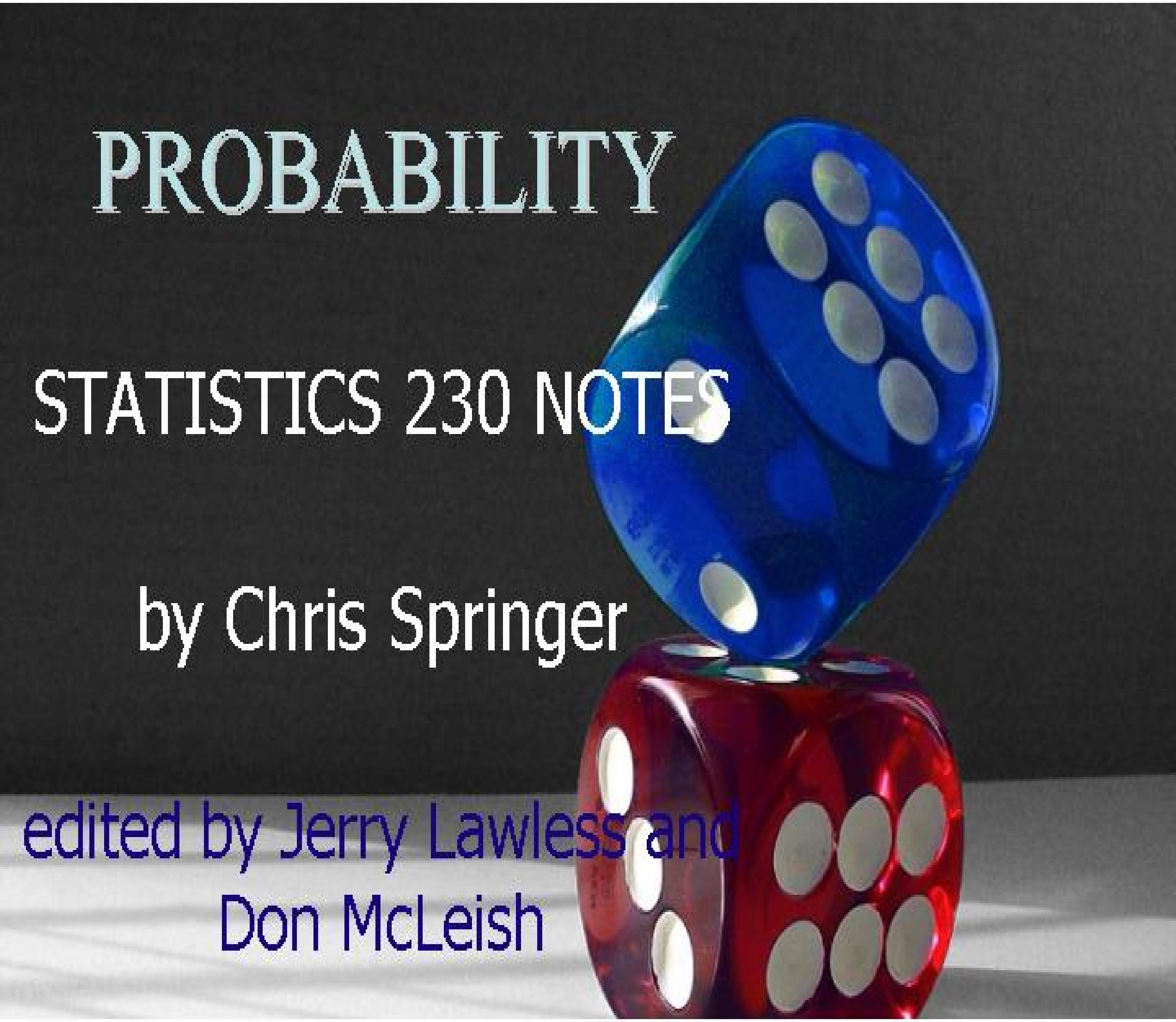
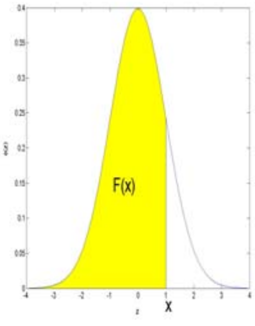
PROBABILITY

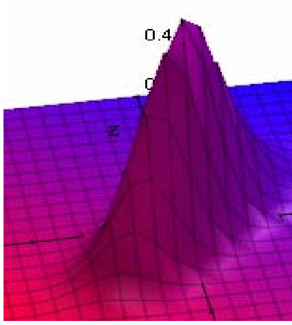
STATISTICS 230 NOTES

by Chris Springer

edited by Jerry Lawless and Don McLeish







STATISTICS 230 COURSE NOTES

Chris Springer, revised by Jerry Lawless and Don McLeish

JANUARY 2006

Contents

1. Introduction to Probability

In some areas, such as mathematics or logic, results of some process can be known with certainty (e.g., 2+3=5). Most real life situations, however, involve variability and uncertainty. For example, it is uncertain whether it will rain tomorrow; the price of a given stock a week from today is uncertain[1]; the number of claims that a car insurance policy holder will make over a one-year period is uncertain. Uncertainty or "randomness" (meaning variability of results) is usually due to some mixture of two factors: (1) variability in populations consisting of animate or inanimate objects (e.g., people vary in size, weight, blood type etc.), and (2) variability in processes or phenomena (e.g., the random selection of 6 numbers from 49 in a lottery draw can lead to a very large number of different outcomes; stock or currency prices fluctuate substantially over time).

Variability and uncertainty make it more difficult to plan or to make decisions. Although they cannot usually be eliminated, it is however possible to describe and to deal with variability and uncertainty, by using the theory of probability. This course develops both the theory and applications of probability.

It seems logical to begin by defining probability. People have attempted to do this by giving definitions that reflect the uncertainty whether some specified outcome or "event" will occur in a given setting. The setting is often termed an "experiment" or "process" for the sake of discussion. To take a simple "toy" example: it is uncertain whether the number 2 will turn up when a 6-sided die is rolled. It is similarly uncertain whether the Canadian dollar will be higher tomorrow, relative to the U.S. dollar, than it is today. Three approaches to defining probability are:

1. The **classical** definition: Let the **sample space** (denoted by S) be the set of all possible distinct outcomes to an experiment. The probability of some event is

$$\frac{\text{number of ways the event can occur}}{\text{number of outcomes in S}},$$

provided all points in S are equally likely. For example, when a die is rolled the probability of getting a 2 is $\frac{1}{6}$ because one of the six faces is a 2.

[1]"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality" Albert Einstein, 1921.

2. The **relative frequency** definition: The probability of an event is the proportion (or fraction) of times the event occurs in a very long (theoretically infinite) series of repetitions of an experiment or process. For example, this definition could be used to argue that the probability of getting a 2 from a rolled die is $\frac{1}{6}$.

3. The **subjective probability** definition: The probability of an event is a measure of how sure the person making the statement is that the event will happen. For example, after considering all available data, a weather forecaster might say that the probability of rain today is 30% or 0.3.

Unfortunately, all three of these definitions have serious limitations.

Classical Definition:

What does "equally likely" mean? This appears to use the concept of probability while trying to define it! We could remove the phrase "provided all outcomes are equally likely", but then the definition would clearly be unusable in many settings where the outcomes in S did not tend to occur equally often.

Relative Frequency Definition:

Since we can never repeat an experiment or process indefinitely, we can never know the probability of any event from the relative frequency definition. In many cases we can't even obtain a long series of repetitions due to time, cost, or other limitations. For example, the probability of rain today can't really be obtained by the relative frequency definition since today can't be repeated again.

Subjective Probability:

This definition gives no rational basis for people to agree on a right answer. There is some controversy about when, if ever, to use subjective probability except for personal decision-making. It will not be used in Stat 230.

These difficulties can be overcome by treating probability as a mathematical system defined by a set of axioms. In this case we do not worry about the numerical values of probabilities until we consider a specific application. This is consistent with the way that other branches of mathematics are defined and then used in specific applications (e.g., the way calculus and real-valued functions are used to model and describe the physics of gravity and motion).

The mathematical approach that we will develop and use in the remaining chapters assumes the following:

- probabilities are numbers between 0 and 1 that apply to outcomes, termed "events",

- each event may or may not occur in a given setting.

Chapter 2 begins by specifying the mathematical framework for probability in more detail.

Exercises

1. Try to think of examples of probabilities you have encountered which might have been obtained by each of the three “definitions".

2. Which definitions do you think could be used for obtaining the following probabilities?

   (a) You have a claim on your car insurance in the next year.

   (b) There is a meltdown at a nuclear power plant during the next 5 years.

   (c) A person’s birthday is in April.

3. Give examples of how probability applies to each of the following areas.

   (a) Lottery draws

   (b) Auditing of expense items in a financial statement

   (c) Disease transmission (e.g. measles, tuberculosis, STD’s)

   (d) Public opinion polls

2. Mathematical Probability Models

2.1 Sample Spaces and Probability

Consider some phenomenon or process which is repeatable, at least in theory, and suppose that certain events (outcomes) $A_1, A_2, A_3, \ldots$ are defined. We will often term the phenomenon or process an **“experiment"** and refer to a single repetition of the experiment as a **“trial"**. Then the probability of an event A, denoted $P(A)$, is a number between 0 and 1.

If probability is to be a useful mathematical concept, it should possess some other properties. For example, if our “experiment” consists of tossing a coin with two sides, Head and Tail, then we might wish to consider the events A_1 = “Head turns up” and A_2 = “Tail turns up”. It would clearly not be desirable to allow, say, $P(A_1) = 0.6$ and $P(A_2) = 0.6$, so that $P(A_1) + P(A_2) > 1$. (Think about why this is so.) To avoid this sort of thing we begin with the following definition.

Definition 1 *A **sample space** S is a set of distinct outcomes for an experiment or process, with the property that in a single trial, one and only one of these outcomes occurs. The outcomes that make up the sample space are called **sample points**.*

A sample space is part of the probability model in a given setting. It is not necessarily unique, as the following example shows.

Example: Roll a 6-sided die, and define the events

$$a_i = \text{number } i \text{ turns up} \quad (i = 1, 2, 3, 4, 5, 6)$$

Then we could take the sample space as $S = \{a_1, a_2, a_3, a_4, a_5, a_6\}$. However, we could also define events

$E =$ even number turns up

$O =$ odd number turns up

and take $S = \{E, O\}$. Both sample spaces satisfy the definition, and which one we use would depend on what we wanted to use the probability model for. In most cases we would use the first sample space.

Sample spaces may be either **discrete** or **non-discrete**; S is discrete if it consists of a finite or countably infinite set of simple events. The two sample spaces in the preceding example are discrete. A sample space $S = \{1, 2, 3, \dots\}$ consisting of all the positive integers is also, for example, discrete, but a sample space $S = \{x : x > 0\}$ consisting of all positive real numbers is not. For the next few chapters we consider only discrete sample spaces. This makes it easier to define mathematical probability, as follows.

Definition 2 *Let $S = \{a_1, a_2, a_3, \dots\}$ be a discrete sample space. Then **probabilities** $P(a_i)$ are numbers attached to the a_i's $(i = 1, 2, 3, \dots)$ such that the following two conditions hold:*

(1) $0 \le P(a_i) \le 1$

(2) $\sum_i P(a_i) = 1$

The set of values $\{P(a_i),\ \ i = 1, 2, \dots\}$ is called a **probability distribution on** S.

Definition 3 *An event in a discrete sample space is a subset $A \subset S$. If the event contains only one point, e.g. $A_1 = \{a_1\}$ we call it a **simple event**. An event A made up of two or more simple events such as $A = \{a_1, a_2\}$ is called a **compound event.***

Our notation will often not distinguish between the point a_i and the simple event $A_i = \{a_i\}$ which has this point as its only element, although they differ as mathematical objects. The condition (2) in the definition above reflects the idea that when the process or experiment happens, some event in S must occur (see the definition of sample space). The probability of a more general event A (not necessarily a simple event) is then defined as follows:

Definition 4 *The probability $P(A)$ of an event A is the sum of the probabilities for all the simple events that make up A.*

For example, the probability of the compound event $A = \{a_1, a_2, a_3\} = P(a_1) + P(a_2) + P(a_3)$. The definition of probability does not say what numbers to assign to the simple events for a given setting, only what properties the numbers must possess. In an actual situation, we try to specify numerical values that make the model useful; this usually means that we try to specify numbers that are consistent with one or more of the empirical "definitions" of Chapter 1.

Example: Suppose a 6-sided die is rolled, and let the sample space be $S = \{1, 2, 3, 4, 5, 6\}$, where 1 means the number 1 occurs, and so on. If the die is an ordinary one, we would find it useful to define probabilities as

$$P(i) = 1/6 \quad \text{for} \quad i = 1, 2, 3, 4, 5, 6,$$

because if the die were tossed repeatedly (as in some games or gambling situations) then each number would occur close to $1/6$ of the time. However, if the die were weighted in some way, these numerical values would not be so useful.

Note that if we wish to consider some compound event, the probability is easily obtained. For example, if A = "even number" then because $A = \{2, 4, 6\}$ we get $P(A) = P(2) + P(4) + P(6) = 1/2$.

We now consider some additional examples, starting with some simple "toy" problems involving cards, coins and dice and then considering a more scientific example.

Remember that in using probability we are actually constructing mathematical models. We can approach a given problem by a series of three steps:

(1) Specify a sample space S.

(2) Assign numerical probabilities to the simple events in S.

(3) For any compound event A, find $P(A)$ by adding the probabilities of all the simple events that make up A.

Many probability problems are stated as "Find the probability that ...". To solve the problem you should then carry out step (2) above by assigning probabilities that reflect long run relative frequencies of occurrence of the simple events in repeated trials, if possible.

Some Examples

When S has few points, one of the easiest methods for finding the probability of an event is to list all outcomes. In many problems a sample space S with equally probable simple events can be used, and the first few examples are of this type.

Example: Draw 1 card from a standard well-shuffled deck (13 cards of each of 4 suits - spades, hearts, diamonds, clubs). Find the probability the card is a club.

Solution 1: Let S = { spade, heart, diamond, club}. (The points of S are generally listed between brackets {}.) Then S has 4 points, with 1 of them being "club", so P(club) = $\frac{1}{4}$.

Solution 2: Let S = {each of the 52 cards}. Then 13 of the 52 cards are clubs, so

$$P(\text{club}) = \frac{13}{52} = \frac{1}{4}.$$

Figure 2.1: 9 tosses of two coins each

Note 1: A sample space is not necessarily unique, as mentioned earlier. The two solutions illustrate this. Note that in the first solution the event A = "the card is a club" is a simple event, but in the second it is a compound event.

Note 2: In solving the problem we have assumed that each simple event in S is equally probable. For example in Solution 1 each simple event has probability $1/4$. This seems to be the only sensible choice of numerical value in this setting. (Why?)

Note 3: The term "odds" is sometimes used. The odds of an event is the probability it occurs divided by the probability it does not occur. In this card example the odds in favour of clubs are 1:3; we could also say the odds against clubs are 3:1.

Example: Toss a coin twice. Find the probability of getting 1 head. (In this course, 1 head is taken to mean **exactly** 1 head. If we meant at least 1 head we would say so.)

Solution 1: Let $S = \{HH, HT, TH, TT\}$ and assume the simple events each have probability $\frac{1}{4}$. (If your notation is not obvious, please explain it. For example, HT means head on the 1^{st} toss and tails on the 2^{nd}.) Since 1 head occurs for simple events HT and TH, we get P (1 head) = $\frac{2}{4} = \frac{1}{2}$.

Solution 2: Let S = { 0 heads, 1 head, 2 heads } and assume the simple events each have probability $\frac{1}{3}$. Then P(1 head) = $\frac{1}{3}$.

Which solution is right? Both are mathematically "correct". However, we want a solution that is useful in terms of the probabilities of events reflecting their relative frequency of occurrence in repeated trials. In that sense, the points in solution 2 are not equally likely. The outcome 1 head occurs more often than either 0 or 2 heads in actual repeated trials. You can experiment to verify this (for example of the nine replications of the experiment in Figure 2.1, 2 heads occurred 2 of the nine times, 1 head occurred 6 of the 9 times. For more certainty you should replicate this experiment many times. You can do this without benefit of coin at http://shazam.econ.ubc.ca/flip/index.html). So we say solution 2 is incorrect for ordinary physical coins though a better term might be "incorrect model". If we were determined to use the sample space in solution 2, we could do it by assigning appropriate probabilities to each point. From solution 1, we can see that 0 heads would have a probability of $\frac{1}{4}$, 1 head $\frac{1}{2}$, and 2 heads $\frac{1}{4}$. However, there seems to be little point using a sample space whose points are not equally probable when one with equally probable points is readily available.

Example: Roll a red die and a green die. Find the probability the total is 5.

Solution: Let (x, y) represent getting x on the red die and y on the green die.
Then, with these as simple events, the sample space is

$$
\begin{array}{llllll}
S = \{ & (1,1) & (1,2) & (1,3) & \cdots & (1,6) \\
 & (2,1) & (2,2) & (2,3) & \cdots & (2,6) \\
 & (3,1) & (3,2) & (3,3) & \cdots & (3,6) \\
 & & -- & -- & -- & \\
 & (6,1) & (6,2) & (6,3) & \cdots & (6,6)\}
\end{array}
$$

The sample points giving a total of 5 are (1,4) (2,3) (3,2), and (4,1).
Therefore P (total is 5) = $\frac{4}{36}$

Example: Suppose the 2 dice were now identical red dice. Find the probability the total is 5.

Solution 1: Since we can no longer distinguish between (x, y) and (y, x), the only distinguishable points in S are :

$$
\begin{array}{llllll}
S = \{ & (1,1) & (1,2) & (1,3) & \cdots & (1,6) \\
 & & (2,2) & (2,3) & \cdots & (2,6) \\
 & & & (3,3) & \cdots & (3,6) \\
 & & & & .. & .. \\
 & & & & & (6,6)\}
\end{array}
$$

Using this sample space, we get a total of 5 from points $(1, 4)$ and $(2, 3)$ only. If we assign equal probability $\frac{1}{21}$ to each point (simple event) then we get P(total is 5) = $\frac{2}{21}$.
At this point you should be suspicious since $\frac{2}{21} \neq \frac{4}{36}$. The colour of the dice shouldn't have any effect on what total we get, so this answer must be wrong. The problem is that the 21 points in S here are not equally likely. If this experiment is repeated, the point (1, 2) occurs twice as often in the long run as the point (1,1). The only sensible way to use this sample space would be to assign probability weights $\frac{1}{36}$

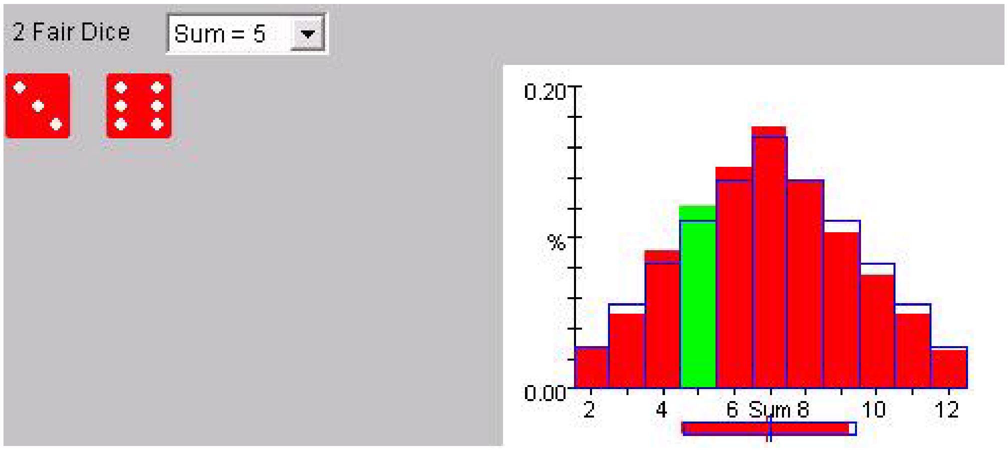


Figure 2.2: Results of 1000 throws of 2 dice

to the points (x, x) and $\frac{2}{36}$ to the points (x, y) for $x \neq y$. Of course we can compare these probabilities with experimental evidence. On the website http://www.math.duke.edu/education/postcalc/probability/dice/index.html you may throw dice up to 10,000 times and record the results. For example on 1000 throws of two dice (see Figure 2.2), there were 121 occasions when the sum of the values on the dice was 5, indicating the probability is around 121/1000 or 0.121 This compares with the true probability $4/36 = 0.111$.

A more straightforward solution follows.

Solution 2: Pretend the dice can be distinguished even though they can't. (Imagine, for example, that we put a white dot on one die, or label one of them 1 and the other as 2.) We then get the same 36 sample points as in the example with the red die and the green die. Hence

$$P(\text{total is } 5) = \frac{4}{36}$$

But, you argue, the dice were identical, and you cannot distinguish them! The laws determining the probabilities associated with these two dice do not, of course, know whether your eyesight is so keen that you can or cannot distinguish the dice. These probabilities must be the same in either case. In many cases, when objects are indistinguishable and we are interested in calculating a probability, the calculation is made easier by pretending the objects can be distinguished.

This illustrates a common pitfall in using probability. When treating objects in an experiment as distinguishable leads to a different answer from treating them as identical, the points in the sample space for identical objects are usually not "equally likely" in terms of their long run relative frequencies. It is generally safer to pretend objects can be distinguished even when they can't be, in order to get equally likely sample points.

While the method of finding probability by listing all the points in S can be useful, it isn't practical when there are a lot of points to write out (e.g., if 3 dice were tossed there would be 216 points in S). We need to have more efficient ways of figuring out the number of outcomes in S or in a compound event without having to list them all. Chapter 3 considers ways to do this, and then Chapter 4 develops other ways to manipulate and calculate probabilities.

To conclude this chapter, we remark that in some settings we rely on previous repetitions of an experiment, or on scientific data, to assign numerical probabilities to events. Problems 2.6 and 2.7 below illustrate this. Although we often use "toy" problems involving things such as coins, dice and simple games for examples, probability is used to deal with a huge variety of practical problems. Problems 2.6 and 2.7, and many others to be discussed later, are of this type.

2.2 Problems on Chapter 2

2.1 Students in a particular program have the same 4 math profs. Two students in the program each independently ask one of their math profs for a letter of reference. Assume each is equally likely to ask any of the math profs.

a) List a sample space for this "experiment".

b) Use this sample space to find the probability both students ask the same prof.

2.2 a) List a sample space for tossing a fair coin 3 times.

b) What is the probability of 2 consecutive tails (but not 3)?

2.3 You wish to choose 2 different numbers from 1, 2, 3, 4, 5. List all possible pairs you could obtain and find the probability the numbers chosen differ by 1 (i.e. are consecutive).

2.4 Four letters addressed to individuals W, X, Y and Z are randomly placed in four addressed envelopes, one letter in each envelope.

(a) List a 24-point sample space for this experiment.

(b) List the sample points belonging to each of the following events:
A: "W's letter goes into the correct envelope";
B: "no letters go into the correct envelopes";
C: "exactly two letters go into the correct envelopes";
D: "exactly three letters go into the correct envelopes".

(c) Assuming that the 24 sample points are equally probable, find the probabilities of the four events in (b).

2.5 (a) Three balls are placed at random in three boxes, with no restriction on the number of balls per box; list the 27 possible outcomes of this experiment. Assuming that the outcomes are all equally probable, find the probability of each of the following events:

A: "the first box is empty";

B: "the first two boxes are empty";

C: "no box contains more than one ball".

(b) Find the probabilities of events A, B and C when three balls are placed at random in n boxes $(n \geq 3)$.

(c) Find the probabilities of events A, B and C when r balls are placed in n boxes $(n \geq r)$.

2.6 **Diagnostic Tests**. Suppose that in a large population some persons have a specific disease at a given point in time. A person can be tested for the disease, but inexpensive tests are often imperfect, and may give either a "false positive" result (the person does not have the disease but the test says they do) or a "false negative" result (the person has the disease but the test says they do not).

In a random sample of 1000 people, individuals with the disease were identified according to a completely accurate but expensive test, and also according to a less accurate but inexpensive test. The results for the less accurate test were that

- 920 persons without the disease tested negative
- 60 persons without the disease tested positive
- 18 persons with the disease tested positive
- 2 persons with the disease tested negative.

(a) Estimate the fraction of the population that has the disease and tests positive using the inexpensive test.

(b) Estimate the fraction of the population that has the disease.

(c) Suppose that someone randomly selected from the population tests positive using the inexpensive test. Estimate the probability that they actually have the disease.

2.7 **Machine Recognition of Handwritten Digits**. Suppose that you have an optical scanner and associated software for determining which of the digits $0, 1, ..., 9$ an individual has written in a square box. The system may of course be wrong sometimes, depending on the legibility of the handwritten number.

(a) Describe a sample space S that includes points (x, y), where x stands for the number actually written, and y stands for the number that the machine identifies.

(b) Suppose that the machine is asked to identify very large numbers of digits, of which $0, 1, ..., 9$ occur equally often, and suppose that the following probabilities apply to the points in your sample space:
$p(0, 6) = p(6, 0) = .004$; $p(0, 0) = p(6, 6) = .096$
$p(5, 9) = p(9, 5) = .005$; $p(5, 5) = p(9, 9) = .095$
$p(4, 7) = p(7, 4) = .002$; $p(4, 4) = p(7, 7) = .098$
$p(y, y) = .100$ for $y = 1, 2, 3, 8$
Give a table with probabilities for each point (x, y) in S. What fraction of numbers is correctly identified?

3. Probability – Counting Techniques

Some probability problems can be attacked by specifying a sample space $S = \{a_1, a_2, \ldots, a_n\}$ in which each simple event has probability $\frac{1}{n}$ (i.e. is “equally likely"). Thus, if a compound event A consists of r simple events, then $P(A) = \frac{r}{n}$. To use this approach we need to be able to count the number of events in S and in A, and this can be tricky. We review here some basic ways to count outcomes from “experiments". These approaches should be familiar from high school mathematics.

3.1 General Counting Rules

There are two basic rules for counting which can deal with most problems. We phrase the rules in terms of “jobs" which are to be done.

1. The **Addition Rule:**

   Suppose we can do job 1 in p ways and job 2 in q ways. Then we can do either job 1 or job 2, but not both, in $p + q$ ways.

For example, suppose a class has 30 men and 25 women. There are $30 + 25 = 55$ ways the prof. can pick one student to answer a question.

2. The **Multiplication Rule:**

   Suppose we can do job 1 in p ways and an unrelated job 2 in q ways. Then we can do both job 1 and job 2 in $p \times q$ ways.

For example, to ride a bike, you must have the chain on both a front sprocket and a rear sprocket. For a 21 speed bike there are 3 ways to select the front sprocket and 7 ways to select the rear sprocket.

This linkage of OR with addition and AND with multiplication will occur throughout the course, so it is helpful to make this association in your mind. The only problem with applying it is that questions do not always have an AND or an OR in them. You often have to play around with re-wording the question for yourself to discover implied AND’s or OR’s.

Example: Suppose we pick 2 numbers at random from digits 1, 2, 3, 4, 5 with replacement. (Note: "with replacement" means that after the first number is picked it is "replaced" in the set of numbers, so it could be picked again as the second number.) Let us find the probability that one number is even. This can be reworded as: "The first number is even AND the second is odd, OR, the first is odd AND the second is even." We can then use the addition and multiplication rules to calculate that there are $(2 \times 3) + (3 \times 2) = 12$ ways for this event to occur. Since the first number can be chosen in 5 ways AND the second in 5 ways, S contains $5 \times 5 = 25$ points. The phrase "at random" in the first sentence means the numbers are equally likely to be picked.

$$\text{Therefore } P(\text{one number is even}) = \frac{12}{25}$$

When objects are selected and replaced after each draw, the addition and multiplication rules are generally sufficient to find probabilities. When objects are drawn **without** being replaced, some special rules may simplify the solution.

Problems:

3.1.1 (a) A course has 4 sections with no limit on how many can enrol in each section. 3 students each randomly pick a section. Find the probability:

(i) they all end up in the same section

(ii) they all end up in different sections

(iii) nobody picks section 1.

(b) Repeat (a) in the case when there are n sections and s students $(n \geq s)$.

3.1.2 Canadian postal codes consist of 3 letters alternated with 3 digits, starting with a letter (e.g. N2L 3G1). For a randomly constructed postal code, what is the probability:

(a) all 3 letters are the same?

(b) the digits are all even or all odd? Treat 0 as being neither even nor odd.

3.1.3 Suppose a password has to contain between six and eight digits, with each digit either a letter or a number from 1 to 9. There must be at least one number present.

(a) What is the total number of possible passwords?

(b) If you started to try passwords in random order, what is the probability you would find the correct password for a given situation within the first 1,000 passwords you tried?

3.2 Permutation Rules

Suppose that n distinct objects are to be “drawn" sequentially, or ordered from left to right in a row. (Order matters; objects are drawn without replacement)

1. *The number of ways to arrange n distinct objects in a row is*

$$n(n-1)(n-2)\cdots(2)(1) = n!$$

Explanation: We can fill the first position in n ways. Since this object can’t be used again, there are only $(n-1)$ ways to fill the second position. So we keep having 1 fewer object available after each position is filled.

Statistics is important, and many games are interesting largely because of the extraordinary rate of growth of the function $n!$ in n.For example

n	0	1	2	3	4	5	6	7	8	9	10
$n!$	1	1	2	6	24	120	720	5040	40320	362880	3628800

which means that for many problems involving sampling from a deck of cards or a reasonably large population, counting the number of cases is virtually impossible. There is an approximation to $n!$ which is often used for large n, called Stirling’s formula which says that $n!$ is *asymptotic* to $n^n e^{-n}\sqrt{2\pi n}$. Here, two sequences a_n and b_n are called asymptotically equal if $a_n/b_n \to 1$ as $n \to \infty$ (intuitively, the percentage error in using Stirling’s approximation goes to zero as $n \to \infty$).For example the error in Stirling’s approximation is less than 1% if $n \geq 8$.

2. *The number of ways to arrange r objects selected from n distinct objects is*

$$n(n-1)(n-2)\cdots(n-r+1),$$

using the same reasoning as in #1, and noting that for the r^{th} selection, $(r-1)$ objects have already been used. Hence there are

$$n-(r-1) = n-r+1$$

ways to make the r^{th} selection. We use the symbol $n^{(r)}$ to represent $n(n-1)\cdots(n-r+1)$ and describe this symbol as “n taken to r terms”. E.g. $6^{(3)} = 6\times5\times4 = 120$.

While $n^{(r)}$ only has a physical interpretation when n and r are positive integers with $n \geq r$, it still has a mathematical meaning when n is not a positive integer, as long as r is a non-negative integer. For example

$$(-2)^{(3)} = (-2)(-3)(-4) = -24 \text{ and}$$
$$1.3^{(2)} = (1.3)(0.3) = 0.39$$

We will occasionally encounter such cases in this course but generally n and r will be non-negative integers with $n \geq r$. In this case, we can re-write $n^{(r)}$ in terms of factorials.

$$n^{(r)} = n(n-1)\cdots(n-r+1)\left[\frac{(n-r)(n-r-1)\cdots(2)(1)}{(n-r)(n-r-1)\cdots(2)(1)}\right] = \frac{n!}{(n-r)!}$$

Note that

$$n^{(0)} = \frac{n!}{(n-0)!} = 1.$$

The idea in using counting methods is to break the experiment into pieces or "jobs" so that counting rules can be applied. There is usually more than one way to do this.

Example: We form a 4 digit number by randomly selecting and arranging 4 digits from 1, 2, 3,…7 without replacement. Find the probability the number formed is (a) even (b) over 3000 (c) an even number over 3000.

Solution: Let S be the set of all possible 4 digit numbers using digits 1, 2, …, 7 without repetitions. Then S has $7^{(4)}$ points. (We could calculate this but it will be easier to leave it in this form for now and do some cancelling later.)

(a) For a number to be even, the last digit must be even. We can fill this last position with a 2, 4, or 6; i.e. in 3 ways. The first 3 positions can be filled by choosing and arranging 3 of the 6 digits not used in the final position. i.e. in $6^{(3)}$ ways. Then there are $3 \times 6^{(3)}$ ways to fill the final position AND the first 3 positions to produce an even number.

$$\text{Therefore } P(\text{even}) = \frac{3 \times 6^{(3)}}{7^{(4)}} = \frac{3 \times 6^{(3)}}{7 \times 6^{(3)}} = \frac{3}{7}$$

Another way to do this problem is to note that the four digit number is even if and only if (iff) the last digit is even. The last digit is equally likely to be any one of the numbers 1, ..., 7 so

$$P(\text{even}) = P(\text{last digit is even}) = \frac{3}{7}$$

(b) To get a number over 3000, we require the first digit to be 3, 4, 5, 6, or 7; i.e. it can be chosen in 5 ways. The remaining 3 positions can be filled in $6^{(3)}$ ways.

$$\text{Therefore } P(\text{number} > 3000) = \frac{5 \times 6^{(3)}}{7^{(4)}} = \frac{5}{7}.$$

Another way to do this problem is to note that the four digit number is over 3000 iff the first digit is one of 3, 4, 5, 6 or 7. Since each of 1, ..., 7 is equally likely to be the first digit, we get P(number $>$ 3000) = $\frac{5}{7}$.

Note that in both (a) and (b) we dealt with positions which had restrictions first, before considering positions with no restrictions. This is generally the best approach to follow in applying counting techniques.

(c) This part has restrictions on both the first and last positions. To illustrate the complication this introduces, suppose we decide to fill positions in the order 1 then 4 then the middle two. We can fill position 1 in 5 ways. How many ways can we then fill position 4? The answer is either 2 or 3 ways, depending on whether the first position was filled with an even or odd digit. Whenever we encounter a situation such as this, we have to break the solution into separate cases. One case is where the first digit is even. The positions can be filled in 2 ways for the first (i.e. with a 4 or 6), 2 ways for the last, and then $5^{(2)}$ ways to arrange 2 of the remaining 5 digits in the middle positions. This first case then occurs in $2 \times 2 \times 5^{(2)}$ ways. The second case has an odd digit in position one. There are 3 ways to fill position one (3, 5, or 7), 3 ways to fill position four (2, 4, or 6), and $5^{(2)}$ ways to fill the remaining positions. Case 2 then occurs in $3 \times 3 \times 5^{(2)}$ ways. We need case 1 OR case 2.

$$\begin{aligned}\text{Therefore } P(\text{even number} > 3000) &= \frac{2 \times 2 \times 5^{(2)} + 3 \times 3 \times 5^{(2)}}{7^{(4)}} \\ &= \frac{13 \times 5^{(2)}}{7 \times 6 \times 5^{(2)}} = \frac{13}{42}\end{aligned}$$

Another way to do this is to realize that we need only to consider the first and last digit, and to find P(first digit is ≥ 3 **and** last digit is even). There are $7 \times 6 = 42$ different choices for (first digit, last digit) and it is easy to see there are 13 choices for which first digit ≥ 3, last digit is even (5×3 minus the impossible outcomes (4, 4) and (6, 6)). Thus the desired probability is $\frac{13}{42}$.

Exercise: Try to solve part (c) by filling positions in the order 4, 1, middle. You should get the same answer.

Exercise: Can you spot the flaw in the following?
There are $3 \times 6^{(3)}$ ways to get an even number (part (a))
There are $5 \times 6^{(3)}$ ways to get a number $\geq$ 3000 (part (b))
By the multiplication rule there are $[3 \times 6^{(3)}] \times [5 \times 6^{(3)}]$ ways to get a number which is even and $>$ 3000. (Read the conditions in the multiplication rule carefully, if you believe this solution.)

Here is another useful rule.

3. *The number of distinct arrangements of n objects when n_1 are alike of one type, n_2 alike of a 2^{nd} type, ..., n_k alike of a k^{th} type* $(\text{with } n_1 + n_2 + \cdots + n_k = n)$ is

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

For example: We can arrange A_1A_2B in 3! ways. These are

$$\begin{array}{lll} A_1A_2B, & A_1BA_2, & BA_1A_2 \\ A_2A_1B, & A_2BA_1, & BA_2A_1 \end{array}$$

However, as soon as we remove the subscripts on the A'^{s}, the second row is the same as the first row. I.e., we have only 3 distinct arrangements since each arrangement appears twice as the A_1 and A_2 are interchanged. In general, there would be n! arrangements if all n objects were distinct. However each arrangement would appear n_1! times as the 1^{st} type was interchanged with itself, n_2! times as the 2^{nd} type was interchanged with itself, etc. Hence only

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

of the n! arrangements are distinct.

Example: 5 men and 3 women sit together in a row. Find the probability that

(a) the same gender is at each end

(b) the women all sit together.

What are you assuming in your solution? Is it likely to be valid in real life?

Solution: If we treat the people as being 8 objects – $5M$ and $3W$, our sample space will have $\frac{8!}{5!3!} = 56$ points.

(a) To get the same gender at each end we need either

$$\mathbf{M}------\mathbf{M} \quad \text{OR} \quad \mathbf{W}------\mathbf{W}$$

The number of distinct arrangements with a man at each end is $\frac{6!}{3!3!} = 20$, since we are arranging $3M$'s and $3W$'s in the middle 6 positions. The number with a woman at each end is $\frac{6!}{5!1!} = 6$. Thus

$$P(\text{same gender at each end}) = \frac{20+6}{56} = \frac{13}{28}$$

assuming each arrangement is equally likely.

(b) Treating WWW as a single unit, we are arranging 6 objects – $5M$'s and 1 WWW. There are $\frac{6!}{5!1!} = 6$ arrangements. Thus,

$$P(\text{women sit together}) = \frac{6}{56} = \frac{3}{28}.$$

Our solution is based on the assumption that all points in S are equally probable. This would mean the people sit in a purely random order. In real life this isn't likely, for example, since friends are more likely to sit together.

Problems:

3.2.1 Digits 1, 2, 3, ..., 7 are arranged at random to form a 7 digit number. Find the probability that

(a) the even digits occur together, in any order

(b) the digits at the 2 ends are both even or both odd.

3.2.2 The letters of the word EXCELLENT are arranged in a random order. Find the probability that

(a) the same letter occurs at each end.

(b) X, C, and N occur together, in any order.

(c) the letters occur in alphabetical order.

3.3 Combinations

This deals with cases where order does not matter; objects are drawn without replacement.
The number of ways to choose r objects from n is denoted by $\binom{n}{r}$ (called "n choose r"). For n and r both non-negative integers with $n \geq r$,

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

Proof: From result 2 earlier, the number of ways to choose r objects from n and arrange them from left to right is $n^{(r)}$. Any choice of r objects can be arranged in $r!$ ways, so we must have

$$\text{(Number of way to choose } r \text{ objects from } n)\times r! = n^{(r)}$$

This gives $\binom{n}{r} = \frac{n^{(r)}}{r!}$ as the number of ways to choose r objects.

Note that $\binom{n}{r}$ loses its physical meaning when n is not a non-negative integer $\geq r$. However it is defined mathematically, provided r is a non-negative integer, by $n^{(r)}/r!$.

$$\text{e.g.,} \left(\begin{array}{c} \frac{1}{2} \\ 3 \end{array} \right) = \frac{(\frac{1}{2})^{(3)}}{3!} = \frac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2})}{3!} = \frac{1}{16}$$

Example: In the Lotto 6/49 lottery, six numbers are drawn at random, without replacement, from the numbers 1 to 49. Find the probability that

(a) the numbers drawn are 1, 2, 3, 4, 5, 6 (in some order)

(b) no even number is drawn.

Solution:

(a) Let the sample space S consist of all combinations of 6 numbers from 1, ..., 49; there are $\binom{49}{6}$ of them. Since 1, 2, 3, 4, 5, 6 consist of one of these 6-tuples, $P(\{1, 2, 3, 4, 5, 6\}) = 1/\binom{49}{6}$, which equals about 1 in 13.9 million.

(b) There are 25 odd and 24 even numbers, so there are $\binom{25}{6}$ choices in which all the numbers are odd.

Therefore P (no even number) $=$ P (all odd numbers)
$= \binom{25}{6} / \binom{49}{6}$

which is approximately equal to 0.0127.

Example: Find the probability a bridge hand (13 cards picked at random from a standard deck) has

(a) 3 aces

(b) at least 1 ace

(c) 6 spades, 4 hearts, 2 diamonds, 1 club

(d) a 6-4-2-1 split between the 4 suits

(e) a 5-4-2-2 split.

Solution: Since order of selection does not matter, we take S to have $\binom{52}{13}$ points.

(a) We can choose 3 aces in $\binom{4}{3}$ ways. We also have to choose 10 other cards from the 48 non-aces. This can be done in $\binom{48}{10}$ ways. Hence $P(3 \text{ aces}) = \frac{\binom{4}{3}\binom{48}{10}}{\binom{52}{13}}$

(b) **Solution 1:** At least 1 ace means 1 ace or 2 aces or 3 aces or 4 aces. Calculate each part as in (a) and use the addition rule to get

$$\frac{\binom{4}{1}\binom{48}{12} + \binom{4}{2}\binom{48}{11} + \binom{4}{3}\binom{48}{10} + \binom{4}{4}\binom{48}{9}}{\binom{52}{13}}$$

Solution 2: If we subtract all cases with 0 aces from the $\binom{52}{13}$ points in S we are left with all points having at least 1 ace. This gives

$$P(\geq 1 \text{ ace}) = \frac{\binom{52}{13} - \binom{4}{0}\binom{48}{13}}{\binom{52}{13}} = 1 - P(0 \text{ aces}).$$

(The term $\binom{4}{0}$ can be omitted since $\binom{4}{0} = 1$, but was included here to show that we were choosing 0 of the 4 aces.)

Solution 3: This solution is incorrect, but illustrates a common error. Choose 1 of the 4 aces then any 12 of the remaining 51 cards. This guarantees we have at least 1 ace, so

$$P(\geq 1 \text{ ace}) = \frac{\binom{4}{1}\binom{51}{12}}{\binom{52}{13}}$$

The flaw in this solution is that it counts some points more than once by partially keeping track of order. For example, we could get the ace of spades on the first choice and happen to get the ace of clubs in the last 12 draws. We also could get the ace of clubs on the first draw and then get the ace of spades in the last 12 draws. Though in both cases we have the same outcome, they would be counted as 2 different outcomes.

(c) Choose the 6 spades in $\binom{13}{6}$ ways and the hearts in $\binom{13}{4}$ ways and the diamonds in $\binom{13}{2}$ ways and the clubs in $\binom{13}{1}$ ways.

$$\text{Therefore } P(6S-4H-2D-1C) = \frac{\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}}$$

(d) The split in (c) is only 1 of several possible 6-4-2-1 splits. In fact, filling in the numbers 6, 4, 2 and 1 in the spaces above each suit $\overline{S}\ \overline{H}\ \overline{D}\ \overline{C}$ defines a 6-4-2-1 split. There are $4!$ ways to do this, and then $\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}$ ways to pick the cards from these suits.

$$\text{Therefore } P(6-4-2-1 \text{ split}) = \frac{4!\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}}$$

(e) This is the same as (d) except the numbers 5-4-2-2 are not all different. There are $\frac{4!}{2!}$ different arrangements of 5-4-2-2 in the spaces $\overline{S}\ \overline{H}\ \overline{D}\ \overline{C}$.

$$\text{Therefore } P(5-4-2-2 \text{ split}) = \frac{\frac{4!}{2!}\binom{13}{5}\binom{13}{4}\binom{13}{2}\binom{13}{2}}{\binom{52}{13}}$$

Problems:

3.3.1 A factory parking lot has 160 cars in it, of which 35 have faulty emission controls. An air quality inspector does spot checks on 8 cars on the lot.

(a) Give an expression for the probability that at least 3 of these 8 cars will have faulty emission controls.

(b) What assumption does your answer to (a) require? How likely is it that this assumption holds if the inspector hopes to catch as many cars with faulty controls as possible?

3.3.2 In a race, the 15 runners are randomly assigned the numbers $1, 2, \cdots, 15$. Find the probability that

(a) 4 of the first 6 finishers have single digit numbers.

(b) the fifth runner to finish is the 3rd finisher with a single digit number.

(c) number 13 is the highest number among the first 7 finishers.

3.4 Problems on Chapter 3

3.1 Six digits from 2, 3, 4, $\ldots$, 8 are chosen and arranged in a row without replacement. Find the probability that

(a) the number is divisible by 2

(b) the digits 2 and 3 appear consecutively in the proper order (i.e. 23)

(c) digits 2 and 3 appear in the proper order but not consecutively.

3.2 Suppose r passengers get on an elevator at the basement floor. There are n floors above (numbered 1, 2, 3, $\ldots$, n) where passengers may get off.

(a) Find the probability

(i) no passenger gets off at floor 1

(ii) passengers all get off at different floors $(n \geq r)$.

(b) What assumption(s) underlies your answer to (a)? Comment briefly on how likely it is that the assumption(s) is valid.

3.3 There are 6 stops left on a subway line and 4 passengers on a train. Assume they are each equally likely to get off at any stop. What is the probability

(a) they all get off at different stops?

(b) 2 get off at one stop and 2 at another stop?

3.4 Give an expression for the probability a bridge hand of 13 cards contains 2 aces, 4 face cards (Jack, Queen or King) and 7 others. You might investigate the various permutations and combinations relating to card hands using the Java applet at $http://www.wcrl.ars.usda.gov/cec/java/comb.htm$

3.5 The letters of the word STATISTICS are arranged in a random order. Find the probability

(a) they spell statistics

(b) the same letter occurs at each end.

3.6 Three digits are chosen in order from 0, 1, 2, . . . , 9. Find the probability the digits are drawn in increasing order; (i.e., the first $<$ the second $<$ the third) if

(a) draws are made without replacement

(b) draws are made with replacement.

3.7 **The Birthday Problem.** [2] Suppose there are r persons in a room. Ignoring February 29 and assuming that every person is equally likely to have been born on any of the 365 other days in a year, find the probability that no two persons in the room have the same birthday. Find the numerical value of this probability for $r = 20, 40$ and 60. There is a graphic Java applet for illustrating the frequency of common birthdays at http://www-stat.stanford.edu/%7Esusan/surprise/Birthday.html

3.8 You have n identical looking keys on a chain, and one opens your office door. If you try the keys in random order then

(a) what is the probability the kth key opens the door?

(b) what is the probability one of the first two keys opens the door (assume $n \geq 3$)?

(c) Determine numerical values for the answer in part (b) for the cases $n = 3, 5, 7$.

3.9 From a set of $2n + 1$ consecutively numbered tickets, three are selected at random without replacement. Find the probability that the numbers of the tickets form an arithmetic progression. [The *order* in which the tickets are selected does *not* matter.]

3.10 The 10,000 tickets for a lottery are numbered 0000 to 9999. A four-digit winning number is drawn and a prize is paid on each ticket whose four-digit number is any *arrangement* of the number drawn. For instance, if winning number 0011 is drawn, prizes are paid on tickets numbered 0011, 0101, 0110, 1001, 1010, and 1100. A ticket costs $1 and each prize is $500.

(a) What is the probability of winning a prize (i) with ticket number 7337? (ii) with ticket number 7235? What advice would you give to someone buying a ticket for this lottery?

(b) Assuming that all tickets are sold, what is the probability that the operator will lose money on the lottery?

[2]" My birthday was a natural disaster, a shower of paper full of flattery under which one almost drowned" Albert Einstein, 1954 on his seventy-fifth birthday.

3.11 (a) There are 25 deer in a certain forested area, and 6 have been caught temporarily and tagged. Some time later, 5 deer are caught. Find the probability that 2 of them are tagged. (What assumption did you make to do this?)

(b) Suppose that the total number of deer in the area was unknown to you. Describe how you could estimate the number of deer based on the information that 6 deer were tagged earlier, and later when 5 deer are caught, 2 are found to be tagged. What estimate do you get?

3.12 **Lotto 6/49**. In Lotto 6/49 you purchase a lottery ticket with 6 different numbers, selected from the set $\{1, 2, ..., 49\}$. In the draw, six (different) numbers are randomly selected. Find the probability that

(a) Your ticket has the 6 numbers which are drawn. (This means you win the main Jackpot.)

(b) Your ticket matches exactly 5 of the 6 numbers drawn.

(c) Your ticket matches exactly 4 of the 6 numbers drawn.

(d) Your ticket matches exactly 3 of the 6 numbers drawn.

3.13 (**Texas Hold-em**) Texas Hold-em is a poker game in which players are each dealt two cards face down (called your hole or pocket cards), from a standard deck of 52 cards, followed by a round of betting, and then five cards are dealt face up on the table with various breaks to permit players to bet the farm. These are communal cards that anyone can use in combination with their two pocket cards to form a poker hand. Players can use any five of the face-up cards and their two cards to form a five card poker hand. Probability calculations for this game are not only required at the end, but also at intermediate steps and are quite complicated so that usually simulation is used to determine the odds that you will win given your current information, so consider a simple example. Suppose we were dealt 2 Jacks in the first round.

(a) What is the probability that the next three cards (face up) include at least one Jack?

(b) Given that there was no Jack among these next three cards, what is the probability that there is at least one among the last two cards dealt face-up?

(c) What is the probability that the 5 face-up cards show two Jacks, given that I have two in my pocket cards?

4. Probability Rules and Conditional Probability

4.1 General Methods

In the mathematical definition of probability, an arbitrary event A is merely some subset of the sample space S. The following rules hold:

1. $P(S) = 1$

2. For any event $A,\ 0 \leq P(A) \leq 1$

It is also obvious from our definitions in Chapter 2 that if A and B are two events with $A \subseteq B$ (that is, all of the simple events in A are also in B), then $P(A) \leq P(B)$.

It is often helpful to use elementary ideas of set theory in dealing with probability; as we show in this chapter, this allows certain rules or propositions about probability to be proved. Before going on to specific rules, we'll review Venn diagrams for sets. In the drawings below, think of all points in S being contained in the rectangle, and those points where particular events occur being contained in circles. We begin by considering the union $(A \cup B)$, intersection $(A \cap B)$ and complement $(\bar{A})$ of sets (see Figure 4.3). At the URL http://stat-www.berkeley.edu/users/stark/Java/Venn.htm, there is an interesting applet which allows you to vary the area of the intersection and construct Venn diagrams for a variety of purposes.

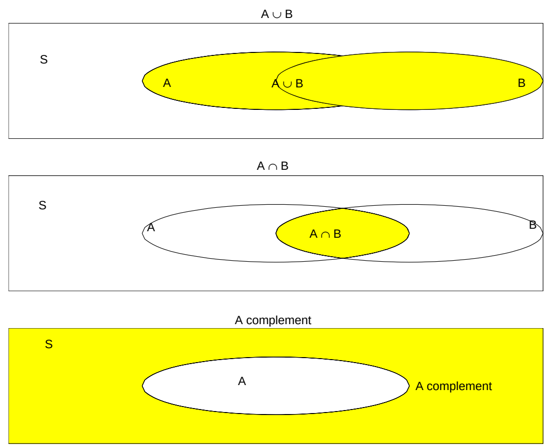


Figure 4.3: **Top panel:** $A \cup B$ means A OR B (or possibly both) occurs.$A \cup B$ is shaded.
Middle panel: $A \cap B$ (usually written as AB in probability) means A and B both occur. $A \cap B$ is shaded
Lower panel: $\bar{A}$ means A does not occur. $\bar{A}$ is shaded

Example:

Suppose for students finishing 2A Math that 22% have a math average $\geq$ 80%, 24% have a STAT 230 mark $\geq$ 80%, 20% have an overall average $\geq$ 80%, 14% have both a math average and STAT 230 $\geq$ 80%, 13% have both an overall average and STAT 230 $\geq$ 80%, 10% have all 3 of these averages $\geq$ 80%, and 67% have none of these 3 averages $\geq$ 80%. Find the probability a randomly chosen math student finishing 2A has math and overall averages both $\geq$ 80% and STAT 230 $<$ 80%.

Solution: When using rules of probability it is generally helpful to begin by labeling the events of interest.

$$\begin{aligned} \text{Let} \quad A &= \{\text{math average} \geq 80\%\} \\ B &= \{\text{overall average } \geq 80\%\} \\ C &= \{\text{STAT 230} \geq 80\%\} \end{aligned}$$

In terms of these symbols, we are given $P(A) = .22$, $P(B) = .20$, $P(C) = .24$, $P(AC) = .14$, $P(BC) = .13$, $P(ABC) = .1$, and $P(\bar{A}\bar{B}\bar{C}) = .67$. We are asked to find $P(AB\bar{C})$, the shaded region in Figure 4.4 Filling in this information on a Venn diagram, in the order indicated by (1), (2), (3), etc.

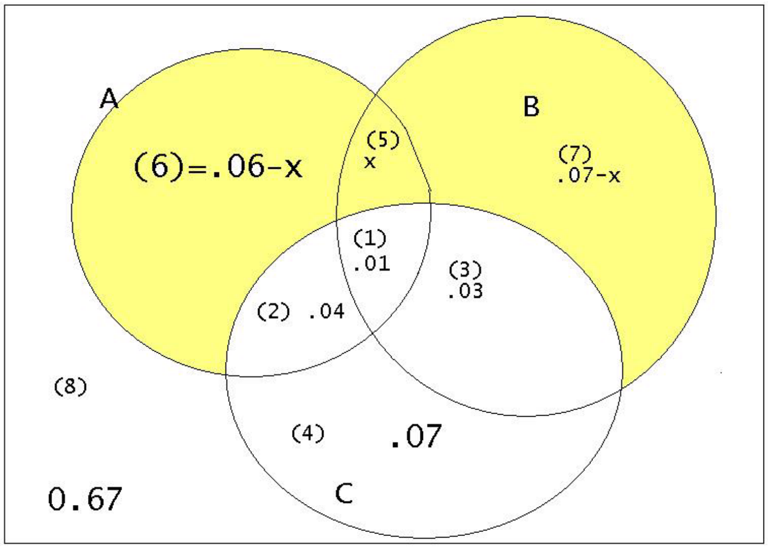


Figure 4.4: Venn Diagram for Math Averages Example

(1) given
(2) $P(AC) - P(ABC)$
(3) $P(BC) - P(ABC)$
(4) $P(C) - P(AC) - .03$
(5) unknown
(6) $P(A) - P(AC) - x$
(7) $P(B) - P(BC) - x$
(8) given

(Usually, we start filling in at the centre and work our way out.)
Adding all probabilities and noting that $P(S) = 1$, we can solve to get $x = .06 = P(AB\bar{C})$.

Problems:

4.1.1 In a typical year, 20% of the days have a high temperature $> 22^o$C. On 40% of these days there is no rain. In the rest of the year, when the high temperature $\leq 22^o$C, 70% of the days have no rain. What percent of days in the year have rain and a high temperature $\leq 22^o$C?

4.1.2 According to a survey of people on the last Ontario voters list, 55% are female, 55% are politically to the right, and 15% are male and politically to the left. What percent are female and politically to the right? Assume voter attitudes are classified simply as left or right.

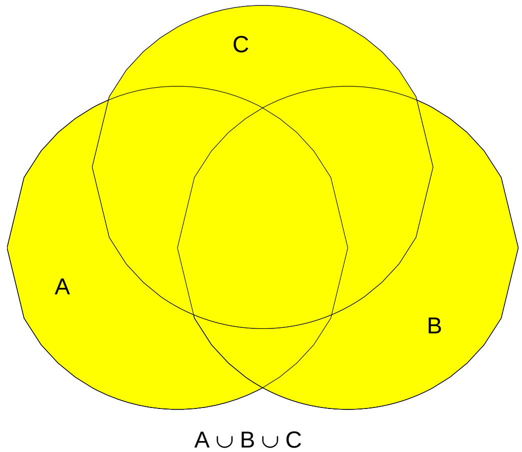


Figure 4.5: The union $A \cup B \cup C$

4.2 Rules for Unions of Events

In addition to the two rules which govern probabilities listed in Section 4.1, we have the following

3. (*probability of unions*)

$$\text{(a)} \qquad P(A \cup B) = P(A) + P(B) - P(AB)$$

This can be obtained by using a Venn diagram. Each point in $A \cup B$ must be counted once. Since points in AB are counted twice - once in $P(A)$ and once in $P(B)$ - they need to be subtracted once.

$$\text{(b)} \qquad P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(AB) + P(AC) + P(BC)] + P(ABC)$$

(see Figure 4.5)

$$\text{(c)} \qquad P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = \sum P(A_i) - \sum P(A_iA_j) + \sum P(A_iA_jA_k) - \sum P(A_iA_jA_kA_l) + \cdots$$

(where the subscripts are all different)
This generalization is seldom used in Stat 230.

Definition 5 *Events A and B are **mutually exclusive** if $AB = \phi$ (the null set)*

Since mutually exclusive events A and B have no common points, $P(AB) = 0$.

In general, events $A_1, A_2 \cdots, A_n$ are mutually exclusive if $A_i A_j = \phi$ for all $i \neq j$. This means that there is no chance of 2 or more of these events occurring together. For example, if a die is rolled twice, the events

$$A = \{\text{2 occurs on the 1st roll}\} \text{ and}$$
$$B = \{\text{the total is 10}\}$$

are mutually exclusive. In the case of mutually exclusive events, rule 3 above simplifies to rule 4 below.

Exercise:
Think of some pairs of events and classify them as being mutually exclusive or not mutually exclusive.

4. (*unions of mutually exclusive events*)

(a) Let A and B be mutually exclusive events. Then $P(A \cup B) = P(A) + P(B)$

(b) In general, let $A_1, A_2, \cdots A_n$ be mutually exclusive.
Then $P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum\limits_{i=1}^{n} P(A_i)$
Proof: Use rule 3 above

5. (*probability of complements*) $P(A) = 1 - P(\bar{A})$

Proof:

A and $\bar{A}$ are mutually exclusive so

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

But

$$A \cup \bar{A} = S \text{ and } P(S) = 1$$

$$\text{Therefore } P(A) + P(\bar{A}) = 1$$
$$P(A) = 1 - P(\bar{A})$$

This result is useful whenever $P(\bar{A})$ is easier to obtain than $P(A)$.

Example: Two ordinary dice are rolled. Find the probability that at least one of them turns up a 6.
Solution 1: Let A = { 6 on the first die }, B = { 6 on the second die } and note (rule 3) that

$$\begin{aligned} P(\geq \text{one } 6) &= P(A \cup B) \\ &= P(A) + P(B) - P(AB) \\ &= \tfrac{1}{6} + \tfrac{1}{6} - \tfrac{1}{36} \\ &= \tfrac{11}{36} \end{aligned}$$

Solution 2:

$$\begin{aligned} P(\geq \text{one } 6) &= 1 - P(\text{no 6 on either die}) \\ &= 1 - \tfrac{25}{36} \\ &= \tfrac{11}{36} \end{aligned}$$

Example: Roll a die 3 times. Find the probability of getting at least one 6.

