CSE 312

Foundations of Computing II

Lecture 20: Two-parameter Estimation and Properties of Estimators

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Aleks Jovcic

Slide Credit: Based on Stefano Tessaro's slides for 312 19au incorporating ideas from Anna Karlin, Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

Agenda

- MLE Practice
- Two-parameter Estimation
- Properties of Estimators
  - Biased Estimators
  - Consistent Estimators

MLE for exponential distribution

$$\sum_{i=1}^{n} x_i = x_1 + x_2$$

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

Given n samples $x_1, \dots, x_n$ from an Exponential distribution with unknown parameter θ

The **likelihood** function of independent observations $x_1, \dots, x_n$ is

$$\mathcal{L}(x_1, \dots, x_n | \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i}$$

Find the MLE $\hat{\theta}$

$$\ln \prod_{i=1}^{n} \theta e^{-\theta x_i} = \sum_{i=1}^{n} \ln\left(\theta e^{-\theta x_i}\right)$$

$$= \sum_{i=1}^{n} \ln(\theta) - \theta x_i$$

$$= n\ln(\theta) - \sum_{i=1}^{n} \theta x_i$$

$$\log(ab) = \log(a) + \log(b)$$
$$\log(a/b) = \log(a) - \log(b)$$
$$\log\left(a^b\right) = b\log(a)$$

$$\frac{\partial}{\partial \theta} \ln L(x_1, \ldots x_n \mid \theta)$$

$$= \frac{n}{\hat{\theta}} - \sum_{i=1}^{n} x_i = 0$$

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

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

General Recipe

1. **Input** Given n iid samples $x_1, \ldots, x_n$ from parametric model with parameters θ.
2. **Likelihood** Define your likelihood $\mathcal{L}(x_1, \ldots., x_n | \theta)$.
   - For discrete $\mathcal{L}(x_1, \ldots., x_n | \theta) = \prod_{i=1}^{n} \Pr(x_i\,; \theta)$
   - For continuous $\mathcal{L}(x_1, \ldots., x_n | \theta) = \prod_{i=1}^{n} f(x_i\,; \theta)$
3. **Log** Compute $\ln \mathcal{L}(x_1, \ldots., x_n | \theta)$
4. **Differentiate** Compute $\frac{\partial}{\partial \theta} \ln \mathcal{L}(x_1, \ldots., x_n | \theta)$
5. **Solve for** $\hat{\theta}$ by setting derivative to 0 and solving for max.

