Chapter 2

1. Give R assignment statements that set the variable z to

   (a) x^{a^b}
   (b) $(x^a)^b$
   (c) $3x^3 + 2x^2 + 6x + 1$ (try to minimise the number of operations required)
   (d) the digit in the second decimal place of x (hint: use `floor(x)` and/or `%%`)
   (e) $z + 1$

2. Give R expressions that return the following matrices and vectors

   (a) $(1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1)$
   (b) $(1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5)$
   (c) $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$
   (d) $\begin{pmatrix} 0 & 2 & 3 \\ 0 & 5 & 0 \\ 7 & 0 & 0 \end{pmatrix}$

3. Suppose `vec` is a vector of length 2. Interpreting `vec` as the coordinates of a point in $\mathbb{R}^2$, use R to express it in polar coordinates. You will need (at least one of) the inverse trigonometric functions: `acos(x)`, `asin(x)`, and `atan(x)`.

4. Use R to produce a vector containing all integers from 1 to 100 that are not divisible by 2, 3, or 7.

5. Suppose that `queue <- c("Steve", "Russell", "Alison", "Liam")` and that `queue` represents a supermarket queue with Steve first in line. Using R expressions update the supermarket queue as successively:

   (a) Barry arrives;
   (b) Steve is served;
   (c) Pam talks her way to the front with one item;
   (d) Barry gets impatient and leaves;
   (e) Alison gets impatient and leaves.

   For the last case you should not assume that you know where in the queue Alison is standing.

   Finally, using the function `which(x)`, find the position of Russell in the queue.

   Note that when assigning a text string to a variable, it needs to be in quotes. We formally introduce text in Section **??**.

6. Which of the following assignments will be successful? What will the vectors `x`, `y`, and `z` look like at each stage?

```
rm(list = ls())
x <- 1
x[3] <- 3
y <- c()
y[2] <- 2
y[3] <- y[1]
y[2] <- y[4]
z[1] <- 0
```

1

Chapter 3

1. Consider the function $y = f(x)$ defined by

x	≤ 0	$\in (0, 1]$	> 1
$f(x)$	$-x^3$	x^2	$\sqrt{x}$

   Supposing that you are given x, write an R expression for y using `if` statements.

   Add your expression for y to the following program, then run it to plot the function f.

   ```
   # input
   x.values <- seq(-2, 2, by = 0.1)

   # for each x calculate y
   n <- length(x.values)
   y.values <- rep(0, n)
   for (i in 1:n) {
       x <- x.values[i]
       # your expression for y goes here
       y.values[i] <- y
   }

   # output
   plot(x.values, y.values, type = "l")
   ```

   Your plot should look like Figure 1. Do you think f has a derivative at 1? What about at 0?

   We remark that it is possible to vectorise the program above, using the `ifelse` function.

2. Let $h(x, n) = 1 + x + x^2 + \cdots + x^n = \sum_{i=0}^{n} x^i$. Write an R program to calculate $h(x, n)$ using a `for` loop.

3. The function $h(x, n)$ from Exercise 2 is the finite sum of a geometric sequence. It has the following explicit formula, for $x \neq 1$,
$$h(x, n) = \frac{1 - x^{n+1}}{1 - x}.$$
   Test your program from Exercise 2 against this formula using the following values

x	n	$h(x, n)$
0.3	55	1.428571
6.6	8	4243335.538178

   You should use the computer to calculate the formula rather than doing it yourself.

4. First write a program that achieves the same result as in Exercise 2 but using a `while` loop. Then write a program that does this using vector operations (and no loops).

   If it doesn't already, make sure your program works for the case $x = 1$.

5. To rotate a vector $(x, y)^T$ anticlockwise by θ radians, you premultiply it by the matrix
$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$
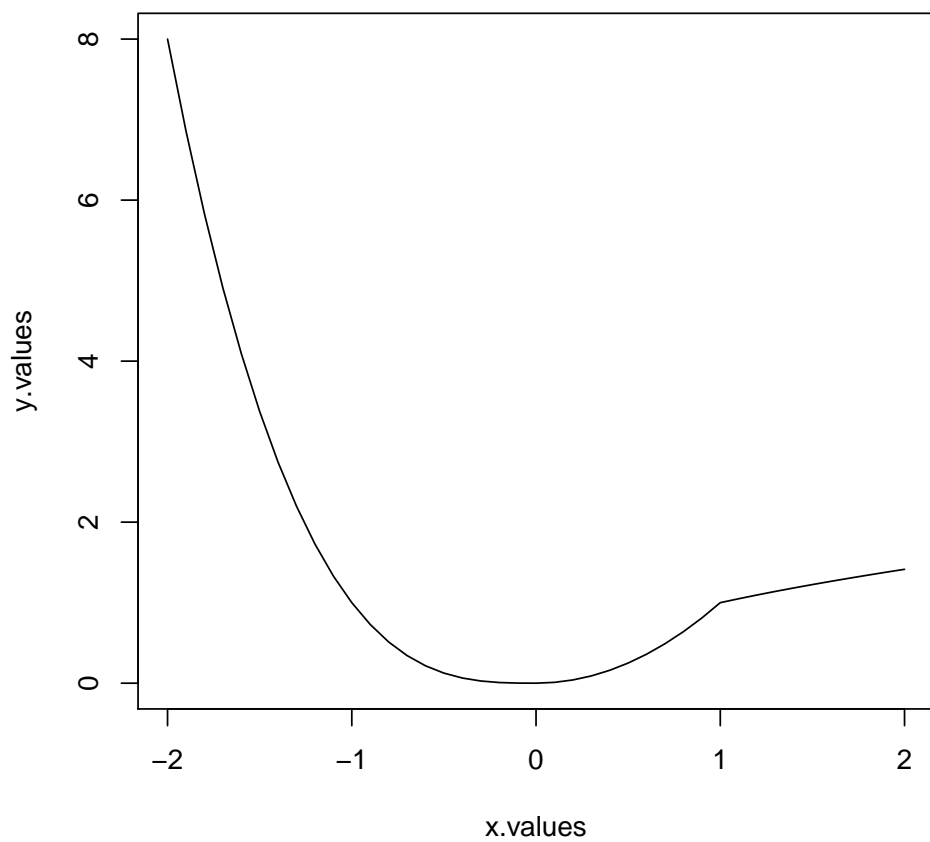
   Write a program in R that does this for you.

Figure 1: The graph produced by Exercise 1.

6. Given a vector `x`, calculate its geometric mean using both a for loop and vector operations. (The geometric mean of $x_1, \ldots, x_n$ is $\left(\prod_{i=1}^{n} x_i\right)^{1/n}$.)

   You might also like to have a go at calculating the harmonic mean, $\left(\sum_{i=1}^{n} 1/x_i\right)^{-1}$, and then check that if the x_i are all positive, the harmonic mean is always less than or equal to the geometric mean, which is always less than or equal to the arithmetic mean.

7. How would you find the sum of every third element of a vector `x`?

8. How does program `quad2.r` (Exercise **??**) behave if `a2` is 0 and/or `a1` is 0? Using `if` statements, modify `quad2.r` so that it gives sensible answers for all possible (numerical) inputs.

9. Chart the flow through the following two programs.

   (a) The first program is a modification of the example from Section **??**, where x is now an array. You will need to keep track of the value of each element of x, namely $x[1]$, $x[2]$, etc.

   ```
   # program spuRs/resources/scripts/threexplus1array.r

   x <- 3
   for (i in 1:3) {
     show(x)
     if (x[i] %% 2 == 0) {
       x[i+1] <- x[i]/2
     } else {
       x[i+1] <- 3*x[i] + 1
     }
   }
   show(x)
   ```

   (b) The second program implements the Lotka-Volterra model for a 'predator-prey' system. We suppose that $x(t)$ is the number of prey animals at the start of a year t (rabbits) and $y(t)$ is the number of predators (foxes), then the Lotka-Volterra model is:

   $$\begin{aligned}
   x(t+1) &= x(t) + b_r \cdot x(t) - d_r \cdot x(t) \cdot y(t); \\
   y(t+1) &= y(t) + b_f \cdot d_r \cdot x(t) \cdot y(t) - d_f \cdot y(t);
   \end{aligned}$$

   where the parameters are defined by:

   b_r is the natural birth rate of rabbits in the absence of predation;

   d_r is the death rate per encounter of rabbits due to predation;

   d_f is the natural death rate of foxes in the absence of food (rabbits);

   b_f is the efficiency of turning predated rabbits into foxes.

   ```
   # program spuRs/resources/scripts/predprey.r
   # Lotka-Volterra predator-prey equations
   br <- 0.04    # growth rate of rabbits
   dr <- 0.0005 # death rate of rabbits due to predation
   df <- 0.2     # death rate of foxes
   bf <- 0.1     # efficiency of turning predated rabbits into foxes
   x <- 4000
   y <- 100
   while (x > 3900) {
     # cat("x =", x, " y =", y, "\n")
     x.new <- (1+br)*x - dr*x*y
     y.new <- (1-df)*y + bf*dr*x*y
     x <- x.new
     y <- y.new
   }
   ```

Note that you do not actually need to know anything about the program to be able to chart its flow.

10. Write a program that uses a loop to find the minimum of a vector x, without using any predefined functions like `min(...)` or `sort(...)`.

    You will need to define a variable, x.min say, in which to keep the smallest value you have yet seen. Start by assigning `x.min <- x[1]` then use a `for` loop to compare x.min with `x[2]`, `x[3]`, etc. If/when you find `x[i] < x.min`, update the value of x.min accordingly.

11. Write a program to merge two sorted vectors into a single sorted vector.

    Do not use the `sort(x)` function, and try to make your program as efficient as possible. That is, try to minimise the number of operations required to merge the vectors.

12. The game of craps is played as follows. First, you roll two six-sided dice; let x be the sum of the dice on the first roll. If $x = 7$ or 11 you win, otherwise you keep rolling until either you get x again, in which case you also win, or until you get a 7 or 11, in which case you lose.

    Write a program to simulate a game of craps. You can use the following snippet of code to simulate the roll of two (fair) dice:

    ```
    x <- sum(ceiling(6*runif(2)))
    ```

13. Suppose that $(x(t), y(t))$ has polar coordinates $(\sqrt{t}, 2\pi t)$. Plot $(x(t), y(t))$ for $t \in [0, 10]$. Your plot should look like Figure 2.
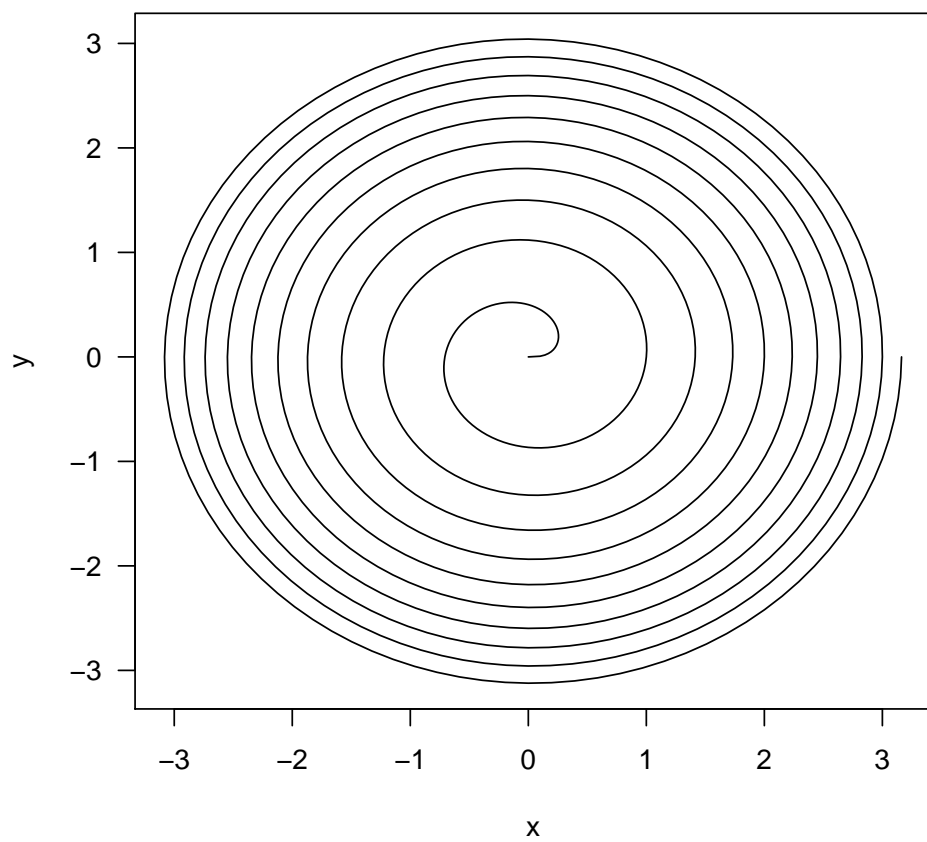
Figure 2: The output from Exercise 13.

Chapter 4

1. Here are the first few lines of the files `age.txt` and `teeth.txt`, taken from the database of a statistically minded dentist:

```
ID      Age
1       18
2       19
3       17
.       .
.       .
.       .


ID      Num Teeth
1       28
2       27
3       32
.       .
.       .
.       .
```

Write a program in R to read each file, and then write an amalgamated list to the file `age_teeth.txt`, of the following form:

```
ID      Age      Num Teeth
1       18       28
2       19       27
3       17       32
.       .        .
.       .        .
.       .        .
```

2. The function `order(x)` returns a permutation of `1:length(x)` giving the order of the elements of `x`. For example

```
> x <- c(1.1, 0.7, 0.8, 1.4)
> (y <- order(x))

[1] 2 3 1 4

> x[y]

[1] 0.7 0.8 1.1 1.4
```

Using `order` or otherwise, modify your program from Exercise 1 so that the output file is ordered by its second column.

3. Devise a program that outputs a table of squares and cubes of the numbers `1` to `n`. For `n <- 7` the output should be as follows:

```
> source("../scripts/square_cube.r")

  number  square    cube

       1       1       1
       2       4       8
       3       9      27
       4      16      64
```

```
     5      25     125
     6      36     216
     7      49     343
```

4. Write an R program that prints out the standard multiplication table:

```
> source("../scripts/mult_table.r")

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    2    3    4    5    6    7    8
[2,]    2    4    6    8   10   12   14   16
[3,]    3    6    9   12   15   18   21   24
[4,]    4    8   12   16   20   24   28   32
[5,]    5   10   15   20   25   30   35   40
[6,]    6   12   18   24   30   36   42   48
[7,]    7   14   21   28   35   42   49   56
[8,]    8   16   24   32   40   48   56   64
[9,]    9   18   27   36   45   54   63   72
      [,9]
[1,]    9
[2,]   18
[3,]   27
[4,]   36
[5,]   45
[6,]   54
[7,]   63
[8,]   72
[9,]   81
```

Hint: generate a matrix mtable that contains the table, then use show(mtable).

5. Use R to plot the hyperbola $x^2 - y^2/3 = 1$, as in Figure 3.

The hyperbola x^2 − y^2/3 = 1

asymptote y = sqrt(3)*x
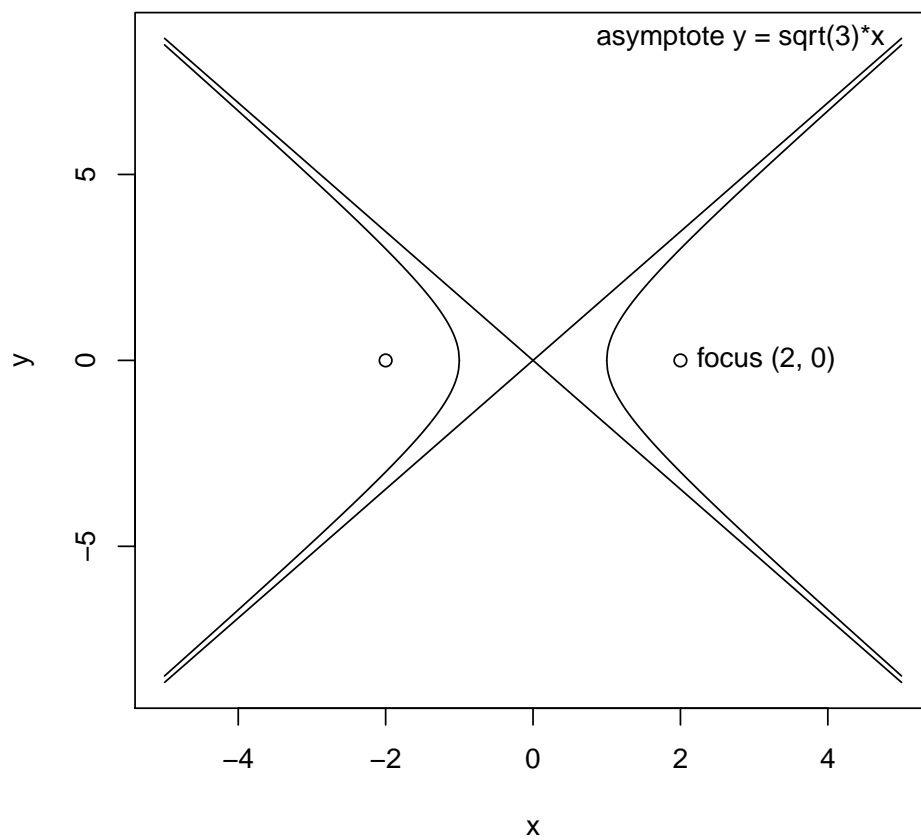
○ focus (2, 0)

y

x

Figure 3: The hyperbola $x^2 - y^2/3 = 1$; see Exercise 5.

