Lecture 4: Wrap up of Bayes net representation

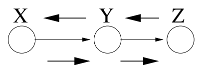
- An example of Bayes ball
- Markov blanket, moral graph
- Independence maps and perfect maps
- Practical considerations

Recall from last time

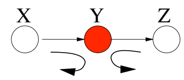
- A Bayes net can be viewed as an *independence map (I-map)* for some distribution
- The I-map property means that the distribution factorizes according to the graph structure of the net
- But the graph can have more arcs than necessary!
- The Bayes ball algorithm can be used to determine which variables are conditionally independent in the presence of evidence
- Knowing conditional independencies will help us provide faster inference

Recall: Base rules for propagating information

- *Head-to-tail*

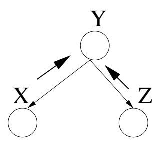


Y unknown, path unblocked

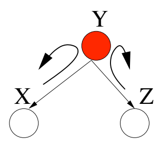


Y known, path blocked

- *Tail-to-tail*

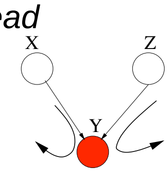


Y unknown, path unblocked

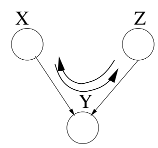


Y known, path blocked

- *Head-to-head*



Y unknown, path BLOCKED

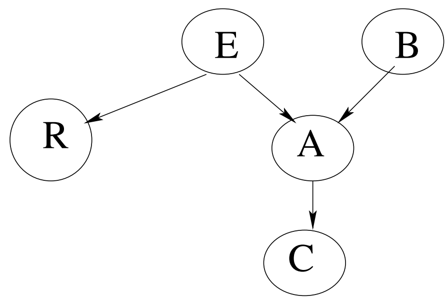


Y known, path UNBLOCKED

d-separation

- Suppose we want to show that a conditional independence relation, $X \perp\!\!\!\perp Z | Y$, is implied by a DAG G in which X, Y, Z are non-intersecting sets of nodes.
- A path is said to be **blocked** if it includes a node such that:
  1. the arrows in the path do *not* meet head-to-head at the node, and the node is in the conditioning set Y (this covers the head-to-tail and tail-to-tail cases)
  2. the arrows do meet head-to-head and neither the node nor its descendents are in Y
- If, given the set of conditioning nodes Y, all paths from any node in X to any node in Z are blocked, then X is **d-separated (directed-separated)** from Z given Y

Example: The alarm network



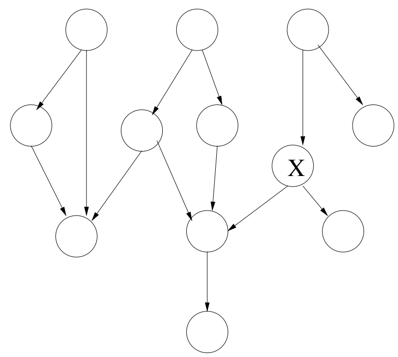
Is $R \perp\!\!\!\perp C | A$?

Important results

- "Soundness": If a joint distribution p factorizes according to a DAG G, and if X, Y and Z are subsets of nodes such that Y d-separates X and Z in G, then p satisfies $X \perp\!\!\!\perp Z | Y$.
- "Completeness": if Y does <u>*not*</u> d-separate X and Z in DAG G, then there exists at least one distribution p which factorizes over G and in which $X \not\!\perp\!\!\!\perp Z | Y$

Isolating a node

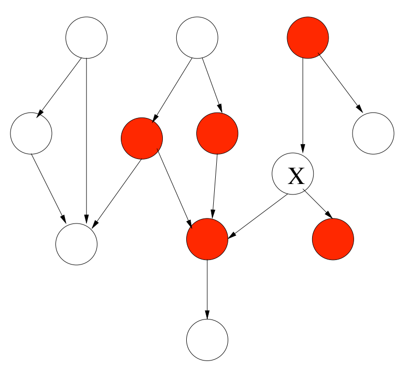
Suppose we want the smallest set of nodes U such that X is independent of all other nodes in the network given U: $X \perp\!\!\!\perp (\{X_1 \ldots X_n\} - \{X\} - U) \,|U$. What should U be?



Markov blanket

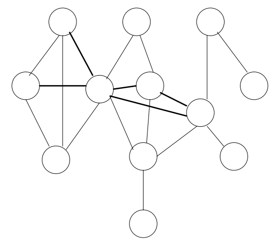
- Clearly, at least X's parents and children should be in U
- But this is not enough if there are v-structures; U will also have to include X's "spouses" - i.e. the other parents of X's children

The set U consisting of X's parents, children and other parents of its children is called the **Markov blanket** of X.



