This table of contents and subsequent notes in this file are for educational purposes only, specifically for personal use and study. All rights remain with the original owner/creator of these notes, as watermarked in the footer of the subsequent notes. Please do not sell or otherwise improperly distribute or use anything in this file, and don't say it's my fault when you get in trouble for it. Good luck with 419!

Table of Contents

- **BEFORE CONTINUING**
  - For improved navigability, this table of contents is designed to correspond to this PDF document's page numbers, NOT the original page numbers marked at the bottom of some pages.
  - Note 6.2 occurs twice; this is an error in the original notes and is therefore intentionally recreated in this table of contents.

This table of contents and subsequent notes in this file are for educational purposes only, specifically for personal use and study. All rights remain with the original owner/creator of these notes, as watermarked in the footer of the subsequent notes. Please do not sell or otherwise improperly distribute or use anything in this file, and don't say it's my fault when you get in trouble for it. Good luck with 419!

This table of contents and subsequent notes in this file are for educational purposes only, specifically for personal use and study. All rights remain with the original owner/creator of these notes, as watermarked in the footer of the subsequent notes. Please do not sell or otherwise improperly distribute or use anything in this file, and don't say it's my fault when you get in trouble for it. Good luck with 419!

This table of contents and subsequent notes in this file are for educational purposes only, specifically for personal use and study. All rights remain with the original owner/creator of these notes, as watermarked in the footer of the subsequent notes. Please do not sell or otherwise improperly distribute or use anything in this file, and don't say it's my fault when you get in trouble for it. Good luck with 419!

1. Foundations of Probability

(Reading Exercises: Montgomery and Runger Chapter 1 -Sections 2.1, 2.3 and 2.4)

Learning outcomes:

You will be able to do the following:

- Distinguish between a population and a sample
- Distinguish between a random and a deterministic experiment.
- Distinguish between a simulation model, a deterministic model and a probability model.
- Distinguish between probability theory and statistics
- Understand why probability is so critically important to the advancement of most kinds of electrical engineering research and design.
- Define an event.
- Derive new events by taking subsets, unions, intersections, and/or complements of already existing events.
- State the definitions of specific kinds of events, namely empty events, mutually exclusive (or disjoint) events, and exhaustive events.
- State the formal definition of probability.
- State the three ways of assigning a probability to an event.
- State the probability axioms and their corollaries,
- Apply the axioms to determine probabilities of various events.
- Get lots of practice calculating probabilities of various events

1.1 Basic Definitions: Sample spaces and events

Definitions:

- **Experiment:** “A scientific procedure undertaken to make a discovery or demonstrate a known fact.
- **Outcome**: the result of an experiment
- **Deterministic experiment:** Outcomes are predictable when the experiment is repeated under the same conditions.
- **Random experiment:** Outcomes are unpredictable when the experiment is repeated under the same conditions.
- **Sample space:** Denoted S, is a collection of all possible outcomes of an experiment
- **Discrete sample space:** S contains discrete numbers of elements (outcomes)
  - **Finite sample space:** contains a finite number of discrete elements
  - **Infinite sample space:** contains an infinite number of discrete elements
- **Continuous sample space:** S contains intervals, for example, $S = \{[0,1],[2,3]\}$.

  An outcome can fall anywhere in an interval, for instance $[0,1]$

Models of Systems

Definition:

- **Model:** A description of a real physical system, process or phenomenon that we want to analyze. A model is used to help us explain observed behavior of a physical system, process or phenomenon using a set of simple and understandable rules. These rules can be used to predict the outcome of experiments involving a given physical situation.
- **Deterministic models (Mathematical models):** the solution of a set of mathematical equations specifies the exact outcome of the experiment.
- **Probabilistic Models:** In many practical situations, we are not able to accurately model some or all aspects of a physical system or process due to **inaccuracies, uncertainties, randomness in measurements,** etc.
- **Computer simulation model:** A computer program is used to simulate or mimic the dynamics of a system

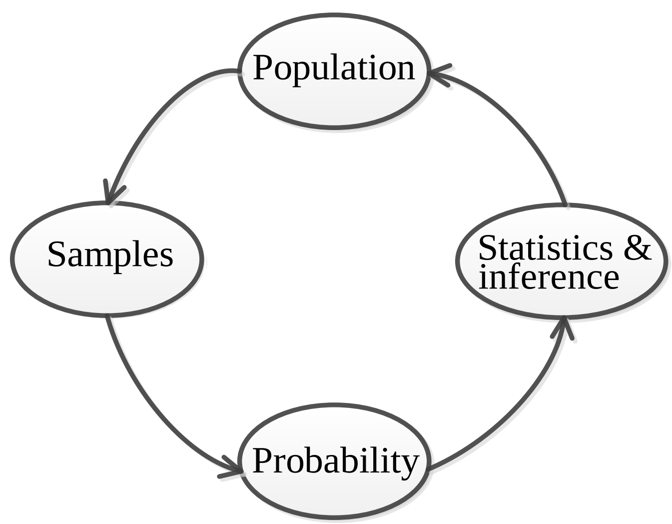
Populations and Samples

Definitions:

- **Population:** a collection of all objects or elements under study
- **Sample:** a subset of a population selected for studying or testing
- **Probability theory:** The study of the mathematical rules that govern random events
- **Statistics:** The application of probability theory to the collection, analysis, and description of random data for the purpose of making inferences, judgements or conclusions about a population

Example: An electronic manufacturing company of transistors can tolerate, in the long run, 5% defective production. A quality control engineer in that company must determine whether the defective rate is within a tolerable range. The engineer accomplishes this by taking, for example, 1000 transistors from the assembly line and analyzing them.

- **Population:** The totality of all possible transistors coming out of the assembly line.
- **Sample**: The 1000 transistors taken from the assembly line.
- The engineer uses **probability theory** to determine whether the defective transistors, in the long-run, meet the 5% rate or less.
- The engineer can also determine the chances of obtaining more than 5% defective transistors.
- The engineer uses **Statistics** to make **inference** regarding the acceptance or rejection of the production process at a certain confidence level.



1.2 Basic Set Theory:

Set theory provides a basic mathematical tool for studying probability. Set theory may be considered as the **algebra of probability theory**. The math involved in probability theory relies on elements of sets. The table below summarizes some correspondence of set algebra and probability theory.

Set Algebra	**Probability Theory**
Set	Event
Universal set	Sample space
Element	Outcome

Basic Set Definitions and notations

Definition:

- **Set:** An unordered collection of events (or objects or elements). Sets can contain items of mixed types or other sets.
- **Subsets:** If every element of a set A is also an element of set B, we say that A is a subset of B, and we write, $A \subset B$ or $B \supset A$.
- **Equal sets:** $A = B$ ~~or~~ $\cancel{A \subseteq B}$ ~~and~~ $\cancel{B \subseteq A}$. That is all the elements of A are also the elements of B and vice versa.

Definitions:

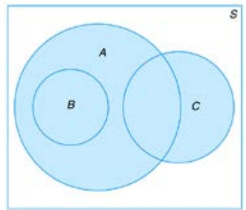
- **Universal Set**: The same as a sample space. All other sets are subsets of the universal set.
- **Event:** A subset of the universal set or sample space S and is usually denoted with capital letters, such as $A, B, C, \ldots$
- **Empty or null Set:** $S = \varnothing$
- $x \in S$
- $x \notin S$

Examples:

- $A = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$
- $B = \{1,\ 2,\ 3,\ 4\}$ and $C = \{4,\ 1,\ 3,\ 2\}$
- $x = 5$ and $x = 7$
- The event A that the chip fails before the end of the third year (the warranty period) is the subset $A = \{t \mid 0 \le t < 3\}$.

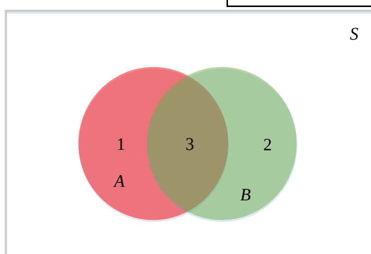
1.3 Basic Set Operations and Venn Diagrams

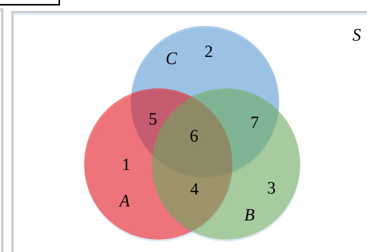
- Venn diagrams
  Venn diagrams are used to graphically represent (or visualize) sets and set operations.



- **Union operation:** The **union** of two disjoint sets A and B is a set that contains every element that is either in A or in B or in both. We denote the union operation of the sets A and B as

$$\boxed{C = A \bigcup B}$$





$A \bigcup B =$ all shaded regions $\quad A \bigcup B \bigcup C =$ all shaded regions

Example:

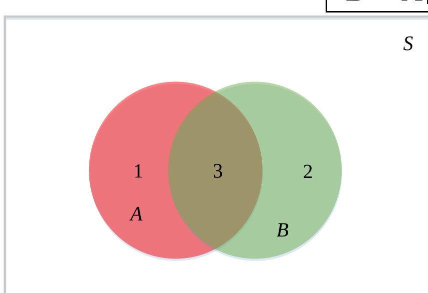
$A = \{1,\ 2,\ 6\}$ and $B = \{1,\ 3,\ 4,\ 5,\ 6\}$

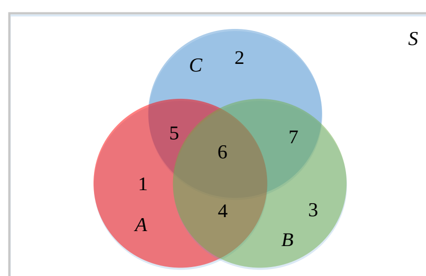
$\{1,\ 2,\ 3\} \bigcup \{6,\ 7,\ 8\} = \{1,\ 2,\ 3,\ 6,\ 7,\ 8\}$

- Analogy with logic circuit operations

- **Intersection operation:** The intersection of two sets A and B contains only elements that are in A and in B. We denote the intersection of the sets A and B as

$$\boxed{D = A \bigcap B}$$





$A \bigcap B =$ region 3 $\quad A \bigcap B \bigcap C =$ Region 6

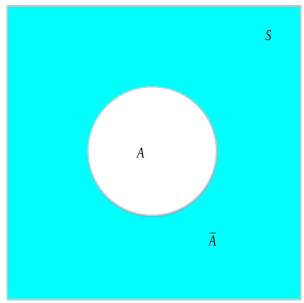
Example:

$\{1,\ 2,\ 3,\ 7,\ 8\} \cap \{6,\ 7,\ 8\} \ = \ \{7,\ 8\}$

Let $A = \{1,2,5,6\}$ and $B = \{2,4,6\}$. Then, $A \cap B = \{2,6\}$.

- Analogy with logic operations

- **Complement operation:** The complement of the set A, denoted A^c or $\overline{A}$, is the set containing all elements in S that are not in A
  Denoted A^c or $\overline{A}$



$\overline{A} = A^c = A' =$ complement (shaded area)

<u>**Example:**</u>

$S \ = \ \{1,\ 2,\ \ldots,\ 10\}$

$\{1,\ 2,\ 3,\ 4,\ 5\}^c = \{6,\ 7,\ 8,\ 9,\ 10\}$

$\{2,\ 4,\ 6,\ 8,\ 10\}^c = \{1,\ 3,\ 5,\ 7,\ 9\}$

- Analogy with logic operations

1.4 Basic Set Identities or Laws

They are used to simply complex set operations.

- **Commutative property:**

1. $A \bigcup B = B \bigcup A$	2. $A \bigcap B = B \bigcap A$

Example: Use Venn diagrams

- **Associative property:**

1. $(A \cup B) \cup C = A \cup (B \cup C)$	2. $(A \cap B) \cap C = A \cap (B \cap C)$

Example: Use Venn diagrams

- **Distributive property:**

$$\boxed{(A \cup B) \cap C = (A \cap C) \cup (B \cap C)}$$

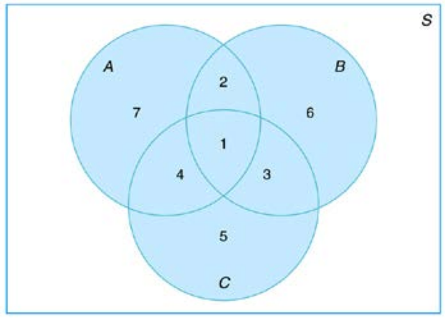
Example: Use Venn diagrams

- **DeMorgan's Laws:**

1. $\overline{A \bigcup B} = \overline{A} \bigcap \overline{B}$	2. $\overline{A \bigcap B} = \overline{A} \bigcup \overline{B}$

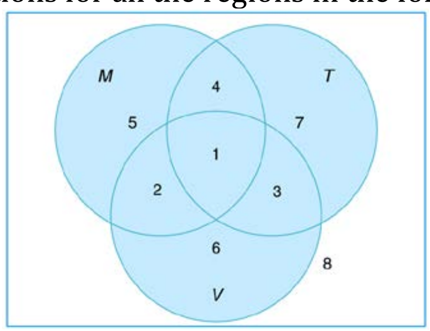
Use Venn diagrams

Example: Write the set operations for each region represented by the numbers 1, 2, 3, 4, 5, 6, 7 and 8 in the Venn diagrams below. The solutions are summarized in the table below.



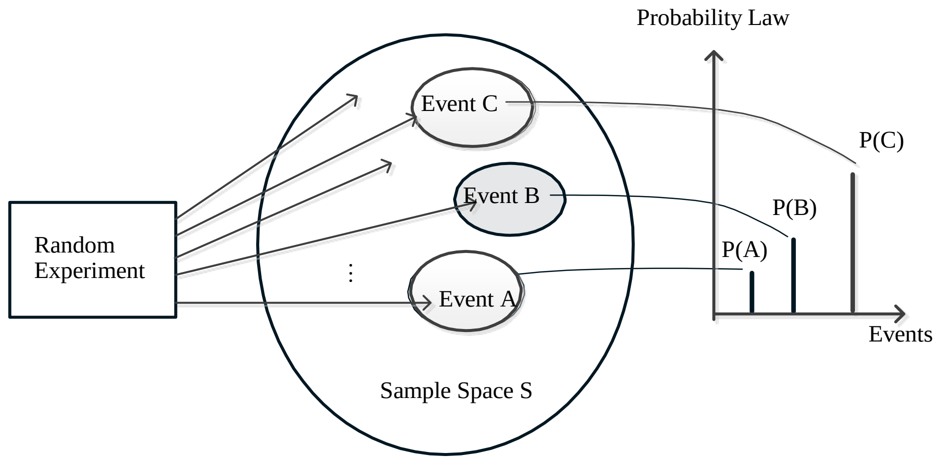
Region 1: $A\cap B\cap C$	Region 2: $A\cap B\cap \bar{C}$	Region3: $\bar{A}\cap B\cap C$	Region 4: $A\cap C\cap \bar{B}$
Region 5: $\overline{A\cup B}\cap C$	Region 6: $\overline{A\cup C}\cap B$	Region 7: $A\cap \overline{B\cup C}$	

Exercise: Write the set operations for all the regions in the following Venn diagram



1.5 Probability Model:

In this section, we study the concept of **probability**, the **rules of probability**, and **probability models**. A probability model is illustrated in the diagram below.



Probability Model

A probability model comprises the following:

- A random experiment - this produces outcomes or sample points
- A sample space, S - This contains the set of all possible outcomes of the random experiment.
- Probabilities assigned to each of the sample points, making sure they sum up to 1.
- Events, denoted $A_i,\ i=1,2,\ldots,n$ - A group or collection of outcomes or sample points that form subsets of the sample space
- Probability rule - a function, which assigns a nonnegative number to each event. This number is called the probability of an event and is written as $P[\mathrm{A}_i]$, $i=1,2,\ldots,n$. The probability encodes our knowledge or belief about the collective "likelihood" of the event to occur. For example, if we are absolutely certain an event will occur, we say the probability is 100% - $P[\text{event}]=1$.

Example:

1.6 Probability Axioms and Corollaries

Axioms:

Definition:
Axioms are facts and do not need proofs.

- Axioms are properties that the probability of an event satisfies.
- Axioms do not provide a means of specifying the probabilities.

Probability and Axioms of Probability (Andrey N. Kolmogorov 1903-1987)

Definition:

Probability is a (real-valued) function, denoted $P[\cdot]$, that assigns to each event, A, in the sample space S, a number $P[A]$, called the probability of the event A.

A probability must satisfy the following three axioms:

- **Axiom 1:** For every event A, $P(A) \geq 0$.
- **Axiom 2:** For the sure or certain event S, $P[S] = 1$.
- **Axiom 3:** For any number of disjoint (mutually exclusive) events $A_1, A_2, A_3, \ldots$, $A_i \cap A_j = \emptyset$, for $i \neq j$,

$$P[A_1 \cup A_2 \cup \cdots] = \sum_{k=1}^{\infty} P[A_k]$$

Corollaries of Probability Axioms

Definition:

Corollaries are properties that can be derived from axioms.

The following is a list of corollaries of probability:

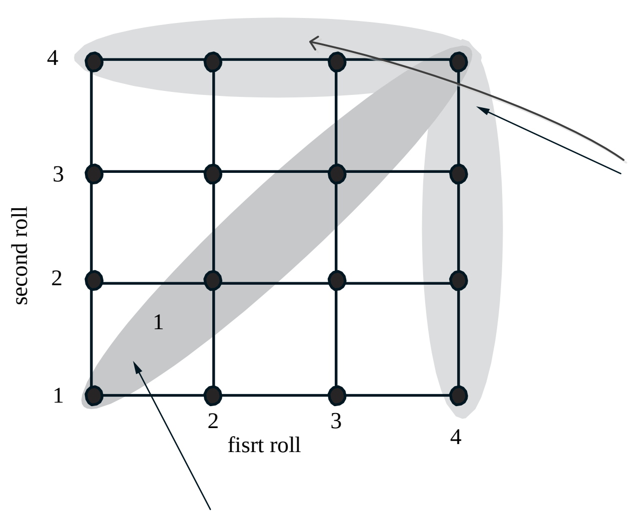
1. $P[A^c] = P[S] - P[A] = 1 - P[A]$ or $P[A] = 1 - P[A^c]$
2. For every event $0 \leq P[A] \leq 1$
3. $P[\emptyset] = 0$
4. If events A and B are such that $A \subseteq B$, then $P[A] \leq P[B]$ and $P[B - A] = P[B] - P[A]$.
5. For non-disjoint events, that is, $A \cap B \neq \emptyset \Rightarrow P[A \cap B] \neq 0$, we have $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

1.7 Methods of Assigning Probabilities

- **Subjective approach:**
  This is a judgement approach. Here, probability is defined as the degree of belief that we hold in the occurrence of an event (non-repeatable experiments). Examples are horse race and stock price, etc.
- **Relative frequency approach**:
  The relative frequency approach involves taking the following three steps in order to assign the probability of an event A:
  (1) Perform an experiment, an indefinite number of times, $n \to \infty$.
  (2) Count the number of times the event A of interest occurs and denote this number as n_A.
  (3) Then, the probability of event A occurring equals:
  $$P[A] = \lim_{n\to\infty} \frac{n_A}{n} = \lim_{n\to\infty} P_n[A]$$
- **Classical Approach**
  The classical approach assumes that all distinct (elementary) outcomes are equally likely, that is, they have equal probability of occurring. Let the sample space, S, be the set of all possible distinct outcomes to an experiment. The probability of some event, A, is defined as
  $$P[A] = \frac{\text{Number of times or ways event } A \text{ can occur}}{\text{Number of all possible outcomes in } S}$$

<u>**Example:**</u> Two 4-sided die rolls

Experiments involving rolling of 2, 4-sided dice



Prob{event that he fisrt roll is equal to the second} = 4/16

Prob{Event that at least one roll is a 4} = 7/16

Example: An urn contains 10 identical balls numbered $0,1,2,\ldots,9$. A random experiment involves selecting a ball from the urn. Find the probability of the following events:

1. $A = \{\text{number of balls selected is odd}\}$
2. $B = \{\text{number of balls selected is a multiple of 3}\}$
3. $C = \{\text{number of balls selected is less than 5}\}$
4. $D = A \cap B$
5. $E = A \cup B$

Example: A fair coin tossed three times. Find the probability of exactly two heads in three tosses.

$$P\left[\{\text{exactly 2 heads in 3 tosses}\}\right] = P\left[\{HHT, HTH, THH\}\right] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

Exercise: (a) One head in first toss, (b) At least two heads,

Some Major Pioneers of Probability and Statistics:

Thomas Bayes 1701– 1761, English Statistician, Philosopher & Minister) -Wikipedia.com 	*Daniel Bernoulli (1700 – 1782, Mathematician & Physicist) - Wikipedia.com* 
Siméon Denis Poisson 1781 – 1840, French Mathematician, Engineer & Physicist) -Wikipedia.com 	*Johann Carl Friedrich Gauss (1777– 1855, German Mathematician) - Wikipedia.com* 
Andrey NikolaevichKolmogorov (1903 – 1987, Russian Mathematician) - Wikipedia.com 	*William Sealy Gosset (1876 – 1937, English Statistician) – Wikipedia.com* 

2. **Conditional Probability and Independence**
(Reading Exercises: Montgomery and Runger Chapter 2 - Sections 2.5-2.8)

Conditional probability provides us with a way to reason about the outcome of an experiment, based on **partial or prior information.** Here are some examples of situations we have in mind:

(1) In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?
(2) In a word guessing game, the first letter of the word is a "t". What is the likelihood that the second letter is an "h"?
(3) How likely is it that a person has a disease given that a medical test was negative?
(4) A spot shows up on a radar screen. How likely is it that it corresponds to an aircraft?

2.1 Conditional Probability

Example: An electrical engineering lab has 20 probes of which 3 are defective. A student selects 2 probes randomly (one after the other), what is the probability that both are defective?

Solution: Define events $A=\{\text{First probe is defective}\}$ and $B=\{\text{Second probe is defective}\}$. The probability that both are defective is

$$P[A\cap B]=P[A]P[B\,|\,A];$$

When we first probe, the probability is $P[A]=\frac{3}{20}$. After picking the first, there are 19 probes remaining, of which 2 are bad, therefore, the probability that the second is defective given that the first is defective, is

$$P[B\,|\,A]=\frac{2}{19}$$

$$P[A\cap B]=\frac{3}{20}\times\frac{2}{19}=\frac{3}{190}.$$

Definition:

- Consider two events A and B. The probability of A occurring, given that B has occurred is the ***conditional probability*** defined as

$$P[A\,|\,B]=\frac{P[A\cap B]}{P[B]}\Rightarrow P[A\cap B]=P[A\,|\,B]P[B]$$

- Similarly, the probability of B occurring, given that A has occurred is the ***conditional probability*** defined as

$$P[B\,|\,A]=\frac{P[A\cap B]}{P[A]}\Rightarrow P[A\cap B]=P[B\,|\,A]P[A]$$

- The joint probability of A and B is defined as the intersection

  $$P[A \cap B] = P[A \mid B]P[B] = P[B \mid A]P[A]$$

Multiplication Rule:

Consider a set A comprising subsets $A_1, A_2, \ldots, A_n$. According to the multiplication rule, the joint probability is given by the following product:

$$P[A_1 A_2 \ldots A_n] = P[A_1]P[A_2 \mid A_1]P[A_3 \mid A_1 A_2]\cdots P[A_n \mid A_1 A_2 \ldots A_{n-1}]$$

Example: Consider events $D = \{\text{flight departs on time}\}$ and $A = \{\text{flight arrives on time}\}$. The probability of a flight departing on time is $P[D] = 0.83$ and the probability of arriving on time is $P[A] = 0.82$. The joint probability that a flight departs and arrives on time is $P[D \cap A] = 0.78$. (a) What is the probability that the flight will arrive on time if it departed on time? (answer: 0.94) (b) What is the probability that a flight departed on time if it arrived on time? (answer: 0.95)

Example: A box contains 5000 IC chips, of which 1000 are manufactured by company X and the rest by company Y. Ten percent of the chips made by company X are defective and 5% of the chips made by company Y are defective. If a randomly select chip is found to be defective, what is the probability that it came from company X?

