Finish Probability Theory + Bayesian Networks

Computer Science cpsc322, Lecture 9

(Textbook Chpt 6.1-2-3)

June, 5, 2012

Lecture Overview

- **Finish Intro to Probability**
  – Chain Rule and Bayes' Rule
  – Marginal and Conditional Independence
- **Bayesian Networks**
  – Build sample BN
  – Intro Inference, Compactness, Semantics
  – Implied Cond. Independences in a Bnet
  – Stronger Independence assumptions :More compact Distributions and Structures

Recap Joint Distribution

H = True   H = False

- 3 binary random variables: **P(H,S,F)**
  - **H   dom(H)={h, ¬h}**   has heart disease, does not have…
  - **S   dom(S)={s, ¬s}**   smokes, does not smoke
  - **F   dom(F)={f, ¬f}**   high fat diet, low fat diet

Recap Joint Distribution

Joint Prob. Distribution (JPD)

- 3 binary random variables: **P(H,S,F)**
  - **H dom(H)={h, ¬h}** has heart disease, does not have…
  - **S dom(S)={s, ¬s}** smokes, does not smoke
  - **F dom(F)={f, ¬f}** high fat diet, low fat diet

	f		**¬ f**	
	s	**¬ s**	**s**	**¬ s**
h	.015	.007	.005	.003
¬ h	.21	.51	.07	.18

$\Sigma\ 1$

$2^3 - 1$

$2^K - 1$

Recap Marginalization

P(H,S,F)

	f: s	f: ¬ s	¬ f: s	¬ f: ¬ s
h	.015	.007	.005	.003
¬ h	.21	.51	.07	.18

$$P(H,S) = \sum_{x \in dom(F)} P(H,S,F=x)$$

P(H,S)?

	s	¬s	
h	.02	.01	.03
¬h	.28	.69	.97
	.3	.7	

P(H)?

P(S)?

Recap Conditional Probability

P(H,S)	**s**	**¬ s**	P(H)
h	.02	.01	.03
¬ h	.28	.69	$\boxed{.97}$
P(S)	.30	.70	

$$P(S \mid H) = \frac{P(S,H)}{P(H)}$$

$$P(s \mid \neg h) = \frac{P(s, \neg h)}{\boxed{P(\neg h)}}$$

Two probability distributions for S

$\boxed{P(S \mid H)}$	**s**	**¬ s**	
h	.666	.333	Σ 1
¬ h	.29	.71	Σ 1

P(H|S)

do this as an exercise

Recap Conditional Probability (cont.)

$$P(S \mid H) = \frac{P(S,H)}{P(H)}$$

P(S | H, F)

P(X_1 . . . X_n | Y_1 . . . Y_k)

binary

Two key points we covered in previous lecture

- We derived this equality from a possible world semantics of probability
- It is not a probability distributions but….. set of prob. distrib.

- One for each configuration of the conditioning var(s)

If conditioned by k binary vars, set 2^k prob. distributions

Product Rule

- Definition of conditional probability:
  - $P(X_1 \mid X_2) = P(X_1 , X_2) / P(X_2)$
- Product rule gives an alternative, more intuitive formulation:
  - $P(X_1 , X_2) = \boxed{P(X_2)}\, P(X_1 \mid X_2) = P(X_1)\, P(X_2 \mid X_1)$
- Product rule general form:

$$\mathbf{P}(X_1, \ldots, X_n) = P(x_1 \ldots X_t, X_{t+1} \ldots X_n)$$

$$= \mathbf{P}(X_1,\ldots,X_t)\, \mathbf{P}(X_{t+1} \ldots X_n \mid X_1,\ldots,X_t)$$

Chain Rule

- Product rule general form:

$\mathbf{P}(X_1, \ldots, X_n) =$

$= \mathbf{P}(X_1,\ldots,X_t)\ \mathbf{P}(X_{t+1} \ldots X_n \mid X_1,\ldots,X_t)$

- Chain rule is derived by successive application of product rule:

$t = n-1$

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

$= \mathbf{P}(X_1,\ldots,X_{n-1})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1})$

$= \mathbf{P}(X_1,\ldots,X_{n-2})\ \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) = \ldots.$

$= \mathbf{P}(X_1)\ \mathbf{P}(X_2 \mid X_1) \ldots \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1})$

$= \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$

Chain Rule: Example

P(*cavity* , *toothache*, *catch*) =

$P(cavity) * P(toothache \mid cavity) *$

$* P(catch \mid cavity, toothache)$

P(*toothache*, *catch*, *cavity*) =

$P(toothache) * P(catch \mid toothache) * P(cavity \mid toothache, catch)$

these and the other four decompositions are OK

Using conditional probability

- Often you have causal knowledge (forward from cause to evidence):
  - For example
    - ✓ P(symptom | disease)
    - ✓ P(light is off | status of switches and switch positions)
    - ✓ P(alarm | fire)
  - In general: P(evidence e | hypothesis h)

- ... and you want to do evidential reasoning (backwards from evidence to cause):
  - For example
    - ✓ P(disease | symptom)
    - ✓ P(status of switches | light is off and switch positions)
    - ✓ P(fire | alarm)
  - In general: P(hypothesis h | evidence e)

Bayes Rule

- By definition, we know that :

$$P(h \mid e) = \frac{P(h \wedge e)}{P(e)} \qquad P(e \mid h) = \frac{P(e \wedge h)}{P(h)}$$

- We can rearrange terms to write

$$P(h \wedge e) = P(h \mid e) \times P(e) \qquad (1)$$

$$P(e \wedge h) = P(e \mid h) \times P(h) \qquad (2)$$

- But

$$P(h \wedge e) = P(e \wedge h) \qquad (3)$$

- From (1) (2) and (3) we can derive
- **Bayes Rule**

$$P(h \mid e) = \frac{P(e \mid h) P(h)}{P(e)}$$

Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = ?$
- If there is a fire, the alarm will almost always ring
- On average, we have a fire every 10 years
- The fire alarm rings. What is the probability there is a fire?

Bayes Rule

$$P(h \mid e) = \frac{P(e \mid h) P(h)}{P(e)}$$

Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = 1/365$

- If there is a fire, the alarm will almost always ring
  - $P(alarm|fire) = 0.999$

- On average, we have a fire every 10 years
  - $P(fire) = 1/3650$

- The fire alarm rings. What is the probability there is a fire?
  - Take a few minutes to do the math!

0.999	0.9	0.0999	0.1

Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = 1/365$
- If there is a fire, the alarm will almost always ring
  - $P(alarm|fire) = 0.999$
- On average, we have a fire every 10 years
  - $P(fire) = 1/3650$
- The fire alarm rings. What is the probability there is a fire?
- $P(fire|\text{alarm}) = \frac{P(alarm|fire) \times P(fire)}{P(alarm)} = \frac{0.999 \times 1/3650}{1/365} = 0.0999$
  - Even though the alarm rings the chance for a fire is only about 10%!

Recap Chain Rule

$$P(H,S,F) = P(H) * P(S \mid H) * P(F \mid H,S)$$

$$\cancel{P(H)} * \frac{P(S,H)}{\cancel{P(H)}} * \frac{P(F,H,S)}{\cancel{P(H,S)}}$$

Bayes Theorem

$$\boxed{P(S \mid H) = \frac{P(S,H)}{P(H)}}$$

$$\boxed{P(H \mid S) = \frac{P(S,H)}{P(S)}}$$

substitute

rewrite

$$P(H \mid S)\,P(S) = P(S,H)$$

$$\boxed{P(S \mid H) = \frac{P(H \mid S)P(S)}{P(H)}}$$

Lecture Overview

- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - **Marginal** and Conditional **Independence**
- Bayesian Networks
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - Implied Cond. Independences in a Bnet
  - Stronger Independence assumptions :More compact Distributions and Structures

Do you always need to revise your beliefs?

…… when your knowledge of **Y**'s value doesn't affect your belief in the value of **X**

DEF. Random variable **X** is marginal independent of random variable **Y** if, for all $x_i \in dom(X)$, $y_k \in dom(Y)$,

$$P(X = x_i \mid Y = y_k) = P(X = x_i)$$

Consequence:

$$P(X = x_i, Y = y_k) = P(X = x_i \mid Y = y_k)\, P(Y = y_k) =$$

$$= P(X = x_i)\, P(Y = y_k)$$

Marginal Independence: Example

- X and Y are independent iff: $P(X) = P(X|Y) = \frac{P(X,Y)}{P(Y)}$

$\boxed{\mathbf{P}(X|Y) = \boxed{\mathbf{P}(X)}}$ or $\boxed{\mathbf{P}(Y|X) = \mathbf{P}(Y)}$ or $\boxed{\mathbf{P}(X, Y) = \mathbf{P}(X)\,\mathbf{P}(Y)}$

- That is new evidence Y(or X) does not affect current belief in X (or Y)

- Ex: **P**(*Toothache, Catch, Cavity, Weather*)
  = $\boxed{\mathbf{P}(\textit{Toothache, Catch, Cavity})}$ $\boxed{P(\text{weather})}$

|dom| = 4

Sunny
Cloudy
rainy
Snowy

- JPD requiring 32 entries is reduced to two smaller ones (8 and 4)

Joint prob. distribution

In our example are Smoking and Heart Disease marginally Independent ?

What our probabilities are telling us….?

P(H,S)	**s**	**¬ s**	P(H)
h	.02	.01	.03
¬ h	.28	.69	.97
P(S)	.30	.70	

$$P(S|H) \stackrel{?}{=} P(S)$$

No!

P(S|H)	**s**	**¬ s**
h	.666	.334
¬ h	.29	.71

$=?$

$=?$

Lecture Overview

- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - Marginal and **Conditional Independence**
- Bayesian Networks
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - Implied Cond. Independences in a Bnet
  - Stronger Independence assumptions :More compact Distributions and Structures

Conditional Independence

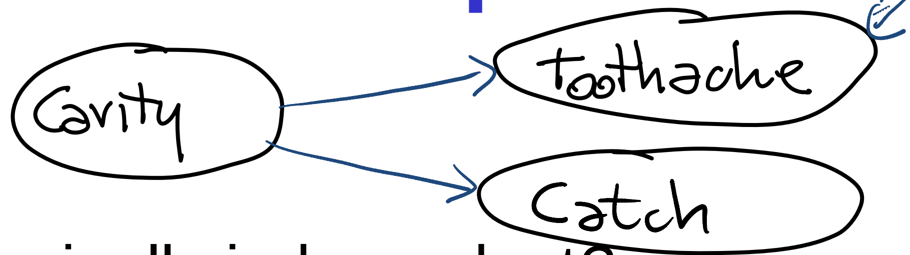
- With marg. Independence, for *n* independent random vars, $O(2^n) \rightarrow O(n)$

$$P(X_1 \ldots X_n) = P(X_1) \times \ldots \times P(X_n)$$

- Absolute independence is powerful **but** when you model a **particular domain**, it is rare..
- Dentistry is a large field with hundreds of variables, few of which are independent (e.g.,*Cavity, Heart-disease*).
- What to do?

Look for weaker form of independence

- **P**(*Toothache, Cavity, Catch*)



- Are *Toothache* and *Catch* marginally independent?

$P(\;|\;) = P(Toothache)$? NO

- BUT If I have a cavity, does the probability that the probe catches depend on whether I have a toothache? NO

  (1) P(*catch* | *toothache, cavity*) = $P(catch \mid cavity)$

- What if I haven't got a cavity?

  (2) P(*catch* | *toothache,¬cavity*) = $P(catch \mid \neg cavity)$

> - *Each is directly caused by the cavity, but neither has a direct effect on the other*

Conditional independence

- In general, *Catch* is conditionally independent of *Toothache* given *Cavity*:

1 **P**(*Catch* | *Toothache,Cavity*) = **P**(*Catch* | *Cavity*)

- Equivalent statements:

2 **P**(*Toothache* | *Catch, Cavity*) = **P**(*Toothache* | *Cavity*)

3 **P**(*Toothache, Catch* | *Cavity*) = **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*)

$$P(X,Y|Z) = P(X|Z)\,P(Y|Z)$$

Proof of equivalent statements

1 If $\boxed{P(X|YZ) = P(X|Z)} \Rightarrow$

$\Rightarrow$ A $\dfrac{P(X,Y,Z)}{P(Y,Z)} = \dfrac{P(X,Z)}{P(Z)} \Rightarrow$ 2

$\Rightarrow \dfrac{P(X,Y,Z)}{P(X,Z)} = \dfrac{P(Y,Z)}{P(Z)} \Rightarrow \boxed{P(Y|X,Z) = P(Y|Z)}$

3 $P(X,Y|Z) = \dfrac{P(X,Y,Z)}{P(Z)} \underset{\text{from A}}{\Rightarrow} \dfrac{P(Y,Z)\,P(X,Z)}{P(Z)} \cdot \dfrac{1}{P(Z)}$

$= \dfrac{P(Y,Z)}{P(Z)} \cdot \dfrac{P(X,Z)}{P(Z)} = \boxed{P(Y|Z) \cdot P(X|Z)}$

Conditional Independence: Formal Def.

