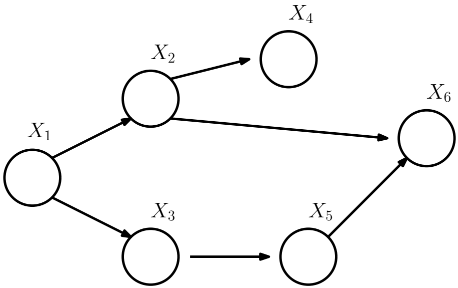
Recap

- Given a directed acyclic graph (DAG) G,

$$p(x_1, x_2, \ldots, x_n) := \prod_{i=1}^{n} p(x_i | x_{\pi_i}).$$

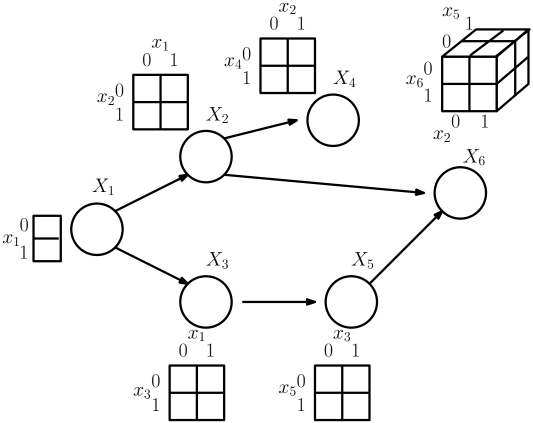


$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3)p(x_6|x_2, x_5)$$

Recap

- Given a directed acyclic graph (DAG) G,

$$p(x_1, x_2, \ldots, x_n) := \prod_{i=1}^{n} p(x_i | x_{\pi_i}).$$



Recap

Which independence assumptions are we exactly making by using a DAG model with a structure described by G? Important because

- we should know exactly what model assumptions we are making;
- this information will be helpful in designing inference algorithms later on.

Recap

Conditional independence

- X_1 and X_2 are conditionally independent given X_3, written $X_1 \perp\!\!\!\perp X_2 | X_3$, iff
  - $p(x_1, x_2|x_3) = p(x_1|x_3)p(x_2|x_3)$, or equivalently
  - $p(x_1|x_2, x_3) = p(x_1|x_3)$.

  for all x_3 such that $p(x_3) > 0$. Given X_3, there is no further relationship between X_1 and X_2.
- Similarly, for sets of random variables, X_A and X_B are conditionally independent given X_C iff

$$p(x_A, x_B|x_C) = p(x_A|x_C)p(x_B|x_C)$$

  or

$$p(x_A|x_B, x_C) = p(x_A|x_C)$$

  for all x_C such that $p(x_C) > 0$.

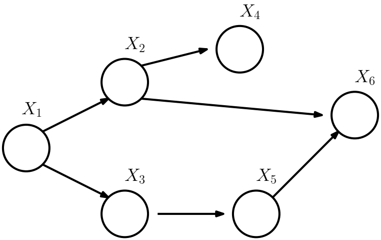
DAG and conditional independence

Compare

$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3)p(x_6|x_2, x_5)$$

and

$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_3, x_2, x_1) \cdots p(x_6|x_5, \ldots, x_1)$$



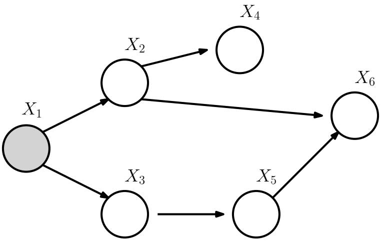
DAG and conditional independence

Compare

$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3)p(x_6|x_2, x_5)$$

and

$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_3, x_2, x_1) \cdots p(x_6|x_5, \ldots, x_1)$$



$X_3 \perp\!\!\!\perp X_2|X_1$. Exercise: verify this using definition.

