Bayesian Networks

Lyle Ungar

Heavily adapted from slides by Mitch Marcus

Learning objectives

Semantics of belief nets: conditional independence
Active trails
D-separation / Markov separation
Belief net construction

Bayesian Networks

- **A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions**

- **Syntax:**
  - A set of nodes, one per random variable
  - A set of directed edges (link ≈ "directly influences"), yielding a directed, acyclic graph
  - A conditional distribution for each node given its parents:
    P *$(X_i$| Parents $(X_i))$*

- **In the simplest case, the conditional distribution is represented as a *conditional probability table* (CPT) giving the distribution over X_i for each combination of parent values**

Why not just specify the full probability distribution?

The Required-By-Law Burglar Example

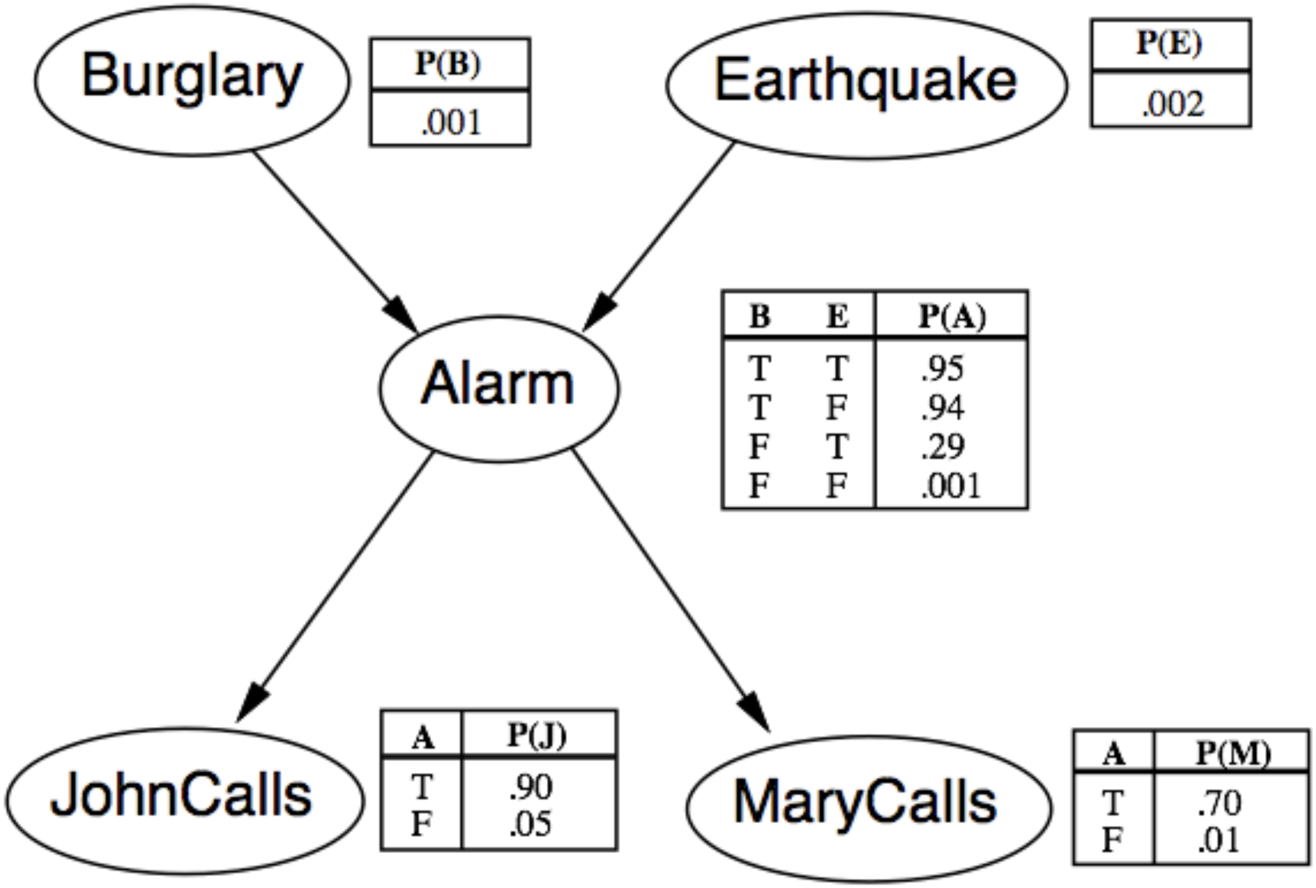
I'm at work, and my neighbor John calls to say my burglar alarm is ringing, but my neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: *Burglary, Earthquake, Alarm, JohnCalls, MaryCalls*

