Regression Analysis

- **Regression analysis** is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.
- In order to understand regression analysis fully, it is essential to comprehend the following terms:
  - **Dependent Variable:** This is the main factor that you are trying to understand or predict.
  - **Independent Variables:** These are the factors that you hypothesize have an impact on your dependent variable.
- Suppose we need to study the influence of your math and biology grades in high school on your first year university grades in math, physics, and chemistry. In this example, the independent variables are the high school math and biology grades, and the dependent variables are the university grades in math, physics, and chemistry.
- To answer the question, we need data. The data, in our example is obtained from the registrar's office. In other cases, it is collected through surveys. The data, thus collected, is called a **random sample.** The random sample is analyzed and conclusions drawn are generalized on the population.
- Some other examples involving two variables are
  - The weight of a newly born child and the age of pregnancy
  - The sell of ice-cream and the weather temperature
  - Your GPA and study hours per week.
  - Cholesterol levels and heart attacks
  - Gas prices and distances traveled by drivers.
- In this lecture, we consider only two variables; X the independent and Y the dependent.

Basic Definitions and Terminology

- First, let us introduce some basic definitions about the random sample

  The ***sample mean*** $\hat{\mu}_X$ is defined as $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i$

- The ***sample variance*** $\hat{\sigma}_X^2$, when the population mean μ is unknown is defined as:

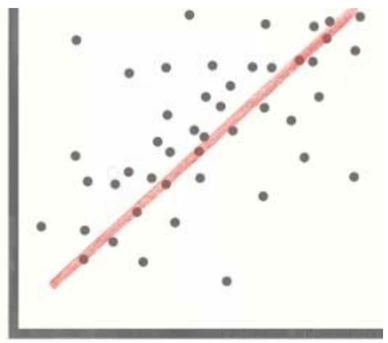
$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} \; ; \; \hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2}$$

is the *sample standard deviation.*

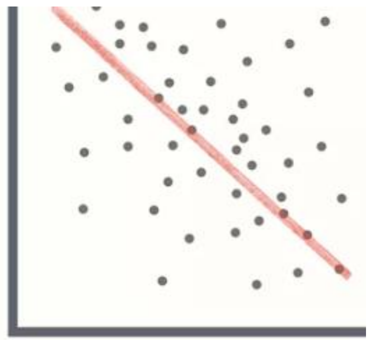
- The correlation coefficient between sampled measurements x and y is

$$\rho_{X,Y} = \frac{C_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{1}{(n-1)\hat{\sigma}_X \hat{\sigma}_Y}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y) \; ; \; C_{XY} \text{ sample covariance}$$

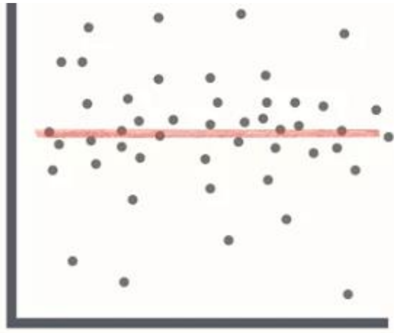
- The correlation coefficient between two random variables (X) and (Y) is a measure of association between X and Y.
- ρ_{XY} is bounded between $-1 \le \rho_{XY} \le 1$. **The magnitude of the correlation coefficient indicates the strength of the association.** For example, a correlation of $\rho_{XY} = 0.9$ suggests a strong, positive association between two variables, whereas a correlation of $\rho_{XY} = -0.2$ suggest a weak, negative association. A correlation close to zero suggests no linear association between two variables.



Positive Correlation



Negative Correlation



No Correlation

Linear Regression

Suppose in a certain experiment we take measurements in pairs, i.e. $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. We suspect that the data can fit a straight line of the form $y = \alpha x + \beta$.

Suppose that the line is to be fitted to the (n) points and let ($\in$) denote the sum of the squares of the vertical distances at the (n) points, then

$$\in \;=\; \sum_{i=1}^{n}\left[y_i - (\alpha x_i + \beta)\right]^2$$

The method of least squares specifies the values of α and β that minimize $\in$.

$$\frac{\partial \in}{\partial \alpha} = -2\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)\,x_i = 0 \;\Rightarrow\; \beta\sum_{i=1}^{n}x_i + \alpha\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}x_i y_i$$

$$\frac{\partial \in}{\partial \beta} = -2\sum_{i=1}^{n}(y_i - \alpha x_i - \beta) = 0 \;\Rightarrow\; n\beta + \alpha\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}y_i$$

In matrix form, these equations are:

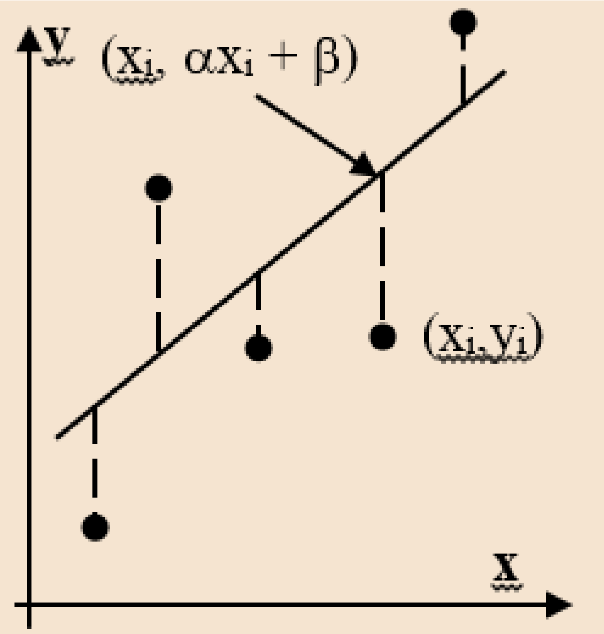
$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}\begin{pmatrix}\beta \\ \alpha\end{pmatrix} = \begin{pmatrix}\sum y_i \\ \sum x_i y_i\end{pmatrix}$$; These two equations are called the *normal equations*.

Solving the above two equations for the two unknowns, we get:

$$\alpha = \frac{C_{XY}}{\hat{\sigma}_X^2} = \frac{1}{(n-1)\hat{\sigma}_X^2}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)\,(y_i - \hat{\mu}_Y)\,, \qquad \beta = \hat{\mu}_Y - \alpha\hat{\mu}_X$$

where C_{XY}, is the sample covariance between x and y, $\hat{\sigma}_X^2$ is the sample variance of the X measurements (as defined earlier), $\hat{\mu}_X$ is the average value of the X measurements, and $\hat{\mu}_Y$ is the average value of the Y measurements.

Finally, the sample correlation coefficient can be calculated as $\rho_{X,Y} = \dfrac{C_{XY}}{\hat{\sigma}_X\hat{\sigma}_Y}$



x_i	$(x_i)^2$	y_i	$x_i y_i$
x_1	$(x_1)^2$	y_1	$x_1 y_1$
x_2	$(x_2)^2$	y_2	$x_2 y_2$
x_3	$(3)^2$	3	$x_3 y_3$
.	.	.	.
x_n	$(x_n)^2$	y_n	$x_n y_n$
$\sum x_i$	$\sum (x_i)^2$	$\sum y_i$	$\sum x_i y_i$

Fitting a Polynomial by the Method of Least Squares:

- Suppose now that instead of simply fitting a straight line to (n) plotted points, we wish to fit a polynomial of the form:

$$y = \beta_1 + \beta_2 x + \beta_3 x^2$$

- The method of least squares specifies the constants β_1, β_2 and β_3 so that the sum of the squares of errors $\in$ is minimized.

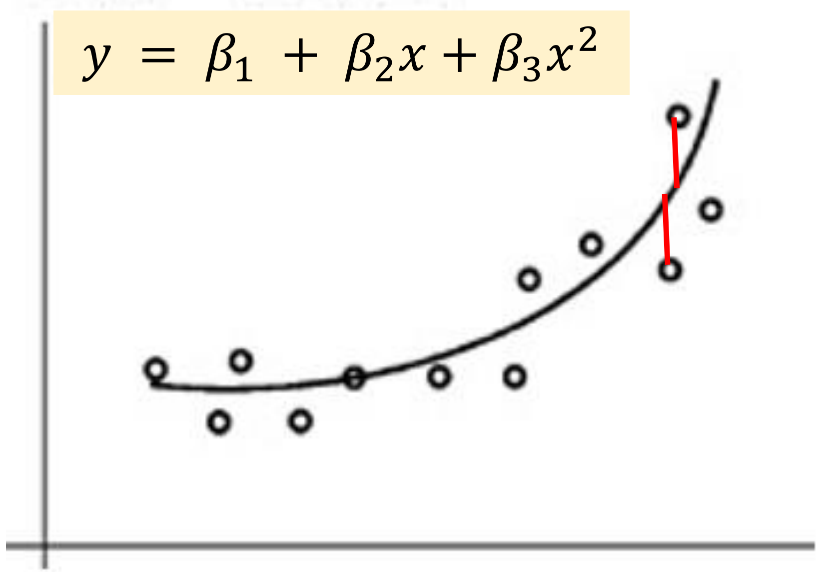
$$\in = \sum_{i=1}^{n}\left[y_i - (\beta_1 + \beta_2 x_i + \beta_3 x_i^2)\right]^2$$

- Taking partial derivatives of $\in$ with respect to β_1, β_2 and β_3, setting the derivative to zero and solving, we get the following set of normal equations

$$\beta_1 n + \beta_2 \sum_{i=1}^{n} x_i + \beta_3 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \quad \text{.........} (1)$$

$$\beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^2 + \beta_3 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i \quad \text{.........} (2)$$

$$\beta_1 \sum_{i=1}^{n} x_i^2 + \beta_2 \sum_{i=1}^{n} x_i^3 + \beta_3 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i \quad \text{.........} (3)$$



- In matrix form, these equations are:

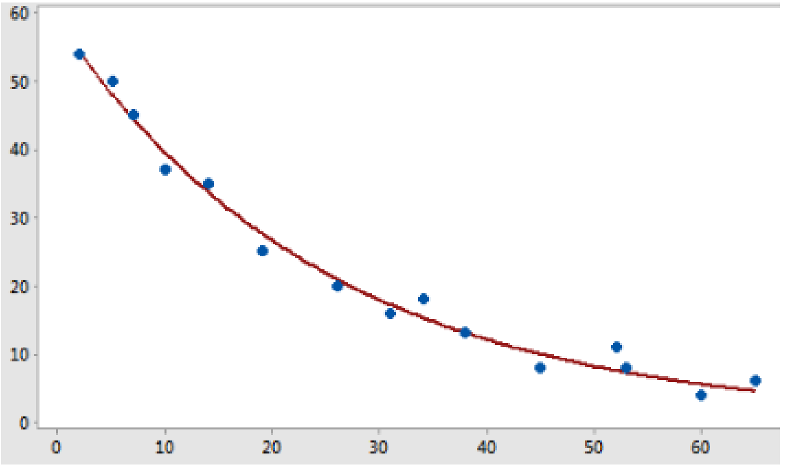
$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix}$$

- Then these equations can be solved, simultaneously for β_1, β_2 and β_3 .

Fitting an Exponential by the Method of Least Squares:

- Suppose that we suspect the data to fit an exponential equation of the form:

$$y = a\,e^{bx} \qquad \ldots\ldots\ldots (1)$$



- Taking the natural logarithm of (1)

$$\ln(y) = \ln(a) + \ln(e^{bx})$$

$$\ln(y) = \ln(a) + b\,x$$

$$y' = \beta' + \alpha'\,x \qquad \ldots\ldots\ldots (2)$$

- As we can see, equation (2) has the same form of the linear regression considered earlier where $y = \beta + \alpha x$. Hence, the solution involves the following steps
  - Take the natural logarithm of each measurement y_i.
  - The new pairs of the data now become $(x_1, \ln y_1), (x_2, \ln y_2), \ldots (x_n, \ln y_n)$.
  - Solve this regression model for β' and α'.
  - $\beta' = \ln(a) \Rightarrow a = e^{\beta'}$
  - $\alpha' = b$

More Regression Models

EXAMPLE: Suppose that the polynomial to be fitted to a set of (n) points is y = b x. It can be shown that:

$$b = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2$$

$$y = \alpha x + \beta$$

EXAMPLE: Let $y = ax^b$.

Taking the ln of both sides, we get ln y = ln a + b ln x, hence transformed into the linear model

y' = β' + α' x' *(Linear regression)*

where: y' = ln y , β' = ln a , α' = b , x' = ln x

EXAMPLE: Let $y = 1 - e^{\frac{-x^b}{a}}$

Manipulation of this equation yields: $\ln\left[\ln\left(\frac{1}{1-y}\right)\right] = -\ln a + b \ln x$

which is in the standard form: y' = β' + α' x' *(Linear regression)*

EXAMPLE: Let $y = \frac{L}{1 + e^{a+bx}}$.

$\ln\left(\frac{L - y}{y}\right) = a + b\,x$, which is in the standard linear form: y' = β' + α' x' *(Linear regression)*

EXAMPLE: The cumulative number of coronavirus cases recorded in a certain city over a 10-day period is shown in the table.

a. Assuming that a simple linear regression model is appropriate, fit the regression model relating the number of coronavirus cases (y) to the time in days (x).

b. What is the expected number of cases y on the 20'th day?

c. Find the correlation coefficient between x and y.

Day	Number of Coronavirus cases
1	60
2	72
3	84
3	91
5	97
6	106
7	115
8	117
9	134
10	161

Solution:

a. The linear regression model to be fit is $y = \beta + \alpha x$

$$\text{Here, } \sum_{i=1}^{10} x_i = 55, \ \sum_{i=1}^{12} x_i^2 = 385, \ \sum_{i=1}^{12} x_i y_i = 6498, \ \sum_{i=1}^{12} y_i = 1037$$

The equation parameters are given by: α = 9.6303, β = 50.7333. y = 50.7333+9.6303x

b. After 20 days, the linear model predicts a number at: y = 50.7333+9.6303x (20) = 244.

c. The correlation coefficient between the x and y data is

$$\rho_{X,Y} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)\hat{\sigma}_X \hat{\sigma}_Y} = 0.977.$$

EXAMPLE: The number of coronavirus cases recorded in a certain city over a 10-day period is shown in the table below.

Day	Number of Coronavirus cases
1	599
2	673
3	784
4	833
5	1001
6	1169
7	1328
8	1382
9	1557
10	1815

a. Fit the regression model relating the number of coronavirus cases (y) as a function of time in days (x) using a linear, a quadratic, and an exponential model.

b. What is the expected number of cases y on the 20'th day using each model?

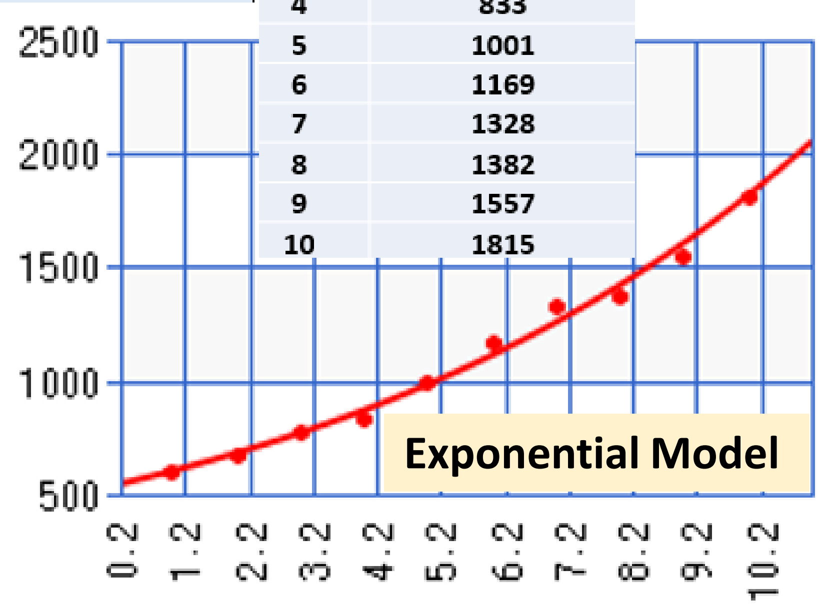
c. Use each one of the model to predict the number of cases on day 10.

Solution:

a. The linear regression model to be fit is $y = \alpha x + \beta = 131.9697x + 388.2667$

The quadratic model is: $y = \beta_1 + \beta_2 x + \beta_3 x^2 = 522.8500 + 64.6780x + 6.1174x^2$

The exponential model is: $y = a\, e^{bx} = 533.779e^{0.122645x}$

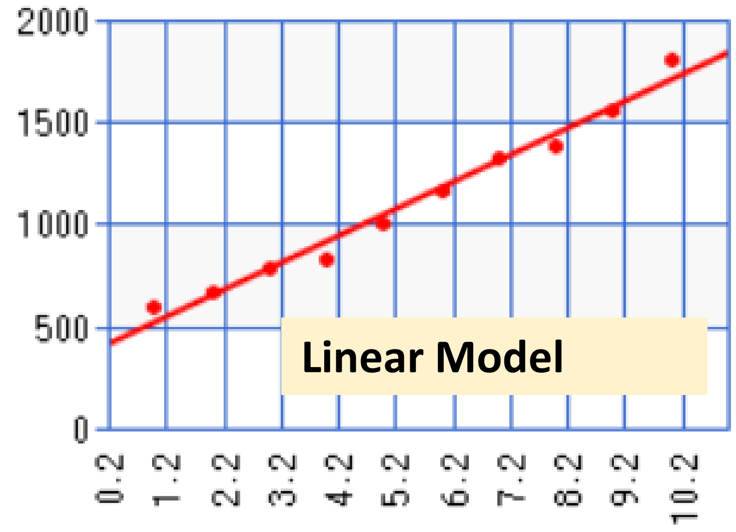


b. After 20 days,

the linear model prediction: $y = 131.9697(20) + 388.2667 = 3028$

the quadratic model prediction: $y = 522.8500 + 64.6780(20) + 6.1174(20)^2 = 4260$

the exponential model prediction: $y = a\, e^{bx} = 533.779e^{0.122645(20)} = 6203$



c. On day 10,

the linear model prediction: $y = 131.9697(10) + 388.2667 = 1707$

the quadratic model prediction: $y = 522.8500 + 64.6780(10) + 6.1174(10)^2 = 1780$

the exponential model prediction: $y = 533.779e^{0.122645(10)} = 1820$

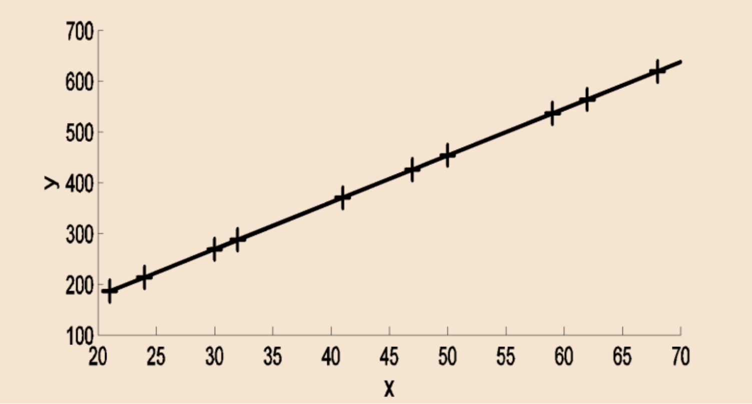
EXAMPLE: The number of pounds of steam used per month by a chemical plant is thought to be related to the average ambient temperature (in ˚F) for that month. The past year's usage and temperature are shown in the following table

Month	Temp.	Usage	Month	Temp.	Usage
Jan.	21	185	July	68	621
Feb.	24	214	Aug.	74	675
Mar.	32	288	Sept.	62	562
Apr.	47	424	Oct.	50	452
May	50	454	Nov.	41	373
June	59	539	Dec.	30	273

a. Assuming that a simple linear regression model is appropriate, fit the regression model relating steam usage (y) to the average temperature (x).

b. What is the expected usage when the average temperature is 55 F

c. Find the correlation coefficient between x and y.



Solution:

a. The linear regression model to be fit is $y = \alpha x + \beta$

Here, $\sum_{i=1}^{12} x_i = 558$, $\sum_{i=1}^{12} x_i^2 = 29256$, $\sum_{i=1}^{12} x_i y_i = 265771$, $\sum_{i=1}^{12} y_i = 5060$

The equation parameters are given by: α = 9.2182, β = - 7.3126. The minimum value of the mean square error calculated using $\text{MMSE} = \sum_{i=1}^{n} [y_i - (\alpha x_i + \beta)]^2 = 38.1315$.

b. when the temperature is 55 F˚, the linear model predicts a usage of y=9.2182*55 -7.3126 = 499.69. (Note that this temperature is not one of those that appear in the table, yet the model can predict the usage at this temperature).

c. The correlation coefficient between the x and y data is:

$$\rho_{X,Y} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)\hat{\sigma}_X \hat{\sigma}_Y} = 0.9999$$

This is very close to 1 meaning that the data are highly correlated (we know that when y is linearly related to x, $\rho_{X,Y} = 1$

Jointly Gaussian Random Variables

Theorem: Let X_1 and X_2 be two jointly Gaussian random variables. Define a linear transformation of the form $\begin{matrix} Y_1 = a_1 X_1 + a_2 X_2 \\ Y_2 = b_1 X_1 + b_2 X_2 \end{matrix}$. The new random variables Y_1 and Y_2 are jointly Gaussian.