Solution 1:
Let $A = \{\geq \text{one } 6\}$. Then $\bar{A} = \{\text{no } 6^{\text{'s}}\}$.
Using counting arguments, there are 6 outcomes on each roll, so S has $6 \times 6 \times 6 = 216$ points. For $\bar{A}$ to occur we can't have a 6 on any roll. Then $\bar{A}$ can occur in $5 \times 5 \times 5 = 125$ ways.

$$\text{Therefore } P(\bar{A}) = \frac{125}{216}. \quad \text{Hence } P(A) = 1 - \frac{125}{216} = \frac{91}{216}$$

Solution 2: Can you spot the flaw in this?
Let A = {6 occurs on 1^{st} roll}
B = {6 occurs on 2^{nd} roll}
C = {6 occurs on 3^{rd} roll}.
Then $P(\geq \text{one } 6) = P(A \cup B \cup C)$
= $P(A) + P(B) + P(C)$
= $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

You should have noticed that A, B, and C are <u>not</u> mutually exclusive events, so we should have used

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

Each of AB, AC, and BC occurs only once in the 36 point sample space for those two rolls.

$$\text{Therefore } P(A \cup B \cup C) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{36} - \frac{1}{36} - \frac{1}{36} + \frac{1}{216} = 91/216.$$

Note: Rules 3, 4, and (indirectly) 5 link the concepts of addition, unions and complements. The next segment will consider intersection, multiplication of probabilities, and a concept known as independence. Making these linkages will make problem solving and the construction of probability models easier.

Problems:

4.2.1 Let $A, B,$ and C be events for which

$$P(A) = 0.2, P(B) = 0.5, P(C) = 0.3 \text{ and } P(AB) = 0.1$$

(a) Find the largest possible value for $P(A \cup B \cup C)$

(b) For this largest value to occur, are the events A and C mutually exclusive, not mutually exclusive, or is this unable to be determined?

4.2.2 Prove that $P(A \cup B) = 1 - P(\overline{A}\,\overline{B})$ for arbitrary events A and B in S.

4.3 Intersections of Events and Independence

Dependent and Independent Events:

Consider these two groups of pairs of events.

		Group 1
A	=	{airplane engine fails in flight}
B	=	{airplane reaches its destination safely}
or		(when a fair coin is tossed twice)
A	=	{H is on 1st toss}
B	=	{H on both tosses}.

		Group 2
A	=	{a coin toss shows heads}
B	=	{a bridge hand has 4 aces}.
or		(when a fair coin is tossed twice)
A	=	{H on 1st toss}
B	=	{H on 2nd toss}

What do the pairs A, B in each group have in common? In group 1 the events are related so that the occurrence of A affects the chances of B occurring. In group 2, whether A occurs or not has no effect on B's occurrence.

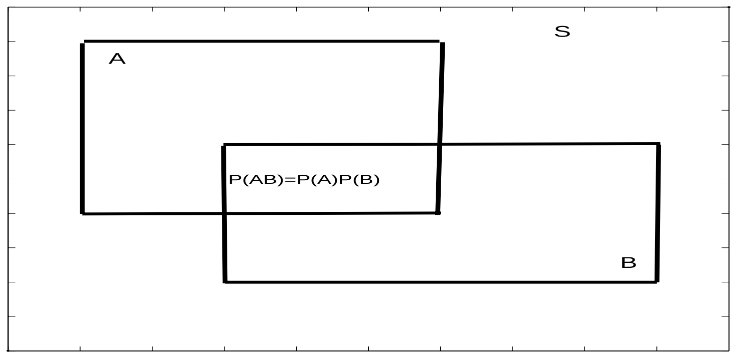


Figure 4.6: Independent Events A, B

We call the pairs in group 1 dependent events, and those in group 2 independent events. We formalize this concept in the mathematical definition which follows.

Definition 6 *Events A and B are* ***independent*** *if and only if $P(AB) = P(A)P(B)$. If they are not independent, we call the events* ***dependent****.*

If two events are independent, then the "size" of their intersection as measured by the probability measure is required to be the product of the individual probabilities. This means, of course, that the intersection must be non-empty, and so the events are not mutually exclusive. For example in the Venn diagram depicted in Figure 4.6, $P(A) = 0.3, P(B) = 0.4$ and $P(AB) = 0.12$ so in this case the two events are independent.

For another example, suppose we toss a fair coin twice. Let A = {head on 1st toss} and B = {head on 2nd toss}. Clearly A and B are independent since the outcome on each toss is unrelated to other tosses, so $P(A) = \frac{1}{2}, P(B) = \frac{1}{2}, P(AB) = \frac{1}{4} = P(A)P(B)$.
However, if we roll a die once and let A = {the number is even} and B = {number > 3} the events will be dependent since

$$P(A) = \frac{1}{2}, P(B) = \frac{1}{2}, P(AB) = P(4 \text{ or } 6 \text{ occurs}) = \frac{2}{6} \neq P(A)P(B).$$

(Rationale: B only happens half the time. If A occurs we know the number is 2, 4, or 6. So B occurs $\frac{2}{3}$ of the time when A occurs. The occurrence of A does affect the chances of B occurring so A and B are <u>not</u> independent.)

When there are more than 2 events, the above definition generalizes to:

Definition 7 *The events $A_1, A_2, \cdots, A_n$ are independent if and only if*

$$P(A_{i_1}, A_{i_2}, \cdots, A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

for all sets $(i_1, i_2, \cdots, i_k)$ of distinct subscripts chosen from $(1, 2, \cdots, n)$

For example, for $n = 3$, we need

$$\begin{aligned}
P(A_1A_2) &= P(A_1)P(A_2),\\
P(A_1A_3) &= P(A_1)P(A_3),\\
P(A_2A_3) &= P(A_2)P(A_3)
\end{aligned}$$

and

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$$

Technically, we have defined "mutually independent" events, but we will shorten the name to "independent" to reduce confusion with "mutually exclusive."
The definition of independence works two ways. If we can find $P(A), P(B)$, and $P(AB)$ then we can determine whether A and B are independent. Conversely, if we know (or assume) that A and B are independent, then we can use the definition as a rule of probability to calculate $P(AB)$. Examples of each follow.

Example: Toss a die twice. Let A = {first toss is a 3} and B = {the total is 7}. Are A and B independent? (What do you think?) Using the definition to check, we get $P(A) = \frac{1}{6}, P(B) = \frac{6}{36}$ (points (1,6), (2,5), (3,4), (4,3), (5,2) and (6,1) give a total of 7) and $P(AB) = \frac{1}{36}$ (only the point (3,4) makes AB occur).
Therefore, $P(AB) = P(A)P(B)$ and so A and B are independent events.
Now suppose we change B to the event {total is 8}.
Then

$$P(A) = \frac{1}{6}, \;\; P(B) = \frac{5}{36} \;\text{ and }\; P(AB) = \frac{1}{36}$$

Therefore $P(AB) \neq P(A)P(B)$

and consequently A and B are dependent events.

This example often puzzles students. Why are they independent if B is a total of 7 but dependent for a total of 8? The key is that regardless of the first toss, there is always one number on the 2nd toss which makes the total 7. Since the probability of getting a total of 7 started off being $\frac{6}{36} = \frac{1}{6}$, the outcome of the 1st toss doesn't affect the chances. However, for any total other than 7, the outcome of the 1st toss does affect the chances of getting that total (e.g., a first toss of 1 guarantees the total cannot be 8).

Example: A (pseudo) random number generator on the computer can give a sequence of independent random digits chosen from $S = \{0, 1, \ldots, 9\}$. This means that (i) each digit has probability of $\frac{1}{10}$ of being any of $0, 1, \ldots, 9$, and (ii) the outcomes for the different trials are independent of one another. We call this type of setting an "experiment with independent trials". Determine the probability that

(a) in a sequence of 5 trials, all the digits generated are odd

(b) the number 9 occurs for the first time on trial 10.

Solution:

(a) Define the events A_i = {digits from trial i is odd}, $i = 1, \ldots, 5$.
Then

$$\begin{aligned} P(\text{all digits are odd}) &= P(A_1A_2A_3A_4A_5) \\ &= \prod_{i=1}^{5} P(A_i), \end{aligned}$$

since the A_i's are mutually independent. Since $P(A_i) = .5$, we get P(all digits are odd) = $.5^5$.

(b) Define events A_i = {9 occurs on trial i}, $i = 1, 2, \ldots$. Then we want

$$\begin{aligned} P(\bar{A}_1\bar{A}_2\ldots\bar{A}_9A_{10}) &= P(\bar{A}_1)P(\bar{A}_2)\ldots P(\bar{A}_9)P(A_{10}) \\ &= (.9)^9(.1), \end{aligned}$$

because the A_i's are independent, and $P(A_i) = .1 = 1 - P(\bar{A}_i)$.

Note: We have used the fact here that if A and B are independent events, then so are $\bar{A}$ and B. To see this note that

$$\begin{aligned} P(\bar{A}B) &= P(B) - P(AB) && \text{(use a Venn diagram)} \\ &= P(B) - P(A)P(B) && \text{(since } A \text{ and } B \text{ are independent)} \\ &= (1 - P(A))P(B) \\ &= P(\bar{A})P(B). \end{aligned}$$

Note: We have implicitly assumed independence of events in some of our earlier probability calculations. For example, suppose a coin is tossed 3 times, and we consider the sample space

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

Assuming that the outcomes on the three tosses are independent, and that

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

on any single toss, we get that

$$P(HHH) = P(H)P(H)P(H) = (\frac{1}{2})^3 = \frac{1}{8}.$$

Similarly, all the other simple events have probability $\frac{1}{8}$. Note that in earlier calculations we assumed this was true without thinking directly about independence. However, it is clear that if somehow the 3 tosses were not independent then it might be a bad idea to assume each simple event had probability $\frac{1}{8}$. (For example, instead of heads and tails, suppose H stands for "rain" and T stands for "no rain" on a given day; now consider 3 consecutive days. Would you want to assign a probability of $\frac{1}{8}$ to each of the 8 simple events?)

Note: The definition of independent events can thus be used either to check for independence or, if events are known to be independent, to calculate $P(AB)$. Many problems are not obvious, and scientific study is needed to determine if two events are independent. For example, are the events A and B independent if, for a random child living in a country,
A = {live within 5 km. of a nuclear power plant}
B = {a child has leukemia}?
Such problems, which are of considerable importance, can be handled by methods in later statistics courses.

Problems:

4.3.1 A weighted die is such that $P(1) = P(2) = P(3) = 0.1, P(4) = P(5) = 0.2$, and $P(6) = 0.3$.

(a) If the die is thrown twice what is the probability the total is 9?

(b) If a die is thrown twice, and this process repeated 4 times, what is the probability the total will be 9 on exactly 1 of the 4 repetitions?

4.3.2 Suppose among UW students that 15% speaks French and 45% are women. Suppose also that 20% of the women speak French. A committee of 10 students is formed by randomly selecting from UW students. What is the probability there will be at least 1 woman and at least 1 French speaking student on the committee?

4.3.3 Prove that $\overline{A}$ and $\overline{B}$ are independent events if and only if $\overline{A}$ and B are independent.

4.4 Conditional Probability

In many situations we may want to determine the probability of some event A, while knowing that some other event B has already occurred. For example, what is the probability a randomly selected person is over 6 feet tall, given that they are female? Let the symbol $P(A|B)$ represent the probability that event A occurs, when we know that B occurs. We call this the conditional probability of A given B. While we will give a definition of $P(A|B)$, let's first consider an example we looked at earlier, to get some sense of why $P(A|B)$ is defined as it is.

Example: Suppose we roll a die once. Let A = {the number is even} and B = {number > 3}. If we know that B occurs, that tells us that we have a 4, 5, or 6. Of the times when B occurs, we have an even number $\frac{2}{3}$ of the time. So $P(A|B) = \frac{2}{3}$. More formally, we could obtain this result by calculating $\frac{P(AB)}{P(B)}$, since $P(AB) = P(4 \textit{ or } 6) = \frac{2}{6}$ and $P(B) = \frac{3}{6}$.

Definition 8 *the conditional probability of event A, given event B, is*

$$P(A|B) = \frac{P(AB)}{P(B)}, \textit{provided } \; P(B) \neq 0.$$

Note: If A and B are independent,

$$\begin{aligned} P(AB) &= P(A)P(B) \text{ so} \\ P(A|B) &= \frac{P(A)P(B)}{P(B)} = P(A). \end{aligned}$$

This makes sense, and can be taken as an equivalent definition of independence; that is, A and B are independent iff $P(A|B) = P(A)$. You should investigate the behaviour of the conditional probabilities as we move the events around on the web-site http://stat-www.berkeley.edu/%7Estark/Java/Venn3.htm.

Example: If a fair coin is tossed 3 times, find the probability that if at least 1 Head occurs, then exactly 1 Head occurs.

Solution: Define the events A = {1 Head}, B = {at least 1 Head}. What we are being asked to find is $P(A|B)$. This equals $P(AB)/P(B)$, and so we find

$$P(B) = 1 - P(0 \text{ Heads}) \ = \frac{7}{8}$$

and

$$\begin{aligned} P(AB) = P\{1 \text{ Head}\} &= P(\{HTT, THT, TTH\}) \\ &= \tfrac{3}{8} \end{aligned}$$

using either the sample space with 8 points, or the fact that the 3 tosses are independent. Thus,

$$P(A|B) = \frac{\frac{3}{8}}{\frac{7}{8}} = \frac{3}{7}.$$

Example: The probability a randomly selected male is colour-blind is .05, whereas the probability a female is colour-blind is only .0025. If the population is 50% male, what is the fraction that is colour-blind?

Solution: Let

C = {person selected is colour-blind}

M = {person selected is male}

F = {person selected is female}

We are asked to find $P(C)$. We are told that

$$P(C|M) = 0.05, \quad P(C|F) = 0.0025, \qquad P(M) = 0.5 = P(F).$$

To get $P(C)$ we can therefore use

$$\begin{aligned} P(C) &= P(CM) + P(CF) \\ &= P(C|M)P(M) + P(C|F)P(F) \\ &= (0.05)(0.5) + (0.0025)(0.05) \\ &= 0.02625. \end{aligned}$$

4.5 Multiplication and Partition Rules

The preceding example suggests two more probability rules, which turn out to be extremely useful. They are based on breaking events of interest into pieces.

6. *Multiplication Rules*

   Let $A, B, C, D, \ldots$ be arbitrary events in a sample space. Then

$$\begin{aligned} P(AB) &= P(A)P(B|A) \\ P(ABC) &= P(A)P(B|A)P(C|AB) \\ P(ABCD) &= P(A)P(B|A)P(C|AB)P(D|ABC) \end{aligned}$$

and so on.

Proof:
The first rule comes directly from the definition $P(B|A)$. The right hand side of the second rule equals (assuming $P(AB) \neq 0$)

$$P(AB)P(C|AB) = P(AB)\frac{P(CAB)}{P(AB)} = P(ABC),$$

and so on.

Partition Rule
Let $A_1, \ldots, A_k$ be a partition of the sample space S into disjoint (mutually exclusive) events such that

$$A_1 \cup A_2 \cup \cdots \cup A_k = S.$$

Let B be an arbitrary event in S. Then

$$\begin{aligned} P(B) &= P(BA_1) + P(BA_2) + \cdots + P(BA_k) \\ &= \sum_{i=1}^{k} P(B|A_i)P(A_i) \end{aligned}$$

Proof: Look at a Venn diagram to see that $BA_1, \ldots, BA_k$ are mutually exclusive, with $B = (BA_1) \cup \cdots \cup (BA_k)$.

Example: In an insurance portfolio 10% of the policy holders are in Class A_1 (high risk), 40% are in Class A_2 (medium risk), and 50% are in Class A_3 (low risk). The probability a Class A_1 policy has a claim in a given year is .10; similar probabilities for Classes A_2 and A_3 are .05 and .02. Find the probability that if a claim is made, it is for a Class A_1 policy.

Solution: For a randomly selected policy, let

B = {policy has a claim }

A_i = {policy is of Class A_i}, $i = 1, 2, 3$

We are asked to find $P(A_1|B)$. Note that

$$P(A_1|B) = \frac{P(A_1B)}{P(B)}$$

and that

$$P(B) = P(A_1B) + P(A_2B) + P(A_3B).$$

We are told that

$$P(A_1) = 0.10, P(A_2) = 0.40, P(A_3) = 0.50$$

and that

$$P(B|A_1) = 0.10, P(B|A_2) = 0.05, P(B|A_3) = 0.02.$$

Thus

$$\begin{aligned} P(A_1B) = P(A_1)P(B|A_1) &= .01 \\ P(A_2B) = P(A_2)P(B|A_2) &= .02 \\ P(A_3B) = P(A_3)P(B|A_3) &= .01 \end{aligned}$$

Therefore $P(B) = .04$ and $P(A_1|B) = .01/.04 = .25$.

Tree Diagrams

Tree diagrams can be a useful device for keeping track of conditional probabilities when using multiplication and partition rules. The idea is to draw a tree where each path represents a sequence of events. On any given branch of the tree we write the conditional probability of that event given all the events on branches leading to it. The probability at any node of the tree is obtained by multiplying the probabilities on the branches leading to the node, and equals the probability of the intersection of the events leading to it.

For example, the immediately preceding example could be represented by the tree in Figure 4.7. Note that the probabilities on the terminal nodes must add up to 1.

Here is another example involving diagnostic tests for disease. See if you can represent the problem by a tree.

Example. Testing for HIV

Tests used to diagnose medical conditions are often imperfect, and give false positive or false negative results, as described in Problem 2.6 of Chapter 2. A fairly cheap blood test for the Human Immunodeficiency Virus (HIV) that causes AIDS (Acquired Immune Deficiency Syndrome) has the following characteristics: the false negative rate is 2% and the false positive rate is 0.5%. It is assumed that around .04% of Canadian males are infected with HIV.

Find the probability that if a male tests positive for HIV, he actually has HIV.

Solution: Suppose a male is randomly selected from the population, and define the events

A = {person has HIV}

B = {blood test is positive}

We are asked to find $P(A|B)$. From the information given we know that

$$P(A) = .0004, \quad P(\bar{A}) = .9996$$
$$P(B|A) = .98, \quad P(B|\bar{A}) = .005$$

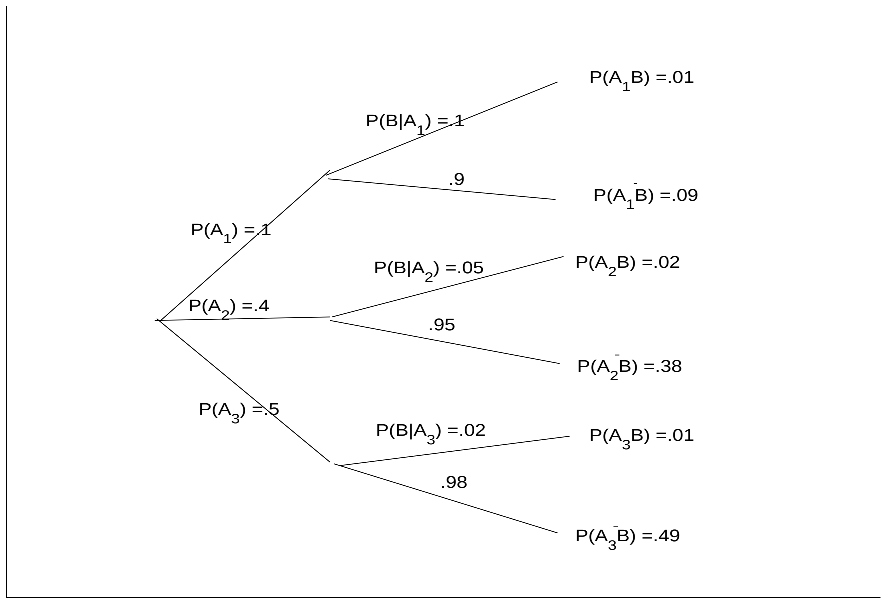


Figure 4.7:

Therefore we can find

$$\begin{array}{rrcll} & P(AB) & = & P(A)P(B|A) & = \; .000392 \\ & P(\bar{A}B) & = & P(\bar{A})P(B|\bar{A}) & = \; .004998 \\ \text{Therefore} & P(B) & = & P(AB)+P(\bar{A}B) & = \; .00539 \end{array}$$

and

$$P(A|B) = \frac{P(AB)}{P(B)} = .0727$$

Thus, if a randomly selected male tests positive, there is still only a small probability (.0727) that they actually have HIV!

Exercise: Try to explain in ordinary words why this is the case.

Note: Bayes Theorem

By using the definition of conditional probability and the multiplication rule, we get that

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

This result is called Bayes Theorem, after a mathematician[3] who proved it in the 1700's. It is a very trivial theorem, but it has inspired approaches to problems in statistics and other areas such as machine

[3](Rev) Thomas Bayes (1702-1761) was an English Nonconformist minister, turned Presbyterian. He may have been

learning, classification and pattern recognition. In these areas the term "Bayesian methods" is often used.

Problems:

4.4.1 If you take a bus to work in the morning there is a 20% chance you'll arrive late. When you go by bicycle there is a 10% chance you'll be late. 70% of the time you go by bike, and 30% by bus. Given that you arrive late, what is the probability you took the bus?

4.4.2 A box contains 4 coins – 3 fair coins and 1 biased coin for which P(heads) = .8. A coin is picked at random and tossed 6 times. It shows 5 heads. Find the probability this coin is fair.

4.4.3 At a police spot check, 10% of cars stopped have defective headlights and a faulty muffler. 15% have defective headlights and a muffler which is satisfactory. If a car which is stopped has defective headlights, what is the probability that the muffler is also faulty?

4.6 Problems on Chapter 4

4.1 If A and B are mutually exclusive events with $P(A) = 0.25$ and $P(B) = 0.4$, find the probability of each of the following events:

$$\bar{A};\;\; \bar{B};\;\; A \cup B;\;\; A \cap B;\;\; \bar{A} \cup \bar{B};\;\; \bar{A} \cap \bar{B};\;\; \overline{A \cap B}.$$

4.2 Three digits are chosen at random with replacement from $0, 1, ..., 9$; find the probability of each of the following events.

	C: "the digits are all nonzero";
A: "all three digits are the same";	D: "the digits all exceed 4";
B: "all three digits are different";	E "digits all have the same parity (all odd or all even)".

Then find the probability of each of the following events, which are combinations of the previous five events:

$$BE;\;\; B \cup D;\;\; B \cup D \cup E;\;\; (A \cup B)D;\;\; A \cup (BD).$$

Show the last two of these events in Venn diagrams.

4.3 Let A and B be events defined on the same sample space, with $P(A) = 0.3$, $P(B) = 0.4$ and $P(A|B) = 0.5$. Given that event B does not occur, what is the probability of event A?

---

tutored by De Moivre. His famous paper introducing this rule was published after his death. "Bayesians" are statisticians who opt for a purely probabilistic view of inference. All unknowns obtain from some distribution and ultimately, the distribution says it all.

4.4 A die is loaded to give the probabilities:

number	1	2	3	4	5	6
probability	.3	.1	.15	.15	.15	.15

The die is thrown 8 times. Find the probability

(a) 1 does not occur

(b) 2 does not occur

(c) neither 1 nor 2 occurs

(d) both 1 and 2 occur.

4.5 Events A and B are independent with $P(A) = .3$ and $P(B) = .2$. Find $P(A \cup B)$.

4.6 Students A, B and C each independently answer a question on a test. The probability of getting the correct answer is .9 for A, .7 for B and .4 for C. If 2 of them get the correct answer, what is the probability C was the one with the wrong answer?

4.7 70% of the customers buying at a certain store pay by credit card. Find the probability

(a) 3 out of 5 customers pay by credit card

(b) the 5th customer is the 3rd one to pay by credit card.

4.8 Let E and F be independent with $E = A \cup B$ and $F = AB$. Prove that either $P(AB) = 0$ or else $P\left(\overline{A}\,\overline{B}\right) = 0$.

4.9 In a large population, people are one of 3 genetic types A, B and C: 30% are type A, 60% type B and 10% type C. The probability a person carries another gene making them susceptible for a disease is .05 for A, .04 for B and .02 for C. If ten unrelated persons are selected, what is the probability at least one is susceptible for the disease?

4.10 Two baseball teams play a best-of-seven series, in which the series ends as soon as one team wins four games. The first two games are to be played on A's field, the next three games on B's field, and the last two on A's field. The probability that A wins a game is 0.7 at home and 0.5 away. Find the probability that:

(a) A wins the series in 4 games; in 5 games;

(b) the series does not go to 6 games.

4.11 A population consists of F females and M males; the population includes f female smokers and m male smokers. An individual is chosen at random from the population. If A is the event that this individual is female and B is the event he or she is a smoker, find necessary and sufficient conditions on f, m, F and M so that A and B are independent events.

4.12 An experiment has three possible outcomes A, B and C with respective probabilities p, q and r, where $p + q + r = 1$. The experiment is repeated until either outcome A or outcome B occurs. Show that A occurs before B with probability $p/(p + q)$.

4.13 In the game of craps, a player rolls two dice. They win at once if the total is 7 or 11, and lose at once if the total is 2, 3, or 12. Otherwise, they continue rolling the dice until they either win by throwing their initial total again, or lose by rolling 7.
Show that the probability they win is 0.493.
(Hint: You can use the result of Problem 4.12)

4.14 A researcher wishes to estimate the proportion p of university students who have cheated on an examination. The researcher prepares a box containing 100 cards, 20 of which contain Question A and 80 Question B.

Question A: Were you born in July or August?

Question B: Have you ever cheated on an examination?

Each student who is interviewed draws a card at random with replacement from the box and answers the question it contains. Since only the student knows which question he or she is answering, confidentiality is assured and so the researcher hopes that the answers will be truthful[4]. It is known that one-sixth of birthdays fall in July or August.

(a) What is the probability that a student answers ‘yes’?

(b) If x of n students answer ‘yes’, estimate p.

(c) What proportion of the students who answer ‘yes’ are responding to Question B?

4.15 **Diagnostic tests.** Recall the discussion of diagnostic tests in Problem 2.6 for Chapter 2. For a randomly selected person let $D =$ ‘person has the disease’ and $R =$ ‘the test result is positive’. Give estimates of the following probabilities: $P(R|D)$, $P(R|\bar{D})$, $P(R)$.

4.16 **Slot machines.** Standard slot machines have three wheels, each marked with some number of symbols at equally spaced positions around the wheel. For this problem suppose there are 10 positions on each wheel, with three different types of symbols being used: flower, dog, and house. The three wheels spin independently and each has probability 0.1 of landing at any

[4]"A foolish faith in authority is the worst enemy of truth" Albert Einsten, 1901.

position. Each of the symbols (flower, dog, house) is used in a total of 10 positions across the three wheels. A payout occurs whenever all three symbols showing are the same.

(a) If wheels 1, 2, 3 have 2, 6, and 2 flowers, respectively, what is the probability all three positions show a flower?

(b) In order to minimize the probability of all three positions showing a flower, what number of flowers should go on wheels 1, 2 and 3? Assume that each wheel must have at least one flower.

4.17 **Spam detection 1**. Many methods of spam detection are based on words or features that appear much more frequently in spam than in regular email. Conditional probability methods are then used to decide whether an email is spam or not. For example, suppose we define the following events associated with a random email message.

Spam	=	"Message is spam"
Not Spam	=	"Message is not spam ("regular")"
A	=	"Message contains the word Viagra"

If we know the values of the probabilities P(Spam), $P(A|$ Spam) and $P(A|$ Not Spam), then we can find the probabilities P(Spam$|A)$ and P(Not Spam$|A)$.

(a) From a study of email messages coming into a certain system it is estimated that P(Spam) = .5, $P(A|$Spam) = .2, and $P(A|$Not Spam) = .001. Find P(Spam$|A)$ and P(Not Spam$|A)$.

(b) If you declared that any email containing the word Viagra was Spam, then find what fraction of regular emails would be incorrectly identified as Spam.

4.18 **Spam detection 2.** The method in part (b) of the preceding question would only filter out 20% of Spam messages. (Why?) To increase the probability of detecting spam, we can use a larger set of email "features"; these could be words or other features of a message which tend to occur with much different probabilities in spam and in regular email. (From your experience, what might be some useful features?) Suppose we identify n binary features, and define events

A_i = feature i appears in a message.

We will assume that $A_1, \ldots, A_n$ are independent events, given that a message is spam, and that they are also independent events, given that a message is regular.

Suppose $n = 3$ and that

$P(A_1\|$Spam) = .2	$P(A_1\|$Not Spam) = .005
$P(A_2\|$Spam) = .1	$P(A_2\|$Not Spam) = .004
$P(A_3\|$Spam) = .1	$P(A_3\|$Not Spam) = .005

Assume as in the preceding question that P(Spam) = .5.

(a) Suppose a message has all of features 1, 2, and 3 present. Determine P(Spam $|A_1A_2A_3)$.

(b) Suppose a message has features 1 and 2 present, but feature 3 is not present. Determine P(Spam $|A_1A_2\bar{A}_3$).

(c) If you declared as spam any message with one or more of features 1, 2 or 3 present, what fraction of spam emails would you detect?

4.19 **Online fraud detection.** Methods like those in problems 4.17 and 4.18 are also used in monitoring events such as credit card transactions for potential fraud. Unlike the case of spam email, however, the fraction of transactions that are fraudulent is usually very small. What we hope to do in this case is to "flag" certain transactions so that they can be checked for potential fraud, and perhaps to block (deny) certain transactions. This is done by identifying features of a transaction so that if F = "transaction is fraudulent", then

$$r = \frac{P(\text{feature present}|F)}{P(\text{feature present}|\bar{F})}$$

is large.

(a) Suppose $P(F)$ =0.0005 and that $P(\text{feature present}|\bar{F}) = .02$. Determine $P(F|$ feature present) as a function of r, and give the values when $r = 10$,30 and 100.

(b) Suppose $r = 100$ and you decide to flag transactions with the feature present. What percentage of transactions would be flagged? Does this seem like a good idea?

5. Review of Useful Series and Sums

The preceding chapters have introduced ways to calculate the probabilities of random events, based on various assumptions. You may have noticed that many of the problems you've encountered are actually similar, despite the contexts being different. Our approach in the next few chapters will be to classify some of these common types of problems and develop general methods for handling them. Rather than working each problem out as if we'd never seen one like it before, our emphasis will now shift to checking whether the problem is one of these general types.

5.1 Series and Sums

Before starting on the probability models of the next chapter, it's worth reviewing some useful results for series, and for summing certain series algebraically. We'll be making use of them in the next few chapters, and many, such as the geometric series, have already been used.

1. **Geometric Series:**

$$a + ar + ar^2 + \cdots + ar^{n-1} = \frac{a\left(1 - r^n\right)}{1 - r} = \frac{a\left(r^n - 1\right)}{r - 1} \text{ for } r \neq 1$$

   If $|r| < 1$, then

$$a + ar + ar^2 + \cdots = \tfrac{a}{1-r}$$

2. **Binomial Theorem:** There are various forms of this theorem. We'll use the form

$$(1+a)^n = \sum_{x=0}^{n} \binom{n}{x} a^x.$$

   In a more general version, when n is not a positive integer,

$$(1+a)^n = \sum_{x=0}^{\infty} \binom{n}{x} a^x \quad \text{if } |a| < 1.$$

   (We have already defined $\binom{n}{x}$ when n is not a positive integer.) While the binomial theorem may not look like a summation result, we'll use it to evaluate series of the form $\sum\limits_{x=0}^{\infty} \binom{n}{x} a^x$.

3. **Multinomial Theorem:** A generalization of the binomial theorem is

$$(a_1 + a_2 + \cdots + a_k)^n = \sum \frac{n!}{x_1! x_2! \cdots x_k!} \; a_1^{x_1} a_2^{x_2} \cdots a_k^{x_k}.$$

with the summation over all $x_1, x_2, \cdots, x_k$ with $\sum x_i = n$. The case $k = 2$ gives the binomial theorem in the form

$$(a_1 + a_2)^n = \sum_{x_1=0}^{n} \binom{n}{x_1} a_1^{x_1} a_2^{n-x_1}$$

.

4. **Hypergeometric Identity:**

$$\sum_{x=0}^{\infty} \binom{a}{x} \binom{b}{n-x} = \binom{a+b}{n}.$$

There will not be an infinite number of terms if a and b are positive integers since the terms become 0 eventually. For example

$$\binom{4}{5} = \frac{4^{(5)}}{5!} = \frac{(4)(3)(2)(1)(0)}{5!} = 0$$

Sketch of Proof:

$$(1+y)^{a+b} = (1+y)^a \; (1+y)^b.$$

Expand each term by the binomial theorem and equate the coefficients of y^n on each side.

5. **Exponential series:**

$$e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

(Recall, also that

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n)$$

6. **Special series involving integers:**

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \left[\frac{n(n+1)}{2}\right]^2$$

Example: Find

$$\sum_{x=0}^{\infty} x\,(x-1)\binom{a}{x}\binom{b}{n-x}$$

Solution: For $x=0$ or 1 the term becomes 0, so we can start summing at $x=2$. For $x \geq 2$, we can expand $x!$ as $x(x-1)(x-2)!$

$$\sum_{x=0}^{\infty} x(x-1)\binom{a}{x}\binom{b}{n-x} = \sum_{x=2}^{\infty} x(x-1)\frac{a!}{x(x-1)(x-2)!(a-x)!}\binom{b}{n-x}.$$

Cancel the $x(x-1)$ terms and try to re-group the factorial terms as "something choose something".

$$\frac{a!}{(x-2)!(a-x)!} = \frac{a(a-1)(a-2)!}{(x-2)!\,[(a-2)-(x-2)]!} = a(a-1)\binom{a-2}{x-2}.$$

Then

$$\sum_{x=0}^{\infty} x(x-1)\binom{a}{x}\binom{b}{n-x} = \sum_{x=2}^{\infty} a(a-1)\binom{a-2}{x-2}\binom{b}{n-x}.$$

Factor out $a(a-1)$ and let $y=x-2$ to get

$$a(a-1)\sum_{y=0}^{\infty}\binom{a-2}{y}\binom{b}{n-(y+2)} = a(a-1)\binom{a+b-2}{n-2}$$

by the hypergeometric identity.

5.2 Problems on Chapter 5

(Solutions to 5.1 and 5.2 are given in Chapter 10.)

5.1 Show that

$$\sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x} = np$$

(use the binomial theorem).

5.2 I have a quarter which turns up heads with probability 0.6, and a fair dime. Both coins are tossed simultaneously and independently until at least one shows heads. Find the probability that both the dime and the quarter show heads at the same time.

5.3 Some other summation formulas can be obtained by differentiating the above equations on both sides. Show that $a + 2ar + 3ar^2 + \cdots = \frac{a}{(1-r)^2}$ by starting with the geometric series formula. Assume $|r| < 1$.

5.4 Players A and B decide to play chess until one of them wins. Assume games are independent with $P(A$ wins) = .3, $P(B$ wins) = .25 and P(draw) = .45 on each game. If the game ends in a draw another game will be played. Find the probability A wins before B.

6. Discrete Random Variables and Probability Models

6.1 Random Variables and Probability Functions

Probability models are used to describe outcomes associated with random processes. So far we have used sets $A, B, C, \ldots$ in sample spaces to describe such outcomes. In this chapter we introduce numerical-valued variables $X, Y, \ldots$ to describe outcomes. This allows probability models to be manipulated easily using ideas from algebra, calculus, or geometry.

A random variable (r.v.) is a numerical-valued variable that represents outcomes in an experiment or random process. For example, suppose a coin is tossed 3 times; then

$$X = \text{Number of Heads that occur}$$

would be a random variable. Associated with any random variable is a **range** or domain A, which is the set of possible values for the variable. For example, the random variable X defined above has range $A = \{0, 1, 2, 3\}$.

Random variables are denoted by capital letters like $X, Y, \ldots$ and their possible values are denoted by $x, y, \ldots$. This gives a nice short-hand notation for outcomes: for example, "$X = 2$" in the experiment above stands for "2 heads occurred".

Random variables are always defined so that if the associated random process or experiment is carried out, then one and only one of the outcomes "$X = x$"$(x \in A)$ occurs. In other words, the possible values for X form a partition of the points in the sample space for the experiment. In more advanced mathematical treatments of probability, a random variable is defined as a function on a sample space, as follows:

Definition 9 *A **random variable** is a function that assigns a real number to each point in a sample space S.*

To understand this definition, consider the experiment in which a coin is tossed 3 times, and suppose that we used the sample space

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

Then each of the outcomes "$X = x$" (where X = number of heads) represents an event (either simple or compound) and so a real number x can be associated with each point in S. In particular, the point HHH corresponds to $x = 3$; the points THH, HTH, HHT correspond to $x = 2$; the points HTT, THT, TTH correspond to $x = 1$; and the point TTT corresponds to $x = 0$.

As you may recall, a function is a mapping of each point in a domain into a unique point. e.g. The function $y = x^3$ maps the point $x = 2$ in the domain into the point $y = 8$ in the range. We are familiar with this rule for mapping being defined by a mathematical formula. However, the rule for mapping a point in the sample space (domain) into the real number in the range of a random variable is most often given in words rather than by a formula. As mentioned above, we generally denote random variables, in the abstract, by capital letters $(X, Y$, etc.) and denote the actual numbers taken by random variables by small letters $(x, y$, etc.). You may, in your earlier studies, have had this distinction made between a <u>function</u> (X) and the <u>value</u> of a function (x).

Since "$X = x$" represents an outcome of some kind, we will be interested in its probability, which we write as $P(X = x)$. To discuss probabilities for random variables, it is easiest if they are classified into two types, according to their ranges:

Discrete random variables take integer values or, more generally, values in a countable set (recall that a set is countable if its elements can be placed in a one-one correspondence with a subset of the positive integers).

Continuous random variables take values in some interval of real numbers.

Examples might be:

<u>Discrete</u>	<u>Continuous</u>
number of people in a car	total weight of people in a car
number of cars in a parking lot	distance between cars in a parking lot
number of phone calls to 911	time between calls to 911.

In theory there could also be mixed r.v.'s which are discrete-valued over part of their range and continuous-valued over some other portion of their range. We will ignore this possibility here and concentrate first on discrete r.v.'s. Continuous r.v.'s are considered in Chapter 9.

Our aim is to set up general models which describe how the probability is distributed among the possible values a random variable can take. To do this we define for any discrete random variable X the probability function.

Definition 10 *The* ***probability function*** *(p.f.) of a random variable X is the function*

$$f(x) = P(X = x), \text{ defined for all } x \in A.$$

The set of pairs $\{(x, f(x)) : x \in A\}$ is called the **probability distribution** of X. All probability functions must have two properties:

1. $f(x) \geq 0$ for all values of x (i.e. for $x \in A$)
2. $\sum\limits_{\text{all } x \in A} f(x) = 1$

By implication, these properties ensure that $f(x) \leq 1$ for all x. We consider a few "toy" examples before dealing with more complicated problems.

Example 1: Let X be the number obtained when a die is thrown. We would normally use the probability function $f(x) = 1/6$ for $x = 1, 2, 3, \cdots, 6$. In fact there probably is no absolutely perfect die in existence. For most dice, however, the 6 sides will be close enough to being equally likely that $f(x) = 1/6$ is a satisfactory model for the distribution of probability among the possible outcomes.

Example 2: Suppose a "fair" coin is tossed 3 times, and let X be the number of heads occurring. Then a simple calculation using the ideas from earlier chapters gives

$$f(x) = P(X = x) = \frac{\binom{3}{x}}{8} \quad \text{for } x = 0, 1, 2, 3.$$

Note that instead of writing $f(0) = \frac{1}{8}, f(1) = \frac{3}{8}, f(2) = \frac{3}{8}, f(3) = \frac{1}{8}$, we have given a simple algebraic expression.

Example 3: Find the value of k which makes $f(x)$ below a probability function.

x	0	1	2	3
$f(x)$	k	$2k$	0.3	$4k$

Using $\sum\limits_{x=0}^{3} f(x) = 1$ gives $7k + 0.3 = 1$. Hence $k = 0.1$.

While the probability function is the most common way of describing a probability model, there are other possibilities. One of them is by using the **cumulative distribution function** (c.d.f.).

Definition 11 *The cumulative distribution function of X is the function usually denoted by $F(x)$*

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

defined for all real numbers x.

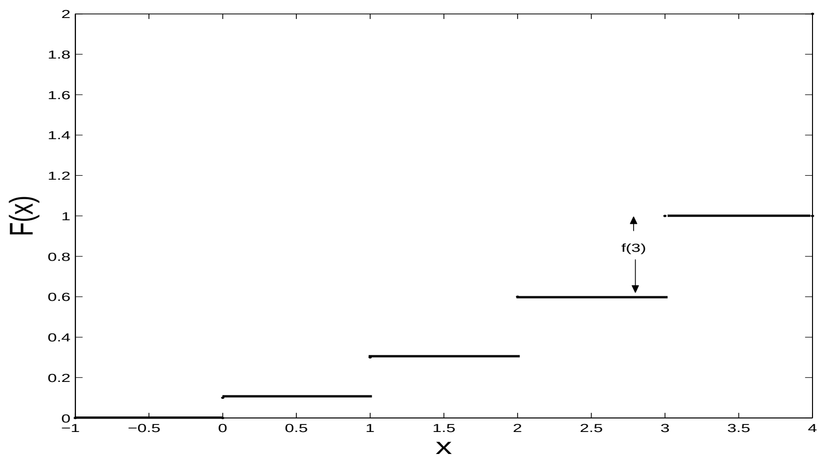


Figure 6.1: A simple cumulative distribution function

In the last example, with $k = 0.1$, we have for $x \in A$

x	0	1	2	3
$F(x)$	0.1	0.3	0.6	1

since, for instance,

$$F(2) = P(X \leq 2) = f(0) + f(1) + f(2) = 0.6.$$

Similarly, $F(x)$ can be defined for real numbers x not in the range of the random variable, for example

$$F(2.5) = F(2) = 0.6 \text{ and } F(3.8) = 1.$$

The c.d.f. for this example is plotted in Figure 6.1.

In general, $F(x)$ can be obtained from $f(x)$ by the fact that

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u).$$

A c.d.f. $F(x)$ has certain properties, just as a probability function $f(x)$ does. Obviously, since it represents a probability, $F(x)$ must be between 0 and 1. In addition it must be a non-decreasing function (e.g. $P(X \leq 8)$ cannot be less than $P(X \leq 7)$). Thus we note the following **properties of a c.d.f.** $F(x)$**:**

1. $F(x)$ is a non-decreasing function of x.

2. $0 \leq F(x) \leq 1$ for all x.

3. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

We have noted above that $F(x)$ can be obtained from $f(x)$. The opposite is also true; for example the following result holds:

If X takes on integer values then for values x such that $x \in A$ and $x-1 \in A$,

$$f(x) = F(x) - F(x-1)$$

This says that $f(x)$ is the size of the jump in $F(x)$ at the point x.
To prove this, just note that

$$F(x) - F(x-1) = P(X \le x) - P(X \le x-1) = P(X = x).$$

When a random variable has been defined it is sometimes simpler to find its probability function (p.f.) $f(x)$ first, and sometimes it is simpler to find $F(x)$ first. The following example gives two approaches for the same problem.

Example: Suppose that N balls labelled $1, 2, \ldots, N$ are placed in a box, and n balls $(n \le N)$ are randomly selected without replacement. Define the r.v.

$$X = \text{largest number selected}$$

Find the probability function for X.

Solution 1: If $X = x$ then we must select the number x plus $n-1$ numbers from the set $\{1, 2, \ldots, x-1\}$. (Note that this means we need $x \ge n$.) This gives

$$f(x) = P(X = x) = \frac{\binom{1}{1}\binom{x-1}{n-1}}{\binom{N}{n}} = \frac{\binom{x-1}{n-1}}{\binom{N}{n}} \quad x = n, n+1, \ldots, N$$

Solution 2: First find $F(x) = P(X \le x)$. Noting that $X \le x$ if and only if all n balls selected are from the set $\{1, 2, \ldots, x\}$, we get

$$F(x) = \frac{\binom{x}{n}}{\binom{N}{n}} \quad x = n, n+1, \ldots N$$

We can now find

$$\begin{aligned} f(x) &= F(x) - F(x-1) \\ &= \frac{\binom{x}{n} - \binom{x-1}{n}}{\binom{N}{n}} \\ &= \frac{\binom{x-1}{n-1}}{\binom{N}{n}} \end{aligned}$$

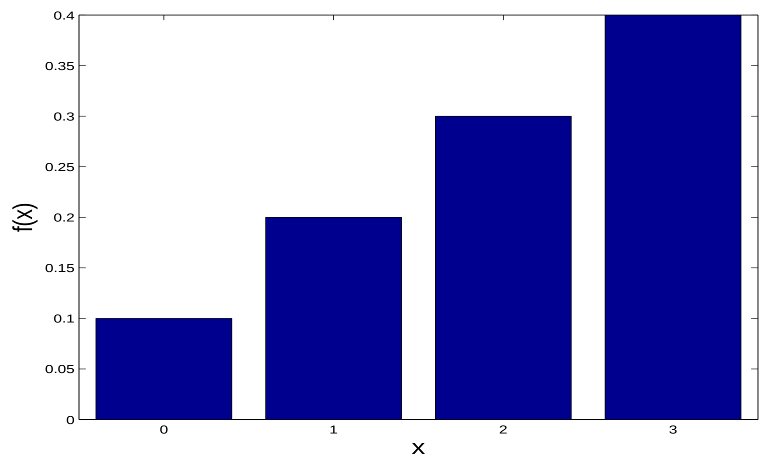


Figure 6.2: Probability histogram for $f(x) = \frac{x+1}{10}, x = 0, 1, 2, 3$

as before.

Remark: When you write down a probability function or a cumulative distribution function, don't forget to give the range of the function (i.e. the possible values of the random variable). This is part of the function's definition.

Sometimes we want to graph a probability function $f(x)$. The type of graph we will use most is called a (probability) **histogram**. For now, we'll define this only for r.v.'s whose range is some set of consecutive integers $\{0, 1, 2, \dots\}$. A histogram of $f(x)$ is then a graph consisting of adjacent bars or rectangles. At each x we place a rectangle with base on $(x - .5, x + .5)$ and with height $f(x)$. In the above Example 3, a histogram of $f(x)$ looks like that in Figure 6.2.

Notice that the areas of these rectangles correspond to the probabilities, so for example $P(1 \leq X \leq 3)$ is the sum of the area of the three rectangles above the points $1, 2,$ and 3. In general, probabilities are depicted by areas.

Model Distributions:

Many processes or problems have the same structure. In the remainder of this course we will identify common types of problems and develop probability distributions that represent them. In doing this it is important to be able to strip away the particular wording of a problem and look for its essential features. For example, the following three problems are all essentially the same.

(a) A fair coin is tossed 10 times and the "number of heads obtained" (X) is recorded.

(b) Twenty seeds are planted in separate pots and the "number of seeds germinating" (X) is recorded.

(c) Twelve items are picked at random from a factory's production line and examined for defects. The number of items having no defects (X) is recorded.

What are the common features? In each case the process consists of "trials" which are repeated a stated number of times - 10, 20, and 12. In each repetition there are two types of outcomes - heads/tails, germinate/don't germinate, and no defects/defects. These repetitions are independent (as far as we can determine), with the probability of each type of outcome remaining constant for each repetition. The random variable we record is the number of times one of these two types of outcome occurred.
Six model distributions for discrete r.v.'s will be developed in the rest of this chapter. Students often have trouble deciding which one (if any) to use in a given setting, so be sure you understand the physical setup which leads to each one. Also, as illustrated above you will need to learn to focus on the essential features of the situation as well as the particular content of the problem.

Statistical Computing
A number of major software systems have been developed for probability and statistics. We will use a system called R, which has a wide variety of features and which has Unix and Windows versions. Appendix 6.1 at the end of this chapter gives a brief introduction to R, and how to access it. For this course, R can compute probabilities for all the distributions we consider, can graph functions or data, and can simulate random processes. In the sections below we will indicate how R can be used for some of these tasks.

Problems:

6.1.1 Let X have probability function $\begin{array}{c|ccc} x & 0 & 1 & 2 \\ \hline f(x) & 9c^2 & 9c & c^2 \end{array}$. Find c.

6.1.2 Suppose that 5 people, including you and a friend, line up at random. Let X be the number of people standing between you and your friend. Tabulate the probability function and the cumulative distribution function for X.

6.2 Discrete Uniform Distribution

We define each model in terms of an abstract "physical setup", or setting, and then consider specific examples of the setup.
Physical Setup: Suppose X takes values $a, a+1, a+2, \cdots, b$ with all values being equally likely. Then X has a discrete uniform distribution, on $[a, b]$.

Illustrations:

1. If X is the number obtained when a die is rolled, then X has a discrete uniform distribution with $a=1$ and $b=6$.

2. Computer random number generators give uniform $[1,N]$ variables, for a specified positive integer N. These are used for many purposes, e.g. generating lottery numbers or providing automated random sampling from a set of N items.

Probability Function: There are $b-a+1$ values X can take so the probability at each of these values must be $\frac{1}{b-a+1}$ in order that $\sum_{x=a}^{b} f(x)=1$. Therefore

$$f(x)=\begin{cases} \frac{1}{b-a+1}; & x=a,a+1,\cdots,b \\ 0; & \text{otherwise} \end{cases}$$

Problem 6.2.1
Let X be the largest number when a die is rolled 3 times. First find the c.d.f., $F(x)$, and then find the probability function, $f(x)$.

6.3 Hypergeometric Distribution

Physical Setup: We have a collection of N objects which can be classified into two distinct types. Call one type "success"[5] (S) and the other type "failure" (F). There are r successes and $N-r$ failures. Pick n objects at random **without replacement**. Let X be the number of successes obtained. Then X has a hypergeometric distribution.

Illustrations:

1. The number of aces X in a bridge hand has a hypergeometric distribution with $N=52,\ r=4$, and $n=13$.

2. In a fleet of 200 trucks there are 12 which have defective brakes. In a safety check 10 trucks are picked at random for inspection. The number of trucks X with defective brakes chosen for inspection has a hypergeometric distribution with $N=200, r=12, n=10$.

[5]"If A is a success in life, then A equals x plus y plus z. Work is x; y is play; and z is keeping your mouth shut." Albert Einstein, 1950

Probability Function: Using counting techniques we note there are $\binom{N}{n}$ points in the sample space S if we don't consider order of selection. There are $\binom{r}{x}$ ways to choose the x success objects from the r available and $\binom{N-r}{n-x}$ ways to choose the remaining $(n-x)$ objects from the $(N-r)$ failures. Hence

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

The range of values for x is somewhat complicated. Of course, $x \geq 0$. However if the number, n, picked exceeds the number, $N-r$, of failures, the difference, $n-(N-r)$ must be successes. So $x \geq \max(0, n-N+r)$. Also, $x \leq r$ since we can't get more successes than the number available. But $x \leq n$, since we can't get more successes than the number of objects chosen. Therefore $x \leq \min(r, n)$.

Example: In Lotto 6/49 a player selects a set of six numbers (with no repeats) from the set $\{1, 2, \ldots, 49\}$. In the lottery draw six numbers are selected at random. Find the probability function for X, the number from your set which are drawn.

Solution: Think of your numbers as the S objects and the remainder as the F objects. Then X has a hypergeometric distribution with $N = 49, r = 6$ and $n = 6$, so

$$P(X = x) = f(x) = \frac{\binom{6}{x}\binom{43}{6-x}}{\binom{49}{6}}, \text{ for } x = 0, 1, \ldots, 6$$

For example, you win the jackpot prize if $X = 6$; the probability of this is $\binom{6}{6}/\binom{49}{6}$, or about 1 in 13.9 million.

Remark: Hypergeometric probabilities are tedious to compute using a calculator. The R functions $dhyper$ and $phyper$ can be used to evaluate $f(x)$ and the c.d.f $F(x)$. In particular, $dhyper(x, r, N-r, n)$ gives $f(x)$ and $phyper(x, r, N-r, n)$ gives $F(x)$. Using this we find for the Lotto 6/49 problem here, for example, that $f(6)$ is calculated by typing $dhyper(6, 6, 43, 6)$ in R, which returns the answer 7.151124×10^{-8} or $1/13, 983, 186$.

For all of our model distributions we can also confirm that $\sum_{\text{all } x} f(x) = 1$. To do this here we use a summation result from Chapter 5 called the hypergeometric identity. Letting $a = r, b = N-r$ in that identity we get

$$\sum_{\text{all } x} f(x) = \sum \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum \binom{r}{x}\binom{N-r}{n-x} = \frac{\binom{r+N-r}{n}}{\binom{N}{n}} = 1$$

Problems:

6.3.1 A box of 12 tins of tuna contains d which are tainted. Suppose 7 tins are opened for inspection and none of these 7 is tainted.

a) Calculate the probability that none of the 7 is tainted for $d = 0, 1, 2, 3$.

b) Do you think it is likely that the box contains as many as 3 tainted tins?

6.3.2 Derive a formula for the hypergeometric probability function using a sample space in which order of selection is considered.

6.4 Binomial Distribution

Physical Setup:

Suppose an “experiment" has two types of distinct outcomes. Call these types “success” (S) and “failure” (F), and let their probabilities be p (for S) and $1 - p$ (for F). Repeat the experiment n **independent** times. Let X be the number of successes obtained. Then X has what is called a **binomial distribution**. (We write $X \sim Bi(n, p)$ as a shorthand for “X is distributed according to a binomial distribution with n repetitions and probability p of success”.) The individual experiments in the process just described are often called “trials”, and the process is called a Bernoulli[6] process or a binomial process.

Illustrations:

1. Toss a fair die 10 times and let X be the number of sixes that occur. Then $X \sim Bi(10, 1/6)$.

2. In a microcircuit manufacturing process, 60% of the chips produced work (40% are defective). Suppose we select 25 independent chips and let X be the number that work. Then $X \sim Bi(25, .6)$.

Comment: We must think carefully whether the physical process we are considering is closely approximated by a binomial process, for which the key assumptions are that (i) the probability p of success is constant over the n trials, and (ii) the outcome (S or F) on any trial is independent of the outcome on the other trials. For Illustration 1 these assumptions seem appropriate. For Illustration 2

[6]After James (Jakob) Bernoulli (1654 – 1705), a Swiss member of a family of eight mathematicians, who started his professional life in the ministry, and is responsible Bernoulli’s Theorem as well as many of the combinatorial results in these notes.

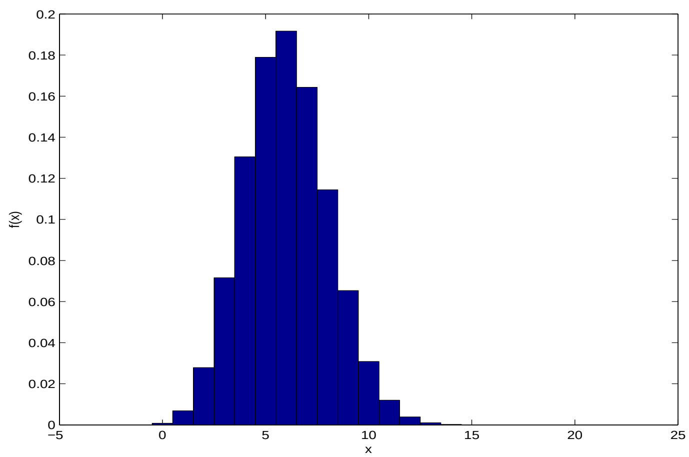


Figure 6.3: The Binomial$(20, 0.3)$ probability histogram.

we would need to think about the manufacturing process. Microcircuit chips are produced on "wafers" containing a large number of chips and it is common for defective chips to cluster on wafers. This could mean that if we selected 25 chips from the same wafer, or from only 2 or 3 wafers, that the "trials" (chips) might not be independent.

Probability Function: There are $\frac{n!}{x!(n-x)!} = \binom{n}{x}$ different arrangements of x S's and $(n-x)$ F's over the n trials. The probability for each of these arrangements has p multiplied together x times and $(1-p)$ multiplied $(n-x)$ times, in some order, since the trials are independent. So each arrangement has probability $p^x(1-p)^{n-x}$.

$$\text{Therefore } f(x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, 2, \cdots, n.$$

Checking that $\sum f(x) = 1$:

$$\begin{aligned} \sum_{x=0}^{n} f(x) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (1-p)^n \sum_{x=0}^{n} \binom{n}{x} \left(\frac{p}{1-p}\right)^x \\ &= (1-p)^n \left(1 + \frac{p}{1-p}\right)^n \text{ by the binomial theorem} \\ &= (1-p)^n \left(\frac{1-p+p}{1-p}\right)^n = 1^n = 1. \end{aligned}$$

We graph in Figure 6.3 the probability function for the Binomial distribution with parameters $n = 20$ and $p = 0.3$. Although the formula for $f(x)$ may seem complicated this shape is typically, increasing to a maximum value near np and then decreasing thereafter.

Computation: Many software packages and some calculators give binomial probabilities. In R we use the function $dbinom(x, n, p)$ to compute $f(x)$ and $pbinom(x, n, p)$ to compute the corresponding c.d.f. $F(x) = P(X \leq x)$.

Example Suppose that in a weekly lottery you have probability .02 of winning a prize with a single ticket. If you buy 1 ticket per week for 52 weeks, what is the probability that (a) you win no prizes, and (b) that you win 3 or more prizes?

Solution: Let X be the number of weeks that you win; then $X \sim Bi(52, .02)$. We find

(a) $P(X = 0) = f(0) = \binom{52}{0}(.02)^0(.98)^{52} = .350$

$$\begin{aligned}
\text{(b)} \quad P(X \geq 3) &= 1 - P(X \leq 2) \\
&= 1 - f(0) - f(1) - f(2) \\
&= 0.0859
\end{aligned}$$

(Note that $P(X \leq 2)$ is given by the R command $pbinom(2, 52, .02)$.)

Comparison of Binomial and Hypergeometric Distributions:
These distributions are similar in that an experiment with 2 types of outcome (S and F) is repeated n times and X is the number of successes. The key difference is that the binomial requires independent repetitions with the same probability of S, whereas the draws in the hypergeometric are made from a fixed collection of objects **without** replacement. The trials (draws) are therefore not independent. For example, if there are $n = 10$ S objects and $N - r = 10$ F objects, then the probability of getting an S on draw 2 depends on what was obtained in draw 1. If these draws had been made **with** replacement, however, they would be independent and we'd use the binomial rather than the hypergeometric model. If N is large and the number, n, being drawn is relatively small in the hypergeometric setup then we are unlikely to get the same object more than once even if we do replace it. So it makes little practical difference whether we draw with or without replacement. This suggests that when we are drawing a fairly small proportion of a large collection of objects the binomial and the hypergeometric models should produce similar probabilities. As the binomial is easier to calculate, it is often used as an approximation to the hypergeometric in such cases.

Example: Suppose we have 15 cans of soup with no labels, but 6 are tomato and 9 are pea soup. We randomly pick 8 cans and open them. Find the probability 3 are tomato.

