

# Reasoning Under Uncertainty: More on BNets structure and construction

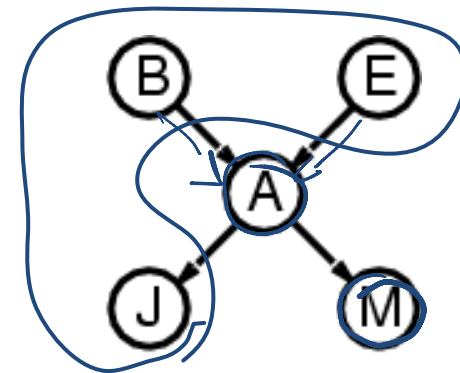
Computer Science cpsc322, Lecture 28

*(Textbook Chpt 6.3)*

~

March, 24, 2010

# Belief networks Recap




- By considering causal dependencies, we order variables in the joint.
- Apply.....*chain rule*..... and simplify

$$P(B, E, A, J, M) = P(B) P(E) \underline{P(A|B, E)} P(J|A) \boxed{P(M|A)}$$

*why  $M \perp (B, E, J) \text{ given } A$   $P(M, B, E, J, A)$*


- Build a directed acyclic graph (DAG) in which the parents of each var.  $X$  are those vars on which  $X$  directly depends.
- By construction, a var is independent from it non-descendant given its parents. *why?*

# Belief Networks: open issues

- **Independencies:** Does a BNet encode more independencies than the ones specified by construction? *yes* 
- **Compactness:** We reduce the number of probabilities from  $2^n$  to  $O(n 2^K)$  *n vars*  $K$  parents

In some domains we need to do better than that!

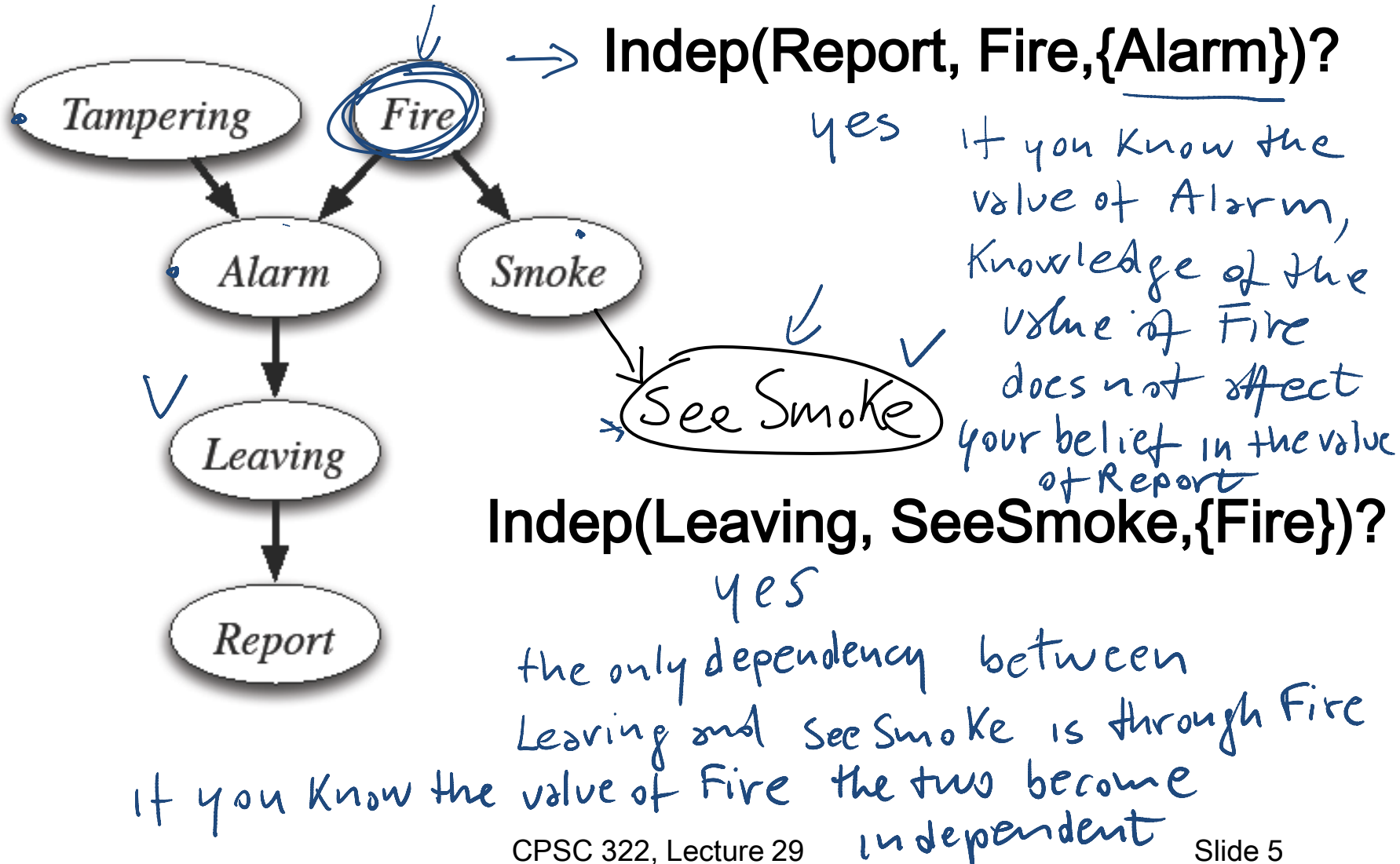
- Still too many and often there are no data/experts for accurate assessment 