Network topology ideally reflects causal knowledge:

1. A burglar can set the alarm off
2. An earthquake can set the alarm off
3. The alarm can cause Mary to call
4. The alarm can cause John to call

Burglar belief network



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Local Markov Assumption (LMA)

- **LMA: A variable X is independent of its non-descendants given its parents:**
  - $(X \perp \text{NonDescendants}_X | \text{Parents}_X)$

- **The LMA lets us decompose the joint in terms of the CPTs:**

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

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

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

Conditional Independence Properties

- **Symmetry:** $(X \perp Y | Z) \longrightarrow (Y \perp X | Z)$
- **Decomposition:** $(X \perp Y, W | Z) \longrightarrow (X \perp Y | Z)$
- **Weak union:** $(X \perp Y, W | Z) \longrightarrow (X \perp Y | Z, W)$
- **Contraction:** $(X \perp W | Y, Z), (X \perp Y | Z) \longrightarrow (X \perp Y, W | Z)$

Burglar belief network

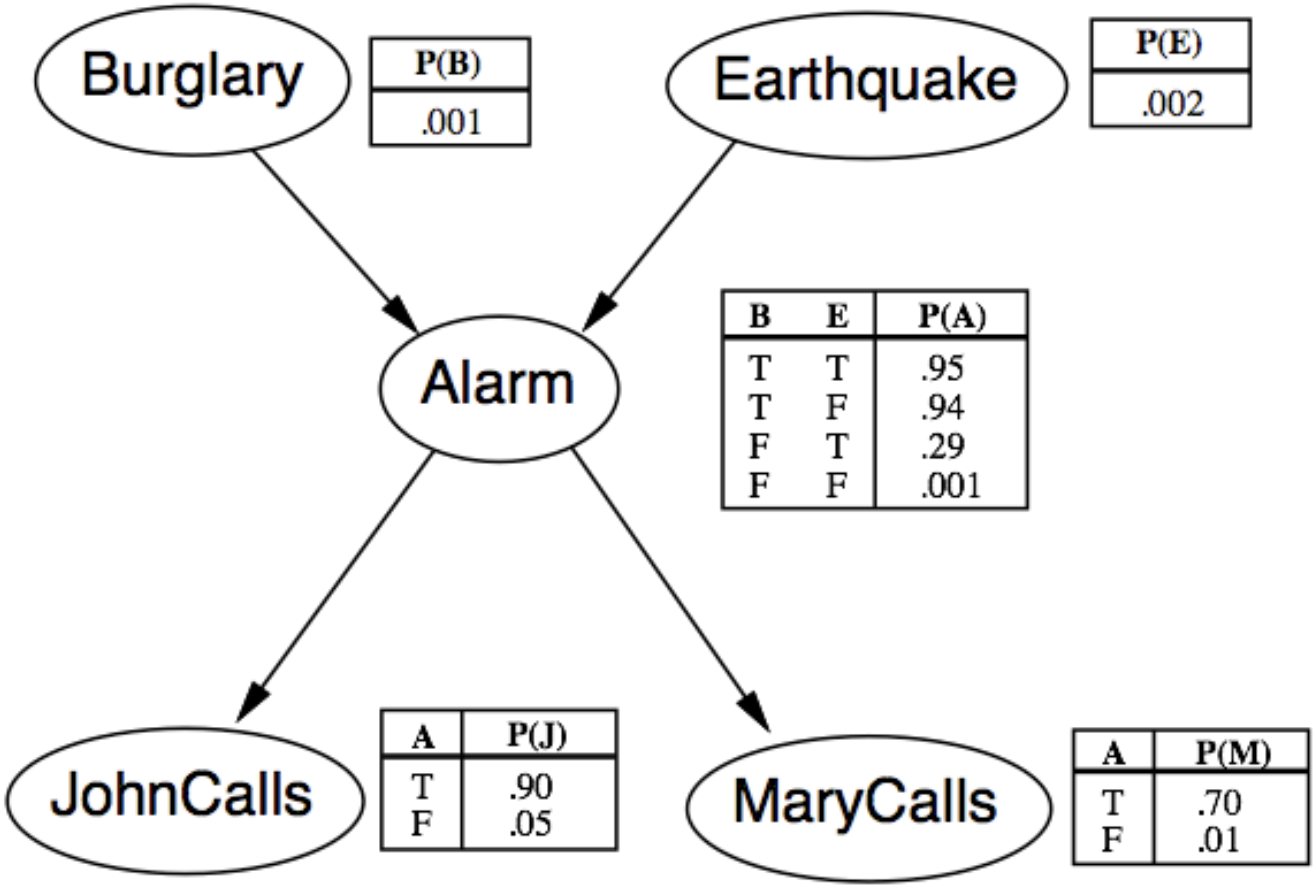
$B \perp J$?

