Introduction to Artificial Intelligence (AI)

Computer Science cpsc502, Lecture 7

Oct, 4, 2011

Today Oct 4

Finish R&R systems in deterministic environments

- Logics
  - Reasoning with individuals and relations
  - Full Propositional Logics and First-Order Logics

- **Start R&R systems in Stochastic environments**
  - Bayesian Networks Representation

R&Rsys we'll cover in this course

Environment

Problem		Deterministic	Stochastic
Static	Constraint Satisfaction	*Vars + Constraints* Arc Consistency Search SLS	
	Query	*Logics* → Propositional → First Order → . . . . Search	*Belief Nets* Var. Elimination Approx. Inference Temporal. Inference
Sequential	Planning	*STRIPS* actions precs effects Search	aka influence diagrams *Decision Nets* Var. Elimination *Markov Processes* Value Iteration

Representation

Reasoning Technique

Logics: Big Picture

First Order Logic

$\forall X \exists Y p(X,Y) \Longleftrightarrow \neg q(Y)$

$p(a_1, a_2)$

$\neg q(a_5)$

Datalog

$p(X) \leftarrow q(X) \wedge r(X,Y)$

$r(X,Y) \leftarrow s(Y)$

$s(a_1), q(a_2)$

Propositional Logic

$\neg(p \vee q) \rightarrow (r \wedge s \wedge f),$

p, r

PDCL

$p \leftarrow s \wedge f$

$r \leftarrow s \wedge q \wedge p$

r

p

Representation and Reasoning in Complex domains

- In complex domains expressing knowledge with **propositions** can be quite limiting

p $\boxed{up}\_s_2$
q $\boxed{up}\_s_3$
r $\boxed{ok}\_cb_1$
s $\boxed{ok}\_cb_2$
$live\_w_1$
$connected\_w_1\_w_2$

- It is often natural to consider **individuals** and their **properties**

$up(s_2)$
$up(s_3)$
$ok(cb_1)$
$ok(cb_2)$
$live(w_1)$
$connected(w_1, w_2)$

There is no notion that the system can reason about

$\boxed{up\_s_2 \quad up\_s_3}$ up are about the same property

$live\_w_1$
$connected\_w_1\_w_2$ w_1 are about the same individual

What do we gain….

By breaking propositions into relations applied to individuals?

- Express **knowledge** that **holds for set of individuals** (by introducing variables)

$$live(W) \text{ <- } connected\_to(W,W1) \wedge live(W1) \wedge wire(W) \wedge wire(W1).$$

- We can **ask generic queries** (i.e., containing variables)

$? \; connected\_to(W, w_1)$

Datalog: a relational rule language

It expands the syntax of PDCL….

A variable is a symbol starting with an upper case letter

X Y

A constant is a symbol starting with lower-case letter or a sequence of digits.

alan w1

A term is either a variable or a constant.

A predicate symbol is a symbol starting with lower-case letter.

in part-of live

Datalog Syntax (cont')

so it includes propositions

An atom is a symbol of the form p or $p(t_1 \ldots. t_n)$ where p is a predicate symbol and t_i are terms

sunny in(alan, X)

A definite clause is either an atom (a fact) or of the form:

$$h \leftarrow b_1 \wedge \ldots \wedge b_m$$

where h and the b_i are atoms (Read this as ``h if b.'')

in(X, Y) ← in(X, Z) ∧ part_of(Z, Y)

A knowledge base is a set of definite clauses

Datalog: Top Down Proof

Extension of TD for PDCL.

How do you deal with variables?

Example:

KB

in(alan, r123).

part_of(r123,cs_building).

in(X,Y) <- part_of(Z,Y) & in(X,Z).

Query: in(alan, cs_building).

yes <- in(alan, cs_building).

yes <- part-of(Z, csbldg) ∧ in(alan, Z)

Datalog: queries with variables

in(alan, r123).

part_of(r123,cs_building).

in(X,Y) <- in(X,Z). & part_of(Z,Y)

one answer

yes(r123)

yes(x1) ← in(alan,Z) & part-of(Z,X1)

Query: in(alan, X1).

Yes(X1) <- in(alan, X1).

yes(cs_building)

Logics: Big Picture

First Order Logic

$\forall X \exists Y p(X,Y) \Longleftrightarrow \neg q(Y)$

$p(a_1, a_2)$

$-q(a_5)$

Datalog

$p(X) \leftarrow q(X) \land r(X,Y)$

$r(X,Y) \leftarrow s(Y)$

$s(a_1), q(a_2)$

Propositional Logic

$\neg(p \lor q) \rightarrow (r \land s \land f),$

p, r

PDCL

$p \leftarrow s \land f$

$r \leftarrow s \land q \land p$

r

p

Today Oct 4

Finish R&R systems in deterministic environments

- Logics
  - Reasoning with individuals and relations
  - Full Propositional Logics and First-Order Logics

- **Start R&R systems in Stochastic environments**
  - Bayesian Networks Representation

Full Propositional Logics

DEFs.

Literal: an atom or a negation of an atom $p \quad \neg q \quad r$

Clause: is a disjunction of literals $p \vee \neg r \vee q$

Conjunctive Normal Form (CNF): a conjunction of clauses $(p) \wedge (q \vee \neg r) \wedge (\neg q \vee p)$

INFERENCE: $KB \overset{?}{\models} \alpha$ formula

- **Convert all formulas in KB and $\neg\alpha$ in CNF**
- **Apply Resolution Procedure (at each step combine two clauses containing complementary literals into a new one)** $p \vee q \quad r \vee \neg q \rightarrow p \vee r$
- **Termination**
  - **No new clause can be added** $KB \not\vdash \alpha$
  - **Two clause resolve into an empty clause** $KB \vdash \alpha$

Propositional Logics: Satisfiability (SAT problem)

Does a set of formulas have a model? Is there an interpretation in which all the formulas are true?