**Solution:** Make stronger (approximate) *independence assumptions* 

# Lecture Overview

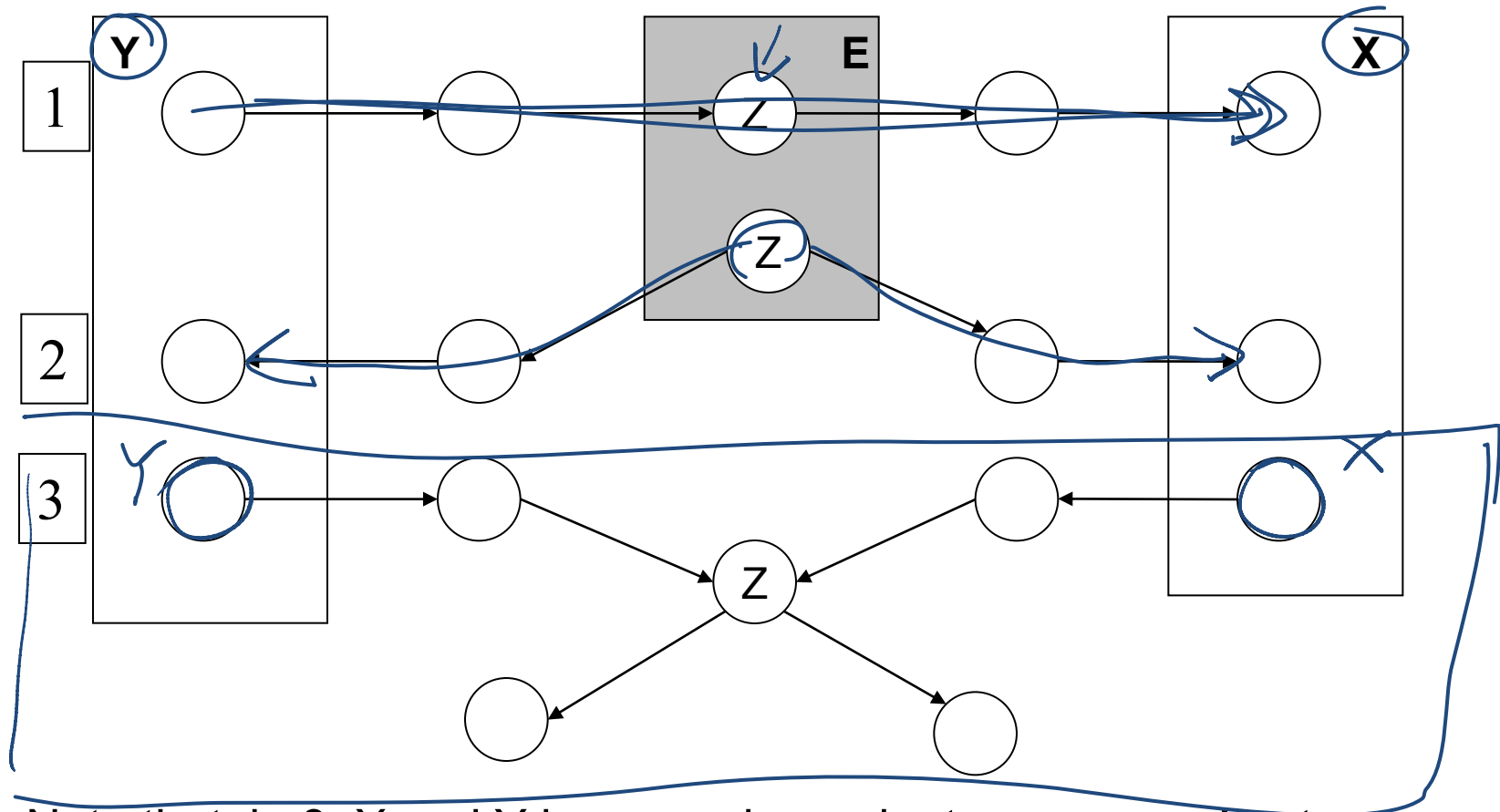
- Implied Conditional Independence relations in a Bnet
- Compactness: Making stronger Independence assumptions
  - Representation of Compact Conditional Distributions
  - Network structure( Naïve Bayesian Classifier)

