

Additional (advanced) exercises

Unlike the exercises in the notes, whose solutions are provided in Chapter 19, you must attempt these questions yourself before I will share my answer with you. You can do so during the live sessions, either in person or via Zoom. Try, struggle, ask questions, and learn!

1 Brownian motion

Write a function that simulates a random walk:

1. From a starting position of $x = 0$ and $y = 0$, move a virtual particle by a distance of 1 in a random direction.
2. Repeat $n = 1000$ times and plot the track of the particle as a line.
3. Repeat steps 1 and 2 for $N = 500$ virtual particles and visualise their final positions on a scatter plot.

2 Diffusion

Using the code from exercise 1:

1. Repeat step 1.3 for $n = 250, 500, 1000$ and 2000 iterations and visualise the final positions on a 2×2 panel grid of scatter plots. Adjust the axis limits so that all four panels are plotted at the same scale.
2. Plot the marginal distributions of the x -values as kernel density estimates and empirical cumulative distribution functions.
3. Visualise the bivariate datasets as 2-dimensional KDEs.

3 Summary statistics

Using the code from the previous exercises:

1. Construct a table with the means, medians, standard deviations and interquartile ranges of the synthetic datasets (x -values) of exercise 2.
2. Plot the four datasets as box plots.

4 Probability

Suppose that a password needs to contain exactly 8 characters and must include at least one lowercase letter, uppercase letter and digit. How many such passwords are possible?

5 Bernoulli variables

1. Draw a random number from a uniform distribution between 0 and 1 and store this number in a variable (p , say).
2. draw $n = 20$ additional numbers from the same uniform distribution and count how many of these values are less than p .

3. Repeat steps 1 and 2 $N = 100$ times, store the results in a vector (x , say), and plot the results as a histogram.
4. Suppose that you did not know the value of p , then you could estimate it (as a new variable \hat{p} , say) from the data x . To this end, compute the binomial density (or ‘likelihood’) of all values in x for $\hat{p} = 0.5$ and sample size n .
5. Now take the sum of the logarithms of the binomial likelihoods of x . Call this sum LL (for ‘log-likelihood’).
6. Repeat step 5 for a regularly spaced sequence of \hat{p} -values. Then plot the resulting LL -values against those \hat{p} -values.
7. Approximately which value of \hat{p} corresponds to the maximum value for LL ? How does this compare to the true value of p ?

By completing this exercise, you have numerically extended the procedure described in Section 5.1 of the notes.

6 Type-2 errors

1. Draw a random number from a binomial distribution with $n = 20$ and $p = 0.52$.
2. Apply a binomial test comparing $H_0 : p = 0.5$ with $H_a : p \neq 0.5$. Do the data pass the test?
3. Repeat steps 1 and 2 $N = 1000$ times. What percentage of the datasets pass the test?
4. Repeat steps 1 through 3 for $n = 200$.
5. Repeat step 4 for a range of values from $n = 20$ to $n = 2000$. Plot the probability of rejection against n .
6. Compare the results of step 5 with a manual calculation of the probability of committing a Type-2 error as a function of sample size.

7 Confidence intervals

1. Draw $N = 10$ random numbers from a binomial distribution with $p = 0.55$ and $n = 20$. Construct 95% confidence intervals for p for each of these numbers.
2. Plot these confidence intervals against n as error bars using R’s `arrows()` function.
3. How many of the confidence intervals in step 2 overlap with $p_0 = 0.5$?
4. Repeat step 2 for $n = 200$. How many of these new confidence intervals fall outside $p = 0.5$? Do so again for $n = 2000$.

8 Random sampling in 2D

1. Generate two vectors (x and y , say) of 1000 random numbers between 0 and 500 and visualise them on a scatter plot. Add a grid of lines at every 20 units of x and y .
2. Count the number of items falling in each square bin of the graticule contained in the interval from $x = 0$ to $x = 100$ and from $y = 0$ to $y = 100$. Plot these numbers as a histogram.
3. Calculate the mean and variance of the resulting dataset of counts. Repeat steps 1 and 2 several times before drawing conclusions.