TABLE OF CONTENTS

Syllabus

Course Catalog

3 Credit hours (3 h lectures). Probability principles and set theory. Random variables. Operations on random variables. Various distribution functions. Random processes: temporal and spectral characterization. Response of linear time-invariant systems to random inputs.

Textbook

Peyton Z. Peebles (2001). *Probability, Random Variables and Random Signal Principles,* 4th ed. McGraw Hill.

References

1. Roy D. Yates and David J. Goodman (2004). *Probability and stochastic processes*. 2nd ed. Wiley.
2. Leon-Garcia (2008). *Probability and Random Processes for Electrical Engineering*. 3nd ed. Prentice Hall.
3. Geoffrey Grimmett and David Stirzaker (2001). *Probability and Random Processes*. 3nd ed. Oxford University Press.

Instructor

Name: Dr. Mohammad M. Banat

Email Address: banat@just.edu.jo

Prerequisites

Prerequisites by topic Calculus, Signal Analysis

Prerequisites by course Math 102, EE 260

Prerequisite for EE 450

Topics Covered

Week	Topics	Chepters in Text
1-2	Introduction to Probability Theory	2
3-6	Random Variables and Distribution and Density Functions	3
7	Operations on a Single Random Variables	4
8-9	Multiple Random Variables	5
10-12	Operations on Multiple Random Variables	6
12-14	Random Processes	7
15-16	Spectral Analysis and Filtering of Random Processes	8

Objectives and Outcomes

Objectives	*Outcomes*
1. Know and apply the basic probability principles [a]	1.1. Recognize the role of probability in science and engineering [a] 1.2. Understand basics of set theory [a] 1.3. Understand the axioms of probability [a]
2. Know and apply the basic principles concerning single and multiple Random Variables [a]	2.1. Understand the concepts of discrete and continuous single and multiple random variables [a] 2.2. Understand the concepts of distribution and density functions [a] 2.3. Understand and apply the concepts of moments and moment generating functions [a] 2.4. Be able to determine probabilities using distribution and density functions [a] 2.5. Be able to perform random variable transformations [a]
3. Know and apply the basic time/frequency domain principles concerning Random Processes [a]	3.1. Understand the concept of a random process [a] 3.2. Be able to characterize random processes in the time domain [a] 3.3. Be able to characterize random processes in the frequency domain [a]
4. Know and apply the basic time/frequency domain input-output relationships concerning Linear time invariant systems with random inputs [a]	4.1. Be able to use time domain input/output relationships of linear time invariant systems with random inputs [a] 4.2. Be able to use frequency domain input/output relationships of linear time invariant systems with random inputs [a]

Contribution of Course to Meeting the Professional Component

The course contributes to building the fundamental basic concepts and applications of probability and random processes in Electrical Engineering.

RELATIONSHIP TO PROGRAM OUTCOMES (%)

1	2	3	4	5	6	7
100						

Evaluation

Assessment Tool	**Expected Due Date**	**Weight**
Mid-Term Exam	Sat. 8 August 2020	25%
Class Work		25%
Final Exam		50%

I. INTRODUCTION

I.1. Set Definitions

Definition I-1

A set is a collection of objects (defined as elements).

Example I.1

A set can consist of integer numbers from 1 to 10.

A set can consist of small alphabet letters from a to z.

Set Elements

When object a is an element of A, we write $a \in A$.

When object a is not an element of A, we write $a \notin A$.

Tabular Method

$A = \{x, y, z, w, u\}$

Rule Method

$A = \{\text{integers from 1 to 5}\}$

$B = \{x \mid x > 0\}$

This methods is useful when the set size is large.

Countable Sets

A set is said to be countable if all its elements can be put in one-to-one correspondence with the natural numbers, which are the integers 1, 2, 3, etc.

Example I.2

The set $C = \{0, 1/4, 1/2, \cdots\}$ is countable; because we can create a one-to-one correspondence with the natural numbers.

If a set is not countable it is called uncountable.

Example I.3

The set $D = \{x \mid 0 \le x < 10\}$ is uncountable; because we cannot create a one-to-one correspondence with the natural numbers.

Empty Set

A set is said to be empty if it has no elements. The empty set is often called the null set. Empty set is often denoted by the symbol ϕ.

Example I.4

The set $E = \{\text{real integers whose squares are negative}\}$ is empty; because squares of real integers cannot be negative. In this case $E = \phi$.

Finite Sets

A finite set is one that is either empty or has elements that can be counted, with the counting process terminating.

Example I.5

Set C and D in the above examples are infinite.

Set E is empty, and is therefore finite.

The set of numbers on the six faces of a fair die is finite.

The set of student names in this EE 360 class is finite.

Subsets

A set A is said to be a subset of another set, B, if all elements of A are also elements of B, in which case we write $A \subseteq B$. With this definition, it is possible that the two sets are equal (i.e., they have all the same elements), in which case $A \subseteq B$ and at the same time $B \subseteq A$. If on the other hand, A is a subset of B and there are some elements of B that are not in A, then we say that A is a proper subset of B and we write $A \subset B$.

Example I.6

Let $F = \{\text{Numbers on the six faces of a fair die}\}$.

Let $G = \{2,4,6\}$.

Let $H = \{\text{Positive integers that are smaller than } 7\}$.

Then we have:

$G \subset F$. $H \subseteq F$. $F \subseteq H$. $H = F$.

We can also write:

$F \supset G$. $F \supseteq H$. $H \supseteq F$. $F = H$.

Exercise I.1

Provide a set X; such that $X \subset F$ and $X \supset G$.

Universal Set

The universal set (or sample space) S is the set of all objects under consideration in a given problem.

Example I.7

Set F in Example I.6 is the universal set of the experiment of rolling a die.

Complement

The complement of a set A, written $\overline{A}$, is the set of all elements in S that are not in A.

Example I.8

Set G in Example I.6 has the complement $\overline{G} = \{1,3,5\}$ in the experiment of rolling a die.

Difference Set

For two sets A and B that satisfy $A \subset B$, the difference set, written $B - A$, is the set of elements in B that are not in A.

Example I.9

Let A consist of integer numbers from 1 to 10.

Let B consist of integer numbers from 7 to 15.

$A - B = \{1,2,3,4,5,6\}$ and $B - A = \{11,12,13,14,15\}$.

Note that $A - B \neq B - A$.

Union of Sets

For any two sets A and B the union of the two sets, $A \cup B$, is the set of all elements that are contained in either A or B. Union is sometimes expressed as $A + B$.

Example I.10

The union of set $A = \{\text{red,blue,orange}\}$ and set $B = \{\text{white,blue}\}$ is the set $C = A \cup B = \{\text{red,blue,orange,white}\} = \{\text{white,blue,red,orange}\}$.

Intersection of Sets

For any two sets A and B the intersection of the two sets, $A \cap B$, is the set of all elements that are contained in both A and B. Intersection is sometimes expressed as AB.

Example I.11

The intersection of set $A = \{\text{red,blue,orange}\}$ and set $B = \{\text{white,blue}\}$ is the set $D = A \cap B = \{\text{blue}\}$.

Mutually Exclusive Sets

Two sets A and B are said to be mutually exclusive, or disjoint, if and only if they have no common elements, in which case $A \cap B = \phi$.

Example I.12

Sets $A = \{1,2,3\}$ and $B = \{\text{x,y}\}$ are mutually exclusive, and $A \cap B = \phi$.

Sets $A = \{\text{red,blue,orange}\}$ and $B = \{\text{white,blue}\}$ are not mutually exclusive; because $A \cap B = \{\text{blue}\} \neq \phi$.

Exhaustive Sets

A collection of sets $A_1, A_2, \cdots, A_n$ are said to be exhaustive if each element in the universal set is contained in at least one of the sets in the collection. In such a case $A_1 \cup A_2 \cup \cdots \cup A_n = S$.

Example I.13

Let $S = \{x = \text{integer},\ 0 < x < 10\}$, $A_1 = \{1,2,3,6\}$, $A_2 = \{4,5,6,7,8\}$ and $A_3 = \{4,8,9\}$. Sets A_1, A_2 and A_3 are exhaustive, but they not mutually exclusive.

Example I.14

Let $S = \{x = \text{integer},\ 0 < x < 10\}$, $A_1 = \{1,2,3,6\}$, $A_2 = \{4,5,7,8\}$ and $A_3 = \{9\}$. Sets A_1, A_2 and A_3 are exhaustive and mutually exclusive.

Venn Diagrams

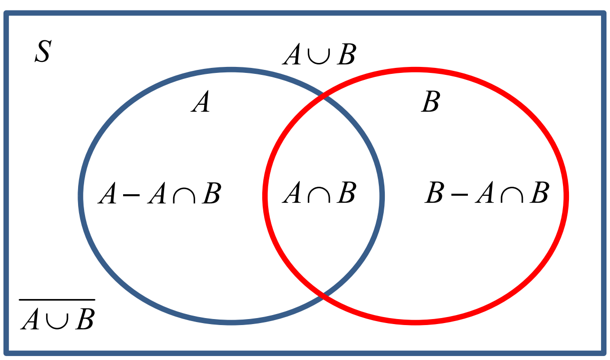


Figure I.1: Venn Diagram

Exercise I.2

Show that:

$\phi \subseteq A \subseteq S$, $A \subseteq A$	If $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$
Generally, $A - B \neq B - A$	If $A - B = B - A$, then $A = B$

$A \cup A = A,\ A \cap A = A$	$A \cup B = B \cup A,\ A \cap B = B \cap A$
$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \subseteq B \Rightarrow \begin{cases} A \cap B = A \\ A \cup B = B \end{cases}$	$\phi \subseteq A \subseteq S$
$\begin{cases} \phi \cap A = \phi \\ S \cap A = A \end{cases}$	$\begin{cases} \phi \cup A = A \\ S \cup A = S \end{cases}$
$\begin{cases} \overline{(\overline{A})} = A \\ \overline{S} = \phi, \overline{\phi} = S \end{cases}$	$\begin{cases} A \cup \overline{A} = S \\ A \cap \overline{A} = \phi \end{cases}$
DeMorgan's first law: $\overline{A \cup B} = \overline{A} \cap \overline{B}$	DeMorgan's second law: $\overline{A \cap B} = \overline{A} \cup \overline{B}$

I.2. Experiments, Sample Spaces, and Events

A deterministic signal is one that may be represented by parameter values, such as a sinusoid, which may be perfectly reconstructed given an amplitude, frequency, and phase. Stochastic (random) signals, such as noise, do not have this property.

The theory of probability provides tools to model and analyze phenomena that occur in many diverse fields, such as communications, signal processing, control, and computers. Perhaps the major reason for studying probability and random processes is to be able to model complex systems and phenomena.

Definition I-2

An experiment is a (quite often hypothetical) procedure we perform that produces some result.

Definition I-3

An outcome is a possible result of an experiment. If a fair coin is tossed five times, an outcome could be $HHTHT$.

Definition I-4

The sample space is the collection or set of all distinct (collectively exhaustive and mutually exclusive) outcomes of an experiment. The letter S is used to designate the sample space, which is the universal set of outcomes of an experiment. A sample space is called discrete if it is a finite or a countably infinite set. It is called continuous or a continuum otherwise.

Example I.15

Consider flipping a fair coin once, where fair means that the coin is not biased in weight to a particular side. There are two possible outcomes, namely, a head or a tail. Thus, the sample space S consists of two outcomes, H to indicate that the outcome of the coin toss was a head and T to indicate that the outcome of the coin toss was a tail. We may write $S=\{H,T\}$.

Example I.16

A cubical die with numbered faces is rolled and the result observed. The sample space consists of six possible outcomes: $1,\ 2,\ \cdots\ ,\ 6$. Note that these numbers represent the six faces of the cubical die. We may write $S=\{1,2,3,4,5,6\}$.

Definition I-5

An event is some set of outcomes of an experiment. For example, the event C in the experiment of tossing a fair coin five time might be $C=\{\text{outcomes with an even number of heads}\}$. All events of an experiment are subsets of the sample space.

Example I.17

Consider the experiment of rolling two dice and observing the results. The sample space consists of the 36 outcomes: $(1,\ 1),\ (1,2),\ \cdots,\ (6,6)$; the first component in the ordered pair indicates the number on the first die, and the second component indicates the number on the second die. Several interesting events can be defined from this experiment, such as

$A=\{\text{sum of the two numbers}=4\}$, $B=\{\text{the two numbers are identical}\}$,

$C=\{\text{the first number is larger than the second}\}$.

Imagine that we conduct two experiments, with each consisting of rolling a single die. The sample spaces (S_1 and S_2) for each of the two experiments are identical, namely, the same as Example I.16. We may now consider the sample space, S, of the original experiment to be the combination of the sample spaces, S_1 and S_2, which consists of all possible combinations of the elements of both S_1 and S_2. This is an example of a **combined sample space**.

Example I.18

Let us flip a coin until a tails occurs. The experiment is then terminated. The sample space consists of a collection of sequences of coin tosses. Label these outcomes as $\xi_n,\ n=1,2,3,\cdots$. The final toss in any particular sequence is a tail and terminates the sequence. All the preceding tosses prior to the occurrence of the tail must be heads.

The possible outcomes that may occur are

$\xi_1=(T),\ \xi_2=(H,\ T),\ \xi_3=(H,H,\ T),\ \cdots$

Note that in this case, n can extend to infinity. This is another example of a combined sample space resulting from conducting independent but identical experiments. In this example, the sample space is **countably infinite**, while in the previous examples sample spaces were finite.

Example I.19

Consider a random number generator that selects a number in an arbitrary manner from the semi-closed interval [0, 1). The sample space consists of all real numbers x for which $0 \le x < 1$. This is an example of an experiment with a **continuous sample space**. We can define events on a continuous space as well, such as

$$\underbrace{C=\left\{x=\frac{1}{2}\right\}}_{\text{one point: finite}} \qquad \underbrace{A=\left\{x<\frac{1}{2}\right\}}_{\text{upper-bounded interval: infinite}} \qquad \underbrace{B=\left\{\left|x-\frac{1}{2}\right|<\frac{1}{4}\right\}}_{\text{upper- and lower-bounded interval: infinite}}$$

Other examples of experiments with continuous sample spaces include the measurement of the voltage of thermal noise in a resistor and the measurement of the (x,y,z) position of an oxygen molecule in the atmosphere.

A particular experiment can often be represented by more than one sample space. The choice of a particular sample space depends upon the questions that are to be answered concerning the experiment.

Example I.20

Consider the experiment of rolling two dice and observing the results.

- If the dice are distinguishable, and we are interested in what numbers show on the upper faces of the dice, then the sample space consist of the 36 ordered pairs $\{(1,1),(1,2),\cdots,(6,6)\}$.
- If the dice are indistinguishable, and we are interested in what numbers show on the upper faces of the dice, then the sample space consist of the 21 ordered pairs $\{(1,1),(1,2),\cdots,(1,6),(2,2),(2,3),\cdots,(6,6)\}$.
- If we are interested in the sum of the two numbers showing on the upper faces of the two dice, then the sample space consist of the 11 numbers $\{2,3,\cdots,12\}$.

I.3. Basic Combinatorial Analysis

I.3.A. Permutations

The factorial of a non-negative integer number r is given by

$$r!=r(r-1)\cdots(1) \tag{I.1}$$

When $r=0$ or $r=1$, the factorial is one:

$$0!=1 \tag{I.2}$$

$$1! = 1 \tag{I.3}$$

Suppose that D is a set with n elements.

A permutation of size $k \in \{0,1,\cdots,n\}$ from D is an ordered sequence of distinct elements of D; i.e., a sequence of the form $\{x_1, x_2, \cdots, x_k\}$ where $x_i \in D$ for each i and $x_i \neq x_j$ for $i \neq j$.

Example I.21

Let $Q = \{A, B, C, D\}$.

- Permutations of length 1 are: $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$.
- Permutations of length 2 are: $\{A, B\}$, $\{A, C\}$, $\{A, D\}$, $\{B, A\}$, $\{B, C\}$, …, $\{D, C\}$.

Exercise I.3

In Example I.21, write down all permutations of all possible lengths.

The number of permutations of length k from an n-element set is

$$\begin{aligned} n^{(k)} &= \frac{n!}{(n-k)!} \\ &= n(n-1)\cdots(n-k+1) \end{aligned} \tag{I.4}$$

Exercise I.4

In Exercise I.3, determine the numbers of permutations.

Exercise I.5

What is the number of permutations of length n from an n-element set?

I.3.B. Powers of Real Numbers

Let $a \in \mathbb{R}$, $s \in \mathbb{R}$ and $n \in \mathbb{N}$. Define

$$a^{(s,k)} = a(a+s)(a+2s)\cdots(a+(k-1)s) \tag{I.5}$$

Example I.22

Let $a = 3.5$, $s = 1.2$ and $k = 5$. Then,

$$\begin{aligned} 3.5^{(1.2,5)} &= 3.5(4.7)(5.9)(7.1)(8.3) \\ &= 5719.45115 \end{aligned}$$

Note that

$$a^{(0,k)} = \underbrace{a(a)\cdots(a)}_{k \text{ terms}} = a^k \tag{I.6}$$

$$\begin{aligned} a^{(-1,k)} &= a(a-1)\cdots(a-k+1) \\ &= a^{(k)} \\ &= \text{Falling Power of } a \end{aligned} \tag{I.7}$$

$$\begin{aligned} a^{(1,k)} &= a(a+1)\cdots(a+k-1) \\ &= \text{Rising Power of } a \end{aligned} \tag{I.8}$$

$$1^{(1,k)} = k! \tag{I.9}$$

I.3.C. Combinations

A combination of size $k \in \{0,1,\cdots,n\}$ from D is an unordered sequence of distinct elements of D; i.e., a sequence of the form $\{x_1, x_2, \cdots, x_k\}$ where $x_i \in D$ for each i and $x_i \neq x_j$ for $i \neq j$. A combination of size k from D corresponds to an unordered sample of size k chosen without replacement from the population D. For each combination of size k from D, there are $k!$ distinct orderings of the elements of that combination. Hence, the number of combinations is equal to

$$\begin{aligned} C_{n,k} &= \frac{n!}{k!(n-k)!} \\ &= \binom{n}{k} \end{aligned} \tag{I.10}$$

The number $\binom{n}{k}$ is known as the binomial coefficient.

$$\binom{n}{k} = 0, \text{ if } k > n \tag{I.11}$$

$$\binom{n}{k} = 0, \text{ if } k < 0 \tag{I.12}$$

$$\binom{n}{n} = 1 \tag{I.13}$$

$$\binom{n}{0} = 1 \tag{I.14}$$

$$\binom{n}{k} = \binom{n}{n-k} \tag{I.15}$$

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \tag{I.16}$$

I.3.D. Binomial Theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \tag{I.17}$$

I.3.E. Multinomial Theorem

The number of ways to partition a set of n elements into k subsets of sizes $n_1, n_2 \cdots, n_k$ is equal to the multinomial coefficient, given by

$$\begin{aligned}\binom{n}{n_1, n_2, \cdots, n_k} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{k-1}}{n_k} \\ &= \frac{n!}{(n_1)!(n_2)!\cdots(n_k)!}\end{aligned} \tag{I.18}$$

$$(a_1 + a_2 + \cdots + a_k)^n = \underbrace{\sum_{n_1+n_2+\cdots+n_k=n}^{n}}_{\binom{n+k-1}{n} \text{ cases}} \binom{n}{n_1, n_2, \cdots, n_k} a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k} \tag{I.19}$$

Example I.23

$$(a_1 + a_2 + a_3)^4 = \underbrace{\binom{n}{0,0,4} a_1^0 a_2^0 a_3^4 + \binom{n}{0,1,3} a_1^0 a_2^1 a_3^3 + \binom{n}{0,2,2} a_1^0 a_2^2 a_3^2 + \cdots + \binom{n}{4,0,0} a_1^4 a_2^0 a_3^0}_{15 \text{ cases}}$$

Exercise I.6

In an exam with 10 participants, the first, second, and third highest marks are noted. How many outcomes are there?

Example I.24

Four husbands and their wives are to be seated on eight chairs. How many seating arrangements are there in each of the following cases:

- There are no restrictions.

$8! = 40320$

- Men must sit together and women must sit together.

$(8\times 3\times 2\times 1)(4\times 3\times 2\times 1)=1152=(2)(4!)^2$

- Men cannot sit together and women cannot sit together.

$(2)(4!)(4!)=1152$

- Men must sit together.

