Machine Learning 10-601

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

February 23, 2015

Today:

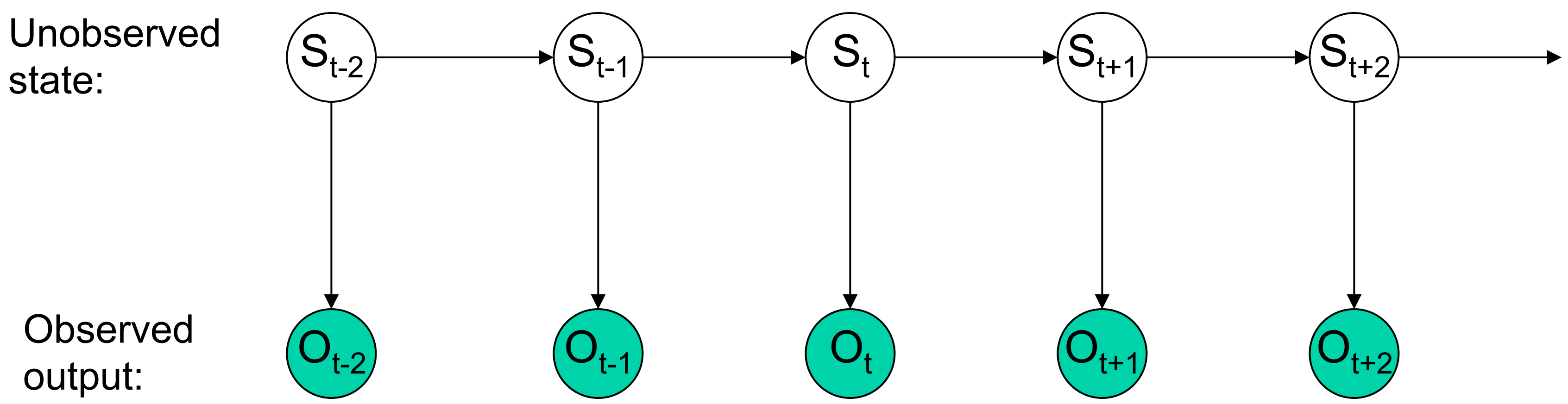
- Graphical models
- Bayes Nets:
  - Representing distributions
  - Conditional independencies
  - Simple inference
  - Simple learning

Readings:

- Bishop chapter 8, through 8.2
- Mitchell chapter 6

Bayes Network for a Hidden Markov Model

Implies the future is conditionally independent of the past, given the present

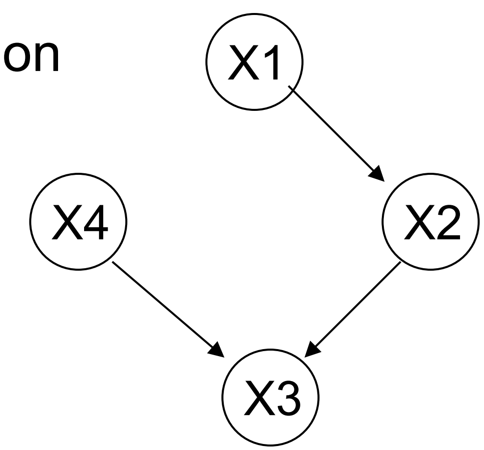


$$P(S_{t-2}, O_{t-2}, S_{t-1}, \ldots, O_{t+2}) =$$

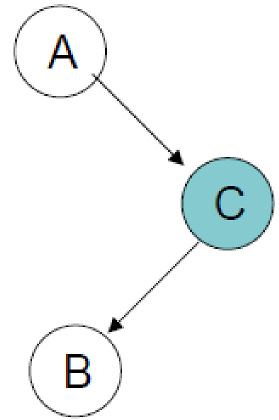
Conditional Independence, Revisited

- We said:
  - Each node is conditionally independent of its non-descendents, given its immediate parents.

