Lecture Notes in Applied Probability

2005

Wayne F. Bialas
Department of Industrial Engineering
University at Buffalo

Department of Industrial Engineering
University at Buffalo
301 Bell Hall
Buffalo, New York 14260
USA

E-mail: bialas@buffalo.edu
Web: http://www.buffalo.edu/~bialas

Date printed: May 12, 2005

CONTENTS

ACKNOWLEDGEMENTS

Contained herein is a bit of the wisdom of the greatest scholars in probability theory. Some of them were my professors.

These notes have benefited the most from the many contributions of my students and teaching assistants. In particular, I am grateful to Mark Chew, Alfred Guiffrida and Nishant Mishra for several contributions to these notes. I must also thank Yu-Shuan Lee Brown, Rajarshi Chakraborty, Kevin Cleary, Shiju David, Satyaki Ghosh-Dastidar, Judah Eisenhandler, Harold T. Gerber, Michael Holender, Dawn Holmes, Derek Hoiem, Simin Huang, Mona Khanna, Steven Kieffer, Keith Klopfle, Lindsay Koslowski, Taneka Lawson, Andrew Lenkei, Nguk Ching Ling, Angelo Morgante, Xiaofeng Nie, James Pasquini, Nikola Petrovic, Amol Salunkhe, Lynne Makrina Tkaczevski, Elif Tokar and Min Zhang for finding numerous mistakes in previous editions.

Any remaining errors and omissions are mine.

This manual was typeset using Donald Knuth's TeX with Leslie Lamport's LaTeX enhancements as implemented by Eberhard Mattes. A LaTeX style file developed by H. Partel was modified and used to format this document. The Adobe Acrobat™ driver, *dvipdfm,* written by Mark A. Wicks, and AFPL Ghostscript were used to produce the printed output. I would not have dared to prepare these notes without these tools.

WFB

Buffalo, New York
May 12, 2005

TO THE STUDENT

These notes are for a course devoted to describing those processes around us that appear to produce random outcomes. Probability theory provides a way to mathematically represent chance.

We will show you how sets can be used to express events, and then assign, in a systematic way, a measure of uncertainty to each of those sets. That measure will be called *probability*.

For some problems, we can assign probabilities to events by simply counting equally likely outcomes. For that, we provide some counting techniques. When outcomes are not equally likely, or when they arise from a continuum of possible values, we will use the mathematical theory of measure.

As engineers and scientists, we are always more comfortable working with real numbers. For that purpose, we will introduce the concept of *random variables* and *random vectors* to map our abstract probability problems into Euclidean real space. There we will quickly discover analogies between allocating probability and allocating physical mass. Point masses and continuous mass densities arise as a natural consequence. The concepts of center of mass (center of gravity) and moment of inertia, become the *mean* and *variance* of our probability distributions.

But, ultimately, the goal of any good mathematical model is to serve as a tool for reaching conclusions about the processes being modeled. We will do that, too, with examples ranging from coin tossing to proving one of the most important results in all of science, the Central Limit Theorem.

What is certain, is that a course in probability asks that you change your view of everything around you. The universe is not a deterministic clockwork, but a harmony of random possibilities.

A fool must now and then be right by chance.
– WILLIAM COWPER, *Conversation* (c. 1790)

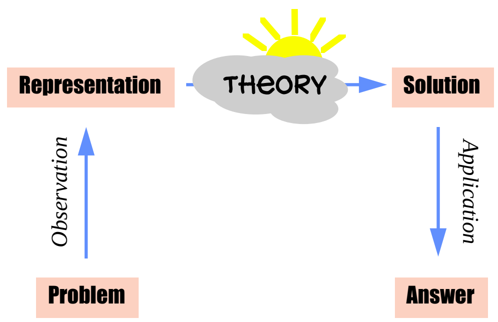
1

SETS, SAMPLE SPACES AND EVENTS

There are no little events in life.
– AMELIA E. BARR, *All the Days of My Life* (1913)

Mathematical Models

What is a mathematical model?



Deterministic models

For specified values of an experiment's input parameters, the result is a known constant.

Example: $V = IR$

Established experimentally by George Simon Ohm in 1827[1]

Nondeterministic models

Also called *probabilistic models* or *stochastic models*.

Examples:

[1]To place this date in perspective, University at Buffalo was founded in 1846.

- the number of α-particles emitted from a piece of radioactive material in one minute.
- the outcome of a single toss of a fair die.
- the number of students who will earn an A in EAS305.

Sets

Basic definitions

We will need to use **set theory** to describe possible outcomes and scenarios in our model of uncertainty.

Definition 1.1. *A* **set** *is a collection of elements.*

> **Notation:** Sets are denoted by capital letters ($A, B, \mathbf{\Omega}, \ldots$) while their elements are denoted by small letters ($a, b, \omega, \ldots$). If x is an element of a set A, we write $x \in A$. If x is not an element of A, we write $x \notin A$.

A set can be described

- by listing the set's elements

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

- by describing the set in words

  "A is the set of all real numbers between 0 and 1, inclusive."

- by using the notation $\{\omega : \text{specification for } \omega\}$

$$A = \{x : 0 \leq x \leq 1\}$$

  or

$$A = \{0 \leq x \leq 1\}$$

There are two sets of special importance:

Definition 1.2. *The* **universe** *is the set containing all points under consideration and is denoted by* Ω*.*

Question: What is the difference (if any) between Ø and $\{$Ø$\}$?

Definition 1.3. *The* **empty set** *(the set containing no elements) is denoted by* Ø*.*

Definition 1.4. *If every point of set A belongs to set B, then we say that A is a* **subset** *of B (B is a* **superset** *of A). We write*

$$A \subseteq B \qquad\qquad B \supseteq A$$

Now we have a way to say A is **identical** to B...

Definition 1.5. *$A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

Example: Let $\Omega = \mathbb{R}$, the set of all real numbers. Define the sets

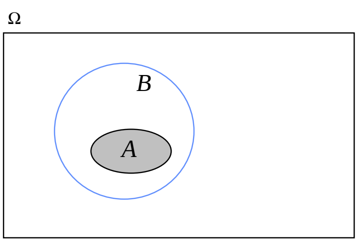
$$\begin{aligned} A &= \{x : x^2 + x - 2 = 0\} \\ B &= \{x : (x-3)(x^2 + x - 2) = 0\} \\ C &= \{-2, 1, 3\} \end{aligned}$$

Then $A \subseteq B$ and $B = C$.

Definition 1.6. *A is a* **proper subset** *of B if $A \subseteq B$ and $A \neq B$. We write $A \subsetneq B$.*

Note: Some authors use the symbol $\subset$ for *subset* and other use $\subset$ for *proper subset*. We avoid the ambiguity here by using $\subseteq$ for *subset* and $\subsetneq$ for *proper subset*.

Venn diagrams

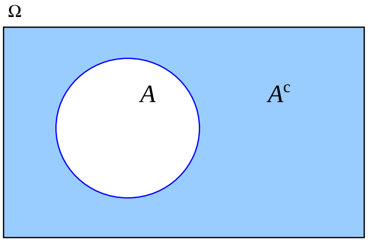


Set operations

Complementation

$$A^c = \{\omega \in \Omega : \omega \notin A\}$$

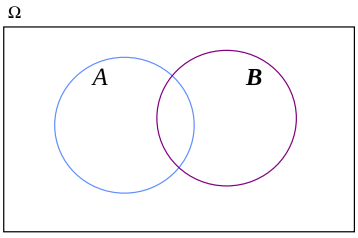
Example: Here is the region representing A^c in a Venn diagram:



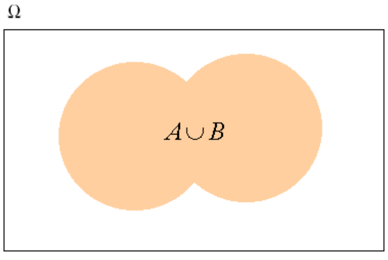
Union

$$A \cup B = \{\omega \in \mathbf{\Omega} \,:\, \omega \in A \text{ or } \omega \in B \text{ (or both)}\}$$

Example: Consider the following Venn diagram:



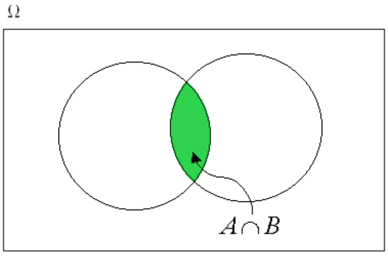
then the region representing $A \cup B$ is:



Intersection

$$A \cap B = \{\omega \in \mathbf{\Omega} \,:\, \omega \in A \text{ and } \omega \in B\}$$

Example: Using the same Venn diagram as above the region representing $A \cap B$ is given by:



Note that some textbooks use the notation AB instead of $A \cap B$ to denote the intersection of two sets.

Definition 1.7. *Two sets A and B are* **disjoint** *(or* **mutually exclusive***) if $A \cap B = \emptyset$.*

These basic operations can be extended to any finite number of sets.

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

and

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

You can show that

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

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

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

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

Note: (a) and (b) are the *commutative laws*

Note: (c) and (d) are the *associative laws*

Set identities

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

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

(g) $A \cap \emptyset = \emptyset$

(h) $A \cup \emptyset = A$

(i) $(A \cup B)^c = A^c \cap B^c$

(j) $(A \cap B)^c = A^c \cup B^c$

(k) $(A^c)^c = A$

Note: (e) and (f) are called the *Distributive Laws*

Note: (i) and (j) are called *DeMorgan's Laws*

Proving statements about sets

Venn diagrams can only illustrate set operations and basic results. You cannot prove a true mathematical statement regarding sets using Venn diagrams. But they can be used to disprove a false mathematical statement.

Here is an example of a simple set theorem and its proof.

Theorem 1.1. *If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.*

Proof.

	Discussion	*Reasons*
(1)	If $x \in A$ then $x \in B$	Definition of $A \subseteq B$
(2)	Since $x \in B$ then $x \in C$	Definition of $B \subseteq C$
(3)	If $x \in A$ then $x \in C$	By statements (1) and (2)
(4)	Therefore $A \subseteq C$	Definition of $A \subseteq C$ ■

Note: A proof is often concluded with the symbol ■ (as in the above proof) or the letters *QED*.

Many mathematicians prefer to write proofs in paragraph form. For example, the proof of Theorem 1.1 would become:

Proof. Choose an $x \in A$. Since $A \subseteq B$, $x \in A$ implies $x \in B$, from the definition of subset. Furthermore, since $B \subseteq C$, $x \in B$ implies $x \in C$. Therefore, every element $x \in A$ is also an element of C. Hence $A \subseteq C$. ■

A proof of the distributive law for sets

Here is a proof of one of the distributive laws for sets. The proof is from a textbook on set theory by Flora Dinkines[2]

Theorem 1.2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof.

	Discussion	*Reasons*
(1)	If $x \in A \cap (B \cup C)$, then $x \in A$ and $x \in B \cup C$.	Definition of $\cap$.
(2)	Since $x \in B \cup C$, then $x \in B$ or $x \in C$.	Definition of $\cup$.
(3)	Therefore $x \in A$ and $x \in B$, *or* $x \in A$ and $x \in C$.	By (1) and (2).
(4)	Hence $x \in A \cap B$ *or* $x \in A \cap C$.	Definition of $\cap$.
(5)	$x \in (A \cap B) \cup (A \cap C)$.	Definition of $\cup$.

Therefore every element of $A \cap (B \cup C)$ is also an element of $(A \cap B) \cup (A \cap C)$, giving us

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C).$$

We now prove the inclusion in the opposite direction:

[2]Dinkines, F., *Elementary Theory of Sets*, Meredith Publishing, 1964.

(6)	If $y \in (A \cap B) \cup (A \cap C)$, then $y \in (A \cap B)$ or $y \in (A \cap C)$.	Definition of $\cup$.
(7)	If $y \in A \cap B$ then $y \in A$ and $y \in B$.	Definition of $\cap$.
(8)	If $y \in A \cap C$ then $y \in A$ and $y \in C$.	Same as (7).
(9)	In either case (7) or (8), $y \in A$ and y is an element of one of the sets B or C.	Statements (7) and (8)
(10)	Therefore $y \in A$ and $y \in B \cup C$.	Since y is in B or C, it is in the union.
(11)	Therefore $y \in A \cap (B \cup C)$.	Definition of $\cap$.

Therefore every element of $(A \cap B) \cup (A \cap C)$ is also an element of $A \cap (B \cup C)$, giving us

$$A \cap (B \cup C) \supseteq (A \cap B) \cup (A \cap C).$$

Therefore, using the definition of set equality, we have

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

■

Elementary sets

Let Ω be a universe with three subsets A, B, and C. Consider the following subsets of Ω:

A	B	C		
0	0	0	$A^c \cap B^c \cap C^c$	$= S_0$
0	0	1	$A^c \cap B^c \cap C$	$= S_1$
0	1	0	$A^c \cap B \cap C^c$	$= S_2$
0	1	1	$A^c \cap B \cap C$	$= S_3$
1	0	0	$A \cap B^c \cap C^c$	$= S_4$
1	0	1	$A \cap B^c \cap C$	$= S_5$
1	1	0	$A \cap B \cap C^c$	$= S_6$
1	1	1	$A \cap B \cap C$	$= S_7$

The sets $S_0, S_1, \ldots, S_7$ are called the **elementary sets** of A, B and C.

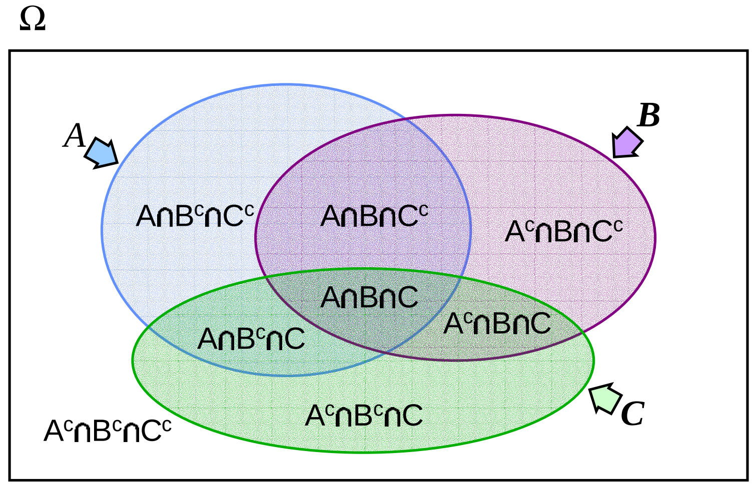


Figure 1.1: An example of elementary sets for three subsets

For example, the binary sequence (0,1,1) corresponds to $S_3 = A^c \cap B \cap C$. This is the subset of elements that are

0:	not in A	and
1:	in B	and
1:	in C	

Figure 1.1 illustrates all of the $2^3 = 8$ elementary sets for A, B and C using a Venn diagram.

Elementary sets have two very useful properties:

- Elementary sets are pairwise disjoint. That is, $S_i \cap S_j = \text{Ø}$ for any $i \neq j$.
- Any set operation involving A, B, and C can always be represented as a *union* of the elementary sets for A, B and C. For example

$$(A \cup B) \cap C = S_3 \cup S_5 \cup S_7$$

Cardinality of sets

The *cardinality* of a set A (denoted by $|A|$) is the number of elements in A.

For some (but not all) experiments, we have

- $|\Omega|$ is finite.
- Each of the outcomes of the experiment are equally likely.

In such cases, it is important to be able to enumerate (i.e., count) the number of elements of subsets of Ω.

Sample spaces and events

Using set notation, we can now introduce the first components of our probability model.

Definition 1.8. *A* **sample space**, Ω, *is the set of all possible outcomes of an experiment.*

Definition 1.9. *An* **event** *is any subset of a sample space.*

Example: Consider the experiment of rolling a die once.

◇ What should be Ω for this experiment?

$\Omega = \{$ ⚀ ⚁ ⚂ ⚃ ⚄ ⚅ $\}$

If you thought the answer was $\Omega = \{1, 2, 3, 4, 5, 6\}$, then you just (re)invented the concept of a *random variable*. This will be discussed in a later chapter.

◇ Identify the subsets of Ω that represent the following events:

1. The die turns up even.

   $A = \{$ ⚁ ⚃ ⚅ $\}$

2. The die turns up less than 5.

   $B = \{$ ⚀ ⚁ ⚂ ⚃ $\}$

3. The die turns up 6.

   $C = \{$ ⚅ $\}$

Example: Consider the experiment of burning a light bulb until it burns out. The outcome of the experiment is the life length of the bulb.

◇ What should be Ω for this experiment?

$$\Omega = \{x \in \mathbb{R} : x \geq 0\}$$

For the lack of a better representation, we resorted to assigning a nonnegative real number to each possible outcome of the experiment.

◇ Identify the subsets of Ω that represent the following events:

1. The bulb burns out within one and one-half hours.

$$A = \{x \in \mathbb{R} : 0 \leq x < 1.5\}$$

2. The bulb lasts at least 5 hours.

$$B = \{x \in \mathbb{R} : x \geq 5\}$$

3. The bulb lasts exactly 24 hours.

$$C = \{24\}$$

Sets of events

It is sometimes useful to be able to talk about the set of all possible events $\mathfrak{F}$. The set $\mathfrak{F}$ is very different than the set of all possible outcomes, Ω. Since an event is a set, the set of all events is really a set of sets. Each element of $\mathfrak{F}$ is a subset of Ω.

For example, suppose we toss a two-sided coin twice. The sample space is

$$\Omega = \{HH, HT, TH, TT\}.$$

If you were asked to list all of the possible events for this experiment, you would need to list all possible subsets of Ω, namely

$$\begin{aligned}
A_1 &= \{\{HH\}\} \\
A_2 &= \{\{HT\}\} \\
A_3 &= \{\{TH\}\} \\
A_4 &= \{\{TT\}\} \\
A_5 &= \{\{HH\}, \{HT\}\} \\
A_6 &= \{\{HH\}, \{TH\}\}
\end{aligned}$$

$$
\begin{aligned}
A_7 &= \{\{HH\}, \{TT\}\} \\
A_8 &= \{\{HT\}, \{TH\}\} \\
A_9 &= \{\{HT\}, \{TT\}\} \\
A_{10} &= \{\{TH\}, \{TT\}\} \\
A_{11} &= \{\{HH\}, \{HT\}, \{TH\}\} \\
A_{12} &= \{\{HH\}, \{HT\}, \{TT\}\} \\
A_{13} &= \{\{HH\}, \{TH\}, \{TT\}\} \\
A_{14} &= \{\{HT\}, \{TH\}, \{TT\}\} \\
A_{15} &= \{\{HH\}, \{HT\}, \{TH\}, \{TT\}\} = \Omega \\
A_{16} &= \{\} = \emptyset
\end{aligned}
$$

In this case

$$\mathfrak{F} = \{A_1, A_2, \ldots, A_{16}\}$$

Definition 1.10. *For a given set A, the set of all subsets of A is called the* **power set** *of A, and is denoted by 2^A.*

Theorem 1.3. *If A is a finite set, then $|2^A| = 2^{|A|}$*

Proof. Left to reader.

If Ω is a sample space, then we often use the set of all possible events $\mathfrak{F} = 2^\Omega$. When Ω contains a finite number of outcomes, this often works well. However, for some experiments, Ω contains an infinite number of outcomes (either countable or uncountable) and $\mathfrak{F} = 2^\Omega$ is much too large. In those cases, we select a smaller collection of sets, $\mathfrak{F}$, to represent those events we actually need to consider. How to properly select a smaller $\mathfrak{F}$, is found in many graduate level probability courses.

Self-Test Exercises for Chapter 1

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S1.1 The event that corresponds to the statement, "at least one of the events A, B and C occurs," is

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

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

(C) $(A \cap B \cap C)^{\mathsf{c}}$

(D) $(A \cup B \cup C)^{\mathsf{c}}$

(E) none of the above

S1.2 The event that corresponds to the statement that exactly one of the events A, B or C occurs is

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

(B) $(A \cap B \cap C)^c$

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

(D) $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$

(E) none of the above.

S1.3 The set $(A \cap B) \cup (A \cap B^{\mathsf{c}})$ is equivalent to

(A) Ø

(B) Ω

(C) A

(D) B

(E) none of the above.

S1.4 The set $(A \cap B) \cup A$ is equivalent to

(A) $(A \cup B) \cap A$

(B) B

(C) $A \cap B$

(D) Ø

(E) none of the above.

S1.5 For any sets A and B,

(A) $(A \cap B) \subseteq B$

(B) $A \subseteq (A \cup B)$

(C) $A^c \subseteq (A \cap B)^c$

(D) all of the above are true.

(E) none of the above are true.

S1.6 For any events A and B, the event $(A \cup B^c) \cap B$ equals

(A) Ø

(B) Ω

(C) $A \cup B$

(D) $A \cap B$

(E) none of the above.

S1.7 If $A \cap B = \text{Ø}$ then $(A^c \cup B^c)^c$ equals

(A) Ø

(B) A

(C) B

(D) Ω

(E) none of the above.

Questions for Chapter 1

1.1 Suppose that the universe consists of the positive integers from 1 through 10. Let $A = \{2, 3, 4\}$, $B = \{3, 4, 5\}$ and $C = \{5, 6, 7\}$. List the elements of the following sets:

(a) $A^{\mathsf{c}} \cap B$

(b) $A^{\mathsf{c}} \cup B$

(c) $(A^{\mathsf{c}} \cup B^{\mathsf{c}})^{\mathsf{c}}$

(d) $[A \cup (B \cap C)^{\mathsf{c}}]^{\mathsf{c}}$

1.2 Suppose that the universe, **Ω**, is given by $\mathbf{\Omega} = \{x : 0 \leq x \leq 2\}$. Let $A = \{x : 0.5 \leq x \leq 1\}$. and $B = \{x : 0.25 \leq x \leq 1.5\}$. Describe the following sets:

(a) $(A \cup B)^{\mathsf{c}}$

(b) $A \cup B^{\mathsf{c}}$

(c) $(A \cap B)^{\mathsf{c}}$

(d) $A^{\mathsf{c}} \cap B$

1.3 Which of the following relationships are true? In each case, either prove your result using basic set relationships or disprove by providing a counterexample.

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

(b) $(A \cup B) = (A \cap B^{\mathsf{c}}) \cup B$

(c) $A^{\mathsf{c}} \cap B = A \cup B$

(d) $(A \cup B)^{\mathsf{c}} \cap C^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}} \cap C^{\mathsf{c}}$

(e) $(A \cap B) \cap (B^{\mathsf{c}} \cap C) = \emptyset$

1.4 Identify the set obtained by the following set operations, where Ω is the universe and A is a subset of Ω.

(a) $A \cap A^{\mathsf{c}}$

(b) $A \cup A^{\mathsf{c}}$

(c) $A \cap \emptyset$

(d) $A \cup \emptyset$

(e) $\Omega \cap \emptyset$

(f) $\Omega \cup \emptyset$

(g) $A \cap \Omega$

(h) $A \cup \Omega$

1.5 Let

$A =$ set of all even integers,

$B =$ set of all integers that are integral multiples of 3,

$C =$ set of all integers that are integral multiples of 4, and

$J =$ set of all integers

Describe the following sets:

(a) $A \cap B$

(b) $A \cup B$

(c) $B \cap C$

(d) $A \cup J$

(e) $A \cup C$

1.6 Prove the *Law of Absorption*, i.e., show that for any sets A and B, the following holds:

$$A \cup (A \cap B) = A \cap (A \cup B) = A.$$

1.7 A box of N light bulbs has r ($r \leq N$) bulbs with broken filaments. These bulbs are tested, one by one, until a defective bulb is found.

(a) Describe the sample space for the above experiment.

(b) Suppose that the above bulbs are tested, one by one, until *all* defectives have been found. Describe the sample space for this experiment.

1.8 During a 24–hour period, at some time, X, a switch is placed in the "ON" position. Subsequently, at some future time, Y, (still during the same 24–hour period) the switch is placed in the "OFF" position. Assume that X and Y are measured in hours on the continuous time axis with the beginning of the time period as the origin. The outcome of the experiment consists of the pair of numbers (X, Y).

(a) Describe the sample space.

(b) Describe and sketch in the XY–plane the following events:

1. The circuit is on for less than one hour.
2. The circuit is on at time z where z is some instant of time during the 24–hour period.
3. The circuit is turned on before time r and turned off after time s ($r < s$).
4. The circuit is on exactly three times as long as it is off.

1.9 During the course of a day, a machine produces three items, each of whose quality, defective or nondefective, is determined at the end of the day. Describe the sample space generated by a day's production.

1.10 A space probe is sent to Mars. If it fails, another probe is sent. If the second probe fails, another follows. This process continues until a successful probe is launched. Describe the sample space for this experiment.

1.11 Let A, B and C be three events associated with an experiment. Express the following verbal statements in set notation:

(a) At least one of the events occurs.

(b) Exactly one of the events occurs.

(c) Exactly two of the events occurs.

(d) At least one event occurs, but no more than two of the events occur.

1.12 (†) (provided by Nishant Mishra) Provide a theorem that is true for Venn diagrams but false for an arbitrary universe Ω. Your theorem should only use sets and basic set operations. Do not add other attributes of elements of $\mathbb{R}^2$ such as the distance between points.

2

ENUMERATION

Do not count your chickens before they are hatched.
– AESOP, *Fables* (600 BC)

This section is based on Chapter 3 of a text by K. L. Chung[1]

So you think you know how to count? Consider the following problem: Suppose you are the chief engineer at the *Ellicott Dominoes* domino factory. Dominoes are rectangular tiles used in some parlor games and they look like Figure 2.1. Typically,

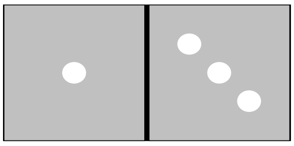


Figure 2.1: A single domino

dominoes have each half of a tile painted with 1 to 9 dots (called "pips") or a blank for a total of 10 possible choices.

Of course, a domino painted (1,4) can be turned 180 degrees to look like a domino painted (4,1). So there is no need to have one assembly line producing (1,4) dominoes and another producing (4,1) dominoes.[2] How many different dominoes does your factory actually produce? (Try this problem on your friends and see how many different answers you get.)

[1]Chung, K. L., *Elementary probability theory with stochastic processes*, Springer-Verlag, New York, 1975.

[2]Just have a person at the end of the assembly line rotating half of the (4,1) the dominoes by 180 degrees.

In this chapter, we will present a systematic way to approach counting problems. We will classify counting problems into four categories, and solve each type. In practice, most counting problems can be approached using one or a combination of these four basic types.

Fundamental Rule

The following rule forms the foundation for solving any counting problem:

Multiplication Principle: Suppose there are several multiple choices to be made. There are m_1 alternatives for the first choice, m_2 alternatives for the second, m_3 for the third, etc. The total number of alternatives for the whole set of choices is equal to

$$(m_1)(m_2)(m_3)\cdots.$$

Example: Suppose a engineering class has five women and four men. You would like to select a research team consisting of one man and one woman. There are four choices for selecting the man and five choices for selecting the woman. The total number of possible research teams is, therefore, $4 \cdot 5 = 20$

Sampling

The Urn Model: Imagine an urn that contains n balls marked 1 to n. There are four disciplines under which the balls can be drawn from the urn and examined.

I. *Sampling with replacement and with ordering.*

A total of m balls are drawn sequentially. After a ball is drawn, it is replaced in the urn and may be drawn again. The order of the balls is noted. Hence the sequence $1, 3, 2, 3$ is distinguished from $3, 3, 1, 2$. Since there are n possible ways to draw the first ball, n possible ways to draw the second, etc., the total number of possible sequences is

$$n^m.$$

Example: Suppose you give a quiz with 5 questions, each with a possible answer of True or False. Therefore, there are two ways to answer the first question, two ways to answer the second question, and so on. The total number of possible ways to answer all of the questions on the exam is

$$2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32.$$

II. *Sampling without replacement and with ordering.*

A total of m balls are drawn sequentially with $m \leq n$. After a ball is drawn, it is not replaced in the urn and hence cannot be drawn again. The order of the balls is noted. Since there are n possible ways to draw the first ball, $n-1$ possible ways to draw the second, $n-2$ possible ways to draw the third, etc., the total number of possible sequences is

$$n(n-1)(n-2)\cdots(n-m+1).$$

For the special case when $m = n$, the total number of possible ways to draw n balls is

$$n(n-1)(n-2)\cdots(2)(1) \equiv n!.$$

Note that when $m = n$, this is equivalent to the problem of counting the number of ways you can arrange n different balls in a row.

Computational Hint: We can compute

$$n(n-1)(n-2)\cdots(n-m+1) = \frac{n!}{(n-m)!}.$$

By definition, $0! \equiv 1$ so that when $n = m$, we have

$$n(n-1)(n-2)\cdots(n-n+1) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!.$$

Example: Suppose you have five different birthday cards, and you have three friends (named Bill, Mary, and Sue) with birthdays this week. You want to send one card to each friend. The number of possible choices for Bill's card is 5. The card for Mary's must come from the remaining 4 cards. Finally, the number of possible choices for Sue's card is 3. Hence the total number of ways you can assign five birthday cards to your three friends is

$$5 \cdot 4 \cdot 3 = 60.$$

It's interesting to note that the sequence of friends (Bill, then Mary, then Sue) is arbitrary and does not affect the answer.

Here is another way to look at this problem: Consider the empty mailboxes of your three friends,

$$\underset{B}{\underline{?}} \quad \underset{M}{\underline{?}} \quad \underset{S}{\underline{?}}$$

If the cards are labelled a,b,c,d,e, then one possible assignment is

$$\frac{\text{c}}{B} \quad \frac{\text{d}}{M} \quad \frac{\text{a}}{S}$$

There are 5 ways to fill the first position (i.e., Bill's mailbox), 4 ways to fill the second position and 3 ways to fill the third.

Example: Using the above example, we can compute the number of ways one can arrange the three letters x, y, z. This time we have 3 positions to fill with 3 different items.

$$\frac{?}{1} \quad \frac{?}{2} \quad \frac{?}{3}$$

So a possible arrangement is

$$\frac{\text{y}}{1} \quad \frac{\text{z}}{2} \quad \frac{\text{x}}{3}$$

and there are $3 \cdot 2 \cdot 1 = 3! = 6$ possible assignments.

III. *Sampling without replacement and without ordering.*

A total of m balls ($m \leq n$) are drawn from the urn. The balls are not replaced after each drawing and the order of the drawings is not observed. Hence, this case can be thought of as reaching into the urn and drawing m balls simultaneously. The total number of ways a subset of m items can be drawn from a set of n elements is

$$\frac{n!}{m!(n-m)!} \equiv \binom{n}{m}. \tag{1}$$

This number is often referred to as the **binomial coefficient**.

To derive the above formula, let x denote the number of ways to choose m items from n items without replacement and without ordering. We can find the value of x by considering Case II as a *two-stage* process:

Stage 1 Grab m items from n without replacement and without ordering.

Stage 2 Given the m items selected in Stage 1, arrange the m items in order.

Combined, the two stages implement Case II (sampling without replacement and *with* ordering). Stage 1 (selecting the m balls) can be performed x ways, our unknown quantity. Stage 2 (ordering the m balls) can be performed $m!$ ways (using Case II for m items drawn from m items).

Furthermore, we know, using Case II (with m items drawn from n items), that the entire process of Stage 1 followed by Stage 2 can be performed in $\frac{n!}{(n-m)!}$ ways. Hence, we must have

$$x \cdot m! = \frac{n!}{(n-m)!}$$

In other words,

$$[\text{Stage 1}] \cdot [\text{Stage 2}] = [\text{Case II}].$$

Solving for x then yields our result.

Computational Hint: Note that

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

So, for example,

$$\binom{5}{3} = \binom{5}{2} \quad \text{and} \quad \binom{9}{4} = \binom{9}{5}.$$

Also note that if n is large, brute force computation of all three factorial terms of $\binom{n}{m}$ can be a challenge. To overcome this, note that

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n(n-1)(n-2)\cdots(n-m+1)}{m!}.$$

with only m terms in both the numerator and the denominator. Hence,

$$\binom{10}{2} = \frac{10!}{2!8!} = \frac{10 \cdot 9}{2!} = 45$$

and

$$\binom{100}{97} = \binom{100}{3} = \frac{100!}{3!97!} = \frac{100 \cdot 99 \cdot 98}{3!} = 161700.$$

Example: A consumer survey asks: *Select exactly two vegetables from the following list:*

○ *peas* ○ *carrots* ○ *lima beans* ○ *corn*

The total number of ways you can select two items from the four is $\binom{4}{2} = 6$.

IV. *Sampling with replacement and without ordering.*

This case is perhaps the most difficult to perceive as a sampling discipline, yet it has important applications. A total of m balls are drawn sequentially from the urn. Each ball is replaced after being drawn. However, the sequence of the draws is ignored. Therefore, drawing $3, 3, 1, 2$ is the same as drawing $1, 3, 2, 3$. The number of ways that m balls can be drawn from an urn containing n balls in this fashion is

$$\binom{m+n-1}{m}.$$

Let's prove that this formula works. First, note that to describe a possible outcome from this scenario, we only need to know how many times each ball was drawn. For example, to describe $3, 3, 1, 2$ or $1, 3, 2, 3$ we only need to know that Ball 1 was drawn once, Ball 2 was drawn once, and Ball 3 was drawn twice. This arrangement can be represented as follows:

	1	2	3
1 2 3 3	$\surd$	$\surd$	$\surd\surd$

For another example using this scheme, the arrangements $2, 2, 2, 3$ or $2, 3, 2, 2$ and their counterparts would be represented as

	1	2	3
2 2 2 3		$\surd\surd\surd$	$\surd$

Now notice that any of these outcomes can be represented by only specifying the arrangement of the four check marks ($\surd$) among the *inner* two vertical lines ($|$). For example,

	1	2	3	*Coded Representation*
1 2 3 3	$\surd$	$\surd$	$\surd\surd$	$\surd\|\surd\|\surd\surd$
2 2 2 3		$\surd\surd\surd$	$\surd$	$\|\surd\surd\surd\|\surd$

So, for example,

$$||\surd\surd\surd\surd$$

would represent $3, 3, 3, 3$.

To count the total number of possible outcomes, note that our coded representation has six positions

$$\underline{?}\ \ \underline{?}\ \ \underline{?}\ \ \underline{?}\ \ \underline{?}\ \ \underline{?}$$

Any four of these can be selected for the check marks ($\surd$) and the remaining two will be filled with the two vertical lines ($|$). The number of ways we can select 4 positions from 6 positions is $\binom{6}{4} = 15$. Therefore, there are 15 possible ways we can select 4 balls from an urn containing 3 balls if we allow replacement and disregard ordering.

In general, if there are n balls and m drawings, we will need m check marks ($\surd$). To represent all possible outcomes, our table will have n columns.

	1	2	$\cdots$	n	*Coded Representation*
$k_1\ k_2\ \cdots\ k_n$	$\surd\surd$	$\surd$	$\cdots$	$\surd$	$\surd\surd\vert\surd\vert\cdots\vert\surd$

Therefore, our coded representation will have m check marks and $n - 1$ vertical lines. So we need $(n - 1) + m$ positions for both check marks and vertical lines

$$\underbrace{\underline{?}\ \ \underline{?}\ \ \underline{?}\ \ \cdots\ \ \underline{?}}_{\substack{(n-1)+m \\ \text{positions}}}$$

The number of ways we can select m positions for check marks from $(n - 1) + m$ positions is

$$\binom{n+m-1}{m}$$

Example: Mary wants to buy a dozen cookies. She likes peanut butter cookies, chocolate chip cookies and sugar cookies. The baker will allow her to select any number of each type to make a dozen. In this case, our table has three columns with twelve check marks. So the following arrangement:

Peanut Butter	Chocolate Chip	Sugar	*Coded Representation*
$\surd\surd$	$\surd\surd\surd$	$\surd\surd\surd\surd\surd\surd\surd$	$\surd\surd\vert\ \surd\surd\surd\vert\surd\surd\surd\surd\surd\surd\surd$

would represent the outcome when Mary selects 2 peanut butter, 3 chocolate chip and 7 sugar cookies. In this case, we are arranging 14 symbols in a row:

2 vertical lines and 12 check marks. So the number of possible different boxes of one dozen cookies is

$$\binom{12+3-1}{12} = \binom{14}{12} = \binom{14}{2} = \frac{14 \cdot 13}{2 \cdot 1} = 91.$$

Permutation of n items distinguished by groups

An extension of case III is the situation where one has n balls with n_1 of them painted with color number 1, n_2 of them painted with color number 2,…,n_k of them painted with color number k. We have

$$\sum_{i=1}^{k} n_i = n.$$

If the balls are arranged in a row, the number of distinguishable arrangements is given by

$$\frac{n!}{n_1! n_2! \cdots n_k!}.$$

This quantity is a generalization of the binomial coefficient and is called the **multinomial coefficient**. Note that when $k = 2$, this situation is the same as case III.

Example: Consider the seven characters, A, A, A, B, B, C, C. The number of distinct arrangements of these seven letters in a row is

$$\frac{7!}{3!2!2!} = 210.$$

Occupancy

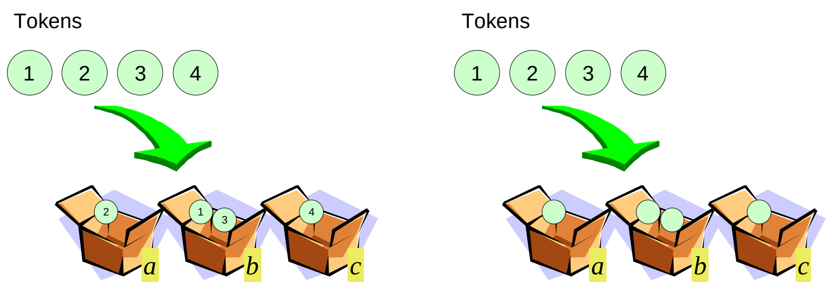
The Token Model: In sampling problems, we think of n balls placed in an urn from which m balls are drawn. A direct analogy can be made with so-called *occupancy problems*. In such models, we are given n boxes marked 1 to n, and m tokens numbered 1 to m. There are four disciplines for placing the tokens in the boxes. Each of these four scenarios correspond to one of the four sampling cases.

The reason for introducing occupancy problems becomes clear as one encounters different counting problems. Some may be easier to conceptualize as occupancy problems rather than sampling problems. Nevertheless, the number of distinguishable ways in which the tokens may occupy the boxes is computed in precisely the same manner as in the sampling models.

The connection between the two formulations is easy to remember if you note the following relationships:

Sampling Problems	**Occupancy Problems**
ball	box
number of drawing	number on token
j^{th} draw yields ball k	j^{th} token is placed in box k

For example, suppose we have a sampling problem with 3 balls (labelled a, b, c) being drawn 4 times. This would correspond to an occupancy problem with 3 boxes and 4 tokens (See Figure 2.2). Moreover, the following outcomes from each



(a) Token labels observed

(b) Token labels not observed

Figure 2.2: The occupancy problem

model would be equivalent:

Sampling	
Drawing	Ball
1	b
2	a
3	b
4	c

Occupancy	
Box	Token(s)
a	2
b	1,3
c	4

Sampling cases I to IV can now be viewed as

I*. Each box may contain any number of tokens and the labels on the tokens are observed.

II*. No box may contain more than one token and the labels on the tokens are observed.

III*. No box may contain more than one token and the labels on the tokens are not observed.

IV*. Each box may contain any number of tokens and the labels on the tokens are not observed.

The scheme with check marks and vertical bars that we used for Case IV is now readily apparent. Each column of the table represents a box, and each check mark represents a token.

Example: Suppose a school has five classrooms (with different room numbers) and three teachers, Ms. Smith, Mr. Thomas and Ms. Warren. For some courses, more than one teacher might be assigned to the same classroom.

I*. If more than one teacher can be assigned to a classroom, and it matters which teacher teaches in a particular room, then the number of ways you can assign teachers to classrooms is

$$n^m = 5^3 = 125.$$

II*. If only one teacher can be assigned to a classroom, and it matters which teacher teaches in a particular room, then the number of ways you can assign teachers to classrooms is

$$n(n-1)\cdots(n-m+1) = 5 \cdot 4 \cdot 3 \cdot = 60.$$

III*. If only one teacher can be assigned to a classroom, and it does not matter which teacher teaches in a particular room, then the number of ways you can assign teachers to classrooms is

$$\binom{n}{m} = \binom{5}{3} = 10.$$

IV*. If more than one teacher can be assigned to a classroom, and it does not matter which teacher teaches in a particular room, then the number of ways you can assign teachers to classrooms is

$$\binom{m+n-1}{m} = \binom{3+5-1}{3} = 35.$$

Self-Test Exercises for Chapter 2

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S2.1 *Bill's Donut Shop* bakes 10 varieties of doughnuts. How many different boxes of 12 doughnuts can a customer select from the 10 varieties? *Note: Assume that Bill's Shop has at least 12 of each type of donut.*

(A) 12!

(B) $\binom{21}{12}$

(C) 10!

(D) 12^{10}

(E) none of the above.

S2.2 A teacher has a class of 30 students. Each day, she selects a team of 6 students to present their homework solutions to the rest of the class. How many different teams of 6 students can she select?

(A) $\binom{30}{6}$

(B) 30^6

(C) $30!/24!$

(D) 6^{30}

(E) none of the above

S2.3 A box is partitioned into 3 numbered cells. Two indistinguishable balls are dropped into the box. Each cell may contain *at most* one ball. The total number of distinguishable arrangements of the balls in the numbered cells is

(A) 1

(B) 3

(C) 6

(D) 9

(E) none of the above.

S2.4 A professor wants to use three colors (red, green and blue) to draw an equilateral triangle on a sheet of paper. Each edge of the triangle is a single color, but any color can be used more than once. How many different triangles can be drawn?

(A) 3^3

(B) 3!

(C) $\binom{5}{3}$

(D) $\binom{5}{3} + 1$

(E) none of the above.

S2.5 A professor has an equilateral triangle cut from a sheet of cardboard. He wants to use three colors (red, green and blue) to color the edges. Each edge of the triangle is a single color, but any color can be used more than once. How many different triangles can be produced?

(A) 3^3

(B) 3!

(C) $\binom{5}{3}$

(D) $\binom{5}{3} + 1$

(E) none of the above.

S2.6 Two white balls are dropped randomly into four labelled boxes, one at a time. A box may contain more than one ball. The number of distinguishable arrangements of the balls in the boxes is

(A) $\binom{4}{2}$

(B) $\binom{7}{2}$

(C) 4

(D) 4^2

(E) none of the above.

S2.7 One white ball and one black ball are dropped randomly into four labelled boxes, one at a time. A box may contain more than one ball. The number of distinguishable arrangements of the balls in the boxes is

(A) $\binom{4}{2}$

(B) $\binom{7}{2}$

(C) 4

(D) 4^2

(E) none of the above.

S2.8 A class has ten students. The teacher must select three students to form a committee. How many different possible committees can be formed?

(A) 10^3

(B) 3^{10}

(C) $10!/7!$

(D) $\binom{10}{3}$

(E) none of the above

S2.9 A student club has ten members. The club must elect a president, vice-president and treasurer from the membership. How many ways can these three positions be filled?

(A) 10^3

(B) 3^{10}

(C) $10!/7!$

(D) $\binom{10}{3}$

(E) none of the above

S2.10 A candy company is introducing a new, bite-sized, chocolate candy called "XYZ’s". Each candy is covered with a sugar coat that is colored a single color: either red, green, or blue. Also each candy is marked with a single letter: either “X,” “Y” or “Z’.

How many different possible candies are there?

(A) 9

(B) 3^3

(C) $\binom{9}{3}$

(D) 3!

(E) none of the above.

S2.11 An ice cream shop offers 20 different flavors of ice cream. Their menu says that their "Big Dish" contains five scoops of ice cream. You may order the five scoops in any way, ranging from five different flavors to five scoops of the same flavor. How many different "Big Dishes" are possible?

(A) 20^5

(B) $\binom{20}{5}$

(C) $\binom{24}{5}$

(D) 20!/5!

(E) none of the above

S2.12 There are 4 machines in a factory. If each is worker assigned to one machine, how many ways can 4 workers be assigned to the machines?

(A) 4^2

(B) 4!

(C) $\binom{7}{4}$

(D) 4^4

(E) none of the above

Questions for Chapter 2

2.1 A domino piece is marked by two numbers. The pieces are symmetrical so that the number pair is not ordered. How many different pieces can be made using the numbers $1, 2, \ldots, n$?

2.2 At a bridge table, the 52 cards are partitioned into four equal groups. How many different playing situations are there?

2.3 How many ways can a poker hand of 5 cards be drawn from a 52 card deck so that each card is a different number or face (i.e., different, ignoring suits)?

2.4 Throw three indistinguishable dice. How many distinguishable results of the throw are there?

2.5 From eight persons, how many committees of three members may be chosen? Suppose that among the eight persons, exactly four are women. How many three-person committees can be chosen so that at least one member is female?