Sometimes, two variables might not be marginally independent. However, they *become* independent after we observe some third variable

DEF. Random variable **X** is conditionally independent of random variable **Y** given random variable **Z** if, for all $x_i \in dom(X)$, $y_k \in dom(Y)$, $z_m \in dom(Z)$

$$P(X= x_i \mid Y= y_k , Z= z_m) = P(X= x_i \mid Z= z_m)$$

That is, knowledge of **Y**'s value doesn't affect your belief in the value of **X,** given a value of **Z**

Conditional independence: Use

- Write out full joint distribution using chain rule:

  P(*Cavity, Catch, Toothache*)

  = **P**(*Toothache* | *Catch, Cavity*) **P**(*Catch* | *Cavity*) **P**(*Cavity*)

  = **P**(*Toothache* | cavity) **P**(*Catch* | *Cavity*) **P**(Cavity)

  2 2 1

  how many probabilities? $2^3 - 1 = 7$

  $2 + 2 + 1 = 5$

- The use of conditional independence often reduces the size of the representation of the joint distribution from exponential in *n* to linear in *n*. **What is n?** # of vars

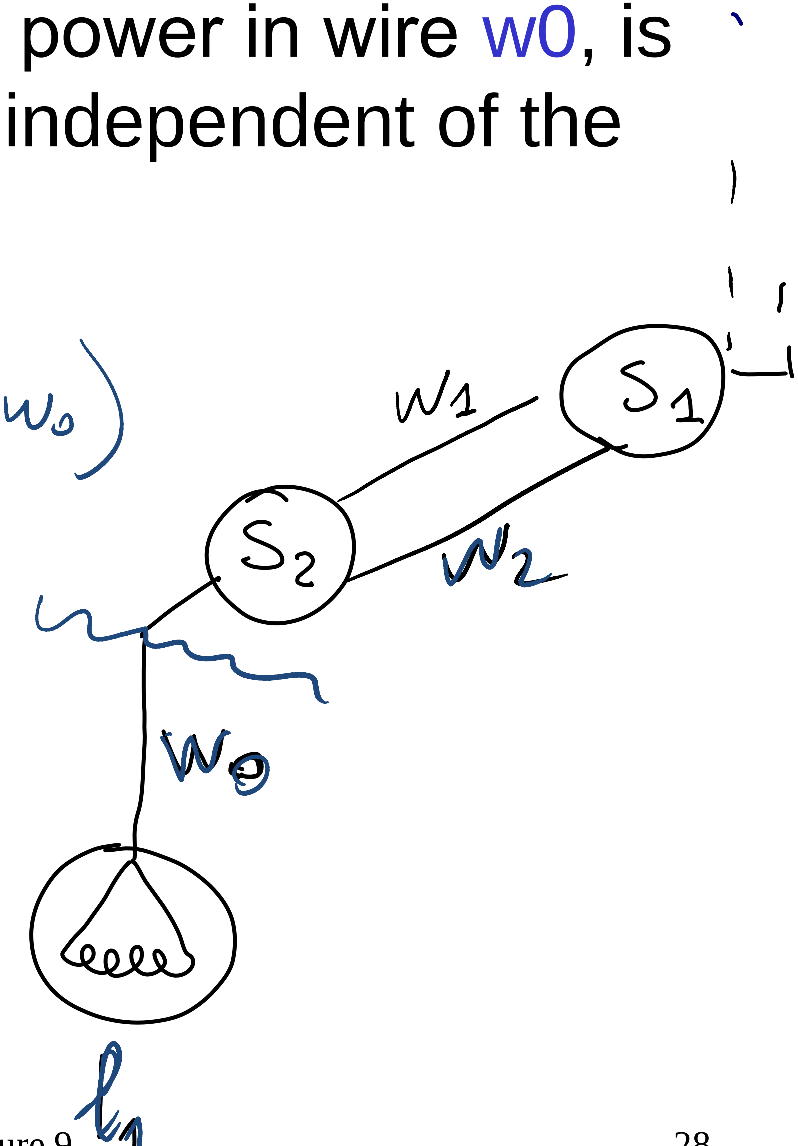
- **Conditional independence** is our **most basic** and **robust** form of **knowledge** about **uncertain environments**.

Conditional Independence Example 2

- Given whether there is/isn't power in wire w0, is whether light l1 is lit or not, independent of the position of switch s2?

$$P(l_1 \mid s_2, w_0) \stackrel{?}{=} P(l_1 \mid w_0)$$

yes!



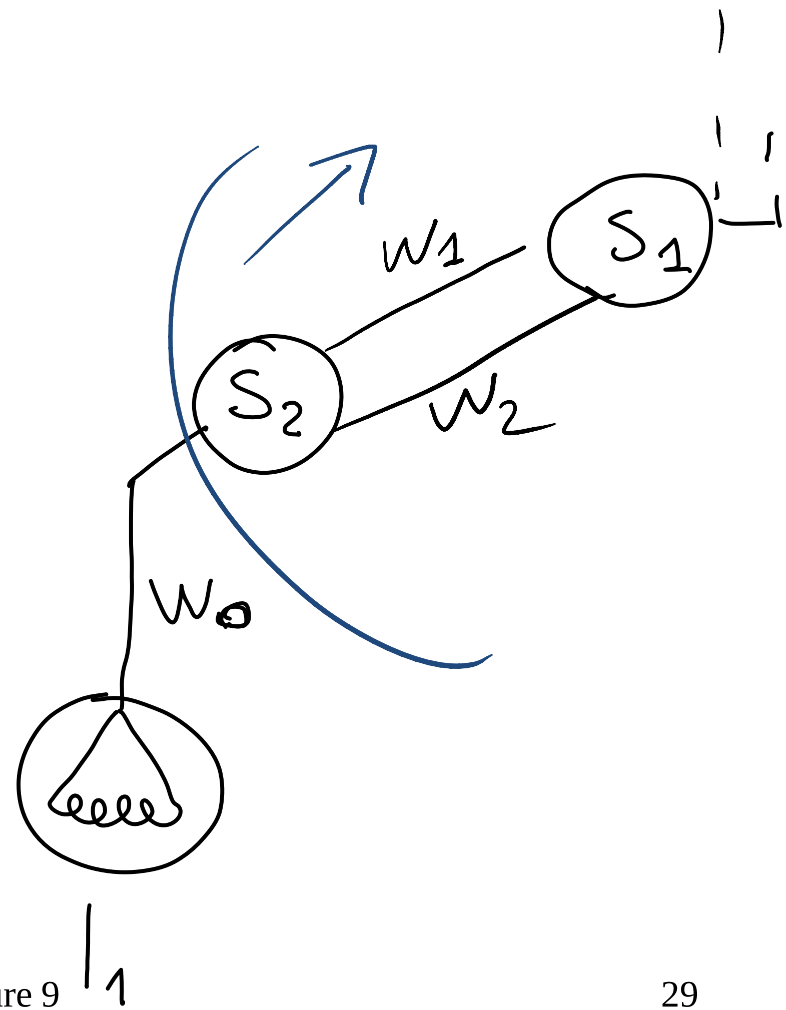
Conditional Independence Example 3

- Is every other variable in the system independent of whether light l1 is lit, given whether there is power in wire w0 ?

$$P(s_1 \mid l_1, w_0) = P(s_1 \mid w_0)$$

w_1
w_2
$\vdots$

yes!



Learning Goals for Prob. Intro

- **You can:**
- Given a joint, compute distributions over any subset of the variables
- Prove the formula to compute $P(h|e)$
- Derive the **Chain Rule** and **the Bayes Rule**
- Define **Marginal Independence**
- Define and use **Conditional Independence**

Where are we? (Summary)

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every possible world
- Queries can be answered by summing over possible worlds
- For nontrivial domains, we must find a way to reduce the joint distribution size
- **Independence** *(rare)* and **conditional independence** *(frequent)* provide the tools

Key points Recap

- We model the environment as a set of …. random vars

  $X_1 \dots X_n$ JPD $P(X_1 \dots X_n)$

- Why the joint is not an adequate representation ?

- “Representation, reasoning and learning” are “exponential” in ….. ~~#~~ vars

- **Solution:** Exploit marginal&**conditional** independence

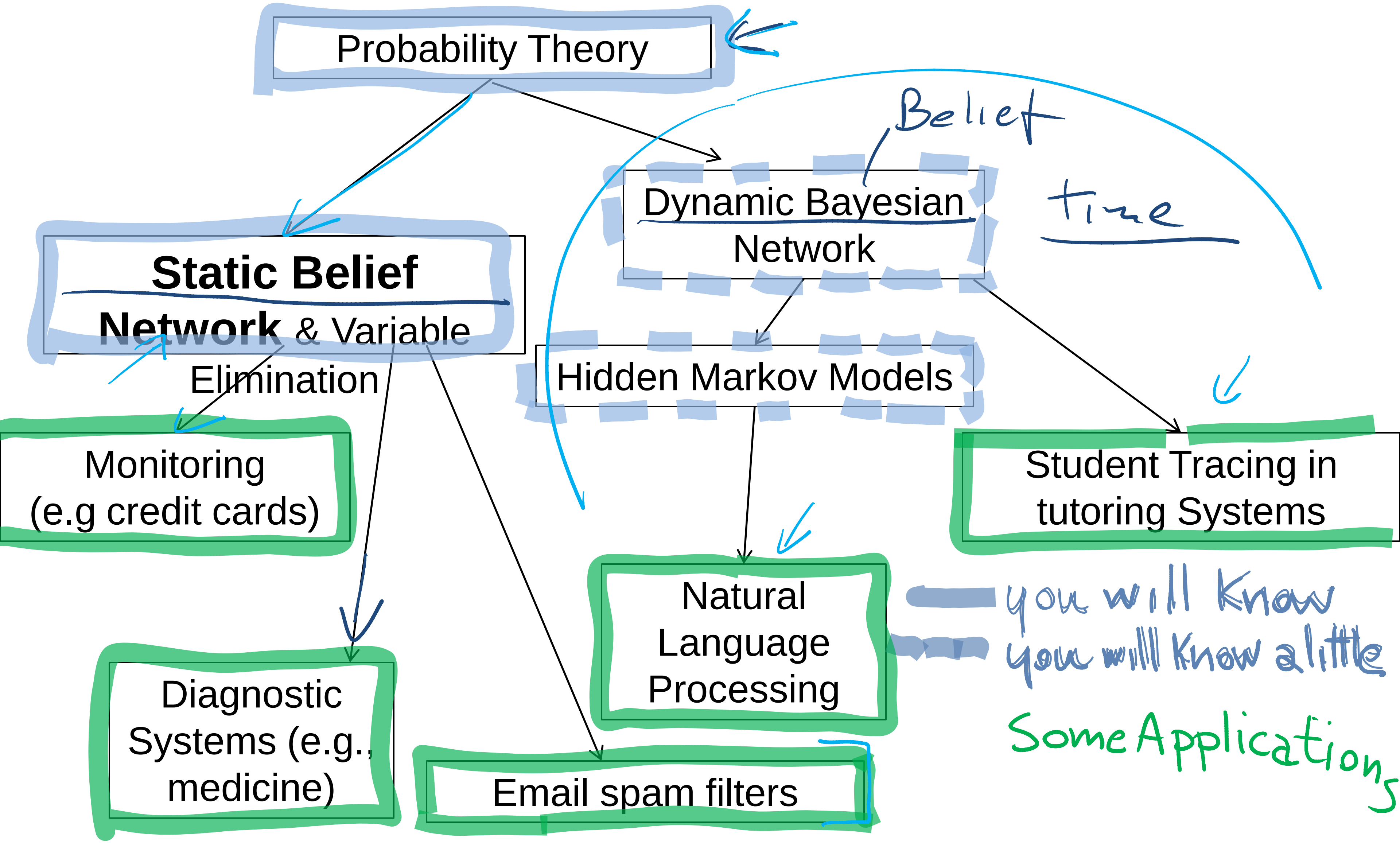
  $\boxed{P(X|Y)} = P(X)$ $\boxed{P(X|YZ)} = P(X|Z)$

- But how does independence allow us to simplify the joint? CHAIN RULE!

