Engsci 760 Decision-Making and Modelling Under Uncertainty

Problem Set 2

Oliver Maclaren
oliver.maclaren@auckland.ac.nz

Lecture 5: Representation and independence for probability models

Problem 1

Consider the undirected graph

$$
\begin{tikzcd}
A \arrow[r, dash] & B \arrow[r, dash] \arrow[d, dash] & C \\
& D &
\end{tikzcd}
$$

- Write down a (u-) separation graph property implied by the above graph corresponding to a single node separating two other nodes.
- Write down a (u-) separation graph property implied by the above graph corresponding to a single node separating sets of nodes (one set may be a single-node set but at least one should contain two nodes)
- Write down the corresponding (probabilistic) conditional independence relations for the above graph properties using $\perp\!\!\!\perp$ notation
- Write down the corresponding conditional probability function properties for the above graph properties.

Problem 2

Draw an undirected graph in which a set of at least two nodes separates two other sets, each containing at least two nodes. Write down corresponding conditional independence and factorisation properties.

Problem 3

Consider the DAG

$$
\begin{tikzcd}
A \arrow[r] \arrow[dr] & B \arrow[r] \arrow[d] & C \\
& D \arrow[ur] &
\end{tikzcd}
$$

- Write down the ancestors, descendants, parents, and children of each node
- Given a probability distribution over the nodes, write down the Markovian factorisation of the joint probability function implied by the DAG
- Is $C \perp\!\!\!\perp A \mid B$?

Problem 4

Using the Markov factorisation interpretation of a DAG (i.e. the DAG as expressing a series of conditional probability function properties) and standard probability theory, prove the product decomposition rule for a set of N random variables (not necessarily ordered), x_1, x_2,...,x_N:

$$p(x_1, x_2, ..., x_N) = \prod_{i=1}^{N} p(x_i \mid \mathrm{pa}(x_i))$$

where $\mathrm{pa}(x_i))$ are the parents of x_i in the DAG.

Problem 5

Look up http://www.dagitty.net/.

Lecture 6: DAGs continued

Note for the questions 6, 7 and 8 below. Consider the undirected arrangement:

$$\begin{tikzcd} & B \arrow[dl, dash] \arrow[dr, dash] & \\ A & & C \end{tikzcd}$$

When considering the relationship between A and C in the DAG versions of the above diagram, the 'middle' variable B is called: a 'mediator' in a 'chain', a 'confounder' in a 'fork', and a 'collider' in a … collider. See next.

Problem 6

Consider the DAG

$$\begin{tikzcd} & B \arrow[dl] \arrow[dr] & \\ A & & C \end{tikzcd}$$

- Which of these is true: $A \perp\!\!\!\perp C \mid B$ or $A \perp\!\!\!\perp C$?
- Is B a confounder, collider or mediator variable with respect to the relationship between A and C?
- Give an example of a 'real world' relationship between three variables that might satisfy the above diagram

Problem 7

Consider the DAG

$$\begin{tikzcd} & B & \\ A \arrow[ur] & & C \arrow[ul] \end{tikzcd}$$

- Which of these is true: $A \perp\!\!\!\perp C \mid B$ or $A \perp\!\!\!\perp C$?
- Is B a confounder, collider or mediator variable with respect to the relationship between A and C?
- Give an example of a 'real world' relationship between three variables that might satisfy the above diagram

Problem 8

$$\begin{tikzcd} & B \arrow[dr] & \\ A \arrow[ur] & & C \end{tikzcd}$$

- Which of these is true: $A \perp\!\!\!\perp C \mid B$ or $A \perp\!\!\!\perp C$?
- Is B a confounder, collider or mediator variable with respect to the relationship between A and C?
- Give an example of a 'real world' relationship between three variables that might satisfy the above diagram

Problem 9

Consider the DAG

$$\begin{tikzcd} A \arrow[r] & B \arrow[d] & C \arrow[l] \\ & D & \end{tikzcd}$$

- Suppose we are interested in the relationship between A and C. What does d-separation say about the effect of conditioning on D?

Problem 10

In a classic 1993 paper called "Toward a Clearer Definition of Confounding", Clarice Weinberg considered questions related to the following diagram (among others):

$$\begin{tikzcd} A \arrow[r] & B \arrow[r] \arrow[d] & C \\ & D & \end{tikzcd}$$

The question can be phrased as: if we want to understand the (causal) relationship between A and C, should we condition on D? It meets one "classical" definition of a confounder in epidemiology, but Weinberg showed that controlling for D is generally a bad idea.

We need an extra ingredient to our basic d-separation rules to understand why. This arises, essentially, because d-separation deals with whether a set of paths are either 'fully blocked' or not, but *not* whether they are 'partially blocked'.

This extra rule can be stated as: **conditioning on a descendant of a variable is like 'partially' conditioning on that variable**.

This rule applies to all three key types of junction: chains (mediators), forks (confounders), and colliders. Note that it already appears in the d-separation conditions for colliders in the form: conditioning on a descendant of a collider (partially) opens the junction, i.e. the path is no longer *fully* blocked. The implications for mediators and confounders can be illustrated by answering the following questions:

- Explain, using this new rule and d-separation, why D would be a bad thing to condition on when we are interested in the relationship between A and C.
- How would your answer change if the arrow from A to B was pointed from B to A instead?