$(4!)^2(5)=2880$

- Men cannot sit together.

$(4!)^2(5)=2880$

- Each family must sit together.

$(8)(6)(4)(2)=384$

Example I.25

Five engineering books, four science books, and three history books are arranged on a bookshelf. Find the number of arrangements in each of the following cases:

- There are no restrictions.

$12!=479{,}001{,}600$

- The books in each subject must be together.

$(5!\times 4!\times 3!)(3!)=103{,}680$

- The engineering must be together.

$(5!\times 7!)(8)=4{,}838{,}400$

Example I.26

Find the number of distinct arrangements of letter in the following words:

- random

$6!=720$

- signals

$\binom{7}{2}(5!)=(21)(120)=2520$

- statistics

$\binom{10}{3}\binom{7}{3}\binom{4}{2}(2)(1) = (120)(35)(6)(2)(1) = 50400$

Example I.27

How many solutions does the equation $x + y = r$ for integer x, y, and r, $r > 0$ in the following three cases:

- $x, y \geq 0$

$r + 1$

- $x, y \geq 0$, $x \neq y$

$r + 1$ if r is odd and r if r is even

- $x, y > 0$

$r - 1$

Example I.28

How many solutions does the equation $x + y + z = 10$ for integer x, y, and z in the following three cases:

- $x, y, z \geq 0$

$11 + 10 + \cdots 1 = 66$

- $x, y, z \geq 0$, $x \neq y$, $y \neq z$, $x \neq z$

$8 + 8 + 6 + 6 + 4 + 4 + 4 + 4 + 2 + 2 = 48$

- $x, y, z > 0$

$8 + 7 + \cdots + 1 = 36$

I.4. Axioms of Probability

Probability is a function of an event that produces a numerical quantity that measures the likelihood of that event.

There are several ways to assign probabilities to events. All events can have probabilities.

Axiom 1:

For any event A, $\Pr(A) \geq 0$ (a negative probability does not make sense).

$$\Pr(A) \geq 0, \forall A \tag{I.20}$$

Axiom 2:

If S is the sample space for a given experiment, $\Pr(S)=1$ (probabilities are normalized so that the maximum value is unity).

$$\Pr(S)=1 \tag{I.21}$$

Axiom 3a:

If $A \cap B = \phi$, then $\Pr\{A \cup B\} = \Pr(A) + \Pr(B)$.

$$\Pr\{A \cup B\} \begin{cases} = \Pr(A) + \Pr(B), & A \cap B = \phi \\ \neq \Pr(A) + \Pr(B), & A \cap B \neq \phi \end{cases} \tag{I.22}$$

Corollary I.1

Consider $M < \infty$ sets $A_1, A_2, \cdots, A_M$ that are mutually exclusive, $A_i \cap A_j = \phi; \forall i \neq j$,

$$\Pr\left(\bigcup_{i=1}^{M} A_i\right) = \sum_{i=1}^{M} \Pr(A_i), \text{if } A_i \cap A_j = \phi; \forall i \neq j \tag{I.23}$$

Axiom 3b:

For an infinite number of mutually exclusive sets, A_i, $i = 1, 2, 3, \cdots$, $A_i \cap A_j = \phi; \forall i \neq j$,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i), \text{if } A_i \cap A_j = \phi; \forall i \neq j \tag{I.24}$$

Theorem I.1

For any sets A and B (not necessarily mutually exclusive),

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \tag{I.25}$$

I.4.A. Venn Diagrams

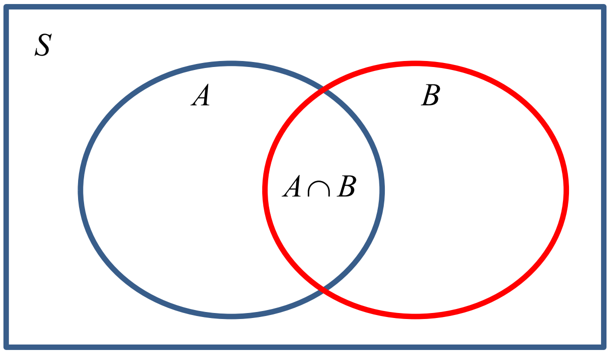


Figure I.2: Venn Diagram to Prove Theorem I.1

Note that $\Pr(A)+\Pr(B)$ involves adding the intersection region twice.

Theorem I.2

$$\Pr(\overline{A}) = 1 - \Pr(A) \tag{I.26}$$

Theorem I.3

$$A \subset B \Rightarrow \Pr(A) \le \Pr(B) \tag{I.27}$$

I.4.B. Atomic Outcomes:

Events that cannot be decomposed into simpler events.

Often, atomic outcomes are assigned equal probabilities.

If there are M mutually exclusive collectively exhaustive atomic outcomes $\xi_1, \xi_2, \cdots, \xi_M$, we could assign (in case there is no information about the likelihood of the outcomes).

$$\Pr(\xi_m) = \frac{1}{M}, \forall m \tag{I.28}$$

For example, (I.28) applies for the numbers on the six faces of a standard cubic die, and the probability of each number is one sixth (1/6). However, (I.28) does not apply if some faces of the die are larger than other faces.

Obviously, when the M outcomes are mutually exclusive collectively exhaustive, we have

$$\begin{gathered} \xi_i \cap \xi_j = \phi, \forall i \ne j \\ \bigcup_{m=1}^{M} \xi_m = S \Rightarrow \Pr\left(\bigcup_{m=1}^{M} \xi_m\right) = 1 \end{gathered} \tag{I.29}$$

Example I.29

Consider the coin flipping experiment of Example I.15. In this case, there are only two atomic events, $\xi_1 = H$ and $\xi_2 = T$. Provided the coin is fair (again, not biased towards one side or the other), we have every reason to believe that these two events should be equally probable. These outcomes are mutually exclusive and collectively exhaustive (provided we rule out the possibility of the coin landing on its edge). According to our theory of probability, these events should be assigned probabilities

$$\Pr(H) = \Pr(T) = \frac{1}{2}$$

Example I.30

Consider the dice rolling experiment of Example I.16. If the die is not loaded, the six possible faces of the cubic die are reasonably taken to be equally likely to appear, in which case, the probability assignment is

$$\Pr(1) = \Pr(2) = \cdots = \Pr(6) = \frac{1}{6}$$

From this assignment we can determine the probability of more complicated events, such as

$$\begin{aligned}\Pr(\text{even number is rolled}) &= \Pr(2 \cup 4 \cup 6) \\ &= \Pr(2)+\Pr(4)+\Pr(6) \\ &= \frac{1}{6}+\frac{1}{6}+\frac{1}{6} \\ &= \frac{1}{2}\end{aligned}$$

Example I.31

In Example I.17, a pair of dice were rolled. In this experiment, the most basic outcomes are the 36 different combinations of the six atomic outcomes of the previous example. Again, each of these atomic outcomes is assigned a probability of 1/36. Next, suppose we want to find the probability of the event $A = \{\text{sum of two dice} = 5\}$. Then,

$$\begin{aligned}\Pr(A) &= \Pr\left\{(1,4) \cup (2,3) \cup (3,2) \cup (4,1)\right\} \\ &= \Pr(1,4) + \Pr(2,3) + \Pr(3,2) + \Pr(4,1) \\ &= \frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36} \\ &= \frac{1}{9}\end{aligned}$$

$$A = \left\{\text{two numbers are identical}\right\}$$

I.4.C. Relative Frequency Approach

The relative frequency approach requires that the experiment we are concerned with be **repeatable**.

The probability of an event, can be assigned by repeating the experiment a large number of times and observing how many times the event actually occurs.

$$\Pr(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

To get an exact measure of the probability of an event using the relative frequency approach, we must repeat the event an **infinite** number of times.

I.5. Joint and Conditional Probabilities

$$\Pr(A, B) = \Pr(A \cap B) \qquad \text{(I.30)}$$

The above can be extended to more than two events.

If A and B are mutually exclusive, then their joint probability is zero $\Rightarrow \Pr(A, B) = \Pr(\phi) = 0$.

Both events (sets) A and B can be expressed in terms of atomic outcomes.

We then write $A \cap B$ as the set of those atomic outcomes that is common to both and calculate the probabilities of each of these outcomes.

Alternatively, we can use the relative frequency approach:

$$\Pr(A, B) = \lim_{n\to\infty} \frac{n_{A,B}}{n} \qquad \text{(I.31)}$$

Example I.32

A standard deck of playing cards has 52 cards that can be divided in several manners. There are *four suits* (spades, hearts, diamonds, and clubs), each of which has 13 cards (ace, 2, 3, 4, . . . , 10, jack, queen, king). There are *two red suits* (hearts and diamonds) and *two black suits* (spades and clubs). Also, the jacks, queens, and kings are referred to as *face cards*, while the others are *number cards*.

Suppose the cards are sufficiently shuffled (randomized) and one card is drawn from the deck. The experiment has 52 atomic outcomes corresponding to the 52 individual cards that could have been selected. Hence, each atomic outcome has a probability of 1/52.

Define the events: A = {red card selected}, B = {number card selected}, and C = {heart selected}.

Since the event A consists of 26 atomic outcomes (there are 26 red cards), then Pr(A) = 26/52 = 1/2.

Likewise, Pr(B) = 40/52 = 10/13 and Pr(C) = 13/52 = 1/4.

Events A and B have 20 outcomes in common, hence Pr(A,B) = 20/52 = 5/13.

Likewise, Pr(A,C) = 13/52 = 1/4 and Pr(B,C) = 10/52 = 5/26.

It is interesting to note that in this example, Pr(A,C) = Pr(C). This is because $C \subset A$ and as a result $A \cap C = C$.

Often the occurrence of one event may be dependent upon the occurrence of another.

In the previous example, the event A = {a red card is selected} had a probability of Pr(A) = 1/2. If it is known that event C = {a heart is selected} has occurred, then the event A is now certain (probability equal to 1), since all cards in the heart suit are red.

Likewise, if it is known that the event C did not occur, then there are 39 cards remaining, 13 of which are red (all the diamonds). Hence, the probability of event A in that case becomes 1/3.

Clearly, the probability of event A depends on the occurrence of event C.

We say that the probability of A is conditional on C. The probability of A given knowledge that the event C has occurred is referred to as the **conditional probability** of A given C.

The shorthand notation $\Pr(A \mid C)$ is used to denote the probability of the event A given that the event C has occurred, or simply the probability of A given C.

Definition I-6

For two events A and B, the probability of A conditioned on knowing that B has occurred is

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}, \Pr(B) \neq 0 \tag{I.32}$$

This definition of conditional probability does indeed satisfy the axioms of probability.

$$\begin{aligned} \Pr(A,B) &= \Pr(A \mid B)\Pr(B) \\ &= \Pr(B \mid A)\Pr(A) \end{aligned} \tag{I.33}$$

$$\begin{aligned} \Pr(A,B,C) &= \Pr(A \mid B,C)\Pr(B,C) \\ &= \Pr(A \mid B,C)\Pr(B \mid C)\Pr(C) \end{aligned} \tag{I.34}$$

Example I.33

Consider the experiment of drawing cards from a standard deck. Suppose that we select two cards at random from the deck.

When we select the second card, we do not return the first card to the deck. We are selecting cards *without replacement*.

The probabilities associated with selecting the second card are slightly different if we have knowledge of which card was drawn on the first selection.

Let A = {first card was a spade} and B = {second card was a spade}. The probability of the event A can be calculated as in the previous example to be Pr(A) = 13/52 = 1/4. Likewise, if we have no knowledge of what was drawn on the first selection, the probability of the event B is the same, Pr(B) = 1/4. To calculate the joint probability of A and B, we have to do some counting.

When we select the first card there are 52 possible outcomes.

Since this card is not returned to the deck, there are only 51 possible outcomes for the second card.

Hence, this experiment of selecting two cards from the deck has $52 * 51$ possible outcomes each of which is equally likely and has a probability of $1/(52*51)$. Therefore,

$$\Pr\left(A,B\right)=\frac{13(12)}{52(51)}=\frac{1}{17}$$

The conditional probability of the second card being a spade given that the first card is a spade is then

$$\begin{aligned}\Pr\left(A,B\right)&=\Pr\left(B\mid A\right)\Pr\left(A\right)\\&=\left(\frac{12}{51}\right)\left(\frac{1}{4}\right)\end{aligned}$$

$$\Pr\left(B\mid A\right)=\frac{\Pr\left(A,B\right)}{\Pr\left(A\right)}=\frac{1/17}{1/4}=\frac{4}{17}$$

However, calculating this conditional probability directly is probably easier than calculating the joint probability. Given that we know the first card selected was a spade, there are now 51 cards left in the deck, 12 of which are spades, thus

$$\Pr\left(B\mid A\right)=\frac{12}{51}=\frac{4}{17}$$

Once this is established, then the joint probability can be calculated as

$$\Pr\left(A,B\right)=\Pr\left(B\mid A\right)\Pr\left(A\right)=\frac{4}{17}\frac{1}{4}=\frac{1}{17}$$

Example I.34

In a game of poker, you are dealt five cards from a standard 52 card deck. What is the probability that you are dealt a flush in spades? (A flush is when you are dealt all five cards of the same suit.)

- What is the probability of a flush in spades?
- What is the probability of a flush in any suit?

Let A_i be the event $\{i$th card dealt to us is a spade$\}$, $i=1,2,3,4,5$. Then

$$\Pr\left(\text{flush in spades}\right)=\Pr\left(A_1,A_2,A_3,A_4,A_5\right)$$

$$\Pr\left(A_1\right)=\frac{1}{4}$$

$$\Pr\left(A_1,A_2\right)=\Pr\left(A_2\mid A_1\right)\Pr\left(A_1\right)=\frac{12}{51}\frac{1}{4}=\frac{1}{17}$$

$$\Pr\left(A_1,A_2,A_3\right)=\Pr\left(A_3\mid A_1,A_2\right)\Pr\left(A_1,A_2\right)=\frac{11}{50}\frac{1}{17}=\frac{11}{850}$$

$$\Pr\left(A_1,A_2,A_3,A_4\right)=\Pr\left(A_4\mid A_1,A_2,A_3\right)\Pr\left(A_1,A_2,A_3\right)=\frac{10}{49}\frac{11}{850}=\frac{11}{4165}$$

$$\Pr\left(A_1,A_2,A_3,A_4,A_5\right)=\Pr\left(A_5\mid A_1,A_2,A_3,A_4\right)\Pr\left(A_1,A_2,A_3,A_4\right)=\frac{9}{48}\frac{11}{4165}=\frac{33}{66,640}$$

$$\Pr\left(\text{flush in any suit}\right)=4\left(\frac{33}{66,640}\right)=\frac{33}{16,660}$$

I.6. Bayes's Theorem

Theorem I.4

For any events A and B such that $\Pr(B)\neq 0$,

$$\Pr\left(A,B\right)=\Pr\left(A\middle|B\right)\Pr(B)=\Pr\left(B\middle|A\right)\Pr(A)\Rightarrow$$

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

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

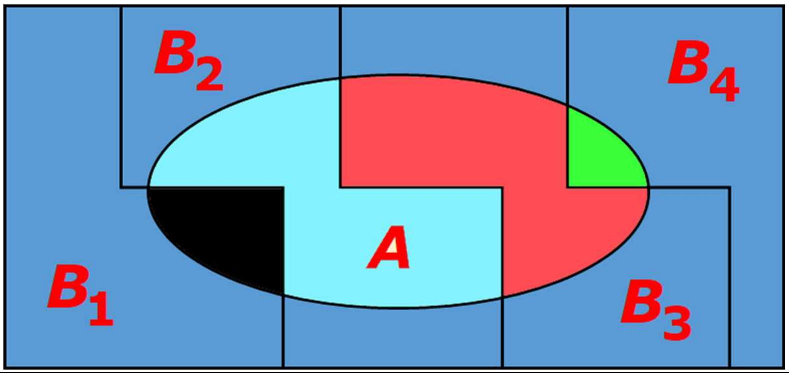
Theorem I.5: Theorem of Total Probability

Let $B_1,B_2,\cdots,B_n$ be a set of mutually exclusive and exhaustive events. That is, $B_i\cap B_j=\phi$ for all $i\neq j$ and

$$\bigcup_{i=1}^{n}B_i=S\Rightarrow\sum_{i=1}^{n}\Pr(B_i)=1$$

Then

$$\Pr(A) = \sum_{i=1}^{n} \Pr\left(A \mid B_i\right) \Pr\left(B_i\right)$$



Theorem I.6: Bayes's Theorem

Let $B_1, B_2, \cdots, B_n$ be a set of mutually exclusive and exhaustive events. Then

$$\Pr\left(B_i \mid A\right) = \frac{\Pr\left(A, B_i\right)}{\Pr(A)} = \frac{\Pr\left(A, B_i\right)}{\sum_{i=1}^{n} \Pr\left(A \mid B_i\right) \Pr\left(B_i\right)}$$

Example I.35

A certain auditorium has 30 rows of seats. Row 1 has 11 seats, while Row 2 has 12 seats, Row 3 has 13 seats, and so on to the back of the auditorium where Row 30 has 40 seats. A prize is to be given away by randomly selecting a row (with equal probability of selecting any of the 30 rows) and then randomly selecting a seat within that row (with each seat in the row equally likely to be selected).

- Find the probability that Seat 15 was selected given that Row 20 was selected.
- Find the probability that Row 20 was selected given that Seat 15 was selected.
- Find the probability that Row 5 was selected given that Seat 15 was selected.

The first task is straightforward. Given that Row 20 was selected, there are 30 possible seats in Row 20 that are equally likely to be selected. Hence, Pr(Seat 15|Row 20) = 1/30.

Without the help of Bayes's theorem, finding the probability that Row 20 was selected given that we know Seat 15 was selected would seem to be a formidable problem.

Using Bayes's theorem,

$$\Pr(\text{Row } 20|\text{Seat } 15)=\frac{\Pr(\text{Seat } 15|\text{Row } 20)\Pr(\text{Row } 20)}{\Pr(\text{Seat } 15)}$$

The two terms in the numerator on the right-hand side are both equal to 1/30. The term in the denominator is calculated using the help of the theorem of total probability.

$$\Pr(\text{Seat } 15)=\left(\sum_{k=5}^{30}\frac{1}{k+10}\right)\frac{1}{30}=0.0342$$

With this calculation completed, the a posteriori probability of Row 20 being selected given seat 15 was selected is given by:

$$\Pr(\text{Row } 20|\text{Seat } 15)=\frac{\frac{1}{30}\frac{1}{30}}{0.0342}=0.0325$$

Note that the a priori probability that Row 20 was selected is 1/30 = 0. 0333. Therefore, the additional information that Seat 15 was selected makes the event that Row 20 was selected slightly less likely.

Using Bayes's theorem again,

$$\Pr(\text{Row } 5|\text{Seat } 15)=\frac{\Pr(\text{Seat } 15|\text{Row } 5)\Pr(\text{Row } 5)}{\Pr(\text{Seat } 15)}$$

$$\Pr(\text{Seat } 15|\text{Row } 5)=\frac{1}{15}$$

$$\begin{aligned}\Pr(\text{Seat } 15,\text{Row } 5)&=\Pr(\text{Seat } 15|\text{Row } 5)\Pr(\text{Row } 5)\\&=\frac{1}{15}\frac{1}{30}\end{aligned}$$

The term in the denominator is the same as the corresponding one in the previous part.

$$\Pr(\text{Seat } 15)=\left(\sum_{k=5}^{30}\frac{1}{k+10}\right)\frac{1}{30}=0.0342$$

Therefore,

$$\Pr(\text{Row } 5|\text{Seat } 15)=\frac{\frac{1}{15}\frac{1}{30}}{0.0342}=0.065$$

The additional information that Seat 15 was selected makes the event that Row 5 was selected more likely.

In some sense, this may be counterintuitive, since we know that if Seat 15 was selected, there are certain rows that could not have been selected (i.e., Rows 1–4 have fewer than 15 seats) and, therefore, we might expect Row 20 to have a slightly higher probability of being selected compared to when we have no information about which seat was selected.

Note that the event that Seat 15 was selected makes some rows much more probable, while it makes others less probable and a few rows now impossible.

I.7. Independence

In Example I.35, it was seen that observing one event can change the probability of the occurrence of another event. In that particular case, knowing that Seat 15 was selected, lowered the probability that Row 20 was selected. We say that the event $A = \{\text{Row 20 was selected}\}$ is statistically dependent on the event $B = \{\text{Seat 15 was selected}\}$.

