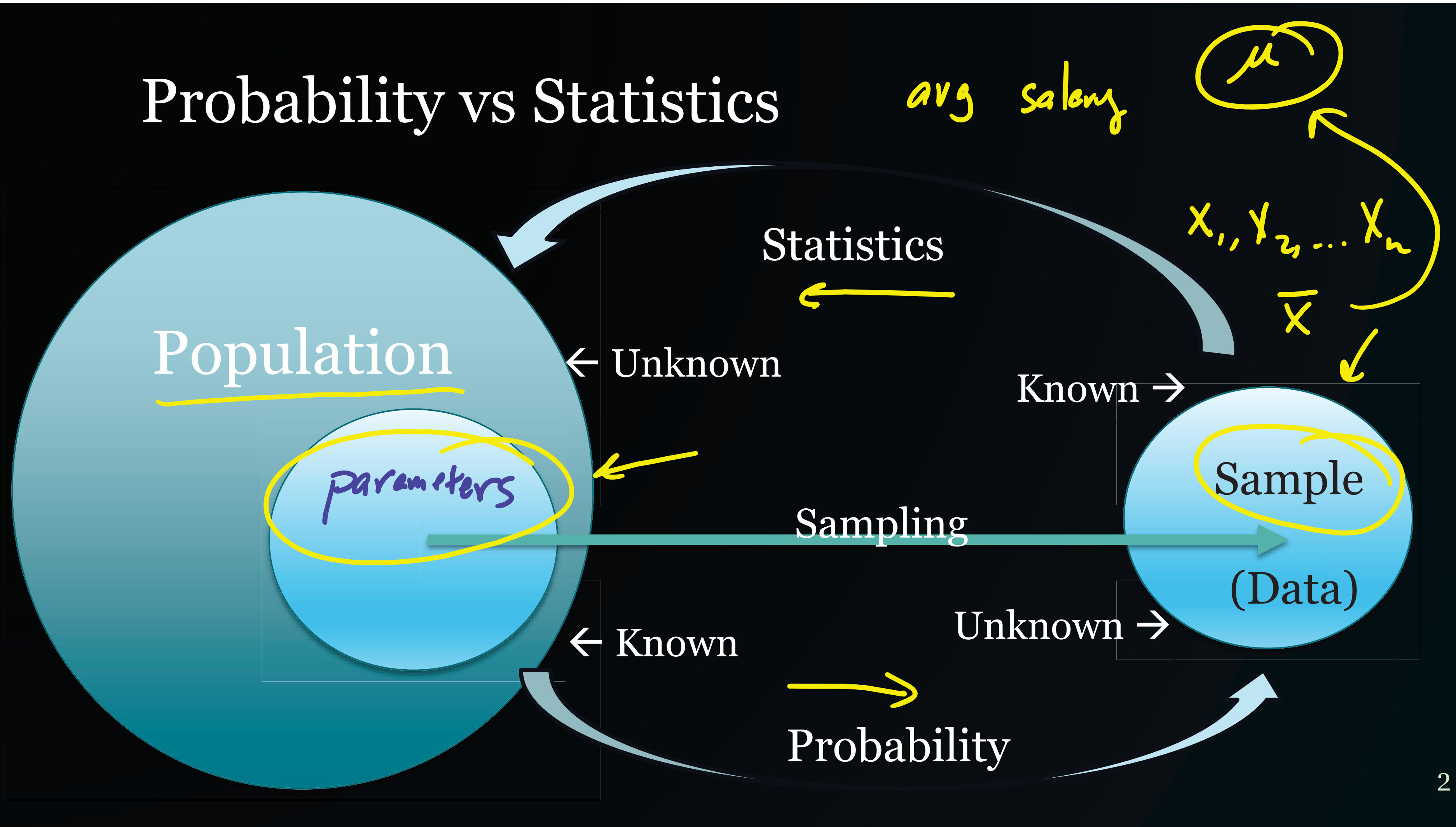
Point Estimation (MLE) (6.4)

Notes

Probability vs Statistics

avg salary



parameter $\to \boxed{\sigma^2} = E[(X-\mu)^2]$

$\to \mu$

Some new terms

statistic $\to \bar{X}$

value of

Sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample variance: $\boxed{\boxed{s^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Sample standard deviation: s = $\sqrt{s^2}$

$\sigma = \sqrt{\sigma^2}$

notes

in STATISTICS

Data	Population
Statistics	Parameters
known	unknown (want to estimate)
$\bar{X}$ $\bar{x}$ $\rightarrow$	μ
s^2 $\rightarrow$	σ^2

MLE　　MOM　　θ

μ

Point Estimation

Let's say we are given a distribution (family), and have random samples from this distribution, but don't know the value of the parameter, θ.

