Parametric Estimators

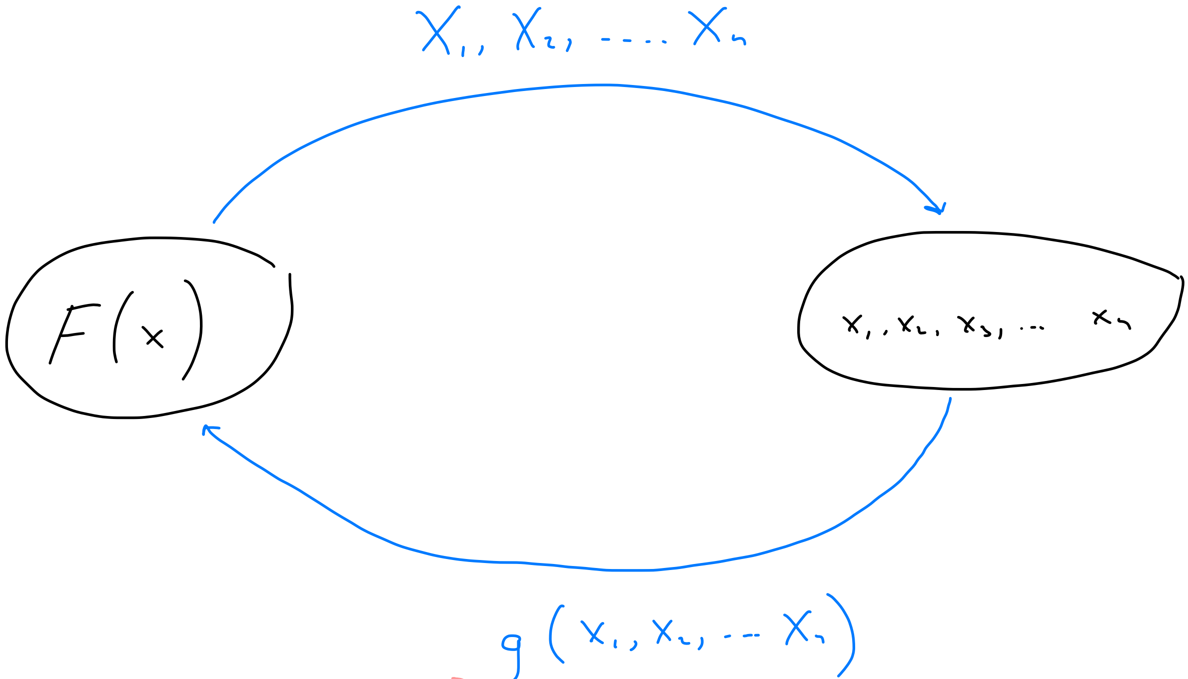
Stat 510

- PARAMETRIC MODELS
- CHECKING ASSUMPTIONS
- METHOD OF MOMENTS
  - EXAMPLES
- LIKELIHOOD
- MAXIMUM LIKELIHOOD
  - EXAMPLES
  - EQUIVARIANT
  - OPTIMIZATION
- PARAMETRIC BOOTSTRAP

$$X_1, X_2, X_3, \ldots\ldots X_n \sim F$$

How do we "infer" F?

LEARN



STATISTIC

↳ A FUNCTION OF DATA

↳ CAN BE A RANDOM VARIABLE

$X_1, X_2, \ldots\ldots X_n \sim F$

- How to estimate F ?

STATISTICAL MODEL

- Parametric Estimation → MOM, MLE

- Nonparametric Estimation

$X_1, X_2, \ldots X_n \sim F$

PARAMETRIC MODEL

STATISTICAL

$$\mathcal{F} = \{ f(x;\theta) : \theta \in \Theta \}$$

DENSITY

PARAMETER THAT DETERMINES DENSITY

PARAMETER SPACE

$X_1, X_2, \ldots X_n \sim F$

Assume $f \in \mathcal{F}$ where

$$\mathcal{F} = \left\{ f(x;p) = p^{x}(1-p)^{x-1}, \quad 0<p<1 \right\}$$

$X_1, X_2, \ldots X_n \sim \text{Bernoulli}(p)$

Problem reduced to estimating p

$$X_1, X_2, \ldots X_n \sim \text{Normal}(\mu, \sigma^2)$$

Assume $f \in \mathcal{F}$ where

$$\mathcal{F} = \left\{ f(x;\mu,\sigma^2) : \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \overbrace{\mu \in \mathbb{R}, \sigma > 0}^{\text{Parameter space}} \right\}$$

If parameter of interest

Nuisance parameter

$$X_1, X_2, \ldots X_n \sim N(\mu, \sigma^2)$$

Point Estimation

Parameter of interest

USE $\hat{\theta}$ AS "BEST GUESS" FOR θ.

