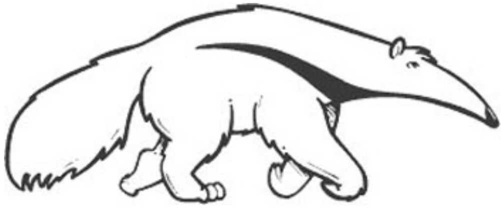
CS 276: Causal and Probabilistic Reasoning

Rina Dechter, UCI

Lecture 1: Introduction



Class Information

Course Topics
Probabilistic Graphical Models, Structural causal models,The Causal Hierarchy.

1.Representing independencies by graphs. d-seperation.
2.Algorithms (Bucket-elimination, Join-trees, The induced-width.).
3.Sampling schemes for graphical models (MCMC, IS)
4. AND/OR search
5.Structural Causal Models; Identification of Causal Effect;
6.The Back-Door and Front-Door Criteria and the Do-Calculus.
7.Linear Causal Models.
8.Counterfactuals.
9.Algorithms for identification. The ID algorithm.
10.Learning Bayesian networks and Causal graphs (causal discovery).

Class page: https://ics.uci.edu/~dechter/courses/ics-276/fall_2024/

Grading

- **Four or five homeworks (the highest 4 will count)**
- Project: Class presentation and a report: Students will present a paper and write a report

Textbooks

[P] Judea Pearl, Madelyn Glymour, Nicholas P. Jewell,
Causal Inference in Statistics: A Primer,
Cambridge Press, 2016.
[C] Judea Pearl,
Causality: Models, Reasoning, and Inference,
Cambridge Press, 2009.
[W] Judea Pearl, Dana Mackenzie,
The Book of Why,
Basic books, 2018.

- [Darwiche] Adnan Darwiche, "Modeling and Reasonin with Bayesian Networks“
- [Dechter] Rina Dechter, "Reasoning with Probabilistic and Deterministic Graphical Models: Exact Algorithms"

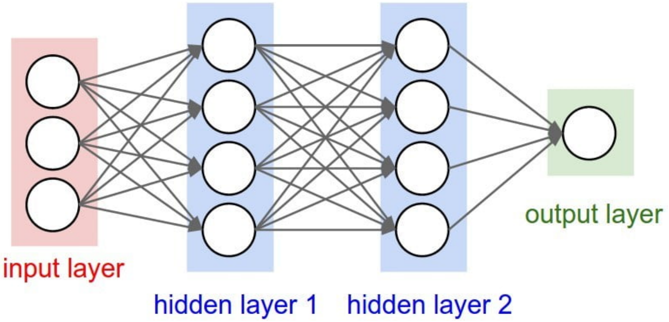
Outline

- Graphical Models
- Inference Tasks
- Basic probabilisty

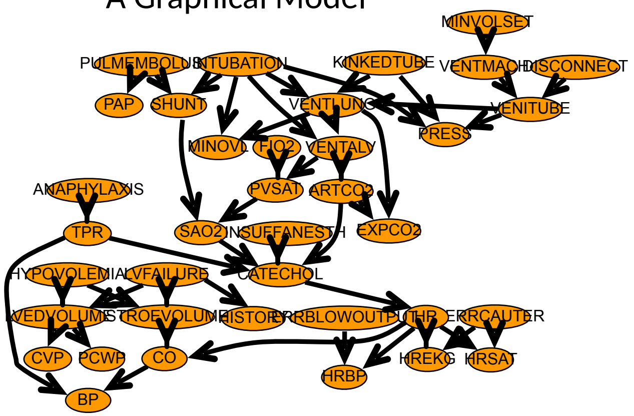
The Primary AI Challenges

- **Machine Learning** focuses on replicating humans learning
- **Automated reasoning** focuses on replicating how people reason.
- Large Language Models (LLMs)

A neural network



A Graphical Model



Automated Reasoning

Medical Doctor

Lawyer

Policy Maker

Queries:

- Prediction: what will happen?
- Diagnosis: what had happened?
- Situation assessment: What is going on?
- Planning, decision making: what to do?
- Explanation: need causal models
- Counterfactuals: What if? need Structural causal models

Same with any common-sense agent

Automated Reasoning

Causal Models **Data**

answers

Queries:

- Prediction
- Diagnosis
- Situation assessment
- Planning, decision making
- Explanation, causal effect
- Counterfactuals

Knowledge is huge, so How to identify what's relevant? → **Causal Graphical Models**

* ***The field of Automated Reasoning** developing general purpose formalisms (languages, models) that enable us to represent knowledge in such a way that we can exploit the relevance and causal relationship quickly.

Answer query in the 3 levels of the causal hierarchy

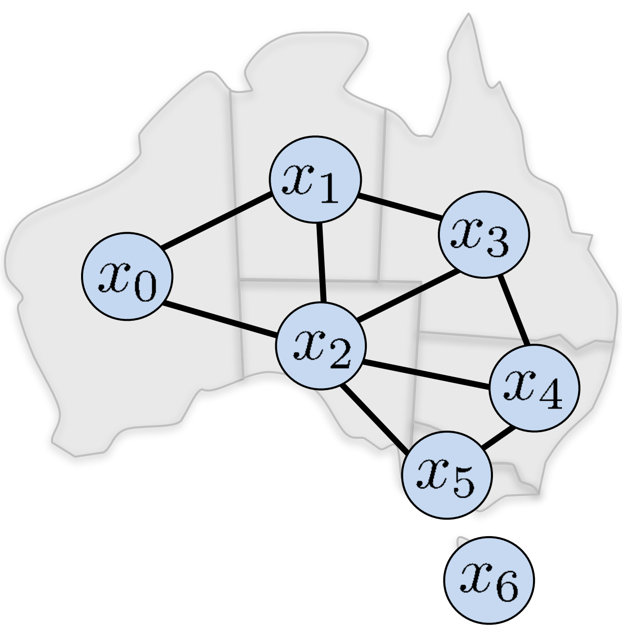
Graphical Models

Describe structure and interdependence in a model of the world

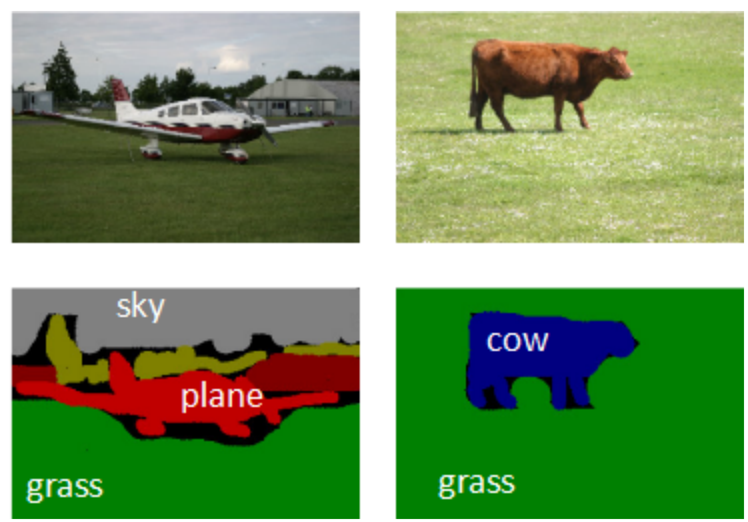
Examples:

- Markov Random Fields: correlations

Map coloring & constraint satisfaction problems



Semantic segmentation: fine-grain object recognition



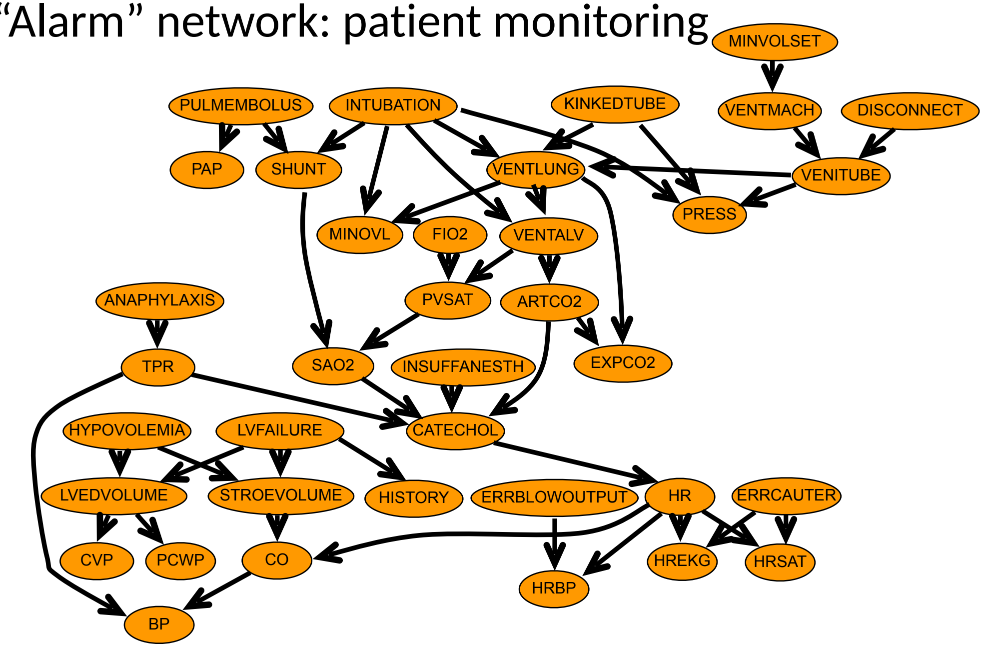
Graphical Models

Describe structure and interdependence in a model of the world

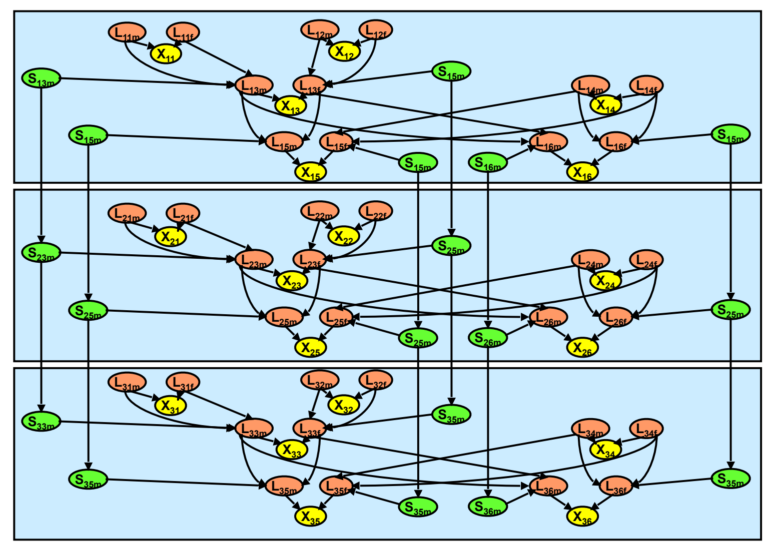
Examples:

- Markov Random Fields: correlations
- Bayesian Networks: conditional dependence

“Alarm” network: patient monitoring



Pedigree network: genetic inheritance



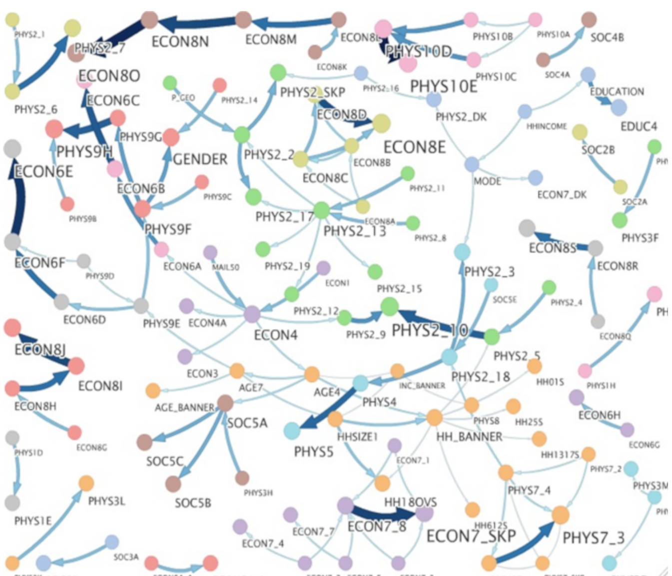
Graphical Models

Describe structure and interdependence in a model of the world

Examples:

- Markov Random Fields: correlations
- Bayesian Networks: conditional dependence
- Causal Networks: effect of intervention – what would happen if?

