CS 4824/ECE 4424: Graphical Models I

Acknowledgement:
Many of these slides are derived from Tom Mitchell, Pascal Poupart, Pieter Abbeel, Eric Eaton, Carlos Guestrin, William Cohen, and Andrew Moore.

Graphical Models

- **Key Idea**:
  - Express sets of conditional independence assumptions via graph structure
  - Graph structure plus associated parameters define joint probability distribution over set of variables/ nodes

Conditional independence

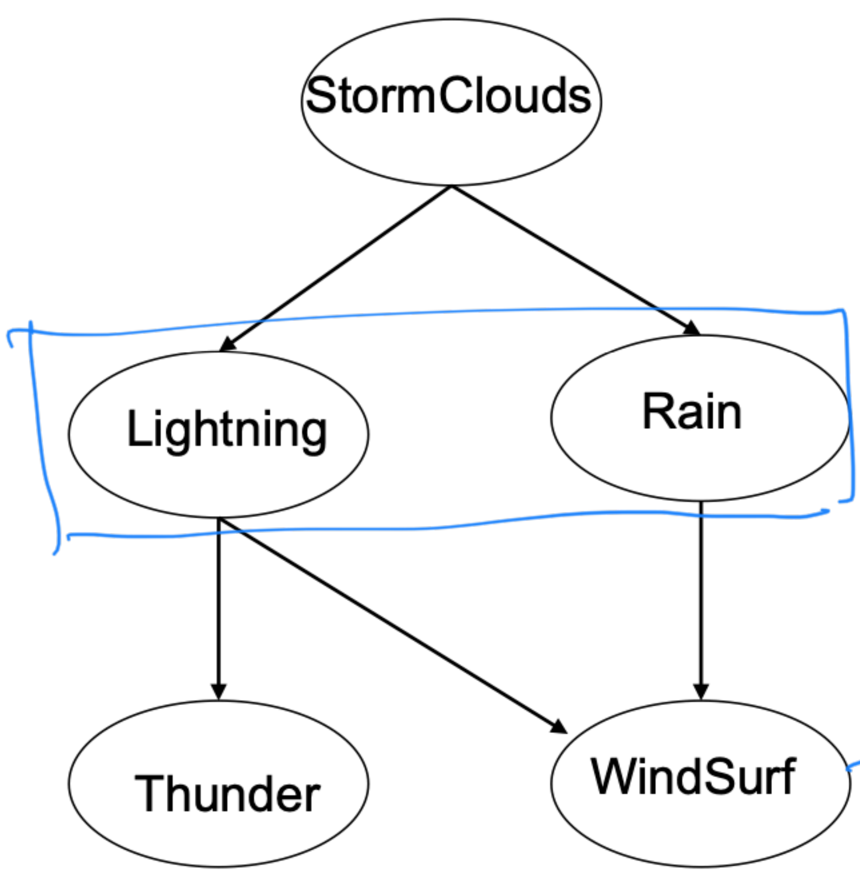
- X is conditionally independent of Y given Z, if the probability distribution governing X is independent of the value of Y given the value of Z
  - $(\forall i, j, k)\ P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z = z_k)$
  - Or equivalently, $\boxed{P(X|Y Z) = P(X|Z)}$
- P(Thunder|Rain, Lightning) = P(Thunder|Lightning)
  - That is, Thunder and Rain are conditionally independent given Lightning.

Marginal independence

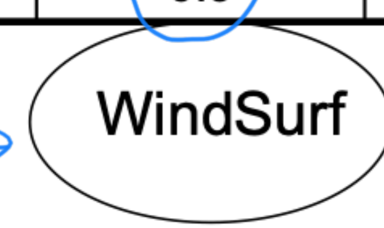
- X is marginally independent of Y if
  - $(\forall i, j)\ \boxed{P(X = x_i, Y = y_j)} = \boxed{P(X = x_i)P(Y = y_j)}$
- Equivalently if
  - $(\forall i, j)\ P(X = x_i \mid Y = y_j) = P(X = x_i)$
- Equivalently if
  - $(\forall i, j)\ P(Y = y_i \mid X = x_j) = P(Y = y_i)$

