Math 341 - Lecture 2 01/31/19

$X = \langle 0, 0, 1, 0, 1, 0 \rangle$

This all you see in the real world.

$\mathcal{F}$ = iid Bernouilli

$$X_1, \dots, X_6 \overset{iid}{\sim} \text{Bern}(\theta) = \theta^x (1-\theta)^{1-x}$$

Three Goals in Statistical Inference

1. Point-Estimation ~ Best Guess of θ
2. Confidence Set - internal of plausible values of θ.
3. Test Theories about θ (hypothesis tests)

$H_0 : \theta = \frac{1}{4{,}000{,}000}$

$$p(x_1, \dots, x_6; \theta) \overset{iid}{=} \prod_{i=1}^{6} p(x_i; \theta)$$

$$p((0,0,1,0,1,0); \theta = \underbrace{(\theta^0 (1-\theta)^{1-0})}_{1-\theta} \underbrace{(\theta^0 (1-\theta)^{1-0})}_{1-\theta} \underbrace{(\theta^1 (1-\theta)^{1-1})}_{\theta}$$

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

What if $\theta = 1/2 = (1/2)^6 = .0156$

$\theta = 1/7 \cdot (1/7)^2 (3/4)^4 = .0198$

$$p(x_1, \dots, x_6; \theta) = \mathcal{L}(\theta; x_1, \dots, x_6)$$

$$\underbrace{p(\vec{x}; \theta)}_{\text{jmf/jdf}} = = \underbrace{\mathcal{L}(\theta, \vec{x})}_{\text{likelihood}} \in (0,1) ?$$

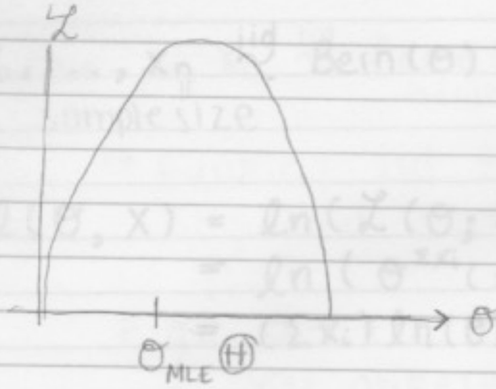
^ "hat" usually denotes an estimate.

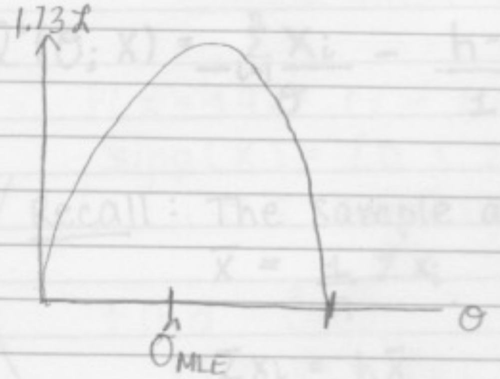
$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\text{avgmax}} \{\mathcal{L}(\theta; \vec{x})\}$$

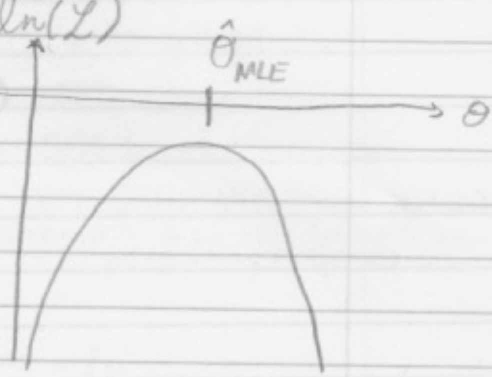
Maximum Likelihood Estimate

$$= \underset{\theta \in \Theta}{\text{avgmax}} \{ g(\mathcal{L}(\theta, x))$$

where g is a strictly increasing function.







$$\ell(\theta; x) = h(\theta^2(1-\theta)^4)$$
$$= 2h\theta + 4h(1-\theta)$$

$$\ell'(\theta, x) = \frac{2}{\theta} - \frac{4}{1-\theta} \overset{set}{=} 0$$

$$\Rightarrow \frac{2}{\theta} = \frac{4}{1-\theta} \Rightarrow (2(1-\theta)) = 4\theta$$

$$\Rightarrow 2 - 2\theta = 4\theta$$

$$2 = 6\theta \Rightarrow \hat{\theta}$$

$X_1, \dots, X_n \overset{iid}{\sim} \text{Bern}(\theta)$ Find $\hat{\theta}_{MLE}$

sample size

$$\ell(\theta, X) = \ln(\mathcal{L}(\theta; X)) = \ln\left(\prod_{i=1}^{n} \theta^{x}(1-\theta)^{1-x}\right)$$
$$= \ln\left(\theta^{\Sigma x_i}(1-\theta)^{\frac{h-\Sigma x_i}{\Sigma x_i}}\right)$$
$$= (\Sigma x_i)\ln(\theta) + (h - \Sigma x_i)\ln(1-\theta)$$

$$\ell'(\theta; X) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{h - \Sigma x_i}{1-\theta} \overset{set}{=} 0$$

