Lecture 6
Statistical Inference

- Sampling Distribution
- Point Estimator

Contents

Model for Samples: Random sampling

Random Sample

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

Observations in a random sample are also known as

independent and identically distributed (***i.i.d.***)

random variables

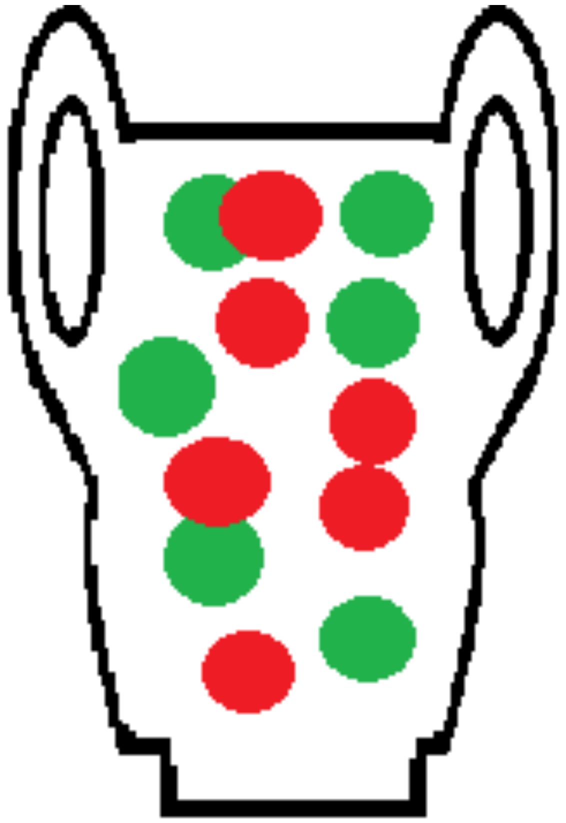
Statistic

A **statistic** is any function of the observations in a random sample.

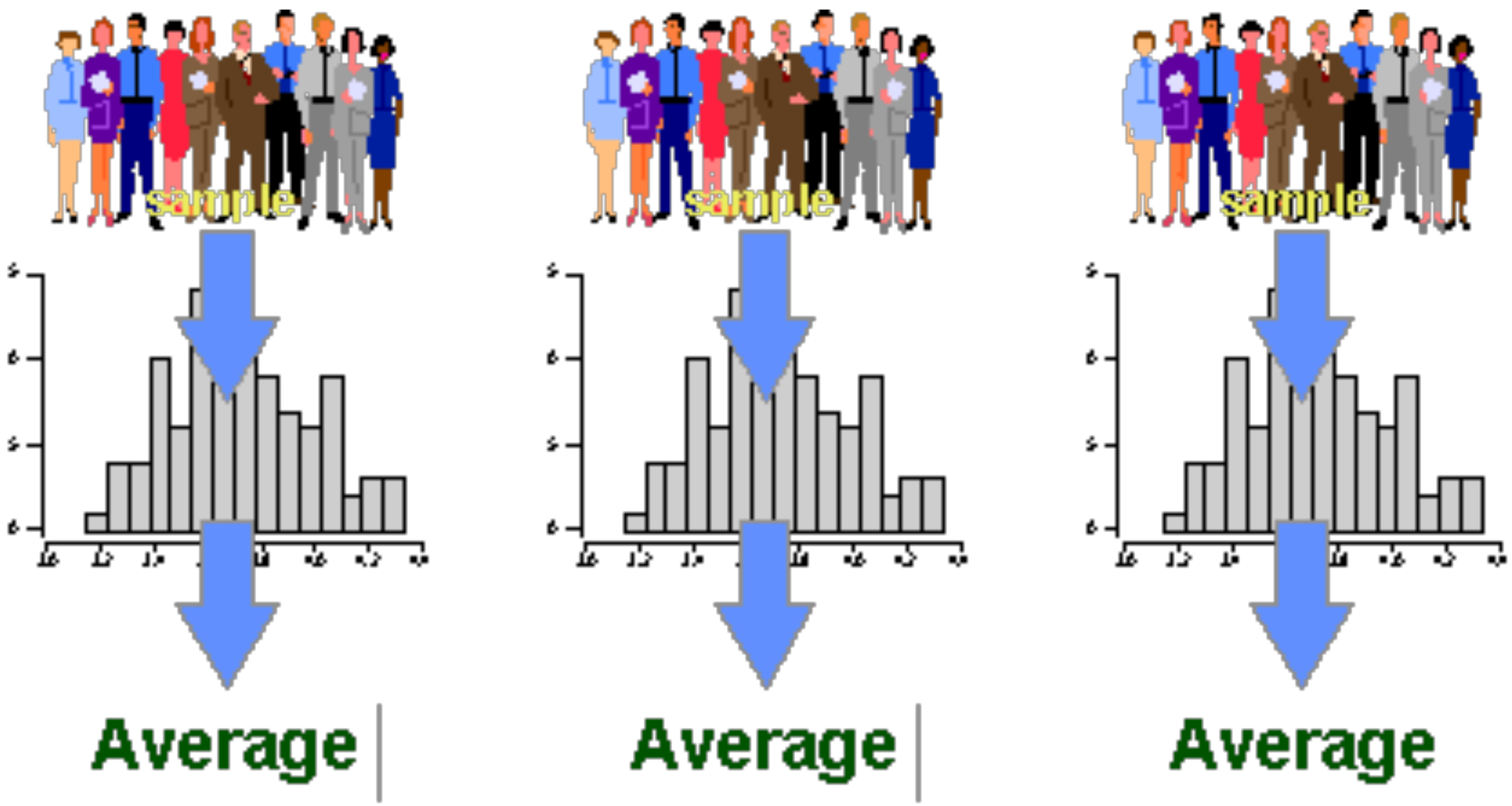
e.g., $X_1, X_2, \cdots, X_n \rightarrow \overline{X}, S^2$

The probability distribution of a statistic is called a **sampling distribution.**

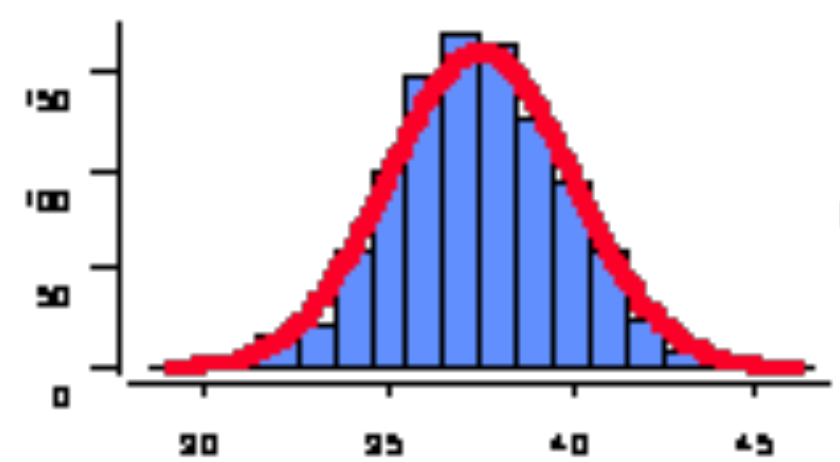
Class activity: urn model



Sampling distribution



The Sampling Distribution...



...is the distribution of a statistic across an infinite number of samples

Why sampling distribution?

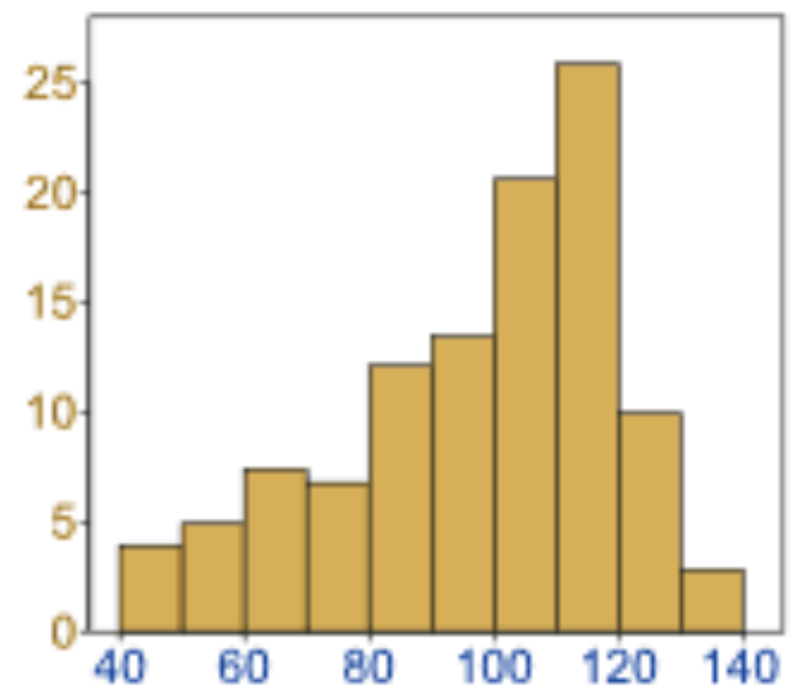
The probability distribution of a statistic is called a **sampling distribution.**

Statistical inference is concerned with making decisions about a population based on the information contained in a *random sample* from that population.