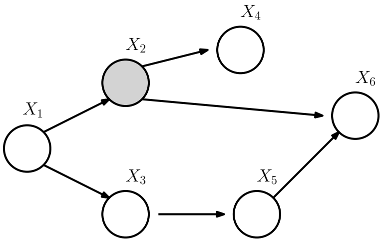
DAG and conditional independence

Compare

$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3)p(x_6|x_2, x_5)$$

and

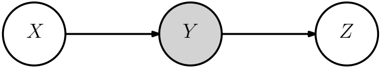
$$p(x_1, \ldots, x_6) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_3, x_2, x_1) \cdots p(x_6|x_5, \ldots, x_1)$$



$X_4 \perp\!\!\!\perp \{X_1, X_3\}|X_2$. Exercise: verify this using definition.

Conditional independence and d-separation

Three canonical graphs - cascade (Markov property)



$$p(x,y,z) = p(x)p(y|x)p(z|y) \Rightarrow X \perp\!\!\!\perp Z|Y,$$

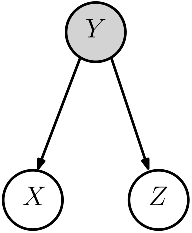
since

$$\begin{aligned} p(z|x,y) =& \frac{p(x,y,z)}{p(x,y)} = \frac{p(x)p(y|x)p(z|y)}{p(x)p(y|x)} \\ =& p(z|y). \end{aligned}$$

e.g. X "past", Y "present", Z "future"

Conditional independence and d-separation

Three canonical graphs - common parent



$$p(x, y, z) = p(y)p(x|y)p(z|y) \Rightarrow X \perp\!\!\!\perp Z|Y,$$

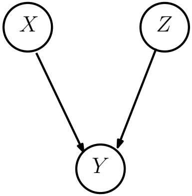
since

$$p(x, z|y) = \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y).$$

e.g. X “shoe size”, Z “gray hair or not”, Y “age”

Conditional independence and d-separation

Three canonical graphs - v-structure



$$p(x, y, z) = p(x)p(y|x, z)p(z) \Rightarrow X \perp\!\!\!\perp Z,$$

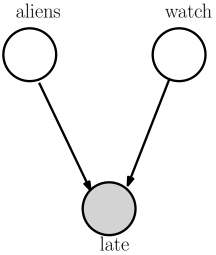
since

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

Can we claim $X \perp\!\!\!\perp Z|Y$? No. In fact observing Y can induce dependence between X and Z.

Conditional independence and d-separation

Three canonical graphs - v-structure



Alice is late for lunch with Bob.

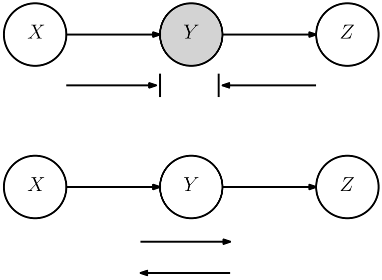
Conditional independence and d-separation

The Bayes ball algorithm

- Decide whether a given conditional statement $X_A \perp\!\!\!\perp X_B | X_C$ is true for a DAG G.
- Convert to a "reachability" algorithm: shade the nodes X_C, place a ball at each of the nodes X_A, let the ball bounce around G according to some rules. If none of the balls can reach any of the nodes in X_B, we assert $X_A \perp\!\!\!\perp X_B | X_C$.

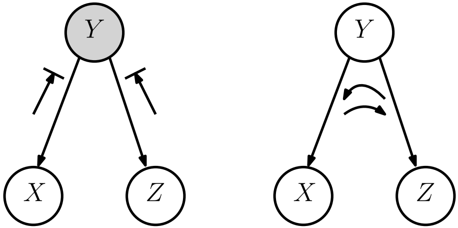
Conditional independence and d-separation

Three canonical graphs - cascade



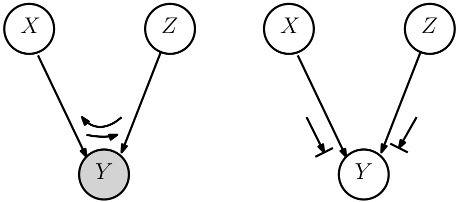
Conditional independence and d-separation