Lecture Overview

- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - Marginal and Conditional Independence
- **Bayesian Networks**
  - **Build sample BN**
  - Intro Inference, Compactness, Semantics
  - Implied Cond. Independences in a Bnet
  - Stronger Independence assumptions :More compact Distributions and Structures

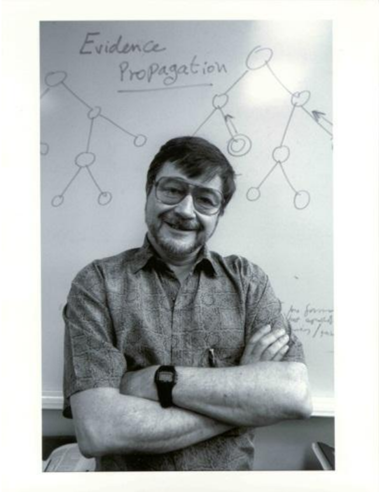
Answering Query under Uncertainty



Bayesian Network Motivation

- We want a representation and reasoning system that is based on conditional (and marginal) independence
  - Compact yet expressive representation
  - Efficient reasoning procedures
- Bayesian (Belief) Networks are such a representation
  - Named after Thomas Bayes (ca. 1702 –1761)
  - Term coined in 1985 by Judea Pearl (1936 –)
  - Their invention changed the primary focus of AI from logic to probability!

Thomas Bayes



Judea Pearl

Pearl just received the ACM Turing Award (widely considered the "Nobel Prize in Computing“) for his contributions to Artificial Intelligence!

Belief Nets: Burglary Example

- There might be a **burglar** in my house

B

- The **anti-burglar alarm** in my house may go off

A

- I have an agreement with two of my neighbors, **John** and **Mary**, that they **call** me if they hear the alarm go off when I am at work

M J

- **Minor earthquakes** may occur and sometimes the set off the alarm.

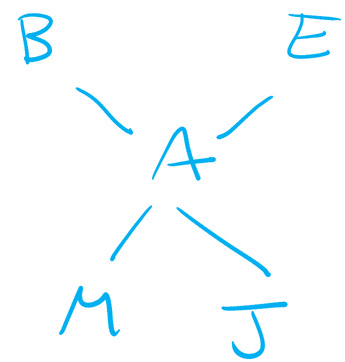
E

- **Variables:** B A M J E $n=5$

- **Joint** has 2^5-1 entries/probs 2^n-1

Belief Nets: Simplify the joint

- Typically order vars to reflect causal knowledge (i.e., causes *before effects)*
  - A burglar (B) can set the alarm (A) off
  - An earthquake (E) can set the alarm (A) off
  - The alarm can cause Mary to call (M)
  - The alarm can cause John to call (J)



$$P(B, E, A, M, J)$$

- Apply Chain Rule

marginal indep.

conditional indep.

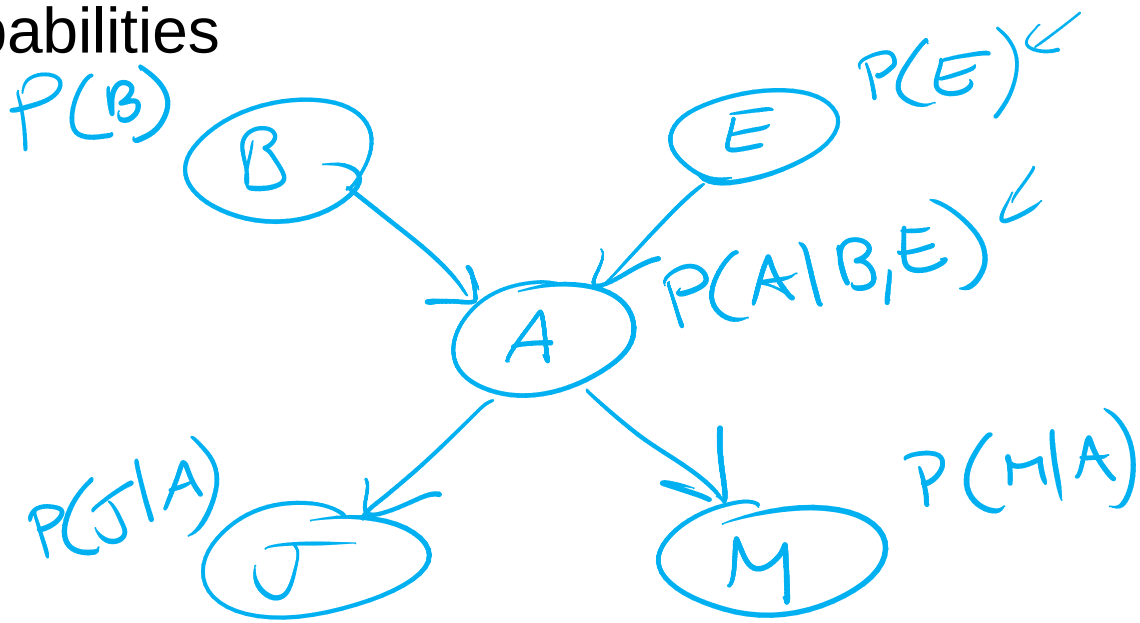
$$P(B)\ P(E \mid \cancel{B})\ P(A \mid B, E)\ P(M \mid A\,\cancel{E}\,\cancel{B})\ P(J \mid \cancel{M}\,A\,\cancel{E}\,\cancel{B})$$

- Simplify according to marginal&conditional independence

Belief Nets: Structure + Probs

$$P(B) * P(E) * P(A \mid \boxed{B, E}) * P(M \mid A) * P(J \mid A)$$

- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities



- Directed Acyclic Graph (DAG)

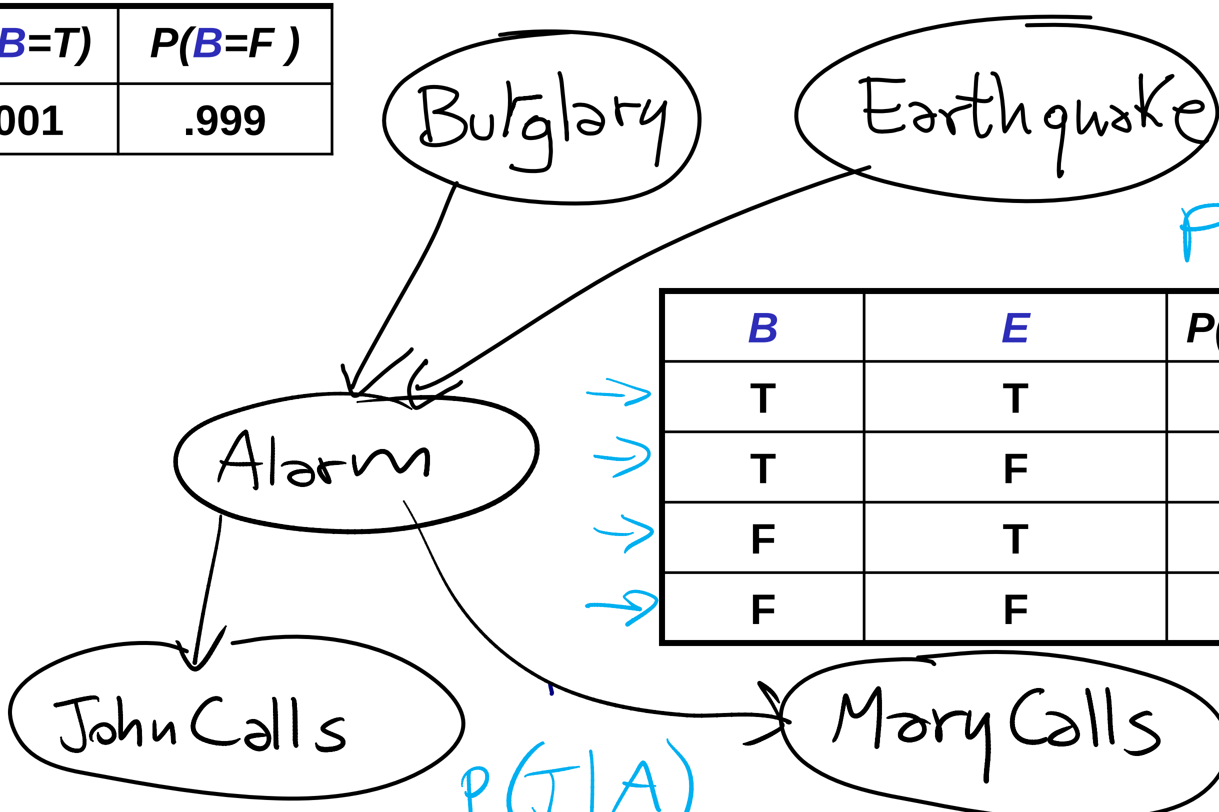
Burglary: complete BN

P(B)

P(B=T)	P(B=F)
.001	.999

P(E)

P(E=T)	P(E=F)
.002	.998



P(A | B, E)

B	E	P(A=T | B,E)	P(A=F | B,E)
T	T	.95	.05
T	F	.94	.06
F	T	.29	.71
F	F	.001	.999

P(J | A)

A	P(J=T | A)	P(J=F | A)
T	.90	.10
F	.05	.95

P(M | A)

A	P(M=T | A)	P(M=F | A)
T	.70	.30
F	.01	.99

call for any other reasons

Lecture Overview

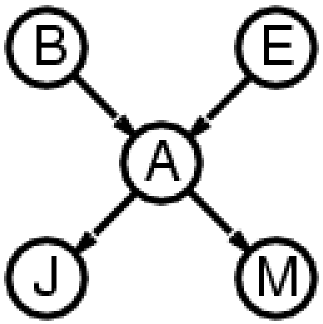
- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - Marginal and Conditional Independence
- Bayesian Networks
  - Build sample BN
  - **Intro Inference, Compactness, Semantics**
  - Implied Cond. Independences in a Bnet
  - Stronger Independence assumptions :More compact Distributions and Structures

BREAK
30 MINS

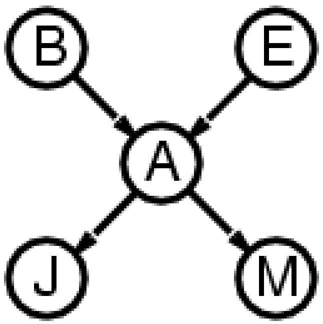
Burglary Example: Bnets inference

Our BN can answer any probabilistic query that can be answered by processing the joint!

- (Ex1) I'm at work,
  - neighbor John calls to say my alarm is ringing,
  - neighbor Mary doesn't call.
  - No news of any earthquakes.
  - Is there a burglar?