Proof: The joint pdf of Y_1 and Y_2 can be determined as (discussed in the previous chapter)

$$f_{y1,y2}(y_1\ ,y) = \frac{f_{x1,x2}(x_1\ ,x_2)}{|J|} \quad \mathbf{(1)}$$

Note that $J = \left| \frac{\partial y_1\ \partial y_2}{\partial x_1\ \partial x_2} \right| = \begin{vmatrix} \frac{\partial y_1}{dx_1} & \frac{\partial y_1}{dx_2} \\ \frac{\partial y_2}{dx_1} & \frac{\partial y_2}{dx_2} \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix};\ ,\ |J| > 0$

Since X_1 and X_2 are jointly Gaussian, and since J is a constant (J > 0), then from (1), Y_1 and Y_2 are jointly Gaussian. The marginal pdf's are evaluated from the joint pdf

$$f_{y1}(y_1) = \int_{-\infty}^{\infty} f_{y1,y2}(y_1\ ,y_2)\, dy_2 \qquad f_{y2}(y_2) = \int_{-\infty}^{\infty} f_{y1,y2}(y_1\ ,y_2)\, dy_1$$

These marginal pdf's are Gaussian

Therefore, any linear combination of Gaussian random variables is Gaussian.

Linear Transformation of a Single Gaussian Random Variable

EXAMPLE: The profit Y of a manufacturing plant is related to the demand X by the relationship $Y = aX + b$. Let X be a Gaussian r.v with mean μ_X variance σ_X^2. Find $f_Y(y)$.

SOLUTION: $Y = aX + b$ is a monotonic function.

$$f_Y(y) = \frac{f_X(x)}{|dy/dx|}; \qquad \left|\frac{dy}{dx}\right| = |a|; \qquad x = \frac{y-b}{a}|$$

$$f_Y(y) = \frac{1}{a} \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{\frac{-(x-\mu_X)^2}{2\sigma_X^2}} = \frac{1}{\sqrt{2\pi(a\sigma_X)^2}} e^{\frac{-(\frac{y-b}{a}-\mu_X)^2}{2\sigma_X^2}}$$

$$= \frac{1}{\sqrt{2\pi(a\sigma_X)^2}} e^{\frac{-(y-(b+a\mu_X))^2}{2(a\sigma_X)^2}} = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{\frac{-(y-\mu_Y)^2}{2\sigma_Y^2}}$$

Therefore, $Y = aX + b$ **is Gaussian with mean** $\mu_Y = a\mu_x + b$ **and variance** $\sigma_Y^2 = a^2\sigma_X^2$

Result: A linear transformation of a Gaussian random variable is also Gaussian

Result: If $Y = a_1X_1 + a_2X_2$, then $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{\frac{-(y-\mu_y)^2}{2\sigma_Y^2}}$; (X_1,X_2 jointly Gaussian)

Where, $\mu_Y = a_1\mu_1 + a_2\mu_2$

$$\sigma_Y^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + 2a_1a_2\sigma_{X_1}\sigma_{X_2}\rho_{X_1X_2}$$

When the random variables are uncorrelated or independent, the second term becomes zero.

Proof: Here, we re-derive the variance of Y, considered in an earlier lecture

$$\mu_Y = a_1\mu_{X_1} + a_2\mu_{X_2}$$

$$\sigma_Y^2 = E\{(Y - \mu_Y)^2\} = E\{(a_1X_1 + a_2X_2 - a_1\mu_{X_1} - a_2\mu_{X_2})^2\}$$

$$= E\{[a_1(X_1 - \mu_{X_1}) + a_2(X_2 - \mu_{X_2})]^2\}$$

$$= E\{a_1^2(X_1 - \mu_{X_1})^2\} + E\{a_2^2(X_2 - \mu_{X_2})^2\} + 2a_1a_2E\{(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})\}$$

$$\Rightarrow \sigma_Y^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + 2a_1a_2\sigma_{X_1}\sigma_{X_2}\rho_{X_1X_2}$$

Where, $\rho_{X_1X_2} = \frac{E\{(X_1 - \mu_{X_1})((X_2 - \mu_{X2}))\}}{\sigma_{X_1}\sigma_{X_2}}$ is the correlation coefficient.

A similar result can be obtained when Y is a linear function of more than two Gaussian random variables.

Remark: For a Gaussian random variable X with mean μ_X and variance σ_X^2, we recall the following two results when evaluating probabilities:

$$P(X \le x_0) = \Phi\left(\frac{x_0 - \mu_X}{\sigma_X}\right), \qquad P(x_0 \le X \le x_1) = \Phi\left(\frac{x_1 - \mu_X}{\sigma_X}\right) - \Phi\left(\frac{x_0 - \mu_X}{\sigma_X}\right)$$

EXAMPLE: Let X_1 and X_2 be two Gaussian random variables such that: $\mu_1 = 0$, $\sigma_1^2 = 4$, $\mu_2 = 10$, $\sigma_2^2 = 9$, $\rho_{1,2} = 0.25$. Define $Y = 2X_1 + 3X_2$

a. Find the mean and variance of Y
b. Find $P(Y \le 35)$.

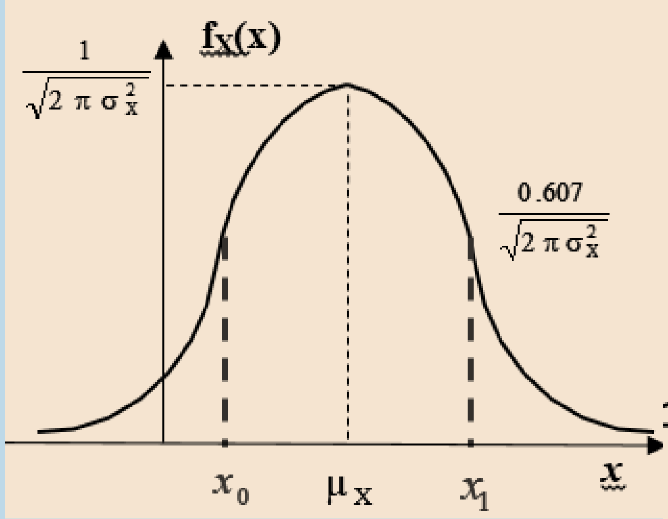
SOLUTION:

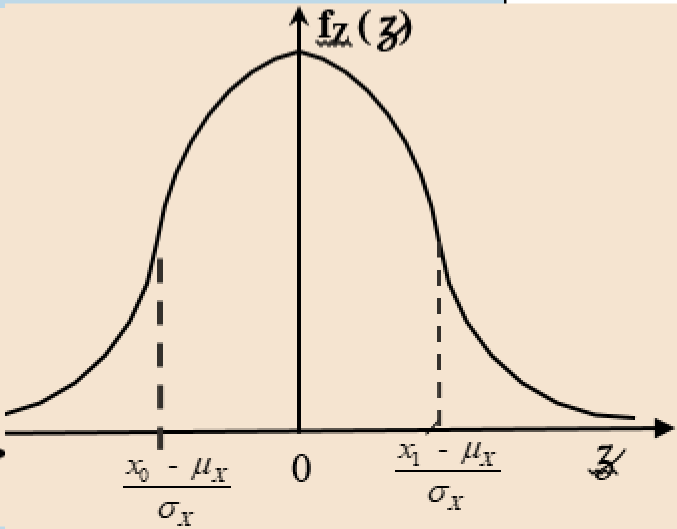
$$\mu_Y = 2\mu_1 + 3\mu_2 = 2(0) + 3(10) = 30$$

$$\sigma_Y^2 = 4\sigma_1^2 + 9\sigma_2^2 + 2(2)(3)(\sigma_1)(\sigma_2)\rho_{1,2}$$

$$= 4(4) + 9(9) + 2(2)(3)(2)(3)(0.25) = 115$$

$$P(Y < 35) = \Phi\left(\frac{35 - 30}{\sqrt{115}}\right) = \Phi(0.466) = 0.6794$$





Theorem: Let $X_1, X_2, \ldots, X_n$ be a sequence of independent Gaussian random variables, each with mean μ_i and variance σ_i^2. Define

$$Y = C_1X_1 + C_2X_2 + \cdots + C_nX_n,\ C_1,\ C_2, C_n \text{ are constants}$$

Then Y has a Gaussian distribution with mean and variance given by:

$$\mu_Y = C_1\mu_1 + C_2\mu_2 + \ldots\ldots\ldots + C_n\mu_n$$
$$\sigma_Y^2 = C_1^2\sigma_1^2 + C_2^2\sigma_2^2 + \ldots + C_n^2\sigma_n^2$$

Recall that when the random variables are independent, then they are uncorrelated. Meaning that the correlation coefficients are zero.

$$Y = a_1X_1 + a_2X_2$$

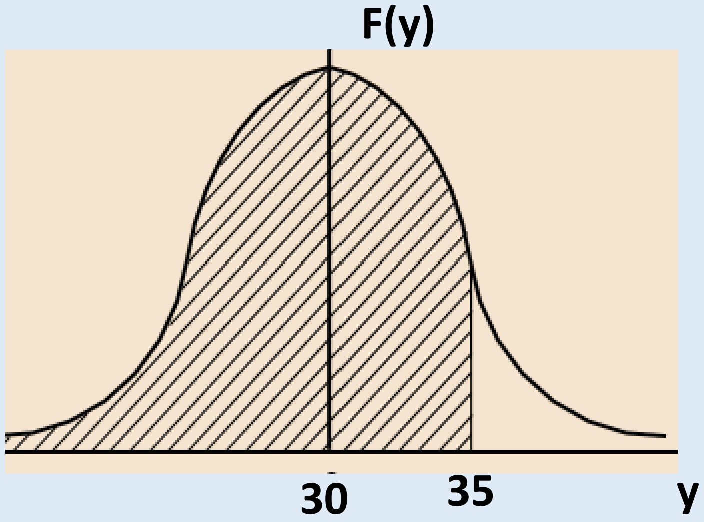
$$\sigma_Y^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + 2\,a_1a_2\sigma_{X_1}\sigma_{X2}\rho_{X_1X_2}$$

EXAMPLE: Let X_1 and X_2 be two independent Gaussian random variables such that: $\mu_1 = 0$, $\sigma_1^2 = 4$, $\mu_2 = 10$, $\sigma_2^2 = 9$. Define $Y = 2X_1 + 3X_2$

c. Find the mean and variance of Y
d. Find $P(Y \leq 35)$.

SOLUTION:

$$\mu_Y = 2\mu_1 + 3\mu_2 = 2(0) + 3(10) = 30$$
$$\sigma_Y^2 = 4\sigma_1^2 + 9\sigma_2^2 = 4(4) + 9(9) = 97$$
$$P(Y < 35) = \Phi\left(\frac{35-30}{\sqrt{97}}\right) = \Phi(0.5077) = 0.6942.$$



Theorem: Let $X_1, X_2, \ldots, X_n$ be a sequence of independent Gaussian random variables, each with mean μ and variance σ^2 (iid). Define the sample mean (sample average) as $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Then, $\hat{\mu}$ has a Gaussian distribution with mean and variance given by:

$$E(\hat{\mu}) = \mu\,,\quad Var(\hat{\mu}) = \sigma^2/n\,.$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

Proof: Rewrite $\hat{\mu}$ in the form:

$$\hat{\mu} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$$

which has the form: $Y = C_1X_1 + C_2X_2 + \cdots + C_nX_n$, where $X_1, X_2, \ldots, X_n$ are iid Gaussian r.v's.

The mean and variance of $\hat{\mu}$ are:

$$E(\hat{\mu}) = \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu = \mu$$

$$\sigma_Y^2 = \left(\frac{1}{n}\right)^2\sigma^2 + \left(\frac{1}{n}\right)^2\sigma^2 + \ldots + \left(\frac{1}{n}\right)^2\sigma^2 = \frac{\sigma^2}{n}$$

Lemma: If $X_1, X_2, \ldots, X_n$ are a sequence of independent Gaussian random variables, each with mean μ and variance σ^2, then $Z = \dfrac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}$ is the standard Gaussian variable with mean zero and variance one.

EXAMPLE: The weights of cement bags are normally distributed with a mean of (50) kg and a standard deviation of 2 kg.

a. What is the probability that one randomly selected cement bag will weigh more than 51 kg?

b. What is the probability that 5 randomly selected cement bags will have a mean weight of more than 51 kg

c. Find n, such that the probability that the mean weight of n randomly selected cement bags be larger than 51 kg is less than 0.01.

SOLUTION:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

a. $P(X > 51) = 1 - \Phi(\frac{51-50}{2}) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085$

b. Sample average: $\hat{\mu} = (X_1 + X_2 + ... X_5)/5$; is a random variable with mean and variance

$$E(\hat{\mu}) = \frac{1}{n}\mu + \frac{1}{n}\mu + ... + \frac{1}{n}\mu = \mu = 50,\ Var(\hat{\mu}) = \sigma^2/n = (2)^2/5 = 0.8$$

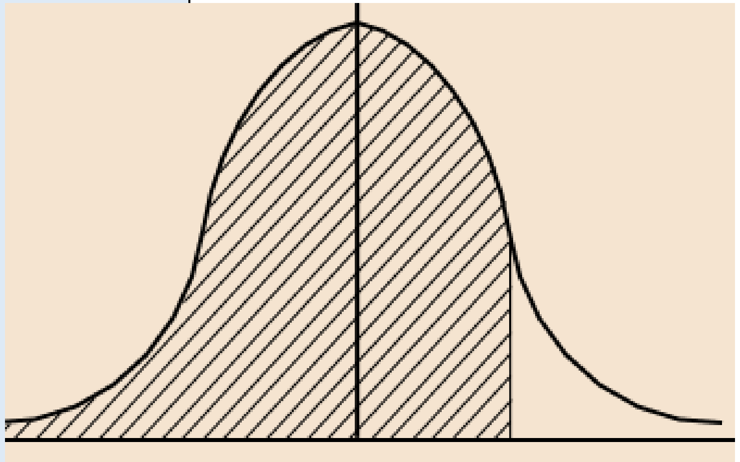
$$P(\hat{\mu} > 51) = 1 - \Phi(\frac{51-50}{\sqrt{0.8}}) = 1 - \Phi(1.118) = 1 - 0.8682 = 0.1318$$

c. $\hat{\mu} = (X_1 + X_2 + ... X_n)/\mathrm{n}$ is a random variable with mean and variance

$$E(\hat{\mu}) = \frac{1}{n}\mu + \frac{1}{n}\mu + ... + \frac{1}{n}\mu = \mu = 50\ ,\ Var(\hat{\mu}) = \sigma^2/n = (2)^2/n = 4/n$$

Need to find n such that $P(\hat{\mu} > 51) = 1 - \Phi(\frac{51-50}{\sqrt{4/n}}) < 0.01 \Rightarrow \Phi(\frac{51-50}{\sqrt{4/n}}) > 0.99$

From the Tables, we get $\Phi(u) = 0.99 \Rightarrow u = 2.3263 = \frac{51-50}{\sqrt{4/n}} \Rightarrow \mathrm{n} \geq 22$



$\mathbf{E}(\hat{\boldsymbol{\mu}}) = \mu,$

EXAMPLE: The monthly rent of a two-bedroom apartment in the city of Ramallah is a random variable, X, that follows the Gaussian distribution with a mean of $ 600 and standard deviation of $ 50. The monthly rent of a similar apartment in the neighbouring city of Al-Birah is also a random variable, Y that follows the Gaussian distribution with a mean of $ 500 and a standard deviation of $ 80. If the number of available rental apartments in Al-Birah is double than that in Ramallah. Find the probability of renting an apartment with a rent less than $ 540.

SOLUTION:

Let R be the monthly rent (irrespective of the city), then

R=P(selecting Ramallah)(R|Ramallah)+P(selecting Birah)(R|Birah) $\Rightarrow R = \frac{1}{3}X + \frac{2}{3}Y$.

R is a Gaussian random variable with mean and variance

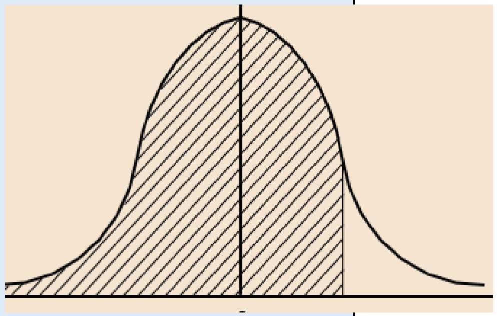
P=1/3
Ramallah
$X \sim N(600, 50^2)$

P=2/3
Birah
$Y \sim N(500, 80^2)$

$$\mu_R = \frac{1}{3}\mu_X + \frac{2}{3}\mu_Y = \frac{1}{3}(600) + \frac{2}{3}(500) = \$533.33$$

$$\sigma_R^2 = \frac{1}{9}\sigma_X^2 + \frac{4}{9}\sigma_Y^2 = \frac{1}{9}(50)^2 + \frac{4}{9}(80)^2 = 3122.22 \Rightarrow \sigma_R = \$55.87$$

$$P(\mathrm{R} < 540) = \Phi\left(\frac{540 - \mu_R}{\sigma_R}\right) = \Phi\left(\frac{540 - 533.33}{55.87}\right) = \Phi(0.1193) = 0.5475$$



EXAMPLE: Soft-drink cans are filled by an automated filling machine. The mean fill volume is 330 ml and the standard deviation is 1.5 ml. Assume that the fill volumes of the cans are independent Gaussian random variables. What is the probability that the average volume of 10 cans selected at random from this process is less than 328 ml.

SOLUTION: Average fill Volume: $\hat{\mu}=(X_1+X_2+...X_n)/n$; this quantity is a random variable.

Mean and variance of $\hat{\mu}$

$$E(\hat{\mu}) = \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu = \mu \qquad Var(\hat{\mu}) = \sigma^2/n = (1.5)^2/10 = 0.225$$

The random variable $\hat{\mu}$ is Gaussian with mean 330 and variance 0.225.

$$P(\hat{\mu} < 328) = \Phi\left(\frac{328-330}{\sqrt{0.225}}\right) = \Phi(-4.21) = 1.27 * 10^{-5}$$

The Central Limit Theorem

Main result from previous lecture.

Theorem: Let $X_1, X_2, \ldots, X_n$ be a sequence of independent Gaussian random variables, each with mean μ and variance σ^2 (iid). Define the sample mean (sample average) as

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

Then, $\hat{\mu}$ has a Gaussian distribution with mean and variance given by:

$$E(\hat{\mu}) = \mu \quad Var(\hat{\mu}) = \sigma^2/\text{n}$$

$$\boldsymbol{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = N(\mu_X, \sigma_X^2/n)}\textbf{, for any n}$$

The Central Limit Theorem:

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables, each with mean μ_x and variance σ_x^2, then the sample mean defined as: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **approaches a normal distribution** as $n \to \infty$, with mean and variance given by: $E\{\hat{\mu}\} = \mu_x$, $Var(\hat{\mu}) = \sigma_x^2/n$. That is, the limiting form of the distribution of: $Z = \dfrac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}}$ $as\ n \to \infty$, is the standard normal distribution.

- In many cases of practical interest, if $n \geq 30$, the normal approximation will be satisfactory regardless of the shape of the population or the nature of the distribution (discrete or continuous).
- The theorem works well even for small samples $n = 4$, $n=5$, when the population has a continuous distribution as illustrated in the following example.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \to N(\mu_X, \sigma_X^2/n)\ as\ n \to \infty$$

EXAMPLE: Let $X_1,\ X_2,\ X_n$ be three independent uniform random variables over the interval (0, 1). Find and sketch the pdf of $Y = X_1 + X_2 + X_3$.