	X	Y	Total
Total # of chips	1000	4000	5000
Probability of events	$P[X] = 0.2$	$P[Y] = 0.8$	1
# of Defectives: $0.1 \times X + 0.05 \times Y$	100	200	300 $P[D] = 0.06$
Conditional probability of defect	$P[D \mid X] = 0.1$	$P[D \mid Y] = 0.05$	

$P[D \cap X] = 0.02$	$P[D \cap Y] = 0.04$
$P[X \mid D] = \dfrac{P[D \cap X]}{P[D]} = 0.33$	$P[Y \mid D] = \dfrac{P[D \cap Y]}{P[D]} = 0.67$

2.2 Independence

Definitions:

- Two events, A and B, are independent if one of the following is true:

$$P[A|B]=P[A] \text{ or } P[B|A]=P[B] \text{ or } P[A\cap B]=P[A]P[B]$$

- Multiple events $\{A_i\}_{i=1}^{n}$ are independent if

$$P[A_1\cap A_2\cdots\cap A_n]=P[A_1]P[A_2]\cdots P[A_n]$$

Example: Consider three events with corresponding probabilities given by the table:

Event A: College student	Event B: Smoker	Event C: Heart disease
$P[A]=0.7$	$P[B]=0.1$	$P[C]=0.05$
$P[A\cap C]=0.035$	$P[B\cap C]=0.03$	

$$P[C|A]=\frac{P[C\cap A]}{P[A]}=\frac{0.035}{0.7}=0.05=P[C]\Rightarrow A \ \& \ C \text{ independent}$$

$$P[C|B]=\frac{P[C\cap B]}{P[B]}=\frac{0.03}{0.1}=0.3\neq P[C]\Rightarrow B \ \& \ \text{C are dependent}$$

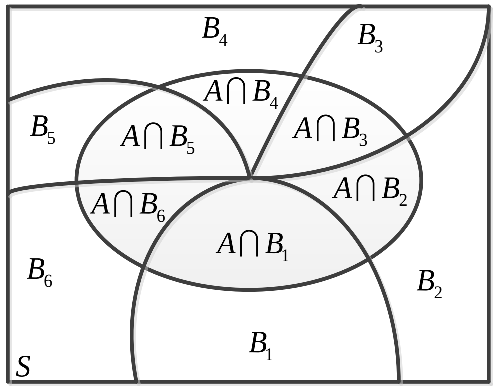
2.3 Theorem of Total Probability

Objective: To compute the probability of an event $A\in S$, in terms of conditional probabilities.

Divide and conquer

- Partition the sample space into disjoint sample spaces $B_1, B_2, \cdots, B_n$ as shown in the Venn diagram below. Then the event A may be expressed as the sum

$$A=\sum_{i=1}^{n} A\cap B_i$$



- Suppose we are given the probabilities $P[A \mid B_i]$, and $P[B_i]$ for each i. Then

$$P[A \cap B_i] = P[A \mid B_i] P[B_i],$$

- Compute $P[A]$ in terms of $P[A \mid B_i]$, and $P[B_i]$ for $i = 1, 2 \cdots, n$, that is,

$$\boxed{P[A] = \sum_{k=1}^{n} A \cap B_k = \sum_{k=1}^{n} P[A \mid B_k] P[B_k]}$$

  - The above expression is known as the ***Law of total probabilities***

<u>**Example:**</u> Two balls are drawn in succession (without replacement) from an urn containing 2 black balls and 3 white balls. Find the probability the second ball drawn is white.

<u>**Solution:**</u> Define the following events:

$$B_1 = \{1^{st} \text{ ball is black}\}, \quad W_1 = \{1^{st} \text{ ball is white}\}, \quad W_2 = \{2^{nd} \text{ ball is white}\}$$

Define event $W = \{B_1 \cap W_2 \text{ or } W_1 \cap W_2\}$. Then

$$P[W] = P[W_2 \mid B_1] P[B_1] + P[W_2 \mid W_1] P[W_1] = \frac{3}{4} \times \frac{2}{5} + \frac{2}{4} \times \frac{3}{5} = 0.6$$

<u>**Example:**</u> In a manufacturing plant, three machines, B_1, B_2, and B_3 make 30%, 30% and 40%, respectively, of the products. It is known that some of these products are defective. The defective rates from the three machines, B_1, B_2, and B_3, are 10%, 4% and 7%, repectively. Using the theorem of total probability, find the probability a selected product is defective. (answer: $P[D] = 0.07$). Hint: Define event $D = \{\text{selected product is defective}\}$. From the given information we have the following:

$P[B_1] = 0.3$ $\qquad$ $P[B_2] = 0.3$ $\qquad$ $P[B_2] = 0.4$

$P[D \mid B_1] = 0.1$ $P[D \mid B_2] = 0.04$ $P[D \mid B_3] = 0.07$

2.4 Bayes' Rule

We are given disjoint events $B_i,\ i = 1, 2, \ldots, n$, that are partitions of of event A. Suppose we know the following probabilities:

- Probabilities $P[B_i],\ i = 1, 2, \ldots, n$, are called **a priori probabilities**. They are measures of initial beliefs
- Probabilities $P[B_i \mid A],\ i = 1, 2, \ldots, n$, are conditional probabilities called **a posteriori probabilities.** They are measures of belief given that A occurred

Given conditional probabilities probabilities $P[A \mid B_i],\ i = 1, 2, \ldots, n$, we wish to compute the a posteriori probability $P[B_j \mid A],\ j = 1, 2, \ldots, n$. The expression for finding this probability is known as Bayes' Rule

Bayes' Rule:

$$P[B_j \mid A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A \mid B_j]P[B_j]}{P[A]} = \frac{P[A \mid B_j]P[B_j]}{\sum_{k=1}^{n} P[A \mid B_k]P[B_k]}, \quad 1 \le j \le n$$

Example: In the previous example,

$P[B_1] = 0.3$ $P[B_2] = 0.3$ $P[B_2] = 0.4$

$P[D \mid B_1] = 0.1$ $P[D \mid B_2] = 0.04$ $P[D \mid B_3] = 0.07$

What is the probability that a defective product is from machine B_1 (i.e, $P[B_1 \mid \mathrm{D}] = 0.428$)?

What is the probability that a defective product is from machine B_2?

Example:

Coin A is Fair, that is, 50% heads and 50% tails, that is,

$P[\text{heads} \mid A] = 0.5$ and $P[\text{tails} \mid A] = 0.5$

Coin B is rigged with 75% heads and 25% tails, that is,

$P[\text{heads} \mid B] = 0.75$ and $P[\text{tails} \mid B] = 0.25$

Coin A or coin B is selected, with equal probability, it is tossed, and heads appear. Given that heads appeared, what is the probability that coin B was selected? In other words, find $P[B|\text{heads}]$ (answer: $\frac{3}{5}$). Show whether events B and $\{Heads\}$ are independent or not

Hint: Apply generalized Bayes' Rule:

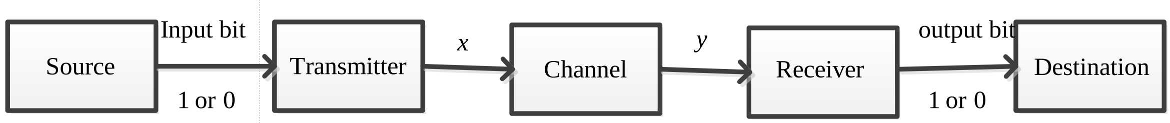
Example: Two machines: A produces 60% of the products of which 2% is defective and B produces 40% of the products of which 4% is defective. A product is examined and found to be defective. (a) Find the probability $P[D]$, that it is defective. (answer: $P[D]=0.28$) (b) Find the probability $P[A|D]$ that it came from A given that it is defective. (answer: $P[A|D]=0.57$) (e) Show that the probability that it came from B given that it is defective is $P[B|D]=0.429$.

Answers: Define event D, the product is defective.

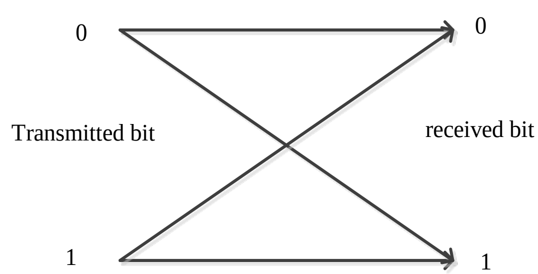
$P[A]=0.6, \quad P[B]=0.4. \; P[D|A]=0.02, \quad P[D|B]=0.04.$

$$P[A|D]=\frac{P[D|A]P[A]}{P[D]}=\frac{0.02\times 0.6}{0.028}=0.43, \quad P[B|D]=\frac{P[D|B]P[B]}{P[D]}=\frac{0.04\times 0.4}{0.028}=0.57$$

Example: The block diagram below is a simple digital, binary communication system. The system transmits a bit 0 or a bit 1. The channel corrupts the signal, which may cause errors at the receiver. The errors may cause bits to be switched.



Binary Communication System



These events and their probabilities are summarized in the table below.

Events	Probabilities
A_0: Transmitted bit is 0 A_1: Transmitted bit is 1 $B_0 \vert A_0$: Received bit 0 when is 0 is transmitted $B_1 \vert A_1$: Received bit 1 when 1 is transmitted $B_1 \vert A_0$: Received bit 1 when is 0 is transmitted $B_0 \vert A_1$: Received bit 0 when is 1 is transmitted	$P[A_0] = 1-p$ $P[A_1] = p$ $P[B_0 \mid A_1] = \varepsilon \Rightarrow P[B_1 \mid A_1] = 1-\varepsilon$ $P[B_1 \mid A_0] = \varepsilon \Rightarrow P[B_0 \mid A_0] = 1-\varepsilon$

<u>**Example:**</u> Game show (Monty Hall Problem) – there are 3 doors labeled Door1, Door2 and Door3. Behind one door is a car, behind each other door is a goat. The objective is to win the car. You select a door which remains closed and the host opens one of the 2 remaining doors to show a goat. Would you remain with your first choice or switch to the other closed door?

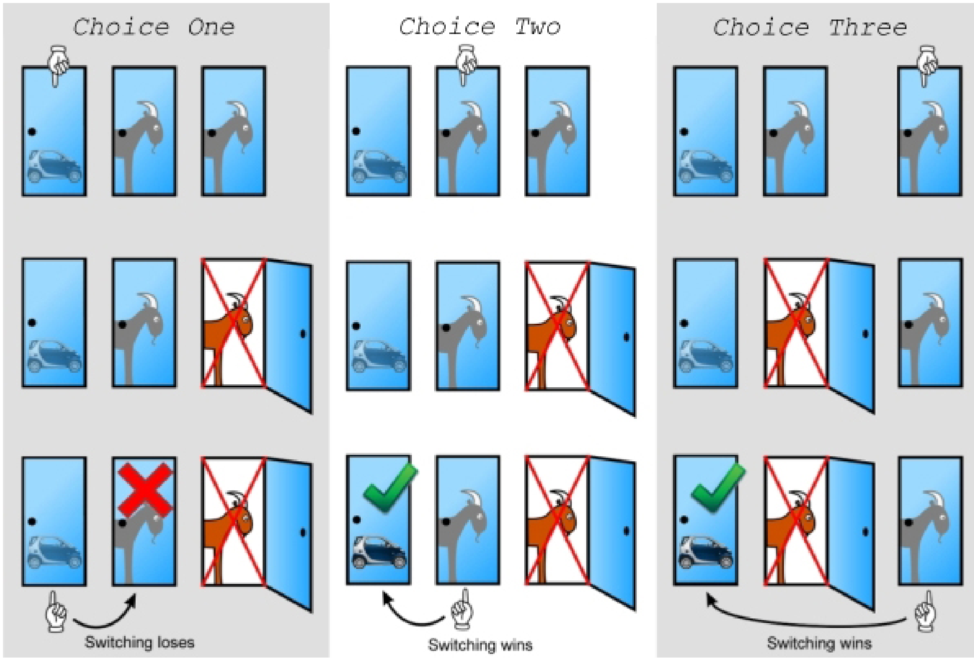
Fact:

Partitions	Events	Door1	Door2	Door3	Probability
Partition 1	B_1	C	G	G	1/3
Partition 2	B_2	G	C	G	1/3
Partition3	B_3	G	G	C	1/3

Assume you always choose Door1. The probability of the car is $P[C] = 1/3$. Host opens door on goat (Door 2 or 3). The events are summarized in the table below.

Events	Door1	Door2	Door3	Probability
$B_1 \mid \text{Host}$	C	G	G	$1/3$
$B_2 \mid \text{Host}$	G	C	G	$1/3$
$B_3 \mid \text{Host}$	G	G	C	$1/3$

Reference: http://www.bing.com/images/search?q=monty+hall+problem



The probability of winning the car given that you do not switch is

$$P[C \mid \text{no switch and host}] = P[B_1 \mid \text{Host}] = 1/3$$

If you switch, there are two possibilities

$$P[C \mid \text{switch and host}] = P[B_2 \text{ or } B_3 \mid \text{Host}] = P[B_2] + P[B_3] = 2/3$$

3. Sequential Experiments and Principles of Counting (Combinatorics)

(Reading Exercises: Montgomery and Runger Chapter 2 - Sections 2.2 and class notes)

Applications:

- The study of all possible arrangements of discrete objects,
- Algorithm complexity analysis,
- Resource allocation & scheduling, for example internet resources and frequency resources in communications (for example wireless communications),
- Security analysis (for example, assignment of IP and security codes),

Definition: A sequential experiment is one that consists of a sequence of sub-experiments or tasks $A_1, A_2, \ldots, A_K$, where each subsequent task is dependent on the previous. Furthermore, there are n ways of completing sub-task A_i, for each $i = 1, 2, \cdots K$.

Problem: We need to determine the probability of an event in a sequential experiment.

- The classical approach for computing the probability of an event A requires counting of all possible outcomes,

$$P[A] = \frac{\text{number of occurences of event } A}{\text{total number of occurrences of all events}} = \frac{n_A}{n_S}$$

For sequential experiments, we need the counts n_A and n_S, which can be extremely difficult task.

Learning Objectives:

You will:

- Know and be able to apply the multiplication principle.
- Know how to count objects when the objects are sampled with replacement (repetition).
- Know how to count objects when the objects are sampled without replacement (no repetition).
- Know and be able to use the permutation formula to count the number of ordered arrangements of n objects taken n at a time.
- Know and be able to use the permutation formula to count the number of ordered arrangements of n objects taken $k < n$ at a time.
- Know and be able to use the combination formula to count the number of unordered subsets of k objects taken from n objects.
- Know and be able to use the combination formula to count the number of distinguishable permutations of n objects, in which k are of the objects are of one type and $n-k$ are of another type.

- Understand and be able to count the number of distinguishable permutations of n objects, when the objects are of more than two types.
- Know how to apply the methods learned in this section to new counting problems.

Counting methods: These are systematic methods used to count the total number of ways to complete an entire experiment

The following is a list of counting methods most frequently used:

- Tree diagram
- Multiplication rule
- Permutations
- Combinations
- Distinguishable Permutations

3.1 Method 1: Tree diagram

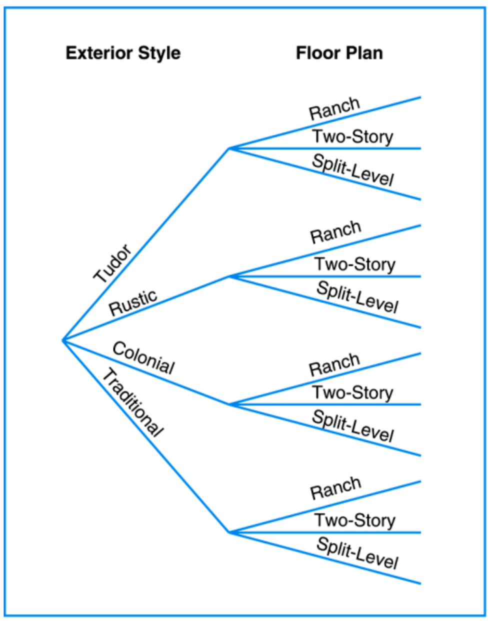
Example: A property developer offers prospective home buyers a choice of 4 exterior styles: Tudor, Rustic, Colonial and Traditional. The developer also offers 3 floor plans, Ranch, Two-storey and Split level.

Solution: Choosing a property involves two sub-tasks:

(1) Sub-task A_1 - selecting exterior styles in $n_1 = 4$ ways;

(2) Sub-task A_2 - selecting floor plans in $n_2 = 3$ ways.

A tree diagram is shown below.



Counting the number of branches on the right-hand side of the tree, gives 12 different ways the buyer may choose a property. Each different way corresponds to an event.

3.2 Method 2: Multiplication rule

Multiplication Rule: Suppose an experiment (or procedure) consists of a sequence of sub-tasks $A_1, A_2, \ldots, A_K$; each completed one after the other. Furthermore, the sub-tasks can be completed in $n_1, n_2, \ldots, n_K$ ways, respectively. According to the multiplication rule, the total number of ways to complete the entire experiment is given by the product

$$n_S = n_1 \times n_2 \times \ldots \times n_K$$

Conditions:

- Duplication is permissible
- Order is important

Example: Consider the previous example involving the choice of a property. The total of ways to select a property is

$$n_S = n_1 \times n_2 = 12$$

Example: The computer parts example - Different combinations. We want to build a computer from parts. We can buy the motherboard from two companies (A, B), RAM from four companies (A, C, D, E), hard drive from three companies (B, D, F) and graphics card from two companies (G, H). For each part, we are equally likely to buy from any one of the companies that manufactures that part. What is the probability of the event X, that we build a computer that has at least one part from company D?

Solution: The number of choices for the individual events are $n_1 = 2,\ n_2 = 4,\ n_3 = 3,\ n_4 = 2$.
The total number of possible choices: $n_s = n_1 \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 2 = 48$.
For "at least one part from D", consider the complement $\bar{D}$, "no part from D"; see table below. It is best to tabulate the given quantities:

Motherboard	RAM	HD	GPU
A, B	A, C, E; no D	B, F; no D	G, H
$n_1 = 2$	$n_2 = 3$	$n_3 = 2$	$n_4 = 2$

$$N_{\bar{D}} = n_1 \times n_2 \times n_3 \times n_4 = 2 \times 3 \times 2 \times 2 = 24$$
$$\Rightarrow P[X] = 1 - P[\bar{D}] = 1 - \frac{N_{\bar{D}}}{N} = \frac{24}{48} = 0.5$$

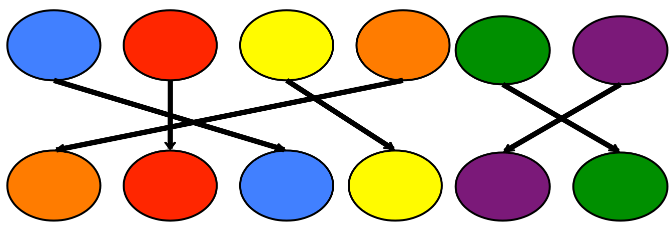
Exercise: Try to use the tree diagram approach to confirm the above result.

Exercise: For the computer parts above answer the following questions: (a) What is the probability that only one part is ordered from company A (b) At least one part is ordered from company A.

3.3 Method 3: Permutations

Definition:
A permutation is the number of ordered arrangements of r objects selected from n distinct objects $(r \le n)$.

Conditions: Permutation arrangement
- Order matters
- Duplications, repetitions or replacements are not permissible



Example: The set $\{3,1,2\}$, is a permutation of the set $S=\{1,2,3\}$.

Suppose we have n distinct objects to arrange in a particular order. The following gives the total number of possible permutations:

- $k=n$: Ordering (arranging or choosing) n elements – no replacement
$$n_S = n\times(n-1)\times(n-2)\times\cdots\times 1 = n!$$

- $k\neq n$: Ordering (or choosing) k elements – no replacement
$$n_S = n\times(n-1)\times(n-2)\times\cdots\times(n-k+1) = \frac{n!}{(n-k)!} = {}_nP_k$$

Note: if replacement is allowed for $k=n$, then

$$n_S = n\times n\times\ldots\times n = n^k$$

Permutation can also be thought of as a process of placing k people or objects into offices, positions or slots with one person in each office, position or slot. Repetition is not allowed since only one person can occupy a room. In that case, the first position can be filled in n ways, the second in $(n-1)$ ways and so on. The number of subsequent ways will reduce by one each time.

Example: How many possible Alberta license plates could be stamped if each license plate is required to have exactly 3 letters and 4 numbers?

Solutions: (a) Repetitions allowed: $175,760,000$ license plates, (b) No repetitions - $78,624,000$

3.4 Method 4: Combinations

Definition:
A Combination is a permutation in which objects are arranged in any order. The number of unordered subsets of n objects taken k (for $k \le n$) at a time, is given by

$$n_S = \binom{n}{k} = \frac{n!}{(n-k)!k!} = {}_nC_k$$

- The notation $\binom{n}{k}$, is read "from n chose k"; (like the LOTO 649)
- The k represents the number of objects one would like to select (without replacement and without regard to order) from the n objects one has.

Conditions: Combination arrangements
- Order is not important
- Duplication is not permissible

Example: A box contains 75 good IC chips and 25 defective chips. If 12 chips are selected at random, find the probability that at least one chip is defective.

Solution: Let A denote the event that at least one chip is defective. Consider the complement $\overline{A}$, the event that no chip is defective. Then all 12 chips will come from the 75 good chips. The number of combinations of 12 from 75 good chips is

$$n_{\overline{A}} = \binom{75}{12}$$

The overall number of combinations of 12 good chips from a total of 100 (good and defective) chips is

$$n_s = \binom{100}{12}$$

The probability of the event that there is no defective chip is

$$P\left[\overline{A}\right] = \frac{n_{\overline{A}}}{n_s} = \frac{75!}{12!63!} \times \frac{12!88!}{100!} = \frac{75!88!}{63!100!}$$

The probability of the event that there is at least one good chip is the complement,

$$P\left[A\right] = 1 - P\left[\overline{A}\right]$$

Example: Lotto 649

3.5 Method 5: Distinguishable Permutations

Definitions: Consider a box containing n objects of which k_1 are of one kind, k_2 are of a second kind, and so on, k_K are of a K^{th} kind $\left(\text{let } n = k_1 + k_2 + \cdots + k_K\right)$.

- Each k_i group is defined as a partition (so there are K partitions)
- In each partition, the objects are indistinguishable
- Objects in different partitions are distinguishable

- To determine the overall number of distinguishable permutations of the n objects, we need to first compute the number of combinations $n_1, n_2, \ldots, n_K$, for the individual partitions, respectively. Since the elements in each group are indistinguishable, order does not matter so we use combinations for each partition as follows:

$$n_1 = \binom{n}{k_1};\ n_2 = \binom{n-k_1}{k_2};\ n_3 = \binom{n-k_1-k_2}{k_3};\ \ldots;\ n_K = \binom{n-k_1-k_2-\cdots-k_{K-1}}{k_K}.$$

- We then apply the multiplication rule to yield a total of distinguishable permutations as

$$n_S = n_1 \times n_2 \times \ldots \times n_K .$$

Example: A box contains a total of n fruits; k_1 of which are oranges and k_2 are apples $\left(n = k_1 + k_2\right)$. How many distinguishable permutations are there?

1. Number of combinations in Partition A_1 (k_1 oranges of n fruits) is

$$n_1 = \binom{n}{k_1} = \frac{n!}{k_1!\left(n-k_1\right)!}$$

2. Number of combinations in Partition A_2 (k_2 apples of n fruits) is

$$n_2 = \binom{n-k_1}{k_2} = \frac{\left(n-k_1\right)!}{k_2!\left(n-k_1-k_2\right)!}$$

3. Total number of distinguishable permutations is

$$n_S = n_1 \times n_2 = \frac{n!}{k_1!\left(n-k_1\right)!} \times \frac{\left(n-k_1\right)!}{k_2!\left(n-k_1-k_2\right)!} = \frac{n!}{k_1!k_2!}$$

Conclusion: For a box containing k_1 identical objects and k_2 identical objects of another kind, the total number of distinguishable permutations can be computed as

$$n_S = \frac{n!}{k_1!k_2!}$$

Generalization:

Consider a box containing n objects comprising K distinguishable partitions $A_1, A_2, \ldots A_K$.

