

# **Chapter 6. Point Estimation**

# **Chapter 7. Interval Estimation**

---

Math 3215 Spring 2024

Georgia Institute of Technology

## **Section 6.3.**

### **Order Statistics**

---

## Order Statistics

Order statistics are the observations of the random sample, arranged, or ordered, in magnitude from the smallest to the largest.

### Example

The values  $x_1 = 0.62$ ,  $x_2 = 0.98$ ,  $x_3 = 0.31$ ,  $x_4 = 0.81$ , and  $x_5 = 0.53$  are the  $n = 5$  observed values of five independent trials of an experiment with PDF  $f(x) = 2x$  for  $0 < x < 1$ .

The observed order statistics are  $y_1 = 0.31$ ,  $y_2 = 0.53$ ,  $y_3 = 0.62$ ,  $y_4 = 0.81$ ,  $y_5 = 0.98$

The sample median is  $y_3 = 0.62$

The sample range is  $[y_1, y_5] = [0.31, 0.98]$

## Order Statistics

Let  $X_1, X_2, X_3, X_4, X_5$  be independent random variables with density  $f(x) = 2x$  for  $0 < x < 1$ .

Let  $Y_1 \leq Y_2 \leq Y_3 \leq Y_4 \leq Y_5$  be the order statistics of  $X_i, i = 1, 2, 3, 4, 5$ .

What is the distribution of  $Y_1$ ?

$$\begin{aligned} F_{Y_1}(x) &= P(Y_1 \leq x) = 1 - P(Y_1 > x) = 1 - P(X_1 > x, \dots, X_5 > x) \\ &= 1 - (P(X_1 > x))^5 \\ &= 1 - (1 - P(X_1 \leq x))^5 \\ &= 1 - (1 - x^2)^5 \end{aligned}$$

for  $0 < x < 1$



## Order Statistics

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x 2t dt = x^2$$

Let  $X_1, X_2, X_3, X_4, X_5$  be independent random variables with density  $f(x) = 2x$  for  $0 < x < 1$ .

Let  $Y_1 \leq Y_2 \leq Y_3 \leq Y_4 \leq Y_5$  be the order statistics of  $X_i, i = 1, 2, 3, 4, 5$ .

What is the distribution of  $Y_5$ ?

$$\begin{aligned} F_{Y_5}(x) &= P(Y_5 \leq x) = P(X_1 \leq x, \dots, X_5 \leq x) \\ &= P(X_1 \leq x) \cdot \dots \cdot P(X_5 \leq x) \\ &= (P(X_1 \leq x))^5 = (x^2)^5 \\ &= x^{10} \quad \text{for } 0 < x < 1. \end{aligned}$$

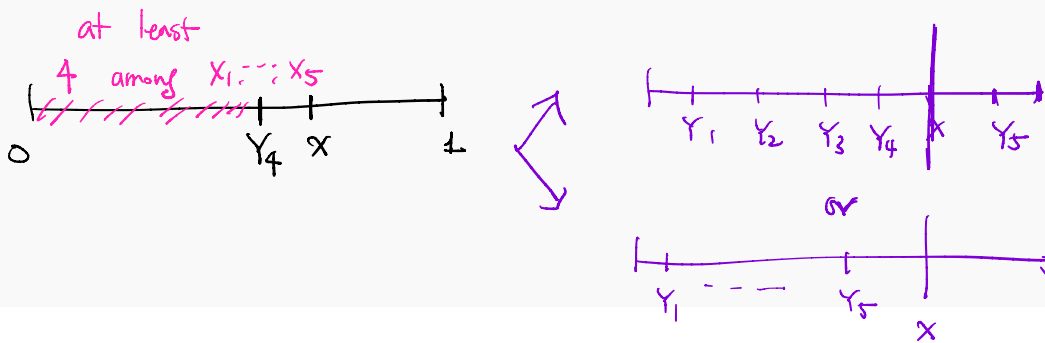
$$f_{Y_5}(x) = 10 \cdot x^9 \quad \text{for } 0 < x < 1.$$

## Order Statistics

Let  $X_1, X_2, X_3, X_4, X_5$  be independent random variables with density  $f(x) = 2x$  for  $0 < x < 1$ .