- Does this rule give us all of the conditional independence relations implied by the Bayes network?
  - No!
  - E.g., X1 and X4 are conditionally indep given {X2, X3}
  - But X1 and X4 not conditionally indep given X3
  - For this, we need to understand D-separation



Easy Network 1: Head to Tail

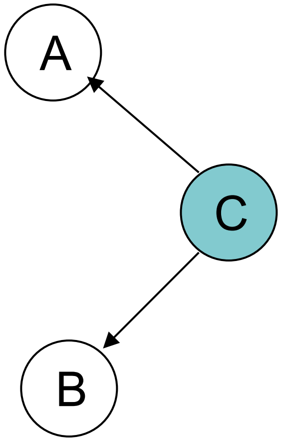


prove A cond indep of B given C?

ie., $p(a,b|c) = p(a|c)\ p(b|c)$

let's use $p(a,b)$ as shorthand for $p(A=a, B=b)$

Easy Network 2: Tail to Tail

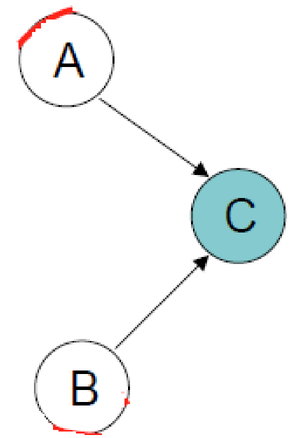


prove A cond indep of B given C?   ie., $p(a,b|c) = p(a|c)\ p(b|c)$

let's use $p(a,b)$ as shorthand for $p(A=a, B=b)$

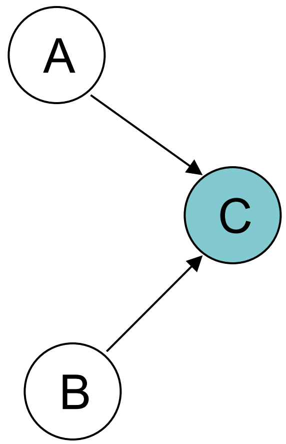
Easy Network 3: Head to Head

prove A cond indep of B given C?   ie., $p(a,b|c) = p(a|c)\, p(b|c)$



let's use $p(a,b)$ as shorthand for $p(A=a, B=b)$

Easy Network 3: Head to Head



prove A cond indep of B given C?   NO!

Summary:

- $p(a,b)=p(a)p(b)$
- $p(a,b|c)$ NotEqual $p(a|c)p(b|c)$

Explaining away.

e.g.,

- A=earthquake
- B=breakIn
- C=motionAlarm

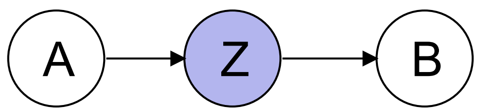
X and Y are conditionally independent given Z, **<u>if and only if</u>** X and Y are D-separated by Z.

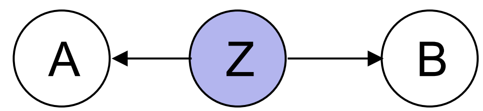
[Bishop, 8.2.2]

Suppose we have three sets of random variables: X, Y and Z

X and Y are **<u>D-separated</u>** by Z (and therefore conditionally indep, given Z) iff every path from every variable in X to every variable in Y is **<u>blocked</u>**

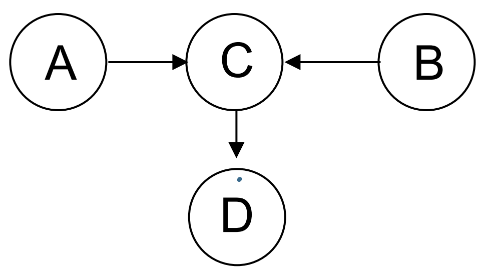
A path from variable X to variable Y is **blocked** if it includes a node in Z such that either





1. arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

2. or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z



X and Y are **D-separated** by Z (and therefore conditionally indep, given Z) iff every path from every variable in X to every variable in Y is **blocked**

A path from variable A to variable B is **blocked** if it includes a node such that either

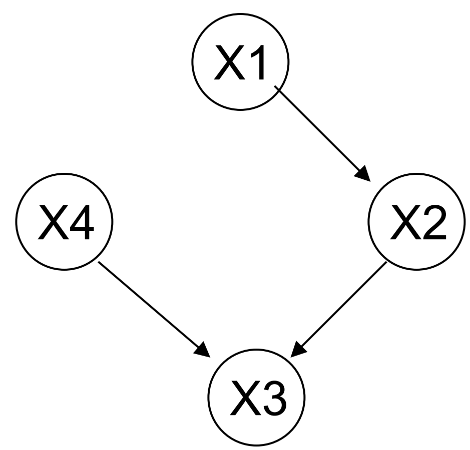
1. arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

2. or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z

X1 indep of X3 given X2?

X3 indep of X1 given X2?