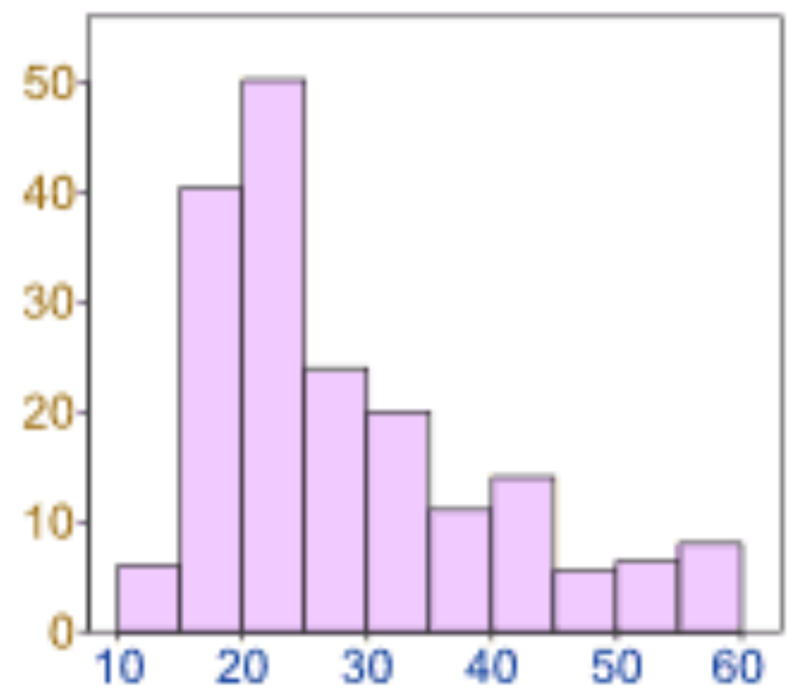
Sampling distribution is the link between probability and statistics.

Empirical distribution

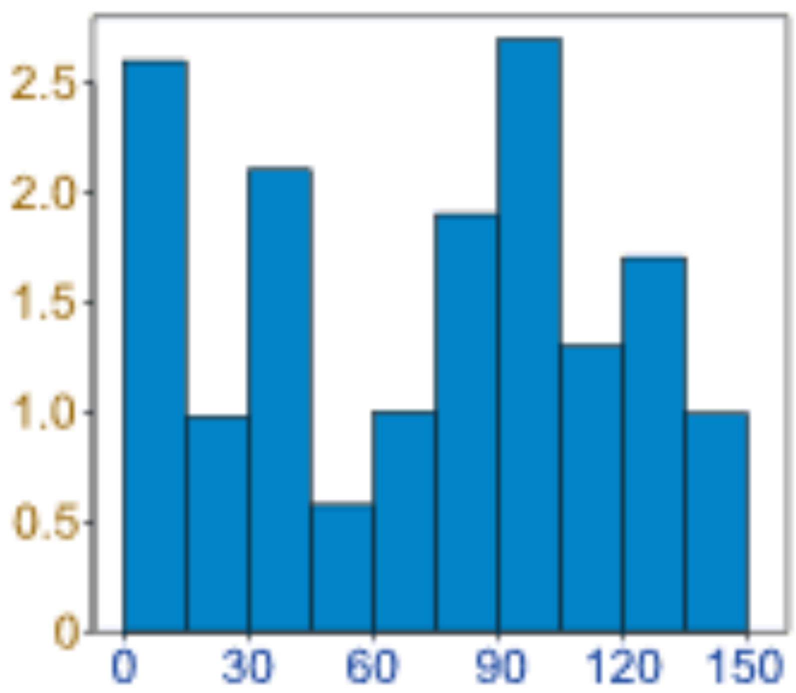
- **Data can be "distributed" (spread out) in different ways**



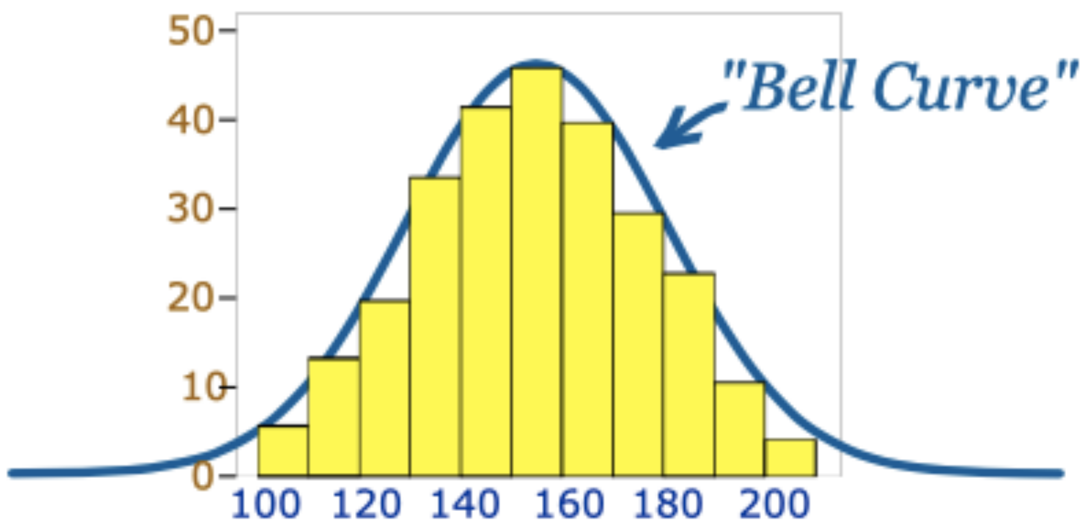
It can be spread out more on the left



Or more on the right



Or it can be all jumbled up



A Normal Distribution

Model sampling distribution

- Relationships between Bernoulli and Binomial distributions

$$X_i \sim BERN(p), i = 1, 2, \ldots, n$$

$$\bar{X} = \sum_{i=1}^{n} X_i \sim BIN(n, p)$$

- In this setting
- Each time X_i is the outcome of each draw:

  = 1, if black, otherwise = 0
- $\bar{X}$ is the number of black stones
- Multiple experiments $\bar{X}$ is different and has variability

Alternative view

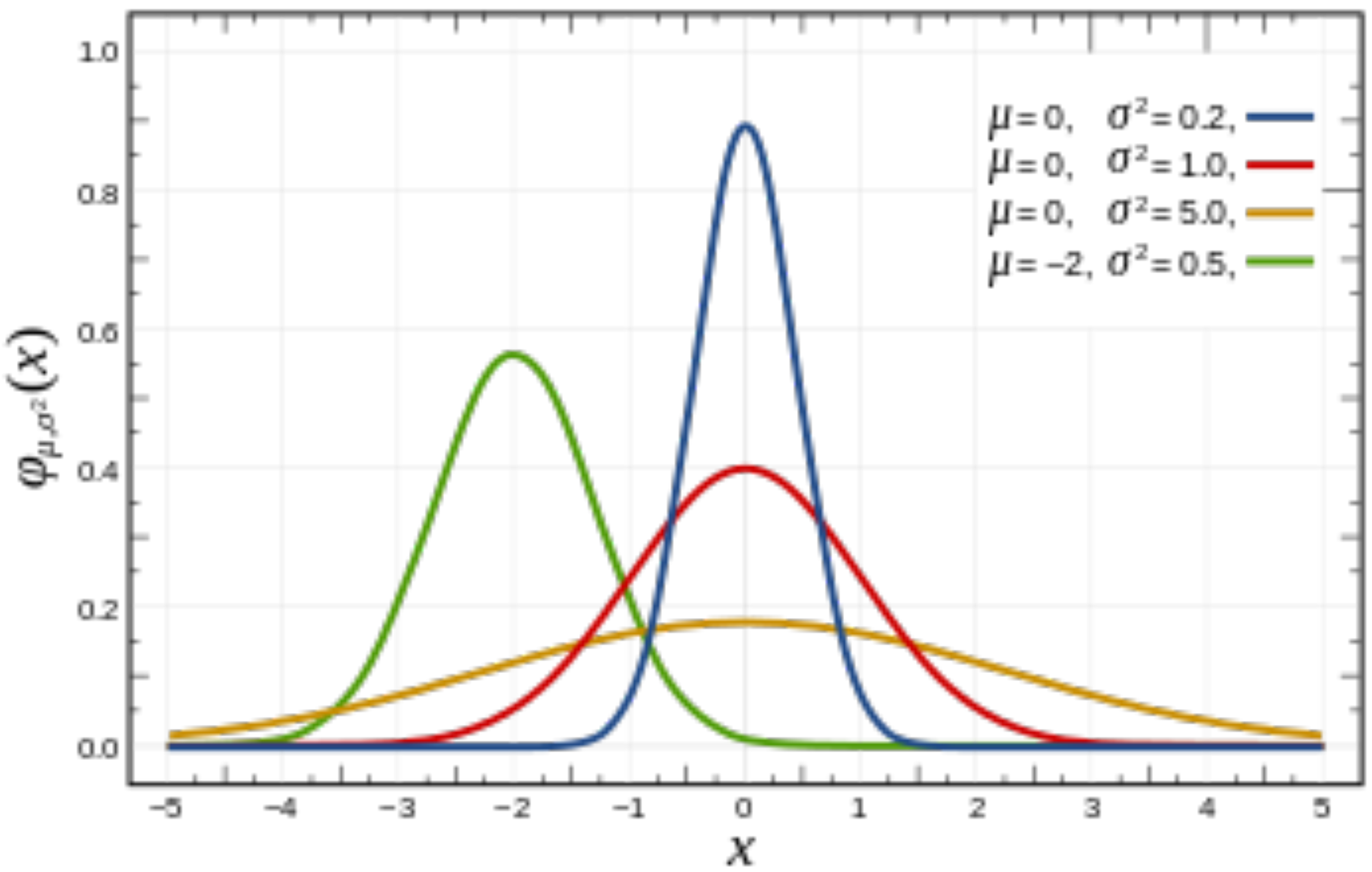
- Sample proportion is the percentage of black stones