$B \perp J | A$?

$J \perp M | A$?

$B \perp E | A$?

Explaining away



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Active Trails

A trail $\{X_1,X_2,\cdots,X_k\}$ in the graph (no cycles) is an **active trail** if for each consecutive triplet in the trail:

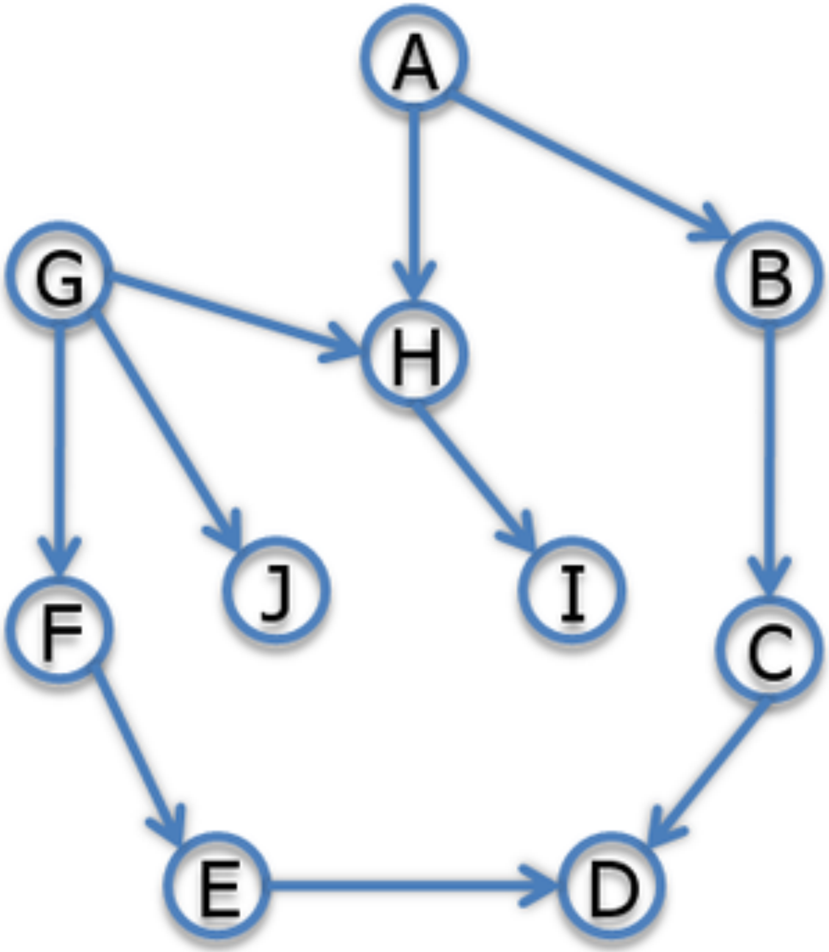
- $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$, and X_i is not observed
  $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, and X_i is not observed
  $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, and X_i is not observed
  $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, and X_i is observed or one of its descendants is observed

Variables connected by active trails are not conditionally independent

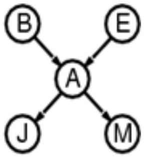
D-separation

- Variables X_i and X_j are independent if there is no *active trail* between X_i and X_j .
  - given a set of observed variables $O \subset \{X_1, \cdots, X_m\}$

Examples



Semantics



- ***Local* semantics give rise to *global* semantics**
- **Local semantics:** given its parents, each node is conditionally independent of everything except its descendants
- **Global semantics (why?):**

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

$$P(J \land M \land A \land \neg B \land \neg E) =$$

$$P(\neg B)P(\neg E)P(A|\neg B \land \neg E)P(J|A)P(M|A)$$

Belief Network Construction Algorithm

1. **Choose an ordering of variables $X_1, \dots , X_n$**
2. **For i = 1 to n**
   a. add X_i to the network
   b. select parents from $X_1, \dots , X_{i-1}$ such that

$$P(X_i|Parents(X_i)) = P(X_i|X_1, \dots, X_{i-1})$$

Belief Network Construction Details

1) **Order the variables (any ordering will do)**

   A, B, C, D, E,..

