1. Mid-term 2: March 28.

2. Arrangements: This + Next lecture: new material.
Third lecture: Mid-term 2 prepare lecture,

3. Post Practice exam this Thur.

4. Post new assignment this Thur,
as long as you submit,
you'll get full credits.

Theorem [Central limit Theorem].

- Have a population distribution P, with (μ, σ^2).
- Have n i.i.d samples $X_1, \cdots, X_n \sim P$.
- $\bar{X} = \frac{1}{n} \sum_i X_i$

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \to N(0, 1).$$

$\bar{X}$ is approximately $N(\mu, \frac{\sigma^2}{n})$.

Extension Theorem.

- Two group of populations, one with (μ_1, σ_1^2)
  another with (μ_2, σ_2^2).

- $X_1, \dots, X_{n_1} \overset{i.i.d.}{\sim} (\mu_1, \sigma_1^2)$. $\quad Y_1, \dots, Y_{n_2} \overset{i.i.d.}{\sim} (\mu_2, \sigma_2^2)$.

- $\bar{X} - \bar{Y}$ is also approximately $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$.

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \to N(0, 1)$$

ISyE 3770, Spring 2024
Statistics and Applications

Point Estimation

Instructor: Jie Wang
H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Tech

jwang3163@gatech.edu
Office: ISyE Main 447

Outline

- **Estimator: Definition**
- **Basic properties**
- **Methods for finding point estimators**

population distribution P

$\downarrow$

$\underbrace{X_1, \cdots, X_n}$ $\overset{i.i.d.}{\sim} P$

$\downarrow$

$h(X_1, \cdots, X_n)$: Statistics

$\downarrow$

sampling distribution

Questions we aim to address

- **What is a good estimator?**
- **How to find estimators?**

Data → **Statistics** ← **Sampling distribution**

Statistics → **Estimator**

Sampling distribution → **Estimator**

Estimator

Suppose X is a random variable with $f(x;\theta)$ as the pdf. If $X_1, X_2, \ldots X_n$ is a random sample of size n from X, the statistic

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

Is called a **point estimator** of θ.

• I don't know value of θ in practice

function of random sample

• $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X$

After the sample has been selected, $\hat{\Theta}$ takes on a particular numerical value called the **point estimate** of θ.

Parameter: μ **Estimator:** $\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ **Estimate:** $\bar{x} = \dfrac{25+30+29+31}{4} = 28.75$

Note that $\hat{\Theta}$ is a random variable because it is a statistic (function of random variables)

Internet service provider

A. population
B. Sample
C. Parameter
D. Statistic

- **Two Internet providers**
- **Observe download rate is as follows (mbp)**

Provider 1	5.34	5.16	5.043	4.661	4.521	5.25	5.245
Provider 2	5.363	4.797	5.28	4.666	4.927	5.286	5.37
Provider 1	5.276	4.508	4.558	5.478	4.919	4.708	
Provider 2	5.109	5.113	5.157	5.145	4.801	4.948	

- What's the difference of their rate?

- What's the difference of their rate?
- Samples
  - First service provider $X_i, i = 1, 2, \ldots, n_1$ $n_1 = 13$.
  - Second service provider $Y_i, i = 1, 2, \ldots, n_2$ $n_2 = 13$
- Assumption
  - $X_i \sim N(\mu_1, \sigma_1^2)$
  - $Y_i \sim N(\mu_2, \sigma_2^2)$

{ Point estimator
  Point estimate.

- **Parameters of interest:** $\mu_1 - \mu_2$
- **Estimator:** $\bar{X} - \bar{Y}$
- **Estimate: 4.9744- 5.0740 = -0.0996 (mbp)**: point estimate

- **How accurate is the estimate?**
- **Is the estimator (method) unbiased?**

Basic properties of estimators

Standard error of estimator

→ random variable

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \boxed{\sqrt{V(\hat{\Theta})}}$. If the standard error involves unknown parameters that can be esti- **ties to sampling distribution** mated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error,** denoted by $\hat{\sigma}_{\hat{\Theta}}$.

$h(X_1, \ldots, X_n)$

Data → **Statistics** ← **Sampling distribution**

Estimator $\hat{\theta}$

point estimate

Internet service provider

- **Two Internet providers**

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

- **Observe download rate is as follows (mbp)**

Provider 1	5.34	5.16	5.043	4.661	4.521	5.25	5.245
Provider 2	5.363	4.797	5.28	4.666	4.927	5.286	5.37
Provider 1	5.276	4.508	4.558	5.478	4.919	4.708	
Provider 2	5.109	5.113	5.157	5.145	4.801	4.948	

Estimated

- What's the ~~standard~~ error of the estimator for the difference of their rate?

- What's the difference of their rate?
- Samples
  - First service provider $X_i, i = 1, 2, \dots, n_1$
  - Second service provider $Y_i, i = 1, 2, \dots, n_2$
- Assumptions
  - $X_i \sim N(\mu_1, \sigma_1^2)$
  - $Y_i \sim N(\mu_2, \sigma_2^2)$
- **Parameters of interest: $\mu_1 - \mu_2$**
- **Estimator: $\bar{X} - \bar{Y}$**

$$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{Var(X_i)}{n_1} + \frac{Var(Y_i)}{n_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- **Standard error of the estimator**

Estimated error should be used! $\Rightarrow$ 1. replace σ_1^2 with S_1^2. 2. replace σ_2^2 with S_2^2.

