Math 382 Lecture Notes
Probability and Statistics

Anwar Hossain and Oleg Makhnin

August 21, 2015

Contents

Chapter 1

Probability in the World Around Us

Probability theory is a tool to describe uncertainty. In science and engineering, the world around us is described by mathematical models. Most mathematical models are *deterministic*, that is, the model output is supposed to be known uniquely once all the inputs are specified. As an example of such model, consider the Newton's law $F = ma$ connecting the force F acting on an object of mass m resulting in the acceleration a. Once F and m are specified, we can determine exactly the object's acceleration.[1]

What is wrong with this model from practical point of view? Most obviously, the inputs in the model (F and m) are not precisely known. They may be measured, but there's usually a measurement error involved. Also, the model itself might be approximate or might not take into account all the factors influencing the model output. Finally, roundoff errors are sure to crop up during the calculations. Thus, our predictions of planetary motions, say, will be imperfect in the long run and will require further corrections as more recent observations become available.

At the other end of the spectrum, there are some phenomena that seem to completely escape any attempts at the rational description. These are random phenomena – ranging from lotteries to the heat-induced motion of the atoms. Upon closer consideration, there are still some laws governing these phenomena. However, they would not apply on case by case basis, but rather to the results of many repetitions. For example, we cannot predict the result of one particular lottery drawing, but we can calculate probabilities of certain outcomes. We cannot describe the velocity of a single atom, but we can say something about the behavior of the velocities in the ensemble of all atoms.

This is the stuff that *probabilistic* models are made of. Another example of a field where probabilistic models are routinely used is actuarial science. It deals with lifetimes of humans and tries to predict how long any given person is expected to live, based on other variables describing the particulars of his/her life. Of course, this expected life span is a poor prediction when applied to any given person, but it works rather well when applied to many persons. It can help to decide the rates the insurance company should charge for covering any given person.

Today's science deals with enormously complex models, for example, the models of Earth's climate (there are many of them available, at different levels of complexity and

[1] Now you are to stop and think: what are the factors that will make this model more uncertain?

resolution). The models should also take into account the uncertainties from many sources, including our imperfect knowledge of the current state of Earth, our imperfect understanding of all physical processes involved, and the uncertainty about future scenarios of human development.[2]

Understanding and communicating this uncertainty is greatly aided by the knowledge of the rules of probability.

The authors thank Lynda Ballou for contributing some examples and exercises, and Brian Borchers for valuable comments.

[2]Not the least, our ability to calculate the output of such models is also limited by the current state of computational science.

Chapter 2

Probability

2.1 What is Probability

Probability theory is the branch of mathematics that studies the possible outcomes of given events together with the outcomes' relative likelihoods and distributions. In common usage, the word "probability" is used to mean the chance that a particular event (or set of events) will occur expressed on a linear scale from 0 (impossibility) to 1 (certainty), also expressed as a percentage between 0 and 100%. The analysis of data (possibly generated by probability models) is called **statistics**.

Probability is a way of summarizing the uncertainty of statements or events. It gives a numerical measure for the degree of certainty (or degree of uncertainty) of the occurrence of an event.

Another way to define probability is the ratio of the number of favorable outcomes to the total number of all possible outcomes. This is true if the outcomes are assumed to be equally likely. The collection of all possible outcomes is called the **sample space**.

If there are n total possible outcomes in a sample space $\mathcal{S}$, and m of those are favorable for an event A, then probability of event A is given as

$$P(A) = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}} = \frac{n(A)}{n(S)} = \frac{m}{n}$$

Example 2.1. *Find the probability of getting a 3 or 5 while throwing a die.*

Solution. Sample space $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ and event $A = \{3, 5\}$.
We have $n(A) = 2$ and $n(S) = 6$.
So, $P(A) = n(A)/n(S) = 2/6 = 0.3333$ □

Axioms of Probability

All probability values are positive numbers not greater than 1, i.e. $0 \leq p \leq 1$. An event that is not likely to occur or impossible has probability zero, while an event that's certain to occur has probability one.

Examples: $P(\text{A pregnant human being a female}) = 1$
$P(\text{A human male being pregnant}) = 0$.

Definition 2.1.

Random Experiment: A random experiment is the process of observing the outcome of a chance event.

Outcome: The elementary outcomes are all possible results of the random experiment.

Sample Space(SS): The sample space is the set or collection of all the outcomes of an experiment and is denoted by $\mathcal{S}$.

Example 2.2.
a) Flip a coin once, then the sample space is: $\mathcal{S} = \{H, T\}$
b) Flip a coin twice, then the sample space is: $\mathcal{S} = \{HH, HT, TH, TT\}$

We want to assign a numerical weight or **probability** to each outcome. We write the probability of A_i as $P(A_i)$. For example, in our coin toss experiment, we may assign $P(H) = P(T) = 0.5$. Each outcome comes up half the time.

2.2 Review of set notation

Definition 2.2. Complement

The complement of event A is the set of all outcomes in a sample that are not included in the event A. The complement of event A is denoted by A'.

If the probability that an event occurs is p, then the probability that the event does not occur is $q = (1 - p)$. i.e. probability of the complement of an event $= 1-$ probability of the event.

$$\text{i.e. } P(A') = 1 - P(A)$$

Example 2.3. *Find the probability of not getting a 3 or 5 while throwing a die.*

Solution. Sample space $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ and event $B = \{1, 2, 4, 6\}$.

$$n(B) = 4 \text{ and } n(S) = 6$$

$$\text{So, } P(B) = n(B)/n(S) = 4/6 = 0.6667$$

On the other hand, A (described in Example 2.1) and B are complementary events, i.e. $B = A'$.

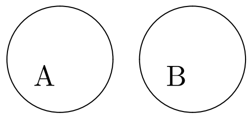
So, $P(B) = P(A') = 1 - P(A) = 1 - 0.3333 = 0.6667$ □

Definition 2.3. Intersections of Events

The event $A \cap B$ is the **intersection** of the events A and B and consists of outcomes that are contained within both events A and B. The probability of this event, is the probability that both events A and B occur [but not necessarily at the same time]. In the future, we will abbreviate intersection as AB.

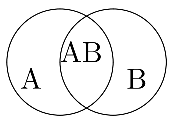
Definition 2.4. Mutually Exclusive Events

Two events are said to be mutually exclusive if $AB = \emptyset$ (i.e. they have empty intersection) so that they have no outcomes in common.



Definition 2.5. Unions of Events

The event $A \cup B$ is the **union** of events A and B and consists of the outcomes that are contained within at least one of the events A and B. The probability of this event $P(A \cup B)$, is the probability that at least one of the events A and B occurs.



Venn diagram

Venn diagram is often used to illustrate the relations between sets (events). The sets A and B are represented as circles; operations between them (intersections, unions and complements) can also be represented as parts of the diagram. The entire sample space $\mathcal{S}$ is the bounding box. See Figure 2.1

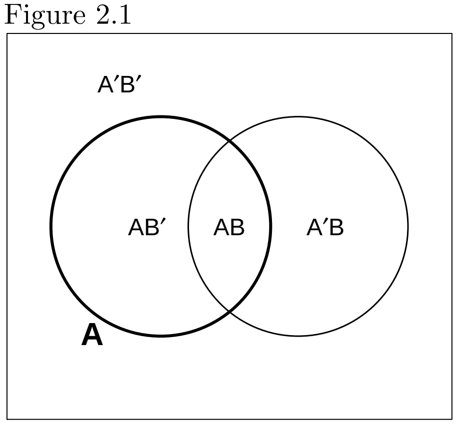


Figure 2.1: Venn diagram of events A (in bold) and B, represented as insides of circles, and various intersections

Example 2.4. Set notation

Suppose a set $\mathcal{S}$ consists of points labeled 1, 2, 3 and 4. We denote this by $\mathcal{S} = \{1, 2, 3, 4\}$.
If $A = \{1, 2\}$ and $B = \{2, 3, 4\}$, then A and B are subsets of $\mathcal{S}$, denoted by $A \subset \mathcal{S}$ and $B \subset \mathcal{S}$ (B is contained in $\mathcal{S}$). We denote the fact that 2 is an element of A by $2 \in A$.

The union of A and B, $A \cup B = \{1, 2, 3, 4\}$. If $C = \{4\}$, then $A \cup C = \{1, 2, 4\}$. The intersection $A \cap B = AB = \{2\}$. The complement $A' = \{3, 4\}$. $\square$

Distributive laws

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

and

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

De Morgan's Law

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

$$(AB)' = A' \cup B'$$

Exercises

2.1.

Use the Venn diagrams to illustrate Distributive laws and De Morgan's law.

2.2.

Simplify the following (Draw the Venn diagrams to visualize)

a) $(A')'$

b) $(AB)' \cup A$

c) $(AB) \cup (AB')$

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

2.3.

Represent by set notation the following events

a) both A and B occur

b) exactly one of A, B occurs

c) at least one of A, B, C occurs

d) at most one of A, B, C occurs

2.4.

The sample space consists of eight capital letters (outcomes), A, B, C ,..., H. Let V = event that the letter represents a vowel, and L = event that the letter is made of straight lines. Describe the outcomes that comprise

a) VL

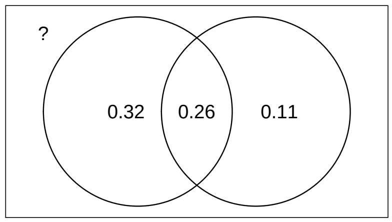
b) $V \cup L'$

c) $V'L'$

Ways to represent probabilities:

- **Venn diagram**

  We may write the probabilities inside the elementary pieces within a Venn diagram. For example, $P(AB') = 0.32$ and
  $P(A) = P(AB) + P(AB')$ [why?] $= 0.58$ The relative sizes of the pieces do not have to match the numbers.



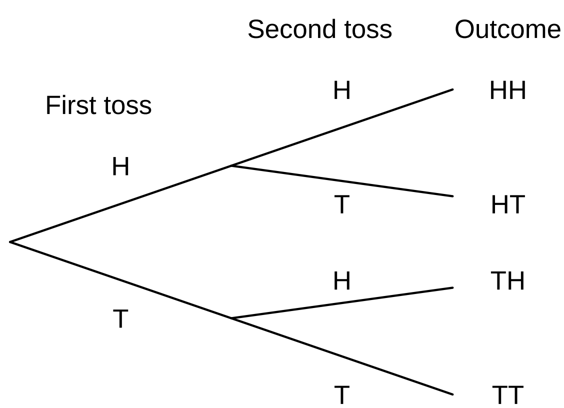
- **Two-way table**

  This is a popular way to represent statistical data. The cells of the table correspond to the intersections of row and column events. Note that the contents of the table add up accross rows and columns of the table. The bottom-right corner of the table contains $P(\mathcal{S}) = 1$

	B	B'	
A	0.26	0.32	0.58
A'	0.11	?	0.42
	0.37	0.63	1

- **Tree diagram**

  A tree diagram may be used to show the sequence of choices that lead to the complete description of outcomes. For example, when tossing two coins, we may represent this as follows



  A tree diagram is also often useful for representing conditional probabilities (see below).

2.5.
Out of all items sent for refurbishing, 40% had mechanical defects, 50% had electrical defects, and 25% had both.
Denoting $A = \{\text{an item has a mechanical defect}\}$ and
$B = \{\text{an item has an electrical defect}\}$, fill the probabilities into the Venn diagram and determine the quantities listed below.

a) $P(A)$

b) $P(AB)$

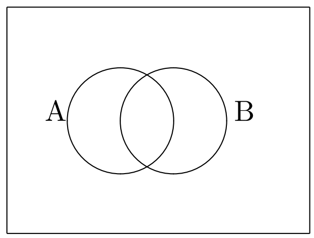
c) $P(A'B)$

d) $P(A'B')$

e) $P(A \cup B)$

f) $P(A' \cup B')$

g) $P([A \cup B]')$



2.6.
Do the following satisfy the definitions of probability? If not, explain why.

a) $P(A) = 0.3$, $P(B) = 0.5$ and $P(AB') = 0.4$.

b) $P(A) = 0.4$, $P(B) = 0.6$ and $P(AB) = 0.2$.

c) $P(A) = 0.7$, $P(B) = 0.6$ and $P(AB) = 0.2$.

2.7.
For tossing a six-sided die, find the following probabilities (assume equally likely outcomes).

a) Probability to get both a number more than 3 **and** an even number.

b) Probability to get a number less than 4 **or** an odd number.

2.8.
When tossing two six-sided dice, find the following probabilities (assume equally likely outcomes).

a) Probability that the first die shows a number more than 3 **and** the second one shows an even number.

b) Probability that the first die shows a number less than 4 **or** the second one shows an odd number.

2.9.

a) Suppose that $P(A \cup B) = 0.8$ and $P(A' \cup B) = 0.7$. Find $P(B)$. [**Hint:** Fill out either a Venn diagram or a two-way table.]

b) Suppose that $P(A) = 0.4$ and $P(B) = 0.3$. What are the possible values for $P(A'B')$?

2.10.
A sample of mutual funds was classified according to whether a fund was up or down last year (A and A') and whether it was investing in international stocks (B and B'). The probabilities of these events and their intersections are represented in the two-way table below.

	B	B'	
A	0.33	?	?
A'	?	?	0.52
	0.64	?	1

a) Fill out all the ? marks.

b) Find the probability of $A \cup B$

2.3 Types of Probability

There are three ways to define probability, namely classical, empirical and subjective probability.

Definition 2.6. Classical probability

Classical or *theoretical* probability is used when each outcome in a sample space is equally likely to occur. The classical probability for an event A is given by

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

Example 2.5.

Roll a die and observe that $P(A) = P(\text{rolling a 3}) = 1/6$.

Definition 2.7. Empirical probability

Empirical (or *statistical*) probability is based on observed data. The empirical probability of an event A is the *relative frequency* of event A, that is

$$P(A) = \frac{\text{Frequency of event A}}{\text{Total number of observations}}$$

Example 2.6.
The following are the counts of fish of each type, that you have caught before.

Fish Types	Blue gill	Red gill	Crappy	Total
Number of times caught	13	17	10	40

Estimate the probability that the next fish you catch will be a Blue gill.

$$P(\text{Blue gill}) = 13/40 = 0.325 \quad \square$$

Example 2.7.
Based on genetics, the proportion of male children among all children conceived should be around 0.5. However, based on the statistics from a large number of live births, the probability that a child being born is male is about 0.512. □

The empirical probability definition has a weakness that it depends on the results of a particular experiment. The next time this experiment is repeated, you are likely to get a somewhat different result.

However, as an experiment is repeated many times, the empirical probability of an event, based on the combined results, approaches the theoretical probability of the event.[1]

Subjective Probability: Subjective probabilities result from intuition, educated guesses, and estimates. For example, given a patient's health and extent of injuries a doctor may feel that the patient has a 90% chance of a full recovery.

Regardless of the way probabilities are defined, they always follow the same laws, which we will explore starting with the following Section.

2.4 Laws of Probability

As we have seen in the previous section, the probabilities are not always based on the assumption of equal outcomes.

Definition 2.8. Axioms of Probability

For an experiment with a sample space $\mathcal{S} = \{e_1, e_2, \ldots, e_n\}$ we can assign probabilities $P(e_1), P(e_2), \ldots, P(e_n)$ provided that

a) $0 \leq P(e_i) \leq 1$

b) $P(\mathcal{S}) = \sum_{i=1}^{n} P(e_i) = 1.$

If a set (event) A consists of outcomes $\{e_1, e_2, \ldots, e_k\}$, then

$$P(A) = \sum_{i=1}^{k} P(e_i)$$

This definition just tells us which probability assignments are legal, but not necessarily which ones would work in practice. However, once we have assigned the probability to each outcome, they are subject to further rules which we will describe below.

Theorem 2.1. Complement Rule

For any event A,

$$P(A') = 1 - P(A) \tag{2.1}$$

[1] This is called *Law of Large Numbers*

Theorem 2.2. Addition Law

If A and B are two different events then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.2}$$

Proof. Consider the Venn diagram. $P(A \cup B)$ is the probability of the sum of all sample points in $A \cup B$. Now $P(A) + P(B)$ is the sum of probabilities of sample points in A and in B. Since we added up the sample points in $(A \cap B)$ twice, we need to subtract once to obtain the sum of probabilities in $(A \cup B)$, which is $P(A \cup B)$. □

Example 2.8. *Probability that John passes a Math exam is 4/5 and that he passes a Chemistry exam is 5/6. If the probability that he passes both exams is 3/4, find the probability that he will pass at least one exam.*

Solution. Let M = John passes Math exam, and C = John passes Chemistry exam.

$$P(\text{John passes at least one exam}) = P(M \cup C) =$$

$$= P(M) + P(C) - P(M \cap C) = 4/5 + 5/6 - 3/4 = 53/60$$

□

Corollary. If two events A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B).$$

This follows immediately from (2.2). Since A and B are mutually exclusive, $P(A \cap B) = 0$.

Example 2.9. *What is the probability of getting a total of 7 or 11, when two dice are rolled?*

	1	2	3	4	5	6
1	(1,1)	(1,2)				(1,6)
2						
3						
4						
5						
6						(6,6)

Solution. Let A be the event that the total is 7 and B be the event that it is 11. The sample space for this experiment is

$$\mathcal{S} = \{(1,1),(1,2),......,(2,1),(2,2),.........,(6,6)\}, \qquad n(\mathcal{S}) = 36$$

$$A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \text{ and } n(A) = 6.$$

So, $P(A) = 6/36 = 1/6$.

$$B = \{(5,6),(6,5)\} \text{ and } n(B) = 2$$

So, $P(B) = 2/36 = 1/18$.
Since we cannot have a total equal to both 7 and 11, A and B are mutually exclusive, i.e. $P(A \cap B) = 0$.
So, we have $P(A \cup B) = P(A) + P(B) = 1/6 + 1/18 = 2/9$. □

Exercises

2.11.
Two cards are drawn from a 52-card deck, without replacement. What is the probability that both are greater than 2 and less than 8?

2.12.
A permutation of the word "white" is chosen at random. Find the probability that it begins with a vowel. Also, find the probability that it ends with a consonant, and the probability that it begins with a vowel *and* ends with a consonant.

2.13.
Find the probability that a leap year will have 53 Sundays.

2.14.
As a foreign language, 40% of the students took Spanish and 30% took French, while 60% took at least one of these languages. What percent of students took both Spanish and French?

2.15.
In a class of 100 students, 30 major in Mathematics. Moreover, of the 40 females in the class, 10 major in Mathematics. If a student is selected at random from the class, what is the probability that the student will be a male or will major in Mathematics (or both)?

2.16.
Suppose that $P(A) = 0.4$, $P(B) = 0.5$ and $P(AB) = 0.2$. Find the following:

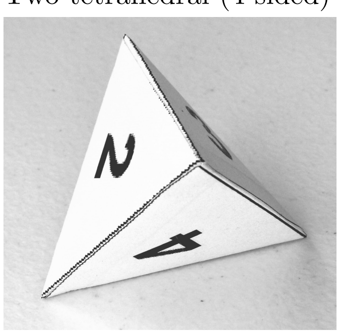
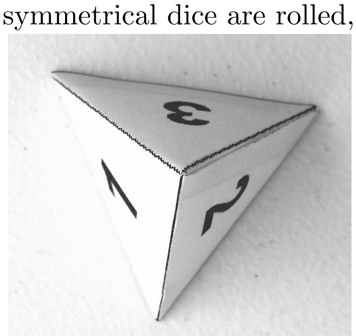
a) $P(A \cup B)$

b) $P(A'B)$

c) $P[A'(A \cup B)]$

d) $P[A \cup (A'B)]$

2.17.
Two tetrahedral (4-sided) symmetrical dice are rolled, one after the other.

a) Find the probability that both dice will land on the same number.

b) Find the probability that each die will land on a number less than 3.

c) Find the probability that the two numbers will differ by at most 1.

d) Will the answers change if we rolled the dice simultaneously?

2.5 Counting Rules useful in Probability

In some experiments it is helpful to list the elements of the sample space systematically by means of a tree diagram, see page 13.

In many cases, we shall be able to solve a probability problem by counting the number of points in the sample space without actually listing each element.

Theorem 2.3. Multiplication principle

If one operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.

Example 2.10. *How large is the sample space when a pair of dice is thrown?*

Solution. The first die can be thrown in $n_1 = 6$ ways and the second in $n_2 = 6$ ways. Therefore, the pair of dice can land in $n_1 n_2 = 36$ possible ways. □

Theorem 2.3 can naturally be extended to more than two operations: if we have n_1, n_2,...,n_k consequent choices, then the total number of ways is $n_1 n_2 \cdots n_k$.

The term *permutations* refers to an arrangement of objects when the order matters (for example, letters in a word).

Theorem 2.4. Permutations

The number of permutations of n distinct objects taken r at a time is

$$_nP_r = \frac{n!}{(n-r)!}$$

Example 2.11.

From among ten employees, three are to be selected to travel to three out-of-town plants A, B, and C, one to each plant. Since the plants are located in different cities, the order in which the employees are assigned to the plants is an important consideration. In how many ways can the assignments be made?

Solution. Because order is important, the number of possible distinct assignments is

$$_{10}P_3 = \frac{10!}{7!} = 10(9)(8) = 720.$$

In other words, there are ten choices for plant A, but then only nine for plant B, and eight for plant C. This gives a total of 10(9)(8) ways of assigning employees to the plants. □

The term *combination* refers to the arrangement of objects when order does not matter. For example, choosing 4 books to buy at the store in any order will leave you with the same set of books.

Theorem 2.5. Combinations

The number of distinct subsets or combinations of size r that can be selected from n distinct objects, $(r \leq n)$, is given by

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} \tag{2.3}$$

Proof. Start with picking ordered sets of size r. This can be done in ${}_nP_r = \frac{n!}{(n-r)!}$ ways. However, many of these are the re-orderings of the same basic set of objects. Each distinct set of r objects can be re-ordered in ${}_rP_r = r!$ ways. Therefore, we need to divide the number of permutations ${}_nP_r$ by $r!$, thus arriving at the equation (2.3). □

Example 2.12.
In the previous example, suppose that three employees are to be selected from among the ten available to go to the same plant. In how many ways can this selection be made?

Solution. Here, order is not important; we want to know how many subsets of size $r = 3$ can be selected from $n = 10$ people. The result is

$$\binom{10}{3} = \frac{10!}{3!\,7!} = \frac{10(9)(8)}{3(2)(1)} = 120$$

□

Example 2.13.
A package of six light bulbs contains 2 defective bulbs. If three bulbs are selected for use, find the probability that none of the three is defective.

Solution. P(none are defective) =

$$= \frac{\text{number of ways 3 nondefectives can be chosen}}{\text{total number of ways a sample of 3 can be chosen}} = \frac{\binom{4}{3}}{\binom{6}{3}} = \frac{1}{5}$$

□

Example 2.14.
In a poker hand consisting of 5 cards, find the probability of holding 2 aces and 3 jacks.

Solution. The number of ways of being dealt 2 aces from 4 is $\binom{4}{2} = 6$ and the number of ways of being dealt 3 jacks from 4 is $\binom{4}{3} = 4$.
The total number of 5-card poker hands, all of which are equally likely is

$$\binom{52}{5} = 2,598,960$$

Hence, the probability of getting 2 aces and 3 jacks in a 5-card poker hand is $P(C) = (6 * 4)/2,598,960$ □

Example 2.15.

A university warehouse has received a shipment of 25 printers, of which 10 are laser printers and 15 are inkjet models. If 6 of these 25 are selected at random to be checked by a particular technician, what is the probability that exactly 3 of these selected are laser printers? At least 3 inkjet printers?

Solution. First choose 3 of the 15 inkjet and then 3 of the 10 laser printers. There are $\binom{15}{3}$ and $\binom{10}{3}$ ways to do it, and therefore

$$P(\text{exactly 3 of the 6}) = \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}} = 0.3083$$

(b) P(at least 3)

$$= \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}} + \frac{\binom{15}{4}\binom{10}{2}}{\binom{25}{6}} + \frac{\binom{15}{5}\binom{10}{1}}{\binom{25}{6}} + \frac{\binom{15}{6}\binom{10}{0}}{\binom{25}{6}} = 0.8530$$

□

Theorem 2.6. Partitions

The number of ways of partitioning n distinct objects into k groups containing $n_1, n_2, \ldots, n_k$ objects respectively, is

$$\frac{n!}{n_1!\, n_2! \ldots n_k!}$$

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

Note that when there are $k = 2$ groups, we will obtain combinations.

Example 2.16.

Consider 10 engineers to be split into 3 groups to be assigned to 3 plants. If we are to send 5 people to Plant A, 3 people to Plant B, and 2 people to Plant C, then the total number of assignments is

$$\frac{10!}{5!\,3!\,2!} = 2520 \quad \square$$

Exercises

2.18.

An incoming lot of silicon wafers is to be inspected for defectives by an engineer in a microchip manufacturing plant. Suppose that, in a tray containing 20 wafers, 4 are defective. Two wafers are to be selected randomly for inspection. Find the probability that neither is defective.

2.19.

A person draws 5 cards from a shuffled pack of 52 cards. Find the probability that the person has at least 3 aces. Find the probability that the person has at least 4 cards of the same suit.

2.20.
Three people enter the elevator on the basement level. The building has 7 floors. Find the probability that all three get off at different floors.

2.21.
In a group of 7 people, each person shakes hands with every other person. How many handshakes did occur?

2.22.
In a lottery, 6 numbers are drawn out of 45. You hit a jackpot if you guess all 6 numbers correctly, and get $400 if you guess 5 numbers out of 6 correctly. What are the probabilities of each of those events?

2.23.
A marketing director considers that there's "overwhelming agreement" in a 5-member focus group when either 4 or 5 people like or dislike the product.[a] If, in fact, the product's popularity is 50% (so that all outcomes are equally likely), what is the probability that the focus group will be in "overwhelming agreement" about it? Is the marketing director making a judgement error in declaring such agreement "overwhelming"?

2.24.
A die is tossed 5 times. Find the probability that we will have 4 of a kind.

2.25.
There are 21 Bachelor of Science programs at New Mexico Tech. Given 21 areas from which to choose, in how many ways can a student select:

a) A major area and a minor area?

b) A major area and two minors (regardless of order)?

2.26.
In a math modeling class, we have 15 students and want to split them into 3 groups, 5 students each, to do group projects. How many possible group assignments are there?

2.27.
If a group consist of 8 men and 6 women, in how many ways can a committee of 5 be selected if:

a) The committee is to consist of 3 men and 2 women.

b) There are no restrictions on the number of men and women on the committee.

c) There must at least one man.

d) There must be at least one of each sex.

2.28.
From a box containing 5 chocolates and 4 hard candies, a child takes a handful of 4 (at random). What is the probability that exactly 3 of the 4 are chocolates?

2.29.
Suppose we have a lot of 40 transistors of which 8 are defective. If we sample without replacement, what is the probability that we get 4 good transistors in the first 5 draws?

Pascal's triangle and binomial coefficients

Long before Pascal, this triangle has been described by several Oriental scholars. It was used in the budding discipline of probability theory by the French mathematician Blaise Pascal (1623-1662). The construction begins by writing 1's along the sides of a triangle and then filling it up row by row so that each number is a sum of the two numbers immediately above it.

$$\begin{array}{c}
1\\
1\qquad 1\\
1\qquad 2\qquad 1\\
1\qquad 3\qquad 3\qquad 1\\
1\qquad 4\qquad 6\qquad 4\qquad 1\\
1\qquad 5\qquad 10\qquad 10\qquad 5\qquad 1
\end{array}$$

A step in construction

The number in each cell represents the number of downward routes from the vertex to that point (can you explain why?). It is also a number of ways to choose r objects out of n (can you explain why?), that is, $\binom{n}{r}$.

$$\begin{array}{c}
1\\
1\quad 1\\
1\quad 2\quad 1\\
1\quad 3\quad 3\quad 1\\
1\quad 4\quad 6\quad 4\quad 1\\
1\quad 5\quad 10\quad 10\quad 5\quad 1\\
1\quad 6\quad 15\quad 20\quad 15\quad 6\quad 1\\
1\quad 7\quad 21\quad 35\quad 35\quad 21\quad 7\quad 1\\
1\quad 8\quad 28\quad 56\quad 70\quad 56\quad 28\quad 8\quad 1\\
1\quad 9\quad 36\quad 84\quad 126\quad 126\quad 84\quad 36\quad 9\quad 1\\
1\quad 10\quad 45\quad 120\quad 210\quad 252\quad 210\quad 120\quad 45\quad 10\quad 1
\end{array}$$

The first 10 rows

The combinations numbers are also called *binomial coefficients* and are seen in Calculus. Namely, they are the terms in the expansion

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$$

Note that, if you let $a = b = 1/2$, then on the right-hand side of the sum you will get the probabilities

$$P(a \text{ is chosen } r \text{ times and } b \text{ is chosen } n-r \text{ times}) = \frac{\binom{n}{r}}{2^n}$$

and on the left-hand side you will have 1 (the total of all probabilities).

2.30.
A housewife is asked to rank four brands A, B, C, and D of household cleaner according to her preference, number one being the one she prefers most, etc. she really has no preference among the four brands. Hence, any ordering is equally likely to occur.

a) Find the probability that brand A is ranked number one.
b) Find the probability that brand C is number 1 and D is number 2 in the rankings.
c) Find the probability that brand A is ranked number 1 or number 2.

2.31.
On a given day, 8 soccer games are played. How many different outcomes are possible, if it's known that 4 games are won by the home team, 2 by the visiting team and 2 are drawn?

2.32.
In how many ways can one arrange the letters of the word ADVANTAGE so that the three A's are adjacent to each other?

2.33.
How many distinct "words" can be formed by permuting the letters in the word PROBABILITY?

2.34.
Eight tires of different brands are ranked 1 to 8 (best to worst) according to mileage performance. If four of these tires are chosen at random by a customer, find the probability that the best tire among the four selected by the customer is actually ranked third among the original eight.

2.35.
A drawer contains 3 white and 2 brown socks. Two socks are taken at random. What is the probability that you got two socks of the same color?

2.36.
For password security, it is often recommended that users choose passwords that contain at least two digits, some capital letters, etc. Calculate and compare the available number of passwords when using the following conditions:

a) A 6-letter password only consisting of lowercase letters.
b) A 6-letter password consisting of lowercase and capital letters, with at least 2 capital letters.
c) A 6-letter password consisting of lowercase and capital letters and some digits, with at least 1 capital letter and at least 1 digit.

2.6 Conditional probability and independence

Humans often have to act based on incomplete information. If your boss has looked at you gloomily, you might conclude that something's wrong with your job performance. However, if you know that she just suffered some losses in the stock market, this extra information may change your assessment of the situation. Conditional probability is a tool for dealing with additional information like this.

Conditional probability is the probability of an event occurring given the knowledge that another event has occurred. The conditional probability of event A occurring, given that event B has occurred is denoted by $P(A|B)$ and is read "probability of A given B".

Definition 2.9. Conditional probability

The conditional probability of event A given B is

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)} \text{ for } P(B) > 0 \tag{2.4}$$

Reduced sample space approach

In case when all the outcomes are equally likely, it is sometimes easier to find conditional probabilities directly, without having to apply equation (2.4). If we already know that B has happened, we need only to consider outcomes in B, thus reducing our sample space to B. Then,

$$P(A \,|\, B) = \frac{\text{Number of outcomes in } AB}{\text{Number of outcomes in } B}$$

For example, $P(\text{a die is } 3 \,|\, \text{ a die is odd}) = 1/3$ and
$P(\text{a die is } 4 \,|\, \text{ a die is odd}) = 0$.

Example 2.17.
Let $A = \{\text{a family has two boys}\}$ and $B = \{\text{a family of two has at least one boy}\}$ Find $P(A \,|\, B)$.

Solution. The event B contains the following outcomes: (B, B), (B, G) and (G, B). Only one of these is in A. Thus, $P(A \,|\, B) = 1/3$.

However, if I know that the family has two children, and I see one of the children and it's a boy, then the probability suddenly changes to 1/2. There is a subtle difference in the language and this changes the conditional probability![2] □

Statistical reasoning

Suppose I pick a card at random from a pack of playing cards, without showing you. I ask you to guess which card it is, and you guess the five of diamonds. What is the probability that you are right? Since there are 52 cards in a pack, and only one five of diamonds, the probability of the card being the five of diamonds is 1/52.

Next, I tell you that the card is red, not black. Now what is the probability that you are right? Clearly you now have a better chance of being right than you had before. In fact, your chance of being right is twice as big as it was before, since only half of the 52

[2] Always read the fine print!

cards are red. So the probability of the card being the five of diamonds is now 1/26. What we have just calculated is a conditional probability–the probability that the card is the five of diamonds, given that it is red.

If we let A stand for the card being the five of diamonds, and B stand for the card being red, then the conditional probability that the card is the five of diamonds given that it is red is written $P(A|B)$.

In our case, $P(A \cap B)$ is the probability that the card is the five of diamonds and red, which is 1/52 (exactly the same as P(A), since there are no black fives of diamonds!). P(B), the probability that the card is red, is 1/2. So the definition of conditional probability tells us that $P(A|B) = 1/26$, exactly as it should. In this simple case we didn't really need to use a formula to tell us this, but the formula is very useful in more complex cases.

If we rearrange the definition of conditional probability, we obtain the *multiplication rule* for probabilities:

$$P(A \cap B) = P(A|B)P(B) \tag{2.5}$$

The next concept, *statistical independence* of events, is very important.

Definition 2.10. Independence

The events A and B are called (statistically) independent if

$$P(A \cap B) = P(A)P(B) \tag{2.6}$$

Another way to express independence is to say that the knowledge of B occurring does not change our assessment of $P(A)$. This means that $P(A|B) = P(A)$. (The probability that a person is female given that he or she was born in March is just the same as the probability that the person is female.)

Equation (2.6) is often called *simplified multiplication rule* because it can be obtained from (2.5) by substituting $P(A|B) = P(A)$.

Example 2.18.

For a coin tossed twice, denote H_1 the event that we got Heads on the first toss, and H_2 is the Heads on the second. Clearly, $P(H_1) = P(H_2) = 1/2$. Then, counting the outcomes, $P(H_1H_2) = 1/4 = P(H_1)P(H_2)$, therefore H_1 and H_2 are independent events. This agrees with our intuition that the result of the first toss should not affect the chances for H_2 to occur. □

The situation of the above example is very common for repeated experiments, like rolling dice, or looking at random numbers etc.

Definition 2.10 can be extended to more than two events, but it's fairly difficult to describe.[3] However, it is often used in this context:

If events $A_1, A_2, ..., A_k$ are independent, then

$$P(A_1A_2...A_k) = P(A_1) \times P(A_2) \times ... \times P(A_k) \tag{2.7}$$

[3] For example, the relation $P(ABC) = P(A)P(B)P(C)$ does not guarantee that the events A, B, C are independent.

For example, if we tossed a coin 5 times, the probability that all are Heads is $P(H_1) \times P(H_2) \times ... \times P(H_5) = (1/2)^5 = 1/32$. However, this calculation also extends to outcomes with unequal probabilities.

Example 2.19.
Three bits (0 or 1 digits) are transmitted over a noisy channel, so they will be flipped independently with probability 0.1 each. What is the probability that
a) At least one bit is flipped
b) *Exactly* one bit is flipped?

Solution. a) Using the complement rule, $P(\text{at least one}) = 1 - P(\text{none})$. If we denote F_k the event that kth bit is flipped, then $P(\text{no bits are flipped}) = P(F_1' F_2' F_3') = (1 - 0.1)^3$ due to independence. Then,

$$P(\text{at least one}) = 1 - 0.9^3 = 0.271$$

b) Flipping exactly one bit can be accomplished in 3 ways:

$$P(\text{exactly one}) = P(F_1 F_2' F_3') + P(F_1' F_2 F_3') + P(F_1' F_2' F_3) = 3(0.1)(1 - 0.1)^2 = 0.243$$

It is slightly smaller than the one in part (a). □

Self-test questions

Suppose you throw two dice, one after the other.

a) What is the probability that the first die shows a 2?
b) What is the probability that the second die shows a 2?
c) What is the probability that both dice show a 2?
d) What is the probability that the dice add up to 4?
e) What is the probability that the dice add up to 4 given that the first die shows a 2?
f) What is the probability that the dice add up to 4 and the first die shows a 2?

Answers:

a) The probability that the first die shows a 2 is 1/6.
b) The probability that the second die shows a 2 is 1/6.
c) The probability that both dice show a 2 is (1/6)(1/6) = 1/36 (using the special multiplication rule, since the rolls are independent).
d) For the dice to add up to 4, there are three possibilities–either both dice show a 2, or the first shows a 3 and the second shows a 1, or the first shows a 1 and the second shows a 3. Each of these has a probability of (1/6)(1/6) 3= 1/36 (using the special multiplication rule, since the rolls are independent). Hence the probability that the dice add up to 4 is 1/36 + 1/36 + 1/36 = 3/36 = 1/12 (using the special addition rule, since the outcomes are mutually exclusive).
e) If the first die shows a 2, then for the dice to add up to 4 the second die must also show a 2. So the probability that the dice add up to 4 given that the first shows a 2 is 1/6.

f) Note that we cannot use the simplified multiplication rule here, because the dice adding up to 4 is not independent of the first die showing a 2. So we need to use the full multiplication rule. This tells us that probability that the first die shows a 2 and the dice add up to 4 is given by the probability that the first die shows a 2, multiplied by the probability that the dice add up to 4 given that the first die shows a 2. This is $(1/6)(1/6) = 1/36$.
Alternatively, see part (c). □

Example 2.20. Trees in conditional probability
Suppose we are drawing marbles from a bag that initially contains 7 red and 3 green marbles. The drawing is without replacement, that is after we draw the first marble, we do not put it back. Let's denote the events

$$R_1 = \{\text{ the first marble is red }\} \qquad R_2 = \{\text{ the second marble is red }\}$$

$$G_1 = \{\text{ the first marble is green }\} \qquad \text{and so on.}$$

Let's fill out the tree representing the consecutive choices. See Figure 2.2.

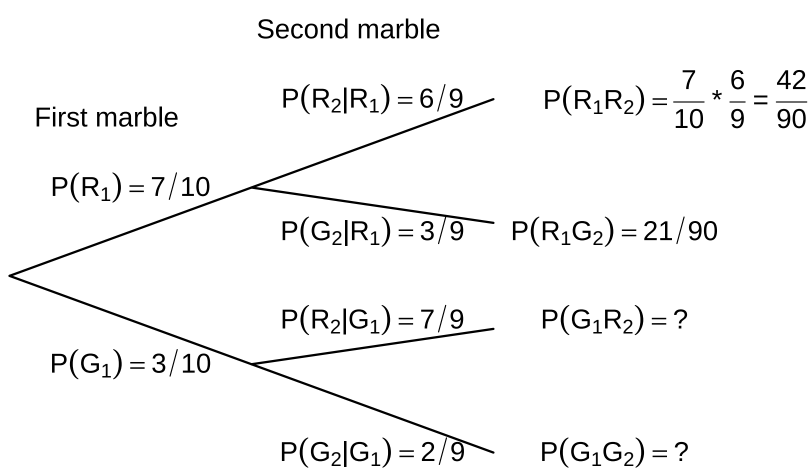


Figure 2.2: Tree diagram for marble choices

The conditional probability $P(R_2 \mid R_1)$ can be obtained directly from reasoning that after we took the first red marble, there remain 6 red and 3 green marbles. On the other hand, we could use the formula (2.4) and get

$$P(R_2 \mid R_1) = \frac{P(R_2R_1)}{P(R_1)} = \frac{42/90}{7/10} = \frac{2}{3}$$

where the probability $P(R_2R_1)$ – same as $P(R_1R_2)$ – can be obtained from counting the outcomes

$$P(R_1R_2) = \frac{\binom{7}{2}}{\binom{10}{2}} = \frac{\frac{7*6}{2*1}}{\frac{10*9}{2*1}} = \frac{42}{90} = \frac{7}{15}$$

Now, can you tell me what $P(R_2)$ and $P(R_1 \mid R_2)$ are? Maybe you know the answer already. However, we will get back to this question in Section 2.7. □

Example 2.21.
Suppose that of all individuals buying a certain digital camera, 60% include an optional memory card in their purchase, 40% include a set of batteries, and 30% include both a card and batteries. Consider randomly selecting a buyer and let A={memory card purchased} and B= {battery purchased}. Then find $P(A|B)$ and $P(B|A)$.

Solution. From given information, we have $P(A) = 0.60, P(B) = 0.40$, and P(both purchased) = $P(A \cap B)$ =0.30. Given that the selected individual purchased an extra battery, the probability that an optional card was also purchased is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.30}{0.40} = 0.75$$

That is, of all those purchasing an extra battery, 75% purchased an optional memory card. Similarly

$$P(\text{battery} \,|\, \text{memory card}) = P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{0.30}{0.60} = 0.50$$

Notice that $P(A|B) \neq P(A)$ and $P(B|A) \neq P(B)$, that is, the events A and B are dependent. □

Exercises

2.37.
A year has 53 Sundays. What is the conditional probability that it is a leap year?

2.38.
The probability that a majority of the stockholders of a company will attend a special meeting is 0.5. If the majority attends, then the probability that an important merger will be approved is 0.9. What is the probability that a majority will attend **and** the merger will be approved?

2.39.
Let events A, B have positive probabilities. Show that, if $P(A \,|\, B) = P(A)$ then also $P(B \,|\, A) = P(B)$.

2.40.
The cards numbered 1 through 10 are placed in a hat, mixed up, then one of the cards is drawn. If we are told that the number on the drawn card is at least five, then what is the probability that it is ten?

2.41.
In the roll of a fair die, consider the events $A = \{2, 4, 6\} =$ "even numbers" and $B = \{4, 5, 6\} =$"high scores". Find the probability that die showing an even number given that it is a high score.

2.42.
There are two urns. In the first urn there are 3 white and 2 black balls and in the second urn there 1 white and 4 black balls. From a randomly chosen urn, one ball is drawn. What is the probability that the ball is white?

2.43.
The level of college attainment of US population by racial and ethnic group in 1998 is given in the following table[b]

Racial or Ethnic Group	Number of Adults (Millions)	Percentage with Associate's Degree	Percentage with Bachelor's Degree	Percentage with Graduate or Professional Degree
Native Americans	1.1	6.4	6.1	3.3
Blacks	16.8	5.3	7.5	3.8
Asians	4.3	7.7	22.7	13.9
Hispanics	11.2	4.8	5.9	3.3
Whites	132.0	6.3	13.9	7.7

The percentages given in the right three columns are conditional percentages.

a) How many Asians have had a graduate or professional degree in 1998?

b) What percent of all adult Americans has had a Bachelor's degree?

c) Given that the person had an Associate's degree, what is the probability that the person was Hispanic?

