Instructions

- Please upload **both** *.Rmd* and compiled *html* files through MyCourses. Save them as Lastname_Firstname_A2.Rmd and Lastname_Firstname_A2.html
- All calculations must be done in `R`. Your document should be fully reproducible. Text, code and ouput must be shown in the compiled document.
- Be sure to put meaningful axis labels for all your plots using the arguments `xlab, ylab, main` in the `plot` function.
- Bonus questions are optional

Q1 You are interested in placing a bet that McGill will win its next soccer match. You talk to your local bookie and she tells you that if the outcome of the random experiment is:

- {McGill Wins} she will give you $100.
- {McGill Loses} you have to give her $110.
- {McGill Ties} she will give you $300.

You then talk to your local statistician and she tells you the probability of each outcome of the random experiment.

- $Pr(McGill\ Wins) = .55$
- $Pr(McGill\ Loses) = .44$
- $Pr(McGill\ Ties) = .01$

(a) Define the random variable for this bet.

(b) Plot the probability mass function. Label the axes appropriately.

(c) Plot the cumulative distribution function. Label the axes appropriately.

(d) On average how much to you expect to make from this bet?

(e) What is the standard deviation (σ) and variance (σ^2) of this random variable?

Q2 You are going on a long flight and pack your I-pod which has 10 songs on it. One of which is favorite song. You put your I-pod on “Shuffle” which means it will pick each song with the same probability and independently. Unfortunately you forgot to charge your I-pod so you only have enough battery life to listen to 30 songs. Define a “success” to be your favorite song being played.

(a) You are interested in modeling the total number of “successes” that will occur during your trip. What type of random variable would be appropriate to model this random experiment? What are its parameters?

(b) Plot the PMF and CDF of this random variable.

(c) Simulate 100 draws from this random variable, and plot a histogram of these draws using the `arm::discrete.hist` function. How does it compare to true PMF in (b)? (`arm` is a package that you need to install using the `R` command `install.packages("arm")`, and `discrete.hist` is a function from the `arm` package)

(d) Plot the empirical cumulative distribution function on the same plot as the true CDF (see the `ecdf` function). Are they different? Explain.

(d) Repeat parts (c) and (d) but now simulate 10,000 draws. Compare the true PMF and CDF to their empirical counterparts. Do you notice any difference between the results from 100 draws, and from 10,000 draws? Explain.

Q3 Calculate the following probabilities.

(a) Probability that a normal random variable with mean 22 and variance 25

  (i) lies between 16.2 and 27.5

  (ii) is greater than 29

  (iii) is less than 17

  (iv) is less than 15 or greater than 25

(b) Probability that in 60 tosses of a fair coin the head comes up

  (i) 20,25 or 30 times

  (ii) less than 20 times

  (iii) between 20 and 30 times

(c) A random variable X has Poisson distribution with mean 7. Find the probability that

  (i) X is less than or equal to 5

  (ii) less than 4

  (iii) X is greater than 10 (strictly)

  (iv) X is between 4 and 16

Q4 Consider the following three random variables and their distributions: $X \sim Exponential(10)$, $Y \sim Exponential(5)$ and $Z \sim Exponential(1)$

(a) Plot, on the same graph, the probability density function (PDF) of X, Y and Z. Each PDF should be a different color. Add a legend. Add meaningful axis labels.

(b) Plot, on the same graph, the cumulative distribution function (CDF) of X, Y and Z. Each CDF should be a different color. Add a legend. Add meaningful axis labels.

(c) What do you notice when comparing the three CDFs? Comment on any patterns you see and explain why this is the case. (think about what the parameter of the exponential distribution means)

(c) Find the expected value and variance for X, Y and Z.

(d) Simulate 10,000 samples from X, where each sample consists of 15 random draws from $X \sim Exponential(10)$. For each sample, calculate the mean. Plot a histogram and a boxplot of these means.

(e) Calculate the mean and standard deviation of the sample means from (d).

Q5 Simulation of a poisson approximation of binomial

(a) Write and run `R` code with various n and p to see how the errors compare as n increases and p decreases, by calculating actual binomial probabilities as well as Poisson probabilities with $\lambda = np$. The error here would be defined as the difference in probabilities between the binomial and poisson for a given value of n and p.

(b) Plot the error as a function of np. What do you notice?

Q6 You roll a die 100 times and get just 10 sixes - Is the die fair?

(a) What is the probability of getting just 10 sixes?

(b) What is the probability of getting 10 or fewer sixes?

(c) Plot the PMF and CDF

(d) Simulate the described experiment 1000 times and compute the CDF. Compare it to the theoretical one.

(e) Repeat the exercise with 1,000,000 simulations. What do you notice?

Q7 Moment Generating functions ($M_X(t)$)

(a) Write your own function in `R` which takes as input the variables t, n, p and outputs the value of the moment generating function of the binomial distribution. Plot $M_X(t)$ vs. t for $n = 25$ and $p = 0.25$, over a range of t.

(b) Write your own function in `R` which takes as input the variables t, n, p and outputs the value of $M_X^{(1)}(t)$ of the binomial distribution.

(c) Use the function in (b) to calculate the expected value. Compare it with the expected value formula for the binomial distribution.

(d) (Bonus) Use the `stats::deriv` function in `R` to calculate $M_X^{(1)}(t)$. Use this to calculate the expected value and show that its equal to the expected value formula for the binomial distribution