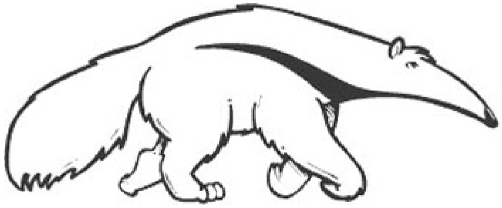
Bayesian Networks: Compact Probabilistic Reasoning

CS171, Summer 1 Quarter, 2019

Introduction to Artificial Intelligence

Prof. Richard Lathrop



Read Beforehand: R&N Ch. 14.1-14.5

You will be expected to know

- Basic concepts and vocabulary of Bayesian networks.
  - Nodes represent random variables.
  - Directed arcs represent (informally) direct influences.
  - Conditional probability tables, P(Xi | Parents(Xi)).

- Given a Bayesian network:
  - Write down the full joint distribution it represents.

- Given a full joint distribution in factored form:
  - Draw the Bayesian network that represents it.

- Given a variable ordering and some background assertions of conditional independence among the variables:
  - Write down the factored form of the full joint distribution, as simplified by the conditional independence assertions.

- Use the network to find answers to probability questions about it.

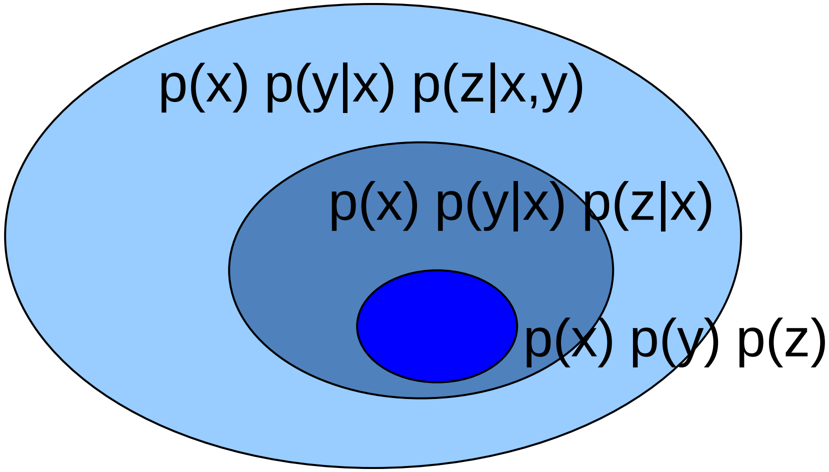
Why Bayesian Networks?

- Probabilistic Reasoning
  - Knowledge Base : Joint distribution over all random variables
  - Reasoning: Compute probability of states of the world
    - Find the most probable assignments
    - Compute marginal / conditional probability

- Why Bayesian Net?
  - Manipulating full joint distribution is very hard!
  - Exploit conditional independence properties
  - Bayesian Network usually more compact & feasible
    - Probabilistic Graphical Models
    - Tool for Reasoning, Computation
    - Probabilistic Reasoning based on the Graph

Conditional independence

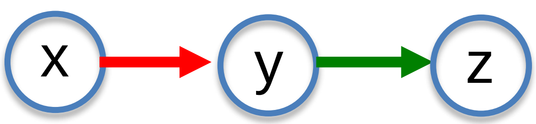
- Recall: chain rule of probability
  - $p(x,y,z) = p(x)\, p(y|x)\, p(z|x,y)$
- *Some* of these models are conditionally independent
  - e.g., $p(x,y,z) = p(x)\, p(y|x)\, p(z|x)$
- *Some* models may have even *more* independence
  - E.g., $p(x,y,z) = p(x)\, p(y)\, p(z)$
- The more independence and conditional independence, the more compactly we can represent and reason over the joint probability distribution.



Bayesian networks

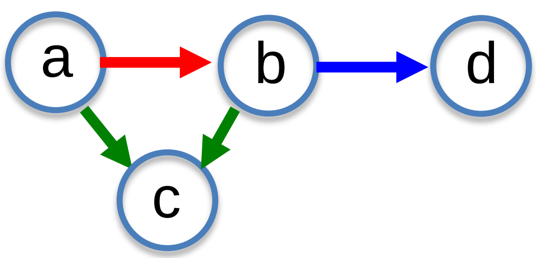
- Directed graphical model
- Nodes associated with variables
- "Draw" independence in conditional probability expansion
  - Parents in graph are the RHS of conditional
- Example:

$$p(x, y, z) = p(x)\, p(y \,|\, x)\, p(z \,|\, y)$$

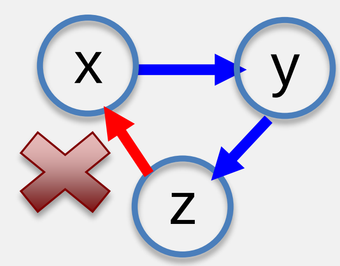


- Example:

$$p(a, b, c, d) = p(a)\, p(b \,|\, a)\, p(c \,|\, a, b)\, p(d \,|\, b)$$



Graph must be ***acyclic***

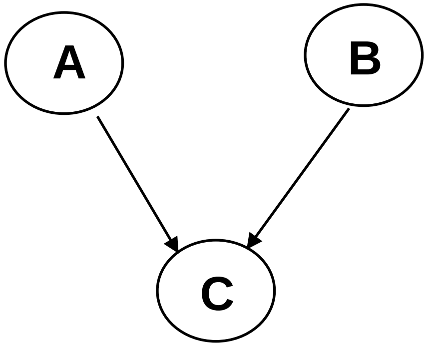


Corresponds to an order over the variables (chain rule)

Bayesian Network

- Specifies a joint distribution in a structured form:

P(A)
0.50



P(B)
0.67

$$p(A,B,C) = p(C|A,B)p(A)p(B) \longleftrightarrow$$

(A and B are independent of all variables;
C depends on A and B)

A	B	P(C)
t	t	0.2
t	f	0.4
f	t	0.3
f	f	0.3

- Dependence/independence shown by a directed graph:
  - Node = random variable
  - Directed Edge = conditional dependence
  - Absence of Edge = conditional independence

- Allows concise view of joint distribution relationships:
  - Graph nodes and edges show conditional relationships between variables.
  - Tables provide probability data.