Bayesian network

- A directed acyclic graph defining a joint probability distribution over a set of variables
- Each node denotes a random variable
- A conditional probability distribution (CPD) is associated with each node N, defining P(N | Parents(N))



Parents	P(W|Pa)	P(¬W|Pa)
L, R	0	1.0
L, ¬R	0	1.0
¬L, R	0.2	0.8
¬L, ¬R	0.9	0.1



The joint distribution over all variables in the network is defined in terms of these CPD's, plus the graph

Bayesian network more formally

- A Bayes network represents the joint probability distribution over a collection of random variables

- A Bayes network is a directed acyclic graph and a set of CPD's
  - Each node denotes a random variable
  - Edges denote dependencies
  - CPD for each node $\boxed{X_i}$ defines $\boxed{P(X_i \mid Pa(X_i))}$
  - The joint distribution over all variables is defined as

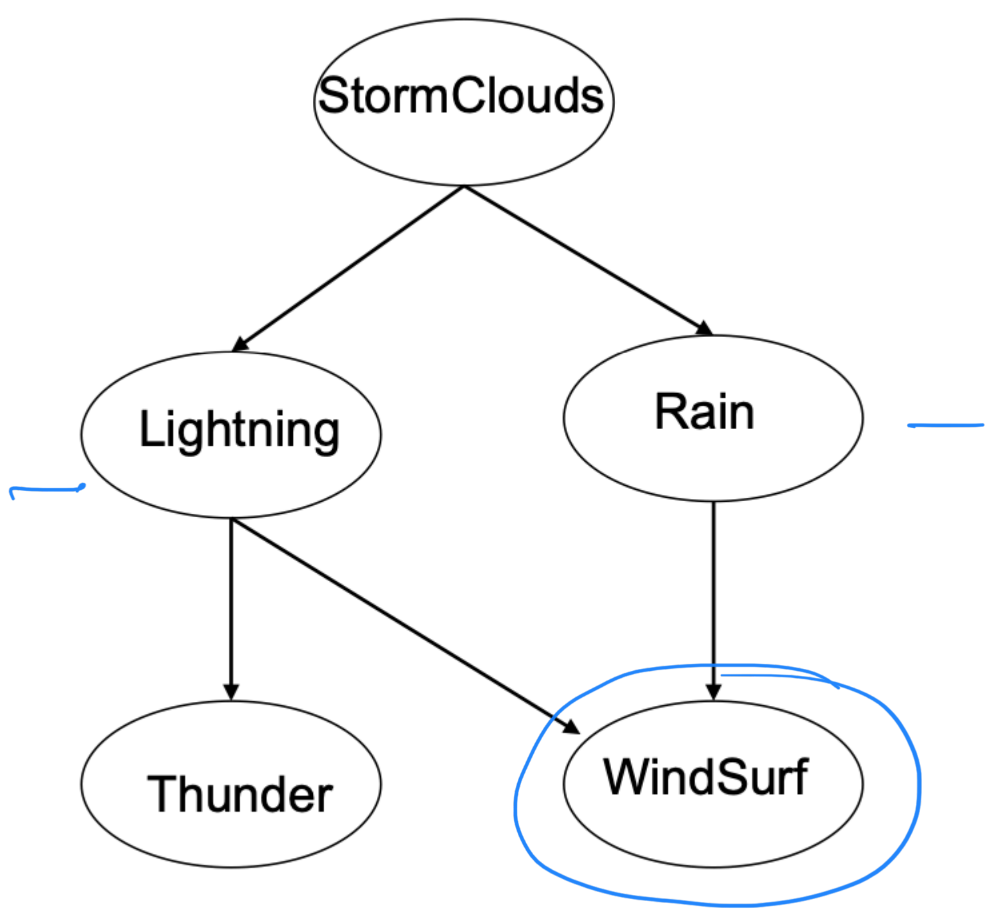
$$\boxed{P(X_1 \dots X_n)} = \prod_i P(X_i \mid Pa(X_i))$$

