BIRZEIT UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF ELECTRICAL ENGINEERING

ENEE 331

ENGINEERING PROBABILITY & STATISTICS

LECTURE NOTES

BY

DR. WA'EL HASHLAMOUN

JUNE, 2009

CHAPTER I

FUNDAMENTAL CONCEPTS OF PROBABILITY

Basic Definitions:

We start our treatment of probability theory by introducing some basic definitions.

Experiment:

By an experiment, we mean any procedure that:

1- Can be repeated, theoretically, an infinite number of times.

2- Has a well-defined set of possible outcomes.

Sample Outcome:

Each of the potential eventualities of an experiment is referred to as a sample outcome(s).

Sample Space:

The totality of sample outcomes is called the sample space (S).

Event:

Any designated collection of sample outcomes, including individual outcomes, the entire sample space and the null space, constitute an event.

Occur:

An event is said to occur if the outcome of the experiment is one of the members of that event.

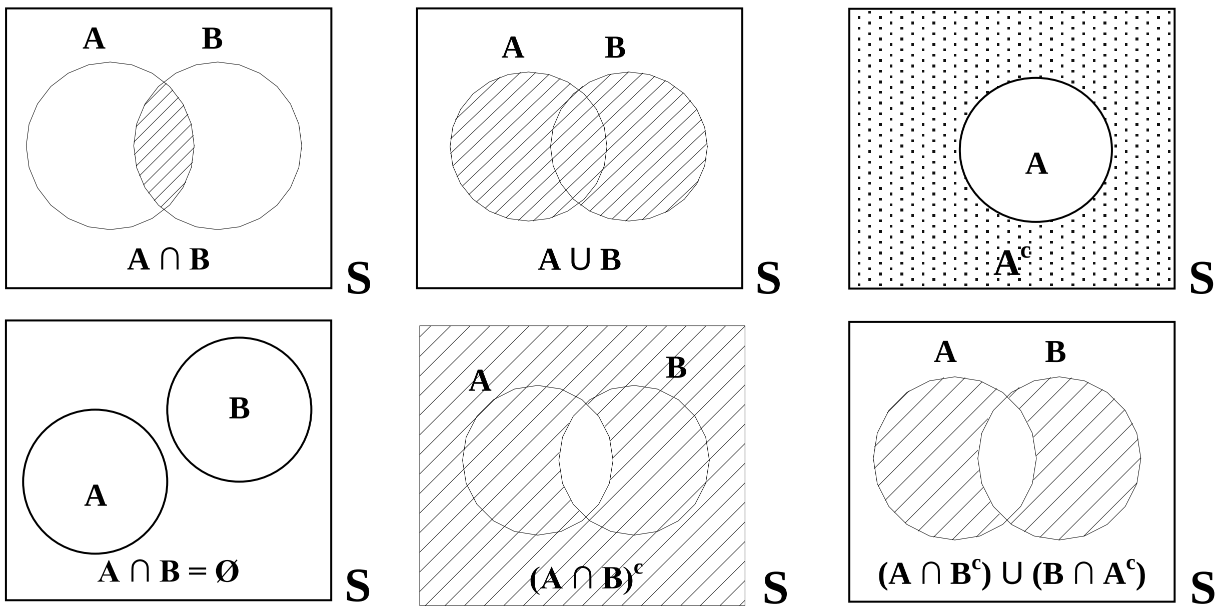
EXAMPLE (2-1):
Consider the experiment of flipping a coin three times.
a- What is the sample space?
b- Which sample outcomes make up the event:
A : Majority of coins show heads.

SOLUTION:
a- Sample Space (S) = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}
b- A = {HHH, HHT, HTH, THH}

Algebra of Events:

Let A and B be two events defined over the sample space S, then:

- The *intersection* of A and B, ($A \cap B$), is the event whose outcome belongs to both A and B.
- The *union* of A and B, ($A \cup B$), is the event whose outcome belongs to either A or B or both.
- Events A and B are said to be *Mutually Exclusive* (or disjoint) if they have no outcomes in common, that is $A \cap B = \emptyset$, where $\emptyset$ is the null set (a set which contains no outcomes).
- The *complement* of A (A^c or $\bar{A}$) is the event consisting of all outcomes in S other than those contained in A.
- *Venn Diagram* is a graphical format often used to simplify the manipulation of complex events.



- **De Morgan's Laws:**

  Use Venn diagrams to show that:

  1- $(A \cap B)^c = A^c \cup B^c$

  2- $(A \cup B)^c = A^c \cap B^c$

EXAMPLE (2-2):

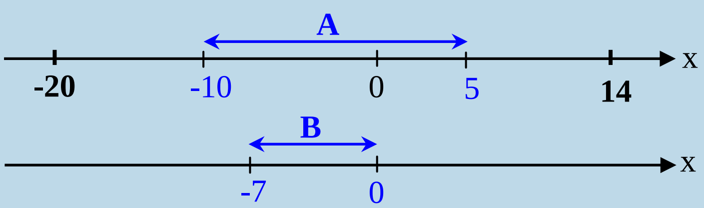
An experiment has its sample space specified as:
S = {1, 2, 3, ……, 48, 49, 50}. Define the events
A : set of numbers divisible by 6
B : set of elements divisible by 8
C : set of numbers which satisfy the relation 2^n , n = 1, 2, 3,…
Find: 1- A, B, C 2- A ∪ B ∪ C 3- A ∩ B ∩ C

SOLUTION:

1- Events A, B, and C are:
A = {6, 12, 18, 24, 30, 36, 42, 48}
B = {8, 16, 24, 32, 40, 48}
C = {2, 4, 8, 16, 32}
2- A ∪ B ∪ C = {6, 12, 18, 24, 30, 36, 42, 48,
8, 16, 32, 40,
2, 4}
3- A ∩ B ∩ C = { Ø }

EXAMPLE (2-3):

The sample space of an experiment is:
$S = \{-20 \le x \le 14\}$. If $A = \{-10 \le x \le 5\}$ and $B = \{-7 \le x \le 0\}$ find.
1- A ∪ B 2- A ∩ B



SOLUTION:

1- $A \cup B = \{-10 \le x \le 5\}$
2- $A \cap B = \{-7 \le x \le 0\}$

- **Definitions of Probability:**

Four definitions of probability have evolved over the years:

- ***Definition I: Classical (a priori)***

If the sample space S of an experiment consists of finitely many outcomes (points) that are equally likely, then the probability of event A, P(A) is:

$$P(A) = \frac{\text{Number of outcomes in A}}{\text{Number of outcomes in S}}$$

Thus in particular, P(S) = 1

- ***Definition II: Relative Frequency (a posteriori)***

Let an experiment be repeated (n) times under identical conditions then, the relative frequency:

$$P(A) = \lim_{n\to\infty} \frac{f(A)}{n} = \frac{\text{Number of times A occurs}}{\text{Number of trials}}$$

f(A) is called the frequency of (A)

Clearly $0 \le \frac{f(A)}{n} \le 1$

$\frac{f(A)}{n} = 0$ if (A) does not occur in the sequence of trials

$\frac{f(A)}{n} = 1$ if (A) occurs on each of the (n) trials

EXAMPLE (2-4):

In digital data transmission, the bit error probability is (p). If 10,000 bits are transmitted over a noisy communication channel and 5 bits were found to be in error, find the bit error probability (p).

SOLUTION:

According to the relative frequency definition we can estimate (p) as: $(p) = \frac{5}{10{,}000}$

- ***Definition III: Subjective***

Probability is defined as a person's measure of belief that some given event will occur.

Example:

What is the probability of establishing an independent Palestinian state in the next 2 years?
Any number we might come up with would be our own personal (subjective) assessment of the situation.

- ***Definition IV: Axiomatic***

Given a sample space (S), with each event (A) of (S) (subset of S) there is associated a number P(A), called the probability of (A), such that the following axioms of probability are satisfied:

1- $P(A) \ge 0$; Probability is nonnegative

2- $P(S) = 1$; Probability of the sample space is a certain

3- For the mutually exclusive events (A) and (B) $(A \cap B = \varnothing)$
$P(A \cup B) = P(A) + P(B)$; $(A \cap B = \varnothing)$

4- If (S) is infinite (has infinitely many points), axiom (3) is to be replaced by:
$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots$
where $A_1, A_2, A_3 \ldots\ldots$ are mutually exclusive events
$(A_1 \cap A_2 = \varnothing \quad A_1 \cap A_3 = \varnothing \quad A_2 \cap A_3 = \varnothing \quad \ldots\ldots)$

- **Basic Theorems for Probability:**

1- $P(A^c) = 1 - P(A)$

Proof: $S = A \cup A^c$

$P(S) = P(A) + P(A^c)$

$1 = P(A) + P(A^c) \quad \Rightarrow \quad P(A^c) = 1 - P(A)$

2- $P(\emptyset) = 0$

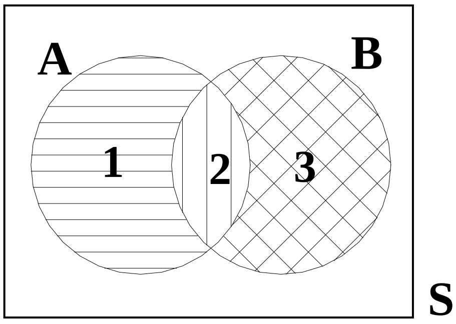
Proof:

$S = S \cup S^c$

$S = S \cup \emptyset$; $S^c = \emptyset$

$P(S) = P(S) + P(\emptyset) \quad \Rightarrow \quad P(\emptyset) = 0$

3- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Proof:

For events (A) and (B) in a sample space:

$\{A \cup B\} = \{A \cap B^c\} \cup \{A \cap B\} \cup \{B \cap A^c\}$

$\{A \cup B\} = \quad \mathbf{1} \quad \cup \quad \mathbf{2} \quad \cup \quad \mathbf{3}$

Where events (1) and (2) and (3) are mutually exclusive

$P(A \cup B) = P(1) + P(2) + P(3)$

$P(A) = P(1) + P(2)$

$P(B) = P(2) + P(3)$

$\Rightarrow P(A \cup B) = \{P(1) + P(2)\} + \{P(2) + P(3)\} - \{P(2)\}$

$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- ***Theorem:***

If A, B, and C are three events, then:

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

EXAMPLE (2-5):

One integer is chosen at random from the numbers {1, 2, ……, 50}. What is the probability that the chosen number is divisible by 6? Assume all 50 outcomes are equally likely.

SOLUTION:

$S = \{1, 2, 3, \ldots\ldots\ldots\ldots, 50\}$

$A = \{6, 12, 18, 24, 30, 36, 42, 48\}$

$$P(A) = \frac{\text{Number of elements in A}}{\text{Number of elements in S}} = \frac{8}{50}$$

EXAMPLE (2-6):

If the probability of occurrence of an even number is twice as likely as that of an odd number in Example (2-5). Find P(A); A is defined above.

SOLUTION:

$P(S) = P(even) + P(odd) = 1$;

Let (P) be the probability of occurrence of an odd number,

then (2P) will be the probability of occurrence of an even number.

$(25)(2P) + (25)(P) = 1$

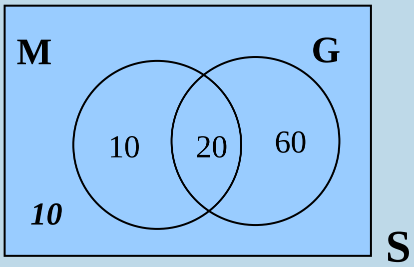
$(50 + 25)(P) = 1 \;\rightarrow\; P = \dfrac{1}{75}$

$P(A) = 8 \times 2P = \dfrac{16}{75}$

EXAMPLE (2-7):

Suppose that a company has 100 employees who are classified according to their marital status and according to whether they are college graduates or not. It is known that 30% of the employees are married, and the percent of graduate employees is 80%. Moreover, 10 employees are neither married nor graduates. What proportion of married employees are graduates?

SOLUTION:



Let: M : set of married employees
G : set of graduate employees
N(.) : number of members in any set (.)

$\rightarrow$ $N(S) = 100$
$N(M) = 0.3 \times 100 = 30$
$N(G) = 0.8 \times 100 = 80$
$N(M \cup G)^c = 10$
$\rightarrow$ $N(M \cup G) = 100 - 10 = 90$
$N(M \cup G) = N(M) + N(G) - N(M \cap G)$
$90 = 30 + 80 - N(M \cap G)$
$N(M \cap G) = 30 + 80 - 90 = 20$
$\rightarrow$ Two third of the married employees in the company are graduates.

EXAMPLE (2-8):

An experiment has two possible outcomes; the first occurs with probability (P), the second with probability (P^2), find (P).

SOLUTION:

$P(S) = 1$
$P + P^2 = 1$
$P^2 + P - 1 = 0$

$$P = \frac{-1 + \sqrt{5}}{2} \quad ; \qquad \text{(only the positive root is taken)}$$

EXAMPLE (2-9):
A sample space "S" consists of the integers 1 to 6 inclusive. Each outcome has an associated probability proportional to its magnitude. If one number is chosen at random, what is the probability that an even number appears?

SOLUTION:
Sample Space "S" = {1 , 2 , 3 , 4 , 5 , 6}
Event (A) = {2 , 4 , 6}
$P(A) = P(2) + P(4) + P(6)$

$$P(S) = 1 = \sum_{i=1}^{6} p(i) = \sum_{i=1}^{6} \alpha(i) = \frac{6(6+1)}{2}\alpha = 1$$

➔ The proportionality constant $\alpha = \frac{1}{21}$

$$P(A) = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{12}{21}$$

EXAMPLE (2-10):

Let (A) and (B) be any two events defined on (S). Suppose that P(A) = 0.4, P(B) = 0.5, and P(A ∩ B) = 0.1.

Find the probability that:

1- (A) or (B) but not both occur.
2- None of the events (A) or (B) will occur.
3- At least one event will occur.
4- Both events occur.

SOLUTION:

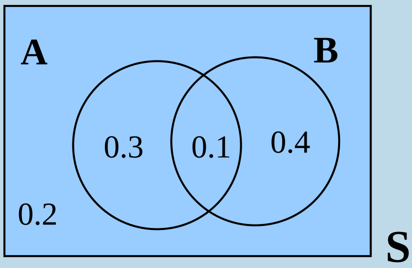
$P(A) = P[(A \cap B^c) \cup (A \cap B)]$

Using Venn diagram:

P(A) only = 0.3

P(B) only = 0.4

1- P(A or B only) = 0.3 + 0.4 = 0.7



Note that:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) = 0.4 + 0.5 - 0.1 = 0.8$

2- P(none) = $P(\overline{A \cup B}) = 0.2$

3- P(at least one) = $P(A \cup B) = 0.8$

4- P(both) = $P(A \cap B) = 0.1$

- **Discrete Probability Functions:**

If the sample space generated by an experiment contains either a *finite* or a *countable infinite* number of outcomes, then it is called a *discrete* sample space.

Any probability assignment on that space such that:

a- $P(s_i) \geq 0$

b- $\sum_{s_i \in S} P(s_i) = 1$

is said to be a discrete probability function.

If (A) is an event defined on (S), then $P(A) = \sum_{s_i \in A} P(s_i)$

For example, the sample space, “S” = {1 , 2 , 3 , 4 , 5 , 6} is countably finite, while the set of positive integers, “S” = {1 , 2 , 3 , ……...} is countably infinite.

EXAMPLE (2-11):

The outcome of an experiment is either a success with probability p or a failure with probability (1-p). If the experiment is to be repeated until a success comes up for the first time. Let X be the number of times the experiment is performed then the discrete probability function for the countably infinite sample space is

$P(x) = p(1-p)^{x-1}$; x = 1, 2, …
What is the probability that a success occurs on an an odd-numbered trial?

SOLUTION:
The sample space for the experiment is S={1, 2, 3, ….}
Let A be the event that a success occurs on an odd numbered trial. Then A consists of the sample points: A = {1, 3, 5, …}
P(A) = P(1) + P(3) + P(5) + …
$P(A) = p(1-p)^{1-1} + p(1-p)^{3-1} p(1-p)^{5-1} + ...$
$P(A) = p(1+p^2+p^4+...) = \frac{p}{1-p^2}$, by virtue of the geometric series $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$
In the special case when p = ½, P(A) becomes
➔ P(A) = $\frac{1}{2} \times \frac{1}{1-\frac{1}{4}}$ ➔ P(A) = $\frac{2}{3}$

EXAMPLE (2-12):

The discrete probability function for the countably infinite sample space S = {1, 2, 3, …} is:
$P(x) = \frac{C}{x^2}$; x = 1, 2, 3, ……
a- Find the constant "C" so that P(x) is valid discrete probability function.
b- Find the probability that the outcome of the experiment is a number less than 4.

SOLUTION:

a. By Axiom 2, P(S) = 1
$\sum_{x=1}^{\infty} \frac{C}{x^2} = 1$ ➔ $C \sum_{x=1}^{\infty} \frac{1}{x^2} = 1$ ➔ $C \frac{\pi^2}{6} = 1$, ➔ $C = \frac{6}{\pi^2}$

b. The event "A" is A = {1 , 2 , 3}
P(A) = P(1) + P(2) + P(3)
➔ $P(A) = \frac{6}{\pi^2}\left(\frac{1}{(1)^2} + \frac{1}{(2)^2} + \frac{1}{(3)^2}\right) = \frac{49}{6\pi^2} = 0.827$

- **Continuous Probability Functions:**

If the sample space associated with an experiment is an interval of real numbers, then (S) has an *uncountable infinite number* of points and (S) is said to be *continuous*.

Let $f(x)$ be a real-valued function defined on (S) such that:
a- $f(x) \geq 0$
b- $\int_{\text{All x}} f(x)\, dx = 1$
The function f(x) that satisfies these conditions is called a continuous probability function.
If (A) is an event defined on (S), then $P(A) = \int_{x \in A} f(x)\, dx$
For example, the sample space S = $\{1 \leq x \leq 2\}$ is uncountably infinite.

EXAMPLE (2-13):
Let the sample space of an experiment be:

"S" = $\{1 \le x \le 2\}$. The continuous probability function defined over "S" is:

$$f(x) = \frac{k}{x^2}\ , \quad 1 \le x \le 2 .$$

a- Find (k) so that $f(x)$ is a valid probability function.

b- Find $P(x \le 1.5)$

SOLUTION:

a- $P(S) = \int_1^2 f(x)\,dx = 1 \Rightarrow \int_1^2 \frac{k}{x^2}\,dx = 1 \ \Rightarrow k = 2$

b- $P(x \le 1.5) = \int_1^{1.5} \frac{k}{x^2}\,dx = \frac{2}{3}$

EXAMPLE (2-14):

The length of a pin that is a part of a wheel assembly is supposed to be 6 cm. The machine that stamps out the parts makes them 6 + x cm long, where x varies from pin to pin according to the probability function:

$f(x) = k(x + x^2)$; $0 \le x \le 2$

where (k) is a constant. If a pin is longer than 7 cm, it is unusable. What proportion of pins produced by this machine will be unusable?

SOLUTION:

$$P(S) = \int f(x)\,dx = 1$$

$$k\int_0^2 (x + x^2)\,dx = 1$$

$$k\left[\frac{x^2}{2} + \frac{x^2}{2}\right]_0^2 = 1 \ \Rightarrow k = \frac{6}{28}$$

A cotter pin is not accepted if the error $x \ge 1$cm.

$$P(x \ge 1) = \int_1^2 k(x + x^2)\,dx$$

$$= k\left[\frac{x^2}{2} + \frac{x^2}{2}\right]_1^2 = \frac{6}{28}\left[\frac{4}{2} + \frac{8}{3} - \frac{1}{2} - \frac{1}{3}\right] = \frac{23}{28}$$

$P(x \ge 1)$ = p(pin length ≥ 7 cm) = $\frac{23}{28}$

Conditional Probabilities and Statistical Independence:

- ***Definition:***

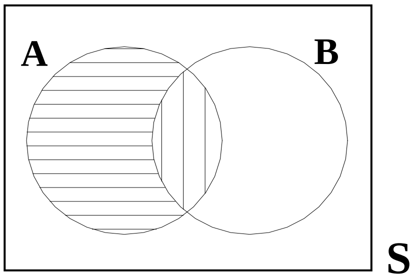
Given two events (A) and (B) with P(A) and P(B) > 0. We define the Conditional Probability of (A) given (B) has occurred as:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

and the probability of (B) given (A) has occurred as:

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

In (2), (A) serves as a new (reduced) sample space
and P(B/A) is the fraction of (A) which corresponds to (A ∩ B).



EXAMPLE (2-15):

A sample space (S) consists of the integers 1 to n inclusive. Each has an associated probability proportional to its magnitude. One integer is chosen at random, what is the probability that number 1 is chosen given that the number selected is in the first (m) integers.

SOLUTION:

Let (A) be the event "number 1" occurs

(A) = {1}

(B) the event "outcome belongs to the first m integers"

(B) = {1 , 2 , 3 , … , m}

$$\sum_{i=1}^{n} P_i = \sum_{i=1}^{n} \alpha\, i = 1 \Rightarrow \alpha \sum_{i=1}^{n} i = 1 \;\rightarrow\; \alpha \frac{n(n+1)}{2} = 1 \Rightarrow \alpha = \frac{2}{n(n+1)}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(1)}{P(B)} = \frac{\alpha}{\sum_{i=1}^{m} P_i} = \frac{\alpha}{\alpha \sum_{i=1}^{m} i} = \frac{1}{\frac{m(m+1)}{2}} \;\rightarrow\; P(A/B) = \frac{2}{m(m+1)}$$

A priori probability: $P(A) = \frac{2}{n(n+1)}$

A posteriori probability: $P(A/B) = \frac{2}{m(m+1)}$

Clearly P(A/B) > P(A) due to the additional information given by event (B).

Theorem: Multiplication Rule

If (A) and (B) are events in a sample space (S) and P(A) ≠ 0, P(B) ≠ 0, then:

P(A ∩ B) = P(A) P(B/A) = P(B) P(A/B)

For three events A, B, and C:

P(A ∩ B ∩ C) = P(A) P(B/A) P(C/B,A)

EXAMPLE (2-16):

A certain computer becomes inoperable if two components A and B both fail. The probability that A fails is 0.001 and the probability that B fails is 0.005. However, the probability that B fails increases by a factor of 4 if A has failed. Calculate the probability that:

a- The computer becomes inoperable.
b- A will fail if B has failed.

SOLUTION:

P(A) = 0.001

$P(B) = 0.005$

$P(B/A) = 4 \times 0.005 = 0.020$

a- The system fails when both A and B fail, i.e.,

$$P(A \cap B) = P(A)\,P(B/A)$$

$$P(A \cap B) = 0.001 \times 0.020 = 0.00002$$

b- $P(A \cap B) = P(A)\,P(B/A) = P(B)\,P(A/B)$

$$P(A/B) = \frac{0.001 \times 0.020}{0.005} = 0.004$$

EXAMPLE (2-17):

A box contains 20 non-defective (N) items and 5 defective (D) items. Three items are drawn without replacement.

a. Find the probability that the sequence of objects obtained is (NND) in the given order.

b. Find the probability that exactly one defective item is obtained.

SOLUTION:

a. $P(N \cap N \cap D) = P(N) \times P(N/N) \times P(D/N,N)$

$$P(NND) = \left(\frac{20}{25}\right)\left(\frac{20-1}{25-1}\right)\left(\frac{5}{25-2}\right) = \left(\frac{20}{25}\right)\left(\frac{19}{24}\right)\left(\frac{5}{23}\right)$$

b. One defective item is obtained, when any one of the following sequences occurs:

(NND), (NDN), (DNN)

The probability of getting one defective item is the sum of the probabilities of these sequences and is given as:

$$\left(\frac{20}{25}\right)\left(\frac{19}{24}\right)\left(\frac{5}{23}\right) + \left(\frac{20}{25}\right)\left(\frac{5}{24}\right)\left(\frac{19}{23}\right) + \left(\frac{5}{25}\right)\left(\frac{20}{24}\right)\left(\frac{19}{23}\right) = (3)\left(\frac{20}{25}\right)\left(\frac{19}{24}\right)\left(\frac{5}{23}\right)$$

Later in Chapter 2, we will see that Part (b) can be solved using the *hyper-geometric distribution.*

Definition: Statistical Independence

Two events (A) and (B) are said to be statistically independent if:

$P(A \cap B) = P(A)\,P(B)$

From this definition we conclude that:

$$P(A/B) = \frac{P(A)\,P(B)}{P(B)} = P(A) \Rightarrow \text{a posteriori probability} = \text{a priori probability}$$

$$P(B/A) = \frac{P(A)\,P(B)}{P(A)} = P(B)$$

This means that the probability of (A) does not depend on the occurrence or nonoccurrence of (B) and vice versa. Hence, the given information does not change our initial perception about the two given probabilities.

Independence of Three Events:

Events (A), (B) and (C) are independent if the following conditions are satisfied:

- $P(A \cap B) = P(A)\,P(B)$
- $P(A \cap C) = P(A)\,P(C)$
- $P(B \cap C) = P(B)\,P(C)$
- $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

EXAMPLE (2-18):

Let $S = \{1 , 2 , 3 , 4\}$; $P_i = \frac{1}{4}$. $A = \{1 , 2\}$ and $B = \{2 , 3\}$. Are (A) and (B) independent?

SOLUTION:

$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$ ➔ $(A \cap B) = \{2\}$, $P(A \cap B) = \frac{1}{4}$

➔ $P(A \cap B) = P(A)\, P(B)$ ➔ ***Events are independent***

EXAMPLE (2-19):

Consider an experiment in which the sample space contains four outcomes $\{S_1, S_2, S_3, S_4\}$ such that $P(S_i) = \frac{1}{4}$. Let events (A), (B) and (C) be defined as:

$A = \{S_1, S_2\}$, $B = \{S_1, S_3\}$, $C = \{S_1, S_4\}$

Are these events independent?

SOLUTION:

$P(A) = P(B) = P(C) = \frac{1}{2}$

$(A \cap B) = \{S_1\}$; $(A \cap C) = \{S_1\}$; $(B \cap C) = \{S_1\}$; $(A \cap B \cap C) = \{S_1\}$

$P(A \cap B) = \frac{1}{4}$; $P(A \cap C) = \frac{1}{4}$; $P(B \cap C) = \frac{1}{4}$; $P(A \cap B \cap C) = \frac{1}{4}$

Check the conditions:

$P(A \cap B) = \frac{1}{4} = P(A)\, P(B) = \frac{1}{2} \times \frac{1}{2}$; $P(A \cap C) = \frac{1}{4} = P(A)\, P(C) = \frac{1}{2} \times \frac{1}{2}$

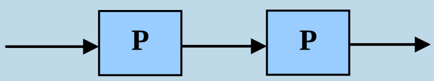
$P(B \cap C) = \frac{1}{4} = P(B)\, P(C) = \frac{1}{2} \times \frac{1}{2}$

$P(A \cap B \cap C) = \frac{1}{4} \neq P(A)\, P(B)\, P(C) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

➔ **Events are not independent** ***(even though the pair wise conditions of independence are satisfied)***

EXAMPLE (2-20): **"Reliability of a series system"**

Suppose that a system is made up of two components connected in series, each component has a probability (P) of working "**Reliability**". What is the probability that the system works assuming that components work independently?



SOLUTION:

P(system works) = P(component 1 works $\cap$ component 2 works)

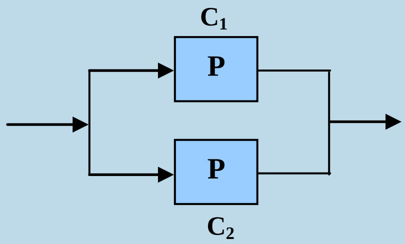
P(system works) = $P \times P = P^2$

*** The probability that the system works is also known as the "Reliability" of the system.***

EXAMPLE (2-21): **"Reliability of a parallel system"**

Suppose that a system is made up of two components connected in parallel. The system works if at least one component works properly. If each component has a probability (P) of working "**Reliability**" and components work independently, find the probability that the system works.

SOLUTION:



Reliability of the system = P(system works)

P(system works) $= P(C_1$ or C_2 or both C_1 and C_2 works)

$= P(C_1 \cup C_2)$

$= P(C_1) + P(C_2) - P(C_1 \cap C_2)$

$= P + P - (P \times P) = 2P - P^2$

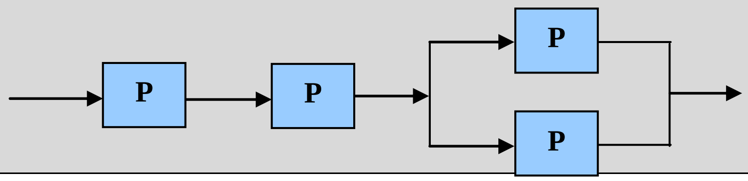
**** This system fails if both components fail.***

EXERCISE:

A pressure control apparatus contains 4 electronic tubes. The apparatus will not work unless all tubes are operative. If the probability of failure of each tube is 0.03, what is the probability of failure of the apparatus assuming that all components work independently?

EXERCISE: "Mixed system"

Find the reliability of the shown mixed system, assuming that all components work independently, and P is the reliability (probability of working) of each component.



EXAMPLE (2-22):

A coin may be fair or it may have two heads. We toss it (n) times and it comes up heads on each occasion. If our initial judgment was that both options for the coin (fair or both sides heads) were equally likely (probable), what is our revised judgment in the light of the data?

