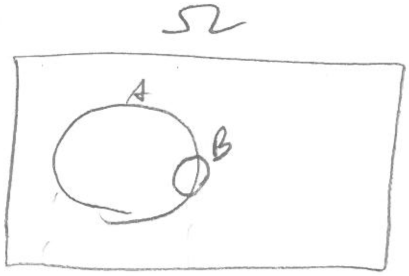
3

We will solve these problems using the Bayesian setup. Unfortunately, we will introduce new problems... but this's life...

Recall $P(A) = 0.2$ A: smoking

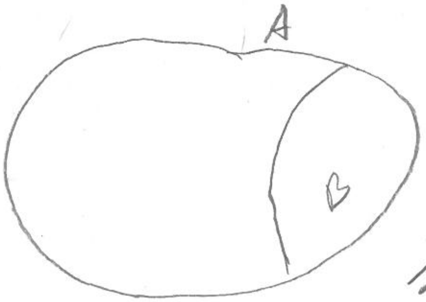
$P(B) = 0.06$ B: Lung Cancer

$P(A, B) = 0.036$



$P(\text{lung cancer} \mid \text{smoking})$

conditional universe... don't care about anything else!

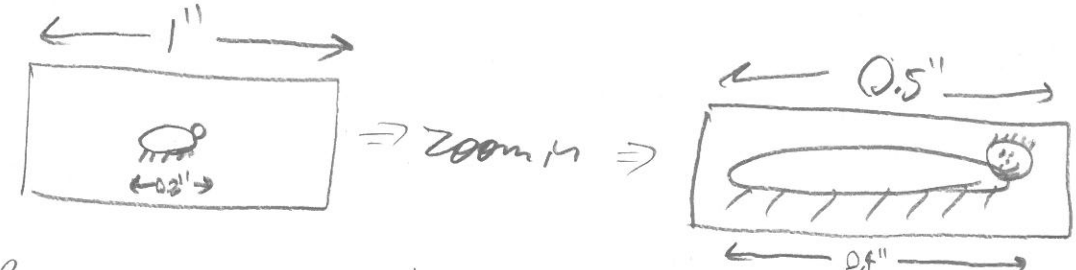


$$P(B|A) = c\, P(A,B) = \frac{1}{\underbrace{0.2}_{5}} P(A,B) = 0.18 = \frac{1}{P(A)} P(A,B) = \frac{P(A,B)}{P(A)}$$

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

proportional

same shape, but need to figure out magnification factor



Bug appears twice as large!

Def. of Cond. Prob.

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

$$\text{zoom} = 2 = \frac{1''}{0.5''} = \frac{\text{Original size of scope}}{\text{new size of scope}}$$

learning a clue over A

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

Bayesian Conditionalism

$\Rightarrow$ better prob.

A

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

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

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

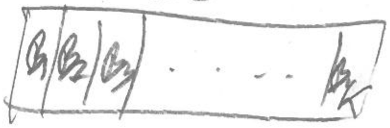
Since $P(A) = P(A,B) + P(A,B^c)$ addition rule of disjoint events

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

$$\Rightarrow P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \quad \text{Bayes Thm for two disjoint events}$$

If $B_1, B_2, \ldots, B_K$ are mutually exclusive and coll. exh.

Ω



i.e. $\bigcup_{i=1}^{K} B_i = \Omega \quad B_i \cap B_j = \phi \ \forall i \neq j$

$$\Rightarrow P(B_k|A) = \frac{P(A|B_k)}{\sum_{k=1}^{K} P(A|B_k)P(B_k)} \quad \text{Bayes Thm}$$