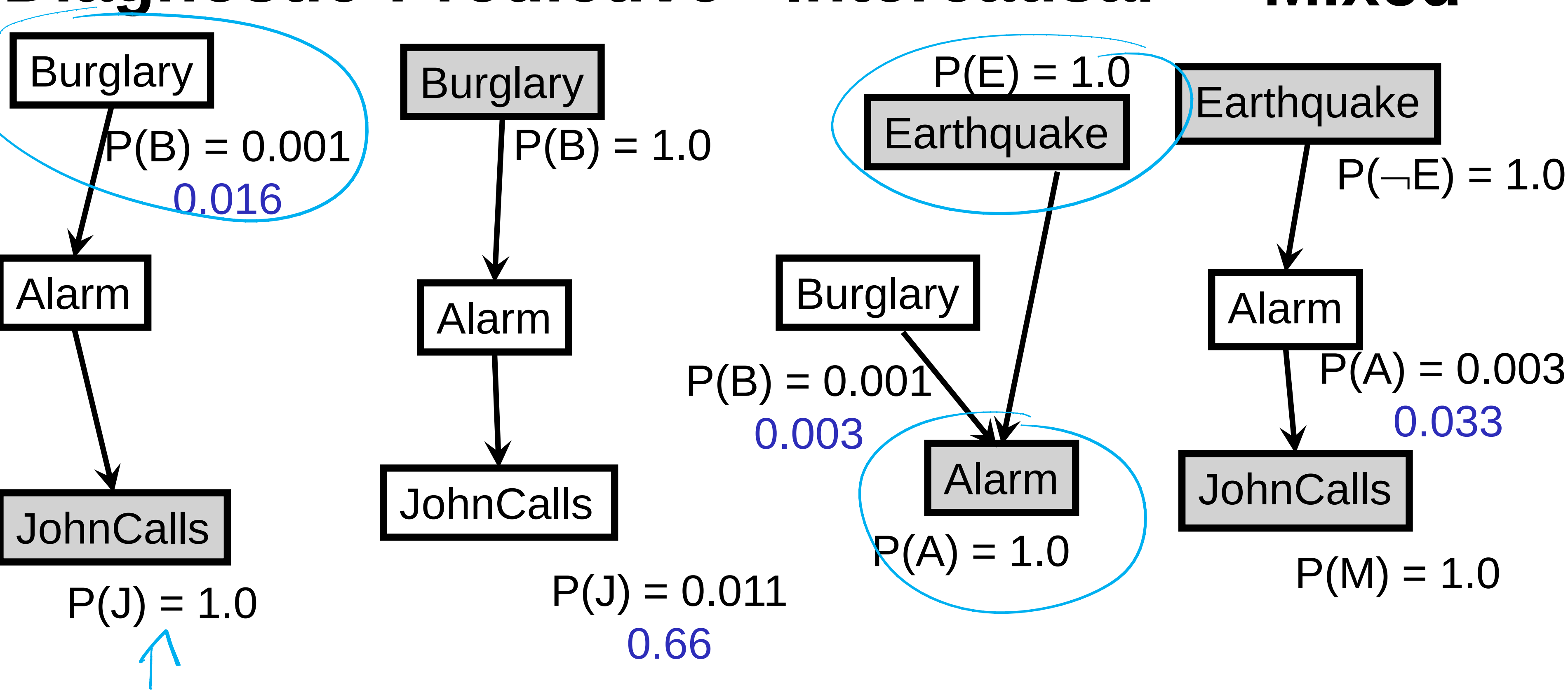


- (Ex2) I'm at work, Try this
  - Receive message that neighbor John called ,
  - News of minor earthquakes.
  - Is there a burglar?



Bayesian Networks – Inference Types

Diagnostic **Predictive** **Intercausal** **Mixed**



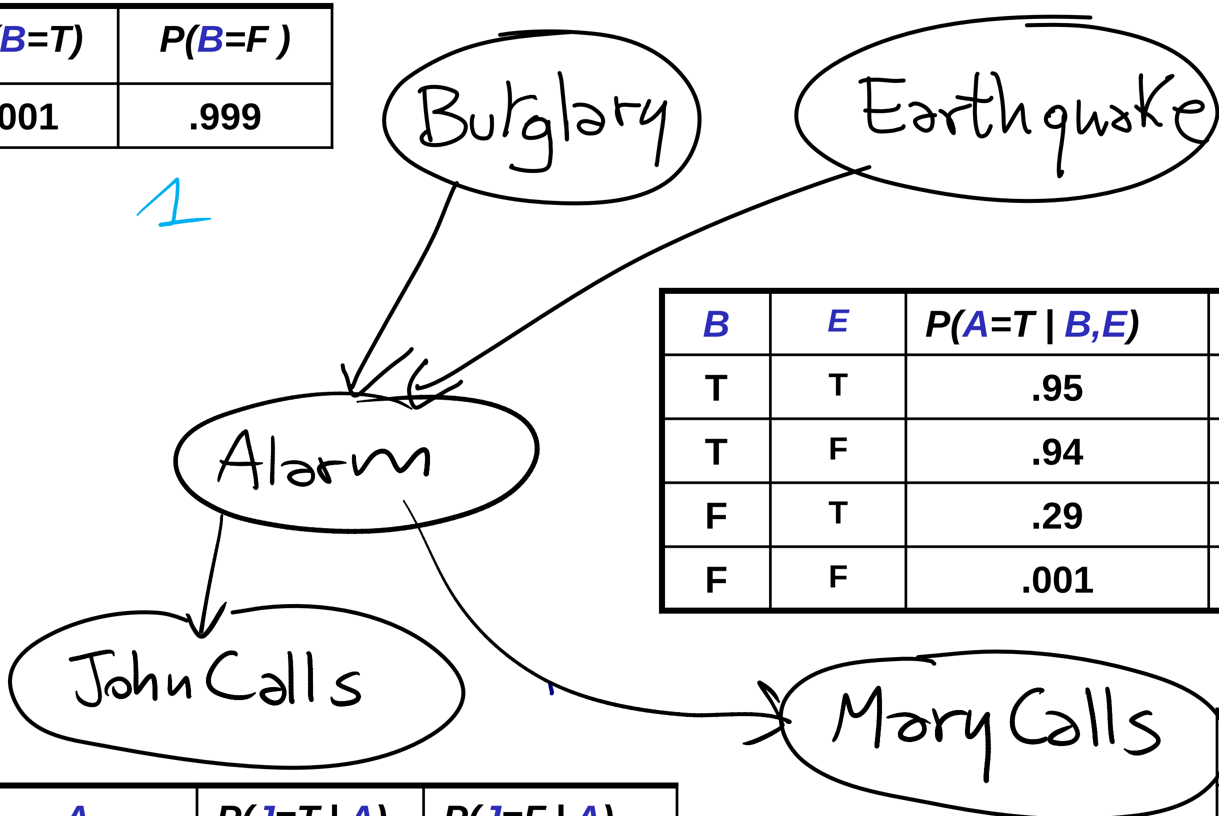
BNnets: Compactness

P(B=T)	*P(B=F)*
.001	.999

1

P(E=T)	*P(E=F)*
.002	.998

1



B	*E*	*P(A=T | B,E)*	*P(A=F | B,E)*
T	T	.95	.05
T	F	.94	.06
F	T	.29	.71
F	F	.001	.999

4

A	*P(J=T | A)*	*P(J=F | A)*
T	.90	.10
F	.05	.95

2

A	*P(M=T | A)*	*P(M=F | A)*
T	.70	.30
F	.01	.99

2

BNet

$2+2+4+1+1=10$

$|JPD| = 2^5 - 1$

BNets: Compactness

Conditional Probability Table

In General:

A $\boxed{\text{CPT}}$ for boolean X_i with k boolean parents has 2^k rows for the combinations of parent values

Each row requires one number p_i for X_i = *true* (the number for X_i = *false* is just $1-p_i$)

for each node

If each variable has no more than k parents, the complete network requires $O(n\,2^k)$ numbers

For $k \ll n$, this is a substantial improvement,

- the numbers required grow linearly with n, vs. $O(2^n)$ for the full joint distribution

BNets: Construction General Semantics

- The full joint distribution can be defined as the product of conditional distributions:
- $$\boldsymbol{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \boldsymbol{P}(X_i \mid X_1, \ldots, X_{i-1})$$ (chain rule)
- Simplify according to marginal&conditional independence
- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities

$$\boldsymbol{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \boldsymbol{P}(X_i \mid Parents(X_i))$$

How to build a Bayesian network

- Define a total order over the random variables: $(X_1, \ldots, X_n)$
- If we apply the chain rule, we have

  $P(X_1, \ldots, X_n) = \prod^{n}_{i=1} P(X_i \mid X_1, \ldots, X_{i-1})$

Predecessors of X_i in the total order defined over the variables

- Define as parents of random variable X_i in the Belief network a minimal set of its predecessors $Parents(X_i)$ such that
- $P(X_i \mid X_1, \ldots, X_{i-1}) = P(X_i \mid Parents(X_i))$

X_i is conditionally independent from all its other predecessors given *$Parents(X_i)$*

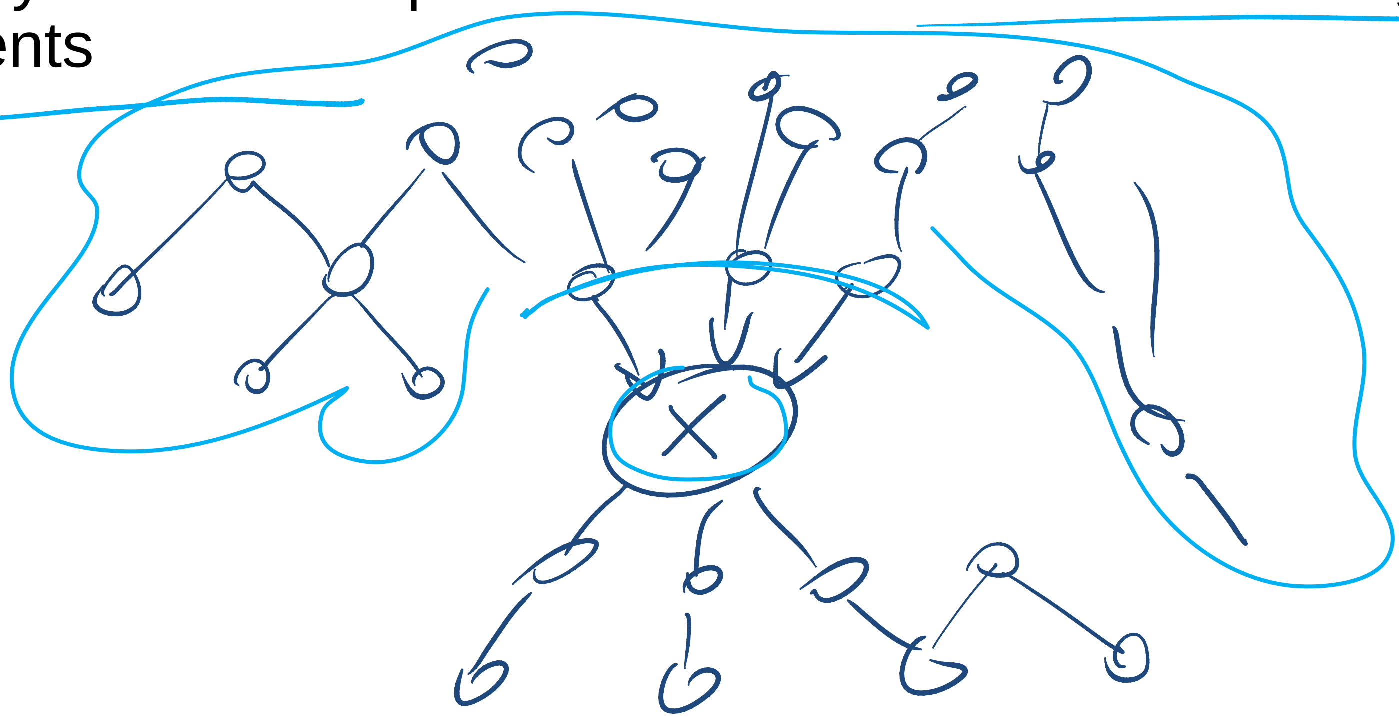
- Putting it all together, in a Belief network
- $P(X_1, \ldots, X_n) = \prod^{n}_{i=1} P(X_i \mid Parents(X_i))$

A Belief network defines a factorization over the JDP for its variables, based on existing conditional independencies among these variables

BNets: Construction General Semantics (cont')

