§4 Graphical Models

Outline

1. Introduction (graphical models: conditional independence, graphs)
2. Directed acyclic graphs (DAGs)
3. Moralising a DAG
4. Factorisation theorem; Markov blankets and full-conditional distributions
5. Summary

1. Introduction

Why graphical models?

Most realistic applications involves many inter-connected random variables. We want an easy way

- to model and visualise the relationships between these random variables
- to figure out the properties of the model, e.g. conditional independence structure
- to simplify the fitting of the model

What is a graphical model?

A graphical model has 2 features:

1. it is a *probability model* for multiple random variables
2. *(conditional) independence structure* of the model is characterised by a *graph*

Conditional independence

Two variables, X and Y, are *marginally independent* (denoted by $X \perp\!\!\!\perp Y$) if

$$p(x,y) = p(x)\,p(y) \quad \forall x,y.$$

Equivalently, as $p(x,y) = p(x)p(y|x)$, variables X and Y are independent if

$$p(y \mid x) = p(y) \quad \forall y \text{ and } \forall x \text{ s.t. } p(x) > 0 \text{ ,}$$

ie knowing X tells me nothing about Y, and vice versa

Given three variables X, Y and Z, we say that X and Y are *conditionally independent* given Z, denoted by $X \perp\!\!\!\perp Y \mid Z$, if

$$p(x,y \mid z) = p(x \mid z)p(y \mid z)$$

(Does not necessarily mean that X and Y are marginally independent, and vice versa)

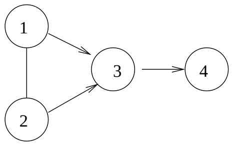
Example

$X =$ height of child
$Y =$ mathematical ability of child
$Z =$ age of child

Is $X \perp\!\!\!\perp Y$?
Is $X \perp\!\!\!\perp Y \mid Z$?

Chapter 1 of Whittaker (1990) discussed some interesting examples of conditional independence: mathematics marks, infant survival, Markov chains, regression, random effects models, sufficient statistics......

Graphs



A graph consists of two sets: a set of nodes (or vertices) and a set of edges. Each edge connects a pair of nodes.

In a probabilistic graph: each node represents a random variable; the edges express association between these random variables.

Edges may be directed (arrows) or undirected (lines). A graph is called

- *undirected* if all its edges are undirected
- *directed* if all its edges are directed
- *mixed* if it contains directed and undirected edges

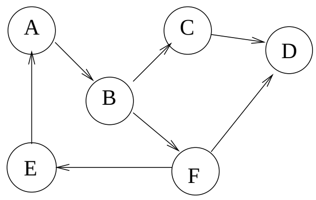
Undirected graphical models are also called *Markov random fields*.
Directed graphical models are also called *Bayesian networks*.

2. Directed Acyclic Graphs (DAGs)

What is a DAG?

A DAG is a directed graph that contains no directed cycles.

Is this a DAG?



Why DAGs?

DAGs are useful for visualising and investigating conditional dependence, e.g. causal relationship, between random variables.

In most studies of interacting variables, there is a lack of symmetry in the variables: it makes sense to say that X might 'cause' (influence) Y, but not to say that Y might 'cause' (influence) X.

- Smoking may 'cause' lung cancer, but not vice versa
- Parents' genotypes 'cause' child's genotype, not vice versa
- And another example: parameters 'cause' data, not vice versa

How to build DAGs?

- Each random variable in the model is represented by a node
- An arrow pointing from one node to another indicates that the first variable 'causes' (influences) the second

Example 4.1

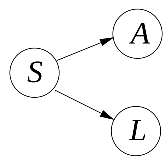
A = arterial disease (present/absent)

L = lung cancer (present/absent)

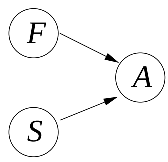
S = smoking amount (pack-years)

F = fat consumption (g per week)

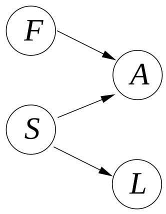
There is evidence smoking 'causes' (increases risk of) both lung cancer and arterial disease:



There is evidence that fat consumption is also related to arterial disease:



Combining these two sub-models, we obtain:



Is

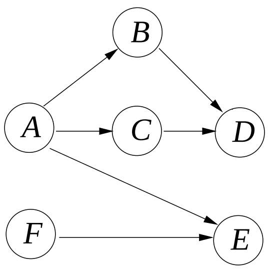
- F marginally independent of S?
- L and A are conditionally independent given S ($L \perp\!\!\!\perp A \mid S$)?
- $F \perp\!\!\!\perp S \mid A$?
- $F \perp\!\!\!\perp L \mid S$?
- $F \perp\!\!\!\perp L \mid A$?

These are not easy to answer without some ‘rules’.