2.44.
Given that $P(A) = 0.3$, $P(B) = 0.5$ and $P(B \mid A) = 0.4$, find the following

a) $P(AB)$

b) $P(A \mid B)$

c) $P(A' \mid B)$

d) $P(A \mid B')$

2.45.
During the Spring semester, the probability that Johnny was late to school was 0.15. Also, the probability it rained in the morning was 0.2. Finally, the probability it rained **and** Johnny was late to school was 0.1.

a) Find the probability that Johnny was late to school if it rained that morning.

b) Find the probability that Johnny was late to school if it didn't rain that morning.

c) Are the events {Late} and {Rained} independent? Explain.

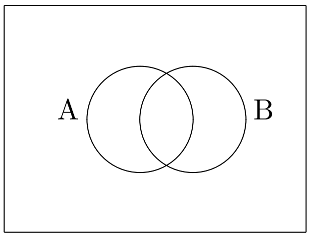
2.46.
The dealer's lot contains 40 cars arranged in 5 rows and 8 columns. We pick one car at random. Are the events $A = \{$the car comes from an odd-numbered row$\}$ and $B = \{$the car comes from one of the last 4 columns$\}$ independent? Prove your point of view.

2.47.
You have sent applications to two colleges. If you are considering your chances to be accepted to either college as 60%, and believe the results are statistically independent, what is the probability that you'll be accepted to at least one?
How will your answer change if you applied to 5 colleges?

2.48.
Show that, if the events A and B are independent, then so are A' and B'.

2.49.
In a high school class, 50% of the students took Spanish, 25% took French and 30% of the students took neither.

Let A = event that a randomly chosen student took Spanish, and B = event that a student took French. Fill in either the Venn diagram or a 2-way table and answer the questions:



	B	B'	
A			
A'			

a) Describe in words the meaning of the event AB'. Find the probability of this event.

b) Are the events A, B independent? Explain with numbers why or why not.

c) If it is known that the student took Spanish, what are the chances that she also took French?

2.50.
Suppose that the events A and B are independent with $P(A \cup B) = 0.7$ and $P(A') = 0.4$. Find $P(A)$.

2.51.
Error-correcting codes are designed to withstand errors in data being sent over communication lines. Suppose we are sending a binary signal (consisting of a sequence of 0's and 1's), and during transmission, any bit may get flipped with probability p, independently of any other bit. However, we might choose to repeat each bit 3 times. For example, if we want to send a sequence 010, we will code it as 000111000. If one of the three bits flips, say, the receiver gets the sequence 001111000, he will still be able to decode it as 010 by majority voting. That is, reading the first three bits, 001, he will interpret it as an attempt to send 000. However, if two of the three bits are flipped, for example 011, this will be interpreted as an attempt to send 111, and thus decoded incorrectly.

What is the probability of a bit being decoded incorrectly under this scheme?[c]

2.52. $\star$
One half of all female physicists are married. Among those married, 50% are married to other physicists, 29% to scientists other than physicists and 21% to nonscientists. Among male physicists, 74% are married. Among them, 7% are married to other physicists, 11% to scientists other than physicists and 82% to nonscientists.[d] What percent of all physicists are female? [**Hint:** This problem can be solved as is, but if you want to, assume that physicists comprise 1% of all population.]

2.53. $\star$
Give an example of events A, B, C such that they are pairwise independent (i.e. $P(AB) = P(A)P(B)$ etc.) but $P(ABC) \neq P(A)P(B)P(C)$. [**Hint:** You may build them on a sample space with 4 elementary outcomes.]

2.7 Bayes Rule

Events $B_1, B_2, \ldots, B_k$ are said to be a **partition** of the sample space $\mathcal{S}$ if the following two conditions are satisfied.

a) $B_i B_j = \emptyset$ for each pair i, j

b) $B_1 \cup B_2 \cup \cdots \cup B_k = \mathcal{S}$

This situation often arises when the statistics are available in subgroups of a population. For example, an insurance company might know accident rates for each age group B_i. This will give the company conditional probabilities $P(A \mid B_i)$ (if we denote $A =$ event of accident).

Question: if we know all the conditional probabilities $P(A \mid B_i)$, how do we find the *un*conditional $P(A)$?

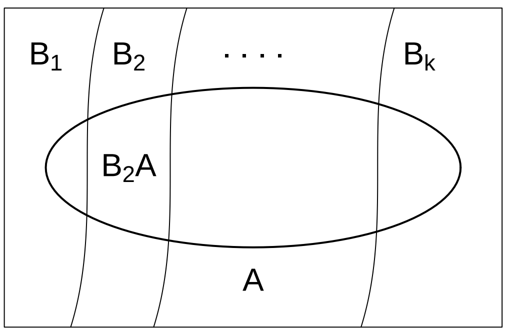


Figure 2.3: Partition $B_1, B_2, \ldots, B_k$ and event A (inside of the oval).

Consider a case when $k = 2$:
The event A can be written as the union of mutually exclusive events AB_1 and AB_2, that is

$$A = AB_1 \cup AB_2 \qquad \text{it follows that} \qquad P(A) = P(AB_1) + P(AB_2)$$

If the conditional probabilities of $P(A|B_1)$ and $P(A|B_2)$ are known, that is

$$P(A|B_1) = \frac{P(AB_1)}{P(B_1)} \qquad \text{and} \qquad P(A|B_2) = \frac{P(AB_2)}{P(B_2)},$$

then $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$.

Suppose we want to find probability of the form $P(B_1|A)$, which can be written as

$$P(B_1|A) = \frac{P(AB_1)}{P(A)} = \frac{P(A|B_1)\ P(B_1)}{P(A)},$$

therefore

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)}$$

This calculation generalizes to $k > 2$ events as follows.

Theorem 2.7. Bayes Rule

If $B_1, B_2, \dots, B_k$ form a partition of the sample space $\mathcal{S}$ such that $P(B_i) \neq 0$ for $i = 1, 2, \dots, k$, then for any event A of $\mathcal{S}$,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i) \tag{2.8}$$

Subsequently,

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)} \tag{2.9}$$

The equation (2.8) is often called *Law of Total Probability.*

Example 2.22.
A rare genetic disease (occuring in 1 out of 1000 people) is diagnosed using a DNA screening test. The test has *false positive rate* of 0.5%, meaning that $P(\text{test positive} \,|\, \text{no disease}) = 0.005$. Given that a person has tested positive, what is the probability that this person actually has the disease?
First, guess the answer, then read on.

Solution. Let's reason in terms of actual numbers of people, for a change. Imagine 1000 people, 1 of them having the disease. How many out of 1000 will test positive? One that actually has the disease, and about 5 disease-free people who would test false positive.[4] Thus, $P(\text{disease} \,|\, \text{test positive}) \approx 1/6$.

It is left as an exercise for the reader to write down the formal probability calculation. □

Example 2.23.
At a certain assembly plant, three machines make 30%, 45%, and 25%, respectively, of the products. It is known from the past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected.

a) What is the probability that it is defective?

b) If a product were chosen randomly and found to be defective, what is the probability that it was made by machine 3?

Solution. Consider the following events:

A: the product is defective

B_1 : the product is made by machine 1,

B_2 : the product is made by machine 2,

B_3 : the product is made by machine 3.

[4] a) Of course, of any actual 1000 people, the number of people having the disease and the number of people who test positive will vary randomly, so our calculation only makes sense when considering averages in a much larger population. b) There's also a possibility of a false negative, i.e. person having the disease and the test coming out negative. We will neglect this, quite rare, event.

Applying additive and multiplicative rules, we can write

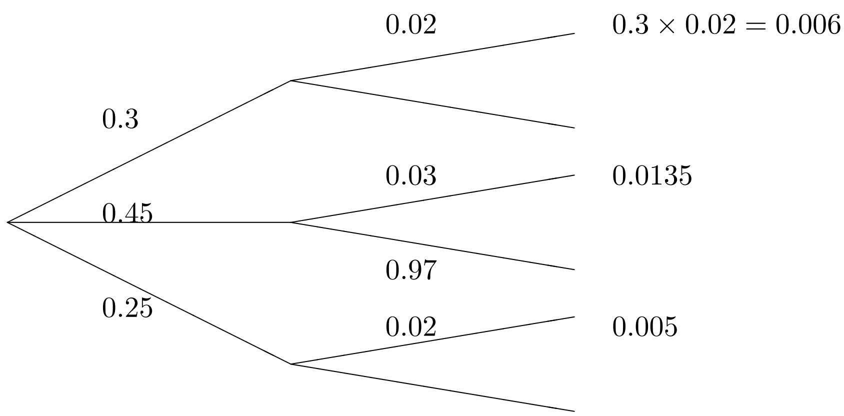
(a) $P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) =$

$= (0.3)(0.02) + (0.45)(0.03) + (0.25)(0.02) = 0.006 + 0.0135 + 0.005 = 0.0245$

(b) Using Bayes' rule

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(A)} = \frac{0.005}{0.0245} = 0.2041$$

This calculation can also be represented using a tree. Here, the first branching represents probabilities of the events B_i, and the second branching represents conditional probabilities $P(A \mid B_i)$. The probabilities of intersections, given by the products, are on the right. $P(A)$ is their sum.



□

Exercises

2.54.

Lucy is undecided as to whether to take a Math course or a Chemistry course. She estimates that her probability of receiving an A grade would be $\frac{1}{2}$ in a math course, and $\frac{2}{3}$ in a chemistry course. If Lucy decides to base her decision on the flip of a fair coin, what is the probability that she gets an A?

2.55.

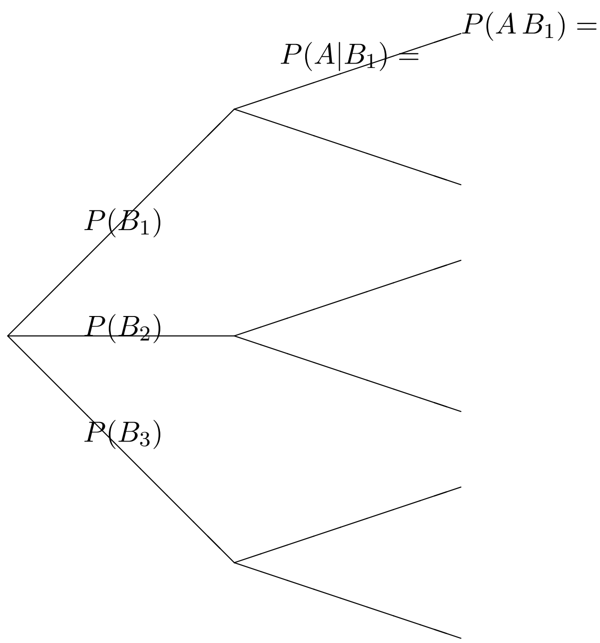
Of the customers at a gas station, 70% use regular gas, and 30% use diesel. Of the customers who use regular gas, 60% will fill the tank completely, and of those who use diesel, 80% will fill the tank completely.

a) What percent of all customers will fill the tank completely?

b) If a customer has filled up completely, what is the probability it was a customer buying diesel?

2.56.
For an on-line electronics retailer, 5% of customers who buy Zony digital cameras will return them, 3% of customers who buy Lucky Star digital cameras will return them, and 8% of customers who buy any other brand will return them. Also, among all digital cameras bought, there are 20% Zony's and 30% Lucky Stars.
Fill in the tree diagram and answer the questions.

(a) What percent of **all** cameras are returned?
(b) If the camera was just returned, what is the probability it is a Lucky Star?
(c) What percent of all cameras sold were Zony **and** were not returned?



2.57.
In 2004, 57% of White households directly and/or indirectly owned stocks, compared to 26% of Black households and 19% of Hispanic households.[e] The data for Asian households is not given, but let's assume the same rate as for Whites. Additionally, 77% of households are classified as either White or Asian, 12% as African American, and 11% as Hispanic.

a) What proportion of all families owned stocks?

b) If a family owned stock, what is the probability it was White/Asian?

2.58.
Drawer one has five pairs of white and three pairs of red socks, while drawer two has three pairs of white and seven pairs of red socks. One drawer is selected at random, and a pair of socks is selected at random from that drawer.

a) What is the probability that it is a white pair of socks.

b) Suppose a white pair of socks is obtained. What is the probability that it came from drawer two?

2.59.
Three newspapers, A, B, and C are published in a certain city. It is estimated from a survey that that of the adult population: 20% read A, 16% read B, 14% read C, 8% read both A and B, 5% read both A and C, 4% read both B and C, 2% read all three. What percentage reads at least one of the papers? Of those that read at least one, what percentage reads both A and B?

2.60.
Suppose $P(A|B) = 0.3$, $P(B) = 0.4$, $P(B|A) = 0.6$. Find:

a) $P(A)$

b) $P(A \cup B)$

2.61. $\star$
This is the famous *Monty Hall problem*.[f] A contestant on a game show is asked to choose among 3 doors. There is a prize behind one door and nothing behind the other two. You (the contestant) have chosen one door. Then, the host is flinging one other door open, and there's nothing behind it. What is the best strategy? Should you switch to the remaining door, or just stay with the door you have chosen? What is your probability of success (getting the prize) for either strategy?

2.62. $\star$
There are two children in a family. We overheard about one of them referred to as a boy.

a) Find the probability that there are 2 boys in the family.

b) Suppose that the oldest child is a boy. Again, find the probability that there are 2 boys in the family.[g] [Why is it different from part (a)?]

Chapter exercises

2.63.
At a university, two students were doing well for the entire semester but failed to show up for the final exam. Their excuse was that they traveled out of state and had a flat tire. The professor gave them the exam in separate rooms, with one question worth 95 points: "which tire was it?". Find the probability that both students mentioned the same tire.[h]

2.64.
In firing the company's CEO, the argument was that during the six years of her tenure, for the last three years the company's market share was lower than for the first three years. The CEO claims bad luck. Find the probability that, given six random numbers, the last three are the lowest among six.

Notes

[a] Taken from Leonard Mlodinow, *The Drunkard's Walk*

[b] Source: US Department of Education, National Center for Education Statistics, as reported in *Chronicle of Higher Education Almanac, 1998-1999*, 2000.

[c] see David MacKay, *Information Theory, Inference, and Learning Algorithms*, 640 pages, Published September 2003.
Downloadable from `http://www.inference.phy.cam.ac.uk/itprnn/book.html`

[d]Laurie McNeil and Marc Sher. *The dual-career-couple problem.* Physics Today, July 1999.

[e]According to `http://www.highbeam.com/doc/1G1-167842487.html`, Consumer Interests Annual, January 1, 2007 by Hanna, Sherman D.; Lindamood, Suzanne

[f]There are some interesting factoids about this in Mlodinow's book, including Marylin vos Savant's column in *Parade* magazine and scathing replies from academics, who believed that the probability was 50%. Vos Savant did it again in 2011 with another probability question that seems, however, intentionally ambiguously worded.

[g]Puzzle cited by Martin Gardner, mentioned in *Math Horizons*, Sept. 2010. See also the discussion at `http://www.stat.columbia.edu/~cook/movabletype/archives/2010/05/hype\_about\_cond.html`

[h]This example is also from Mlodinow's book.

Chapter 3

Discrete probability distributions

3.1 Discrete distributions

In this chapter, we will consider random quantities that are usually called **random variables**.

Definition 3.1. Random variable

A random variable (RV) is a number associated with each outcome of some random experiment.

One can think of the shoe size of a randomly chosen person as a random variable. We have already seen the example when a die was rolled and a number was recorded. This number is also a random variable.

Example 3.1.
Toss two coins and record the number of heads: 0, 1 or 2. Then the following outcomes can be observed.

Outcome	TT	HT	TH	HH
Number of heads	0	1	1	2

The random variables will be denoted with capital letters $X, Y, Z, ...$ and the lowercase x would represent a particular value of X. For the above example, $x = 2$ if heads comes up twice. Now we want to look at the probabilities of the outcomes. For the probability that the random variable X has the value x, we write $P(X = x)$, or just $p(x)$.

For the coin flipping random variable X, we can make the table:

x	0	1	2
$p(x)$	1/4	1/2	1/4

This table represents the *probability distribution* of the random variable X.

Definition 3.2. Probability mass function

A random variable X is said to be **discrete** if it can take on only a finite or countable[a] number of possible values x. In this case,

a) $P(X = x) = p_X(x) \geq 0$

b) $\sum_x P(X = x) = 1$, where the sum is over all possible x

The function $p_X(x)$ or simply $p(x)$ is called *probability mass function* (PMF) of X.

[a] A set is called *countable* if it can be enumerated with positive integers $1, 2, 3, \ldots$. Most frequently we will use integers themselves, or nonnegative integers as possible values of X. Note, however, that the set of all rational fractions m/n, where both m and n are integers, is also countable.

What does this actually mean? A discrete probability function is a function that can take a discrete number of values (not necessarily finite). There is no mathematical restriction that discrete probability functions only be defined at integers, but we will use integers in many practical situations. For example, if you toss a coin 6 times, you can get 2 heads or 3 heads but not 2.5 heads.

Each of the discrete values has a certain probability of occurrence that is between zero and one. That is, a discrete function that allows negative values or values greater than one is not a PMF. The condition that the probabilities add up to one means that one of the values has to occur.

Example 3.2.
A shipment of 8 similar microcomputers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability mass function for the number of defectives.

Solution. Let X be a random variable whose values x are the possible numbers of defective computers purchased by school. Then x must be $0, 1$ or 2. Then,

$$P(X = 0) = \frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28}$$

$$P(X = 1) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28}$$

$$P(X = 2) = \frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \frac{3}{28}$$

Thus, the probability mass function of X is

x	0	1	2
$p(x)$	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$

□

Definition 3.3. Cumulative distribution function

The cumulative distribution function (CDF) $F(x)$ for a random variable X is defined as

$$F(x) = P(X \leq x)$$

If X is discrete,

$$F(x) = \sum_{y \leq x} p(y)$$

where $p(x)$ is the probability mass function.

Properties of discrete CDF

a) $\lim_{x \to -\infty} F(x) = 0$

b) $\lim_{x \to \infty} F(x) = 1$

c) $F(x)$ is non-decreasing

d) $p(x) = F(x) - F(x-) = F(x) - \lim_{y \uparrow x} F(y)$

In words, CDF of a discrete RV is a step function, whose jumps occur at the values x for which $p(x) > 0$ and are equal in size to $p(x)$. It ranges from 0 on the left to 1 on the right.

Example 3.3.
Find the CDF of the random variable from Example 3.2. Using $F(x)$, verify that $P(X = 1) = 15/28$.

Solution. The CDF of the random variable X is:
$F(0) = p(0) = \frac{10}{28}$
$F(1) = p(0) + p(1) = \frac{25}{28}$
$F(2) = p(0) + p(1) + p(2) = \frac{28}{28} = 1.$

Hence,

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 10/28 & \text{for } 0 \leq x < 1 \\ 25/28 & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases} \tag{3.1}$$

Now, $P(X = 1) = p(1) = F(1) - F(0) = \frac{25}{28} - \frac{10}{28} = \frac{15}{28}$. □

Graphically, $p(x)$ can be represented as a *probability histogram* where the heights of the bars are equal to $p(x)$.

Exercises

3.1.
Suppose that two dice are rolled independently, with outcomes X_1 and X_2. Find the distribution of the random variable $Y = X_1 + X_2$. [**Hint:** It's easier to visualize all the ourcomes if you make a two-way table.]

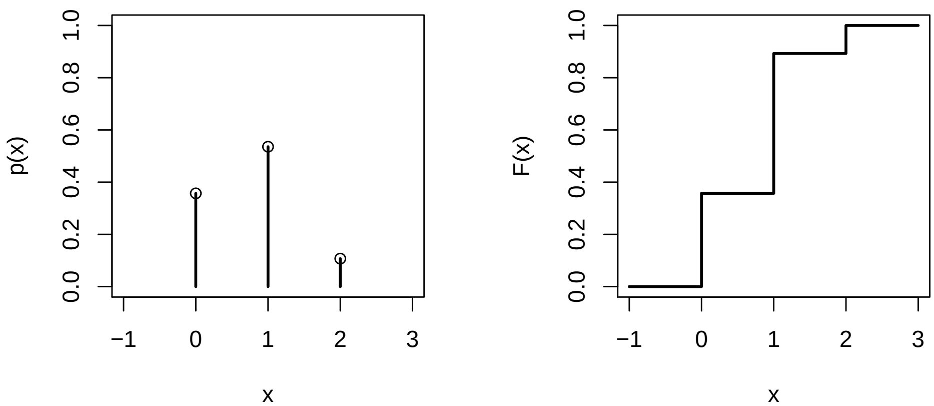


Figure 3.1: PMF and CDF for Example 3.3

3.2.
What constant c makes $p(x)$ a valid PMF?

a) $p(x) = c$ for $x = 1, 2, ..., 5$.

b) $p(x) = c(x^2 + 1)$ for $x = 0, 1, 2, 3$.

c) $p(x) = cx\binom{3}{x}$ for $x = 1, 2, 3$.

3.3.
Are the following valid PMF's? If yes, find constant k that makes it so.

a) $p(x) = (x - 2)/k$ for $x = 1, 2, ..., 5$

b) $p(x) = (x^2 - x + 1)/k$ for $x = 1, 2, ..., 5$

c) $p(x) = \dfrac{k}{2^x}$ for $x = -1, 0, 1, 2$

3.4.
With reference to the previous problem find an expression for the values of $F(x)$, that is CDF of X.

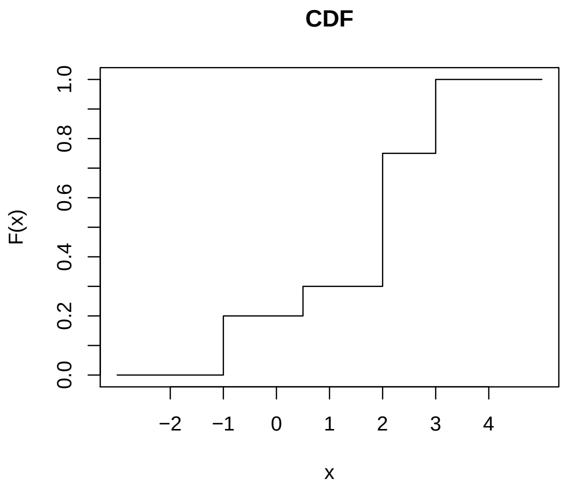
3.5.
For an on-line electronics retailer, X = the number of Zony digital cameras returned per day follows the distribution given by

x	0	1	2	3	4	5
$p(x)$	0.05	0.1	?	0.2	0.25	0.1

(a) Fill in the "?"
(b) Find $P(X > 3)$
(c) Find the CDF of X (make a table).

3.6.
Out of 5 components, 3 are domestic and 2 are imported. 3 components are selected at random (without replacement). Calculate the PMF for X = number of domestic components picked (make a table).

3.7.
The CDF of a discrete random variable X is shown in the plot below.



Find the probability mass function $p_X(x)$ (make a table)

3.2 Expected values of Random Variables

One of the most important things we'd like to know about a random variable is: what value does it take on average? What is the average price of a computer? What is the average value of a number that rolls on a die?

The value is found as the average of all possible values, *weighted* by how often they occur (i.e. probability)

Definition 3.4. Expected value (mean)

The mean or expected value of a discrete random variable X with probability mass function $p(x)$ is given by

$$\mathbb{E}(X) = \sum_x x\, p(x)$$

We will sometimes use the notation $\mathbb{E}(X) = \mu$.

Theorem 3.1. Expected value of a function

If X is a discrete random variable with probability mass function $p(x)$ and if $g(x)$ is a real valued function of x, then

$$\mathbb{E}[g(X)] = \sum_x g(x)p(x).$$

Definition 3.5. Variance

The variance of a random variable X with expected value μ is given by

$$V(X) = \sigma^2 = \mathbb{E}\,(X-\mu)^2 = \mathbb{E}\,(X^2) - \mu^2,$$

where

$$\mathbb{E}\,(X^2) = \sum_x x^2\, p(x).$$

The variance defines the average (or expected) value of the squared difference from the mean.
If we use $V(X) = \mathbb{E}\,(X-\mu)^2$ as a definition, we can see that

$$V(X) = \mathbb{E}\,(X-\mu)^2 = \mathbb{E}\,(X^2 - 2\mu X + \mu^2) = \mathbb{E}\,(X^2) - 2\mu\mathbb{E}\,(X) + \mu^2 = \mathbb{E}\,(X^2) - \mu^2$$

due to the linearity of expectation (see Theorem 3.2 below).

Definition 3.6. Standard deviation

The standard deviation of a random variable X is the square root of the variance, and is given by

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mathbb{E}\,(X-\mu)^2}$$

The mean describes the *center* of the probability distribution, while standard deviation describes the *spread*. Larger values of σ signify a distribution with larger variation. This will be undesirable in some situations, e.g. industrial process control, where we would like the manufactured items to have identical characteristics. On the other hand, a *degenerate* random variable X that has $P(X = a) = 1$ for some value of a is not random at all, and it has the standard deviation of 0.

Example 3.4.
The number of fire emergencies at a rural county in a week, has the following distribution

x	0	1	2	3	4
$P(X = x)$	0.52	0.28	0.14	0.04	0.02

Find $\mathbb{E}\,(X)$, $V(X)$ and σ.

Solution. From Definition 3.4, we see that

$$\mathbb{E}\,(X) = 0(0.52) + 1(0.28) + 2(0.14) + 3(0.04) + 4(0.02) = 0.76 = \mu$$

and from definition of $\mathbb{E}\,(X^2)$, we get

$$\mathbb{E}\,(X^2) = 0^2(0.52) + 1^2(0.28) + 2^2(0.14) + 3^2(0.04) + 4^2(0.02) = 1.52$$

Hence, from Definition 3.5, we get

$$V(X) = \mathbb{E}\,(X^2) - \mu^2 = 1.52 - (0.76)^2 = 0.9424$$

Now, from Definition 3.6, the standard deviation $\sigma = \sqrt{0.9424} = 0.9708$. $\square$

Theorem 3.2. Linear functions

For any random variable X and constants a and b,

a) $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$

b) $V(aX+b) = a^2\, V(X) = a^2\,\sigma^2$

c) $\sigma_{aX+b} = |a|\,\sigma$.

d) For several RV's, $X_1, X_2, ..., X_k$,

$$\mathbb{E}(X_1 + X_2 + ... + X_k) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + ... + \mathbb{E}(X_k)$$

Example 3.5.
Let X be a random variable having probability mass function given in Example 3.4. Calculate the mean[1] and variance of $g(X) = 4X + 3$.

Solution. In Example 3.4, we found $\mathbb{E}(X) = \mu = 0.76$ and $V(X) = 0.9424$. Now, using Theorem 3.2,

$$\mathbb{E}(g(X)) = 4\mathbb{E}(X) + 3 = 4(0.76) + 3 = 3.04 + 3 = 6.04$$

$$\text{and} \qquad V(g(X)) = 4^2\, V(X) = 16(0.9424) = 15.08$$

□

Theorem 3.3. Chebyshev Inequality

Let X be a random variable with mean μ and a variance σ^2. Then for any positive k,

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

The inequality in the statement of the theorem is equivalent to

$$P(\mu - k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2}$$

To interpret this result, let $k = 2$, for example. Then the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ must contain at least $1 - \frac{1}{k^2} = 1 - \frac{1}{4} = \frac{3}{4}$ of the probability mass for the random variable.

Chebyshev inequality is useful when the mean and variance of a RV are known and we would like to calculate estimates of some probabilities. However, these estimates are usually quite crude.

Example 3.6.
The performance period of a certain car battery is known to have a mean of 30 months and standard deviation of 5 months.

a) Estimate the probability that a car battery will last at least 18 months.

b) Give a range of values to which at least 90% of all batteries' lifetimes will belong.

[1] Note that in general $\mathbb{E}(g(X)) \neq g(\mathbb{E}\,X)$, the equality is guaranteed only if g is a linear function!

Solution. (a) Let X be the battery performance period. Calculate k such that the value of 18 is k standard deviations below the mean: $18 = 30 - 5k$, therefore $k = (30 - 18)/5 = 2.4$. From Chebyshev's theorem we have

$$P(30 - 5k < X < 30 + 5k) > 1 - 1/k^2 = 1 - 1/2.4^2 = 0.826$$

Thus, at least 82.6% of batteries will make it to 18 months. (However, in reality this percentage could be much higher, depending on distribution.)

(b) From Chebyshev's theorem we have

$$P(\mu - k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2}$$

According to the problem set $1 - \frac{1}{k^2} = 0.90$ and solve for k, we get $k = \sqrt{10} = 3.16$. Hence, the desired interval is between $30 - 3.16(5)$ and $30 + 3.16(5) = 14.2$ to 45.8 months. □

Example 3.7.
The number of customers per day at a certain sales counter, X, has a mean of 20 customers and standard deviation of 2 customers. The probability distribution of X is not known. What can be said about the probability that X will be between 16 and 24 tomorrow?

Solution. We want $P(16 \le X \le 24) = P(15 < X < 25)$. From Chebyshev's theorem

$$P(\mu - k\sigma < X < \mu + k\sigma) \ge 1 - \frac{1}{k^2}$$

given $\mu = 20, \sigma = 2$ we set $\mu - k\sigma = 15$ and hence $k = 2.5$. Thus, $P(16 \le X \le 24) \ge 1 - \frac{1}{6.25} = 0.84$.
So, tomorrow's customer total will be between 16 and 24 with probability at least 0.84. □

Exercises

3.8.
Timmy is selling cholocates door to door. The probability distribution of X, the number of cholocates he sells in each house, is given by

x	0	1	2	3	4
$P(X = x)$	0.45	0.25	0.15	0.1	0.05

Find the expected value and standard deviation of X.

3.9.
In the previous exercise, suppose that Timmy earns 50 cents for school from each purchase. Find the expected value and standard deviation of his earnings per house.

3.10.
For the exercise 3.8, calculate the random variables $g_1(X) = 2X - 1$ and $g_2(X) = X^3$. [Hint: keep the same probability values and change all the X-values into $g(X)$.] For which of these functions you can claim that $\mathbb{E}[g(X)] = g[\mathbb{E}(X)]$? Verify numerically.

3.11.
A dollar coin, a quarter, a nickel and a dime are tossed. I get to pocket all the coins that came up heads. What are my expected winnings?

3.12.
Consider X with the distribution of a random digit, $p(x) = 1/10$, $x = 0, 1, 2, ..., 9$

a) Find the mean and standard deviation of X.

b) According to Chebyshev's inequality, estimate the probability that a random digit will be between 1 and 8, inclusive. Compare to the actual probability.

3.13.
In the *Numbers* game, two players choose a random number between 1 and 6, and compute the absolute difference.
That is, if Player 1 gets the number Y_1, and Player 2 gets Y_2, then they find

$$X = |Y_1 - Y_2|$$

a) Find the distribution of the random variable X (make a table). [Hint: consider all outcomes (y_1, y_2).]

b) Find the expected value and variance of X, and $\mathbb{E}(X^3)$

c) If Player 1 wins whenever the difference is 3 or more, and Player 2 wins whenever the difference is 2 or less, who is more likely to win?

d) If Player 1 bets \$1, what is the value that Player 2 should bet to make the game fair?

3.14.
According to `ScanUS.com`, the number of cars per household in an Albuquerque neighborhood was distributed as follows

x	0	1	2	3+
$P(X = x)$	0.047	0.344	0.402	0.207

`3+` really means 3 or more, but let's assume that there are no more than 3 cars in any household.

Find the expected value and standard deviation of X.

3.15.
For the above Problem, the web site really reported the average of 1.9 cars per household. This is higher than the answer for the Problem 3.14. Probably, it's due to the fact that we limited the number of cars by 3.

Suppose we limit the number of cars by 4. This means the distribution will look like

x	0	1	2	3	4
$p(x)$	0.047	0.344	0.402	p_3	p_4

where $p_3 + p_4 = 0.207$. Assuming that $\mathbb{E}(X) = 1.9$, reverse-engineer this information to find p_3 and p_4.

3.16.

The frequencies of electromagnetic waves in the upper ionosphere observed in the vicinity of earthquakes have the mean 1.7 kHz, and standard deviation of 0.2 kHz. According to Chebyshev inequality,

a) What percent of all observed waves is guaranteed to be contained in the interval 1.4 to 2.0 kHz?

b) Give an interval that would contain at least 95% of all such observed waves.

3.17.

Find the mean and variance of the given PMF $p(x) = 1/k$, where $x = 1, 2, 3, ..., k$.

3.18.

Show that the function defined by $p(x) = 2^{-x}$ for $x = 1, 2, 3, ...$ can represent a probability mass function of a random variable X. Find the mean and the variance of X.

3.19.

For $t > 0$ show that $p(x) = e^{-t}(1 - e^{-t})^{x-1},\ x = 1, 2, 3, ...$ can represent a probability mass function. Also, find $\mathbb{E}(X)$ and $V(X)$.

3.20. $\star$

The average salary of the employees in a firm is 80 thousand dollars, and the standard deviation is 100 thousand. Given that the salary can't be negative, what can you say about the proportion of the employees who earn more than 150 thousand?

3.21. $\star$ "Baker's problem"

A shopkeeper is selling the quantity X (between 0 and 3) of a certain item per week, with a given probability distribution:

x	0	1	2	3
$p(x)$	0.05	0.2	0.5	0.25

For each item bought, the profit is \$50. On the other hand, if the item is stocked, but was not bought, then the cost of upkeep, insurance etc. is \$20. At the beginning of the week, the shopkeeper stocks a items.
For example, if 3 items were stocked, then the expected profit can be calculated from the following table:

Y = Profit

y	$-\$60$	\$10	\$80	\$150
$p(y)$	0.05	0.2	0.5	0.25

a) What is the expected profit if the shopkeeper stocked $a = 3$ items?

b) What is the expected profit if the shopkeeper stocked $a = 1$ and $a = 2$ items? [You'll need to produce new tables for Y first.]

c) Which value of a maximizes the expected profit?

3.3 Bernoulli distribution

Let X be the random variable denoting the condition of the inspected item. Agree to write $X = 1$ when the item is defective and $X = 0$ when it is not. (This is a convenient notation because, once we inspect n such items, $X_1, X_2, ..., X_n$ denoting their condition, the total number of defectives will be given by $X_1 + X_2 + ... + X_n$.)

Let p denote the probability of observing a defective item. The probability distribution of X, then, is given by

x	0	1
$p(x)$	$q = 1 - p$	p

Such a random variable is said to have a *Bernoulli distribution.* Note that

$$\mathbb{E}\,(X) = \sum xp(x) = 0 \times p(0) + 1 \times p(1) = 0(q) + 1(p) = p \qquad \text{and}$$

$$\mathbb{E}\,(X^2) = \sum x^2\, p(x) = 0(q) + 1(p) = p.$$

Hence, $V(X) = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = p - p^2 = pq$.

3.4 Binomial distribution

Now, let us inspect n items and count the total number of defectives. This process of repeating an experiment n times is called **Bernoulli trials**. The Bernoulli trials are formally defined by the following properties:

a) The result of each trial is either a success or a failure

b) The probability of success p is constant from trial to trial.

c) The trials are independent

d) The random variable X is defined to be the number of successes in n repeated trials

This situation applies to many random processes with just two possible outcomes: a heads-or-tails coin toss, a made or missed free throw in basketball etc[2]. We arbitrarily call one of these outcomes a "success" and the other a "failure".

Definition 3.7. Binomial RV

Assume that each Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

$$P(X = k) = \binom{n}{k} p^k\, q^{n-k}, \qquad k = 0, 1, 2, \ldots, n.$$

The mean and variance of the binomial distribution are

$$\mathbb{E}\,(X) = \mu = np \quad \text{and} \quad V(X) = \sigma^2 = npq.$$

We can notice that the mean and variance of the Binomial are n times larger than those

[2]However, we have to make sure that the probability of success remains constant. Thus, for example, wins or losses in a series of football games may not be a Bernoulli experiment!

of the Bernoulli random variable.

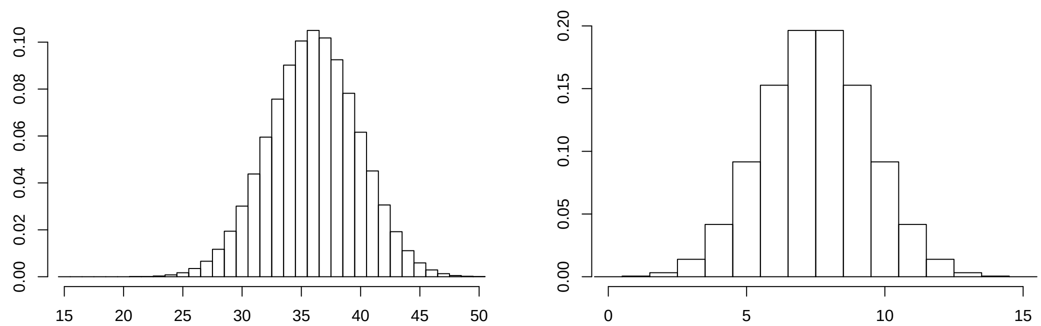


Figure 3.2: Binomial PMF: left, with $n = 60$, $p = 0.6$; right, with $n = 15$, $p = 0.5$

Note that Binomial distribution is symmetric when $p = 0.5$. Also, two Binomials with the same n and $p_2 = 1 - p_1$ are mirror images of each other.

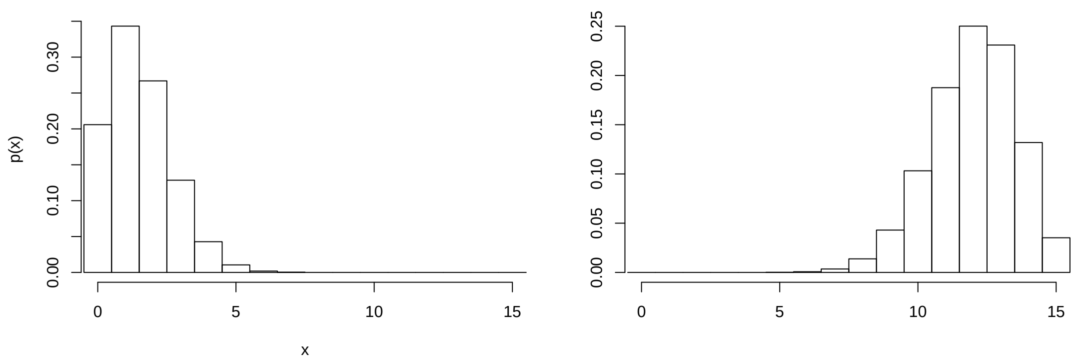


Figure 3.3: Binomial PMF: left, with $n = 15$, $p = 0.1$; right, with $n = 15$, $p = 0.8$

Example 3.8.
The probability that a certain kind of component will survive a shock test is 0.75. Find the probability that

a) exactly 2 of the next 8 components tested survive,

b) at least 2 will survive,

c) at most 6 will survive.

Solution. (a) Assuming that the tests are independent and $p = 0.75$ for each of the 8 tests, we get

$$P(X = 2) = \binom{8}{2} (0.75)^2\,(0.25)^{8-2} = \frac{8!}{2!\,(8-2)!}\,0.75^2\,0.25^6 =$$

$$= \frac{40320}{2 \times 720}\,(0.5625)(0.000244) = 0.003843$$

(b)

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

$$= 1 - [8(0.75)(0.000061) + 0.000002] = 1 - 0.000386 \approx 0.9996$$

(c)
$$P(X \le 6) = 1 - P(X \ge 7) = 1 - [P(X = 7) + P(X = 8)]$$
$$= 1 - [0.2669 + 0.1001] = 1 - 0.367 = 0.633$$

□

Example 3.9.
It has been claimed that in 60% of all solar heating installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one-third in
(a) four of five installations;
(b) at least four of five installations?

Solution.

$$\text{(a)} \quad P(X = 4) = \binom{5}{4}(0.60)^4\,(0.4)^{5-4} = 5(0.1296)(0.4) = 0.2592$$

$$\text{(b)} \quad P(X = 5) = \binom{5}{5}(0.60)^5\,(0.40)^{5-5} = 0.60^5 = 0.0777$$

Hence, $P(\text{reduction for at least four}) = P(X \ge 4) = 0.2592 + 0.0777 = 0.3369$ □

Exercises

3.22.
There's 50% chance that a mutual fund return on any given year will beat the industry's average. What proportion of funds will beat the industry average for at least 4 out of 5 last years?

3.23.
Biologists would like to catch Costa Rican glass frogs for breeding. There is 75% probability that a glass frog they catch is male. If 10 glass frogs of a certain species are caught, what are the chances that they will have at least 2 male and 2 female frogs? What is the expected value of the number of female frogs caught?

3.24.
A 5-member focus group are testing a new game console. Suppose that there's 50% chance that any given group member approves of the new console, and their opinions are independent of each other.

a) Calculate and fill out the probability distribution for X = number of group members who approve of the new console.

b) Calculate $P(X \ge 3)$.

c) How does your answer in part (b) change when there's 70% chance that any group member approves of the new console?

3.25.
Suppose that the four engines of a commercial airplane were arranged to operate independently and that the probability of in-flight failure of a single engine is 0.01. Find:

a) Probability of no failures on a given flight.

b) Probability of at most one failure on a given flight.

c) The mean and variance for the number of failures on a given flight.

3.26.

Suppose a television contains 60 transistors, 2 of which are defectives. Five transistors are selected at random, removed and inspected. Approximate

a) probability of selecting no defectives,

b) probability of selecting at least one defective.

c) The mean and variance for the number of defectives selected.

3.27.

A train is made up of 50 railroad cars. Each car may need service with probability 0.05. Let X be the total number of cars in the train that need service.

a) Find the mean and standard deviation of X.

b) Find the probability that no cars need service.

c) Find the probability that at least two cars need service.

3.28.

Show that mean and variance of the binomial random variable X are np and npq respectively.

3.29.

If a thumb-tack is flipped, then the probability that it will land point-up is 1/3. If this thumb-tack is flipped 6 times, then find:

a) the probability that it lands point-up on exactly 2 flips,

b) at least 2 flips,

c) at most 4 flips.

3.30.

The proportion of people with type A blood in a certain city is reported to be 0.20. Suppose a random group of 20 people is taken and their blood types are to be checked. What is the probability that there are at least 4 people who have type A blood in the sample? What is the probability that at most 5 people in the group have type A blood?

3.31.

A die and a coin are tossed together. Let us define success as the event that the die shows an odd number and the coin shows a head. We repeat the experiment 5 times. What is the probability of exactly 3 successes?

3.5 Geometric distribution

In the case of Binomial distribution, the number of trials was a fixed number n, and the variable of interest was the number of successes. It is sometimes of interest to count instead how many trials are required to achieve a specified number of successes.

The number of trials Y required to obtain the first success is called a *Geometric random variable* with parameter p.