Solution: The correct solution uses hypergeometric, and is (with X = number of tomato soups picked)

$$f(3) = P(X = 3) = \frac{\binom{6}{3}\binom{9}{5}}{\binom{15}{8}} = 0.396.$$

If we incorrectly used binomial, we'd get

$$f(3) = \binom{8}{3}\left(\frac{6}{15}\right)^3\left(\frac{9}{15}\right)^5 = 0.279$$

As expected, this is a poor approximation since we're picking over half of a fairly small collection of cans.
However, if we had 1500 cans - 600 tomato and 900 pea, we're not likely to get the same can again even if we did replace each of the 8 cans after opening it. (Put another way, the probability we get a tomato soup on each pick is very close to .4, regardless of what the other picks give.) The exact, hypergeometric, probability is now $\frac{\binom{600}{3}\binom{900}{5}}{\binom{1500}{8}} = .2794$. Here the binomial probability,

$$\binom{8}{3}\left(\frac{600}{1500}\right)^3\left(\frac{900}{1500}\right)^5 = 0.279$$

is a very good approximation.
Problems:

6.4.1 Megan audits 130 clients during a year and finds irregularities for 26 of them.

a) Give an expression for the probability that 2 clients will have irregularities when 6 of her clients are picked at random,

b) Evaluate your answer to (a) using a suitable approximation.

6.4.2 The flash mechanism on camera A fails on 10% of shots, while that of camera B fails on 5% of shots. The two cameras being identical in appearance, a photographer selects one at random and takes 10 indoor shots using the flash.

(a) Give the probability that the flash mechanism fails exactly twice. What assumption(s) are you making?

(b) Given that the flash mechanism failed exactly twice, what is the probability camera A was selected?

6.5 Negative Binomial Distribution

Physical Setup:
The setup for this distribution is almost the same as for binomial; i.e. an experiment (trial) has two distinct types of outcome (S and F) and is repeated independently with the same probability, p, of success each time. Continue doing the experiment until a specified number, k, of success have been obtained. Let X be the number of failures obtained before the k^{th} success. Then X has a negative binomial distribution. We often write $X \sim NB(k,p)$ to denote this.

Illustrations:

(1) If a fair coin is tossed until we get our 5^{th} head, the number of tails we obtain has a negative binomial distribution with $k = 5$ and $p = \frac{1}{2}$.

(2) As a rough approximation, the number of half credit failures a student collects before successfully completing 40 half credits for an honours degree has a negative binomial distribution. (Assume all course attempts are independent, with the same probability of being successful, and ignore the fact that getting more than 6 half credit failures prevents a student from continuing toward an honours degree.)

Probability Function: In all there will be $x + k$ trials (x F's and k S's) and the last trial must be a success. In the first $x + k - 1$ trials we therefore need x failures and $(k - 1)$ successes, in any order. There are $\frac{(x+k-1)!}{x!(k-1)!} = \binom{x+k-1}{x}$ different orders. Each order will have probability $p^k(1-p)^x$ since there must be x trials which are failures and k which are success. Hence

$$f(x) = \binom{x+k-1}{x} p^k (1-p)^x; \quad x = 0, 1, 2, \cdots$$

Note: An alternate version of the negative binomial distribution defines X to be the total number of trials needed to get the k^{th} success. This is equivalent to our version. For example, asking for the probability of getting 3 tails before the 5^{th} head is exactly the same as asking for a total of 8 tosses in order to get the 5^{th} head. You need to be careful to read how X is defined in a problem rather than mechanically "plugging in" numbers in the above formula for $f(x)$.

Checking that $\sum f(x) = 1$ requires somewhat more work for the negative binomial distribution. We first re-arrange the $\binom{x+k-1}{x}$ term,

$$\binom{x+k-1}{x} = \frac{(x+k-1)^{(x)}}{x!} = \frac{(x+k-1)(x+k-2)\cdots(k+1)(k)}{x!}$$

Factor a (-1) out of each of the x terms in the numerator, and re-write these terms in reverse order,

$$\begin{aligned}\binom{x+k-1}{x} &= (-1)^x \frac{(-k)(-k-1)\cdots(-k-x+2)(-k-x+1)}{x!} \\ &= (-1)^x \frac{(-k)^{(x)}}{x!} = (-1)^x \binom{-k}{x}\end{aligned}$$

Then (using the binomial theorem)

$$\begin{aligned}\sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} \binom{-k}{x} (-1)^x p^k (1-p)^x \\ &= p^k \sum_{x=0}^{\infty} \binom{-k}{x} [(-1)(1-p)]^x = p^k [1 + (-1)(1-p)]^{-k} \\ &= p^k p^{-k} = 1\end{aligned}$$

Comparison of Binomial and Negative Binomial Distributions

These should be easily distinguished because they reverse what is specified or known in advance and what is variable.

Binomial:	Know the number, n, of repetitions in advance. Don't know the number of successes we'll obtain until after the experiment.
Negative Binomial:	Know the number, k, of successes in advance. Don't know the number of repetitions needed until after the experiment.

Example:
The fraction of a large population that has a specific blood type T is .08 (8%). For blood donation purposes it is necessary to find 4 people with type T blood. If randomly selected individuals from the population are tested one after another, then (a) What is the probability y persons have to be tested to get 5 type T persons, and (b) What is the probability that over 80 people have to be tested?

Solution:
Think of a type T person as a success (S) and a non-type T as an F. Let Y = number of persons who have to be tested and let X = number of non-type T persons in order to get 5 S's. Then $X \sim NB(k = 5, p = .08)$ and

$$P(X = x) = f(x) = \binom{x+4}{x}(.08)^5(.92)^x \quad x = 0, 1, 2, \ldots$$

We are actually asked here about $Y = X + 5$. Thus

$$\begin{aligned} P(Y = y) &= P(X = y - 5) \\ &= f(y-5) = \tbinom{y-1}{y-5}(.08)^5(.92)^{y-5} \quad y = 5, 6, 7, \ldots \end{aligned}$$

Thus we have the answer to (a) as given above, and
(b)

$$P(Y > 80) = P(X > 75) = 1 - P(X \le 75)$$
$$= 1 - \sum_{x=0}^{75} f(x) = .2235$$

Note: Calculating such probabilities is easy with R. To get $f(x)$ we use $dnbinom(x, k, p)$ and to get $F(x) = P(X \le x)$ we use $pnbinom(x, k, p)$.

Problems:

6.5.1 You can get a group rate on tickets to a play if you can find 25 people to go. Assume each person you ask responds independently and has a 20% chance of agreeing to buy a ticket. Let X be the total number of people you have to ask in order to find 25 who agree to buy a ticket. Find the probability function of X.

6.5.2 A shipment of 2500 car headlights contains 200 which are defective. You choose from this shipment without replacement until you have 18 which are not defective. Let X be the number of defective headlights you obtain.

(a) Give the probability function, $f(x)$.

(b) Using a suitable approximation, find $f(2)$.

6.6 Geometric Distribution

Physical Setup:
This is a special case of the negative binomial distribution with $k = 1$, i.e., an experiment is repeated independently with two types of outcome (S and F) each time, and the same probability, p, of success each time. Let X be the number of failures obtained before the first success.

Illustrations:

(1) The probability you win a lottery prize in any given week is a constant p. The number of weeks **before** you win a prize for the first time has a geometric distribution.

(2) If you take STAT 230 until you pass it and attempts are independent with the same probability of a pass each time, then the number of failures would have a geometric distribution. (These assumptions are unlikely to be true for most persons! Why is this?)

Probability Function:
There is only the one arrangement with x failures followed by 1 success. This arrangement has probability

$$f(x) = (1-p)^x p; \quad x = 0, 1, 2, \cdots$$

which is the same as $f(x)$ for $NB(k = 1, p)$.
Checking that $\sum f(x) = 1$, we will be evaluating a geometric series,

$$\begin{aligned}\sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} (1-p)^x p = p + (1-p)p + (1-p)^2 p + \cdots \\ &= \tfrac{p}{1-(1-p)} = \tfrac{p}{p} = 1\end{aligned}$$

Note: The names of the models so far derive from the summation results which show $f(x)$ sums to 1. The geometric distribution involved a geometric series; the hypergeometric distribution used the hypergeometric identity; both the binomial and negative binomial distributions used the binomial theorem.

Bernoulli Trials

The binomial, negative binomial and geometric models involve trials (experiments) which:

(1) are independent

(2) have 2 distinct types of outcome (S and F)

(3) have the same probability of "success" (S) each time.

Such trials are known as Bernoulli trials; they are named after an 18th century mathematician.

Problem 6.6.1

Suppose there is a 30% chance of a car from a certain production line having a leaky windshield. The probability an inspector will have to check at least n cars to find the first one with a leaky windshield is .05. Find n.

6.7 Poisson Distribution from Binomial

The **Poisson**[7] **distribution** has probability function (p.f.) of the form

$$f(x) = e^{-\mu}\frac{\mu^x}{x!} \quad x = 0, 1, 2, \ldots$$

where $\mu > 0$ is a parameter whose value depends on the setting for the model. Mathematically, we can see that $f(x)$ has the properties of a p.f., since $f(x) \geq 0$ for $x = 0, 1, 2, \ldots$ and since

$$\begin{aligned}\sum_{x=0}^{\infty} f(x) &= e^{-\mu}\sum_{x=0}^{\infty}\frac{\mu^x}{x!} \\ &= e^{-\mu}(e^{\mu}) = 1\end{aligned}$$

The Poisson distribution arises in physical settings where the random variable X represents the number of events of some type. In this section we show how it arises from a binomial process, and in the following section we consider another derivation of the model.

We will sometimes write $X \sim$ Poisson (μ) to denote that X has the p.f. above.

[7] After Siméon Denis Poisson (1781-1840), a French mathematician who was supposed to become a surgeon but, fortunately for his patients, failed medical school for lack of coordination. He was forced to do theoretical research, being too clumsy for anything in the lab. He wrote a major work on probability and the law, *Recherchés sur la probabilité des jugements en matière criminelle et matière civile (1837),* discovered the Poisson distribution (called law of large numbers) and to him is ascribed one of the more depressing quotes in our discipline "Life is good for only two things: to study mathematics and to teach it"

Physical Setup: One way the Poisson distribution arises is as a limiting case of the binomial distribution as $n \to \infty$ and $p \to 0$. In particular, we keep the product np fixed at some constant value, μ, while letting $n \to \infty$. This automatically makes $p \to 0$. Let us see what the limit of the binomial p.f. $f(x)$ is in this case.

Probability Function: Since $np = \mu$, Therefore $p = \frac{\mu}{n}$ and

$$\begin{aligned}
f(x) &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n^{(x)}}{x!}\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x} \\
&= \frac{\mu^x}{x!}\frac{\overbrace{n(n-1)(n-2)\cdots(n-x+1)}^{x \text{ terms}}}{(n)(n)\quad(n)\quad\cdots\quad(n)}\left(1-\frac{\mu}{n}\right)^{n-x} \\
&= \frac{\mu^x}{x!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-x+1}{n}\right)\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x} \\
&= \frac{\mu^x}{x!}(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x} \\
\lim_{n\to\infty} f(x) &= \frac{\mu^x}{x!}\underbrace{(1)(1)(1)\cdots(1)}_{x \text{ terms}} e^{-\mu}(1)^{-x} \quad \left(\text{since } e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n\right) \\
&= \tfrac{\mu^x e^{-\mu}}{x!}; \quad x = 0, 1, 2, \cdots
\end{aligned}$$

(For the binomial the upper limit on x is n, but we are letting $n \to \infty$.) This result allows us to use the Poisson distribution with $\mu = np$ as a close approximation to the binomial distribution $Bi(n, p)$ in processes for which n is large and p is small.

Example: 200 people are at a party. What is the probability that 2 of them were born on Jan. 1?

Solution: Assuming all days of the year are equally likely for a birthday (and ignoring February 29) and that the birthdays are independent (e.g. no twins!) we can use the binomial distribution with $n = 200$ and $p = 1/365$ for X = number born on January 1, giving

$$f(2) = \binom{200}{2}\left(\frac{1}{365}\right)^2\left(1-\frac{1}{365}\right)^{198} = .086767$$

Since n is large and p is close to 0, we can use the Poisson distribution to approximate this binomial probability, with $\mu = np = \frac{200}{365}$, giving

$$f(2) = \frac{\left(\frac{200}{365}\right)^2 e^{-\left(\frac{200}{365}\right)}}{2!} = .086791$$

As might be expected, this is a very good approximation.

Notes:

(1) If p is close to 1 we can also use the Poisson distribution to approximate the binomial. By interchanging the labels "success" and "failure", we can get the probability of "success" (formerly labelled "failure") close to 0.

(2) The Poisson distribution used to be very useful for approximating binomial probabilities with n large and p near 0 since the calculations are easier. (This assumes values of e^x to be available.) With the advent of computers, it is just as easy to calculate the exact binomial probabilities as the Poisson probabilities. However, the Poisson approximation is useful when employing a calculator without a built in binomial function.

(3) The R functions $dpois(x, \mu)$ and $ppois(x, \mu)$ give $f(x)$ and $F(x)$.

Problem 6.7.1
An airline knows that 97% of the passengers who buy tickets for a certain flight will show up on time. The plane has 120 seats.

a) They sell 122 tickets. Find the probability that more people will show up than can be carried on the flight. Compare this answer with the answer given by the Poisson approximation.

b) What assumptions does your answer depend on? How well would you expect these assumptions to be met?

6.8 Poisson Distribution from Poisson Process

We now derive the Poisson distribution as a model for the number of events of some type (e.g. births, insurance claims, web site hits) that occur in time or in space. To this end, we use the "order" notation $g(\Delta t) = o(\Delta t)$ as $\Delta t \to 0$ to mean that the function g approaches 0 faster than Δt as Δt approaches zero, or that

$$\frac{g(\Delta t)}{\Delta t} \to 0 \text{ as } \Delta t \to 0.$$

For example $g(\Delta t) = (\Delta t)^2 = o(\Delta t)$ but $(\Delta t)^{1/2}$ is not $o(\Delta t)$.

Physical Setup:
Consider a situation in which events are occurring randomly over time (or space) according to the following conditions:

1. Independence: the number of occurrences in non-overlapping intervals are independent.

2. Individuality: for sufficiently short time periods of length Δt, the probability of 2 or more events occurring in the interval is close to zero i.e. events occur singly not in clusters. More precisely, as $\Delta t \to 0$, the probability of two or more events in the interval of length Δt must go to zero faster than $\Delta t \to 0$. or that

$$P(\text{2 or more events in } (t, t + \Delta t)) = o(\Delta t) \text{ as } \Delta t \to 0.$$

3. Homogeneity or Uniformity: events occur at a uniform or homogeneous rate λ over time so that the probability of one occurrence in an interval $(t, t + \Delta t)$ is approximately $\lambda \Delta t$ for small Δt for any value of t. More precisely,

$$P(\text{one event in } (t, t + \Delta t)) = \lambda \Delta t + o(\Delta t).$$

These three conditions together define a **Poisson Process.**

Let X be the number of event occurrences in a time period of length t. Then it can be shown (see below) that X has a Poisson distribution with $\mu = \lambda t$.

Illustrations:

(1) The emission of radioactive particles from a substance follows a Poisson process. (This is used in medical imaging and other areas.)

(2) Hits on a web site during a given time period often follow a Poisson process.

(3) Occurrences of certain non-communicable diseases sometimes follow a Poisson process.

Probability Function: We can derive the probability function $f(x) = P(X = x)$ from the conditions above. We are interested in time intervals of arbitrary length t, so as a temporary notation, let $f_t(x)$ be the probability of x occurrences in a time interval of length t. We now relate $f_t(x)$ and $f_{t+\Delta t}(x)$. From that we can determine what $f_t(x)$ is.
To find $f_{t+\Delta t}(x)$ we note that for Δt small there are only 2 ways to get x event occurrences by time $t + \Delta t$. Either there are x events by time t and no more from t to $t + \Delta t$ or there are $x - 1$ by time t and 1 more from t to $t + \Delta t$. (since $P\left(\geq 2 \text{ in time } \Delta t\right) = 0(\Delta t)$, other possibilities are negligible if Δt is small). This and condition 1 above (independence) imply that

$$f_{t+\Delta t}(x) \doteq f_t(x)(1 - \lambda \Delta t) + f_t(x - 1)(\lambda \Delta t)$$

Re-arranging gives $\frac{f_{t+\Delta t}(x)-f_t(x)}{\Delta t} \doteq \lambda\,[f_t(x-1)-f_t(x)]$.
Taking the limit as $\Delta t \to 0$ we get

$$\frac{d}{dt}f_t(x) = \lambda\,[f_t(x-1)-f_t(x)]$$

This "differential-difference" equation can be "solved" by using the "boundary" conditions $f_0(0)=1$ and $f_0(x)=0$ for $x=1,2,3,\cdots$. You can confirm that

$$f_t(x) = f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}; \quad x=0,1,2,\cdots$$

satisfies these conditions and the equation above, even though you don't know how to solve the equations. If we let $\mu = \lambda t$, we can re-write $f(x)$ as $f(x) = \frac{\mu^x e^{-\mu}}{x!}$, which is the Poisson distribution from Section 6.7. That is:

> In a Poisson process with rate of occurrence λ, the number of event occurrences X in a time interval of length t has a Poisson distribution with $\mu = \lambda t$.

Interpretation of μ and λ:
λ is referred to as the **intensity or rate of occurrence** parameter for the events. It represents the average rate of occurrence of events per unit of time (or area or volume, as discussed below). Then $\lambda t = \mu$ represents the average number of occurrences in t units of time. It is important to note that the value of λ depends on the units used to measure time. For example, if phone calls arrive at a store at an average rate of 20 per hour, then $\lambda = 20$ when time is in hours and the average in 3 hours will be 3×20 or 60. However, if time is measured in minutes then $\lambda = 20/60 = 1/3$; the average in 180 minutes (3 hours) is still $(1/3)(180) = 60$.

Examples:

(1) Suppose earthquakes recorded in Ontario each year follow a Poisson process with an average of 6 per year. The probability that 7 earthquakes will be recorded in a 2 year period is $f(7) = \frac{12^7 e^{-12}}{7!} = .0437$. We have used $\lambda = 6$ and $t = 2$ to get $\mu = 12$.

(2) At a nuclear power station an average of 8 leaks of heavy water are reported per year. Find the probability of 2 or more leaks in 1 month, if leaks follow a Poisson process.

Solution: Assuming leaks satisfy the conditions for a Poisson process and that a month is $1/12$ of a year, we'll use the Poisson distribution with $\lambda = 8$ and $t = 1/12$, so $\mu = 8/12$. Thus

$$\begin{aligned} P(X \geq 2) &= 1 - P(X<2) \\ &= 1 - [f(0)+f(1)] \\ &= 1 - \left[\frac{(8/12)^0 e^{-8/12}}{0!} + \frac{\left(\frac{8}{12}\right)^1 e^{-8/12}}{1!}\right] = .1443 \end{aligned}$$

Random Occurrence of Events in Space

The Poisson process also applies when "events" occur randomly in space (either 2 or 3 dimensions). For example, the "events" might be bacteria in a volume of water or blemishes in the finish of a paint job on a metal surface. If X is the number of events in a volume or area in space of size v and if λ is the average number of events per unit volume (or area), then X has a Poisson distribution with $\mu = \lambda v$.

For this model to be valid, it is assumed that the Poisson process conditions given previously apply here, with "time" replaced by "volume" or "area". Once again, note that the value of λ depends on the units used to measure volume or area.

Example: Coliform bacteria occur in river water with an average intensity of 1 bacteria per 10 cubic centimeters (cc) of water. Find (a) the probability there are no bacteria in a 20cc sample of water which is tested, and (b) the probability there are 5 or more bacteria in a 50cc sample. (To do this assume that a Poisson process describes the location of bacteria in the water at any given time.)

Solution: Let X = number of bacteria in a sample of volume v cc. Since λ = 0.1 bacteria per cc (1 per 10cc) the p.f. of X is Poisson with $\mu = .1v$,

$$f(x) = e^{-.1v}\frac{(.1v)^x}{x!} \quad x = 0, 1, 2, \ldots$$

Thus we find

(a) With $v = 20, \mu = 2$ so $P(X = 0) = f(0) = e^{-2} = .135$

(b) With $v = 50, \mu = 5$ so $f(x) = e^{-5}5^x/x!$ and $P(X \geq 5) = 1 - P(X \leq 4) = .440$

(Note: we can use the R command $ppois(4, 5)$ to get $P(X \leq 4)$.)

Exercise: In each of the above examples, how well are each of the conditions for a Poisson process likely to be satisfied?

Distinguishing Poisson from Binomial and Other Distributions

Students often have trouble knowing when to use the Poisson distribution and when not to use it. To be certain, the 3 conditions for a Poisson process need to be checked. However, a quick decision can often be made by asking yourself the following questions:

1. Can we specify in advance the maximum value which X can take?
   If we can, then the distribution is <u>not</u> Poisson. If there is no fixed upper limit, the distribution

might be Poisson, but is certainly not binomial or hypergeometric, e.g. the number of seeds which germinate out of a package of 25 does not have a Poisson distribution since we know in advance that $X \leq 25$. The number of cardinals sighted at a bird feeding station in a week might be Poisson since we can't specify a fixed upper limit on X. At any rate, this number would not have a binomial or hypergeometric distribution.

2. Does it make sense to ask how often the event did <u>not</u> occur?
   If it does make sense, the distribution is not Poisson. If it does not make sense, the distribution might be Poisson. For example, it does not make sense to ask how often a person did not hiccup during an hour. So the number of hiccups in an hour might have a Poisson distribution. It would certainly not be binomial, negative Binomial, or hypergeometric. If a coin were tossed until the 3^{rd} head occurs it does make sense to ask how often heads did not come up. So the distribution would not be Poisson. (In fact, we'd use negative binomial for the number of non-heads; i.e. tails.)

Problems:

6.8.1 Suppose that emergency calls to 911 follow a Poisson process with an average of 3 calls per minute. Find the probability there will be

a) 6 calls in a period of $2\frac{1}{2}$ minutes.

b) 2 calls in the first minute of a $2\frac{1}{2}$ minute period, given that 6 calls occur in the entire period.

6.8.2 Misprints are distributed randomly and uniformly in a book, at a rate of 2 per 100 lines.

(a) What is the probability a line is free of misprints?

(b) Two pages are selected at random. One page has 80 lines and the other 90 lines. What is the probability that there are exactly 2 misprints on each of the two pages?

6.9 Combining Models

While we've considered the model distributions in this chapter one at a time, we will sometimes need to use two or more distributions to answer a question. To handle this type of problem you'll need to be very clear about the characteristics of each model. Here is a somewhat artificial illustration. Lots of other examples are given in the problems at the end of the chapter.

Example: A very large (essentially infinite) number of ladybugs is released in a large orchard. They scatter randomly so that on average a tree has 6 ladybugs on it. Trees are all the same size.

a) Find the probability a tree has > 3 ladybugs on it.

b) When 10 trees are picked at random, what is the probability 8 of these trees have > 3 ladybugs on them?

c) Trees are checked until 5 with > 3 ladybugs are found. Let X be the total number of trees checked. Find the probability function, $f(x)$.

d) Find the probability a tree with > 3 ladybugs on it has exactly 6.

e) On 2 trees there are a total of t ladybugs. Find the probability that x of these are on the first of these 2 trees.

Solution:

a) If the ladybugs are randomly scattered the most suitable model is the Poisson distribution with $\lambda = 6$ and $v = 1$ (i.e. any tree has a "volume" of 1 unit), so $\mu = 6$ and

$$\begin{aligned} P(X > 3) &= 1 - P(X \le 3) = 1 - [f(0) + f(1) + f(2) + f(3)] \\ &= 1 - \left[\tfrac{6^0 e^{-6}}{0!} + \tfrac{6^1 e^{-6}}{1!} + \tfrac{6^2 e^{-6}}{2!} + \tfrac{6^3 e^{-6}}{3!}\right] = .8488 \end{aligned}$$

b) Using the binomial distribution where "success" means > 3 ladybugs on a tree, we have $n = 10, p = .8488$ and

$$f(8) = \binom{10}{8}(.8488)^8(1 - .8488)^2 = .2772$$

c) Using the negative binomial distribution, we need the number of successes, k, to be 5, and the number of failures to be $(x - 5)$. Then

$$\begin{aligned} f(x) &= \tbinom{x-5+5-1}{x-5}(.8488)^5(1 - .8488)^{x-5} \\ &= \tbinom{x-1}{x-5}(.8488)^5(1 - .8488)^{x-5} \text{ or } \tbinom{x-1}{4}(.8488)^5(.1512)^{x-5}; \;\; x = 5, 6, 7, \cdots \end{aligned}$$

d) This is conditional probability. Let A = { 6 ladybugs} and B = { $>$ 3 ladybugs }. Then

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(6 \text{ ladybugs})}{P(\; > \; 3 \text{ ladybugs})} = \frac{\frac{6^6 e^{-6}}{6!}}{0.8488} = 0.1892.$$

e) Again we need to use conditional probability.

$$P(x\text{on } 1^{\text{st}} \text{ tree}|\text{total of } t) = \frac{P(x \text{ on } 1^{\text{st}} \text{ tree and total of } t)}{P(\text{total of } t)}$$

$$
\begin{aligned}
&= \frac{P(x \text{ on } 1^{\text{st}} \text{ tree and } t{-}x \text{ on } 2^{\text{nd}} \text{ tree})}{P(\text{ total of } t)} \\
&= \frac{P(x \text{ on } 1^{\text{st}} \text{ tree}).\ P(t{-}x \text{ on } 2^{\text{nd}} \text{ tree})}{P(\text{total of } t)}
\end{aligned}
$$

Use the Poisson distribution to calculate each, with $\mu = 6 \times 2 = 12$ in the denominator since there are 2 trees.

$$
\begin{aligned}
P(x \text{ on } 1^{\text{st}} \text{ tree}|\text{total of } t) &= \frac{\left(\frac{6^x e^{-6}}{x!}\right)\left(\frac{6^{t-x} e^{-6}}{(t-x)!}\right)}{\frac{12^t e^{-12}}{t!}} \\
&= \frac{t!}{x!(t-x)!}\left(\frac{6}{12}\right)^x \left(\frac{6}{12}\right)^{t-x} \\
&= \binom{t}{x}\left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{t-x}, \quad x = 0, 1, \cdots, t.
\end{aligned}
$$

Caution: Don't forget to give the range of x. If the total is t, there couldn't be more than t ladybugs on the 1^{st} tree.

Exercise: The answer to (e) is a binomial probability function. Can you reach this answer by general reasoning rather than using conditional probability to derive it?

Problems:

6.9.1 In a Poisson process the average number of occurrences is λ per minute. Independent 1 minute intervals are observed until the first minute with no occurrences is found. Let X be the number of 1 minute intervals required, including the last one. Find the probability function, $f(x)$.

6.9.2 Calls arrive at a telephone distress centre during the evening according to the conditions for a Poisson process. On average there are 1.25 calls per hour.

(a) Find the probability there are no calls during a 3 hour shift.

(b) Give an expression for the probability a person who starts working at this centre will have the first shift with no calls on the 15^{th} shift.

(c) A person works one hundred 3 hour evening shifts during the year. Give an expression for the probability there are no calls on at least 4 of these 100 shifts. Calculate a numerical answer using a Poisson approximation.

6.10 Summary of Single Variable Discrete Models

Name	Probability Function
Discrete Uniform	$f(x) = \frac{1}{b-a+1};\ x = a, a+1, a+2, \cdots b$
Hypergeometric	$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}};\ x = \max(0, n-(N-r)), \cdots, \min(n, r)$
Binomial	$f(x) = \binom{n}{x} p^x (1-p)^{n-x};\ x = 0, 1, 2, \cdots n$
Negative Binomial	$f(x) = \binom{x+k-1}{x} p^k (1-p)^x;\ x = 0, 1, 2, \cdots$
Geometric	$f(x) = p(1-p)^x;\ x = 0, 1, 2, \cdots$
Poisson	$f(x) = \frac{e^{-\mu}\mu^x}{x!};\ x = 0, 1, 2, \cdots$

6.11 Appendix: R Software

The R software system is a powerful tool for handling probability distributions and data concerning random variables. The following short notes describe basic features of R; further information and links to other resources are available on the course web page. Unix and Windows versions of R are available on Math Faculty undergraduate servers, and free copies can be downloaded from the web.

You should make yourself familiar with R, since some problems (and most applications of probability) require computations or graphics which are not feasible by hand.

```
Some R Basics

R is a statistical software system that has excellent numerical,
graphical and statistical capabilities. There are Unix and Windows
versions. These notes are a very brief introduction to a few of the
features of R. Web resources have much more information. Links can
be found on the Stat 230 web page. You can also download a Unix or
Windows version of R for free.


1.PRELIMINARIES
R is invoked on Math Unix machines by typing R. The R prompt
is >. R objects include variables, functions, vectors, arrays, lists and
other items. To see online documentation about something, we use the help
function. For example, to see documentation on the function mean(), type
help(mean). In some cases help.search() is helpful.

The assignment symbol is <- : for example,
      x<- 15   assigns the value 15 to variable x.

To quit an R session, type  q()

2.VECTORS
Vectors can consist of numbers or other symbols; we will consider only
numbers here. Vectors are defined using c(): for example,
   x<- c(1,3,5,7,9)
defines a vector of length 5 with the elements given. Vectors and other
classes of objects possess certain attributes. For example, typing
```

length(x) will give the length of the vector x. Vectors are a convenient way to store values of a function (e.g. a probability function or a c.d.f) or values of a random variable that have been recorded in some experiment or process.

3.ARITHMETIC

The following R commands and responses should explain arithmetic operations.

```
> 7+3
[1] 10
> 7*3
[1] 21
> 7/3
[1] 2.333333
> 2^3
[1] 8
```

4.SOME FUNCTIONS

Functions of many types exist in R. Many operate on vectors in a transparent way, as do arithmetic operations. (For example, if x and y are vectors then x+y adds the vectors element-wise; thus x and y must be the same length.) Some examples, with comments, follow.

```
> x<- c(1,3,5,7,9)   # Define a vector x
> x            # Display x
[1] 1 3 5 7 9
> y<- seq(1,2,.25)    #A useful function for defining a vector whose
                      elements are an arithmetic progression
> y
[1] 1.00 1.25 1.50 1.75 2.00
> y[2]   # Display the second element of vector y
[1] 1.25
> y[c(2,3)]    # Display the vector consisting of the second and
                   third elements of vector y.
[1] 1.25 1.50
```

```
> mean(x)      #Computes the mean of the elements of vector x
[1] 5
> summary(x)    # A useful function which summarizes features of
                   a vector x
 Min. 1st Qu. Median Mean 3rd Qu. Max.
    1       3      5    5       7    9
> var(x)    # Computes the (sample) variance of the elements of x
[1] 10
> exp(1)    # The exponential function
[1] 2.718282
> exp(y)
[1] 2.718282 3.490343 4.481689 5.754603 7.389056
> round(exp(y),2)   # round(y,n) rounds the elements of vector y to
                       n decimals
[1] 2.72 3.49 4.48 5.75 7.39
> x+2*y
[1]  3.0  5.5  8.0 10.5 13.0
```

5. GRAPHS

To open a graphics window in Unix, type x11(). Note that in R, a graphics window opens automatically when a graphical function is used.
There are various plotting and graphical functions. Two useful ones are

```
plot(x,y)  # Gives a scatterplot of x versus y; thus x and y must
              be vectors of the same length.

hist(x)    # Creates a frequency histogram based on the values in
              the vector x. To get a relative frequency histogram
              (areas of rectangles sum to one) use hist(x,prob=T).
```

Graphs can be tailored with respect to axis labels, titles, numbers of plots to a page etc. Type help(plot), help(hist) or help(par) for some information.

To save/print a graph in R using UNIX, you generate the graph you would like to save/print in R using a graphing function like plot() and type:

```
dev.print(device,file="filename")
```

where device is the device you would like to save the graph to (i.e. x11) and filename is the name of the file that you would like the graph saved to. To look at a list of the different graphics devices you can save to, type help(Devices).

To save/print a graph in R using Windows, you can do one of two things.

a) You can go to the File menu and save the graph using one of several formats (i.e. postscript, jpeg, etc.). It can then be printed. You may also copy the graph to the clipboard using one of the formats and then paste to an editor, such as MS Word. Note that the graph can be printed directly to a printer using this option as well.

b) You can right click on the graph. This gives you a choice of copying the graph and then pasting to an editor, such as MS Word, or saving the graph as a metafile or bitmap. You may also print directly to a printer using this option as well.

6.DISTRIBUTIONS

There are functions which compute values of probability or probability density functions, cumulative distribution functions, and quantiles for various distributions. It is also possible to generate (pseudo) random samples from these distributions. Some examples follow for Binomial and Poisson distributions. For other distribution information, type help(rhyper), help(rnbinom) and so on. Note that R does not have any function specifically designed to generate random samples from a discrete uniform distribution (although there is one for a continous uniform distribution). To generate n random samples from a discrete UNIF(a,b), use sample(a:b,n,replace=T).

```
> y<- rbinom(10,100,0.25)    # Generate 10 random values from the Binomial
                               distribution Bi(100,0.25). The values are
```

```
                                          stored in the vector y.
> y    # Display the values
 [1] 24 24 26 18 29 29 33 28 28 28
> pbinom(3,10,0.5)     # Compute P(Y<=3) for a Bi(10,0.5) random variable.
[1] 0.171875
> qbinom(.95,10,0.5)   # Find the .95 quantile (95th percentile) for
[1] 8                    Bi(10,0.5).

> z<- rpois(10,10)    # Generate 10 random values from the Poisson
                        distribution Poisson(10). The values are stored in the
                        vector z.
> z   # Display the values
 [1]  6  5 12 10  9  7  9 12  5  9
> ppois(3,10)     # Compute P(Y<=3) for a Poisson(10) random variable.
[1] 0.01033605
> qpois(.95,10)   # Find the .95 quantile (95th percentile) for
[1] 15              Poisson(10).
```

To illustrate how to plot the probability function for a random variable, a Bi(10,0.5) random variable is used.

```
# Assign all possible values of the random variable, X ~ Bi(10,0.5)
x <- seq(0,10,by=1)

# Determine the value of the probability function for possible values of X
x.pf <- dbinom(x,10,0.5)

# Plot the probability function
barplot(x.pf,xlab="X",ylab="Probability Function",
names.arg=c("0","1","2","3","4","5","6","7","8","9","10"))
```

6.12 Problems on Chapter 6

6.1 Suppose that the probability $p(x)$ a person born in 1950 lives at least to certain ages x is as given in the table below.

	x:	30	50	70	80	90
Females		.980	.955	.910	.595	.240
Males		.960	.920	.680	.375	.095

(a) If a female lives to age 50, what is the probability she lives to age 80? To age 90? What are the corresponding probabilities for males?

(b) If 51% of persons born in 1950 were male, find the fraction of the total population (males and females) that will live to age 90.

6.2 Let X be a non-negative discrete random variable with cumulative distribution function

$$F(x) = 1 - 2^{-x} \text{ for } x = 0, 1, 2, ...$$

(a) Find the probability function of X.

(b) Find the probability of the event $X = 5$; the event $X \geq 5$.

6.3 Two balls are drawn at random from a box containing ten balls numbered $0, 1, ..., 9$. Let random variable X be the *larger* of the numbers on the two balls and random variable Y be their *total*.

(a) Tabulate the p.f. of X and of Y if the sampling is *without* replacement.

(b) Repeat (a) if the sampling is *with* replacement.

6.4 Let X have a geometric distribution with $f(x) = p(1-p)^x$; $x = 0, 1, 2, \cdots$. Find the probability function of R, the remainder when X is divided by 4.

6.5 (a) Todd decides to keep buying a lottery ticket each week until he has 4 winners (of some prize). Suppose 30% of the tickets win some prize. Find the probability he will have to buy 10 tickets.

(b) A coffee chain claims that you have a 1 in 9 chance of winning a prize on their “roll up the edge" promotion, where you roll up the edge of your paper cup to see if you win. If so, what is the probability you have no winners in a one week period where you bought 15 cups of coffee?

(c) Over the last week of a month long promotion you and your friends bought 60 cups of coffee, but there was only 1 winner. Find the probability that there would be this few (i.e. 1 or 0) winners. What might you conclude?

6.6 An oil company runs a contest in which there are 500,000 tickets; a motorist receives one ticket with each fill-up of gasoline, and 500 of the tickets are winners.

(a) If a motorist has ten fill-ups during the contest, what is the probability that he or she wins at least one prize?

(b) If a particular gas bar distributes 2,000 tickets during the contest, what is the probability that there is at least one winner among the gas bar's customers?

6.7 **Jury selection**. During jury selection a large number of people are asked to be present, then persons are selected one by one in a random order until the required number of jurors has been chosen. Because the prosecution and defense teams can each reject a certain number of persons, and because some individuals may be exempted by the judge, the total number of persons selected before a full jury is found can be quite large.

(a) Suppose that you are one of 150 persons asked to be present for the selection of a jury. If it is necessary to select 40 persons in order to form the jury, what is the probability you are chosen?

(b) In a recent trial the numbers of men and women present for jury selection were 74 and 76. Let Y be the number of men picked for a jury of 12 persons. Give an expression for $P(Y = y)$, assuming that men and women are equally likely to be picked.

(c) For the trial in part (b), the number of men selected turned out to be two. Find $P(Y \leq 2)$. What might you conclude from this?

6.8 A waste disposal company averages 6.5 spills of toxic waste per month. Assume spills occur randomly at a uniform rate, and independently of each other, with a negligible chance of 2 or more occurring at the same time. Find the probability there are 4 or more spills in a 2 month period.

6.9 Coliform bacteria are distributed randomly and uniformly throughout river water at the average concentration of one per twenty cubic centimetres of water.

(a) What is the probability of finding exactly two coliform bacteria in a 10 cubic centimetre sample of the river water?

(b) What is the probability of finding at least one coliform bacterium in a 1 cubic centimetre sample of the river water?

(c) In testing for the concentration (average number per unit volume) of bacteria it is possible to determine cheaply whether a sample has **any** (i.e. 1 or more) bacteria present or not.

Suppose the average concentration of bacteria in a body of water is λ per cubic centimetre. If 10 independent water samples of 10 c.c. each are tested, let the random variable Y be the number of samples with **no** bacteria. Find $P(Y = y)$.

(d) Suppose that of 10 samples, 3 had no bacteria. Find an estimate for the value of λ.

6.10 In a group of policy holders for house insurance, the average number of claims per 100 policies per year is $\lambda = 8.0$. The number of claims for an individual policy holder is assumed to follow a Poisson distribution.

(a) In a given year, what is the probability an individual policy holder has at least one claim?

(b) In a group of 20 policy holders, what is the probability there are no claims in a given year? What is the probability there are two or more claims?

6.11 Assume power failures occur independently of each other at a uniform rate through the months of the year, with little chance of 2 or more occurring simultaneously. Suppose that 80% of months have no power failures.

a) Seven months are picked at random. What is the probability that 5 of these months have no power failures?

b) Months are picked at random until 5 months without power failures have been found. What is the probability that 7 months will have to be picked?

c) What is the probability a month has more than one power failure?

6.12 a) Let $f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, and keep $p = \frac{r}{N}$ fixed. (e.g. If N doubles then r also doubles.) Prove that $\lim_{N \to \infty} f(x) = \binom{n}{x} p^x (1-p)^{n-x}$.

b) What part of the chapter is this related to?

6.13 Spruce budworms are distributed through a forest according to a Poisson process so that the average is λ per hectare.

a) Give an expression for the probability that at least 1 of n one hectare plots contains at least k spruce budworms.

b) Discuss briefly which assumption(s) for a Poisson process may not be well satisfied in this situation.

6.14 A person working in telephone sales has a 20% chance of making a sale on each call, with calls being independent. Assume calls are made at a uniform rate, with the numbers made in non-overlapping periods being independent. On average there are 20 calls made per hour.

a) Find the probability there are 2 sales in 5 calls.

b) Find the probability exactly 8 calls are needed to make 2 sales.

c) If 8 calls were needed to make 2 sales, what is the probability there was 1 sale in the first 3 of these calls?

d) Find the probability of 3 calls being made in a 15 minute period.

6.15 A bin at a hardware store contains 35 forty watt lightbulbs and 70 sixty watt bulbs. A customer wants to buy 8 sixty watt bulbs, and withdraws bulbs without replacement until these 8 bulbs have been found. Let X be the number of 40 watt bulbs drawn from the bin. Find the probability function, $f(x)$.

6.16 During rush hour the number of cars passing through a particular intersection has a Poisson distribution with an average of 540 per hour.

a) Find the probability there are 11 cars in a 30 second interval and the probability there are 11 or more cars.

b) Find the probability that when 20 disjoint 30 second intervals are studied, exactly 2 of them had 11 cars.

c) We want to find 12 intervals having 11 cars in 30 seconds.

(i) Give an expression for the probability 1400 30 second intervals have to be observed to find the 12 having the desired traffic flow.

(ii) Use an approximation which involves the Poisson distribution to evaluate this probability and justify why this approximation is suitable.

6.17 (a) Bubbles are distributed in sheets of glass, as a Poisson process, at an intensity of 1.2 bubbles per square metre. Let X be the number of sheets of glass, in a shipment of n sheets, which have no bubbles. Each sheet is 0.8m^2. Give the probability function of X.

(b) The glass manufacturer wants to have at least 50% of the sheets of glass with no bubbles. How small will the intensity λ need to be to achieve this?

6.18 Random variable X takes values 1,2,3,4,5 and has c.d.f.

x	0	1	2	3	4	5
$F(x)$	0	$.1k$	.2	$.5k$	k	$4k^2$

Find k, $f(x)$ and $P(2 < X \leq 4)$. Draw a histogram of $f(x)$.

6.19 Let random variable Y have a *geometric distribution* $P(Y = y) = p(1 - p)^y$ for $y = 0, 1, 2, \ldots$.

(a) Find an expression for $P(Y \geq y)$, and show that $P(Y \geq s+t|Y \geq s) = P(Y \geq t)$ for all non-negative integers s, t.

(b) What is the most probable value of Y?

(c) Find the probability that Y is divisible by 3.

(d) Find the probability function of random variable R, the *remainder* when Y is divided by 3.

6.20 **Polls and Surveys**. Polls or surveys in which people are selected and their opinions or other characteristics are determined are very widely used. For example, in a survey on cigarette use among teenage girls, we might select a random sample of n girls from the population in question, and determine the number X who are regular smokers. If p is the fraction of girls who smoke, then $X \sim Bi(n, p)$. Since p is unknown (that is why we do the survey) we then estimate it as $\hat{p} = X/n$. (In probability and statistics a “hat" is used to denote an estimate of a model parameter based on data.) The binomial distribution can be used to study how “good" such estimates are, as follows

(a) Suppose $p = .3$ and $n = 100$. Find the probability $P(.27 \leq \frac{X}{n} \leq .33)$. Many surveys try to get an estimate X/n which is within 3% (.03) of p with high probability. What do you conclude here?

(b) Repeat the calculation in (a) if $n = 400$ and $n = 1000$. What do you conclude?

(c) If $p = .5$ instead of .3, find $P(.47 \leq \frac{X}{n} \leq .53)$ when $n = 400$ and 1000.

(d) Your employer asks you to design a survey to estimate the fraction p of persons age 25-34 who download music via the internet. The objective is to get an estimate accurate to within 3%, with probability close to .95. What size of sample (n) would you recommend?

6.21 **Telephone surveys**. In some “random digit dialing" surveys, a computer phones randomly selected telephone numbers. However, not all numbers are “active" (belong to a telephone account) and they may belong to businesses as well as to individual or residences.

Suppose that for a given large set of telephone numbers, 57% are active residential or individual numbers. We will call these “personal" numbers.

Suppose that we wish to interview (over the phone) 1000 persons in a survey.

(a) Suppose that the probability a call to a personal number is answered is .8, and that the probability the person answering agrees to be interviewed is .7. Give the probability distribution for X, the number of calls needed to obtain 1000 interviews.

(b) Use R software to find $P(X \leq x)$ for the values $x = 2900,\ 3000,\ 3100,\ 3200$.

(c) Suppose instead that 3200 randomly selected numbers were dialed. Give the probability distribution for Y, the number of interviews obtained, and find $P(Y \geq 1000)$.

(Note: The R functions *pnbinom* and *pbinom* give negative binomial and binomial probabilities, respectively.)

7. Expectation, Averages, Variability

7.1 Summarizing Data on Random Variables

When we return midterm tests, someone almost always asks what the average was. While we could list out all marks to give a picture of how students performed, this would be tedious. It would also give more detail than could be immediately digested. If we summarize the results by telling a class the average mark, students immediately get a sense of how well the class performed. For this reason, "summary statistics" are often more helpful than giving full details of every outcome.

To illustrate some of the ideas involved, suppose we were to observe cars crossing a toll bridge, and record the number, X, of people in each car. Suppose in a small study data on 25 cars were collected. We could list out all 25 numbers observed, but a more helpful way of presenting the data would be in terms of the **frequency distribution** below, which gives the number of times (the "frequency") each value of X occurred.

X	Frequency Count	Frequency
1	卌 |	6
2	卌 |||	8
3	卌	5
4	|||	3
5	||	2
6	|	1

We could also draw a *frequency* histogram of these frequencies:

Frequency distributions or histograms are good summaries of data because they show the variability in the observed outcomes very clearly. Sometimes, however, we might prefer a single-number summary. The most common such summary is the average, or arithmetic mean of the outcomes. The mean

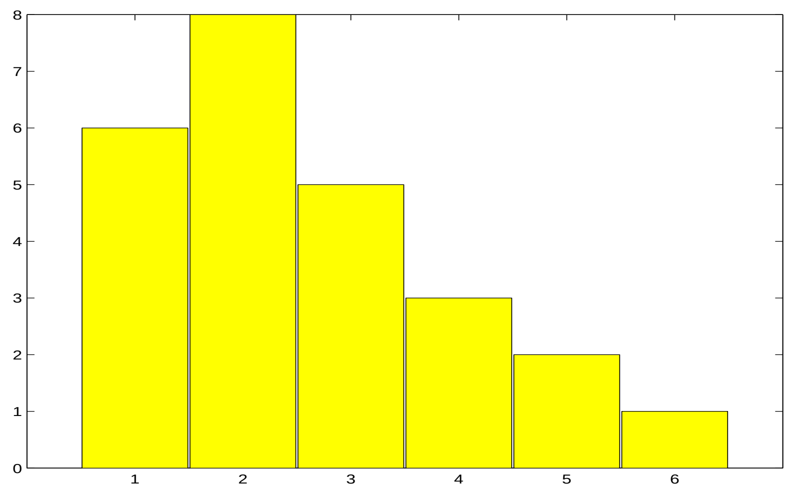


Figure 7.1: Frequency Histogram

of n outcomes $x_1, \ldots, x_n$ for a random variable X is $\sum_{i=1}^{n} x_i/n$, and is denoted by $\bar{x}$. The arithmetic mean for the example above can be calculated as

$$\frac{(6 \times 1) + (8 \times 2) + (5 \times 3) + (3 \times 4) + (2 \times 5) + (1 \times 6)}{25} = \frac{65}{25} = 2.60$$

That is, there was an average of 2.6 persons per car. A set of observed outcomes $x_1, \ldots, x_n$ for a random variable X is termed a **sample** in probability and statistics. To reflect the fact that this is the average for a particular sample, we refer to it as the **sample mean**. Unless somebody deliberately "cooked" the study, we would not expect to get precisely the same sample mean if we repeated it another time. Note also that $\bar{x}$ is not in general an integer, even though X is.

Two other common summary statistics are the median and mode.

Definition 12 *The* ***median*** *of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.*

If these 25 results were written in order, the 13^{th} outcome would be a 2. So the median is 2. By convention, we go half way between the middle two values if there are an even number of observations.

Definition 13 *The* ***mode*** *of the sample is the value which occurs most often. In this case the mode is 2. There is no guarantee there will be only a single mode.*

7.2 Expectation of a Random Variable

The statistics in the preceding section summarize features of a sample of observed X-values. The same idea can be used to summarize the probability distribution of a random variable X. To illustrate, consider the previous example, where X is the number of persons in a randomly selected car crossing a toll bridge.

Note that we can re-arrange the expression used to calculate $\overline{x}$ for the sample, as

$$\frac{(6\times 1)+(8\times 2)+\cdots+(1\times 6)}{25} = (1)\left(\frac{6}{25}\right)+(2)\left(\frac{8}{25}\right)+(3)\left(\frac{5}{25}\right)+(4)\left(\frac{3}{25}\right)+(5)\left(\frac{2}{25}\right)+(6)\left(\frac{1}{25}\right)$$

$$= \sum_{x=1}^{6} x \times \text{fraction of times } x \text{ occurs}$$

Now suppose we know that the probability function of X is given by

x	1	2	3	4	5	6
$f(x)$	.30	.25	.20	.15	.09	.01

Using the relative frequency "definition" of probability, if we observed a very large number of cars, the fraction (or relative frequency) of times $X = 1$ would be .30, for $X = 2$, this proportion would be .25, etc. So, *in theory*, (according to the probability model) we would expect the mean to be

$$(1)(.30) + (2)(.25) + (3)(.20) + (4)(.15) + (5)(.09) + (6)(.01) = 2.51$$

if we observed an infinite number of cars. This "theoretical" mean is usually denoted by μ or $E(X)$, and requires us to know the distribution of X. With this background we make the following mathematical definition.

Definition 14 *The expectation (also called the mean or the expected value) of a discrete random variable X with probability function $f(x)$ is*

$$E(X) = \sum_{\text{all x}} x f(x).$$

The expectation of X is also often denoted by the Greek letter μ. The expectation of X can be thought of physically as the average of the X-values that would occur in an infinite series of repetitions of the process where X is defined. This value not only describes one aspect of a probability distribution, but is also very important in certain types of applications. For example, if you are playing a casino game in which X represents the amount you win in a single play, then $E(X)$ represents your average winnings (or losses!) per play.

Sometimes we may not be interested in the average value of X itself, but in some function of X. Consider the toll bridge example once again, and suppose there is a toll which depends on the number of car occupants. For example, a toll of $1 per car plus 25 cents per occupant would produce an average toll for the 25 cars in the study of Section 7.1 equal to

$$(1.25)\left(\frac{6}{25}\right)+(1.50)\left(\frac{8}{25}\right)+(1.75)\left(\frac{5}{25}\right)+(2.00)\left(\frac{3}{25}\right)+(2.25)\left(\frac{2}{25}\right)+(2.50)\left(\frac{1}{25}\right)=\$1.65$$

If X has the theoretical probability function $f(x)$ given above, then the average value of this $(.25X + 1) toll would be defined in the same way, as,

$$(1.25)(.30)+(1.50)(.25)+(1.75)(.20)+(2.00)(.15)+(2.25)(.09)+(2.50)(.01)=\$1.6275$$

We call this the expectation of $(0.25X+1)$ and write $E\,(0.25X+1)=1.6275.$
As a further illustration, suppose a toll designed to encourage car pooling charged $\$12/x^2$ if there were x people in the car. This scheme would yield an average toll, in theory, of

$$\left(\frac{12}{1}\right)(.30)+\left(\frac{12}{4}\right)(.25)+\left(\frac{12}{9}\right)(.20)+\left(\frac{12}{16}\right)(.15)+\left(\frac{12}{25}\right)(.09)+\left(\frac{12}{36}\right)(.01)=\$4.7757$$

that is,

$$E\left(\frac{12}{X^2}\right)=4.7757$$

is the "expectation" of $\left(\frac{12}{X^2}\right)$.
With this as background, we can now make a formal definition.

Definition 15 *The **expectation** of some function $g(X)$ of a random variable X with probability function $f(x)$ is*

$$E\left[g(X)\right]=\sum_{\text{all } x} g(x)f(x)$$

Notes:

(1) You can interpret $E[g(X)]$ as the average value of $g(X)$ in an infinite series of repetitions of the process where X is defined.

(2) $E\left[g(X)\right]$ is also known as the "expected value" of $g(X)$. This name is somewhat misleading since the average value of $g(X)$ may be a value which $g(X)$ never takes - hence unexpected!

(3) The case where $g(x) = X$ reduces to our earlier definition of $E(X)$.

(4) Confusion sometimes arises because we have two notations for the mean of a probability distribution: μ and $E(X)$ mean the same thing. There is nothing wrong with this.

(5) When calculating expectations, look at your answer to be sure it makes sense. If X takes values from 1 to 10, you should know you've made an error if you get $E(X) > 10$ or $E(X) < 1$. In physical terms, $E(X)$ is the balance point for the histogram of $f(x)$.

Let us note a couple of mathematical properties of expectation that can help to simplify calculations.

Properties of Expectation:
If your linear algebra is good, it may help if you think of E as being a linear operator. Otherwise you'll have to remember these and subsequent properties.

1. For constants a and b,

$$E\left[ag(X) + b\right] = aE\left[g(X)\right] + b$$

Proof:
$$\begin{aligned} E\left[ag(X) + b\right] &= \sum_{\text{all } x} \left[ag(x) + b\right] f(x) \\ &= \sum_{\text{all } x} \left[ag(x)f(x) + bf(x)\right] \\ &= a \sum_{\text{all } x} g(x)f(x) + b \sum_{\text{all } x} f(x) \\ &= aE\left[g(X)\right] + b \quad \left(\text{since } \sum_{\text{all } x} f(x) = 1\right) \end{aligned}$$

2. For constants a and b and functions g_1 and g_2, it is also easy to show

$$E\left[ag_1(X) + bg_2(X)\right] = aE\left[g_1(X)\right] + bE\left[g_2(X)\right]$$

Don't let expectation intimidate you. Much of it is common sense. For example, property 1 says $E(13) = 13$ if we let $a = 0$ and $b = 13$. The expectation of a constant b is of course equal to b. It also says $E\left(2X\right) = 2E(X)$ if we let $a = 2$, $b = 0$, and $g(X) = X$. This is obvious also. Note, however, that for $g(x)$ a nonlinear function, it is not true that $E[g(X)] = g(E(X))$; this is a common mistake. (Check this for the example above when $g(X) = 12/X^2$.)

7.3 Some Applications of Expectation

Because expectation is an average value, it is frequently used in problems where costs or profits are connected with the outcomes of a random variable X. It is also used a lot as a summary statistic for

probability distributions; for example, one often hears about the expected life (expectation of lifetime) for a person or the expected return on an investment.
The following are examples.

Example: Expected Winnings in a Lottery
A small lottery sells 1000 tickets numbered $000, 001, \ldots, 999$; the tickets cost \$10 each. When all the tickets have been sold the draw takes place: this consists of a simple ticket from 000 to 999 being chosen at random. For ticket holders the prize structure is as follows:

- Your ticket is drawn - win \$5000.
- Your ticket has the same first two number as the winning ticket, but the third is different - win \$100.
- Your ticket has the same first number as the winning ticket, but the second number is different - win \$10.
- All other cases - win nothing.

Let the random variable X represent the winnings from a given ticket. Find $E(X)$.

Solution: The possible values for X are 0, 10, 100, 5000 (dollars). First, we need to find the probability function for X. We find (make sure you can do this) that $f(x) = P(X = x)$ has values

$$f(0) = 0.900, \quad f(10) = 0.090, \quad f(100) = .009, 0pt \quad f(5000) = .001$$

The expected winnings are thus the expectation of X, or

$$E(X) = \sum_{\text{all x}} xf(x) = \$6.80$$

Thus, the gross expected winnings per ticket are \$6.80. However, since a ticket costs \$10 your expected net winnings are negative, -\$3.20 (i.e. an expected loss of \$3.20).

Remark: For any lottery or game of chance the expected net winnings per play is a key value. A fair game is one for which this value is 0. Needless to say, casino games and lotteries are never fair: the net winnings for a player are negative.

Remark: The random variable associated with a given problem may be defined in different ways but the expected winnings will remain the same. For example, instead of defining X as the amount won we could have defined $X = 0, 1, 2, 3$ as follows:

$X = 3$	all 3 digits of number match winning ticket
$X = 2$	1st 2 digits (only) match
$X = 1$	1st digit (but not the 2nd) match
$X = 0$	1st digit does not match

Now, we would define the function $g(x)$ as the winnings when the outcome $X = x$ occurs. Thus,

$$g(0) = 0, \qquad g(1) = 10, \qquad g(2) = 100, \qquad g(3) = 5000$$

The expected winnings are then

$$E(g(X)) = \sum_{x=0}^{3} g(x)f(x) = \$6.80,$$

the same as before.

Example: Diagnostic Medical Tests
Often there are cheaper, less accurate tests for diagnosing the presence of some conditions in a person, along with more expensive, accurate tests. Suppose we have two cheap tests and one expensive test, with the following characteristics. All three tests are positive if a person has the condition (there are no "false negatives"), but the cheap tests give "false positives".
Let a person be chosen at random, and let D = {person has the condition}. The three tests are

Test 1:	P (positive test $\vert\overline{D}$) = .05; test costs \$5.00
Test 2:	P (positive test $\vert\overline{D}$) = .03; test costs \$8.00
Test 3:	P (positive test $\vert\overline{D}$) = 0; test costs \$40.00

We want to check a large number of people for the condition, and have to choose among three testing strategies:

(i) Use Test 1, followed by Test 3 if Test 1 is positive.

(ii) Use Test 2, followed by Test 3 if Test 2 is positive.

(iii) Use Test 3.

Determine the expected cost per person under each of strategies (i), (ii) and (iii). We will then choose the strategy with the lowest expected cost. It is known that about .001 of the population have the condition (i.e. $P(D) = .001, P(\overline{D}) = .999$).

Solution: Define the random variable X as follows (for a random person who is tested):

$$X = 1 \qquad \text{if the initial test is negative}$$
$$X = 2 \qquad \text{if the initial test is positive}$$

Also let $g(x)$ be the total cost of testing the person. The expected cost per person is then

$$E[g(X)] = \sum_{x=1}^{2} g(x)f(x)$$

The probability function $f(x)$ for X and function $g(x)$ differ for strategies (i), (ii) and (iii). Consider for example strategy (i). Then

$$\begin{aligned} P(X = 2) &= P\,(\text{initial test positive}) \\ &= P(D) + P(\text{positive}|\overline{D})P(\overline{D}) \\ &= .001 + (.05)(.999) = 0.0510 \end{aligned}$$

The rest of the probabilities, associated values of $g(X)$ and $E[g(X)]$ are obtained below.

(i) $f(1) = P(X = 1) = 1 - f(2) = 1 - 0.0510 = 0.949$ (see $f(2)$ below)
$f(2) = 0.0510$ (obtained above)
$g(1) = 5 \qquad g(2) = 45$
$E[g(X)] = 5(.949) + 45(.0510) = \$7.04$

(ii) $f(2) = .001 + (.03)(.999) = .03097$
$f(1) = 1 - f(2) = .96903$
$g(1) = 8 \qquad g(2) = 48$
$E[g(X)] = 8(.96903) + 48(.03097) = \$9.2388$

(iii) $f(2) = .001, f(1) = .999$
$g(0) = g(1) = 40$
$E[g(X)] = \$40.00$

Thus, its cheapest to use strategy (i).

Problem:

7.3.1 A lottery has tickets numbered 000 to 999 which are sold for \$1 each. One ticket is selected at random and a prize of \$200 is given to any person whose ticket number is a permutation of the selected ticket number. All 1000 tickets are sold. What is the expected profit or loss to the organization running the lottery?

