Graphical Models

Pradeep Ravikumar

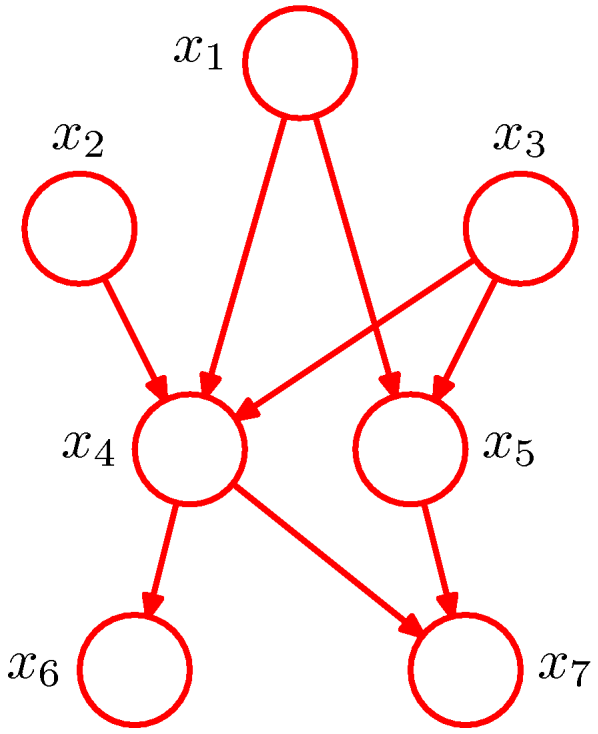
Co-instructor: Manuela Veloso

Machine Learning 10-701

Some Slides courtesy of Carlos Guestrin

iid to dependent data

- Graphical Models
  - general dependence



Applications

- Speech recognition
- Diagnosis of diseases
- Study Human genome
- Robot mapping
- Modeling fMRI data
- Fault diagnosis
- Modeling sensor network data
- Modeling protein-protein interactions
- Weather prediction
- Computer vision
- Statistical physics
- Many, many more …

Joint Distributions

- Consider a set of random variables $\{X_1, X_2, \ldots, X_n\}$
- Let x_i be a realization of random variable X_i; x_i may be real, discrete, vector in vector space; for now assume variables are discrete.
- Probability Mass Function $p(x_1, \ldots, x_n) := P(X_1 = x_1, \ldots, X_n = x_n)$.
- Shorthand: $X = (X_1, \ldots, X_n)$, $x = (x_1, \ldots, x_n)$, so that $p(x) = P(X = x)$.

Joint Distributions and Tables

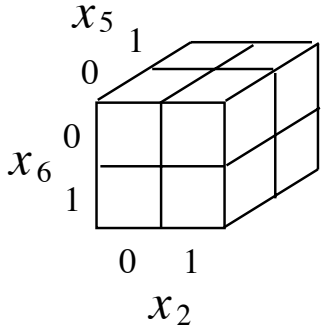
- Graphical Models are an answer to the following key question: how do we represent the joint distribution p(x)?

- One way is by using an n-dimensional table

  - a separate cell for each setting of variables {x_1, ... x_n}, with value p(x_1,...,x_n)

  - If each variable takes r values, we need to store r^n values : exponential in n

  - For large values of r,n ; such a representation is out!



Graphical Models: The Why

- Compact way of representing distributions among random variables
- Visually appealing way (accessible to domain experts) of representing distributions among random variables
- They represent distributions among random variables — probability theory
- They use graphs to do so — graph theory
    - A marriage of graph and probability theory

Graphical Models

- Key Idea:
  - **Conditional independence (CI)** assumptions useful
  - Recall Naïve Bayes uses CI assumptions, but is extreme!
  - Graphical models express more flexible sets of conditional independence assumptions via graph structure

- Two types of graphical models:
  - Directed graphs (aka Bayesian Networks)
  - Undirected graphs (aka Markov Random Fields)

Topics in Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

Topics in Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

Representation of Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?
- Can be characterized in two ways
  - using graph theory, or using probability theory; and both are equivalent!
  - Graphical models are thus said to draw from a marriage of graph theory and probability theory
  - Graph Theory View of representation: The graph specifies the factorization i.e. algebraic structure of joint distribution
  - Probability Theory View of representation: The graph specifies conditional independence assumptions satisfied by joint distribution

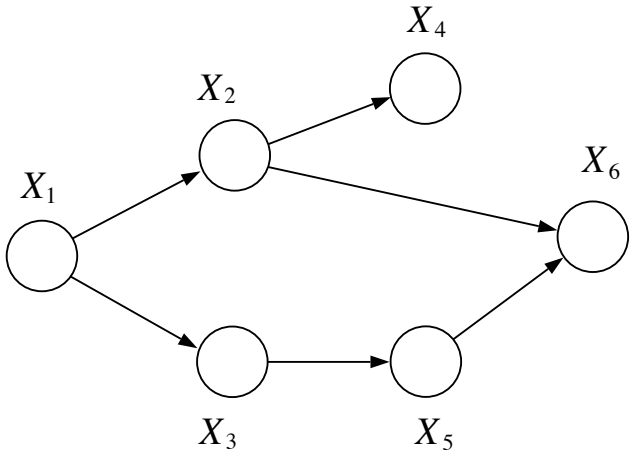
Directed - Bayesian Networks

- Representation
  - Which joint probability distributions does a graphical model represent?
  - **Factorization View**

Given a directed acyclic graph (DAG) a BN specifies the following joint distribution

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

Directed Graphs and Joint Probabilities



- Directed Graph is a pair $G = (V, E)$, where V is a set of nodes, E is a set of directed (also called oriented) edges. We will assume G is acyclic.

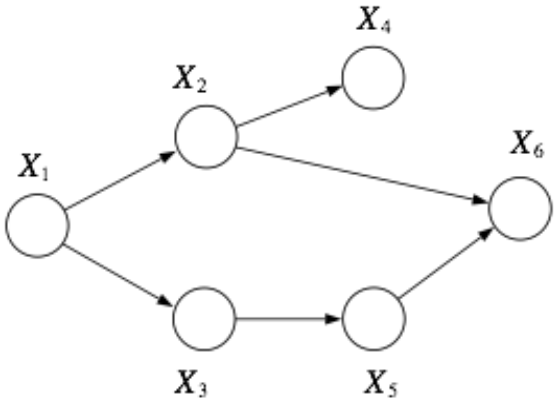
- Each node $i \in V$ is associated with a random variable X_i.

- Letting $V = \{1, 2, \ldots, n\}$, the set of random variables is $\{X_1, X_2, \ldots, X_n\}$.

- We will use node i and the associated random variable X_i interchangeably (though graph nodes and random variables are different formal objects!)

Factorization View of BNs

- Each node $i \in V$ has a set of parents π_i
  - $\pi_6 = \{X_2, X_5\}$.

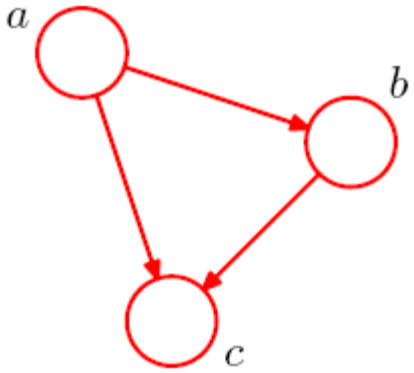


$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_{\pi_i}).$$

Directed Graphical Models: **set** of distributions
that **factorize** according to a given directed acyclic graph G

Directed - Bayesian Networks

- Representation
  - Which joint probability distributions does a graphical model represent?
  - **Conditional Independence View**



Recall: Conditional Independence

- X is **conditionally independent** of Y given Z:

  probability distribution governing X is independent of the value of Y, given the value of Z

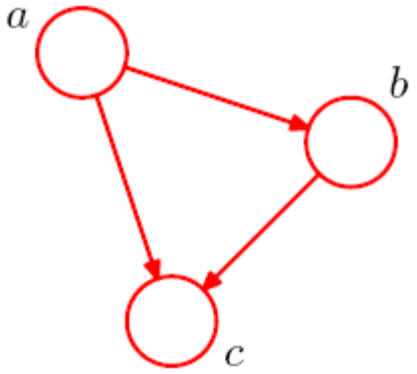
$$(\forall x, y, z) P(X = x | Y = y, Z = z) = P(X = x | Z = z)$$

- Compactly: $P(X \mid Y, Z) = P(X \mid Z)$

- Equivalent to: $P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z)$

Directed - Bayesian Networks

- Representation
  - Which joint probability distributions does a graphical model represent?
  - **Conditional Independence View**



For any arbitrary distribution,

Chain rule:

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

More generally:

$$p(\mathbf{X}) = \prod_{i=1}^{n} p(X_n|X_{n-1}, \ldots, X_1)$$

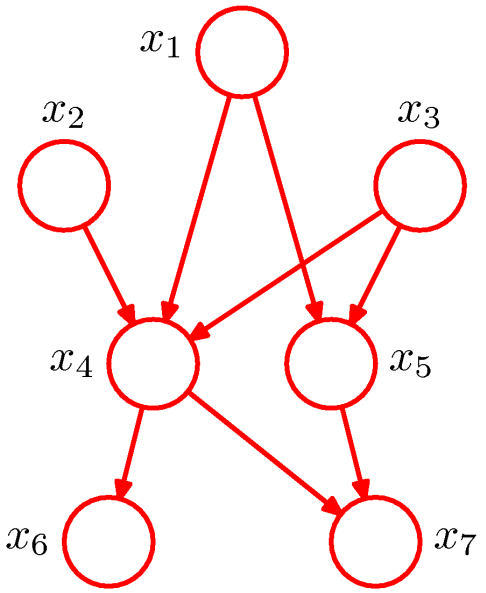
Fully connected directed graph between X_1, …, X_n

Directed - Bayesian Networks

- Representation
  - Which joint probability distributions does a graphical model represent?

