EE4491
Probability Theory and Applications

A/P Li Kwok Hung

EEE-S1-B1b-62

Tel: 6790-5028

ekhli@ntu.edu.sg

EE4491

Probability Theory and Applications

Course Overview

Course Overview

Academic Units	3 (Total 39 contact hours)
General Structure	3 hours of Lectures per week (including flexible tutorial hours)
Pre-requisite	EE2010/IM2004 *Signals and Systems*
Description	The course consists of three main modules:
	1. Probability and Random Variables (Part A) 2. Random Processes, Correlation Functions and Spectral Densities (Part B) 3. Applications to Estimation and Detection (Part C)
	Probability will be introduced using the relative-frequency approach and the axiomatic approach. Both continuous and discrete random variables will be studied. Some elements of statistics will also be included.

Course Overview (cont’d)

Learning Outcomes	Upon the completion of the course, you will be able to
	1. Tell the difference between the relative-frequency approach and the axiomatic approach in probability and conditional probability; 2. Explain the concepts of continuous and discrete random variables; 3. Compute mean values, expected values and moments; 4. Find the sample mean and the sample variance; 5. Describe the difference between random variables and random processes; 6. Evaluate the correlation functions and spectral densities of random processes; 7. Determine the minimum-cost estimation and maximum-likelihood estimation; 8. Study signal detection in binary communication and performance of the optimum receiver.

Course Overview (cont'd)

Continuous Assessment (CA)	40%	Assignment – 10% Quiz 1 in week 6 – 10% Short Project – 10% Quiz 2 in week 10 – 10%
Final Examination	60%	

Note: Due to the Covid-19 issue, the weighting % of CA and exam may be changed. You will be informed when adjustments are necessary.

E4491 Weekly Study Guide

Week no.	Topics
1	Definition of Probability, Relative Frequency and Axiomatic Approaches
2	Conditional Probability, Independence, Combined Experiments
3	Concepts of Random Variables, Discrete Random Variables, Continuous Random Variables
4	Distribution Functions, Density Functions, Expected Values
5	Two and more Random Variables, Conditional Probability, Independence, Correlation
6	Sample Mean, Sample Variance, Sampling Distribution
7	Random Variables and Random Processes

E4491 Weekly Study Guide (cont'd)

Week no.	Topics
8	Continuous-Time and Discrete-Time Random Processes, Stationary and Nonstationary Random Processes
9	Autocorrelation Functions, Cross-Correlation Functions, Spectral Density, Relation of Spectral Density to Autocorrelation Function
10	Optimum Estimation, Minimum-Cost Estimation, Maximum-Likelihood Estimation
11	Linear Estimation, Mean-Square Error, Principle of Orthogonality, Linear Least-Square Estimation
12	Binary Communication, the Neyman-Pearson Criterion for Signal Detection
13	Detection in Gaussian Noise, Performance of Optimum Receiver

Recommended Readings

Textbook

A Papoulis and S U Pillai, *Probabilities, Random Variables, and Stochastic Processes*, 4^{th} ed., McGraw-Hill, New York 2002.

References

1. G R Cooper and C D McGillen, *Probability Methods of Signals and System Analysis*, 3^{rd} ed., Oxford University Press, 1999.
2. B P Lathi and Z Ding, *Modern Digital and Analog Communication Systems*, International 4^{th} ed., Oxfold University Press, 2010.
3. R D Yates and D J Goodman, *Probability and Stochastic Processes*, 2^{nd} ed., John Wiley, 2005.

Random Experiments and Events (1)

- The term ***experiment*** is used in probability theory to describe a process whose outcome cannot be fully predicted. Very often, the conditions under which it is performed cannot be predetermined with sufficient accuracy and completeness.
- Some examples of such experiments are tossing a coin, rolling a die, and drawing a card from a deck.

Random Experiments and Events (2)

- An experiment may have several identifiable ***outcomes***. For example, rolling a die has six possible outcomes (1, 2, 3, 4, 5, and 6).
- An ***event*** is a subset of outcomes that share some common characteristics. An event occurs if the outcome of the experiment belongs to the specific subset of outcomes defining the event.
- In the experiment of rolling a die, for example, the event “even number on a throw" can result from any one of three outcomes (viz., 2, 4, and 6).

Example 1

Experiment	Possible Outcomes
Tossing a coin	Heads (H), tails (T)
Throwing a die	1, 2, 3, 4, 5, 6
Throwing a die	Even, odd
Drawing a card	Any of the 52 possible cards
Drawing a card	Less than or equal to 10, greater than 10
Observing a voltage	Greater than or equal to 0, less than 0
Observing a voltage	Greater than or equal to 2 volts, less than 2 volts
Observing a voltage	Between V_1 and V_2, not between V_1 and V_2

Elementary Events and Composite Events

- An ***elementary*** event is one for which there is only one outcome.
- When a coin is tossed, the event of getting a head or the event of getting a tail can be obtained in only one way.
- Likewise, when a die is rolled, the event of getting any integer from 1 to 6 can be achieved in only one way.
- On the other hand, it is possible to define events associated with more than one elementary outcome. For example, when a die is rolled, observing an even number can be achieved in 3 different ways. This is a ***composite*** event.

Discrete and Continuous Outcomes

- When the number of outcomes of an experiment is countable, the outcomes are said to be ***discrete***.
- "Countable" means the outcomes can be placed one-to-one correspondence with the integers. Tossing a coin results in

  Heads (H) → 1

  Tails (T) → 2
- However, there are many experiments in which the outcomes are NOT countable. In the case of observing a random voltage, the outcomes are said to be ***continuous***. Note that the concept of an elementary event does not apply in this case.

Measure the Likelihood of Events

- If the outcome of an experiment is uncertain before the experiment is performed, the possible outcomes are ***random events***.
- To each random event, it is possible to assign a number, called the ***probability*** of the that event, to measure its likelihood of occurrence.
- Very often, these numbers are assumed based on our intuition about the experiment.
- For example, if we roll a die, we would expect that the possible outcomes of 1 to 6 are equally likely. Therefore, we would assume the probabilities of these 6 elementary events to be 1/6.

Definitions of Probability

- There are many different definitions for probability that have been proposed and used with varying degrees of success. They all suffer from deficiencies in concept or application.
- Of the various approaches to probability, the two that appear to be most useful are the ***relative-frequency approach*** and the ***axiomatic approach***.
- The relative-frequency approach is useful because it attempts to link some physical significance to the concept of probability and makes it possible to relate probabilistic concepts to the real world.

The Relative-Frequency Approach (1)

- It is closely related to the frequency of occurrence of the defined events. For any given event, the frequency of occurrence is used to define the ***probability*** of that event.
- Usually, the obtained probabilities are based on our intuition about the experiment or on the assumption that events are equally likely.

The Relative-Frequency Approach (2)

- Consider an experiment that is performed N times. There are 4 possible outcomes, corresponding to the 4 elementary events A, B, C, and D. Let N_A, N_B, N_C, and N_D be, respectively, the numbers of times of occurrence. Obviously,

$$N_A + N_B + N_C + N_D = N \qquad (1)$$

- We now define the relative frequency of A as

$$r(A) = \frac{N_A}{N}$$

- From (1), it is clear that

$$r(A) + r(B) + r(C) + r(D) = 1 \qquad (2)$$

The Relative-Frequency Approach (3)

- Now imagine that N increases without limit. Due to a phenomenon know as ***statistical regularity***, the relative frequency tends to be a stable value, Pr(A). That is,

$$\Pr(A) = \lim_{N \to \infty} r(A) \tag{3}$$

- From the relation given above, it follows that

$$\Pr(A) + \Pr(B) + \Pr(C) + \Pr(D) = 1 \tag{4}$$

The Relative-Frequency Approach (4)

These concepts can be extended and summarized by the following statements:

1. $0 \leq \Pr(A) \leq 1$.
2. For a complete set of *M* **mutually exclusive** events,
$$\Pr(A) + \Pr(B) + \Pr(C) + \cdots + \Pr(M) = 1.$$
3. An ***impossible event*** is represented by $\Pr(A) = 0$.
4. A ***certain event*** is represented by $\Pr(A) = 1$.

Example 2

Assume that a large bin contains an assortment of resistors of different values, which are thoroughly mixed. In particular, let there be 100 resistors having a marked value of 1 Ω, 500 resistors marked 10 Ω, 150 resistors marked 100 Ω, and 250 resistors marked 1000 Ω. If one resistor is randomly pulled out, there are 4 possible outcomes corresponding to different resistor values. Determine the probability of each of these events.

Answer

$$N = 100 + 500 + 150 + 250 = 1000$$

$$\Pr(1\,\Omega) = \frac{100}{1000} = 0.1 \qquad \Pr(10\,\Omega) = \frac{500}{1000} = 0.5$$

$$\Pr(100\,\Omega) = \frac{150}{1000} = 0.15 \qquad \Pr(1000\,\Omega) = \frac{250}{1000} = 0.25$$

Example 3

Consider again the bin of resistors and specify that in addition to have different values, they also have different power ratings (1 W, 2W and 5 W) shown below. Determine the probability of each power rating.

Power Rating	$1\ \Omega$	$10\ \Omega$	$100\ \Omega$	$1000\ \Omega$	Sum of Row
1 W	50	300	90	0	440
2 W	50	50	0	100	200
5 W	0	150	60	150	360
Sum of Column	100	500	150	250	1000

__Answer__ $\Pr(1\ \mathrm{W}) = \frac{440}{1000} = 0.44$ $\Pr(2\ \mathrm{W}) = \frac{200}{1000} = 0.2$

$\Pr(5\ \mathrm{W}) = \frac{3600}{1000} = 0.36$

Observations

- Pr(10 Ω, 5 W) = 150/1000 = 0.15
- Pr(10 Ω) = 0.5 and Pr(5 W) = 0.36
- Thus,

$$\Pr(10\ \Omega, 5\ \mathrm{W}) \neq \Pr(10\ \Omega) \times \Pr(5\ \mathrm{W})$$

- Hence, the joint probability on the LHS is not the product of the marginal probabilities on the RHS in general.

Sets and Elements

- The framework of the axiomatic approach is closely related to the set theory. Let us first review some of the elementary concepts of set theory.
- A ***set*** is a collection of objects known as ***elements***. It will be denoted as

$$A = \{\alpha_1, \alpha_2, \cdots, \alpha_n\} \tag{5}$$

where the set is A and the elements are α_1, α_2, …, α_n. We may say $\alpha_i \in A, i = 1, 2, \cdots, n$.

- For example, the set A may consist of the integers from 1 to 6 so that $\alpha_1 = 1$, $\alpha_2 = 2$, …, $\alpha_6 = 6$ are the elements.

Subsets

- A ***subset*** of A is any set whose elements are also elements of A.
- For example,

$$B = \{2, 4, 6\}$$

is a subset of

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

- The general notation for indicating this relationship is $B \subset A$.
- Note that every set D is a subset of itself, i.e., $D \subset D$.

Space and Null sets

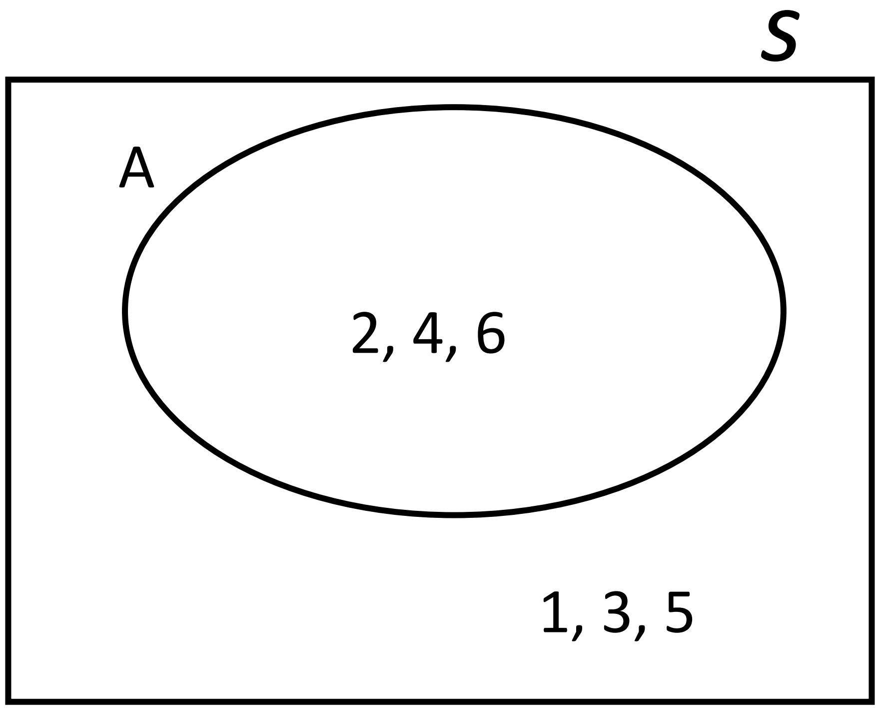
- All sets of interest in probability theory have elements taken from the largest set called a ***space*** and designated as S. Hence, all sets will be subsets of the space S.
- The ***null set*** or ***empty set*** $\varnothing$ has no elements in it. Hence, $\varnothing \subset D$, where D is any set.
- Suppose that the elements of a space consists of the 6 faces of a die. Thus,

$$S = \{1, 2, 3, 4, 5, 6\}$$

There are altogether $2^6 = 64$ subsets may be formed. (Can you list all of them?)

Venn Diagrams

- One of the reasons for using set theory to develop probability concepts is that many important operations have been developed for sets. In particular, ***Venn diagrams*** have simple graphical representations that aid in visualizing and understanding these operations.
- The space is represented by a rectangle and the sets are represented by closed plane figures.



Set Equality

- Set A and set B are ***equal*** if and only if every element of A is an element of B and every element of B is an element of A. This is,

$$A = B \;\Rightarrow\; A \subset B \text{ and } B \subset A$$

Union

- The union of two sets is a set consisting of all the elements that are elements of A or B or of both.
- Since the associativity law holds, the union of more than two sets can be written without parentheses. That is,

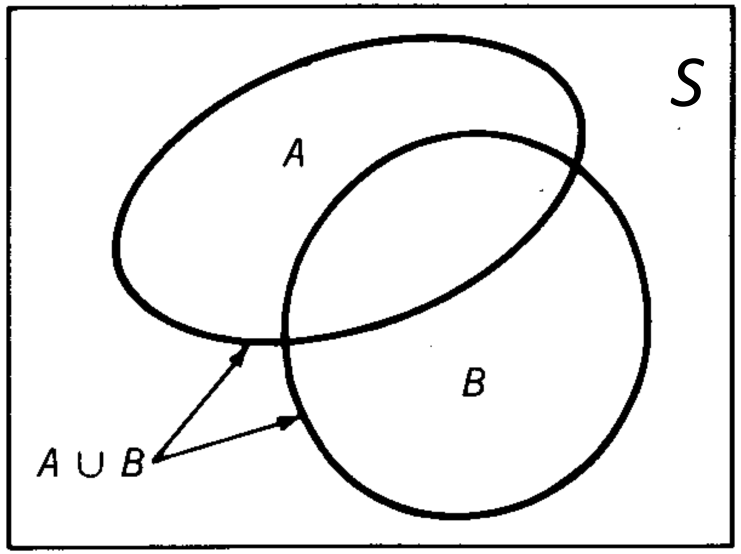
$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$$

- The commutative law also holds, so that

$$A \cup B = B \cup A, \qquad A \cup A = A$$

$$A \cup \varnothing = A, \qquad A \cup S = S$$

$$A \cup B = A \quad \text{if } B \subset A$$



Intersection (1)

- The intersection of two sets is the set consisting of all the elements that are common to both sets.

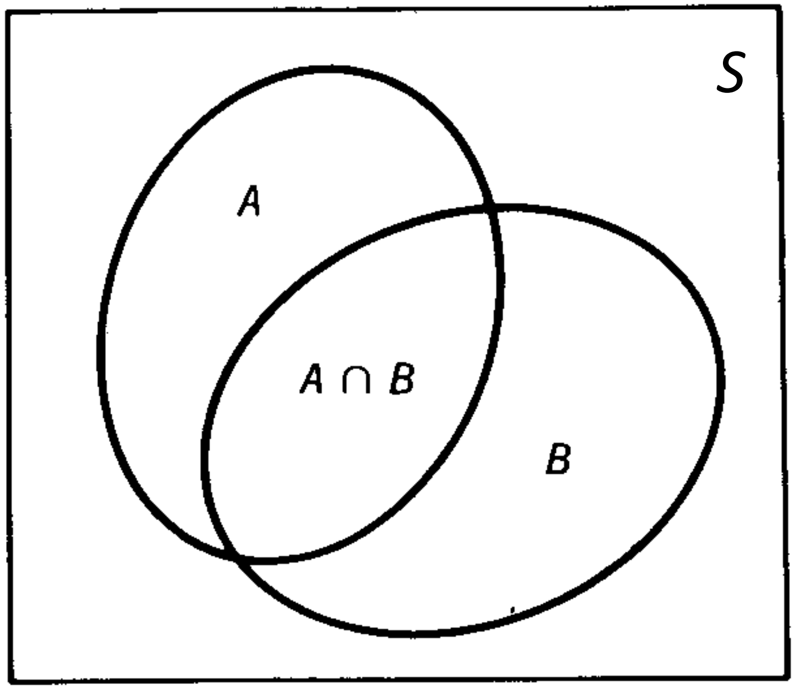
$$A \cap B = B \cap A$$

$$A \cap A = A$$

$$A \cap \varnothing = \varnothing$$

$$A \cap S = A$$

$$A \cap B = B \quad \text{if } B \subset A$$



Intersection (2)

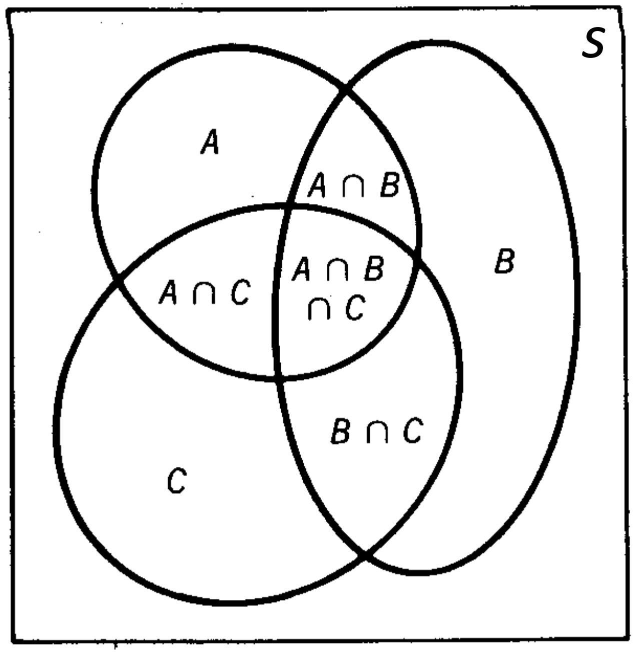
- If there are more than two sets involved in the intersection, the Venn diagram shows that

$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{(Distributive law)}$$

- Two sets *A* and *B* are ***mutually exclusive*** or ***disjoint*** if

$$A \cap B = \emptyset$$



Complement

- The ***complement*** of a set A is a set containing all the elements of S that are not in A. It is written as $\bar{A}$. It is clear that