7.4 Means and Variances of Distributions

Its useful to know the means, $\mu = E(X)$ of probability models derived in Chapter 6.

Example: (Mean of binomial distribution)
Let $X \sim Bi(n,p)$. Find $E(X)$.

Solution:

$$\begin{aligned} \mu &= E(X) = \sum_{x=0}^{n} x\tbinom{n}{x}p^x(1-p)^{n-x} \\ &= \sum_{x=0}^{n} x\tfrac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \end{aligned}$$

When $x = 0$ the value of the expression is 0. We can therefore begin our sum at $x = 1$. Provided $x \neq 0$, we can expand $x!$ as $x(x-1)!$ (so it is important to eliminate the term when $x = 0$).

$$\text{Therefore}\mu = \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!\,[(n-1)-(x-1)]!}pp^{x-1}(1-p)^{(n-1)-(x-1)}$$

$$= np(1-p)^{n-1}\sum_{x=1}^{n}\binom{n-1}{x-1}\left(\frac{p}{1-p}\right)^{x-1}$$

Let $y = x - 1$ in the sum, to get

$$\begin{aligned} \mu &= np(1-p)^{n-1}\sum_{y=0}^{n-1}\tbinom{n-1}{y}\left(\tfrac{p}{1-p}\right)^{y} \\ &= np\,(1-p)^{n-1}\left(1+\tfrac{p}{1-p}\right)^{n-1} \quad \text{(binomial theorem)} \\ &= np\,(1-p)^{n-1}\,\tfrac{(1-p+p)^{n-1}}{(1-p)^{n-1}} = np \end{aligned}$$

Exercise: Does this result make sense? If you try something 100 times and there is a 20% chance of success each time, how many successes do you expect to get, on average?

Example: (Mean of the Poisson distribution)
Let X have a Poisson distribution where λ is the average rate of occurrence and the time interval is of length t. Find $\mu = E(X)$.

Solution:
The probability function of X is $f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$. Then $\mu = E(X) = \sum\limits_{x=0}^{\infty} x \frac{(\lambda t)^x e^{-\lambda t}}{x!}$.
As in the binomial example, we can eliminate the term when $x = 0$ and expand $x!$ as $x(x-1)$ for $x = 1, 2, \cdots, \infty$.

$$\begin{aligned} \mu &= \sum_{x=1}^{\infty} x \frac{(\lambda t)^x e^{-\lambda t}}{x(x-1)!} = \sum_{x=1}^{\infty} (\lambda t) e^{-\lambda t} \frac{(\lambda t)^{x-1}}{(x-1)!} \\ &= (\lambda t)\, e^{-\lambda t} \sum_{x=1}^{\infty} \frac{(\lambda t)^{x-1}}{(x-1)!} \end{aligned}$$

Let $y = x - 1$ in the sum.

$$\mu = (\lambda t) e^{-\lambda t} \sum_{y=0}^{\infty} \frac{(\lambda t)^y}{y!} = (\lambda t) e^{-\lambda t} e^{\lambda t} \text{ since } e^x = \sum_{y=0}^{\infty} \frac{x^y}{y!}$$

$$\text{Therefore } \mu = \lambda t.$$

Note that we used the symbol $\mu = \lambda t$ earlier in connection with the Poisson model; this was because we knew (but couldn't show until now) that $E(X) = \mu$.

Exercise: These techniques can also be used to work out the mean for the hypergeometric or negative binomial distributions. Looking back at how we proved that $\sum f(x) = 1$ shows the same method of summation used to find μ. However, in Chapter 8 we will give a simpler method of finding the means of these distributions, which are $E(X) = nr/N$ (hypergeometric) and $E(X) = k(1-p)/p$ (negative binomial).

Variability:
While an average is a useful summary of a set of observations, or of a probability distribution, it omits another important piece of information, namely the amount of variability. For example, it would be possible for car doors to be the right width, on average, and still have no doors fit properly. In the case of fitting car doors, we would also want the door widths to all be close to this correct average. We give a way of measuring the amount of variability next. You might think we could use the average difference between X and μ to indicate the amount of variation. In terms of expectation, this would be $E(X - \mu)$. However, $E(X - \mu) = E(X) - \mu$ (since μ is a constant) = 0. We soon realize that to measure variability we need a function that is the same sign for $X > \mu$ and for $X < \mu$. We now define

Definition 16 *The* ***variance*** *of a r.v* X *is* $E\left[(X-\mu)^2\right]$*, and is denoted by* σ^2 *or by Var* (X).

In words, the variance is the average square of the distance from the mean. This turns out to be a very useful measure of the variability of X.

Example: Let X be the number of heads when a fair coin is tossed 4 times. Then $X \sim Bi\left(4, \frac{1}{2}\right)$ so $\mu = np = (4)\left(\frac{1}{2}\right) = 2$. Without doing any calculations we know $\sigma^2 \leq 4$ because X is always between 0 and 4. Hence it can never be further away from μ than 2. This makes the average square of the distance from μ at most 4. The values of $f(x)$ are

x	0	1	2	3	4
$f(x)$	1/16	4/16	6/16	4/16	1/16

since $f(x) = \binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x} = \binom{4}{x}\left(\frac{1}{2}\right)^4$

The value of $Var(X)$ (i.e. σ^2) is easily found here:

$$\begin{aligned}
\sigma^2 &= E\left[(X-\mu)^2\right] = \sum_{x=0}^{4}(x-\mu)^2 f(x) \\
&= (0-2)^2\left(\tfrac{1}{16}\right) + (1-2)^2\left(\tfrac{4}{16}\right) + (2-2)^2\left(\tfrac{6}{16}\right) + (3-2)^2\left(\tfrac{4}{16}\right) + (4-2)^2\left(\tfrac{1}{16}\right) \\
&= 1
\end{aligned}$$

If we keep track of units of measurement the variance will be in peculiar units; e.g. if X is the number of heads in 4 tosses of a coin, σ^2 is in units of heads^2! We can regain the original units by taking (positive) $\sqrt{\text{variance}}$. This is called the standard deviation of X, and is denoted by σ, or as $SD(X)$.

Definition 17 *The standard deviation of a random variable X is* $\sigma = \sqrt{E\left[(X-\mu)^2\right]}$

Both variance and standard deviation are commonly used to measure variability.

The basic definition of variance is often awkward to use for mathematical calculation of σ^2, whereas the following two results are often useful:

$$\begin{aligned}
(1)\quad \sigma^2 &= E\left(X^2\right) - \mu^2 \\
(2)\quad \sigma^2 &= E\left[X(X-1)\right] + \mu - \mu^2
\end{aligned}$$

Proof:

(1) Using properties of expectation,

$$\begin{aligned}
\sigma^2 &= E\left[(X-\mu)^2\right] = E\left[X^2 - 2\mu X + \mu^2\right] \\
&= E\left(X^2\right) - 2\mu E(X) + \mu^2 \quad \text{(since } \mu \text{ is constant)} \\
&= E\left(X^2\right) - 2\mu^2 + \mu^2 \quad \text{(Therefore } E(X) = \mu\text{)} \\
&= E\left(X^2\right) - \mu^2
\end{aligned}$$

(2) $$\text{since} X^2 = X(X-1) + X$$

$$\begin{aligned} \text{Therefore} \quad E\left(X^2\right) - \mu^2 &= E\left[X\left(X-1\right) + X\right] - \mu^2 \\ &= E\left[X(X-1)\right] + E(X) - \mu^2 \\ &= E\left[X(X-1)\right] + \mu - \mu^2 \end{aligned}$$

Formula (2) is most often used when there is an $x!$ term in the denominator of $f(x)$. Otherwise, formula (1) is generally easier to use.

Example: (Variance of binomial distribution)
Let $X \sim Bi(n,p)$. Find Var (X).

Solution: The probability function for the binomial is

$$f(x) = \frac{n!}{x!(n-x)!}\, p^x \left(1-p\right)^{n-x}$$

so we'll use formula (2) above,

$$E\left[X(X-1)\right] = \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} \; p^x (1-p)^{n-x}$$

If $x = 0$ or $x = 1$ the value of the term is 0, so we can begin summing at $x = 2$. For $x \neq 0$ or 1, we can expand the $x!$ as $x(x-1)(x-2)!$

$$\text{Therefore } E\left[X\left(X-1\right)\right] = \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x}$$

Now re-group to fit the binomial theorem, since that was the summation technique used to show $\sum f(x) = 1$ and to derive $\mu = np$.

$$E\left[X(X-1)\right] = \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!\left[(n-2)-(x-2)\right]!} p^2 p^{x-2} (1-p)^{(n-2)-(x-2)}$$

$$= n(n-1)p^2(1-p)^{n-2} \sum_{x=2}^{n} \binom{n-2}{x-2} \left(\frac{p}{1-p}\right)^{x-2}$$

Let $y = x - 2$ in the sum, giving

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

$$\begin{aligned} &= \quad n(n-1)p^2(1-p)^{n-2} \left(1 + \tfrac{p}{1-p}\right)^{n-2} \\ &= \quad n(n-1)p^2(1-p)^{n-2} \tfrac{(1-p+p)^{n-2}}{(1-p)^{n-2}} = n(n-1)p^2 \end{aligned}$$

Then

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

Remember that the variance of a binomial distribution is $np(1-p)$, since we'll be using it later in the course.

Example: (Variance of Poisson distribution) Find the variance of the Poisson distribution.
Solution: The probability function of the Poisson is

$$f(x)=\frac{\mu^x e^{-\mu}}{x!}$$

from which we obtain

$$\begin{aligned}E\left[X(X-1)\right] &= \sum_{x=0}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x!}\\ &= \sum_{x=2}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x(x-1)(x-2)!},\ \text{setting the lower limit to 2 and expanding } x!\\ &= \mu^2 e^{-\mu}\sum_{x=2}^{\infty}\frac{\mu^{x-2}}{(x-2)!}\end{aligned}$$

Let $y=x-2$ in the sum, giving

$$\begin{aligned}E\left[X(X-1)\right] &= \mu^2 e^{-\mu}\sum_{y=0}^{\infty}\frac{\mu^y}{y!}=\mu^2e^{-\mu}e^{\mu}=\mu^2 \text{ so}\\ \sigma^2 &= E\left[X(X-1)\right]+\mu-\mu^2\\ &= \mu^2+\mu-\mu^2=\mu\end{aligned}$$

(*For the Poisson distribution, the variance equals the mean.*)

Properties of Mean and Variance
If a and b are constants and $Y=aX+b$, then

$$\mu_Y=a\mu_X+b \text{ and } \sigma_Y^2=a^2\sigma_X^2$$

(where μ_X and σ_X^2 are the mean and variance of X and μ_Y and σ_Y^2 are the mean and variance of Y).
Proof:
We already showed that $E(aX+b)=aE(X)+b$.

i.e. $\mu_Y = a\mu_X + b$, and then

$$\begin{aligned}\sigma_Y^2 &= E\left[(Y-\mu_Y)^2\right] = E\left\{[(aX+b)-(a\mu_X+b)]^2\right\}\\ &= E\left[(aX-a\mu_X)^2\right] = E\left[a^2(X-\mu_X)^2\right]\\ &= a^2E\left[(X-\mu_X)^2\right] = a^2\sigma_X^2\end{aligned}$$

This result is to be expected. Adding a constant, b, to all values of X has no effect on the amount of variability. So it makes sense that Var $(aX+b)$ doesn't depend on the value of b. Also since variance is in squared units, multiplication by a constant results in multiplying the variance by the constant squared. A simple way to relate to this result is to consider a random variable X which represents a temperature in degrees Celsius (even though this is a continuous random variable which we don't study until Chapter 9). Now let Y be the corresponding temperature in degrees Fahrenheit. We know that

$$Y = \frac{9}{5}X + 32$$

and it is clear if we think about it that $\mu_Y = (\frac{9}{5})\mu_X + 32$ and that $\sigma_Y^2 = (\frac{9}{5})^2\sigma_X^2$.

Problems:

7.4.1 An airline knows that there is a 97% chance a passenger for a certain flight will show up, and assumes passengers arrive independently of each other. Tickets cost \$100, but if a passenger shows up and can't be carried on the flight the airline has to refund the \$100 and pay a penalty of \$400 to each such passenger. How many tickets should they sell for a plane with 120 seats to maximize their expected ticket revenues after paying any penalty charges? Assume ticket holders who don't show up get a full refund for their unused ticket.

7.4.2 A typist typing at a constant speed of 60 words per minute makes a mistake in any particular word with probability .04, independently from word to word. Each incorrect word must be corrected; a task which takes 15 seconds per word.

(a) Find the mean and variance of the time (in seconds) taken to finish a 450 word passage.

(b) Would it be less time consuming, on average, to type at 45 words per minute if this reduces the probability of an error to .02?

7.5 Moment Generating Functions

We have now seen two functions which characterize a distribution, the probability function and the cumulative distribution function. There is a third type of function, the *moment generating function*, which uniquely determines a distribution. The moment generating function is closely related to other transforms used in mathematics, the Laplace and Fourier transforms.

Definition 18 *Consider a discrete random variable X with probability function $f(x)$. The moment generating function (m.g.f.) of X is defined as*

$$M(t) = E(e^{tX}) = \sum_x e^{tx} f(x),$$

assuming that this sum exists.

The *moments* of a random variable X are the expectations of the functions X^r for $r = 1, 2, \ldots$. The expectation $E(X^r)$ is called r^{th} moment of X. The mean $\mu = E(X)$ is therefore the first moment, $E(X^2)$ the second and so on. It is often easy to find the moments of a probability distribution mathematically by using the moment generating function. This often gives easier derivations of means and variances than the direct summation methods in the preceding section. The following theorem gives a useful property of m.g.f.'s.

Theorem 19 *Let the random variable X have m.g.f. $M(t)$. Then*

$$E(X^r) = M^{(r)}(0) \quad r = 1, 2, \ldots$$

where $M^{(r)}(0)$ stands for $d^r M(t)/dt^r$ evaluated at $t = 0$.

Proof:
$M(t) = \sum\limits_x e^{tx} f(x)$ and if the sum converges, then

$$\begin{aligned} M^{(r)}(t) &= \tfrac{d}{dt^r} \sum_x e^{tx} f(x) \\ &= \sum_x \tfrac{d}{dt^r}(e^{tx}) f(x) \\ &= \sum_x x^r e^{tx} f(x) \end{aligned}$$

Therefore $M^{(r)}(0) = \sum_x x^r f(x) = E(X^r)$, as stated.
This sometimes gives a simple way to find the moments for a distribution.

Example 1. Suppose X has a Binomial(n, p)distribution. Then its moment generating function is

$$\begin{aligned} M(t) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

Therefore

$$
\begin{aligned}
M'(t) &= npe^t(pe^t + 1 - p)^{n-1} \\
M''(t) &= npe^t(pe^t + 1 - p)^{n-1} + n(n-1)p^2e^{2t}(pe^t + 1 - p)^{n-2}
\end{aligned}
$$

and so

$$
\begin{aligned}
E(X) &= M'(0) = np, \\
E(X^2) &= M"(0) = np + n(n-1)p^2 \\
\text{Var}(X) &= E(X^2) - E(X)^2 = np(1-p)
\end{aligned}
$$

Exercise. Poisson distribution

Show that the Poisson distribution with probability function

$$f(x) = e^{-\mu}\mu^x/x! \quad x = 0, 1, 2, \ldots$$

has m.g.f. $M(t) = e^{-\mu+\mu e^t}$. Then show that $E(X) = \mu$ and $Var(X) = \mu$.

The m.g.f. also uniquely identifies a distribution in the sense that two different distributions cannot have the same m.g.f. This result is often used to find the distribution of a random variable. For example if I can show somehow that the moment generating function of a random variable X is

$$e^{2(e^t-1)}$$

then I know, from the above exercise that the random variable must have a Poisson(2) distribution. Moment generating functions are often used to identify a given distribution. If two random variables have the same moment generating function, they have the same distribution (so the same probability function, cumulative distribution function, moments, etc.). Of course the moment generating functions must match for all values of t, in other words they agree as *functions*, not just at a few points. Moment generating functions can also be used to determine that a sequence of distributions gets closer and closer to some limiting distribution. To show this (aleit a bit loosely), suppose that a sequence of probability functions $f_n(x)$ have corresponding moment generating functions

$$M_n(t) = \sum_x e^{tx} f_n(x)$$

Suppose moreover that the probability functions $f_n(x)$ converge to another probability function $f(x)$ pointwise in x as $n \to \infty$. This is what we mean by convergence of discrete distributions. Then since

$$f_n(x) \to f(x) \text{ as } n \to \infty \text{ for each } x, \tag{7.1}$$

$$\sum_x e^{tx} f_n(x) \to \sum_x e^{tx} f(x) \text{ as } n \to \infty \text{ for each } t \tag{7.2}$$

which says that $M_n(t)$ converges to $M(t)$ the moment generating function of the limiting distribution. It shouldn't be too surprising that a very useful converse to this result also holds. (This is strictly an aside and may be of interest only to those with a thing for infinite series, but is it always true that becuase the individual terms in a series converge as in (7.1) does this guarantee that the sum of the series also converges (7.2)?)

Suppose conversely that X_n has moment generating function $M_n(t)$ and $M_n(t) \to M(t)$ for each t such that $M(t) < \infty$. For example we saw in Chapter 6 that a Binomial(n,p) distribution with very large n and very small p is close to a Poisson distribution with parameter $\mu = np$. Consider the moment generating function of such a binomial random variable

$$\begin{aligned} M(t) &= (pe^t + 1 - p)^n \\ &= \{1 + p(e^t - 1)\}^n \\ &= \{1 + \frac{\mu}{n}(e^t - 1)\}^n \end{aligned} \tag{7.3}$$

Now take the limit of this expression as $n \to \infty$. Since in general

$$(1 + \frac{c}{n})^n \to e^c$$

the limit of (7.3) as $n \to \infty$ is

$$e^{\mu(e^t-1)} = e^{-\mu+\mu e^t}$$

and this is the moment generating function of a Poisson distribution with parameter μ. This shows a little more formally than we did earlier that the binomial(n,p) distribution with (small) $p = \mu/n$ approaches the Poisson(μ) distribution as $n \to \infty$.

7.6 Problems on Chapter 7

7.1 Let X have probability function $f(x) = \begin{cases} \frac{1}{2x} \text{ for } x = 2, 3, 4, 5, \text{ or } 6 \\ 11/40 \text{ for } x = 1 \end{cases}$ Find the mean and variance for X.

7.2 A game is played where a fair coin is tossed until the first tail occurs. The probability x tosses will be needed is $f(x) = .5^x;\ \ x = 1, 2, 3, \cdots$. You win \$$2^x$ if x tosses are needed for $x = 1, 2, 3, 4, 5$ but lose \$256 if $x > 5$. Determine your expected winnings.

7.3 Diagnostic tests. Consider diagnostic tests like those discussed above in the example of Section 7.3 and in Problem 15 for Chapter 4. Assume that for a randomly selected person, $P(D) = .02$, $P(R|D) = 1$, $Pr(R|\overline{D}) = .05$, so that the inexpensive test only gives false positive, and not false negative, results.

Suppose that this inexpensive test costs \$10. If a person tests positive then they are also given a more expensive test, costing \$100, which correctly identifies all persons with the disease. What is the expected cost per person if a population is tested for the disease using the inexpensive test followed, if necessary, by the expensive test?

7.4 Diagnostic tests II. Two percent of the population has a certain condition for which there are two diagnostic tests. Test A, which costs \$1 per person, gives positive results for 80% of persons with the condition and for 5% of persons without the condition. Test B, which costs \$100 per person, gives positive results for all persons with the condition and negative results for all persons without it.

(a) Suppose that test B is given to 150 persons, at a cost of \$15,000. How many cases of the condition would one expect to detect?

(b) Suppose that 2000 persons are given test A, and then only those who test positive are given test B. Show that the expected cost is \$15,000 but that the expected number of cases detected is much larger than in part (a).

7.5 The probability that a roulette wheel stops on a red number is 18/37. For each bet on "red" you win the amount bet if the wheel stops on a red number, and lose your money if it does not.

(a) If you bet \$1 on each of 10 consecutive plays, what is your expected winnings? What is your expected winnings if you bet \$10 on a single play?

(b) For each of the two cases in part (a), calculate the probability that you made a profit (that is, your "winnings" are positive, not negative).

7.6 Slot machines. Consider the slot machine discussed above in Problem 16 for Chapter 4. Suppose that the number of each type of symbol on wheels 1, 2 and 3 is as given below:

	Wheel		
Symbols	1	2	3
Flower	2	6	2
Dog	4	3	3
House	4	1	5

If all three wheels stop on a flower, you win \$20 for a \$1 bet. If all three wheels stop on a dog, you win \$10, and if all three stop on a house, you win \$5. Otherwise you win nothing.

Find your expected winnings per dollar spent.

7.7 Suppose that n people take a blood test for a disease, where each person has probability p of having the disease, independent of other persons. To save time and money, blood samples from k people are pooled and analyzed together. If none of the k persons has the disease then the test will be negative, but otherwise it will be positive. If the pooled test is positive then each of the k persons is tested separately (so $k+1$ tests are done in that case).

(a) Let X be the number of tests required for a group of k people. Show that

$$E(X) = k + 1 - k(1-p)^k.$$

(b) What is the expected number of tests required for n/k groups of k people each? If $p = .01$, evaluate this for the cases $k = 1, 5, 10$.

(c) Show that if p is small, the expected number of tests in part (b) is approximately $n(kp + k^{-1})$, and is minimized for $k \doteq p^{-1/2}$.

7.8 A manufacturer of car radios ships them to retailers in cartons of n radios. The profit per radio is \$59.50, less shipping cost of \$25 per carton, so the profit is \$ $(59.5n - 25)$ per carton. To promote sales by assuring high quality, the manufacturer promises to pay the retailer $\$200X^2$ if X radios in the carton are defective. (The retailer is then responsible for repairing any defective radios.) Suppose radios are produced independently and that 5% of radios are defective. How many radios should be packed per carton to maximize expected <u>net</u> profit per carton?

7.9 Let X have a geometric distribution with probability function
$f(x) = p(1-p)^x;\;\; x = 0, 1, 2, ...$

(a) Calculate the m.g.f. $M(t) = E\left(e^{tX}\right)$, where t is a parameter.

(b) Find the mean and variance of X.

(c) Use your result in (b) to show that if p is the probability of "success" (S) in a sequence of Bernoulli trials, then the expected number of trials until the first S occurs is $1/p$. Explain why this is "obvious".

7.10 **Analysis of Algorithms: Quicksort.** Suppose we have a set S of distinct numbers and we wish to sort them from smallest to largest. The quicksort algorithm works as follows: When $n = 2$ it just compares the numbers and puts the smallest one first. For $n > 2$ it starts by choosing a random "pivot" number from the n numbers. It then compares each of the other $n-1$ numbers with the pivot and divides them into groups S_1 (numbers smaller than the pivot) and $\bar{S}_1$ (numbers bigger than the pivot). It then does the same thing with S_1 and $\bar{S}_1$ as it did with S, and repeats this recursively until the numbers are all sorted. (Try this out with, say $n = 10$ numbers to see how it works.) In computer science it is common to analyze such algorithms by finding the expected number of comparisons (or other operations) needed to sort a list. Thus, let

$$C_n = \text{expected number of comparisons for lists of length } n$$

(a) Show that if X is the number of comparisons needed,

$$C_n = \sum_{i=1}^{n} E(X| \text{ initial pivot is } i\text{th smallest number}) \left(\frac{1}{n}\right)$$

(b) Show that

$$E(X|\text{initial pivot is } i\text{th smallest number}) = n - 1 + C_{i-1} + C_{n-i}$$

and thus that C_n satisfies the recursion (note $C_0 = C_1 = 0$)

$$C_n = n - 1 + \frac{2}{n}\sum_{k=1}^{n-1} C_k \quad n = 2, 3, \ldots$$

(c) Show that

$$(n+1)C_{n+1} = 2n + (n+2)C_n \quad n = 1, 2, \ldots$$

(d) (Harder) Use the result of part (c) to show that for large n,

$$\frac{C_{n+1}}{n+1} \sim 2 \text{ log } (n+1)$$

(Note: $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$) This proves a result from computer science which says that for Quicksort, $C_n \sim O(n \; log \; n)$.

7.11 Find the distributions that corresponds to the following moment-generating functions:
(a) $M(t) = \frac{1}{3e^{-t}-2}$
(b) $M(t) = e^{2(e^t-1)}$

7.12 Find the moment generating function of the discrete uniform distribution X on $\{a, a+1, ..., b\}$;

$$P(X = x) = \frac{1}{b-a+1}, \text{for } x = a, a+1, ..., b.$$

What do you get in the special case $a = b$ and in the case $b = a + 1$? Use the moment generating function in these two cases to confirm the expected value and the variance of X.

7.13 Let X be a random variable taking values in the set $\{0, 1, 2\}$ with moments $E(X) = 1$, $E(X^2) = 3/2$.

(a) Find the moment generating function of X

(b) Find the first six moments of X

(c) Find $P(X = i), i = 0, 1, 2$.

(d) Show that any probability distribution on $\{0, 1, 2\}$ is completely determined by its first two moments.

8. Discrete Multivariate Distributions

8.1 Basic Terminology and Techniques

Many problems involve more than a single random variable. When there are multiple random variables associated with an experiment or process we usually denote them as $X, Y, \ldots$ or as $X_1, X_2, \ldots$. For example, your final mark in a course might involve X_1=your assignment mark, X_2=your midterm test mark, and X_3 =your exam mark. We need to extend the ideas introduced for single variables to deal with multivariate problems. Here we only consider discrete multivariate problems, though continuous multivariate variables are also common in daily life (e.g. consider a person's height X and weight Y, or X_1 =the return from Stock 1, X_2 =return from stock 2). To introduce the ideas in a simple setting, we'll first consider an example in which there are only a few possible values of the variables. Later we'll apply these concepts to more complex examples. The ideas themselves are simple even though some applications can involve fairly messy algebra.

Joint Probability Functions:

First, suppose there are two r.v.'s X and Y, and define the function

$$\begin{aligned} f(x, y) &= P(X = x \text{ and } Y = y) \\ &= P(X = x, Y = y). \end{aligned}$$

We call $f(x, y)$ the joint probability function of (X, Y). In general,

$$f(x_1, x_2, \cdots, x_n) = P(X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \ldots \text{ and } X_n = x_n)$$

if there are n random variables $X_1, \ldots, X_n$.

The properties of a joint probability function are similar to those for a single variable; for two r.v.'s we have $f(x, y) \geq 0$ for all (x, y) and

$$\sum_{\text{all(x,y)}} f(x, y) = 1.$$

Example: Consider the following numerical example, where we show $f(x,y)$ in a table.

$f(x,y)$		$x=0$	$x=1$	$x=2$
y	1	.1	.2	.3
	2	.2	.1	.1

for example $f(0,2) = P(X = 0 \text{ and } Y = 2) = 0.2$. We can check that $f(x,y)$ is a proper joint probability function since $f(x,y) \geq 0$ for all 6 combinations of (x,y) and the sum of these 6 probabilities is 1. When there are only a few values for X and Y it is often easier to tabulate $f(x,y)$ than to find a formula for it. We'll use this example below to illustrate other definitions for multivariate distributions, but first we give a short example where we need to find $f(x,y)$.

Example: Suppose a fair coin is tossed 3 times. Define the r.v.'s X = number of Heads and $Y = 1(0)$ if $H(T)$ occurs on the first toss. Find the joint probability function for (X,Y).

Solution: First we should note the range for (X,Y), which is the set of possible values (x,y) which can occur. Clearly X can be 0, 1, 2, or 3 and Y can be 0 or 1, but we'll see that not all 8 combinations (x,y) are possible.
We can find $f(x,y) = P(X = x, Y = y)$ by just writing down the sample space
$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$ that we have used before for this process.
Then simple counting gives $f(x,y)$ as shown in the following table:

$f(x,y)$		$x=0$	$x=1$	$x=2$	$x=3$
y	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
	1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

For example, $(X,Y) = (0,0)$ if and only if the outcome is TTT; $(X,Y) = (1,0)$ iff the outcome is either THT or TTH.
Note that the range or joint p.f. for (X,Y) is a little awkward to write down here in formulas, so we just use the table.

Marginal Distributions:

We may be given a joint probability function involving more variables than we're interested in using. How can we eliminate any which are not of interest? Look at the first example above. If we're only interested in X, and don't care what value Y takes, we can see that

$$P(X = 0) = P(X = 0, Y = 1) + P(X = 0, Y = 2),$$

so $P(X = 0) = f(0, 1) + f(0, 2) = 0.3$. Similarly

$$P(X = 1) = f(1, 1) + f(1, 2) = .3 \text{ and}$$
$$P(X = 2) = f(2, 1) + f(2, 2) = .4$$

The distribution of X obtained in this way from the joint distribution is called the marginal probability function of X:

x	0	1	2
$f(x)$	.3	.3	.4

In the same way, if we were only interested in Y, we obtain

$$P(Y = 1) = f(0, 1) + f(1, 1) + f(2, 1) = .6$$

since X can be 0, 1, or 2 when $Y = 1$. The marginal probability function of Y would be:

y	1	2
$f(y)$	.6	.4

Our notation for marginal probability functions is still inadequate. What is $f(1)$? As soon as we substitute a number for x or y, we don't know which variable we're referring to. For this reason, we generally put a subscript on the f to indicate whether it is the marginal probability function for the first or second variable. So $f_1(1)$ would be $P(X = 1) = .3$, while $f_2(1)$ would be $P(Y = 1) = 0.6$. An alternative notation that you may see is $f_X(x)$ and $f_Y(y)$.
In general, to find $f_1(x)$ we add over all values of y where $X = x$, and to find $f_2(y)$ we add over all values of x with $Y = y$. Then

$$f_1(x) = \sum_{\text{all } y} f(x, y) \text{ and}$$
$$f_2(y) = \sum_{\text{all } x} f(x, y).$$

This reasoning can be extended beyond two variables. For example, with three variables (X_1, X_2, X_3),

$$f_1(x_1) = \sum_{\text{all } (x_2,x_3)} f(x_1, x_2, x_3) \text{ and}$$

$$f_{1,3}(x_1, x_3) = \sum_{\text{all } x_2} f(x_1, x_2, x_3) = P(X_1 = x_1, X_3 = x_3)$$

where $f_{1,3}(x_1, x_3)$ is the marginal joint distribution of (X_1, X_3).

Independent Random Variables:

For events A and B, we have defined A and B to be independent if and only if $P(AB) = P(A)\ P(B)$. This definition can be extended to random variables (X, Y): two random variables are independent if their joint probability function is the product of the marginal probability functions.

Definition 20 *X and Y are **independent** random variables iff $f(x, y) = f_1(x)f_2(y)$ for all values (x, y)*

Definition 21 *In general, $X_1, X_2, \cdots, X_n$ are independent random variables iff*

$$f(x_1, x_2, \cdots, x_n) = f_1(x_1)f_2(x_2). \cdots .f_n(x_n) \textit{ for all } x_1, x_2, ...x_n$$

In our first example X and Y are not independent since $f_1(x)f_2(y) \neq f(x, y)$ for any of the 6 combinations of (x, y) values; e.g., $f(1, 1) = .2$ but $f_1(1)f_2(1) = (0.3)\ (0.6) \neq 0.2$. Be careful applying this definition. You can only conclude that X and Y are independent after checking all (x, y) combinations. Even a single case where $f_1(x)f_2(y) \neq f(x, y)$ makes X and Y dependent.

Conditional Probability Functions:

Again we can extend a definition from events to random variables. For events A and B, recall that $P(A|B) = \frac{P(AB)}{P(B)}$. Since $P(X = x|Y = y) = P(X = x, Y = y)/P(Y = y)$, we make the following definition.

Definition 22 *The conditional probability function of X given $Y = y$ is $f(x|y) = \frac{f(x,y)}{f_2(y)}$.*
Similarly, $f(y|x) = \frac{f(x,y)}{f_1(x)}$ (provided, of course, the denominator is not zero).

In our first example let us find $f(x|Y=1)$.

$$f(x|Y=1)=\frac{f(x,1)}{f_2(1)}.$$

This gives:

x	0	1	2
$f(x\|Y=1)$	$\frac{.1}{.6}=\frac{1}{6}$	$\frac{.2}{.6}=\frac{1}{3}$	$\frac{.3}{.6}=\frac{1}{2}$

As you would expect, marginal and conditional probability functions are probability functions in that they are always ≥ 0 and their sum is 1.

Functions of Variables:

In an example earlier, your final mark in a course might be a function of the 3 variables X_1, X_2, X_3 - assignment, midterm, and exam marks[8]. Indeed, we often encounter problems where we need to find the probability distribution of a function of two or more r.v.'s. The most general method for finding the probability function for some function of random variables X and Y involves looking at every combination (x,y) to see what value the function takes. For example, if we let $U=2(Y-X)$ in our example, the possible values of U are seen by looking at the value of $U=2(y-x)$ for each (x,y) in the range of (X,Y).

		x		
	u	0	1	2
y	1	2	0	-2
	2	4	2	0

$$\begin{aligned}
\text{Then}\quad P(U=-2) &= P(X=2 \text{ and } Y=1)=f(2,1)=.3\\
P(U=0) &= P(X=1 \text{ and } Y=1, \text{ or } X=2 \text{ and } Y=2)\\
&= f(1,1)+f(2,2)=.3\\
P(U=2) &= f(0,1)+f(1,2)=.2\\
P(U=4) &= f(0,2)=.2
\end{aligned}$$

The probability function of U is thus

[8]"Don't worry about your marks. Just make sure that you keep up with the work and that you don't have to repeat a year. It s not necessary to have good marks in everything" Albert Einstein in letter to his son, 1916.

u	-2	0	2	4
$f(u)$	.3	.3	.2	.2

For some functions it is possible to approach the problem more systematically. One of the most common functions of this type is the total. Let $T = X + Y$. This gives:

		x		
t		0	1	2
	1	1	2	3
y				
	2	2	3	4

Then $P(T = 3) = f(1, 2) + f(2, 1) = .4$, for example. Continuing in this way, we get

t	1	2	3	4
$f(t)$	.1	.4	.4	.1

(We are being a little sloppy with our notation by using "f" for both $f(t)$ and $f(x, y)$. No confusion arises here, but better notation would be to write $f_T(t)$ for $P(T = t)$.) In fact, to find $P(T = t)$ we are simply adding the probabilities for all (x, y) combinations with $x + y = t$. This could be written as:

$$f(t) = \sum_{\substack{\text{all }(x,y)\\ \text{with } x+y=t}} f(x, y).$$

However, if $x + y = t$, then $y = t - x$. To systematically pick out the right combinations of (x, y), all we really need to do is sum over values of x and then substitute $t - x$ for y. Then,

$$f(t) = \sum_{\text{all } x} f(x, t - x) = \sum_{\text{all } x} P(X = x, Y = t - x)$$

So $P(T = 3)$ would be

$$P(T = 3) = \sum_{\text{all } x} f(x, 3 - x) = f(0, 3) + f(1, 2) + f(2, 1) = 0.4.$$

(note $f(0, 3) = 0$ since Y can't be 3.)

We can summarize the method of finding the probability function for a function $U = g(X, Y)$ of two random variables X and Y as follows:

Let $f(x, y) = P(X = x, Y = y)$ be the probability function for (X, Y). Then the probability function for U is

$$f_U(u) = P(U = u) = \sum_{\substack{\text{all}(x,y):\\ g(x,y)=u}} f(x, y)$$

This can also be extended to functions of three or more r.v.'s $U = g(X_1, X_2, \ldots, X_n)$:

$$f_U(u) = P(U = u) = \sum_{\substack{(x_1,\ldots,x_n): \\ g(x_1,\ldots,x_n)=u}} f(x_1, \ldots, x_n).$$

(**Note:** Do not get confused between the functions f and g in the above: $f(x, y)$ is the joint probability function of the r.v.'s X, Y whereas $U = g(X, Y)$ defines the "new" random variable that is a function of X and Y, and whose distribution we want to find.)

Example: Let X and Y be independent random variables having Poisson distributions with averages (means) of μ_1 and μ_2 respectively. Let $T = X + Y$. Find its probability function, $f(t)$.

Solution: We first need to find $f(x, y)$. Since X and Y are independent we know

$$f(x, y) = f_1(x) f_2(y)$$

Using the Poisson probability function,

$$f(x, y) = \frac{\mu_1^x e^{-\mu_1}}{x!} \frac{\mu_2^y e^{-u_2}}{y!}$$

where x and y can equal 0, 1, 2, ... Now,

$$P(T = t) = P(X + Y = t) = \sum_{\text{all } x} P(X = x, Y = t - x).$$

Then

$$\begin{aligned} f(t) &= \sum_{\text{all } x} f(x, t - x) \\ &= \sum_{x=0}^{t} \frac{\mu_1^x e^{-\mu_1}}{x!} \frac{\mu_2^{t-x} e^{-\mu_2}}{(t - x)!} \end{aligned}$$

To evaluate this sum, factor out constant terms and try to regroup in some form which can be evaluated by one of our summation techniques.

$$f(t) = \mu_2^t e^{-(\mu_1+\mu_2)} \sum_{x=0}^{t} \frac{1}{x!(t - x)!} \left(\frac{\mu_1}{\mu_2}\right)^x$$

If we had a $t!$ on the top inside the $\sum\limits_{x=0}^{t}$, the sum would be of the form $\sum\limits_{x=0}^{t}\binom{t}{x}\left(\frac{\mu_1}{\mu_2}\right)^x$. This is the right hand side of the binomial theorem. Multiply top and bottom by $t!$ to get:

$$\begin{aligned} f(t) &= \frac{\mu_2^t e^{-(\mu_1+\mu_2)}}{t!}\sum_{x=0}^{t}\binom{t}{x}\left(\frac{\mu_1}{\mu_2}\right)^x \\ &= \frac{\mu_2^t e^{-(\mu_1+\mu_2)}}{t!}(1+\frac{\mu_1}{\mu_2})^t \text{ by the binomial theorem.} \end{aligned}$$

Take a common denominator of μ_2 to get

$$f(t) = \frac{\mu_2^t\ e^{-(\mu_1+\mu_2)}}{t!}\frac{(\mu_1+\mu_2)^t}{\mu_2^t} = \frac{(\mu_1+\mu_2)^t}{t!}e^{-(\mu_1+\mu_2)},\ \text{for}\ \ t = 0, 1, 2, \cdots$$

Note that we have just shown that the sum of 2 independent Poisson random variables also has a Poisson distribution.

Example: Three sprinters, A, B and C, compete against each other in 10 independent 100 m. races. The probabilities of winning any single race are .5 for A, .4 for B, and .1 for C. Let X_1, X_2 and X_3 be the number of races A, B and C win.

(a) Find the joint probability function, $f(x_1, x_2, x_3)$

(b) Find the marginal probability function, $f_1(x_1)$

(c) Find the conditional probability function, $f(x_2|x_1)$

(d) Are X_1 and X_2 independent? Why?

(e) Let $T = X_1 + X_2$. Find its probability function, $f(t)$.

Solution: Before starting, note that $x_1 + x_2 + x_3 = 10$ since there are 10 races in all. We really only have two variables since $x_3 = 10 - x_1 - x_2$. However it is convenient to use x_3 to save writing and preserve symmetry.

(a) The reasoning will be similar to the way we found the binomial distribution in Chapter 6 except that there are now 3 types of outcome. There are $\frac{10!}{x_1!x_2!x_3!}$ different outcomes (i.e. results for races 1 to 10) in which there are x_1 wins by A, x_2 by B, and x_3 by C. Each of these arrangements has a probability of (.5) multiplied x_1 times, (.4) x_2 times, and (.1) x_3 times in some order;

i.e., $(.5)^{x_1}(.4)^{x_2}(.1)^{x_3}$

Therefore
$$f(x_1,x_2,x_3)=\frac{10!}{x_1!x_2!x_3!}(.5)^{x_1}(.4)^{x_2}(.1)^{x_3}$$

The range for $f(x_1,x_2,x_3)$ is triples (x_1,x_2,x_3) where each x_i is an integer between 0 and 10, and where $x_1+x_2+x_3=10$.

(b) It would also be acceptable to drop x_3 as a variable and write down the probability function for X_1,X_2 only; this is

$$f(x_1,x_2) \;=\; \frac{10!}{x_1!x_2!(10-x_1-x_2)!}\;(.5)^{x_1}\;(.4)^{x_2}\;(.1)^{10-x_1-x_2},$$

because of the fact that X_3 must equal $10-X_1-X_2$. For this probability function $x_1=0,1,\cdots,10;\; x_2=0,1,\cdots,10$ and $x_1+x_2\le 10$. This simplifies finding $f_1(x_1)$ a little . We now have $f_1(x_1)=\sum\limits_{x_2} f(x_1,x_2)$. The limits of summation need care: x_2 could be as small as 0, but since $x_1+x_2\le 10$, we also require $x_2\le 10-x_1$. (For example if $x_1=7$ then B can win $0,1,2$, or 3 races.) Thus,

$$\begin{aligned}f_1(x_1)&=\sum_{x_2=0}^{10-x_1}\frac{10!}{x_1!x_2!(10-x_1-x_2)!}(.5)^{x_1}(.4)^{x_2}(.1)^{10-x_1-x_2}\\ &==\frac{10!}{x_1!}(.5)^{x_1}(.1)^{10-x_1}\sum_{x_2=0}^{10-x_1}\frac{1}{x_2!(10-x_1-x_2)!}\left(\frac{.4}{.1}\right)^{x_2}\end{aligned}$$

(**Hint:** In $\binom{n}{r}=\frac{n!}{r!(n-r)!}$ the 2 terms in the denominator add to the term in the numerator, if we ignore the ! sign.) Multiply top and bottom by $[x_2+(10-x_1-x_2)]!=(10-x_1)!$ This gives

$$\begin{aligned}f_1(x_1)&=\frac{10!}{x_1!(10-x_1)!}(0.5)^{x_1}(0.1)^{10-x_1}\sum_{x_2=0}^{10-x_1}\binom{10-x_1}{x_2}\left(\frac{0.4}{0.1}\right)^{x_2}\\ &=\binom{10}{x_1}(0.5)^{x_1}(0.1)^{10-x_1}(1+\frac{.4}{.1})^{10-x_1}\quad\text{(again using the binomial theorem)}\\ &=\binom{10}{x_1}(0.5)^{x_1}(0.1)^{10-x_1}\frac{(0.1+0.4)^{10-x_1}}{(0.1)^{10-x_1}}=\binom{10}{x_1}(0.5)^{x_1}(0.5)^{10-x_1}\end{aligned}$$

Here $f_1(x_1)$ is defined for $x_1=0,1,2,\cdots,10$.

Note: While this derivation is included as an example of how to find marginal distributions by summing a joint probability function, there is a much simpler method for this problem. Note that each race is either won by A ("success") or it is not won by A ("failure"). Since the races are independent and X_1 is now just the number of "success" outcomes, X_1 must have a binomial distribution, with $n = 10$ and $p = .5$.

Hence $f_1(x_1) = \binom{10}{x_1}(.5)^{x_1}(.5)^{10-x_1}$; for $x_1 = 0, 1, \ldots, 10$, as above.

(c) Remember that $f(x_2|x_1) = P(X_2 = x_2|X_1 = x_1)$, so that

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{\frac{10!}{x_1!x_2!(10-x_1-x_2)!}(.5)^{x_1}(.4)^{x_2}(.1)^{10-x_1-x_2}}{\frac{10!}{x_1!(10-x_1)!}(.5)^{x_1}(.5)^{10-x_1}}$$
$$= \frac{(10-x_1)!}{x_2!\,(10-x_1-x_2)!}\frac{(.4)^{x_2}\,(.1)^{10-x_1-x_2}}{(.5)^{x_2}\,(.5)^{10-x_1-x_2}} = \binom{10-x_1}{x_2}\left(\frac{4}{5}\right)^{x_2}\left(\frac{1}{5}\right)^{10-x_1-x_2}$$

For any given value of x_1, x_2 ranges through $0, 1, \ldots, (10 - x_1)$. (So the range of X_2 depends on the value x_1, which makes sense: if B wins x_1 races then the most A can win is $10 - x_1$.)

Note: As in (b), this result can be obtained more simply by general reasoning. Once we are given that A wins x_1 races, the remaining $(10 - x_1)$ races are all won by either B or C. For these races, B wins $\frac{4}{5}$ of the time and C $\frac{1}{5}$ of the time, because $P(B \text{ wins}) = .4$ and $P(C) = .1$; i.e., B wins 4 times as often as C. More formally

$$P(B \text{ wins } |B \text{ or} C \text{ wins}) = 0.8.$$

$$\text{Therefore } f(x_2|x_1) = \binom{10-x_1}{x_2}\left(\frac{4}{5}\right)^{x_2}\left(\frac{1}{5}\right)^{10-x_1-x_2}$$

from the binomial distribution.

(d) X_1 and X_2 are clearly not independent since the more races A wins, the fewer races there are for B to win. More formally,

$$f_1(x_1)f_2(x_2) = \binom{10}{x_1}(.5)^{x_1}(.5)^{10-x_1}\binom{10}{x_2}(.4)^{x_2}(.6)^{10-x_2} \neq f(x_1, x_2)$$

(In general, if the range for X_1 depends on the value of X_2, then X_1 and X_2 cannot be independent.)

(e) If $T = X_1 + X_2$ then

$$f(t) = \sum_{x_1} f\,(x_1, t - x_1)$$
$$= \sum_{x_1=0}^{t} \frac{10!}{x_1!(t-x_1)!\underbrace{(10 - x_1 - (t - x_1))!}_{(10-t)!}}(.5)^{x_1}(.4)^{t-x_1}(.1)^{10-t}$$

The upper limit on x_1 is t because, for example, if $t = 7$ then A could not have won more than 7 races. Then

$$f(t) = \frac{10!}{(10-t)!}(.4)^t(.1)^{10-t}\sum_{x_1=0}^{t}\frac{1}{x_1!(t-x_1)!}\left(\frac{.5}{.4}\right)^{x_1}$$

What do we need to multiply by on the top and bottom? Can you spot it before looking below?

$$\begin{aligned} f(t) &= \frac{10!}{t!(10-t)!}(.4)^t(.1)^{10-t}\sum_{x_1=0}^{t}\frac{t!}{x_1!(t-x_1)!}\left(\frac{.5}{.4}\right)^{x_1} \\ &= \binom{10}{t}(.4)^t(.1)^{10-t}\left(1+\frac{.5}{.4}\right)^t \\ &= \binom{10}{t}(.4)^t(.1)^{10-t}\frac{(.4+.5)^t}{(.4)^t} = \binom{10}{t}(.9)^t(.1)^{10-t} \text{ for } t = 0, 1, \dots, 10. \end{aligned}$$

Exercise: Explain to yourself how this answer can be obtained from the binomial distribution, as we did in the notes following parts (b) and (c).

The following problem is similar to conditional probability problems that we solved in Chapter 4. Now we are dealing with events defined in terms of random variables. Earlier results give us things like

$$P(Y = y) = \sum_{\text{all}x} P(Y = y|X = x)P(X = x) = \sum_{\text{all}x} f(y|x)f_1(x)$$

Example: In an auto parts company an average of μ defective parts are produced per shift. The number, X, of defective parts produced has a Poisson distribution. An inspector checks all parts prior to shipping them, but there is a 10% chance that a defective part will slip by undetected. Let Y be the number of defective parts the inspector finds on a shift. Find $f(x|y)$. (The company wants to know how many defective parts are produced, but can only know the number which were actually detected.)

Solution: Think of $X = x$ being event A and $Y = y$ being event B; we want to find $P(A|B)$. To do this we'll use

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

We know $f_1(x) = \frac{\mu^x e^{-\mu}}{x!} = P(X = x)$. Also, for a given number x of defective items produced, the number, Y, detected has a binomial distribution with $n = y$ and $p = .9$, assuming each inspection takes place independently. Then

$$f(y|x) = \binom{x}{y}(.9)^y(.1)^{x-y} = \frac{f(x,y)}{f_1(x)}.$$

Therefore

$$f(x,y) = f_1(x)f(y|x) = \frac{\mu^x e^{-\mu}}{x!}\frac{x!}{y!(x-y)!}(.9)^y(.1)^{x-y}$$

To get $f(x|y)$ we'll need $f_2(y)$. We have

$$f_2(y) = \sum_{\text{all } x} f(x,y) = \sum_{x=y}^{\infty} \frac{\mu^x e^{-\mu}}{y!(x-y)!}(.9)^y(.1)^{x-y}$$

($x \geq y$ since the number of defective items produced can't be less than the number detected.)

$$= \frac{(.9)^y e^{-\mu}}{y!} \sum_{x=y}^{\infty} \frac{\mu^x (.1)^{x-y}}{(x-y)!}$$

We could fit this into the summation result $e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots$ by writing μ^x as $\mu^{x-y}\mu^y$. Then

$$\begin{aligned}
f_2(y) &= \frac{(.9\mu)^y e^{-\mu}}{y!} \sum_{x=y}^{\infty} \frac{(.1\mu)^{x-y}}{(x-y)!} \\
&= \frac{(.9\mu)^y e^{-\mu}}{y!} \left[\frac{(.1\mu)^0}{0!} + \frac{(.1\mu)^1}{1!} + \frac{(.1\mu)^2}{2!} + \cdots\right] \\
&= \frac{(.9\mu)^y e^{-\mu}}{y!} e^{.1\mu} = \frac{(.9\mu)^y e^{-.9\mu}}{y!} \\
f(x|y) &= \frac{f(x,y)}{f_2(y)} = \frac{\frac{\mu^x e^{-\mu}(.9)^y(.1)^{x-y}}{y!((x-y)!}}{\frac{(.9)^y \mu^y e^{-.9\mu}}{y!}} \\
&= \frac{(.1\mu)^{x-y} e^{-.1\mu}}{(x-y)!} \text{ for } x = y, y+1, y+2, \cdots
\end{aligned}$$

Problems:

8.1.1 The joint probability function of (X,Y) is:

		x		
$f(x,y)$		0	1	2
	0	.09	.06	.15
y	1	.15	.05	.20
	2	.06	.09	.15

a) Are X and Y independent? Why?

b) Tabulate the conditional probability function, $f(y|X=0)$.

c) Tabulate the probability function of $D = X - Y$.

8.1.2 In problem 6.14, given that x sales were made in a 1 hour period, find the probability function for Y, the number of calls made in that hour.

8.1.3 X and Y are independent, with $f(x) = \binom{x+k-1}{x} p^k (1-p)^x$ and $f(y) = \binom{y+\ell-1}{y} p^\ell (1-p)^y$. Let $T = X + Y$. Find the probability function, $f(t)$. You may use the result $\binom{a+b-1}{a} = (-1)^a \binom{-b}{a}$.

8.2 Multinomial Distribution

There is only this one multivariate model distribution introduced in this course, though other multivariate distributions exist. The multinomial distribution defined below is very important. It is a generalization of the binomial model to the case where each trial has k possible outcomes.

Physical Setup: This distribution is the same as binomial except there are k types of outcome rather than two. An experiment is repeated independently n times with k distinct types of outcome each time. Let the probabilities of these k types be $p_1, p_2, \cdots, p_k$ each time. Let X_1 be the number of times the 1st type occurs, X_2 the number of times the 2nd occurs, $\cdots$, X_k the number of times the k^{th} type occurs. Then $(X_1, X_2, \cdots, X_k)$ has a multinomial distribution.

Notes:

(1) $p_1 + p_2 + \cdots + p_k = 1$

(2) $X_1 + X_2 + \cdots + X_k = n$,

If we wish we can drop one of the variables (say the last), and just note that X_k equals $n - X_1 - X_2 - \cdots - X_{k-1}$.

Illustrations:

(1) In the example of Section 8.1 with sprinters A,B, and C running 10 races we had a multinomial distribution with $n = 10$ and $k = 3$.

(2) Suppose student marks are given in letter grades as A, B, C, D, or F. In a class of 80 students the number getting A, B, ..., F might have a multinomial distribution with $n = 80$ and $k = 5$.

Joint Probability Function: The joint probability function of $X_1, \ldots, X_k$ is given by extending the argument in the sprinters example from $k = 3$ to general k. There are $\frac{n!}{x_1!x_2!\cdots x_k!}$ different outcomes of the n trials in which x_1 are of the 1st type, x_2 are of the 2nd type, etc. Each of these arrangements has probability $p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$ since p_1 is multiplied x_1 times in some order, etc.

$$\text{Therefore } f(x_1, x_2, \cdots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

The restriction on the x_i's are $x_i = 0, 1, \cdots, n$ and $\sum_{i=1}^{k} x_i = n$.
As a check that $\sum f(x_1, x_2, \cdots, x_k) = 1$ we use the multinomial theorem to get

$$\sum \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} = (p_1 + p_2 + \cdots + p_k)^n = 1.$$

We have already seen one example of the multinomial distribution in the sprinter example. Here is another simple example.

Example: Every person is one of four blood types: A, B, AB and O. (This is important in determining, for example, who may give a blood transfusion to a person.) In a large population let the fraction that has type A, B, AB and O, respectively, be p_1, p_2, p_3, p_4. Then, if n persons are randomly selected from the population, the numbers X_1, X_2, X_3, X_4 of types A, B, AB, O have a multinomial distribution with $k = 4$ (In Caucasian people the values of the p_i's are approximately $p_1 = .45, p_2 = .08, p_3 = .03, p_4 = .44$.)

Remark: We sometimes use the notation $(X_1, \ldots, X_k) \sim Mult(n; p_1, \ldots, p_k)$ to indicate that $(X_1, \ldots, X_k)$ have a multinomial distribution.

Remark: For some types of problems its helpful to write formulas in terms of $x_1, \ldots, x_{k-1}$ and $p_1, \ldots, p_{k-1}$ using the fact that

$$x_k = n - x_1 - \cdots - x_{k-1} \quad \text{and} \quad p_k = 1 - p_1 - \cdots - p_{k-1}.$$

In this case we can write the joint p.f. as $f(x_1, \ldots, x_{k-1})$ but we must remember then that $x_1, \ldots, x_{k-1}$ satisfy the condition $0 \le x_1 + \cdots + x_{k-1} \le n$.
The multinomial distribution can also arise in combination with other models, and students often have trouble recognizing it then.

Example: A potter is producing teapots one at a time. Assume that they are produced independently of each other and with probability p the pot produced will be "satisfactory"; the rest are sold at a lower price. The number, X, of rejects before producing a satisfactory teapot is recorded. When 12 satisfactory teapots are produced, what is the probability the 12 values of X will consist of six 0's, three 1's, two 2's and one value which is ≥ 3?

Solution: Each time a "satisfactory" pot is produced the value of X falls in one of the four categories $X = 0, X = 1, X = 2, X \ge 3$. Under the assumptions given in this question, X has a geometric distribution with

$$f(x) = p(1-p)^x; \text{ for } x = 0, 1, 2, \cdots$$

so we can find the probability for each of these categories. We have $P(X = x) = f(x)$ for $0, 1, 2$, and we can obtain $P(X \ge 3)$ in various ways:

a)

$$P\left(X \geq 3\right) = f(3) + f(4) + f(5) + \cdots = p(1-p)^3 + p(1-p)^4 + p(1-p)^5 + \cdots$$
$$= \frac{p(1-p)^3}{1-(1-p)} = (1-p)^3$$

since we have a geometric series.

b)

$$P\left(X \geq 3\right) = 1 - P\left(X < 3\right) = 1 - f(0) - f(1) - f(2).$$

With some re-arranging, this also gives $(1-p)^3$.

c) The only way to have $X \geq 3$ is to have the first 3 pots produced all being rejects. Therefore $P\left(X \geq 3\right) = P$ (3 consecutive rejects) = $(1-p)(1-p)(1-p) = (1-p)^3$

Reiterating that each time a pot is successfully produced, the value of X falls in one of 4 categories $(0, 1, 2, or\ \geq 3)$, we see that the probability asked for is given by a multinomial distribution, Mult$(12; f(0), f(1), f(2), P(X \geq 3))$:

$$\begin{aligned} f(6,3,2,1) &= \tfrac{12!}{6!3!2!1!}\left[f(0)\right]^6 \left[f(1)\right]^3 [f(2)]^2 [P\left(X \geq 3\right)]^1 \\ &= \tfrac{12!}{6!3!2!1!} p^6 \left[p(1-p)\right]^3 \left[p(1-p)^2\right]^2 \left[(1-p)^3\right]^1 \\ &= \tfrac{12!}{6!3!2!1!} p^{11}(1-p)^{10} \end{aligned}$$

Problems:

8.2.1 An insurance company classifies policy holders as class A,B,C, or D. The probabilities of a randomly selected policy holder being in these categories are .1, .4, .3 and .2, respectively. Give expressions for the probability that 25 randomly chosen policy holders will include

(a) 3A's, 11B's, 7C's, and 4D's.

(b) 3A's and 11B's.

(c) 3A's and 11B's, given that there are 4D's.

8.2.2 Chocolate chip cookies are made from batter containing an average of 0.6 chips per c.c. Chips are distributed according to the conditions for a Poisson process. Each cookie uses 12 c.c. of batter. Give expressions for the probabilities that in a dozen cookies:

(a) 3 have fewer than 5 chips.

(b) 3 have fewer than 5 chips and 7 have more than 9.

(c) 3 have fewer than 5 chips, given that 7 have more than 9.

8.3 Markov Chains

Consider a sequence of (discrete) random variables $X_1, X_2, \ldots$ each of which takes integer values $1, 2, \ldots N$ (called *states*). We assume that for a certain matrix P (called the *transition probability matrix*), the conditional probabilities are given by corresponding elements of the matrix; i.e.

$$P[X_{n+1} = j | X_n = i] = P_{ij}, i = 1, \ldots N, j = 1, \ldots N$$

and furthermore that the chain only uses the last state occupied in determining its future; i.e. that

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_1, X_{n-2} = i_2 ... X_{n-l} = i_l] = P[X_{n+1} = j | X_n = i] = P_{ij}$$

for all $j, i, i_1, i_2, \ldots i_l$, and $l = 2, 3, ...$ Then the sequence of random variables X_n is called a *Markov*[9] *Chain*. Markov Chain models are the most common simple models for dependent variables, and are used to predict weather as well as movements of security prices. They allow the future of the process to depend on the present state of the process, but the past behaviour can influence the future only through the present state.

Example. Rain-No rain

Suppose that the probability that tomorrow is rainy given that today is not raining is α (and it does not otherwise depend on whether it rained in the past) and the probability that tomorrow is dry given that today is rainy is β. If tomorrow's weather depends on the past only through whether today is wet or dry, we can define random variables

$$X_n = \begin{cases} 1 & \text{if} \quad \text{Day } n \text{ is wet} \\ 0 & \text{if} \quad \text{Day } n \text{ is dry} \end{cases}$$

(beginning at some arbitrary time origin, day $n = 0$). Then the random variables $X_n, n = 0, 1, 2, ...$ form a Markov chain with $N = 2$ possible states and having probability transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

[9] After Andrei Andreyevich Markov (1856-1922), a Russian mathematician, Professor at Saint Petersburg University. Markov studied sequences of mutually dependent variables, hoping to establish the limiting laws of probability in their most general form and discovered Markov chains, launched the theory of stochastic processes. As well, Markov applied the method of continued fractions, pioneered by his teacher Pafnuty Chebyshev, to probability theory, completed Chebyschev's proof of the central limit theorem (see Chapter 9) for independent non-identically distributed random variables. For entertainment, Markov was also interested in poetry and studied poetic style.