Terminology for DAGs

- *Parents* of a node are the nodes immediately ‘upstream’ (arrows point from parents)
- *Children* of a node are the nodes immediately ‘downstream’ (arrows point to children)
- *Ancestors* of a node are all ‘upstream’ nodes (can get from ancestor to node by following arrows)
- *Descendants* of a node are all ‘downstream’ nodes (can get from node to descendants by following arrows)
- *Founders* are nodes without parents

Example 4.2

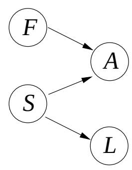


Which nodes are the

- parents of D?
- children of A?
- ancestors of D?
- descendants of A?
- founders?

Some properties of DAGs:

1. A node is conditionally independent of its ancestors given its parents
2. Founders are marginally independent
3. Parents of a node are conditionally dependent given that node



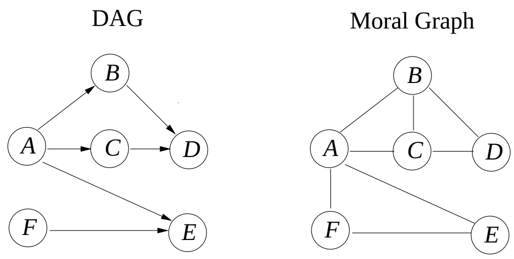
So, $F \perp\!\!\!\perp S$ is true, but $F \perp\!\!\!\perp S \mid A$ is not true.

This does not tell us whether $F \perp\!\!\!\perp L \mid S$ or $F \perp\!\!\!\perp L \mid A$.

To answer these questions, convert the DAG into its corresponding Conditional Independence Graph (a type of undirected graph).

3. Moralising a DAG

Converting a DAG to its corresponding conditional independence (c.i.) graph is called "moralising" the DAG



- moralisation: "marry" the parents, ie connect them with undirected edges
- drop the directions on other edges

To determine whether $X_1 \perp\!\!\!\perp X_2 \mid X_3$ (here each X_i can be a set of random variables):

- first, keep (X_1, X_2, X_3) and their ancestors in the original DAG to form a new (partial) DAG;
- secondly, convert this (partial) DAG to a c.i. graph by 'moralising';
- finally, use the global Markov property of c.i. graphs to determine conditional independence

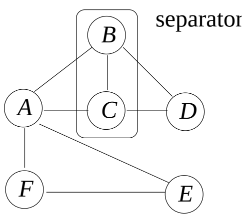
The global Markov property of c.i. graphs

A separating set, S, of two nodes (or two sets of nodes) N_1 and N_2 is a set of nodes that blocks all paths between N_1 and N_2.

The global Markov property of c.i. graphs:

$$N_1 \perp\!\!\!\perp N_2 \mid S$$

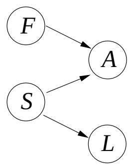
Example 4.3



Here, (B, C) is a separating set for nodes A and D, but B or C by itself is not. Thus $A \perp\!\!\!\perp D \mid B, C$.

Are there any other conditional independencies shown in this graph?

Example 4.4

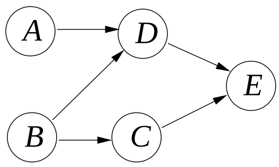


For the lung cancer and arterial disease example, convert the DAG into its corresponding c.i. graph and determine whether or not

- $F \perp\!\!\!\perp S$?
- $L \perp\!\!\!\perp A \mid S$?
- $F \perp\!\!\!\perp S \mid A$?
- $F \perp\!\!\!\perp L \mid S$?
- $F \perp\!\!\!\perp L \mid A$?

Example 4.5

$A =$ salt intake
$B =$ smoking
$C =$ lung cancer
$D =$ heart disease
$E =$ death



By moralising the graph, work out whether or not

1. $A \perp\!\!\!\perp B$?
2. $C \perp\!\!\!\perp A \mid B$?
3. $D \perp\!\!\!\perp C \mid A, B$?
4. $E \perp\!\!\!\perp A, B \mid C, D$?
5. $C \perp\!\!\!\perp A \mid B, E$?
6. $D \perp\!\!\!\perp C \mid A, B, E$?

4. Factorisation theorem; Markov blankets and full-conditional distributions

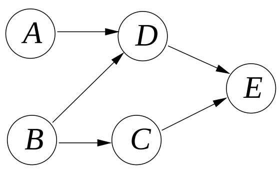
The Factorisation Theorem

Write the random variables in a probability model as $X_1, \ldots, X_K$, say, and let $\mathbf{X} = (X_1, \ldots, X_K)$. If the model is represented by a DAG, the joint probability distribution, $p(\mathbf{X})$, of all the random variables, $\mathbf{X}$, in the model can be calculated using *Factorisation Theorem*:

$$p(\mathbf{X}) = \prod_{k=1}^{K} p\left(X_k \mid \text{parents}\left[X_k\right]\right)$$

Note: We shall need this when we want to fit a Bayesian model.

Example 4.6



What is the joint distribution of (A, B, C, D, E)?

$$\begin{aligned} p(A, B, C, D, E) &= p(A)p(B)p(C \mid B) \\ &\quad \times p(D \mid A, B)p(E \mid C, D) \end{aligned}$$

(since A and B have no parents; B is the only parent of $C \ldots$)

Why is the Factorisation Theorem true?