If the description of the auditorium were changed so that each row had an equal number of seats (e.g., all 30 rows had 20 seats each), then observing the event $B = \{\text{Seat 15 was selected}\}$ would not give us any new information about the likelihood of the event $A = \{\text{Row 20 was selected}\}$. In that case, we say that the events A and B are statistically **independent**.

Mathematically, two events A and B are statistically independent if

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

Note that if $\Pr(A \mid B) = \Pr(A)$, then the following two conditions hold

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

$$\begin{aligned}\Pr(A, B) &= \Pr(A \mid B)\Pr(B) \\ &= \Pr(A)\Pr(B)\end{aligned}$$

Furthermore, if $\Pr(A \mid B) \neq \Pr(A)$, then the other two conditions do not hold.

Definition I-7

Two events are statistically independent if and only if

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

Example I.36

Consider the experiment of tossing two distinguishable dice and observing the numbers that appear on the two upper faces. For convenience, let the first die tossed being red and the second being blue. Let

$A = \{\text{number on the red die is less than or equal to 2}\}$

$B = \{\text{number on the blue die is greater than or equal to } 4\}$

$C = \{\text{the sum of the numbers on the two dice is } 3\}$

Let's compare joint probabilities with products of single event probabilities.

$$\Pr(A) = 1/3$$
$$\Pr(B) = 1/2$$
$$\Pr(C) = 1/18$$

Multiplying each two probabilities above results in

$$\Pr(A)\Pr(B) = \frac{1}{6}$$

$$\Pr(A)\Pr(C) = \frac{1}{54}$$

$$\Pr(B)\Pr(C) = \frac{1}{36}$$

Of the 36 atomic outcomes of the experiment, six belong to the event $A \cap B$, and hence,

$$\Pr(A, B) = \frac{1}{6}$$

Since $\Pr(A, B) = \Pr(A)\Pr(B)$, we conclude that the events A and B are independent.

What about the events A and C ?

Of the 36 possible atomic outcomes of the experiment, two belong to the event $A \cap C$, and hence,

$$\Pr(A, C) = \frac{1}{18}$$

Since $\Pr(A, C) \neq \Pr(A)\Pr(C)$, the events A and C are not independent.

Finally, we look at the pair of events B and C . Clearly, B and C are mutually exclusive. If the white die shows a number greater than or equal to 4, there is no way the sum can be 3.

Hence, $\Pr(B, C) = 0 \neq \Pr(B)\Pr(C)$, and these two events are dependent.

The previous example brings out a point that is worth elaborating on. It is a common mistake to equate mutual exclusiveness with independence. Mutually exclusive events are not the same thing as independent events. In fact, for two events A and B for which $\Pr(\mathrm{A}) \neq 0$ and $\Pr(\mathrm{B}) \neq 0$, A

and B can never be both independent and mutually exclusive. Thus, mutually exclusive events are necessarily statistically dependent.

Note that intersecting events need not be independent. Consider the following example.

Example I.37

Consider the experiment of tossing a fair die and observing the number that appears on the upper face. Consider the events $A=\{2,4,6\}$ and $B=\{2,4,5\}$.

Obviously, $\Pr(A)=\Pr(B)=1/2$, and hence, $\Pr(A)\Pr(B)=1/4$.

However, the joint event $C=A\cap B=\{2,4\}$ has the probability $\Pr(C)=1/3$, which is not equal to the product of the individual event probabilities. Therefore, A and B are dependent events even though they are intersecting.

Definition I-8

The events A, B and C are mutually independent if each pair of events is independent; that is,

$$\begin{gathered}\Pr(A,B)=\Pr(A)\Pr(B)\\ \Pr(A,C)=\Pr(A)\Pr(C),\\ \Pr(B,C)=\Pr(B)\Pr(C)\end{gathered}$$

and in addition,

$$\Pr(A,B,C)=\Pr(A)\Pr(B)\Pr(C).$$

Definition I-9

The events $A_1,A_2,\cdots,A_n$ are independent if any subset of $k<n$ of these events are independent, and in addition

$$\Pr(A_1,A_2,\cdots,A_n)=\Pr(A_1)\Pr(A_2)\cdots\Pr(A_n).$$

Suppose we have some time waveform $X(t)$ which represents a noisy signal that we wish to sample at various points in time, $t_1,t_2,\cdots,t_n$ such that $A_i=X(t_i)$. In some cases, we have every reason to believe that the value of the noise at one point in time does not affect the value of the noise at another point in time. Hence, we assume that these events are independent and write

$$\Pr\left(A_1,A_2,\cdots,A_n\right)=\Pr\left(A_1\right)\Pr\left(A_2\right)\cdots\Pr\left(A_n\right)$$

I.8. Discrete Random Variables

Suppose we conduct an experiment which has some sample space S. Furthermore, let ξ be some outcome defined on the sample space S. It is useful to define functions $X=f(\xi)$ of the outcome

ξ. The function f has as its domain all possible outcomes associated with the experiment. The range of the function f will depend upon how it maps outcomes to numerical values but in general will be the set of real numbers or a subset of the set of real numbers.

Definition I-10

A random variable is a real-valued function of the elements of a sample space S. Given an experiment with sample space S, the random variable X maps each possible outcome $\xi \in S$ to a real number $X = f(\xi)$ as specified by some rule.

If the mapping $X(\xi)$ is such that the random variable X takes on a finite or countably infinite number of values, then we refer to X as a discrete random variable; whereas, if the range of $X(\xi)$ is an uncountably infinite number of points, we refer to X as a continuous random variable.

Since $X = f(\xi)$ is a random variable whose numerical value depends on the outcome of an experiment, we give X a probabilistic description by stating the probabilities that the variable X takes on a specific value or values (e.g., $\Pr(X = 3)$ or $\Pr(X > 8)$). For now, we will focus on random variables that take on discrete values and will describe these random variables in terms of probabilities of the form $\Pr(X = x)$.

Definition I-11

The probability mass function (PMF) $P_X(x)$ of a random variable X is a function that assigns a probability to each possible value of the random variable X. The probability that the random variable X takes on the specific value x is the value of the probability mass function for x. That is,

$$P_X(x) = \Pr(X = x)$$

Example I.38

A discrete random variable may be defined for the random experiment of flipping a coin. The sample space of outcomes is $S = \{H, T\}$. We could define the random variable X to be $X(H) = 0$ and $X(T) = 1$. That is, the sample space is mapped to the set $\{0,1\}$ by the random variable X.

Assuming a fair coin, the resulting probability mass function is

$$P_X(0) = \Pr(X = 0) = \Pr(H) = \frac{1}{2}$$

$$P_X(1) = \frac{1}{2}$$

Note that the mapping is not unique and we could have just as easily mapped the sample space to any other pair of real numbers (e.g., $\{-1,+1\}$).

Example I.39

Suppose we repeat the experiment of flipping a fair coin n times and observe the sequence of heads and tails. A random variable Y could be defined to be the number of times tails occurs in n trials. It turns out that the probability mass function for this random variable is

$$P_Y(k) = \binom{n}{k}\left(\frac{1}{2}\right)^n, \quad k = 0,1,\cdots,n$$

The details of how this PMF is obtained will be deferred until later in this section.

Example I.40

Again, let the experiment be the flipping of a coin, and this time we will continue repeating the trials until the first time a heads occurs. The random variable Z will represent the number of times until the first occurrence of a heads. In this case, the random variable Z can take on any positive integer value $1 \le Z < \infty$. The probability mass function of the random variable Z can be worked out as follows:

$$\begin{aligned} P_Z(n) &= \Pr(Z = n) = \Pr(n-1 \text{ tails followed by one heads}) \\ &= \left(\Pr(T)\right)^{n-1}\Pr(H) \\ &= \left(\frac{1}{2}\right)^{n-1}\frac{1}{2} \\ &= \left(\frac{1}{2}\right)^n \end{aligned}$$

Hence,

$$P_Z(n) = \left(\frac{1}{2}\right)^n, \; n = 1,2,3,\cdots$$

Note that the following must always hold for discrete random variables

$$0 \le P_X(x) \le 1$$

$$\sum_x P_X(x) = 1$$

I.8.A. Bernoulli Random Variable

Consider an experiment with the sample space $S = \{0,1\}$. Let

$$P_X(1) = p$$
$$P_X(0) = 1 - p$$

X is called a Bernoulli random variable.

I.8.B. Binomial Random Variable

Consider repeating a Bernoulli trial n times, where the outcome of each trial is independent of all others. We say that the repeated experiment has a sample space of $S_n = \{0,1\}^n$, which is referred to as a Cartesian space. That is, outcomes of the repeated trials are represented as n element vectors whose elements are taken from S.

Consider, for example, the outcome

$$\xi_k = \left(\underbrace{\underbrace{1,1,\cdots,1}_{k}, \underbrace{0,0,\cdots,0}_{n-k}}_{n} \right)$$

The probability of this outcome occurring is

$$\Pr\{\xi_k\} = p^k (1-p)^{n-k}$$

The probability does not change if we shuffle the digits. The order of the 1s and 0s in the sequence is irrelevant.

Let the random variable X represent the number of times the outcome 1 has occurred in the sequence of n trials. This is known as a binomial random variable and takes on integer values from 0 to n.

To find the probability mass function of the binomial random variable, let A_k be the set of all outcomes that have exactly k 1s and $n-k$ 0s. Note that all outcomes in this event occur with the same probability. Furthermore, all outcomes in this event are mutually exclusive.

$$\begin{aligned} P_X(k) &= \Pr(A_k) \\ &= (\#\text{ of outcomes in } A_k) \,.\, (\text{probability of each outcome in } A_k) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

The binomial coefficient is given by

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

As a check, we verify that this probability mass function is properly normalized:

$$\sum_{k=0}^{n} P_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$
$$= (p+1-p)^n$$
$$= 1$$

In this calculation, we have used the binomial expansion

$$(\alpha+\beta)^n = \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-k}$$

I.8.C. Poisson Random Variable

Consider a binomial random variable X where the number of repeated trials n is very large. In that case, evaluating the binomial coefficients can pose numerical problems. If the probability of success in each individual trial p is very small, then the binomial random variable can be well approximated by a Poisson random variable. That is, the Poisson random variable is a limiting case of the binomial random variable. Formally, let n approach infinity and p approach zero in such a way that

$$\lim_{n\to\infty} np = \alpha$$

Then the binomial probability mass function converges to the form

$$P_X(m) = \frac{\alpha^m}{m!} e^{-\alpha}, \quad m = 0,1,2,\cdots$$

which is the probability mass function of a Poisson random variable. We see that the Poisson random variable is properly normalized by noting that

$$\sum_{m=0}^{\infty} P_X(m) = 1$$

The number of customers arriving at a cashier in a store during some time interval may be well modeled as a Poisson random variable, as may the number of data packets arriving at a given node in a computer network.

I.8.D. Geometric Random Variable

Consider repeating a Bernoulli trial until the first occurrence of the outcome ξ_0. If X represents the number of times the outcome ξ_1 occurs before the first occurrence of ξ_0, then X is a geometric random variable whose probability mass function is

$$P_X(k) = (1-p)\, p^k, \quad k = 0,1,2,\cdots$$

We might also formulate the geometric random variable in a slightly different way. Suppose X counted the number of trials that were performed until the first occurrence of ξ_0. Then the probability mass function would take on the form

$$P_X(k) = (1-p)\,p^{k-1}, \quad k = 1, 2, \cdots$$

The geometric random variable can also be generalized to the case where the outcome ξ_0 must occur exactly m times. We can derive the form of the probability mass function for the generalized geometric random variable from what we know about binomial random variables.

For the m-th occurrence of ξ_0 to occur on the k-th trial, then the first $k-1$ trials must have had $m-1$ occurrences of ξ_0 and $k-m$ occurrences of ξ_1. Then

$$\begin{aligned} P_X(k) &= \Pr\left(\{m-1 \text{ occurrences of } \xi_0 \text{ in } k-1 \text{ trials}\} \cap \{\xi_0 \text{ occurs on the } k^{\text{th}} \text{ trial}\}\right) \\ &= \binom{k-1}{m-1} p^{k-m}(1-p)^{m-1}(1-p) \\ &= \binom{k-1}{m-1} p^{k-m}(1-p)^{m}, \quad k = m, m+1, \cdots \end{aligned}$$

This generalized geometric random variable sometimes goes by the name of a Pascal random variable or the negative binomial random variable.

***

II. Random Variables, Distributions, and Density Functions

II.1. Introduction

Discrete random variables have just been described by their probability mass functions. While this description works fine for discrete random variables, it is inadequate to describe random variables that take on a continuum of values.

In this chapter, we introduce the cumulative distribution function as an alternative description of random variables that is appropriate for describing continuous as well as discrete random variables. The probability density function is also covered.

Consider a discrete random variable X that takes on values from the set $\{0, 1/N, 2/N, \cdots, (N-1)/N\}$ with equal probability. Then

$$P_X\left(\frac{k}{N}\right) = \frac{1}{N}, \quad k = 0, 1, \cdots, N-1$$

This is the type of random variable that is produced by "random" number generators in software packages like MATLAB and Mathematica. In these cases, N is taken to be a fairly large number so that it appears that the random number can be anything in the continuous range [0, 1). Consider the limiting case as $N \to \infty$; so that the random variable can truly fall anywhere in the interval [0,1). Then

$$\begin{aligned} P_X\left(\frac{k}{N}\right) &= \lim_{N\to\infty} \frac{1}{N} \\ &= 0 \end{aligned}$$

That is, each point has zero probability of occurring. Yet, something has to occur! This problem is common to continuous random variables, and it is clear that the probability mass function is not a suitable description for such a random variable.

II.2. The Cumulative Distribution Function (CDF)

Definition II-1

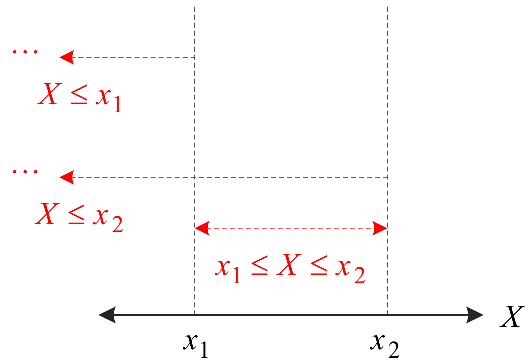
The cumulative distribution function (CDF) of a random variable X is given by

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

CDF Properties

- Since the CDF is a probability, it must take on values between 0 and 1.
- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.

- For $x_1 \le x_2 \Rightarrow \{X \le x_1\} \subseteq \{X \le x_2\} \Rightarrow F_X(x_1) \le F_X(x_2)$. This implies that the CDF is a monotonic non-decreasing function.



- $\{X \le x_2\} = \{X \le x_1\} \cup \{x_1 \le X \le x_2\} \Rightarrow F_X(x_2) = F_X(x_1) + \Pr(\{x_1 \le X \le x_2\})$.
- $\Pr(\{x_1 \le X \le x_2\}) = F_X(x_2) - F_X(x_1)$.

Example II.1

Which of the following mathematical functions can be the CDF of some random variable?

a. $F_X(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)$

b. $F_X(x) = \left(1 - e^{-x}\right)u(x)$

c. $F_X(x) = e^{-x^2}$

d. $F_X(x) = x^2 u(x)$

To determine this, we need to check that $F_X(-\infty) = 0$, $F_X(\infty) = 1$ and that the function is monotonically increasing in between. The first two functions satisfy these properties and thus are valid CDFs. The third function is decreasing for positive values of x, while the forth function takes on values greater than 1 and $F_X(\infty) \ne 1$.

Let's return to the computer random number generator that generates N possible values from the set $\{0, 1/N, 2/N, \cdots, (N-1)/N\}$ with equal probability. The CDF of this random variable is illustrated below.

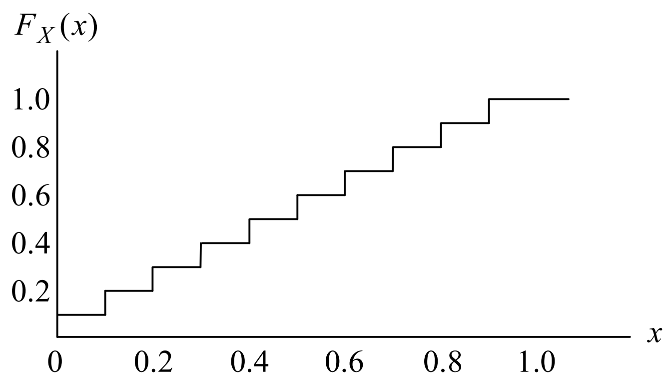


Figure II.1: CDF of the random variable X for $N = 10$

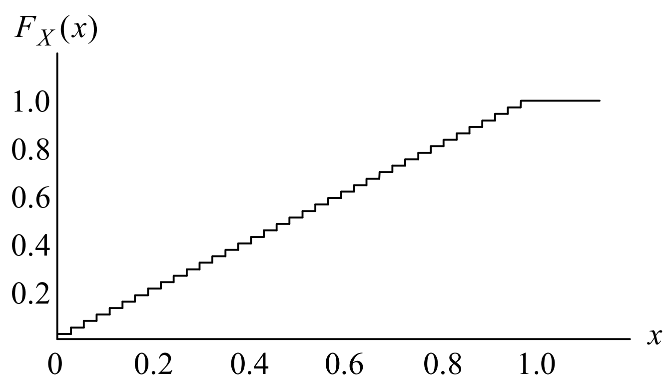


Figure II.2: CDF of the random variable X for $N = 50$

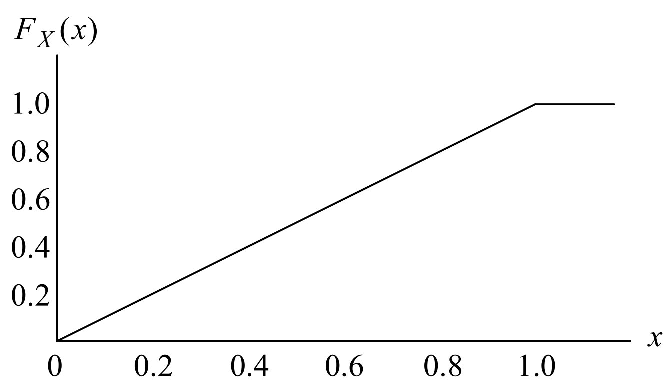


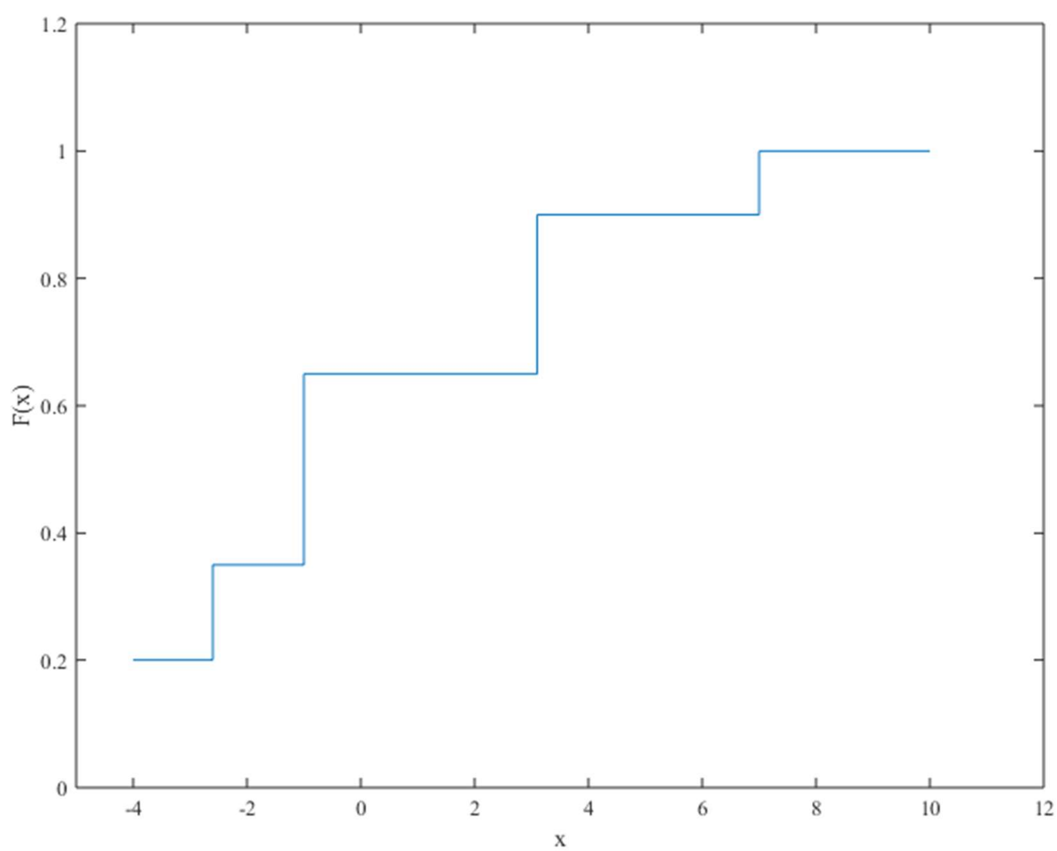
Figure II.3: CDF of the random variable X for $N \to \infty$

Note that when $N \to \infty$ we have

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$$

Example II.2

Suppose a random variable has the value $\{-4,-2.6,-1,3.1,7\}$ with probabilities $\{0.2,0.15,0.3,0.25,0.1\}$. Sketch the CDF.