Chapter 5

1. The (Euclidean) length of a vector $v = (a_0, \ldots, a_k)$ is the square root of the sum of squares of its coordinates, that is $\sqrt{a_0^2 + \cdots + a_k^2}$. Write a function that returns the length of a vector.

2. In Exercise .2 you wrote a program to calculate $h(x, n)$, the sum of a finite geometric series. Turn this program into a *function* that takes two arguments, x and n, and returns $h(x, n)$.

   Make sure you deal with the case $x = 1$.

3. In this question we simulate the rolling of a die. To do this we use the function `runif(1)`, which returns a 'random' number in the range (0,1). To get a random integer in the range $\{1, 2, 3, 4, 5, 6\}$, we use `ceiling(6*runif(1))`, or if you prefer, `sample(1:6,size=1)` will do the same job.

   (a) Suppose that you are playing the gambling game of the Chevalier de Méré. That is, you are betting that you get at least one six in 4 throws of a die. Write a program that simulates one round of this game and prints out whether you win or lose.

   Check that your program can produce a different result each time you run it.

   (b) Turn the program that you wrote in part (a) into a function `sixes`, which returns `TRUE` if you obtain at least one six in n rolls of a fair die, and returns `FALSE` otherwise. That is, the argument is the number of rolls n, and the value returned is `TRUE` if you get at least one six and `FALSE` otherwise.

   How would you give n the default value of 4?

   (c) Now write a program that uses your function `sixes` from part (b), to simulate N plays of the game (each time you bet that you get at least 1 six in n rolls of a fair die). Your program should then determine the proportion of times you win the bet. This proportion is an estimate of the *probability* of getting at least one 6 in n rolls of a fair die.

   Run the program for $n = 4$ and $N = 100$, 1000, and 10000, conducting several runs for each N value. How does the *variability* of your results depend on N?

   The probability of getting no 6's in n rolls of a fair die is $(5/6)^n$, so the probability of getting at least one is $1 - (5/6)^n$. Modify your program so that it calculates the theoretical probability as well as the simulation estimate and prints the difference between them. How does the *accuracy* of your results depend on N?

   You may find the `replicate` function useful here.

   (d) In part (c), instead of processing the simulated runs as we go, suppose we first store the results of every game in a file, then later postprocess the results.

   Write a program to write the result of all N runs to a textfile `sixes_sim.txt`, with the result of each run on a separate line. For example, the first few lines of the textfile could look like

   ```
   TRUE
   FALSE
   FALSE
   TRUE
   FALSE
   .
   .
   ```

   Now write another program to read the textfile `sixes_sim.txt` and again determine the proportion of bets won.

   This method of saving simulation results to a file is particularly important when each simulation takes a very long time (hours or days), in which case it is good to have a record of your results in case of a system crash.

4. Consider the following program and its output

```
# Program spuRs/resources/scripts/err.r

# clear the workspace
rm(list=ls())

random.sum <- function(n) {
    # sum of n random numbers
    x[1:n] <- ceiling(10*runif(n))
    cat("x:", x[1:n], "\n")
    return(sum(x))
}

x <- rep(100, 10)
show(random.sum(10))
show(random.sum(5))


> source("../scripts/err.r")

x: 9 6 4 1 1 5 4 4 7 9
[1] 50
x: 2 8 9 8 9
[1] 536
```

Explain what is going wrong and how you would fix it.

5. For $r \in [0, 4]$, the *logistic map* of $[0, 1]$ into $[0, 1]$ is defined as $f(x) = rx(1 - x)$.

Given a point $x_1 \in [0, 1]$ the sequence $\{x_n\}_{n=1}^{\infty}$ given by $x_{n+1} = f(x_n)$ is called the *discrete dynamical system* defined by f.

Write a function that takes as parameters x_1, r, and n, generates the first n terms of the discrete dynamical system above, and then plots them.

The logistic map is a simple model for population growth subject to resource constraints: if x_n is the population size at year n, then x_{n+1} is the size at year $n + 1$. Type up your code, then see how the system evolves for different starting values x_1 and different values of r.

Figure 4 gives some typical output.

6. The Game of Life is a cellular automaton and was devised by the mathematician J.H. Conway in 1970. It is played on a grid of cells, each of which is either alive or dead. The grid of cells evolves in time and each cell interacts with its eight neighbours, which are the cells directly adjacent horizontally, vertically, and diagonally.

At each time step cells change as follows:

- A live cell with fewer than two neighbours dies of loneliness.
- A live cell with more than three neighbours dies of overcrowding.
- A live cell with two or three neighbours lives on to the next generation.
- A dead cell with exactly three neighbours comes to life.

The initial pattern constitutes the first generation of the system. The second generation is created by applying the above rules simultaneously to every cell in the first generation: births and deaths all happen simultaneously. The rules continue to be applied repeatedly to create further generations.

Theoretically the Game of Life is played on an infinite grid, but in practice we use a finite grid arranged as a torus. That is, if you are in the left-most column of the grid then your left-hand neighbours are in the right-most column, and if you are in the top row then your neighbours above are in the bottom row.
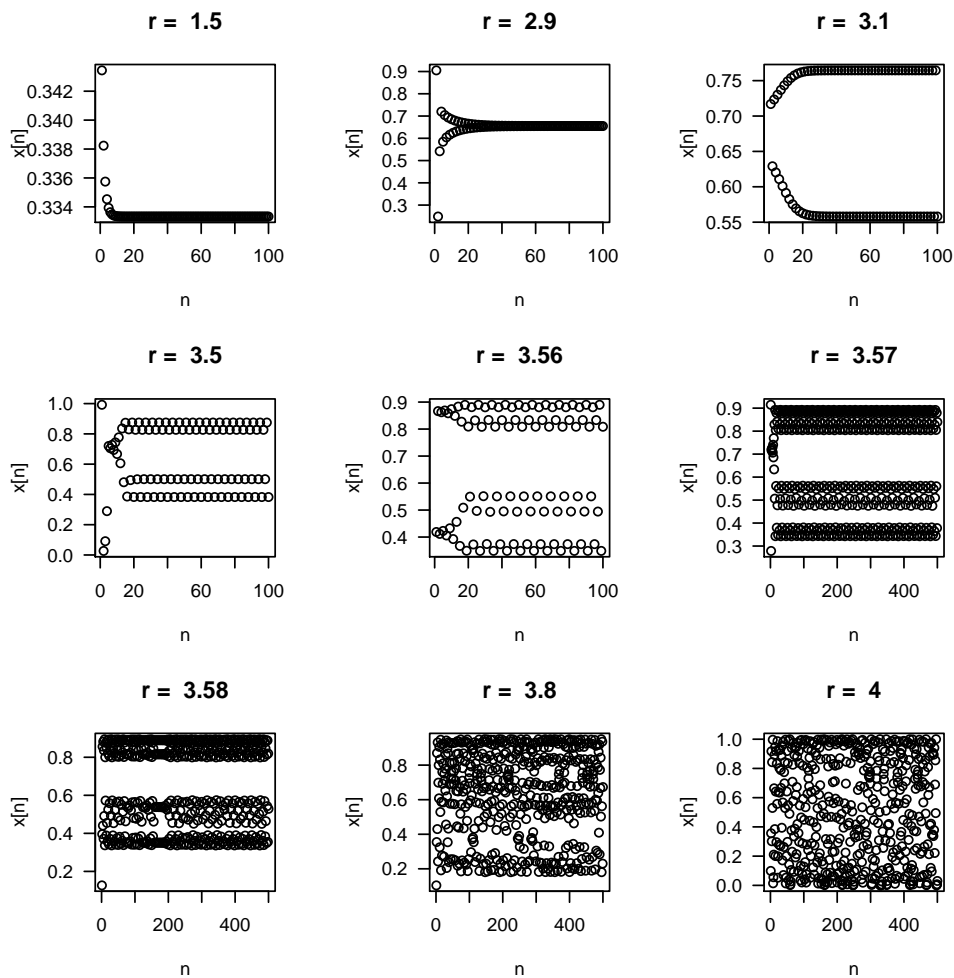
Figure 4: The logistic map described in Exercise 5.

Here is an implementation of the Game of Life in R. The grid of cells is stored in a matrix `A`, where `A[i,j]` is 1 if cell (i, j) is alive and 0 otherwise.

```
# program spuRs/resources/scripts/life.r

neighbours <- function(A, i, j, n) {
    # A is an n*n 0-1 matrix
    # calculate number of neighbours of A[i,j]
    .
    .
    .
}

# grid size
n <- 50

# initialise lattice
A <- matrix(round(runif(n^2)), n, n)

finished <- FALSE
while (!finished) {
    # plot
    plot(c(1,n), c(1,n), type = "n", xlab = "", ylab = "")
    for (i in 1:n) {
        for (j in 1:n) {
            if (A[i,j] == 1) {
                points(i, j)
            }
        }
    }

    # update
    B <- A
    for (i in 1:n) {
        for (j in 1:n) {
            nbrs <- neighbours(A, i, j, n)
            if (A[i,j] == 1) {
                if ((nbrs == 2) | (nbrs == 3)) {
                    B[i,j] <- 1
                } else {
                    B[i,j] <- 0
                }
            } else {
                if (nbrs == 3) {
                    B[i,j] <- 1
                } else {
                    B[i,j] <- 0
                }
            }
        }
    }
    A <- B

    ## continue?
```
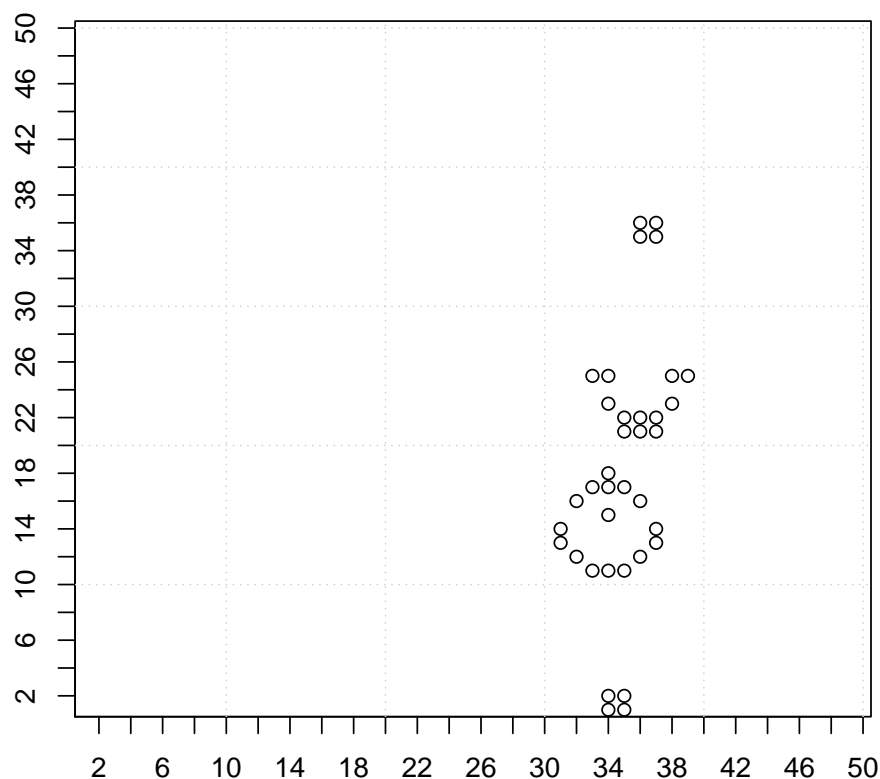
Figure 5: The glider gun, from Exercise 6.

```
    #input <- readline("stop? ")
    #if (input == "y") finished <- TRUE
}
```

Note that this program contains an infinite loop! To stop it you will need to use the escape or stop button (Windows or Mac) or control-C (Unix). Alternatively, uncomment the last two lines. To get the program to run you will need to complete the function `neighbours(A, i, j, n)`, which calculates the number of neighbours of cell (i, j). (The program `forest_fire.r` in Section ?? uses a similar function of the same name, which you may find helpful.)

Once you get the program running, you might like to initialise it using the glider gun, shown in Figure 5 (see `glidergun.r` in the `spuRs` package). Many other interesting patterns have been discovered in the Game of Life.[1]

7. The number of ways you can choose r things from a set of n, ignoring the order in which they are chosen, is $\binom{n}{r} = n!/(r!(n-r)!)$. Let x be the first element of the set of n things. We can partition the collection of possible size r subsets into those that contain x and those that don't: there must be $\binom{n-1}{r-1}$ subsets of the first type and $\binom{n-1}{r}$ subsets of the second type. Thus

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}.$$

---

[1]M. Gardner, *Wheels, Life, and Other Mathematical Amusements.* Freeman, 1985.

Using this and the fact that $\binom{n}{n} = \binom{n}{0} = 1$, write a recursive function to calculate $\binom{n}{r}$.

8. A classic puzzle called the *Towers of Hanoi* uses a stack of rings of different sizes, stacked on one of 3 poles, from the largest on the bottom to the smallest on top (so that no larger ring is on top of a smaller ring). The object is to move the stack of rings from one pole to another by moving one ring at a time so that larger rings are never on top of smaller rings.

   Here is a recursive algorithm to accomplish this task. If there is only one ring, simply move it. To move n rings from the pole `frompole` to the pole `topole`, first move the top $n-1$ rings from `frompole` to the remaining `sparepole`, then move the last and largest from `frompole` to the empty `topole`, then move the $n-1$ rings on `sparepole` to `topole` (on top of the largest).

   The following program implements this algorithm. For example, if there are initially 8 rings, we then move them from pole 1 to pole 3 by calling `moverings(8,1,3)`.

   ```
   # Program spuRs/resources/scripts/moverings.r

   # Tower of Hanoi

   moverings <- function(numrings, frompole, topole) {
     if (numrings == 1) {
       cat("move ring 1 from pole", frompole,
           "to pole", topole, "\n")
     } else {
       sparepole <- 6 - frompole - topole # clever
       moverings(numrings - 1, frompole, sparepole)
       cat("move ring", numrings, "from pole", frompole,
           "to pole", topole, "\n")
       moverings(numrings - 1, sparepole, topole)
     }
     return(invisible(NULL))
   }
   ```

   Check that the algorithm works for the cases `moverings(3, 1, 3)` and `moverings(4, 1, 3)`, then satisfy yourself that you understand why it works.

   Use mathematical induction to show that, using this algorithm, moving a stack of n rings will require exactly $2^n - 1$ individual movements.

Chapter 6

1. From the `spuRs` package you can obtain the dataset `ufc.csv`, with forest inventory observations from the University of Idaho Experimental Forest. Try to answer the following questions:

   (a) What are the species of the three tallest trees? Of the five fattest trees? (Use the `order` command.)

   (b) What are the mean diameters by species?

   (c) What are the two species that have the largest third quartile diameters?

   (d) What are the two species with the largest median slenderness (height/diameter) ratios? How about the two species with the smallest median slenderness ratios?

   (e) What is the identity of the tallest tree of the species that was the fattest on average?

2. Create a list in R containing the following information:

   - your full name,
   - gender,
   - age,
   - a list of your 3 favourite movies,
   - the answer to the question 'Do you support the United Nations?', and
   - a list of the birth day and month of your immediate family members including you (identified by first name).

   Do the same for three close friends, then write a program to check if there are any shared birthdays or names in the four lists.

   Produce a table of birthdays by birth month and a table of the mean number of immediate family members by gender.

3. Using the tree growth data (Section **??**, available from the `spuRs` package), plot tree age versus height for each tree, broken down by habitat type. That is, create a grid of 5 plots, each showing the trees from a single habitat.

   For the curious, the habitats corresponding to codes 1 through 5 are: Ts/Pach, Ts/Op, Th/Pach, AG/Pach, and PA/Pach. These codes refer respectively to the climax tree species, which is the most shade-tolerant species that can grow on the site, and the dominant understorey plant. Ts refers to *Thuja plicata* and *Tsuga heterophylla*, Th refers to just *Thuja plicata*, AG is *Abies grandis*, PA is *Picea engelmanii* and *Abies lasiocarpa*, Pach is *Pachistima myrsinites*, and Op is the nasty *Oplopanaz horridurn*. *Abies grandis* is considered a major climax species for AG/Pach, a major seral species for Th/Pach and PA/Pach, and a minor seral species for Ts/Pach and Ts/Op. Loosely speaking, a community is *seral* if there is evidence that at least some of the species are temporary, and *climax* if the community is self-regenerating.[2]

4. Pascal's triangle.

   Suppose we represent Pascal's triangle as a list, where item n is row n of the triangle. For example, Pascal's triangle to depth four would be given by

   ```
   list(c(1), c(1, 1), c(1, 2, 1), c(1, 3, 3, 1))
   ```

   The n-th row can be obtained from row $n-1$ by adding all adjacent pairs of numbers, then prefixing and suffixing a 1.

   Write a function that, given Pascal's triangle to depth n, returns Pascal's triangle to depth $n+1$. Verify that the eleventh row gives the binomial coefficients $\binom{10}{i}$ for $i = 0, 1, \ldots, 10$.

---

[2]R. Daubenmire, 1952. Forest Vegetation of Northern Idaho and Adjacent Washington, and Its Bearing on Concepts of Vegetation Classification, *Ecological Monographs* **22**, 301–330.

