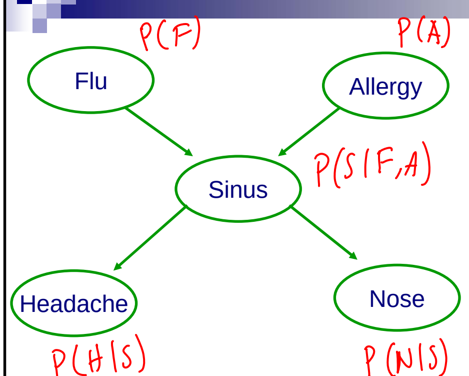
Readings:
K&F: 3.1, 3.2, 3.3.1, 3.3.2

BN Semantics 2 –
Representation Theorem
The revenge of d-separation

Graphical Models – 10708
Carlos Guestrin
Carnegie Mellon University
September 17th, 2008

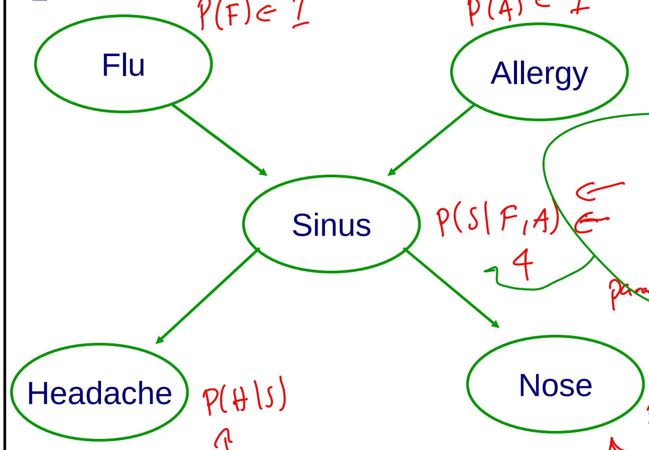
Factored joint distribution - Preview



$P(F,A,S,H,N) =$
$P(F) \cdot P(A) \cdot P(S|F,A)$
$P(H|S) \cdot P(N|S)$

Number of parameters

binary vars



$P(F) \Leftarrow 1$

$P(A) \Leftarrow 1$

$P(F,A,S,H,N)$
explicitly
$2^5 - 1 = 31$

$P(S|F,A) \Leftarrow$ 4

S \ FA	tt	tf	ft	ff
t	.9	.7	.8	.2
f	.1	.3	.2	.8

params $1+1+1+1 = 4$

$P(H|S)$ 2

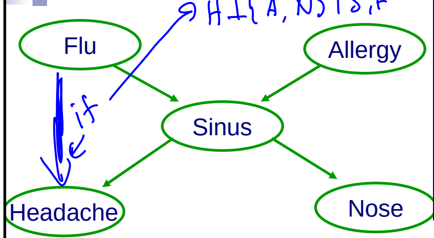
$P(N|S)$ 2

with BN: 10

The independence assumption

Before new edge $H \perp \{A,F,N\} | S$

$H \perp \{A,N\} | S,F$



Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents

($X_i \perp NonDescendants_{X_i} \mid Pa_{X_i}$)

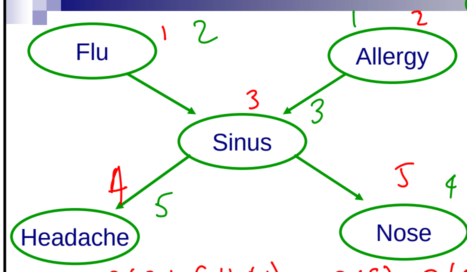
Flu: $Pa_{Flu} = \emptyset$
$NonDescendants_{Flu} = \{A\}$
$F \perp A$

Nose: $Pa_{Nose} = \{S\}$
$NonDescendants_{Nose} = \{F,A,H\}$
$N \perp \{F,A,H\} | S$

Sinus: $Pa_{Sinus} = \{F,A\}$
$NonDescendants_{Sinus} = \emptyset$
NO ASSUMPTIONS ABOUT
$S \perp ?? | F,A$

Joint distribution

topological orderings
interpret a BN
particular chain rule order
+ indep. assumptions



$P(F,A,S,H,N) = P(F)P(A)\,P(S|FA)\cdot P(H|S)\cdot P(N|S)$

Local Markov Assumption
suppose P ← truth satisfies

$P(FASHN) = P(F)\;P(A|F)\;P(S|FA)\cdot P(H|SFA)\;P(N|SFAH)$

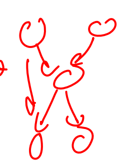
$A \perp F$ → $P(A)$ $P(S|FA)$ $H \perp FA | S$ → $P(H|S)$ $N \perp FAH | S$ → $P(N|S)$

Why can we decompose? Local Markov Assumption

Assumption!