Impact of COVID & assistance
on mental health (survey)



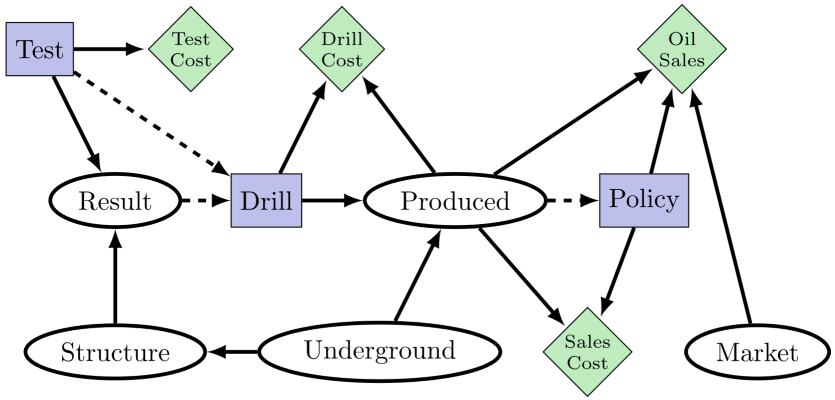
Graphical Models

Describe structure and interdependence in a model of the world

Examples:

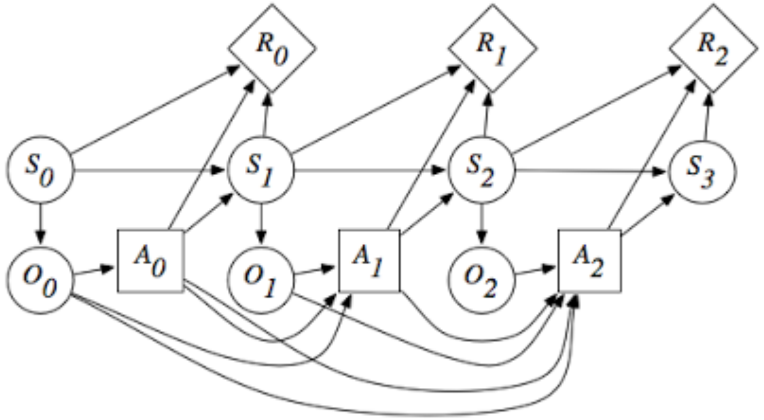
- Markov Random Fields: correlations
- Bayesian Networks: conditional dependence
- Causal Networks: effect of intervention – what would happen if?
- Influence Diagrams: actions and rewards – what should we do if?

"Oil Wildcatter" Decision Network

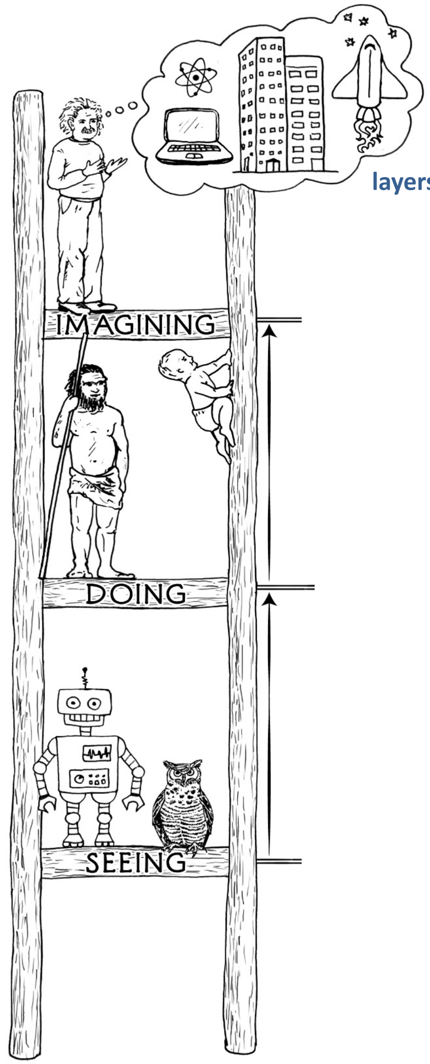


(Partially Observable) Markov Decision Process

(Planning, Reinforcement Learning)



Why Causality?



layers/rungs of the causal hierarchy

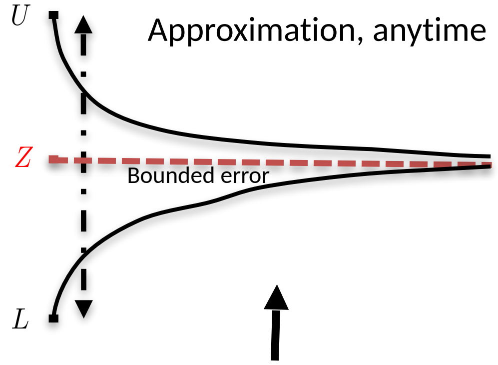
This course
last part (20%)

This course
main part (40%)

This course,
first part (40%)

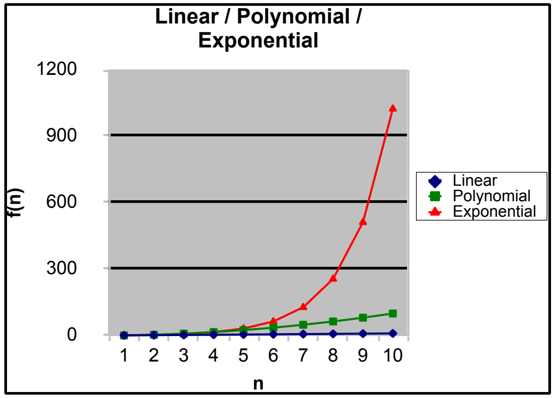
Complexity of Automated Reasoning

- Prediction
- Diagnosis
- Planning and scheduling
- Probabilistic Inference
- Explanation
- Decision-making
- Causal reasoning



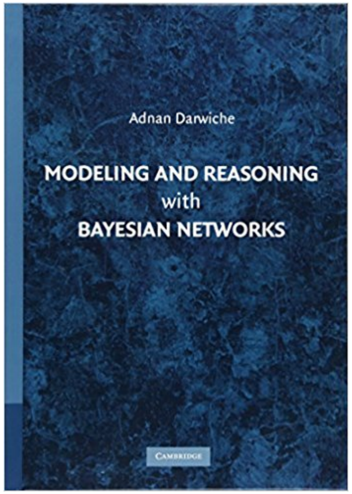
Reasoning is computationally hard

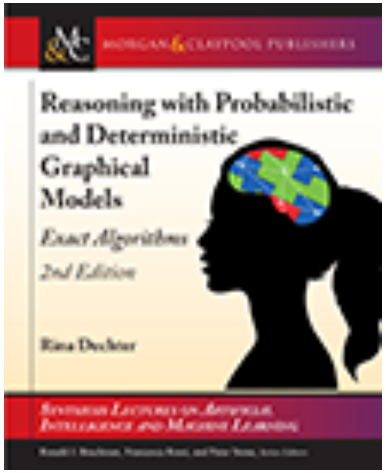
Complexity is exponential



Reasoning models is hard. Reasoning with functions is easy. So?

Books on Graphical Models & Causality

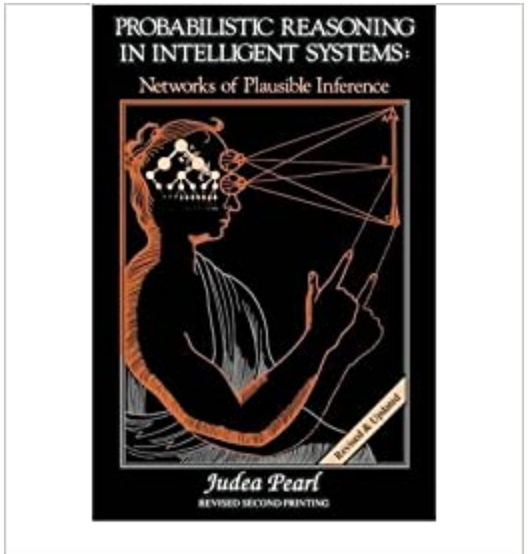


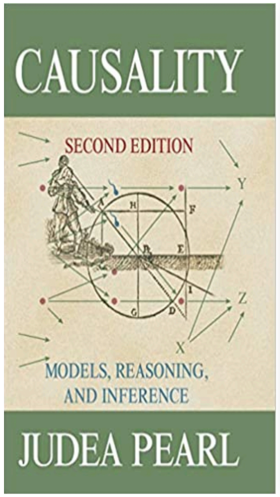


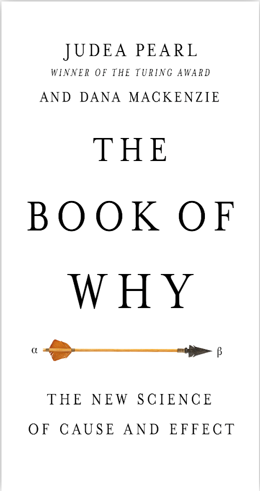
[Class page]

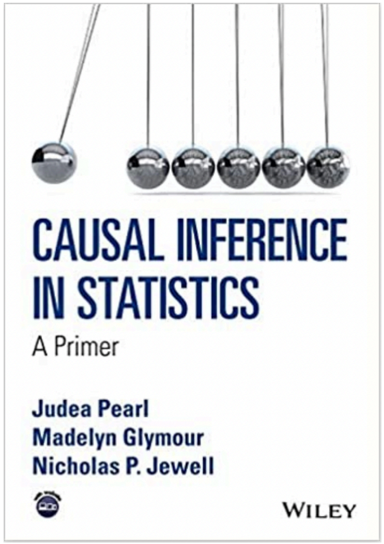
2009

2018









intervention & counterfactuals

Why graphical models?