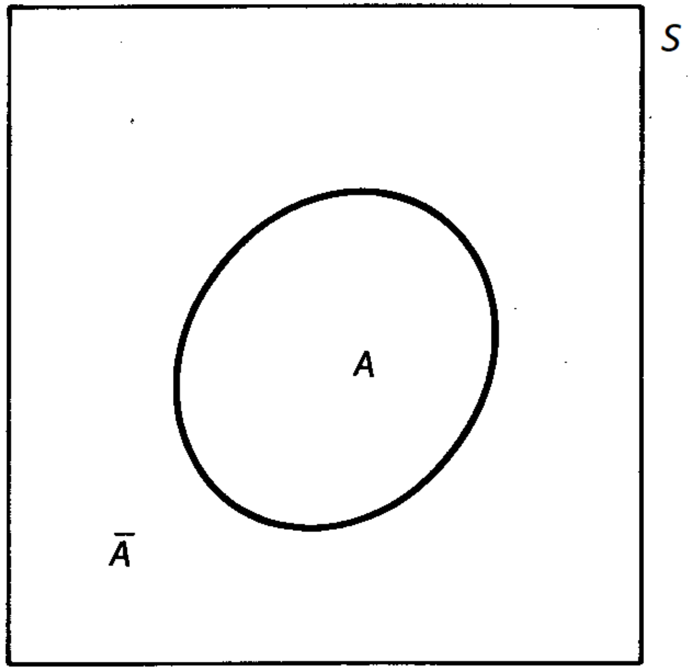
$$\bar{\varnothing} = S, \qquad \bar{S} = \varnothing$$

$$\overline{\left(\bar{A}\right)} = A, \qquad A \cup \bar{A} = S$$

$$A \cap \bar{A} = \varnothing, \quad \bar{A} \subset \bar{B} \text{ if } B \subset A$$

- Two additional relations that are usually referred to as ***DeMorgan's laws*** are

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}, \qquad \overline{(A \cap B)} = \bar{A} \cup \bar{B}$$



Differences

- The difference of two sets, $A - B$, is a set consisting of the elements in A that are not in B. The difference may also be expressed as

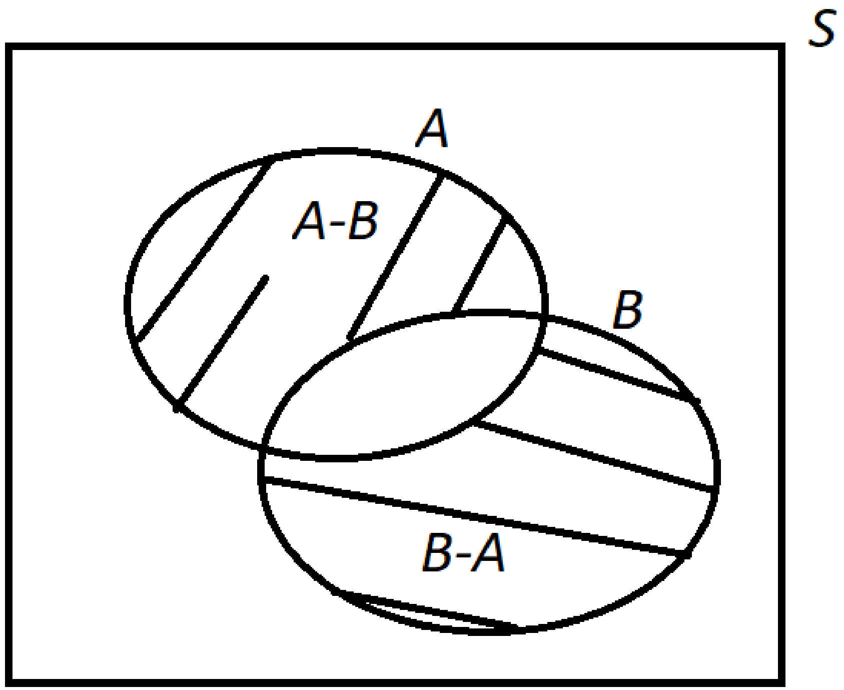
$$A - B = A \cap \bar{B} = A - (A \cap B)$$

- Using the Venn diagram, we see that

$$(A - B) \cup B = A \cup B$$

$$A - A = \varnothing, \qquad A - \varnothing = A$$

$$A - S = \varnothing, \qquad S - A = \bar{A}$$



The Axiomatic Approach (1)

- We define a ***probability space*** *S* whose elements are all the outcome from an experiment. The various subsets of *S* can be identified with events. In particular, the space *S* corresponds to the ***certain event*** and the empty set $\varnothing$ corresponds to the ***impossible event***. Any event consisting of a single element is called an ***elementary event***.
- The next step is to assign a ***probability*** to each event. The probabilities are assigned so as to satisfy the following three conditions or ***axioms***:

$$\Pr(A) \geq 0 \tag{6}$$

$$\Pr(S) = 1 \tag{7}$$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \quad \text{if } A \cap B = \varnothing \tag{8}$$

The Axiomatic Approach (2)

- Accordingly, the whole body of probability theory can be deduced from these axioms. It should be emphasized that axioms are postulates and it is meaningless to try to prove them.
- The only possible test of their validity is whether the resulting theory adequately represents the real world.

Example 4

Show that Pr(∅) = 0.

<u>Answer</u>

Since $S \cap \emptyset = \emptyset$, these two sets are mutually exclusive. Using (8), we have

$$\Pr(S \cup \emptyset) = \Pr(S) = \Pr(S) + \Pr(\emptyset)$$

Hence,

$$\Pr(\emptyset) = 0$$

Example 5

Show that $Pr(\bar{A}) = 1 - Pr(A)$.

<u>Answer</u>

Since $A \cap \bar{A} = \emptyset$, these two sets are mutually exclusive. It follows from (7) that

$$\Pr(A \cup \bar{A}) = \Pr(S) = 1$$

From (8), we see that

$$\Pr(A \cup \bar{A}) = \Pr(A) + \Pr(\bar{A})$$

Hence, $\Pr(\bar{A}) = 1 - \Pr(A)$.

Conditional Probability (1)

- The concept of conditional probability on the basis of the relative frequency of one event when another event is specified to have occurred.
- In the axiomatic approach, conditional probability is a defined quantity. If an event B is assumed to have a nonzero probability, then the conditional probability of an event A, given B, is defined as

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \qquad \text{if } \Pr(B) > 0 \tag{9}$$

- We may conveniently rewrite $\Pr(A \cap B)$ as $\Pr(A, B)$ or $\Pr(AB)$.

Observations

- If A and B are mutually exclusive, then $A \cap B = \emptyset$ and $\Pr(A \cap B) = 0$. Hence, the conditional probability is $\Pr(A|B) = 0$.
- If $A \subset B$, then $A \cap B = A$. Hence,
$$\Pr(A|B) = \frac{\Pr(A)}{\Pr(B)} \geq \Pr(A)$$
- If $B \subset A$, then $A \cap B = B$. Thus,
$$\Pr(A|B) = \frac{\Pr(B)}{\Pr(B)} = 1.$$

Conditional Probability (2)

- So far we have not shown that conditional probabilities are really probabilities.
- In the relative-frequency approach they are clearly probabilities in that they could be defined as ratios of the numbers of favourable occurrences to the total number of trials. [see slide #19.]
- In the axiomatic approach, conditional probabilities are defined quantities. Hence, it is possible to verify independently their validity as probabilities. We are going to do the proof in the next few slides.

The First and Second Axioms

- The first axiom is $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \geq 0$
- This axiom is certainly true since both $\Pr(A \cap B)$ and $\Pr(B)$ are positive. Their ratio must be positive.
- The second axiom is $\Pr(S|B) = 1$
- Note that $\Pr(S|B) = \frac{\Pr(S \cap B)}{\Pr(B)} = \frac{\Pr(B)}{\Pr(B)} = 1.$

The Third Axiom

- $\Pr(A \cup C|B) = \Pr(A|B) + \Pr(C|B)$, where $A \cap C = \emptyset$.
- The LHS can be written as

$$\Pr(A \cup C|B) = \frac{\Pr[(A \cup C) \cap B]}{\Pr(B)} \tag{10}$$

- The numerator on the RHS of (10) is

$$\Pr[(A \cup C) \cap B] = \Pr[(A \cap B) \cup (C \cap B)] \tag{11}$$

- Since $(A \cap B)$ and $(C \cap B)$ are also mutually exclusive, (11) becomes

$$\Pr[(A \cup C) \cap B] = \Pr(A \cap B) + \Pr(C \cap B) \tag{12}$$

- Substituting (12) into (10), we get the desired result.

Example 6

Let the experiment be the throwing of a single die. The outcomes are the integers from 1 to 6. Then define

$$A = \{1, 2\} \quad \textit{and} \quad B = \{2, 4, 6\}$$

Determine the conditional probabilities $Pr(A|B)$ *and* $Pr(B|A)$.

<u>Answer</u>

$$\Pr(A) = \frac{2}{6} = \frac{1}{3} \qquad \Pr(B) = \frac{3}{6} = \frac{1}{2} \qquad \Pr(A \cap B) = \frac{1}{6}$$

Example 6 (Cont'd)

The conditional probabilities are given by

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

and

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{1/6}{1/3} = \frac{1}{2}$$

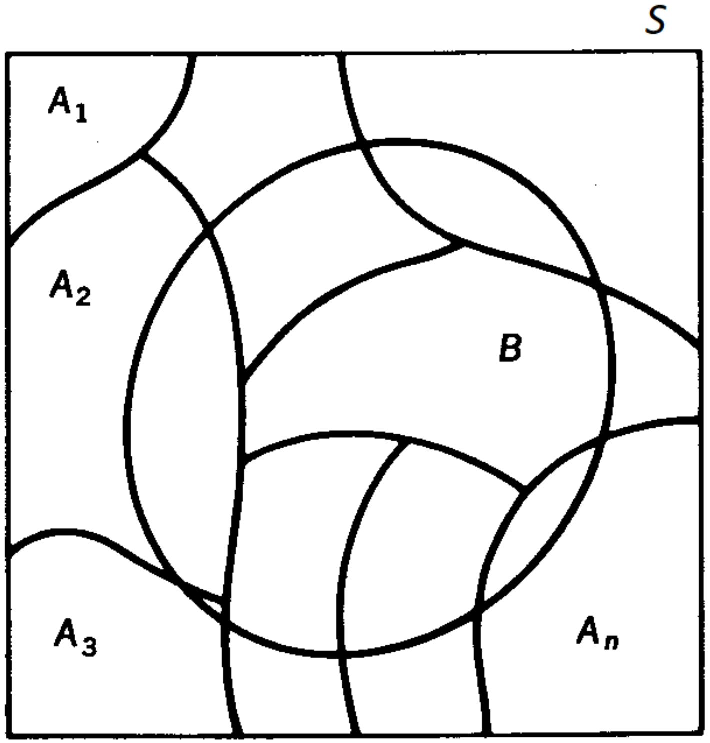
The results are intuitively correct.

Total Probabilities (1)

- Suppose there are *n* mutually exclusive events $A_1, A_2, \cdots, A_n$ and an arbitrary event *B*. The events A_i occupy the whole space *S* so that

$$A_1 \cup A_2 \cup \cdots \cup A_n = S \qquad (13)$$

- Since A_i and A_j, $i \neq j$, are mutually exclusive, it follows that $B \cap A_i$ and $B \cap A_j$ are also mutually exclusive.



Total Probabilities (2)

Hence,

$$\begin{aligned} B &= B \cap S \\ &= B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) \end{aligned}$$

are *n* mutually exclusive events. Using (8), we have

$$\begin{aligned} \Pr(B) &= \Pr(B \cap A_1) + \Pr(B \cap A_2) + \cdots + \Pr(B \cap A_n) \\ &= \sum_{i=1}^{n} \Pr(B \cap A_i) \end{aligned}$$

Total Probabilities (3)

- Note that

$$\Pr(B \cap A_i) = \Pr(B|A_i)\Pr(A_i)$$

- The final form is obtained as

$$\begin{aligned}
&\Pr(B) \\
&= \Pr(B|A_1)\Pr(A_1) + \Pr(B|A_2)\Pr(A_2) + \cdots + \Pr(B|A_n)\Pr(A_n) \\
&= \sum_{i=1}^{n} \Pr(B|A_i)\Pr(A_i)
\end{aligned} \tag{14}$$

- Pr(B) is called the ***total probability***, in terms of various conditional probabilities.

Example 7

- *Consider a resistor box consisting of 6 bins. Each bin contains an assortment of resistors as follows:*

Ohms	Bin Number						
	1	2	3	4	5	6	Row total
10 Ω	500	0	200	800	1200	1000	3700
100 Ω	300	400	600	200	800	0	2300
1000 Ω	200	600	200	600	0	1000	2600
Col total	1000	1000	1000	1600	2000	2000	8600

- *If one of the bins is selected at random, a single resistor drawn from that bin at random. What is the probability that the resistor chosen will be 10 Ω?*

Example 7 (cont'd)

<u>Answer</u>

- Let A_i be the event that the *i*-th bin is selected.

$$\Pr(A_i) = \frac{1}{6}, \quad i = 1, 2, \cdots, 6$$

- The event *B* is the event that the 10-Ω resistor is chosen.
- Using (14), we have

$$\Pr(B) = \sum_{i=1}^{6} \Pr(B|A_i)\Pr(A_i)$$

After substitutions,

$$\Pr(B) = \frac{500}{1000}\left(\frac{1}{6}\right) + \frac{0}{1000}\left(\frac{1}{6}\right) + \cdots + \frac{1000}{2000}\left(\frac{1}{6}\right) = 0.3833$$

Prior and Conditional Probabilities

- The probabilities $\Pr(A_i)$ in (14) are called ***a priori probabilities*** or ***prior probabilities*** because they describe the probabilities of the events A_i ***before*** any experiment is performed.
- ***After*** the experiment is conducted, and event B observed, the probabilities that describe the events A_i are the ***conditional probabilities*** $\Pr(A_i|B)$.

Bayes' Theorem

- Note that
$$\Pr(A_i \cap B) = \Pr(A_i|B)\Pr(B) = \Pr(B|A_i)\Pr(A_i)$$
- The second equality above can now be written as
$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B)}, \qquad \Pr(B) \neq 0 \tag{15}$$
- Substituting (14) into (15) yields
$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{j=1}^{n} \Pr\left(B|A_j\right)\Pr(A_j)} \tag{16}$$
- The conditional probability $\Pr(A_i|B)$ is often called the ***a posteriori*** probabilities because it applies after the experiment is performed. Either (15) and (16) is referred to as ***Bayes' theorem***.

Example 8

Refer to Example 7. Suppose the resistor chosen from the resistor box is found to be a 10-Ω resistor. What is the probability that it came from Bin 3?

<u>Answer</u>

Using (15), we have

$$\Pr(A_3|B) = \frac{\Pr(B|A_3)\,P(A_3)}{\Pr(B)}$$
$$= \frac{(200/1000)(1/6)}{0.3833} = 0.087$$

Exercise

Using the table in Example 7, find the probabilities:

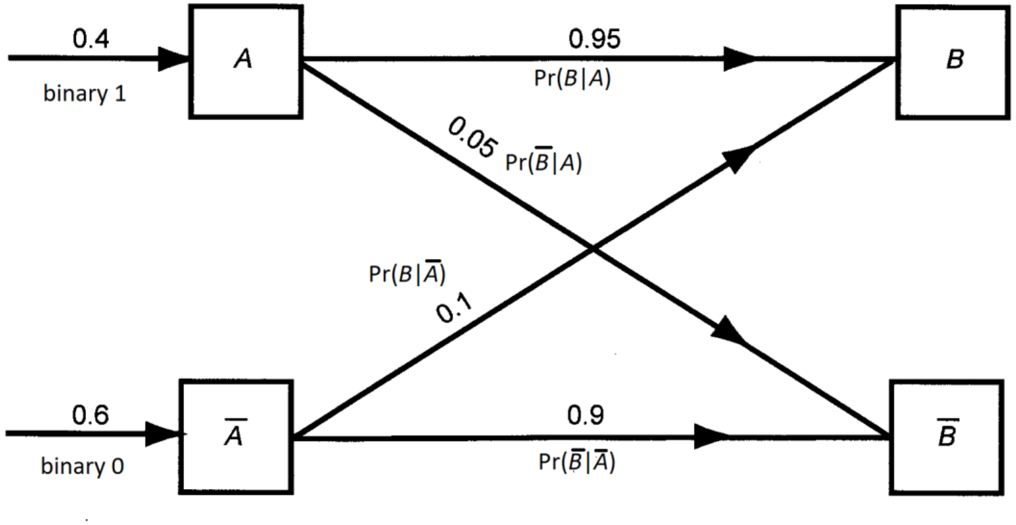
a) A 1000-Ω resistor that is selected came from bin 3.
b) A 10-Ω resistor that is selected came from bin 5.

<u>*Answer*</u>

0.1067, 0.2609

Example 9

A simple binary communication channel carries messages by using only two signals, binary 0 and binary 1. We assume that 40% of the time a binary 1 is transmitted. The probability that a transmitted binary 0 is correctly received is 0.90, and the probability that a transmitted 1 is correctly received is 0.95. Determine (a) the probability of a binary 1 being received, and (b) given a binary 1 is received, the probability that binary 1 was actually transmitted.



Example 9 (Cont'd)

<u>Answer</u>

Let A be the event that binary 1 is transmitted and $\bar{A}$ be the event that binary 0 is transmitted. Similarly, let B be the event that binary 1 is received and $\bar{B}$ be the event that binary 0 is received.

The information given in the problem statement suggests us

$$\Pr(A) = 0.4, \qquad \Pr(\bar{A}) = 1 - 0.4 = 0.6$$

$$\Pr(B|A) = 0.95, \quad \Pr(\bar{B}|A) = 0.05$$

$$\Pr(\bar{B}|\bar{A}) = 0.90, \quad \Pr(B|\bar{A}) = 0.10$$

Example 9 (Cont'd)

- For part (a) we want to find $\Pr(B)$. Using the total probability yields

$$\begin{aligned}\Pr(B) &= \Pr(B|A)\Pr(A) + \Pr(B|\bar{A})\Pr(\bar{A}) \\ &= (0.95)(0.4) + (0.1)(0.6) \\ &= 0.44\end{aligned}$$

The probability of interest in part (b) is $\Pr(A|B)$, and this can be found using Bayes' theorem:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} = \frac{(0.95)(0.4)}{0.44} = 0.863$$

Statistical Independence

- In general, the knowledge of $\Pr(A)$ and $\Pr(B)$ is not sufficient to determine $\Pr(A \cap B)$.
- This is so because $\Pr(A \cap B)$ deals with joint behavior of the two events whereas $\Pr(A)$ and $\Pr(B)$ are probabilities associated with individual events and do not yield information on their joint behavior.
- Let us then consider a special case in which the occurrence or nonoccurrence of one does not affect the occurrence or nonoccurrence of the other. In this situation events *A* and *B* are called ***statistically independent*** or simply ***independent***

Definition of Independence

- Two events *A* and *B* are said to be ***independent*** if and only if

$$\Pr(A \cap B) = \Pr(A, B) = \Pr(AB) = \Pr(A)\Pr(B) \qquad (17)$$

- Care should be taken in extending the concept of independence to more than two events. In the case of three events, A_1, A_2, and A_3, for example, they are mutually independent if and only if

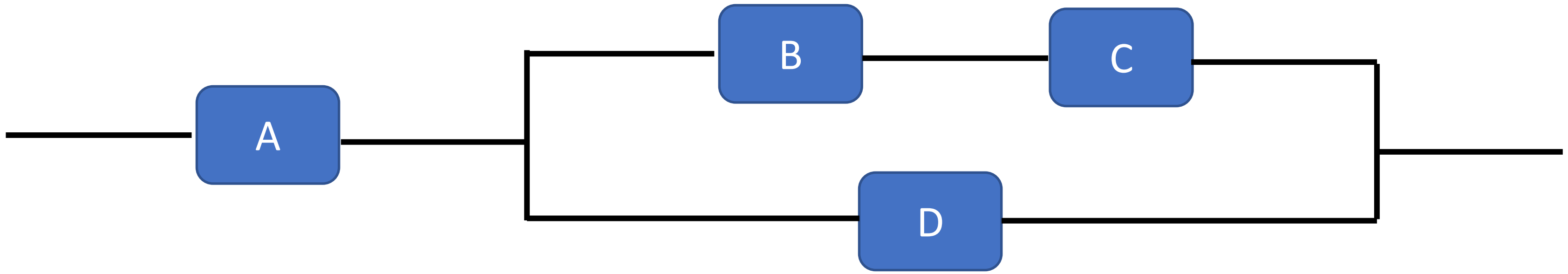
$$\Pr(A_j A_k) = \Pr(A_j)\Pr(A_k),\ \ j \neq k,\ \ j, k = 1, 2, 3 \qquad (18)$$

and

$$\Pr(A_1 A_2 A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3) \qquad (19)$$

Example 10

In a switching circuit below, the switches are assumed to operate randomly and independently. If each switch has a probability of 0.2 of being closed, find the probability that there exists a complete path through the circuit.



