

# Stat0009 ICA1 (2023)

## Introduction

- Made available UCL Week 10, 3 Nov 2023.
- Due 17 November 2023, 13:00 London time.
- You are to work by yourself.
- This ICA consists of 5 questions, worth 35 points;
  - another 5 points will be given based on presentation, so that the total available points is **40**
  - You **must** hand in the html file **and** the Rmd/Qmd/ipynb source file.
- This ICA is worth 30 percent of your grade for this module.
- Some parts of the questions will require you to write and run R/Python code.
- The ICA must be completed in R Markdown/Quarto/Jupyter Notebook and typeset using Markdown with Latex code, just like the way our module content is generated. You can choose to use either R or Python.
- As usual, part of the grading will depend on the clarity and presentation of your solutions.
- You are to do this assignment by yourself, without any help from others.
- You are allowed to use any materials and code that was presented so far.
- Do not search the internet for answers to the ICA.
- Do not use ChatGPT or similar AI type assistance.
- Do not use any fancy code or packages imported from elsewhere.
- Conduct yourselves honorably.

## Plagiarism and collusion

Please familiarize yourself with the following excerpt on plagiarism and collusion from the student handbook (<https://tsoo-math.github.io/ucl/Plagiarism-Collusion.pdf>)

By ticking the submission declaration box in Moodle you are agreeing to the following declaration:

**Declaration:** I am aware of the UCL Statistical Science Department's regulations on plagiarism for assessed coursework. I have read the guidelines in the student handbook and understand what constitutes plagiarism. I hereby affirm that the work I am submitting for this in-course assessment is entirely my own.

## Anonymous Marking

Please do **not** write your name anywhere on the submission. Please include only your **student number** as the proxy identifier.

## Questions

### 1.) The cooked books [10 points]

The number of arrivals to boutique shop each day can be modeled with a Poisson random variable with mean  $\lambda > 0$ , which we hope to estimate. We are given a data set  $x_1, \dots, x_{300}$ , but we know that the owner has cooked (<https://www.investopedia.com/terms/c/cookthebooks.asp#:~:text=Cook%20the%20books%20is%20a,up%20its%20earnings%20or%20profit.>) the books, and will falsify the data so that  $x_i \neq 0$ ; specifically, they will record a entry of 1, whenever no customers arrived, but will faithfully enter records otherwise.

- With the cooked data set, how would you estimate  $\lambda$ ? Justify your answer, and explain why your estimator is reasonable. [4 points]
- Test your method, in the case and  $\lambda = 1$ ; that is, generate say  $N = 100$  samples of cooked data, and see how close (on average) your estimates are to the true  $\lambda = 1$ . [3 points]

- Now, use your method to estimate  $\lambda$  for the given data (<https://tsoo-math.github.io/ucl4/cook.csv>). [3 points]

## 2.) The half parabola [5 points]

Consider the region  $R$  bounded by the  $x$ -axis and the function  $f(x) = \frac{3}{2}(1 - x^2)$ , for  $x \in [0, 1]$ .

- Write code that samples a point  $U$ , that uniformly at random on  $R$ . [2 points]
  - Plot of a probability histogram of the location of the  $x$ -coordinates of the sampled points. [1 points]
- Let  $X$  be the  $x$ -coordinate for  $U$ . What is the pdf for  $X$ ? Explain. [1 point]
- If  $Y$  is the  $y$ -coordinate for  $U$ , then are  $X$  and  $Y$  independent? Explain. [1 point]

## 3.) The Casino [10 marks]

An inexperienced gambler starts with blackjack, betting five pounds, so only wins (5 pounds) with probability 0.43; if they win, they bet again, and if they lose, they switch to roulette, where they always bet (1 pound) on 0, so that they have a  $1/(36 + 1)$  chance of winning (35 pounds); again they switch back to blackjack if they lose, and stay on roulette if they win.

- Model the gambler with a Markov chain. Write down an explicit transition matrix. [5 points]
  - Simulate the Markov chain, for  $N = 100000$  steps/bets
  - Compute,  $S_N$ , their winning (possibly negative) after bet  $N$ .
- Use theory to estimate the average  $S_N/N$ , and compare with your simulation above. [5 points]

## 4.) The Poisson collector [5 points]

Consider a Poisson point process on the real line of unit intensity, which we think of as dust. We start at the origin, and are interested in vacuuming, and we do it in a greedy way, so that we always go to the closest piece of dust. For background (not necessary for the completion nor understanding of this question) see: On the greedy random walk (<https://link.springer.com/article/10.1007/s11134-011-9246-x>).

- By running simulations, estimate the probability that, initially, you will head in the positive direction twice, successively.

## 5.) Coupling [5 points]

Let  $\epsilon > 0$ . Suppose that  $X$  is a Poisson random variable of mean  $\lambda > 0$ , and  $Y_\epsilon$  is a Poisson random variable of mean  $\lambda + \epsilon$ . Use the power of coupling to upper bound the total variational distance of  $X$  and  $Y$  as a function of  $\epsilon$ , and show that the distance converges to zero, as  $\epsilon \rightarrow 0$ .

## Endnotes

- Q5 is an old-school maths-type exercise, with no coding.
- Q4: this is a Poisson point process on the entire real line  $\mathbb{R} = (-\infty, \infty)$ , which can be constructed by joining two independent Poisson point processes on the half line  $[0, \infty)$ , at the origin.
- Q1: If necessary, you are free to use numerical solvers in R; that is, you do not need to code Newton's method yourself, but you are welcome to.
- I will create a submission portal on Moodle soon.
- source (<https://tsoo-math.github.io/ucl4/2023-ica1-stat9-release.Rmd>)
- Version: 11 November 2023