- Partition A_1 with k_1 identical elements
- Partition A_2 with k_2 identical elements
- And so, on
- Partition A_K, with k_K identical elements

The total number of distinct permutations or number of ways to arrange a set of n elements involving K sets of sizes $k_1, k_2, \ldots, k_K$, respectively, is

$$n_S = \underbrace{\binom{n}{k_1}}_{\text{ways to place objecs of type1}} \times \overbrace{\binom{n-k_1}{k_2}}^{\text{ways to place objecs of type2}} \times\cdots\times \underbrace{\binom{n-k_1-k_2-\cdots-k_{K-1}}{k_K}}_{\text{ways to place objecs of type k}} = \frac{n!}{k_1!k_2!\cdots k_k!};$$

$$(n-k_1-k_2\cdots-k_{K-1}-k_K)! = 0! = 1$$

Therefore, the probability of picking one permutation or choice is

$$P[\text{one choice}] = \frac{1}{n_S} = \frac{k_1!k_2!\cdots k_K!}{n!}$$

Exercise: A lab technician wants to assign 7 students to 2 double-seat benches and 1 triple-seat bench during an electronic circuit lab session. In how many was can these students be seated?

Solution: Each double-seat can take two students, therefore, $n_1 = n_2 = 2$ and one triple-seat can take 3 students, therefore, $n_3 = 3$. The total number of students to be seated is $n = 7$. The total number of possible partitions is, therefore, equal to

$$n_s = \frac{n!}{n_1!n_2!n_3!} = \frac{7!}{2!2!3!} = 210$$

Exercise: In a bag, we have $k_1 = 1$ red ball, $k_2 = 2$ green balls, $k_3 = 3$ blue balls. What is the number of distinguishable permutations? What is the probability of one permutation?

Exercise: A batch of 50 chips contains 10 defective chips. Suppose 12 are selected at random and tested. What is the probability of selecting exactly 5 defective chips?

<u>Example:</u> How many strings can be formed by permuting the letters in SUCCESS?

<u>Solution:</u> The word SUCCESS contains

$$3\times S \to 3 \text{ indistinguishable letters} \to k_1 = 3$$
$$1\times \text{U} \to 1 \text{ indistinguishable letter} \to k_2 = 1$$
$$2\times C \to 2 \text{ indistinguishable letters} \to k_3 = 2$$
$$1\times E \to 1 \text{ indistinguishable letter} \to k_4 = 1$$

The total number of strings (or distinct permutations) is

$$n_S = \frac{7!}{3!1!2!1!} = 420$$

<u>Example:</u> The captain of a ship sends signals by arranging 4 orange flags and 3 blue flags on a vertical pole. How many different signals could the captain possibly send?

<u>Solution:</u> The four orange flags are not distinguishable among themselves, and the three blue flags are not distinguishable among themselves. We need to count the number of distinguishable permutations when the two colors are the only features that make the flags distinguishable. The number of possible signals is

$$n_s = \frac{7!}{4!3!} = \frac{7\times 6\times 5}{3\times 2} = 35$$

<u>Example:</u> A box contains 2 black balls and 3 white balls. Two balls are selected at random from the box.

(a) What is the probability that they are both black?
(b) If 3 balls are selected, what is the probability that two are black and the third is white?

<u>Solution:</u> - Use of partitions. Partition 1 contains 2 black balls ($n_1 = 2$) and Partition 2 contains 3 white balls ($n_2 = 3$). The total number of balls is $n = 5$.

(a) Overall experiment: pick 2 balls from a total $n = 5$,

$$n_s = \binom{n}{k} = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

- Event A: pick 2 black balls from Partition 1 and 0 black balls from Partition 2

$$n_A = n_1 \times n_2 = \binom{2}{2}\binom{3}{0} = 1$$

- The probability of selecting 2 black balls, both black, is

$$P\left[\{\text{selecting 2 black balls}\}\right] = \frac{n_A}{n_S} = \frac{1}{10}$$

.

(b)When 3 balls are selected at random from the total $n = 5$,

$$n_s = \binom{n}{k} = \binom{5}{3} = \frac{5!}{3!2!} = 10$$

When 2 of the selected balls are black 1 has to be from 3 white balls. Therefore,

$$n_A = \binom{2}{2}\binom{3}{1} = 3$$

The probability of selecting 2 black balls and 1 white ball is

$$P\left[\{\text{selecting 2}\}\right] = \frac{n_A}{n_S} = \frac{3}{10}.$$

<u>Concluding Remarks on Permutation and Combinations:</u>

The main difference in the definition of a permutation and a combination is whether order is important.

- Permutation: order is important
- Combination: order is not important
- For distinguishable partitions, we need to compute the number of combinations within each partition and apply the multiplication rule to obtain the overall permutation

4. Discrete Random Variables and Probability Distributions:
(Reading Exercises: Montgomery and Runger Section 2.9, 3.1-3.2, 3.4-3.8)

Learning outcomes:

You will be able to
- Define a discrete random variable.
- Define a discrete probability mass function and a discrete cumulative distribution function.
- Describe the important discrete random variables and use probability mass function to find probabilities for practical situations.
- Apply the material learned in this section to new problems.

Definition:
Given a random experiment with sample space S, a random variable is a function that associates a unique real number to each outcome (or element) that belongs in the sample space S.
- A random variable is denoted by upper case letters, for example, X
- The numerical value that a random variable take is denoted by lower case, for example, $X = x$.

Definition:
Let S be a sample space. A discrete random variable is one that takes on either
- A finite number of values $x_1, x_2 \ldots, x_n$ or
- An infinite number of values $x_1, x_2 \ldots$ from S.

Example:
A coin is tossed 3 times and the sequence of heads (H) and tails (T) is noted. The sample space comprises outcomes, which are all possible combinations of heads and tails, that is,

$$S = \{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT\}$$

Denote an outcome by the symbol ξ and let random variable X denote the number of heads in the three tosses. The random variable, X, maps each outcome onto a real number. We see that X can take on four values as illustrated in the table below.

ξ	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\xi)$	3	2	2	2	1	1	1	0

$$X : \xi \Rightarrow x \in \{0, 1, 2, 3\}$$

4.1 Probability Mass Function (PMF)

4.1.1 Characterization of Discrete Random Variables:

Random variables are characterized by their probability distributions. For discrete random variables the probability distribution functions are referred to as Probability Mass Functions.

Definition: Consider a discrete random variable X. The probability that X takes on a particular value, $X = x$, is denoted $p_X(x)$, that is,

$$p_X(x) = P[X = x]$$

Suppose that X can take on any value from a set $x \in \{x_1, x_2, \cdots\}$. The probability mass function (PMF) is the collection of the probabilities of all the values X can take, that is,

$$p_X(x) = P[X = x],\ \forall\ x \in \{x_1, x_2, \cdots\}$$

Properties of PMF: Since $p_X(x)$ is a probability, it must satisfy the following axioms:

(i) $p_X(x) \geq 0,\ \forall\ x \in \{x_1, x_2, \cdots\}$

(ii) $\sum_x p_X(x) = 1$

(iii) $p_X(x) = 0$ if $x \notin \{x_1, x_2, \cdots\}$.

Example: A shipment of 20 RAM chips from a company to a lab contains 3 that are defective. If 2 of these chips are randomly picked, find the PMF for the number of defective RAM chips.

Solution: Define random variables X and Y, and events $A_1 = \{X \text{ defective chips are selected}\}$; $A_2 = \{Y = 2 - X \text{ non-defective chips are selected}\}$. The values the random variables can take are

$$X = x \in \{0,1,2\};\ \ Y = y = 2 - x \in \{0,\ 1,\ 2\}$$

For event A_1 (or Partition 1), the number of possible ways x chips can be selected from 3 defective chips is the combination $n_1 = \binom{3}{x},\ x = 0,\ 1,\ 2.$

For event A_2 (or Partition 2), the number of possible ways y chips can be selected from 17 non-defective chips is the combination $n_2 = \binom{17}{y} = \binom{17}{2-x},\ x = 0,\ 1,\ 2.$

The experiment involves picking 2 chips out of 20. The number of elements of the sample space is, therefore, the combination $n_S = \binom{20}{2}$.

The PMF of the random variable X, is

$$p(x) = P[X = x] = \frac{n_1 \times n_2}{n_S} = \frac{\binom{3}{x}\binom{17}{2-x}}{\binom{20}{2}}$$

$$p(0) = P[X = 0] = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95} = \frac{136}{190}$$

$$p(1) = P[X = 1] = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$$

$$p(2) = P[X = 2] = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$$

Check:

$$p(0) + p(1) + p(2) = 1$$

The PMF (probability distribution) can be represented in tabular form as

x	0	1	2
$p(x)$	$\frac{136}{190}$	$\frac{51}{190}$	$\frac{3}{190}$

4.2 Cumulative Distribution Function (CDF) for Discrete Random Variable:

Definition: The Cumulative Distribution Function (CDF), denoted $F_X(x)$, of a discrete random variable, X, with probability mass function $p_X(x)$, is a function defined as the probability that X does not exceed a value x, that is,

$$F_X(x) = P[X \le x] = \sum_{y=-\infty}^{x} p_X(y), \quad -\infty < x < \infty$$

Properties of the CDF:

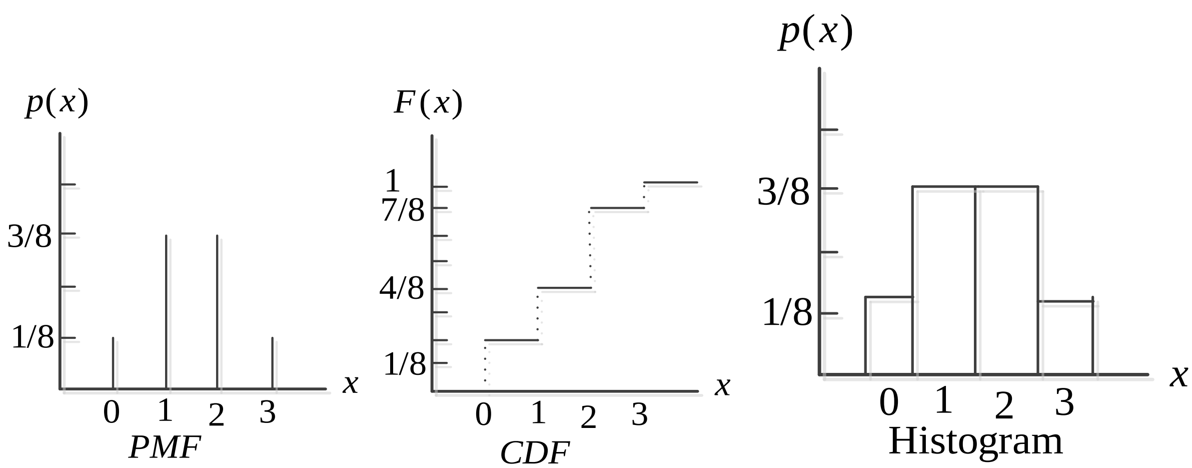
(1) $F(-\infty)=0 \Rightarrow$ impossible and $F(\infty)=1 \Rightarrow$ certain ,

(2) $F(x)$ is monotonically non-decreasing in x.

Example: The table below represents the PMF and the CDF of random variable X , representing the number of heads in 3 tosses of a fair coin. There are 8 events in total.

Random variable	$x=0$	$x=1$	$x=2$	$x=3$
PMF	$p_X(0)=1/8$	$p_X(1)=3/8$	$p_X(2)=3/8$	$p_X(3)=1/8$
CDF	$F_X(0)=1/8$	$F_X(1)$	$F_X(2)$	$F_X(3)$

The PMF and CDF can be represented graphically as illustrated in the figures below.



Graphical methods of representing probability distributions

Example: Find the CDFs, $F_X(0)$ and $P[-3 \le X \le 1]$, for the PMF of random variable X in the table below.

$X=x$	-8	-3	-1	0	1	4	6
$p_X(x)$	0.13	0.15	0.17	0.20	0.15	0.11	0.09

$$F_X(0)=P[X \le 0]=0.13+0.15+0.17+0.20=0.65$$
$$P[-3 \le X \le 1]=0.15+0.17+0.20+0.15=0.67$$

Example: The table below provides the PMF and CDF of 2 dice rolls when the random variable X represents the sum of the two numbers that face up.

$$p_X(x) = \frac{\text{\# of ways 2 dices can sum to } x}{\text{\# of all possible ways 2 dices can result in}}$$

$$F_X(x) = \sum_{i=2}^{x} p_X(i)$$

Probability Mass Function & CDF of two dice roll

x	$p_X(x)$	$F_X[x]$
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36

4.3 Important Discrete Random Variables and their Distribution Functions:

Counting Processes:

There are three counting processes – (1) Binomial random variables, (2) Geometric random variable and (3) Poisson random variables. Fundamental to counting processes is the Bernoulli trial and Bernoulli random variable.

Bernoulli trial: (Daniel Bernoulli – 1700-1782, Swiss Mathematician and Physicist)

Definition: A Bernoulli trial is a single random experiment in which there can be only two outcomes (success or failure). Probability of success is $0 < p < 1$ and of failure is $q = 1 - p$.

Examples of Bernoulli Trials: (1) Tossing of a fair coin (outcomes are H or T) and (2) classification of products as defective or non-defective (outcomes are D or N).

Bernoulli random variable:

Definition: A Bernoulli random variable, X, is a random variable whose only possible values are 1 and 0, and whose probability mass density function is

$$p_X(x) = p^x (1-p)^{1-x}, \quad x \in \{0,1\}$$
$$p_X(0) = 1 - p$$
$$p_X(1) = p$$

- We denote the distribution as $Ber\left(p\right)$.

Binomial Random Variable:

Consider the following:

1. An experiment is performed, in the same way, n times,
2. Each experiment is a Bernoulli trial,
3. The trials generate n independent Bernoulli random variables,
4. The probability of success, denoted p, is the same for each trial. The probability of failure is $q = 1 - p$.

Definition: A Binomial random variable, X, is the number of successes (or failures) in n independent Bernoulli trials with p the probability of success and q the probability of failure at each trial. Let X_i denote the outcome at the ith trial, then the number of successes is the summation

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$\begin{aligned} X_i &= 1 \rightarrow \text{success in the i}^{th}\text{ experiment} \\ &= 0 \rightarrow \text{success in the i}^{th}\text{ experiment} \end{aligned}$$

We note the following:

- The number of ways of arranging x successes in n trials is the combination $\binom{n}{x}$
- All arrangements of x successes and $(n-x)$ failures are equally likely with probability $p^x(1-p)^{n-x}$
- The PMF is, therefore, given by (all the combinations)

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

- The above PMF is known as a Binomial distribution and is denoted as $Bin(n, p)$ with parameters n and p, where $n =$ 1,2, . . . with $0 \le p \le 1$.
- The CDF of a Binomial random variable is given by

$$F_X\left(x\right) = P\left[X \le x\right] = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y} \quad x = 0,1,\ldots,n$$

Example: A manufacturing process results in probability $p = 0.05$ of generating defective products. If we select 3 products in succession from the manufacturing process and inspect them, what are the probabilities for the different numbers of defective products?

Solution: Each product selection is a Bernoulli trial because it has two outcomes: defective $(D-\text{failure})$ and non-defective $(N-\text{success})$. The number of failures (or successes) is a Bernoulli process because

(1) There are 3 trials
(2) Each trial is a Bernoulli trial
(3) The probability of defective product, $p = 0.05$, is the same for each trial.
(4) The 3 trials are independent. The fact that the first product is D or N has no bearing on the second.

The following table summarizes all the possible probabilities:

Outcome	x	$P[X]$
NNN	0	$(1-0.05)^3$
NND	1	$(1-0.05)^2 \times 0.05$
NDN	1	$(1-0.05)^2 \times 0.05$
DNN	1	$(1-0.05)^2 \times 0.05$
DND	2	$(1-0.05) \times 0.05^2$
NDD	2	$(1-0.05) \times 0.05^2$
DDN	2	$(1-0.05) \times 0.05^2$
DDD	3	0.05^3

Now, applying the formula $p_X(x) = P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0,\ 1,\ 2,\ 3$, we obtain

$$p_X(0) = \binom{3}{0} p^0 (1-p)^{3-0} = (1-0.05)^3$$
$$p_X(1) = \binom{3}{1} p^1 (1-p)^{3-1} = 3(1-0.05)^2 \times 0.05$$
$$p_X(2) = \binom{3}{2} p^2 (1-p)^{3-2} = 3(1-0.05) \times 0.05^2$$
$$p_X(3) = \binom{3}{3} p^3 (1-p)^{3-3} = 0.05^3$$

These are the same results as in the table above.

Exercise: Consider an experiment where a coin tossed three times. Evaluate the PMF and CDF of the number of heads.

Geometric Random Variable:

Definition: A geometric random variable X is the number of trials of any experiment (comprising n independent Bernoulli trials) until the first success. Therefore, the PMF of a geometric random variable X is given by

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

This PMF, denoted by $Geo(p)$, is known as a geometric distribution with parameter $p,\ 0 < p \le 1$.

Consider a sequential experiment in which we repeat independent Bernoulli trials until the occurrence of the 1st success. Let random variable X denote the number of trials until the 1st success. Let $p \triangleq P\left[\text{'}1^{st}\text{success after } x \text{ trials}\right]$. Then, the PMF is

$$\begin{aligned}
p_X(x) &\triangleq P\left[\text{'}1^{st}\text{success after } x \text{ trials}\right] \\
&= P\left[\text{'failure in } (x-1) \text{ previous trials and success at the } x^{th} \text{ trial}\right] \\
&= P\left[\text{'fail at } 1^{st}\right] P\left[\text{'fail at } 2^{nd}\right] \cdots P\left[\text{'fail at } (x-1)^{th}\right] P\left[\text{'fail at } x^{th}\right] \\
&= (1-p)(1-p)\cdots(1-p)\, p \\
&= (1-p)^{x-1} p \equiv \text{Geometric Probability Law}
\end{aligned}$$

Example: Let X denote the number of times a paging message needs to be transmitted until success.

(a) Find the PMF of $X \in \{1, 2, \ldots, m\}$.

(b) Find the probability that X is an even number.

Solution: (a) Denote events F and S as failure and success, respectively. The PMF is

$$p_X(m) = P[X = m] = P[F \cap F \cap \cdots \cap F \cap S] = P[F]P[F]\cdots P[F]P[S] = (1-p)^{m-1} p.$$

Solution: (b)

$$\begin{aligned}
P[X \text{ is even}] &= P[X = 2] + P[X = 4] + \cdots = \sum_{k=1}^{\infty} p(2k) = p\sum_{k=1}^{\infty}(1-p)^{2k-1} \\
&= p\sum_{k=1}^{\infty}(1-p)^{2k-2+1} = p(1-p)\sum_{k=1}^{\infty}\left((1-p)^2\right)^{k-1} \\
&= p\frac{1-p}{1-(1-p)^2} = \frac{1-p}{2-p}
\end{aligned}$$

Memoryless Property:

Theorem: A geometric distribution has the following memoryless property for all non-negative integers m and n,

$$P\left(X \geq m+n \mid X \geq m\right) = P\left[X \geq n\right]$$

Proof:

$$\begin{aligned} P\left[X \geq m+n \mid X \geq m\right] &= \frac{P\left[\{X \geq m+n\} \cap \{X \geq m\}\right]}{P\left[X \geq m\right]} \\ &= \frac{P\left[X \geq m+n\right]}{P\left[X \geq m\right]} = \frac{(1-p)^{m+n}}{(1-p)^{m}} \\ &= (1-p)^{n} \\ &= P\left[X \geq n\right] \end{aligned}$$

Interpretation:
Suppose you are told that there have been m failures initially. Then the chance of at least n more failures before the first success; is exactly the same as if you started the experiment for the first time and the information of initial m failures is irrelevant.

Example: What is the probability that more than K trials are required before 1st success?

Solution:

$$\begin{aligned} P\left[\{i > K\}\right] &= P\left[\{K+1 \text{ or } K+2 \text{ or } \cdots\}\right] \\ &= P\left[K+1\right] + P\left[K+2\right] + \cdots = \sum_{m=1}^{\infty} P\left[K+m\right] \\ &= \sum_{m=1}^{\infty} (1-p)^{K+m-1} p = p(1-p)^{K} \sum_{m=1}^{\infty} (1-p)^{m-1} \\ &= \frac{p(1-p)^{K}}{1-(1-p)} = (1-p)^{K} \end{aligned}$$

Poisson Random Variables

Consider events occurring independently of each other, with an average number of events λ, in some fixed interval Δt.

Definition: A random variable X, is said to be a Poisson RV if it counts the number of events that occur, "completely at random", with a mean value λ in a specified time interval Δt.

Examples: A Poisson random variable can be used to model the following events:

- The number of calls arriving at a phone exchange centre in a specified time interval,
- The number of data packets arriving at a router/server in a specified time interval,
- The number of multipath arriving at a wireless radio receiver, in a specified time interval
- The number of photons arriving at a CCD pixel in some exposure time (astronomy observation)
- Number of customers arriving at a checkout counter, in a specified time interval
- The number of accidents occurring at an intersection, in a specified time interval.

PMF of Poisson RV

- Let X represent the number of events,
- α rate of arrival (average number of arrivals per unit time)
- Δt the time interval of interest
- $\lambda = \alpha \times \Delta t$ is the mean or average number of events in an interval of interest.

Definition: A random variable X has a Poisson distribution, with parameter $\lambda > 0$, if its PMF is given by

$$P[X = x] = \frac{\lambda^x}{x!} e^{-\lambda}, \; x = 0, 1, 2, \ldots$$

Example: The number of queries arriving in a Δt seconds interval, at a call center is a Poisson RV with an average of 4 queries per minute.

(a) Find the probability of more than 4 queries in 10 seconds.
(b) Find the Probability of less than 5 queries in 2 minutes.
(c) Find the Probability of exactly 4 queries in 2 minutes.

Solution:

(a) $\Delta t = 10\text{s}; \quad \alpha = \dfrac{4 \text{ querries}}{\min} = \dfrac{4}{60} \dfrac{\text{querries}}{s}; \quad \Rightarrow \lambda = \dfrac{4}{60} \times 10 = \dfrac{2}{3}$

$$\begin{aligned}
P[X > 4] &= 1 - P[X \le 4] \\
&= 1 - P[X = 0 \text{ or } X = 1 \text{ or } X = 2 \text{ or } X = 3 \text{ or } X = 4] \\
&= 1 - \left(P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4]\right) \\
&= 1 - \left(\frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \frac{\lambda^2}{2!} e^{-\lambda} + \frac{\lambda^3}{3!} e^{-\lambda} + \frac{\lambda^4}{4!} e^{-\lambda}\right) \\
&= 1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \frac{\lambda^4}{24}\right) = 6.33 \times 10^{-4}
\end{aligned}$$

(b) $\lambda = \alpha \times \Delta t = \left(\frac{4}{60}\right) \times (2 \times 60) = 8$

$$P[X < 5] = P[X \le 4] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4]$$

$$= e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \frac{\lambda^4}{24}\right) = 0.1$$

(c) $P[X = 4] = e^{-8}\frac{8^4}{24} = 0.06$

Example: Data packets arrive at a multiplexer at random and at an average rate of 1.2 per second.

(a) Find the probability of 5 messages arriving in a 2-seconds interval.

(b) For how long can the operation of the multiplexer be interrupted, if the probability of losing one or more packets is to be no more than 0.05?

Solution: Times of arrival form a Poisson process with rate $\alpha = \frac{1.2}{s}$.