Example 10 (Cont'd)

- From the figure, it is easy to see that there exists a complete path if

$$A \cap [(B \cap C) \cup D]$$

- Its probability is given by

$$\Pr(A \cap [(B \cap C) \cup D]) = \Pr(\mathrm{A})\Pr[(B \cap C) \cup D]$$

- Using the result from Q4 of Tutorial no. 1, we have

$$\begin{aligned}\Pr[(B \cap C) \cup D] &= \Pr(B \cap C) + \Pr(D) - \Pr(B \cap C \cap D) \\ &= (0.2)(0.2) + (0.2) - (0.2)(0.2)(0.2) \\ &= 0.232\end{aligned}$$

- The final result is

$$\Pr(\mathrm{A})\Pr[(B \cap C) \cup D] = (0.2)(0.232) = 0.0464$$

Conditional Probability and Independence

- From the definition of conditional probability,

$$\Pr(A \mid B) = \frac{\Pr(AB)}{\Pr(B)} \qquad \text{if } \Pr(B) > 0$$

- From the definition of independence,

$$\Pr(AB) = \Pr(A)\Pr(B)$$

- From these two given conditions, we have

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B)}{\Pr(B)} = \Pr(A)$$

- That is, in the event that *A* and *B* are independent, it implies that the occurrence of *B* has no effect on the occurrence or nonoccurrence of *A*.

Example 11

- In a large number of trials of a random experiment, let n_A and n_B be, respectively, the numbers of occurrences of two outcomes *A* and *B*, and let n_{AB} be the number of times both *A* and *B* occur.
- Using the relative-frequency interpretation, the ratios n_A/n and n_B/n tend to be $P(A)$ and $P(B)$, respectively, as *n* becomes large. Similarly, n_{AB}/n tends to be $P(AB)$.
- Let us now confine our attention to only those outcomes in which *A* is realized. If *A* and *B* are independent, we expect that the ratio n_{AB}/n_A also tends to $P(B)$ as n_A becomes large. The independence assumption then leads to the observation that

$$\frac{n_{AB}}{n_A} \cong \Pr(B) = \frac{n_B}{n}$$

Example 11 (Cont'd)

This then gives

$$\frac{n_{AB}}{n} \cong \frac{n_A}{n} \times \frac{n_B}{n}$$

or, in the limit as *n* becomes large,

$$\Pr(AB) = \Pr(A)\Pr(B)$$

This is also the definition of independence.

Combined Experiments (1)

- So far, the probability space S was associated with a single experiment. The concept is too restricted to deal with many practical problems. It is necessary to extend it for more realistic cases.
- Consider a situation in which two experiments are performed. For example, one experiment might be throwing a die and the other one tossing a coin. We would like to find the probability of getting, say, a "3" on the die and a "tail" on the coin.
- In other situations the second experiment might be simply a repeated trial of the first experiment.
- The two experiments, taken together, form a ***combined experiment***.

Combined Experiments (2)

- Let the first experiment have a space S_1 and the second experiment a space S_2.
- Designate the elements of S_1 and S_2 as

$$S_1 = \{\alpha_1, \alpha_2, \cdots, \alpha_n\} \qquad \text{and} \qquad S_2 = \{\beta_1, \beta_2, \cdots, \beta_m\}$$

- These two spaces form a new space, called the ***cartesian product space***, whose elements are all the nm ordered pairs

$$(\alpha_1, \beta_1), (\alpha_1, \beta_2), \cdots, \left(\alpha_i, \beta_j\right), \cdots, (\alpha_n, \beta_m)$$

- The cartesian product space may be denoted as

$$S = S_1 \times S_2$$

to distinguish it from the previous product or intersection.

Combined Experiments (3)

- It is necessary to define the events of the new probability space. If A_1 is a subset considered to be an event in S_1 and A_2 is a subset to be an event in S_2, then $A = A_1 \times A_2$ is an event in S.
- In order to specify the probability of the event A, it is necessary to consider whether the two experiments are independent. If they are independent, the probability in the product space is simply the product of the probabilities in the original spaces. That is,

$$\Pr(A) = \Pr(A_1 \times A_2) = \Pr(A_1)\Pr(A_2)$$

- It is possible to generalize the above ideas in a straightforward manner to situations in which there are more than two experiments.

Example 12

- Consider the cartesian product space for a combined experiment of throwing a die and tossing a coin. The two spaces are

$$S_1 = \{1, 2, 3, 4, 5, 6\} \qquad \text{and} \qquad S_2 = \{H, T\}$$

- Thus, the cartesian product space $S = S_1 \times S_2$ has a total of 12 elements:

$$S = \{(1, H), (1, T), (2, H), (2, T), \cdots, (6, H), (6, T)\}$$

- Let $A_1 = \{2, 4, 6\}$ and $A_2 = \{T\}$. Then $A = A_1 \times A_2$ is an event of S. Thus, the probability of getting an even number on the die and a tail on the coin is

$$\Pr(A) = \Pr(A_1) \Pr(A_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

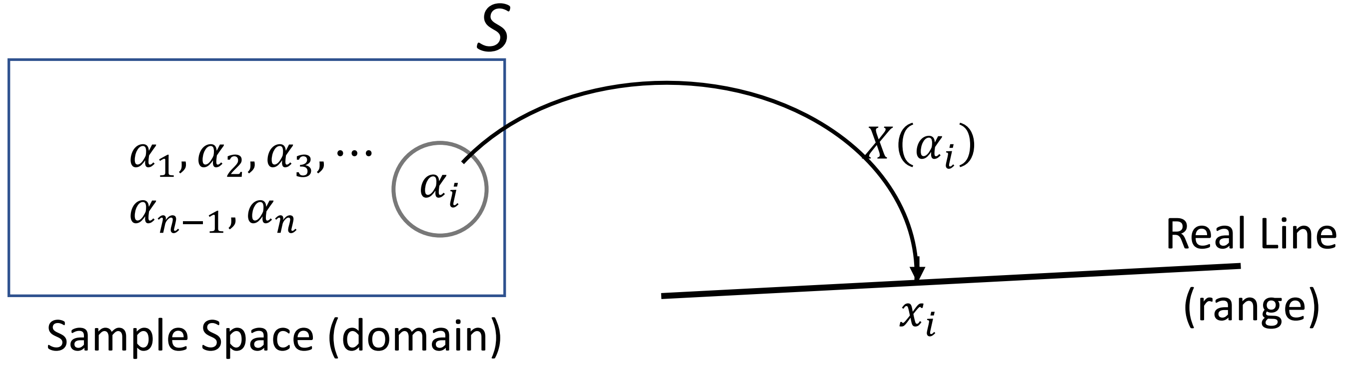
Motivations of Using Random Variables (1)

- Our interest in the study of a random phenomenon is in the statements we can make concerning the events that can occur, and these statements are made based on probabilities assigned to simple outcomes. Basic concepts have been developed in the previous studies, but a systematic and unified procedure is needed to facilitate making these statements, which can be quite complex.
- One of the solutions is to require that each of the possible outcomes of a random experiment be represented by a real number. In this way, when the experiment is performed, each outcome is identified by its assigned real number rather than by its physical description.

Motivations of Using Random Variables (2)

- For example, when the possible outcomes of a random experiment consist of success and failure, we arbitrarily assign **1** to the event '**success**' and **0** to the event '**failure**'. The associated sample space has 1, 0 as its sample points instead of success and failure, and the statement 'the outcome is 1' means 'the outcome is success'.
- This procedure not only permits us to replace a sample space of arbitrary elements by a new sample space having only real numbers as its elements but also enables us to use arithmetic means for probability calculations.
- Furthermore, most problems in science and engineering deal with quantitative measures. The real-number assignment procedure is thus a natural unifying agent.

Concept of a Random Variable



Here we expand the concept of events by mapping the space (or sample space) to the real line. Such a mapping provides a numerical designation for events so that we may use numerical valued functions instead of event (or set) functions.

Definition of Random Variables

- A random variable is a real-valued function defined over a sample space. If the sample space for the random experiment is denoted as S, then a function X that assigns a real number $X(\alpha_i)$ to every outcome or point $\alpha_i \epsilon S$ is called a ***random variable***.
- $X(\alpha_i)$ is a function whose domain is the sample space and whose range is the set of real numbers.
- We will denote the random variable by an upper-case letter (e.g., $X, Y, \text{or } Z$), and any specific value of the random variable with a lower-case letter (e.g., $x, y, \text{or } z$). Furthermore, the notation $\{X = x\}$ will denote the subset of S such that $X(\alpha) = x$.

Discrete and Continuous Random Variables

- A ***discrete random variable*** is a random variable that has a range with only a finite or countably infinite number of values on the real line. A discrete random variable is usually, but not necessarily, defined over a finite set or a countably infinite set of outcomes in the sample space.
- If a random variable can assume any value within a specified range (possible infinite), then it is designated as a ***continuous random variable***.

Example 13

Our experiment is the rolling of a die. With $f_1, f_2, \cdots, f_6$ its six faces, we do the mapping

$$X(f_i) = 10i \tag{20}$$

That is, $X(f_1) = 10, X(f_2) = 20, \cdots, X(f_6) = 60$.

Another possible mapping is

$$Y(f_i) = \begin{cases} +1, & i = 2, 4, 6 \\ -1, & i = 1, 3, 5 \end{cases} \tag{21}$$

Example 14

A telephone call occurs at random in the interval $(0, T)$. The space S consists of all numbers t in this interval. We define the random variable $X(t)$ as follows:

$$X(t) = \begin{cases} 1, & \text{if } t_1 \leq t \leq t_2 \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

where the interval (t_1, t_2) is located within $(0, T)$.

Bernoulli Trials

A large number of practical situations can be described by the repeated performance of a random experiment with the following basic characteristics: a sequence of trials is performed so that

- for each trial, there are only two possible outcomes, say, success and failure;
- the probabilities of the occurrence of these outcomes remain the same throughout the trials; and
- the trials are carried out independently.

Trials performed under these conditions are called ***Bernoulli trials***.

Binomial Distribution (1)

- A random experiment is repeated *n* times. These *n* trials form an *n*-fold combined experiment. The product space can be designated by

$$S^n = S \times S \times \cdots \times S \qquad (n \text{ times})$$

- The outcomes are *n*-tuples of outcomes of the component experiments.
- We focus on a particular Bernoulli trial in which

$$\Pr(\text{success}) = p, \;\; \Pr(\text{failure}) = 1 - p = q$$

- Suppose the success is mapped to 1 and the failure is mapped to 0. We are interested in the number of successes and failures in a succession of Bernoulli trails.

Binomial Distribution (2)

- Let A_k be the event that there are k successes and $n - k$ failures in n trails. The event A_k consists of k 1's and $n - k$ 0's in all possible n-tuple sequences.
- The number of possible sequences in A_k equals

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{23}$$

The symbol $\binom{n}{k}$ is called the ***binomial coefficient***.

- Note that the binomial theorem gives

$$\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n \tag{24}$$

Binomial Distribution (3)

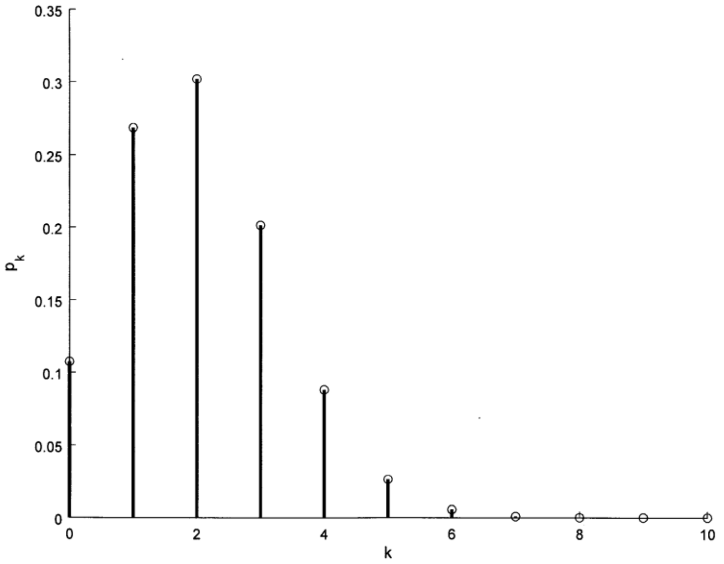
- From the previous study, we see that

$$\Pr(A_k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \cdots, n \tag{25}$$

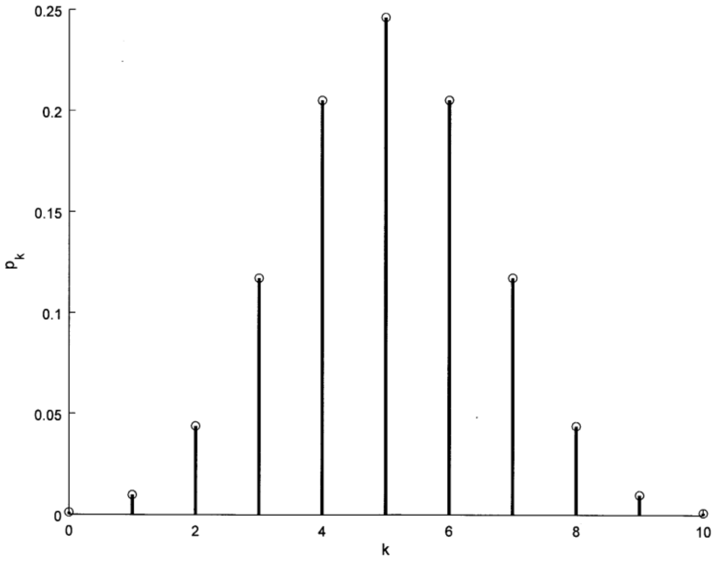
- By axioms 2 and 3, these $(n+1)$ probabilities must sum to 1 as these $(n+1)$ mutually exclusive events A_k's exhaust the whole product space S^n. To be specific,

$$\sum_{k=0}^{n} \Pr(A_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1 \tag{26}$$

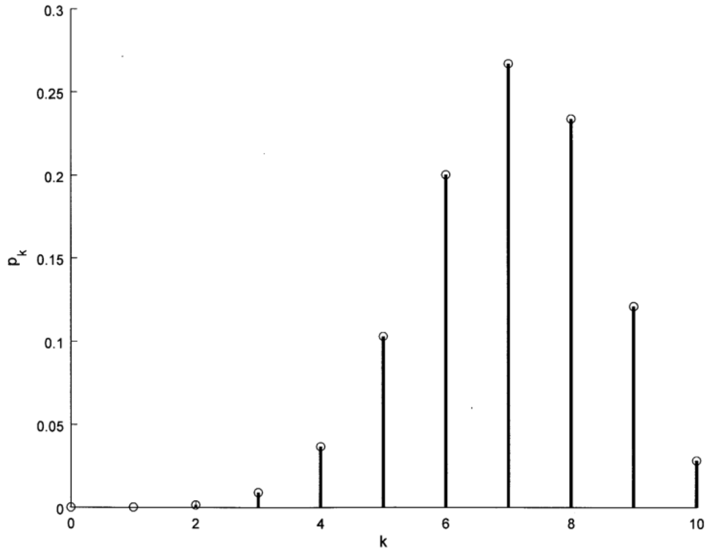
Binomial Distribution (4)



n = 10, p = 0.2



n = 10, p = 0.5



n = 10, p = 0.7

Example 15

- *A pair of fair dice are tossed ten times. What is the probability that the dice total seven points exactly four times?*

<u>Answer</u>

- The event *success* with which we are concerned consists of the six parts (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1). The probability of success is

$$p = \frac{6}{36} = \frac{1}{6}$$

- The probability of failure is thus

$$q = 1 - p = \frac{5}{6}$$

- Hence, the probability that event *A* occurs four times and $\bar{A}$ occurs six times is

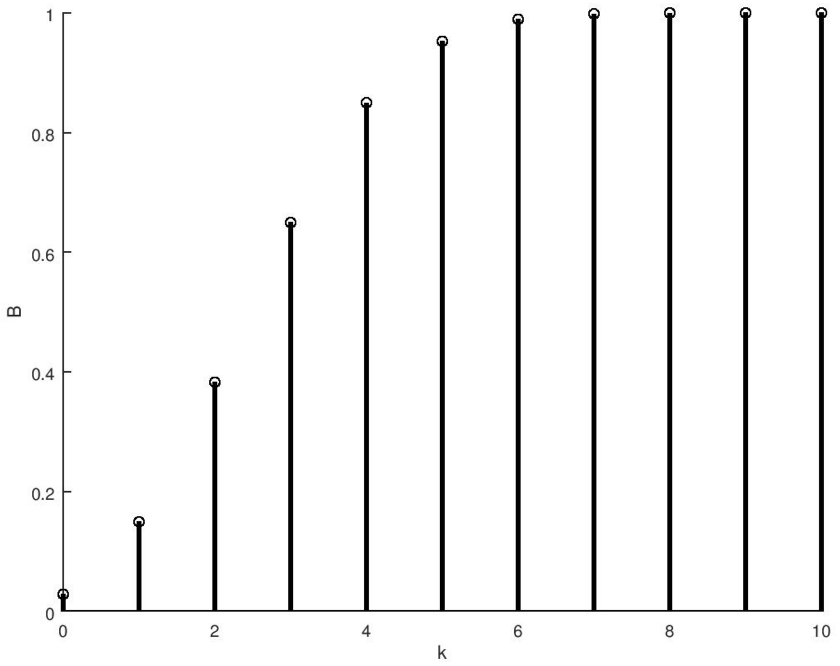
$$\Pr(\text{total seven points exactly four times}) = \binom{10}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^6 = 0.0543$$

Cumulative Binomial Distribution

- The probability that at most *k* successes occur in *n* trials is given by

$$B(k,n;p) = \sum_{r=0}^{k}\binom{n}{r} p^r q^{n-r} \tag{27}$$

- These numbers form what is called the ***cumulative binomial distribution***.
- When the numbers *n* and *k* are large, even a programmable calculator takes a long time to calculate such binomial probabilities in (27). We shall obtain a useful formula for approximating them.