2.6 A ballot lists 10 candidates. You may vote for any three of them. If you vote for exactly three candidates, how many different ways can you mark your ballot?

2.7 An election ballot lists 10 candidates. Each voter is allowed 4 votes. According to the "bullet" voting system, a voter must place 4 check marks on the ballot, and may assign more than one check mark to any candidate(s) up to a total of four marks. How many different ways can the ballot be marked?

2.8 A multiple choice test contains 10 questions with each question having 4 possible answers. How many different ways could a student answer the entire test?

2.9 (provided by Mark Chew) Jack Dukey enters the Dip-'n-Donut Shop to buy a dozen doughnuts for himself and his co-workers in the Parks Department. He notices that they have the following inventory of his favorite doughnuts: Bavarian Creme (21), Strawberry (30), Peanut Krunch (11) and Powdered Jelly (15).

(a) How many different ways can Jack order a dozen doughnuts? Assume that ordering 6 Bavarian Creme and 6 Peanut Krunch is the same as ordering 6 Peanut Krunch and 6 Bavarian Creme.

(b) If Jack wants 2 Peanut Krunch for himself, and his boss in the Parks Department wants 3 Bavarian Creme for his coffee break, how many different ways can Jack order a dozen doughnuts?

2.10 Consider a box partitioned into 3 cells. Two balls are dropped into the box, each independently of the other, in such a way that every cell has an equal chance of receiving a ball. A cell may contain more than one ball. If the balls are indistinguishable, list all possible arrangements.

Reanswer the question if no cell can have more than one ball.

2.11 A classroom in the Bell Elementary School contains 16 seats with four rows of four seats each. A class is composed of 4 girls and 12 boys. How many ways can the 4 girls be seated among the 12 boys if the students are distinguished only by sex?

2.12 Fifty students have rented 3 busses for a class trip. Each bus can hold 50 riders. How many different ways can the students sign up for busses?

2.13 (See also Question 2.6.) A ballot lists 10 candidates. You must pick three candidates and rank them according to your preference, A, B, or C with A being your most preferred candidate. How many different ways can you mark your ballot?

2.14 Prove the binomial theorem: For any real numbers x and y and integer $n \geq 1$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

2.15 You are given n balls of different colors. How many distinguishable ways can you. . .

(a) arrange all of them in a row?

(b) choose m of them simultaneously?

(c) divide them (perhaps unequally) between two identical urns?

(d) divide them (perhaps unequally) between two children (and each child cares about the colors he or she receives)?

3

PROBABILITY

The scientific imagination always restrains itself within the limits of probability.
– THOMAS HENRY HUXLEY, *Reflection* (c 1880)

Probability Spaces

In Chapter 1, we showed how you can use sample spaces Ω and events $A \subseteq \Omega$ to describe an experiment. In this chapter, we will define what we mean by the *probability of an event*, $P(A)$ and the mathematics necessary to consistently compute $P(A)$ for all events $A \subseteq \Omega$.

For infinite sample spaces, Ω, (for example, $\Omega = \mathbb{R}$) we often cannot reasonably compute $P(A)$ for all $A \subseteq \Omega$. In those cases, we restrict our definition of $P(A)$ to some smaller subset of events, $\mathfrak{F}$, that does not necessarily include *every* possible subset of Ω. The mathematical details are usually covered in advanced courses in probability theory.

Measuring sets

Definition 3.1. *The* **sample space**, Ω, *is the set of all possible outcomes of an experiment.*

Definition 3.2. *An* **event** *is any subset of the sample space.*

How would you measure a set?

- finite sets
- countable sets
- uncountable sets

Definition 3.3. *A* **probability measure**, $P(\cdot)$, *is a function, defined on subsets of* Ω, *that assigns to each event* E *a real number* $P(E)$ *such that the following three probability axioms hold:*

1. *$P(E) \geq 0$ for all events E*
2. *$P(\Omega) = 1$*
3. *If $E \cap F = \emptyset$ (i.e., if the events E and F are mutually exclusive), then*

$$P(E \cup F) = P(E) + P(F)$$

The three specifications in the above definition are often called the *Kolmogorov axioms* for probability measures. However, a great deal of probability theory has its foundations in a mathematical field called *measure theory* with important results from many mathematicians, including Emile Borel, Henri Lebesgue and Eugene Dynkin.

Note that every event $A \subseteq \Omega$ is assigned one and only one number $P(A)$. In section 3, we will show that $0 \leq P(A) \leq 1$ for every $A \subseteq \Omega$. In mathematical terms, $P(\cdot)$ is a *function* that assigns to each member of $\mathfrak{F}$, the set of events, a single number from the interval $[0, 1]$. That is

$$P : \mathfrak{F} \to [0, 1].$$

In calculus, you worked with real-valued functions $f(\cdot)$. That is,

$$f : \mathbb{R} \to \mathbb{R}.$$

Such functions assign each number in $\mathbb{R}$ a single number from $\mathbb{R}$. For example, $f(x) = x^3$, maps $x = 2$ into $f(2) = 8$.

Here, the function, $P(\cdot)$, is a function whose argument is a subset of Ω (namely an event) rather than a number. In mathematics we call functions that assign real numbers to subsets, *measures*. That's why we refer to $P(\cdot)$ as a *probability measure*.

Interpretation of probability

How do you assign the real numbers in Definition 3.3?

Logical probability This can be done when Ω is a finite set

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

and each of the outcomes in Ω is equally likely. That is

$$P(\{\omega_i\}) = \frac{1}{n} \quad \text{for } i = 1, \ldots, n$$

In that case, for any $A \subseteq \Omega$ assign

$$P(A) = \frac{|A|}{n}$$

For example, consider the outcome of rolling a fair six-sided die once, In that case

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$$

where

$$\omega_i = \text{the die shows } i \text{ pips}$$

for $i = 1, 2, \ldots, 6$. So we can assign, for example,

$$P(\{\omega_i\}) = \frac{1}{6} \quad \text{for } i = 1, \ldots, 6$$

$$P(\{\text{The die is even}\}) = P(\{\omega_2, \omega_4, \omega_6\}) = \frac{3}{6}$$

and, in general,

$$P(A) = \frac{|A|}{6} \quad \text{for any } A \subseteq \Omega$$

Logical probability is only useful when the mechanism producing the random outcomes has symmetry so that one could argue that there is no reason for one outcome to occur rather than other. Some examples where logical probability can be used are:

tossing a fair coin The coin is symmetrical and there is no reason for heads to turn up rather than tails.

rolling a fair die The die is symmetrical and there is no reason for one side to turn up rather than another.

drawing a card randomly from a deck of playing cards Each card is physically the same, and there is no reason for one card to be drawn rather than another.

As you can see, the notion of logical probability has limited applications.

Experimental probability This interpretation requires that the experiment can be identically reproduced over and over again, indefinitely. Each replicate is called an *experimental trial.* We also must be able to assume that the outcome of any one of the trials does not affect the probabilities associated with the outcomes of any of the other trials. As an example, suppose we examine transistors being produced by an assembly line, with each transistor an experimental trial. The outcome of one of our experimental trials is whether that transistor is either good or bad. Consider the sequence of outcomes for the repeated experiment, with

$$\begin{aligned} T_i &= 0 \quad \text{if transistor } i \text{ is bad} \\ T_i &= 1 \quad \text{if transistor } i \text{ is good} \end{aligned}$$

In this case, $\Omega = \{\text{bad}, \text{good}\}$. Let $A = \{\text{good}\}$. To assign $P(A)$ using logical probability, consider the infinite sequence of experimental outcomes

$$T_1, T_2, \ldots$$

and assign

$$P(A) = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} T_i}{n}$$

There are some important mathematical issues associated with defining $P(A)$ in this way. In particular, the type of limit that you defined in calculus cannot be used. In calculus, there always existed some K, determined in advance, such that for all $n > K$ the quantity

$$\frac{\sum_{i=1}^{n} T_i}{n}$$

was arbitrarily close to $P(A)$. That concept is not suitable here.

Suppose, for instance, that $P(A)$ equaled $\frac{1}{2}$. It is possible, although not likely, that you could examine billions of transistors and find that all of them are bad. In that case, $\frac{\sum_{i=1}^{n} T_i}{n}$ would be zero, and quite far away from the eventual limit of $\frac{1}{2}$ as n gets bigger than billions.

In fact, it can be proven that for a prespecified small number $\delta > 0$, there does exist some integer K such that for all $n > K$

$$\left| \frac{\sum_{i=1}^{n} T_i}{n} - P(A) \right| < \delta.$$

The value of K is, itself, random and cannot be predetermined. We can, however, say that $P(K < \infty) = 1$. The proof requires measure theory, and is typically covered in a graduate-level probability course.

Some examples where experimental probability can be used are:

quality assurance Specifying the probability that a single microprocessor from a semiconductor fabrication facility will be defective.

interarrival times for queues Stating the probability that the time between any two consecutive customers is greater than five minutes.

hitting a home run Specifying the probability that a baseball player will hit a home run on any attempt.

Subjective probability There are some cases where we would like to use probability theory, but it is impossible to argue that the experiment can be replicated. For example, consider the experiment of taking EAS305, with the outcome being the grade that you get in the course. Let

$$A = \{\text{you get an “A” in EAS305}\}$$

The procedure of determining the subjective probability of A, is to construct an equivalent lottery consisting of an urn, and an unlimited quantity of red and white balls. The decision maker is asked to place in the urn any number of red balls and white balls he or she desires, with the goal of making the event of drawing a red ball equivalent to getting an “A” in EAS305.

There is a mathematical theory (complete with axioms and deductions) regarding the existence of such equivalent lotteries.

Once the decision maker has decided how many balls of each type to place in the urn, we simply count the balls in the urn and assign

$$P(A) = \frac{\text{number of red balls in the urn}}{\text{total number of balls in the urn}}$$

Not surprisingly, subjective probability has a lot of critics. But sometimes it is the only way we can give meaning to the number $P(A)$.

Some examples where subjective probability can be used are:

weather Specifying the probability that it will be warm and sunny tomorrow.

auctions Stating the probability that painting by Monet will sell for one million dollars at an art auction.

football Specifying the probability that the Buffalo Bills will win the Super Bowl next year.

Examples using logical probability

Example: A box contains three light bulbs, and one of them is bad. If two light bulbs are chosen at random from the box, what is the probability that one of the two bulbs is bad?

Solution 1: Assume that the bulbs are labelled A, B, and C and that light bulb A is bad. If the two bulbs are selected one at a time, there are $(3)(2) = 6$ possible outcomes of the experiment:

$$\begin{array}{cc} AB & BA \\ AC & CA \\ CB & BC \end{array}$$

Note that 4 of the 6 outcomes result in choosing A, the bad light bulb. If these 6 outcomes are equally likely (are they?), then the probability of selecting the bad bulb is $\frac{4}{6} = \frac{2}{3}$

Solution 2: If the two bulbs are selected simultaneously (i.e., without ordering), there are $\binom{3}{2} = 3$ possible outcomes of the experiment:

$$AB \quad AC \quad CB$$

and 2 of the 3 outcomes result in choosing bad bulb A. If these 3 outcomes are equally likely (are they?), then the probability of selecting the bad bulb is $\frac{2}{3}$

Example: A box contains *two* light bulbs, and one of them is bad. Suppose a first bulb is chosen, tested and then placed back in the box. Then a second bulb is chosen, and tested. Note that the same bulb might be selected twice. What is the probability that exactly one of the two tested bulbs is bad?

"Solution:" Suppose the two bulbs are labelled A and B and bulb A is the bad light bulb, There are three possible outcomes:

Both bulbs are good
Both bulbs are bad
Exactly one bulb is bad

So the probability that exactly one bulb is bad is $\frac{1}{3}$. (Actually, this solution is wrong. Why it is wrong?)

Relationships between English and set notation

- $P(\text{not } E) = P(E^{\mathsf{C}})$
- $P(E \text{ or } F) = P(E \cup F)$
- $P(E \text{ and } F) = P(E \cap F)$

Deductions from the axioms

Theorem 3.1. $P(A^{\mathsf{C}}) = 1 - P(A)$

Proof.

$$\begin{array}{rcll} P(\Omega) & = & 1 & \text{from Axiom 2} \\ P(A \cup A^c) & = & 1 & \text{since } \Omega = A \cup A^c \\ P(A) + P(A^c) & = & 1 & \text{since } A \cap A^c = \emptyset \\ P(A^c) & = & 1 - P(A) & \end{array}$$

■

Theorem 3.2. $P(A) \leq 1$ *for all* $A \subseteq \Omega$

Proof.

$$\begin{array}{rcll} P(A) & = & 1 - P(A^c) & \text{from Theorem 3.1} \\ P(A) & \leq & 1 & \text{since } P(A^c) \geq 0 \text{ (Axiom 1)} \end{array}$$

■

Theorem 3.3. $A \subseteq B \Rightarrow P(A) \leq P(B)$

Proof.

$$\begin{array}{rcll} A \cup B & = & B & \text{since } A \subseteq B \\ (A \cup B) \cap (A \cup A^c) & = & B & \text{since } A \cup A^c = \Omega \\ A \cup (B \cap A^c) & = & B & \text{DeMorgan's law} \\ P(A \cup (B \cap A^c)) & = & P(B) & \text{See Question 3.31} \\ P(A) + P(B \cap A^c) & = & P(B) & A \cap (B \cap A^c) = \emptyset \\ P(A) & \leq & P(B) & \text{since } P(B \cap A^c) \geq 0 \text{ (Axiom 1)} \end{array}$$

■

Example: In a class of 100 students it is known that 80 of the students like hamburger. Consider the experiment of picking a student at random from the class of 100. Let A be the event that the selected student is one of those who likes hamburger. In this case $|A| = 80$ and $|\Omega| = 100$ and we can use logical probability to determine $P(A) = 80/100 = 0.80$.

Let B be the event that the selected student is one of those who likes broccoli. Suppose $|B|$ is unknown.

◇ Describe the set $A \cap B$.

Solution: $A \cap B$ is the event that the student selected likes hamburger *and* broccoli.

◇ Show that $A \cap B \subseteq A$.

Solution: Pick an element $\omega \in A \cap B$. That implies that $\omega \in A$ and $\omega \in B$. Hence, $\omega \in A$. Therefore, $A \cap B \subseteq A$.

◇ Even if $|B|$ is unknown we know that $P(A \cap B) \leq 0.80$. Why?

Solution: For any events E and F, $E \subseteq F$ implies $P(E) \leq P(F)$. Since $A \cap B \subseteq A$, we have $P(A \cap B) \leq P(A) = 0.80$.

Theorem 3.4. *For any finite number of pairwise disjoint sets $A_1, A_2, \ldots, A_n$ we have*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

Proof. (By induction)

Is it true for $n = 2$?

Yes, that's Axiom 3.

Assume it's true for $n = k$, is it true for $n = k + 1$?

Assume for k pairwise disjoint sets $A_1, A_2, \ldots, A_k$ we have

$$P(A_1 \cup \cdots \cup A_k) = P(A_1) + \cdots + P(A_k)$$

Let A_{k+1} be an event such that $A_i \cap A_{k+1} = \emptyset$ for all $i = 1, \ldots, k$. Then

$$\begin{aligned}
(A_1 \cup \cdots \cup A_k) \cap A_{k+1} &= \emptyset \\
&\quad \text{since } A_i \cap A_{k+1} = \emptyset \text{ for all } i \\
P((A_1 \cup \cdots \cup A_k) \cup A_{k+1}) &= P(A_1 \cup \cdots \cup A_k) + P(A_{k+1}) \\
&\quad \text{since } (A_1 \cup \cdots \cup A_k) \cap A_{k+1} = \emptyset \\
P((A_1 \cup \cdots \cup A_k) \cup A_{k+1}) &= (P(A_1) + \ldots + P(A_k)) + P(A_{k+1}) \\
&\quad \text{from induction hypothesis} \\
P(A_1 \cup \cdots \cup A_k \cup A_{k+1}) &= P(A_1) + \ldots + P(A_{k+1}) \\
&\quad \text{Associative law}
\end{aligned}$$

■

Theorem 3.5. *For any two events A and B*

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

Proof.

$$\begin{aligned}
A \cup B &= A \cup (A^c \cap B) \\
&\quad \text{Distributive law for sets} \\
\text{and} & \\
B &= (A \cap B) \cup (A^c \cap B) \\
&\quad \text{Distributive law for sets} \\
\text{Therefore} & \\
P(A \cup B) &= P(A) + P(A^c \cap B) \\
&\quad \text{from Axiom 3} \\
P(B) &= P(A \cap B) + P(A^c \cap B) \\
&\quad \text{from Axiom 3}
\end{aligned}$$

Subtracting the second equation from the first produces the result. ■

Example: A hydraulic piston can have only two types of defects:

1. a surface defect, and/or
2. a structural defect

A single piston may have one, both or neither defect. From quality control records we see that

- The probability that a piston has a surface defect is 0.5.

- The probability that a piston has a structural defect is 0.4.
- The probability that a piston has some type of defect is 0.6

To find the probability that a piston has *both* a structural and a surface defect, first define the following events:

$$\begin{aligned} S &= \{\text{a piston has a surface defect}\} \\ T &= \{\text{a piston has a structural defect}\} \end{aligned}$$

We want to find $P(S \cap T)$.

We know that

$$\begin{aligned} P(S) &= 0.5 \\ P(T) &= 0.4 \\ P(S \cup T) &= 0.6 \end{aligned}$$

Therefore, using Theorem 3.5,

$$\begin{aligned} P(S \cup T) &= P(S) + P(T) - P(S \cap T) \\ 0.6 &= 0.4 + 0.5 - P(S \cap T) \\ P(S \cap T) &= 0.4 + 0.5 - 0.6 = 0.3 \end{aligned}$$

Theorem 3.6. $P(\emptyset) = 0$

Proof.

$$\begin{aligned} P(\emptyset) &= 1 - P(\emptyset^c) && \text{from Theorem 3.1} \\ P(\emptyset) &= 1 - P(\Omega) && \text{since } \emptyset^c = \Omega \\ P(\emptyset) &= 1 - 1 && \text{from Axiom 2} \\ P(\emptyset) &= 0 \end{aligned}$$

■

In order to allow the extension of Axiom (3) in Definition 3.3 to apply to a *countable* collection of disjoint sets (rather than just a *finite* collection), the third probability axiom is replaced by

3.* If $A_1, A_2, \ldots$ is a countable sequence of disjoint sets, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Example: Consider the experiment of counting the number of accidents in a factory in a month. Let E_i be the event that exactly i accidents occur in a month with $i = 0, 1, 2, \ldots$. Suppose that

$$P(E_i) = \frac{e^{-1}}{i!} \quad \text{for } i = 0, 1, 2, \ldots$$

and we want to find the probability that the factory has at least one accident in a month.

That is, we want to compute $P\left(\cup_{i=1}^{\infty} E_i\right)$.

Since $\sum_{i=0}^{\infty} \frac{\alpha^i}{i!} = e^{\alpha}$ for any α, we have

$$\begin{aligned}
\sum_{i=0}^{\infty} P(E_i) &= \sum_{i=0}^{\infty} \frac{e^{-1}}{i!} \\
&= e^{-1} \sum_{i=0}^{\infty} \frac{1}{i!} \\
&= e^{-1} e^{1} = 1
\end{aligned}$$

Using axiom 3.* and the fact that $E_i \cap E_j = \emptyset$ for any $i \neq j$, we get

$$\begin{aligned}
P\left(\bigcup_{i=0}^{\infty} E_i\right) &= \sum_{i=0}^{\infty} P(E_i) = 1 \\
P\left(E_0 \cup \left[\bigcup_{i=1}^{\infty} E_i\right]\right) &= 1 \\
P(E_0) + P\left(\bigcup_{i=1}^{\infty} E_i\right) &= 1 \qquad \text{Since } E_0 \cap \left[\cup_{i=1}^{\infty} E_i\right] = \emptyset \\
P\left(\bigcup_{i=1}^{\infty} E_i\right) &= 1 - P(E_0) = 1 - e^{-1}
\end{aligned}$$

THE BIRTHDAY PROBLEM

Suppose k randomly selected individuals are seated in a room. What is the probability that at least two of them will have the same birthday? How large must k be so that the probability is 0.50 that at least two of them will have the same birthday?

To solve this problem, we will make two assumptions:

1. there are $n = 365$ days in a year
2. each person is assigned a birthday with any of the $n = 365$ days being equally likely.

To illustrate the issues, we'll first solve the problem for $k = 3$ and $n = 365$. Later, we'll extend our solution for any k, and any n.

Define the events

$$\begin{aligned} E &= \{\text{at least 2 birthdays are the same}\} \\ F &= \{\text{all of the birthdays are different}\} \end{aligned}$$

and notice that $E = F^c$. The question asks us to compute $P(E)$. But $P(E) = 1 - P(F)$ and it turns out that $P(F)$ is easier to compute.

To compute $P(F)$, we'll use an approach that is common for most problems involving logical probability:

Step 1 Count the number of equally likely outcomes in Ω.

Step 2 Count the number of elements in $F \subseteq \Omega$.

Step 3 Divide the number obtained in Step 2 by the number in Step 1.

It sounds simple, that is until you begin identifying and counting equally likely elements in Ω. Some textbooks[1] say something like "If we imagine the k people to be ordered, there are $(365)^k$ corresponding permutations of their k birthdays." Why *ordered*? Why not?

Suppose $k = 3$. Let's suppose the individuals are ordered. Then every element of the sample space can be written as an ordered triple $\omega = (d_1, d_2, d_3)$, where d_1, d_2 and d_3 are the birthdays of the three individuals. We can begin to write down all of the elements of Ω as follows:

$$\begin{aligned} \Omega = \{&(\texttt{Jan01,Jan01,Jan01}), \\ &(\texttt{Jan01,Jan01,Jan02}), \\ &(\texttt{Jan01,Jan01,Jan03}), \\ &(\texttt{Jan01,Jan01,Jan04}), \ldots\} \end{aligned}$$

The assignment of birthdays with ordered individuals can be represented as a sampling problem. The number of ways we can draw three balls from 365 *with* replacement and *with* ordering is $(365)^3$.

[1]For just one example of many, see Richard J. Larsen and Morris Marx, *An introduction to probability and its applications*, Prentice-Hall, 1985.

Continuing with this reasoning, the number of ways that we can assign birthdays to individuals with everyone getting a *different* birthday is equal to the number of ways we can draw three balls from 365 *without* replacement and *with* ordering. And that number is $(365)!/(365-3)!$.

So for $k = 3$ we get

$$P(E) = 1 - P(F) = 1 - \frac{(365)!/(365-3)!}{(365)^3}$$

This turns out to be the correct answer. Furthermore, if we perform the same computations for any n and $k \leq n$, the general correct answer is

$$P(E) = 1 - P(F) = 1 - \frac{(n)!/(n-k)!}{(n)^k}$$

To answer the question "how large must k be?," the following table lists $P(E)$ for several different values of k with $n = 365$:

k	$P(E)$
15	0.253
22	0.475
23	0.507
40	0.891
50	0.970
70	0.999

and we see that we need at least 23 persons to insure that $P(E)$ is at least 0.50.

But why would we get the *wrong* answer if we don't order the k individuals? Let's see what this would mean for $k = 3$.

Identifying the elements of Ω without ordering the individuals is equivalent to listing all of the ways we can draw 3 balls from an urn containing 365 balls, *with* replacement and *without* ordering. We can do this by creating a table with 365 columns (one for each calendar day) and asking the question, "how many ways can we arrange 3 check marks among the 365 columns?"

Jan01	Jan02	Jan03	...
$\surd\surd\surd$			
$\surd\surd$	$\surd$		
$\surd\surd$		$\surd$	
...etc.			

Using this approach, each element of Ω is distinguishable by the *number of times* someone in the group is assigned the birthday `Jan01`, `Jan02`, etc., not *who in the group* is assigned `Jan01`, `Jan02`, etc.. There is nothing wrong with identifying the elements of the sample space Ω in this way. But we would wrong to assume that such elements are equally likely.

We can again compute $P(E)$, but this time, using the *unordered individual* assumption. In this case, the total number of elements in Ω is

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

and the number of elements in F is

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

So that

$$P(E) = 1 - P(F) = 1 - \frac{(n)!(n-1)!}{(n+k-1)!(n-k)!}$$

To show why the *unordered individual assumption* is wrong, let's use all of our formulas to solve a slightly modified version of the birthday problem:

The birthday problem on the planet Yak

There are two aliens from outer space in a room. They are from a planet called Yak. Yak has only two days in its calendar year. The names of the days are "Heads" and "Tails".

Once again, assume that birthdays are randomly assigned, with each day equally likely. Assigning birthdays at random to the aliens is the same as each alien tossing a fair coin to determine it's birthday. So we can use our formulas with values $n = 2$ and $k = 2$.

Under the *ordered individual model* there are four elements in Ω, that is

$$\Omega = \{HH, HT, TH, TT\},$$

and

$$P(E) = 1 - \frac{(2)!/(2-2)!}{(2)^2} = \frac{1}{2}$$

$$\begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array}$$

Figure 3.1: Sample Space for Rolling a Die Twice

Under the *unordered individual model* there are three elements in Ω,

$$\Omega = \{2\text{H}, 2\text{T}, 1\text{H}1\text{T}\}$$

that is, "two heads", "two tails" and "one head one tail." And the corresponding computation for $P(E)$ yields

$$P(E) = 1 - P(F) = 1 - \frac{(2)!(2-1)!}{(2+2-1)!(2-2)!} = \frac{2}{3}$$

And we see that the second (and faulty) formulation results in the same mistake that Jean d'Alembert[2] made in 1754.

Moral of the story: Sampling Case IV (*with replacement and without ordering*) usually cannot be used to enumerate equally likely outcomes resulting from multiple, repeated random assignments.[3]

Conditional Probability and Independence

Conditional probability

Example: Roll a die twice. The sample space can be represented as in Figure 3.1. Define the events

$$\begin{aligned} A &= \{\text{first roll is less than or equal to the second roll}\} \\ B &= \{\text{the sum of the two rolls is 4}\} \end{aligned}$$

[2]Jean d'Alembert *Encyclopédie ou dictionnaire raisonné des sciences, des arts and des métiers,* Volume 4, 1754.

[3]For an important exception, see the work of Bose and Einstein in statistical mechanics.

Since each of the outcomes is equally likely, we can use logical probability to get $P(B) = 1/12$. Suppose you are told that the first roll is less than or equal to the second. With this additional information, what is the probability that the two rolls sum to 4?

Definition 3.4. *The* **probability of an event** A **given the event** B*, denoted* $P(A \mid B)$*, is the value that solves the equation*

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

Note: If $P(B) > 0$, this is equivalent to

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Notation: Sometimes we write $P(A \cap B) = P(AB)$.

Theorem 3.7. *For any events* $A_1, A_2, \ldots, A_n$

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2) \cdots P(A_n \mid A_1 A_2 \cdots A_{n-1})$$

$P(\cdot \mid B)$ is a probability measure

Remember that $P(\cdot)$ is a probability measure if and only if

1. $P(E) \geq 0$ for all events E
2. $P(\Omega) = 1$
3. If $E \cap F = \varnothing$ then $P(E \cup F) = P(E) + P(F)$

Suppose $P(B) > 0$ and consider $P(\cdot \mid B)$

- For any event $E \subseteq \Omega$

$$\begin{aligned} P(E \mid B) &= \frac{P(E \cap B)}{P(B)} \\ &= \frac{[\text{something} \geq 0]}{[\text{something} > 0]} \geq 0 \end{aligned}$$

Therefore $P(E \mid B) \geq 0$ for all events E

- Also, we have

$$\begin{aligned} P(\Omega \mid B) &= \frac{P(\Omega \cap B)}{P(B)} \\ &= \frac{P(B)}{P(B)} = 1 \end{aligned}$$

  Therefore $P(\Omega \mid B) = 1$

- And for any two events E and F with $E \cap F = \emptyset$

$$\begin{aligned} P(E \cup F \mid B) &= \frac{P((E \cup F) \cap B)}{P(B)} \\ &= \frac{P((E \cap B) \cup (F \cap B)}{P(B)} \\ &= \frac{P(E \cap B) + P(F \cap B)}{P(B)} \\ &= \frac{P(E \cap B)}{P(B)} + \frac{P(F \cap B)}{P(B)} \\ &= P(E \mid B) + P(F \mid B) \end{aligned}$$

  Therefore if $E \cap F = \emptyset$ then $P(E \cup F \mid B) = P(E \mid B) + P(F \mid B)$

- Hence $P(\cdot \mid B)$ is a probability measure.
- We can think of $P(\cdot \mid B)$ as the updated version of $P(\cdot)$ given B.
- In fact, every event $A \subseteq \Omega$ produces an updated probability measure $P(\cdot \mid A)$.
- When interpreting $P(E \mid A)$ we are asking the "what if" question, "what is the probability of E if the event A occurs."
- We can evaluate and use $P(E \mid A)$ and $P(E \mid B)$ simultaneously for different events, A and B.

Independence

Definition 3.5. *The events A and B are* **independent** *if and only if*

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

Note: If $P(B) > 0$ this is equivalent to saying that A and B are independent if and only if

$$P(A \mid B) = P(A).$$

Theorem 3.8. *If A and B are independent events then*

- *A and B^c are independent events,*
- *A^c and B are independent events, and*
- *A^c and B^c are independent events.*

Proof. Suppose A and B are independent events. Then

$$\begin{aligned}
P(A \cap (B \cup B^c)) &= P(A) \\
P((A \cap B) \cup (A \cap B^c)) &= P(A) \\
P(A \cap B) + P(A \cap B^c) &= P(A) \\
P(A)P(B) + P(A \cap B^c) &= P(A) \\
P(A \cap B^c) &= P(A) - P(A)P(B) \\
P(A \cap B^c) &= P(A)(1 - P(B)) \\
P(A \cap B^c) &= P(A)P(B^c)
\end{aligned}$$

Therefore A and B^c are independent events.

The other two results follow by interchanging the roles of A and B. ■

Bayes' theorem

Definition 3.6. *A* **complete set of alternatives** *is a collection of events*

$$B_1, B_2, \ldots, B_n$$

such that

1. *$B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$, and*
2. *$B_i \cap B_j = \emptyset$ for any $i \neq j$.*

Theorem 3.9. (Law of Total Probability) *Given an event A and a complete set of alternatives $B_1, B_2, \ldots, B_n$ then*

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

Example: (from Meyer[4]) Suppose that among six bolts, two are shorter than a specified length. If two bolts are chosen at random, what is the probability that the two short bolts are picked? Let

$$A_i = \{\text{the } i\text{th chosen bolt is short}\}.$$

The proper solution is obtained by writing

$$P(A_1 \cap A_2) = P(A_2 \,|\, A_1)P(A_1) = \tfrac{1}{5} \cdot \tfrac{2}{6} = \tfrac{1}{15}.$$

The common but *incorrect* solution is obtained by writing

$$P(A_1 \cap A_2) = P(A_2)P(A_1) = \tfrac{1}{5} \cdot \tfrac{2}{6} = \tfrac{1}{15}.$$

Although the answer is correct, the value of $P(A_2)$ is not $\frac{1}{5}$. But the value of $P(A_2 \,|\, A_1)$ is $\frac{1}{5}$.

To evaluate $P(A_2)$ properly, we note that $\{A_1, A_1^c\}$ is a complete set of alternatives. We can the use the law of total probability, and write

$$\begin{aligned} P(A_2) &= P(A_1)P(A_2 \,|\, A_1) + P(A_1^c)P(A_2 \,|\, A_1^c) \\ &= \tfrac{1}{5} \cdot \tfrac{2}{6} + \tfrac{2}{5} \cdot \tfrac{4}{6} = \tfrac{1}{3}. \end{aligned}$$

Theorem 3.10. **(Bayes' Theorem)** *Let $B_1, B_2, \ldots, B_n$ be a complete set of alternatives, then*

$$P(B_k \,|\, A) = \frac{P(A \,|\, B_k)P(B_k)}{\sum_{i=1}^{n} P(A \,|\, B_i)P(B_i)}$$

Example: Thousands of little snails live in the *Snail Garden* between the University at Buffalo Natural Sciences Complex and Talbert Hall. A concrete stairway divides the garden into two sections.

When it rains, the snails bravely travel from one section of the garden to the other by crawling across the steps of the concrete stairway. It is a treacherous journey. Many of the snails are crushed by the feet of unthinking students and faculty walking up and down the stairway.

[4]Meyer, P., *Introductory probability theory and statistical applications,* Addison-Wesley, Reading MA, 1965.

There are two types of snails, *smart* snails and *not-so-smart* snails. A *smart* snail crawls along the vertical portion of a step to avoid being crushed. A *not-so-smart* snail crawls on the horizontal part of a step.

Suppose that half of the snails are *smart* snails, and the other half are *not-so-smart* snails. The probability that a *smart* snail will be crushed is 0.10. The probability that a *not-so-smart* snail will be crushed is 0.40. The event that any one snail will be crushed is independent from all others.

- ◇ To find the probability that a single snail (chosen at random) will successfully cross the concrete steps uncrushed define the events

$$\begin{aligned} S &= \{\text{The snail is a } \textit{smart} \text{ snail}\} \\ S^c &= \{\text{The snail is a } \textit{not-so-smart} \text{ snail}\} \\ A &= \{\text{The snail is successful}\} \end{aligned}$$

  Note that $S \cup S^c = \Omega$ and $S \cap S^c = \emptyset$. Therefore, using the Law of Total Probability

$$\begin{aligned} P(A) &= P(A \mid S)P(S) + P(A \mid S^c)P(S^c) \\ &= (0.90)(0.50) + (0.60)(0.50) = 0.75 \end{aligned}$$

- ◇ Suppose that after the end of a rainstorm, you select, at random, a snail that has successfully crossed the concrete steps. What is the probability that it is a *smart* snail? In this case, we want to find

$$\begin{aligned} P(S \mid A) &= \frac{P(A \mid S)P(S)}{P(A \mid S)P(S) + P(A \mid S^c)P(S^c)} \\ &= \frac{(0.90)(0.50)}{(0.90)(0.50) + (0.60)(0.50)} = 0.60 \end{aligned}$$

  The probability of selecting a smart snail *before* the rainstorm was 0.50. The probability *after* the rainstorm is 0.60.

  What is the probability of finding a smart snails after *two* rainstorms?

Independence of more than two events

The definition of independence can be extended to more than two events.

Definition 3.7. *A finite collection of events $A_1, A_2, \ldots, A_n$ are* **independent** *if and only if*

$$P(A_{k_1} \cap \cdots \cap A_{k_j}) = P(A_{k_1}) \cdots P(A_{k_j})$$

for all $2 \leq j \leq n$ and all $1 \leq k_1 < \cdots < k_j \leq n$.

Note: For $n = 3$, the events A_1, A_2, A_3 are independent if and only if

$$P(A_1 \cap A_2) = P(A_1)P(A_2) \tag{1}$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3) \tag{2}$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3) \tag{3}$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \tag{4}$$

All of the above must hold for A_1, A_2, A_3 to be independent. If only the first three conditions hold, then we say that A_1, A_2, A_3 are **pairwise independent.**

Three events A, B and C can be pairwise independent, but not independent, as shown by the following example:

Example: Toss a fair coin twice, and assign probability $\frac{1}{4}$ to each of the outcomes HH, HT, TH, TT. Let

$$\begin{aligned} A &= \{\text{first toss is a head}\} \\ B &= \{\text{second toss is a head}\} \\ C &= \{\text{first toss equals second toss}\} \end{aligned}$$

Then

$$\begin{aligned} P(A \cap B) &= \tfrac{1}{4} = P(A)P(B) \\ P(A \cap C) &= \tfrac{1}{4} = P(A)P(C) \\ P(B \cap C) &= \tfrac{1}{4} = P(B)P(C) \\ P(A \cap B \cap C) &= \tfrac{1}{4} \neq P(A)P(B)P(C) \end{aligned}$$

Hence A, B, C are pairwise independent but not independent.

It is also possible to have three events A_1, A_2 and A_3 where equation (4) holds, but equations (1), (2) and (3) all fail. This is demonstrated by the following example:

Example: Suppose that two dice are tossed, and let Ω be all ordered pairs (i, j) for $i, j = 1, 2, \ldots, 6$ with probability of $\frac{1}{36}$ assigned to each point. Let

$$\begin{aligned} A &= \{\text{second die is 1, 2 or 5}\} \\ B &= \{\text{second die is 4, 5 or 6}\} \\ C &= \{\text{the sum of the faces is 9}\} \end{aligned}$$

Then

$$\begin{aligned} P(A \cap B) = &\ \tfrac{1}{6} &\neq P(A)P(B) = \tfrac{1}{4} \\ P(A \cap C) = &\ \tfrac{1}{36} &\neq P(A)P(C) = \tfrac{1}{18} \\ P(B \cap C) = &\ \tfrac{1}{12} &\neq P(B)P(C) = \tfrac{1}{18} \\ P(A \cap B \cap C) = &\ \tfrac{1}{36} &= P(A)P(B)P(C) \end{aligned}$$

Example: Suppose you roll a fair die repeatedly until a *six* appears. Assume all the outcomes from each roll are independent. Let

$$\begin{aligned} S_k &= \{\text{toss } k \text{ is a six}\} \\ A_n &= \{\text{the first six occurs on toss } n\} \end{aligned}$$

Then

$$P(S_k) = \tfrac{1}{6} \quad P(S_k^c) = \tfrac{5}{6}$$

for $k = 1, 2, \ldots$.

Furthermore,

$$A_n = S_1^c \cap S_2^c \cap \cdots \cap S_{n-1}^c \cap S_n$$

for $n = 1, 2, \ldots$.

$$\begin{aligned} P(A_n) &= P(S_1^c)P(S_2^c) \cdots P(S_{n-1}^c)P(S_n) \\ &= \left(\tfrac{5}{6}\right)^{n-1} \tfrac{1}{6} \end{aligned}$$

for $n = 1, 2, \ldots$.

We can also compute,

$$\begin{aligned} & P(\text{first 6 occurs after toss } k) \\ &= \sum_{n=k+1}^{\infty} P(A_n) \\ &= \sum_{n=k+1}^{\infty} \left(\tfrac{5}{6}\right)^{n-1} \tfrac{1}{6} \\ & \quad \text{letting } n' = n - (k+1) \\ &= \sum_{n'=0}^{\infty} \left(\tfrac{5}{6}\right)^{k+n'} \tfrac{1}{6} \end{aligned}$$

$$\begin{aligned}
&= \tfrac{1}{6}\left(\tfrac{5}{6}\right)^k \textstyle\sum_{n'=0}^{\infty}\left(\tfrac{5}{6}\right)^{n'} \\
&= \tfrac{1}{6}\left(\tfrac{5}{6}\right)^k \cdot 6 \\
&= \left(\tfrac{5}{6}\right)^k
\end{aligned}$$

for $k = 0, 1, 2, \ldots$. Which makes sense, since the first 6 occurs after toss k if and only if you get k non-6's in a row. So

$$\begin{aligned}
\{\text{first 6 occurs after toss } k\} &= S_1^c \cap S_2^c \cap \cdots \cap S_k^c \\
P(\text{first 6 occurs after toss } k) &= P(S_1^c)P(S_2^c)\cdots P(S_k^c) \\
&= \left(\tfrac{5}{6}\right)^k
\end{aligned}$$

for $k = 0, 1, 2, \ldots$.

Self-Test Exercises for Chapter 3

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S3.1 If $P(A) = 0.3$, $P(B) = 0.3$ and $P(A \mid B) = 0.3$ then

(A) A and B are mutually exclusive

(B) $P(A \cap B) = 0$

(C) A and B are independent

(D) all of the above

(E) none of the above.

S3.2 Let $P(A) = P(B) = 2p$ and let $P(A \cup B) = 3p$. Which of the following are always true?

(A) $p = 0$

(B) A and B are independent

(C) $P((A \cap B)^{\mathsf{c}}) = 1 - p$

(D) A and B are mutually exclusive

(E) none of the above.

S3.3 If $A \subseteq B$, then which of the following statements are always true?

(A) A and B are independent

(B) $P(A) > P(B)$

(C) A and B are mutually exclusive

(D) $A \cup B = A$

(E) none of the above

S3.4 If $P(A) = p > 0$, $P(B) = p > 0$ and $P(A \cup B) = 2p$ then $P(B \mid A \cup B)$ equals

(A) 1

(B) p

(C) $1/2$

(D) p^2

(E) none of the above

S3.5 If $P(A) = p$, $P(B) = p$ and $P(A \cap B) = p^2$ then

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

(B) $A = B$

(C) A and B are independent

(D) all of the above

(E) none of the above.

S3.6 Two fair coins are tossed once. One coin is painted red, and the other is painted blue. Let R denote the outcome of the red coin (Heads or Tails). Let B denote the outcome of the blue coin. Then

$$P(R = \text{Heads} \mid R \neq B)$$

is equal to

(A) $1/4$

(B) $1/3$

(C) $1/2$

(D) 1

(E) none of the above.

S3.7 Two indistinguishable six-sided fair dice are rolled. What is the probability that they both turn up the same value?

(A) $1/72$

(B) $1/36$

(C) $1/12$

(D) $1/6$

(E) none of the above.

S3.8 Suppose $P(A) = 0$. Then

(A) $A = \emptyset$

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

(C) $P(A \cup B) = 0$ for any $B \subseteq \Omega$.

(D) all of the above are true

(E) none of the above are true.

S3.9 Which of the following statements are *always* true?

(A) $P(\Omega) = 1$

(B) If $P(A) < P(B)$ then $A \subseteq B$

(C) If $P(A) = 0$ then $A = \emptyset$

(D) all of the above

(E) none of the above.

S3.10 Suppose $P(A) = p$, $P(B) = q$. If $A \subseteq B$ and $A \neq B$ then

(A) $p \geq q$

(B) $p \leq q$

(C) $p > q$

(D) $p < q$

(E) none of the above.

S3.11 There is one lock on a door and the key is among the six different ones you usually carry in your pocket. Someone places a seventh, useless key, in your pocket. What is the probability that the first key you try will open the door?

(A) $1/6$

(B) $1/2$

(C) $1/3$

(D) $2/3$

(E) none of the above.

S3.12 A multiple choice test has 12 questions, with each question having 5 possible answers. If a student randomly picks the answer to each question, what is the probability that the student will answer all questions correctly?

(A) $1/(12^5)$

(B) $1/(5^{12})$

(C) $1/\binom{12}{5}$

(D) $1/\binom{16}{5}$

(E) none of the above.

S3.13 Two balls are dropped randomly into four labelled boxes, one at a time. A box may contain more than one ball. The probability that both balls fall in the same box is

(A) $4/\binom{7}{2}$

(B) $1/4$

(C) $1/16$

(D) it depends on whether or not the balls are different colors

(E) none of the above.

S3.14 A student club has four officers, three of them are seniors and one of them is a junior. Three of the students will be picked at random to go on a trip. What is the probability that the junior stays home?

(A) $1/4$

(B) $1/\binom{6}{3}$

(C) $1/4!$

(D) $1/4^3$

(E) none of the above.

S3.15 Suppose that five tiles labelled with the letters **C, C, C, A, T** are placed in a container. If three tiles are selected from the container (without replacement), what is the probability that they can be arranged to spell the word **CAT**?

(A) 1

(B) $1/3$

(C) $1/\binom{5}{3}$

(D) $3/\binom{5}{2}$

(E) none of the above.

S3.16 Suppose A is any event. Then $P(A) + P(A^c)$ equals

(A) 0

(B) $P(A)$

(C) $P(A^c)$

(D) 1

(E) none of the above.

S3.17 If $A \subseteq B$ with $P(A) = P(B) > 0$ then

(A) $P(A^c \cap B) = 0$

(B) $A = B$

(C) $A^c \cap B = \emptyset$

(D) all of the above are true

(E) none of the above are true.

S3.18 Which of the following statements are *always* true?

(A) $P(\Omega) = 1$

(B) If $A \subseteq B$ then $P(A) \leq P(B)$

(C) $P(A \cap B^c) \leq P(A)$ for any event B

(D) all of the above are true.

(E) none of the above are true.

S3.19 Suppose $P(A) = 1$ and $P(B) = 0.25$. Which of the following statements are *always* true?

(A) A and B are independent

(B) $A \cap B = B$

(C) $B \subseteq A$

(D) all of the above are true.

(E) none of the above are true.

S3.20 A multiple choice test has 5 questions, with each question having 4 possible answers. If a student randomly picks the answer to each question, what is the probability that the student will answer all questions correctly?

(A) $1/(4^5)$

(B) $1/4$

(C) $1/\binom{5}{4}$

(D) $1/\binom{8}{4}$

(E) none of the above.

S3.21 A fair die is tossed twice. Given that a 2 appears on the first toss, the probability that the sum of the two tosses is 7 equals

(A) $1/36$

(B) $1/18$

(C) $1/9$

(D) $1/6$

(E) none of the above.