- **<u>Tables are concise!!</u>**
  - <u>P(¬ A) is not shown since it can be inferred as (1 − P(A)), etc.</u>

Bayesian Networks

- Structure of the graph ⇔ Conditional independence relations

In general,

$$p(X_1, X_2, \ldots X_N) = \prod p(X_i \mid parents(X_i))$$

The full joint distribution — The graph-structured approximation

- <u>Requires that graph is acyclic (no directed cycles)</u>

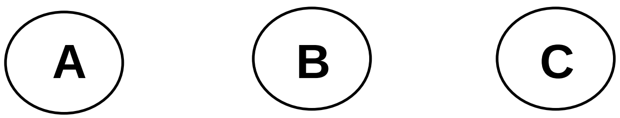
- 2 components to a Bayesian network
  - The graph structure (conditional independence assumptions)
  - The numerical probabilities (for each variable given its parents)

- Also known as belief networks, graphical models, causal networks

- <u>**Parents in the graph ⇔ conditioning variables (RHS) in the formula**</u>

Examples of 3-way Bayesian Networks

A, B, and C are independent.

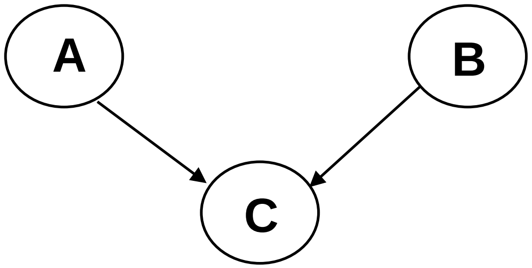


Marginal Independence:
p(A,B,C) = p(A) p(B) p(C)

Parents in the graph ⇔ conditioning variables (RHS)

Examples of 3-way Bayesian Networks

A and B directly influence C.



Independent Causes:
p(A,B,C) = p(C|A,B)p(A)p(B)

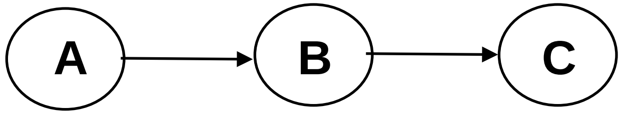
"Explaining away" effect:
Given C, observing A makes B less likely
e.g., earthquake/burglary/alarm example

A and B are (marginally) independent
but become dependent once C is known

<u>**Parents in the graph ⇔ conditioning variables (RHS)**</u>

Examples of 3-way Bayesian Networks

A directly influences B;
B directly influences C;
but A influences C only
indirectly through B.



Markov dependence:
p(A,B,C) = p(C|B) p(B|A)p(A)

Parents in the graph ⇔ conditioning variables (RHS)

Burglar Alarm 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
  - J = John calls to report the alarm
  - M = Mary calls to report the alarm

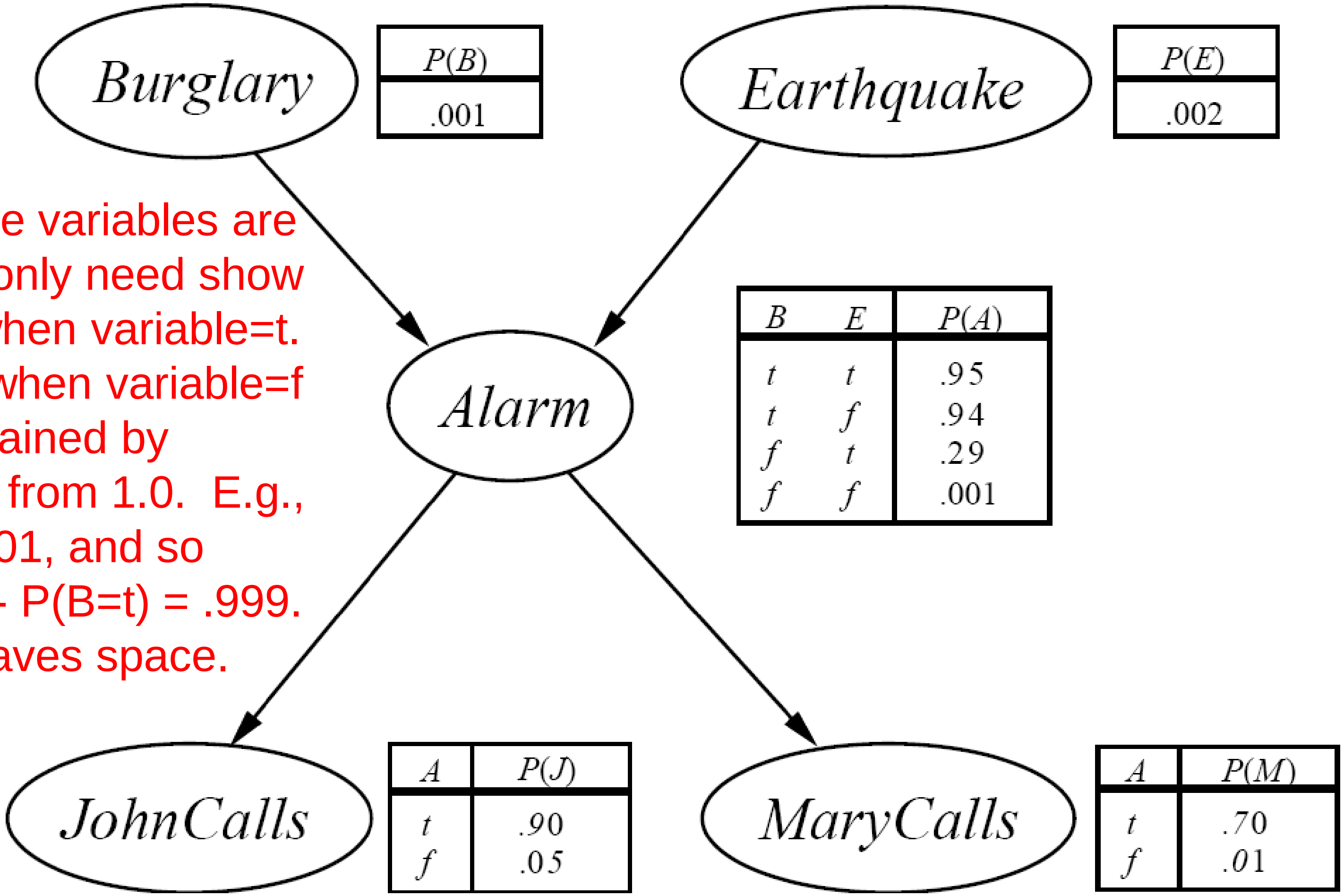
  - What is P(B | M, J) ? (for example)

  - 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?