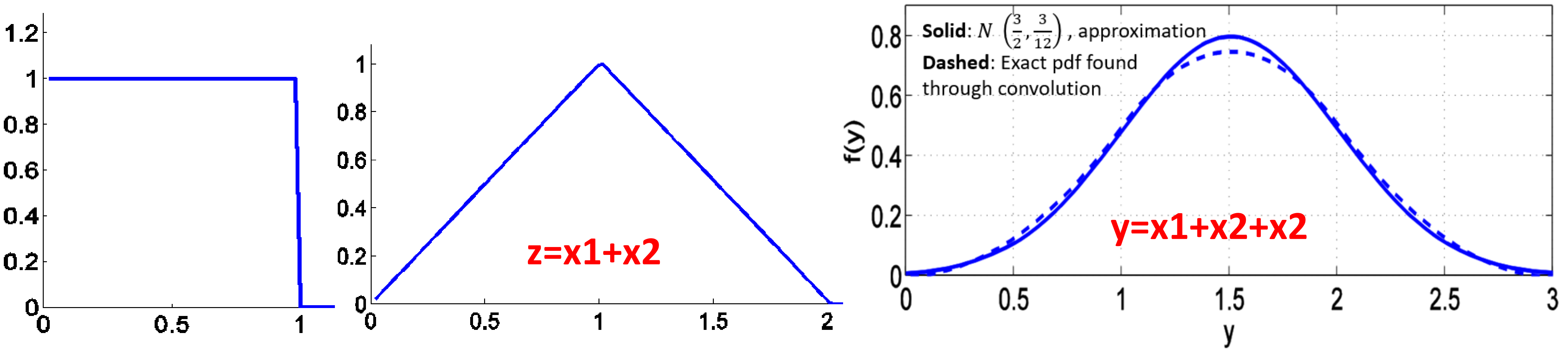
SOLUTION: First, we find the pdf of Z=X_1+ X_2 by convolving the pdf of X_1 with that for X_2. Then, we convolve the pdf of Z with that of X_3. The result is:

$$f_Y(y) = \begin{cases} 0 & y \le 0 \\ y^2/2 & 0 < y \le 1 \\ 3y - y^2 - 3/2 & 1 < y \le 2 \\ (3-y)^2/2 & 2 < y \le 3 \\ 0 & y > 3 \end{cases}$$

Mean and Variance of Y: $Y = X_1 + X_2 + X_3$

The mean Y is: $\mu_Y = 3E(X) = 3\left(\frac{a+b}{2}\right) = 3\left(\frac{0+1}{2}\right) = \frac{3}{2}$,

$$Var(Y) = \sigma_Y^2 = 3(\sigma_x)^2 = 3\frac{(b-a)^2}{12} = 3\frac{(1-0)^2}{12} = \frac{3}{12}$$



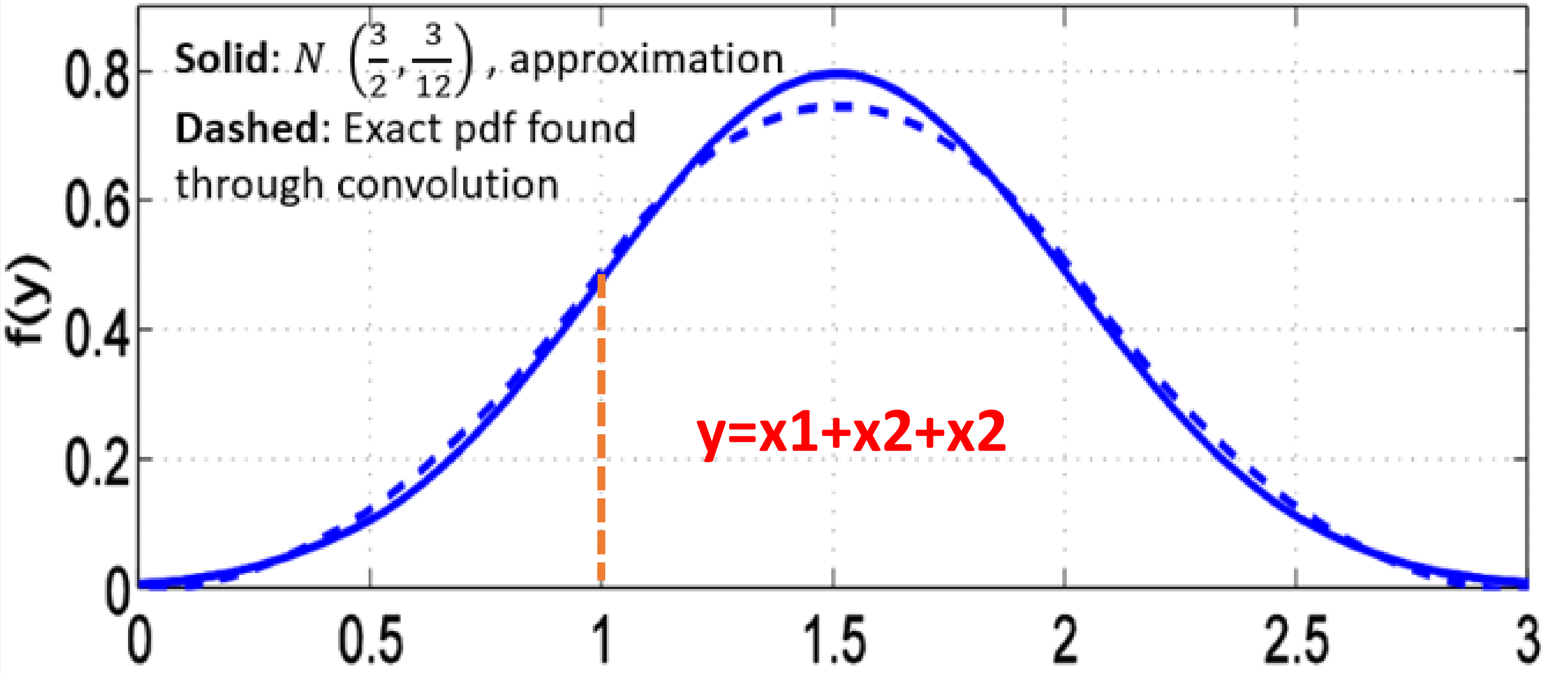
In the figure below we plot the pdf's of X_1, $Z = X_1 + X_2$ and $Y = X_1 + X_2 + X_3$. In addition, for the sake of comparison, we plot the pdf of a Gaussian distribution with the same mean 3/2 and variance 3/12 as that of Y. It is very clear that even for n=3, f(y) is very close to the Gaussian curve.
Now let us calculate P(0≤ Y ≤ 1) using the exact formula and the approximation.

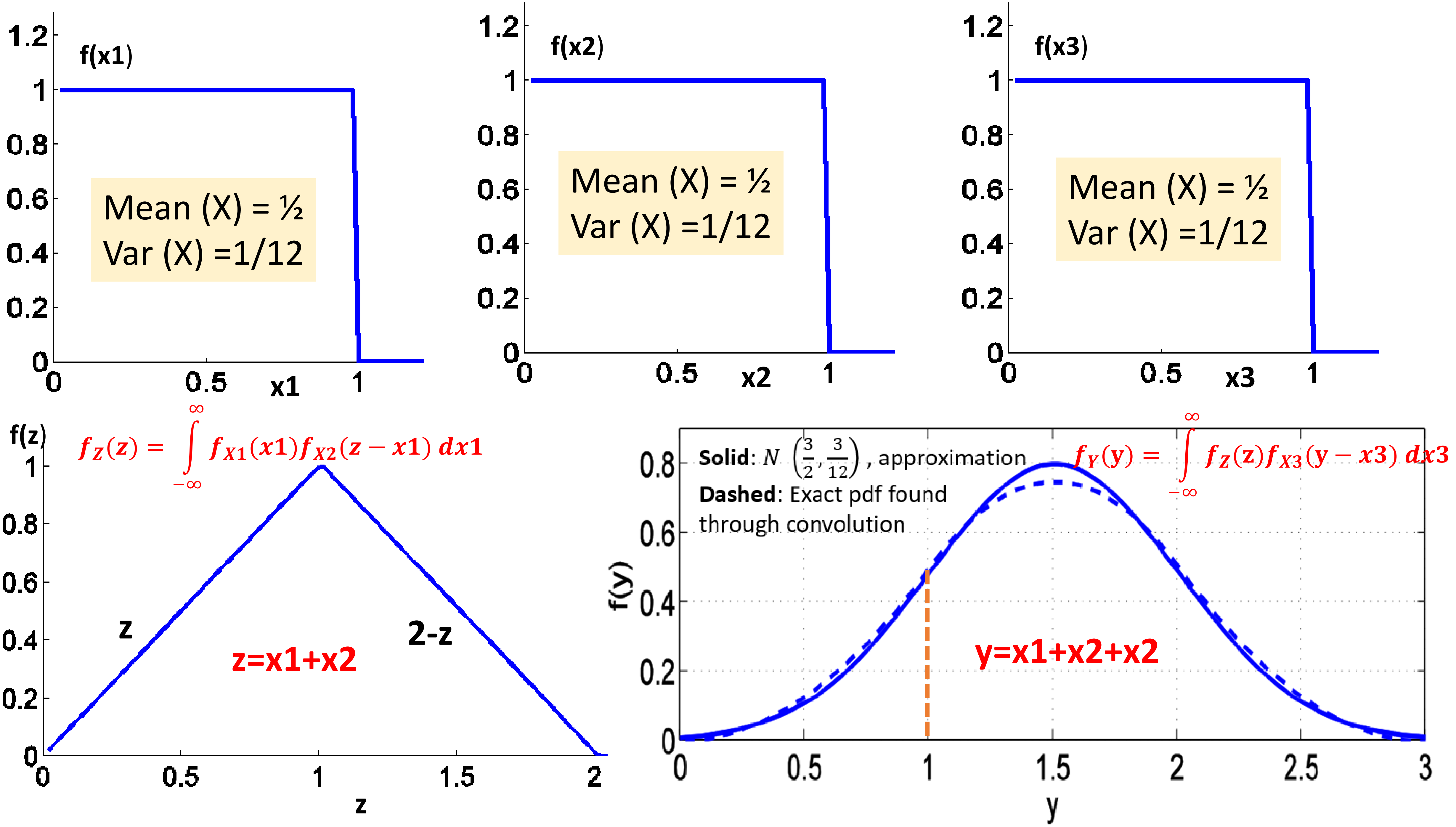
$$P(0 \le Y \le 1) = \int_0^1 y^2/2\,dy = 0.1666; \text{ Exact probability}$$

$$\frac{P(0 \le Y' \le 1)}{P(0 \le Y \le 1)} = \frac{0.1574}{0.1666} = 94.44\%$$

$$P(0 \le Y' \le 1) = \Phi\left(\frac{1-1.5}{\sqrt{0.25}}\right) - \Phi\left(\frac{0-1.5}{\sqrt{0.25}}\right) = 0.1587\text{-}0.0013 = 0.1574.; \text{ Gaussian approximation}$$

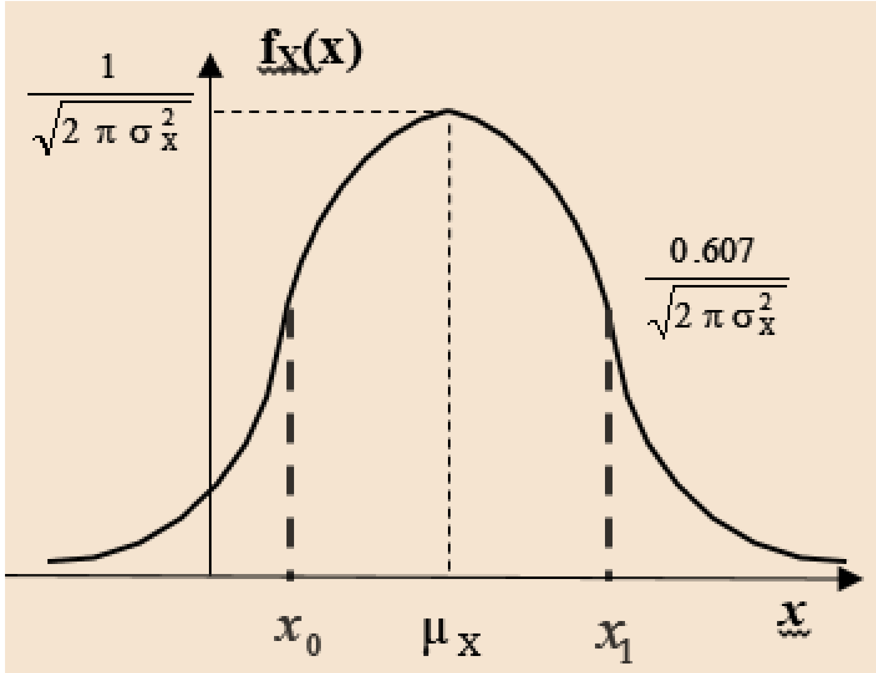
$$f_Y(y) = \begin{cases} 0 & y \le 0 \\ y^2/2 & 0 < y \le 1 \\ 3y - y^2 - 3/2 & 1 < y \le 2 \\ (3-y)^2/2 & 2 < y \le 3 \\ 0 & y > 3 \end{cases}$$

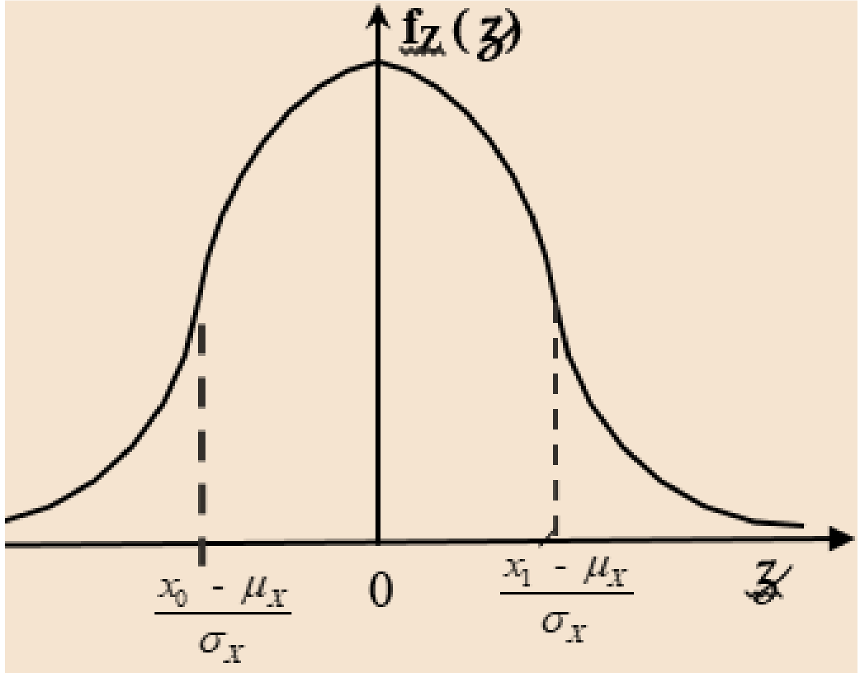



f(x1)
Mean (X) = ½
Var (X) =1/12
x1
f(x2)
Mean (X) = ½
Var (X) =1/12
x2
f(x3)
Mean (X) = ½
Var (X) =1/12
x3
f(z)
$f_Z(z) = \int_{-\infty}^{\infty} f_{X1}(x1) f_{X2}(z - x1)\, dx1$
z
2-z
z=x1+x2
z
Solid: $N\left(\frac{3}{2}, \frac{3}{12}\right)$, approximation
Dashed: Exact pdf found through convolution
$f_Y(y) = \int_{-\infty}^{\infty} f_Z(z) f_{X3}(y - x3)\, dx3$
f(y)
y=x1+x2+x2
y

Remark: For a Gaussian random variable X with mean μ_X and variance σ_X^2, we recall the following two results when evaluating probabilities:

$$P(X \le x_0) = \Phi\left(\frac{x_0 - \mu_X}{\sigma_X}\right), \qquad P(x_0 \le X \le x_1) = \Phi\left(\frac{x_1 - \mu_X}{\sigma_X}\right) - \Phi\left(\frac{x_0 - \mu_X}{\sigma_X}\right)$$





EXAMPLE: An electronic company manufactures resistors that have a mean resistance of 100 Ω and a standard deviation of 10 Ω. Find the probability that a random sample of n = 25 resistors will have an average resistance less than 95 Ω.

SOLUTION:

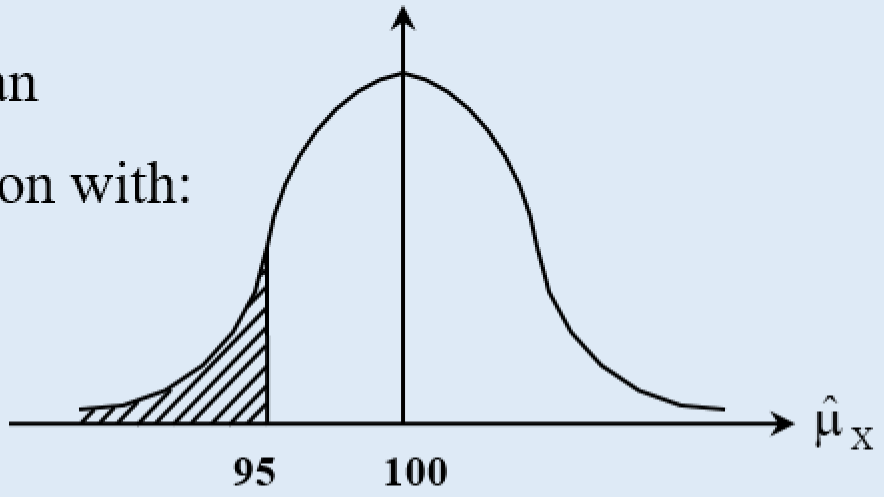
With n =25, we can approximate the sample mean

$\hat{\mu} = (X_1 + X_2 + \dots X_{25}) / 25$ by a normal distribution with:

Mean: $E\{\hat{\mu}\} = (\mu + \mu + \dots \mu) / 25 = \mu = 100$

$$Var(\hat{\mu}_X) = \hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{10^2}{25} = 4 \Rightarrow \hat{\sigma}_X = 2$$

$$P(\hat{\mu}_X < 95) = \Phi\left(\frac{95 - 100}{2}\right) = \Phi(-2.5) = 0.00621$$



$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow N(\mu_X, \sigma_X^2/n) \ as\ n \rightarrow \infty$$

Remark: Note that in this example, the distribution of the manufactured resistors is not known; only the mean and variance are known.

EXAMPLE: Suppose that the random variable X has a uniform distribution over the interval $0 \leq X \leq 1$. A random sample of size 30 is drawn from this distribution.

a. Find the probability distribution of the sample mean $\hat{\mu} = \left(\sum_{i=1}^{n} X_i \right) \Big/ n$

b. Find $P(\hat{\mu}_X) < 0.52$.

SOLUTION: Since X has a continuous uniform distribution, and since n = 30, then the probability density function of the sample mean $\hat{\mu}_X$ is approximately normal with:

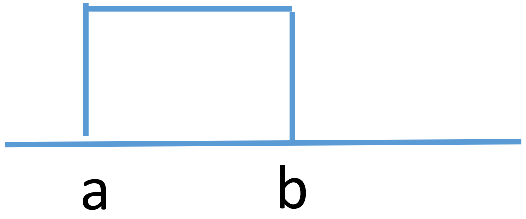
$$\text{Mean: } E(\hat{\mu}_X) = E(X) = \left(\frac{a+b}{2} \right) = \left(\frac{0+1}{2} \right) = \frac{1}{2},$$

$$Var(\hat{\mu}_X) = \hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{(b-a)^2}{(12)n} = \frac{(1-0)^2}{(12)(30)} = \frac{1}{360} \Rightarrow \hat{\sigma}_X = \sqrt{1/360} = 0.0527$$

$$P(\hat{\mu}_X < 0.52) = \Phi\left(\frac{0.52 - \mu_X}{\hat{\sigma}_X} \right) = \Phi\left(\frac{0.52 - 0.5}{0.0527} \right) = \Phi(0.379) = 0.648027$$

Remark: Note that in this example, the distribution of the sampled data is known. From the distribution, we can determine its mean and variance.

The Uniform Distribution



Mean = (a + b)/2
Var = (b-a)^2/12

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow N(\mu_X, \sigma_X^2/n) \text{ as } n \rightarrow \infty$$

EXAMPLE: Suppose that X is a discrete distribution, which assumes the two values 1 and 0 with equal probability. A random sample of size 50 is drawn from this distribution.

a. Find the probability distribution of the sample mean $\hat{\mu} = \left(\sum_{i=1}^{n} X_i \right) \Big/ n$

b. Find $P(\hat{\mu}_X) < 0.6$.