S3.22 If $A \cap B \cap C = \text{Ø}$ then which of the following must be true

(A) $A \cap B = \text{Ø}$

(B) $P(A \cap B) = 0$

(C) $P(A \cap B \cap C) = 0$

(D) all of the above are true

(E) none of the above.

S3.23 Four fair dice are rolled simultaneously. The probability that exactly two of them turn up with a value less than or equal to 4 is

(A) $24/81$

(B) $1/3$

(C) $4/9$

(D) $1/2$

(E) none of the above

S3.24 If $P(A) \leq P(A^{\mathsf{c}})$ then

(A) $0 \leq P(A) \leq \frac{1}{2}$

(B) $P(A) = 0$

(C) $A \subseteq A^{\mathsf{c}}$

(D) $P(A)$ can be any value between 0 and 1

(E) none of the above.

S3.25 If $P(A) = \frac{1}{4}$ and $P(B) = \frac{1}{4}$ then

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

(B) $P(A \cup B) \leq \frac{1}{2}$

(C) $A \cap B = \text{Ø}$

(D) $A = B$

(E) none of the above.

S3.26 The probability that the birthdays of twelve people will fall in twelve different calendar months (assuming equal probabilities for the twelve months) is

(A) $12!/(12^{12})$

(B) $1/(12!)$

(C) $1/12$

(D) $1/(12^{12})$

(E) none of the above.

S3.27 In a screw factor, there are three machines A, B and C that produce screws. Machine A produces 20% of the daily output. Machine B produces 30% of the daily output. Machine C produces the remaining 50% of the daily output. Of their output, 5%, 4% and 2% of the screws are defective, respectively. A screw is chosen at random from a day's production and found to be defective. What is the probability that it came from machine A?

1. 0.01
2. 0.032
3. 0.2
4. 0.3125
5. none of the above.

S3.28 A box contains fifty light bulbs, exactly five of which are defective. If five bulbs are chosen at random from the box, the probability that all of them are good is equal to

(A) $(0.01)^5$

(B) $1/\binom{50}{5}$

(C) $5/\binom{50}{5}$

(D) $(45!)^2/(40!50!)$

(E) none of the above.

S3.29 Suppose A and B are independent events with $P(A) = p$ and $P(B) = q$. Then $P(A \cap B^c)$ equals

(A) pq

(B) $1 - pq$

(C) $p - pq$

(D) $q - pq$

(E) none of the above.

S3.30 Suppose A and B are independent events with $P(A) = p$ and $P(B) = q$. Then which of the following statements are true:

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

(B) $P(A \cap B^c) = p - pq$

(C) $P(A^c \cap B) = q - pq$

(D) all of the above are true

(E) none of the above are true.

S3.31 An urn contains 5 red balls and 5 black balls. One ball is drawn from the urn at random and then thrown away. A second ball is then drawn from the urn. The probability that the *second* ball is red equals

(A) $1/3$

(B) $4/9$

(C) $1/2$

(D) $5/9$

(E) none of the above.

S3.32 Suppose birthdays are assigned to people randomly with each of the 365 days of the year equally likely. A classroom has 364 students. What is the probability that they all have different birthdays?

(A) 0

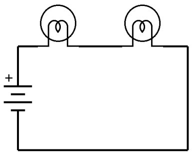
(B) $365!/(365)^{364}$

(C) $1/365$

(D) $(365!364!)/728!$

(E) none of the above.

S3.33 A box contains 10 incandescent lamps. Five of them are good, five are bad. The following circuit contains two such incandescent lamps in series and a (good) battery sufficient to light both lamps:



Both lamps must be good in order for either to light. Two lamps are randomly chosen from the box. The probability that the lamps will light is

(A) $2/9$

(B) $1/4$

(C) $1/2$

(D) $3/4$

(E) none of the above.

S3.34 Which of the following statements (A) (B) or (C) are *not* always true?

(A) $P(E \cap F) = P(E)P(F)$ for any events E and F

(B) $P(\Omega) = 1$

(C) $P(E) \geq 0$ for any event E

(D) all of the above are always true.

(E) none of the above are always true.

S3.35 Suppose

$$P(A \mid B) = 0.4 \qquad P(A \mid B^c) = 0.8 \qquad P(B) = 0.5$$

Then $P(B \mid A)$ equals

(A) 0.2

(B) $1/3$

(C) 0.5

(D) 0.6

(E) none of the above.

S3.36 A study of automobile accidents produced the following data:

Model year	Proportion of all vehicles	Probability of involvement in an accident
2002	0.10	0.02
2003	0.20	0.01
Other	0.70	0.05

You know that an automobile has been involved in an accident. What is the probability that the model year for this automobile is 2003?

(A) 0.0200

(B) 0.0390

(C) 0.0513

(D) 0.3333

(E) none of the above.

Questions for Chapter 3

3.1 Consider the sample space $\Omega = \{1, 2, \ldots, n\}$. Let $|A|$ denote the number of elements in the set A. For any event, A, define $P(A) \equiv |A|/n$. Show that $P(\cdot)$ satisfies the three probability axioms.

3.2 Let P be a probability measure. Show that if $A \subset B$ then

$$P(B \setminus A) = P(B) - P(A).$$

Hint: $B = A \cup (B \setminus A)$.

3.3 There is one lock on the door and the correct key is among the six different ones you usually carry in your pocket. However, in haste, you've lost a key at random. What is the probability that you can still open the door? What is the probability that the first key you try will open the door?

3.4 There are two locks on the door and the keys are among the six different ones you usually carry in your pocket. However, in haste, you've lost one. What is the probability that you can still open the door? What is the probability that the first two keys you try will open the door? (*Note:* There are two interpretations.)

3.5 A convict is planning to escape prison by leaving through the prison sewer system. There are 4 manholes in the prison, but only one leads to the outside. The prisoner plans to enter each of the manholes at random, never reentering an unsuccessful manhole. What is the probability that the prisoner must try exactly 1 manhole? 2 manholes? 3 manholes? 4 manholes?

3.6 If $P(S) = p$ and $P(T) = q$ with $p > 0$, and the events S and T are independent, what is the value of $P(S \cap T \mid S)$?

3.7 Suppose $P(S) = p$ and $P(T) = q$ with $0 < p < 1$ and $0 < q < 1$. What is the value of $P(S \cup T \mid T)$?

3.8 Suppose that the land of a square kingdom, Ω, is divided into three strips, A, B and C of equal area, and suppose that the value per unit is in the ratio of 1:3:2. For any piece of land, S, in this kingdom, the relative value with respect to that of the kingdom is then given by the formula

$$V(S) = \frac{P(S \cap A) + 3P(S \cap B) + 2P(S \cap C)}{2}$$

where $P(D) = (\text{area of } D)/(\text{area of } \Omega)$ for any region D. Show that V is a probability measure, i.e., show that V satisfies the three probability axioms.

3.9 Consider a sample space, Ω, with events A, B and C and probability measure $P(\cdot)$, such that

$$P(A) = 0.3 \qquad P(B) = 0.2 \qquad P(C) = 0.3$$

$$P(A \cup B) = 0.5 \qquad P(A \cup C) = 0.5 \qquad P(B \cup C) = 0.4$$

$$P(A \cup B \cup C) = 0.6$$

(a) Compute each of the following, justifying your answer:

$$P(A \cap B \cap C) \quad P(A^{\mathsf{c}} \cup B^{\mathsf{c}}) \quad P(A \cap B \cap C^{\mathsf{c}})$$

$$P(A \cap B) \qquad P(A \mid B)$$

(b) Are A and B mutually exclusive? Justify your answer.

3.10 (from Selvin[5]) It's "Let's Make a Deal" – the famous TV show starring Monty Hall!

> *Monty Hall:* One of the three boxes labelled A, B and C contains the keys to that new car worth $50,000. The other two are empty. If you choose the box containing the keys, you win the car.
>
> *Contestant:* Gasp!
>
> *Monty Hall:* Select one of these boxes.
>
> *Contestant:* I'll take box B.
>
> *Monty Hall:* Now box A and box C are on the table and here is box B (contestant grips box B tightly). The keys might be in your box! I'll give you $500 for the box.
>
> *Contestant:* No thank you.
>
> *Monty Hall:* I'll give you $2000.
>
> *Contestant:* No, I think I'll keep this box.
>
> *Monty Hall:* I'll do you a favor and open one of the remaining boxes on the table (he opens box A). It's empty! (*Audience:* applause). Now either box C or your box B contains the car keys. Since there are two boxes left, the probability of your box containing the keys is now $\frac{1}{2}$. I'll give you $10,000 cash for your box.
>
> *Contestant:* No, but I'll trade you my box B for box C on the table.

Is Monty right? Assuming Monty knows which of the three boxes contains the keys, is the contestant's probability of winning the car 0.5? Why does the contestant want to trade boxes? (*Hint:* The contestant is a former EAS305 student.)[6]

---

[5]This problem has appeared in various forms, and there are several claims for its origin. To my knowledge, the original problem appeared in the paper by Steve Selvin, "On The Monty Hall Problem," *The American Statistician*, August 1975, Vol. 29, No. 3.

[6]To Monty Hall's credit, once Selvin's paper appeared in *American Statistician*, it is claimed that Monty wrote a letter to Steve Selvin. In that letter, Monty provided a short, but clear, statement of

3.11 Given two mutually exclusive events A and B, and a probability measure $P(\cdot)$ such that $P(A) > 0$ and $P(B) > 0$, can A and B be independent? Why?

3.12 A warehouse receives the same model of a coffee maker from two different factories, A and B. It is known that any one coffee maker has an equal chance of originating from A or B. However, past history has shown that 0.10 of the coffee makers from A are defective while 0.50 of the coffee makers from B are defective.

(a) You know that two particular coffee makers came from the same factory. One is defective. What is the probability that the other one is also defective?

(b) You chose another coffee maker at random from those in the warehouse. You test it and discover it is defective. What is the probability that it came from Factory B?

3.13 A computer can send any one of three characters (A, B or C) to a printer. During each transmission, each character has an equal chance of being sent. However, due to transmission errors, it is possible that the printer will not print the same character as transmitted. The probability that an A will be incorrectly printed as a B is 0.2. The probability that an C will be incorrectly printed as a B is 0.5. Other printer errors are impossible.

(a) Suppose that a single A is transmitted. What is the probability that it is correctly printed?

(b) Suppose that two characters are transmitted independently. What is the probability that two A's will be printed?

(c) Suppose that a single character is transmitted. If a B is printed, what is the probability that a B was transmitted?

3.14 John throws six dice and wins if he scores at least one ace. Mary throws twelve dice and wins if she scores at least two aces. Who has the advantage?

3.15 A person gives you a six-sided die and tells you that it is a fair die. You roll the die 100 times and 99 of the 100 rolls turns up 4. Should you accept the claim that die is a fair die? Why or why not?

---

the correct answer to this question. Monty also cautioned that there is a difference between the actual probability that the keys are in B and the contestant's *perceived* probability. The text of Monty's letter is available on several web sites, but without any verification of its authenticity.

3.16 A professor parks illegally on campus every night and has received twelve tickets for overnight parking. Every one of the 12 tickets was given on either a Tuesday or Thursday. Find the probability of this event assuming tickets are given out on random days. Was his renting a garage only for Tuesdays and Thursdays justified?

3.17 Let n balls be placed "randomly" into n cells. What is the probability that each cell is occupied?

3.18 A parking lot attendant noticed that in a section where there are 10 parking spaces arranged in a row, there are 6 cars parked, but in such a way so that the four empty places were adjacent to each other. Assuming that car owners pick a parking place at random among the spaces available, what is the probability of such an event?

3.19 Consider a box partitioned into 3 cells. Two balls are dropped into the box, each independently of the other, in such a way that every cell has an equal chance of receiving a ball. A cell may contain more than one ball.

(a) If the balls are indistinguishable, list all possible arrangements.

(b) What is the probability of each arrangement?

(c) Suppose that no cell can have more than one ball. Reanswer (a) and (b).

3.20 (from Feller[7]) In a certain family, four boys take turns at washing dishes. Out of a total of four broken dishes, three were caused by the youngest and he was thereafter called "clumsy." Was he justified in attributing the frequency of his breakage to chance? *Hint:* Before the four dishes are broken, adopt the following decision rule: *If the youngest brother breaks three or more of four broken dishes, we will announce that the four dishes were not randomly broken and, thus, he is clumsy.* Using this decision rule, what is the probability of incorrectly announcing that the youngest is clumsy, when the process of breaking dishes is actually random?

3.21 *de Méré's paradox:* Which is more probable: (a) to get at least one ace with four dice, or (b) to get at least one double ace in 24 throws of two dice?

[7]This is a problem from the text: William Feller, *An introduction to probability theory and its applications*, Wiley, 1950. Feller is often regarded as the first person to organize probability theory as a self-contained subject.

B	G	G	B
B	G	G	B
B	B	B	B
B	B	B	B

(B = Boy, G = Girl)

Figure 3.2: Example of a 'Block' for Bell Elementary School

3.22 A bus starts with r passengers and stops at n bus stops ($r \leq n$). If no new passengers are picked up, what is the probability that no two passengers leave at the same stop? What assumptions are you making?

3.23 After a night on the town, the neighborhood postal carrier decides to deliver n letters addressed to n different individuals. However, in his state-of-mind, the carrier places the letters randomly into the n mailboxes, one letter per box.

(a) How many different ways could the postal carrier deliver the letters?

(b) What is the probability that everyone gets the correct letter?

(c) What is the probability that at least two people get the wrong letters?

3.24 A classroom in the Bell Elementary School contains 16 seats with four rows of four seats each. A class is composed of 4 girls and 12 boys.

(a) The students in the class are assigned seats at random. What is the probability that the four girls will be seated together in a block of two rows and two columns somewhere in the classroom. (See Figure 3.2 for an example of such a block).

(b) If the seats are assigned randomly, what is the probability that each row contains exactly one girl?

3.25 Fifty students have rented 3 busses for a class trip. Each bus can hold 50 riders. Answer the following questions assuming that each student chooses his bus for the trip at random:

(a) What is the probability that every bus has at least one rider?

(b) What is the probability that all students sign up for the same bus?

3.26 If an ordinary deck of 52 cards is dealt to four hands of 13 cards each, what is the probability that the player who gets the ace of spades will also get the ace of hearts?

3.27 A class of 100 students vote for class president among two candidates. If each student casts his vote independently and at random for one of the two candidates, what is the probability of a tied election?

3.28 In a family, n boys take turns at washing dishes. Out of a total of d broken dishes. What is the probability that the youngest child breaks exactly x dishes?

3.29 Let Ω be a sample space with probability measure $P(\cdot)$. Show that for any events A, B and C

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & +P(A \cap B \cap C) \end{aligned}$$

3.30 Show that Theorem 3.3 is still true even if we replace $\subseteq$ by *proper subset* (i.e., $A \subset B$ and $A \neq B$). Provide an example where $A \subset B$ (with $A \neq B$) and $P(A) = P(B)$.

3.31 We know that

$$A = B \;\Rightarrow\; P(A) = P(B)$$

Note that this is used in many of the proofs in this chapter. Provide an example to show that the converse is not true. That is.

$$A = B \;\not\Leftarrow\; P(A) = P(B)$$

3.32 A community of $20,000$ voters is composed of $10,000$ Republicans, $5,000$ Democrats and $5,000$ independents. From a recent survey, it is known that 80% of the Republicans favor Presidential candidate JK, while 5% of the Democrats favor JK, and 20% of the independents favor him.

(a) What is the probability that a voter chosen at random from the community will favor JK?

(b) Suppose a voter is chosen from the community and it is learned to favor JK. What is the probability that the voter is a Republican?

3.33 A printing machine capable of printing any of the 26 letters (A through Z) is operated by electrical impulses, each character being produced by a unique impulse. Suppose the machine has the probability 0.9 of correctly producing the character corresponding to the impulse received, independent of past behavior. If it prints the wrong character, the probabilities that any of the 25 other characters will actually appear are equal.

(a) Suppose that one of the 26 impulses is chosen at random and fed into the machine, and that the character C is printed. What is the probability that the impulse chosen was the one designed to produce a C?

(b) Suppose that one of the 26 impulses is chosen at random and fed into the machine *twice*, and the character C is printed both times. What is the probability that the impulse chosen was the one designed to produce a C?

3.34 Let A and B be two events. Suppose that $P(A) = 0.4$ while $P(A \cup B) = 0.7$. Let $P(B) = p$.

(a) For what choice(s) of p is $P(A \cap B) = 0$?

(b) For what choice(s) of p are A and B indpendent?

3.35 Mary, a university student, has a basket that contains 5 blue socks, 4 red socks and 2 green socks. In the dark, she picks two socks at random from the basket. What's the probability that both socks are the same color?

3.36 The *Colorful LED Company* manufacturers both green and red light emitting diodes (LED's). When turned off, the red and green LED's look alike. Only when an electrical current is applied, can it be determined whether the LED gives off green or red light.

LED's are packaged and shipped in bulk using large cardboard barrels.

On a busy production day, one barrel of red LED's was produced. In addition, accidentally, a second barrel was filled with a mixture of LED's, one fourth green and three quarters red. The barrels have no distinguishing markings.

(a) Suppose that one of the two barrels is chosen at random. What is the probability that it is the barrel with a mixture of green and red LED's?

(b) Suppose that one of the two barrels is chosen at random and then an LED is tested. What is the probability that it is a green LED?

(c) Suppose that one of the two barrels is chosen at random, an LED is tested, and it turns out to be RED. What is the probability that it is the barrel with a mixture of green and red LED's?

3.37 You have been captured by aliens from outer space and held in a one-room cell on their planet. The room has three doors labelled A, B and C. Two of the doors lead to corridors that return you back to your one-room cell.

Only one of the three doors leads to an escape spacecraft. Because of your engineering degree, you know how to expertly pilot the spacecraft back to Earth.

(a) You decide to try each door (one at a time) at random. Assume that after a failed escape attempt you never choose a previously chosen door. Let E_k denote the event that you require exactly k escape attempts. Find the values of $P(E_1)$, $P(E_2)$ and $P(E_3)$.

(b) Anticipating your plan to escape, each time you return to your cell, the space aliens zap you with a electromagnetic beam that causes you to forget which doors you have previously chosen. Hence, every time you attempt to escape, you choose randomly from all three doors. Find the values of $P(E_1)$, $P(E_2)$ and $P(E_3)$.

(c) Assuming that the aliens zap you with their anti-memory beam (as in part (b)), you may need more than three attempts to escape. What is the probability that you need more than three attempts?

(d) Assume that the aliens zap you with their anti-memory beam (as in part (b)). Find a formula to compute the value of $P(E_k)$ for any positive integer k.

3.38 *The Belchee Seed Company* sells flower seeds and specializes in red petunias and white petunias. Before they are grown, red and white petunia seeds are indistinguishable except through DNA testing.

At the end of the season, the company has only four barrels of petunia seeds remaining. Two of the barrels contain only white petunia seeds, one contains only red petunia seeds, and the fourth barrel contains a mixture of 10% red petunia seeds and 90% white petunia seeds.

Just before leaving the company, an angry employee removed the labels from all four barrels of seeds so that they are now indistinguishable (unless, of course, you perform a DNA test on the seeds).

(a) Suppose you choose a barrel at random. What is the probability that it is the barrel with the mixture of red and white seeds? Are you using logical, experimental or subjective probability?

(b) You choose a single petunia seed at random from the randomly chosen barrel. What is the probability that it is a red petunia seed?

(c) You now perform a DNA test on that single petunia seed (the one chosen at random in part (b.), from the barrel that was randomly chosen in part (a.)). You determine that it is a *white* petunia seed. What is the probability you selected the barrel containing a mixture of red and white seeds?

3.39 Jelly beans are packaged in one-pound bags, with each bag containing jelly beans that are all of the same color. In order to conduct a probability experiment, you are given one bag each of red, green, purple and yellow jelly beans.

The purple and yellow jelly beans are placed in a blue bowl. The green jelly beans are likewise placed in a white bowl.

Exactly one half of the red jelly beans (one-half pound) are placed in the blue bowl and the other half are placed in the white bowl. The beans in both bowls are thoroughly mixed together.

(a) In the experiment, you will be blindfolded. Then you will choose a bowl at random and then pick a jelly at random from that bowl. Let

$$\begin{aligned} B &= \{\text{the BLUE bowl is chosen}\} \\ W &= \{\text{the WHITE bowl is chosen}\} \end{aligned}$$

Also let

$$\begin{aligned} R &= \{\text{a RED jelly bean is chosen}\} \\ G &= \{\text{a GREEN jelly bean is chosen}\} \\ P &= \{\text{a PURPLE jelly bean is chosen}\} \\ Y &= \{\text{a YELLOW jelly bean is chosen}\} \end{aligned}$$

Do the events B and W form a complete set of alternatives? Briefly explain why. Do the events R, G, P and Y form a complete set of alternatives? Briefly explain why.

(b) While blindfolded, you choose a bowl at random and then pick a jelly bean from that bowl at random. What is the probability that it is a RED jelly bean?

(c) A friend (who is not blindfolded) tells you that you did, indeed, choose a RED jelly bean? What is the probability that you selected the BLUE bowl?

(d) You are still blindfolded. So you do not know precisely which bowl you selected in part (b) and (c). You draw out a *second* jelly bean from the same bowl as you selected in parts (b) and (c). Given that the first jelly bean was RED, what is the probability that the *second* jelly bean is also RED?

3.40 Show that if $P(A) = P(B) = 1$ then $P(A \cap B) = 1$.

3.41 (†) Suppose that you have n white balls and m red balls which you plan to distribute between two boxes. After you have distributed all of the balls between the two boxes, a box will be picked at random and a ball randomly chosen from the selected box. How should you initially distribute the balls so as to maximize the probability of drawing a white ball?

3.42 (†) We define $P(A \mid B)$ to be the number x that solves the equation

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

We do not define $P(A \mid B)$ as

$$(5) \qquad P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Here is one reason why... Consider the following equation

$$(6) \qquad P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B)$$

Show that if we use Definition 3.4, then Equation (6) holds for any events A, B and C. But if we try to use (5) to define conditional probabilities, then Equation (6) doesn't always hold. Why?

3.43 (†) *The 9-1-1 New York Lottery Problem.* Jackpot winners of the daily New York Lottery numbers game, must match a sequence of three numbers, with each number independently selected at random from the digits 0 through 9. On September 11, 2002, the numbers drawn were 9-1-1, the same as the date of the drawing. This was regarded by many as a "miraculous" outcome.

For this question, assume the following:

- Two sets of three numbers are drawn twice each day. That is, the lottery game is played twice a day, once mid-day and once in the evening.
- The numbers can match the calendar date only in the months of January to September.

What is the probability that the lottery numbers will match the calendar date at least once in a year?

3.44 Mr. Pop has a computer with an E-mail account, and he gets a lot of "spam" messages. The "spam" E-mail is of two types:

junk mail that is advertising to sell a product, or

adult content mail that is promoting web sites containing explicit adult material.

Mr. Pop has carefully analyzed his E-mail messages. Each message he receives belongs to one of the following three categories

$$\begin{aligned} B_1 &= \{\text{the message is ``junk mail''}\} \\ B_2 &= \{\text{the message is ``adult content mail''}\} \\ B_3 &= (B_1 \cup B_2)^c = \{\text{the message is regular E-mail}\} \end{aligned}$$

Note that every message is assigned to one and only one category. Mr. Pop estimates that the probability that a single message is assigned to a category is given by

$$P(B_1) = 0.6 \qquad P(B_2) = 0.1 \qquad P(B_3) = 0.3$$

That is, 60% of his E-mail is "junk mail," 10% is "adult content mail," and the remaining 30% is useful, regular E-mail. Upon further analysis, Mr. Pop has determined that most of the "junk mail" contains the text character '\$', and that almost all of the "adult mail" contains the text string ''XXX". Regular mail usually contains neither string. Being a former EAS305 student, Mr. Pop defined the events

$$\begin{aligned} A_\$ &= \{\text{the message contains '\$'}\} \\ A_X &= \{\text{the message contains "XXX"}\} \end{aligned}$$

and determined that

$$\begin{array}{ll} P(A_\$ \mid B_1) = 0.8 & P(A_X \mid B_1) = 0.0 \\ P(A_\$ \mid B_2) = 0.2 & P(A_X \mid B_2) = 0.9 \\ P(A_\$ \mid B_3) = 0.1 & P(A_X \mid B_3) = 0.0 \end{array}$$

(a) Use the law of total probability to find $P(A_\$)$, that is, the probability that an E-mail message chosen at random contains the character '\$'.

(b) An E-mail message arrives and contains the character '\$'. Use Bayes' theorem to find the probability that the message is "junk mail."

(c) An E-mail message arrives and contains the character string "XXX". Use Bayes' theorem to find the probability that the message is "adult content mail."

4

RANDOM VARIABLES

He stares in vain for what awaits him,
And sees in Love a coin to toss;
– EDWIN ARLINGTON ROBINSON, *The Man Against the Sky* (1921)

Basic Concepts

Definition 4.1. *A* **random variable** *is a function* $X(\cdot)$ *which assigns to each element* ω *of* Ω *a real value* $X(\omega)$.

Notation:

(1) $$P(\{\omega \in \Omega : X(\omega) = a\}) \equiv P(\{X(\omega) = a\}) \equiv P(X = a)$$

Cumulative Distribution Function

Definition 4.2. *For any random variable* X*, we define the* **cumulative distribution function (CDF)***,* $F_X(a)$*, as*

$$F_X(a) \equiv P(X \le a).$$

Properties of any cumulative distribution function

1. $\lim_{a\to\infty} F_X(a) = 1$
2. $\lim_{a\to-\infty} F_X(a) = 0$
3. $F_X(a)$ is a nondecreasing function of a.
4. $F_X(x)$ is a right-continuous function of a. In other words, $\lim_{x\downarrow a} F_X(x) = F_X(a)$

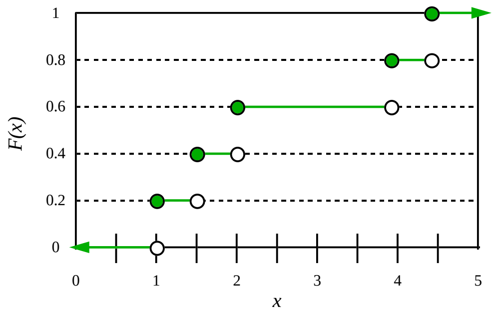


Figure 4.1: An example of a CDF for a discrete random variable

Theorem 4.1. *For any random variable X and real values $a < b$*

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

Discrete Random Variables

Definition 4.3. *The* **probability mass function (pmf)** *of a discrete random variable X is given by*

$$p_X(x_i) \equiv P(X = x_i).$$

Properties of any probability mass function

1. $p_X(x) \geq 0$ for every x
2. $\displaystyle\sum_{\text{all } x_i} p_X(x_i) = 1$
3. $\displaystyle P(E) = \sum_{x_i \in E} p_X(x_i)$

Examples of discrete distributions

The binomial distribution

Consider a sequence of n independent, repeated trials, with each trial having two possible outcomes, *success* or *failure*. Let p be the probability of a success for any single trial. Let X denote the number of successes on n trials. The random variable X is said to have a **binomial distribution** and has probability mass function

(2) $$p_X(k) = \binom{n}{k} p^k (1-p)^{(n-k)} \qquad \text{for } k = 0, 1, \ldots, n.$$

Let's check to make sure that if X has a binomial distribution, then $\sum_{k=0}^{n} p_X(k) = 1$. We will need the binomial expansion for any polynomial:

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

So

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

The geometric distribution

Recall the "success/failure" model that was used to derive the binomial distribution, but, this time, we want to find the probability that the first success occurs on the kth trial. Let p be the probability of a success for any single trial and let Y denote the number of trials needed to obtain the first success.

Then the probability mass function for Y is given by

$$p_Y(k) = (1-p)^{k-1} p \qquad \text{for } k = 1, 2, \ldots.$$

and the random variable Y is said to have a *geometric distribution* with parameter p where $0 \leq p \leq 1$.

Some facts about geometric series When working with the geometric distribution, the following results are often helpful:

Theorem 4.2. *For any* $0 < \theta < 1$,

$$\sum_{j=0}^{\infty} \theta^j = \frac{1}{1-\theta}$$

Proof.

$$\begin{aligned}
\sum_{j=0}^{\infty} \theta^j - \sum_{j=1}^{\infty} \theta^j &= 1 \\
\theta^0 = \sum_{j=0}^{\infty} \theta^j - \sum_{j=0}^{\infty} \theta^{j+1} &= 1 \\
\sum_{j=0}^{\infty} \theta^j - \theta \sum_{j=0}^{\infty} \theta^j &= 1 \\
(1-\theta) \sum_{j=0}^{\infty} \theta^j &= 1 \\
\sum_{j=0}^{\infty} \theta^j &= \frac{1}{1-\theta}
\end{aligned}$$

■

Theorem 4.3. *For any* $0 < \theta < 1$,

$$\sum_{j=0}^{\infty} j\theta^j = \frac{\theta}{(1-\theta)^2}$$

Proof. Using Theorem 4.2

$$\begin{aligned}
\sum_{j=0}^{\infty} \theta^j &= \frac{1}{1-\theta} \\
\frac{d}{d\theta} \sum_{j=0}^{\infty} \theta^j &= \frac{d}{d\theta} \frac{1}{1-\theta}
\end{aligned}$$

$$\begin{aligned}
\sum_{j=0}^{\infty} \frac{d}{d\theta}\theta^j &= \frac{1}{(1-\theta)^2} \\
\sum_{j=1}^{\infty} j\theta^{j-1} &= \frac{1}{(1-\theta)^2} \\
\theta^{-1}\sum_{j=1}^{\infty} j\theta^j &= \frac{1}{(1-\theta)^2} \\
\sum_{j=1}^{\infty} j\theta^j &= \frac{\theta}{(1-\theta)^2} \\
\sum_{j=0}^{\infty} j\theta^j &= \frac{\theta}{(1-\theta)^2}
\end{aligned}$$

■

Example: Using Theorem 4.2, we can show that the probability mass function for a geometric random variable must sum to one. Suppose

$$p_Y(k) = (1-p)^{k-1}p \qquad \text{for } k = 1, 2, \ldots.$$

then

$$\begin{aligned}
\sum_{k=1}^{\infty} p_Y(k) &= \sum_{k=1}^{\infty} (1-p)^{k-1}p \\
&= \sum_{j=0}^{\infty} (1-p)^j p \\
&= \frac{1}{1-(1-p)}p = 1
\end{aligned}$$

The Poisson distribution

The random variable X is said to have a **Poisson distribution** with parameter $\alpha > 0$ if X has a probability mass function given by

(3) $$p_X(k) = e^{-\alpha}\frac{\alpha^k}{k!} \qquad \text{for } k = 0, 1, 2, \ldots$$

Continuous Random Variables

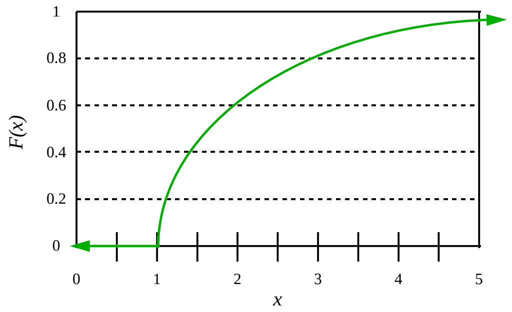


Figure 4.2: An example of a CDF for a continuous random variable

Definition 4.4. *The* **probability density function (pdf)** *of a continuous random variable X is given by*

$$f_X(a) \equiv \frac{d}{da} F_X(a).$$

Properties of any probability density function

1. $f_X(x) \geq 0$ for every x
2. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
3. $P(E) = \int_E f_X(x)\, dx$

Theorem 4.4. *If X is a continuous random variable then*

$$P(X < a) = P(X \leq a).$$

Theorem 4.5. *If X is a continuous random variable, then*

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\, dx. \tag{4}$$

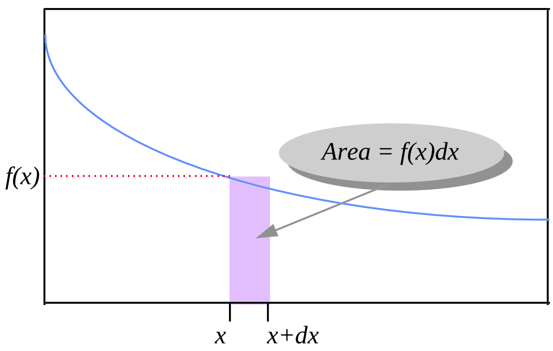


Figure 4.3: The differential mass interpretation of $f_X(x)\,dx$

Differential mass interpretation of the pdf

$$P(X \in [x, x+dx)) = f_X(x)dx \tag{5}$$

Comparison of pmf's and pdf's

A comparison of the properties of the probability mass function and probability density function.

pmf	**pdf**
$p_X(x) \geq 0$	$f_X(x) \geq 0$
$P(X = x) = p_X(x)$	$P(x \leq X < x + dx) = f_X(x)\,dx$
$\sum_{\text{all } x} p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
$F_X(a) = \sum_{x \leq a} p_X(x)$	$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$
$\sum_{x \in E} p_X(x) = P(E)$	$\int_E f_X(x)\,dx = P(E)$

Examples of continuous distributions

The uniform distribution

The random variable X is said to have a **uniform distribution** if it has the probability density function given by

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

The exponential distribution

The random variable X is said to have an **exponential distribution** with parameter $\alpha > 0$ if it has the probability density function given by

(7) $$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Expectation

Definition 4.5. *The* **expected value** *of a random variable X is given by*

$$E(X) = \sum_{\text{all } x} x p_X(x)$$

+ for discrete X, and

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

for continuous X.

Note:

- $E(X)$ is a number.
- The expected value of X is often called the "mean of X."
- The value of $E(X)$ can be interpreted as the *center of mass* or *center of gravity* of the probability mass distribution for X.

Example: Roll a fair die once. Let X equal the number of pips. Then

$$\begin{aligned} E(X) &= \sum_{\text{all } x} x p_X(x) \\ &= (1)p_X(1) + (2)p_X(2) + (3)p_X(3) + (4)p_X(4) + (5)p_X(5) + (6)p_X(6) \\ &= (1)\tfrac{1}{6} + (2)\tfrac{1}{6} + (3)\tfrac{1}{6} + (4)\tfrac{1}{6} + (5)\tfrac{1}{6} + (6)\tfrac{1}{6} = 3.5 \end{aligned}$$

Example: Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x)\, dx \\ &= \int_{-\infty}^{0} x0\, dx + \int_{0}^{1} x2x\, dx + \int_{1}^{+\infty} x0\, dx \\ &= \int_{0}^{1} 2x^2\, dx = \left.\frac{2x^3}{3}\right|_0^1 = \tfrac{2}{3} - 0 = \tfrac{2}{3} \end{aligned}$$

Finding $E(X)$ for some specific distributions

Binomial distribution

Let X be a binomial random variable with parameters n and $0 \leq p \leq 1$. We have

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n.$$

We can find $E(X)$ as follows:

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

Let $s = k - 1$

$$\begin{aligned} E(X) &= \sum_{s=0}^{n-1} n \binom{n-1}{s} p^{s+1} (1-p)^{(n-s-1)} \\ &= np \sum_{s=0}^{n-1} \binom{n-1}{s} p^{s} (1-p)^{(n-s-1)} \\ &= np \end{aligned}$$

Poisson distribution

Let X be a Poisson random variable with parameter $\alpha > 0$. We have

$$p_X(k) = e^{-\alpha} \frac{\alpha^k}{k!} \qquad \text{for } k = 0, 1, 2, \ldots$$

The expected value for X can be found as follows:

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

Let $s = k - 1$

$$\begin{aligned} E(X) &= \sum_{s=0}^{\infty} e^{-\alpha} \frac{\alpha^{s+1}}{s!} \\ &= \alpha \sum_{s=0}^{\infty} e^{-\alpha} \frac{\alpha^{s}}{s!} \\ &= \alpha \end{aligned}$$

Geometric distribution

Let X be a geometric random variable with parameter $0 \le p \le 1$. We have

$$p_X(k) = (1-p)^{k-1} p \qquad \text{for } k = 1, 2, \ldots.$$

To find $E(X)$ let $q = (1-p)$ and then,

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k q^{k-1} p \\ &= p \sum_{k=1}^{\infty} k q^{k-1} \\ &= p \frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

Exponential distribution

Let X be an exponential random variable with parameter $\alpha > 0$. We have

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

We can get $E(X)$ as follows:

$$E(X) = \int_0^{\infty} x \alpha e^{-\alpha x}\, dx$$

Integrate this by parts, letting $dv = \alpha e^{-\alpha x}\, dx$ and $x = u$. Hence $v = -e^{-\alpha x}$ and $du = dx$. This produces

$$E(X) = -x e^{-\alpha x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\alpha x}\, dx = \frac{1}{\alpha}$$

Other measures of central tendency

Definition 4.6. *The* **median** *of a random variable X, denoted by $m_{\frac{1}{2}}(X)$, is given by*

$$P(X \leq m_{\frac{1}{2}}(X)) = \frac{1}{2}.$$

Definition 4.7. *The* **mode** *of a random variable X, denoted by* $\text{Mode}(X)$*, is given by*

$$p_X(\text{Mode}(X)) = \max_x p_X(x) \qquad \textit{if } X \textit{ is discrete}$$

$$f_X(\text{Mode}(X)) = \sup_x f_X(x) \qquad \textit{if } X \textit{ is continuous}$$

Self-Test Exercises for Chapter 4

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S4.1 If X is a discrete random variable with

$$P(X = -1) = P(X = 0) = P(X = 1) = 1/3$$

Then the expected value of X, $E(X)$, equals

(A) 0

(B) $1/3$

(C) $1/2$

(D) any of the values in the set $\{-1, 0, 1\}$

(E) none of the above are true.

S4.2 Suppose X is a continuous random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Then $P(X < 0.50)$ equals

(A) 0

(B) 0.25

(C) 0.50

(D) $\sqrt{0.50}$

(E) none of the above.

S4.3 Suppose X is a random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/3 & \text{if } 0 \le x < 1 \\ 2/3 & \text{if } 1 \le x < 2 \\ 1 & \text{if } x \ge 2 \end{cases}$$

Then $P(X = 2)$ equals

(A) 0

(B) 1/3

(C) 2/3

(D) 1

(E) none of the above.

S4.4 Suppose X is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 1/2 & \text{if } 0 \le x \le c \\ 0 & \text{otherwise} \end{cases}$$

Then the value of c is

(A) 1/2

(B) 1

(C) 2

(D) $+\infty$

(E) none of the above.

S4.5 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \lambda e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

The value of λ must equal

(A) $1/2$

(B) 1

(C) 2

(D) e^{2x}

(E) none of the above.

S4.6 Suppose a fair coin is flipped 3 times. The probability that exactly two heads turn up is

(A) $\frac{1}{2}$

(B) $\frac{1}{4}$

(C) $\frac{3}{8}$

(D) $\frac{2}{3}$

(E) none of the above.

S4.7 Let X be a random variable with cumulative distribution function $F_X(a)$. Which of the following statements are *always* true:

(A) $\lim_{a\to\infty} F_X(a) = 0$

(B) $P(X = a) = F_X(a) - F_X(a-1)$

(C) $0 \leq F_X(a) \leq 1$ for every a

(D) all of the above are true

(E) none of the above are true.

S4.8 The length X (in inches) of a steel rod is a random variable with probability density function given by

$$f_X(x) = \begin{cases} \dfrac{x-9}{2} & \text{if } 9 \leq x \leq 11 \\ 0 & \text{otherwise} \end{cases}$$

A rod is considered defective if its length is less than 10 inches. If 2 independent steel rods are examined, the probability that *exactly one* is defective is

(A) $1/4$

(B) $3/8$

(C) $1/16$

(D) $3/16$

(E) none of the above.

S4.9 An airplane can hold 100 passengers. Passenger behavior has been extensively studied over the years. It is known that the probability is 0.01 that a person making an airline reservation will never show up for his or her flight. Assume that a person decides independently whether or not to show up for a flight. If the airline permits 101 persons to make reservations for a given flight, the probability that someone will be denied a seat is

(A) 0

(B) 0.01

(C) $(0.99)^{101}$

(D) 1

(E) none of the above.

S4.10 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

If $P(X > 1) = e^{-2}$, then α must equal

(A) -1

(B) $1/2$

(C) 1

(D) 2

(E) none of the above.

S4.11 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} x/2 & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Then $P(X \leq 1)$ equals

(A) $1/8$

(B) $1/4$

(C) $1/3$

(D) $1/2$

(E) none of the above.

S4.12 Let X be a discrete random variable with probability mass function

$$P(X = k) = \frac{e^{-1}}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

Then $P(X > 1)$ equals

(A) e^{-1}

(B) $1 - e^{-1}$

(C) $1 - 2e^{-1}$

(D) 1

(E) none of the above.

S4.13 You have two light bulbs. The life length of each bulb (in hours) is a continuous random variable X with probability density function

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

The two bulbs are allowed to burn independently of each other. The probability that both bulbs will *each* last at least one hour is

(A) 0

(B) $(1 - e^{-2})^2$

(C) e^{-4}

(D) $2e^{-2}$

(E) none of the above.

S4.14 Let X be a continuous random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{2}(x+1) & \text{if } -1 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Let $f_X(x)$ denote the probability density function for X. The value of $f_X(0)$ is

(A) 0.00

(B) 0.50

(C) 0.75

(D) 1.00

(E) none of the above.

S4.15 Let X be a random variable with cumulative distribution function $F_X(x)$. Suppose $F_X(0) = 0.3$ and $F_X(1) = 0.5$. Then $P(0 < X \le 1)$ equals

(A) 0

(B) 0.2

(C) 0.3

(D) 0.5

(E) none of the above.

S4.16 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $E(X)$ denote the expected value of X. The value of $E(X)$ is

(A) 0

(B) 3/8

(C) 3/4

(D) 3/2

(E) none of the above.

S4.17 Suppose that an integrated circuit has a life length X (in units of 1000 hours) that is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 3e^{-3x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

An integrated circuit considered defective if its life length is less than 3000 hours. The life lengths of any two integrated circuits are independent random variables. A batch of 100 integrated circuits are tested. The expected number of defective items is

(A) e^{-3}

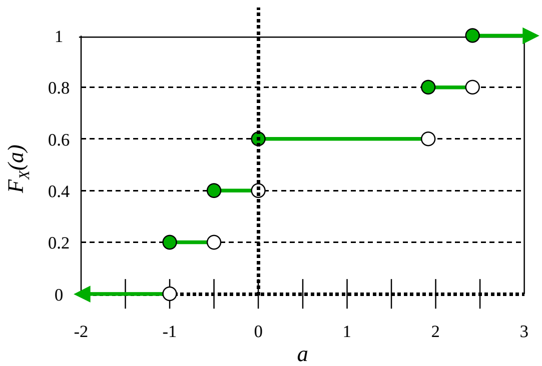
(B) $(1 - e^{-3})$

(C) $(1 - e^{-9})$

(D) $100(1 - e^{-9})$

(E) none of the above.

S4.18 Let X be a discrete random variable with cumulative distribution function, $F(a) \equiv P(X \leq a)$, given by the following graph:



The value of $P(-1 \leq X \leq 1)$ is

(A) 0.2

(B) 0.4

(C) 0.6

(D) 0.8

(E) none of the above.

S4.19 An unfair coin is tossed 4 times. The probability of a head on one toss is $1/4$. What is the probability of getting exactly two heads on 4 tosses?

(A) $\binom{4}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2$

(B) $\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2$

(C) $\left(\frac{1}{4}\right)^2$

(D) $1-\left(\frac{3}{4}\right)^2$

(E) none of the above.

Questions for Chapter 4

4.1 Let X be a continuous random variable having the probability density function

$$f_X(x) = \begin{cases} \frac{3}{2}(1-x^2) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute and plot the cumulative distribution function for X.

(b) Find $P(X \ge 0.5)$.

(c) Find $E(X)$.

4.2 The angle of ascent of an erratic rocket is a random variable X having the cumulative distribution function given by

$$F_X(x) = \begin{cases} 1 & \text{if } x > \pi/2 \\ \sin x & \text{if } 0 \le x \le \pi/2 \\ 0 & \text{if } x < 0 \end{cases}$$

(a) Compute $P(X \le \pi/8)$.

(b) Find the probability density function $f_X(x)$ for X.

(c) Find $E(X)$.

4.3 Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} \frac{3}{2}\sqrt{x} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let F denote the cumulative distribution function of X.

(a) Compute $F(0.25)$, $F(1.3)$, and $F(-0.25)$.

(b) Compute $P(-1 \leq X \leq 0.25)$.

(c) Find $E(X)$.