# Bnets: Entailed (in)dependencies



# Conditional Independencies

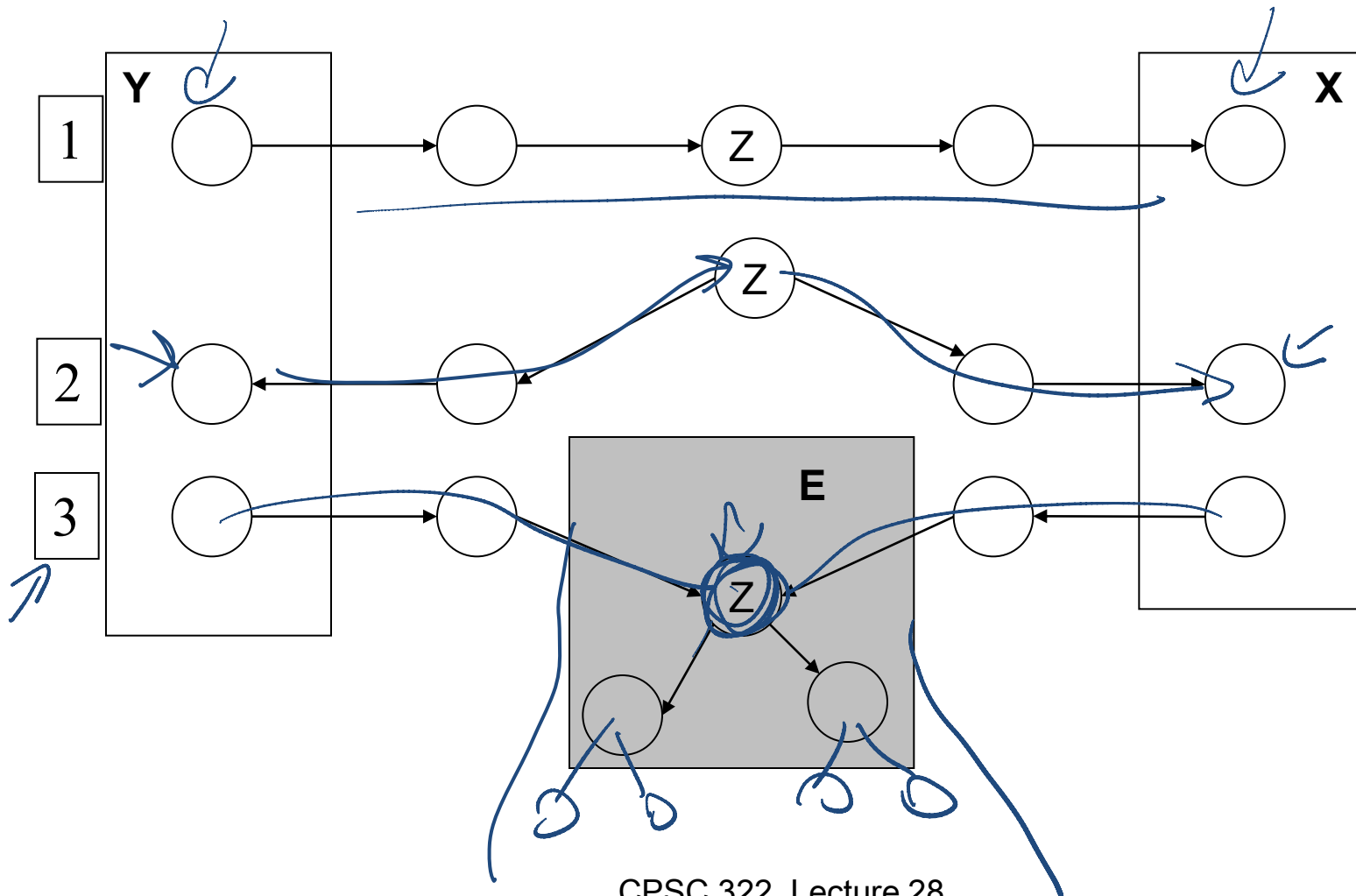
Or, blocking paths for probability propagation. Three ways in which a path between X to Y can be blocked, (1 and 2 given evidence E )