SOLUTION:

Let A : event representing coin is fair

B : event representing coin with two heads

C : outcome of the experiment $\underbrace{H\,H\,H\,H\,H \ldots H}_{\text{n times}}$

A priori probabilities:

$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$

➔ We need to find $P(A/C) = ?$

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A)\,P(C/A)}{P(C)}$$

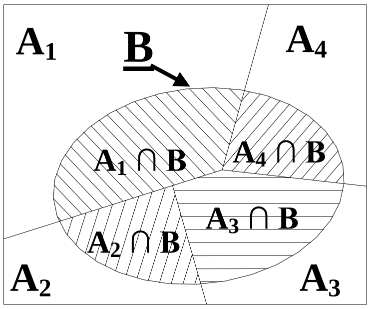
$$P(A/C) = \frac{P(A)\,P(H\,H\,H \ldots H / \text{fair coin})}{P(A)\,P(H\,H\,H \ldots H / \text{fair coin}) + P(B)\,P(H\,H\,H \ldots H / \text{coin with two heads})}$$

$$P(A/C) = \frac{\frac{1}{2}\left(\frac{1}{2}\right)^n}{\frac{1}{2}\left(\frac{1}{2}\right)^n + \frac{1}{2}(1)} = \frac{\left(\frac{1}{2}\right)^n}{\left(\frac{1}{2}\right)^n + 1} = \frac{1}{1+2^n}$$

$$P(B/C) = 1 - P(A/C) = 1 - \frac{1}{1+2^n} = \frac{2^n}{1+2^n}$$

Theorem of Total Probability:

Let $A_1, A_2, \ldots, A_n$ be a set of events defined over (S) such that:
$S = A_1 \cup A_2 \cup \ldots \cup A_n$; $A_i \cap A_j = \emptyset$ for $i \neq j$, and $P(A_i) > 0$ for $i = 1, 2, 3, \ldots n$.
For any event (B) defined on (S):
$P(B) = P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + \ldots\ldots + P(A_n)\, P(B/A_n)$



Proof:

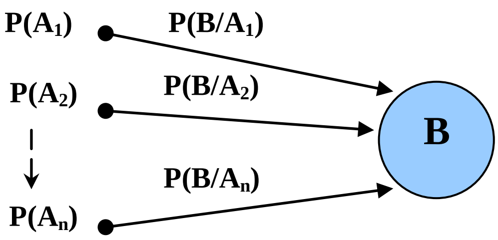
For events (A) and (B) in a sample space:
$B = \{A_1 \cap B\} \cup \{A_2 \cap B\} \cup \{A_3 \cap B\} \cup \{A_4 \cap B\}$

Since these events are disjoint, then:

$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + P(A_4 \cap B)$

But $P(A \cap B) = P(A)\, P(B/A) = P(B)\, P(A/B)$

$P(B) = P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + P(A_3)\, P(B/A_3) + P(A_4)\, P(B/A_4)$

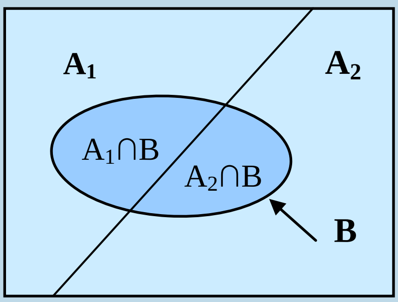


EXAMPLE (2-23):

If female students constitute 30% of the student body in the Faculty of Engineering and 40% of them have A GPA > 80, while 25 % of the male students have their GPA > 80. What is the probability that a person selected at random will have a GPA > 80?

SOLUTION:

A_1 = Event representing the selected person is a female
A_2 = Event representing the selected person is a male
B = Event representing GPA > 80
$P(A_1) = 0.3$
$P(A_2) = 0.7$
$B = (A_1 \cap B) \cup (A_2 \cap B) \Rightarrow P(B) = P(A_1 \cap B) + P(A_2 \cap B)$
$P(B) = P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2)$
$P(B) = (0.3 \times 0.4) + (0.7 \times 0.25)$
$P(B) = 0.295$



Baye's Theorem:

If $A_1, A_2, A_3, \ldots\ldots, A_n$ are disjoint events defined on (S), and (B) is another event defined on (S) (same conditions as above), then:

$$P(A_j/B) = \frac{P(A_j)\,P(B/A_j)}{\sum_{i=1}^{n} P(A_i)\,P(B/A_i)} = \frac{P(A_j \cap B)}{P(B)}$$

EXAMPLE (2-24):

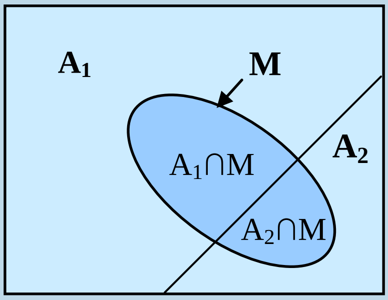
Suppose that when a machine is adjusted properly, 50% of the items produced by it are of high quality and the other 50% are of medium quality. Suppose, however, that the machine is improperly adjusted during 10% of the time and that under these conditions 25% of the items produced by it are of high quality and 75% are of medium quality.

a- Suppose that one item produced by the machine is selected at random, find the probability that it is of medium quality.

b- If one item is selected at random, and found to be of medium quality, what is the probability that the machine was adjusted properly.

SOLUTION:

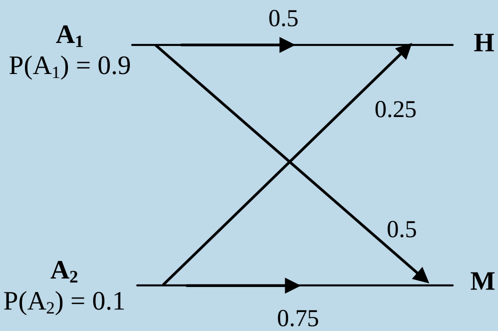
A_1 = Event representing machine is properly adjusted
A_2 = Event representing machine is improperly adjusted
H = Event representing item is of high quality
M = Event representing item is of medium quality
From the problem statement we have:
$P(A_1) = 0.9$; $P(A_2) = 0.1$
$P(H/A_1) = 0.5$; $P(H/A_2) = 0.25$
$P(M/A_1) = 0.5$; $P(M/A_2) = 0.75$



a- $P(M) = P(A_1 \cap M) + P(A_2 \cap M)$
$P(M) = P(A_1)\,P(M/A_1) + P(A_2)\,P(M/A_2)$
$\boldsymbol{P(M) = (0.9)(0.5) + (0.1)(0.75) = 0.525}$

b- $$P(A_1/M) = \frac{P(A_1 \cap M)}{P(M)} = \frac{P(A_1)\,P(M/A_1)}{P(M)}$$

$$\boldsymbol{P(A_1/M) = \frac{(0.9)(0.5)}{(0.525)} = 0.8571}$$

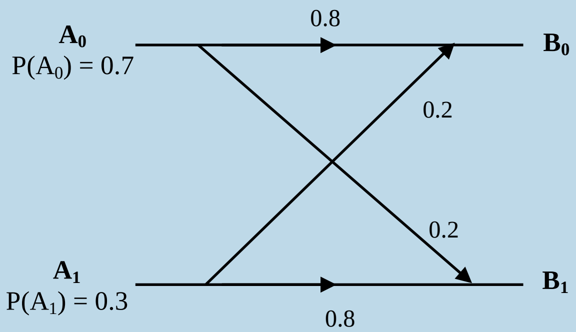


EXAMPLE (2-25):

Consider the problem of transmitting binary data over a noisy communication channel. Due to the presence of noise, a certain amount of transmission error is introduced. Suppose that the probability of transmitting a binary 0 is 0.7 (70% of transmitted digits are zeros) and there is a 0.8 probability that a given 0 or 1 being received properly.

a- What is the probability of receiving a binary 1.
b- If a 1 is received, what is the probability that a 0 was sent.

SOLUTION:



A_0 = Event representing 0 is sent
A_1 = Event representing 1 is sent
B_0 = Event representing 0 is received
B_1 = Event representing 1 is received

From the problem statement we have:

$P(A_0) = 0.7$; $P(A_1) = 0.3$
$P(B_0/A_0) = 0.8$; $P(B_0/A_1) = 0.2$
$P(B_1/A_0) = 0.2$; $P(B_1/A_1) = 0.8$

a- $P(B_1) = P(A_0)\, P(B_1/A_0) + P(A_1)\, P(B_1/A_1)$
$\boldsymbol{P(B_1) = (0.7)(0.2) + (0.3)(0.8) = 0.38}$
$\boldsymbol{P(B_0) = 1 - P(B_1) = 0.62}$

b- $$P(A_0/B_1) = \frac{P(A_0 \cap B_1)}{P(B_1)} = \frac{P(A_0)\, P(B_1/A_0)}{P(B_1)}$$

$$\boldsymbol{P(A_0/B_1) = \frac{(0.7)(0.2)}{(0.38)} = 0.3684}$$

EXAMPLE (2-26):

In a factory, four machines produce the same product. Machine A_1 produces 10% of the product, A_2 20%, A_3 30%, and A_4 40%. The proportion of defective items produced by the machines follows:

A_1: 0.001 ; A_2: 0.005 ; A_3: 0.005 ; A_4: 0.002

An item selected at random is found to be defective, what is the probability that the item was produced by machine A_1?

SOLUTION:

Let D be the event: Selected item is defective

$P(D) = P(A_1)\, P(D/A_1) + P(A_2)\, P(D/A_2) + P(A_3)\, P(D/A_3) + P(A_4)\, P(D/A_4)$
$P(D) = (0.1 \times 0.001) + (0.2 \times 0.005) + (0.3 \times 0.005) + (0.4 \times 0.002)$
$\boldsymbol{P(D) = 0.0034}$

$$P(A_1/D) = \frac{P(A_1)\, P(D/A_1)}{P(D)} = \frac{(0.1)\,(0.001)}{(0.0034)} = \frac{0.0001}{0.0034} = \frac{1}{34}$$

- **Counting techniques:**

Here we introduce systematic counting of sample points in a sample space. This is necessary for computing the probability P(A) in experiments with a finite sample space (S) consisting of (n) equally likely outcomes. Then each outcome has probability $\left(\frac{1}{n}\right)$.

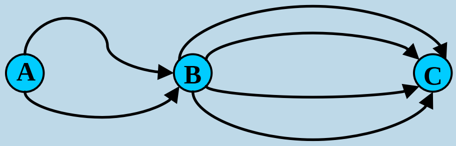
And if (A) consists of (m) outcomes, then $P(A) = \frac{m}{n}$

- ***Multiplication Rule:***

If operation A can be performed in n_1 different ways and operation B in n_2 different ways, then the sequence (operation A , operation B) can be performed in $n_1 \times n_2$ different ways.

EXAMPLE (2-27):

There are two roads between A and B and four roads between B and C. How many different routes can one travel between A and C.



SOLUTION:

n = 2 x 4 = 8

- **Permutation:**

Consider an urn having (n) distinguishable objects (numbered 1 to n). We perform the following two experiments:

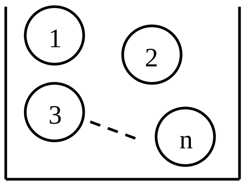
1- Sampling without replacement:

An object is drawn; its number is recorded and then put aside, another object is drawn; its number is recorded and then put aside, the process is repeated (k) times. The total number of ordered sequences $\{x_1, x_2, \ldots\ldots, x_k\}$ *(repetition is not allowed)* called permutation is:

$N = n (n - 1) (n - 2) \ldots\ldots (n - k + 1)$

$$N = \frac{n!}{(n-k)!} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $n! = n (n - 1) (n - 2) \ldots\ldots (3) (2) (1)$



2- Sampling with replacement:

If in the previous experiment, each drawn object is dropped back into the urn and the process is repeated (k) times. The number of possible sequences $\{x_1, x_2, \ldots\ldots, x_k\}$ of length (k) that can be formed from the set of (n) distinct objects *(repetition allowed):*

$$N = n^k \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

EXAMPLE (2-28):

How many different five-letter computer passwords can be formed:

a- If a letter can be used more than once.
b- If each word contains each letter no more than once.

SOLUTION:

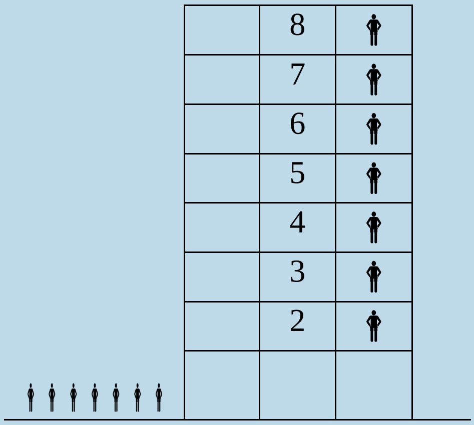
a- $N = (26)^5$

b- $N = \frac{26!}{(26-5)!}$

EXAMPLE (2-29):

An apartment building has eight floors (numbered 1 to 8). If seven people get on the elevator on the fist floor, what is the probability that:

a- All get off on different floors?
b- All get off on the same floor?



SOLUTION:

Number of points in the sample space:
First person can get off at any of the 7 floors.

Person (2) can get off at any of the 7 floors and so on.

➔ The number of ways people can get off:

(N) = $7 \times 7 \times 7 \times 7 \times 7 \times 7 \times 7 = 7^7$

a- Here the problem is to find the number of permutations of 7 objects taking 7 at a time.

$$P = \frac{7!}{7^7}$$

b- Here there are 7 ways whereby all seven persons get off on the same floor.

$$P = \frac{7}{7^7}$$

EXAMPLE (2-30):

If the number of people getting on the elevator on the first floor is 3:

a- Find the probability they get off the elevator on different floors.
b- Find the probability they get off the elevator on the same floor.

SOLUTION:

Number of points in the sample space (N) = $7 \times 7 \times 7 = 7^3$

a- $P = \frac{7 \times 6 \times 5}{7^3}$

b- $P = \frac{7}{7^3}$

EXAMPLE (2-31):

If the number of floors is 5 (numbered 1 to 5) and the number of people getting on the elevator is 8. Find the probability that exactly 2 people get off the elevator on each floor.

SOLUTION:

Number of points in the sample space (N) = $4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4^8$

$$P = \frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}\binom{2}{2}}{4^8}$$

EXAMPLE (2-32):

To determine an "odd man out", (n) players each toss a fair coin. If one player's coin turns up differently from all the others, that person is declared the odd man out. Let (A) be the event that some one is declared an odd man out.

a- Find P(A)
b- Find the probability that the game is terminated with an odd man out after (k) trials

SOLUTION:

a- $P(A) = \dfrac{\text{Number of outcomes in event (A)}}{\text{Number of possible sequences}}$

number of outcomes leading to an odd man out:
(n – 1) Heads and one Tail
(n – 1) Tails and one Head

$$P(A) = \frac{2\,n}{2^n} = \frac{n}{2^{n-1}}$$

with an odd man out, a success is obtained and the game is over.

$$\overbrace{\underset{\substack{H\\T}}{O}\ \underset{\substack{H\\T}}{O}\ \underset{\substack{H\\T}}{O}\ \underset{\substack{H\\T}}{O}\ \underset{\substack{H\\T}}{O}\cdots\underset{\substack{H\\T}}{O}}^{n}$$

$$\left.\begin{matrix} H\,H\,H\,H\,H\ \dots\ T \\ H\,H\,H\,H\ \dots\ T\,H \\ H\,H\,H\ \dots\ T\,H\,H \\ \vdots \\ T\,H\,H\,H\,H\ \dots\ H \end{matrix}\right\} \Rightarrow (n)$$

$$\left.\begin{matrix} T\,T\,T\,T\,T\ \dots\ H \\ T\,T\,T\,T\ \dots\ H\,T \\ T\,T\,T\ \dots\ H\,T\,T \\ \vdots \\ H\,T\,T\,T\,T\ \dots\ T \end{matrix}\right\} \Rightarrow (n)$$

b- A second trial is needed when the experiment ends with a failure:
➔ P(a second trial is needed) = 1 – P(A)
For (k) trials:

$$P(\underbrace{F\,F\,F\,F\dots F}_{k-1\ Trials}\ \underline{\underline{S}}) = P(F)^{k-1}\ P(S)$$

$$P(\underbrace{F\,F\,F\,F\dots F}_{k-1\ Trials}\ \underline{\underline{S}}) = [1-P(A)]^{k-1}\ P(A)$$

- ## Combination:

In permutation, the order of the selected objects is essential. In contrast, a combination of a given objects means any selection of one or more objects without regard to order.
The number of combinations of (n) different objects, taken (k) at a time, without repetition is the number of sets that can be made up from the (n) given objects, each set containing (k) different objects and no two sets containing exactly the same (k) objects.
The number is:

$$\boldsymbol{\binom{n}{k} = \frac{n!}{k!\,(n-k)!}}$$

Note that:

$$\underbrace{\begin{pmatrix}\text{Arrange (k) objects}\\ \text{selected from (n)}\end{pmatrix}}_{N} \text{ is the same as } \underbrace{\begin{pmatrix}\text{First select (k)}\\ \text{objects from (n)}\end{pmatrix}}_{\binom{n}{k}} \text{ and then } \underbrace{\begin{pmatrix}\text{Arrange the (k)}\\ \text{selected objects}\end{pmatrix}}_{k!}$$

$$N = \binom{n}{k} \times k! \quad \text{where } N = \frac{n!}{(n-k)!} \quad ➔ \quad \binom{n}{k} = \frac{N}{k!} = \frac{n!}{k!\,(n-k)!}$$

EXAMPLE (2-33):

From four persons (set of elements), how many committees (subsets) of two members (elements) may be chosen?

SOLUTION:

Let the persons be identified by the initials A, B, C and D
Subsets: (A , B) , (A , C) , (A , D) , (B , C) , (B , D) , (C , D)

$$N = \binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$$

Missing sequences: (A , A) , (B , B) , (C , C) , (D , D) ➔ *(repetition is not allowed)*
Missing sequences: (B , A) , (C , A) , (D , A)
(C , B) , (D , B) , (D , C) ➔ *(order is not important)*

EXAMPLE (2-34):

Consider the rolling of a die twice, how many pairs of numbers can be formed for each case?

SOLUTION:
n = 6 and k = 2
Case I: Permutation
a- With repetition
$N = n^k = 6^2 = 36$

b- Without repetition

$$N = \frac{n!}{(n-k)!} = \frac{6!}{(6-2)!} = 30$$

Case I: Combination

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{6!}{2!(6-2)!} = 15$$

D_1 \ D_2	1	2	3	4	5	6
1	***(1,1)***	**(1,2)**	**(1,3)**	**(1,4)**	**(1,5)**	**(1,6)**
2	**(2,1)**	***(2,2)***	**(2,3)**	**(2,4)**	**(2,5)**	**(2,6)**
3	**(3,1)**	**(3,2)**	***(3,3)***	**(3,4)**	**(3,5)**	**(3,6)**
4	**(4,1)**	**(4,2)**	**(4,3)**	***(4,4)***	**(4,5)**	**(4,6)**
5	**(5,1)**	**(5,2)**	**(5,3)**	**(5,4)**	***(5,5)***	**(5,6)**
6	**(6,1)**	**(6,2)**	**(6,3)**	**(6,4)**	**(6,5)**	***(6,6)***

EXAMPLE (2-35):

In how many ways can we arrange 5 balls numbered 1 to 5 in 10 baskets each of which can accommodate one ball?

SOLUTION:

$$\text{The number of ways } (N) = \frac{n!}{(n-k)!} = \frac{10!}{(10-5)!} = \frac{10!}{5!}$$

NOTE:
If we remove the numbers of the balls so that the balls are no longer distinguishable, then:

$$\text{The number of ways } \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{10!}{5!(10-5)!} = \frac{10!}{5!\ 5!}$$

This is because the permutation within the 5 balls is no longer needed.

Arrangement of Elements of Two Distinct Types
When a set contains only elements of two distinct types, type (1) consists of k elements and type (2) consists of (n-k) elements, then the number of different arrangements of all the

elements in the set is given by the binomial coefficient. Suppose, for example, that we have k ones and (n-k) zeros to be arranged in a row, then the number of binary numbers that can be formed is $\binom{n}{k}$. If n = 4 and k = 1, then the possible binary numbers are (0001, 0010, 0100, 1000).

Exercise: How many different binary numbers of five digits can be formed from the numbers 1, 0? List these numbers.

Exercise: How many different binary numbers of five digits can be formed from the numbers 1, 0 such that each number contains two ones? List these numbers.

Exercise: In how many ways can a group of five persons be seated in a row of 10 chairs?

The Multinomial Coefficient:
The number of ways to arrange n items of which n_1 are of one type, n_2 of a second type, …, n_k of a k'th type is given by $N = \binom{n}{n_1 \quad n_2 \quad . \quad . \quad n_k} = \dfrac{n!}{n_1!n_2!...n_k!}$

- ***Comments: Stirling's formula***
  Computing n! can be approximated by: $n! \sim \sqrt{2\pi}\, n^{n+1/2}\, e^{-n}$

CHAPTER II

SINGLE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

- **Definition:**

  A real-valued function whose domain is the sample space is called a *random variable* (r.v).

- The random variable is given an uppercase letter X, Y, Z, … while the values assumed by this random variable are given lowercase letters x, y, z, …
- The whole idea behind the r.v is a one to one *mapping* from the sample space on the real line via the mapping function X(s).
- Associated with each discrete r.v (X) is a *Probability Mass Function P(X = x)*. This density function is the sum of all probabilities associated with the outcomes in the sample space that get mapped into (x) by the mapping function (random variable X).
- Associated with each continuous r.v (X) is a *Probability Density Function (pdf)* $f_X(x)$.This $f_X(x)$ is not the probability that the random variable (X) takes on the value (x), rather $f_X(x)$ is a continuous curve having the property that:

$$P(a \le X \le b) = \int_a^b f_X(x)\, dx$$

- **Definition:**

  The cumulative distribution function of a r.v (X) defined on a sample space (S) is given by:
  $F_X(x) = P\{X \le x\}$

- ***Properties of $F_X(x)$***

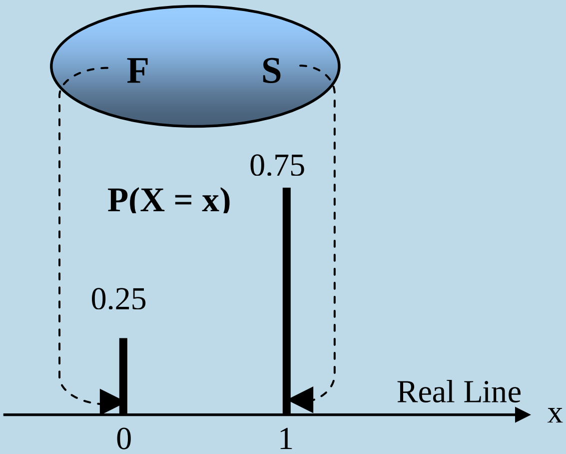
  1- $F_X(-\infty) = 0$
  2- $F_X(\infty) = 1$
  3- $0 \le F_X(x) \le 1$
  4- $F_X(x_1) \le F_X(x_2)$ if $x_1 \le x_2$
  5- $F_X(x^+) = F_X(x)$ function is continuous from the right
  6- $P\{x_1 \le X \le x_2\} = F_X(x_2) - F_X(x_1)$

EXAMPLE (3-1):

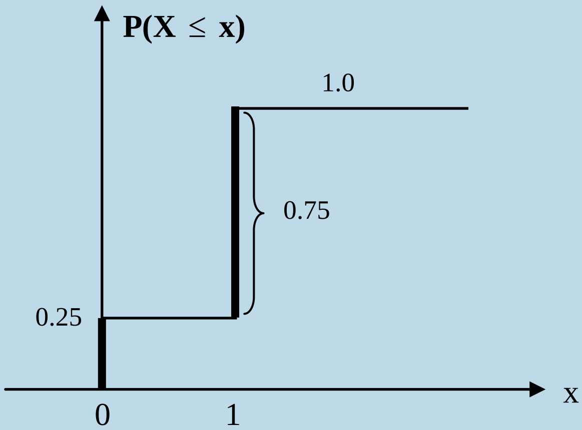
A chance experiment has two possible outcomes, a success with probability 0.75 and a failure with probability 0.25. Mapping function (random variable X) is defined as:
x = 1 if outcome is a success
x = 0 if outcome is a failure

SOLUTION:

$P(X < 0) = 0$; $P(X \le 0) = 0.25$; $P(X < 1) = 0.25$; $P(X \le 1) = 1$



Probability Mass Function



Cumulative Distribution Function

EXAMPLE (3-2):

Let the above experiment be conducted three times in a row.

a- Find the sample space.

b- Define a random variable (X) as X = number of successes in the three trials.

c- Find the probability mass function P(X = x).

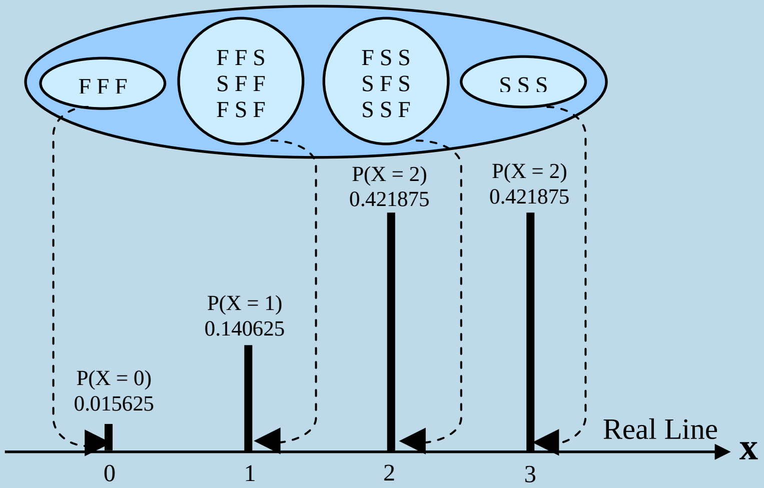
d- Find the cumulative distribution function $F_X(x) = P\{X \leq x\}$

SOLUTION:

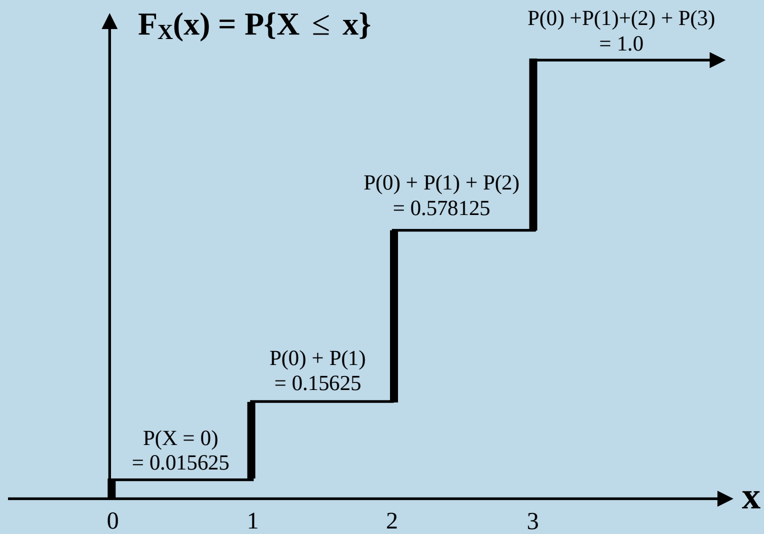
In the table below we show the possible outcomes and the mapping process:

Sample Outcome	$P(s_i)$	x	P(X = x)
F F F	$(0.25)^3$	0	$(0.25)^3 = 0.015625$
F F S	$(0.75)\,(0.25)^2$	1	$3 \times (0.75)\,(0.25)^2 = 0.140625$
S F F	$(0.75)\,(0.25)^2$		
F S F	$(0.75)\,(0.25)^2$		
S S F	$(0.75)^2\,(0.25)$	2	$3 \times (0.75)^2\,(0.25) = 0.421875$
S F S	$(0.75)^2\,(0.25)$		
F S S	$(0.75)^2\,(0.25)$		
S S S	$(0.75)^3$	3	$(0.75)^3 = 0.421875$

Probability Mass Function



Cumulative Distribution Function



<u>Binomial Distribution</u>: (n) = number of trials ; (x) number of successes in (n) trials

EXAMPLE (3-3):

Suppose that 5 people including you and your friend line up at random. Let (X) denote the number of people standing between you and your friend. Find the probability mass function for the random variable (X).

SOLUTION:

Number of different ways by which the 5 people can arrange themselves = 5!
This is the total number of points in the sample space.

Let (A) denote you,
(B) denote your friend.
The random variable (X) assumes four possible values 0, 1, 2, 3 as shown below:

$$\left.\begin{matrix}\text{A B O O O}\\ \text{O A B O O}\\ \text{O O A B O}\\ \text{O O O A B}\end{matrix}\right\} \Rightarrow (X=0)$$

$$\left.\begin{matrix}\text{A O B O O}\\ \text{O A O B O}\\ \text{O O A O B}\end{matrix}\right\} \Rightarrow (X=1)$$

$$\left.\begin{matrix}\text{A O O B O}\\ \text{O A O O B}\end{matrix}\right\} \Rightarrow (X=2)$$

$$\left.\text{A O O O B}\right\} \Rightarrow (X=3)$$

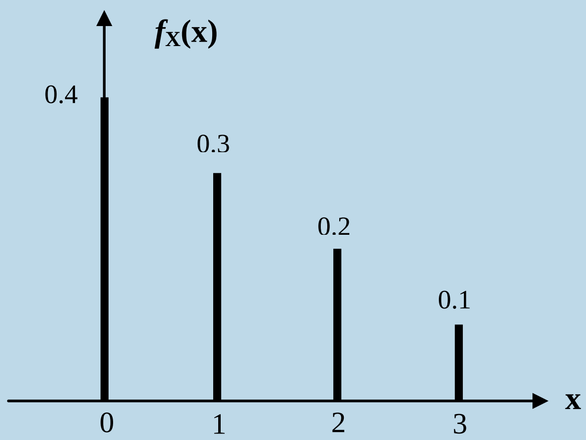
Any Sequence similar to what is shown can be done in: $\underbrace{2!}_{\text{you and your friend}} \times \underbrace{3!}_{\text{the other people}}$

$$P(X=0)=\frac{4\times 2!\times 3!}{5!}=0.4$$

$$P(X=1)=\frac{3\times 2!\times 3!}{5!}=0.3$$

$$P(X=2)=\frac{2\times 2!\times 3!}{5!}=0.2$$

$$P(X=3)=\frac{1\times 2!\times 3!}{5!}=0.1$$



Probability Mass Function

Continuous Random Variables and Distribution:

- ***Definition:***

A random variable and its distribution are called of continuous type if the corresponding cumulative distribution function $F_X(x)$ can be given by an integral of the form:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$$

where $f_X(x)$ is the probability density function related to $F_X(x)$ by:

$$f_X(x) = \frac{d}{dx} F_X(x)$$

- ***Properties of $f_X(x)$***

1- $f_X(x) \geq 0$; nonnegative

2- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

3- $P\{x_1 \leq X \leq x_2\} = \int_{x1}^{x2} f_X(u)\,du$; Probability is the area under the $f_X(x)$ curve between x_1 and x_2.