(Stochastic) Local Search Algorithms can be used for this task!

Evaluation Function: number of unsatisfied clauses

WalkSat: One of the simplest and most effective algorithms:

Start from a randomly generated interpretation

- Pick an unsatisfied clause
- Pick a proposition to flip (randomly 1 or 2)
  1. To minimize # of unsatisfied clauses
  2. Randomly

Full First-Order Logics (FOLs)

We have constant symbols, predicate symbols and function symbols

So interpretations are much more complex (but the same basic idea – one possible configuration of the world)

constant symbols => individuals, entities

predicate symbols => relations

function symbols => functions

INFERENCE:

- Semidecidable: algorithms exists that says yes for every entailed formulas, but no algorithm exists that also says no for every non-entailed sentence
- Resolution Procedure can be generalized to FOL

Today Oct 4

Finish R&R systems in deterministic environments

- Logics
  - Reasoning with individuals and relations
  - Full Propositional Logics and First-Order Logics

- Start R&R systems in Stochastic environments
  - **Bayesian Networks Representation**

R&Rsys we'll cover in this course

Environment

Problem		Deterministic	Stochastic
Static	Constraint Satisfaction	*Vars + Constraints* Arc Consistency SLS Search	
	Query	*Logics* → Propositional → First Order → . . . . Search	*Belief Nets* Var. Elimination Approx. Inference Temporal. Inference
Sequential	Planning	*STRIPS* actions preCs effects Search	aka influence diagrams *Decision Nets* Var. Elimination *Markov Processes* Value Iteration

Representation

Reasoning Technique

Intro to Probability (Motivation)

- *Will it rain in 10 days? Was it raining 98 days ago?*
- *Right now, how many people are in this building (DMP)? At UBC? ….Yesterday?*

- AI agents (and humans ☹) are not omniscient (know everything)

  they are ignorant

- And the problem is not only predicting the future or "remembering" the past

  also current state

Intro to Probability (Key points)

- Are agents all ignorant/uncertain to the same degree? No
  it is subjective
- Should an agent act only when it is certain about relevant knowledge?
- (not acting usually has implications)
- So agents need to **represent and reason about their ignorance/ uncertainty**

Probability as a formal measure of uncertainty/ignorance

- Belief in a proposition *f* (e.g., *it is snowing outside, there are 321 people in this bldg*) can be measured in terms of a number between 0 and 1 – this is the probability of *f*
  - The probability *f* is 0 means that *f* is believed to be definitely false
  - The probability *f* is 1 means that *f* is believed to be definitely true
  - Using 0 and 1 is purely a convention.

Random Variables

- A random variable is a **variable** like the ones we have seen in **CSP** and **Planning**, but the agent can be **uncertain about its value**.
- As usual
  - The domain of a random variable *X*, written *dom(X),* is the set of values *X* can take
  - values are mutually exclusive and exhaustive

Examples (Boolean and discrete)

outside Raining

T F

#-of-people-in-DMP

$[0 - 10^4]$

Random Variables (cont')

- A tuple of random variables $<X_1,\ldots.,X_n>$ is a complex random variable with domain..

$dom(X_1) \times \ldots \times dom(X_n)$

- Assignment $X=x$ means X has value x

outsideRaining = T

- A proposition is a Boolean formula made from assignments of values to variables

Examples

$\vee$ OR

outsideRaining = T $\wedge$ #people-rm = 47

AND

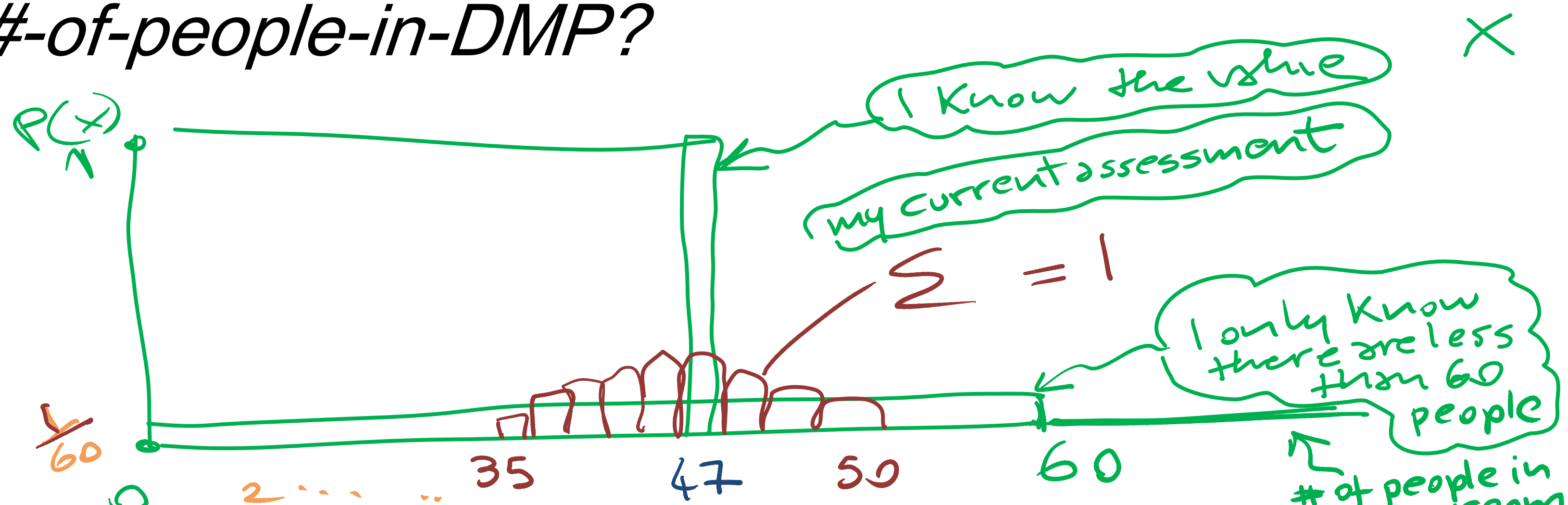
Probability Distributions

- A probability distribution **P** on a random variable **X** is a function *dom(X) - > [0,1]* such that *x -> P(X=x)*

dom(cavity) = [T, F]

cavity?
- T → .2 P(cavity = T)
- F → .8 P(cavity = F)

#-of-people-in-DMP?