(a) Let X denote the number of messages arriving in a $\Delta t = 2\,\text{sec}$ interval. Then X is Poisson with mean number $\lambda = \alpha \times \Delta t = 1.2 \times 2 = 2.4$

$$P[X = 5] = \frac{\lambda^5}{5!}e^{-\lambda} = \frac{2.4^5}{5!}e^{-2.4} = 0.06$$

(b) Let the number of messages be denoted by random variable Y. Y is Poisson with $\lambda = \alpha \times \Delta t = 1.2 \times \Delta t$. We need to find time interval Δt.

$$P[\{\text{at least one message}\}] = P[Y \ge 1] = 1 - P[Y = 0] = 1 - \frac{2.4^0}{0!}e^{-1.2\times\Delta t} \le 0.05$$

$$\Rightarrow e^{-1.2\times\Delta t} \ge 0.95 \Rightarrow -1.2 \times \Delta t \ge \ln(0.95) = -0.05129$$

$$\Rightarrow \Delta t = 0.043\,\text{sec}$$

Discrete Uniform Distribution:

Definition: A discrete uniform random variable X, is one that is equally likely to take any integer value in a finite interval $x \in [k,\ l]$,

- X has a PMF defined as

$$p_X(x) = P[X = x] = \begin{cases} \frac{1}{l - k + 1}, & x = k, k+1, k+2, \ldots, l \\ 0, & \text{otherwise} \end{cases}$$

- A uniform random variable X, has a constant PMF over a finite range.

Useful Series in Probability:

$\sum_{k=1}^{n} q^{k-1} = \frac{1-q^n}{1-q}, \quad (q>0)$	$\sum_{k=1}^{\infty} q^{k-1} = \frac{1}{1-q}, \quad (0<q<1)$	$\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}, \quad (0<q<1)$
$\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3}, \quad (0<q<1)$	$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$	$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$
$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$	$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n$ Binomial Theorem	$\sum_{x=k}^{l} \frac{x}{l-k+1} = \frac{l+k}{2}$

4.4 Mathematical expectation of discrete random variables:

(Reading Exercises: Yates and Goodman –Section 3.5-3.7)

Learning Outcomes:

You will be able to
- Define the expected value of a discrete random variable
- Define the expected value of a function of a discrete random variable.
- Apply the properties of mathematical expectations.
- Derive a formula for the mean of the special random variables.
- Define the variance and standard deviation of a discrete random variable.
- Apply a shortcut formula for the variance of a discrete random variable.
- Calculate the mean and variance of a linear function of a discrete random variable.
- Understand the steps involved in each of the proofs in the lesson.
- Apply the methods learned in this section to new problems.

Definition: Let X be a discrete random variable with a possible set of values $S = \{x_1, x_2, \ldots\}$ and PMF $p_X(x)$. The expected value or mean or average value of X, denoted $E[X]$ or μ_X is the weighted average defined as

$$\mu_X = E[X] = \sum_{x \in S} x p_X(x)$$

- The sum in the definition above, is known as the mathematical expectation of X.
- The expected value of X is also referred to as the first moment of X.
- μ_X is a measure of location of the PMF

Example: What is the average toss of a fair, six-sided die?

Example: Let random variable X denote the number of credit cards owned by an electrical engineering student. Using the data in the table below, find the expected number of credit cards a student will possess.

x	0	1	2	3	4	5	6
$p_X(x)$	0.08	0.28	0.38	0.16	0.06	0.03	0.01

Example: Find the mean of a Binomial random variable X, with parameters n and p.

Solution:

$$E\{X\} = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{i=0}^{n} E[x_i] = np$$

Example: Find the mean of a geometric RV, X, with $p = P[\text{Success}]$.

Solution: Use identity $\sum_{i=1}^{\infty} iq^i = \frac{q}{(1-q)^2}$

$$E[X] = \sum_{x=1}^{\infty} xp_X(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p = p\frac{d}{dp}\sum_{k=1}^{\infty}(1-p)^x = p\frac{d}{dp}\left(\frac{1}{1-(1-p)}\right) = \frac{1}{p}$$

Example: Find the mean of a discrete uniform random variable X. Using the series identity, we obtain

$$E(X) = \sum_{x=k}^{l} x\frac{1}{l-k+1} = \frac{k+l}{2}$$

Example: A quality control engineer is inspecting a batch of 7 electronic components. The batch consists of 4 good components and 3 defective components. The engineer takes a sample of 3 components. (a) Find the PMF of the good components and (b) Find the expected value of the number of good components in this sample.

Solution: Use partitions good and defective. Let X represent the number of good components in the sample. There are 2 distinguishable sets with two combinations: x out of 4 good components and $3-x$ out of 3 defective components.

- For set 1, $n_1 = \binom{4}{x}$ possible number of combinations,
- For set 2, $n_2 = \binom{3}{3-x}$ possible number of combinations.
- The total number of possible combinations is $n = \binom{7}{3}$. The probability distribution (PMF) of X is

$$p_X(x) = \frac{n_1 \times n_2}{n} = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \quad x = 0,1,2,3$$

- The result is provided in the table below.

x	0	1	2	3
$p(x)$	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

- The expected value of X is

$$\mu_X = 0\times\left(\frac{1}{35}\right)+1\times\left(\frac{2}{35}\right)+2\times\left(\frac{18}{35}\right)+3\times\left(\frac{4}{35}\right)=\frac{17}{7}=1.7$$

- This implies that if a sample size of 3 components is taken over and over again, there will be on average, 1.7 good components.

Expected Value of a Function of a Discrete Random Variable:

Definition: consider a discrete random variable X, with probability mass density function $p_X(x)$. The expected value of a function $h(X)$, of random variable X is defined as

$$\mu_{h(X)} = E\left[h(X)\right] = \sum_{x\in S} h(x)\, p_X(x)$$

Example: Suppose the number of cars passing through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution (PMF):

x	4	5	6	7	8	9
$p_X(x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $h(x) = 2X - 1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected (or average) earnings for this particular time period.

4.5 Properties of Mathematical Expectations

Property 1: Consider a random variable X with PMF $p_X(x)$. Then the expectation of a function $g(X)$, of X is calculated by the following formula:

$$E\left[g(X)\right] = \sum_{x\in Range(X)} g(x)\, p_X(x)$$

Property 2: Consider a random variable X with expectation $E[X]$. Consider a linear transformation $Y = aX + b$, of X, where a and b are two constants. Then the expectation of Y is given by

$$E[Y] = aE[X] + b$$

Examples:

4.6 Variance and Standard Deviation of Discrete Random Variables:

(Reading Exercises: Yates and Goodman Section 3.8)

Definition: Consider a discrete random variable, X, with PMF $p_X(x)$ and expected value μ_X. The variance of X, denoted $Var[X]$ or σ_X^2 is defined as

$$\sigma_X^2 = E\left[\left(X-\mu_X\right)^2\right] = \sum_{x\in S}\left(x-\mu_X\right)^2 p_X\left(x\right)$$

Definition: The standard deviation of a random variable X is the square-root of the variance,

$$\sigma_X = \sqrt{E\left[\left(X-\mu_X\right)^2\right]} = \sqrt{\sigma_X^2}\ .$$

- The variance of a random variable is the second central moment of X
- The variance (or the standard deviation) of a random variable characterizes the variability or spread in the distribution of the random variable
- It gives a description of the shape of the distribution.

Example: The quiz scores for a particular student are 22, 25, 20, 18, 12, 20, 24, 20, 20, 25, 24, 25 and 18. Find the variance and standard deviation.

Value	12	18	20	22	24	25
Frequency	1	2	4	1	2	3
Probability	0.08	0.15	0.31	0.08	0.15	0.23

Shortcut formula for variance:

$$\sigma_X^2 = E\left[\left(X-\mu_X\right)^2\right] = E\left[X^2\right] - 2\mu_X E\left[X\right] + \mu_X^2 = E\left[X^2\right] - \mu_X^2$$

$E\left[X^2\right]$ is the Second Moment of X ; also called the mean-square value of X

Example:

Variance of a Function of a Discrete RV

Definition: Consider a discrete random variable X , with probability mass function $p_X\left(x\right)$. Let $Z = h\left(X\right)$ be a function of X with expected value $\mu_{h(X)}$. The variance of the discrete random variable $Z = h\left(X\right)$ is defined as

$$\sigma_{h(X)}^2 = E\left[\left(h\left(X\right)-\mu_{h(X)}\right)^2\right] = \sum_{x\in S}\left(h\left(X\right)-\mu_{h(X)}\right)^2 p_X\left(x\right)$$

Examples: Calculate the variance of $h\left(X\right) = 2X+3$, where X is a random variable with probability distribution

x	4	5	6	7	8	9
$p\left(x\right)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Solution:

$$\mu_{h(X)} = E\left[h(X)\right] = \sum_{x=0}^{3}(2x+3)p_X(x) = 6$$

$$\sigma_{h(X)}^2 = E\left[\left(h(X)-\mu_{h(X)}\right)^2\right] = \sum_{x=0}^{3}\left(h(X)-\mu_{h(X)}\right)^2 p_X(x) = \sum(2x+3-6)^2 p_X(x) = 4$$

Example: A Square-Law Device (diode) has input noise voltage X, that is uniformly distributed with values in the set $S=\{-3,-1,+1,+3\}$. The output of a square-law device is $Z = X^2$. Find $E[Z]$.

Solution: Since X is uniform we can assume the values are equally likely, therefore, the PMF is $p_X(x)=\frac{1}{4}$. Then

$$E[Z] = \sum_{x=-3}^{3} x^2 p_X(x) = (-3)^3\frac{1}{4}+(-1)^2\frac{1}{4}+(1)^2\frac{1}{4}+(3)^3\frac{1}{4} = 5$$

Exercise: Suppose the output of the square-law device is now $Z=(2X+10)^2$. Show that $E[Z]=120$

Properties of Variance:

Consider a discrete random variable, X, with variance σ_X^2. Consider a linear transformation $Y = aX + b$, of X, where a and b are two constants. Then the variance of Y is given by

$$\sigma_Y^2 = a^2\sigma_X^2$$

- The variance of a constant $Y=b$, is $\sigma_Y^2 = 0$.

Examples:

Exercises:

5. Continuous RVs and Probability Density Function:
(Reading Exercises: Montgomery and Runga Chapter 4)

You will

- Understand the concept of a probability density function of a continuous random variable.
- Know the formal definition of a probability density function of a continuous random variable.
- Know how to find the probability that a continuous random variable takes on values in some interval.
- Know the formal definition of a cumulative distribution function of a continuous random variable.
- Know how to find the cumulative distribution function of a continuous random variable from its probability density function.
- Know how to extend the definitions of the mean, variance and standard deviation to functions of a continuous random variable.
- Be able to apply the methods learned in the section to new problems.

5.1 Probability Distribution Function

Definition: A random variable is continuous if its set of possible values belong to an entire interval of numbers.

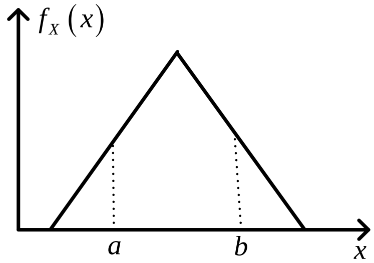
Definition: A continuous random variable is said to have a probability density function (PDF), denoted $f_X(x)$, if

- $f_X(x) \geq 0$ (non-negative) for all $x \in \mathbb{R}$ (all real numbers x)
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- $P[a < x < b] = P[a < x \leq b] = P[a \leq x < b] = P[a \leq x \leq b] = \int_a^b f_X(x)\,dx$

Remarks:

- A continuous random variable is described by a probability density function
- With continuous random variables we talk about the probability of X being in some interval, $a \leq x \leq b$

$$P[a \leq X \leq b] = \int_a^b f_X(x)\,dx$$



- The probability $P[a \le X \le b]$, is the area under the PDF in the interval $a \le X \le b$.
- A continuous random variable has a probability of zero, to assume exactly any one of its values, $P[X = a] = 0$.

Examples: Suppose the error in the reaction temperature, in $^{\circ}C$, for a controlled laboratory experiment is a continuous random variable X having the PDF

$$f_X(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Verify that $f_X(x)$ is a valid pdf and then find $P[0 < X \le 1]$.

Solution:

It is a valid PDF because $f_X(x) \ge 0$ for all $x \in [-1,\ 2]$ and the area underneat it is equal to 1. The probability is

$$P[0 < X \le 1] = \int_0^1 \frac{x^2}{3} dx = \left. \frac{x^3}{9} \right|_0^1 = \frac{1}{9}$$

5.2 Cumulative Distribution Function

Definition: The cumulative distribution function (CDF), denoted $F_X(x)$, of a continuous random variable X, with pdf $f_X(x)$ is defined as the probability that X is less than or equal to a given value x, that is,

$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(y)\, dy$$

- For every x, $F_X(x)$ is the area under the PDF curve to the left of x.
- If the derivative of $F_X(x)$ bexists, then the PDF is related to the CDF as follows:

- $f_X(x) = \dfrac{dF_X(x)}{dx}$

Properties of CDF:

- $F_X(x)$ is a non-decreasing function from left to right,
- $F_X(x)$ is a continuous function,
- $F_X(\infty)=1$,
- $F_X(-\infty)=0$,
- $P[a < X \le b] = F_X(b) - F_X(a)$
- $P[X > x] = 1 - P[X \le x] = 1 - F_X(x)$

Example: A random variable has the probability density function

$$f_X(x) = \begin{cases} x^2/3, & -1 < x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the CDF $F_X(x)$, and evaluate the probability $P[0 < X \le 1]$.

Solution: For the interval $-\infty < X < -1$ $F_X(x) = 0$. For the interval $-1 \le X < 2$, we have

$$F_X(x) = \int_{-1}^{x} \frac{y^2}{3} dy = \left.\frac{y^3}{9}\right|_{-1}^{x} = \frac{x^3+1}{9}$$

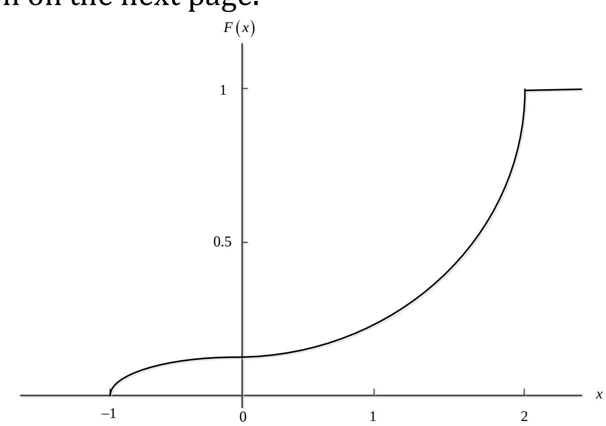
For $X \ge 2$,

$$F_X(x) = \left.\frac{x^3+1}{9}\right|_{x=2} = 1$$

Therefore,

$$F_X(x) = \begin{cases} 0, & x < -1 \\ \dfrac{x^3+1}{9}, & -1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

A rough sketch is shown on the next page.



$$P[0 < X \le 1] = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

Example: The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the pdf of the winning (low) bid is

$$f_X(x) = \begin{cases} \dfrac{5}{8b}, & \dfrac{2}{5}b \le x \le 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the CDF $F_X(x)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Solution: For $\frac{2}{5}b \le x < 2b$,

$$F_X(x) = \int_{\frac{2b}{5}}^{x} \frac{5}{8b}\,dy = \left.\frac{5y}{8b}\right|_{\frac{2b}{5}}^{x} = \frac{5x}{8b} - \frac{1}{4} \Rightarrow$$

For the interval $x < \frac{2}{5}b$, $F_X(x) = 0$ and for $x \ge 2b$, $F_X(x) = 1$. Therefore,

$$F_X(x) = \begin{cases} 0, & x < \dfrac{2}{5}b \\ \dfrac{5x}{8b} - \dfrac{1}{4}, & \dfrac{2}{5}b \le x < 2b \\ 1, & x \ge 2b \end{cases}$$

The probability that the winning bid is less than b is

$$P[X < b] = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$$

5.3 PDF of the Transformation of Random Variables

Consider a continuous random variable, X, with PDF $f_X(x)$. To find the PDF of $Y = aX + b$ we proceed as follows:

$$y = ax + b \Rightarrow x = \frac{y-b}{a}$$

$$F_Y(y) = P[Y \le y = ax + b] \equiv P\left[X \le x = \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$$

Therefore, the PDF of Y is given by

$$f_Y(y) = \frac{dF_X\left(\dfrac{y-b}{a}\right)}{dy} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

In general, suppose we are given the PDF, $f_X(x)$, of a random variable X and we want to find the PDF of a linear function of X, say $Y = g(X)$. First, we solve the equation $y = g(x)$ to obtain $x = g^{-1}(y)$. The PDF of the transformation Y is determined as follows:

$$F_Y(y) = P\left[Y \le y = g(x)\right] \equiv P\left[X \le x = g^{-1}(y)\right] = F_X\left(g^{-1}(y)\right)$$

$$f_Y(y) = \frac{dF_X(g^{-1}(y))}{dy} = f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right|$$

- $\left|\frac{dg^{-1}(y)}{dy}\right|$ is called the Jacobian of the transformation

<u>Examples:</u>

5.4 Expected Value of a Continuous Random Variable:

<u>Definition:</u> Consider a continuous random variable X, with PDF $f_X(x)$. The expected (mean or average) value of X, denoted $E[X]$ or μ_X is defined as

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

<u>Example:</u> Let random variable X represent the life expectancy (in hours) of a certain electronic component and has probability density function

$$f_X(x) = \begin{cases} \frac{20,000}{x^3}, & x > 100 \\ 0, & \text{elsewhere} \end{cases}$$

What is the expected life of this component? Otherwise, if I purchase many of these components, on average, how long will they last?

<u>Solution:</u>

$$\mu_X = E[X] = \int_{100}^{\infty} x \frac{20,000}{x^3} dx = -\frac{20,000}{x}\Bigg|_{100}^{\infty} = 200 \text{ hours}$$

Expected Value of a Function of a Continuous RV:

<u>Definition:</u> Consider a continuous random variable X, with PDF $f_X(x)$. The expected value of the function $h(X)$ of the random variable X, is defined as

$$\mu_{h(X)} = E\left[h(X)\right] = \int_{-\infty}^{\infty} h(x)\, f_X(x)\,dx$$

Examples: Let a continuous random variable X, have probability density function

$$f_X(x) = \begin{cases} \dfrac{x^2}{3}, & -1 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the expected value of $h(X) = 4X + 3$

Solution:

$$E[4X+3] = \int_{-1}^{2} \frac{(4x+3)x^2}{3} dx = \frac{1}{3}\int_{-1}^{2}\left(4x^3 + 3x^2\right)dx = \frac{1}{3}\left(x^4 + x^3\right)\Big|_{-1}^{2} = 8$$

5.5 Variance and Standard Deviation of Continuous Random Variables:

Definition: Consider a continuous random variable X with PDF $f_X(x)$ and mean μ_X. The variance of X, denoted $Var(X)$ or σ_X^2 is defined as

$$\sigma_X^2 = Var(X) = E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\,dx$$

Definition: The standard deviation of a random variable X is the square root of the variance,

$$\sigma_X = E\left[(X - \mu_X)^2\right] = \sqrt{\sigma_X^2}$$

- The variance is also known as the second central moment of X
- The variance (or standard deviation) of a random variable characterizes the variability or spread of the distribution.
- It gives a description of the shape of the distribution.

Shortcut formula for variance:

The shortcut formula is obtained by simplifying the expression

$$\sigma_X^2 = E\left[(X - \mu_X)^2\right] = E\left[X^2\right] - 2\mu_X E[X] + \mu_X^2 = E\left[X^2\right] - \mu_X^2$$

- $E\left[X^2\right]$ is the second moment or mean-square value of X

Example: The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

$$f_X(x) = \begin{cases} 2(x-1), & 1 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the mean and variance of X.

Solution:

$$\mu_X = 2\int_1^2 x(x-1)dx = \frac{5}{3}.$$

$$E\left[X^2\right] = 2\int_1^2 x^2(x-1)dx = \frac{17}{6}$$

$$\sigma_X^2 = E\left[X^2\right] - \mu_X^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}$$

Definition: Let X be a random variable with probability density function $f_X(x)$. Let $h(X)$ be a function of X with expected value $\mu_{h(X)}$. The variance of the function of a continuous random variable $h(X)$ is defined as

$$\sigma_{h(X)}^2 = E\left[\left(h(X) - \mu_{h(X)}\right)^2\right] = \int_{-\infty}^{\infty}\left(h(X) - \mu_{h(X)}\right)^2 f_X(x)dx$$

Example: Let a continuous random variable X have probability density function

$$f_X(x) = \begin{cases} \dfrac{x^2}{3}, & -1 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the variance of $h(X) = 4X + 3$.

Solution: From the previous example $\mu_{h(x)} = 8$

$$\sigma_{h(X)}^2 = E\left[(4X+3-8)^2\right] = \int_{-1}^2 (4x-5)^2 \frac{x^2}{3}dx = \int_{-1}^2 \frac{16x^4 - 40x^3 + 25x^2}{3}dx$$

$$= \frac{16 \times 11}{5} - 10 \times 5 + 25 = \frac{176 - 125}{15} = \frac{51}{5}$$

Properties of expectations:

Consider two random variable X and Y with expected values μ_X and μ_Y. Let a, b and c be constants. Then for

- $Y = a = \text{constant} \Rightarrow E[Y] = \mu_Y = a$ and $\sigma_Y^2 = 0$
- $Y = aX + b \Rightarrow \mu_Y = E[aX + b] = a\mu_X + b$ and $\sigma_Y^2 = a^2\sigma_X^2$

5.6 Important Continuous Random Variables in Electrical Engineering:

(Reading Exercises: Montgomery and Runger – Section 4.4)

Uniform Random Variable:

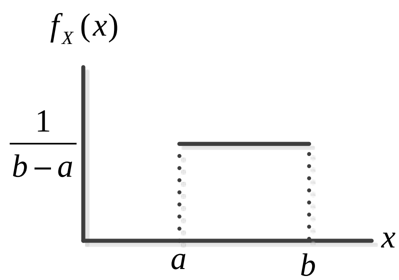
Definition: A continuous uniform random variable X, is one that is equally likely (with equal probability) to assume values in a finite interval $a \le X \le b$. The PDF is equal to a constant c, in the interval $a \le X \le b$,

$$f_X(x) = \begin{cases} c & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Since the area under the PDF is equal to 1, we can show that

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

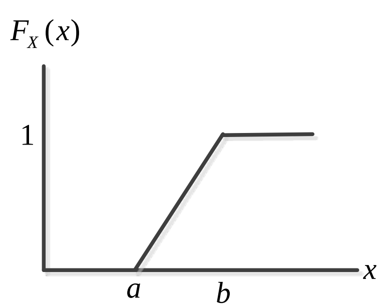
A sketch of the PDF is shown in the figure below.



CDF of Uniform Random Variables:

The CDF is obtained by integrating the PDF, in three regions.

$$F_X(x) = \begin{cases} 0, & x < a \\ \displaystyle\int_a^x \frac{1}{b-a}\,dy = \frac{x-a}{b-a}, & a \le x < b \\ 1 & x \ge b \end{cases}$$



Uniform Probability Distribution

Exercise: Show that a random variable, uniformly distributed in the interval $b \le x \le a$, has mean and variance given by $E[X] = (a+b)/2$ and $\sigma_X^2 = (a-b)^2/12$.

Example: Resistors are manufactured with a tolerance of $\pm 10\%$. Within this tolerance region, they are approximately uniformly distributed. What is the probability that a nominal 1000Ω resistor has a value between 900Ω and 1010Ω?
Solution: For a nominal 1000Ω resistor, the tolerance region is $900 \le X \le 1100$, therefore, the PDF is

$$f_X(x) = \begin{cases} \dfrac{1}{200}, & 900 \le X \le 1100 \\ 0, & \text{otherwise} \end{cases}$$

The probability that the resistance lies in the range $900 \le X \le 1010$ is

$$P[900 \le X \le 1010\Omega] = \int_{900}^{1010} \frac{1}{200} dx = 0.55.$$

Exponential Random Variables:
The exponential distribution is often used to model the time interval between occurrences of events or the lifetime of many devices or systems in practice.