$n = 10, p = 0.3$

Poisson Distribution (1)

- If the number *n* of Bernoulli trials is very large and if the probability *p* of success in each trial is very small, then the binomial probability can be approximated by

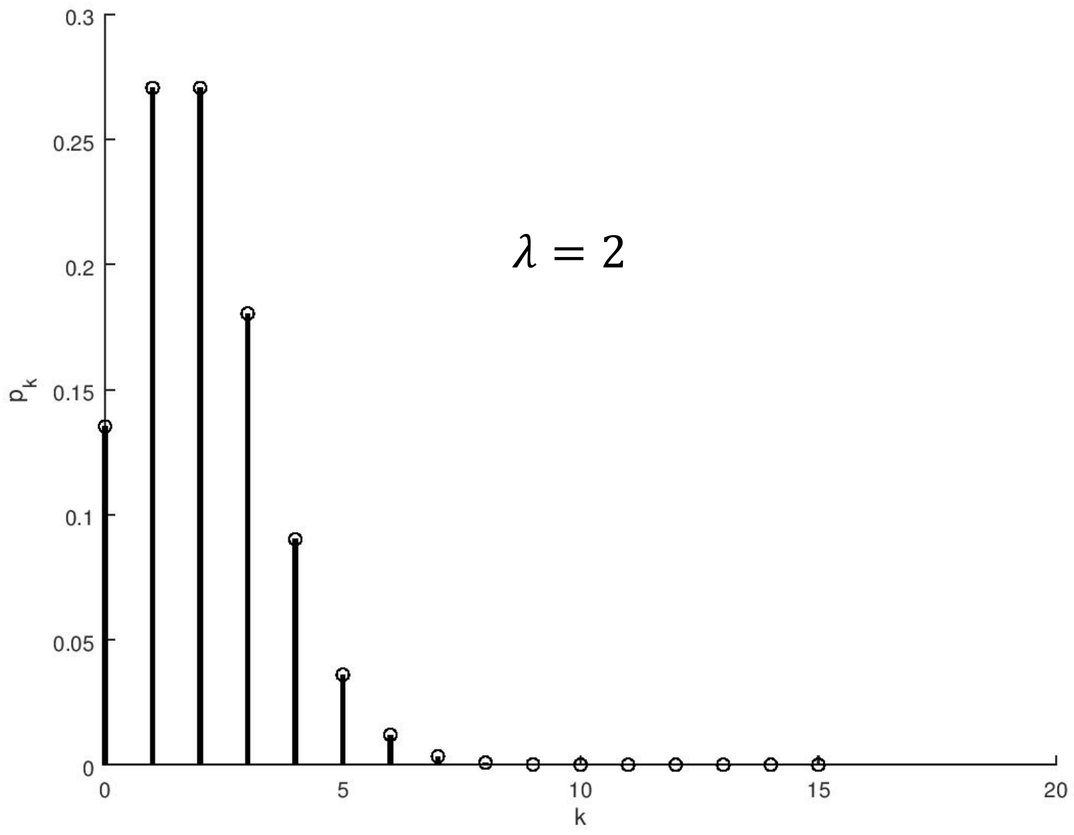
$$\Pr(A_k) = p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda = np, \qquad k = 0, 1, 2, 3, \cdots \tag{28}$$

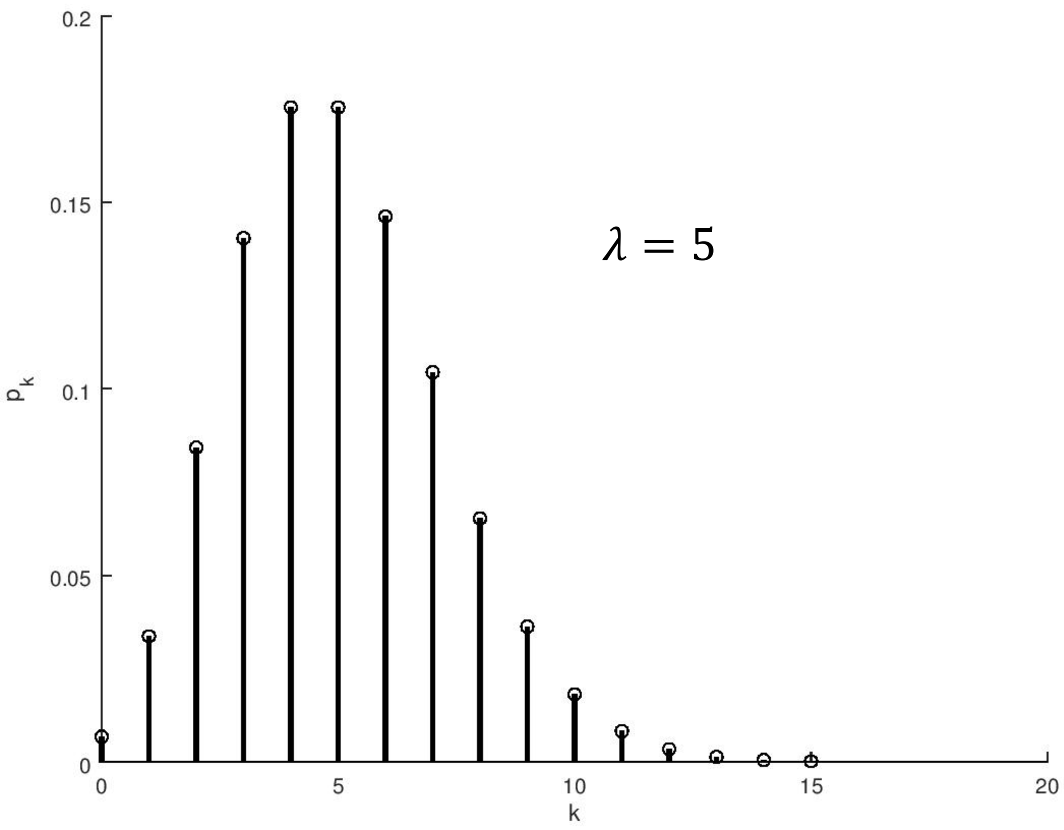
- The probabilities p_k make up the Poisson distribution. They add up to

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \times e^{\lambda} = 1$$

by virtue of the power series expansion of e^{λ} learnt in calculus. For reasons to appear later, the parameter $\lambda = np$ is called the mean or expected value of success.

Poisson Distribution (2)





Example 16

Suppose that the probability of a transistor manufactured by a certain plant being defective is 0.015. What is the probability that there is at most one defective transistor in a batch of 100?

Answer

The expected number of defective transistors in 100 is

$$\lambda = np = (100)(0.015) = 1.5$$

Therefore,

$$p_0 + p_1 = \frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} = 0.558$$

Distribution Functions (1)

- Let X be a random variable defined in S and x be any allowed value of the random variable. The ***distribution function*** or ***cumulative distribution function*** (CDF) is defined to be

$$F_X(x) = \Pr(X \leq x)$$

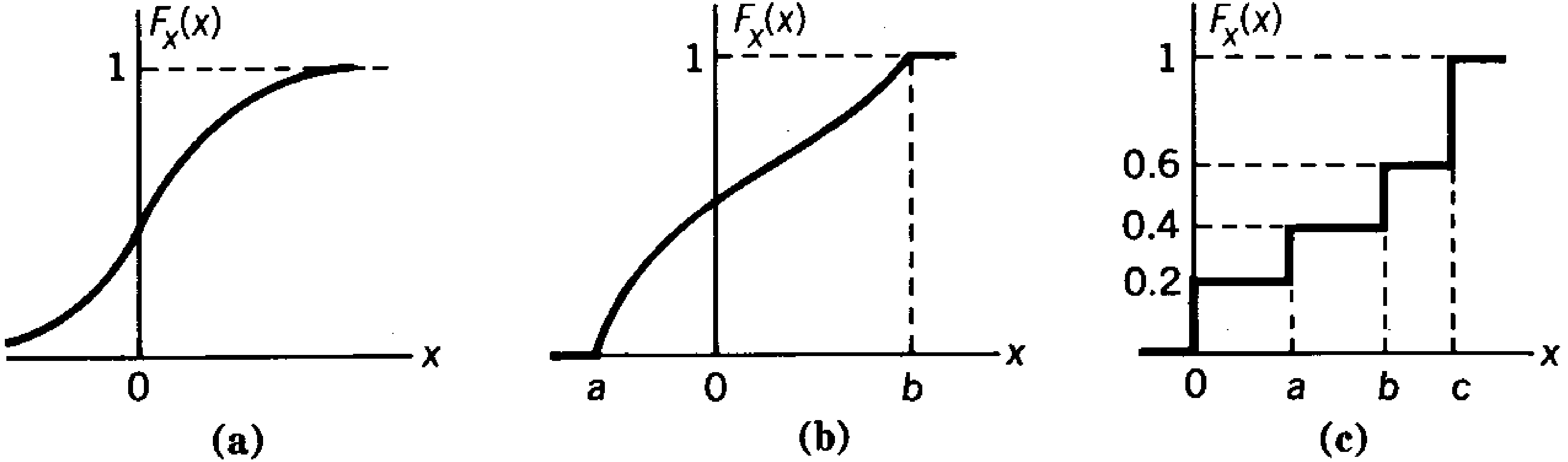
- The subscript X denotes the random variable while the argument x is a real value. In much of the subsequent discussion it is convenient to suppress the subscript X when no confusion will result.
- Since the CDF is a probability, it must satisfy the basic axioms and must have the same properties as the probabilities discussed before.

Distribution Functions (2)

The properties of CDFs are summarized as follows:

1. $0 \leq F_X(x) \leq 1 \qquad -\infty < x < \infty$
2. $F_X(-\infty) = 0$
3. $F_X(\infty) = 1$
4. $F_X(x)$ is nondecreasing as x increases
5. $\Pr(a < X \leq b) = F_X(b) - F_X(a)$

Distribution Function (3)



Continuous random variable
Ranging from $-\infty$ to $+\infty$

Continuous random variable
Ranging from a to b

Discrete random variable
Possible values are 0, *a*, *b* and *c*

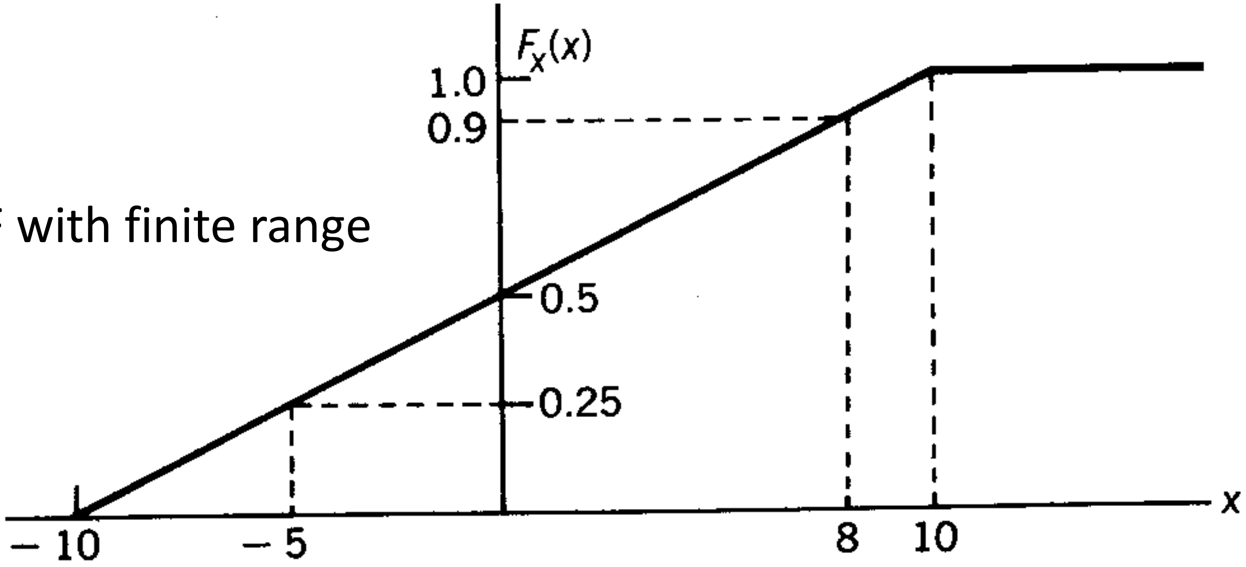
Distribution Function (4)

- The CDF can also be used to express the probability of the event that the random variable X is greater than x. It follows that

$$\begin{aligned}\Pr(X > x_1) &= 1 - \Pr(X \le x_1) \\ &= 1 - F_X(x_1)\end{aligned} \qquad (29)$$

Example 17

A CDF with finite range

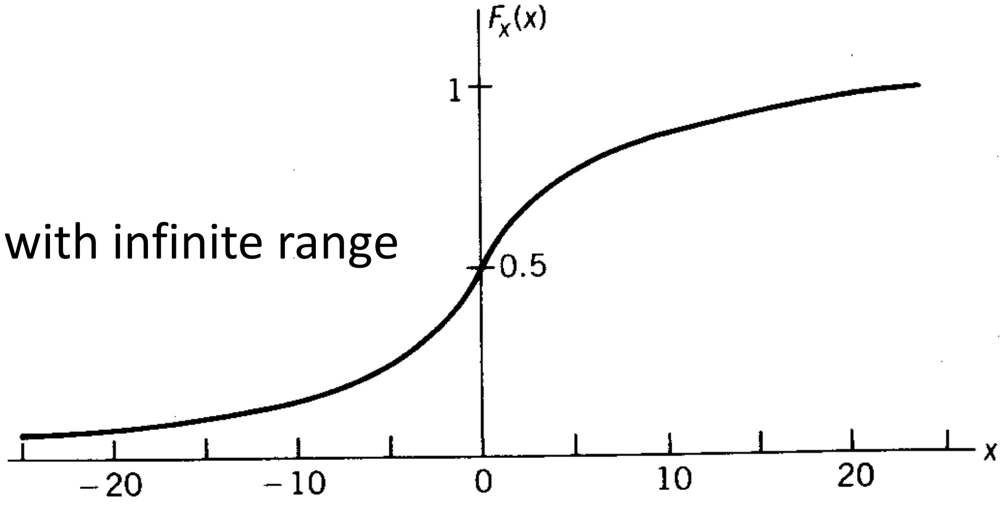


The CDF satisfies all the requirements.

$$\Pr(X \le -5) = 0.25$$
$$\Pr(X > 5) = 1 - 0.25 = 0.75$$
$$\Pr(X > 8) = 1 - 0.9 = 0.1$$
$$\Pr(-5 < X \le 8) = 0.9 - 0.25 = 0.65$$
$$\Pr(X > 0) = 1 - \Pr(X \le 0) = 0.5$$

Example 18

A CDF with infinite range



This is another CDF with infinite range.

$$\Pr(X \leq -5) = 0.25$$
$$\Pr(X > -5) = 0.75$$
$$\Pr(X > 8) = 1 - 0.8222 = 0.1778$$
$$\Pr(-5 < X \leq 8) = 0.8222 - 0.25 = 0.5722$$
$$\Pr(X > 0) = 1 - \Pr(X \leq 0) = 0.5$$

Exercise

A random experiment consists of flipping six fair coins and taking the random variable X to be the number of heads.

a) Sketch the CDF for X.

b) What is the probability that X is less than 3.5?

c) What is the probability that the random variable X is greater than 2.5?

d) What is the probability $Pr(1.5 < X \leq 5.0)$?

<u>Answer</u>

(b) 0.6563; (b) 0.6563; (c) 0.875

Exercise

A random variable X has a CDF given by

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}$$

a) *Find the probability that* $X > 0.5$.

b) *Compute the probability that* $X \leq 0.25$.

c) *Determine* $Pr(0.3 < X \leq 0.7)$.

Answer

(a) 0.6065; (b) 0.2212; (c) 0.2442

Example 19 --- Poisson Distribution

Recall the Poisson distribution,

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \lambda = np, \qquad k = 0, 1, 2, 3, \cdots$$

Let *X* be Poisson distributed. Its CDF is given by

$$F_X(x) = \Pr(X \leq x)$$

When $x < 0$,

$$F_X(x) = \Pr(X \leq x) = 0$$

When $0 \leq x < 1$,

$$F_X(x) = \Pr(X \leq x) = \Pr(X = 0) = e^{-\lambda}$$

Example 19 (Cont'd)

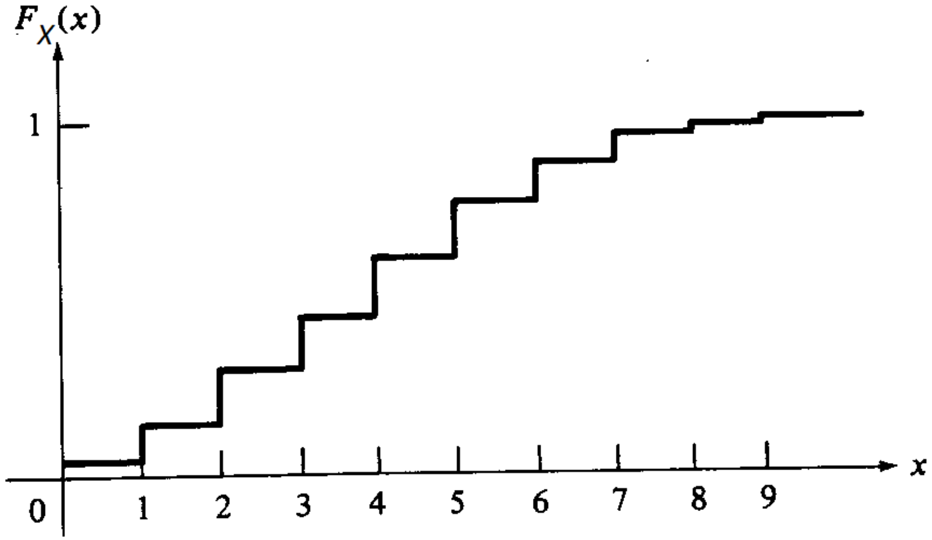
When $1 \leq x < 2$,

$$F_X(x) = \Pr(X = 0) + \Pr(X = 1)$$
$$= e^{-\lambda} + \lambda e^{-\lambda}$$

Continuing this, we see that, for $x \geq 0$,

$$F_X(x) = \sum_{k=0}^{[x]} \Pr(X = k) = \sum_{k=0}^{[x]} \frac{\lambda^k}{k!} e^{-\lambda}$$

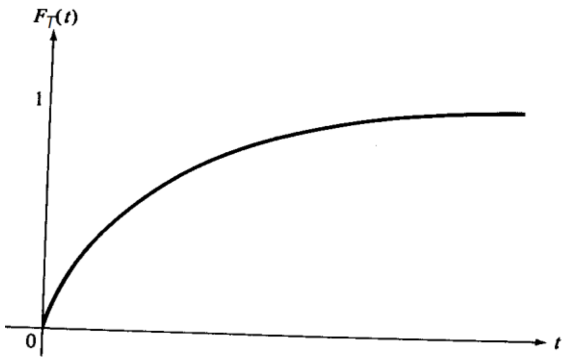
where $[x]$ stands for the greatest integer in x. Obviously, the CDF is nondecreasing with jumps at integer values of x.



Example 20

Diode of a certain kind are tested to see how long they can last before burning out under typical stressful conditions. An experiment is the test of an individual diode starting at time 0, and the outcome is the time *T* at which it fails. The CDF of this random variable is

$$F_T(t) = \Pr(T \le t) = \begin{cases} 0, & t < 0 \\ 1 - \exp(-\mu t), & t \ge 0 \end{cases}$$



for some constant μ.