5. Horse racing.

   The following is an excerpt from the file `racing.txt` (available in the `spuRs` archive), which has details of nine horse races run in the U.K. in July 1998.

```
1 0 54044 4.5    53481 4     53526 4     53526 3.5   53635 3     53792
1 1 54044 1.375 53481 1.5    53635 1.5   53635 1.375 53928 1.25 54026
1 0 54044 1.75   53481 1.625 53792 1.625 53792 1.75  53936
1 0 54044 14     53481 20    53635 20    53635 16    53868 20    54026
1 0 54044 20     53481 25    53635 25    53635
1 0 54044 33     53481 50    53635 50    53635 66    53929
1 0 54044 20     53481 25    53635 25    53635 33    53792 50    54045
2 1 55854 6      55709 7     56157 7     56157
2 0 55854 6      55138 6.5   55397 6.5   55397 7     55825 7     56157
...
```

   In each row, the first number gives the race number. There is one line for each horse in each race. The next number is 0 or 1 depending on whether the horse lost or won the race. Numbers then come in pairs (t_i, p_i), $i = 1, 2, \ldots$, where t_i is a time and p_i a price. That is, the odds on the horse at time t_i were $p_i : 1$.

   Import this data into an object with the following structure:

   - A list with one element per race.

   - Each race is a list with one element per horse.

   - Each horse is a list with three elements: a logical variable indicating win/loss, a vector of times, and a vector of prices.

   Write a function that, given a single race, plots log price against time for each horse, on the same graph. Highlight the winning horse in a different colour.

6. Here is a recursive program that prints all the possible ways that an amount x (in cents) can be made up using Australian coins (which come in 5, 10, 20, 50, 100, and 200 cent denominations). To avoid repetition, each possible decomposition is ordered.

```
# Program spuRs/resources/scripts/change.r

change <- function(x, y.vec = c()) {
  # finds possible ways of making up amount x using Australian coins
  # x is given in cents and we assume it is divisible by 5
  # y.vec are coins already used (so total amount is x + sum(y.vec))
  if (x == 0) {
    cat(y.vec, "\n")
  } else {
    coins <- c(200, 100, 50, 20, 10, 5)
    new.x <- x - coins
    new.x <- new.x[new.x >= 0]
    for (z in new.x) {
      y.tmp <- c(y.vec, x - z)
      if (identical(y.tmp, sort(y.tmp))) {
        change(z, y.tmp)
      }
    }
  }
  return(invisible(NULL))
}
```

Rewrite this program so that instead of writing its output to the screen it returns it as a list, where each element is a vector giving a possible decomposition of x.

Chapter 7

1. The *slenderness* of a tree is defined as the ratio of its height over its diameter, both in metres.[3] Slenderness is a useful metric of a tree's growing history, and indicates the susceptibility of the tree to being knocked over in high winds. Although every case must be considered on its own merits, assume that a slenderness of 100 or above indicates a high-risk tree for these data. Construct a boxplot, a histogram, and a density plot of the slenderness ratios by species for the Upper Flat Creek data. Briefly discuss the advantages and disadvantages of each graphic.

2. Among the data accompanying this book, there is another inventory dataset, called `ehc.csv`. This dataset is also from the University of Idaho Experimental Forest, but covers the East Hatter Creek stand. The inventory was again a systematic grid of plots. Your challenge is to produce and interpret Figure **??** using the East Hatter Creek data.

3. Regression to the mean.

   Consider the following very simple genetic model. A population consists of equal numbers of two sexes: male and female. At each generation men and women are paired at random, and each pair produces exactly two offspring, one male and one female. We are interested in the distribution of height from one generation to the next. Suppose that the height of both children is just the average of the height of their parents, how will the distribution of height change across generations?

   Represent the heights of the current generation as a dataframe with two variables, `m` and `f`, for the two sexes. The command `rnorm(100, 160, 20)` will generate a vector of length 100, according to the normal distribution with mean 160 and standard deviation 20 (see Section **??**). We use it to randomly generate the population at generation 1:

   ```
   pop <- data.frame(m = rnorm(100, 160, 20), f = rnorm(100, 160, 20))
   ```

   The command `sample(x, size = length(x))` will return a random sample of size `size` taken from the vector `x` (without replacement). (It will also sample with replacement, if the optional argument `replace` is set to `TRUE`.) The following function takes the dataframe `pop` and randomly permutes the ordering of the men. Men and women are then paired according to rows, and heights for the next generation are calculated by taking the mean of each row. The function returns a dataframe with the same structure, giving the heights of the next generation.

   ```
   next.gen <- function(pop) {
     pop$m <- sample(pop$m)
     pop$m <- apply(pop, 1, mean)
     pop$f <- pop$m
     return(pop)
   }
   ```

   Use the function `next.gen` to generate nine generations, then use the lattice function `histogram` to plot the distribution of male heights in each generation, as in Figure 6. The phenomenon you see is called *regression to the mean*.

   Hint: construct a dataframe with variables height and generation, where each row represents a single man.

4. Reproduce Figure **??** using `lattice` graphics.

---

[3]For example, Y. Wang, S. Titus, and V. LeMay. 1998. Relationships between tree slenderness coefficients and tree or stand characteristics for major species in boreal mixedwood forests. *Canadian Journal of Forest Research* **28**, 1171–1183.
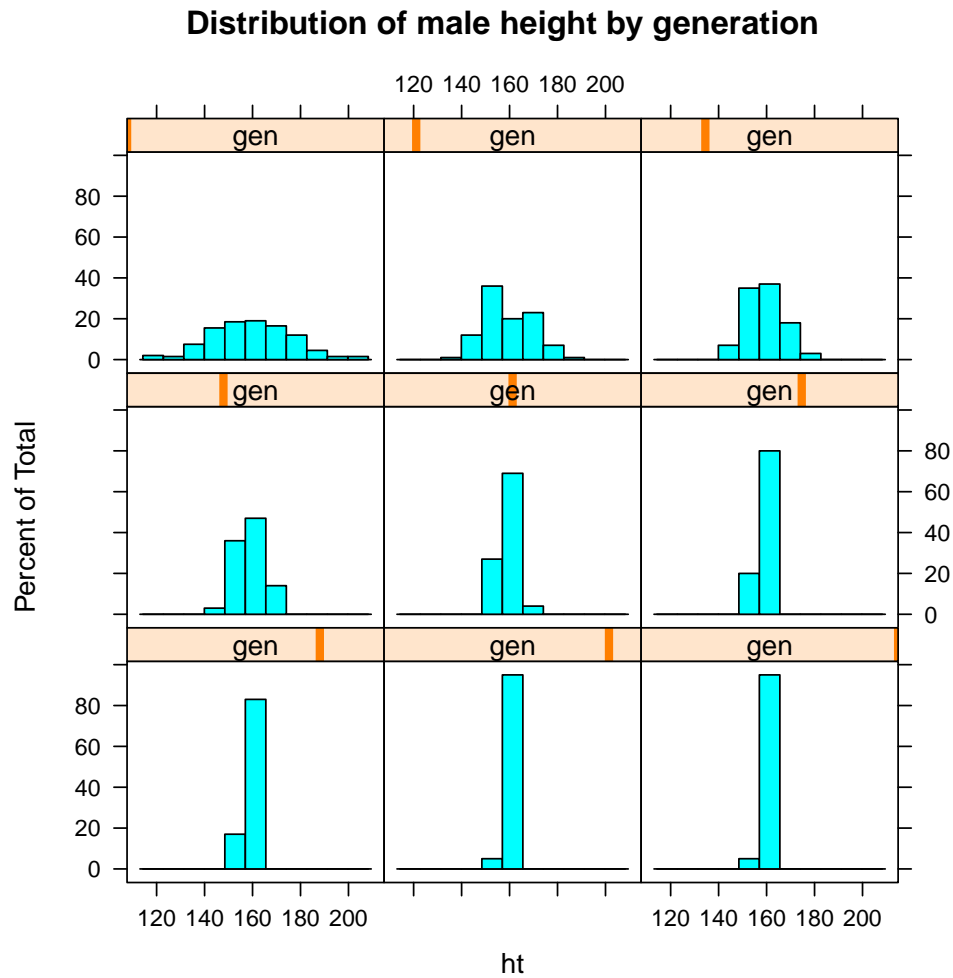
Distribution of male height by generation

Figure 6: Regression to the mean: male heights across nine generations. See Exercise 3.

Chapter 8

1. Student records.

   Create an S3 class `studentRecord` for objects that are a list with the named elements 'name', 'subjects completed', 'grades', and 'credit'.

   Write a `studentRecord` method for the generic function `mean`, which returns a weighted GPA, with subjects weighted by credit. Also write a `studentRecord` method for `print`, which employs some nice formatting, perhaps arranging subjects by year code.

   Finally create a further class for a cohort of students, and write methods for `mean` and `print` which, when applied to a cohort, apply `mean` or `print` to each student in the cohort.

2. Let `Omega` be an ordered vector of numbers and define a subset of `Omega` to be an ordered subvector. Define a class `set` for subsets of `Omega` and write functions that perform union, intersection, and complementation on objects of class `set`.

   Do not use R's built-in functions `union`, `intersect`, `setdiff`, or `setequal`.

3. Continued fractions.

   A continued fraction has the form

   $$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}.$$

   The representation is finite if all the a_k are zero for $k \geq k_0$.

   Write a class for continued fractions (with a finite representation). Now write functions to convert numbers from decimal to continued fraction form, and back.

Chapter 9

1. In single precision four bytes are used to represent a floating point number: 1 bit is used for the sign, 8 for the exponent, and 23 for the mantissa.

   What are the largest and smallest non-zero positive numbers in single precision (including denormalised numbers)?

   In base 10, how many significant digits do you get using single precision?

2. What is the relative error in approximating π by 22/7? What about 355/113?

3. Suppose x and y can be represented without error in double precision. Can the same be said for x^2 and y^2?

   Which would be more accurate, $x^2 - y^2$ or $(x - y)(x + y)$?

4. To calculate $\log(x)$ we use the expansion

   $$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots.$$

   Truncating to n terms, the error is no greater in magnitude than the last term in the sum. How many terms in the expansion are required to calculate $\log 1.5$ with an error of at most 10^{-16}? How many terms are required to calculate $\log 2$ with an error of at most 10^{-16}?

   Using the fact that $\log 2 = 2 \log \sqrt{2}$, suggest a better way of calculating $\log 2$.

5. The sample variance of a set of observations $x_1, \ldots, x_n$ is given by $S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n - 1) = (\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)/(n - 1)$, where $\bar{x} = \sum_{i=1}^{n} x_i/n$ is the sample mean.

   Show that the second formula is more efficient (requires fewer operations) but can suffer from catastrophic cancellation. Demonstrate catastrophic cancellation with an example sample of size $n = 2$.

6. Horner's algorithm for evaluating the polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ consists of re-expressing it as
   $$a_0 + x(a_1 + x(a_2 + \cdots + x(a_{n-1} + x a_n) \cdots)).$$

   How many operations are required to evaluate $p(x)$ in each form?

7. This exercise is based on the problem of sorting a list of numbers, which is one of the classic computing problems. Note that R has an excellent sorting function, `sort(x)`, which we will not be using.

   To judge the effectiveness of a sorting algorithm, we count the number of *comparisons* that are required to sort a vector `x` of length `n`. That is, we count the number of times we evaluate logical expressions of the form `x[i] < x[j]`. The fewer comparisons required, the more efficient the algorithm.

   Selection sort The simplest but least efficient sorting algorithm is selection sort. The selection sort algorithm uses two vectors, an unsorted vector and a sorted vector, where all the elements of the sorted vector are less than or equal to the elements of the unsorted vector. The algorithm proceeds as follows:

   1 Given a vector x, let the initial unsorted vector u be equal to x, and the initial sorted vector s be a vector of length 0.

   2 Find the smallest element of u then remove it from u and add it to the end of s.

   3 If u is not empty then go back to step 2.

   Write an implementation of the selection sort algorithm. To do this you may find it convenient to write a function that returns the *index* of the smallest element of a vector.

   How many comparisons are required to sort a vector of length n using the selection sort algorithm?

   Insertion sort Like selection sort, insertion sort uses an unsorted vector and a sorted vector, moving elements from the unsorted to the sorted vector one at a time. The algorithm is as follows:

1 Given a vector x, let the initial unsorted vector u be equal to x, and the initial sorted vector s be a vector of length 0.

2 Remove the last element of u and insert it into s so that s is still sorted.

3 If u is not empty then go back to step 2.

Write an implementation of the insertion sort algorithm. To insert an element a into a sorted vector $s = (b_1, \ldots, b_k)$ (as per step 2 above), you do not usually have to look at every element of the vector. Instead, if you start searching from the end, you just need to find the first i such that $a \geq b_i$, then the new sorted vector is $(b_1, \ldots, b_i, a, b_{i+1}, \ldots, b_k)$.

Because inserting an element into a sorted vector is usually quicker than finding the minimum of a vector, insertion sort is usually quicker than selection sort, but the actual number of comparisons required depends on the initial vector x. What are the worst and best types of vector x, with respect to sorting using insertion sort, and how many comparisons are required in each case?

Bubble sort Bubble sort is quite different from selection sort and insertion sort. It works by repeatedly comparing adjacent elements of the vector $x = (a_1, \ldots, a_n)$, as follows:

1 For $i = 1, \ldots, n - 1$, if $a_i > a_{i+1}$ then swap a_i and a_{i+1}.

2 If you did any swaps in step 1, then go back and do step 1 again.

Write an implementation of the bubble sort algorithm and work out the minimum and maximum number of comparisons required to sort a vector of length n.

Bubble sort is not often used in practice. Its main claim to fame is that it does not require an extra vector to store the sorted values. There was a time when the available memory was an important programming consideration, and so people worried about how much storage an algorithm required, and bubble sort is excellent in this regard. However at present computing speed is more of a bottleneck than memory, so people worry more about how many operations an algorithm requires.

If you like bubble sort then you should look up the related algorithm *gnome sort*, which was named after the garden gnomes of Holland and their habit of rearranging flower pots.

Quick sort The quick sort algorithm is (on average) one of the fastest sorting algorithms currently available and is widely used. It was first described by C.A.R. Hoare in 1960. Quick sort uses a 'divide-and-conquer' strategy: it is a recursive algorithm that divides the problem into two smaller (and thus easier) problems. Given a vector $x = (a_1, \ldots, a_n)$, the algorithm works as follows:

1 If $n = 0$ or 1 then x is sorted so stop.

2 If $n > 1$ then split $(a_2, \ldots, a_n)$ into two subvectors, l and g, where l consists of all the elements of x less than a_1, and g consists of all the elements of x greater than a_1 (ties can go in either l or g).

3 Sort l and g. Call the sorted subvectors $(b_1, \ldots, b_i)$ and $(c_1, \ldots, c_j)$, respectively, then the sorted vector x is given by $(b_1, \ldots, b_i, a_1, c_1, \ldots, c_j)$.

Implement the quick sort algorithm using a recursive function.

It can be shown that on average the quick sort algorithm requires $O(n \log n)$ comparisons to sort a vector of length n, though its worst-case performance is $O(n^2)$. Also, it is possible to implement quick sort so that it uses memory efficiently while remaining quick.

Two other sorting algorithms that also require on average $O(n \log n)$ comparisons are *heap sort* and *merge sort*.

8. Use the `system.time` function to compare the programs `primedensity.r` and `primesieve.r`, given in Chapter **??**.

9. For $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_n)^T$, the *convolution* of $\mathbf{x}$ and $\mathbf{y}$ is the vector $\mathbf{z} = (z_1, \ldots, z_{2n})^T$ given by

$$z_k = \sum_{i=\max\{1, k-n\}}^{\min\{k, n\}} x_i \cdot y_{k-i}.$$

Write two programs to convolve a pair of vectors, one using loops and the other using vector operations, then use `system.time` to compare their speed.

10. Use the `system.time` function to compare the relative time that it takes to perform addition, multiplication, powers, and other simple operations. You may wish to perform each more than once!

Chapter 10

1. Draw a function $g(x)$ for which the fixed-point algorithm produces the oscillating sequence $1, 7, 1, 7, \ldots$ when started with $x_0 = 7$.

2. (a) Suppose that $x_0 = 1$ and that for $n \geq 0$

$$x_{n+1} = \alpha x_n.$$

   Find a formula for x_n. For which values of α does x_n converge, and to what?

   (b) Consider the fixed-point algorithm for finding x such that $g(x) = x$:

$$x_{n+1} = g(x_n).$$

   Let c be the fixed point, so $g(c) = c$. The first-order Taylor approximation of g about the point c is

$$g(x) \approx g(c) + (x - c)g'(c).$$

   Apply this Taylor approximation to the fixed-point algorithm to give a recurrence relation for $x_n - c$.

   What condition on the function g at the point c will result in the convergence of x_n to c?

3. Use `fixedpoint` to find the fixed point of $\cos x$. Start with $x_0 = 0$ as your initial guess (the answer is 0.73908513).

   Now use `newtonraphson` to find the root of $\cos x - x$, starting with $x_0 = 0$ as your initial guess. Is it faster than the fixed-point method?