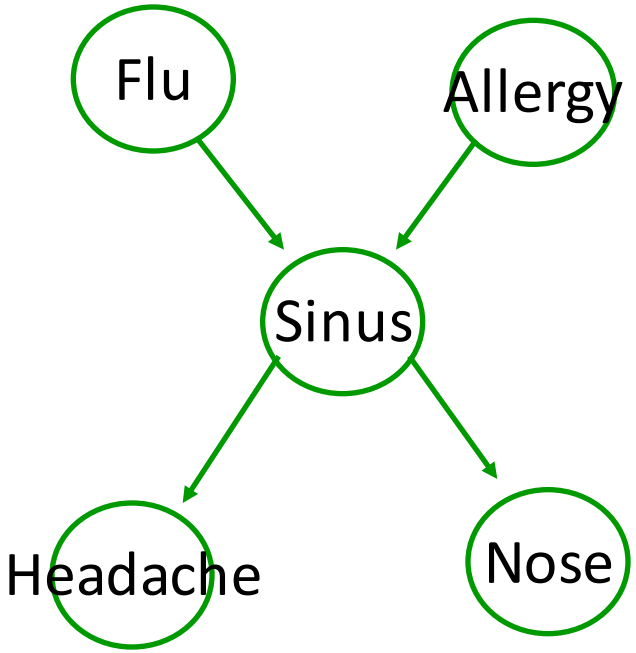
Absence of edges in a graphical model conveys useful information.

$$p(x_1, x_2, \ldots, x_6) =$$
$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)$$
$$p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$



Bayesian Networks Example

- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches



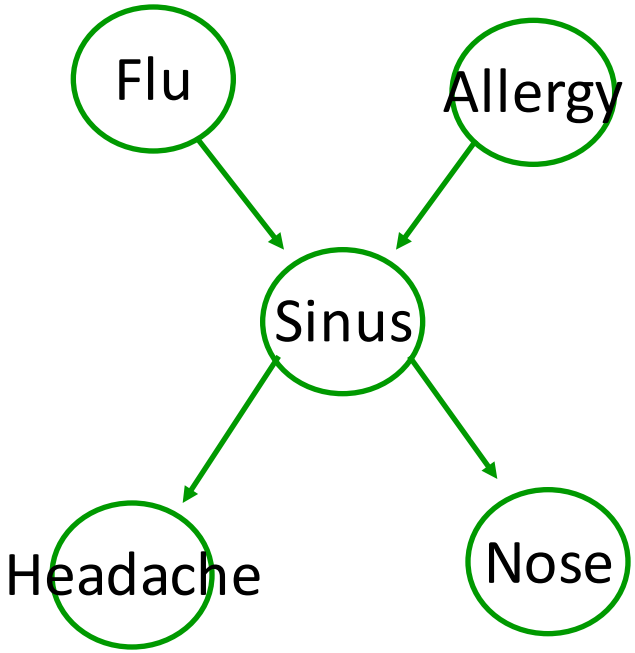
- We have a Directed Acyclic Graph

- **Local Markov Assumption:** If you have no sinus infection, then flu has no influence on headache (flu causes headache but only through sinus)

Markov independence assumption

Local Markov Assumption: A variable X is independent of its (non-descendants – parents) given its parents

	parents	non-desc	assumption
S	F,A	-	-
H	S	F,A,N	$H \perp \{F,A,N\} \mid S$
N	S	F,A,H	$N \perp \{F,A,H\} \mid S$
F	-	A	$F \perp A$
A	-	F	$A \perp F$



Markov independence assumption

Local Markov Assumption: A variable X is independent of its (non-descendants – parents) given its parents

Joint distribution:

$P(F, A, S, H, N)$

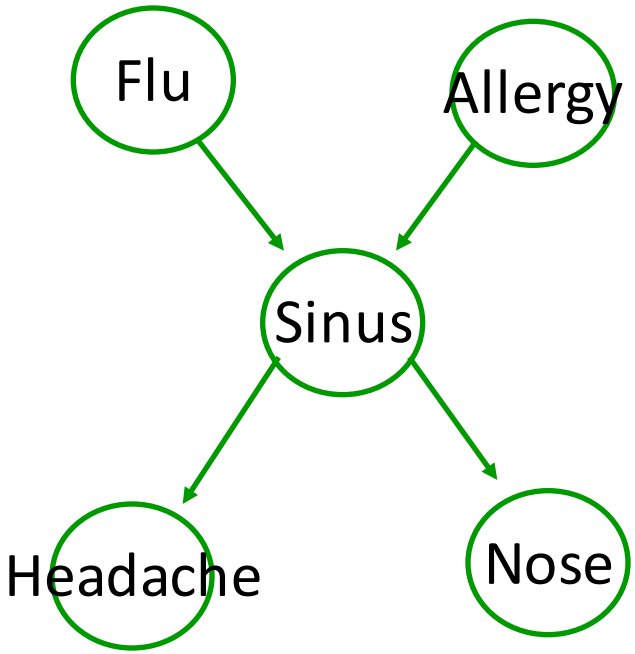
$= P(F)\ P(A|F)\ P(S|F,A)\ P(H|S,F,A)\ P(N|S,F,A,H)$

Chain rule

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

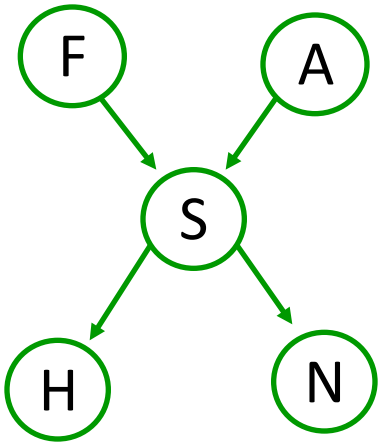
Markov Assumption

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



How many parameters in a BN?

- Discrete variables $X_1, \ldots, X_n$
- Directed Acyclic Graph (DAG)
  - Defines parents of X_i, $\mathbf{Pa}_{X_i}$
- CPTs (Conditional Probability Tables)
  - $P(X_i \mid \mathbf{Pa}_{Xi})$



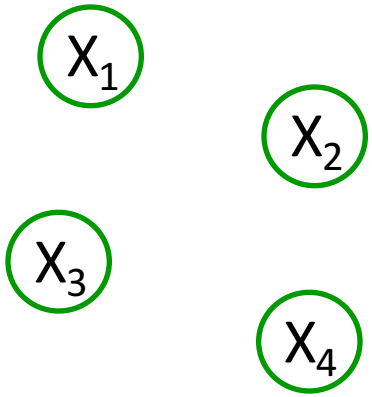
E.g. X_i = S, $\mathbf{Pa}_{Xi}$ = {F, A}