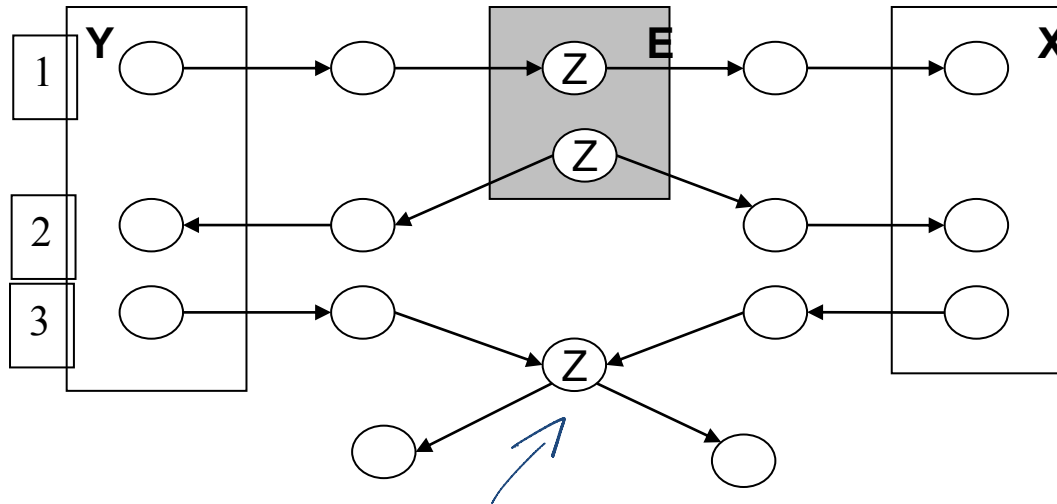
Note that, in 3, X and Y become dependent as soon as I get evidence on Z or on *any of its descendants*