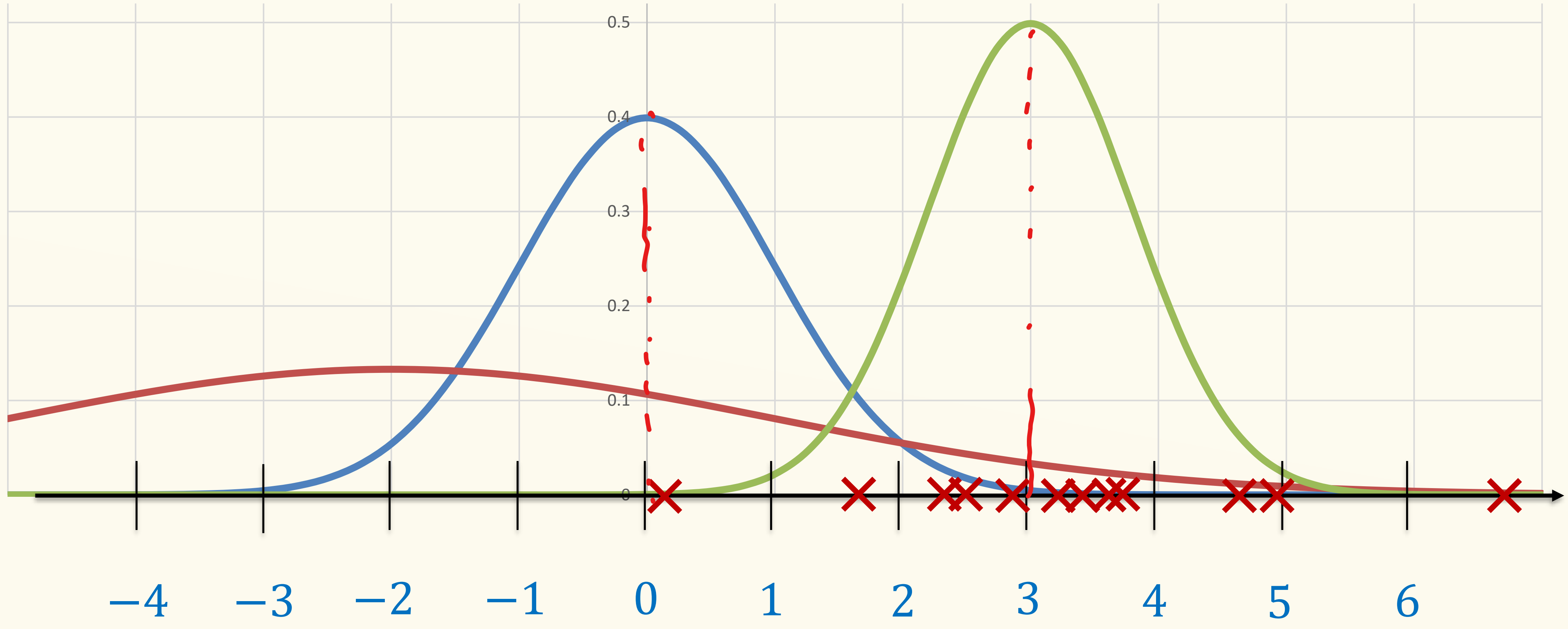
Generally, you need to do a second derivative test to verify it is a maximum, but we won't ask you to do that in CSE 312.

Agenda

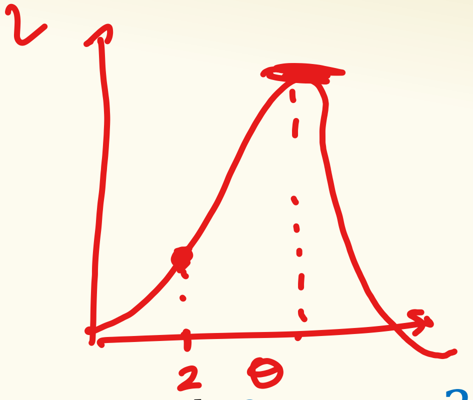
- MLE Practice
- Two-parameter Estimation
- Properties of Estimators
  - Biased Estimators
  - Consistent Estimators

Next: n samples $x_1, \dots, x_n \in \mathbb{R}$ from Gaussian $\mathcal{N}(\mu, \sigma^2)$.

Most likely μ and σ^2?