EXAMPLE (3-4):

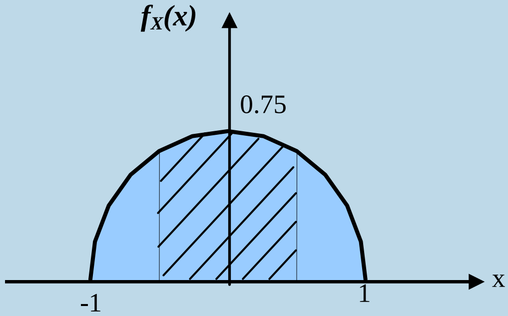
Let (X) have the pdf: $f_X(x) = 0.75\,(1 - x^2)$; $\{-1 \leq x \leq 1\}$

1- Verify that $f_X(x)$ is indeed a valid pdf.

2- Find:

a- $F_X(x)$

b- $P\{-\frac{1}{2} \leq X \leq \frac{1}{2}\}$



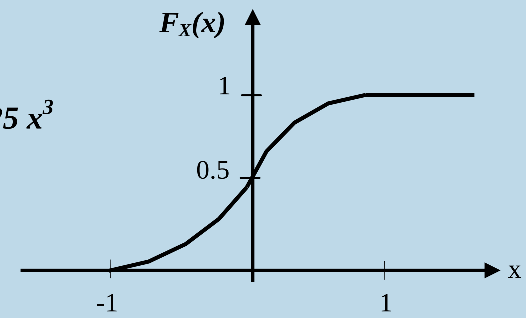
Probability Density Function

SOLUTION:

1- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ ➔ $2\int_{0}^{1} 0.75\,(1 - x^2)\,dx$

$$= 2 \times (0.75\,u - 0.75\frac{u^3}{3}\Big|_0^1 = 2(0.75 - 0.25) = 1.0$$

2-a) $F_X(x) = \int_{-\infty}^{x} 0.75\,(1 - u^2)\,du$ $= \boldsymbol{0.5 + 0.75x - 0.25\,x^3}$



Cumulative Distribution Function

2-b) $P\{-\frac{1}{2} \leq X \leq \frac{1}{2}\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} 0.75\,(1 - u^2)\,du$

$$= F_X(\tfrac{1}{2}) - F_X(-\tfrac{1}{2}) = \boldsymbol{0.6875}$$

EXERCISE:

Find x_0 such that $F_X(x) = P\{X \leq x_0\} = 0.95$

SOLUTION:

$P\{X \leq x_0\} = 0.5 + 0.75x_0 - 0.25\,x_0^3 = 0.95$ ➔ $3x_0 - x_0^3 = 1.8$ ➔ $x_0 \cong 0.73$

Mean and Variance of a Distribution:

- ***Definition:***

The mean value or expected value of a random variable (X) is defined as:

$$\mu_X = E\{X\} = \sum x_i \; P(X = x_i) \qquad \textit{if x is discrete}$$

$$\mu_X = E\{X\} = \int_{-\infty}^{\infty} x \, f_X(x) \, dx \qquad \textit{if x is continuous}$$

- ***Definition:***

The variance of a random variable (X) is defined as:

$$\sigma_X^2 = E\{(X - \mu_x)^2\} = \sum (X - \mu_x)^2 \; P(X = x_i) \quad \textit{if x is discrete}$$

$$\sigma_X^2 = E\{(X - \mu_x)^2\} = \int_{-\infty}^{\infty} (x - \mu_x)^2 \; f_X(x) \, dx \qquad \textit{if x is continuous}$$

$$\sigma_X = \sqrt{\sigma_X^2} \qquad \textit{is the standard deviation}$$

The variance is the measure of the spread of the distribution.

- ***Definition:***

For any random variable (X) and any continuous function Y = g (X), the expected value of g(X) is defined as:

$$E\{g(X)\} = \sum g(x_i) \, P(X = x_i) \qquad \textit{if x is discrete}$$

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) \, f_X(x) \, dx \qquad \textit{if x is continuous}$$

- ***Theorem:***

Let (X) be a random variable with mean μ_X, then:

$$\sigma_X^2 = E(X^2) - \mu_X^2$$

Proof:

$$\sigma_X^2 = E\{(X - \mu_X)^2\} = \int_{-\infty}^{\infty} (x - \mu_X)^2 \; f_X(x) \, dx$$

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x^2 - 2x\mu_X + \mu_X^2) \, f_X(x) \, dx$$

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 \; f_X(x) \, dx - 2\mu_X \int_{-\infty}^{\infty} x \, f_X(x) \, dx + \mu_X^2 \int_{-\infty}^{\infty} f_X(x) \, dx$$

$$\sigma_X^2 = E(X^2) - 2\mu_X \mu_X + \mu_X^2$$

$$\boldsymbol{\sigma_X^2 = E(X^2) - \mu_X^2}$$

Illustration:

1. The center of mass for a system of particles of masses m_1, m_2 ..., m_n placed at x_1, x_2 …, x_n is:

$$x_{cm} = \frac{1}{\sum m_i} (x_1 m_1 + x_2 m_2 + ... \; x_n m_n)$$

If we let $m_1 = p_1$, $m_2 = p_2$, …, then:

$$x_{cm} = x_1 p_1 + x_2 p_2 + ... \; x_n p_n \quad \text{(The mean of a discrete distribution)}$$

2. If $\rho(x)$ is the density of a rigid body along the x-axis, then the center of mass is:

$$x_{cm} = \frac{1}{M}\int x\,\rho(x)\,dx$$

Where $M = \int \rho(x)\,dx$

Again, if $\rho(x)$ is replaced by $f_X(x)$, the pdf function, then:

$x_{cm} = \int x\, f_X(x)\,dx$ is the mean of continuous distribution.

Moment of Inertia:

3. If the particles in (1) above rotate with angular velocity (w), then the moment of inertia is evaluated as:

$$I = \sum_{i=1}^{n} m_i\, x_i^2$$

With m_i replaced by p_i, we have:

$$I = \sum_{i=1}^{n} p_i\, x_i^2$$

4. If the rigid body in (2) rotates with angular velocity (w), then:

$$I = \int x^2\,\rho(x)\,dx \;\Rightarrow\; E(x^2) = \int x^2\, f_X(x)\,dx$$

5. The variance $E\{(X-\mu_X)^2\}$ parallels the moment of inertia about the center of mass.

"Recall the parallel axis theorem"

$$I = I_{cm} + M\,h^2$$

$$E(x^2) = \sigma^2 + E^2(x)$$

EXAMPLE (3-5):

In the kinetic theory of gases, the distance (x) that a molecule travels between collisions is described by the exponential density function

$$f_X(x) = \frac{1}{\lambda}\, e^{\frac{-x}{\lambda}} \quad x > 0$$

a- The mean free path defined as the average distance between collisions is calculated as:

$$\text{Mean Free Path} = \mu_X = E\{X\} = \int_0^{\infty} x\, f_X(x)\,dx$$

$$= \int_0^{\infty} x\left(\frac{1}{\lambda}\right) e^{\frac{-x}{\lambda}}\,dx = \lambda = \frac{1}{\sqrt{2}\,\pi\, d^2\, N/V}$$

Where (N/V) is the number of molecules per unit volume and (d) is the molecular diameter.

b- If the average speed of a molecule is $\bar{v}$ m/s, then the average collision rate is

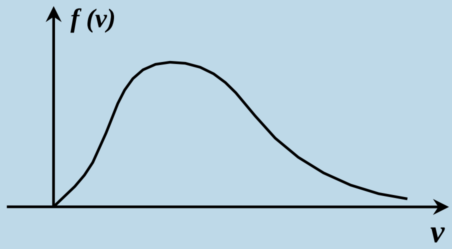
$$\text{Rate} = \frac{\bar{v}}{\lambda}$$

EXAMPLE (3-6):

Maxwell's Distribution Law:

The speed of gas molecules follows the distribution:

$$f(v) = 4\pi\left(\frac{M}{2\pi R T}\right)^{\frac{3}{2}} v^2 \; e^{\frac{-M v^2}{2RT}} \qquad v \geq 0$$



Where v is the molecular speed
T is the gas temperature in Kelvin
R is the gas constant (8.31 J/mol.K)
M is the molecular mass of the gas

a- Find the average speed, $\bar{v}$
b- Find the root mean square speed v_{rms}
c- Find the most probable speed

SOLUTION:

a- $\bar{v} = E(v) = \int_0^\infty v \, f(v)\, dv = \sqrt{\frac{8RT}{\pi M}}$

b- $E\{v^2\} = (v_{rms})^2 = \int_0^\infty v^2 f(v)\; dv = \frac{3RT}{M} \quad \Rightarrow \quad v_{rms} = \sqrt{\frac{3RT}{M}} \quad ; \quad rms = \sqrt{E(v^2)}$

c- The most probable speed is the speed at which $f(v)$ attains its maximum value. Therefore, we differentiate $f(v)$ with respect to (v), set the derivative to zero and solve for the maximum. The result is:

Most probable speed = $\sqrt{\frac{2RT}{M}}$

Root ➔ $\sqrt{\quad}$
Mean ➔ E(.)
Square ➔ v^2
➔ $rms = \sqrt{E(v^2)}$

Exercise

The radial probability density function for the ground state of the hydrogen atom (the pdf of the electron position from the atom) is given by

$$f(r) = \frac{4}{a^3} r^2 e^{-2r/a} \quad \text{for } r > 0$$

where a is the Bohr radius (a = 52.9 pm).

a. What is the distance from the center of the atom that the electron is most likely to be found?
b. Find the average value of r?, (the mean distance of the electron from the center of the atom).
c. What is the probability that the electron will be found within a sphere of radius a centered at the origin?

- ***Theorem:***

Let (X) be a random variable with mean μ_X and variance σ_X^2.

Define Y = aX + b ; (a) and (b) are real constants, then:

$\mu_Y = a\,\mu_x + b$ ………… (a)

$\sigma_Y^2 = a^2\,\sigma_X^2$ ………… (b)

Proof:

a- $\mu_Y = E\{aX + b\}$

$$= \int_{-\infty}^{\infty} (ax + b)\, f_X(x)\, dx$$

$$= a\int_{-\infty}^{\infty} x\, f_X(x)\, dx + b\int_{-\infty}^{\infty} f_X(x)\, dx \;\Rightarrow\; \boldsymbol{\mu_Y = a\,\mu_x + b}$$

b- $\sigma_Y^2 = E\{(Y - \mu_Y)^2\}$

$$= E\{[(ax + b) - (a\mu_X + b)]^2\} \quad = \quad E\{[a(x - \mu_X)]^2\}$$

$$= a^2 E\{(x - \mu_X)^2\} \;\Rightarrow\; \boldsymbol{\sigma_Y^2 = a^2\,\sigma_X^2}$$

EXAMPLE (3-7):

Find the mean and the variance of the binomial distribution considered earlier (Example 3-2) with n = 3 and P(S) = 0.75

SOLUTION:

Mean = $\mu_X = E\{X\} = \sum x_i\, P(X = x_i)$

x	**P(X = x)**	**x . P(X = x)**
0	0.015625	0
1	0.140625	0.140625
2	0.421875	0.843750
3	0.421875	1.265625
$\sum$		***2.25***

$\sum x_i\, P(X = x_i)$ = ***2.25 = 3 x 0.75***

➔ $E(X) = n\,p$ ***= number of trials x probability of a success***

Variance = $\sigma_X^2 = E(X^2) - [E(X)]^2$; $E\{X^2\} = \sum x_i^2\, P(X = x_i)$

x	**x^2**	**P(X = x)**	**x^2 . P(X = x)**
0	0	0.015625	0
1	1	0.140625	0.140625
2	4	0.421875	1.687500
3	9	0.421875	3.796875
$\sum$			***5.625***

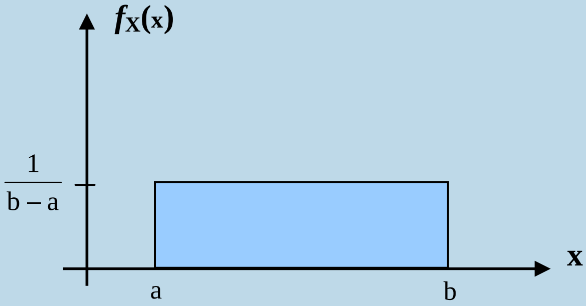
$\boldsymbol{\sigma_X^2 = 5.625 - (2.25)^2 = 0.5625 = 3 \times 0.75 \times 0.25}$

= number of trials x probability of success x probability of Failure ➔ $\boldsymbol{\sigma_X^2 = n\,p\,(1 - p)}$

EXAMPLE (3-8):

Find the mean and the variance of the uniform distribution shown in the figure.

SOLUTION:



$$\text{Mean} = \mu_X = E\{X\} = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

$$\mu_X = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

$$\text{Var}(X) = \sigma_X^2 = E(X^2) - [E(X)]^2$$

$$E\{X^2\} = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

$$\sigma_X^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

EXAMPLE (3-9):

Let $Z = \frac{X - \mu_X}{\sigma_X}$ *(Standardized r.v.)*, show that the mean of (Z) is zero and the variance is 1.

SOLUTION:

Z can be written as: $Z = \frac{X}{\sigma_X} - \frac{\mu_X}{\sigma_X} = aX + b$

$$\text{Mean} = \mu_Z = E\{Z\} = \frac{1}{\sigma_X} E\{(X - \mu_X)\} = \frac{1}{\sigma_X}\{E(X) - E(\mu_X)\} = 0$$

$$\text{Var}(Z) = \sigma_Z^2 = \frac{1}{\sigma_X^2}\sigma_X^2 = 1$$

EXAMPLE (3-10):

Let X be a discrete random variable with the following pmf: $P(X=0) = 0.4$, $P(X=1) = 0.3$ $P(X=2) = 0.2$, $P(X=3) = 0.1$. Find the mean and variance of X.

SOLUTION:

$$\mu_X = E\{X\} = \sum x_i\, P(X = x_i) = 0(0.4)+1(0.3)+2(0.2)+3(0.1) = 1$$

$$E\{X^2\} = \sum x_i^2\, P(X = x_i) = 0(0.4) + 1(0.3) + (2)^2(0.2) + (3)^2(0.1) = 2$$

$$\text{Variance} = \sigma_X^2 = E(X^2) - [E(X)]^2 = 2 - 1 = 1$$

- ***Some useful properties of expectation:***
- $E\{a\} = a$; a is a constant
- $E\{a\, g(X)\} = a\, E\{g(X)\}$; a is a constant
- $E\{g_1(X) + g_2(X)\} = E\{g_1(X)\} + E\{g_2(X)\}$

- ***The median and the mode:***
- ***Definition:***

For a continuous random variable (X), the median of the distribution of (X) is defined to be a point (x_0) such that:
$P(X < x_0) = P(X \geq x_0)$

- ***Definition:***

If a random variable (X) has a pdf $f_X(x)$, then the value of (x) for which $f_X(x)$ is maximum is called the mode of the distribution.

EXAMPLE (3-11):

Find the median and the mode for the random variable X with pdf: $f_X(x) = 2xe^{-x^2}$, x >0

SOLUTION:

The *median* is a point (x_0) such that

$$\int_0^{x_0} 2xe^{-x^2}dx = \int_{x_0}^{\infty} 2xe^{-x^2}dx = ½.$$

(x_0) is the solution to $e^{-x_0^2} = 0.5$ which results in (x_0) = 0.832554

To find the *mode* we differentiate $f_X(x)$ with respect to x and set the derivative to zero

$\frac{df(x)}{dx} = 2e^{-x^2} - 4x^2e^{-x^2} = 0$, the solution of which is $x = 1/\sqrt{2}$.

- **Common Discrete Random Variables:**

I. The Binomial Distribution

- ***Definition:***

A random experiment consisting of (n) repeated trials such that:

a- The trials are independent.
b- Each trial results in only two possible outcomes, a success and a failure.
c- The probability of a success (p) on each trial remains constant

Is called a binomial experiment.

The r.v (X) that equals the number of trials that results in a success has a binomial distribution with parameters (n) and (p).

The probability mass function of (X) is:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad x = 0, 1, 2, ... , n$$

Theorem:

If (X) is a binomial r.v with parameters (n) and (p), then:

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

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

Proof:

First we show that $\mu_X = n\,p$

$$\mu_X = E(X) = \sum_{x=0}^{n} x \binom{n}{x} (p)^x (1-p)^{n-x} = \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} (p)^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} (p)^x (1-p)^{n-x}$$

Let u= x -1. In terms of u, the summation above can be expressed as

$$\mu_X = \sum_{u=0}^{n-1} \frac{n(n-1)!}{u!(n-1-u)!} (p)^{u+1} (1-p)^{n-1-u}$$

Now substitute m = n-1, and take n and p out of the summation we get

$$\mu_X = np \sum_{u=0}^{m} \frac{m!}{u!(m-u)!} (p)^u (1-p)^{m-u}$$

The summation on the right hand side equals 1 since this is the summation of probabilities of a binomial distribution with parameters m and p. The mean value of X is then

$$\mu_X = np$$

This result simply states that the mean value of a binomial random variable with parameters (n, p) equals the number of the times the experiment is repeated times the probability of a success on each trial.

To find the variance of X we find it convenient to first find E(X(X-1)) as follows:

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1) \binom{n}{x} (p)^x (1-p)^{n-x} = \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} (p)^x (1-p)^{n-x}$$

$$= \sum_{x=2}^{n} x(x-1) \frac{n!}{x(x-1)(x-2)!(n-x)!} (p)^x (1-p)^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} (p)^x (1-p)^{n-x}$$

As we did before, let u= x-2 or x = u+2. The summation above becomes

$$= \sum_{u=0}^{n-2} \frac{n(n-1)(n-2)!}{u!(n-2-u)!} (p)^{u+2} (1-p)^{n-2-u}$$

Next let m = n-2 and take out of the summation n, (n-1) and p^2, we get

$$= n(n-1)p^2 \sum_{u=0}^{m} \frac{m!}{u!(m-u)!} (p)^u (1-p)^{m-u}$$

Again, the summation on the right hand side equals 1 since it represents the sum of probabilities for a binomial distribution with parameters m and p. Therefore,

$$E(X(X-1)) = n(n-1)p^2$$

But

$$E(X(X-1)) = E(X^2 - X) = E(X^2) - E(X)$$

Or, $E(X^2) = E(X) + E(X(X-1))$

From which we conclude that:

$$\sigma_x^{\ 2} = E(X^2) - (\mu_X)^2 = np + n(n-1)p^2 - (np)^2$$

This simplifies to

$$\sigma_x^{\ 2} = np(1-p)$$

which concludes the proof.

EXAMPLE (3-12):

Suppose that the probability that any particle emitted by a radioactive material will penetrate a certain shield is 0.02. If 10 particles are emitted. Find the probability that:

a- Exactly one particle will penetrate the shield.

b- At least two particles will penetrate the shield.

SOLUTION:

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x} \quad ; \quad p=0.02 \quad ; \quad n=10$$

a- $P(X=1)=\binom{10}{1}(0.02)^1(1-0.02)^{10-1}$

b- $P(X\geq 2)=\sum_{x=2}^{10}\binom{10}{x}(0.02)^x(1-0.02)^{10-x}$

Also: $P(X = 0) + P(X = 1) + P(X\geq 2) = 1$

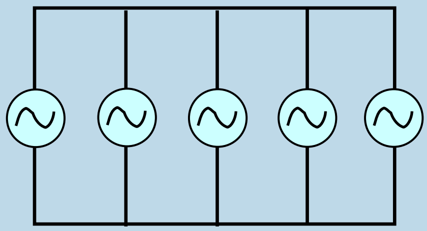
➔ $P(X\geq 2) = 1 - [P(X = 0) + P(X = 1)]$

$$P(X\geq 2)=1-\left[\binom{10}{0}(0.02)^0(1-0.02)^{10-0}+\binom{10}{1}(0.02)^1(1-0.02)^{10-1}\right]$$

EXAMPLE (3-13):

Consider the parallel system shown in the figure. The system fails if at least three of the five machines making up the system fail. Find the reliability of the system assuming that the probability of failure of each unit is 0.1 over a given period of time.

SOLUTION:



Let (X) be the number of machines in failure.

(X) has a binomial distribution.

P(system fails) = P(number of machines in failure ≥ 3)

$= P(x\geq 3)$

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

P(system fails) = 0.00856 ; when p = 0.1

➔ Reliability = 1 – P(Failure) = 0.99144

EXAMPLE (3-14):

The process of manufacturing screws is checked every hour by inspecting 10 screws selected at random from the hour's production. If one or more screws are found defective, the production process is halted and carefully examined. Otherwise the process continues. From past experience it is known that 1% of the screws produced are defective. Find the probability that the process is not halted.

SOLUTION:

Let (X) be the number of defective items in the sample.

P(system is not halted) = P(X = 0) = P(number of defective items is zero)

$$=\binom{10}{0}(p)^0(1-p)^{10-0}$$

$$=\binom{10}{0}(0.01)^{0}(0.99)^{10-0}=(0.99)^{10}=0.9043$$

EXAMPLE (3-15):

Thirty students in a class compare birthdays. What is the probability that:

a- 5 of the students have their birthday in January?

b- 5 of the students have their birthday on January 1st?

c- At least one student is born in January?

SOLUTION:

a- $P(\text{success})=\frac{1}{12}$; $P(\text{failure})=\frac{11}{12}$, Number of trials (n) = 30

Required number of successes (k) = 5

$$P(5\,\text{successes in 30 trials})=\binom{n}{k}(p)^{k}(1-p)^{n-k}$$

$$P(5\,\text{successes in 30 trials})=\binom{30}{5}\left(\frac{1}{12}\right)^{5}\left(\frac{11}{12}\right)^{30-5}=\binom{30}{5}\left(\frac{1}{12}\right)^{5}\left(\frac{11}{12}\right)^{25}$$

b- $P(\text{success})=\frac{1}{365}$; $P(\text{failure})=\frac{364}{365}$, Number of trials (n) = 30

Required number of successes (k) = 5

$$P(5\,\text{successes in 30 trials})=\binom{n}{k}(p)^{k}(1-p)^{n-k}=\binom{30}{5}\left(\frac{1}{365}\right)^{5}\left(\frac{364}{365}\right)^{25}$$

c- $P(\text{success})=\frac{1}{12}$; $P(\text{failure})=\frac{11}{12}$

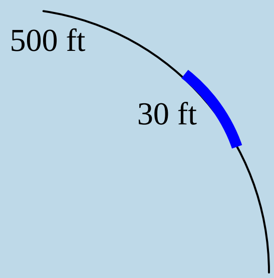
$$P(X\geq 1)=1-P(X=0)=\left[1-\binom{30}{0}\left(\frac{1}{12}\right)^{0}\left(\frac{11}{12}\right)^{30-0}\right]=1-0.0735=0.9265$$

EXAMPLE (3-16):

The captain of a navy gunboat orders a volley of 25 missiles to be fired at random along a 500-foot stretch of shoreline that he hopes to establish as a beach head. Dug into the beach is a 30-foot long bunker serving as the enemy's first line of defense.

a. What is the probability that exactly three shells will hit the bunker?

b. Find the number of shells expected to hit the bunker.



SOLUTION:

$$P(\text{success})=\frac{30}{500}=0.06$$

$$P(3\,\text{successes in 25 shells})=\binom{n}{k}(p)^{k}(1-p)^{n-k}$$

For p = 0.06 and n = 25

$$P(3\,\text{successes in 25 shells})=\binom{25}{3}(0.06)^{3}(1-0.06)^{25-3}=\binom{25}{3}(0.06)^{3}(0.94)^{22}$$

b. E(x) = n p = 25(0.06)= 1.5.

II. The Geometric Distribution

Let the outcome of an experiment be either a success with probability (p) or a failure with probability (1 – p). Let (X) be the number of times the experiment is performed to the first occurrence of a success. Then (X) is a discrete random variable with integer values ranging from one to infinity. The probability mass function of (X) is:

$$P(X = x) = P(\underbrace{F\,F\,F\,F \ldots F}_{x-1}\,S) = P(F)^{x-1}P(S)$$

$$= (1 - p)^{x-1}\,(p) \quad ; \quad x = 1, 2, 3, \ldots\ldots$$

- ***Theorem:***

The mean and the variance of (X) are:

$$\mu_X = E(X) = \frac{1}{p}$$

$$\sigma_X^2 = Var(X) = \frac{1-p}{p^2}$$

EXAMPLE (3-17):

Let the probability of occurrence of a flood of magnitude greater than a critical magnitude in a given year be 0.02. Assuming that floods occur independently, determine the "return period" defined as the average number of years between floods.

SOLUTION:

(X) has a geometric distribution with p = 0.01

$$\mu_X = E(X) = \frac{1}{p} = \frac{1}{0.02} = 50 \text{ years}$$

EXAMPLE (3-18):

Show that the mean value of the geometric distribution = 1/p and the variance is $\sigma_X^2 = \frac{1-p}{p^2}$, where p is the probability of a success.

SOLUTION:

$$\mu_X = E(X) = \sum_{x=1}^{\infty} x\, p(1-p)^{x-1} = p\{1+2(1-p)+3(1-p)^2 + 4(1-p)^3 + \ldots\}$$

Recall the geometric series

$$1 + u + u^2 + u^3 + \ldots = \frac{1}{1-u}$$

Differentiating both sides with respect to u, we get

$$1 + 2u + 3u^2 + \ldots = \frac{1}{(1-u)^2}$$

Making use of this result (with u =1-p), the expected value of X becomes

$$\mu_X = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

To find the variance, we first find $E(X(X-1))$ as

$$E(X(X-1)) = \sum_{x=1}^{\infty} x(x-1)\, p(1-p)^{x-1} = p\{2(1)(1-p) + 3(2)(1-p)^2 + 4(3)(1-p)^3 + \ldots\}$$

$$E(X(X-1)) = p(1-p)\{2(1)+3(2)(1-p)^1 + 4(3)(1-p)^2 + \ldots\}$$

Differentiating the geometric series twice with respect to u

$$2+3(2)u+4(3)u^2\ldots=\frac{2}{(1-u)^3}$$

Making use of this result (with u=1-p), we get

$$E(X(X\text{-}1)) \;=\; p(1\text{-}p)\frac{2}{p^3}=\frac{2(1-P)}{p^2}$$

But, $E(X(X-1))=E(X^2-X)=E(X^2)-E(X)$

Or, $E(X^2)=E(X)+E(X(X-1))$

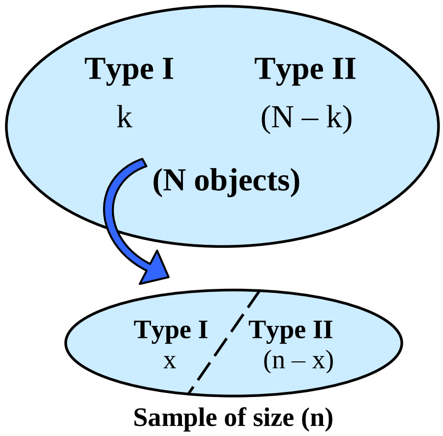
From which we conclude that:

$$\sigma_x{}^2=E(X^2)\text{-}(\mu_X)^2=\frac{1}{p}+\frac{2(1-p)}{p^2}-\frac{1}{p^2}=\frac{(1-p)}{p^2}$$

Hyper-geometric Distribution

Consider the sampling without replacement of a lot of (N) items, (k) of which are of one type and (N – k) of a second type. The probability of obtaining (x) items in a selection of (n) items without replacement obeys the hyper-geometric distribution:

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$



$x = 0 , 1 , 2 , \;\ldots\ldots\ldots , \min(n , k)$

NOTE:

$p=\dfrac{k}{N}$ is the ratio of items of type (I) to the total population

- ***Theorem:***

The mean and the variance of the hyper-geometric random variable are:

$$\mu_X=E(X)=n\frac{k}{N}=n\,p$$

$$\sigma_X^2=Var(X)=\frac{n\,k\,(N-k)\,(N-n)}{N^2\,(N-1)}=n\left(\frac{k}{N}\right)\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$=n\left(\frac{k}{N}\right)\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)=n\,p\,(1-p)\left(\frac{N-n}{N-1}\right)$$

EXAMPLE (3-19):

Fifty small electric motors are to be shipped. But before such a shipment is accepted, an inspector chooses 5 of the motors randomly and inspects them. If none of these tested motors are defective, the lot is accepted. If one or more are found to be defective, the entire shipment is inspected. Suppose that there are, in fact, three defective motors in the lot. What is the probability that the entire shipment is inspected?

SOLUTION:

Let (X) be the number of defective motors found, then (X) assumes the values (0 , 1 , 2 , 3).