Combine domain knowledge with learning and data

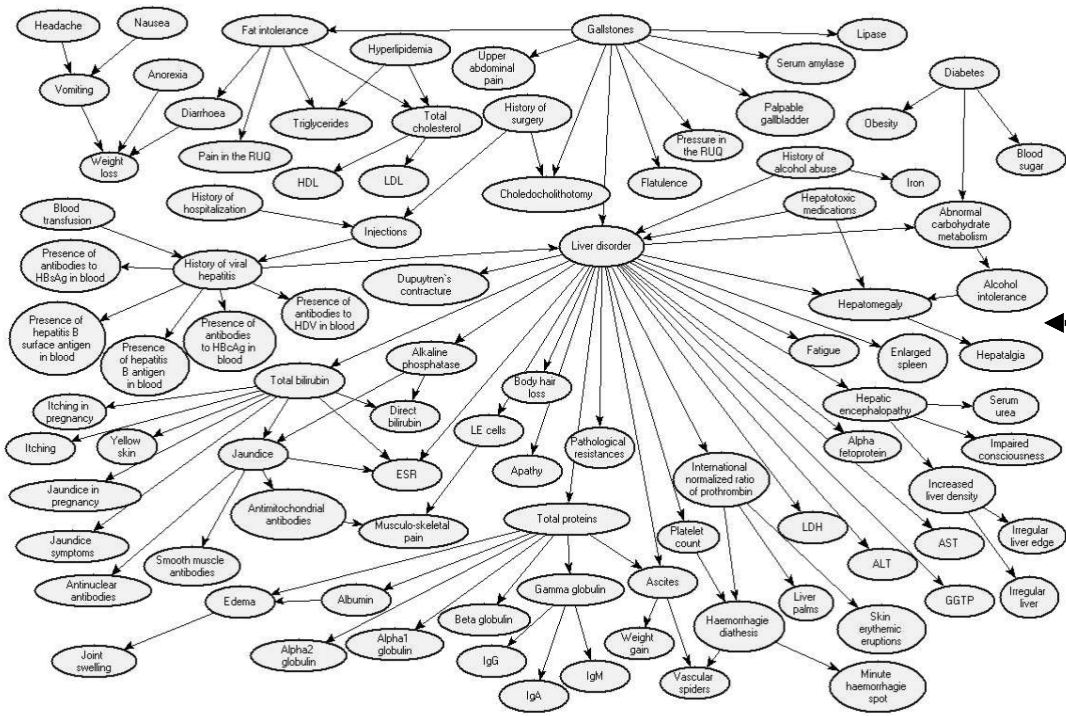
- Domain knowledge
  - Problem structure: potential causation or interactions
  - Model parameters: known dependency mechanisms, probabilities

- Learning and data
  - Identify (in)dependence from data
  - Estimate model parameters to explain observations

- Scalable and Composable
  - Models over large systems may be composed of smaller parts
  - Efficient representation allows learning from relatively few data

Graphical Models

Example: diagnosing liver disease (Onisko et al., 1999)



Queries:

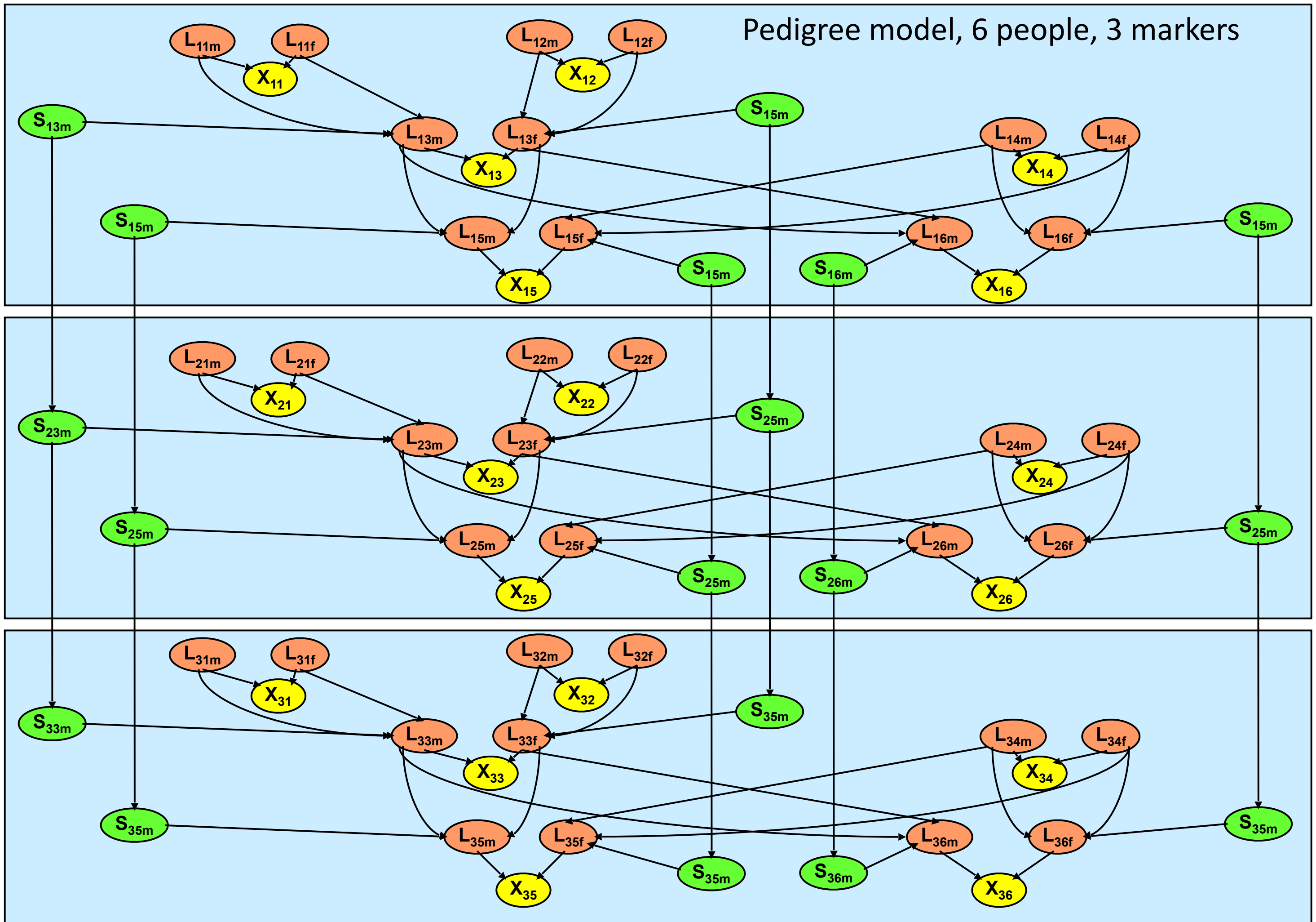
- Prediction
- Diagnosis, explanation
- Situation assessment
- Planning, decision making
- Counterfactual reasoning

Automated Reasoning:

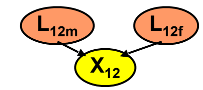
- Develop methods to answer these questions.
- Learning the models: from experts and data.

Ex: Model composability

Large models may be defined by many repeated, interrelated structures

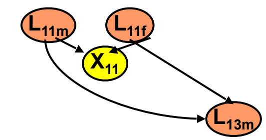


Individual's genotype

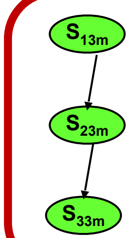


determines phenotype

Each parent



passes one gene copy to their child



Gene position correlates which copy is inherited over several genes

Example domains for graphical models

- Natural Language processing
  - Information extraction, semantic parsing, translation, topic models, …
- Computer vision
  - Object recognition, scene analysis, segmentation, tracking, …
- Computational biology
  - Pedigree analysis, protein folding and binding, sequence matching, …
- Networks
  - Webpage link analysis, social networks, communications, citations, ….
- Robotics
  - Planning & decision making
- Social sciences, man-machine interaction requires causality

In more details…

Bayesian networks

Use **independence** and **conditional independence** to simplify a **joint probability**

- Joint probability, p(X=x,Y=y,Z=z)
  - The probability that event (x,y,z) happens.

- Conditional probability
  - The chain rule of probability tells us

    p(X=x,Y=y,Z=z) = p(X=x) p(Y=y | X=x) p(Z=z | X=x,Y=y)

    (x,y,z all happen) (x happens) (y happens given x happened) (z happens given x,y happened)

  - Can use any order, e.g. (Z,X,Y):

    p(X=x,Y=y,Z=z) = p(Z=z) p(X=x | Z=z) p(Y=y | X=x,Z=z)

Independence

- X, Y independent:
  - p(X|Y) = p(X) or p(Y|X) = p(Y) (if p(Y), p(X) > 0)
  - Intuition: knowing X has no information about Y (or vice versa)
  - Leads to: p(X=x,Y=y) = p(X=x) p(Y=y) for all x,y
  - Shorthand: p(X,Y) = P(X) P(Y)

Independent probability distributions:

A	P(A)
0	0.4
1	0.6

B	P(B)
0	0.7
1	0.3

C	P(C)
0	0.1
1	0.9

This reduces representation size!

Joint:

A	B	C	P(A,B,C)
0	0	0	.4 * .7 * .1
0	0	1	.4 * .7 * .9
0	1	0	.4 * .3 * .1
0	1	1	…
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Independence

- X, Y independent:
  - p(X|Y) = p(X) or p(Y|X) = p(Y) (if p(Y), p(X) > 0)
  - Intuition: knowing X has no information about Y (or vice versa)
  - p(X=x,Y=y) = p(X=x) p(Y=y) for all x,y
  - Shorthand: p(X,Y) = P(X) P(Y)

Independent probability distributions:

A	P(A)
0	0.4
1	0.6

B	P(B)
0	0.7
1	0.3

C	P(C)
0	0.1
1	0.9

Joint:

A	B	C	P(A,B,C)
0	0	0	0.028
0	0	1	0.252
0	1	0	0.012
0	1	1	0.108
1	0	0	0.042
1	0	1	0.378
1	1	0	0.018
1	1	1	0.162

This reduces representation size!

Note: it is hard to "read" independence from the joint distribution. We can "test" for it, however.

Conditional Independence

- X, Y independent given Z
  - $p(X=x,Y=y|Z=z) = p(X=x|Z=z)\ p(Y=y|Z=z)$ for all x,y,z
  - Equivalent: $p(X|Y,Z) = p(X|Z)$ or $p(Y|X,Z) = p(Y|Z)$ (if all > 0)
  - Intuition: X has no additional info about Y beyond Z's

- Example

X = height — p(height|reading, age) = p(height|age)

Y = reading ability — p(reading|height, age) = p(reading|age)

Z = age

Height and reading ability are dependent (not independent), but are conditionally independent given age

Conditional Independence

- X, Y independent given Z
  - p(X=x,Y=y|Z=z) = p(X=x|Z=z) p(Y=y|Z=z)   for all x,y,z
  - Equivalent: p(X|Y,Z) = p(X|Z)   or p(Y|X,Z) = p(Y|Z)
  - Intuition: X has no additional info about Y beyond Z's

- Example: Dentist

$$(T \perp\!\!\!\perp D \,|\, C)?$$

Is T conditionally independent of C given D?