	F=f, A=f	F=t, A=f	F=f, A=t	F=t,A=t
S=t	0.9	0.8	0.7	0.3
S=f	0.1	0.2	0.3	0.7

n variables, K values, max d parents/node $O(n \times K \times K^d)$

Two (trivial) special cases

Fully disconnected graph



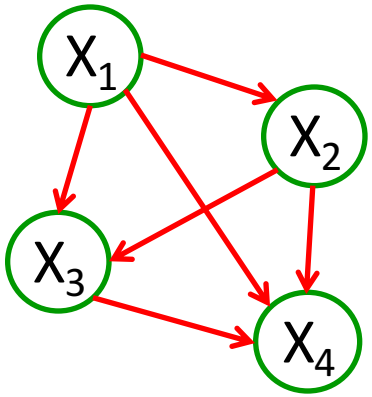
X_i

parents: ϕ

non-descendants: $X_1,\ldots,X_{i-1}, X_{i+1},\ldots, X_n$

$X_i \perp X_1,\ldots,X_{i-1},X_{i+1},\ldots, X_n$

Fully connected graph



X_i

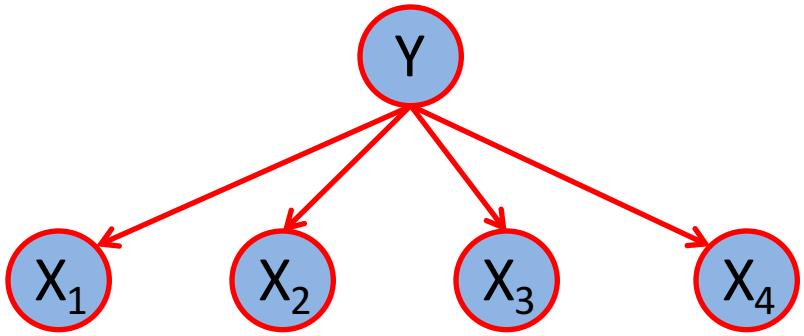
parents: $X_1, \ldots, X_{i-1}$

non-descendants: ϕ

No independence assumption

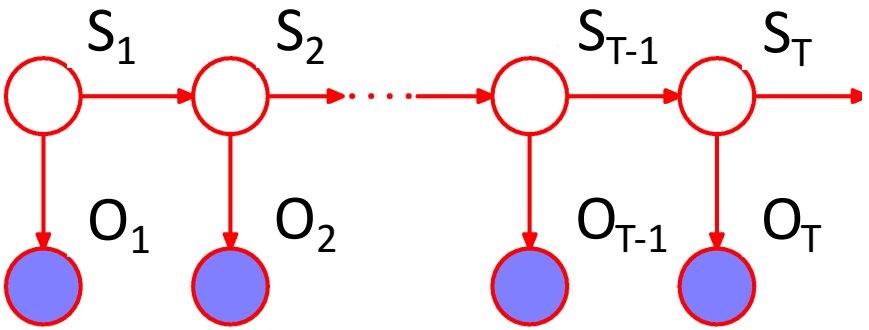
Bayesian Networks Example

- Naïve Bayes $X_i \perp X_1,\ldots,X_{i-1},X_{i+1},\ldots, X_n \mid Y$



$P(X_1,\ldots,X_n,Y) =$
$P(Y)P(X_1|Y)\ldots P(X_1|Y)$

- HMM



$$p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) = p(S_1)\prod_{t=2}^T p(S_t|S_{t-1})\prod_{t=1}^T p(O_t|S_t)$$

Explaining Away

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

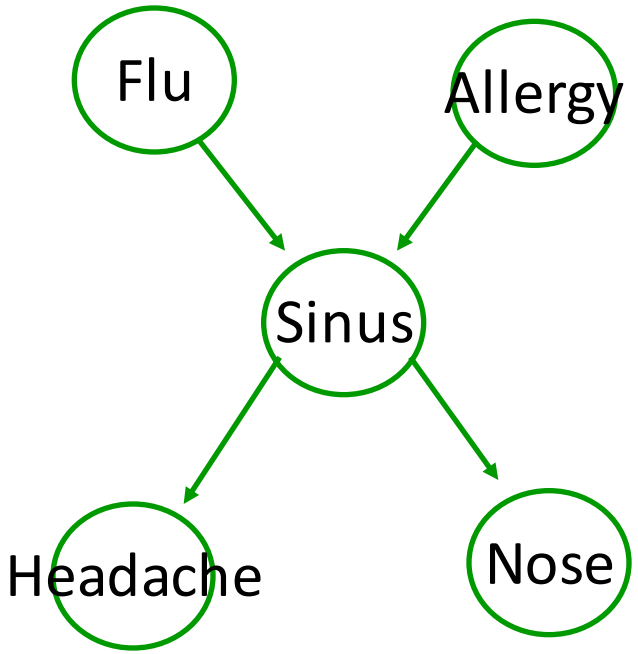
$F \perp A$ $\quad P(F|A=t) = P(F)$

$F \perp A|S$?
$P(F|A=t,S=t) = P(F|S=t)$? **No!**

$P(F=t|S=t)$ is high,
but $P(F=t|A=t,S=t)$ not as high
since A = t explains away S=t

Infact, $P(F=t|A=t,S=t) < P(F=t|S=t)$

$F \perp A|N$? **No!**



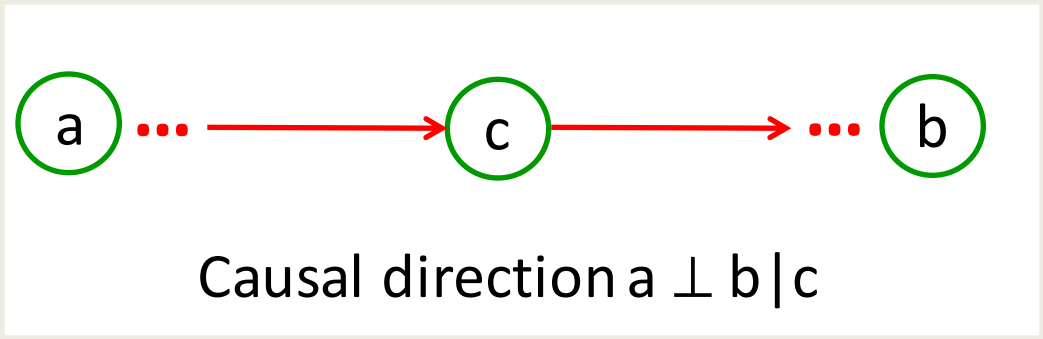
Independencies encoded in BN

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

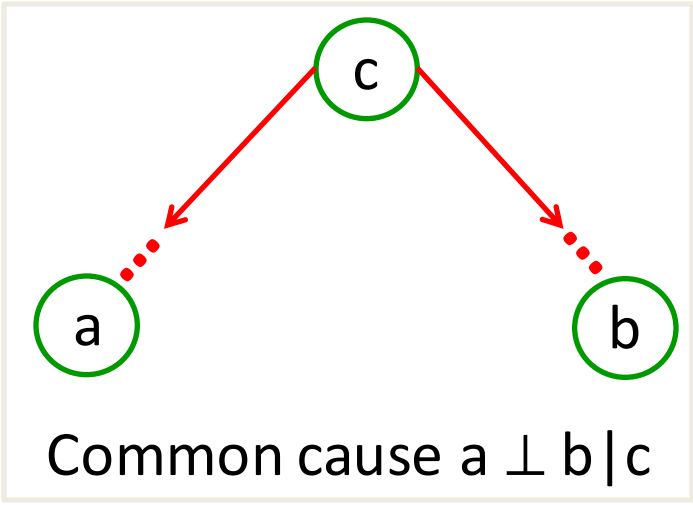
- What are the independencies encoded by a BN?
  - Only assumption is local Markov
  - But many others can be derived using the algebra of conditional independencies!!!

D-separation

- a is D-separated from b by c iff $a \perp b \mid c$
- Three important configurations

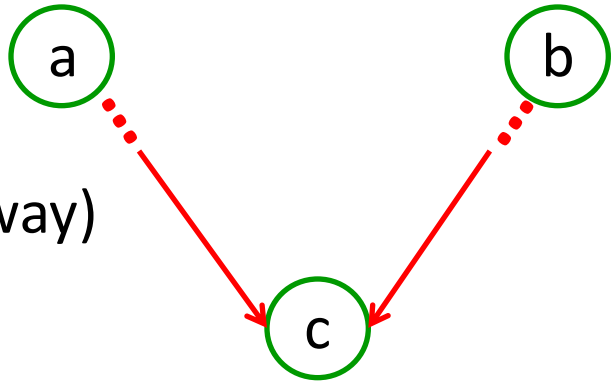


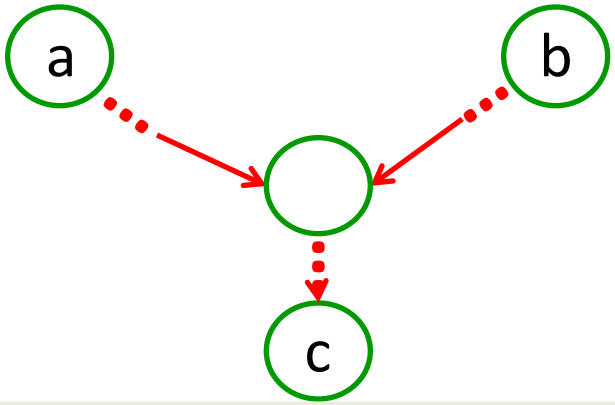
Causal direction $a \perp b \mid c$



Common cause $a \perp b \mid c$

V-structure
(Explaining away)
$a \not\perp b \mid c$





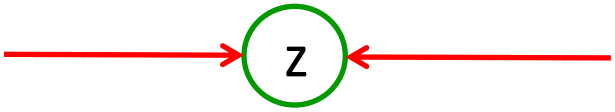
D-separation

- A, B, C – non-intersecting set of nodes
- A is D-separated from B by C
  iff $A \perp B \mid C$
  iff all paths between nodes in A & B are "blocked"
  iff Path contains a node z such that