2) **Pop the first item off the list, and determine what the minimal set of parents is**

   i) ***add A*** **(trivial)**

   ii) ***add B***

   is P(B|A) = P(B) and P(B|~A) = P(B)

   if either equality test fails, then add a link from A to B

Belief Network Construction Details

iii) ***add C***

is $P(C) = P(C|A,B) = P(C|\sim A,B) = P(C|A,\sim B) = P(C|\sim A,\sim B)$

if so, C Is *not* linked to anything; proceed to test D.

otherwise there are one or more links

is $P(C) = P(C|A) = P(C|\sim A)$

if not, then there is a link from A to C and we also test:

is $P(C|A) = P(C|A,B) = P(C|A, \sim B)$ and $P(C|\sim A) = P(C|\sim A,B) = P(C|\sim A, \sim B)$

if so then we only need a link from A to C and proceed to add D

otherwise keep both, then check:

is $P(C) = P(C|B) = P(C|\sim B)$

if not, then there is a link from B to C and we also test:

is $P(C|B) = P(C|B,A) = P(C|B, \sim A)$ and $P(C|\sim B) = P(C|A,\sim B) = P(C|\sim A, \sim B)$

if so then we only need a link from B to C

if not, both links are present.

iv) ***add D***

things keep getting uglier, but the same idea continues.

Comment about previous slide

There are four possible structures for adding C.

C doesn't depend on A or B

C depends on A but not B

C depends on B but not A

C depends on both

if P(C) = P(C|A) = P(C| ~A) then most likely C does not depend on A

(the special case we still need to check is that A alone tells us nothing about C, but A and B together are informative)

similarly, but giving us different information

if P(C) = P(C|B) = P(C| ~B) then most likely C does not depend on B

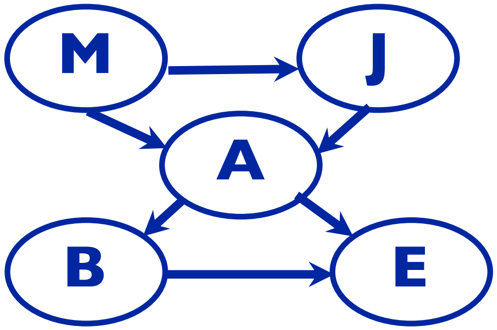
This needs to be checked even if we know C doesn't depend on A

I'm not sure you need to check the special case again, but just to be safe:

(the special case we still need to check is that B alone tells us nothing about C, but A and B together are informative)

Construction Example

Suppose we choose the ordering M, J, A, B, E



$P(J|M)=P(J)?$

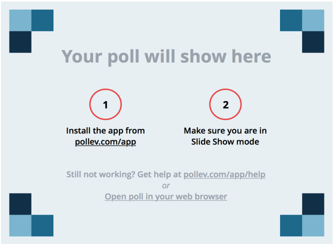
$P(A|J,M)=P(A)?$ $P(A|J,M)=P(A|J)?$

$P(B|A,J,M)=P(B)?$

$P(B|A,J,M)=P(B|A)?$

$P(E|B,A,J,M)=P(E|A)?$

$P(E|B,A,J,M)=P(E|A,B)?$



Your poll will show here

1 Install the app from pollev.com/app

2 Make sure you are in Slide Show mode

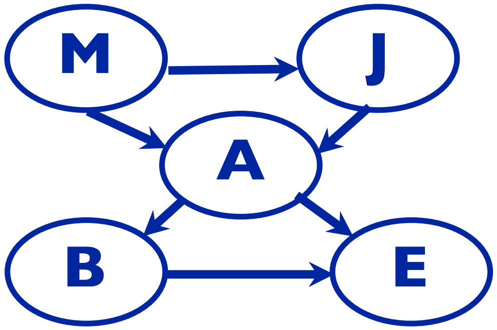
Still not working? Get help at pollev.com/app/help