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{Var(\bar{X} - \bar{Y})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

S_1^2, S_2^2 are sample variance for provider 1 and 2

Exercise

An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

• Estimated error:

41.60, 41.48, 42.34, 41.95, 41.86,
42.18, 41.72, 42.26, 41.81, 42.04

$$\hat{\sigma}_{\bar{X}} = \sqrt{\frac{S^2}{10}} = 0.089$$

• $X_1, \cdots, X_n \underset{\sim}{i.i.d.} N(\mu, \sigma^2)$

• point estimator: $\bar{X} = \frac{1}{n}\sum_i X_i$

What is the estimator for the conductivity?

• point estimate: $\hat{\mu} = \frac{41.60 + 41.48 + \cdots + 42.04}{10} = 41.924$

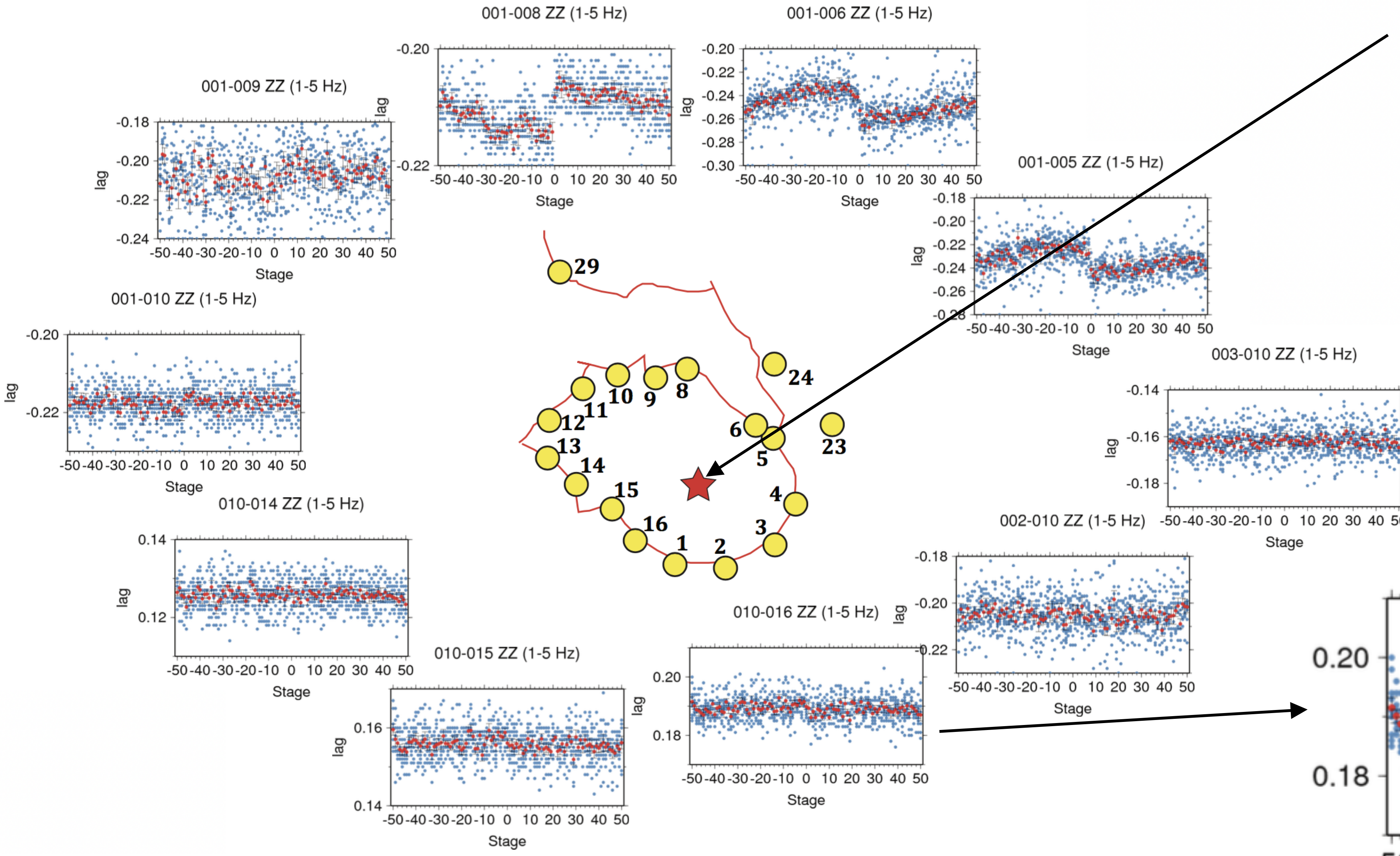
What is the standard error of the estimator?

• Standard error: $\sigma_{\bar{X}} = \sqrt{Var(\bar{X})} = \sqrt{Var(\frac{1}{n}\sum_i X_i)} = \sqrt{\frac{Var(X_i)}{n}} = \sqrt{\frac{\sigma^2}{10}}$