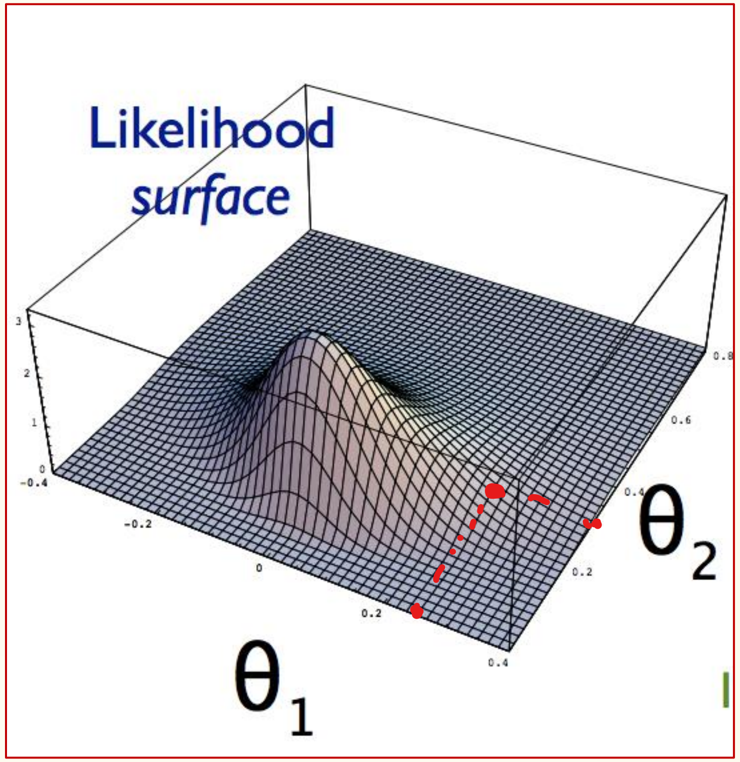
Two-parameter optimization



$$\log(ab) = \log(a) + \log(b)$$
$$\log(a/b) = \log(a) - \log(b)$$
$$\log(a^b) = b\log(a)$$

Normal outcomes $x_1, \ldots, x_n$

Goal: estimate $\theta_1 = \mu$ = expectation and $\theta_2 = \sigma^2$ = variance



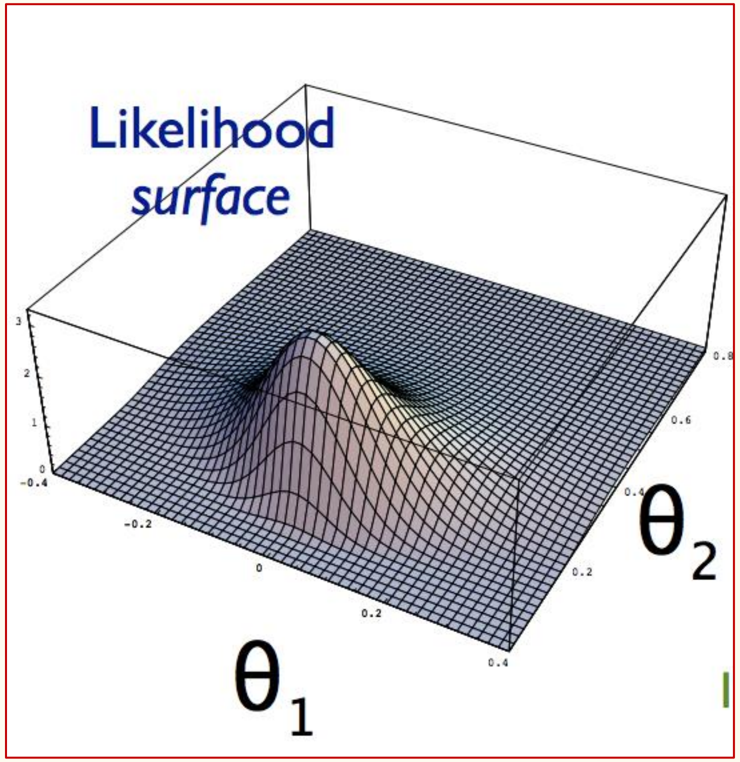
$$L(x_1, \ldots, x_n | \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}}$$

$$\ln L(x_1, \ldots, x_n | \theta_1, \theta_2) =$$

Two-parameter optimization

Normal outcomes $x_1, \dots, x_n$

Goal: estimate $\theta_1 = \mu$ = expectation and $\theta_2 = \sigma^2$ = variance



$$L(x_1, \dots, x_n | \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}}$$

$$\ln L(x_1, \dots, x_n | \theta_1, \theta_2) =$$

$$= -n\frac{\ln(2\pi\,\theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2}$$

Two-parameter estimation

$$\ln L(x_1, \dots, x_n | \theta_1, \theta_2) = -n \frac{\ln(2\pi\,\theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2}$$