Moral graphs

Given a DAG G, we define the **moral graph of G** to be an undirected graph U over the same set of vertices, such that the edge (X, Y) is in U if X is in Y's Markov blanket



- If G is an I-map of p, then U will also be an I-map of p
- But many independencies are lost when going to a moral graph
- Moral graphs will prove to be useful when we talk about inference.

Perfect maps

A DAG G is a **perfect map** of a distribution p if it satisfies the following property:

$$X \perp\!\!\!\perp Z | Y \Leftrightarrow Y \text{ d-separates } X \text{ and } Z$$

- A perfect map captures all the independencies of a distribution
- Perfect maps are unique, up to DAG equivalence
- How can we construct a perfect map for a distribution?

Example

- Consider a distribution over 4 random variable X, Y, Z, W such that:
  - $X \perp\!\!\!\perp Y | \{Z, W\}$
  - $Z \perp\!\!\!\perp W | \{X, Y\}$
- Can you find an I-map for this distribution?
- Can you find a perfect map?

Example

- Consider a distribution over 4 random variable X, Y, Z, W such that:
  - $X \perp\!\!\!\perp Y | \{Z, W\}$
  - $Z \perp\!\!\!\perp W | \{X, Y\}$
- Can you find an I-map for this distribution?
- Can you find a perfect map?

  Some distributions do not have perfect maps!

Representing distributions more compactly

- Sometimes the conditional probabilities at a node can be specified more compactly than through a table.
- E.g., If it rains, this influences the probability that you get wet, but only if you go outside.
- In this case, the CPD can be represented more succinctly using a *tree*
- People often find it natural to write tree-structured CPDs
- Other compact representations are also possible (e.g. rules)

Example: Pathfinder (Heckerman, 1991)

- Medical diagnostic system for lymph node diseases
- Large net! 60 diseases, 100 symptoms and test results, 14000 probabilities
- Network built by medical experts
  - 8 hours to determine the variables
  - 35 hours for network topology
  - 40 hours for probability table values
- Experts found it easy to invent causal links and probabilities
- Pathfinder is now *outperforming world experts* in diagnosis
- Commercialized by Intellipath and Chapman Hall Publishing; extended to other medical domains

Typical applications for Bayes nets

- Medical diagnosis
- Bioinformatics (data integration)
- Risk assessment
- Environmental science (e.g., wildlife habitat viability, risk of foreign species invasion)
- Analysis of demographic data
- In general, diagnosis and causal reasoning tasks
- Many commercial packages available (e.g. Netica, Hugin, WinMine, ...)
- Sometimes Bayes net technology is incorporated in business software

Constructing graphical models in practice

Usually, we do not construct graphical models based on knowledge of the joint probability distribution p. We have some vague idea of the dependencies in the world, and we need to make that precise using a graph.

This involves several steps:

- Formulating the problem
- Choosing random variables
- Choosing independence relations
- Assigning probabilities in the CPDs

Choosing random variables

- Variables must be *precise*. What are the values, how are they defined, and how are they measured?
  E.g. *Weather* - what values will it take? When do we assign the *bitter-cold* value?
- If the variables are continuous and we discretize them, a coarse discretization may introduce additional dependencies (more on continuous variables later)
- There are several kinds of variables:
  - Observable
  - Sometimes observable (e.g. medical tests)
  - Hidden - these may or may not be useful to include, depending on the other independencies that they generate

Choosing the graph structure

- For a Bayes net, if we have information about causality, using causal connections will make the graph sparser.
- There is often a trade-off between the precision of the model, and the size/sparsity of the graph

Choosing the parameters of the model: Bayes nets

- Conditional probabilities could come from a few sources:
  - An expert
    - People hate picking numbers!
    - Having a good network structure usually makes it easier to elicit numbers from people too.
  - An approximate analysis (e.g. in card games)
  - Guessing
  - Learning
- Bad news: In all these cases, the numbers are approximate!
- Good news: the numbers usually do not matter all that much.
- Sensitivity analysis can help in deciding whether certain numbers are critical or not for the conclusions

Important factors when choosing probabilities

- Avoid assigning zero probability to any events, unless you are absolutely certain they cannot occur
- The relative values (or ordering) of conditional probabilities for $p(x|x_{\pi_i})$, given different values of X_{π_i} is important
- Having probabilities that are orders of magnitude different can cause problems in the network