$Pa(X_i)$
$= \text{Parent}(X_i)$

immidiate parents of X_i

Bayesian network: conditional independence

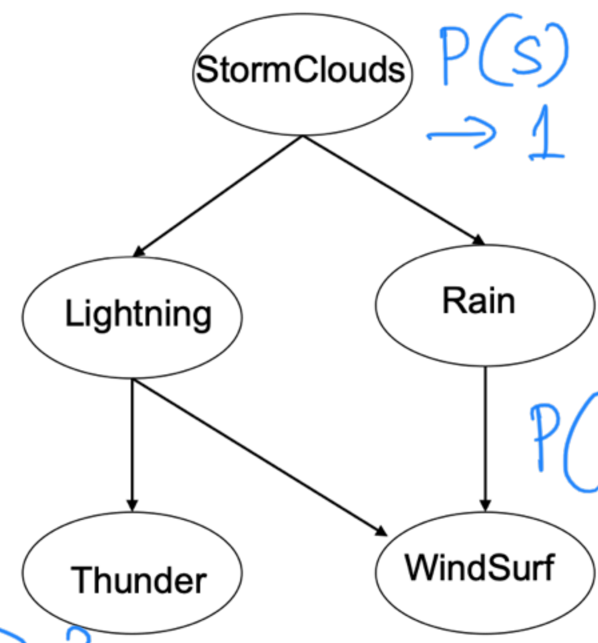
- What can we say about conditional independence in a Bayes Net?
- Each node is conditionally independent of its non-descendents, given only its immediate parents



Parents	P(W|Pa)	P(¬W|Pa)
L, R	0	1.0
L, ¬R	0	1.0
¬L, R	0.2	0.8
¬L, ¬R	0.9	0.1

Bayesian network

- CPD for each node X_i describes $P(X_i \mid Pa(X_i))$



$P(S) \rightarrow 1$

$P(L|S) \rightarrow 2$

$P(R|S) \rightarrow 2$

$P(T|L) \rightarrow 2$

Parents	P(W|Pa)	P(¬W|Pa)
L, R	0	1.0
L, ¬R	0	1.0
¬L, R	0.2	0.8
¬L, ¬R	0.9	0.1

$P(W|L,R) \rightarrow 4$

- Chain rule of probability

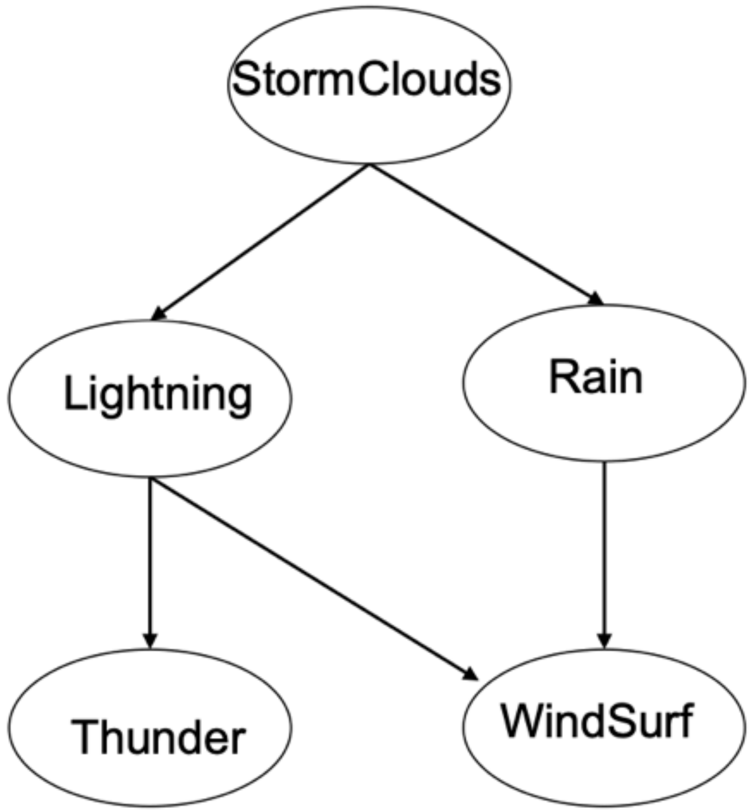
$$\underbrace{P(S, L, R, T, W)}_{\text{Full joint distribution}} = \underbrace{P(S)}\underbrace{P(L \mid S)}\underbrace{P(R \mid S, L)}\underbrace{P(T \mid S, L, R)}\underbrace{P(W \mid S, L, R, T)}$$

Full joint distribution how many parameters? $2^5 - 1 = 31$

- But in Bayes net: $\underbrace{P(X_1 \ldots X_n)} = \prod_i P(X_i \mid Pa(X_i))$ ~ 11 parameters

$P(S,L,R,T,W) = P(S) \cdot P(L|S)\, P(R|S)\, P(T|L)\, P(W|L,R)$

How many parameters?