The probability that its lifetime is longer than *c* is

$$\Pr(T > c) = \exp(-\mu c), \qquad c > 0$$

Density Functions (1)

- Although the CDF is a complete description of the probability model for a single random variable, it is not the most convenient form for many calculations of interest.
- For some cases, it may be preferable to use the derivative of $F_X(x)$ rather than $F_X(x)$ itself.
- The derivative is called the ***density function*** or ***probability density function*** (PDF). When exists, it is defined by

$$f_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta \to 0} \frac{F_X(x+\Delta) - F_X(x)}{\Delta} \tag{30}$$

Density Functions (2)

- The physical significance of the PDF is best described in terms of the probability element, $f_X(x)\,dx$. It is interpreted as

$$f_X(x)\,dx = \Pr(x < X \leq x + dx) \tag{31}$$

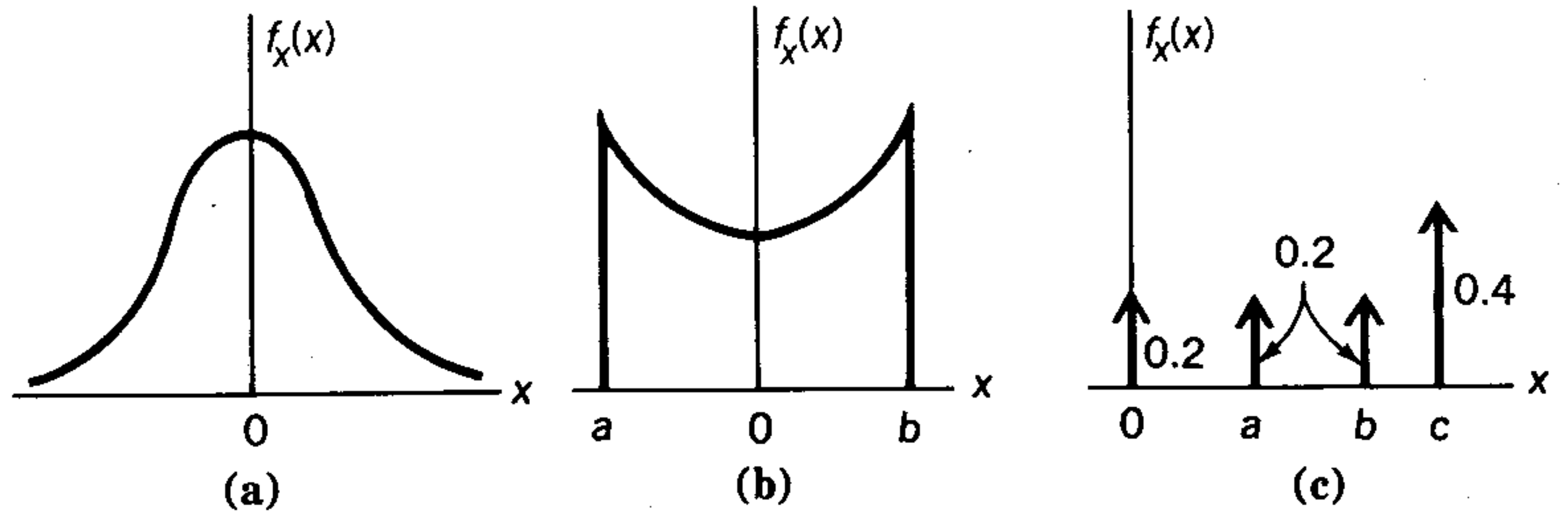
- In other words, $f_X(x)\,dx$ is the probability of the event that the random variable X lies in the interval $(x, x + dx]$.
- Since $f_X(x)$ is a density function and not a probability, it can be greater than 1. However, it must be nonnegative.
- Again, the subscripts denoting the random variable in $f_X(x)$ may be deleted when no confusion results.

Density Functions (3)

- The properties of PDFs are summarized as follows:

1. $f_X(x) \geq 0, \qquad -\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
3. $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$
4. $\int_{x_1}^{x_2} f_X(x) dx = \Pr(x_1 < X \leq x_2)$

Density Functions (4)

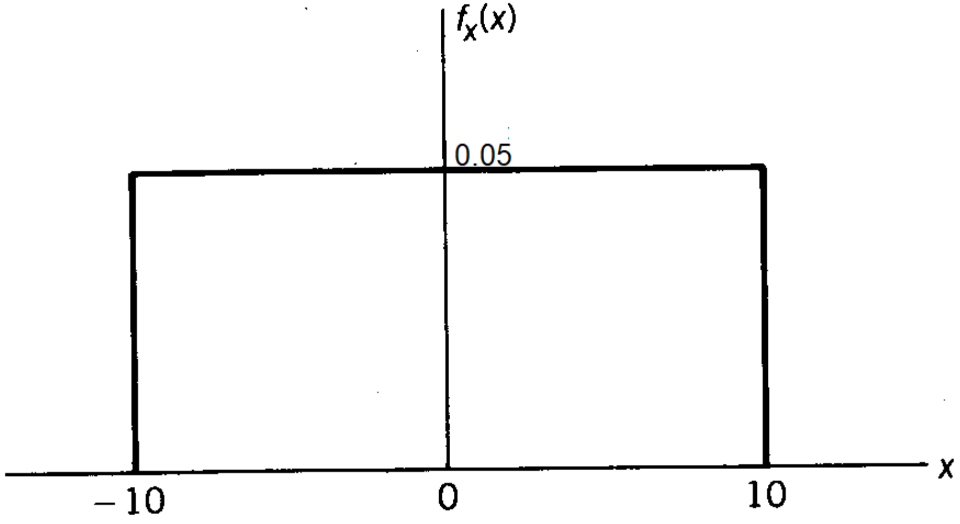


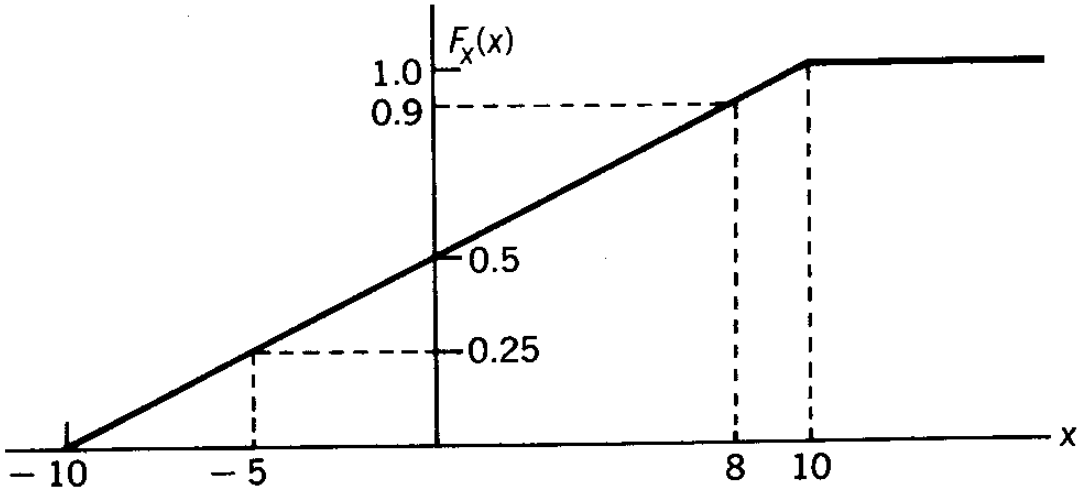
- Referring to Slide #86, the corresponding PDFs are shown here.
- Note that the PDF for a discrete random variable consists of impulses, each impulse with an area equal to the height of the jump at the same location.

Example 21 --- Uniform Distribution

The PDF of a uniformly distributed random variable is given by

$$
\begin{aligned}
f_X(x) &= 0 && x \le -10 \\
&= 0.05 && -10 < x \le 10 \\
&= 0 && x > 10
\end{aligned}
$$





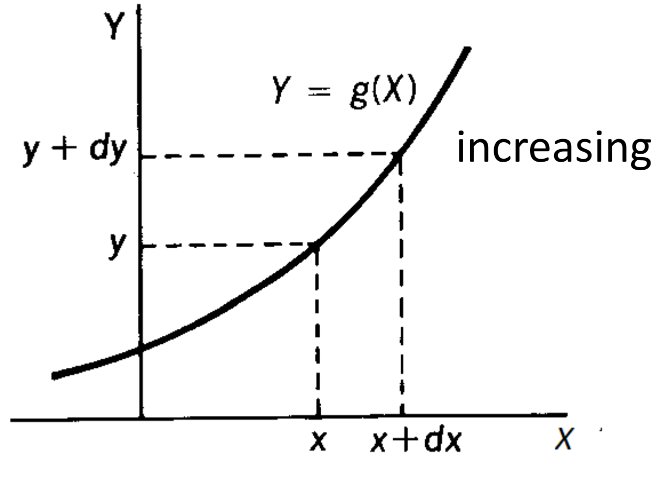
Transformation of Variables (1)

This is a common problem in the analysis of engineering systems. Suppose a random variable

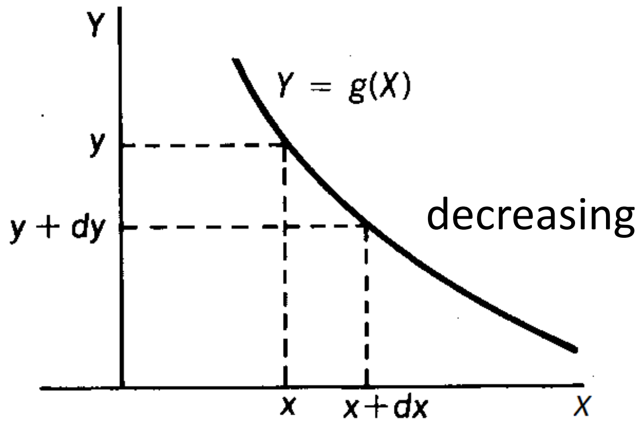
$$Y = g(X) \tag{32}$$

is functionally related to another random variable X whose PDF $f_X(x)$ is known. It is desired to determine the PDF of Y.

In order to determine $f_Y(y)$, let the random variable Y be a single-valued real function of X. It may be monotonically increasing (see Fig. (a)) or monotonically decreasing (see Fig. (b)).



(a)



(b)

Transformation of Variables (2)

Let us focus on Fig. (a) first. Whenever the random variable X lies in $(x, x + dx)$, the random variable Y will lie in $(y, y + dy)$. Hence, these two events have the same probability as

$$f_X(x)dx = f_Y(y)dy$$

The desired PDF becomes

$$f_Y(y) = f_X(x)\frac{dx}{dy} \tag{33}$$

Of course, on the RHS of (33), x must be replaced by its corresponding function of y.

Transformation of Variables (3)

When $g(X)$ is a monotonically decreasing function of x, as shown in Fig. (b), a similar result is obtained except that the derivative is negative. Since the PDFs and the probabilities must be nonnegative, (33) must be modified to yield the general form

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| \tag{34}$$

Example 22

Assume that we have a random variable X whose PDF is $f_X(x)$ is known. Consider another random variable

$$Y = -2X$$

It follows that

$$^{dy}/_{dx} = -2$$

From the previous discussion, we have

$$f_Y(y) = \frac{1}{|dy/dx|} f_X\left(-\frac{y}{2}\right) = \frac{1}{2} f_X\left(-\frac{y}{2}\right)$$

Thus , it is very easy to find the PDF of any random variable that is simply a scaled version of another random variable with known PDF.

Example 23

Assuming that the random variable X has a PDF of the form

$$f_X(x) = e^{-x}u(x)$$

where $u(x)$ is the unit step function starting at $x = 0$. Now consider another random variable Y that is related to X by

$$Y = X^3$$

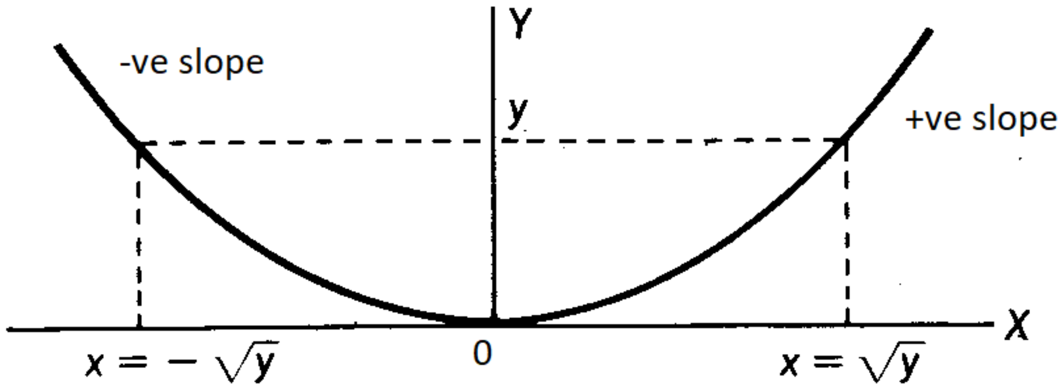
Since y and x are monotonically related, it follows that

$$\frac{dy}{dx} = 3x^2 \quad \text{or} \quad \frac{dx}{dy} = \frac{1}{3x^2} = \frac{1}{3y^{2/3}}$$

Thus, the PDF of Y is

$$f_Y(y) = \frac{1}{3y^{2/3}} e^{-y^{1/3}} u(y)$$

Example 24



Let the functional relationship be

$$Y = X^2$$

It is observed that $g(X) = X^2$ has a region in which the slope is positive and other region in which the slope is negative. In this case, the regions may be considered separately and the corresponding PDFs added.

Note that $\left|\frac{dx}{dy}\right| = \frac{1}{2\sqrt{y}}$. Since there are two roots ($x = \pm\sqrt{y}$) for every $y > 0$, the desired PDF is

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})]u(y)$$

Exercise

The PDF of a random variable has the form

$$f_X(x) = Ae^{-2x}u(x)$$

where $u(x)$ is the unit step function.

a) Find the value of A.

b) Find the probability that $X < 1$.

c) Find the probability that $X \geq 0.5$.

<u>Answer</u>

(a) $A = 2$; (b) $1 - e^{-2}$; (c) 0.3679

Exercise

A random variable Y *is related to* X *by*

$$Y = 6X + 3$$

The PDF of X *is*

$$f_X(x) = Ae^{-2x}u(x)$$

where A is a constant. Find the PDF of Y*.*

<u>Answer</u>

$$f_Y(y) = \frac{2}{|6|}e^{-2\left(\frac{y-3}{6}\right)}u\left(\frac{y-3}{6}\right) = \frac{1}{3}e^{-\frac{y-3}{3}}u(y-3)$$

Mean Values

- For random variables, it is often necessary to find the average value by integrating over the range of possible values that the random variable may assume. Such an operation is called the ***ensemble average*** or ***mathematical expectation***.

$$E\{X\} = \int_{-\infty}^{\infty} x\, f_X(x) dx = \mu_X \tag{35}$$

- The symbol $\mu_X = E\{X\}$ denotes the expected value of X.
- The expected value of any function of X, say $g(x)$, can also be obtained by

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)\, f_X(x) dx \tag{36}$$

Moments

The kth ***moment*** of the random variable X is defined as

$$E\{X^k\} = \int_{-\infty}^{\infty} x^k f_X(x)dx \tag{37}$$

When $k = 2$, we have the ***second moment*** or ***mean-square value***

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 f_X(x)dx \tag{38}$$

which is proportional to the average power and its square root is called the ***rms value***.

Central Moments

The kth ***central moment*** of a random variable X is defined as

$$E\{(X-\mu_X)^k\} = \int_{-\infty}^{\infty}(x-\mu_X)^k f_X(x)dx \qquad (39)$$

The first central moment is

$$E\{X-\mu_X\} = E\{X\} - \mu_X = 0$$

The second central moment is called the ***variance*** and is usually denoted by σ^2. Hence,

$$\sigma^2 = E\{(X-\mu_X)^2\} = \int_{-\infty}^{\infty}(x-\mu_X)^2 f_X(x)dx \qquad (40)$$

Variance

Since expectation is a linear operator, we have

$$\begin{aligned}\sigma^2 &= E\{(X-\mu)^2\} \\ &= E\{X^2 - 2\mu X + \mu^2\} \\ &= E\{X^2\} - 2\mu E\{X\} + \mu^2 \\ &= E\{X^2\} - 2\mu^2 + \mu^2 \\ &= E\{X^2\} - \mu^2\end{aligned} \tag{41}$$

It is observed that the variance is the difference between the mean-square value and the square of the mean value.

The square root of the variance, denoted as σ, is known as the ***standard derivation***.

Example 25

Consider a uniform PDF

$$f_X(x) = \begin{cases} \dfrac{1}{20}, & 20 < x \leq 40 \\ 0, & \text{otherwise} \end{cases}$$

The mean value of this random variable is

$$\mu = \int_{20}^{40} x\left(\frac{1}{20}\right)dx = \frac{1}{40}[40^2 - 20^2] = 30$$

The mean-square value is

$$E\{X^2\} = \int_{20}^{40} x^2\left(\frac{1}{20}\right)dx = \frac{1}{60}[40^3 - 20^3] = 933.3$$

Hence, the variance is $\sigma^2 = 933.3 - 30^2 = 33.3$

Exercise

Consider the PDF

$$f_X(x) = kx[u(x) - u(x-1)]$$

a) *Determine the value of* k.
b) *Calculate the expected value of* X.
c) *Find the mean-square value of* X.
d) *Compute the variance of* X.

<u>Answer</u>

(a) $k = 2;$ (b) $\mu_X = 2/3;$ (c) $E\{X^2\} = \frac{1}{2};$ (d) σ^2=1/18.