Theorem 3.4. Geometric RV

The probability mass function for a Geometric random variable is

$$g(y;p) := P(Y=y) = (1-p)^{y-1}p, \qquad y = 1,2,3,\ldots$$

Its CDF is

$$F(y) = 1 - q^y, \qquad y = 1,2,3,\ldots, \qquad q = 1-p$$

Its mean and variance are

$$\mu = \frac{1}{p} \qquad \text{and} \qquad \sigma^2 = \frac{1-p}{p^2}$$

Proof. To achieve the first success on yth trial means to have the first $y-1$ trials to result in failures, and the last yth one a success, and then by independence of trials,

$$P(FF...FS) = q^{y-1}p$$

Now the CDF

$$F(y) = P(Y \le y) = 1 - P(Y > y)$$

The latter means that all the trials up to and including the yth one, resulted in failures, which equals $P(y \text{ failures in a row}) = q^y$ and we get the CDF subtracting this from 1.

The mean $\mathbb{E}(Y)$ can be found by differentiating a geometric series:

$$\mathbb{E}(Y) = \sum_{y=1}^{\infty} yp(y) = \sum_{y=1}^{\infty} yp(1-p)^{y-1} = p\sum_{y=1}^{\infty} y(1-p)^{y-1} =$$

$$= p\sum_{y=1}^{\infty} \frac{d}{dq}q^y = p\,\frac{d}{dq}\sum_{y=1}^{\infty} q^y = p\left[\frac{d}{dq}(1+q+q^2+q^3+\cdots-1)\right] =$$

$$= p\left\{\frac{d}{dq}\left[(1-q)^{-1}\right] - \frac{d}{dq}(1)\right\} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

The variance can be calculated by differentiating a geometric series twice:

$$\mathbb{E}\{Y(Y-1)\} = \sum_{y=2}^{\infty} y(y-1)p\,q^{y-1} = pq\sum_{y=2}^{\infty}\frac{d^2}{dq^2}(q^y) =$$

$$= pq\frac{d^2}{dq^2}\left[\sum_{y=0}^{\infty} q^y\right] = pq\,\frac{d^2}{dq^2}(1-q)^{-1} = pq\,\frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

$$\text{Hence} \qquad \mathbb{E}(Y^2) = \frac{2q}{p^2} + \frac{1}{p} \qquad \text{and} \qquad V(Y) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}$$

□

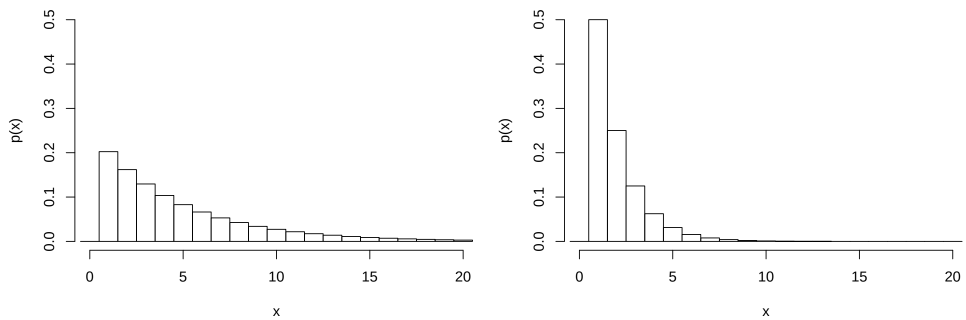


Figure 3.4: Geometric PMF: left, with $p = 0.2$; right, with $p = 0.5$

Example 3.10.
For a certain manufacturing process it is known that, on the average, 1 in every 100 items is defective. What is the probability that the first defective item found is the fifth item inspected? What is the average number of items that should be sampled before the first defective is found?

Solution. Using the geometric distribution with $x = 5$ and $p = 0.01$, we have
$g(5; 0.01) = (0.01)(0.99)^4 = 0.0096$.

Mean number of items needed is $\mu = 1/p = 100$. □

Example 3.11.
If the probability is 0.20 that a burglar will get caught on any given job, what is the probability that he will get caught no later than on his fourth job?

Solution. Substituting $y = 4$ and $p = 0.20$ into the geometric CDF, we get
$P(Y \leq 4) = 1 - 0.8^4 = 0.5904$ □

Exercises

3.32.
The probability to be caught while running a red light is estimated as 0.1. What is the probability that a person is first caught on his 10th attempt to run a red light? What is the probability that a person runs a red light at least 10 times without being caught?

3.33.
A computing center is interviewing people until they find a qualified person to fill a vacant position. The probability that any single applicant is qualified is 0.15.

a) Find the expected number of people to interview.

b) Find the probability the center will need to interview between 4 and 8 people (inclusive).

3.34.
From past experience it is known that 3% of accounts in a large accounting population are in error. What is the probability that the first account in error is found on the 5th try? What is the probability that the first account in error occurs in the first five accounts audited?

3.35.
A rat must choose between five doors, one of which contains chocolate. If the rat chooses the wrong door, it is returned to the starting point and chooses again (randomly), and continues until it gets the chocolate. What is the probability of the rat getting chocolate on the second attempt?

3.36.
If the probability of a success is 0.01, how many trials are necessary so that probability of at least one success is greater than 0.5?

3.6 Negative Binomial distribution

Let Y denote the number of the trial on which the rth success occurs in a sequence of independent Bernoulli trials, with p the probability of success. Such Y is said to have *Negative Binomial distribution.* When $r = 1$, we will of course obtain the Geometric distribution.

Theorem 3.5. Negative Binomial RV

The PMF of the Negative Binomial random variable Y is

$$nb(y; r, p) := P(Y = y) = \binom{y-1}{r-1} p^r\, q^{y-r}, \qquad y = r, r+1, \ldots$$

The mean and variance of Y are:

$$\mathbb{E}\,(Y) = \frac{r}{p} \qquad \text{and} \qquad V(Y) = \frac{rq}{p^2}.$$

Proof. We have $P(Y = y) =$

$$= P[\text{First } y-1 \text{ trials contain } r-1 \text{ successes and } y\text{th trial is a success}] =$$

$$= \binom{y-1}{r-1} p^{r-1}\, q^{y-r} \times p = \binom{y-1}{r-1} p^r\, q^{y-r}, \qquad y = r, r+1, r+2, \ldots$$

The proof for the mean and variance uses the properties of the independent sums to be discussed in Section 5.4. However, note at this point that both μ and σ^2 are r times larger than those of the Geometric distribution. □

Example 3.12.
In an NBA championship series, the team which wins four games out of seven will be the winner. Suppose that team A has probability 0.55 of winning over the team B, and the teams A and B face each other in the championship games.
(a) What is the probability that team A will win the series in six games?
(b) What is the probability that team A will win the series?

Solution.
(a) $nb(6; 4, 0.55) = \binom{5}{3}\ (0.55)^4\, (1 - 0.55)^{6-4} = 0.1853.$
(b) P(team A wins the championship series) =

$$= nb(4; 4, 0.55) + nb(5; 4, 0.55) + nb(6; 4, 0.55) + nb(7; 4, 0.55) =$$

$$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083$$

□

Example 3.13.
A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new childbirth regimen. She anticipates that 20% of all couples she asks will agree. What is the probability that 15 couples must be asked before 5 are found who agree to participate?

Solution. Substituting $x = 15, p = 0.2, r = 5$, we get

$$nb(15; 5, 0.2) = \binom{14}{4}(0.2)^5(0.8)^{15-5} = 0.034$$

□

Exercises

3.37.
Biologists catch Costa Rican glass frogs for breeding. There is 75% probability that a glass frog they catch is male. Biologists would like to have at least 2 female frogs. What is the expected value of the total number of frogs caught, until they reach their goal? What is the probability that they will need exactly 6 frogs to reach their goal?

3.38.
Jim is a high school baseball player. He has 0.25 batting average, meaning that he makes a hit in 25% of his tries ("at-bats")[3]. What is the probability that Jim makes his second hit of the season on his sixth at-bat?

3.39.
A telemarketer needs to sell 3 insurance policies before lunch. He estimates the probability of a sale as 0.1. How many calls, on average, does he need to make before lunch? What is the probability that he needs exactly 25 calls to reach his goal?

3.40.
In the best-of-5 series, Team A has 60% chance to win any single game, and the outcomes of the games are independent. Find the probability that Team A will win the series (i.e. will win the majority of the games).

3.41.
For Problem 3.40, find the expected duration of the series (regardless of which team wins). [**Hint:** First, fill out the table containing d, $p(d)$ – the distribution of the duration D. For example, $P(D = 3) = P(\text{team A wins in 3}) + P(\text{team B wins in 3})$]

[3] For baseball experts, let's exclude the possibility of a walk.

3.7 Poisson distribution

It is often useful to define a random variable that counts the number of events that occur within certain specified boundaries. For example, the average number of telephone calls received by customer service within a certain time limit. The Poisson distribution is often appropriate to model such situations.

Definition 3.8. Poisson RV

A random variable X with a Poisson distribution takes the values $x = 0, 1, 2, \ldots$ with a probability mass function

$$pois(x; \mu) := P(X = x) = \frac{e^{-\mu}\,\mu^x}{x!}$$

where μ is the parameter of the distribution.[a]

[a] Some textbooks use λ for the parameter. We will use λ for the intensity of the Poisson process, to be discussed later.

Theorem 3.6. Mean and variance of Poisson RV

For Poisson RV with parameter μ,

$$\mathbb{E}\,(X) = V(X) = \mu.$$

Proof. Recall the Taylor series expansion of e^x:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Now,

$$\mathbb{E}\,(X) = \sum x * pois(x, \mu) = \sum_{x=0}^{\infty} x\,\frac{e^{-\mu}\,\mu^x}{x!} = \sum_{x=1}^{\infty} \frac{x\,e^{-\mu}\,\mu\,\mu^{x-1}}{x(x-1)!} =$$

$$= \mu\,e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} = \mu\,e^{-\mu}\left[1 + \frac{\mu}{1!} + \frac{\mu^2}{2!} + \frac{\mu^3}{3!}\ldots\right] = \mu\,e^{-\mu}\,e^{\mu} = \mu$$

To find $\mathbb{E}\,(X^2)$, let us consider the factorial expression $\mathbb{E}\,[X(X-1)]$.

$$\mathbb{E}\,[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)\,\frac{e^{-\mu}\,\mu^x}{x!} = \sum_{x=2}^{\infty} x(x-1)\,\frac{\mu^2\,e^{-\mu}\,\mu^{x-2}}{x(x-1)(x-2)!}$$

$$= \mu^2\,e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!} = \mu^2\,e^{-\mu}\,e^{\mu} = \mu^2$$

Therefore, $\mathbb{E}\,[X(X-1)] = \mathbb{E}\,(X^2) - \mathbb{E}\,(X) = \mu^2$. Now we can solve for $\mathbb{E}\,(X^2)$ which is $\mathbb{E}\,(X^2) = \mathbb{E}\,[X(X-1)] + \mathbb{E}\,(X) = \mu^2 + \mu$.
Thus,

$$V(X) = \mathbb{E}\,(X^2) - [\mathbb{E}\,(X)]^2 = \mu^2 + \mu - \mu^2 = \mu.$$

$\square$

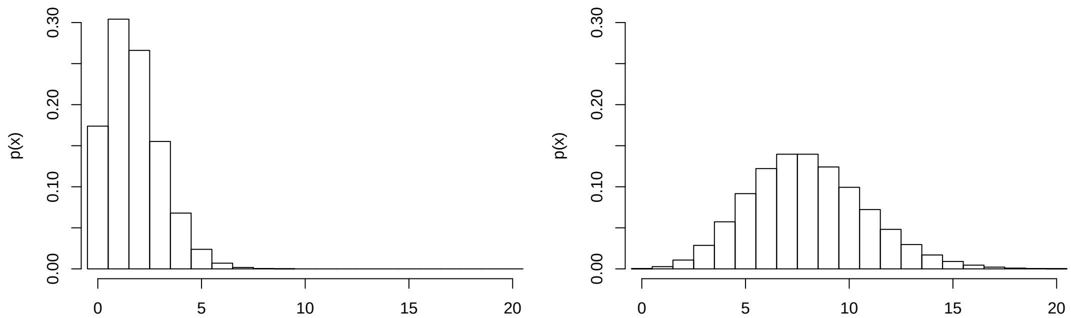


Figure 3.5: Poisson PMF: left, with $\mu = 1.75$; right, with $\mu = 8$

Example 3.14.
During World War II, the Nazis bombed London using V-2 missiles. To study the locations where missiles fell, the British divided the central area of London into 576 half-kilometer squares.[i] The following is the distribution of counts per square

Number of missiles in a square	Number of squares	Expected (Poisson) Number of squares
0	229	227.5
1	211	211.3
2	93	98.1
3	35	30.4
4	7	7.1
5 and over	1	1.6
Total	576	576.0

Are the counts suggestive of Poisson distribution?

Solution. The total number of missiles is $1(211) + 2(93) + 3(35) + 4(7) + 5(1) = 535$ and the average number per square, $\mu = 0.9288$. If the Poisson distribution holds, then the expected number of 0 squares (out of 576) will be

$$576 \times P(X = 0) = 576 \times \frac{e^{-0.9288}\, 0.9288^0}{0!} = 227.5$$

The same way, fill out the rest of the expected counts column. As you can see, the data match the Poisson model very closely!

Poisson distribution is often mentioned as a distribution of *spatial randomness*. As a result, British command were able to conclude that the missiles were unguided. □

Using the CDF

Knowledge of CDF (cumulative distribution function) is useful for calculating probabilities of the type $P(a \le X \le b)$. In fact,

$$P(a < X \le b) = F_X(b) - F_X(a) \tag{3.2}$$

(you have to carefully watch strict and non-strict inequalities). We might use CDF tables (see Appendix) to calculate such probabilities. Nowadays, CDF's of popular distributions are built into various software packages.

Example 3.15.
During a laboratory experiment, the average number of radioactive particles passing through a counter in one millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond? What is the probability of **at least** 6 particles?

Solution. Using the Poisson distribution with $x = 6$ and $\mu = 4$, we get

$$pois(6; 4) = \frac{e^{-4}4^6}{6!} = 0.1042$$

Alternatively, using the CDF, $P(X = 6) = P(5 < X \leq 6) = F(6) - F(5)$. Using the Poisson table, $P(X = 6) = 0.8893 - 0.7851 = 0.1042$.

To find $P(X \geq 6)$, use $P(5 < X \leq \infty) = F(\infty) - F(5) = 1 - 0.7851 = 0.2149$ □

Poisson approximation for Binomial

Poisson distribution was originally derived as a limit of Binomial when $n \to \infty$ while $p = \mu/n$, with fixed μ. We can use this fact to estimate Binomial probabilities for large n and small p.

Example 3.16.
At a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and the accidents are independent of each other. For a given period of 400 days, what is the probability that
(a) there will be an accident on only one day?
(b) there are at most two days with an accident?

Solution. Let X be a binomial random variable with $n = 400$ and $p = 0.005$. Thus $\mu = np = (400)(0.005) = 2$. Using the Poisson approximation,

a) $P(X = 1) = \frac{e^{-2}\,2^1}{1!} = 0.271$

b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-2}\,2^0}{0!} + \frac{e^{-2}\,2^1}{1!} + \frac{e^{-2}\,2^2}{2!}$
$= 0.1353 + 0.271 + 0.271 = 0.6766$

□

Exercises

3.42.
Number of cable breakages in a year is known to have Poisson distribution with $\mu = 0.32$.

a) Find the mean and standard deviation of the number of cable breakages in a year.

b) According to Chebyshev's inequality, what is the upper bound for $P(X \geq 2)$?

c) What is the exact probability $P(X \geq 2)$, based on Poisson model?

3.43.
Bolted assemblies on a hull of spacecraft may become loose with probabiity 0.005. There are 96 such assemblies on board. Assuming that assemblies behave statistically independently, find the probability that there is at most one loose assembly on board.

3.44.
At a barber shop, expected number of customers per day is 8. What is a probability that, on a given day, between 7 and 9 customers (inclusive) show up? At least 3 customers?

3.45.
Poisson distribution can be derived by considering Binomial with n large and p small. Compare computationally

a) Binomial with $n = 20$, $p = 0.05$: find $P(X = 0), P(X = 1)$ and $P(X = 2)$.

b) Repeat for Binomial with $n = 200$, $p = 0.005$

c) Poisson with $\mu = np = 1$ [Note that μ matches the expected value for both (a) and (b).]

d) Compare the standard deviations for distributions in (a)-(c)

3.46.
An airline finds that 5% of the people making reservations on a certain flight will not show up for the flight. If the airline sells 160 tickets for a flight with 155 seats, what is the probability that the flight ends up overbooked, i.e. more that 155 people will show up? [**Hint:** Use the Poisson approximation for the number of people who will **not** show up.]

3.47.
A region experiences, on average, 7.5 earthquakes (magnitude 5 or higher), per year. Assuming Poisson distribution, find the probability that

a) between 5 and 9 earthquakes will happen in a year;

b) at least one earthquake will happen in a given **month**.

c) Find the mean and standard deviation of the number of earthquakes per year.

3.48.
A plumbing company estimates to get the average of 60 service calls per week. Assuming Poisson distribution, find the probability that, in a given week

a) it gets exactly 60 service calls;

b) it gets between 55 and 59 service calls.

3.49.
A credit card company estimates that, on average, 0.18% of all its internet transactions are fraudulent. Out of 1000 transactions,

a) find the mean and standard deviation of the number of fraudulent transcations,

b) approximate the probability that at least one transaction will be fraudulent,

c) approximate the probability that 3 or less transactions will be fraudulent.

3.8 Hypergeometric distribution

Consider the **Hypergeometric experiment**, that is, one that possesses the following two properties:

a) A random sample of size n is selected *without replacement* from N items.

b) Of the N items overall, k may be classified as successes and $N - k$ are classified as failures.

We will be interested, as before, in the number of successes X, but now the probability of success is not constant (why?).

Theorem 3.7.

The PMF of the hypergeometric random variable X, the number of successes in a random sample of size n selected from N items of which k are labeled success and $N-k$ labeled failure, is

$$hg(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, ..., \min(n, k)$$

The mean and variance of the hypergeometric distribution are $\mu = n\frac{k}{N}$ and $\sigma^2 = n\left(\frac{k}{N}\right)\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$

We have already seen such a random variable: see Example 3.2. Here are some more examples.

Example 3.17.
Lots of 40 components each are called unacceptable if they contain as many as 3 defectives or more. The procedure for sampling the lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

Solution. Using the above distribution with $n = 5, N = 40, k = 3$ and $x = 1$, we can find the probability of obtaining one defective to be

$$hg(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011$$ □

Example 3.18.
A shipment of 20 tape recorders contains 5 that are defective. If 10 of them are randomly chosen for inspection, what is the probability that 2 of the 10 will be defective?

Solution. Subsituting $x = 2, n = 10, k = 5$, and $N = 20$ into the formula, we get
$P(X=2) = \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}} = \frac{10(6435)}{184756} = 0.348$ □

Note that, if we were sampling *with replacement*, we would have Binomial distribution (why?) with $p = k/N$. In fact, if N is much larger than n, then the difference between Binomial and Hypergeometric distribution becomes small.

Exercises

3.50.
Out of 10 construction facilities, 4 are in-state and 6 are out of state. Three facilities are earmarked as test sites for a new technology. What is the probability that 2 out of 3 are out of state?

3.51.
A box contains 8 diodes, among them 3 are of new design. If 4 diodes are picked randomly for a circuit, what is the probability that at least one is of new design?

3.52.
There are 25 schools in a district, 10 of which are performing below standard. Five schools are selected at random for an in-depth study. Find:

a) Probability that in your sample, no schools perform below standard.

b) Probability of selecting at least one that performs below standard.

c) The mean and variance for the number of the schools that perform below standard.

3.53.
A small division, consisting of 6 women and 4 men, picks "employee of the month" for 3 months in a row. Suppose that, in fact, a random person is picked each month. Let X be the number of times a woman was picked. Calculate the distribution of X (make a table with all possible values), for the cases

a) No repetitions are allowed.

b) Repetitions are allowed (the same person can be picked again and again).

c) Compare the results.

3.54.
A jar contains 50 red marbles and 30 blue marbles. Four marbles were selected at random. Find the probability to obtain at least 3 red marbles, if the sampling was

a) without replacement;

b) with replacement.

c) Compare the results.

3.9 Moment generating function

We saw in an earlier section that, if $g(Y)$ is a function of a random variable Y with PMF $p(y)$, then

$$\mathbb{E}\left[g(Y)\right] = \sum_y g(y)p(y)$$

The expected value of the exponential function e^{tY} is especially important.

Definition 3.9. Moment generating function

The moment generating function (MGF) of a random variable Y is

$$M(t) = M_Y(t) = \mathbb{E}\left(e^{tY}\right) = \sum_y e^{ty}p(y)$$

The expected values of powers of random variables are often called *moments*. For example, $\mathbb{E}\left(Y\right)$ is the first moment of Y, and $\mathbb{E}\left(Y^2\right)$ is the second moment of Y. When

$M(t)$ exists, it is differentiable in a neighborhood of $t = 0$, and the derivatives may be taken inside the expectation. Thus,

$$M'(t) = \frac{dM(t)}{dt} = \frac{d}{dt}\,\mathbb{E}\,[e^{tY}] = \mathbb{E}\left[\frac{d}{dt}\,e^{tY}\right] = \mathbb{E}\,[Y\,e^{tY}]$$

Now if we set $t = 0$, we have $M'(0) = \mathbb{E}\,Y$. Going on the second derivative,

$$M''(t) = \mathbb{E}\,[Y^2 e^{tY}]$$

and hence $M''(0) = \mathbb{E}\,(Y^2)$. In general, $M^{(k)}(0) = \mathbb{E}\,(Y^k)$.

Theorem 3.8. Properties of MGF's

a) Uniqueness: Let X and Y be two random variables with moment generating functions $M_X(t)$ and $M_Y(t)$, respectively, If $M_X(t) = M_Y(t)$ for all values of t, in some neighborhood of 0, then X and Y have the same probability distribution.

b) $M_{X+b}(t) = e^{bt} M_X(t)$.

c) $M_{aX}(t) = M_X(at)$

d) If $X_1, X_2, \ldots, X_n$ are independent random variables with moment generating functions $M_1(t), M_2(t), \ldots, M_n(t)$, respectively, and $Y = X_1 + X_2 + \cdots + X_n$, then
$$M_Y(t) = M_1(t) \times M_2(t) \times \cdots \times M_n(t).$$

Example 3.19.
Evaluate the moment generating function for the geometric distribution

Solution. From definition,

$$M(t) = \sum_{x=1}^{\infty} e^{tx} p q^{x-1} = \frac{p}{q} \sum_{x=1}^{\infty} (qe^t)^x$$

On the right, we have an infinite geometric series with first term qe^t and the ratio qe^t. Its sum is $\displaystyle\sum_{x=1}^{\infty} (qe^t)^x = \frac{qe^t}{1 - qe^t}$. We obtain

$$M(t) = p\,e^t \left[\frac{1}{1 - qe^t}\right]$$ □

Exercises

3.55.

Find $M_X(t)$ for random variables X given by

a) $p(x) = 1/3,\ x = -1, 0, 1$

b) $p(x) = \left(\frac{1}{2}\right)^{x+1},\ x = 0, 1, 2, \ldots$

c) $p(x) = \frac{1}{8}\binom{3}{x}, \ x = 0, 1, 2, 3$

3.56.

a) Find the MGF of the Bernoulli distribution.

b) Apply the property (d) of Theorem 3.8 to calculate the MGF of the Binomial distribution. [Hint: Binomial random variable Y with parameters n, p can be represented as $Y = X_1 + X_2 + ... + X_n$, where X's are independent and each has Bernoulli distribution with parameter p.]

3.57.
Apply the property (d) of Theorem 3.8 and Example 3.19 to calculate the MGF of Negative Binomial distribution.

3.58.
Use the derivatives of MGF to calculate the mean and variance of geometric distribution.

3.59.
Suppose that MGF of a random variable X was found equal to

$$M(t) = \frac{1}{1 - t^2}$$

Using the properties of MGF, find $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$.

3.60. $\star$

a) Compute the MGF of Poisson distribution.

b) Using the property (d) of Theorem 3.8, describe the distribution of a sum of two independent Poissons, one with mean μ_1 and another with mean μ_2.

Chapter 4

Continuous probability distributions

4.1 Continuous random variables and their probability distributions

All of the random variables discussed previously were discrete, meaning they can take only a finite (or, at most, countable) number of values. However, many of the random variables seen in practice have more than a countable collection of possible values. For example, the metal content of ore samples may run from 0.10 to 0.80. Such random variables can take any value in an interval of real numbers. Since the random variables of this type have a continuum of possible values, they are called **continuous random variables**.[1]

Definition 4.1. Density (PDF)

The function $f(x)$ is a **probability density function** (PDF) for the continuous random variable X, defined over the set of real numbers $\mathcal{R}$, if

a) $f(x) \geq 0, \text{ for all } x$

b) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$

c) $\displaystyle P(a \leq X \leq b) = \int_a^b f(x)\,dx.$

What does this actually mean? Since continuous probability functions are defined for an infinite number of points over a continuous interval, the probability at a single point is always zero. Probabilities are measured over intervals, not single points. That is, the area under the curve between

[1]Even though the tools we will use to describe continuous RV's are different from the tools we use for discrete ones, practically there is not an enormous gulf between them. For example, a physical measurement of, say, wavelength may be continuous. However, when the measurements are recorded (either on paper or in computer memory), they will take a finite number of values. The number of values will increase if we keep more decimals in the recorded quantity. With rounding we can **discretize** the problem, that is, reduce a continuous problem to a discrete one, whose solution will hopefully be "close enough" to the continuous one. In order to see if we have discretized a problem in a right way, we still need to know something about the nature of continuous random variables.

two distinct points defines the probability for that interval. This means that the height of the probability function can in fact be greater than one. The property that the integral must equal one is equivalent to the property for discrete distributions that the sum of all the probabilities must equal one.

Probability mass function (PMF) vs. Probability Density Function (PDF)

Discrete probability functions are referred to as probability mass functions and continuous probability functions are referred to as probability density functions. The term probability functions covers both discrete and continuous distributions. When we are referring to probability functions in generic terms, we may use the term probability density functions to mean both discrete and continuous probability functions.

Example 4.1.
Suppose that the error in the reaction temperature, in $^\circ C$, for a controlled laboratory experiment is a continuous random variable X having the density

$$f(x) = \begin{cases} \frac{x^2}{3} & \text{for } -1 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Verify condition (b) of Definition 4.1.
(b) Find $P(0 < X < 1)$.

Solution. (a) $\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$

(b) $P(0 < X < 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9}|_0^1 = \frac{1}{9}$. □

Definition 4.2. CDF

The **cumulative distribution function (CDF)** $F(x)$ of a continuous random variable X, with density function $f(x)$, is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \tag{4.1}$$

As an immediate consequence of equation (4.1) one can write these two results:

(a) $P(a < X \leq b) = F(b) - F(a)$ [2]

(b) $f(x) = F'(x)$, if the derivative exists.

Example 4.2.
For the density function of Example 4.1, find $F(x)$ and use it to evaluate $P(0 < X < 1)$.

Solution. For $-1 < x < 2$, we have

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} \frac{t^2}{3} dt = \left.\frac{t^3}{9}\right|_{-1}^{x} = \frac{x^3+1}{9},$$

[2]Note that the same relation holds for discrete RV's but in the continuous case $P(a \leq X \leq b)$, $P(a < X \leq b)$ and $P(a < X < b)$ are all the same. Why?

Therefore,

$$F(x) = \begin{cases} 0 & x \leq -1 \\ \frac{x^3+1}{9} & \text{for } -1 < x < 2 \\ 1 & x \geq 2. \end{cases}$$

Now, $P(0 < X < 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$, which agrees with the result obtained using the density function in Example 4.1. □

Example 4.3.
The time X in months until failure of a certain product has the PDF

$$f(x) = \begin{cases} \frac{3x^2}{64} \exp\left(-\frac{x^3}{64}\right) & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find $F(x)$ and evaluate $P(2.84 < X < 5.28)$

Solution. $F(x) = 1 - \exp\left(-\dfrac{x^3}{64}\right)$, and $P(2.84 \leq X \leq 5.28) = 0.5988$ □

Example 4.4.
The life length of batteries X (in hundreds of hours) has the density

$$f(x) = \begin{cases} \frac{1}{2}\, e^{-\frac{x}{2}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability that the life of a battery of this type is less than 200 or greater than 400 hours.

Solution. Let A denote the event that X is less than 2, and let B denote the event that X is greater than 4. Then

$$\begin{aligned} P(A \cup B) = P(A) + P(B)\ (\texttt{why?}) &= \int_0^2 \frac{1}{2}\, e^{-\frac{x}{2}}\, dx + \int_4^\infty \frac{1}{2}\, e^{-\frac{x}{2}}\, dx \\ &= (1 - e^{-1}) + (e^{-2}) = 1 - 0.368 + 0.135 = 0.767 \end{aligned}$$

□

Example 4.5.
Refer to Example 4.4. Find the probability that a battery of this type lasts more than 300 hours, given than it already has been in use for more than 200 hours.

Solution. We are interested in $P(X > 3|X > 2)$; and by the definition of conditional probability,

$$P(X > 3|X > 2) = \frac{P(X > 3, X > 2)}{P(X > 2)} = \frac{P(X > 3)}{P(X > 2)}$$

because the intersection of the events $(X > 3)$ and $(X > 2)$ is the event $(X > 3)$. Now

$$\frac{P(X > 3)}{P(X > 2)} = \frac{\displaystyle\int_3^\infty \frac{1}{2}\, e^{-x/2} dx}{\displaystyle\int_2^\infty \frac{1}{2}\, e^{-x/2} dx} = \frac{e^{-\frac{3}{2}}}{e^{-1}} = e^{-\frac{1}{2}} = 0.606$$

□

Example 4.6.
For each of the following functions,
(i) find the constant c so that $f(x)$ is a PDF of a random variable X, and
(ii) find the distribution function $F(x)$.

a) $$f(x) = \begin{cases} \frac{x^3}{4} & \text{for } 0 < x < c \\ 0 & \text{elsewhere} \end{cases}$$

b) $$f(x) = \begin{cases} \frac{3}{16}x^2 & \text{for } -c < x < c \\ 0 & \text{elsewhere} \end{cases}$$

c) $$f(x) = \begin{cases} 4x^c & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

d) $$f(x) = \begin{cases} \frac{c}{x^{3/4}} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Answers. a) $c = 2$ and $F(x) = \frac{x^4}{16}$, $0 < x < 2$.

b) $c = 2$ and $F(x) = \frac{x^3}{16} + \frac{1}{2}$, $-2 < x < 2$.

c) $c = 3$ and $F(x) = x^4$, $0 < x < 1$.

d) $c = \frac{1}{4}$ and $F(x) = x^{1/4}$, $0 < x < 1$.

□

Exercises

4.1.
The lifetime of a vacuum cleaner, in years, is described by

$$f(x) = \begin{cases} x/4 & \text{for } 0 < x < 2 \\ (4-x)/4 & \text{for } 2 \le x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability that the lifetime of a vacuum cleaner is
(a) less than 2.5 years
(b) between 1 and 3 years.

4.2.
The demand for an antibiotic from a local pharmacy is given by a random variable X with CDF

$$F(x) = \begin{cases} 1 - \frac{2500}{(x+50)^2} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

a) Find the probability that the demand is at least 50 doses
b) Find the probability that the demand is between 40 and 80 doses
c) Find the density function of X.

4.3.
The proportion of warehouse items claimed within 1 month is given by a random variable X with density

$$f(x) = \begin{cases} c(x+1) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find c to make this a legitimate density function.
(b) Find the probability that the proportion of items claimed will be between 0.5 and 0.7.

4.4.
The waiting time, in minutes, between customers coming into a store is a continuous random variable with CDF

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \exp(-x/2) & \text{for } x \geq 0 \end{cases}$$

Find the probability of waiting less than 1.5 minutes between successive customers

a) using the cumulative distribution of X;

b) using the probability density function of X (first, you have to find it).

4.5.
A continuous random variable X that has a density function given by

$$f(x) = \begin{cases} \frac{1}{5} & \text{for } -1 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

a) Show that the area under the curve is equal to 1.

b) Find $P(0 < X < 2)$.

c) Find c such that $P(X < c) = 1/2$. [This is called a *median* of the distribution.]

4.6.
A continuous random variable X that has a density function given by

$$f(x) = \frac{c}{1+x^2} \quad \text{for } -\infty < x < \infty$$

a) Find the constant c to make this a legitimate density function.

b) Find $P(-3 < X < 3)$.

c) Calculate the CDF of X. Verify that $\lim_{x\to\infty} F(x) = 1$

4.2 Expected values of continuous random variables

The expected values of continuous RV's are obtained using formulas similar to those of discrete ones. However, the summation is now replaced by integration.

Definition 4.3. Expected value

The **expected value** or **mean** of a continuous random variable X that has a probability density function $f(x)$ is given by

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x) dx$$

Theorem 4.1. Expected value of a function

If X is a continuous random variable with probability density function $f(x)$, and if $g(x)$ is any real-valued function of X, then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

Definition 4.4. Variance

Let X be a random variable with probability density function $f(x)$ and mean $\mathbb{E}\, X = \mu$. The variance of X is

$$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \mathbb{E}(X^2) - \mu^2$$

Definition 4.5. Median

The *median* of a probability distribution is defined as solution m to the equation (F is the CDF)

$$F(m) = 0.5$$

Example 4.7.
Suppose that X has density function given by

$$f(x) = \begin{cases} 3x^2 & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find the mean and variance of X
(b) Find mean and variance of $u(X) = 4X + 3$.
(c) Find the median of X

Solution. (a) From the above definitions,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_0^1 x\,(3x^2) dx = \int_0^1 3x^3 dx = 3\frac{x^4}{4}\bigg]_0^1 = \frac{3}{4} = 0.75$$

$$\text{Now,} \qquad \mathbb{E}(X^2) = \int_0^1 x^2\,(3x^2) dx = \int_0^1 3x^4 dx = 3\frac{x^5}{5}\bigg]_0^1 = \frac{3}{5} = 0.6$$

Hence, $\sigma^2 = \mathbb{E}(X^2) - \mu^2 = 0.6 - (0.75)^2 = 0.6 - 0.5625 = 0.0375$

(b) From Theorem 3.2, we get

$$\mathbb{E}(u(X)) = \mathbb{E}(4X+3) = 4\mathbb{E}(X) + 3 = 4(0.75) + 3 = 6$$

and
$$V(u(X)) = V(4X+3) = 16[V(X)] + 0 \text{ (why?) } = 16(0.0375) = 0.6$$

(c)
$$F(x) = \int_0^x 3y^2\, dy = y^3\ ]_0^x = x^3.$$

Now, solve $F(m) = 0.5 = m^3$, hence $m = 0.5^{1/3} = 0.7937$

□

Note that, according to the Theorem 3.2, $\mathbb{E}\, g(X) = g(\mathbb{E}\, X)$ when g is a linear function, that is, $g(x) = a + bx$. What happens when g is not linear?

Example 4.8.
Suppose that X has density function given by

$$f(x) = \begin{cases} (x+1)/2 & \text{for } -1 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find the expected value of $g(X) = X^3$
(b) Is it true that $\mathbb{E}(X^3) = (\mathbb{E}\, X)^3$?

Solution. (a) $\mathbb{E}(X^3) = \int_{-1}^1 x^3 f(x)\, dx = \int_{-1}^1 x^3(x+1)/2\, dx = \frac{1}{2}\int_{-1}^1 (x^4 + x^3)\, dx = 1/5$

(b) Since $\mathbb{E}\, X = \int_{-1}^1 x(x+1)/2\, dx = 1/3$, then $\mathbb{E}(X^3) \neq (\mathbb{E}\, X)^3$. □

Exercises

4.7.
For the density described in Exercise 4.3, find the mean and standard deviation of X.

4.8.
For a random variable X with the density

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

a) Find the mean of X

b) Find $V(X)$

c) Find $\mathbb{E}(X^4)$

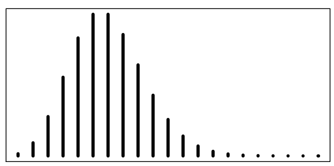
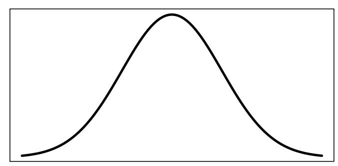
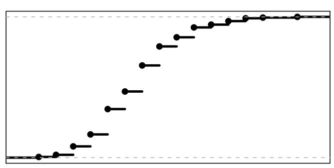
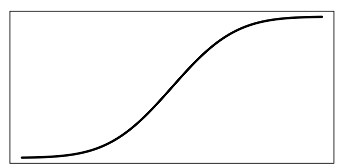
4.9.
For a random variable X with the density

$$f(x) = \begin{cases} 2 - x & \text{for } 0 < x < c \\ 0 & \text{elsewhere} \end{cases}$$

a) Find c that makes f a legitimate density function

b) Find the mean of X

Discrete and Continuous random variables

	Discrete	Continuous
Probability	Probability function  $p(x) = P(X = x)$	Density  $f(x) = \frac{d}{dx} P(X \le x) = F'(x)$ $P(X = x)$ is 0 for any x
CDF $F(x) = P(X \le x)$ $P(a < X \le b) =$ $= F(b) - F(a)$	Is a ladder function 	Is continuous 
Mean $\mathbb{E}\,(X) = \mu_X$	$\sum xp(x)$	$\int xf(x)\,dx$
Mean of a function $\mathbb{E}\,g(X)$	$\sum g(x)p(x)$	$\int g(x)f(x)\,dx$
Variance $\sigma_X^2 = \mathbb{E}\,(X^2) - \mu^2$	$\sum (x-\mu)^2 p(x)$	$\int (x-\mu)^2 f(x)\,dx$

4.10.
For the density described in Exercise 4.1,

a) find the mean and standard deviation of X;

b) Use Chebyshev inequality to estimate the probability that X is between 1 and 3 years. Compare with the answer to Exercise 4.1.

4.11.
For the density described in Exercise 4.5,

a) find the mean and standard deviation of X.

b) Discretize the problem by assigning equal probabilities for each integer between -1 and 4. Re-calculate the mean and standard deviation and compare the results to (a).

4.12.
For a random variable X with the CDF

$$F(x) = \begin{cases} x^3/8 & \text{for } 0 < x < 2 \\ 0, & x \leq 0 \\ 1, & x \geq 2 \end{cases}$$

a) Find the mean of X

b) Find $V(X)$

c) Find the median of X and compare it to the mean

4.13.
The waiting time X, in minutes, between successive customers coming into a store is given by

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 2\exp(-2x) & \text{for } x \geq 0 \end{cases}$$

a) Find the average time between customers

b) Find $\mathbb{E}(e^X)$

4.14.
The PDF for a random variable X is given by

$$f(x) = \begin{cases} 12x^2(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

a) Find the mean of X

b) Find the median of X

c) Find the *mode* of X, that is the point x where $f(x)$ is the highest.

d) Is the function $f(x)$ symmetric? Explain. Sketch $f(x)$ and mark the mean, median, and mode.

4.3 Uniform distribution

One of the simplest continuous distributions is the continuous uniform distribution. This distribution is characterized by a density function that is flat and thus the probability is uniform in a finite interval, say $[a, b]$. The density function of the continuous uniform random variable X on the interval $[a, b]$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

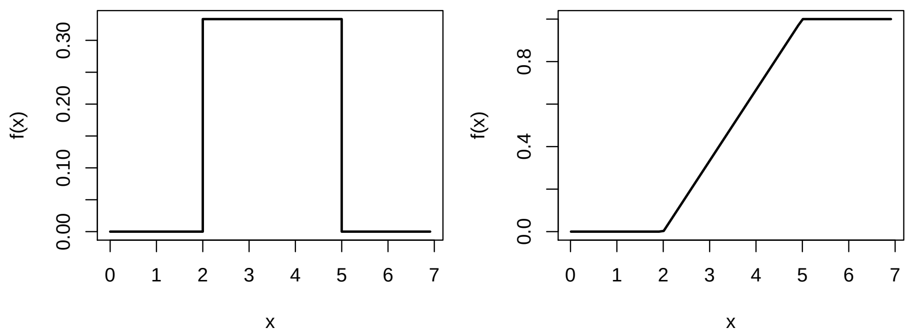


Figure 4.1: Left: uniform density, right: uniform CDF, $a = 2$, $b = 5$

The CDF of a uniformly distributed X is given by

$$F(x) = \int_a^x \frac{1}{b-a}\,dt = \frac{x-a}{b-a}, \quad a \le x \le b$$

The mean and variance of the uniform distribution are

$$\mu = \frac{b+a}{2} \quad \text{and} \quad \sigma^2 = \frac{(b-a)^2}{12}.$$

Example 4.9.
Suppose that a large conference room for a certain company can be reserved for no more than 4 hours. However, the use of the conference room is such that both long and short conferences occur quite often. In fact, it can be assumed that length X of a conference has a uniform distribution on the interval $[0, 4]$.

a) What is the probability density function of X?

b) What is the probability that any given conference lasts at least 3 hours?

Solution. (a) The appropriate density function for the uniformly distributed random variable X in this situation is

$$f(x) = \begin{cases} 1/4 & \text{for } 0 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

(b)

$$P(X \ge 3) = \int_3^4 \frac{1}{4}\,dx = \frac{1}{4}.$$

□

Example 4.10.
The failure of a circuit board interrupts work by a computing system until a new board is delivered. Delivery time X is uniformly distributed over the interval of at least one but no more than four days. The cost C of this failure and interruption consists of a fixed cost C_0 for the new part and a cost that increases proportionally to X^2, so that

$$C = C_0 + C_1 X^2$$

(a) Find the probability that the delivery time is two or more days.
(b) Find the expected cost of a single failure, in terms of C_0 and C_1.

Solution. a)
$$f(x) = \begin{cases} \frac{1}{4} & \text{for } 1 \le x \le 5 \\ 0 & \text{elsewhere} \end{cases}$$

Thus,
$$P(X \ge 2) = \int_2^5 \frac{1}{4}\, dx = \frac{1}{4}(5-2) = \frac{3}{4}$$

b) We know that
$$\mathbb{E}(C) = C_0 + C_1 \mathbb{E}(X^2)$$

so it remains for us to find $\mathbb{E}(X^2)$. This value could be found directly from the definition or by using the variance and the fact that $\mathbb{E}(X^2) = V(X) + \mu^2$. Using the latter approach, we find
$$\mathbb{E}(X^2) = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 = \frac{(5-1)^2}{12} + \left(\frac{1+5}{2}\right)^2 = \frac{31}{3}$$

Thus, $\mathbb{E}(C) = C_0 + C_1\left(\frac{31}{3}\right)$. □

Exercises

4.15.
For a digital measuring device, rounding errors have Uniform distribution, between -0.05 and 0.05 mm.

a) Find the probability that the rounding error is between -0.01 and 0.03mm
b) Find the expected value and the standard deviation of the rounding error.
c) Calculate and plot the CDF of the rounding errors.