The Causal Bayesian Network

<u>Generally, order variables so that resulting graph reflects assumed causal relationships.</u>



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

Because the variables are binary, we only need show the value when variable=t. The value when variable=f may be obtained by subtracting from 1.0. E.g., P(B=t) = .001, and so P(B=f) = 1 - P(B=t) = .999. Doing so saves space.

Only requires 10 probabilities!

Constructing a Bayesian Network: Step 1

- Order the variables in terms of influence (may be a partial order)

  e.g., {E, B} -> {A} -> {J, M}

  Generally, order variables to reflect the assumed causal relationships.

- Now, apply the chain rule, and simplify based on assumptions

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

  ≈ P(J, M | A) P(A| E, B) P(E) P(B)

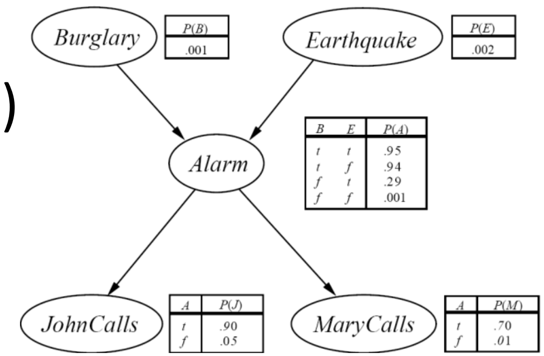
  ≈ P(J | A) P(M | A) P(A| E, B) P(E) P(B)

These conditional independence assumptions are reflected in the graph structure of the Bayesian network

Constructing this Bayesian Network: Step 2

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

  <u>**Parents in the graph ⇔ conditioning variables (RHS)**</u>

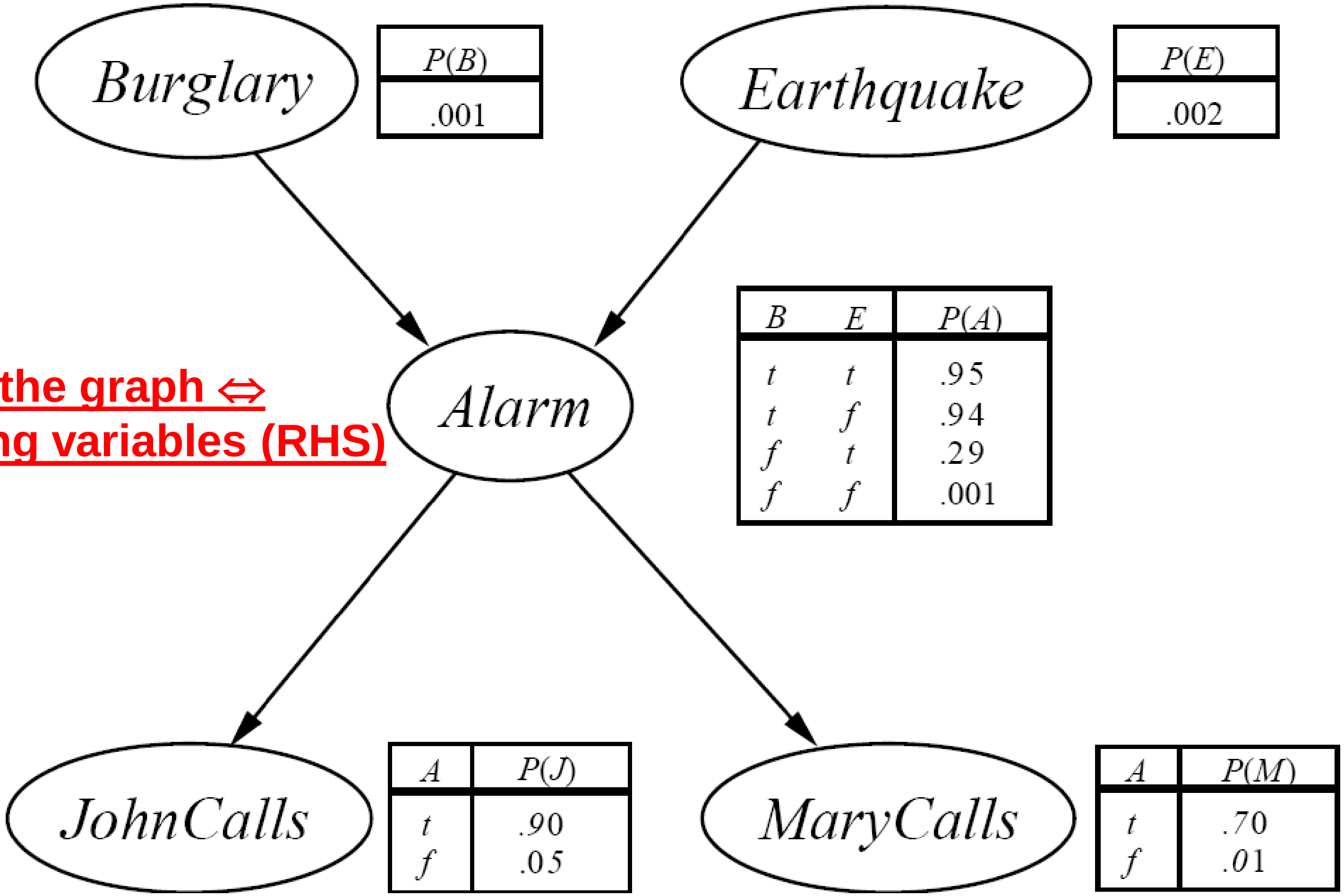


- There are 3 conditional probability tables (CPDs) to be determined: P(J | A), P(M | A), P(A | E, B)
  - Requiring 2 + 2 + 4 = 8 probabilities

- And 2 marginal probabilities P(E), P(B) -> 2 more probabilities

- Where do these probabilities come from?
  - Expert knowledge
  - From data (relative frequency estimates)
  - Or a combination of both - see discussion in Section 20.1 and 20.2 (optional)