# Or ....Conditional Dependencies

In 1,2,3 X Y are dependent

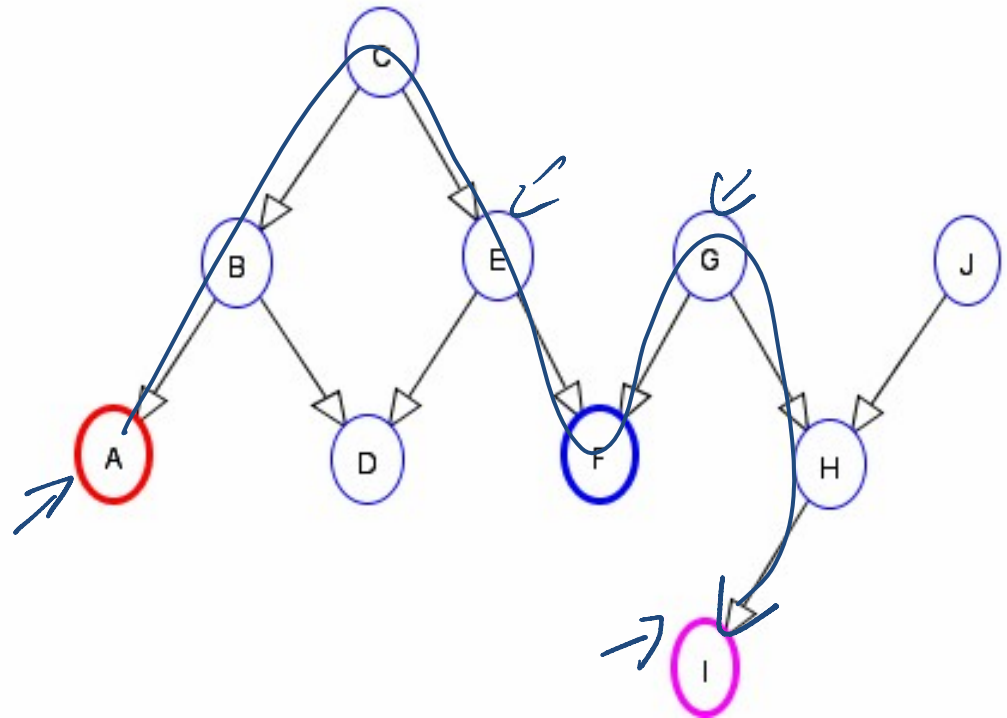


# In/Dependencies in a Bnet : Example 1



Is A conditionally independent of I given F?


false





The logo for AIspace, featuring the letters 'AI' in a blue circle with a magnifying glass over it, followed by the word 'space' in a yellow, stylized font.

# Lecture Overview

- Implied Conditional Independence relations in a Bnet
- Compactness: Making stronger Independence assumptions
  - Representation of Compact Conditional Distributions 
  - Network structure( Naïve Bayesian Classifier)

# More on Construction and Compactness: Compact Conditional Distributions

Once we have established the topology of a Bnet, we still need to specify the conditional probabilities

How?

- From Data
- From Experts

To facilitate acquisition, we aim for compact representations for which data/experts can provide accurate assessments

# More on Construction and Compactness: Compact Conditional Distributions

From JointPD  $2^n$  to  $n 2^K$

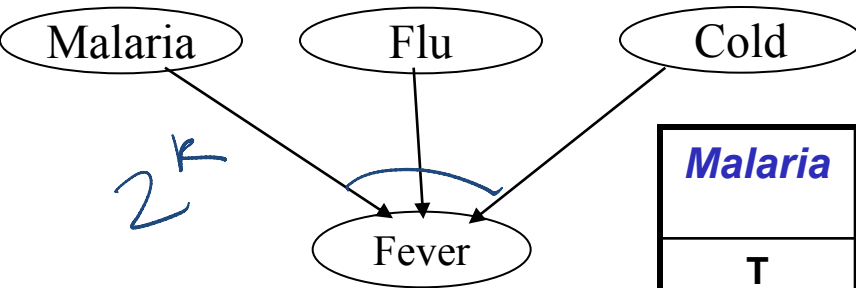
But still, CPT grows exponentially with number of parents

In **realistic model of internal medicine** with 448 nodes and 906 links 133,931,430 values are required!

And often there are no data/experts for accurate assessment

# Effect with multiple non-interacting causes

What do we need to specify?



<i>Malaria</i>	<i>Flu</i>	<i>Cold</i>	$P(\text{Fever}=T \mid ..)$	$P(\text{Fever}=F \mid ..)$
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F		

What do you think data/experts could easily tell you?  
*when only one cause is present*

More difficult to get info to assess more complex conditioning....

# Solution: Noisy-OR Distributions

- Models multiple non interacting causes
- Logic OR with a probabilistic twist.
- Logic OR Conditional Prob. Table.

<i>Malaria</i>	<i>Flu</i>	<i>Cold</i>	$P(\text{Fever}=T \mid ..)$	$P(\text{Fever}=F \mid ..)$
T	T	T	1	0
T	T	F	1	0
T	F	T	1	0
T	F	F	1	0
F	T	T	1	0
F	T	F	1	0
F	F	T	1	0
F	F	F	0	1

# Solution: Noisy-OR Distributions

The Noisy-OR model allows for uncertainty in the ability of each cause to generate the effect (e.g.. one may have a cold without a fever)

