COMS30035, Machine learning: Probabilistic Graphical Models 3

James Cussens
james.cussens@bristol.ac.uk

Department of Computer Science, SCEEM
University of Bristol

October 13, 2020

Agenda

- Conditional independence
- d-separation

Conditional independence

- A random variable x is independent of another random variable y *conditional on* a set of random variables S if and only if:

$$P(x, y|S) = P(x|S)P(y|S) \tag{1}$$

Equivalently:

$$P(x|S) = P(x|y, S) \tag{2}$$

- The DAG for a BN encodes conditional independence relations.

Conditional independence

- A random variable x is independent of another random variable y *conditional on* a set of random variables S if and only if:

$$P(x, y|S) = P(x|S)P(y|S) \tag{1}$$

Equivalently:

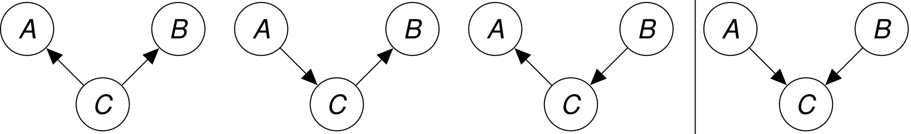
$$P(x|S) = P(x|y, S) \tag{2}$$

- The DAG for a BN encodes conditional independence relations.
- Some of the following slides are modified versions of those made available by David Barber,
- who has written a great (freely available) book on Bayesian machine learing [Bar12]

Independence $\perp$ in Bayesian Networks – Part I

All Bayesian networks with three nodes and two links:

$A \perp B \mid C$ $\qquad$ $A \not\perp B \mid C$

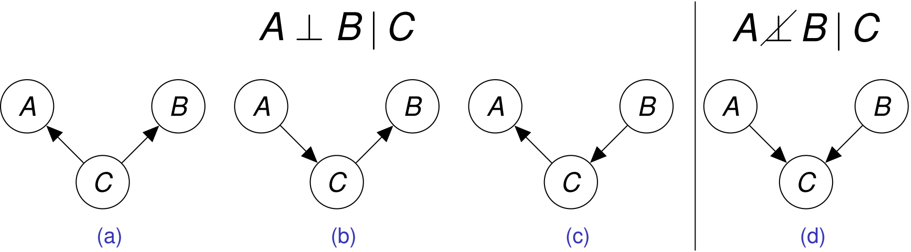


(a) (b) (c) (d)

- In (a), (b) and (c), A and B are conditionally independent given C.
  - (a) $p(A,B|C) = \frac{p(A,B,C)}{p(C)} = \frac{p(A|C)p(B|C)p(C)}{p(C)} = p(A|C)p(B|C)$
  - (b) $p(A,B|C) = \frac{p(A)p(C|A)p(B|C)}{p(C)} = \frac{p(A,C)p(B|C)}{p(C)} = p(A|C)p(B|C)$
  - (c) $p(A,B|C) = \frac{p(A|C)p(C|B)p(B)}{p(C)} = \frac{p(A|C)p(B,C)}{p(C)} = p(A|C)p(B|C)$
- In (d) the variables A, B are conditionally dependent given C, $p(A,B|C) \propto p(A,B,C) = p(C|A,B)p(A)p(B)$.

James Cussens
james.cussens@bristol.ac.uk

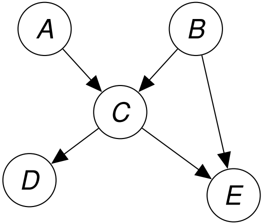
Independence $\perp$ in Bayesian Networks – Exercises



- Show that in (d), we have $A \perp B$.
- For each of (a), (b) and (c), assume that each variable is binary, and find parameters so that $A \not\perp B$

Paths and colliders

$$p(A, B, C, D, E) = p(A)p(B)p(C|A, B)p(D|C)p(E|B, C)$$

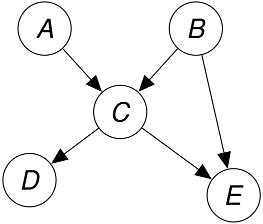


- A node is a *collider* on some path if both arrows point into it on that path.
- C is a collider on the path (A, C, B) but is not a collider on the path (A, C, E) or on any of the following paths: (A, C, E, B), (D, C, B) or (D, C, E).

d-separation

- If all paths from node x to node y are *blocked given nodes* S then x and y are *d-separated* by S.
- A path is blocked by S if at least one of the following is the case:
  1. there is a collider on the path that is not in S and none of its descendants are in S
  2. there is a non-collider on the path that is in S.
- If x and y are *d-separated* by S then $x \perp y | S$ for any probability distribution which factorises according to the DAG.
- Let's do some *d*-separation exercises.

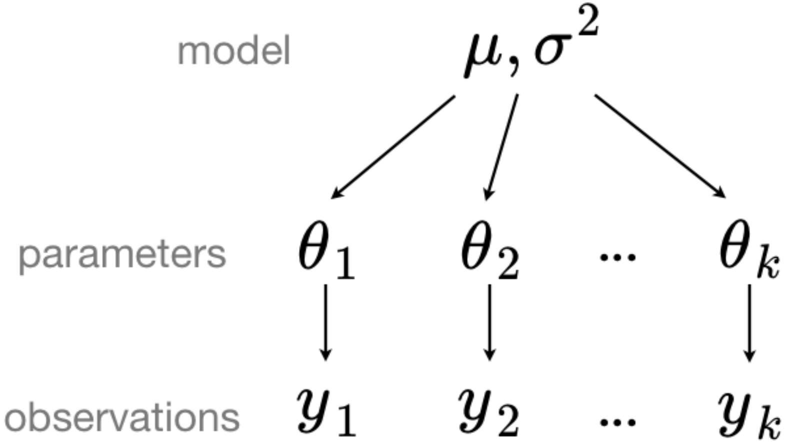
Checking for *d*-separation



A path is blocked by S if at least one of the following is the case:

1. there is a collider on the path that is not in S and none of its descendants are in S
2. there is a non-collider on the path that is in S.

Hierarchical regression revisited



$$P(\theta, y, \mu, \sigma^2) = P(\mu, \sigma^2) \prod_{i=1}^{k} P(y_i|\theta_i) P(\theta_i|\mu, \sigma^2)$$

David Barber.
Bayesian Reasoning and Machine Learning.
Cambridge University Press, 2012.