4. A picture is worth a thousand words.

   The function `fixedpoint_show.r` below is a modification of `fixedpoint` that plots intermediate results. Instead of using the variables `tol` and `max.iter` to determine when the algorithm stops, at each step you will be prompted to enter `"y"` at the keyboard if you want to continue. There are also two new inputs, `xmin` and `xmax`, which are used to determine the range of the plot. `xmin` and `xmax` have defaults `x0 - 1` and `x0 + 1`, respectively.

```
# program spuRs/resources/scripts/fixedpoint_show.r
# loadable spuRs function

fixedpoint_show <- function(ftn, x0, xmin = x0-1, xmax = x0+1) {
  # applies fixed-point method to find x such that ftn(x) == x
  # x0 is the starting point
  # subsequent iterations are plotted in the range [xmin, xmax]

  # plot the function
  x <- seq(xmin, xmax, (xmax - xmin)/200)
  fx <- sapply(x, ftn)
  plot(x, fx, type = "l", xlab = "x", ylab = "f(x)",
    main = "fixed point f(x) = x", col = "blue", lwd = 2)
  lines(c(xmin, xmax), c(xmin, xmax), col = "blue")

  # do first iteration
  xold <- x0
  xnew <- ftn(xold)
  lines(c(xold, xold, xnew), c(xold, xnew, xnew), col = "red")
  lines(c(xnew, xnew), c(xnew, 0), lty = 2, col = "red")

  # continue iterating while user types "y"
```

```
    cat("last x value", xnew, " ")
    continue <- readline("continue (y or n)? ") == "y"
    while (continue) {
      xold <- xnew;
      xnew <- ftn(xold);
      lines(c(xold, xold, xnew), c(xold, xnew, xnew), col = "red")
      lines(c(xnew, xnew), c(xnew, 0), lty = 2, col = "red")
      cat("last x value", xnew, " ")
      continue <- readline("continue (y or n)? ") == "y"
    }

    return(xnew)
  }
```

Use `fixedpoint_show` to investigate the fixed points of the following functions:

(a) $\cos(x)$ using $x_0 = 1, 3, 6$

(b) $\exp(\exp(-x))$ using $x_0 = 2$

(c) $x - \log(x) + \exp(-x)$ using $x_0 = 2$

(d) $x + \log(x) - \exp(-x)$ using $x_0 = 2$

5. Below is a modification of `newtonraphson` that plots intermediate results, analogous to `fixedpoint_show` above. Use it to investigate the roots of the following functions:

(a) $\cos(x) - x$ using $x_0 = 1, 3, 6$

(b) $\log(x) - \exp(-x)$ using $x_0 = 2$

(c) $x^3 - x - 3$ using $x_0 = 0$

(d) $x^3 - 7x^2 + 14x - 8$ using $x_0 = 1.1, 1.2, \ldots, 1.9$

(e) $\log(x)\exp(-x)$ using $x_0 = 2$.

```
# program spuRs/resources/scripts/newtonraphson_show.r
# loadable spuRs function

newtonraphson_show <- function(ftn, x0, xmin = x0-1, xmax = x0+1) {
  # applies Newton-Raphson to find x such that ftn(x)[1] == 0
  # x0 is the starting point
  # subsequent iterations are plotted in the range [xmin, xmax]

  # plot the function
  x <- seq(xmin, xmax, (xmax - xmin)/200)
  fx <- c()
  for (i in 1:length(x)) {
    fx[i] <- ftn(x[i])[1]
  }
  plot(x, fx, type = "l", xlab = "x", ylab = "f(x)",
    main = "zero f(x) = 0", col = "blue", lwd = 2)
  lines(c(xmin, xmax), c(0, 0), col = "blue")

  # do first iteration
  xold <- x0
  f.xold <- ftn(xold)
  xnew <- xold - f.xold[1]/f.xold[2]
  lines(c(xold, xold, xnew), c(0, f.xold[1], 0), col = "red")
```

```
   # continue iterating while user types "y"
   cat("last x value", xnew, " ")
   continue <- readline("continue (y or n)? ") == "y"
   while (continue) {
     xold <- xnew;
     f.xold <- ftn(xold)
     xnew <- xold - f.xold[1]/f.xold[2]
     lines(c(xold, xold, xnew), c(0, f.xold[1], 0), col = "red")
     cat("last x value", xnew, " ")
     continue <- readline("continue (y or n)? ") == "y"
   }

   return(xnew)
}
```

6. Write a program, using both `newtonraphson.r` and `fixedpoint.r` for guidance, to implement the secant root-finding method:
$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}.$$

First test your program by finding the root of the function $\cos x - x$. Next see how the secant method performs in finding the root of $\log x - \exp(-x)$ using $x_0 = 1$ and $x_1 = 2$. Compare its performance with that of the other two methods.

Write a function `secant_show.r` that plots the sequence of iterates produced by the secant algorithm.

7. Adaptive fixed-point iteration.

To find a root a of f we can apply the fixed-point method to $g(x) = x + cf(x)$, where c is some non-zero constant. That is, given x_0 we put $x_{n+1} = g(x_n) = x_n + cf(x_n)$.

From Taylor's theorem we have

$$\begin{aligned} g(x) &\approx g(a) + (x - a)g'(a) \\ &= a + (x - a)(1 + cf'(a)) \end{aligned}$$

so

$$g(x) - a \approx (x - a)(1 + cf'(a)).$$

Based on this approximation, explain why $-1/f'(a)$ would be a good choice for c.

In practice we don't know a so cannot find $-1/f'(a)$. At step n of the iteration, what would be your best guess at $-1/f'(a)$? Using this guess for c, what happens to the fixed-point method? (You can allow your guess to change at each step.)

8. The iterative method for finding the fixed point of a function works in very general situations. Suppose $A \subset \mathbb{R}^d$ and $f : A \to A$ is such that for some $0 \le c < 1$ and all vectors $\mathbf{x}, \mathbf{y} \in A$,

$$\|f(\mathbf{x}) - f(\mathbf{y})\|_d \le c\|\mathbf{x} - \mathbf{y}\|_d.$$

It can be shown that for such an f there is a unique point $\mathbf{x}_* \in A$ such that $f(\mathbf{x}_*) = \mathbf{x}_*$. Moreover for any $\mathbf{x}_0 \in A$, the sequence defined by $\mathbf{x}_{n+1} = f(\mathbf{x}_n)$ converges to $\mathbf{x}_*$. Such a function is called a *contraction mapping*, and this result is called the *contraction mapping theorem*, which is one of the fundamental results in the field of *functional analysis*.

Modify the function `fixedpoint(ftn, x0, tol, max.iter)` given in Section ??, so that it works for any function `ftn(x)` that takes as input a numerical vector of length $d \ge 1$ and returns a numerical vector of length d. Use your modified function to find the fixed points of the function f below, in the region $[0, 2] \times [0, 2]$.

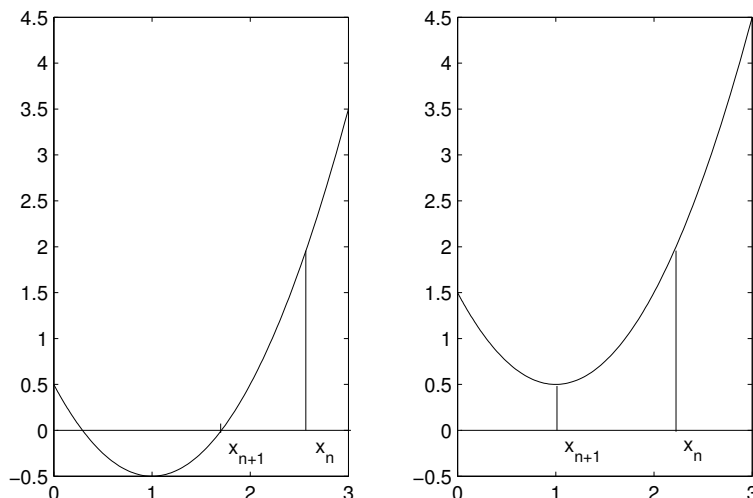$$f(x_1, x_2) = (\log(1 + x_1 + x_2), \log(5 - x_1 - x_2)).$$

Figure 7: The iterative root-finding scheme of Exercise 9

9. For $f : \mathbb{R} \to \mathbb{R}$, the Newton-Raphson algorithm uses a sequence of linear approximations to f to find a root. What happens if we use quadratic approximations instead?

Suppose that x_n is our current approximation to f, then a quadratic approximation to f at x_n is given by the second-order Taylor expansion:

$$f(x) \approx g_n(x) = f(x_n) + (x - x_n)f'(x_n) + \tfrac{1}{2}(x - x_n)^2 f''(x_n).$$

Let x_{n+1} be the solution of $g_n(x) = 0$ that is closest to x_n, assuming a solution exists. If $g_n(x) = 0$ has no solution, then let x_{n+1} be the point at which g_n attains either its minimum or maximum. Figure 7 illustrates the two cases.

Implement this algorithm in R and use it to find the fixed points of the following functions:

(a) $\cos(x) - x$ using $x_0 = 1, 3, 6$.

(b) $\log(x) - \exp(-x)$ using $x_0 = 2$.

(c) $x^3 - x - 3$ using $x_0 = 0$.

(d) $x^3 - 7x^2 + 14x - 8$ using $x_0 = 1.1, 1.2, \ldots, 1.9$.

(e) $\log(x)\exp(-x)$ using $x_0 = 2$.

For your implementation, assume that you are given a function `ftn(x)` that returns the vector $(f(x), f'(x), f''(x))$. Given x_n, if you rewrite g_n as $g_n(x) = a_2 x^2 + a_1 x + a_0$ then you can use the formula $(-a_1 \pm \sqrt{a_1^2 - 4a_2 a_0})/2a_2$ to find the roots of g_n and thus x_{n+1}. If g_n has no roots then the min/max occurs at the point $g_n'(x) = 0$.

How does this algorithm compare to the Newton-Raphson algorithm?

10. How do we know $\pi = 3.1415926$ (to 7 decimal places)? One way of finding π is to solve $\sin(x) = 0$. By definition the solutions to $\sin(x) = 0$ are $k\pi$ for $k = 0, \pm 1, \pm 2, \ldots$, so the root closest to 3 should be π.

(a) Use a root-finding algorithm, such as the Newton-Raphson algorithm, to find the root of $\sin(x)$ near 3. How close can you get to π? (You may use the function `sin(x)` provided by R.)

The function $\sin(x)$ is *transcendental*, which means that it cannot be written as a rational function of x. Instead we have to write it as an infinite sum:

$$\sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

(This is the infinite order Taylor expansion of $\sin(x)$ about 0.) In practice, to calculate $\sin(x)$ numerically we have to truncate this sum, so any numerical calculation of $\sin(x)$ is an approximation. In particular the function `sin(x)` provided by R is only an approximation of $\sin(x)$ (though a very good one).

(b) Put
$$f_n(x) = \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!}.$$

Write a function in R to calculate $f_n(x)$. Plot $f_n(x)$ over the range $[0, 7]$ for a number of values of n, and verify that it looks like $\sin(x)$ for large n.

(c) Choose a large value of n, then find an approximation to π by solving $f_n(x) = 0$ near 3. Can you get an approximation that is correct up to 6 decimal places? Can you think of a better way of calculating π?

11. The astronomer Edmund Halley devised a root-finding method faster than the Newton-Raphson method, but which requires second derivative information. If x_n is our current solution then

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n) - (f(x_n)f''(x_n)/2f'(x_n))}.$$

Let m be a positive integer. Show that applying Halley's method to the function $f(x) = x^m - k$ gives

$$x_{n+1} = \left(\frac{(m-1)x_n^m + (m+1)k}{(m+1)x_n^m + (m-1)k} \right) x_n.$$

Use this to show that, to 9 decimal places, $59^{1/7} = 1.790518691$.

12. The bisection method can be generalised to deal with the case $f(x_l)f(x_r) > 0$, by *broadening* the bracket. That is, we reduce x_l and/or increase x_r, and try again. A reasonable choice for broadening the bracket is to double the width of the interval $[x_l, x_r]$, that is (in pseudo-code)

$$
\begin{aligned}
m &\leftarrow (x_l + x_r)/2 \\
w &\leftarrow x_r - x_l \\
x_l &\leftarrow m - w \\
x_r &\leftarrow m + w
\end{aligned}
$$

Incorporate bracket broadening into the function `bisection` given in Section **??**. Note that broadening is *not guaranteed* to find x_l and x_r such that $f(x_l)f(x_r) \le 0$, so you should include a limit on the number of times it can be tried.

Use your modified function to find a root of

$$f(x) = (x-1)^3 - 2x^2 + 10 - \sin(x),$$

starting with $x_l = 1$ and $x_r = 2$.

Chapter 11

1. Let p be the quadratic $p(x) = c_0 + c_1 x + c_2 x^2$. Simpson's rule uses a quadratic to approximate a given function f over two adjacent intervals, then uses the integral of the quadratic to approximate the integral of the function.

   (a) Show that
   $$\int_{-h}^{h} p(x)\, dx = 2hc_0 + \frac{2}{3}c_2 h^3;$$

   (b) Write down three equations that constrain the quadratic to pass through the points $(-h, f(-h))$, $(0, f(0))$, and $(h, f(h))$, then solve them for c_0 and c_2;

   (c) Show that
   $$\int_{-h}^{h} p(x)\, dx = \frac{h}{3}(f(-h) + 4f(0) + f(h)).$$

2. Suppose $f : [0, 2\pi] \to [0, \infty)$ is continuous and $f(0) = f(2\pi)$. For $(x, y) \in \mathbb{R}^2$ let (R, θ) be the polar coordinates of (x, y), so $x = R\cos\theta$ and $y = R\sin\theta$. Define the set $A \subset \mathbb{R}^2$ by

   $$(x, y) \in A \text{ if } R \le f(\theta).$$

   We consider the problem of calculating the area of A.

   We approximate the area of A using triangles. For small ϵ, the area of the triangle with vertices $(0, 0)$, $(f(\theta)\cos\theta, f(\theta)\sin\theta)$ and $(f(\theta+\epsilon)\cos(\theta+\epsilon), f(\theta+\epsilon)\sin(\theta+\epsilon))$ is $\sin(\epsilon)f(\theta)f(\theta+\epsilon)/2 \approx \epsilon f(\theta)f(\theta+\epsilon)$ (since $\sin(x) \approx x$ near 0). Thus the area of A is approximately

   $$\sum_{k=0}^{n-1} \sin(2\pi/n)f(2\pi k/n)f(2\pi(k+1)/n)/2$$

   $$\approx \sum_{k=0}^{n-1} \pi f(2\pi k/n)f(2\pi(k+1)/n)/n. \tag{1}$$

   See, for example, Figure 8.

   Write a program to numerically calculate this polar-integral using the summation formula (1).

   Check numerically (or otherwise) that as $n \to \infty$ the polar-integral (1) converges to $\frac{1}{2}\int_0^{2\pi} f^2(x)\, dx$. Use $f_1(x) = 2$ and $f_2(x) = 4\pi^2 - (x - 2\pi)^2$ as test cases.

3. The standard normal distribution function is given by
   $$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

   For $p \in [0, 1]$, the $100p$ standard normal percentage point is defined as that z_p for which

   $$\Phi(z_p) = p.$$

   Using the function `Phi(z)` from Example **??**, calculate z_p for $p = 0.5$, 0.95, 0.975, and 0.99.

   Hint: express the problem as a root-finding problem.

4. Consider
   $$I = \int_0^1 \frac{3}{2}\sqrt{x}\, dx = 1.$$

   Let T_n be the approximation to I given by the trapezoid rule with a partition of size n and let S_n be the approximation given by Simpson's rule with a partition of size n.

   Let $n_T(\epsilon)$ be the smallest value of n for which $|T_n - I| \le \epsilon$ and let $n_S(\epsilon)$ be the smallest value of n for which $|S_n - I| \le \epsilon$. Plot $n_T(\epsilon)$ and $n_S(\epsilon)$ against ϵ for $\epsilon = 2^{-k}$, $k = 2, \ldots, 16$.
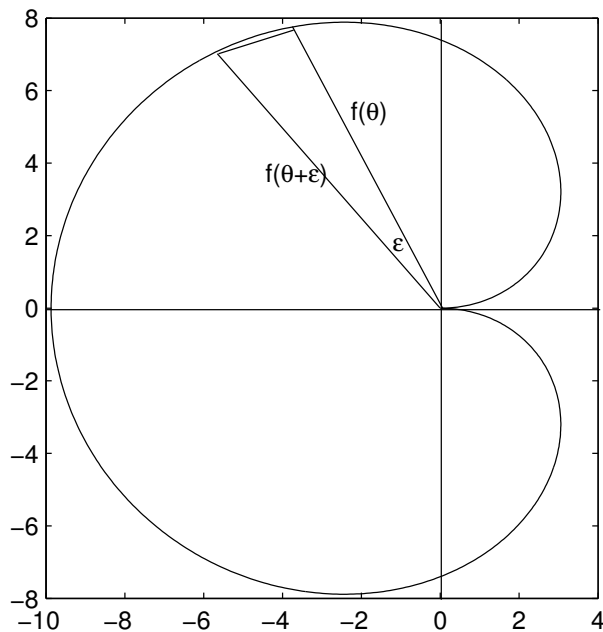
Figure 8: Integration using polar coordinates, as per Exercise 2.

5. Let $T(n)$ be the estimate of $I = \int_a^b f(x)dx$ obtained using the trapezoidal method with a partition of size n. If f has a continuous second derivative, then using Taylor's theorem one can show that the error $E(n) = |I - T(n)| = O(1/n^2)$. This suggests a method for improving the trapezoid method: if $T(n) \approx I + c/n^2$ and $T(2n) \approx I + c/(2n)^2$, for some constant c, then

$$R(2n) = (4T(2n) - T(n))/3 \approx (4I + c/n^2 - I - c/n^2)/3 = I.$$

That is, the errors cancel.