The PDF and CDF of Exponential Random Variables:

PDF: The PDF of an exponential random variable, with parameter λ, is given by

$$f_X(x) = \begin{cases} 0, & 0 < 0 \\ \lambda e^{-\lambda x}, & x \ge 0 \end{cases}$$

CDF: The CDF of an exponential random variable, obtained by integrating the PDF, is given by

$$F_X(x) = \int_{-\infty}^{x} f_x(x) dx = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

Example: The interval of service (the duration from beginning to completion of service) for a customer in a line at the bank has an exponential distribution with parameter $\lambda = 0.033 s^{-1}$. You are next to be served. (a) What is the probability that 15 seconds or less will elapse until you are finished being served? (b) What is the probability that one minute or greater will elapse until you are finished being served?
Solution: Let the waiting time be Δt. Then the CDF of the waiting time is

$$F_{\Delta t}(\Delta t) = \left(1 - e^{-0.033\Delta t}\right) u(\Delta t)$$

$$P\left[\Delta t \le 15s\right] = F(15) - F(0) = 0.388$$

$$P\left[\Delta t \ge 60s\right] = 1 - P\left[\Delta t \le 60s\right]$$
$$= 1 - F(60) + F(0) = 0.138$$

Example: The amount of time one spends in a bank is exponentially distributed with parameter $\lambda = 10$ minutes. (a) What is the probability that a customer will spend more than 15 minutes in the bank? (b) What is the probability that a customer will spend more than 15 minutes in the bank given that he/she is still in the bank after 10 minutes?

Solution:

$$P\left[X > 15\right] = e^{-15\lambda} = e^{-15/10} = 0.22$$

$$P\left[X > 15 \middle| X > 10\right] = P\left[X > 5\right] = e^{-5/10} = 0.604$$

Mean of and Mean-squared Value of Exponential Random Variables:
The mean value of an exponential random variable is given by

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{\lambda}$$

$$E\left[X^2\right] = \int_0^{\infty} x^2 e^{-\lambda x} dx$$

Integrating by parts with $u = \lambda x^2$ and $dv = e^{-\lambda x}dx \Rightarrow du = 2\lambda x dx$ and $v = -\frac{1}{\lambda}e^{-\lambda x}$, yields

$$E\left[X^2\right] = \int_0^{\infty} x^2 e^{-\lambda x} dx = \lim_{r\to\infty}\left(\left[-x^2 e^{-\lambda x}\right]\Big|_0^r + 2\int_0^r x e^{-\lambda x} dx\right)$$

$$= \lim_{r\to\infty}\left(\left[-x^2 e^{-\lambda x} - \frac{2}{\lambda} x e^{-\lambda x} - \frac{2}{\lambda^2} e^{-\lambda x}\right]\Bigg|_0^r\right) = \frac{2}{\lambda^2}$$

Variance: Using the short-cut formula, the variance is

$$\sigma_X^2 = E\left(X^2\right) - \left(E(X)\right)^2 = \frac{1}{\lambda^2}$$

Exponential random variables are sometimes used to model the time to failure of equipment or devices.

Gaussian (or Normal) Random Variable:
(Reading Exercises: Montgomery and Runga – Section 4.5 - 4.6)

The Gaussian random variable is the single most widely used in practice.

- Many physical phenomena in electrical engineering are modeled very well by the Gaussian distribution.

- If the model of a continuous RV is unknown, we can conveniently use the Gaussian model.

Definition: A random variable X, with mean μ_X and variance σ_X^2, is said to be Gaussian if its PDF is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_x^2} \right), \quad -\infty \le x \le \infty$$

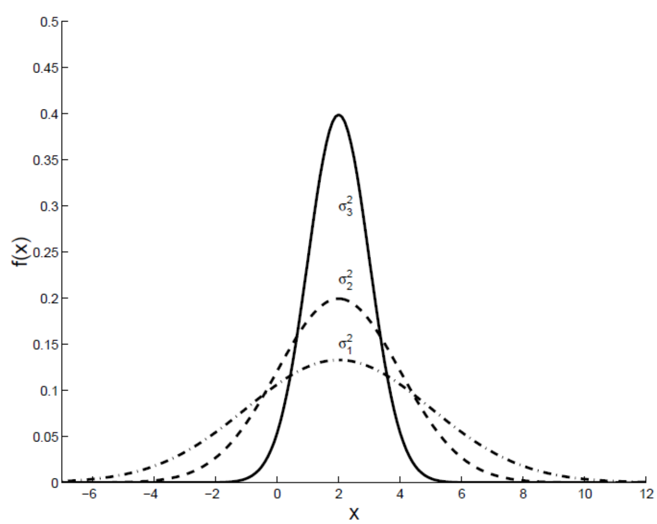
The Gaussian random variable is characterized by only two parameters:

- The mean μ_X and variance σ_X^2 or the standard deviation σ_X.

Properties of the Gaussian PDF:

- The curve is bell shaped
- The curve is symmetric around the mean μ_X
- When the variance σ_X^2, decreases the shape of the curve is peakier
- When the variance σ_X^2, increases the shape of the curve is more spread out

Example plots of a Gaussian PDF with, mean $\mu_x = 2$, and different variances are shown in the figure below.



Gaussian PDF: $\sigma_1 > \sigma_2 > \sigma_3$

Standard Normal (Standard Gaussian) Distribution:

For a Gaussian random variable Y with mean μ_Y and variance σ_Y^2, the PDF is given

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_y^2}\right), \quad -\infty \le y \le \infty$$

The above is called the general Gaussian distribution. Suppose we normalized the general Gaussian random variable to obtain another random variable, X , as shown below.

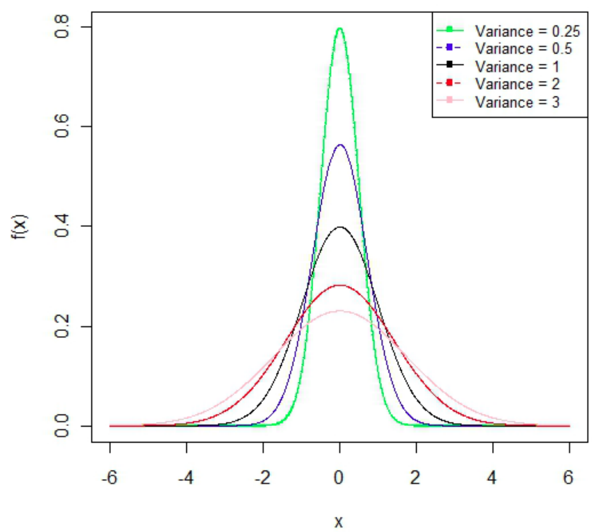
$$X = \frac{Y-\mu_Y}{\sigma_Y}$$

The normalized random variable has a mean of zero ($\mu_X = 0$) and a unit variance ($\sigma_X^2 = 1$).

Definition: A Gaussian random variable with zero mean and unit variance is referred to as a standard Gaussian random variable and has PDF given by.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty \le x \le \infty$$

The plots below illustrate examples of the standard Gaussian random variable.



Example plots of Standard Gaussian random variables

CDF of Stamdard and Regular Gaussian Random Variables:

The CDF for a standard Gaussian random variable X , is defined as

$$F_X(x) = P[X \le x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(x) \quad \text{(standard Gaussian)}$$

The CDF of the regular Gaussian random variable can be expressed in terms of $\Phi(\bullet)$ by making the following change of variable: $u = \dfrac{y - \mu_Y}{\sigma_Y} \Rightarrow dy = \sigma_Y du$, which yields

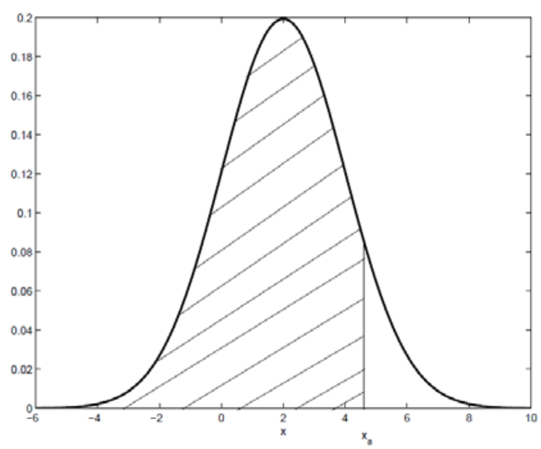
$$\begin{aligned} F_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_Y} \int_{-\infty}^{y} \exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_y^2}\right) dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y-\mu_Y}{\sigma_Y}} \exp\left(-\frac{u^2}{2}\right) du = \Phi\left(\frac{y-\mu_Y}{\sigma_Y}\right) \end{aligned} \quad \text{(regular Gaussian)}$$

The Q-Function:

The Q-function (or Marcum Q-function) denoted $Q(\bullet)$, is the complement of the standard Gaussian CDF,

$$Q(x) = 1 - P[X \le x] = P[X > x] = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp\left(-\frac{u^2}{2}\right) du$$

As illustrated in the plot below, the Q-function $Q(x_a)$, represents the area under the right tail of the standard normal function (the region $X > x_a$) below.



Right tail: $P[X > x_a]$; Shaded region: $P[X \le x_a] = 1 - P[X > x_a]$

<u>Remarks:</u>

- There is no closed form solution for the Q-Function integral
- The Q-Function is tabulated and is used extensively in Electrical Engineering.
- The CDF of the regular Gaussian random variable can be easily expressed in terms of the Q-function as follows:

$$F_X(x) = P[X \le x] = 1 - P[X > x] = \frac{1}{\sqrt{2\pi}\sigma_x}\int_x^{\infty} \exp\left(-\frac{(y-\mu_X)^2}{2\sigma_x^2}\right)dy$$
$$= \frac{1}{\sqrt{2\pi}}\int_{\frac{x-\mu_X}{\sigma_x}}^{\infty} \exp\left(-\frac{y^2}{2}\right)dy = 1 - Q\left(\frac{x-\mu_Y}{\sigma_x}\right)$$

Properties of the Q-function:

- $Q(-\gamma) = 1 - Q(\gamma), \ \ \gamma > 0$
- $Q(\infty) = 0$
- $Q(-\infty) = 1$
- $Q(0) = 0.5$
- $Q(x)$ is monotonically decreasing with x

Applications of the Gaussian Distribution:

Example: A certain type of battery lasts, on average, 3 years with a standard deviation of 0.5 years. Assuming battery lives are Gaussian distributed. Find the probability that a given battery will last less than 2.3 years.

Solution: Given $\mu_X = 3$ and $\sigma_X = 0.5$, find the probability that $X < 2.3$,

$$P[X \le 2.3] = 1 - P[X > 2.3] = 1 - \frac{1}{\sqrt{2\pi} \times 0.5}\int_{2.3}^{\infty} \exp\left(-\frac{(x-3)^2}{2 \times 0.5^2}\right)dx$$
$$= 1 - Q\left(\frac{2.3-3}{0.5}\right) = 1 - [1 - Q(1.4)] = 0.0808$$

Example: For the same battery life problem, what is the probability that a given battery will last between 1.5 to 2 years?

Solution: We need to find the following probability:

$$P[1.5 \le X \le 2] = P[X > 1.5] - P[X > 2] = Q\left(\frac{1.5-3}{0.5}\right) - Q\left(\frac{2-3}{0.5}\right)$$
$$= 1 - Q\left(\frac{1.5}{0.5}\right) - \left[1 - Q\left(\frac{1}{0.5}\right)\right] = Q(2) - Q(3)$$
$$= 0.0214$$

Example: What is the probability that a given battery will last between 3 to 4 years?

Solution:

$$P[3 \le X \le 4] = P[X > 3] - P[X > 4] = Q\left(\frac{3-3}{0.5}\right) - Q\left(\frac{4-3}{0.5}\right)$$
$$= 0.5 - Q(2) = 0.5 - 0.022275 = 0.47725$$

Example: A company manufactures electrical resistors with mean 3 Ohms and standard deviation 0.005 Ohms. A buyer sets specifications on the resistance to be 3.0 ± 0.01 Ohms. On the average, what percentage of resistors will be rejected?

Solution: We are required to find $P[3.01 < X < 2.99]$.

$$P[3.01 > X < 2.99] = P[x < 2.99] + P[x > 3.01] = 1 - P[x > 2.99] + P[x > 3.01]$$
$$= 1 - \left(1 - Q\left(\frac{3-2.99}{0.005}\right)\right) + Q\left(\frac{3.01-3}{0.005}\right) = 2Q(2) = 0.0455$$

The percentage of resistors rejected is 45.5%.

Example: A company manufactures electrical resistors with mean 3 Ohms. The standard deviation is 0.1Ω. Find a value of the tolerance d, such that 95% of all manufactured resistors fall in the range $3 \pm d\ \Omega$.

Solution: Assume Gaussian distribution. The region for 95% is $3-d \le X \le 3+d$. We need to find d such that the probability $P[3-d \le X \le 3+d]$ is equal to 95%`

$$P[3-d \le X \le 3+d] = P[X \ge 3-d] - P[X \ge 3+d]$$
$$= Q\left(\frac{3-d-3}{0.1}\right) - Q\left(\frac{3+d-3}{0.1}\right)$$
$$= 1 - Q\left(\frac{d}{0.1}\right) - Q\left(\frac{d}{0.1}\right) = 1 - 2Q\left(\frac{d}{0.1}\right) = 0.95 \Rightarrow Q\left(\frac{d}{0.1}\right) = 0.025$$
$$Q(10d) = 0.025 \Rightarrow d = 0.196\Omega$$

Example*:* Signal Detection - assume that in detection of a radar signal, with background noise, follow a Gaussian distribution with a mean of 0 V and a standard deviation of 0.45 V. The radar system assumes that an enemy airplane is detected if the received voltage exceeds 0.9 V. What is the probability of False Alarm?

Solution:

$$P_{FA} = P[\text{Received voltage} > 0.9V \mid \text{no enemy airplane exists}]$$
$$= P[\text{noise} > 0.9\text{V}]$$
$$= Q\left[\frac{0.9}{0.45}\right] = Q(2) = 0.02275$$

6. BASIC RELIABILITY CALCULATIONS:
(Reading Exercises: Class Notes)

Reliability may be described in several ways as follows:

- Reliability is a measure of the quality of a product.
- Reliability is the probability of the product working for a specified period.
- Reliability is the ability of a product to perform a required function without failure, under stated conditions for a specified time period.

Learning outcomes:

You will be able to
- Define Failure distribution (PDF and CDF)
- Define Reliability and its properties
- Define Failure rate and its properties
- Know the formal definition of Mean time to failure (MTTF) and its derivation
- Evaluate reliability, failure rate and MTTF for a given distribution
- Evaluate the reliabilities of systems involving series and parallel connections

In order to increase reliability, the causes of failure need to be identified and addressed in the design stage, if possible. Otherwise, reliability must be designed into the product itself.

"Unreliability" has several unfortunate consequences that can be damaging to a company. Below are some examples.

- Safety can be impacted
- Competitiveness can be impacted
- Profit margins can be reduced
- Cost of repair and maintenance can be hefty
- Delays can be experienced further up the supply chain
- Reputation can be damaged
- Good will can be impacted

In reliability analysis, the underlying random variable, denoted T, is the lifetime of the product. In the study of reliability, we are concerned with the probability that a component or system will survive beyond a stated time, that is, $T > t$, i.e., there is no failure in the time interval ($0 \le T \le t$).

6.1 Failure distribution function (Failure CDF and failure PDF)

<u>Definition:</u> Failure CDF, denoted $F_T(t)$, is the probability that a component or system will fail in the interval $0 \le T \le t$,

$$F_T(t) = P[T \le t] = \int_0^t f_T(t)dt$$

<u>Definition:</u> Failure PDF, denoted $f_T(t)$, is the derivative of failure CDF,

$$f_T(t) = \frac{dF_T(t)}{dt}$$

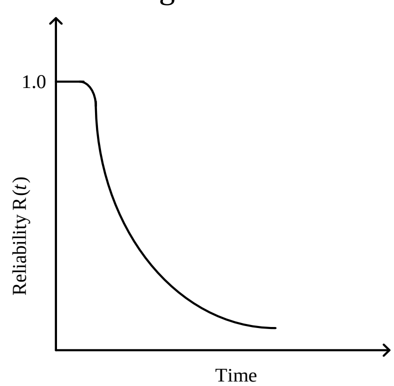
6.2 Reliability function:

<u>Definition:</u> If a component or product has lifetime T, with failure distribution function $F_T(t)$, then the reliability function of the component or product, denoted $R_T(t)$, is the probability that failure occurs later, after time t,

$$R_T(t) = \mathrm{P}[T > t] = 1 - P[T \le t] = 1 - F_T(t)$$

- Reliability is the complement of failure
- Reliability function is also known as survival function

A typical reliability curve is shown in the figure below.



We can derive the following properties from the curve above:

<u>Properties of the reliability function:</u>

- $R_T(t)$ is a decreasing function of t. That is, if $t_1 < t_2$ then $R_T(t_2) < R_T(t_1)$.
- $R_T(0) = P[T > 0] = 1$ (Initial life, new born, etc.)
- $R_T(\infty) = 0$ (Nothing survives forever).
- $0 \le R_T(t) \le 1$.

<u>Failure PDF in Terms of Reliability Function:</u>
Taking the derivative of the reliability function gives,

$$R_T'(t)\frac{dR_T(t)}{dt}=\frac{d\left(1-F_T(t)\right)}{dt}=-f_T(t)\Rightarrow f_T(t)=-R_T'(t)$$

<u>Definition:</u> Failure PDF is the negative of the derivative of the reliability function.

6.2 Conditional probability of failure

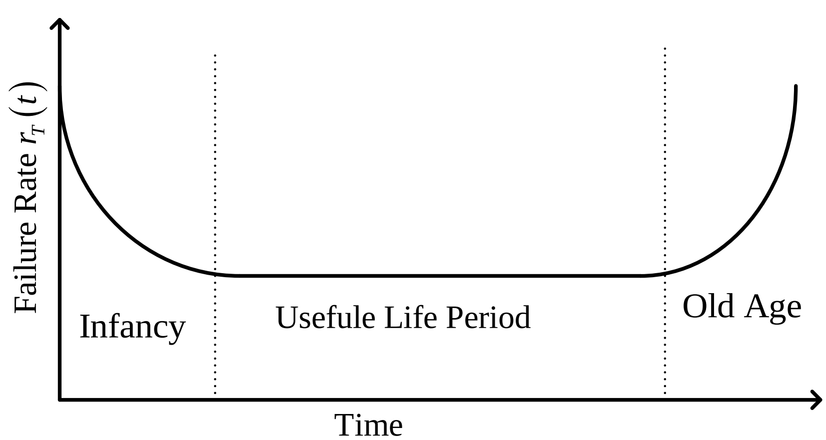
<u>Definition:</u> Conditional probability of failure is the probability of the product failing in the interval $t \le T \le t+\Delta t$, given that it did not fail prior to time t,

$$P\left[t<T\le t+\Delta t \mid T>t\right]=\frac{P\left[\{t<T\le t+\Delta t\}\cap\{T>t\}\right]}{P\left[T>t\right]}=\frac{P\left[t<T\le t+\Delta t\right]}{P\left[T>t\right]}=\frac{F_T(t+\Delta t)-F_T(t)}{R_T(t)}$$

<u>Failure rate:</u> Failure rate, denoted, $r_T(t)$, is the rate at which the conditional probability of failure changes with time. That is,

$$r_T(t)=\lim_{\Delta t\to 0}\frac{F_T(t+\Delta t)-F_T(t)}{\Delta t}\frac{1}{R_T(t)}=\frac{dF_T(t)}{dt}\frac{1}{R_T(t)}$$
$$=\frac{f_T(t)}{R_T(t)}=\frac{f_T(t)}{1-F_T(t)},\quad t>0,\ F_T(t)<1$$

The failure rate function curve is, in general, represented by a bath tub curve, as shown below.



Bathtub Model

The bathtub curve represents the failure rate of a product during its lifecycle, which comprises three stages:

- Early failure period (or infant mortality) – this is the debugging period in which faults are fixed.
- Useful life period (constant failure rate period) – this is where failures are rare.
- Old age period - failures occur due to wear-out.

6.3 Mean Time to Failure ($MTTF$)

Definition: Mean time to failure, denoted $MTTF$, is the mean or average time to failure (also known as average lifetime), is defined as the expected value of the product's lifetime T,

$$MTTF \triangleq E\{T\} = \int_0^\infty t f_T(t)dt$$

Substituting $f_T(t) = -R_T'(t)$ into $\int_0^\infty t f_T(t)dt$ and integrating by parts, we obtain

$$\int_0^\infty t f_T(t)dt = -\int_0^\infty t R_T'(t)dt = tR_T(t)\Big|_0^\infty + \int_0^\infty R_T(t)dt$$

If we consider that $R_T(t)$ decays much faster than t approaching infinity, then the first term is zero. Therefore,

$$MTTF \triangleq E\{T\} = \int_0^\infty R_T(t)dt = \int_0^\infty [1 - F_T(t)]dt, \quad \text{where } R_T(t) = 1 - F_T(t) = \int_t^\infty f_T(t)dt$$

Summary of Useful Formulas

Expression in terms of	$f_T(t)$	$F_T(t)$	$R_T(t)$
$f_T(t)$		$\dfrac{dF_T(t)}{dt}$	$-\dfrac{dR_T(t)}{dt}$
$F_T(t)$	$\int_0^t f_T(t)dt$		$1 - R_T(t)$
$R_T(t)$	$\int_t^\infty f_T(t)dt$	$1 - F_T(t)$	
$r_T(t)$	$\dfrac{f_T(t)}{\int_t^\infty f_T(t)}$	$\dfrac{dF_T(t)/dt}{1 - F_T(t)}$	$-\dfrac{dR_T(t)/dt}{R_T(t)}$

<u>Exponential Failure Distribution (PDF)</u>

The exponential distribution is the most commonly used distribution in reliability engineering and it models the useful life (constant) portion of the bath-tub curve.

<u>Definition:</u> For an exponentially distributed lifetime, the reliability function is,

$$R_T(t) = e^{-\lambda t}\ .$$

The PDF, CDF, reliability function, failure rate and MTTF for an exponential failure rate are all summarized in the table below.

PDF	CDF	Reliability	Failure rate	MTTF
$f_T(t) = \lambda e^{-\lambda t}$	$F_T(t) = 1 - e^{-\lambda t}$	$R_T(t) = e^{-\lambda t}$	$r(t) = \dfrac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$	$\int_0^{\infty} R_T(t)dt = \dfrac{1}{\lambda}$

<u>Example:</u>
Suppose we want to measure the time it takes to run a washing machine before it fails. If the distribution is exponential with mean $100,000 \text{ hours}$. What is the probability that the machine will fail during its first 50,000 hours:

<u>Solution:</u>

$$\mu_T = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{\mu_T} = \frac{1}{100,000 \text{ hours}}$$

$$P\left[X \le 50,000\right] = \int_0^{50,000} \frac{1}{100,000} \exp\left(-\frac{1}{100,000} x\right) dx$$

$$= -\exp\left(-\frac{x}{100,000}\right)\Bigg|_0^{50,000} = 1 - e^{-\frac{1}{2}} \approx 0.3935$$

<u>Weibull Distribution (PDF):</u> (Montgomery and Runga Section 4.9)

The Weibull PDF, used to model the bath tub curve, is given by

$$f_T(t) = \lambda\beta(\beta t)^{\beta-1} \exp\left(-\lambda t^{\beta}\right), \quad \beta > 0,\ \lambda > 0,\ t > 0.$$

The parameter β is called the shape factor.

- For $\beta = 1$, we have an exponential distribution and the failure rate is constant.
$$f_T(t) = \lambda \exp(-\lambda t).$$
- For $\beta < 1$, the failure rate decreases with time.
- For $\beta > 1$, the failure rate increases with time.