Let  $Y_1 \leq Y_2 \leq Y_3 \leq Y_4 \leq Y_5$  be the order statistics of  $X_i, i = 1, 2, 3, 4, 5$ .

What is the distribution of  $Y_4$ ?



$$\begin{aligned}
 F_{Y_4}(x) &= P(Y_4 \leq x) \\
 &= P(X_1 \leq x, X_2 \leq x, \dots, X_5 \leq x) \\
 &\quad + \\
 &\quad P(4 \text{ of } X_i \leq x, 1 \text{ of } X_i > x) \\
 &= x^{10} + P(X_1 > x, \text{ the others } \leq x) \\
 &\quad + P(X_2 > x, \text{ " } \leq x) \\
 &\quad \vdots \\
 &\quad + P(X_5 > x, \text{ the others } \leq x)
 \end{aligned}$$

$$\begin{aligned}
&= x^{10} + 5 \cdot P(X_1 > x) \cdot P(X_2 \leq x) P(X_3 \leq x) P(X_4 \leq x) P(X_5 \leq x) \\
&= x^{10} + 5 \cdot (1-x^2) (x^2)^4 \\
&\quad \binom{5}{1} = \binom{5}{4}
\end{aligned}$$

$$F_{Y_3}(x) = P(Y_3 \leq x)$$

$$\begin{aligned}
&= P(Y_5 \leq x) + P(\underline{Y_4 \leq x < Y_5}) + P(\underline{Y_3 \leq x < Y_4}) \\
&= (P(X_i \leq x))^5 + \binom{5}{1} (P(X_i \leq x))^4 \cdot P(X_i > x) \\
&\quad + \binom{5}{2} (P(X_i \leq x))^3 (P(X_i > x))^2
\end{aligned}$$

## Order Statistics

In general, if  $Y_1 < Y_2 < \dots < Y_n$  are the order statistics of independent random variables  $X_i$  for  $i = 1, 2, \dots, n$  with PDF  $f$ ,

then the distribution of  $Y_i$  is

## Order Statistics

$$P(X_i \leq x) = x^2$$

### Example

Let  $X_1, X_2, X_3, X_4, X_5$  be independent random variables with density  $f(x) = 2x$  for  $0 < x < 1$ .

Let  $Y_1 \leq Y_2 \leq Y_3 \leq Y_4 \leq Y_5$  be the order statistics of  $X_i, i = 1, 2, 3, 4, 5$ .

Find  $P(Y_3 \leq 4)$ .

Find the density of  $Y_3$ .

$$P(Y_4 \leq x < Y_5)$$

$$F_{Y_3}(x) = \underbrace{P(X_1 \leq x)^5}_{P(Y_5 \leq x)} + \underbrace{\binom{5}{1} P(X_1 \leq x)^4 \cdot (1 - P(X_1 \leq x))}_{P(Y_3 \leq x < Y_4)}$$

$$= (x^2)^5 + \binom{5}{1} (x^2)^4 \cdot (1 - x^2) + \binom{5}{2} (x^2)^3 (1 - x^2)^2$$

$$P(Y_3 \leq 4) = F_{Y_3}(4) = 4^{10} + 5 \cdot 4^8 \cdot (1 - 16) + 10 \cdot 4^6 \cdot (1 - 16)^2$$

$$f_{Y_3}(x) = \frac{d}{dx} F_{Y_3}(x) = 10 \cdot x^9 + 5 \cdot 8 \cdot x^7 (1 - x^2) + 5 \cdot x^8 \cdot (-2x) + 10 \cdot 6 \cdot x^5 (1 - x^2)^2 + 10 \cdot x^6 \cdot 2(1 - x^2)(-2x)$$

## Exercise

Let  $Y_1 < Y_2 < \dots < Y_{19}$  be the order statistics of  $n = 19$  independent observations from the exponential distribution with mean  $\theta$ .

Find the PDF of  $Y_1$ .

Find  $\mathbb{E}[F(Y_1)]$  where  $F$  is the CDF of the exponential distribution.

