Math 341 Lec 2 1/31/18

L1

$$\mathcal{L}(\theta; x) = p(\underbrace{x}_{\text{data}}; \underbrace{\theta}_{\text{parameter}})$$

what is the "likelihood" of these parameters given this data — what is the prob. of the data given these values of the params

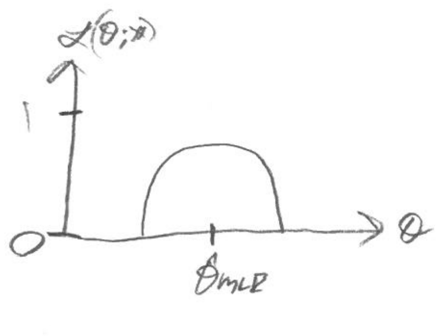
Numerically equivalent, conceptually two different universes

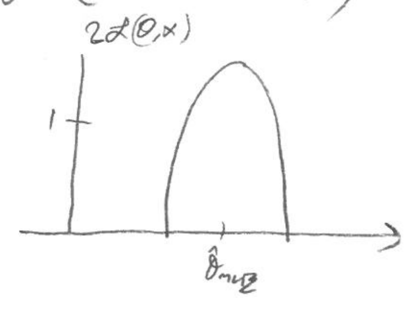
$$\hat{\theta}_{MLE} := \underset{\theta \in \Theta}{\text{argmax}} \{\mathcal{L}(\theta; x)\}$$ i.e. the most likely parameter value assuming this data

Note: a positive monotonic transformation g is one that preserves the following:

$$x > y \Rightarrow g(x) > g(y)$$ i.e. a strictly increasing function

If $\hat{\theta}_{MLE}$ is the best because it maximizes $\mathcal{L}(\theta; x)$ then it will also maximize $g(\mathcal{L}(\theta; x))$



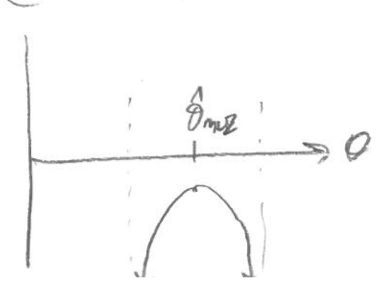


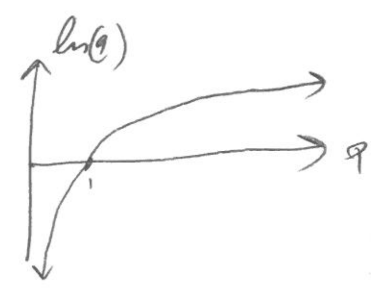
let

$\ell(\theta; x) :=$

$\ln(\mathcal{L}(\theta; x))$

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





$X_1, \ldots X_6 \overset{iid}{\sim} Bern(\theta)$ sample $\langle 0, 0, 1, 0, 1, 0 \rangle$

Let's get $\hat{\theta}_{MLE}$ for this sample... what do you think it is?

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

$f(x_1 \ldots x_6; \theta) = \mathcal{L}(\theta; x_1, \ldots, x_6)$

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

$\prod \Rightarrow \sum$

prod $\Rightarrow$ sum

usually (HARD) usually (EASY)

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

$$= \ln\theta \sum x_i + \left(6 - \sum x_i\right)\ln(1-\theta)$$

Recall $\bar{x} = \frac{1}{n}\sum x_i \Rightarrow \sum x_i = n\bar{x}$

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

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

Now... we maximize...

use calculus!

$$0 \overset{set}{=} \frac{d}{d\theta}\left[\quad\right] = 6\left(\bar{x}\left(\frac{1}{\theta}\right) + (1-\bar{x})\frac{1}{1-\theta}\right)$$

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

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

from Ps2#1

$$0 = \bar{x}(1-\theta) - (1-\bar{x})\theta = \bar{x} - \cancel{\bar{x}\theta} - \theta + \cancel{\bar{x}\theta} \Rightarrow \boxed{\theta = \bar{x} = \hat{\theta}_{MLE}} = \hat{p}$$

In our case $\bar{x} = \frac{2}{6} = \frac{1}{3}$

Our calc. above was independent of $n=6$. It works for all n.

$\hat{\theta}_{MLE} = \bar{x}$ estimate, $\hat{\theta}_{MLE} = \bar{X}$ "estimator" analogous notation!

MLE's aren't the only estimators but they have really nice properties

"Consistency"

(1) $\hat{\theta}_{MLE} \xrightarrow{p} \theta$

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

$\Rightarrow$ $\hat{\theta}_{MLE}$ becomes arbitrarily close to θ with high n

(2) Asymptotic Normality

$$\hat{\theta}_{MLE} \xrightarrow{d} N\left(\theta, SE[\hat{\theta}_{MLE}]^2\right)$$

(3) Efficiency $SE[\hat{\theta}_{MLE}]$ is the theoretically lowest std error for all consistent estimators (Cramer-Rao bound)

In the Bernoulli case $\hat{\theta}_{MLE} = \bar{X}$

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

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta) \Rightarrow \hat{\theta}_{MLE} = \bar{X}$

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Geom}(\theta) := \underbrace{(1-\theta)^x \theta}$

different parametrization than in HW2#1

Here, X is the # of failures before the first success.

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

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

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

$$\ell(\theta; x) = \ln(\quad) = n \ln\theta + \sum x_i \ln(1-\theta)$$

$$\ell'(\theta; x) = \frac{n}{\theta} - \frac{\sum x_i}{1-\theta} \overset{\text{set}}{=} 0 \Rightarrow n(1-\theta) - \theta \sum x_i = 0$$

$$\Rightarrow n - n\theta = \theta \sum x_i$$

$$\Rightarrow n = \theta\left(\sum x_i + n\right)$$

$$\Rightarrow \hat{\theta}_{MLE} = \frac{n}{\sum x_i + n} = \frac{n}{\bar{x} n + n} = \frac{1}{\bar{x}+1}$$

What would it be for alt parameterization? $\frac{1}{\bar{X}}$ since $\bar{X}$ always gets one more trial

$SE[\hat{\theta}_{MLE}] = SE\left[\frac{1}{\bar{X}+1}\right]$ <... HARDER but totally possible

Inference Based on MLE's (frequentist method)

---

(1) Pt Estimate $\hat{\theta}_{MLE}$

(2) Conf set $CI_{\theta, 1-\alpha} := \left[\hat{\theta}_{MLE} \pm z_{\frac{\alpha}{2}} \hat{SE}[\hat{\theta}_{MLE}]\right]$

(3) Hyp testing: $H_0 : \theta = \theta_0$

$H_a : \theta \neq \theta_0$ your theory

Retainment Region $:= \left[\theta_0 \pm z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}] \big|_{\theta_0}\right]$

if $\hat{\theta}_{MLE} \in$ Retainment Region $\Rightarrow$ Fail to reject

Observe Data $\rightarrow$ Pick $\mathcal{F} \rightarrow$ Do Inference via MLE

What's wrong with this????