I. z in C, and

OR

II. neither z nor any of its descendants is in C, and

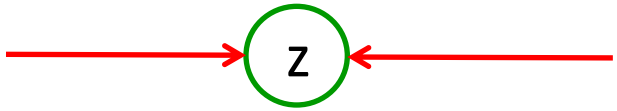


D-separation Example

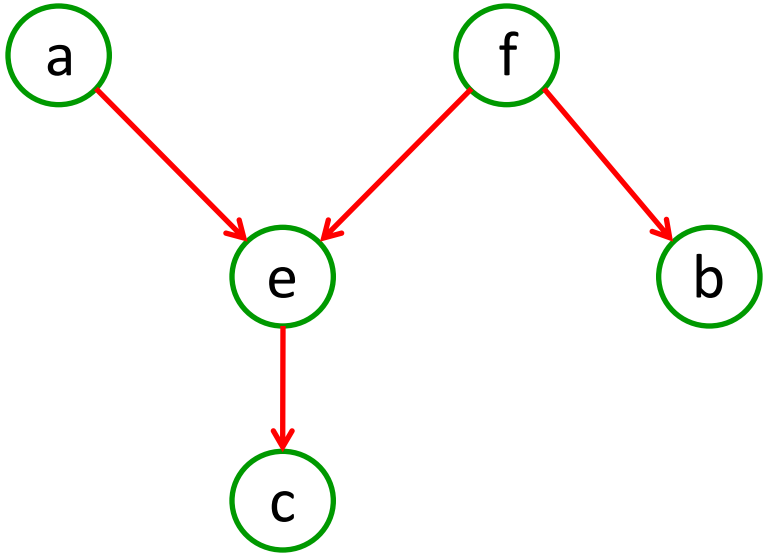
A is D-separated from B by C if every path between A and B contains a node z such that either

And z in C

or



And neither z nor its descendants are in C



$a \perp b \mid f$?

Yes, Consider z = f or z = e

$a \perp b \mid c$?

No, Consider z = e

Representation Theorem

- Set of distributions that **factorize** according to the graph - F
- Set of distributions that respect **conditional independencies** implied by d-separation properties of graph – I

Important because: **Given independencies of *P* can get BN structure *G***

Important because: **Read independencies of *P* from BN structure *G***

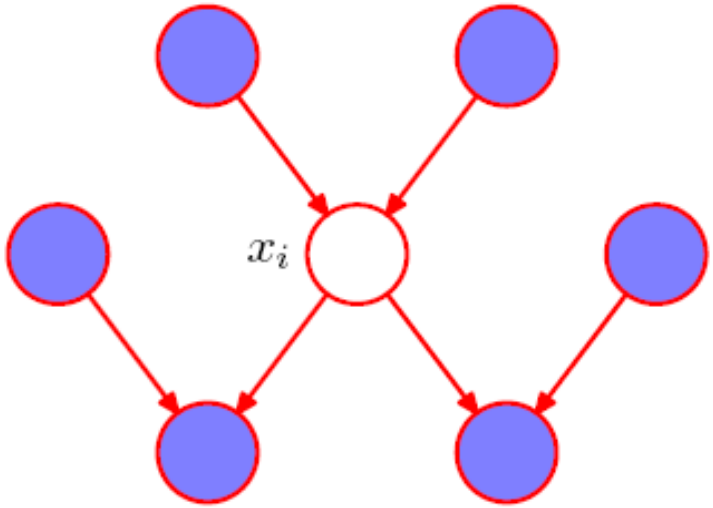
Markov Blanket

- Conditioning on the Markov Blanket, node i is independent of all other nodes.

$$p(\mathbf{x}_i|\mathbf{x}_{\{j\neq i\}}) = \frac{p(x_1,\ldots,x_n)}{\sum_i p(x_1,\ldots,x_n)} = \frac{\prod_k p(x_k|pa(x_k))}{\sum_i \prod_k p(x_k|pa(x_k))}$$

Only terms that remain are the ones which involve i

$$p(x_i|pa(x_i)) \qquad p(x_k|pa(x_k) \ni i)$$



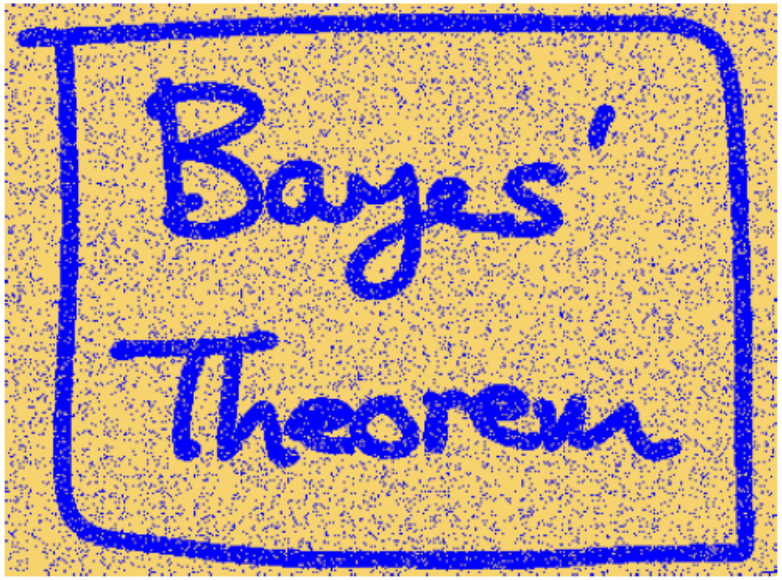
- Markov Blanket of node i - Set of parents, children and co-parents of node i

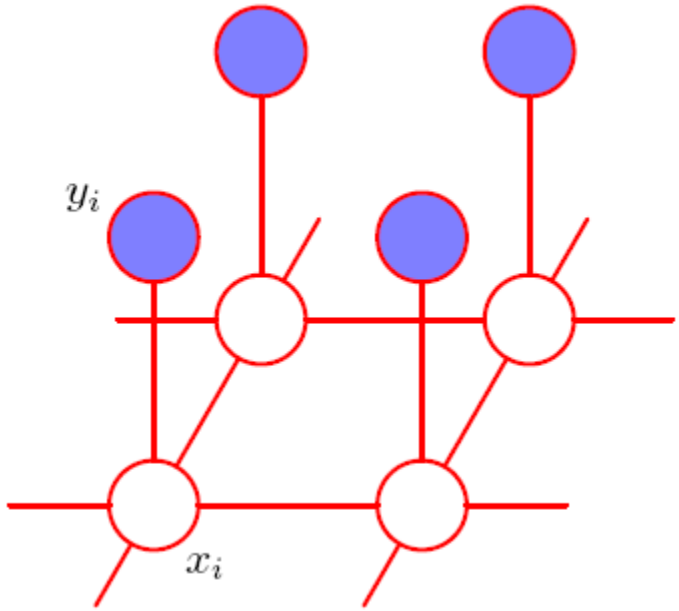
Undirected – Markov Random Fields

- Popular in statistical physics, computer vision, sensor networks, social networks, protein-protein interaction network
- Example – Image Denoising

x_i – value at pixel i

y_i – observed noisy value

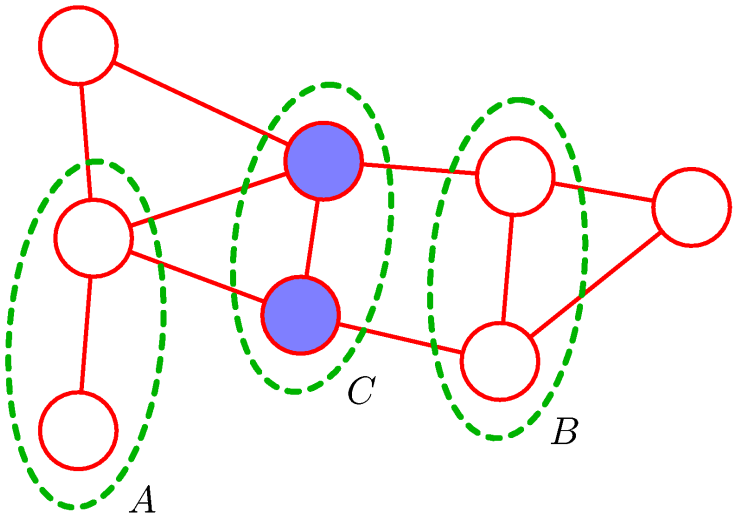




Representation of MRFs: Conditional Independence View

- No directed edges
- Conditional independence dictated by graph separation (much simpler than “d separation”)
- A, B, C – non-intersecting set of nodes
- $A \perp B \mid C$ if all paths between nodes in A & B are “blocked”

  i.e. path contains a node z in C.