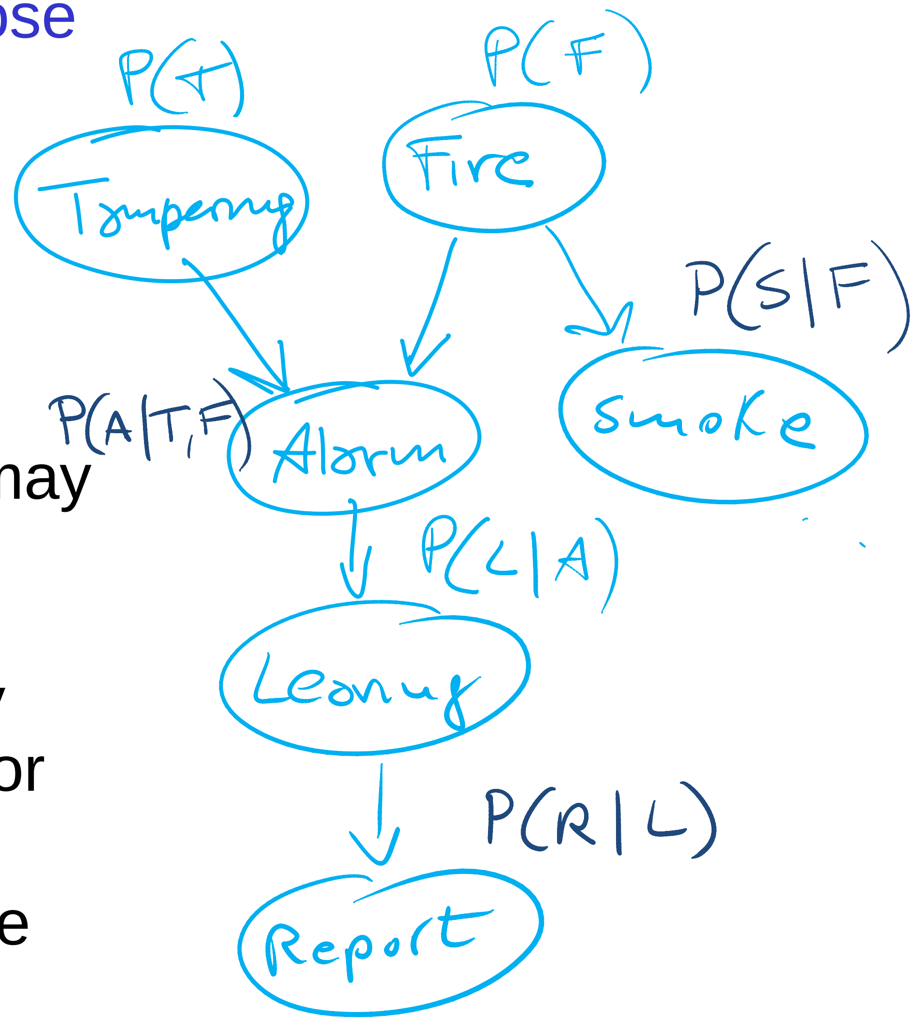
$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

- Every node is independent from its non-descendants given it parents



Other Examples: Fire Diagnosis (textbook Ex. 6.10)

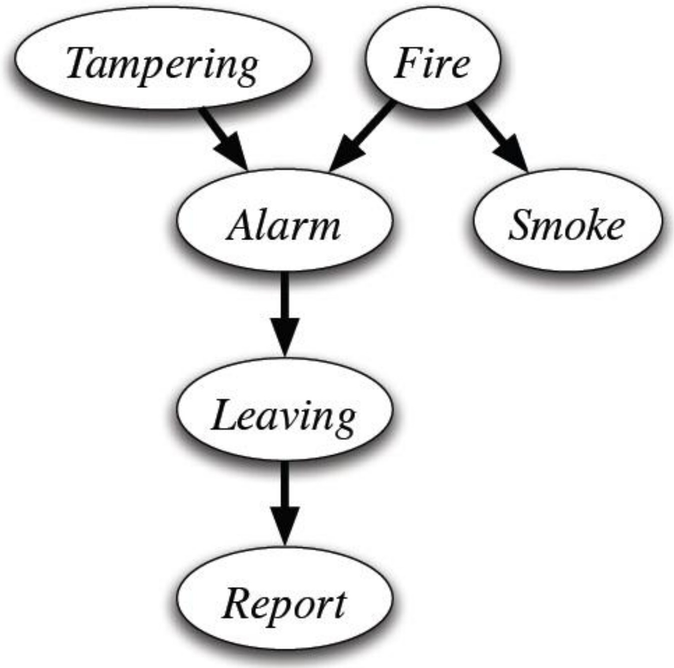
- Suppose you want to diagnose whether there is a fire in a building
- you receive a noisy report about whether everyone is leaving the building.
- if everyone is leaving, this may have been caused by a fire alarm.
- if there is a fire alarm, it may have been caused by a fire or by tampering
- if there is a fire, there may be smoke raising from the bldg.



Example for BN construction: Fire Diagnosis

- We are not done yet: must specify the Conditional Probability Table (CPT) for each variable. All variables are Boolean.
- How many probabilities do we need to specify for this Bayesian network?
  - This time taking into account that probability tables have to sum to 1 6 12 20 2^6-1

Example for BN construction: Fire Diagnosis



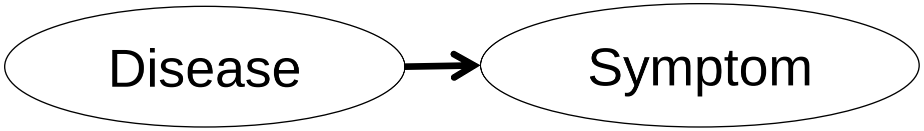
- We are not done yet: must specify the Conditional Probability Table (CPT) for each variable. All variables are Boolean.
- How many probabilities do we need to specify for this Bayesian network?
  - ✓ P(Tampering): 1 probability
  - ✓ P(Alarm|Tampering, Fire): 4 (independent)

    1 probability for each of the 4 instantiations of the parents
  - ✓ In total: 1+1+4+2+2+2 = 12 (compared to 2^6 -1= 63 for full JPD!)

Recap of BN construction with a small example

- Which (conditional) probability tables do we need?

P(D) P(D|S) P(S|D) P(D,S)

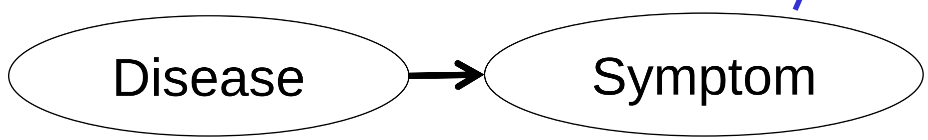


Recap of BN construction with a small example

- Which conditional probability tables do we need?
  - P(D) and P(S|D)
  - In general: for each variable X in the network: P(X|Pa(X))

P(D=t)

Disease *D*	*P(S=t|D)*
t	
f	



Other Examples (cont')

- Make sure you explore and understand the **Fire Diagnosis** example (we'll expand on it to study Decision Networks)

- **Electrical Circuit** example (textbook ex 6.11)

- **Patient's wheezing and coughing** example (ex. 6.14)