**Section 6.4.**  
**Maximum Likelihood and Method of**  
**Moments Estimation**

---

## Estimators

We consider an experiment whose outcome follows a certain distribution.

Assume that the distribution is characterized by a PDF  $f(x; \theta)$  where  $\theta$  is an unknown parameter.

Let  $\Omega$  be the set of all possible parameters. We call it the parameter space.

Suppose we repeat the  $n$  independent experiment and observe the outcome (sample)  $X_1, X_2, \dots, X_n$ .

Our goal is to estimate  $\theta$  based on the samples.

$X_1, X_2, \dots, X_n \sim \text{i.i.d. Exp}(\lambda)$ ,  $\lambda$  : unknown

Assume  $\lambda \in \Omega$

- $\lambda \in [0, 1]$
- $\lambda = 1, \text{ or } 2 \text{ or } 3$
- $\lambda > 0$
- $\lambda > 0$

Estimate



## Estimators

### Definition

For a function  $u : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $u(X_1, X_2, \dots, X_n)$  is called an estimator (or a point estimator) of  $\theta$ .

## Estimators

### Example

i.i.d. Ber( $p$ )

Suppose  $X_1, X_2, \dots, X_n$  are independent Bernoulli RVs with unknown success probability  $p \in [0, 1]$ .

Parameter Space  $\Omega = [0, 1]$ .

We want to find an estimator of  $p$  which maximizes its possibility.

$$\hat{p} = u(x_1, x_2, \dots, x_n)$$

Possibility =  $\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid p)$  ← function of  $p \in [0, 1]$ .

$$= \mathbb{P}(X_1 = x_1) \mathbb{P}(X_2 = x_2) \dots \mathbb{P}(X_n = x_n)$$

0 or 1

$$= \left( p^{x_1} \cdot (1-p)^{1-x_1} \right) \left( p^{x_2} \cdot (1-p)^{1-x_2} \right) \dots \left( p^{x_n} \cdot (1-p)^{1-x_n} \right)$$
$$= p^{\overbrace{x_1 + x_2 + \dots + x_n}^A} (1-p)^{n - (x_1 + \dots + x_n)} = p^A \cdot (1-p)^{n-A}$$

Likelihood function

$$g(p) = \log(p^A \cdot (1-p)^{n-A}) = A \cdot \log p + (n-A) \cdot \log(1-p)$$

$$g'(p) = A \cdot \frac{1}{p} + (n-A) \cdot \frac{1}{1-p} \cdot (-1) = 0$$

$$\frac{A}{p} = \frac{n-A}{1-p}, \quad A(1-p) = (n-A) \cdot p, \quad A = n \cdot p$$

$$p = \frac{1}{n} \cdot A = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \hat{p}_{MLE}$$

maximum likelihood estimator

## Maximum Likelihood Estimators

### Definition

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution  $f(x; \theta_1, \dots, \theta_m)$ ,  $(\theta_1, \dots, \theta_m) \in \Omega$ .

The **likelihood function** of  $\theta_1, \dots, \theta_m$  is  $f(x_1, x_2, \dots, x_n; \theta_1, \dots, \theta_m)$

The **maximum likelihood estimators (MLE)** of  $\theta_1, \dots, \theta_m$  are

$\hat{\theta}_1, \dots, \hat{\theta}_m$  which maximize  $f(\downarrow)$

PMF  $\sim$  PDF

$\downarrow$

$\downarrow$

## Maximum Likelihood Estimators

**Example**  $X_1, \dots, X_n \sim \text{i.i.d. Exp}(\lambda = \frac{1}{\theta})$

Suppose  $X_1, X_2, \dots, X_n$  are independent exponential with mean  $\theta \in (0, \infty)$ .