Joint Probability Distributions P(< $X_1, \ldots, X_n$ >)

- Probability distribution over the variable Cartesian product of multiple random variables
  - Think of a joint distribution over n variables as an n-dimensional table
  - Each entry, indexed by $X_1 = x_1, \ldots, X_n = x_n$ corresponds to $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$
  - The sum of entries across the whole table is 1

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

for n Boolean vars

24

entries

cavity	*toothache*	*catch*	$\mu(w)$
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

Joint Prob. Distribution (JPD): Example2

3 binary random variables: **P(H,S,F)**

H=true H=false

- **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

$\Sigma 1$

$2^3 - 1$ $2^K - 1$

Marginalization

$P(H,S,F)$

	f		¬ f	
	s	¬ s	s	¬ 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	P(H)
h	.02	.01	.03
¬ h	.28	.69	.97
P(S)	.3	.7	

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

P(S|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.)

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

$P(S \mid H, F)$

$$P(\underbrace{X_1 \ldots, X_n}\mid Y_1 \ldots Y_k)$$

binary

- 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

Chain Rule

$$\underline{P(H,S,F)} = P(H) * P(S|H) * P(F|H,S)$$

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

Bayes Theorem

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

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

substitute

rewrite

$$P(H|S)\,P(S) = \underline{P(S,H)}$$

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

Do you always need to revise your beliefs?

NO 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 $x_i \in dom(X)$, $y_k \in dom(Y)$,

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

Marginal Independence: Example

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

P(*X*|*Y*) = **P**(*X*) or **P**(*Y*|*X*) = **P**(*Y*) or **P**(X, Y) = **P**(*X*) **P**(*Y*)

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

4 possible values

Sunny
Cloudy
rainy
snowy

Ex: **P**(*Toothache, Catch, Cavity, Weather*)
= **P**(*Toothache, Catch, Cavity*) P(weather)

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

Joint prob. distribution

Conditional Independence

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

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

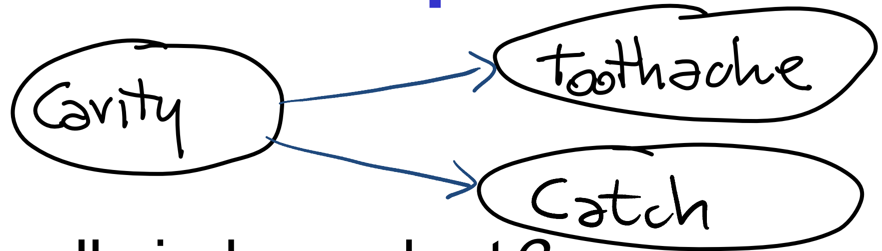
Absolute independence is powerful **but** when you model a **particular domain**, it is **rare**

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?

$P(\ \mid\ ) = P(\text{Toothache})$?

BUT If I have a cavity, does the probability that the probe catches depend on whether I have a toothache?

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

What if I haven't got a cavity?

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

STOP HERE 04/10/11

- *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 $\boxed{P(X|YZ) = P(X|Z)} \Rightarrow$

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

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

$$\text{③}\ P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)} \xRightarrow{\text{from } A} \frac{P(Y,Z)\,P(X,Z)}{P(Z)} \cdot \frac{1}{P(Z)}$$

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

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:

$\mathbf{P}(Cavity, Catch, Toothache)$

$= \mathbf{P}(Toothache \mid Catch, Cavity)\, \mathbf{P}(Catch \mid Cavity)\, \mathbf{P}(Cavity)$

$= \underbrace{\mathbf{P}(Toothache \mid \text{Cavity})}_{2}\, \underbrace{\mathbf{P}(Catch \mid Cavity)}_{2}\, \underbrace{\mathbf{P}(\text{Cavity})}_{1}$

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*. **n is the number of vars**

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

Key points Recap

- We model the environment as a set of random vars

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

- Why the joint is not an adequate representation ?

"Representation, reasoning and learning" are "exponential" in the number of variables

Solution: Exploit marginal&**conditional** independence

$P(X|Y) = P(X)$ $P(X|YZ) = P(X|Z)$

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

Today Oct 4

Finish R&R systems in deterministic environments

- Logics
  - Reasoning with individuals and relations
  - Full Propositional Logics and First-Order Logics

- **Start R&R systems in Stochastic environments**
  - Bayesian Networks Representation

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.

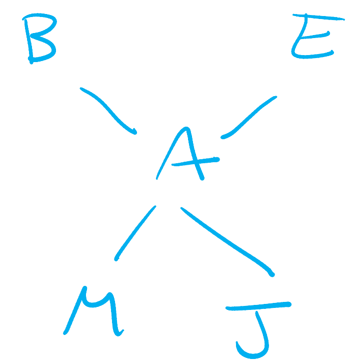
E

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.