This is called Richardson's deferred approach to the mean. Show that $R(2n)$ is precisely $S(2n)$, that is, Simpson's rule using a partition of size $2n$.

Chapter 12

1. In the golden-section algorithm, suppose that you start with $x_l = 0$, $x_m = 0.5$, and $x_r = 1$, and that at each step if $x_r - x_m > x_m - x_l$, then $y = (x_m + x_r)/2$, while if $x_r - x_m \le x_m - x_l$, then $y = (x_l + x_m)/2$.

   In the worst-case scenario, for this choice of y, by what factor does the width of the bracketing interval reduce each time? In the worst-case scenario, is this choice of y better or worse than the usual golden-section rule?

   What about the best-case scenario?

2. Use the golden-section search algorithm to find all of the local maxima of the function

$$f(x) = \begin{cases} 0, & x = 0 \\ |x|\log(|x|/2)e^{-|x|}, & \text{o/w} \end{cases}$$

   within the interval $[-10, 10]$.

   Hint: plotting the function first will give you a good idea where to look.

3. Write a version of function `gsection` that plots intermediate results. That is, plot the function being optimised, then at each step draw a vertical line at the positions x_l, x_r, x_m, and y (with the line at y in a different colour).

4. The Rosenbrock function is a commonly used test function, given by

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2.$$

   You can plot the function in the region $[-2, 2] \times [-2, 5]$ using the code below (Figure 9). It has a single global minimum at $(1, 1)$.

```
# program spuRs/resources/scripts/Rosenbrock.r

Rosenbrock <- function(x) {
  g <- (1 - x[1])^2 + 100*(x[2] - x[1]^2)^2
  g1 <- -2*(1 - x[1]) - 400*(x[2] - x[1]^2)*x[1]
  g2 <- 200*(x[2] - x[1]^2)
  g11 <- 2 - 400*x[2] + 1200*x[1]^2
  g12 <- -400*x[1]
  g22 <- 200
  return(list(g, c(g1, g2), matrix(c(g11, g12, g12, g22), 2, 2)))
}

x <- seq(-2, 2, .1)
y <- seq(-2, 5, .1)
xyz <- data.frame(matrix(0, length(x)*length(y), 3))
names(xyz) <- c('x', 'y', 'z')
n <- 0
for (i in 1:length(x)) {
  for (j in 1:length(y)) {
    n <- n + 1
    xyz[n,] <- c(x[i], y[j], Rosenbrock(c(x[i], y[j]))[[1]])
  }
}
library(lattice)
print(wireframe(z ~ x*y, data = xyz, scales = list(arrows = FALSE),
  zlab = 'f(x, y)', drape = T))
```
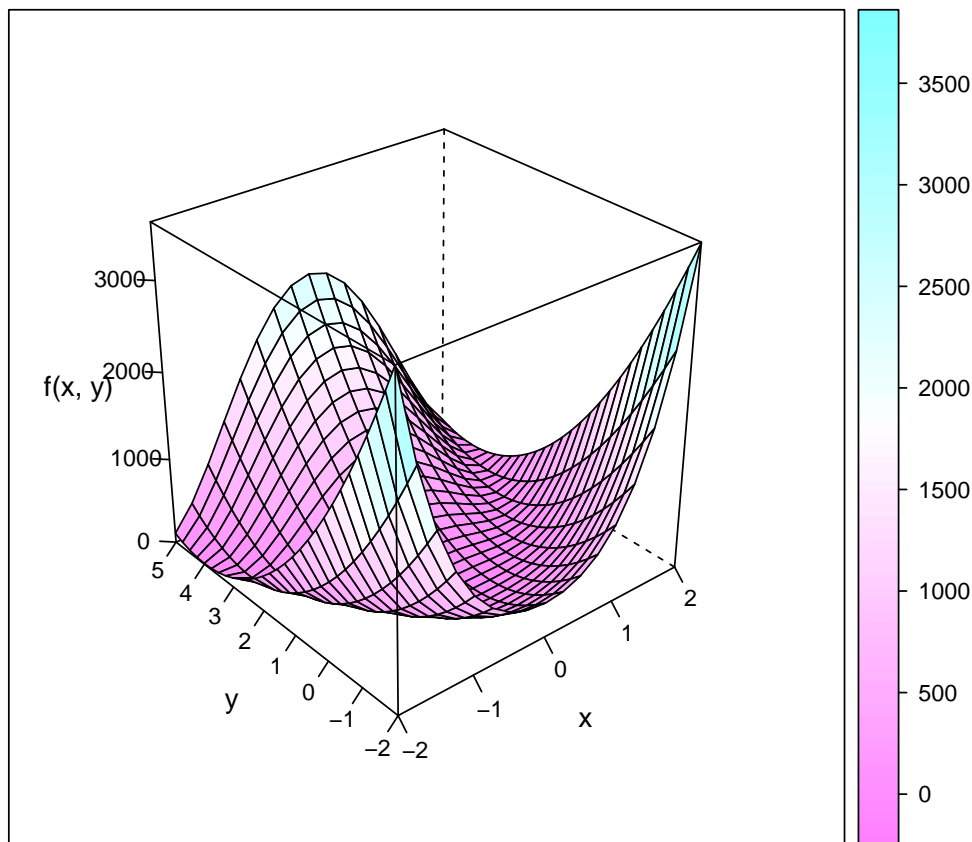
Figure 9: The Rosenbrock function. See Exercise 4.

Use the function `contour` to form a contour-plot of the Rosenbrock function over the region $[-2, 2] \times [-2, 5]$. Next modify `ascent` so that it plots each step on the contour-plot, and use it to find the *minimum* of the Rosenbrock function (that is, the maximum of $-f$), starting at $(0, 3)$. Use a tolerance of 10^{-9} and increase the maximum number of iterations to at least 10,000.

Now use `newton` to find the minimum (no need to use $-f$ in this case). Plot the steps made by the Newton method and compare them with those made by the steepest descent method.

5. Suppose $f : \mathbb{R}^d \to \mathbb{R}$. Since $\partial f(\mathbf{x})/\partial x_i = \lim_{\epsilon \to 0}(f(\mathbf{x} + \epsilon \mathbf{e}_i) - f(\mathbf{x}))/\epsilon$, we have for small ϵ

$$\frac{\partial f(\mathbf{x})}{\partial x_i} \approx \frac{f(\mathbf{x} + \epsilon \mathbf{e}_i) - f(\mathbf{x})}{\epsilon}.$$

In the same way, show that for $i \neq j$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \approx \frac{f(\mathbf{x} + \epsilon \mathbf{e}_i + \epsilon \mathbf{e}_j) - f(\mathbf{x} + \epsilon \mathbf{e}_i) - f(\mathbf{x} + \epsilon \mathbf{e}_j) + f(\mathbf{x})}{\epsilon^2}$$

and

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i^2} \approx \frac{f(\mathbf{x} + 2\epsilon \mathbf{e}_i) - 2f(\mathbf{x} + \epsilon \mathbf{e}_i) + f(\mathbf{x})}{\epsilon^2}.$$

Figure 10: A function with a number of local maxima: see Exercise 6.

(a) Test the accuracy of these approximations using the function $f(x, y) = x^3 + xy^2$ at the point $(1, 1)$. That is, for a variety of ϵ, calculate the approximate gradient and Hessian, and see by how much they differ from the true gradient and Hessian.

In R real numbers are only accurate to order 10^{-16} (try `1+10^{-16} == 1`). Thus the error in estimating $\partial f(\mathbf{x})/\partial x_i$ is of the order $10^{-16}/\epsilon$. For example, if $\epsilon = 10^{-8}$ then the error will be order 10^{-8}. It is worse for second-order derivatives: the error in estimating $\partial^2 f(\mathbf{x})/\partial x_i \partial x_j$ is of the order $10^{-16}/\epsilon^2$. Thus if $\epsilon = 10^{-8}$ then the error will be order 1. We see that we have a trade-off in our choice of ϵ: too large and we have a poor approximation of the limit; too small and we suffer rounding error.

(b) Modify the steepest ascent method, replacing the gradient with an approximation. Apply your modified algorithm to the function $f(x, y) = \sin(x^2/2 - y^2/4) \cos(2x - \exp(y))$, using the same starting points as in Example **??**.

How does the algorithm's behaviour depend on your choice of ϵ? You might find it helpful to plot each step, as in Exercise 4.

6. A simple way of using local search techniques to find a global maximum is to consider several different starting points, and hope that for one of them its local maximum is in fact the global maximum. If you have no idea where to start, then randomisation can be used to choose the starting point.

Consider the function

$$f(x, y) = -(x^2 + y^2 - 2)(x^2 + y^2 - 1)(x^2 + y^2)(x^2 + y^2 + 1)(x^2 + y^2 + 2)$$
$$\times \left(2 - \sin(x^2 - y^2) \cos(y - \exp(y))\right).$$

It has several local maxima in the region $[-1.5, 1.5] \times [-1.5, 1.5]$. Using several randomly chosen starting points, use steepest ascent to find all of the local maxima of f, and thus the global maximum. You can use the command `runif(2, -1.5, 1.5)` to generate a random point (x, y) in the region $[-1.5, 1.5] \times [-1.5, 1.5]$.

A picture of f is given in Figure 10. Note that f has been truncated below at -3.

7. This question follows on from Example **??**.

The three parameters of the Richards curve give a concise summary of the growth behaviour of a tree. In practice, the optimal management of a timber plantation requires knowledge of how different trees grow in different conditions, so that you can choose which trees to put where, and how far apart.

The table `trees.csv` contains information on spruce trees grown in a number of different sites. Each tree has ID of the form $x.y.z$, where x gives the site the tree is from and y a location within that site. Fit the Richards curve to all the trees given in the table, then for each tree plot the point (a, b) on a graph, where (a, b, c) are the curve parameters. Label each point according to the site the tree comes from: can you see any relation between the site a tree is from and the parameters of its Richards curve?

Hint: to print the character 1 at point (a, b) use `text(a, b, '1')`.

Chapter 13

1. List the sample space for the following random experiment. First you toss a coin. Then, if you get a head, you throw a single die.

2. Blood is of differing types or blood groups: O, A, B, and AB. Not all are compatible for transfusion purposes. Any recipient can receive the blood from a donor with the same blood group or from a donor with type O blood. A recipient with type AB blood can receive blood of any type. No other combinations will work. Consider an experiment which consists of drawing a litre of blood and determining its type for each of the next two donors who enter a blood bank.

    (a) List the possible (ordered) outcomes of this experiment.

    (b) List the outcomes where the second donor can receive the blood of the first.

    (c) List the outcomes where each donor can receive the blood of the other.

3. (a) The number of alpha particles emitted by a radioactive sample in a fixed time interval is counted. Give the sample space for this experiment.

    (b) The elapsed time is measured until the first alpha particle is emitted. Give the sample space for this experiment.

4. An experiment is conducted to determine what fraction of a piece of metal is gold. Give the sample space for this experiment.

5. A box of n components has r $(r < n)$ components which are defective. Components are tested one by one until all defective components are found, and the number of components tested is observed. Describe the sample space for this experiment.

6. Let A, B, C be three arbitrary events. Find expressions for the events that, of A, B, and C,

    (a) Only B occurs.

    (b) Both B and C, but not A, occur.

    (c) All three events occur.

    (d) At least one occurs.

    (e) None occur.

7. Using the probability axioms show that

$$\mathbb{P}(\bar{A} \cap \bar{B}) = 1 - \mathbb{P}(A \cup B).$$

You may find it helpful to draw a Venn diagram of A and B.

8. Is it possible to have an assignment of probabilities such that $\mathbb{P}(A) = 2/3$, $\mathbb{P}(B) = 1/5$, and $\mathbb{P}(A \cap B) = 1/4$?

9. When an experiment is performed, one and only one of the events A, B, or C will occur. Find $\mathbb{P}(A)$, $\mathbb{P}(B)$, and $\mathbb{P}(C)$ under each of the following assumptions:

    (a) $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C)$.

    (b) $\mathbb{P}(A) = \mathbb{P}(B)$ and $\mathbb{P}(C) = 1/4$.

    (c) $\mathbb{P}(A) = 2\mathbb{P}(B) = 3\mathbb{P}(C)$.

10. Consider a sample space $\Omega = \{a, b, c, d, e\}$ in which the following events are defined $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{d\}$, $E = \{e\}$. We are given a number of alternative probability measures on this sample space. It seems that in some of them an error has been made with the figures. Find those cases in which an error has been made, indicating why it must be an error. In those cases where there are no apparent errors, find $\mathbb{P}(E)$.

(a) $\mathbb{P}(A \cup B \cup C \cup D) = 0.5$, $\mathbb{P}(B \cup C \cup D) = 0.6$.

(b) $\mathbb{P}(A \cup B) = 0.3$, $\mathbb{P}(C \cup D) = 0.5$.

(c) $\mathbb{P}(A \cup B) = 0.6$, $\mathbb{P}(C) = 0.4$.

(d) $\mathbb{P}(A \cup B \cup C) = 0.7$, $\mathbb{P}(A \cup B) = \mathbb{P}(B \cup C) = 0.3$.

11. Let A and B be events in a sample space such that $\mathbb{P}(A) = \alpha$, $\mathbb{P}(B) = \beta$, and $\mathbb{P}(A \cap B) = \gamma$. Find an expression for the probabilities of the following events in terms of α, β, and γ.

(a) $\bar{A} \cap B$.

(b) $A \cap \bar{B}$.

(c) $\bar{A} \cap \bar{B}$.

12. If the occurrence of B makes A more likely, does the occurrence of A make B more likely?

13. Suppose that $\mathbb{P}(A) = 0.6$. What can you say about $\mathbb{P}(A|B)$ when

(a) A and B are mutually exclusive?

(b) A is a subset of B?

(c) B is a subset of A?

14. If A and B are events such that $\mathbb{P}(A) = 0.4$ and $\mathbb{P}(A \cup B) = 0.7$, find $\mathbb{P}(B)$ if A and B are

(a) Mutually exclusive.

(b) Independent.

15. Determine the conditions under which an event A is independent of its subset B.

16. How many times should a fair coin be tossed in order that the probability of observing at least one head is at least 0.99?

17. A random sample of size n is taken from a population of size N. Write down the number of distinct samples when sampling is:

(a) Ordered, with replacement.

(b) Ordered, without replacement.

(c) Unordered, without replacement.

(d) Unordered, with replacement.

(Note that (d) is hard; it may help to write down a few special cases first.)

18. Suppose that both a mother and father carry genes for blood types A and B. They each pass one of these genes to a child and each gene is equally likely to be passed. We assume they pass genes independently. The child will have blood type A if both parents pass their A genes, type B if both pass their B genes, and type AB if one A and one B gene are passed. What are the probabilities that a child of these parents has type A blood? Type B? Type AB?

19. An individual plays roulette using the following system. He bets $1 that the roulette wheel will come up black. If he wins, he quits. If he loses he makes the same bet a second time but now he bets $2. Then irrespective of the result, he quits. What is the sample space for this experiment? Assuming he has a probability of $^1/_2$ of winning each bet, what is the probability that he goes home a winner?

20. Two dice are rolled. What is the probability that at least one is a six? If the two faces are different, what is the probability that one is a six? If the sum is seven what is the probability that one die shows a six?

21. A woman has two children. Assume that all possible outcomes for the sexes are equally likely. What is the probability that she has two boys given that

(a) The eldest is a boy.

(b) At least one is a boy.

22. Two archers A and B shoot at the same target. Suppose A hits the target with probability 0.65 and independently B hits the target with probability 0.5.

(a) Given only one of the archers hits the target, what is the probability it was A?

(b) Given at least one of them hits, what is the probability that B hits?

23. A diagnostic test is used to determine whether or not a person has a certain disease. If the test is positive, then it is assumed the person has the disease, if negative that they don't have it. However the test is not 100% accurate. If a diseased person is tested, it still gives a negative result 5% of the time (a false negative) and when testing a person free of the disease, it gives a false positive 10% of the time. Suppose we choose someone at random from a population in which only 1 person in 50 has the disease.

(a) Find the probability that their test result is positive.

(b) Find the probability that their test result is misleading.

(c) Find the probability that they actually have the disease if they test positive.

24. There are two bus lines which travel between towns A and B. Bus line A runs late 20% of the time, while bus line B runs late 50% of the time. You travel three times as often by line A as you do by line B. On a certain day you arrive late. What is the probability that you used bus line B that day?

25. An electronic system receives signals as input and sends out appropriate coded messages as output.

The system consists of 3 converters (C_1, C_2, and C_3), 2 monitors (M_1 and M_2), and a perfectly reliable three-way switch for connecting the input to the converters. The incoming signal is encoded by one or more of the converters and the monitors check whether the conversion is correct.

Initially the signal is fed into C_1. If M_1 passes the conversion, the coded message is sent out. If M_1 rejects the conversion, the input is switched to C_2 and the conversion is checked by M_2. If M_2 passes the conversion, the coded message is sent out. If M_2 rejects the conversion, the input is switched to C_3 and the coded message is sent out without any further checks.

Each of the converters has probability 0.9 of correctly coding the incoming message. Each of the monitors has probability 0.8 of rejecting a wrongly coded message and also probability 0.8 of passing a correctly coded message.

Show that the probability of a correct output from the system is about 0.968.