Exercise

A random variable X *has a PDF*

$$f_X(x) = \frac{1}{4}[u(x+2) - u(x-2)]$$

where $u(x)$ *is the unit step function. Suppose the random variable* $Y = X^2$.

a) *Find the mean value of* Y.

b) *Compute the mean-square value of* Y.

c) *Determine the variance of Y.*

<u>Answer</u>

(a) 4/3; (b) 16/5; (c) 64/45

Example 26 --- Standard Gaussian

Consider a PDF

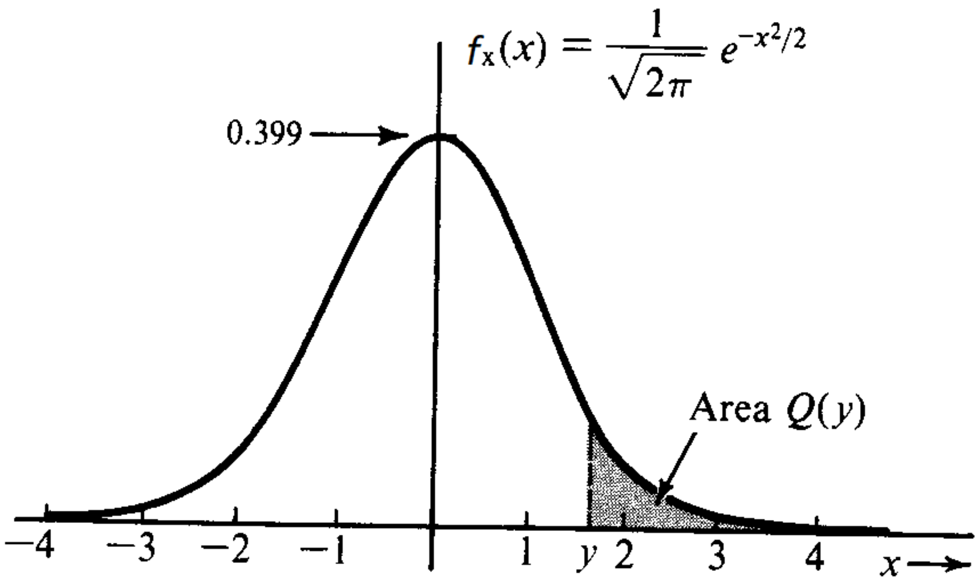
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{42}$$

This is the well-known ***standard Gaussian*** or ***standard normal*** PDF, which has zero mean and unit variance. That is,

$$\mu_X = E\{X\} = 0$$

and

$$\sigma^2 = E\{X^2\} - \mu_X^2 = E\{X^2\} = 1$$



Example 26 (Cont'd)

The CDF $F_X(x)$ is

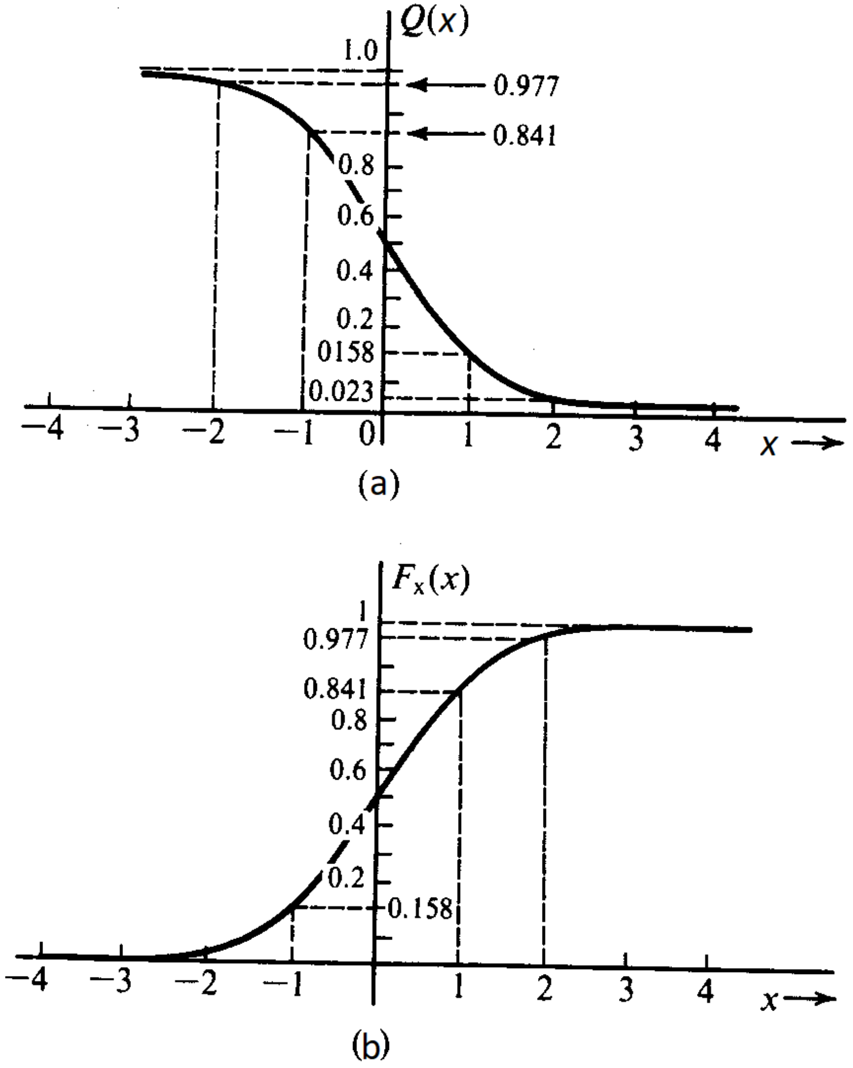
$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du \qquad (43)$$

This integral cannot be evaluated in a closed form and must be computed numerically. It is convenient to use $Q(x)$, defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-u^2/2}\, du, \qquad (44)$$

to express $F_X(x)$. Obviously,

$$F_X(x) = 1 - Q(x) \qquad (45)$$



Exercise

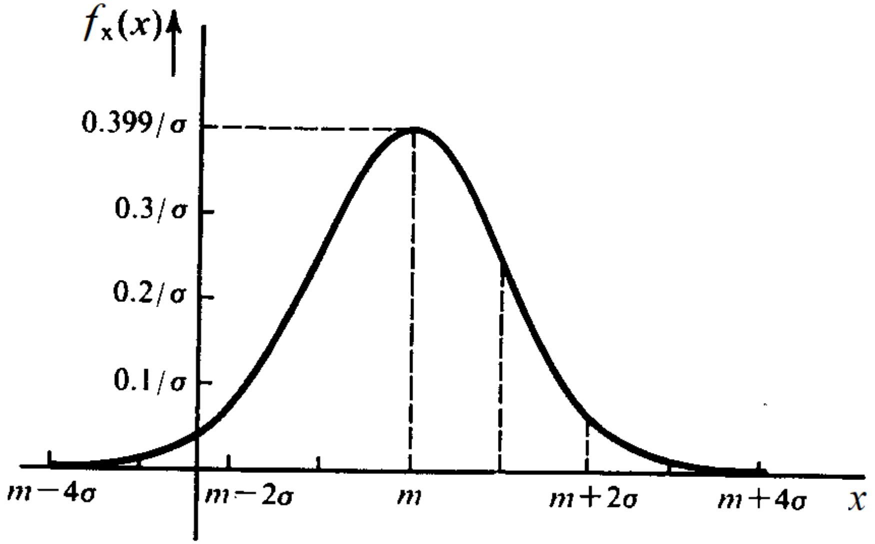
A more general Gaussian PDF is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2} \qquad (46)$$

where m is the mean value and σ^2 is the variance.

a) *Find the CDF of X in terms of $Q(x)$.*

b) *Assuming that $m > 0$. Determine $Pr(X > 0)$.*



Answer

(a) $1 - Q\left(\frac{x-m}{\sigma}\right)$; (b) $Q\left(-\frac{m}{\sigma}\right) = 1 - Q\left(\frac{m}{\sigma}\right)$

Two or More Random Variables

- In previous studies, we introduced the concept of a single random variable and its CDF and PDF. These important concepts will now be developed, with elaboration, for two random variables. The extension to more random variables will require no further effort.
- For example, we may want to find the relation between the input and output of a linear-time-invariant (LTI) system, either at the same time instant or at two different time instants.
- Another example is the light from a common source falling on two adjacent photodetectors. We count the numbers n_1 and n_2 of photoelectrons emitted by each photodetector during an interval $(0, T)$. Note that we cannot assume that n_1 and n_2 are independent.

Joint Cumulative Distribution Functions

Let the two random variables be *X* and *Y*. We define their ***joint CDF*** as

$$F_{XY}(x, y) = \Pr(X \ \leq x, Y \leq y) \tag{47}$$

Note that this is simply the probability of the event $\{X \leq x\} \cap \{Y \leq y\}$.

The properties of the joint CDF are summarized as follows:

1) $0 \leq F_{XY}(x, y) \leq 1, \qquad -\infty < x < \infty, -\infty < y < \infty$
2) $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = F_{XY}(-\infty, -\infty) = 0$
3) $F_{XY}(\infty, \infty) = 1$
4) $F_{XY}(x, y)$ is a nondecreasing function as either *x* or *y*, or both, increase
5) $F_{XY}(\infty, y) = F_Y(y), \qquad F_{XY}(x, \infty) = F_X(x)$

In item 5 above, $F_X(x)$ and $F_Y(y)$ are called ***marginal*** CDFs.

Exercise

Show that

$$\Pr(x_1 < X \leq x_2, y_1 < Y \leq y_2)$$
$$= F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1) \qquad (48)$$

Joint Probability Density Functions

By differentiating the joint CDF, we have the joint PDF

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} \tag{49}$$

The order of differentiation is not important. The probability element is

$$f_{XY}(x, y)dx\,dy = \Pr(x < X \leq x + dx, y < Y \leq y + dy) \tag{50}$$

The expected value of any function $g(X, Y)$ can be computed as

$$E\{g(X, Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f_{XY}(x, y)\, dx\, dy \tag{51}$$

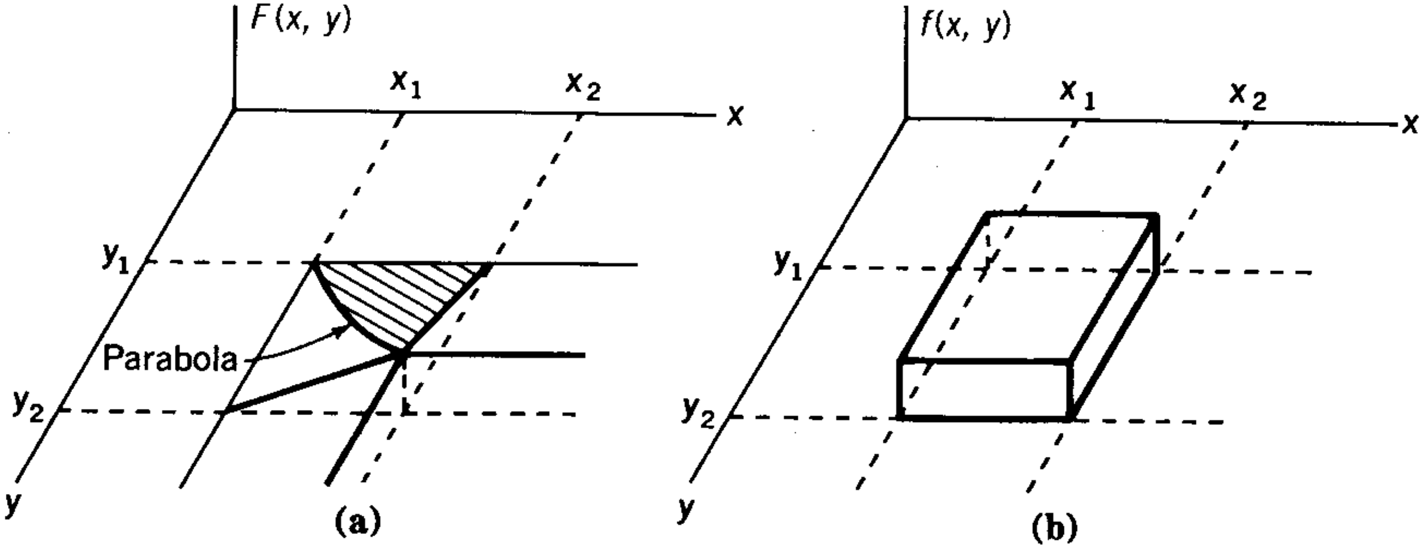
Properties of Joint PDFs

The properties of the joint PDFs are quite similar to those of a single random variable. They are summarized as follows:

1. $f_{XY}(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$
3. $F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u, v) du dv$
4. $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy\,, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$
5. $\Pr(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y) dx dy$

Note that item 2 implies that the volume under the joint PDF is unity.

Example 27



Consider a pair of random variables X and Y having a joint PDF that is constant between (x_1, x_2) *and* (y_1, y_2). *Thus,*

$$f_{XY}(x, y) = \begin{cases} \dfrac{1}{(x_2 - x_1)(y_2 - y_1)}, & x_1 < x \leq x_2 \quad y_1 < y \leq y_2 \\ 0, & \textit{otherwise} \end{cases}$$

a) Determine $E\{XY\}$.

b) Find the marginal PDFs $f_X(x)$ *and* $f_Y(y)$.

Example 27 (cont'd)

<u>Answer</u>

a) The correlation of *X* and *Y* is given by

$$E\{XY\} = \iint_{-\infty}^{\infty} xy f_{XY}(x,y)dx\,dy$$

$$= \frac{(x_2^2 - x_1^2)(y_2^2 - y_1^2)}{4(x_2 - x_1)(y_2 - y_1)}$$

$$= \frac{1}{4}(x_1 + x_2)(y_1 + y_2)$$

b) $f_X(x) = \int_{y_1}^{y_2} f_{XY}(x,y)dy = \frac{y_2 - y_1}{(x_2 - x_1)(y_2 - y_1)} = \frac{1}{x_2 - x_1} \qquad x \in (x_1, x_2)$

$f_Y(y) = \int_{x_1}^{x_2} f_{XY}(x,y)dx = \frac{x_2 - x_1}{(x_2 - x_1)(y_2 - y_1)} = \frac{1}{y_2 - y_1} \qquad y \in (y_1, y_2)$

Exercise

The random variables X and Y have a joint PDF given by

$$f_{XY}(x,y) = A\, exp[-(3x+4y)]\, u(x)u(y)$$

where $u(\cdot)$ is the unit step function.

a) Find the value of A.

b) Determine $Pr(X > 0.5, Y > 0.25)$.

c) Compute the expected value of XY.

Answer [hint: $\frac{\Gamma(\alpha)}{\lambda^{\alpha}} = \int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx$]

a) $A = 12$; b) 0.0821; c) 0.0833

Conditional Probability (1)

It is easy to see that

$$F_X(x|Y \le y) = \Pr(X \le x|Y \le y)$$
$$= \frac{\Pr(X \le x, Y \le y)}{\Pr(Y \le y)}$$
$$= \frac{F_{XY}(x,y)}{F_Y(y)} \tag{52}$$

Similarly,

$$F_X(x|y_1 < Y \le y_2) = \frac{\Pr(X \le x, y_1 < Y \le y_2)}{\Pr(y_1 < Y \le y_2)} = \frac{F_{XY}(x, y_2) - F_{XY}(x, y_1)}{F_Y(y_2) - F_Y(y_1)} \tag{53}$$

Conditional Probability (2)

A more interesting case is

$$F_X(x|Y=y) = \lim_{\Delta y \to 0} \frac{F_{XY}(x,\ \ y+\Delta y) - F_{XY}(x,y)}{F_Y(y+\Delta y) - F_Y(y)}$$

As $\Delta y \to 0$,

$$F_X(x|Y=y) = \frac{\partial F_{XY}(x,y)/\partial y}{\partial F_Y(y)/\partial y} = \frac{\int_{-\infty}^{x} f_{XY}(u,y)du}{f_Y(y)} \tag{54}$$

The corresponding conditional PDF is given by

$$f_X(x|Y=y) = \frac{\partial F_X(x|Y=y)}{\partial x} = \frac{f_{XY}(x,y)}{f_Y(y)} = {\color{blue} f_X(x|y)} \tag{56}$$

By interchanging *X* and *Y*, we have

$$f_Y(y|X=x) = \frac{f_{XY}(x,y)}{f_X(x)} = {\color{blue} f_Y(y|x)} \tag{57}$$

Bayes' Theorem (continuous version)

From (56) and (57), one can obtain the continuous version of Bayes' theorem

$$f_Y(y|x) = \frac{f_X(x|y)f_Y(y)}{f_X(x)} \tag{58}$$

We can also obtain the total probability from (56) and (57) as follows:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-\infty}^{\infty} f_X(x|y)f_Y(y)dy \tag{59}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \int_{-\infty}^{\infty} f_Y(y|x)f_X(x)dx \tag{60}$$

Example 28

Suppose we are given a joint PDF

$$f_{XY}(x,y) = \begin{cases} \frac{6}{5}(1 - x^2 y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & otherwise \end{cases}$$

Find $f_X(x)$, $f_Y(y)$, $f_X(x|y)$ and $f_Y(y|x)$.

<u>Answer</u>

Integrating the joint PDF with respect to *y* yields

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy = \int_0^1 \frac{6}{5}(1 - x^2 y) dy = \frac{6}{5}\left(1 - \frac{x^2}{2}\right)$$

for $0 \le x \le 1$ and 0 otherwise.

Example 28 (cont'd)

Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_0^1 \frac{6}{5}(1 - x^2 y)dx = \frac{6}{5}\left(1 - \frac{y}{3}\right)$$

for $0 \le y \le 1$ and 0 otherwise.

The two conditional PDFs are

$$f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1-x^2 y}{1-y/3} \qquad 0 \le x \le 1, 0 \le y \le 1$$

and

$$f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1-x^2 y}{1-x^2/2} \qquad 0 \le x \le 1, 0 \le y \le 1$$

Both conditional PDFs are 0 otherwise.

Communication Systems (1)

In a digital communication system, the observed quantity at some time instant is the sum of two parts:

$$Y = X + N \tag{61}$$

where *X* is the signal component and *N* is the noise component. Note that *Y* is the only quantity that can be observed.

Given the observed value of *Y*, it is desired to find the conditional PDF of X, $f_X(x|y)$, which is a reasonable estimate of *X* when *X* can only be observed in the presence of noise.

From Bayes' theorem, we have

$$f_X(x|y) = \frac{f_Y(y|x)\, f_X(x)}{f_Y(y)}$$

Communication Systems (2)

Let us consider $f_Y(y|x)$ in detail. If *X* is given, then the only randomness about *Y* is the noise *N*, and it is assume that its PDF $f_N(n)$ is known. Hence,

$$f_Y(y|x) = f_N(y - x) \tag{62}$$

- The desired conditional PDF can now be written as

$$f_X(x|y) = \frac{f_N(y-x)\ f_X(x)}{f_Y(y)} = \frac{f_N(y-x)\ f_X(x)}{\int_{-\infty}^{\infty} f_N(y-x)\ f_X(x)dx} \tag{63}$$

- The denominator in (63) is obtained from (60) as

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\ dx = \int_{-\infty}^{\infty} f_Y(y|x) f_X(x) dx \\ &= \int_{-\infty}^{\infty} f_N(y-x)\ f_X(x)\ dx \end{aligned}$$

Communication Systems (3)

If the prior PDF $f_X(x)$ and the noise PDF $f_N(n)$ are given, it is possible to determine the conditional PDF $f_X(x|y)$ from (63). When a particular value of *Y* is observed, say $Y = y_1$, then the value of x for which $f_X(x|y_1)$ is maximum will be a good estimate for the true value of *X*.

Example 29

Suppose that the prior PDF of the signal random variable *X* is

$$f_X(x) = b\exp(-bx)\,u(x)$$

The noise is assumed to be additive and Gaussian with mean zero and variance σ_N^2 as

$$f_N(n) = \frac{1}{\sigma_N\sqrt{2\pi}}\,e^{-n^2/2\sigma_N^2}$$

Note that

$$f_Y(y) = \int_{-\infty}^{\infty} f_N(y-x)\, f_X(x)\, dx$$

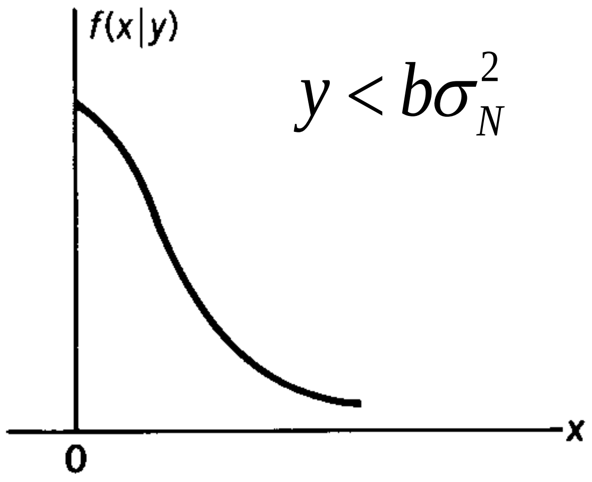
does not contain any useful information for locating the maximum of $f_X(x|y)$. That is, it is not a function of x.