$\sum_{i=1}^{n} X_i$ Binomial with parameters n=50 and p=1/2

SOLUTION: Since n = 50 (> 30), then the probability density function of the sample mean $\hat{\mu}_X$ is approximately normal with:

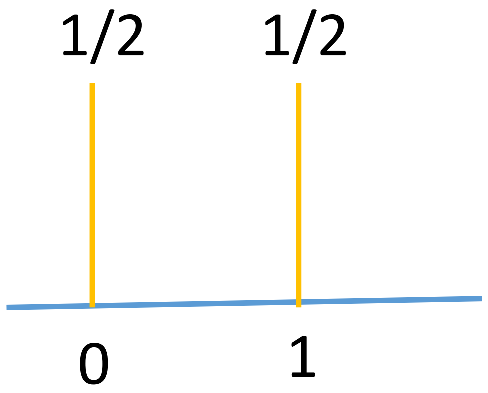
$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow N(\mu_X, \sigma_X^2/n) \; as \; n \rightarrow \infty$$

$$E(\hat{\mu}_X) = \mu_X = 0(1/2) + (1)(1/2) = 1/2$$

$$Var(\hat{\mu}_X) = \hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{(0-1/2)^2(1/2) + (1-1/2)^2(1/2)}{50} = \frac{1}{200} \Rightarrow \hat{\sigma}_X = \sqrt{1/200} = 0.0707$$

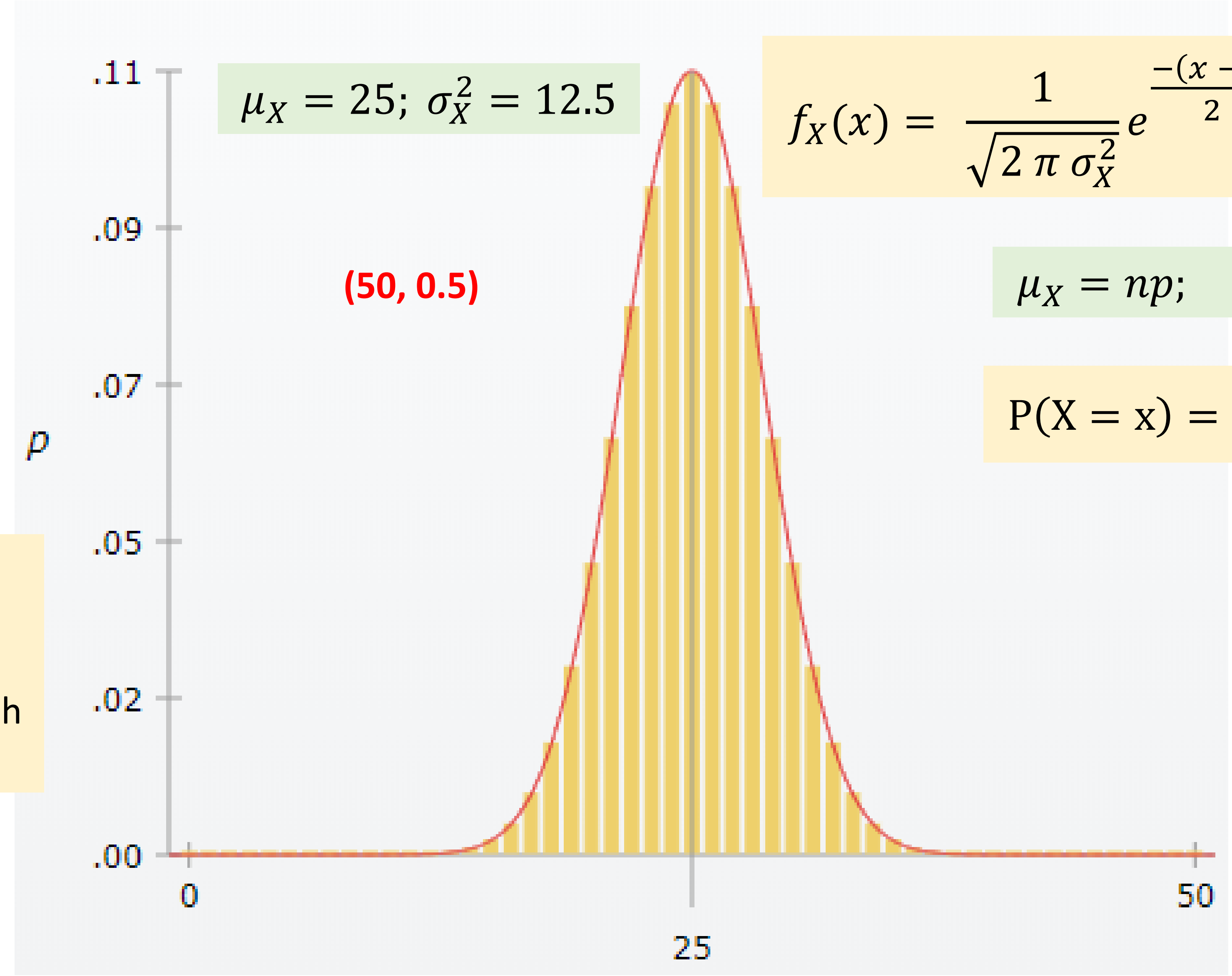
$$P(\hat{\mu}_X < 0.6) = \Phi\left(\frac{0.6 - \mu_X}{\hat{\sigma}_X}\right) = \Phi\left(\frac{0.6 - 0.5}{0.0707}\right) = \Phi(1.414) = 0.92073$$

SOLUTION: Note that in this example, X has a discrete distribution. However, since n is large, we have used the continuous Gaussian distribution to approximate the distribution of the discrete variable $\hat{\mu}_X$.



Mean (X) = ½
Var (X) =1/4

Normal Approximation of the Binomial Distribution



$\mu_X = 25;\ \sigma_X^2 = 12.5$

$$f_X(x) = \frac{1}{\sqrt{2\,\pi\,\sigma_X^2}} e^{\frac{-(x-\mu_X)^2}{2\,\sigma_X^2}}; \quad -\infty < x < \infty$$

$$\mu_X = np; \quad \sigma_X^2 = np(1-p)$$

$$\mathrm{P(X = x)} = \binom{n}{x} p^x (1-p)^{n-x}$$

The Binomial with parameters n=50 and p=1/2 along with the Gaussian distribution with same parameter.

Source: https://digitalfirst.bfwpub.com/stats_applet/stats_applet_2_cltbinom.html

EXAMPLE: The lifetime of a special type of chargeable battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, then it is replaced by a newly charged one. Assume we have 25 such battery replacements, the lifetime of which are independent. Approximate the probability that at least 1100 hours of use can be obtained.

SOLUTION: Let $X_1, X_2, \ldots, X_{25}$ be the lifetimes of the batteries.

Let $Y = X_1 + X_2 + \ldots\ldots + X_{25}$ be the overall lifetime of the system

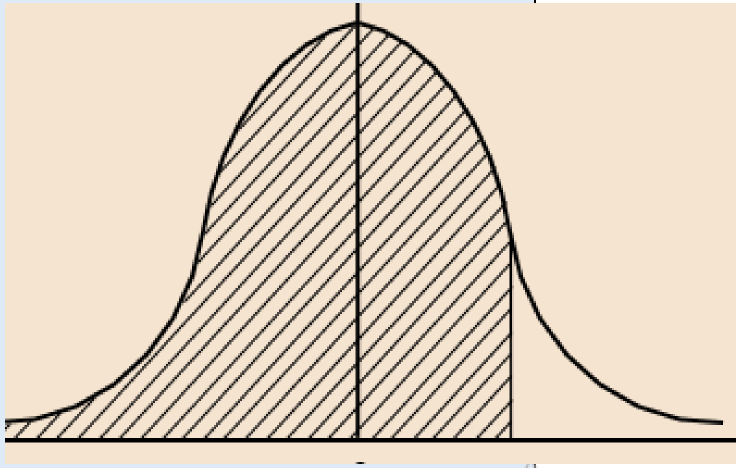
Since X_i are independent, then Y will be approximately normal with mean and variance:

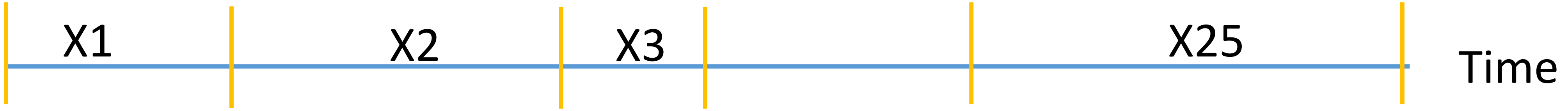
$$\mu_Y = \mu_1 + \mu_2 + \ldots + \mu_{25} = 25\mu = (25)(40) = 1000$$

$$\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{25}^2 = 25\sigma_X^2 = (25)\left[(20)^2\right] = 10000$$

$$P(Y > 1100) = 1 - P(Y \le 1100) = 1 - \Phi\left(\frac{1100 - \mu_Y}{\hat{\sigma}_Y}\right)$$

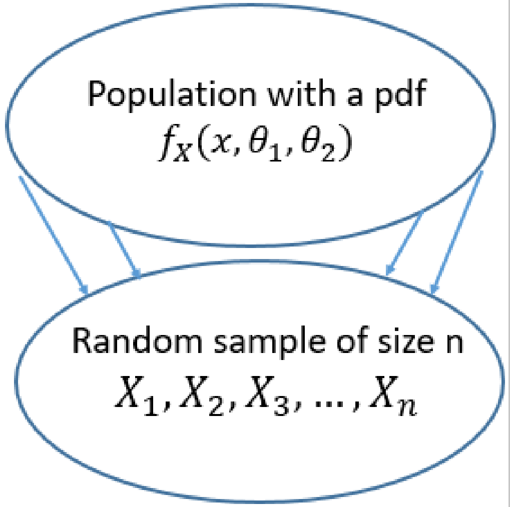
$$= 1 - \Phi\left(\frac{1100 - 1000}{\sqrt{10000}}\right) = 1 - \Phi(1) = 0.158655$$





Estimation of Parameters

- The field of ***statistical inference*** consists of those methods used to make decisions or to draw conclusions about a population. These methods utilize the information contained in a random sample taken from a population in drawing conclusions.
- The ***population*** consists of all the conceivable items, observations, or measurements in a group. In most cases, it is not practical to obtain all the measurements in a given population (eligible voters, the unemployed, people below poverty line, Birzeit University students, high school teachers, …)
- For example, suppose we need to find the average height and standard deviation of the university male and female students. The population here is all university students. It is evident that to get exact results, we need to take the height of all students and compute the average and the standard deviation (these are the population parameters).
- In practice, **a random sample** of size n is drawn from the university population. The heights of the selected students are taken, and then the mean and standard deviation of the sample are calculated. The sample mean and standard deviation are used to describe the actual mean and standard deviation
- Estimates of population parameters derived from a subset of the measurements in a sample drawn from the underlying population are called ***sample statistics***

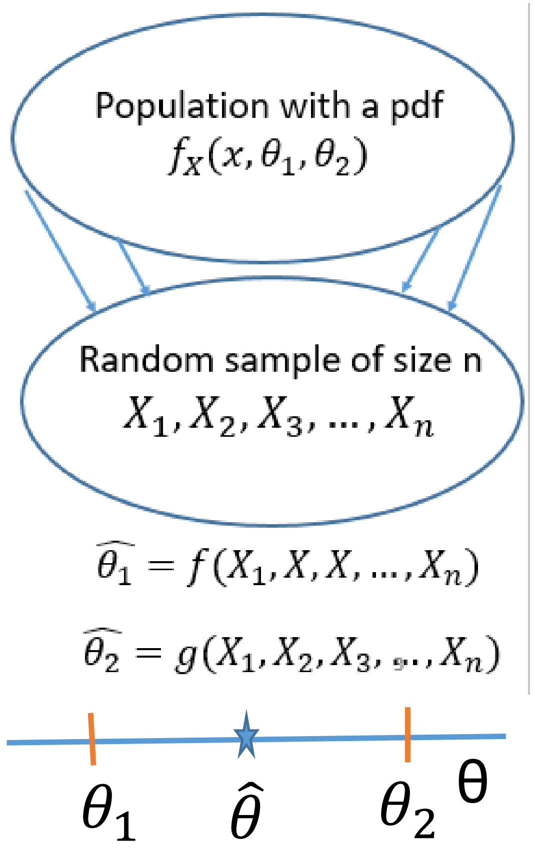


$$\widehat{\theta_1} = f(X_1, X, X, \dots, X_n)$$

$$\widehat{\theta_2} = g(X_1, X_2, X_3, \dots, X_n)$$

Estimation of Parameters

- **Statistical inference may be divided into two major areas:** ***Parameter estimation*** and ***hypotheses testing***.
- In this chapter, we focus on parameter estimation and consider hypothesis testing in the next chapter.
- For populations, we define numbers called ***parameters*** that characterize important properties of the distributions, like **the mean and standard deviation of a normal distribution, the probability of success p in the binomial distribution, the rate of arrival in the Poisson process, and the end points a and b of the uniform distribution**.
- ***Estimation*** represents ways or processes of learning and determining the population parameter based on the model fitted to the data.
- **Point estimation, interval estimation**, and **hypothesis testing** are three main ways of learning about the population parameter from the sample statistics.



$$\hat{\theta} = f(X_1, X_2, \ldots, X_n)$$

$$P(\theta_1 \le \theta \le \theta_2) \ge 1 - \alpha \; ; \; 0 < \alpha < 1$$

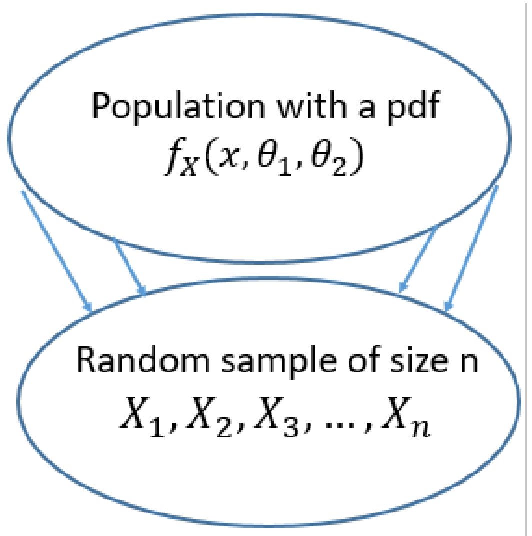
Point and interval estimation

Estimation of Parameters

- In **point estimation**, we estimate the unknown parameter using a **single number** that is calculated from the sample data.
  - **The point estimate** of the height of students based on the random sample would be a number like 175cm for male students and 165 cm for female students.
- In **interval estimation**, we estimate an unknown parameter using an **interval of values** that is likely to contain the true value of that parameter (and state how confident we are that this interval indeed captures the true value of the parameter).
  - A **confidence interval** would be like:

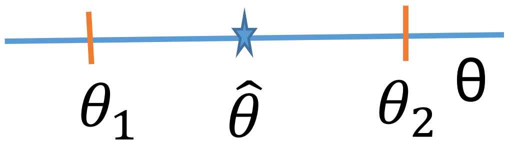
P (height of male students falls between 173cm and 177cm) > 0.95.

- In **hypothesis testing**, we begin with a claim about the population (usually, called the null hypothesis), and we check **whether or not the data** obtained from the sample **provide evidence in favor or against this claim.**
  - The **hypothesis testing** would test the null hypothesis: Height of male students = 175 cm versus the alternative hypothesis Height > 175 cm



$\widehat{\theta_1} = f(X_1, X, X, \dots, X_n)$

$\widehat{\theta_2} = g(X_1, X_2, X_3, \dots, X_n)$



$\hat{\theta} = f(X_1, X_2, \dots, X_n)$

$P(\theta_1 \le \theta \le \theta_2) \ge 1 - \alpha\,;\; 0 < \alpha < 1$

Point and interval estimation

Formal Definitions and Terminology

- In statistics, we take a set of n independent measurements $(X_1, X_2, \ldots, X_n)$ of size n from a distribution X (population) for which the pdf is $f_X(x, \theta_1, \theta_2)$ by performing that experiment n times.
- The random variables $X_1, X_2 \ldots, X_n$, called a ***random sample***, have the same distribution $f_X(x, \theta_1, \theta_2)$ and are assumed to be independent. **Joint pdf = product of marginal pdf's**
- **The purpose is to draw conclusions from the properties of the *sample* about properties of the distribution of the corresponding X (*the population*).**
- For populations, we define numbers called ***parameters***, denoted (θ_1, θ_2) that characterize important properties of the distributions. Here, the pdf is explicitly expressed in terms of the parameter as $f_X(x, \theta_1, \theta_2)$. These parameters are **unknown**.
- The unknown parameters (θ_1, θ_2) are estimated by some appropriate functions of the observations $\widehat{\theta_1} = f(X_1, X_2, \ldots, X_n)\,;\ \widehat{\theta_2} = g(X_1, X_2, \ldots, X_n)$

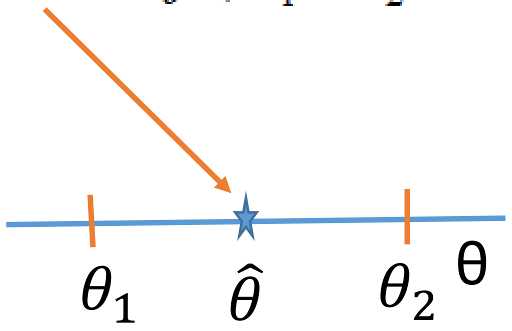
Formal Definitions and Terminology

- The function $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$, of the observable sample data, that is used to estimate the unknown population parameter is called a ***statistic*** or an ***estimator***. A particular value of the estimator is called an *estimate* of θ.
- A probability distribution of a statistic is called its ***sampling distribution*** $f_{\hat{\theta}}(\widehat{\theta})$
- We consider two types of parameter estimation, ***point estimation*** and ***interval estimation***.
- **Examples of parameters**: height of male and female university students, the percentage of smokers among high school students, the compression strength of concrete, the percentage of students who favor e-learning techniques, …

Point Estimation

- Point estimation involves the use of the sample data to calculate a single value, which is to serve as a best guess for an unknown parameter. In other words, a point estimate of some unknown population parameter (θ) is a single numerical value

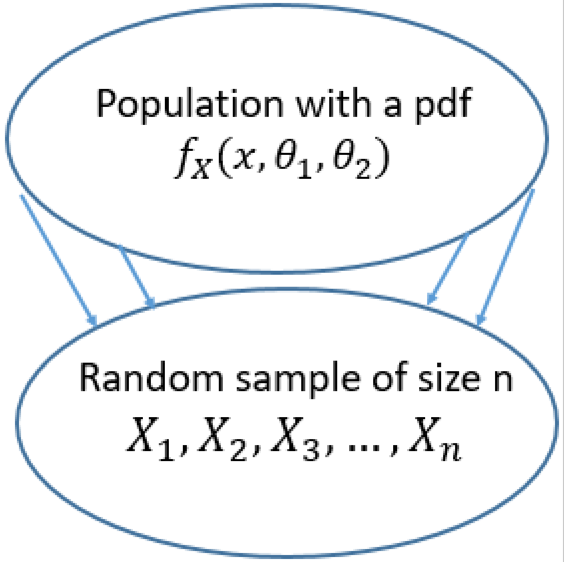
$$\hat{\theta} = f(X_1, X_2, \ldots, X_n) \ .$$



$$\hat{\theta} = f(X_1, X_2, \ldots, X_n)$$

$$P(\theta_1 \le \theta \le \theta_2) \ge 1 - \alpha \ ; \ 0 < \alpha < 1$$

Point and interval estimation



$$\widehat{\theta_1} = f(X_1, X, X, \ldots, X_n)$$

$$\widehat{\theta_2} = g(X_1, X_2, X_3, \ldots, X_n)$$

The estimator $\hat{\theta}$ is a random variable with a sampling distribution $f_{\hat{\theta}}(\widehat{\theta})$. This estimator should have certain desirable properties that makes it close to the true value in a probabilistic sense.

- It should be unbiased
- Should have a small variance
- Should have a small mean squared error.

These properties are considered next.

Desirable Properties of Point Estimators

- An estimator should be close to the true value of the unknown parameter.
- **Definition:** A point estimator $(\hat{\theta})$ is ***unbiased*** estimator of (θ) if $E(\hat{\theta}) = \theta$.
- If the estimator is *biased*, then $E(\hat{\theta}) - \theta = B$ is called the bias of the estimator $(\hat{\theta})$.
- Let $\hat{\theta}_A$, $\hat{\theta}_B$ be two *unbiased* estimators of (θ). A logical principle of estimation when selecting among several estimators is to chooses the one that has the ***minimum variance***.
- **Definition:** If we consider all unbiased estimators of (θ), the one with the smallest variance is called the ***minimum variance unbiased estimator* (MVUE)**.
- When $Var(\hat{\theta}_A) < Var(\hat{\theta}_B)$, $\hat{\theta}_A$ is called more ***efficient*** than $\hat{\theta}_B$.
- Recall that the variance $Var(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta})]^2\}$ is a measure of the imprecision of the estimator (a measure of the spread of the data around the mean value)

Mean Squared Error of an Estimator

- **Definition**: the mean square error of an estimator $(\hat{\theta})$ of the parameter (θ) is defined as:

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

- This measure of goodness takes into account both the bias and imprecision.
- $MSE(\hat{\theta})$ can also be expressed as:

$$MSE(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2\} = E\{[(\hat{\theta} - E(\hat{\theta})) + ((\underbrace{E(\hat{\theta}) - \theta}_{=B}))]^2\}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2 + 2\ B\ \ E\{(\underbrace{\hat{\theta} - E(\hat{\theta})}_{=0})\} + B^2 = Var(\hat{\theta}) + B^2$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + B^2$$

- **Definition***:* An estimator whose variance and bias go to zero as the number of observations goes to infinity is called ***consistent.***

EXAMPLE: Let X_1 and X_2 be a random sample of size two from a population with mean μ_x and variance σ_X^2. Two estimators for μ_x are proposed: $\hat{\mu}_1 = \dfrac{X_1 + X_2}{2}$ and $\hat{\mu}_2 = \dfrac{X_1 + 2X_2}{3}$.

Which estimator is better and in what sense?