P(entire shipment is inspected) = P(X ≥ 1) = 1 – P(X = 0)

$$P(X = x) = \frac{\binom{3}{x}\binom{47}{5-x}}{\binom{50}{5}}$$

$$P(X = 0) = \frac{\binom{3}{0}\binom{47}{5}}{\binom{50}{5}} = 0.72 \quad \textbf{\textit{(The lot is accepted)}}$$

P(X ≥ 1) = 1 – 0.72 = 0.28

EXAMPLE (3-20):

A committee of seven members is to be formed at random from a class with 25 students of whom 15 are girls. Find the probability that:

a- No girls are among the committee
b- All committee members are girls
c- The majority of the members are girls

SOLUTION:

Let (X) represents the number of girls in the committee.

a- $P(X = 0) = \frac{\binom{15}{0}\binom{10}{7}}{\binom{25}{7}}$

b- $P(X = 7) = \frac{\binom{15}{7}\binom{10}{0}}{\binom{25}{7}}$

c- P(majority are girls) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)

$$= \sum_{x=4}^{7} \frac{\binom{15}{x}\binom{10}{7-x}}{\binom{25}{7}}$$

Theorem:

For large (N), one can use the approximation:

$$\boldsymbol{P(X = x) \cong \binom{n}{x} P^{x}(1-P)^{n-x} \quad ; \quad P = \frac{k}{N}}$$

This approximation gives very good results if $\frac{n}{N} \le 0.1$, for the example above:

$$P = \frac{3}{50} = 0.06 \quad \Rightarrow \quad P(X = 0) \cong \binom{5}{0}(0.06)^{0}(1-0.06)^{5-0} = 0.733$$

III. Poisson Distribution

- ***Definition:***

A discrete random variable (X) is said to have a *Poisson distribution* if it has the following probability mass function:

$$P(X = x) = e^{-b}\frac{b^x}{x!} \quad ; \quad x = 0, 1, 2, \dots\dots\dots \qquad \text{where (b) is a positive constant.}$$

To verify that this is, indeed, a valid probability mass function we need to show that:

$$\sum_{x=0}^{\infty} e^{-b}\frac{b^x}{x!} = 1$$

The left hand side is expanded as:

$$\sum_{x=0}^{\infty} e^{-b}\frac{b^x}{x!} = e^{-b}\sum_{x=0}^{\infty} 1 + \frac{b}{1!} + \frac{b^2}{2!} + \frac{b^3}{3!} + \dots$$

The summation on the right side is easily recognized as the power series expansion of e^b. Therefore,

$$\sum_{x=0}^{\infty} e^{-b}\frac{b^x}{x!} = e^{-b}e^{b} = 1.$$

- ***Theorem:***

If (X) is a Poisson r.v with parameter (b), then its mean and variance are:

$$\boldsymbol{\mu_X = E(X) = b}$$
$$\boldsymbol{\sigma_X^2 = Var(X) = b}$$

Proof:

First we find the mean value of X. The method we follow is quite similar to the one used to find the mean and variance of the binomial distribution.

$$E(X) = \sum_{x=0}^{\infty} x e^{-b}\frac{b^x}{x!} = \sum_{x=1}^{\infty} x e^{-b}\frac{b^x}{x!}$$
$$= \sum_{x=1}^{\infty} x e^{-b}\frac{b^x}{x(x-1)!} = \sum_{x=1}^{\infty} e^{-b}\frac{b^x}{(x-1)!}$$

Let u = x-1 (or x = u + 1), and change the index of the summation from x to u. The result is

$$E(X) = \sum_{u=0}^{\infty} e^{-b}\frac{b^{u+1}}{u!} = b\sum_{u=0}^{\infty} e^{-b}\frac{b^u}{u!}$$

As was shown earlier, the summation on the right side equals 1. Therefore,

$$E(X) = b$$

which completes the proof.

To find the variance, we first find E(X(X-1))

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)e^{-b}\frac{b^x}{x!} = \sum_{x=2}^{\infty} x(x-1)e^{-b}\frac{b^x}{x(x-1)(x-2)!}$$

$$E(X(X-1)) = \sum_{x=2}^{\infty} e^{-b}\frac{b^x}{(x-2)!}$$

Let u = x-2 in the above summation , or x = u + 2, then

$$E(X(X-1)) = \sum_{u=0}^{\infty} e^{-b}\frac{b^{u+2}}{u!} = b^2\sum_{u=0}^{\infty} e^{-b}\frac{b^u}{u!} = b^2$$

But, $E(X(X-1)) = E(X^2) - E(X)$

Or, $E(X^2) = E(X(X-1)) + E(X)$

The variance of X can, therefore, be obtained as

$$\sigma_X^2 = E(X^2) - \mu_X^2 = E(X(X-1)) + E(X) - \mu_X^2$$

$$\sigma_X^2 = b^2 + b - b^2 = b$$

- ***Poisson Process:***

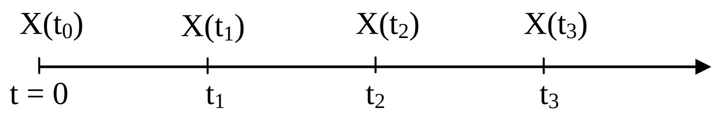
Consider a counting process in which events occur at a rate of (λ) occurrence per unit time. Let X(t) be the number of occurrences recorded in the interval (0 , t), we define the *Poisson process* by the following assumptions:

1- X(0) = 0 , i.e., we begin the counting at time t = 0.

2- For non-overlapping time intervals (0 , t_1) , (t_2 , t_3), the number of occurrences {X(t_1) – X(0)} and {X(t_3) – X(t_2)} are independent.

3- The probability distribution of the number of occurrences in any time interval depends only on the length of that interval.

4- The probability of an occurrence in a small time interval (Δt) is approximately ($\lambda\ \Delta$t).



Using the above assumptions, one can show that the probability of exactly (x) occurrences in any time interval of length (T) follows the Poisson distribution and,

$$\boldsymbol{P(X = x) = e^{-\lambda T}\,\frac{(\lambda T)^x}{x!}} \qquad ; \quad x = 0 , 1 , 2 , 3 , \ldots\ldots\ldots$$

- ***Theorem:***

Let (b) be a fixed number and (n) any arbitrary positive integer. For each nonnegative integer (x):

$$\lim_{n\to\infty}\binom{n}{x}(p)^x(1-p)^{n-x} = e^{-b}\,\frac{b^x}{x!} \qquad ; \text{ where } p = b/n$$

<u>EXAMPLE (3-21):</u>

Messages arrive to a computer server according to a Poisson distribution with a mean rate of 10 messages/hour.

a- What is the probability that 3 messages will arrive in one hour.
b- What is the probability that 6 messages will arrive in 30 minutes.

<u>SOLUTION:</u>

a- λ = 10 messages/hour ➔ T = 1 hour

$$P(X = x) = e^{-10\times 1}\,\frac{(10\times 1)^x}{x!} = e^{-10}\,\frac{(10)^x}{x!} \qquad ; \quad x = 0 , 1 , 2 , 3 , \ldots\ldots\ldots$$

$$P(X = 3) = e^{-10}\,\frac{(10)^3}{3!}$$

b- λ = 10 messages/hour ➔ T = 0.5 hour

$$P(X = x) = e^{-10\times\frac{1}{2}}\,\frac{(10\times\frac{1}{2})^x}{x!} = e^{-5}\,\frac{(5)^x}{x!} \qquad ; \quad x = 0 , 1 , 2 , 3 , \ldots\ldots\ldots$$

$$P(X=6) = e^{-5}\frac{(5)^6}{6!}$$

EXAMPLE (3-22):

The number of cracks in a section of a highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile.

a- What is the probability that there are no cracks in 5 miles of highway?
b- What is the probability that at least one crack requires repair in ½ miles of highway?
c- What is the probability that at least one crack in 5 miles of highway?

SOLUTION:

a- $\lambda = 2$ cracks/mile ➔ T = 5 miles

$$P(X=x) = e^{-2\times 5}\frac{(2\times 5)^x}{x!} = e^{-10}\frac{(10)^x}{x!} \quad ; \quad x = 0, 1, 2, 3, \ldots\ldots\ldots$$

$$P(X=0) = e^{-10}$$

b- $\lambda = 2$ cracks/mile ➔ T = 5 miles

$$P(X=x) = e^{-2\times\frac{1}{2}}\frac{(2\times\frac{1}{2})^x}{x!} = e^{-1}\frac{(1)^x}{x!} = \frac{e^{-1}}{x!} \quad ; \quad x = 0, 1, 2, 3, \ldots\ldots\ldots$$

$$P(X\geq 1) = \sum_{x=1}^{\infty}\frac{e^{-1}}{x!} = [1 - P(X=0)] = 1 - e^{-1}$$

c- $\lambda = 2$ cracks/mile ➔ T = 5 miles

$$P(X=x) = e^{-2\times 5}\frac{(2\times 5)^x}{x!} = e^{-10}\frac{(10)^x}{x!} \quad ; \quad x = 0, 1, 2, 3, \ldots\ldots\ldots$$

$$P(X\geq 1) = \sum_{x=1}^{\infty}\frac{e^{-10}(10)^x}{x!} = [1 - P(X=0)] = 1 - e^{-10}$$

EXAMPLE (3-23):

Given 1000 transmitted bits, find the probability that exactly 10 will be in error. Assume that the bit error probability is $\frac{1}{365}$.

SOLUTION:

X: random variable representing number of bits in error.

Exact solution:

$P(\text{bit error}) = \frac{1}{365}$; Number of trials (n) = 1000

Required number of bits in error (k) = 10

$$P(X=10) = \binom{n}{k}(p)^k(1-p)^{n-k} = \binom{1000}{10}\left(\frac{1}{365}\right)^{10}\left(\frac{364}{365}\right)^{990}$$

Approximate solution:

$$P(X=x) = e^{-b}\frac{b^x}{x!} \quad ; \quad b = n\,p = 1000\times\frac{1}{365} = \frac{1000}{365}$$

$$P(X=10) = e^{-b}\frac{b^{10}}{10!}$$

Exercise:

Perform the computation and compare the difference

- **Common Continuous Random Variables:**

I. Exponential Distribution:

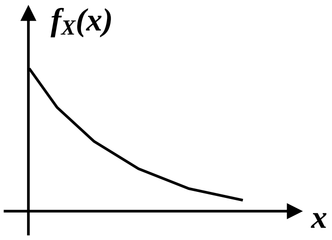
- ***Definition:***

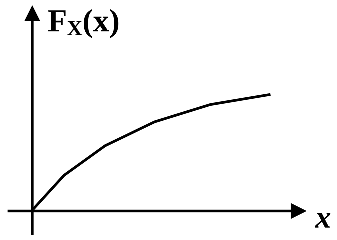
It is said that a random variable (X) has an exponential distribution with a parameter λ ($\lambda > 0$) if (X) has a continuous distribution for which the pdf $f_X(x)$ is given as:

$$f_X(x) = \lambda e^{-\lambda x} \qquad ; \qquad x \geq 0$$

The cumulative distribution function is:

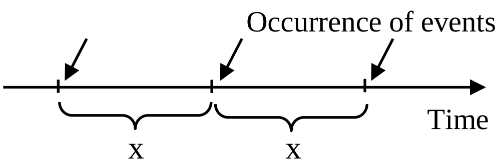
$$F_X(x) = 1 - e^{-\lambda x} \qquad ; \qquad x \geq 0$$





The exponential distribution is often used in a practical problem to represent the distribution of the time that elapses before the occurrence of some event. It has been used to represent the such periods of time as the period for which a machine or an electronic component will operate without breaking down, the period required to take care of a customer at some service facility, and the period between the arrivals of two successive customers at a facility.

If the event being considered occurs in accordance with a Poisson process, then both the waiting time until an event will occur and the period of time between any two successive events will have exponential distribution.



- ***Theorem:***

If the random variable (X) has an exponential distribution with parameter (λ), then:

$$\mu_X = E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \quad \text{and}$$

$$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$\sigma_X^2 = E(x^2) - E^2(x) = \frac{1}{\lambda^2}$$

Exercise

The number of telephone calls that arrive at a certain office is modeled by a Poisson random variable. Assume that on the average there are five calls per hour.

a. What is the average (mean) time between phone calls?
b. What is the probability that at least 30 minutes will pass without receiving any phone call?

c. What is the probability that there are exactly three calls in an observation interval of two consecutive hours?
d. What is the probability that there is exactly one call in the first hour and exactly two calls in the second hour of a two-hour observation interval?

EXAMPLE (3-24):

Suppose that the depth of water, measured in meters, behind a dam is described by an exponential random variable with pdf:

$$f_X(x) = \begin{cases} \frac{1}{13.5} e^{\frac{-x}{13.5}} & x > 0 \\ 0 & o.w \end{cases}$$

There is an emergency overflow at the top of the dam that prevents the depth from exceeding 40.6 m. There is a pipe placed 32.0 m below the overflow that feeds water to a hydroelectric generator (turbine).

a- What is the probability that water is wasted though emergency overflow?
b- What is the probability that water will be too low to produce power?
c- Given that water is not wasted in overflow, what is the probability that the generator will have water to derive it?

SOLUTION:

a- P(water wasted through emergency) = $P(X \geq 40.6\,\text{m}) = \int_{40.6}^{\infty} \frac{1}{13.5} e^{\frac{-x}{13.5}} dx = e^{-3}$

b- P(water too low to produce power) = P(x < 8.6 m) $= \left(1 - e^{-0.637}\right) = 0.47$

c- P(generator has water to derive it / water is not wasted) = P(x > 8.6 / x < 40.6)

$$= \frac{P(x > 8.6 \cap x < 40.6)}{P(x < 40.6)} = \frac{P(8.6 < x < 40.6)}{P(x < 40.6)} = \frac{\int_{8.6}^{40.6} \frac{1}{13.5} e^{\frac{-x}{13.5}} dx}{\int_{0}^{40.6} \frac{1}{13.5} e^{\frac{-x}{13.5}} dx = e^{-3}} = 0.504$$

II. Rayleigh Distribution:

The Rayleigh density and distribution functions are:

$$f_X(x) = \frac{2}{b} x e^{-x^2/b} \quad ; \quad x \geq 0$$

$$F_X(x) = 1 - e^{\frac{-x^2}{b}} \quad ; \quad x \geq 0$$

The Rayleigh pdf describes the envelope of white noise when passed through a band pass filter. It is used in the analysis of errors in various measurement systems.

- ***Theorem:***

$$\mu_X = E(X) = \sqrt{\frac{\pi b}{4}} \quad \text{and} \quad \sigma_X^2 = Var(X) = \frac{b(4 - \pi)}{4}$$

III. Cauchy Random Variable:

This random variable has:

$$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2}, \; F_X(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x}{\alpha}\right)$$

Exercise: Prove that the mean and variance of the Rayleigh distribution are as given in the theorem above.

Exercise: Find the mode and the median of the Rayleigh distribution

Exercise: Find the mean and variance of the Cauchy distribution.

Gaussian (Normal) Distribution:

- ***Definition:***

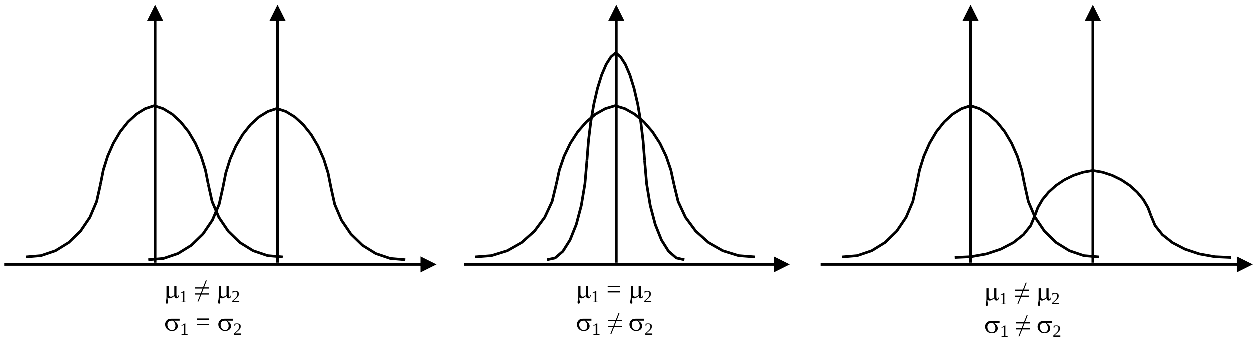
A random variable (X) with pdf:

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

has a normal distribution with parameters (μ_X) and (σ_X^2) where $-\infty < x < \infty$ and $\sigma_X^2 \geq 0$.
Furthermore:

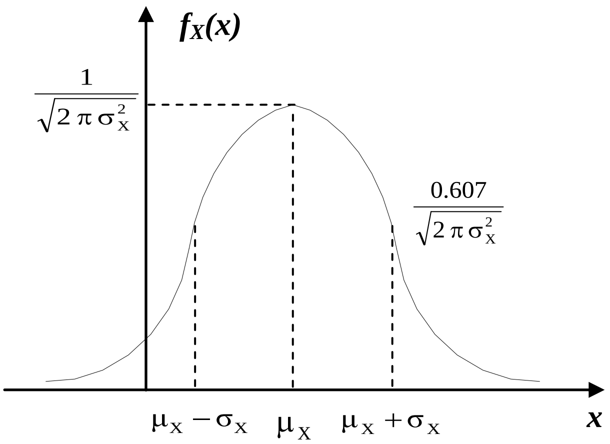
$E(x) = \mu_X$; $Var(x) = \sigma_X^2$

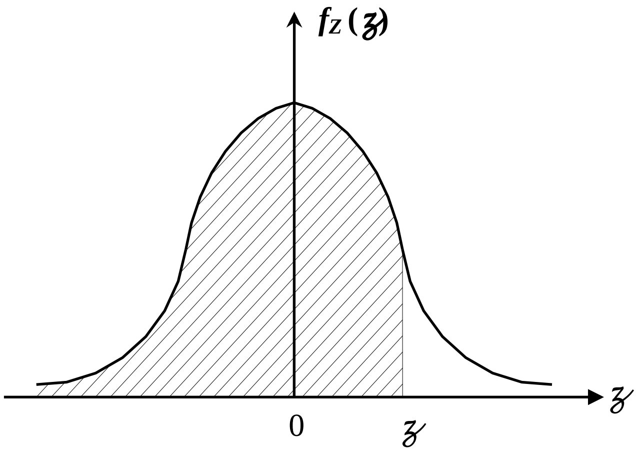
Infinite number of normal distributions can be formed by different combination of parameters.



- ***Definition:***

A normal random variable with mean zero and variance one is called a standard normal random variable. A standard normal random variable is denoted as Z.





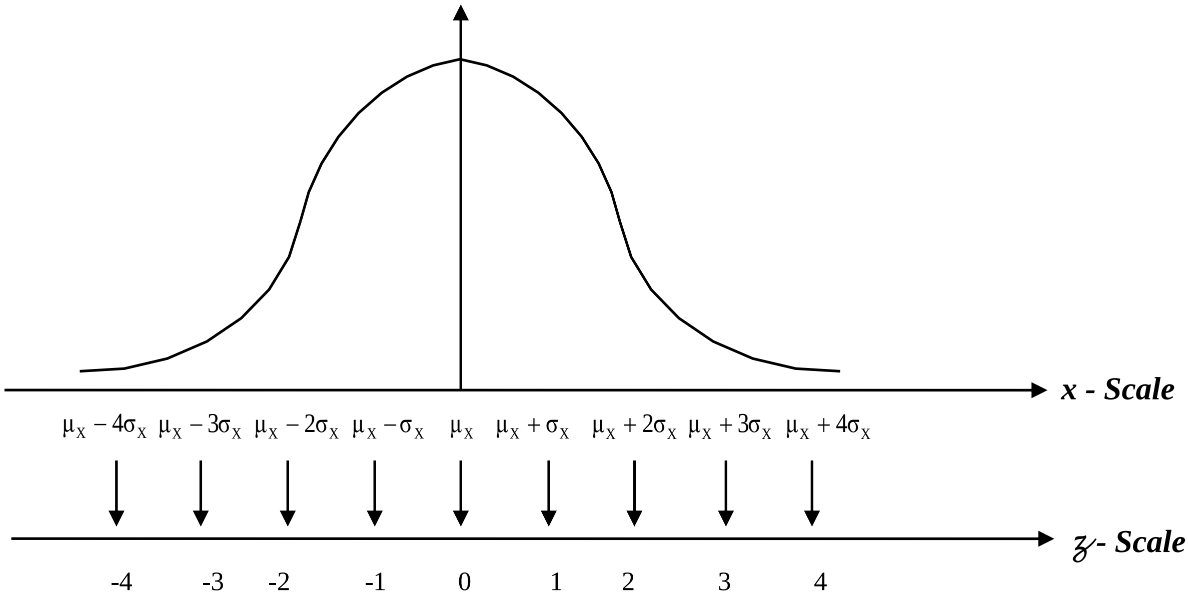
- ***Definition:***

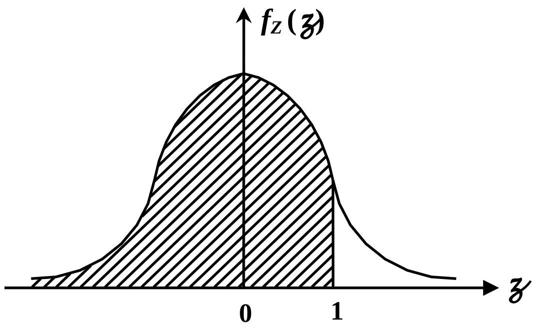
The function $\Phi(z) = P\{Z \leq z\}$ is used to denote the cumulative distribution function of a standard normal random variable:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}\, du$$

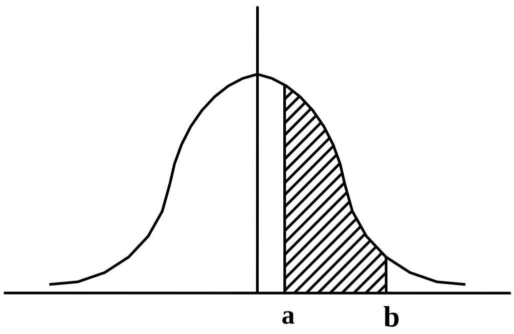
This function is tabulated for $z \geq 0$

For $z < 0$; $\Phi(z) = 1 - \Phi(-z)$

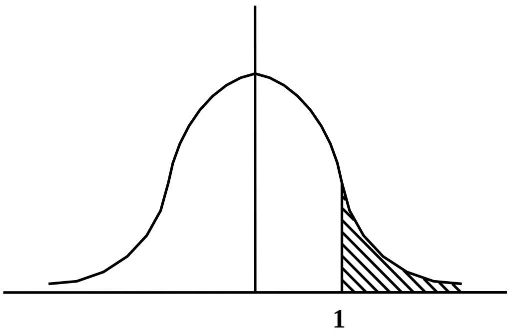




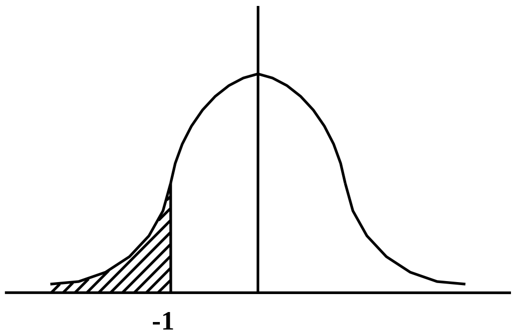
$\Phi(1)$



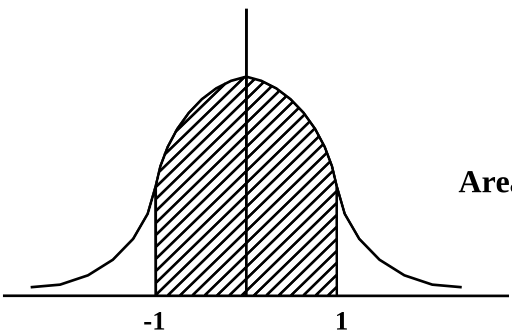
$\Phi(b) - \Phi(a)$



Area = $1 - \Phi(1)$



$\Phi(-1) = 1 - \Phi(1)$



Area = $\Phi(1) - \Phi(-1)$

$= \Phi(1) - [1 - \Phi(1)]$

$= 2\Phi(1) - 1$

- ***Cumulative Distribution Function:***

$$P(X \le x) = F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma_X^2}}\, e^{\frac{-(x-\mu_X)^2}{2\sigma_X^2}}\, dx$$

$$\text{Let } u = \left(\frac{x-\mu_X}{\sigma_X}\right) \;\Rightarrow\; du = \frac{dx}{\sigma_X} \;\Rightarrow\; dx = \sigma_X du$$

$$F_X(x) = \int_{-\infty}^{\frac{x-\mu_X}{\sigma_X}} \frac{1}{\sqrt{2\pi\sigma_X^2}}\, e^{\frac{-u^2}{2}}\, \sigma_X\, du$$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{\frac{-u^2}{2}}\, du$$

$$\Phi(z) = \Phi\left(\frac{Z-\mu_X}{\sigma_X}\right)$$

Therefore, we conclude that:

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

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

EXAMPLE (3-25):

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 mA and variance 4 $(mA)^2$. What is the probability that a measurement will exceed 13 mA?

SOLUTION:

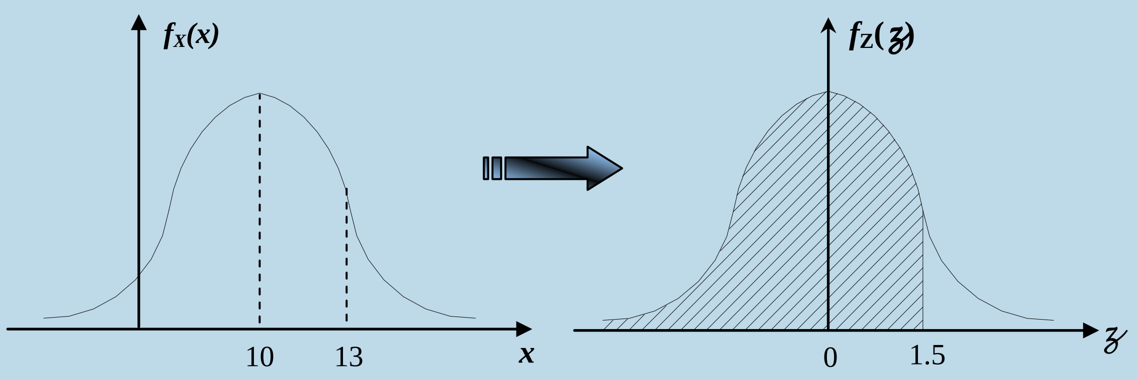
X = current in mA

$$Z = \left(\frac{X-\mu_X}{\sigma_X}\right) = \left(\frac{X-10}{2}\right)$$

$$P(X > 13) = P\left\{Z = \left(\frac{X-10}{2}\right) > \left(\frac{13-10}{2}\right) = 1.5\right\}$$

$P(X > 13) = P\{Z > 1.5\} = 1 - \Phi(z)$ ➔ *From tables:*

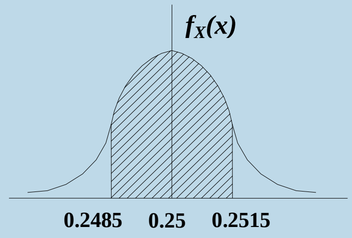
$= 1 - 0.93319 = 0.06681$



EXAMPLE (3-26):

The diameter of a shaft in an optical storage drive is normally distributed with mean 0.25 inch and standard deviation of 0.0005 inch. The specifications on the shaft are 0.25 ± 0.0015 inch. What proportion of shafts conforms to specifications?

SOLUTION:



$$P(0.2485 < X < 0.2515)$$

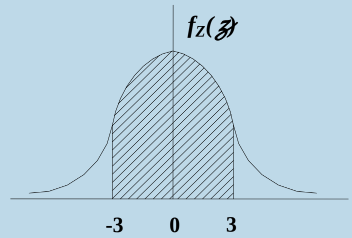
$$= P\left\{\left(\frac{0.2485 - 0.2500}{0.0005}\right) < Z < \left(\frac{0.2515 - 0.2500}{0.0005}\right)\right\}$$

$$= P\{-3 < Z < 3\}$$

$= \Phi(3) - \Phi(-3)$

$= 2\,\Phi(3) - 1$ ➔ *From tables:*

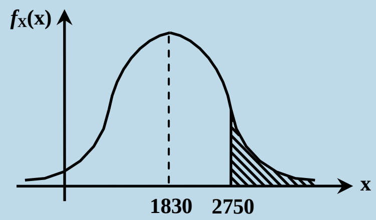
$= \Phi(3) - \Phi(-3) = (2 \times 0.99865) - 1 = 0.9973$



EXAMPLE (3-27):

Assume that the height of clouds above the ground at some location is a Gaussian random variable (X) with mean 1830 m and standard deviation 460 m. find the probability that clouds will be higher than 2750 m.

SOLUTION:



$$P(X > 2750) = 1 - P(X \le 2750)$$

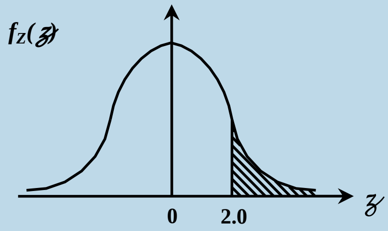
$$= 1 - P\left\{Z \le \left(\frac{2750 - 1830}{460}\right)\right\}$$

$$= 1 - P(Z \le 2.0)$$

$= 1 - \Phi(2.0)$ ➔ From tables:

$= 1 - 0.9772$

$$P(X > 2750) = 0.0228$$



Exercise

The tensile strength of paper is modeled by a normal distribution with a mean of 35 pounds per square inch and a standard deviation of 2 pounds per square inch.

a. If the specifications require the tensile strength to exceed 33 lb/in^2 , what is the probability that a given sample will pass the specification test?
b. If 10 samples undergo the specification test, what is the probability that at least 9 will pass the test?
c. If 20 samples undergo the test, what is the expected number of samples that pass the test?

Exercise

The rainfall over Ramallah district follows the normal distribution with a mean of 600 mm and a standard deviation of 80 mm. The rainfall is distributed over 500 km^2 area. Find:

1. The probability of obtaining a rainwater volume less than 206 MCM (MCM = Million Cubic Meter)
2. Find the mean and the standard deviation of the volume (V) of rainfall in MCM.
3. Flooding condition will be considered if the rainfall is higher than 900 mm.
   Find the probability of flooding for any given year.