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

- Claim: $\bar{X}$ is approximately normal distributed with mean p and variance = $\frac{p(1-p)}{n}$

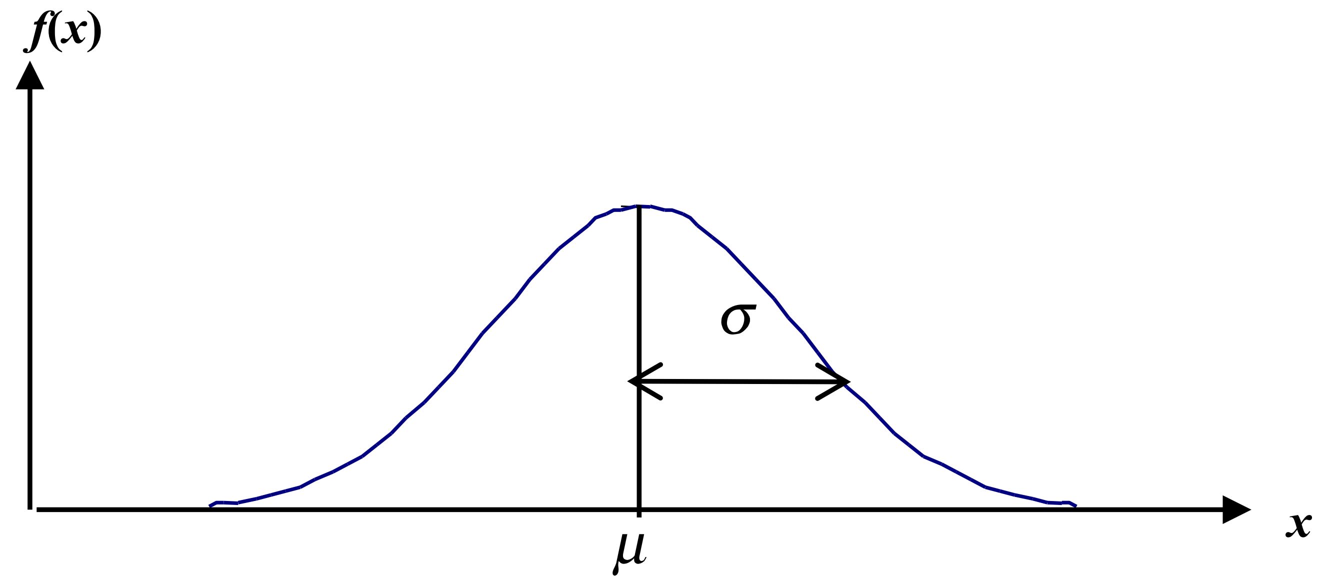
Sampling distribution describes the distribution of sample mean

Normal Distribution



Normal Distribution

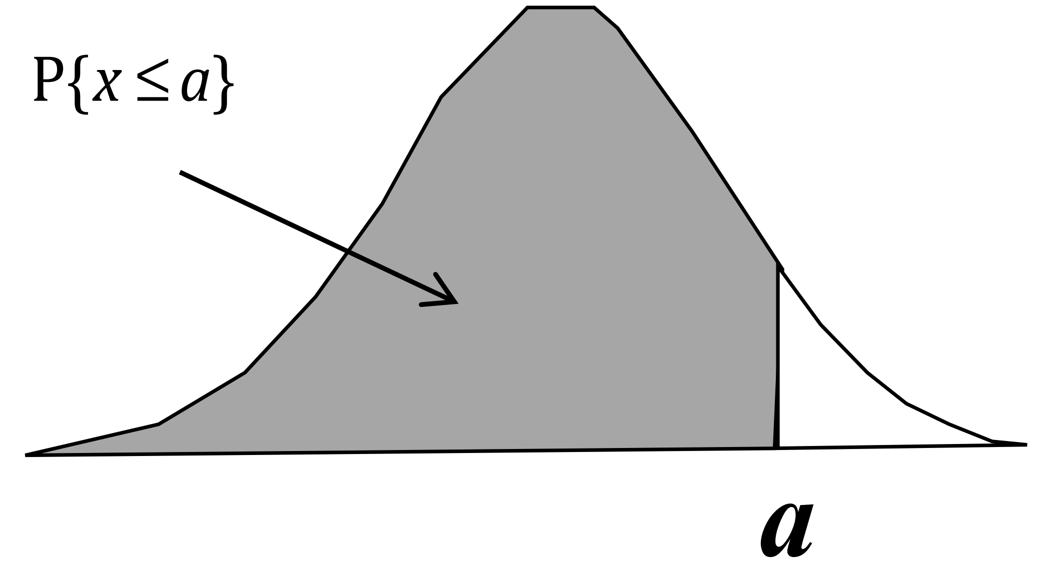
$$X \sim N(\mu, \sigma^2); \qquad -\infty < x < +\infty$$



$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$E(x) = \mu \qquad Var(x) = \sigma^2$$

$$\mathrm{P}\{x \le a\} = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$$



Important Fact

- **Fact:** If x_1, x_2 are independently normally distributed variables, then

$$y = x_1 + x_2$$

also follows the normal distribution:

$$y \sim N(\mu_1+\mu_2, \sigma_1^2 + \sigma_2^2)$$

Special case

$Var(aX) = a^2 Var(X)$

- **Making normal assumption about samples:**

  Xi' s are **normally** independently distributed (a random sample from a Normal distribution with the known variance)

$$X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2) \rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- **Proof?**

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

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

$\bar{X} = \frac{X_1+X_2}{2} \sim N(\mu, \frac{2\sigma^2}{4}) = N(\mu, \frac{\sigma^2}{2})$

Sampling distribution of <u>sample mean</u> is normal, when samples are normal

Example

1. The design of the machine has fill volume 300 mls, and variance 9ml. An engineer takes a random sample of 25 cans, what's the sampling distribution of mean filling volume of a can of soft drink?

• from scipy.stats import norm

• norm.cdf(-3.33)

2. The engineer finds the sample mean of fill volume to be 298 mls. Is this considered to be normal?

1. $\mu = 300, \quad \sigma^2 = 9 \quad n = 25$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(300, \frac{9}{25}\right)$$

$Z \sim N(0,1)$

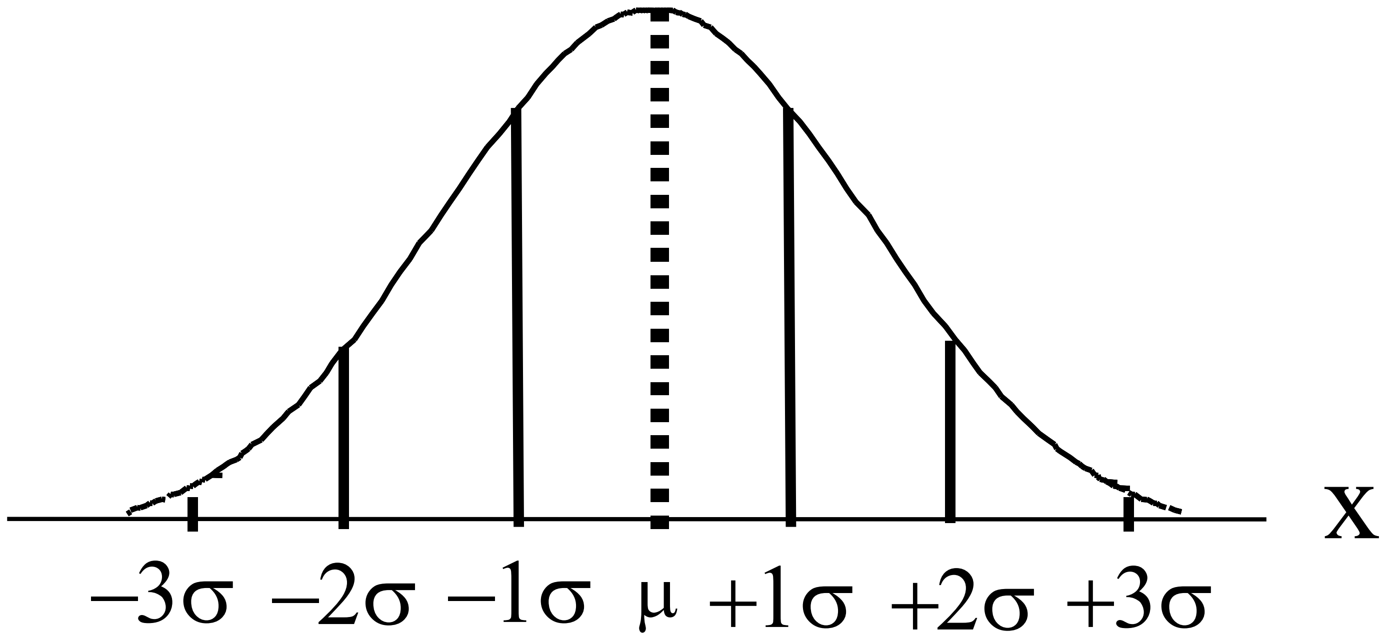
2. $$\Pr(\bar{X} \leq 298) = \Pr\left(\frac{\bar{X}-300}{3/5} \leq \frac{298-300}{3/5}\right) = \Pr(Z \leq -3.33)$$