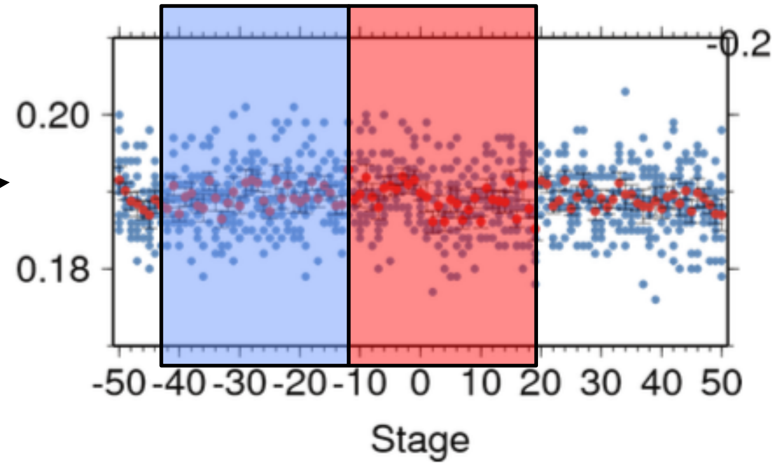
A real-world example

- **Detecting changes using sliding windows, sample mean difference**

Old faithful geyser



Sliding windows



Unbiased Estimator

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$h(X_1, \ldots, X_n)$.

$$\boxed{E(\hat{\Theta})} = \theta$$

ties to sampling distribution

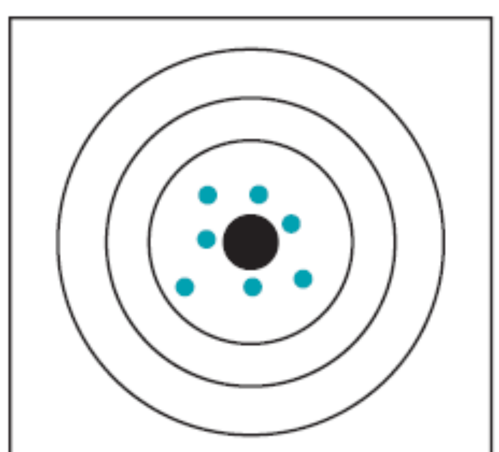
If the estimator is not unbiased, then the difference

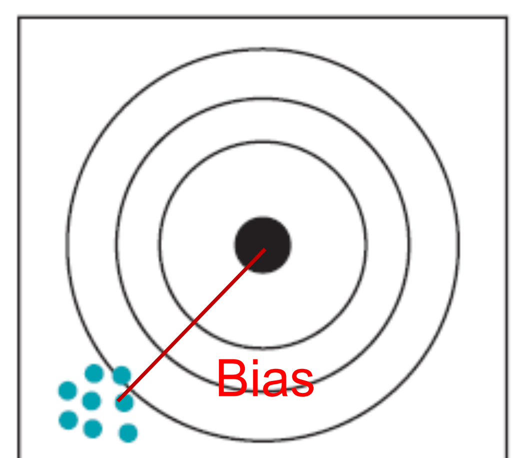
$$E(\hat{\Theta}) - \theta$$

is called the **bias** of the estimator $\hat{\Theta}$.

if $E[\hat{\theta}] - \theta = 0$, say $\hat{\theta}$ is unbiased.

if $E[\hat{\theta}] - \theta \neq 0$, say $\hat{\theta}$ is biased, bias $= E[\hat{\theta}] - \theta$.





Sample mean is unbiased estimator

- Assume $x_1, \ldots, x_n \sim N(\mu, \sigma^2)$
- Then $\bar{x}$ is an unbiased estimator of μ

point estimator : $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

proof: To verify $E[\bar{X}] = \mu$.

- $E[\bar{X}] = E[\frac{1}{n} \sum_i X_i] = \frac{1}{n} \sum_i E[X_i]$.

$= \frac{1}{n} \sum_i \mu = \mu$.

- $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow E[\bar{X}] = \mu$

Sample variance is unbiased estimator

- Assume $x_1, \dots, x_n \sim N(\mu, \sigma^2)$
- Then S^2 is an unbiased estimator of σ^2

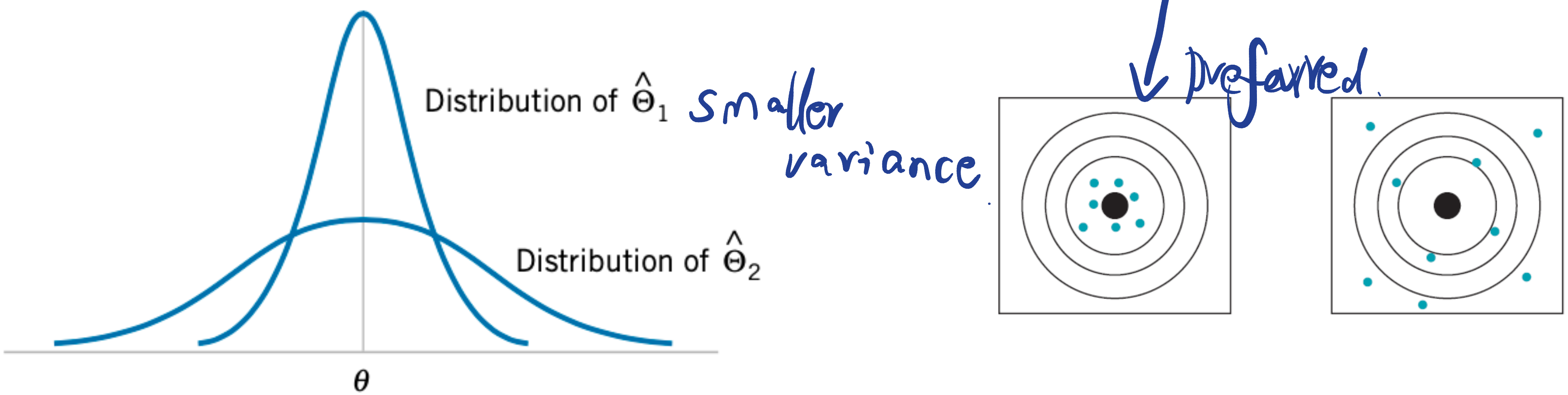
- point estimator $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$,

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

- $E[S^2] = \sigma^2$.

