Announcements

• “midterm one” will be posted soon
  – This “midterm” is no different in format from the assignments so far

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 \textbf{not} blocked is “active”)

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

Put more generally, paths are blocked by
1) A tail-tail node in the conditioning set
2) A head-tail node in the conditioning set
3) A head-head node \textbf{not} in the conditioning set, AND that has no descendants in the conditioning set.
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:

a) The arrows meet either tail-to-tail or head-to-tail at a node in $C$.

b) 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 $C$.

---

A $\nabla B | C$

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 $C$.

---

Grounded example of a Bayesian Network

From Koller and Friedman

$A \perp B | F$

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 $F$. 

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(NODE) if no parents).

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

• For each $X_i : X_i \perp \text{NonDescendants}(X_i \mid \text{Parents}(X_i))$
  – Notice no mention of factorization
  – Notice no mention of observed (shaded nodes)
  – Call such independence assertions for a graph, G, $I(G)$
  – Call such independence assertions for a distribution, P, $I(P)$

A few notes on notation and independence

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

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

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

This generalizes to:

$\left( A \perp B, C, \ldots \right) \Rightarrow P(A|\ldots, B, C, \ldots) = P(A|\ldots, C, \ldots)$

Going from independence to a factorization

From Koller and Friedman

For $P(I,D,G,L,S)$, what does $I(Graph)$ tells us?

Recall one version of DAG semantics is

$X_i \perp \text{NonDescendants}(X_i) \mid \text{Parents}(X_i)$
Example going from I-map to a factorization
From Koller and Friedman

For $P(I,D,G,L,S)$, $I(Graph)$ tells us

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

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

Recall one version of DAG semantics is

$X_j \perp \text{NonDescendants}(X_j) \mid \text{Parents}(X_j)$

We can write the joint distribution as conditioning on non-descendants if we maintain a sensible "lexigraphical 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-descendants.

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

Conditional independence in distributions and graphs

Let $I(P)$ be the set of independence assertions of the form $(X \perp Y|Z)$ 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 independance represented in $G$ are true.
(There could be some more in $P$ that $G$ does not reveal).
Summary on the equivalence of the two interpretations of directed graphical models

Factorization semantics
Factors are $p(\text{node} \mid \text{parents})$

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