$\approx 0.04\%$

Standard Normal Distribution

$$X \sim N(\mu, \sigma^2); \qquad -\infty < x < +\infty$$

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right\}$$

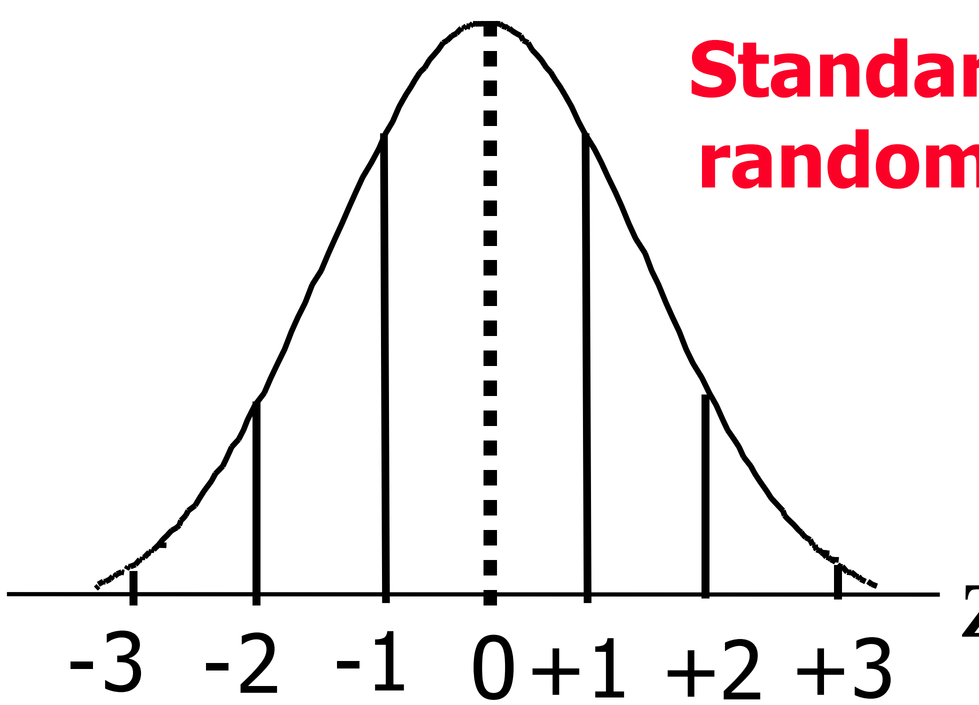


Map any X into Z

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

$$Z \sim N(0, 1^2); \qquad -\infty < z < +\infty$$

$$f(z; \mu = 0, \sigma^2 = 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2} \right)$$



Standard normal random variable

$Pr(|z|>1) = 0.32$ $\quad Pr(|z|>2) = 0.05$

$Pr(|z|>3) = 0.0026$

Exercise

- Example: Three shafts are made and assembled in a machine. The length of each shaft, in centimeters, is distributed as follows:

  Shaft 1: ~ N (75, 0.09)
  Shaft 2: ~ N (60, 0.16)
  Shaft 3: ~ N (25, 0.25)

  Assume the shafts' length are independent to each other:
  (a) What is the distribution of the linkage?
  (b) What is the probability that the linkage will be longer than 160.5 cm?

$Y \sim N(160, 0.5)$

(b)

$Pr(Y > 160.5) = Pr\left(Z > \frac{160.5-160}{\sqrt{0.5}}\right)$

$= Pr\left(Z > \frac{\sqrt{2}}{2}\right) = Pr(Z > 0.7)$

$= 1 - \Phi(0.7)$

≈ 0.24

Exercise: Airport Check-in

The amount of time that a customer spends waiting in the airport check-in counter is a normal random variable with mean 8.2 minutes and standard deviation 1.5 minutes. Suppose that a random sample of 49 customers is observed.

1. Find the probability that the average waiting time for these customers is:
(a) Less than 10 minutes;
(b) Between 5 and 10 minutes.

2. What is a value such that 90% of chance, average wait time will wait shorter than that?

General case: Central Limit Theorem (CLT)

Sampling distribution of <u>sample mean</u> is normal, even when samples are NOT normal

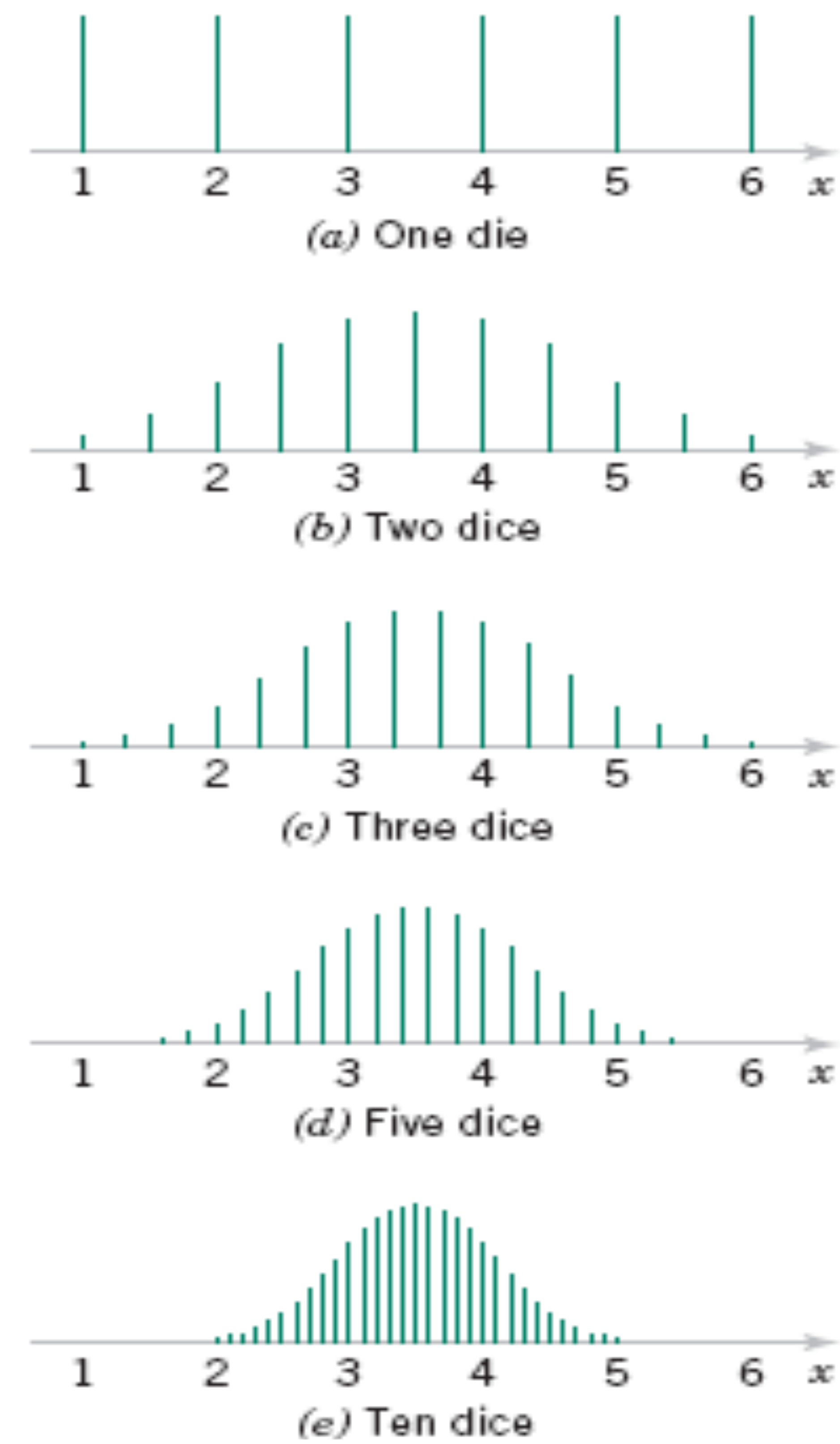


Figure 7-1 Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

Rule of thumb: when n > ~~30~~ 15 this works pretty well.

CENTRAL LIMIT THEOREM If all samples of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. This approximation improves with larger samples.

Exercise: Coffee

For the "number of coffee drink / day" question, assume the number of coffee drink / day is a random variable with mean 1 and variance 0.5.

There are 58 responses of survey. What the sampling distribution of the sample mean?

$\mu = 1 \quad \sigma^2 = 0.5 \quad n = 58$

$\bar{X} \sim N(1, \frac{0.5}{58})$

Sampling Distribution of Sample Mean With Known Variance

One Population:

Xi' s are normally independently distributed (a random sample from a Normal distribution with the known variance)

$$X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2) \rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Two Populations:

Two independent random samples from two Normal distributions with the known variances

$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1}) \quad \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

$$\left.\begin{array}{l} X_1, X_2, \cdots, X_{n_1} \sim N(\mu_1, \sigma_1^2) \\ \\ Y_1, Y_2, \cdots, Y_{n_2} \sim N(\mu_2, \sigma_2^2) \end{array}\right\} \rightarrow \bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$

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

$\bar{X}_1 \sim N(5000, \frac{40^2}{16})$ $\bar{X}_2 \sim N(5050, \frac{30^2}{25})$ $\bar{X}_1 - \bar{X}_2 \sim N(-50, \frac{40^2}{16} + \frac{30^2}{25})$ $Pr(\bar{X}_1 - \bar{X}_2 \geq 25)$

Aircraft Engine Life

- The effective life of a component, X1, used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The engine manufacture design a new component X2, which increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Assume X1 and X2 is fairly close to a normal distribution. Suppose n1 = 16 samples of old components, and n2 = 25 samples from the new components, are selected. What is the probability that the difference in two sample means is at least 25 hours?

