Lecture 2

01/30/20

We observe "data" $x = \langle 0,0,1,0,1,0 \rangle$
realization from a random process.
Assume a parametric model:

$$\tilde{F} = \text{iid Bernoulli} = \{ P(X;\theta) : \theta \in \Theta$$

PMF

$$= P(x) = P(X;\theta) \quad = \{ \theta^x (1-\theta)^{1-x} : \theta \in (0,1) \}$$

constant θ needed to compute the probability

$$f(x) = f(x;a) = \sin(ax)$$

we'd like to learn about θ (inference)

$$P(X;\theta) \overset{\tilde{F}}{=} P(\langle 0,0,1,0,1,0 \rangle;\theta) = (\theta^0(1-\theta)^{1-0}) \cdot (\theta^0(1-\theta)^1) \cdots (\theta^0(1-\theta)^1)$$

$$= \theta^2(1-\theta)^4$$

What if $\theta = .5 \rightarrow P(X;\theta) = .5^2(1-.5)^4 = .0156$

What if $\theta = .25 \rightarrow P(X;\theta) = .25^2(1-.25)^4 = .0198$

"Likelihood Function" $\qquad$ "Probability"

$$\mathcal{L}(\theta;X) = P(X;\theta)$$

$\mathcal{L}$: θ — Input, X — constant $\qquad$ P: X — input, θ — constant

(the likelihood of "seeing" the parameter at a certain value)

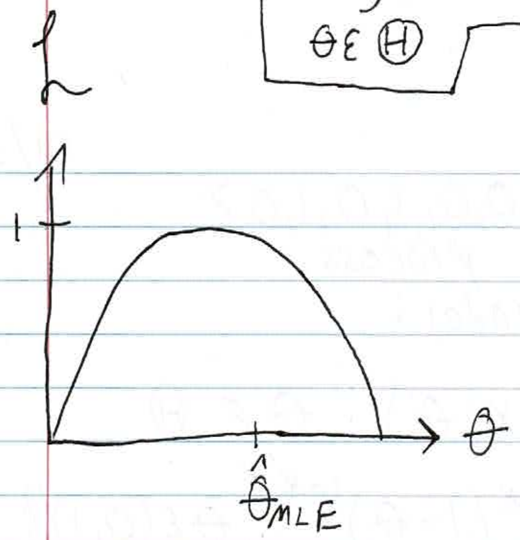
$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta;X)\} = \underset{\theta \in \Theta}{\operatorname{argmax}} \{ g(\mathcal{L}(\theta;X)) \}$$

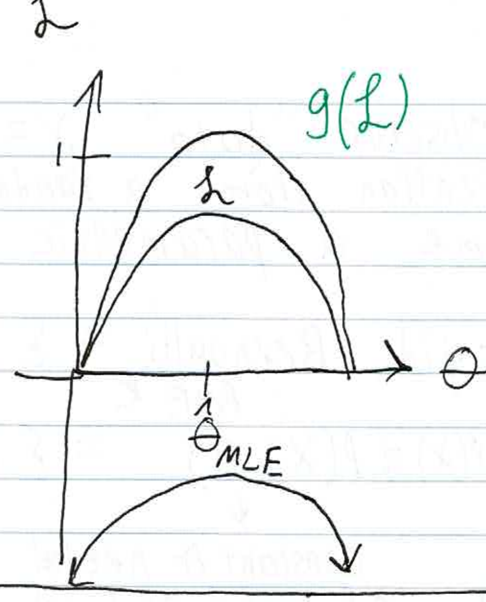
MLE — maximum likelihood estimate

argmax — argument that results in the maximum

where g is a strictly increasing function

$$\boxed{\int_{\theta \in \Theta} \mathcal{L}(\theta;x)\,d\theta \overset{?}{=} 1 \quad NO}$$





$$\ell(\theta;x) = \underbrace{\ln(\mathcal{L}(\theta;x))}_{\text{log-likelihood}} = \ln\left(\prod_{i=1}^{6} \theta^{x_i}(1-\theta)^{1-x_i}\right) =$$

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

Should be $(6-6\bar{x})$?

$$= \ln(\theta)\sum_{i=1}^{6} x_i + \ln(1-\theta)\sum_{i=1}^{6} 1-x_i = \ln(\theta)6\bar{x} + \ln(1-\theta)(6-\bar{x})$$

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

$$= \ln(\theta;x)$$

Sample Avg.

