

Announcements

Assignment three posted. It is due Friday, Feb 24 (truthfully Feb 28).

Three random variables summary

In cases one and two, a and b were not independent until the observation of c “blocked” the (connection) path from a to b .

(From Koller and Friedman, a path that is not blocked is “active”)

In case three, if c is not observed, the path is blocked. Observing c made the connection (path) active.

d-Separation (Pearl, 88)

“d” stands for
“directed”

Generalizes the examples we have been studying.

Consider non-overlapping subsets A , B , C of nodes of a graph.

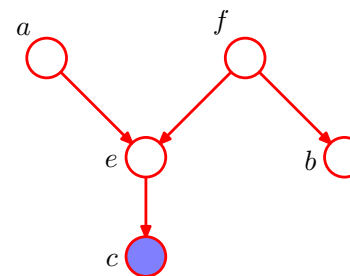
Consider all paths from nodes in A to nodes in B .

A path is blocked if either:

- The arrows meet either tail-to-tail or head-to-tail at a node in C .
- The arrows meet head-to-head at some node that is not in C , nor are any of its descendants in C .

If all paths are blocked, then A and B are independent given C .

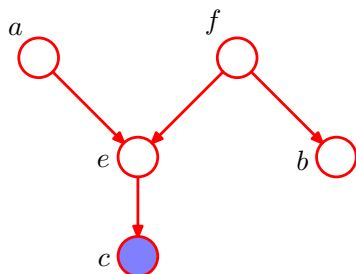
d-Separation (example one)



$(A=\{a\}, B=\{b\}, \text{ and } C=\{c\})$

Does this graph encode $A \perp B | C$?

d-Separation (example one)

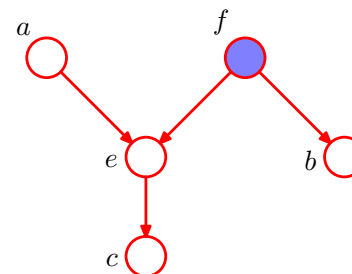


$$A \not\perp B | C$$

The path is not blocked by e because, although it is head-to-head, it has a descendant, c , in the conditioning set.

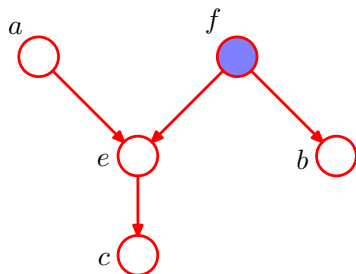
The path is not blocked by f because it is tail-to-tail, and f is not in C .

d-Separation (example two)



Does this graph encode $A \perp B | F$?

d-Separation (example two)



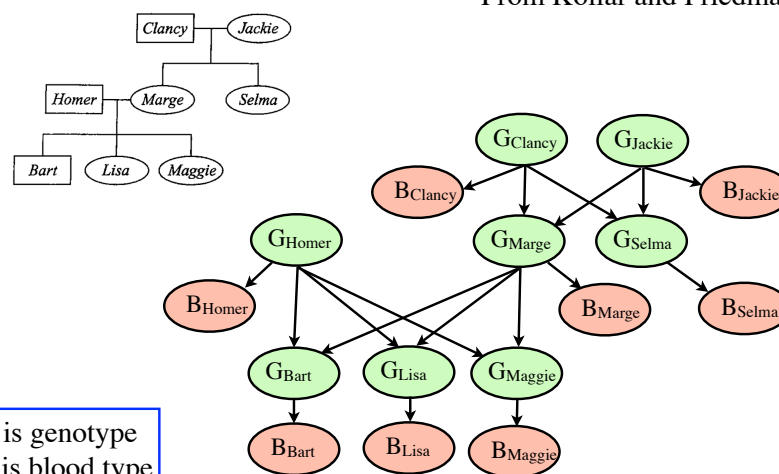
$$A \perp B | F$$

The path is blocked by e because it is head-to-head, and neither it, nor any of its descendants are in the conditioning set.

The path is also blocked by f because it is tail-to-tail, and f is in C .

Grounded example of a Bayesian Network

From Kollar and Friedman



G is genotype
B is blood type

Bayesian network semantics

- Represents a factorization of $p()$
 - Random variables are nodes
 - Factors are CPD (conditional probability distributions) for child given parent (just $p(\text{NODE})$ if no parents).