Fixed, unknown

Given $X_1, X_2, \ldots X_n \sim F$

Point estimator $\longrightarrow$ $\hat{\theta}_n = \underbrace{g}_{\text{Function}}(\underbrace{X_1, X_2, \ldots X_n}_{\text{Data}})$

Random Variable

Sampling distribution...

Checking Assumptions

$$\underbrace{X_1, X_2, \dots X_n \sim N(\mu, \sigma^2)}$$

$$X_1, X_2, \dots X_n \underbrace{\sim N(\mu, \sigma^2)}_{\text{Assumption}}$$

- Graphical
- Formal Tests

Method of Moments

DISTRIBUTION IS A FUNCTION OF THE PARAMETER

Parameter $\theta = (\theta_1, \ldots, \theta_k)$

j^{th} moment $\alpha_j \equiv \alpha_j(\theta) = E_\theta\left[X^j\right] = \int x^j \, dF_\theta(x)$

MOMENTS ARE FUNCTIONS OF THE PARAMETER

j^{th} sample moment $\hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$

METHOD OF MOMENTS MoM

THE MoM ESTIMATOR $\hat{\theta}_n$ IS THE VALUE OF θ SUCH THAT

$$\alpha_1(\hat{\theta}_n) = \hat{\alpha}_1$$

$$\alpha_2(\hat{\theta}_n) = \hat{\alpha}_2$$

SAMPLE SECOND MOMENT

SECOND MOMENT

$$\vdots$$

$$\alpha_k(\hat{\theta}_n) = \hat{\alpha}_k$$

k EQUATIONS, k UNKNOWNS $\longrightarrow$ SOLVE

Example

$$X_1, X_2, \ldots, X_n \sim \text{Geom}(p)$$

$$\alpha_1 = \mathbb{E}[X] = 1/p$$

$$\hat{\alpha}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{X}$$

$$1/\hat{p} = \bar{X} \Rightarrow \hat{p} = 1/\bar{X} \quad \checkmark$$

MOM ESTIMATOR

Example

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

$$\alpha_1 = \mathbb{E}[X] = \mu$$

$$\alpha_2 = \mathbb{E}[X^2] = \mathbb{V}[X] + (\mathbb{E}[X])^2 = \sigma^2 + \mu^2$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