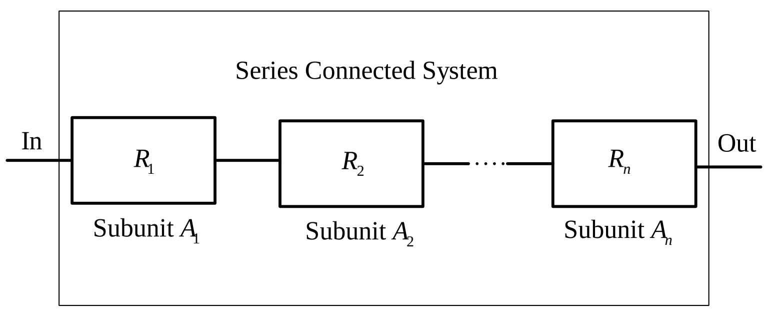
6.4 Reliability of Systems:

In practice, an electrical product or system is made up of multiple components and we often want to know the reliability of such a system. The reliability of the product depends on how the components are inter-connected. Two basic types of inter-connections are series (series systems) and parallel (parallel systems). More complex systems will comprise a combination of series and parallel connections.

Series Systems:
When a system consists of n components connected in series, the entire system fails as soon as any one component fails. The probability that subunit A_i functions correctly is equal to its reliability $R_i,\ i = 1, 2, \ldots n$. A series connected system will only operate properly if all its components (or sub-units) function properly. If the components operate completely independently, the reliability (being probability) of this model is the product of the reliabilities of the individual components, that is,

$$R_s(t) = \prod_{i=1}^{n} R_i(t)$$

The overall system reliability is, therefore, determined by that of the least reliable component.



Example: A system consists of n components that operate independently. Assume the component lifetimes are exponential random variables with rates $\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_n$. Let $R_j(t),\ j = 1, 2, \cdots, n$ denote the reliability of the j^{th} component in a series connection. Then the overall reliability function is the product of all individual reliability functions, that is,

$$R_s(t) = \prod_{j=1}^{n} R_j(t) = \mathrm{e}^{-t\sum_{j=1}^{n}\lambda_j} \Rightarrow \lambda_s = \sum_{j=1}^{n} \lambda_j$$

The system life time is also exponentially distributed with the mean rate λ_s, being equal to the sum of the individual lifetimes.

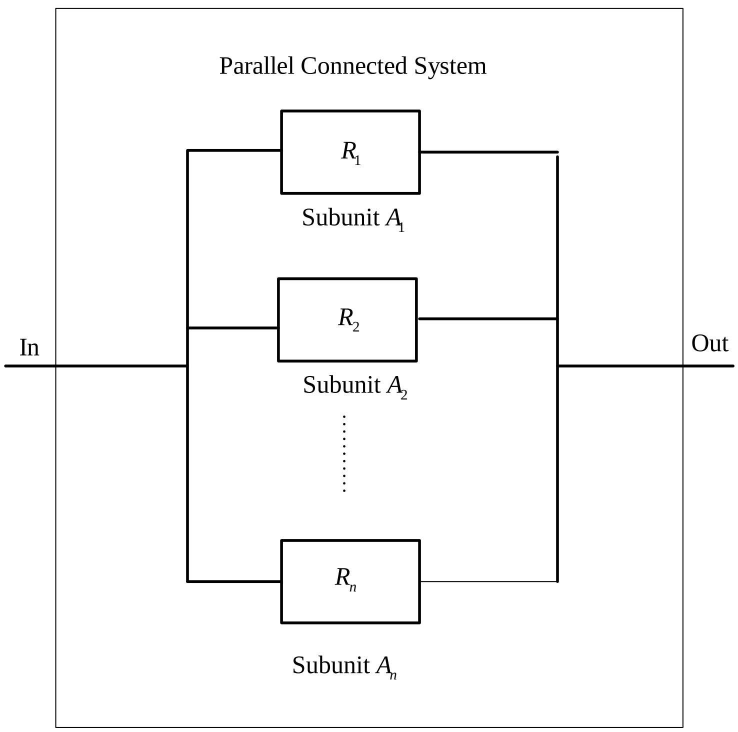
Parallel Systems:

When a system consists of n components connected in parallel, the entire system fails only if all components fail. Let $R_i,\ i=1,2,\ldots n$ denote the reliability of the i^{th} subunit in a parallel connection. For this case, we consider the complement (the probability that the i^{th} component will fail) $1-R_i(t)$. The probability that all the components will fail (that is, the system will fail) is the product

$$1-R_S(t) = \left(1-R_1(t)\right)\left(1-R_2(t)\right)\cdots\left(1-R_n(t)\right)=\prod_{i=1}^{n}\left(1-R_i(t)\right)$$

The system reliability is, therefore, equal to

$$R_S(t) = 1-\prod_{i=1}^{n}\left(1-R_i(t)\right)$$



Example: A parallel-connected system consists of n components and the component lifetimes are exponential random variables with rates $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$. Find the reliability.

$$R_s(t)=1-\prod_{j=1}^{n}\left(1-\mathrm{e}^{-\lambda_j t}\right)$$

7. Bivariate Random Variables
(Reading Exercises: Montgomery and Runger – Chapter 5)

In some experiments, it may be desirable to simultaneously measure outcomes of two or more random variables $X = \{X_1, X_2, \cdots, X_n\} \in \mathbb{R}^n$. Often in electrical engineering, we want to investigate the joint behavior of multiple random variables. The focus in this course, however, is on bivariate (two) random variables. In order to study the joint behavior of two random variables we must know how to handle joint probability distributions.

Learning outcomes:

You will

- Know the formal definition of a joint probability mass function of two discrete random variables.
- Know how to use a joint probability mass function to find the probability of a specific event.
- Know how to find a marginal probability mass function of a discrete random variable X from the joint probability mass function of X and Y.
- Know the conditions for independence of two random variables X and Y.
- Be able to find the expectation of a function of the discrete random variables X and Y using their joint probability mass function.
- Be able to find the means and variances of the discrete random variables X and Y using their joint probability mass function.
- Be able to compute covariance and correlation coefficient inorder to infer the kind of relationship between two random variables.
- Be able to apply the methods learned in the section to solve new problems.

7.1 Joint Probability Distribution of Discrete Random Variables

Joint PMF

Definition: Consider two discrete random variables X and Y, defined on the sample space of an experiment. Their joint PMF, denoted $p_{XY}(x, y)$ is defined, for all real numbers x and y, as

$$p_{XY}(x, y) = P[X = x \text{ and } Y = y]$$

Properties of the joint PMF:

1. $p_{XY}(x, y) \geq 0 \ \forall (x, y)$
2. $\sum_x \sum_y p_{XY}(x, y) = 1$

- Any function that satisfies these two properties is a joint PMF.

Joint CDF of Discrete Random Variables

Definition: Consider two discrete random variables X and Y, defined on the sample space of an experiment. Their joint CDF, denoted $F_{XY}(x,y)$ is defined, for all real numbers x and y, as

$$F_{XY}(x,y) = P[X \le x, Y \le y] = \sum_{u=-\infty}^{x} \sum_{v=-\infty}^{y} p_{XY}(u,v)$$

Example: For the joint PMF in the table below, calculate $P[X+Y \le 2]$.

PMF	$X=0$	$X=1$
$Y=1$	0.09	0.06
$Y=2$	0.05	0.08
$Y=3$	0.08	0.24
$Y=4$	0.10	0.30

Solution:

$$P[X+Y \le 2] = P[X=0, Y=1] + P[X=0, Y=2] + P[X=1, Y=1]$$
$$= 0.09 + 0.05 + 0.06 = 0.2$$

Example: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find

(a) The joint probability mass density function $p_{XY}(x,y)$.

(b) $P[(X,Y) \in A]$, where A is the region $\{(x,y) \mid x+y \le 1\}$

Solution: Since only two ball pens are selected, the possible pairs of values (the region) are (x,y): $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, $(0,2)$, $(2,0)$

(a) The probability $p_{XY}(x,y)$ represents the probability that x blue and y red pens are selected. There are 3 distinguishable partitions – blue, red and green.

- The total number of equally likely ways of selecting any 2 pens from the 8 pens is $n = \binom{8}{2} = 28$.
- The number of equally likely ways of selecting x blue pens from 3 is $n_1 = \binom{3}{x}$
- The number of equally likely ways of selecting y red pens from 2 is $n_2 = \binom{2}{y}$.

- The number of equally likely ways of selecting green pens from 3 after selecting x blue and y red pens is $n_3 = \binom{3}{2-x-y}$.

Therefore, the PMF is

$$p_{XY}(x,y) = \frac{n_1 n_2 n_3}{n} = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}, \quad x = 0,1,2;\ y = 0,1,2;\ 0 \le x + y \le 2$$

$p(x,y)$	$x=0$	1	2
$y=0$	$p(0,0)=\frac{3}{28}$	$p(1,0)=\frac{9}{28}$	$p(2,0)=\frac{3}{28}$
1	$p(0,1)=\frac{3}{14}$	$p(1,1)=\frac{3}{14}$	$p(2,1)=0$
2	$p(0,2)=\frac{1}{28}$	$p(1,2)=0$	$p(2,2)=0$

(b) $P[(X,Y)\in A] = P[X+Y\le 1] = p(0,0)+p(0,1)+p(1,0) = \frac{9}{14}$

The tabulated PMF is a valid PMF since the sum of all the entries is equal to one.

7.2 Continuous Random Variables

You will

- Know the formal definition of a joint probability density function of two continuous random variables.
- Be able to use a joint probability density function to find the probability of a specific event.
- Know how to find a marginal probability density function of a continuous random variable from the joint probability density function of two random variables.
- Know how to find the means and variances of the continuous random variables using their joint probability density function.
- Know the formal definition of a conditional probability density function of a continuous random variable Y given a continuous random variable X.
- Know how to calculate the conditional mean and conditional variance of a continuous random variable Y given a continuous random variable X.
- Be able to compute covariance and correlation coefficient inorder to infer the kind of relationship between two random variables
- Be able to apply the methods learned in the section to solve new problems.

Joint PDF:

Definition: Let X and Y be two continuous random variables defined on the sample space of an experiment. The joint pdf, denoted $f_{XY}(x,y)$, is a function with the following properties:

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

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

(3) $P[(X,Y) \in A] = \iint_A f_{XY}(x,y)\,dxdy$

- Any function that satisfies these two properties is a joint PDF.

Joint CDF

Definition: The joint CDF for two continuous random variables, X and Y, is defined as

$$F_{XY}(x,y) = P[X \leq x,\ Y \leq y] = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u,v)\,dudv$$

- The joint PDF is the second partial derivative with respect to the two random variables,

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

Properties of the Joint CDF:

$$F_{XY}(-\infty,y) = 0 \qquad F_{XY}(x,-\infty) = 0 \qquad F_{XY}(-\infty,-\infty) = 0$$
$$F_{XY}(\infty,\infty) = 1 \qquad F_{XY}(x,\infty) = F_X(x) \qquad F_{XY}(\infty,y) = F_Y(y)$$

7.3 Marginal PDFs:

Definition: For discrete random variables the marginal or individual probability mass functions of X and Y are defined, respectively, as

$$p_X(x) = \sum_y p_{XY}(x,y), \quad p_Y(x) = \sum_x p_{XY}(x,y)$$

Example: Find the marginal PDFs of the random variables X and Y for the joint PMF given by the table below.

$p_{XY}(x,y)$	$x=0$	1	2
$y=0$	$p(0,0)=\frac{3}{28}$	$p(1,0)=\frac{9}{28}$	$p(2,0)=\frac{3}{28}$
1	$p(0,1)=\frac{3}{14}$	$p(1,1)=\frac{3}{14}$	$p(2,1)=0$
2	$p(0,2)=\frac{1}{28}$	$p(1,2)=0$	$p(2,2)=0$

Solution:

x	0	1	2
$p_X(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$
y	0	1	2
$p_Y(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

Definition: For continuous random variables, the marginal or individual probability density functions of X and Y are defined, respectively, as

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

Example: Find the marginal pdfs for $f_{XY}(x,y)=\frac{2}{5}(2x+3y)$ for $0\le x\le 1$ and $0\le y\le 1$

Answers: $f_X(x)=\frac{1}{5}(4x+3)$ and $f_Y(y)=\frac{2}{5}(1+3y)$

7.4 Conditional Probability Distribution:

Conditional CDFs

Recall that

$$P[B\mid A]=\frac{P[A\cap B]}{P[A]}=\frac{P[AB]}{P[A]}.$$

This definition also applies to the conditional CDF of X and Y. Joint CDF is given below.

$$F_{X|Y}(x|Y) = P[X \le x | Y] = \frac{P[X \le x, Y]}{P[Y]}.$$

Similarly, the CDF of Y conditioned on X is given by

$$F_{Y|X}(Y | X) = P[Y \le y | X] = \frac{P[Y \le y, X]}{P[X]}$$

Conditional PDF

Like the joint conditional CDF, the joint conditional PDF is defined as

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}, \quad p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)} \quad \text{(Discrete)}$$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}, \quad f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \text{(Continuous)}$$

Example: Use the table below to find the conditional PMF of X given that $Y = 1$ and use it to find $P[X = 0 | Y = 1]$.

$p_{XY}(x,y)$	$x = 0$	1	2
$y = 0$	$p(0,0) = \frac{3}{28}$	$p(1,0) = \frac{9}{28}$	$p(2,0) = \frac{3}{28}$
1	$p(0,1) = \frac{3}{14}$	$p(1,1) = \frac{3}{14}$	$p(2,1) = 0$
2	$p(0,2) = \frac{1}{28}$	$p(1,2) = 0$	$p(2,2) = 0$

Solution:

$$p_{X|1}(x|1) = \frac{p_{XY}(x,1)}{p_Y(1)}$$

$$p_Y(1) = \sum_{x=0}^{2} p_{XY}(x,1) = \frac{3}{14} + \frac{3}{14} = \frac{3}{7} \quad \Rightarrow \quad p_X(x|1) = \frac{p_{XY}(x,1)}{p_Y(1)} = \frac{7}{3} p_{XY}(x,1)$$

$$p_{X|1}(0|1) = \frac{p_{XY}(0,1)}{p_Y(1)} = \frac{7}{3} \times \frac{3}{14} = \frac{1}{2} \qquad p_{X|1}(1|1) = \frac{p_{XY}(1,1)}{p_Y(1)} = \frac{7}{3} \times \frac{3}{14} = \frac{1}{2}$$

$$p_{X|1}(2|1) = \frac{p_{XY}(2,1)}{p_Y(1)} = \frac{7}{3} \times 0 = 0 \qquad P[X = 0 | Y = 1] = \frac{p_{XY}(0,1)}{p_Y(1)} = \frac{1}{2}$$

Example: The joint pdf is

$$f_{XY}(x,y)=\begin{cases}10xy^2, & 0<x<y<1\\ 0, & \text{elsewhere}\end{cases}$$

Solution:

$$f_X(x)=10\int_x^1 xy^2dy=\frac{10}{3}x\left(1-x^3\right),\ \text{for } 0<x<1;$$
$$f_Y(y)=10\int_0^y xy^2dx=5y^4,\ \text{for } 0<y<1$$
$$f_{Y|X}(y\,|\,x)=\frac{f_{XY}(y,x)}{f_X(x)}=\frac{3y^2}{1-x^3},\ \ \text{for } 0<x<1 \text{ and } 0<y<1$$
$$P\left(Y>\frac{1}{2}\,|\,X=\frac{1}{4}\right)=\int_{\frac{1}{2}}^1 f_{Y|X}\left(y\,|\,x=\frac{1}{4}\right)dy=\int_{\frac{1}{2}}^1\frac{3y^2dy}{1-\left(\frac{1}{4}\right)^3}=\frac{8}{9}$$

Example: Consider the following joint density function:

$$f_{XY}(x,y)=\begin{cases}x\left(1+3y^2\right), & 0<x<1,\ 0<y<1,\\ 0, & \text{elsewhere.}\end{cases}$$

Find the marginal density functions $f_X(x)$ and $f_Y(y)$, and the conditional density function $f_{X|Y}(x\,|\,y)$.

Solution:

$$f_X(x)=x\int_0^1\left(1+3y^2\right)dy=x\left(y+y^3\right)\Big|_{y=0}^{y=1}=2x$$
$$f_Y(y)=\int_0^1 x\left(1+3y^2\right)dx=\frac{x^2}{2}\left(1+3y^2\right)\Big|_{x=0}^{x=1}=\frac{1+3y^2}{2}$$
$$f_{X|Y}(x\,|\,y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{x\left(1+3y^2\right)}{\left(1+3y^2\right)/2}=2x$$
$$P\left[\frac{1}{4}<x<\frac{1}{2}\middle|Y=\frac{1}{3}\right]=\int_{1/4}^{1/2}2xdx=\frac{3}{16}$$

Conditional Expectations:

Definition: If X and Y are two random variables, the conditional expectation of X given $Y = y$, is defined as

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx, \quad \text{(continuous)}$$

$$E[X \mid Y = y] = \sum_{x} x f_{X|Y}(x \mid y), \quad \text{(discrete)}$$

Example:

A soft-drink machine has a random supply Y (with measurements in gallons) at the beginning of given day. Let X denote the amount (in dollars) of soft drink sold during the day. The machine is not resupplied during the day. It has been observed that X and Y have joint PDF

$$f_{XY}(x, y) = \begin{cases} \frac{1}{2}, & 0 \le x \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the conditional expectation of amount of sales X given that $Y = 1$

Solution:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{y}, & 0 \le x \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

For $Y = 1$, we have

$$f_{X|Y}(x \mid y) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

$$E[X \mid Y = 1] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx = \int_{0}^{1} x dx$$

$$= \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}$$

7.5 Independence of Random Variables and Functions:

Definition: Two random variables, X and Y, are said to be statistically independent if one of the following holds:

$$F_{XY}(x,y)=F_X(x)F_Y(y), \quad f_{XY}(x,y)=f_X(x)f_Y(y);$$
$$f_{X|Y}(x|y)=f_X(x), \qquad f_{Y|X}(y|x)=f_Y(y).$$

Examples: In the previous example,

$$f_{X|Y}(x|y)=f_X(x)=2x.$$

Therefore, random variables X and Y, are independent.

7.6 Expected Values of Functions of Bivariate Random Variables:

Definition: Consider two random variables X and Y with joint probability distribution function $p_{XY}(x,y)$ (PMF) or $f_{XY}(x,y)$ (PDF). If $g(X,Y)$ is a real-valued function of (X,Y), then the expected value of $g(X,Y)$ is

$$E[g(X,Y)]=\sum_x\sum_y g(x,y)p_{XY}(x,y) \quad \text{(discrete)}$$

$$E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(X,Y)f_{XY}(x,y)\,dxdy \quad \text{(continuous)}$$

Definition: Consider two independent random variables X and Y with means μ_X and μ_Y, respectively. Then the expected value of their product is

$$r_{XY}=E[XY]=E[X]E[Y]=\mu_X\mu_Y$$

Definition: If two random variables, X and Y, are independent then, for any two functions $g(X)$ and $h(Y)$,

$$E[g(X)h(Y)]=E[g(Y)]E[h(Y)]$$

Covariance:

Definition: Consider two random variables X and Y with means μ_X and μ_Y. The covariance of X and Y, denoted $Cov(X,Y)$ or simply C_{XY}, is defined as

$$C_{XY}=E[(X-\mu_X)(Y-\mu_Y)]=\sum_y\sum_x(x-\mu_X)(y-\mu_Y)\,p_{XY}(x,y), \quad \text{(discrete)}$$

$$C_{XY}=E[(X-\mu_X)(Y-\mu_Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_X)(Y-\mu_Y)\,f_{XY}(x,y)\,dxdy \quad \text{(Continuous)}$$

$$C_{XY} = E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right] = E\left[XY\right] - \mu_X \mu_Y = r_{XY} - \mu_X \mu_Y \quad \text{(Short-cut formula)}$$

- Covariance provides information regarding the nature of a linear relationship between the two random variables X and Y.
- If there is no linear relationship or the two random variables are uncorrelated, then $C_{XY} = 0$.

Examples:

Correlation Coefficient:

Definition: Consider two random variables X and Y with C_{XY} and individual standard deviations σ_X and σ_Y. The correlation coefficient of X and Y, denoted ρ_{XY} is defined as

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1$$

ρ_{XY} has no unit and is used widely in Engineering.

-
    - $\rho_{XY} = 1$ X and Y have a linear relationship (correlated) with a positive slope.
    - $\rho_{XY} = 1$: X and Y have a linear relationship (correlated) with a negative slope.
    - $\rho_{XY} = 0$: X and Y have no linear relationship (uncorrelated)
    - Statistically independent random variables are uncorrelated, but the converse is not necessarily true.

Examples: For the joint PMF in the table below, show that the two random variables are uncorrelated $C_{XY} = 0$ but they are dependent, that is, $p_{XY}(x,y) \ne p_X(x)\, p_Y(y)$.

	$X=-1$	$X=0$	$X=1$	Totals
$Y=-1$	$1/8$	$1/8$	$1/8$	$3/8$
$Y=0$	$1/8$	0	$1/8$	$2/8$
$Y=1$	$1/8$	$1/8$	$1/8$	$3/8$
Totals	$3/8$	$2/8$	$1/8$	1

Exercise: A travel agent keeps track of the number of customers who call and the number of trips booked on any one day. Let X denote the number of calls, Y the number of trips booked and $p_{XY}(x,y)$ the joint PMF. Records show the following:

$p_{XY}(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$
$y=0$	0.04	0.08	0.12	0.10
$y=1$	0	0.06	0.20	0.16
$y=2$	0	0	0.12	0.10
$y=3$	0	0	0	0.02

Verify that $E[XY]=2.20$

x	0	1	2	3
$p_X(x)$	0.04	0.14	0.44	0.28

y	0	1	2	3
$p_Y(y)$	0.34	0.42	0.22	0.02

$\mu_X=1.86$	$\mu_Y=0.92$	$C_{XY}=0.4888$	$\sigma_X^2=0.9604$	$\sigma_Y^2=0.6336$	$\rho_{XY}=0.6266$

Example: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint PDF

$$f_{XY}(x,y)=\begin{cases}8xy, & 0\le y\le x\le 1\\ 0, & \text{elsewhere}\end{cases}$$

Find the covariance of X and Y, and the correlation coefficient.

Solution: First, we compute the marginal PDFs:

$$f_X(x)=\int_0^x 8xy\,dy=\begin{cases}4x^3, & 0\le x\le 1,\\ 0, & \text{elsewhere.}\end{cases}\qquad f_Y(y)=\int_y^1 8xy\,dy=\begin{cases}4y\left(1-y^2\right), & 0\le y\le 1,\\ 0, & \text{elsewhere.}\end{cases}$$

Next, we compute the individual mean values

$$\mu_X=E[X]=\int_0^1 4x^4dx=\frac{4}{5},\quad \mu_Y=E[Y]=\int_0^1 4y^2\left(1-y^2\right)dy=\frac{8}{15}.$$

Using the joint PDF, we find

$$E[XY]=\int_0^1\int_y^1 8x^2y^2dxdy=\frac{4}{9}$$

The covariance is equal to

$$C_{XY} = E[XY] - \mu_X \mu_Y = \frac{4}{9} - \frac{4}{5} \times \frac{8}{15} = \frac{4}{225}$$

$$C_{XY} = E[XY] - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}$$

To find the correlation coefficient, we need the individual variances. We can use the short-cut formula:

$$E\left[X^2\right] = \int_0^1 4x^5 dx = \frac{2}{3}, \qquad E\left[Y^2\right] = \int_0^1 4y^3\left(1-y^2\right)dy = \frac{1}{3}$$

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75}, \quad \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}$$

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}} = 0.4924$$

7.7 Sums of Random Variables:

7.7.1 Mean and Variance of Sums of Two Random Variables

Definition: Consider two random variables X and Y. The mean and variance of their sum are

- $E[X \pm Y] = \mu_X \pm \mu_Y$
- $Var(X \pm Y) = \sigma_X^2 + \sigma_Y^2 \pm 2C_{XY}$ (if not independent)
- $Var(X \pm Y) = \sigma_X^2 + \sigma_Y^2$ (if independent)

Consider n independent random variables $X_1, X_2, \ldots, X_n$ with individual PDFs $f_i(x_i)$, $i = 1, 2, \ldots, n$. It is desired to determine the PDF and the statistics of their sum $Y = \sum_{i=1}^{n} X_i$.

