Assignment 1

(MA6.102) Probability and Random Processes, Monsoon 2023

Date: 7 August 2023, Due on 14 August 2023 (Monday).

INSTRUCTIONS

- Discussions with other students are not discouraged. However, all write-ups must be done individually with your own solutions.
- Any plagiarism when caught will be heavily penalized.
- Be clear and precise in your writing. Also, clearly state the assumptions made (if any) that are not specified in the question.
- Coding portion can be done in Python/Matlab. There will be a moss check, code copying will result in a straight 0.
- Submit a zipped folder (rollnumber.zip) containing your handwritten solutions(pdf), code and pdf of plots.

Problem 1 (10 Marks)**.** Let A and B be sets and $pow(S)$ denote the power set of a set S.
(a) Prove that

$$pow(A \bigcap B) = pow(A) \bigcap pow(B).$$

(b) Prove that

$$pow(A) \bigcup pow(B) \subseteq pow(A \bigcup B)$$

with equality holding if and only if one of A or B is a subset of the other.

Problem 2 (10 Marks)**.** Let A and B be finite sets such that $A \subseteq B$. Then, compute the value of the expression:

$$\sum_{\forall C: A \subseteq C \subseteq B} (-1)^{|C \setminus A|}$$

(where $C \setminus A = \{x \in C : x \notin A\}$ and $|C \setminus A|$ represents cardinality of set $C \setminus A$).

Problem 3 (10 Marks)**.** If A and B are independent events, then does it imply that A^c and B^c are independent events as well. If it's true give a proof for the same, or give a counterexample to disprove it. Also, is it true that A^c and B are independent events?

Problem 4 (10 Marks)**.** One of the largest sea turtles, the Green Sea Turtle are found in warm tropical to subtropical waters. There are N Green sea turtles, from which a random sample of size m are captured and tagged. Random sample means that all ${}^{N}C_m$ sets of m Green sea turtles are equally likely. The captured Green sea turtles are returned to the population, and then a new sample is drawn, this time with size n. This is an important method that is widely used in ecology, known as capture-recapture. What is the probability that exactly k of the n Green sea turtles in the new sample were previously tagged?
(Assume that a Green sea turtle that was previously captured has the same probability of being captured again and $k \leq m, n$)

Problem 5 (10 Marks)**.** a) (Birthday Paradox). There are k people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (we exclude February 29), and that people's birthdays are independent (we assume there are no twins in the room). What is the probability that two or more people in the group have the same birthday? For what minimum k is the probability greater than $\frac{1}{2}$?

b) (Modified Birthday Paradox) In the previous part, we assumed that birthdays on all the 365 days of the year are equally likely. In reality, some days are slightly more likely than others. Let $p_1, p_2, \ldots, p_{365}$ be the birthday probabilities, where p_i is the probability of having birthday on the i^{th} day of the year (February 29 is still excluded, no offense). Let $k \geq 2$ be the number of people.
(i) Find a simple expression for the probability that there is at least one birthday match.
(ii) Explain why it makes sense that P(at least one birthday match) is minimized when
$p_i = \frac{1}{365}, \forall i$.

Problem 6 (10 Marks)**.** A train coach with 50 seats (each seat having been alloted to a unique passenger) is waiting at the platform. Those 50 passengers are lined up in queue. Crazy enough, the first person in queue chooses to idly occupy one of the chairs (all seats are equally likely). Each subsequent traveller takes their assigned seat, if it's available, or otherwise occupies one of the available seats randomly. What is the probability that the last person standing in the queue gets to sit on his assigned seat?

Problem 7 (10 Marks)**.** This problem will provide intuition to why continuous probabilistic models are useful. In this problem we will approximate π. One method to estimate the value of π (3.141592...) is by using a Monte Carlo method. This methods consists of drawing on a canvas a square with an inner circle. We then generate a large number of random points within the square and count how many fall in the enclosed circle. The probability that that a random point lies in the circle is equal to $\frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}$ Simulate n random points to get the experimental probability and multiply it by 4 to get an estimate of π. Plot estimated value of π versus n for $1 \leq n \leq 100000$.

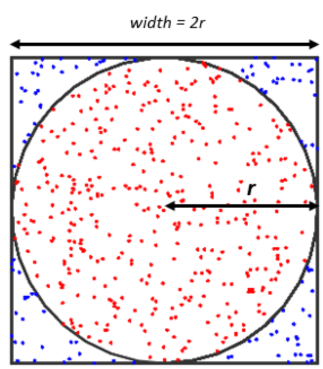
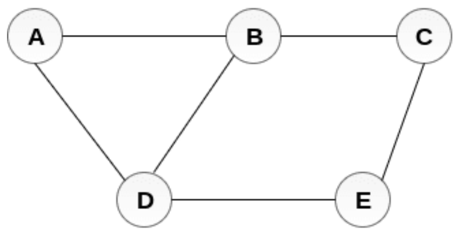


Fig. 1. Monte Carlo simulation

Problem 8 (10 Marks)**.** i) When you generate a random number in any Programming language, it usually returns a value between 0 and 1. The random number generated has an equal probability of lying anywhere between 0 and 1. Now use this random number generated in [0, 1] to simulate a coin flip with a 0.3 probability of being heads and 0.7 of being tails. Plot no. of heads vs no. of tosses (n) for $1 \leq n \leq 100000$.

ii) This problem will use probability to rank websites as done by Google search. Consider the undirected graph G shown in the figure where each of the nodes represents a website and each edge represents a hyperlink between two websites (Here we assume that if website A has hyperlink to website B, then B also has hyperlink to A to simplify code). Simulate a random walk with 20 steps (You start at a random node and in each step you randomly select one of the edges that has the current vertex and move to the other vertex of the selected edge, a collection of such steps constitute a walk) and maintain a counter for how many times you visited each node. After performing $n = 1000$ walks, rank the nodes (websites) in descending order of the amount of times they have been visited.



Undirected Graph

$$\begin{array}{c} \\ \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \\ \mathbf{E} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Adjacency Matrix

Fig. 2. Graph G