We need to find a solution $\hat{\theta}_1, \hat{\theta}_2$ to

$$\begin{cases} \dfrac{\partial}{\partial \theta_1} \ln L(x_1, \dots, x_n | \theta_1, \theta_2) = 0 \\ \dfrac{\partial}{\partial \theta_2} \ln L(x_1, \dots, x_n | \theta_1, \theta_2) = 0 \end{cases}$$

MLE for Expectation

$$\ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = -n \frac{\ln(2\pi\,\theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial}{\partial \theta_1} \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) =$$

MLE for Expectation

$$\ln L(x_1, \dots, x_n | \theta_1, \theta_2) = -n \frac{\ln(2\pi \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial}{\partial \theta_1} \ln L(x_1, \dots, x_n | \theta_1, \theta_2) = \frac{1}{\theta_2} \sum_{i}^{n} (x_i - \theta_1) = 0$$

$$\hat{\theta}_1 = \frac{\sum_i^n x_i}{n}$$

In other words, MLE of expectation is the *sample mean* of the data, regardless of θ_2

What about the variance?

MLE for Variance

$$\ln L(x_1, \ldots, x_n | \hat{\theta}_1, \theta_2) = -n \frac{\ln(2\pi\, \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \hat{\theta}_1)^2}{2\theta_2}$$

$$= -n \frac{\ln 2\pi}{2} - n \frac{\ln \theta_2}{2} - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2$$

$$\frac{\partial}{\partial \theta_2} \ln L(x_1, \ldots, x_n | \theta_1, \hat{\theta}_1) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2 = 0$$

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2$$

In other words, MLE of variance is what's called the *population variance* of the data set.

Likelihood – Continuous Case

Definition. The **likelihood** of independent observations $x_1, \ldots, x_n$ is

$$L(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

Normal outcomes $x_1, \ldots, x_n$

$$\hat{\theta}_\mu = \frac{\sum_i^n x_i}{n}$$

MLE estimator for **expectation**

$$\hat{\theta}_{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_\mu)^2$$

MLE estimator for **variance**

Agenda

$\frac{1}{b-a} = \frac{1}{\theta - 0} = \frac{1}{\theta}$

3, 7, 12, 2, 9

- MLE Practice
- Two-parameter Estimation
- **Properties of Estimators**
  - Biased Estimators
  - Consistent Estimators

MLE of the uniform $(0, \theta)$

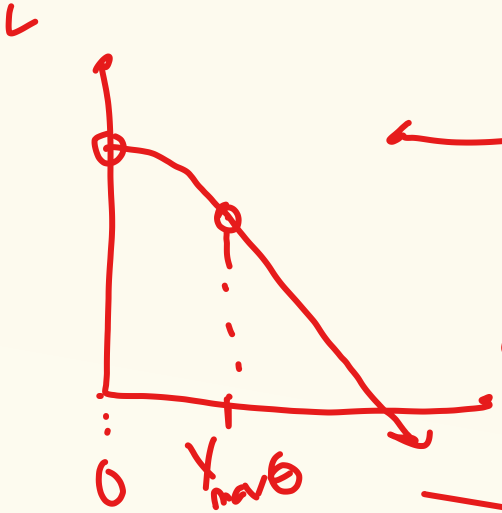
$L(X_1 \dots X_n \mid \theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n} = \theta^{-n}$

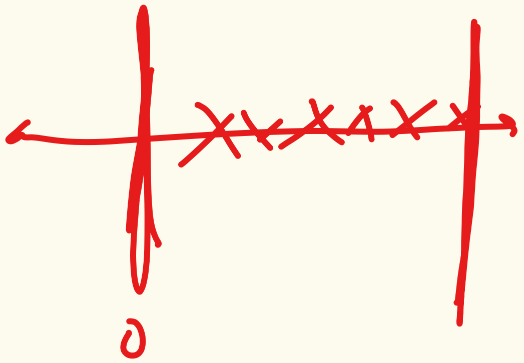
$\ln L(X_1 \dots X_n \mid \theta) = -n \ln(\theta)$

