BMIG 5003 Computational Methods for Biomedical Informatics
Horacio Gómez-Acevedo, PhD
Fall 2023
Assignment #5
Due day: Nov 21st, 12:00 pm CST

Submission:
All your answers should be placed in the Assigned Box Folder. I will accept only Python and text files, NOT jupyter notebooks. Feel free to use any following modules: `numpy, pandas, plotly, scikit-learn, statsmodels, regex, bnlearn, and random`.

1. A survey was conducted in a school about the favorite subject including Math (M), English (E), and Science (S). Twenty percent of students were in the 1st grade, 50% in the 6th grade and 30% in the 11th grade. Their preferences are tabulated in Table 1. Let's denote by θ the (random) variable grade, and by S the subject.

	M	E	S
1st Grade	0.3	0.6	0.1
6th Grade	0.1	0.3	0.6
11th Grade	0.3	0.6	0.1

Table 1: Favorite Subjects per grade distribution

   (a) What is the prior distribution $P(\theta)$?
   (b) Using Bayes' theorem. Calculate $P(\theta = \text{"6th grade"} | S = \text{"}M\text{"})$.
   (c) Implement a Python program in which the user provides the prior distribution, and returns the posterior probability $P(\theta = \text{"11th Grade"} | S = s)$ for each value of s.

2. Consider the Markov chain model depicted in figure 1.

   (a) Find out what is the missing value.
   (b) Write down the transition probability matrix (you may consider writing it down in Python as a list of lists).
   (c) If the initial probability distribution of the states is $\boldsymbol{\pi}(0) = (0.3, 0.2, 0.1, 0.3, 0.1)$, what is the probability distribution of the states after the Markov model runs 15 steps.

3. Using the `survey.txt` file split the rows into two files (say `survey1.txt` and `survey2.txt`) with around 50% of the total rows each following this strategy:

   Strategy: Randomly select (with equal probability) which row goes into each file.

   Use **bnlearn** to answer the following questions

   (a) Find the best DAG structure that fits `survey`i`.txt` ($i = 1, 2$).
   (b) Print out each DAG. Describe the differences (if any)

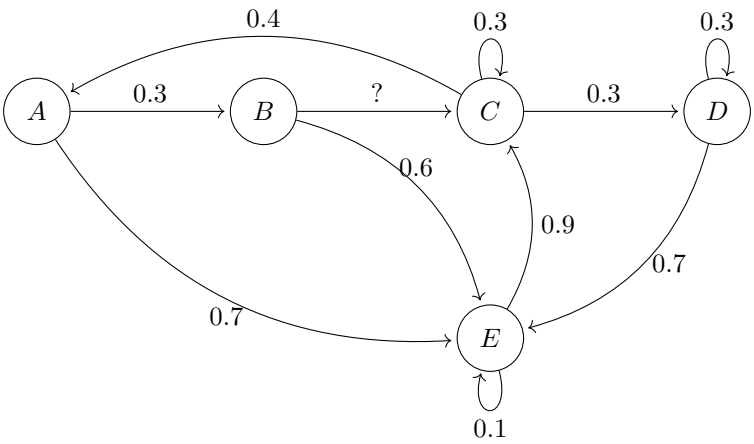

Figure 1: Markov Chain