- Several other examples on

Realistic BNet: Liver Diagnosis

JPD

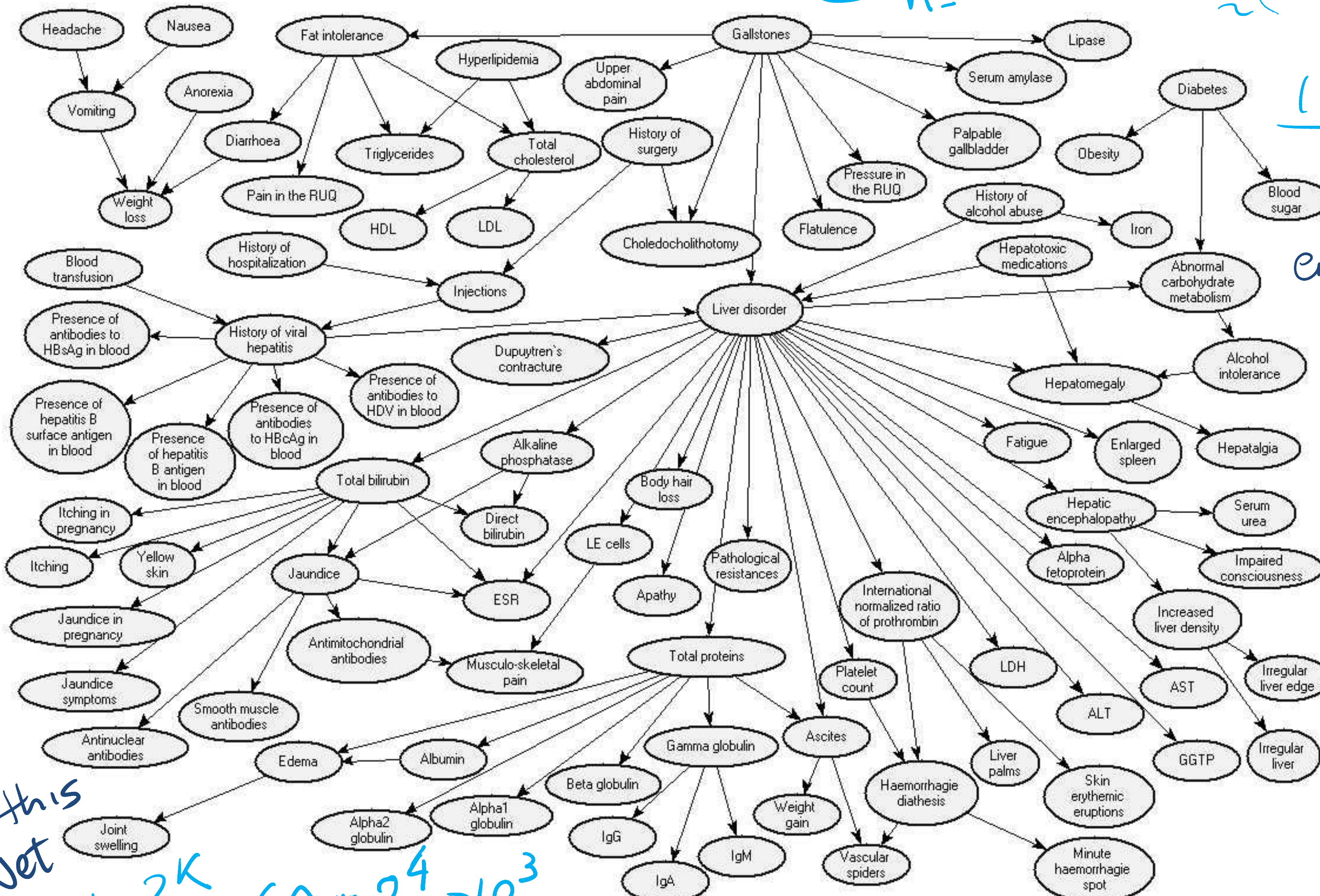
Source: Onisko et al., 1999

~60 nodes

$n = \sim 60$ $\sim 2^{60}$ $= (2^{10})^6$

10^{18}

entries

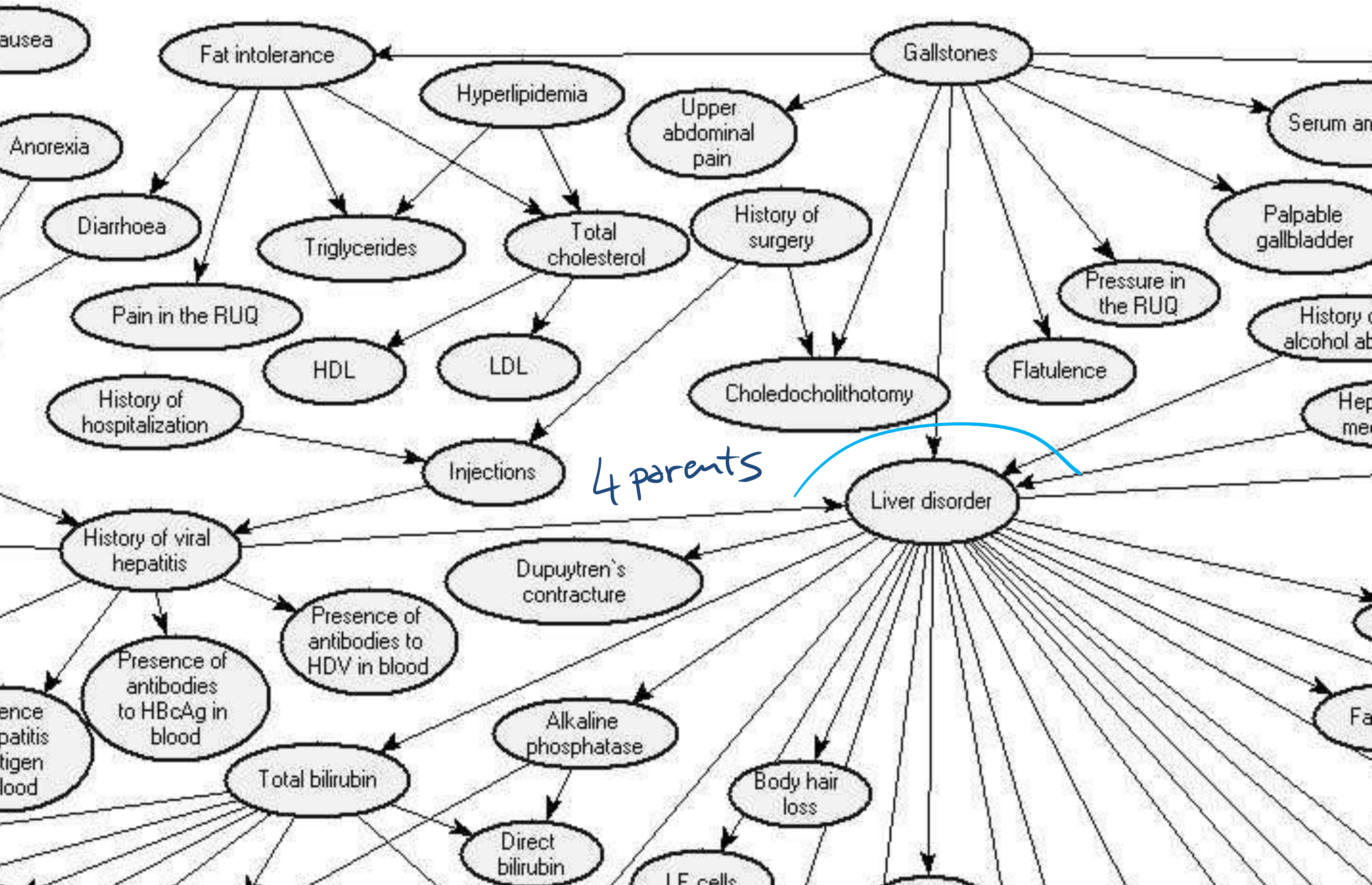


for this BNet

$n\,2^K$ $\quad 60 * 2^4 = 10^3$

Realistic BNet: Liver Diagnosis

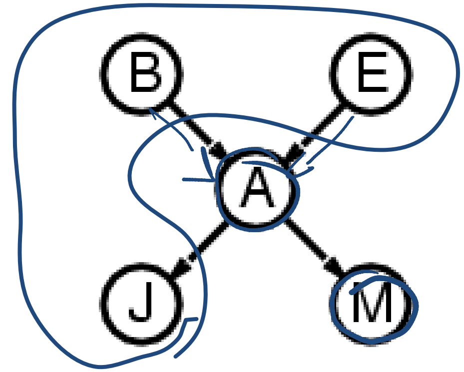
Source: Onisko et al., 1999



Belief network summary

- A belief network is a directed acyclic graph (DAG) that effectively expresses independence assertions among random variables.
- The parents of a node *X* are those variables on which *X* directly depends.
- Consideration of causal dependencies among variables typically help in constructing a Bnet

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 $\underline{\text{indep}}$ (B, E, J) given $\underline{A}$ $\quad P(M, \underbrace{B, E, J}, A)$

- 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 form 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

n vars

k parents

- Compactness: We reduce the number of probabilities from $O(2^n)$ to $O(n2^k)$

In some domains we need to do better than that!

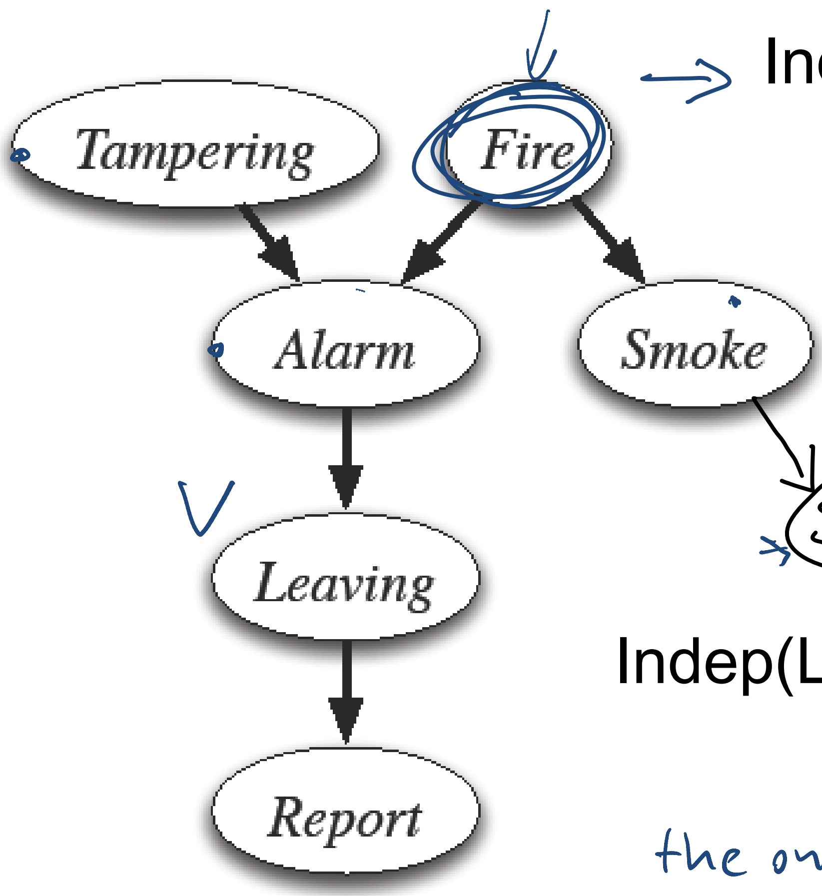
- Still too many and often there are no data/experts for accurate assessment

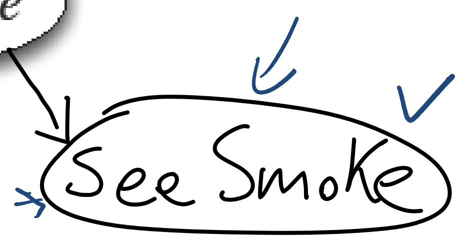
Solution: Make stronger (approximate) independence assumptions

Lecture Overview

- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - Marginal and Conditional Independence
- Bayesian Networks
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - **Implied Cond. Independences in a Bnet**
  - Stronger Independence assumptions :More compact Distributions and Structures

Bnets: Entailed (in)dependencies





Indep(Report, Fire,{Alarm})?

yes

If you know the value of Alarm, knowledge of the value of Fire does not affect your belief in the value of Report

Indep(Leaving, SeeSmoke,{Fire})?

