- inference goals of MLE: (the frequentist model)

① Point estimation for θ : find $\hat{\theta}_{MLE}$ itself

② confidence sets for θ : confidence intervals CI
(region of values)

$$CI_{p,1-\alpha} = \hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

parameter, level of confidence, standard normal

Bernoulli Model inference via samples

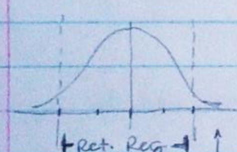
$$CI_{\theta,1-\alpha} := \left[\hat{\theta}_{MLE} \pm Z_{\frac{\alpha}{2}} \, SE\left[\hat{\theta}_{MLE}\right] \right]$$

③ Testing theories about θ : Hypothesis testing

$H_0 : \theta = \theta_0$
$H_a : \theta \neq \theta_0$

Retention Region $:= \left[\theta_0 \pm Z_{\frac{\alpha}{2}} \, SE(\hat{\theta}_{MLE}) \right]$

* Observe Data → Pick $\mathcal{I}$ → do inference via MLE



* Model can only tell us to reject θ_0 not how to adjust

Lecture 3

- issues of iid Beroulli model / frequentism

① Small data sets,
one outcome implies exclusion of other result which isnt true ex: $<0,0,0>$ but $\hat{\theta}_{MLE} \neq 0$

② if $\Theta \neq (0,1)$ ex. $\Theta = [02, 08]$; doesn't account for restricted parameter space

③ independent of confidence interval
- over repeated experiments $1-\alpha$ of CI will cover θ
- before you do experiments $p(\theta \in CI) = 1 - \alpha$

for given CI, frequentism doesn't work

④ testing ; $\hat{\theta}_{MLE} \in$ Retain. Reg. => dont Reject H_0

$\hat{\theta}_{MLE} \in$ Retain. Reg. => Reject H_0

$H_0 : \theta = \theta_0$ we want $P(H_0 \mid data)$ which isnt possible.

$H_\alpha : \theta \neq \theta_0$ $P(H_\alpha \mid data)$

ex: $x = \langle 0, 0, 1, 0, 1, 0 \rangle$

$\hat{\theta}_{MLE} = \bar{x} = 1/3$

$$CI_{\theta, 95\%} = \left[\underbrace{1/3}_{\hat{\theta}_{MLE}} \pm Z_{\frac{\alpha}{2}} SE(\hat{\theta}_{MLE}) \right] = \left[\frac{1}{3} \pm 2 \sqrt{\frac{1}{3}\left(1 - \frac{1}{3}\right)} \right] = [-0.60, 1.26]$$

cant happen

$\alpha = 5\%$

$\frac{\alpha}{2} = 2.5\%$

$Z_{2.5\%} = 2$ $SE(\hat{\theta}_{MLE}) = \sqrt{\theta(1-\theta)}$

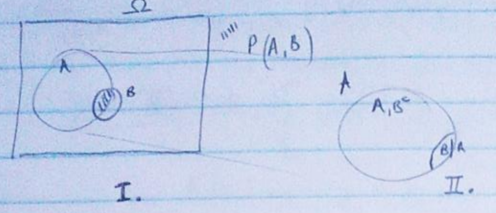
$X_1 \ldots X_n \overset{iid}{\sim} Bern(\theta)$ $\theta \in \Theta = (0, 1)$

* the sample is not large enough for $\hat{\theta}_{MLE} \xrightarrow{d} N$

Note: Zoom factor : $\frac{\text{original}}{\text{final size}}$

Lec 3 cont.

Bayes Rule / Theorem



	I	II (zoom)
	$P(\Omega) = 1$	
	$P(A) = 0.2$	$P(A) = 1$
	$P(B) = 0.06$	$P(B) = 0.06 \cdot$ Constant
	$P(A,B) = 0.036$	$P(A,B) = 0.036 \cdot$ Constant $= P(B\|A)$

so $P(B|A) = c\, P(A,B) = \left(\frac{P(\Omega)}{P(A)}\right) P(A,B) = \frac{1}{P(A)} P(A,B)$

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

Also $P(A)\,P(B|A) = P(A,B)$
$P(B)\,P(A|B) = P(B,A)$ so $P(A)\,P(B|A) = P(B)\,P(A|B)$

- hence $\left[P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} \right]$ Bayes rule

By law of total probability: $P(A) = P(A,B) + P(A,B^c)$

so

- $P(B_k|A) = \frac{P(A|B)\,P(B)}{P(A,B) + P(A,B^c)} = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \left[\frac{P(A|B_k)\;P(B_k)}{\sum_{i=1}^{n} P(A|B_i)\,P(B_i)} \right]$

* given that all B_i partition the space — Bayes theorem

$P(\underset{\text{parameter}}{B} \mid \underset{\text{data/evidence}}{A}) = \frac{P(A|B)\,P(B) \leftarrow \text{prior}}{P(A)}$

posterior — probability of seeing the data — prior distribution

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

bayesian conditionalism

• Odds

odds for A : $\text{Odds}(A) = \dfrac{P(A)}{1-P(A)}$

odds against A : $\text{OddsA}(A) = \dfrac{1}{\text{odds}(A)} = \dfrac{1-P(A)}{P(A)}$

Range of $\underset{\text{oddsA}(A)}{\text{odds}(A)}$: $(0, \infty)$ if prob $(A) \neq 0$

odds of posterior =

$$\underbrace{\frac{P(B|A)}{P(B^c|A)}}_{\text{posterior odds}} = \underbrace{\frac{P(A|B)}{P(A|B^c)}}_{\text{likelihood ration}} \underbrace{\frac{P(B)}{P(B^c)}}_{\text{prior odds}}$$

• ex: two r.v. X, Y

supp[X]	1	2	3	4	5	6
1						
2						
3						
4						

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

Generally $= \left[\dfrac{P(X=x, Y=y)}{P(Y=y)}\right]$

$$P_y(Y) = \sum_{X \in \text{Supp}(X)} P_{x,y}(X,Y)$$

$$\left[f_y(Y) = \int_{X \in \text{Supp}(X)} f_{x,y}(x,y)\, dx\right]$$

Marginalization
law of total prob

$$P(2|5) = \frac{P(X=2, Y=5)}{P(Y=5)} = \frac{\left(\frac{1}{6} \cdot \frac{1}{4}\right)}{\left(\frac{1}{6} \cdot \frac{1}{4}\right) + \left(\frac{1}{6} \cdot \frac{1}{4}\right) + \left(\frac{1}{6} \cdot \frac{1}{4}\right) + \left(\frac{1}{6} \cdot \frac{1}{4}\right)} = \frac{1}{4}$$

For 341

$$P(\theta | x) = \frac{P(x|\theta)\, P(\theta)}{P(x)}$$

Such that $P(\theta) \xrightarrow{\text{data}(x)} P(\theta|x)$

* What if θ is not fixed?