The desired conditional PDF can now be written as

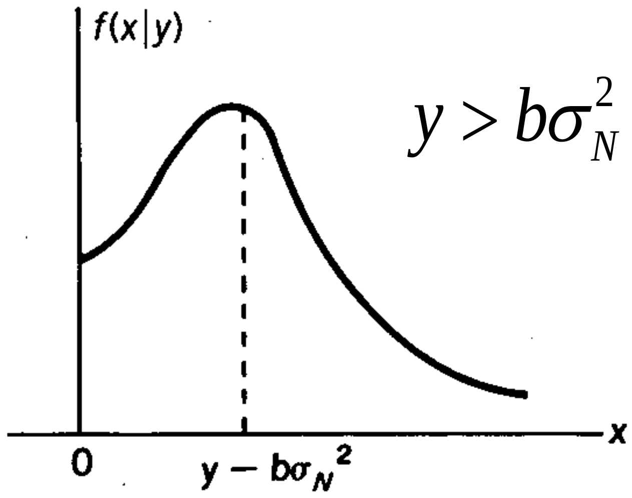
$$f_X(x|y) = \frac{f_N(y-x)\, f_X(x)}{f_Y(y)} = \frac{be^{-bx}}{\sigma_N\sqrt{2\pi}\, f_Y(y)} \exp\left[-\frac{(y-x)^2}{2\sigma_N^2}\right] u(x)$$

This may also be written as

$$f_X(x|y) = \frac{b}{\sigma_N\sqrt{2\pi}\, f_Y(y)} \exp\left[-\frac{x^2 - 2(y - b\sigma_N^2)x + y^2}{2\sigma_N^2}\right] u(x)$$



(a)



(b)

It was noted when a particular value of $Y = y$ is observed, a reasonable estimate for the true value of X is that value of x which maximizes $f_X(x|y)$. Since the conditional PDF is maximum when the exponent part $x^2 - 2(y - b\sigma_N^2)x + y^2$ is a minimum, it follows that this value of x can be determined by taking the derivative and set it equal to zero. Thus

$$2x - 2(y - b\sigma_N^2) = 0$$

or

$$x = y - b\sigma_N^2 \tag{64}$$

If $y < b\sigma_N^2$, as seen from Fig. (a) of the previous page, there is no point of zero slope on $f_X(x|y)$ and the largest value occurs at the boundary $x = 0$, yielding $\hat{x} = 0$. Otherwise, when $y > b\ \sigma_N^2$, the appropriate estimate for *X* is $\hat{x} = y - b\sigma_N^2$, as seen in Fig. (b).

Exercise

Suppose X and Y are two random variables with joint PDF

$$f_{XY}(x,y) = D(x+y)$$

over the rectangular region $0 \leq x \leq 1$ *and* $0 \leq y \leq 1$*, and 0 elsewhere.*

a) *Find the value of D such that* $f_{XY}(x,y)$ *is a valid PDF.*

b) *Determine* $Pr(X > Y|Y = 0.5)$.

c) *Compute the probability that Y is less than or equal to 0.5 given that X is 0.5.*

<u>Answer</u>

a) *D* = 1; b) 0.625; c) 0.375

Exercise

A random signal X is uniformly distributed between 6 and 10 V. It is received in the presence of Gaussian noise N with mean 0 and variance 4.

a) If the observed value $Y = X + N$ is 4, find the best estimate of the signal amplitude.

b) Repeat (a) if the received value Y is 8.

c) Repeat (a) if the received value Y is 12.

Answer

a) 6; b) 8; c) 10

Statistical Independence

The joint PDF for ***statistically independent*** random variables can always be factored into the two marginal PDFs. That is,

$$f_{XY}(x,y) = f_X(x) f_Y(y) \tag{65}$$

Accordingly, the correlation is

$$E\{XY\} = E\{X\}E\{Y\} = \mu_X \mu_Y \tag{66}$$

If $\mu_X = 0$ or $\mu_Y = 0$, then the result is zero.

Statistical independence implies that

$$f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x) \tag{67}$$

and

$$f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y) \tag{68}$$

Exercise

X and Y are two independent random variables. X is Gaussian distributed with mean 1 and unity variance, while Y is also Gaussian with mean 2 and variance 4. Find the probability that $XY > 0$.

<u>Answer</u>

0.7078 + 0.0252 = 0.7330

Correlation and Covariance

We have previously defined the ***correlation*** as

$$E\{XY\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)\,dx\,dy \tag{69}$$

If the mean values are subtracted from individual random variables, it is known as the ***covariance***

$$E\{(X-\mu_X)(Y-\mu_Y)\} = E\{XY\} - \mu_X\mu_Y \tag{70}$$

If it is desired to express the degree to which two random variables are correlated without regard to the magnitude of either one, we have the ***correlation coefficient***

$$\rho = E\left\{\left[\frac{X-\mu_X}{\sigma_X}\right]\left[\frac{Y-\mu_Y}{\sigma_Y}\right]\right\} \tag{71}$$

Sum of Two Independent Random Variables

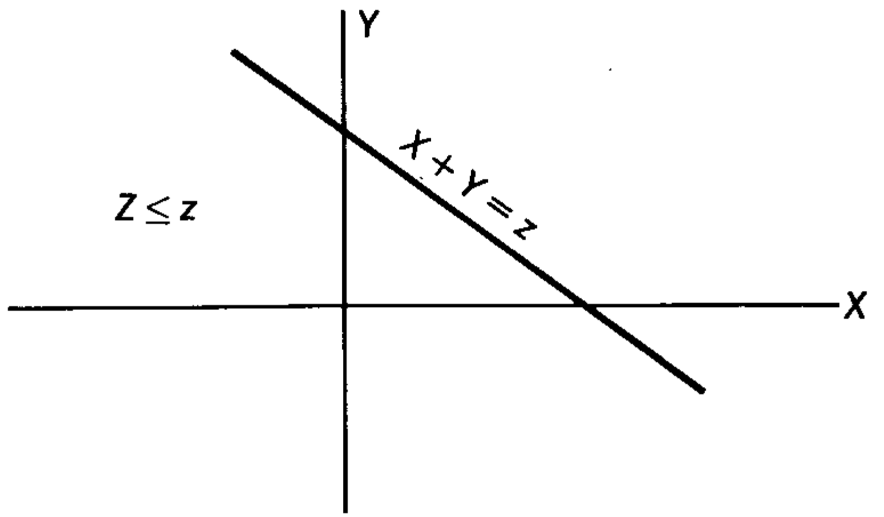
Let X and Y be two statistically independent random variables with PDFs $f_X(x)$ and $f_Y(y)$. Let the sum be

$$Z = X + Y \tag{72}$$

It is desired to find the PDF of Z, $f_Z(z)$. Let us consider the CDF of Z,

$$F_Z(z) = \Pr(Z \leq z) = \Pr(X + Y \leq z)$$

The CDF can be obtained by integrating the joint PDF $f_{XY}(x, y)$ over the region below the line $x + y = z$.



Thus,

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(x, y)\, dx\, dy \tag{73}$$

For the special case that X and Y are statistically independent, the joint PDF is factorable as

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} f_Y(y) \left[\int_{-\infty}^{z-y} f_X(x) dx \right] dy \end{aligned}$$

The PDF of Z can be obtained by differentiating $F_Z(z)$ with respect z. Hence,

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y) dy \tag{74}$$

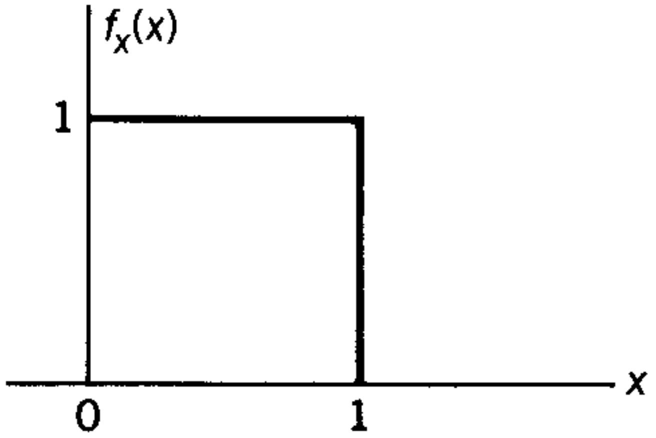
This is observed that the PDF of Z is simply the ***convolution*** of X and Y.

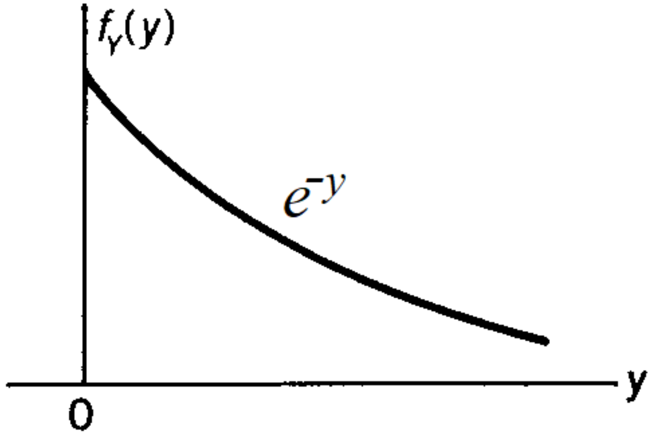
Exercise

Show that the previous convolution integral in (74) can be written as

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \qquad (75)$$

Example 30 -- Convolution





Consider two independent random variables *X* and *Y* with PDFs

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

and

$$f_Y(y) = e^{-y} u(y)$$

The convolution can be carried out in three steps.

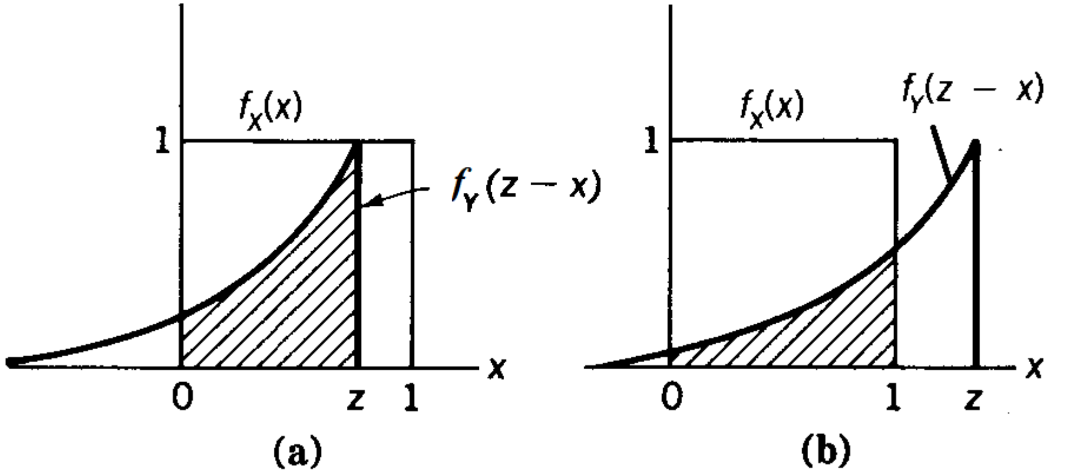
- When $z \leq 0, f_Z(z) = 0$ since $f_X(x) = 0$ in this region.

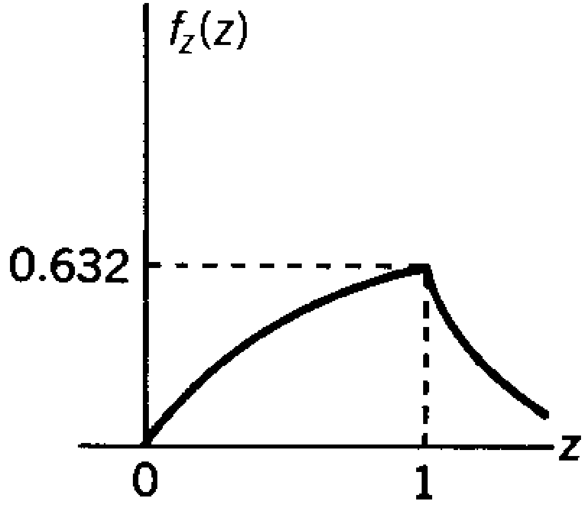
- When $0 < z \leq 1$, as shown in Fig. (a),

$$f_Z(z) = \int_0^z (1)e^{-(z-x)}\,dx = 1 - e^{-z}$$

- When $z > 1$, as shown in Fig. (b),

$$f_Z(z) = \int_0^1 (1)e^{-(z-x)}\,dx = (e-1)e^{-z}$$





Exercise

If X and Y are independent random variables, show that the PDF of

$$Z = X - Y$$

is

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z+y)\,dy \tag{76}$$

or

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x-z)\,dx \tag{77}$$

Exercise

Let X and Y be two statistically independent random variables with PDFs

$$f_X(x) = 5e^{-5x}u(x)$$

and

$$f_Y(y) = 2e^{-2y}u(y)$$

For the random variable $Z = X + Y$, find

a) the value of $f_Z(0)$

b) the value of Z for which $f_Z(z)$ is a maximum

c) the probability that Z is greater than 1.0.

<u>Answer</u>

a) 0; b) 0.305; (c) 0.221

Characteristic Functions

The ***characteristic function*** of a random variable *X* is defined to be

$$\varphi_X(\omega) = E\{e^{j\omega X}\} = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x}\, dx \tag{78}$$

It is observed that the RHS of (78), except for a plus sign in the exponent, is the Fourier transform of the PDF f_X(x). The difference in sign for the characteristic function is traditional rather than fundamental. It makes no essential difference in the application or properties of the transform.

When the convolution arises, we can apply what we learnt in *Signals and Systems* and use characteristic functions to transform PDFs into the frequency domain. After the frequency operations, the desired PDF can be obtained by using the inverse Fourier transform.

Example 31

Let X and Y be two statistically independent random variables with PDFs and characteristic functions

$$f_X(x) = 5e^{-5x}u(x) \leftrightarrow \frac{5}{5 - j\omega}$$

and

$$f_Y(y) = 2e^{-2y}u(y) \leftrightarrow \frac{2}{2 - j\omega}$$

Hence, the characteristic function of $Z = X + Y$ is given by

$$\varphi_Z(\omega) = \varphi_X(\omega)\varphi_Y(\omega) = \frac{5}{5-j\omega} \times \frac{2}{2-j\omega}$$

- Using partial fractions, we have

$$\varphi_Z(\omega) = \frac{10/3}{2 - j\omega} - \frac{10/3}{5 - j\omega}$$

- Taking the inverse Fourier transform, we obtain the PDF of Z as

$$f_Z(z) = \frac{10}{3}[e^{-2z} - e^{-5z}]u(z)$$

Generating Moments

Another application of the characteristic function is to generate the moments of a random variable. Note that if $\varphi_X(\omega)$ is differentiated with respect to ω, the result is

$$\frac{d\varphi_X(\omega)}{d\omega} = \frac{d}{d\omega}\int_{-\infty}^{\infty} f_X(x)e^{j\omega x}\,dx = \int_{-\infty}^{\infty}(jx)f_X(x)e^{j\omega x}\,dx$$

Hence, by setting $\omega = 0$ on both sides, we can see

$$\left[\frac{d\varphi_X(\omega)}{d\omega}\right]_{\omega=0} = \int_{-\infty}^{\infty}(jx)f_X(x)\,dx = j\mu_X \tag{79}$$

Higher-order moments can also be generated by repeating the differentiations as

$$E\{X^k\} = \frac{1}{j^k}\left[\frac{d^k\varphi_X(\omega)}{d\omega^k}\right]_{\omega=0} \tag{80}$$

Example 32

Again using the previous example,

$$f_X(x) = 5e^{-5x}u(x) \leftrightarrow \varphi_X(\omega) = \frac{5}{5 - j\omega}$$

It is easy to compute the first two moments,

$$E\{X\} = \frac{1}{j}\left[\frac{d\varphi_X(\omega)}{d\omega}\right]_{\omega=0} = \frac{1}{j}\left[\frac{5j}{(5-j\omega)^2}\right]_{\omega=0} = \frac{1}{5}$$

and

$$E\{X^2\} = \frac{1}{j^2}\left[\frac{d^2\varphi_X(\omega)}{d\omega^2}\right]_{\omega=0} = \frac{1}{j^2}\left[\frac{(5j)(-2)(-j)}{(5 - j\omega)^3}\right]_{\omega=0} = \frac{2}{25}$$

The variance is therefore $\sigma^2 = E\{X^2\} - E^2\{X\} = \frac{2}{25} - \left(\frac{1}{5}\right)^2 = \frac{1}{25}$.

Probability and Statistics

- Probability and statistics are often considered to be one and the same subject. They are linked in courses and textbooks.
- However, they are really two different areas of study even through statistics relies heavily on probability concepts.
- ***Statistics*** is defined as the science of assembling, classifying, tabulating, and analysing data or facts.

Two Branches of Statistics

- ***Descriptive statistics***: involves collecting , grouping, and presenting data in a way that can be easily understood or assimilated.
- ***Statistical inference***: uses the data to draw conclusions about, or estimate parameters of, the environment from which the data came.

Area Classifications in Statistics

a) ***Sampling theory*** – deals with problems associated with selecting samples from some collection of data that is too large to be examined completely.
b) ***Estimation theory*** – is concerning with making some estimate or prediction based on the data that is available.
c) ***Hypothesis test*** – attempts to decide which of two or more hypotheses about the data is true.
d) ***Curve fitting (regression)*** – attempts to find mathematical expressions that best represent the data.
e) ***Analysis of variance*** – attempts to assess the significance of variation in the data and the relation of these variation to the physical situations from which the data arose.

In this course, we will limit our attention to some simple concepts associated with sampling theory, analysis of variance and linear regression.

Number of Samples

- In many cases, the number of potential items is so large that it would be impractical to test every one of them. Problems of this type can be solved by sampling a subset of all possible items.
- A sufficient number of samples must be taken in order to obtain an answer in which one has reasonable confidence. On the other hand, it may be very expensive and time consuming to take excessive samples. Hence, one of the purposes is to determine how many samples are required for a given degree of confidence in the result.

Terminology

- The collection of data that is being studied is known as the ***population***.
- The number of items that make up the population is designated as N, which is also called the ***size of the population***. If N is not very large, then its exact value may be significant in our calculation. On the other hand, if N is very large, it is often convenient to assume that it is infinite. It is generally believed that the calculations for infinite populations are somewhat easier to carry out.
- A ***sample*** is simply part of the population that has been selected at random. That is, all members of the population are equally likely to be chosen. The number of items in the sample is denoted as n and is called the ***size of the sample***.

Sample Mean

- For our purposes of description, let us assume that we have a random sample of size *n* drawn from a population of size *N*. Each member of the sample has a numerical value that is designated by $x_1, x_2, \cdots, x_n$. The sample mean is simply the average of the sample values as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{81}$$

where the x_i's are the particular values in the sample.

- In general, we are interested in describing the statistical properties of arbitrary random samples rather than those of any particular one. In this case, the sample mean becomes a random variable as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{82}$$

where X_i is the *i*-th random variable from the population with PDF $f_X(x)$.

Unbiased Estimator

- Suppose the true mean of the population is denoted by m_X. Hopefully, the sample mean in (82) will be close to m_X.
- Since the sample mean is a random variable, its expectation is

$$E\{\bar{X}\} = \frac{1}{n}\sum_{i=1}^{n} E\{X_i\} = m_X \tag{83}$$

- The sample mean has its expected value same as the population mean. Therefore, it is an ***unbiased estimator***, implying that the mean of the estimate of the parameter is the same as the true value of the parameter.