$$\begin{aligned} \text{Likelihood function} &= \text{joint PDF} \stackrel{\text{indep.}}{=} \text{product of PDFs} \\ &= \left(\frac{1}{\theta} e^{-\frac{x_1}{\theta}}\right) \cdot \left(\frac{1}{\theta} e^{-\frac{x_2}{\theta}}\right) \cdots \left(\frac{1}{\theta} e^{-\frac{x_n}{\theta}}\right) \\ &= \theta^{-n} e^{-\frac{1}{\theta}(x_1 + \dots + x_n)} = \theta^{-n} e^{-\frac{A}{\theta}} \end{aligned}$$

$$g(\theta) = \log(\text{Likelihood function}) = -n \cdot \log \theta - \frac{A}{\theta}$$

$$g'(\theta) = -\frac{n}{\theta} + \frac{A}{\theta^2} = 0 \quad n \cdot \theta = A = (x_1 + x_2 + \dots + x_n)$$

$$\hat{\theta} = \frac{1}{n}(x_1 + \dots + x_n) \quad : \quad \underline{\underline{\text{MLE}}}$$

## Maximum Likelihood Estimators

### Example

Suppose  $X_1, X_2, \dots, X_n$  are <sup>i.i.d.</sup> normal with mean  $\theta_1 \in \mathbb{R}$  and variance  $\theta_2 \in (0, \infty)$ .

$x_1, x_2, \dots, x_n$  : given observation

$$\begin{aligned} \text{Likelihood function} &= f(x_1, x_2, \dots, x_n) \quad (\text{joint PDF}) \\ &= \left( \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_1-\theta_1)^2}{2\theta_2}} \right) \cdot \left( \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_2-\theta_1)^2}{2\theta_2}} \right) \cdots \left( \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_n-\theta_1)^2}{2\theta_2}} \right) \\ &= (2\pi\theta_2)^{-\frac{n}{2}} e^{-\frac{1}{2\theta_2} \cdot \left( \sum_{i=1}^n (x_i - \theta_1)^2 \right)} \end{aligned}$$

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

$$0 = \frac{\partial g}{\partial \theta_1} = -\frac{1}{2\theta_2} \cdot \sum_{i=1}^n 2(x_i - \theta_1) \cdot (-1) = 0$$

$$0 = \frac{\partial g}{\partial \theta_2} = -\frac{1}{2} \cdot \frac{1}{\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2 = 0$$

Maximize  $g(\theta_1, \theta_2)$

$$\sum_{i=1}^n (x_i - \theta_1) = 0$$

$$\sum_{i=1}^n x_i - n \cdot \theta_1 = 0$$

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

$$\hat{\theta}_1 = \frac{1}{n} (x_1 + \dots + x_n)$$

## Unbiased estimators

### Definition

If an estimator  $u(X_1, X_2, \dots, X_n)$  of  $\theta$  satisfies  $\mathbb{E}[u(X_1, X_2, \dots, X_n)] = \theta$ , we say the estimator is **unbiased**.

Example  $X_1, X_2, \dots, X_n \sim$  i.i.d. Normal  
mean  $\theta_1$ , var  $\theta_2$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{sample mean})$$

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\theta}_1)^2 = \frac{1}{n} \sum_{i=1}^n \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2$$

$$\mathbb{E}[\hat{\theta}_1] = \frac{1}{n} \mathbb{E}[X_1 + \dots + X_n] = \frac{1}{n} \cdot n \cdot \mathbb{E}[X_i] = \theta_1$$

$\Rightarrow \hat{\theta}_1$  is unbiased.

$$\mathbb{E}[\hat{\theta}_2] \neq \theta_2$$

$\Rightarrow \hat{\theta}_2$  is biased.

Sample variance

$$= \frac{n}{n-1} \theta_2$$

$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\theta}_1)^2$  is unbiased

$$\mathbb{E}[S^2] = \theta_2$$

## Recall

$X_1, X_2, \dots, X_n \sim \text{i.i.d.}$  with common  $\begin{cases} \text{PMF} \\ \text{PDF} \end{cases} f(x; \theta)$   
 $\Downarrow$  from these

$x_1, x_2, \dots, x_n \leftarrow$  given

Likelihood function = joint PDF/PMF

$$= f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$$

= a function of  $\theta$

$\hat{\theta}$  = Maximum Likelihood Estimator

= maximizer of likelihood function.

$$= \text{arg max } f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$$

= in terms of  $x_1, \dots, x_n$

## Unbiased estimators

### Example

Suppose  $X_1, X_2, \dots, X_n$  are normal with mean  $\theta_1 \in \mathbb{R}$  and variance  $\theta_2 \in (0, \infty)$ .