Representation of MRFs: Factorization View

- Joint distribution factorizes according to the graph

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

$\mathcal{C}$ is the set of maximal cliques in the graph

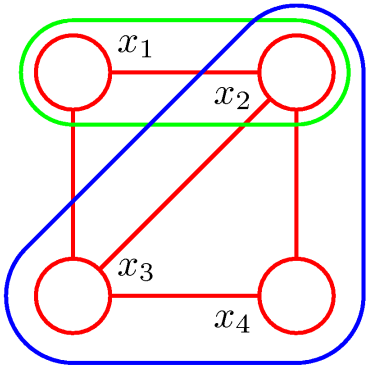
$\psi_C(x_C)$ is a potential function on the clique x_C

→ Arbitrary positive function

normalization factor

$$Z = \sum_x \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

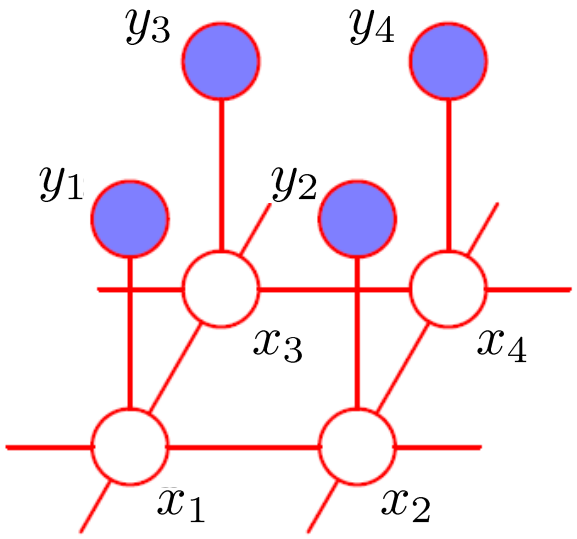
typically NP-hard to compute



Clique, $x_C = \{x_1, x_2\}$

Maximal clique
$x_C = \{x_2, x_3, x_4\}$

MRF Example



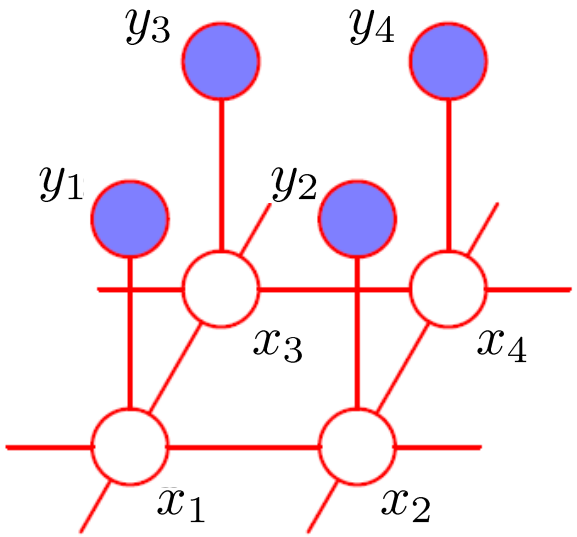
$$P(x,y) \propto \Psi(x_1,x_2)\Psi(x_1,x_3)\Psi(x_2,x_4)\Psi(x_3,x_4)\prod_{i=1}^{4}\Psi(x_i,y_i)$$

Often $\psi_C(\mathbf{x}_C) = \exp\{-E(\mathbf{x}_C)\}$

→ Energy of the clique

$$p(\mathbf{x}) = \prod_{C\in\mathcal{C}} \exp\{-E(\mathbf{x}_C)\} = \exp\{-\sum_{C\in\mathcal{C}} E(\mathbf{x}_C)\}$$

MRF Example



Ising model:

cliques are edges $x_C = \{x_i, x_j\}$
binary variables $x_i \in \{-1, 1\}$

$$\psi_C(\mathbf{x}_C) = \exp\{\beta x_i x_j\}$$

1 if $x_i = x_j$
-1 if $x_i \neq x_j$

$$p(\mathbf{x}) = \prod_{(i,j)\in E} \exp\{\beta x_i x_j\} = \exp\{\sum_{(i,j)\in E} \beta x_i x_j\}$$

Probability of assignment is higher if neighbors x_i and x_j are same

Hammersley-Clifford Theorem

- Set of distributions that factorize according to the graph - F
- Set of distributions that respect conditional independencies implied by graph-separation – I

I ⇨ F

Important because: **Given independencies of *P* can get MRF structure *G***

I ⇦ F

Important because: **Read independencies of *P* from MRF structure *G***

What you should know…

- Graphical Models: Directed Bayesian networks, Undirected Markov Random Fields
  - A compact **representation** for large probability distributions
  - Two different (but equivalent) views of representation of graphical models: conditional independence (graph separation or d-separation) & factorization
  - Each graph is a set of distributions (that factor a certain way, or equivalently that satisfy a set of conditional independence constraints)

Topics in Graphical Models

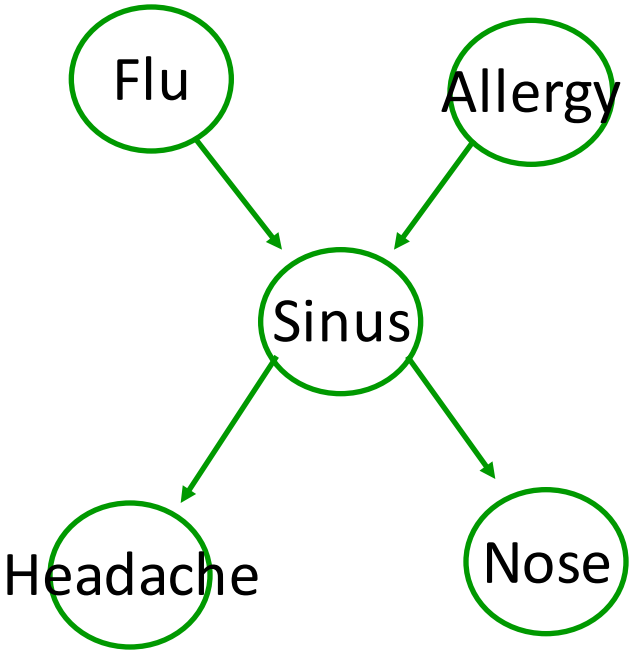
- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

Inference

- Possible queries:

1) Marginal distribution e.g. P(S)

   Posterior distribution e.g. P(F|H=1)



2) Most likely assignment of nodes

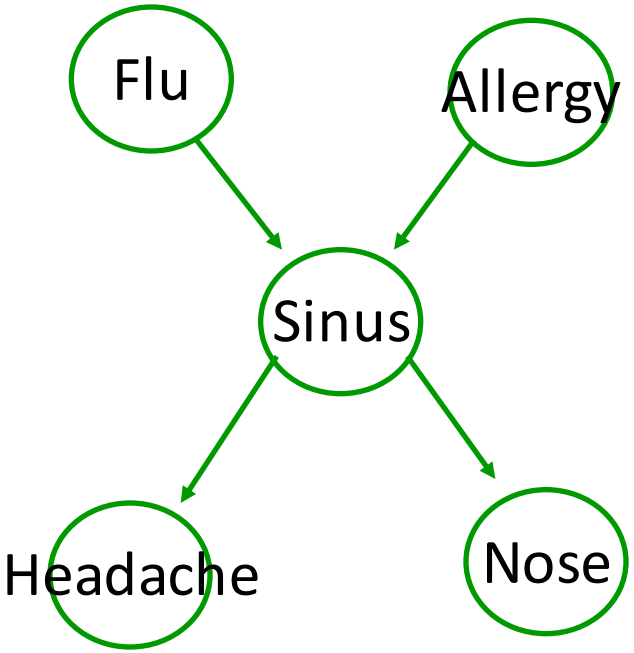
$$\arg\max_{f,a,s,n} P(F=f,A=a,S=s,N=n|H=1)$$

Inference

- Possible queries:

1) Marginal distribution e.g. P(S)

   Posterior distribution e.g. P(F|H=1)



P(F|H=1) ?

$$P(F|H=1) = \frac{P(F, H=1)}{P(H=1)}$$

$$= \frac{P(F, H=1)}{\sum_f P(F=f,H=1)}$$

$$\propto P(F, H=1)$$

will focus on computing this, posterior will follow with only constant times more effort