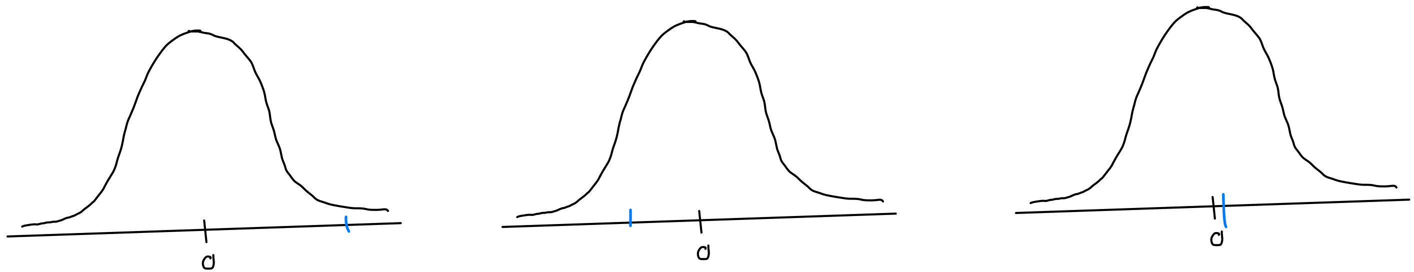
$$\Longrightarrow$$

$$\hat{\mu} = \overline{X}_n$$

$$\hat{\sigma}^2 = \frac{1}{n} \left(\sum X_i - \overline{X}_n \right)^2$$

Likelihood

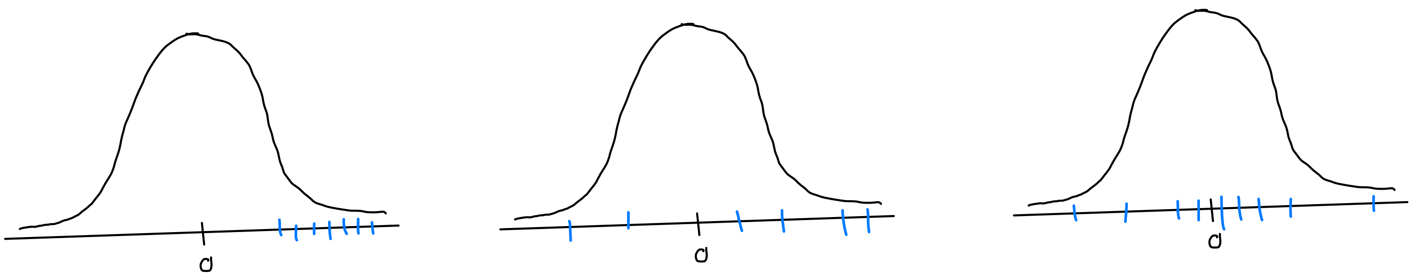
Consider $X \sim \underbrace{N(0,1)}_{\text{Known}}$



More likely to be sampled

LIKELIHOOD

CONSIDER $X_1 \dots X_n \sim \underbrace{N(0,1)}_{\text{KNOWN}}$

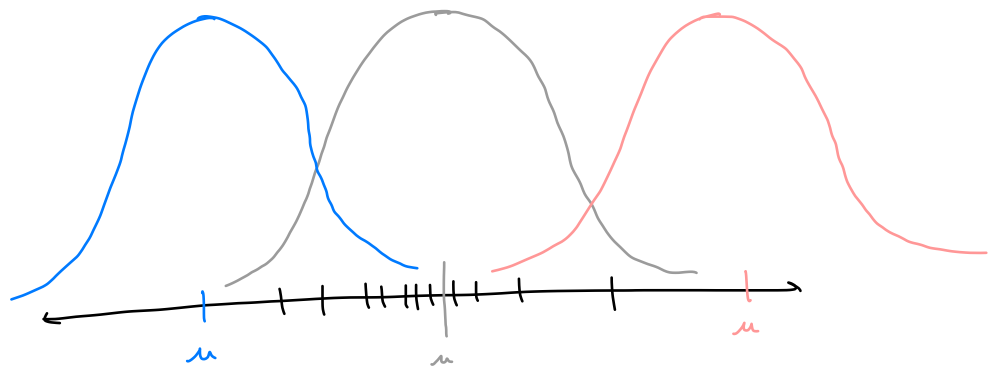


MORE LIKELY TO BE SAMPLED

LIKELIHOOD

Assume $X_1, \dots, X_n \sim N(\mu, 1)$

UNKNOWN



LIKELIHOOD FUNCTION

$X_1 \dots X_n$ IID WITH PDF $f(x;\theta)$

$$\mathcal{L}_n(\theta) = \underbrace{f(x_1, x_2, \dots x_n ; \theta)}_{\text{JOINT DENSITY IF } \theta \text{ KNOWN}} \underset{\text{IID}}{=} \prod_{i=1}^{n} f(x_i, \theta)$$