## Exercise

A random sample  $X_1, X_2, \dots, X_n$  of size  $n$  is taken from a Poisson distribution with a mean of  $\theta \in (0, \infty)$ .

Find the MLE for  $\theta$ .

## **Section 7.1.**

### **Confidence Intervals for Means**

---

Observation  $X_1, X_2, \dots, X_n$  +  $\sigma^2$  known  
 $\Rightarrow$  Estimate  $\mu$

## Two-sided confidence intervals

(i.i.d.) Consider a random sample  $X_1, X_2, \dots, X_n$  from a normal distribution  $N(\mu, \sigma^2)$ . common dis

We are interested in the closeness of the maximum likelihood estimator  $\bar{X}$  to the unknown mean  $\mu$ .

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

Suppose  $\sigma^2$  is known.

Our goal is to find an interval  $[a, b]$  such that  $P(\bar{X} \in [a, b]) = 1 - \alpha$ , where  $\alpha > 0$ .

Let  $z_{\frac{\alpha}{2}} > 0$  be such that  $P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ . (Here,  $Z$  is standard normal.)

Since  $(\bar{X} - \mu)/(\sigma/\sqrt{n})$  is standard normal, we have

$$P(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}) = 1 - \alpha.$$

$$P(a \leq \bar{X} \leq b) = 0.95$$

$$X_1, \dots, X_n \sim \text{i.i.d } N(\mu, \sigma^2)$$

$$= P\left(\frac{a - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{b - \mu}{\sigma/\sqrt{n}}\right)$$

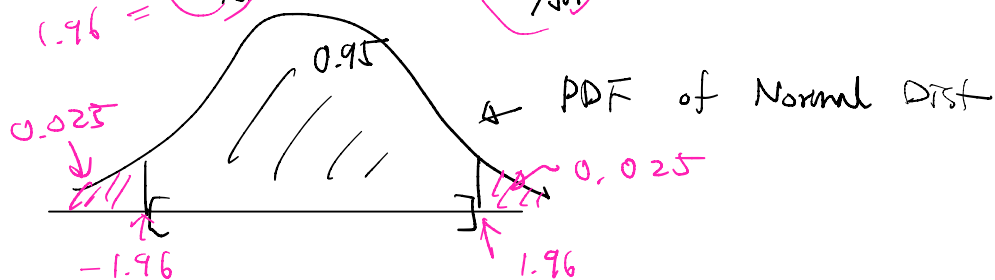
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$E[\bar{X}] = \mu$$

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

$$= \Phi\left(\frac{b - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right)$$

$= 1.96$



$$b = \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

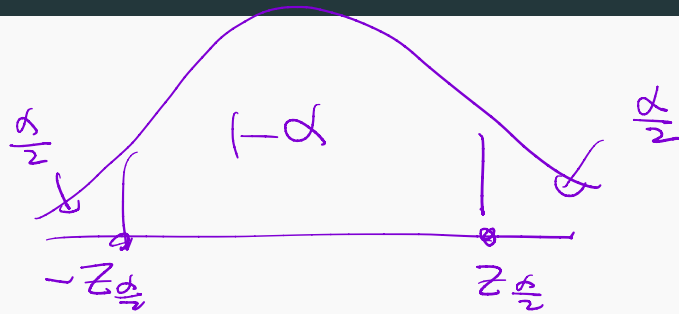
$$a = \mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow P\left(\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

With 95%,  $\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

## Two-sided confidence intervals



Solving for  $\bar{X}$ , we get

$$P\left(\mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

$$\Rightarrow \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

## Two-sided confidence intervals

### Definition

The interval  $\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$  is called the **two-sided confidence interval** for unknown  $\mu$  with **confidence coefficient**  $1 - \alpha$  (for  $100(1 - \alpha)\%$ ).

$$0.95, \quad \alpha = 0.05, \quad z_{\frac{\alpha}{2}} = 1.96$$

## Two-sided confidence intervals

### Example

Let  $X$  equal the length of life of a 60-watt light bulb marketed by a certain manufacturer.