The Resulting Bayesian 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

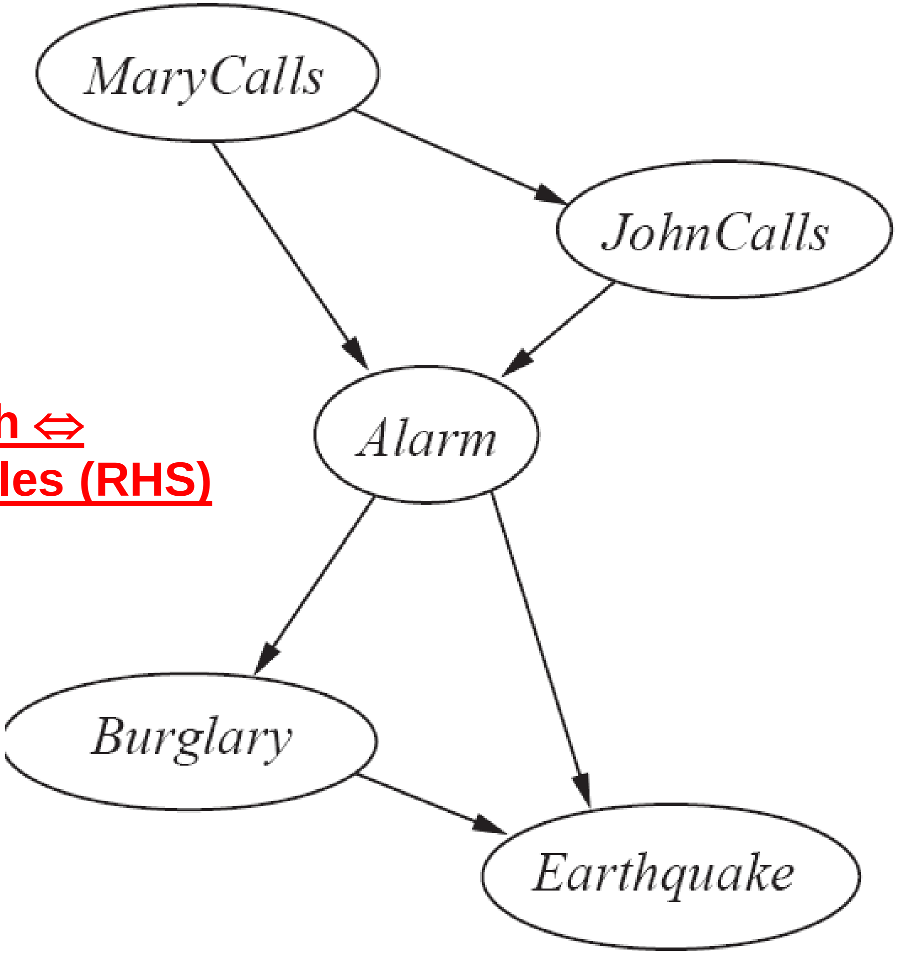
A	P(M)
t	.70
f	.01

Parents in the graph ⇔ conditioning variables (RHS)

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

Generally, order variables so that resulting graph reflects assumed causal relationships.

The Bayesian Network From a Different Variable Ordering

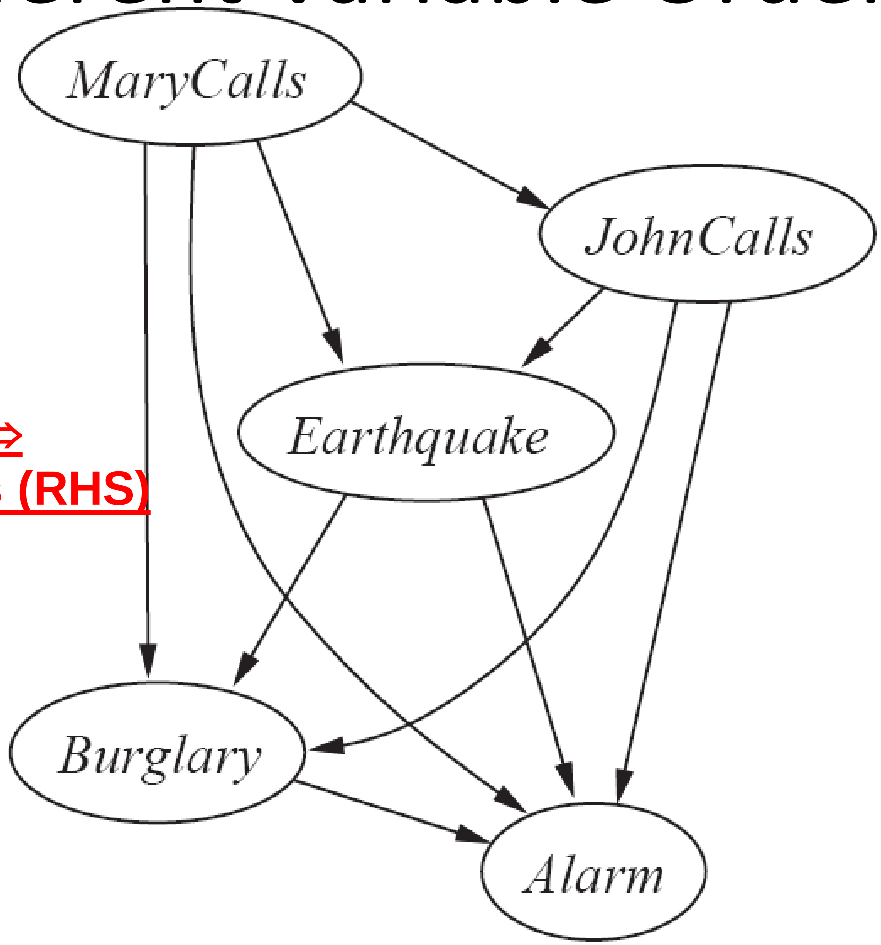


Parents in the graph ⇔ conditioning variables (RHS)

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

Generally, order variables so that resulting graph reflects assumed causal relationships.

The Bayesian Network From a Different Variable Ordering



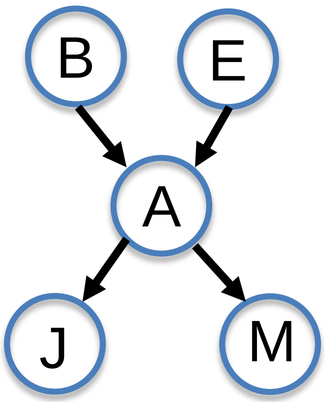
Parents in the graph ⇔ conditioning variables (RHS)

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

Generally, order variables to reflect the assumed causal relationships.