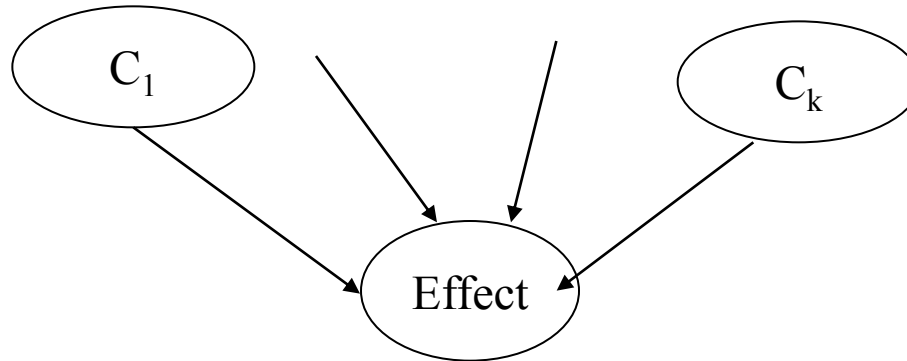
<i>Malaria</i>	<i>Flu</i>	<i>Cold</i>	$P(\text{Fever}=T \mid ..)$	$P(\text{Fever}=F \mid ..)$
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F	0	1

non zeros

Two assumptions

1. All possible causes are listed
2. For each of the causes, whatever inhibits it to generate the target effect is independent from the inhibitors of the other causes

# Noisy-OR: Derivations



For each of the causes, whatever inhibits it to generate the target effect is independent from the inhibitors of the other causes

- Independent Probability of failure  $q_i$  for each cause alone:

- $P(\text{Effect}=\text{F} \mid C_i = \text{T}, \text{ and no other causes}) = q_i$
- $P(\text{Effect}=\text{F} \mid C_1 = \text{T}, \dots, C_j = \text{T}, C_{j+1} = \text{F}, \dots, C_k = \text{F}) = \prod_{i=2}^k q_i$
- $P(\text{Effect}=\text{T} \mid C_1 = \text{T}, \dots, C_j = \text{T}, C_{j+1} = \text{F}, \dots, C_k = \text{F}) = 1 - \prod_{i=2}^k q_i$



9i

# Noisy-OR: Example

$$P(\text{Fever}=F | \text{Cold}=T, \text{Flu}=F, \text{Malaria}=F) = 0.6$$

$$P(\text{Fever}=F | \text{Cold}=F, \text{Flu}=T, \text{Malaria}=F) = 0.2$$

$$P(\text{Fever}=F | \text{Cold}=F, \text{Flu}=F, \text{Malaria}=T) = 0.1$$

Model of internal medicine

133,931,430 → 8,254

using Noisy-ORs

$$P(\text{Effect}=F | C_1 = T, \dots, C_j = T, C_{j+1} = F, \dots, C_k = F) = \prod_{i=1}^j q_i$$

Malaria	Flu	Cold	$P(\text{Fever}=T   ..)$	$P(\text{Fever}=F   ..)$
→ T	T	T	.988	$0.1 \times 0.2 \times 0.6 = 0.012$
→ T	T	F	→ .98	$0.2 \times 0.1 = 0.02$
T	F	T	.94	$0.6 \times 0.1 = 0.06$
→ T	F	F	0.9	0.1 ←
F	T	T	.88	$0.2 \times 0.6 = 0.12$
→ F	T	F	0.8	0.2 ←
→ F	F	T	0.4	0.6 ←
F	F	F	0	1.0

required

3 in this example

- Number of probabilities linear in .....

# Lecture Overview

- Implied Conditional Independence relations in a Bnet
- **Compactness: Making stronger Independence assumptions**
  - Representation of Compact Conditional Distributions
  - **Network structure ( Naïve Bayesian Classifier)**

# Naïve Bayesian Classifier

A very simple and successful Nets that allow to classify entities in a set of classes  $C$ , given a set of attributes

## Example:

- Determine whether an **email** is spam (only two classes spam=T and spam=F)
- Useful attributes of an email ? *words contained in the email*

## Assumptions

- The value of each attribute depends on the classification
- (Naïve) The attributes are independent of each other given the classification