Four more examples on point estimators are fully explained in the next lecture, entitled "examples on point estimators)

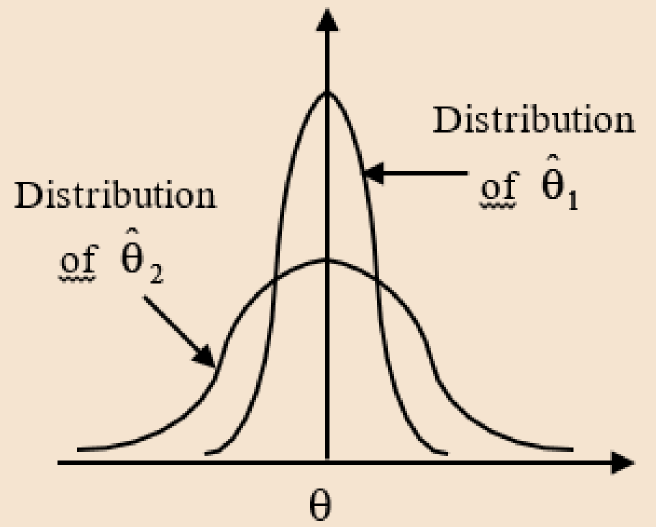
SOLUTION: First, we check for the un-biasedness of the two estimators

$E(\hat{\mu}_1) = E(\dfrac{X_1 + X_2}{2}) = \dfrac{\mu_x + \mu_x}{2} = \mu_x$. Therefore, $\hat{\mu}_1$ is an unbiased estimator of μ_x.

(First bias: $B_1 = E(\hat{\mu}_1) - \mu_1 = 0$)

$E(\hat{\mu}_2) = E(\dfrac{X_1 + 2X_2}{3}) = \dfrac{\mu_x + 2\mu_x}{3} = \mu_x$. Therefore, $\hat{\mu}_2$ is also an unbiased estimator of μ_x

(Second bias: $B_2 = E(\hat{\mu}_2) - \mu_2 = 0$) **Both estimators are unbiased.**



Next, Now, we evaluate the variance of each one of the two estimators:

$$Var(\hat{\mu}_1) = Var(\frac{X_1 + X_2}{2}) = \frac{1}{4}\sigma_x^2 + \frac{1}{4}\sigma_x^2 = \frac{1}{2}\sigma_x^2$$

$$Var(\hat{\mu}_2) = Var(\frac{X_1 + 2X_2}{3}) = \frac{1}{9}\sigma_x^2 + \frac{4}{9}\sigma_x^2 = \frac{5}{9}\sigma_x^2$$

Since $Var(\hat{\mu}_1) = \dfrac{1}{2}\sigma_x^2 < Var(\hat{\mu}_2) = \dfrac{5}{9}\sigma_x^2$, the first estimator is more efficient.

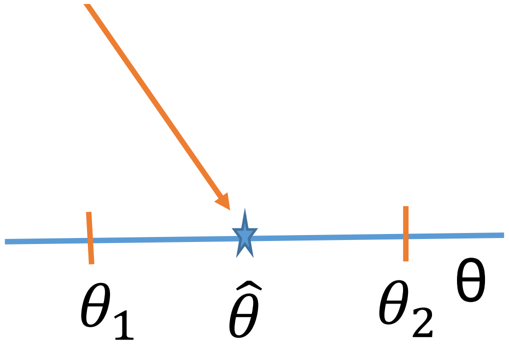
The sampling distribution of the statistic $f_{\hat{\theta}}(\widehat{\theta})$.

$$MSE(\hat{\mu}_1) = Var(\hat{\mu}_1) + B_1^2 = \frac{1}{2}\sigma_x^2, \quad MSE(\hat{\mu}_2) = Var(\hat{\mu}_2) + B_2^2 = \frac{5}{9}\sigma_x^2, \Rightarrow MSE(\hat{\mu}_1) < MSE(\hat{\mu}_2)$$

Point Estimation

Point estimation involves the use of the sample data to calculate a single value, which is to serve as a best guess for an unknown parameter. In other words, a point estimate of some unknown population parameter (θ) is a single numerical value

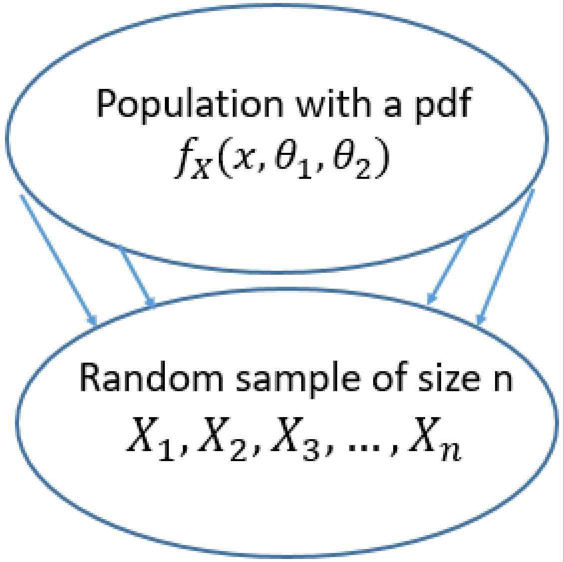
$$\hat{\theta} = f(X_1, X_2, \ldots, X_n) \; .$$



$$\hat{\theta} = f(X_1, X_2, \ldots, X_n)$$

$$P(\theta_1 \leq \theta \leq \theta_2) \geq 1 - \alpha \; ; \; 0 < \alpha < 1$$

Point and interval estimation



$$\widehat{\theta_1} = f(X_1, X, X, \ldots, X_n)$$

$$\widehat{\theta_2} = g(X_1, X_2, X_3, \ldots, X_n)$$

The estimator $\hat{\theta}$ is a random variable with a sampling distribution $f_{\hat{\theta}}(\widehat{\theta})$. This estimator should have certain desirable properties that makes it close to the true value in a probabilistic sense.

- It should be unbiased
- Should have a small variance
- Should have a small mean squared error.

These properties are considered next.

In **interval estimation**, we estimate an unknown parameter using an **interval of values** that is likely to contain the true value of that parameter (and state how confident we are that this interval indeed captures the true value of the parameter)

Examples of Point Estimators

Unknown Parameter (θ)	Statistic $(\hat{\Theta})$	Remarks
μ_X	$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i$	Used to estimate the mean regardless of whether the variance is known or unknown.
σ_X^2	$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$	Used to estimate the variance when the mean is unknown.
σ_X^2	$\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2$	Used to estimate the variance, when the mean is known.
p	$\hat{P} = \frac{x}{n}$	Used to estimate the probability of a success in a binomial distribution. n: sample size, x: number of successes in the sample
$\mu_{X1} - \mu_{X2}$	$\hat{\mu}_{X1} - \hat{\mu}_{X2} = \sum_{i=1}^{n}\frac{x_{1i}}{n_1} - \sum_{i=1}^{n}\frac{x_{2i}}{n_2}$	Used to estimate the difference in the means of two populations.
$p_1 - p_2$	$\hat{P}_1 - \hat{P}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$	Used to estimate the difference in the proportions of two populations.

Desirable Properties of Point Estimators

- An estimator should be close to the true value of the unknown parameter.
- **Definition:** A point estimator $(\hat{\theta})$ is ***unbiased*** estimator of (θ) if $E(\hat{\theta}) = \theta$.
- If the estimator is *biased*, then $E(\hat{\theta}) - \theta = B$ is called the bias of the estimator $(\hat{\theta})$.
- Let $\hat{\theta}_A$, $\hat{\theta}_B$ be two *unbiased* estimators of (θ). A logical principle of estimation when selecting among several estimators is to chooses the one that has the ***minimum variance***.
- **Definition:** If we consider all unbiased estimators of (θ), the one with the smallest variance is called the ***minimum variance unbiased estimator*** **(MVUE)**.
- When $Var(\hat{\theta}_A) < Var(\hat{\theta}_B)$, $\hat{\theta}_A$ is called more ***efficient*** than $\hat{\theta}_B$.
- Recall that the variance $Var(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta})]^2\}$ is a measure of the imprecision of the estimator (a measure of the spread of the data around the mean value)

Mean Squared Error of an Estimator

- **Definition**: the mean square error of an estimator $(\hat{\theta})$ of the parameter (θ) is defined as:

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

- This measure of goodness takes into account both the bias and imprecision. $MSE(\hat{\theta})$ can also be expressed as:

$$MSE(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2\} \quad = \quad E\{[(\hat{\theta} - E(\hat{\theta})) + ((\underbrace{E(\hat{\theta}) - \theta}_{=B}))]^2\}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2 + 2\ B\ \ E\{(\underbrace{\hat{\theta} - E(\hat{\theta})}_{=\ 0})\} + B^2 = Var(\hat{\theta}) + B^2$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + B^2$$

- **Definition:** An estimator whose variance and bias go to zero as the number of observations goes to infinity is called ***consistent.***

EXAMPLE: Let X_1 and X_2 be a random sample of size two from a population with mean μ_x and variance σ_X^2 . Two estimators for μ_x are proposed: $\hat{\mu}_1 = \frac{X_1 + X_2}{2}$ and $\hat{\mu}_2 = \frac{X_1 + 2X_2}{3}$.

Which estimator is better and in what sense?

SOLUTION: First, we check for the un-biasedness of the two estimators

$E(\hat{\mu}_1) = E(\frac{X_1 + X_2}{2}) = \frac{\mu_x + \mu_x}{2} = \mu_x$. Therefore, $\hat{\mu}_1$ is an unbiased estimator of μ_x.

(First bias: $B_1 = E(\hat{\mu}_1) - \mu_1 = 0$)

$E(\hat{\mu}_2) = E(\frac{X_1 + 2X_2}{3}) = \frac{\mu_x + 2\mu_x}{3} = \mu_x$. Therefore, $\hat{\mu}_2$ is also an unbiased estimator of μ_x

(Second bias: $B_2 = E(\hat{\mu}_2) - \mu_2 = 0$) **Both estimators are unbiased.**

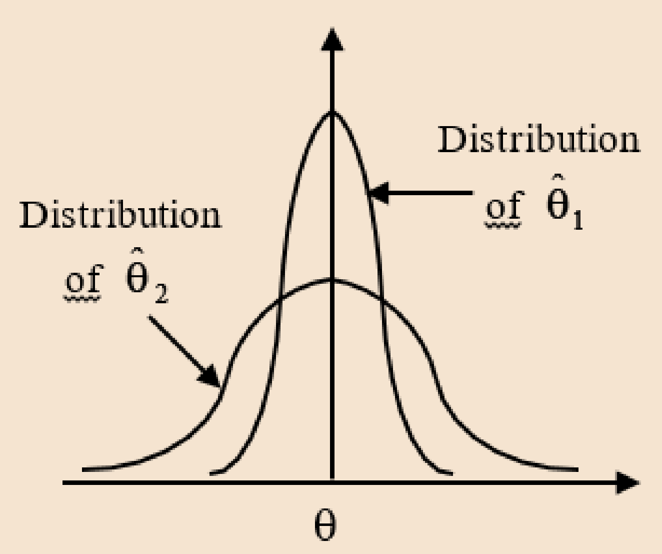
Next, Now, we evaluate the variance of each one of the two estimators:

$$Var(\hat{\mu}_1) = Var(\frac{X_1 + X_2}{2}) = \frac{1}{4}\sigma_x^2 + \frac{1}{4}\sigma_x^2 = \frac{1}{2}\sigma_x^2$$

$$Var(\hat{\mu}_2) = Var(\frac{X_1 + 2X_2}{3}) = \frac{1}{9}\sigma_x^2 + \frac{4}{9}\sigma_x^2 = \frac{5}{9}\sigma_x^2$$

Since $Var(\hat{\mu}_1) = \frac{1}{2}\sigma_x^2 < Var(\hat{\mu}_2) = \frac{5}{9}\sigma_x^2$, the first estimator is more efficient.

$$MSE(\hat{\mu}_1) = Var(\hat{\mu}_1) + B_1^2 = \frac{1}{2}\sigma_x^2, \; MSE(\hat{\mu}_2) = Var(\hat{\mu}_2) + B_2^2 = \frac{5}{9}\sigma_x^2, \Rightarrow MSE(\hat{\mu}_1) < MSE(\hat{\mu}_2)$$



The sampling distribution of the statistic $f_{\hat{\theta}}(\widehat{\theta})$.

EXAMPLE: Find the expected value and the variance of the sample mean

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

SOLUTION: Let μ_X and σ_X^2 are the mean and variance of the population parameters.

The estimator $\hat{\mu}_X$ is used to estimate the population mean μ_x.

$$E\{\hat{\mu}_X\} = \frac{1}{n} E\left\{\sum_{i=1}^{n} x_i\right\} = \frac{1}{n}\left\{\sum_{i=1}^{n} E\{x_i\}\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n} \mu_x\right\} = \frac{1}{n}(n\mu_x) = \mu_x$$

(**unbiased estimator** $B = E(\hat{\mu}) - \mu = 0$)

The variance of $\hat{\mu}_X$ is

$$Var\{\hat{\mu}_X\} = Var\left(\frac{1}{\mathrm{n}}\sum_{i=1}^{n} x_i\right) \; \frac{1}{n^2} Var\left\{\sum_{i=1}^{n} x_i\right\} = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{n\sigma_{\mathrm{X}}^2}{n^2} = \frac{\sigma_{\mathrm{X}}^2}{n}.$$

Remark: Note $Var\{\hat{\mu}_X\}$ tends to zero as n tends to infinity. Therefore, $\hat{\mu}_X$ is a **consistent estimator**.

EXAMPLE: Show that the sample variance

$$\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2$$

(when the mean is known) is an unbiased estimator of the population variance σ_X^2

SOLUTION: Need to verify that $E\{\hat{\sigma}_X^2\} = \sigma_X^2$

$$E\left(\hat{\sigma}_X^2\right) = \frac{1}{n}\sum_{i=1}^{n}E(X_i - \mu_X)^2$$

Note that $E(X_i - \mu_X)^2 = \sigma_X^2$; since the mean is known . Therefore,

$$E\left(\hat{\sigma}_X^2\right) = \frac{1}{n}\sum_{i=1}^{n}\sigma_X^2 = \frac{n\sigma_X^2}{n} = \sigma_X^2 .$$

Therefore, $\hat{\sigma}_X^2$ is an unbiased estimator of σ_X^2 .

EXAMPLE: Show that the sample variance $\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$ (*when the mean is unknown*) is an unbiased estimator of the population variance σ_X^2

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

(x_1, x_1)	(x_1, x_2)	(x_1, x_3)
(x_2, x_1)	(x_2, x_2)	(x_2, x_3)
(x_3, x_1)	(x_3, x_2)	(x_3, x_3)

Total Terms n^2
Diagonal: n
Off Diagonal $n^2 - n$
$= n(n-1)$

SOLUTION: Need to verify that $E\{\hat{\sigma}_X^2\} = \sigma_X^2$

A computationally simpler expression for the sample variance is

$$\hat{\sigma}_X^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} \Rightarrow E\{\hat{\sigma}_X^2\} = \frac{1}{n(n-1)} E\left\{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right\} = \frac{1}{n(n-1)}\left\{n\sum_{i=1}^{n} E(x_i^2) - E\left(\sum_{i=1}^{n} x_i\right)^2\right\}$$

Note that since $E\{x_i^2\} = \mu_X^2 + \sigma_X^2$, then $n\sum_{i=1}^{n} E(x_i^2) = n^2(\mu_X^2 + \sigma_X^2)$

$$E\left(X_i^2\right) = E(X_i X_i) = \mu_X^2 + \sigma_X^2$$

$$E\left(X_i X_j\right) = \mu_X^2$$

$$E\left(\sum_{i=1}^{n} x_i\right)^2 = E\left(\sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j\right) = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(x_i x_j)$$

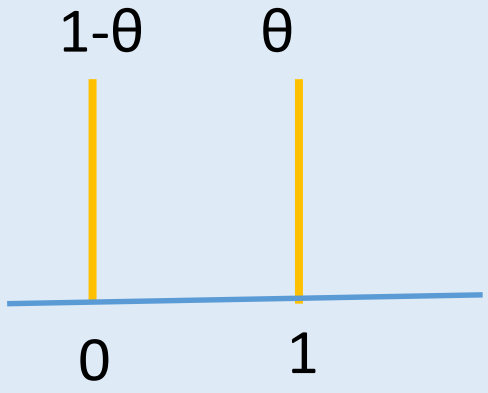
The double summation contains n^2 elements n terms are such that i=j, and $(n^2-n)=n(n-1)$ are such that $i \neq j$. When i=j, $E\{x_i^2\} = \mu_X^2 + \sigma_X^2$, and when $i \neq j$, $E(x_i x_j) = E(x_i)E(x_j) = \mu_X^2$ since the random variables are independent.

Therefore, $\sum_{i=1}^{n}\sum_{j=1}^{n} E(x_i x_j) = n(\mu_X^2 + \sigma_X^2) + n(n-1)\mu_X^2 = n\sigma_X^2 + n^2\mu_X^2$

$$E\{\hat{\sigma}_X^2\} = \frac{1}{n(n-1)}\left\{n^2(\mu_X^2 + \sigma_X^2) - n\sigma_X^2 - n^2\mu_X^2\right\} = \frac{1}{n(n-1)}\left\{n^2\sigma_X^2 - n\sigma_X^2\right\} = \sigma_X^2$$

EXAMPLE: Consider a random sample of size n taken from a discrete distribution, the pmf of which is given by: $f(x) = \theta^x (1-\theta)^{1-x}$, x = 0, 1. Two estimators for θ are proposed $\hat{\theta}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{\theta}_2 = \frac{n\overline{X}+1}{n+2}$

a. Which one of these two estimators is an unbiased estimator of the parameter θ?
b. Which one has a smaller variance?



SOLUTION: First, we find the mean and variance of the distribution X.

$$E(X) = \mu_x = (0)(1-\theta) + (1)(\theta) = \theta \,; \quad E(X^2) = (0)(1-\theta) + (1)(\theta) = \theta$$

$$Var(X) = \sigma_X^2 = E(X^2) - (\mu_x)^2 = \theta - \theta^2 = \theta(1-\theta)$$

Mean (X) = θ
Var (X) =θ(1-θ)

Expected values of the two estimators

$$E\left(\hat{\theta}_1\right) = E\left(\overline{X}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{n\theta}{n} = \theta \Rightarrow \text{ unbiaded estimator} \Rightarrow B_1 = 0$$

$$E\left(\hat{\theta}_2\right) = \frac{1}{n+2} E\left\{n\overline{X}+1\right\} = \frac{1}{n+2}\left\{nE\left(\overline{X}\right)+1\right\} = \frac{1}{n+2}\left\{n\theta+1\right\} = \frac{n\theta+1}{n+2} \Rightarrow \text{ Biased estimator}$$

(Second bias: $B_2 = E(\hat{\mu}_2) - \mu_2 = \frac{n\theta+1}{n+2} - \theta = \frac{1-2\theta}{n+2}$).

The bias approaches 0 as n goes to infinity

Next, Now, we evaluate the variance of each one of the two estimators:

$$Var\{\hat{\theta}_1\} = Var\{\overline{X}\} = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} Var\left\{\sum_{i=1}^{n} x_i\right\} = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{n\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n} = \frac{\theta(1-\theta)}{n}$$

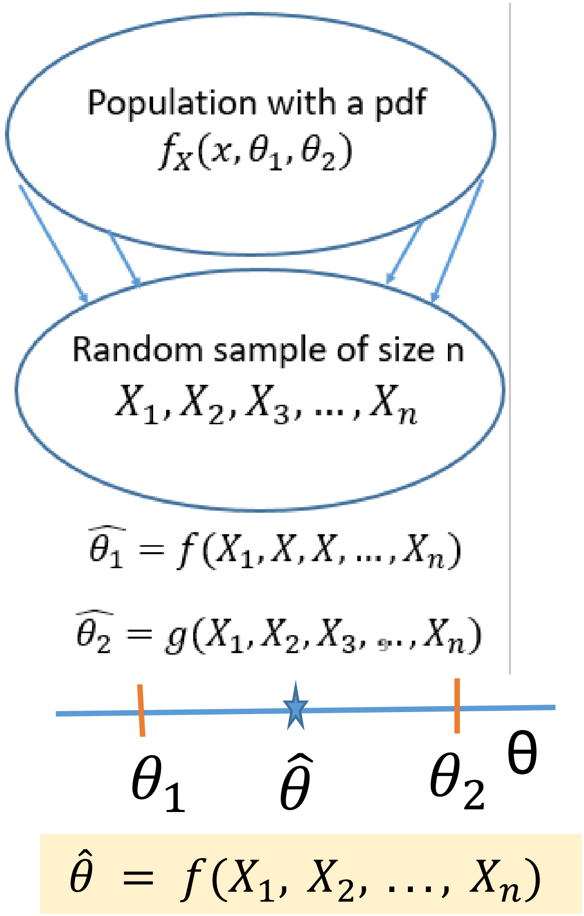
$$Var\{\hat{\theta}_2\} = Var\left(\frac{n\overline{X}+1}{n+2}\right) = \frac{n^2}{(n+2)^2} Var\{\overline{X}\} = \frac{n^2}{(n+2)^2}\frac{\theta(1-\theta)}{n} = \frac{n\theta(1-\theta)}{(n+2)^2}$$

Since $Var(\hat{\theta}_2) = \dfrac{n\theta(1-\theta)}{(n+2)^2} < Var(\hat{\theta}_1) = \dfrac{\theta(1-\theta)}{n}$, the second estimator is more efficient.

$$MSE(\hat{\mu}_1) = Var(\hat{\mu}_1) + B_1^2 = \frac{\theta(1-\theta)}{n} \quad MSE(\hat{\mu}_2) = Var(\hat{\mu}_2) + B_2^2 = \frac{n\theta(1-\theta)}{(n+2)^2} + \left(\frac{1-2\theta}{n+2}\right)^2$$

$$\mathrm{Y} = \mathrm{aX} + \mathrm{b} \Rightarrow \mu_Y = a\mu_X + b, o_Y^2 = a^2 o_X^2,$$

Maximum Likelihood (ML) Estimation
(A Method for Obtaining Point Estimators)

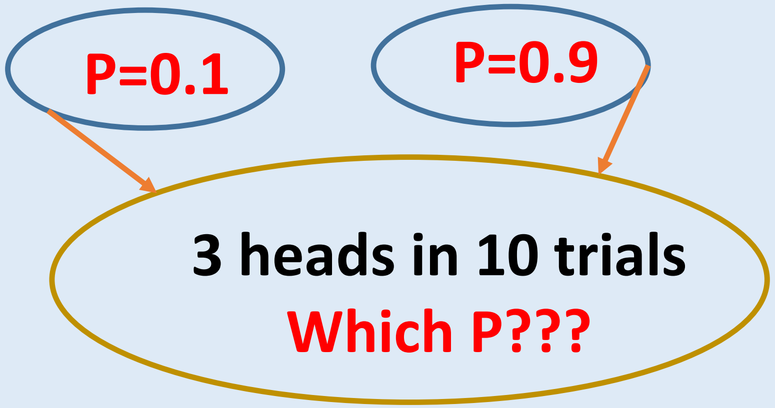


- Point Estimation deals with the **method** of estimating an unknown parameter of a population based on random samples from the same population. In the parameter space, it is represented as a point. Hence the name **point estimation.** Desirable properties of a point estimator was addressed in a previous lecture.
- The assumption here is that the parameter to be estimated is a constant with one value, and the sample statistic computed from the sample is estimating that value exactly.
- Maximum Likelihood is one method of obtaining point estimators..
- In this lecture, we will explain this method and present a number of illustrative examples.