Number of Probabilities Needed (1)

- Joint distribution



Full joint distribution:
2^5 = 32 probabilities

Structured distribution:
specify 10 parameters

E	B	A	J	M	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	J	M	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

Number of Probabilities Needed (2)

- Consider n binary variables
- Unconstrained joint distribution requires $O(2^n)$ probabilities
- If we have a Bayesian network, with a maximum of k parents for any node, then we need $O(n\ 2^k)$ probabilities
- Example
  - Full unconstrained joint distribution
    - $n = 30$, $k = 4$: need 10^9 probabilities for full joint distribution
  - Bayesian network
    - $n = 30$, $k = 4$: need 480 probabilities

Example of Answering a Simple Query

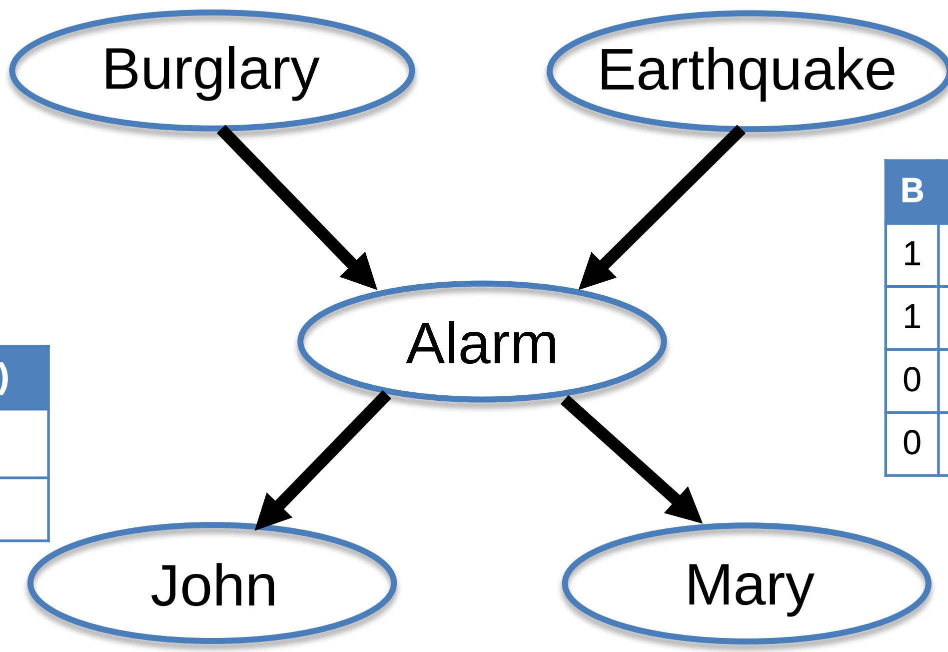
- What is $P(\neg j, m, a, \neg e, b) = P(J = false \wedge M{=}true \wedge A{=}true \wedge E{=}false \wedge B{=}true)$

$P(J, M, A, E, B) \approx P(J \mid A)\, P(M \mid A)\, P(A \mid E, B)\, P(E)\, P(B)$; by conditional independence

$$P(\neg j, m, a, \neg e, b) \approx P(\neg j \mid a)\, P(m \mid a)\, P(a \mid \neg e, b)\, P(\neg e)\, P(b)$$
$$= 0.10 \times 0.70 \times 0.94 \times 0.998 \times 0.001 \approx .0000657$$