Parents	P(W|Pa)	P(¬W|Pa)
L, R	0	1.0
L, ¬R	0	1.0
¬L, R	0.2	0.8
¬L, ¬R	0.9	0.1

- In full joint distribution?

- Given this Bayes net?

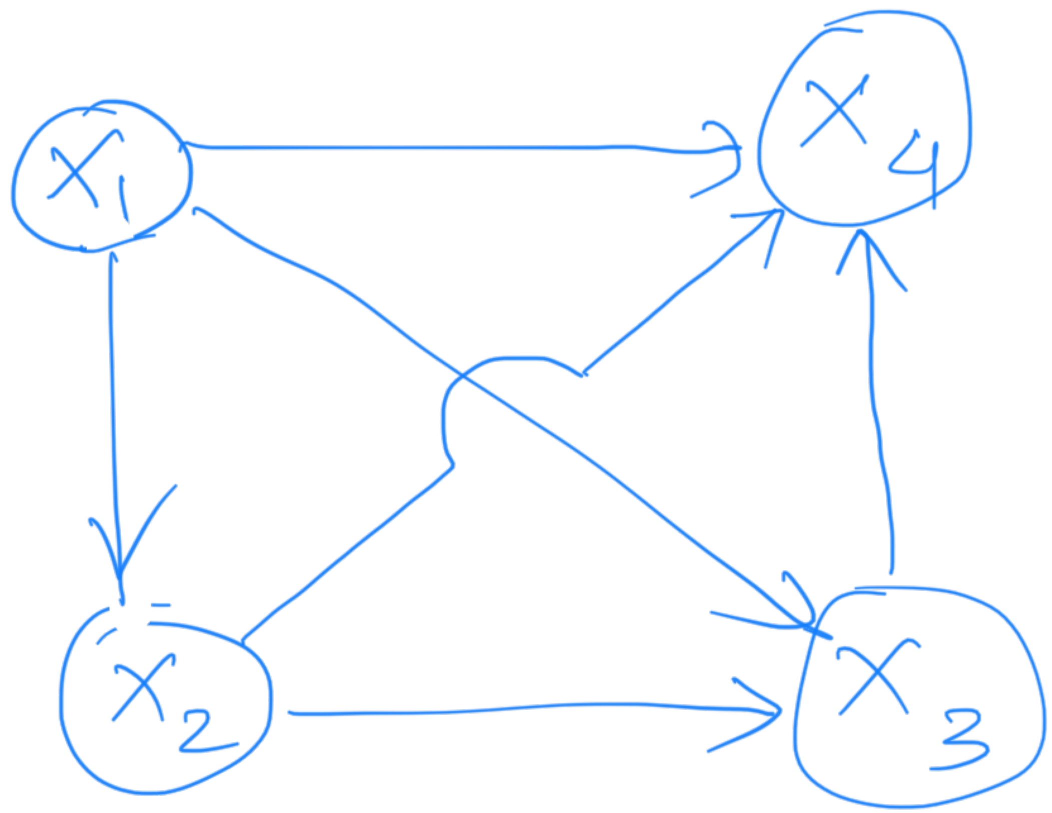
Algorithm for constructing Bayes network

- Choose an ordering over variables, e.g. $\boxed{X_1, X_2, \ldots, X_n}$
- For i=1 to n
  - Add X_i to the network
  - Select parents $Pa(X_i)$ as minimal subset of $X_1 \ldots X_{i-1}$ such that $\boxed{P(X_i \mid Pa(X_i)) = P(X_i \mid X_1 \ldots X_{i-1})}$

- Notice this choice of parents assures
  $$P(X_1 \ldots X_n) = \prod_i P(X_i \mid X_1 \ldots X_{i-1}) = \boxed{\prod_i P(X_i \mid Pa(X_i))}$$

What is the Bayes net for $X_1, X_2, \ldots, X_n$ with NO assumed conditional independence ?