The best way to handle this is through the moment generating functions.

Learning Outcomes:

You will

- Know the definition of and how to calculate the moment-generating function of a linear combination of n independent random variables.
- Understand the steps involved in each of the proofs in this section.
- Be able to determine the PDF of a linear combination of n independent random variables from the moment-generating function.

- Understand and be able to apply the central limit theorem
- To be able to apply the methods learned to new problems.

7.7.2 Moment Generating Functions

Single Random Variable:

Definition: Consider a single random variable X, with PDF $f_X(x)$. The moment generating function of X, denoted $\phi_X(u)$, is defined as the expected value of the complex function e^{uX}, that is,

$$\phi_X(u) = E\left[e^{uX}\right] = \int_{-\infty}^{\infty} e^{uX} f_X(x)\,dx$$

The above integral may be interpreted as the inverse Fourier transform of the PDF $f_X(x)$. In that case, the probability density function may be interpreted as the Fourier transform of the moment generating function, that is,

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

Remarks:

- The moment generating function and the pdf may be considered Fourier transform pairs.
- In many cases it may be easier to obtain the moment generating functions of random variables than it is to find the PDF directly.
- Furthermore, knowing the moment generating function, it is easier to evaluate the expected values (or moments) of the random variable.

Expected value of X (first moment):

Taking the first derivative of the moment generating function $\phi_X(u)$, gives

$$\frac{d\phi_X(u)}{du} = \int_{-\infty}^{\infty} f_X(x)\frac{d}{du}\left(e^{ux}\right)dx = \int_{-\infty}^{\infty} x f_X(x) e^{ux}\,dx$$

If we evaluate the first derivative at $u = 0$, we obtain the expression for the mean of X,

$$\left.\frac{d\phi_X(u)}{du}\right|_{u=0} = \int_{-\infty}^{\infty} x f_X(x)\,dx = \mu_X$$

To obtain the second moment of X, we evaluate the second derivative at $u = 0$, that is,

$$\left.\frac{d^2\phi_X(u)}{du^2}\right|_{u=0} = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = E\left[X^2\right]$$

In general, we can obtain the m^th moment, $E\left[X^m\right]$, by evaluating we need to take the m^{th} derivative of the moment generating function $u=0$, that is,

$$\left.\frac{d^m\phi_X(u)}{du^m}\right|_{u=0} = \int_{-\infty}^{\infty} x^m f_X(x)\,dx = E\left[X^m\right]$$

Two Random Variable:

Definition: The moment generating function of the sum of two independent random variables, X and Y, is defined as

$$\phi_{XY}(u,v) = E\left[e^{(uX+vY)}\right] = \phi_X(u)\phi_Y(v)$$

- The moment generating function of the sum of two independent random variables is equal to the product of their individual moment generating functions

Definition: The joint PDF, defined as the Fourier transform of $\phi_{XY}(u,v)$, is the convolution of the individual PDF's,

$$f_{XY}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(ux+vy)}\phi_X(u)\phi_Y(v)\,dudv = f_X(x) * f_Y(y)$$

- According to the Fourier transform theory, the Fourier transform of the product of two functions is equal to the convolution of their Fourier transforms.
- In general, given the PDFs of n independent random variables, $f_i(x_i)$, $i = 1, 2, \ldots, n$, the joint PDF of the sum these random variables is equal to the convolution of all the PDFs of the individual random variables, that is,

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) * f_2(x_2) * \ldots * f_n(x_n)$$

Properties of Moment-Generating Functions:

- Consider $Y = X + a$. Then $\phi_Y(u) = e^{au}\phi_X(u)$
- Consider $Y = aX$. Then $\phi_Y(u) = \phi_X(au)$

- Consider $Y=\sum_{i=1}^{n} X_i$. Then $\phi_Y(u)=\phi_{X_1}(u),\phi_{X_2}(u),\ldots,\phi_{X_n}(u)$

7.8 The Central Limit Theorem:

Let $X_1, X_2, \ldots, X_n$ be n independent random variables with arbitrary probability distribution functions. The central limit theorem tells us about the probability distribution of their sum as n approaches infinity.

Definition: Let $X_1, X_2, \ldots, X_n$ be n independent random variables with arbitrary probability distribution functions and identical mean, $\mu_i = \mu$, and variance $\sigma_i^2 = \sigma^2$.

Then, according to the central limit theorem, the distribution of their sum, $X=\sum_{i=1}^{n} X_i$ converges to the Gaussian distribution as n increases.

- The mean of the sum $\mu_X = \sum_{i=1}^{n} \mu_i = n\mu$
- The variance of the sum is

$$\sigma_X^2 = \sum_{i=1}^{n} \sigma_i^2$$

Definition: Consider the sum of the normalized random variables,

$$Y = \sum_{i=1}^{n} \frac{x_i - \mu_i}{\sqrt{n}\sigma_i}$$

Then, according to the central limit theorem, as the number of random variables increase, Y tends to the standard Gaussian random variable ($\mu_X = 0$ and $\sigma_Y^2 = 1$).

Rule of Thumb: If $n > 30$, the Central Limit Theorem can be used.

8 STATISTICS:

(Reading Exercises: Montgomery and Runger – Chapter 7)

So far, we have focused on theoretical models where knowledge of the relevant probability distribution functions is assumed to be known. Statistics deals with the real world where probability distributions are unknown; all we have are measured or collected data. Given a set of random data, we are interested in determining the sample mean, sample variance and the empirical distribution of the data.

Consider a set of n independent random data samples $\{x_1,\ x_2,\ \ldots,\ x_n\}$, with identical but unknown distributions. We say that the data set consists of n independent and identically distributed (IID) random variables.

Definition: A statistic is any quantity whose value can be calculated from a given set of random data samples. • A statistic is a function, $U\left(x_1,\ x_2,\ \ldots,\ x_n\right)$, of the collected random data • A statistic is a random variable and is denoted by an uppercase letter; a lowercase letter is used to represent the value of the statistic.

Learning outcomes:

You will • Know the formal definition for sample mean and sample variance; and how to find them from randomly collected data • Know the formal definition for sample covariance and sample correlation coefficient; and how to find them from randomly collected data • Empirical distribution functions from randomly collected data • Understand and be able to create a quantile-quantile (q-q) plot.

8.1 Statistics and their Distributions:

Given a set of random data, we want to determine estimates of the first and second order statistics – the mean and variance

Sample Mean – First order statistic

Consider a set of random samples, $X = \{x_1,\ x_2,\ \ldots,\ x_n\}$, consisting of n independent random variables that have identical mean $\mu_{X_i} = \mu_X$ and identical variance $\sigma^2_{X_i} = \sigma^2_X$. However, the true mean is unknown and needs to be estimated from these samples. A statistic that is often used to estimate the true mean is the sample mean, denoted $\overline{X}$, and defined as

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

The mean and variance of the sample mean are given, respectively, by

$$\mu_{\overline{X}} = E\left[\overline{X}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \mu_X$$

$$\sigma_{\overline{X}}^2 = E\left[\left(\overline{X} - \mu_{\overline{X}}\right)^2\right] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right]\right)^2\right] = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i - \mu_X\right)^2\right]\right) = \frac{\sigma_X^2}{n}$$

- The mean of the sample mean is equal to the true mean of X.
- The variance of the sample mean reduces to zero as the sample size n, is increased.
- The accuracy of the estimated mean improves with increasing sample size.

Example: It is required to infer whether the population of resistors from a production line is within a tolerance range $1000\Omega \pm 10\%$. We pick a random sample of 10 resistors and measure their resistances in ohms, which are provided in the table below

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
900	1013	939	1062	1017	996	970	1079	1065	1049

Solution: Compute the sample mean

$$\overline{X} = \frac{1}{10}\sum_{j=1}^{10} X_j = 1009\ \Omega$$

The sample mean is $\overline{X} = 1009\ \Omega$ and is, therefore, within the tolerance range $1000\ \Omega \pm 10\%$

Sample Variance

The sample variance is an estimate of the true variance of X.

Definition: Consider a set of random data samples, $X = \{x_1,\ x_2,\ \ldots, x_n\}$, consisting of n IID random variables. The sample variance may be computed in one of two ways:

Mean is known:

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_X\right)^2$$

Mean is unkown:

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{or}$$

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

Exercise: Prove that for

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2 \text{ , } E\left[S_X^2\right] = \frac{n-1}{n}\sigma_X^2$$

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ , } E\left[S_X^2\right] = \sigma_X^2$$

The sample variance can be expressed in a short-cut formula form as following:

$$S_X^2 \triangleq \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \bar{X}\right)^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n}x_j^2 - 2\bar{X}\sum_{j=1}^{n}x_j + \sum_{i=1}^{n}\bar{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{j=1}^{n}x_j^2 - \frac{2}{n}\left(\sum_{j=1}^{n}x_j\right)^2 + \frac{1}{n}\left(\sum_{j=1}^{n}x_j\right)^2\right) = \frac{n\sum_{j=1}^{n}x_j^2 - \left(\sum_{j=1}^{n}x_j\right)^2}{n(n-1)}$$

The sample standard deviation is defined as

$$S_X = \sqrt{\frac{n\sum_{j=1}^{n}X_j^2 - \left(\sum_{j=1}^{n}X_j\right)^2}{n(n-1)}}$$

For the previous example, the sample variance is

$$S_X^2 = \frac{10\sum_{i=1}^{10}X_i^2 - \left(\sum_{i=1}^{10}X_i\right)^2}{10\times 9} = \frac{102118060 - 101808100}{90} = 3444\Omega^2 \Rightarrow S_X = 58.7\Omega$$

Sample Covariance and Correlation Coefficient:

In order to infer whether two populations have any relationship, we collect random samples $X = \{X_i\}$ and $Y = \{Y_i\}$, $i = 1, 2, \ldots, n$, from each population and then estimate the covariance and the correlation coefficient. These two statistics will tell us whether the two populations are correlated.

Definition: Consider two collected random data $X = \{X_i\}$ and $Y = \{Y_i\}$, $i = 1, 2, \ldots, n$, from two populations. The sample covariance is defined as

$$C_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(y_i - \overline{Y}\right) = \frac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n(n-1)}$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} {}_iY$$

Sample Correlation Coefficient: The sample correlation coefficient is defined as

$$\rho_{XY} = \frac{C_{XY}}{S_X S_Y}, \quad -1 \le \rho_{XY} \le +1$$

- If $\rho_{XY} = 0$ then X and Y are uncorrelated or independent.
- If $\rho_{XY} = \pm 1$, then X and Y are correlated or linearly related.

Example: Find the correlation coefficient for the data pairs in the table below.

X_i	0.68	0.72	1.27	2.01	2.63	3.06	3.15	4.00	4.03	4.50
Y_i	12.45	9.93	6.64	10.14	8.93	13.34	11.56	16.72	19.62	15.03

Solution:

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2.6, \quad \overline{Y} = \frac{1}{10}\sum_{i=1}^{10} Y_i = 12.44$$

$$S_X^2 = \frac{10\sum_{i=1}^{10} X_i^2 - \sum_{i=1}^{10} X_i^2}{10\times 9} = 1.95 \Rightarrow S_X = 1.4$$

$$S_Y^2 = \frac{10\sum_{i=1}^{10} Y_i^2 - \left(\sum_{i=1}^{10} Y_i\right)^2}{10\times 9} = 16.07 \Rightarrow S_Y = 4.0$$

$$C_{XY} = \frac{10\sum_{i=1}^{10} X_i Y_i - \sum_{i=1}^{10} X_i \sum_{i=1}^{10} Y_i}{10 \times 9} = 3.86$$

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} = \frac{3.86}{1.4 \times 4} = 0.7$$

- There is a strong correlation between the two populations

8.2 Empirical Distribution Functions:

Suppose a random variable, X , has a CDF $F_X(x)$, which we do not know. However, we have several IID experimental data samples $X = \{X_1, \ldots, X_n\}$.

Definition: The empirical CDF is defined as

$$\tilde{F}_X(x) = \frac{\text{Number of samples, } X_1, \ldots, X_n, \text{ no greater than } x}{n}$$

Example: Obtain the empirical distribution of the resistance samples in the table below.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
900	1013	939	1062	1017	996	970	1079	1065	1049

Solution: List the samples in ascending order, then compute $\tilde{F}_X(x)$

$Y_1 = X_1$	$Y_2 = X_3$	$Y_3 = X_7$	$Y_4 = X_6$	$Y_5 = X_2$	$Y_6 = X_5$	$Y_6 = X_{10}$	$Y_6 = X_4$	$Y_6 = X_9$	$Y_6 = X_8$
900	939	970	996	1013	1017	1049	1062	1065	1079
$\tilde{F}_X(x) = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

8.3 Quantile-Quantile (Q-Q) Plot :

When we collect data samples, we often want to know the probability distribution. A quantile-quantile (Q-Q) plot is used to determine whether a set of experimental data fits some specified theoretical distribution. Quantile-Quantile plots are an excellent graphical tool for comparing a sample data set to a theoretical distribution, most often, the standard Gaussian model.

Procedure for Computing Quantiles:

Consider the case where we collect a set of n experimental data samples, $X_1, \dots, X_n$, and we want to see how well this data set fits a standard Gaussian CDF function $F_X(x)$. The following are the steps for computing the quantiles:

- **Data (or Sample) Quantile:**

  The sample quantile is obtained by sorting the values in the data sample set into ascending order and denoting them as $X_{\min} = X_{(1)}, \dots, X_{(n)} = X_{\max}$. The ordered data set is known as the sample quantile.

  Example: The table below shows a data set and its sample quantile.

Data Set	$X_i = 50, 76, 92, 83, 105, 102, 109, 106, 91, 110, 89$
Sample quantile	$X_{(i)} = 50, 76, 83, 89, 91, 92, 102, 105, 106, 109, 110$

- **Theoretical Quantile:**

Definition: The p^{th} quantile is a number q_p $(-\infty < q_p < \infty)$, such that

$$P\left[X \le q_p\right] = p$$

The parameter p is called the p^{th} percentile, where $0 \le p \le 1$

- We divide the interval $[0, n]$ into n equal bins and assume the k^{th} percentile will lie between adjacent points $\frac{k}{n}$ and $\frac{k+1}{n}$, $k = 0, 1, \dots, n$. The percentile values lie in the intervals $\left(\frac{0}{n}, \frac{1}{n}\right), \left(\frac{1}{n}, \frac{2}{n}\right), \dots, \left(\frac{n-1}{n}, 1\right)$ $\left(\frac{0}{n}, \frac{1}{n}\right), \left(\frac{1}{n}, \frac{2}{n}\right), \cdots, \left(\frac{n-1}{n}, 1\right)$.
  If we want the percentile values to line on the mid point of each interval, then the i^{th} percentile is computed as $p_i = \frac{i-0.5}{n}$, $i = 1, 2, \dots, n$.
- Define the i^{th} quantile z_i (also known as a z-score), such that

$$F_Z(z_i) = P\left[Z \le z_i\right] = \frac{i-0.5}{n}, \quad i = 1, 2, \cdots, n$$

Testing whether our data fits a standard Gaussian distribution:

If we want to test whether our data fits a standard Gaussian distribution, we would use the Gaussian CDF as follows:

$$P[Z \le z_i] = F_Z(z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} \exp\left(-\frac{x^2}{2}\right) dx = 1 - Q(z_i) = \frac{i-0.5}{n}$$

$$Q(z_i) = 1 - \frac{i-0.5}{n},\ i = 1, 2, \ldots, n$$

- We obtain the z_i values from the Q-function table that satisfy the above equation. The theoretical quantile is the set of values of z-scores.

For the data collected above, the sample and theoretical quantiles are summarized in the table below.

i	$X_{(i)}$	$(i-0.5/n)$	$Q(z_i) = 1 - \left(\frac{i-0.5}{n}\right)$	z_i
1	50	0.05	0.95	–1.65
2	76	0.15	0.85	–1.05
3	83	0.25	0.75	–0.675
4	89	0.35	0.65	–0.375
5	91	0.45	0.55	–0.125
6	92	0.55	0.45	0.125
7	102	0.65	0.35	0.375
8	105	0.75	0.25	0.675
9	106	0.85	0.15	1.05
10	109	0.95	0.05	1.65

- If we are testing for a regular Gaussian distribution, we would have to compute the sample mean and sample standard deviation. We find the z_i values from the CDF

$$F_Z[z_i] = 1 - Q\left(z_i = \frac{X_{(i)} - \overline{X}}{S_X}\right) = \frac{i-0.5}{n} \Rightarrow Q(z_i) = 1 - \frac{i-0.5}{n}$$

Finally, we express the sample quantile $X_{(i)}$ versus the theoretical quantile z_i as

$$X_{(i)} = S_X z_i + \overline{X}$$

Definition: A Q-Q plot is a plot of the sample quantile $X_{(i)}$ quantiles versus the theoretical quantiles z_i, or vice versa.

- If the Q-Q plot is approximately linear, then our data are approximately Gaussian distributed.

Exercise: (a) Sketch the Q-Q-plot and conclude whether or not our data is coming from a Guassian population.

Example: The table below gives the sample quantiles and theoretical quantiles for a standard Gaussian distribution.

Sample Quantile $X_{(i)}$	i	$\frac{i-0.5}{n}$	Theoretical Quantile $z_i = F^{-1}\left(\left(i-0.5\right)/n\right)$	$X_{(i)} = S_X z_i + \bar{X}$ (Regular Gaussian)
-1.96	1	0.1	-1.28	
-0.78	2	0.3	-0.52	
0.31	3	0.5	0.00	
1.15	4	0.7	0.52	
1.62	5	0.9	1.28	

Example: We are given the following random data set:

$$X = \{25.0,\ 25.0,\ 27.7,\ 25.9,\ 25.9,\ 21.7,\ 22.8,\ 28.9,\ 26.4,\ 22.4\}$$

Compute $\bar{X}$ and S_X, fill in the corresponding values in the last column of the table below.

i	$X_{(i)}$	$p_i = \left(i - 0.5/n\right)$	z_i	$X_{(i)} = S_X z_i + \bar{X}$
1	21.7	0.05	−1.65	
2	22.4	0.15	−1.05	
3	22.8	0.25	−0.675	
4	25.0	0.35	−0.375	
5	25.0	0.45	−0.125	
6	25.9	0.55	0.125	
7	25.9	0.65	0.375	
8	26.4	0.75	0.675	
9	27.7	0.85	1.05	
10	28.9	0.95	1.65	

Plot $X_{(i)}$ versus z_i and $X_{(i)} = S_X z_i + \overline{X}$ on the same graph. How do they relate?

Monte Carlo Simulations:

A Monte Carlo simulation is a method for carrying out complex computer simulations of systems undergoing random perturbations. Suppose we have a system described in terms of input-output equations. The input or system parameters have random aspects (random noise, unknown component values, etc.). We can obtain a probabilistic description of the system by modeling the random quantities with random number generators on a computer, repeatedly solving the system equations as we draw different realizations for the random quantities and computing some statistics of the desired responses or outputs.

9 Estimation Theory and Applications:

(Reading Exercises: Montgomery and Runger – Sections 7.1, 7.2, 7.3.1, 7.3.2, 7.3.3, 7.3.5, 7.4.2)

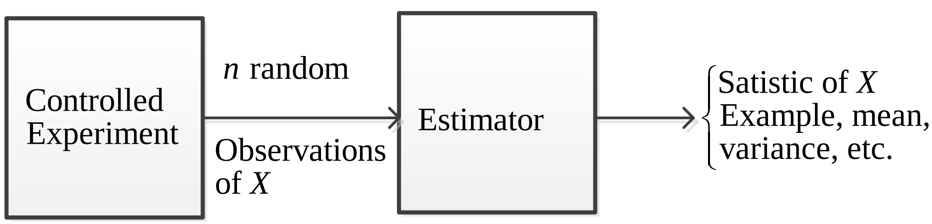
When we acquire samples taken from a population, we may need to estimate certain statistics or parameters of interest that help us make inferences about the population.

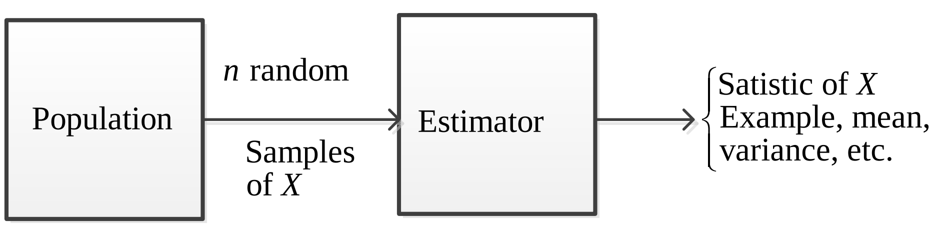
<u>Learning Outcomes:</u>

You will

- Know how to find a maximum likelihood estimator of a population parameter.
- Know how to check to see if an estimator is unbiased and/or minimum variance for a parameter.
- Understand the difference between point estimation and interval estimation
- Be able to establish a confidence interval for a sample mean
- Understand the steps involved in each of the proofs in the section.
- Be able to apply the methods learned in the section to new problems.

<u>Definition:</u> An estimator is a procedure used to obtain an estimate of a desired parameter from a set of random data samples drawn from a controlled experiment or a population.





For example, we can observe data samples X from n geostationary satellites and use them to calculate estimates of our location.

9.1 Desirable Properties of Good Estimators:

Bias:

Definition: Let $\hat{\theta}$ denote an estimate of a desired parameter. The bias, denoted as B, of the estimator is defined as the difference between the expected value of the estimate and the true value of value of the parameter, that is,

$$B = E\left[\hat{\theta}\right] - \theta$$

- If $B = 0$, then the estimator is said to be unbiased.
- If $B \neq 0$, the estimator is said to be biased.
- The desired estimator is the unbiased estimator.

Variance:

Definition: The variance of an estimate of parameter θ, is defined as

$$Var\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2\right]$$

- It is desirable that the variance of the estimator, which is a measure of the spread of values of the estimate around the true parameter, be as small as possible – minimum variance estimator.

Definition: A consistent estimator is one whose variance and bias go to zero as the number of samples approach infinity.

Definition: The mean-squared error of an estimate of parameter θ, is defined as

$$MSE\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = Var\left(\hat{\theta}\right) + B^2$$

Conclusion:

- For overall accuracy, we must find a minimum variance, unbiased estimator (MVUE).

Sample Mean and Sample Variance:

Taking the expectation of the sample mean gives

$$\mu_{\overline{X}} = E\left[\overline{X}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \mu_X$$

Therefore, the sample mean is unbiased.

Consider the two samples variances defined earlier. For the first definition, let us first simplify as follows:

$$
\begin{aligned}
S_X^2 &\triangleq \frac{1}{n}\sum_{j=1}^{n}\left(x_j-\bar{X}\right)^2=\frac{1}{n}\sum_{j}^{n}\left(x_j-\mu_X+\mu_X-\bar{X}\right)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(x_j-\mu_X\right)^2-\frac{2}{n}\sum_{j=1}^{n}\left(x_j-\mu_X\right)\left(\bar{X}-\mu_X\right)+\frac{1}{n}\sum_{i=1}^{n}\left(\bar{X}-\mu_X\right)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(x_j-\mu_X\right)^2-2\left(\bar{X}-\mu_X\right)^2+\left(\bar{X}-\mu_X\right)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(x_j-\mu_X\right)^2-\left(\bar{X}-\mu_X\right)^2
\end{aligned}
$$

We have used the relationship

$$
\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu_X\right)=\bar{X}-\mu_X
$$

Taking the expectation, and noting that the samples are independent, gives

$$
\begin{aligned}
\sigma_X^2 &=\frac{1}{n}\sum_{j=1}^{n}E\left[\left(x_j-\mu_X\right)^2\right]-E\left[\left(\bar{X}-\mu_X\right)^2\right]\\
&=\sigma_X^2-\frac{\sigma_X^2}{n}=\frac{n-1}{n}\sigma_X^2\neq\sigma_X^2
\end{aligned}
$$

The sample variance in this case is biased.