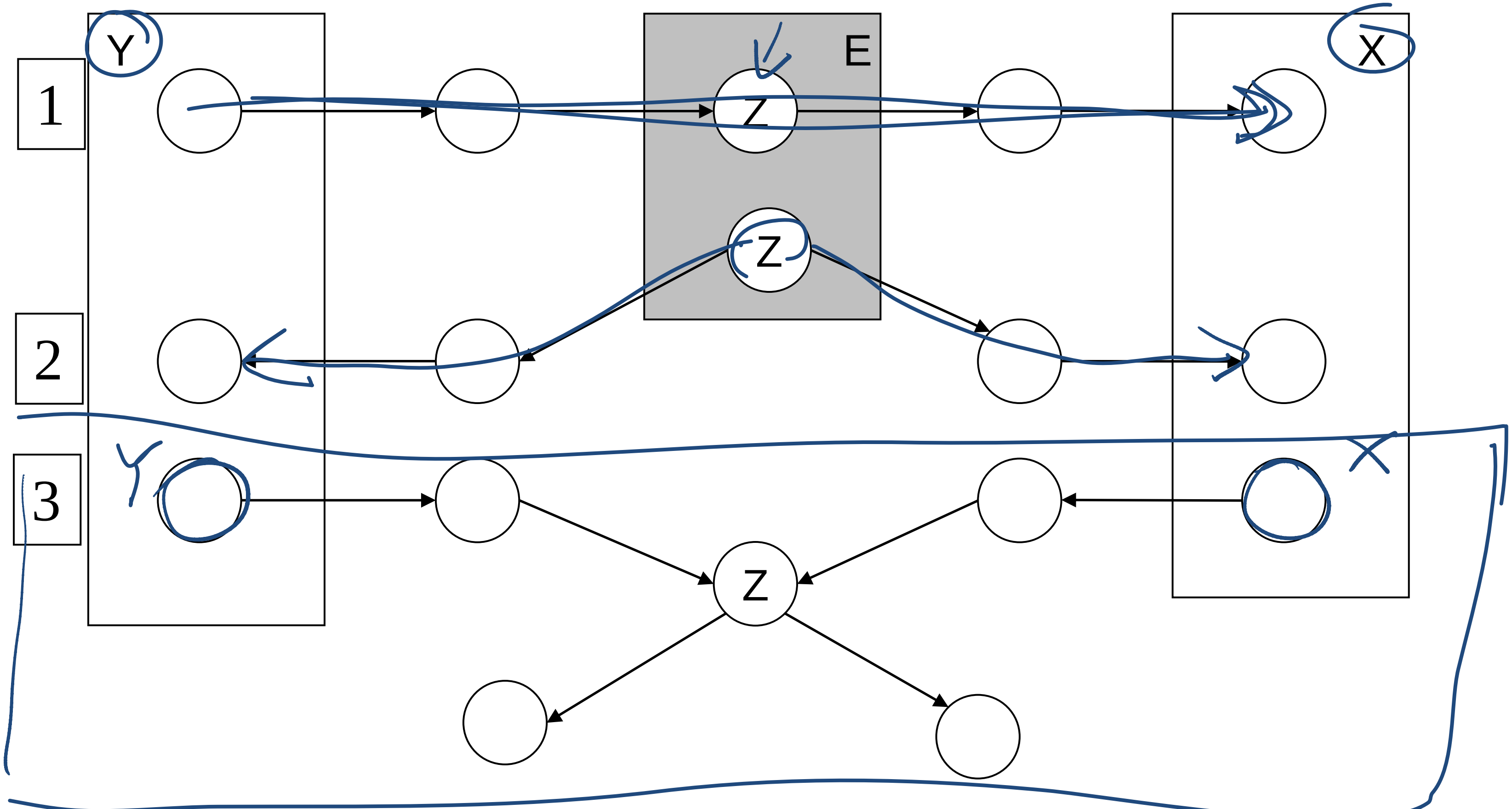
yes

the only dependency between Leaving and SeeSmoke is through Fire

If you know the value of Fire the two become independent

Conditional Independencies

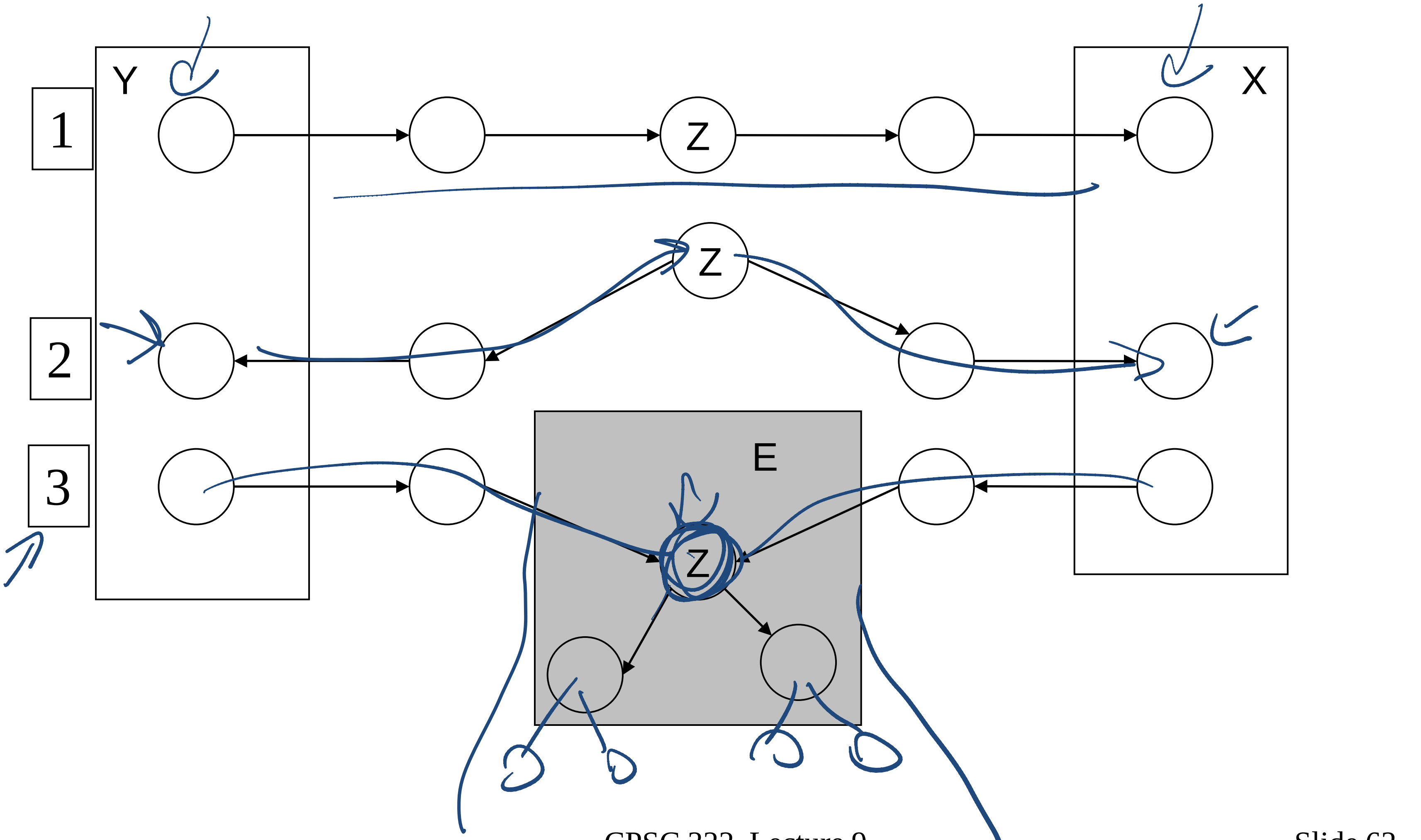
- Or, blocking paths for probability propagation. Three ways in which a path between X to Y can be blocked, (1 and 2 given evidence E)



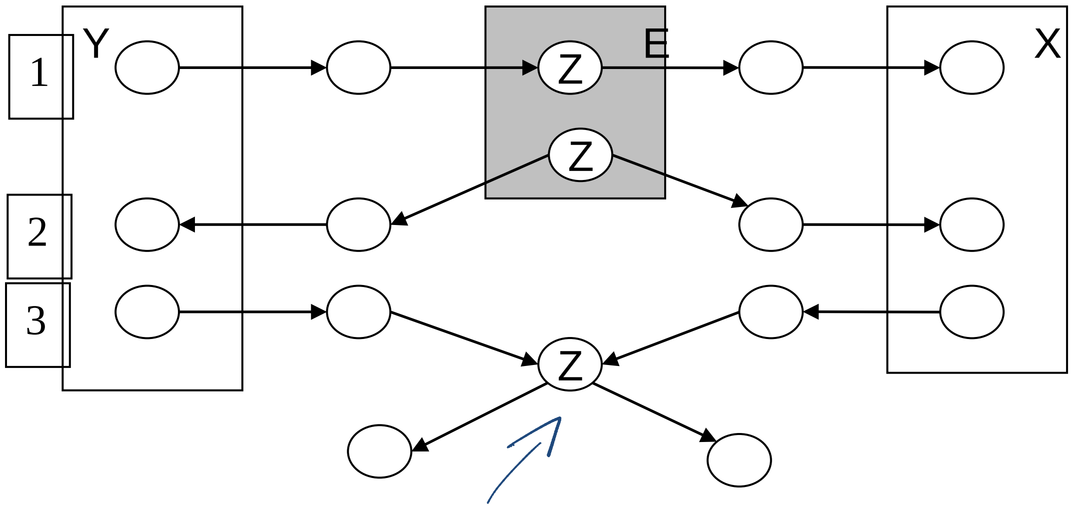
Note that, in 3, X and Y become dependent as soon as I get evidence on Z or on *any of its descendants*

Or ….Conditional Dependencies

In 1, 2, 3 X Y are dependent

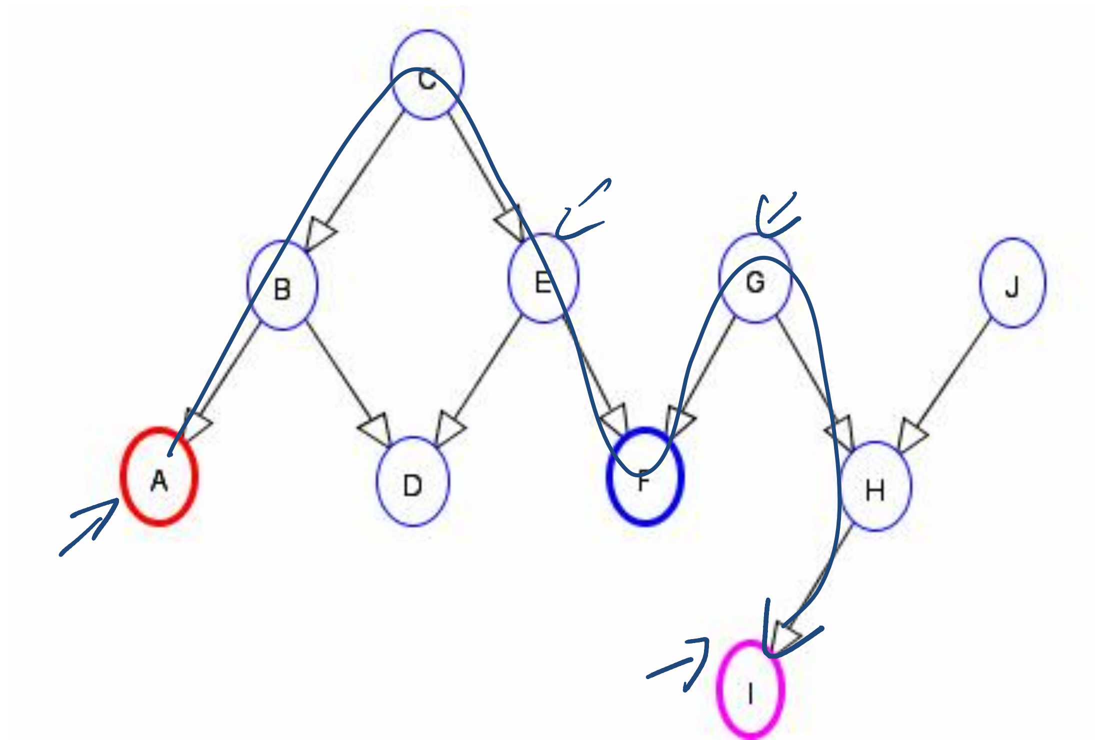


In/Dependencies in a Bnet : Example 1

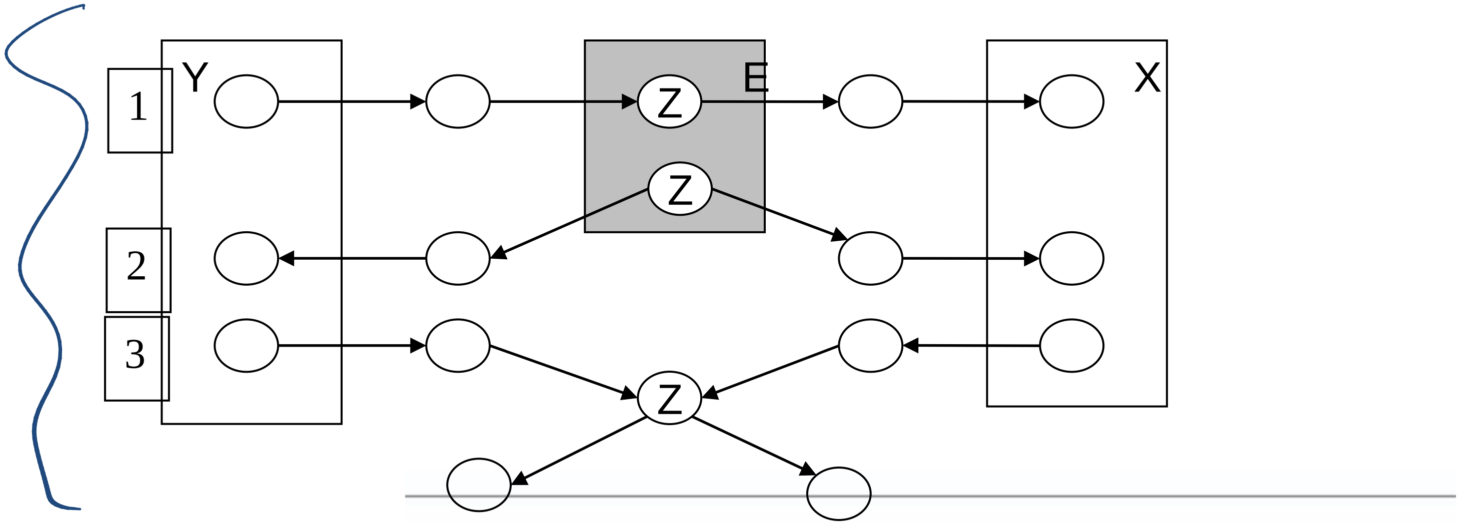


Is A conditionally independent of I given F?

false

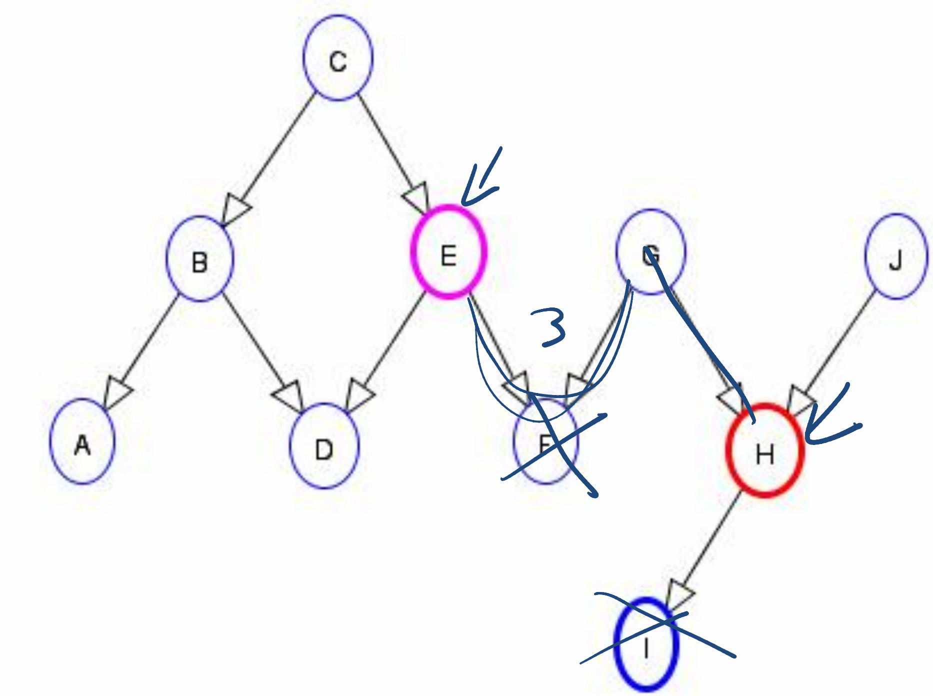


In/Dependencies in a Bnet : Example 2



Is H conditionally independent of E given ~~I~~?

true



Lecture Overview

- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - Marginal and Conditional Independence
- Bayesian Networks
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - Implied Cond. Independences in a Bnet
  - Stronger Independence assumptions: **More compact Distributions** and Structures

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

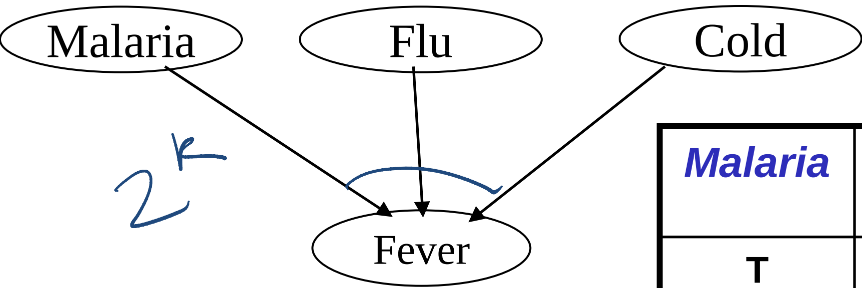
From JointPD 2^n to $n2^K$

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?