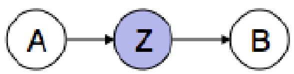
X4 indep of X1 given X2?

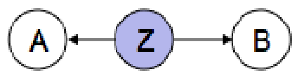


X and Y are **D-separated** by Z (and therefore conditionally indep, given Z) iff every path from any variable in X to any variable in Y is **blocked** by Z

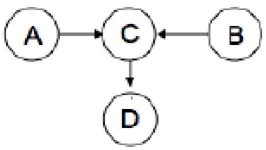
A path from variable A to variable B is **blocked** by Z if it includes a node such that either

1.arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z





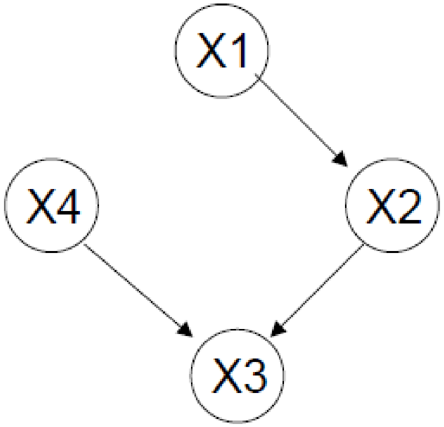
2.the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z



X4 indep of X1 given X3?

X4 indep of X1 given {X3, X2}?

X4 indep of X1 given {}?



X and Y are **D-separated** by Z (and therefore conditionally indep, given Z) iff every path from any variable in X to any variable in Y is **blocked**

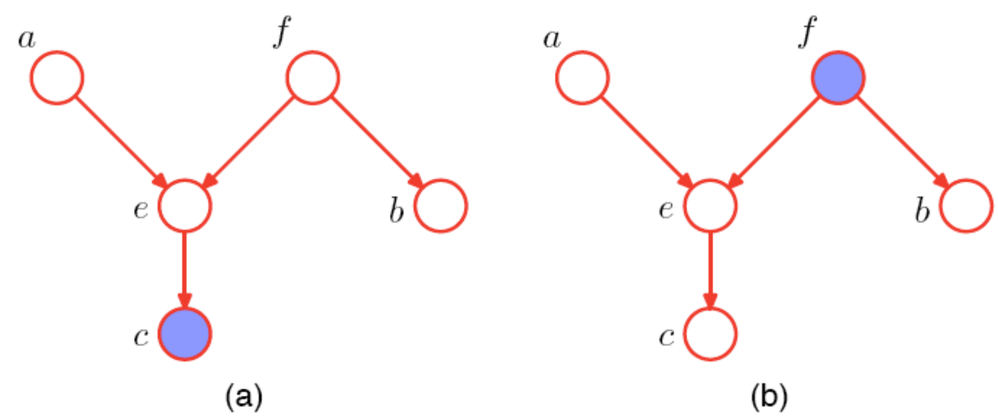
A path from variable A to variable B is **blocked** if it includes a node such that either

1. arrows on the path meet either head-to-tail or tail-to-tail at the node and this node is in Z

2. or, the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in Z

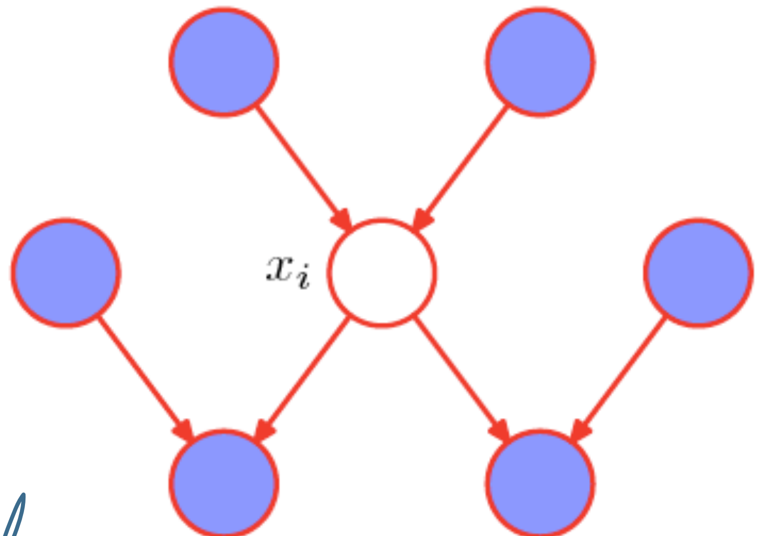
a indep of b given c?

a indep of b given f ?



