Lec 2
Feb 2

$X \sim Deg(C) := \{ C$ w.p 1, $Sup[X] = \{C\}$

also be a vector

In the real world you see $X = <0,0,1,0,1,0>$
Then, you pick a curly-F (ex. a parametric model)
But you don't know thetha! So you have to guess thetha. This guessing is "inference".
There are typically three goals of "statistical inference".

(1) Point estimation. Give me your best guess of theta (one value).

(2) confidence sets. Give me a range of likely theta's.

(3) Theory testing. Evaluate a theory about the value of theta.

* Assume curly f = Bernoulli. Once you make an assumption of the parametric model, you can compute the JMF or JDF:

B/c it's discrete

Compute

$$p(X;\theta) = \prod_{i=1}^{6} p(X_i;\theta)$$

left with $1-\theta$

$$P(<0,0,1,0,1,0>;\theta) = (\theta^0(1-\theta)^{1-0})(\theta^0(1-\theta)^{1-0})(\theta^1(1-\theta)^{1-1})\ldots = \theta^2(1-\theta)^4$$

left w/ theta

plug in the last answer you found

if $\theta = .5 \rightarrow P(x;\theta) = 0.5^2(1-0.5)^4 = 0.0156$
if $\theta = 0.25 \rightarrow P(X;\theta) = 0.25^2(1-0.25)^4 = 0.0198$

$\Rightarrow$ theta = 0.25 seem "more likely" than theta = 0.5.

$\mathcal{L}(\theta;x) = P(X;\theta)$ ← probability of the data with theta known

likelihood function, probability of theta given x known or the likelihood of "seeing" the parameter at a certain value

Parameter and Supp[X] **NOT** the same thing

$\int_{\Theta} \mathcal{L}(\theta;x)\,d\theta$ = no rule

$\sum_{Supp[X]} P(X;\theta) = 1$

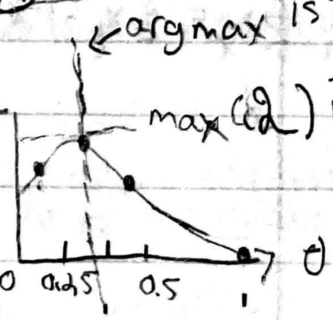
$\sum_{\Theta} \mathcal{L}(\theta;x) = \infty$

"hat means "estimator" representing your guess

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

~~maximum~~ ?

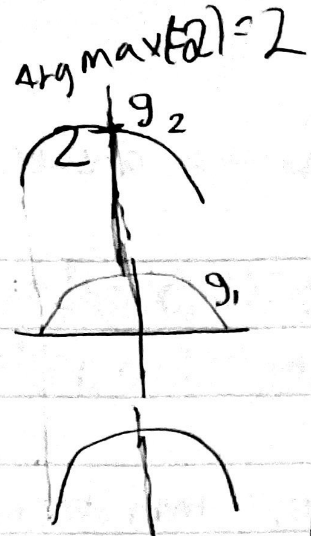
is the x value corresponding is where it's maximal



Let g be a strictly increasing function $= \underset{\theta \in \Theta}{\text{argmax}} \{g(\mathcal{L}(\theta,x))\}$

The $\hat{\theta}$ estimates to $.33\bar{3}$.

graph on the next page



* if g function is strictly increasing the argmax remains the same.

log-likelihood

let $g = \ln$

$= \text{argmax}\{\ln(\mathcal{L}(\theta; x))\}$ estimator is large

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

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

product

indep.

$$\ell(\theta; x) = \ln(P(X; \theta)) = \ln\left(\prod_{i=1}^{n} p(x_i, \theta)\right) = \sum_{i=1}^{n} \ln(f(x; \theta))$$

The log of product becomes the sum of logs $\sum_{i=1}$

$n = 6$

In our example of $x = <0, 0, 1, 0, 1, 0> \ldots$

$$\ell(\theta; x) = \sum_{i=1}^{6} \ln(\underbrace{\theta^{x_i}}_{f} \underbrace{(1-\theta)^{1-x_i}}_{g}) = \sum_{i=1}^{6} (\underbrace{x_i}_{g'(x)} \underbrace{\ln(\theta)}_{f(x)} + \underbrace{(1-x_i)}_{g(x)} \underbrace{\ln(1-\theta)}_{f'(x)})$$

use product Rule output

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

Side Note:

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

?