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\Rightarrow \sum x_i = n\bar{x} = 6\bar{x}$

$$\hat{\theta}_{MLE} = \operatorname{argmax}\{\mathcal{L}(\theta,x)\} = \operatorname{argmax}\{\ln(\mathcal{L}(\theta;x))\} = \operatorname{argmax}\{\ell(\theta;x)\}$$

1) Find $\ell(\theta;x)$
2) Find $\frac{d}{d\theta}[\ell(\theta;x)] = \ell'(\theta;x)$
3) $\ell'(\theta;x) \overset{set}{=} 0$
4) Solve for $\hat{\theta}_{MLE}$

$$l'(\theta;x) = 6\left(\frac{\bar{x}}{\theta} - \frac{1-\bar{x}}{1-\theta}\right) \overset{set}{=} 0$$

$$\Rightarrow \frac{\bar{x}}{\theta} = \frac{1-\bar{x}}{1-\theta} \Rightarrow \bar{x}(1-\theta) = (1-\bar{x})\theta$$

$$\Rightarrow \bar{x} - \bar{x}\theta = \theta - \bar{x}\theta$$

$$+\bar{x}\theta \qquad +\bar{x}\theta$$

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

$x = \langle 0,0,1,0,1,0 \rangle \qquad \bar{x} = \frac{1}{3} = \hat{\theta}_{MLE}$

$X_1, \ldots, X_n \overset{iid}{\sim} Bern(\theta)$

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

MLE is not the only point estimation Strategy but it is common, as it has nice properties

① $\hat{\theta}_{MLE} \xrightarrow{P} \theta$

"Consistency" i.e. it converges to the true value in n

② $\hat{\theta}_{MLE} \approx N(\theta, SE[\hat{\theta}_{MLE}]^2)$

"Asymptotic Normality"

③ Among all consistent estimators $\hat{\theta}_{MLE}$ has lowest variance

"Efficiency"

Consider $X \sim Geometric(\theta) = (1-\theta)^x \theta$

$\tilde{F}$ = iid Geometric

$$\mathcal{L}(\theta; x) = \prod_{i=1}^{n} (1-\theta)^{x_i} \theta = (1-\theta)^{\Sigma x_i} \theta^n$$

$$\ell(\theta; x) = \sum_{i=1}^{n} \ln((1-\theta)^{x_i}\theta) = \sum_{i=1}^{n} x_i \ln(1-\theta) + \ln(\theta)$$

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

$$\ell'(\theta; x) = n\left(-\frac{\bar{x}}{1-\theta} + \frac{1}{\theta}\right) \overset{set}{=} 0 \Rightarrow \frac{1}{\theta} = \frac{\bar{x}}{1-\theta}$$

$$\Rightarrow 1-\theta = \theta\bar{x} \Rightarrow 1 = \theta\bar{x} + \theta$$

$$1 = \theta(\bar{x}+1)$$

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

Let's say $\bar{x} = 99$ (99 failures happen before 1 success)

$$\hat{\theta}_{MLE} = \frac{1}{1+99} = 1\%$$

If $\bar{x} = 1$ (waiting for 1 failure)

$$\hat{\theta}_{MLE} = \frac{1}{1+1} = 50\%$$

Property ② from MLE properties

$$\hat{\theta}_{MLE} \overset{d}{\approx} N(\theta, SE[\hat{\theta}_{MLE}]^2) \approx N(\hat{\theta}_{MLE}, SE[\hat{\theta}_{MLE}] \Big|_{\theta = \hat{\theta}_{MLE}}$$

$\tilde{F}$ = iid Bernoulli :

$\hat{\theta}_{MLE} = \bar{X}$ $\quad\quad \hat{\theta}_{MLE} \approx N(\theta, SE[X]^2) = \cancel{N(\theta,}$

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

$$\hat{\theta}_{MLE} \approx N\left(\bar{X}, \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}^{\,2}\right)$$

For iid Geometric : $\hat{\theta}_{MLE} = \frac{1}{1+\bar{X}}$, $SE[\hat{\theta}_{MLE}] = SE\left[\frac{1}{1+\bar{X}}\right]$