4.16.
The capacitances of "1mF" (microfarad) capacitors are, in fact, Uniform$[0.95, 1.05]$ mF.

a) What proportion of capacitors are 0.98 mF or above?
b) What proportion of capacitors are within 0.03 of the nominal value?

4.17.
For X having a Uniform$[-1, 4]$ distribution, find the mean and variance. Then, use the formula for variance and a little algebra to find $\mathbb{E}(X^2)$.

4.18.
Suppose the radii of spheres R have a uniform distribution on $[2, 3]$. Find the mean volume. $(V = \frac{4}{3}\pi R^3)$. Find the mean surface area. $(A = 4\pi R^2)$.

4.4 Exponential distribution

Definition 4.6. Exponential distribution

The continuous random variable X has an exponential distribution, with parameter β, if its density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta}\, e^{-\frac{x}{\beta}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The mean and variance of the exponential distribution are

$$\mu = \beta \text{ and } \sigma^2 = \beta^2.$$

The distribution function for the exponential distribution has the simple form:

$$F(t) = P(X \leq t) = \int_0^t \frac{1}{\beta}\, e^{-\frac{x}{\beta}} dx = 1 - e^{-\frac{t}{\beta}} \qquad \text{for } t \geq 0$$

The **failure rate function r(t)** is defined as

$$r(t) = \frac{f(t)}{1 - F(t)}, \qquad t > 0 \tag{4.2}$$

Suppose that X, with density f, is a lifetime of an item. Consider the proportion of items currently alive (at the time t) that will fail in the next time interval $(t, t + \Delta t]$, where Δt is small. Thus, by the conditional probability formula,

$$P\{\text{die in the next } (t, t + \Delta t] \mid \text{ currently alive}\} =$$

$$= \frac{P\{X \in (t, t + \Delta t]\}}{P(X > t)} \approx \frac{f(t)\Delta t}{1 - F(t)} = r(t)\Delta t$$

so the rate at which the items fail is $r(t)$.

For the exponential case,

$$r(t) = \frac{f(t)}{1 - F(t)} = \frac{1/\beta\, e^{-t/\beta}}{e^{-t/\beta}} = \frac{1}{\beta}$$

Note that the failure rate $\lambda = \frac{1}{\beta}$ of an item with exponential lifetime does not depend on the item's age. This is known as the *memoryless property* of exponential distribution. The exponential distribution is the only continuous distribution to have a constant failure rate.

In reliability studies, the mean of a positive-valued distribution, is also called *Mean Time To Fail* or MTTF. So, we have exponential MTTF $= \beta$.

Relationship between Poisson and exponential distributions

Suppose that certain events happen at the rate λ, so that the average (expected) number of events on the interval $[0, t]$ is $\mu = \lambda t$. If we assume that the number of events on $[0, t]$ has Poisson distribution, then the probability of no events up to time t is given by

$$pois(0, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Thus, if the time of first failure is denoted X, then

$$P(X \leq t) = 1 - P(X > t) = 1 - e^{-\lambda t}$$

We see that $P(X \leq t) = F(t)$, the CDF for X, has the form of an exponential CDF. Here, $\lambda = \frac{1}{\beta}$ is again the failure rate. Upon differentiating, we see that the density of X is given by

$$f(t) = \frac{dF(t)}{dt} = \frac{d(1 - e^{-\lambda t})}{dt} = \lambda e^{-\lambda t} = \frac{1}{\beta}\, e^{-t/\beta}$$

and thus X has an exponential distribution.

Some natural phenomena have a constant failure rate (or occurrence rate) property; for example, the arrival rate of cosmic ray alpha particles or Geiger counter ticks. The exponential model works well for interarrival times (while the Poisson distribution describes the total number of events in a given period).

Example 4.11.
A downtime due to equipment failure is estimated to have Exponential distribution with the mean $\beta = 6$ hours. What is the probability that the next downtime will last between 5 and 10 hours?

Solution. $P(5 < X < 10) =$

$= F(10) - F(5) = 1 - \exp(-10/6) - [1 - \exp(-5/6)] = 0.2457$ □

Example 4.12.
The number of calls to the call center has Poisson distribution with parameter $\lambda = 4$ calls per minute. What is the probability that we have to wait more than 20 seconds for the next call?

Solution. The waiting time between calls, X, has exponential distribution with parameter $\beta = 1/\lambda = 1/4$. Then, $P(X > \frac{1}{3}) = 1 - F(\frac{1}{3}) = e^{-4/3} = 0.2636$ □

Exercises

4.19.
Prove another version of the memoryless property of the exponential distribution,

$$P(X > t + s \,|\, X > t) = P(X > s).$$

Thus, an item that is t years old has the same probabilistic properties as a brand-new item. [**Hint:** Use the definition of conditional probability and the expression for exponential CDF.]

4.20.
The 1-hour carbon monoxide concentrations in a big city are found to have an exponential distribution with a mean of 3.6 parts per million (ppm).

(a) Find the probability that a concentration will exceed 9 ppm.

(b) A traffic control policy is trying to reduce the average concentration. Find the new target mean β so that the probability in part (a) will equal 0.01

(c) Find the median of the concentrations from part (a).

4.21.
Customers come to a barber shop as a Poisson process with the frequency of 3 per hour. Suppose Y_1 is the time when first customer comes.

a) Find the expected value and the standard deviation of Y_1

b) Find the probability that the store is idle for at least first 30 minutes after opening.

4.5 The Gamma distribution

The Gamma distribution derives its name from the well-known gamma function, studied in many areas of mathematics. This distribution plays an important role in both queuing theory and reliability problems. Time between arrivals at service facilities, and time to failure of component parts and electrical systems, often are nicely modeled by the Gamma distribution.

Definition 4.7. Gamma function

The **gamma function**, for $\alpha > 0$, is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$$

$\Gamma(k) = (k-1)!$ for integer k.

Definition 4.8. Gamma distribution

The continuous random variable X has a gamma distribution, with *shape parameter* α and *scale parameter* β, if its density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{for } \ x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The mean and variance of the Gamma distribution are

$$\mu = \alpha\beta \quad \text{and} \quad \sigma^2 = \alpha\beta^2.$$

Note: When $\alpha = 1$, the Gamma reduces to the exponential distribution. Another well-known statistical distribution, chi-square, is also a special case of the Gamma.

Uses of the Gamma Distribution Model

a) The gamma is a flexible life distribution model that may offer a good fit to some sets of failure data, or other data where positivity is enforced.

b) The gamma does arise naturally as the time-to-failure distribution for a system with standby exponentially distributed backups. If there are
$n-1$ standby backup units and the system and all backups have exponential lifetimes with mean β, then the total lifetime has a Gamma distribution with $\alpha = n$. Note: when α is a positive integer, the Gamma is sometimes called *Erlang distribution.* The Erlang distribution is used frequently in queuing theory applications.

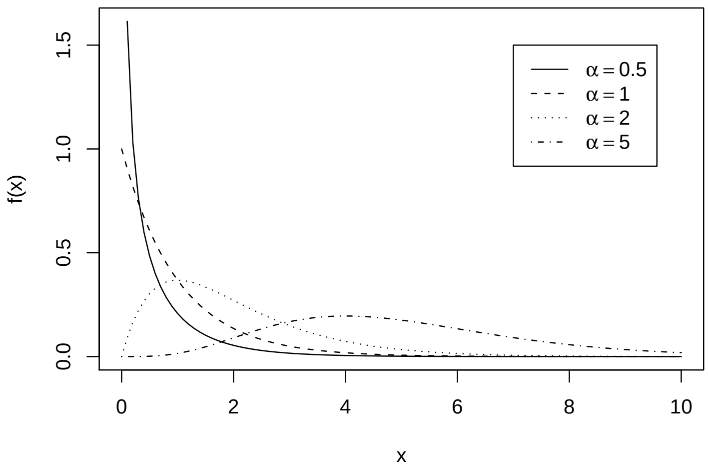


Figure 4.2: Gamma densities, all with $\beta = 1$

c) A simple and often used property of sums of identically distributed, independent gamma random variables will be stated, but not proved, at this point. Suppose that $X_1, X_2, \ldots, X_n$ represent independent gamma random variables with parameters α and β, as just used. If $Y = \sum_{i=1}^{n} X_i$ then Y also has a gamma distribution with parameters $n\alpha$ and β. Thus, we see that $\mathbb{E}(Y) = n\alpha\beta$, and $V(Y) = n\alpha\beta^2$.

Example 4.13.
The total monthly rainfall (in inches) for a particular region can be modeled using Gamma distribution with $\alpha = 2$ and $\beta = 1.6$. Find the mean and variance of the monthly rainfall.

Solution. $\mathbb{E}(X) = \alpha\beta = 3.2$, and variance $V(X) = \alpha\beta^2 = 2(1.6^2) = 5.12$ □

4.5.1 Poisson process

Following our discussion about Exponential distribution, the latter is a good model for the waiting times between randomly occurring events. Adding independent Exponential RV's will result in the *Poisson process.*

The Poisson process was first studied[3] in 1900's when modeling the observation times of radioactive particles recorded by Geiger counter. It consists of the consecutive event times Y_1, Y_2,... such that the interarrival times $X_1 = Y_1$, $X_2 = Y_2 - Y_1$,... have independent Exponential distributions. (The observations start at the time $t = 0$.)

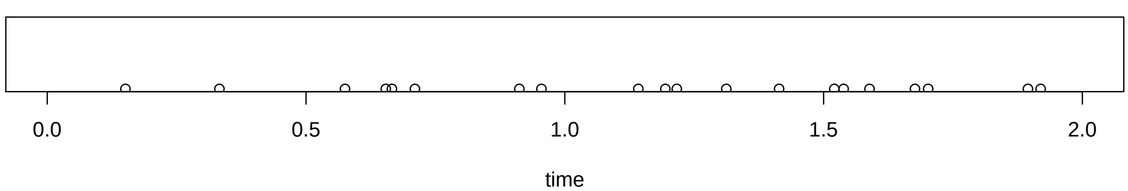


Figure 4.3: Events of a Poisson process

[3]not by Poisson!

From the property (c) above, the kth event time has Gamma distribution with $\alpha = k$. As in Section 4.4, the average number of particles to appear during $[0, t)$ has Poisson distribution with the mean $\mu = \lambda t$ where the **rate** or **intensity** $\lambda = 1/\beta$.

The same way, the number of events on any given interval of time, say, $(t_1, t_2]$ follows the Poisson distribution with the mean $\mu = \lambda(t_2 - t_1)$. Thus, the expected number of events to be observed equals the intensity times the length of the observation period.

Note the units: if the rate λ is measured in events per hour (say), that is, the unit is `hours`$^{-1}$, then the mean time between events is measured in `hours`.

The Gamma CDF (for integer α) can be derived using this relationship. Suppose Y_k is the time to wait for kth event. Then it is Gamma ($\alpha = k$, β) random variable. On one hand, the probability that this event happens before time t is the CDF $F(t)$. On the other hand, this will happen if and only if there is a total of at least k events on the interval $[0, t]$:

$$F(t) = P(Y_k \le t) = P(N(t) \ge k) \tag{4.3}$$

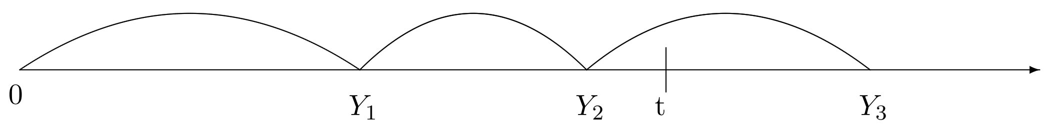


Figure 4.4: Illustration of the principle "$Y_k \le t$ if and only if $N(t) \ge k$", here $k = 2$.

Here, $N(t)$ is the number of events on the $[0, t]$ interval. According to Poisson process, $N(t)$ has Poisson distribution with the mean $\mu = \lambda t = t/\beta$. Thus,

$$P(Y_k \le t) = P(N \ge k) = 1 - P(N < k) = 1 - \sum_{i=0}^{k-1} e^{-t/\beta} \frac{(t/\beta)^i}{i!} \tag{4.4}$$

This is an interesting link between continuous and discrete distributions! In particular, when $k = 1$, we get back the familiar exponential CDF, $F(t) = 1 - \exp(-t/\beta)$.

Example 4.14.
For the situation in Example 4.13, find the probability that the total monthly rainfall exceeds 5 inches.

Solution. $P(Y > 5) = 1 - F(5) = 1 - (1 - P(N < k)) = P(N < k)$ where $k = \alpha = 2$. Equation 4.4 yields $P(Y > 5) = e^{-5/1.6}(1 + 5/1.6) = 0.181$ □

Exercises

4.22.

Customers come to a barber shop with the frequency of 3 per hour. Suppose Y_4 is the time when 4th customer has come.

a) Find the expected value and the standard deviation of Y_4

b) Find the probability that the 4th customer comes within the 1st hour.

4.23.
A truck has 2 spare tires. Under intense driving conditions, tire blowouts are determined to approximately follow a Poisson process with the intensity of 1.2 per 100 miles. Let X be the total distance the truck can go with 2 spare tires.

a) Find the expected value and the standard deviation of X

b) Find the probability that the truck can go at least 200 miles

4.24.
Differentiate Equation 4.4 for $k = 2$ to show that you indeed will get the Gamma density function with $\alpha = 2$.

4.25.
The time X between successive visits to a repair shop is estimated to have Gamma distribution with $\alpha = 2$ and $\beta = 50$ days.

a) Find the expected value and the standard deviation of X.

b) Find the probability that 80 days pass without a visit.

4.26.
The bicycle sales at a store follow a Poisson process with the rate of 0.1 sales per working hour.

a) Find the probability of having exactly 3 bicycle sales over the course of 30 hours.

b) What is the average time between bicycle sales?

c) Describe the distribution of the time between bicycle sales.

4.27.
The counts of user requests incoming to a server are approximated by a Poisson process with the intensity of 560 per second.

a) Describe the distribution of time between requests

b) Find the probability that, during the next 10 ms (= 0.01 sec), between 4 and 6 requests (inclusive) will arrive.

4.28.
For X having a Gamma distribution with $\alpha = 3.5$ and $\beta = 4$,

a) Can you apply the Poisson process formulas to obtain the CDF of X? Do it or explain why you can't.

b) Calculate $\mathbb{E}\, X^5$

4.6 Normal distribution

The most widely used of all the continuous probability distributions is the *normal distribution* (also known as Gaussian). It serves as a popular model for measurement errors, particle displacements under Brownian motion, stock market fluctuations, human intelligence and many other things. It is also used as an approximation for Binomial (for large n) and Gamma (for large α) distributions.

The normal density follows the well-known symmetric bell-shaped curve. The curve is centered at the mean value μ and its spread is, of course, measured by the standard deviation σ. These two parameters, μ and σ^2, completely determine the shape and center of the normal density function.

Definition 4.9.

The normal random variable X has the PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \qquad \text{for } -\infty < x < \infty$$

It will be denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$

The normal random variable with $\mu = 0$ and $\sigma = 1$ is said to have the *standard normal distribution* and will be called Z. Its density becomes $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$. Direct integration would show that $\mathbb{E}(Z) = 0$ and $V(Z) = 1$.

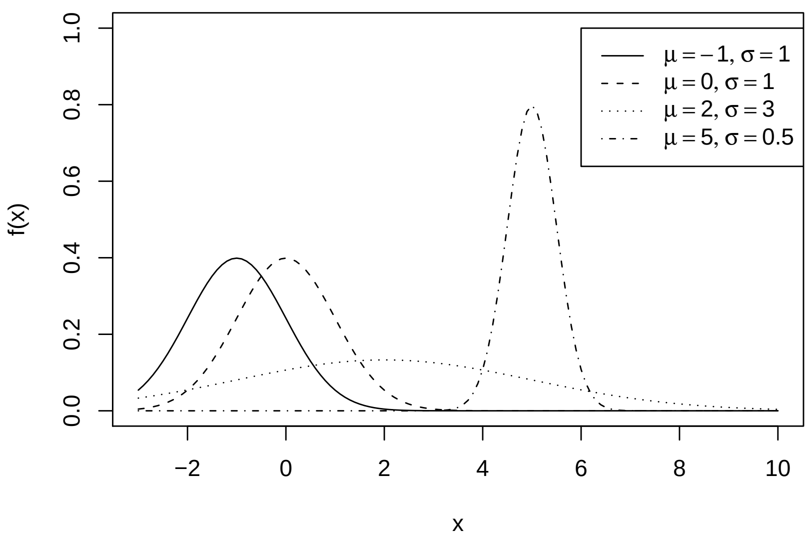


Figure 4.5: Normal densities

Usefulness of Z

We are able to transform the observations of any normal random variable X to a new set of observations of a standard normal random variable Z. This can be done by means of the transformation

$$Z = \frac{X - \mu}{\sigma}.$$

Example 4.15.

Popular (and controversial) IQ scores are scaled to have the mean $\mu = 100$ and standard deviation $\sigma = 15$. Then, if a person has an IQ of 115, it can be transformed into Z-score as $z = (115 - 100)/15 = 1$ and expressed as "one standard deviation above the mean". A lot of standardized test scores (like SAT) follow the same principle. □

The values of the CDF of Z can be obtained from Table A. Namely,

$$F(z) = \begin{cases} 0.5 + \mathtt{TA}(z), & z \geq 0 \\ 0.5 - \mathtt{TA}(|z|), & z < 0 \end{cases}$$

where $\mathtt{TA}(z) = P(0 < Z < z)$ denotes table area of z. The second equation follows from the symmetry of the Z distribution.

Table A allows us to calculate probabilities and percentiles associated with normal random variables, as the direct integration of normal density is not possible.

Example 4.16.
If Z denotes a standard normal variable, find
(a) $P(Z \leq 1)$ (b) $P(Z > 1)$ (c) $P(Z < -1.5)$ (d) $P(-1.5 \leq Z \leq 0.5)$.
(e) Find a number, say z_0, such that $P(0 \leq Z \leq z_0) = 0.49$

Solution. This example provides practice in using Normal probability Table. We see that

a) $P(Z \leq 1) = P(Z \leq 0) + P(0 \leq Z \leq 1) = 0.5 + 0.3413 = 0.8413.$

b) $P(Z > 1) = 0.5 - P(0 \leq Z \leq 1) = 0.5 - 0.3413 = 0.1587$

c) $P(Z < -1.5) = P(Z > 1.5) = 0.5 - P(0 \leq Z \leq 1.5) = 0.5 - 0.4332 = 0.0668.$

d) $P(-1.5 \leq Z \leq 0.5) = P(-1.5 \leq Z \leq 0) + P(0 \leq Z \leq 0.5)$

$= P(0 \leq Z \leq 1.5) + P(0 \leq Z \leq 0.5) = 0.4332 + 0.1915 = 0.6247.$

e) To find the value of z_0 we must look for the given probability of 0.49 on the area side of Normal probability Table. The closest we can come is at 0.4901, which corresponds to a Z value of 2.33. Hence $z_0 = 2.33$.

□

Example 4.17.
For $X \sim \mathcal{N}(50, 10^2)$, find the probability that X is between 45 and 62.

Solution. The Z- values corresponding to X = 45 and X = 62 are

$$Z_1 = \frac{45 - 50}{10} = -0.5 \quad \text{and} \quad Z_2 = \frac{62 - 50}{10} = 1.2.$$

Therefore, $P(45 \leq X \leq 62) = P(-0.5 \leq Z \leq 1.2) = \mathtt{TA}(1.2) + \mathtt{TA}(0.5) = 0.3849 + 0.1915 = 0.5764$ □

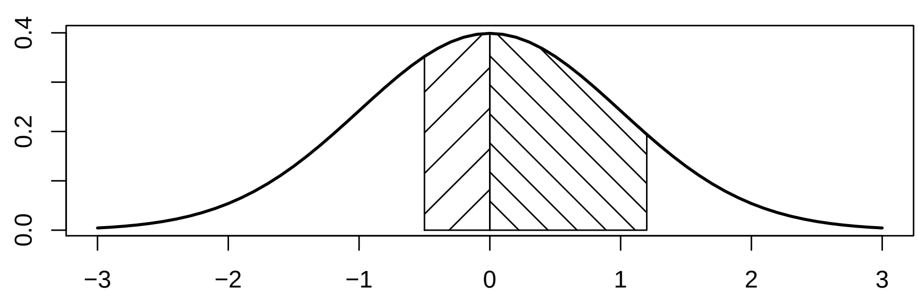


Figure 4.6: Splitting a normal area into two Table Areas

Example 4.18.
Given a random variable X having a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X is greater than 362.

Solution. To find $P(X > 362)$, we need to evaluate the area under the normal curve to the right of $x = 362$. This can be done by transforming $x = 362$ to the corresponding Z-value. We get

$$z = \frac{x-\mu}{\sigma} = \frac{362-300}{50} = 1.24$$

Hence $P(X > 362) = P(Z > 1.24) = P(Z < -1.24) = 0.5 - \texttt{TA}(1.24) = 0.1075$. □

Example 4.19.
A diameter X of a shaft produced has a normal distribution with parameters $\mu = 1.005, \sigma = 0.01$. The shaft will meet specifications if its diameter is between 0.98 and 1.02 cm. Which percent of shafts will not meet specifications?

Solution.

$$1 - P(0.98 < X < 1.02) = 1 - P\left(\frac{0.98-1.005}{0.01} < Z < \frac{1.02-1.005}{0.01}\right)$$

$= 1 - (0.4938 + 0.4332) = 0.0730$ □

4.6.1 Using Normal tables in reverse

Definition 4.10. Percentile

A pth *percentile* of a random variable X is the point q that leaves the area of $p/100\%$ to the left. That is, q is the solution for the equation

$$P(X \le q) = p/100\%$$

For example, the median (introduced in Exercise 4.20) is the 50th percentile of a probability distribution.

We will discuss how to find percentiles of normal distribution. The previous two examples were solved by going first from a value of x to a z-value and then computing the desired area. In the next example we reverse the process and begin with a known area, find the z-value, and then determine x by rearranging the equation $z = \frac{x-\mu}{\sigma}$ to give

$$x = \mu + \sigma z$$

Using the Normal Table calculations, it's straightforward to show the following

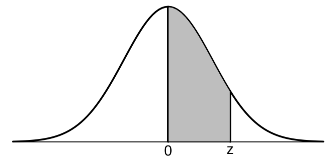
The famous 68% - 95% rule

For a Normal population, 68% of all values lie in the interval $[\mu - \sigma, \mu + \sigma]$, and 95% lie in $[\mu - 2\sigma, \mu + 2\sigma]$.
In addition, 99.7% of the population lies in $[\mu - 3\sigma, \mu + 3\sigma]$.

Example 4.20.
Using the situation in Example 4.19, a diameter X of a shaft had $\mu = 1.005, \sigma = 0.01$. Give an interval that would contain 95% of all diameters.

Solution. The interval is $\mu \pm 2\sigma = 1.005 \pm 2(0.01)$, that is, from 0.985 to 1.025. □

Table A: standard normal probabilities



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Example 4.21.
The SAT Math exam is scaled to have the average of 500 points, and the standard deviation of 100 points. What is the cutoff score for the top 10% of the SAT takers?

Solution. In this example we begin with a known area, find the z-value, and then find x from the formula $x = \mu + \sigma z$. The 90th percentile corresponds to the 90% area under the normal curve to the left of x. Thus, we also require a z-value that leaves 0.9 area to the left and hence, the Table Area of 0.4. From Table A, $P(0 < Z < 1.28) = 0.3997$. Hence

$$x = 500 + 100(1.28) = 628$$

Therefore, the cutoff for the top 10% is 628 points. □

Example 4.22.
Let X = monthly sick leave time have normal distribution with parameters $\mu = 200$ hours and $\sigma = 20$ hours.

a) What percentage of months will have sick leave below 150 hours?
b) What amount of time x_0 should be budgeted for sick leave so that the budget will not be exceeded with 80% probability?

Solution. (a) $P(X < 150) = P(Z < -2.5) = 0.5 - 0.4938 = 0.0062$
(b) $P(X < x_0) = P(Z < z_0) = 0.8$, which leaves a table area for z_0 of 0.3. Thus, $z_0 = 0.84$ and hence $x_0 = 200 + 20(0.84) = 216.8$ hours □

Quantile-Quantile (Q-Q) plots

If X is normal (μ, σ^2) distribution, then

$$X = \mu + \sigma Z$$

and there is a perfect linear relationship between X and Z. This is a graphical method for checking normality.

The details of this method will be considered in Chapter 7.

4.6.2 Normal approximation to Binomial

As another example of using the Normal distribution, consider the Normal approximation to Binomial distribution. This will be also used when discussing sample proportions.

Theorem 4.2. Normal approximation to Binomial

If X is a Binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the random variables

$$Z_n = \frac{X - np}{\sqrt{npq}}$$

approach the standard Normal as n gets large.

We already know one Binomial approximation (by Poisson). It mostly applies when the Binomial distribution in question has a skewed shape, that is, when p is close to 0 or 1. When the shape of Binomial distribution is close to symmetric, the Normal appoximation will work better. Practically, we will require that **both** np and $n(1 - p) \geq 5$.

Example 4.23.

Suppose X is Binomial with parameters $n = 15$, and $p = 0.4$, then $\mu = np = (15)(4) = 6$ and $\sigma^2 = npq = 15(0.4)(0.6) = 3.6$. Suppose we are interested in the probability that X assumes a value from 7 to 9 inclusive, that is, $P(7 \leq X \leq 9)$. The exact probability is given by

$$P(7 \leq X \leq 9) = \sum_{7}^{9} bin(x; 15, 0.4) = 0.1771 + 0.1181 + 0.0612 = 0.3564$$

For Normal approximation we find the area between $x_1 = 6.5$ and $x_2 = 9.5$ using z-values which are

$$z_1 = \frac{x_1 - np}{\sqrt{npq}} = \frac{x_1 - \mu}{\sigma} = \frac{6.5 - 6}{1.897} = 0.26,$$

and

$$z_2 = \frac{9.5 - 6}{1.897} = 1.85$$

The value 0.5 we add or subtract is called *continuity correction*. It arises when we try to approximate a distribution with integer values (here, Binomial) through the use of a continuous distribution (here, Normal). Shown in Fig.4.7, the sum over the discrete set $\{7 \leq X \leq 9\}$ is approximated by the integral of the continuous density from 6.5 to 9.5.

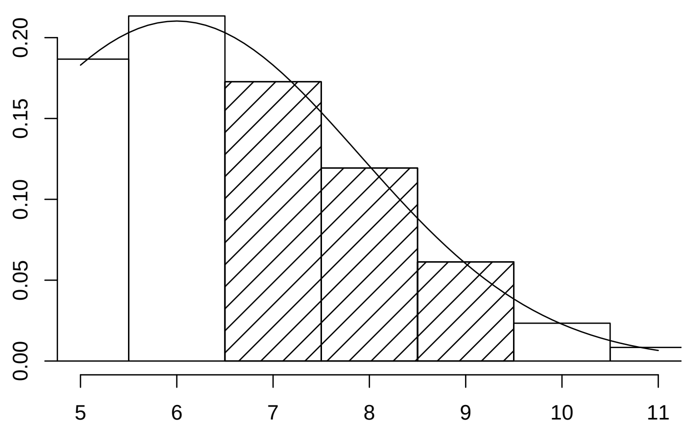


Figure 4.7: continuity correction

Now,

$$P(7 \leq X \leq 9) = P(0.26 < Z < 1.85) = 0.4678 - 0.1026 = 0.3652$$

therefore, the normal approximation provides a value that agrees very closely with the exact value of 0.3564. The degree of accuracy depends on both n and p. The approximation is very good when n is large and if p is not too near 0 or 1. $\square$

Example 4.24.

The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that at most 30 survive?

Solution. Let the binomial variable X represent the number of patients that survive. Since $n = 100$ and $p = 0.4$, we have

$$\mu = np = (100)(0.4) = 40$$

and

$$\sigma^2 = npq = (100)(0.4)(0.6) = 24,$$

also $\sigma = \sqrt{\sigma^2} = 4.899$. To obtain the desired probability, we compute z-value for $x = 30.5$[4]. Thus,

$$z = \frac{x - \mu}{\sigma} = \frac{30.5 - 40}{4.899} = -1.94,$$

and the probability of at most 30 of the 100 patients surviving is $P(X \leq 30) \approx P(Z < -1.94) = 0.5 - 0.4738 = 0.0262$. □

Example 4.25.
A fair coin ($p = 0.5$) is tossed 10,000 times, and the number of Heads X is recorded. What are the values that contain X with 95% certainty?

Solution.
We have $\mu = np = 10,000(0.5) = 5,000$ and $\sigma = \sqrt{10,000(0.5)(1-0.5)} = 50$. We need to find x_1 and x_2 so that $P(x_1 \leq X \leq x_2)$. Since the mean of X is large, we will neglect the continuity correction.

Since we will be working with Normal approximation, let's find z_1 and z_2 such that

$$P(z_1 \leq Z \leq z_2) = 0.95$$

The solution is not unique, but we can choose the values of $z_{1,2}$ that are symmetric about 0. This will mean finding z such that $P(0 < Z < z) = 0.475$. Using Normal tables "in reverse" we will get $z = 1.96$. Thus, $P(-1.96 < Z < 1.96) = 0.95$.

Next, transforming back into X, use the formula $x = \mu + \sigma z$, so

$$x_1 = 5000 + 50(-1.96) = 4902 \quad \text{and} \quad x_2 = 5000 + 50(1.96) = 5098$$

Thus, with a large likelihood, our Heads count will be within 100 of the expected value of 5,000.

This is an example of the famous 68% - 95% rule. □

Exercises

4.29.
Given a standard normal distribution Z, find

a) $P(0 < Z < 1.28)$

b) $P(-2.14 < Z < 0)$

c) $P(Z > -1.28)$

d) $P(-2.3 < Z < -0.75)$

[4] To set this up correctly, remember to include the value of 30 because it's already included in the inequality. For $P(X < 30)$ you would have used 29.5

e) the value z_0 such that $P(Z > z_0) = 0.25$

4.30.
Given a normal distribution with $\mu = 30$ and $\sigma = 6$, find

a) the normal curve area to the right of $x = 17$

b) the normal curve area to the left of $x = 22$

c) the normal curve area between $x = 32$ and $x = 41$

d) the value of x that has 80% of the normal curve area to the left

e) the two values of x that contain the middle 75% of the normal curve area.

4.31.
Given the normally distributed variable X with mean 18 and standard deviation 2.5, find

a) $P(X < 15)$

b) the value of k such that $P(X < k) = 0.2236$

c) the value of k such that $P(X > k) = 0.1814$

d) $P(17 < X < 21)$.

4.32.
A soft drink machine is regulated so that it discharges an average of 200 milliliters (ml) per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 ml,

a) what fraction of the cups will contain more than 224 ml?

b) what is the probability that a cup contains between 191 and 209 milliliters?

c) how many cups will probably overflow if 230 ml cups are used for the next 1000 drinks?

d) below what value do we get the smallest 25% of the drinks?

4.33.
A company pays its employees an average wage of \$15.90 an hour with a standard deviation of \$1.50. If the wages were approximately normally distributed and paid to the nearest cent,

a) What percentage of workers receive wages between \$13.75 and \$16.22 an hour?

b) What is the cutoff value for highest paid 5% of the employees?

4.34.
A solar panel produces, on average, 34.5 kWh (kilowatt-hours) per month, with standard deviation of 2.5 kWh.

a) Find the probability that the panel output will be between 35 and 38 kWh in a month.

b) Find an interval, symmetric about the mean (that is, $[\mu - a, \mu + a]$ for some a), that contains 72% of monthly kWh values.

4.35.
The likelihood that a job application will result in an interview is estimated as 0.1. A grad student has mailed 40 applications. Find the probability that she will get at least 3 interviews,

a) Using the Normal approximation.

b) Using the Poisson approximation.

c) Find the exact probability. Which approximation has worked better? Why?

4.36.
It is estimated that 33% of individuals in a population of Atlantic puffins have a certain recessive gene. If 90 individuals are caught, estimate the probability that there will be between 30 and 40 (inclusive) with the recessive gene.

4.7 Weibull distribution

Earlier we learned that Gamma is a generalization of Exponential distribution (in fact, when $\alpha = 1$ we get Exponential). The Weibull distribution is another such generalization. Like Gamma, it has positive values and is, therefore, suitable as a model of reliability and lifetimes, among other things.

The easiest way to look at the Weibull distribution is through its CDF

$$F(x) = 1 - \exp[-(x/\beta)^\gamma],\ x > 0 \tag{4.5}$$

Note: if $\gamma = 1$ then we get the Exponential distribution. The parameter β has the dimension of time and γ is dimensionless.

By differentiating the CDF, we get the Weibull density.

Definition 4.11. Weibull distribution

The Weibull RV has the density function

$$f(x) = \frac{\gamma x^{\gamma-1}}{\beta^\gamma} \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right], \qquad x > 0$$

and the CDF

$$F(x) = 1 - \exp[-(x/\beta)^\gamma],\ x > 0$$

Its mean is $\mu = \beta\,\Gamma\left(1 + \frac{1}{\gamma}\right)$ and variance is $\sigma^2 = \beta^2\,\Gamma\left(1 + \frac{2}{\gamma}\right) - \mu^2$

The Weibull distribution with $\gamma > 1$ typically has an asymmetric shape with a peak in the middle and the long right "tail". Shapes of Weibull density are shown in Fig. 4.8 for various values of γ.

Regarding the computation of the mean: the Gamma function of non-integer parameter is, generally, not easy to find. Note only that $\Gamma(0.5) = \sqrt{\pi}$, and we can use the recursive relation $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ to compute the Gamma function for $\alpha = 1.5, 2.5$ etc. Also, for large γ, $\Gamma\left(1 + \frac{1}{\gamma}\right) \approx \Gamma(1) = 1$

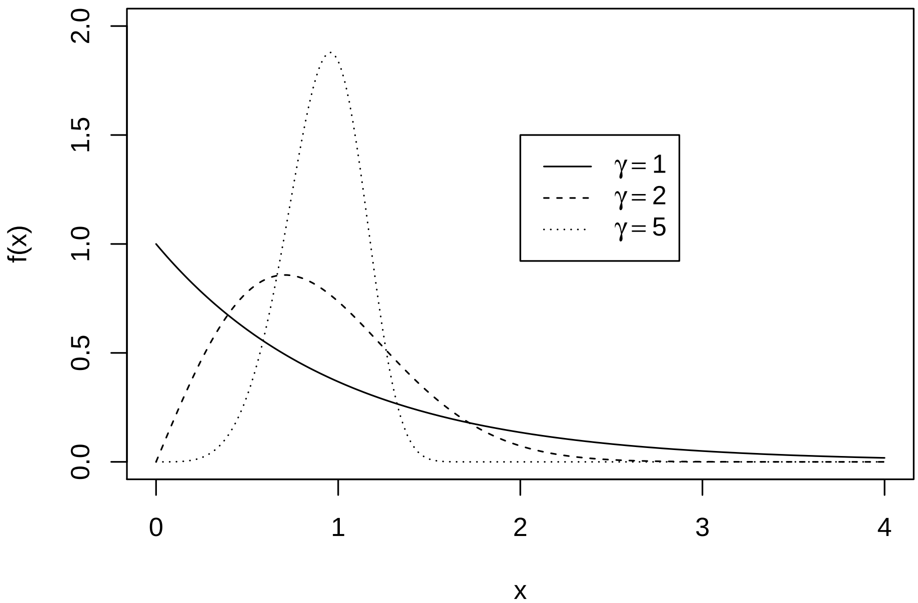


Figure 4.8: Weibull densities, all with $\beta = 1$

Example 4.26.
The duration of subscription to the Internet services is modeled by the Weibull distribution with parameters $\gamma = 2$ and $\beta = 15$ months.

a) Find the average duration.

b) Find the probability that a subscription will last longer than 10 months.

Solution.
(a) $\mu = 15\,\Gamma(1.5) = 15(0.5)\Gamma(0.5) = 7.5\sqrt{\pi} = 13.29$

(b) $P(X > 10) = 1 - F(10) = \exp[-(10/15)^2] = 0.6412$ □

Exercises

4.37.
The time it takes for a server to respond to a request is modeled by the Weibull distribution with $\gamma = 2/3$ and $\beta = 15$ milliseconds.

a) Find the average time to respond.

b) Find the probability that it takes less than 12 milliseconds to respond.

c) Find the 70th percentile of the response times.

4.38.
The lifetimes of refrigerators are assumed to follow Weibull distribution with parameters $\beta = 7$ years and $\gamma = 4$. Find:

a) The proportion of refrigerators with lifetime between 2 and 5 years.

b) If a refrigerator has already worked for 2 years, what is the probability that it will work for at least 3 more years?

4.39.
The tensile strength (in MPa) of titanium rods is estimated to follow Weibull distribution with $\gamma = 10.2$ and $\beta = 415$.

a) Find the critical value c so that only 5% of rods will break before reaching the load c.

b) What proportion of rods will have tesnile strength above 450 MPa?

4.8 Moment generating functions for continuous case

The moment generating function of a continuous random variable X with a pdf of $f(x)$ is given by

$$M(t) = \mathbb{E}\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

when the integral exists. For the exponential distribution, this becomes

$$M(t) = \int_0^{\infty} e^{tx} \frac{1}{\beta} e^{-x/\beta} dx = \int_0^{\infty} e^{-x(1/\beta - t)} \frac{1}{\beta} dx = \frac{1}{(1/\beta - t)\beta} = \frac{1}{1 - \beta t}$$

For properties of MGF's, see Section 3.9

Exercises

4.40.
Calculate MGF for the distribution with a given PDF

a) $f(x) = \dfrac{1}{b} \exp[-(x-a)/b], \quad x > a$

b) $f(x) = \exp[-(x+2)], \quad x > -2$

c) $f(x) = 2x \exp(-2x), \quad x > 0$

d) $f(x) = b \exp[-b(x-7)], \quad x > 7$

e)
$$f(x) = \frac{1}{96} x^3 \exp(-x/2), \quad x > 0$$

f)
$$f(x) = \frac{b^3 x^2}{2} \exp(-bx), \quad x > 0$$

4.41.

a) Calculate the MGF for the Standard Normal distribution.

b) Using properties (b) and (c) of Theorem 3.8, find the MGF for the RV $X \sim \mathcal{N}(\mu, \sigma^2)$

4.42.
It is known that Gamma RV `Gamma`(α, β) for integer $\alpha = n$ is the sum of n independent copies of Exponential RV. Calculate MGF for the Gamma distribution and check the property given by Theorem 3.8(d), p. 63

Chapter 5

Joint probability distributions

5.1 Bivariate and marginal probability distributions

All of the random variables discussed previously were one dimensional, that is, we consider random quantities one at a time. In some situations, however, we may want to record the simultaneous outcomes of several random variables.

Examples:

a) We might measure the amount of precipitate A and volume V of gas released from a controlled chemical experiment, giving rise to a two-dimensional sample space.

b) A physician studies the relationship between weekly exercise amount and resting pulse rate of his patients.

c) An educator studies the relationship between students' grades and time devoted to study.

If X and Y are two discrete random variables, the probability that X equals x while Y equals y is described by $p(x, y) = P(X = x, Y = y)$. That is, the function $p(x, y)$ describes the probability behavior of the pair X, Y. It is not enough to know only how X or Y behave on their own (which is described by their *marginal* probability functions).

Definition 5.1. Joint PMF

The function $p(x, y)$ is a joint probability mass function of the discrete random variables X and Y if

a) $p(x, y) \geq 0$ for all pairs (x, y),

b) $\sum_x \sum_y p(x, y) = 1$,

c) $P(X = x, Y = y) = p(x, y)$.

For any region A in the xy-plane, $P[(X, Y) \text{ belongs to } A] = \displaystyle\sum_{(x,y)\in A} p(x, y)$.

Definition 5.2. Marginal PMF

The marginal probability functions of X and Y respectively are given by

$$p_X(x) = \sum_y p(x,y) \quad \text{and} \quad p_Y(y) = \sum_x p(x,y)$$

Example 5.1.
If two dice are rolled independently, then the numbers X and Y on the first and second die, respectively, will each have marginal PMF $p(x) = 1/6$ for $x = 1, 2, ..., 6$.

The joint PMF is $p(x,y) = 1/36$, so that $p(x) = \sum_{y=1}^{6} p(x,y)$ □

Example 5.2.
Consider X = person's age and Y = income. The data are abridged from the US Current Population Survey.[j] For the purposes of this example, we replace the age and income groups by their midpoints. For example, the first row represents ages 25-34 and the first column represents incomes $0-$10,000.

	Y, income 5	20	40	60	85	Total
X, age 30	0.049	0.116	0.084	0.039	0.032	0.320
40	0.042	0.093	0.081	0.045	0.061	0.322
50	0.047	0.102	0.084	0.053	0.072	0.358
Total	0.139	0.310	0.249	0.137	0.165	1.000

Here, the joint PMF is given inside the table and the marginal PMF's of X and Y are row and column totals, respectively.

For example, $p(30, 60) = 0.039$ and $p_Y(40) = 0.084 + 0.081 + 0.084 = 0.249$. □

For continuous random variables, the PMF's turn into densities, and summation into integration.

Definition 5.3. Joint density, marginal densities

The function $f(x,y)$ is a **joint probability density function** for the continuous random variables X and Y if

a) $f(x,y) \geq 0$, for all (x,y)

b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$.

c) $P[(X,Y) \in A] = \iint_A f(x,y)\,dx\,dy$ for any region A in the xy-plane.[a]

The **marginal probability density functions** of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

[a] Note that,even if X, Y are each continuous RV's, this does not always mean that the joint density exists. For example, X is Uniform[0,1] and $Y = X$. For this reason, X, Y satisfying this definition might be called **jointly continuous**

When X and Y are continuous random variables, the joint density function $f(x, y)$ describes the likelihood that the pair (X, Y) belongs to the neighborhood of the point (x, y). It is visualized as a surface lying above the xy plane.

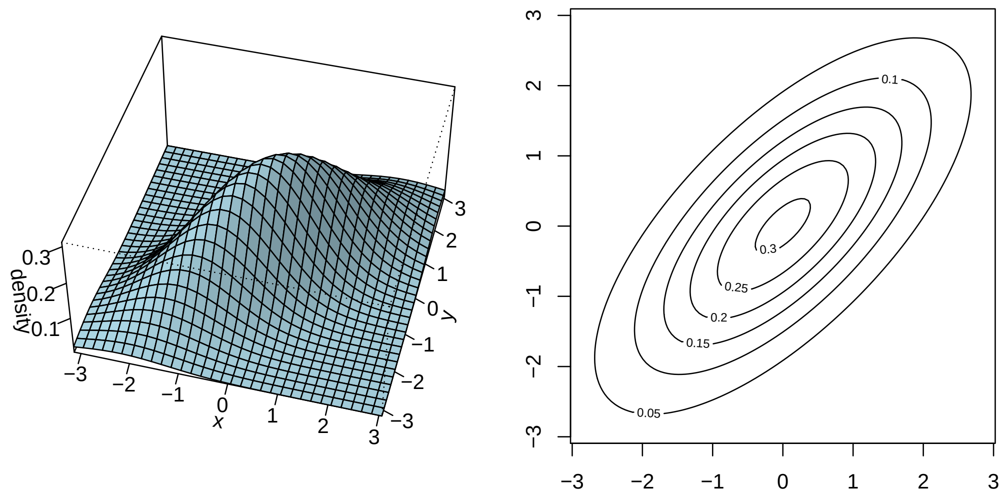


Figure 5.1: An example of a joint density function. Left: surface plot. Right: contour plot.

