Problem sheet 3

Quantitative Methods

Preparation: Watch the following three (short) video tutorials on random variables, discrete and continuous random variables, and probability density functions.

A. In R, the function `runif()` simulates a continuous random variable X that is said to follow a *Uniform distribution* on the interval $(0, 1)$. The function `dunif(x)` calculates the probability density of X at `x`.

1. By changing the input `n` of `runif(n)`, draw 100, 1,000, and 100,000 numbers from the Uniform distribution. Draw a histogram of each output using the `hist()` function. What do you notice as `n` increases? What does this tell you about the Uniform distribution?

2. Use `dunif()` to calculate the density of X for all numbers between -1 and 2, incremented by 0.01 (i.e., create a sequence of numbers between -1 and 2 that are incremented by 0.01 and calculate the density for each number in the sequence). Visualise your output using either `plot()` or `ggplot()`.

3. What is $\mathbb{P}(X = 0.37629)$?

B. During an election, you go from door to door in your local community to encourage your neighbours to vote in the election. Let X be the random variable indicating whether a randomly selected neighbour decides to vote or not, taking the value 1 if they vote and 0 otherwise. Due to your eloquence and charisma, the probability that you will convince a neighbour to vote is 0.7. X is said to follow a *Bernoulli distribution*, which has one parameter $p = \mathbb{P}(X = 1)$.

1. Imagine that you are able to speak to 100 people in your community over the course of the period leading up to the election. Using the `sample()` function in R, estimate the total number of people who end up voting after speaking to you and plot the result.

2. In R, the function `rbinom()` simulates a random variable with a *Binomial distribution* and it takes three inputs: the number of random variables we want to generate (in our example, the 100 people you speak to before the election), the number of trials per random variable (in our case, you only speak to each person once), and the probability of 'success' in any single trial ($p = 0.7$). By setting the three inputs to 100, 1, and 0.7, respectively, compare the output(s) of the `rbinom()` function to your answer to the previous question. What is the relation between a Bernoulli and Binomial random variable?

3. What is the substantive difference between `rbinom(100, 1, 0.7)` and `rbinom(1, 100, 0.7)`? Give your answer both in terms of probability theory and in relation to the empirical example given above.

C. A *Poisson distribution* expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate λ and independently of the time since the last event. Let X be the number of global pandemics taking place over a one-year period, with on average one pandemic occurring every year. Then X follows a Poisson distribution with a mean rate parameter $\lambda = 1$. In R, the function `dpois(x, lambda)` calculates the probability density of X at `x` for a chosen parameter value `lambda`.

1. For the example given above (and assuming independence of events), what is the probability of 4 global pandemics occurring in a given year?
2. Calculate the density of X for all numbers between 0 and 20, incremented by 1. Visualise your results.
3. Repeat the previous exercise for `lambda` = 3, 7, and 10. What do you notice?