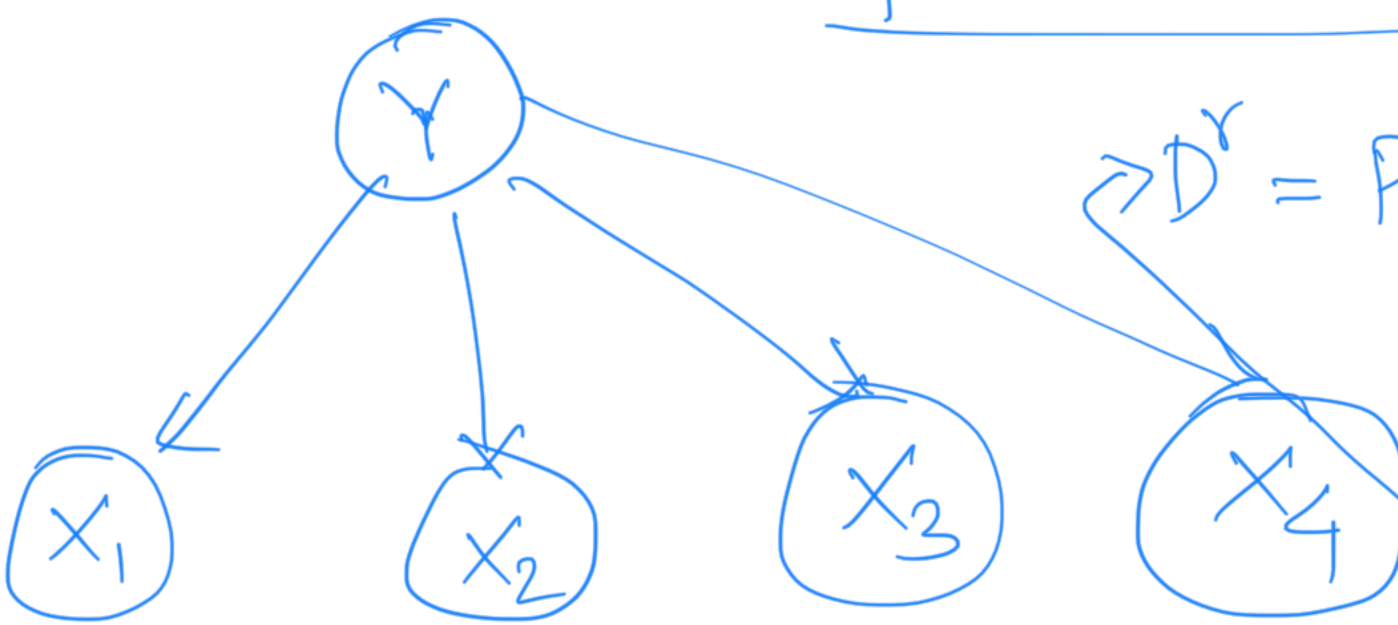
$$P(X_1 \ X_2 \ X_3 \ X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1 X_2)$$
$$P(X_4|X_1 X_2 X_3)$$



What is the Bayes net for Naïve Bayes?

$$P(Y \mid x_1, x_2, x_3, x_4)$$

$$\boxed{P(Y, x_1, x_2, x_3, x_4) = P(Y)\, P(x_1|Y)\, P(x_2|Y)\, P(x_3|Y)\, P(x_4|Y)}$$