Markov Blanket

The Markov blanket of a node $\mathbf{x}_i$ comprises the set of parents, children and co-parents of the node. It has the property that the conditional distribution of $\mathbf{x}_i$, conditioned on all the remaining variables in the graph, is dependent only on the variables in the Markov blanket.



co-parent = other side
of x_i's colliders

from [Bishop, 8.2]

What You Should Know

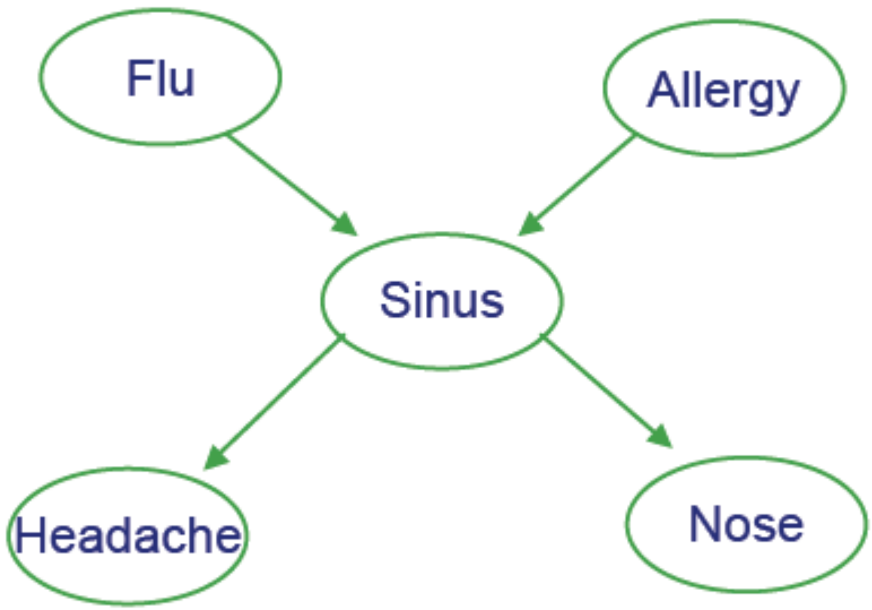
- Bayes nets are convenient representation for encoding dependencies / conditional independence
- BN = Graph plus parameters of CPD's
  - Defines joint distribution over variables
  - Can calculate everything else from that
  - Though inference may be intractable
- Reading conditional independence relations from the graph
  - Each node is cond indep of non-descendents, given only its parents
  - X and Y are conditionally independent given Z if Z D-separates every path connecting X to Y
  - Marginal independence : special case where Z={}

Inference in Bayes Nets

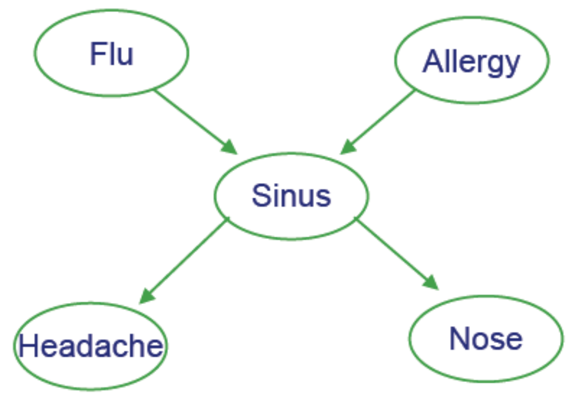
- In general, intractable (NP-complete)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
  - Or for singly connected graphs (ie., no undirected loops)
    - Belief propagation
- Sometimes use Monte Carlo methods
  - Generate many samples according to the Bayes Net distribution, then count up the results
- Variational methods for tractable approximate solutions

Example

- Bird flu and Allegies both cause Sinus problems
- Sinus problems cause Headaches and runny Nose



Prob. of joint assignment: easy

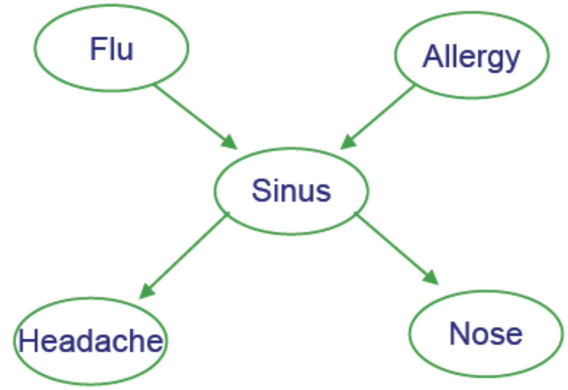


- Suppose we are interested in joint assignment <F=f,A=a,S=s,H=h,N=n>

What is P(f,a,s,h,n)?

