2/2/17

last time : $P(X_1, \dots, X_n ; \theta) \rightarrow$ fix parameter (θ) and calculate the probability of stuff happening

$\parallel$

$\mathcal{L}(\theta ; X_1, \dots, X_n) \rightarrow$ Try to figure out the model
Try to infer the parameter
Try all values and see which is best

- $\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta ; X_1, \dots, X_n)\}$

$= \underset{\theta \in \Theta}{\operatorname{argmax}} \{\ell(\theta ; X_1, \dots, X_n)\}$

Ex : $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$

data = 0,0,1,0,1,0 & we know $\theta = \frac{1}{3}$. Let's prove it.

- $\ell(\underset{X_1 \dots X_n}{\theta ; X}) = \ln\left(\prod_{i=1}^{6} \theta^{X_i}(1-\theta)^{1-X_i}\right) = \sum_{i=1}^{6} \ln\left(\theta^{X_i}(1-\theta)^{1-X_i}\right)$

$= \sum_{i=1}^{6}\left[X_i \ln(\theta) + (1-X_i)\ln(1-\theta)\right]$

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

Def: $\bar{X}_i = \frac{1}{n}\sum X_i \implies \sum X_i = n\bar{X}$

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

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

Trying to find Maximum so take derivative set $=0$

$\frac{d}{d\theta}(\ \ ) = 6\left(\frac{\bar{X}}{\theta} - \frac{1-\bar{X}}{1-\theta}\right) \overset{set}{=} 0$

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

$\bar{X} - \cancel{\theta\bar{X}} - \theta + \cancel{\theta\bar{X}} = 0$

$\bar{X} - \theta = 0$

$\bar{X} = \theta$

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

* To move from $\prod \to \sum$ makes taking derivative easier.

EX: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Geo(\theta) \overset{pmf}{:=} (1-\theta)^x \theta$ alternative parametrization of geometric

: x = # of failures before the stopping success

: $Supp(X) = \{0, 1, \ldots\} = \mathbb{N}_0$

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

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

$$\ell(\theta ; x) = \sum \ln((1-\theta)^{x_i} \theta)$$

$$= \ln(1-\theta) \sum x_i + n \ln(\theta)$$

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

$$\ell'(\theta ; x) = \frac{n}{\theta} - \frac{n \bar{x}}{1-\theta} = 0$$

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

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

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

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

$$\boxed{\hat{\theta}_{MLE} = \frac{1}{\bar{x}+1}} \quad \boxed{\checkmark\checkmark}$$

Properties of MLE

1. Consistency : $\hat{\theta}_{MLE} \xrightarrow{P} \theta$ (true θ)

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

: Whatever you demand of closeness from θ to $\hat{\theta}_{MLE}$, add more data and you'll get there.

2. Asymptotic Normality : $\hat{\theta}_{MLE} \xrightarrow{\text{distribution}} N(\theta, \underbrace{SE}_{\text{Standard Error}}(\hat{\theta}_{MLE})^2)$

: If you want a reasonable range (Confidence Interval)

: You're safe to perform testing hypothesis.

3. Efficiency : $\hat{\theta}_{MLE}$ has the lowest standard error theoretically possible.

EX : $\theta = 0.6 \pm 0.01$ * MLE

$\theta = 0.6 \pm 17$

Inference with MLEs

1. Point Estimate $\hat{\theta}_{MLE}$
2. Confidence Set $\left(I_{\theta, 1-\alpha} = [\hat{\theta}_{MLE} \pm Z_{\frac{\alpha}{2}} \, S.E. [\hat{\theta}_{MLE}]] \right)$
3. Hypothesis testing : level α $H_0 : \theta = \theta_0$ ($\neq$ reject)

$H_1 : \theta \neq \theta_0$ (= reject)

: retainment region $= [\theta_0 \pm Z_{\frac{\alpha}{2}} \, S.E(\hat{\theta}_{MLE})]$

So : Observe data $\rightarrow$ Pick $\underset{\text{model}}{\mathcal{F}}$ $\rightarrow$ Do inference with your MLE.

Holes / Problems in the System.

1. EX : $\mathcal{F}$ = Bernoulli , $X = 0, 0, 0$, $\hat{\theta}_{MLE} = \bar{X} = \frac{0+0+0}{3} = 0$

So the $S.E[\hat{\theta}_{MLE}] = \sqrt{\hat{\theta}_{MLE}(1 - \hat{\theta}_{MLE})} = 0$.

This tells us NOTHING. No inference can be made. $\boxed{XX}$

2. What if I have prior knowledge about Θ? I can't use it.

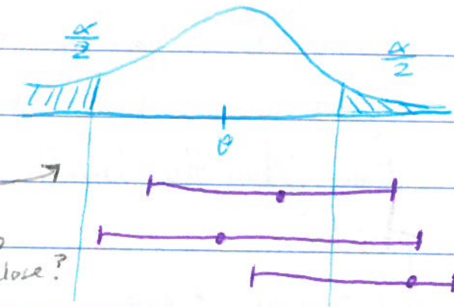
EX : Guessing a professor's age.

You know, they're not 5 or 100 years old, but not a part of the testing - Can't use it.