In this limiting case, the random variable is a continuous random variable and takes on values in the range [0, 1) with equal probability. It is referred to as a uniform random variable. Continuous random variables have a continuous CDF, while discrete random variables have a discontinuous CDF with a staircase type of function.

Example II.3

Suppose a random variable has a CDF given by $F_X(x) = \left(1 - e^{-x}\right)u(x)$. Find the following quantities:

$$\begin{aligned}\Pr(X > 5) &= \Pr(5 < X < \infty) \\ &= 1 - F_X(5) \\ &= 1 - (1 - e^{-5})u(5) = e^{-5}\end{aligned}$$

$$\begin{aligned}\Pr(3 < \mathrm{X} < 7) &= \Pr(3 \le \mathrm{X} \le 7) \\ &= e^{-3} - e^{-7}\end{aligned}$$

$$\Pr(X > 5 \mid X < 7) = \frac{\Pr\left(\{X > 5\} \cap \{X < 7\}\right)}{\Pr(X < 7)} = \frac{\Pr(5 < X < 7)}{\Pr(X < 7)} = \frac{F_X(7) - F_X(5)}{F_X(7)} = \frac{e^{-5} - e^{-7}}{1 - e^{-7}}$$

For a discrete random variable, and for $x_k \le x < x_{k+1}$

$$F_X(x) = \sum_{i=1}^{k} P_X(x_i)u(x - x_i)$$

II.3. The Probability Density Function

Definition II-2

The probability density function (PDF) of the random variable X evaluated at the point x is

$$\begin{aligned}f_X(x) &= \lim_{\varepsilon \to 0} \frac{\Pr(x \le X < x + \varepsilon)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{F_X(x + \varepsilon) - F_X(x)}{\varepsilon} \\ &= \frac{d}{dx} F_X(x)\end{aligned}$$

II.3.A. PDF Properties

$$f_X(x) \ge 0$$

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

$$F_X(x) = \int_{-\infty}^{x} f_X(\gamma)d\gamma = \int_{-\infty}^{x} f_X(\xi)d\xi$$
$$= \int f_X(x)dx$$

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(\gamma)d\gamma$$
$$= \Pr\left(X \le x_0\right)$$

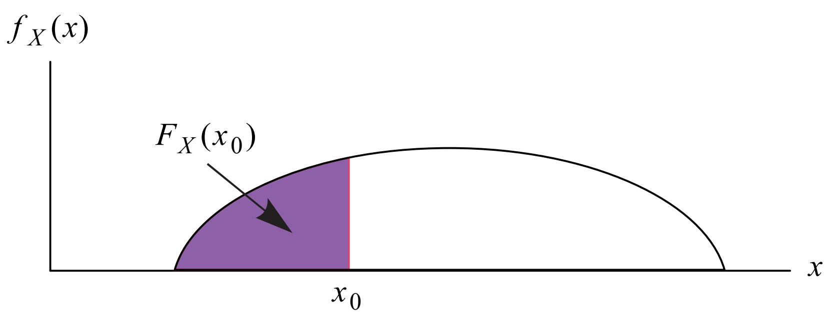


Figure II.4: Area under PDF

$$\int_{-\infty}^{\infty} f_X(x)dx = F_X(\infty) \tag{II.1}$$
$$= 1$$

$$\int_a^b f_X(x)dx = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx$$
$$= F_X(b) - F_X(a) \tag{II.2}$$
$$= \Pr(a < X \le b)$$

Example II.4

Which of the following are valid probability density functions?

a. $f_X(x) = e^{-x}u(x)$ $\quad f_X(x) = \frac{1}{a}e^{-x/b}u(x),\ b > 0,\ a = b$

b. $f_X(x) = e^{-|x|}$

c. $f_X(x) = \frac{1}{a}e^{-|x|},\ a = 2$ $\quad f_X(x) = \frac{1}{a}e^{-|x/b|},\ a = |2b|$

d. $f_X(x) = \begin{cases} \frac{3}{4}(x^2 - 1), & |x| < 2 \\ 0, & \text{otherwise} \end{cases}$

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

f. $f_X(x) = 2xe^{-x^2}u(x)$

The function in *b* is not properly normalized, and is not a PDF.

The function in *d* takes on negative values, and is not a PDF.

The functions in *a*, *e* and *f* are valid PDFS.

II.3.B. The Gaussian (Normal) Random Variable

Definition II-3

A Gaussian random variable is one whose probability density function can be written in the general form

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

The Gaussian PDF is centered about the point $x = m$ and has a width that is proportional to σ.

When $m = 0$ and $\sigma = 1$, X is called a "standard normal" random variable. It is standard practice to introduce a shorthand notation to describe a Gaussian random variable $X \sim \mathcal{N}\left(m, \sigma^2\right)$.

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(\gamma-m)^2}{2\sigma^2}} d\gamma$$

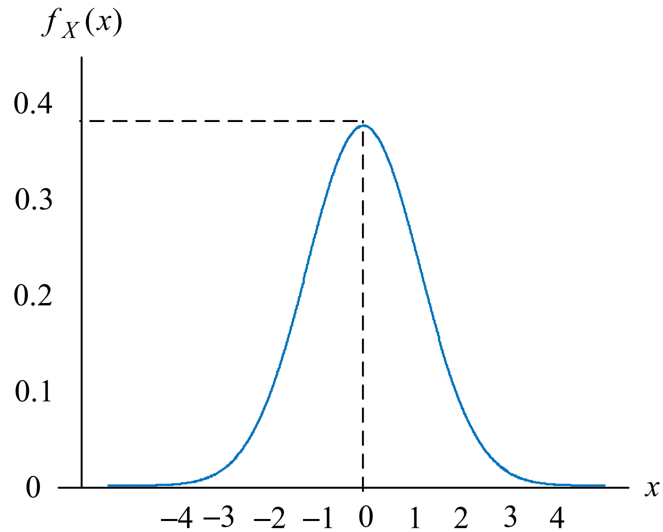


Figure II.5: PDF of a standard normal random variable

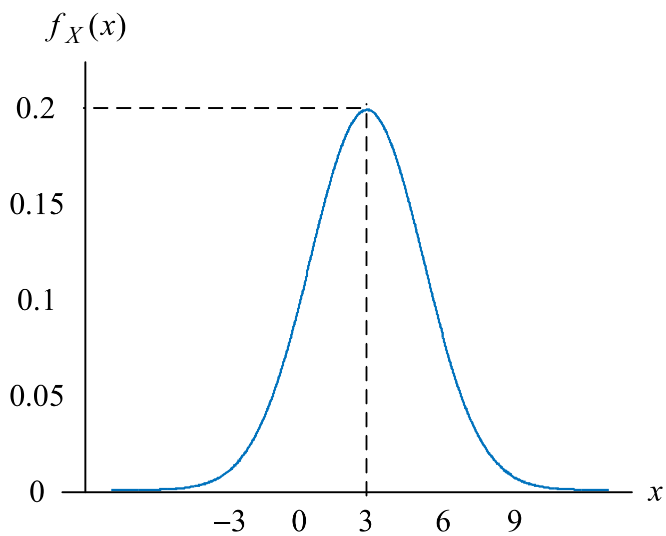


Figure II.6: PDF of a normal random variable with mean 3 and variance 4

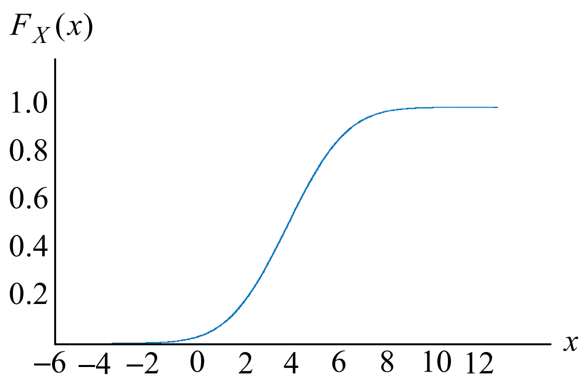


Figure II.7: CDF of a normal random variable with mean 3 and variance 4

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(\gamma-m)^2}{2\sigma^2}} d\gamma$$

For a standard gaussian random variable

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\gamma^2}{2}} d\gamma = 1 - Q(x)$$

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-\gamma^2} d\gamma \tag{II.3}$$

$$\begin{aligned} \text{erfc}(x) &= \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-\gamma^2} d\gamma \\ &= 1 - \text{erf}(x) \end{aligned} \tag{II.4}$$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{\gamma^2}{2}} d\gamma \tag{II.5}$$

$$F_X(x) = 1 - Q\left(\frac{x-m}{\sigma}\right) \tag{II.6}$$

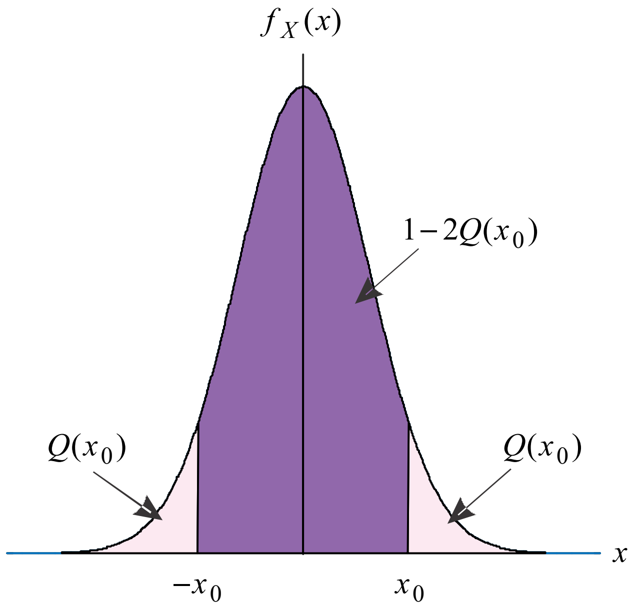


Figure II.8: Q function

$$Q(-x_0) = 1 - Q(x_0)$$

Example II.5

A random variable has a PDF given by

$$f_X(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{(x+3)^2}{8}}$$

$m = -3$, $\sigma = 2$.

Determine

$$\Pr\left(\left|X-3\right| > 6\right) = \Pr(X > 9 \text{ or } X < -3)$$

$$\begin{aligned}
\Pr\left(\left|X+3\right| < 2\right) &= \Pr(-5 < X < -1) \\
&= F_X(-1) - F_X(-5) \\
&= 1 - Q\left((-1+3)/2\right) - \left[1 - Q\left((-5+3)/2\right)\right] \\
&= 1 - Q(1) - \left[1 - Q(-1)\right] \\
&= \Pr(X > -5) - \Pr(X > -1) \\
&= Q(-1) - Q(1) \\
&= 1 - 2Q(1)
\end{aligned}$$

$$\Pr\left(|X-2|>1\right)=\Pr(X<1)+\Pr(X>3)$$
$$=1-Q(2)+Q(3)$$

II.3.C. Uniform Random Variable

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

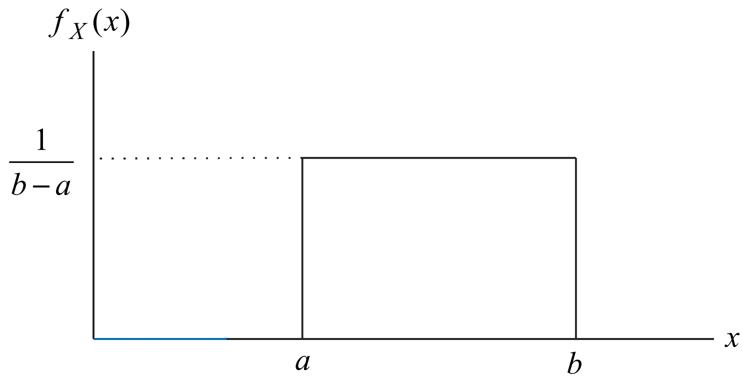


Figure II.9: PDF of a uniform random variable

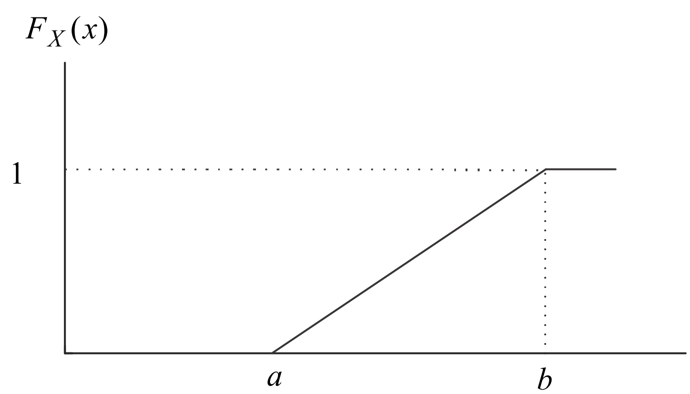


Figure II.10: CDF of a uniform random variable

II.3.D. Exponential Random Variable

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

$$F_X(x)=\left(1-e^{-\frac{x}{b}}\right)u(x)$$

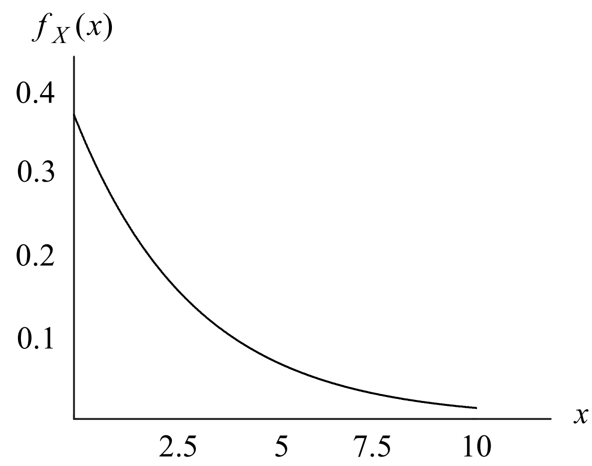


Figure II.11: PDF of an exponential random variable with $b=3$

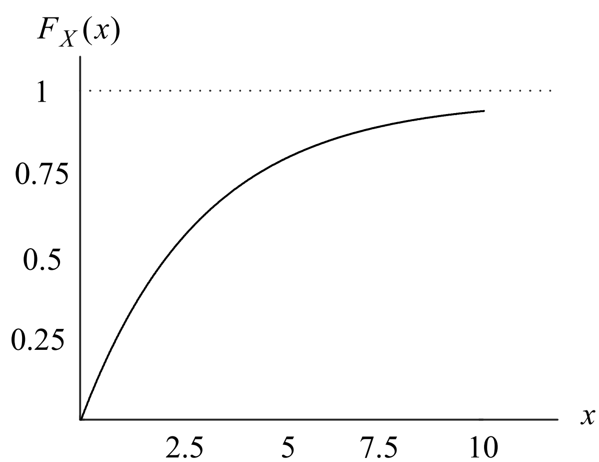


Figure II.12: CDF of an exponential random variable with $b=3$

II.3.E. Laplace Random Variable

$$f_X(x)=\frac{1}{2b}e^{-\frac{|x|}{b}},\quad b>0$$

$$F_X(x)=\begin{cases}\frac{1}{2}e^{\frac{x}{b}}, & x<0\\ 1-\frac{1}{2}e^{-\frac{x}{b}}, & x\ge 0\end{cases}$$

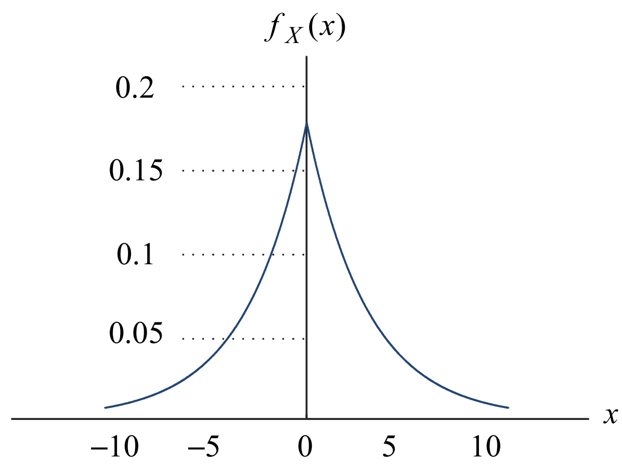


Figure II.13: PDF of a Laplace random variable with $b = 3$

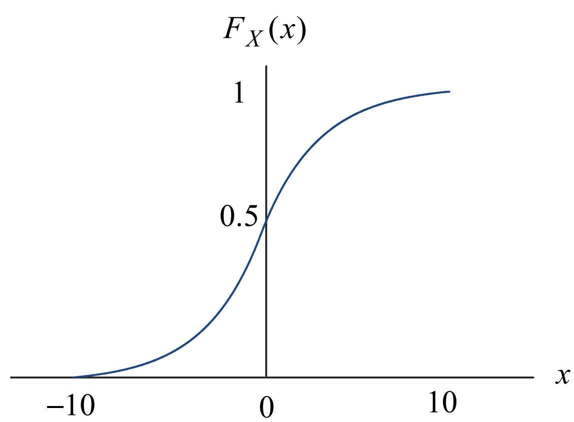


Figure II.14: CDF of a Laplace random variable with $b = 3$

II.3.F. Gamma Random Variable

$$f_X(x) = \frac{\left(\frac{x}{b}\right)^{c-1} e^{-\frac{x}{b}}}{b\Gamma(c)} u(x), \quad b > 0, c > 0$$

$$F_X(x) = \frac{\gamma\left(c, \frac{x}{b}\right)}{\Gamma(c)} u(x)$$

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$$

When α is an integer,

$$\Gamma(\alpha) = (\alpha - 1)!, \quad \alpha > 0$$

In other words,

$$\alpha! = \alpha\Gamma(\alpha)$$

For integer and non-integer α,

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \alpha > 0$$

$$\gamma(\alpha, \beta) = \int_0^{\beta} e^{-t} t^{\alpha-1} dt$$

Special Cases

1. c is integer: Erlang Random Variable
2. $b = 2$ and c is half integer: χ^2 Random Variable.
3. $c = 1$: Exponential Random Variable.

II.3.G. Erlang Random Variable

$$f_X(x) = \frac{\left(\frac{x}{b}\right)^{n-1} e^{-\frac{x}{b}}}{b(n-1)!} u(x)$$

$$F_X(x) = \left[1 - e^{-\frac{x}{b}} \sum_{m=0}^{n-1} \frac{\left(\frac{x}{b}\right)^m}{m!}\right] u(x)$$

The Erlang distribution plays a fundamental role in the study of wireline telecommunication networks. In fact, this random variable plays such an important role in the analysis of trunked telephone systems that the amount of traffic on a telephone line is measured in Erlangs.

II.3.H. Chi-Squared Random Variable

$$f_X(x) = \frac{x^{c-1} e^{-\frac{x}{2}}}{2^c \Gamma(c)} u(x)$$

$$F_X(x) = \frac{\gamma\left(c, \frac{x}{2}\right)}{\Gamma(c)} u(x)$$

II.3.I. Rayleigh Random Variable

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} u(x), \quad \sigma > 0$$

$$F_X(x) = \left(1 - e^{-\frac{x^2}{2\sigma^2}}\right) u(x)$$

The Rayleigh distribution arises when studying the magnitude of a complex number whose real and imaginary parts both follow a zero-mean Gaussian distribution. The Rayleigh distribution arises often in the study of noncoherent communication systems and also in the study of land mobile communication channels, where the phenomenon known as fading is often modeled using Rayleigh random variables.