(often, we use θ as the generic term for an unknown parameter).

Definition: The **parameter space, Ω**, is the range of all possible values of θ.

S　$\mathcal{X}$

Ω : parameter space

p

$n = 13$

E.g.

$X \sim \text{Exp}(\theta)$, $\quad \Omega = \{\theta : \theta > 0\}$

$X \sim \text{Binom}(n, \theta)$ $\quad \Omega = \{\theta : 0 < \theta < 1\}$

$X \sim \text{Geom}(\theta)$ $\quad X \sim \text{Geom}(p)$ $\quad$ " $\quad$ "

$X \sim N(\theta, 1)$

$\Omega = \{\theta : -\infty < \theta < \infty\}$

s^2

$X \sim N(\theta_1, \theta_2)$

X x

Point Estimation

iid

- Goal: Estimate $\theta \in \mathbf{\Omega}$
- We will observe n samples, X_1, X_2, …, X_n, and estimate θ with n sample values, x_1, x_2, …, x_n.

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ μ

The statistic, $u(X_1, X_2, \ldots X_n)$, is an **estimator** of θ.

e.g. $\bar{x} = 4.3$

Using the values from the observations, we can find an **estimate** of θ, $u(x_1, x_2, \ldots x_n)$.

notes

median

use $\bar{X}$ to estimate μ

estimator

Simple example of Point Estimate

What was the true mean score on Midterm 2? (μ)

- Population: All students in Math 463/Stat 400
- Right now: only have 30 grades.
- $\bar{x} = \frac{1}{30}\sum_{i=1}^{30} x_i = 85$

estimate of μ

This is a point estimate of μ.

Binomial example

Suppose that I perform an experiment 10 times and define success as 1. I don't know p. We would like to estimate the parameter, p.

- If I get a sample: $\underbrace{1,1,1,1,1,0,1,1,1,1}$. What is the best estimate for p?

9 s   1 f $\boxed{.9}$

$X \quad Y \qquad X_1, X_2$

sample size n

Joint pmf/pdf

- Assuming that X_1, X_2, … X_n are independent and identically distributed, we know that the joint pmf (or pdf) is equal to the product of the pmf/pdfs.

- Bivariate: if X and Y are independent,

$$\underbrace{f(x,y) = f(x)f(y)}$$

Joint pmf/pdf

mutually

For multiple independent observations from the same distribution (iid), the **joint distribution** (joint pdf or joint pmf) is the product of the individual pdf/pmfs.

$X_1, X_2, \dots X_n$ iid

$$f(x_1, x_2, \dots x_n) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) = \prod_{i=1}^{n} f(x_i)$$

notes

adithya__r: = so for 10 flips, we got 9 successes, and got p= 0.9. If n is increased to say 1000 flips, p is likly to be 'more accurate'. But instead of increasing 'n', what if we just used n = 10 flips but performed the trial of 10 flips multiple times, averaging the found p value for that?

The Likelihood function

Statistics

X $\boxed{f(x)}$

The likelihood function looks exactly like the joint pdf (or pmf).

It is obtained through finding the joint distribution.

$X \sim \text{Binom}(n,p)$ $f(x) = \boxed{\binom{n}{x} p^x (1-p)^{n-x}}$

It is a function of θ, not of x_i.

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

semicolon

$f_X(x)$

$f(x;p) =$ " "

$f(x,p) =$ " "

$f(x|p) =$ " "

Probability: Know the value of parameters. Calculate probability of observing some data.

Statistics: Know what data look like. Come up with an estimate of parameters.

$$\prod_{i=1}^{3} x_i = x_1 \cdot x_2 \cdot x_3$$

Binomial Example (MLE)

$\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$ $\boxed{.9}$

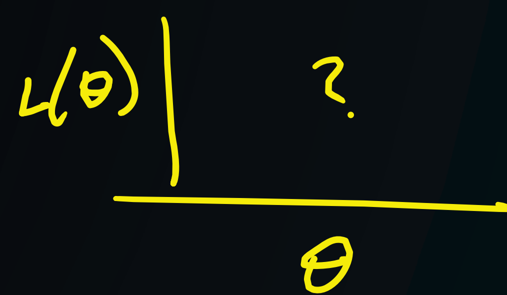
- Flipping a loaded coin 10 times. It shows heads on 9 of 10 flips.