Variance of a Point Estimator

If two estimators are unbiased, the one with smaller variance is preferred.



smaller variance

preferred

Figure 7-1 **The sampling distributions of two unbiased estimators**

$$\boxed{\mathrm{var}(\hat{\Theta}_1)} < \mathrm{var}(\hat{\Theta}_2)$$

ties to sampling distribution

$\hat{\theta}_1$, $\hat{\theta}_2$.

if $E[\hat{\theta}_1] = E[\hat{\theta}_2] = \theta$

prefer $\hat{\theta}_1$ if

$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

1. Mid-term 2: March 28.

practice exam and Solu:

2. HW 5. As long as submit, receive full credit.

3. Upload solution to exercise of slides.

Mean Square Error (MSE)

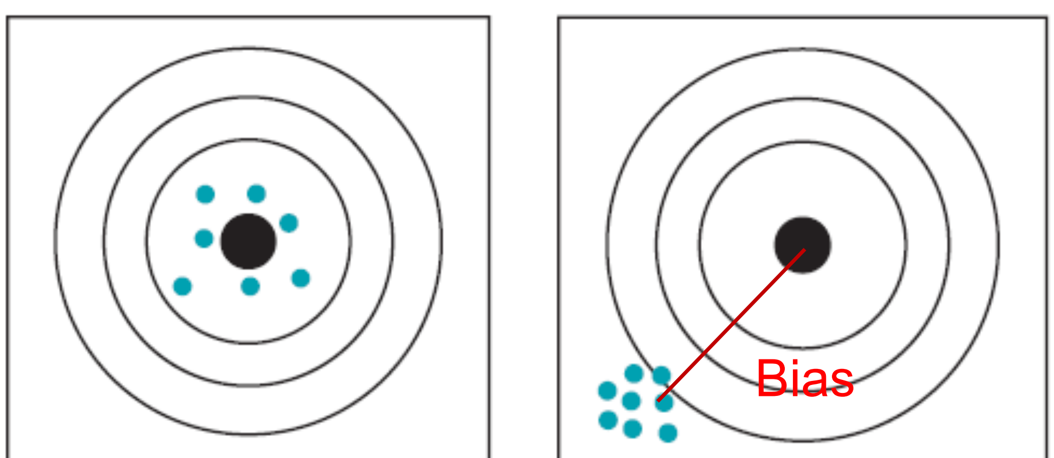
The **mean square error** of an estimator $\hat{\Theta}$ of the parameter θ is defined as

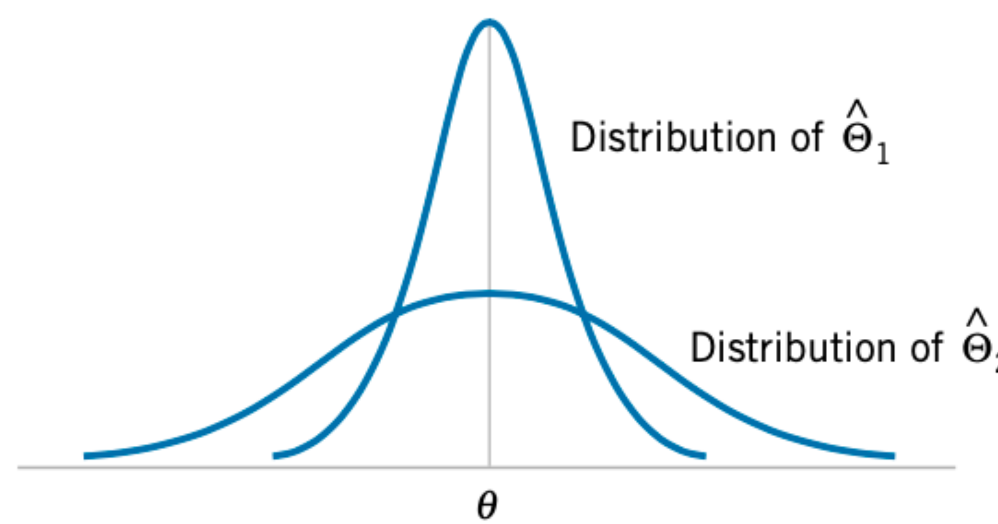
$$\text{MSE}(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2 \qquad (7\text{-}3)$$

according to sampling distribution of $\hat{\theta}$

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2 = \left[E(\hat{\Theta} - \Theta)\right]^2 + \text{var}(\hat{\Theta} - \Theta)$$

$$MSE(\hat{\Theta}) = \left[\text{Bias}(\hat{\Theta})\right]^2 + \text{var}(\hat{\Theta})$$





Example: find bias and variance of estimator

Let X_1, X_2 be independent random variables with mean μ and variance σ^2. Suppose that we have two estimators of μ:

$MSE(\hat{\theta}_1) = MSE(\hat{\theta}_2)$

$$\hat{\Theta}_1 = \frac{X_1 + X_2}{2}$$

$$\hat{\Theta}_2 = \frac{X_1 + 3X_2}{4}$$

$(bias(\hat{\theta}))^2 + Var(\hat{\theta})$

(a) To verify unbiasedness, to show

$E[\hat{\theta}_1] = \mu$?