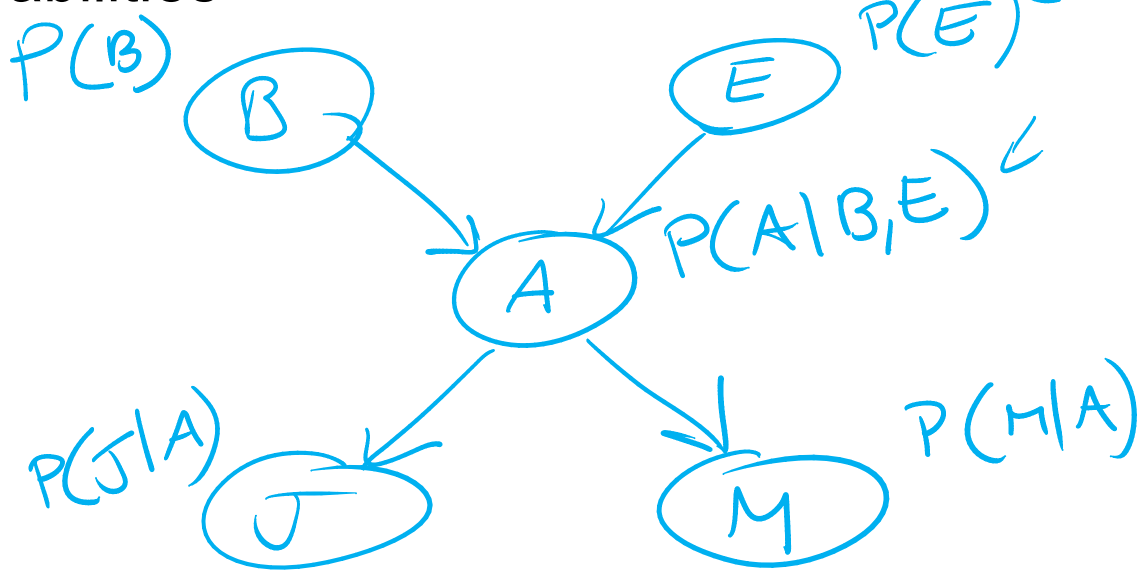
$$P(B)\,P(E|\cancel{B})\,P(A|B,E)\,P(M|A,\cancel{E},\cancel{B})\,P(J|\cancel{M},A,\cancel{E},\cancel{B})$$

- 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)

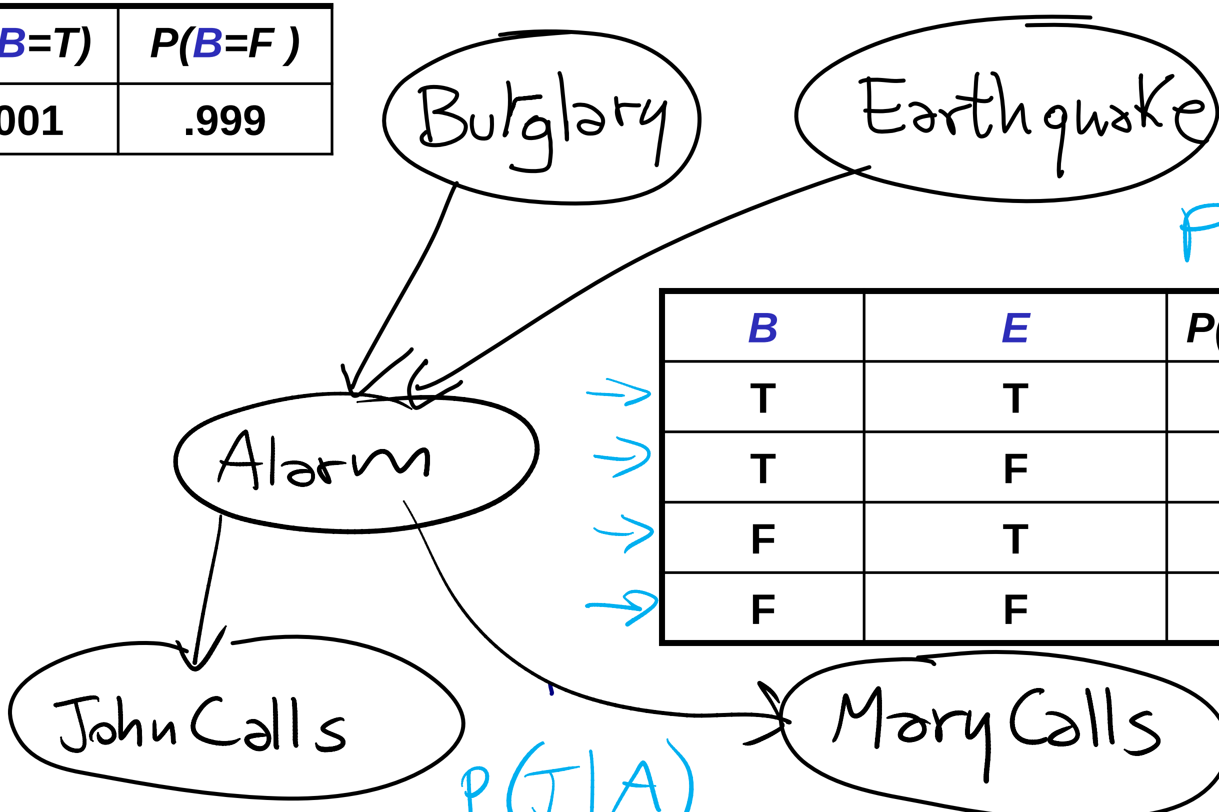
Burglary: complete BN

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

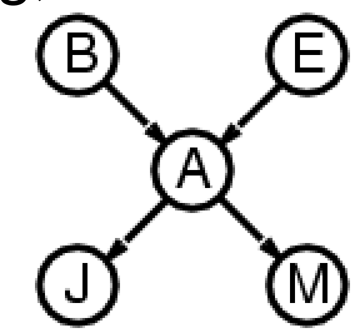
call for any other reasons

Burglary Example: Bnets inference

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

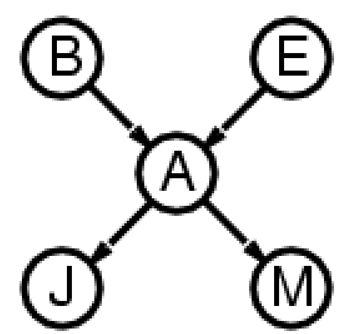
(Ex1) I'm at work,

- neighbor John calls to say my alarm is ringing,
- neighbor Mary doesn't call.
- No news of any earthquakes.
- Is there a burglar?