26. The dice game craps is played as follows. The player throws two dice, and if the sum is seven or eleven, then he wins. If the sum is two, three, or twelve, then he loses. If the sum is anything else, then he continues throwing until he either throws that number again (in which case he wins) or he throws a seven (in which case he loses). Calculate the probability that the player wins.

27. If you toss a coin four times, the probability of getting four heads in a row is $(0.5)^4 = 0.0625$. Suppose that we toss a coin twenty times; what is the probability that we get a sequence of four heads in a row at some point?

Write a program to estimate this probability. Your answer should be greater than $1 - (15/16)^5$. (Why?)

Your program should have the following structure:

(a) A function `four.n.twenty()` that simulates twenty coin tosses, then checks to see if there are four heads in a row.

(b) A function `four.n.twenty.prob(N)` that executes `four.n.twenty()` N times and returns the proportion of times there were four heads in a row.

Using `four.n.twenty.prob(N)`, what can you do to be confident that your answer is accurate to two decimal places?

To simulate twenty coin tosses you can use the command `round(runif(20))`, then interpret a 1 as a head and a 0 as a tail. One way to structure `four.n.twenty()` is to first generate a sequence of twenty coin tosses, then for $i = 1, \ldots, 17$ check to see if tosses $(i, i+1, i+2, i+3)$ are all heads. Suppose that you use 1 for a head and 0 for a tail, and that `coins` is a vector of 0's and 1's of length n, then `coins` is a sequence of n heads if and only if `prod(coins) == 1`.

Chapter 14

1. Suppose you throw two dice. What values can the following random variables take?

    (a) The minimum face value showing;

    (b) The absolute difference between the face values showing;

    (c) The ratio: minimum face value/other face value.

    Assuming all outcomes in the sample space are equally likely, what are the probability mass functions for these random variables? Give these in a table format and also do a rough sketch.

    Calculate the mean of each random variable.

2. The following is the probability mass function of a discrete random variable X

x	1	2	3	4	5
$\mathbb{P}(X = x)$	$2c$	$3c$	c	$4c$	$5c$

    (a) What is the value of c?

    (b) Find $\mathbb{P}(X \leq 4)$ and $\mathbb{P}(1 < X < 5)$.

    (c) Calculate $\mathbb{E}X$ and $\operatorname{Var}X$.

3. A game consists of first tossing an unbiased coin and then rolling a six-sided die. Let the random variable X be the score that is obtained by adding the face value of the die and the number of heads obtained (0 or 1). List the possible values of X and calculate its pmf.

4. This question concerns an experiment with sample space

    $$\Omega = \{a, b, c, d\}.$$

    (a) List all possible events for this experiment.

    (b) Suppose that $\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = \mathbb{P}(\{d\}) = 1/4$.
    Find two independent events and two dependent events.

    (c) Define random variables X, Y, and Z as follows

ω	a	b	c	d
$X(\omega)$	1	1	0	0
$Y(\omega)$	0	1	0	1
$Z(\omega)$	1	0	0	0

    Show that X and Y are independent and that X and Z are dependent.

    (d) Let $W = X + Y + Z$. What is the probability mass function (pmf) of W? What is its mean and variance?

5. A discrete random variable has pmf $f(x) = k(1/2)^x$ for $x = 1, 2, 3$; $f(x) = 0$ for all other values of x. Find the value of k and then the mean and variance of the random variable.

6. Consider the discrete random variable X that takes the values $0, 1, 2, \ldots, 9$ each with probability $1/10$. Let Y be the remainder obtained after dividing X^2 by 10 (e.g., if $X = 9$ then $Y = 1$). Y is a function of X and so is also a random variable. Find the pmf of Y.

7. Consider the discrete probability distribution defined by

    $$p(x) = \mathbb{P}(X = x) = \frac{1}{x(x + 1)} \quad \text{for } x = 1, 2, 3, \ldots$$

    (a) Let $S(n) = \mathbb{P}(X \leq n) = \sum_{x=1}^{n} p(x)$. Using the fact that $\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$, find a formula for $S(n)$ and thus show that p is indeed a pmf.

(b) Write down the formula for the mean of this distribution. What is the value of this sum?

8. For some fixed integer k, the random variable Y has probability mass function (pmf)

$$p(y) = \mathbb{P}(Y = y) = \begin{cases} c(k-y)^2 & \text{for } y = 0, 1, 2, \ldots, k-1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of c? (Your answer will depend on k.)
    Hint: $\sum_{i=1}^{n} i^2 = n(n+1)(2n+1)/6$.

(b) Give a formula for the cumulative distribution function (cdf) $F(y) = \mathbb{P}(Y \leq y)$ for $y = 0, 1, 2, \ldots, k-1$.

(c) Write a function in R to calculate $F(y)$. Your function should take y and k as inputs and return $F(y)$. You may assume that k is an integer greater than 0 and that $y \in \{0, 1, 2, \ldots, k-1\}$.

9. Toss a coin 20 times and let X be the length of the longest sequence of heads. We wish to estimate the probability function p of X. That is, for $x = 1, 2, \ldots, 20$ we wish to estimate

$$p(x) = \mathbb{P}(X = x).$$

Here is a function `maxheads(n.toss)` that simulates X (using `n.toss = 20`).

```
maxheads <- function(n.toss) {
  # returns the length of the longest sequence of heads
  # in a sequence of n.toss coin tosses
  n_heads = 0   # length of current head sequence
  max_heads = 0 # length of longest head sequence so far
  for (i in 1:n.toss) {
    # toss a coin and work out length of current head sequence
    if (runif(1) < 0.5) { # a head, sequence of heads increases by 1
      n_heads <- n_heads + 1
    } else { # a tail, sequence of heads goes back to 0
      n_heads <- 0
    };
    # see if current sequence of heads is the longest
    if (n_heads > max_heads) {
      max_heads <- n_heads
    }
  }
  return(max_heads)
}
```

Use `maxheads(20)` to generate an iid sample $X_1, \ldots, X_N$ then estimate p using

$$\hat{p}(x) = \frac{|\{X_i = x\}|}{N}.$$

Print out your estimate as a table like this

```
  x   p_hat(x)
--------------
  0    0.0010
  1    0.0500
  .      .
  .      .
 20    0.0000
```

As a supplementary exercise try rewriting the function `maxheads` using the R function `rle`.

10. Suppose the rv X has continuous pdf $f(x) = 2/x^2$, $1 \le x \le 2$. Determine the mean and variance of X and find the probability that X exceeds 1.5.

11. Which of the following functions are probability density functions for a continuous random variable X?

    (a)
    $$f(x) = \begin{cases} 5x^4 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

    (b)
    $$f(x) = \begin{cases} 2x & -1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

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

    (d)
    $$f(x) = \begin{cases} 2x/9 & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

    For those that are pdfs, calculate $\mathbb{P}(X \le 1/2)$.

12. Suppose a continuous random variable Y has pdf
    $$f(y) = \begin{cases} 3y^2 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

    (a) Sketch this pdf and find $\mathbb{P}(0 \le Y \le 1/2)$ and $\mathbb{P}(1/2 \le Y \le 1)$.
    (b) Find the cdf $F_Y(y)$ of Y.

13. Suppose a continuous random variable Z has pdf
    $$f_Z(z) = \begin{cases} z - 1 & 1 \le z \le 2 \\ 3 - z & 2 \le z \le 3 \\ 0 & \text{otherwise} \end{cases}$$

    (a) Sketch this pdf and find $\mathbb{P}(Z \le 3/2)$ and $\mathbb{P}(3/2 \le Z \le 5/2)$.
    (b) Find the cdf $F_Z(z)$ of Z.

14. A random variable X has cdf
    $$F_X(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x} & 0 < x < \infty \end{cases}$$

    (a) Sketch this cdf.
    (b) Is X continuous or discrete? What are the possible values of X?
    (c) Find $\mathbb{P}(X \ge 2)$, $\mathbb{P}(X \le 2)$, and $\mathbb{P}(X = 0)$.

15. A random variable X has cdf
    $$F_X(x) = \begin{cases} x/2 & 0 < x \le 1 \\ x - 1/2 & 1 < x < 3/2 \end{cases}$$

    (a) Sketch this cdf.
    (b) Is X continuous or discrete? What are the possible values of X?

(c) Find $\mathbb{P}(X \le 1/2)$ and $\mathbb{P}(X \ge 1/2)$.

(d) Find a number m such that $\mathbb{P}(X \le m) = \mathbb{P}(X \ge m) = 1/2$ (the *median*).

16. For Exercises 12–15 above, try to guess the mean by judging the 'centre of gravity' of the pdf. Then check your guess by evaluating the mean analytically.

17. Consider two continuous random variables X and Y with pdfs

$$ f_X(x) = \begin{cases} 4x^3 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} $$

$$ f_Y(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases} $$

(a) Sketch both these pdfs and try to guess the means of X and Y. Check your guesses by actually calculating the means.

(b) From the sketches, which random variable do you think would be more variable, X or Y? Check your guess by actually calculating the variances.

18. It is known that a good model for the variation, from item to item, of the quality of a certain product is the random variable X with probability density function $f(x) = 2x/\lambda^2$ for $0 \le x \le \lambda$. Here λ is a parameter that depends on the manufacturing process, and can be altered.

During manufacture, each item is tested. Items for which $X > 1$ are passed, and the rest are rejected. The cost of a rejected item is $c = a\lambda + b$ and the profit on a passed item is $d - c$, for constants a, b, and d.

Find λ such that the expected profit is maximised.

19. The variable X has pdf $f(x) = \frac{1}{8}(6 - x)$ for $2 \le x \le 6$. A sample of two values of X is taken. Denoting the lesser of the two values by Y, use the cdf of X to write down the cdf of Y. Hence obtain the pdf and mean of Y. Show that its median is approximately 2.64. (The median is the point m for which $\mathbb{P}(Y \le m) = 0.5$.)

20. Discrete and continuous are not the only possible types of random variable. For example, what sort of distribution is the time spent waiting in a bank queue? If we suppose that there is a strictly positive probability of waiting no time at all, then the cumulative distribution function will have a jump at 0. However, if there are people ahead of you, the time you wait could be any value in $(0, \infty)$, so that this part of the cumulative distribution will be a continuous function. Thus this distribution is a mixture of a discrete and continuous part.

Let X be the length of time that a customer is in the queue, and suppose that

$$ F(x) = 1 - pe^{-\lambda x} \text{ for } x \ge 0, \lambda > 0 \text{ and } 0 < p < 1. $$

Find $\mathbb{P}(X = 0)$ and the cdf for $X \mid X > 0$. Hence, find the mean queuing time, noting that (from the Law of Total Probability)

$$
\begin{aligned}
\mathbb{E}X &= \mathbb{E}(X \mid X = 0)\mathbb{P}(X = 0) + \mathbb{E}(X \mid X > 0)\mathbb{P}(X > 0) \\
&= 0 + \mathbb{E}(X \mid X > 0)\mathbb{P}(X > 0).
\end{aligned}
$$

21. Let $X_1, \ldots, X_n$ be an iid sample with mean μ and variance σ^2. Show that you can write

$$ (n - 1)S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\mu - \overline{X})^2. $$

Now suppose that $\mathbb{E}(X_i - \mu)^4 < \infty$, and use the Weak Law of Large Numbers to show that

$$ S^2 \xrightarrow{\mathbb{P}} \sigma^2 \text{ as } n \to \infty. $$

Chapter 15

1. The probability of recovery from a certain disease is 0.15. Nine people have contracted the disease. What is the probability that at most 2 of them recover? What is the expected number that will recover?

2. On a multiple choice exam with five possible answers for each of the ten questions, what is the probability that a student would get three or more correct answers just by guessing (choosing an answer at random)? What is the expected number of correct answers the student would get just by guessing?

3. An airline knows that on average 10% of people making reservations on a certain flight will not show up. So they sell 20 tickets for a flight that can only hold 18 passengers.

   (a) Assuming individual reservations are independent, what is the probability that there will be a seat available for every passenger that shows up?

   (b) Now assume there are 15 flights of this type in one evening. Let N_0 be the number of these flights on which everyone who shows up gets a seat and N_1 be the number of these flights that leave just one disgruntled person behind. What are the distributions of N_0 and N_1? What are their means and variances?

   (c) The independence assumption in (a) is not really very realistic. Why? Try to describe what might be a more realistic model for this situation.

4. In the board game Monopoly you can get out of jail by throwing a double (on each turn you throw two dice). Let N be the number of throws required to get out of jail this way. What is the distribution of N, $\mathbb{E}(N)$, and $\text{Var}(N)$?

5. A couple decides to keep having children until they have a daughter. That is, they stop when they get a daughter even if she is their first child. Let N be the number of children they have. Assume that they are equally likely to have a boy or a girl and that the sexes of their children are independent.

   (a) What is the distribution of N? $\mathbb{E}(N)$? $\text{Var}(N)$?

   (b) Write down the probabilities that N is 1, 2, or 3.

   Another couple decides to do the same thing but they don't want an only child. That is they have two children and then only keep going if they haven't yet had a daughter. Let M be the number of children they have.

   (c) Calculate $\mathbb{P}(M = 1)$, $\mathbb{P}(M = 2)$, and $\mathbb{P}(M = 3)$.

   (d) Explain why we must have $\mathbb{P}(N = i) = \mathbb{P}(M = i)$ for any $i \geq 3$.

   (e) Using the above information calculate $\mathbb{E}(M)$.
   Hint: use the known value of $\mathbb{E}(N)$ and consider the difference $\mathbb{E}(N) - \mathbb{E}(M)$.

6. A random variable $Y \sim \text{pois}(\lambda)$ and you are told that λ is an integer.

   (a) Calculate $\mathbb{P}(Y = y)/\mathbb{P}(Y = y + 1)$ for $y = 0, 1, \ldots$

   (b) What is the most likely value of Y?
   Hint: what does it mean if the ratio in (a) is less than one?

7. If X has a Poisson distribution and $\mathbb{P}(X = 0) = 0.2$, find $\mathbb{P}(X \geq 2)$.

8. Suppose $X \sim \text{pois}(\lambda)$.

   (a) Find $\mathbb{E}X(X - 1)$ and thus show that $\text{Var}\, X = \lambda$.

   (b) Using the fact that $\text{binom}(n, \lambda/n)$ probabilities converge to $\text{pois}(\lambda)$ probabilities, as $n \to \infty$, again show that $\text{Var}\, X = \lambda$.

9. Large batches of components are delivered to two factories, A and B. Each batch is subjected to an acceptance sampling scheme as follows:

Factory A: Accept the batch if a random sample of 10 components contains less than two defectives. Otherwise reject the batch.

Factory B: Take a random sample of five components. Accept the batch if this sample contains no defectives. Reject the batch if this sample contains two or more defectives. If the sample contains one defective, take a further sample of five and accept the batch if this sample contains no defectives.

If the fraction defective in the batch is p, find the probabilities of accepting a batch under each scheme.

Write down an expression for the average number sampled in factory B and find its maximum value.

10. A new car of a certain model may be assumed to have X minor faults where X has a Poisson distribution with mean μ. A report is sent to the manufacturer listing the faults for each car that has at least one fault. Write down the probability function of Y, the number of faults listed on a randomly chosen report card and find $\mathbb{E}(Y)$. Given $\mathbb{E}(Y) = 2.5$, find μ correct to two decimal places.

11. A contractor rents out a piece of heavy equipment for t hours and is paid \$50 per hour. The equipment tends to overheat and if it overheats x times during the hiring period, the contractor will have to pay a repair cost \$$x^2$. The number of times the equipment overheats in t hours can be assumed to have a Poisson distribution with mean $2t$. What value of t will maximise the expected profit of the contractor?

12. Calculating binomial probabilities using a recursive function.

Let $X \sim \text{binom}(k, p)$ and let $f(x, k, p) = \mathbb{P}(X = x) = \binom{k}{x} p^x (1 - p)^{k-x}$ for $0 \leq x \leq k$ and $0 \leq p \leq 1$. It is easy to show that

$$
\begin{aligned}
f(0, k, p) &= (1 - p)^k; \\
f(x, k, p) &= \frac{(k - x + 1)p}{x(1 - p)} f(x - 1, k, p) \text{ for } x \geq 1.
\end{aligned}
$$

Use this to write a recursive function `binom.pmf(x, k, p)` that returns $f(x, k, p)$.

You can check that your function works by comparing it with the built-in function `dbinom(x, k, p)`.

13. An airline is selling tickets on a particular flight. There are 50 seats to be sold, but they sell $50 + k$ as there are usually a number of cancellations.

Suppose that the probability a customer cancels is $p = 0.1$ and assume that individual reservations are independent. Suppose also that the airline makes a profit of \$500 for each passenger who travels (does not cancel and does get a seat), but loses \$100 for each empty seat on the plane and loses \$500 if a customer does not get a seat because of overbooking. The loss because of an empty seat is due to the fixed costs of flying a plane, irrespective of how many passengers it has. The loss if a customer does not get a seat represents both an immediate cost—for example they may get bumped up to first class—as well as a the cost of lost business in the future.

What value of k maximises the airline's expected profit?

14. Write a program to calculate $\mathbb{P}(X + Y + Z = k)$ for arbitrary discrete non-negative rv's X, Y, and Z.

Chapter 16

1. A random variable U has a $U(a,b)$ distribution if $\mathbb{P}(U \in (u,v)) = (v-u)/(b-a)$ for all $a \le u \le v \le b$. Show that if $U \sim U(a,b)$ then so is $a+b-U$.