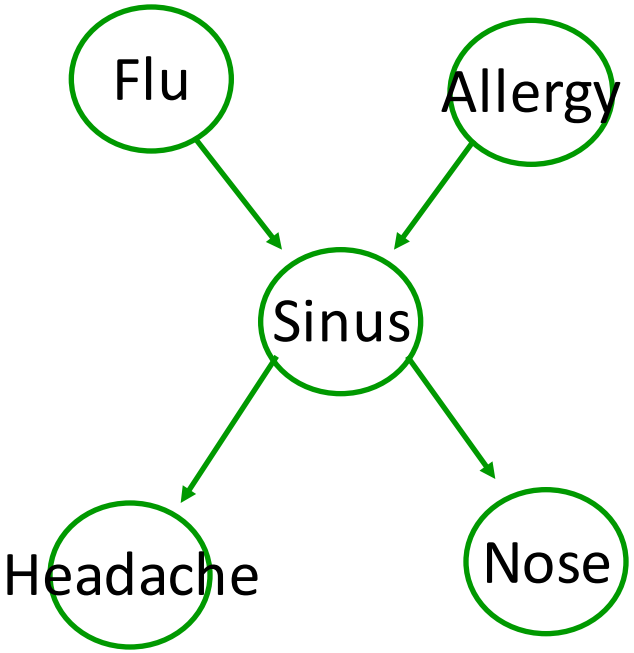
Marginalization

Need to marginalize over other vars

$$P(S) = \sum_{f,a,n,h} P(f,a,S,n,h)$$

$$P(F,H=1) = \sum_{a,s,n} P(F,a,s,n,H=1)$$

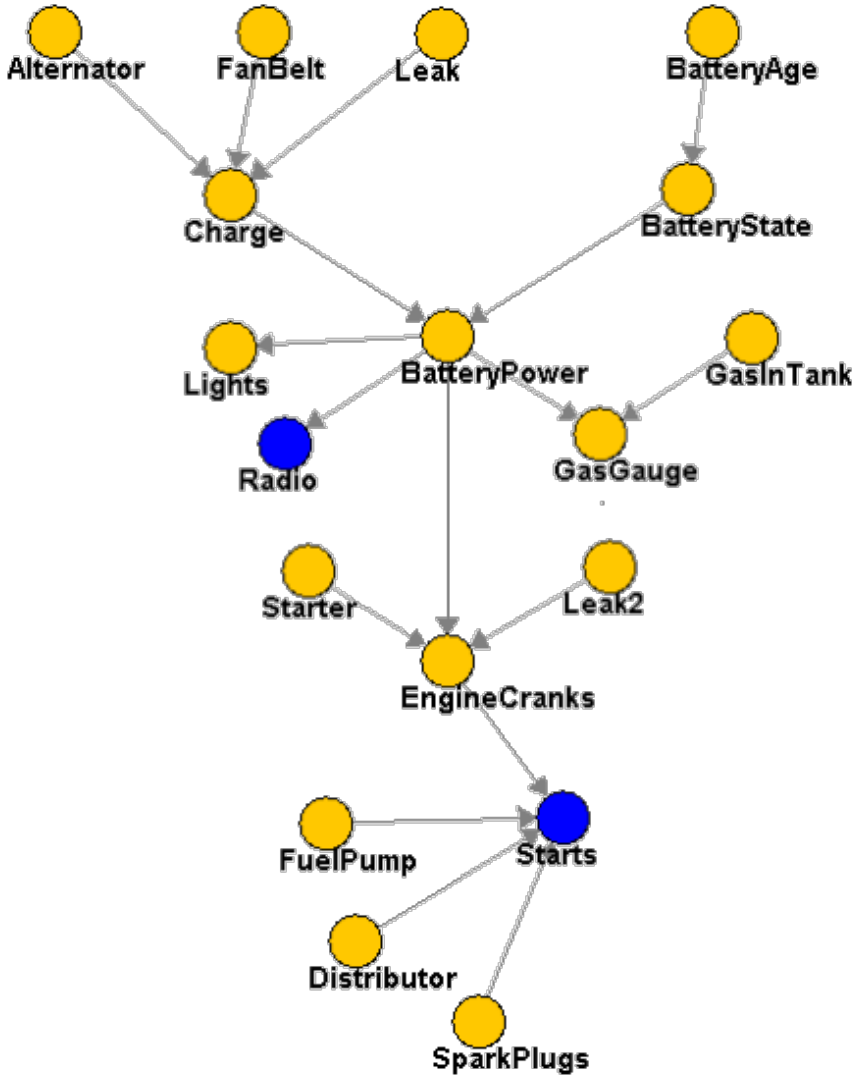
2^3 terms



To marginalize out n binary variables, need to sum over 2^n terms

Inference seems exponential in number of variables!
Actually, inference in graphical models is NP-hard ☹

Bayesian Networks Example



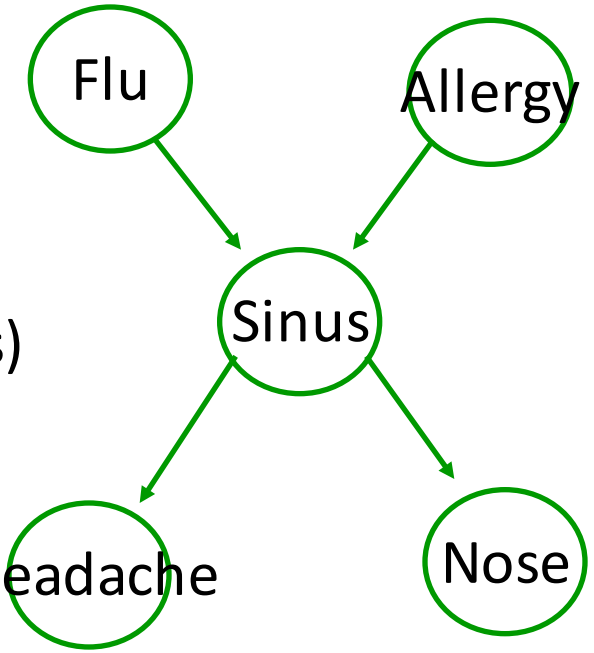
- 18 binary attributes
- Inference
  - P(BatteryAge|Starts=f)
- need to sum over 2^{16} terms!
- Not impressed?
  - HailFinder BN – more than 3^{54} = 58149737003040059690390169 terms

Fast Probabilistic Inference

$$P(F,H=1) = \sum_{a,s,n} P(F,a,s,n,H=1)$$

$$= \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s) \sum_{n} P(n|s)$$



Push sums in as far as possible

Distributive property: $x_1z + x_2z = z(x_1+x_2)$

2 multiply 1 multiply

Fast Probabilistic Inference

$$P(F,H=1) = \sum_{a,s,n} P(F,a,s,n,H=1)$$

8 values x 4 multiplies

$$= \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s) \sum_{n} P(n|s)$$

$\sum_{n} P(n|s) = 1$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s)$$

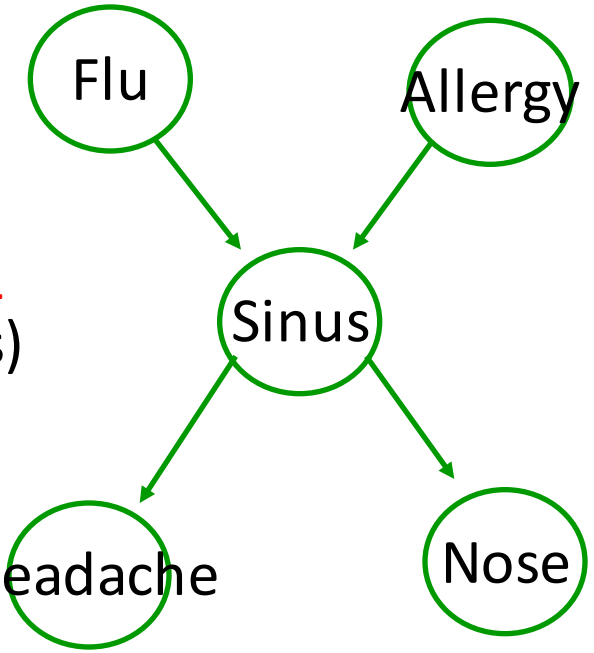
4 values x 1 multiply

$$= P(F) \sum_{a} P(a)\, g_1(F,a)$$

2 values x 1 multiply

$$= P(F)\, g_2(F)$$

1 multiply



32 multiplies vs. 7 multiplies

2^n vs. $n\,2^k$

k – scope of largest factor

(Potential for) exponential reduction in computation!

Fast Probabilistic Inference – Variable Elimination

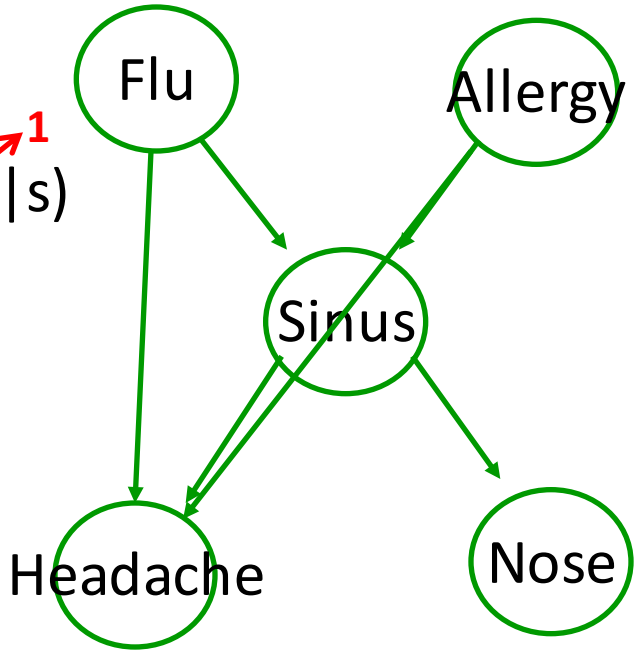
$$P(F,H=1) = \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s) \sum_{n} P(n|s)$$

$\sum_{n} P(n|s) = 1$

$g_1(F,a)$

$g_2(F)$



(Potential for) exponential reduction in computation!