Example 5.3.
A certain process for producing an industrial chemical yields a product that contains two main types of impurities. Suppose that the joint probability distribution of the impurity concentrations (in mg/l) X and Y is given by

$$f(x, y) = \begin{cases} 2(1-x) & \text{for } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Verify the condition (b) of Definition 5.3
(b) Find $P(0 < X < 0.5, 0.4 < Y < 0.7)$[1]
(c) Find the marginal probability density functions for X and Y.

Solution. (a) Condition $\iint f(x, y)\, dx\, dy = 1$ can be verified by integrating one of the densities in part (c).
(b)

$$P(0 < X < 0.5,\ 0.4 < Y < 0.7) = \int_{0.4}^{0.7} \int_{0}^{0.5} 2(1-x) dx\, dy = 0.225$$

(c)

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 2(1-x) dy = 2(1-x), \qquad 0 < x < 1$$

and $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 2(1-x) dx = 1, \qquad 0 < y < 1$ □

[1]Recall that for continuous RV's, the choice of the $<$ of $\leq$ sign in the inequality does not matter.

5.2 Conditional probability distributions

Definition 5.4. Conditional PMF or density

For a pair of **discrete** RV's, the conditional PMF of X given Y is

$$p(x \mid y) = \frac{p(x,y)}{p_Y(y)} \text{ for } y \text{ such that } p_Y(y) > 0$$

For a pair of **continuous** RV's with joint density $f(x,y)$, the conditional density function of X given $Y = y$ is defined as

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} \text{ for } y \text{ such that } f_Y(y) > 0$$

and the conditional density of Y given $X = x$ is defined by

$$f(y|x) = \frac{f(x,y)}{f_X(x)} \text{ for } x \text{ such that } f_X(x) > 0$$

For discrete RV's, the conditional probability distribution of X given Y fixes a value of Y. For example, conditioning on $Y = 0$, produces

$$P(X = 0 \mid Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)}$$

Example 5.4.
Using the data from Example 5.2,

	Y, income					
	5	20	40	60	85	Total
X, age 30	0.049	0.116	0.084	0.039	0.032	0.320
40	0.042	0.093	0.081	0.045	0.061	0.322
50	0.047	0.102	0.084	0.053	0.072	0.358
Total	0.139	0.310	0.249	0.137	0.165	1.000

Calculate the conditional PMF of Y given $X = 30$.

Solution. Conditional PMF of Y given $X = 30$, will give the distribution of incomes in that age group. Divide all of the row $X = 30$ by its marginal and obtain

	5	20	40	60	85	Total
X, age 30	$\frac{0.049}{0.320} = 0.153$	$\frac{0.116}{0.320} = 0.362$	$\frac{0.084}{0.320} = 0.263$	$\frac{0.039}{0.320} = 0.122$	$\frac{0.032}{0.32} = 0.1$	1

The conditional PMF's will add up to 1. □

Example 5.5. *Uniform distribution on a rectangle*
X and Y will have (joint) uniform distribution on a rectangle R (or any other shape), if their joint PDF is constant:

$f(x, y) = k$ on R and 0 elsewhere. It's clear that $k = 1/Area(R)$ in order to satisfy the definition of joint PDF. If $R = [a, b] \times [c, d]$ is a rectangle, then $k = 1/[(b - a)(d - c)]$.

Example 5.6.
Suppose that X, Y are Uniform on the $[0, 2] \times [0, 2]$ square. Find the probability that $X + Y \leq 3$.

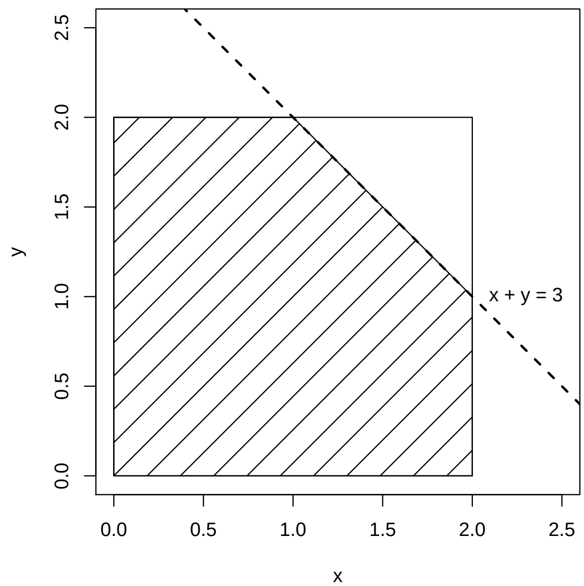


Figure 5.2: Areas in the Example 5.6

Solution. The shaded area corresponds to the inequality $X + Y \leq 3$; you can integrate over this area, or simply notice that, since the density is constant k, the integral should be equal to $k \cdot Area(shaded) = (1/4)(4 - 0.5) = 7/8$. □

Example 5.7.
The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces is

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

(a) Find the marginal densities.
(b) Find the conditional densities $f(x|y)$ and $f(y|x)$.

Solution. (a) By definition,

$$f_X(x) = \int_x^1 10xy^2 dy = \frac{10}{3}x(1 - x^3), \quad 0 < x < 1$$

$$f_Y(y) = \int_0^y 10xy^2 dx = 5y^4, \quad 0 < y < 1$$

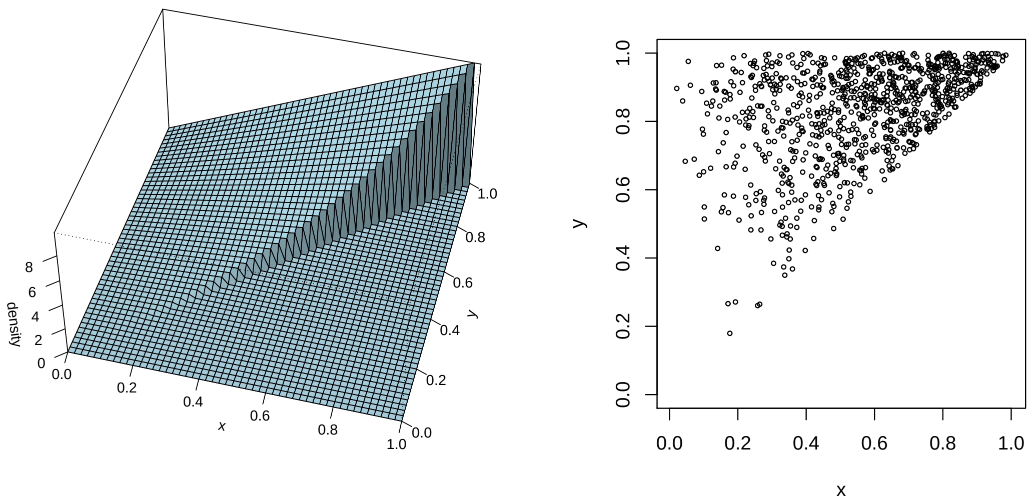


Figure 5.3: Left: Joint density from Example 5.7, right: a typical sample from this distribution

(b) Now

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{10xy^2}{(10/3)x(1-x^3)} = \frac{3y^2}{(1-x^3)},\ 0 < x < y < 1$$

and

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{10xy^2}{5y^4} = \frac{2x}{y^2},\ 0 < x < y < 1$$

For the last one, say, treat y as fixed (given) and x is the variable. □

5.3 Independent random variables

Definition 5.5. Independence

The random variables X and Y are said to be *statistically independent* iff

$$p(x,y) = p_X(x)p_Y(y) \text{ for discrete case}$$

and

$$f(x,y) = f_X(x)f_Y(y) \ \text{ for continuous case}$$

This definition of independence agrees with our definition for the events, $P(AB) = P(A)P(B)$. For example, if two dice are rolled independently, then the numbers X and Y on the first and second die, respectively, will each have PMF $p(x) = 1/6$ for $x = 1, 2, ..., 6$.

The joint PMF will then be $p(x,y) = p_X(x)p_Y(y) = (1/6)^2 = 1/36$.

Example 5.8.
Show that the random variables in Example 5.3 are independent.

Solution. Here,

$$f(x,y) = \begin{cases} 2(1-x) & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

We have $f_X(x) = 2(1-x)$ and $f_Y(y) = 1$ from Example 5.3, thus

$$f_X(x) f_Y(y) = 2(1-x)(1) = 2(1-x) = f(x,y)$$

for $0 < x, y < 1$ and 0 elsewhere. Hence, X and Y are independent random variables. □

Exercises

5.1.
Suppose that the rolls of two dice, X_1 and X_2 have joint PMF

$$p(i,j) = P(X_1 = i, X_2 = j) = 1/36$$

a) Are random variables X_1, X_2 independent? Explain.

b) Are the events $A = \{X_1 \leq 3\}$ and $B = \{X_2 \geq 3\}$ independent? Explain.

c) Are the events $C = \{X_1 + X_2 \leq 3\}$ and $D = \{X_1 - X_2 \geq 3\}$ independent? Explain.

5.2.
X and Y have the following joint density:

$$f(x,y) = \begin{cases} k & \text{for } 0 \leq x \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

a) Calculate the constant k that makes f a legitimate density.

b) Calculate the marginal densities of X and Y.

5.3.
The joint distribution for the number of total sales $=X_1$ and number of electronic equipment sales $=X_2$ per hour for a wholesale retailer are given below

X_2	0	1	2	
$X_1 = 0$	0.1	0	0	
$X_1 = 1$	0.1	0.2	0	
$X_1 = 2$	0.1	?	0.15	

a) Fill in the "?"

b) Compute the marginal probability function for X_2. (That is, find $P(X_2 = i)$ for every i.)

c) Find the probability that both $X_1 \leq 1$ and $X_2 \leq 1$.

d) Find the *conditional* probability distribution for X_2 given that $X_1 = 2$. (That is, find $P(X_2 = i \mid X_1 = 2)$ for every i.)

e) Are X_1, X_2 independent? Explain.

5.4.

X and Y have the following joint density:

$$f(x,y) = \begin{cases} kxy & \text{for } \ 0 \leq x, y \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

a) Calculate the constant k that makes f a legitimate density.

b) Calculate the marginal densities of X and Y.

c) Are X, Y independent? Explain.

5.5.

A point lands into $[0,1] \times [0,1]$ square with random coordinates X, Y independent, having Uniform$[0,1]$ distribution each.

a) What is the probability that the distance from the point to the origin is less than 1, that is, $P(X^2 + Y^2 < 1)$?

b) Find the conditional density of X given that $Y = 0.5$

5.6.

The random variables X, Y have joint density $f(x,y) = e^{-(x+y)}, \quad x, y > 0$

a) Are X, Y independent? Explain.

b) Find $P(X < 3, Y > 2)$

5.7.

A point is chosen at random from a rectangle with sides 3 and 5. That is, let X, Y be uniformly distributed on $[0,3] \times [0,5]$. Find the probability that the point is located at least one meter away from the boundary of the rectangle.

5.8.

A point is chosen at random from a circle with radius 2. That is, X, Y are uniformly distributed on this circle. Find the CDF of R= distance from a point to the center of the circle.

5.9.

A point is chosen at random from a circle with radius 1. That is, X, Y are uniformly distributed on this circle. Find the marginal densities of X and Y. Are X, Y independent?

5.4 Expected values of functions

Definition 5.6. Expected values

Suppose that the discrete RV's (X, Y) have a joint PMF $p(x, y)$. If $g(x, y)$ is any real-valued function, then

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{x}\sum_{y} g(x, y)p(x, y).$$

The sum is over all values of (x, y) for which $p(x, y) > 0$.

If (X, Y) are continuous random variables, with joint PDF $f(x, y)$, then

$$\mathbb{E}\left[g(X, Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\, f(x, y) dx\, dy.$$

Definition 5.7. Covariance

The *covariance* between two random variables X and Y is given by

$$Cov(X, Y) = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right],$$

where $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$.

The covariance helps us assess the relationship between two variables. Positive covariance means *positive association* between X and Y meaning that, as X increases, Y also tends to increase. Negative covariance means *negative association.*

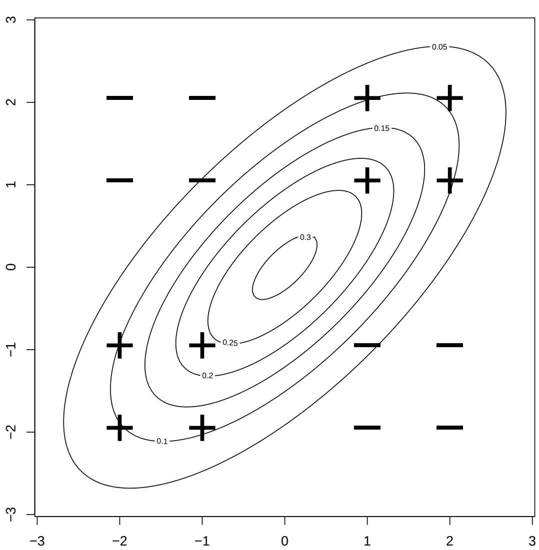


Figure 5.4: Explanation of positive covariance

In Figure 5.4, positive covariance is achieved, since pairs of x, y with positive products have higher densities than those with the negative products.

This definition also extends our notion of variance as $Cov(X, X) = V(X)$.

While covariance measures the direction of the association between two random variables, its magnitude is not directly interpretable. *Correlation coefficient*, introduced below, measures the strength of the association and has some nice properties.

Definition 5.8. Correlation

The *correlation coefficient* between two random variables X and Y is given by

$$\rho = \frac{Cov(X, Y)}{\sqrt{V(X)\, V(Y)}}$$

Properties of correlation:

- The correlation coefficient lies between -1 and $+1$.
- The correlation coefficient is dimensionless (while covariance has dimension of XY).
- If $\rho = +1$ or $\rho = -1$, then Y must be a linear function of X.
- The correlation coefficient does not change when X or Y are linearly transformed (e.g. when you change the units from miles to ångströms.)
- However, the correlation coefficient is not a good indicator of a *nonlinear* relationship.

The following Theorem simplifies the computation of covariance. Compare it to the variance identity $V(X) = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2$.

Theorem 5.1. Covariance

$$Cov(X, Y) = \mathbb{E}\,(XY) - \mathbb{E}\,(X)\mathbb{E}\,(Y)$$

Example 5.9.

Let X and Y have the following joint PMF

Y	0	1	
$X = 0$	0.1	0.25	0.35
$X = 1$	0.35	0.3	0.65
	0.45	0.55	1

Find the covariance and correlation between X and Y.

Solution. From the marginals, $\mathbb{E}\,(X) = 0.35(0) + 0.65(1) = 0.65$ and $\mathbb{E}\,(Y) = 0.55$. $\mathbb{E}\,(XY) = 0(0.1 + 0.25 + 0.35) + 1(0.3) = 0.3$. Therefore, $Cov(X, Y) = 0.3 - 0.65(0.55) = -\mathbf{0.0575}$.

Next, $V(X) = \mathbb{E}\,(X^2) - (\mathbb{E}\,X)^2 = 0.65 - 0.65^2 = 0.2275$ and $V(Y) = 0.2475$. Finally, $\rho = corr(X, Y) = -0.0575/\sqrt{0.2275(0.2475)} = -\mathbf{0.2423}$ $\square$

Example 5.10.

Let X denote the proporion of calls to the Support Instruction Center (SIC) about computers and Y the proportion of calls to SIC about projectors. It is estimated that X and Y have a joint density

$$f(x, y) = \begin{cases} c & \text{for } x > 0, y > 0, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the constant c that makes f a legitimate density. Then, find the covariance and correlation of X and Y.

Solution. We first compute the marginal density functions (sketching the density may help you set up the limits of integration). They are: $f_X(x) = \int_0^{1-x} c\,dy = c(1-x)$, and $f_Y(y) = \int_0^{1-y} c\,dx = c(1-y)$. Then, integrating one of them, say $\int_0^1 f_X(x)\,dx = \int_0^1 c(1-x)\,dx = c/2 = 1$, we get $c = 2$. Thus,

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

and you can notice that $f_Y = f_X$ here. From the marginal density functions, we get

$$\mathbb{E}(X) = \int_0^1 x * 2(1-x)dx = \frac{1}{3} \qquad \text{and } \mathbb{E}(Y) = \mathbb{E}(X)$$

Now, we are ready to calculate covariance. From the joint density function given, we have

$$\mathbb{E}(XY) = \int_0^1 \int_0^{1-y} xy * 2\,dx\,dy = \frac{1}{12}.$$

Then
$$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{12} - \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = -\frac{1}{36}$$

To find correlation ρ, we first need to find variances of X and Y.

$$\mathbb{E}(X^2) = \int_0^1 x^2 * 2(1-x)\,dx = \frac{1}{6} \qquad \text{and } \mathbb{E}(Y^2) = \int_0^1 y^2 * 2(1-y)\,dy = \mathbb{E}(X^2)$$

Thus $V(X) = 1/6 - (1/3)^2 = 1/18 = V(Y)$.

Finally, $\rho = \dfrac{-1/36}{\sqrt{1/18}\sqrt{1/18}} = -1/2$ □

Theorem 5.2. Covariance and independence

> If random variables X and Y are independent, then $Cov(X, Y) = 0$.

Proof. We will show the proof for the continuous case; the discrete case follows similarly. For independent X, Y,

$$\mathbb{E}(XY) = \iint xy\, f(x,y)dx\, dy = \iint x f_X(x) y f_Y(y) dx\, dy =$$

$$= \left(\int x f_X(x)dx\right)\left(\int y f_Y(x)dy\right) = \mathbb{E}(X)\,\mathbb{E}(Y)$$

Therefore, $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$. □

Of course, if covariance is 0, then so is the correlation coefficient. Such random variables are called *uncorrelated*. The converse of this Theorem is not true, meaning that **zero covariance does not necessarily imply independence**.

5.4.1 Variance of sums

Recall that $\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c$, regardless of the joint distribution of X, Y. The same is not true for variances. However, the following Theorem simplifies calculation of variance in certain cases.

Theorem 5.3. Variance of sums

If X and Y are random variables and $U = aX + bY + c$, then

$$V(U) = V(aX + bY + c) = a^2 V(X) + b^2 V(Y) + 2ab\, Cov(X,Y)$$

If X and Y are independent then $V(U) = V(aX + bY) = a^2 V(X) + b^2 V(Y)$

Example 5.11.
If X and Y are random variables with variances $V(X) = 2$, $V(Y) = 4$, and covariance $Cov(X,Y) = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

Solution. By Theorem 5.3,

$$V(Z) = \sigma_Z^2 = V(3X - 4Y + 8) = 9V(X) + 16V(Y) - 24\, Cov(X,Y)$$

so $V(Z) = (9)(2) + (16)(4) - 24(-2) = 130$. □

Corollary. If the random variables X and Y are independent, then

$$V(X+Y) = V(X) + V(Y)$$

Note. Theorem 5.3 and the above Corollary naturally extend to more than 2 random variables. If X_1, X_2, ..., X_n are all independent RV's, then

$$V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ... + V(X_n)$$

Example 5.12.
We have discussed in Chapter 3 that the Binomial random variable Y with parameters n, p can be represented as $Y = X_1 + X_2 + ... + X_n$. Here X_i are independent Bernoulli (0/1) random variables with $P(X_i = 1) = p$.

It was found that $V(X_i) = p(1-p)$. Then, using the above Note, $V(Y) = V(X_1) + V(X_2) + ... + V(X_n) = np(1-p)$, which agrees with the formula for Binomial variance in Section 3.4.

The same reasoning applies to Gamma RV's. If $Y = X_1 + X_2 + ... + X_n$, where X_i are independent Exponentials, each with mean β, then we know that $V(X_i) = \beta^2$ and Y has Gamma distribution with $\alpha = n$. Then, $V(Y) = V(X_1) + V(X_2) + ... + V(X_n) = n\beta^2$. □

Example 5.13.
A very important application of Theorem 5.3 is the calculation of variance of the **sample mean**

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{Y}{n}$$

where X_i are independent and identically distributed RV's (representing a sample of measurements), and Y denotes the total of all measurements.

Suppose that $V(X_i) = \sigma^2$ for each i. Then

$$V(\overline{X}) = \frac{V(Y)}{n^2} = \frac{V(X_1) + V(X_2) + ... + V(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

This means that $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, that is, **the mean of n independent measurements is $\sqrt{n}$ more precise than a single measurement.** □

Example 5.14.
The error in a single permeability measurement has the standard deviation of 0.01 millidarcies (md). If we made 8 independent measurements, how large is the error we should expect from their mean?

Solution. $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 0.01/\sqrt{8} \approx 0.0035 md$ □

Exercises

5.10.

Y	0	1	2	
$X = 0$	0.1	0	0	
$X = 1$	0.1	0.2	0	
$X = 2$	0.1	0.35	0.15	

Find the covariance and correlation between X and Y.

5.11.
For X, Y as in Exercise 5.10, find $\mathbb{E}(X^2)$ and $\mathbb{E}[(X-2)e^Y]$.

5.12.
X and Y have the following joint density:

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

a) Calculate $\mathbb{E}\left(X^2 Y\right)$.

b) Calculate $\mathbb{E}\left(X/Y\right)$.

5.13.
Using the density in Problem 5.12, find the covariance and correlation between X and Y.

5.14.
The random variables X, Y have the following joint density function:

$$f(x,y) = \begin{cases} k, & 0 \le x, \quad 0 \le y, \quad \texttt{and} \quad x+y \le 2 \\ 0 & \texttt{elsewhere} \end{cases}$$

Sketch the region where f is positive and answer the following questions:

a) Find the constant k that makes f a true density function.

b) Find the marginal density of X.

c) Find the probability that $X + Y > 1$

d) **Set up, do not evaluate** the expression for the expected value of $U = X^2\sqrt{1+Y^3}$.

5.15.
Ten people get into an elevator. Assume that their weights are independent, with the mean 150 lbs and standard deviation 30 lbs.

a) Find the expected value and the standard deviation of their total weight.

b) Assuming Normal distribution, find the probability that their combined weight is less than 1700 pounds.

5.16.
While estimating speed of light in a transparent medium, an individual measurement X is determined to be unbiased (that is, the mean of X equals the unknown speed of light), but the measurement error, assessed as the standard deviation of X, equals 35 kilometers per second (km/s).

a) In an experiment, 20 independent measurements of the speed of light were made. What is the standard deviation of the mean of these measurements?

b) How many measurements should be made so that the error in estimating the speed of light (measured as $\sigma_{\overline{X}}$) will decrease to 5 km/s?

5.17.
Near-Earth asteroids (NEA) are being surveyed. The average mass of these is 250 tons, with a standard deviation of 180 tons. Suppose that 3 NEAs are randomly selected.

a) Find the expected value and standard deviation of the total mass of these NEAs, assuming that their masses are independent.

b) How does your answer change if you assume that that each pair of masses have correlation of $\rho = 0.5$?

5.18.
A part is composed of two segments. One segment is produced with the mean length 4.2cm and standard deviation of 0.1cm, and the second segment is produced with the mean length 2.5cm and standard deviation of 0.05cm. Assuming that the production errors are independent, calculate the mean and standard deviation of the total part length.

5.19.
Random variables X and Y have means 3 and 5, and variances 0.5 and 2, respectively. Further, the correlation coefficient between X and Y equals -0.5. Find the mean and variance of $U = X + Y$ and $W = X - Y$.

5.20. $\star$
Find an example of uncorrelated, but not independent random variables. [**Hint:** Two discrete RV's with 3 values each are enough.]

5.5 Conditional Expectations*

Definition 5.9. Conditional Expectation

If X and Y are any two random variables, the conditional expectation of X given that $Y = y$ is defined to be

$$\mathbb{E}\left(X \mid Y = y\right) = \int_{-\infty}^{\infty} x f(x|y) dx$$

if X and Y are jointly continuous, and

$$\mathbb{E}\left(X \mid Y = y\right) = \sum_{x} x\, p(x|y)$$

if X and Y are jointly discrete.

Note that $\mathbb{E}(X|Y = y)$ is a number depending on y. If now we allow y to vary randomly, we get a random variable denoted by $\mathbb{E}(X|Y)$. The concept of conditional expectation is useful when we have only a partial information about X, as in the following example.

Example 5.15.
Suppose that random variable X is the number rolled on a die, and $Y = 0$ when $X \leq 3$ and $Y = 1$ otherwise. Thus, Y carries partial information about X, namely, whether $X \leq 3$ or not.

a) Compute the conditional expectation $\mathbb{E}\left(X \mid Y = 0\right)$.

b) Describe the random variable $\mathbb{E}(X \mid Y)$.

Solution. (a) The conditional distributions of X are given by

$$P(X = x \mid Y = 0) = \frac{P(X = x, Y = 0)}{P(Y = 0)} = \frac{1/6}{1/2} = 1/3$$

for $x = 1, 2, 3$, and
$P(X = x \mid Y = 1) = 1/3$ for $x = 4, 5, 6$.

Thus, $\mathbb{E}(X \mid Y = 0) = (1/3)(1 + 2 + 3) = 2$ and $\mathbb{E}(X \mid Y = 1) = (1/3)(4 + 5 + 6) = 5$

(b) $\mathbb{E}(X \mid Y)$ is 2 or 5, depending on Y. Each value may happen with probability 1/2. Thus, $P[\mathbb{E}(X \mid Y) = 2] = 0.5$ and $P[\mathbb{E}(X \mid Y) = 5] = 0.5$ □

Theorem 5.4. Expectation of expectation

Let X and Y denote random variables. Then

$$(a) \quad \mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)]$$

$$(b) \quad V(X) = \mathbb{E}[V(X|Y)] + V[\mathbb{E}(X|Y)]$$

Proof. (Part (a) only.)
Let X and Y have joint density $f(x, y)$ and the marginal densities $f_X(x)$ and $f_Y(y)$, respectively. Then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x|y) f(y) dx\, dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f(x|y) dx \right] f(y) dy$$

$$= \int_{-\infty}^{\infty} \mathbb{E}(X|Y = y) f(y) dy = \mathbb{E}[\mathbb{E}(X|Y)]$$

□

Example 5.16.
Suppose we are interested in the total weight X of occupants in a car. Let the number of occupants equal Y, and each occupant weighs 150 lbs on average.[2] Then $\mathbb{E}(X \mid Y = y) = 150y$. Suppose Y has the following distribution

y	1	2	3	4
$p(y)$	0.62	0.28	0.07	0.03
$150y$	150	300	450	600

Then $\mathbb{E}(X \mid Y)$ has the distribution with values given in the last row of the table, and probabilities identical to $p(y)$. We can verify by straightforward calculation that $\mathbb{E}(X \mid Y) = \mathbb{E}(150Y) = 226.5$. Then the Theorem says that $\mathbb{E}(X) = 226.5$ as well, so we don't even have to know the distribution of occupant weights Y, only its mean (150). □

[2] We will assume that the number of occupants is independent of occupants' weights.

Exercises

5.21.
For the random variables X and Y from Example 5.15, verify the identity in part (a) of the Theorem 5.4.

5.22.
Suppose that the number of lobsters caught in a trap follows the distribution

y	0	1	2	3
$p(y)$	0.5	0.3	0.15	0.05

and the average weight of lobster is 1.7 lbs, with variance 0.25 lbs^2. Find the expected value and the variance of the total catch in one trap. Assume independence of lobsters' weights.

5.23.
In the following table from US Social Security Administration[k], the survival probabilities $P(X \geq a)$ for US males are given, where a is the current age.

a, age	0	10	20	30	40	50	60
$P(X \geq a)$	1	0.992	0.987	0.975	0.959	0.928	0.860
a, age	70	80	90	100	110		
$P(X \geq a)$	0.734	0.499	0.172	0.087	0.001		

a) Find $P(X \geq 80 \,|\, X \geq 60)$

b) Find $\mathbb{E}\,(X \,|\, X \geq 60)$.[3] Since the data are only given once per decade, approximate the age at death at the midpoint, e.g. if $60 \leq X < 70$ then count it as $X = 75$. Explain why the result is higher than unconditional $\mathbb{E}\,(X)$.

5.24.
Tires from Manufacturer A last 50 thousand miles, on average, with the standard deviation 2.887 thousand miles; and those from Manufacturer B last 60 thousand miles, on average, with standard deviation 2.887 too. You pick a tire at random, with 50% chance it comes from A and 50% chance it comes from B. Find the expected lifetime of your tire, and standard deviation of the lifetime.

Verify your calculations assuming that tire A has Uniform[45,55] lifetime and tire B has Uniform[55,65] lifetime; so that the "random tire" will have Uniform[45,65] lifetime.

[3] The related quantity, $\mathbb{E}\,(X \,|\, X \geq 60) - 60$ is called *life expectancy at 60*

Chapter 6

Functions of Random Variables

6.1 Introduction

At times we are faced with a situation where we must deal not with the random variable whose distribution is known but rather with some function of that random variable. For example, we might know the distribution of particle sizes, and would like to infer the distribution of particle weights.

In the case of a simple linear function, we have already asserted what the effect is on the mean and variance. What has been omitted was what actually happens to the distribution.

We will discuss several methods of obtaining the distribution of $Y = g(X)$ from known distribution of X. The CDF method and the transformation method are most frequently used. The CDF method is all-purpose and flexible. The transformation method is typically faster (when it works).

6.1.1 Simulation

One use of the above methods is to generate random variables with a given distribution. This is important in simulation studies. Suppose that we have a complex operation that involves several components. Suppose that each component is described by a random variable and that the outcome of the operation depends on the components in a complicated way. One approach to analyzing such a system is to simulate each component and calculate the outcome for the simulated values. If we repeat the simulation many times, then we can get an idea of the probability distribution of the outcomes. Some examples of simulation are given in Labs.

6.2 Method of distribution functions (CDF)

The CDF method is straightforward and very versatile. The procedure is to derive the CDF for $Y = g(X)$ in terms of both the CDF of X, $F(x)$, and the function g, while also noting how the range of possible values changes. This is done by starting with the computation of $P(Y < y)$ and inverting this into a statement that can often be expressed in terms of the CDF of X.

If we also need to find the density of Y, we can do this by differentiating its CDF.

Example 6.1.
Suppose X has cdf given by $F(x) = 1 - e^{-\lambda x}$, so that X is Exponential with the mean $1/\lambda$. Let $Y = bX$ where $b > 0$. Note that the range of Y is the same as the range of X, namely $(0, \infty)$.

$$P(Y < y) = P(bX < y) = P(X < y/b) =$$

(Since $b > 0$, the inequality sign does not change.)

$$= 1 - e^{-\lambda y/b} = 1 - e^{-(\lambda/b)y}$$

The student should recognize this as CDF of the exponential distribution with the mean b/λ. We already knew that the mean would be b/λ, but we did not know that Y also has an exponential distribution. $\square$

Example 6.2.
Suppose X has a uniform distribution on $[a, b]$ and $Y = cX + d$, with $c > 0$. Find the CDF of Y.

Solution. Recall that $F(t) = (t - a)/(b - a)$. Note that the range of Y is $[ca + d, cb + d]$. We have

$$P(Y < t) = P(cX + d < t) = P(X < (t - d)/c) = F((t - d)/c)$$

$$= ((t - d)/c - a)/(b - a) = (t - d - ac)/(c(b - a))$$

With a little algebra, this can be shown to be the uniform CDF on $[ca + d, cb + d]$. $\square$

This example shows that certain simple transformations do not change the distribution type, only the parameters. Sometimes, however, the change is dramatic.

Example 6.3.
Show that if X has a uniform distribution on the interval $[0, 1]$ then
$Y = -\ln(1 - X)$ has an exponential distribution with mean 1.

Solution. Recall that for the uniform distribution on $(0, 1)$, $P(X < x) = x$. Also, note that the range of Y is $(0, \infty)$.

$$P(Y < t) = P(-\ln(1 - X) < t) = P(\ln(1 - X) > -t) =$$

$$= P\left(1 - X > e^{-t}\right) = P\left(X < 1 - e^{-t}\right) = 1 - e^{-t}$$

Incidentally, note that if X has a uniform distribution on $(0, 1)$, then so does $W = 1 - X$. (See exercises.) $\square$

Example 6.4.
The pdf of X is given by

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

Find the pdf of $U = 40(1 - X)$.

Solution.

$$F(u) = P(U \le u) = P[40(1-X) \le u] = P\left(X > 1 - \frac{u}{40}\right) = 1 - P(X \le 1 - \frac{u}{40})$$
$$= 1 - F_X\left(1 - \frac{u}{40}\right) = 1 - \int_0^{1-u/40} f(x)dx = 1 - \left(1 - \frac{u}{40}\right)^3.$$

Therefore,

$$f(u) = F_U'(u) = \frac{3}{40}\left(1 - \frac{u}{40}\right)^2, \quad \text{for} \quad 0 \le u \le 40$$

□

Exercises

6.1.
Show that if X has a uniform distribution on $[0,1]$, then so does $1 - X$.

6.2.
Let X have a uniform distribution on $[0,1]$. Let $Y = X^{1/3}$.

a) Find the distribution of Y.

b) Find the mean of Y using the result in (a).

c) Find the mean of Y using the formula $\mathbb{E}\, g(X) = \int g(x) f(x)\, dx$.

6.3.
Using the CDF method, show that the Weibull random variable Y (with some parameter $\gamma > 0$, and $\beta = 1$) can be obtained from Exponential X (with the mean 1) as $Y = X^{1/\gamma}$.

6.4.
Suppose the radii of spheres have a normal distribution with mean 2.5 and variance $\frac{1}{12}$. Find the median volume and median surface area.

6.5.
Let X have a uniform distribution on $[0,1]$. Show how you could define $H(x)$ so the $Y = H(X)$ would have a Poisson distribution with mean 1.3.

6.6.
A point lands into $[0,1] \times [0,1]$ square with random coordinates X, Y independent, having Uniform$[0,1]$ distribution each. Use the CDF method to find the distribution of $U = \max(X, Y)$.

6.7.
Let X, Y be independent, standard Normal RV's. Find the distribution of $Z = \sqrt{X^2 + Y^2}$. You can interpret this as the distance from a random point (X, Y) to the origin. [**Hint:** Use the polar coordinates.]

6.3 Method of transformations

Theorem 6.1. Transformations: discrete

> Suppose that X is a discrete random variable with probability mass function $p(x)$. Let $Y = h(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = h(x)$ can be uniquely solved for x, say $x = w(y)$. Then the PMF of Y is $p_Y(y) = p_X[w(y)]$.

For a discrete RV, the probabilities will stay the same and only the values of X will change to the values of Y. In case the function h is not one-to-one, you should also take care to aggregate the values that might appear several times.

Example 6.5.
Let X be a geometric random variable with PMF

$$p(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1}, \quad x = 1, 2, 3, \ldots$$

Find the distribution of the random variable $Y = X^2$.

Solution. Since the values of X are all positive, the transformation defines a one-to-one correspondence between the x and y values, $y = x^2$ and $x = \sqrt{y}$. Hence,

$$p_Y(y) = p_X(\sqrt{y}) = \frac{3}{4}\left(\frac{1}{4}\right)^{\sqrt{y}-1}, \qquad y = 1, 4, 9, ... \qquad \square$$

For continuous RV's, the transformation formula originates from the change of variable formula for integrals.

Theorem 6.2. Transformations: continuous

Suppose that X is a continuous random variable with density $f_X(x)$. Let $y = h(x)$ define a one-to-one transformation that can be uniquely solved for x, say $x = w(y)$, and $J = w'(y)$ exists (it is called the **Jacobian** of the transformation). Then the density of $Y = h(X)$ is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X[w(y)] \times |J|$$

Example 6.6.
Let X be a continuous random variable with probability distribution

$$f(x) = \begin{cases} x/12 & \text{for } 1 \le x \le 5 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability distribution of the random variable $Y = 2X - 3$.

Solution. The inverse solution of $y = 2x - 3$ yields $x = (y + 3)/2$, from which we obtain $J = w'(y) = \frac{dx}{dy} = \frac{1}{2}$. Therefore, using the above Theorem 6.2, we find the density function of Y to be

$$f_Y(y) = \frac{1}{12}\left(\frac{y+3}{2}\right)\frac{1}{2} = \frac{y+3}{48}, \qquad -1 < y < 7$$

$\square$

Example 6.7.
Let X be a Uniform$[0, 1]$ random variable. Find the distribution of $Y = X^5$.

Solution. Inverting, $x = y^{1/5}$, and $dx/dy = (1/5)y^{-4/5}$. Thus, we obtain

$$f_Y(y) = 1 \times (1/5)y^{-4/5} = (1/5)y^{-4/5}, \qquad 0 < y < 1 \qquad \square$$

Example 6.8.
Let X be a continuous random variable with density

$$f(x) = \begin{cases} \frac{x+1}{2} & \text{for } -1 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the density of the random variable $Y = X^2$.

Solution. The inversion of $y = x^2$ yields $x_{1,2} = \pm\sqrt{y}$, from which we obtain $J_1 = w_1'(y) = \frac{dx_1}{dy} = \frac{1}{2\sqrt{y}}$ and $J_2 = w_2'(y) = \frac{dx_2}{dy} = -\frac{1}{2\sqrt{y}}$. We cannot directly use Theorem 6.2 because the function $y = x^2$ is not one-to-one. However, we can split the range of X into two parts $(-1, 0)$ and $(0, 1)$ where the function is one-to-one. Then, we will just add the results.

Thus, we find the density function of Y to be

$$f_Y(y) = |J_1| f(\sqrt{y}) + |J_2| f(-\sqrt{y})$$

$$= \frac{1}{2\sqrt{y}}\left(\frac{\sqrt{y}+1}{2} + \frac{-\sqrt{y}+1}{2}\right) = \frac{1}{2\sqrt{y}} \text{ for } 0 \le y \le 1$$ □

Example 6.9. Location and Scale parameters
Suppose that X is some standard distribution (for example, Standard Normal, or maybe Exponential with $\beta = 1$) and $Y = a + bX$, or, solving for X,

$$X = \frac{Y - a}{b}$$

Then a is called **location (or shift) parameter** and b is **scale** parameter.

Let X have the density $f(x)$. Then the density of Y can be obtained from Theorem 6.2 as

$$f_Y(y) = f(x)\left|\frac{dx}{dy}\right| = \frac{1}{|b|} f\left(\frac{y-a}{b}\right) \tag{6.1}$$

For example, let X be Exponential with the mean 1, and $Y = bX$. Then $f(x) = e^{-x}$, $x > 0$, and (6.1) gives

$$f_Y(y) = (1/b)e^{-y/b}, \quad y > 0$$

That is, Y is Exponentially distributed with the mean b. This agrees with the result of Example 6.1.

Another example of location and scale parameters is provided by Normal distribution: if Z is standard Normal, then $Y = \mu + \sigma Z$ produces Y a Normal (μ, σ^2) random variable. Thus, μ is the location and σ is the scale parameter.
Formula (6.1) also provides a faster way to solve some of the above Examples. □

Exercises

6.8.
Suppose that $Y = \cos(\pi X)$ where the RV X is given by the table

x	-2	-1	0	2	3
$p(x)$	0.1	0.2	0.3	0.3	0.1

Find the distribution of Y (make a table).

6.9.
The random variable X has a distribution given by the table

x	-1	0	1	2
$p(x)$	0.1	0.2	0.3	0.4

Find the distribution of a random variable $Y = X^2 - 1$.

6.10.
Let X be a continuous random variable with density

$$f(x) = \begin{cases} \frac{2}{3}(x+1) & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the density of the random variable $Y = X^2$.

6.11.
Use the methods of this section to show that linear functions of normal random variables again have a normal distribution. Let $Y = a + bX$, where X is normal with the mean μ and variance σ^2. How do the mean and variance of Y relate to those of X? Again, use the methods of this section.

6.12.
The so-called *Pareto* random variable X with parameters 10 and 2 has the density function

$$f(x) = \frac{10}{x^2}, \qquad x > 10$$

Write down the density function of $Y = 4X - 20$ (do not forget the limits!)

6.13.
X is a random variable with Uniform $[0, 5]$ distribution.

a) Compute the density function for $Y = \sqrt{X}$, do not forget the limits!

b) Find the expected value of Y.

6.14.
Let X be a random variable with Uniform $[0, 1]$ distribution. Use formula (6.1) to calculate the density of $Y = 5X + 7$. What is the distribution of Y? (That is, give the name and parameters.)

6.15.
Re-do Example 6.4 (p. 112) using the transform (Jacobian) method.

6.16.
For the following distributions identify the parameters as location or scale parameters, or neither:

a) Weibull, parameter β.

b) Weibull, parameter γ.

c) Uniform on $[-\theta, \theta]$, parameter θ.

d) Uniform on $[b, b + 1]$, parameter b.

6.4 Central Limit Theorem

Sample mean (average of all observations) plays a central role in statistics. We have discussed the variance of the sample mean in Section 5.4. Here are more facts about the behavior of the sample mean.

From the linear properties of the expectation, it's clear that

$$\mathbb{E}\,(\overline{X}) = \mathbb{E}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{n\mu}{n} = \mu.$$

Summarizing the above, we obtain

Definition 6.1. Sample mean

A group of independent random variables from some distribution is called a *sample*, usually denoted as

$$X_1, X_2, ..., X_n.$$

Sample mean, denoted $\overline{X}$, is

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

If $\mathbb{E}\,(X_i) = \mu$ and $V(X_i) = \sigma^2$ for all i, then the mean and variance of sample mean are

$$\mathbb{E}\,(\overline{X}) = \mu \quad \text{and} \quad V(\overline{X}) = \sigma^2/n$$

But we are not only able to find the mean and variance of $\overline{X}$, but to describe (albeit approximately) its entire distribution!

Theorem 6.3. CLT

Let $\overline{X}$ be the mean of a sample coming from some distribution with mean μ and variance σ^2. Then, for large n, $\overline{X}_n$ is approximately Normal with mean μ and variance σ^2/n.

