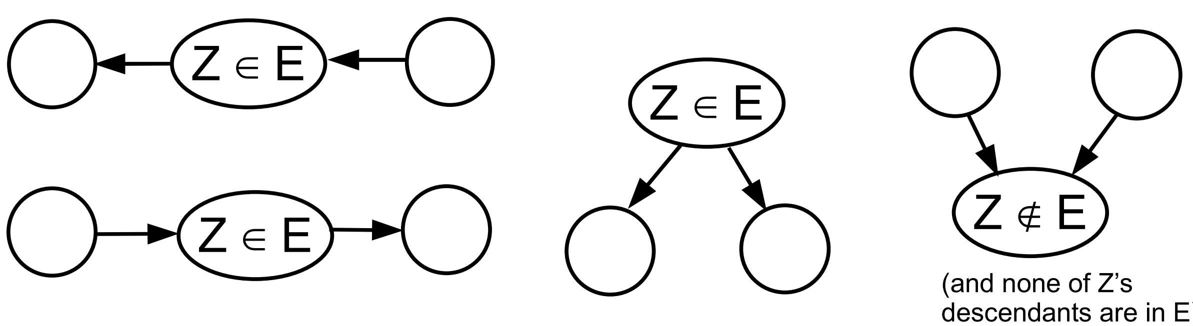
More Bayes Nets

- Remember that for any joint distribution, there are many Bayes nets we could use to represent it. It is always okay to have more edges than necessary. The *fewer* edges in a Bayes net, the *more* constraints are imposed, because fewer variables are allowed to directly depend on each other.

- In **causal reasoning**, we know some causes (nodes higher up in the Bayes net) and predict their effects. In **evidential reasoning**, we observe some effects (nodes lower down in the Bayes net) and reason about their causes. (Note that not every situation is purely causal or purely evidential; we might know some causes and some evidence.)

- Two effect variables have a **common cause** if they are both descendants of the same cause variable. Two cause variables have a **common effect** if they are both ancestors of the same effect variable. When X and Y have a common effect on Z, and the value of Z is known, observing X can **explain away** the value of Z, thereby giving information about Y.

- Two nodes X and Y are **d-separated** by a set of nodes E if (no matter what the CPTs are) they are conditionally independent given E. X and Y are d-separated given E if every trail (undirected path) between X and Y is "inactive" (i.e. carries no information). A trail is inactive if it contains any of the following four structures:



  Therefore X and Y are d-separated by E if *each* of the trails between them contains (at least) one of those four structures.

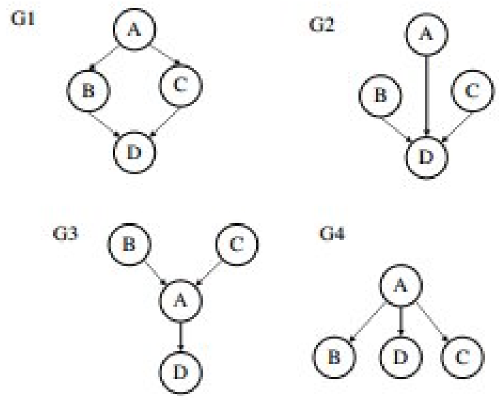
- A **topological order** of the nodes in a Bayes net is a linear order in which no node precedes any of its parents.

- In **approximate inference**, probabilities are estimated using methods such as sampling, instead of computing them analytically. This is because in large networks, exact inference methods such as variable elimination can take a very long time.

- In **rejection sampling**, samples are drawn from the entire network without regard to the observed evidence, and any samples inconsistent with the evidence are thrown out.
- In **likelihood weighting**, the evidence variables are fixed to their observed values, and the remaining variables are sampled given the values of their parents. These samples are not drawn from the correct distribution (because variables can be dependent on non-parents too), so each sample is weighted by the likelihood of the evidence given its parents.

Exercises

The first two questions relate to the set of networks:



1. Following is a list of conditional independence statements. For each statement, name which (if any) of the graph structures G1 – G4 imply it.

   (a) A is conditionally independent of B given C

   (b) A is conditionally independent of B given D

   (c) B is conditionally independent of D given A

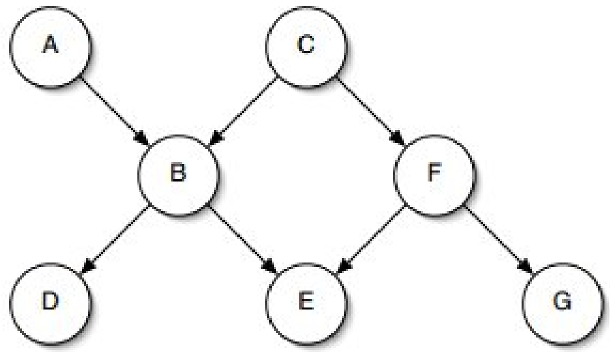
   (d) B is conditionally independent of D given C

   (e) B is independent of C

   (f) B is conditionally independent of C given A

2. How many independent parameters are required to specify a Bayesian network given each of the graph structures G1 – G4? Assume the nodes are binary.

3. The next set of questions question is about the following network:



(a) Is A independent of C?

(b) Is C independent of E?

(c) Is D independent of C?

(d) Name a set of variables that, if given as evidence, would change your answer to part (a).

(e) Name a set of variables that, if given as evidence, would change your answer to part (b).

(f) Name a set of variables that, if given as evidence, would change your answer to part (c).

(g) If all the nodes are binary, how many parameters would be required to specify all the CPTs in this network? (Remember that if p is specified, it is not necessary to specify $1 - p$ as well.)

(h) Give an expression for $\Pr(D|C)$ given probabilities that are stored in the CPTs. Don't include any unnecessary terms.

(i) What factor is created if we eliminate B first in the course of a variable elimination query to compute $\Pr(A|G)$?

(j) What is the Markov blanket of B?

(k) Imagine you're doing likelihood weighting to compute $\Pr(E = e|A = a)$. What weight would you have to assign to sample $< a, b, c, d, e, f, g >$?