Central Limit Theorem (CLT) for two populations

Approximate Sampling Distribution of a Difference in Sample Means

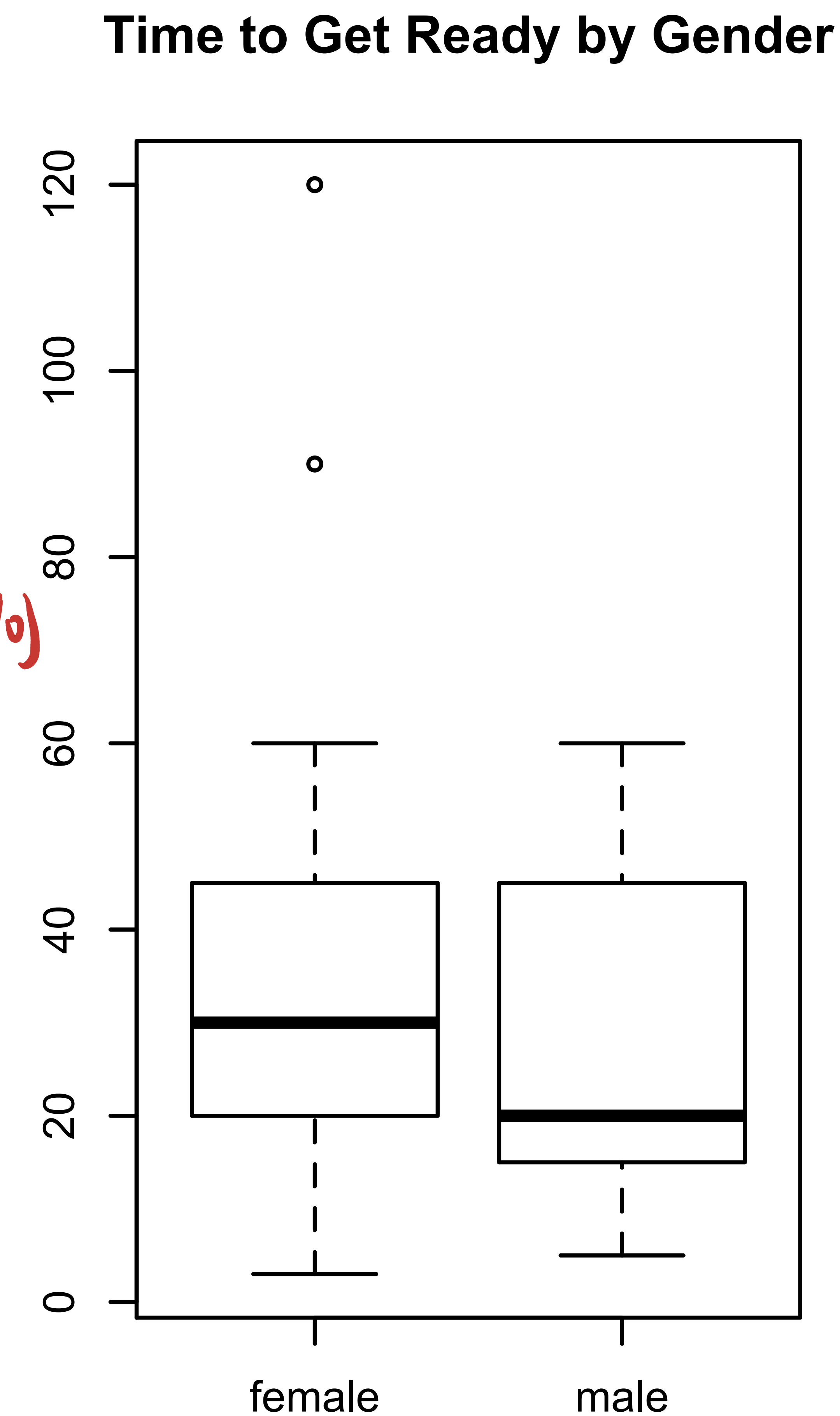
If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}4)$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

Example: Time in the morning

- The time for students to get ready in the morning, for male and female students.
- There are n1 = 33 girls and n2 = 25 guys who provides the answer
- Assume the time for girl is a random variable with mean 30 minutes, and the time for guy is a random variable with mean 20 minutes. The standard deviation for both of them is 10 minutes.
- What is the probability that the difference in two sample means is at least 10 minutes?



$Pr(\bar{X}_F - \bar{X}_M \geq 10)$

$= 0.5$

$\bar{X}_{female} \sim N(30, \frac{10^2}{33})$ $\quad \bar{X}_{male} \sim N(20, \frac{10^2}{25})$

Of course the variance is unknown…

$\bar{X}_F - \bar{X}_M \sim N(10, \frac{10^2}{33} + \frac{10^2}{25})$

Contents

Outline

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

Questions we aim to address

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

$$\begin{tikzcd}
\textbf{Data} \arrow[r] & \textbf{Statistics} \arrow[dr] & \textbf{Sampling distribution} \arrow[l] \arrow[d] \\
& & \textbf{Estimator}
\end{tikzcd}$$

Estimator

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

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

Is called a **point estimator** of θ.

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

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

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

Internet service provider

- **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?

Google Fiber

xfinity

- What's the difference of their rate?
- Samples
  - First service provider $X_i, i = 1, 2, \dots, n_1$
  - Second service provider $Y_i, i = 1, 2, \dots, n_2$
- Assumption
  - $X_i \sim N(\mu_1, \sigma_1^2)$
  - $Y_i \sim N(\mu_2, \sigma_2^2)$
- **Parameters of interest:** $\mu_1 - \mu_2$
- **Estimator:** $\bar{X} - \bar{Y}$
- **Estimate: 4.9744- 5.0740 = -0.0996 (mbp)**

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

Basic properties of estimators

Standard error of estimator

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 estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error,** denoted by $\hat{\sigma}_{\hat{\Theta}}$.

ties to sampling distribution

$$
\begin{tikzcd}
\textbf{Data} \arrow[r] & \textbf{Statistics} \arrow[dr] & \textbf{Sampling distribution} \arrow[l] \arrow[d] \\
 & & \textbf{Estimator}
\end{tikzcd}
$$

Internet service provider

- **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 standard error of the estimator for the difference of their rate?

- What's the difference of their rate?
- Samples
  - First service provider $X_i, i = 1, 2, \ldots, n_1$
  - Second service provider $Y_i, i = 1, 2, \ldots, n_2$
- 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:** $\overline{X} - \overline{Y}$

- **Standard error of the estimator**

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

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:

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

$\overline{X}$: sample mean.

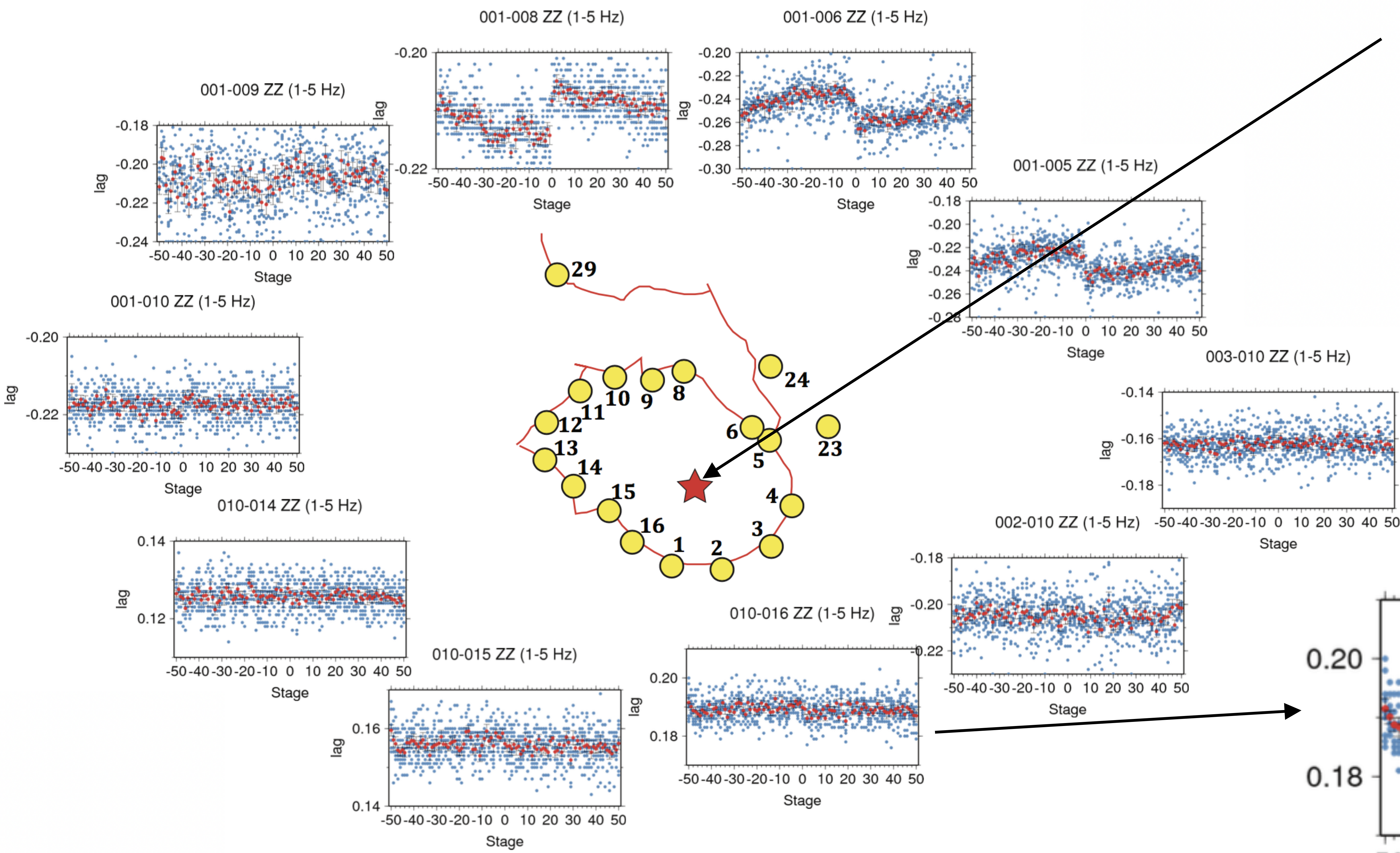
What is the estimator for the conductivity?