Properties of the Transition Matrix P

Note that $P_{ij} \geq 0$ for all i, j and $\sum_{j=1}^{N} P_{ij} = 1$ for all i. This last property holds because given that $X_n = i$, X_{n+1} must occupy one of the states $j = 1, 2, ..., N$.

The distribution of X_n

Suppose that the chain is started by randomly choosing a state for X_0 with distribution $P[X_0 = i] = q_i, i = 1, 2, \ldots N$. Then the distribution of X_1 is given by

$$\begin{aligned} P(X_1 = j) &= \sum_{i=1}^{N} P(X_1 = j, X_0 = i) \\ &= \sum_{i=1}^{N} P(X_1 = j | X_0 = i) P(X_0 = i) \\ &= \sum_{i=1}^{N} P_{ij} q_i \end{aligned}$$

and this is the $j'th$ element of the vector $q'P$ where $\underline{q}$ is the column vector of values q_i. To obtain the distribution at time $n = 1$, premultiply the transition matrix P by a vector representing the distribution at time $n = 0$. Similarly the distribution of X_2 is the vector $\underline{q}'P^2$ where P^2 is the product of the matrix P with itself and the distribution of X_n is $\underline{q}'P^n$. Under very general conditions, it can be shown that these probabilities converge because the matrix P^n converges pointwise to a limiting matrix as $n \to \infty$. In fact, in many such cases, the limit does not depend on the initial distribution $\underline{q}$ because the limiting matrix has all of its rows identical and equal to some vector of probabilities $\underline{\pi}$. Identifying this vector $\underline{\pi}$ when convergence holds is reasonably easy.

Definition

A *limiting distribution* of a Markov chain is a vector ($\underline{\pi}$ say) of long run probabilities of the individual states so

$$\pi_i = \lim_{t \to \infty} P[X_t = i].$$

Now let us suppose that convergence to this distribution holds for a particular initial distribution $\underline{q}$ so we assume that

$$\underline{q}'P^n \to \underline{\pi}' \text{ as } n \to \infty.$$

Then notice that

$$\underline{q}'P^n P \to \underline{\pi}'P$$

but also

$$\underline{q}'P^n P = \underline{q}'P^{n+1} \to \underline{\pi}' \text{ as } n \to \infty$$

so $\underline{\pi}'$ must have the property that

$$\underline{\pi}'P = \underline{\pi}'$$

Any limiting distribution must have this property and this makes it easy in many examples to identify the limiting behaviour of the chain.

Definition 23 *A* stationary distribution *of a Markov chain is the column vector ($\underline{\pi}$ say) of probabilities of the individual states such that $\underline{\pi}'P = \underline{\pi}'$.*

Example: (weather continued)

Let us return to the weather example in which the transition probabilities are given by the matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

What is the long-run proportion of rainy days? To determine this we need to solve the equations

$$\underline{\pi}'P = \underline{\pi}'$$

$$\begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix}$$

subject to the conditions that the values π_0, π_1 are both probabilities (non-negative) and add to one. It is easy to see that the solution is

$$\pi_0 = \frac{\beta}{\alpha+\beta}$$
$$\pi_1 = \frac{\alpha}{\alpha+\beta}$$

which is intuitively reasonable in that it says that the long-run probability of the two states is proportional to the probability of a switch to that state from the other. So the long-run probability of a dry day is the limit

$$\pi_0 = \lim_{n\to\infty} P(X_n = 0) = \frac{\beta}{\alpha+\beta}.$$

You might try verifying this by computing the powers of the matrix P^n for $n = 1, 2, ...$ and show that P^n approaches the matrix

$$\begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$$

as $n \to \infty$. There are various mathematical conditions under which the limiting distribution of a Markov chain unique and independent of the initial state of the chain but roughly they assert that the chain is such that it forgets the more and more distant past.

Example (Gene Model) TA simple form of inheritance of traits occurs when a trait is governed by a pair of genes A and a.An individual may have an AA of an Aa combination (in which case they are indistinguishable in appearance, or "A dominates a”$)$. Let us call an AA individual *dominant*, aa, *recessive* and Aa *hybrid*. When two individuals mate, the offspring inherits one gene of the pair from each parent, and we assume that these genes are selected at random. Now let us suppose that two individuals of opposite sex selected at random mate, and then two of their offspring mate, etc. Here the state is determined by a pair of individuals, so the states of our process can be considered to be objects like (AA, Aa) indicating that one of the pair is AA and the other is Aa (we do not distinguish the order of the pair, or male and female-assuming these genes do not depend on the sex of the individual)

Number	State
1	(AA, AA)
2	(AA, Aa)
3	(AA, aa)
4	(Aa, Aa)
5	$Aa, aa)$
6	(aa, aa)

For example, consider the calculation of $P(X_{t+1} = j|X_t = 2)$. In this case each offspring has probability $1/2$ of being a dominant AA, and probability of $1/2$ of being a hybrid (Aa). If two offspring are selected independently from this distribution the possible pairs are (AA, AA), (AA, Aa), (Aa, Aa) with probabilities $1/4, 1/2, 1/4$ respectively. So the transitions have probabilities below:

	(AA, AA)	(AA, Aa)	(AA, aa)	(Aa, Aa)	(Aa, aa)	(aa, aa)
(AA, AA)	1	0	0	0	0	0
(AA, Aa)	.25	.5	0	.25	0	0
(AA, aa)	0	0	0	1	0	0
(Aa, Aa)	.0625	.25	.125	.25	.25	.0625
(Aa, aa)	0	0	0	.25	.5	.25
(aa, aa)	0	0	0	0	0	1

and transition probability matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ .25 & .5 & 0 & .25 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ .0625 & .25 & .125 & .25 & .25 & .0625 \\ 0 & 0 & 0 & .25 & .5 & .25 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

What is the long-run behaviour in such a system? For example, the two-generation transition probabilities are given by

$$P^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.3906 & 0.3125 & 0.0313 & 0.1875 & 0.0625 & .01156 \\ 0.0625 & 0.25 & 0.125 & 0.25 & 0.25 & 0.0625 \\ 0.1406 & 0.1875 & 0.0312 & 0.3125 & 0.1875 & 0.14063 \\ 0.01562 & 0.0625 & 0.0313 & 0.1875 & 0.3125 & 0.3906 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

which seems to indicate a drift to one or other of the extreme states 1 or 6. To confirm the long-run behaviour calculate and :

$$P^{100} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.75 & 0 & 0 & 0 & 0 & 0.25 \\ 0.5 & 0 & 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0 & 0.5 \\ 0.25 & 0 & 0 & 0 & 0 & 0.75 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

which shows that eventually the chain is absorbed in either of state 1 or state 6, with the probability of absorption depending on the initial state. This chain, unlike the ones studied before, has more than one possible stationary distributions, for example, $\pi' = (1, 0, 0, 0, 0, 0)$ and $\pi' = (0, 0, 0, 0, 0, 1)$,and in these circumstances the chain does not have the same limiting distribution regardless of the initial state.

8.4 Expectation for Multivariate Distributions: Covariance and Correlation

It is easy to extend the definition of expectation to multiple variables. Generalizing $E[g(X)] = \sum_{\text{all } x} g(x)f(x)$ leads to the definition of expected value in the multivariate case

Definition 24

$$E[g(X,Y)] = \sum_{\text{all } (x,y)} g(x,y)f(x,y)$$

and

$$E[g(X_1, X_2, \cdots, X_n)] = \sum_{\text{all } (x_1,x_2,\cdots,x_n)} g(x_1, x_2, \cdots x_n) f(x_1, \cdots, x_n)$$

As before, these represent the average value of $g(X,Y)$ and $g(X_1,\dots,X_n)$.

Example: Let the joint probability function, $f(x,y)$, be given by

	$f(x,y)$	$x=0$	1	2
	1	.1	.2	.3
y	2	.2	.1	.1

Find $E(XY)$ and $E(X)$.

Solution:

$$\begin{aligned} E(XY) &= \sum_{\text{all }(x,y)} xyf(x,y) \\ &= (0\times 1\times .1)+(1\times 1\times .2)+(2\times 1\times .3)+(0\times 2\times .2)+(1\times 2\times .1)+(2\times 2\times .1) \\ &= 1.4 \end{aligned}$$

To find $E(X)$ we have a choice of methods. First, taking $g(x,y)=x$ we get

$$\begin{aligned} E(X) &= \sum_{\text{all }(x,y)} xf(x,y) \\ &= (0\times .1)+(1\times .2)+(2\times .3)+(0\times .2)+(1\times .1)+(2\times .1) \\ &= 1.1 \end{aligned}$$

Alternatively, since $E(X)$ only involves X, we could find $f_1(x)$ and use

$$E(X)=\sum_{x=0}^{2} xf_1(x)=(0\times .3)+(1\times .3)+(2\times .4)=1.1$$

Example: In the example of Section 8.1 with sprinters A, B, and C we had (using only X_1 and X_2 in our formulas)

$$f(x_1,x_2)=\frac{10!}{x_1!x_2!(10-x_1-x_2)!}(.5)^{x_1}(.4)^{x_2}(.1)^{10-x_1-x_2}$$

where A wins x_1 times and B wins x_2 times in 10 races. Find $E(X_1X_2)$.

Solution: This will be similar to the way we derived the mean of the binomial distribution but, since this is a multinomial distribution, we'll be using the multinomial theorem to sum.

$$E\left(X_1X_2\right)=\sum x_1x_2f(x_1,x_2)=\sum_{\substack{x_1\neq 0\\ x_2\neq 0}} x_1x_2\frac{10!}{x_1(x_1-1)!x_2(x_2-1)!(10-x_1-x_2)!}(.5)^{x_1}(.4)^{x_2}(.1)^{10-x_1-x_2}$$
$$=\sum_{\substack{x_1\neq 0\\ x_2\neq 0}}\frac{(10)(9)(8!)}{(x_1-1)!(x_2-1)!\left[(10-2)-(x_1-1)-(x_2-1)\right]!}(.5)(.5)^{x_1-1}(.4)(.4)^{x_2-1}(.1)^{(10-2)-(x_1-1)-(x_2-1)}$$
$$=(10)(9)(.5)(.4)\sum_{\substack{x_1\neq 0\\ x_2\neq 0}}\frac{8!}{(x_1-1)!(x_2-1)!\left[8-(x_1-1)-(x_2-1)\right]!}(.5)^{x_1-1}(.4)^{x_2-1}(.1)^{8-(x_1-1)-(x_2-1)}$$

Let $y_1 = x_1 - 1$ and $y_2 = x_2 - 1$ in the sum and we obtain

$$\begin{aligned}E\left(X_1X_2\right)&=(10)(9)(.5)(.4)\sum_{(y_1,y_2)}\frac{8!}{y_1!y_2!(8-y_1-y_2)!}(.5)^{y_1}(.4)^{y_2}(.1)^{8-y_1-y_2}\\&=18(.5+.4+.1)^8=18\end{aligned}$$

Property of Multivariate Expectation: It is easily proved (make sure you can do this) that

$$E\left[ag_1(X,Y)+bg_2(X,Y)\right]=aE\left[g_1(X,Y)\right]+bE\left[g_2(X,Y)\right]$$

This can be extended beyond 2 functions g_1 and g_2, and beyond 2 variables X and Y.

Relationships between Variables:

Independence is a "yes/no" way of defining a relationship between variables. We all know that there can be different types of relationships between variables which are dependent. For example, if X is your height in inches and Y your height in centimetres the relationship is one-to-one and linear. More generally, two random variables may be related (non-independent) in a probabilistic sense. For example, a person's weight Y is not an exact linear function of their height X, but Y and X are nevertheless related. We'll look at two ways of measuring the strength of the relationship between two random variables. The first is called covariance.

Definition 25 *The **covariance** of X and Y, denoted $\mathrm{Cov}(X,Y)$ or σ_{XY}, is*

$$\mathrm{Cov}(X,Y)=E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

For calculation purposes this definition is usually harder to use than the formula which follows, which is proved noting that

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E\left[(X-\mu_X)(Y-\mu_Y)\right] = E\left(XY-\mu_X Y - X\mu_Y + \mu_X\mu_Y\right) \\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\
&= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\
\text{Therefore } \mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y)
\end{aligned}$$

Example:
In the example with joint probability function

	$f(x,y)$	x = 0	1	2
y	1	.1	.2	.3
	2	.2	.1	.1

find Cov (X,Y).

Solution: We previously calculated $E(XY) = 1.4$ and $E(X) = 1.1$. Similarly, $E(Y) = (1 \times .6) + (2 \times .4) = 1.4$

$$\text{Therefore Cov}(X,Y) = 1.4 - (1.1)(1.4) = -.14$$

Exercise: Calculate the covariance of X_1 and X_2 for the sprinter example. We have already found that $E(X_1X_2)$ = 18. The marginal distributions of X_1 and of X_2 are models for which we've already derived the mean. If your solution takes more than a few lines you're missing an easier solution.

Interpretation of Covariance:

(1) Suppose large values of X tend to occur with large values of Y and small values of X with small values of Y. Then $(X-\mu_X)$ and $(Y-\mu_Y)$ will tend to be of the same sign, whether positive or negative. Thus $(X-\mu_X)(Y-\mu_Y)$ will be positive. Hence Cov $(X,Y) > 0$. For example in Figure 8.2 we see several hundred points plotted. Notice that the majority of the points are in the two quadrants (lower left and upper right) labelled with "+" so that for these $(X-\mu_X)(Y-\mu_Y) > 0$. A minority of points are in the other two quadrants labelled

"-" and for these $(X-\mu_X)(Y-\mu_Y) < 0$. Moreover the points in the latter two quadrants appear closer to the mean (μ_X, μ_Y) indicating that on average, over all points generated $average((X-\mu_X)(Y-\mu_Y)) > 0$. Presumably this implies that over the joint distribution of (X,Y), $E[(X-\mu_X)(Y-\mu_Y)] > 0$ or $Cov(X,Y) > 0$.

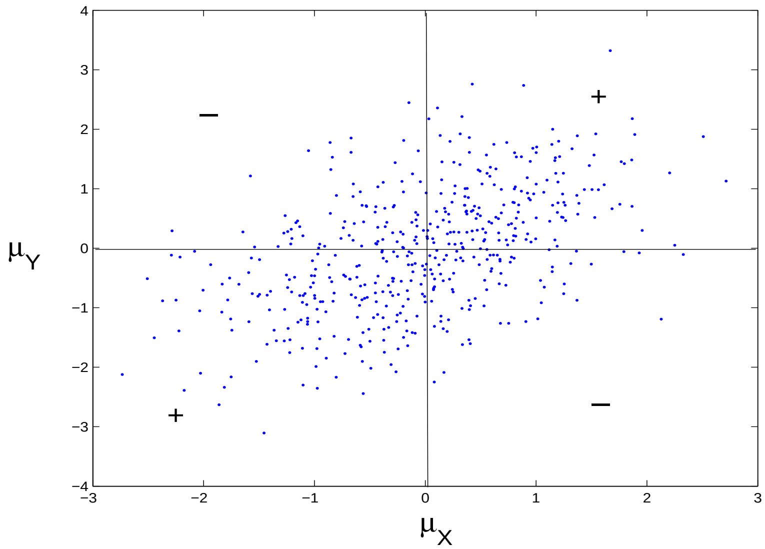


Figure 8.2: Random points (X,Y) with covariance 0.5, variances 1.

For example of X =person's height and Y =person's weight, then these two random variables will have positive covariance.

(2) Suppose large values of X tend to occur with small values of Y and small values of X with large values of Y. Then $(X-\mu_X)$ and $(Y-\mu_Y)$ will tend to be of opposite signs. Thus $(X-\mu_X)(Y-\mu_Y)$ tends to be negative. Hence Cov $(X,Y) < 0$. For example see Figure 8.3

For example if X =thickness of attic insulation in a house and Y =heating cost for the house, then $Cov(X,Y) < 0$.

Theorem 26 *If X and Y are independent then Cov $(X,Y) = 0$.*

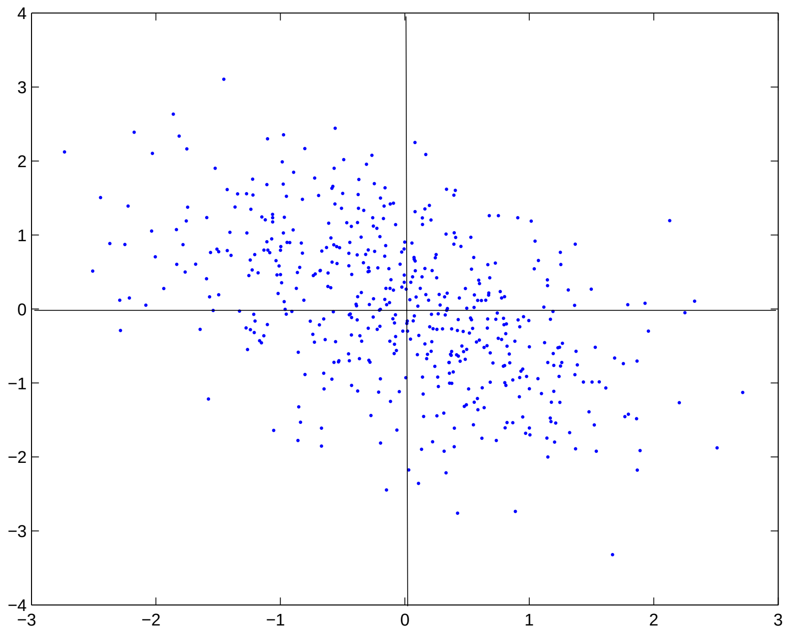


Figure 8.3: Covariance=-0.5, variances=1

Proof: Recall $E\left(X-\mu_X\right)=E(X)-\mu_X=0$. Let X and Y be independent.
Then $f(x,y)=f_1(x)f_2(y)$.

$$
\begin{aligned}
\operatorname{Cov}(X,Y) &= E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]=\sum_{\text{all } y}\left[\sum_{\text{all } x}\left(x-\mu_X\right)\left(y-\mu_Y\right)f_1(x)f_2(y)\right] \\
&= \sum_{\text{all } y}\left[\left(y-\mu_Y\right)f_2(y)\sum_{\text{all } x}\left(x-\mu_X\right)f_1(x)\right] \\
&= \sum_{\text{all } y}\left[\left(y-\mu_Y\right)f_2(y)E\left(X-\mu_X\right)\right] \\
&= \sum_{\text{all } y} 0 = 0
\end{aligned}
$$

The following theorem gives a direct proof the result above, and is useful in many other situations.

Theorem 27 *Suppose random variables X and Y are independent. Then, if $g_1(X)$ and $g_2(Y)$ are any two functions,*

$$E[g_1(X)g_2(Y)]=E[g_1(X)]E[g_2(Y)].$$

Proof: Since X and Y are independent, $f(x,y) = f_1(x)f_2(y)$. Thus

$$\begin{aligned} E[g_1(X)g_2(Y)] &= \sum_{\text{all}(x,y)} g_1(x)g_2(y)f(x,y) \\ &= \sum_{\text{all } x}\sum_{\text{all } y} g_1(x)f_1(x)g_2(y)f_2(y) \\ &= [\sum_{\text{all } x} g_1(x)f_1(x)][\sum_{\text{all } y} g_2(y)f_2(y)] \\ &= E[g_1(X)]E[g_2(Y)] \end{aligned}$$

$\square$

To prove result (3) above, we just note that if X and Y are independent then

$$\begin{aligned} \text{Cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E(X-\mu_X)E(Y-\mu_Y) = 0 \end{aligned}$$

Caution: This result is <u>not</u> reversible. If Cov $(X,Y) = 0$ we can not conclude that X and Y are independent. For example suppose that the random variable Z is uniformly distributed on the values $\{-1, -0.9, ...0.9, 1\}$ and define $X = \sin(2\pi Z)$ and $Y = \cos(2\pi Z)$. It is easy to see that Cov$(X,Y) = 0$ but the two random variables X, Y are clearly related because the points (X,Y) are always on a circle.

Example: Let (X,Y) have the joint probability function $f(0,0) = 0.2$, $f(1,1) = 0.6$, $f(2,0) = 0.2$; i.e. (X,Y) only takes 3 values.

x	0	1	2
$f_1(x)$	.2	.6	.2

and

y	0	1
$f_2(y)$	.4	.6

are marginal probability functions. Since $f_1(x)f_2(y) \neq f(x,y)$, therefore, X and Y are not independent. However,

$$\begin{aligned} E(XY) &= (0\times 0\times .2) + (1\times 1\times .6) + (2\times 0\times .2) = .6 \\ E(X) &= (0\times .2) + (1\times .6) + (2\times .2) = 1 \text{ and } E(Y) = (0\times .4) + (1\times .6) = .6 \end{aligned}$$

Therefore $\text{Cov}\,(X,Y) = E(XY) - E(X)E(Y) = .6 - (1)(.6) = 0$

So X and Y have covariance 0 but are not independent. If Cov $(X,Y) = 0$ we say that X and Y are uncorrelated, because of the definition of correlation [10] given below.

[10]" The finest things in life include having a clear grasp of correlations. " Albert Einstein, 1919.

(4) The actual numerical value of Cov (X, Y) has no interpretation, so covariance is of limited use in measuring relationships.

Exercise:

(a) Look back at the example in which $f(x, y)$ was tabulated and Cov $(X, Y) = -.14$. Considering how covariance is interpreted, does it make sense that Cov (X, Y) would be negative?

(b) Without looking at the actual covariance for the sprinter exercise, would you expect Cov (X_1, X_2) to be positive or negative? (If A wins more of the 10 races, will B win more races or fewer races?)

We now consider a second, related way to measure the strength of relationship between X and Y.

Definition 28 *The **correlation coefficient** of X and Y is* $\rho = \frac{\text{Cov }(X,Y)}{\sigma_X \sigma_Y}$

The correlation coefficient measures the strength of the linear relationship between X and Y and is simply a rescaled version of the covariance, scaled to lie in the interval $[-1, 1]$. You can attempt to guess the correlation between two variables based on a scatter diagram of values of these variables at the web page
`http://statweb.calpoly.edu/chance/applets/guesscorrelation/GuessCorrelation.html`
For example in Figure 8.4 I guessed a correlation of -0.9 whereas the true correlation coefficient generating these data was $\rho = -0.91$.

Properties of ρ:

1) Since σ_X and σ_Y, the standard deviations of X and Y, are both positive, ρ will have the same sign as Cov (X, Y). Hence the interpretation of the sign of ρ is the same as for Cov (X, Y), and $\rho = 0$ if X and Y are independent. When $\rho = 0$ we say that X and Y are uncorrelated.

2) $-1 \leq \rho \leq 1$ and as $\rho \to \pm 1$ the relation between X and Y becomes one-to-one and linear.

Proof: Define a new random variable $S = X + tY$, where t is some real number. We'll show that the fact that Var$(S) \geq 0$ leads to 2) above. We have

$$\begin{aligned}
\text{Var }(S) &= E\{(S - \mu_S)^2\} \\
&= E\{[(X + tY) - (\mu_X + t\mu_Y)]^2\} \\
&= E\{[(X - \mu_X) + t(Y - \mu_Y)]^2\} \\
&= E\{(X - \mu_X)^2 + 2t(X - \mu_X)(Y - \mu_Y) + t^2(Y - \mu_Y)^2\} \\
&= \sigma_X^2 + 2t\text{Cov}(X, Y) + t^2\sigma_Y^2
\end{aligned}$$

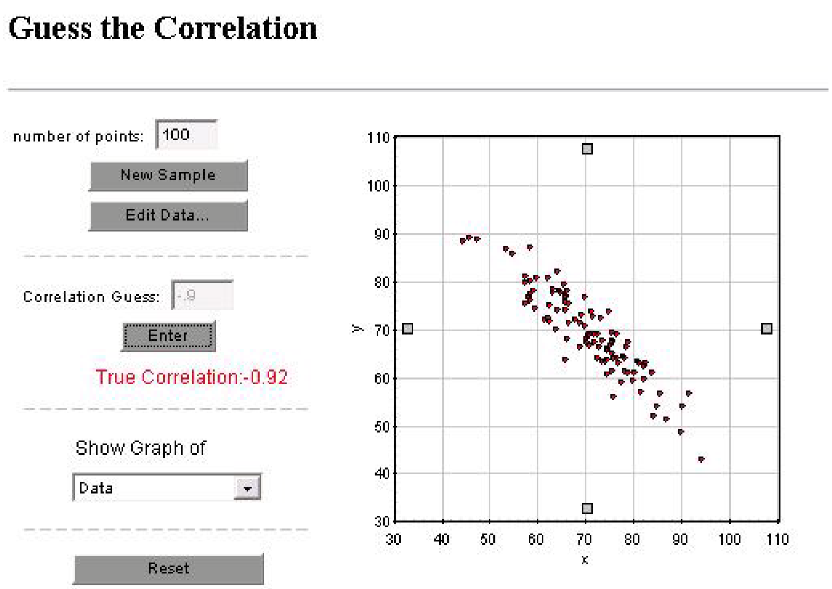


Figure 8.4: Guessing the correlation based on a scatter diagram of points

Since $Var(S) \geq 0$ for any real number t, this quadratic equation must have at most one real root (value of t for which it is zero). Therefore

$$(2\text{Cov}(X,Y))^2 - 4\sigma_X^2\sigma_Y^2 \leq 0$$

leading to the inequality

$$\left|\frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}\right| \leq 1$$

To see that $\rho \pm 1$ corresponds to a one-to-one linear relationship between X and Y, note that $\rho = \pm 1$ corresponds to a zero discriminant in the quadratic equation. This means that there exists one real number t^* for which

$$\text{Var }(S) = \text{Var }(X + t^*Y) = 0$$

But for Var$(X + t^*Y)$ to be zero, $X + t^*Y$ must equal a constant c. Thus X and Y satisfy a linear relationship.

Exercise: Calculate ρ for the sprinter example. Does your answer make sense? (You should already have found Cov (X_1, X_2) in a previous exercise, so little additional work is needed.)

Problems:

8.4.1 The joint probability function of (X, Y) is:

$f(x,y)$		$x=0$	$x=1$	$x=2$
y	0	.06	.15	.09
	1	.14	.35	.21

Calculate the correlation coefficient, ρ. What does it indicate about the relationship between X and Y?

8.4.2 Suppose that X and Y are random variables with joint probability function:

$f(x,y)$		$x=2$	$x=4$	$x=6$
y	-1	1/8	1/4	p
	1	1/4	1/8	$\frac{1}{4} - p$

(a) For what value of p are X and Y uncorrelated?

(b) Show that there is no value of p for which X and Y are independent.

8.5 Mean and Variance of a Linear Combination of Random Variables

Many problems require us to consider linear combinations of random variables; examples will be given below and in Chapter 9. Although writing down the formulas is somewhat tedious, we give here some important results about their means and variances.

Results for Means:

1. $E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$, when a and b are constants. (This follows from the definition of expectation.) In particular, $E(X + Y) = \mu_X + \mu_Y$ and $E(X - Y) = \mu_X - \mu_Y$.

2. Let a_i be constants (real numbers) and $E(X_i) = \mu_i$. Then $E\left(\sum a_i X_i\right) = \sum a_i \mu_i$. In particular, $E\left(\sum X_i\right) = \sum E(X_i)$.

3. Let $X_1, X_2, \cdots, X_n$ be random variables which have mean μ. (You can imagine these being some sample results from an experiment such as recording the number of occupants in cars travelling over a toll bridge.) The sample mean is $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. Then $E\left(\overline{X}\right) = \mu$.

Proof: From (2), $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E\left(X_i\right) = \sum_{i=1}^{n} \mu = n\mu$. Thus

$$E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}E\left(\sum X_i\right) = \frac{1}{n}n\mu = \mu$$

Results for Covariance:

1. Cov $(X, X) = E\left[\left(X-\mu_X\right)\left(X-\mu_X\right)\right] = E\left[(X-\mu)^2\right] = Var\ (X)$
2. Cov $(aX+bY, cU+dV) = ac\,Cov\ (X,U) + ad\,Cov\ (X,V) + bc\,Cov\ (Y,U) + bd\,Cov\,(Y,V)$ where a, b, c, and d are constants.

Proof:

$$\begin{aligned} Cov\,(aX+bY, cU+dV) &= E\left[\left(aX+bY-a\mu_X-b\mu_Y\right)\left(cU+dV-c\mu_U-d\mu_V\right)\right] \\ &= E\left\{\left[a\left(X-\mu_X\right)+b\left(Y-\mu_Y\right)\right]\left[c\left(U-\mu_U\right)+d\left(V-\mu_V\right)\right]\right\} \\ &= acE\left[\left(X-\mu_X\right)\left(U-\mu_U\right)\right] + adE\left[\left(X-\mu_X\right)\left(V-\mu_V\right)\right] \\ &+ bcE\left[\left(Y-\mu_Y\right)\left(U-\mu_U\right)\right] + bdE\left[\left(Y-\mu_Y\right)\left(V-\mu_V\right)\right] \\ &= ac\text{ Cov }(X,U) + ad\text{ Cov }(X,V) + bc\text{ Cov }(Y,U) + bd\text{ Cov }(Y,V) \end{aligned}$$

This type of result can be generalized, but gets messy to write out.

Results for Variance:

1. **Variance of a linear combination:**

$$\text{Var }(aX+bY) = a^2\text{ Var }(X) + b^2\text{ Var}(Y) + 2ab\text{ Cov }(X,Y)$$

Proof:

$$\begin{aligned} \text{Var }(aX+bY) &= E\left[\left(aX+bY-a\mu_X-b\mu_Y\right)^2\right] \\ &= E\left\{\left[a\left(X-\mu_X\right)+b\left(Y-\mu_Y\right)\right]^2\right\} \\ &= E\left[a^2\left(X-\mu_X\right)^2 + b^2\left(Y-\mu_Y\right)^2 + 2ab\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right] \\ &= a^2E\left[\left(X-\mu_X\right)^2\right] + b^2E\left[\left(Y-\mu_Y\right)^2\right] + 2abE\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right] \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\text{ Cov }(X,Y) \end{aligned}$$

Exercise: Try to prove this result by writing $Var\ (aX + bY)$ as Cov $(aX + bY, aX + bY)$ and using properties of covariance.

2. **Variance of a sum of independent random variables:** Let X and Y be independent. Since Cov $(X, Y) = 0$, result 1. gives

$$\text{Var } (X + Y) = \sigma_X^2 + \sigma_Y^2;$$

i.e., for independent variables, the *variance of a sum is the sum of the variances*. Also note

$$\text{Var } (X - Y) = \sigma_X^2 + (-1)^2\sigma_Y^2 = \sigma_X^2 + \sigma_Y^2;$$

i.e., for independent variables, the variance of a difference is the sum of the variances.

3. **Variance of a general linear combination:** Let a_i be constants and Var $(X_i) = \sigma_i^2$. Then

$$\text{Var } \left(\sum a_i X_i\right) = \sum a_i^2 \sigma_i^2 + 2\sum_{i<j} a_i a_j \text{ Cov } (X_i, X_j)\,.$$

This is a generalization of result 1. and can be proved using either of the methods used for 1.

4. **Variance of a linear combination of independent:** Special cases of result 3. are:

   a) If $X_1, X_2, \cdots, X_n$ are independent then Cov $(X_i, X_j) = 0$, so that

   $$\text{Var } \left(\sum a_i X_i\right) = \sum a_i^2 \sigma_i^2.$$

   b) If $X_1, X_2, \cdots, X_n$ are independent and all have the same variance σ^2, then

   $$\text{Var } \left(\overline{X}\right) = \sigma^2/n$$

Proof of 4 (b): $\overline{X} = \frac{1}{n}\sum X_i$. From 4(a), Var $(\sum X_i) = \sum_{i=1}^{n} Var\ (X_i) = n\sigma^2$. Using Var $(aX + b) = a^2\ Var\ (X)$, we get:

$$\text{Var } \left(\overline{X}\right) = \text{Var } \left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\text{Var}\left(\sum X_i\right) = \frac{n\sigma^2}{n^2} = \sigma^2/n.$$

Remark: This result is a very important one in probability and statistics. To recap, it says that if $X_1, \ldots, X_n$ are independent r.v.'s with the same mean μ and some variance σ^2, then the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ has

$$E(\bar{X}) = \mu$$
$$\text{Var } (\bar{X}) = \sigma^2/n$$

This shows that the average $\bar{X}$ of n random variables with the same distribution is less variable than any single observation X_i, and that the larger n is the less variability there is. This explains mathematically why, for example, that if we want to estimate the unknown mean height μ in a population of people, we are better to take the average height for a random sample of $n = 10$ persons than to just take the height of one randomly selected person. A sample of $n = 20$ persons would be better still. There are interesting applets at the url *http://users.ece.gatech.edu/users/gtz/java/samplemean/notes.html* and *http://www.ds.unifi.it/VL/VL_EN/applets/BinomialCoinExperiment.html* which allows one to sample and explore the rate at which the sample mean approaches the expected value. In Chapter 9 we will see how to decide how large a sample we should take for a certain degree of precision. Also note that as $n \to \infty, Var(\bar{X}) \to 0$, which means that $\bar{X}$ becomes arbitrarily close to μ. This is sometimes called the "law of averages". There is a formal theorem which supports the claim that for large sample sizes, sample means approach the expected value, called the "law of large numbers".

Indicator Variables

The results for linear combinations of random variables provide a way of breaking up more complicated problems, involving mean and variance, into simpler pieces using indicator variables; an indicator variable is just a binary variable (0 or 1) that indicates whether or not some event occurs. We'll illustrate this important method with 3 examples.

Example: Mean and Variance of a Binomial R.V.

Let $X \sim Bi(n, p)$ in a binomial process. Define new variables X_i by:

X_i = 0 if the i^{th} trial was a failure

X_i = 1 if the i^{th} trial was a success.

i.e. X_i indicates whether the outcome "success" occurred on the i^{th} trial. The trick we use is that the total number of successes, X, is the sum of the X_i's:

$$X = \sum_{i=1}^{n} X_i.$$

We can find the mean and variance of X_i and then use our results for the mean and variance of a sum to get the mean and variance of X. First,

$$E\left(X_i\right)=\sum_{x_i=0}^{1} x_i f\left(x_i\right)=0 f(0)+1 f(1)=f(1)$$

But $f(1)=p$ since the probability of success is p on each trial. Therefore $E\left(X_i\right)=p$. Since $X_i=0$ or 1, $X_i=X_i^2$, and therefore

$$E\left(X_i^2\right)=E\left(X_i\right)=p.$$

Thus

$$\operatorname{Var}\left(X_i\right)=E\left(X_i^2\right)-\left[E\left(X_i\right)\right]^2=p-p^2=p(1-p).$$

In the binomial distribution the trials are independent so the X_i's are also independent. Thus

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} E\left(X_i\right)=\sum_{i=1}^{n} p=np \\ \operatorname{Var}(X) &= \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} \operatorname{Var}\left(X_i\right)=\sum_{i=1}^{n} p(1-p)=np(1-p) \end{aligned}$$

These, of course, are the same as we derived previously for the mean and variance of the binomial distribution. Note how simple the derivation here is!

Remark: If X_i is a binary random variable with $P(X_i=1)=p=1-P(X_i=0)$ then $E(X_i)=p$ and Var$(X_i)=p(1-p)$, as shown above. (Note that $X_i \sim Bi(1,p)$ is actually a binomial r.v.) In some problems the X_i's are not independent, and then we also need covariances.

Example: Let X have a hypergeometric distribution. Find the mean and variance of X.

Solution: As above, let us think of the setting, which involves drawing n items at random from a total of N, of which r are "S" and $N-r$ are "F" items. Define

$$X_i=\begin{cases} 0 & \text{if } i^{\text{th}} \text{ draw is a failure } (F) \text{ item} \\ 1 & \text{if } i^{\text{th}} \text{ draw is a success } (S) \text{ item.} \end{cases}$$

Then $X=\sum_{i=1}^{n} X_i$ as for the binomial example, but now the X_i's are dependent. (For example, what we get on the first draw affects the probabilities of S and F for the second draw, and so on.) Therefore we need to find Cov(X_i, X_j) for $i \neq j$ as well as $E(X_i)$ and Var(X_i) in order to use our formula for the variance of a sum.

We see first that $P(X_i = 1) = r/N$ for each of $i = 1, \ldots, n$. (If the draws are random then the probability an S occurs in draw i is just equal to the probability position i is an S when we arrange r S's and $N-r$ F's in a row.) This immediately gives

$$E(X_i) = r/N$$
$$\mathrm{Var}(X_i) = \frac{r}{N}(1 - \frac{r}{N})$$

since

$$\mathrm{Var}(X_i) = E(X_i^2) - E(X_i)^2 = E(X_i) - E(X_i)^2.$$

The covariance of X_i and $X_j (i \neq j)$ is equal to $E(X_iX_j) - E(X_i)E(X_j)$, so we need

$$\begin{aligned} E(X_iX_j) &= \sum_{x_i=0}^{1}\sum_{x_j=0}^{1} x_ix_jf(x_i,x_j) \\ &= f(1,1) \\ &= P(X_i=1, X_j=1) \end{aligned}$$

The probability of an S on both draws i and j is just

$$r(r-1)/[N(N-1)] = P(X_i = 1)P(X_j = 1|X_i = 1)$$

Thus,

$$\begin{aligned} \mathrm{Cov}\ (X_i, X_j) &= E(X_iX_j) - E(X_i)E(X_j) \\ &= \tfrac{r(r-1)}{N(N-1)} - \left(\tfrac{r}{N}\right)\left(\tfrac{r}{N}\right) = \left(\tfrac{r}{N}\right)\left(\tfrac{r-1}{N-1} - \tfrac{r}{N}\right) \\ &= -\tfrac{r(N-r)}{N^2(N-1)} \end{aligned}$$

(Does it make sense that Cov (X_i, X_j) is negative? If you draw a success in draw i, are you more or less likely to have a success on draw j?) Now we find $E(X)$ and Var(X). First,

$$E(X) = E\left(\sum X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n}\left(\frac{r}{N}\right) = n\left(\frac{r}{N}\right)$$

Before finding Var (X), how many combinations X_i, X_j are there for which $i < j$? Each i and j takes values from $1, 2, \cdots, n$ so there are $\binom{n}{2}$ different combinations of (i, j) values. Each of these can only be written in 1 way to make $i < j$. Therefore There are $\binom{n}{2}$ combinations with $i < j$ (e.g. if $i = 1, 2, 3$ and $j = 1, 2, 3$, the combinations with $i < j$ are (1,2) (1,3) and (2,3). So there are $\binom{3}{2} = 3$ different combinations.)

Now we can find

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\mathrm{Var}\,(X_i) + 2\sum_{i<j}\mathrm{Cov}\,(X_i, X_j) \\ &= n\tfrac{r(N-r)}{N^2} + 2\tbinom{n}{2}\left[-\tfrac{r(N-r)}{N^2(N-1)}\right] \\ &= n\left(\tfrac{r}{N}\right)\left(\tfrac{N-r}{N}\right)\left[1 - \tfrac{(n-1)}{(N-1)}\right] \quad \left(\text{since } 2\tbinom{n}{2} = \tfrac{2n(n-1)}{2} = n(n-1)\right) \\ &= n\left(\tfrac{r}{N}\right)\left(1 - \tfrac{r}{N}\right)\left(\tfrac{N-n}{N-1}\right) \end{aligned}$$

In the last two examples, we know $f(x)$, and could have found $E(X)$ and Var(X) without using indicator variables. In the next example $f(x)$ is not known and is hard to find, but we can still use indicator variables for obtaining μ and σ^2. The following example is a famous problem in probability.

Example: We have N letters to N different people, and N envelopes addressed to those N people. One letter is put in each envelope at random. Find the mean and variance of the number of letters placed in the right envelope.

Solution:

$$\text{Let } X_i = \begin{cases} 0; & \text{if letter } i \text{ is not in envelope } \ i \\ 1; & \text{if letter } i \text{ is in envelope } \ i. \end{cases}$$

Then $\sum\limits_{i=1}^{N} X_i$ is the number of correctly placed letters. Once again, the X_i's are dependent (Why?).

First $E\left(X_i\right) = \sum\limits_{x_i=0}^{1} x_i f(x_i) = f(1) = \frac{1}{N} = E\left(X_i^2\right)$ (since there is 1 chance in N that letter i will be put in envelope i) and then,

$$\text{Var}\left(X_i\right) = E\left(X_i\right) - \left[E\left(X_i\right)\right]^2 = \frac{1}{N} - \frac{1}{N^2} = \frac{1}{N}\left(1 - \frac{1}{N}\right)$$

Exercise: Before calculating cov (X_i, X_j), what sign do you expect it to have? (If letter i is correctly

placed does that make it more or less likely that letter j will be placed correctly?)

Next, $E\left(X_i X_j\right) = f(1,1)$ (As in the last example, this is the only non-zero term in the sum.) Now, $f(1,1) = \frac{1}{N}\frac{1}{N-1}$ since once letter i is correctly placed there is 1 chance in $N-1$ of letter j going in envelope j.

$$\text{Therefore } E\left(X_i X_j\right) = \frac{1}{N(N-1)}$$

For the covariance,

$$\begin{aligned}
\text{Cov }(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) = \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)\left(\frac{1}{N}\right) \\
&= \frac{1}{N}\left(\frac{1}{N-1} - \frac{1}{N}\right) = \frac{1}{N^2(N-1)} \\
E\left(\sum_{i=1}^N X_i\right) &= \sum_{i=1}^N E(X_i) = \sum_{i=1}^N \frac{1}{N} = \left(\frac{1}{N}\right) N = 1 \\
\text{Var}\left(\sum_{i=1}^N X_i\right) &= \sum_{i=1}^N \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j) \\
&= \sum_{i=1}^N \frac{1}{N}\left(1 - \frac{1}{N}\right) + 2\binom{N}{2}\frac{1}{N^2(N-1)} \\
&= N\frac{1}{N}\left(1 - \frac{1}{N}\right) + 2\binom{N}{2}\frac{1}{N^2(N-1)} \\
&= 1 - \frac{1}{N} + 2\frac{N(N-1)}{2}\frac{1}{N^2(N-1)} = 1
\end{aligned}$$

(Common sense often helps in this course, but we have found no way of being able to say this result is obvious. On average 1 letter will be correctly placed and the variance will be 1, regardless of how many letters there are.)

Problems:

8.5.1 The joint probability function of (X, Y) is given by:

		x		
	$f(x,y)$	0	1	2
y	0	.15	.1	.05
	1	.35	.2	.15

Calculate $E(X)$, Var (X), Cov (X, Y) and Var $(3X-2Y)$. You may use the fact that $E(Y) = .7$ and Var (Y) = .21 without verifying these figures.

8.5.2 In a row of 25 switches, each is considered to be "on" or "off". The probability of being on is .6 for each switch, independently of other switch. Find the mean and variance of the number of unlike pairs among the 24 pairs of adjacent switches.

8.5.3 Suppose Var $(X) = 1.69$, Var $(Y) = 4$, $\rho = 0.5$; and let $U = 2X - Y$. Find the standard deviation of U.

8.5.4 Let $Y_0, Y_1, \cdots, Y_n$ be uncorrelated random variables with mean 0 and variance σ^2. Let $X_1 = Y_0 + Y_1$, $X_2 = Y_1 + Y_2, \cdots$, $X_n = Y_{n-1} + Y_n$. Find Cov (X_{i-1}, X_i) for $i = 1, 2, 3, \cdots, n$ and Var $\left(\sum_{i=1}^{n} X_i\right)$.

8.5.5 A plastic fabricating company produces items in strips of 24, with the items connected by a thin piece of plastic:

$\boxed{\text{Item 1}}-\boxed{\text{Item 2}}- \ldots -\boxed{\text{Item 24}}$

A cutting machine then cuts the connecting pieces to separate the items, with the 23 cuts made independently. There is a 10% chance the machine will fail to cut a connecting piece. Find the mean and standard deviation of the number of the 24 items which are completely separate after the cuts have been made. (Hint: Let $X_i = 0$ if item i is not completely separate, and $X_i = 1$ if item i is completely separate.)

8.6 Multivariate Moment Generating Functions

Suppose we have two possibly dependent random variables (X, Y) and we wish to characterize their joint distribution using a moment generating function. Just as the probability function and the cumulative distribution function are, in tis case, functions of two arguments, so is the moment generating function.

Definition 29 *The joint moment generating function of* (X, Y) *is*

$$M(s,t) = E\{e^{sX+tY}\}$$

Recall that if X, Y happen to be independent, $g_1(X)$ and $g_2(Y)$ are any two functions,

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]. \tag{8.4}$$

and so with $g_1(X) = e^{sX}$ and $g_2(Y) = e^{tY}$ we obtain, for independent random variables X, Y

$$M(s,t) = M_X(s)M_Y(t)$$

the product of the moment generating functions of X and Y respectively.

There is another labour-saving property of moment generating functions for independent random variables. Suppose X, Y are independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Suppose you wish the moment generating function of the sum $Z = X + Y$. One could attack this problem by first determining the probability function of Z,

$$\begin{aligned}
f_Z(z) = P(Z = z) &= \sum_{\text{all } x} P(X = x, Y = z - x) \\
&= \sum_{\text{all } x} P(X = x)P(Y = z - x) \\
&= \sum_{\text{all } x} f_X(x)f_Y(z - x)
\end{aligned}$$

and then calculating

$$E(e^{tZ}) = \sum_{\text{all } z} e^{tZ} f_Z(z).$$

Evidently lots of work! On the other hand recycling (8.4) with

$$\begin{aligned}
g_1(X) &= e^{tX} \\
g_2(Y) &= e^{tY}
\end{aligned}$$

gives

$$M_Z(t) = Ee^{t(X+Y)} = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t).$$

Theorem 30 *The moment generating function of the sum of independent random variables is the product of the individual moment generating functions.*

For example if both X and Y are independent with the same (Bernoulli) distribution

$$\begin{array}{r|ll}
x = & 0 & 1 \\
f(x) = & 1-p & p
\end{array}$$

then both have moment generating function

$$M_X(t) = M_Y(t) = (1 - p + pe^t)$$

and so the moment generating function of the sum Z is $M_X(t)M_Y(t) = (1-p+pe^t)^2$. Similarly if we add another independent Bernoulli the moment generating function is $(1-p+pe^t)^3$ and in general the sum of n independent Bernoulli random variables is $(1-p+pe^t)^n$, the moment generating function of a Binomial(n, p) distribution. This confirms that the sum of independent Bernoulli random variables has a Binomial(n, p) distribution.

8.7 Problems on Chapter 8

8.1 The joint probability function of (X, Y) is given by:

$f(x,y)$	x = 0	1	2
y = 0	.15	.1	.05
1	.35	.2	.15

a) Are X and Y independent? Why?

b) Find $P(X > Y)$ and $P(X = 1|Y = 0)$

8.2 For a person whose car insurance and house insurance are with the same company, let X and Y represent the number of claims on the car and house policies, respectively, in a given year. Suppose that for a certain group of individuals, $X \sim$ Poisson (mean $= .10$) and $Y \sim$ Poisson (mean $= .05$).

(a) If X and Y are independent, find $P(X+Y > 1)$ and find the mean and variance of $X+Y$.

(b) Suppose it was learned that $P(X = 0, Y = 0)$ was very close to $.94$. Show why X and Y cannot be independent in this case. What might explain the non-independence?

8.3 Consider Problem 2.7 for Chapter 2, which concerned machine recognition of handwritten digits. Recall that $p(x, y)$ was the probability that the number actually written was x, and the number identified by the machine was y.

(a) Are the random variables X and Y independent? Why?

(b) What is $P(X = Y)$, that is, the probability that a random number is correctly identified?

(c) What is the probability that the number 5 is incorrectly identified?

8.4 Blood donors arrive at a clinic and are classified as type A, type O, or other types. Donors' blood types are independent with P (type A) = p, P (type O) = q, and P (other type) = $1 - p - q$. Consider the number, X, of type A and the number, Y, of type O donors arriving before the 10^{th} other type.

a) Find the joint probability function, $f(x, y)$

b) Find the conditional probability function, $f(y|x)$.

8.5 Slot machine payouts. Suppose that in a slot machine there are $n+1$ possible outcomes $A_1, \ldots, A_{n+1}$ for a single play. A single play costs \$1. If outcome A_i occurs, you win $\$a_i$, for $i = 1, \ldots, n$. If outcome A_{n+1} occurs, you win nothing. In other words, if outcome A_i $(i = 1, \ldots, n)$ occurs your net profit is $a_i - 1$; if A_{n+1} occurs your net profit is - 1.

(a) Give a formula for your expected profit from a single play, if the probabilities of the $n+1$ outcomes are $p_i = P(A_i),\ i = 1, \ldots, n+1$.

(b) The owner of the slot machine wants the player's expected profit to be negative. Suppose $n = 4$, with $p_1 = .1,\ p_2 = p_3 = p_4 = .04$ and $p_5 = .78$. If the slot machine is set to pay \$3 when outcome A_1 occurs, and \$5 when either of outcomes A_2, A_3, A_4 occur, determine the player's expected profit per play.

(c) The slot machine owner wishes to pay da_i dollars when outcome A_i occurs, where $a_i = \frac{1}{p_i}$ and d is a number between 0 and 1. The owner also wishes his or her expected profit to be \$.05 per play. (The player's expected profit is -.05 per play.) Find d as a function of n and p_{n+1}. What is the value of d if $n = 10$ and $p_{n+1} = .7$?

8.6 Bacteria are distributed through river water according to a Poisson process with an average of 5 per 100 c.c. of water. What is the probability five 50 c.c. samples of water have 1 with no bacteria, 2 with one bacterium, and 2 with two or more?

8.7 A box contains 5 yellow and 3 red balls, from which 4 balls are drawn at random without replacement. Let X be the number of yellow balls on the first two draws and Y the number of yellow balls on all 4 draws.

a) Find the joint probability function, $f(x, y)$.

b) Are X and Y independent? Justify your answer.

8.8 In a quality control inspection items are classified as having a minor defect, a major defect, or as being acceptable. A carton of 10 items contains 2 with minor defects, 1 with a major defect, and 7 acceptable. Three items are chosen at random without replacement. Let X be the number selected with minor defects and Y be the number with major defects.

a) Find the joint probability function of X and Y.

b) Find the marginal probability functions of X and of Y.

c) Evaluate numerically $P(X = Y)$ and $P(X = 1|Y = 0)$.

8.9 Let X and Y be discrete random variables with joint probability function $f(x, y) = k\frac{2^{x+y}}{x!y!}$ for $x = 0, 1, 2, \cdots$ and $y = 0, 1, 2, \cdots$, where k is a positive constant.

a) Derive the marginal probability function of X.

b) Evaluate k.

c) Are X and Y independent? Explain.

d) Derive the probability function of $T = X + Y$.

8.10 **"Thinning" a Poisson process.** Suppose that events are produced according to a Poisson process with an average of λ events per minute. Each event has a probability p of being a "Type A" event, independent of other events.

(a) Let the random variable Y represent the number of Type A events that occur in a one-minute period. Prove that Y has a Poisson distribution with mean λp. (Hint: let X be the total number of events in a 1 minute period and consider the formula just before the last example in Section 8.1).

(b) Lighting strikes in a large forest region occur over the summer according to a Poisson process with $\lambda = 3$ strikes per day. Each strike has probability .05 of starting a fire. Find the probability that there are at least 5 fires over a 30 day period.

8.11 In a breeding experiment involving horses the offspring are of four genetic types with probabilities:

Type	1	2	3	4
Probability	3/16	5/16	5/16	3/16

A group of 40 independent offspring are observed. Give expressions for the following probabilities:

(a) There are 10 of each type.

(b) The total number of types 1 and 2 is 16.

(c) There are exactly 10 of type 1, given that the total number of types 1 and 2 is 16.

8.12 In a particular city, let the random variable X represent the number of children in a randomly selected household, and let Y represent the number of female children. Assume that the probability a child is female is 0.5, regardless of what size household they live in, and that the marginal distribution of X is as follows:

$$f(0) = .20, f(1) = .25, f(2) = .35, f(3) = .10, f(4) = .05,$$
$$f(5) = .02, f(6) = .01, f(7) = .01, f(8) = .01$$

(a) Determine $E(X)$.

(b) Find the probability function for the number of girls Y in a randomly chosen family. What is $E(Y)$?

8.13 In a particular city, the probability a call to a fire department concerns various situations is as given below:

1. fire in a detached home - $p_1 = .10$
2. fire in a semi detached home - $p_2 = .05$
3. fire in an apartment or multiple unit residence - $p_3 = .05$
4. fire in a non-residential building - $p_4 = .15$
5. non-fire-related emergency - $p_5 = .15$
6. false alarm - $p_6 = .50$

In a set of 10 calls, let $X_1, ..., X_6$ represent the numbers of calls of each of types $1, ..., 6$.

(a) Give the joint probability function for $X_1, ..., X_6$.

(b) What is the probability there is at least one apartment fire, given that there are 4 fire-related calls?

(c) If the average costs of calls of types $1, ..., 6$ are (in $100 units) 5, 5, 7, 20, 4, 2 respectively, what is the expected total cost of the 10 calls?

8.14 Suppose $X_1, \ldots, X_n$ have joint p.f. $f(x_1, \ldots, x_n)$. If $g(x_1, \ldots, x_n)$ is a function such that $a \leq g(x_1, \ldots, x_n) \leq b$ for all $(x_1, \ldots, x_n)$ in the range of f,

then show that $a \leq E[g(X_1, \ldots, X_n)] \leq b$.

8.15 Let X and Y be random variables with Var $(X) = 13$, Var$(Y) = 34$ and $\rho = -0.7$. Find Var$(X - 2Y)$.

8.16 Let X and Y have a trinomial distribution with joint probability function

$$f(x,y) = \frac{n!}{x!y!(n-x-y)!} p^x q^y (1-p-q)^{n-x-y}; \quad \begin{array}{l} x = 0, 1, \cdots, n \\ y = 0, 1, \cdots, n \end{array}$$

and $x + y \leq n$. Let $T = X + Y$.

a) What distribution does T have? Either explain why or derive this result.

b) For the distribution in (a), what is $E(T)$ and Var(T)?

c) Using (b) find Cov(X, Y), and explain why you expect it to have the sign it does.

8.17 Jane and Jack each toss a fair coin twice. Let X be the number of heads Jane obtains and Y the number of heads Jack obtains. Define $U = X + Y$ and $V = X - Y$.

a) Find the means and variances of U and V.

b) Find Cov (U, V)

c) Are U and V independent? Why?

8.18 A multiple choice exam has 100 questions, each with 5 possible answers. One mark is awarded for a correct answer and 1/4 mark is deducted for an incorrect answer. A particular student has probability p_i of knowing the correct answer to the i^{th} question, independently of other questions.

a) Suppose that on a question where the student does not know the answer, he or she guesses randomly. Show that his or her total mark has mean $\sum p_i$ and variance $\sum p_i (1 - p_i) + \frac{(100 - \sum p_i)}{4}$.

b) Show that the total mark for a student who refrains from guessing also has mean $\sum p_i$, but with variance $\sum p_i (1 - p_i)$. Compare the variances when all p_i's equal (i) .9, (ii) .5.

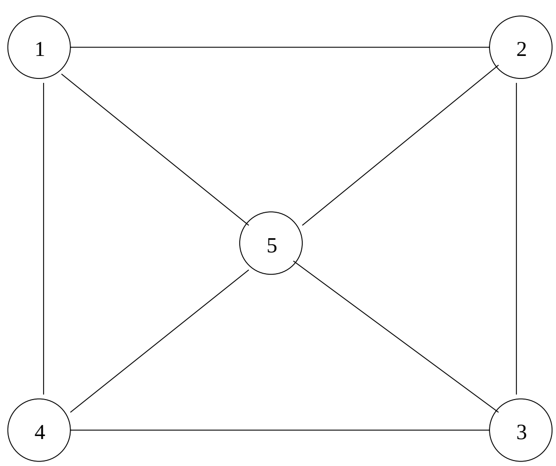
8.19 Let X and Y be independent random variables with $E(X) = E(Y) = 0$, Var$(X) = 1$ and Var $(Y) = 2$. Find Cov$(X + Y, X - Y)$.

8.20 An automobile driveshaft is assembled by placing parts A, B and C end to end in a straight line. The standard deviation in the lengths of parts A, B and C are 0.6, 0.8, and 0.7 respectively.

(a) Find the standard deviation of the length of the assembled driveshaft.

(b) What percent reduction would there be in the standard deviation of the assembled driveshaft if the standard deviation of the length of part B were cut in half?

8.21 The inhabitants of the beautiful and ancient canal city of Pentapolis live on 5 islands separated from each other by water. Bridges cross from one island to another as shown.



On any day, a bridge can be closed, with probability p, for restoration work. Assuming that the 8 bridges are closed independently, find the mean and variance of the number of islands which are completely cut off because of restoration work.

8.22 A Markov chain has a *doubly stochastic* transition matrix if both the row sums and the column sums of the transition matrix P are all 1. Show that for such a Markov chain, the uniform distribution on $\{1, 2, \ldots, N\}$ is a stationary distribution.

8.23 A salesman sells in three cities A,B, and C. He never sells in the same city on successive weeks. If he sells in city A, then the next week he always sells in B. However if he sells in either B or C, then the next week he is twice as likely to sell in city A as in the other city. What is the long-run proportion of time he spends in each of the three cities?

8.24 Find

$$\lim_{n \to \theta} P^n$$

where

$$P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

8.25 Suppose X and Y are independent having Poisson distributions with parameters λ_1 and λ_2 respectively. Use moment generating functions to identify the distribution of the sum $X + Y$.

8.26 Waterloo in January is blessed by many things, but not by good weather. There are never two nice days in a row. If there is a nice day, we are just as likely to have snow as rain the next day. If we have snow or rain, there is an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. Taking as states the kinds of weather R, N, and S. the transition probabilities P are as follows

$$P = \left(\begin{array}{cccc} & \text{R} & \text{N} & \text{S} \\ \text{R} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \text{N} & \frac{1}{2} & 0 & \frac{1}{2} \\ \text{S} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{array} \right)$$

If today is raining, find the probability of Rain, Nice, Snow three days from now. Find the probabilities of the three states in five days, given (1) today is raining (ii) today is nice (iii) today is snowing.

8.27 (**One-card Poker**) A card game, which, for the purposes of this question we will call Metzler Poker, is played as follows. Each of 2 players bets an initial \$1 and is dealt a card from a deck of 13 cards numbered 1-13. Upon looking at their card, each player then decides (unaware of the other's decision) whether or not to increase their bet by \$5 (to a total stake of \$6). If both increase the stake ("raise"), then the player with the higher card wins both stakes-i.e. they get their money back as well as the other player's \$6. If one person increases and the other does not, then the player who increases automatically wins the pot (i.e. money back+\$1). If neither person increases the stake, then it is considered a draw-each player receives their own \$1 back. Suppose that Player A and B have similar strategies, based on threshold numbers {*a*,*b*} they have chosen between 1 and 13. A chooses to raise whenever their card is greater than or equal to *a* and B whenever B's card is greater than or equal to *b*.