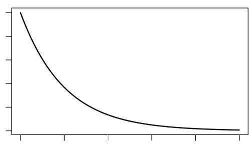
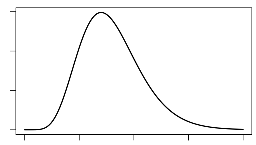
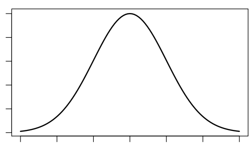
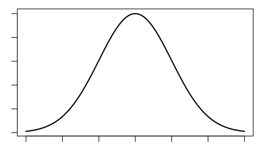
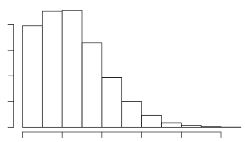
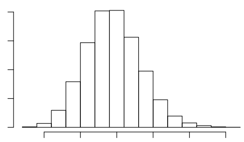
Here we mention (without proof, which can be obtained using the moment generating functions) some properties of the sums of independent random variables.

What do these have in common?
The sum of independent Normal RV's is always Normal. The shape of the sum distribution for other independent RV's starts resembling Normal as n increases.

The Central Limit Theorem (CLT) ensures the similar property for most general distributions. However, it holds *in the limit*, that is, as n gets large (practically, $n > 30$ is usually enough). According to it, the sums of independent RV's approach normal distribution. The same holds for averages, since they are sums divided by n.

If $n < 30$, the approximation is good only if the population distribution is not too different from a normal. If the population is normal, the sampling distribution of $\overline{X}$ will follow a normal distribution exactly, no matter how small the sample size.[1]

[1] There are some cases of the so-called "heavy-tailed" distributions for which the CLT does not hold, but they will not be discussed here.

	Distribution of X_i	Distribution of $Y = X_1 + X_2 + \ldots + X_n$ (indep.)
Exponential		$\mapsto$ Gamma 
Normal		$\mapsto$ Normal 
Poisson		$\mapsto$ Poisson 

Example 6.10.

The average voltage of the batteries is 9.2V and standard deviation is 0.25V. Assuming normal distribution and independence, what is the distribution of total voltage $Y = X_1 + \ldots + X_4$? Find the probability that the total voltage is above 37.

Solution. The mean is $4 \times 9.2 = 36.8$. The variance is $4 \times 0.25^2 = 0.25$. Furthermore, Y itself will have a normal distribution.
Using z-scores, $P(Y > 37) = P(Z > (37 - 36.8)/0.5) = P(Z > 0.4) = 0.5 - 0.1554 = 0.345$ from Normal table, p. 85. $\square$

Example 6.11.

An electrical firm manufactures light bulbs with average lifetime equal to 800 hours and standard deviation of lifetimes equal 400 hours. Approximate the probability that a random sample of 16 bulbs will have an average life of less than 725 hours.

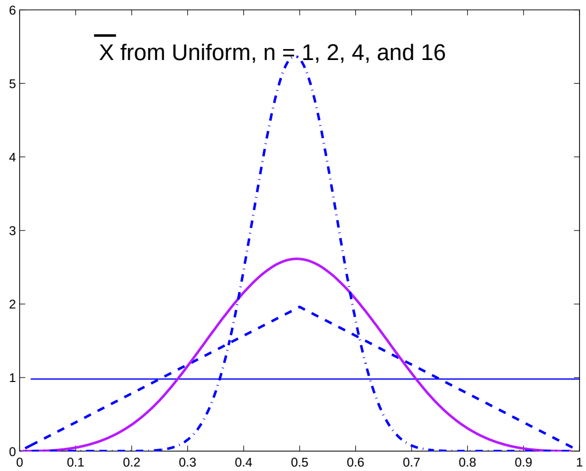
Solution. The sampling distribution of $\overline{X}$ will be approximately normal, with mean $\mu_{\overline{X}} = 800$ and $\sigma_{\overline{X}} = \frac{400}{\sqrt{16}} = 100$. Therefore,

$$P(\overline{X} < 725) \approx P\left(Z < \frac{725 - 800}{100}\right) = P(Z < -0.75) = 0.5 - 0.2734 = 0.2266$$

$\square$

Dependence on n

As n increases, two things happen to the distribution of $\overline{X}$: it is becoming sharper (due to the variance decreasing) and also the shape is becoming more and more Normal. For example, if X_i are Uniform[0,1], then the density of $\overline{X}$ behaves as follows:



Example 6.12.
The fracture strengths of a certain type of glass average 14 (thousands of pounds per square inch) and have a standard deviation of 2. What is the probability that the average fracture strength for 100 pieces of this glass exceeds 14.5?

Solution. By the central limit theorem the average strength $\overline{X}$ has approximately a normal distribution with mean= 14 and standard deviation, $\sigma = \frac{2}{\sqrt{100}} = 0.2$. Thus,

$$P(\overline{X} > 14.5) \approx P\left(Z > \frac{14.5 - 14}{0.2}\right) = P(Z > 2.5) = 0.5 - 0.4938 = 0.0062$$

from normal probability Table. □

6.4.1 CLT examples: Binomial

Historically, CLT was first discovered in case of Binomial distribution. Since Binomial Y is a sum of n independent Bernoulli RV's, CLT applies and says that $\overline{X} = Y/n$ is approximately Normal, mean p and variance $p(1-p)/n$. In this case, $\hat{p} := Y/n$ is called *sample proportion.* The Binomial Y itself is also approximately Normal with mean np and variance $np(1-p)$, as was discussed earlier in Section 4.6.2.

Example 6.13.
A fair ($p = 0.5$) coin is tossed 500 times.

a) What is the expected proportion of Heads?

b) What is the typical deviation from the expected proportion?

c) What is the probability that the sample proportion is between 0.46 and 0.54?

Solution. (a) We have $\mathbb{E}(\hat{p}) = p = 0.5$ and $\sigma_{\hat{p}} = \sqrt{p(1-p)/n} = \sqrt{0.25/500} = 0.0224$.

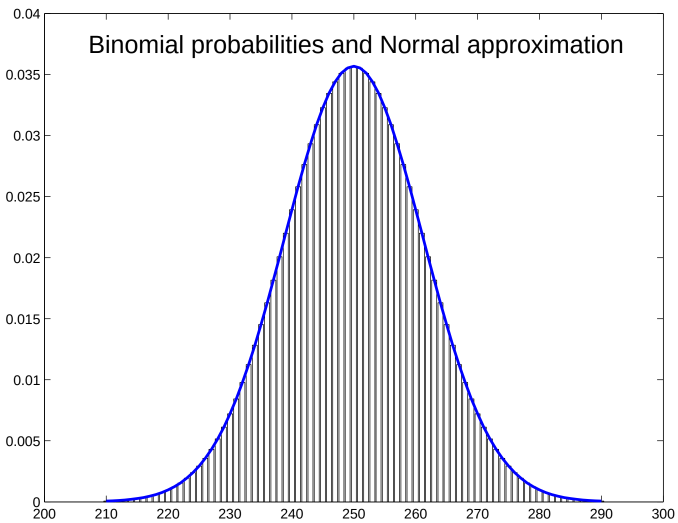
(b) For example, the *empirical rule* states that about 68% of a normal distribution is contained within one standard deviation of its mean. Here, the 68% interval is about

0.5 ± 0.0224, or 0.4776 to 0.5224.

(c)

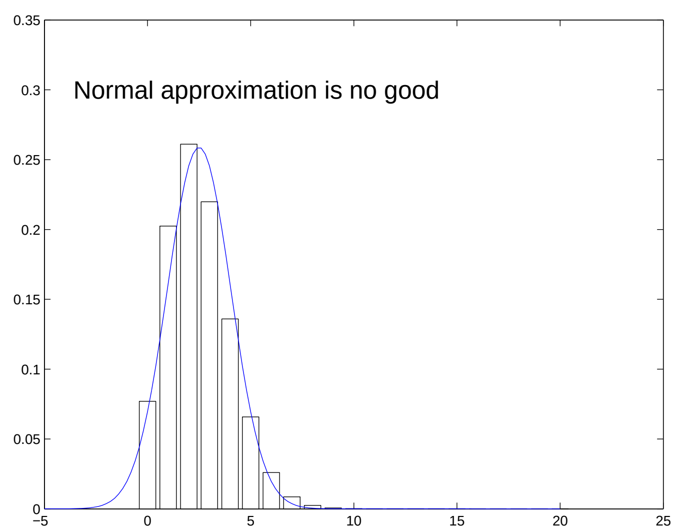
$$P(0.46 \leq Y \leq 0.54) \approx P\left(\frac{0.46-0.5}{0.0224} < Z < \frac{0.54-0.5}{0.0224}\right) =$$

$= P(-1.79 < Z < 1.79) = 2P(0 < Z < 1.79) = 0.927$ □



Normal approximation for $n = 500$ and $p = 0.5$

Normal approximation is not very good when np is small. Here's an example with $n = 50$ and $p = 0.05$:



Exercises

6.17.
The average concentration of potassium in county soils was determined as 85 ppm, with standard deviation 30 ppm. If $n = 20$ samples of soils are taken, find the probability that their average potassium concentration will be in the "medium" range (80 to 120 ppm).

6.18.
The heights of students have a mean of 174.5 centimeters (cm) and a standard deviation of 6.9 cm. If a random sample of 25 students is obtained, determine

a) the mean and standard deviation of $\overline{X}$;

b) the probability that the sample mean will fall between 172.5 and 175.8 cm;

c) the 70th percentile of the $\overline{X}$ distribution.

6.19.
The measurements of an irregular signal's frequency have mean of 20 Hz and standard deviation of 5 Hz. 50 independent measurements are done.

a) Find the probability that the average of these 50 measurements will be within 1 unit of the theoretical mean 20.

b) How many measurements should be done to ensure that the probability in part (a) equals 0.9?

6.20.
A process yields 10% defective items. If 200 items are randomly selected from the process, what is the probability that the sample proportion of defectives

a) exceeds 13%?

b) is less than 8%?

6.21.
The weight X_i of a Giant Siamese Frog has an approximately normal distribution with the mean of 215g and standard deviation of 40g.

a) Find the probability that a single frog weighs is between 200 and 230g.

b) Find the 85th percentile of the frogs' weights.

c) Let $\overline{X}$ be the average weight of 25 frogs. Find the mean and standard deviation of $\overline{X}$. Assume that the frogs' weights are independent.

d) Approximate the probability that $\overline{X}$ is between 200 and 230g. Compare to part (a)

6.22.
The proportion of office workers who frequently check Facebook at work is believed to be 0.8. When $n = 100$ office workers are observed, X of them will be found checking Facebook.

a) Find the mean and standard deviation of X.

b) Find the Normal approximation for the probability that X will be between 75 and 87 (inclusive).

Chapter 7

Descriptive statistics

The goal of statistics is somewhat complementary to that of the probability. Probability answers the question of what data are likely to be obtained from known probability distributions.
Statistics answers the opposite question: what kind of probability distributions are likely to have generated the data at hand?

Descriptive statistics are the ways to summarize the data set, to represent its tendencies in a concise form and/or describe them graphically.

7.1 Sample and population

We will usually refer to the given data set as a *sample* and denote its entries as $X_1, X_2, ..., X_n$. The objects whose measurements are represented by X_i are often called *experimental units* and are usually assumed to be sampled randomly from a larger *population* of interest. The probability distribution of X_i is then referred to as *population distribution*.

Definition 7.1. Population and sample

> *Population* is the collection of all objects of interest. *Sample* is the collection of objects from the population picked for the study.
>
> A *simple random sample* (SRS) is a sample for which each object in the population has the same probability to be picked as any other object, and is picked independently of any other object.

Example 7.1.

a) We would like to learn the public opinion regarding a tax reform. We set up phone interviews with $n = 1000$ people. Here, the population (which we really would like to learn about) is all U.S. adults, and the sample (which are the objects, or individuals we actually get), is the 1000 people contacted.

   For some really important matters, the U.S. Census Bureau tries to reach every single American, but this is practically impossible.

b) The gas mileage of a car is investigated. Suppose that we drive $n = 20$ times starting with a full tank of gas, until it's empty, and calculate the average gas mileage after each trip. Here, the population is *all potential trips* between fillups on this car to be made (under usual driving conditions) and the sample is the 20 trips actually made.

Usually, we require that our sample be a simple random sample (SRS) so that we can extend our findings to the entire population of interest. This means that no part of the population is preferentially selected for, or excluded from the study. It is amazing that, given proper sampling procedures, we can sometimes tell a lot about a large population after sampling only a small fraction of it!

Bias often occurs when the sample is not an SRS. For example, *self-selection bias* occurs when subjects volunteer for the study. Medical studies that pay for participation may attract lower-income volunteers. A questionnaire issued by a website will represent only the people that visit that website etc.

The ideal way to implement an SRS is to create a list of all objects in a population, and then use a random number generator to pick the objects to be sampled. In practice, this is usually very difficult to accomplish.

In the future, we will always assume that we are dealing with an SRS, unless otherwise noted. Thus, we will obtain a sequence of independent and identically distributed (IID) random variables $X_1, X_2, ..., X_n$ from the population distribution we are studying.

7.2 Graphical summaries

The most popular graphical summary for a numeric data set is a *histogram.*

Definition 7.2.

The histogram of the data set $X_1, X_2, \ldots, X_n$ is a bar chart representing the *classes* (or *bins*) on the x-axis and frequencies (or proportions) on the y-axis.

Bins should be of equal width so that all bars would visually be on the same level.[1] The construction of a histogram is easier to show by example.

Example 7.2.

Old Faithful is a famous geyser in Yellowstone National Park. The data recorded represent waiting times between eruptions (in minutes). There are $n = 272$ observations. The first ten observations are 79, 54, 74, 62, 85, 55, 88, 85, 51, 85. Using the bins 41-45, 46-50 etc we get

Bin	41-45	46-50	51-55	56-60	61-65	66-70	71-75
Count	4	22	33	24	14	10	27

Bin	76-80	81-85	86-90	91-95	96-100
Count	54	55	23	5	1

The choice of bins of course affects the appearance of a histogram. With too many bins, the graph becomes hard to read, and with too little bins, a lot of information is lost. We would generally recommend to use more bins for larger sample sizes; but not too many bins, so that the histogram keeps a smooth appearance. Some authors recommend the number of bins no higher than $\sqrt{n}$ where n is the sample size.

[1]Bins can be of unequal width but then some adjustment to their heights must be made.

Describing the shape of a histogram, we may note its features as being symmetric, or maybe skewed (left or right); having one "bulge" (*mode*) - that is, unimodal distribution, or two modes - that is, bimodal distribution etc. The Old Faithful data have bimodal shape. Some skewed histogram shapes are shown in Fig. 8.1

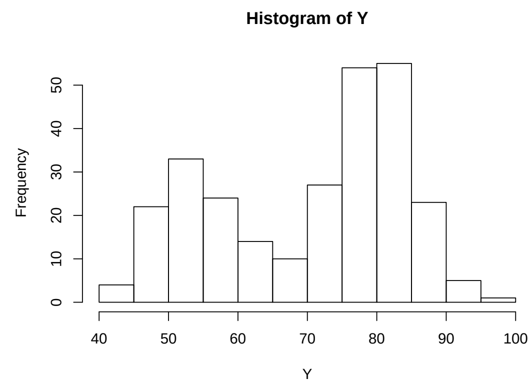


Figure 7.1: histogram of Old Faithful data

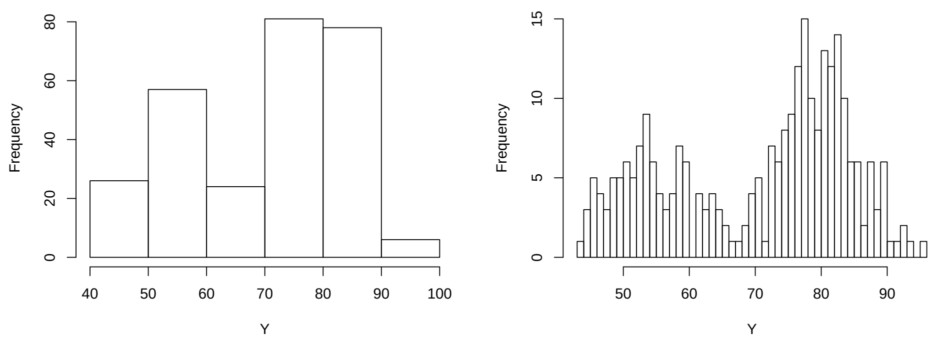


Figure 7.2: histograms of Old Faithful data: bins too wide, bins too narrow

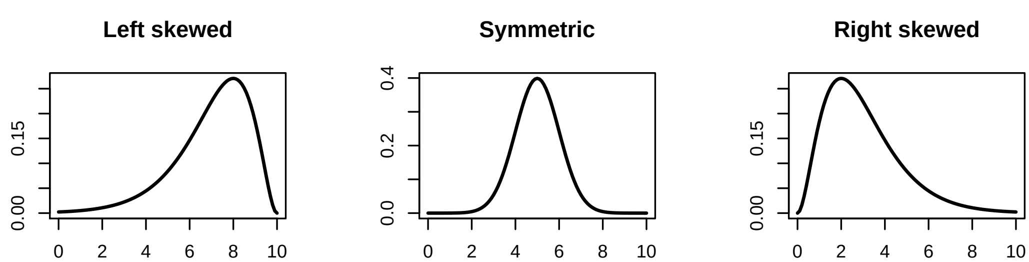


Figure 7.3: Symmetric and skewed shapes

7.3 Numerical summaries

7.3.1 Sample mean and variance

The easiest and most popular summary for a data set is its mean $\overline{X}$. The mean is a *measure of location* for the data set. We often need also a measure of spread. One such measure is the *sample standard deviation.*

Definition 7.3. Sample variance and standard deviation

The *sample variance* is denoted as S^2 and equals to

$$S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^n (X_i^2) - n\overline{X}^2}{n-1} \tag{7.1}$$

Sample standard deviation S is the square root of S^2.

A little algebra may show that both expressions in the formula (7.1) are equivalent. Denominator in the formula is $n-1$ which is called *degrees of freedom.* A simple explanation is that the calculation starts with n numbers and is then constrained by finding $\overline{X}$, thus $n-1$ degrees of freedom are left. Note that if $n = 1$ then the calculation of sample variance is not possible.

The sample mean and standard deviation are counterparts of the mean and standard deviation of a probability distribution. Further we will use them as the estimates of the unknown mean and standard deviation of a probability distribution (or a population).

Example 7.3.
The heights of last 8 US presidents are (in cm)[1] : 185, 182, 188, 188, 185, 177, 182, 193. Find the mean and standard deviation of these heights.

Solution. The average height is $\overline{X} = 185$. To make the calculations more compact, let's subtract 180 from each number, as it will not affect the standard deviation: $5, 2, 8, 8, 5, -3, 2, 13$, and $\overline{X} = 5$. Then, $\sum X_i^2 = 364$ and we get $S^2 = \dfrac{364 - 5^2(8)}{8-1} = 23.43$ and $S = \sqrt{23.43} = 4.84$. □

7.3.2 Percentiles

Definition 7.4.

The pth percentile (or quantile) of a data set is a number q such that $p\%$ of the entire sample are below this number. It can be calculated as $r = ((n+1)p/100)$th smallest number in the sample.

The algorithm for calculating pth percentile is then as follows.[2]

a) Order the sample, from smallest to largest, denote these as $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.
b) Calculate $r = (n+1)p/100$, let $k = \lfloor r \rfloor$ be the integer part of r.
c) If interpolation is desired, take $X_{(k)} + (r-k)[X_{(k+1)} - X_{(k)}]$,
   If interpolation is not needed, take $X_{(r^*)}$ where r^* is the rounded value of r.

Generally, if the sample size n is large, the interpolation is not needed.[3]

The 50-th percentile is known as *median.* It is, along with the mean, a measure of center of the data set.

---

[2]see e.g. http://www.itl.nist.gov/div898/handbook/prc/section2/prc252.htm

[3]Software note: different books and software packages may have different ways to interpret the fractional value of $(n+1)p/100$, so the percentile results might vary.

Example 7.4.
Back to the example of US presidents: find the median and 22nd percentile of the presidents' heights.

Solution. The ordered data are 177, 182, 182, 185, 185, 188, 188, 193. For $n = 8$ we have two "middle observations": ranked 4th and 5th, these are both 185. Thus, the median is 185 (accidentally we have seen that $\overline{X} = 185$ also).

To find 22nd percentile, take $r = (n+1)p = 9(0.22) = 1.98$, round it to 2. Then, take 2nd ranked observation, which is 182. □

Mean and median

The mean and median are popular measures of center. For a symmetric data set, both give roughly the same result. However, for a skewed data set, they might produce fairly different results. For the right-skewed distribution, `mean > median`, and for the left-skewed, `mean < median`.

The median is *resistant to outliers*. This means that the unusually high or low observations do not greatly affect the median. The mean $\overline{X}$ is not resistant to outliers.

Mean of a function

We can define the mean of any function g of our data as

$$\overline{g(X)} = \frac{g(X_1) + g(X_2) + ... + g(X_n)}{n}$$

Similarly to the properties of the expected values (see Theorem 3.2), we have the following properties:

a) $\overline{aX + b} = a\overline{X} + b$

b) but, generally, $\overline{g(X)} \neq g(\overline{X})$

c) For sample standard deviation, $S_{aX+b} = |a| S_X$

Exercises

7.1.
The temperature data one morning from different weather stations in the vicinity of Socorro were
71.9, 73.7, 72.3, 74.6, 72.8, 67.5, 72.0 (in $°F$)

a) Find the mean and standard deviation of temperatures

b) Find the median and 86th percentile.

c) Suppose that the last measurement came from Magdalena Ridge and became equal to 41.7 instead of 72.0. How will this affect the mean and the median, respectively?

d) Re-calculate the above answers if the temperature is expressed in Celcius. [**Hint:** you do not have to do it from scratch!]

7.2.

The heights of the last 20 US presidents are, in cm: 185, 182, 188, 188, 185, 177, 182, 193, 183, 179, 175, 188, 182, 178, 183, 180, 182, 178, 170, 180.

a) Make a histogram of the heights, choosing bins wisely.

b) Calculate mean and the median, compare. How do these relate to the shape of the histogram?

7.3.

The permeabilities of 12 oil pumping locations, in millidarcies, are: 0.07, 0.17, 0.06, 0.09, 0.17, 0.18, 0.04, 0.07, 0.02, 0.57, 0.71, 0.05.

a) Make a histogram of the permeabilities, choosing bins wisely.

b) Calculate mean and the median, compare. How do these relate to the shape of the histogram?

c) Find standard deviation of permeabilities.

7.4.

Several runners have completed a 1 mile race, with these results: 4.35, 4.51, 4.18, 4.56, 4.10, 3.75 (in minutes).

a) Find the average time of these runners.

b) Find the average speed (note: you will have to find each runner's individual speed, first).

c) Compare the answers to (a) and (b): why is mean speed **not** equal to the inverse of mean running time?

7.5.

Here are samples of temperature measurements at three cities (on random days).[m] They've been ordered for your convenience.

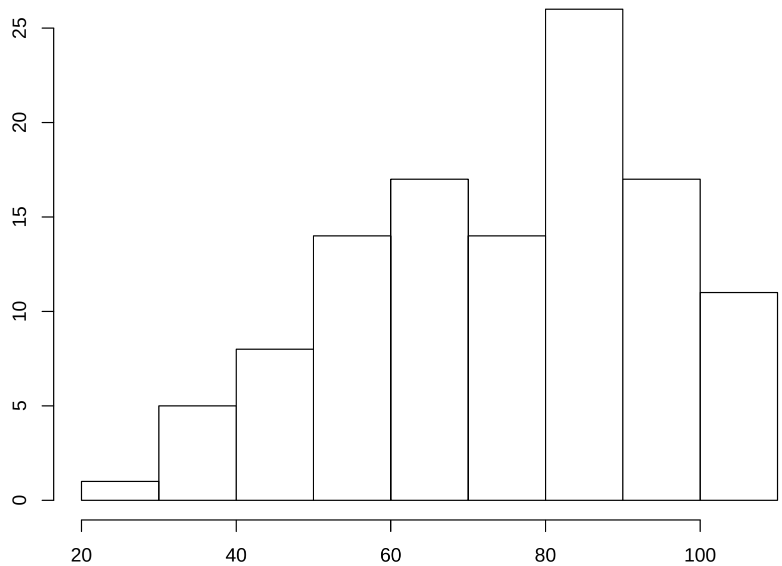
```
Albuquerque, NM
28 28 30 34 36 42 48 48 52 55 56 65 66 69 70 74 76 77 80 83

San Francisco, CA
46 47 48 51 52 52 53 55 57 58 58 58 59 60 60 61 61 62 62 64

Anchorage, AK
9 15 24 26 28 30 32 32 33 33 34 38 41 48 52 54 55 58 59 63
```

Plot the histograms overlayed over one another (using different colors, maybe) or directly above one another with a common scale. Also, calculate the means and standard deviations. Based on these numbers and graphs, compare the climates of these 3 cities.

7.6.
The following histogram was obtained for the distribution of 113 final grades in a math course.



a) Estimate the fraction of grades below 60.

b) Estimate the median grade (by the way, the average was 75.09).

c) Estimate the standard deviation visually [**Hint:** Recall the 68-95% rule]

d) Comment on the shape of the plot.

Chapter 8

Statistical inference

8.1 Introduction

In previous sections we emphasized properties of the sample mean. In this section we will discuss the problem of estimation of population parameters, in general. A **point estimate** of some population parameter θ is a single value $\hat{\theta}$ of a statistic. For example, the value $\overline{X}$ is the point estimate of population parameter μ. Similarly, $\hat{p} = \dfrac{X}{n}$ is a point estimate of the true proportion p in a binomial experiment.

Statistical inference deals with the question: can we infer something about the unknown population parameters (e.g., μ, σ or p)? Two major tools for statistical inference are *confidence intervals* (they complement a point estimate with a margin of error) and *hypothesis tests* that try to prove some statement about the parameters.

8.1.1 Unbiased Estimation

What are the properties of desirable estimators? We would like the sampling distribution of $\hat{\theta}$ to have a mean equal to the parameter estimated. An estimator possessing this property is said to be **unbiased**.

Definition 8.1.

A statistic $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if

$$\mathbb{E}\,(\hat{\theta}) = \theta.$$

The unbiased estimators are correct "on average", while actual samples yield results higher or lower than the true value of the parameter

On the other hand, biased estimators would consistently overestimate or underestimate the target parameter.

Example 8.1.
We have seen (p. 117) that $\mathbb{E}\,(\overline{X}) = \mu$, therefore $\overline{X}$ is an unbiased estimate of μ. □

Example 8.2.
One reason that the sample variance $S^2 = \sum (X_i - \overline{X})^2/(n-1)$ is divided by $n-1$ (instead of n) is the unbiasedness property. Indeed, it can be shown that $\mathbb{E}\,(S^2) = \sigma^2$. However, $\mathbb{E}\,(S) \neq \sigma$. □

8.2 Confidence intervals

The **confidence interval (CI)** or **interval estimate** is an interval within which we would expect to find the "true" value of the parameter.

Interval estimates, say, for population mean, are often desirable because the point estimate $\overline{X}$ varies from sample to sample. Instead of a single estimate for the mean, a confidence interval generates a lower and an upper bound for the mean. The interval estimate provides a measure of uncertainty in our estimate of the true mean μ. The narrower the interval, the more precise is our estimate.

Confidence limits are evaluated in terms of a confidence level.[1] Although the choice of confidence level is somewhat arbitrary, in practice 90%, 95%, and 99% intervals are often used, with 95% being the most commonly used.

An easy way to find CI's is when an estimate $\hat{\theta}$ is *asympotically Normal*, that is $\hat{\theta} \approx \mathcal{N}(\theta, \sigma^2_{\hat{\theta}})$. In that case,[2] there is about 95% chance that

$$\hat{\theta} - 2\sigma_{\hat{\theta}} < \theta < \hat{\theta} + 2\sigma_{\hat{\theta}}$$

The following is an example of how this works.

Theorem 8.1. CI for the mean

If $\overline{X}$ is the mean of a random sample of size n from a population with known variance σ^2, an approximate $(1-\alpha)100\%$ confidence interval[a] for μ is given by

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \tag{8.1}$$

where $z_{\alpha/2}$ is the Z-value leaving an area of $\alpha/2$ to the right.

[a] The CI will be exact if the population distribution is Normal

Proof. Central Limit Theorem (CLT) claims that, regardless of the initial distribution, the sample mean $\overline{X} = (X_1 + ... + X_n)/n$ will be approximately Normal:

$$\overline{X} \approx \text{ Normal } (\mu, \sigma^2/n)$$

for n reasonably large (usually $n \geq 30$ is considered enough).

Suppose that a confidence level $C = 100\%(1-\alpha)$ is given. Then, find $z_{\alpha/2}$ such that

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \quad Z \text{ is a standard Normal RV}$$

Due to the symmetry of Z-distribution, we need to find the z-value with the upper tail probability $\alpha/2$. That is, table area $\texttt{TA}(z_{\alpha/2}) = 0.5 - \alpha/2$.

[1] On a technical note, a 95% confidence interval does not mean that there is a 95% probability that the interval contains the true mean. The interval computed from a given sample either contains the true mean or it does not. Instead, the level of confidence is associated with the method of calculating the interval. For example, for a 95% confidence interval, if many samples are collected and a confidence interval is computed for each, in the long run about 95% of these intervals would contain the true mean.

[2] In this case, $\hat{\theta}$ will also be an unbiased estimate of θ.

Then, using CLT, $Z \approx \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$, therefore

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Solving for μ, we obtain the result. □

Notes:

(a) If σ is unknown, it can replaced by S, the sample standard deviation, with no serious loss in accuracy for the large sample case. Later, we will discuss what happens for small samples.

(b) This CI (and many to follow) has the following structure

$$\overline{X} \pm m$$

where m is called *margin of error*.

Example 8.3.
The drying times, in hours, of a certain brand of latex paint are

```
3.4  2.5  4.8  2.9  3.6  2.8  3.3  5.6
3.7  2.8  4.4  4.0  5.2  3.0  4.8
```

Compute the 95% confidence interval for the mean drying time. Assume that $\sigma = 1$.

Solution. We compute $\overline{X} = 3.79$ and $z_{\alpha/2} = 1.96$
($\alpha = 0.05$, upper-tail probability = 0.025, table area = 0.5 − 0.025 = 0.475)

Then, using (8.1), the 95% C.I. for the mean is

$$3.79 \pm 1(1.96)/\sqrt{15} = 3.79 \pm 0.51$$ □

Example 8.4.
The average zinc concentration recovered from a sample of zinc measurements in 36 different locations in the river is found to be 2.6 milligrams per liter. Find the 95% and 99% confidence intervals for the mean zinc concentration μ. Assume that the population standard deviation is 0.3.

Solution. The point estimate of μ is $\overline{X} = 2.6$. For 95% confidence, $z_{\alpha/2} = 1.96$. Hence, the 95% confidence interval is

$$2.6 - 1.96\frac{0.3}{\sqrt{36}} < \mu < 2.6 + 1.96\frac{0.3}{\sqrt{36}} = (2.50, 2.70)$$

For a 99% confidence, $z_{\alpha/2} = 2.575$ and hence the 99% confidence interval is

$$2.6 - 2.575\frac{0.3}{\sqrt{36}} < \mu < 2.6 + 2.575\frac{0.3}{\sqrt{36}} = (2.47, 2.73)$$

We see that a wider interval is required to estimate μ with a higher degree of confidence. □

Example 8.5.
An important property of plastic clays is the amount of shrinkage on drying. For a certain type of plastic clay 45 test specimens showed an average shrinkage percentage of 18.4 and a standard deviation of 1.2. Estimate the "true" average shrinkage μ for clays of this type with a 95% confidence interval.

Solution. For these data, a point estimate of μ is $\overline{X} = 18.4$. The sample standard deviation is $S = 1.2$. Since n is fairly large, we can replace σ by S.
Hence, 95% confidence interval for μ is

$$18.4 - 1.96\frac{1.2}{\sqrt{45}} < \mu < 18.4 + 1.96\frac{1.2}{\sqrt{45}} = (18.05, 18.75)$$

Thus we are 95% confident that the true mean lies between 18.05 and 18.75. □

Sample size calculations

In practice, another problem often arises: how many data should be collected to determine an unknown parameter with a given accuracy? That is, let m be the desired size of the margin of error, for a given confidence level $100\%(1-\alpha)$

$$m = \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.2}$$

What is the sample size n to achieve this goal?

To do this, assume that some estimate of σ is available. Then, solving for n,

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{m}\right)^2$$

Example 8.6.
We would like to estimate the pH of a certain type of soil to within 0.1, with 99% confidence. From past experience, we know that the soils of this type usually have pH in the 5 to 7 range. Find the sample size necessary to achieve our goal.

Solution. Let us take the reported 5 to 7 range as the $\pm 2\sigma$ range. This way, the crude estimate of σ is $(7-5)/4 = 0.5$. For 99% confidence, we find the upper tail area $\alpha/2 = (1-0.99)/2 = 0.005$, thus $z_{\alpha/2} = 2.576$, and $n = (2.576 \times 0.5/0.1)^2 \approx 166$ □

Exercises

8.1.
We toss a coin n times, let X be the total number of Heads observed. Let the probability of a Head in a single toss be p. Is $\hat{p} = X/n$ an unbiased estimate of p? Explain your reasoning.

8.2.
In a school district, they would like to estimate the average reading rate of first-graders. After selecting a random sample of $n = 65$ readers, they obtained sample mean of 53.4 words per minute (wpm), and standard deviation of 33.9 wpm.[n] Calculate a 98% confidence interval for the average reading rate of all first-graders in the district.

8.3.
A random sample of 200 calls initiated while driving had a mean duration of 3.5 minutes with standard deviation 2.2 minutes. Find a 99% confidence interval for the mean duration of telephone calls initiated while driving.

8.4.

a) Bursting strength of a certain brand of paper is supposed to have Normal distribution with $\mu = 150$ kPa and $\sigma = 15$ kPa. Give an interval that contains about 95% of all bursting strength values

b) Assuming now that the true μ and σ are unknown, the researchers collected a sample of $n = 100$ paper bags and measured their bursting strength. They obtained $\overline{X} = 148.4$ kPa and $S = 18.9$ kPa. Calculate the 95% C.I. for the mean bursting strength.

c) Sketch a Normal density curve with $\mu = 150$, $\sigma = 15$, with both of your intervals shown on the x-axis. Compare the intervals' widths.

8.5.
In determining the mean viscosity of a new type of motor oil, the lab needs to collect enough observations to approximate the mean within ± 0.2 SAE grade, with 96% confidence. The standard deviation typical for this type of measurement is 0.4. How many samples of motor oil should the lab test?

8.6.
The times to react to a pistol start were measured for a sample of 100 experienced swimmers, yielding a mean of 0.214 sec and standard deviation 0.036 sec. Find a 95% confidence interval for the average reaction time for the population of all experienced swimmers.

8.7.
A new petroleum extraction method was tested on 60 wells. The average improvement in total extraction was 18.3%. Assuming that standard deviation of the improvement was $\sigma = 10.3\%$, find the 96% CI for the "true" (i.e. all possible future wells) average improvement in total extraction by the new method.

8.8.

a) Show that, for $n = 2$, S^2 is an unbiased estimate of σ^2, that is, $\mathbb{E}\,(S^2) = \sigma^2$ [**Hint:** use the fact that $\mathbb{E}\,(X_i^2) = (\mathbb{E}\,X_i)^2 + Var(X_i)$]

b)$\star$ Show $\mathbb{E}\,(S^2) = \sigma^2$ for any $n \geq 2$.

8.3 Statistical hypotheses

Definition 8.2.

> A **statistical hypothesis** is an assertion or conjecture concerning one or more population parameters.

The goal of a statistical hypothesis test is to **make a decision** about an unknown parameter (or parameters). This decision is usually expressed in terms of rejecting or accepting a certain value of parameter or parameters.

Some common situations to consider:

- Is the coin fair? That is, we would like to test if $p = 1/2$ where $p = P(Heads)$.
- Is the new drug more effective than the old one? In this case, we would like to compare two parameters, e.g. the average effectiveness of the old drug versus the new one.

In making the decision, we will compare the statement (say, $p = 1/2$) with the available data and will reject the claim $p = 1/2$ if it contradicts the data. In the subsequent sections we will learn how to set up and test the hypotheses in various situations.

Null and alternative hypotheses

A statement like $p = 1/2$ is called the **Null hypothesis** (denoted by H_0). It expresses the idea that the parameter (or a function of parameters) is equal to some fixed value. For the coin example, it's

$$H_0 : p = 1/2$$

and for the drug example it's

$$H_0 : \mu_1 = \mu_2$$

where μ_1 is the mean effectiveness of the old drug compared to μ_2 for the new one. **Alternative hypothesis** (denoted by H_A) seeks to disprove the null. For example, we may consider *two-sided* alternatives

$$H_A : p \neq 1/2 \qquad \text{or, in the drug case, } H_A : \mu_1 \neq \mu_2$$

8.3.1 Hypothesis tests of a population mean

Steps of a Hypothesis Test

a) Null Hypothesis $H_0 : \mu = \mu_0$

b) Alternative Hypothesis $H_A : \mu \neq \mu_0$, or $H_A : \mu > \mu_0$, or $H_A : \mu < \mu_0$.

c) Critical value: $z_{\alpha/2}$ for two-tailed or z_α for one-tailed test, for some chosen *significance level* α. (Here, α is the false positive rate, i.e. how often you will reject H_0 that is, in fact, true.)

d) Test Statistic $z = \dfrac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}$

e) Decision Rule: Reject H_0 if

$$\begin{array}{rl} |z| > z_{\alpha/2} & \text{for two-tailed} \\ z > z_\alpha & \text{for right-tailed} \\ z < -z_\alpha & \text{for left-tailed} \end{array}$$

or, using p-value (see below), Reject H_0 when p-value $< \alpha$

f) Conclusion in the words of the problem.

A **null hypothesis** H_0 for the population mean μ is a statement that designates the value μ_0 for the population mean to be tested. It is associated with an **alternative hypothesis** H_A, which is a statement incompatible with the null. A **two-sided** (or **two-tailed**) hypothesis setup is

$$H_0 : \mu = \mu_0 \text{ versus } H_A : \mu \neq \mu_0$$

for a specified value of μ_0, and a **one-sided** (or **one-tailed**) hypothesis setup is either

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu > \mu_0 \qquad \text{(right-tailed test)}$$

or

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu < \mu_0 \qquad \text{(left-tailed test)}$$

Definition 8.3. P-values

A data set can be used to measure the plausibility of a null hypothesis H_0 through the calculation of a p-value.[a] The smaller the p-value, the less plausible is the null hypothesis.

Rejection Rule: Given the *significance level* α,

Reject H_0 when p-value $< \alpha$

otherwise **Accept** H_0.

[a]Do not confuse p-value with notation for proportion p

Calculation of P-values

For the two-tailed alternative hypothesis, P-value $= 2 \times\ P(Z > |z|)$.

For the right-tailed hypothesis, $H_A : \mu > \mu_0$, P-value $= P(Z > z)$

For the left-tailed hypothesis, $H_A : \mu < \mu_0$, P-value $= P(Z < z)$

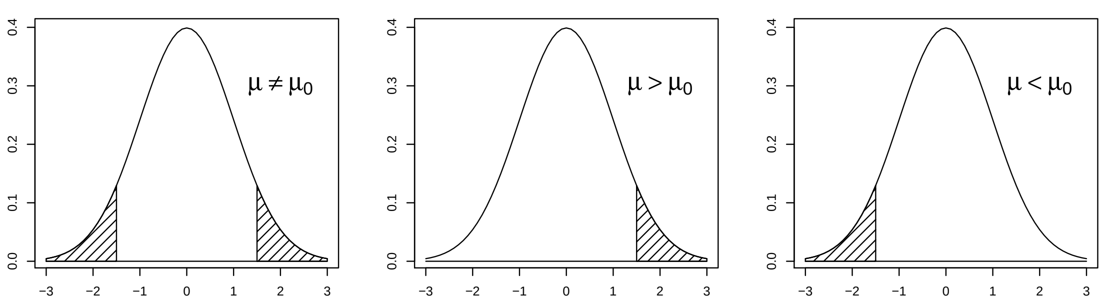


Figure 8.1: P-value calculation for different H_A

Picking significance level α

Since p-value is the probability of "extreme" results (tails) in your sample[3], the choice of α reflects our definition of "extreme". In fact, α is the proportion of samples for which H_0

[3]Contrary to a popular misconception, p-value is **not** the probability that H_0 is true.

is true, but nevertheless will be rejected by our test (known as *Type I error*, or proportion of *false positives*). Decreasing α decreases the proportion of false positives, but also makes it harder to reject H_0. Usually p-value < 0.01 is considered "strong evidence" against H_0, and p-value around 0.10 as "weak evidence". Which α to use as the threshold for rejecting H_0 may depend on how important it is to avoid false positives. Many people think that $\alpha = 0.05$ provides a good practical choice.

Example 8.7.
A manufacturer of sports equipment has developed a new synthetic fishing line that he claims has a mean breaking strength of 8.0 kg with a standard deviation of 0.5 kg. A random sample of 50 lines is tested and found to have a mean breaking strength of 7.80 kg. Test the hypothesis that $\mu = 8$ against the alternative that $\mu \neq 8$. Use $\alpha = 0.01$ level of significance.

Solution.

a) $H_0 : \mu = 8$

b) $H_A : \mu \neq 8$

c) $\alpha = 0.01$ and hence critical value $z_{\alpha/2} = 2.57$

d) Test statistic:
$$z = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} = \frac{\sqrt{50}(7.8 - 8)}{0.5} = -2.83$$

e) Decision: reject H_0 since $|-2.83| > 2.57$.

f) Conclusion: there is evidence that the mean breaking strength is **not** 8 kg (in fact, it's lower).

Decision based on P-value:
Since the test in this example is two-sided, the p-value is double the area.

$$\text{P-value} = P(|Z| > 2.83) = 2\,[0.5 - \mathtt{TA}(2.83)] = 2(0.5 - 0.4977) = 0.0046$$

which allows us to reject the null hypothesis that $\mu = 8$ kg at 0.01 level of significance. □

Example 8.8.
A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Solution.

a) $H_0 : \mu = 70$ years.

b) $H_A : \mu > 70$ years.

c) $\alpha = 0.05$ and $z_\alpha = 1.645$

d) Test statistic:
$$z = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} = \frac{\sqrt{100}(71.8 - 70)}{8.9} = 2.02$$

e) Decision: Reject H_0 if $2.02 > 1.645$, since $2.02 > 1.645$, we reject H_0.

f) Conclusion: We conclude that the mean life span today is greater than 70 years.

Decision based on P-value:
Since the test in this example is one-sided, the desired p-value is the area to the right of $z = 2.02$. Using Normal Table, we have

$$\text{P-value} = P(Z > 2.02) = 0.5 - 0.4783 = 0.0217.$$

Conclusion: Reject H_0. □

Example 8.9.
The nominal output voltage for a certain electrical circuit is 130V. A random sample of 40 independent readings on the voltage for this circuit gave a sample mean of 128.6V and a standard deviation of 2.1V. Test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. Use a 5% significance level.