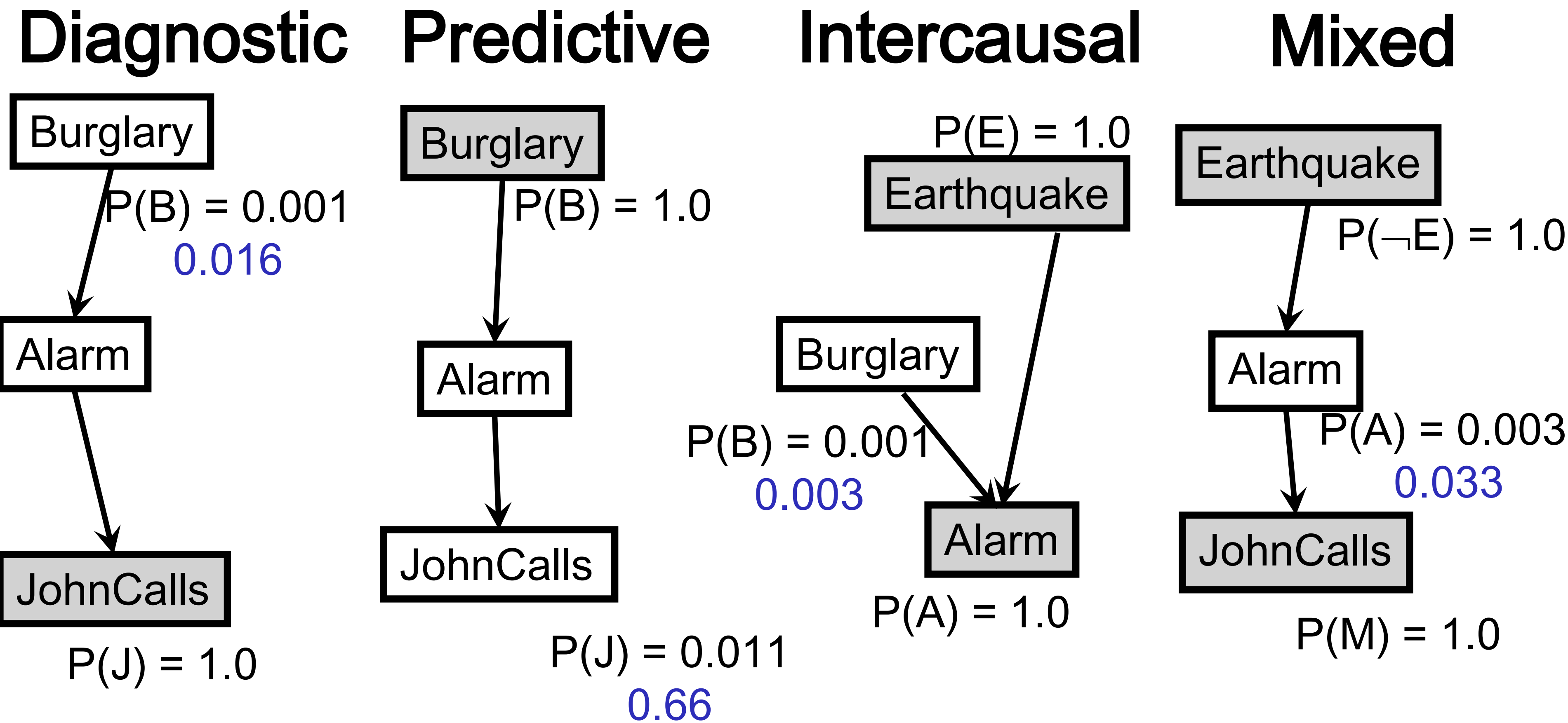


(Ex2) I'm at work,

- Receive message that neighbor John called ,
- News of minor earthquakes.
- Is there a burglar?