4.4 Let X be a discrete random variable with probability mass function

$$p_X(-1) = p \qquad p_X(0) = 1 - 2p \qquad p_X(1) = p$$

The value of p is unknown. Identify *all possible* values for p.

4.5 Let Y be a random variable with probability density function

$$f_Y(y) = \begin{cases} 3y^2 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the cumulative distribution function for Y.

(b) Compute $P\left(Y \leq \frac{1}{2}\right)$, $P\left(Y \geq \frac{1}{2}\right)$, $P\left(Y \leq 0\right)$, and $P\left(\frac{1}{3} \leq Y \leq \frac{1}{2}\right)$.

(c) Find $E(Y)$.

4.6 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} cx & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of c.

4.7 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 1/\theta & \text{if } 0 \leq x \leq \alpha \\ 0 & \text{otherwise} \end{cases}$$

What relationship must hold between θ and α?

4.8 The angle of ascent of an erratic rocket is a random variable X having the cumulative distribution function given by

$$F_X(x) = \begin{cases} 1 & \text{if } x > \pi/4 \\ \sin x & \text{if } 0 \le x \le \pi/4 \\ 0 & \text{if } x < 0 \end{cases}$$

Note: This is an example of a *mixed probability distribution* with both discrete and continuous probability mass. There is a positive point mass at $\pi/4$.

(a) Compute $P(X \le \pi/8)$.

(b) Find the probability density function $f_X(x)$ for X and the probability mass at $\pi/4$.

(c) Find $E(X)$.

4.9 Suppose that D, the daily demand for an item, is a random variable with the following probability distribution:

$$P(D = k) = \frac{c2^k}{k!} \qquad k = 1, 2, 3, \ldots$$

(a) Evaluate the constant c.

(b) Compute the expected demand.

4.10 The probability is p that there will be at least one accident in a certain factory in any given week. If the number of weekly accidents in this factory is distributed with a Poisson distribution, what is the expected number of weekly accidents?

4.11 *True or False:* If X has a Poisson distribution and $P(X = 0) = e^{-1}$, then $P(X = 1) = e^{-1}$.

4.12 Let X be a discrete random variable with probability mass function

$$\begin{array}{ll} p_X(-2) = 0.2 & p_X(-1) = 0.4 \\ p_X(0) = 0.2 & p_X(+2) = 0.2 \end{array}$$

Draw a plot of the cumulative distribution function (CDF) for X.

4.13 Suppose that X is a random variable with the following probability density function:

$$f_X(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute $P(X > 0.5)$.

(b) Let $F_X(x)$ denote the cumulative distribution function for X. Find $F_X(0.5)$.

(c) Compute $P(-0.5 \leq X \leq 0.5)$.

4.14 The *Classey Gassey* gasoline company produces gasoline at two plants (A and B) and then delivers the gasoline by truck to gasoline stations where it is sold.

The octane rating of gasoline produced at Plant A is a continuous random variable X with probability density function

$$f_X(x) = \begin{cases} 0.005(x-80) & \text{if } 80 < x < 100 \\ 0 & \text{otherwise} \end{cases}$$

The octane rating of gasoline produced at Plant B is a continuous random variable Y with probability density function

$$f_Y(y) = \begin{cases} 0.02(y-80) & \text{if } 80 < y < 90 \\ 0 & \text{otherwise} \end{cases}$$

The *grade* of the gasoline is determined by its octane rating. If the octane rating is r, then the gasoline is

$$\begin{array}{rl} \text{Low grade if} & r < 85 \\ \text{High grade if} & r \geq 85 \end{array}$$

When gasoline is delivered, there is an equal probability that it was produced at Plant A or Plant B.

You are the owner of a *Classey Gassey* gasoline station. A truck has arrived with today's shipment of gasoline.

(a) Given only the above information, what is the probability that the shipment is *Low grade* gasoline?

(b) You actually test today's shipment of gasoline and discover that it is, in fact, *Low grade* gasoline. What is the probability that the shipment came from Plant A?

4.15 Suppose that X is a random variable with the following probability density function:

$$f_X(x) = \begin{cases} cx & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

(a) What is the value of the constant c?

(b) Find $P(X < 0.5)$.

4.16 Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 2/x^2 & \text{if } 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Let F denote the cumulative distribution function of X.

(a) Compute $F(-11)$, $F(\frac{3}{2})$, and $F(14)$.

(b) Find $E(X)$.

4.17 Let X be a discrete random variable with probability mass function

$$\begin{array}{lll} p_X(-4) = 1/4 & p_X(-\pi) = 1/4 & p_X(0) = 1/4 \\ p_X(4) = 1/8 & p_X(\pi) = 1/8 & \end{array}$$

Let F denote the cumulative distribution function of X.

(a) Compute $F(-5)$, $F(-3)$, $F(\pi)$ and $F(2\pi)$.

(b) Compute $E(X)$.

4.18 The manager of a bakery knows that the number of chocolate cakes he can sell on any given day is a random variable with probability mass function

$$p(x) = \frac{1}{6} \qquad \text{for } x = 0, 1, 2, 3, 4 \text{ and } 5.$$

He also knows that there is a profit of \$1.00 on each cake that he sells and a loss (due to spoilage) of \$0.40 on each cake that he does not sell. Assuming that each cake can be sold only on the day it is made, find the baker's expected profit for

(a) a day when he bakes 5 chocolate cakes. *Hint:* Define the random variable Y as the actual daily profit (or loss). Using the probability distribution for X (the number of customer asking for cakes) find the probability distribution for Y. The probability distribution for Y will depend on the number of cakes you decide to bake that day. You can then answer the question using the random variable Y. Be careful to consider the possibility that the demand for your cakes, X, may exceed your supply.

(b) a day when he bakes 4 chocolate cakes.

(c) How many chocolate cakes should he bake to maximize his expected profit?

4.19 If a contractor's profit on a construction job is a random variable with probability density function

$$f_X(x) = \begin{cases} \frac{1}{18}(x+1) & -1 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

where the units are \$1000, what is her expected profit?

4.20 (†) A large bag contains pretzels with 4 different shapes: heart, square, circle and star. Billy plans to reach into the bag and draw out one pretzel at a time, stopping only when he has at least one pretzel of each type. Let the random variable X denote the number of pretzels that Billy will draw from the bag. You can assume that there is an infinite number of pretzels in the bag with equal proportions of each shape.

(a) Find $P(X = k)$ for $k = 4, 5, 6, \ldots$.

(b) Find $E(X)$.

5

FUNCTIONS OF RANDOM VARIABLES

When you do dance, I wish you
A wave o'the sea, that you might ever do
Nothing but that, move still, still so,
And own no other function.
– WILLIAM SHAKESPEARE, *The Winter's Tale* (c 1600)

Basic Concepts

Two of the primary reasons for considering random variables are:

1. They are convenient, and often obvious, representations of experimental outcomes.
2. They allow us to restrict our attention to sample spaces that are subsets of the real line.

There is a third, and probably more important reason for using random variables:

3. They permit algebraic manipulation of event definitions.

This chapter deals with this third "feature" of random variables. As an example, let X be the random variable representing the Buffalo, New York temperature at 4:00 PM next Monday in degrees Fahrenheit. This random variable might be suitable for the use of Buffalonians, but our neighbors in Fort Erie, Ontario prefer the random variable Y representing the same Buffalo temperature, except in degrees Celsius. Furthermore, they are interested in the event

$$A = \{\omega \in \Omega : Y(\omega) \leq 20\}$$

where Ω is the sample space of the underlying experiment. Note that Ω is being mapped by both of the random variables X and Y into the real line.

In order to compute $P(A)$, we could evaluate the cumulative distribution function for Y, $F_Y(y)$, at the point $y = 20$. But suppose we only know the cumulative distribution function for X, namely $F_X(x)$. Can we still find $P(A)$?

The answer is "yes." First, notice that for any $\omega \in \mathbf{\Omega}$,

(1) $$Y(\omega) = \tfrac{5}{9}\left(X(\omega) - 32\right).$$

We can then redefine the event A in terms of X as follows:

$$\begin{aligned} A &= \{\omega \in \mathbf{\Omega} : Y(\omega) \le 20\} \\ &= \left\{\omega \in \mathbf{\Omega} : \tfrac{5}{9}(X(\omega) - 32) \le 20\right\} \\ &= \{\omega \in \mathbf{\Omega} : X(\omega) - 32 \le 36\} \\ &= \{\omega \in \mathbf{\Omega} : X(\omega) \le 68\}, \end{aligned}$$

So, knowing a relationship like Equation 1 between $X(\omega)$ and $Y(\omega)$ for every $\omega \in \mathbf{\Omega}$, allows a representation of the event A in terms of either the random variable X or the random variable Y.

When finding the probability of the event A, we can use the shortcut notation presented in Chapter 5:

$$\begin{aligned} P(A) &= P(Y \le 20) \\ &= P(\tfrac{5}{9}(X - 32) \le 20) \\ &= P(X - 32 \le 36) \\ &= P(X \le 68) \\ &= F_X(68). \end{aligned}$$

So, if we know the probability distribution for X, we can compute probabilities of events expressed in terms of the random variable Y.

The general concept surrounding the above example is the notion of a *function of a random variable*. To show you what we mean by this, let the function $h(x)$ be defined by

$$h(x) = \tfrac{5}{9}(x - 32),$$

which is nothing more than the equation of a straight line. If X is a random variable, we also have

$$Y = h(X)$$

as a random variable.[1] The function $h(x)$ maps merely points in the real line (the sample space for X) into points in the real line (the sample space for Y) as in Figure 5.1.

[1]Note that we should have really written this as $Y(\omega) = h(X(\omega))$. In order to simplify the notation, mathematicians usually will drop the (ω).

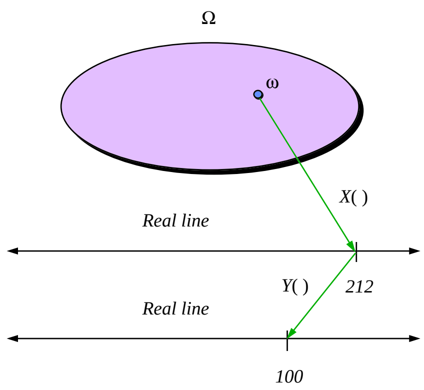


Figure 5.1: The mapping $Y(\omega) = h(X(\omega))$

The random variable X serves as a "middleman" in the mapping shown in Figure 5.1. It may seem, at first, more obvious to define Y as mapping directly from Ω to the real numbers. But consider the case where we have already defined X and then are given a functional relationship between X and Y. Under those circumstances, it would be very handy to have a tool that would enable us to compute probabilities involving Y using the facts we already know about X.

Referring to Figure 5.2, you can see that what is happening is really quite simple. Suppose we are given a cumulative distribution function for the random variable X, the Buffalo temperature in degrees Fahrenheit. Using this, we can compute $P(X \leq t)$ for any value of t. This gives us the cumulative distribution function $F_X(t)$ and allows us to compute the probability of many other events.

Returning to our example, the folks in Fort Erie are asking us to compute $P(Y \leq 20)$. To oblige, we ask ourselves, "What are all of the possible outcomes of the experiment (i.e., measuring Monday's 4 PM temperature) that would result in announcing that the temperature in degrees Celsius in 20 or less?" Call this event A and compute the probability of A from the distribution of the random variable X.

An even more general situation appears in Figure 5.3. In this case, suppose we are given a random variable X. Also assume that we are given the cumulative

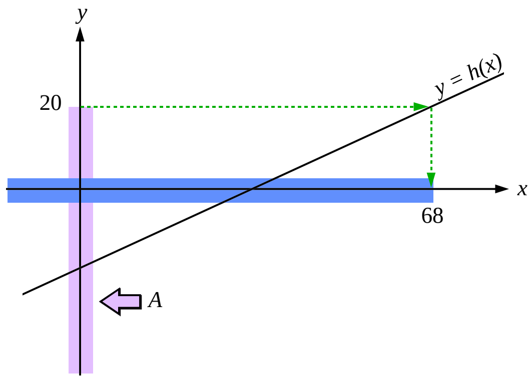


Figure 5.2: Representing the event $\{Y \leq 20\}$

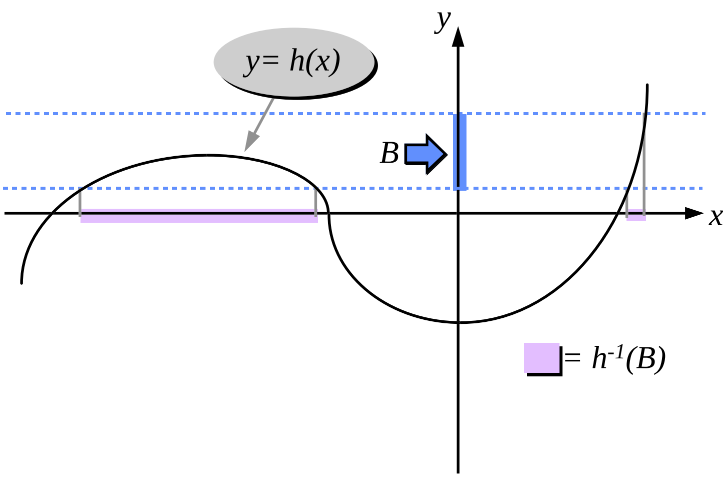


Figure 5.3: A more complicated example

distribution function for X so that we are capable of computing the probability measure of any subset of the sample space for X. If we then define the random variable $Y = h(X)$ (where h is the function shown in Figure 5.3) we can compute the probability measure of any subset B in the sample space for Y by determining which outcomes of X will be mapped by h into B. The probability of any of these particular outcomes of X is, therefore, also the probability of observing Y taking on a value in B.

Some examples

Discrete case

Example: Let X be a discrete random variable with probability mass function

$$\begin{aligned} p_X(-2) &= \tfrac{1}{5} \\ p_X(-1) &= \tfrac{1}{5} \\ p_X(0) &= \tfrac{1}{5} \\ p_X(1) &= \tfrac{1}{5} \\ p_X(2) &= \tfrac{1}{5} \end{aligned}$$

Let $Y = X^2$. We would like to find the probability distribution for Y.

Solution: There is a pleasant fact that helps when working with discrete random variables:

Theorem 5.1. *A function of any discrete random variable is also a discrete random variable.*

Unfortunately, we cannot make such a statement for continuous random variables, as we will see in a later example.

To solve the problem, we should first ask ourselves, "Since X can take on the values -2, -1, 0, 1 or 2, what values can Y assume?" We see that the function $h(x) = x^2$ maps the point $x = -2$ into the point 4, $x = -1$ into 1, $x = 0$ into 0, $x = 1$ into 1 and $x = 2$ into 4. So all of the probability mass for the random variable Y is assigned to the set $\{0, 1, 4\}$. Such a set is called the *support* for the random variable Y.

Definition 5.1. *Let A be the smallest subset of $\mathbb{R}$ such that $P(Y \in A) = 1$. Then A is called the* **support** *of the distribution for the random variable Y.*

Since X will take on the value -2 with probability $\frac{1}{5}$ and the value 2 with probability $\frac{1}{5}$, and since each of these mutually exclusive events will result in Y taking on the value 4, we can say that

$$p_Y(4) = p_X(-2) + p_X(2) = \tfrac{2}{5}.$$

Similarly, $p_Y(1) = \frac{2}{5}$ and $p_Y(0) = \frac{1}{5}$. Hence the probability distribution for Y is given by the probability mass function:

$$\begin{aligned} p_Y(0) &= \tfrac{1}{5} \\ p_Y(1) &= \tfrac{2}{5} \\ p_Y(4) &= \tfrac{2}{5} \end{aligned}$$

Continuous case

Example: Let X be a uniformly distributed random variable on the interval $(-1, 1)$, i.e., the density for X is given by

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

and the cumulative distribution function is

$$F_X(x) = \begin{cases} 0 & \text{if } x < -1 \\ \dfrac{x+1}{2} & \text{if } -1 \leq x \leq 1 \\ 1 & \text{if } x > 1. \end{cases}$$

Find the distribution for the random variable $Y = X^2$.

Solution: For continuous random variables, it is easier to deal with the cumulative distribution functions. Let's try to find $F_Y(y)$. First, we should ask ourselves, "Since X can take on only values from -1 to $+1$, what values can Y take on?" Since $Y = X^2$, the answer is "Y can assume values from 0 to 1." With no elaborate computation, we now know that

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ ? & \text{if } 0 \leq y \leq 1 \\ 1 & \text{if } y > 1. \end{cases}$$

Now suppose that y is equal to a value in the only remaining unknown case, that is $0 \leq y \leq 1$. We then have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(X^2 \leq y\right) \\ &= P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) \\ &= F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) \\ &= \frac{\sqrt{y}+1}{2} - \frac{-\sqrt{y}+1}{2} \\ &= \sqrt{y}. \end{aligned}$$

So,

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ \sqrt{y} & \text{if } 0 \leq y \leq 1 \\ 1 & \text{if } y > 1 \end{cases}$$

Finding the probability density function for Y is now easy. That task is left to you.

Example: Let X be a uniformly distributed random variable on the interval $(0, 2)$, i.e., the density for X is given by

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

and the cumulative distribution function is

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 1 & \text{if } x > 2 \end{cases}$$

Let $h(x)$ be the greatest integer less than or equal to x, i.e., let h be given by the step function shown in Figure 5.4. Find the distribution for $Y = h(X)$.

Solution: Since X takes on the values ranging continuously from 0 to 2, Y can take on only the integers 0, 1 and 2 (why 2?). We immediately know that the random variable Y is discrete, and we have our first example of a continuous random variable being mapped into a discrete random variable. Look at Figure 5.4 again and try to visualize why this happened.

To find the probability mass function for Y, first let's try to find $p_Y(0)$. We know that Y takes on the value 0 when X takes on any value from 0 to 1 (but not including

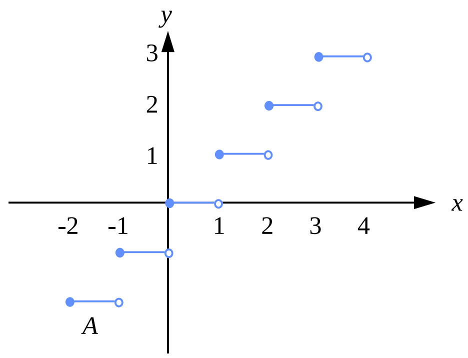


Figure 5.4: The greatest integer function

1). Hence,

$$p_Y(0) = \int_0^1 \frac{1}{2}\,dx = \left.\frac{x}{2}\right|_0^1 = \frac{1}{2}.$$

Similarly,

$$p_Y(1) = \int_1^2 \frac{1}{2}\,dx = \frac{1}{2}$$

and

$$p_Y(2) = \int_2^2 \frac{1}{2}\,dx = 0,$$

which is good thing because we ran out of probability mass after the first two computations.

So we have the answer,

$$p_Y(0) = p_Y(1) = \tfrac{1}{2}.$$

Invertible Functions of Continuous Random Variables

First we need a few concepts about inverse functions from calculus:

Definition 5.2. *A function $h : \mathbb{R} \to \mathbb{R}$ is called* **one-to-one**, *if $h(a) = h(b)$ implies $a = b$. A function $h : \mathbb{R} \to \mathbb{R}$ is called* **onto**, *if for every $y \in \mathbb{R}$ there exists some $x \in \mathbb{R}$ such that $y = h(x)$.*

If a function $h(x)$ is both one-to-one and onto, then we can define the inverse function $h^{-1} : \mathbb{R} \to \mathbb{R}$ such that $h^{-1}(y) = x$ if and only if $h(x) = y$.

Suppose $h(x)$ has an inverse and is increasing (or decreasing). Then, if X is a continuous random variable, and $Y = h(X)$, we can develop a general relationship between the probability density function for X and the probability density function for Y.

If $h(x)$ a strictly increasing function and $h(x)$ has an inverse, then $h^{-1}(x)$ is nondecreasing. Therefore, for any $a \leq b$, we have $h^{-1}(a) \leq h^{-1}(b)$. This produces the following relationship between the cumulative distribution function for X and the cumulative distribution function for Y:

$$\begin{aligned} F_Y(y) = P[Y \leq y] &= P[h(X) \leq y] \\ F_Y(y) &= P[h^{-1}(h(X)) \leq h^{-1}(y)] \\ F_Y(y) &= P[X \leq h^{-1}(y)] \\ F_Y(y) &= F_X(h^{-1}(y)) \end{aligned}$$

To get the relationship between the densities, take the first derivative of both sides with respect to y and use the chain rule:

$$\begin{aligned} \frac{d}{dy} F_Y(y) &= \frac{d}{dy} F_X(h^{-1}(y)) \\ f_Y(y) &= f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y) \end{aligned}$$

We have just proven the following theorem:

Theorem 5.2. *Suppose X is a continuous random variable with probability density function $f_X(x)$. If $h(x)$ is a strictly increasing function with a differentiable inverse $h^{-1}(\cdot)$, then the random variable $Y = h(X)$ has probability density function*

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

We will leave it to you to find a similar result for the case when $h(x)$ is decreasing rather than increasing.

Note: This transformation technique is actually a special case of a more general technique from calculus. If one considers the function $Y = h(X)$ as a transformation $(x) \to (y)$, the monotonicity property of h allows us to uniquely solve

$y = h(x)$. That solution is $x = h^{-1}(y)$. We can then express the probability density function of Y as

$$f_Y(y) = f_X(h^{-1}(y))J(y)$$

where $J(y)$ is the following 1×1 determinant:

$$J(y) = \left|\frac{\partial x}{\partial y}\right| = \left|\frac{\partial}{\partial y}h^{-1}(y)\right| = \frac{\partial}{\partial y}h^{-1}(y)$$

since $h^{-1}(y)$ is also an increasing function of y and, thus, has a nonnegative first derivative with respect to y.

In calculus, when solving transformation problems with n variables, the resulting $n \times n$ determinant is called the *Jacobian* of the transformation.

Example: Let X be a continuous random variable with density for X given by

$$f_X(x) = \begin{cases} |x|/4 & \text{if } -2 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = h(X) = X^3$. Find the probability density function for Y without explicitly finding the cumulative distribution function for Y.

Solution: The support for X is $[-2, 2]$. So the support for Y is $[-8, 8]$.

Since $h^{-1}(y) = y^{\frac{1}{3}}$, we can use Theorem 5.2 to get

$$\begin{aligned} f_Y(y) &= f_X(h^{-1}(y))\frac{d}{dy}h^{-1}(y) \\ &= f_X(y^{\frac{1}{3}})\frac{d}{dy}y^{\frac{1}{3}} \\ &= f_X(y^{\frac{1}{3}})(\tfrac{1}{3})y^{-\frac{2}{3}} \end{aligned}$$

When $y < -8$ or $y > 8$ we have $y^{\frac{1}{3}} < -2$ or $y^{\frac{1}{3}} > 2$. In those cases, $f_X(y^{\frac{1}{3}}) = 0$ hence $f_Y(y)$ is zero.

When $-8 \le y \le 8$

$$\begin{aligned} f_Y(y) &= \tfrac{1}{4}|y^{\frac{1}{3}}|(\tfrac{1}{3})y^{-\frac{2}{3}} \\ &= \tfrac{1}{12}|y^{\frac{1}{3}}|\tfrac{1}{\sqrt[3]{y^2}} \\ &= \tfrac{1}{12}|y^{-\frac{1}{3}}| \end{aligned}$$

Summarizing

$$f_Y(y) = \begin{cases} |y^{-\frac{1}{3}}|/12 & \text{if } -8 \le y \le 8 \\ 0 & \text{otherwise} \end{cases}$$

Expected values for functions of random variables

Theorem 5.3. *Let $Y = h(X)$.*

If X is a discrete random variable with probability mass function $p_X(x)$, then

$$E(Y) = E(h(X)) = \sum_{\text{all } x} h(x) p_X(x)$$

If X is a continuous random variable with probability density function $f_X(x)$, then

$$E(Y) = E(h(X)) = \int_{-\infty}^{+\infty} h(x) f_X(x)\, dx$$

Question: Toss a coin. Let $X = 1$ if heads, $X = 0$ if tails. Let $h(X)$ denote your winnings.

$$h(x) = \begin{cases} -5 & x = 0 \\ 10 & x = 1 \end{cases}$$

Find $E(h(X))$, your expected winnings.

Question: Let X, the life length of a transistor, be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 2e^{-2x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

Suppose you earn $\ln(x) + x^2 - 3$ dollars if the transistor lasts exactly x hours. What are your expected earnings?

Theorem 5.4. *$E(a) = a$ if a is a constant.*

Proof. We will provide proofs for both the discrete and continuous cases.

We are really finding $E(h(X))$ where $h(x) = a$ for all x.

Suppose X is a discrete random variable with probability mass function

$$p_X(x_k) = P(X = x_k) = p_k \qquad \text{for } k = 1, 2, \ldots.$$

Then

$$\begin{aligned} E(h(X)) &= \sum_k h(x_k)p_k \\ E(a) &= \sum_k a p_k \\ &= a\sum_k p_k = a(1) = a \end{aligned}$$

Suppose X is a continuous random variable with probability density function $f_X(x)$. Then

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x) f_X(x)\,dx \\ E(a) &= \int_{-\infty}^{\infty} a f_X(x)\,dx \\ &= a\int_{-\infty}^{\infty} f_X(x)\,dx = a(1) = a \end{aligned}$$

■

Theorem 5.5. *$E(aX) = aE(X)$ if a is a constant.*

Proof. We are finding $E(h(X))$ where $h(x) = ax$.

Suppose X is a discrete random variable with probability mass function

$$p_X(x_k) = P(X = x_k) = p_k \qquad \text{for } k = 1, 2, \ldots.$$

Then

$$\begin{aligned} E(h(X)) &= \sum_k h(x_k)p_k \\ E(aX) &= \sum_k a x_k p_k \\ &= a\sum_k x_k p_k = a(E(X)) \end{aligned}$$

Suppose X is a continuous random variable with probability density function $f_X(x)$. Then

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x) f_X(x)\,dx \\ E(aX) &= \int_{-\infty}^{\infty} a x f_X(x)\,dx \\ &= a\int_{-\infty}^{\infty} x f_X(x)\,dx = a(E(X)) \end{aligned}$$

■

Theorem 5.6. *$E(X+b) = E(X) + b$ if b is a constant.*

Proof. We are finding $E(h(X))$ where $h(x) = x + b$.

Suppose X is a discrete random variable with probability mass function

$$p_X(x_k) = P(X = x_k) = p_k \qquad \text{for } k = 1, 2, \ldots.$$

Then

$$\begin{aligned} E(h(X)) &= \sum_k h(x_k)p_k \\ E(X+b) &= \sum_k (x_k + b)p_k = \sum_k (x_k p_k + b p_k) \\ &= \sum_k x_k p_k + \sum_k b p_k \\ &= \sum_k x_k p_k + b \sum_k p_k = E(X) + b(1) \end{aligned}$$

Suppose X is a continuous random variable with probability density function $f_X(x)$. Then

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x) f_X(x)\,dx \\ E(X+b) &= \int_{-\infty}^{\infty} (x+b) f_X(x)\,dx = \int_{-\infty}^{\infty} (x f_X(x) + b f_X(x))\,dx \\ &= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} b f_X(x)\,dx = E(X) + b(1) \end{aligned}$$

■

Theorem 5.7. *$E[h_1(X) + h_2(X)] = E[h_1(X)] + E[h_2(X)]$ for any functions $h_1(X)$ and $h_2(X)$*

Proof. We are finding $E(h(X))$ where $h(x) = h_1(x) + h_2(x)$.

Suppose X is a discrete random variable with probability mass function

$$p_X(x_k) = P(X = x_k) = p_k \qquad \text{for } k = 1, 2, \ldots.$$

Then

$$E(h(X)) = \sum_k h(x_k)p_k$$

$$
\begin{aligned}
E[h_1(X) + h_2(X)] &= \sum_k [h_1(x_k) + h_2(x_k)]p_k \\
&= \sum_k [h_1(x_k)p_k + h_2(x_k)p_k] \\
&= \sum_k h_1(x_k)p_k + \sum_k h_2(x_k)p_k \\
&= E[h_1(X)] + E[h_2(X)]
\end{aligned}
$$

Suppose X is a continuous random variable with probability density function $f_X(x)$. Then

$$
\begin{aligned}
E(h(X)) &= \int_{-\infty}^{\infty} h(x) f_X(x)\,dx \\
E[h_1(X) + h_2(X)] &= \int_{-\infty}^{\infty} (h_1(x) + h_2(x)) f_X(x)\,dx \\
&= \int_{-\infty}^{\infty} (h_1(x) f_X(x) + h_2(f_X(x))\,dx \\
&= \int_{-\infty}^{\infty} h_1(x) f_X(x)\,dx + \int_{-\infty}^{\infty} h_2(x) f_X(x)\,dx \\
&= E[h_1(X)] + E[h_2(X)]
\end{aligned}
$$

■

Notice in the all of the above proofs the close relationship of

$$
\begin{array}{rcl}
\text{Discrete } X & \text{and} & \text{Continuous } X \\
\sum_k & \text{and} & \int_{-\infty}^{\infty} \\
P(X = k) = p_X(x_k) = p_k & \text{and} & P(x \le X \le x + dx) = f_X(x)\,dx
\end{array}
$$

Also notice that Theorem 5.6 is really only a corollary to Theorem 5.7. Simply use Theorem 5.7 with $h_1(x) = x$ and $h_2(x) = b$ and you get Theorem 5.6.

The kth Moment of X

If we let $h(x) = x^k$ and consider $E[h(X)]$, we get the following definition:

Definition 5.3. *The quantity $E(X^k)$ is defined as the* ***k^{th} moment about the origin*** *of the random variable X.*

Variance

Recall that $E(X)$ is simply a number. If we let $h(x) = (x - E(X))^2$ and compute $E(h(X))$ we get a very useful result:

Definition 5.4. *The* **variance** *of a random variable X is*

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

.

Note that

$$\text{Var}(X) = \sum_{\text{all } x} (x - E(X))^2 p_X(x)$$

for a discrete random variable X, and

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x)\, dx$$

for a continuous random variable X.

Note: The value of Var(X) is the second moment of the probability distribution about the expected value of X. This can be interpreted as the *moment of inertia* of the probability mass distribution for X.

Definition 5.5. *The* **standard deviation** *of a random variable X is given by*

$$\sigma_X \equiv \sqrt{\text{Var}(X)}.$$

Notation: We often write

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

Theorem 5.8. Var$(X) = E(X^2) - [E(X)]^2$

Proof. Using the properties of expected value, we get

$$\begin{aligned}
\text{Var}(X) &= E[(X - E(X))^2] \\
&= E[X^2 - 2XE(X) + [E(X)]^2] \quad \text{expand the square} \\
&= E[X^2] + E[-2XE(X)] + E[[E(X)]^2] \quad \text{from Theorem 5.7} \\
&= E[X^2] - 2E(X)E[X] + E[[E(X)]^2] \quad \text{from Theorem 5.5}
\end{aligned}$$

$$\begin{aligned}
&= E[X^2] - 2E(X)E[X] + [E(X)]^2 \quad \text{from Theorem 5.4} \\
&= E[X^2] - 2[E[X]]^2 + [E(X)]^2 \\
&= E[X^2] - [E[X]]^2
\end{aligned}$$

■

Theorem 5.9. $\text{Var}(aX) = a^2\text{Var}(X)$

Proof. Using Theorem 5.8, we get

$$\begin{aligned}
\text{Var}(aX) &= E[(aX)^2] - [E(aX)]^2 \\
&= E[a^2X^2] - [E(aX)]^2 \\
&= a^2E[X^2] - [aE(X)]^2 \quad \text{Using Theorem 5.5} \\
&= a^2E[X^2] - a^2[E(X)]^2 \\
&= a^2(E[X^2] - [E(X)]^2) \\
&= a^2\text{Var}(X) \quad \text{Using Theorem 5.8}
\end{aligned}$$

■

Theorem 5.10. $\text{Var}(X + b) = \text{Var}(X)$

Proof. Using Definition 5.4, we get

$$\begin{aligned}
\text{Var}(X + b) &= E[((X + b) - E[X + b])^2] \\
&= E[(X + b - E[X] - b)^2] \\
&= E[(X - E[X])^2] \\
&= \text{Var}(X)
\end{aligned}$$

■

The proof of the following result is left to the reader:

Theorem 5.11. *Let $Y = h(X)$ then*

$$\text{Var}(Y) = E\left[[h(X) - E(h(X))]^2\right]$$

Example: Suppose X is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = h(X) = 1/X$. Then

$$\begin{aligned} E(Y) &= \int_0^1 h(x)(3x^2)\,dx \\ &= \int_0^1 \frac{1}{x}(3x^2)\,dx \\ &= \frac{3}{2}x^2\Big|_0^1 = \frac{3}{2} \end{aligned}$$

Then we can use $E(Y) = E(h(X)) = \frac{3}{2}$ to compute

$$\begin{aligned} \mathrm{Var}(Y) &= \int_0^1 (h(x) - E(h(X)))^2(3x^2)\,dx \\ &= \int_0^1 \left(\frac{1}{x} - \frac{3}{2}\right)^2 (3x^2)\,dx \\ &= \int_0^1 \left(\frac{1}{x^2} - \frac{3}{x} + \frac{9}{4}\right)(3x^2)\,dx \\ &= \int_0^1 3\,dx - \int_0^1 9x\,dx + \int_0^1 \frac{27}{4}x^2\,dx = \frac{3}{4} \end{aligned}$$

Approximating $E(h(X))$ and $\mathrm{Var}(h(X))$

This section is based on material in a text by P. L. Meyer.[2]

We have already shown that, given a random variable X and function $Y = h(X)$, we can compute $E(Y) = E(h(X))$ directly using only the probability distribution for X. This avoids the need for explicitly finding the probability distribution for the random variable $Y = h(X)$ if we are only computing expected values.

However, if the function $h(\cdot)$ is very complicated, then even using these short-cut methods for computing $E(h(X))$ and $\mathrm{Var}(h(X))$ may result in difficult integrals. Often, we can get good results using the following approximations:

Theorem 5.12. *Let* X *be a random variable with* $E(X) = \mu$ *and* $\mathrm{Var}(X) = \sigma^2$. *Suppose* $Y = h(X)$. *Then*

$$E(Y) \approx h(\mu) + \frac{h''(\mu)}{2}\sigma^2 \tag{2}$$

$$\mathrm{Var}(Y) \approx [h'(\mu)]^2\sigma^2 \tag{3}$$

[2] Meyer, P., *Introductory probability theory and statistical applications,* Addison-Wesley, Reading MA, 1965.

Proof. (This is only a sketch of the proof.)

Expand the function h in a Taylor series about $x = \mu$ to two terms. We obtain,

$$Y = h(\mu) + (X-\mu)h'(\mu) + \frac{(X-\mu)^2 h''(\mu)}{2} + R_1$$

where R_1 is the remainder of the expansion. Discard the remainder term R_1 and take the expected value of both sides to get Equation (2).

For the second approximation, expand h in a Taylor series about $x = \mu$ to only one term to get

$$Y = h(\mu) + (X-\mu)h'(\mu) + R_2$$

where R_2 is the remainder. Discard the term R_2 and take the variance of both sides to get Equation (3). ■

A common mistake when using expected values is to assume that $E(h(X)) = h(E(X))$. This is not generally true.

For example, suppose we are selling a product, and the random variable X represents the monthly demand (in terms of the number of items sold). Then $E(X) = \mu$ represents the expected number of items sold in a month.

Now suppose, $Y = h(X)$ is the actual profit from selling X items in a month. Often, the number $h(\mu)$ is incorrectly interpreted as the expected profit. That is, the expected sales figure is "plugged into" the profit function, believing the result is the expected profit.

In fact, the expected profit is actually $E(Y) = E(h(X))$. Unless h is linear, this will not be the same as $h(\mu) = h(E(X))$.

Equation (2) in Theorem 5.12 provides an approximation for the error if you incorrectly use $h(\mu)$ rather than $E(h(X))$.

$$\begin{aligned} E(Y) &\approx h(\mu) + \frac{h''(\mu)}{2}\sigma^2 \\ h(\mu) - E(Y) &\approx -\frac{h''(\mu)}{2}\sigma^2 \end{aligned}$$

Therefore, if you incorrectly state that $h(\mu)$ is your expected monthly profit, your answer will be wrong by approximately

$$\Delta = -\frac{h''(\mu)}{2}\sigma^2$$

Note that if h is linear, $h''(x) = 0$, and $\Delta = 0$. This result for the linear case is also provided by Theorems 5.5, 5.6 and 5.7.

Example: (Meyer) Under certain conditions, the surface tension of a liquid (dyn/cm) is given by the formula $S = 2(1 - 0.005T)^{1.2}$, where T is the temperature of the liquid (degrees centigrade).

Suppose T is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 3000t^{-4} & \text{if } t \geq 10 \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$\begin{aligned} E(T) &= \int_{10}^{\infty} 3000t^{-3}\,dt = 15 \quad \text{(degrees centigrade)} \\ \text{Var}(T) &= E(T^2) - (15)^2 \\ &= \int_{10}^{\infty} 3000t^{-2}\,dt - 225 = 75 \quad \text{(degrees centigrade)}^2 \end{aligned}$$

In order to use Equations (2) and (3), we will use

$$\begin{aligned} h(t) &= 2(1 - 0.005t)^{1.2} \\ h'(t) &= -0.012(1 - 0.005t)^{0.2} \\ h''(t) &= 0.000012(1 - 0.005t)^{-0.8} \end{aligned}$$

Since $\mu = 15$, we have

$$\begin{aligned} h(15) &= 1.82 \\ h'(15) &= 0.01 \\ h''(15) &= 0.000012(1 - 0.005(15))^{-0.8} \approx 0^+ \end{aligned}$$

and, using Theorem 5.12, results in

$$\begin{aligned} E(S) &\approx h(15) + 75h''(15) = 1.82 \quad \text{(dyne/cm)} \\ \text{Var}(S) &\approx 75[h'(15)]^2 = 0.87 \quad \text{(dyne/cm)}^2 \end{aligned}$$

Chebychev's Inequality

Chebychev's inequality, given in the following Theorem 5.13, is applicable for any random variable X with finite variance. For a given $\text{Var}(X)$, the inequality establishes a limit on the amount of probability mass that can be found at distances far away from $E(X)$ (the boondocks[3] of the probability distribution).

Theorem 5.13. *Let X be a discrete or continuous random variable with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Then for any $a > 0$*

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Proof. We will prove this result when X is a continuous random variable. (The proof for X discrete is similar using summations rather than integrals, and $f_X(x)\,dx$ replaced by the probability mass function for X.)

We have for any $a > 0$

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\,dx \\
&= \int_{-\infty}^{\mu-a} (x-\mu)^2 f_X(x)\,dx + \int_{\mu-a}^{\mu+a} (x-\mu)^2 f_X(x)\,dx + \int_{\mu+a}^{+\infty} (x-\mu)^2 f_X(x)\,dx \\
&= \int_{-\infty}^{\mu-a} (x-\mu)^2 f_X(x)\,dx + \boxed{\text{something} \geq 0} + \int_{\mu+a}^{+\infty} (x-\mu)^2 f_X(x)\,dx \\
&\geq \int_{-\infty}^{\mu-a} (x-\mu)^2 f_X(x)\,dx + \int_{\mu+a}^{+\infty} (x-\mu)^2 f_X(x)\,dx
\end{aligned}$$

since $\int_{\mu-a}^{\mu+a} (x-\mu)^2 f_X(x)\,dx \geq 0$.

Note that $(x-\mu)^2 \geq a^2$ for every $x \in (-\infty, \mu - a] \cup [\mu + a, +\infty)$. Hence, for all x over this region of integration, we have $(x-\mu)^2 f_X(x) \geq a^2 f_X(x)$. Therefore,

$$\begin{aligned}
\int_{-\infty}^{\mu-a} (x-\mu)^2 f_X(x)\,dx &\geq \int_{-\infty}^{\mu-a} a^2 f_X(x)\,dx \qquad \text{and} \\
\int_{\mu+a}^{+\infty} (x-\mu)^2 f_X(x)\,dx &\geq \int_{\mu+a}^{+\infty} a^2 f_X(x)\,dx.
\end{aligned}$$

[3]**boon·docks** (bo͞on'dŏks') *plural noun.* Rural country; the backwoods. *Example usage:* The 1965 Billy Joe Royal Song: "Down in the Boondocks;" *Source:* The American Heritage® Dictionary of the English Language, Fourth Edition, Houghton Mifflin Company.

So, we get

$$\begin{aligned}
\sigma^2 &\geq \int_{-\infty}^{\mu-a} (x-\mu)^2 f_X(x)\,dx + \int_{\mu+a}^{+\infty} (x-\mu)^2 f_X(x)\,dx \\
&\geq \int_{-\infty}^{\mu-a} a^2 f_X(x)\,dx + \int_{\mu+a}^{+\infty} a^2 f_X(x)\,dx \\
&= a^2 P(X \leq \mu - a) + a^2(P(X \geq \mu + a) \\
&= a^2 P(|X-\mu| \geq a)
\end{aligned}$$

■

Note: X can be discrete. An equivalent way to state Chebychev's inequality is

> Let X be a discrete or continuous random variable with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Then for any $a > 0$
>
> $$P(|X-\mu| < a) \geq 1 - \frac{\sigma^2}{a^2}$$

Example:[4]

The number of customers who visit a car dealer's showroom on a Saturday morning is a random variable, X. Using extensive prior experience, the manager, a former EAS305 student, has determined that the mean and variance of X are 18 and 6.25 respectively.

What is the probability that there will be between 8 and 28 customers on any given Saturday?

Solution We are given $E(X) = \mu = 18$ and $\text{Var}(X) = \sigma^2 = 6.25$, so

$$\begin{aligned}
P(8 < X < 28) &= P(8 - 18 < X - 18 < 28 - 18) \\
&= P(-10 < X - 18 < 10) \\
&= P(|X - 18| < 10) \\
&= P(|X - \mu| < 10) \geq 1 - \frac{6.25}{10^2} = 0.9375
\end{aligned}$$

This is the best statement we can make without knowing more about the distribution for X.

[4]Thanks to Dr. Linda Chattin at Arizona State University for this example.

The Law of Large Numbers

This is an important application of Chebychev's inequality.

Theorem 5.14. *Let an experiment be conducted n times, with each outcome of the experiment independent of all others. Let*

$$\begin{aligned} X_n &\equiv \textit{number of times the event } A \textit{ occurs} \\ p &\equiv P(A) \end{aligned}$$

Then for any $\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{X_n}{n} - p\right| < \epsilon\right) = 1.$$

The above result is the *weak law of large numbers.* The theorem guarantees that, as n gets large, the probability that the ratio X_n/n is observed to be "close" to p approaches one.

There is a related result called the *strong law of large numbers* that makes a similar (but stronger statement) about the entire sequence of random variables

$$\{X_1, \tfrac{1}{2}X_2, \tfrac{1}{3}X_3, \tfrac{1}{4}X_4, \ldots\}$$

The strong law states that, with probability one, this sequence converges to p. In this case, we often say that the sequence $\{\frac{1}{n}X_n\}$ converges to p **almost surely**.

The Normal Distribution

Definition 5.6. *The random variable X is said to have a* **normal distribution** *if it has probability density function*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \textit{for } -\infty < x < \infty$$

Note that X has two parameters, μ and σ.

Theorem 5.15. *If $X \sim N(\mu, \sigma^2)$ then $E(X) = \mu$. and* $\mathrm{Var}(X) = \sigma^2$.

We often say that "X is distributed normally with mean μ and variance σ^2"

We write

$$X \sim N(\mu, \sigma^2)$$

One of the amazing facts about the normal distribution is that the cumulative distribution function

(4) $$F_X(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx$$

does not exist in closed form. In other words, there exists no function in closed form whose derivative is

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

The only way to compute probabilities for normally distributed random variables is to evaluate the cumulative distribution function in Equation 4 using numerical integration.

If $X \sim N(\mu, \sigma^2)$, there is an important property of the normal distribution that allows us to convert any probability statement, such as $P(X \leq a)$, to an equivalent probability statement involving only a normal random variable with mean zero and variance one. We can then refer to a table of $N(0, 1)$ probabilities, and thus avoid numerically integrating Equation 4 for every conceivable pair of values μ and σ. Such a table is provided in Table I of Appendix A.

Definition 5.7. *The* **standard normal** *random variable, Z, has probability density function*

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

i.e., $Z \sim N(0, 1)$.

We denote the cumulative distribution function for Z by $\Phi(a) = P(Z \leq a)$.