Unbiased Estimator = Good Estimator (?)

- It is desirable for the sample mean to be an unbiased estimate of the true population mean. However, the “stability” of this estimator is equally important. Since the sample mean is a random variable, it will have a random value fluctuating around the true value as different samples are drawn. It is important to know the magnitude of this fluctuation.
- Hence, we need to determine the variance of the sample mean. This is initially done for the case in which the population size is very much greater than the sample size.

Variance of Sample Mean

- Assuming that $N \gg n$ so that the characteristics of the population do not change as the sample is drawn. It is equivalent to saying that $N = \infty$.
- Thus

$$
\begin{aligned}
\operatorname{var}[\bar{X}] &= E\{\bar{X}^2\} - E^2\{\bar{X}\} \\
&= E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] - m_X^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\{X_i X_j\} - m_X^2
\end{aligned}
\tag{84}
$$

- Since X_i and X_j, $i \neq j$, are values of different items in the population, it is reasonable to assume that they are statistically independent. Hence,

$$E\{X_i X_j\} = \begin{cases} m_X^2, & i \neq j \\ E\{X^2\}, & i = j \end{cases} \tag{85}$$

- Using (85) in (84) leads to

$$\begin{aligned} \mathrm{var}(\bar{X}) &= \frac{1}{n^2}[nE\{X^2\} + (n^2 - n)m_X^2] - m_X^2 \\ &= \frac{1}{n}[E\{X^2\} - m_X^2] = \frac{\sigma_X^2}{n} \end{aligned} \tag{86}$$

- Note that σ_X^2 is the true variance of the population. The variance of the sample mean can be made small by increasing the sample size n.

Remarks

- Recall that the basic reason for assuming that the population size is very large is to ensure that the statistical characteristics of the population do not change as we withdraw the members of the sample. If the requirement of a large population size is not satisfied, we can maintain the population characteristics by replacing an item that is withdrawn after it has been selected. Sampling done in this way is said to be ***sampling with replacement***.
- If the population size is not large and we do not replace the selected items, the variance for the sample mean, which is simply quoted here without proof, is

$$\mathrm{var}(\bar{X}) = \frac{\sigma_X^2}{n}\left(\frac{N-n}{N-1}\right) \qquad (87)$$

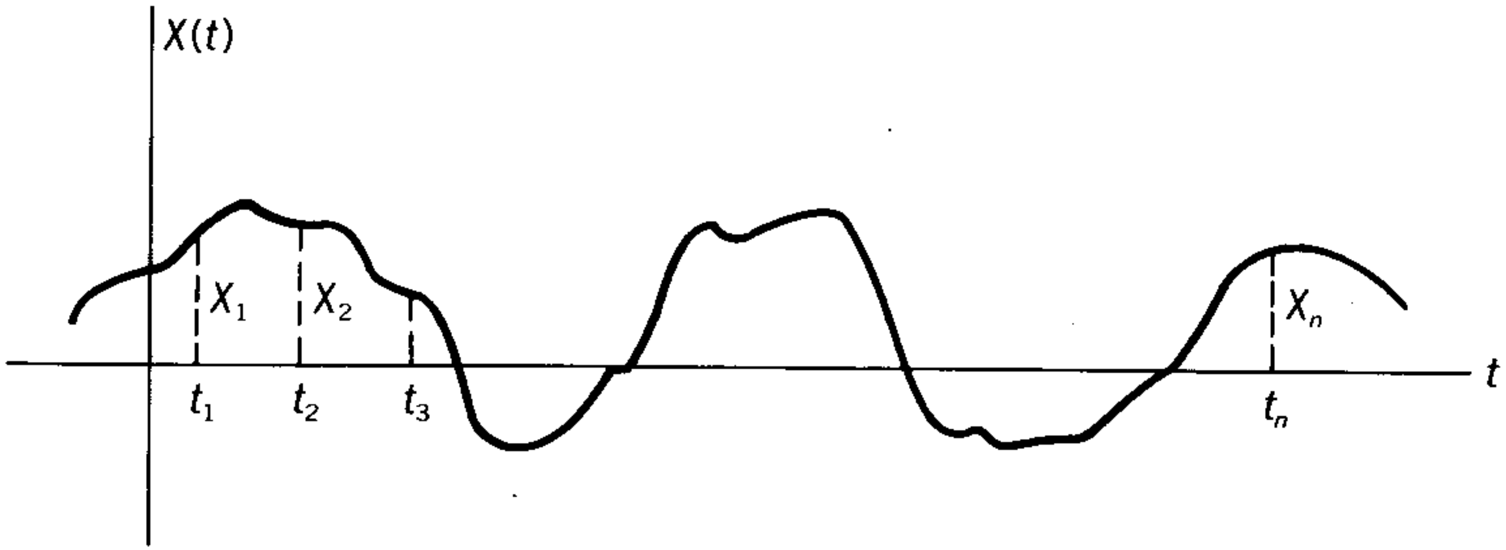
Exercise

In (87), the variance of the sample mean for finite population size without replacement is given.

a) *Discuss the resulting* $var(\bar{X})$ *when* $N \to \infty$. *It should be consistent with the expression in (86),*

b) *When* $N = n$, $var(\bar{X}) = 0$. *Discuss the implication of its meaning.*

Example 33



- *Suppose we have a random waveform with true mean of 10 and true variance of 9 shown above. The value of this waveform* $X(t)$ *is being sampled at equally spaced time instants* $t_1, t_2, \cdots, t_n$. *In the general situation, these sample values are random variables and are denoted by* $X_i = X(t_i)$ *for* $i = 1, 2, \cdots$. *We estimate the mean value of this waveform using the sample mean in (82).*

- *We assume that the waveform lasts forever, so that the population size is infinite.*
- *In addition, we would like to find a suitable sample size to estimate the mean value of this waveform with a standard derivation that is only one percent of the true mean value.*

<u>Answer</u>

According to the description, we have

$$\operatorname{var}(\bar{X}) = \frac{\sigma_X^2}{n} = \frac{9}{n} \tag{88}$$

and

$$\sigma_{\bar{X}} = 0.01 \times 10 = 0.1 \tag{89}$$

Thus, using (88) and (89), we find

$$n = \frac{9}{\sigma_{\bar{X}}^2} = \frac{9}{0.01} = 900$$

Observations

- The result in Example 33 indicates that the sample size must be quite large in most case of sampling an infinite population, or in sampling with replacement, if it is desired to obtain a sample mean with a small variance.
- Estimating the mean value of the random waveform with the specified variance does not necessarily imply that the estimate is really within one percent of the true mean.
- As the sample size is large ($n = 900$) in this case, it is reasonable to use the central limit theorem to check how good the estimator is.

Is It a Good Estimator?

- Since the estimated mean is related to the sum of a large number of independent random variables, the sum is nearly Gaussian regardless of the individual PDFs of X_i's. Hence,

$$\begin{aligned}
\Pr(9.9 < \bar{X} \leq 10.1) &= Q\left(\frac{9.9 - 10}{0.1}\right) - Q\left(\frac{10.1 - 10}{0.1}\right) \\
&= Q(-1) - Q(1) \\
&= 1 - 2Q(1) \\
&= 1 - 2(0.1587) = 0.6826
\end{aligned}$$

- It is observed that there is a significant probability (= 0.3174) that the estimate of the population mean is actually more than one percent away from the true population mean.

Example 34

There is a population of 100 bipolar transistors for which one wishes to estimate the mean value of the current gain, β. If the true population mean m_β is 120 and the true population variance is $\sigma_\beta^2 = 25$. How large a sample size is required to obtain a sample mean that has a standard deviation equal to one percent of the true mean?

<u>Answer</u>

The desired variance is given by

$$\mathrm{var}\left(\bar{\beta}\right) = (0.01 \times 120)^2 = 1.44$$

From (87), we also have

$$\mathrm{var}\left(\bar{\beta}\right) = \frac{25}{n}\left(\frac{100 - n}{100 - 1}\right) = 1.44$$

- The equation may be solved to yield $n = 14.92$, which implies a sample size of 15 as *n* is an integer.
- In this case, we cannot find the probability that the sample mean is within one percent of the true population mean because $n = 15$ is not large enough to justify the central limit theorem.
- As a rule of thumb, it is often assumed that a sample size of at least 30 is required to make the Gaussian assumption.

Exercise

An endless production line is producing solid-state diodes and every 100th diode is tested for reverse current I_{-1} and forward current I_1 at diode voltages of -1 and $+1$, respectively.

a) *If the random variable I_{-1} has a true value of 10^{-6} and a variance of 10^{-11}, how many diodes must be tested to obtain a sample mean whose standard deviation is 5% of the true mean?*

b) *If the random variable I_1 has a true mean value of 0.1 and a variance of 0.0025, how many diodes must be tested to obtain a sample mean whose standard deviation is 1% of the true mean?*

c) *If the larger of the two numbers found in (a) and (b) is used for both tests, what will be their standard deviations.*

Answers: (a) 4000; (b) 2500; (c) 5×10^{-6}, 7.905×10^{-4}

Sample Variance

- In addition to the sample mean, we are also interested in estimating its variance. A knowledge of the variance is important because it indicates something about the spread of values around the mean.
- For example, it is not sufficient to test resistors and find that the sample mean is very close to the desired resistance value. If the standard deviation of the resistance value is vary large, then regardless of how close the sample is, many of the resistors can be quite far from the desired value.
- The sample variance is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad (90)$$

where $\bar{X}$ is the sample mean.

- The expected value of S^2 can be obtained by expanding the squared term in (90) and taking the expected value of each term in the expansion. The result is

$$E\{S^2\} = \sigma_X^2 \tag{91}$$

where σ_X^2 is the true variance of the population.

- Note that the reason for using $1/(n-1)$ rather than $1/n$ in (90) is to make the estimator unbiased.
- The variance of S^2 can be found from

$$\mathrm{var}(S^2) = E\{(S^2 - \sigma_X^2)^2] \tag{92}$$

- Upon expanding the RHS and carrying out expectation term by term, it can be found to be

$$\mathrm{var}(S^2) = \frac{1}{n}\left[\mu_4 - \frac{n-3}{n-1}\sigma_X^4\right] \tag{93}$$

where μ_4 is the fourth central moment of X, given by

$$\mu_4 = E\{(X - m_X)^4\} \tag{94}$$

- Equation (93) shows again that the variance of S^2 is an inverse function of n.
- In principle, the distribution of S^2 can be derived with use of techniques discussed previous. It is, however, a tedious process because of the complex nature of the expression for S^2 as defined by (90).
- For the case in which population X is distributed according to $N(m_X, \sigma_X^2)$, i.e., Gaussian distributed with mean m_X and variance σ_X^2, it can be shown that $(n-1)S^2/\sigma_X^2$ has a chi-squared (χ^2) distribution with $(n-1)$ degrees of freedom.

χ^2 Distribution

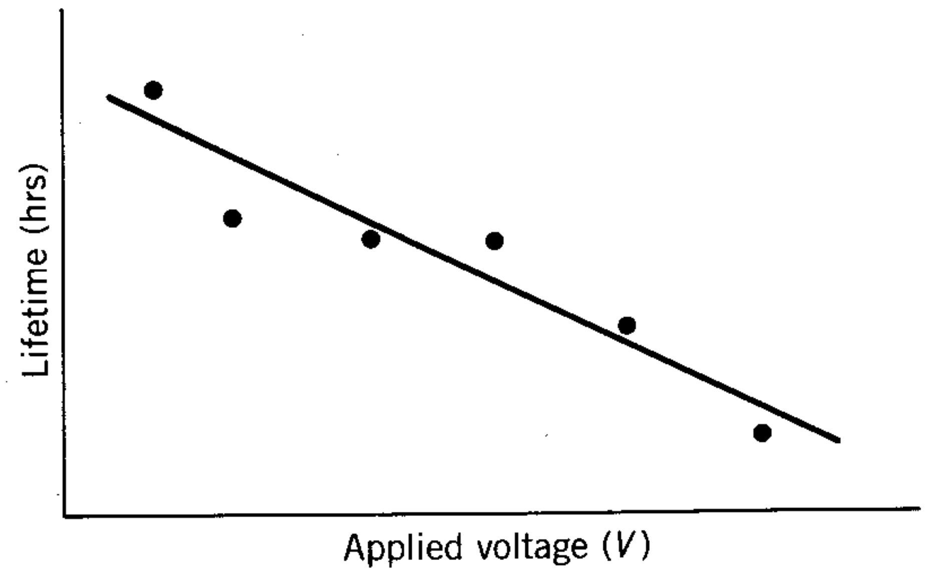
- The χ^2 distribution contains one parameter *n* with the PDF of the form

$$f_X(x) = \frac{1}{2^{\left(\frac{n}{2}\right)}\Gamma(n/2)} x^{(n/2)-1}\, e^{-\frac{x}{2}} u(x) \tag{95}$$

- The parameter *n* is called the ***degrees of freedom***. The utility of this distribution arises from the fact that a sum of the squares of *n* independent standardized normal random variables has a χ^2 distribution with *n* degrees of freedom.

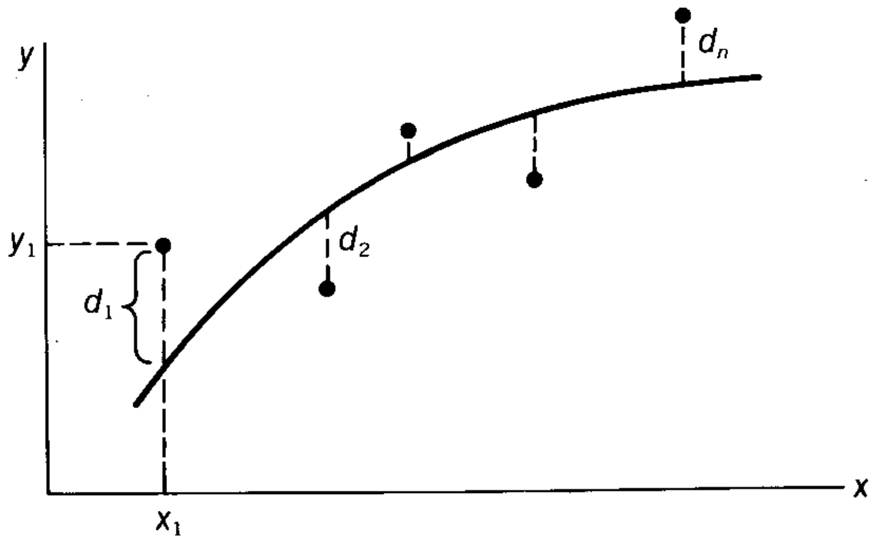
Scatter Diagram

Consider two variables *x* and *y*. Data are collected related to these two variables. For a sample size of n, we would have values of one variable denoted as $x_1, x_2, \dots, x_n$ and corresponding values of the other variable as $y_1, y_2, \dots, y_n$. For the scatter diagram displayed, each *x*-value might be an applied voltage and each *y*-value the corresponding lifetime.



Curve Fitting

- The general problem of find a mathematical relationship to represent the data is called ***curve fitting***. The resulting curve is called a ***regression curve*** and the mathematical equation is the ***regression equation***.
- In order to find the "best" regression equation, it is necessary to establish a criterion that will be used to justify it.



Least-Squares Regression Curve

- The difference between the regression curve and the corresponding value of *y* at any *x* is designated as $d_i,\ i = 1, 2, \dots, n$. The criterion of goodness of fit that is employed here is

$$d_1^2 + d_2^2 + \cdots + d_n^2 = \text{a minimum} \tag{96}$$

- Such a criterion leads to a ***least-squares regression curve*** and is the criterion that is most often employed.
- Note that the least-squares criterion weights errors on either side of the regression curve equally and also weights large errors more than small errors.

Linear Regression

- Once the criterion is decided, the next step is to select the type of the regression equation that is to be fitted to the data. The general choice is a polynomial of the form

$$y = a + bx + cx^2 + \cdots + kx^j \tag{97}$$

- Our subsequent discussion here is limited to using a first degree of polynomial in order to preserve simplicity while conveying the essential aspects of the method. This technique is referred to as ***linear regression***.
- The linear regression equation becomes

$$y = a + bx \tag{98}$$

where a and b are the constants to be determined.

- In order to minimize the expression

$$\sum_{i=1}^{n}[y_i - (a + bx_i)]^2 = \text{a minimum} \tag{99}$$

we would differentiate partially with respect to a and b and set the derivatives equal to zero. This leads to two equations that may be solved simultaneously for the values of a and b. The equations are

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i \tag{100}$$

and

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \tag{101}$$

- The resulting values of a and b are

$$b = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{102}$$

and

$$a = \frac{\sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i}{n} \tag{103}$$

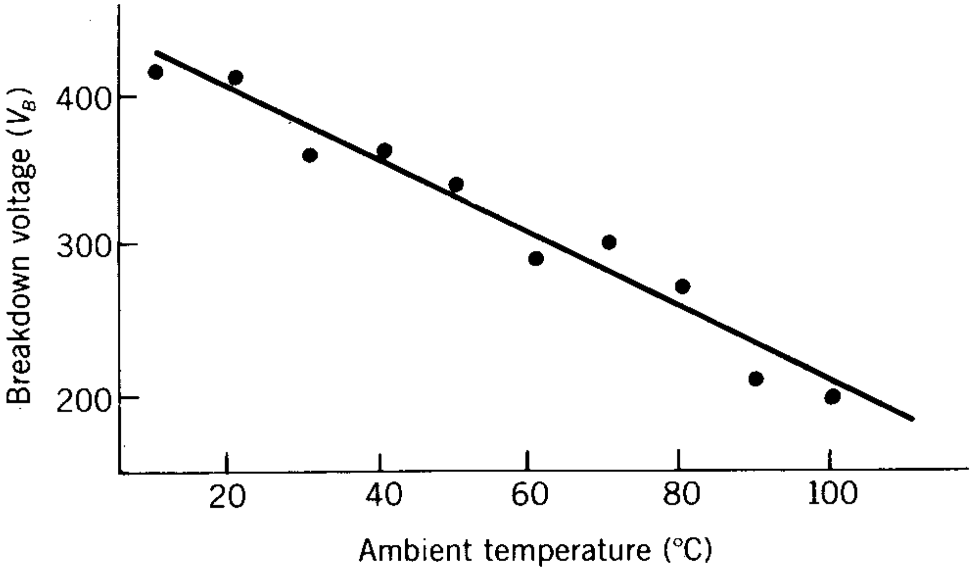
- Although these equations are fairly complicated expressions, they can be evaluated readily by computers or programmable calculators.

Example 35

i	1	2	3	4	5	6	7	8	9	10
T (deg C)	10	20	30	40	50	60	70	80	90	100
Vb (V)	420	410	360	360	340	290	300	270	210	200

- A manufacturer of capacitors wishes to determine the relationship between the breakdown voltage and the ambient temperature in which the capacitor operates. He tests 10 capacitors at different temperature and obtains the data above.
- The calculated values of $a = 451.33$ and $b = -2.406$. Hence, the mathematical relationship is

$$V_b = 451.33 - 2.406T$$



Exercise

Four light bulbs are tested to establish a relationship between lifetime and operating voltage. The resulting data shows in the following table.

i	1	2	3	4
Voltage (V)	105	110	115	120
LifeTime (Hr)	1200	1000	920	750

Find the coefficients of the linear regression curve and plot it with the scatter diagram. Determine the expected lifetime of a light bulb operating at a voltage of 100 volts.

Answer

$a = 4185$; $b = -28.6$; 1325 hrs