Assume that the distribution of  $X$  is  $N(\mu, 1296)$ .

If a random sample of  $n = 27$  bulbs is tested until they burn out, yielding a sample mean of  $\bar{X} = 1478$  hours,

then what is a 95% confidence interval for  $\mu$ ?

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

$$95\% \Rightarrow \alpha = 0.05, \quad z_{\frac{\alpha}{2}} = 1.96$$

$$\sigma = 36 = \sqrt{1296}, \quad n = 27$$

$$\left[ 1478 - 1.96 \cdot \frac{36}{\sqrt{27}}, 1478 + 1.96 \cdot \frac{36}{\sqrt{27}} \right]$$

## Two-sided confidence intervals

If we **cannot assume** that the distribution from which the sample arose is **normal**, we can still obtain an approximate confidence interval for  $\mu$ .

By the central limit theorem, provided that  $n$  is large enough, the ratio  $(\bar{X} - \mu)/(\sigma/\sqrt{n})$  has the approximate normal distribution  $N(0, 1)$  **even though the underlying distribution is not normal**.

$$\text{If } X_i \sim \text{i.i.d. normal} \Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\text{If } X_i \sim \text{i.i.d. non-normal} \Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Rightarrow N(0, 1) \text{ as } n \rightarrow \infty$$

## Two-sided confidence intervals

### Example

Let  $X$  equal the amount of orange juice (in grams per day) consumed by an American. Suppose it is known that the standard deviation of  $X$  is  $\sigma = 96$ .

An orange growers' association took a random sample of  $n = 576$  Americans and found that they consumed, on average,  $\bar{X} = 133$  grams of orange juice per day.



## When $\sigma^2$ is unknown

Even though  $\sigma^2$  is unknown and the sample distribution is not normal, if  $n$  is large enough ( $n \geq 30$ ), then

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}, \quad S^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

is approximately the standard normal.

Thus, a  $100(1 - \alpha)\%$  confidence interval is

If the sample distribution is normal, then  $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$  has a student's  $t$  distribution with degree of freedom  $n - 1$ .

## When $\sigma^2$ is unknown

### Example

Let  $X$  equal the amount of butterfat in pounds produced by a typical cow during a 305-day milk production period between her first and second calves.

Assume that the distribution of  $X$  is  $N(\mu, \sigma^2)$ .

To estimate, a farmer measured the butterfat production for  $n = 20$  cows and  $\bar{X} = 507.50$  and  $S = 89.75$ .

## Exercise

A random sample of size 16 from the normal distribution  $N(\mu, 25)$  yielded  $\bar{X} = 73.8$ .  
Find a 95% confidence interval for  $\mu$ .

**Section 7.2.**  
**Confidence Intervals for the**  
**Difference of Two Means**

---

## Confidence Intervals for the Difference of Two Means

Let  $X_1, X_2, \dots, X_{n_X}$  and  $Y_1, Y_2, \dots, Y_{n_Y}$  be, respectively, two independent random samples of sizes  $n_X$  and  $n_Y$  from the two normal distributions  $N(\mu_X, \sigma_X^2)$  and  $N(\mu_Y, \sigma_Y^2)$ .

The distribution of  $W = \bar{X} - \bar{Y}$  is

The confidence interval for  $\mu_X - \mu_Y$  is

This also works when

## Confidence Intervals for the Difference of Two Means

### Example

The length of life of brand  $X$  light bulbs is assumed to be  $N(\mu_X, 784)$ .

The length of life of brand  $Y$  light bulbs is assumed to be  $N(\mu_Y, 627)$  and independent of  $X$ .

If a random sample of  $n_X = 56$  brand  $X$  light bulbs yielded a mean of  $\bar{X} = 937.4$  hours and a random sample of size  $n_Y = 57$  brand  $Y$  light bulbs yielded a mean of  $\bar{Y} = 988.9$  hours,

find a 90% confidence interval for  $\mu_X - \mu_Y$ .