$X = 9$

Let $X \sim Binom(10, \theta)$

$$f(9;\theta) = \binom{10}{9}\theta^{9}(1-\theta)^{1} = \boxed{L(\theta)}$$

x 'p'



- Now, given that I have gotten 9 successes, what value of theta makes this expression the largest (most likely)? How can I find that value?

$$f(9;\theta) = \binom{10}{9}\theta^9(1-\theta)^1 = L(\theta)$$

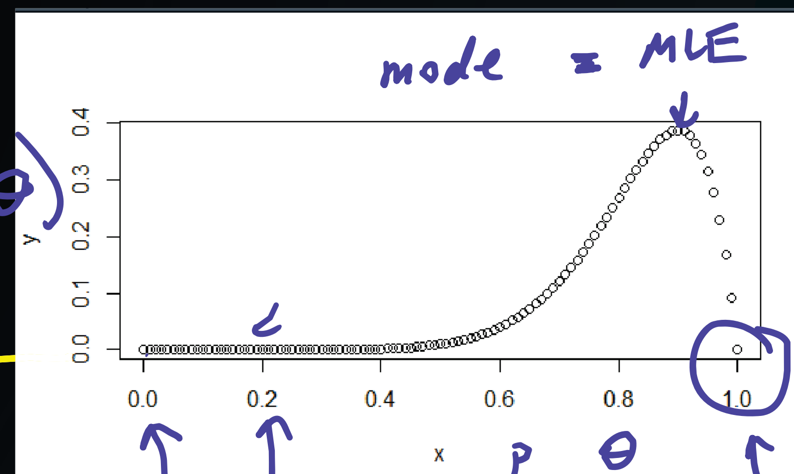
θ

```
x = seq(from = 0, to = 1, by = 0.01)
y = 10*x^9 * (1-x)
plot(x,y)
```

0 θ 1

$L(\theta)$

mode = MLE



p θ

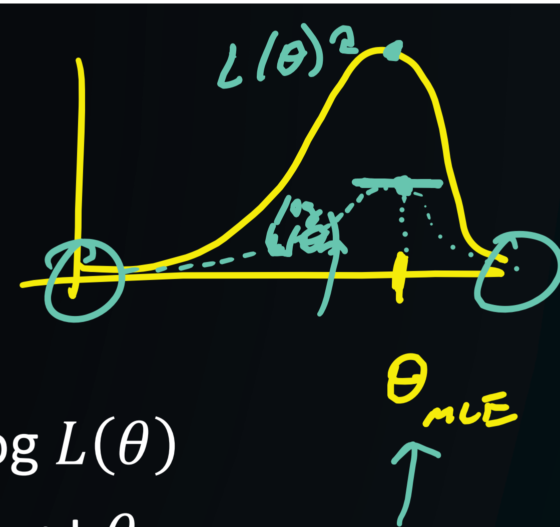
p=0.5

p=1

log means natural log

Using Calculus to find MLE

ln means natural log



Step 1: Find $L(\theta)$

Step 2: Take the (natural) log of $L(\theta)$, $\log L(\theta)$

Step 3: Take first derivative of $\log L(\theta)$ w.r.t θ.

Step 4: Set expression equal to 0

Step 5: Solve for θ

In case you forgot…

Suppose $f(x)$ is a function of x that is twice differentiable at a stationary point x_0.

1. If $f''(x_0) > 0$, then f has a local minimum at x_0.
2. If $f''(x_0) < 0$, then f has a local maximum at x_0.

notes

$\log(ab) = \log a + \log b$ $\quad f(x_i;\theta)$ pdf cont

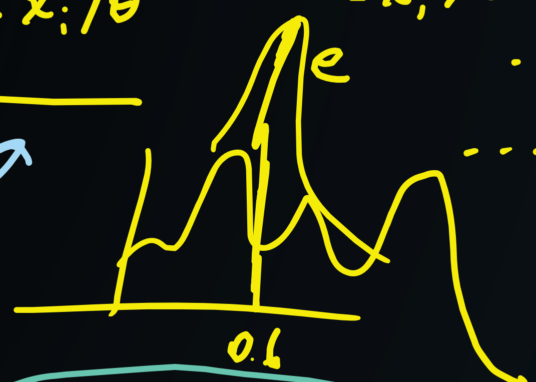
MLE Example

Let $X_1, X_2, \dots X_n$, be iid $\sim f(x;\theta) = \theta^{-2} x e^{-x/\theta}$, $x > 0$, $\theta > 0$

a) Find the MLE of θ, $\hat{\theta}$.