Again, hard to "read" from the joint probabilities; only from the conditional probabilities.

Like independence, reduces representation size!

People knows dependence information
But not the actual numbers.

Joint prob:

T	D	C	P(T,D,C)
0	0	0	0.576
0	0	1	0.008
0	1	0	0.144
0	1	1	0.072
1	0	0	0.064
1	0	1	0.012
1	1	0	0.016
1	1	1	0.108

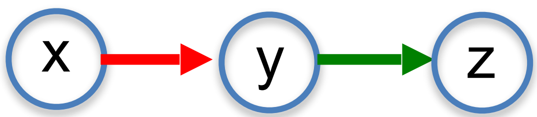
Conditional prob:

T	D	C	P(T|D,C)
0	0	0	0.90
0	0	1	0.40
0	1	0	0.90
0	1	1	0.40
1	0	0	0.10
1	0	1	0.60
1	1	0	0.10
1	1	1	0.60

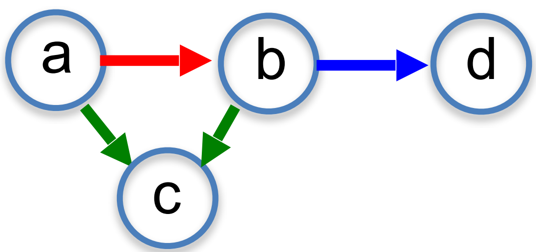
Bayesian networks

- Directed graphical model
- Nodes associated with variables
- "Draw" independence in conditional probability expansion
  - Parents in graph are the RHS of conditional

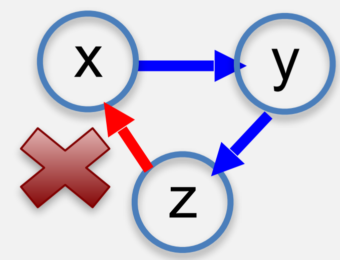
- Ex: $p(x, y, z) = p(x)\, p(y \mid x)\, p(z \mid y)$



- Ex: $p(a, b, c, d) = p(a)\, p(b \mid a)\, p(c \mid a, b)\, p(d \mid b)$



Graph must be ***acyclic***



Corresponds to an order over the variables (chain rule)

Example adapted from Pearl

Example

- Consider the following 5 binary variables:
  - B = a burglary occurs at your house
  - E = an earthquake occurs at your house
  - A = the alarm goes off
  - W = Watson calls to report the alarm
  - H = Mrs. Hudson calls to report the alarm

  - What is P(B | H=1, W=1) ? (for example)

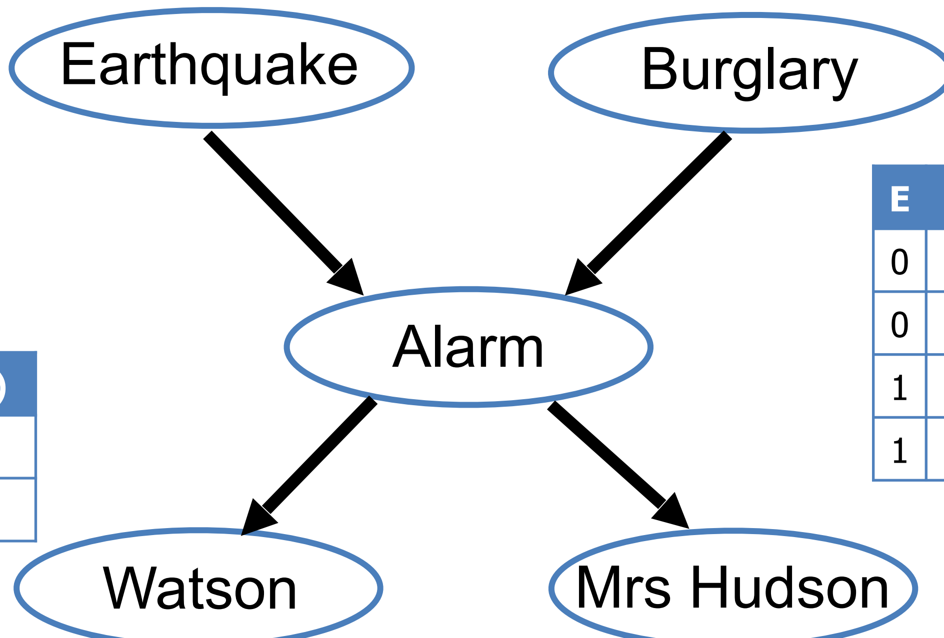
  - We can use the full joint distribution to answer this question
    - Requires $2^5 = 32$ probabilities

    - Can we use prior domain knowledge to come up with a Bayesian network that requires fewer probabilities?

Constructing a Bayesian network

- Given $p(W, H, A, E, B) = p(E)\ p(B)\ p(A|E, B)\ p(W|A)\ p(H|A)$
- Define probabilities: 1 + 1 + 4 + 2 + 2
- Where do these come from?
  - Expert knowledge; estimate from data; some combination

P(E)
0.002



P(B)
0.001

E	B	P(A|E,B)
0	0	0.001
0	1	0.29
1	0	0.94
1	1	0.95

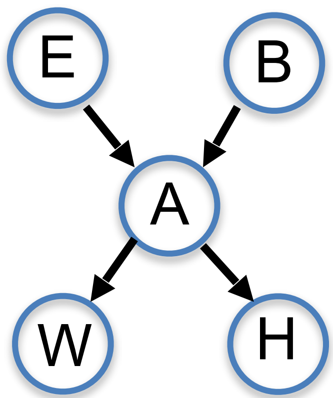
"CPT" = conditional probablility table

A	P(W|A)
0	0.05
1	0.90

A	P(H|A)
0	0.01
1	0.70

Constructing a Bayesian network

- Joint distribution



Full joint distribution:
$2^5 = 32$ probabilities

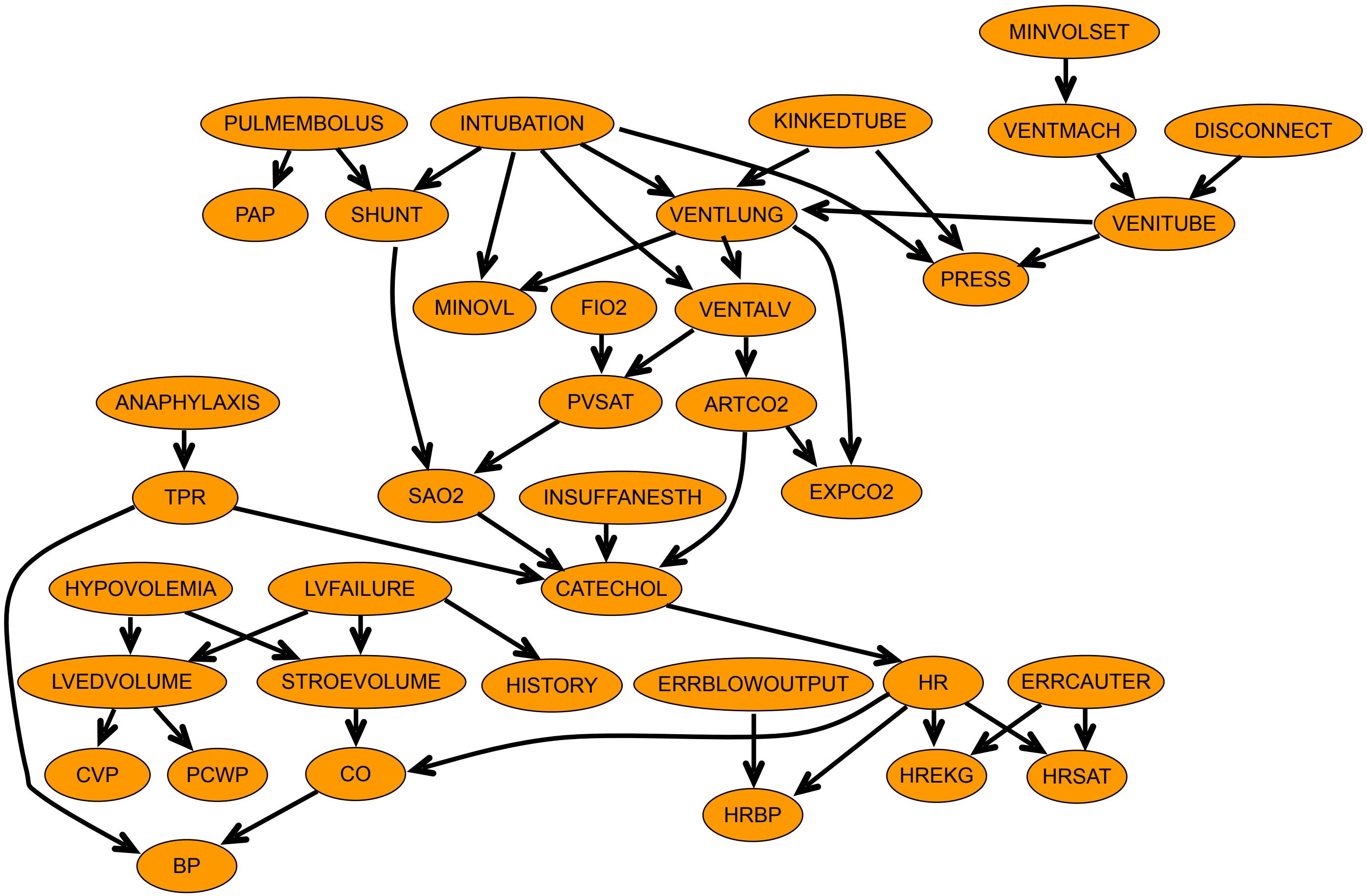
Structured distribution:
specify 10 parameters

E	B	A	W	H	P(…)
0	0	0	0	0	.93674
0	0	0	0	1	.00133
0	0	0	1	0	.00005
0	0	0	1	1	.00000
0	0	1	0	0	.00003
0	0	1	0	1	.00002
0	0	1	1	0	.00003
0	0	1	1	1	.00000
0	1	0	0	0	.04930
0	1	0	0	1	.00007
0	1	0	1	0	.00000
0	1	0	1	1	.00000
0	1	1	0	0	.00027
0	1	1	0	1	.00016
0	1	1	1	0	.00025
0	1	1	1	1	.00000

E	B	A	W	H	P(…)
1	0	0	0	0	.00946
1	0	0	0	1	.00001
1	0	0	1	0	.00000
1	0	0	1	1	.00000
1	0	1	0	0	.00007
1	0	1	0	1	.00004
1	0	1	1	0	.00007
1	0	1	1	1	.00000
1	1	0	0	0	.00050
1	1	0	0	1	.00000
1	1	0	1	0	.00000
1	1	0	1	1	.00000
1	1	1	0	0	.00063
1	1	1	0	1	.00037
1	1	1	1	0	.00059
1	1	1	1	1	.00000

Alarm network [Beinlich et al., 1989]

The "alarm" network (Patient monitoring):
37 variables, 509 parameters (rather than $2^{37} = 10^{11}$!)