(a) Suppose B always raises (so that *b=1*). What is the expected value of A's win or loss for the different possible values of *a=1,2,...,13.*

(b) Suppose *a* and *b* are arbitrary. Given that both players raise, what is the probability that A wins? What is the expected value of A's win or loss?

(c) Suppose you know that *b=11.* Find your expected win or loss for various values of *a* and determine the optimal value. How much do you expect to make or lose per game under this optimal strategy?

8.28 (**Searching a database**) Suppose that we are given *3* records, R_1, R_2, R_3 *initially* stored in that order. The cost of accessing the *j'th* record in the list is *j* so we would like the more frequently accessed records near the front of the list. Whenever a request for record *j* is processed, the "move-to-front" heuristic stores R_j at the front of the list and the others in the original order. For example if the first request is for record 2, then the records will be re-stored in the order R_2, R_1, R_3. Assume that on each request, record j is requested with probability p_j, for $j = 1, 2, 3$.

(a) Show that if X_t =the permutation that obtains after j requests for records (e.g. $X_2 = (2, 1, 3)$), then X_t is a Markov chain.

(b) Find the stationary distribution of this Markov chain. (Hint: what is the probability that X_t takes the form $(2, *, *)$?).

(c) Find the expected long-run cost per record accessed in the case $p_1, p_2, p_3 = 0.1, 0.3, 0.6$ respectively.

(d) How does this expected long-run cost compare with keeping the records in random order, and with keeping them in order of decreasing values of p_j(only possible if we know p_j).

9. Continuous Probability Distributions

9.1 General Terminology and Notation

A **continuous random variable** is one for which the range (set of possible values) is an interval (or a collection of intervals) on the real number line. Continuous variables have to be treated a little differently than discrete ones, the reason being that $P(X = x)$ has to be zero for each x, in order to avoid mathematical contradiction. The distribution of a continuous random variable is called a **continuous probability distribution**. To illustrate, consider the simple spinning pointer in Figure 9.1. where all numbers in the interval (0,4] are equally likely. The probability of the pointer stopping precisely at the number x must be zero, because if not the total probability for $R = \{x : 0 < x \leq 4\}$ would be infinite, since the set R is non-countable. Thus, for a continuous random variable the probability at each point is 0. This means we can no longer use the probability function to describe a distribution. Instead there are two other functions commonly used to describe continuous distributions.

Cumulative Distribution Function:

For discrete random variables we defined the c.d.f., $F(x) = P(X \leq x)$. This function still works for continuous random variables. For the spinner, the probability the pointer stops between 0 and 1 is 1/4 if all values x are equally “likely"; between 0 and 2 the probability is 1/2, between 0 and 3 it is 3/4; and so on. In general, $F(x) = x/4$ for $0 < x \leq 4$.

Also, $F(x) = 0$ for $x \leq 0$ since there is no chance of the pointer stopping at a number ≤ 0, and $F(x) = 1$ for $x > 4$ since the pointer is certain to stop at number below x if $x > 4$.

Most properties of a c.d.f. are the same for continuous variables as for discrete variables. These are:

1. $F(-\infty) = 0$;and $F(\infty) = 1$
2. $F(x)$ is a non-decreasing function of x
3. $P(a < X \leq b) = F(b) - F(a)$.

Note that, as indicated before, for a continuous distribution, we have $P(X = a) = P(a \leq X \leq a) = F(a) - F(a) = 0$. Also, since the probability is 0 at each point:

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

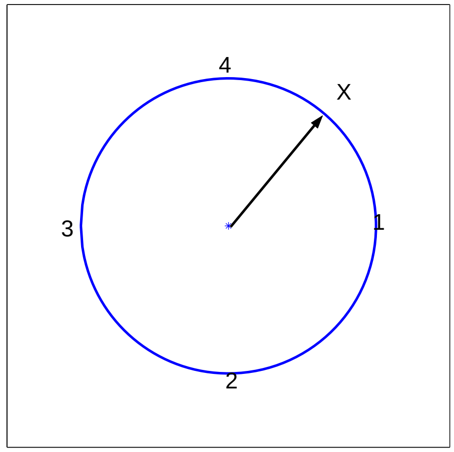


Figure 9.1: Spinner: a device for generating a continuous random variable (in a zero-gravity, virtually frictionless environment)

(For a discrete random variable, each of these 4 probabilities could be different.). For the continuous distributions in this chapter, we do not worry about whether intervals are open, closed, or half-open since the probability of these intervals is the same.

Probability Density Function (p.d.f.): While the c.d.f. can be used to find probabilities, it does not give an intuitive picture of which values of x are more likely, and which are less likely. To develop such a picture suppose that we take a short interval of X-values, $[x, x + \Delta x]$. The probability X lies in the interval is

$$P(x \le X \le x + \Delta x) = F(x + \Delta x) - F(x).$$

To compare the probabilities for two intervals, each of length Δx, is easy. Now suppose we consider what happens as Δx becomes small, and we divide the probability by Δx. This leads to the following definition.

Definition 31 *The **probability density function** (p.d.f.) $f(x)$ for a continuous random variable X is the derivative*

$$f(x) = \frac{dF(x)}{dx}$$

where $F(x)$ is the c.d.f. for X.

We will see that $f(x)$ represents the relative likelihood of different x-values. To do this we first note some properties of a p.d.f. It is assumed that $f(x)$ is a continuous function of x at all points for which $0 < F(x) < 1$.

Properties of a probability density function

1. $P(a \le X \le b) = F(b) - F(a) = \int_a^b f(x)dx$
   This follows from the definition of $f(x)$.

2. $f(x) \ge 0$
   (Since $F(x)$ is non-decreasing).

3. $\int_{-\infty}^{\infty} f(x)dx = \int_{\text{all}x} f(x)dx = 1$
   This is because $P(-\infty \le X \le \infty) = 1$.

4. $F(x) = \int_{-\infty}^{x} f(u)du$.
   This is just property 1 with $a = -\infty$.

To see that $f(x)$ represents the relative likelihood of different outcomes, we note that for Δx small,

$$P(x \le X \le x + \Delta x) = F(x + \Delta x) - F(x) \doteq f(x)\Delta x.$$

Thus, $f(x) \ne P(X = x)$ **but** $f(x)\Delta x$ is the approximate probability that X is inside the interval $[x, x + \Delta x]$. A plot of the function $f(x)$ shows such values clearly and for this reason it is very common to plot the p.d.f.'s of continuous random variables.

Example: Consider the spinner example, where

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{x}{4} & \text{for } 0 < x \le 4 \\ 1 & \text{for } x > 4 \end{cases}$$

Thus, the p.d.f. is $f(x) = F'(x)$, or

$$f(x) = \frac{1}{4} \text{ for } 0 < x < 4.$$

and outside this interval the p.d.f. is 0.Figure 9.2 shows the probability density function $f(x)$; for obvious reasons this is called a "uniform" distribution.

Remark: Continuous probability distributions are, like discrete distributions, mathematical **models**. Thus, the distribution assumed for the spinner above is a model, though it seems likely it would be a good model for many real spinners.

Remark: It may seem paradoxical that $P(X = x) = 0$ for a continuous r.v. and yet we record the outcomes $X = x$ in real "experiments" with continuous variables. The catch is that all measurements have finite precision; they are in effect discrete. For example, the height $60 + \pi$ inches is within the

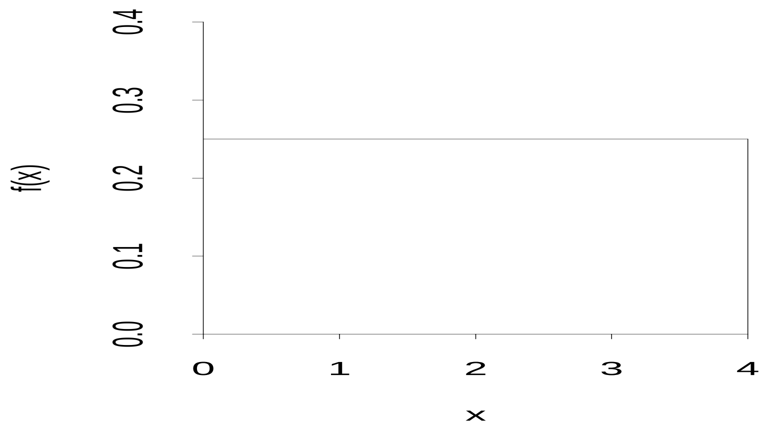


Figure 9.2: Uniform p.d.f.

range of the height X of people in a population but we could never observe the outcome $X = 60 + \pi$ if we selected a person at random and measured their height.

To summarize, in measurements we are actually observing something like

$$P(x - .5\Delta \leq X \leq x + .5\Delta),$$

where Δ may be very small, but not zero. The probability of this outcome is **not** zero: it is (approximately) $f(x)\Delta$.

We now consider a more complicated mathematical example of a continuous random variable Then we'll consider real problems that involve continuous variables. Remember that it is always a good idea to sketch or plot the p.d.f. $f(x)$ for a r.v.

Example:

Let $f(x) = \begin{cases} kx^2; & 0 < x \leq 1 \\ k(2-x); & 1 < x < 2 \\ 0; & \text{otherwise} \end{cases}$ be a p.d.f.

Find

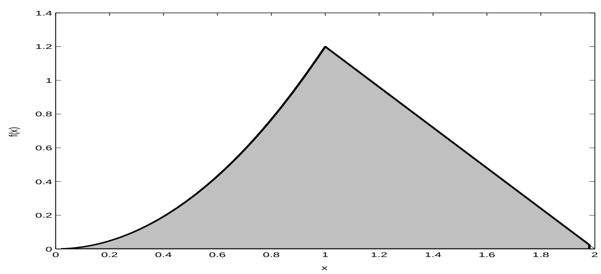
a) k

b) $F(x)$

c) $P\left(1/2 < X < 1\frac{1}{2}\right)$

Solution:

a) Set $\int_{-\infty}^{\infty} f(x)dx = 1$ to solve for k. When finding the area of a region bounded by different functions we split the integral into pieces.



(We normally wouldn't even write down the parts with $\int 0dx$)

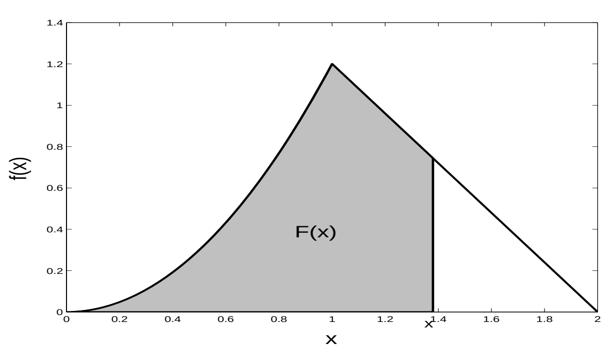
$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{0} 0dx + \int_0^1 kx^2dx + \int_1^2 k(2-x)dx + \int_2^{\infty} 0dx \\
&= 0 + k\int_0^1 x^2dx + k\int_1^2 (2-x)dx + 0 \\
&= k\frac{x^3}{3}\Big|_0^1 + k\left(2x - \frac{x^2}{2}\Big|_1^2\right) \\
&= \frac{5k}{6} = 1 \\
& \text{Therefore } k = \frac{6}{5}
\end{aligned}$$

b) Doing the easy pieces, which are often left out, first:

$$\begin{aligned}
F(x) &= 0 \text{ if } x \leq 0 \\
\text{and } F(x) &= 1 \text{ if } x \geq 2 \ \ \text{(since all probability is below } x \text{ if } x \text{ is a number above 2.)}
\end{aligned}$$

$$\begin{aligned}
&\text{For } 0 < x < 1 \quad P(X \leq x) = \int_0^x \tfrac{6}{5}z^2dz = \tfrac{6}{5} \times \tfrac{x^3}{3}\Big|_0^x = \tfrac{2x^3}{5} \\
&\text{For } 1 < x < 2, \quad P(X \leq x) = \int_0^1 \tfrac{6}{5}x^2dx + \int_1^x \tfrac{6}{5}(2-x)\,dx
\end{aligned}$$

(see shaded area below)



$$\begin{aligned} &= \tfrac{6}{5}\tfrac{x^3}{3}\ |_0^1 + \tfrac{6}{5}\left(2x - \tfrac{x^2}{2}\ |_1^x\right. \\ &= \tfrac{12x-3x^2-7}{5} \end{aligned}$$

i.e.

$$F(x) = \begin{cases} 0; & x \le 0 \\ 2x^3/5; & 0 < x \le 1 \\ \frac{12x-3x^2-7}{5}; & 1 < x < 2 \\ 1; & x \ge 2 \end{cases}$$

As a rough check, since for a continuous distribution there is no probability at any point, $F(x)$ should have the same value as we approach each boundary point from above and from below.
e.g.

$$\begin{aligned} &\text{As } x \to 0^+, \quad \tfrac{2x^3}{5} \to 0 \\ &\text{As } x \to 1^-, \quad \tfrac{2x^3}{5} \to \tfrac{2}{5} \\ &\text{As } x \to 1^+, \quad \tfrac{12x-3x^2-7}{5} \to \tfrac{2}{5} \\ &\text{As } x \to 2^-, \quad \tfrac{12x-3x^2-7}{5} \to 1 \end{aligned}$$

This quick check won't prove your answer is right, but will detect many careless errors.

c)

$$\begin{aligned} P\left(\tfrac{1}{2} < X < 1\tfrac{1}{2}\right) &= \int_{1/2}^{1\frac{1}{2}} f(x)dx \\ &\text{or} \quad F\left(1\tfrac{1}{2}\right) - F\left(\tfrac{1}{2}\right) \quad \text{(easier)} \\ &= \frac{12\left(\frac{3}{2}\right)-3\left(\frac{3}{2}\right)^2-7}{5} - \frac{2\left(\frac{1}{2}\right)^3}{5} = 4/5 \end{aligned}$$

Defined Variables or Change of Variable:

When we know the p.d.f. or c.d.f. for a continuous random variable X we sometimes want to find the p.d.f. or c.d.f. for some other random variable Y which is a function of X. The procedure for doing this is summarized below. It is based on the fact that the c.d.f. $F_Y(y)$ for Y equals $P(Y \le y)$, and this can be rewritten in terms of X since Y is a function of X. Thus:

1) Write the c.d.f. of Y as a function of X.

2) Use $F_X(x)$ to find $F_Y(y)$. Then if you want the p.d.f. $f_Y(y)$, you can differentiate the expression for $F_Y(y)$.

3) Find the range of values of y.

Example: In the earlier spinner example,

$$\begin{aligned} f(x) &= \tfrac{1}{4}; \quad 0 < x \leq 4 \\ \text{and } F(x) &= \tfrac{x}{4}; \quad 0 < x \leq 4 \end{aligned}$$

Let $Y = 1/X$. Find $f(y)$.

Solution:

$$\begin{aligned} F_Y(y) &= P\left(Y \leq y\right) = P\left(\tfrac{1}{X} \leq y\right) = P\left(X \geq \tfrac{1}{y}\right) \\ &= 1 - P\left(X < 1/y\right) \\ &= 1 - F_X\left(1/y\right) \quad \text{(this completes step (1))} \end{aligned}$$

For step (2), we can do either:

$$\begin{aligned} F_Y(y) &= 1 - \frac{\left(\frac{1}{y}\right)}{4} \quad \text{(substituting } \tfrac{1}{y} \text{ for } x \text{ in } F_X(x)) \\ &= 1 - \tfrac{1}{4y} \\ \text{Therefore } f_Y(y) &= \tfrac{d}{dy} F_Y(y) = \tfrac{1}{4y^2}; \quad \tfrac{1}{4} \leq y < \infty \end{aligned}$$

(As x goes from 0 to 4, $y = \frac{1}{x}$ goes between ∞ and $\frac{1}{4}$.)

$$\begin{aligned} \text{or : } \quad f_Y(y) &= \tfrac{d}{dy} F_Y(y) = \tfrac{d}{dy}\left(1 - F_X\left(1/y\right)\right) \\ &= -\tfrac{d}{dy} F_X(1/y) = -\tfrac{d}{dx} F_X(1/y) \tfrac{dx}{dy}\Big|_{x=1/y} \quad \text{(chain rule)} \\ &= -f_X\left(1/y\right)\left(-\tfrac{1}{y^2}\right) = -\tfrac{1}{4}\left(-\tfrac{1}{y^2}\right) = \tfrac{1}{4y^2}; \quad \tfrac{1}{4} \leq y < \infty \end{aligned}$$

Generally if $F_X(x)$ is known it is easier to substitute first, then differentiate. If $F_X(x)$ is in the form of an integral that can't be solved, it is usually easier to differentiate first, then substitute $f_X(x)$.

Extension of Expectation, Mean, and Variance to Continuous Distributions

Definition 32 *When X is continuous, we still define*

$$E(g(X)) = \int_{\text{all } x} g(x) f(x) dx.$$

With this definition, all of the earlier properties of expectation and variance still hold; for example with $\mu = E(X)$,

$$\sigma^2 = \text{Var}(X) = E[(X-\mu)^2] = E(X^2) - \mu^2.$$

(This definition can be justified by writing $\int_{\text{all } x} g(x)f(x)dx$ as a limit of a Riemann sum and recognizing the Riemann sum as being in the form of an expectation for discrete random variables.)

Example: In the spinner example with $f(x) = \frac{1}{4};\ \ 0 < x \le 4$

$$\begin{aligned}
\mu &= \textstyle\int_0^4 x\frac{1}{4}dx = \frac{1}{4}\left(\frac{x^2}{2}\right)|_0^4 = 2 \\
E\left(X^2\right) &= \textstyle\int_0^4 x^2\frac{1}{4}dx = \frac{1}{4}\left(\frac{x^3}{3}\right)|_0^4 = \frac{16}{3} \\
\sigma^2 &= E\left(X^2\right) - \mu^2 = \tfrac{16}{3} - 4 = 4/3
\end{aligned}$$

Example: Let X have p.d.f.

$$f(x) = \begin{cases} \frac{6x^2}{5}; & 0 < x \le 1 \\ \frac{6}{5}(2-x); & 1 < x < 2 \\ 0; & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned}
\mu &= \int_{\text{all } x} xf(x)dx = \int_0^1 x\frac{6}{5}x^2dx + \int_1^2 x\frac{6}{5}(2-x)dx \quad \text{(splitting the integral)} \\
&= \frac{6}{5}\left[\frac{x^4}{4}\,|_0^1 + \left(x^2 - \frac{x^3}{3}\right)|_1^2\right] = 11/10 \text{ or } 1.1 \\
E(X^2) &= \int_0^1 x^2\frac{6}{5}x^2dx + \int_1^2 x^2\frac{6}{5}(2-x)dx \\
&= \frac{6}{5}\left(\frac{x^5}{5}\,|_0^1 + 2\left(\frac{x^3}{3}\right)|_1^2 - \frac{x^4}{4}\,|_1^2\right) = \frac{67}{50} \\
\sigma^2 &= E\left(X^2\right) - \mu^2 = \frac{67}{50} - \left(\frac{11}{10}\right)^2 = \frac{13}{100} \text{ or } 0.13
\end{aligned}$$

Problems:

9.1.1 Let X have p.d.f. $f(x) = \begin{cases} kx^2; & -1 < x < 1. \\ 0; & \text{otherwise} \end{cases}$ Find

a) k

b) the c.d.f., $F(x)$

c) $P(-.1 < X < .2)$

d) the mean and variance of X.

e) let $Y = X^2$. Derive the p.d.f. of Y.

9.1.2 A continuous distribution has c.d.f. $F(x) = \frac{kx^n}{1+x^n}$ for $x > 0$, where n is a positive constant.

(a) Evaluate k.

(b) Find the p.d.f., $f(x)$.

(c) What is the median of this distribution? (The median is the value of x such that half the time we get a value below it and half the time above it.)

9.2 Continuous Uniform Distribution

Just as we did for discrete r.v.'s, we now consider some special types of continuous probability distributions. These distributions arise in certain settings, described below. This section considers what we call uniform distributions.

Physical Setup:
Suppose X takes values in some interval [a,b] (it doesn't actually matter whether interval is open or closed) with all subintervals of a fixed length being equally likely. Then X has a **continuous uniform distribution**. We write $X \sim U\,[a,b]$.

Illustrations:

(1) In the spinner example $X \sim U(0,4]$.

(2) Computers can generate a random number X which appears as though it is drawn from the distribution $U(0,1)$. This is the starting point for many computer simulations of random processes; an example is given below.

The probability density function and the cumulative distribution function:
Since all points are equally likely (more precisely, intervals contained in $[a,b]$ of a given length, say 0.01, all have the same probability), the probability density function must be a constant $f(x) = k$; $a \le x \le b$ for some constant k. To make $\int_a^b f(x)dx = 1$, we require $k = \frac{1}{b-a}$.

$$\text{Therefore } f(x) = \frac{1}{b-a} \text{ for } \quad a \le x \le b$$

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x > b \end{cases}$$

Mean and Variance:

$$\begin{aligned}
\mu &= \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\left(\frac{x^2}{2}\,|_a^b\right) = \frac{b^2-a^2}{2(b-a)} \\
&= \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2} \\
E(X^2) &= \int_a^b x^2\frac{1}{b-a}dx = \frac{1}{(b-a)}\left(\frac{x^3}{3}\,|_a^b\right) \\
&= \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)\left(b^2+ab+a^2\right)}{3(b-a)} = \frac{b^2+ab+a^2}{3} \\
\sigma^2 &= E\left(X^2\right) - \mu^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} \\
&= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}
\end{aligned}$$

Example: Suppose X has the continuous p.d.f.

$$f(x) = .1e^{-.1x} \qquad\qquad x > 0$$

(This is called an exponential distribution and is discussed in the next section. It is used in areas such as queueing theory and reliability.) We'll show that the new random variable

$$Y = e^{-.1X}$$

has a uniform distribution, $U(0,1)$. To see this, we follow the steps in Section 9.1:

$$\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(e^{-.1X} \le y) \\
&= P(X \ge -10\;ln\;y) \\
&= 1 - P(X < -10\;ln\;y) \\
&= 1 - F_X(-10\;ln\;y)
\end{aligned}$$

Since $F_X(x) = \int_0^x .1e^{-.1u}du = 1 - e^{-.1x}$ we get

$$\begin{aligned}
F_Y(y) &= 1 - (1 - e^{-.1(-10ln\;y)}) \\
&= y \;\text{ for } 0 < y < 1
\end{aligned}$$

(The range of Y is (0,1) since $X > 0$.) Thus $f_Y(y) = F_Y'(y) = 1(0 < y < 1)$ and so $Y \sim U(0,1)$.

Many computer software systems have "random number generator" functions that will simulate observations Y from a $U(0,1)$ distribution. (These are more properly called **pseudo-random number generators** because they are based on deterministic algorithms. In addition they give observations Y that have finite precision so they cannot be **exactly** like continuous $U(0,1)$ random variables. However, good generators give Y's that appear indistinguishable in most ways from $U(0,1)$ r.v.'s.) Given such a generator, we can also simulate r.v.'s X with the exponential distribution above by the following algorithm:

1. Generate $Y \sim U(0,1)$ using the computer random number generator.
2. Compute $X = -10\, ln\, Y$.

Then X has the desired distribution. This is a particular case of a method described in Section 9.4 for generating random variables from a general distribution. In R software the command $runif(n)$ produces a vector consisting of n independent $U(0,1)$ values.

Problem:

9.2.1 If X has c.d.f. $F(x)$, then $Y = F(X)$ has a uniform distribution on [0,1]. (Show this.) Suppose you want to simulate observations from a distribution with $f(x) = \frac{3}{2}x^2; \quad -1 < x < 1$, by using the random number generator on a computer to generate $U[0,1)$ numbers. What value would X take when you generated the random number .27125?

9.3 Exponential Distribution

The continuous random variable X is said to have an **exponential distribution** if its p.d.f. is of the form

$$f(x) = \lambda e^{-\lambda x} \qquad\qquad x > 0$$

where $\lambda > 0$ is a real parameter value. This distribution arises in various problems involving the time until some event occurs. The following gives one such setting.

Physical Setup: In a Poisson process for events in time let X be the length of time we wait for the first event occurrence. We'll show that X has an exponential distribution. (Recall that the <u>number</u> of occurrences in a fixed time has a Poisson distribution. The difference between the Poisson and exponential distributions lies in what is being measured.)

Illustrations:

(1) The length of time X we wait with a Geiger counter until the emission of a radioactive particle is recorded follows an exponential distribution.

(2) The length of time between phone calls to a fire station (assuming calls follow a Poisson process) follows an exponential distribution.

Derivation of the probability density function and the c.d.f.

$$\begin{aligned} F(x) = P(X \le x) &= P \text{ (time to } 1^{\text{st}} \text{ occurrence} \le x) \\ &= 1 - P \text{ (time to } 1^{\text{st}} \text{ occurrence} > x \text{)} \\ &= 1 - P \text{ (no occurrences in the interval } (0, x)) \end{aligned}$$

Check that you understand this last step. If the time to the first occurrence $> x$, there must be no occurrences in $(0, x)$, and vice versa.

We have now expressed $F(x)$ in terms of the number of occurrences in a Poisson process by time x. But the number of occurrences has a Poisson distribution with mean $\mu = \lambda x$, where λ is the average rate of occurrence.

$$\text{Therefore } F(x) = 1 - \frac{\mu^0 e^{-\mu}}{0!} = 1 - e^{-\mu}.$$

Since $\mu = \lambda x, F(x) = 1 - e^{-\lambda x};$ for $x > 0$. Thus

$$f(x) = \frac{d}{dx} F(x) = \lambda e^{-\lambda x}; \text{for } x > 0$$

which is the formula we gave above.

Alternate Form: It is common to use the parameter $\theta = 1/\lambda$ in the exponential distribution. (We'll see below that $\theta = E(X)$.) This makes

$$\begin{aligned} F(x) &= 1 - e^{-x/\theta} \\ \text{and } f(x) &= \tfrac{1}{\theta} e^{-x/\theta} \end{aligned}$$

Exercise:

Suppose trees in a forest are distributed according to a Poisson process. Let X be the distance from an arbitrary starting point to the nearest tree. The average number of trees per square metre is λ. Derive $f(x)$ the same way we derived the exponential p.d.f. You're now using the Poisson distribution in 2 dimensions (area) rather than 1 dimension (time).

Mean and Variance:

Finding μ and σ^2 directly involves integration by parts. An easier solution uses properties of **gamma functions**, which extends the notion of factorials beyond the integers to the positive real numbers.

Definition 33 ***The Gamma Function:*** *$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$ is called the gamma function of α, where $\alpha>0$.*

Note that α is 1 more than the power of x in the integrand. e.g. $\int_0^\infty x^4e^{-x}dx=\Gamma(5)$. There are 3 properties of gamma functions which we'll use.

1. $\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$ for $\alpha>1$
   Proof: Using integration by parts,
   $$\int_0^\infty x^{\alpha-1}e^{-x}dx=-x^{\alpha-1}e^{-x}|_0^\infty+(\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}dx$$
   and provided that $\alpha>1$, $x^{\alpha-1}e^{-x}|_0^\infty=0$. Therefore
   $$\int_0^\infty x^{\alpha-1}e^{-x}dx=(\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}dx$$
2. $\Gamma(\alpha)=(\alpha-1)!$ if α is a positive integer.
   Proof: It is easy to show that $\Gamma(1)=1$. Using property 1. repeatedly, we obtain $\Gamma(2)=1\Gamma(1)=1,\Gamma(3)=2\Gamma(2)=2!,\Gamma(4)=3\Gamma(3)=3!$,etc.
   Generally,. $\Gamma(n+1)=n!$ for integer n.
3. $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$
   (This can be proved using double integration.)

Returning to the exponential distribution:

$$\mu=\int_0^\infty x\frac{1}{\theta}e^{-x/\theta}dx$$

Let $y=\frac{x}{\theta}$. Then $dx=\theta dy$ and

$$\begin{aligned}\mu &= \int_0^\infty ye^{-y}\theta dy=\theta\int_0^\infty y^1e^{-y}dy=\theta\Gamma(2)\\ &= \theta\end{aligned}$$

Note: Read questions carefully. If you're given the average **rate** of occurrence in a Poisson process, that is λ. If you're given the average **time** you wait for an occurrence, that is θ.
To get $\sigma^2=\mathrm{Var}(X)$, we first find

$$\begin{aligned}E\left(X^2\right) &= \int_0^\infty x^2\tfrac{1}{\theta}e^{-x/\theta}dx\\ &= \int_0^\infty \theta^2y^2\tfrac{1}{\theta}e^{-y}\theta dy=\theta^2\int_0^\infty y^2e^{-y}dy\\ &= \theta^2\Gamma(3)=2!\theta^2=2\theta^2\\ \text{Therefore } \sigma^2 &= E\left(X^2\right)-\mu^2=2\theta^2-\theta^2=\theta^2\end{aligned}$$

Example:

Suppose #7 buses arrive at a bus stop according to a Poisson process with an average of 5 buses per hour. (i.e. $\lambda = 5$/hr. So $\theta = \frac{1}{5}$ hr. or 12 min.) Find the probability (a) you have to wait longer than 15 minutes for a bus (b) you have to wait more than 15 minutes longer, having already been waiting for 6 minutes.

Solution:

a) $P(X > 15) = 1 - P(X \leq 15) = 1 - F(15)$
= $1 - \left(1 - e^{-15/12}\right) = e^{-1.25} = .2865$

b) If X is the total waiting time, the question asks for the probability

$$P(X > 21 | X > 6) = \frac{P(X > 21 \text{ and } X > 6)}{P(X > 6)} = \frac{P(X > 21)}{P(X > 6)}$$

$$= \frac{1 - \left(1 - e^{-21/12}\right)}{1 - \left(1 - e^{-6/12}\right)} = \frac{e^{-21/12}}{e^{-6/12}} = e^{-15/12} = e^{-1.25} = .2865$$

Does this surprise you? The fact that you're already waited 6 minutes doesn't seem to matter. This illustrates the "memoryless property" of the exponential distribution:

$$P(X > a + b | X > b) = P(X > a)$$

Fortunately, buses don't follow a Poisson process so this example needn't cause you to stop using the bus.

Problems:

9.3.1 In a bank with on-line terminals, the time the system runs between disruptions has an exponential distribution with mean θ hours. One quarter of the time the system shuts down within 8 hours of the previous disruption. Find θ.

9.3.2 Flaws in painted sheets of metal occur over the surface according to the conditions for a Poisson process, at an intensity of λ per m^2. Let X be the distance from an arbitrary starting point to the second closest flaw. (Assume sheets are of infinite size!)

(a) Find the p.d.f., $f(x)$.

(b) What is the average distance to the second closest flaw?

9.4 A Method for Computer Generation of Random Variables.

Most computer software has a built-in "pseudo-random number generator" that will simulate observations U from a $U(0,1)$ distribution, or at least a reasonable approximation to this uniform distribution. If we wish a random variable with a non-uniform distribution, the standard approach is to take a suitable function of U. By far the simplest and most common method for generating non-uniform variates is based on the inverse cumulative distribution function. For arbitrary c.d.f. $F(x)$, define $F^{-1}(y)$ =min $\{x; F(x) \geq y\}$. This is a real inverse (i.e. $F(F^{-1}(y)) = F^{-1}(F(y)) = y$) in the case that the c.d.f. is continuous and strictly increasing, so for example for a continuous distribution. However, in the more general case of a possibly discontinuous non-decreasing c.d.f. (such as the c.d.f. of a discrete distribution) the function continues to enjoy at least some of the properties of an inverse. F^{-1} is useful for generating a random variables having c.d.f. $F(x)$ from U, a uniform random variable on the interval $[0,1]$.

Theorem 34 *If F is an arbitrary c.d.f. and U is uniform on $[0,1]$ then the random variable defined by $X = F^{-1}(U)$ has c.d.f. $F(x)$.*

Proof:
The proof is a consequence of the fact that

$$[U < F(x)] \subset [X \leq x] \subset [U \leq F(x)] \text{ for all } x.$$

You can check this graphically be checking, for example, that if $[U < F(x)]$ then $[F^{-1}(U) \leq x]$ (this confirms the left hand "$\subset''$). Taking probabilities on all sides of this, and using the fact that $P[U \leq F(x)] = P[U < F(x)] = F(x)$, we discover that $P[X \leq x] = F(x)$.
The relation $X = F^{-1}(U)$ implies that $F(X) \geq U$ and for any point $z < X$, $F(z) < U$. For example, for the rather unusual looking piecewise linear cumulative distribution function in Figure 9.3, we find the solution $X = F^{-1}(U)$ by drawing a horizontal line at U until it strikes the graph of the c.d.f. (or where the graph would have been if we had joined the ends at the jumps) and then X is the $x-coordinate$ of this point. This is true in general, X is the coordinate of the point where a horizontal line first strikes the graph of the c.d.f. We provide one simple example of generating random variables by this method, for the geometric distribution.

Example: A geometric random number generator

For the Geometric distribution, the cumulative distribution function is given by

$$F(x) = 1 - (1-p)^{x+1}, \quad \text{for } x = 0, 1, 2, ...$$

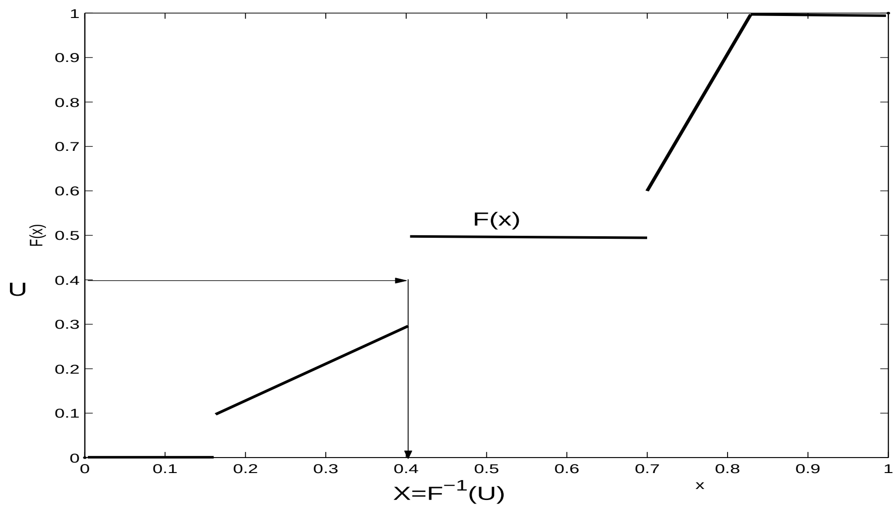


Figure 9.3: Inverting a c.d.f.

Then if U is a uniform random number in the interval $[0,1]$, we seek an integer X such that

$$F(X-1) < U \le F(X)$$

(you should confirm that this is the value of X at which the above horizontal line strikes the graph of the c.d.f) and solving these inequalities gives

$$\begin{aligned} 1-(1-p)^X &< U \le 1-(1-p)^{X+1} \\ (1-p)^X &> 1-U \ge (1-p)^{X+1} \\ X\ln(1-p) &> \ln(1-U) \ge (X+1)\ln(1-p) \\ X &< \frac{\ln(1-U)}{\ln(1-p)} \le X+1 \end{aligned}$$

so we compute the value of

$$\frac{\ln(1-U)}{\ln(1-p)}$$

and round down to the next lower integer.

Exercise: An exponential random number generator.

Show that the inverse transform method above results in the generator for the exponential distribution

$$X = -\frac{1}{\lambda}\ln(1-U)$$

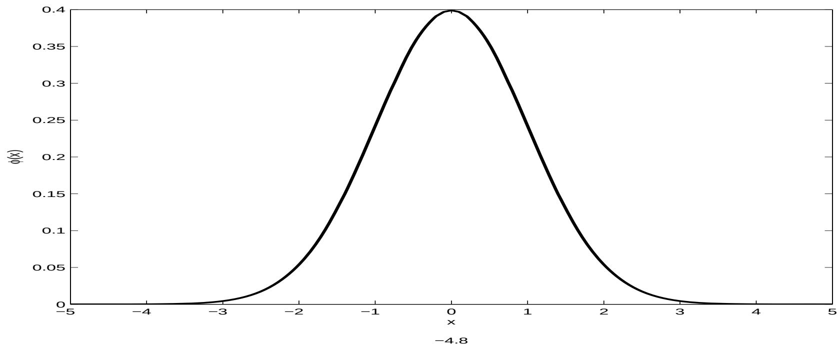


Figure 9.4: The standard normal probability density function

9.5 Normal Distribution

Physical Setup:

A random variable X defined on $(-\infty, \infty)$ has a normal distribution if it has probability density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

where $-\infty < \mu < \infty$ and $\sigma > 0$ are parameters. It turns out (and is shown below) that $E(X) = \mu$ and Var$(X) = \sigma^2$ for this distribution; that is why its p.d.f. is written using the symbols μ and σ. We write

$$X \sim N(\mu, \sigma^2)$$

to denote that X has a normal distribution with mean μ and variance σ^2 (standard deviation σ).

The normal distribution is the most widely used distribution in probability and statistics. Physical processes leading to the normal distribution exist but are a little complicated to describe. (For example, it arises in physics via statistical mechanics and maximum entropy arguments.) It is used for many processes where X represents a physical dimension of some kind, but also in many other settings. We'll see other applications of it below. The shape of the p.d.f. $f(x)$ above is what is often termed a "bell shape" or "bell curve", symmetric about 0 as shown in Figure 9.4.(you should be able to verify the shape without graphing the function)

Illustrations:

(1) Heights or weights of males (or of females) in large populations tend to follow normal distributions.

(2) The logarithms of stock prices are often assumed to be normally distributed.

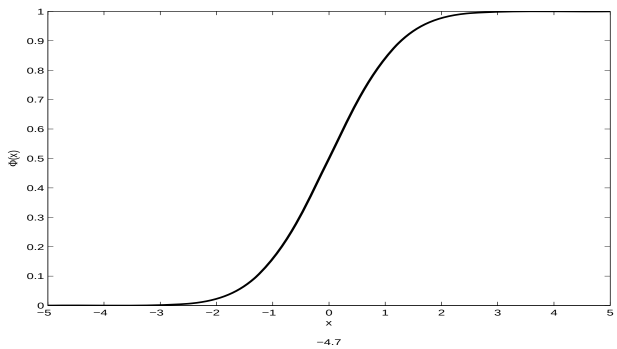


Figure 9.5: The standard normal c.d.f.

The cumulative distribution function: The c.d.f. of the normal distribution $N(\mu, \sigma^2)$ is

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} dy.$$

as shown in Figure 9.5. This integral cannot be given a simple mathematical expression so numerical methods are used to compute its value for given values of x, μ and σ. This function is included in many software packages and some calculators.

In the statistical packages R and S-Plus we get $F(x)$ above using the function $pnorm(x, \mu, \sigma)$. Before computers, people needed to produce tables of probabilities $F(x)$ by numerical integration, using mechanical calculators. Fortunately it is necessary to do this only for a single normal distribution: the one with $\mu = 0$ and $\sigma = 1$. This is called the **“standard" normal distribution** and denoted $N(0, 1)$.

It is easy to see that if $X \sim N(\mu, \sigma^2)$ then the “new" random variable $Z = (X - \mu)/\sigma$ is distributed as $Z \sim N(0, 1)$. (Just use the change of variables methods in Section 9.1.) We’ll use this to compute $F(x)$ and probabilities for X below, but first we show that $f(x)$ integrates to 1 and that $E(X) = \mu$ and

$\text{Var}(X) = \sigma^2$. For the first result, note that

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \text{ where we let } z = (x-\mu)/\sigma) \\
&= 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y} \frac{dy}{\sqrt{2}y^{\frac{1}{2}}} \text{ Note:} y = \frac{1}{2}z^2;\ dy = \frac{dy}{\sqrt{2}y^{\frac{1}{2}}} \\
&= \frac{1}{\sqrt{\pi}} \int_0^{\infty} y^{-\frac{1}{2}} e^{-y} dy \\
&= \frac{1}{\sqrt{\pi}} \Gamma(\frac{1}{2}) \qquad \text{(gamma function)} \\
&= 1
\end{aligned}$$

Mean, Variance, Moment generating function: Recall that an odd function, $f(x)$, has the property that $f(-x) = -f(x)$. If $f(x)$ is an odd function then $\int_{-\infty}^{\infty} f(x)dx = 0$, provided the integral exists. Consider

$$E(X-\mu) = \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Let $y = x - \mu$. Then

$$E(X-\mu) = \int_{-\infty}^{\infty} y \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} dy,$$

where $y\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{y^2}{2\sigma^2}}$ is an odd function so that $E(X-\mu) = 0$. But since $E(X-\mu) = E(X) - \mu$, this implies

$$E(X) = \mu,$$

and so μ is the mean. To obtain the variance,

$$\begin{aligned}
\text{Var}(X) &= E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= 2\int_{\mu}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \ \ (\text{ since the function is symmetric about } \mu).
\end{aligned}$$

We can obtain a gamma function by letting $y = \frac{(x-\mu)^2}{2\sigma^2}$.

$$\begin{aligned}
\text{Then } (x-\mu)^2 &= 2\sigma^2 y \\
(x-\mu) &= \sigma\sqrt{2y} \ \ (x > \mu, \text{ so the positive root is taken}) \\
dx &= \frac{\sigma\sqrt{2}dy}{2\sqrt{y}} = \frac{\sigma}{\sqrt{2y}} dy
\end{aligned}$$

Then

$$
\begin{aligned}
\text{Var}(X) &= 2\int_0^\infty \left(2\sigma^2 y\right)\frac{1}{\sigma\sqrt{2\pi}}e^{-y}\left(\frac{\sigma}{\sqrt{2y}}dy\right)\\
&= \frac{2\sigma^2}{\sqrt{\pi}}\int_0^\infty y^{1/2}e^{-y}dy = \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}}\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \frac{2\sigma^2\left(\frac{1}{2}\right)\sqrt{\pi}}{\sqrt{\pi}}\\
&= \sigma^2
\end{aligned}
$$

and so σ^2 is the variance. We now find the moment generating function of the $N(\mu,\sigma^2)$ distribution. If X has the $N(\mu,\sigma^2)$ distribution, then

$$
\begin{aligned}
M_X(t) = E(e^{Xt}) &= \int_{-\infty}^\infty e^{xt}f(x)dx\\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty e^{xt}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2\sigma^2}(x^2-2\mu x-2xt\sigma^2+\mu^2)}dx\\
&= \frac{e^{\mu t+\sigma^2t^2/2}}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2\sigma^2}\{x^2-2(\mu+t\sigma^2)x+(\mu+t\sigma^2)^2\}}dx\\
&= \frac{e^{\mu t+\sigma^2t^2/2}}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2\sigma^2}\{x-(\mu+t\sigma^2)\}^2}dx\\
&= e^{\mu t+\sigma^2t^2/2}
\end{aligned}
$$

where the last step follows since

$$
\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2\sigma^2}\{x-(\mu+t\sigma^2)\}^2}dx
$$

is just the integral of a $N(\mu+t\sigma^2,\sigma^2)$ probability density function and is therefore equal to one. This confirms the values we already obtained for the mean and the variance of the normal distribution

$$
\begin{aligned}
M_X'(0) &= e^{\mu t+\sigma^2t^2/2}(\mu+t\sigma^2)|_{t=0} = \mu\\
M_X''(0) &= \mu^2+\sigma^2 = E(X^2)
\end{aligned}
$$

from which we obtain

$$
Var(X) = \sigma^2.
$$

Finding Normal Probabilities Via $N(0,1)$ Tables As noted above, $F(x)$ does not have an explicit closed form so numerical computation is needed. The following result shows that if we can compute the c.d.f. for the standard normal distribution $N(0,1)$, then we can compute it for any other normal distribution $N(\mu,\sigma^2)$ as well.

Theorem 35 *Let $X \sim N(\mu, \sigma^2)$ and define $Z = (X - \mu)/\sigma$. Then $Z \sim N(0, 1)$ and*

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= F_Z(\tfrac{x-\mu}{\sigma}). \end{aligned}$$

Proof: The fact that $Z \sim N(0, 1)$ has p.d.f.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

follows immediately by change of variables. Alternatively, we can just note that

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \qquad (z = \frac{x-\mu}{\sigma}) \\ &= F_Z(\frac{x-\mu}{\sigma}) \qquad \square \end{aligned}$$

A table of probabilities $F_Z(z) = P(Z \le z)$ is given on the last page of these notes. A space-saving feature is that only the values for $z > 0$ are shown; for negative values we use the fact that $N(0, 1)$ p.d.f. is symmetric about 0. The following examples illustrate how to get probabilities for Z using the tables.

Examples: Find the following probabilities, where $Z \sim N(0, 1)$.

(a) $P(Z \le 2.11)$

(b) $P(Z \le 3.50)$

(c) $P(Z > 1.06)$

(d) $P(Z < -1.06)$

(e) $P(-1.06 < Z < 2.11)$

Solution:

a) Look up 2.11 in the table by going down the left column to 2.1 then across to the heading .01. We find the number .9826. Then $P(Z \le 2.11) = F(2.11) = .9826$. See Figure 9.6.

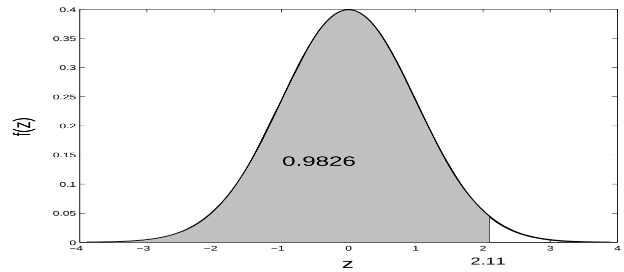


Figure 9.6:

b) $P(Z \leq 3.40) = F(3.40) = .9996631$

c) $P(Z > 1.06) = 1 - P(Z \leq 1.06) = 1 - F(1.06) = 1 - .8554 = .1446$

d) Now we have to use sy

$P(Z < -1.0$ $06) = .1446$

See Figure 9.7.

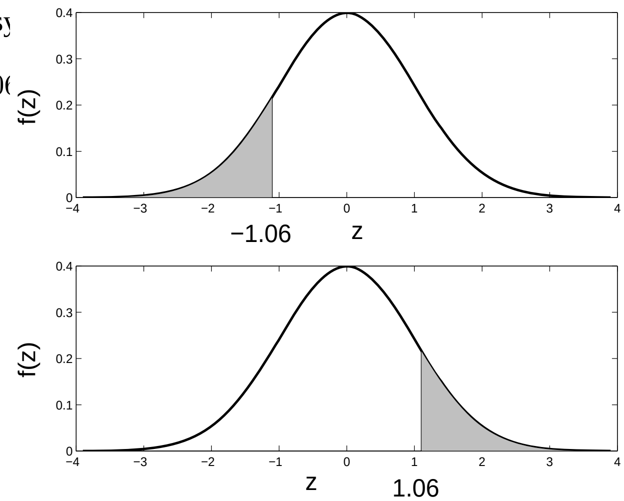


Figure 9.7:

e) $P(-1.06 < Z < 2.11) = F(2.11) - F(-1.06)$
$= F(2.11) - P(Z \leq -1.06) = F(2.11) - [1 - F(1.06)]$
$= .9826 - (1 - .8554) = .8380$

In addition to using the tables to find the probabilities for given numbers, we sometimes are given the probabilities and asked to find the number. With R or S-Plus software , the function qnorm (p, μ, σ) gives the 100 p-th percentile (where $0 < p < 1$). We can also use tables to find desired values.

Examples:

a) Find a number c such that $P(Z < c) = .85$

b) Find a number d such that $P(Z > d) = .90$

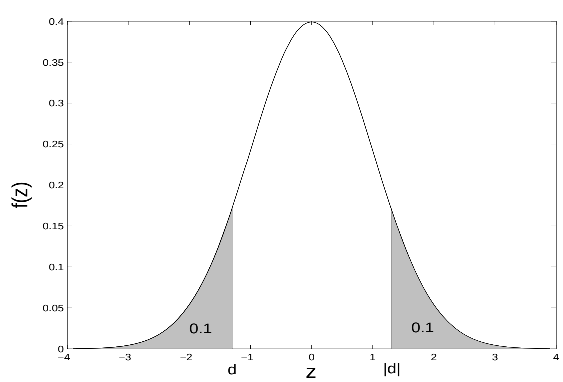
c) Find a number b such that $P(-b < Z < b) = .95$

Solutions:

a) We can look in the body of the table to get an entry close to .8500. This occurs for z between 1.03 and 1.04; $z = 1.04$ gives the closest value to .85. For greater accuracy, the table at the bottom of the last page is designed for finding numbers, given the probability. Looking beside the entry .85 we find $z = 1.0364$.

b) Since $P(Z > d) = .90$ we have $F(d) = P(Z \leq d) = 1 - P(Z > d) = .10$. There is no entry for which $F(z) = .10$ so we again have to use symmetry, since d will be negative.

$P(Z \leq d) = P(Z \geq |d|)$
$= 1 - F(|d|) = .10$
Therefore $F(|d|) = .90$
Therefore $|d| = 1.2816$
Therefore $d = -1.2816$



The key to this solution lies in recognizing that d will be negative. If you can picture the situation it will probably be easier to handle the question than if you rely on algebraic manipulations.

Exercise: Will a be positive or negative if $P(Z > a) = .05$? What if $P(Z < a) = .99$?

c) If $P(-b < Z < b) = .95$ we again use symmetry.

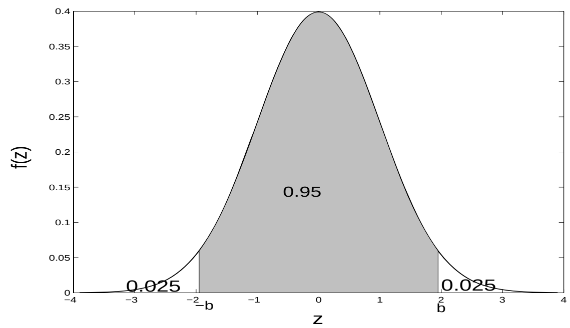


Figure 9.8:

The probability outside the interval $(-b, b)$ must be .05, and this is evenly split between the area above b and the area below $-b$.

$$\begin{aligned}\text{Therefore } P(Z < -b) &= P(Z > b) = .025 \\ \text{and } P(Z \le b) &= .975\end{aligned}$$

Looking in the table, $b = 1.96$.

To find $N(\mu, \sigma^2)$ probabilities in general, we use the theorem given earlier, which implies that if $X \sim N(\mu, \sigma^2)$ then

$$\begin{aligned}P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)\end{aligned}$$

where $Z \sim N(0, 1)$.

Example: Let $X \sim N(3, 25)$.

a) Find $P(X < 2)$

b) Find a number c such that $P(X > c) = .95$.

Solution:

a)

$$P(X<2)=P\left(\frac{X-\mu}{\sigma}<\frac{2-3}{5}\right)=P(Z<-.20)=1-P(Z<.20)$$
$$=1-F(.20)=1-.5793=.4207$$

b)

$$P(X>c)=P\left(\frac{X-\mu}{\sigma}>\frac{c-3}{5}\right)=P\left(Z>\frac{c-3}{5}\right)=.95$$

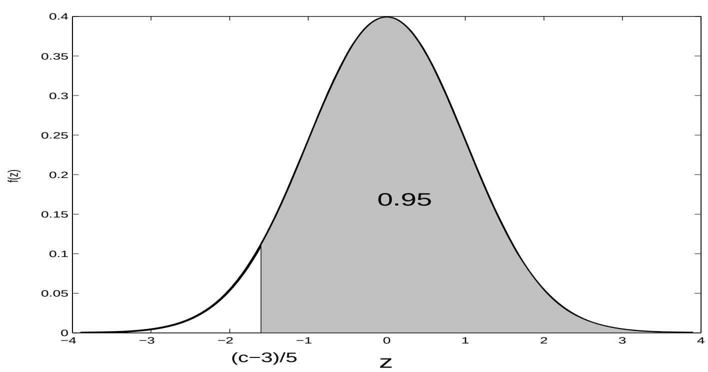


Figure 9.9:

$$\begin{aligned}\text{Therefore } \tfrac{c-3}{5} &= -1.6449\\ \text{and } c &= -5.2245\end{aligned}$$

Gaussian Distribution: The normal distribution is also known as the Gaussian[11] distribution. The notation $X \sim G(\mu,\sigma)$ means that X has Gaussian (normal) distribution with mean μ and standard deviation σ. So, for example, if $X \sim N(1,4)$ then we could also write $X \sim G(1,2)$.

Example: The heights of adult males in Canada are close to normally distributed, with a mean of 69.0 inches and a standard deviation of 2.4 inches. Find the 10th and 90th percentiles of the height distribution. (Recall that the a-th percentile is such that a% of the population has height less that this value.)

[11] After Johann Carl Friedrich Gauss (1777-1855), a German mathematician, physicist and astronomer, discoverer of Bode's Law, the Binomial Theorem and a regular 17-gon. He discovered the prime number theorem while an 18 year-old student and used least-squares (what is called statistical regression in most statistics courses) to predict the position of Ceres.

Solution: We are being told that if X is the height of a randomly selected Canadian adult male, then $X \sim G(69.0, 2.4)$, or equivalently $X \sim N(69.0, 5.76)$. To find the 90th percentile c, we use

$$\begin{aligned} P(X \leq c) &= P\left(\frac{X - 69.0}{2.4} \leq \frac{c - 69.0}{2.4}\right) \\ &= P\left(Z \leq \frac{c - 69.0}{2.4}\right) = .90 \end{aligned}$$

From the table we see $P(Z \leq 1.2816) = .90$ so we need

$$\frac{c - 69.0}{2.4} = 1.2816,$$

which gives $x = 72.08$ inches. Similarly, to find c such that $P(X \leq c) = .10$ we find that $P(Z \leq -1.2816) = .10$, so we need

$$\frac{c - 69.0}{2.4} - -1.2816,$$

or $c = 65.92$ inches, as the 10th percentile.

Linear Combinations of Independent Normal Random Variables

Linear combinations of normal r.v.'s are important in many applications. Since we have not covered continuous multivariate distributions, we can only quote the second and third of the following results without proof. The first result follows easily from the change of variables method.

1. Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, where a and b are constant real numbers. Then $Y \sim N(a\mu + b, a^2\sigma^2)$
2. Let $X \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y \sim N\left(\mu_2, \sigma_2^2\right)$ be independent, and let a and b be constants.
   Then $aX + bY \sim N\left(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2\right)$.
   In general if $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ are independent and a_i are constants,
   then $\sum a_i X_i \sim N\left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2\right)$.
3. Let $X_1, X_2, \cdots, X_n$ be independent $N\left(\mu, \sigma^2\right)$ random variables.
   Then $\sum X_i \sim N\left(n\mu, n\sigma^2\right)$ and $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$.

   Actually, the only new result here is that the distributions are normal. The means and variances of linear combinations of r.v.'s were previously obtained in section 8.3.

Example: Let $X \sim N(3, 5)$ and $Y \sim N(6, 14)$ be independent. Find $P\left(X > Y\right)$.

Solution: Whenever we have variables on both sides of the inequality we should collect them on one side, leaving us with a linear combination.

$$\begin{aligned} &P\left(X > Y\right) = P\left(X - Y > 0\right) \\ &X - Y \sim N\left(3 - 6, 5 + 14\right) \text{ i.e. } N(-3, 19) \\ &P\left(X - Y > 0\right) = P\left(Z > \tfrac{0 - (-3)}{\sqrt{19}} = .69\right) = 1 - F(.69) = .2451 \end{aligned}$$

Example: Three cylindrical parts are joined end to end to make up a shaft in a machine; 2 type A parts and 1 type B. The lengths of the parts vary a little, and have the distributions: $A \sim N(6, .4)$ and $B \sim N(35.2, .6)$. The overall length of the assembled shaft must lie between 46.8 and 47.5 or else the shaft has to be scrapped. Assume the lengths of different parts are independent. What percent of assembled shafts have to be scrapped?

Exercise: Why would it be wrong to represent the length of the shaft as 2A + B? How would this length differ from the solution given below?

Solution: Let L, the length of the shaft, be $L = A_1 + A_2 + B$.

Then

$$L \sim N(6 + 6 + 35.2,\ \ .4 + .4 + .6) = N(47.2, 1.4)$$

and so

$$\begin{aligned} P(46.8 < L < 47.5) &= P\left(\tfrac{46.8-47.2}{\sqrt{1.4}} < Z < \tfrac{47.5-47.2}{\sqrt{1.4}}\right) \\ &= P(-.34 < Z < .25) = .2318. \end{aligned}$$

i.e. 23.18% are acceptable and 76.82% must be scrapped. Obviously we have to find a way to reduce the variability in the lengths of the parts. This is a common problem in manufacturing.

Exercise: How could we reduce the percent of shafts being scrapped? (What if we reduced the variance of A and B parts each by 50%?)

Example: The heights of adult females in a large population is well represented by a normal distribution with mean 64 in. and variance 6.2 in^2.

(a) Find the proportion of females whose height is between 63 and 65 inches.

(b) Suppose 10 women are randomly selected, and let $\bar{X}$ be their average height (i.e. $\bar{X} = \sum_{i=1}^{10} X_i/10$, where $X_1, \ldots, X_{10}$ are the heights of the 10 women). Find $P(63 \le \bar{X} \le 65)$.

(c) How large must n be so that a random sample of n women gives an average height $\bar{X}$ so that $P(|\bar{X} - \mu| \le 1) \ge .95$?

Solution:

(a) $X \sim N(64, 6.2)$ so for the height X of a random woman,

$$
\begin{aligned}
P(63 \leq X \leq 65) = P\left(\frac{63-64}{\sqrt{6.2}} \leq \frac{X-\mu}{\sigma} \leq \frac{65-64}{\sqrt{6.2}}\right) \\
= P(-0.402 \leq Z \leq 0.402) \\
= 0.312
\end{aligned}
$$

(b) $\bar{X} \sim N\left(64, \frac{6.2}{10}\right)$ so

$$
\begin{aligned}
P(63 \leq \bar{X} \leq 65) &= P\left(\frac{63-64}{\sqrt{.62}} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \frac{65-64}{\sqrt{.62}}\right) \\
&= P(-.1.27 \leq Z \leq 1.27) \\
&= 0.796
\end{aligned}
$$

(c) If $\bar{X} \sim N\left(64, \frac{6.2}{n}\right)$ then

$$
\begin{aligned}
P(|\bar{X}-\mu| \leq 1) &= P(|\bar{X}-64| \leq 1) \\
&= P(63 \leq \bar{X} \leq 65) \\
&= P\left(\frac{63-64}{\sqrt{6.2/n}} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \frac{65-64}{\sqrt{6.2/n}}\right) \\
&= P(-0.402\sqrt{n} \leq Z \leq 0.402\sqrt{n}) = .95
\end{aligned}
$$

iff $.402\sqrt{n} = 1.96$. (This is because $P(-1.96 \leq Z \leq 1.96) = .95)$. So $P(|\bar{X}-64| \leq 1) \geq .95$ iff $0.402\sqrt{n} \geq 1.96$ which is true if $n \geq (1.96/.402)^2$, or $n \geq 23.77$. Thus we require $n \geq 24$ since n is an integer.

