Due: Monday 17 May, 10pm. Submit via Canvas

ENGSCI760 2021 – Assignment 2

Joint Distributions (5 marks)

1. Suppose we independently roll two fair 6-sided dice (#1 and #2), and define three events as follows: let A be the event that the sum of the two dice is 7; let B be the event that die #1 is 2 or less; let C be the event that the die #2 is odd.
   a) Are the events A, B and C pairwise independent? Explain why / why not?
   b) Are the events A, B and C mutually independent? Explain why / why not?

Markov Chains (10 marks)

2. A car dealership, which operates only during weekends, is selling the latest model electric vehicle: the Ohm. They currently have space to stock up to four Ohms on their lot. The dealership can buy these cars from the head office at a price of $45,000 per vehicle, and they sell them to the consumer for $50,000 per vehicle. They estimate that demand for Ohms, each weekend, will follow the distribution in the table below. (As is independent between weekends.)

Demand	Probability
0	0.1
1	0.3
2	0.3
3	0.2
4 or more	0.1

At the end of each weekend, the manager of the dealership must decide whether to place an order for up to four vehicles, which will be delivered by the following Saturday, with a fixed delivery fee of $6,000, regardless of the number ordered. There is also an insurance cost for storing vehicles in the lot over the week, which is $100 per vehicle per week. Currently the manager only places an order for Ohms if they have none remaining in stock, ordering 4 Ohms.

   a) What is the one-step Markov transition matrix for the number of Ohms in stock at the end of each weekend.
   b) What is the limiting distribution for the number of Ohms in stock at the end of each weekend. (You don't need to work this out by hand, but write down the equations, and provide a screenshot of your implementation.)
   c) Suppose at the end of a given weekend there are 2 Ohms in stock. Given the current reorder policy, what's the expected number of weeks before there are 2 Ohms in stock at the end the weekend again? (You don't need to work this out by hand, but write down the equations, and provide a screenshot of your implementation.)
   d) Given the current reorder policy, compute the expected weekly profit the dealership would make in steady-state.
   e) Propose a better reorder policy and demonstrate that its steady-state weekly profit is higher than the current policy.

Hidden Markov Model (15 Marks)

3. You have a friend who, despite his fat fingers, refuses to turn on autocorrect on his phone. You received the following txt messages from him one evening.
   - `cljlx ypi ktxwf a pwfi psti vgicien aabdwucg vpd me and vtiex voe zoicw`
   - `qe qzby yii tl gp tp yhr cpozwdt fwstqurzby`
   - `qee ypi xfjvkjv ygetw ib ulur vae`
   - `wgrrr zrw uiu`
   - `hpq fzr qee ypi vrpm grfw`
   - `qe zfr xtztvkmh`
   - `wgzf tjmr will uiu xjoq jp ywfw`

   Develop a hidden Markov model using the King James Bible to determine what he was trying to say.

   a) Using the file 'bible.txt' and function **createTransitions** in 'HMM.py' available on Canvas create a transition matrix and a vector of prior probabilities for lowercase letters (do not worry about where in the word the letter is observed when constructing the prior distribution).
   b) Using the function **createEmissions** and the keyboard adjacency matrix provided in **main**, create a matrix of emission probabilities (assume that there is a 0.5 probability of your friend hitting the correct letter, and the probabilities of hitting any adjacent letters are equal.
   c) Using the function **HMM** in 'HMM.py' implement the Viterbi algorithm to determine the most likely sequence of states **x** given an observed vector **y**. Utilise the prior distribution, transition probabilities and emission probabilities found in (a) and (b).
   d) Using the function **main**, determine what was intended to be typed.

   For questions (a), (b) and (c) above, detail and describe (using comments in the code) the changes you made to the code. Submit your final commented code.

Bayesian Network (10 Marks)

4. Anadarko is investigating the possibility of drilling for oil off New Zealand's coast. They are particularly interested in one site, but have come to you for advice on the probabilities of finding oil.

   In order to get a better understanding of the oil resource, Anadarko can perform two tests:
   - a seismic study can be carried out to determine the geological structure of the area;
   - a test well can be drilled.

   There are three types of geological structures (A,B,C) that are possible, with a prior distribution of:

$$f_G(g) = \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ [0.2 \quad 0.4 \quad 0.4] \end{array}$$

   From historic data, Anadarko estimates the probability of no (N), low (L), or high (H) oil, given the geological structure to be:

$$f_{O|G=g}(o) = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{N} & \mathrm{L} & \mathrm{H} \end{array} \\ \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.5 & 0.3 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \end{array}$$

   Also given the geological structure, the probability of the seismic test returning a positive/negative (+/-) result is given in the table below:

$$f_{S|G=g}(s) = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{cc} + & - \end{array} \\ \begin{bmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \\ 0.8 & 0.2 \end{bmatrix} \end{array}$$

   Finally, the likelihood that the test well will return a positive/negative (+/-) result depends on the amount of oil present, as given below:

$$f_{T|O=o}(t) = \begin{array}{c} \\ \mathrm{N} \\ \mathrm{L} \\ \mathrm{H} \end{array} \begin{array}{c} \begin{array}{cc} + & - \end{array} \\ \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} \end{array}$$

   (Note that the positive / negative result for the tests do not directly indicate the presence of oil, they are simple outcomes for the test variables.)

   a) Draw a Bayesian network for the situation given above.
   b) Explain the concept of conditional independence. Use as an example the relationship between the oil outcome and the seismic test, if you already know the geology.
   c) Utilise 'belief_propagation.py' or 'belief_propagation.m' to determine the probability distribution for finding oil, given the following test results:
      i. Seismic test is positive; test well not drilled.
      ii. Seismic test is negative; test well is positive.
      iii. Test well is negative; seismic test not performed.