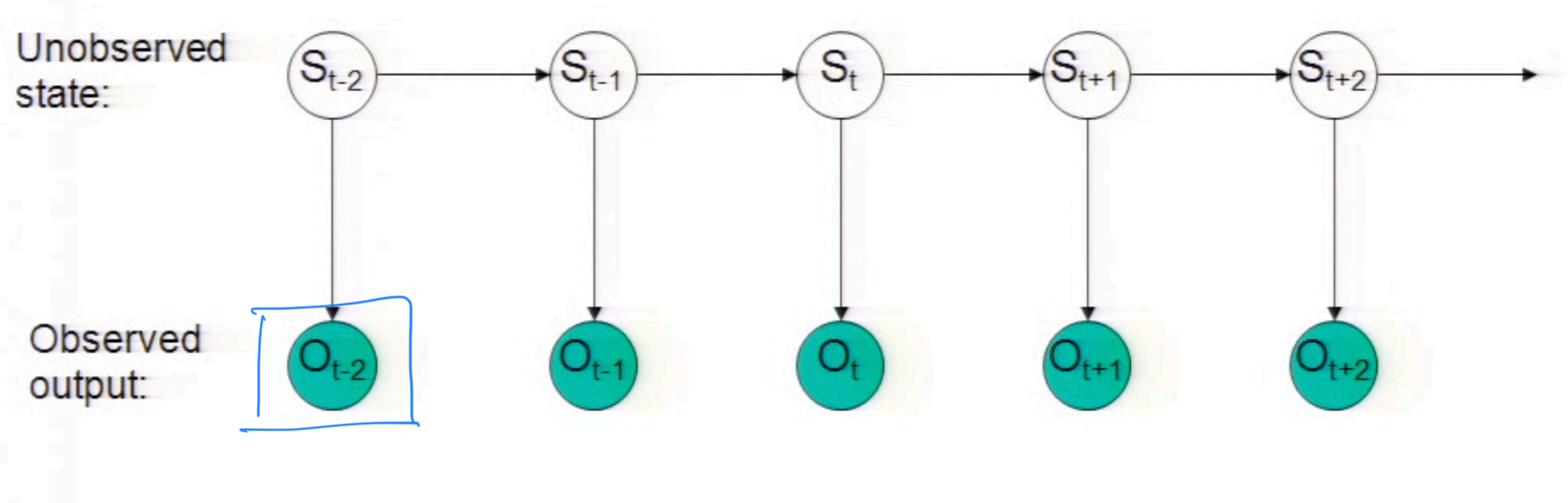
$$D^r = P(Y=1, X_1=a, X_2=b, X_3=c, x_4=d) + P(Y=0, X_1=a, X_2=b, X_3=c, x_4=d)$$

$$P(Y=1 \mid X_1=a, X_2=b, X_3=c, X_4=d)$$

$$= \frac{P(Y=1, X_1=a, X_2=b, X_3=c, X_4=d)}{D^r}$$

Hidden Markov Model

- Assume the future is conditionally independent of the past, given the present



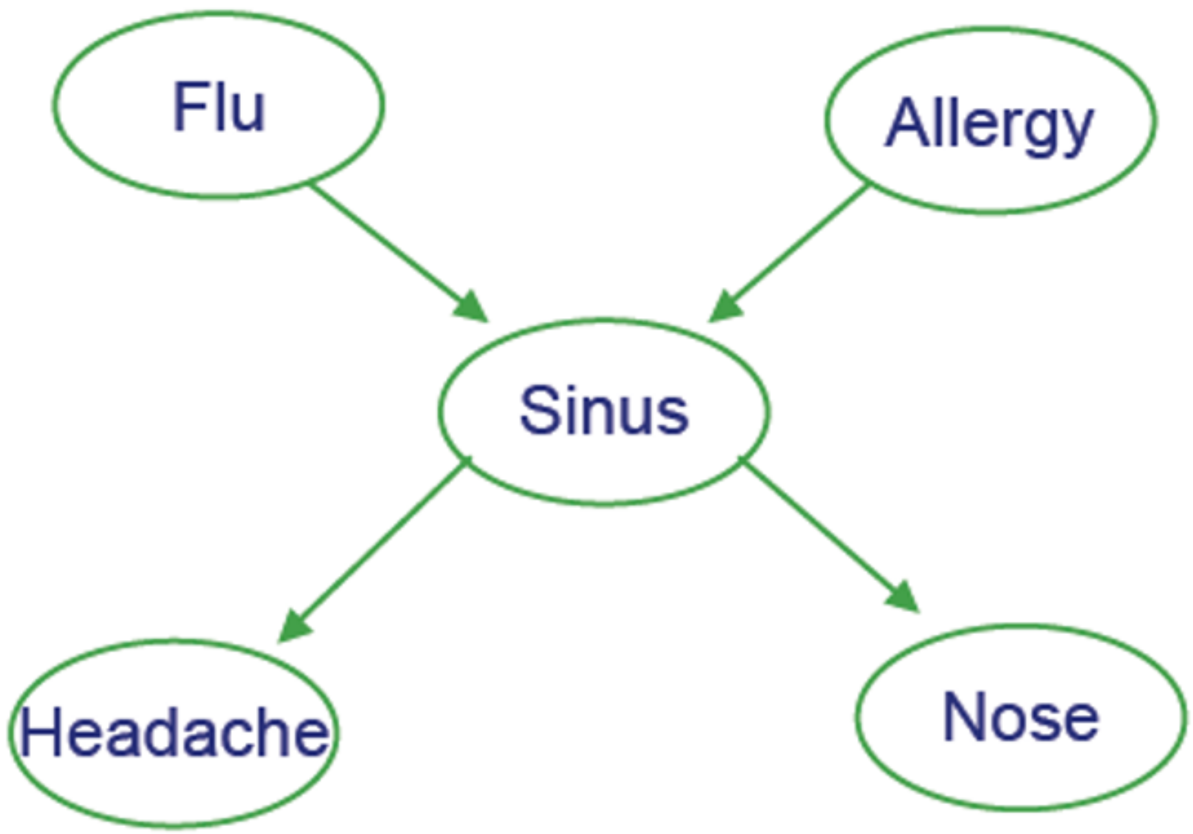
$$P(S_{t-2}, O_{t-2}, S_{t-1}, \ldots, O_{t+2}) = P(S_{t-2})\, P(O_{t-2} \mid S_{t-2})\, P(S_{t-1} \mid S_{t-2})\, P(O_{t-1} \mid S_{t-1})\, P(S_t \mid S_{t-1})$$