Graphical models

A *graphical model* consists of:

- $X = \{X_1, \ldots, X_n\}$ bles
- $D = \{D_1, \ldots, D_n\}$ ains (we'll assume discrete)
- $F = \{f_{\alpha_1}, \ldots, f_{\alpha_m}\}$ ions or "factors"

and a *combination operator*

Example:

$A \in \{0,1\}$

$B \in \{0,1\}$

$C \in \{0,1\}$

$f_{AB}(A,B), \quad f_{BC}(B,C)$

The *combination operator* defines an overall function from the individual factors,

e.g., "*" : $F(A,B,C) = f_{AB}(A,B) \cdot f_{BC}(B,C)$

Notation:

Discrete X_i : values called "states"

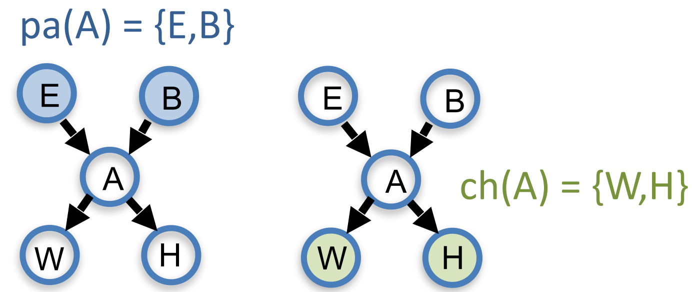
"Tuple" or "configuration": states taken by a set of variables

"Scope" of f: set of variables that are arguments to a factor f

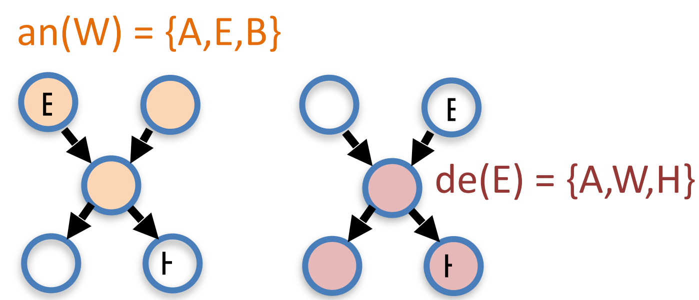
often index factors by their scope, e.g., $f_\alpha(X_\alpha), \quad X_\alpha \subseteq X$

Some terminology

- Parents & Children
  - Parents   pa(A) = {E,B}
  - Children ch(A) = {W,H}



- Ancestors & Descendants
  - Ancestors an(W) = {A,E,B}
  - Descendants de(E) = {A,W,H}



- Roots & Leaves
- Paths
  - Directed paths, undirected paths

Outline

- Graphical Models
- **Inference Tasks**
- Variable Elimination

Inference

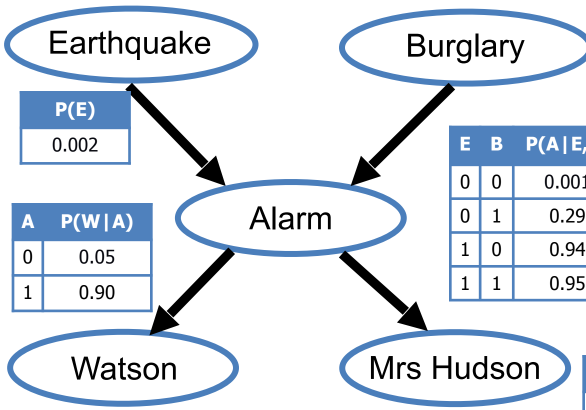
Enable us to answer **queries** about our model

- Some probabilities are directly accessible
- Some are only **implicit**, and require computation

p(B=1) = .001
Explicitly in model parameters

p(A=1) = ?
Implicit only:
p(A=1|E=0,B=0) p(E=0) p(B=0) +
p(A=1|E=1,B=0) p(E=1) p(B=0) + …

p(W=1) = ?
Implicit:
p(W=1|A=0) p(A=0) + p(W=1|A=1) p(A=1)
p(A) =? (may need to compute recursively!)



P(E)
0.002

P(B)
0.001

E	B	P(A|E,B)
0	0	0.001
0	1	0.29
1	0	0.94
1	1	0.95

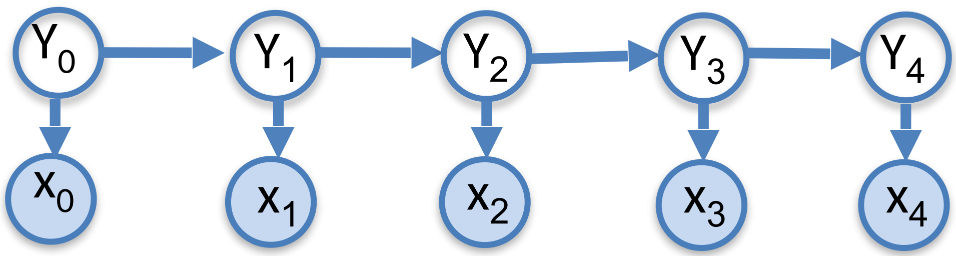
A	P(W|A)
0	0.05
1	0.90

A	P(H|A)
0	0.01
1	0.70

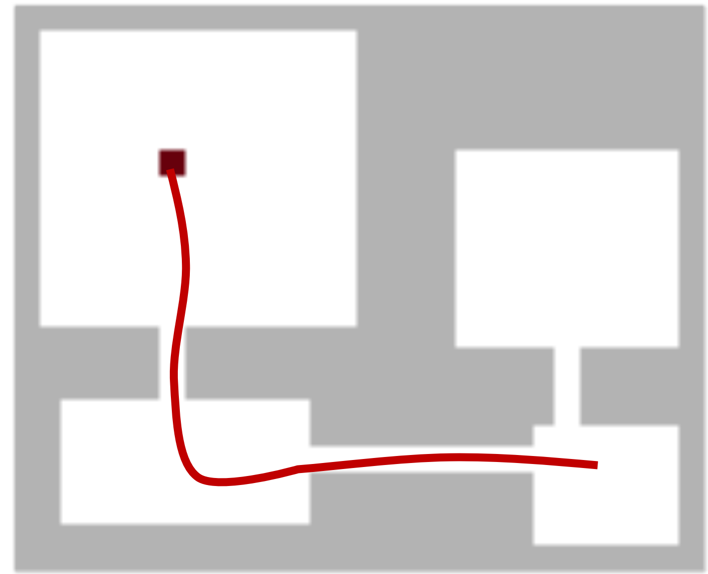
Types of queries

Ex: Robot position over time

Y_t : robot location at time t



x_t : noisy observations



Summation Query (marginal probailitiies, probability of evidence):

$$p(Y_t | x_0 \dots x_t) \propto \sum_{y_{t-1}} \dots \sum_{y_0} p(Y_t, x_t, y_{t-1}, \dots, x_0)$$

What do my model and observations tell me about my uncertainty?

Maximization Query (MAP: maximum a posteriori estmation):

$$\mathbf{y}^* = \arg \max_{y_0, \dots, y_t} p(y_t, x_t, y_{t-1}, \dots, x_0)$$

What is the most probable value of the unobserved variables?

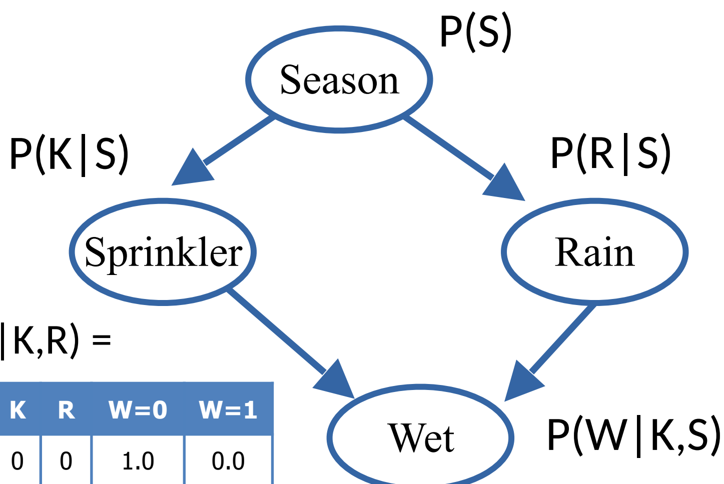
Causal Bayesian networks

- Typical BNs capture conditional independence
- May not correspond to causation; but if so:

Causal Effect Query ("Intervention"):

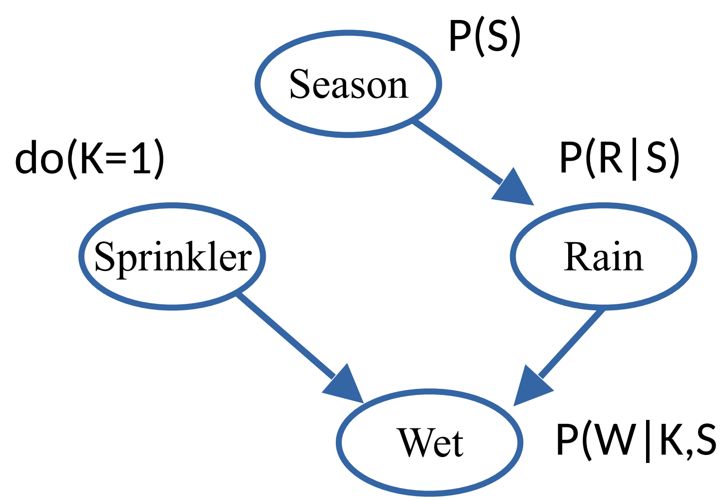
$$p(W|\text{do}(K=1))$$

What is the probability when we intervene to turn on the sprinkler?



P(W|K,R) =

K	R	W=0	W=1
0	0	1.0	0.0
0	1	0.2	0.8
1	0	0.1	0.9
1	1	0.01	0.99



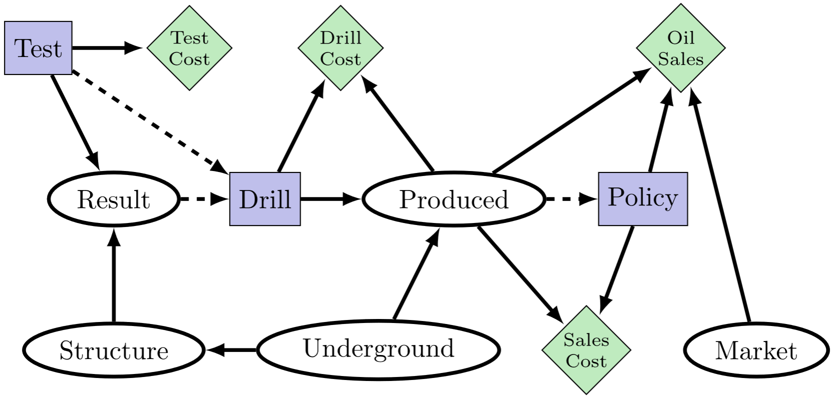
Influence diagrams

Random variables, plus **actions** (policy) and **utilities** (outcome values)

Maximum Expected Utility Query:

What actions should I take in a given situation?
What is the expected value of my policy over the actions?