What is the standard error of the estimator? $\sqrt{\frac{\sigma^2}{n}} \approx \sqrt{\frac{S^2}{n}}$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$: point estimator of population variance.

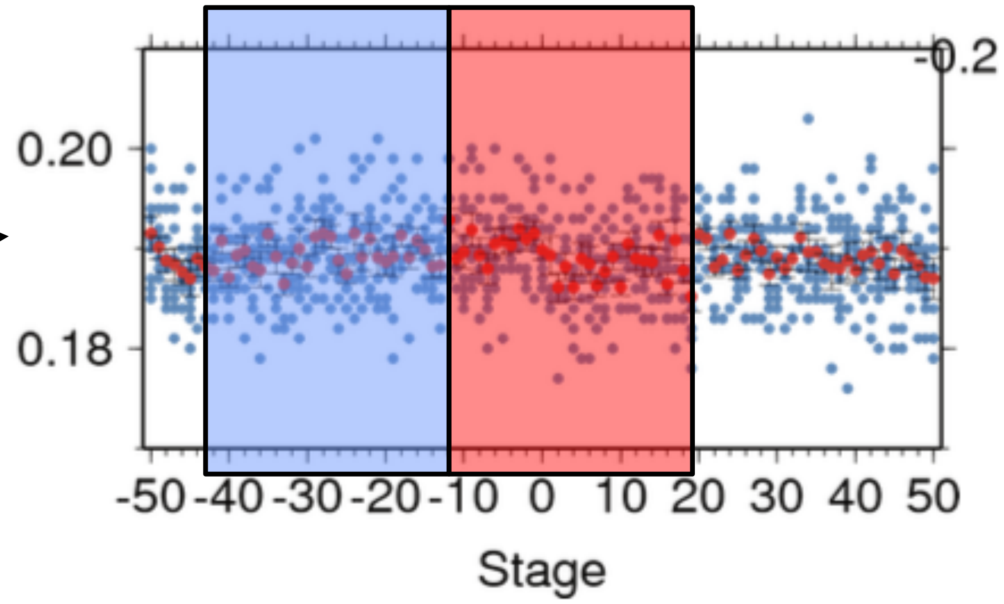
A real-world example

- **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

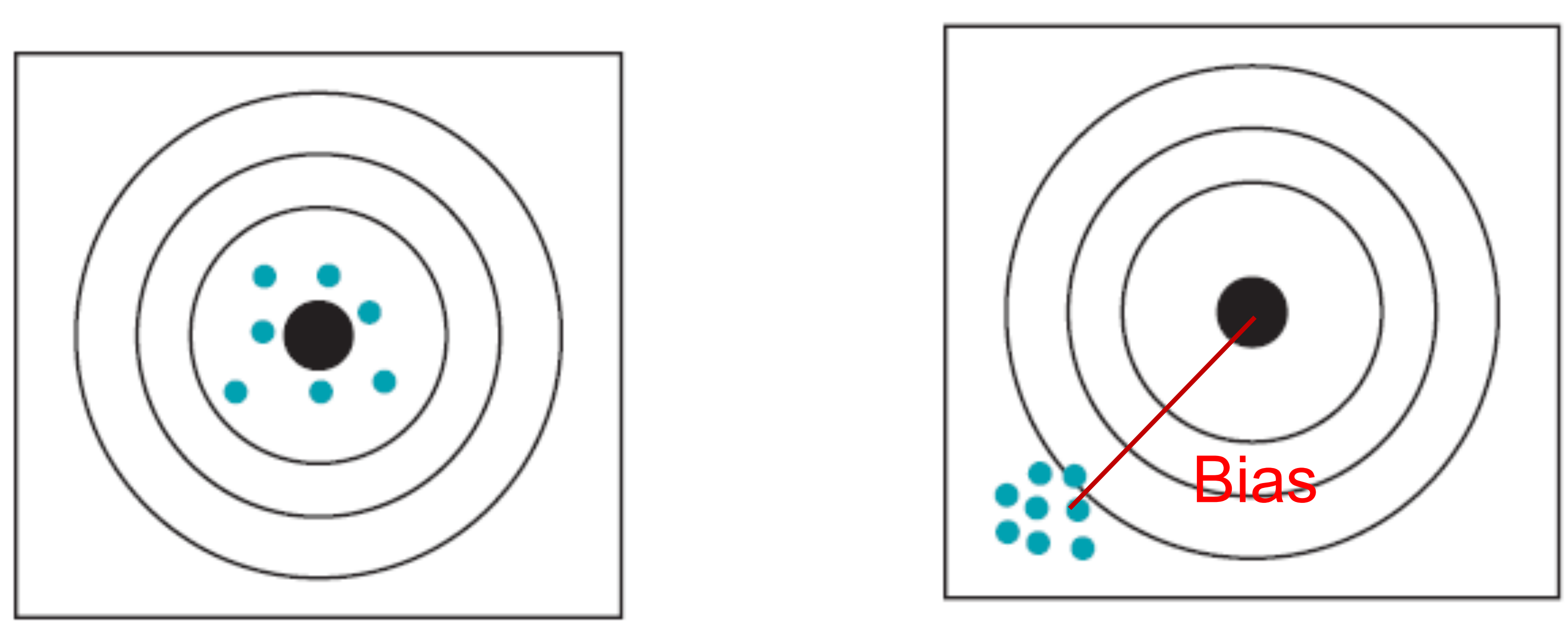
$$\boxed{E(\hat{\Theta})} = \theta$$

ties to sampling distribution

If the estimator is not unbiased, then the difference

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

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



Sample mean is unbiased estimator

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

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

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

Variance of a Point Estimator

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

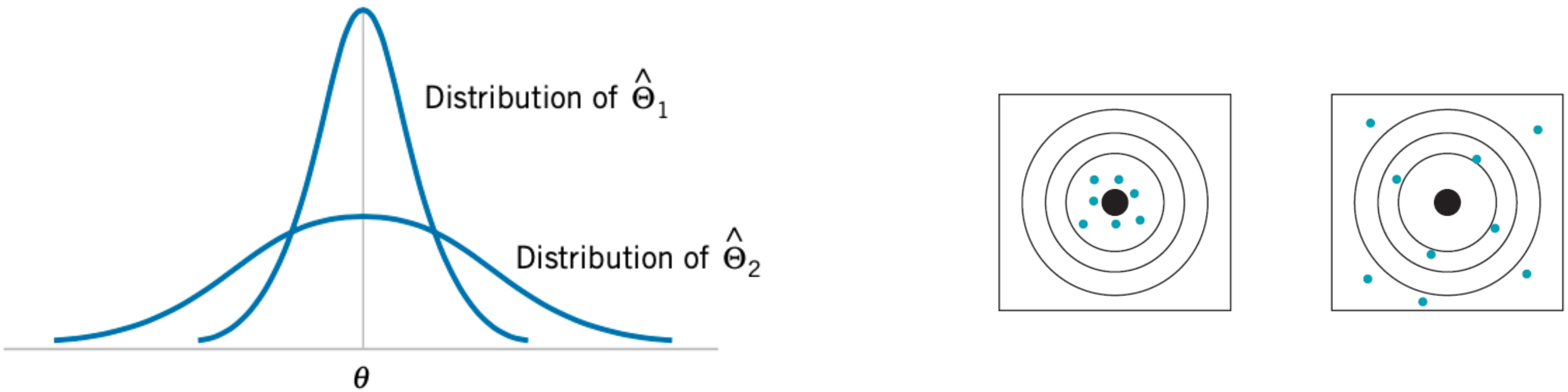


Figure 7-1 The sampling distributions of two unbiased estimators

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

ties to sampling distribution

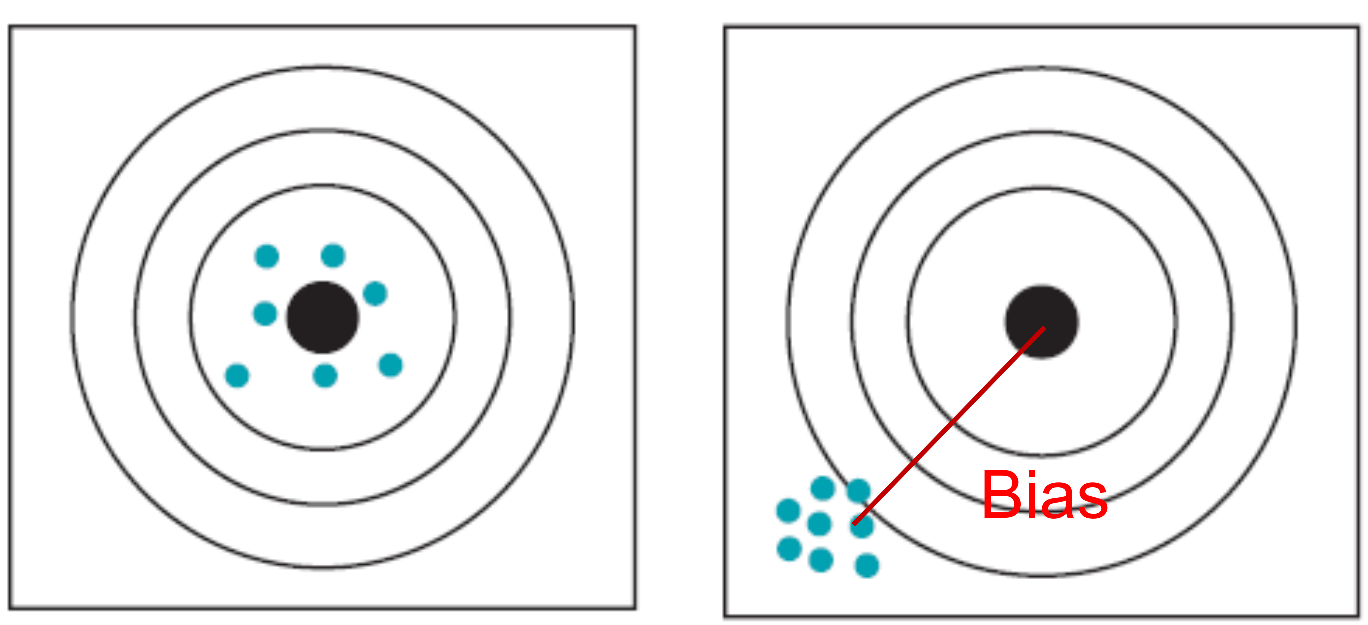
Mean Square Error (MSE)

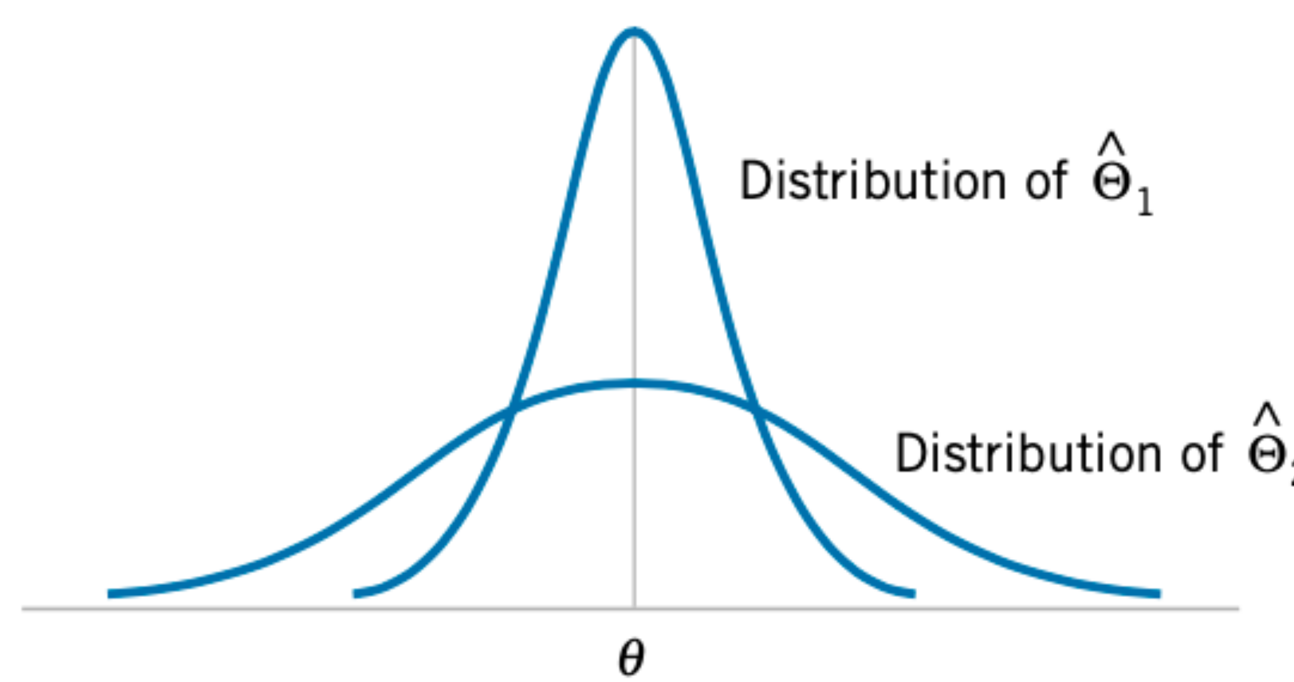
The **mean square error** of an estimator $\hat{\Theta}$ of the parameter θ is defined as

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

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

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





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:

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