Remark: This shows that if we were to select a random sample of $n = 24$ persons, then their average height $\bar{X}$ would be with 1 inch of the average height μ of the whole population of women. So if we did not know μ then we could estimate it to within ± 1 inch (with probability .95) by taking this small a sample.

Exercise: Find how large n would have to be to make $P(|\bar{X}-\mu| \leq .5) \geq .95$.

These ideas form the basis of statistical sampling and estimation of unknown parameter values in populations and processes. If $X \sim N(\mu, \sigma^2)$ and we know roughly what σ is, but don't know μ, then we can use the fact that $\bar{X} \sim N(\mu, \sigma^2/n)$ to find the probability that the mean $\bar{X}$ from a sample of size n will be within a given distance of μ.

Problems:

9.5.1 Let $X \sim N(10, 4)$ and $Y \sim N(3, 100)$ be independent. Find the probability

a) $8.4 < X < 12.2$

b) $2Y > X$

c) $\overline{Y} < 0$ where $\overline{Y}$ is the sample mean of 25 independent observations on Y.

9.5.2 Let X have a normal distribution. What percent of the time does X lie within one standard deviation of the mean? Two standard deviations? Three standard deviations?

9.5.3 Let $X \sim N(5,4)$. An independent variable Y is also normally distributed with mean 7 and standard deviation 3. Find:

(a) The probability $2X$ differs from Y by more than 4.

(b) The minimum number, n, of independent observations needed on X so that $P\left(|\overline{X}-5| < 0.1\right) \geq .98$. $\quad$ ($\overline{X} = \sum\limits_{i=1}^{n} X_i/n$ is the sample mean)

9.6 Use of the Normal Distribution in Approximations

The normal distribution can, under certain conditions, be used to approximate probabilities for linear combinations of variables having a non-normal distribution. This remarkable property follows from an amazing result called the central limit theorem. There are actually several versions of the central limit theorem. The version given below is one of the simplest.

Central Limit Theorem (CLT):

The major reason that the normal distribution is so commonly used is that it tends to approximate the distribution of sums of random variables. For example, if we throw n fair dice and S_n is the sum of the outcomes, what is the distribution of S_n? The tables below provide the number of ways in which a given value can be obtained. The corresponding probability is obtained by dividing by 6^n. For example on the throw of $n = 1$ dice the probable outcomes are 1,2,...,6 with probabilities all $1/6$ as indicated in the first panel of the histogram in Figure 9.10.

If we sum the values on two fair dice, the possible outcomes are the values 2,3,...,12 as shown in the following table and the probabilities are the values below:

Values	2	3	4	5	6	7	8	9	10	11	12
Probabilities$\times 36$	1	2	3	4	5	6	5	4	3	2	1

The probability histogram of these values is shown in the second panel. Finally for the sum of the values on three independent dice, the values range from 3 to 18 and have probabilities which, when multiplied by 6^3 result in the values

1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1

to which we can fit three separate quadratic functions one in the middle region and one in each of the

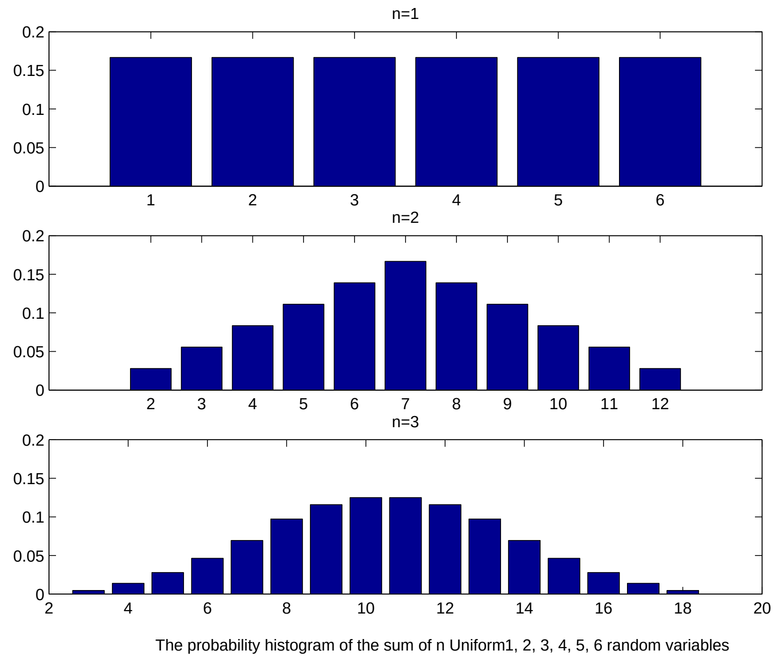


Figure 9.10: The probability histogram of the sum of n discrete uniform {1,2,3,4,5,6}Random variables

two tails. The histogram of these values shown in the third panel of Figure 9.10. and already resembles a normal probability density function.In general, these distributions show a simple pattern. For $n = 1$, the probability function is a constant (polynomial degree 0). For $n = 2$, two linear functions spliced together. For $n = 3$, the histogram can be constructed from three quadratic pieces (polynomials of degree $n - 1$). These probability histograms rapidly approach the shape of the normal probability density function, as is the case with the sum or the average of independent random variables from most distributions. You can simulate the throws of any number of dice and illustrate the behaviour of the sums on at the url *http://www.math.csusb.edu/faculty/stanton/probstat/clt.html.*
Let $X_1, X_2, \cdots, X_n$ be independent random variables all having the same distribution, with mean μ and variance σ^2. Then as $n \to \infty$,

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right) \tag{9.5}$$

and

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \tag{9.6}$$

This is actually a rough statement of the result since, as $n \to \infty$, both the $N\left(n\mu, n\sigma^2\right)$ and $N\left(\mu, \sigma^2/n\right)$

distributions fail to exist. (The former because both $n\mu$ and $n\sigma^2 \to \infty$, the latter because $\frac{\sigma^2}{n} \to 0$.) A precise version of the results is:

Theorem 36 *If $X_1, X_2, \cdots, X_n$ be independent random variables all having the same distribution, with mean μ and variance σ^2, then as $n \to \infty$, the cumulative distribution function of the random variable*

$$\frac{\sum X_i - n\mu}{\sigma\sqrt{n}}$$

approaches the $N(0,1)$ c.d.f. Similarly, the c.d.f. of

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the standard normal c.d.f.

Although this is a theorem about limits, we will use it when n is large, but finite, to approximate the distribution of $\sum X_i$ or $\overline{X}$ by a normal distribution, so the rough version of the theorem in (9.5) and (9.6) is adequate for our purposes.

Notes:

(1) This theorem works for essentially all distributions which X_i could have. The only exception occurs when X_i has a distribution whose mean or variance don't exist. There are such distributions, but they are rare.

(2) We will use the Central Limit Theorem to approximate the distribution of sums $\sum\limits_{i=1}^{n} X_i$ or averages $\bar{X}$. The accuracy of the approximation depends on n (bigger is better) and also on the actual distribution the X_i's come from. The approximation works better for small n when X_i's p.d.f. is close to symmetric.

(3) If you look at the section on linear combinations of independent normal random variables you will find two results which are very similar to the central limit theorem. These are:

For $X_1, \cdots, X_n$ independent and $N\left(\mu, \sigma^2\right)$, $\sum X_i \sim N\left(n\mu, n\sigma^2\right)$, and $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$.

Thus, if the X_i's themselves have a normal distribution, then $\sum X_i$ and $\overline{X}$ have exactly normal distributions for all values of n. If the X_i's do not have a normal distribution themselves, then $\sum X_i$ and $\overline{X}$ have approximately normal distributions when n is large. From this distinction you should be able to guess that if the X_i's distribution is somewhat normal shaped the approximation will be good for smaller values of n than if the X_i's distribution is very non-normal in shape. (This is related to the second remark in (2)).

Example: Hamburger patties are packed 8 to a box, and each box is supposed to have 1 Kg of meat in it. The weights of the patties vary a little because they are mass produced, and the weight X of a single patty is actually a random variable with mean $\mu = 0.128$ kg and standard deviation $\sigma = 0.005$ kg. Find the probability a box has at least 1 kg of meat, assuming that the weights of the 8 patties in any given box are independent.

Solution: Let $X_1, \ldots, X_8$ be the weights of the 8 patties in a box, and $Y = X_1 + \cdots + X_8$ be their total weight. By the Central Limit Theorem, Y is approximately $N(8\mu, 8\sigma^2)$; we'll assume this approximation is reasonable even though $n = 8$ is small. (This is likely ok because X's distribution is likely fairly close to normal itself.) Thus $Y \sim N(1.024, .0002)$ and

$$\begin{aligned} P(Y > 1) &= P\left(Z > \tfrac{1-1.024}{\sqrt{.0002}}\right) \\ &= P(Z > -1.702) \\ &\doteq .9554 \end{aligned}$$

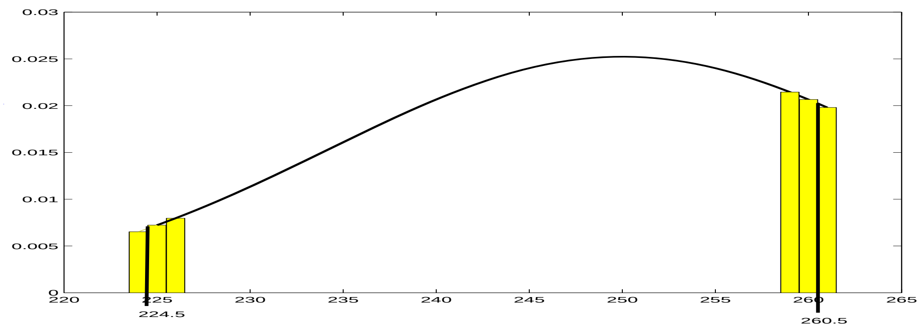
(We see that only about 95% of the boxes actually have 1 kg or more of hamburger. What would you recommend be done to increase this probability to 95%?)

Example: Suppose fires reported to a fire station satisfy the conditions for a Poisson process, with a mean of 1 fire every 4 hours. Find the probability the 500^{th} fire of the year is reported on the 84^{th} day of the year.

Solution: Let X_i be the time between the $(i-1)^{\text{st}}$ and i^{th} fires (X_1 is the time to the 1^{st} fire). Then X_i has an exponential distribution with $\theta = 1/\lambda = 4$ hrs, or $\theta = 1/6$ day. Since $\sum\limits_{i=1}^{500} X_i$ is the time until the 500th fire, we want to find $P\left(83 < \sum\limits_{i=1}^{500} X_i \leq 84\right)$. While the exponential distribution is not close to normal shaped, we are summing a large number of independent exponential variables. Hence, by the central limit theorem, $\sum X_i$ has approximately a $N\left(500\mu, 500\sigma^2\right)$ distribution, where $\mu = E(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i)$.
For exponential distributions, $\mu = \theta = 1/6$ and $\sigma^2 = \theta^2 = 1/36$ so

$$P\left(83 < \sum X_i \leq 84\right) = P\left(\frac{83 - \frac{500}{6}}{\sqrt{\frac{500}{36}}} < Z \leq \frac{84 - \frac{500}{6}}{\sqrt{\frac{500}{36}}}\right)$$
$$= P\left(-.09 < Z \leq .18\right) = .1073$$

Example: This example is frivolous but shows how the normal distribution can approximate even sums of discrete r.v.'s. In an orchard, suppose the number X of worms in an apple has probability function:



x	0	1	2	3
$f(x)$	.4	.3	.2	.1

Find the probability a basket with 250 apples in it has between 225 and 260 (inclusive) worms in it.

Solution:

$$\begin{aligned}
\mu &= E(X) = \sum_{x=0}^{3} x f(x) = 1 \\
E\left(X^2\right) &= \sum_{x=0}^{3} x^2 f(x) = 2 \\
\text{Therefore } \sigma^2 &= E\left(X^2\right) - \mu^2 = 1
\end{aligned}$$

By the central limit theorem, $\sum\limits_{i=1}^{250} X_i$ has approximately a $N\left(250\mu, 250\sigma^2\right)$ distribution, where X_i is the number of worms in the i^{th} apple.
i.e.

$$\sum X_i \sim N(250, 250)$$

$$P\left(225 \le \sum X_i \le 260\right) = P\left(\frac{225-250}{\sqrt{250}} \le Z \le \frac{260-250}{\sqrt{250}}\right)$$
$$= P(-1.58 \le Z \le .63) = .6786$$

While this approximation is adequate, we can improve its accuracy, as follows. When X_i has a discrete distribution, as it does here, $\sum X_i$ will always remain discrete no matter how large n gets. So the distribution of $\sum X_i$, while normal shaped, will never be precisely normal. Consider a probability histogram of the distribution of $\sum X_i$, as shown in Figure 9.6. (Only part of the histogram is shown.) The area of each bar of this histogram is the probability at the x value in the centre of the interval. The smooth curve is the p.d.f. for the approximating normal distribution. Then $\sum\limits_{x=225}^{260} P\left(\sum X_i = x\right)$ is the total area of all bars of the histogram for x from 225 to 260. These bars actually span continuous x

values from 224.5 to 260.5. We could then get a more accurate approximation by finding the area under the normal curve from 224.5 to 260.5.

i.e. $P\left(225 \le \sum X_i \le 260\right) = P\left(224.5 < \sum X_i < 260.5\right)$

$$\begin{aligned} &= P\left(\tfrac{224.5-250}{\sqrt{250}} < Z < \tfrac{260.5-250}{\sqrt{250}}\right) \\ &= P\left(-1.61 < Z < .66\right) = .6917 \end{aligned}$$

Unless making this adjustment greatly complicates the solution, it is preferable to make this "**continuity correction**".

Notes:

(1) A continuity correction should not be applied when approximating a continuous distribution by the normal distribution. Since it involves going halfway to the next possible value of x, there would be no adjustment to make if x takes real values.

(2) Rather than trying to guess or remember when to add .5 and when to subtract .5, it is often helpful to sketch a histogram and shade the bars we wish to include. It should then be obvious which value to use.

Example: Normal approximation to the Poisson Distribution

Let X be a random variable with a Poisson(λ) distribution and suppose λ is large. For the moment suppose that λ is an integer and recall that if we add λ independent Poisson random variables, each with parameter 1, then the sum has the Poisson distribution with parameter λ. In general, a Poisson random variable with large expected value can be written as the sum of a large number of independent random variables, and so the central limit theorem implies that it must be close to normally distributed. We can prove this using moment generating functions. In Section 7.5 we found the moment generating function of a Poisson random variable X

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

Then the standardized random variable is

$$Z = \frac{X-\lambda}{\sqrt{\lambda}}$$

and this has moment generating function

$$\begin{aligned} M_Z(t) &= E(e^{Zt}) = E(e^{t\left(\frac{X-\lambda}{\sqrt{\lambda}}\right)}) \\ &= e^{-t\sqrt{\lambda}}E(e^{Xt/\sqrt{\lambda}}) \\ &= e^{-t\sqrt{\lambda}}M_X(t/\sqrt{\lambda}) \end{aligned}$$

This is easier to work with if we take logarithms,

$$\begin{aligned}\ln(M_Z(t)) &= -t\sqrt{\lambda} - \lambda + \lambda e^{t/\sqrt{\lambda}} \\ &= \lambda(e^{t/\sqrt{\lambda}} - 1 - \frac{t}{\sqrt{\lambda}}).\end{aligned}$$

Now as $\lambda \to \infty$,

$$\frac{t}{\sqrt{\lambda}} \to 0$$

and

$$e^{t/\sqrt{\lambda}} = 1 + \frac{t}{\sqrt{\lambda}} + \frac{1}{2}\frac{t^2}{\lambda} + o(\lambda^{-1})$$

so

$$\begin{aligned}\ln(M_Z(t)) &= \lambda(e^{t/\sqrt{\lambda}} - 1 - \frac{t}{\sqrt{\lambda}}). \\ &= \lambda(\frac{t^2}{2\lambda} + o(\lambda^{-1})) \\ &\to \frac{t^2}{2} \text{ as } \lambda \to \infty.\end{aligned}$$

Therefore the moment generating function of the standardized Poisson random variable Z approaches $e^{t^2/2}$, the moment generating function of the standard normal and this implies that the Poisson distribution approaches the normal as $\lambda \to \infty$.

Normal approximation to the Binomial Distribution

It is well-known that the binomial distribution, at least for large values of n, resembles a bell-shaped or normal curve. The most common demonstration of this is with a mechanical device common in science museums called a "Galton board" or "Quincunx"[12] which drop balls through a mesh of equally spaced pins (see Figure 9.11 and the applet at `http://javaboutique.internet.com/BallDrop/`). Notice that if balls either go to the right or left at each of the 8 levels of pins, independently of the movement of the other balls, then $X =$number of moves to right has a $Bin(8, \frac{1}{2})$ distribution. If the balls are dropped from location 0 (on the $x-$axis) then the ball eventually rests at location $2X - 8$ which is approximately normally distributed since X is approximately normal.

The following result is easily proved using the Central Limit Theorem.

Theorem 37 *Let X have a binomial distribution, $Bi(n, p)$. Then for n large, the r.v.*

$$W = \frac{X - np}{\sqrt{np(1-p)}} \qquad \textit{is approximately } N(0,1)$$

[12]The word comes from Latin quinque (five) unicia (twelve) and means five twelfths.

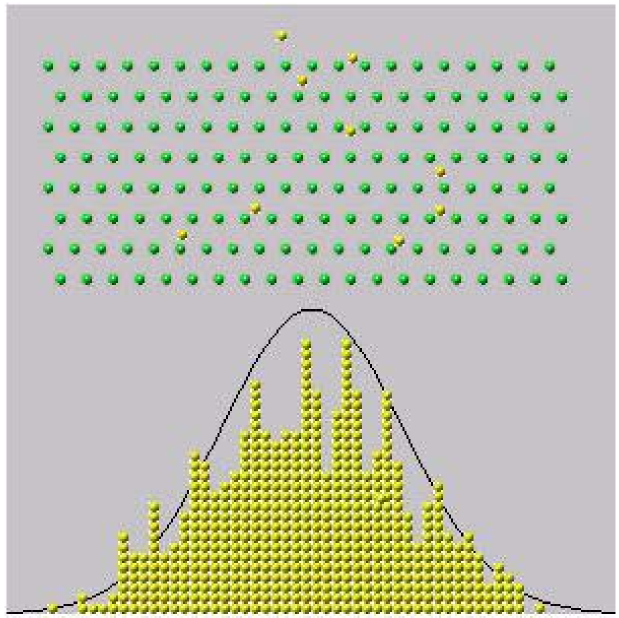


Figure 9.11: A "Galton Board" or "Quincunx"

Proof: We use indicator variables $X_i (i = 1, \ldots, n)$ where $X_i = 1$ if the ith trial in the binomial process is an "S" outcome and 0 if it is an "F" outcome. Then $X = \sum_{i=1}^{n} X_i$ and we can use the CLT. Since

$$\mu = E(X_i) = p, \text{ and } \sigma^2 = \text{Var}(X_i) = p(1-p)$$

we have that as $n \to \infty$

$$\frac{\sum X_i - np}{\sqrt{np(1-p)}} = \frac{X - np}{\sqrt{np(1-p)}}$$

is $N(0, 1)$, as stated. $\square$

An alternative proof uses moment generating functions and is essentially a proof of this particular case of the Central Limit Theorem. Recall that the moment generating function of the binomial random variable X is

$$M_X(t) = (1 - p + pe^t)^n.$$

As we did with the standardized Poisson random variable, we can show with some algebraic effort that the moment generating function of W

$$E(e^{Wt}) \to e^{-t^2/2} \text{ as } n \to \infty$$

proving that the standardized binomial random variable W approaches the standard normal distribution.

Remark: We can write the normal approximation either as $W \sim N(0,1)$ or as $X \sim N(np, np(1-p))$.

Remark: The continuity correction method can be used here. The following numerical example illustrates the procedure.

Example: If (i) $X \sim Bi(n=20, p=.4)$, use the theorem to find the approximate probability $P(4 \le X \le 12)$ and (ii) if $X \sim Bi(100, .4)$ find the approximate probability $P(34 \le X \le 48)$. Compare the answer with the exact value in each case.

Solution (i) By the theorem above, $X \sim N(8, 4.8)$ approximately. Without the continuity correction,

$$\begin{aligned} P(4 \le X \le 12) &= P\left(\tfrac{4-8}{\sqrt{4.8}} \le \tfrac{X-9}{\sqrt{4.8}} \le \tfrac{12-8}{\sqrt{4.8}}\right) \\ &\doteq P(-1.826 \le Z \le 1.826) = 0.932 \end{aligned}$$

where $Z \sim N(0,1)$. Using the continuity correction method, we get

$$\begin{aligned} P(4 \le X \le 12) &\doteq P\left(\tfrac{3.5-8}{\sqrt{4.8}} \le Z \le \tfrac{12.5-8}{\sqrt{4.8}}\right) \\ &= P(-2.054 \le Z \le 2.054) = 0.960 \end{aligned}$$

The exact probability is $\sum\limits_{x=4}^{12} \binom{20}{x} (.4)^x (.6)^{20-x}$, which (using the R function $pbinom(\;\;)$) is .963. As expected the continuity correction method gives a more accurate approximation.
(ii) $X \sim N(40, 24)$ approximately so without the continuity correction

$$\begin{aligned} P(34 \le X \le 48) &\doteq P\left(\tfrac{34-40}{\sqrt{24}} \le Z \le \tfrac{48-40}{\sqrt{24}}\right) \\ &= P(-1.225 \le Z \le 1.033) = .838 \end{aligned}$$

With the continuity correction

$$\begin{aligned} P(34 \le X \le 48) &\doteq P\left(\tfrac{33.5-40}{\sqrt{24}} \le Z \le \tfrac{48-40}{\sqrt{24}}\right) \\ &= P(-1.327 \le Z \le 1.735) = .866 \end{aligned}$$

The exact value, $\sum\limits_{x=34}^{48} f(x)$, equals .866 (to 3 decimals). The error of the normal approximation decreases as n increases, but it is a good idea to use the CC when it is convenient.

Example: Let p be the proportion of Canadians who think Canada should adopt the US dollar.

a) Suppose 400 Canadians are randomly chosen and asked their opinion. Let X be the number who say yes. Find the probability that the proportion, $\frac{X}{400}$, of people who say yes is within .02 of p, if p is .20.

b) Find the number, n, who must be surveyed so there is a 95% chance that $\frac{X}{n}$ lies within .02 of p. Again suppose p is .20.

c) Repeat (b) when the value of p is unknown.

Solution:

a) $X \sim Bi\,(400, .2)$. Using the normal approximation we take

$X \sim$ Normal with mean $np = (400)(.2) = 80$ and variance $np(1-p) = (400)(.2)(.8) = 64$

If $\frac{X}{400}$ lies within $p \pm .02$, then $.18 \le \frac{X}{400} \le .22$, so $72 \le X \le 88$. Thus, we find

$$\begin{aligned} P\,(72 \le X \le 88) &\doteq P\left(\frac{71.5-80}{\sqrt{64}} < Z < \frac{88.5-80}{\sqrt{64}}\right) \\ &= P\,(-1.06 < Z < 1.06) = .711 \end{aligned}$$

b) Since n is unknown, it is difficult to apply a continuity correction, so we omit it in this part. By the normal approximation,

$$X \sim N\,(np = .2n, np(1-p) = .16n)$$

Therefore,

$$\frac{X}{n} \sim N\left(\frac{0.2n}{n} = 0.2, \frac{0.16n}{n^2} = \frac{0.16}{n}\right) \quad (\text{Recall Var }(aX) = a^2 \text{ Var }(X))$$

$P\left(.18 \le \frac{X}{n} \le .22\right) = .95$ is the condition we need to satisfy. This gives

$$\begin{aligned} P\left(\frac{.18-.2}{\sqrt{\frac{.16}{n}}} \le Z \le \frac{.22-.2}{\sqrt{\frac{.16}{n}}}\right) &= .95 \\ P\left(-.05\sqrt{n} \le Z \le .05\sqrt{n}\right) &= 0.95 \end{aligned}$$

Therefore, $F\,(.05\sqrt{n}) = .975$ and so $.05\sqrt{n} = 1.9600$ giving $n = 1536.64$. In other words, we need to survey 1537 people to be at least 95% sure that $\frac{X}{n}$ lies within .02 either side of p.

c) Now using the normal approximation to the binomial, approximately $X \sim N\,(np, np(1-p))$ and so

$$\frac{X}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

We wish to find n such that

$$\begin{aligned}
0.95 &= P\left(p - .02 \le \frac{X}{n} \le p + .02\right) \\
&= P\left(\frac{p - .02 - p}{\sqrt{\frac{p(1-p)}{n}}} \le Z \le \frac{p + .02 - p}{\sqrt{\frac{p(1-p)}{n}}}\right) \\
&= P\left(\frac{-.02}{\sqrt{\frac{p(1-p)}{n}}} \le Z \le \frac{.02}{\sqrt{\frac{p(1-p)}{n}}}\right)
\end{aligned}$$

As is part (b),

$$\begin{aligned}
F\left(\frac{.02}{\sqrt{\frac{p(1-p)}{n}}}\right) &= .975, \\
\frac{.02\sqrt{n}}{\sqrt{p(1-p)}} &= 1.96
\end{aligned}$$

Solving for n,

$$n = \left(\frac{1.96}{.02}\right)^2 p(1-p)$$

Unfortunately this does not give us an explicit expression for n because we don't know p. The way out of this dilemma is to find the maximum value $\left(\frac{1.96}{.02}\right)^2 p(1-p)$ could take. If we choose n this large, then we can be sure of having the required precision in our estimate, $\frac{X}{n}$, for any p. It's easy to see that $p(1-p)$ is a maximum when $p = \frac{1}{2}$. Therefore we take

$$n = \left(\frac{1.96}{.02}\right)^2 \left(\frac{1}{2}\right)\left(1 - \frac{1}{2}\right) = 2401$$

i.e., if we survey 2401 people we can be 95% sure that $\frac{X}{n}$ lies within .02 of p, regardless of the value of p.

Remark: This method is used when poll results are reported in the media: you often see or hear that "this poll is accurate to with 3 percent 19 times out of 20". This is saying that n was big enough so that $P(p - .03 \le X/n \le p + .03)$ was 95%. (This requires n of about 1067.)

Problems:

9.6.1 Tomato seeds germinate (sprout to produce a plant) independently of each other, with probability 0.8 of each seed germinating. Give an expression for the probability that at least 75 seeds out of 100 which are planted in soil germinate. Evaluate this using a suitable approximation.

9.6.2 A metal parts manufacturer inspects each part produced. 60% are acceptable as produced, 30% have to be repaired, and 10% are beyond repair and must be scrapped. It costs the manufacturer \$10 to repair a part, and \$100 (in lost labour and materials) to scrap a part. Find the approximate probability that the total cost associated with inspecting 80 parts will exceed \$1200.

9.7 Problems on Chapter 9

9.1 The diameters X of spherical particles produced by a machine are randomly distributed according to a uniform distribution on [.6,1.0] (cm). Find the distribution of Y, the volume of a particle.

9.2 A continuous random variable X has p.d.f.

$$f(x) = k(1 - x^2) \qquad -1 \le x \le 1.$$

(a) Find k and the c.d.f. of X. Graph $f(x)$ and the c.d.f.

(b) Find the value of c such that $P(-c \le X \le c) = .95$.

9.3 a) When people are asked to make up a random number between 0 and 1, it has been found that the distribution of the numbers, X, has p.d.f. close to

$$f(x) = \begin{cases} 4x; & 0 < x \le 1/2 \\ 4\,(1-x)\,; & \frac{1}{2} < x < 1 \end{cases}$$

(rather than the $U[0,1]$ distribution which would be expected). Find the mean and variance of X.

b) For 100 "random" numbers from the above distribution find the probability their sum lies between 49.0 and 50.5.

c) What would the answer to (b) be if the 100 numbers were truly $U[0,1]$?

9.4 Let X have p.d.f. $f(x) = \frac{1}{20}; \quad -10 < x < 10$, and let $Y = \frac{X+10}{20}$. Find the p.d.f. of Y.

9.5 A continuous random variable X which takes values between 0 and 1 has probability density function

$$f(x) = (\alpha + 1)\,x^{\alpha}; \quad 0 < x < 1$$

a) For what values of α is this a p.d.f.? Explain.

b) Find $P\left(X \le \frac{1}{2}\right)$ and $E(X)$

c) Find the probability density function of $T = 1/X$.

9.6 The magnitudes of earthquakes in a region of North America can be modelled by an exponential distribution with mean 2.5 (measured on the Richter scale).

(a) If 3 earthquakes occur in a given month, what is the probability that none exceed 5 on the Richter scale?

(b) If an earthquake exceeds 4, what is the probability it also exceeds 5?

9.7 A certain type of light bulb has lifetimes that follow an exponential distribution with mean 1000 hours. Find the median lifetime (that is, the lifetime x such that 50% of the light bulbs fail before x).

9.8 The examination scores obtained by a large group of students can be modelled by a normal distribution with a mean of 65% and a standard deviation of 10%.

(a) Find the percentage of students who obtain each of the following letter grades:

$$A(\geq 80\%),\ B(70-80\%),\ C(60-70\%),\ D(50-60\%),\ F(<50\%)$$

(b) Find the probability that the average score in a random group of 25 students exceeds 70%.

(c) Find the probability that the average scores of two distinct random groups of 25 students differ by more than 5%.

9.9 The number of litres X that a filling machine in a water bottling plant deposits in a nominal two litre bottle follows a normal distribution $N(\mu, \sigma^2)$, where $\sigma = .01$ (litres) and μ is the setting on the machine.

(a) If $\mu = 2.00$, what is the probability a bottle has less than 2 litres of water in it?

(b) What should μ be set at to make the probability a bottle has less than 2 litres be less than .01?

9.10 A turbine shaft is made up of 4 different sections. The lengths of those sections are independent and have normal distributions with μ and σ: (8.10, .22), (7.25, .20), (9.75, .24), and (3.10, .20). What is the probability an assembled shaft meets the specifications $28 \pm .26$?

9.11 Let $X \sim G(9.5, 2)$ and $Y \sim N(-2.1, 0.75)$ be independent.

Find:

(a) $P(9.0 < X < 11.1)$

(b) $P(X + 4Y > 0)$

(c) a number b such that $P(X > b) = .90$.

9.12 The amount, A, of wine in a bottle $\sim N\left(1.05l, .0004l^2\right)$ (Note: l means liters.)

a) The bottle is labelled as containing $1l$. What is the probability a bottle contains less than $1l$?

b) Casks are available which have a volume, V, which is $N(22l, .16l^2)$. What is the probability the contents of 20 randomly chosen bottles will fit inside a randomly chosen cask?

9.13 In problem 8.18, calculate the probability of passing the exam, both with and without guessing if (a) each p_i = .45; (b) each $p_i = .55$.
What is the best strategy for passing the course if (a) $p_i = .45$ (b) $p_i = .55$?

9.14 Suppose that the diameters in millimeters of the eggs laid by a large flock of hens can be modelled by a normal distribution with a mean of 40 mm. and a variance of 4 mm^2. The wholesale selling price is 5 cents for an egg less than 37 mm in diameter, 6 cents for eggs between 37 and 42 mm, and 7 cents for eggs over 42 mm. What is the average wholesale price per egg?

9.15 In a survey of n voters from a given riding in Canada, the proportion $\frac{x}{n}$ who say they would vote Conservative is used to estimate p, the probability a voter would vote P.C. (x is the number of Conservative supporters in the survey.) If Conservative support is actually 16%, how large should n be so that with probability .95, the estimate will be in error at most .03?

9.16 When blood samples are tested for the presence of a disease, samples from 20 people are pooled and analysed together. If the analysis is negative, none of the 20 people is infected. If the pooled sample is positive, at least one of the 20 people is infected so they must each be tested separately; i.e., a total of 21 tests is required. The probability a person has the disease is .02.

a) Find the mean and variance of the number of tests required for each group of 20.

b) For 2000 people, tested in groups of 20, find the mean and variance of the total number of tests. What assumption(s) has been made about the pooled samples?

c) Find the approximate probability that more than 800 tests are required for the 2000 people.

9.17 Suppose 80% of people who buy a new car say they are satisfied with the car when surveyed one year after purchase. Let X be the number of people in a group of 60 randomly chosen new car buyers who report satisfaction with their car. Let Y be the number of satisfied owners in a second (independent) survey of 62 randomly chosen new car buyers. Using a suitable approximation, find $P(|X - Y| \geq 3)$. A continuity correction is expected.

9.18 Suppose that the unemployment rate in Canada is 7%.

(a) Find the approximate probability that in a random sample of 10,000 persons in the labour force, the number of unemployed will be between 675 and 725 inclusive.

(b) How large a random sample would it be necessary to choose so that, with probability $.95$, the proportion of unemployed persons in the sample is between 6.9% and 7.1%?

9.19 **Gambling.** Your chances of winning or losing money can be calculated in many games of chance as described here.

Suppose each time you play a game (or place a bet) of \$1 that the probability you win (thus ending up with a profit of \$1) is .49 and the probability you lose (meaning your "profit" is -\$1) is .51

(a) Let X represent your profit after n independent plays or bets. Give a normal approximation for the distribution of X.

(b) If $n = 20$, determine $P(X \geq 0)$. (This is the probability you are "ahead" after 20 plays.) Also find $P(X \geq 0)$ if $n = 50$ and $n = 100$. What do you conclude?

Note: For many casino games (roulette, blackjack) there are bets for which your probability of winning is only a little less than .5. However, as you play more and more times, the probability you lose (end up "behind") approaches 1.

(c) Suppose now you are the casino. If all players combined place $n = 100,000$ \$1 bets in an evening, let X be your profit. Find the value c with the property that $P(X > c) = .99$. Explain in words what this means.

9.20 **Gambling: Crown and Anchor.** Crown and Anchor is a game that is sometimes played at charity casinos or just for fun. It can be played with a "wheel of fortune" or with 3 dice, in which each die has its 6 sides labelled with a crown, an anchor, and the four card suits club, diamond, heart and spade, respectively. You bet an amount (let's say \$1) on one of the 6 symbols: let's suppose you bet on "heart". The 3 dice are then rolled simultaneously and you win $\$t$ if t hearts turn up ($t = 0, 1, 2, 3$).

(a) Let X represent your profits from playing the game n times. Give a normal approximation for the distribution of X.

(b) Find (approximately) the probability that $X > 0$ if (i) $n = 10$, (ii) $n = 50$.

9.21 **Binary classification**. Many situations require that we "classify" a unit of some type as being one of two types, which for convenience we will term Positive and Negative. For example, a diagnostic test for a disease might be positive or negative; an email message may be spam or not spam; a credit card transaction may be fraudulent or not. The problem is that in many cases we cannot tell for certain whether a unit is Positive or Negative, so when we have to decide which a unit is, we may make errors. The following framework helps us to deal with these problems.

For a randomly selected unit from the population being considered, define the indicator random variable

$$Y = I(\text{unit is Positive})$$

Suppose that we cannot know for certain whether $Y = 0$ or $Y = 1$ for a given unit, but that we can get a measurement X with the property that

$$\text{If } Y = 1, \qquad X \sim N(\mu_1, \sigma_1^2)$$
$$\text{If } Y = 0, \qquad X \sim N(\mu_0, \sigma_0^2)$$

where $\mu_1 > \mu_0$. We now decide to classify units as follows, based on their measurement X: select some value d between μ_0 and μ_1, and then

- if $X \geq d$, classify the unit as Positive
- if $X < d$, classify the unit as Negative

(a) Suppose $\mu_0 = 0,\ \mu_1 = 10,\ \sigma_0 = 4,\ \sigma_1 = 6$ and $d = 5$. Find the probability that

   (i) If a unit is really Positive, they are wrongly classified as Negative. (This is called the “false negative" probability.)

   (ii) If a unit is really Negative, they are wrongly classified as Positive. (This is called the “false positive" probability.)

(b) Repeat the calculations if $\mu_0 = 0,\ \mu_1 = 10$ as in (a), but $\sigma_1 = 3,\ \sigma_2 = 3$. Explain in plain English why the false negative and false positive misclassification probabilities are smaller than in (a).

9.22 **Binary classification and spam detection**. The approach in the preceding question can be used for problems such as spam detection, which was discussed earlier in Problems 4.17 and 4.18. Instead of using binary features as in those problems, suppose that for a given email message we compute a measure X, designed so that X tends to be high for spam messages and low for regular (non-spam) messages. (For example X can be a composite measure based on the presence or absence of certain words in a message, as well as other features.) We will treat X as a continuous random variable.

Suppose that for spam messages, the distribution of X is approximately $N(\mu_1, \sigma_1^2)$, and that for regular messages, it is approximately $N(\mu_0, \sigma_0^2)$, where $\mu_1 > \mu_0$. This is the same setup as for Problem 9.21. We will filter spam by picking a value d, and then filtering any message for which $X \geq d$. The trick here is to decide what value of d to use.

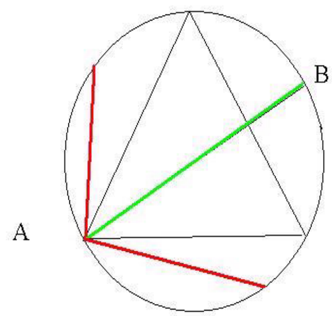


Figure 9.12: Bertrand's Paradox

(a) Suppose that $\mu_0 = 0,\ \mu_1 = 10,\ \sigma_1 = 3,\ \sigma_2 = 3$. Calculate the probability of a false positive (filtering a message that is regular) and a false negative (not filtering a message that is spam) under each of the three choices (i) $d = 5$ (ii) $d = 4$ (iii) $d = 6$.

(b) What factors would determine which of the three choices of d would be best to use?

9.23 **Random chords of a circle**. Given a circle, find the probability that a chord chosen at random be longer than the side of an inscribed equilateral triangle. For example in Figure 9.12, the line joining A and B satisfies the condition, the other lines do not. This is called Bertrand's paradox (see the Java applet at http://www.cut-the-knot.org/bertrand.shtml) and there various possible solutions, depending on exactly how you interpret the phrase "a chord chosen at random". For example, since the only important thing is the position of the second point relative to the first one, we can fix the point A and consider only the chords that emanate from this point. Then it becomes clear that 1/3 of the outcomes (those with angle with the tangent at that point between 60 and 120 degrees) will result in a chord longer than the side of an equilateral triangle. But a chord is fully determined by its midpoint. Chords whose length exceeds the side of an equilateral triangle have their midpoints inside a smaller circle with radius equal to 1/2 that of the given one. If we choose the midpoint of the chord at random and uniformly from the points within the circle, what is the probability that corresponding chord has length greater than the side of the triangle? Can you think of any other interpretations which lead to different answers?

9.24 **A model for stock returns**. A common model for stock returns is as follows: the number of trades N of stock XXX in a given day has a Poisson distribution with parameter λ. At each trade, say the i'th trade, the change in the price of the stock is X_i and has a normal distribution with mean 0 and variance σ^2, say and these changes are independent of one another and independent of N. Find the moment generating function of the total change in stock price over the day. Is this a distribution that you recognise? What is its mean and variance?

9.25 Let $X_1, X_2, ..., X_n$ be independent random variable with a Normal distribution having mean 1, and variance 2. Find the moment generating function for

(a) X_1

(b) $X_1 + X_2$

(c) $S_n = X_1 + X_2 + ... + X_n$

(d) $n^{-1/2}(S_n - n)$

10. Solutions to Section Problems

3.1.1 (a) Each student can choose in 4 ways and they each get to choose. So S has $4 \times 4 \times 4 = 64$ points.

(i) The first student can choose in 4 ways and the others then only have 1 section they can go in.
Therefore $P\,(\text{all in same section}) = \frac{4\times1\times1}{64} = 1/16$.

(ii) The first to pick has 4 ways to choose, the next has 3 sections left, and the last has 2 sections left.
Therefore P (different sections) = $\frac{4\times3\times2}{64} = 3/8$.

(iii) Each has 3 ways to choose a section.
Therefore P (nobody in section 1) = $\frac{3\times3\times3}{64} = 27/64$

(b) Now S has n^s points

(i) P (all in same section) = $n \times 1 \times 1 \times ... \times 1/n^s = 1/n^{s-1}$

(ii) P (different sections) = $n(n-1)(n-2)...(n-s+1)/n^s = \frac{n^{(s)}}{n^s}$.

(iii) P (nobody in section 1) = $(n-1)(n-1)(n-1)...(n-1)/n^s = \frac{(n-1)^s}{n^s}$

3.1.2 (a) There are 26 ways to choose each of the 3 letters, so in all the letters can be chosen in $26\times26\times26$ ways. If all letters are the same, there are 26 ways to choose the first letter, and only 1 way to choose the remaining 2 letters. So P (all letters the same) is $\frac{26\times1\times1}{26^3} = 1/26^2$.

(b) There are $10 \times 10 \times 10$ ways to choose the 3 digits. The number of ways to choose all even digits is $4 \times 4 \times 4$. The number of ways to choose all odd digits is $5 \times 5 \times 5$. Therefore P (all even or all odd) = $\frac{4^3+5^3}{10^3} = .189$.

3.1.3 (a) There are 35 symbols in all (26 letters + 9 numbers). The number of different 6-symbol passwords is $35^6 - 26^6$ (we need to subtract off the 26^6 arrangements in which only letters are used, since there must be at least one number). Similarly, we get the number of 7-symbol and 8-symbol passwords as $35^7 - 26^7$ and $35^8 - 26^8$. The total number of possible passwords is then

$$(35^6 - 26^6) + (35^7 - 26^7) + (35^8 - 26^8).$$

(b) Let N be the answer to part (a) (the total no. of possible passwords). Assuming you never try the same password twice, the probability you find the correct password within the first 1,000 tries is $1000/N$.

3.2.1 There are $7!$ different orders

(a) We can stick the even digits together in $3!$ orders. This block of even digits plus the 4 odd digits can be arranged in $5!$ orders.
Therefore P (even together) = $\frac{3!5!}{7!} = 1/7$

(b) For even at ends, there are 3 ways to fill the first place, and 2 ways to fill the last place and $5!$ ways to arrange the middle 5 digits. For odd at ends there are 4 ways to fill the first place and 3 ways to fill the last place and $5!$ ways to arrange the middle 5 digits. P (even or odd ends) = $\frac{(3)(2)(5!)+(4)(3)(5!)}{7!} = \frac{3}{7}$.

3.2.2 The number of arrangements in S is $\frac{9!}{3!2!}$

(a) E at each end gives $\frac{7!}{2!}$ arrangements of the middle 7 letters.
L at each end gives $\frac{7!}{3!}$ arrangements of the middle 7 letters.
Therefore P (same letter at ends) = $\frac{\frac{7!}{2!}+\frac{7!}{3!}}{\frac{9!}{3!2!}} = \frac{1}{9}$

(b) The X, C and N can be "stuck" together in $3!$ ways to form a single unit. We can then arrange the $3E$'s, $2L$'s, T, and (XCN) in $\frac{7!}{3!2!}$ ways.
Therefore $P(XCN$ together) =$\frac{\frac{7!}{3!2!}\times 3!}{\frac{9!}{3!2!}} = \frac{1}{12}$.

(c) There is only 1 way to arrange the letters in the order CEEELLNTX.
Therefore P (alphabetical order) = $\frac{1}{\frac{9!}{3!2!}} = \frac{12}{9!}$

3.3.1 (a) The 8 cars can be chosen in $\binom{160}{8}$ ways. We can choose x with faulty emission controls and $(8-x)$ with good ones in $\binom{35}{x}\binom{125}{8-x}$ ways.
Therefore P (at least 3 faulty) = $\frac{\sum_{x=3}^{8}\binom{35}{x}\binom{125}{8-x}}{\binom{160}{8}}$ since we need x = 3 or 4 or .... or 8.

(b) This assumes all $\binom{160}{8}$ combinations are equally likely. This assumption probably doesn't hold since the inspector would tend to select older cars or those in bad shape.

3.3.2 (a) The first 6 finishes can be chosen in $\binom{15}{6}$ ways. Choose 4 from numbers 1,2, ..., 9 in $\binom{9}{4}$ ways and 2 from numbers 10, ..., 15 in $\binom{6}{2}$ ways.
Therefore P (4 single digits in top 6) = $\frac{\binom{9}{4}\binom{6}{2}}{\binom{15}{6}} = \frac{54}{143}$.

(b) Need 2 single digits and 2 double digit numbers in 1^{st} 4 and then a single digit. This occurs in $\binom{9}{2}\binom{6}{2} \times 7$ ways.

Therefore P (5^{th} is 3^{rd} single digit) = $\frac{\binom{9}{2}\binom{6}{2}\times 7}{\binom{15}{4}\times 11} = \frac{36}{143}$. (since we can choose 1^{st} 4 in $\binom{15}{4}$ ways and then have 11 choices for the 5^{th})

Alternate Solution: There are $15^{(5)}$ ways to choose the first 5 in order. We can choose in order, 2 double digit and 3 single digit finishers in $6^{(2)}9^{(3)}$ ways, and then choose which 2 of the first 4 places have double digit numbers in $\binom{4}{2}$ ways.

Therefore P (5^{th} is 3^{rd} single digit) =$\frac{6^{(2)}9^{(3)}\binom{4}{2}}{15^{(5)}} = \frac{36}{143}$.

(c) Choose 13 in 1 way and the other 6 numbers in $\binom{12}{6}$ ways. (from 1,2, ..., 12).

Therefore P (13 is highest) = $\frac{\binom{12}{6}}{\binom{15}{7}} = \frac{28}{195}$.

Alternate Solution: From the $\binom{13}{7}$ ways to choose 7 numbers from 1,2, ..., 13 subtract the $\binom{12}{7}$ which don't include 13 (i.e. all 7 chosen from 1,2, ..., 12).

Therefore P (13 is highest) = $\frac{\binom{13}{7}-\binom{12}{7}}{\binom{15}{7}} = \frac{28}{195}$

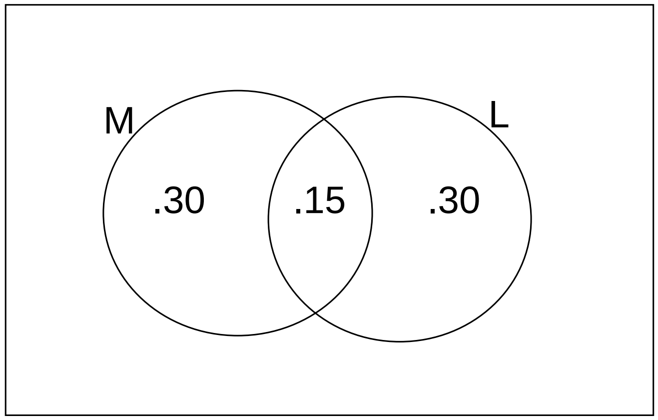
4.1.1 Let $R = \{\text{rain}\}$ and $T = \{\text{temp.} > 22°C\}$, and draw yourself a Venn diagram. Then

$P(T\overline{R}) = .4 \times .2 = .08$
$P(\overline{T}\ \overline{R}) = .7 \times .8 = .56$

(Note that the information that 40% of days with temp $>\ 22°$ have no rain is not needed to solve the question).

Therefore $P(R\overline{T}) = 24\%$
This result is to be expected since 80% of days have a high temperature $\leq\ 22°C$ and 30% of these days have rain.

4.1.2 $P(ML) = .15$
$P(M) = .45$
$P(L) = .45$ See the Figure below:



The region outside the circles represents females to the right. To make $P(S) = 1$. we need $P(FR) = .25$.

4.2.1 (a) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$
$= 1 - .1 - [P(AC) + P(BC) - P(ABC)]$
Therefore $P(A \cup B \cup C) = .9$ is the largest value, and this occurs when $P(AC) + P(BC) - P(ABC) = 0$.

(b) Since ABC is contained within both AC and BC we know $P(ABC) \leq P(AC)$ and $P(ABC) \leq P(BC)$. Thus $P(AC) + P(BC) - P(ABC) = 0$ iff $P(AC) = P(BC) = P(ABC) = 0$.
While A and C could be mutually exclusive, it can't be determined for sure.

4.2.2

$$\begin{aligned} P(A \cup B) = P(A \text{ or } B \text{ occur}) &= 1 - P(A \text{ doesn't occur AND } B \text{ doesn't occur}) \\ &= 1 - P(\bar{A}\bar{B}). \end{aligned}$$

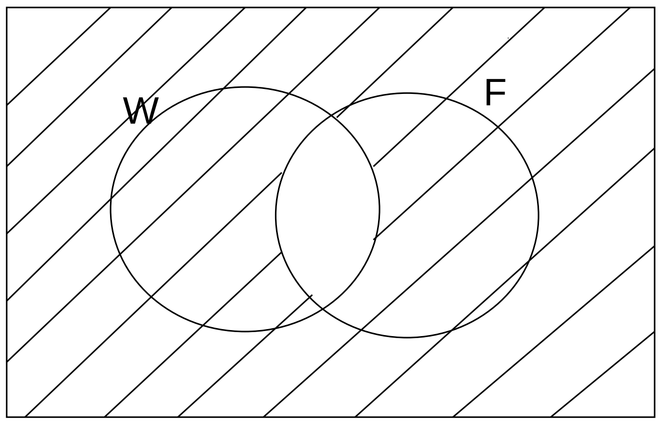
Alternatively, (look at a Venn diagram), $S = (A \cup B) \cup (\bar{A}\bar{B})$ is a partition, so $P(S) = 1 \Rightarrow P(A \cup B) + P(\bar{A}\bar{B}) = 1$.

4.3.1 (a) Points giving a total of 9 are: (3,6), (4,5), (5,4) and (6,3). The probabilities are (.1)(.3) = .03 for (3,6) and for (6,3), and (.2)(.2) = .04 for (4,5) and for (5,4).
Therefore P (3, 6) or (4, 5) or (5, 4) or (6, 3)) = .03 + .04 + .04 + .03 = .14.

(b) There are $\binom{4}{1}$ arrangements with 1 nine and 3 non-nines. Each arrangement has probability $(.14)(.86)^3$.
Therefore P (nine on 1 of 4 repetitions) = $\binom{4}{1}(.14)(.86)^3 = .3562$

4.3.2 Let W = {at least 1 woman} and F = {at least 1 French speaking student}.

$$P(WF) = 1 - P(\overline{WF}) = 1 - P(\overline{W} \cup \overline{F}) = 1 - \left[P(\overline{W}) + P(\overline{F}) - P(\bar{W}\bar{F})\right] \text{ (see figure)}$$



Venn diagram, Problem 4.3.2

But $P(\overline{WF}) = P$ (no woman and no French speaking student)$= P$ (all men who don't speak French)
P (woman who speaks French) =P (woman)P(French|woman)= $.45 \times .20 = .09$.
From Venn diagram, P (man without French) = .49.

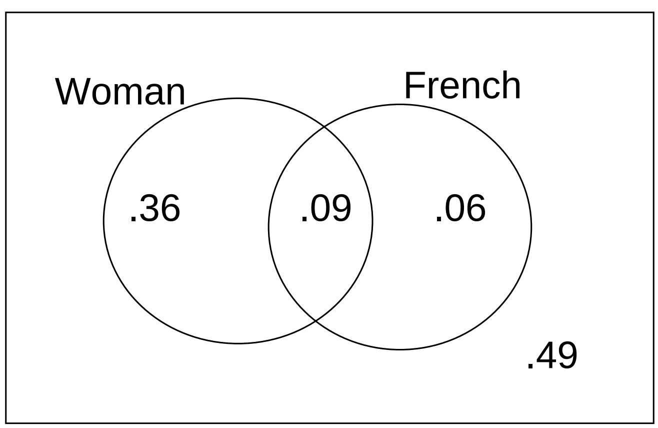


Figure 10.1: P(woman who speaks french)

$$P(\overline{W}\,\overline{F}) = (.49)^{10} \text{ and } P(\overline{W}) = (.55)^{10}; P(\overline{F}) = (.85)^{10}$$

Therefore $P(WF) = 1 - \left[(.55)^{10} + (.85)^{10} - (.49)^{10}\right] = 0.8014.$

4.3.3 From a Venn diagram: (1) $P(\overline{A}B) = P(B) - P(AB)$ (2) $P(\overline{A}\,\overline{B}) = P(\overline{A \cup B})$

$P(\overline{A}\,\overline{B}) = P(\overline{A})P(\overline{B})$
$\Leftrightarrow P(\overline{A \cup B}) = P(\overline{A})P(\overline{B})$
$\Leftrightarrow 1 - P(A \cup B) = P(\overline{A})P(\overline{B})$
$\Leftrightarrow 1 - [P(A) + P(B) - P(AB)] = P(\overline{A})\,[1 - P(B)]$
$\Leftrightarrow [1 - P(A)] - [P(B) - P(AB)] = P(\overline{A}) - P(\overline{A})P(B)$
$\Leftrightarrow P(\overline{A}) - P(\overline{A}B) = P(\overline{A}) - P(\overline{A})P(B)$
$\Leftrightarrow P(\overline{A})P(B) = P(\overline{A}B)$

Therefore $\overline{A}$ and $\overline{B}$ are independent iff $\overline{A}$ and B are independent.

4.4.1 Let B = $\{\text{bus}\}$ and L = $\{\text{late}\}$.

$P(B|L) = \frac{P(BL)}{P(L)} = \frac{P(L|B)P(B)}{P(L|B)P(B)+P(L|\bar{B})P(\bar{B})}$
$= \frac{(.3)(.2)}{(.3)(.2)+(.7)(.1)} = 6/13$

4.4.2 Let F = $\{\text{fair}\}$ and H = $\{5\text{ heads}\}$

$$P(F|H) = \frac{P(FH)}{P(H)} = \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|\bar{F})P(\bar{F})}$$
$$= \frac{(\frac{3}{4})\binom{6}{5}(\frac{1}{2})^6}{(\frac{3}{4})\binom{6}{5}(\frac{1}{2})^6 + (\frac{1}{4})\binom{6}{5}(.8)^5(.2)^1} = 0.4170$$

4.4.3 Let H = { defective headlights}, M = {defective muffler}

$$P(M/H) = \frac{P(MH)}{P(H)} = \frac{P(MH)}{P(MH \cup \overline{M}H)} = \frac{.1}{.1 + .15} = .4$$

5.1 $\sum\limits_{x=0}^{n} \binom{n}{x} a^x = (1+a)^n$

Differentiate with respect to a on both sides:

$\sum\limits_{x=0}^{n} x\binom{n}{x} a^{x-1} = n(1+a)^{n-1}$. Multiply by a to get $\sum\limits_{x=0}^{n} x\binom{n}{x} a^x = na(1+a)^{n-1}$

Let $a = \left(\frac{p}{1-p}\right)$. Then $\sum\limits_{x=0}^{n} x\binom{n}{x}\left(\frac{p}{1-p}\right)^x =^n \left(\frac{p}{p-1}\right)\left(1 + \frac{p}{1-p}\right)^{n-1} = \frac{np}{(10p)^n}(1)^{n-1}$

Multiply by $(1-p)^n$:

$$\sum_{x=0}^{n} x\binom{n}{x}\left(\frac{p}{1-p}\right)^x (1-p)^n = \sum_{x=0}^{n} x\binom{n}{x} p^x (1-p)^{n-x} = \frac{np}{(1-p)^n}(1-p)^n = np$$

5.2 Let Q = {heads on quarter} and D ={heads on dime}. Then

$$\begin{aligned} P(\text{Both heads at same time}) &= P(QD \cup \overline{Q}\,\overline{D}QD \cup \overline{Q}\,\overline{D}\,\overline{Q}\,\overline{D}\,QD \cup \cdots) \\ &= (.6)(.5) + (.4)(.5)(.6)(.5) + (.4)(.5)(.4)(.5)(.6)(.5) + \cdots \\ &= \frac{(.6)(.5)}{1-(.4)(.5)} = 3/8 (\text{using } a + ar + ar^2 + \cdots = \frac{a}{1-r} \text{with } r = (.4)(.5)) \end{aligned}$$

6.1.1 We need $f(x) \geq 0$ and $\sum\limits_{x=0}^{2} f(x) = 1$

$$9c^2 + 9c + c^2 = 10c^2 + 9c = 1$$

$$\text{Therefore} \quad 10c^2 + 9c - 1 = 0$$

$$(10c-1)(c+1) = 0$$

$$c = 1/10 \text{ or } -1$$

But if $c = -1$ we have $f(1) < 0$ ... impossible. Therefore $c = .1$

6.1.2 We are arranging $Y\ F\ O\ O\ O$ where Y = {you}, F = {friend}, O = {other}. There are $\frac{5!}{3!} = 20$ distinct arrangements.

$X = 0$: $Y\ F\ O\ O\ O, \cdots, O\ O\ O\ Y F$ has 4 arrangements with Y first and 4 with F first.
$X = 1$: $YOFOO, \cdots, OOYOF$ has 3 arrangements with Y first and 3 with F first.
$X = 2$: $YOOFO, OYOOF$ has 2 with Y first and 2 with F.
$X = 3$: $YOOOF$ has 1 with Y first and 1 with F.

x	0	1	2	3
$f(x)$	.4	.3	.2	.1
$F(x)$	.4	.7	.9	1

6.2.1

$$\begin{aligned} F(x) &= P\,(\text{largest number} \leq x) \\ &= P\,(\text{all numbers} \leq x) \\ &= P\,(\text{1st roll} \leq x \text{ and 2nd roll} \leq x \text{ and 3rd roll} \leq x) \\ &= \left(\tfrac{x}{6}\right)\left(\tfrac{x}{6}\right)\left(\tfrac{x}{6}\right) \quad (\text{since } P\,(\text{any roll} \leq x) = x/6) \\ \text{Therefore } F(x) &= \begin{cases} \frac{x^3}{216} & \text{for } x = 1, 2, ..., 6 \\ 0 & \text{for } x < 1 \\ 1 & \text{for } x > 6 \end{cases} \end{aligned}$$

$$\begin{aligned} f(x) = F(x) - F(x-1) &= \frac{x^3 - (x-1)^3}{216} = \frac{[x-(x-1)][x^2 + x(x-1) + (x-1)^2]}{216} \\ &= \frac{3x^2 - 3x + 1}{216}; \quad x = 1, 2, ..., 6 \end{aligned}$$

6.3.1 (a) Using the hypergeometric distribution,

$$f(0) = \frac{\binom{d}{0}\binom{12-d}{7}}{\binom{12}{7}}$$

d	0	1	2	3
$f(0)$	1	$5/12$	$5/33$	$5/110$

(b) While we could find none tainted if d is as big as 3, it is not likely to happen. This implies the box is not likely to have as many as 3 tainted tins.

6.3.2 Considering order, there are $N^{(n)}$ points in S. We can choose which x of the n selections will have "success" in $\binom{n}{x}$ ways. We can arrange the x "successes" in their selected positions in $r^{(x)}$ ways and the $(n-x)$ "failures" in the remaining positions in $(N-r)^{(n-x)}$ ways.