Maximum Likelihood (ML) Estimation
(Method for Obtaining Point Estimators)

Motivating Example: The probability p of a success in a binomial experiment may be **0.1** or it may be **0.9**. To resolve the uncertainty, the experiment was repeated 10 times and 3 successes were observed. What will be your estimate for p in light of the experiment outcome?

Solution: Let us calculate the probability of getting 3 successes in 10 trials for the two possible values of p using the binomial distribution



$$P(x=3;0.1)=\binom{10}{3}(0.1)^3(1-0.1)^7=0.0574$$

$$P(x=3;0.9)=\binom{10}{3}(0.9)^3(1-0.9)^7=8.748*10^{-6}$$

Therefore, we observe that p=0.1 has a higher probability of producing the outcome and our estimate for p would be $\hat{p}=0.1$.

Maximum Likelihood (ML) Estimation

Motivating Example: Let p be the probability of a success in a binomial distribution. This probability is unknown. To estimate p, the experiment is performed 10 times and 3 successes are observed. Find a maximum likelihood estimate for p.

Solution: Any value of $0 \le p \le 1$ is likely to produce the three successes in the 10 trials. But there is a specific value, $\hat{p}$, to be estimated, that has the highest probability of producing the result. This value of p is called the ***maximum likelihood estimate***.

The probability of getting 3 successes in 10 trials for any value of p is:

$$f(p) = P(x = 3; p) = \binom{10}{3} p^3 (1-p)^7$$

To find the specific value of p that maximizes $f(p)$, we differentiate $f(p)$ with respect to p, set the derivative to zero, and solve for $\hat{p}$

$$\frac{df(p)}{dp} = \binom{10}{3} [3p^2(1-p)^7 + 7p^3(1-p)^6(-1)] = 0$$

Solving for p, we get $\hat{p} = 3/10$.

$0 \le p \le 1$

3 heads in 10 trials
Which P???

How to Obtain the Maximum Likelihood Estimator

The maximum likelihood estimator selects the parameter $\hat{\theta}$ due to which the measurements $X_1, X_2, \dots, X_n$ occur with the largest possible probability. The following steps summarize the procedure for obtaining a maximum likelihood estimator for a continuous parameter θ based on a random sample of measurements $X_1, X_2, \dots, X_n$ of size n

- Form the joint pdf of the measurements $X_1, X_2, \dots, X_n$ (expressed in terms of θ). The joint pdf is also known as the ***likelihood function.***

$$\boldsymbol{L(\theta) = f(X_1, X_2, \dots, X_n; \theta)}$$

- Since the observations are independent, the joint pdf is the product of the marginal pdf's

$$\boldsymbol{L(\theta) = f(X_1, \theta) f(X_2, \theta) \dots, f(X_n; \theta)}$$

- The maximum likelihood technique looks for that value $(\hat{\theta})$ of the parameter θ that maximizes the joint pdf of the samples. A necessary condition for the maximum likelihood estimator of (θ) is:

$$\frac{\partial}{\partial\theta} L(\theta) = 0 \text{ or equivalently } \frac{\partial}{\partial\theta} \ln\{L(\theta)\} = 0$$

  Note that this step is justified since ln(u) is a monotonically increasing function in u.

- Solve for $\hat{\theta}$ that maximizes $\boldsymbol{L(\theta)}$. The solution to $\frac{\partial}{\partial\theta} \ln\{L(\theta)\} = 0$ is the desired maximum likelihood estimator.

EXAMPLE: Given a random sample $X_1, X_2, ..., X_n$ of size n taken from a discrete distribution, the pmf of which is given by: $f(x) = \theta^x(1-\theta)^{1-x}$, x = 0, 1.

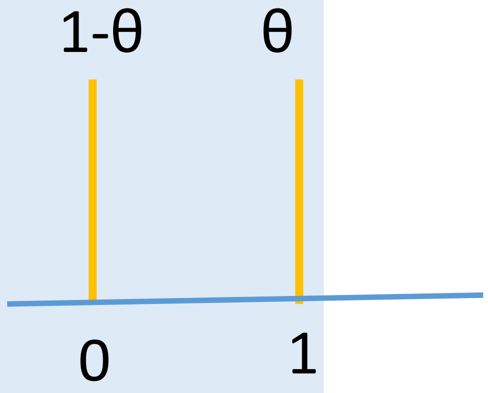
- Use the ML technique to find an estimator $\hat{\theta}$ for θ.
- Is this estimator unbiased?

SOLUTION: Form the likelihood function as a product of the marginal pdf's of $X_i, i = 1, 2, ...n$

$$L(x,\theta) = f(x_1,\theta) f(x_2,\theta)...f(x_n,\theta)$$

$$L(x,\theta) = \left[\theta^{x_1}(1-\theta)^{1-x_1}\right]\left[\theta^{x_2}(1-\theta)^{1-x_2}\right]...\left[\theta^{x_n}(1-\theta)^{1-x_n}\right] = \theta^{(x_1+x_2+...x_n)}(1-\theta)^{n-(x_1+x_2+...x_n)}$$

$$\ln L(x,\theta) = (x_1 + x_2 + ...x_n)\ln\theta + \left[n-(x_1+x_2+...x_n)\right]\ln(1-\theta)$$



Differentiate w.r.t to θ, set derivative to zero, we get

$$\frac{d}{d\theta}\ln L(x,\theta) = \frac{\sum_{i=1}^{n} X_i}{\theta} + \frac{-n+\sum_{i=1}^{n} X_i}{1-\theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Mean (X) = θ
Var (X) =θ(1-θ)

Now, we calculate the parameters of X_i.

$$E(X_i) = \mu_x = (0)(1-\theta) + (1)(\theta) = \theta \,;$$

Therefore, $E\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{n\theta}{n} = \theta \Rightarrow$ unbiaded estimator $\Rightarrow Bias = 0$

EXAMPLE: Given a random sample $X_1, X_2, ..., X_n$ of size (n) taken from a distribution X with pdf

$f(x) = (\alpha + 1)x^{\alpha}$, $0 < x < 1$. Use the ML technique to find an estimator for α.

SOLUTION: The likelihood function is

$$L(x,\alpha) = f(x_1,\alpha)f(x_2,\alpha)...f(x_n,\alpha)$$

$$L(x,\alpha) = (\alpha+1)x_1^{\alpha}...(\alpha+1)x_n^{\alpha} = (\alpha+1)^n x_1^{\alpha}...x_n^{\alpha}$$

$$\ln L(x,\alpha) = n\ln(\alpha+1) + \alpha\ln x_1 ... + \alpha\ln x_n$$

Differentiating with respect to α and setting the derivative to zero, we get

$$\frac{d}{d\alpha}\ln L(\alpha) = \frac{n}{\hat{\alpha}+1} + \ln x_1 ... + \ln x_n = 0$$

Solving for $\hat{\alpha}$ we get

$$\hat{\alpha} = \frac{n}{-\ln x_1 ... - \ln x_n} - 1 = \frac{1}{(-\sum_{i=1}^{n}\ln x_i)/n} - 1$$. **Estimator**, (note that $\ln x_i < 0$ since $0 < x < 1$)

Now, suppose that the random sample yield the observations $\{0.52, 0.6, 0.55, 0.58, 0.5\}$. Then, $\hat{\alpha}$

$$\hat{\alpha} = \frac{5}{-\ln 0.52 - \ln 0.6 - \ln 0.55 - \ln 0.58 - \ln 0.5} - 1 = \frac{5}{3} - 1 = \frac{2}{3}$$. (**Point estimate** of α)

EXAMPLE: Given a random sample $X_1, X_2, ..., X_n$ of size (n) taken from a Gaussian population with parameters μ_X and σ_X^2. Use the ML technique to find estimators for the cases:

a. The mean μ_X when the variance σ_X^2 is known.

b. The variance σ_X^2 when the mean μ_X is known.

Solution: Form the likelihood function as a product of the pdf's of the n Gaussian observations

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{\frac{-(x_i-\mu_X)^2}{2\sigma_X^2}} = \frac{e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2}}}{\left(2\pi\sigma_X^2\right)^{\frac{n}{2}}} \Rightarrow \ln(L) = -\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2} - \frac{n}{2}\ln\left(2\pi\sigma_X^2\right) \quad (1)$$

a- Take the derivative of (1) w.r.t μ_X (treating σ_X^2 as a constant), set derivative to 0.

$$\frac{\partial}{\partial\mu_X}\{\ln L\} = 0 \Rightarrow -\frac{2}{2\sigma_X^2}(-1)\sum_{i=1}^{n}(x_i-\mu_X) = 0 \Rightarrow \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}(\mu_X) = 0$$

$$\Rightarrow \hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n}x_i; \qquad \textbf{ML Estimator (unbiased)} \quad (2)$$

b. Take the derivative of (1) w.r.t σ_X^2 (*treating* μ_X *as a constant*), set the derivative to 0.

$$\frac{\partial}{\partial\sigma_X^2}\{\ln L\} = 0 \Rightarrow \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2$$

(Note the division by n since μ_X is known). In the previous lecture, we proved that this estimator is **unbiased**.

EXAMPLE: Given a random sample $X_1, X_2, \ldots, X_n$ of size (n) taken from a Gaussian population with parameters μ_X and σ_X^2. Use the ML technique to find estimators for the case when the mean μ_X and variance σ_X^2 are both assumed unknown.

Solution: Form the likelihood function as a product of the pdf's of the n Gaussian observations

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\,\pi\,\sigma_X^2}}\, e^{\frac{-(x_i-\mu_X)^2}{2\sigma_X^2}} = \frac{e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2}}}{\left(2\,\pi\,\sigma_X^2\right)^{\frac{n}{2}}} \Rightarrow \ln(L) = -\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2} - \frac{n}{2}\ln\left(2\,\pi\,\sigma_X^2\right) \quad (1)$$

Set $\frac{\partial}{\partial \mu_X}\ln L = 0$, $\frac{\partial}{\partial \sigma_X^2}\ln L = 0$ and solve for μ_X and σ_X^2. The result is

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

The ML for the mean is unbiased. However, the ML of the variance is biased since

$$E\left(\hat{\sigma}_{ML}^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2\right) = \left(\frac{n-1}{n}\right)\sigma_X^2;$$ An unbiased estimator is

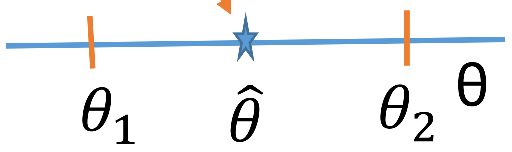
$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$. In the previous lecture, we proved that this estimator is **unbiased**

Interval Estimators for the Mean and Variance

An interval estimate of an unknown parameter (θ) is an interval of the form $\theta_1 \leq \theta \leq \theta_2$ where the end points θ_1 and θ_2 depend on the numerical value of the parameter to be estimated for a particular sample. From the sampling distribution of $(\hat{\theta})$ we will be able to determine values of θ_1 and θ_2 such that:

$$P(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha \;\; ; \; 0 < \alpha < 1$$

point estimate



$$P(\theta_1 \leq \theta \leq \theta_2) \geq 1 - \alpha \, ; \; 0 < \alpha < 1$$

where,

θ:	the unknown parameter
$(1-\alpha)$:	is the confidence coefficient
α:	is called the confidence level.
θ_1 and θ_2:	lower and the upper confidence limits on θ

In point estimation, we estimate the unknown parameter using a single number that is calculated from the sample data. $\hat{\theta} = f(X_1, X_2, \ldots, X_n)$

Confidence Interval on the Mean: (Variance Known)

Suppose that the population of interest, X, follows the Gaussian distribution $X \sim N(\mu_X, \sigma_X^2)$, μ_X is unknown and σ_X^2 is known.

A random sample of n measurements $X_1, X_2, \ldots, X_n$ is drawn from a Gaussian distribution with an unknown mean and a known variance. The objective is to construct 100(1-α)% confidence interval on the mean.

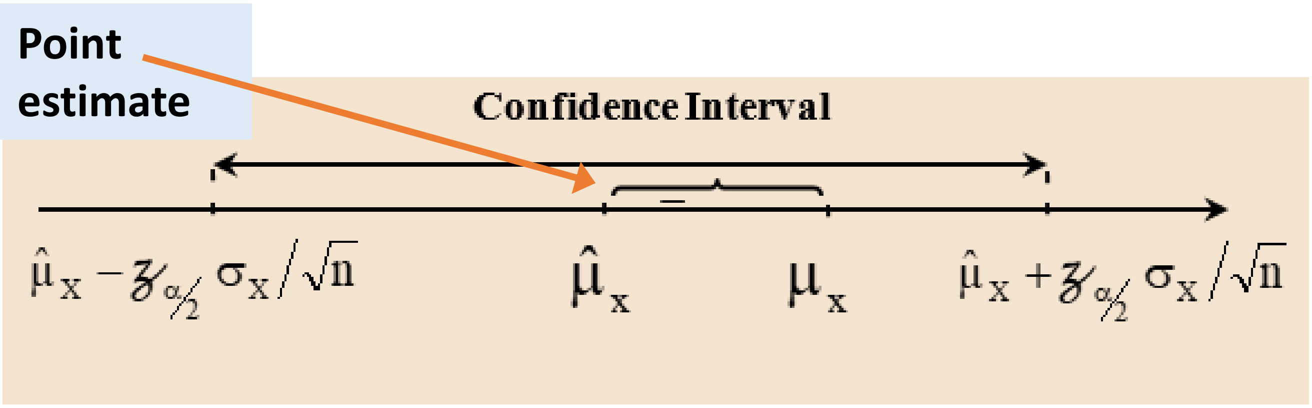
The sampling distribution of $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N(\mu_X, \sigma_X^2 / n)$

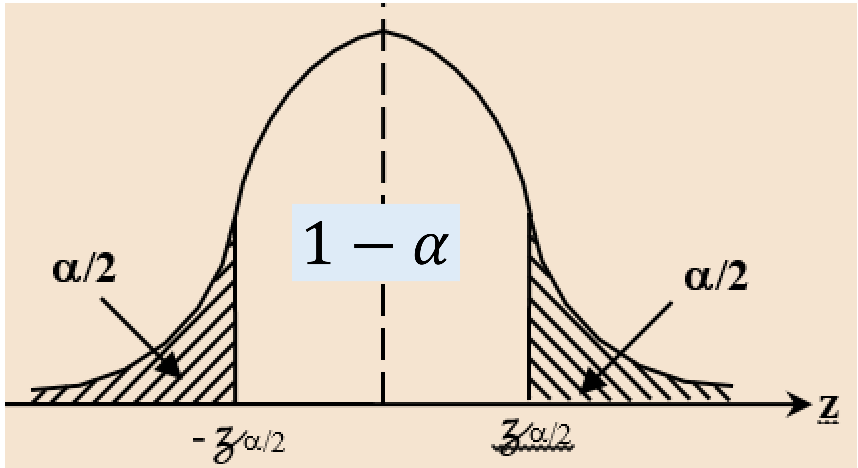
Therefore, the distribution of the statistic $Z = \dfrac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}} \sim N(0,1)$

$$P\{-z_{\alpha/2} \le Z \le z_{\alpha/2}\} = 1-\alpha \Rightarrow P\left\{-z_{\alpha/2} \le \frac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}} \le z_{\alpha/2}\right\} = 1-\alpha$$

$P\{\hat{\mu}_X - z_{\alpha/2}\,\sigma_X/\sqrt{n} \le \mu_X \le \hat{\mu}_X + z_{\alpha/2}\,\sigma_X/\sqrt{n}\} = 1-\alpha$; 100(1–α)% confidence interval on μ_X

where $z_{\alpha/2}$ is the upper 100(α/2)% point of the standard normal.





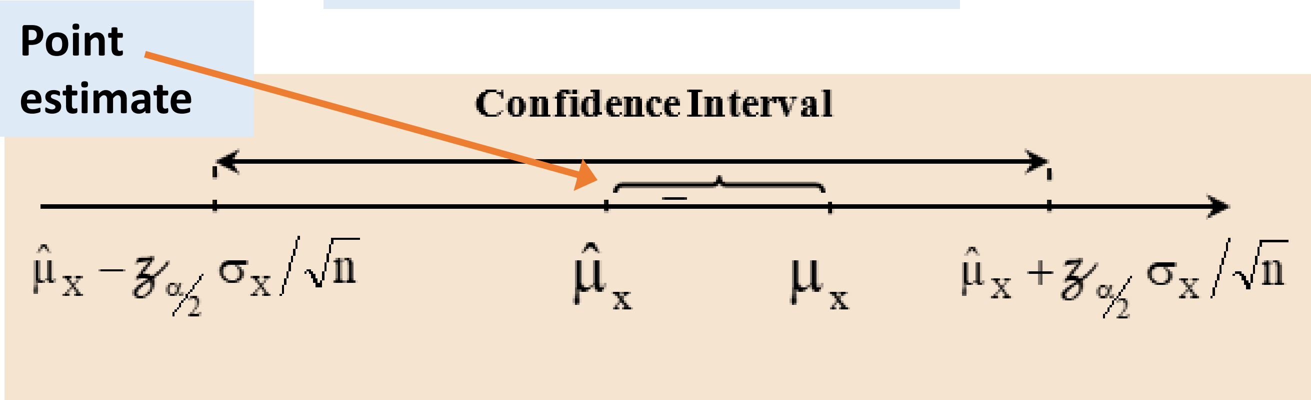
Choice of the Sample Size

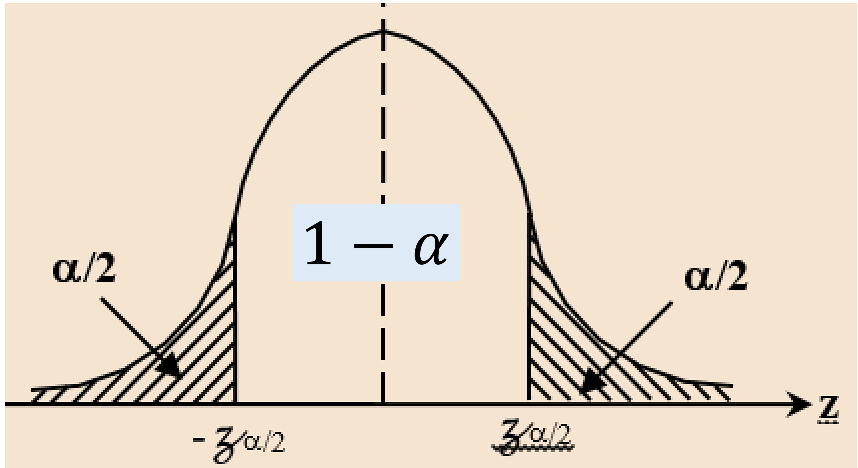
The definition above means that in using $\hat{\mu}_X$ to estimate μ_X, the error $E = |\hat{\mu}_X - \mu_X|$ is less than or equal to $z_{\alpha/2}\,\sigma_X/\sqrt{n}$ with confidence $100(1-\alpha)$. In situations where the sample size can be controlled, we can choose (n) so that we are $100(1-\alpha)\%$ confident that the error in estimating μ_X is less than a specified error (E).

n is chosen such that $E = z_{\alpha/2}\,\sigma_X/\sqrt{n} \Rightarrow n = \left(\dfrac{z_{\alpha/2}\sigma_X}{E}\right)^2$

$$P\{-z_{\alpha/2}\,\sigma_X/\sqrt{n} \le \hat{\mu}_X - \mu_X \le z_{\alpha/2}\,\sigma_X/\sqrt{n}\} = 1-\alpha$$

$$\mathrm{E} = |\hat{\mu}_X - \mu_X| \le z_{\alpha/2}\,\sigma_X/\sqrt{n}$$





EXAMPLE: The following samples are drawn from a population that is known to be Gaussian.