The "oil wildcatter" problem:



e.g., [Raiffa 1968; Shachter 1986]

Chance variables: $X = x_1, \ldots, x_n$

Decision variables: $D = d_1, \ldots d_m$

CPDs for chance variables: $P_i = P(x_i | x_{\text{pa}_i}),$

Reward components: $r = \{r_1, \ldots, r_j\}$

Utility function: $u(X) = \sum_i r_i(X)$

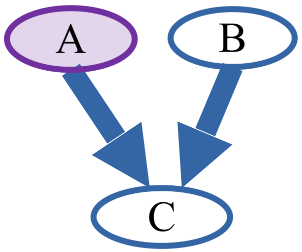
Structural Causal Models

Deterministic mechanisms involving (random) underlying causes

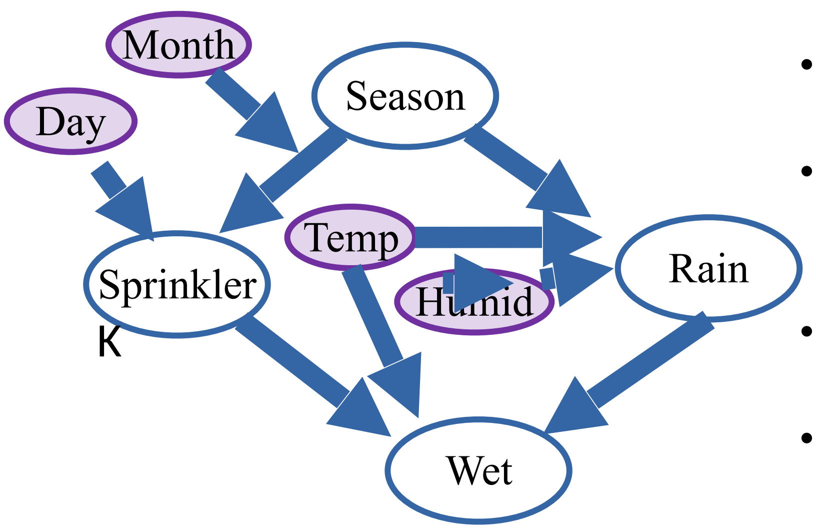
Unobservable random variables: {A}

Observable variables: {B,C}

Deterministic mechanism: → C = f(A,B)



Ex: Sprinkler



- p(S): season a function of (unobserved) month
- p(K|S): sprinkler on due to watering schedule: randomness in K due to (unobserved) day of week
- p(R|S) caused by humidity and temperature
- p(W|R,K) also caused by humidity and temperature (effects of evaporation, etc.)

Structural Causal Models

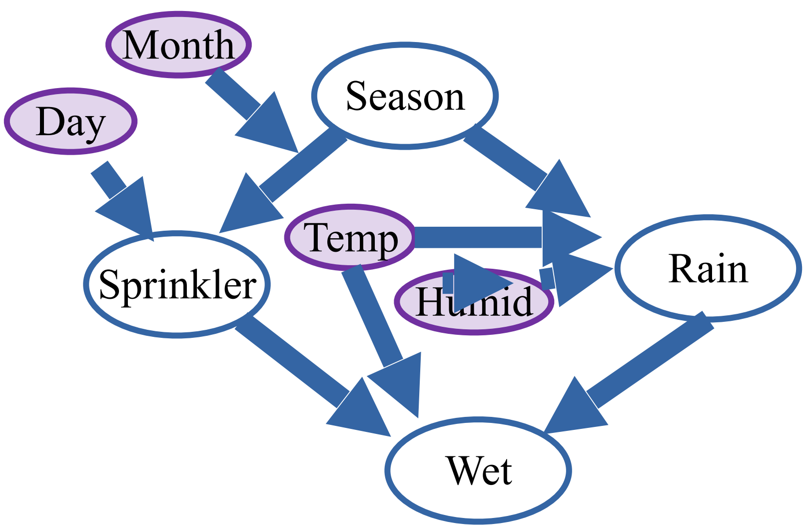
Deterministic mechanisms involving (random) underlying causes

Counterfactual Query:

Probability of an event in contradiction with the observations
*What **would have happened** if the sprinkler had been turned off?*

Requires that we transfer information about random outcomes that happened, to a different setting

Ex: Sprinkler



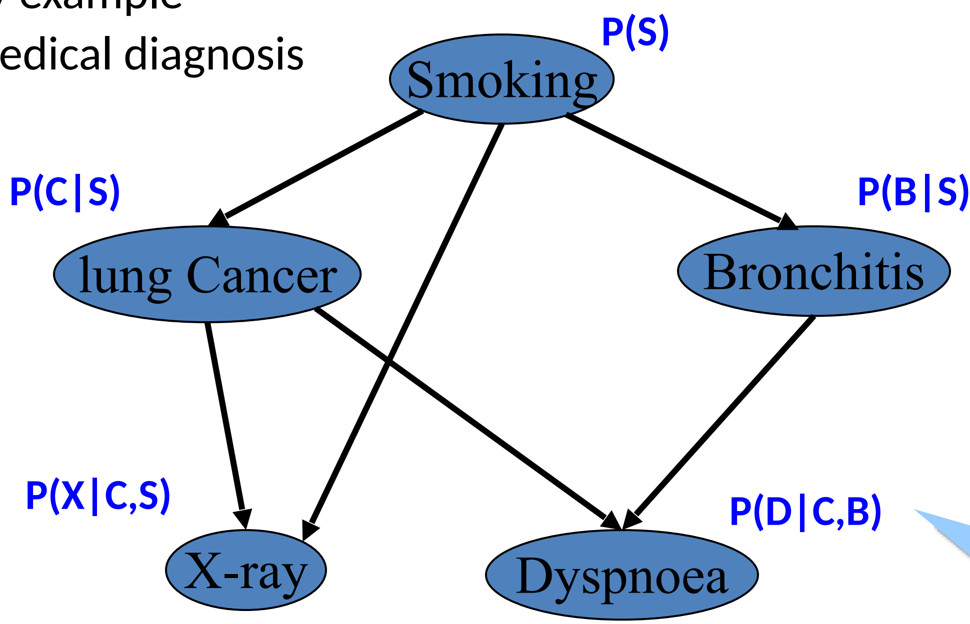
Observe the sprinkler is on & grass is wet: (K=1,W=1)

What is the probability it would still be wet if we had turned the sprinkler off?

Observing K=1 tells us it is more likely to be summer;
Observing K=1,W=1 tells us it is not too hot & dry.

Then, apply this knowledge to compute the counterfactual: $p(W_{K=0} \mid K=1, W=1)$

An early example
From medical diagnosis



BN = (G, Θ)

CPD:

C	B	P(D|C,B)	
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

$P(S, C, B, X, D) = P(S)\, P(C|S)\, P(B|S)\, P(X|C,S)\, P(D|C,B)$

Combination: Product
Marginalization: sum/max

- Posterior marginals, probability of evidence, MPE

Is this a causal model?

- $P(D=0) = \sum_{S,L,B,X} P(S) \cdot P(C|S) \cdot P(B|S) \cdot P(X|C,S) \cdot P(D|C,B$

MAP(P)= $max_{S,L,B,X}$ P(S)· P(C|S)· P(B|S)· P(X|C,S)· P(D|C,B)

Constraint Networks

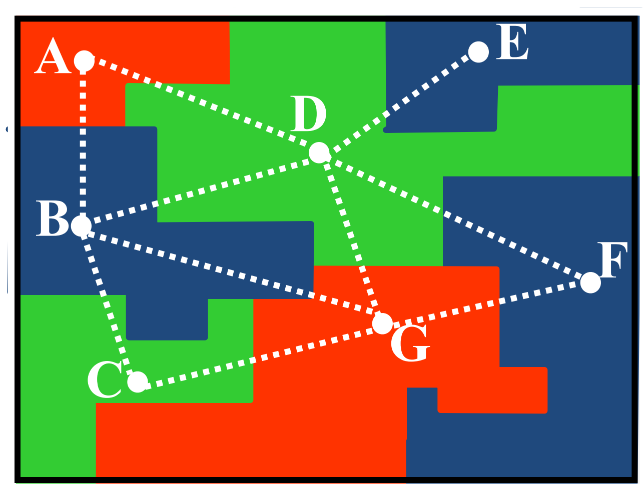
Example: map coloring

Variables - countries (A,B,C,etc.)

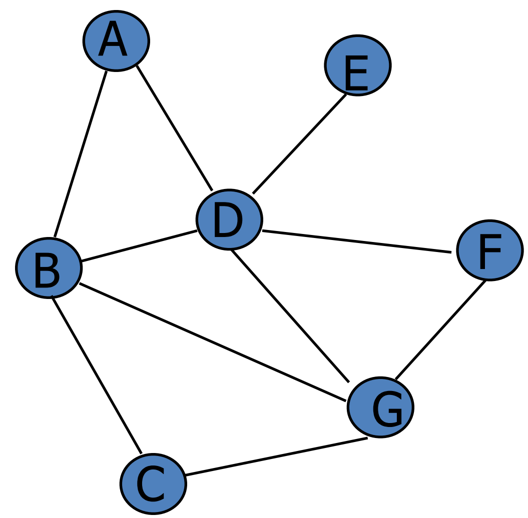
Values - colors (red, green, blue)

Constraints: $A \neq B$, $A \neq D$, $D \neq E$, etc.

A	B
red	green
red	yellow
green	red
green	yellow
yellow	green
yellow	red



Constraint graph



Propositional Reasoning

Example: party problem

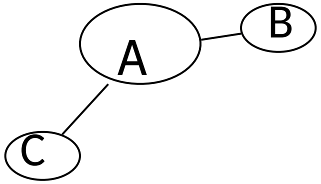
= A
= B

- If Alex goes, then Becky goes: $A \rightarrow B$
- If Chris goes, then Alex goes: $C \rightarrow A$

= C
= A

- **Question:**

Is it possible that Chris goes to the party but Becky does not?

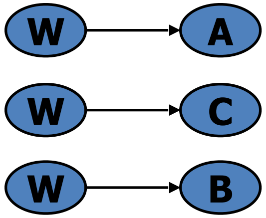


Is the *propositional theory*
$\phi = \{A \rightarrow B, C \rightarrow A, \neg B, C\}$ satisfiable?