$E[\hat{\theta}_2] = \mu$?

$E[\hat{\theta}_1] = E[\frac{X_1+X_2}{2}]$

$= \frac{1}{2}E[X_1] + \frac{1}{2}E[X_2] = \mu$

(a) Are both estimators unbiased estimators of μ?

There is not a unique unbiased estimator!

(b) What is the variance of each estimator?

(b). $Var(\hat{\theta}_1) = Var(\frac{X_1+X_2}{2})$

$= Var(\frac{X_1}{2}) + Var(\frac{X_2}{2})$

$= \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2)$

$= \frac{1}{2}\sigma^2$

(c) What's the MSE of two estimators?

$Var(\hat{\theta}_2) = Var(\frac{1}{4}X_1 + \frac{3}{4}X_2) = Var(\frac{1}{4}X_1) + Var(\frac{3}{4}X_2) = \frac{10}{16}\sigma^2$

Compare the MSE of estimators

Let $X_1, X_2, \ldots X_7$ denote a random sample from a population with mean μ and variance σ^2. Calculate the MSE of the following estimators of μ.

$(\text{bias}(\hat{\theta}))^2 + \text{Var}(\hat{\theta})$

$$\hat{\Theta}_1 = \frac{\sum_{i=1}^{7} X_i}{7}$$

$$\hat{\Theta}_2 = \frac{2X_1 - X_6 + X_4}{2}$$

$$\hat{\Theta}_3 = \frac{4X_2 + 2X_3 - 2X_5}{2} \quad X$$

$$\begin{cases} \text{Var}(\hat{\Theta}_1) = \frac{\sigma^2}{7} \\ \text{Var}(\hat{\Theta}_2) = \text{Var}(X_1) + \text{Var}(-\frac{1}{2}X_6) + \text{Var}(\frac{1}{2}X_4) \end{cases}$$

$= \frac{3}{2}\sigma^2$

- **Is either estimator unbiased?**
- **Which estimator is best? In what sense is it best?**

$$E[\hat{\Theta}_3] = \frac{4}{2}E[X_2] + \frac{2}{2}E[X_3] - \frac{2}{2}E[X_5]$$

$$= 2\mu$$

Example

Suppose $X \sim Uniform(\theta, 3\theta)$, $\theta > 0$

$X_1, \dots, X_n \overset{i.i.d.}{\sim} X$

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

- Show that $\frac{\bar{X}}{2}$ is an unbiased estimator of θ
- Calculate the MSE of $\frac{\bar{X}}{2}$ and $\bar{X}$

$MSE(\bar{X}) = \underbrace{(bias(\bar{X}))^2}_{(E[\bar{X}]-\theta)^2 \to (2\theta-\theta)^2 \to \theta^2} + \underbrace{Var(\bar{X})}_{\frac{\sigma^2}{n} \to \frac{1}{n}\cdot\frac{1}{12}(3\theta-\theta)^2 \to \frac{\theta^2}{3n}}$

Suppose

$X \sim Uni(a, b)$

$E[X] = \frac{1}{2}(a+b)$

$Var(X) = \frac{1}{12}(b-a)^2$

- $MSE\left(\frac{\bar{X}}{2}\right) = \cancel{\left(bias\left(\frac{\bar{X}}{2}\right)\right)^2} + Var\left(\frac{\bar{X}}{2}\right)$

$= \frac{1}{4}Var(\bar{X}) = \frac{1}{4n}Var(X)$

$= \frac{1}{4n} \cdot \frac{1}{12} \cdot (3\theta - \theta)^2$

$= \frac{\theta^2}{12n}$.

$MSE(\bar{X})$

$= \left(1 + \frac{1}{3n}\right)\theta^2$

$X_1, \dots, X_n \overset{i.i.d.}{\sim} X$

$Var(X) = \sigma^2$

$\Rightarrow Var(\bar{X}) = \frac{\sigma^2}{n}$.

1. Two internet providers.

2. Ground truth difference of the mean download rate.

parameter. θ

3. All data collected from users

population

4. Observed data collected from users, sample size n

random sample

5. Constructed estimator for the unknown

point estimator difference of mean download rate.

statistics. $\hat{\theta}$

bias
variance
MSE

Methods for Finding Estimators

- **Assume a distribution for the samples**
- **Estimate the parameter of the distribution**
- **Several methods**
  - **Maximum likelihood**
  - **Method of moment**

Baseball team

- The weight for a baseball team players are

  {150, 143, 132, 160, 175, 190, 123, 154}

- Assume their weights are uniformly distributed over an interval [a, b]

$a = \min_i X_i \qquad b = \max_i X_i$

- What are good estimators for a? for b?