7.31	10.80	11.27	11.91	5.51	8.00	9.03	14.42	10.24	10.91

a. Find a 95% confidence interval on the mean if the variance of the population is known to be 4.

b. Find the sample size if we want to be 95% confident that the error is less than 0.2.

SOLUTION: Clearly, the sample size is n=10. A 95% confidence interval takes the form

$$P(\hat{\mu}_X - z_{\alpha/2}\,\sigma_X/\sqrt{n} \le \mu_X \le \hat{\mu}_X + z_{\alpha/2}\,\sigma_X/\sqrt{n}) = 1-\alpha$$

Here, the sample average is: $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94$

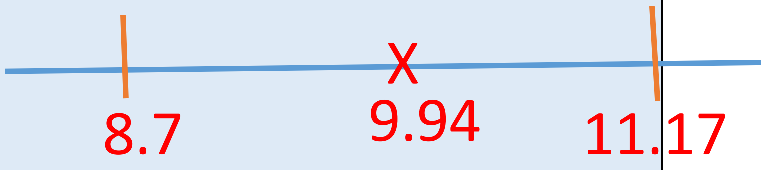
$E = |\hat{\mu}_X - \mu_X| \le z_{\alpha/2}\,\sigma_X/\sqrt{n}$

$\alpha = 0.05 \;\Rightarrow\; \alpha/2 = 0.025$. From the table of the Gaussian cumulative distribution function, we find that $\Phi(-1.96) = 0.025 \;\Rightarrow\; z_{\alpha/2} = 1.96$. The confidence interval is

$$P\left\{9.94 - \frac{1.96\times\sqrt{4}}{\sqrt{10}} \le \mu_X \le 9.94 + \frac{1.96\times\sqrt{4}}{\sqrt{10}}\right\} = 0.95 \;\Rightarrow\; P\{8.70 \le \mu_X \le 11.1796\} = 0.95$$

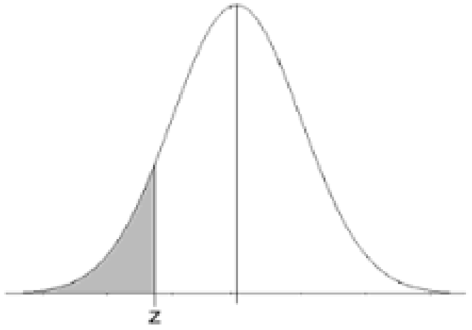
With n=10, we are 95% confident that the error is bounded by

$$E = z_{\alpha/2}\,\sigma_X/\sqrt{n} = 1.96*2/\sqrt{10} = 1.239$$



The sample size for n to have an error <0.2, is $n = \left(\frac{z_{\alpha/2}\sigma_X}{E}\right)^2 = \left(\frac{1.96*2}{0.2}\right)^2 = 385$

Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367

EXAMPLE: The rainfall in a region is normally distributed with a mean value μ and a standard deviation $\sigma = 25$ cm. Over a certain period, the following rain gauge readings (in cm) were collected

115.4	99.5	110.2	79.1	187.6	106.4	101.7	112.5	138.7	117.5	99.1	134.1

a- Find the width of a 95% confidence interval on the mean.

b- If the width of the 95% confidence interval on the mean is to be reduced to 10 cm, how large should the sample size be to get this result?

SOLUTION: Clearly, we have a sample of size n=12. A 95% confidence interval takes the form

$$P(\hat{\mu}_X - z_{\alpha/2}\,\sigma_X/\sqrt{n} \le \mu_X \le \hat{\mu}_X + z_{\alpha/2}\,\sigma_X/\sqrt{n}) = 1-\alpha$$

$$\mathrm{E} = |\hat{\mu}_X - \mu_X| \le z_{\alpha/2}\,\sigma_X/\sqrt{n}$$

Here, the sample average is: $\hat{\mu}_X = \frac{1}{12}\sum_{i=1}^{n} x_i = 116.81$

$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$. From the table of the Gaussian cumulative distribution function, we find that $\Phi(-1.96) = 0.025 \Rightarrow z_{\alpha/2} = 1.96$. The confidence interval is

$$P\left\{116.81 - \frac{(1.96)(25)}{\sqrt{12}} \le \mu_X \le 116.8167 + \frac{(1.96)(25)}{\sqrt{12}}\right\} = 0.95 \Rightarrow P\{113.98 \le \mu_X \le 119.64\} = 0.95$$

With n=12, we are 95% confident that the error is bounded by $\mathrm{E} = z_{\alpha/2}\,\sigma_X/\sqrt{n} = 1.96 * 25/\sqrt{12} = 14.14$

The sample size for n to have an error < 10, is $n = \left(\frac{z_{\alpha/2}\sigma_X}{E}\right)^2 = \left(\frac{1.96*25}{10}\right)^2 = 25$

Confidence Interval on the Mean: (Variance Unknown)

- Suppose that the population of interest has a normal distribution with unknown mean μ_X and an unknown variance σ_X^2.
- Estimating the mean, when the variance is known, was considered in the previous lecture. In that case, we used the statistic $Z = \dfrac{\hat{\mu}_X - \mu}{\sigma_X / \sqrt{n}}$ to establish the confidence interval.
- Note that with σ_X known, $Z = \dfrac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}} \sim N(0,1)$.

The number of independent pieces of information that go into the estimate of a parameter are called the degrees of freedom

- ***Definition:*** Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ_X and unknown variance σ_X^2. The quantity $T = \dfrac{\hat{\mu}_X - \mu}{\hat{\sigma}_X / \sqrt{n}}$ has a T-distribution with $k = n-1$ degrees of freedom, where $\hat{\sigma}_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$ is the sample variance and $\hat{\mu}_X = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean.

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow X_n = n\hat{\mu}_X - \sum_{j=1}^{n-1} X_j$$

- When the variance is unknown, the sample standard deviation used in the definition of T is no longer a constant, but rather a random variable. As such, T does not follow the Gaussian distribution.

The probability density function of the T-distribution is given here, without proof.

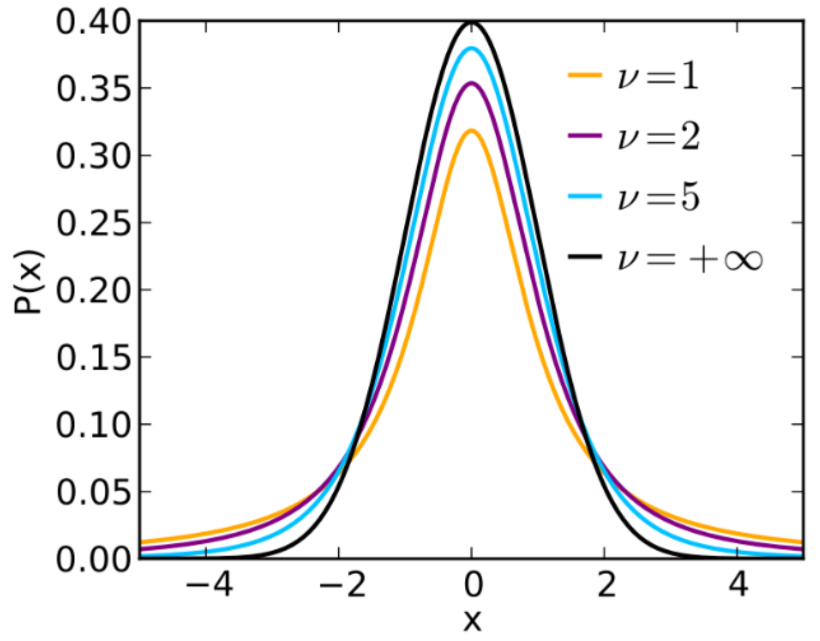
$$f_T(t) = \frac{\Gamma\left(\frac{(k+1)}{2}\right)}{\sqrt{\pi\, k}\;\Gamma\left(\frac{k}{2}\right)\left(\frac{t^2}{k+1}\right)^{\frac{k+1}{2}}} \qquad -\infty < t < \infty$$

$$T = \frac{\hat{\mu}_X - \mu}{\hat{\sigma}_X/\sqrt{n}}$$

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

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

- The T distribution is similar to the $Z \sim N(0,1)$ distribution in that it is symmetrical and bell-shaped, but more spread out (has a higher variance).
- The exact shape of the T distribution is determined by one parameter, **(k = n-1)**, called the "degrees of freedom. Therefore, there is a T distribution for each value of n.
- The mean of the t-distribution is zero and the variance $\sigma_T^2 = \frac{k}{k-2}$.
- As $n \to \infty$, the t-distribution converges to the normal distribution.
- The maximum is reached at the mean value.



A plot of the *t* probability density function for 4 different values of the degrees of freedom k = n-1, Wikipedia

Confidence on the mean when the variance is known takes the form

$$P(\hat{\mu}_X - z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \le \mu_X \le \hat{\mu}_X + z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}) = 1-\alpha$$

***Definition*: Confidence interval on the mean when the variance is unknown**

If $\hat{\mu}_X$ and $\hat{\sigma}_X$ are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ_X^2, the $100(1-\alpha)$ % confidence interval on μ_X is:

$$P\left\{\hat{\mu}_X - t_{\alpha/2,\,n-1}\frac{\hat{\sigma}_X}{\sqrt{n}} \le \mu_X \le \hat{\mu}_X + t_{\alpha/2,\,n-1}\frac{\hat{\sigma}_X}{\sqrt{n}}\right\} = 1-\alpha$$

where $t_{\alpha/2,n-1}$ is the upper $100(\alpha/2)$ % point of the T-distribution with $(n-1)$ degrees of freedom.

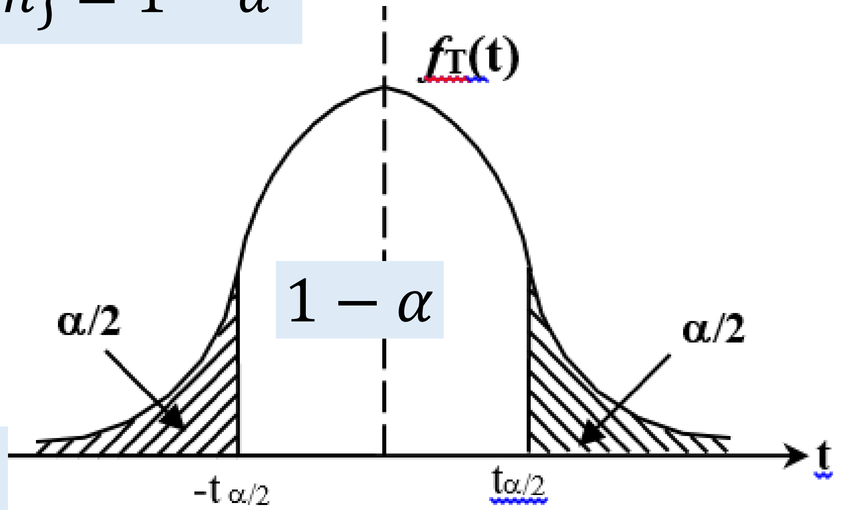
Derivation: $P\{-t_{\alpha/2,n-1} \le T \le t_{\alpha/2,n}$

$$P\{|\hat{\mu}_X - \mu_X| \le t_{\alpha/2,\mathrm{n}-1}\,\sigma_X/\sqrt{n}\} = 1-\alpha$$

$$P\left\{-t_{\alpha/2,n-1} \le \frac{\hat{\mu}_X - \mu_X}{\hat{\sigma}_X/\sqrt{n}} \le t_{\alpha/2,n-1}\right\} = 1-\alpha$$

$$P\left\{\hat{\mu}_X - t_{\alpha/2,\,n-1}\frac{\hat{\sigma}_X}{\sqrt{n}} \le \mu_X \le \hat{\mu}_X + t_{\alpha/2,\,n-1}\frac{\hat{\sigma}_X}{\sqrt{n}}\right\} = 1-\alpha$$

$$P\{-t_{\alpha/2,\mathrm{n}-1}\,\sigma_X/\sqrt{n} \le \hat{\mu}_X - \mu_X \le t_{\alpha/2,\mathrm{n}-1}\,\sigma_X/\sqrt{n}\} = 1-\alpha$$



EXAMPLE: For the following samples drawn from a normal population:

7.31	10.80	11.27	11.91	5.51	8.00	9.03	14.42	10.24	10.91

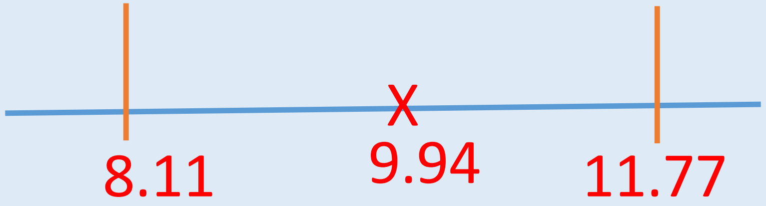
Find 95% confidence interval for the mean if the variance of the population is unknown.

SOLUTION: The sample size n = 10 and the degrees of freedom k = n-1=9.

The sample average and the sample variance are

$$P\{|\hat{\mu}_X - \mu_X| \leq 1.83\} = 0.95$$

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94; \qquad \hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 = 6.51$$



A 95% confidence interval on the mean takes the form

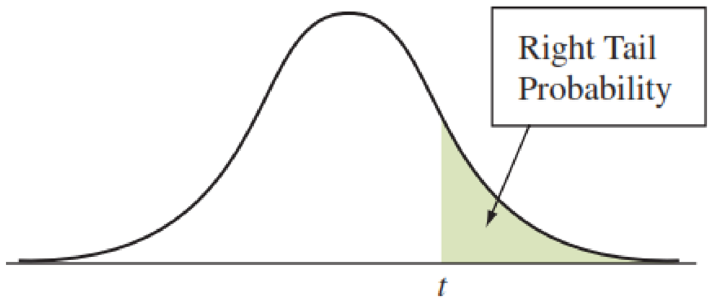
$$P\left\{\hat{\mu}_X - t_{\alpha/2\,,\,n\text{-}1}\frac{\hat{\sigma}_X}{\sqrt{n}} \leq \mu_X \leq \hat{\mu}_X + t_{\alpha/2\,,\,n\text{-}1}\frac{\hat{\sigma}_X}{\sqrt{n}}\right\} = 1-\alpha$$

From the table of the T-distribution, we can obtain $t_{\alpha/2,n-1} = t_{0.025,9} = 2.263$ as: Therefore,

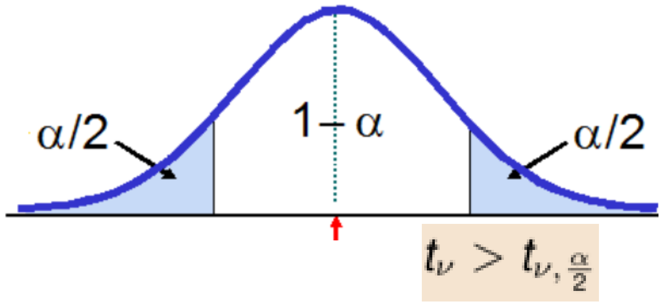
$$P\left\{9.94 - 2.263\frac{\sqrt{6.51}}{\sqrt{10}} \leq \mu_X \leq 9.94 + 2.263\frac{\sqrt{6.51}}{\sqrt{10}}\right\} = 0.95 \Rightarrow P\{8.11 \leq \mu_X \leq 11.77\} = 0.95$$

TABLE B t Distribution Critical Values

		Confidence Level					
		80%	90%	95%	98%	99%	99.8%
		Right-Tail Probability					
df=n-1	df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
	1	3.078	6.314	12.706	31.821	63.656	318.289
	2	1.886	2.920	4.303	6.965	9.925	22.328
	3	1.638	2.353	3.182	4.541	5.841	10.214
	4	1.533	2.132	2.776	3.747	4.604	7.173
	5	1.476	2.015	2.571	3.365	4.032	5.894
	6	1.440	1.943	2.447	3.143	3.707	5.208
	7	1.415	1.895	2.365	2.998	3.499	4.785
	8	1.397	1.860	2.306	2.896	3.355	4.501
	9	1.383	1.833	2.262	2.821	3.250	4.297
	10	1.372	1.812	2.228	2.764	3.169	4.144
	11	1.363	1.796	2.201	2.718	3.106	4.025
	12	1.356	1.782	2.179	2.681	3.055	3.930
	13	1.350	1.771	2.160	2.650	3.012	3.852



Shaded area $= \frac{\alpha}{2} = P(t_\nu > t_{\nu,\frac{\alpha}{2}})$



EXAMPLE: A civil engineer is testing the compressive strength of concrete. He tests 12 specimens and obtains the following data (in psi)

2216 2237 2249 2204 2225 2301 2281 2283 2318 2255 2275 2295

a. Find point estimates for the mean and variance of the strength
b. Construct a 95% confidence interval on the mean strength

SOLUTION: The sample size n = 12 and the degrees of freedom k = 12-1=11.

The sample average and the sample variance are

$$\hat{\mu}_X = \frac{1}{12}\sum_{i=1}^{n} x_i = 2261; \qquad \hat{\sigma}_X^2 = \frac{1}{11}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 = 1309 \Rightarrow \hat{\sigma}_X = 36.18$$

A 95% confidence interval on the mean takes the form

$$P\left\{\hat{\mu}_X - t_{\alpha/2\,,\,n\text{-}1}\frac{\hat{\sigma}_X}{\sqrt{n}} \le \mu_X \le \hat{\mu}_X + t_{\alpha/2\,,\,n\text{-}1}\frac{\hat{\sigma}_X}{\sqrt{n}}\right\} = 1-\alpha$$

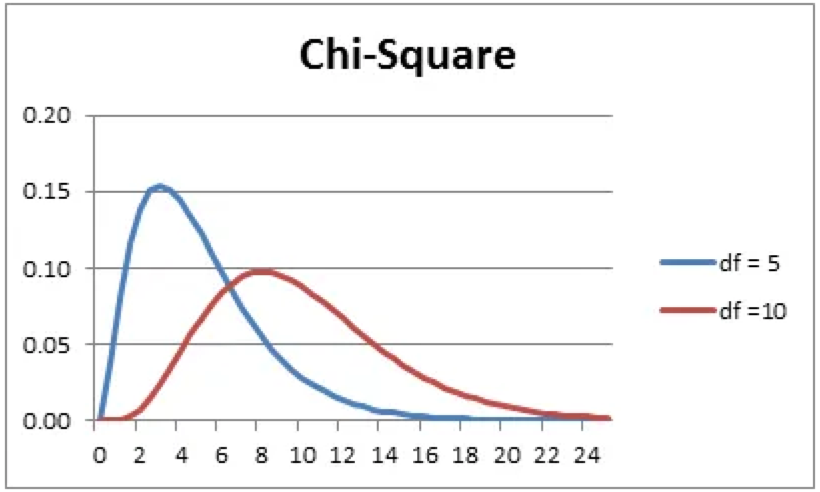
$$P\{|\hat{\mu}_X - \mu_X| \le 22\} = 0.95$$

From the table of the T-distribution, we can obtain $t_{\alpha/2,n-1} = t_{0.025,1} = 2.2$ as: Therefore,

$$P\left\{2261 - 2.2\frac{36.18}{\sqrt{12}} \le \mu_X \le 2261 + 2.2\frac{36.18}{\sqrt{12}}\right\} = 0.95 \Rightarrow P\{2238 \le \mu_X \le 2283\} = 0.95$$

Confidence Interval on the Variance of a Normal Population

- The χ^2 Distribution: Let $Z_1, Z_2, \ldots, Z_n$ be n independent and identically standard normal random variables (each with mean 0 and variance 1). The random variable χ^2 with n degrees of freedom is defined as:
- $\chi^2 = (Z_1)^2 + (Z_2)^2 + \ldots + (Z_n)^2$.
- Its mean and variance are $E(\chi^2) = n$, $Var(\chi^2) = 2n$
- This distribution is positive valued and skewed to the right.