A general Bayes net

- Set of random variables $X_1 \ldots X_n$
- Directed acyclic graph DAG
  - Loops OK
  - but no directed cycles
  - NO! $X_1 \leftarrow X_3$, X_2



- CPTs
  - with each X_i conditional Probability table $P(X_i | Pa_{X_i})$
- Joint distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

- **Local Markov Assumption:**
  - A variable X is independent of its non-descendants given its parents and only its parents – **(Xi ⊥ NonDescendantsXi | PaXi)**

Questions????

- What distributions can be represented by a BN?
- What BNs can represent a distribution?
- What are the independence assumptions encoded in a BN?
  - in addition to the local Markov assumption

$A \rightarrow B \rightarrow C \rightarrow D$

Local Markov

$A \perp C \mid B$

$AB \perp D \mid C$

derived

derived independencies

$A \perp D \mid B$

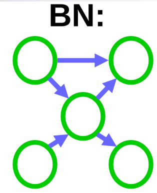
Independencies in Problem

World, Data, reality:

True distribution *P* contains independence assertions

$I(P)$

BN:



Graph *G* encodes local independence assumptions

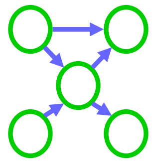
$I_\ell(G)$

Key Representational Assumption:

$I_\ell(G) \subseteq I(P)$

Today: The Representation Theorem – True Independencies to BN Factorization

BN:



local

Encodes independence assumptions $I_\ell(G)$

If conditional independencies in BN are subset of conditional independencies in *P*

$I_\ell(G) \subseteq I(P)$

Obtain

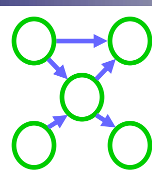
P can be represented exactly by:

Joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Today: The Representation Theorem – BN Factorization to True Independencies

BN:



local

Encodes independence assumptions $I_\ell(G)$

if you can write P this way

If joint probability distribution:

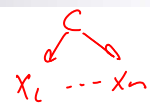
$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain

Then conditional independencies in BN are subset of conditional independencies in *P*

$I_\ell(G) \subseteq I(P)$

Let's start proving it for naïve Bayes – From True Independencies to BN Factorization

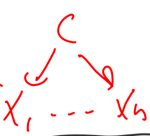


- Independence assumptions:
  - X_i independent given C
- Let's assume that P satisfies independencies must prove that P factorizes according to BN:
  - $P(C,X_1,\ldots,X_n) = P(C) \prod_i P(X_i|C)$
- Use chain rule!

$$P(X_i \mid C \ldots X_1 \ldots X_{i-1}) = P(X_i \mid C)$$

Let's start proving it for naïve Bayes – From BN Factorization to True Independencies

- Let's assume that P factorizes according to the BN:
  - $P(C,X_1,\ldots,X_n) = P(C) \prod_i P(X_i|C)$ ← start from here
- Prove the independence assumptions: prove



  - X_i independent given C
  - Actually, $(\mathbf{X} \perp \mathbf{Y} \mid C)$, $\forall$ $\mathbf{X},\mathbf{Y}$ subsets of $\{X_1,\ldots,X_n\}$

$\mathbf{X} = \{X_1, X_2\}$ $\quad$ $\mathbf{Y} = \{X_3, X_4\}$ $\quad$ $X_1 X_2 \perp X_3 X_4 \mid C$

$$P(X_1 \ldots X_4 \mid C) = \frac{P(X_1 \ldots X_4 C)}{P(C)} = P(X_1|C)\,P(X_2|C)\,P(X_3|C)\,P(X_4|C)$$

$$P(X_1X_2X_3X_4|C) = P(X_1X_2|C) \cdot P(X_3X_4|C)$$

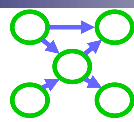
$$P(X_1X_2|C) = \frac{P(X_1X_2C)}{P(C)} = \frac{\sum_{X_3X_4} P(X_1 \ldots X_4 C)}{P(C)} = \frac{1}{\cancel{P(C)}} \sum_{x_3x_4} \cancel{P(C)}\, P(X_1|C)\,P(X_2|C) \cdot P(x_3|C)\,P(x_4|C)$$

$$= P(X_1|C) \cdot P(X_2|C) \sum_{x_3x_4} P(x_3|C) \cdot P(x_4|C) = P(X_1|C)\,P(X_2|C) \Big[\sum_{x_3} \cancel{P(x_3|C)}\Big]\Big[\sum_{x_4} \cancel{P(x_4|C)}\Big]$$

$$P(X_1X_2|C) = P(X_1|C)\,P(X_2|C) \quad \| \quad P(X_3X_4|C) = P(X_3|C)\,P(X_4|C)$$