Solution.

a) $H_0 : \mu = 130$

b) $H_A : \mu < 130$

c) $\alpha = 0.05$ and $z_\alpha = -1.645$

d) Test statistic: $$z = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} = \frac{\sqrt{40}(128.6 - 130)}{2.1} = -4.22$$

e) Decision: Reject H_0 since $-4.22 < -1.645$.

f) Conclusion: We conclude that the average output voltage is less than 130.

Decision based on p-value:

$$\text{P-value} = P(Z < -4.22) = (0.5 - 0.4990) = 0.001.$$

As a result, the evidence in favor of H_A is even stronger than that suggested by the 0.05 level of significance. (P-value is very small!) □

Exercises

8.9.
It is known that the average height of US adult males is about 173 cm, with standard deviation of about 6 cm. Assume that the heights follow Normal distribution.

Referring to Exercise 7.2, the average height of 20 last US presidents was 181.9 cm. Are the presidents taller than the average? Test at the level $\alpha = 0.05$ and also compute the p-value.

8.10.
In an industrial process, nanotubes should have the average diameter of 5 angstrom. The typical variance for the nanotubes obtained in this process is 0.2 angstrom.

The sample of 50 nanotubes was studied with the observed average diameter of 5.12 angstrom. Is the evidence that the process average is different from 5 angstrom?

8.11.

A biologist knows that, under normal conditions, the average length of a leaf of a certain full-grown plant is 4 inches, with the standard deviation of 0.6 inches. A sample of 45 leaves from the plants that were given a new type of plant food had an average length of 4.2 inches. Is there reason to believe that the new plant food is responsible for a change in the average growth of leaves? Use $\alpha = 0.02$.

8.12.

In the situation of Exercise 8.7, test $H_0 : \mu = 0$ against $H_A : \mu > 0$. Test at the level $\alpha = 0.04$ and also compute the p-value.

8.13.

Is it more difficult to reject H_0 when the significance level is smaller? Suppose that the p-value for a test was 0.023. Would you reject H_0 at the level $\alpha = 0.05$? At $\alpha = 0.01$?

8.14.

It is well known that the "normal" human temperature is 98.6°. If a sample of 75 healthy adults is collected and the sample mean was 98.3°, can we claim that 98.6 is a plausible value for the mean temperature of all adults? Assume that $\sigma = 0.8°$. Make your decision based on the p-value.

8.4 The case of unknown σ

8.4.1 Confidence intervals

Frequently, we are attempting to estimate the mean of a population when the variance is unknown. Suppose that we have a random sample from a normal distribution, then the random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is said to have a (Student)° **T-distribution** with $n-1$ degrees of freedom. Here, S is the sample standard deviation.

With σ unknown, T should be used instead of Z to construct a confidence interval for μ. The procedure is same as for known σ except that σ is replaced by S and the standard normal distribution is replaced by the T-distribution.

T-distribution is also symmetric, but has somewhat "heavier tails" than Z. This is because of extra uncertainty of not knowing σ.

Definition 8.4. CI for mean, unknown σ

If $\overline{X}$ and S are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $(1-\alpha)100\%$ confidence interval for μ is

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $n-1$ degrees of freedom leaving an area of $\alpha/2$ to the right. (See Table B.)

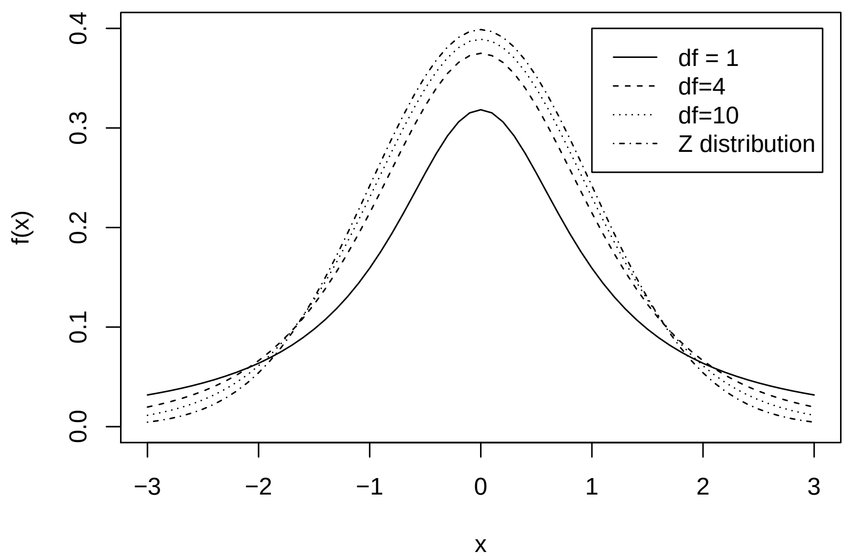


Figure 8.2: T distribution for different values of df = degrees of freedom

Normality assumption becomes more important as n gets smaller. As a practical rule, we will not trust the confidence intervals based on small samples (generally, $n < 30$) that are strongly skewed or have outliers.

On the other hand, we already noted that for large n we could simply use Z-distribution for the C.I. calculation. This is justified by the fact that $t_{\alpha/2}$ values approach $z_{\alpha/2}$ values as n gets larger.

Example 8.10.

The contents of 7 similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6 liters. Find a 95% confidence interval for the mean volume of all such containers, assuming an approximate normal distribution.

Solution. The sample mean and standard deviation for the given data are $\overline{X} = 10.0$ and S = 0.283. Using the t-Table, we find $t_{0.025} = 2.447$ for 6 degrees of freedom. Hence the 95% confidence interval for μ is

$$10.0 - 2.447\frac{0.283}{\sqrt{7}} < \mu < 10.0 + 2.447\frac{0.283}{\sqrt{7}},$$

which reduces to $9.74 < \mu < 10.26$ □

Example 8.11.

A random sample of 12 graduates of a certain secretarial school typed an average of 79.3 words per minute (wpm) with a standard deviation of 7.8 wpm. Assuming a normal distribution for the number of words typed per minute, find a 99% confidence interval for the average typing speed for all graduates of this school.

Solution. The sample mean and standard deviation for the given data are $\overline{X} = 79.3$ and S = 7.8. Using the t-Table, we find $t_{0.005} = 3.106$ with 11 degrees of freedom. Hence the

95% confidence interval for μ is

$$79.3 - 3.106\frac{7.8}{\sqrt{12}} < \mu < 79.3 + 3.106\frac{7.8}{\sqrt{12}},$$

which reduces to $72.31 < \mu < 86.30$.
We are 99% confident that the interval 72.31 to 86.30 includes the true average typing speed for all graduates. □

8.4.2 Hypothesis test

When sample sizes are small and population variance is unknown, use the test statistic

$$t = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S},$$

with $n-1$ degrees of freedom.

Steps of a Hypothesis Test

a) Null Hypothesis $H_0 : \mu = \mu_0$

b) Alternative Hypothesis $H_A : \mu \neq \mu_0$, or $H_A : \mu > \mu_0$, or $H_A : \mu < \mu_0$.

c) Critical value: $t_{\alpha/2}$ for two-tailed or t_α for one-tailed test.

d) Test Statistic $t = \dfrac{\sqrt{n}(\overline{X} - \mu_0)}{S}$ with $n-1$ degrees of freedom

e) Decision Rule: Reject H_0 if

$$\begin{aligned} |t| > t_{\alpha/2} &\quad \text{for two-tailed} \\ t > t_\alpha &\quad \text{for right-tailed} \\ t < -t_\alpha &\quad \text{for left-tailed} \end{aligned}$$

or, using p-value, Reject H_0 when p-value $< \alpha$

f) Conclusion.

Example 8.12.
Engine oil was stated to have the mean viscosity of $\mu_0 = 85.0$. A sample of $n = 25$ viscosity measurements resulted in a sample mean of $\overline{X} = 88.3$ and a sample standard deviation of $S = 7.49$. What is the evidence that the mean viscosity is not as stated? Use $\alpha = 0.1$.

Solution.

a) $H_0 : \mu = 85.0$

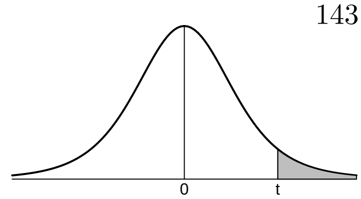
b) $H_A : \mu \neq 85.0$

c) $\alpha = 0.1$ and $t_{\alpha/2} = 1.711$ with 24 degrees of freedom.

d) Test statistic:

$$t = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} = \frac{\sqrt{25}(88.3 - 85.0)}{7.49} = 2.203$$

Table B: Critical points of the t-distribution



Degrees of freedom	Upper tail probability						
	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

e) Decision: Reject H_0 since $2.203 > 1.711$.

f) Conclusion: We conclude that the average viscosity is not equal to 85.0

Decision based on P-value:
Since the test in this example is two sided, the desired p-value is twice the tail area. Therefore, using software[4] with `df` = 24, we have

$$\text{P-value} = 2 \times P(T > 2.203) = 2(0.0187) = 0.0374,$$

which allows us to reject the null hypothesis that $\mu = 85$ at 0.1 level of significance. If we used Table B, we would locate 2.203 between two table values 2.064 and 2.492, concluding that the p-value is between $2(0.025) = 0.05$ and $2(0.01) = 0.02$ and reach the same conclusion.

Conclusion: In summary, we conclude that there is fairly strong evidence that the mean viscosity is not equal to 85.0 □

Example 8.13.

A sample of $n = 20$ cars driven under varying highway conditions achieved fuel efficiencies with a sample mean of $\overline{X} = 34.271$ miles per gallon (mpg) and a sample standard deviation of $S = 2.915$ mpg. Test the hypothesis that the average highway mpg is less than 35 with $\alpha = 0.05$.

Solution.

a) $H_0 : \mu = 35.0$

b) $H_A : \mu < 35.0$

c) $\alpha = 0.05$ and $t_\alpha = 1.729$ with 19 degrees of freedom.

d) Test statistic:

$$t = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} = \frac{\sqrt{20}(34.271 - 35.0)}{2.915} = -1.119$$

e) Decision: since $-1.119 > -1.729$, we do not reject H_0.

f) Conclusion: There is no evidence that the average highway mpg is any less than 35.0

Decision based on P-value:

$$\text{P-value} = P(T < -1.119) = P(T > 1.119) > 0.10,$$

(using $df = 19$ and critical point $t = 1.328$, which corresponds to the upper-tail area 0.10), thus p-value $> \alpha = 0.05$, do not reject H_0. □

[4] For example, R syntax `1 - pt(2.203, df = 24)`

8.4.3 Connection between Hypothesis tests and C.I.'s

We can test a two-sided hypothesis

$$H_0 : \mu = \mu_0 \text{ vs. } H_A : \mu \neq \mu_0$$

at the level α, using a confidence interval with the confidence level $100\%(1-\alpha)$. If we found the $100\%(1-\alpha)$ C.I. for the mean μ, and μ_0 belongs to it, we accept H_0, otherwise we reject H_0.

This way, the C.I. is interpreted as the range of "plausible" values for μ. The false positive rate in this case will be equal to $\alpha = 1 - C/100\%$

Example 8.14.
Reconsider Example 8.12. There, we had to test $H_0 : \mu = 85.0$ with the data $n = 25$, $\overline{X} = 88.3$ and $S = 7.49$, at the level $\alpha = 0.1$. Is there evidence that the mean average viscosity is not 85.0?

Solution. If we calculate a 90% C.I. ($90\% = 100\%(1-\alpha)$), we get

$$88.3 \pm 1.711 \frac{7.49}{\sqrt{25}} = 88.3 \pm 2.6 \text{ or } (85.7, 90.9)$$

Since 85.0 does not belong to this interval, there is evidence that the "true" mean viscosity is **not** 85.0 (in fact, it's higher).
We arrived at the same conclusion as in Example 8.12. □

8.4.4 Statistical significance vs Practical significance

Statistical significance sometimes has little to do with practical significance. Statistical significance (i.e. a small p-value) is only concerned with the amount of evidence to reject H_0. It does not directly reflect the size of the effect itself. Confidence intervals are more suitable for that.

For example, in testing the effect of a new medication for lowering cholesterol, we might find that the confidence interval for the average decrease μ is $(1.2, 2.8)$ units (mg/dL). Since the C.I. has positive values we proved $H_A : \mu > 0$. However, the decrease of 1.2 to 2.8 units might be too small in practical terms to justify developing this new drug.

Exercises

8.15.
In determining the gas mileage of a new model of hybrid car, the independent research company collected information from 14 randomly selected drivers. They obtained the sample mean of 38.4 mpg, with the standard deviation of 5.2 mpg. Obtain a 99% C.I. for μ.
What is the meaning of μ in this problem? What assumptions are necessary for your C.I. to be correct?

8.16.

This problem is based on the well-known Newcomb data set for the speed of light.[p] It contains the measurements (in nanoseconds) it took the light to bounce inside a network of mirrors. The numbers given are the time recorded minus 24,800 ns. We will only use the first ten values.

```
28  26  33  24  34  -44  27  16  40  -2
```

Some mishaps in the experimental procedure led to the two unusually low values (-44 and -2). Calculate the 95% C.I.'s for the mean in case when

a) all the values are used

b) the two outliers are removed

Which of the intervals will you trust more and why?

8.17.

The following data were collected for salinity of water from a sample of municipal sources (in parts per thousand)

```
0.5  0.5  0.6  0.6  0.8  0.8  0.8  0.9  1.0  1.1  1.3
```

Find a 98% confidence interval for the average salinity in all municipal sources in the sampling area.

8.18.

A job placement director claims that mean starting salary for nurses is \$25 per hour. A random sample of 10 nurses' salaries has a mean \$21.6 and a standard deviation of \$4.7 per hour. Is there enough evidence to reject the director's claim at $\alpha = 0.01$?

Repeat the exercise for the following results: sample mean \$21.6 and a standard deviation of \$0.47. What is your answer now? What can you conclude about the role of "noise" (standard deviation) in statistical testing?

8.19.

Refer to the Exercise 8.14. If, now, a sample of only 25 adults is collected and the sample mean was still 98.3°, with sample standard deviation $S = 0.8°$, what will your conclusion be? Comparing to the answer to Exercise 8.14, what can you conclude about the role of sample size in statistical testing?

8.20.

Suppose that 95% CI for the mean of a large sample was computed and equaled $[10.15, 10.83]$. What will be your decision about the hypothesis $H_0 : \mu = 10$ vs $H_A : \mu \neq 10$ at 5% level of significance? At 10% level? At 1% level?

8.21.

For the situation in Example 8.7 (fishing line strength), test the hypotheses using the C.I. approach.

8.5 C.I. and hypothesis tests for comparing two population means

Two-sample problems:

- The goal of inference is to compare the response in two groups.
- Each group is considered to be a sample from a distinct population.
- The responses in each group are independent of those in the other group.

Suppose that we have two independent samples, from two distinct populations. Here is the notation that we will use to describe the two populations:

population	Variable	Mean	Standard deviation
1	X_1	μ_1	σ_1
2	X_2	μ_2	σ_2

We want to compare the two population means, either by giving a confidence interval for $\mu_1 - \mu_2$ or by testing the hypothesis of difference, $H_0 : \mu_1 = \mu_2$. Inference is based on two independent random samples. Here is the notation that describes the samples:

sample	sample size	sample mean	sample st.dev.
1	n_1	$\overline{X}_1$	S_1
2	n_2	$\overline{X}_2$	S_2

If independent samples of size n_1 and n_2 are drawn at random from two populations, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, the sampling distribution of the differences of the means $\overline{X}_1 - \overline{X}_2$, is normally distributed with mean $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$ and variance $\sigma_D^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2$. Then, the two-sample Z statistic

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_D}$$

has the standard normal $N(0,1)$ sampling distribution.

Usually, population standard deviations σ_1 and σ_2 are not known. We estimate them by using sample standard deviations S_1 and S_2. But then the Z-statistic will turn into (approximately) T-statistic, with degrees of freedom equal to the smaller of $n_1 - 1$ or $n_2 - 1$.[5]

Further, if we are testing $H_0 : \mu_1 = \mu_2$, then $\mu_1 - \mu_2 = 0$. Thus, we obtain the confidence intervals and hypothesis tests for $\mu_1 - \mu_2$.

The $100\%(1-\alpha)$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \qquad T \text{ has } \mathtt{df} = \min(n_1, n_2) - 1 \tag{8.3}$$

[5] Picking $\mathtt{df} = \min(n_1, n_2) - 1$ is actually *conservative* (leads to an over-estimation of p-value) and there are other more complicated formulas for the number of degrees of freedom that may be used in different books and software packages.

Steps of a Hypothesis Test

a) Null Hypothesis $H_0 : \mu_1 = \mu_2$

b) Alternative Hypothesis $H_A : \mu_1 \neq \mu_2$, or $H_A : \mu_1 > \mu_2$, or $H_A : \mu_1 < \mu_2$.

c) Critical value: $t_{\alpha/2}$ for two-tailed or t_α for one-tailed test, for some chosen *significance level* α.

d) Test Statistic $t = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$

e) Decision Rule: Reject H_0 if

$$\begin{array}{cc} |t| > t_{\alpha/2} & \text{for two-tailed} \\ t > t_\alpha & \text{for right-tailed} \\ t < -t_\alpha & \text{for left-tailed} \end{array}$$

or, using p-value, Reject H_0 when p-value $< \alpha$.
P-value is calculated similarly to 1-sample T-test, but now with
$\texttt{df} = \min(n_1, n_2) - 1$.

f) Conclusion in the words of the problem.

Example 8.15.
A study of iron deficiency among infants compared samples of infants following different feeding regimens. One group contained breast-fed infants, while the other group were fed a standard baby formula without any iron supplements. Here are the data on blood hemoglobin levels at 12 months of age:

Group	n	$\overline{X}$	s
Breast-fed	23	13.3	1.7
Formula	19	12.4	1.8

(a) Is there significant evidence that the mean hemoglobin level is higher among breast-fed babies?
(b) Give a 95% confidence interval for the mean difference in hemoglobin level between the two populations of infants.

Solution. (a) $H_0 : \mu_1 - \mu_2 = 0$ vs $H_A : \mu_1 - \mu_2 > 0$, where μ_1 is the mean of the Breast-fed population and μ_2 is the mean of the Formula population. The test statistic is

$$t = \frac{13.3 - 12.4}{\sqrt{\frac{1.7^2}{23} + \frac{1.8^2}{19}}} = \frac{0.9}{0.544} = 1.654$$

with 18 degrees of freedom. The p-value is $P(T > 1.654) = 0.058$ using software. Using Table B, we see that 1.654 is between table values 1.330 and 1.734, which gives upper-tail probability between 0.05 and 0.10. This is not quite significant at 5% level.
(b) The 95% confidence interval is

$$0.9 \pm 2.101(0.544) = 0.9 \pm 1.1429 = (-0.2429, 2.0429)$$

□

Standard Error

All previous formulas involving t-distribution have a common structure. For example, (8.3) can be re-written as

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2}\, \mathtt{SE}_{\overline{X}_1-\overline{X}_2},$$

where the quantity $\mathtt{SE}_{\overline{X}_1-\overline{X}_2} = \sqrt{S_1^2/n_1 + S_2^2/n_2}$ is called the *Standard Error.* Likewise, the one-sample confidence interval for the mean is

$$\overline{X} \pm t_{\alpha/2}\, \mathtt{SE}_{\overline{X}},$$

where $\mathtt{SE}_{\overline{X}} = s/\sqrt{n}$.

Likewise, the formulas for the t-statistic are

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\mathtt{SE}_{\overline{X}_1-\overline{X}_2}} \text{ for 2-sample, and } t = \frac{\overline{X} - \mu_0}{\mathtt{SE}_{\overline{X}}} \text{ for 1-sample situation.}$$

We will see a lot of similar structure in the CI and hypothesis testing formulas in the future. The value of standard error is often reported by the software when you request CI's or hypothesis tests.

8.5.1 Matched pairs

Sometimes, we are comparing data that come in pairs of matched observations. A good example of this are "before" and "after" studies. They present the measurement of some quantity for the same set of subjects before and after a certain treatment has been administered. Another example of this situation is twin studies for which pairs of identical twins are selected and one twin (at random) is given a treatment, while the other is serving as a *control* (that is, does not receive any treatment, or maybe receives a fake treatment, *placebo*, to eliminate psychological effects).

When the same (or somehow related) subjects are used, we should not consider the measurements independent. This is the *Matched Pairs* design. In this case, we would compute `Difference` = `Before` − `After` or `Treatment` − `Control` and just do a one-sample test for the mean difference.

Example 8.16.

The following are the left hyppocampus volumes (in cm^3) for a group of twin pairs, one is affected by schizophrenia, and the other is not[q]

Pair number	1	2	3	4	5	6	7	8	9	10	11	12
Unaffected	1.94	1.44	1.56	1.58	2.06	1.66	1.75	1.77	1.78	1.92	1.25	1.93
Affected	1.27	1.63	1.47	1.39	1.93	1.26	1.71	1.67	1.28	1.85	1.02	1.34
Difference	0.67	-0.19	0.09	0.19	0.13	0.40	0.04	0.10	0.50	0.07	0.23	0.59

Is there evidence that the LH volumes for schizophrenia-affected people are different from the unaffected ones?

Solution. Since the twins' LH volumes are clearly not independent (if one is large the other is likely to be large, too – positive correlation!), we cannot use the 2-sample procedure.

However, we can just compute the differences (Unaffected − Affected) and test for the mean difference to be equal to 0. That is,

$$H_0 : \mu = 0 \text{ versus } H_A : \mu \neq 0$$

where μ is the "true" average difference, and $\overline{X}, S$ are computed for the sample of differences.

Given that $\overline{X} = 0.235$ and $S = 0.254$, let's test these hypotheses at $\alpha = 0.10$. We obtain $t = (0.235 - 0)/(0.254/\sqrt{12}) = 3.20$. From the t-table with $df = 11$ we get p-value between $2(0.005) = 0.01$ and $2(0.001) = 0.002$. At $\alpha = 0.05$, we Reject H_0, thus stating that there is a significant difference between LH volumes of normal and schizophrenic people. □

Exercises

More exercises for this section are located at the end of this Chapter.

8.22.

In studying how humans pick random objects, the subjects were presented a population of rectangles and have used two different sampling methods. They then calculated the average areas of the sampled rectangles for each method. Their results were

	mean	st.dev.	n
Method 1	10.8	4.0	16
Method 2	6.1	2.3	16

Calculate the 99% C.I. for the difference of "true" means by the two methods. Is there evidence that the two methods produce different results?

8.23.

The sports research lab studies the effects of swimming on maximal volume of oxygen uptake.

For 8 volunteers, the maximal oxygen uptake was measured before and after the 6-week swimming program. The results are as follows:

Before	2.1	3.3	2.0	1.9	3.5	2.2	3.1	2.4
After	2.7	3.5	2.8	2.3	3.2	2.1	3.6	2.9

Is there evidence that the swimming program has increased the maximal oxygen uptake?

8.24.

Visitors to an electronics website rated their satisfaction with two models of printers/scanners, on the scale of 1 to 5. The following statistics were obtained:

	n	mean	st.dev.
Model A	31	3.6	1.5
Model B	65	4.2	0.9

At the level of 5%, test the hypothesis that both printers would have the same average rating in the general population, that is, $H_0 : \mu_A = \mu_B$. Also, calculate the 95% confidence interval for the mean difference $\mu_A - \mu_B$.

8.6 Inference for Proportions

8.6.1 Confidence interval for population proportion

In this Chapter, we will consider estimating the proportion p of items of certain type, or maybe some probability p. The unknown population proportion p is estimated by the **sample proportion**

$$\hat{p} = \frac{X}{n}.$$

We know (from CLT, Section 6.4) that if the sample size is sufficiently large, $\hat{p}$ has approximately normal distribution, with mean $\mathbb{E}(\hat{p}) = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. Based on this, we obtain the confidence intervals and hypothesis tests for proportion.

Theorem 8.2. CI for proportion

For a random sample of size n from a large population with unknown proportion p of successes, the $(1-\alpha)100\%$ confidence interval for p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

8.6.2 Test for a single proportion

To test the hypothesis $H_0 : p = p_0$, use the z-statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

In terms of a standard normal Z, the approximate p-value for a test of H_0 is

$P(Z > z)$ against $H_A : p > p_0$,
$P(Z < z)$ against $H_A : p < p_0$,
$2P(Z > |z|)$ against $H_A : p \neq p_0$.

In practice, Normal approximation works well when both X and $n - X$ are at least 10.

Example 8.17.
The French naturalist Count Buffon once tossed a coin 4040 times and obtained 2048 heads. Test the hypothesis that the coin was balanced.

Solution. To assess whether the data provide evidence that the coin was not balanced, we test $H_0 : p = 0.5$ versus $H_A : p \neq 0.5$.

The test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.5069 - 0.50}{\sqrt{0.50(1-0.5)/4040}} = 0.88$$

From Z chart we find $P(Z < 0.88) = 0.8106$. Therefore, the p-value is $2(1 - 0.8106) = 0.38$. The data are compatible with balanced coin hypothesis.

Now we will calculate a 99% confidence interval for p. The $z_{\alpha/2} = 2.576$ from the normal table. Hence, the 99% CI for p is

$$\begin{aligned}\hat{p} &= 0.5069 \pm 2.576\sqrt{\frac{(0.5069)(1-0.5069)}{4040}} = 0.5069 \pm (2.576)(0.00786)\\ &= 0.5069 \pm 0.0202 = (0.4867, 0.5271)\end{aligned}$$

□

Sample size computation

To set up a study (e.g. opinion poll) with a guarantee not to exceed a certain maximum amount of error, we can solve for n in the formula for error margin m

$$m = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}, \quad \text{therefore } n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{m}\right)^2$$

Since $\hat{p}$ is not known prior to the study (a "Catch-22" situation), we might try to find n that will guarantee the desired maximum error margin m, no matter what p is. It turns out that using $p = 1/2$ is the worst possible case, i.e. produces the maximum margin of error.

Thus, we should use

$$n = \frac{1}{4}\left(\frac{z_{\alpha/2}}{m}\right)^2 \text{ if } p \text{ is completely unknown, or} \qquad n = p^*(1-p^*)\left(\frac{z_{\alpha/2}}{m}\right)^2,$$

if some estimate p^* of p is available.

Example 8.18.
How many people should be polled in order to provide a 98% margin of error equal to $\pm 1\%$,

a) assuming no knowledge of p

b) assuming that the proportion of interest should not exceed 0.1.

Solution. a) Since we do not have a prior knowledge of p, use

$$n = \frac{1}{4}\left(\frac{z_{\alpha/2}}{m}\right)^2 = \frac{1}{4}\left(\frac{2.33}{0.01}\right)^2 = 13572 \text{ people!}$$

Note that we converted m from 1% to 0.01

b) Now, making similar calculations, we get

$$n = (0.1)(1-0.1)\left(\frac{z_{\alpha/2}}{m}\right)^2 = 0.09\left(\frac{2.33}{0.01}\right)^2 = 4883$$

That is, the required sample size is much smaller.

□

8.6.3 Comparing two proportions*

We will call the two groups being compared Population 1 and Population 2, with population proportions of successes p_1 and p_2. Here is the notation we will use in this section:

Population	pop. prop.	sample	# successes	sample prop.
1	p_1	n_1	X_1	$\hat{p}_1 = X_1/n_1$
2	p_2	n_2	X_2	$\hat{p}_2 = X_2/n_2$

To compare the two proportions, we use the difference between the two sample proportions: $\hat{p}_1 - \hat{p}_2$. Therefore, when n_1 and n_2 are large, $\hat{p}_1 - \hat{p}_2$ is approximately normal with mean $\mu = p_1 - p_2$ and standard deviation $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$. Note that for unknown p_1 and p_2 we replace them by $\hat{p}_1$ and $\hat{p}_2$ respectively.

Definition 8.5. Inference for two proportions

The $(1-\alpha)100\%$ confidence interval for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

To test the hypothesis $H_0 : p_1 - p_2 = 0$, we use the test statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\mathtt{SE}_{\hat{p}}},$$

where $\quad \mathtt{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad$ and $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$.

Example 8.19.
To test the effectiveness of a new pain relieving drug, 80 patients at a clinic were given a pill containing the drug and 80 others were given a placebo. At the 0.01 level of significance, what can we conclude about the effectiveness of the drug if the first group 56 of the patients felt a beneficial effect while 38 out of those who received placebo felt a beneficial effect?

Solution. $H_0 : p_1 - p_2 = 0$ and $H_A : p_1 - p_2 > 0$

$z = 2.89$, where $\hat{p}_1 = \dfrac{56}{80} = 0.7$ and $\hat{p}_2 = \dfrac{38}{80} = 0.475$ and $\hat{p} = \dfrac{56+38}{80+80} = 0.5875$

P-value $= P(Z > 2.89) = 0.0019$

Since the p-value is less than 0.01, the null hypothesis must be rejected, so we conclude that the drug is effective. □

Exercises

8.25.
A nutritionist claims that at 75% of the preschool children in a certain country have protein deficient diets. A sample survey reveals that 206 preschool children in a sample of 300 have protein deficient diets. Test the claim at the 0.02 level of significance. Also, compute a 98% confidence interval.

8.26.
In a survey of 200 office workers, 165 said they were interrupted three or more times an hour by phone messages, faxes etc. Find and interpret a 90% confidence interval for the population proportion of workers who are interrupted three or more times an hour.

8.27.
You would like to design a poll to determine what percent of your peers volunteer for charities. You have no clear idea of what the value of p is going to be like, and you'll be satisfied with the 90% margin of error equal to $\pm 10\%$. Find the sample size needed for your study.

8.28.
In an opinion poll, out of a sample of 300 people, 182 were in support of Proposition Z. At the level of 5%, test the hypothesis that more than half of population support Proposition Z. Also, find the p-value.

8.29.
In random samples of 200 tractors from one assembly line and 400 tractors from another, there were, respectively, 16 tractors and 20 tractors which required extensive adjustments before they could be shipped. At the 5% level of significance, can we conclude that there is a difference in the quality of the work of the two assembly lines?

8.30.
In a survey of customer satisfaction on `amazon.com`, 86 out of 120 customers of Supplier A gave it 5 stars, and 75 out of 136 customers of Supplier B gave it 5 stars. Is there evidence that customers are more satisfied with one supplier than the other? Also, compute a 95% confidence interval for the difference of two proportions.

Chapter Exercises

For each of the questions involving hypothesis tests, state the null and alternative hypotheses, compute the test statistic, determine the p-value, make the decision and summarize the results in plain English. Use $\alpha = 0.05$ unless otherwise specified.

8.31.
Two brands of batteries are tested and their voltages are compared. The summary statistics are below. Find and interpret a 95% confidence interval for the true difference in means.

	mean	st.dev.	n
Brand 1	9.2	0.3	25
Brand 2	8.9	0.6	27

8.32.
You are studying yield of a new variety of tomato. In the past, yields of similar types of tomato have shown a standard deviation of 8.5 lbs per plant. You would like to design a study that will determine the average yield within a 90% error margin of ± 2 lbs. How many plants should you sample?

8.33.
College Board claims[r] that in 2010, public four-year colleges charged, on average, $7,605 per year in tuition and fees for in-state students. A sample of 20 public four-year colleges collected in 2011 indicated a sample mean of $8,039 and the sample standard deviation was $1,950. Is there sufficient evidence to conclude that the average in-state tuition has increased?

8.34.
The weights of grapefruit follow a normal distribution. A random sample of 12 new hybrid grapefruit had a mean weight of 1.7 pounds with standard deviation 0.24 pounds. Find a 95% confidence interval for the mean weight of the population of the new hybrid grapefruit.

8.35.
The Mountain View Credit Union claims that the average amount of money owed on their car loans is \$ 7,500. Suppose a random sample of 45 loans shows the average amount owed equals \$8,125, with standard deviation \$4,930. Does this indicate that the average amount owed on their car loans is not \$7,500? Use a level of significance $\alpha = 0.01$. Would your conclusion have changed if you used $\alpha = 0.05$?

8.36.
An overnight package delivery service has a promotional discount rate in effect this week only. For several years the mean weight of a package delivered by this company has been 10.7 ounces. However, a random sample of 12 packages mailed this week gave the following weights in ounces:

```
12.1  15.3   9.5  10.5  14.2  8.8  10.6  11.4  13.7  15.0   9.5  11.1
```

Use a 1% level of significance to test the claim that the packages are averaging more than 10.7 ounces during the discount week.

8.37.
Some people claim that during US elections, the taller of the two major party candidates tends to prevail. Here are some data on the last 15 elections (heights are in cm).

Year	2008	2004	2000	1996	1992	1988	1984	1980
Winning candidate	185	182	182	188	188	188	185	185
Losing candidate	175	193	185	187	188	173	180	177

Year	1976	1972	1968	1964	1960	1956	1952
Winning candidate	177	182	182	193	183	179	179
Losing candidate	183	185	180	180	182	178	178

Test the hypothesis that the winning candidates tend to be taller, on average.

8.38.
An item in USA Today reported that 63% of Americans owned a mobile browsing device. A survey of 143 employees at a large school showed that 85 owned a mobile browsing device. At $\alpha = 0.02$, test the claim that the percentage is the same as stated in USA Today.

8.39.
A poll by CNN revealed that 47% of Americans approve of the job performance of the President. The poll was based on a random sample of 537 adults.

a) Find the 95% margin of error for this poll.

b) Based on your result in part (a), test the hypothesis $H_0 : p = 0.5$ where p is the proportion of all American adults that approve of the job performance of the President. *Do not compute the test statistic and p-value.*

c) Would you have also reached the same conclusion for $H_0 : p = 0.45$?

8.40.
Find a poll cited in a newspaper, web site or other news source, with a mention of the sample size and the margin of error. (For example, `rasmussenreports.com` frequently discuss their polling methods.) Confirm the margin of error presented by the pollsters, using your own calculations.

Notes

[i]see e.g. http://forgetomori.com/2009/skepticism/seeing-patterns/

[j]see http://www.census.gov/hhes/www/cpstables/032010/perinc/new01_001.htm

[k]abridged from http://www.ssa.gov/oact/STATS/table4c6.html

[l]see http://en.wikipedia.org/wiki/Heights_of_Presidents_of_the_United_States_and_presidential_candidates

[m]see http://academic.udayton.edu/kissock/http/Weather/

[n]see http://www.readingonline.org/articles/bergman/wait.html

[o]"Student" [William Sealy Gosset] (March 1908). "The probable error of a mean". Biometrika 6 (1): 1-25.

[p]For example, see http://www.stat.columbia.edu/~gelman/book/data/light.asc

[q]example from "Statistical Sleuth"

[r]http://www.collegeboard.com/student/pay/add-it-up/4494.html

Chapter 9

Linear Regression

In science and engineering, there is often a need to investigate the relationship between two continuous random variables.

Suppose that, for every case observed, we record two variables, X and Y. The *linear* relationship between X and Y means that $\mathbb{E}(Y) = b_0 + b_1 X$.

X variable is usually called *predictor* or *independent variable* and Y variable is the *response* or *dependent variable*; the parameters are *slope* b_1 and *intercept* b_0.

Example 9.1.

Imagine that we are opening an ice cream stand and would like to be able to predict how many customers we will have. We might use the temperature as a predictor. We decided to collect data over a 30-week period from March to July.[s]

Week	1	2	3	4	5	6	7	8	9	10
Mean temp	41	56	63	68	69	65	61	47	32	24
Consumption	0.386	0.374	0.393	0.425	0.406	0.344	0.327	0.288	0.269	0.256

Week	11	12	13	14	15	16	17	18	19	20
Mean temp	28	26	32	40	55	63	72	72	67	60
Consumption	0.286	0.298	0.329	0.318	0.381	0.381	0.47	0.443	0.386	0.342

Week	21	22	23	24	25	26	27	28	29	30
Mean temp	44	40	32	27	28	33	41	52	64	71
Consumption	0.319	0.307	0.284	0.326	0.309	0.359	0.376	0.416	0.437	0.548

The following *scatterplot* is made to graphically investigate the relationship.
There indeed appears to be a straight-line trend. We will discuss fitting the equation a little later.

9.1 Correlation coefficient

We already know the correlation coefficient between two random variables,

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

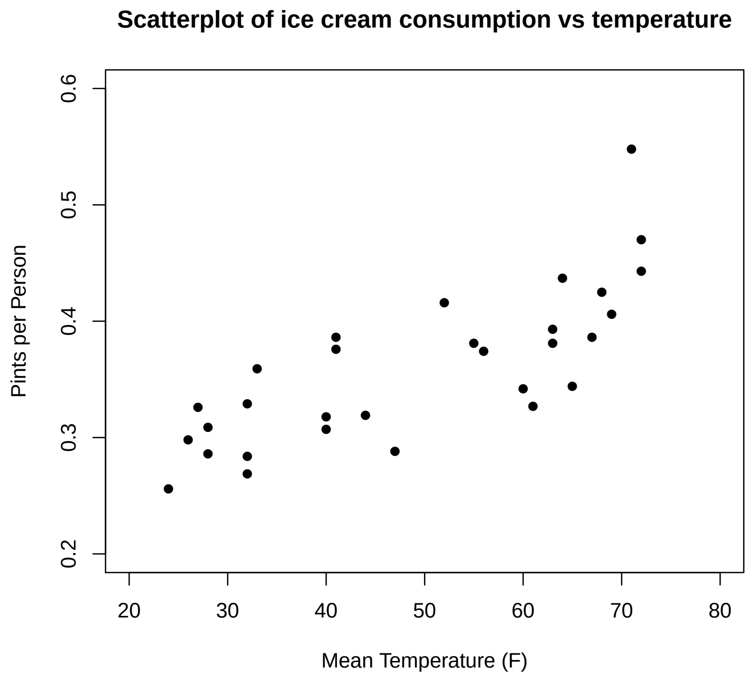


Figure 9.1: Scatterplot of ice cream data

Now, let's consider its sample analog, **sample correlation coefficient**

$$r = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \ \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \equiv \frac{SS_{XY}}{\sqrt{SS_X SS_Y}}$$

You can recognize the summation on top as a discrete version of $Cov(X, Y)$ and the sums on the bottom as part of the computation for the sample variances of X, Y. These are

$$SS_{XY} = \sum XY - \frac{\sum X \sum Y}{n},$$

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{n}, \quad SS_Y = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

All sums are taken from 1 to n.
For example, the sample variance of X is $S_X^2 = SS_X/(n-1)$.

Let's review the properties of the correlation coefficient ρ and its sample estimate, r:

- the sign of r points to positive (when X increases, Y increases too) or negative (when one increases, the other decreases) relationship
- $-1 \leq r \leq 1$, with $+1$ being a perfect positive and -1 a perfect negative relationship
- $r \approx 0$ means no linear relationship between X and Y (caution: there can still be a non-linear relationship!)
- r is dimensionless, and it does not change when X or Y are linearly transformed.

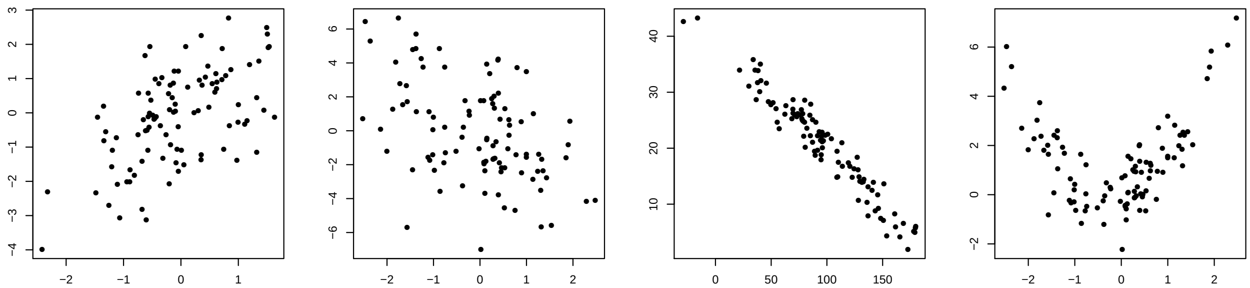


Figure 9.2: Correlations (left to right): `0.57, -0.45, -0.97, 0.13`

9.2 Least squares regression line

The complete regression equation is

$$Y_i = b_0 + b_1 X_i + \varepsilon_i, \qquad i = 1, ..., n$$

where the errors ε_i are assumed to be independent, $\mathcal{N}(0, \sigma^2)$.

To find the "best fit" line, we choose $\hat{b}_0$ and $\hat{b}_1$ that minimize the sum of squared residuals

$$SSE = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

(SSE is for the Sum of Squared Errors, however the quantities $Y_i - \hat{b}_0 - \hat{b}_1 X_i$ are usually referred to as *residuals*.)

To find the minimum, we would calculate partial derivatives of SSE with respect to b_0, b_1. Solving the resulting system of equations, we get the following

Theorem 9.1. Least squares estimates

The estimates for the regression equation $Y_i = b_0 + b_1 X_i + \varepsilon_i, \quad i = 1, ..., n$ are:

$$\text{Slope } \hat{b}_1 = \frac{SS_{XY}}{SS_X} = r\frac{S_Y}{S_X} \qquad \text{and Intercept } \hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}$$

Example 9.2.
To illustrate the computations, let's consider another data set. Here, X = amount of tannin in the larva food, and Y = growth of insect larvae.[t]

X	0	1	2	3	4	5	6	7	8
Y	12	10	8	11	6	7	2	3	3

Estimate the regression equation and correlation coefficient.

Solution.
$\sum X = 36$, $\sum Y = 62$, $\sum X^2 = 204$, $\sum Y^2 = 536$, $\sum XY = 175$.

Therefore,

$\overline{X} = 36/9 = 4, \qquad \overline{Y} = 62/9 = 6.89, \qquad SS_X = 204 - 36^2/9 = 60,$

$SS_Y = 536 - 62^2/9 = 108.9, \quad SS_{XY} = 175 - 36(62)/9 = -73$

and finally,
$\hat{b}_1 = -73/60 = -1.22, \quad \hat{b}_0 = 6.89 - (-1.22)4 = 11.76, \quad r = -0.903$

Thus, we get the equation

$$\hat{Y} = 11.76 - 1.22X$$

that is interpretable as a prediction for *any* given value X. In practice, the accuracy of prediction depends on X, see the next Section. □

Example 9.3.

For the data in Example 9.1,

a) Calculate and plot the least squares regression line

b) Predict the consumption when $X = 50°F$.

Solution.

(a) For the hand calculation, $\overline{X} = 49.1$, $\overline{Y} = 0.3594$, $SS_X = 7820.7$ $SS_Y = 0.1255$ and $SS_{XY} = 24.30$. We obtain the following estimates (can also be done by a computer)

$$\hat{b}_0 = 0.2069, \quad \hat{b}_1 = 0.003107 \quad \text{and} \quad r = 0.776$$

These can be used to plot the regression line (Fig. 9.3) and make predictions. Can you interpret the slope and the intercept for this problem in plain English?