Therefore $f(x) = \frac{\binom{n}{x} r^{(x)} (N-r)^{(n-x)}}{N^{(n)}}$ with x ranging from max $(0, n-(N-r))$ to min (n, r)

6.4.1 (a) Using hypergeometric, with $N = 130, r = 26, n = 6$,

$$f(2) = \frac{\binom{26}{2}\binom{104}{4}}{\binom{130}{6}} \quad (= .2506)$$

(b) Using binomial as an approximation,

$$f(2) = \binom{6}{2}\left(\frac{26}{130}\right)^2\left(\frac{104}{130}\right)^4 = 0.2458$$

6.4.2 (a) P (fail twice)
= $P(A)P$ (fail twice $|A$) + $P(B)P$ (fail twice $|B$)
= $\left(\frac{1}{2}\right)\binom{10}{2}(.1)^2(.9)^8 + \left(\frac{1}{2}\right)\binom{10}{2}(.05)^2(.95)^8 = .1342.$
Where A = { camera A is picked }
and B = { camera B is picked }

This assumes shots are independent with a constant failure probability.

(b) $P(A|$ failed twice) = $\frac{P(A \text{ and fail twice})}{P(\text{fail twice})} = \frac{\left(\frac{1}{2}\right)\binom{10}{2}(.1)^2(.9)^8}{.1342} = .7219$

6.5.1 We need $(x-25)$ "failures" before our 25th "success".

$$f(x) = \binom{x-1}{x-25}(.2)^{25}(.8)^{x-25} \text{ or } \binom{x-1}{24}(.2)^{25}(.8)^{x-25};\ x = 25, 26, 27, \ldots$$

6.5.2 (a) In the first $(x+17)$ selections we need to get x defective (use hypergeometric distribution) and then we need a good one on the $(x+18)^{\text{th}}$ draw.

$$\text{Therefore } f(x) = \frac{\binom{200}{x}\binom{2300}{17}}{\binom{2500}{x+17}} \times \frac{2283}{2500-(x+17)};\ \ x = 0, 1, \cdots, 200$$

(b) Since 2500 is large and we're only choosing a few of them, we can approximate the hypergeometric portion of $f(x)$ using binomial

$$f(2) \doteq \binom{19}{2}\left(\frac{200}{2500}\right)^2\left(1-\frac{200}{2500}\right)^{17} \times \frac{2283}{2481} = .2440$$

6.6.1 Using geometric,

$$P(x \text{ not leaky found before first leaky}) = (0.7)^x(0.3) = f(x)$$

$$P(X \geq n-1) = \ f(n-1) + f(n) + f(n+1) + ...$$

$$= (0.7)^{n-1}(0.3) + (0.7)^n(0.3) + (0.7)^{n+1}(0.3) + ... = \frac{(.7)^{n-1}(.3)}{1-.7}$$

$$= (.7)^{n-1} = 0.05$$

$$(n-1)\log .7 = \log .05;\ \ \text{so } n \ = \ 9.4$$

At least 9.4 cars means 10 or more cars must be checked
Therefore $n \ = \ 10$

6.7.1 (a) Let X be the number who don't show. Then

$$X \ \sim \ Bi(122, .03)$$

$$P(\text{not enough seats}) \ = \ P(X = 0 \text{ or } 1)$$

$$= \binom{122}{0}(.03)^0(.97)^{122} + \binom{122}{1}(.03)^1(.97)^{121} = 0.1161$$

(To use a Poisson approximation we need p near 0. That is why we defined "success" as not showing up).

For Poisson, $\mu = np = (122)(.03) = 3.66$

$$f(0) + f(1) = e^{-3.66} + 3.66e^{-3.66} = 0.1199$$

(b) Binomial requires all passengers to be independent as to showing up for the flight, and that each passenger has the same probability of showing up. Passengers are not likely independent since people from the same family or company are likely to all show up or all not show. Even strangers arriving on an earlier incoming flight would not miss their flight independently if the flight was delayed. Passengers may all have roughly the same probability of showing up, but even this is suspect. People travelling in different fare categories or in different classes (e.g. charter fares versus first class) may have different probabilities of showing up.

6.8.1 (a)

$$\lambda = 3, \;\; t = 2\frac{1}{2}, \;\; \mu = \lambda t = 7.5$$

$$f(6) = \frac{7.5^6 e^{-7.5}}{6!} = 0.1367$$

(b)

$$P(2 \text{ in 1st minute}|6 \text{ in } 2\frac{1}{2} \text{ minutes})$$

$$= \frac{P(2 \text{ in 1st min. and 6 in } 2\frac{1}{2} \text{ min.})}{P(6 \text{ in } 2\frac{1}{2} \text{ min})}$$

$$= \frac{P(2 \text{ in 1st min. and 4 in last } 1\frac{1}{2} \text{ min})}{P(6 \text{ in } 2\frac{1}{2} \text{ min.})}$$

$$= \frac{\left(\frac{3^2 e^{-3}}{2!}\right)\left(\frac{4.5^4 e^{-4.5}}{4!}\right)}{\left(\frac{7.5^6 e^{-7.5}}{6!}\right)} = \binom{6}{2}\left(\frac{3}{7.5}\right)^2\left(\frac{4.5}{7.5}\right)^4 = .3110$$

Note this is a binomial probability function.

6.8.2 Assuming the conditions for a Poisson process are met, with lines as units of “time":

(a) $\lambda = .02$ per line; $t = 1$ line; $\mu = \lambda t = .02$

$$f(0) = \frac{\mu^0 e^{-\mu}}{0!} = e^{-.02} = .9802$$

(b) $\mu_1 = 80 \times .02 = 1.6$; $\mu_2 = 90 \times .02 = 1.8$

$$\left[\frac{\mu_1^2 e^{-\mu_1}}{2!}\right]\left[\frac{\mu_2^2 e^{-\mu_2}}{2!}\right] = .0692$$

6.9.1 Consider a 1 minute period with no occurrences as a "success". Then X has a geometric distribution. The probability of "success" is

$$f(0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}.$$

$$\text{Therefore } \; f(x) = (e^{-\lambda})(1 - e^{-\lambda})^{x-1}; \; x = 1, 2, 3, ...$$

(There must be $(x-1)$ failures before the first success.)

6.9.2 (a) $\mu = 3 \times 1.25 = 3.75$

$$f(0) = \frac{3.75^0 e^{-3.75}}{0!} = .0235$$

(b) $\left(1 - e^{-3.75}\right)^{14} e^{-3.75}$, using a geometric distribution

(c) Using a binomial distribution

$$f(x) = \binom{100}{x} \left(e^{-3.75}\right)^x \left(1 - e^{-3.75}\right)^{100-x}$$

Approximate by Poisson with $\mu = np = 100e^{-3.75} = 2.35$
$f(x) \doteq e^{-2.35} \frac{2.35^x}{x!}$ (n large, p small)
Thus, $P(X \geq 4) = 1 - P(X \leq 3) = 1 - .789 = .211$.

7.3.1 There are 10 tickets with all digits identical. For these there is only 1 prize. There are 10×9 ways to pick a digit to occur twice and a different digit to occur once. These can be arranged in $\frac{3!}{2!\,1!} = 3$ different orders; i.e. there are 270 tickets for which 3 prizes are paid. There are $10 \times 9 \times 8$ ways to pick 3 different digits in order. For each of these 720 tickets there will be 3! prizes paid.

The organization takes in \$1,000.

$$\text{Therefore } E(\text{Profit}) = \left[(1000 - 200) \times \frac{10}{1000}\right] + \left[(1000 - 600) \times \frac{270}{1000}\right]$$
$$+ \left[(1000 - 1200) \times \frac{720}{1000}\right] = -\$28$$

i.e., on average they lose \$28.

7.4.1 Let them sell n tickets. Suppose X show up. Then $X \sim Bi(n, .97)$. For binomial, $\mu = E(X) = np = .97n$
If $n \leq 120$, revenues will be $100X$, and $E(100X) = 100E(X) = 97n$. This is maximized for $n = 120$. Therefore Max. expected revenue is \$11,640.
For $n = 121$, revenues are $100X$, less \$500 if all 121 show up.

i.e., $100 \times 121 \times .97 - 500\, f(121)$
$= \; 11,737 - 500(.97)^{121}$ = \$11,724.46 is expected.
For $n = 122$, revenues are $100X$, less \$500 if 121 show up, less \$1000 if all 122 show.
i.e., $100 \times 122 \times .97 - 500\binom{122}{121}(.97)^{121}(.03) - 1000(.97)^{122}$
$= \; \$11,763.77$ is expected.
For $n = 123$, revenues are $100X$, less \$500 if 121 show, less \$1,000 if 122 show, less \$1500 if all 123 show.
i.e. $100 \times 123 \times .97 - 500\binom{123}{121}(.97)^{121}(.03)^2$
$-1000\binom{123}{122}(.97)^{122}(.03) - 1500(.97)^{123}$
$= \; \$11,721.13$ is expected.
Therefore They should sell 122 tickets.

7.4.2 (a) Let X be the number of words needing correction and let T be the time to type the passage. Then $X \sim Bi(450, .04)$ and $T = 450 + 15X$.

X has mean $np = 18$ and variance $np(1-p) = 17.28$.
$E(T) = E(450 + 15X) = 450 + 15E(X) = 450 + (15)(18) = 720$
Var(T) = Var$(450 + 15X) = 15^2$ Var$(X) = 3888$.

(b) At 45 words per minute, each word takes $1\frac{1}{3}$ seconds. $X \sim Bi(450, .02)$ and
$T = \left(450 \times 1\frac{1}{3}\right) + 15X = 600 + 15X$
$E(X) = 450 \times .02 = 9$; $\;E(T) = 600 + (15)(9) = 735$, so it takes longer on average.

8.1.1 (a) The marginal probability functions are:

x	0	1	2
$f_1(x)$	.3	.2	.5

and

y	0	1	2
$f_2(y)$	.3	.4	.3

Since $f_1(x)\;f_2(y) \;\neq\; f(x, y)$ for all (x, y)
Therefore X and Y are not independent.
e.g. $f_1(1)\; f_2(1) = 0.08 \neq 0.05$

(b)

$$f(y|X=0) \;=\; \frac{f(0,y)}{f_1(0)} \;=\; \frac{f(0,y)}{.3}$$

y	0	1	2
$f(y\|X=0)$	.3	.5	.2

(c)

$$\begin{array}{cc|ccc} & & & x & \\ & d & 0 & 1 & 2 \\ \hline & 0 & 0 & 1 & 2 \\ y & 1 & -1 & 0 & 1 \\ & 2 & -2 & -1 & 0 \end{array}$$

d	-2	-1	0	1	2
$f(d)$	$.06$	$.24$	$.29$	$.26$	$.15$

(e.g. $P(D=0) = f(0,0) + f(1,1) + f(2,2)$)

8.1.2

$$f(y|x) = \frac{f(x,y)}{f_1(x)}$$

$$\begin{aligned} f(x,y) &= P(y \text{ calls}) P(x \text{ sales} | y \text{ calls}) \\ &= \left(\frac{20^y e^{-20}}{y!}\right)\left(\binom{y}{x}(.2)^x(.8)^{y-x}\right) \\ &= \frac{(20^{y-x})(.8)^{y-x}}{(y-x)!} \cdot \frac{(20)^x(.2)^x}{x!} e^{-20};\ x = 0,1,2,...;\ y = x, x+1, x+2, ... \end{aligned}$$

(y starts at x since no. of calls $\geq$ no. of sales).

$$\begin{aligned} f_1(x) &= \sum_{y=x}^{\infty} f(x,y) = \frac{[(20)(.2)]^x}{x!} e^{-20} \sum_{y=x}^{\infty} \frac{[(20)(.8)]^{y-x}}{(y-x)!} \\ &= \frac{4^x e^{-20}}{x!}\left[\frac{16^0}{0!} + \frac{16^1}{1!} + \frac{16^2}{2!} + ...\right] = \frac{4^x e^{-20}}{x!} \cdot e^{16} \end{aligned}$$

$$= \frac{4^x e^{-4}}{x!}$$

Therefore $f(y|x) = \dfrac{\frac{16^{y-x}}{(y-x)!}\ \frac{4^x}{x!}\ e^{-20}}{\frac{4^x e^{-4}}{x!}} = \dfrac{16^{y-x} e^{-16}}{(y-x)!};\ y = x, x+1, x+2, ...$

8.1.3

$$
\begin{aligned}
f(x,y) = f(x)f(y) &= \binom{x+k-1}{x}\binom{y+\ell-1}{y}p^{k+\ell}(1-p)^{x+y} \\
f(t) &= \sum_{x=0}^{t} f(x, y=t-x) \\
&= \sum_{x=0}^{t}\binom{x+k-1}{x}\binom{t-x+\ell-1}{t-x}p^{k+\ell}(1-p)^{t} \\
&= \sum_{x=0}^{t}(-1)^{x}\binom{-k}{x}(-1)^{t-x}\binom{-\ell}{t-x}p^{k+\ell}(1-p)^{t} \\
&= (-1)^{t}p^{k+\ell}(1-p)^{t}\sum_{x=0}^{t}\binom{-k}{x}\binom{-\ell}{t-x} \\
&= (-1)^{t}p^{k+\ell}(1-p)^{t}\binom{-k-\ell}{t} \text{ using the hypergeometric identity} \\
&= \binom{t+k+\ell-1}{t}p^{k+\ell}(1-p)^{t};\;\; t=0,1,2,\cdots
\end{aligned}
$$

using the given identity on $(-1)^t\binom{-k-\ell}{t}$. (T has a negative binomial distribution)

8.2.1 (a) Use a multinomial distribution.

$$f(3,11,7,4) \;=\; \frac{25!}{3!\;11!\;7!\;4!}(.1)^3(.4)^{11}(.3)^7(.2)^4$$

(b) Group C's and D's into a single category.

$$f(3,11,11) \;=\; \frac{25!}{3!\;11!\;11!}(.1)^3(.4)^{11}(.5)^{11}$$

(c) Of the 21 non D's we need $3A$'s, 11 B's and $7C$'s. The (conditional) probabilities for the non-D's are: 1/8 for A, 4/8 for B, and $3/8$ for C.
(e.g. $P(A|\overline{D}) \;=\; P(A)/P(\overline{D}) \;=\; .1/.8 = 1/8$)
Therefore $f(3\;A's,\;11B's\;7C's|4D's) \;=\; \frac{21!}{3!11!7!}(\frac{1}{8})^3(\frac{4}{8})^{11}(\frac{3}{8})^7$.

8.2.2 $\mu = .6 \times 12 = 7.2$

$p_1 = P$ (fewer than 5 chips) = $\sum_{x=0}^{4}\frac{7.2^x e^{-7.2}}{x!}$

$p_2 = P$ (more than 9 chips) = $1 - \sum_{x=0}^{9}\frac{7.2^x e^{-7.2}}{x!}$

(a) $\binom{12}{3}p_1^3(1-p_1)^9$

(b) $\frac{12!}{3!7!2}p_1^3\, p_2^7\, (1-p_1-p_2)^2$

(c) Given that 7 have $>$ 9 chips, the remaining 5 are of 2 types - under 5 chips, or 5 to 9 chips

$$P(<5|\leq 9 \text{ chips}) = \frac{P(<5 \text{ and } \leq 9)}{P(\leq 9)} = \frac{p_1}{1-p_2}.$$

Using a binomial distribution,

P (3 under 5|7 over 9) = $\binom{5}{3}\left(\frac{p_1}{1-p_2}\right)^3\left(1-\frac{p_1}{1-p_2}\right)^2$

8.4.1

x	0	1	2
$f_1(x)$	.2	.5	.3

y	0	1
$f_2(y)$	.3	.7

$$\begin{aligned}
E(X) &= (0\times .2)+(1\times .5)+(2\times .3) = 1.1\\
E(Y) &= (0\times .3)+(1\times .7) = .7\\
E(X^2) &= (0^2\times .2)+(1^2\times .5)+(2^2\times .3) = 1.7;\ E(Y^2) = .7\\
\text{Var}(X) &= 1.7-1.1^2 = .49;\ \ \text{Var}(Y) = .7-.7^2 = .21\\
E(XY) &= (1\times 1\times .35)+(2\times 1\times .21) = .77\\
\text{Cov}(X,Y) &= .77-(1.1)(.7) = 0
\end{aligned}$$

Therefore $\rho = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = 0$

While $\rho = 0$ indicates X and Y may be independent (and indeed are in this case), it does not prove that they are independent. It only indicates that there is no linear relationship between X and Y.

8.4.2

(a)

x	2	4	6
$f_1(x)$	3/8	3/8	1/4

y	-1	1
$f_2(y)$	$\frac{3}{8}+p$	$\frac{5}{8}-p$

$E(X) = \left(2\times\frac{3}{8}\right)+\left(4\times\frac{3}{8}\right)+\left(6\times\frac{1}{4}\right) = 15/4;\ E(Y) = -\frac{3}{8}-p+\frac{5}{8}-p = \frac{1}{4}-2p;$

$E(XY) = \left(-2\times\frac{1}{8}\right)+\left(-4\times\frac{1}{4}\right)+\cdots+\left(6\times\left(\frac{1}{4}-p\right)\right) = \frac{5}{4}-12p$

Cov$(X,Y) = 0 = E(XY)-E(X)E(Y) \Rightarrow \frac{5}{4}-12p = \frac{15}{16}-\frac{15}{2}p$

Therefore $p = 5/72$

(b) If X and Y are independent then Cov$(X,Y) = 0$, and so p must be 5/72. But if $p = 5/72$ then

$$f_1(2)f_2(-1) = \frac{3}{8}\times\frac{4}{9} = \frac{1}{6} \neq f(2,-1)$$

Therefore X and Y cannot be independent for any value of p

8.5.1

$x =$	0	1	2
$f_1(x) =$	0.5	0.3	0.2

$$E(X) = (0 \times .5) + (1 \times .3) + 2 \times .2) = 0.7$$

$$E(X^2) = (0^2 \times .5) + (1^2 \times .3) + (2^2 \times .2) = 1.1$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 0.61$$

$$E(XY) = \sum_{\text{all } x,y} xyf(x,y)\text{and this has only two non-zero terms}$$

$$= (1 \times 1 \times 0.2) + (2 \times 1 \times .15) = 0.5$$

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0.01$$

$$\text{Var}(3X - 2Y) = 9\text{Var}(X) + (-2)^2\text{Var}(Y) + 2(3)(-2)\text{Cov}(X,Y)$$

$$= 9(.61) + 4(.21) - 12(.01) = 6.21$$

8.5.2 Let $X_i = \begin{cases} 0, & \text{if the ith pair is alike} \\ 1, & \text{if the ith pair is unlike} \end{cases}$

$$E(X_i) = \sum_{x_i=0}^{1} x_i f(x_i) = 1f(1) = P(\text{ON OFF } \cup \text{ OFF ON})$$

$$= (.6)(.4) + (.4)(.6) = .48$$

$$E(X_i^2) = E(X_i) = .48 \text{ (for } X_i = 0 \text{ or } 1)$$

$$\text{Var }(X_i) = .48 - (.48)^2 = .2496$$

Pairs are independent if they have no common points, but may not be independent if the pairs are adjacent.

$$E(X_i X_{i+1}) = \sum x_i x_{i+1}\, f(x_i,\, x_{i+1})$$

$$= 1 \times 1 \times f(1,1) = P(\text{ON OFF ON } \cup \text{ OFF ON OFF})$$

$$= (.6)(.4)(.6) + (.4)(.6)(.4) = .24$$

$$\text{Cov }(X_i,\, X_{i+1}) = E(X_i\, X_{i+1}) - E(X_i)E(X_{i+1})$$

$$= .24 - (.48)^2 = .0096$$

$$E(\sum X_i) = \sum_{i=1}^{24} E(X_i) = 24 \times .48 = 11.52$$

$$\text{Var }(\sum X_i) = \sum_{i=1}^{24} \text{ Var }(X_i) + 2\sum_{i=1}^{23} \text{ Cov }(X_i,\ X_{i+1}) = (24 \times .2496) + (2 \times 23 \times .0096)$$
$$= 6.432$$

8.5.3 $\rho = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} = 0.5 \Rightarrow \text{Cov}(X,Y) = 0.5\sqrt{1.69 \times 4} = 1.3$
$\text{Var}(U) = \text{Var}(2X - Y) = 4\sigma_X^2 + \sigma_Y^2 - 4\text{Cov}(X,Y) = 5.56$
Therefore $s.d.(U) = 2.36$

8.5.4 $\text{Cov }(X_{i-1}, X_i) = \text{Cov }(Y_{i-2} + Y_{i-1}, Y_{i-1} + Y_i)$
$\text{Cov }(Y_{i-2}, Y_{i-1}) + \text{Cov }(Y_{i-2}, Y_i) + \text{Cov }(Y_{i-1}, Y_{i-1}) + \text{Cov }(Y_{i-1}, Y_i)$
$= 0 + 0 + \text{Var }(Y_{i-1}) + 0 = \sigma^2$
Also, $\text{Cov }(X_i, X_j) = 0$ for $j \neq i \pm 1$ and $\text{Var}(X_i) = \text{Var}(Y_{i-1}) + \text{Var}(Y_i) = 2\sigma^2$
$\text{Var }(\sum X_i) = \sum \text{Var}(X_i) + 2\sum_{i=2}^{n} \text{Cov }(X_{i-1}, X_i) = n(2\sigma^2) + 2(n-1)\sigma^2 = (4n-2)\sigma^2$

8.5.5 Using X_i as defined,
$E(X_i) = \sum_{x_i=0}^{1} x_i f(x_i) = f(1) = E\left(X_i^2\right)$ since $X_i = X_i^2$
$E\left(X_1\right) = E\left(X_{24}\right) = .9$ since only 1 cut is needed
$E\left(X_2\right) = E\left(X_3\right) = \cdots = E\left(X_{23}\right) = .9^2 = .81$ since 2 cuts are needed.
Var(X_1) = Var$(X_{24}) = .9 - .9^2 = .09$
Var(X_2) = Var$(X_3) = \cdots$ = Var$(X_{23}) = .81 - .81^2 = .1539$
$\text{Cov }(X_i, X_j) = 0$ if $j \neq i \pm 1$ since there are no common pieces and cuts are independent.
$E\left(X_i X_{i+1}\right) = \sum x_i x_{i+1} f\left(x_i, x_{i+1}\right) = f(1,1)$
(product is 0 if either x_i or x_{i+1} is a 0)

$$= \begin{cases} .9^2 & \text{for } i = 1 \text{ or } 23\text{........2 cuts needed} \\ .9^3 & \text{for } i = 2, .., 22\text{........3 cuts needed} \end{cases}$$

$$\text{Cov }(X_i, X_{i+1}) = E\left(X_i X_{i+1}\right) - E(X_i)E\left(X_{i+1}\right)$$

$$= \begin{cases} .9^2 - (.9)(.9^2) = .081 & \text{for } i = 1 \text{ or } 23 \\ .9^3 - (.9^2)(.9^2) = .0729 & \text{for } i = 2, \cdots, 22 \end{cases}$$

$E\left(\sum_{i=1}^{24} X_i\right) = \sum_{i=1}^{24} E(X_i) = (2 \times .9) + (22 \times .81) = 19.62$
Var $\left(\sum_{i=1}^{24} X_i\right) = \sum_{i=1}^{24}$ Var $(X_i) + 2\sum_{i<j}$ Cov(X_i, X_j)
$= (2 \times .09) + (22 \times .1539) + 2\left[(2 \times .081) + (21 \times .0729)\right]$
$= 6.9516$
Therefore s.d.$(\sum X_i) = \sqrt{6.9516} = 2.64$

9.1.1 (a) $\int_{-1}^{1} kx^2 dx = k\frac{x^3}{3}|_{-1}^{1} = \frac{2k}{3} = 1$
Therefore $k = 3/2$

(b) $F(x) = \begin{cases} 0; & x \leq -1 \\ \int_{-1}^{x} \frac{3}{2}x^2 dx = \frac{x^3}{2}|_{-1}^{x} = \frac{x^3}{2} + \frac{1}{2}; & -1 < x < 1 \\ 1; & x \geq 1 \end{cases}$

(c) $P(-.1 < X < .2) = F(.2) - F(-.1) = .504 - .4995 = .0045$

(d) $E(X) = \int_{-1}^{1} x \times \frac{3}{2}x^2 dx = \frac{3}{2} \int_{-1}^{1} x^3 dx = \frac{3}{8}x^4|_{-1}^{1} = 0; \quad E(X^2) = \int_{-1}^{1} x^2 \frac{3}{2}x^2 dx = \frac{3}{10}x^5|_{-1}^{1} = 3/5$ Therefore $\text{Var}\ (X) = E(X^2) - \mu^2 = 3/5$

(e) $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$
$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \left[\frac{(\sqrt{y})^3}{2} + \frac{1}{2}\right] - \left[\frac{(-\sqrt{y})^3}{2} + \frac{1}{2}\right] = y^{3/2}$
$f(y) = \frac{d}{dy} F_Y(y) = \frac{3}{2}\sqrt{y};\ 0 \leq y < 1$

9.1.2 (a) $F(\infty) = 1 = \lim_{x\to\infty} \frac{kx^n}{1+x^n} = \lim_{x\to\infty} \frac{k}{\frac{1}{x^n}+1} = k$
Therefore $k = 1$

(b) $f(x) = \frac{d}{dx}F(x) = \frac{nx^{n-1}}{(1+x^n)^2};\ \ x > 0$

(c) Let m be the median. Then $F(m) = .5 = \frac{m^n}{1+m^n}$
Therefore $m^n = 1$ and so the median is 1

9.2.1 $F(x) = \int_{-1}^{x} \frac{3}{2}x^2 dx = \frac{x^3+1}{2}$.
If $y = F(x) = \frac{x^3+1}{2}$ is a random number between 0 and 1, then

$$x = (2y - 1)^{1/3}$$

For $y = .27125$ we get $x = (-.4574)^{1/3} = -.77054$

9.3.1 Let the time to disruption be X.

$$\text{Then } P(X \leq 8) = F(8) = 1 - e^{-8/\theta} = .25$$

Therefore $e^{-8/\theta} = .75$. Take natural logs.
$\theta = -\frac{8}{\ln\ .75} = 27.81$ hours.

9.3.2 (a) $F(x) = P$ (distance $\leq x$) = $1 - P$ (distance $> x$)
= $1 - P$ (0 flaws or 1 flaw within radius x)
The number of flaws has a Poisson distribution with mean $\mu = \lambda\pi x^2$
$F(x) = 1 - \frac{\mu^0 e^{-\mu}}{0!} - \frac{\mu^1 e^{-\mu}}{1!} = 1 - e^{-\lambda\pi x^2}\left(1 + \lambda\pi x^2\right)$
$f(x) = \frac{d}{dx}F(x) = 2\lambda^2\pi^2 x^3 e^{-\lambda\pi x^2};\ \ x > 0$

(b) $\mu = E(X) = \int_0^\infty x 2\lambda^2\pi^2 x^3 e^{-\lambda\pi x^2} = \int_0^\infty 2\lambda^2\pi^2 x^4 e^{-\lambda\pi x^2} dx$
Let $y = \lambda\pi x^2$. Then $dy = 2\lambda\pi x dx$, so $dx = \frac{dy}{2\sqrt{\lambda\pi y}}$
$\mu = \int_0^\infty 2y^2 e^{-y} \frac{dy}{2\sqrt{\lambda\pi y}} = \frac{1}{\sqrt{\lambda\pi}} \int_0^\infty y^{3/2} e^{-y} dy$
$= \frac{1}{\sqrt{\lambda\pi}}\Gamma\left(\frac{5}{2}\right) = \frac{1}{\sqrt{\lambda\pi}}\left(\frac{3}{2}\right)\Gamma\left(\frac{3}{2}\right) = \frac{1}{\sqrt{\lambda\pi}}\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \frac{\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}}{\sqrt{\lambda\pi}} = \frac{3}{4\sqrt{\lambda}}$

9.5.1 (a) $P\left(8.4 < X < 12.2\right) = P\left(\frac{8.4-10}{2} < Z < \frac{12.2-10}{2}\right)$.

= $P(-.8 < Z < 1.1)$
= $F(1.1) - F(-.8)$
= $F(1.1) - [1 - F(.8)]$
= $.8643 - (1 - .7881) = .6524$

(see Figure 10.2)

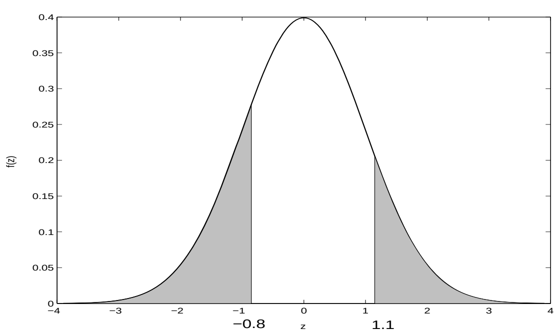


Figure 10.2:

(b) $2Y - X \sim N\left(2(3) - 10 = -4,\ 2^2(100) + (-1)^2(4) = 404\right)$.
$P(2Y > X) = P(2Y - X > 0)$
$= P(Z > \frac{0-(-4)}{\sqrt{404}} = .20) = 1 - F(.20) = 1 - .5793 = .4207$

(c) $\overline{Y} \sim N(3,\ \frac{100}{25} = 4)$
$P(\overline{Y} < 0) = P(Z < \frac{0-3}{2} = -1.5)$
$= P(Z > 1.5) = 1 - F(1.5)$
$= 1 - .9332 = .0668$ (see Figure 10.3)

9.5.2 $P\left(|X - \mu| < \sigma\right) = P\left(-\sigma < X - \mu < \sigma\right) = P\left(-1 < Z < 1\right)$

$= F(1) - [1 - F(1)] = .8413 - (1 - .8413) = 68.26\%$ (about 2/3)

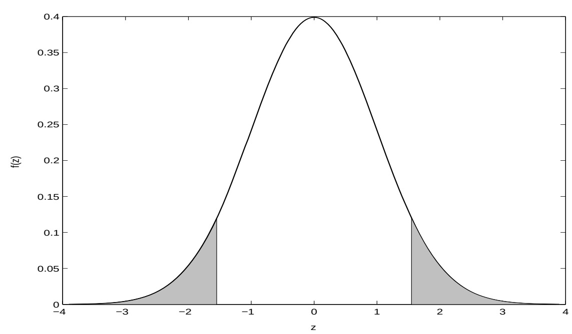


Figure 10.3:

$P(|X-\mu| < 2\sigma) = P(-2\sigma < X - \mu < 2\sigma) = P(-2 < Z < 2)$

$= F(2) - [1 - F(2)] = .9772 - (1 - .9772) = 95.44\%$ (about 95%)
Similarly, $P(|X-\mu| < 3\sigma) = P(-3 < Z < 3) = 99.73\%$ (over 99%)

9.5.3 (a) $2X - Y \sim N(2(5) - 7 = 3,\ 2^2(4) + 9 = 25)$
$P(|2X - Y| > 4) = P(2X - Y > 4) + P(2X - Y < -4)$
$= P\left(Z > \frac{4-3}{5} = .20\right) + P\left(Z < \frac{-4-3}{5} = -1.40\right)$
$= .42074 + .08076 = .5015$

(b) $P\left(\left|\overline{X} - 5\right| < 0.1\right) = P\left(|Z| < \frac{0.1}{2/\sqrt{n}}\right) = .98$
(since ...$\overline{X} \sim N(5, 4/n)$)
$F\left(\frac{0.1}{2/\sqrt{n}}\right) = .99$
Therefore $.05\sqrt{n} = 2.3263$
Therefore $n = 2164.7$. Take $n = 2165$ observations.

9.6.1 Let X be the number germinating.
Then $X \sim Bi(100, .8)$.
$P(X \geq 75) = \sum_{x=75}^{100} \binom{100}{x}(.8)^x(.2)^{100-x}$. (see Figure 10.4)
Approximate using a normal distribution with $\mu = np = 80$ and $\sigma^2 = np(1-p) = 16$.
$P(X \geq 75) \simeq P(X > 74.5)$
$= P(Z > \frac{74.5-80}{4} = -1.375)$
$\simeq F(1.38) = .9162$
Possible variations on this solution include calculating
$F(1.375)$ as $\frac{F(1.37)+F(1.38)}{2}$ and realizing that $X \leq 100$ means
$P(X \geq 75) \simeq P(74.5 < X < 100.5)$. However
$P(X \geq 100.5) \simeq P(Z > \frac{100.5-80}{4} = 5.125) \simeq 0$ so we get the same answer as before.

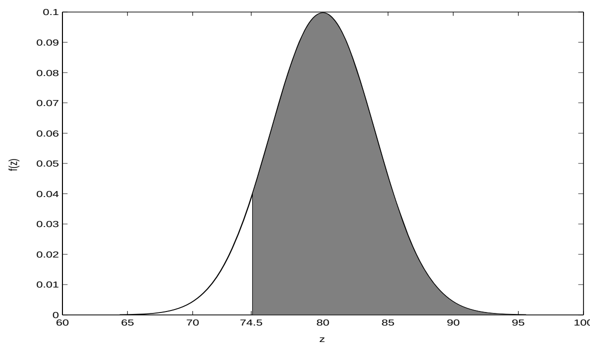


Figure 10.4:

9.6.2 Let X_i be the cost associated with inspecting part i

$E\left(X_i\right) = (0 \times .6) + (10 \times .3) + (100 \times .1) = 13$

$E\left(X_i^2\right) = \left(0^2 \times .6\right) + \left(10^2 \times .3\right) + \left(100^2 \times .1\right) = 1030$

$\text{Var}\left(X_i\right) = 1030 - 13^2 = 861$

By the central limit theorem

$\sum_{i=1}^{80} X_i \sim N(80 \times 13 = 1040, 80 \times 861 = 68,880)$ approx.

Since $\sum X_i$ increases in \$10 increments,

$P(\sum X_i > 1200) \simeq P\left(Z > \frac{1205-1040}{\sqrt{68,880}} = 0.63\right) = .2643$

11. Answers to End of Chapter Problems

Chapter 2:

2.1 (a) Label the profs A, B, C and D.

$$S = \{AA, AB, AC, AD, BA, BB, BC, BD, CA, CB, CC, CD, DA, DB, DC, DD\}$$

(b) 1/4

2.2 (a) $\{HHH,\ HHT,\ HTH,\ THH,\ HTT,\ THT,\ TTH,\ TTT\}$ (b) $\frac{1}{4}$;

2.3 (1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5); 0.4;

2.4 (c) $\frac{1}{4}, \frac{3}{8}, \frac{1}{4}, 0$

2.5 (a) $\frac{8}{27}, \frac{1}{27}, \frac{2}{9}$ (b) $\frac{(n-1)^3}{n^3}, \frac{(n-2)^3}{n^3}, \frac{n^{(3)}}{n^3}$ (c) $\frac{(n-1)^r}{n^r}, \frac{(n-2)^r}{n^r}, \frac{n^{(r)}}{n^r}$

2.6 (a) .018 (b) .020 (c) 18/78 = .231

2.7 (b) .978

Chapter 3:

3.1 (a) 4/7 (b) 5/42 (c) 5/21;

3.2 (a) (i) $\frac{(n-1)^r}{n^r}$ (ii) $\frac{n^{(r)}}{n^r}$

(b) All n^r outcomes are equally likely. That is, all n floors are equally likely to be selected, and each passenger's selection is unrelated to each other person's selection. Both assumptions are doubtful since people may be travelling together (e.g. same family) and the floors may not have equal traffic (e.g. more likely to use the stairs for going up 1 floor than for 10 floors);

3.3 (a) 5/18 (b) 5/72;

3.4 $\frac{\binom{4}{2}\binom{12}{4}\binom{36}{7}}{\binom{52}{13}}$

3.5 (a) 1/50,400 (b) 7/45;

3.6 (a) 1/6 (b) 0.12;

3.7 Values for $r = 20, 40$ and 60 are .589, .109 and .006.

3.8 (a) $\frac{1}{n}$ (b) $\frac{2}{n}$

3.9 $\frac{1+3+\cdots+(2n-1)}{\binom{2n+1}{3}} = \frac{n^2}{\binom{2n+1}{3}}$

3.10 (a) (i) .0006 (ii) .0024 (b) $\frac{10^{(4)}}{10^4} = .504$

3.11 (a) $\binom{6}{2}\binom{19}{3}/\binom{25}{5}$ (b) 15

3.12 (a) $1/\binom{49}{6}$ (b) $\binom{6}{5}\binom{43}{1}/\binom{49}{6}$ (c) $\binom{6}{4}\binom{43}{2}/\binom{49}{6}$ (d) $\binom{6}{3}\binom{43}{3}/\binom{49}{6}$

3.13

(a) $1 - \frac{\binom{48}{3}}{\binom{50}{3}}$ (b) $1 - \frac{\binom{45}{2}}{\binom{47}{2}}$ (c)$\frac{\binom{48}{3}}{\binom{50}{5}}$

Chapter 4:

4.1 .75, .6, .65, 0, 1, .35, 1

4.2 A- .01, B - .72, C - $.9^3$, D - $.5^3$, E - $.5^2$

4.3 $\frac{1}{6}$

4.4 (a) 0.0576 (b) 0.4305 (c) 0.0168 (d) 0.5287

4.5 .44

4.6 0.7354

4.7 (a) .3087 (b) .1852;

4.9 .342

4.10 (a) .1225, .175 (b) .395

4.11 $(\frac{f}{F}) = (\frac{m}{M})$

4.14 (a) $\frac{1}{30} + \frac{4P}{5}$ (b) $p = \frac{(30x/n)-1}{24}$ (c) $\frac{24p}{1+24p}$

4.15 .9, .061, .078

4.16 (a) .024 (b) 8 on any one wheel and 1 on the others

4.17 (a) .995 and .005 (b) .001

4.18 (a) .99995 (b) .99889 (c)

$$0.2 + 0.1 + 0.1 - (.2)(.1) - (.2)(.1) - (.1)(.1) + (.2)(.1)(.1) = 0.352$$

4.19 (a) $\frac{r}{r+1999}$; $0.005, 0.0148, 0.0476$ (b) 2.1%

Chapter 5:

5.4 .545

Chapter 6:

6.1 (a) .623, .251; for males, .408, .103 (b) .166

6.2 (a) $f(0) = 0, f(x) = 2^{-x}\ (x = 1, 2, \dots)$ (b) $f(5) = \frac{1}{32};\ P(X \geq 5) = \frac{1}{16}$

6.4 $\frac{p(1-p)^r}{1-(1-p)^4};\ \ r = 0, 1, 2, 3$

6.5 (a) .0800 (b) .171 (c) .00725

6.6 (a) .010 (b) .864

6.7 (a)$\frac{4}{15}$ (b) $\binom{74}{y}\binom{76}{12-y}/\binom{150}{12}$ (c) .0176

6.8 0.9989

6.9 (a) .0758 (b) .0488 (c) $\binom{10}{y}(e^{-10\lambda})^y(1-e^{-10\lambda})^{10-y}$ (d) $\lambda = .12$

6.10 (a) .0769 (b) 0.2019; 0.4751

6.11 (a) 0.2753 (b) 0.1966 (c) 0.0215

6.12 (b) enables us to approximate hypergeometric distribution by binomial distribution when n is large and p near 0.

6.13 (a) $1 - \left[\sum_{x=0}^{k-1} \frac{\lambda^x e^{-\lambda}}{x!}\right]^n$ (b) (Could probably argue for other answers also). Budworms probably aren't distributed at a uniform rate over the forest and may not occur singly

6.14 (a) .2048 (b) .0734 (c) .428 (d) .1404

6.15 $\frac{\binom{35}{x}\binom{70}{7}}{\binom{105}{x+7}} \frac{63}{98-x};\ \ x = 0, 1, \cdots, 35$

6.16 (a) .004264; .006669 (b) .0032 (c) (i) $\binom{1399}{11}(.004264)^{12}(.995736)^{1388}$ (ii) 9.336×10^{-5}

On the first 1399 attempts we essentially have a binomial distribution with $n = 1399$ (large) and $p = .004264$ (near 0)

6.17 (a) $\binom{n}{x}\left(e^{-0.96}\right)^x\left(1-e^{-0.96}\right)^{n-x};\ \ x = 0, 1, \cdots, n$ (b) $\lambda \leq 0.866$ bubbles per m^2

6.18 0.5;

x	0	1	2	3	4	5
$f(x)$	0	.05	.15	.05	.25	.5

; 0.3

6.19 (a) $(1-p)^y$ (b) $Y = 0$ (c) $p/[1-(1-p)^3]$ (d) $P(R = r) = \frac{p(1-p)^r}{1-(1-p)^3}$ for $r = 0, 1, 2$

6.20 (a) .555 (b) .809; .965 (c) .789; .946 (d) $n = 1067$

6.21 (a) $\binom{x-1}{999}(.3192^{1000})(.6808^{x-1000})$ (b) .002, .051, .350, .797 (c) $\binom{3200}{y}(.3192^y)(.6808^{3200-y})$; .797

Chapter 7:

7.1 2.775; 2.574375

7.2 -$3

7.3 $16.90

7.4 (a) 3 cases (b) 32 cases

7.5 (a) - 10/37 dollars in both cases (b) .3442; 0.4865

7.6 $.94

7.7 (b) $n + \frac{n}{k} - n(1-p)^k$, which gives $1.01n, 0.249n, 0.196n$ for $k = 1, 5, 10$

7.8 50

7.9 (a) $\frac{p}{1-(1-p)e^t}$; (b) $\frac{1-p}{p}; \frac{1-p}{p^2}$

7.11 (a) Expand $M(t)$ in a power series in powers of e^t, i.e. $M_X(t) = \frac{1}{3}e^t + \frac{2}{9}e^{2t} + \frac{4}{27}e^{3t} + \frac{8}{81}e^{4t} + \frac{16}{243}e^{5t} + ...$Then $P(X = j)$ =coefficient of $e^{jt} = \frac{1}{3}(\frac{2}{3})^{j-1}, j = 1, 2, ...$ (b) Similarly $M_X(t) = e^{-2} + 2e^{-2}e^t + 2e^{-2}e2^t + \frac{4}{3}e^{-2}e^{3t} + \frac{2}{3}e^{-2}e^{4t} + \frac{4}{15}e^{-2}e^{5t} + ...$ Then $P(X = j) = e^{-2}\frac{2^j}{j!}, j = 0, 1, ...$

Chapter 8:

8.1 (a) no $f(1,0) \neq f_1(1)f_2(0)$ (b) 0.3 and 1/3

8.2 (a) mean = 0.15, variance = 0.15

8.3 (a) No (b) 0.978 (c) .05

8.4 (a) $\frac{(x+y+9)!}{x!y!9!}\, p^x\, q^y(1-p-q)^{10}$ (b) $\binom{x+y+9}{y}\, q^y(1-q)^{x+10};\ y = 0, 1, 2, ...;$ 6 .0527

8.5 (b) - .10 dollars (c) $d = .95/n$

8.7 $\frac{\binom{5}{x}\binom{3}{2-x}\binom{5-x}{y-x}\binom{1+x}{2+x-y}}{\binom{8}{4}};\ x = 0, 1, 2,;\ y = max(1, x), x+1, x+2;$ (b) note e.g. that $f_1(0) \neq 0$; $f_2(3) \neq 0$, but $f(0,3) = 0$

8.8 (a) $\frac{\binom{2}{x}\binom{1}{y}\binom{7}{3-x-y}}{\binom{10}{3}};\ x = 0, 1, 2,$ and $y = 0, 1$ (b) $f_1(x) = \binom{2}{x}\binom{8}{3-x}/\binom{10}{3};\ x = 0, 1, 2;$ $f_2(y) = \binom{1}{y}\binom{9}{3-y}/\binom{10}{3};\ y = 0, 1$ (c) 49/120 and 1/2

8.9 (a) $k\frac{2^x e^2}{x!};\ x = 0, 1, 2, ...$ (b) e^{-4} (c) Yes. $f(x,y) = f_1(x)f_2(y)$
(d) $\frac{4^t e^{-4}}{t!};\ t = 0, 1, 2, \ldots$

8.10 (b) .468

8.11 (a) $\frac{40!}{(10!)^4}\left(\frac{3}{16}\right)^{20}\left(\frac{5}{16}\right)^{20}$ (b) $\binom{40}{16}(1/2)^{40}$ (c) $\binom{16}{10}\left(\frac{3}{8}\right)^{10}\left(\frac{5}{8}\right)^{6}$

8.12 (a) 1.76 (b) $P(Y = y) = \sum_{x=y}^{8}\binom{x}{y}(\frac{1}{2})^x f(x);\ E(Y) = 0.88 = \frac{1}{2}E(X)$

8.13 (a) Multinomial (b) .4602 (c) $5700

8.15 207.867

8.16 (a) $Bi(n, p+q)$ (b) $n(p+q)$ and $n(p+q)(1-p-q)$ (c) $-npq$

8.17 (a) $\mu_U = 2, \mu_V = 0,\ \sigma_U^2 = \sigma_V^2 = 1$ (b) 0
(c) no. e.g. $P(U = 0) \neq 0; P(V = 1) \neq 0; P(U = 0$ and $V = 1) = 0$

8.19 -1

8.20 (a) 1.22 (b) 17.67%

8.21 $p^3(4+p); 4p^3(1-p^3) + p^4(1-p^4) + 8p^5(1-p^2)$

8.22

8.23 The transition matrix is

$$P=\begin{bmatrix} 0 & 1 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

from which, solving $\pi' P = \pi'$ and rescaling so that the sum of the probabilities is one, we obtain $\pi' = (0.4, 0.45, 0.15)$, the long run fraction of time spent in cities A,B,C respectively.

8.24 By arguments similar to those in section 8.3, the limiting matrix has rows all identically π' where the vector π' are the stationary probabilities satisfying $\pi' P = \pi'$ and

$$P=\begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

The solution is $\pi' = (0.1, 0.6, 0.3)$ and the limit is

$$\begin{bmatrix} 0.1 & 0.6 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.6 & 0.3 \end{bmatrix}$$

8.25 With $Z = X + Y$,

$$\begin{aligned} M_Z(t) = Ee^{t(X+Y)} &= E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t) = \exp(-\lambda_1 + \lambda_1 e^t)\exp(-\lambda_2 + \lambda_2 e^t) \\ &= \exp(-(\lambda_1 + \lambda_2) + (\lambda_1 + \lambda_2)e^t) \end{aligned}$$

and since this is the MGF of a Poisson($\lambda_1 + \lambda_2$) distribution, this must be the distribution of Z.

8.26 If today is raining, the probability of Rain, Nice, Snow three days from now is obtainable from the first row of the matrix P^3, i.e. $(0.406\ 0.203\ \ 0.391)$.The probabilities of the three states in five days, given (1) today is raining (ii) today is nice (iii) today is snowing are the three rows of the matrix P^5. In this case call rows are identical to three decimals; they are all equal the equilibrium distribution $\pi' = (0.400\ 0.200\ 0.400)$.

(b) If *a*>*b*, and both parties raise then the probability B wins is

$$\frac{13-b}{2(13-a)} < \frac{1}{2}$$

and the probability A wins is 1 minus this or $\frac{13-2a+b}{2(13-a)}$. If $a \le b$, then the probability A wins is

$$\frac{13-a}{2(13-b)}.$$

8.27 (a) In the special case *b=1*, count the number of possible pairs (i,j) for which $A=i\geq a$ and $B=j>A$.

1	$(A=12)$
2	$(A=11)$
:	
$13-a$	$(A=a)$
$\frac{(13-a)(13-a+1)}{2}$	Total

This leads to

$$P(B>A,A\geq a)=\frac{(13-a)(13-a+1)}{2(13^2)}$$

Similarly, since the number of pairs (A,B) for which $A\geq a$, and $B<a$ is $(13-a+1)(a-1)$, we have

$$\begin{aligned}P(A>B,A\geq a)&=P(A>B,A\geq a,B\geq a)+P(A>B,A\geq a,B<a)\\&=\frac{(13-a)(13-a+1)}{2(13^2)}+\frac{(13-a+1)(a-1)}{13^2}=\frac{(14-a)(a+11)}{2(13^2)}\end{aligned}$$

Therefore, in case *b=1*, the expected winnings of A are

$$\begin{aligned}&-1P(\text{B raises, A does not})-6P(\text{both raise, B wins})+6P(\text{both raise, A wins})\\&=-1P(A<a)-6P(B>A,A\geq a)+6P(A>B,A\geq a)\\&=-1\times\frac{a-1}{13}-6\times\frac{(13-a)(13-a+1)}{2(13^2)}+6\times\frac{(14-a)(a+11)}{2(13^2)}\\&=-\frac{6}{169}a^2+\frac{77}{169}a-\frac{71}{169}=-\frac{1}{169}\left(a-1\right)\left(6a-71\right)\end{aligned}$$

:whose maximum (over real a) is at $77/12$ and over integer a, at 6 or 7. For *a=1,2,...,13* this gives expected winnings of *0, 0.38462, 0.69231, 0.92308, 1.0769, 1.1538, 1.1538, 1.0769, 0.92308, 0.69231, 0.38462, 0, -0.46154* respectively, and the maximum is for *a=6* or 7.

(b) We want $P(A>B,A\geq a,B\geq b)$. Count the number of pairs (i,j) for which $A\geq a$ and $B\geq b$ and $A>B$. Assume that $b\leq a$.

1	$(B=12)$
2	$(B=11)$
:	:
$13-a$	$(B=a)$
$(a-b)(13-a+1)$	$(b\leq B<a)$

for a total of

$$\frac{(13-a)(13-a+1)}{2} + (a-b)(13-a+1) = \frac{1}{2}(14-a)(13+a-2b)$$

and

$$P(A > B, A \geq a, B \geq b) = \frac{(14-a)(13+a-2b)}{2(13^2)}$$

Similarly

$$P(A < B, A \geq a, B \geq b) = P(A < B, A \geq a) = \frac{(13-a)(13-a+1)}{2(13^2)}$$

Therefore the expected return to A (still assuming $b \leq a$) is

$$\begin{aligned} &-1P(A{<}\, a, B \geq b) + 1P(A \geq a, B < b) + 6P(A > B, A \geq a, B \geq b) - 6P(A < B, A \geq a, B \geq b) \\ &\quad = -1\frac{(a-1)(13-b+1)}{13^2} + 1\frac{(b-1)(13-a+1)}{13^2} + 6\frac{(14-a)(13+a-2b)}{2(13^2)} - 6\frac{(13-a)(13-a+1)}{2(13^2)} \\ &\quad = \frac{1}{13^2}(71-6a)(a-b) \end{aligned}$$

If $b > a$ then the expected return to B is obtained by switching the role of a, b above, namely

$$\frac{1}{13^2}(71-6b)(b-a)$$

and so the expected return to A is

$$\frac{1}{13^2}(71-6b)(a-b)$$

In general, then the expected return to A is

$$\frac{1}{13^2}(71-6\max(a,b))(a-b)$$

(c) By part (b), A's possible expected profit per game for *a=1,2,...,13 and* $b = 11$ is

$$\frac{1}{13^2}(71-6\max(a,11))(a-11) = -\frac{6}{13^2}\left(\max(a,11) - \frac{71}{6}\right)(a-11)$$

For $a = 1, 2, ...13$ these are, respectively,-0.2959, -0.2663, -0.2367, -0.2071, -0.1775, -0.1479, -0.1183,-0.0888, -0.0592, -0.0296, 0, -0.0059, -0.0828. There is no strategy that provides a positive expected return. The optimal is the break-even strategy *a=11*. (Note: in this two-person zero-sum game, *a= 11* and *b=11* is a minimax solution)

8.28 (a) Show that you can determine the probability of the various values of X_{t+1} knowing only the state X_t (and without knowing the previous states).

(b) For example the long-run probability of the state (i, j, k) is, with $q_i = p_i/(1-p_i)$,

$$q_i p_j$$

(c) The probability that record j is in position $k = 1, 2, 3$ is, with $Q = \sum_{i=1}^{3} q_i$, $p_j, (Q - q_j)p_j, 1 - p_j(1 + Q - q_j)$ respectively. The expected cost of accessing a record in the long run is

$$\sum_{j=1}^{3} (p_j^2 + 2p_j^2(Q - q_j) + 3p_j(1 - p_j(1 + Q - q_j))) \tag{10.7}$$

Substitute $p_1 = 0.1, p_2 = 0.3, p_3 = 0.6$ so $q_1 = \frac{1}{9}, q_2 = \frac{3}{7}, q_3 = \frac{6}{4}$ and $Q = \frac{1}{9} + \frac{3}{7} + \frac{6}{4} = 2.0397$ and (10.7) is 1.7214.

(d) If they are in random order, the expected cost= $1(\frac{1}{3}) + 2(\frac{1}{3}) + 3(\frac{1}{3}) = 2$. If they are ordered in terms of decreasing p_j, expected cost is $p_3^2 + 2p_2^2 + 3p_1^2 = 0.57$

Chapter 9:

9.1 $f(y) = (\frac{5}{6})(\frac{6}{\pi})^{\frac{1}{3}} y^{-\frac{2}{3}}$ for $.036\pi \le y \le \frac{\pi}{6}$

9.2 (a) $k = .75$; $F(x) = .75\left(\frac{2}{3} + x - \frac{x^3}{3}\right)$ for $-1 \le x \le 1$

(b) Find c such that $c^3 - 3c + 1.9 = 0$. This gives $c = .811$

9.3 (a) 1/2, 1/24 (b) 0.2828 (c) 0.2043

9.4 $f(y) = 1;\ 0 < y < 1$

9.5 (a) $\alpha > -1$ (b) $0.5^{\alpha+1}$, $\frac{\alpha+1}{\alpha+2}$ (c) $\frac{\alpha+1}{t^{\alpha+2}};\ 1 < t < \infty$

9.6 (a) $(1 - e^{-2})^3$ (b) $e^{-.4}$

9.7 $1000 \log 2 = 693.14$

9.8 (a) .0668, .2417, .3829, .2417, .0668 (b) .0062 (c) .0771

9.9 (a) .5 (b) $\mu \ge 2.023$

9.10 .4134

9.11 (a) .3868 (b) .6083 (c) 6.94

9.12 (a) .0062 (b) .9927

9.13 (a) .2327, .1841 (b) .8212, .8665; Guess if $p_i = 0.45$, don't guess if $p_i = 0.55$

9.14 6.092 cents

9.15 574

9.16 (a) 7.6478, 88.7630 (b) 764.78, 8876.30, people within pooled samples are independent and each pooled sample is independent of each other pooled sample. (c) 0.3520

9.17 .5969

9.18 (a) .6728 (b) 250,088

9.19 (a) $X \sim N(-.02n, .9996n)$ (b) $P(X \geq 0) = .4641, .4443, .4207$ (using table) for $n = 20, 50, 100$ The more you play, the smaller your chance of winning. (c) 1264.51

With probability .99 the casino's profit is at least \$1264.51.

9.20 (a) X is approximately $N\left(-\frac{n}{2}, \frac{5n}{12}\right)$ (b) (i)$P(X > 0) = P(Z > 2.45) = 0.0071.$ (ii) $P(X > 0) = P(Z > 5.48) \simeq 0.$

9.21 (a) (i) .202 (ii) .106 (b) .045, .048

9.22 (a) False positive probabilities are .048, .091, .023 for (i), (ii), (iii). False negative probabilities are .023, .091, .048.

9.23

9.24 Let Y =total change over day. Given $N = n$, Y has a Normal$(0, n\sigma^2)$ distribution and therefore

$$E[e^{tY}|N = n] = \exp(n\sigma^2 t^2/2)$$

$$M_Y(t) = E[e^{tY}] = \sum_n E[e^{tY}|N = n]P(N = n) = e^{-\lambda}\sum_n \exp(n\sigma^2 t^2/2)\frac{\lambda^n}{n!}$$

$$= e^{-\lambda}\sum_n \frac{(e^{\sigma^2 t^2/2}\lambda)^n}{n!} = \exp(-\lambda + e^{\sigma^2 t^2/2}\lambda)$$

Not a MGF in this course at least. The mean is $M_Y'(0) = 0$ and the variance is $M''_Y(0) = \lambda\sigma^2$.

(a) $\exp(t + t^2)$

(b) $\exp(2t + 2t^2)$

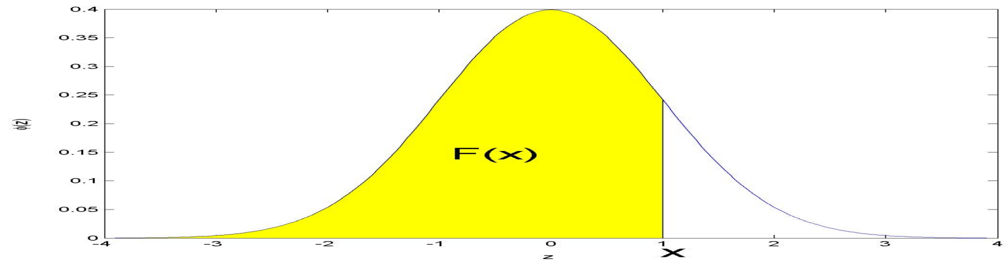
(c) $\exp(nt + nt^2)$

(d) $\exp(t^2)$

Summary of Distributions

Discrete				
Notation and Parameters	**Probability function** $f(x)$	**Mean**	**Variance**	**Moment generating function** $M_X(t)$
Binomial(n,p) $x=0,1,...,n$ $0<p<1, q=1-p$	$\binom{n}{x}\ p^x q^{n-x}$	np	npq	$(pe^t+q)^n$
Bernoulli(p) $x=0,1$ $0<p<1, q=1-p$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	(pe^t+q)
Negative Binomial(k,p) $x=0,1,...$ $0<p<1, q=1-p$	$\binom{x+k-1}{x}p^k q^x$	$\frac{kq}{p}$	$\frac{kq}{p^2}$	$(\frac{p}{1-qe^t})^k$
Geometric(p) $x=0,1,...$ $0<p<1, q=1-p$	pq^x	$\frac{q}{p}$	$\frac{q}{p^2}$	$(\frac{p}{1-qe^t})$
Hypergeometric(N,r,n) $x=0,1,...$min(r,n) $r<N, n<N$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\frac{nr}{N}$	$n\frac{r}{N}(1-\frac{r}{N})\frac{N-n}{N-1}$	intractible
Poisson(λ)	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$e^{\lambda(e^t-1)}$
Continuous	**p.d.f.** $f(x)$	**Mean**	**Variance**	**Moment generating function** $M_X(t)$
Uniform(a,b) $a<x<b$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$
Exponential(θ) $0<x,\ \ 0<\theta$	$\frac{1}{\theta}e^{-x/\theta}$	θ	θ^2	$\frac{1}{1-\theta t}$
Normal(μ,σ^2) $-\infty<x<\infty$... $-\infty<\mu<\infty,\ \ \sigma^2>0$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2	$e^{\mu t+\sigma^2t^2/2}$

Probabilities for Standard Normal N(0,1) Distribution



The table gives the values of $F(x)$ for $x \geq 0$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983