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
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.

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
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)

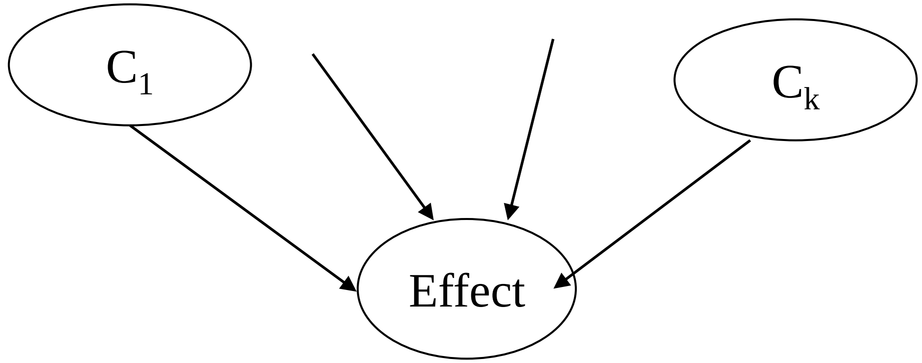
Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F	0	1

u on zeros

Two assumptions

1. All possible causes a 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(Effect=F | C_i = T, *and no other causes*) = q_i
- P(Effect=F | C_1 = T,.. C_j = T, C_{j+1} = F,., C_k = F)= $\prod_{i=1}^{j} q_i$
- P(Effect=T | C_1 = T,.. C_j = T, C_{j+1} = F,., C_k = F) = $1 -$

Noisy-OR: Example

q_i

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

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

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

Model of internal medicine

$\underbrace{133{,}931{,}430}$ → $\underbrace{8{,}254}$

using Noisy-ORs

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

Malaria	*Flu*	*Cold*	*P(Fever=T | ..)*	*P(Fever=F|..)*
T	**T**	**T**	.988	$\underline{0.1}$ x $\underline{0.2}$ x $\underline{0.6}$ **= 0.012**
T	**T**	**F**	.98	$\underline{0.2}$ x $\underline{0.1}$ **= 0.02**
T	**F**	**T**	.94	0.6 x 0.1**=0.06**
T	**F**	**F**	**0.9**	**0.1**
F	**T**	**T**	.88	0.2 x 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

- Finish Intro to Probability
  - Chain Rule and Bayes' Rule
  - Marginal and Conditional Independence
- Bayesian Networks
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - Implied Cond. Independences in a Bnet
  - Stronger Independence assumptions: More compact Distributions and Structures

Naïve Bayesian Classifier

A very simple and successful Bnets 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

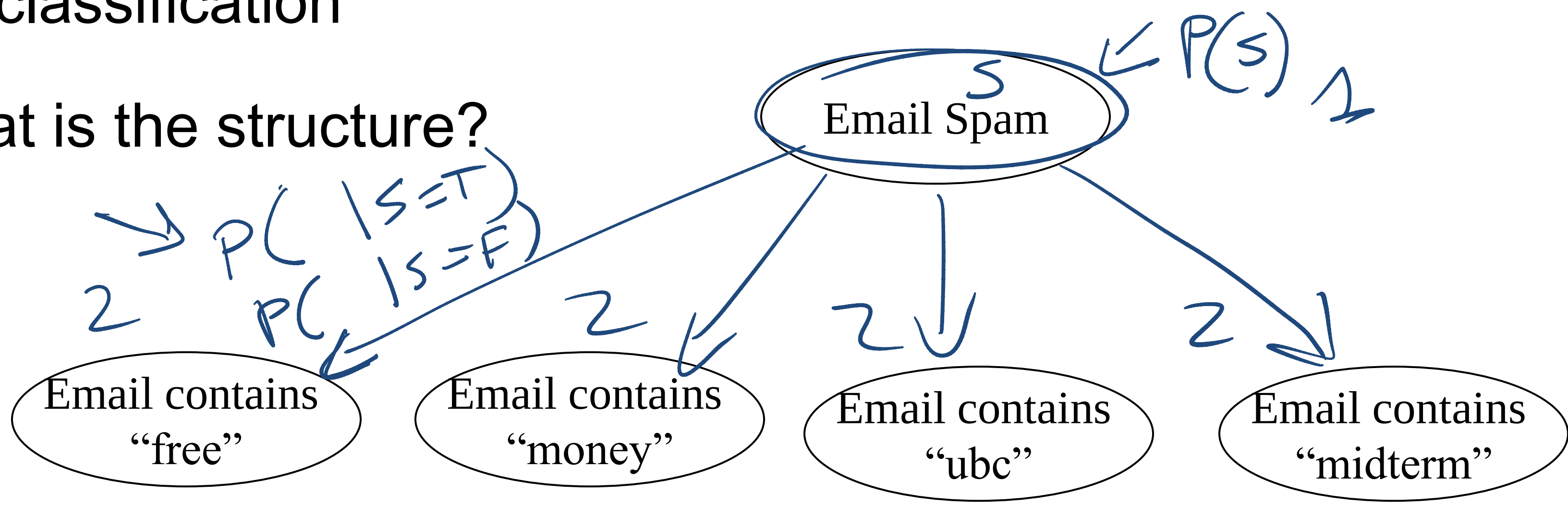
P("bank" | "account" , spam=T) = P("bank" | spam=T)

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?



words

10^5 in English

Number of parameters?

$2 \cdot 10^5$ vs 2^{10^5} joint

Easy to acquire?

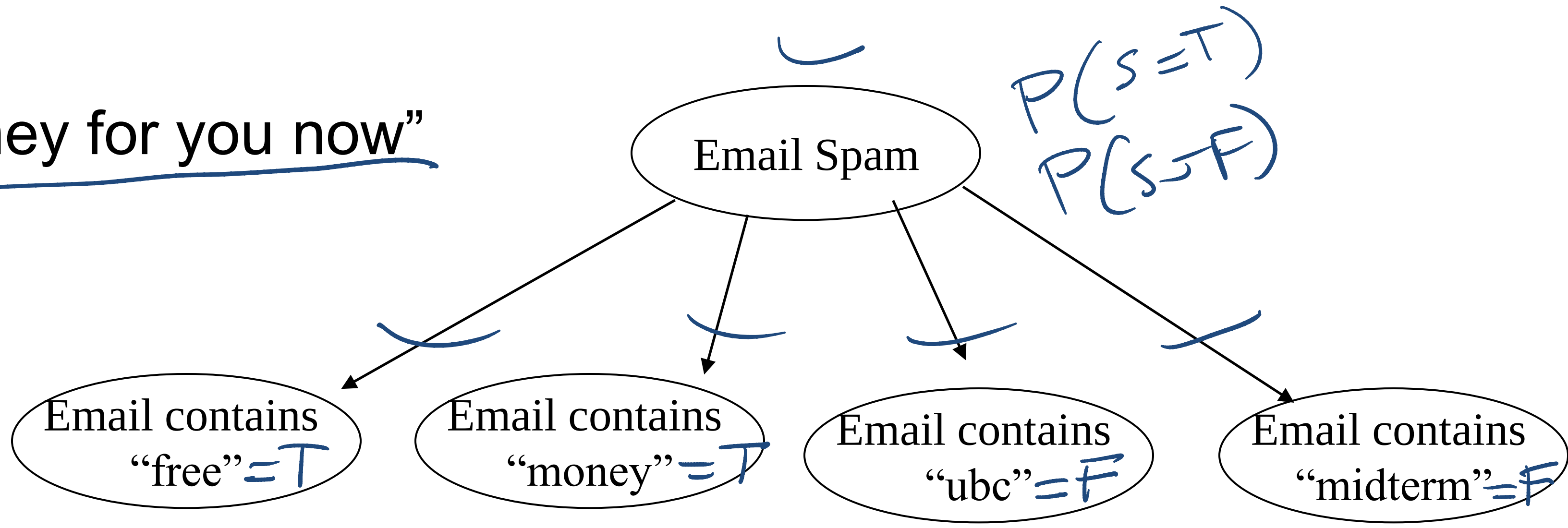
If you have a large collection of emails for which you know if they are spam or not……

NB Classifier for Email Spam: Usage

Most likely class given set of observations

Is a given Email *E* spam?

“free money for you now”



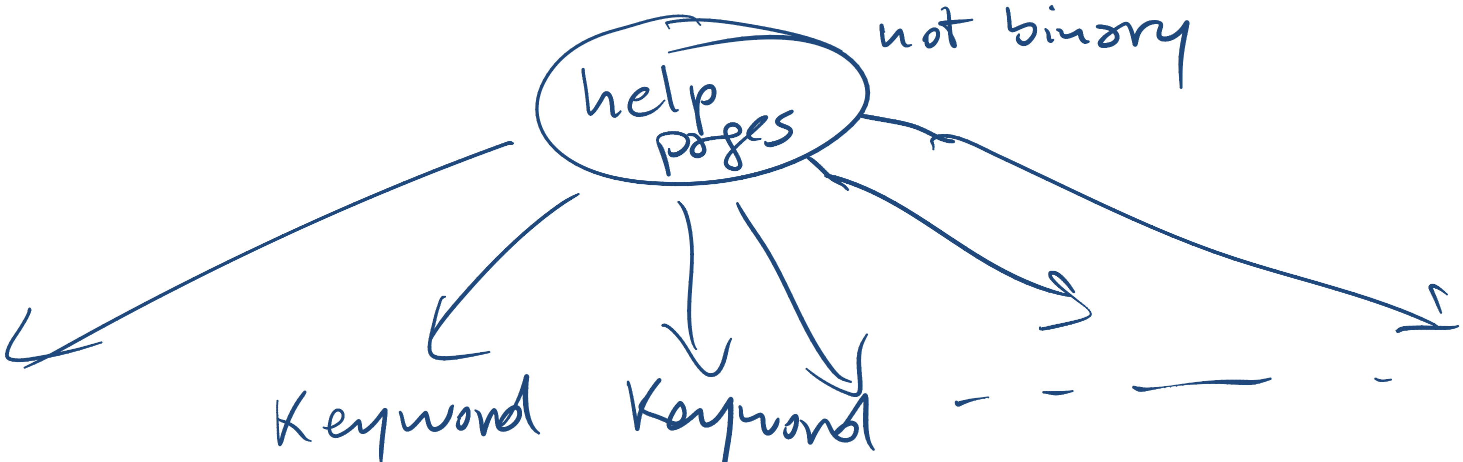
Email is a spam if……. $P(S=T) > P(S=F)$

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 part-2

- **You can:**
- Given a Belief Net,
- Compute the representational saving in terms on number of probabilities required
- 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 Exercises Reasoning under Uncertainty:**
  - **Ex 6.A**: conditional independence and intro to belief networks
  - **Ex 6.B:** more on belief networks
- **Assignment 3** is due now !
- Assignment 4 will be available on Wed and will be self assessed.

Learning Goals for today's class

- **You can:**
- Build a Belief Network for a simple domain
- Classify the types of inference

Diagnostic, Predictive, Intercausal, Mixed

- Compute the representational saving in terms on number of probabilities required

Lecture Overview

- Recap with Example
  - Marginalization
  - Conditional Probability
  - Chain Rule

- Bayes' Rule
- Marginal Independence
- Conditional Independence

our most basic and robust form of knowledge about uncertain environments.

Big Picture: R&R systems

- **Environment**

Problem		Deterministic	Stochastic
Static	Constraint Satisfaction	*Vars + Constraints* Arc Consistency Search → for CSP SLS	
	Query	*Logics* CSP Search → for Inference	*Belief Nets* Var. Elimination
Sequential	Planning	*STRIPS* CSP Search → for complex planning	*Decision Nets* Var. Elimination *Markov Processes* Value Iteration

Representation

Reasoning Technique