or

Open poll in your web browser

Construction Example

Suppose we choose the ordering M, J, A, B, E



P(J|M)=P(J)? **No**

P(A|J,M)=P(A)? P(A|J,M)=P(A|J)? **No**

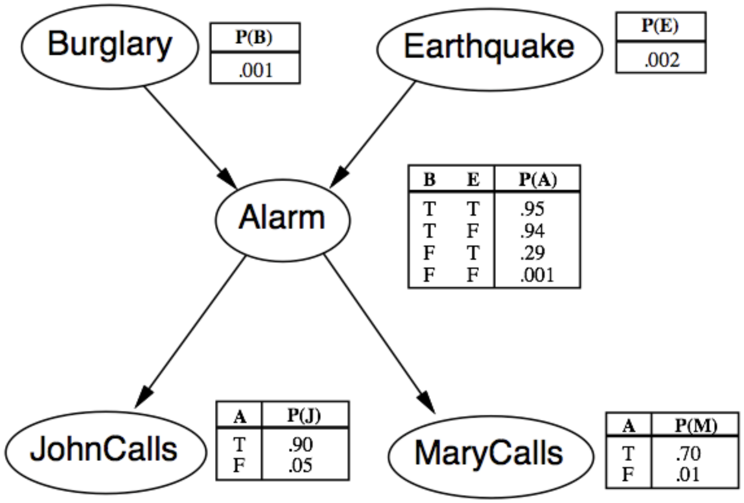
P(B|A,J,M)=P(B)? **No**

P(B|A,J,M)=P(B|A)? *Yes*

P(E|B,A,J,M)=P(E|A)? **No**

P(E|B,A,J,M)=P(E|A,B)? *Yes*

Lessons from the example



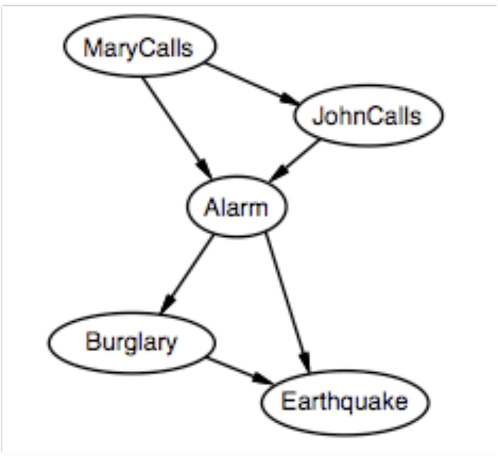
P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01



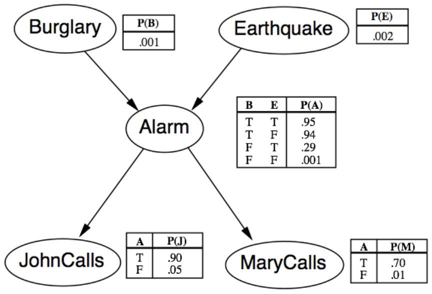
- Network less compact: 13 numbers (compared to 10)
- Ordering of variables can make a big difference!
- Intuitions about causality are useful

How to build a belief net?

- **Ask people**
  - Try to model causality in structure
  - Estimate probabilities from data
- **Automatic construction**
  - How to cast belief network building as search?

How to learn a belief net?

- **Pick the loss function that you want to minimize**
  - - log(likelihood) + const * (# of parameters)
    - Parameters, θ, are the numbers in the conditional probability tables
  - Likelihood = $p(X|\theta) = \Pi_i\, p(x_i|\theta)$
    - Probability of the data given the model
- **Do stochastic gradient descent (or annealing)**
  - Randomly change the network structure by one link
  - Re-estimate the parameters
  - Accept the change if the loss function is lower
    - Or sometimes even if it is higher