ASSUMES DATA OBSERVED

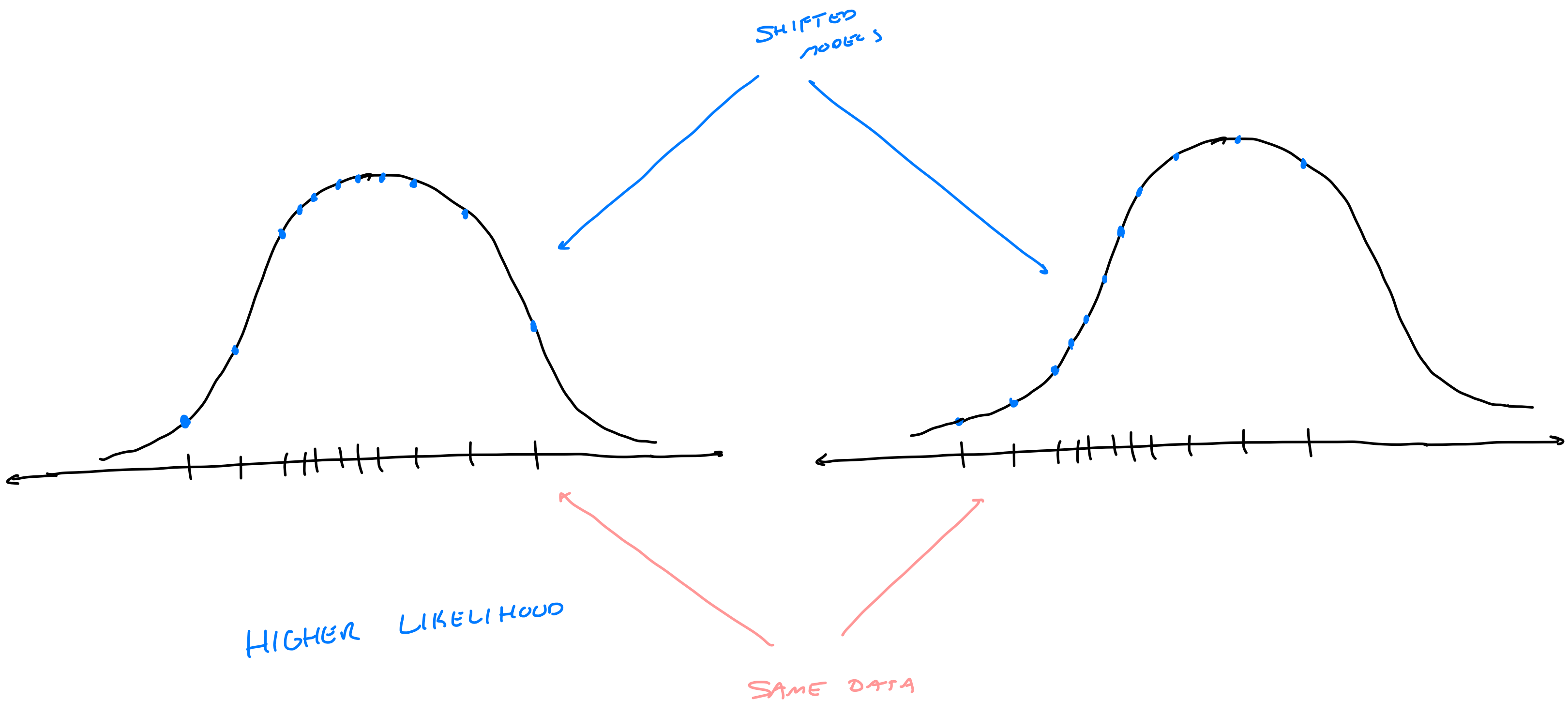
FUNCTION OF $\underline{\theta}$

$$\mathcal{L}_n(\theta) : \Theta \longrightarrow [0, \infty)$$

DOES NOT NEED TO INTEGRATE TO 1

LOG-LIKELIHOOD

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta)$$


SHIFTED MODELS
HIGHER LIKELIHOOD
SAME DATA

Maximum Likelihood Estimator

MLE

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}} \; \mathcal{L}_n(\theta)$$

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}} \; \log \mathcal{L}_n(\theta)$$

EXAMPLE

$X_1, \dots, X_n \sim \text{Geom}(p)$

$$f(x;p) = (1-p)^{x-1} p \qquad x = 1, 2, \dots \qquad p \in (0,1)$$

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

$$\log \mathcal{L}(p) = n \log p + \left(\sum x_i - n\right) \log(1-p)$$

$$\frac{d}{dp} \log \mathcal{L}(p) = \frac{n}{p} - \frac{\sum x_i - n}{1-p} = 0$$

$$\Longrightarrow \quad \hat{p} = 1/\overline{X}$$

SOME OTHER CHECKS

Example

$X_1 \dots X_n \sim N(\mu, \sigma^2)$

$$\mathcal{L}(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

TWO PARAMETERS

$$\left.\begin{aligned} \frac{d}{d\mu} \log \mathcal{L}(\mu, \sigma) &= 0 \\ \frac{d}{d\sigma} \log \mathcal{L}(\mu, \sigma) &= 0 \end{aligned}\right\} \text{SOLVE FOR } \mu, \sigma$$