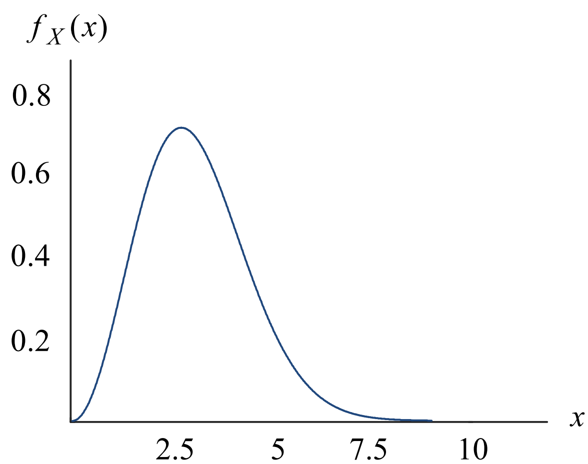


Figure II.15: PDF of a Rayleigh random variable with $\sigma = 2$

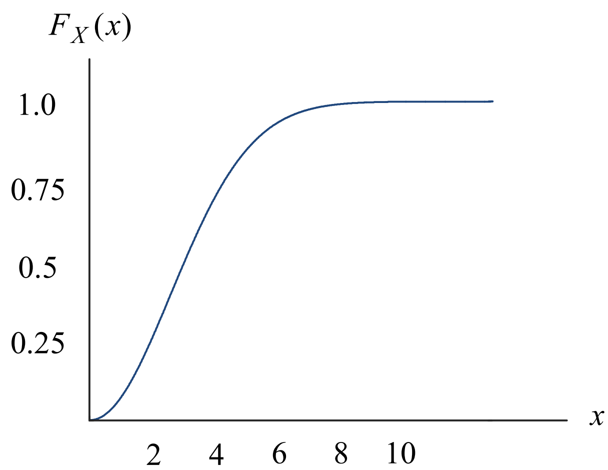


Figure II.16: CDF of a Rayleigh random variable with $\sigma = 2$

II.3.J. Rician Random Variable

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2+a^2}{2\sigma^2}} I_0\left(\frac{ax}{\sigma^2}\right) u(x), \quad a > 0, \sigma > 0$$

In this expression, the function $I_0(x)$ is the modified Bessel function of the first kind of order zero, which is defined by

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x\cos(\theta)} d\theta$$

Marcum's Q-function which describes the CDF of a Rician random variable. It is defined by

$$Q(\alpha, \beta) = \int_\beta^\infty z e^{-\frac{z^2+a^2}{2}} I_0(\alpha z) dz$$

$$F_X(x) = 1 - Q\left(\frac{a}{\sigma}, \frac{x}{\sigma}\right)$$

II.3.K. Cauchy Random Variable

$$f_X(x) = \frac{b}{\pi\left(b^2 + (x-a)^2\right)}, \quad b > 0$$

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x-a}{b}\right)$$

II.4. Conditional Distribution and Density Functions

Definition II-4

The conditional cumulative distribution function of a random variable X conditioned on that the event A has occurred is given by:

$$F_{X|A}(x) = \Pr(X \le x \mid A) = \frac{\Pr\left(\{X \le x\} \cap A\right)}{\Pr(A)}, \quad \Pr(A) \ne 0$$

Example II.6

Suppose a random variable X is uniformly distributed over the interval $[0,1)$ so that its CDF is given by:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

Let $A = \left\{X < \frac{1}{2}\right\}$, then $F_{X|A}(x) = \dfrac{\Pr\left(\{X \le x\}, \{X < 1/2\}\right)}{\Pr\left(\{X < 1/2\}\right)}$

$$x < 0 \Rightarrow \{X \le x\} \cap \{X < 1/2\} = \{X \le x \mid x < 0\} \Rightarrow F_{X|A}(x) = 0$$

$$0 \le x \le 1/2 \Rightarrow \{X \le x\} \cap \{X < 1/2\} = \{X \le x \mid 0 \le x \le 1/2\} \Rightarrow F_{X|A}(x) = \frac{x}{1/2} = 2x$$

$$x > 1/2 \Rightarrow \{X \le x\} \cap \{X < 1/2\} = \{X < 1/2\} \Rightarrow F_{X|A}(x) = 1$$

Then

$$F_{X|A}(x) = \begin{cases} 0, & x < 0 \\ 2x, & 0 \le x \le 1/2 \\ 1, & x > 1/2 \end{cases}$$

If $A = \{a < X \le b\}$ and $a < b$, then

$$F_{X|A}(x) = \begin{cases} 0, & x < a \\ \dfrac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}, & a < x \le b \\ 1, & x > b \end{cases}$$

Definition II-5

The conditional probability density function of a random variable X conditioned on some event A is

$$f_{X|A}(x) = \frac{d}{dx} F_{X|A}(x)$$

$$f_{X|\{a \le X < b\}}(x) = \begin{cases} \dfrac{f_X(x)}{\Pr(a \le X < b)}, & a \le x < b \\ 0, & \text{otherwise} \end{cases}$$

Example II.7

Let X be a random variable representing the length of time we spend waiting in the grocery store checkout line. Suppose the random variable X has an exponential PDF given by

$$f_X(x) = \frac{1}{c} e^{-\frac{x}{c}} u(x)$$

Let $c = 3$. What is the PDF for the amount of time we spend waiting in line given that we have already been waiting for 2 minutes?

$A = \{X > 2\}$. Use equation in Definition II-5 with $a = 2$ and $b = \infty$.

$$\Pr(A) = 1 - F_X(2) = e^{-\frac{2}{3}}$$

Therefore,

$$f_{X|\{a \le X < b\}}(x) = \frac{f_X(x)}{e^{-\frac{2}{3}}} u(x-2) = \frac{1}{3} e^{-\frac{x-2}{3}} u(x-2)$$

***

III. Operations on a Single Random Variable

In this chapter we introduce several mathematical operations that can be applied to single random variables.

III.1. Expected Value of a Random Variable

Definition III-1

The expected value of a random variable X which has a PDF $f_X(x)$ is defined as

$$\begin{aligned} \mathrm{E}\left[X\right] &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \overline{X} \\ &= \mu_X \end{aligned} \tag{III.1}$$

For a discrete random variable

$$f_X(x) = \sum_k P_X(x_k)\delta(x - x_k) \tag{III.2}$$

Then, the expected value of a discrete random variable is

$$\mathrm{E}\left[X\right] = \sum_k x_k P_X(x_k) \tag{III.3}$$

For example, an exam is held for the 22 students in the 360 summer 2020 class. Grades are distributed as follows:

7 students got 70

4 students got 80

2 students got 90

6 students got 60

3 students got 50

$$\begin{aligned} \overline{X} &= \frac{7(70) + 4(80) + 2(90) + 6(60) + 3(50)}{22} \\ &= \frac{7}{22}(70) + \frac{4}{22}(80) + \frac{2}{22}(90) + \frac{6}{22}(60) + \frac{3}{22}(50) = 68.18 \end{aligned}$$

Example III.1

Consider a discrete random variable that has the values $\{1,2,4,7,11\}$, with respective probabilities $\{0.35,0.1,0.15,0.2,0.2\}$. The mean value of this random variable is equal to

$$\begin{aligned}\mu &= 0.35(1)+0.1(2)+0.15(4)+0.2(18)\\ &= 0.35+0.2+0.6+3.6\\ &= 4.75\end{aligned}$$

Example III.2

Consider a random variable that is uniform over the interval $[-4,9]$. The mean value of this random variable is equal to

$$\begin{aligned}\mu &= \frac{1}{13}\int_{-4}^{9} x dx\\ &= \frac{1}{13}\frac{x^2}{2}\Big|_{-4}^{9}\\ &= \frac{1}{13}\left(\frac{9^2}{2}-\frac{(-4)^2}{2}\right)\\ &= 2.5\end{aligned}$$

Example III.3

Consider a random variable X that has the PDF

$$f_X(x)=\begin{cases}\frac{x}{18}, & 0\le x\le 6\\ 0, & \text{otherwise}\end{cases}$$

The mean value of this random variable is equal to

$$\begin{aligned}\mathrm{E}[X] &= \frac{1}{18}\int_0^6 x^2 dx\\ &= \frac{1}{18}\frac{6^3}{3}\\ &= 4\end{aligned}$$

Example III.4

Consider a random variable that has an exponential PDF given by

$$f_X(x) = \frac{1}{\gamma} e^{-\frac{x}{\gamma}} u(x)$$

$$\begin{aligned} \mathrm{E}[X] &= \frac{1}{\gamma} \int_0^\infty x e^{-\frac{x}{\gamma}} dx \\ &= \gamma \end{aligned}$$

Example III.5

Consider a Poisson random variable that has the PMF given by

$$P_X(k) = \frac{\alpha^k e^{-\alpha}}{k!}, k = 0, 1, 2, \cdots$$

The expected value of this random variable is found as follows:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=0}^{\infty} k \frac{\alpha^k e^{-\alpha}}{k!} \\ &= \sum_{k=1}^{\infty} k \frac{\alpha^k e^{-\alpha}}{k!} \\ &= \alpha e^{-\alpha} \sum_{k=1}^{\infty} \frac{\alpha^{k-1}}{(k-1)!} \\ &= \alpha e^{-\alpha} \sum_{m=0}^{\infty} \frac{\alpha^m}{m!} \\ &\alpha \end{aligned}$$

Example III.6

Consider a Rayleigh random variable with the PDF

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} u(x)$$

The mean is calculated as follows:

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

$$=\sigma\sqrt{2}\int_0^\infty \sqrt{y}e^{-y}dy$$

$$=\sigma\sqrt{\frac{\pi}{2}}$$

Definition III-2

Given a random variable X with PDF $f_X(x)$, the expected value of a function $g(X)$ of that random variable is given by

$$\mathrm{E}\left[g(X)\right]=\int_{-\infty}^{\infty} g(x)f_X(x)dx \tag{III.4}$$

For a discrete random variable, this definition reduces to

$$\mathrm{E}\left[g(x)\right]=\sum_k g(x_k)P_X(x_k) \tag{III.5}$$

Theorem III.1

For any constants a and b,

$$\begin{aligned}\mathrm{E}\left[aX+b\right]&=\int_{-\infty}^{\infty}(ax+b)f_X(x)dx\\&=a\int_{-\infty}^{\infty}xf_X(x)dx+b\int_{-\infty}^{\infty}f_X(x)dx\\&=a\,\mathrm{E}\left[X\right]+b\end{aligned} \tag{III.6}$$

$$E\left[\sum_m g_m(X)\right]=\sum_m E\left[g_m(X)\right] \tag{III.7}$$

III.2. Moments

Definition III-3

The nth moment of a random variable X is defined as

$$E\left[X^n\right] = \int_{-\infty}^{\infty} x^n f_X(x)dx \qquad \text{(III.8)}$$

For a discrete random variable, this definition reduces to

$$E\left[X^n\right] = \sum_k x_k^n P_X(x_k) \qquad \text{(III.9)}$$

The zeroth moment is simply the area under the PDF and hence must be 1 for any random variable.

The first moment is what we previously referred to as the mean, while the second moment is the mean squared value.

For some random variables, the second moment might be a more meaningful characterization than the first. For example, suppose X is a sample of a noise waveform. We might expect that the distribution of the noise is symmetric about zero (i.e., just as likely to be positive as negative) and hence the first moment will be zero. So if we are told that X has a zero mean, this merely says that the noise does not have a bias. On the other hand, the second moment of the random noise sample is in some sense a measure of the strength of the noise.

III.2.A. Mean Square Value

$$\begin{aligned}\overline{X^2} &= E\left[X^2\right] \\ &= \int_{-\infty}^{\infty} x^2 f_X(x)dx\end{aligned} \quad \text{mean of the square} \qquad \text{(III.10)}$$

III.2.B. Root Mean Square (RMS) Value

$$\begin{aligned}X_{\text{rms}} &= \sqrt{\mathrm{E}\left[X^2\right]} \\ &= \sqrt{\overline{X^2}}\end{aligned} \quad \text{root of the mean of the square} \qquad \text{(III.11)}$$

Example III.7

Consider a discrete random variable that has the values $\{1,2,4,7,11\}$, with respective probabilities $\{0.35,0.1,0.15,0.2,0.2\}$. The mean square value of this random variable is equal to

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= 0.35(1)^2 + 0.1(2)^2 + 0.15(4)^2 + 0.2(7)^2 + 0.2(11)^2 \\ &= 0.35 + 0.4 + 2.4 + 9.8 + 24.2 \\ &= 37.15\end{aligned}$$

The RMS value is equal to

$$X_{\text{rms}} = \sqrt{37.15} = 6.095$$

Example III.8

Consider a discrete binomial random variable with the PMF

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots$$

The first moment (mean) is calculated as follows:

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{kn!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{m=0}^{n-1} \frac{(n-1)!}{m!(n-1-m)!} p^m (1-p)^{n-1-m} \\
&= np \sum_{m=0}^{n-1} \binom{n-1}{m} p^m (1-p)^{n-1-m} \\
&= np
\end{aligned}$$

The second moment can be calculated as follows:

$$\mathrm{E}\left[X^2\right] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}$$

Note that we can use the identity $k^2 = k(k-1) + k$ to get

$$\begin{aligned}
\mathrm{E}\left[X^2\right] = &\sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\
&+ \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

The second sum is the mean, which has been calculated the above. The first sum can be calculated similarly to the calculation of the mean, resulting in

$$
\begin{aligned}
\sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} &= \sum_{k=2}^{n} \frac{k(k-1)n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{k(k-1)n!}{k(k-1)(k-2)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{m=0}^{n-2} \binom{n-2}{m} p^{m} (1-p)^{n-2-m} \\
&= n(n-1)p^2
\end{aligned}
$$

Adding the two results above produces

$$
\begin{aligned}
\mathrm{E}\left[X^2\right] &= np + n(n-1)p^2 \\
&= n^2 p^2 + np(1-p)
\end{aligned}
$$

Example III.9

Consider a discrete random variable that has the values $\{1,2,4,7,11\}$, with respective probabilities $\{0.35,0.1,0.15,0.2,0.2\}$. The 3rd moment of this random variable is equal to

$$
\begin{aligned}
\mathrm{E}\left[X^3\right] &= 0.35(1)^3 + 0.1(2)^3 + 0.15(4)^3 + 0.2(7)^3 + 0.2(11)^3 \\
&= 0.35 + 0.8 + 9.6 + 68.6 + 266.2 \\
&= 345.55
\end{aligned}
$$

III.3. Central Moments

Definition III-4

The nth central moment of a random variable X is defined as

$$E\left[(X-\mu_X)^n\right]=\int_{-\infty}^{\infty}(x-\mu_X)^n f_X(x)dx \qquad \text{(III.12)}$$

For discrete random variables, this definition reduces to

$$E\left[(X-\mu_X)^n\right]=\sum_k (x_k-\mu_X)^n P_X(x_k) \qquad \text{(III.13)}$$

III.3.A. Variance

$$\begin{aligned}\sigma_X^2 &= E\left[(X-\mu_X)^2\right]\\ &= E\left[X^2\right]-\mu_X^2\\ &= \overline{X^2}-\overline{X}^2\end{aligned} \qquad \text{(III.14)}$$

III.3.B. Standard Deviation

$$\sigma_X=\sqrt{E\left[(X-\mu_X)^2\right]} \qquad \text{(III.15)}$$

Example III.10

Consider a discrete random variable that has the values $\{1,2,4,7,11\}$, with respective probabilities $\{0.35,0.1,0.15,0.2,0.2\}$. The mean square value of this random variable is equal to

$$\begin{aligned}\mathrm{E}\left[X^2\right]&=0.35(1)^2+0.1(2)^2+0.15(4)^2+0.2(7)^2+0.2(11)^2\\ &=0.35+0.4+2.4+9.8+24.2\\ &=37.15\end{aligned}$$

Using the mean from Example III.1, the variance is equal to

$$\begin{aligned}\sigma_X^2&=37.15-(4.75)^2\\ &=14.5875\end{aligned}$$

The standard deviation is equal to

$$\begin{aligned}\sigma_X&=\sqrt{14.5875}\\ &\approx 3.819\end{aligned}$$

III.4. Conditional Expected Values

Definition III-5

The expected value of a random variable X conditioned on some event A is

$$\mathrm{E}\left[X \mid A\right] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx \tag{III.16}$$

For a discrete random variable, this definition reduces to

$$\mathrm{E}\left[X \mid A\right] = \sum_k x_k P_{X|A}(x_k) \tag{III.17}$$

Similarly, the conditional expectation of a function $g(X)$ of a random variable X conditioned on the event A is

$$\mathrm{E}\left[g(X) \mid A\right] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)dx \tag{III.18}$$

For a discrete random variable, this becomes

$$\mathrm{E}\left[g(X) \mid A\right] = \sum_k g(x_k) P_{X|A}(x_k) \tag{III.19}$$

Example III.11

Consider a standard Gaussian random variable X. Let $A = \{X > 0\}$.

$$f_{X|A}(x) = \frac{f_X(x)}{\Pr\{X > 0\}} u(x)$$
$$= \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} u(x)$$

Conditioned on A, the expected value is equal to

$$\mathrm{E}\left[X \mid A\right] = \mathrm{E}\left[X \mid X > 0\right]$$
$$= \sqrt{\frac{2}{\pi}} \int_0^{\infty} x e^{-\frac{x^2}{2}} dx$$
$$= \sqrt{\frac{2}{\pi}}$$

III.5. Transformations of Random Variables

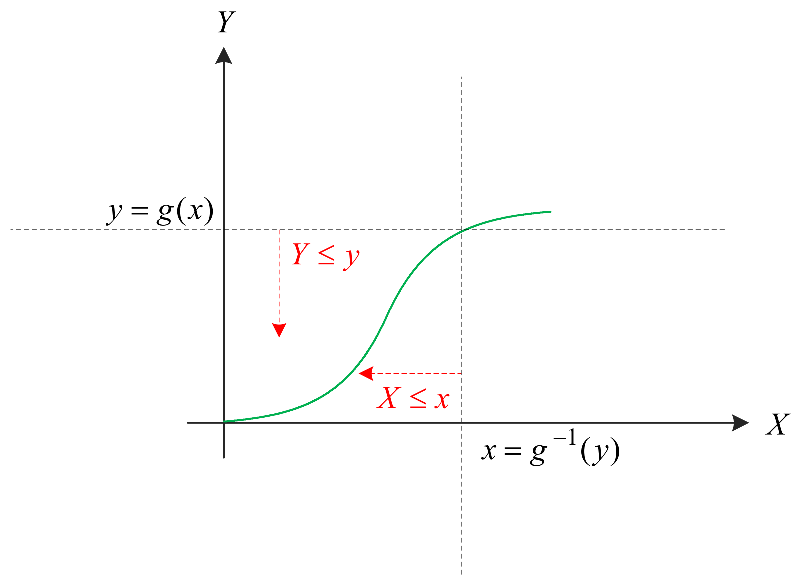
III.5.A. Monotonically Increasing Functions

Assume that Y is a continuous, one-to-one, and monotonically increasing function of X.

$$Y = g(X)$$

$$X = g^{-1}(Y)$$

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(g(X) \le y) \\ &= \Pr\left(X \le g^{-1}(y)\right) \\ &= F_X\left(g^{-1}(y)\right) \end{aligned}$$



Then,

$$F_Y(y) = F_X\left(g^{-1}(y)\right)$$

Note that

$$F_X(x) = F_Y\left(g(x)\right)$$

Differentiating with respect to y produces

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\frac{d}{dy}g^{-1}(y)$$
$$= f_X\left(x\right)\frac{dx}{dy}\bigg|_{x=g^{-1}(y)}$$
$$= \frac{f_X\left(x\right)}{\frac{dy}{dx}}\Bigg|_{x=g^{-1}(y)}$$

Example III.12

Consider a Gaussian random variable X with mean μ and variance σ^2. A new random variable is formed as $Y = aX + b$, where $a > 0$ (so that the transformation is monotonically increasing).

$$\frac{dy}{dx} = a$$

$$x = \frac{y-b}{a}$$

$$f_Y(y) = \frac{f_X\left(\frac{y-b}{a}\right)}{a}$$

Substituting $x = \frac{y-b}{a}$, we get

$$f_Y(y) = \frac{1}{a\sqrt{2\pi\sigma^2}}e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}$$
$$= \frac{1}{\sqrt{2\pi a^2\sigma^2}}e^{-\frac{\left(y-(a\mu+b)\right)^2}{2a^2\sigma^2}}$$

Note that Y is Gaussian with mean $a\mu + b$ and variance $a^2\sigma^2$.