Problem 11

Consider the DAG

$$\begin{tikzcd} & B \arrow[dl] \arrow[dr] & \\ A \arrow[rr] & & C \end{tikzcd}$$

- What does the DAG look like if you intervene on A?
- Derive expressions for $p(c|a)$ and $p(c|\mathrm{do}(a))$

Problem 12

Consider the DAG describing relationships between the season (summer, winter etc.), whether the nearby sprinkler is on, whether it's raining, whether the footpath is wet and whether the footpath is slippery:

$$
\begin{tikzcd}
& \text{Season} \arrow[dl] \arrow[dr] & \\
\text{Sprinkler} \arrow[dr] & & \text{Rain} \arrow[dl] \\
& \text{Wet} \arrow[d] & \\
& \text{Slippery} &
\end{tikzcd}
$$

- What does the DAG look like if you intervene on 'Sprinkler'?
- Given this diagram would you expect $p(\text{Slippery}|\text{Sprinkler})$ and $p(\text{Slippery}|\text{do}(\text{Sprinkler}))$ to be the same?

Lecture 7: Markov models in time

Problem 13

Consider the transition matrix

$$\mathbf{P} = \begin{bmatrix} 1/5 & 4/5 \\ 2/5 & * \end{bmatrix}$$

- What is the entry denoted '$*$' ?
- Calculate the 2- and 3-step transition matrices

Problem 14

Consider the transition matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- Is this a valid transition matrix?
- Calculate the 2- and 3-step transition matrices
- Given initial distribution $\mu_0 = [1/3, 1/3, 1/3]$, find the marginal at the next time step, μ_1.
- Given μ_1 from before, find μ_2
- Use $\mathbf{P}_2$ to 'evolve' μ_0 to μ_2 directly and verify your answer is the same.
- State the relevant property of homogeneous Markov chains that justifies the above sort of result in the general case.

Problem 15

Consider a Markov process $X_0, X_1, ...$, with state space $\{0, 1, 2\}$ and transition matrix:

$$\mathbf{P} = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.9 & 0.1 & 0.0 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$

and suppose the initial distribution is $\mu_0 = [0.3, 0.4, 0.3]$.

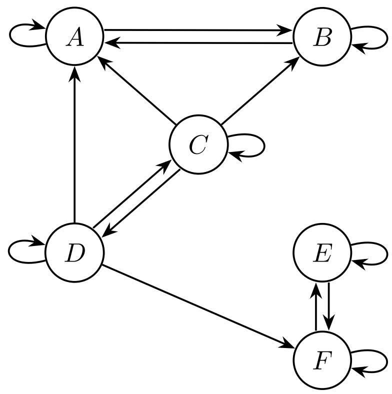
- Calculate the joint probability $\mathbb{P}(X_0 = 1, X_1 = 1, X_2 = 2)$
- Calculate the joint probability $\mathbb{P}(X_0 = 0, X_1 = 1, X_2 = 1)$

Problem 16

Prove the Chapman-Kolmogorov equation from first principles.

Problem 17

Consider the transition diagram



- Find the communication classes
- Find the trapping sets
- Is the Markov chain irreducible?
- Are there any absorbing states?

Lecture 8: Markov models in time continued

Problem 18

Consider the transition matrix

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix}$$

- Find the invariant distribution
- By hand or computer, compute $\mathbf{P}^n$ for a series of increasing values of n. What does this indicate?
- Describe the difference between an invariant and a limiting distribution.

Problem 19

Consider the transition matrix

$$\mathbf{P} = \begin{bmatrix} 0.40 & 0.50 & 0.10 \\ 0.05 & 0.70 & 0.25 \\ 0.05 & 0.50 & 0.45 \end{bmatrix}$$

- Find the invariant distribution
- Suppose the distribution you found above is also a limiting distribution. Write down the form the limiting transition matrix $\mathbf{P}^n, n \to \infty$ takes.

Problem 20

Given an invariant/limiting distribution and a utility function over the states of a Markov chain:

- Describe how to compute the expected utility under this distribution.
- Give a concrete example (i.e. with numbers!)

Suppose instead that you only had a long series of samples (i.e. state realisations) from a Markov chain and a utility function over states but didn't know the invariant/limiting distribution. How might you calculate the expected utility? Would you expect this to be close to the answer obtained given the invariant/limiting distribution?

Problem 21

Suppose you have the following sequence of observations assumed to be realisation from a Markov chain with state space $\{1, 2, 3\}$:

$$1, 2, 1, 1, 1, 2, 3, 1, 1, 2, 1, 3, 1, 1, 3, 2, 2, 1, 1, 3$$

- Write down an expression that you could use to estimate the transition matrix
- Estimate the transition matrix
- Assuming the Markov chain is ergodic, calculate an approximation to the expected utility for a utility function $u(i) = 2i + i^2$

Problem 22

Suppose you observed a process assumed to be Markov and found the following matrix of counts

$$\mathbf{C} = \begin{bmatrix} 412 & 54 & 165 \\ 54 & 19 & 25 \\ 165 & 24 & ? \end{bmatrix}$$

- Is it possible to determine the missing entry? Why/why not?
- Given the missing entry is 143, determine the associated transition matrix
- If the data from this question are observations of the same process as in the previous question, which matrix would you prefer to use to estimate the true process? You can justify this informally.
- Calculate an approximate expected utility using this data for the same utility function as in the previous question

Miscellaneous

Problem 23

Based on Lectures 6 and 7, describe a simple, potentially very inefficient, 'direct sampling' algorithm for calculating a distribution over a future state value of a Markov chain given a collection of past state values.

Problem 24

Read the supplementary material at the end of Lecture 8 (e.g. forwards-backwards algorithm) and try to implement some hidden Markov model algorithms [not examinable!].