- ***Remark:***
  The area under the Gaussian curve within (k) standard deviations of the mean is given in the following table:

k	**Area** $P(\mu_X - k\sigma_X \le X \le \mu_X + k\sigma_X)$
1	0.6826
2	0.9544
3	0.9973
4	0.99994 ➔

Total probability outside an interval of 4 standard deviations on each side of the mean is only 0.00006

- **Normal Approximation of the Binomial and Poisson Distribution:**

- ***Theorem: De-Moiver-Laplace***

For large (n) the binomial distribution

$$\binom{n}{x} p^x (1-p)^{n-x} \sim \frac{1}{\sqrt{2\,\pi\,n\,p\,q}}\, e^{\frac{-(x-np)^2}{2npq}} \qquad \textit{(\sim : asymptotically equal)}$$

Which is a normal distribution with mean (n p) and variance (n p q). Therefore, if (X) is a binomial r.v, then $Z = \left(\frac{X - n\,p}{\sqrt{n\,p\,q}}\right)$ is approximately a standard normal r.v.

The theorem gives better results when (n p > 5) and (n p q > 5)

$$P(a \le X \le b) = \sum_{x=a}^{b} \binom{n}{x} p^x (1-p)^{n-x} \sim \Phi(\beta) - \Phi(\alpha) \quad \text{where:}$$

$$\beta = \left(\frac{b - n\,p}{\sqrt{n\,p\,q}}\right) \quad \text{and} \quad \alpha = \left(\frac{a - n\,p}{\sqrt{n\,p\,q}}\right)$$

EXAMPLE (3-28):

Consider a binomial experiment with n = 1000 and p = 0.2. if X is the number of successes, find the probability that $X \le 240$.

SOLUTION:

Exact solution: $P(X \le 240) = \sum_{x=0}^{240} \binom{1000}{x} (0.2)^x (1-0.2)^{1000-x}$

Applying the Demoiver-Laplace theorem:

$$P(X < 240) = \Phi\left(\frac{240 - 1000 \times 0.2}{\sqrt{1000 \times 0.2 \times 0.8}}\right) = \Phi(3.162) = 0.999216$$

- ***Theorem:***

If (X) is a Poisson r.v with E(X) = b and Var(X) = b, then: $Z = \left(\dfrac{X - b}{\sqrt{b}}\right)$

is approximately a standard normal r.v. *The approximation is good for (b > 5).*

$$e^{-b}\frac{b^x}{x!} \rightarrow \frac{1}{\sqrt{2\pi b}} e^{\frac{-(x-b)^2}{2b}}$$

EXAMPLE (3-29):

Assume the number of asbestos particles in a cm^3 of dust follow a Poisson distribution with a mean of 1000. If a cm^3 of dust is analyzed, what is the probability that less than 950 particles are found in 1 cm^3?

SOLUTION:

Exact solution: $P(X \le 950) = \sum_{x=0}^{950} e^{-1000}\dfrac{(1000)^x}{x!}$

Approximate: $P(X < 950) = P\left\{Z \le \dfrac{950 - 1000}{\sqrt{1000}}\right\} = P\{Z \le -1.58\} = 0.057$

Transformation of Random Variables:

Let (X) be a random variable with a pdf $f_X(x)$. If Y = g(X) is a function of (X), then (Y) is a random variable. Its pdf is to be determined. The function g(X) is a single valued function of its argument.

I. Discrete Case:

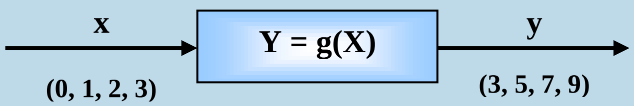
EXAMPLE (3-30):

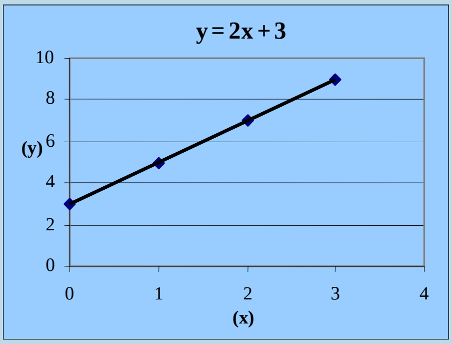
Let (X) be a binomial r.v with parameters (n = 3) and (p = 0.75). Let Y = g(x) = 2X + 3
P(Y = y) = P(X = x)

SOLUTION:

The table below shows the (x) and (y) values and their probabilities.

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





x	y	P(X = x)	P(Y = y)
0	3	$(1-p)^3$	$(1-p)^3$
1	5	$3\,p\,(1-p)^2$	$3\,p\,(1-p)^2$
2	7	$3\,p^2\,(1-p)$	$3\,p^2\,(1-p)$
3	9	p^3	p^3

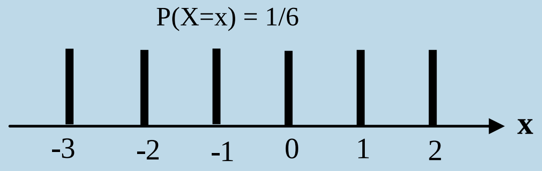
EXAMPLE (3-31):

Let (X) has the distribution $P\{X = x\} = \dfrac{1}{6}$; x = -3 , -2 , -1 , 0 , 1 , 2

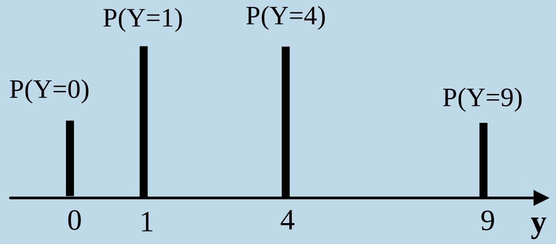
Define $Y = g(x) = X^2$. Find the pdf of the random variable Y.

SOLUTION:

x	y	P(X = x)	P(Y = y)
7	9	1/6	1/6
-2	4	1/6	1/6
-1	1	1/6	1/6
0	0	1/6	1/6
1	1	1/6	1/6
2	4	1/6	1/6



Probability Density Function $f_X(x)$



Probability Density Function $f_Y(y)$

The distribution of Y is:

P(Y = 0) = 1/6
P(Y = 1) = 2/6
P(Y = 4) = 2/6
P(Y = 9) = 1/6

II. Continuous Case:

Let Y = g(X) be a monotonically increasing or decreasing function of (x).

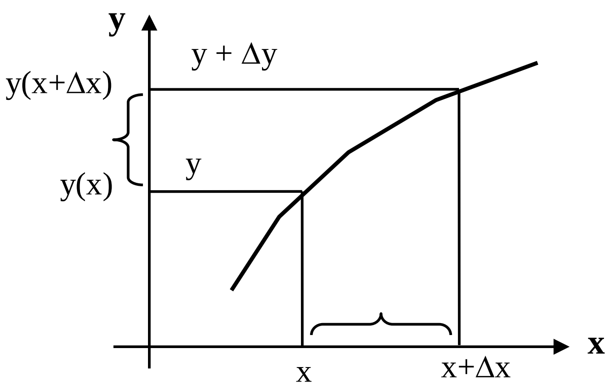
$$P(x < X < x + \Delta x) = P\{y(x) < Y < y(x + \Delta x)\}$$

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

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

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

$y_1 < y < y_2$



EXAMPLE (3-32):

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

SOLUTION:

$$f_Y(y) = 2\frac{f_X(x)}{|dy/dx|} \qquad 0 \le y \le \infty \quad ; \quad \left|\frac{dy}{dx}\right| = 2x$$

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

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

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

EXAMPLE (3-33):

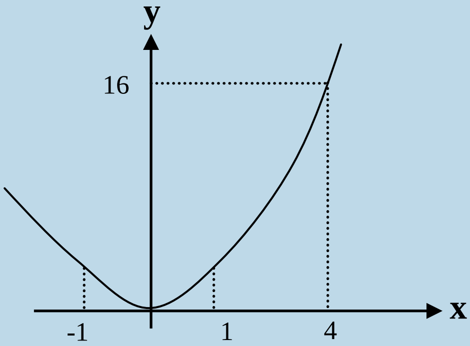
Let (X) be a uniform r.v in the interval (-1 , 4). If Y = X^2. Find $f_Y(y)$

SOLUTION:

For $(-1 \le X \le 1)$: ➔ $f_Y(y) = 2\frac{f_X(x)}{|dy/dx|} = \frac{2 \times 1/5}{2\,x} = \frac{1}{5\sqrt{y}}$

For $(1 < X \le 4)$: ➔ $f_Y(y) = \frac{f_X(x)}{|dy/dx|} = \frac{1/5}{2\,x} = \frac{1}{10\sqrt{y}}$

$$f_Y(y) = \begin{cases} \dfrac{1}{5\sqrt{y}} & 0 < y \le 1 \\ \dfrac{1}{10\sqrt{y}} & 1 < y \le 16 \\ 0 & \text{Otherwise} \end{cases}$$



EXAMPLE (3-32):

Let (X) be a r.v with the exponential pdf: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$

Let Y = 2X + 3. Find $f_Y(y)$ and the region over which it is defined.

SOLUTION:

$$f_Y(y) = \frac{f_X(x)}{|dy/dx|} \quad ; \quad \left|\frac{dy}{dx}\right| = 2$$

$$f_Y(y) = \frac{f_X(x)}{2} \text{ , but } x = \frac{y-3}{2} \quad \Rightarrow \quad f_Y(y) = \frac{f_X\left(\frac{y-3}{2}\right)}{2}$$

$$f_Y(y) = \begin{cases} \frac{\lambda}{2} e^{-\lambda(\frac{y-3}{2})} & \frac{y-3}{2} > 0 \\ 0 & \frac{y-3}{2} < 0 \end{cases} \quad \rightarrow \quad f_Y(y) = \begin{cases} \frac{\lambda}{2} e^{-\lambda\left(\frac{y-3}{2}\right)} & y > 3 \\ 0 & y < 3 \end{cases}$$

NOTE: $P(3 < Y \le 5) = P(0 < X \le 1) = \int_0^1 \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda}$

EXAMPLE (3-33):

Let (X) be a Gaussian r.v with mean (μ_X) variance (σ_X^2)
Let (Y) = aX + b be any r.v. Find $f_Y(y)$

SOLUTION:

$$Y = aX + b \quad \rightarrow \quad \left|\frac{dy}{dx}\right| = a \text{ and } x = \frac{y-b}{a}$$

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

but from previous results we have: $\mu_Y = a\mu_x + b$ and $\sigma_Y^2 = a^2\sigma_X^2$ hence,

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

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

SPECIAL CASE: The Standard Normal

If (X) is a Gaussian r.v with mean (μ_X) variance (σ_X^2), then:

the r.v $Z = \left(\frac{X-\mu_X}{\sigma_X}\right)$ is a Gaussian r.v with mean ($\mu_Z = 0$) variance ($\sigma_Z^2 = 1$).

That is **(Z) is a standard normal random variable**.

GENERAL RESULT:

A linear transformation of a Gaussian random variable is also Gaussian.

CHAPTER I I I

PROBABILITY DISTRIBUTIONS FOR MORE THAN ONE RANDOM VARIABLE

In certain experiments we may be interested in observing several quantities as they occur, such as carbon content (X) and hardness (Y) of steel; input (X) to a system and output (Y) at a given time too.

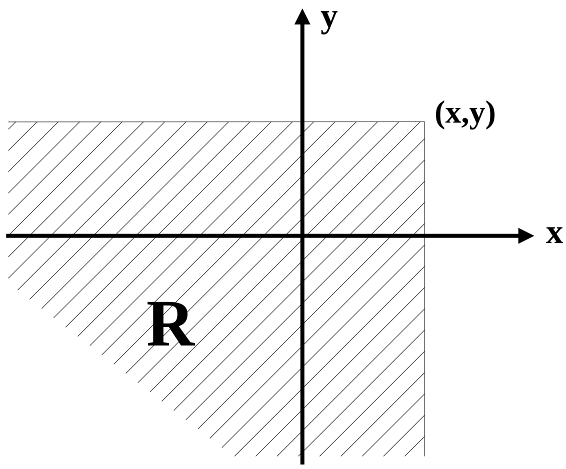
- If we observe two quantities (X) and (Y), each trial gives a pair of values X = x and Y = y, (x,y) which represents a point (x,y) in the x-y plane.

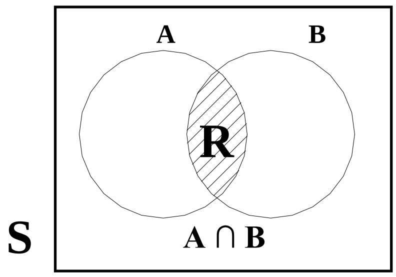
- The joint cumulative distribution function of two r.v X and Y is defined as:

$F_{XY}(x,y) \triangleq P\{X \le x , Y \le y\}$

Event (A) = $\{X \le x\}$

Event (B) = $\{Y \le y\}$





I. <u>Discrete Two Dimensional Distribution:</u>

A random variable (x,y) and its distribution are called discrete if (x,y) can assume only countably finite or at most countably infinite pairs of values (x_1,y_1), (x_2,y_2), ……

The joint probability mass function of (X) and (Y) is:

$P_{ij} = P\{X = x_i , Y = y_j\}$ such that

$$F_{XY}(x,y) = \sum_{x_i \le x} \sum_{y_j \le y} P_{ij}$$

and
$$\sum_{i} \sum_{j} P_{ij} = 1$$

II. <u>Continuous Two Dimensional Distribution:</u>

A random variable (x,y) and its distribution are called continuous if $F_{XY}(x,y)$ can be given by:

$$F_{XY}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(x, y)\, dx\, dy$$

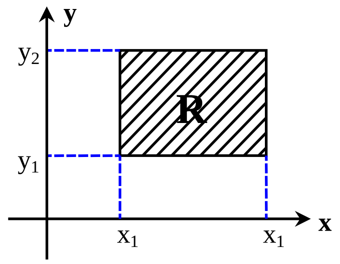
where $f_{XY}(x,y)$ is the joint probability density function *(f being continuous and nonnegative)*

- ***Properties of the joint pdf:***

1- $f_{XY}(x,y) \geq 0$

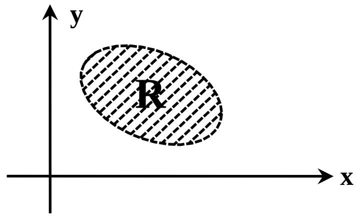
2- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$

3- $P(x_1 < X \leq x_2 , y_1 < Y \leq y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y)\,dx\,dy$



and in general:

$$P(x , y \in R) = \iint_R f_{XY}(x,y)\,dx\,dy$$



- **Marginal Distributions of a Discrete Distribution:**

$P(X = x_i) = P(X = x_i , Y \text{ arbitrary})$

$= \sum_y P(X = x_i , Y = y_j)$

This is the probability that (X) may assume a value (x), while (Y) may assume any value which we ignore.

Likewise:

$P(Y = y_j) = \sum_x P(X = x_i , Y = y_j)$

- **Marginal Distributions of a Continuous Distribution:**

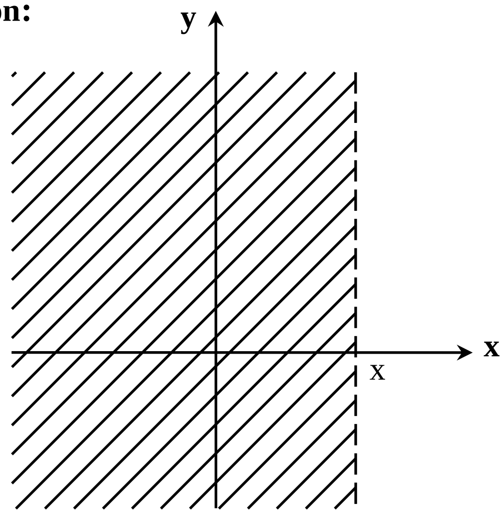
For a continuous distribution we have:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\right)dx$$

but $f_X(x) = \dfrac{d}{dx} F_X(x)$

➔ $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$; Marginal pdf

➔ $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$; Marginal pdf



Region such that:
$\{X \leq x\}$

- **Independence of Random Variable:**

- ***Theorem:***

Two random variables (X) and (Y) are said to be independent if:

$F_{XY}(x,y) = F_X(x)\,F_Y(y)$ holds for all (x,y), or equivalently:

$f_{XY}(x,y) = f_X(x)\,f_Y(y)$

Proof:

$F_{XY}(x,y) = P\{X \leq x , Y \leq y\}$

Let: A: event $\{X \leq x\}$

B: event $\{Y \leq y\}$

A and B are independent if:

$P(A \cap B) = P(A)\,P(B)$

$P(X \leq x , Y \leq y) = P(X \leq x)\,P(Y \leq y)$

- **Conditional Densities:**

Let (X) and (Y) be discrete random variables. The conditional probability density function of (Y) given (X = x), that is the probability that (Y) takes on the value (y) given that (X = x), is given by:

$$f_{Y/X}(y) = P(Y = y / X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

If (X) and (Y) are continuous, then the conditional pdf of (Y) given (X = x) is given by:

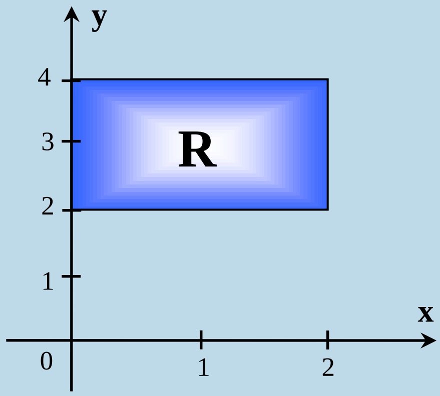
$$f_{Y/X}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

EXAMPLE (4-1):

Let (X) and (Y) be continuous random variables with joint pdf:

$$f_{X,Y}(x,y) = \frac{1}{8}(6 - x - y) \quad ; \quad 0 < x \le 2 \,, \; 2 < y \le 4$$

1- Find $f_X(x)$ and $f_Y(y)$.
2- Find $f_{Y/X}(y)$.
3- Find $P(2<y\le 3)$
4- Find $P(2<y\le 3 / x=1)$



SOLUTION:

1- $$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$= \int_2^4 \frac{1}{8}(6 - x - y)\,dy = \frac{1}{8}(6 - 2x) \quad ; \quad 0 < x \le 2$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_0^2 \frac{1}{8}(6 - x - y)\,dx = \frac{1}{4}(5 - y) \quad ; \quad 2 < y \le 4$$

2- $$f_{Y/X}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \frac{\frac{1}{8}(6 - x - y)}{\frac{1}{8}(6 - 2x)} = \frac{(6 - x - y)}{(6 - 2x)} \quad ; \quad 0 < x \le 2 \quad , \quad 2 < y \le 4$$

3- $$P(2<Y\le 3) = \int_2^3 f_Y(y)\,dy = \int_2^3 \frac{1}{4}(5 - y)\,dy = \frac{5}{8}$$

4- $$f_{Y/X}(y / x = 1) = \frac{(5 - y)}{4} \quad ; \quad 2 < y \le 4$$

$$P(2<Y\le 3 / X=1) = \int_2^3 \frac{5 - y}{4}\,dy = \frac{5}{8}$$

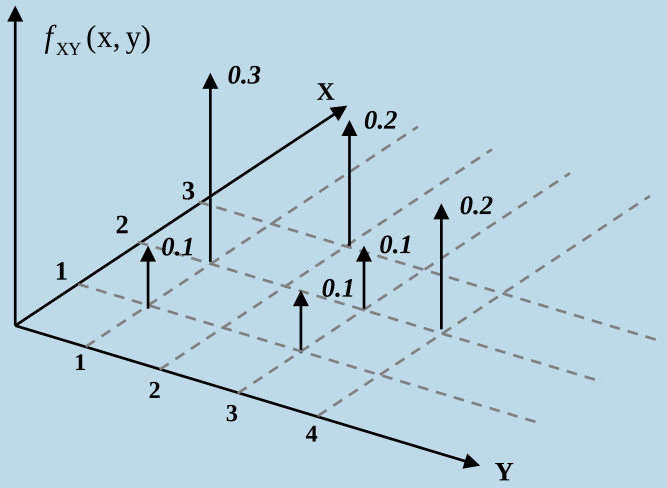
Exercise:

1- Find $P(2<Y\le 3 / 0\le X\le 1)$
2- Find μ_X, μ_Y, σ_X^2, and σ_Y^2
3- Are X and Y independent?

EXAMPLE (4-2):

Suppose that the random variable (X) can take only the values (1 , 2 , 3) and the random variable (Y) can take only the values (1 , 2 , 3 , 4). The joint pdf is shown in the table.

X \ Y	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0



1- Find $f_X(x)$ and $f_Y(y)$.

2- Find $P(X \geq 2)$

3- Are (X) and (Y) independent.

SOLUTION:

1- $P(X = 1) = 0.2$ $\qquad P(Y = 1) = 0.4$

$P(X = 2) = 0.6$ $\qquad P(Y = 2) = 0.2$

$P(X = 3) = 0.2$ $\qquad P(Y = 3) = 0.2$

$P(Y = 4) = 0.2$

$\sum = \mathbf{1.0}$ $\qquad \sum = \mathbf{1.0}$

2- $P(X \geq 2) = P(X = 2) + P(X = 3) = 0.6 + 0.2 = 0.8$

3- Check all pairs (x,y) for:

$P(X = x, Y = y) \overset{?}{=} P(X = x)\ P(Y = y)$

$P(X = 1 , Y = 1) = 0.1 \neq (0.2 \times 0.4) = 0.08$ ➔ *we do not continue*

➔ ***X and Y are not independent***

Exercise:

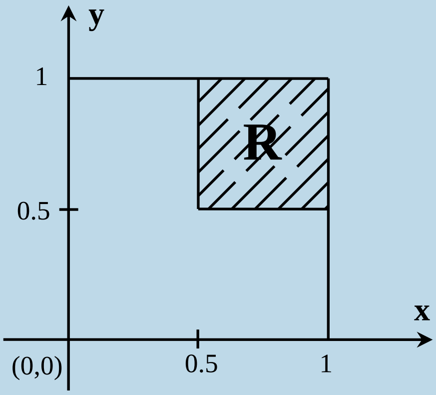
– Find μ_X and σ_X^2

EXAMPLE (4-3):

Let (X) and (Y) have the joint pdf:

$$f_{XY}(x,y)=\begin{cases} k\,x\,y\;, & 0\le x\le 1,\ 0\le y\le 1,\\ 0\;, & \text{otherwise}\end{cases}$$

a- Find (k) so that $f_X(x)$ is a proper pdf.
b- Find P(X > 0.5 , y > 0.5)
c- Find $f_X(x)$ and $f_Y(y)$
d- Are (X) and (Y) independent.



SOLUTION:

a- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy=1 \;\Rightarrow\; k\int_0^1 y\left(\int_0^1 x\,dx\right)dy=1$

$$k\int_0^1 y\left(\frac{x^2}{2}\Big|_0^1\right)dy=\frac{k}{2}\int_0^1 y\,dy=\frac{k}{2}\,\frac{y^2}{2}\Big|_0^1=\frac{k}{4} \;\Rightarrow\; \therefore \frac{k}{4}=1 \;\Rightarrow\; k=4$$

b- $P(X > 0.5 , y > 0.5) = \int_{0.5}^{1}\int_{0.5}^{1} 4\,x\,y\,dx\,dy$

$$4\left(\frac{x^2}{2}\Big|_{0.5}^{1}\right)\left(\frac{y^2}{2}\Big|_{0.5}^{1}\right)=(1-0.25)\times(1-0.25)=0.75\times0.75=0.5625$$

c- $f_X(x)=\int_0^1 f_{XY}(x,y)\,dy$

$$f_X(x)=\int_0^1 4\,x\,y\,dy \;=\; 4\,x\,\frac{y^2}{2}\Big|_0^1=2\,x$$

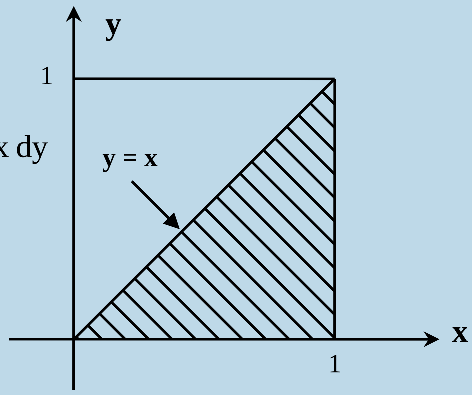
$$f_Y(y)=\int_0^1 4\,x\,y\,dx \;=\; 4\,y\,\frac{x^2}{2}\Big|_0^1=2\,y$$

Since $f_{XY}(x,y) = f_X(x)\, f_Y(y)$

➔ $4\,x\,y \;=\; (2\,x)\,(2\,y)$ ➔ (X) and (Y) are independent.

EXAMPLE (4-5):

For example (4-3), find P(X > Y).



SOLUTION:

$$P(Y<X)=\iint_R 4\,x\,y\,dx\,dy \;=\int_0^1\int_0^x 4\,x\,y\,dy\,dx \;=\int_0^1\int_y^1 4\,x\,y\,dx\,dy$$

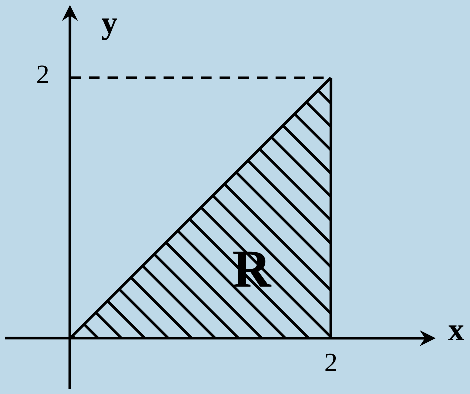
$$=\int_0^1 4\,x\left(\frac{y^2}{2}\Big|_0^x\right)dx \;=\int_0^1 4\,x\,\frac{x^2}{2}\,dx=2\int_0^1 x^3\,dx$$

$$P(Y<X)=2\left(\frac{x^4}{4}\Big|_0^1\right)=\frac{1}{2}$$

EXAMPLE (4-4):

Two random variables (X) and (Y) have the joint pdf:

$$f_{XY}(x,y)=\begin{cases}\frac{5}{16}x^2 y & , \quad 0<y<x<2 \\ 0 & , \quad \text{otherwise}\end{cases}$$



a- Verify that $f_{XY}(x,y)$ is a valid pfd.
b- Find the marginal density functions of X and Y.
c- Are X and Y statistically independent?
d- Find P{X<1} , P{Y<0.5} , P{XY<1}

SOLUTION:

a- $$\iint_R f_{XY}(x,y)\,dx\,dy = \int_0^2\int_0^x \frac{5}{16}x^2 y\,dy\,dx$$

$$\frac{5}{16}\int_0^2 x^2\left(\int_0^x y\,dy\right)dx = \frac{5}{16}\int_0^2 x^2\left(\frac{y^2}{2}\bigg|_0^x\right)dx = \frac{5}{16}\int_0^2 \frac{x^4}{2}dx = \frac{5}{16}\times\frac{1}{2}\times\frac{x^5}{5}\bigg|_0^2 = \frac{5}{16}\times\frac{1}{2}\times\frac{32}{5}=1$$

b- $$f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \Rightarrow f_X(x)=\int_0^x \frac{5}{16}x^2\,y\,dy = \frac{5}{16}x^2\frac{y^2}{2}\bigg|_0^x = \frac{5}{32}x^4$$

$$f_X(x)=\begin{cases}\frac{5}{32}x^4 & , \quad 0<x<2 \\ 0 & , \quad \text{otherwise}\end{cases} \Rightarrow \text{check} \int_0^2 \frac{5}{32}x^4\,dx = \frac{5}{32}\frac{x^5}{2}\bigg|_0^2 = 1 \quad \Rightarrow \text{OK}$$

$$f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx \Rightarrow f_Y(y)=\int_y^2 \frac{5}{16}x^2\,y\,dx = \frac{5}{16}y\frac{x^3}{3}\bigg|_y^2 = \frac{5}{48}y(8-y^3)$$

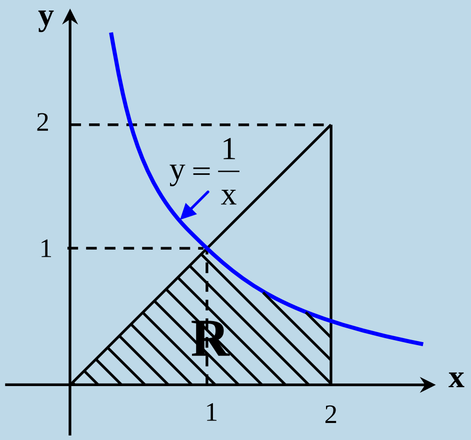
$$f_Y(y)=\begin{cases}\frac{5}{48}y(8-y^3) & , \quad 0<y<2 \\ 0 & , \quad \text{otherwise}\end{cases} \Rightarrow \int_0^2 \frac{5}{48}(8y-y^4)\,dy = \frac{5}{48}\left(4y^2-\frac{y^5}{5}\bigg|_0^2\right) = 1 \quad \Rightarrow \text{OK}$$

c- Since $f_{XY}(x,y) \neq f_X(x)\,f_Y(y) \Rightarrow$ (X) and (Y) are not statistically independent.

d- $$P\{X<1\} = \int_0^1 f_X(x)\,dx = \int_0^1 \frac{5}{32}x^4\,dx = \frac{5}{32}\frac{x^5}{5}\bigg|_0^1 = \frac{1}{32} = 0.03125$$

$$P\{Y<0.5\} = \int_0^{0.5} f_Y(y)\,dy = \int_0^{0.5}\frac{5}{48}y(8-y^3)\,dy$$

$$= \frac{5}{48}\left(4y^2 - \frac{y^4}{4}\bigg|_0^{0.5}\right) = \frac{105}{1024} = 0.1025$$



$$P\{XY<1\} = P\{Y<\frac{1}{X}\}$$

$$P\{Y<\frac{1}{X}\} = \iint_R f_{XY}(x,y)\,dx\,dy$$

$$P\{Y<\frac{1}{X}\} = \int_0^1\int_0^x \frac{5}{16}x^2 y\,dy\,dx + \int_1^2\int_0^{\frac{1}{x}} \frac{5}{16}x^2 y\,dy\,dx$$

=1/32 + 5/32 = 6/32

Addition of Means and Variances:

- ***Review: (Basic operations on a single random variable)***

  - $E(X) = \int_{-\infty}^{\infty} x \, f_X(x)\, dx$
  - $E\{g(X)\} = \int_{-\infty}^{\infty} g(X) \, f_X(x)\, dx$
  - $E\{g_1(X) + g_2(X)\} = E\{g_1(X)\} + E\{g_2(X)\}$
  - If Y = aX + b

    E(Y) = a E(X) + b ➔ $\mu_Y = a\,\mu_X + b$ ➔ $\sigma_Y^2 = a^2 \sigma_X^2$

- ***Definition:***

  The expected value of a function g(x,y) of two random variables (X) and (Y) is:

  $E\{g(x,y)\} = \sum_{x_i} \sum_{y_j} g(x_i, y_j)\, P(X = x_i, Y = y_j)$; X and Y are discrete

  $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, f_{XY}(x, y)\, dx\, dy$; X and Y are continuous

  Since summation and integration are linear processes, we have:

  $E\{a\, g_1(x,y) + b\, g_2(x,y)\} = a\, E\{g_1(x,y)\} + b\, E\{g_2(x,y)\}$

- ***Theorem: <u>Addition of Means</u>***

  The mean or expected value of a sum of random variables is the sum of the expectations.