Example III.13

Let X be an exponential random variable with $f_X(x) = 2e^{-2x}u(x)$. Let $Y = X^3$. Determine $f_Y(y)$.

$$\frac{dy}{dx} = 3x^2$$

$$x = y^{\frac{1}{3}}$$

Solution:

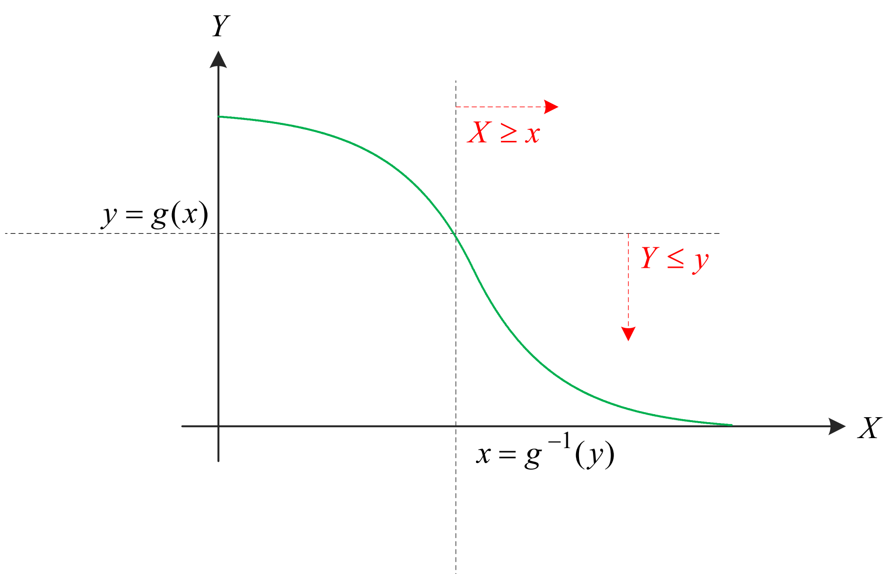
$$f_Y(y) = \left.\frac{f_X(x)}{3x^2}\right|_{x=y^{\frac{1}{3}}}$$

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

Note that Y is not an exponential random variable.

III.5.B. MONOTONICALLY DECREASING FUNCTIONS

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(g(X) \le y) \\ &= \Pr\left(X \ge g^{-1}(y)\right) \\ &= 1 - F_X\left(g^{-1}(y)\right) \end{aligned}$$



Differentiating with respect to y produces

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

$$= -f_X(x)\frac{dx}{dy}\bigg|_{x=g^{-1}(y)}$$

$$= -\frac{f_X(x)}{\frac{dy}{dx}}\Bigg|_{x=g^{-1}(y)}$$

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|\bigg|_{x=g^{-1}(y)}$$

$$= \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}\Bigg|_{x=g^{-1}(y)}$$

III.5.C. NON-MONOTONIC FUNCTIONS

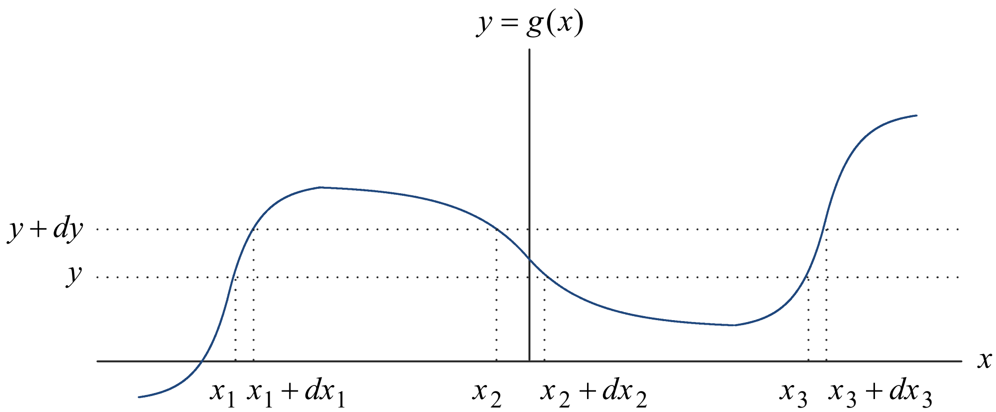


Figure III.1: Non-monotonic function

In this case, we cannot associate the event $\{Y \le y\}$ with events of the form $\{X \le g^{-1}(y)\}$ or $\{X \ge g^{-1}(y)\}$. To avoid this problem, we calculate the PDF of Y directly, rather than first finding the CDF.

Note that

$$\Pr(y \le Y < y + dy) = f_Y(y)dy$$

$$\{y \le Y < y+dy\} = \left[\bigcup_{i:x_i \in X^+} \{x_i \le X < x_i + dx_i\}\right] \cup \left[\bigcup_{i:x_i \in X^-} \{x_i + dx_i < X \le x_i\}\right]$$

Since each of the events on the right-hand side is mutually exclusive, the probability of the union is simply the sum of the probabilities, so that

$$f_Y(y) = \sum_{x_i \in X^+} f_X(x_i)dx_i + \sum_{x_i \in X^-} f_X(x_i)(-dx_i)$$

$$f_Y(y) = \sum_{x_i} f_X(x_i)\left|\frac{dx}{dy}\right|\Bigg|_{x_i = g^{-1}(y)}$$

$$= \sum_{x_i} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|\Big|_{x_i = g^{-1}(y)}}$$

Example III.14

Suppose X is a Gaussian random variable with zero mean and variance σ^2. Let $Y = X^2$. For any positive value of y, $y = x^2$ has two real roots, namely, $\pm\sqrt{y}$. For negative values of y, there are no real roots. Using the last result above,

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

$$= \frac{f_X\left(+\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)}{2\sqrt{y}}u(y)$$

For a zero-mean Gaussian PDF, $f_X(x)$ is an even function so that $f_X\left(+\sqrt{y}\right) = f_X\left(-\sqrt{y}\right)$. Therefore,

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X\left(\sqrt{y}\right)u(y)$$

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

Hence, Y is a gamma random variable.

Example III.15

Suppose X is an exponential random variable with a PDF $f_X(x) = e^{-x}u(x)$. Let

$$\begin{aligned} Y &= g(X) \\ &= \text{floor}(X) \\ &= k, \quad k \le X < k+1 \end{aligned}$$

The PMF of Y for $k = 0, 1, \cdots$ is

$$\begin{aligned} \mathrm{P}_Y(k) &= \Pr(k \le X < k+1) \\ &= \int_k^{k+1} e^{-x} dx \\ &= e^{-k} - e^{-(k+1)} \end{aligned}$$

III.6. Characteristic Functions

Let

$$g(X) = e^{j\omega X}$$

The characteristic function of a random variable X is given by

$$\begin{aligned} \Phi_X(\omega) &= \mathrm{E}\left[g(X)\right] \\ &= \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx \end{aligned} \tag{III.20}$$

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-j\omega x} d\omega \tag{III.21}$$

Note that

$$\mathcal{F}\left\{f_X(x)\right\} = \Phi_X(-\omega) \tag{III.22}$$

We can get the PDF of a random variable from its characteristic function through an inverse Fourier transform operation.

Example III.16

Suppose X is an exponential random variable with a PDF $f_X(x) = e^{-x}u(x)$. The characteristic function of X is found to be

$$\Phi_X(\omega) = \int_0^{\infty} e^{j\omega x} e^{-x} dx$$
$$= \frac{1}{1 - j\omega}$$

This result assumes that ω is a real quantity. $f_X(x) = e^{-x} u(x)$

Let $f_Y(y) = ae^{-ay} u(y)$. Note that a must be positive. Note also that $f_Y(y) = af_X(ay)$. Using the scaling property of the Fourier transform, the characteristic function of Y is given by

$$\Phi_Y(\omega) = a \frac{1}{|a|} \Phi_X\left(\frac{\omega}{a}\right)$$
$$= \frac{a}{a - j\omega}$$

Let $f_Z(z) = ae^{-a(z-b)} u(z-b)$. Note that $f_Z(z) = f_Y(z-b)$. Using the shifting property of the Fourier transform, the characteristic function of Z is given by

$$\Phi_Z(\omega) = \Phi_Y(\omega) e^{-j\omega b}$$
$$= \frac{ae^{-j\omega b}}{a - j\omega}$$

Example III.17

Suppose X is a binomial random variable with a PDF $f_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k)$. The characteristic function of X is found to be

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} e^{j\omega x} \left(\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k) \right) dx$$

Interchanging the orders of the summation and integration operators, we get

$$\Phi_X(\omega) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \int_{-\infty}^{\infty} \delta(x-k) e^{j\omega x} dx$$

Using the sifting property of the delta function,

$$\Phi_X(\omega) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{j\omega k}$$

Combining the two terms that are raised to power k inside the summation operator,

$$\Phi_X(\omega) = \sum_{k=0}^{n} \binom{n}{k} \left(pe^{j\omega}\right)^k (1-p)^{n-k}$$

$$= \left(1 - p + pe^{j\omega}\right)^n$$

Example III.18

Suppose X is a standard gaussian random variable. The characteristic function of X can be found as follows:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{j\omega x} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 - 2j\omega x}{2}} dx$$

We complete the square in the exponent to get

$$\Phi_X(\omega) = e^{-\frac{\omega^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - j\omega)^2}{2}} dx$$

The integrand in the last expression above looks like the PDF of a unit variance Gaussian random variable with a mean of $j\omega$, and since the integral is over all values of x, the integration must be unity. However, since x is a real random variable, it cannot have a complex mean, and the above argument is mathematically wrong.

Nevertheless, with some mathematical manipulations, the integral above can be shown to produce an answer of unity. The resulting characteristic function is

$$\Phi_X(\omega) = e^{-\frac{\omega^2}{2}}$$

It can be shown that for a Gaussian random variable with a mean of μ and a variance of σ^2, the characteristic function is

$$\Phi_X(\omega) = e^{j\omega\mu - \frac{\omega^2\sigma^2}{2}}$$

$$\frac{d}{d\omega}\Phi_X(\omega) = \frac{d}{d\omega}\left(\int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx\right)$$

$$= \int_{-\infty}^{\infty} \frac{d}{d\omega}\left(e^{j\omega x}\right) f_X(x)dx$$

$$= \int_{-\infty}^{\infty} jxe^{j\omega x} f_X(x)dx$$

$$-j\frac{d}{d\omega}\Phi_X(\omega) = \int_{-\infty}^{\infty} xe^{j\omega x} f_X(x)dx$$

$$-j\frac{d}{d\omega}\Phi_X(\omega)\bigg|_{\omega=0} = \int_{-\infty}^{\infty} xf_X(x)dx$$

$$= \mathrm{E}\left[X\right]$$

For any random variable whose characteristic function is differentiable at $\omega = 0$,

$$\mathrm{E}\left[X\right] = -j\frac{d}{d\omega}\Phi_X(\omega)\bigg|_{\omega=0} \tag{III.23}$$

$$\mathrm{E}\left[X^k\right] = (-j)^k \frac{d^k}{d\omega^k}\Phi_X(\omega)\bigg|_{\omega=0} \tag{III.24}$$

Example III.19

Suppose Y is an exponential random variable with the PDF $f_Y(y) = ae^{-ay}u(y)$. The characteristic function of Y is (see Example III.16):

$$\Phi_Y(\omega) = \frac{a}{a - j\omega}$$

The 1st derivative of $\Phi_Y(\omega)$ is

$$\frac{d}{d\omega}\Phi_Y(\omega) = \frac{ja}{(a - j\omega)^2}$$

Thus, the 1st moment is

$$\begin{aligned} \mathrm{E}\left[Y\right] &= -j\frac{d}{d\omega}\Phi_Y(\omega\Big|_{\omega=0} \\ &= \frac{a}{(a-j\omega)^2}\Big|_{\omega=0} \\ &= \frac{1}{a} \end{aligned}$$

Note that the kth derivative of $\Phi_Y(\omega)$ is

$$\frac{d^k}{d\omega^k}\Phi_Y(\omega) = \frac{j^k k!a}{(a-j\omega)^{k+1}}$$

Thus, the kth moment is

$$\begin{aligned} \mathrm{E}\left[Y^k\right] &= (-j)^k\frac{d^k}{d\omega^k}\Phi_Y(\omega\Big|_{\omega=0} \\ &= \frac{k!a}{(a-j\omega)^{k+1}}\Big|_{\omega=0} \\ &= \frac{k!}{a^k} \end{aligned}$$

Specifically, suppose the characteristic function is expanded in the form

$$\Phi_X(\omega) = \sum_{k=0}^{\infty}\phi_k\omega^k \qquad \text{(III.25)}$$

Then,

$$\mathrm{E}\left[X^k\right] = (-j)^k k!\phi_k \qquad \text{(III.26)}$$

Example III.20

Consider a zero-mean Gaussian random variable X with variance σ^2. The characteristic function of X is (see Example III.18):

$$\Phi_X(\omega) = e^{-\frac{\sigma^2\omega^2}{2}}$$

Using Taylor series expansion,

$$\Phi_X(\omega) = \sum_{n=0}^{\infty} \frac{\left(-\frac{\sigma^2\omega^2}{2}\right)^n}{n!}$$
$$= \sum_{n=0}^{\infty} \frac{(-1)^n \sigma^{2n}}{2^n n!} \omega^{2n}$$

The coefficients of the power series expansion in (III.25) are given by

$$\phi_k = \begin{cases} \dfrac{j^k \left(\frac{\sigma}{\sqrt{2}}\right)^k}{\left(\frac{k}{2}\right)!}, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

Using (III.26),

$$\mathrm{E}\left[X^k\right] = \begin{cases} \dfrac{k!}{\left(\frac{k}{2}\right)!}\left(\frac{\sigma}{\sqrt{2}}\right)^k, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

III.7. Moment Generating Functions (MGF)

The moment generating function $M_X(u)$ of a nonnegative random variable X is

$$g(X) = e^{uX}$$

$$M_X(u) = \mathrm{E}\left[e^{uX}\right] = \int_0^{\infty} f_X(x) e^{ux} dx \qquad \text{(III.27)}$$

$$\mathrm{E}\left[X^k\right] = \frac{d^k}{du^k} M_X(u)\bigg|_{u=0} \qquad \text{(III.28)}$$

$$\begin{aligned} M_X(u) &= \sum_{k=0}^{\infty} \frac{\mathrm{E}\left[X^k\right]}{k!} u^k \\ &= \sum_{k=0}^{\infty} m_k u^k \end{aligned} \tag{III.29}$$

Example III.21

Consider an Erlang random variable with a PDF of the form

$$f_X(x) = \frac{x^{n-1} e^{-x}}{(n-1)!} u(x)$$

$$M_X(u) = \frac{1}{(1-u)^n}$$

The first two moments are found as follows:

$$\begin{aligned} \mathrm{E}\left[X\right] &= \frac{d}{du} \frac{1}{(1-u)^n}\bigg|_{u=0} \\ &= \frac{n}{(1-u)^{n+1}}\bigg|_{u=0} \\ &= n \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[X^2\right] &= \frac{d^2}{du^2} \frac{1}{(1-u)^n}\bigg|_{u=0} \\ &= \frac{n(n+1)}{(1-u)^{n+2}}\bigg|_{u=0} \\ &= n(n+1) \end{aligned}$$

Using the first two moments, the variance can be found to be equal to

$$\begin{aligned} \sigma_X^2 &= \mathrm{E}\left[X^2\right] - \left(\mathrm{E}\left[X\right]\right)^2 \\ &= \overline{X^2} - \overline{X}^2 \\ &= n(n+1) - n^2 \\ &= n \end{aligned}$$

The k th moments is found to be equal to

$$\begin{aligned} \mathrm{E}\left[X^k\right] &= \frac{d^k}{du^k}\frac{1}{(1-u)^n}\bigg|_{u=0} \\ &= n(n+1)\cdots(n+k-1) \\ &= \frac{(n+k-1)!}{(n-1)!} \end{aligned}$$

Theorem III.2

Markov Inequality

Suppose that X is a nonnegative random variable, then

$$\Pr\left(X \geq x_0\right) \leq \frac{\mathrm{E}[X]}{x_0}$$

Proof:

For nonnegative random variables, the expected value is

$$\begin{aligned} \mathrm{E}[X] &= \int_0^\infty x f_X(x) dx \\ &= \int_0^{x_0} x f_X(x) dx + \int_{x_0}^\infty x f_X(x) dx \\ &\geq \int_{x_0}^\infty x f_X(x) dx \\ &\geq x_0 \int_{x_0}^\infty f_X(x) dx \\ &= x_0 \Pr\left(X \geq x_0\right) \end{aligned}$$

Example III.22

Suppose that the average life span of a person was 75 years. The probability of a person living to be 110 years old would then be bounded by

$$\begin{aligned} \Pr\left(X \geq 110\right) &\leq \frac{75}{110} \\ &= 0.6818 \end{aligned}$$

Of course, we know that in fact very few people live to be 110 years old, and hence, this bound is almost useless to us.

Theorem III.3

Chebyshev's Inequality

Let X be a random with mean μ_X and variance σ_X^2. The probability that X takes on a value that is removed from the mean by more than x_0 is given by

$$\Pr\left(|X-\mu_X| \geq x_0\right) \leq \frac{\sigma_X^2}{x_0^2}$$

Proof:

Chebyshev's inequality is a direct result of Markov's inequality. Note that the event $\left\{|X-\mu_X| \geq x_0\right\}$ is equivalent to the event $\left\{(X-\mu_X)^2 \geq x_0^2\right\}$. Applying Markov's inequality to the latter event results in

$$\Pr\left((X-\mu_X)^2 \geq x_0^2\right) \leq \frac{\mathrm{E}\left[(X-\mu_X)^2\right]}{x_0^2}$$

$$= \frac{\sigma_X^2}{x_0^2}$$

Note that Chebyshev's inequality gives a bound on the two-sided tail probability, whereas Markov's inequality applies to the one-sided tail probability. Also, Chebyshev's inequality can be applied to any random variable, and not only to non-negative random variables.

EXAMPLE 4.27: Continuing the previous example, suppose that in addition to a mean of 75 years, the human life span had a standard deviation of 5 years. In this case,

$$\Pr(X \geq 110) \leq \Pr(X \geq 110) + \Pr(X \leq 40) = \Pr(|X-75| \geq 35).$$

Now the Chebyshev inequality can be applied to give

$$\Pr(|X-75| \geq 35) \leq \left(\frac{5}{35}\right)^2 = \frac{1}{49}.$$

While this result may still be quite loose, by using the extra piece of information provided by the variance, a better bound is obtained.

***

III.7-Moment Generating Functions (MGF)

III.7-Moment Generating Functions (MGF)

IV. Pairs of Random Variables

IV.1. Joint Cumulative Distribution Functions

Definition IV-1

The joint cumulative distribution function of a pair of random variables X and Y is

$$F_{X,Y}(x,y) = \Pr(X \le x, Y \le y) \tag{IV.1}$$

That is, the joint CDF is the joint probability of the two events $\{X \le x\}$ and $\{Y \le y\}$.