Variable Elimination – Order can make a HUGE difference

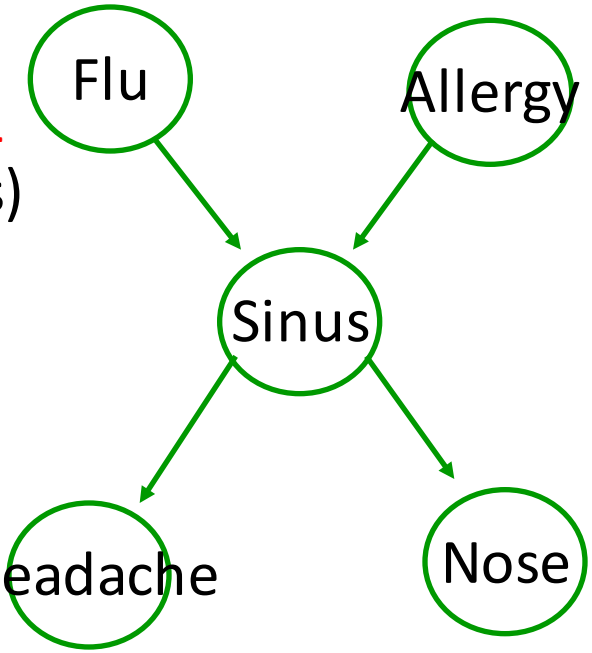
$$P(F,H=1) = \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s) \sum_{n} P(n|s)$$

$\sum_{n} P(n|s) = 1$

$g_1(F,a) = \sum_{s} P(s|F,a)P(H=1|s)$

$g_2(F) = \sum_{a} P(a)\, g_1(F,a)$



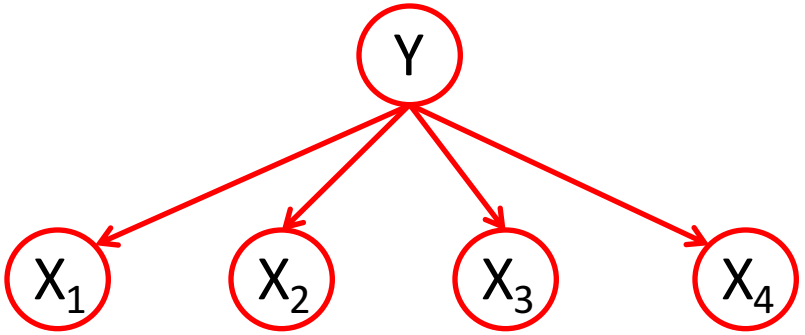
$$P(F,H=1) = P(F) \sum_{a} P(a) \sum_{n} \sum_{s} P(s|F,a)P(n|s)P(H=1|s)$$

$g(F,a,n) = \sum_{s} P(s|F,a)P(n|s)P(H=1|s)$

3 – scope of largest factor

(Potential for) exponential reduction in computation!

Variable Elimination – Order can make a HUGE difference



$$P(X_1) = \sum_{Y,X_2,\ldots,X_n} P(Y)P(X_1|Y)\prod_{i=2}^{n} P(X_i|Y)$$

$$= \sum_{Y,X_3,\ldots,X_n} P(Y)P(X_1|Y)\prod_{i=3}^{n} P(X_i|Y)\underbrace{\sum_{X_2} P(X_2|Y)}_{g(Y)}$$

1 – scope of largest factor

$$= \sum_{X_2,\ldots,X_n} \underbrace{\sum_{Y} P(Y)P(X_1|Y)\prod_{i=2}^{n} P(X_i|Y)}_{g(X_1,X_2,..,X_n)}$$

n – scope of largest factor

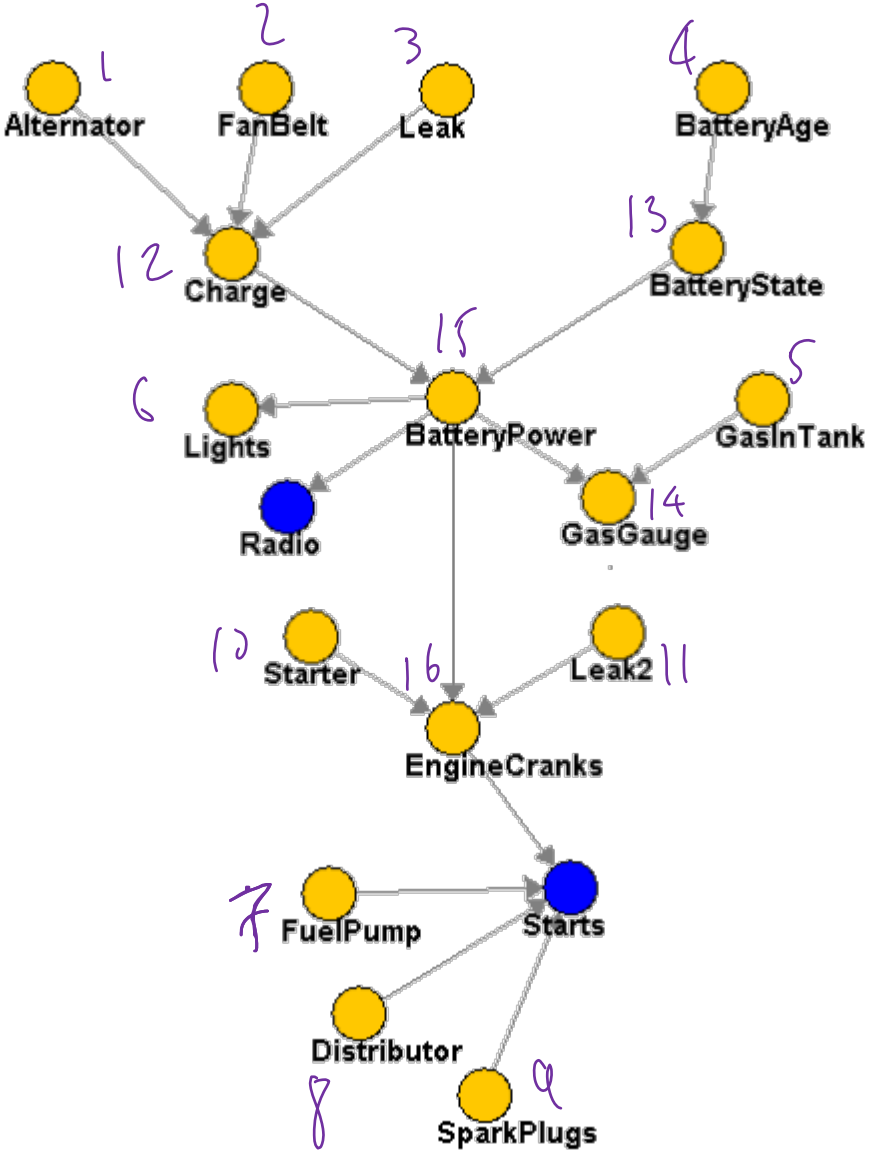
Variable Elimination Algorithm

- Given BN – DAG and CPTs (initial factors – $p(x_i|pa_i)$ for i=1,..,n)
- Given evidence e, set of variables X, suppose we want to compute P(X|e). We first compute P(X,e)
- Instantiate evidence e e.g. set H=1
- Choose an ordering on the variables e.g., X_1, …, X_n
- For i = 1 to n, If $X_i \notin \{X,e\}$
  - Collect factors $g_1,\ldots,g_k$ that include X_i
  - Generate a new factor by eliminating X_i from these factors

$$g = \sum_{X_i} \prod_{j=1}^{k} g_j$$

  - Variable X_i has been eliminated!
  - Remove $g_1,\ldots,g_k$ from set of factors but add g
- Eliminate X to compute P(e); combine to obtain P(X|e)

Complexity for (Poly)tree graphs



Variable elimination order:

- Consider undirected version (ignore edge directions)
- Start from "leaves" up
- find topological order
- eliminate variables in that order

Does not create any factors bigger than original CPTs

For polytrees, inference is linear in # variables (vs. exponential in general)!