Inference in Bayes Nets

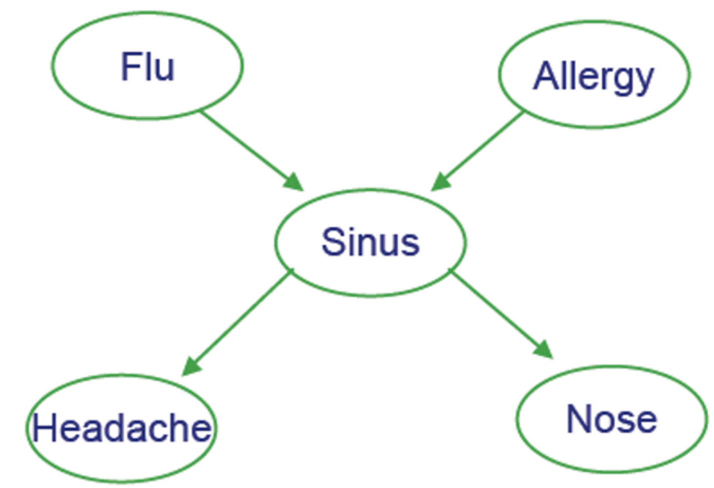
- In general, intractable (NP-complete)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
  - Or for singly connected graphs (ie., no undirected loops)
    - Belief propagation
- Sometimes use Monte Carlo methods
  - Generate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

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 $\langle F=f, A=a, S=s, H=h, N=n \rangle$

- What is $\boxed{P(f,a,s,h,n)}$?

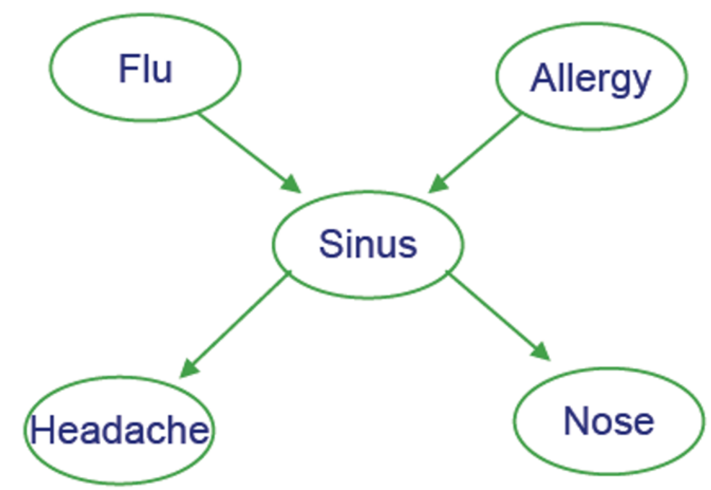
$$= P(f)\, P(a)\, P(s|a,f)\, P(h|s)\, P(n|s)$$