On the next slide, we derive the pdf of the each component Z_i^2

The shape of the Chi-square distribution depends on the number of degrees of freedom

$$f_Z(z) = \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{z^2}{2}}$$

$$Y = Z^2 \Rightarrow f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{\frac{-y}{2}} \;;\; y \geq 0$$

$$E(Y) = 1, E(Z^2) = 1$$

$$\text{Var}(Y) = 2$$

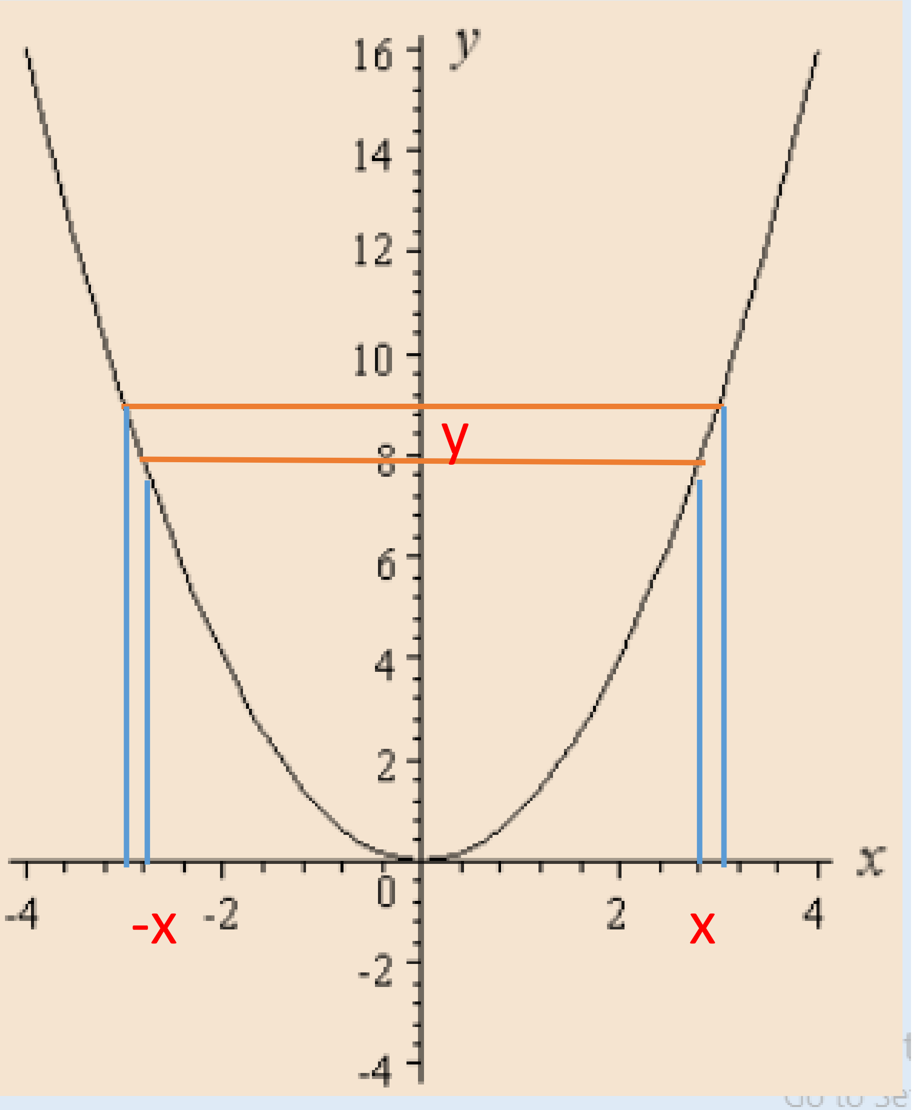
EXAMPLE: Let (X) be a Gaussian r.v with mean 0 variance 1. Define $Y = X^2$. Find $f_Y(y)$

SOLUTION: From the figure, we note that

$$P(y < Y < y + \Delta y) = 2P(x < X < x + \Delta x)$$

$$f_Y(y)\Delta y = 2 f_X(x)\Delta x$$

$$f_Y(y) = 2 f_X(x)\left|\frac{\Delta x}{\Delta y}\right| = \frac{2f_X(x)}{\left|\frac{\Delta y}{\Delta x}\right|} = \frac{2f_X(x)}{\left|\frac{dy}{dx}\right|} ; y \geq 0$$



Here, $y = x^2$; $\left|\frac{dy}{dx}\right| = |2x|$

$$f_Y(y) = \frac{2}{|2x|}\frac{1}{\sqrt{2\,\pi}}e^{\frac{-(x)^2}{2}} \quad , x = \sqrt{y}$$

$$f_Y(y) = \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{\frac{-y}{2}}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}e^{\frac{-y}{2}} \quad ; \quad y \geq 0$$

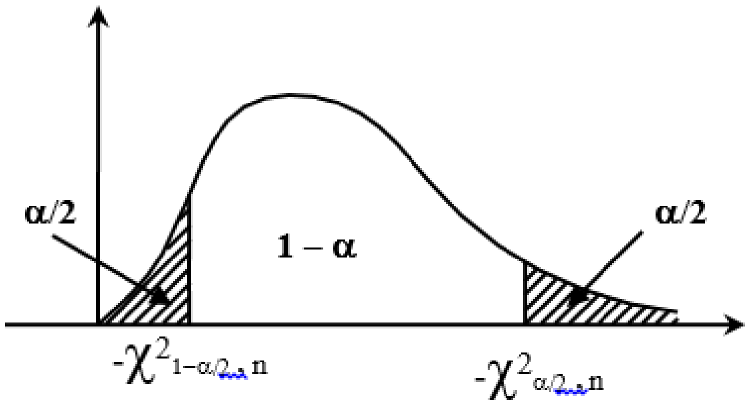
Confidence Interval on the Variance: (Mean Known)

- Let $X_1, X_2, ..., X_n$ be n independent and identically distributed normal random variables (each with mean μ_X and variance σ_X^2). The random variable χ^2 can be expressed as:

$$\chi^2 = \left(\frac{x_1 - \mu_X}{\sigma_X}\right)^2 + \left(\frac{x_2 - \mu_X}{\sigma_X}\right)^2 ...+ \left(\frac{x_n - \mu_X}{\sigma_X}\right)^2 = \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(x_i - \mu_X)^2$$

-

$$\chi^2 = \frac{n}{\sigma_X^2}\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2\right\} = \frac{n}{\sigma_X^2}\hat{\sigma}_X^2, \quad \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2$$



Confidence Interval: The confidence interval is constructed as

$$P\{\chi^2_{1-\alpha/2\,,\,n} \le \chi^2 \le \chi^2_{\alpha/2\,,\,n}\} = 1-\alpha \;\Rightarrow\; P\left\{\chi^2_{1-\alpha/2\,,\,n} \le \frac{n\,\hat{\sigma}_X^2}{\sigma_X^2} \le \chi^2_{\alpha/2\,,\,n}\right\} = 1-\alpha$$

$$P\left\{\frac{n\,\hat{\sigma}_X^2}{\chi^2_{\alpha/2\,,\,n}} \le \sigma_X^2 \le \frac{n\,\hat{\sigma}_X^2}{\chi^2_{1-\alpha/2\,,\,n}}\right\} = 1-\alpha$$

where $\chi^2_{\alpha/2\,,\,n}$ and $\chi^2_{1-\alpha/2\,,\,n}$ are the upper and lower 100(α/2)% points of the chi-square distribution with (n) degrees of freedom, respectively.

Confidence Interval on the Variance of a Normal Population: (Mean Unknown)

Let $X_1, X_2, \ldots, X_n$ be n independent and identically standard normal random variables (each with mean μ_X and variance σ_X^2). The random variable χ^2 with (n-1) degrees of freedom can be expressed as:

$$\chi^2 = \frac{n-1}{\sigma_X^2}\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2\right\} = \frac{n-1}{\sigma_X^2}\hat{\sigma}_X^2, \quad \hat{\sigma}_X^2 = \frac{1}{\text{n-1}}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$$

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

Definition: If $\hat{\sigma}_X^2$ is the sample variance from a random sample of (n) observations from a normal distribution with an unknown mean and an unknown variance σ_X^2, then a $100(1-\alpha)\%$ confidence interval on σ_X^2 is:

$$\frac{(n-1)\ \hat{\sigma}_X^2}{\chi^2_{\alpha/2\ ,\ \text{n-1}}} \le \sigma_X^2 \le \frac{(n-1)\ \hat{\sigma}_X^2}{\chi^2_{1-\alpha/2\ ,\ \text{n-1}}}$$

The number of independent pieces of information that go into the estimate of a parameter are called the degrees of freedom

where $\chi^2_{\alpha/2\ ,\ \text{n-1}}$ and $\chi^2_{1-\alpha/2\ ,\ \text{n-1}}$ is the upper and lower $100(\alpha/2)\%$ point of the chi-square distribution with (n – 1) degrees of freedom, respectively.

$\chi^2_{0.025,\,10} = 20.483$

$\chi^2_{0.975,\,10} = 3.247$

0.95 *confidence*

$\chi^2_{0.025\,,\,9} = 19.023$

$\chi^2_{0.975\,,\,9} = 2.7$

0.95 *confidence*

$\chi^2_{0.005,5} = 16.750$

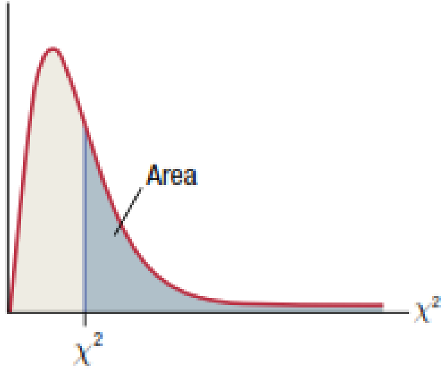
$\chi^2_{0.995,5} = 0.412$

0.99 *confidence*

$\chi^2_{0.05,\,11} = 19.675$

$\chi^2_{0.95,\,11} = 4.575$

0.90 *confidence*



Area to the Right of the Critical Value of χ^2

df	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718

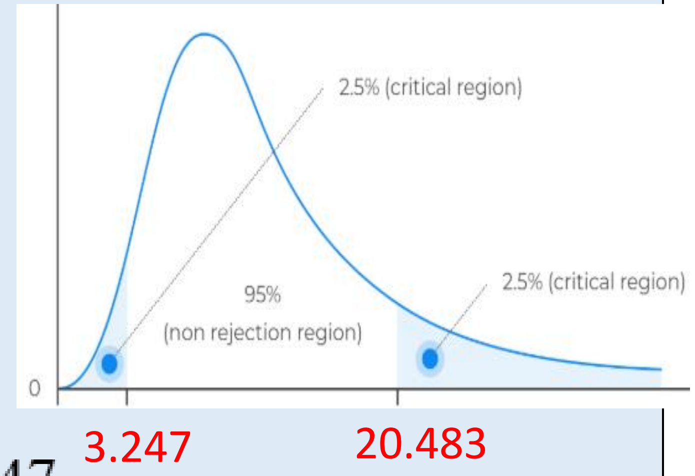
EXAMPLE: For the following samples drawn from a normal population:

7.31	10.80	11.27	11.91	5.51	8.00	9.03	14.42	10.24	10.91

Find 95% confidence interval for estimation of the variance if the mean of the population is known to be 10.

SOLUTION: From the sample we calculate the sample variance using the known mean of 10

$$\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2 = \frac{1}{10}\sum_{i=1}^{n}(x_i - 10)^2 = 5.866$$



From tables of χ^2 -distribution:

Number of degree of freedom = n = 10 (since the mean is known)

$$\alpha = 0.05 \rightarrow \alpha/2 = 0.025 \Rightarrow \chi^2_{0.025\,,\,10} = 20.483 \text{ and } \chi^2_{0.975\,,\,10} = 3.247$$

$$P\left\{\frac{n\,\hat{\sigma}_X^2}{\chi^2_{\alpha/2\,,\,n}} \le \sigma_X^2 \le \frac{n\,\hat{\sigma}_X^2}{\chi^2_{1-\alpha/2\,,\,n}}\right\} = 1-\alpha \Rightarrow P\left\{\frac{10\times 5.866}{20.483} \le \sigma_X^2 \le \frac{10\times 5.866}{3.247}\right\} = 0.95$$

$$P\left\{2.863 \le \sigma_X^2 \le 18.065\right\} = 0.95$$

EXAMPLE: For the following samples drawn from a normal population:

7.31	10.80	11.27	11.91	5.51	8.00	9.03	14.42	10.24	10.91

Find 95% confidence interval for estimation of the variance if the mean of the population is unknown.

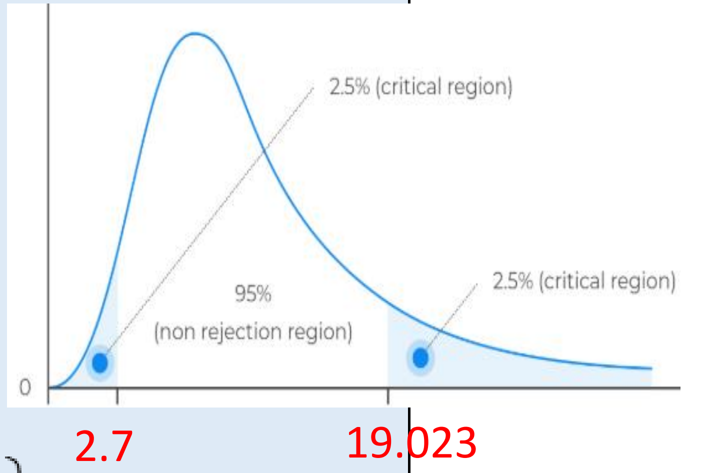
SOLUTION: From the sample we calculate the sample mean and sample variance:

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94 ; \quad \hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 = \frac{1}{9}\sum_{i=1}^{n}(x_i - 9.94)^2 = 6.51$$

From tables of χ^2 -distribution:

Number of degree of freedom = n = 10 – 1 = 9

$$\alpha = 0.05 \rightarrow \alpha/2 = 0.025 \Rightarrow \chi^2_{0.025\,,\,9} = 19.023 \text{ and } \chi^2_{0.975\,,\,9} = 2.7$$



$$P\left\{\frac{(n-1)\,\hat{\sigma}_X^2}{\chi^2_{\alpha/2\,,\,n\text{-}1}} \le \sigma_X^2 \le \frac{(n-1)\,\hat{\sigma}_X^2}{\chi^2_{1-\alpha/2\,,\,n\text{-}1}}\right\} = 1-\alpha \Rightarrow P\left\{\frac{9\times 6.51}{19.023} \le \sigma_X^2 \le \frac{9\times 6.51}{2.7}\right\} = 0.95$$

$$P\{3.0799 \le \sigma_X^2 \le 21.7\} = 0.95$$

EXAMPLE: A civil engineer is testing the compressive strength of concrete. He tests 12 specimens and obtains the following data (in psi)

2216 2237 2249 2204 2225 2301 2281 2283 2318 2255 2275 2295

a. Find point estimates for the mean and variance of the strength
b. Construct a 90% confidence interval on the variance of the strength

SOLUTION: The sample size n = 12 and the degrees of freedom k = 12-1=11.

The sample average and the sample variance are

$$\hat{\mu}_X = \frac{1}{12}\sum_{i=1}^{n} x_i = 2261; \qquad \hat{\sigma}_X^2 = \frac{1}{11}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 = 1309 \Rightarrow \hat{\sigma}_X = 36.18$$

A 90% confidence interval on the variance takes the form ($\alpha = 0.1$, $\alpha/2 = 0.05$)

$$P\left\{\frac{(n-1)\ \hat{\sigma}_X^2}{\chi^2_{\alpha/2,\, n\text{-}1}} \le \sigma_X^2 \le \frac{(n-1)\ \hat{\sigma}_X^2}{\chi^2_{1-\alpha/2\ ,\, n\text{-}1}}\right\} = 1-\alpha \ \Rightarrow\ P\left\{\frac{11\times1309}{19.675} \le \sigma_X^2 \le \frac{11\times1309}{4.575}\right\} = 0.95$$

$$P\left\{732 \le \sigma_X^2 \le 3147\right\} = 0.95 \ \Rightarrow\ P\left\{27.05 \le \sigma_X \le 56.098\right\}$$

Confidence Interval on a Binomial Proportion

The Binomial Distribution: A trial experiment is repeated n times under identical conditions. The probability of a success in any given trial is p. If X is the random variable representing the number of successes in the n trials, then X follows the binomial distribution

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}, x=0,1,\ldots,n$$

1	2	3	4	.	n
(S,F)	(S,F)	(S,F)	(S,F)	(S,F)	(S,F)

The mean and variance of X are:

$$\mu_X = E(X) = n\,p$$

$$Var(x) = \sigma_X^2 = n\,p\,(1-p)$$

$$\binom{n}{x}p^x(1-p)^{n-x} \simeq \frac{1}{\sqrt{2\pi np(1-p)}}e^{\frac{-(x-np)^2}{2np(1-p)}}$$

$$X \rightarrow N(np, np(1-p))$$

The estimator for p is $\hat{p} = X/n$

The mean and variance of $\hat{p} = X/n$ are

$$E(\hat{p}) = E(X/n) = np/p = p$$

Unbiased Estimator

$$Var(\hat{p}) = Var(X/n) = Var(X)/n^2 = np(1-p)/n^2 = p(1-p)/n$$

Consistent Estimator: Var →0 as n→ infinity (MVUE)

For large n, the central limit theorem applies and X can be approximated by a Gaussian distribution. Since the difference between X and $\hat{p} = X/n$ is a constant, then $\hat{p}$ can also be approximated by a Gaussian distribution (p is not too close to 0 or 1 and n is large; n p > 5 and n p (1 – p > 5).

To find a 100(1 – α)% confidence interval on the binomial proportion using the normal approximation, we construct the statistic:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \Rightarrow P\left\{-z_{\alpha/2} \le Z \le z_{\alpha/2}\right\} = 1 - \alpha$$

$$\boldsymbol{Z = (X - \mu_x)/\sigma_X \rightarrow N(0,1)}$$

$$P\left\{-z_{\alpha/2} \le \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right\} = 1 - \alpha \Rightarrow P\left\{\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\} = 1 - \alpha$$

The last equation expresses the upper and lower limits of the confidence interval in terms of the unknown parameter p.

Wald Method: This method replaces (p) by $\hat{p}$ in the lower and upper bounds

$$P\left\{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right\} = 1 - \alpha$$

This method is widely used, however careful study however reveals that it is flawed and inaccurate for a large range of n and p and is not recommended as a general method.

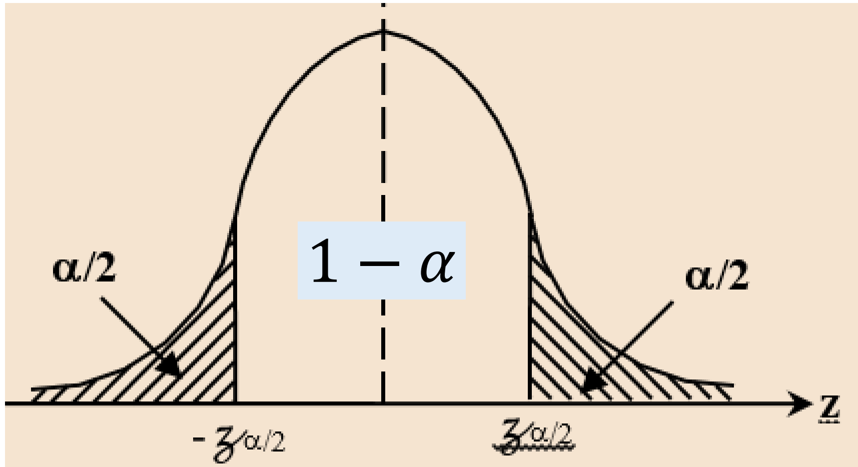
Wilson Score method: This has been suggested as an alternative method. It has been shown to be accurate for most parameter values.

It does not make the approximation as in the Ward method. Rather, it is a more complex method and involves solving a quadratic equation in p. The bounds are the roots of

$$\left|p-\hat{p}\right| = +z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

. The confidence interval on p is

$$p = \frac{\hat{p}+\dfrac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}+\dfrac{z_{\alpha/2}^2}{4n^2}}}{1+\dfrac{z_{\alpha/2}^2}{n}}$$



EXAMPLE: In a random sample of 85 automobile engine crankshafts bearings, 10 have a surface finish that is rougher than the specifications allow. Construct a 95% confidence interval for (p) using both the Wald method and the Wilson Score method.

Solution: $z_{\alpha/2} = z_{0.025} = 1.96$, $\hat{p} = \frac{X}{n} = \frac{10}{85} = 0.12$

Wald Method:

$$P\left\{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right\} = 1-\alpha$$

$$P\left\{0.12 - 1.96\sqrt{\frac{0.12(1-0.12)}{85}} \le p \le 0.12 + 1.96\sqrt{\frac{0.12(1-0.12)}{85}}\right\} = 0.95 \Rightarrow P\{0.05 \le p \le 0.19\} = 0.95$$

Wilson Score Method

$$p = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}} = \frac{0.12 + \frac{(1.96)^2}{2(85)} \pm 1.96\sqrt{\frac{0.12(1-0.12)}{85} + \frac{(1.96)^2}{4(85)^2}}}{1 + \frac{(1.96)^2}{85}}$$

$P\{0.101 \le p \le 0.1719\} = 0.95$ **Tighter upper and lower bounds**