$= ?$ Beyond this course

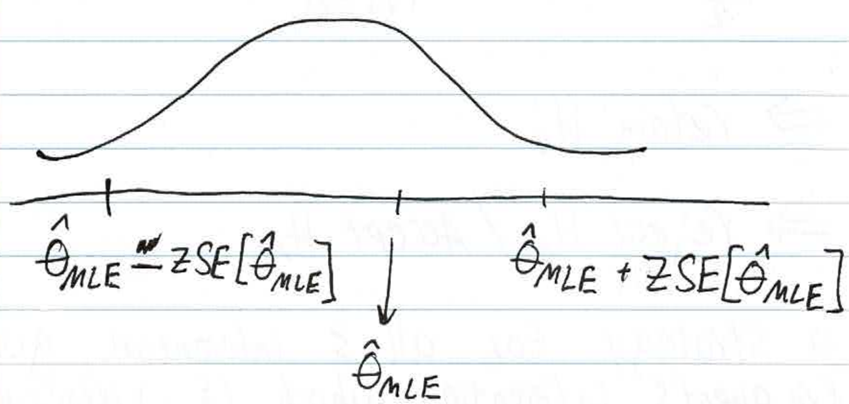
---

3 goals of inference :

1) Point est. $\theta \approx \hat{\theta}_{MLE}$

2) Confidence Interval for θ of size $1-\alpha$

$$CI_{\theta, 1-\alpha} = [\hat{\theta}_{MLE} \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}] \Big|_{\theta = \hat{\theta}_{MLE}}$$



3) Testing (also called Hypothesis Testing)

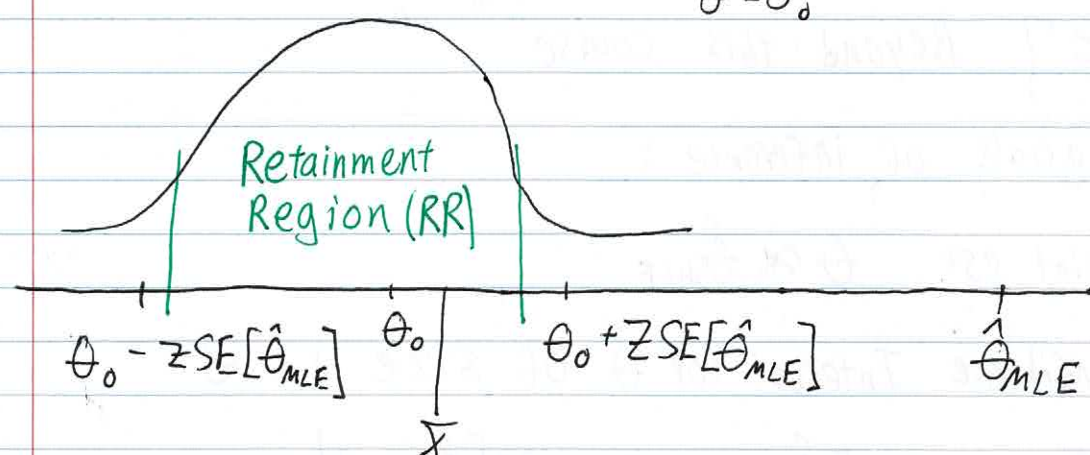
next page →

idea "Null Hypothesis" $\leftarrow H_0: \theta_0$ → some specific value

Alternative Hypothesis $\leftarrow H_A: \theta \neq \theta_0$

Assume my theory is true and let the data tell me if I'm right or wrong

$$\hat{\theta}_{MLE} \approx N\left(\theta_0, SE[\hat{\theta}_{MLE}]^2\Big|_{\theta=\theta_0}\right)$$



$$RR_{\theta_0, 1-\alpha} = \left[\theta_0 \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\right]\Big|_{\theta=\theta_0}$$

If $\hat{\theta}_{MLE} \in RR \Rightarrow$ retain H_0

If $\hat{\theta}_{MLE} \notin RR \Rightarrow$ reject H_0 / Accept H_A

We have a strategy for all 3 inferential goals. we've done frequent's inferential which is historically classic.