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

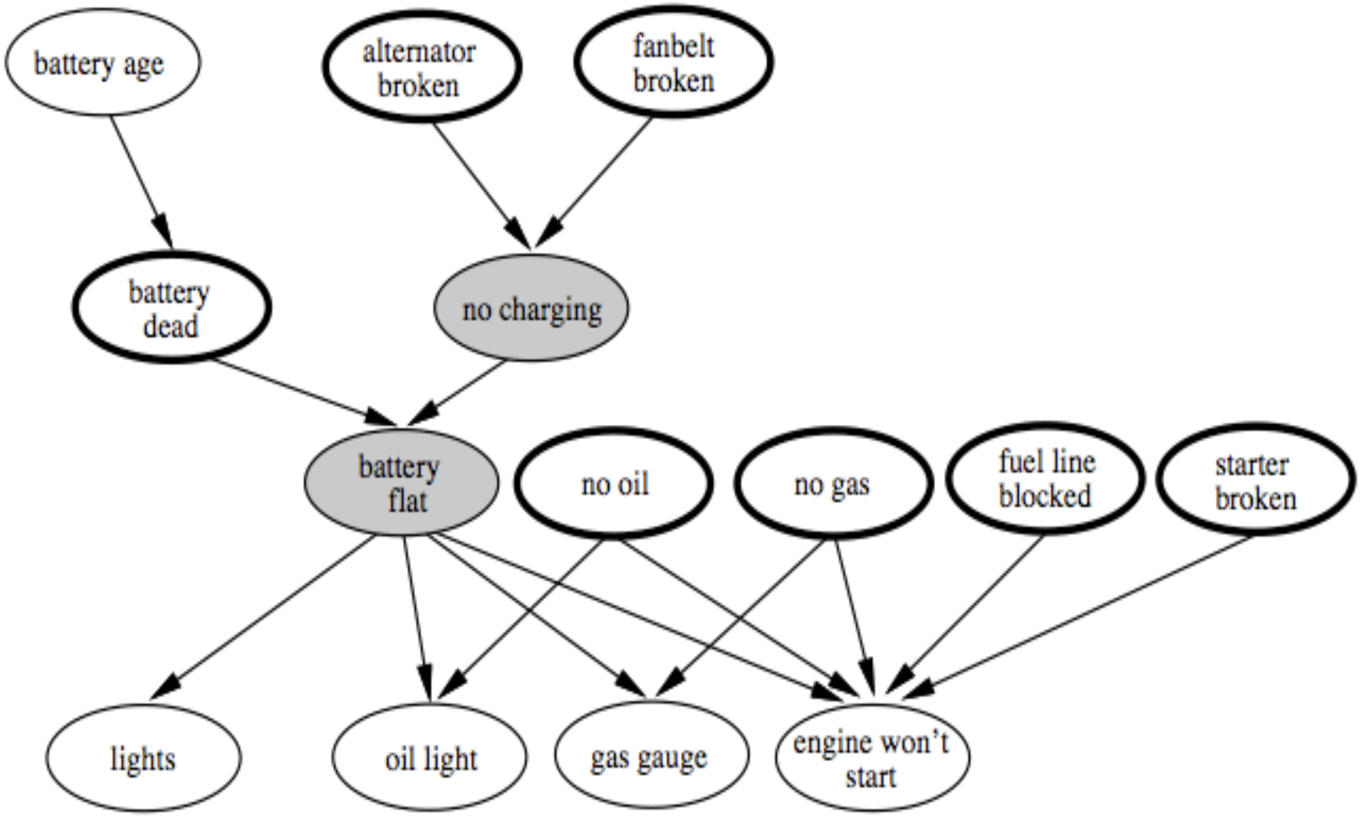
A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Hidden/latent variables

- **Sometimes not every variable is observable**
- **In this case, to do estimation, use EM**
  - If you know the values of the latent variable, then it is trivial to estimate the model parameters (here, the values in the conditional probability tables)
  - If you know the model (the probabilities), you can find the expected values of the non-observed variables
- **Thus, given a known structure, one can find a (local) maximum likelihood model.**

Belief Network with Hidden Variables: Example



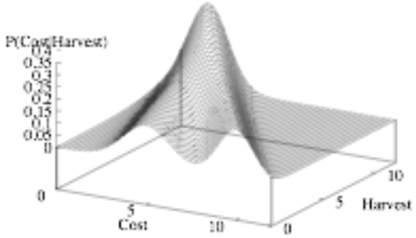
Belief Net Extensions

- **Hidden variables**
- **Decision (action) variables**
- **Continuous distributions**
- **“Plate” models**
  - E.g. Latent Dirichlet Allocation (LDA)
- **Dynamical Belief Nets (DBNs)**
  - E.g. Hidden Markov Models (HMMs)

What is the simplest Belief net with hidden variables?