IV.1.A. Joint CDF Properties

$$0 \le F_{X,Y}(x,y) \le 1 \tag{IV.2}$$

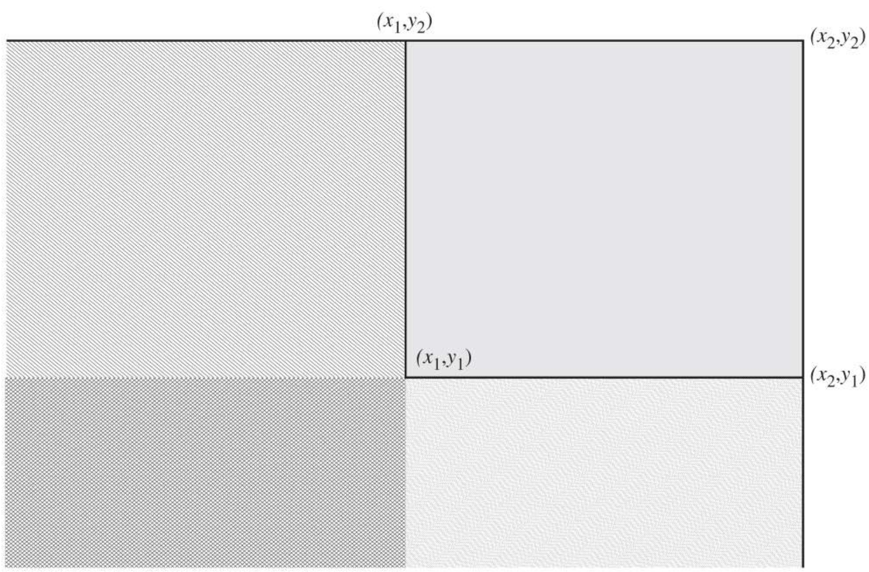
$$F_{X,Y}(-\infty,y) = F_{X,Y}(x,-\infty) = F_{X,Y}(-\infty,-\infty) = 0 \tag{IV.3}$$

$$F_{X,Y}(\infty,\infty) = 1 \tag{IV.4}$$

$$\begin{aligned} F_{X,Y}(\infty,y) &= F_Y(y) \\ F_{X,Y}(x,\infty) &= F_X(x) \end{aligned} \tag{IV.5}$$

For $x_1 \le x_2$ and $y_1 \le y_2$, $\{X \le x_1\} \cap \{Y \le y_1\}$ is a subset of $\{X \le x_2\} \cap \{Y \le y_2\}$ so that $F_{X,Y}(x_1,y_1) \le F_{X,Y}(x_2,y_2)$.

$F_X(x)$ and $F_Y(y)$ are referred to as the marginal CDFs of X and Y, respectively.



IV.1-Joint Cumulative Distribution Functions

$$F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$$

$$\begin{aligned} &\Pr(x_1 < X \le x_2, y_1 < Y \le y_2) \\ &\quad = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1) \\ &\quad \ge 0 \end{aligned} \quad \text{(IV.6)}$$

EXAMPLE 5.1: One of the simplest examples (conceptually) of a pair of random variables is one that is uniformly distributed over the unit square (i.e., $0 < x < 1, 0 < y < 1$). The CDF of such a random variable is

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ x & 0 \le x \le 1,\ y > 1 \\ y & x > 1,\ 0 \le y \le 1 \\ xy & 0 \le x \le 1,\ 0 \le y \le 1 \\ 1 & x > 1,\ y > 1 \end{cases}$$

Even this very simple example leads to a rather cumbersome function. Nevertheless, it is straightforward to verify that this function does indeed satisfy all the properties of a joint CDF. From this joint CDF, the marginal CDF of X can be found to be

$$F_X(x) = F_{X,Y}(x,\infty) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Hence, the marginal CDF of X is also a uniform distribution. The same statement holds for Y as well.

IV.2. Joint Probability Density Functions

Definition IV-2

The joint probability density function of a pair of random variables X and Y evaluated at the point (x, y) is given by

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{Y,Y}(x,y) \quad \text{(IV.7)}$$

Based on (IV.7),

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(\alpha,\beta)\, d\alpha\, d\beta \quad \text{(IV.8)}$$

EXAMPLE 5.2: From the joint CDF given in Example 5.1, it is easily found (by differentiating the joint CDF with respect to both x and y) that the joint PDF for a pair of random variables uniformly distributed over the unit square is

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Note how much simpler the joint PDF is to specify than is the joint CDF.

IV.2.A. JOINT PDF PROPERTIES

$$f_{X,Y}(x,y) \geq 0 \quad \text{(IV.9)}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1 \quad \text{(IV.10)}$$

$$\begin{aligned} F_X(x) &= F_{X,Y}(x,\infty) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy \end{aligned} \quad \text{(IV.11)}$$

$$\begin{aligned} f_X(x) &= \frac{d}{dx}F_X(x)f_Y(y) \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \end{aligned}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \quad \text{(IV.12)}$$

$$\Pr(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)dxdy \quad \text{(IV.13)}$$

$$\Pr\left((X,Y) \in A\right) = \iint_A f_{X,Y}(x,y)dxdy \quad \text{(IV.14)}$$

Example IV.1

Suppose X and Y are jointly uniformly distributed over the unit circle ($r=1$). That is, $f_{X,Y}(x,y)$ is equal to a constant c over all the points (x,y) that satisfy $x^2+y^2\le 1$:

$$f_{X,Y}(x,y)=\begin{cases} c, & x^2+y^2\le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$c\pi r^2=1$$

The constant c can be found as follows:

$$\iint\limits_{x^2+y^2=1} c\,dx\,dy=1$$

$$c=\frac{1}{\pi}$$

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

$$=\frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy$$

$$=\frac{2}{\pi}\sqrt{1-x^2},\quad |x|\le 1$$

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

Example IV.2

Let

$$f_{X,Y}(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$$

The probability that the point (x,y) falls inside the unit circle is calculated as follows:

$$\Pr(X^2+Y^2\le 1)=\frac{1}{2\pi}\iint_{x^2+y^2\le 1} e^{-\frac{x^2+y^2}{2}}\,dxdy$$

Converting the double integral to polar coordinates,

$$\Pr(X^2+Y^2\le 1)=\frac{1}{2\pi}\int_0^{2\pi}\int_0^1 re^{-\frac{r^2}{2}}\,drd\theta$$

$$=\int_0^1 re^{-\frac{r^2}{2}}\,dr$$

$$=1-e^{-\frac{1}{2}}$$

EXAMPLE 5.5: Now suppose that a pair of random variables has the joint PDF given by

$$f_{X,Y}(x,y)=c\exp\left(-x-\frac{y}{2}\right)u(x)u(y).$$

First, the constant c is found using the normalization integral

$$\int_0^\infty\int_0^\infty c\exp\left(-x-\frac{y}{2}\right)dxdy=1\Rightarrow c=\frac{1}{2}.$$

Next, suppose we wish to determine the probability of the event $\{X>Y\}$. This can be viewed as finding the probability of the pair (X,Y) falling in the region A that is now defined as $A=\{(x,y):x>y\}$. This probability is calculated as

$$\Pr(X>Y)=\iint_{x>y} f_{X,Y}(x,y)\,dxdy=\int_0^\infty\int_y^\infty \frac{1}{2}\exp\left(-x-\frac{y}{2}\right)dxdy$$

$$=\int_0^\infty \frac{1}{2}\exp\left(-\frac{3y}{2}\right)dy=\frac{1}{3}.$$

$$f_{X,Y}(x,y)=\frac{1}{2}e^{-\left(x+\frac{y}{2}\right)}u(x)u(y)$$

$$=e^{-x}u(x)\frac{1}{2}e^{-\frac{y}{2}}u(y)$$

Definition IV-3

The joint probability mass function for a pair of discrete random variables X and Y is given by $P_{X,Y}(x_m, y_n) = \Pr\left(\{X = x_m\} \cap \{Y = y_n\}\right)$.

IV.2.B. PMF Properties

$$0 \le P_{X,Y}(x_m, y_n) \le 1 \tag{IV.15}$$

$$\sum_{m=1}^{M} \sum_{n=1}^{N} P_{X,Y}(x_m, y_n) = 1 \tag{IV.16}$$

$$\sum_{n=1}^{N} P_{X,Y}(x_m, y_n) = P_X(x_m) \tag{IV.17}$$

$$\sum_{m=1}^{M} P_{X,Y}(x_m, y_n) = P_Y(y_n) \tag{IV.18}$$

$$\Pr\left((X,Y) \in A\right) = \sum_{(x,y)\in A} P_{X,Y}(x, y) \tag{IV.19}$$

Example

Let X be the number on the upper face of a fair die after throwing it. Let Y be the number on the upper face of another fair die after throwing it. Let event $A = \{X < 3, Y \text{ is even}\} = \{(1,2),(2,2),(1,4),(2,4),(1,6),(2,6)\}$.

$\Pr\left((X,Y) \in A\right) = \frac{6}{36} = \frac{1}{6}$.

$$f_{X,Y}(x, y) = \sum_{m=1}^{M} \sum_{n=1}^{N} P_{X,Y}(x_m, y_n)\delta(x - x_m)\delta(y - y_n) \tag{IV.20}$$

$$F_{X,Y}(x, y) = \sum_{m=1}^{M} \sum_{n=1}^{N} P_{X,Y}(x_m, y_n)u(x - x_m)u(y - y_n) \tag{IV.21}$$

EXAMPLE 5.7: A pair of discrete random variables N and M have a joint PMF given by

$$P_{N,M}(n,m) = \frac{(n+m)!}{n!m!}\frac{a^n b^m}{(a+b+1)^{n+m+1}}, \quad m=0,1,2,3,\ldots, \quad n=0,1,2,3,\ldots.$$

The marginal PMF of N can be found by summing over m in the joint PMF:

$$P_N(n) = \sum_{m=0}^{\infty} P_{N,M}(n,m) = \sum_{m=0}^{\infty} \frac{(n+m)!}{n!m!}\frac{a^n b^m}{(a+b+1)^{n+m+1}}.$$

To evaluate this series, the following identity is used:

$$\sum_{m=0}^{\infty} \frac{(n+m)!}{n!m!} x^m = \left(\frac{1}{1-x}\right)^{n+1}.$$

The marginal PMF then reduces to

$$P_N(n) = \frac{a^n}{(a+b+1)^{n+1}} \sum_{m=0}^{\infty} \frac{(n+m)!}{n!m!}\frac{b^m}{(a+b+1)^m}$$

$$= \frac{a^n}{(a+b+1)^{n+1}} \left(\frac{1}{1-\dfrac{b}{a+b+1}}\right)^{n+1} = \frac{a^n}{(1+a)^{n+1}}.$$

Likewise, by symmetry, the marginal PMF of M is

$$P_M(m) = \frac{b^m}{(1+b)^{m+1}}.$$

Hence, the random variables M and N both follow a geometric distribution.

IV.3. Conditional CDFs, PMFs and PDFs

IV.3.A. Discrete Random Variables

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$
$$P_{X|Y}(x \mid y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} \qquad \text{(IV.22)}$$

EXAMPLE 5.8: Using the joint PMF given in Example 5.7, along with the marginal PMF found in that example, it is found that

$$P_{N|M}(n|m) = \frac{P_{M,N}(m,n)}{P_M(m)} = \frac{(n+m)!}{n!m!}\frac{a^n b^m}{(a+b+1)^{n+m+1}}\frac{(1+b)^{m+1}}{b^m}$$
$$= \frac{(n+m)!}{n!m!}\frac{a^n(1+b)^{m+1}}{(a+b+1)^{n+m+1}}.$$

Note that the conditional PMF of N given M is quite different than the marginal PMF of N. That is, knowing M changes the distribution of N.

IV.3.B. CONTINUOUS RANDOM VARIABLES

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{(IV.23)}$$

EXAMPLE 5.9: A certain pair of random variables has a joint PDF given by

$$f_{X,Y}(x,y) = \frac{2abc}{(ax+by+c)^3}u(x)u(y)$$

for some positive constants a, b, and c. The marginal PDFs are easily found to be

$$f_X(x) = \int_0^\infty f_{X,Y}(x,y)\,dy = \frac{ac}{(ax+c)^2}\,u(x)$$

and

$$f_Y(y) = \int_0^\infty f_{X,Y}(x,y)\,dx = \frac{bc}{(by+c)^2}\,u(y).$$

The conditional PDF of X given Y then works out to be

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2a(by+c)^2}{(ax+by+c)^3}\,u(x).$$

The conditional PDF of Y given X could also be determined in a similar way:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2b(ax+c)^2}{(ax+by+c)^3}\,u(y).$$

IV.4. Expected Values Involving Joint Random Variables

$$\mathrm{E}\left[g(X,Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy \quad \text{(IV.24)}$$

For discrete random variables,

$$E\left[g(X,Y)\right] = \sum_m \sum_n g(x_m,y_n)P_{X,Y}(x_m,y_n) \quad \text{(IV.25)}$$

Definition IV-4

The correlation between two random variables is defined as

$$\begin{aligned} R_{X,Y} &= \mathrm{E}\left[XY\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy \end{aligned} \tag{IV.26}$$

Two random variables that have a correlation of zero are said to be orthogonal.

Definition IV-5

The covariance between two random variables is

$$\begin{aligned} Cov(X,Y) &= E\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= R_{X,Y} - \mu_X\mu_Y \end{aligned} \tag{IV.27}$$

Definition IV-6

The correlation coefficient of two random variables is defined as

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} \tag{IV.28}$$

$$\left|\rho_{XY}\right| \le 1 \tag{IV.29}$$

$$\mathrm{E}\left[g(X)\,|\,Y\right] = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x\,|\,y)dx \tag{IV.30}$$

Definition IV-7

The joint characteristic function is defined as

$$\Phi_{X,Y}(\omega_1,\omega_2) = \mathrm{E}\left[e^{j(\omega_1 X+\omega_2 Y)}\right] \tag{IV.31}$$

$$E\left[X^m Y^n\right] = (-j)^{m+n}\frac{\partial^m}{\partial\omega_1^m}\frac{\partial^n}{\partial\omega_2^n}\Phi_{X,Y}(\omega_1,\omega_2)\Big|_{\omega_1=\omega_2=0} \tag{IV.32}$$

Definition IV-8

The joint MGF is defined as

$$M_{X,Y}(u_1,u_2) = \mathrm{E}\left[e^{u_1 X+u_2 Y}\right] \tag{IV.33}$$

$$E\left[X^m Y^n\right] = \frac{\partial^m}{\partial u_1^m}\frac{\partial^n}{\partial u_2^n} M_{X,Y}(u_1,u_2)\Big|_{u_1=u_2=0} \tag{IV.34}$$

If $\mathrm{E}\left[XY\right] = \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]$, then X and Y are uncorrelated.

If $\mathrm{E}\left[XY\right] = 0$, then X and Y are orthogonal.

IV.5. Independent Random Variables

$$\begin{aligned} F_{X,Y}(x,y) &= \Pr(X \le x, Y \le y) \\ &= \Pr(X \le x)\Pr(Y \le y) \\ &= F_X(x)F_Y(y) \end{aligned} \tag{IV.35}$$

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \tag{IV.36}$$

Note that when X and Y are independent then

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= f_X(x) \end{aligned} \tag{IV.37}$$

and

$$\mathrm{E}\left[XY\right] = \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right] \tag{IV.38}$$

$$\begin{aligned} R_{X,Y} &= \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right] \\ &= \overline{X}\overline{Y} \\ &= \mu_X \mu_Y \end{aligned} \tag{IV.39}$$

$$\begin{aligned} Cov(X,Y) &= R_{X,Y} - \mu_X \mu_Y \\ &= 0 \end{aligned} \tag{IV.40}$$

EXAMPLE 5.13: Returning once again to the joint PDF of Example 5.10, we saw in that example that the marginal PDF of X is

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

while the conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \sqrt{\frac{2}{3\pi}}\exp\left(-\frac{2}{3}\left(x-\frac{y}{2}\right)^2\right).$$

Clearly, these two random variables are not independent.

EXAMPLE 5.14: Suppose the random variables X and Y are uniformly distributed on the square defined by $0 \le x, y \le 1$. That is

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

The marginal PDFs of X and Y work out to be

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

These random variables are statistically independent since $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Theorem IV.1

Let X and Y be two independent random variables and consider forming two new random variables $U_1 = g_1(X)$ and $U_2 = g_2(Y)$. These new random variables U_1 and U_2 are also independent.

IV.6. Transformations of Pairs of Random Variables

IV.6.A. PDF of the Sum of Two Independent Random Variables

Let $Z = X + Y$, then

$$\begin{aligned}\Phi_Z(\omega) &= \mathrm{E}\left[e^{j\omega Z}\right] \\ &= \mathrm{E}\left[e^{j\omega(X+Y)}\right] \\ &= \mathrm{E}\left[e^{j\omega X}\right]\mathrm{E}\left[e^{j\omega Y}\right] \\ &= \Phi_X(\omega)\Phi_Y(\omega)\end{aligned} \qquad \text{(IV.41)}$$

Then

$$\begin{aligned} f_Z(z) &= f_X(z) * f_Y(z) \\ &= \int_{-\infty}^{\infty} f_X(z-\xi) f_Y(\xi) d\xi \end{aligned} \tag{IV.42}$$

EXAMPLE 5.20: Suppose X and Y are independent and both have exponential distributions,

$$f_X(x) = a\exp(-ax)u(x), \qquad f_Y(y) = b\exp(-by)u(y).$$

The PDF of $Z = X + Y$ is then found by performing the necessary convolution:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy = ab\int_{-\infty}^{\infty} \exp(-a(z-y))\exp(-by)u(z-y)u(y)dy$$

$$= abe^{-az}\int_0^z \exp((a-b)y)\,dy\,u(z)$$

$$= \frac{ab}{a-b}\left[e^{-az}e^{(a-b)y}\Big|_{y=0}^{y=z} \right] u(z) = \frac{ab}{a-b}\left[e^{-by} - e^{-az}\right]u(z).$$

This result is valid assuming that $a \neq b$. If $a = b$, then the convolution works out to be

$$f_Z(z) = a^2 z e^{-az} u(z).$$

IV.6.B. PDF of Functions of Two Independent Random Variables

Let $Z = g(X,Y)$, then

$$\begin{aligned} \Phi_Z(\omega) &= \mathrm{E}\left[e^{j\omega g(X,Y)} \right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j\omega g(x,y)} f_{X,Y}(x,y) dxdy \end{aligned} \tag{IV.43}$$

Then

$$\begin{aligned} f_Z(z) &= \mathcal{F}^{-1}\{\Phi_Z(\omega)\} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_Z(\omega) e^{-j\omega z} d\omega \end{aligned} \tag{IV.44}$$

EXAMPLE 5.21: Suppose X and Y are independent, zero-mean, unit variance Gaussian random variables. The PDF of $Z = X^2 + Y^2$ can be found using either of the methods described thus far. Using characteristic functions,

$$\Phi_Z(\omega) = E\left[e^{j\omega(X^2+Y^2)}\right] = E\left[e^{j\omega X^2}\right]E\left[e^{j\omega Y^2}\right].$$

The expected values are evaluated as follows:

$$E\left[e^{j\omega X^2}\right] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{j\omega x^2} e^{-x^2/2}\,dx$$

$$= \frac{1}{\sqrt{1-2j\omega}} \int_{-\infty}^{\infty} \sqrt{\frac{1-2j\omega}{2\pi}} e^{-(1-2j\omega)x^2/2}\,dx = \frac{1}{\sqrt{1-2j\omega}}.$$

The last step is accomplished using the normalization integral for Gaussian functions. The other expected value is identical to the first since X and Y have identical distributions. Hence,

$$\Phi_Z(\omega) = \left(\frac{1}{\sqrt{1-2j\omega}}\right)^2 = \frac{1}{1-2j\omega}.$$

The PDF is found from the inverse Fourier transform to be

$$f_Z(z) = \frac{1}{2}\exp\left(-\frac{x}{2}\right)u(x).$$

EXAMPLE 5.22: Suppose X and Y are independent zero-mean, unit-variance Gaussian random variables and we want to find the PDF of $Z = Y/X$. Conditioned on $X = x$, the transformation $Z = Y/x$ is a simple linear transformation and

$$f_{Z|X}(z|x) = |x| f_Y(xz) = \frac{|x|}{\sqrt{2\pi}} \exp\left(-\frac{x^2z^2}{2}\right).$$

Multiplying the conditional PDF by the marginal PDF of X and integrating out x gives the desired marginal PDF of Z.