Consider now, the second definition. We may write the expression as

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

The expression in the winged brackets is the same as the first definition. Therefore, taking the expected value gives

$$
E\left[S_X^2\right]=\frac{1}{n-1}\sum_{i=1}^{n}E\left[(X_i-\bar{X})\right]^2=\frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}E\left[(X_i-\bar{X})^2\right]\right\}=\frac{n}{n-1}\left(\frac{n-1}{n}\sigma_X^2\right)=\sigma_X^2
$$

The expected value of the sample mean in this case is unbiased.

9.2 Methods for Obtaining Estimators:

There are two types of parameters to be estimated – deterministic and random parameters. Based on the parameter of interest, there are two basic estimators:

- Minimum mean square estimators (MMSE): used when the parameter to be estimated is random.
- Maximum likelihood estimators (MLE): used when the parameter is deterministic.

For example, we want to estimate a parameter θ, from a random sample $X+V$, where V is Gaussian noise.

- If θ is a deterministic (constant) parameter, then use the MLE.

- If θ is a random variable, then use the MMSE.

In this course, the focus is on deterministic parameter estimation.

9.2.1 Maximum Likelihood Estimator:

Consider a set of collected IID random samples $X=\{X_1,\ldots,X_n\}$, that contain a desired deterministic parameter θ. The joint conditional pdf $f\left(x_1,\ldots x_n\,|\,\theta\right)$, given θ, is known to as the likelihood function. The MLE estimate is the parameter that maximizes the likelihood function. Since the random variables are IID, the joint conditional PDF (the likelihood function) is the product of the individual PDFs,

$$f\left(x_1,x_2,\ldots,x_n\,|\,\theta\right)=\prod_{i=1}^{n} f\left(x_i\,|\,\theta\right)$$

A parameter that maximizes the logarithm of a function will also maximize the function itself.
Taking the natural log of the likelihood function gives

$$L\left(\mathbf{x}\,|\,\theta\right)=\ln f\left(x_1,x_2,\ldots,x_n\,|\,\theta\right)=\sum_{i=1}^{n} L\left(x_i\,|\,\theta\right)$$

The parameter, θ, that maximizes the log likelihood function must satisfy the following condition:

$$\frac{\partial}{\partial\theta}L\left(\mathbf{x}\,|\,\theta\right)=\frac{\partial}{\partial\theta}\ln f\left(x_1,x_2,\ldots,x_n\,|\,\theta\right)=0$$

Estimation of the Mean of with known variance:

Consider a set of n IID samples, $X=\{X_1,\ldots,X_n\}$, which are known to be Gaussian. The variance, σ_X^2, of each sample is assumed known but the mean, μ_X, is unknown. The likelihood function is

$$f(\mathbf{x}\,|\,\mu_X)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left(-\frac{(x_i-\mu_X)^2}{2\sigma_X^2}\right)=\left(2\pi\sigma_X^2\right)^{-\frac{n}{2}}\exp\left(-\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2}\right)$$

Taking the natural log gives

$$L(\mathbf{x}\,|\,\mu_X)=-\frac{n}{2}\ln\left(2\pi\sigma_X^2\right)-\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2}$$

Taking the derivative with respect to μ_X and equating to zero gives

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

Therefore, the maximum likelihood estimate of the mean is the sample mean,

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

The ML estimate of the mean is unbiased since $E\left[\bar{X}\right]=\mu_X$. It is also minimum variance for large sample size, since $\sigma_{\bar{X}}^2=\sigma_X^2/n\to 0$ with increasing n.

Estimation of the variance with known Mean:

Consider a set of n IID samples, $X=\{X_1,\ldots,X_n\}$, which are known to be Gaussian. The mean, μ_X, of each sample is assumed known but the variance, σ_X^2, is unknown. Find the MLE for the variance σ_X^2, of n IID samples, which are known to be Gaussian. The mean of each sample μ_X, is assumed known. The likelihood function is

$$\ell(\mathbf{x}\,|\,\sigma_X^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left(-\frac{(x_i-\mu_X)^2}{2\sigma_X^2}\right)=\left(2\pi\sigma_X^2\right)^{-\frac{n}{2}}\exp\left(-\sum_{i=1}^{n}\frac{(x_i-\mu_X)^2}{2\sigma_X^2}\right)$$

Taking the derivative with respect to σ_X^2 and equating it to zero gives

$$\frac{\partial L\left(\mathbf{x}\,|\,\sigma_X^2\right)}{\partial \sigma_X^2}=\sum_{i=1}^{n}\frac{\left(x_i-\mu_X\right)^2}{2\left(\sigma_X^2\right)^2}-\frac{n}{2}\frac{1}{2\sigma_X^2}2\pi=0$$

Solving the above equation, the estimate of the variance is the sample variance

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

The ML estimate of the variance is biased for small sample size since $E\left[S_X^2\right]=\sigma_X^2-\sigma_X^2/n$. However, for large sample size, it is unbiased since $E\left[S_X^2\right]=\sigma_X^2-\sigma_X^2/n\to\sigma_X^2$ as $n\to\infty$.

Estimation of the Mean and Variance:

Consider a set of n IID samples, $X=\left\{X_1,\ldots,X_n\right\}$, which are known to be Gaussian. The variance, σ_X^2, and the mean, μ_X, are unknown. Equating to zero, the partial derivatives of the log likelihood function with respect to the mean and variance, respectively, and then solving the two equations simultaneously for the estimates gives

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

9.3 Interval Estimation and Confidence Interval:

Instead of seeking a single value that we designate to be the estimate of the parameter of interest, we may want to specify an interval that is highly likely to contain the true value of the parameter. For example, we can assume that the desired parameter lies inside an interval $l\left(X\right)\le\theta\le u\left(X\right)$. The objective then is to find the interval limits, such that the interval contains θ with some degree of confidence or acceptable probability, say $1-\alpha$. That is,

$$P\left[l\left(\mathbf{x}\right)\le\theta\le u\left(\mathbf{x}\right)\right]=1-\alpha$$

Definition: Let α denote the probability that the parameter θ is not in the interval $\left[l\left(\mathbf{x}\right),\ u\left(\mathbf{x}\right)\right]$. Then,

- $1-\alpha$ is called the degree of confidence, which may be also measured in percentage

- The width of $\left[l(\mathbf{x}),\ u(\mathbf{x})\right]$ is called the confidence interval and is a measure of the accuracy of the estimator. A narrow interval corresponds to a higher accuracy of the estimated parameter.

9.3.1 Interval Estimation of the Mean with Known Variance:

Consider a set of IID random samples, $X = X_1, \ldots, X_n$ obtained from a Gaussian population where the mean is unknown but the true variance σ_X^2 , is known. Denote the true mean as μ_X . Let us use the sample mean $\bar{X}$, as an estimate of the true mean. We know that

$$E\left[\bar{X}\right] = \mu_X , \quad \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}.$$

Define the normalized sample estimate

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}}$$

According to the central limit theorem, the normalized sample estimate Z , is a standard Gaussian random variable. The distribution being symmetric around zero, we can write the estimation interval as $z_c \le Z \le z_c$. Given α , we can determine z_c by evaluating the probability that the normalized sample mean is inside this interval. That is,

$$P\left[-z_c \le Z \le z_c\right] = \int_{-z_c}^{z_c} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1 - 2Q\left(z_c\right) = 1 - \alpha \Rightarrow Q\left(z_c\right) = \alpha/2$$

We use the Q-function table to find $z_c = Q^{-1}\left(\frac{\alpha}{2}\right)$. For example, for $\alpha = 0.05$, $Q\left(z_c\right) = 0.025$ and the table gives $z_c \simeq 1.96$.

<u>**Confidence interval for the true mean:**</u>

In order to find the confidence interval for the true mean, we make the following substitution:

$$-z_c \le Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \le z_c$$

Upon simplification, we obtain the confidence interval for the true mean as

$$\boxed{\bar{X} - \frac{\sigma_X}{\sqrt{n}} z_c \leq \mu_X \leq \bar{X} + \frac{\sigma_X}{\sqrt{n}} z_c}$$

This interval is valid even if the set of random samples are not Gaussian as long as the sample size n is large enough to invoke the central limit theorem.

Example: (a) Find the 99% confidence interval for the batch of data in the table below. (b) How will the 99% confidence limits change if the number of samples is increased to 100? The true variance is 4.

X_i	7.31	10.8	11.27	11.91	5.51	8.0	9.03	14.42	10.24	10.91

Solution: The sample mean is $\bar{X} = 9.94$ for 10 samples. For a confidence level of 99%, $\alpha = 0.01$ and $Q(z_c) = 0.5\alpha = 0.005$. From the Q-function table, $z_c \simeq 2.6$. The confidence interval for the true mean is

$$9.94 - \frac{\sqrt{4}}{\sqrt{10}} \times 2.6 \leq \mu_X \leq 9.94 + \frac{\sqrt{4}}{\sqrt{10}} \times 2.6 \Rightarrow 8.3 \leq \mu_X \leq 11.58$$

If we increase the number of samples to 100, we obtain

$$9.94 - \frac{\sqrt{4}}{\sqrt{100}} \times 2.6 \leq \mu_X \leq 9.94 + \frac{\sqrt{4}}{\sqrt{100}} \times 2.6 \Rightarrow 9.12 \leq \mu_X \leq 9.42$$

We observe that the confidence interval is tighter, and the accuracy is, therefore, increased with sample size.

10 Hypothesis Testing

(Reading Exercises: Montgomery and Runger – Sections 9.1.1-9.1.3, 9.1.5, 9.2.1 9.3.1)

A statistical Hypothesis is an assertion or conjecture concerning one or more populations. We take samples from the population and examine them to prove whether a hypothesis is true.

<u>**Definition**</u>: Statistical or hypotheses tests are procedures that enable us to decide either to reject or to accept claims or assertions about a population.

- Hypothesis testing is a systematic way to test claims or ideas about a population.

<u>**Learning outcomes:**</u> Students are expected to be able to do the following:

You will be able to

- Define null hypothesis and alternative hypothesis,
- Define Type I error and Type II error
- Identify the steps involved in the hypothesis testing procedure.
- Define level of significance and test statistic.
- Determine critical regions for known/unknown mean and known/unknown variance.
- Describe the Student distribution
- Distinguish between one-sided and two-sided hypothesis tests.

10.1 Binary Hypothesis Testing:

In binary hypothesis testing, we are testing two possibilities – reject or accept the hypothesis. There are two possibilities for which the hypothesis is true, and two possibilities for which the hypothesis is untrue.

Consider the two models (a) null hypothesis denoted H_0 (accept hypothesis) and (b) alternate hypothesis denoted H_1 (reject hypothesis). For example, we can partition the space of all possible data sets into two regions – acceptance region and rejection region. In the test, four possible situations can occur:

- H_0 is true but we reject it and accept H_1. This is referred to as Type I error (false alarm).
- H_0 is false but we accept it and reject H_1. This is referred to as Type II error (false acceptance or missed detection.
- H_0 is true and we accept it. This is correct decision.
- H_0 is true and we accept it. This is correct decision.

Probability of False Alarm:

This is the probability of Type I error, defined as the conditional probability

$$P_{FA} = P\left[H_1 \text{ is declared true} \mid H_0\right] = \alpha$$

- The probability of committing a Type I error (false alarm), denoted as α , is called the significance level.

Probability of Miss:

This is the probability of Type II error, defined as the conditional probability

$$P_{MISS} = P\left[H_0 \text{ is declared true} \mid H_1\right] = \beta$$

- The probability of committing a Type II error, denoted as β, is called the power of the test.
- We have control of α but not β.

Definition: A plot of P_{MISS} versus P_{FA} is known as the receiver operating curve (ROC)

- It shows the trade-off between probability of miss and probability of false alarm.

Hypothesis Testing Procedure:

Following are the steps for a binary hypothesis test:

1. Identify the parameter of interest
2. Formulate the null (H_0) and alternate hypothesis (H_1).
3. Collect samples from the population of interest
4. Compute the appropriate test statistic
5. Select a significance level α, based on rejection criteria
6. Determine the critical value y_c based on the significance level
7. Establish the acceptance region, $-y_c \le Y \le y_c$ or alternatively, the critical region, $Y < -y_c$ and $Y > y_c$
8. Accept the null hypothesis (H_0) if the test statistic is in the acceptance region, otherwise reject if it is in the critical region.
9. Draw engineering conclusions.

Two-sided versus one-sided tests:

In a two-sided test, there are two critical (or rejection) regions for the alternative hypothesis– one in the right tail and the other in the left tail of the probability distribution curve.

In a one-sided test, there is only one critical (rejection) region for the alternative hypothesis– either in the right tail or in the left tail of the probability distribution curve.

Hypothesis Testing of Mean (known variance):

Two-sided test: Test of the Mean when the variance is known.
We want to test a claim that the mean value of a population is $\mu_X = \mu_0$. To test this hypothesis, we collect a set of random samples $X = \{X_1, \ldots, X_n\}$, from the population. The population variance is known to be equal to σ_X^2 .

(1) The parameter of interest is μ_0

(2) The null hypothesis is $\mu_X = \mu_0$ and the alternative hypothesis is $\mu_X \neq \mu_0$. We formulate the binary hypotheses as follows:

$$\begin{aligned} H_0: &\qquad \mu_X = \mu_0 \\ H_1: &\quad \mu_X \neq \mu_0 \ \ (\mu_X < \mu_0 \text{ or } \mu_X > \mu_0) \end{aligned}$$

(3) We Collect n independent samples, $X_1, \ldots, X_n$ from the population of interest

(4) Since we are testing the mean, the appropriate test statistic if the sample mean $\bar{X}$

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

In accordance with the central limit theorem, $\bar{X}$ is a general Gaussian random variable. We compute the normalized test statistic Y, which a standard Gaussian variable,

$$Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

This normalization enables us to use the Q-function table.

(5) We select a significance level by setting the Type I error probability α, to some very low value. We find the critical value y_c, that satisfies the following probability:

$$P\left[-y_c \le Y \le y_c\right] = 1-\alpha = 1-2Q\left(y_c\right) \Rightarrow Q\left(y_c\right) = 0.5\alpha$$

We then look up the value of y_c for which $Q\left(y_c\right) = 0.5\alpha$.

(6) We establish the acceptance region for the sample mean as follows:

$$-y_c \le Y = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \le y_c \quad \Rightarrow$$

$$\boxed{\mu_0 - \frac{\sigma_X}{\sqrt{n}} y_c \le \overline{X} \le \mu_0 + \frac{\sigma_X}{\sqrt{n}} y_c}$$

(7) We make the decision as follows:

- Accept μ_0 (hypothesis H_0) if the value of the sample mean $\overline{X}$ falls inside the acceptance region.
- Reject μ_0 (hypothesis H_1) if the value of the sample mean $\overline{X}$ falls in the critical region.

<u>**Example**</u>: Suppose a manufacturing line produces resistors that are supposed to be 10Ω. Ten resistors are taken from the production line and measured, with the following results:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
9.86	9.90	9.93	9.95	9.96	9.97	9.98	10.01	10.02	10.04

Assume each measurement is the actual resistance, $\mu_X = R$, plus a Gaussian measurement error that has mean zero and a variance of 0.1. Test the hypothesis that the resistance is $10\ \Omega$ versus the hypothesis that it is not $10\ \Omega$ at a significance level of 0.05 (95% confidence).

<u>**Solution:**</u> The hypotheses are

$$\begin{aligned} H_0 &: \quad \mu_X = 10 \\ H_1 &: \quad \mu_X \ne 10 \ \ (\mu_X < 10 \text{ or } \mu_X > 10) \end{aligned}$$

Test statistic: The sample mean is

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 9.962$$

Acceptance (or Critical) interval: The acceptance interval for the sample estimate is obtained as

$$-y_c \le Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le y_c \;\Rightarrow\; \mu_0 - \frac{\sigma_X}{\sqrt{n}} y_c \le \bar{X} \le \mu_0 + \frac{\sigma_X}{\sqrt{n}} y_c$$

For $\alpha = 0.05$, we find from the Q-function table that $y_c \simeq 1.975$.

$$10 - \frac{\sqrt{0.1}}{\sqrt{10}} y_c \le \bar{X} \le 10 + \frac{\sqrt{0.1}}{\sqrt{10}} y_c \;\Rightarrow\; 9.8025 \le \bar{X} \le 10.1975$$

Decision: Since $\bar{X} = 9.962$ falls inside this acceptance interval, we accept the hypothesis that the resistance is 10Ω.

Example: A batch of 100 resistors have an average of 102Ω . Assuming a population standard deviation of 8Ω, test whether the population mean is 100Ω at a significance level 0.95 $(\alpha = 0.05)$.

Solution:

Hypotheses: $\begin{cases} H_0: & \mu_X = 100 \\ H_1: & \mu_X \ne 100 \end{cases}$

Test statistic: For the test statistic, we compute the sample mean,

$$\bar{X} = \frac{1}{100}\sum_{i=1}^{100} X_i = 102$$

Acceptance region: For $\alpha = 0.05$, $Q(y_c) = 0.5\alpha = 0.025$ and from the Q-function table $y_c \simeq 1.96$

$$100 - \frac{8}{10}\times 1.96 \le \bar{X} \le 100 + \frac{8}{10}\times 1.96 \;\Rightarrow\; 98.432 \le \bar{X} \le 101.568.$$

Decision procedure: Since $\bar{X} = 102$ is outside the acceptance region $98.432 \le \bar{X} \le 101.568$, we reject the hypothesis.

Example: We have a production line of resistors that are supposed to be $100\ \Omega$. The standard deviation is assumed to be $\sigma_X = 8\ \Omega$. We are given the acceptance region for the sample mean as $98 \le \overline{X} \le 102$ and a sample size of 100. Find the corresponding significance level or false alarm probability α.

Solution:

Hypotheses: $\begin{cases} H_0: & \mu_X = 100 \\ H_1: & \mu_X \neq 100 \ \ (\mu_X < 100 \text{ or } \mu_X > 100) \end{cases}$

Test statistic: We first compute the sample mean $\overline{X}$,

$$\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i$$

We know that the sample mean has a Gaussian distribution with standard deviation $\sigma_X/\sqrt{n}$. The probability of Type I error is the sum of the areas under the two tails (the critical regions), that is,

$$\begin{aligned}\alpha &= P\left[\overline{X} < 98\right] + P\left[\overline{X} > 102\right] = 1 - P\left[\overline{X} > 98\right] + P\left[\overline{X} > 102\right] \\ &= 1 - Q\left(\frac{98-100}{8/\sqrt{100}}\right) + Q\left(\frac{102-100}{8/\sqrt{100}}\right) = 2Q\left(\frac{2}{8/\sqrt{100}}\right) = 2Q(2.5) = 0.0124\end{aligned}$$

Acceptance region:

$$100 - \frac{8}{10}\times 1.96 \le \overline{X} \le 100 + \frac{8}{10}\times 1.96 \quad \Rightarrow \quad 98.432 \le \overline{X} \le 101.568.$$

Note that $\overline{X} = 102$ is outside the acceptance region, $98.432 \le \overline{X} \le 101.568$, so this hypothesis would have been rejected.

One-sided Test of Sample Mean (variance known):

There are two possible one-sided hypothesis testing problems. The first one is

$$\begin{cases} H_0: & \mu_X = \mu_0 \\ H_1: & \mu_X > \mu_0 \end{cases}$$

This is a one-sided test where the acceptance region is on the left side (critical region is on the right side) of the test statistic. For this hypothesis, the acceptance region is

$$\overline{X} \le \mu_0 + \frac{\sigma_X}{\sqrt{n}} \mathrm{y}_c$$

The second one is

$$\begin{cases} H_0: & \mu_X = \mu_0 \\ H_1: & \mu_X > \mu_0 \end{cases}$$

This is a one-sided test where the acceptance region is on the right side (critical region is on the left side) of the test statistic.

For the above hypothesis the acceptance region is

$$\overline{X} \ge \mu_0 - \frac{\sigma_X}{\sqrt{n}} \mathrm{y}_c$$

Example: A quality control engineer finds that a sample of 100 light bulbs had an average life-time of 470 hours. Assuming a population standard deviation of 25 hours, test whether the population mean is $\mu_X = 48$ hours versus the alternate hypothesis $\mu_X < 48$ hours at a significance level of $0.95 \Rightarrow \alpha = 0.05$.

Solution:

Hypotheses: $\begin{cases} H_0: & \mu_X = 480 \\ H_1: & \mu_X < 480 \end{cases}$

Test statistic: For the test statistic we first compute the sample mean $\overline{X}$,

$$\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i = 470$$

Acceptance region: For $\alpha = 0.05$, $Q(y_c) = \alpha = 0.05$ and from the Q-function table $y_c = 1.65$.

$$\overline{X} \ge 480 - 1.65 \times \frac{25}{10} \quad \Rightarrow \quad \overline{X} \ge 475.875$$

Decision: Since $\overline{X} = 470$ is outside the acceptance region, we reject the hypothesis.

Exercise: A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilograms. Suppose a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Test the hypothesis that the mean is 8

kilograms versus the alternative that the mean is not 8 kilograms. Use a significance level of $\alpha = 0.01$.

Tests Concerning Sample Mean (Variance unknown):

Hypotheses: $\begin{cases} H_0: & \mu_X = \mu_0 \\ H_1: & \mu_X > \mu_0 \end{cases}$

The test statistic: If the standard deviation is known then the normalized test statistic $\bar{Y}$, is Gaussian,

$$Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

However, since the standard deviation is unknown, it must be estimated. The unbiased estimate is the sample variance given by

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \Rightarrow S_X = \sqrt{S_X^2}$$

In this case the normalized test statistic is given by

$$T = \frac{\bar{X} - \mu_0}{\mathrm{S}_X/\sqrt{n}}$$

The normalized test statistic T, in this case, satisfies the Student t-distribution with $v = n-1$ degrees of freedom.

- The student t-distribution arises when we are estimating the mean and the variance is unknown.
- If we use n samples to estimate the mean and variance, then the degrees of freedom is $v = n-1$
- Like the Q-function, the t-distribution is symmetric with respect to zero. Therefore, we may define the significance level as

$$1 - P\left[-t_c \le T \le t_c\right] = \alpha \Rightarrow P\left[T > t_c\right] = \frac{\alpha}{2}$$

The t-distribution is also tabulated and so the parameter t_c may be obtained from the table in a similar way as the Q-function. The acceptance region is given by

$$\boxed{\mu_0 - \frac{S_X}{\sqrt{n}} t_c \le \bar{X} \le \mu_0 + \frac{S_X}{\sqrt{n}} t_c}$$

For a given significance level α, we look up the value for t_c and determine the acceptance region. We accept the hypothesis if $\bar{X}$ falls inside the acceptance region, otherwise, we reject the hypothesis.

Example: Test the hypothesis that the average content of containers of a particular lubricant is 10 liters if the contents of a random sample of 10 containers are $10.2,\ 9.7,\ 10.1,\ 10.3,\ 10.1,\ 9.8,\ 9.9,\ 10.4,\ 10.3,\ 9.8$ liters. Use a confidence level of $0.99\ (\alpha = 0.01)$ and assume that the distribution of contents is Gaussian.

Solution: $\alpha = 0.01$ **and** $v = 9$. For $P[T > t_c] = \frac{\alpha}{2} = 0.005$, from t-tables, $t_c = 3.25$

$$\begin{cases} H_0: & \mu_X = 10, \\ H_1: & \mu_X \neq 10. \end{cases}$$

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 10.06$$

$$S_X = \sqrt{\frac{1}{9}\sum_{i=1}^{10}\left(X_i - 10.06\right)^2} = 0.246$$

$$10 - \frac{S_X}{\sqrt{n}} t_c \le \bar{X} \le 10 + \frac{S_X}{\sqrt{n}} t_c$$

$$10 - \frac{0.246}{\sqrt{10}} \times 3.25 \le \bar{X} \le 10 + \frac{0.246}{\sqrt{10}} \times 3.25 \Rightarrow 9.75 \le \bar{X} \le 10.25$$

Since the value of $\bar{X}$ falls within the critical interval, we accept the hypothesis that $\mu_X = 10$ liters.