2. Show that if $X \sim \exp(\lambda)$ and $Y \sim \exp(\mu)$, independently of X, then $Z = \min\{X,Y\} \sim \exp(\lambda + \mu)$.
   Hint: $\min\{X,Y\} > z \iff X > z$ and $Y > z$.

3. The time to failure of a new type of light bulb is thought to have an exponential distribution.

   Reliability is defined as the probability that an article will not have failed by a specified time. If the reliability of this type of light bulb at 10.5 weeks is 0.9, find the reliability at 10 weeks.

   One hundred bulbs of this type are put in a new shop. All the bulbs that have failed are replaced at 20-week intervals and none are replaced at other times. If R is the number of bulbs that have to be replaced at the end of the first interval, find the mean and variance of R.

   Explain why this result will hold for *any* such interval and not just the first.

4. The length of a certain type of battery is normally distributed with mean 5.0 cm and standard deviation 0.05 cm. Find the probability that such a battery has a length between 4.92 and 5.08 cm.

   Tubes are manufactured to contain four such batteries. 95% of the tubes have lengths greater than 20.9, and 10% have lengths greater than 21.6 cm. Assuming that the lengths are also normally distributed, find the mean and standard deviation, correct to two decimal places.

   If tubes and batteries are chosen independently, find the probability that a tube will contain four batteries with at least 0.75 cm to spare.

5. A man travels to work by train and bus. His train is due to arrive at 08:45 and the bus he hopes to catch is due to leave at 08:48. The time of arrival of the train has a normal distribution with mean 08:44 and standard deviation three mins; the departure time of the bus is independently normally distributed with mean 08:50 and standard deviation one minute. Calculate the probabilities that:

   - The train is late;
   - The bus departs before the train arrives;
   - In a period of five days there are at least three days on which the bus departs before the train arrives.

6. Suppose $X \sim U(0,1)$ and $Y = X^2$.
   Use the cdf of X to show that $P(Y \le y) = \sqrt{y}$ for $0 < y < 1$, and thus obtain the pdf of Y. Hence or otherwise evaluate $\mathbb{E}(Y)$ and $\text{Var}(Y)$.

7. A mechanical component is only usable if its length is between 3.8 cm and 4.2 cm. It is observed that on average 7% are rejected as undersized, and 7% are rejected oversized. Assuming the lengths are normally distributed, find the mean and standard deviation of the distribution.

8. Telephone calls arrive at a switchboard in accordance with a Poisson process of rate $\lambda = 5$ per hour.

   (a) What is the distribution of N_1 = the number of calls that arrive in any one hour period?
   (b) What is the distribution of N_2 = the number of calls that arrive in any half hour period?
   (c) Find the probability that the operator is idle for the next half hour.

9. Glass sheets have faults called 'seeds', which occur in accordance with a Poisson process at a rate of 0.4 per square metre. Find the probability that rectangular sheets of glass of dimensions 2.5 metres by 1 metre will contain:

   (a) No seeds.
   (b) More than one seed.

If sheets with more than one seed are rejected, find the probability that in a batch of 10 sheets, at most one is rejected.

10. Cars pass through an intersection in accordance with a Poisson process with rate $\lambda = 3$ per minute. A pedestrian takes s seconds to cross at the intersection and chooses to start to cross irrespective of the traffic conditions. Assume that if he is on the intersection when a car passes by, then he is injured. Find the probability that the pedestrian crosses safely for $s = 5$, 10, and 20.

11. We examine blood under a microscope for red blood cell deficiency, using a small fixed volume that will contain on the average five red cells for a normal person. What is the probability that a specimen from a normal person will contain only two red cells or fewer (assume that cells are independently and uniformly distributed throughout the volume)?

12. Defects occur in an optical fibre in accordance with a Poisson process with rate $\lambda = 4.2$ per kilometre. Let N_1 be the number of defects in the first kilometre of fibre and N_2 be the number of defects in the second and third kilometres of fibre.

    (a) What are the distributions of N_1 and N_2?
    (b) Are N_1 and N_2 dependent or independent?
    (c) Let $N = N_1 + N_2$. What is the distribution of N?

13. The time (in hours) until failure of a transistor is a random variable $T \sim \exp(1/100)$.

    (a) Find $\mathbb{P}(T > 10)$.
    (b) Find $\mathbb{P}(T > 100)$.
    (c) It is observed that after 90 hours the transistor is still working. Find the conditional probability that $T > 100$, that is, $\mathbb{P}(T > 100 \,|\, T > 90)$. How does this compare with part (a)? Explain this result.

14. Jobs submitted to a computer system have been found to require a CPU time T, which is exponentially distributed with mean 150 milliseconds. If a job doesn't complete within 90 milliseconds is suspended and put back at the end of the queue. Find the probability that an arriving job will be forced to wait for a second quantum.

15. An insurance company has received notification of five pending claims. Claim settlement will not be complete for at least one year. An actuary working for the company has been asked to determine the size of the reserve fund that should be set up to cover these claims. Claims are independent and exponentially distributed with mean \$2,000. The actuary recommends setting up a claim reserve of \$12,000. What is the probability that the total claims will exceed the reserve fund?

16. Suppose that $X \sim U(0, 1)$.

    (a) Put $Y = h(X)$ where $h(x) = 1 + x^2$. Find the cdf F_Y and the pdf f_Y of Y.
    (b) Calculate $\mathbb{E}Y$ using $\int y f_Y(y)\, dy$ and $\int h(x) f_X(x)\, dx$.
    (c) The function `runif(n)` simulates n iid $U(0,1)$ random variables, thus `1 + runif(n)^2` simulates n iid copies of Y.

    Estimate and plot the pdf of Y using a simulated random sample. Experiment with the bin width to get a good-looking plot: it should be reasonably detailed but also reasonably smooth. How large does your sample have to be to get a decent approximation?

17. Let $N(t)$ be the number of arrivals up to and including time t in a Poisson process of rate λ, with $N(0) = 0$. In this exercise we will verify that $N(t)$ has a pois(λt) distribution.

    We define the Poisson process in terms of the times between arrivals, which are independent with an $\exp(\lambda)$ distribution. The first part of the task is to simulate $N(t)$ by simulating all the arrival times up until time t. Let $T(k)$ be the time of the first arrival, then

    $$T(1) \sim \exp(\lambda) \text{ and } T(k) - T(k-1) \sim \exp(\lambda).$$

Given the arrival times we get $N(t) = k$ where k is such that

$$T(k) \leq t < T(k+1).$$

Thus to simulate $N(t)$ we simulate $T(1), T(2), \ldots$, and keep going until we get $T(n) > t$, then put $N(t) = n - 1$.

Once you have code that can simulate $N(t)$, use it to generate a sample with $\lambda = 0.5$ and $t = 10$. Now check the distribution of $N(t)$ by using the sample to estimate the probability function of $N(t)$. That is, for $x \in \{0, 1, 2, \ldots\}$ (stop at around 20), we calculate $\hat{p}(x)$ = proportion of sample with value x, and compare the estimates with the theoretical Poisson probabilities

$$p(x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!}.$$

An easy way to compare the two is to plot $\hat{p}(x)$ for each x and then on the same graph plot the true probability function using vertical lines with heights $p(x)$,

You might also like to try plotting the *sample path* of a Poisson process. That is, plot $N(t)$ as a function of t.

Chapter 17

1. Using a normal approximation, find the probability that a Poisson variable with mean 20 takes the value 20. Compare this with the true value; to how many decimal places do they agree?

2. Migrating geese arrive at a certain wetland at a rate of 220 per day during the migration season. Suggest a model for X, the number of geese that arrive per hour (assume the arrival rate remains constant throughout the day).

   What is $\mathbb{P}(X > 10)$? Give the answer exactly, based on your model (it is sufficient to express the probability as a finite sum), and approximately, using the Central Limit Theorem.

3. The weights of 20 people are measured, and the resulting sample mean and sample standard deviation are
$$\overline{x} = 71.2 \text{ kg}, \qquad s = 4.9 \text{ kg}.$$
Calculate a 95% CI for the mean μ of the underlying population. Assume that the weights are iid normal.

4. A random sample of size n is taken without replacement from a very large sample of components and r of the sample are found to be defective. Write down an approximate 99% confidence interval for the proportion of the population that are defective stating clearly three reasons why your interval is only approximate.

   If $n = 400$ show that the longest the confidence interval can be is about 0.13.

5. Assume a manager is using the sample proportion $\hat{p}$ to estimate the proportion p of a new shipment of computer chips that are defective. He doesn't know p for this shipment, but in previous shipments it has been close to 0.01, that is 1% of chips have been defective.

   (a) If the manager wants the standard deviation of $\hat{p}$ to be about 0.02, how large a sample should she take based on the assumption that the rate of defectives has not changed dramatically?

   (b) Now suppose something went wrong with the production run and the actual proportion of defectives in the shipment is 0.3, that is 30% are defective. Now what would be the actual standard deviation of $\hat{p}$ for the sample size you choose in (a)?

6. A company fills plastic bottles with orange juice. The bottles are supposed to contain 250 ml. In fact, the contents vary according to a normal distribution with mean $\mu = 242$ ml and standard deviation $\sigma = 12$ ml.

   (a) What is the probability that one bottle contains less than 250 ml?

   (b) What is the probability that the mean contents of a carton with 12 bottles is less than 250 ml?

7. The number of accidents per week at a hazardous intersection follows a Poisson distribution with mean 2.2. We observe the intersection for a full year (52 weeks) and calculate $\overline{X}$ the mean number of accidents per week.

   (a) What is the approximate distribution of $\overline{X}$ according to the Central Limit Theorem?

   (b) What is the approximate probability that $\overline{X}$ is less than 2?

   (c) What is the approximate distribution of T, the total number of accidents in the year?

   (d) What is the probability that there are fewer than 90 accidents at the intersection during the year?

8. A scientist is observing the radioactive decay of a substance. The waiting time between successive decays has an exponential distribution with a mean of 10 minutes.

   (a) What is the probability that the first waiting time exceeds 12 minutes?

   (b) The scientist observes 50 successive waiting times and calculates the mean. What is the probability that this mean exceeds 12 minutes?

(c) In another experiment the scientist waits until the 80th decay. What is the probability that he waits longer than 14 hours?

9. An actuary has received notification that 100 claims on an account have been filed but are still in the course of settlement. The actuary has been asked to determine the size of an appropriate reserve fund for these 100 claims. Claim sizes are independent and exponentially distributed with mean $300. The actuary recommends setting up a claim reserve of $31,000. What is the probability that the total claims will exceed the reserve fund?

Hint: use an appropriate approximation.

10. Suppose that 55% of the voting population are Democrat voters. If 200 people are selected at random from the population, what is the probability that more than half of them are Democrat voters?

11. Approximate the probability that the proportion of heads obtained will be between 0.50 and 0.52 when a fair coin is tossed

    (a) 50 times.
    (b) 500 times.

12. A course can cater for 200 new students. Not all offers to students are accepted, so 250 offers are made based on previous rejection rates. Assume that for this current round of offers the actual rejection rate is 35% and that students make their decisions independently.

    (a) State the distribution of N, the number of students who accept, and state its mean and standard deviation.
    (b) Find the approximate probability that less than 180 students accept.
    (c) Find the approximate probability that more than 200 students accept.

13. A survey of 900 people asked whether they play any competitive sport. In fact only 5% of the surveyed population plays a competitive sport.

    (a) Find the mean and standard deviation of the proportion of the sample who play competitive sport.
    (b) What sample size would be required to reduce the standard deviation of the sample proportion to one-half the value you found in (a)?

14. Cards with different shapes printed on them are used to test if a subject has extrasensory perception (ESP). The subject has to guess the shape on the card being viewed by the experimenter without viewing the card itself. Assume we use a large pack containing cards marked with one of four different shapes in equal proportions. That is, we can assume that on each draw, each shape is equally likely, and that successive draws are independent. We test subjects (who are all just guessing at random) on 800 cards each.

    (a) What is the probability that any one subject guesses correctly on any one trial?
    (b) What are the mean and standard deviation of the proportion of successes among the 800 attempts?
    (c) What is the probability that any one subject is successful in at least 26% of the 800 attempts?
    (d) Assume you decide to do further tests on any subject whose proportion of successes is so large that there is only a probability of 0.02 that they could do that well or better simply by guessing. What proportion of successes must a subject have to meet this standard?
    (e) How many subjects will the researcher need to assess so that the probability at least one of them will be tested further is 0.75?

15. You take a random sample of size n from a population which is uniform on the interval $(0, \theta)$, where θ is an unknown parameter.

    (a) Using the Central Limit Theorem, about which point do you think the distribution of the sample mean will become concentrated as the sample size increases? Consequently, what function of the sample mean would you suggest to estimate θ?

(b) Arguing intuitively, what do you think will happen to the distribution of the sample maximum as the sample size increases?

(c) Suppose that $X \sim U(0, \theta)$; write down the pdf and cdf of X. Hence find the cdf and pdf of the sample maximum.

(d) Calculate the expected value of the sample maximum. Use this result to suggest a function of the sample maximum which would give you an unbiased estimate of the unknown parameter θ.

16. Calculating the confidence interval.

Write a function that takes as input a vector `x`, then returns as output the vector `(m, lb, ub)`, where `m` is the mean and `(lb, ub)` is a 95% confidence interval for `m`. That is

$$
\begin{aligned}
m &= \bar{x}, \\
lb &= \bar{x} - 1.96\sqrt{s^2/n}, \\
ub &= \bar{x} + 1.96\sqrt{s^2/n},
\end{aligned}
$$

where

$$
s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).
$$

Write a program that applies your subroutine to the following sample

`11 52 87 45 39 95 42 38 10 03 48 56`

To four decimal places you should be getting $(43.8333, 27.9526, 59.7140)$.

17. Gaining confidence with confidence intervals.

We know that the $U(-1, 1)$ rv has mean 0. Use a sample of size 100 to estimate the mean and give a 95% confidence interval. Does the confidence interval contain 0?

Repeat the above a large number of times. What percentage of time does the confidence interval contain 0? Write your code so that it produces output similar to the following

```
Number of trials: 10

Sample mean  lower bound  upper bound  contains mean?

   -0.0733     -0.1888      0.0422          1
   -0.0267     -0.1335      0.0801          1
   -0.0063     -0.1143      0.1017          1
   -0.0820     -0.1869      0.0230          1
   -0.0354     -0.1478      0.0771          1
   -0.0751     -0.1863      0.0362          1
   -0.0742     -0.1923      0.0440          1
    0.0071     -0.1011      0.1153          1
    0.0772     -0.0322      0.1867          1
   -0.0243     -0.1370      0.0885          1

100 percent of CI's contained the mean
```

18. Use `rnorm(10, 1, 1)` to generate a sample of 10 independent $N(1, 1)$ random variables. Form a 90% CI for the mean of this sample (using a t distribution). Does this CI include 1?

Repeat the above 20 times. How many times did the CI include 1? How many times do you expect the CI to include 1?

19. A bottle-washing plant has to discard many bottles because of breakages. Bottles are washed in batches of 144. Let X_i be the number of broken bottles in batch i and let p be the probability that a given bottle is broken.

(a) Assuming that each bottle breaks independently of the others, what is the distribution of X_1? Also, what is the distribution of $Y = X_1 + X_2 + \cdots + X_{100}$?

(b) Data are collected from 100 batches of bottles; the total number of broken bottles was 220. Using this data give an estimate and 95% CI for p.

(c) Which would be the more suitable approximation for X_1, a Normal approximation or a Poisson approximation?

20. Consider a normal distribution Y, with mean μ and variance σ^2, truncated so that only observations above some limit a are observed. In Example **??** we used the method of maximum likelihood to estimate μ and σ; in this exercise we use the method of moments.

Let $\mu_X = g_1(\mu, \sigma)$ and $\sigma_X^2 = g_2(\mu, \sigma)$ be the mean and variance of the truncated random variable X. That is,

$$
\begin{aligned}
\mu_X &= \int_a^\infty \frac{x f_Y(x)}{1 - F_Y(a)} dx \text{ and} \\
\sigma_X^2 &= \int_a^\infty \frac{(x - \mu_X)^2 f_Y(x)}{1 - F_Y(a)} dx,
\end{aligned}
$$

where the pdf and df of Y (f_Y and F_Y respectively), depend on μ and σ^2. Given μ and σ^2, μ_X and σ_X^2 can be calculated numerically.

If $X_1, \ldots, X_n$ is a sample from X, then you estimate μ and σ by solving

$$
\begin{aligned}
\hat{\mu}_X &= \overline{X} = g_1(\mu, \sigma), \\
\hat{\sigma}_X^2 &= S^2 = g_2(\mu, \sigma).
\end{aligned}
$$

Put $\theta = (\mu, \sigma)^T$, then this is equivalent to solving $g(\theta) = \theta$, where

$$
g(\theta) = \theta + A \begin{pmatrix} \overline{X} - g_1(\mu, \sigma) \\ S^2 - g_2(\mu, \sigma) \end{pmatrix},
$$

for any non-singular 2×2 matrix A.

One way to solve $g(\theta) = \theta$ is to find an A such that g is a contraction mapping (by trial and error), then use the fixed-point method (see Chapter **??**, Exercise 8). Test your method using the same sample used in Example **??**.

Chapter 18

1. Express 45 in binary.

   Now express 45 mod 16 and 45 mod 17 in binary.

   What can you say about these three binary representations?

2. Find all of the cycles of the following congruential generators. For each cycle identify which seeds X_0 lead to that cycle.