~~$\frac{\partial}{\partial \theta} \ln L(X_1 \dots X_n \mid \theta) = \frac{-n}{\theta}$~~

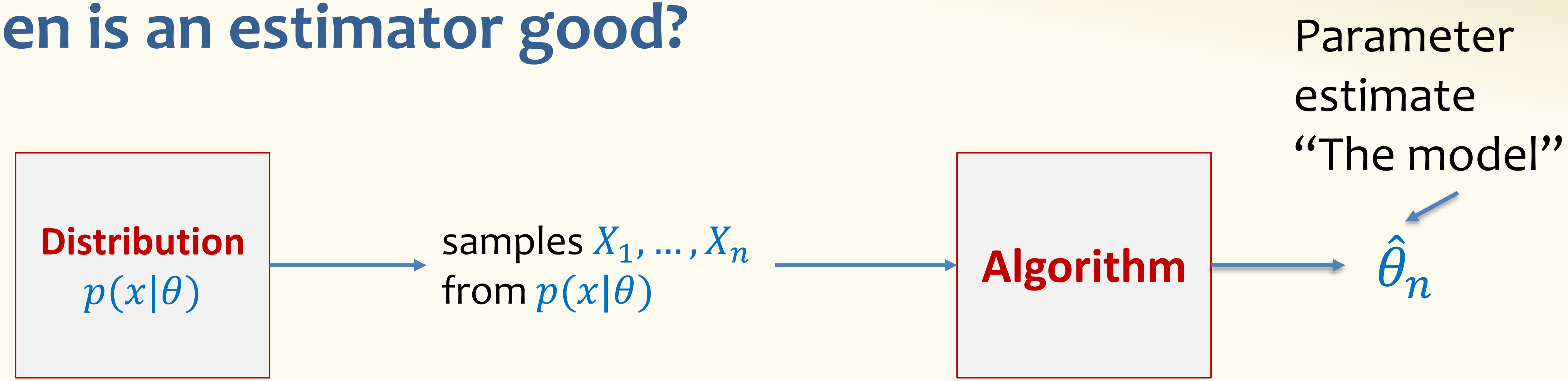
MLE for θ:

$\hat{\theta} = \max(X)$





When is an estimator good?



θ = unknown parameter

Definition. An estimator of parameter θ is an **unbiased estimator**

$$\mathbb{E}(\hat{\theta}_n) = \theta.$$

Example – Coin Flips

Recall: $\hat{\theta} = \frac{n_H}{n}$

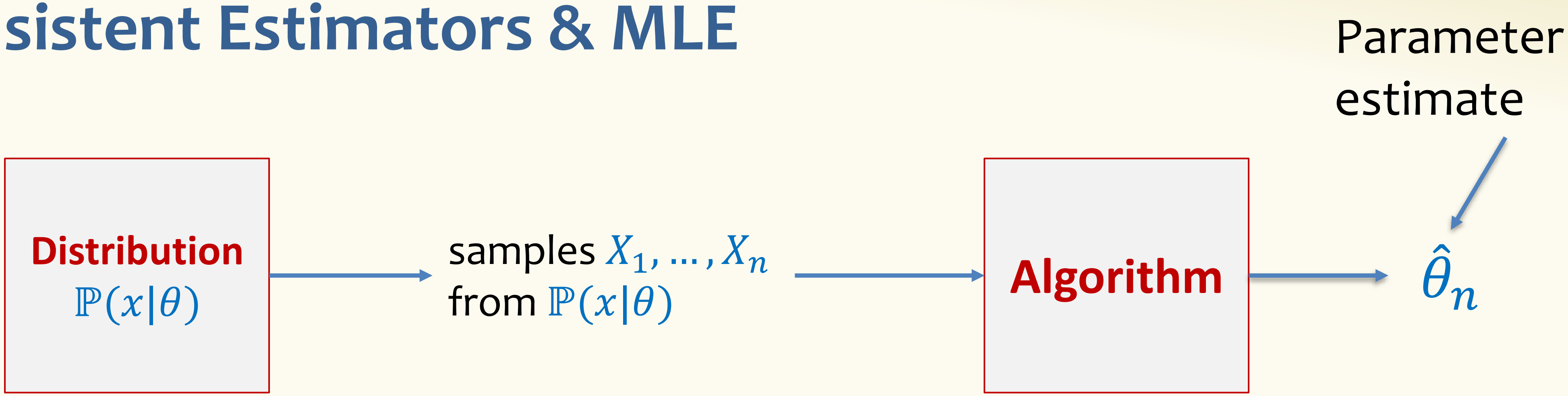
Coin-flip outcomes $x_1, \dots, x_n$, with n_H heads, n_T tails