Bayesian Networks – Inference Types



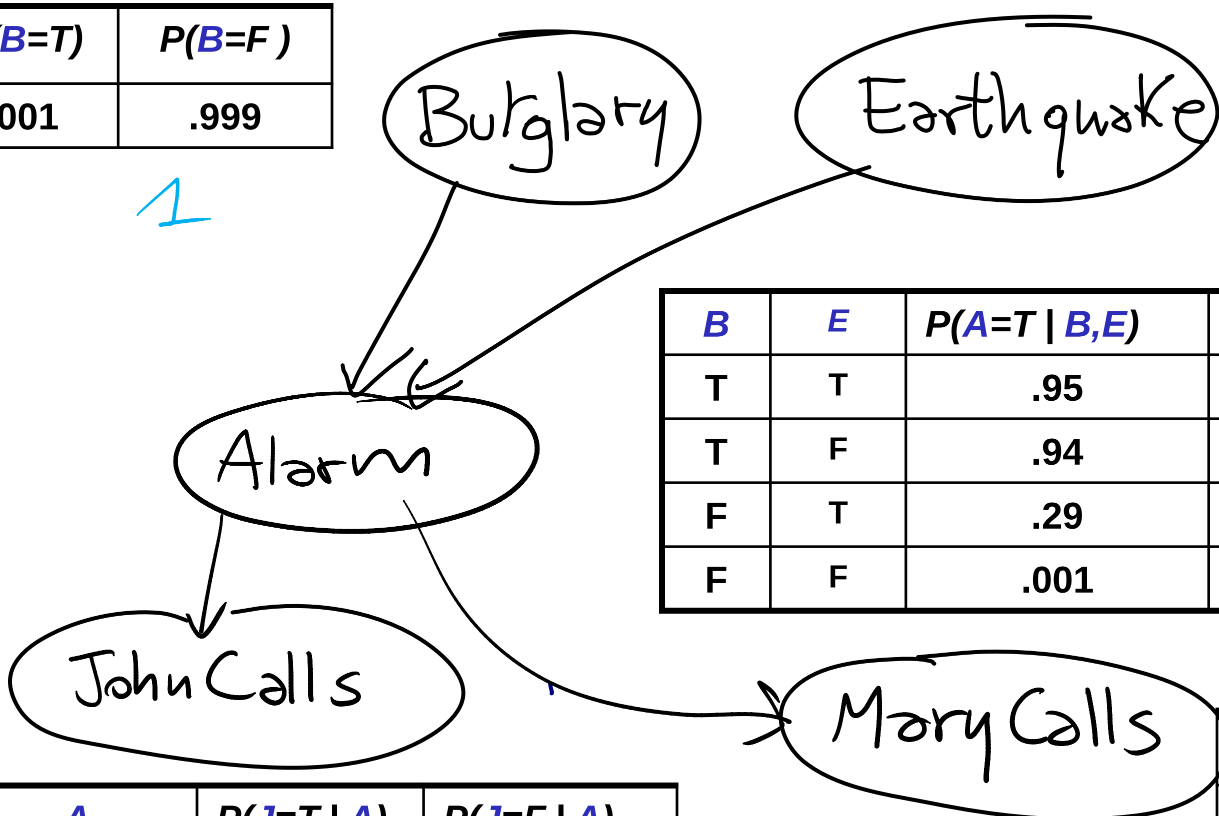
BNnets: Compactness

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 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 $1-p_i$)