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

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

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

$= \frac{Var(X_1)+Var(X_2)}{4} = \frac{\sigma^2}{2}$

$Var(\hat{\theta}_2) = \frac{1}{16} Var(X_1) + \frac{9}{16} Var(X_2)$

$= \frac{10\sigma^2}{16} = \frac{5\sigma^2}{8}$

(a) Are both estimators unbiased estimators of μ?

There is not a unique unbiased estimator!

(b) What is the variance of each estimator?

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

$MSE(\hat{\theta}_1) = (bias(\hat{\theta}_1))^2 + Var(\hat{\theta}_1) = \frac{\sigma^2}{2}$

$MSE(\hat{\theta}_2) = \frac{5\sigma^2}{8}$

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.

$$\hat{\Theta}_1 = \frac{\sum_{i=1}^{7} X_i}{7} \qquad \frac{\sigma^2}{7}$$

$$\hat{\Theta}_2 = \frac{2X_1 - X_6 + X_4}{2} \qquad \mathrm{Var}(\hat{\Theta}_2) = \frac{4\sigma^2 + \sigma^2 + \sigma^2}{4} = 1.5\sigma^2$$

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

$\downarrow$ biased

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

Example

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

- Show that $\dfrac{\overline{X}}{2}$ is an unbiased estimator of θ

- Calculate the MSE of $\dfrac{\overline{X}}{2}$ and $\overline{X}$

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]

- 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

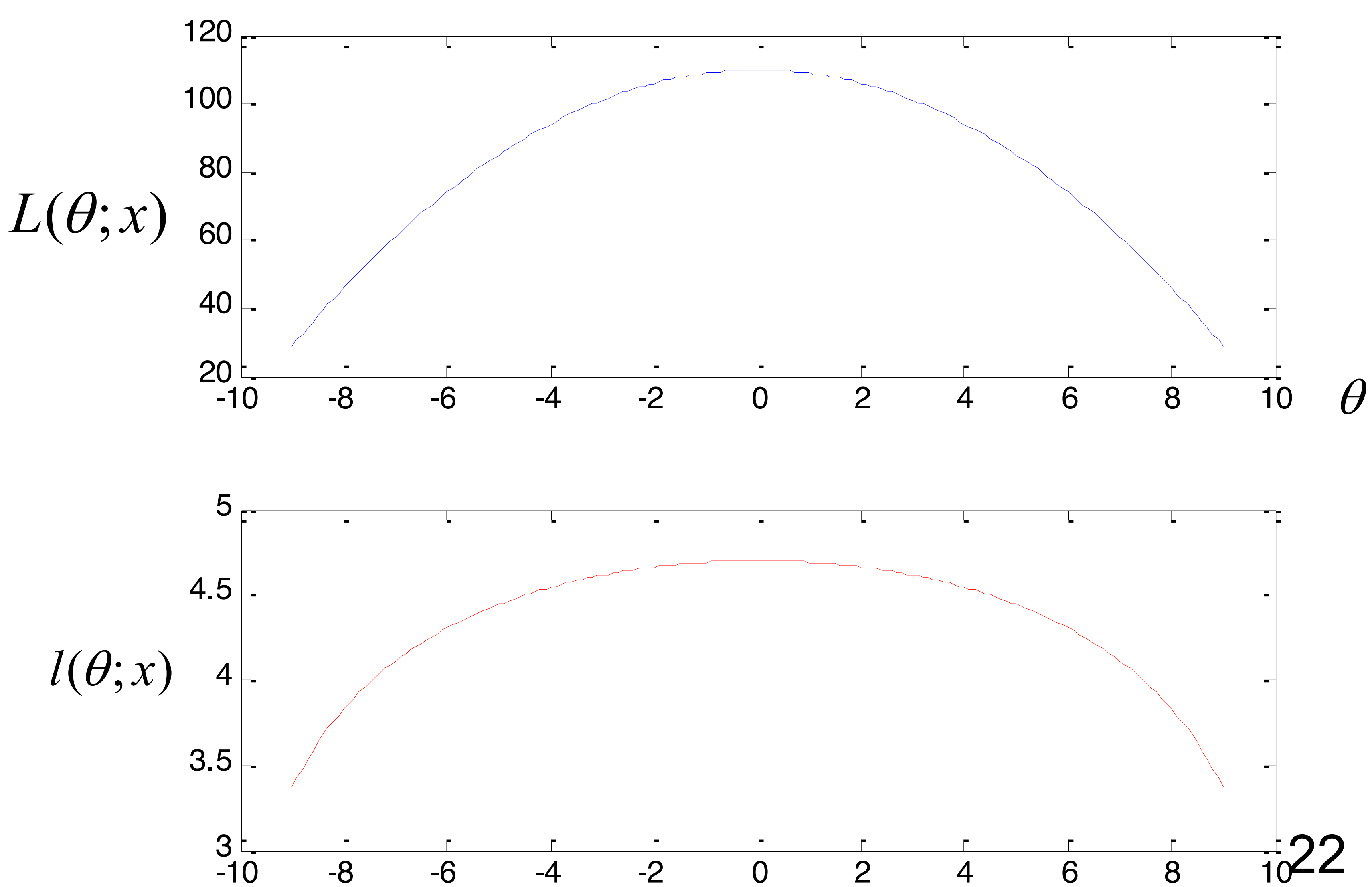
$$L(\theta) = f(x_1;\theta)\cdot f(x_2;\theta)\cdot \cdots \cdot f(x_n;\theta) \tag{7-5}$$

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 $L(\theta)$.