Today: The Representation Theorem

BN:



Encodes independence assumptions

if BN is an I-MAP

If conditional independencies in BN are subset of conditional independencies in *P*

$I_\ell(G) \subseteq I(P)$

Obtain

P

Joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

P

If joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain

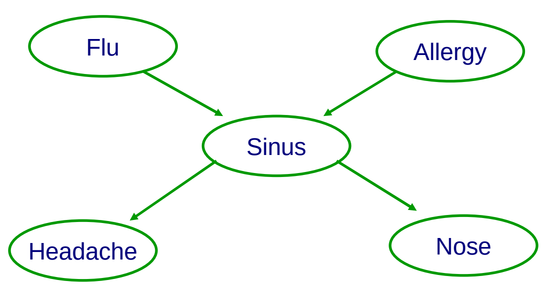
BN I-MAP

Then conditional independencies in BN are subset of conditional independencies in *P*

$I_\ell(G) \subseteq I(P)$

Local Markov assumption & I-maps

- Local independence assumptions in BN structure G: $I_\ell(G)$
- Independence assertions of *P*: $I(P)$
- BN structure G is an ***I-map*** (independence map) if: $I_\ell(G) \subseteq I(P)$

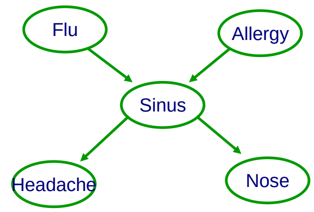


Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents
($X_i \perp \mathrm{NonDescendants}_{X_i} \mid \mathrm{Pa}_{X_i}$)

Factorized distributions

- Given
  - Random vars $X_1,\ldots,X_n$
  - *P* distribution over vars
  - BN structure *G* over same vars
- *P* factorizes according to *G* if

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$



BN Representation Theorem – I-map to factorization

if BN is I MAP

If conditional independencies in BN are subset of conditional independencies in *P*

G is an I-map of P

$I_\ell(G) \subseteq I(P)$

Obtain

P factorizes according to G

Joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

P* factorizes according to *G

BN Representation Theorem – I-map to factorization: **Proof, part 1**

G* is an I-map of *P

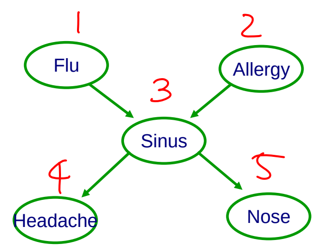
Obtain

P* factorizes according to *G

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Topological Ordering:

- Number variables such that:
  - parent has lower number than child
  - i.e., $X_i \rightarrow X_j \Rightarrow i<j$
  - ***Key: variable has lower number than all of its*** Descendants
- DAGs always have (many) topological orderings
  - find by a modification of breadth first search



BN Representation Theorem – I-map to factorization: **Proof, part 2**

G* is an I-map of *P

Obtain

P* factorizes according to *G

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

- Start with a topological ordering WLOG $X_1, \ldots, X_n$
- Chain rule

$P(X_1 \ldots X_n) = P(X_1)\, P(X_2|X_1) \cdots P(X_n|X_1 \ldots X_{n-1})$

$P(X_i | X_1 \ldots X_{i-1})$

→ I know that $Pa_{X_i} \subseteq X_1 \ldots X_{i-1}$

there are no descendants of X_i in $X_1 \ldots X_{i-1}$ topological ordering

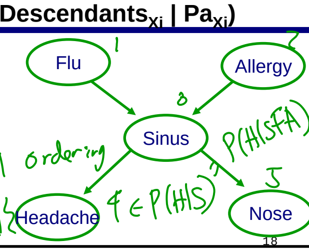
⇒ $P(X_i | X_1 \ldots X_{i-1}) = P(X_i | Pa_{X_i})$

by local Markov assumption

QED

ALL YOU NEED:

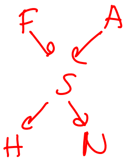
Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents
($X_i \perp \mathrm{NonDescendants}_{X_i} \mid \mathrm{Pa}_{X_i}$)



Defining a BN

estimated from data

- Given a set of variables and conditional independence assertions of P
  $X_1 \ldots X_n$
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
  F A S H N
- For i = 1 to n
  - Add X_i to the network
  - Define parents of X_i, $\mathbf{Pa}_{X_i}$, in graph as the minimal subset of $\{X_1,\ldots,X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1,\ldots,X_{i-1}\}$, given parents $\mathbf{Pa}_{Xi}$
  - Define/learn CPT – $P(X_i \mid \mathbf{Pa}_{Xi})$