Theorem 5.16. *If $X \sim N(\mu, \sigma^2)$, then the random variable*

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Proof. Suppose $X \sim N(\mu, \sigma^2)$. Therefore, the probability density function for X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty$$

Define the function

$$h(x) = \frac{x - \mu}{\sigma}$$

Note that $z = h(x)$ is an invertible, increasing function of x with

$$\begin{aligned} h^{-1}(z) &= \sigma z + \mu \\ \frac{d}{dz}h^{-1}(z) &= \sigma \end{aligned}$$

Using Theorem 5.2, the probability density function for the random variable $Z = h(X)$ is

$$\begin{aligned} f_Z(z) &= f_X(h^{-1}(z))\frac{d}{dz}h^{-1}(z) \\ &= \left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{(\sigma z+\mu)-\mu}{\sigma}\right)^2}\right]\sigma \quad \text{for } -\infty < \sigma z + \mu < \infty \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \quad \text{for } -\infty < z < \infty \end{aligned}$$

■

Note that we couldn't use the cumulative distribution function of X to find the cumulative distribution function of Z since neither cumulative distribution function exists in closed form.

If $X \sim N(\mu, \sigma^2)$, Theorem 5.16 tells us three things about $Z = \dfrac{X-\mu}{\sigma}$

1. $E(Z) = 0$
2. $\text{Var}(Z) = 1$, and
3. Z has a normal distribution

Actually, the first two results would be true even if X did not have a normal distribution:

Lemma 5.17. *Suppose X is any random variable with finite mean $E(X) = \mu$ and finite, nonzero variance* $\text{Var}(X) = \sigma^2$. *Define the random variable*

$$Z = \frac{X-\mu}{\sigma}.$$

Then $E(Z) = 0$ *and* $\text{Var}(Z) = 1$.

Proof. We have

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right]$$

$$
\begin{aligned}
&= E\left[\frac{1}{\sigma}(X-\mu)\right] \\
&= \frac{1}{\sigma}E[X-\mu] \\
&= \frac{1}{\sigma}(E[X]-\mu) \\
&= \frac{1}{\sigma}(0) = 0
\end{aligned}
$$

and

$$
\begin{aligned}
\text{Var}(Z) &= \text{Var}\left[\frac{X-\mu}{\sigma}\right] \\
&= \text{Var}\left[\frac{1}{\sigma}(X-\mu)\right] \\
&= \frac{1}{\sigma^2}\text{Var}[X-\mu] \\
&= \frac{1}{\sigma^2}\text{Var}[X] = 1
\end{aligned}
$$

■

So, it is no surprise that, in Theorem 5.16, $E(Z) = 0$ and $\text{Var}(Z) = 1$. What *is* important in Theorem 5.16 is that if X is normally distributed, then the linear transformation of X, namely,

$$Z = \frac{X-\mu}{\sigma}$$

is also normally distributed.

A distribution that has zero mean and unit variance is said to be in *standard form.* When we perform the above transformation (i.e., $Z = (X-\mu)/\sigma$), we say we are *standardizing* the distribution of X.)

With a little additional work we can show that id X is normally distributed, then any linear function $h(X) = aX + b$ is also normally distributed as long as $a \neq 0$. This fact is provided by the following theorem:

Theorem 5.18. *If $X \sim N(\mu, \sigma^2)$, then for any constants $a \neq 0$ and b, the random variable*

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Proof. This is just a special case of Theorem 5.2. If $a > 0$, then the function $Y = h(X)$ is strictly increasing and if $a < 0$, the function $Y = h(X)$ is strictly decreasing.

We know

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

Suppose $a > 0$, then

$$\begin{aligned} F_Y(y) = P(Y \le y) &= P(aX + b \le y) \\ &= P\left(X \le \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

Take derivatives on both sides and use the chain rule to obtain

$$\begin{aligned} f_Y(y) = \frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) \\ &= f_X\left(\frac{y-b}{a}\right)\frac{1}{a} \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2}\frac{1}{a} \\ &= \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2} \end{aligned}$$

Hence, $Y \sim N(a\mu + b, a^2\sigma^2)$.

Now suppose $a < 0$, then

$$\begin{aligned} F_Y(y) = P(Y \le y) &= P(aX + b \le y) \\ &= P\left(X \ge \frac{y-b}{a}\right) \\ &= P\left(X > \frac{y-b}{a}\right) + P\left(X = \frac{y-b}{a}\right) \\ &= 1 - F_X\left(\frac{y-b}{a}\right) + P\left(X = \frac{y-b}{a}\right) \\ &= \text{and since } X \text{ is continuous}\ldots \\ &= 1 - F_X\left(\frac{y-b}{a}\right) + 0 \end{aligned}$$

Take derivatives on both sides and use the chain rule to obtain

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(1 - F_X\left(\frac{y-b}{a}\right)\right)$$

$$\begin{aligned}
&= -f_X\left(\frac{y-b}{a}\right)\frac{1}{a} \\
&\quad \text{and remembering that } a < 0 \ldots \\
&= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2} \frac{1}{-a} \\
&= \frac{1}{-a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2} \\
&\quad \text{and since } a^2 = (-a)^2 \ldots \\
&= \frac{1}{-a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{-a\sigma}\right)^2}
\end{aligned}$$

So, again, we have $Y \sim N(a\mu + b, a^2\sigma^2)$. ■

Question: What happens to the probability distribution of $Y = aX + b$ when $a = 0$?

Linear transformations rarely preserve the "shape" of the probability distribution. For example, if X is exponentially distributed, Z would not be exponentially distributed. If X was a Poisson random variable, Z would not be a Poisson random variable. But, as provided by Theorem 5.16, if X is normally distributed, then Z is also normally distributed. (See Question 5.6 to see that the uniform distribution *is* preserved under linear transformations.)

Tools for using the standard normal tables

$P(Z \leq a) = \Phi(a)$ is given in the tables for various values of a. It can be shown that since Z is a continuous random variable and the probability density function for Z is symmetric about zero, we have

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

Example: Suppose $X \sim N(0, 4)$. Let's find $P(X \leq 1)$.

$$\begin{aligned}
P(X \leq 1) &= P(X - 0 \leq 1 - 0) \\
&= P\left(\frac{X-0}{2} \leq \frac{1-0}{2}\right)
\end{aligned}$$

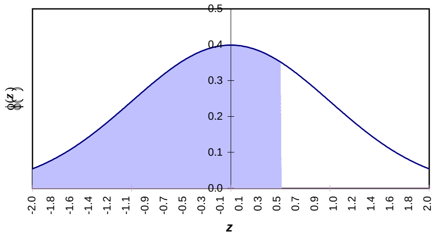
Since $X \sim N(0, 4)$ we know from Theorem 5.16 that

$$Z = \left(\frac{X-0}{2}\right) \sim N(0, 1)$$

so, using Table I in Appendix A, we have

$$\begin{aligned} P(X \le 1) &= P\left(Z \le \frac{1-0}{2}\right) \\ &= P\left(Z \le \tfrac{1}{2}\right) = 0.6915 \end{aligned}$$

This is the area shown in the following figure:



Example: Suppose $X \sim N(2, 9)$. Let's find $P(0 \le X \le 3)$.

$$\begin{aligned} P(0 \le X \le 3) &= P(0-2 \le X-2 \le 3-2) \\ &= P\left(\frac{0-2}{3} \le \frac{X-2}{3} \le \frac{3-2}{3}\right) \end{aligned}$$

Since $X \sim N(2, 9)$ we know that

$$Z = \left(\frac{X-2}{3}\right) \sim N(0, 1)$$

Hence,

$$\begin{aligned} P(0 \le X \le 3) &= P\left(-\tfrac{2}{3} \le Z \le \tfrac{1}{3}\right) \\ &= P\left(Z \le \tfrac{1}{3}\right) - P\left(Z < -\tfrac{2}{3}\right) \end{aligned}$$

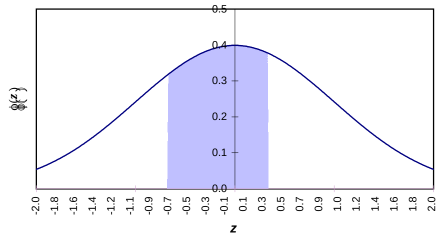
Since the $N(0, 1)$ distribution is symmetric about zero,

$$P\left(Z < -\tfrac{2}{3}\right) = P\left(Z > \tfrac{2}{3}\right) = 1 - P\left(Z \le \tfrac{2}{3}\right)$$

Finally, from Table I, we get

$$\begin{aligned} P(0 \le X \le 3) &= P\left(Z \le \tfrac{1}{3}\right) - \left(1 - P\left(Z \le \tfrac{2}{3}\right)\right) \\ &= 0.63 - (1 - 0.75) = 0.38 \end{aligned}$$

which is the area shown in the following figure:



Self-Test Exercises for Chapter 5

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S5.1 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The value of $P(X^2 \geq \frac{1}{4})$ is

(A) 0

(B) $\frac{1}{4}$

(C) $\frac{1}{2}$

(D) 1

(E) none of the above.

S5.2 Let X be a random variable with $E(X) = 0$ and $E(X^2) = 2$. Then the value of $E[X(1 - X)]$ is

(A) -2

(B) 0

(C) 1

(D) 2

(E) Var(X)

(F) none of the above.

S5.3 Let X be a random variable with $E(X) = 1$ and $E(X^2) = 2.$ Then Var(X) equals

(A) -1

(B) 0

(C) 1

(D) 3

(E) none of the above.

S5.4 Suppose that X is a continuous random variable with $P(X < -2) = \frac{1}{4}$ and $P(X < 2) = \frac{1}{2}$. Then $P(X^2 \leq 4)$ equals.

(A) $\frac{1}{4}$

(B) $\frac{1}{2}$

(C) $\frac{3}{4}$

(D) 1

(E) none of the above.

S5.5 Let X be a random variable with $E(X) = 2.$ Then $E[(X-2)^2]$ equals

(A) 0

(B) 2

(C) Var(X)

(D) $E(X^2)$

(E) none of the above.

S5.6 Let X be a random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Then $P(3X \leq 1)$ equals

(A) $1/6$

(B) $1/4$

(C) $1/3$

(D) $1/2$

(E) none of the above.

S5.7 Let X be a discrete random variable with probability mass function

$$p_X(-1) = p \qquad p_X(0) = 1 - 2p \qquad p_X(1) = p$$

where $0 < p < \frac{1}{2}$. The value of $P(X^2 = 1)$ is

(A) p^2

(B) $2p$

(C) $(1 - 2p)^2$

(D) 1

(E) none of the above.

S5.8 Let X be a discrete random variable with probability mass function

$$p_X(x) = \begin{cases} 1/4 & \text{if } x = -2, -1, 1 \text{ or } 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = 1/X$. Then $P(Y \leq \frac{1}{2})$ equals

(A) 0

(B) $1/4$

(C) $3/4$

(D) 1

(E) none of the above.

S5.9 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 1/4 & \text{if } -2 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = 1/X$. Then $P(Y \leq \frac{1}{2})$ equals

(A) 0

(B) $1/2$

(C) $3/4$

(D) 1

(E) none of the above.

S5.10 Let X be a random variable with

$$\begin{aligned} P(X < -1) &= 1/4 \\ P(X \leq +1) &= 3/4 \end{aligned}$$

Then $P(|X| > 1)$ equals

(A) $1/4$

(B) $1/2$

(C) $3/4$

(D) 1

(E) none of the above.

S5.11 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $P(|X| \leq \frac{1}{4})$ equals

(A) 0

(B) 1/4

(C) 1/2

(D) 1

(E) none of the above.

S5.12 Suppose that X is a random variable with $P(X > 0) = r$ where $0 < r < 1$. Then $P(X^2 \geq 0)$ equals

(A) 0

(B) $\sqrt{r}$

(C) r

(D) 1

(E) none of the above.

S5.13 Suppose X is a random variable with $E(X) = \alpha$ and $\text{Var}(X) = \alpha$, where $\alpha > 0$. Then $E(X^2)$ equals

(A) 0

(B) α

(C) α^2

(D) $\alpha + \alpha^2$

(E) none of the above.

S5.14 A fair coin is tossed 10 times, with each outcome independent of the others. Suppose you win $2 for every head and lose $1 for every tail, Let Y denote your actual net amount won (which may be negative). Then $P(Y \geq 0)$ equals

(A) $2/3$

(B) $848/1024$

(C) $904/1024$

(D) 1

(E) none of the above.

S5.15 The daily demand for chocolate cake in a neighborhood bakery is a random variable X with probability mass function given by

$$\begin{array}{ll} p_X(0) = 0.2 & p_X(1) = 0.5 \\ p_X(2) = 0.2 & p_X(3) = 0.1 \end{array}$$

where X represents the number of cakes requested by customers.

At the beginning of the day, the bakery owner bakes two chocolate cakes at a cost of $2.00 per cake. The cake is sold for $5.00 a cake and all unsold chocolate cakes are thrown away at the end of the day. The expected daily profit from chocolate cake is equal to

(A) 0

(B) \$1.50

(C) \$2.00

(D) \$3.00

(E) none of the above.

S5.16 The daily demand for rye bread in a corner grocery store is a random variable X with probability mass function given by

$$p_X(0) = 0.2 \qquad p_X(1) = 0.2 \qquad p_X(2) = 0.6$$

At the beginning of the day, the store owner bakes two loaves of rye bread at a cost of 20 cents per loaf. The bread is sold for 50 cents a loaf and all unsold loaves are thrown away at the end of the day. The expected daily profit from rye bread is equal to

(A) 0

(B) 20 cents

(C) 30 cents

(D) 40 cents

(E) none of the above.

S5.17 Suppose that an electronic device has a life length X (in units of 1000 hours) that is considered a continuous random variable with probability density function

$$f_X(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Suppose that the cost of manufacturing one such item is \$2.00. The manufacturer sells the item for \$5.00, but guarantees a total refund if $X \leq 1$. What is the manufacturer's expected profit per item?

(A) $5e^{-1}$

(B) $2e^{-1}$

(C) $-2 + 5e^{-1}$

(D) $5 - 2e^{-1}$

(E) none of the above.

S5.18 Let X be a random variable with $E(X) = 2$ and $E(X^2) = 5$. Then Var(X) equals

(A) 1

(B) 3

(C) 7

(D) 9

(E) none of the above.

S5.19 Let X be a random variable with $E(X) = 1$ and Var$(X) = 4$ then

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

equals

(A) 0

(B) 1

(C) 4

(D) 16

(E) none of the above.

S5.20 A fair coin is tossed 3 times. You win \$2 for each head that appears, and you lose \$1 for each tail that appears. Let Y denote your earnings after 3 tosses. The value of $E(Y)$ is

(A) 0

(B) $3/2$

(C) 1

(D) 3

(E) none of the above.

S5.21 Let X be *any* random variable with $E(X) = \mu$ and Var$(X) = \sigma^2$. Consider the random variable

$$Y = \frac{X - \mu}{\sigma}.$$

Which of the following are *always* true:

(A) $E(Y) = 0$ and Var$(Y) = 1$

(B) Y has a normal distribution

(C) $P(Y > 0) = 1/2$

(D) all of the above are true.

(E) none of the above are true.

S5.22 Let X be a normally distributed random variable with $E(X) = 3$ and Var$(X) = 9$. Then $P(X < 0)$ equals

(A) 0

(B) 0.1587

(C) 0.5

(D) 0.8413

(E) none of the above

S5.23 If Y is a random variable with $P(Y \geq 0) = p$, then $P(-2Y < 0)$ equals

(A) p

(B) $1 - p$

(C) 0.5

(D) 0

(E) none of the above.

S5.24 Let X be a random variable with

$$P(X = -1) = 0.5 \quad \text{and} \quad P(X = +1) = 0.5.$$

Then the value of $E(|X|)$ is

(A) 0

(B) 0.5

(C) 1

(D) 2

(E) none of the above are true.

S5.25 The time to wait for service in the *El Gasso Mexican Restaurant* is a continuous random variable X (in minutes) with probability density function

$$f_X(x) = \begin{cases} 5e^{-5x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The restaurant will pay a customer a \$10 reward if they wait more than 5 minutes for service. Let Y denote the reward for any one customer. The value of $E(Y)$, the expected reward, is

(A) 0

(B) 5

(C) 10

(D) $10e^{-25}$

(E) none of the above.

S5.26 Suppose that the random variable X has a normal distribution with mean 1 and variance 4. Then, $P(|X| \leq 1)$ equals

(A) 0.1528

(B) 0.1826

(C) 0.3174

(D) 0.3413

(E) none of the above.

S5.27 Suppose X is a continuous random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Let $\ln(x)$ denote the natural logarithm of x. Define the random variable $Y = \ln(X)$. Then $P(Y < \ln(0.50))$ equals

(A) 0

(B) 0.25

(C) 0.50

(D) $\sqrt{0.50}$

(E) none of the above.

S5.28 The life length X (in years) of a printer costing $200 is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The manufacturer agrees to pay a full refund to the buyer if the printer fails within the first year. If the manufacturer sells 100 printers, how much should he expect to pay in refunds?

(A) $2706.71

(B) $7357.59

(C) $12,642.41

(D) $17,293.29

(E) none of the above.

Questions for Chapter 5

5.1 Let X be the random variable with probability density function

$$f_X(x) = \begin{cases} \frac{3}{2}(1 - x^2) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $h(x) = 1/x$.

(a) Find the probability density function for the random variable $Y = h(X) = 1/X$.

(b) Find $E(X)$, $E(h(X))$ and $h(E(X))$.

(c) Find Var(X).

5.2 Let X be the random variable with probability density function

$$f_X(x) = \begin{cases} 2/x^2 & \text{if } 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $h(x) = \ln x$.

(a) Find the probability density function for the random variable $Y = h(X) = \ln X$.

(b) Find $E(X)$, $E(h(X))$ and $h(E(X))$.

(c) Find Var(X).

5.3 Let X be the random variable with probability density function

$$f_X(x) = \begin{cases} \frac{3}{2}\sqrt{x} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $h(x) = \frac{4}{9}x$.

(a) Find the probability density function for the random variable $Y = h(X) = \frac{4}{9}X$.

(b) Find $E(X)$, $E(h(X))$ and $h(E(X))$.

(c) Find Var(X).

5.4 The price, X, of a commodity is a continuous random variable with the following probability density function

$$f_X(x) = \begin{cases} ax & \text{if } 0 \leq x \leq 1 \\ a & \text{if } 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine the value of a.

(b) Determine the cumulative distribution function for X.

(c) Compute $P(2X^2 \leq 0.5)$.

(d) Compute $E(X)$ and Var(X).

5.5 Let X be a random variable with the following probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for } -\infty < x < \infty$$

The is the normal distribution with parameters $\mu = 0$ and $\sigma^2 = 1$. Find the probability density function for the random variable $Y = X^3$. *Note:* The integral for $f_X(x)$ does not exist in closed form.

5.6 (†) Let X be a uniformly distributed random variable on the interval $[a, b]$, where $a < b$. That is, suppose the cumulative distribution function of X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b. \end{cases}$$

Let

$$Z = \frac{X - r}{s}$$

Where r and s are non-zero constants. Show that Z also has a uniform distribution. *Note:* Be careful of the case where $s < 0$.

5.7 Verify the results of Example 5.5 using the actual probability density function for Y. In other words, let X be the random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the probability density function for the random variable $Y = h(X) = 1/X$.

(b) Using the probability density function for Y find $E(Y)$ and $\text{Var}(Y)$. The results should be the same as Example 5.5.

5.8 Suppose that an electronic device has a life length X (in units of 1000 hours) that is considered as a continuous random variable with probability density function:

$$f_X(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Suppose that the cost of manufacturing one such item is \$2.00. The manufacturer sells the item for \$5.00 but guarantees a total refund if $X \leq 0.9$. What is the manufacturer's expected profit per item?

5.9 Suppose that the continuous random variable X has a probability density function given by

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-x^2} & -\infty \leq x \leq +\infty \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X^2$. Evaluate $E(Y)$.

5.10 Let X be a random variable with density

$$f_X(x) = \begin{cases} \frac{2}{x^2} & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Using the fact that $\frac{d}{dx} \ln(x) = \frac{1}{x}$, compute $E(X)$ and $\mathrm{Var}(X)$.

(b) Compute $E(Y)$, where $Y = X^2 + 1$.

5.11 John and Mary play a game by tossing a fair coin repeatedly and betting on the number of tosses required, X, to obtain the first head. John will pay Mary 2^X dollars when the game is over (i.e., when the first head appears). To make the game fair, Mary will pay an "entry fee" in advance to John equal to her expected winnings. Therefore, Mary's entry fee will be $E(2^X)$.

(a) What is the expected number of tosses required to obtain the first head?

(b) What is Mary's entry fee? Is this fair? Why?

5.12 Due to pressing social commitments, Bill Bell, an industrial engineering student, was not able to study for an examination. The examination consists of 4 multiple choice questions. Each question has 6 possible answers from which to choose, only one of which is correct. Billy plans to answer the questions randomly by rolling a fair die for each question and choosing the response associated with the number rolled. Let X denote the number of questions correctly answered by Billy.

(a) What is the probability mass function for the random variable X?

(b) If the professor assigns a credit of ten points for every correct answer and deducts one point for every incorrect answer, what is Billy's expected score?

(c) The professor deducts points for wrong answers in order to discourage random guessing. If 10 points are credited for each correct answer, how many points should be deducted for each wrong answer so that Billy's expected score from guessing is zero?

5.13 Let X be the number of dots obtained when a fair, common six-sided die is tossed. For what number α is $E(X(\alpha - X))$ zero?

5.14 *(True or False)* If X is a random variable with $E(X) = 1$, then $E(X(X-1))$ equals zero. Justify your answer.

5.15 If X is a random variable with $P(X = -2) = \frac{1}{2}$ and $P(X = +2) = \frac{1}{2}$, what is the value of $E(|X|)$?

5.16 The daily demand for pumpernickel bread in a corner grocery store is a random variable X with probability mass function given by

$$p_X(0) = 0.3; \qquad p_X(1) = 0.1; \qquad p_X(2) = 0.6$$

At the beginning of the day, the store owner bakes k loaves of bread at a cost of 20 cents for the first loaf and 10 cents for each additional loaf. The bread is sold for 30 cents a loaf and all unsold loaves are thrown away at the end of the day.

(a) Determine the probability distribution for the daily profit, R, if the store owner decides to bake one loaf (i.e., $k = 1$).

(b) How many loaves should the grocer bake per day to maximize his expected profit?

5.17 A random variable X represents the weight (in ounces) of an article, and has a uniform distribution on the interval $(36, 56)$. If the weight of the article is less than 44 ounces, it is considered unsatisfactory and thrown away. If its weight is greater than 44 ounces and less than 52 ounces, it is sold for \$12.00. If its weight is greater than 52 ounces, it is sold for \$15.00. What is the expected selling price of an article?

5.18 A distribution closely related to the normal distribution is the *log-normal distribution*. Suppose that $X \sim N(\mu_X, \sigma_X^2)$. Let $Y = e^X$. Then Y is said to have a log-normal distribution. That is, Y is log-normal if and only if $\ln(Y)$ is normal. Find the probability density function for Y.

Note: For this question, your answer is in terms of μ_X and σ_X. These are *not* the expected value and standard deviation of Y. The relationship between the moments of X and the moments of Y is given by

$$\begin{aligned}\mu_X &= \ln(\mu_Y) - \frac{1}{2}\ln\left(\frac{\sigma_Y^2}{\mu_Y^2}+1\right) = \ln(\mu_Y) - \frac{1}{2}\sigma_X^2 \\ \sigma_X^2 &= \ln\left(\frac{\sigma_Y^2}{\mu_Y^2}+1\right)\end{aligned}$$

Examples of random variables that may be represented by the log-normal distribution are river flows, the diameter of a small particle after a crushing process, the size of an organism subject to a number of small impulses, and the length of life of certain items.

5.19 A rocket fuel is to contain a certain percent (say X) of a particular additive. The specifications call for X to be between 30 and 35 percent. The manufacturer will make a net profit on the fuel (per gallon) which is the following function of X:

$$T(X) = \begin{cases} \text{\$0.10 per gallon} & \text{if } 30 < X < 35 \\ \text{\$0.05 per gallon} & \text{if } 35 < X < 40 \text{ or } 25 < X \leq 30 \\ \text{-\$0.10 per gallon} & \text{otherwise} \end{cases}$$

(a) If $X \sim N(33, 9)$, evaluate $E(T)$.

(b) Suppose that the manufacturer wants to increase his expected profit, $E(T)$, by 50%. He intends to do this by increasing his profit (per gallon) on those batches of fuel meeting the specification $30 < X < 35$. What must his new profit function be in order to achieve this goal?

5.20 The Youbee Resistor Company uses a machine to produce resistors. The actual resistance of each resistor is a normally distributed random variable with mean resistance of 1000 ohms and a variance of 625 ohms^2. The cost of producing one resistor is 10 cents. If a resistor has a resistance between 950 and 1050 ohms, it can be sold for 25 cents, otherwise it is scrapped.

(a) What proportion of the resistors must be scrapped?

(b) The company is considering the purchase of a new high-precision machine that will reduce the above production variance to 400 ohms^2 (the mean remains the same). Although the selling price of resistors would remain the same, the cost of producing each resistor will increase to 15 cents if the company decides to purchase the machine (the new machine's capital cost is included in this production cost). If the company wishes to maximize expected profit, should the new machine be purchased?

5.21 Suppose that the breaking strength of cotton fabric (in pounds), say X, is normally distributed with $E(X) = 165$ and $\text{Var}(X) = 9$. Assume furthermore that a sample of this fabric is considered to be defective if $X \leq 162$. What is the probability that a sample of fabric, chosen at random, will be defective?

Note: An immediate objection to the use of the normal distribution may be raised here. It is obvious that X, the strength of cotton fabric, cannot take on negative values. However, a normally distributed random variable may be observed to take on any positive or *negative* value. Nevertheless, the above model (apparently not valid in view of the objection just raised) assigns negligible probability to the event $\{X < 0\}$. That is $P(X < 0) \approx 0$. Compute this probability to verify this property for yourself in the above example.

This situation will occur frequently: A certain random variable that we know cannot take on negative values will be assumed to have a normal distribution anyway. As long as the parameters μ and σ^2 are chosen so that $P(X < 0)$ is essentially zero, such a representation *may* work. In the above example, suppose we hold $E(X)$ fixed at 165. What would $\text{Var}(X)$ have to equal so that $P(X < 0) = 0.10$? Suppose, on the other hand, that we hold $\text{Var}(X)$ fixed at 9. What would $E(X)$ have to be so that $P(X < 0) = 0.10$?

5.22 Let $X_i \sim N(\mu, \sigma^2)$. Find k such that $P(X > \mu - k\sigma) = 0.841$; $= 0.95$; $= 0.975$; $= 0.99$.

5.23 A measuring device provides a reading on a continuous scale whenever a button is pressed. If a signal is present when the button is pressed, the reading, X, will have a normal distribution with mean 10 and variance 9. If only noise is present, the reading will have a normal distribution with mean 5 and variance 9. You settle on the following signal detection rule: If the reading is 7 or greater, you will announce that a signal is present. Otherwise, you will say only noise is present.

(a) If a signal is truly present, and you use the above rule, what is the probability that your announcement is wrong? If only noise is really present, what is the probability that your announcement is wrong?

(b) Suppose that the probability that a signal is truly present when the button is pressed is 0.25. If your reading is greater than 7, what is the probability that a signal is really present?

(c) The value, 7, is called the *threshold*. What should this value be changed to so that your two answers for question (a) are equal?

5.24 An electronic device has a life length X (in years) that is considered a continuous random variable with probability density function

$$f_X(x) = \begin{cases} x/2 & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X > 1)$.

(b) Find $E(X)$.

(c) Suppose that the cost of manufacturing one such item is \$2.00. The manufacturer sells the item for \$10.00, but guarantees a total refund if $X \leq 1$. What is the manufacturer's expected profit per item?

(d) Suppose that the cost of manufacturing one such item is \$2.00 and the customer is guaranteed a total refund if $X \leq 1$. What should be the manufacturer's selling price in order to provide an expected profit of \$4.00 per item?

5.25 The reading of a water level meter is a continuous random variable X with probability density function

$$f_X(x) = \begin{cases} 1/4 & \text{if } -2 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X \leq 1)$.

(b) Find $P(|X| \leq 1)$.

(c) Find $E(X)$.

(d) Find $\text{Var}(X)$.

(e) Let $Y = X^2$. Find the cumulative distribution function for Y.

5.26 The daily demand for pumpkin pies at a corner grocery store is a random variable with probability mass function

$$p_X(k) = \begin{cases} \frac{1}{16}\binom{4}{k} & \text{for } k = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

The cost of baking one pumpkin pie is \$1.00. The store sells the pies for \$3.00, each. At the beginning of every day, the store manager must decide how many pies to bake. At the end of the day, any unsold pies are thrown away. If the demand for pies exceeds the supply, you can only sell those pies that you have already baked for that day. The additional customers leave unsatisfied.

(a) Find $P(X > 0)$.

(b) Find $E(X)$.

(c) Find $\text{Var}(X)$.

(d) If the manager decides to bake three pies, what is the expected profit for that day?

5.27 The input voltage for an electronic component is a continuous random variable X with probability density function

$$f_X(x) = \begin{cases} 1/3 & \text{if } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X \leq 1)$.

(b) Find $P(X^2 \leq 0.25)$.

(c) Find $E(X^2)$.

(d) Let $Y = 2X$. Find the cumulative distribution function for Y.

5.28 The *Youbee Mine* has been used to mine gold ore for several years. The mass (in grams) of actual gold in one cubic centimeter of ore is a random variable

X with probability density function

$$f_X(x) = \begin{cases} cx^3 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of the constant c.

(b) Find the cumulative distribution function for X **and** sketch a graph of the cumulative distribution function.

(c) Each cubic centimeter or ore is worth \$5 if it has more than $\frac{1}{2}$ gram of gold. It is worth \$0 otherwise. Let Y be a random variable denoting the value of a cubic centimeter of ore selected at random. What is the probability mass function for Y?

5.29 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 1/3 & \text{if } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X^2$. Find the cumulative distribution function for Y.

5.30 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 1 & \text{if } -1 < x < 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X^2$. Find $P(Y \leq 0.25)$.

5.31 Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} a(1+x) & \text{if } -1 \leq x < 0 \\ a(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Prove that the value of a must be 1. Use a clear mathematical presentation to show why this must be true.

(b) Find $P(X \leq 0.5)$.

(c) Find $P(X^2 \leq 0.25)$.

(d) Find the cumulative distribution function for X.

5.32 A person accused of a crime is being questioned by the police. A new "lie detector" device is being used that measures the amount of stress in a person's voice. When the person speaks, the device gives a reading, X. If the person is telling the truth, then $X \sim N(2, 1)$. If the person is lying, then $X \sim N(4, 1)$. Therefore, the mean (expected) stress level of a person's voice increases if he is lying.

(a) The police decide on the following rule: If a person's voice stress exceeds 4 when he is answering a question, conclude that he is lying. What is the probability of incorrectly accusing person of lying when he is telling the truth?

(b) When the suspect was originally brought in for questioning, the police thought there was only a 50% chance that he was guilty of the crime. The police ask him the question, "Did you commit the crime?" The suspect replies, "No." The reading on the voice stress meter exceeds 4. What is the probability that he committed the crime?

5.33 Consider a random variable X with possible outcomes, $0, 1, 2, \ldots$ Suppose that

$$P(X = k) = (1 - a)(a^k) \qquad \text{for } k = 0, 1, 2, \ldots$$

(a) Identify all permissible values for the constant a.

(b) Show that for any two positive integers, m and n,

$$P(X > m + n \,|\, X > m) = P(X \geq n).$$

5.34 Check whether the mean and the variance of the following distributions exist:

(a)

$$f_X(x) = \begin{cases} 2x^{-3} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) (The Cauchy distribution) a is a positive constant.

$$f_X(x) = \frac{a}{\pi(a^2 + x^2)} \qquad -\infty < x < \infty.$$

5.35 Diameters of bearings in mass production are normally distributed with mean 0.25 inch and standard deviation 0.01 inch. Bearing specifications call for diameters of 0.24 ± 0.02 inch. What is the probability of a defective bearing being produced?

5.36 Suppose that the continuous random variable X has a probability density function given by

$$f_X(x) = \begin{cases} \dfrac{2}{\sqrt{2\pi}} e^{-x^2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X^2$. Evaluate $E(Y)$.

5.37 Suppose that the life lengths of two electronic devices, D_1 and D_2, have distributions $N(40, 36)$ and $N(45, 9)$, respectively. If the electronic device is to be used for a 45-hour period, which device is to be preferred? If it is to be used for a 50-hour period, which device is to be preferred?

5.38 (†) Suppose random variable Z has a standard normal density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{for } -\infty < z < \infty$$

Let $Y = Z^2$. Find the probability density function for Y. (We say that Y has a *Chi-square distribution with one degree of freedom.*) *Warning:* Y is not strictly increasing or strictly decreasing.

6

RANDOM VECTORS

Basic Concepts

Often, a single random variable cannot adequately provide all of the information needed about the outcome of an experiment. For example, tomorrow's weather is really best described by an array of random variables that includes wind speed, wind direction, atmospheric pressure, relative humidity and temperature. It would not be either easy or desirable to attempt to combine all of this information into a single measurement.

We would like to extend the notion of a random variable to deal with an experiment that results in several observations each time the experiment is run. For example, let T be a random variable representing tomorrow's maximum temperature and let R be a random variable representing tomorrow's total rainfall. It would be reasonable to ask for the probability that tomorrow's temperature is greater than 70° *and* tomorrow's total rainfall is less than 0.1 inch. In other words, we wish to determine the probability of the event

$$A = \{T > 70,\ R < 0.1\}.$$

Another question that we might like to have answered is, "What is the probability that the temperature will be greater than 70° regardless of the rainfall?" To answer this question, we would need to compute the probability of the event

$$B = \{T > 70\}.$$

In this chapter, we will build on our probability model and extend our definition of a *random variable* to permit such calculations.

Definition

The first thing we must do is to precisely define what we mean by "an array of random variables."

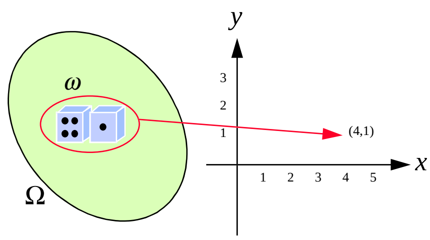
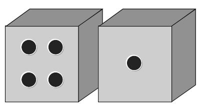


Figure 6.1: A mapping using (X_1, X_2)

Definition 6.1. *Let Ω be a sample space. An* **n-dimensional random variable** *or* **random vector**, *$(X_1(\cdot), X_2(\cdot), \ldots, X_n(\cdot))$, is a vector of functions that assigns to each point $\omega \in \Omega$ a point $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ in n-dimensional Euclidean space.*

Example: Consider an experiment where a die is rolled twice. Let X_1 denote the number of the first roll, and X_2 the number of the second roll. Then (X_1, X_2) is a two-dimensional random vector. A possible sample point in Ω is



that is mapped into the point $(4, 1)$ as shown in Figure 6.1.

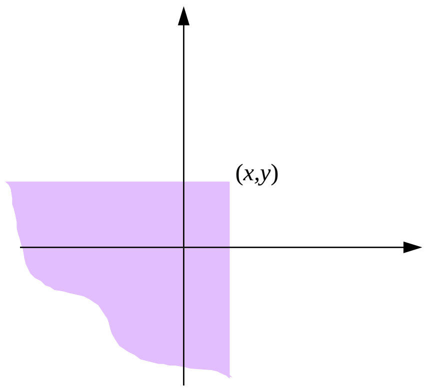


Figure 6.2: The region representing the event $P(X \leq x, Y \leq y)$.

Joint Distributions

Now that we know the definition of a random vector, we can begin to use it to assign probabilities to events. For any random vector, we can define a joint cumulative distribution function for all of the components as follows:

Definition 6.2. *Let* $(X_1, X_2, \ldots, X_n)$ *be a random vector. The* **joint cumulative distribution function** *for this random vector is given by*

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \equiv P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n).$$

In the two-dimensional case, the joint cumulative distribution function for the random vector (X, Y) evaluated at the point (x, y), namely

$$F_{X,Y}(x, y),$$

is the probability that the experiment results in a two-dimensional value within the shaded region shown in Figure 6.2.

Every joint cumulative distribution function must posses the following properties:

1. $\displaystyle\lim_{\text{all } x_i \to -\infty} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = 0$

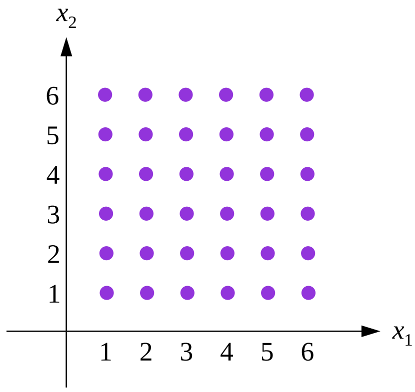


Figure 6.3: The possible outcomes from rolling a die twice.

2. $\lim_{\text{all } x_i \to +\infty} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = 1$

3. As x_i varies, with all other x_j's ($j \neq i$) fixed,

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

   is a nondecreasing function of x_i.

As in the case of one-dimensional random variables, we shall identify two major classifications of vector-valued random variables: *discrete* and *continuous*. Although there are many common properties between these two types, we shall discuss each separately.

Discrete Distributions

A random vector that can only assume at most a countable collection of discrete values is said to be **discrete**. As an example, consider once again the example on page 156 where a die is rolled twice. The possible values for either X_1 or X_2 are in the set $\{1, 2, 3, 4, 5, 6\}$. Hence, the random vector (X_1, X_2) can only take on one of the 36 values shown in Figure 6.3.

If the die is fair, then each of the points can be considered to have a probability mass of $\frac{1}{36}$. This prompts us to define a joint probability mass function for this type of random vector, as follows:

Definition 6.3. *Let $(X_1, X_2, \ldots X_n)$ be a discrete random vector. Then*

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \equiv P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

is the **joint probability mass function** *for the random vector $(X_1, X_2, \ldots, X_n)$.*

Referring again to the example on page 156, we find that the joint probability mass function for (X_1, X_2) is given by

$$p_{X_1,X_2}(x_1, x_2) = \tfrac{1}{36} \qquad \text{for } x_1 = 1, 2, \ldots, 6 \text{ and } x_2 = 1, 2, \ldots, 6$$

Note that for any probability mass function,

$$F_{X_1,X_2,\ldots,X_n}(b_1, b_2, \ldots, b_n) = \sum_{x_1 \le b_1} \sum_{x_2 \le b_2} \cdots \sum_{x_n \le b_n} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n).$$

Therefore, if we wished to evaluate $F_{X_1,X_2}(3.0, 4.5)$ we would sum all of the probability mass in the shaded region shown in Figure 6.3, and obtain

$$F_{X_1,X_2}(3.0, 4.5) = 12\tfrac{1}{36} = \tfrac{1}{3}.$$

This is the probability that the first roll is less than or equal to 3 *and* the second roll is less than or equal to 4.5.

Every joint probability mass function must have the following properties:

1. $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n) \ge 0$ for any $(x_1, x_2, \ldots, x_n)$.
2. $\displaystyle\sum_{\text{all } x_1} \cdots \sum_{\text{all } x_n} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n) = 1$
3. $\displaystyle P(E) = \sum_{(x_1,\ldots,x_n) \in E} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n)$ for any event E.

You should compare these properties with those of probability mass functions for single-valued discrete random variables given in Chapter 5.

Continuous Distributions

Extending our notion of probability density functions to continuous random vectors is a bit tricky and the mathematical details of the problem is beyond the scope of an introductory course. In essence, it is not possible to define the joint density as a derivative of the cumulative distribution function as we did in the one-dimensional case.

Let $\mathbb{R}^n$ denote n-dimensional Euclidean space. We sidestep the problem by defining the density of an n-dimensional random vector to be a function that *when integrated* over the set

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \,:\, x_1 \leq b_1, x_2 \leq b_2, \ldots, x_n \leq b_n\}$$

will yield the value for the cumulative distribution function evaluated at

$$(b_1, b_2, \ldots, b_n).$$

More formally, we have the following:

Definition 6.4. *Let $(X_1, \ldots, X_n)$ be a continuous random vector with joint cumulative distribution function*

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$

The function $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ that satisfies the equation

$$F_{X_1,\ldots,X_n}(b_1, \ldots, b_n) = \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} f_{X_1,\ldots,X_n}(t_1, \ldots, t_n)\, dt_1 \cdots dt_n$$

for all $(b_1, \ldots, b_n)$ is called the **joint probability density function** *for the random vector $(X_1, \ldots, X_n)$.*

Now, instead of obtaining a derived relationship between the density and the cumulative distribution function by using integrals as anti-derivatives, we have enforced such a relationship by the above definition.

Every joint probability density function must have the following properties:

1. $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n) \geq 0$ for any $(x_1, x_2, \ldots, x_n)$.
2. $\displaystyle\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$
3. $\displaystyle P(E) = \int \cdots \int_E f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$ for any event E.

You should compare these properties with those of probability density functions for single-valued continuous random variables given in Chapter 5.

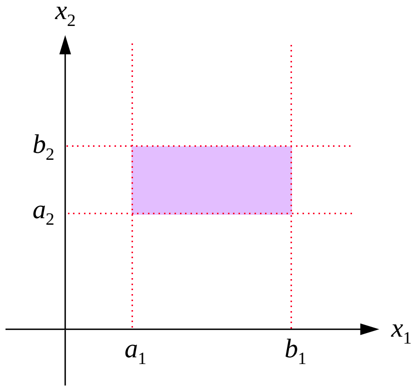


Figure 6.4: Computing $P(a_1 < X_1 \le b_1,\, a_2 < X_2 \le b_2)$.

In the one-dimensional case, we had the handy formula

$$P(a < X \le b) = F_X(b) - F_X(a).$$

This worked for any type of probability distribution. The situation in the multi-dimensional case is a little more complicated, with a comparable formula given by

$$\begin{aligned} &P(a_1 < X_1 \le b_1,\, a_2 < X_2 \le b_2) = \\ &\quad F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2). \end{aligned}$$

You should be able to verify this formula for yourself by accounting for all of the probability masses in the regions shown in Figure 6.4.

Example: Let (X, Y) be a two-dimensional random variable with the following joint probability density function (see Figure 6.5):

$$f_{X,Y}(x, y) = \begin{cases} 2 - y & \text{if } 0 \le x \le 2 \text{ and } 1 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Note that

$$\int_1^2 \int_0^2 (2 - y)\, dx\, dy = 1.$$

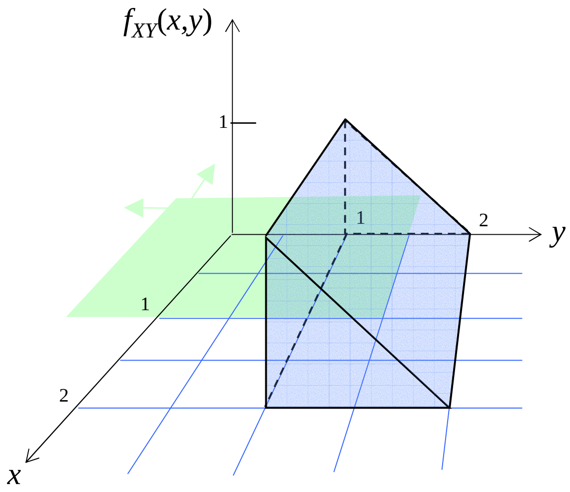


Figure 6.5: A two-dimensional probability density function

Suppose we would like to compute $P(X \leq 1.0,\, Y \leq 1.5)$. To do this, we calculate the volume under the surface $f_{X,Y}(x,y)$ over the region $\{(x,y) \,:\, x \leq 1,\, y \leq 1.5\}$. This region of integration is shown shaded (in green) in Figure 6.5. Performing the integration, we get,

$$\begin{aligned} P(X \leq 1.0,\, Y \leq 1.5) &= \int_{-\infty}^{1.5} \int_{-\infty}^{1.0} f_{X,Y}(x,y)\, dx\, dy \\ &= \int_{1.0}^{1.5} \int_{0}^{1.0} (2-y)\, dx\, dy = \tfrac{3}{8}. \end{aligned}$$

Marginal Distributions

Given the probability distribution for a vector-valued random variable $(X_1, \ldots X_n)$, we might ask the question, "Can we determine the distribution of X_1, disregarding the other components?" The answer is yes, and the solution requires the careful use of English rather than mathematics.

For example, in the two-dimensional case, we may be given a random vector (X,Y) with joint cumulative distribution function $F_{X,Y}(x,y)$. Suppose we would like to find the cumulative distribution function for X alone, i.e., $F_X(x)$? We know that

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

and we are asking for

(1) $$F_X(x) = P(X \le x).$$

But in terms of both X and Y, expression 1 can be read: "the probability that X takes on a value less than or equal to x and Y takes on *any* value." Therefore, it would make sense to say

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= P(X \le x,\, Y \le \infty) \\ &= \lim_{y\to\infty} F_{X,Y}(x,y). \end{aligned}$$

Using this idea, we shall define what we will call the *marginal cumulative distribution function:*

Definition 6.5. *Let* $(X_1, \ldots, X_n)$ *be a random vector with joint cumulative distribution function* $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$*. The* **marginal cumulative distribution function** *for* X_1 *is given by*

$$F_{X_1}(x_1) = \lim_{x_2\to\infty} \lim_{x_3\to\infty} \cdots \lim_{x_n\to\infty} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n).$$

Notice that we can renumber the components of the random vector and call any one of them X_1. So we can use the above definition to find the marginal cumulative distribution function for any of the X_i's.