Consider the model in Examples 4.5 & 4.6.

First, we can always write the joint distribution of a vector of random variables (A, B, C, D, E) as the following sequence of conditional distributions:

$$\begin{aligned} p(A, B, C, D, E) &= p(A)p(B \mid A)p(C \mid A, B) \\ &\quad \times p(D \mid A, B, C) \\ &\quad \times p(E \mid A, B, C, D) \end{aligned}$$

Secondly, our solutions on p.17 show that

$$\begin{aligned} B \perp\!\!\!\perp A &\Rightarrow p(B \mid A) = p(B) \\ C \perp\!\!\!\perp A \mid B &\Rightarrow p(C \mid A, B) = p(C \mid B) \\ D \perp\!\!\!\perp C \mid A, B &\Rightarrow p(D \mid A, B, C) = p(D \mid A, B) \\ E \perp\!\!\!\perp A, B \mid C, D &\Rightarrow p(E \mid A, B, C, D) = p(E \mid C, D) \end{aligned}$$

So, it follows that, for this particular model ,

$$\begin{aligned} p(A, B, C, D, E) &= p(A)p(B)p(C \mid B) \\ &\quad \times p(D \mid A, B)p(E \mid C, D) \end{aligned}$$

Markov blankets

Related to the Factorisation Theorem is a result that will prove useful when using MCMC methods to fit Bayesian models:

$$X_k \perp\!\!\!\perp \mathbf{X}_{\setminus(X_k,\text{bl}[X_k])} \mid \text{bl}[X_k]$$

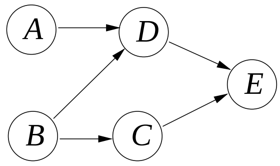
where $\mathbf{X}_{\setminus \mathbf{Y}}$ denotes the vector $\mathbf{X}$ excluding a sub-vector $\mathbf{Y}$, and bl$[X_k]$ is the *Markov blanket* of X_k, given by

$$\begin{aligned} \text{bl}[X_k] &= \text{parents}\,[X_k] \ \cup \ \text{children}\,[X_k] \\ &\quad \cup \ \text{partners}\,[X_k] \end{aligned}$$

(Partners are other parents of X_k's children).

Note: The Markov blanket can also be found from the c.i. graph, where it is just the set of nodes that are directly linked to X_k (ie X_k's neighbours in the graph).

Example 4.7



What is the Markov blanket of C?

We have

$$\begin{aligned} \text{parents}\,[C] &= \{B\} \\ \text{children}\,[C] &= \{E\} \\ \text{partners}\,[C] &= \{D\} \end{aligned}$$

So, the Markov blanket of C is

$$\text{bl}[C] = \{B, E, D\} \ .$$

Full-conditional distributions

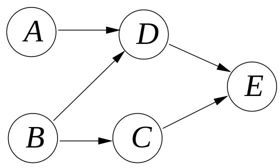
Using this result, it is straightforward to show that

$$p(X_k \mid \mathbf{X}_{\setminus X_k}) \ \propto \ p(X_k \mid \text{parents}\,[X_k]) \times \prod_{W \,\in\, \text{children}[X_k]} p(W \mid \text{parents}\,[W])$$

This is called the *full-conditional distribution* of X_k.

- The full-conditional distribution of X_k depends only on the variables in X_k's Markov blanket
- Full-conditional distributions are needed for fitting Bayesian models using MCMC methods

Example 4.8



What is the full-conditional distribution of C?

We have

$$\begin{aligned} \text{parents}\,[C] &= \{B\} \\ \text{children}\,[C] &= \{E\} \\ \text{parents}\,[E] &= \{C, D\} \end{aligned}$$

So, the full-conditional distribution of C is

$$p(C \mid A, B, D, E) \ \propto \ p(C \mid B) \times p(E \mid C, D) \ .$$

Note:

- The Markov blanket of C is bl$[C] = \{B, E, D\}$
- The joint distribution of the model is

$$\begin{aligned} p(A, B, C, D, E) &= p(A)p(B)p(C \mid B) \\ &\quad \times p(D \mid A, B)p(E \mid C, D) \end{aligned}$$

5. Summary

- DAGs provide a means of representing and visualising complex statistical models. This simplifies model specification and helps communicate essential structure of the model
- Conditional independence (c.i.) assumptions can be read off DAGs through moralising the DAG to a c.i. graph
- The factorisation theorem for DAGs means that we can specify a joint probability model as the product of simple local (parent-child) relationships
- The full-conditional distribution of variable X given all other variables in model depends only on those variables in X's Markov blanket
  - We shall use this property to specify the necessary distributions for implementing MCMC simulation methods

Outline revisited

1. Introduction (graphical models: conditional independence, graphs)
2. Directed acyclic graphs (DAGs)
3. Moralising a DAG
4. Factorisation theorem; Markov blankets and full-conditional distributions
5. Summary

Next week: Hierarchical Models