Method of Maximum Likelihood

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

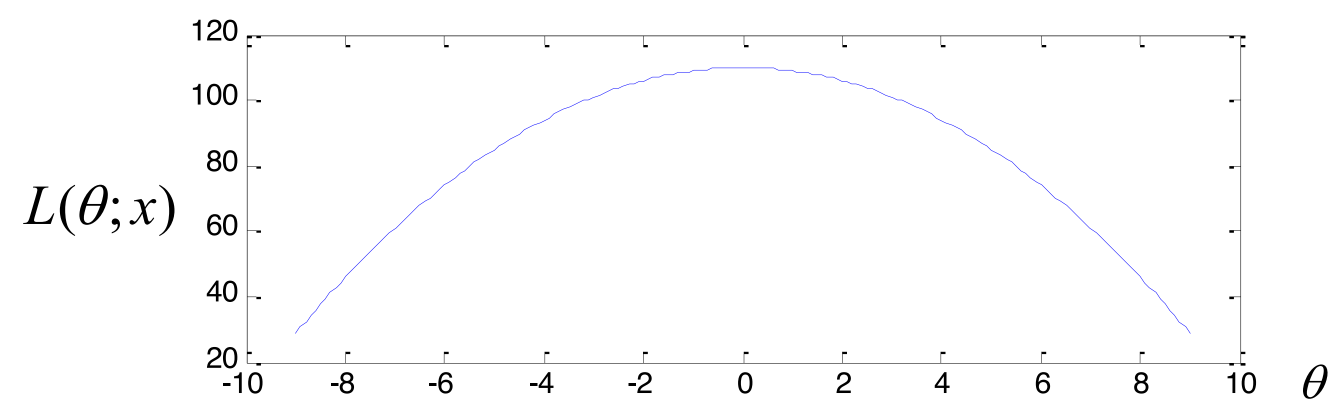
$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \tag{7-5}$$

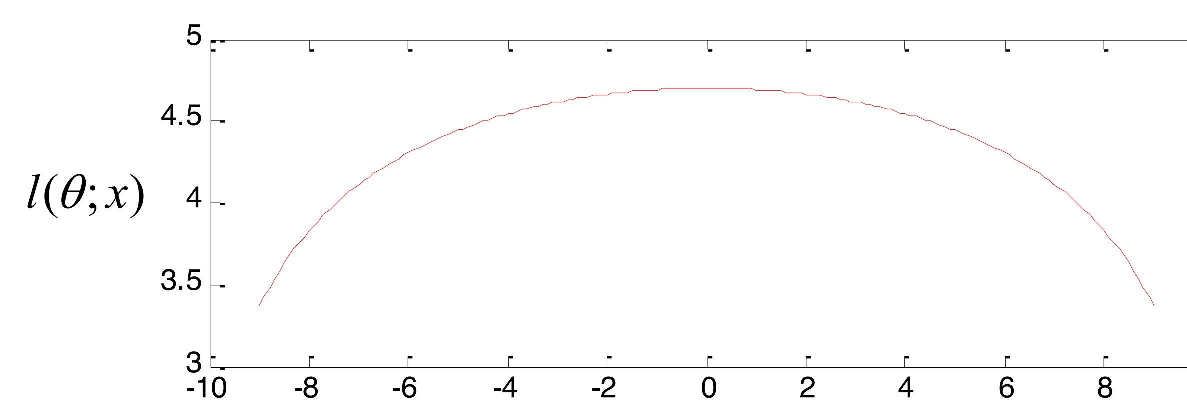
Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

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

$$\boxed{l(\theta;x) = \sum_{i=1}^{n} \log[f(x_i;\theta)]}$$

$$\hat{\Theta}(x) = \arg\max_{\theta} L(\theta;x) = \arg\max_{\theta} l(\theta;x)$$





7-61. A random variable x has probability density function

$$f(x;\theta) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, \qquad 0 < x < \infty, \quad 0 < \theta < \infty$$

Given samples $x_1, \ldots x_n$,
find the maximum likelihood estimator for θ

Step 1. Write down and simplify log-likelihood function

$$\ell(\theta; x) = \sum_{i=1}^{n} \log(f(x_i;\theta)) = \sum_{i=1}^{n} \log\left(\frac{1}{2\theta^3} x_i^2 \exp\left(-\frac{x_i}{\theta}\right)\right)$$

$$= \sum_{i=1}^{n} \left[\cancel{-\log 2} - 3\log\theta + \cancel{2\log x_i} - \frac{x_i}{\theta}\right] + \text{constant}$$

$$= \cancel{-n\log 2} - 3n\log\theta + \cancel{2\sum_{i=1}^{n}\log x_i} - \frac{\sum_{i=1}^{n} x_i}{\theta}$$

Step 2. Compute maximizer $\hat{\theta}(x)$

$$\frac{\partial}{\partial\theta}\ell(\theta; x) = 0$$

$$\Downarrow$$

$$\frac{-3n}{\hat{\theta}(x)} + \frac{\sum_i x_i}{\hat{\theta}(x)^2} = 0$$

$$\boxed{\hat{\theta}(x) = \frac{\sum x_i}{3n} = \frac{1}{3}\bar{X}}$$

Example: Bernoulli

Let X be a Bernoulli random variable. The probability mass function is

$$f(x; p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

• Step 1: write down and simplify likelihood function

$$L(p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

• Step 2: find maximizer of $L(p)$, i.e., maximizer of $\log(L(p))$

$$\longrightarrow \quad \ln L(p) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

$$\longrightarrow \quad \frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0 \longrightarrow \hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Example: normal

Step1:

Let X be normally distributed with unknown μ and known variance σ^2. The likelihood function of a random sample of size n, say $X_1, X_2, \ldots, X_n$, is

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

Now

$$\ln L(\mu) = -(n/2)\ln(2\pi\sigma^2) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\mu = \frac{1}{n}\sum_i X_i$$

and Step2: Find estimator to maximize $\log(L(\mu))$

$$\Rightarrow \quad = \bar{X}$$

$$\frac{d\ln L(\mu)}{d\mu} = \underbrace{(\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)}_{} = 0 = \frac{1}{\sigma^2}\left(\sum_i X_i - n\mu\right)$$

$$\frac{d}{d\mu}\log(L(\mu)) = (-2\sigma^2)^{-1}\sum_{i=1}^{n} 2(X_i-\mu)(-1)$$

$\longrightarrow$ What is the MLE for μ?