  $E(x_1 + x_2 + \ldots\ldots + x_n) = E(x_1) + E(x_2) + \ldots\ldots + E(x_n)$

- ***Theorem: <u>Multiplication of Means</u>***

  The expected value of the product of independent r.v equals the product of the expected values.

  $E(x_1\, x_2 \ldots\ldots X_n) = E(x_1)\, E(x_2) \ldots\ldots E(x_n)$

  <u>*Proof:*</u>

  If (X) and (Y) are *independent* random variables, then $f_{XY}(x,y) = f_X(x)\, f_Y(y)$, so:

  $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y\, f_{XY}(x, y)\, dx\, dy = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = E(X)\, E(Y)$

  And in general, if (X) and (Y) are independent, then:

  $E\{g_1(X)\, g_2(Y)\} = E\{g_1(X)\}\, E\{g_2(Y)\}$

- ***Theorem: <u>Addition of Variances</u>***

- ***Definition:***

  The correlation coefficient between two random variables (X) and (Y) is:

  $\rho_{XY} \triangleq \dfrac{E\{(X - \mu_X)(Y - \mu_Y)\}}{\sigma_X\, \sigma_Y} = \dfrac{\mu_{XY}}{\sigma_X\, \sigma_Y}$

  where μ_{XY} is called the covariance and ρ_{XY} is bounded between $-1 \le \rho_{XY} \le 1$

  when $\rho_{XY} = 0$, (X) and (Y) are said to be uncorrelated.

  when $\rho_{XY} = \pm 1$, (X) and (Y) are said to be fully correlated.

- ***Theorem:***
  Let $Y = a_1X_1 + a_2X_2$, then
  $$\sigma_Y^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + 2a_1a_2\sigma_{X_1}\sigma_{X_2}\rho_{X_1X_2}$$
  Proof:
  $$\sigma_Y^2 = E\{(Y-\mu_Y)^2\} = E\{(a_1X_1 + a_2X_2 - a_1\mu_{X_1} - a_2\mu_{X_2})^2\}$$
  $$= E\{[a_1(X_1-\mu_{X_1}) + a_2(X_2-\mu_{X_2})]^2\}$$
  $$= E\{a_1^2(X_1-\mu_{X_1})^2\} + E\{a_2^2(X_2-\mu_{X_2})^2\} + 2a_1a_2E\{(X_1-\mu_{X_1})(X_2-\mu_{X_2})\}$$
  $$= a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + 2a_1a_2\mu_{XY}$$
  since $\rho_{XY} = \dfrac{\mu_{XY}}{\sigma_X\sigma_Y}$ $\quad\therefore\ \sigma_Y^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + 2a_1a_2\sigma_{X_1}\sigma_{X_2}\rho_{XY}$

- ***Theorem: Multiplication of Means***
  If (X) and (Y) are *independent* random variables, then they are uncorrelated.
  Proof:
  $$\mu_{XY} = E\{(X-\mu_X)(Y-\mu_Y)\}$$
  $$= E\{XY\} - \mu_Y E\{X\} - \mu_X E\{Y\} + \mu_X\mu_Y$$
  $$= E\{XY\} - E\{X\}E\{Y\}$$
  But since (X) and (Y) are *independent*, then $E\{XY\} = E\{X\}E\{Y\}$
  $$\Rightarrow\ \mu_{XY} = 0\ \Rightarrow\ \rho_{XY} = \frac{\mu_{XY}}{\sigma_X\sigma_Y} = 0$$
  This result asserts that if X and Y are independent then they are uncorrelated ($r_{XY} = 0$). However, the converse is not necessarily true. That is, if $r_{XY} = 0$, then X and Y are nor necessarily independent. The only exception is when X and Y are Gaussian. In this case, $r_{XY} = 0$ implied that X and Y are independent.

- ***Theorem:***
  Let $Y = a_1X_1 + a_2X_2$, and (X) and (Y) are *independent* random variables, then
  $$\sigma_Y^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2$$
  This result follows immediately from the above two theorems.

The sum of independent random variables equals the sum of the variances of these variables.

- **Functions of Random Variables:**

- Let (X) and (Y) be random variables with a joint pdf $f_{XY}(x,y)$ and let g(x,y) be any continuous function that is defined for all (x,y). then:
  Z = g(x,y) is a random variable. The objective is to find $f_Z(z)$.
- When (X) and (Y) are discrete random variables, we may obtain the probability mass function P{Z = z} by summing all probabilities for which g(x,y) equals the value of (z) considered, thus:
  $$P(Z=z) = \sum_{g(x,y)=z}\sum P(X=x_i\,,\,Y=y_j)$$
- In the case of continuous random variables (X) and (Y) we find $F_Z(z)$ first:
  $$F_Z(z) = P\{Z \le z\} = \iint_{g(x,y)\le z} f_{XY}(x,y)\,dx\,dy$$
  Then we find: $f_Z(z) = \dfrac{dF_z(z)}{dz}$

- ***Theorem:***

Let Z = X + Y and let (X) and (Y) be *independent* random variables, then;

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z - x)\, dx$$

<u>*Proof:*</u>

$$F_Z(z) = P(Z \le z) = P(X + Y \le z)$$

$$F_Z(z) = P(Y \le z - X) = \iint_R f_{XY}(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{XY}(x, y)\, dx\, dy$$

since (X) and (Y) be *independent* random variables, then

$f_{XY}(x,y) = f_X(x)\, f_Y(y)$, so:

$$F_z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x)\, f_Y(y)\, dx\, dy$$

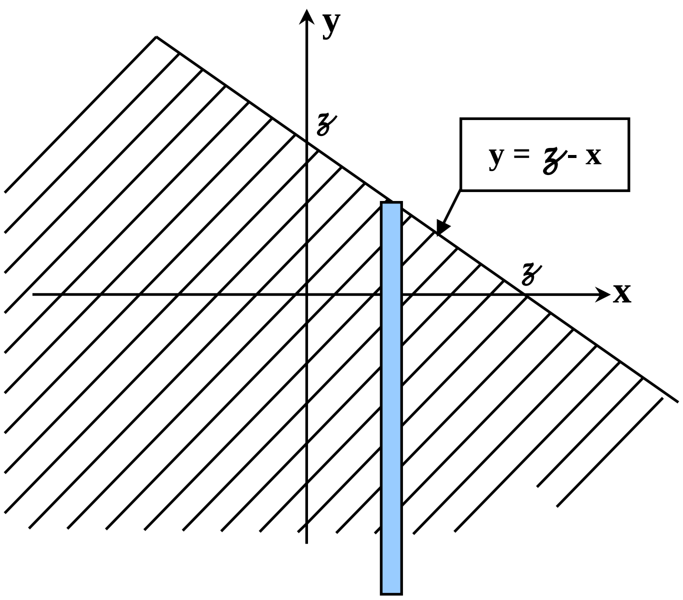
$$F_z(z) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_Y(y)\, dy \right) f_X(x)\, dx$$

$$F_z(z) = \int_{-\infty}^{\infty} f_X(x)\, F_Y(z - x)\, dx$$

$$f_Z(z) = \frac{d\, F_z(z)}{dz}$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z - x)\, dx$$

The Convolution Integral



EXAMPLE (4-6):

Consider the joint pdf shown in the table *(considered before in example 4-1).*
Let Z = X + Y.

1- Find the probability mass function of (Z), P{Z = z}.
2- Find P(X = Y).
3- Find E{XY}

X \ Y	**1**	**2**	**3**	**4**
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

SOLUTION:

1- Possible values of (Z) and their probabilities are shown as follows:

Z	**P(Z = z)**
2	P(X = 1 , Y = 1) = 0.1
3	P(X = 1 , Y = 2) + P(X = 2 , Y = 1) = 0 + 0.3 = 0.3
4	P(X = 1 , Y = 3) + P(X = 3 , Y = 1) + P(X = 2 , Y = 2) = 0.1 + 0 + 0 = 0.1
5	P(X = 1 , Y = 4) + P(X = 2 , Y = 3) + P(X = 3 , Y = 2) = 0 + 0.1 + 0.2 = 0.3
6	P(X = 2 , Y = 4) + P(X = 3 , Y = 3) = 0.2 + 0 = 0.2
7	P(X = 3 , Y = 4) = 0

2- P(Y = X) = summation of probabilities over all values for which x = y.
= P(X=1 , Y=1) + P(X=2 , Y=2) + P(X=3 , Y=3)
= 0.1 + 0 + 0 = 0.1

3- E{XY} $= \sum_{x_i} \sum_{y_j} x_i y_j \, P(X = x_i, Y = y_j)$

= (1)(1) P(X=1 , Y=1) + (1)(3) P(X=1 , Y=3) + (2)(1) P(X=2 , Y=1)
+ (2)(3) P(X=2 , Y=3) + (2)(4) P(X=2 , Y=4) + (3)(2) P(X=3 , Y=2)

= (1)(1) (0.1) + (1)(3) (0.3) + (2)(1) (0.3)
+ (2)(3) (0.1) + (2)(4) (0.2) + (3)(2) (0.2)

E{XY} = 5

Exercise:

Let Z = |X – Y|

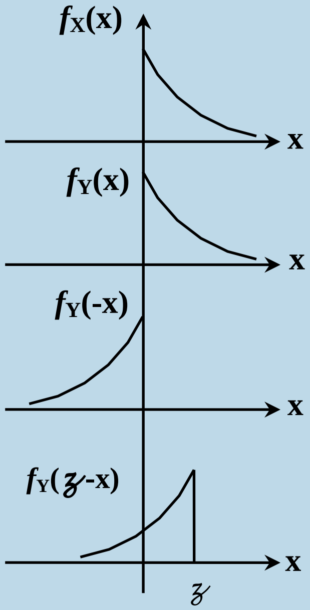
- Find the pmf of Z: P(Z = z)

EXAMPLE (4-7):

Let (X) and (Y) be two independent exponential random variables, such that:

$$f_X(x)=\begin{cases}\alpha\, e^{-\alpha x} , & x>0\\ 0 , & \text{otherwise}\end{cases}\quad ; \quad f_Y(y)=\begin{cases} b\ e^{-by} , & y>0\\ 0 , & \text{otherwise}\end{cases}$$

Let Z = X + Y. Find $f_Z(z)$.



SOLUTION:

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx = \int_0^z a\ e^{-ax}.\ b\ e^{-b(z-x)}\ dx$$

$$f_Z(z)=\int_0^z a\ e^{-ax}.\ b\ e^{-b(z-x)}\ dx = ab\, e^{-bz}\int_0^z e^{-(a-b)x}\ dx,$$

$$f_Z(z)=\frac{ab}{(a-b)}e^{-bz}(1-e^{-(a-b)z})$$

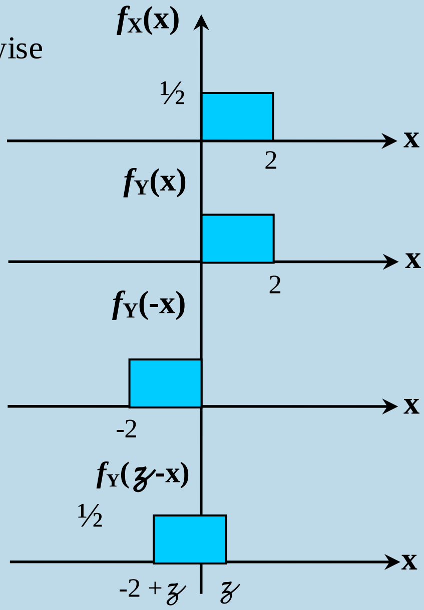
$$f_Z(z)=\frac{ab}{(a-b)}(e^{-bz}-e^{-az})$$

EXAMPLE (4-8):

Let (X) and (Y) be two *identical and independent* random variables, such that:

$$f_X(x)=\begin{cases}\frac{1}{2} , & 0<x<2\\ 0 , & \text{otherwise}\end{cases}\quad ; \quad f_Y(y)=\begin{cases}\frac{1}{2} , & 0<y<2\\ 0 , & \text{otherwise}\end{cases}$$

Let Z = X + Y. Find $f_Z(z)$.



SOLUTION:

Z = X + Y ➔ $0 < z < 4$

- For ($z < 0$) ➔ $f_Z(z) = 0$
- For ($0 < z < 2$)

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)\ f_Y(z-x)\ dx$$

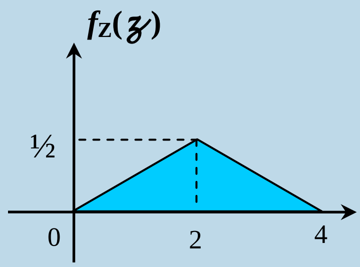
$$f_Z(z)=\int_0^z \frac{1}{2}\times\frac{1}{2}\ dx=\frac{z}{4}$$

- For ($z = 2$)

$$f_Z(z)=\int_0^2 \frac{1}{2}\times\frac{1}{2}\ dx=\left.\frac{x}{4}\right|_0^2=\frac{1}{2}$$

- For ($2 < z < 4$)

$$f_Z(z)=\int_{-2+z}^{2} \frac{1}{2}\times\frac{1}{2}\ dx = \left.\frac{x}{4}\right|_{-2+z}^{2}=\frac{1}{4}[(2-(-2+z)]=\frac{1}{4}[4-z]$$



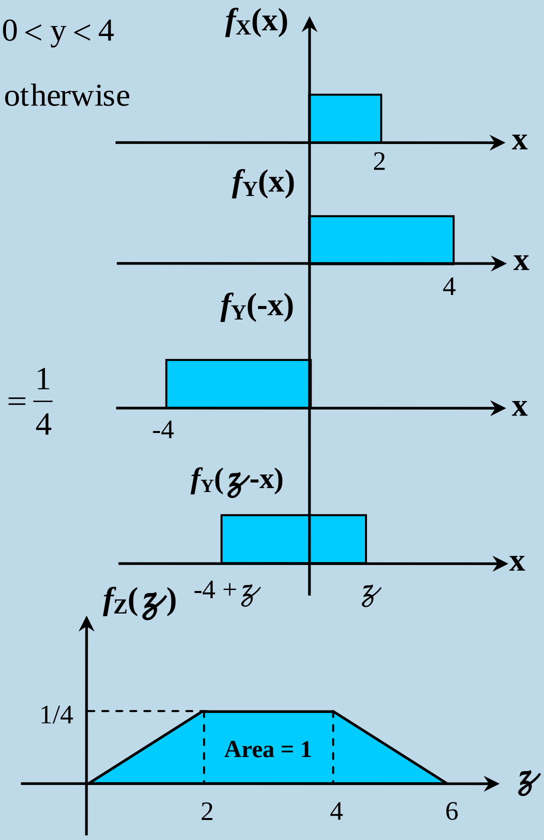
➔ Total area = $\frac{1}{2}\times 4\times\frac{1}{2}=1$ ➔ It is a pdf

EXAMPLE (4-9):

Let (X) and (Y) be two *uniformly distributed and independent* random variables, such that:

$$f_X(x)=\begin{cases}\frac{1}{2} & , \quad 0<x<2\\ 0 & , \quad \text{otherwise}\end{cases} \quad ; \quad f_Y(y)=\begin{cases}\frac{1}{4} & , \quad 0<y<4\\ 0 & , \quad \text{otherwise}\end{cases}$$

Let Z = X + Y. Find $f_Z(z)$.



SOLUTION:

- For $(z<0)$ ➔ $f_Z(z)=0$
- For $(0<z<2)$ ➔ $f_Z(z)=\int_0^z \frac{1}{2}\times\frac{1}{4}\,dx=\frac{1}{8}z$
- For $(2<z<4)$ ➔ $f_Z(z)=\int_0^2 \frac{1}{2}\times\frac{1}{4}\,dx=\frac{1}{8}\times 2=\frac{1}{4}$
- For $(4<z<6)$

$$f_Z(z)=\int_{-4+z}^{2}\frac{1}{2}\times\frac{1}{4}\,dx=\frac{1}{8}x\Big|_{-4+z}^{2}=\frac{1}{8}(6-z)$$

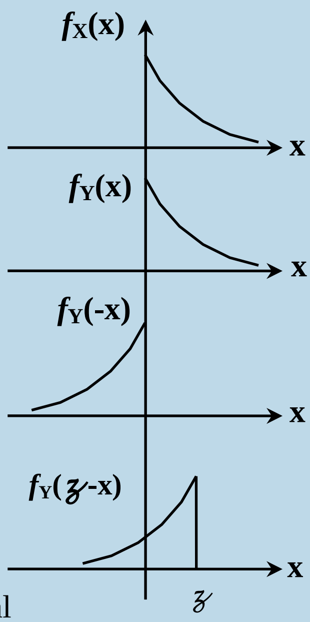
$$f_Z(z)=\begin{cases}\frac{1}{8}z & 0<z<2\\ \frac{1}{4} & 2<z<4\\ \frac{1}{8}(6-z) & 4<z<6\end{cases}$$

EXAMPLE (4-10):

Let (X) and (Y) be two *identical and independent* random variables, such that:

$$f_X(x)=\begin{cases}\alpha\, e^{-\alpha x} & , \quad x>0\\ 0 & , \quad \text{otherwise}\end{cases} \quad ; \quad f_Y(y)=\begin{cases}\alpha\, e^{-\alpha y} & , \quad y>0\\ 0 & , \quad \text{otherwise}\end{cases}$$

Let Z = X + Y. Find $f_Z(z)$.



SOLUTION:

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\,dx=\int_0^z \alpha\, e^{-\alpha x}\,.\,\alpha\, e^{-\alpha(z-x)}\,dx$$

$$f_Z(z)=\int_0^z \alpha^2\, e^{-\alpha x}\,.\,e^{-\alpha z+\alpha x}\,dx$$

$$f_Z(z)=\alpha^2 e^{-\alpha z}\int_0^z dx \;\Rightarrow\; f_Z(z)=\alpha^2 z e^{-\alpha z}$$

Exercise: Find the pdf of $Z = X_1 + X_2 + X_2$ when the variables are independent and identically distributed exponential Random variables having the above given pdf.

Transformation of Multiple Random Variables

Let X_1 and X_2 be two random variables with a joint pdf $f_{x1x2}(x_1, x_2)$ and let $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_1(X_1, X_2)$ be two new random variables , where we assume that g_1 and g_2 have continuous first partial derivatives for all x_1 and x_2.

For a one-to-one mapping, we have

$$P\{ x_1 \le X_1 \le x_1 + \Delta x_1 , x_2 \le X_2 \le x_2 + \Delta x_2 \} = P\{ y_1 \le Y_1 \le y_1 + \Delta y_1 , y_2 \le Y_2 \le y_2 + \Delta y_2 \}$$

$$f_{x1,x2}(x_1, x_2) dx_1\, dx_2 = f_{y1,y2}(y_1, y_2) dy_1\, dy_2$$

$$f_{y1,y2}(y_1, y_2) = \frac{f_{x1,x2}(x_1, x_2)}{|dy_1\, dy_2 / dx_1\, dx_2|} = \frac{f_{x1,x2}(x_1, x_2)}{|J|}$$

The denominator is what is called the Jacobean. It is the ratio of the differential area in the $y_1 - y_2$ plane to the differential area in the $x_1 - x_2$ plane.

$$|J| = \begin{vmatrix} \frac{\partial y_1}{dx_1} & \frac{\partial y_1}{dx_2} \\ \frac{\partial y_2}{dx_1} & \frac{\partial y_2}{dx_2} \end{vmatrix} \quad ; \qquad J \neq 0$$

Therefore, the joint pdf of Y_1 and Y_2 can be determind as

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

Note that :

$$J = \left| \frac{\partial y_1\ \partial y_2}{\partial x_1\ \partial x_2} \right| = \begin{vmatrix} \frac{\partial y_1}{dx_1} & \frac{\partial y_1}{dx_2} \\ \frac{\partial y_2}{dx_1} & \frac{\partial y_2}{dx_2} \end{vmatrix}$$

which can also be re-written in the equivalent form:

$$J = \frac{1}{\left| \frac{\partial x_1\ \partial x_2}{\partial y_1\ \partial y_2} \right|} = \frac{1}{\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}}$$

The marginal pdf's can be found as

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

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

Example 1

Let X_1 and X_2 be two independent exponential R.V with pdf's

$$f_{X1}(x_1) = \begin{cases} \lambda\, e^{-\lambda x_1} & x_1 \geq 0 \\ 0 & x_1 < 0 \end{cases} ;$$

$$f_{X2}(x_2) = \begin{cases} \lambda\, e^{-\lambda x_2} & x_2 \geq 0 \\ 0 & x_2 < 0 \end{cases}$$

Define $Y_1 = X_1 + X_2$

$Y_2 = X_1 / X_2$

a. Find the joint pdf of Y_1 and Y_2.
b. Find the marginal pdf's $f_{y1}(y_1)$ and $f_{y2}(y_2)$.
c. Are Y_1 and Y_2 independent?

Solution

$$f_{X_1X_2}(x_1,x_2) = f_{X_1}(x_1).f_{X_2}(x_2) = \begin{cases} \lambda^2 e^{-\lambda(x_1+x_2)} & ; x_1 \geq 0,\ x_2 \geq 0 \\ 0 & o.w \end{cases}$$

The Jacobean can be calculated as:

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ \frac{1}{x_2} & -\frac{x_1}{x_2^2} \end{vmatrix} = \left|\frac{-x_1}{x_2^2} - \frac{1}{x_2}\right|$$

Therefore , $$f_{y_1,y_2}(y_1, y_2) = \frac{\lambda^2 e^{-\lambda(x_1+x_2)}}{\left|\frac{-x_1}{x_2^2} - \frac{1}{x_2}\right|} \frac{x_2^2}{x_1 + x_2}$$

Solving for x_1 and x_2 in terms of y_1 and y_2, we have

$$x_1 = \frac{y_1 y_2}{1+y_2} \quad , \quad x_2 = \frac{y_1}{1+y_2}$$

$$f_{y_1y_2}(y_1, y_2) = \lambda^2 e^{-\lambda y_1} . \frac{x_2^2}{x_1 + x_2} = \lambda^2 e^{-\lambda y_1} \frac{\frac{{y_1}^2}{(1+y_2)^2}}{y_1}$$

$$= \lambda^2 y_1 e^{-\lambda y_1} \frac{1}{(1+y_2)^2} \qquad \begin{matrix} 0 \leq y_1 < \infty \\ 0 \leq y_2 < \infty \end{matrix}$$

The marginal pdf's are:

$$f_{y_1}(y_1) = \int_0^\infty \lambda^2 y_1 e^{-\lambda y_1} \frac{1}{(1+y_2)^2} dy_2 = \lambda^2 y_1 e^{-\lambda y_1} \underbrace{\int_0^\infty \frac{1}{(1+y_2)^2} dy_2}_{1}$$

$$f_{y_1}(y_1) = \begin{cases} \lambda^2 y_1 e^{-\lambda y_1} & 0 \leq y_1 < \infty \\ 0 & y_1 < 0 \end{cases}$$

$$f_{y_2}(y_2) = \int_0^{\infty} \lambda^2 y_1 e^{-\lambda y_1} \frac{1}{(1+y_2)^2} dy_1 = \frac{1}{(1+y_2)^2} \underbrace{\int_0^{\infty} \lambda^2 y_1 e^{-\lambda y_1} dy_1}_{1}$$

$$f_{y_1}(y_1) = \begin{cases} \frac{1}{(1+y_2)^2} & 0 \le y_2 < \infty \\ 0 & y_2 < 0 \end{cases}$$

Since $f_{y_1 y_2}(y_1, y_2) = f_{y_1}(y_1) \,.\, f_{y_2}(y_2)$, then Y_1 and Y are independent.

Example 2

Let X_1 and X_2 be two independent uniform R.V with a joint pdf

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} 1 & ; 0 \le x_1 \le 1 \,,\, 0 \le x_2 \le 1 \\ 0 & o.w \end{cases}$$

Define $Y_1 = X_1 + X_2$
$Y_2 = X_1 - X_2$

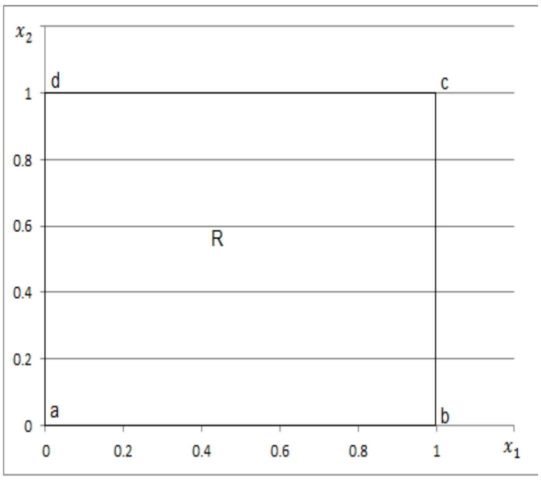
a. Find $f_{y_1 y_2}(y_1, y_2)$ and the region over which it is defined.
b. Find $f_{y_1}(y_1)$ and $f_{y_2}(y_2)$.
c. Are Y_1 and Y_2 independent?