3. Frequentist Confidence Interval Interpretation

$C.I_{\theta, 1-\alpha}\,[0.42, 0.47]$ ← What does this likely mean?

confidence interval ↙   parameter ↓   ↳ coverage level



without knowing how to you know if your θ is close?

(a) If the experiment is repeated "many" times, then you cover ($\theta \in$ interval) $1-\alpha$ proportion of the time.

***Problem:** You only want to do it once!

(b) Before I do my experiment, $P(\theta \in C.I.) = 1-\alpha$

***Problem:** We don't care about before experiment

: Can't tell me anything about interval

∴ largely unsatisfactory.

4. Hypothesis Testing : level α

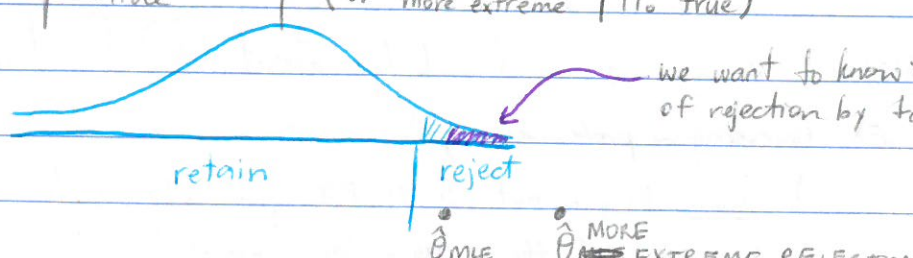
$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

There are 2 outcomes (a) H_0 : Retain

(b) H_0 : Reject

p-value $:= P\left(\begin{array}{l}\text{this data} \\ \text{or "more extreme"}\end{array} \middle| H_0 \text{ true}\right)$



we want to know the strength of rejection by taking integral (probability of null hyp given data.

- Want to know : $P(H_0 | X)$? (Want to know how good interval is)

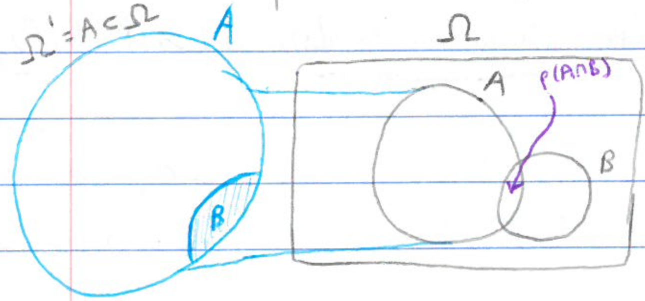
Ex : $X = 0,0,1,0,1,0$, $\hat{\theta} = \frac{1}{3}$

$C.I_{\theta, 95\%} = \left[\frac{1}{3}, \pm 2\sqrt{\frac{1}{3} \cdot \frac{2}{3}} \right] = [-0.60, 1.26]$ BUT, we can't get (−) #s or #s > 1!

used the 2 bc of 95% bc we assume the Normality distribution **BUT** Not converged yet.

* Normal Asymptotic didn't kick in yet bc $n = 6$. (not big enough)

- Motivation for Math 341 : To prove Bayesian > Frequentist approach.



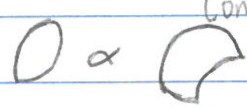
A = smoking $P(A) = 0.2$

B = lung cancer $P(B) = 0.06$

$P(A \cap B) = 0.036$

Conditional Prob. $P(B|A)$.

look at shapes :



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

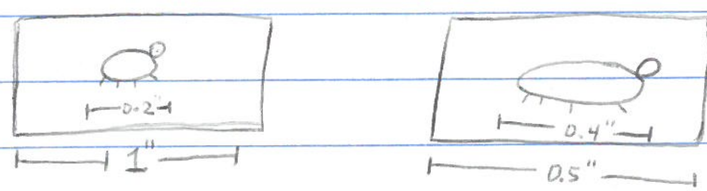
$\Downarrow$

- $P(A \cap B) = c\, P(B|A)$

* The shapes are the same but the scale of the siz are different.

- $\underset{\text{(New size)}}{P(B|A)} = \frac{P(\Omega)}{P(A)} \underset{\text{(Old size)}}{P(A \cap B)} = \frac{P(A \cap B)}{P(A)}$ } Bayes Rule

- Bug example :



$\text{zoom} = \frac{1''}{0.5''} = \frac{\text{Previous Zoom}}{\text{Current Zoom}}$

$\cap = ,$ comma

$$\underbrace{P(A, B)}_{P(A \cap B)} + P(A, \underbrace{\bar{B}}_{B^c})$$

$P(A) = P(A, B) + P(A, \bar{B})$ = Law of Total Probablity

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

posterior prob. — likelihood/prob. of data — Prior Prob.

$$\underbrace{P(B|A)}_{\text{posterior prob.}} = \frac{\underbrace{P(A|B)}_{\text{likelihood/prob. of data}}\ \underbrace{P(B)}_{\text{Prior Prob.}}}{P(A)}$$

B: lung cancer, target of Estimation

A: evidence/data

After ⟵ Before

So, Begin with $P(B)$ with introduction to data $(A) \rightarrow P(B|A)$ — Bayesian Conditionalism