Example (Continued, unknown variance)

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$\Rightarrow \mu = \frac{1}{n}\sum_i x_i$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$\frac{1}{2\sigma^4}\sum(x_i - \mu)^2 = \frac{n}{2\sigma^2}$

$\sum(x_i - \mu)^2 = n\sigma^2$

$\Rightarrow \sigma^2 = \frac{1}{n}\sum(x_i - \mu)^2$

$= \frac{1}{n}\sum(x_i - \bar{x})^2$

The solutions to the above equation yield the maximum likelihood estimators

$$\hat{\mu} = \bar{X} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

MLE: Exponential

Let X be a exponential random variable with parameter λ. The likelihood function of a random sample of size n is:

$f(x;\lambda) = \lambda \cdot e^{-\lambda x}$

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

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

Step 1: simplify log-likelihood function

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\hat{\lambda} = n \Big/ \sum_{i=1}^{n} x_i = 1/\bar{X} \quad \text{(same as moment estimator)}$$

step 2: Compute maximizer of $\log(L(\lambda))$

MLE: Graphical Illustration

The time to failure is exponentially distributed. Eight units are randomly selected and tested, resulting in the following failure time (in hours): $x_1 = 11.96$, $x_2 = 5.03$, $x_3 = 67.40$, $x_4 = 16.07$, $x_5 = 31.50$, $x_6 = 7.73$, $x_7 = 11.10$, and $x_8 = 22.38$.

"peaked" vs **"flat"**

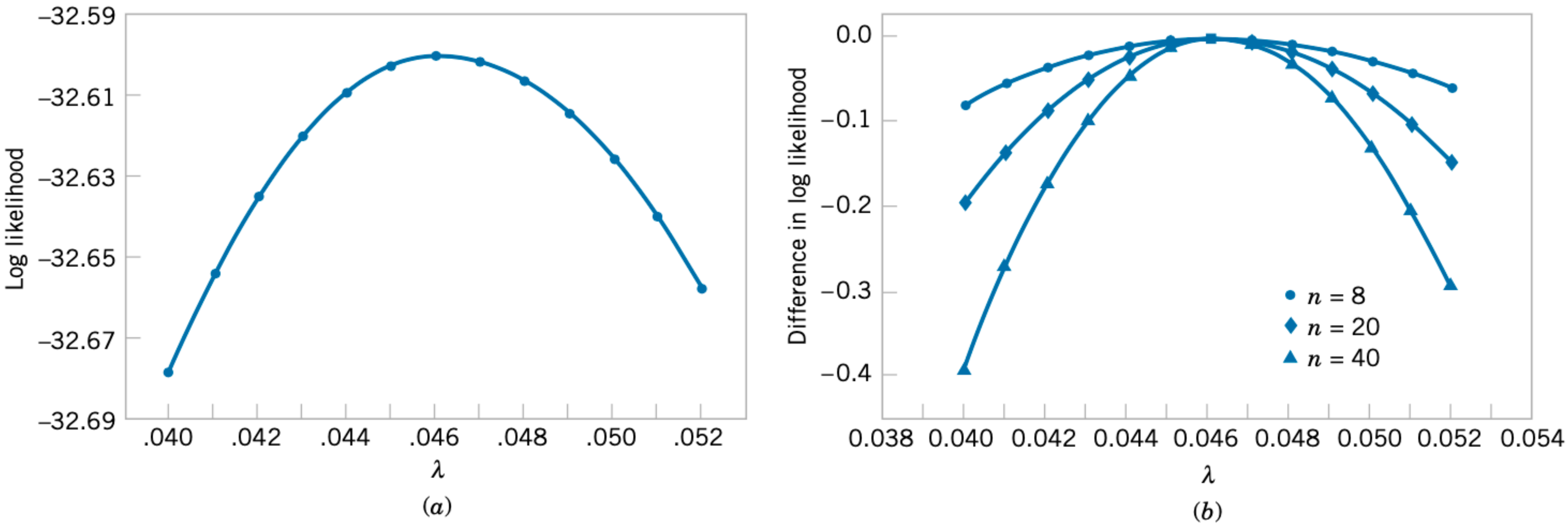


Figure 7-3 Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Log likelihood if $n = 8$, 20, and 40.

Why use maximum likelihood estimator?

It enjoys the following good properties:

Properties of the Maximum Likelihood Estimator

Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\boldsymbol{\Theta}}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\boldsymbol{\Theta}}$ is an approximately unbiased estimator for θ $[E(\hat{\boldsymbol{\Theta}}) \simeq \theta]$,

(2) the variance of $\hat{\boldsymbol{\Theta}}$ is nearly as small as the variance that could be obtained with any other estimator, and

(3) $\hat{\boldsymbol{\Theta}}$ has an approximate normal distribution.

maximum likelihood estimator in some cases is biased!

Complications in Using MLE

- It is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\Theta)/d\Theta = 0$ may be difficult to solve.
- It may not always be possible to use calculus methods directly to determine the maximum of $L(\Theta)$.

Baseball team

- The weight for a baseball team players are

  {150, 143, 132, 160, 175, 190, 123, 154}

- Assume their weights are uniformly distributed over an interval [a, b]

- What are good estimators for a? for b?