P(B)
0.001

P(E)
0.002



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

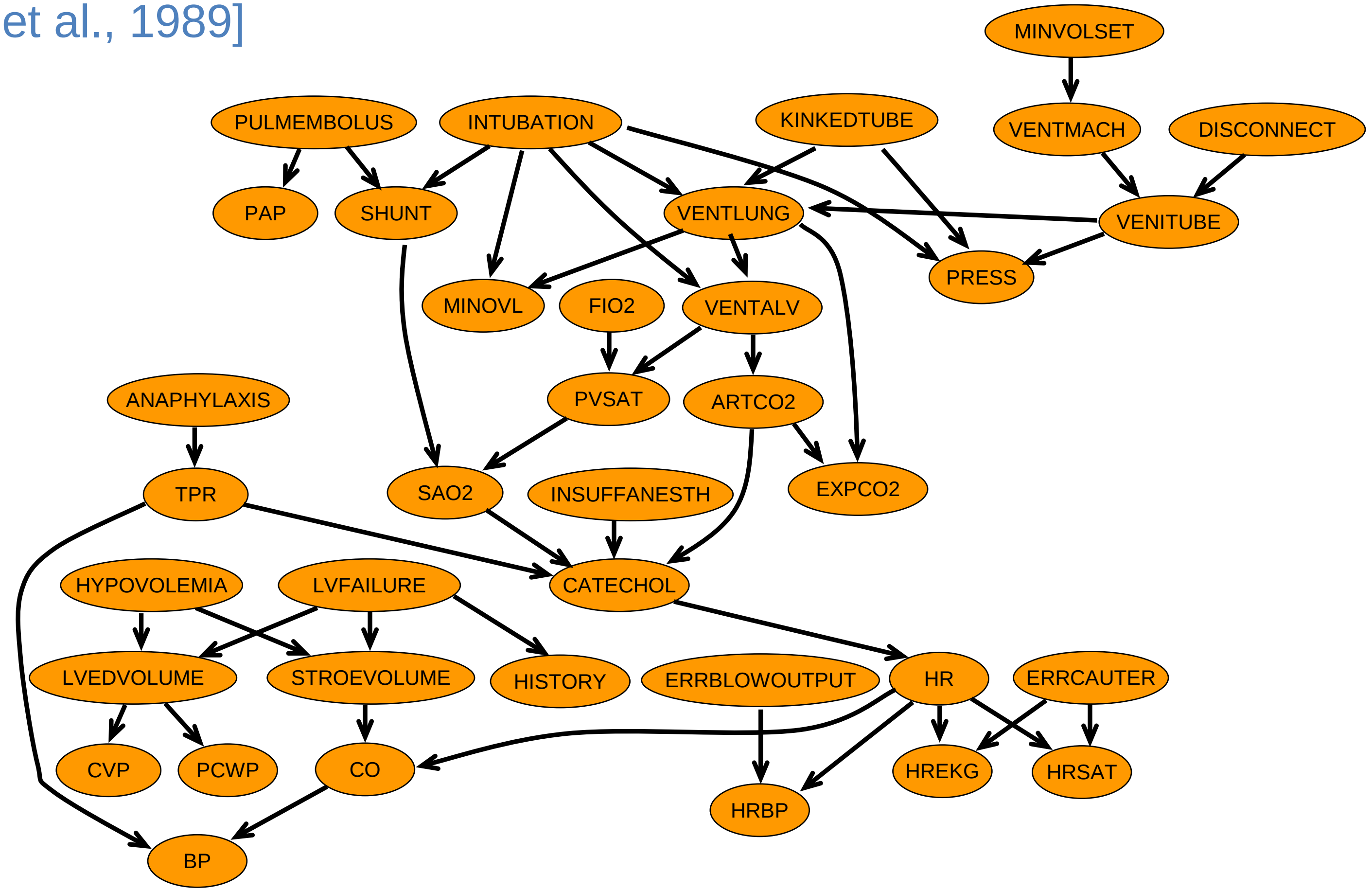
A	P(J|A)
1	0.90
0	0.05

A	P(M|A)
1	0.70
0	0.01

Hospital Alarm network

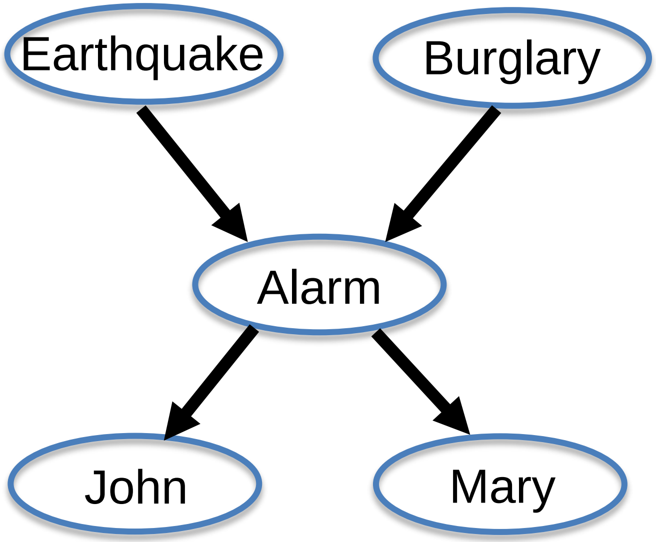
The "alarm" network: 37 variables, 509 parameters (rather than $2^{37} = 10^{11}$!)

[Beinlich et al., 1989]



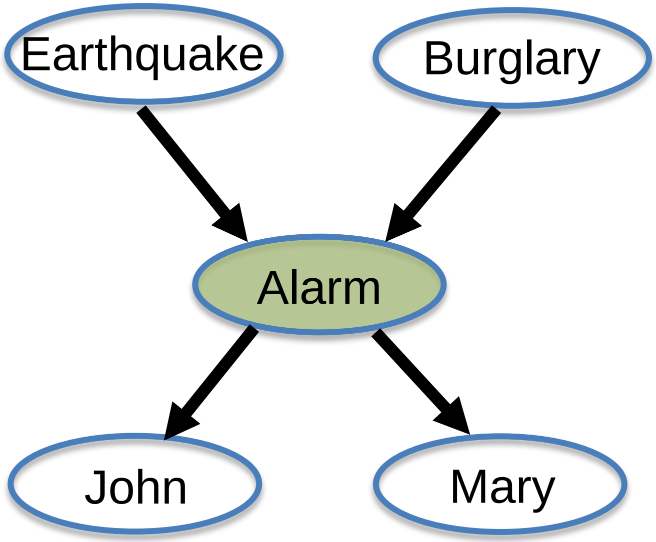
Reasoning in Bayesian networks

- Suppose we observe J
  - Observing J makes A more likely
  - A being more likely makes B more likely



- Suppose we observe A
  - Makes M more likely
- Observe A and J?
  - J doesn't add any more information about M
  - Observing A makes J, M independent
  - P(M | A, J) = P(M | A) ; M is conditionally independent of J given A
- How can we read independence directly from the graph?

Reasoning in Bayesian networks



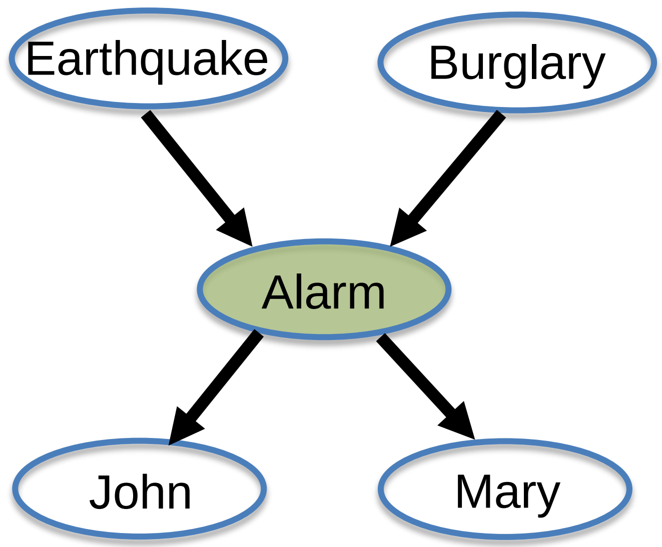
- How are J,M related given A?
  - P(M) = 0.0117
  - P(M|A) = 0.7
  - P(M|A,J) = 0.7
  - Conditionally independent

    (we actually know this by construction!)