Although Definition 6.5 is a nice definition, it is more useful to examine *marginal probability mass functions* and *marginal probability density functions.* For example, suppose we have a discrete random vector (X,Y) with joint probability mass function $p_{X,Y}(x,y)$. To find $p_X(x)$, we ask "What is the probability that $X = x$ regardless of the value that Y takes on? This can be written as

$$\begin{aligned} p_X(x) = P(X = x) &= P(X = x,\, Y = \text{any value}) \\ &= \sum_{\text{all } y} p_{X,Y}(x,y). \end{aligned}$$

Example: In the die example (on page 156)

$$p_{X_1,X_2}(x_1,x_2) = \tfrac{1}{36} \qquad \text{for } x_1 = 1,2,\ldots,6 \text{ and } x_2 = 1,2,\ldots,6$$

To find $p_{X_1}(2)$, for example, we compute

$$p_{X_1}(2) = P(X_1 = 2) = \sum_{k=1}^{6} p_{X_1,X_2}(2,k) = \tfrac{1}{6}.$$

Table 6.1: Joint pmf for daily production

$p_{X,Y}(x,y)$		Y 1	2	3	4	5
	1	.05	0	0	0	0
	2	.15	.10	0	0	0
X	3	.05	.05	.10	0	0
	4	.05	.025	.025	0	0
	5	.10	.10	.10	.10	0

This is hardly a surprising result, but it brings some comfort to know we can get it from all of the mathematical machinery we've developed thus far.

Example: Let X be the total number of items produced in a day's work at a factory, and let Y be the number of defective items produced. Suppose that the probability mass function for (X, Y) is given by Table 6.1. Using this joint distribution, we can see that the probability of producing 2 items with exactly 1 of those items being defective is

$$p_{X,Y}(2,1) = 0.15.$$

To find the marginal probability mass function for the total daily production, X, we sum the probabilities over all possible values of Y for each fixed x:

$$\begin{aligned} p_X(1) &= p_{X,Y}(1,1) = 0.05 \\ p_X(2) &= p_{X,Y}(2,1) + p_{X,Y}(2,2) = 0.15 + 0.10 = 0.25 \\ p_X(3) &= p_{X,Y}(3,1) + p_{X,Y}(3,2) + p_{X,Y}(3,3) = 0.05 + 0.05 + 0.10 = 0.20 \\ &\text{etc.} \end{aligned}$$

But notice that in these computations, we are simply adding the entries all columns for each row of Table 6.1. Doing this for Y as well as X we can obtain Table 6.2 So, for example, $P(Y = 2) = p_Y(2) = 0.275$. We simply look for the result in the margin[1] for the entry $Y = 2$.

The procedure is similar for obtaining marginal probability density functions. Recall that a density, $f_X(x)$, itself is not a probability measure, but $f_X(x)dx$, is. So

[1]Would you believe that this is why they are called *margin*al distributions?

Table 6.2: Marginal pmf's for daily production

$p_{X,Y}(x,y)$		Y 1	2	3	4	5	$p_X(x)$
	1	.05	0	0	0	0	.05
	2	.15	.10	0	0	0	.25
X	3	.05	.05	.10	0	0	.20
	4	.05	.025	.025	0	0	.10
	5	.10	.10	.10	.10	0	.40
$p_Y(y)$		0.4	.275	.225	.10	0	

with a little loose-speaking integration notation we should be able to compute

$$\begin{aligned} f_X(x)\,dx &= P(x \leq X < x + dx) \\ &= P(x \leq X < x + dx,\, Y = \text{any value}) \\ &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy\,dx \end{aligned}$$

where y is the variable of integration in the above integral. Looking at this relationship as

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

it would seem reasonable to define

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

and we therefore offer the following:

Definition 6.6. *Let $(X_1, \ldots, X_n)$ be a continuous random variable with joint probability density function $f_{X_1,\ldots,X_n}$. The* **marginal probability density function** *for the random variable X_1 is given by*

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)\,dx_2 \cdots dx_n.$$

Notice that in both the discrete and continuous cases, we sum (or integrate) over all possible values of the unwanted random variable components.

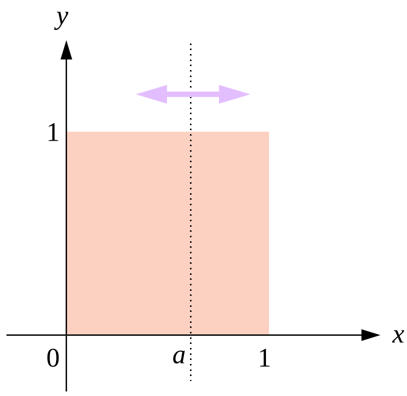


Figure 6.6: The support for the random vector (X, Y) in the example

The trick in such problems is to insure that your limits of integration are correct. Drawing a picture of the region where there is positive probability mass (the *support* of the distribution) often helps.

For the above example, the picture of the support would be as shown in Figure 6.6. If the dotted line in Figure 6.6 indicates a particular value for x (call it a), by integrating over all values of y, we are actually determining how much probability mass has been placed along the line $x = a$. The integration process assigns all of that mass to $x = a$ in one-dimension. Repeating this process for each x yields the desired probability density function.

Example: Let (X, Y) be a two-dimensional continuous random variable with joint probability density function

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

Find the marginal probability density function for X.

Solution: Let's consider the case where we fix x so that $0 \leq x \leq 1$. We compute

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

$$
\begin{aligned}
&= \int_0^1 (x+y)\,dy \\
&= \left(xy + \frac{y^2}{2} \right)\Bigg|_{y=0}^{y=1} \\
&= x + \frac{1}{2}
\end{aligned}
$$

If x is outside the interval $[0,1]$, we have $f_X(x) = 0$.

So, summarizing these computations we find that

$$
f_X(x) = \begin{cases} x + \frac{1}{2} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}
$$

We will leave it to you to check that $f_X(x)$ is, in fact, a probability density functionby making sure that $f_X(x) \ge 0$ for all x and that $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$.

Example: Let (X, Y) be the two-dimensional random variable with the following joint probability density function (see Figure 6.5 on page 162)

$$
f_{X,Y}(x,y) = \begin{cases} 2 - y & \text{if } 0 \le x \le 2 \text{ and } 1 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}
$$

Find the marginal probability density function for X and the marginal probability density function for Y.

Solution: Let's first find the marginal probability density function for X. Consider the case where we fix x so that $0 \le x \le 2$. We compute

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy \\
&= \int_1^2 (2-y)\,dy \\
&= \left(2y - \frac{y^2}{2} \right)\Bigg|_{y=1}^{y=2} \\
&= \frac{1}{2}
\end{aligned}
$$

If x is outside the interval $[0,2]$, we have $f_X(x) = 0$. Therefore,

$$
f_X(x) = \begin{cases} \frac{1}{2} & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}
$$

To find the marginal probability density function for Y, consider the case where we fix y so that $1 \le y \le 2$. We compute

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx \\ &= \int_0^2 (2-y)\,dx \\ &= (2-y)x\big|_{x=0}^{x=2} \\ &= 4-2y \end{aligned}$$

If y is outside the interval $[1,2]$, we have $f_{X,Y}(x,y) = 0$ giving us $f_Y(y) = 0$. Therefore,

$$f_Y(x) = \begin{cases} 4-2y & 1 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

You should also double check that f_X and f_Y are both probability density functions.

Functions of Random Vectors

The technique for computing functions of one-dimensional random variables carries over to the multi-dimensional case. Most of these problems just require a little careful reasoning.

Example: Let (X_1, X_2) be the die-tossing random vector of the example on page 156. Find the probability mass function for the random variable $Z = X_1 + X_2$, the sum of the two rolls.

Solution: We already know that $p_{X_1,X_2}(x_1,x_2) = \frac{1}{36}$ for $x_1 = 1,\ldots,6$ and $x_2 = 1,\ldots,6$. We ask the question, "What are the possible values that Z can take on?" The answer: "The integers $2, 3, 4, \ldots, 12$." For example, Z equals 4 precisely when any one of the mutually exclusive events

$$\begin{aligned} &\{X_1 = 1,\ X_2 = 3\}, \\ &\{X_1 = 2,\ X_2 = 2\}, \\ \text{or} \\ &\{X_1 = 3,\ X_2 = 1\} \end{aligned}$$

occurs. So,

$$p_Z(4) = p_{X_1,X_2}(1,3) + p_{X_1,X_2}(2,2) + p_{X_1,X_2}(3,1) = \tfrac{3}{36}.$$

Continuing in this manner, you should be able to verify that

$$\begin{array}{lll} p_Z(2) = \frac{1}{36}; & p_Z(3) = \frac{2}{36}; & p_Z(4) = \frac{3}{36}; \\ p_Z(5) = \frac{4}{36}; & p_Z(6) = \frac{5}{36}; & p_Z(7) = \frac{6}{36}; \\ p_Z(8) = \frac{5}{36}; & p_Z(9) = \frac{4}{36}; & p_Z(10) = \frac{3}{36}; \\ p_Z(11) = \frac{2}{36}; & p_Z(12) = \frac{1}{36}. & \end{array}$$

Example: Let (X, Y) be the two-dimensional random variable given in the example on page 166. Find the cumulative distribution function for the random variable $Z = X + Y$.

Solution: The support for the random variable Z is the interval $[0, 2]$, so

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ \boxed{?} & \text{if } 0 \leq z \leq 2 \\ 1 & \text{if } z > 2 \end{cases}$$

For the case $0 \leq z \leq 2$, we wish to evaluate

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

In other words, we are computing the probability mass assigned to the shaded set in Figure 6.7 as z varies from 0 to 2.

In establishing limits of integration, we notice that there are two cases to worry about as shown in Figure 6.8:

Case I ($z \leq 1$):

$$\begin{aligned} F_Z(z) &= \int_0^z \int_0^{z-y} (x + y)\, dx\, dy \\ &= \tfrac{1}{3} z^3. \end{aligned}$$

Case II ($z > 1$):

$$\begin{aligned} F_Z(z) &= \int_0^{z-1} \int_0^1 (x + y)\, dx\, dy + \int_{z-1}^1 \int_0^{z-y} (x + y)\, dx\, dy \\ &= z^2 - \tfrac{1}{3} z^3 - \tfrac{1}{3}. \end{aligned}$$

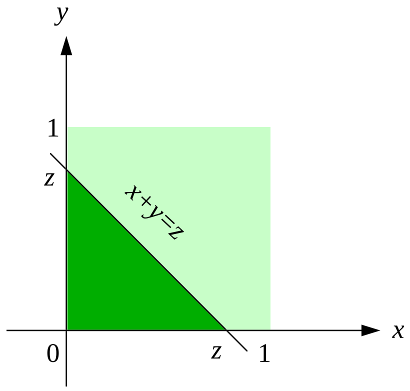


Figure 6.7: The event $\{Z \le z\}$.

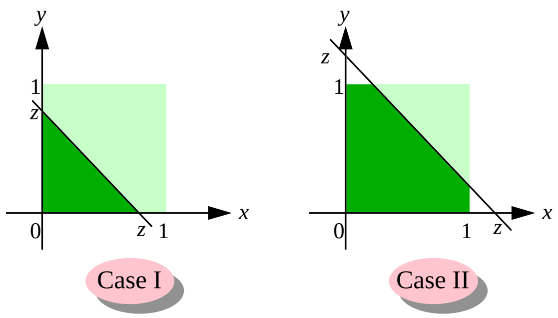


Figure 6.8: Two cases to worry about for the example

These two cases can be summarized as

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ \frac{1}{3}z^3 & \text{if } 0 \le z \le 1 \\ z^2 - \frac{1}{3}z^3 - \frac{1}{3} & \text{if } 1 < z \le 2 \\ 1 & \text{if } z > 2 \end{cases}$$

Notice what we thought at first to be one case ($0 \le z \le 2$) had to be divided into two cases ($0 \le z \le 1$ and $1 < z \le 2$).

Independence of Random Variables

Definition 6.7. *A sequence of n random variables $X_1, X_2, \ldots, X_n$ is* **independent** *if and only if, and*

$$F_{X_1,X_2,\ldots,X_n}(b_1, b_2, \ldots, b_n) = F_{X_1}(b_1)F_{X_2}(b_2)\cdots F_{X_n}(b_n)$$

for all values $b_1, b_2, \ldots, b_n$.

Definition 6.8. *A sequence of n random variables $X_1, X_2, \ldots, X_n$ is a* **random sample** *if and only if*

1. *$X_1, X_2, \ldots, X_n$ are independent, and*
2. *$F_{X_i}(x) = F(x)$ for all x and for all i (i.e., each X_i has the same marginal distribution, $F(x)$).*

We say that a random sample is a vector of **independent and identically distributed (i.i.d.)** *random variables.*

Recall: An event A is independent of an event B if and only if

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

Theorem 6.1. *If X and Y are independent random variables then any event A involving X alone is independent of any event B involving Y alone.*

Testing for independence

Case I: Discrete

A discrete random variable X is independent of a discrete random variable Y if and only if

$$p_{X,Y}(x, y) = [p_X(x)][p_Y(y)]$$

for all x and y.

Case II: Continuous

A continuous random variable X is independent of a continuous random variable Y if and only if

$$f_{X,Y}(x,y) = [f_X(x)][f_Y(y)]$$

for all x and y.

Example: A company produces two types of compressors, grade A and grade B. Let X denote the number of grade A compressors produced on a given day. Let Y denote the number of grade B compressors produced on the same day. Suppose that the joint probability mass function $p_{X,Y}(x,y) = P(X = x, Y = y)$ is given by the following table:

		y		
$p_{X,Y}(x,y)$		0	1	
	0	0.1	0.3	
x	1	0.2	0.1	
	2	0.2	0.1	

The random variables X and Y are not independent. Note that

$$p_{X,Y}(0,0) = 0.1 \neq p_X(0)p_Y(0) = (0.4)(0.5) = 0.2$$

Example: Suppose an electronic circuit contains two transistors. Let X be the time to failure of transistor 1 and let Y be the time to failure of transistor 2.

$$f_{X,Y}(x,y) = \begin{cases} 4e^{-2(x+y)} & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The marginal densities are

$$f_X(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} 2e^{-2y} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We must check the probability density functions for (X,Y), X and Y for all values of (x,y).

For $x \geq 0$ and $y \geq 0$:

$$f_{X,Y}(x,y) = 4e^{-2(x+y)} = f_X(x)f_Y(y) = 2e^{-2x}2e^{-2y}$$

For $x \geq 0$ and $y < 0$:

$$f_{X,Y}(x,y) = 0 = f_X(x)f_Y(y) = 2e^{-2x}(0)$$

For $x < 0$ and $y \geq 0$:

$$f_{X,Y}(x,y) = 0 = f_X(x)f_Y(y) = (0)2e^{-2y}$$

For $x < 0$ and $y < 0$:

$$f_{X,Y}(x,y) = 0 = f_X(x)f_Y(y) = (0)(0)$$

So the random variables X and Y are independent.

Expectation and random vectors

Suppose we are given a random vector (X,Y) and a function $g(x,y)$. Can we find $E(g(X,Y))$?

Theorem 6.2.

$$E(g(X,Y)) = \sum_{\text{all } x}\sum_{\text{all } y} g(x,y)p_{X,Y}(x,y) \qquad \text{if } (X,Y) \text{ is discrete}$$

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\,dy\,dx \qquad \text{if } (X,Y) \text{ is continuous}$$

Example: Suppose X and Y have joint probability density function

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

Let $Z = XY$. To find $E(Z) = E(XY)$ use Theorem 6.2 to get

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy \\
&= \int_0^1\int_0^1 xy(x+y)\,dx\,dy \\
&= \int_0^1\int_0^1 x^2y + xy^2\,dx\,dy \\
&= \int_0^1 \tfrac{1}{3}x^3y + \tfrac{1}{2}x^2y^2\Big|_0^1\,dy \\
&= \int_0^1 \tfrac{1}{3}y + \tfrac{1}{2}y^2\,dy \\
&= \tfrac{1}{6}y^2 + \tfrac{1}{6}y^3\Big|_0^1 = \tfrac{1}{3}
\end{aligned}
$$

We will prove the following results for the case when (X, Y) is a continuous random vector. The proofs for the discrete case are similar using summations rather than integrals, probability mass functions rather than probability density functions.

Theorem 6.3. $E(X + Y) = E(X) + E(Y)$

Proof. Using Theorem 6.2:

$$
\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y) f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,dy\,dx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\right]\,dx + \int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\right]\,dy \\
&= \int_{-\infty}^{\infty} x\left[f_X(x)\right]\,dx + \int_{-\infty}^{\infty} y\left[f_Y(y)\right]\,dy \\
&= E(X) + E(Y)
\end{aligned}
$$

■

Theorem 6.4. *If X and Y are independent, then*

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

Proof. Using Theorem 6.2:

$$\begin{aligned} E[h(X)g(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)g(y)f_{X,Y}(x,y)\,dx\,dy \\ &\quad \text{since } X \text{ and } Y \text{ are independent}\ldots \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)g(y)f_X(x)f_Y(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} g(y)f_Y(y)\left[\int_{-\infty}^{\infty} h(x)f_X(x)\,dx\right]\,dy \\ &= \int_{-\infty}^{\infty} g(y)f_Y(y)\left[E(h(X))\right]\,dy \\ &\quad \text{since } E(h(X)) \text{ is a constant}\ldots \\ &= E(h(X))\int_{-\infty}^{\infty} g(y)f_Y(y)\,dy \\ &= E(h(X))E(g(Y)) \end{aligned}$$

■

Corollary 6.5. *If X and Y are independent, then $E(XY) = E(X)E(Y)$*

Proof. Using Theorem 6.4, set $h(x) = x$ and $g(y) = y$ to get

$$\begin{aligned} E[h(X)g(Y)] &= E(h(X))E(g(Y)) \\ E[XY] &= E(X)E(Y) \end{aligned}$$

■

Definition 6.9. *The* **covariance** *of the random variables X and Y is*

$$\text{Cov}(X,Y) \equiv E[(X - E(X))(Y - E(Y))].$$

Note that $\text{Cov}(X,X) = \text{Var}(X)$.

Theorem 6.6. *For any random variables X and Y*

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$$

Proof. Remember that the variance of a random variable W is defined as

$$\begin{aligned}
\text{Var}(W) &= E[(W - E(W))^2] \\
&\quad \text{Now let } W = X + Y \ldots \\
\text{Var}(X+Y) &= E[(X + Y - E(X+Y))^2] \\
&= E[(X + Y - E(X) - E(Y))^2] \\
&= E[(\{X - E(X)\} + \{Y - E(Y)\})^2] \\
&\quad \text{Now let } a = \{X - E(X)\}, \text{ let } b = \{Y - E(Y)\}, \text{ and} \\
&\quad \text{expand } (a+b)^2 \text{ to get} \ldots \\
&= E[\{X - E(X)\}^2 + \{Y - E(Y)\}^2 + 2\{X - E(X)\}\{Y - E(Y)\}] \\
&= E[\{X - E(X)\}^2] + E[\{Y - E(Y)\}^2] \\
&\qquad + 2E[\{X - E(X)\}\{Y - E(Y)\}] \\
&= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)
\end{aligned}$$

■

Theorem 6.7. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

Proof. Using Definition 6.9, we get

$$\begin{aligned}
\text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
&= E[XY - XE(Y) - E(X)Y + E(X)E(Y)] \\
&= E[XY] + E[-XE(Y)] + E[-E(X)Y] + E[E(X)E(Y)]] \\
&\quad \text{Since } E[X], E[Y], \text{ and } E[XY] \text{ are all constants} \ldots \\
&= E[XY] - E(Y)E[X] - E(X)E[Y] + E(X)E(Y) \\
&= E[XY] - E(X)E(Y)
\end{aligned}$$

■

Corollary 6.8. *If X and Y are independent then* $\text{Cov}(X, Y) = 0$.

Proof. If X and Y are independent then, from Corollary 6.5, $E(XY) = E(X)E(Y)$. We can then use Theorem 6.7:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

■

Corollary 6.9. *If X and Y are independent then* $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$.

Proof. If X and Y are independent then, from Theorem 6.6 and Corollary 6.8:

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)\\ \mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2(0)\\ \mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)\end{aligned}$$

■

Definition 6.10. *The* **covariance** *of two random variables X and Y is given by*

$$\mathrm{Cov}(X,Y) \equiv E\big[(X-E(X))(Y-E(Y))\big]\,.$$

Definition 6.11. *The* **correlation coefficient** *for two random variables X and Y is given by*

$$\rho(X,Y) \equiv \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Theorems 6.10, 6.11 and 6.12 are stated without proof. Proofs of these results may be found in the book by Meyer.[2]

Theorem 6.10. *For any random variables X and Y,*

$$|\rho(X,Y)| \le 1.$$

Theorem 6.11. *Suppose that $|\rho(X,Y)| = 1$. Then (with probability one), $Y = aX + b$ for some constants a and b. In other words: If the correlation coefficient ρ is ± 1, the Y is a linear function of X (with probability one).*

The converse of this theorem is also true:

Theorem 6.12. *Suppose that X and Y are two random variables, such that $Y = aX + b$ where a and b are constants. Then $|\rho(X,Y)| = 1$. If $a > 0$, then $\rho(X,Y) = +1$. If $a < 0$, then $\rho(X,Y) = -1$.*

[2] Meyer, P., *Introductory probability theory and statistical applications*, Addison-Wesley, Reading MA, 1965.

Random vectors and conditional probability

Example: Consider the compressor problem again.

$p_{X,Y}(x,y)$		$y=0$	$y=1$
x	0	0.1	0.3
	1	0.2	0.1
	2	0.2	0.1

Given that no grade B compressors were produced on a given day, what is the probability that 2 grade A compressors were produced?

Solution:

$$\begin{aligned} P(X=2 \,|\, Y=0) &= \frac{P(X=2, Y=0)}{P(Y=0)} \\ &= \frac{0.2}{0.5} = \frac{2}{5} \end{aligned}$$

Given that 2 compressors were produced on a given day, what is the probability that one of them is a grade B compressor?

Solution:

$$\begin{aligned} P(Y=1 \,|\, X+Y=2) &= \frac{P(Y=1, X+Y=2)}{P(X+Y=2)} \\ &= \frac{P(X=1, Y=1)}{P(X+Y=2)} \\ &= \frac{0.1}{0.3} = \frac{1}{3} \end{aligned}$$

Example: And, again, consider the two transistors…

$$f_{X,Y}(x,y) = \begin{cases} 4e^{-2(x+y)} & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Given that the total life time for the two transistors is less than two hours, what is the probability that the first transistor lasted more than one hour?

Solution:

$$P(X > 1 \mid X + Y \le 2) = \frac{P(X > 1, X + Y \le 2)}{P(X + Y \le 2)}$$

We then compute

$$\begin{aligned} P(X > 1, X + Y \le 2) &= \int_1^2 \int_0^{2-x} 4e^{-2(x+y)}\, dy\, dx \\ &= e^{-2} - 3e^{-4} \end{aligned}$$

and

$$\begin{aligned} P(X + Y \le 2) &= \int_0^2 \int_0^{2-x} 4e^{-2(x+y)}\, dy\, dx \\ &= 1 - 5e^{-4} \end{aligned}$$

to get

$$P(X > 1 \mid X + Y \le 2) = \frac{e^{-2} - 3e^{-4}}{1 - 5e^{-4}} = 0.0885$$

Conditional distributions

Case I: Discrete

Let X and Y be random variables with joint probability mass function $p_{X,Y}(x, y)$ and let $p_Y(y)$ be the marginal probability mass function for Y.

We define the **conditional probability mass function** of X given Y as

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

whenever $p_Y(y) > 0$.

Case II: Continuous

Let X and Y be random variables with joint probability density function $f_{X,Y}(x, y)$ and let $f_Y(y)$ be the marginal probability density function for Y.

We define the **conditional probability density function** of X given Y as

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

whenever $f_Y(y) > 0$.

Law of total probability

Case I: Discrete

$$p_X(x) = \sum_y p_{X|Y}(x \mid y) p_Y(y)$$

Case II: Continuous

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

Self-Test Exercises for Chapter 6

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S6.1 Let X_1, X_2, X_3, X_4 be independent and identically distributed random variables, each with $P(X_i = 1) = \frac{1}{2}$ and $P(X_i = 0) = \frac{1}{2}$. Let $P(X_1 + X_2 + X_3 + X_4 = 3) = r$, then the value of r is

(A) $\frac{1}{16}$

(B) $\frac{1}{4}$

(C) $\frac{1}{2}$

(D) 1

(E) none of the above.

S6.2 Let (X, Y) be a discrete random vector with joint probability mass function given by

$$\begin{aligned} p_{X,Y}(0,0) &= 1/4 \\ p_{X,Y}(0,1) &= 1/4 \\ p_{X,Y}(1,1) &= 1/2 \end{aligned}$$

Then $P(Y = 1)$ equals

(A) $1/3$

(B) $1/4$

(C) $1/2$

(D) $3/4$

(E) none of the above.

S6.3 Let (X, Y) be a random vector with joint probability mass function given by

$$p_{X,Y}(1,-1) = \tfrac{1}{4} \qquad p_{X,Y}(1,1) = \tfrac{1}{2} \qquad p_{X,Y}(0,0) = \tfrac{1}{4}$$

Let $p_X(x)$ denote the marginal probability mass function for X. The value of $p_X(1)$ is

(A) 0

(B) $\frac{1}{4}$

(C) $\frac{1}{2}$

(D) 1

(E) none of the above.

S6.4 Let (X, Y) be a random vector with joint probability mass function given by

$$p_{X,Y}(1,-1) = \tfrac{1}{4} \qquad p_{X,Y}(1,1) = \tfrac{1}{2} \qquad p_{X,Y}(0,0) = \tfrac{1}{4}$$

The value of $P(XY > 0)$ is

(A) 0

(B) $\frac{1}{4}$

(C) $\frac{1}{2}$

(D) 1

(E) none of the above.

S6.5 Let (X, Y) be a continuous random vector with joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} 1/2 & \text{if } -1 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $P(X > 0)$ equals

(A) 0

(B) $1/4$

(C) $1/2$

(D) 1

(E) none of the above.

S6.6 Suppose (X, Y) is a continuous random vector with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} e^{-y} & \text{if } 0 \leq x \leq 1 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then $P(X > \frac{1}{2})$ equals

(A) 0

(B) 1/2

(C) $e^{-1/2}$

(D) $\frac{1}{2}e^{-y}$

(E) none of the above.

S6.7 Let (X, Y) be a discrete random vector with joint probability mass function given by

$$\begin{aligned} p_{X,Y}(0, 0) &= 1/3 \\ p_{X,Y}(0, 1) &= 1/3 \\ p_{X,Y}(1, 0) &= 1/3 \end{aligned}$$

Then $P(X + Y = 1)$ equals

(A) 0

(B) $1/3$

(C) $2/3$

(D) 1

(E) none of the above.

S6.8 Let (X, Y) be a discrete random vector with joint probability mass function given by

$$\begin{aligned} p_{X,Y}(0, 0) &= 1/3 \\ p_{X,Y}(0, 1) &= 1/3 \\ p_{X,Y}(1, 0) &= 1/3 \end{aligned}$$

Let $W = \max\{X, Y\}$. Then $P(W = 1)$ equals

(A) 0

(B) $1/3$

(C) $2/3$

(D) 1

(E) none of the above.

S6.9 Suppose (X, Y) is a continuous random vector with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} y & \text{if } 0 \leq x \leq 2 \text{ and} \\ & \ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $E(X^2)$ equals

(A) 0

(B) 1

(C) 4/3

(D) 8/3

(E) none of the above.

S6.10 Suppose (X, Y) is a continuous random vector with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} y & \text{if } 0 \leq x \leq 2 \text{ and} \\ & \ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $P(X < 1, Y < 0.5)$ equals

(A) 0

(B) 1/8

(C) 1/4

(D) 1/2

(E) none of the above.

S6.11 The weights of individual oranges are independent random variables, each having an expected value of 6 ounces and standard deviation of 2 ounces. Let Y denote the total net weight (in ounces) of a basket of n oranges. The variance of Y is equal to

(A) $4\sqrt{n}$

(B) $4n$

(C) $4n^2$

(D) 4

(E) none of the above.

S6.12 Let X_1, X_2, X_3 be independent and identically distributed random variables, each with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. If $P(X_1 + X_2 + X_3 = 3) = r$, then the value of p is

(A) $\frac{1}{3}$

(B) $\frac{1}{2}$

(C) $\sqrt[3]{r}$

(D) r^3

(E) none of the above.

S6.13 If X and Y are random variables with $\text{Var}(X) = \text{Var}(Y)$, then $\text{Var}(X + Y)$ must equal

(A) $2\text{Var}(X)$

(B) $\sqrt{2}(\text{Var}(X))$

(C) $\text{Var}(2X)$

(D) $4\text{Var}(X)$

(E) none of the above.

S6.14 Let (X, Y) be a continuous random vector with joint probability density function

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

Then $P(X > 0)$ equals

(A) $1/(4\pi)$

(B) $1/(2\pi)$

(C) $1/2$

(D) 1

(E) none of the above.

S6.15 Let X be a random variable. Then $E[(X+1)^2] - (E[X+1])^2$ equals

(A) Var(X)

(B) $2E[(X+1)^2]$

(C) 0

(D) 1

(E) none of the above.

S6.16 Let X, Y and Z be independent random variables with

$$E(X) = 1 \quad E(Y) = 0 \quad E(Z) = 1$$

Let

$$W = X(Y + Z).$$

Then $E(W)$ equals

(A) 0

(B) $E(X^2)$

(C) 1

(D) 2

(E) none of the above

S6.17 Toss a fair die twice. Let the random variable X represent the outcome of the first toss and the random variable Y represent the outcome of the second toss. What is the probability that X is odd and Y is even.

(A) $1/4$

(B) $1/3$

(C) $1/2$

(D) 1

(E) none of the above

S6.18 Assume that (X, Y) is random vector with joint probability density function

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

where k is a some positive constant. Let $W = X - Y$. To find the cumulative distribution function for W we can compute

$$F_W(w) = \int_A^B \int_C^D k \, dx \, dy$$

for $0 \le w \le 1$. The limit of integration that should appear in position D is

(A) $\min\{1, x - w\}$

(B) $\min\{1, y - w\}$

(C) $\min\{1, x - y\}$

(D) $\min\{1, y + w\}$

(E) none of the above.

S6.19 Assume that (X, Y) is a random vector with joint probability mass function given by

$$\begin{aligned} p_{X,Y}(-1, 0) &= 1/4 \\ p_{X,Y}(0, 0) &= 1/2 \\ p_{X,Y}(0, 1) &= 1/4 \end{aligned}$$

Define the random variable $W = XY$. The value of $P(W = 0)$ is

(A) 0

(B) $1/4$

(C) $1/2$

(D) $3/4$

(E) none of the above.

S6.20 Let X be a random variable with $E(X) = 0$ and $\text{Var}(X) = 1$. Let Y be a random variable with $E(Y) = 0$ and $\text{Var}(Y) = 4$. Then $E(X^2 + Y^2)$ equals

(A) 0

(B) 1

(C) 3

(D) 5

(E) none of the above.

S6.21 Suppose (X, Y) is a continuous random vector with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} 4xy & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $E\left[\dfrac{1}{XY}\right]$ equals

(A) $+\infty$

(B) 0

(C) $1/2$

(D) 4

(E) none of the above.

S6.22 The life lengths of two transistors in an electronic circuit is a random vector (X, Y) where X is the life length of transistor 1 and Y is the life length of transistor 2. The joint probability density function of (X, Y) is given by

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-(x+2y)} & \text{if } x \ge 0 \text{ and } y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

Then $P(X + Y \le 1)$ equals

(A) $\int_0^1 \int_0^1 2e^{-(x+2y)}\, dx\, dy$

(B) $\int_0^1 \int_0^{1-y} 2e^{-(x+2y)}\, dx\, dy$

(C) $\int_0^1 \int_y^1 2e^{-(x+2y)}\, dx\, dy$

(D) $\int_0^1 \int_{1-y}^1 2e^{-(x+2y)}\, dx\, dy$

(E) none of the above.

Questions for Chapter 6

6.1 A factory can produce two types of gizmos, Type 1 and Type 2. Let X be a random variable denoting the number of Type 1 gizmos produced on a given day, and let Y be the number of Type 2 gizmos produced on the same day. The joint probability mass function for X and Y is given by

$$\begin{array}{ll} p_{X,Y}(1,0) = 0.10; & p_{X,Y}(1,1) = 0.20; \\ p_{X,Y}(1,2) = 0.10; & p_{X,Y}(1,3) = 0.10; \\ p_{X,Y}(2,1) = 0.20; & p_{X,Y}(2,2) = 0.20; \\ p_{X,Y}(3,1) = 0.05; & p_{X,Y}(3,2) = 0.05; \end{array}$$

(a) Compute $P(X \leq 2, Y = 2)$, $P(X \leq 2, Y \neq 2)$ and $P(Y > 0)$.

(b) Find the marginal probability mass functions for X and for Y.

(c) Find the distribution for the random variable $Z = X + Y$ which is the total daily production of gizmos.

6.2 A sample space Ω is a set consisting for four points, $\{\omega_1, \omega_2, \omega_3, \omega_4\}$. A probability measure, $P(\cdot)$ assigns probabilities, as follows:

$$\begin{array}{ll} P(\{\omega_1\}) = \frac{1}{2} & P(\{\omega_2\}) = \frac{1}{8} \\ P(\{\omega_3\}) = \frac{1}{4} & P(\{\omega_4\}) = \frac{1}{8} \end{array}$$

Random variables X, Y and Z are defined as

$$\begin{array}{llll} X(\omega_1) = 0, & X(\omega_2) = 0, & X(\omega_3) = 1, & X(\omega_4) = 1 \\ Y(\omega_1) = 0, & Y(\omega_2) = 1, & Y(\omega_3) = 1, & Y(\omega_4) = 0 \\ Z(\omega_1) = 1, & Z(\omega_2) = 2, & Z(\omega_3) = 3, & Z(\omega_4) = 4 \end{array}$$

(a) Find the joint probability mass function for (X, Y, Z).

(b) Find the joint probability mass function for (X, Y).

(c) Find the probability mass function for X.

(d) Find the probability mass function for the random variable $W = X + Y + Z$.

(e) Find the probability mass function for the random variable $T = XYZ$.

(f) Find the joint probability mass function for the random variable (W, T).

(g) Find the probability mass function for the random variable $V = \max\{X, Y\}$.

6.3 Let (X, Y) have a uniform distribution over the rectangle

$$((0, 0), (0, 1), (1, 1), (1, 0)) .$$

In other words, the probability density function for (X, Y) is given by

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

Find the cumulative distribution function for the random variable $Z = X+Y$.

6.4 Two light bulbs are burned starting at time 0. The first one to fail burns out at time X and the second one at time Y. Obviously $X \le Y$. The joint probability density function for (X, Y) is given by

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-(x+y)} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the region in (x, y)-space for which the above probability density function assigns positive mass.

(b) Find the marginal probability density functions for X and for Y.

(c) Let Z denote the *excess life* of the second bulb, i.e., let $Z = Y - X$. Find the probability density function for Z.

(d) Compute $P(Z > 1)$ and $P(Y > 1)$.

6.5 If two random variables X and Y have a joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{if } x > 0,\ y > 0 \text{ and } x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the probability that both random variables will take on a value less than $\frac{1}{2}$.

(b) Find the probability that X will take on a value less than $\frac{1}{4}$ and Y will take on a value greater than $\frac{1}{2}$.

(c) Find the probability that the *sum* of the values taken on by the two random variables will exceed $\frac{2}{3}$.

(d) Find the marginal probability density functions for X and for Y.

(e) Find the marginal probability density function for $Z = X + Y$.

6.6 Suppose (X, Y, Z) is a random vector with joint probability density function

$$f_{X,Y,Z}(x, y, z) = \begin{cases} \alpha(xyz^2) & \text{if } 0 < x < 1,\ 0 < y < 2 \text{ and } 0 < z < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of the constant α.

(b) Find the probability that X will take on a value less than $\frac{1}{2}$ and Y and Z will both take on values less than 1.

(c) Find the probability density function for the random vector (X, Y).

6.7 Suppose that the random vector (X, Y) has joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} kx(x - y) & \text{if } 0 \le x \le 2 \text{ and } |y| \le x \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the support of the distribution for (X, Y) in the xy-plane.

(b) Evaluate the constant k.

(c) Find the marginal probability density function for Y.

(d) Find the marginal probability density function for X.

(e) Find the probability density function for $Z = X + Y$.

6.8 The length, X, and the width, Y, of salt crystals form a random variable (X, Y) with joint probability density function

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

(a) Find the marginal probability density functions for X and for Y.

6.9 If the joint probability density function of the price, P, of a commodity (in dollars) and total sales, S, (in $10,000$ units) is given by

$$f_{S,P}(s, p) = \begin{cases} 5pe^{-ps} & 0.20 < p < 0.40 \text{ and } s > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal probability density functions for P and for S.

(b) Find the conditional probability density function for S given that P takes on the value p.

(c) Find the probability that sales will exceed $20,000$ units given $P = 0.25$.

6.10 If X is the proportion of persons who will respond to one kind of mail-order solicitation, Y is proportion of persons who will respond to a second type of mail-order solicitation, and the joint probability density function of X and Y is given by

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{5}(x + 4y) & 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal probability density functions for X and for Y.

(b) Find the conditional probability density function for X given that Y takes on the value y.

(c) Find the conditional probability density function for Y given that X takes on the value x.

(d) Find the probability that there will be at least a 20% response to the second type of mail-order solicitation.

(e) Find the probability that there will be at least a 20% response to the second type of mail-order solicitation given that there has only been a 10% response to the first kind of mail-order solicitation.

6.11 An electronic component has two fuses. If an overload occurs, the time when fuse 1 blows is a random variable, X, and the time when fuse 2 blows is a random variable, Y. The joint probability density function of the random vector (X,Y) is given by

$$f_{X,Y}(x,y) = \begin{cases} e^{-y} & 0 \le x \le 1 \text{ and } y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute $P(X+Y \ge 1)$.

(b) Are X and Y independent? Justify your answer.

6.12 The coordinates of a laser on a circular target are given by the random vector (X,Y) with the following probability density function:

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

Hence, (X,Y) has a uniform distribution on a disk of radius one centered at $(0,0)$.

(a) Compute $P(X^2+Y^2 \le 0.25)$.

(b) Find the marginal distributions for X and Y.

(c) Compute $P(X \ge 0,\ Y > 0)$.

(d) Compute $P(X \ge 0 \mid Y > 0)$.

(e) Find the conditional distribution for X given $Y = y$. (*Note:* Be careful of the case $|y| = 1$.)

6.13 Let X and Y be discrete random variables with the support of X denoted by Θ and the support of Y denoted by Φ. Let p_X be the marginal probability mass function for X, let $p_{X|Y}$ be the conditional probability mass function of X given Y, and let $p_{Y|X}$ be the conditional probability mass function of

Y given X. Show that

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_{x \in \Theta} p_{Y|X}(y|x)p_X(x)}$$

for any $x \in \Theta$ and $y \in \Phi$.

6.14 The volumetric fractions of each of two compounds in a mixture are random variables X and Y, respectively. The random vector (X, Y) has joint probability density function

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

(a) Determine the marginal probability density functions for X and for Y.

(b) Are X and Y independent random variables? Justify your answer.

(c) Compute $P(X < 0.5 \,|\, Y = 0.5)$.

(d) Compute $P(X < 0.5 \,|\, Y < 0.5)$.

6.15 Suppose (X, Y) is a continuous random vector with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{if } x > 0, y > 0 \text{ and } x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(X + Y)$.

6.16 Suppose (X, Y, Z) is a random vector with joint probability density function

$$f_{X,Y,Z}(x, y, z) = \begin{cases} \frac{3}{8}(xyz^2) & \text{if } 0 < x < 1, 0 < y < 2, 0 < z < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(X + Y + Z)$.

6.17 Suppose the joint probability density function of the price, P, of a commodity (in dollars) and total sales, S, (in $10,000$ units) is given by

$$f_{S,P}(s, p) = \begin{cases} 5pe^{-ps} & 0.20 < p < 0.40 \text{ and } s > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(S \,|\, P = 0.25)$.

6.18 Let X be the proportion of persons who will respond to one kind of mail-order solicitation, and let Y be the proportion of persons who will respond to a second type of mail-order solicitation. Suppose the joint probability density function of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{2}{5}(x+4y) & 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(X \mid Y = y)$ for each $0 < y < 1$.

6.19 Suppose that (X, Y) is a random vector with the following joint probability density function:

$$f_{X,Y}(x,y) = \begin{cases} 2x & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the marginal probability density function for Y.

6.20 Suppose that (X, Y) is a random vector with the following joint probability density function:

$$f_{X,Y}(x,y) = \begin{cases} 2xe^{-y} & \text{if } 0 \leq x \leq 1 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X > 0.5)$.

(b) Find the marginal probability density function for Y.

(c) Find $E(X)$.

(d) Are X and Y independent random variables? Justify your answer.

6.21 The Department of Metaphysics of Podunk University has 5 students, with 2 juniors and 3 seniors. Plans are being made for a party. Let X denote the number of juniors who will attend the party, and let Y denote the number of seniors who will attend.

After an extensive survey was conducted, it has been determined that the random vector (X, Y) has the following joint probability mass function:

$p_{X,Y}(x,y)$		Y (seniors) 0	1	2	3
X (juniors)	0	.01	.25	.03	.01
	1	.04	.05	.15	.05
	2	.04	.12	.20	.05

(a) What is the probability that 3 *or more* students (juniors and/or seniors) will show up at the party?

(b) What is the probability that more seniors than juniors will show up at the party?

(c) If juniors are charged \$5 for attending the party and seniors are charged \$10, what is the expected amount of money that will be collected from all students?

(d) The two juniors arrive early at the party. What is the probability that no seniors will show up?

6.22 Let (X, Y) be a continuous random vector with joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{if } x \geq 0,\, y \geq 0 \text{ and } x + y \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

(a) Compute $P(X < Y)$.

(b) Find the marginal probability density function for X *and* the marginal probability density function for Y.

(c) Find the conditional probability density function for Y given that $X = \frac{1}{2}$.

(d) Are X and Y independent random variables? Justify your answer.

6.23 Let (X, Y) be a continuous random vector with joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} 4xy & \text{for } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

(a) Find the marginal probability density function for X *and* the marginal probability density function for Y.

(b) Are X and Y independent random variables? Justify your answer.

(c) Find $P(X + Y \leq 1)$.

(d) Find $P(X > 0.5 \,|\, Y \leq 0.5)$.

6.24 Let (X, Y) be a continuous random vector with joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} 6x^2y & \text{for } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal probability density function for X *and* the marginal probability density function for Y.

(b) Are X and Y independent random variables? Justify your answer.

(c) Find $P(X + Y \leq 1)$.

(d) Find $E(X + Y)$.

6.25 A box contains four balls numbered 1, 2, 3 and 4. A game consists of drawing one of the balls at random from the box. It is not replaced. A second ball is then drawn at random from the box.

Let X be the number on the first ball drawn, and let Y be the number on the second ball.

(a) Find the joint probability mass function for the random vector (X, Y).

(b) Compute $P(Y \geq 3 \mid X = 1)$.

6.26 Suppose that (X, Y) is a random vector with the following joint probability density function:

$$f_{X,Y}(x,y) = \begin{cases} 2xe^{-y} & \text{if } 0 < x < 1 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute $P(X > 0.5, Y < 1.0)$.

(b) Compute $P(X > 0.5 \mid Y < 1.0)$.

(c) Find the marginal probability density function for X.

(d) Find $E(X + Y)$.

6.27 Past studies have shown that juniors taking a particular probability course receive grades according to a normal distribution with mean 65 and variance 36. Seniors taking the same course receive grades normally distributed with a mean of 70 and a variance of 36. A probability class is composed of 75 juniors and 25 seniors.

(a) What is the probability that a student chosen at random from the class will receive a grade in the 70's (i.e., between 70 and 80)?

(b) If a student is chosen at random from the class, what is the student's expected grade?

(c) A student is chosen at random from the class and you are told that the student has a grade in the 70's. What is the probability that the student is a junior?

6.28 (†) Suppose X and Y are positive, independent continuous random variables with probability density functions $f_X(x)$ and $f_Y(y)$, respectively. Let $Z = X/Y$.