If each on the n 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

Realistic BNet: Liver Diagnosis

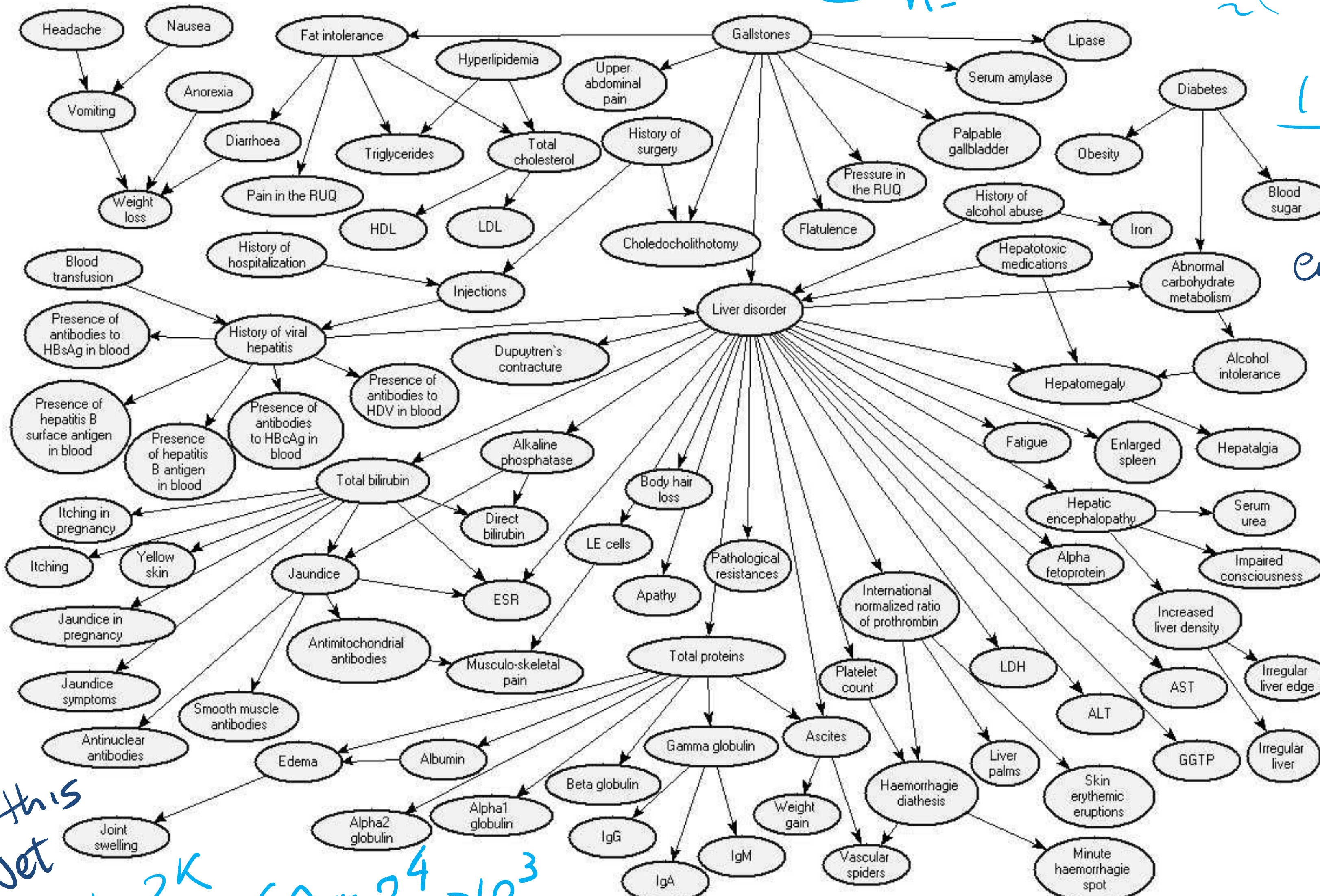
Source: Onisko et al., 1999

~60 nodes

JPD

$n = 60$ $\sim 2^{60}$ $= (2^{10})^6$ $\approx 10^{18}$

10^{18} entries



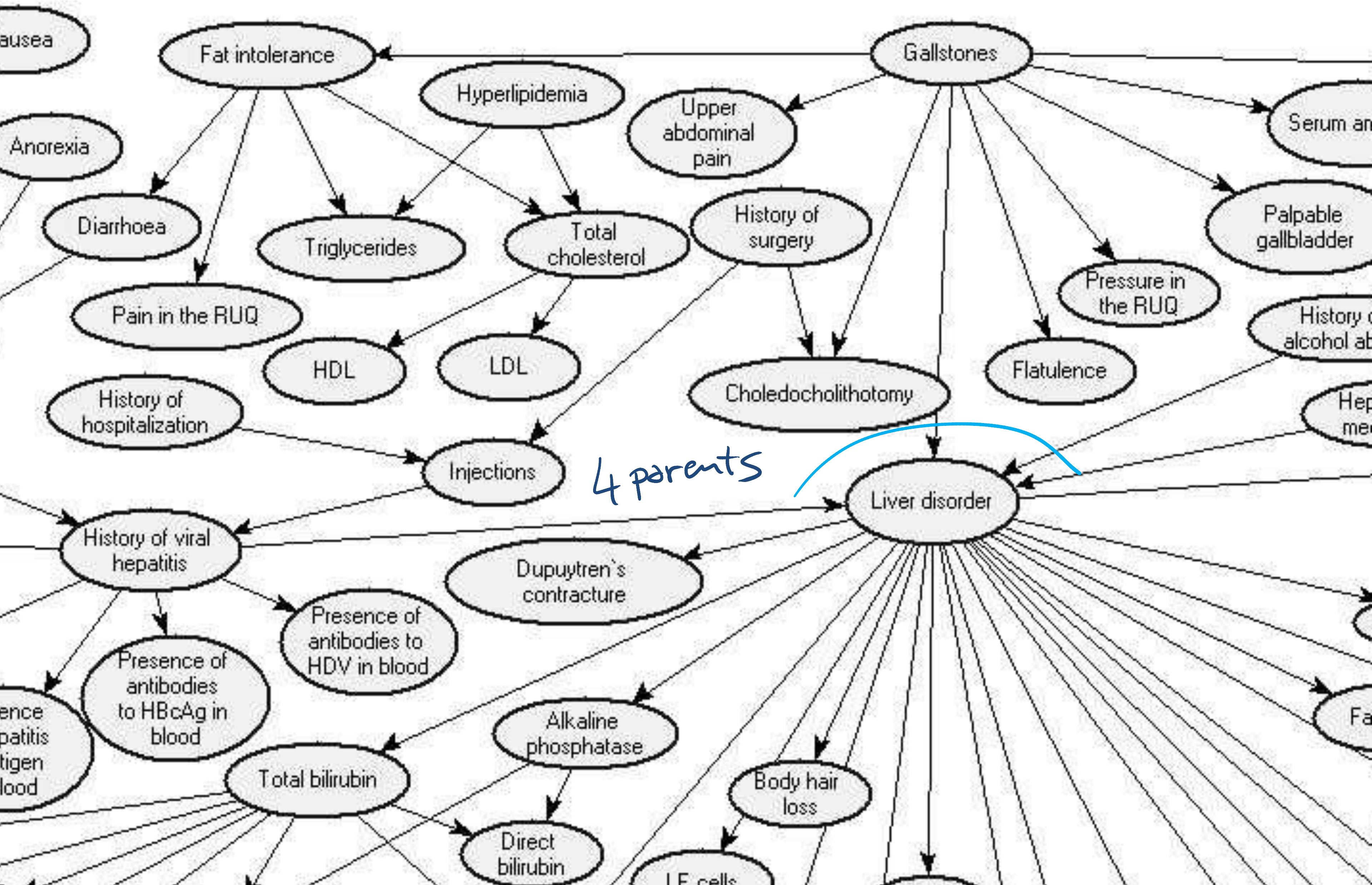
for this BNet

$n\,2^K$

$60 * 2^4 = 10^3$

Realistic BNet: Liver Diagnosis

Source: Onisko et al., 1999



TODO for this Thur

Read Chp 6 of textbook up to Rejection Sampling included

Also Do exercises 6.A and 6.B
http://www.aispace.org/exercises.shtml

BNets: Construction General Semantics

The full joint distribution can be defined as the product of conditional distributions:

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

Simplify according to marginal&conditional independence

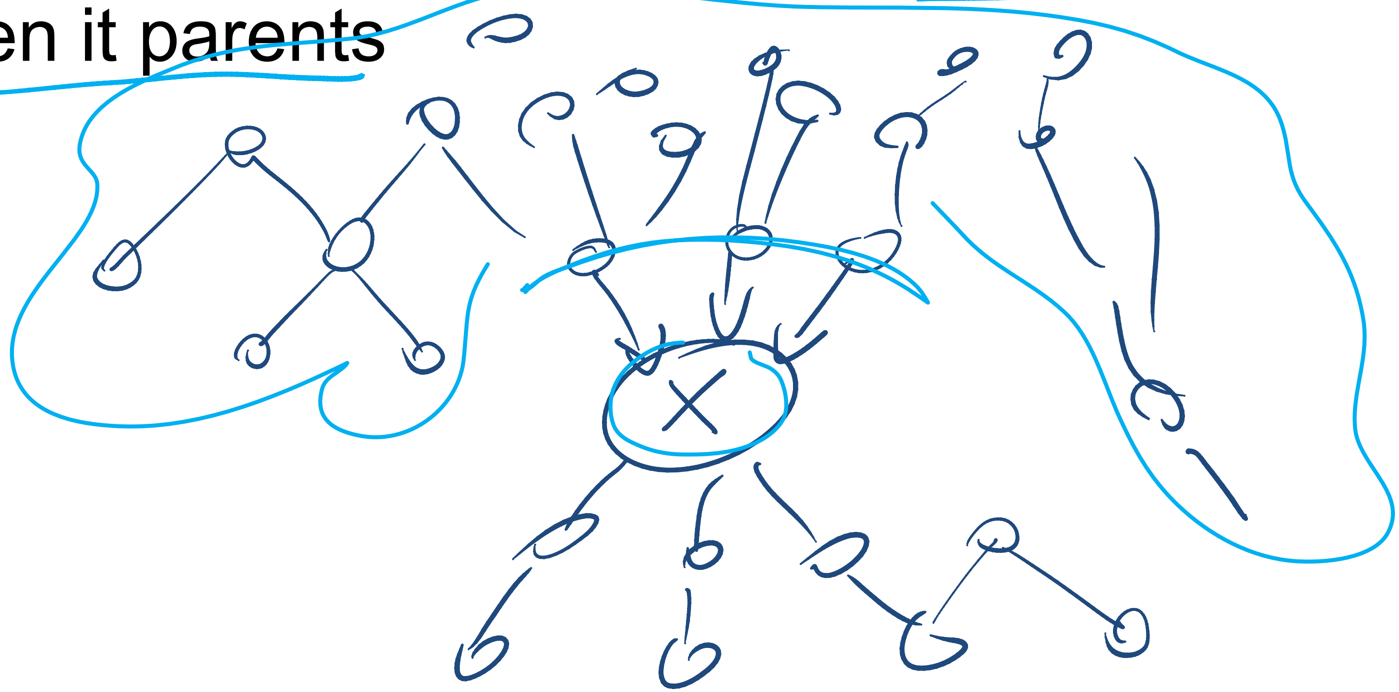
- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities

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

BNets: Construction General Semantics (cont')

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

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



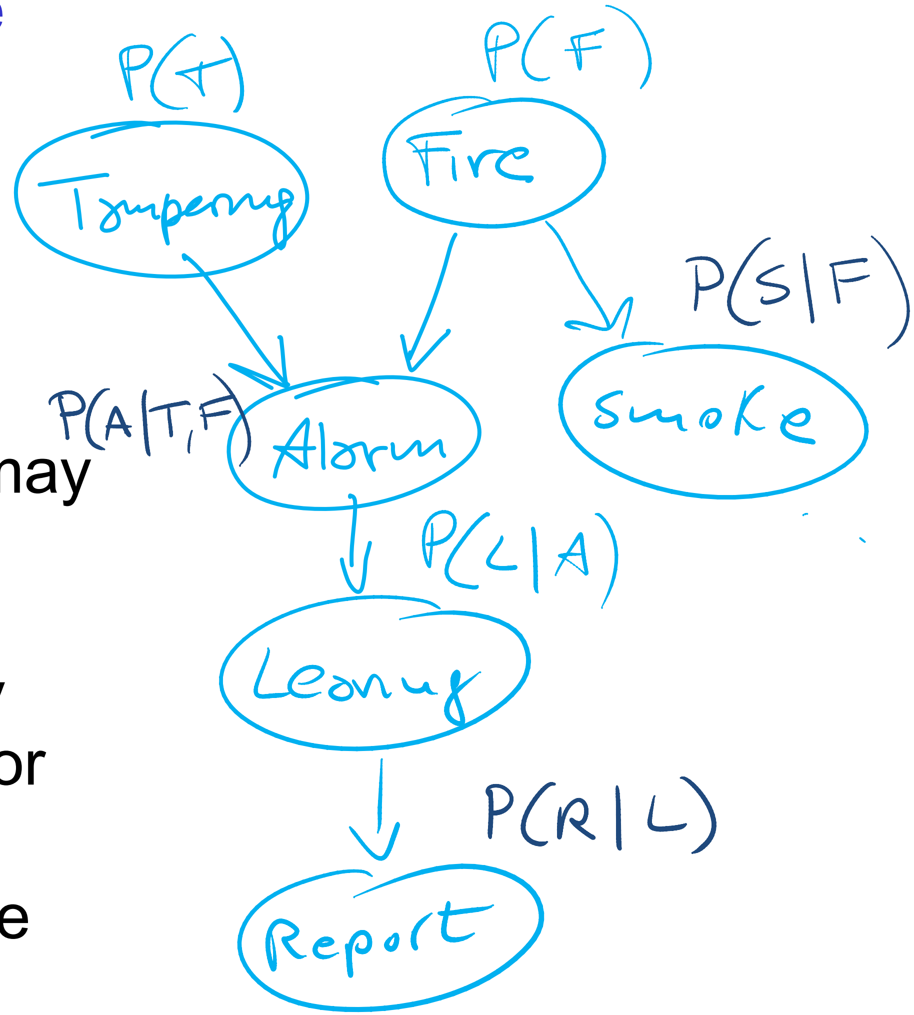
Lecture Overview

- Belief Networks
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - More Examples

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.



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 AIspace