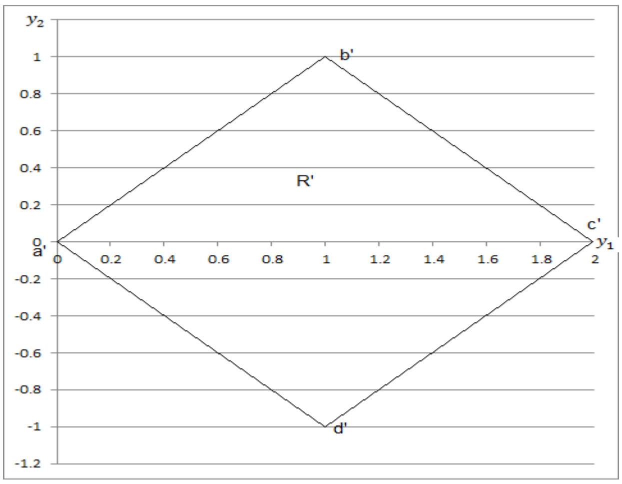
Solution

$$f_{y_1 y_2}(y_1, y_2) = \frac{f_{X_1 X_2}(x_1, x_2)}{|J|}, \qquad J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = |-1 - 1| = 2$$

$$f_{y_1 y_2}(y_1, y_2) = \frac{1}{2} \text{ for } y_1, y_2 \in \grave{R}$$

Where $\grave{R}$ is as shown below





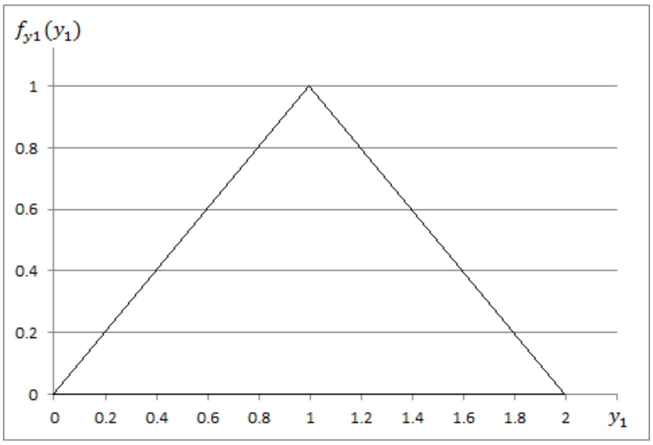
$$f_{y_1}(y_1) = \int_{-y_1}^{y_1} \frac{1}{2}\, dy_2 \qquad 0< y_1 < 1$$

$$= \frac{1}{2}(y_1 - (-y_1)) = y_1$$

$$f_{y_1}(y_1) = \int_{-2+y_1}^{2-y_1} \frac{1}{2}\, dy_2 = \frac{1}{2}[(2 - y_1) - (-2 + y_1)] \qquad 1 < y_1 < 2$$

$$= \frac{1}{2}(4 - 2y_1) = 2 - y_1 \qquad 1<y_1<2$$

$$= \begin{cases} y_1 & 0 \le y_1 \le 1 \\ 2 - y_1 & 1 < y_1 \le 2 \end{cases}$$

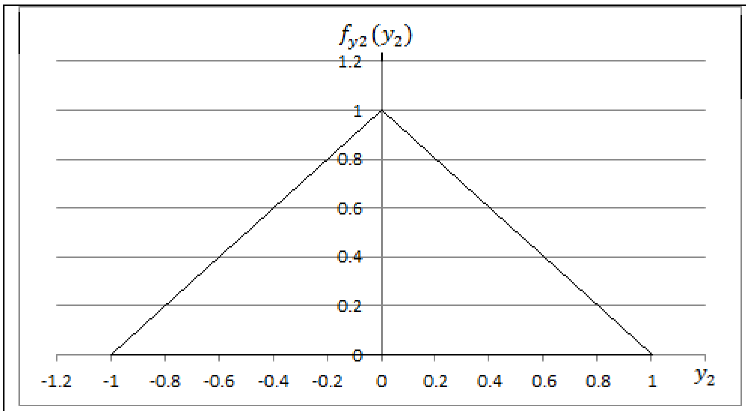


$$f_{y_2}(y_2) = \int_{-y_2}^{2+y_2} \frac{1}{2}\, dy_1 \qquad -1 < Y_2 < 0$$

$$= \frac{1}{2}(2 + y_2 - (-y_2)) = 1+y_2 \qquad -1 < Y_2 < 0$$

$$f_{y_2}(y_2) = \int_{y_2}^{2-y_2} \frac{1}{2}\, dy_1 = \frac{1}{2}[2 - y_2 - y_2] = 1- y_2 \qquad 0 < y_2 < 1$$

$$= \begin{cases} 1 + y_2 & -1 \le y_2 \le 0 \\ 1 - y_2 & 0 < y_2 \le 1 \end{cases}$$



Clearly, since $f_{y_1}(y_1).\, f_{y_2}(y_2) \neq f_{y_1y_2}(y_1, y_2) = \frac{1}{2}$, then Y_1 and Y_2 are not independent

Example 3
Let X_1 and X_2 be two zero –mean, unit variance independent Gaussian random variable. Define the polar random variables R and θ as

$$R = \sqrt{{x_1}^2 + {x_2}^2} \quad , \qquad \theta = \tan^{-1}\frac{x_2}{x_1}$$

Find $f_{R,\theta}(r,\theta)$, $f_R(r)$, $f_\theta(\theta)$

Solution
It can be easily shown that:

$$x_1 = r\cos\theta \qquad\qquad x_2 = r\sin\theta$$

$$\text{The Jacobian } J = \begin{vmatrix} \frac{\partial r}{\partial x_1} & \frac{\partial r}{\partial x_2} \\ \frac{\partial \theta}{\partial x_1} & \frac{\partial \theta}{\partial x_2} \end{vmatrix} = \frac{1}{\begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{vmatrix}}$$

$$J = \frac{1}{\begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}} = \frac{1}{|r(\cos\theta)^2 + r(\sin\theta)^2|} = \frac{1}{r}$$

Therefore,

$$f_{R,\theta}(r,\theta) = \frac{f_{X_1X_2}(x_1, x_2)}{|J|} = r.\frac{1}{\sqrt{2\pi}}e^{-x_1^2/2}.\frac{1}{\sqrt{2\pi}}e^{-x_2^2/2}$$

$$= \frac{r}{2\pi}e^{-(x_1^2+x_2^2)/2}$$

$$= \frac{r}{2\pi}e^{-r^2/2} \qquad 0 \le r < \infty, -\pi \le \theta \le \pi$$

$$f_{R,\theta}(r,\theta) = = \left(\frac{1}{2\pi}\right)\left(re^{-r^2/2}\right)$$

$$f_R(\mathrm{r}) = \int_{-\pi}^{\pi} \frac{1}{2\pi} r e^{-r^2/2}\, d\theta = r e^{-r^2/2}$$

$$f_R(\mathrm{r}) = \begin{cases} r e^{-r^2/2} & 0 \le r < \infty \\ 0 & o.w \end{cases}$$

$$f_\theta(\theta) = \begin{cases} \frac{1}{2\pi} & -\pi \le r < \pi \\ 0 & o.w \end{cases}$$

CHAPTER IV

ELEMENTARY STATISTICS

Basic Definitions and Terminology

- In statistics, we take a *random sample* (X_1, X_2, ……, X_n) of size (n) from a distribution X (population) for which the pdf is $f_X(x)$ by performing that experiment (n) times. The purpose is to draw conclusions from properties of the *sample* about properties of the distribution of the corresponding X (*the population*)

- First let us introduce some basic definitions about the random sample
- The *sample mean* $\hat{\mu}_X$ is defined as :

$$\hat{\mu}_X \ (\text{or } \bar{X}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- The *sample variance* $\hat{\sigma}_X^2$ is defined as:

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

$\hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2}$ is called the *sample standard deviation*.

- A computationally simpler expression for $\hat{\sigma}_X^2$ is:

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

▪ Regression Techniques:

Suppose in a certain experiment we take measurements in pairs, i.e. (x_1,y_1) , (x_2,y_2), … (x_n,y_n). We suspect that the data can be fit in 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

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

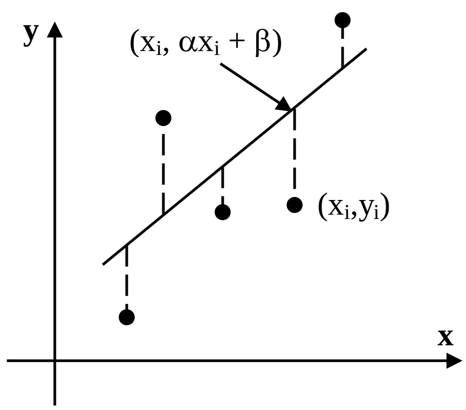
The method of least squares specifies the values of α and β so that $\in$ is minimized.

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

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

$$n\beta + \alpha\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad \text{……… (1)}$$

$$\beta\sum_{i=1}^{n} x_i + \alpha\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad \text{……… (2)}$$



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:

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

where, $C_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y) = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)}$ is the sample covariance between x and y and $\hat{\sigma}_X^2$ is the sample variance of the x measurements (as defined earlier).

$$\beta = \hat{\mu}_Y - \alpha\hat{\mu}_X$$

where, $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\mu}_Y = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the mean of the x and the y measurements respectively.

A useful formula for α may also be taken as:

$$\alpha = \frac{\sum_{i=1}^{n} x_i y_i - n\,\hat{\mu}_X\,\hat{\mu}_Y}{\sum_{i=1}^{n} x_i^2 - n\,\hat{\mu}_X^2}$$ *(Curve passes through $\hat{\mu}_X$ and $\hat{\mu}_Y$.)*

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

- **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 .

$$\beta_1 n + \beta_2\sum_{i=1}^{n} x_i + \beta_3\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \qquad \ldots\ldots\ldots (1)$$

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

$$\beta_1\sum_{i=1}^{n} x_1^2 + \beta_2\sum_{i=1}^{n} x_i^3 + \beta_3\sum_{i=1}^{n} x_1^4 = \sum_{i=1}^{n} x_i^2 y_i \qquad \ldots\ldots\ldots (3)$$

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:

$y = a\,e^{bx}$ ……… (1)

Taking the natural logarithm:

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

$\ln(y) = \ln(a) + b\,x$ ……… (2)

Let $Y' = \ln(y)$; $\beta' = \ln(a)$; $\alpha' = b$

So, equation (2) now becomes $Y' = \beta' + \alpha' x$

Which is the case of the straight line treated first. For each y_i take its natural logarithm to get Y'_i.
The new pairs of the data are (x_1, lny_1), (x_2, lny_2), … (x_n, lny_n), the solution of which is known.

EXAMPLE (5-1):
Suppose that the polynomial to be fitted to a set of (n) points is y = b x. It can be shown that:

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

EXAMPLE (5-2):
Let $y = a\,x^b$.
Taking the ln of both sides, then:
ln y = ln a + b ln x
$y' = \beta' + \alpha' x'$ *(Linear regression)*
where: $y' = \ln y$, $\beta' = \ln a$, $\alpha' = b$, $x' = \ln x$

EXAMPLE (5-3):

If $y = 1 - e^{\frac{-x^b}{a}}$

Manipulation of this equation yields:

$$\ln\ln\left(\frac{1}{1-y}\right) = -\ln a + b\ln x$$

which is the standard form:
$y' = \beta' + \alpha' x'$ *(Linear regression)*

EXAMPLE (5-4):

If $y = \dfrac{L}{1 + e^{a+bx}}$

This form reduces to:

$$\ln\left(\frac{L-y}{y}\right) = a + b\,x$$

which is in the standard form:
$y' = \beta' + \alpha' x'$ *(Linear regression)*

EXAMPLE (5-5):

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

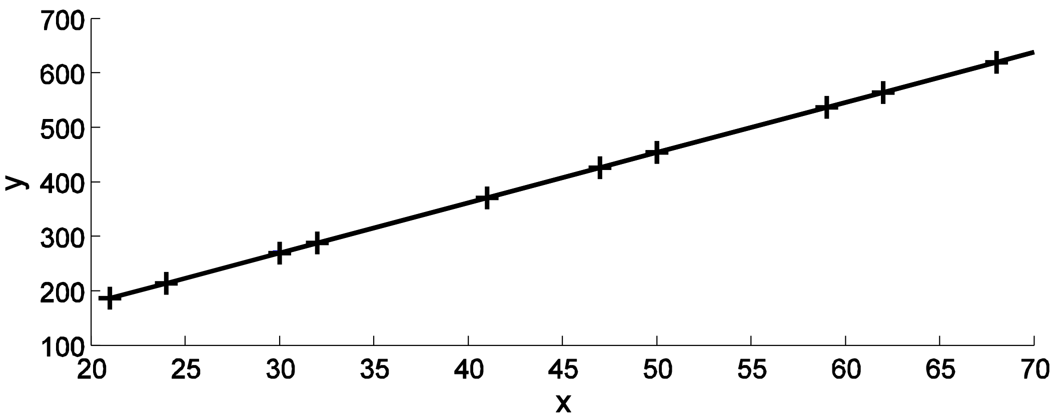
a. Assuming that a simple linear regression model is appropriate, fit the regression model relating steam usage (y) to the average temperature (x).
b. What is the estimate of expected usage when the average temperature is 55 F˚?
c. Find the correlation coefficient between x and y.

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

Here, $\sum_{i=1}^{12} x_i = 558, \sum_{i=1}^{12} x_i^2 = 29256, \sum_{i=1}^{12} x_i y_i = 265607, \sum_{i=1}^{12} y_i = 5060$

The equation parameters are given by: α = 9.2182, β = - 7.3126. The minimum value of the mean square error is MMSE = 38.1315. VERIFY THESE RESULTS.
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 = 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, the correlation coefficient =1).

Now let us try to fit the data in a polynomial $y = \beta_1 + \beta_2 x + \beta_3 x^2$. The equation parameters are: $\beta_1 = -5.0455$, $\beta_2 = 9.1068$, $\beta_3 = 0.0012$. The MMSE = 37.0561.
Note that the second order curve fitting has too little effect on the mean square error, which essentially implies that the linear model is quite satisfactory.
The linear model and the measured data are shown in the figure below



Sums of Random Variables and the Central Limit Theorem

Theorem:

Let $X_1, X_2, \ldots, X_n$ be a sequence of Gaussian random variables, then any linear combination of them is Gaussian, i.e., if

$$Y = C_1X_1 + C_2X_2 + \ldots\ldots + C_nX_n$$

Then the pdf of y is: $f_Y(y) = \dfrac{1}{\sqrt{2\pi\sigma_Y^2}} e^{\frac{-(y-\mu_y)^2}{2\sigma_X^2}}$

where $\mu_Y = C_1\mu_1 + C_2\mu_2 + \ldots\ldots\ldots + C_n\mu_n$

And $\sigma_Y^2 = C_1^2\sigma_1^2 + C_2^2\sigma_2^2 + \ldots + C_n^2\sigma_n^2 + 2C_1C_2\sigma_1\sigma_2\rho_{1,2} + 2C_1C_3\sigma_1\sigma_3\rho_{1,3} + 2C_2C_3\sigma_2\sigma_3\rho_{2,3} + \ldots$

The following example illustrates this theorem for the case when the random variables are dependent Gaussian random variables.

EXAMPLE (5-6):

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

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

SOLUTION:

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

$\sigma_Y^2 = 4\sigma_1^2 + 9\sigma_2^2 + 2(2)(3)(\sigma_1)(\sigma_2)\rho_{1,2} = 4(4) + 9(9) + 2(2)(3)(2)(3)(0.25) = 115$

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

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

$$Y = C_1X_1 + C_2X_2 + \ldots\ldots + C_nX_n$$

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

$$\mu_Y = C_1\mu_1 + C_2\mu_2 + \ldots\ldots\ldots + C_n\mu_n$$

$$\sigma_Y^2 = C_1^2\sigma_1^2 + C_2^2\sigma_2^2 + \ldots + C_n^2\sigma_n^2$$

EXAMPLE (5-7):

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

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

SOLUTION:

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

$\sigma_Y^2 = 4\sigma_1^2 + 9\sigma_2^2 = = 4(4) + 9(9) = 97$

$P(Y < 35) = \Phi\left(\dfrac{35-30}{\sqrt{97}}\right) = \Phi(0.5077) = 0.6942.$

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. Define

$$Y = (X_1 + X_2 + \ldots\ldots + X_n)/n$$

Then Y has a Gaussian distribution with mean and variance given by: $\mu_Y = \mu$, $\sigma_Y^2 = \sigma^2/n$. Y is called the sample mean and will be denoted by $\hat{\mu}$ or $\overline{X}$.

EXAMPLE (5-8):

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:

$\hat{\mu} = (X_1 + X_2 + \ldots X_n)/n$

$E\{\hat{\mu}\} = (\mu + \mu + \ldots \mu)/n = \mu = 330$

$Var(\hat{\mu}) = \sigma^2/n = (1.5)^2/10 = 0.225$

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

$$P(\hat{\mu} < 328) = \Phi\left(\frac{328-330}{\sqrt{0.225}}\right) = \Phi(-4.21) = 1.2769\text{e-}005.$$

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:

$$\hat{\mu} = (X_1 + X_2 + \ldots\ldots + X_n)/n$$

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. If n < 30, the central limit theorem will work well if the distribution of the population is not severely non-normal.

  The theorem works well for small samples n = 4 , 5 when the population has a continuous distribution as illustrated in the following example.

EXAMPLE (5-9):

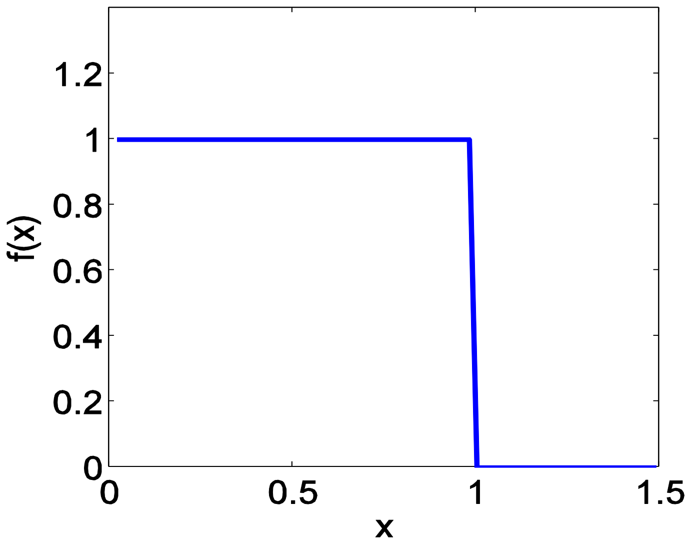
Let $Y = X_1 + X_2 + X_3$, where Xi are uniform over the interval ($0 \leq Xi \leq 1$) and are independent. Find and sketch the pdf of Y.

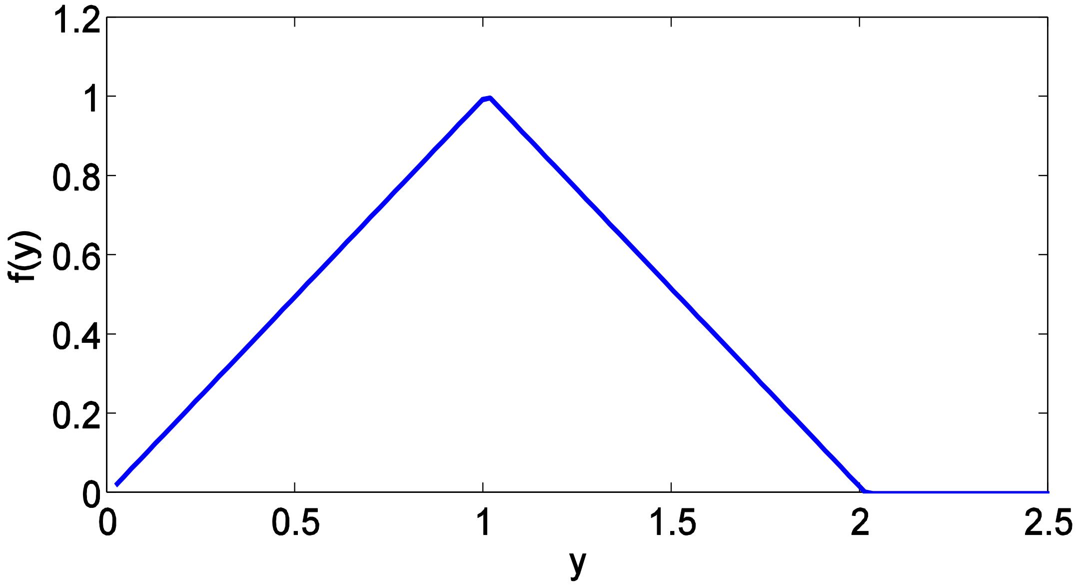
SOLUTION:

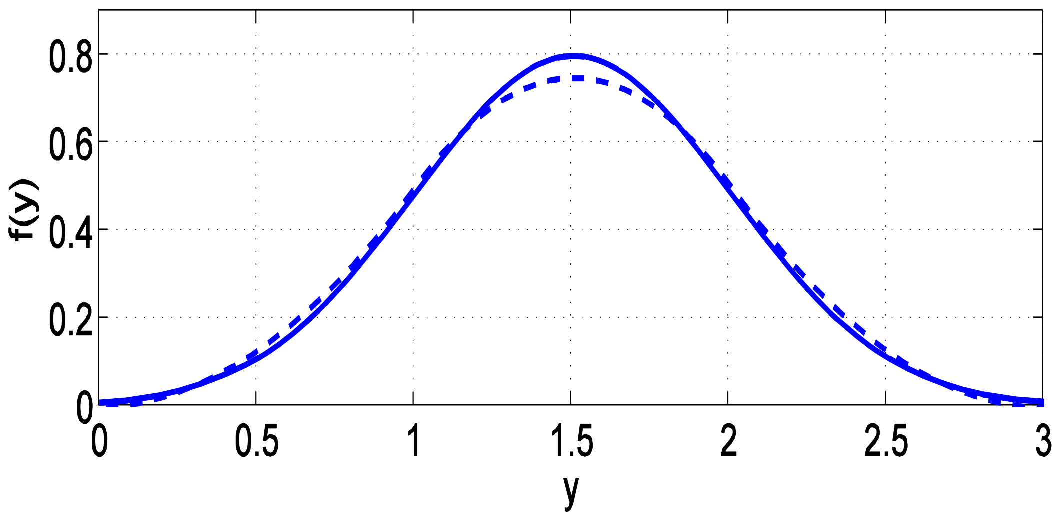
First we find the pdf of ($X_1 + X_2$) by convolving the pdf of X_1 with that for X_2. Then the new pdf is convolved with that for X_3. The result is:

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

The mean and variance of Y are: $\mu_Y=3(1/2)=3/2$, $\sigma_Y^2=3(\sigma_x)^2=3(1/12)=3/12$. In the figure below we plot the pdf's of X and (X_1+X_2). Also, we plot the Gaussian pdf (Solid line) with mean 3/2 and variance 3/12 on the same graph of the pdf for $Y=X_1+X_2+X_3$ (dashed line)







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.

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

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

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

EXAMPLE (5-10):

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:

$\hat{\mu}_X$ is approximately normal with:

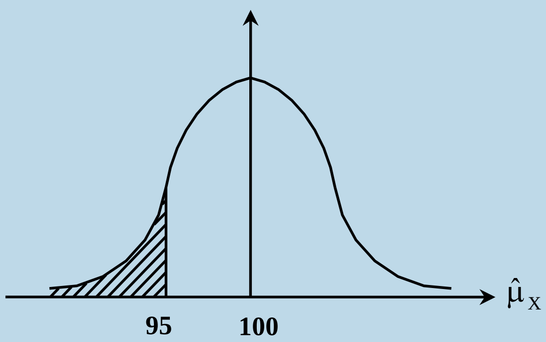
mean = $E(\hat{\mu}_X) = 100$ Ω.

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

$$\hat{\sigma}_X = \sqrt{\frac{\sigma_X^2}{n}} = \sqrt{\frac{10^2}{25}} = 2$$

$$P\{\hat{\mu}_X < 95\} = P\{Z < \frac{95-100}{2}\}$$

$$= \Phi(-2.5) = 0.00621$$



EXAMPLE (5-11):

The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, then it is immediately replaced by a new one. Assume we have 25 such batteries, 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:

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

$$\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{25}^2 = 25\sigma_X^2 = 25*(20)^2 = 10000$$

$$P(Y > 1100) = P(Z > \frac{1100-1000}{\sqrt{10000}}) = P(Z > 1) = 1 - \Phi(1) = 0.158655$$

EXAMPLE (5-12):

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 $\hat{\mu}_X$

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

SOLUTION:

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

$E(\hat{\mu}_X) = E(X) = 1/2.$

$$Var(\hat{\mu}_X) = \hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{1/12}{30} = \frac{1}{360}$$

$$P\{\hat{\mu}_X < 0.52\} = P\{Z < \frac{0.52-0.5}{\sqrt{1/360}} = 0.379\}$$

$$= \Phi(0.379) = 0.648027$$

EXAMPLE (5-13):

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 $\hat{\mu}_X$

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

SOLUTION:

Since n=50 > 30, then we can approximate the sample mean by a normal distribution with:

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

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

$$P\{\hat{\mu}_X < 0.6\} = P\{Z < \frac{0.6-0.5}{\sqrt{1/200}}\} = \Phi(1.414) = 0.92073$$

Estimation of Parameters:

- The field of *statistical inference* consists of those method used to make decisions or to draw conclusions about a population. These methods utilize the information contained in a sample from a population in drawing conclusions.

- 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.

Basic Definitions and Terminology

- In statistics, we take a random sample (x_1, x_2, ……, x_n) of size (n) from a distribution X (population) for which the pdf is $f_X(x)$ by performing that experiment (n) times. The purpose is to draw conclusions from properties of *sample* about properties of the distribution of the corresponding X (*the population*). We do this by calculating *point estimators* or *confidence intervals* or by performing a *test of parameters* or by a *test for distribution functions*.
- For populations we define numbers called *parameters* that characterize important properties of the distributions (μ_X and σ_X^2 in normal distribution, (p) in binomial distribution, λ for the exponential distribution, the rate of arrival in the Poisson distribution, the end points a and b in the uniform distribution). Here, the pdf is explicitly expressed in terms of the parameter as $f_X(x;\theta)$. The unknown parameter (θ) is estimated by some appropriate function of the observations $\hat{\theta} = f(x_1, x_2, ..., x_n)$

  The function $\hat{\theta} = f(x_1, x_2, ..., x_n)$ is called *statistics* 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*

- The random variables X_1, X_2, …, X_n , called a *random sample*, have the same distribution and are assumed to be independent.
- *Estimator*: is a function of the observable sample data that is used to estimate an unknown population parameter.

- We consider two types of estimators, *point estimators* and *interval estimators*.

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 population parameter (θ) is a single numerical value $\hat{\theta} = f(x_1, x_2, ..., x_n)$.

- In the table below we list some examples of point estimators and the parameters that are used to estimate.

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:**

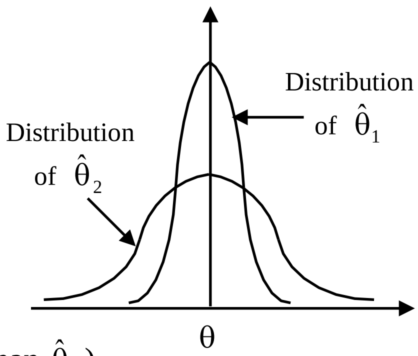
1- An estimator should be close to the true value of the unknown parameter.

- ***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 $(\hat{\theta})$.

2- Let $\hat{\theta}_1, \hat{\theta}_2$ be *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 ($Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, $\hat{\theta}_1$ is called more *efficient* than $\hat{\theta}_2$)

The variance $Var(\hat{\theta}) = E\{[\hat{\theta} - E(\theta)]^2\}$ is a measure of the imprecision of the estimator.

3- 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 (5-14):

Show that the sample mean $\hat{\mu}_X$

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

is an unbiased estimator of the population mean μ_X

SOLUTION:

$$E\{\hat{\mu}_X\} = \frac{1}{n}E\left\{\sum_{i=1}^{n} x_i\right\} = \frac{1}{n}\left\{\sum_{i=1}^{n} E\{x_i\}\right\} = \frac{1}{n}\left\{\sum_{i=1}^{n}\mu_x\right\} = \frac{1}{n}(n\mu_x) = \mu_x \text{ (unbiased estimator)}$$

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

$$Var\{\hat{\mu}_X\} = \frac{1}{n^2}Var\left\{\sum_{i=1}^{n} x_i\right\} = Var\{\hat{\mu}_X\} = \frac{\sigma_X^2}{n}$$ (The variance tends to zero as n tends to infinity.

Therefore, $\hat{\mu}_X$ is unbiased and consistent estimator). When n goes to infinity, $\frac{1}{n}\left\{\sum_{i=1}^{n} x_i\right\} \to \mu_X$

EXAMPLE (5-15):

Show that the sample variance $\hat{\sigma}_X^2$ (*when the mean is unknown*).

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

is an unbiased estimator of the population variance σ_X^2

SOLUTION: A computationally simpler expression for the sample variance is

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

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

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

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

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

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

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

$$\mathrm{E}\{\hat{\sigma}_X^2\} = \sigma_X^2$$

EXAMPLE (5-16):

Let X1 and X2 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:

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

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

Now, we evaluate the variance of the two estimators:

From previous results we know:

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

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

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

CHAPTER V

ESTIMATION THEORY AND APPLICATIONS

Method for Obtaining Point Estimators: *The Maximum Likelihood (ML) Estimator*

Let us start this section with two motivating examples:

EXAMPLE (6-1):

The probability p = P(H) of a coin may be 0.1 or it may be 0.9. To resolve the uncertainty, the coin was tossed 10 times and 3 heads 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

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

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

Therefore, we conclude that p=0.1 has a higher probability of producing the outcome and our estimate for p would be $\hat{p}=0.1$.
Instead, suppose that the experiment resulted in 8 heads, what would be our estimate for p?. Again, we calculate

$P(x=8;0.1)$ =3.645 e-7

$P(x=8;0.9)$ =0.1937

In this case $\hat{p}=0.9$.

EXAMPLE (6-2):

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 were 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 value of p that maximizes f(p), we differentiate f(p) with respect to p, set the derivative to zero, and solve for 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 $\hat{p}=3/10$.

For the sake of comparison, let us compute f(p) at three different values of p; p=0.3, p=0.35, and p=0.25. f(0.3) = 0.2668, f(0.35) = 0.2503, f(0.25) = 0.2522. Hence $\hat{p}=3/10$ has the highest probability of generating the 3 successes in 10 trials.

To find the maximum likelihood estimator of a parameter θ, we base our estimation on (n) statistically independent samples $\{X_1, X_2, \ldots, X_n\}$ taken from the population. The maximum likelihood estimator selects the parameter value which gives the observed data the largest possible probability. The following steps summarize the procedure for obtaining a maximum likelihood estimator for a continuous parameter θ.

- The joint pdf of the samples is (*expressed in terms of θ*)
  $L(\theta) = f\{x_1, x_2, \ldots, x_n ; \theta\} = f\{x_1 ; \theta\} \,.\, f\{x_2 ; \theta\} \ldots\ldots f\{x_n ; \theta\}$ ➔ *due to independent of x_i.*

  $L(\theta)$ is called the *likelihood function*. The maximum likelihood technique looks for that value $(\hat{\theta})$ of the parameter that maximizes the joint pdf of the samples.

- A necessary condition for the maximum likelihood estimator of (θ) is:
  $\frac{d}{d\theta} L(\theta) = 0$ or equivalently $\frac{d}{d\theta} \ln\{L(\theta)\} = 0$
  (The ln(*) is a monotonically increasing function of the argument)
  The following example illustrates this technique.

<u>EXAMPLE (6-3):</u>

Given a random sample 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 assumed known.

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

c- The mean μ_X and variance σ_X^2 when both are assumed unknown.

<u>SOLUTION:</u>

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

a- Set $\frac{d}{d\mu_X} \ln L(\mu) = 0$ ➔ *treating σ_X^2 as a constant.*

$$\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 = 0 \;➔\; \hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \ldots\ldots\ldots (1) \quad \textit{Unbiased Estimator}$$

Thus the ML estimator of the mean is the sample average mentioned earlier.

b- Set $\frac{d}{d\sigma_X^2} \ln L(\sigma_X^2) = 0$ ➔ *treating μ_X as a constant*

The result is $\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2$ ……… (2) *Unbiased Estimator*

Note that: the division is by (n) since we are using the known mean of the distribution

c- Set $\frac{\partial}{\partial\mu_X} \ln L(\mu_X\,,\, \sigma_X^2) = 0$ and $\frac{\partial}{\partial\sigma_X^2} \ln L(\mu_X\,,\, \sigma_X^2) = 0$

This results in: $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$

$\hat{\sigma}_X^2$ is a biased estimator since $E\{\hat{\sigma}_X^2\} = \frac{(n-1)\,\sigma_X^2}{n}$

For this general case, the unbiased estimator of σ_X^2 is: $\hat{\sigma}_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2$

Which is the sample variance introduced earlier.