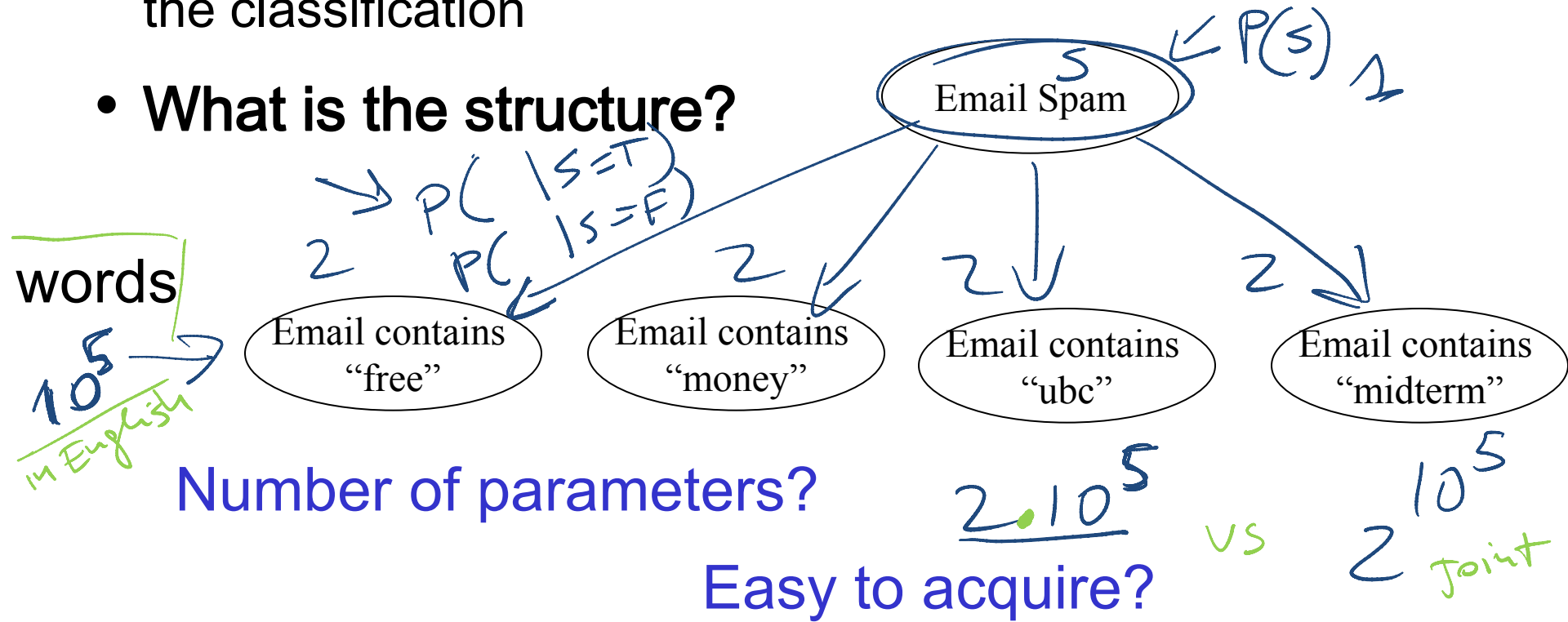
$$P(\text{"bank"} \mid \text{"account"}, \text{spam=T}) \neq P(\text{"bank"} \mid \text{spam=T})$$

# Naïve Bayesian Classifier for Email Spam

## Assumptions

- The value of each attribute depends on the classification
- **(Naïve)** The attributes are independent of each other given the classification