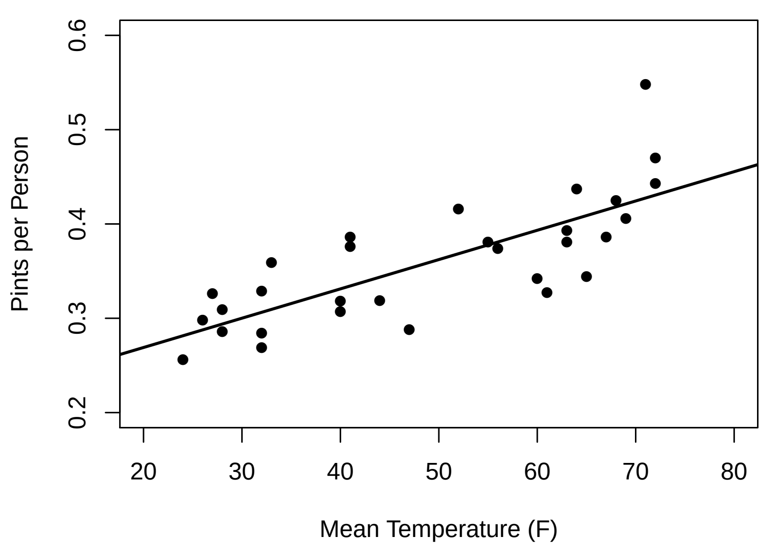


Figure 9.3: Least squares regression line for the ice cream example

(b) $\hat{Y} = \hat{b}_0 + \hat{b}_1 X = 0.2069 + 0.003107(50) = 0.362$ pints per person. □

9.3 Inference for regression

The error variance σ^2 determines the amount of scatter of the Y-values about the line. That is, it reflects the uncertainty of prediction of Y using X.
Its sample estimate is

$$S^2 = \frac{\texttt{SSE}}{n-2} = \frac{\sum_{i=1}^{n}[Y_i - (\hat{b}_0 + \hat{b}_1 X_i)]^2}{n-2} = \frac{(1-r^2)SS_Y}{n-2},$$

where SSE is the Sum of Squared Errors (Residuals). It is divided by $n-2$ because two degrees of freedom have been used up when estimating $\hat{b}_0$, $\hat{b}_1$. The estimate of S can be obtained by hand or using the computer output.

The values $\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$ are called *predicted* or *fitted* values of Y.
The differences

$$\text{Actual} - \text{ Predicted} \equiv Y_i - \hat{Y}_i = e_i, \qquad i = 1, ..., n$$

are called *residuals*.

The least squares esimates for slope and intercept can be viewed as *sample estimates* for the "true" (unknown) slope and intercept. We can apply the same methods we have done for, say, estimating the unknown mean μ. To make confidence intervals and perform hypothesis testing for the slope and intercept, we will need *standard errors* (that is, the estimates of standard deviations) of their estimates.
$100\%(1-\alpha)$ CI's for regression parameters are then found as

$$\texttt{Estimate} \pm t_{\alpha/2}(\texttt{Std.Error}), \qquad t \text{ has } df = n-2$$

The standard errors for slope and intercept can be obtained using the formulas

$$\texttt{SE}_{b_1} = \frac{S}{S_X\sqrt{n-1}} \qquad\qquad \texttt{SE}_{b_0} = S\sqrt{\frac{1}{n} + \frac{(\overline{X})^2}{(n-1)S_X^2}}$$

or using the computer output. Notice that the errors decrease at the familiar $\sqrt{n}$ rate, as sample size n grows.

Example 9.4.
Continuing the analysis of data from Example 9.1, let's examine a portion of computer output (done by R statistical package).

```
              Estimate  Std.Error  t-value   Pr(>|t|)
(Intercept)  0.2069        0.0247    8.375   4.13e-09
X           0.003107    0.000478    6.502   4.79e-07
```

We can calculate confidence intervals and hypothesis tests for the parameters b_0 and b_1.

The 95% C.I. for the slope b_1 is

$$0.003107 \pm 2.048(0.000478) = [0.002128, 0.004086]$$

To test the hypothesis $H_0 : b_1 = 0$ we could use the test statistic

$$t = \frac{\texttt{Estimate}}{\texttt{Std.Error}}$$

For the above data, we have $t = 0.003107/0.000478 = 6.502$, as reported in the table. The p-values for this test can be found using a t-table; they are also reported by the computer. Above, the reported p-value of `4.79e-07` is very small, meaning that the hypothesis $H_0 : b_1 = 0$ is strongly rejected.

Another part of the output will be useful later. This is a so-called *ANOVA (ANalysis Of VAriance) table*[1]:

```
               Df    Sum Sq   Mean Sq   F value     Pr(>F)
temperature     1  0.075514  0.075514     42.28  4.789e-07
Residuals      28  0.050009  0.001786
```

Here, we are interested in the Mean Square of Residuals $S^2 = 0.001786$. Also note that the p-value (here given as `Pr(>F)`) coincides with the T-test p-value for slope. □

9.3.1 Correlation test for linear relationship

To test whether the "true" slope equals 0 (and therefore the linear relationship does not exist), we can use the test statistic $t = \dfrac{b_1}{\mathtt{SE}_{b_1}}$.

In terms of correlation r, the above test can be calculated more easily using the test statistic

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \qquad df = n-2$$

Strictly speaking, this is for testing correlation

$$H_0 : \rho = 0 \text{ versus } H_A : \rho \neq 0$$

but $\rho = 0$ and $b_1 = 0$ are equivalent statements.

Example 9.5.
For a relationship between Population size and Divorce rate in $n = 20$ American cities the correlation of 0.28 was found. Is there a significant linear relationship between Population size and Divorce rate?

Solution.

$$t = 0.28\sqrt{\frac{20-2}{1-0.28^2}} = 1.23 \qquad \text{with } df = 18$$

From T-table (comparing with table value $t = 1.33$), $\mathtt{p-value} > 2(0.1) = 0.2$. Since p-value is larger than our default level $\alpha = 0.05$, do not reject H_0. Thus, we can claim **no significant evidence** of the linear relationship between Population size and Divorce rate. □

9.3.2 Confidence and prediction intervals

In addition to the C.I.'s for b_0 and b_1, we might be interested in the uncertainty of estimating Y-values given the particular value of X.

[1] Can you guess how to find SS_Y from this table?

$100\%(1-\alpha)$ **confidence interval for mean response** $\mathbb{E}(\hat{Y})$ **given** $X = x^*$

$$(\hat{b}_0 + \hat{b}_1 x^*) \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x^* - \overline{X})^2}{(n-1)S_X^2}}$$

$100\%(1-\alpha)$ **prediction interval for a future observation** Y **given** $X = x^*$

$$(\hat{b}_0 + \hat{b}_1 x^*) \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{X})^2}{(n-1)S_X^2}}$$

What is the main difference between *confidence* and *prediction* intervals? Confidence interval is only concerned with the **mean** response $\mathbb{E}(Y)$. That is, it's trying to catch the regression line. Prediction interval is concerned with any future observation. Thus, it is trying to catch all the points in the scatterplot. As a consequence, prediction interval is typically much wider.

Note also that $(n-1)S_X^2 = SS_X$, and both intervals are narrowest when x^* is closest to $\overline{X}$, the center of all data. The least squares fit becomes less reliable as you move to values of X away from the center, especially the areas where there is no X-data.

Example 9.6.
Continuing the analysis of data from Example 9.1, calculate both 95% confidence and prediction intervals for the ice cream consumption when temperature is $70°F$

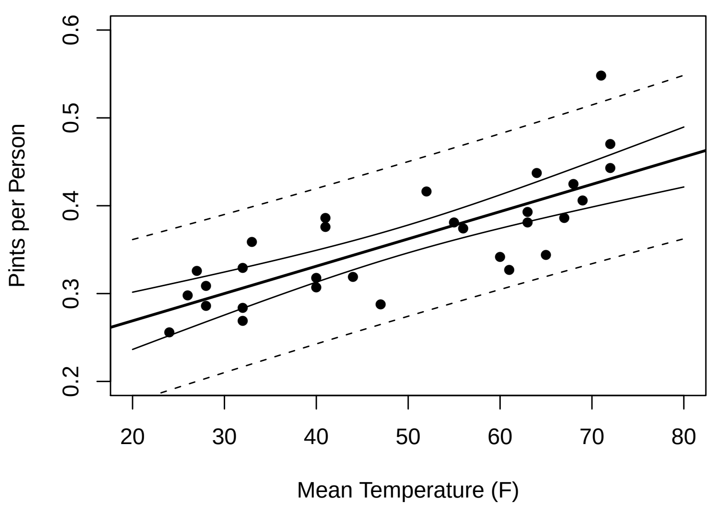


Figure 9.4: Confidence (solid lines) and prediction bands (broken lines) for the ice cream example

Solution. $\hat{Y} = \hat{b}_0 + \hat{b}_1 x^* = 0.2069 + (0.003107)70 = 0.4244$, and using the computer output in Example 9.4, we will get
$S = \sqrt{\texttt{Mean Sq Residuals}} = \sqrt{0.001786} = 0.0423$ and $\dfrac{1}{n} + \dfrac{(x^* - \overline{X})^2}{(n-1)S_X^2} = 0.0892$. Then, with $t_{\alpha/2} = 2.048$ ($\texttt{df} = 28$),

$$\text{CI} \quad 0.4244 \pm 0.0259, \qquad \text{PI} \quad 0.4244 \pm 0.0904$$

For comparison, both intervals are plotted in Fig. 9.4 for various values of x^*. Note that the 95% prediction band (broken lines) contains all but one observation. □

9.3.3 Checking the assumptions

To check the assumption of linear relationship and the constant variance (σ^2) of the residuals, we might make a plot of Residuals $e_i = Y_i - \hat{Y}_i$ versus Predicted (Fitted) values. If there is any trend or pattern in the residuals, then the assumptions for linear regression

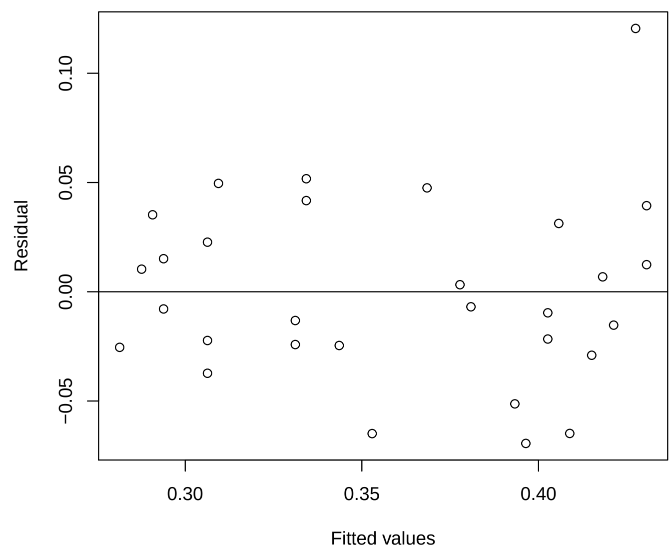


Figure 9.5: Residuals for the ice cream example

are not met. It might tell us, for example, if the size of residuals remains the same when the predicted value changes. Also, it can help spot non-linear behavior, outliers etc.

Such a plot for the ice cream example is given in Fig. 9.5. We do not see any particular trend except possibly one unusually high value (an outlier) in the top right corner.

Exercises

9.1.

In the file `http://www.nmt.edu/~olegm/382book/cars2010.csv`, there are some data on several 2010 compact car models. The variables are: engine displacement (liters), city MPG, highway MPG, and manufacturer's suggested price.

a) Is the car price related to its highway MPG?

b) Is there a relationship between city and highway MPG?

Use scatterplots, calculate and interpret the correlation coefficient, test to determine if there is a linear relationship.
For part (b), also compute and interpret the regression equation. Plot the regression line on the scatterplot. Plot the residuals versus predicted values. Does the model fit well?

9.2.
The following is an illustration of famous Moore's Law for computer chips. X = Year (minus 1900, for ease of computation), Y = number of transistors (in 1000)

X	71	79	83	85	90	93	95
Y	2.3	31	110	280	1200	3100	5500

a) Make a scatterplot of the data. Is the growth linear?

b) Let's try and fit the exponential growth model using a transformation:

$$\text{If} \qquad Y_i = a_0 e^{a_1 X_i} \qquad \text{then} \qquad \ln Y_i = \ln a_0 + a_1 X_i$$

That is, doing the linear regression analysis of $\ln Y$ on X will help recover the exponential growth. Make the regression analysis of $\ln Y$ on X. Does this model do a good job fitting the data?

c) Predict the number of transistors in the year 2005. Did this prediction come true?

9.3.
We are trying to model the relationship between X = Year (1981 = Year 1), and Y = Corn crop in US (in millions of tons). The following data were obtained for years 1981 to 1996:

	mean	st.dev.
Year	8.5	4.76
Crop	112.9	15.06

Additionally, the correlation between X and Y was 0.429

a) Compute the regression equation of Y as a linear function of X.

b) Predict the crop for 1988 ($X = 8$). The actual crop that year was $Y = 84.6$. What was the residual?

9.4.
A head of a large Hollywood company has seen the following values of its market share in the last six years:[u]

```
11.4, 10.6, 11.3, 7.4, 7.1, 6.7
```

Is there statistical evidence of a downward trend in the company's market share?

9.5.
For the Old Faithful geyser, the durations of eruption (X) were recorded, with the interval to the next eruption (Y), both in minutes.

X	3.6	1.8	3.3	2.3	4.5	2.9	4.7	3.6	1.9
Y	79	54	74	62	85	55	88	85	51

Perform the regression analysis of Y on X. Interpret the slope and give a 95% confidence interval for the slope.

9.6.

Are stock prices predictable? The following is a data set of the daily increases (in percent) of the S&P500 companies' stock prices for 21 consecutive business days. Is the next day increase significantly correlated to the last day's increase? [**Hint:** X and Y are both in here!]

```
 0.18  0.21 -0.50  0.13 -1.12  0.02  0.15 -0.10 -0.08 -0.06  0.16
-0.34 -0.60 -0.25 -0.08 -0.61  0.51 -0.09  0.00 -0.60 -0.38
```

9.7.

Does the price of the first-class postal stamp follow linear regression, or some other pattern?[v]

Year (since 1900)	32	58	63	68	71	74	75	78	81	85	88	91	95	99
Price (cents)	3	4	5	6	8	10	13	15	20	22	25	29	32	33
Year (since 1900)	101	102	106	108	111	112								
Price (cents)	34	37	39	42	44	45								

Predict the price in 2020.

9.8.

Bioilogists measured the correlation between arsenate concentration and bacterial cell growth rate, and obtained $r = -0.35$. Will this correlation be significant at $\alpha = 0.05$ level if $n = 25$? What if they obtained the same correlation for $n = 100$?

9.9. $\star$

Using other relationships found in this Chapter, prove that

$$\frac{b_1}{\mathtt{SE}_{b_1}} = r\sqrt{\frac{n-2}{1-r^2}}$$

These are the equivalent expressions for t-test statistic, first expression for testing $b_1 = 0$ and the second one is for testing $\rho = 0$.

Chapter 10

Categorical Data Analysis

In Section 8.6, we learned to compare two population proportions. We can extend this approach to more than two populations (groups) by the means of a chi-square test.

Consider the experiment of randomly selecting n items, each of which belongs to one of k categories (for example, we collect a sample of 100 people and look at their blood types, and there are $k = 4$ types). We will count the number of items in our sample of the type i and denote that X_i. We will refer to X_i as *observed count* for category i. Note that $X_1 + X_2 + ... + X_k = n$.

We will be concerned with estimating or testing the probabilities (or proportions) of ith category, p_i, $i = 1, ..., k$. Also, keep in mind the restriction $\sum_i p_i = 1$.

There are two types of tests considered in this Chapter:

- A test for **goodness-of-fit**, that is, how well do the observed counts X_i fit a given distribution.
- A test for **independence**, for which there are two classification categories (variables), and we are testing the independence of these variables.

10.1 Chi-square goodness-of-fit test

This is a test for the fit of the sample proportions to given numbers. Suppose that we have observations that can be classified into each of k groups (categorical data). We would like to test

$$H_0 : p_1 = p_1^0,\ p_2 = p_2^0,\ \dots\ , p_k = p_k^0$$

$$H_A : \text{ some of the } p_i\text{'s are unequal to } p_i^0\text{'s}$$

where p_i is the probability that a subject will belong to group i and $p_i^0, i = 1, ..., k$ are given numbers. (Note that $\sum p_i = \sum p_i^0 = 1$, so that p_k can actually be obtained from the rest of p_i's.)

Our data (*Observed counts*) are the counts of each category in the sample, $X_1, X_2, ..., X_k$ such that $\sum_{i=1}^k X_i = n$. The total sample size is n. For $k = 2$ we would get $X_1 =$ number of successes, and $X_2 = n - X_1 =$ number of failures, that is, Binomial distribution. For $k > 2$ we deal with *Multinomial distribution.*

For testing H_0, we compare the observed counts X_i to the ones we would expect under null hypothesis, that is,

$$\text{Expected counts} \qquad E_1 = np_1^0, \ \dots , E_k = np_k^0$$

To adjust for the size of each group, we would take the squared difference divided by E_i, that is $(E_i - X_i)^2/E_i$. Adding up, we obtain the

$$\text{Chi-square statistic} \qquad \chi^2 = \sum_{i=1}^{k} \frac{(E_i - X_i)^2}{E_i} \tag{10.1}$$

with $k-1$ degrees of freedom

We would reject H_0 when χ^2 statistic is large (that is, the Observed counts are far from Expected counts). Thus, our test is always *one-sided.* To find the p-value, use χ^2 upper-tail probability table very much like the t-table. See Table C.

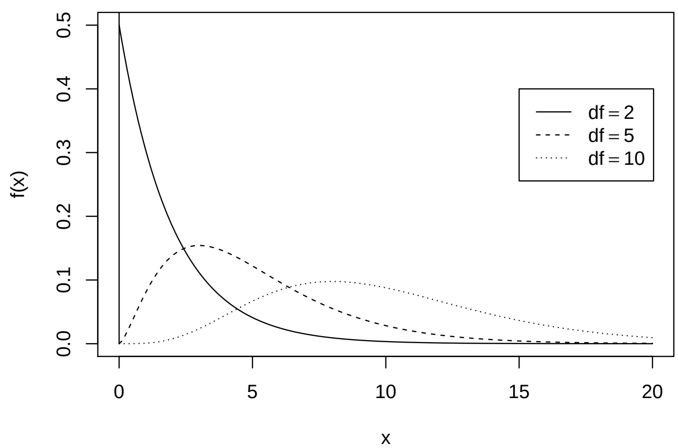


Figure 10.1: Chi-square densities

Assumption for chi-square test: all Expected counts should be ≥ 5 (this is necessary so that the normal approximation for counts X_i holds.) Some details: see below[1]

[1] Chi-square distribution with degrees of freedom $= k$ is related to Normal distribution as follows:

$$\chi^2 = Z_1^2 + Z_2^2 + \dots\dots + Z_k^2,$$

where $Z_1, ..., Z_k$ are independent, standard Normal r.v.'s.

Also, it can be shown that chi-square (`df` $= k$) distribution is simply $Gamma(\alpha = k/2, \beta = 2)$– sorry, this α and the significance level for testing are not the same!

For example, Chi-square(`df` $= 2$) is the same as Exponential ($\beta = 2$). (Why?)
Note that this distribution has positive values and is not symmetric!

Table C: Critical points of the chi-square distribution

	Upper tail probability						
Degrees of freedom	0.100	0.050	0.025	0.010	0.005	0.001	0.0005
1	2.706	3.841	5.024	6.635	7.879	10.828	12.116
2	4.605	5.991	7.378	9.210	10.597	13.816	15.202
3	6.251	7.815	9.348	11.345	12.838	16.266	17.730
4	7.779	9.488	11.143	13.277	14.860	18.467	19.997
5	9.236	11.070	12.833	15.086	16.750	20.515	22.105
6	10.645	12.592	14.449	16.812	18.548	22.458	24.103
7	12.017	14.067	16.013	18.475	20.278	24.322	26.018
8	13.362	15.507	17.535	20.090	21.955	26.124	27.868
9	14.684	16.919	19.023	21.666	23.589	27.877	29.666
10	15.987	18.307	20.483	23.209	25.188	29.588	31.420
11	17.275	19.675	21.920	24.725	26.757	31.264	33.137
12	18.549	21.026	23.337	26.217	28.300	32.909	34.821
13	19.812	22.362	24.736	27.688	29.819	34.528	36.478
14	21.064	23.685	26.119	29.141	31.319	36.123	38.109
15	22.307	24.996	27.488	30.578	32.801	37.697	39.719
16	23.542	26.296	28.845	32.000	34.267	39.252	41.308
17	24.769	27.587	30.191	33.409	35.718	40.790	42.879
18	25.989	28.869	31.526	34.805	37.156	42.312	44.434
19	27.204	30.144	32.852	36.191	38.582	43.820	45.973
20	28.412	31.410	34.170	37.566	39.997	45.315	47.498
21	29.615	32.671	35.479	38.932	41.401	46.797	49.011
22	30.813	33.924	36.781	40.289	42.796	48.268	50.511
23	32.007	35.172	38.076	41.638	44.181	49.728	52.000
24	33.196	36.415	39.364	42.980	45.559	51.179	53.479
25	34.382	37.652	40.646	44.314	46.928	52.620	54.947
30	40.256	43.773	46.979	50.892	53.672	59.703	62.162
40	51.805	55.758	59.342	63.691	66.766	73.402	76.095
60	74.397	79.082	83.298	88.379	91.952	99.607	102.695
80	96.578	101.879	106.629	112.329	116.321	124.839	128.261
100	118.498	124.342	129.561	135.807	140.169	149.449	153.167

Example 10.1.

When studying earthquakes, we recorded the following numbers of earthquakes (1 and above on Richter scale) for 7 consecutive days in January 2008.

Day	1	2	3	4	5	6	7	Total
Count	85	98	79	118	112	135	137	764
Expected	109.1	109.1	109.1	109.1	109.1	109.1	109.1	764

Here, $n = 764$. Is there evidence that the rate of earthquake activity changes during this week?

Solution. If the null hypothesis $H_0 : p_1 = p_2 = ... = p_7$ were true, then each $p_i = 1/7$, $i = 1, ..., 7$. Thus, we can find the expected counts $E_i = 764/7 = 109.1$.

Results: $\chi^2 = 28.8$, $df = 6$, p-value < 0.0005 from Table C.(The highest number there, 24.103, corresponds to upper tail area 0.0005.) Since the p-value is small, we reject H_0 and claim that the earthquake frequency **does** change during the week.[2] □

Example 10.2.

In this example, we will test whether a paricular distribution matches our experimental results. These are the data from the probability board (quincunx), we test if the distribution is really Binomial (as is often claimed). The slots are labeled 0-19. Some slots were merged together (why?)

Slots	0-6	7	8	9	10	11	12	13-19	Total
Observed	16	2	11	18	14	14	7	18	100
Expected	8.4	9.6	14.4	17.6	17.6	14.4	9.6	8.4	100

Solution. The expected counts are computed using Binomial($n = 19, p = 0.5$) distribution, and then multiplying by the $Total = 100$. For example,

$$E_9 = \binom{19}{9} 0.5^9 (1 - 0.5)^{19-9} \times 100 = 17.6$$

Next, $\chi^2 = 26.45$, $df = 7$, and p-value < 0.0005.
Conclusion: Reject H_0, the distribution is not exactly Binomial. □

10.2 Chi-square test for independence

This test is applied to the category probabilities for two variables. Each case is classified according to variable 1 (for example, Gender) and variable 2 (for example, College Major). The data are usually given in a *cross-classification* table (a 2-way table). Let X_{ij} be the observed table counts for row i and column j.

We are interested in testing whether Variable 1 (in r rows) is independent of Variable 2 (in c columns).[3]

[2] We did not specify α for this example. As mentioned earlier, $\alpha = 0.05$ is a good "default" choice. Even if we pick a conservative $\alpha = 0.01$, we would still reject H_0 here.

[3] These are not random variables in the sense of Chapter 3, because they are *categorical*, not numerical.

In this situation, we set up a chi-square statistic following equation (10.1). However, now the table is bigger. The Expected counts will be found using independence assumption, as

$$\text{Expected counts } E_{ij} = \frac{R_i C_j}{n}, \quad i = 1, ..., r \quad j = 1, ..., c$$

where R_i and C_j are the row and column totals.

Theorem 10.1. Chi-square test for independence

To test

$$H_0: \text{ Variable 1 is independent of Variable 2 } \textit{vs}$$

$$H_A: \text{ Variable 1 is } \textbf{not} \text{ independent of Variable 2}$$

we can use the χ^2 random variable with $df = (r-1)(c-1)$, where

$$\text{test statistic} \quad \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(E_{ij} - X_{ij})^2}{E_{ij}} \tag{10.2}$$

Example 10.3.
Suppose that we ordered 50 components from each of the vendors A, B and C, and the results are as follows

	Succeeded	Failed	Total
Vendor A	49	1	50
Vendor B	45	5	50
Vendor C	41	9	50

We would like to investigate whether all the vendors are equally reliable. That is,

$$H_0: \text{ Failure rate is independent of Vendor}$$

$$H_A: \text{ Not all Vendors have the same failure rate}$$

Solution. We'll put all the expected counts into the table

```
Expected counts:
           Succeeded   Failed   Total
Vendor A      45          5       50
Vendor B      45          5       50
Vendor C      45          5       50
 -----------------------------------------------------
Total        135         15      150
```

The χ^2 statistic will have $df = (3-1)(2-1) = 2$.

Here, $\chi^2 = (45-49)^2/45 + (1-5)^2/5 + ... = 7.11$. Since χ^2 statistic is between table values 5.991 and 7.378, the p-value is between 0.025 and 0.05. At the standard $\alpha = 0.05$ we are rejecting H_0. Thus, there **is** evidence that vendors have different failure rates.[4] □

[4]For this particular example, since `df` = 2, there is a more exact p-value calculation based on Exponential distribution: $P(Y > 7.11) = \exp(-7.11/2) = 0.0286$. For `df` $\neq 2$, we can use `R` function `pchisq`, Excel function `chidist` or other software to compute the exact p-values.

Exercises

10.1.
In testing how well people can generate random patterns, the researchers asked everyone in a group of 20 people to write a list of 5 random digits. The results are tabulated below

Digits	0	1	2	3	4	5	6	7	8	9	Total
Observed	6	11	10	13	8	13	7	17	8	7	100

Are the digits completely random or do humans have preference for some particular digits over the others?

10.2.
Forensic statistics. To uncover rigged elections, a variety of statistical tests might be applied. For example, made-up precinct totals are sometimes likely to have an excess of 0 or 5 as their last digits. For a city election, the observers counted that 21 precinct totals had the last digit 0, 18 had the last digit 5, while 102 had some other last digit. Is there evidence that the elections were rigged?

10.3.
In an earlier example of Poisson distribution, we discussed the number of Nazi bombs hitting $0.5 \times 0.5km$ squares in London. The following were counts of squares that have $0, 1, 2, ...$ hits:

number of hits	0	1	2	3	4 and up
count	229	211	93	35	8

Test whether the data fit the Poisson distribution (for $p_1^0, ...p_k^0$ use the Poisson probabilities, with the parameter μ estimated as average number of hits per square, $\mu = 0.9288$).

10.4.
To test the attitudes to a tax reform, the state officials collected data of the opinions of likely voters, along with their income level

	Income Level:		
	Low	Medium	High
For	182	213	203
Against	154	138	110

Do the people with different incomes have significantly different opinions on tax reform? (That is, test whether the Opinion variable is independent of Income variable.)

10.5.
Using exponential distribution, confirm the calculation of chi-square ($\mathtt{df} = 2$) critical points from Table C for upper tail area $\alpha = 0.1$ and $\alpha = 0.005$. Find the point for $\chi^2(\mathtt{df} = 2)$ distribution with $\alpha = 0.2$

Notes

[s] Kotswara Rao Kadilyala (1970). "Testing for the independence of regression disturbances" Econometrica, 38, 97-117. Appears in: *A Handbook of Small Data Sets*, D. J. Hand, et al, editors (1994). Chapman and Hall, London.

[t]from *The R book* by Michael Crawley

[u]Mlodinow again. The director, Sherry Lansing, was subsequently fired only to see several films developed during her tenure, including *Men In Black*, hit it big.

[v]see `http://www.akdart.com/postrate.html`

Appendix

Table D

Binomial CDF: $F(x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$

n = 5

	p												
x	.01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.951	.774	.59	.328	.168	.078	.031	.01	.002	0	0	0	0
1	.999	.977	.919	.737	.528	.337	.187	.087	.031	.007	0	0	0
2	1	.999	.991	.942	.837	.683	.5	.317	.163	.058	.009	.001	0
3	1	1	1	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001
4	1	1	1	1	.998	.99	.969	.922	.832	.672	.41	.226	.049
5	1	1	1	1	1	1	1	1	1	1	1	1	1

n = 6

	p												
x	.01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.941	.735	.531	.262	.118	.047	.016	.004	.001	0	0	0	0
1	.999	.967	.886	.655	.42	.233	.109	.041	.011	.002	0	0	0
2	1	.998	.984	.901	.744	.544	.344	.179	.07	.017	.001	0	0
3	1	1	.999	.983	.93	.821	.656	.456	.256	.099	.016	.002	0
4	1	1	1	.998	.989	.959	.891	.767	.58	.345	.114	.033	.001
5	1	1	1	1	.999	.996	.984	.953	.882	.738	.469	.265	.059
6	1	1	1	1	1	1	1	1	1	1	1	1	1

n = 7

	p												
x	.01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.932	.698	.478	.21	.082	.028	.008	.002	0	0	0	0	0
1	.998	.956	.85	.577	.329	.159	.063	.019	.004	0	0	0	0
2	1	.996	.974	.852	.647	.42	.227	.096	.029	.005	0	0	0
3	1	1	.997	.967	.874	.71	.5	.29	.126	.033	.003	0	0
4	1	1	1	.995	.971	.904	.773	.58	.353	.148	.026	.004	0
5	1	1	1	1	.996	.981	.938	.841	.671	.423	.15	.044	.002
6	1	1	1	1	1	.998	.992	.972	.918	.79	.522	.302	.068
7	1	1	1	1	1	1	1	1	1	1	1	1	1

n = 8

	p												
x	.01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.923	.663	.43	.168	.058	.017	.004	.001	0	0	0	0	0
1	.997	.943	.813	.503	.255	.106	.035	.009	.001	0	0	0	0
2	1	.994	.962	.797	.552	.315	.145	.05	.011	.001	0	0	0
3	1	1	.995	.944	.806	.594	.363	.174	.058	.01	0	0	0
4	1	1	1	.99	.942	.826	.637	.406	.194	.056	.005	0	0
5	1	1	1	.999	.989	.95	.855	.685	.448	.203	.038	.006	0
6	1	1	1	1	.999	.991	.965	.894	.745	.497	.187	.057	.003
7	1	1	1	1	1	.999	.996	.983	.942	.832	.57	.337	.077
8	1	1	1	1	1	1	1	1	1	1	1	1	1

Table D (continued)

n = 9

x	p .01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.914	.63	.387	.134	.04	.01	.002	0	0	0	0	0	0
1	.997	.929	.775	.436	.196	.071	.02	.004	0	0	0	0	0
2	1	.992	.947	.738	.463	.232	.09	.025	.004	0	0	0	0
3	1	.999	.992	.914	.73	.483	.254	.099	.025	.003	0	0	0
4	1	1	.999	.98	.901	.733	.5	.267	.099	.02	.001	0	0
5	1	1	1	.997	.975	.901	.746	.517	.27	.086	.008	.001	0
6	1	1	1	1	.996	.975	.91	.768	.537	.262	.053	.008	0
7	1	1	1	1	1	.996	.98	.929	.804	.564	.225	.071	.003
8	1	1	1	1	1	1	.998	.99	.96	.866	.613	.37	.086
9	1	1	1	1	1	1	1	1	1	1	1	1	1

n = 10

x	p .01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.904	.599	.349	.107	.028	.006	.001	0	0	0	0	0	0
1	.996	.914	.736	.376	.149	.046	.011	.002	0	0	0	0	0
2	1	.988	.93	.678	.383	.167	.055	.012	.002	0	0	0	0
3	1	.999	.987	.879	.65	.382	.172	.055	.011	.001	0	0	0
4	1	1	.998	.967	.85	.633	.377	.166	.047	.006	0	0	0
5	1	1	1	.994	.953	.834	.623	.367	.15	.033	.002	0	0
6	1	1	1	.999	.989	.945	.828	.618	.35	.121	.013	.001	0
7	1	1	1	1	.998	.988	.945	.833	.617	.322	.07	.012	0
8	1	1	1	1	1	.998	.989	.954	.851	.624	.264	.086	.004
9	1	1	1	1	1	1	.999	.994	.972	.893	.651	.401	.096
10	1	1	1	1	1	1	1	1	1	1	1	1	1

n = 11

x	p .01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.895	.569	.314	.086	.02	.004	0	0	0	0	0	0	0
1	.995	.898	.697	.322	.113	.03	.006	.001	0	0	0	0	0
2	1	.985	.91	.617	.313	.119	.033	.006	.001	0	0	0	0
3	1	.998	.981	.839	.57	.296	.113	.029	.004	0	0	0	0
4	1	1	.997	.95	.79	.533	.274	.099	.022	.002	0	0	0
5	1	1	1	.988	.922	.753	.5	.247	.078	.012	0	0	0
6	1	1	1	.998	.978	.901	.726	.467	.21	.05	.003	0	0
7	1	1	1	1	.996	.971	.887	.704	.43	.161	.019	.002	0
8	1	1	1	1	.999	.994	.967	.881	.687	.383	.09	.015	0
9	1	1	1	1	1	.999	.994	.97	.887	.678	.303	.102	.005
10	1	1	1	1	1	1	1	.996	.98	.914	.686	.431	.105
11	1	1	1	1	1	1	1	1	1	1	1	1	1

Table D (continued)

n = 15

x	p .01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.86	.463	.206	.035	.005	0	0	0	0	0	0	0	0
1	.99	.829	.549	.167	.035	.005	0	0	0	0	0	0	0
2	1	.964	.816	.398	.127	.027	.004	0	0	0	0	0	0
3	1	.995	.944	.648	.297	.091	.018	.002	0	0	0	0	0
4	1	.999	.987	.836	.515	.217	.059	.009	.001	0	0	0	0
5	1	1	.998	.939	.722	.403	.151	.034	.004	0	0	0	0
6	1	1	1	.982	.869	.61	.304	.095	.015	.001	0	0	0
7	1	1	1	.996	.95	.787	.5	.213	.05	.004	0	0	0
8	1	1	1	.999	.985	.905	.696	.39	.131	.018	0	0	0
9	1	1	1	1	.996	.966	.849	.597	.278	.061	.002	0	0
10	1	1	1	1	.999	.991	.941	.783	.485	.164	.013	.001	0
11	1	1	1	1	1	.998	.982	.909	.703	.352	.056	.005	0
12	1	1	1	1	1	1	.996	.973	.873	.602	.184	.036	0
13	1	1	1	1	1	1	1	.995	.965	.833	.451	.171	.01
14	1	1	1	1	1	1	1	1	.995	.965	.794	.537	.14
15	1	1	1	1	1	1	1	1	1	1	1	1	1

n = 20

x	p .01	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95	.99
0	.818	.358	.122	.012	.001	0	0	0	0	0	0	0	0
1	.983	.736	.392	.069	.008	.001	0	0	0	0	0	0	0
2	.999	.925	.677	.206	.035	.004	0	0	0	0	0	0	0
3	1	.984	.867	.411	.107	.016	.001	0	0	0	0	0	0
4	1	.997	.957	.63	.238	.051	.006	0	0	0	0	0	0
5	1	1	.989	.804	.416	.126	.021	.002	0	0	0	0	0
6	1	1	.998	.913	.608	.25	.058	.006	0	0	0	0	0
7	1	1	1	.968	.772	.416	.132	.021	.001	0	0	0	0
8	1	1	1	.99	.887	.596	.252	.057	.005	0	0	0	0
9	1	1	1	.997	.952	.755	.412	.128	.017	.001	0	0	0
10	1	1	1	.999	.983	.872	.588	.245	.048	.003	0	0	0
11	1	1	1	1	.995	.943	.748	.404	.113	.01	0	0	0
12	1	1	1	1	.999	.979	.868	.584	.228	.032	0	0	0
13	1	1	1	1	1	.994	.942	.75	.392	.087	.002	0	0
14	1	1	1	1	1	.998	.979	.874	.584	.196	.011	0	0
15	1	1	1	1	1	1	.994	.949	.762	.37	.043	.003	0
16	1	1	1	1	1	1	.999	.984	.893	.589	.133	.016	0
17	1	1	1	1	1	1	1	.996	.965	.794	.323	.075	.001
18	1	1	1	1	1	1	1	.999	.992	.931	.608	.264	.017
19	1	1	1	1	1	1	1	1	.999	.988	.878	.642	.182
20	1	1	1	1	1	1	1	1	1	1	1	1	1

Table E

Poisson CDF: $F(x) = \sum_{k=0}^{x} \frac{e^{-\mu}\mu^k}{k!}$

x	μ .1	.2	.3	.4	.5	.6	.7	.8	.9	1	1.5	2
0	.905	.819	.741	.67	.607	.549	.497	.449	.407	.368	.223	.135
1	.995	.982	.963	.938	.91	.878	.844	.809	.772	.736	.558	.406
2	1	.999	.996	.992	.986	.977	.966	.953	.937	.92	.809	.677
3	1	1	1	.999	.998	.997	.994	.991	.987	.981	.934	.857
4	1	1	1	1	1	1	.999	.999	.998	.996	.981	.947
5	1	1	1	1	1	1	1	1	1	.999	.996	.983
6	1	1	1	1	1	1	1	1	1	1	.999	.995
7	1	1	1	1	1	1	1	1	1	1	1	.999
8	1	1	1	1	1	1	1	1	1	1	1	1

x	μ 2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
0	.082	.05	.03	.018	.011	.007	.004	.002	.002	.001	.001	0
1	.287	.199	.136	.092	.061	.04	.027	.017	.011	.007	.005	.003
2	.544	.423	.321	.238	.174	.125	.088	.062	.043	.03	.02	.014
3	.758	.647	.537	.433	.342	.265	.202	.151	.112	.082	.059	.042
4	.891	.815	.725	.629	.532	.44	.358	.285	.224	.173	.132	.1
5	.958	.916	.858	.785	.703	.616	.529	.446	.369	.301	.241	.191
6	.986	.966	.935	.889	.831	.762	.686	.606	.527	.45	.378	.313
7	.996	.988	.973	.949	.913	.867	.809	.744	.673	.599	.525	.453
8	.999	.996	.99	.979	.96	.932	.894	.847	.792	.729	.662	.593
9	1	.999	.997	.992	.983	.968	.946	.916	.877	.83	.776	.717
10	1	1	.999	.997	.993	.986	.975	.957	.933	.901	.862	.816
11	1	1	1	.999	.998	.995	.989	.98	.966	.947	.921	.888
12	1	1	1	1	.999	.998	.996	.991	.984	.973	.957	.936
13	1	1	1	1	1	.999	.998	.996	.993	.987	.978	.966
14	1	1	1	1	1	1	.999	.999	.997	.994	.99	.983
15	1	1	1	1	1	1	1	.999	.999	.998	.995	.992
16	1	1	1	1	1	1	1	1	1	.999	.998	.996
17	1	1	1	1	1	1	1	1	1	1	.999	.998
18	1	1	1	1	1	1	1	1	1	1	1	.999
19	1	1	1	1	1	1	1	1	1	1	1	1

Table E (continued)

x	μ 9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	0	0	0	0	0	0	0	0	0	0
1	.001	0	0	0	0	0	0	0	0	0	0	0
2	.006	.003	.001	.001	0	0	0	0	0	0	0	0
3	.021	.01	.005	.002	.001	0	0	0	0	0	0	0
4	.055	.029	.015	.008	.004	.002	.001	0	0	0	0	0
5	.116	.067	.038	.02	.011	.006	.003	.001	.001	0	0	0
6	.207	.13	.079	.046	.026	.014	.008	.004	.002	.001	.001	0
7	.324	.22	.143	.09	.054	.032	.018	.01	.005	.003	.002	.001
8	.456	.333	.232	.155	.1	.062	.037	.022	.013	.007	.004	.002
9	.587	.458	.341	.242	.166	.109	.07	.043	.026	.015	.009	.005
10	.706	.583	.46	.347	.252	.176	.118	.077	.049	.03	.018	.011
11	.803	.697	.579	.462	.353	.26	.185	.127	.085	.055	.035	.021
12	.876	.792	.689	.576	.463	.358	.268	.193	.135	.092	.061	.039
13	.926	.864	.781	.682	.573	.464	.363	.275	.201	.143	.098	.066
14	.959	.917	.854	.772	.675	.57	.466	.368	.281	.208	.15	.105
15	.978	.951	.907	.844	.764	.669	.568	.467	.371	.287	.215	.157
16	.989	.973	.944	.899	.835	.756	.664	.566	.468	.375	.292	.221
17	.995	.986	.968	.937	.89	.827	.749	.659	.564	.469	.378	.297
18	.998	.993	.982	.963	.93	.883	.819	.742	.655	.562	.469	.381
19	.999	.997	.991	.979	.957	.923	.875	.812	.736	.651	.561	.47
20	1	.998	.995	.988	.975	.952	.917	.868	.805	.731	.647	.559
21	1	.999	.998	.994	.986	.971	.947	.911	.861	.799	.725	.644
22	1	1	.999	.997	.992	.983	.967	.942	.905	.855	.793	.721
23	1	1	1	.999	.996	.991	.981	.963	.937	.899	.849	.787
24	1	1	1	.999	.998	.995	.989	.978	.959	.932	.893	.843
25	1	1	1	1	.999	.997	.994	.987	.975	.955	.927	.888
26	1	1	1	1	1	.999	.997	.993	.985	.972	.951	.922
27	1	1	1	1	1	.999	.998	.996	.991	.983	.969	.948
28	1	1	1	1	1	1	.999	.998	.995	.99	.98	.966
29	1	1	1	1	1	1	1	.999	.997	.994	.988	.978
30	1	1	1	1	1	1	1	.999	.999	.997	.993	.987
31	1	1	1	1	1	1	1	1	.999	.998	.996	.992
32	1	1	1	1	1	1	1	1	1	.999	.998	.995
33	1	1	1	1	1	1	1	1	1	1	.999	.997
34	1	1	1	1	1	1	1	1	1	1	.999	.999
35	1	1	1	1	1	1	1	1	1	1	1	.999
36	1	1	1	1	1	1	1	1	1	1	1	1