Three canonical graphs - common parent



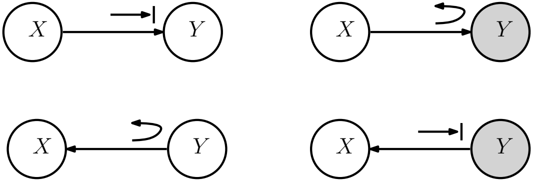
Conditional independence and d-separation

Three canonical graphs - v-structure



Conditional independence and d-separation

When source and destination are the same



Conditional independence and d-separation

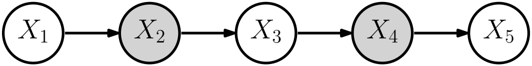
Example 1. Markov chain

$$X_1 \to X_2 \to X_3 \to X_4 \to X_5$$

$X_1 \perp\!\!\!\perp X_5 | X_3$, $X_1 \perp\!\!\!\perp X_4 | X_3$

Conditional independence and d-separation

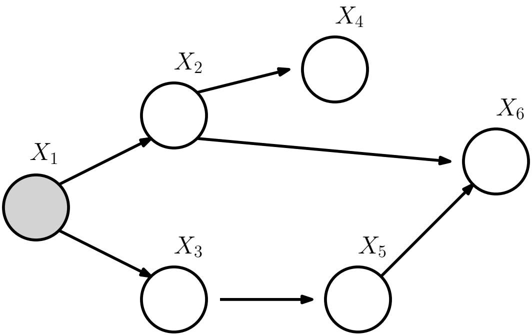
Example 1. Markov chain



$X_1 \perp\!\!\!\perp X_5 | \{X_2, X_4\}$

Conditional independence and d-separation

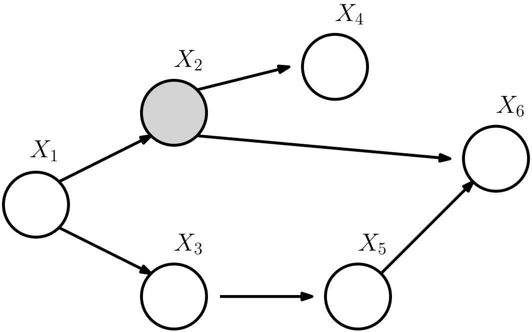
Example 2.



$X_2 \perp\!\!\!\perp X_3 | X_1$

Conditional independence and d-separation

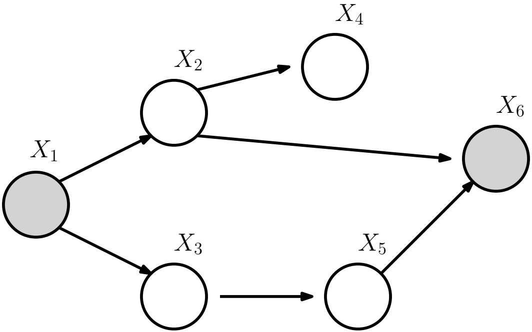
Example 2.



$X_4 \perp\!\!\!\perp \{X_1, X_3\} | X_2$

Conditional independence and d-separation

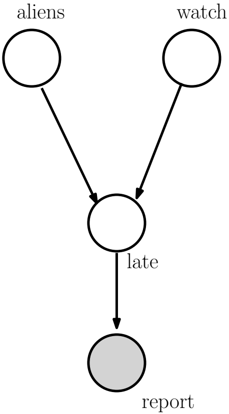
Example 2.



Can we claim $X_2 \perp\!\!\!\perp X_3 | \{X_1, X_6\}$?

Conditional independence and d-separation

Example 3.



alien $\perp\!\!\!\perp$ watch, but cannot assert alien $\perp\!\!\!\perp$ watch$|$report

Two equivalent characterisations

Given a DAG G,

- Generate a family of distributions $\mathcal{D}_1$ as follows:
  - factorisation in terms of conditional probabilities
  - range over all possible choices of numerical values for conditional PMFs
- Generate a family of distributions $\mathcal{D}_2$ as follows:
  - find all conditional independences by running the Bayes ball algorithm
  - consider all possible joint distributions
  - test each against the list of conditional independences; keep the distribution if all satisfied

$\mathcal{D}_1$ and $\mathcal{D}_2$ are the same.