Example: Uniform Distribution MLE

Let X be uniformly distributed on the interval 0 to a.

$$f(x) = 1/a \text{ for } 0 \le x \le a$$

$$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n} = a^{-n} \text{ for } 0 \le x_i \le a$$

$$\frac{dL(a)}{da} = \frac{-n}{a^{n+1}} = -na^{-(n+1)}$$

$$\hat{a} = \max(x_i)$$

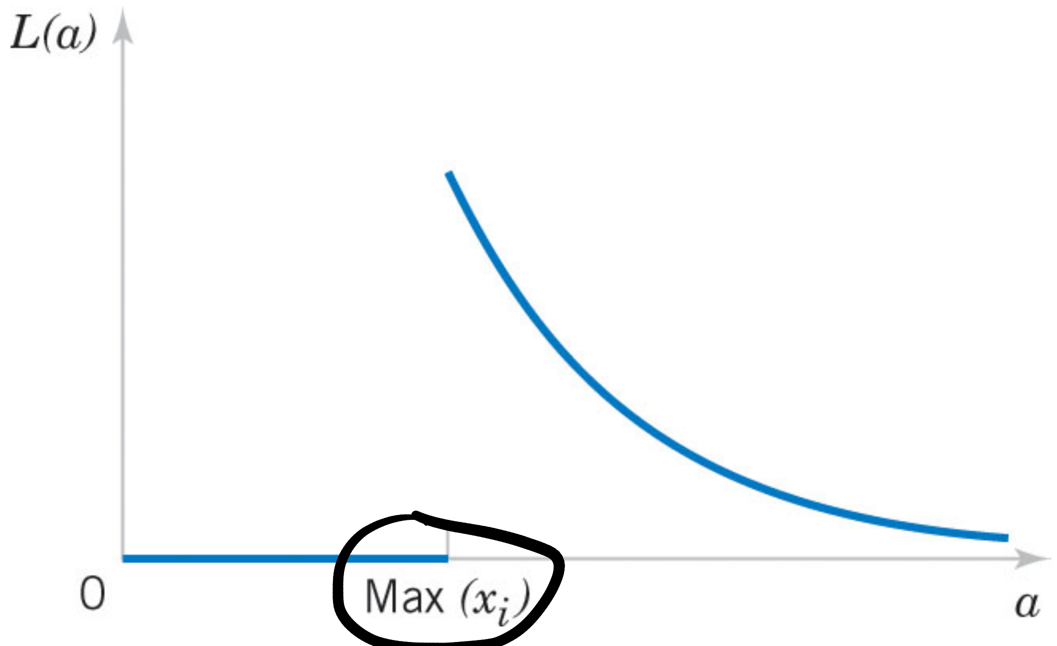


Figure 7-8 **The likelihood function for this uniform distribution**

Calculus methods don't work here because $L(a)$ is maximized at the discontinuity.

Clearly, a cannot be smaller than $\max(x_i)$, thus the MLE is $\max(x_i)$.

Methods of Moments

Population and samples moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

Population moments

$$\mu'_k = \begin{cases} \int_x x^k f(x)\,dx & \textbf{If } \boldsymbol{x} \textbf{ is continuous} \\ \sum_x x^k f(x) & \textbf{If } \boldsymbol{x} \textbf{ is discrete} \end{cases}$$

Sample moments

$$m'_k = \frac{\sum_{i=1}^{n} X_i^k}{n}$$

Method of Moments

- **Equating empirical moments to theoretical moments**

> Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

***m* equations for *m* parameters**

$$\begin{cases} m_1' = \mu_1' \\ m_2' = \mu_2' \\ \vdots \\ m_m' = \mu_m' \end{cases}$$

Example

MoM estimator for exponential parameter?

MoM estimator for normal distribution?

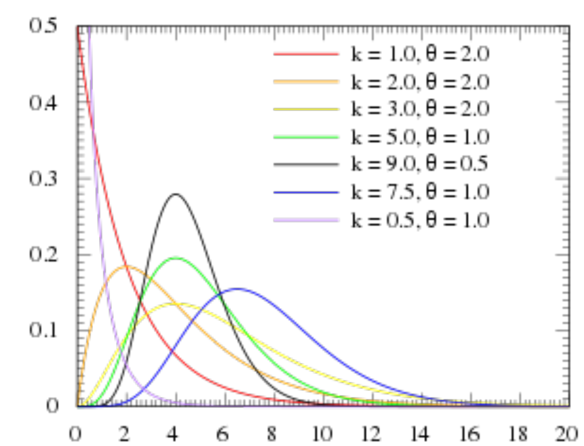
MoM for Gamma distribution

Method of moment estimator for Gamma distribution?

$$f(x_i) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$$

The likelihood function is difficult to differentiate because of the Gamma function $\Gamma(\alpha)$.

$$L(\alpha, \theta) = \left(\frac{1}{\Gamma(\alpha)\theta^\alpha}\right)^n (x_1 x_2 \cdots x_n)^{\alpha-1} \exp\left[-\frac{1}{\theta}\sum x_i\right]$$



We will use method of moment estimator

$$E(X) = \alpha\theta = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$Var(X) = \alpha\theta^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\alpha = \frac{\bar{X}}{\theta}$$

$$\hat{\theta}_{MM} = \frac{1}{n\bar{X}}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

MoM for Gamma distribution, known α

7-61. A random variable x has probability density function

$$f(x;\theta) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, \qquad 0 < x < \infty, \quad 0 < \theta < \infty$$

Given samples $x_1, \ldots x_n$,
find the MoM estimator for θ

Gamma distribution with $\alpha = 3$