- **What is the structure?**

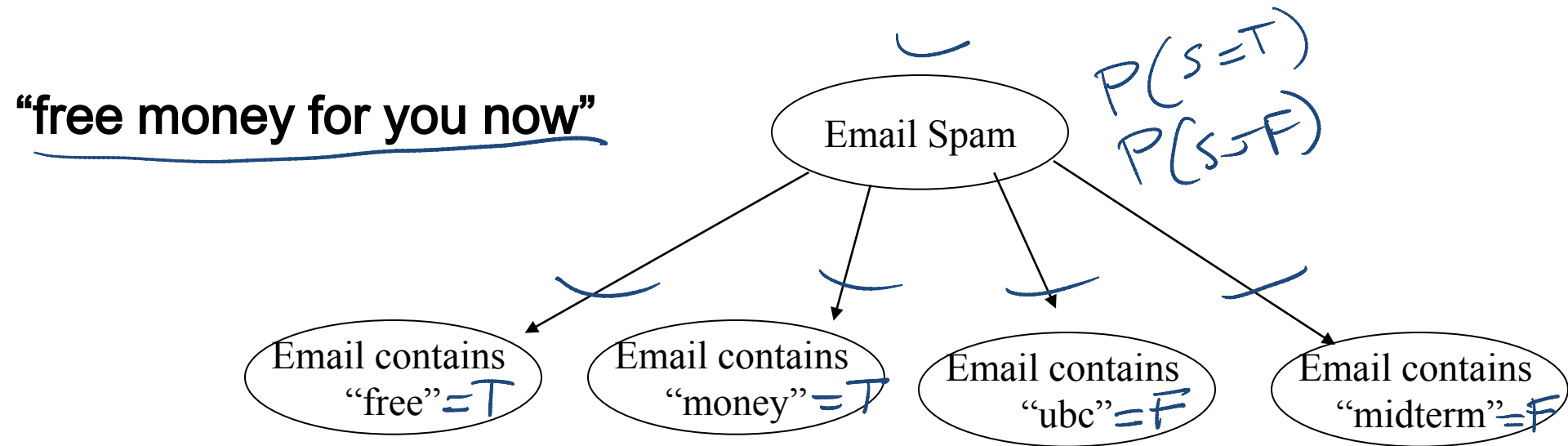


If you have a large collection of emails for which you know if they are spam or not.....

# NB Classifier for Email Spam: Usage

Most likely class given set of observations

Is a given Email  $E$  spam?  $\leftarrow$



Email is a spam if.....

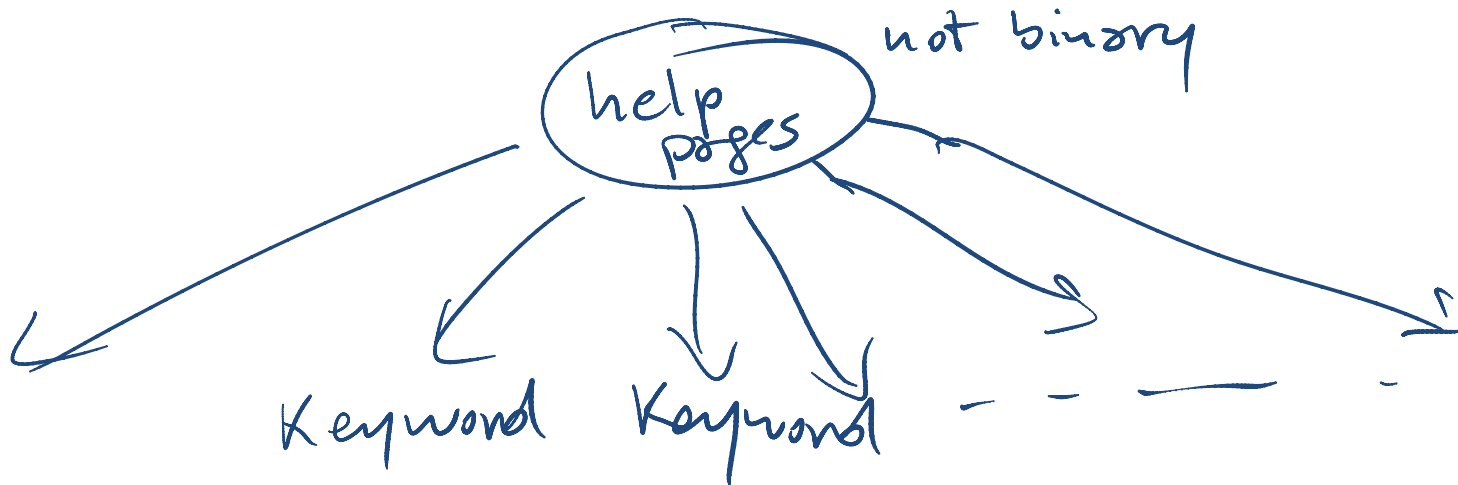
$$P(s=T) > P(s=F)$$

after the two probs are updated in light of the evidence (words in email are set to T)

# For another example of naïve Bayesian Classifier

See textbook ex. 6.16 

help system to determine what help page a user is interested in based on the keywords they give in a query to a help system.



# Learning Goals for today's class

You can:

- Given a Belief Net, determine whether one variable is conditionally independent of another variable, given a set of observations.
- Define and use **Noisy-OR** distributions. Explain assumptions and benefit.
- Implement and use a **naïve Bayesian classifier**. Explain assumptions and benefit.

# Next Class

Bayesian Networks Inference: Variable Elimination

## Course Elements

- Practice Exercise on will be Bnet posted tomorrow.
- Assignment 3 is due on Monday!
- Assignment 4 will be available on Wednesday and due on Apr the 14<sup>th</sup> (last class).