Probabilistic reasoning (directed)

Party example: the weather effect

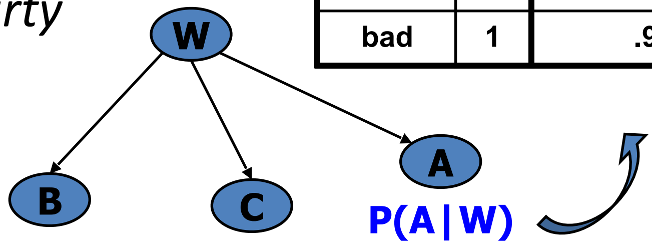
- Alex is-likely-to-go in bad weather
- Chris rarely-goes in bad weather
- Becky is indifferent but unpredictable



Questions:

- *Given bad weather, which group of individuals is most likely to show up at the party?*
- *What is the probability that Chris goes to the party but Becky does not?*

W	A	P(A|W)
good	0	.01
good	1	.99
bad	0	.1
bad	1	.9



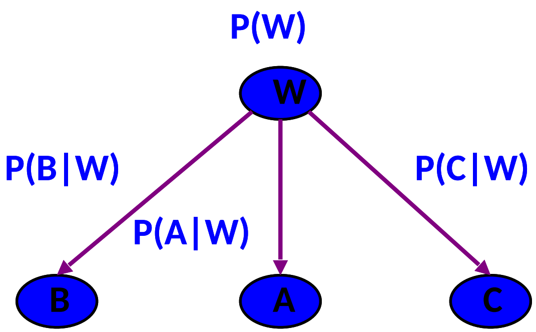
P(W,A,C,B) = P(B|W) · P(C|W) · P(A|W) · P(W)

P(A,C,B|W=bad) = 0.9 · 0.1 · 0.5

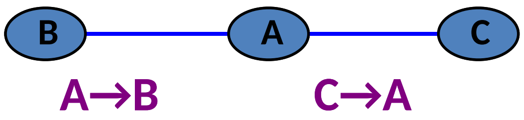
Mixed Probabilistic and Deterministic networks

Alex is-likely-to-go in bad weather
Chris rarely-goes in bad weather
Becky is indifferent but unpredictable

PN



CN



Query:
Is it likely that Chris goes to the party if Becky does not but the weather is bad?

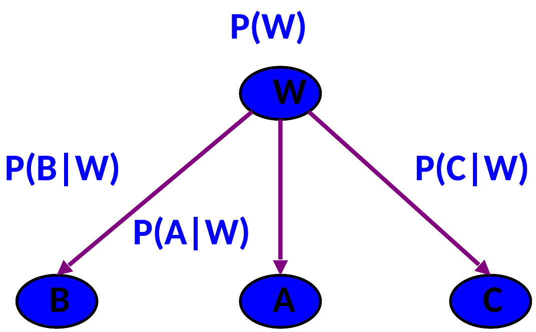
$$P(C, \neg B \mid w = bad, A \rightarrow B, C \rightarrow A)$$

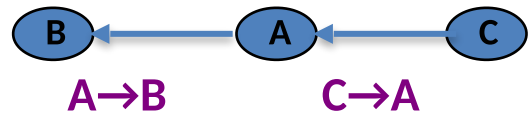
Causal Probabilistic and Deterministic networks

$$P(C, \neg B \,|\, w = bad, A \rightarrow B, C \rightarrow A)$$

Alex is-likely-to-go in bad weather
Chris rarely-goes in bad weather
Becky is indifferent but unpredictable

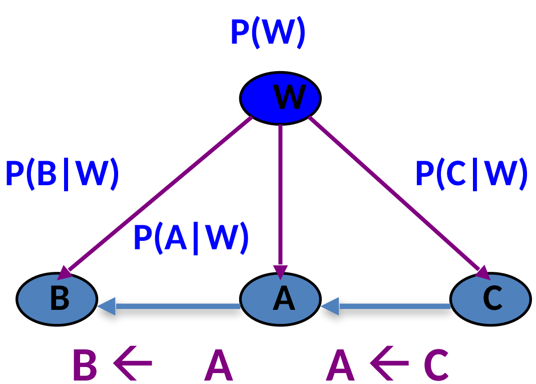
PN





Causal effect query vs obs query:

- *Is it likely that Becky goes to the party if Chris does not?*
- *Is it likely that Becky goes to the party if **we force Chris to go.***



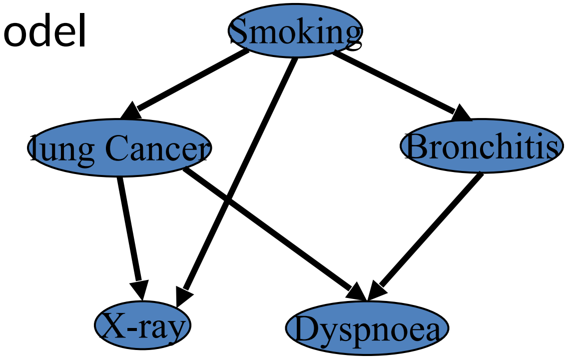
$$P(B \,|\, do(C = go)\,,\; w = bad)$$

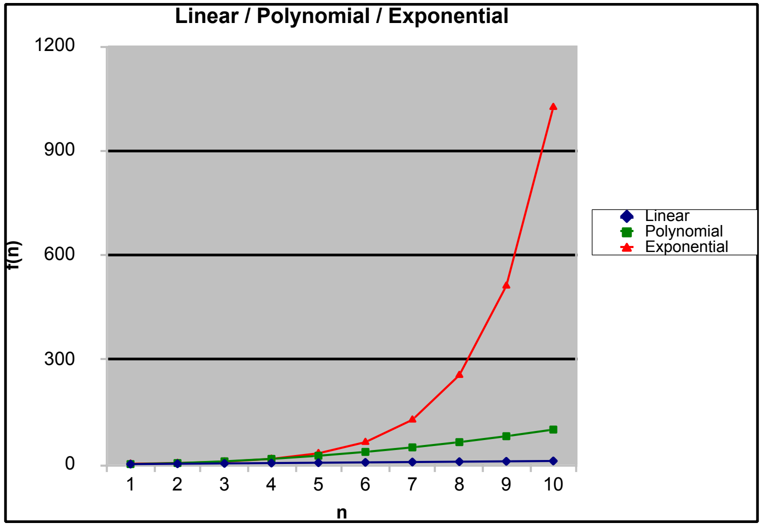
$$P(B \,|\, C = go.\;\; w = bad)$$

Complexity of Reasoning Tasks

- Constraint satisfaction
- Counting solutions
- Combinatorial optimization
- Belief updating
- Most probable explanation
- Decision-theoretic planning
- Causal reasoning

Given a **full** model





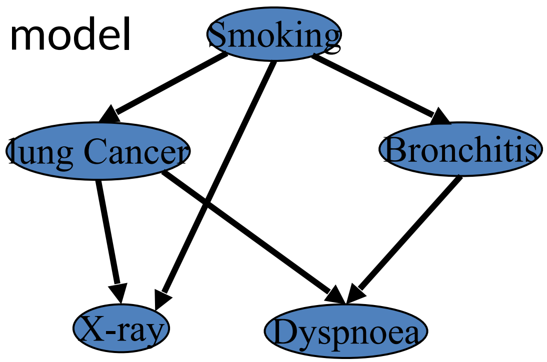
Reasoning is
computationally hard

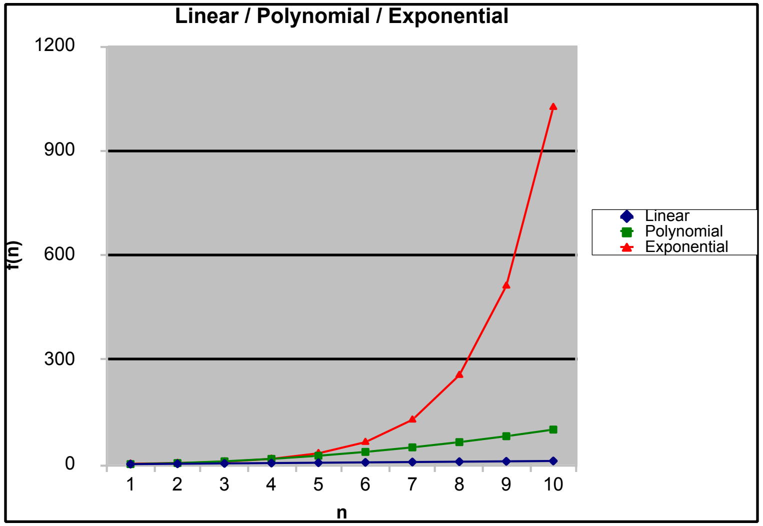
Complexity is
Time and space(memory)

Complexity of Causal Tasks

- Constraint satisfaction
- Counting solutions
- Combinatorial optimization
- Belief updating
- Most probable explanation
- Decision-theoretic planning
- **Causal reasoning**

Given a **partial** model
And **data**...





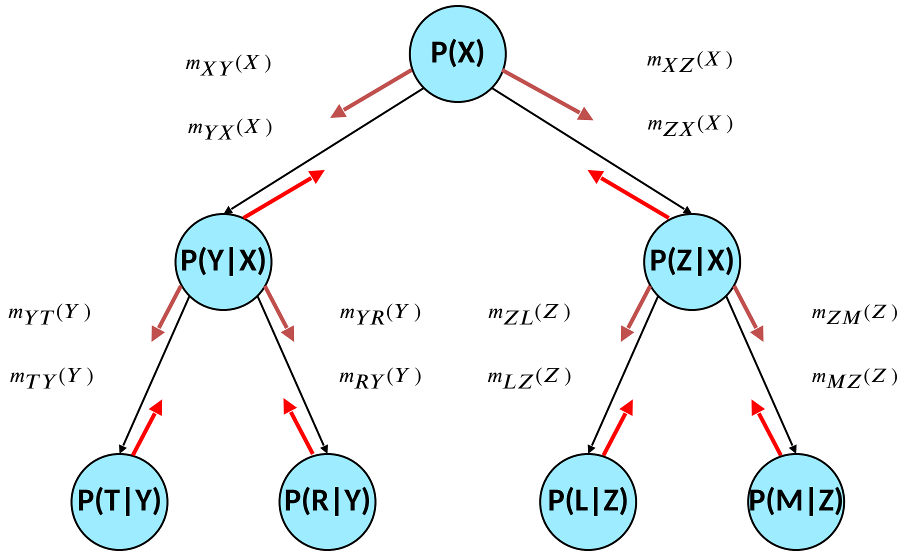
Reasoning is computationally hard

Complexity is
Time and space(memory)

Tree-solving is easy

Belief updating (sum-prod)

CSP – consistency (projection-join)



MPE (max-prod)

#CSP (sum-prod)

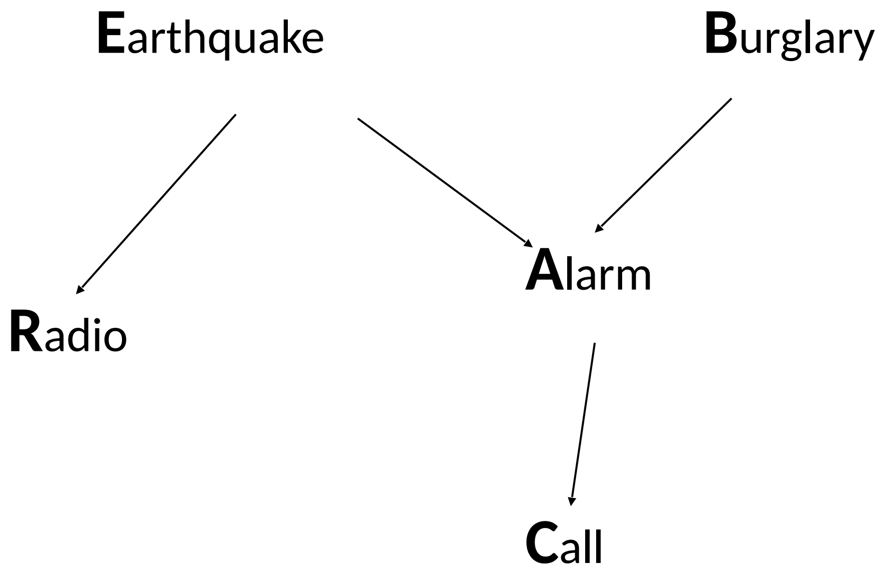
Trees are processed in linear time and memory

Transforming into a Tree

- **By Inference (thinking)**
  - Transform into a single, equivalent tree of sub-problems
- **By Conditioning (guessing)**
  - Transform into many tree-like sub-problems.