let's use p(a,b) as shorthand for p(A=a, B=b)

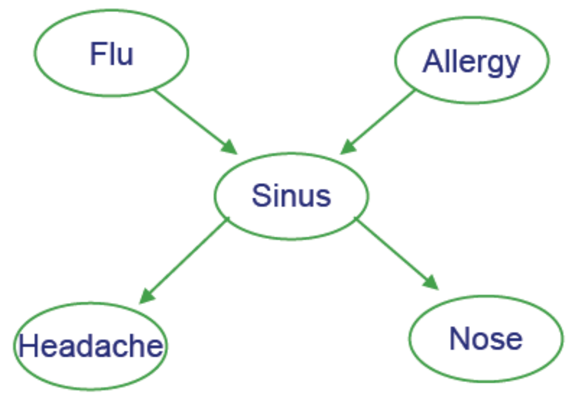
Prob. of marginals: not so easy



- How do we calculate P(N=n) ?

let's use p(a,b) as shorthand for p(A=a, B=b)

Generating a sample from joint distribution: easy



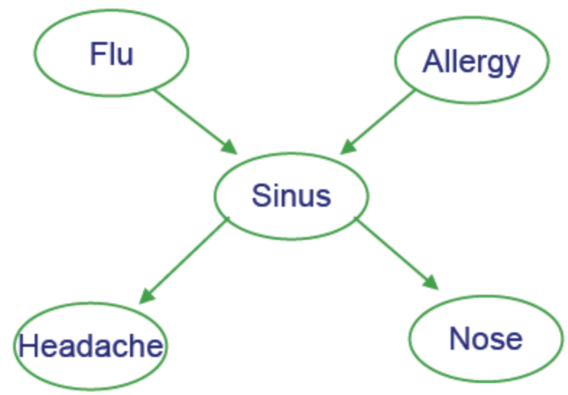
How can we generate random samples drawn according to P(F,A,S,H,N)?

Hint: random sample of F according to $P(F=1) = \theta_{F=1}$:

- draw a value of r uniformly from [0,1]
- if $r<\theta$ then output F=1, else F=0

let's use p(a,b) as shorthand for p(A=a, B=b)

Generating a sample from joint distribution: easy



How can we generate random samples drawn according to P(F,A,S,H,N)?

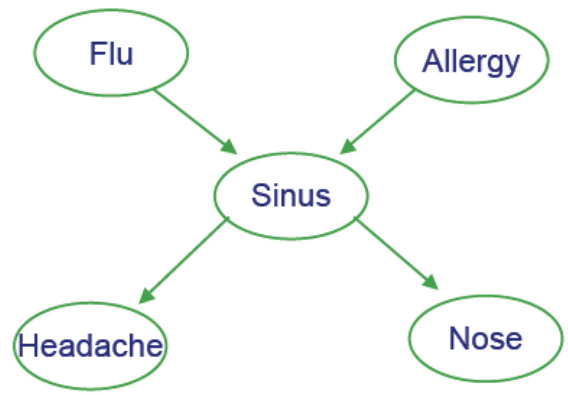
Hint: random sample of F according to $P(F=1) = \theta_{F=1}$:

- draw a value of r uniformly from [0,1]
- if $r<\theta$ then output F=1, else F=0

Solution:

- draw a random value f for F, using its CPD
- then draw values for A, for S|A,F, for H|S, for N|S

Generating a sample from joint distribution: easy



Note we can estimate marginals

like P(N=n) by generating many samples

from joint distribution, then count the fraction of samples for which N=n

Similarly, for anything else we care about P(F=1|H=1, N=0)

→ weak but general method for estimating <u>any</u> probability term…

Inference in Bayes Nets

- In general, intractable (NP-complete)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
  - Or for singly connected graphs (ie., no undirected loops)
    - Variable elimination
    - Belief propagation
- Often use Monte Carlo methods
  - e.g., Generate many samples according to the Bayes Net distribution, then count up the results
  - Gibbs sampling
- Variational methods for tractable approximate solutions

see Graphical Models course 10-708