Recall: The sample average
$$\bar{x} = \frac{1}{n}\Sigma x_i$$
$$\Sigma x_i = h\bar{x}$$

$$\to \frac{n\bar{x}}{\theta} - \frac{n - n\bar{x}}{1-\theta} = 0$$
$$\to \frac{\bar{x}}{\theta} - \frac{1-\bar{x}}{1-\theta} = 0$$
$$\to \frac{\bar{x}}{\theta} = \frac{1-\bar{x}}{1-\theta}$$
$$\to \bar{x}(1-\theta) = \theta(1-\bar{x})$$
$$\to \bar{x} - \bar{x}\theta = \theta - \theta\bar{x}$$
$$\to \boxed{\hat{\theta}_{MLE} = \bar{x}}$$

Math 633 you prove that...

1. $\hat{\theta}_{MLE}$ is "consistent" meaning
   $\hat{\theta}_{MLE} \to \theta$ when n gets large.
2. $\hat{\theta}_{MLE}$ converges to
   $\hat{\theta}_{MLE} \to N(\theta; SE[\hat{\theta}_{MLE}]^2)$
   "Asymptotic Normality"
3. $\hat{\theta}_{MLE}$ is "Efficient" i.e. Minimum/lowest variable among all consistant estimates.

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

parametrize by:

↳ Running iid Bernoullis until first "1" (success). Stop and count # of 0's (failures).

E.g. $\theta = 1\%$.
you observe $\underbrace{0, 0, \ldots, 0}_{49}, \underbrace{1}_{50^{th}} \rightarrow x = 49$

$P(X = 49) = (1 - 1\%)^{49} (1\%) \approx .0061$
$\text{Supp}(X) = \{0, 1, 2, \ldots\}$ — possible values it can output if you can count # of fail before success:
$\theta \in (0, 1) = \Theta$ parameter space.

Find $\hat{\theta}_{MLE}$ log likelihood of $x = \ln$
$\ell(\theta; X) = \ln\left(\prod_{i=1}^{n} (1-\theta)^{x_i} \theta\right) = \ln\left((1-\theta)^{\Sigma x_i} \theta^n\right)$

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

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

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

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

$\rightarrow \hat{\theta}_{MLE} = \frac{1}{\bar{x}+1}$ what if $\bar{x} = 99 \rightarrow \hat{\theta}_{MLE} = 1\%$.
what if $\bar{x} = 0 \rightarrow \hat{\theta}_{MLE} = 100\%$.

* disturbing because not in parameter space.
* scary estimate because "always"

What does MLE property #2 imply?

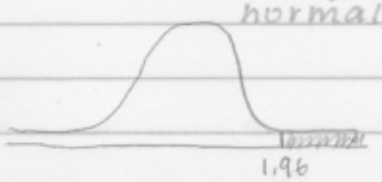
For iid Bernoulli $\hat{\theta}_{MLE} \approx N(\theta, SE(\bar{X})^2)$

$$= N\left(\theta, \left(\sqrt{\frac{\theta(1-\theta)}{n}}\right)^2\right)$$

For iid Geom $\hat{\theta}_{MLE} \approx N(\theta, SE[\frac{1}{1-\bar{x}}]^2)$

Confidence Interval

$$\underset{\substack{\uparrow\\ \text{parameter}}}{CI}_{\theta, 1-\alpha}\ \underset{\text{level}}{} = [\hat{\theta}_{MLE} \pm \overbrace{\underbrace{Z_{\frac{\alpha}{2}}}_{\text{Standard normal quantile due to asymptotic normality.}} SE[\hat{\theta}_{MLE}]\Big|_{\theta=\hat{\theta}_{MLE}}}^{\text{Margin of Error}}]$$



$= [.37, .57] = [.47 \pm .10]$

Hypothesis Test

$H_0: \theta = \theta_0$ - Null

$H_1: \theta \neq \theta_0$ - Alternative

outcome is:

Retain H_0 Reject H_0

can only do this if Random Var. is normal which you get from #2.

Normal when n is large.

Retainment Region $= [\theta_0 \pm Z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\Big|_{\theta=\theta_0}]$

$\hat{\theta}_{MLE} \in$ Retain Region $\rightarrow$ Retain H_0

$\rightarrow$ Reject H_0

MLE is projection in frequentism inference.

Frequentism: Believing

1. θ is fixed and cannot be a r.v.
2. Reliable on Repeated sampling.

1. iid Bernoulli Case

$X = (0,0,0)$

$\hat{\theta}_{MLE} = 0$ — Bad Strategy

$CI_{\theta,1-\alpha} = [0,0] = \{0\}$

2. You know for sure $\theta \in [0.1, 0.2]$

$\hat{\theta}_{MLE} = .14$