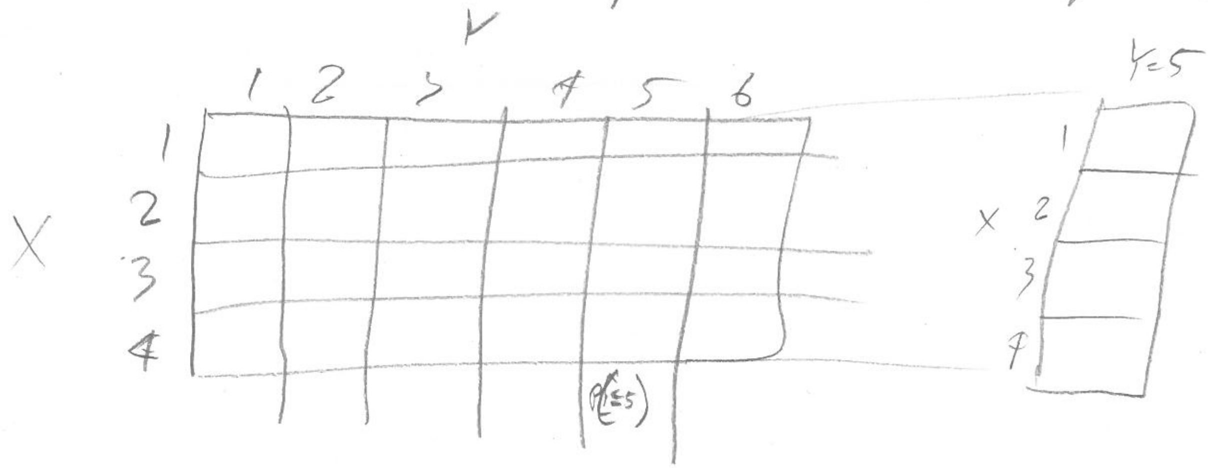
$$P(A) = \sum_{k=1}^{K} P(A|B_k)P(B_k) = \underbrace{\sum_{k=1}^{K} P(A, B_k)}$$

sum operation is called "marginalizing over B" or "marginalizing out B"

or "integrating over B" or "integrating out B"

Bayes Rule & Bayes Thm for r.v.'s

Imagine two r.v.'s X, Y, $Supp(X) = \{1,2,3,4\}$, $Supp(Y) = \{1,2,3,4,5,6\}$



$$P(Y=5) = P(Y=5, X=1) + P(Y=5, X=2) + P(Y=5, X=3) + P(Y=5, X=4) = \sum_{x \in Supp(X)} P(Y=5, X=x)$$

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

in general $P(X=x \mid Y=y) = \dfrac{P(X=x, Y=y)}{P(Y=y)}$

or $P(x \mid y) = \dfrac{P(x,y)}{P(y)}$ ← joint

and $P(y) = \sum_{x \in Supp(X)} p(x,y)$ if X discrete or

$P(y) = \int_{x \in Supp(X)} p(x,y)\, dx$ if X cont.

Back to the story... Let's use Bayes Rule to tell us something about the parameter θ?

$$P(\theta \mid X) = \frac{P(X \mid \theta)\, P(\theta)}{P(X)}$$

What's wrong with this?

(1) If θ is a fixed, parameter $\Rightarrow$ $\theta \sim \mathrm{Deg}(\theta)$

(2) $P(X)$ makes no sense. You cannot calc. prob of X unless know θ. Can marginalization help?

$$P(X) = \sum_{\theta_0 \in \Theta} P(X|\theta_0) P(\theta_0) \quad \text{if } \theta \text{ discrete}$$

$$= \int_{\theta_0 \in \Theta} P(X|\theta_0) P(\theta_0)\, d\theta \quad \text{if } \theta \text{ continuous}$$

If $\theta \sim \mathrm{Deg}(\theta) \Rightarrow \Theta = \{\theta_0\} \Rightarrow$ discrete

$\Rightarrow P(X) = P(X|\theta) \Rightarrow P(\theta|X) = \mathrm{Deg}(\theta)$

So this doesn't help at all!

What if we didn't insist that θ remain fixed?? Let θ itself be a r.v.

$\Longrightarrow$ The Big Leap

What is $P(\theta)$? This is called the "prior". Represents belief about the parameter before any data is seen. Is this real? Frequentists say no! This is the crux of the whole debate!