Compact Conditional Probability Tables

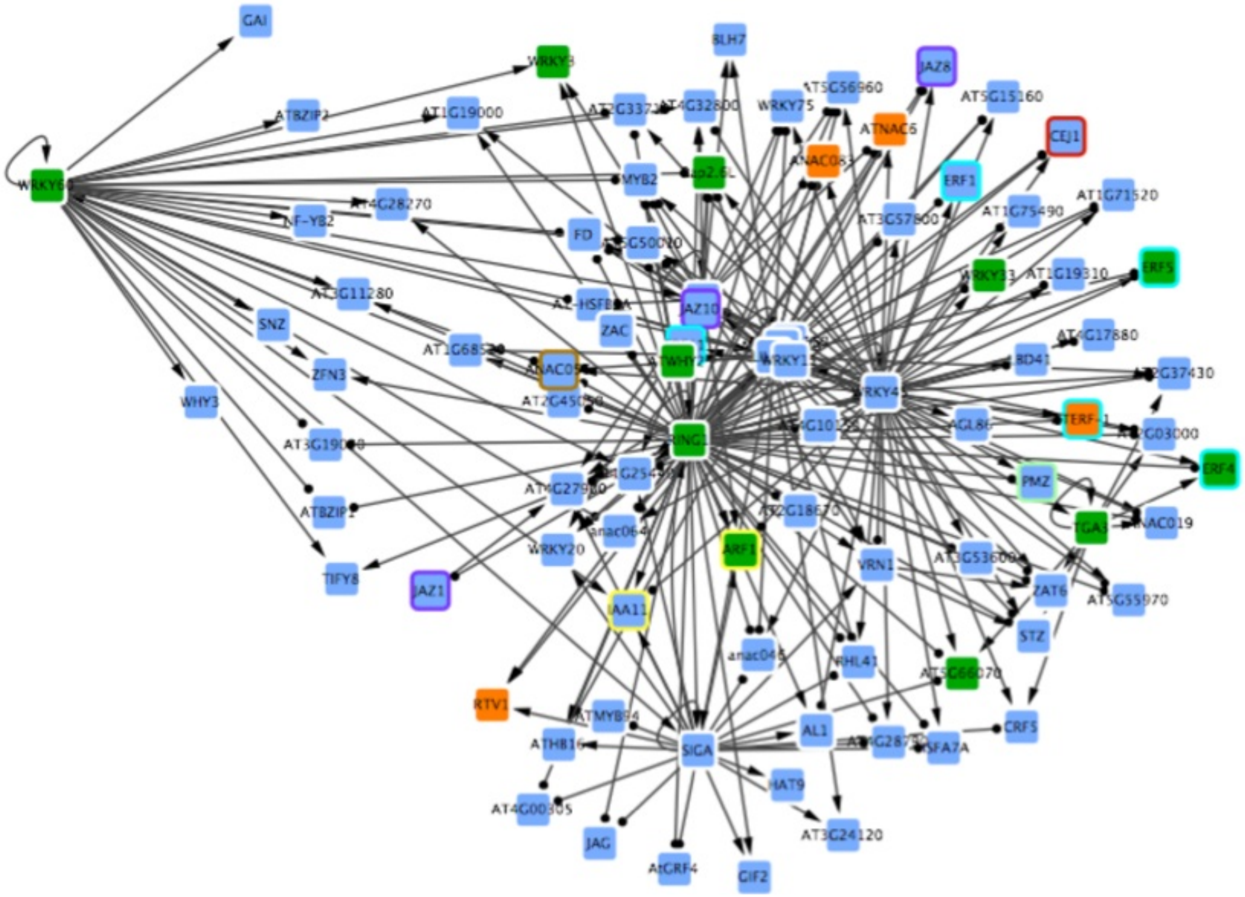
- **CPT size grows exponentially in number of variables**
  - continuous variables have infinite CPTs!
- **Solution: *canonical* distributions**
  - E.g., Boolean functions, “noisy OR”
  - Gaussian and other standard distributions