$P(H|S,A=t) = P(H|S,A=f) \Rightarrow P(H|S) = P(H|S,A) \Leftrightarrow H \perp A|S \in I(P)$

because $I_\ell(G) \subseteq I(P)$

Adding edges doesn't hurt

- **Theorem**: Let **G** be an I-map for ***P***, any DAG **G**' that includes the same directed edges as **G** is also an I-map for ***P***.
  - **Corollary 1**: G' is strictly more expressive than G
  - **Corollary 2**: If G is an I-map for P, then adding edges still an I-map
- **Proof**:

if $I_\ell(G) \subseteq I(P) \Rightarrow I_\ell(G') \subseteq I(P)$

$I_\ell(G) \supset I_\ell(G')$

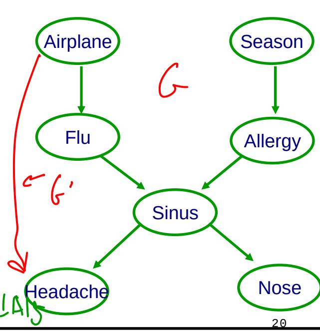
in G $P(X_1 \ldots X_n) = \prod P(X_i|Pa_{X_i})$

added edge $P(H|S)$

enough in G, G' I have $P(H|S,A)$

$P(H|S,A=t) = P(H|S,A=f)$

$\Rightarrow P(H|S) = P(H|S,A) \Leftrightarrow H \perp A|S$



Announcements

- Homework 1:
  - Out today
  - Due in 2 weeks – **beginning of class!**
  - It's hard – start early, ask questions
- Collaboration policy
  - OK to discuss in groups
  - Tell us on your paper who you talked with
  - Each person must write their **own unique paper**
  - No searching the web, papers, etc. for answers, we trust you want to learn
- Audit policy
  - No sitting in, official auditors only, see course website
- Recitation tomorrow
  - Wean 5409, 5pm

BN Representation Theorem – Factorization to I-map

if P factorizes according to BN

If joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain

Then conditional independencies in BN are subset of conditional independencies in *P*

$I_\ell(G) \subseteq I(P)$

P* factorizes according to *G

G* is an I-map of *P

BN Representation Theorem – Factorization to I-map: **Proof**

If joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain →

Then conditional independencies in BN are subset of conditional independencies in *P*

P* factorizes according to *G

G* is an I-map of *P

Homework 1!!!! ☺

The BN Representation Theorem

I-MAP

If conditional independencies in BN are subset of conditional independencies in *P*

Obtain →

factorization

Joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Important because:

we want simplest

Every *P* has at least one BN structure *G*

factorizes

If joint probability distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain →

IMAP

Then conditional independencies in BN are subset of conditional independencies in *P*

Important because:

Read independencies of *P* from BN structure *G*

What you need to know thus far

- Independence & conditional independence
- Definition of a BN
- Local Markov assumption
- The representation theorems
  - Statement: *G* is an I-map for *P* if and only if *P* factorizes according to *G*
  - Interpretation

Independencies encoded in BN

- We said: All you need is the local Markov assumption
  - $(X_i \perp \mathrm{NonDescendants}_{X_i} \mid \mathbf{Pa}_{X_i})$
- But then we talked about other (in)dependencies
  - e.g., explaining away

$A \rightarrow B \rightarrow C \rightarrow D$

$A \perp D \mid B$

$A \rightarrow C \leftarrow B$

$A \perp B$

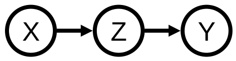
$\neg A \perp B \mid C$

- What are the independencies encoded by a BN?
  - Only assumption is local Markov
  - But many others can be derived using the algebra of conditional independencies!!!

Understanding independencies in BNs – BNs with 3 nodes

Local Markov Assumption:
A variable X is independent of its non-descendants given its parents and only its parents

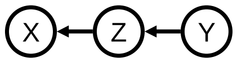
Indirect causal effect:



$Y \perp X \mid Z$

$\neg X \perp Y$

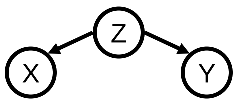
Indirect evidential effect:



$Y \perp X \mid Z$

$\neg X \perp Y$

Common cause:



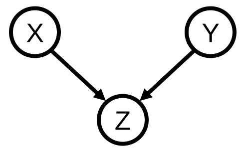
$Y \perp X \mid Z$

$\neg X \perp Y$

all represent same dist.

v-structures

Common effect:



$X \perp Y$

$\neg X \perp Y \mid Z$