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z|X}(z|x) f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{|x|}{\sqrt{2\pi}} \exp\left(-\frac{x^2z^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| \exp\left(-\frac{(1+x^2)z^2}{2}\right) dx$$

$$= \frac{1}{\pi} \int_{0}^{\infty} x \exp\left(-\frac{(1+z^2)x^2}{2}\right) dx = \frac{1}{\pi}\frac{1}{1+z^2}$$

Next, our attention moves to solving a slightly more general class of problems. Given two random variables X and Y, suppose we now create two new random variables W and Z according to some 2×2 transformation of the general form

$$\begin{aligned} Z &= g_1(X,Y) \\ W &= g_2(X,Y) \end{aligned} \tag{IV.45}$$

$$\begin{aligned} f_{Z,W}(z,w) &= f_{X,Y}(x,y)\left|J\begin{pmatrix} x & y \\ z & w \end{pmatrix}\right| \\ &= \frac{f_{X,Y}(x,y)}{\left|J\begin{pmatrix} z & w \\ x & y \end{pmatrix}\right|} \end{aligned} \tag{IV.46}$$

$$J\begin{pmatrix} x & y \\ z & w \end{pmatrix} = \det\begin{bmatrix} \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} \\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} \end{bmatrix} \tag{IV.47}$$

$$J\begin{pmatrix} z & w \\ x & y \end{pmatrix} = \det\begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{bmatrix} \tag{IV.48}$$

EXAMPLE 5.23: A classical example of this type of problem involves the transformation of two independent Gaussian random variables from Cartesian to polar coordinates. Suppose

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right).$$

We seek the PDF of the polar magnitude and phase given by

$$R = \sqrt{X^2+Y^2},$$
$$\Theta = \tan^{-1}(Y/X).$$

The inverse transformation is

$$X = R\cos(\Theta),$$

$$Y = R\sin(\Theta).$$

In this case, the inverse transformation takes on a simpler functional form and so we elect to use this form to compute the Jacobian.

$$J\begin{pmatrix} x & y \\ r & \theta \end{pmatrix} = \det\begin{bmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{bmatrix} = \det\begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix}$$

$$= r\cos^2(\theta) + r\sin^2(\theta) = r$$

The joint PDF of R and Θ is then

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(x,y)\left|J\begin{pmatrix} x & y \\ r & \theta \end{pmatrix}\right|\Bigg|_{\substack{x = h_1(r,\theta) \\ y = h_2(r,\theta)}}$$

$$= \frac{r}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\Bigg|_{\substack{x = r\cos(\theta) \\ y = r\sin(\theta)}}$$

$$= \frac{r}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right), \quad \begin{matrix} r \geq 0 \\ 0 \leq \theta < 2\pi \end{matrix}.$$

Note that in these calculations, we do not have to worry about taking the absolute value of the Jacobian since for this problem the Jacobian ($= r$) is always nonnegative. If we were interested, we could also find the marginal distributions of R and Θ to be

$$f_R(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)u(r) \quad \text{and} \quad f_\Theta(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi.$$

The magnitude follows a Rayleigh distribution while the phase is uniformly distributed over $(0, 2\pi)$.

***

V. Random Processes

V.1. Introduction

Definition V-1

A random process is a function of the elements of a sample space S, as well as another independent variable t. Given an experiment E with sample space S, the random process $X(t)$ maps each possible outcome $\xi \in S$ to a function $x(t,\xi)$ as specified by some rule.

Example V.1

Suppose an experiment consists of flipping a coin. If the outcome is heads $\xi = H$, the random process takes on the functional form $x_H(t) = \sin(\omega_0 t)$; whereas if the outcome is tails $\xi = T$, the realization $x_T(t) = \sin(2\omega_0 t)$ occurs.

Example V.2

Now suppose that an experiment results in a random variable A that is uniformly distributed over $[0,1)$. A random process is then constructed according to $X(t) = A\sin(\omega_0 t)$. Since the random variable is continuous, there are an uncountably infinite number of realizations of the random process.

The mean value of $X(t)$ is calculated as follows:

$$\begin{aligned}
\mathrm{E}\left[X(t)\right] &= \mathrm{E}\left[A\sin(\omega_0 t)\right] \\
&= \mathrm{E}\left[A\right]\sin(\omega_0 t) \\
&= \frac{1}{2}\sin(\omega_0 t)
\end{aligned}$$

Example V.3

Consider the experiment of rolling a standard die and assigning the number on the top face to random variable Z. Let a discrete random sequence be defined as $X(n) = X(n-1) + Z$, where $X(0) = 0$. A possible realization of $X(n)$ is $x(n) = 0,3,4,10,12,17,\cdots$.

Exercise V.1

- Determine the mean value of $X(n)$ in Example V.3 above.

$$\begin{aligned}
\mathrm{E}\left[X(n)\right] &= \mathrm{E}\left[X(n-1) + Z\right] \\
&= \mathrm{E}\left[X(n-1)\right] + \mathrm{E}\left[Z\right]
\end{aligned}$$

$$\mathrm{E}\left[X(0)\right] = 0$$

$$\begin{aligned} \mathrm{E}\left[X(1)\right] &= \mathrm{E}\left[X(0)\right] + \mathrm{E}\left[Z\right] \\ &= \mathrm{E}\left[Z\right] \\ &= 3.5 \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[X(2)\right] &= \mathrm{E}\left[X(1)\right] + \mathrm{E}\left[Z\right] \\ &= 3.5 + \mathrm{E}\left[Z\right] \\ &= 7 \end{aligned}$$

$$\mathrm{E}\left[X(n)\right] = 3.5n$$

- Determine the mean square value of $X(n)$ in Example V.3 above.
- Determine the variance of $X(n)$ in Example V.3 above.
- Replace Z in Example V.3 above with a uniform discrete random variable that takes its values from the set $\{\pm 1, \pm 3\}$, then determine the mean, mean square and variance of $X(n)$.

Example V.4

Let Z_m be a Gaussian random variable with a mean μ_m and a variance σ_m^2, for $m = 0, 1, 2, \cdots$. Let a random sequence be defined as $X(n) = \sum_{m=0}^{n-1} Z_m$. The PDF of $X(n)$ is Gaussian with a mean $\mu_{X(n)} = \sum_{m=0}^{n-1} \mu_m$ and a variance $\sigma_{X(n)}^2 = \sum_{m=0}^{n-1} \sigma_m^2$. If $\mu_m = \mu$ and $\sigma_m^2 = \sigma^2$ (both are constants), then the PDF of the random process $X(n)$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi n \sigma^2}} e^{-\frac{(x - n\mu)^2}{2n\sigma^2}}$$

EXAMPLE 8.10: Now suppose the random process of the previous example is slightly modified. In particular, consider a sine-wave process where the random variable is the phase, Θ, which is uniformly distributed over $[0, 2\pi)$. That is, $X(t) = a\sin(\omega_o t + \Theta)$. For this example, the amplitude of the sine wave, a, is taken to be fixed (not random). The mean function is then

$$\mu_X(t) = E[X(t)] = E[a\sin(\omega_o t + \Theta)] = a\int f_\Theta(\theta)\sin(\omega_o t + \theta)\,d\theta$$

$$= \frac{a}{2\pi}\int_0^{2\pi}\sin(\omega_o t + \theta)\,d\theta = 0,$$

which is a constant. Why is the mean function of the previous example a function of time and this one is not? Consider the member functions of the respective ensembles for the two random processes.

Definition V-2

The autocorrelation function $R_{XX}(t_1,t_2) = \mathrm{E}\left[X(t_1)X(t_2)\right]$ of a continuous random process $X(t)$ is defined as

$$\begin{aligned} R_{XX}(t_1,t_2) &= \mathrm{E}\left[X(t_1)X(t_2)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X_1,X_2}(x_1,x_2;t_1,t_2)\,dx_1 dx_2 \end{aligned} \tag{V.1}$$

EXAMPLE 8.12: Consider the sine wave process with a uniformly distributed amplitude as described in Examples 8.2 and 8.9, where $X(t) = A\sin(\omega_o t)$. The autocorrelation function is found as

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = E[A^2 \sin(\omega_o t_1)\sin(\omega_o t_2)]$$
$$= \tfrac{1}{3}\sin(\omega_o t_1)\sin(\omega_o t_2),$$

or

$$R_{XX}(t, t+\tau) = \tfrac{1}{3}\sin(\omega_o t)\sin(\omega_o(t+\tau)).$$

EXAMPLE 8.13: Now consider the sine wave process with random phase of Example 8.10 where $X(t) = a\sin(\omega_o t + \Theta)$. Then

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = E[a^2 \sin(\omega_o t_1 + \theta)\sin(\omega_o t_2 + \theta)].$$

To aid in calculating this expected value, we use the trigonometric identity

$$\sin(x)\sin(y) = \tfrac{1}{2}\cos(x-y) - \tfrac{1}{2}\cos(x+y).$$

The autocorrelation then simplifies to

$$R_{XX}(t_1, t_2) = \frac{a^2}{2}E[\cos(\omega_o(t_2 - t_1))] + \frac{a^2}{2}E[\cos(\omega_o(t_1 + t_2 + 2\theta))]$$
$$= \frac{a^2}{2}\cos(\omega_o(t_2 - t_1)),$$

or

$$R_{XX}(t, t+\tau) = \frac{a^2}{2}\cos(\omega_o \tau).$$

When

1. $\mathrm{E}\left[X(t)\right]$ is not function of t, and
2. $R_{X,X}(t, t+\tau)$ function of only τ,

The process is classified as wide-sense stationary (WSS).

DEFINITION 8.4: The autocovariance function, $C_{XX}(t_1, t_2)$, of a continuous time random process, $X(t)$, is defined as the covariance of $X(t_1)$ and $X(t_2)$:

$$C_{XX}(t_1, t_2) = \text{Cov}(X(t_1), X(t_2)) = E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))]. \quad (8.6)$$

The definition is easily extended to discrete time random processes.

As with the covariance function for random variables, the autocovariance function can be written in terms of the autocorrelation function and the mean function:

$$C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2). \quad (8.7)$$

DEFINITION 8.5: For a pair of random processes $X(t)$ and $Y(t)$, the crosscorrelation function is defined as

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)]. \quad (8.10)$$

Likewise, the cross-covariance function is

$$C_{XY}(t_1, t_2) = E[(X(t_1) - \mu_X(t_1))(Y(t_2) - \mu_Y(t_2))]. \quad (8.11)$$

EXAMPLE 8.15: Suppose $X(t)$ is a zero-mean random process with autocorrelation function $R_{XX}(t_1, t_2)$. A new process $Y(t)$ is formed by delaying $X(t)$ by some amount t_d. That is, $Y(t) = X(t - t_d)$. Then the crosscorrelation function is

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] = E[X(t_1)X(t_2 - t_d)] = R_{XX}(t_1, t_2 - t_d).$$

In a similar fashion, it is seen that $R_{YX}(t_1, t_2) = R_{XX}(t_1 - t_d, t_2)$ and $R_{YY}(t_1, t_2) = R_{XX}(t_1 - t_d, t_2 - t_d)$.

DEFINITION 8.4: The autocovariance function, $C_{XX}(t_1, t_2)$, of a continuous time random process, $X(t)$, is defined as the covariance of $X(t_1)$ and $X(t_2)$:

$$C_{XX}(t_1, t_2) = \text{Cov}(X(t_1), X(t_2)) = E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))]. \quad (8.6)$$

The definition is easily extended to discrete time random processes.

As with the covariance function for random variables, the autocovariance function can be written in terms of the autocorrelation function and the mean function:

$$C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2). \quad (8.7)$$

DEFINITION 8.5: For a pair of random processes $X(t)$ and $Y(t)$, the crosscorrelation function is defined as

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)]. \quad (8.10)$$

Likewise, the cross-covariance function is

$$C_{XY}(t_1, t_2) = E[(X(t_1) - \mu_X(t_1))(Y(t_2) - \mu_Y(t_2))]. \quad (8.11)$$

EXAMPLE 8.15: Suppose $X(t)$ is a zero-mean random process with autocorrelation function $R_{XX}(t_1, t_2)$. A new process $Y(t)$ is formed by delaying $X(t)$ by some amount t_d. That is, $Y(t) = X(t - t_d)$. Then the crosscorrelation function is

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] = E[X(t_1)X(t_2 - t_d)] = R_{XX}(t_1, t_2 - t_d).$$

In a similar fashion, it is seen that $R_{YX}(t_1, t_2) = R_{XX}(t_1 - t_d, t_2)$ and $R_{YY}(t_1, t_2) = R_{XX}(t_1 - t_d, t_2 - t_d)$.

If the crosscorrelation is zero, the processes are orthogonal.

V.2. Stationary and Ergodic Random Processes

DEFINITION 8.6: A continuous time random process $X(t)$ is *strict sense stationary* if the statistics of the process are invariant to a time shift. Specifically, for any time shift τ and any integer $n \geq 1$,

$$\begin{aligned} &f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) \\ &\quad = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; t_1 + \tau, t_2 + \tau, \ldots, t_n + \tau). \end{aligned} \quad (8.12)$$

DEFINITION 8.7: A random process is *wide sense stationary* (WSS) if the mean function and autocorrelation function are invariant to a time shift. In particular, this implies that

$$\mu_X(t) = \mu_X = \text{constant}, \quad (8.14)$$

$$R_{XX}(t, t + \tau) = R_{XX}(\tau) \quad \text{(function only of } \tau\text{)}. \quad (8.15)$$

All strict sense stationary random processes are also WSS, provided that the mean and autocorrelation function exist. The converse is not true. A WSS process does not necessarily need to be stationary in the strict sense. We refer to a process that is not WSS as *nonstationary*.

EXAMPLE 8.17: Suppose we form a random process $Y(t)$ by modulating a carrier with another random process, $X(t)$. That is, let $Y(t) = X(t)\cos(\omega_o t+\Theta)$ where Θ is uniformly distributed over $[0, 2\pi)$ and independent of $X(t)$. Under what conditions is $Y(t)$ WSS? To answer this, we calculate the mean and autocorrelation function of $Y(t)$.

$$\mu_Y(t) = E[X(t)\cos(\omega_o t + \Theta)] = E[X(t)]E[\cos(\omega_o t + \Theta)] = 0;$$

$$\begin{aligned} R_{YY}(t,t+\tau) &= E[X(t)X(t+\tau)\cos(\omega_o t)\cos(\omega_o(t+\tau))] \\ &= E[X(t)X(t+\tau)]\left\{\tfrac{1}{2}\cos(\omega_o\tau)+\tfrac{1}{2}E[\cos(\omega_o(2t+\tau)+2\Theta)]\right\} \\ &= \tfrac{1}{2}R_{XX}(t,t+\tau)\cos(\omega_o\tau) \end{aligned}$$

While the mean function is a constant, the autocorrelation is not necessarily only a function of τ. The process $Y(t)$ will be WSS provided that $R_{XX}(t,t+\tau) = R_{XX}(\tau)$. Certainly if $X(t)$ is WSS, then $Y(t)$ will be as well.

EXAMPLE 8.18: Let $X(t) = At + B$ where A and B are independent random variables, both uniformly distributed over the interval $(-1, 1)$. To determine whether this process is WSS, calculate the mean and autocorrelation functions:

$$\mu_X(t) = E[At + B] = E[A]t + E[B] = 0;$$

$$\begin{aligned} R_{XX}(t, t + \tau) &= E[(At + B)(A(t + \tau) + B)] \\ &= E[A^2]t(t + \tau) + E[B^2] + E[AB](2t + \tau) = \tfrac{1}{3}t(t + \tau) + \tfrac{1}{3}. \end{aligned}$$

Clearly, this process is not WSS.

DEFINITION 8.8: A WSS random process is ergodic if ensemble averages involving the process can be calculated using time averages of any realization of the process. Two limited forms of ergodicity are

(1) ergodic in the mean: $\langle x(t)\rangle = \mathrm{E}\left[x(t)\right]$;

(2) ergodic in the autocorrelation: $\langle x(t)x(t+\tau)\rangle = E\left[x(t)x(t+\tau)\right]$.

EXAMPLE 8.20: Now consider the sinusoid with random phase $X(t) = a\sin(\omega_o t + \Theta)$, where Θ is uniform over $[0, 2\pi)$. It was demonstrated in Example 8.13 that this process is WSS. But is it ergodic? Given any realization $x(t) = a\sin(\omega_o t + \theta)$, the time average is $\langle x(t)\rangle = \langle a\sin(\omega_o t + \theta)\rangle = 0$. That is, the average value of any sinusoid is zero. So this process is ergodic in the mean since the ensemble average of this process was also zero. Next, consider the sample autocorrelation function:

$$\langle x(t)x(t+\tau)\rangle = a^2\langle \sin(\omega_o t + \theta)\sin(\omega_o t + \omega_o\tau + \theta)\rangle$$
$$= \frac{a^2}{2}\langle\cos(\omega_o\tau)\rangle - \frac{a^2}{2}\langle\cos(2\omega_o t + \omega_o\tau + 2\theta)\rangle = \frac{a^2}{2}\cos(\omega_o\tau).$$

This also is exactly the same expression obtained for the ensemble averaged autocorrelation function. Hence, this process is also ergodic in the autocorrelation.

V.3. Properties of the Autocorrelation Function

$$\begin{aligned} R_{XX}(0) &= \mathrm{E}\left[X^2(t)\right] \\ &= P_X \end{aligned} \tag{V.2}$$

$$R_{XX}(\tau) = R_{XX}(-\tau) \tag{V.3}$$

$$\left|R_{XX}(\tau)\right| \le R_{XX}(0) \tag{V.4}$$

EXAMPLE 8.23: Consider the random process $X(t) = A\cos(\omega_o t) + B\sin(\omega_o t)$, where A and B are independent, zero-mean Gaussian random variables with equal variances of σ^2. This random process is formed as a linear combination of two Gaussian random variables, and hence samples of this process are also Gaussian random variables. The mean and autocorrelation functions of this process are found as

$$\mu_X(t) = E[A\cos(\omega_o t) + B\sin(\omega_o t)] = E[A]\cos(\omega_o t) + E[B]\sin(\omega_o t) = 0,$$
$$\begin{aligned} R_{XX}a(t_1, t_2) &= E[(A\cos(\omega_o t_1) + B\sin(\omega_o t_1))(A\cos(\omega_o t_2) + B\sin(\omega_o t_2))] \\ &= E[A^2]\cos(\omega_o t_1)\cos(\omega_o t_2) + E[B^2]\sin(\omega_o t_1)\sin(\omega_o t_2) \\ &\quad + E[AB]\{\cos(\omega_o t_1)\sin(\omega_o t_2) + \sin(\omega_o t_1)\cos(\omega_o t_2)\} \\ &= \sigma^2\{\cos(\omega_o t_1)\cos(\omega_o t_2) + \sin(\omega_o t_1)\sin(\omega_o t_2)\} \\ &= \sigma^2\cos(\omega_o(t_2 - t_1)). \end{aligned}$$

Note that this process is WSS since the mean is constant and the autocorrelation function depends only on the time difference. Since the process is zero-mean, the first order PDF is that of a zero-mean Gaussian random variable:

$$f_X(x;t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

This PDF is independent of time as would be expected for a stationary random process. Now consider the joint PDF of two samples, $X_1 = X(t)$ and $X_2 = X(t+\tau)$. Since the process is zero-mean, the mean vector is simply the all-zeros vector. The covariance matrix is then of the form

$$\boldsymbol{C_{XX}} = \begin{bmatrix} R_{XX}(0) & R_{XX}(\tau) \\ R_{XX}(\tau) & R_{XX}(0) \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \cos(\omega_o\tau) \\ \cos(\omega_o\tau) & 1 \end{bmatrix}.$$

The joint PDF of the two samples would then be

$$f_{X_1,X_2}(x_1,x_2;t,t+\tau) = \frac{1}{2\pi\sigma^2|\sin(\omega_o\tau)|} \exp\left(-\frac{x_1^2 - 2x_1x_2\cos(\omega_o\tau) + x_2^2}{2\sigma^2\sin^2(\omega_o\tau)}\right).$$

Note once again that this joint PDF is dependent only on time difference, τ, and not on absolute time t. Higher order joint PDFs could be worked out in a similar manner.

V.4. Power Spectral Density

THEOREM 10.1 (Wiener-Khintchine-Einstein): For a wide sense stationary (WSS) random process $X(t)$ whose autocorrelation function is given by $R_{XX}(\tau)$, the PSD of the process is

$$S_{XX}(f) = F(R_{XX}(\tau)) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-j2\pi f\tau}\, d\tau. \tag{10.13}$$

In other words, the autocorrelation function and PSD form a Fourier transform pair.

EXAMPLE 10.2: Let us revisit the random sinusoidal process, $X(t) = A\sin(\omega_o t + \Theta)$, of Example 10.1. This time the PSD function will be calculated by first finding the autocorrelation function.

$$R_{XX}(t,t+\tau) = E[X(t)X(t+\tau)] = E[A^2\sin(\omega_o t + \Theta)\sin(\omega_o(t+\tau) + \Theta)]$$
$$= \frac{1}{2}E[A^2]E[\cos(\omega_o\tau) - \cos(\omega_o(2t+\tau) + 2\Theta)] = \frac{1}{2}E[A^2]\cos(\omega_o\tau)$$