## Confidence Intervals for the Difference of Two Means

Suppose  $\sigma_X^2$  and  $\sigma_Y^2$  are unknown and the sample sizes are small.

Assume that  $\sigma_X^2 = \sigma_Y^2$ .

Then,

$$T = \frac{Z}{\sqrt{U/(n_X + n_Y - 2)}}$$

has a student's  $t$  distribution of degree of freedom  $n_X + n_Y - 2$  where

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma^2/n_X + \sigma^2/n_Y}}, \quad U = \frac{(n_X - 1)S_X^2}{\sigma^2} + \frac{(n_Y - 1)S_Y^2}{\sigma^2}.$$

After simplification,  $T$  can be written as

## Confidence Intervals for the Difference of Two Means

### Example

Suppose that scores on a standardized test in mathematics taken by students from large and small high schools are  $N(\mu_X, \sigma^2)$  and  $N(\mu_Y, \sigma^2)$ , respectively, where  $\sigma$  unknown.

If a random sample of  $n_X = 9$  students from large high schools yielded  $\bar{X} = 81.31$  and  $S_X^2 = 60.76$ , and a random sample of  $n_Y = 15$  students from small high schools yielded  $\bar{Y} = 78.61$  and  $S_Y^2 = 48.24$ .

Find a 95% confidence interval for  $\mu_X - \mu_Y$ .



## **Section 7.3.**

# **Confidence Intervals for Proportions**

---

## Confidence Intervals for Proportions

### Example

In a certain political campaign, one candidate has a poll taken at random among the voting population.

The results are that  $y = 185$  out of  $n = 351$  voters favor this candidate.

Should the candidate feel very confident of winning?

## Confidence Intervals for Proportions

In general, suppose  $Y$  is a random sample from a binomial distribution with the given number of trials  $n$  and an unknown success probability  $p$ .

Note that  $Y/n$  is an unbiased point estimator for  $p$  and

$$\frac{(Y/n) - p}{\sqrt{p(1-p)/n}}$$

has an approximate normal distribution  $N(0, 1)$ .

## One-sided confidence interval

One-sided confidence intervals are sometimes appropriate for  $p$ .

For example, we may be interested in an upper bound on the proportion of defectives in manufacturing some item.

For an upper bound, we consider

$$\mathbb{P} \left( -z_{\alpha} \leq \frac{(Y/n) - p}{\sqrt{(Y/n)(1 - (Y/n))/n}} \right) \approx 1 - \alpha.$$

This gives

A lower bound is

## One-sided confidence interval

### Example

The Wisconsin Department of Natural Resources (DNR) wished to determine the prevalence,  $p$ , of chronic wasting disease (a neurological disease similar to mad cow disease) among its whitetail deer population.

In a particular area of the state, 272 deer were legally killed by hunters in a particular season. A tissue sample from each animal was submitted to the DNR.

Laboratory analysis determined that nine of the deer had chronic wasting disease.

Assume that the 272 harvested deer can, to a good approximation, be regarded as a random sample.

Find a one-sided approximate 95% confidence interval that provides an upper bound for  $p$  in that area of Wisconsin.

## Difference of two proportions

Suppose  $Y_1$  and  $Y_2$  are independent binomial random variables with  $n_1, n_2$  and  $p_1, p_2$ .

To make a statistical inference about the difference  $p_1 - p_2$ , we consider

$$Y_1/n_1 - Y_2/n_2.$$

Its mean is

Its variance is

So the CLT gives

## Difference of two proportions

### Example

Two detergents were tested for their ability to remove stains of a certain type.

An inspector judged the first one to be successful on 63 out of 91 independent trials and the second one to be successful on 42 out of 79 independent trials.

The respective relative frequencies of success are 0.692 and 0.532.

What is an approximate 90% confidence interval for the difference  $p_1 - p_2$  of the two detergents?

## Exercise

A machine shop manufactures toggle levers.

A lever is flawed if a standard nut cannot be screwed onto the threads.

Let  $p$  equal the proportion of flawed toggle levers that the shop manufactures.

There were 24 flawed levers out of a sample of 642 that were selected randomly from the production line.

Find an approximate 95% confidence interval for  $p$ .