Fact. $\hat{\theta}$ is unbiased

i.e., $\mathbb{E}(\hat{\theta}) = p$, where p is the probability that the coin turns out heads.

$$E[\hat{\theta}] = E\left[\frac{n_H}{n}\right] = E\left[\frac{1}{n} \cdot n_H\right] = \frac{1}{n} E[n_H] = \frac{1}{n} np = \boxed{p}$$

Consistent Estimators & MLE



θ = unknown parameter

Definition. An estimator is **unbiased** if $\mathbb{E}(\hat{\theta}_n) = \theta$ for all $n \geq 1$.

Definition. An estimator is **consistent** if $\lim_{n \to \infty} \mathbb{E}(\hat{\theta}_n) = \theta$.

Theorem. MLE estimators are consistent.

(But not necessarily unbiased)

Example – Consistency

Normal outcomes $X_1, \dots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$ Assume: $\sigma^2 > 0$

$$\widehat{\Theta}_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_{\mu}\right)^2$$

MLE – Biased!

$\widehat{\Theta}_{\sigma^2}$ converges to σ^2, as $n \to \infty$.

$\widehat{\Theta}_{\sigma^2}$ is “consistent”

Why is the estimator consistent, but biased?

linearity

$$\mathbb{E}(\widehat{\Theta}_{\sigma^2}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \widehat{\Theta}_{\mu}\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n}X_j + \frac{1}{n^2}\sum_{j=1}^{n}X_j\sum_{k=1}^{n}X_k\right]$$

…

Why is the estimator consistent, but biased?

$E[\hat{\Theta}_{\sigma^2}] = \sigma^2$

linearity

$$\mathbb{E}(\widehat{\Theta}_{\sigma^2}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \widehat{\Theta}_1\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right]$$

…

$$= \left(1 - \frac{1}{n}\right)\sigma^2 = \frac{n-1}{n}\sigma^2$$

Why is the estimator consistent, but biased?

linearity

$$\mathbb{E}(\widehat{\Theta}_{\sigma^2}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \widehat{\Theta}_1\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right]$$

…

$$= \left(1 - \frac{1}{n}\right)\sigma^2 = \frac{n-1}{n}\sigma^2 \rightarrow \sigma^2 \text{ for } n \rightarrow \infty$$

Therefore: $\frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \widehat{\Theta}_1\right)^2\right] = \frac{n}{n-1}\mathbb{E}(\widehat{\Theta}_{\sigma^2}) = \sigma^2$

Bessel's correction

Example – Consistency

$E[\hat{\theta}] = \theta$ $E[\hat{\theta}] \neq \theta$

Normal outcomes $X_1, \ldots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$ Assume: $\sigma^2 > 0$

$$\widehat{\Theta}_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_{\mu}\right)^2$$

MLE – Biased!

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \widehat{\Theta}_{\mu}\right)^2$$

Sample variance – Unbiased!

$\widehat{\Theta}_{\sigma^2}$ converges to σ^2, as $n \to \infty$.

$\widehat{\Theta}_{\sigma^2}$ is "consistent"