Basics of Probabilistic Calculus (Chapter 3)

The Burglary Example



Degrees of Belief

- Assign a degree of belief or probability in $[0,1]$ to each world ω and denote it by $\Pr(\omega)$.
- The belief in, or probability of, a sentence α:

$$\Pr(\alpha) \overset{def}{=} \sum_{\omega \models \alpha} \Pr(\omega).$$

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

Properties of Beliefs

- A bound on the belief in any sentence:

$$0 \leq \Pr(\alpha) \leq 1 \qquad \text{for any sentence } \alpha.$$

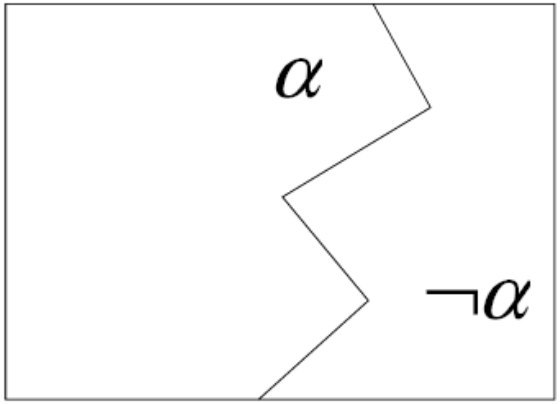
- A baseline for inconsistent sentences:

$$\Pr(\alpha) = 0 \qquad \text{when } \alpha \text{ is inconsistent.}$$

- A baseline for valid sentences:

$$\Pr(\alpha) = 1 \qquad \text{when } \alpha \text{ is valid.}$$

Properties of Beliefs



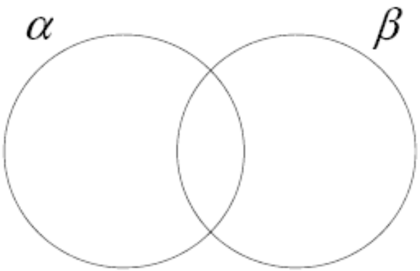
- The belief in a sentence given the belief in its negation:

$$\Pr(\alpha) + \Pr(\neg\alpha) = 1.$$

Example

$$\begin{aligned}
\Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\
\Pr(\neg\text{Burglary}) &= \Pr(\omega_3) + \Pr(\omega_4) + \Pr(\omega_7) + \Pr(\omega_8) = .8
\end{aligned}$$

Properties of Beliefs



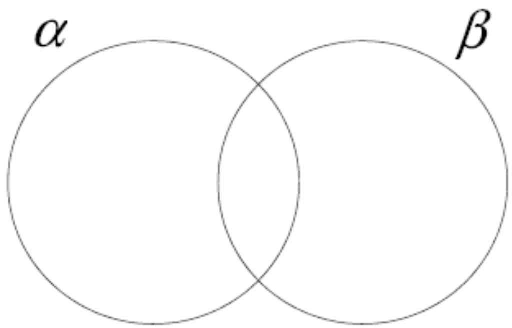
- The belief in a disjunction:

$$\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta) - \Pr(\alpha \wedge \beta).$$

- Example:

$$\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\
\Pr(\text{Earthquake} \wedge \text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) = .02 \\
\Pr(\text{Earthquake} \vee \text{Burglary}) &= .1 + .2 - .02 = .28
\end{aligned}$$

Properties of Beliefs



- The belief in a disjunction:

  $\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta)$ when α and β are mutually exclusive.

Entropy

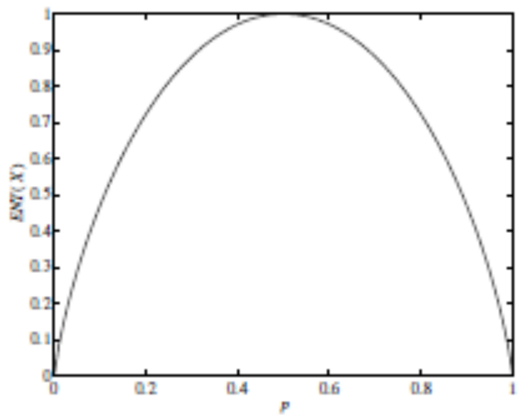
Quantify uncertainty about a variable X using the notion of entropy:

$$\mathrm{ENT}(X) \stackrel{def}{=} -\sum_{x} \Pr(x) \log_2 \Pr(x),$$

where $0 \log 0 = 0$ by convention.

	Earthquake	Burglary	Alarm
true	.1	.2	.2442
false	.9	.8	.7558
ENT(.)	.469	.722	.802

Entropy



- The entropy for a binary variable X and varying $p = \Pr(X)$.
- Entropy is non-negative.
- When $p = 0$ or $p = 1$, the entropy of X is zero and at a minimum, indicating no uncertainty about the value of X.
- When $p = \frac{1}{2}$, we have $\Pr(X) = \Pr(\neg X)$ and the entropy is at a maximum (indicating complete uncertainty).

Bayes Conditioning

Alpha and beta are events

Closed form for Bayes conditioning:

$$\Pr(\alpha|\beta) = \frac{\Pr(\alpha \land \beta)}{\Pr(\beta)}.$$

Defined only when $\Pr(\beta) \neq 0$.

Degrees of Belief

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

$$
\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= .2 \\
\Pr(\neg\text{Burglary}) &= .8 \\
\Pr(\text{Alarm}) &= .2442
\end{aligned}
$$

Belief Change

Burglary is independent of Earthquake

Conditioning on evidence Earthquake:

$$\begin{aligned} \Pr(\text{Burglary}) &= .2 \\ \Pr(\text{Burglary}|\text{Earthquake}) &= .2 \\ \\ \Pr(\text{Alarm}) &= .2442 \\ \Pr(\text{Alarm}|\text{Earthquake}) &\approx .75 \uparrow \end{aligned}$$

The belief in Burglary is not changed, but the belief in Alarm increases.

Belief Change

Earthquake is independent of burglary

Conditioning on evidence Burglary:

$$
\begin{aligned}
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Burglary}) &\approx .905 \uparrow \\
\\
\Pr(\text{Earthquake}) &= .1 \\
\Pr(\text{Earthquake}|\text{Burglary}) &= .1
\end{aligned}
$$

The belief in Alarm increases in this case, but the belief in Earthquake stays the same.

Belief Change

The belief in Burglary increases when accepting the evidence Alarm. How would such a belief change further upon obtaining more evidence?

- Confirming that an Earthquake took place:

$$\begin{aligned} \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\ \Pr(\text{Burglary}|\text{Alarm} \wedge \text{Earthquake}) &\approx .253 \downarrow \end{aligned}$$

  We now have an explanation of Alarm.

- Confirming that there was no Earthquake:

$$\begin{aligned} \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\ \Pr(\text{Burglary}|\text{Alarm} \wedge \neg\text{Earthquake}) &\approx .957 \uparrow \end{aligned}$$

  New evidence will further establish burglary as an explanation.

Conditional Independence

$\Pr$ finds α conditionally independent of β given γ iff

$$\Pr(\alpha|\beta \wedge \gamma) = \Pr(\alpha|\gamma) \quad \text{or } \Pr(\beta \wedge \gamma) = 0.$$

Another definition

$$\Pr(\alpha \wedge \beta|\gamma) = \Pr(\alpha|\gamma)\Pr(\beta|\gamma) \quad \text{or } \Pr(\gamma) = 0.$$

Variable Independence

$\Pr$ finds $\mathbf{X}$ independent of $\mathbf{Y}$ given $\mathbf{Z}$, denoted $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, means that $\Pr$ finds $\mathbf{x}$ independent of $\mathbf{y}$ given $\mathbf{z}$ for all instantiations $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$.

Example

$\mathbf{X} = \{A, B\}$, $\mathbf{Y} = \{C\}$ and $\mathbf{Z} = \{D, E\}$, where A, B, C, D and E are all propositional variables. The statement $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is then a compact notation for a number of statements about independence:

$A \wedge B$ is independent of C given $D \wedge E$;

$A \wedge \neg B$ is independent of C given $D \wedge E$;

$\vdots$

$\neg A \wedge \neg B$ is independent of $\neg C$ given $\neg D \wedge \neg E$;

That is, $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is a compact notation for $4 \times 2 \times 4 = 32$ independence statements of the above form.

Further Properties of Beliefs

Chain rule

$$\begin{aligned} & \Pr(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n) \\ = \;\; & \Pr(\alpha_1 | \alpha_2 \wedge \ldots \wedge \alpha_n) \Pr(\alpha_2 | \alpha_3 \wedge \ldots \wedge \alpha_n) \ldots \Pr(\alpha_n). \end{aligned}$$

Case analysis (law of total probability)

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha \wedge \beta_i),$$

where the events $\beta_1, \ldots, \beta_n$ are mutually exclusive and exhaustive.

Further Properties of Beliefs

Another version of case analysis

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha|\beta_i)\Pr(\beta_i),$$

where the events $\beta_1, \ldots, \beta_n$ are mutually exclusive and exhaustive.

Two simple and useful forms of case analysis are these:

$$\begin{aligned}\Pr(\alpha) &= \Pr(\alpha \wedge \beta) + \Pr(\alpha \wedge \neg\beta) \\ \Pr(\alpha) &= \Pr(\alpha|\beta)\Pr(\beta) + \Pr(\alpha|\neg\beta)\Pr(\neg\beta).\end{aligned}$$

The main value of case analysis is that, in many situations, computing our beliefs in the cases is easier than computing our beliefs in α. We shall see many examples of this phenomena in later chapters.

Further Properties of Beliefs

Bayes rule

$$\Pr(\alpha|\beta) = \frac{\Pr(\beta|\alpha)\Pr(\alpha)}{\Pr(\beta)}.$$

- Classical usage: α is perceived to be a cause of β.
- Example: α is a disease and β is a symptom–
- Assess our belief in the cause given the effect.
- Belief in an effect given its cause, $\Pr(\beta|\alpha)$, is usually more readily available than the belief in a cause given one of its effects, $\Pr(\alpha|\beta)$.