K boolean RV's will have K computations

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)?



$$P(N=n) = \sum_{s} P(N=n \mid S=s)\, P(S=s)$$

chase up the Bayes Net

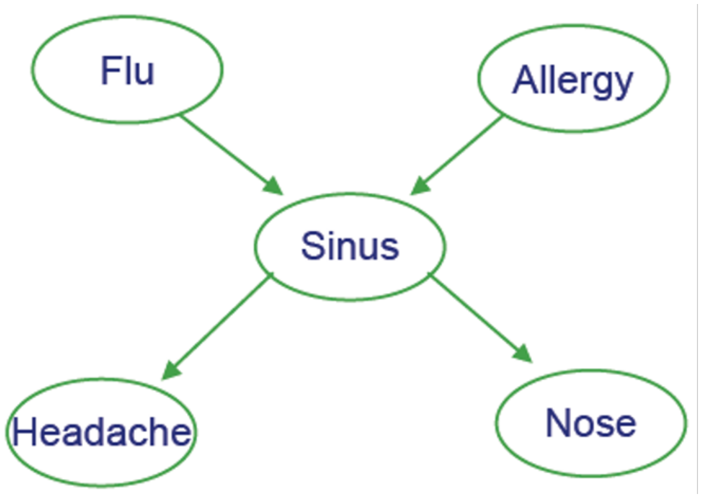
$$P(N=n) = \sum_{f,a,h,s} \underbrace{P(F=f, A=a, H=h, S=s, N=n)}_{P(f)\,P(a)\,P(s|fa)\,P(h|s)\,P(N=n|s)}$$

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

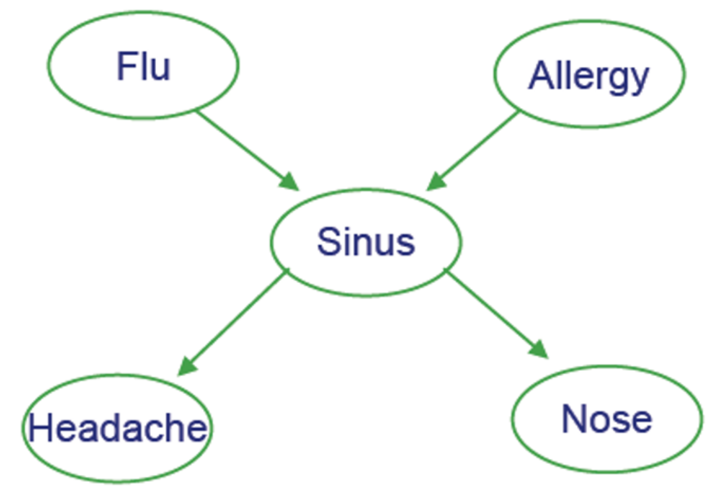
K boolean random variables 2^{K-1} computations

Generating a sample from joint distribution: easy

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



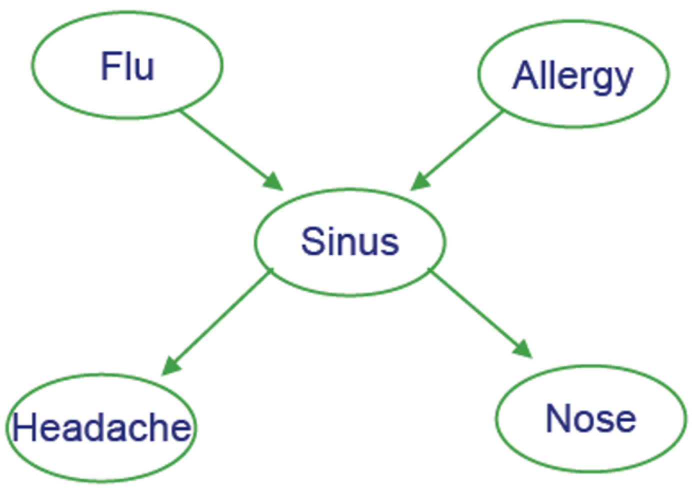
Generating a sample from joint distribution: easy



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

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