- $$p(J, M|a) \propto \sum_{e,b} p(e)\ p(b)\ p(a|e,b)\ p(J|a)\ p(M|a)$$
  $$= \left(\sum_{e,b} p(e,b,a)\right)\ p(J|a)\ p(M|a)$$
  $$= p(a)\ p(J|a)\ p(M|a)$$
  $$= c_a \quad f_a(J)\ g_a(M)$$

Reasoning in Bayesian networks

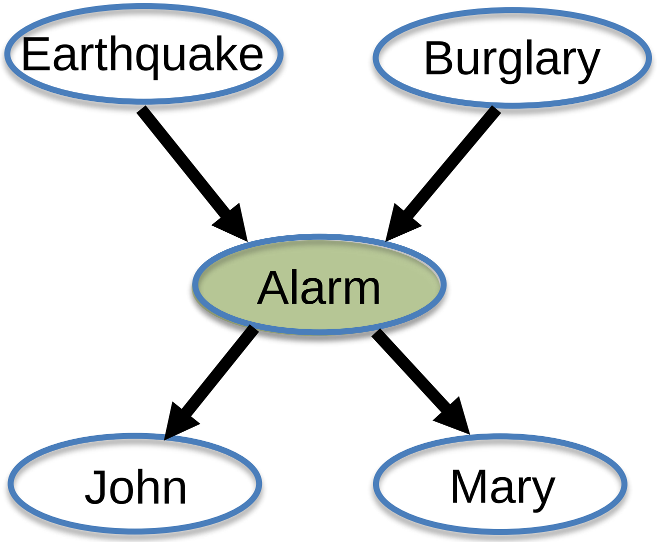
- How are J,B related given A?
  - P(B) = 0.001
  - P(B|A) = 0.3735
  - P(B|A,J) = 0.3735
  - Conditionally independent



- $$\begin{aligned} p(J,B|a) &\propto \sum_{e,m} p(e)\; p(B)\; p(a|e,B)\; p(J|a)\; p(m|a) \\ &= \Big(\sum_e p(e,B,a)\Big)\; p(J|a)\; \Big(\sum_m p(m|a)\Big) \\ &= p(B,a)\; p(J|a) \\ &= \; f_a(B)\;\; g_a(J) \end{aligned}$$

Reasoning in Bayesian networks

- How are E,B related?
  - P(B) = 0.001
  - P(B|E) = 0.001
  - (Marginally) independent



- What about given A?
  - P(B|A) = 0.3735
  - P(B|A,E) = 0.0032
  - Not conditionally independent!
  - The "causes" of A become coupled by observing its value
  - Sometimes called "explaining away"

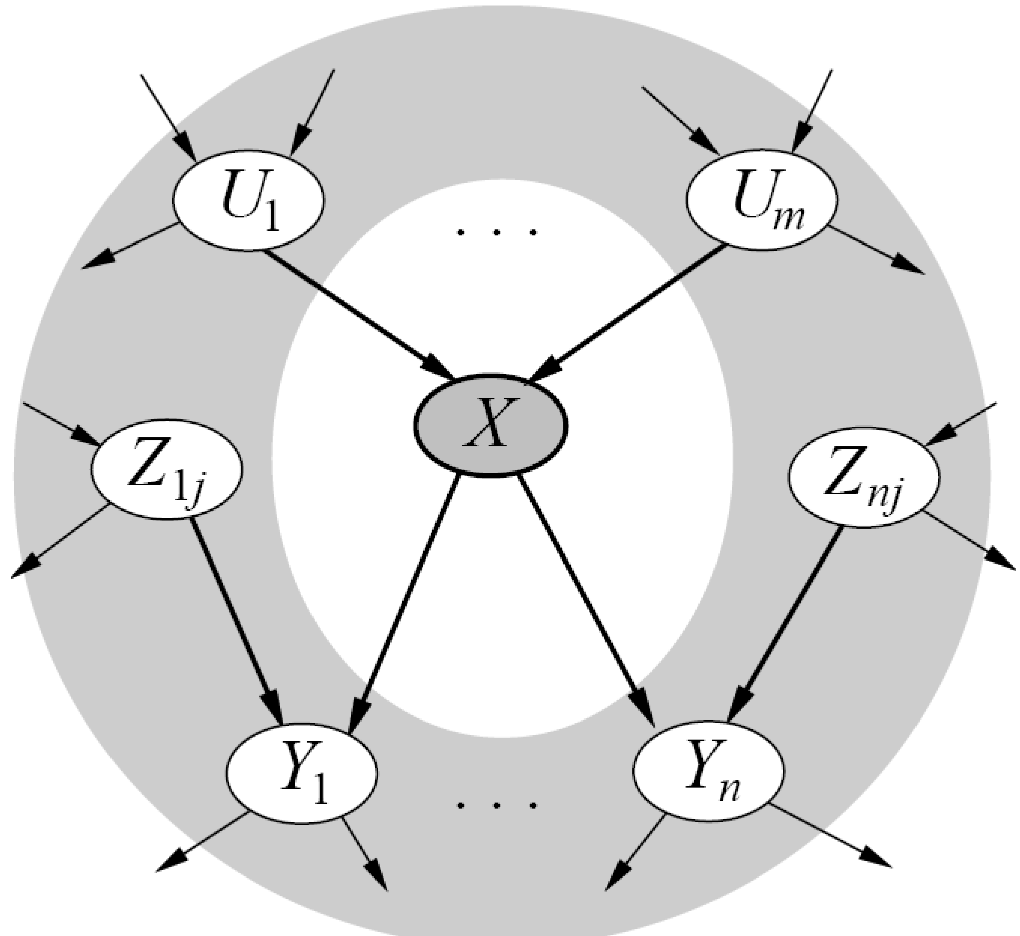
Given a graph, can we "read off" conditional independencies?

The "Markov Blanket" of X (the gray area in the figure)

X is conditionally independent of everything else, GIVEN the values of:

* X's parents
* X's children
* X's children's parents

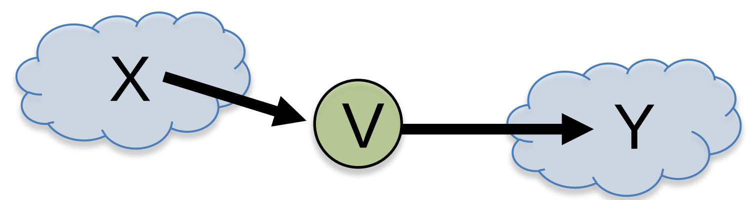
X is conditionally independent of its non-descendants, GIVEN the values of its parents.