Complexity for graphs with loops

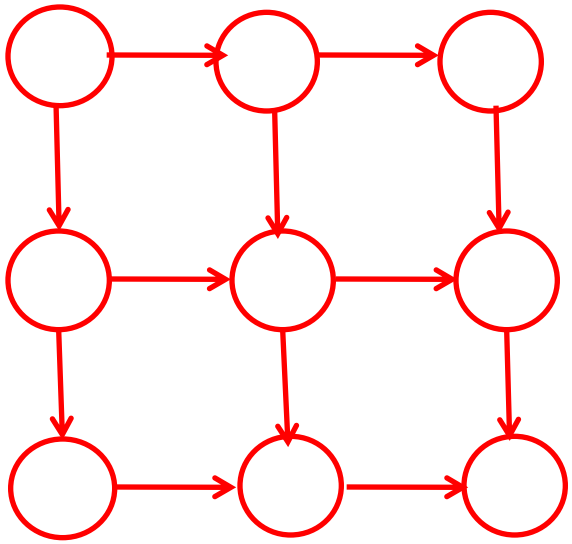
- Loop – undirected cycle

  Linear in # variables but exponential in size of largest factor generated!

  Minimize (over all possible orderings) of size of largest factor generated, minus one, is called the "tree-width of a graph"

  - Equal to one for tree (since ordering from leaves to root introduces factor of size at most two)

Example: Large tree-width with small number of parents



At most 2 parents per node, but tree width is $O(\sqrt{n})$

Compact representation $\not\Rightarrow$ Easy inference ☹

Choosing an elimination order

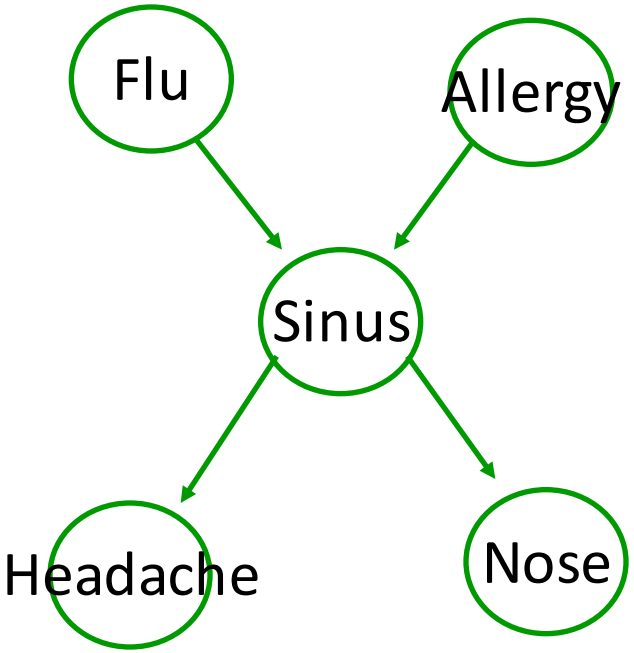
- Choosing best order is NP-complete
  - Reduction from MAX-Clique
- Many good heuristics (some with guarantees)
- Ultimately, can't beat NP-hardness of inference
  - Even optimal order can lead to exponential variable elimination computation
- In practice
  - Variable elimination often very effective
  - Many (many many) approximate inference approaches available when variable elimination too expensive

Inference

- Possible queries:

2) Most likely assignment of nodes

$$\arg\max_{f,a,s,n} P(F=f,A=a,S=s,N=n \mid H=1)$$



Use Distributive property:

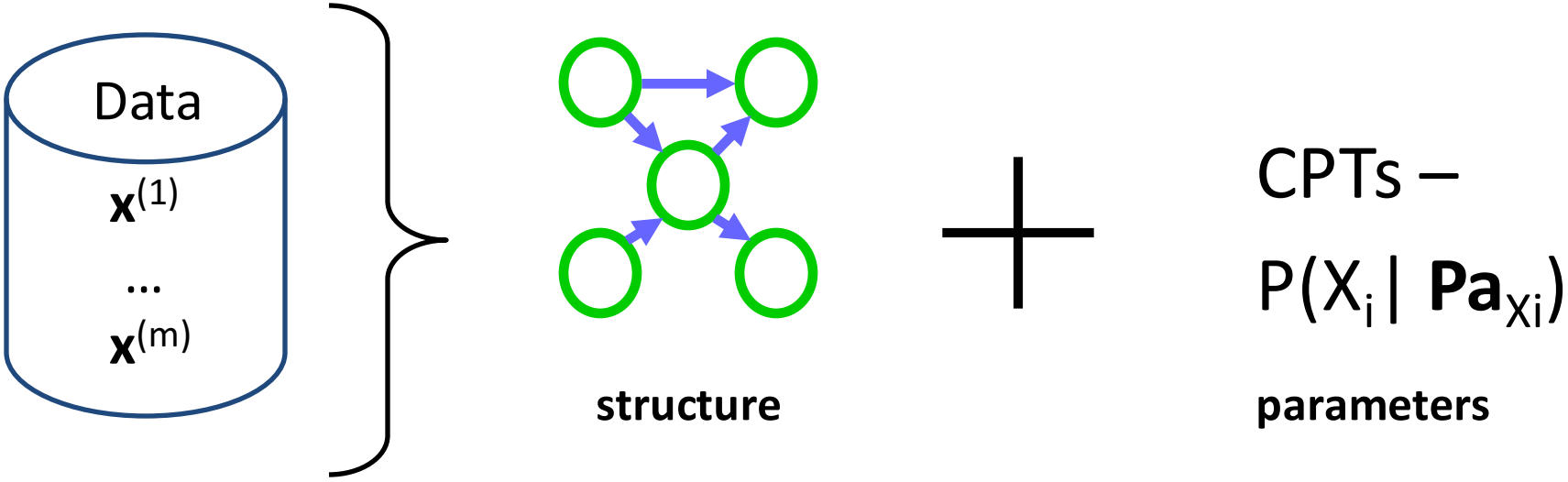
$$\max(x_1 z,\ x_2 z) = z \max(x_1, x_2)$$

2 multiply 1 mulitply

Topics in Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

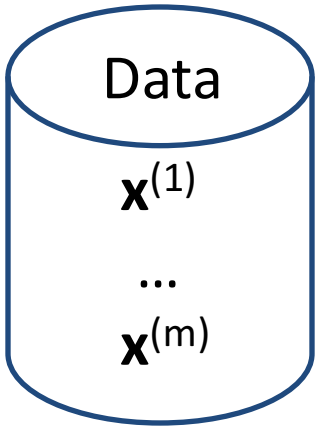
Learning

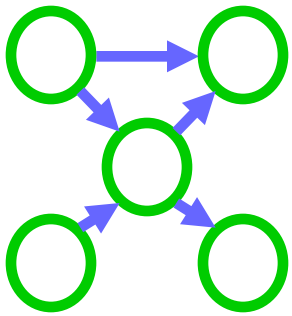


Given set of m independent samples (assignments of random variables),

find the best (most likely?) Bayes Net (Graph Structure + CPTs)

Learning the CPTs (given structure)





For each discrete variable X_k

Compute MLE or MAP estimates for

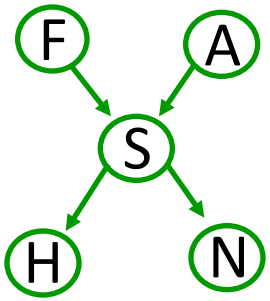
$$p(x_k|\mathrm{pa}_k)$$

Recall

MLE: $$P(X_i = x_i \mid X_j = x_j) = \frac{\mathrm{Count}(X_i = x_i, X_j = x_j)}{\mathrm{Count}(X_j = x_j)}$$

MAP: Add psuedocounts

MLEs decouple for each CPT in Bayes Nets



- Given structure, log likelihood of data

$$\log P(\mathcal{D} \mid \theta_{\mathcal{G}}, \mathcal{G})$$

$$= \log \prod_{j=1}^{m} P(f^{(j)}) P(a^{(j)}) P(s^{(j)} \mid f^{(j)}, a^{(j)}) P(h^{(j)} \mid s^{(j)}) P(n^{(j)} \mid s^{(j)})$$

$$= \sum_{j=1}^{m} [\log P(f^{(j)}) + \log P(a^{(j)}) + \log P(s^{(j)} \mid f^{(j)}, a^{(j)}) + \log P(h^{(j)} \mid s^{(j)}) + \log P(n^{(j)} \mid s^{(j)})]$$

$$= \underbrace{\sum_{j=1}^{m} \log P(f^{(j)})}_{\theta_F} + \underbrace{\sum_{j=1}^{m} \log P(a^{(j)})}_{\theta_A} + \underbrace{\sum_{j=1}^{m} \log P(s^{(j)} \mid f^{(j)}, a^{(j)}) + \sum_{j=1}^{m} \log P(h^{(j)} \mid s^{(j)})}_{\theta_{F,A}} + \underbrace{\sum_{j=1}^{m} \log P(n^{(j)} \mid s^{(j)})}_{\theta_{N|S}}]$$

Depends only on $\theta_{H|S}$

Can compute MLEs of each parameter independently!