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

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

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



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

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

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

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

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

Example: normal

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

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

and

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

$\longrightarrow$ What is the MLE for μ?

Example (Continued, unknown variance)

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

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

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

The solutions to the above equation yield the maximum likelihood estimators

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

MLE: Exponential

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

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

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

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

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

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"**

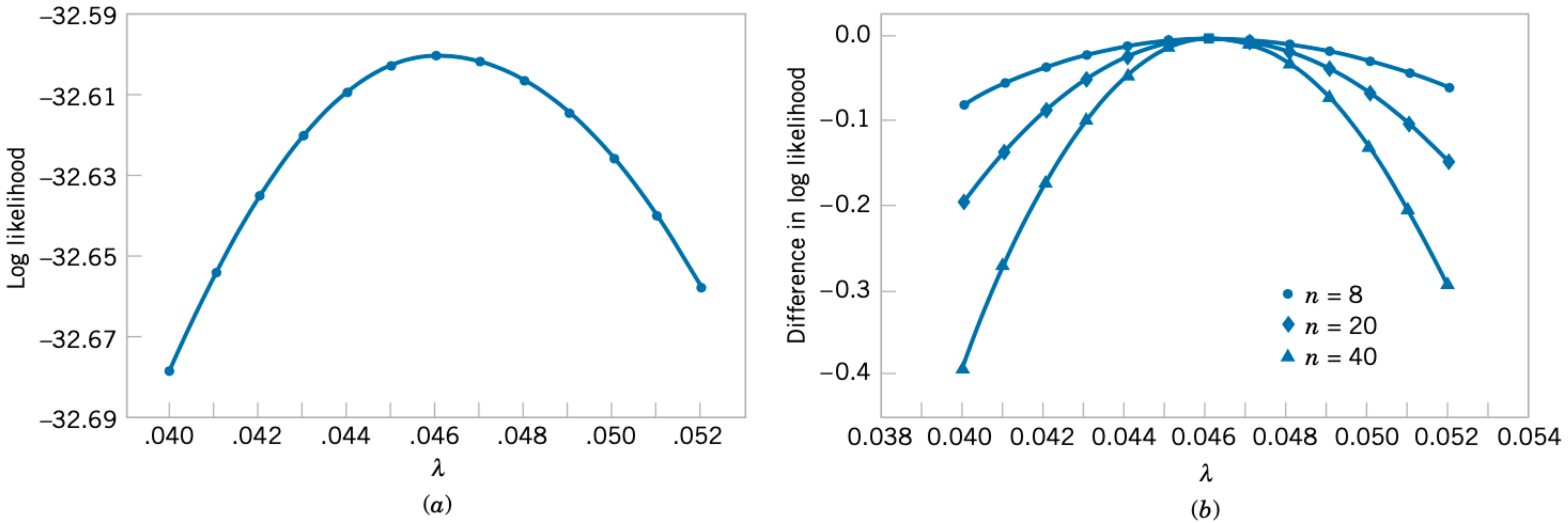


Figure 7-3 Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Log likelihood if $n = 8$, 20, and 40.

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 $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

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

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

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

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.

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

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

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

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

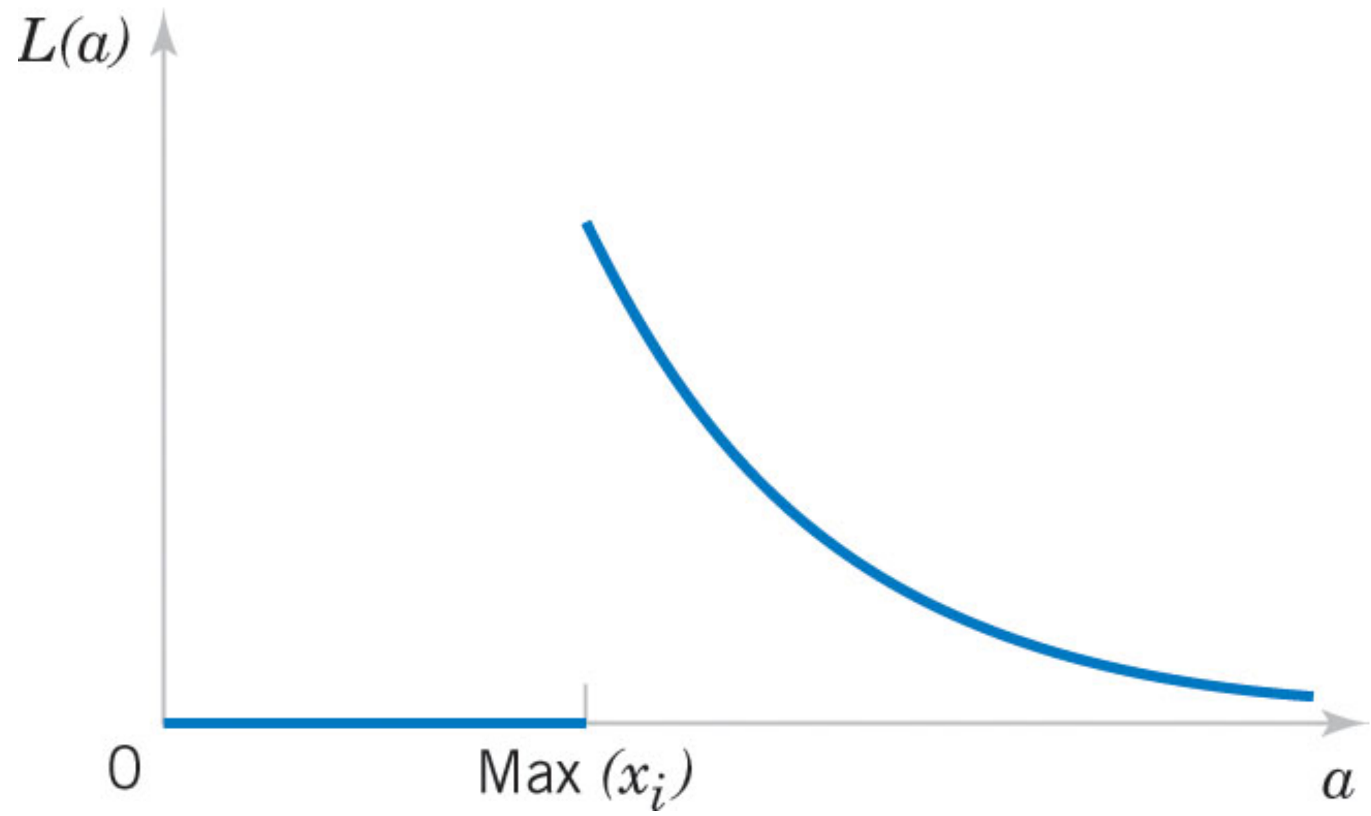


Figure 7-8 The likelihood function for this uniform distribution

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?

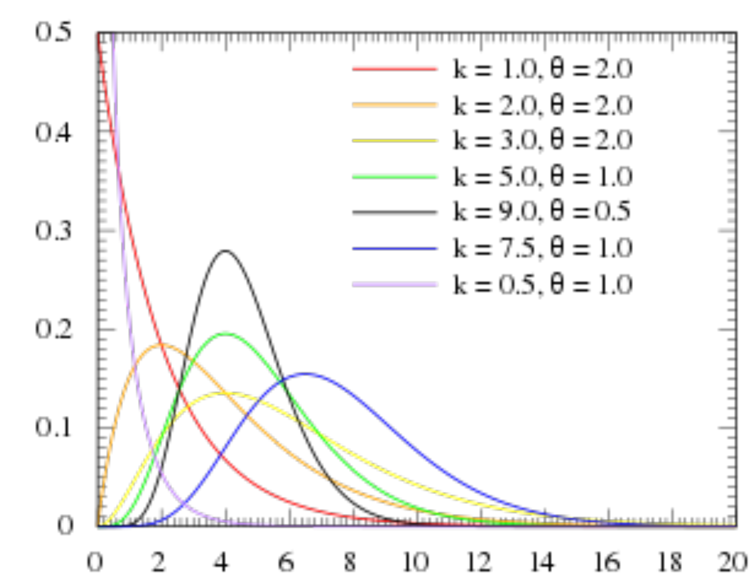
MoM for Gamma distribution

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 $x_1, \dots x_n$,
find the MoM estimator for θ

Gamma distribution with $\alpha = 3$