EXAMPLE (6-4):

Given a random sample 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 estimators for α.

SOLUTION:

The likelihood function is

$$L(\alpha) = f(x_1)f(x_2)\dots f(x_n)$$
$$L(\alpha) = (\alpha+1)x_1^{\alpha}\dots(\alpha+1)x_n^{\alpha} = (\alpha+1)^n x_1^{\alpha}\dots x_n^{\alpha}$$
$$\ln L(\alpha) = n\ln(\alpha+1) + \alpha\ln x_1 \dots + \alpha \ln x_n$$

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

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

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

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

Finding Interval Estimators for the Mean and Variance:

- An interval estimate of an unknown parameter of (θ) is an interval of the form $\theta_1 \le \theta \le \theta_2$ where the end points θ_1 and θ_2 depend on the numerical value of the parameter to be estimated $(\hat{\theta})$ 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:

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

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

I. Confidence Interval on the Mean: (*Variance Known*)

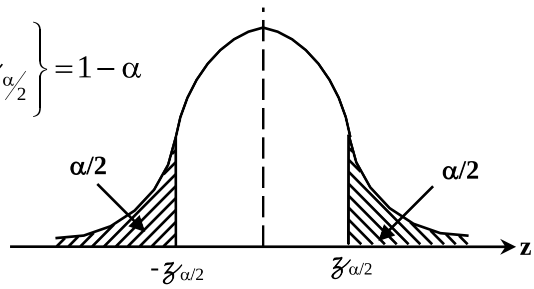
- Suppose that the population of interest has a Gaussian distribution with unknown mean μ_X and known variance σ_X^2.

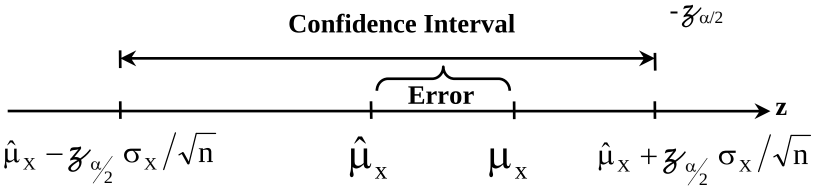
The sampling distribution of $\hat{\mu}_X = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ is Gaussian with mean μ_X and variance $\dfrac{\sigma_X^2}{n}$.

Therefore, the distribution of the statistic $Z = \dfrac{\hat{\mu}_X - \mu_X}{\sigma_X/\sqrt{n}}$ is a *standard normal distribution*.

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

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





- ***Definition:***

If $\hat{\mu}_X$ is the sample mean of a random sample of size (n) from a population with known variance σ_X^2, a $100(1-\alpha)\%$ confidence interval on μ_X is given by:

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

where $z_{\alpha/2}$ is the upper $100(\alpha/2)\%$ 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 $E = z_{\alpha/2}\,\sigma_X/\sqrt{n} \;\Rightarrow\; n = \left(\dfrac{z_{\alpha/2}\,\sigma_X}{E}\right)^2$.

EXAMPLE (6-5):

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

Find the confidence limits for a 95% confidence level if the variance of the population is 4.

SOLUTION:

From the sample we have:
n = 10

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94$$

$$z_{\alpha/2} = 1.96$$

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

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

$$P\{8.70 \le \mu_X \le 11.1796\} = 0.95$$

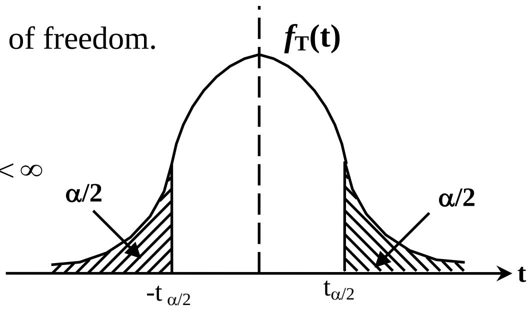
II. Confidence Interval on the Mean: (Variance Unknown)

- Suppose that the population of interest has a normal distribution with unknown mean μ_X and unknown variance σ_X^2.

- ***Definition:***

Let X_1, X_2, ……, X_n be a random sample for 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 (n – 1) degree of freedom.



The t-pfd is: $f_T(t) = \dfrac{\Gamma\left(\frac{(k+1)}{2}\right)}{\sqrt{\pi k}\;\Gamma\left(\frac{k}{2}\right)} \dfrac{1}{\left(\frac{t^2}{k+1}\right)^{\frac{k+1}{2}}}$ $\quad -\infty < t < \infty$

(k) is the number of degrees of freedom.

The mean of the t-distribution is zero and the variance $\dfrac{k}{k-2}$.

The t-distribution is symmetrically and unimodal, the maximum is reached when the mean is 0 (quite similar to normal distribution. As $k \to \infty$, the t-distribution is the normal distribution).

$$P\left\{-t_{\alpha/2,\,n-1} \le T \le t_{\alpha/2,\,n-1}\right\} = 1-\alpha$$

$T = \dfrac{\hat{\mu}_X - \mu}{\hat{\sigma}_X / \sqrt{n}}$ is the t-distribution with (n – 1) degree of freedom

$t_{\alpha/2,\,n-1}$ is the upper 100(α/2)% point of the t- distribution with (n – 1) degree of freedom

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

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

- ***Definition:***

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 – α)% 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(α/2)% point of the t-distribution with (n–1) degrees of freedom.

EXAMPLE (6-6):

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:

From the sample we have:

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94$$

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

From tables of t-distribution:

Number of degrees of freedom = n – 1 = 10 – 1 = 9 = v

$$\alpha = 0.05 \rightarrow \alpha/2 = 0.025 \rightarrow t_{\alpha/2,9} = 2.263$$

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

$$P\left\{9.94 - 2.263\frac{\sqrt{6.51}}{\sqrt{10}} \le \mu_X \le 9.94 + 2.263\frac{\sqrt{6.51}}{\sqrt{10}}\right\} = 0.95$$

$$P\{8.11 \le \mu_X \le 11.77\} = 0.95$$

III. Confidence Interval on the Variance of a Normal Population: (Mean Known)

- When the population is normal, the sampling distribution of:

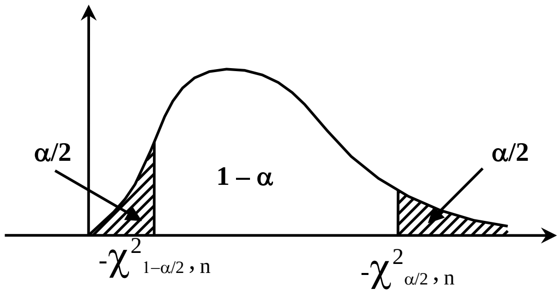
$$\chi^2 = \frac{n\,\hat{\sigma}_X^2}{\sigma_X^2} = \sum_{i=1}^{n}\left(\frac{x_i - \mu_X}{\sigma_X}\right)^2 \quad ; \quad \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2$$

is chi-square with (n) degrees of freedom.
The confidence interval is developed as:

$$P\left\{\chi^2_{1-\alpha/2,n} \le \chi^2 \le \chi^2_{\alpha/2,n}\right\} = 1-\alpha$$

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

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



- ***Definition:***

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

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

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

<u>EXAMPLE (6-7):</u>

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.

<u>SOLUTION:</u>

From the sample we have:

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94 \quad \text{and} \quad \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2 = 5.866$$

From tables of χ^2-distribution:

Number of degree of freedom = n = 10 = ν

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

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

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

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

- When the population is normal, the sampling distribution of:

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

is chi-square with (n – 1) degrees of freedom.

- ***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 – α)% confidence interval on σ_X^2 is:

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

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

<u>EXAMPLE (6-8):</u>

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.

<u>SOLUTION:</u>

From the sample we have:

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 9.94$$

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

From tables of χ^2-distribution:

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

$$\alpha = 0.05 \rightarrow \alpha/2 = 0.025$$

$$\chi^2_{0.025,9} = 19.023 \text{ and } \chi^2_{0.975,9} = 2.7$$

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

$$P\left\{\frac{9\times 6.51}{19.023} \le \sigma_X^2 \le \frac{9\times 6.51}{2.7}\right\} = 0.95$$

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

V. Confidence Interval on a Binomial Proportion:

- Suppose that a random sample of size (n) has been taken from a large population and that X ; (x ≤ n) observations in this sample belong to a class of interest. Then $\hat{P} = x/n$ is a point estimator of the proportion of the population (p) that belongs to this class. Here (n) and (p) are the parameters of a binomial distribution.
  (X) is binomial with mean (np) and variance np(1 – p). Therefore,
  $\hat{P} = x/n$ has a mean (p) and variance $\dfrac{n\,p\,(1-p)}{n^2} = \dfrac{p\,(1-p)}{n}$

- As was mentioned earlier (limiting case of the binomial distribution to the normal distribution) the sampling distribution $\hat{P}$ is approximately normal with mean (p) and variance $\dfrac{p\,(1-p)}{n}$.
  (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:

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

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

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

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

- The solution is to replace (p) by $\hat{P}$ in $\dfrac{p\,(1-p)}{n}$ so that:

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

<u>EXAMPLE (6-9):</u>

In a random sample of 85 automobile engine crankshafts bearings, 10 have a surface finish that is rougher than the specifications allow. A 95% confidence interval for (p) is:

$z_{\alpha/2} = z_{0.025} = 1.96$ and $\hat{P} = \dfrac{x}{n} = \dfrac{10}{85} = 0.12$

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

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

$$P\{0.05 \le p \le 0.19\} = 0.95$$

CHAPTER VI

ENGINEERING DECISION

- **Hypothesis Testing:**

In the last chapter we illustrated how a parameter can be estimated (points or interval estimation) from sample data. However, many problems require that we decide whether to accept or reject a statement about some parameter. The statement is called a *hypothesis*, and the decision-making procedure is called *Hypothesis Testing.*

Two types of error are possible in such a decision process:

1- We decide that the null hypothesis H_0 is false when it is really correct

This is called a *type I error* and its probability is denoted by α

α is called the significance level or size of the test.

2- We decide that the null hypothesis H0 is correct when it is really false

This is called a *type II error* and its probability is denoted by β

- ***Definition:***

The power of the test $(1 - \beta)$ is the probability of accepting the alternative hypothesis when the alternative hypothesis is true.

- **One-Sided and Two-Sided Hypothesis:**

A test of hypothesis such as:

H_0 : $\theta = \theta_0$

H_1 : $\theta \neq \theta_0$

Is called a two-sided test.

Reject H_0 Accept H_1	Accept H_0 $\theta = \theta_0$	Reject H_0 Accept H_1	
$\theta \neq \theta_0$ Rejection Region	Acceptance region	$\theta \neq \theta_0$ Rejection Region	$\hat{\theta}$

H_0 : is known as the null hypothesis.

H_1 : is known as the alternative hypothesis.

Tests such as:

H_0 : $\theta = \theta_0$

H_1 : $\theta > \theta_0$

Accept H_0	Reject H_0 Accept H_1
$\theta = \theta_0$	$\theta > \theta_0$

H_0 : $\theta = \theta_0$

H_1 : $\theta < \theta_0$

Reject H_0 Accept H_1	Accept H_0
$\theta > \theta_0$	$\theta = \theta_0$

Are called one-sided tests.

- α is called the significance level or size of the test.
- The power of the test plotted against the true parameter value is called the Operating Characteristic (OC) curve.

- **Hypothesis Testing on the Mean:** ***Variance Known***

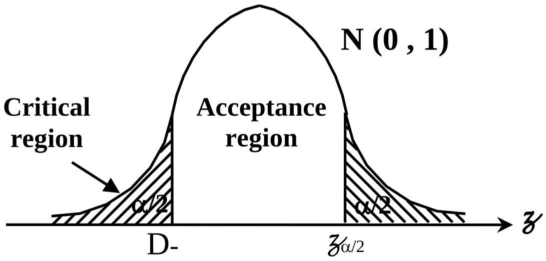
- Suppose that we wish to test the hypothesis:

  $H_0: \quad \mu = \mu_0$
  $H_1: \quad \mu \neq \mu_0$

  Where μ_0 is a specified constant. We have a random sample $X_1, X_2, \ldots\ldots, X_n$ from the population (assumed normal). $\hat{\mu}$ is normal with mean μ_0 and variance σ^2/n when H_0 is assumed true.

- We use the test statistic: $Z = \dfrac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}$; Z is N(0 , 1) when H_0 is assumed true.

- If the level of significance is (α), then the probability is $(1 - \alpha)$ that the test statistic (Z) falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$.



- Reject H_0 if $z > -z_{\alpha/2}$ or $z < z_{\alpha/2}$
  Fail to reject H_0 if $-z_{\alpha/2} < z < z_{\alpha/2}$

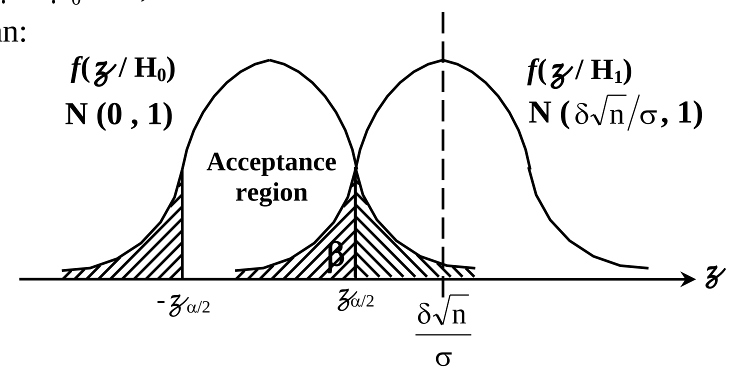
- In the terms of $\hat{\mu}$, we reject H_0 if: $\hat{\mu} > \mu_0 + z_{\alpha/2}\,\sigma/\sqrt{n}$ or $\hat{\mu} < \mu_0 - z_{\alpha/2}\,\sigma/\sqrt{n}$

- Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, when H_1 is true, Z is normal with mean:

  $$E(Z/H_1) = \frac{E(\hat{\mu}/H_1) - \mu_0}{\sigma/\sqrt{n}}$$

  $$E(Z/H_1) = \frac{\mu_0 + \delta - \mu_0}{\sigma/\sqrt{n}} = \frac{\delta}{\sigma/\sqrt{n}}$$

  and unit variance. $Z : N\left(\dfrac{\delta\sqrt{n}}{\sigma}, 1\right)$



  The probability of type II error is the probability that (Z) will fall between $-z_{\alpha/2}$ and $z_{\alpha/2}$. This probability is:

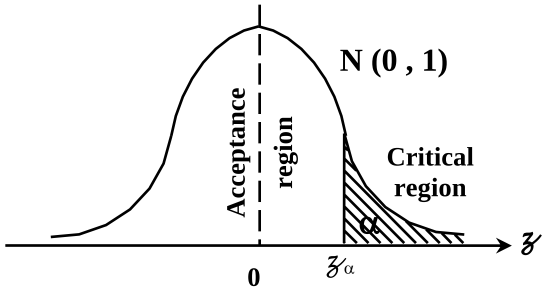
  $$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

  Now if we want to test $\quad H_0: \quad \mu = \mu_0$
  $\qquad H_1: \quad \mu > \mu_0$

  $Z = \dfrac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}$, Z is N(0 , 1) when H_0 is assumed true.

  If (α) is the level of significance, then

  H_0 is rejected if $z > z_\alpha$ and accepted if $z < z_\alpha$

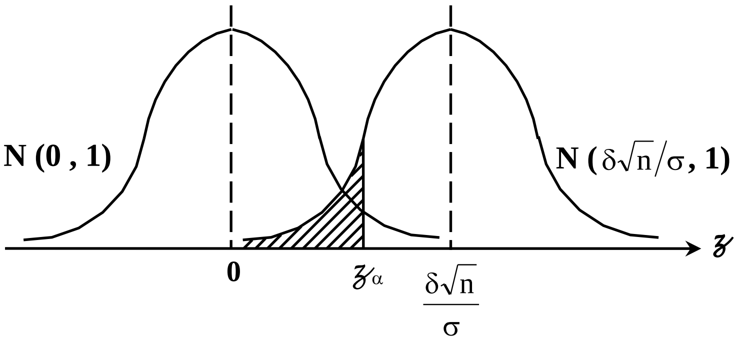


If H_1 is true, that is $\mu = \mu_0 + \delta$, $\delta > 0$, then

The type II error is the probability that z falls between $-\infty$ and z_α.

This probability is:

$$\beta = \Phi\left(z_\alpha - \frac{\delta\sqrt{n}}{\sigma} \right)$$



EXAMPLE (7-1):

Aircrew space systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 cm/s. We know that the standard deviation of burning rate is 2 cm/s. The experimenter decided to specify a type I error probability of significance level of α = 0.05. He selects a random sample of n = 25 and obtains a sample average burning rate of $\hat{\mu}$ = 51.3 cm/s. What conclusions should be drawn?

SOLUTION:

Test H_0 : μ = 50 cm/s , α = 0.05

H_1 : $\mu \neq$ 50 cm/s

Rejected H_0 if $z > 1.96$ or $z < -1.96$

For $\hat{\mu}$ = 51.3 cm/s and σ = 2 cm/s, then

$$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

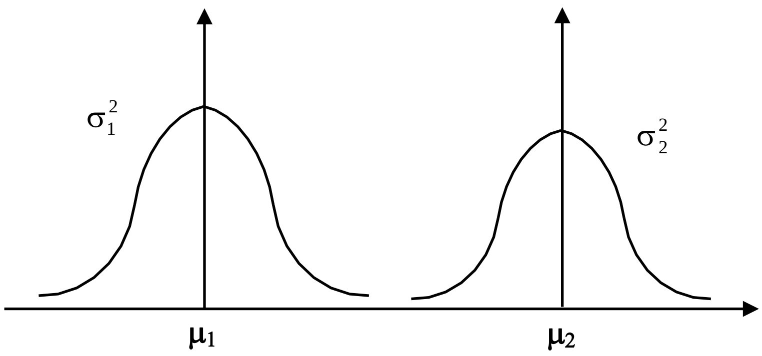
Since 3.25 > 1.96 we reject H_0 and we have strong evidence that the mean burning rate exceeds 50 cm/s.

SUMMARY FOR HYPOTHESIS TESTING PROCEDURE			
NULL HYPOTHESIS	**TEST STATISTIC**	**ALTERNATIVE HYPOTHESIS**	**CRITERIA FOR REJECTION**
$H_0: \mu = \mu_0$ σ^2 known	$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}$ N(0 , 1)	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$\lvert Z \rvert > z_{\alpha/2}$ $Z > z_\alpha$ $Z < -z_\alpha$
$H_0: \mu = \mu_0$ σ^2 unknown	$T = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{n}}$ student t-distribution with (n – 1) degrees of freedom	$H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$	$\lvert t \rvert > t_{\alpha/2, n-1}$ $t > t_{\alpha/2, n-1}$ $t < -t_{\alpha/2, n-1}$
$H_0: \sigma^2 = \sigma_0^2$ μ unknown	$\chi^2 = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2}$ Chi-square distributions with (n – 1) degrees of freedom	$H_1: \sigma^2 \neq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$\chi^2 > \chi^2_{\alpha/2, n-1}$ or $\chi^2 < \chi^2_{1-\alpha/2, n-1}$ $\chi^2 > \chi^2_{\alpha, n-1}$ $\chi^2 < \chi^2_{1-\alpha, n-1}$
$H_0: \sigma^2 = \sigma_0^2$ μ known	$\chi^2 = \frac{n\hat{\sigma}^2}{\sigma_0^2}$ Chi-square distributions with (n) degrees of freedom	$H_1: \sigma^2 \neq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$\chi^2 > \chi^2_{\alpha/2, n}$ or $\chi^2 < \chi^2_{1-\alpha/2, n}$ $\chi^2 > \chi^2_{\alpha, n}$ $\chi^2 < \chi^2_{1-\alpha, n}$
$H_0: p = p_0$	$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ N(0 , 1)	$H_1: p \neq p_0$ $H_1: p > p_0$ $H_1: p < p_0$	$\lvert Z \rvert > z_{\alpha/2}$ $Z > z_\alpha$ $Z < -z_\alpha$

- **Decision Making for Two Samples:**

- The previous chapter presented hypothesis tests and confidence intervals for a single population parameter (the mean μ, the variance σ^2, or the proportion p). Here we extend those results to the case of two independent populations.

- Population (1) has mean μ_1 and variance σ_1^2, population (2) has mean μ_2 and variance σ_2^2. Inferences will be based on two random samples of sizes (n_1) and (n_2).
  That is $X_{11}, X_{12}, \ldots\ldots, X_{1n1}$ is a random sample of (n_1) observations from population 1, and $X_{21}, X_{22}, \ldots\ldots, X_{2n2}$ is a random sample of (n_2) observations from population 2.



- **Inferences for a Difference in Means: *Variances Known***

- ***Assumptions:***

1- $X_{11}, X_{12}, \ldots\ldots, X_{1n}$ is a random sample from population 1.

2- $X_{21}, X_{22}, \ldots\ldots, X_{2n}$ is a random sample from population 2.

3- The two populations presented by X_1 and X_2 are independent.

4- Both populations are normal, or if they are not normal, the conditions for the central limit theorem apply.

- The test statistic $Z = \dfrac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ has an N(0 , 1) distribution.

- **Testing hypothesis on ($\mu_1 - \mu_2$): Variances Known**

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$

Test statistic: $Z = \dfrac{\hat{\mu}_1 - \hat{\mu}_2 - (\Delta_0)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Alternative Hypothesis	Criteria for Rejection
$H_1 : \mu_1 - \mu_2 \neq \Delta_0$	$Z > z_{\alpha/2}$ or $Z < -z_{\alpha/2}$
$H_1 : \mu_1 - \mu_2 > \Delta_0$	$Z > z_{\alpha}$
$H_1 : \mu_1 - \mu_2 < \Delta_0$	$Z < -z_{\alpha}$

- ***Definition: Confidence Interval on the Difference in Two Means: Variances Known.***

  If $\hat{\mu}_1$ and $\hat{\mu}_2$ are the means of independent random samples of sizes (n_1) and (n_2) with known variances σ_1^2 and σ_2^2, then a 100%(1 – α) confidence interval for ($\mu_1 - \mu_2$) is:

$$\left\{\hat{\mu}_1 - \hat{\mu}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \hat{\mu}_1 - \hat{\mu}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right\}$$

  where $z_{\alpha/2}$ is the upper $\alpha/2\%$ point of standard normal distribution.

- **Inferences for a Difference in Means of Two Normal Distributions:** ***Variances Unknown***

- ***Hypothesis tests for the difference in means:***

  <u>CASE I:</u> $\boldsymbol{\sigma_1^2 = \sigma_2^2 = \sigma^2}$

- The pooled estimator of σ^2 denoted by s_p^2 is defined as:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- The statistic $T = \dfrac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)}{s_P\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ has a t-distribution with ($n_1 + n_2 - 2$) degrees of freedom when H_0 is true.

- ***The Two-Sample Pooled t-test:***

  Null hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

  Test Statistic: $T = \dfrac{\hat{\mu}_1 - \hat{\mu}_2 - (\Delta_0)}{s_P\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$

Alternative Hypotheses	Criteria for Rejection
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	$t > t_{\alpha/2, n_1+n_2-2}$ or $t < t_{\alpha/2, n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	$t > t_{\alpha,, n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 < \Delta_0$	$t < -t_{\alpha,, n_1+n_2-2}$

- ***Definition: Confidence Interval on the Difference in Means of Two Normal Distributions: Variances Unknown and Equal.***

  If $\hat{\mu}_1$, $\hat{\mu}_2$, S_1^2, and S_2^2 are the means and variances of two random samples of sizes (n_1) and (n_2) respectively from two independent normal populations with unknown but equal variances, then a 100%(1 – α) confidence interval on the difference in means ($\mu_1 - \mu_2$) is:

$$\left\{\hat{\mu}_1 - \hat{\mu}_2 - t_{\alpha/2, n_1+n_2-2}\, s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le \hat{\mu}_1 - \hat{\mu}_2 + t_{\alpha/2, n_1+n_2-2}\, s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right\}$$

CASE II: $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$

If $H_0: \mu_1 - \mu_2 = \Delta_0$ is true, then the test Statistic $T^* = \dfrac{\hat{\mu}_1 - \hat{\mu}_2 - (\Delta_0)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Is distributed approximately as t with degrees of freedom given by:

$$\nu = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1+1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2+1}} - 2 \qquad \text{if } H_0 \text{ is true.}$$

- ***Definition: Confidence Interval on the Difference in Means of Two Normal Distributions: Variances Unknown and Unequal.***

If $\hat{\mu}_1$, $\hat{\mu}_2$, S_1^2, and S_2^2 are the means and variances of two random samples of sizes (n_1) and (n_2) respectively from two independent normal populations with unknown and unequal variances, then an approximate 100%(1 – α) confidence interval on the difference in means ($\mu_1 - \mu_2$) is:

$$\left\{\hat{\mu}_1 - \hat{\mu}_2 - t_{\alpha/2},\nu\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \hat{\mu}_1 - \hat{\mu}_2 + t_{\alpha/2},\nu\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right\}$$

- **Inferences on the variances of two normal populations:**

Next, we introduce tests and confidence intervals for two population variances. Both populations are assumed normal.

- ***Definition:***

Let X_{11}, X_{12}, ……, X_{1n1} be a random sample from a normal population with mean μ_1 and variance σ_1^2, and let X_{21}, X_{22}, ……, X_{2n2} be a random sample from a second normal population with mean μ_2 and variance σ_2^2. Assume that both normal populations are independent. Let S_1^2 and S_2^2 be the sample variances, then the ratio:

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

has an F distribution with (n_1 – 1) numerator degrees of freedom and (n_2 – 1) denominator degrees of freedom.

- ***Hypothesis testing procedure:***

A hypothesis testing procedure for the equality of two variances is based on the following:

Null hypothesis: $H_0: \sigma_1^2 = \sigma_2^2$

Test Statistic: $F_0 = \dfrac{S_1^2}{S_2^2}$

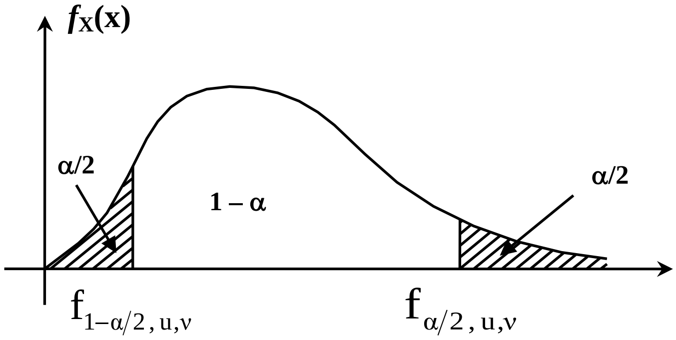
Alternative Hypotheses	Rejection Criterion
$H_1: \sigma_1^2 \neq \sigma_2^2$	$f_0 > f_{\alpha/2, n_1-1, n_2-1}$ or $f_0 < f_{1-\alpha/2, n_1-1, n_2-1}$
$H_1: \sigma_1^2 > \sigma_2^2$	$f_0 > f_{\alpha, n_1-1, n_2-1}$
$H_1: \sigma_1^2 < \sigma_2^2$	$f_0 < f_{1-\alpha, n_1-1, n_2-1}$

- ***Definition: Confidence Interval on the Ratio of Variances of Two Normal Distributions.***

If S_1^2 and S_2^2 are the sample variances of random samples of sizes (n_1) and (n_2) respectively from two independent normal populations with unknown variances σ_1^2 and σ_2^2, then a 100%(1 – α) confidence interval on the ratio $\frac{\sigma_1^2}{\sigma_2^2}$ is:

$$\frac{S_1^2}{S_2^2} f_{1-\alpha/2, n_1-1, n_2-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} f_{\alpha/2, n_1-1, n_2-1}$$

where $f_{\alpha/2, n_1-1, n_2-1}$ and $f_{1-\alpha/2, n_1-1, n_2-1}$ are the upper and lower α/2% points of the F distribution with (n_2 – 1) numerator degrees of freedom and (n_1 – 1) denominator degrees of freedom respectively.



- ***Remark:*** $f_{1-\alpha/2, u, \nu} = \frac{1}{f_{\alpha/2, u, \nu}}$

- **Inferences on Two Population Proportions:**

Now we consider the case where there are two binomial parameters of interest p_1 and p_2 and we wish to draw inferences about these proportions.

- ***Large Sample Test for H_0: $p_1 = p_2$***

Suppose that the two independent random samples of sizes (n_1) and (n_2) are taken from two populations, and let X_1 and X_2 represent the number of observations that belong to the class of interest in the samples. Furthermore, suppose that the normal approximation is applied to each population so that the estimators of the population proportions:

$\hat{P}_1 = \frac{X_1}{n_1}$ and $\hat{P}_2 = \frac{X_2}{n_2}$ have approximate normal distributions.

- ***Hypothesis testing procedure:***

Null hypothesis: $H_0: p_1 = p_2$

Test Statistic: $$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

Alternative Hypotheses	Rejection Criterion
$H_1 : p_1 \neq p_2$	$Z_0 > z_{\alpha/2}$ or $Z_0 < -z_{\alpha/2}$
$H_1 : p_1 > p_2$	$Z_0 > z_{\alpha}$
$H_1 : p_1 < p_2$	$Z_0 < -z_{\alpha}$

- ***Confidence Interval for $p_1 - p_2$:***

The confidence interval for $p_1 - p_2$ can be found from the statistic:

$$Z = \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

which is a standard normal r.v.

The 100%(1 – α) confidence interval on $p_1 - p_2$ is:

$$\hat{P}_1 - \hat{P}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq \hat{P}_1 - \hat{P}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$