Equivalent semantic specification (Proof is in K&F, ch. 3)

- For each X_i : $X_i \perp \text{NonDescendents}(X_i) \mid \text{Parents}(X_i)$
 - Notice no mention of factorization

Conditional independence in distributions and graphs

Let $I(P)$ be the set of independence assertions of the form $(X \perp Y \mid Z)$ that are true for a distribution P .

Let $I(G)$ be the set of independence assertions represented by a DAG, G .

G is an I-map for P if $I(G) \subseteq I(P)$

In other words, all independence represented in G are true.
(There could be some more in P that G does not reveal).

A few notes on notation and independence

We sometimes write $(A \perp B \mid \emptyset)$ for $A \perp B$

Also, we write $(A \perp B, C \mid X)$ for $(A \perp B \mid X)$ and $(A \perp C \mid X)$

Recall that $(A \perp B \mid C)$ means that $P(A \mid B, C) = P(A \mid C)$

This generalizes to:

$$(A \perp B \mid \dots, C, \dots) \Rightarrow P(A \mid \dots, B, C, \dots) = P(A \mid \dots, C, \dots)$$

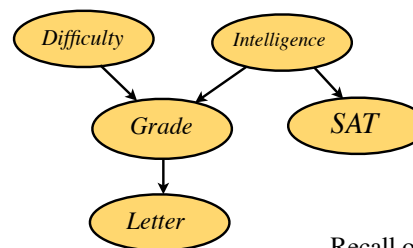
Example going from I-map to a factorization

From Kollar and Friedman

For $P(I, D, G, L, S)$, the $I(\text{Graph})$ tells us

$$(D \perp I \mid \emptyset) \quad (D \perp I \mid S) \quad (L \perp I, D, S \mid G) \quad (G \perp S \mid I, D) \quad (S \perp D, G, L \mid I)$$

(Note that this is not necessarily all relationships that we can extract)



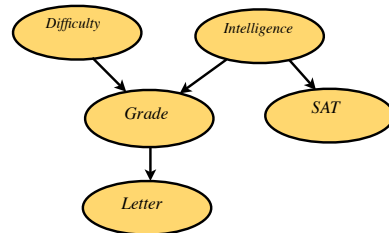
Recall one version of DAG semantics is
 $X_i \perp \text{NonDescendents}(X_i) \mid \text{Parents}(X_i)$

We can write the joint distribution as conditioning on non-descendents if we maintain a sensible "lexicographical order where parents occur before children.

$$P(I, D, G, L, S) = P(I) P(D|I) P(G|I, D) P(L|I, D, G) P(S|I, D, G, L)$$

This means that for each factor, all variables conditioned on are either the parents, or non-descendents.

This means that for each factor, we may have rule that gets rid of some non-descendents.



Example going from I-map to a factorization

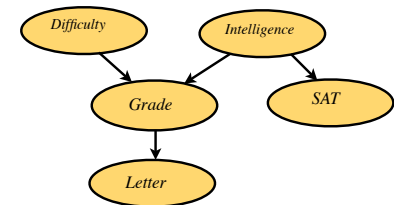
$$P(I, D, G, L, S) = P(I) P(D|I) P(G|I, D) P(L|I, D, G) P(S|I, D, G, L)$$

$$(D \perp I | \emptyset) \Rightarrow P(D|I) = P(D)$$

$$(L \perp I, D, S | G) \Rightarrow P(L|I, D, G) = P(L|G)$$

$$(S \perp D, G, L | I) \Rightarrow P(S|I, D, G, L) = P(S|I)$$

$$\text{So, } P(I, D, G, L, S) = P(I) P(D) P(G|I, D) P(L|G) P(S|I)$$



Summary on the equivalence of the two interpretations of directed graphical models

Factorization semantics

Factors are $p(\text{node} | \text{parents})$

Abstract semantics

$$X_i \perp \text{NonDescendents}(X_i) \mid \text{Parents}(X_i)$$

These are equivalent

Proof of one direction by the one example just completed.

Interesting questions

- Does every probability distribution have a corresponding Bayesian network?

Chain rule says yes

- Given the independence structure of a probability distribution, and a graph that captures them all ($I(G)=I(P)$), is the corresponding graph unique (ignoring isomorphisms)?

Case study of three nodes says no

- Do our graphs faithfully capture the independence structure of our distributions?

TBA