EXAMPLE

$X_1, \ldots, X_n \sim \text{UNIF}(0, \theta)$

$$f(x;\theta) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{o.w} \end{cases}$$

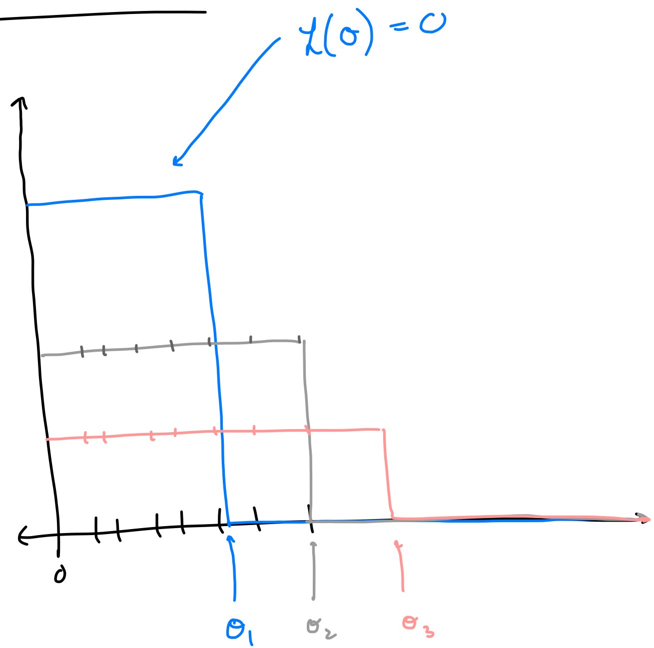
MAXIMUM IN DATA WHICH IS KNOWN

$$\mathcal{L}(\theta) = \begin{cases} (1/\theta)^n & \theta \ge X_{(n)} \\ 0 & \theta < X_{(n)} \end{cases}$$

AFTER $\theta = X_{(n)}$ $\mathcal{L}(\theta)$ DECREASES

SO $\hat{\theta} = X_{(n)}$

INTUITION



$$\mathcal{L}(\theta_2) > \mathcal{L}(\theta_3)$$

EQUIVARIANT

IF $\hat{\theta}_n$ IS MLE FOR θ

THEN $g(\hat{\theta}_n)$ IS MLE FOR $g(\theta)$

Example

$$X_1, \ldots, X_n \sim N(\theta, 1)$$

$\hat{\theta}_n = \overline{X}$ is MLE for θ

$e^{\hat{\theta}_n} = e^{\overline{X}}$ is MLE for e^{θ}

OPTIMIZATION

$$\frac{d}{d\theta} \log \mathcal{L}(\theta) = 0$$

MIGHT NOT HAVE ANALYTIC SOLUTION

- NEWTON RAPHSON
- EM ALGORITHM

$$X_1, X_2, X_3, \ldots\ldots X_n \sim F$$

How do we "infer" F?

LEARN

$$X_1 \ldots\ldots X_n \sim N(\mu, \sigma^2)$$

LEARNED DISTRIBUTION

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

PARAMETRIC BOOTSTRAP

EMPIRICAL DISTRIBUTION

NON-PARAMETRIC BOOTSTRAP

- USE $\hat{F}_n(x)$ AS ESTIMATE OF $F(x)$
- SAMPLE FROM $\hat{F}_n(x)$

PARAMETRIC BOOTSTRAP

- ASSUME PARAMETRIC FORM OF $F_\theta(x)$
- ESTIMATE PARAMETERS
- SAMPLE FROM LEARNED DISTRIBUTION, $F_{\hat{\theta}}(x)$

$$X_1^*, \ldots, X_n^* \sim N(\hat{\mu}, \hat{\sigma}^2)$$