Lec 3 Math 341 1/31/19 Syllabus → Bracha's off. hr only on Thurs. (today) 1

In the real world you see $X = \langle 0, 0, 1, 0, 1, 0 \rangle$, the data.

$\mathcal{F} =$ Here, Bernoulli iid

Then, you pick $\mathcal{F}$, an assumption! But you don't know the parameter, θ!

Figuring out θ is the goal of inference. There are generally 3 sub-goals:

1. Point estimation. Provide best guess of θ
2. Confidence set. Provide a range of possible θ's.
3. Theory testing. Evaluate a theory about θ.

---

e.g. imagine data above and assume $\mathcal{F}$ is iid Bernoulli

$$p(\langle 0,0,1,0,1,0 \rangle ; \theta) = \left(\theta^0 (1-\theta)^1\right)\left(\theta^0 (1-\theta)^1\right) \cdots = \theta^2 (1-\theta)^4$$

if $\theta = 0.5$ $\quad = 0.5^6 = 0.0156$

if $\theta = 0.25$ $\quad = 0.25^2 \, 0.75^4 = 0.0198$

$\theta = 0.5$ is "more likely" than $\theta = 0.25$

the data is fixed, and we want to know how probable the value of θ are.

$$\mathcal{L}(\theta ; x) = p(x ; \theta)$$

"likelihood function": what is the likelihood of "seeing" the parameter at a certain value

$\mathcal{L}$: prob of θ with x known

p: prob of data with θ known

$\mathcal{L}(\theta;x) \in ? \ (0,1)$

$\int_{\theta\in\Theta} \mathcal{L}(\theta;x)\,d\theta \stackrel{?}{=} 1$ No...

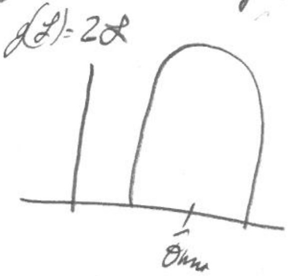
not an interesting question

What is the most likely value of the parameter?

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max}\left\{\mathcal{L}(\theta;x)\right\} = \underset{\theta\in\Theta}{\arg\max}\left\{g(\mathcal{L}(\theta;x))\right\}$$

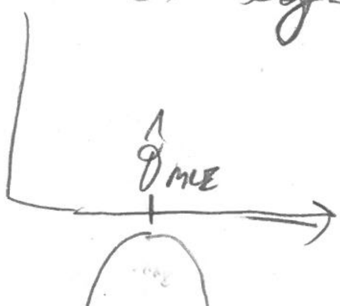
maximum likelihood estimate

where g is a strictly increasing function



$x>y \Rightarrow g(x)>g(y)$

$g(\mathcal{L}) = \ln(\mathcal{L})$ log-likelihood



$\ell(\theta;x) := \ln(\mathcal{L}(\theta;x))$ very useful! Since products become sums!

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max}\left\{\ell(\theta;x)\right\}$$ the usual definition

Back to the example

$X_1,\dots,X_6 \stackrel{iid}{\sim} \text{Bern}(\theta)$

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

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

sample avg

Note $\bar{x} = \frac{1}{n}\sum x_i \Rightarrow \sum x_i = n\bar{x} = 6\bar{x}$

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

need $\hat{\theta} = \text{argmax}\left(\ell(\theta; x)\right)$ How? Take derivative, set $= 0$

$$\frac{d}{d\theta}\left[\ell(\theta; x)\right] = 6\left(\frac{\bar{x}}{\theta} - \frac{1-\bar{x}}{1-\theta}\right) \overset{\text{set}}{=} 0 \Rightarrow \bar{x}(1-\theta) = (1-\bar{x})\theta$$

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

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

$\bar{x} = \frac{2}{6} = \frac{1}{3} \Rightarrow \theta = \frac{1}{3}$ most likely value

$x = \langle 0, 0, 1, 0, 1, 0 \rangle$ makes sense?

MLE is not the only strategy, but it has nice properties:

1. "Consistency" $\hat{\theta}_{MLE} \xrightarrow{P} \theta$ in the limit
   conv. in prob. As $n \to \infty$ $\hat{\theta}_{MLE}$ becomes arbitrarily close to θ
2. "Asymptotic Normality"
   $$\hat{\theta}_{MLE} \overset{d}{\to} N(\theta, SE[\hat{\theta}_{MLE}]^2)$$
3. "Efficiency". Among all consistent estimators, it has minimum variance.

Consider $X \sim \text{Geom}(\theta) := (1-\theta)^x \theta$

this is running Bernoullis until you get a success and counting the # of failures

lets say $\theta = 1\%$

$$\underbrace{0, 0, 0, \ldots\ldots, 0}_{49}, \underbrace{1}_{50\text{th}} \Rightarrow X = 49$$

$Supp(X) = \{0, 1, 2, \ldots\}$

$\theta \in (0, 1) = \Theta$

Same as Bernoulli

$$P(X = 49) = \underbrace{.99 \cdot .99 \cdot \ldots \cdot .99}^{49} \cdot .01 \approx .0061$$

Consider $X_1, \ldots, X_n \overset{iid}{\sim} \text{Geom}(\theta)$, we want $\hat{\theta}_{MLE}$.

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

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

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

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

Does this make sense? Let's say $\bar{x} = 99 \Rightarrow \hat{\theta}_{MLE} = 1\%$

$\bar{x} = 0 \Rightarrow \hat{\theta}_{MLE} = 100\%$

What does MLE property #2 imply?

For Bernoulli (iid)

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

For iid Geom

$$\hat{\theta}_{MLE} \approx N\left(\theta, SE\left[\frac{1}{\bar{X}+1}\right]^2\right)$$ start 633 problem

Solving for an MLE means you can fulfill the 3 goals of inference:

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

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

normal quantile due to asymptotic normality

for iid Bern.

$$= \left[\bar{X} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

③ Hypothesis test

$H_0: \theta = \theta_0$

$H_a: \theta = \theta_1$

Retainment Region $= \left[\theta_0 \pm z_{\frac{\alpha}{2}} \, SE(\hat{\theta}_{MLE})\Big|_{\theta=\theta_0}\right]$

if $\hat{\theta}_{MLE} \in RR \Rightarrow$ Retain H_0

for iid Bern

evaluated at

$$= \left[\theta_0 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right]$$

Frequentist inference:

reliance on repeated sampling and treatment of θ as absolutely fixed

This is the usual way to do "frequentist inference".
What are some problems with this strategy?

(1) iid Bern case $\hat{\theta}_{MLE} = \bar{X}$ what if $X = \langle 0,0,0 \rangle$
$\Rightarrow \hat{\theta}_{MLE} = 0$

Is this a reasonable pt est? Is $[0,0]$ a reasonable CI?

(2) what if you know that $\theta \notin [0.1, 0.2]$. The MLE will ignore this.

(3) Frequentist interpretation of a CI

$CI_{\theta, 95\%} = [0.37, 0.43]$ this means

(a) If you repeated the experiment, 95% of $\hat{\theta}_{MLE}$'s will be inside of the CI created then

(b) Before you begin, there's a 95% chance $\theta \in CI$,

But for this specific $CI = [0.37, 0.43]$, no interpretation applies!

You want $P(\theta \in [0.37, 0.43]) = 95\%$ but θ is a constant! So this is nonsense!

(4) Hypothesis tests end in

Reject Ho, Retain Ho

$Pval := P(\text{seeing } \hat{\theta}_{MLE} \text{ or more extreme} \mid H_0 \text{ true})$