What you should know

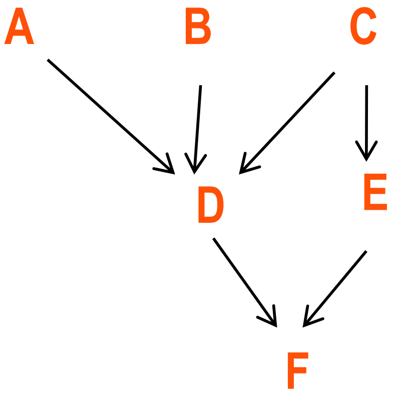
- **Belief nets provide a compact representation of joint probability distributions**
  - Guaranteed to be a consistent specification
  - Graphically show conditional independence
  - Can be extended to include “action” nodes
- **Networks that capture causality tend to be sparser**
- **Estimating belief nets is easy if everything is observable**
  - But requires serious search if the network structure is not known

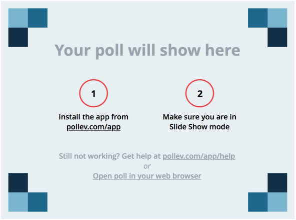
Gene transcriptional response to pathogen infection



Bayesian network questions

- **What does the belief network diagram mean?**
  - Why are the diagrams good?
  - Why are the diagrams bad?
- **How many parameters in the model below?**
  - A) <12 B) 12-16 C)17-20 D) more than 20
  - How does this change if a variable can take on 3 values?
- **What is conditionally independent of what?**
  - P(D|C) = P(D|C,E)?
  - P(D|F) = P(D|F,E)?
  - P(F|D) = P(F|D,A)?
  - P(F|D) = P(F|D,C)?




Your poll will show here

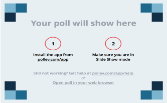
1 Install the app from pollev.com/app

2 Make sure you are in Slide Show mode

Still not working? Get help at pollev.com/app/help
or
Open poll in your web browser

Bayesian network questions

- **How are most belief nets built?**
  - A) interview people
  - B) machine learning
- **How to build a belief net?**
  - Sequentially add variables, checking for conditional independence
  - Is it sufficient to check P(C|A,B) = P(C|A), P(C|B) or P(C)?
    - A) yes
    - B) no



Your poll will show here

1 Install the app from pollev.com/app

2 Make sure you are in Slide Show mode

Still not working? Get help at pollev.com/app/help
or
Open poll in your web browser

- **What makes a belief net better or worse?**

- **How can you build a better belief net?**

What’s a good network? - MDL

- **Uses few bits to code the model**
  - Proportional to the number of parameters in the model
- **Uses few bits to code the data given the model**
  - $\Pi_i P(\mathbf{x}_i)$
  - $\Sigma_i \log(P(\mathbf{x}_i))$ - Entropy of the data
  - $P(\mathbf{x}_i) = P(A)P(B|A)P(C|A,B)P(D|\ldots)$
- **For example**

```
A       B         model as 6 parameters P(A),P(B)
   \  /        P(C|A,B), P(C|~A,B), P(C|A,~B), P(C|~A,~B)
    C             P(A,B,C) = P(A)P(B)P(C|A,B)
```

Not covered this year!

Learn a belief net - observations

A	B
0	0
0	0
0	0
0	1
1	0
1	1
1	1
1	1

Learn a belief net - models

A	B
0	0
0	0
0	0
0	1
1	0
1	1
1	1
1	1

model	model complexity (# parameters)	-log(P(A,B))
A B		
A → B		
B → A		

Learn a belief net

A	B
0	0
0	0
0	0
0	1
1	0
1	1
1	1
1	1

model   complexity   - log(P(A,B))

A   B   2   - log(P(A)) - log(P(B))

8[-P(A)log(P(A)) - P(~A)log(P(~A))] +   - 8 log(1/2)
8[-P(B)log(P(B)) - P(~B)log(P(~B))]   - 8 log(1/2)
= 16

A ➔ B   3   - log(P(A)) - log(P(B|A))

8[-P(A)log(P(A) - P(~A)log(P(~A))] +   - 8 log(1/2)
4[-P(B|A) log(P(B|A))   - P(~B|A) log(P(~B|A))] 4[¾log¾ + ¼log¼]
4[-P(B|~A)log(P(B|~A)) - P(~B|~A)log(P(~B|~A))]   “   “
= 8[1 +0.81] = 14.5

B ➔A   3   - log(P(B)) - log(P(A|B))

P(A)=P(B)= ½
P(B|A) = ¾
P(B|~A)= ¼