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

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

We need to find the argmax of this function...

→ $\hat{\theta} =$ ← this represent estimate (a realization from the estimator)

take derivate of the log likeihood wrt theta and set = 0 and solve

Θ ← parameter space

the realization of 6 cancels

$$\frac{d}{d\theta}\left[\ell(\theta;x)\right] = \cancel{6}\left(\frac{\bar{x}}{\theta} - \frac{1-\bar{x}}{1-\theta}\right) = 0 \qquad = \frac{\bar{X}}{\theta} = \frac{1-\bar{X}}{1-\theta}$$

converse

set it to 0 to find maximum

$$= \bar{x}(1-\theta) = (1-\bar{x})\theta$$
$$= \bar{x} - \cancel{\bar{x}\theta}$$
$$= \theta - \cancel{\bar{x}\theta}$$
$$= \hat{\theta}$$
$$\bar{x} = \frac{2}{6} = .3\bar{3} = 1/3$$

The estimator $\hat{\theta}^{MLE} = \bar{x}$ is a random variable whose realization are the estimates. This random variable has nice properties:

1. $\hat{\theta}_{MLE}$ is "consistent". This means that this estimator can provide arbitrary precision on theta given enough n.
2. $\hat{\theta}_{MLE} \dot{\sim} N(\theta, SE[\hat{\theta}_{MLE}]^2)$ asymptotic normality (SE: standar error)
3. "Efficiency" means that among all consisten estimators, it has minimum variance. going to be a constant

Consider $X \sim Geom(\theta)$: $(1-\theta)^x \theta = (1-\theta)^x\theta =$ $Supp[X] = \{(0,1,2..)\}$, $\Theta = (0,1)$

consider a sequence of iid Bernoulli theta. This random variable tells you the # of failures (realization of zero) before the first success (realization of one). →

If $\theta = 1\%$.

$$\underset{1^{st}}{0}, \underset{2^{nd}}{0}, \underset{3^{rd}}{0} \cdots \underset{44^{th}}{0}, \underset{50^{th}}{1} \qquad X=49$$

$$P(X=49;\theta = 0.99^{49}\ 0.01$$

$\tilde{F}$ = iid geometric . n realization

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

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

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

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

Let's Find the MLE. We take the derivate of the likeihood wrt theta and set it equal to zero and solve.

property 1:

$$\frac{d}{d\theta}[\ell] = \cancel{n}\left(-\frac{\bar{x}}{1-\theta} + \frac{1}{\theta}\right) = 0$$

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

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

Scanned with CamScanner

$\hat{X}$ = estimator of X

$\bar{X}$ =

$$\frac{1}{\theta} - 1 = \bar{X}$$

$$= \frac{1}{\theta} = \bar{X} + 1$$

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

consider $\bar{x} = 49$

$$\hat{\theta}_{MLE} = \frac{1}{49+1} = 2\%$$

Let's examine MLE <u>property #2</u>:

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

In the curly-F: iid Bernoulli case,

$$\hat{\theta}_{MLE} = \bar{X}, \quad SE[\hat{\theta}^{MLE}] = SE[\bar{X}] = \overset{\text{standard deviation}}{\sqrt{var[x]}} \overset{\text{MA241}}{=} \sqrt{\frac{\sigma^2}{n}} = \frac{\sqrt{\theta(1-\theta)}}{n}$$

In the curly-F: iid Geometric case,

$$\hat{\theta}_{MLE} = \frac{1}{\bar{X}+1}, \quad SE\left[\frac{1}{\bar{X}+1}\right] = ?$$

We now use property 2 to attack the other goals of interference:

Confidence Sets: we use a method called the "confidence interval":

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

parameter

level of confidence

standard normal quantile at alpha/2.

For the iid Bernoulli case,

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

letting $1-\alpha = 95\% = \alpha = 5\%$

need to use th z table

$$CI_{\theta, 95\%} = \left[\bar{X} \pm 1.96 \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

$z = 1.96$