(1) $L(\theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \theta^{-2} x_i e^{-x_i/\theta}$

$e^{-x_1/\theta} \cdot e^{-x_2/\theta} \cdots$



$= \theta^{-2n} \prod_{i=1}^{n} x_i \, e^{-\frac{1}{\theta}\sum x_i}$

(2)

$\log L(\theta) = \boxed{-2n \log \theta} + \sum \log x_i + -\frac{1}{\theta}\sum x_i$

$\frac{\partial \log L(\theta)}{\partial \theta} = -2n\left(\frac{1}{\theta}\right) + 0 + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$

$\log\left(\prod x_i\right) = \log(x_1 \cdot x_2 \cdots x_n) = \log x_1 + \log x_2 + \dots \log x_n$

MLE Example continued

$$f(x;\theta) = \theta^{-2} x e^{-\frac{x}{\theta}}, \qquad x > 0, \qquad \theta > 0$$

(3,4)

$$\frac{\partial \log L(\theta)}{\partial \theta} = -2n\left(\frac{1}{\theta}\right) + \cancel{0} + \frac{1}{\theta^2}\sum_{i=1}^{n} X_i = 0$$

$$-2n\theta + \sum_{i=1}^{n} X_i = 0$$

(5)

$$\sum_{i=1}^{n} X_i = 2n\hat{\theta} \Rightarrow \boxed{\frac{\sum_{i=1}^{n} X_i}{2n} = \hat{\theta}}$$

OR

$$\boxed{\bar{X}/2 = \hat{\theta}}$$

MLE Example continued

$\rightarrow \theta$ unknown

pdf

$f(x;\theta) = \boxed{\theta^{-2} x e^{-x/\theta}}$

Find an estimate of θ when

$X_1 = x_1 = 1$

$x_1 = 1$, $x_2 = 0.75$, $x_3 = 2$, $x_4 = 1.5$, $x_5 = 0.75$

$\frac{6}{5} = 1.2$

$\boxed{\bar{X}/2 = \hat{\theta}}$ OR

Estimator

$\boxed{\frac{\bar{x}}{2} = \frac{1.2}{2} = 0.6}$

estimate

Theta = seq(from=0, to = 5, by = 0.01)
Y = theta^(-2)*

MLE Example (for you to practice at home)

Let $X_1, X_2, \dots X_n, \sim$Bern(p). $\boxed{\text{Find the MLE of } p.}$ $\boxed{\text{estimator}}$

$l(\theta)$

$$f(x_i; p) = p^{x_i} (1-p)^{1-x_i}$$

$L(\theta)$

(1) $L(p) = \prod_{i=1}^{n} f(x_i; p)$ joint pmf

pdf

$L(p)$ $= p^{x_1} (1-p)^{1-x_1} \cdot p^{x_2} (1-p)^{1-x_2} \cdot \dots \cdot p^{x_n} (1-p)^{1-x_n}$

$= p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n - \sum_{i=1}^{n} x_i}$

$L(\theta) = f(x_1, x_2, \dots x_n)$

$\log L(p)$ (2) $= \sum x_i \cdot \log(p) + (n - \sum x_i) \log(1-p)$

$$\log L(p) \quad (2) = \sum X_i \cdot \log(p) + \left(n - \sum X_i\right) \log(1-p)$$

$$\frac{\partial \log L(p)}{\partial p} \quad (3,4) = \sum_{i=1}^{n} X_i \cdot \left(\frac{1}{p}\right) + \left(n - \sum_{i=1}^{n} X_i\right)\left(\frac{-1}{1-p}\right) \overset{set}{=} 0$$

$$\frac{\sum X_i}{p} - \frac{n}{1-p} + \frac{\sum X_i}{1-p} = 0$$

(5)

$$p(1-p) \qquad (1-p)\sum X_i - pn + p\sum X_i = 0$$

$$\sum_{i=1}^{n} X_i = np \qquad \boxed{\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}} \qquad \hat{p}_{MLE} = \bar{X}$$

notes