(a) Express the probability density function for Z in terms of an integral that involves only f_X and f_Y.

(b) Now suppose that X and Y can be positive and/or negative. Express the probability density function for Z in terms of an integral that involves only f_X and f_Y. Compare your answer to the case where X and Y are both positive.

7

TRANSFORMS AND SUMS

Transforms

Notation

Throughout this section we will use the following notation:

Definition 7.1. *For any differentiable function $h(u)$, and constant c*

$$h^{(k)}(c) \equiv \frac{d^k}{du^k} h(u) \bigg|_{u=c}.$$

In other words, $h^{(k)}(c)$ is the kth derivative of h evaluated at $u = c$.

Example: Let $h(x) = x^3$ then

$$\begin{aligned} h^{(2)}(4) &= \frac{d^2}{dx^2} h(x) \bigg|_{x=4} \\ &= \frac{d^2}{dx^2} x^3 \bigg|_{x=4} \\ &= 3x^2 \Big|_{x=4} \\ &= 3(4)^2 = 48 \end{aligned}$$

Probability generating functions

Definition 7.2. *Let X be a discrete random variable with support on the nonnegative integers. Then the* **probability generating function** *for X is given by*

$$g_X(z) \equiv E(z^X) \qquad \text{for } |z| < 1.$$

Example: Let X have a Poisson distribution with parameter $\alpha > 0$ Then

$$p_X(k) = \frac{e^{-\alpha}\alpha^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

Recalling that $e^x = \sum_{k=0}^{\infty}(x^k/k!)$, we get

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

Theorem 7.1. *If $g_X(z)$ is the probability generating function of X, then*

$$P(X = k) = \frac{g_X^{(k)}(0)}{k!}$$

Proof. (Informal proof) Let $g^{(k)}(z)$ denote the kth derivative of g with respect to z. We have

$$\begin{aligned} g_X(z) &= E(z^X) \\ &= \sum_{k=0}^{\infty} z^k P(X = k) \end{aligned}$$

Writing out the sum explicitly, we get...

$$\begin{aligned} g_X(z) &= z^0P(X = 0) + z^1P(X = 1) + z^2P(X = 2) + z^3P(X = 3) + \ldots \\ g_X(0) &= (1)P(X = 0) \end{aligned}$$

Now take the first derivative of $g_X(z)$...

$$\begin{aligned} g_X^{(1)}(z) &= P(X = 1) + 2zP(X = 2) + 3z^2P(X = 3) + 4x^3P(X = 4) + \ldots \\ g_X^{(1)}(0) &= P(X = 1) \end{aligned}$$

Another derivative produces...

$$\begin{aligned}
g_X^{(2)}(z) &= 2P(X=2) + 3(2)zP(X=3) + 4(3)z^2P(X=4) + 5(4)z^3P(X=5) + \ldots \\
g_X^{(2)}(0) &= 2P(X=2)
\end{aligned}$$

And once more to see the pattern...

$$\begin{aligned}
g_X^{(3)}(z) &= 3(2)P(X=3) + 4(3)(2)zP(X=4) + 5(4)(3)z^2P(X=5) + \ldots \\
g_X^{(3)}(0) &= 3(2)P(X=3)
\end{aligned}$$

And, in general

$$g_X^{(n)}(0) = n!P(X=n)$$

■

Theorem 7.2. *If $g_X(z)$ is the probability generating function of X, then*

$$E(X) = g_X^{(1)}(1)$$

where $g_X^{(1)}(1)$ is the first derivative of $g_X(z)$ with respect to z, evaluated at $z = 1$.

Proof. As above,

$$\begin{aligned}
g_X^{(1)}(z) &= P(X=1) + 2zP(X=2) + 3z^2P(X=3) + 4x^3P(X=4) + \ldots \\
g_X^{(1)}(1) &= P(X=1) + 2P(X=2) + 3P(X=3) + 4P(X=4) + \ldots \\
&= \sum_{k=0}^{\infty} kP(X=k) = E(X)
\end{aligned}$$

■

Theorem 7.3. *If $X_1, X_2, \ldots, X_n$ are independent random variables with probability generating functions $g_{X_1}(z),\ g_{X_2}(z), \ldots, g_{X_n}(z)$, respectively, then the probability generating function of the random variable $Y = X_1 + X_2 + \cdots + X_n$ is*

$$g_Y(z) = \prod_{i=1}^{n} g_{X_i}(z).$$

Proof. Recall that if X and Y are independent random variables, then

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

Therefore,

$$\begin{aligned} g_Y(z) = E(z^Y) &= E(z^{X_1+X_2+\ldots+X_n}) \\ &= E(z^{X_1}z^{X_2}\cdots z^{X_n}) \\ &= E(z^{X_1})E(z^{X_2})\cdots E(z^{X_n}) \\ &= g_{X_1}(z)g_{X_2}(z)\cdots g_{X_n}(z) \end{aligned}$$

■

Theorem 7.4. *If $g_X(z)$ is the probability generating function of X, and a is a nonnegative, integer constant, then the probability generating function of the random variable $Y = a + X$ is*

$$g_Y(z) = z^a g_X(z).$$

Proof. We have,

$$\begin{aligned} g_Y(z) = E(z^Y) &= E(z^{a+X}) \\ &= E(z^a z^X) \\ &= z^a E(z^X) \\ &= z^a g_X(z) \end{aligned}$$

■

We offer Theorem 7.5 without proof. Except for some technical details, it is a direct result of Theorem 7.1.

Theorem 7.5. *The probability generating function of a random variable uniquely determines its probability distribution.*

Example: Let T have a Bernoulli distribution with parameter p.

$$\begin{aligned} P(T=1) &= p \\ P(T=0) &= 1-p=q \end{aligned}$$

Then the probability generating function of T is given by

$$g_T(z) = z^0 q + z^1 p = q + zp$$

Example: Now suppose, $T_1, T_2, \ldots, T_n$ are n independent Bernoulli random variables each with

$$\begin{aligned} P(T_i = 1) &= p \\ P(T_i = 0) &= 1 - p = q \end{aligned}$$

We know (from Chapter 4) that $X = T_1 + T_2 + \cdots + T_n$ has a Binomial distribution with parameters n and p. That is

$$P(X = k) = \binom{n}{k} p^k q^{(n-k)} \qquad \text{for } k = 0, 1, \ldots, n.$$

One could obtain the probability generating function for X is given by computing

$$g_X(z) = E(z^X) = \sum_{k=0}^{n} z^k \binom{n}{k} p^k q^{(n-k)}$$

But we can use Theorem 7.3 with the T_i's to easily get

$$g_X(z) = g_{T_1}(z) g_{T_2}(z) \cdots g_{T_n}(z) = (q + zp)^n$$

As an exercise, use Theorem 7.2 to get $E(X)$ from $g_X(z)$.

Moment generating functions

Definition 7.3. *The* ***k*th moment about the origin** *of a random variable X is defined as $E(X^k)$.*

Definition 7.4. *Let X be (almost) any random variable. Then the* **moment generating function** *for X is given by*

$$M_X(t) \equiv E(e^{tX}).$$

Example: Let X have a uniform distribution on the interval (a, b). Then

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

Therefore,

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \\ &= \int_a^b e^{tx} \frac{1}{b-a}\,dx \\ &= \frac{1}{t(b-a)} \int_a^b t e^{tx}\,dx \\ &= \frac{1}{t(b-a)} e^{tx} \Big|_a^b \\ &= \frac{e^{tb}-e^{ta}}{t(b-a)} \qquad \text{for } t \neq 0 \end{aligned}$$

Example: Let X have a Poisson distribution with parameter $\alpha > 0$ Then

$$p_X(k) = \frac{e^{-\alpha}\alpha^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

Recalling that $e^x = \sum_{k=0}^{\infty}(x^k/k!)$, we get

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\alpha}\alpha^k}{k!} \\ &= e^{-\alpha} \sum_{k=0}^{\infty} \frac{(\alpha e^t)^k}{k!} \\ &= e^{-\alpha} e^{\alpha e^t} \\ &= e^{\alpha(e^t-1)} \end{aligned}$$

Note that since a Poisson random variable X has support $\{0, 1, 2, \ldots\}$, it can have a probability generating function and a moment generating function. When a random variable has both a probability generating function and a moment generating function, we can derive the following relationship:

$$\begin{aligned} g_X(z) &= E[z^X] \\ & \quad \text{substituting } z = e^t, \text{ we obtain} \\ g_X(e^t) &= E[(e^t)^X] \\ &= E[e^{tX}] = M_X(t) \end{aligned}$$

Since the probability generating function for a Poisson random variable is

$$g_X(z) = e^{\alpha(z-1)}$$

the moment generating function for a Poisson random variable must be

$$M_X(t) = g_X(e^t) = e^{\alpha(e^t-1)}$$

It's important to remember that this relationship will only hold for random variables with support $\{0, 1, 2, \ldots\}$ (that is, those random variables with probability generating functions).

Example: Let X have an exponential distribution with parameter $\alpha > 0$. Then

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \int_0^\infty e^{tx} \alpha e^{-\alpha x}\, dx \\ &= \int_0^\infty \alpha e^{x(t-\alpha)}\, dx \end{aligned}$$

Note that this integral only converges if $t < \alpha$. Continuing,

$$\begin{aligned} M_X(t) &= \left. \frac{\alpha}{t-\alpha} e^{x(t-\alpha)} \right|_0^\infty \\ &= \frac{\alpha}{\alpha - t} \qquad \text{for } t < \alpha \end{aligned}$$

Example: (From Meyer.[1]) Let $X \sim N(\mu, \sigma^2)$. Then

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)$$

Therefore, for every t,

$$M_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tx} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)\, dx$$

Let $(x-\mu)/\sigma = s$; thus $x = \sigma s + \mu$ and $dx = \sigma ds$. Therefore,

$$M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp[t(\sigma s + \mu)] e^{-s^2/2}\, ds$$

[1]Meyer, P., *Introductory probability theory and statistical applications*, Addison-Wesley, Reading MA, 1965.

$$
\begin{aligned}
&= e^{t\mu}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp(-\tfrac{1}{2}[s^2 - 2\sigma ts])\,ds \\
&= e^{t\mu}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\{-\tfrac{1}{2}[(s-\sigma t)^2 - \sigma^2 t^2]\}\,ds \\
&= e^{t\mu+\sigma^2 t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp(-\tfrac{1}{2}[s-\sigma t]^2)\,ds
\end{aligned}
$$

Now let $s - \sigma t = v$; then $ds = dv$ and we obtain

$$
\begin{aligned}
M_X(t) &= e^{t\mu+\sigma^2 t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-v^2/2}\,dv \\
&= e^{(t\mu+\sigma^2 t^2/2)}.
\end{aligned}
$$

Theorem 7.6. *If $M_X(t)$ is the moment generating function of X, then*

$$E(X^k) = M_X^{(k)}(0) \qquad \textit{for } k = 0, 1, 2, \ldots$$

Proof. (Informal proof) Let $M_X^{(k)}(t)$ denote the kth derivative of M_X with respect to t. We have

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= E\left(\sum_{k=0}^{\infty}\frac{(tX)^k}{k!}\right)
\end{aligned}
$$

Writing out the sum explicitly, we get...

$$M_X(t) = E\left(1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \frac{(tX)^4}{4!} + \cdots\right)$$

If the sum in the right hand side converges, we get ...

$$M_X(t) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \frac{t^3 E(X^3)}{3!} + \frac{t^4 E(X^4)}{4!} + \cdots$$

Take the first derivative of both sides to get ...

$$M_X^{(1)}(t) = E(X) + tE(X^2) + \frac{t^2 E(X^3)}{2!} + \frac{t^3 E(X^4)}{3!} + \cdots$$

and when evaluated at $t = 0\ldots$

$$M_X^{(1)}(0) = E(X)$$

Take the second derivative of both sides to get ...

$$M_X^{(2)}(t) = E(X^2) + tE(X^3) + \frac{t^2 E(X^4)}{2!} + \frac{t^3 E(X^5)}{3!} + \cdots$$

and when evaluated at $t = 0\ldots$

$$M_X^{(2)}(0) = E(X^2)$$

Continuing, we see that...

$$M_X^{(k)}(t) = E(X^k) + tE(X^{k+1}) + \frac{t^2 E(X^{k+2})}{2!} + \frac{t^3 E(X^{k+3})}{3!} + \cdots$$

and when evaluated at $t = 0\ldots$

$$M_X^{(k)}(0) = E(X^k)$$

■

Theorem 7.7. *If $X_1, X_2, \ldots, X_n$ are independent random variables with moment generating functions $M_{X_1}(t),\ M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively, then the moment generating function of the random variable $Y = X_1 + X_2 + \cdots + X_n$ is*

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Theorem 7.8. *If $M_X(t)$ is the moment generating function of X, and a and b are constants, then the moment generating function of the random variable $Y = aX + b$ is*

$$M_Y(t) = e^{bt} M_X(at).$$

Theorem 7.9. *The moment generating function of a random variable uniquely determines its probability distribution.*

Example: If $X \sim N(\mu, \sigma^2)$ then the moment generating function for X is

$$M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

So, if someone tells you that the random variable W has moment generating function

$$M_W(t) = e^{-2t+4t^2}$$

then the distribution of W must be $N(-2, 8)$.

And, if you encounter a random Y with moment generating function

$$M_Y(t) = e^{t^2/2}$$

you can immediately say that $Y \sim N(0, 1)$.

Now, suppose you are given that the random variable X has a normal distribution with mean μ and variance σ^2. So the moment generating function for X is

$$M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

Consider the random variable

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

Using Theorem 7.8 with

$$a = \frac{1}{\sigma} \qquad \text{and} \qquad b = -\frac{\mu}{\sigma}$$

we find that the moment generating function for Z is

$$\begin{aligned} M_Z(t) &= e^{bt} M_X(at) \\ &= e^{-t\mu/\sigma} M_X(t/\sigma) \\ &= e^{[-\frac{t\mu}{\sigma}]} e^{[\frac{t}{\sigma}\mu + \frac{t^2}{2\sigma^2}\sigma^2]} \\ &= e^{[-\frac{t\mu}{\sigma} + \frac{t\mu}{\sigma} + \frac{t^2}{2\sigma^2}\sigma^2]} \\ &= e^{t^2/2} \end{aligned}$$

which is the moment generating function of a normal random variable with expected value equal to zero and variance equal to one. Therefore, we have shown that if $X \sim N(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

Other transforms of random variables

Laplace transform

Definition 7.5. *Let X be (almost) any random variable. Then the* **Laplace transform** *for X is given by*

$$L_X(t) \equiv E(e^{-tX}).$$

Note that the moment generating function for a continuous random variable X with probability density function $f_X(x)$ is

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$

and the Laplace transform for the same density function is

$$L_X(t) = \int_{-\infty}^{\infty} e^{-tx} f_X(x)\, dx$$

So the moment generating function of a continuous random variable is simply the Laplace transform of the probability density function with t replaced by $-t$ everywhere in the formula for the transform.

Characteristic function

Definition 7.6. *Let X be any random variable. Then the* **characteristic function** *for X is given by*

$$\psi_X(t) \equiv E(e^{itX})$$

where $i \equiv \sqrt{-1}$.

Linear combinations of normal random variables

Theorem 7.10. *if $X_1, X_2, \ldots, X_n$ are independent normal random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ and $a_1, a_2, \ldots, a_n$ and b are constants, then*

$$Y = b + \sum_{i=1}^{n} a_i X_i \sim N(\mu_Y, \sigma_Y^2)$$

where

$$\mu_Y = b + \sum_{i=1}^{n} a_i \mu_i$$

and

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Example: The weights of snails are normally distributed with mean 6 gm. and variance 4 gm^2. Therefore, it is reasonable to assume that the weights of n snails chosen at random from the population of all snails are n independent identically distributed random variables, $X_1, X_2, \ldots, X_n$, with each $X_i \sim N(6, 4)$.

◇ Suppose that 4 snails are selected at random. and we would like to find the probability that the total weight, $T = X_1 + X_2 + X_3 + X_4$, will exceed 28 gm.?

Since $X_1, X_2, \ldots, X_n$, are iid with each $X_i \sim N(6, 4)$, we have $T = \sum_{i=1}^{4} X_i \sim N(24, 16)$. Therefore,

$$\begin{aligned} P(T > 28) &= P\left(\frac{T - 24}{4} > \frac{28 - 24}{4}\right) \\ &= P(Z > 1) \qquad \text{where } Z \sim N(0, 1) \\ &= 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587 \end{aligned}$$

◇ Once again, suppose that 4 snails are selected at random. Let $\overline{X} = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$ denote the average weight of the 4 snails. Let's find the probability that the observed average weight deviates from the expected weight by more than 2 gm. That is, we want to find $P(|\overline{X} - 6| > 2)$.

Since $X_1, X_2, \ldots, X_n$, are iid with each $X_i \sim N(6, 4)$, we have $\overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i \sim N(6, 1)$. Therefore,

$$\begin{aligned} P(|\overline{X} - 6| > 2) &= 1 - P(|\overline{X} - 6| \leq 2) \\ &= 1 - P(-2 \leq \overline{X} - 6 \leq 2) \\ &= 1 - P\left(\frac{-2}{1} \leq \frac{\overline{X} - 6}{1} \leq \frac{2}{1}\right) \\ &= 1 - P(-2 \leq Z \leq 2) \qquad \text{where } Z \sim N(0, 1) \\ &= 1 - [P(Z \leq 2) - P(Z \leq -2)] \\ &= 1 - [P(Z \leq 2) - (1 - P(Z \leq 2))] \\ &= 2 - 2P(Z \leq 2) = 2 - 2(0.9772) = 0.0456 \end{aligned}$$

The distribution of the sample mean

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Remember that $\overline{X}$, the sample mean, is a random variable. Then

$$\overline{X} \sim N(\mu, \sigma^2/n).$$

The Central Limit Theorem

Theorem 7.11. **(Central Limit Theorem)** *If*

1. $X_1, X_2, \ldots, X_n$ *is a random sample with* $E(X_i) = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2$ *for all* i*, and*
2. $S_n = X_1 + X_2 + \cdots + X_n$*, and*
3. $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$,

then

$$\lim_{n\to\infty} P(Z_n \le z) = \Phi(z).$$

Proof. First note that $E(S_n) = n\mu$ and $\mathrm{Var}(S_n) = n\sigma^2$. We will get the result by showing that the moment generating function for Z_n must approach the moment generating function of a $N(0,1)$ random variable. Since each X_i has the same moment generating function for all i, we can let $M(t)$ denote the common moment generating function of each $X_1, X_2, \ldots, X_n$. Then

$$M_{Z_n}(t) = M_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = M_{\frac{S_n}{\sigma\sqrt{n}} - \frac{\sqrt{n}\mu}{\sigma}}(t)$$

Since $M_{aY+b}(t) = e^{tb}M_Y(at)$ for any random variable Y and any constants a and b, we get

(1) $$M_{Z_n}(t) = e^{-\mu\sqrt{n}t/\sigma} M_{S_n}(t/(\sigma\sqrt{n}))$$

Since each X_i has the same moment generating function for all i, we can let $M(t)$ denote the common moment generating function of each $X_1, X_2, \ldots, X_n$. Therefore, $M_{S_n}(t) = [M(t)]^n$, and

$$M_{Z_n}(t) = e^{-\mu\sqrt{n}t/\sigma}\left[M(t/(\sigma\sqrt{n}))\right]^n$$

We are going to need the following two facts:

(A) $M(t) = 1 + \sum_{k=1}^{\infty} \frac{E(X^k)}{k!} t^k$

(B) $\ln(1+\xi) = \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} \xi^j \qquad \text{for } |\xi| < 1.$

Now, take the natural log of both sides of equation (1) and use fact (A):

$$\begin{aligned} \ln M_{Z_n}(t) &= \frac{-\mu\sqrt{n}t}{\sigma} + n \ln M\left(\frac{t}{\sigma\sqrt{n}}\right) \\ &= \frac{-\mu\sqrt{n}t}{\sigma} + n \ln\left(1 + \sum_{k=1}^{\infty} \frac{E(X^k)}{k!} \left(\frac{t}{\sigma\sqrt{n}}\right)^k\right). \end{aligned}$$

Note that the natural log term looks like fact (B) with

$$\xi = \sum_{k=1}^{\infty} \frac{E(X^k)}{k!} \left(\frac{t}{\sigma\sqrt{n}}\right)^k.$$

As $n \to \infty$, we can make this summation as small as we want. Hence, for suitably large n, the summation (i.e., ξ) will eventually be less than one, permitting us to use fact (B), so

$$\begin{aligned} &\ln M_{Z_n}(t) \\ &= \frac{-\mu\sqrt{n}t}{\sigma} + n \left[\sum_{j=1}^{\infty} \left\{ \frac{(-1)^{j-1}}{j} \left(\sum_{k=1}^{\infty} \frac{E(X^k)}{k!} \left(\frac{t}{\sigma\sqrt{n}}\right)^k \right)^j \right\} \right] \\ &= \frac{-\mu\sqrt{n}t}{\sigma} + n \left[\left(\sum_{k=1}^{\infty} \frac{E(X^k)}{k!} \left(\frac{t}{\sigma\sqrt{n}}\right)^k \right) - \frac{1}{2} \left(\sum_{k=1}^{\infty} \frac{E(X^k)}{k!} \left(\frac{t}{\sigma\sqrt{n}}\right)^k \right)^2 + \cdots \right]. \end{aligned}$$

Collect similar powers of t to obtain

$$\ln M_{Z_n}(t) = \frac{-\mu\sqrt{n}t}{\sigma} + n \left[\left(\frac{E(X)}{\sigma\sqrt{n}} \right) t + \left(\frac{E(X)}{2!\sigma^2 n} - \frac{1}{2} \frac{E(X)^2}{\sigma^2 n} \right) t^2 + \cdots \right].$$

Hence,

$$\ln M_{Z_n}(t) = \left(\frac{(-\mu + E(X))\sqrt{n}}{\sigma} \right) t + \left(\frac{E(X^2) - E(X)^2}{2\sigma^2} \right) t^2 + [\text{other terms}]$$

But $E(X) = \mu$ and $E(X^2) - E(X)^2 = \sigma^2$. Therefore,

$$\ln M_{Z_n}(t) = 0t + \frac{1}{2}t^2 + [\text{other terms}].$$

But the “[other terms]” for t^n with $n \geq 3$ go to zero as $n \to \infty$ because there is a power of n in the denominator of each of the “[other terms]” (expand the series further if you need convincing). Hence,

$$\lim_{n\to\infty} \left(\ln M_{Z_n}(t)\right) = \frac{t^2}{2}.$$

Therefore,

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}$$

which is the moment generating function for $N(0,1)$. So $\lim_{n\to\infty} P(Z_n \leq a) = \Phi(a)$. ■

The DeMoivre-Laplace Theorem

A proof of the Central Limit Theorem was provided by Liapounoff in 1902. But DeMoivre (1733) and Laplace (1812) proved a similar result for the following special case:

Let $T_1, T_2, \ldots, T_n$ be a random sample from a Bernoulli distribution, that is let

$$\begin{aligned} P(T_i = 1) &= p \\ P(T_i = 0) &= 1 - p. \end{aligned}$$

Then $E(T_i) = p$ and $\mathrm{Var}(T_i) = p(1-p)$. We already know that $S_n = \sum_{i=1}^n T_i$ has a Binomial distribution with parameters n and p, that is

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n.$$

However, because of the Central Limit Theorem,[2] we get the following

Theorem 7.12. **(DeMoivre-Laplace Theorem)** *As n gets large (i.e., as the number of Bernoulli trials increases) the distribution of*

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$$

approaches the distribution of a $N(0,1)$ random variable.

[2] Of course, DeMoivre and Laplace didn't know about the Central Limit Theorem and had to prove this in a different way.

Another (more casual) way to state the DeMoivre-Laplace Theorem is to say that as $n \to \infty$, the distribution of the random variable S_n approaches the distribution of a $N(np,\, np(1-p))$ random variable. In the limit, this is not very useful (and not precise), because as $n \to \infty$ both the mean np and variance $np(1-p)$ go to ∞. But for finite n, we get what is often called the "normal approximation to the binomial," as follows:

Corollary 7.13. *Suppose n is an integer and $0 < p < 1$, then for "large" values of n,*

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

This approximation usually works well when $np(1-p) > 9$.

The distribution of the sample mean for large n

Let $X_1, X_2, \ldots, X_n$ be a random sample from (just about any) distribution with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. And recall that the sample mean is defined as

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

Simply using the rules for expected value and variance, we get

$$\begin{aligned} E(\overline{X}) &= \mu \\ \text{Var}(\overline{X}) &= \sigma^2/n \end{aligned}$$

But, furthermore, since $\overline{X}$ is composed of a sum of independent random variables, we can use the Central Limit Theorem to say that the probability distribution of $\overline{X}$ approaches a $N(\mu,\, \sigma^2/n)$ distribution.

Combining this fact with the case discussed on page 211, we can summarize both results as follows:

- If $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Then the probability distribution of the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is $N(\mu,\, \sigma^2/n)$.
- If $X_1, X_2, \ldots, X_n$ is a random sample with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$ but the distribution of the X_i is not normal. Then, because of the Central Limit Theorem, we can still say that the probability distribution of the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is approximately $N(\mu,\, \sigma^2/n)$.

Self-Test Exercises for Chapter 7

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S7.1 Let X be a random variable with probability generating function

$$g(z) = \frac{z + z^5 + z^6}{3}$$

The value of $E(X)$ equals

(A) 0

(B) $\frac{1}{3}$

(C) 1

(D) 4

(E) none of the above.

S7.2 Suppose $X \sim N(2, 9)$ and $Y \sim N(2, 16)$ are independent random variables. Then $P(X + Y < 9)$ equals (approximately)

(A) 0.1587

(B) 0.5199

(C) 0.8413

(D) 0.9772

(E) none of the above.

S7.3 Suppose X is a random variable with moment generating function

$$M_X(t) = E(e^{tX}) = e^{0.5t^2}$$

Let X_1 and X_2 be two independent random variables, each with the same probability distribution as X. Then the moment generating function for $Y = X_1 + X_2$ is

(A) $e^{0.5t^2}$

(B) $2e^{0.5t^2}$

(C) e^{t^2}

(D) $e^{0.5t^4}$

(E) none of the above.

S7.4 An automated assembly line produces incandescent light bulbs. Suppose that a batch of n bulbs are selected at random. Let $T_1, T_2, \ldots, T_n$ be n independent identically distributed random variables with $\{T_i = 1\}$ if light bulb i is good, and $\{T_i = 0\}$ if light bulb i is bad. Then $S = \sum_{i=1}^{n} T_i$ is the total number of good light bulbs in the batch. If p is the probability that any one light bulb is good, then each T_i has probability mass function

$$p_{T_i}(k) = \begin{cases} 1-p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

where $0 < p < 1$. Then $P(S \geq 1)$ equals

(A) $1 - p^n$

(B) $1 - (1-p)^n$

(C) $\int_1^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$

(D) 1

(E) none of the above.

S7.5 If X and Y are independent, normally distributed random variables each with mean 0 and variance 1, then $3X - Y$

(A) has a normal distribution with mean 2 and variance 1

(B) has a normal distribution with mean 0 and variance 10

(C) has a normal distribution with mean 0 and variance 2

(D) has a normal distribution with mean 2 and variance 4

(E) none of the above.

S7.6 Let X and Y be independent random variables with $X \sim N(0,1)$ and $Y \sim N(0,2)$. The value of $P(X > Y)$ is

(A) 0

(B) 0.05

(C) 0.50

(D) 0.95

(E) none of the above.

S7.7 Let X be a random variable with probability generating function

$$g(z) = \frac{z(1+z)}{2}$$

The value of $P(X = 0)$ is

(A) 0

(B) $1/4$

(C) $1/2$

(D) 1

(E) none of the above.

S7.8 Let X_1, X_2, X_3, X_4 be independent random variables, each with $E(X_i) = 0$ and Var$(X_i) = 1$ for $i = 1, 2, 3, 4$. Let $Y = X_1 + X_2 + X_3 + X_4$. Then Var$(Y)$ equals

(A) $1/16$

(B) $1/4$

(C) 4

(D) 16

(E) none of the above.

S7.9 Let X and Y be independent, normally distributed random variables each with mean 0 and variance 1. Then the random variables

$$\begin{aligned} W_1 &= X + Y \\ W_2 &= X - Y \end{aligned}$$

(A) have the same mean

(B) have the same variance

(C) are both normally distributed

(D) all of the above

(E) none of the above.

S7.10 Let X be a random variable with moment generating function

$$M_X(t) = \frac{e^t(1 + e^{2t})}{2}$$

The value of $E(X)$ is

(A) 0

(B) 1

(C) 2

(D) 3

(E) none of the above.

S7.11 Let X be a random variable with moment generating function $M_X(t) = (1 - t)^{-1}$. Let Y be a random variable with moment generating function $M_Y(s) = (1 - s)^{-1}$. If X and Y are independent, then $X + Y + 1$ has a moment generating function equal to

(A) $e^t(1 - t)^{-2}$

(B) $(1 - t)^{-2}$

(C) $2e^t(1 - t)^{-1}$

(D) $e^s(1 - t)(1 - s)$

(E) none of the above.

S7.12 Jack is attempting to fill a 100 gallon water tank using a one gallon bucket. Jack repeatedly goes to a well at the top of a hill, fills his one gallon bucket with water and brings that water down to the 100 gallon tank. The amount of water he carries on any one trip from the well to the tank is normally distributed with mean 1 gallon and standard deviation 0.1 gallon. The amount of water carried on any trip is independent of all other trips. The probability that Jack will need to make more than 100 trips to the well is

(A) 0.05

(B) 0.0636

(C) 0.2580

(D) 0.5

(E) none of the above.

S7.13 Let X and Y be independent random variables with $X \sim N(0,1)$ and $Y \sim N(0,2)$. The value of $P(X - Y < 1)$ equals

(A) 0

(B) 0.5

(C) 1

(D) $P(X + Y < 1)$

(E) none of the above.

S7.14 Let X and Y be independent random variables with $X \sim N(1,2)$ and $Y \sim N(1,2)$. Then $P(|X - Y| \leq 1)$ equals

(A) 0

(B) 0.3830

(C) 0.6915

(D) 1.0

(E) none of the above.

S7.15 Let X be a discrete random variable with probability mass function given by

$$P(X = 0) = \tfrac{1}{3} \qquad P(X = 1) = \tfrac{2}{3}$$

Then the moment generating function for X is

(A) $e^{2t/3}$

(B) $\frac{1}{3}(1 + 2e^t)$

(C) $\frac{1}{2}(1 + e^t)$

(D) $\frac{1}{2}e^t(1 + e^{2t})$

(E) none of the above.

S7.16 Suppose $X_1, X_2, \dots, X_n$ are independent random variables with each $X_i \sim N(2,4)$ for $i = 1, 2, \dots, n$. Let $W = \sum_{i=1}^{n} X_i$ and define the random variable

$$Z = \frac{W - 2n}{c}$$

Then $Z \sim N(0,1)$ if and only if

(A) $c = 1$

(B) $c = n$

(C) $c = 2\sqrt{n}$

(D) $c = 2/\sqrt{n}$

(E) none of the above.

S7.17 Let X be a random variable with moment generating function

$$M_X(t) = \frac{e^t(1 + e^t)}{2}$$

The value of $E(X)$ is

(A) 0

(B) $\frac{2}{3}$

(C) 1

(D) $\frac{3}{2}$

(E) none of the above.

S7.18 Let $X_1, \dots, X_n$ be independent, identically distributed random variables with each X_i having a probability mass function given by

$$P(X_i = 0) = 1 - p \qquad P(X_i = 1) = p$$

where $0 \leq p \leq 1$. Define the random variable

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

Then $E(Y)$ equals

(A) 0

(B) p

(C) p^2

(D) 1

(E) none of the above.

Questions for Chapter 7

7.1 The probability generating function for a random variable X is given by

$$g(z) = \frac{z^{10} + z^{20}}{2}.$$

(a) Find $P(X > 18 \mid X > 13)$.

(b) Find $E(X)$ and $E(X^2 + 1)$.

(c) Find $E(|X - 15|)$.

(d) Find the moment generating function for X.

7.2 A random variable has the moment generating function

$$M(t) = \frac{2}{2 - t}.$$

What is the expected value for this random variable?

7.3 The moment generating function for a random variable X is given by

$$M(t) = (0.5)e^{26t} + (0.25)e^{27t} + (0.25)e^{49t}.$$

(a) Compute $P(X \leq 30)$.

(b) Find $E(X)$.

7.4 Let X be a random variable with nonnegative, integer support and moment generating function

$$M(t) = \frac{p}{1 - e^t(1 - p)}.$$

What is the value of $P(X = 0)$?

7.5 Suppose that X, Y, and Z are independent random variables with $X \sim N(1, 4)$, $Y \sim N(2, 9)$ and $Z \sim N(3, 16)$.

(a) Compute $P(X+Y \le 1)$.

(b) Find the number a such that $P(3-4a < Z < 3+4a) = 0.95$.

(c) Compute $P[(X+Y+Z)^2 \le 25]$.

(d) Compute $P[(3X-Z)^2 > 52]$. 0.3174

7.6 Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables, each with $E(X_i) = 0$ and $\text{Var}(X_i) = 1$ for all $i = 1, 2, \ldots, n$.

(a) Using Chebychev's inequality, find a number a (that will depend on n), such that

$$P(|\overline{X}_n| \le a) \ge 0.95$$

where $\overline{X}_n$ denotes the sample mean of $X_1, X_2, \ldots, X_n$.

(b) Solve question (a) again using the additional information that each of the X_i is normally distributed. *Hint:* Don't use Chebychev's inequality. You should be able to get this answer knowing the exact distribution of X_n.

(c) Compare the answers to (a) and (b). Explain the difference in your answers.

7.7 A large forklift has a 0.5 ton (1000 pound) capacity. Cartons of items have individual weights that are normally distributed with mean 25 pounds and standard deviation 2 pounds. These weights can be assumed to be independent and identically distributed. Approximately how many cartons can the forklift carry so that the probability that its total load exceeds its capacity is only 0.01?

7.8 A fair coin is tossed 5000 times. Approximately what is the probability that at least 1965 of the tosses will turn up heads?

7.9 A one-pound box of breakfast cereal never contains *exactly* one pound of product. The net weight (in pounds) is normally distributed with variance 0.01 and mean μ that can be varied by adjusting the packaging equipment.

(a) What value should μ be adjusted to so that the producer can guarantee that the probability that the net weight of a box is less than one pound is only 0.05?

(b) What is the probability that a carton of 16 cereal boxes produced according to the specifications determined in part (a) will not exceed a total net weight of 19.28 pounds?

7.10 The distribution of test scores on an examination is approximately normal with mean 70 and standard deviation 7.

(a) What is the probability of receiving a grade less than 80?

(b) If 49 students take the test, what is the probability that the average of their scores lies between 68 and 72?

(c) What is the probability that the difference in the scores of two students (picked at random) lies between -5 and $+5$?

7.11 *True or False:* If X and Y are normally distributed random variables with mean 0 and variance 1, then $3X - Y$ has a normal distribution with mean 0 and variance 10.

7.12 *The Phizzi Soft Drink Company's* best selling product is *Phizzi-Soda* that it sells in 12-ounce bottles. When the automated filling machine is working properly, the quantity of Phizzi-Soda in a single bottle is a random variable X. Assume that X is normally distributed with mean 12.1 ounces and variance 0.0025 (ounces)^2.

(a) Let X_1 denote the fill-volume of a single bottle of Phizzi-Soda.

Find $P(X_1 < 12)$.

(b) Suppose that two bottles of Phizzi-Soda are chosen at random. Let X_1 and X_2 be independent random variables denoting the fill-volume of each of the two bottles.

Find $P(X_1 > X_2)$. *Hint:* $P(X_1 > X_2) = P(X_1 - X_2 > 0)$.

(c) Let $\overline{X} = (X_1 + X_2)/2$ be a random variable representing the average fill-volume of the two bottles selected in question (b).

Find $P(\overline{X} < 12)$.

7.13 Let X be a continuous random vector with probability density function given by

$$f_X(x) = \begin{cases} 1/\theta & \text{for } 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases},$$

where θ is a constant with $\theta > 0$. Your answers to the following questions might depend on the constant θ.

(a) Find $\text{Var}(X)$.

(b) Find the moment generating function for X.

7.14 Individual packages of writing paper have weights that are normally distributed each with mean 2 kgm and variance $0.25(\text{kgm})^2$. The weights of all packages can be considered independent, identically distributed random variables.

(a) Suppose that one package of paper is weighed. What is the probability that its weight is greater than 2.5 kgm?

(b) Suppose that ten packages of paper are weighed together. What is the probability that their *total* weight is greater than 25 kgm?

(c) Suppose that ten packages of paper are weighed together. What is the probability that their *average* weight is greater than 2.1 kgm?

7.15 Prove that if $g_X(z) \equiv E(z^x)$ is a probability generating function for X, then $\lim_{z\to 1} g_X(z) = 1$.

7.16 Let X and Y be independent, normally distributed random variables with $X \sim N(1, 9)$ and $Y \sim N(6, 16)$.

(a) Find the number a having the property that $P(Y > a) = 0.975$.

(b) Find $P(X > Y)$.

(c) Compute $P(|X - 1| < 4)$.

(d) Compute $P(0 \leq X < 2)$.

7.17 Referring to the previous question, show that if $Y_1, Y_2, \ldots, Y_n$ are independent log-normal random variables, then

$$W = \prod_{i=1}^{n} Y_i$$

is a log-normal random variable.

7.18 At one stage in the assembly of an article, a cylindrical rod with a circular cross-section has to fit into a circular socket. Quality control measurements show that the distribution of rod diameter and socket diameter are independent and normal with parameters:

Rod diameter:	$\mu = 5.01$ cm.	$\sigma = 0.03$ cm.
Socket diameter:	$\mu = 5.10$ cm.	$\sigma = 0.04$ cm.

If the components are selected at random for the assembly, what proportion of rods will not fit?

7.19 Bill Gutts, one of the wealthiest men on earth, wants to fill one of the swimming pools at his mountain-top home. The elevation of the pool (and his home) is very high so that it is impossible to pump water up the mountain. For that reason, Mr. Gutts plans to transport the water to the mountain-top pool using a fleet of 200 helicopters with each helicopter carrying a small water-tank.

The capacity of the swimming pool is 198869 gallons.

The amount of water transported varies from helicopter to helicopter. But it is well-known that the amount of water (in gallons) carried by a single helicopter can be represented by a random variable that is normally distributed with mean 1000 and variance 1600. The amount of water carried the 200 individual helicopters can be assumed to be a collection of independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ with $n = 200$.

Let $Y = \sum_{i=1}^{n} X_i$ be a random variable denoting the total water carried by all of the $n = 200$ helicopters.

(a) Find $E(Y)$ and $\text{Var}(Y)$

(b) What is the probability that Mr. Gutts can fill his swimming pool using the 200 helicopters?

(c) Helicopters are very expensive, even for a wealthy guy like Mr. Gutts. Suppose Mr. Gutts would be happy if the probability that his swimming pool is completely filled is only 0.50. How many helicopters should he use?

(d) In an E-mail message, Mr. Gutts is given the bad news that the amount of water carried by each helicopter does not have a $N(1000, 1600)$ distribution. Instead it is has a uniform distribution on the interval $(930.718, 1069.282)$. Use the Central Limit Theorem to find the approximate probability that Mr. Gutts can fill his swimming pool if he uses 200 helicopters.

7.20 (†) Suppose X and Y are independent random varaibles. The continuous random variable X has a uniform distribution on the interval $(0, 1)$. That is, X has density function

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

The discrete random variable Y is a Bernoulli random variable with $p = 1/2$. That is,

$$P(Y = k) = \begin{cases} 1/2 & \text{if } k = 0 \\ 1/2 & \text{if } k = 1 \end{cases}$$

Find the probability distribution of the random variable $W = X + Y$.

8

ESTIMATION

Good and bad luck are often mistaken for good and bad jusment
– ANONYMOUS, *Salada* ™ *tea bag tag*

Basic Concepts

Random samples

Definition 8.1. *A* **random sample** *is any sequence,* $X_1, X_2, \ldots, X_n$ *of independent, identically distributed random variables.*

If a continuous random variable X has an unknown parameter θ, we often write it's probability density function as

$$f_X(x; \theta).$$

Example: Let X have an exponential distribution with probability density function

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{1}{\theta}x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where θ is unknown. Then a random sample, $X_1, X_2, \ldots, X_n$ representing n independent observations from this random variable has a joint probability density function given by

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{1}{\theta}(x_1+x_2+\cdots+x_n)} & x_i > 0 \text{ for all } i \\ 0 & \text{otherwise} \end{cases}$$

Statistics

Definition 8.2. *A* **statistic** *is any function,* $h(X_1, X_2, \ldots, X_n)$*, of independent, identically distributed random variables.*

If a statistic is being used to establish statistical inferences about a particular unknown parameter θ, then it is often written as $\hat{\theta}(X_1, X_2, \ldots, X_n)$.

Note: *Any* function of a random sample can be called a statistic.

Example: Let $X_1, X_2, \ldots, X_n$ be a random sample. You can think of the random variables $X_1, X_2, \ldots, X_n$ as representing the future data values that will be obtained by making n independent observations of our experiment.

Then the following are statistics:

1. $\hat{\theta}(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$

   The random variable $\hat{\theta}$ is the average of the observations.

2. $\hat{\theta}(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} X_i$

   The random variable $\hat{\theta}$ is the product of the observations.

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

   The random variable $\hat{\theta}$ is the largest of the observations.

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

   The random variable $\hat{\theta}$ is the first observation, with the others discarded.

There are two important statistics that arise so frequently that they are given special symbols:

Definition 8.3. *Let $X_1, X_2, \ldots, X_n$ be a random sample. The* **sample mean** *is the statistic*

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

and the **sample variance** *is the statistic*

$$s_X^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Point Estimation

The problem

Let X be a random variable with cumulative distribution function $F_X(x;\theta)$ that is completely determined except for the unknown constant θ. A random sample $X_1, X_2, \ldots, X_n$ is to be drawn from this distribution. Suppose that a statistic $\hat{\theta}(X_1, X_2, \ldots, X_n)$ has been already been selected for use.

Let $x_1, x_2, \ldots, x_n$ denote the actual observed values taken on by $X_1, X_2, \ldots, X_n$. Applying the function $\hat{\theta}$ to these observations yields the number $\hat{\theta}(x_1, x_2, \ldots, x_n)$. This number can be thought of as an actual observation of the random variable $\hat{\theta}(X_1, X_2, \ldots, X_n)$. We will call the number $\hat{\theta}(x_1, x_2, \ldots, x_n)$ our **estimate** for the value of θ. We call the random variable $\hat{\theta}(X_1, X_2, \ldots, X_n)$ an **estimator** for θ.

Since any statistic can be called an estimator of any parameter, we will have to find ways to distinguish good estimators from poor estimators.

Unbiased estimators

Definition 8.4. *The estimator $\hat{\theta}$ is an* **unbiased estimator** *of θ if its expected value equals θ, i.e.,*

$$E(\hat{\theta}) = \theta$$

for all θ.

Example: Let X be a discrete random variable with probability mass function given by $p_X(0) = 1 - \theta$ and $p_X(1) = \theta$, where θ is unknown. If $X_1, X_2, \ldots, X_n$ is a random sample from X then

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is an unbiased estimator for θ.

Finding Estimators

Method of maximum likelihood

Definition 8.5. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a discrete random variable X with probability mass function $p_X(x;\theta)$. The* **likelihood function** *for*

$X_1, X_2, \ldots, X_n$ is given by

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} p_X(x_i; \theta).$$

If, instead, X is a continuous random variable with probability density function $f_X(x; \theta)$, then the likelihood function is given by

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta).$$

Maximum likelihood estimators are obtained by finding that value of θ that maximizes L for a given set of observations $x_1, x_2, \ldots, x_n$. Since the value of θ that does this will usually vary with $x_1, x_2, \ldots, x_n$, θ can be thought of as a function of $x_1, x_2, \ldots, x_n$, namely $\hat{\theta}(x_1, x_2, \ldots, x_n)$.

To evaluate the properties of $\hat{\theta}$, we can look at its performance *prior* to actually making the observations $x_1, x_2, \ldots, x_n$. That is we can substitute X_i for x_i in the specification for $\hat{\theta}$ and look at its properties as the statistic

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

For example, one of the properties that we might like to check for is whether $\hat{\theta}$ is an *unbiased* estimator for θ (i.e., check to see if $E(\hat{\theta}) = \theta$).

Self-Test Exercises for Chapter 8

For the following multiple-choice question, choose the best response among those provided. The answer can be found in Appendix B.