   (a) $X_{n+1} = 9X_n + 3 \mod 11$.

   (b) $X_{n+1} = 8X_n + 3 \mod 11$.

   (c) $X_{n+1} = 8X_n + 2 \mod 12$.

3. Here is some pseudo-code of an algorithm for generating a sample $y_1, \ldots, y_k$ from the population $x_1, \ldots, x_n$ *without* replacement ($k \leq n$):

```
for (i in 1:k) {
  { Select j at random from 1:(n+1-i) }
  y[i] <- x[j]
  { Swap x[j] and x[n+1-i] }
}
```

   Implement this algorithm in R. (The built-in implementation is `sample`.)

4. Consider the discrete random variable with pmf given by:

$$\mathbb{P}(X = 1) = 0.1, \quad \mathbb{P}(X = 2) = 0.3, \quad \mathbb{P}(X = 5) = 0.6.$$

   Plot the cdf for this random variable.

   Write a program to simulate a random variable with this distribution, using the built-in function `runif(1)`.

5. How would you simulate a negative binomial random variable from a sequence of Bernoulli trials? Write a function to do this in R. (The built-in implementation is `rnbinom(n, size, prob)`.)

6. For $X \sim \text{Poisson}(\lambda)$ let $F(x) = \mathbb{P}(X \leq x)$ and $p(x) = \mathbb{P}(X = x)$. Show that the probability function satisfies

$$p(x + 1) = \frac{\lambda}{x + 1} p(x).$$

   Using this write a function to calculate $p(0), p(1), \ldots, p(x)$ and $F(x) = p(0) + p(1) + \cdots + p(x)$.

   If $X \in \mathbb{Z}_+$ is a random variable and `F(x)` is a function that returns the cdf F of X, then you can simulate X using the following program:

```
F.rand <- function () {
  u <- runif(1)
  x <- 0
  while (F(x) < u) {
    x <- x + 1
  }
  return(x)
}
```

   In the case of the Poisson distribution, this program can be made more efficient by calculating F just once, instead of recalculating it every time you call the function `F(x)`. By using two new variables, `p.x` and `F.x` for $p(x)$ and $F(x)$ respectively, modify this program so that instead of using the function `F(x)` it updates `p.x` and `F.x` within the `while` loop. Your program should have the form

```
F.rand <- function(lambda) {
  u <- runif(1)
  x <- 0
  p.x <- ?
  F.x <- ?
  while (F.x < u) {
    x <- x + 1
    p.x <- ?
    F.x <- ?
  }
  return(x)
}
```

You should ensure that at the start of the `while` loop you always have `p.x` equal to $p(x)$ and `F.x` equal to $F(x)$.

7. This exercise asks you to verify the function `F.rand` from Exercise 6. The idea is to use `F.rand` to estimate the Poisson probability mass function, and compare the estimates with known values. Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) pois(λ) random variables, then we estimate $p_\lambda(x) = \mathbb{P}(X_1 = x)$ using

$$\hat{p}_\lambda(x) = \frac{|\{X_i = x\}|}{n}.$$

Write a program `F.rand.test(n, lambda)` that simulates n pois(λ) random variables and then calculates $\hat{p}_\lambda(x)$ for $x = 0, 1, \ldots, k$, for some chosen k. Have your program print a table giving $p_\lambda(x)$, $\hat{p}_\lambda(x)$ and a 95% confidence interval for $p_\lambda(x)$, for $x = 0, 1, \ldots, k$.

Finally, modify your program `F.rand.test` so that it also draws a graph of $\hat{p}$ and p, with confidence intervals, similar to Figure **??**.

8. Suppose that X takes on values in the countable set $\{\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots\}$, with probabilities $\{\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots\}$. Suppose also that you are given that $\sum_{i=0}^\infty p_i = p$, then write an algorithm for simulating X.

Hint: first decide whether or not $X \in \{a_0, a_1, \ldots\}$, which occurs with probability p.

9. Suppose that X and Y are independent rv's taking values in $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ and let $Z = X + Y$.

(a) Suppose that you are given functions `X.sim()` and `Y.sim()`, which simulate X and Y. Using these, write a function in R to estimate $\mathbb{P}(Z = z)$ for a given z.

(b) Suppose that instead of `X.sim()` and `Y.sim()` you are given `X.pmf(x)` and `Y.pmf(y)`, which calculate $\mathbb{P}(X = x)$ and $\mathbb{P}(Y = y)$ respectively. Using these, write a function `Z.pmf(z)` to calculate $\mathbb{P}(Z = z)$ for a given z.

(c) Given `Z.pmf(z)` write a function in R to calculate $\mathbb{E}Z$.

Note that we may have $\mathbb{P}(Z = z) > 0$ for all $z \geq 0$. To approximate $\mu = \mathbb{E}Z$ numerically we use $\mu_n^{trunc} = \sum_{z=0}^{n-1} z\mathbb{P}(Z = z) + n\mathbb{P}(Z \geq n) = n - \sum_{z=0}^{n-1}(n - z)\mathbb{P}(Z = z)$. How can we decide how large n needs to be to get a good approximation?

Do you think this method of approximating $\mathbb{E}Z$ is better or worse than simulation?

10. Consider the following program, which performs a simulation experiment. The function `X.sim()` simulates some random variable X, and we wish to estimate $\mathbb{E}X$.

```
# set.seed(7)
# seed position 1
mu <- rep(0, 6)
for (i in 1:6) {
  # set.seed(7)
```

```
    # seed position 2
    X <- rep(0, 1000)
    for (j in 1:1000) {
      # set.seed(7)
      # seed position 3
      X[j] <- X.sim()
    }
    mu[i] <- mean(X)
  }
  spread <- max(mu) - min(mu)
  mu.estimate <- mean(mu)
```

(a) What is the value of `spread` used for?

(b) If we uncomment the command `set.seed(7)` at seed position 3, then what is `spread`?

(c) If we uncomment the command `set.seed(7)` at seed position 2 (only), then what is `spread`?

(d) If we uncomment the command `set.seed(7)` at seed position 1 (only), then what is `spread`?

(e) At which position should we set the seed?

11. (a) Here is some code for simulating a discrete random variable Y. What is the probability mass function (pmf) of Y?

```
Y.sim <- function() {
  U <- runif(1)
  Y <- 1
  while (U > 1 - 1/(1+Y)) {
    Y <- Y + 1
  }
  return(Y)
}
```

Let N be the number of times you go around the while loop when `Y.sim()` is called. What is $\mathbb{E}N$ and thus what is the expected time taken for this function to run?

(b) Here is some code for simulating a discrete random variable Z. Show that Z has the same pmf as Y

```
Z.sim <- function() {
  Z <- ceiling(1/runif(1)) - 1
  return(Z)
}
```

Will this function be faster or slower that `Y.sim()`?

12. People arrive at a shoe store at random. Each person then looks at a random number of shoes before deciding which to buy.

(a) Let N be the number of people that arrive in an hour. Given that $\mathbb{E}N = 10$, what would be a good distribution for N?

(b) Customer i tries on X_i pairs of shoes they do not like before finding a pair they like and then purchase ($X_i \in \{0, 1, \ldots\}$). Suppose that the chance they like a given pair of shoes in 0.8, independently of the other shoes they have looked at. What is the distribution of X_i?

(c) Let Y be the total number of shoes that have been tried on, excluding those purchased. Supposing that each customer acts independently of other customers, give an expression for Y in terms of N and the X_i, then write functions for simulating N, X_i, and Y.

(d) What is $\mathbb{P}(Y = 0)$?

Use your simulation of Y to estimate $\mathbb{P}(Y = 0)$. If your confidence interval includes the true value, then you have some circumstantial evidence that your simulation is correct.

13. Consider the continuous random variable with pdf given by:

$$f(x) = \begin{cases} 2(x-1)^2 & \text{for } 1 < x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

Plot the cdf for this random variable.

Show how to simulate a rv with this cdf using the inversion method.

14. Consider the continuous random variable X with pdf given by:

$$f_X(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2} \quad -\infty < x < \infty.$$

X is said to have a standard logistic distribution. Find the cdf for this random variable. Show how to simulate a rv with this cdf using the inversion method.

15. Let $U \sim U(0,1)$ and let $Y = 1 - U$. Derive an expression for the cdf $F_Y(y)$ of Y in terms of the cdf of U and hence show that $Y \sim U(0,1)$.

16. For a given u, adapt the bisection method from Chapter ?? to write a program to find the root of the function $\Phi(x) - u$ where $\Phi(x)$ is the cdf of the standard normal distribution. (You can evaluate Φ using numerical integration or by using the built-in R function.) Notice that the root satisfies $x = \Phi^{-1}(u)$.

Using the inversion method, write a program to generate observations on a standard normal distribution. Compare the proportion of your observations that fall within the interval $(-1, 1)$ with the theoretical value of 68.3%.

17. The continuous random variable X has the following probability density function (pdf), for some positive constant c,

$$f(x) = \frac{3}{(1+x)^3} \text{ for } 0 \le x \le c.$$

   (a) Prove that $c = \sqrt{3} - 1$.
   (b) What is $\mathbb{E}X$? (Hint: $\mathbb{E}X = \mathbb{E}(X+1) - 1$.)
   (c) What is $\text{Var } X$? (Hint: start with $\mathbb{E}(X+1)^2$.)
   (d) Using the inversion method, write a function that simulates X.

18. The Cauchy distribution with parameter α has pdf

$$f_X(x) = \frac{\alpha}{\pi(\alpha^2 + x^2)} \quad -\infty < x < \infty.$$

Write a program to simulate from the Cauchy distribution using the inversion method.

Now consider using a Cauchy envelope to generate a standard normal random variable using the rejection method. Find the values for α and the scaling constant k that minimise the probability of rejection. Write an R program to implement the algorithm.

Chapter 19

1. Suppose that X and Y are iid $U(0,1)$ random variables.

   (a) What is $\mathbb{P}((X,Y) \in [a,b] \times [c,d])$ for $0 \le a \le b \le 1$ and $0 \le c \le d \le 1$?

      Based on your previous answer, what do you think you should get for $\mathbb{P}((X,Y) \in A)$, where A is an arbitrary subset of $[0,1] \times [0,1]$?

   (b) Let $A = \{(x,y) \in [0,1] \times [0,1] : x^2 + y^2 \le 1\}$. What is the area of A?

   (c) Define the rv Z by
   $$Z = \begin{cases} 1 & \text{if } X^2 + Y^2 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

      What is $\mathbb{E}Z$?

   (d) By simulating Z, write a program to estimate π.

2. Which is more accurate, the hit-and-miss method or the improved Monte-Carlo method? Suppose that $f : [0,1] \to [0,1]$ and we wish to estimate $I = \int_0^1 f(x)\,dx$.

   Using the hit-and-miss method, we obtain the estimate
   $$\hat{I}_{HM} = \frac{1}{n} \sum_{i=1}^n X_i,$$

   where $X_1, \ldots, X_n$ are an iid sample and $X_i \sim \text{binom}(1, I)$ (make sure you understand why this is the case).

   Using the improved Monte-Carlo method, we obtain the estimate
   $$\hat{I}_{MC} = \frac{1}{n} \sum_{i=1}^n f(U_i),$$

   where $U_1, \ldots, U_n$ are an iid sample of $U(0,1)$ random variables.

   The accuracy of the hit-and-miss method can be measured by the standard deviation of $\hat{I}_{HM}$, which is just $1/\sqrt{n}$ times the standard deviation of X_1. Similarly the accuracy of the basic Monte-Carlo method can be measured by the standard deviation of $\hat{I}_{MC}$, which is just $1/\sqrt{n}$ times the standard deviation of $f(U_1)$.

   Show that
   $$\text{Var}\, X_1 = \int_0^1 f(x)\,dx - \left(\int_0^1 f(x)\,dx\right)^2,$$

   and that
   $$\text{Var}\, f(U_1) = \int_0^1 f^2(x)\,dx - \left(\int_0^1 f(x)\,dx\right)^2.$$

   Explain why (in this case at least) the improved Monte-Carlo method is more accurate than the hit-and-miss method.

3. The previous exercise gave a theoretical comparison of the hit-and-miss and improved Monte-Carlo method. Can you verify this experimentally?

   Repeat the example of Section **??** using the improved Monte-Carlo method. How many function calls are required to get 2 decimal places accuracy?

4. The trapezoidal rule for approximating the integral $I = \int_0^1 f(x)\,dx$ can be broken into two steps

   Step 1: $I = \sum_{i=0}^{n-1}$ (Area under the curve from i/n to $(i+1)/n$);

   Step 2: Area under the curve from i/n to $(i+1)/n \approx \frac{1}{2}(f(i/n) + f((i+1)/n)) \times \frac{1}{n}$.

In two dimensions the integral $I = \int_0^1 \int_0^1 f(x, y) \, dx \, dy$ can be broken down as

$$\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \text{(Volume under the surface above the square}$$

$$[i/n, (i+1)/n] \times [j/n, (j+1)/n]).$$

(a) By analogy with the trapezoidal method, suggest a method for approximating the volume under the surface above the square $[i/n, (i+1)/n] \times [j/n, (j+1)/n]$, and thus a method for approximating the two-dimensional integral.

(b) Can you suggest a two-dimensional analogue for the improved Monte-Carlo algorithm?

Chapter 20

1. Write a program to calculate the Monte-Carlo integral of a function `ftn(x)`, using antithetic sampling, then use it to estimate

$$B(z, w) = \int_0^1 x^{z-1}(1-x)^{w-1}\,dx, \text{ for } z = 0.5, w = 2.$$

$B(z, w)$ is called the beta function, and is finite for all $z, w > 0$.

2. Suppose that X has a continuous cdf F, and that F^{-1} is known. Let $U_1, \ldots, U_n$ be iid $U(0, 1)$ rv's and put $X_i = F^{-1}(U_i)$, then we can estimate $\mu = \mathbb{E}X$ and $\sigma^2 = \text{Var}\,X$ using $\overline{X} = n^{-1}\sum_i X_i$ and $S^2 = (n-1)^{-1}\sum_i(X_i - \overline{X})^2$, respectively.

   (a) Show that if $U \sim U(0, 1)$, then

   $$\text{Cov}(F^{-1}(U), F^{-1}(1 - U)) \le 0.$$

   (b) Show how to use antithetic sampling to improve our estimate of μ.

   (c) Suppose that the distribution of X is symmetric about μ. Show that antithetic sampling will *not* improve our estimate of σ^2.

3. Consider the integral

$$I = \int_0^1 \sqrt{1 - x^2}\,dx.$$

   (a) Estimate I using Monte-Carlo integration.

   (b) Estimate I using antithetic sampling, and compute an estimate of the percentage variance reduction achieved by using the antithetic approach.

   (c) Approximate the integrand by a straight line and use a control variate approach to estimate the value of the integral. Estimate the resulting variance reduction achieved.

   (d) Use importance sampling to estimate the integral I. Try using three different importance sampling densities, and compare their effectiveness.

4. Suppose that the rv X has mean μ and can be simulated. Further, suppose that f is a non-linear function, and that we wish to estimate $a = \mathbb{E}f(X)$ using simulation.

   Using $g(x) = f(\mu) + (x - \mu)f'(\mu)$ and tuning parameter $\alpha = 1$, estimate a using control variates. That is, if $X_1, \ldots, X_n$ are an iid sample distributed as X, show that for $\alpha = 1$, the controlled estimate of a is

$$\frac{1}{n}\sum_{i=1}^n f(X_i) - (\overline{X} - \mu)f'(\mu).$$

   Furthermore, using the fact that for x close to μ, $g(x) \approx f(x)$, show that the controlled estimate can be written approximately as

$$\frac{1}{n}\sum_{i=1}^n f(X_i) - f(\overline{X}) + f(\mu).$$

   Finally, derive the optimal (theoretical) value of α.

5. Daily demand for a newspaper is approximately gamma distributed, with mean 10,000 and variance 1,000,000. The newspaper prints and distributes 11,000 copies each day. The profit on each newspaper sold is $1, and the loss on each unsold newspaper is $0.25.

   (a) Express the expected daily profit as a finite integral, then estimate it using both Simpson's method and Monte-Carlo integration.

(b) Improve your Monte-Carlo estimate using importance sampling and/or a control variate.

(c) For m integer valued, a $\Gamma(\lambda, m)$ rv can be written as the sum of m iid $\exp(\lambda)$ rv's. Using this approach to simulate gamma rv's, estimate the expected daily profit using antithetic sampling.

6. Consider estimating $I = \int_0^1 g(x)dx$ by improved Monte-Carlo integration. We showed in Section **??**, that using antithetic variates is equivalent to replacing g by $h(x) = (g(x) + g(1-x))/2$, which averages g with its mirror image around $x = 1/2$. Further variance reduction may be possible by iterating this process on subintervals, as illustrated below.

(a) Let $g(x) = x^4$. Write an R program to calculate the improved Monte-Carlo estimator $\hat{I}$ of I, and to estimate its variance.

(b) Repeat (a) using antithetic variates, and compute the variance reduction achieved.

(c) Using the fact that $h(x) = h(1-x)$, verify that

$$I = \int_0^{1/2} (g(x) + g(1-x))dx.$$

Then verify that over this subinterval you can again replace the integrand by a function which averages its value with the value of its mirror image around $x = 1/4$. Hence verify that

$$I = \int_0^{1/4} (g(x) + g(1-x) + g((1/2) - x) + g((1/2) + x))dx$$

Use this to estimate I and calculate the resulting variance reduction.