S8.1 Suppose that $X_1, X_2, \ldots, X_n$ are independent identically distributed random variables each with marginal probability density function

$$f_{X_i}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

for $-\infty < x < +\infty$, where $\sigma > 0$. Then an unbiased estimator for μ is

(A) $(X_1)(X_2)\cdots(X_n)$

(B) $(X_1 + X_2)^2/2$

(C) $\frac{1}{n}\sum_{i=1}^{n} X_i$

(D) σ

(E) none of the above.

Questions for Chapter 8

8.1 Let X be a random variable with a binomial distribution, i.e.,

$$p_X(k;\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n.$$

Let X_1 be a random sample of size 1 from X.

(a) Show that $\hat{\theta} = X_1/n$ is an unbiased estimator for θ.

(b) Show that $\hat{\theta} = X_1/n$ is the maximum likelihood estimator for θ.

8.2 Let Y be an estimator for θ based on the random sample $X_1, X_2, \ldots, X_n$. Suppose that $E(X_i) = \theta$ and $Y = \sum_{i=1}^n a_i X_i$ where $a_1, a_2, \ldots, a_n$ are constants. What constraint must be placed on $a_1, a_2, \ldots, a_n$ in order for Y to be an unbiased estimator for θ?

8.3 The life of a light bulb is a random variable X which has probability density function

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{1}{\theta}x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from X.

(a) Find an estimator for θ using the method of maximum likelihood.

(b) Is the estimator for part (a) unbiased? Justify your answer.

(c) Find the maximum likelihood estimator for $\eta = 1/\theta$.

8.4 The number of typographical errors, X, on a page of text has a Poisson distribution with parameter λ, i.e.,

$$p_X(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad \text{for } k = 0, 1, 2, \ldots.$$

A random sample of n pages are observed

(a) Find an estimator for λ using the method of maximum likelihood.

(b) Is the estimator for part (a) unbiased? Justify your answer.

8.5 Given the random sample $X_1, X_2, \ldots, X_n$ consider the statistic d^2 formed by averaging the squared differences of all possible pairings of $\{X_i, X_j\}$. There are $\binom{n}{2}$ such pairs. That statistic can be represented as

$$d_X^2 \equiv \frac{1}{\binom{n}{2}} \sum_{i>j} (X_i - X_j)^2$$

Prove that $d_X^2 = 2s_X^2$.

9

PROBABILITY MODELS

Coin Tossing (Binary) Models

Definition 9.1. *A* **Bernoulli trial** *is an experiment with only two outcomes. These outcomes are conventionally called* success *and* failure.

Definition 9.2. *A* **Bernoulli random variable**, T *is a random variable with a probability mass function given by*

$$\begin{aligned} p_T(0) &= 1 - p \\ p_T(1) &= p \end{aligned}$$

Definition 9.3. *Let* $T_1, T_2, \ldots, T_n$ *be a sequence of* n *independent and identically distributed Bernoulli random variables, with* $P(T_i = 1) = p$ *for all* i*. Let* $X = \sum_{i=1}^{n} T_i$*. Then* X *is said to have a* **binomial distribution**.

Note: X denotes the total number of successes from n independent and identically distributed Bernoulli trials.

Theorem 9.1. *If* X *is a binomial random variable with parameters* n *and* p*, then*

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, \ldots, n$$

The binomial distribution

Support: $0, 1, 2, \ldots, n$

Probability mass function: $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
$k = 0, 1, 2, \ldots, n$

Probability generating function: $g(z) = (zp + (1-p))^n$

Moment generating function: $M_X(t) = (e^t p + (1-p))^n$

Moments: $E(X) = np \qquad \text{Var}(X) = np(1-p)$

Notation

If X has a binomial distribution with parameters n and p, we often write $X \sim B(n,p)$.

The geometric distribution

Support: $1, 2, \ldots$

Probability mass function: $p_X(k) = p(1-p)^{k-1} \qquad k = 1, 2, \ldots$

Probability generating function: $g(z) = \dfrac{zp}{1 - z(1-p)}$ for $|z| < \dfrac{1}{1-p}$

Moment generating function: $M_X(t) = \dfrac{e^t p}{1 - e^t(1-p)}$

Moments: $E(X) = 1/p \qquad \text{Var}(X) = (1-p)/p^2$

Note

X denotes the number of independent and identically distributed Bernoulli trials required until the first success (counting the successful trial).

Example: A computer breaks down at least once a day with probability 0.8. It therefore remains operational for an entire day with probability 0.2.

1. What is the probability that it breaks down exactly 3 days out of a week?

   Solution: Let X denote the number of breakdowns, $X \sim B(7,\, 0.8)$. Hence,

   $$P(X = 3) = \binom{7}{3}(0.8)^3(0.2)^4 = 0.287$$

2. What is the probability that it is operational for 3 days in a row before a breakdown?

   Solution: Let Y denote the number of days until a breakdown.

   $$P(Y = 4) = (0.2)^3(0.8) = 0.0064$$

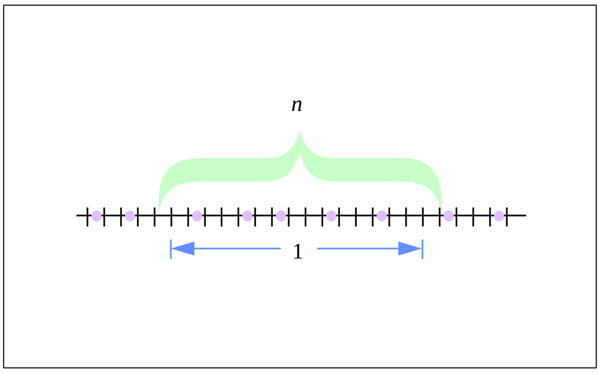
The Poisson distribution

Support:	$0, 1, 2, \ldots$
Probability mass function:	$p_X(k) = \dfrac{e^{-\alpha}\alpha^k}{k!}$ $\quad k = 0, 1, 2, \ldots$
Probability generating function:	$g(z) = e^{\alpha(z-1)}$ for $\|z\| < 1$
Moment generating function:	$M_X(t) = e^{\alpha(e^t-1)}$
Moments:	$E(X) = \alpha \qquad \text{Var}(X) = \alpha$

Theorem 9.2. *If X is a binomial random variable with parameters n and p, then for "large" n and "small" p*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{e^{-np}(np)^k}{k!}$$

Random Particle Models



The expected number of particles per interval of length one is equal to α. Suppose the interval is divided into n equal segments with at most one particle placed in any segment. What is the probability that there are exactly k particles in the unit interval?

Let X denote the number of particles in the unit interval. X is a binomial random variable with parameters n and $p = \alpha/n$. Therefore,

$$P(X = k) \quad = \quad \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k}$$

$$= \left(1 - \frac{\alpha}{n}\right)^{-k} \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \cdots \left(\frac{n-k+1}{n}\right) \frac{\alpha^k}{k!} \left(1 - \frac{\alpha}{n}\right)^n$$

for $k = 0, 1, 2, \ldots, n$.

Taking the limit as n gets large (the segment widths decrease) we obtain

$$\begin{aligned} \lim_{n\to\infty} P(X = k) &= 1(1)(1)\cdots(1)\frac{\alpha^k}{k!}e^{-\alpha} \\ &= \frac{e^{-\alpha}\alpha^k}{k!} \end{aligned}$$

for $k = 0, 1, 2, \ldots$.

This limiting particle process is called the **Poisson Process** with

$$\begin{aligned} \alpha &= \text{expected number of occurrences per unit interval} \\ X &= \text{number of occurrences in a unit interval} \end{aligned}$$

and the probability mass function for X is given by

$$p_X(k) = \frac{\alpha^k e^{-\alpha}}{k!} \qquad k = 0, 1, 2, \ldots$$

Let Y_i denote the interarrival time between occurrences $i-1$ and i (see Figure 9.1). Each of the Y_i are independent and identically distributed random variables with probability density function given by

$$f_{Y_i}(y) = \begin{cases} \alpha e^{-\alpha y} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Question: Suppose a person begins to observe the process at time t_0 chosen independently of the future of the process? What is the distribution of the waiting time to the next occurrence, W (see Figure 9.1)?

Theorem 9.3.

$$f_W(w) = \begin{cases} \alpha e^{-\alpha w} & \text{if } w \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

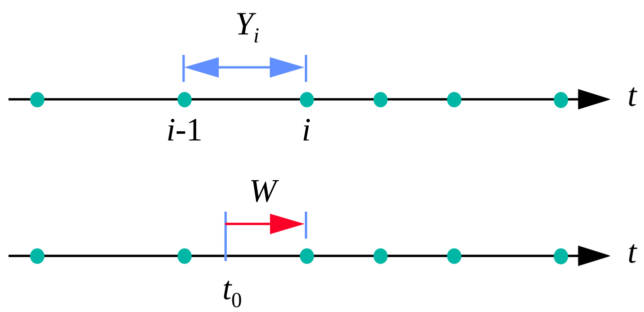


Figure 9.1: Interarrival and waiting times

Self-Test Exercises for Chapter 9

For each of the following multiple-choice questions, choose the best response among those provided. Answers can be found in Appendix B.

S9.1 It costs the Cheapy Flashbulb Company c cents to manufacture a single flashbulb. The company sells each bulb for r cents. The probability that any one bulb is good is $p > 0$ and it is independent of all other bulbs. Flashbulbs are packed n to a box. If any of the bulbs in a box is bad, Cheapy will refund the purchase price of the entire box to the customer. What is the probability that every bulb in a box is good?

(A) 0

(B) p

(C) $1 - p$

(D) p^n

(E) none of the above.

S9.2 Using the information in the previous question, let Y denote the profit from selling a box of Cheapy flashbulbs. Note that Y may be negative. The value of $E(Y)$ is

(A) 0

(B) $n(rp - c)$

(C) $nrp^n - nc$

(D) $nrp^n + nc - 2ncp^n$

(E) none of the above.

S9.3 The number of knots in a piece of lumber is a random variable, X. Suppose that X has a Poisson distribution with $E(X) = \alpha > 0$. If four independent pieces of lumber are examined, the probability that they all have no knots is

(A) $1/4$

(B) $4e^{-\alpha}$

(C) $e^{-4\alpha}$

(D) $\alpha^4 e^{-4\alpha}$

(E) none of the above.

Questions for Chapter 9

9.1 A complicated piece of machinery, when running properly, can bring a profit of c dollars per hour ($c < 2$) to a firm. However, this machine breaks down at unexpected and unpredictable times. The number of breakdowns during any period of length t hours is a random variable with a Poisson distribution with parameter t. If the machine breaks down x times during the t hours, the loss incurred (from the shutdown plus repair costs) equals $(x^2 + x)$. Let P be the random variable representing the total profit from running the machine for t hours.

(a) Find $E(P)$ as a function of t.

(b) Find the length of time the machine should be operated to maximize the expected profit.

9.2 The number of oil tankers, N, arriving at a refinery each day has a Poisson distribution with parameter $\alpha = 3$. Present port facilities can service 2 tankers each day. If more than 2 tankers arrive in a day, the excess tankers must be sent to another port.

(a) On any given day, what is the probability of having to send tankers away?

(b) How much must the present facilities be increased to permit handling all tankers on approximately 80% on the days?

(c) What is the expected number of tankers arriving per day?

(d) What is the expected number of tankers serviced daily?

(e) What is the expected number of tankers turned away daily?

(f) If a fee of \$10000 is collected from each tanker serviced, what is the expected total fees collected in one day?

(g) How much must the present facilities be increased to assure that the expected total fees collected in one day is at least \$30000?

9.3 Suppose cars on a very long test track travel at exactly 50 miles per hour and pass a check point on the track according to a Poisson process at a rate of 5.2 cars/hour.

(a) What is the average distance between cars?

(b) What is the probability that the distance between any two cars is greater than 10 miles?

9.4 Suppose that a book of 620 pages contains 55 typographical errors. If these are randomly distributed throughout the book, what is the probability that 10 pages, selected at random, will be free of errors?

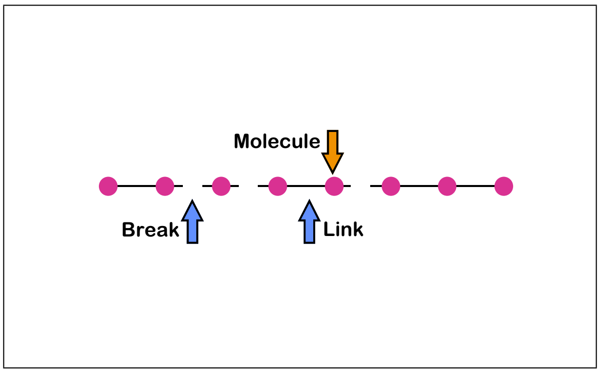
9.5 In forming binary numbers with 32 binary digits, the probability that an incorrect digit will appear is 0.002. If the errors are independent, what is the probability of finding zero, one, or more than one incorrect digit in a binary number? If a computer forms 10^6 such 32-digit numbers per second, what is the probability that an incorrect number is formed during any one-second time period?

9.6 Let X_1, X_2, X_3, X_4, X_5 be independent and identically distributed Bernoulli random variables, each with $P(X_i = 1) = p$. If

$$P\left(\sum_{i=1}^{5} X_i = 5\right) = r,$$

then what is the value of p in terms of r?

9.7 A chromosome can be considered a string of molecules, with a simple link connecting exactly two molecules (see diagram below)



Suppose a section of a chromosome consists of 1000 such links (and 1001 molecules). In recombinant DNA experiments, some of these links are broken, thus breaking the section into smaller chains. Each link has a probability 0.005 of breaking, and each link breaks independently of the others.

(a) Let N be the number of links broken. What is the actual probability mass function for N?

(b) Compute the *approximate* probability that exactly 4 smaller chains result from the breakup.

(c) Consider the number of molecules on the smaller chains that result from the breakup. Approximately what proportion of these chains will contain at least 200 molecules?

9.8 A computer transmits 1000 words per minute. The probability that any one word has a parity error is 0.001. Parity errors are generated independently from one word to the next.

(a) Approximately what is the probability of generating at least 2 words with parity errors in a one minute period?

(b) An error cycle is the time between successive words with parity errors. Approximately what is the probability that an error cycle will be longer than two minutes?

(c) A shift change in computer personnel takes place at 8:00 AM each day. What is the probability that the computer operator on the new shift will wait at least two minutes for the first parity error on that shift?

(d) What is the probability that no parity error occurs between 10:00 and 10:02 AM?

9.9 The number of knots in a piece of lumber has a Poisson distribution with an average of 2 knots per piece. If a piece has any knots, it is considered unsatisfactory. The number of knots from piece to piece can be assumed to be independent.

(a) What is the probability that a piece of lumber is unsatisfactory?

(b) What is the expected number of unsatisfactory pieces in a load of 500 pieces?

(c) What is the probability that all of the 500 pieces are satisfactory?

9.10 Glass rods are manufactured by cutting a continuous glass rod leaving a production machine. Small imperfections appear in the continuous rod according to a Poisson process with an average of two imperfections per foot of glass rod. To produce "perfect" glass rods, the continuous rod is cut at each imperfection.

(a) What is the average length of the cut glass rods?

(b) What fraction of the rods will have a length greater than one foot?

9.11 (†) Suppose X_1 and X_2 are independent Binomial random variables with

$$\begin{aligned} P(X_1 = k) &= \binom{n}{k} p_1^k (1-p_1)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n \\ P(X_2 = k) &= \binom{n}{k} p_2^k (1-p_2)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n \end{aligned}$$

(a) *Prove or give a counterexample:* Suppose $0 < p_i < 1$ for $i = 1, 2$. If $p_1 = p_2$ then $Y = X_1 + X_2$ has a Binomial distribution.

(b) *Prove or give a counterexample:* Suppose $0 < p_i < 1$ for $i = 1, 2$. If $p_1 \neq p_2$ then $Y = X_1 + X_2$ cannot have a Binomial distribution.

(c) Suppose $0 < p_i < 1$ for $i = 1, 2$ and let $Y = X_1 + X_2$. Let $W = Y - n$. Find the limiting distribution for each of the following random variables:

(i) Y as $p_2 \to 0$.

(ii) W as $p_2 \to 1$.

9.12 (†) (from Meyer[1]) In the manufacturing of glass bottles, small particles in the molten glass cause individual bottles to be defective. A single bottle requires 1 kg of molten glass. On average, there are x particles in a batch of 100 kg of molten glass. Assume that: (a) every particle may appear in any bottle with equal probability, and (b) the distribution of any particle among the bottles is independent of all other particles. A defective bottle may have more than one particle. Note that the molten glass required for 100 bottles contains, on average, x particles. If we wanted to compute the expected percentage of defective bottles, the quick (and wrong) answer would be x percent.

(a) Show that for M 100 kg batches of bottles, the number of particles in a randomly chosen bottle, Z, has probability distribution

$$P(Z = k) = \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k} \quad \text{for } k = 0, 1, \ldots, n$$

where $n = xM$ and $\alpha = x/100$.

(b) Show that as $M \to \infty$

$$P(Z = k) \approx \frac{e^{-\alpha} a^k}{k!} \quad \text{for } k = 0, 1, \ldots$$

(c) Show that, for large M, the expected percentage of defective bottles equals $d(x) = 100(1 - e^{x/100})$.

(d) Finally, show that for small x, we have $d(x) \approx x$, but for larger x (say $x = 100$) the "quick" solution is very wrong.

[1] Meyer, P., *Introductory Probability and Statistical Applications,* Addison-Wesley, 1965

9.13 The *Youbee Light Bulb Company* manufactures electric light bulbs. Each bulb has a probability be being defective equal to 0.1. Defects from one bulb to the next can be assumed to be independent random variables.

Let the Bernoulli random variable T_i equal 1 if bulb i is defective and 0 if the bulb is non-defective. Therefore, for a box of 10 bulbs $T_1, T_2, \ldots T_{10}$ are independent, identically distributed random variables with

$$\begin{aligned} P(T_i = 0) &= 0.9 \\ P(T_i = 1) &= 0.1 \end{aligned}$$

for all $i = 1, 2, \ldots, 10$.

(a) Compute $E(T_i)$ and Var(T_i) for every i.

(b) Let

$$X = T_1 + T_2 + \cdots + T_{10}$$

denote the total number of defective bulbs in a box of 10 bulbs. Compute $E(X)$ and Var(X). Clearly justify each step of your answer.

(c) Find $P(X = 0)$.

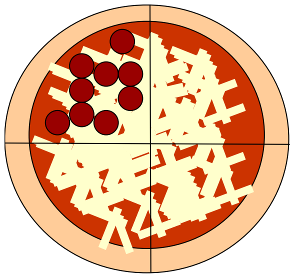
9.14 (†) Mr. Bonkers is selling his house. He decides he will accept the first offer exceeding c dollars, where c is a positive constant. Suppose the offers $X_1, X_2, \ldots$ are independent identically distributed random variables (in dollars) each with cumulative distribution function $F_X(x)$. Let N be the number of offers needed to sell the house. Let Y be the actual selling price of the house.

(a) Find the probability mass function for N.

(b) Find $E(N)$

(c) Find the cumulative distribution function for Y. That is, find $F_Y(y) = P(Y \leq y)$.

9.15 (†) *Pip's Pizza Shoppe* sells pepperoni and plain pizza by the slice. There are four slices in each pizza. Since one fourth of the slices sold by Pip are pepperoni, every pizza that Pip prepares has pepperoni on only one of the four slices. The remainder of the pizza is plain. See the following diagram:



Let the independent random variables $X_1, X_2, \ldots$ represent the slices sold by Pip with $X_i = 1$ if slice i is pepperoni, and $X_i = 0$ if it's plain. Note that a sequence such as

$$X_1 = 1,\ X_2 = 0,\ X_3 = 1,\ X_4 = 0,\ X_5 = 0, \ldots$$

would require that Pip bake a second pizza before selling all of the slices in the first pizza. Assume that Pip never throws away any slices of pizza during the day.

(a) Suppose 20 customers visit Pip's Pizza Shoppe on a particular day. Let the random variable Y denote the number of pizzas that Pip must bake that day. What is the probability distribution for the random variable Y? *Hint:* There is only one pepperoni slice per pie.

(b) Suppose Pip has just taken a freshly baked pizza out of the oven and suppose that there are no other unsold slices. What is the probability that Pip must bake a another pizza before selling all of the slices in the current pizza? *Hint:* Find the probability that he doesn't need to bake another pizza.

(c) (††) What is the probability that Pip must bake a pizza to satisfy the request of the current customer? That is, the current customer wants pepperoni but Pip has no pepperoni slices available *or* the current customer wants plain but Pip has no plain slices available.

9.16 (†) In a radio advertisement, a major oil company says that one fourth of all cars have at least one underinflated tire. Each car has 4 tires. Suppose that any tire becomes underinflated independently and with the same probability for each tire on every car.

(a) Suppose we select a car at random, and then select a tire on that car at random. What is the probability that the tire in underinflated?

(b) Suppose a parking lot contains 10 cars. We plan on testing the air pressure for every tire on every car. Let the random variable X represent the number of tires (out of the 40) that are underinflated. What is the probability distribution for X?

10

APPROXIMATIONS AND COMPUTATIONS

These numerical approximations are from a book by Abramowitz and Stegun.[1]

Approximation for the Normal Integral

Let Z be a **standard normal random variable**, that is the probability density function for Z is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

Let $Q(z) \equiv P(Z > z)$, that is

$$Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

For $z \geq 0$ the following approximation can be used to compute $Q(z)$:

$$Q(z) = \phi(z)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \epsilon(z)$$

where

$$t = \frac{1}{1 + rz} \qquad r = 0.2316419$$
$$b_1 = 0.31938153 \qquad b_2 = -0.356563782 \qquad b_3 = 1.781477937$$
$$b_4 = -1.821255978 \qquad b_5 = 1.330274429$$

With this approximation, $|\epsilon(z)| < 7.5 \times 10^{-8}$.

[1] Abramowitz, M. and Stegun, I. (1970), *Handbook of Mathematical Functions*, Dover Publications, Inc., New York (1970). This handbook is being published as a reference source on the World Wide Web at http://dlmf.nist.gov/.

Approximation for the Inverse Normal Integral

Let Z be a **standard normal random variable**, that is the probability density function for Z is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

Let $Q = P(Z > z)$, that is

$$Q = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

The problem is, for a given $0 < Q < 0.5$, find the corresponding z that satisfies the above integral equation.

For a given $0 < Q < 0.5$, the following approximation can be used to compute z:

$$z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \epsilon(Q)$$

where

$$t = \sqrt{\ln \frac{1}{Q^2}}$$

$$\begin{array}{lll} c_0 = 2.515517 & c_1 = 0.802853 & c_2 = 0.010328 \\ d_0 = 1.432788 & d_1 = 0.189269 & d_2 = 0.001308 \end{array}$$

With this approximation, $|\epsilon(Q)| < 4.5 \times 10^{-4}$.

Questions for Chapter 10

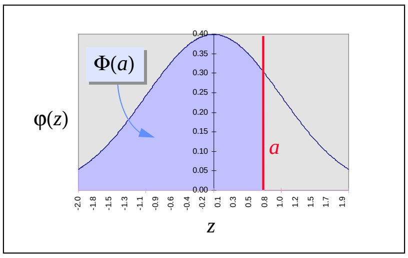
10.1 Write a computer program to generate a table for the cumulative distribution function of the standard normal distribution for $-3.00 \leq z \leq 3.00$ in increments of 0.01.

10.2 Let $Z \sim N(0, 1)$. Write a computer program to find z when $P(Z \leq z) =$ 0.01, 0.05, 0.25, 0.50, 0.95 and 0.99.

Appendix A

TABLES

TABLE I: STANDARD NORMAL CUMULATIVE DISTRIBUTION FUNCTION



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

a	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

TABLE II: SOME COMMON PROBABILITY DISTRIBUTIONS

Name	**Type**	**Parameters**	**Definition**	**Support**	$E(X)$	Var(X)	**mgf**
Bernoulli	disc	$0 \leq p \leq 1$	$p_X(k) = p^k(1-p)^{1-k}$	$k = 0, 1$	p	$p(1-p)$	$e^t p + (1-p)$
binomial	disc	$n; 0 \leq p \leq 1$	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$	$k = 0, 1, \ldots, n$	np	$np(1-p)$	$(e^t p + (1-p))^n$
geometric	disc	$0 \leq p \leq 1$	$p_X(k) = (1-p)^{k-1} p$	$k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Poisson	disc	$\alpha > 0$	$p_X(k) = \frac{e^{-\alpha}\alpha^k}{k!}$	$k = 0, 1, \ldots$	α	α	$e^{\alpha(e^t - 1)}$
uniform	cont	$a < b$	$f_X(x) = \frac{1}{b-a}$	$a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt} - e^{at}}{(b-a)t}$
exponential	cont	$\alpha > 0$	$f_X(x) = \alpha e^{-\alpha x}$	$x \geq 0$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{\alpha}{\alpha - t}$
gamma	cont	$r \geq 1; \alpha > 0$	$f_X(x) = \frac{\alpha}{\Gamma(r)}(\alpha x)^{r-1} e^{-\alpha x}$	$x \geq 0$	$\frac{r}{\alpha}$	$\frac{r}{\alpha^2}$	$\left(\frac{\alpha}{\alpha - t}\right)^r$
normal	cont	$\mu; \sigma > 0$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < \infty$	μ	σ^2	$e^{(t\mu + \sigma^2 t^2/2)}$

Appendix B

ANSWERS TO SELF-TEST EXERCISES

Chapter 1

S1.1	A	S1.2	D	S1.3	C	S1.4	A	S1.5	D
S1.6	D	S1.7	A						

Chapter 2

S2.1	B	S2.2	A	S2.3	B	S2.4	A	S2.5	D
S2.6	E	S2.7	D	S2.8	D	S2.9	C	S2.10	A
S2.11	C	S2.12	B						

Remarks

S2.5 Answer C is acceptable if you allow the triangle to be flipped

S2.6 The correct answer is $\binom{5}{2}$

Chapter 3

S3.1	C	S3.2	C	S3.3	E	S3.4	C	S3.5	C
S3.6	C	S3.7	D	S3.8	B	S3.9	A	S3.10	B
S3.11	E	S3.12	B	S3.13	B	S3.14	A	S3.15	D
S3.16	D	S3.17	A	S3.18	D	S3.19	A	S3.20	A
S3.21	D	S3.22	C	S3.23	A	S3.24	A	S3.25	B
S3.26	A	S3.27	D	S3.28	D	S3.29	C	S3.30	D
S3.31	C	S3.32	B	S3.33	A	S3.34	A	S3.35	B
S3.36	C								

Remarks

S3.5 Must have $p = P(A) = P(B) = P(A \cap B) = 0$. Hence, $P(A \cap B) = P(A)P(B)$.

Chapter 4

S4.1	A	S4.2	B	S4.3	B	S4.4	C	S4.5	C
S4.6	C	S4.7	C	S4.8	B	S4.9	C	S4.10	D
S4.11	B	S4.12	C	S4.13	C	S4.14	B	S4.15	B
S4.16	C	S4.17	D	S4.18	C	S4.19	A		

Chapter 5

S5.1	C	S5.2	A	S5.3	C	S5.4	A	S5.5	C
S5.6	C	S5.7	B	S5.8	C	S5.9	B	S5.10	B
S5.11	B	S5.12	D	S5.13	D	S5.14	B	S5.15	B
S5.16	C	S5.17	C	S5.18	A	S5.19	C	S5.20	B
S5.21	A	S5.22	B	S5.23	E	S5.24	C	S5.25	D
S5.26	D	S5.27	B	S5.28	C				

Remarks

S5.27 Note that $P(\ln(X) < \ln(0.50)) = P(X < 0.50)$ since $\ln(x)$ is an increasing function.

Chapter 6

S6.1	B	S6.2	D	S6.3	E	S6.4	C	S6.5	C
S6.6	B	S6.7	C	S6.8	C	S6.9	C	S6.10	B
S6.11	B	S6.12	C	S6.13	E	S6.14	C	S6.15	A
S6.16	C	S6.17	A	S6.18	D	S6.19	E	S6.20	D
S6.21	D	S6.22	B						

Remarks

S6.19 The correct answer is 1

Chapter 7

S7.1	D	S7.2	C	S7.3	C	S7.4	B	S7.5	B
S7.6	C	S7.7	A	S7.8	C	S7.9	D	S7.10	C
S7.11	A	S7.12	D	S7.13	D	S7.14	B	S7.15	B
S7.16	C	S7.17	D	S7.18	B				

Chapter 8

S8.1	C

Chapter 9

S9.1	D	S9.2	C	S9.3	C

Appendix C

SAMPLE EXAMINATION 1

PART I: MULTIPLE CHOICE QUESTIONS

Instructions: For each of the questions in Part I, choose the best answer among those provided.

In all of the following questions, Ω is a sample space with events A, B and C and probability measure $P(\cdot)$.

1. *(5 points)* Four of your friends are having birthdays on the same date. You purchased four birthday cards. How many different ways can you send cards to your friends so that each gets *one* card?

   (a) $4!$

   (b) $\binom{4}{4}$

   (c) 4

   (d) 4^4

   (e) none of the above.

2. *(5 points) Three* of your friends are having birthdays on the same date. You purchased four birthday cards. How many different ways can you send cards to your friends so that each gets *one* card? (You will have one card left over.)

   (a) $4!$

   (b) $\binom{4}{3}$

   (c) $3!$

   (d) 4^3

   (e) none of the above.

3. *(5 points)* If $A \subseteq B$, $P(A) = 0.7$ and $P(B) = 0.7$, then
   (a) $P(A \cap B) = 0$
   (b) $A = B$
   (c) $A \cap B = \emptyset$
   (d) A and B are independent
   (e) none of the above.

4. *(5 points)* The event which corresponds to the statement, "all of the events A, B and C occur simultaneously," is
   (a) $A \cup B \cup C$
   (b) $A \cap B \cap C$
   (c) $(A \cap B) \cup (A \cap C) \cup (B \cap C)$
   (d) $(A \cup B \cup C)^{\mathsf{c}}$
   (e) none of the above.

5. *(5 points)* If $P(A) = 0.3$, $P(B) = 0.3$ and $P(A \mid B) = 0.3$ then
   (a) A and B are mutually exclusive
   (b) $P(A \cap B) = 0$
   (c) A and B are independent
   (d) all of the above
   (e) none of the above.

6. *(5 points)* If $P(A) = p > 0$, $P(B) = q > 0$ and $A \cap B = \emptyset$ then $P(A \cup B \mid A)$ equals
   (a) 1
   (b) p
   (c) $(p + q)/p$
   (d) $(p + q - pq)/p$
   (e) none of the above

For questions 7 and 8: Assume that X is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} cx & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

---

7. *(5 points)* The value of the constant c is
   (a) 0.5
   (b) 1
   (c) 2
   (d) $\sqrt{2}$
   (e) none of the above.
8. *(5 points)* $P(X > 1)$ equals
   (a) 0.25
   (b) 0.5
   (c) 0.75
   (d) 1
   (e) none of the above.

PART II

Instructions: Show your work in the space provided. **Circle your final answer.**

---

9. After getting your engineering degree, you are now the owner of the most successful restaurant in Buffalo, New York. Your chicken wing recipe is so famous that the White House (in Washington DC) has asked that you send a large order of chicken wings for a important dinner.

   Your chicken wings can be either *hot* or *mild*, and usually you keep the two different types of chicken wings in separate containers. Unfortunately, an engineering intern (who is working for you as a cook), accidentally mixed some hot wings with some mild ones. You now have only three containers of chicken wings:

Container A All hot wings

Container B All mild wings

Container C One half hot and one half mild wings

The White House wants only hot chicken wings, and you need to ship the order immediately. You don't want to accidentally ship **Container B** or **C**.

To make matters worse, the containers are unmarked. The only way you can determine the contents of the containers is by sampling the wings.

(a) *(10 points)* You plan to choose a container at random, and then sample one of the chicken wings from that container. What is the probability that the wing will be mild? *Justify your answer.*

(b) *(10 points)* You actually select a container at random, and sample a chicken wing from that container. It's a hot wing! What is the probability that you chose **Container C** (the one with both hot and mild wings). *Justify your answer using Bayes' theorem.*

(c) *(10 points)* You plan to sample one more wing from the container chosen in part (b). Given that the first wing that you sampled was hot, what is the probability that if you sample a second wing from the same container, it will also be hot? *Justify your answer.*

10. Let X be a random variable with the following probability mass function:

$$\begin{aligned} p_X(-1) &= 0.2 \\ p_X(0) &= 0.4 \\ p_X(+1) &= 0.4 \end{aligned}$$

(a) *(10 points)* Sketch the cumulative distribution function for X.

(b) *(5 points)* Compute $P(X < 0)$.

(c) *(5 points)* Compute $P(X > 1)$.

(d) *(10 points)* Compute $P(X = 0 \mid X \leq 1)$.

Appendix D

SAMPLE EXAMINATION 2

PART I: MULTIPLE CHOICE QUESTIONS

Instructions: For each of the questions in Part I, choose the best answer among those provided.

1. *(5 points)* Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \lambda x^n & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where λ is a constant and n is a nonnegative, integer constant. A relationship which must exist between λ and n is

   (a) $\lambda n = 1$

   (b) $\lambda = n + 1$

   (c) $\lambda = n$

   (d) $\lambda = 2n$

   (e) none of the above.

2. *(5 points)* Let X be a discrete random variable with a probability distribution given by

$$P(X = k) = \left(\tfrac{1}{4}\right)^k \left(1 - \tfrac{1}{4}\right) \qquad \text{for } k = 0, 1, 2, \ldots.$$

Then $P(X \ge k)$ equals

   (a) $\left(\frac{1}{4}\right)^k$

   (b) $\left(\frac{1}{4}\right)^{k+1}$

(c) $\frac{3}{4}$

(d) $\left(\frac{3}{4}\right)^k$

(e) none of the above.

3. *(5 points)* Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} 4e^{-4x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The value of $P(X^2 \geq 16)$ is

(a) $1 - e^{-4}$

(b) e^{-16}

(c) $1 - 4e^{-16}$

(d) e^{-4}

(e) none of the above.

4. *(5 points)* Let X and Y be random variables with $E(X) = 2$ and $E(Y) = 2$. Then, $E[(X - E(Y))^2]$ equals

(a) 0

(b) Var(X)

(c) Var(Y)

(d) Cov(X, Y)

(e) none of the above.

5. *(5 points)* If X is a random variable with $P(X \leq 0) = \alpha$, then $P(-3X < 0)$ equals

(a) 0

(b) $\frac{1}{3}\alpha$

(c) α

(d) $1 - \alpha$

(e) none of the above.

6. *(5 points)* Let X be a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \dfrac{x+1}{2} & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

Then $P(X < 0)$ equals

(a) 0

(b) $\frac{1}{8}$

(c) $\frac{1}{4}$

(d) $\frac{1}{2}$

(e) none of the above

For questions 7 and 8: Assume that (X, Y) is a random vector with joint probability mass function given by

$$\begin{aligned} p_{X,Y}(-1,-1) &= 1/3 \\ p_{X,Y}(0,0) &= 1/3 \\ p_{X,Y}(1,1) &= 1/3 \end{aligned}$$

---

7. *(5 points)* Define the random variable $W = XY$. The value of $P(W = 1)$ is

(a) 0

(b) $1/9$

(c) $1/3$

(d) $2/3$

(e) none of the above.

8. *(5 points)* The value of $E(X + Y)$ is

(a) -1

(b) 0

(c) $1/3$

(d) 1

(e) none of the above.

PART II

Instructions: Show your work in the space provided. **Circle your final answer.**

9. The Youbee School of Engineering has a graduating class of 2 men and 3 women. Plans are being made for a graduation party. Let X denote the number of men from the graduating class who will attend the party, and let Y denote the number of women from the class who will attend.

   After an extensive survey was conducted, it has been determined that the random vector (X, Y) has the following joint probability mass function:

$p_{X,Y}(x, y)$		Y (women) 0	1	2	3
	0	.01	.05	.03	.01
X (men)	1	.05	.25	.15	.05
	2	.04	.20	.12	.04

   (a) *(5 points)* What is the probability that 4 *or more* members of the graduating class will show up at the party?

   (b) *(5 points)* What is the probability that more women than men from the graduating class will show up at the party?

   (c) *(10 points)* If women are charged \$1 for attending the party and men are charged \$2, what is the expected amount of money that will be collected from members of the graduating class?

   (d) *(10 points)* All three women graduates arrive early at the party. What is the probability that no men will show up?

10. Let (X, Y) be a continuous random vector with joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{if } x \geq 0,\ y \geq 0 \text{ and } x + y \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

   (a) *(10 points)* Compute $P\left(X + Y < \frac{1}{2}\right)$.

   (b) *(10 points)* Find the marginal probability density function for X.

   (c) *(10 points)* Find the conditional probability density function for Y given that $X = \frac{1}{2}$.

Appendix E

SAMPLE FINAL EXAMINATION

PART I: MULTIPLE CHOICE QUESTIONS

Instructions: For each of the questions in Part I, choose the best answer among those provided.

In all of the following questions, Ω is a sample space with events A, B and C and probability measure $P(\cdot)$.

1. *(3 points)* Let X and Y be independent random variables with $X \sim N(0,1)$ and $Y \sim N(0,2)$. The value of $P(X > Y)$ is
   (a) 0
   (b) 0.05
   (c) 0.50
   (d) 0.95
   (e) none of the above.

2. *(3 points)* Let $P(A) = r$ and $P(B) = s$ with $A \cap B = \varnothing$. Then $P(A \mid B^{\mathsf{c}})$ equals
   (a) 0
   (b) r
   (c) $r/(1-s)$
   (d) 1
   (e) none of the above.

For questions 3 and 4: A game is played by repeatedly tossing a fair coin. If the coin turns up heads, you win $1, and if the coin turns up tails, you lose $1.

Let Y_n denote your net winnings (i.e., your winnings minus your losses) after n tosses. Note that Y_n may be negative.

---

3. *(3 points)* The value of $E(Y_n)$ is
   (a) 0
   (b) $-n$
   (c) $n/2$
   (d) n
   (e) none of the above.
4. *(3 points)* The value of Var(Y_n) is
   (a) 0
   (b) $\sqrt{n}$
   (c) n
   (d) n^2
   (e) none of the above.
5. *(3 points)* In a class of 10 students, 5 students received an A on the last examination. Two students are chosen at random from the class. The probability that both of the selected students received a grade of A is
   (a) $1/4$
   (b) $2/9$
   (c) $1/3$
   (d) $1/2$
   (e) none of the above.
6. *(3 points)* Let X be a continuous random variable with probability density function
$$f_X(x) = \begin{cases} \alpha e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$
   The value of α
   (a) must be 1

(b) can be any positive real number

(c) can be any positive integer

(d) can be any real number

(e) none of the above.

7. *(3 points)* The set $(A \cup B) \cap (A \cup B^{\mathsf{c}})$ is equivalent to

   (a) Ø

   (b) Ω

   (c) A

   (d) B

   (e) none of the above.

8. *(3 points)* Suppose a chain link has a probability of 0.001 of breaking, independently of all other links. Approximately, what is the probability that a chain with 2000 such links will break. (*Note:* The chain will break if one *or more* links break.)

   (a) e^{-2}

   (b) $1 - e^{-2}$

   (c) $1/2$

   (d) $(0.001)^{2000}$

   (e) none of the above.

9. *(3 points)* The weight of a gum drop (in ounces) is normally distributed with mean 2 and standard deviation $\frac{1}{4}$. A bag contains 10 independent gum drops. The probability that the total weight of the gum drops in the bag exceeds 20 ounces is

   (a) 0

   (b) 0.5

   (c) 1

   (d) $1 - (\frac{1}{4})^{16}$

   (e) none of the above.

10. *(3 points)* Let X be a random variable with probability generating function

$$g(z) = \frac{z(1+z)}{2}$$

The value of $E(X)$ is

(a) 0

(b) $2/3$

(c) 1

(d) $3/2$

(e) none of the above.

11. *(3 points)* Let $X \sim N(1, 8)$. If $Y = X^3$, then $P(Y > 1)$ equals

(a) 0.1587

(b) 0.5000

(c) 0.8413

(d) 1.0000

(e) none of the above.

12. *(3 points)* Let X be a random variable with moment generating function

$$M_X(t) = e^{(e^t - 1)}$$

The moment generating function of the random variable $Y = 2X + 1$ is

(a) e^t

(b) e^{2t}

(c) $e^{(2e^t - 2)}$

(d) $e^{(t + e^{2t} - 1)}$

(e) none of the above.

13. *(3 points)* Let X be a random variable with $E(X) = 2$ and $E(X^2) = 5$. Then Var(X) equals

(a) 1

(b) 3

(c) 7

(d) 9

(e) none of the above.

14. *(3 points)* The weight X (in pounds) of a roll of steel is a random variable with probability density function given by

$$f_X(x) = \begin{cases} \frac{1}{2}(x-499) & \text{if } 499 \leq x \leq 501 \\ 0 & \text{otherwise} \end{cases}$$

A roll is considered defective if its weight is less than 500 pounds. If 3 independent rolls of steel are examined, the probability that *exactly one* is defective is

(a) 1/4

(b) 3/8

(c) 9/64

(d) 27/64

(e) none of the above.

15. *(3 points)* An airplane can hold 100 passengers. It is know that, over many years, 1% of all persons making airline reservations never show up for their flight. Assume that a person decides independently whether or not to show up for a flight. If the airline permits 101 persons to make reservations for a given flight, the probability that someone will be denied a seat is

(a) 0

(b) 0.01

(c) $(0.99)^{101}$

(d) 1

(e) none of the above.

For questions 16 and 17: Suppose that X_1 and X_2 are independent random variables representing the lifetimes of two lightbulbs, each having marginal probability density function

$$f(x) = \begin{cases} 1/2 & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $W = X_1 + X_2$ represent the total lifetime of the two lightbulbs.

---

16. *(3 points)* The value of $E(W)$ is
    (a) $1/2$
    (b) 1
    (c) 2
    (d) $X_1 + X_2$
    (e) none of the above.

17. *(3 points)* The value of $P(W < 2)$ is
    (a) 0
    (b) $1/4$
    (c) $1/2$
    (d) 1
    (e) none of the above.

18. *(3 points)* Suppose $A \subseteq B$. Which of the following statements are *always* true?
    (a) $P(A) \leq P(B)$
    (b) $P(A \cup B) = P(B)$
    (c) $P(A \cap B) = P(A)$
    (d) all of the above
    (e) none of the above.

19. *(3 points)* Let $X_1, \ldots, X_n$ be a random sample with each X_i having a probability mass function given by

$$\begin{aligned} P(X_i = 0) &= 1 - p \\ P(X_i = 1) &= p \end{aligned}$$

where $0 \le p \le 1$. An unbiased estimator for p is given by

(a) $\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

(b) $\hat{p} = \sum_{i=1}^{n} X_i$

(c) $\hat{p} = \sum_{i=1}^{n} X_i^2$

(d) all of the above

(e) none of the above.

20. *(3 points)* A box is partitioned into 3 numbered cells. Two indistinguishable balls are dropped into the box. Any cell may contain more than one ball. The total number of distinguishable arrangements of the balls in the numbered cells is

(a) $\binom{4}{2}$

(b) $\binom{3}{2}$

(c) $\binom{3}{1}$

(d) 4

(e) none of the above.

PART II

Instructions: Show your work in the space provided. **Circle your final answer.**

21. A sociologist is interested in determining the percentage of college students who like classical music. In order to avoid embarrassing any students and help assure a truthful response to the survey, she decides on the following technique: A person is given a bowl which contains 600 red beads and 400 blue beads. The person is instructed to go to a secluded corner, draw out a bead at random, and write down the answer to one of the following questions depending on the color of the bead:

- **If the bead is red,** answer "Yes" or "No" to the question:

Do you like classical music?

- **If the bead is <u>blue</u>,** answer "Yes" or "No" to the question:

  Does your student number end in an odd digit?

The person simply responds "Yes" or "No." No one, except the person being surveyed, knows which question was actually answered. Therefore, the person is likely to be truthful.

(a) *(5 points)* Suppose p is the unknown fraction of students who like classical music. What is the unconditional probability that a person surveyed will answer "Yes?" Your answer should depend on p.

(b) *(5 points)* Let X denote the number of "Yes" responses from a survey of 100 individuals. Determine $P(X = k)$ as a function of k and p.

(c) *(5 points)* Let $Y = (X/60) - (1/3)$. Show that $E(Y) = p$.

(d) *(5 points)* Suppose 40 of the 100 students answered "Yes" to the survey. What is your best guess for the value of p? (*Hint:* Use an unbiased estimator for p.)

22. Let (X, Y) be a continuous random vector with joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{if } x \geq 0, y \geq 0 \text{ and } x + y \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

(a) *(5 points)* Compute $P\Big(X + Y < \frac{1}{2}\Big)$.

(b) *(5 points)* Find the marginal probability density function for X.

(c) *(5 points)* Find the marginal probability density function for Y.

(d) *(5 points)* Find the conditional probability density function for Y given that $X = \frac{1}{2}$.