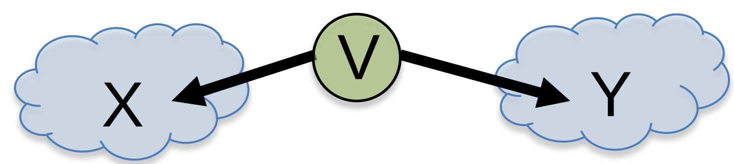
D-Separation

- Prove sets X,Y independent given Z?
- Check all *undirected* paths from X to Y
- A path is "inactive" if it passes through:

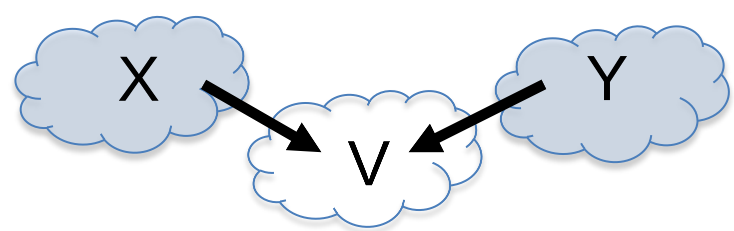
  (1) A "chain" with an observed variable



  (2) A "split" with an observed variable

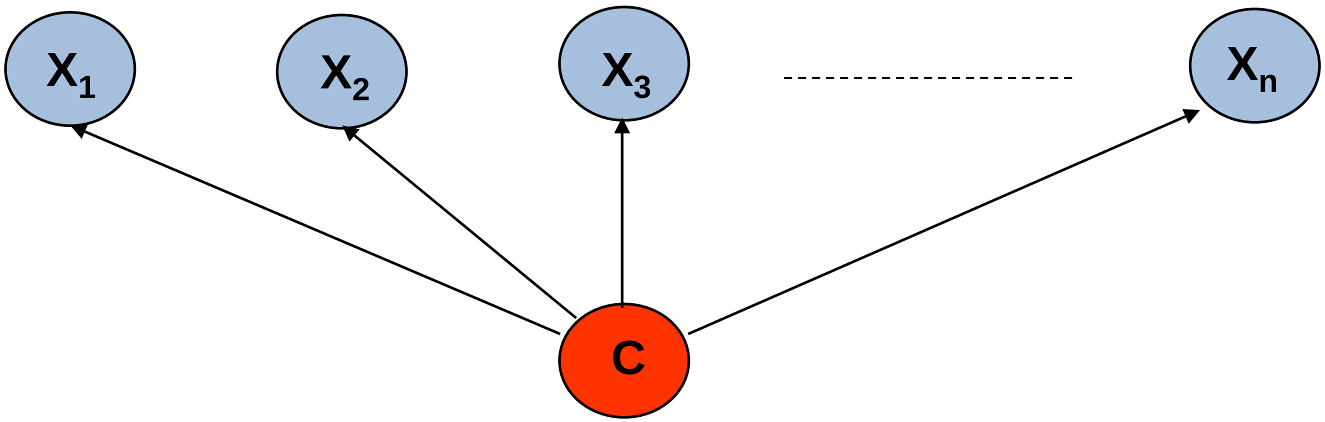


  (3) A "vee" with **only unobserved** variables below it



- If all paths are inactive, conditionally independent!

Naïve Bayes Model



$$P(C \mid X_1,\ldots,X_n) = P(C)\ P(X_1,\ldots,X_n \mid C) / P(X_1,\ldots,X_n)$$
$$= \alpha\ P(C)\ \Pi\ P(X_i \mid C)$$

Normalizing constant α abbreviates normalization

Features X_i are conditionally independent given the class variable C

Widely used in machine learning
e.g., spam email classification: C = spam/not spam, X_i = counts of word_i in emails

Probabilities P(C) and $P(X_i \mid C)$ can be estimated easily from labeled data

Naïve Bayes Model (2)

$$P(C \mid X_1,\ldots X_n) = \alpha\ P(C)\ \Pi\ P(X_i \mid C)$$

Probabilities P(C) and $P(X_i \mid C)$ can be estimated easily from labeled data

$P(C = c_j) \approx$ #(Examples with class label c_j) / #(Examples)

$P(X_i = x_{i,k} \mid C = c_j)$
$\approx$ #(Examples with X_i value $x_{i,k}$ and class label c_j)
/ #(Examples with class label c_j)

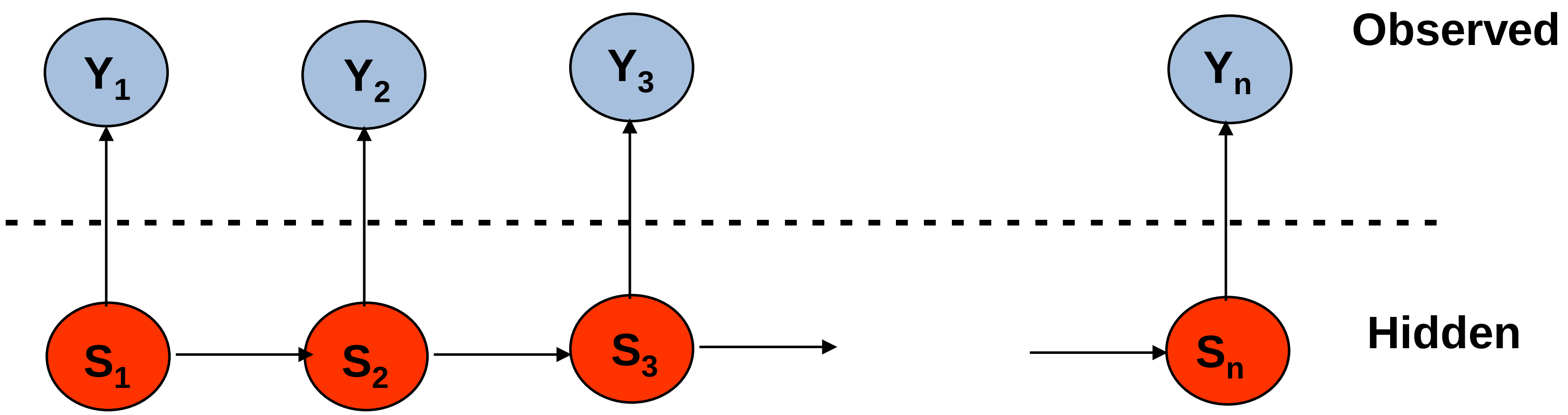
Usually easiest to work with logs

$$\log [\, P(C \mid X_1,\ldots X_n)\,] = \log \alpha + \log P(C) + \Sigma\ \log P(X_i \mid C)$$

DANGER: Suppose ZERO examples with X_i value $x_{i,k}$ and class label c_j ?
An unseen example with X_i value $x_{i,k}$ will NEVER predict class label c_j !

<u>Practical solutions:</u> Pseudocounts, e.g., add 1 to every #() , etc.
<u>Theoretical solutions:</u> Bayesian inference, beta distribution, etc.

Hidden Markov Model (HMM)



Two key assumptions:

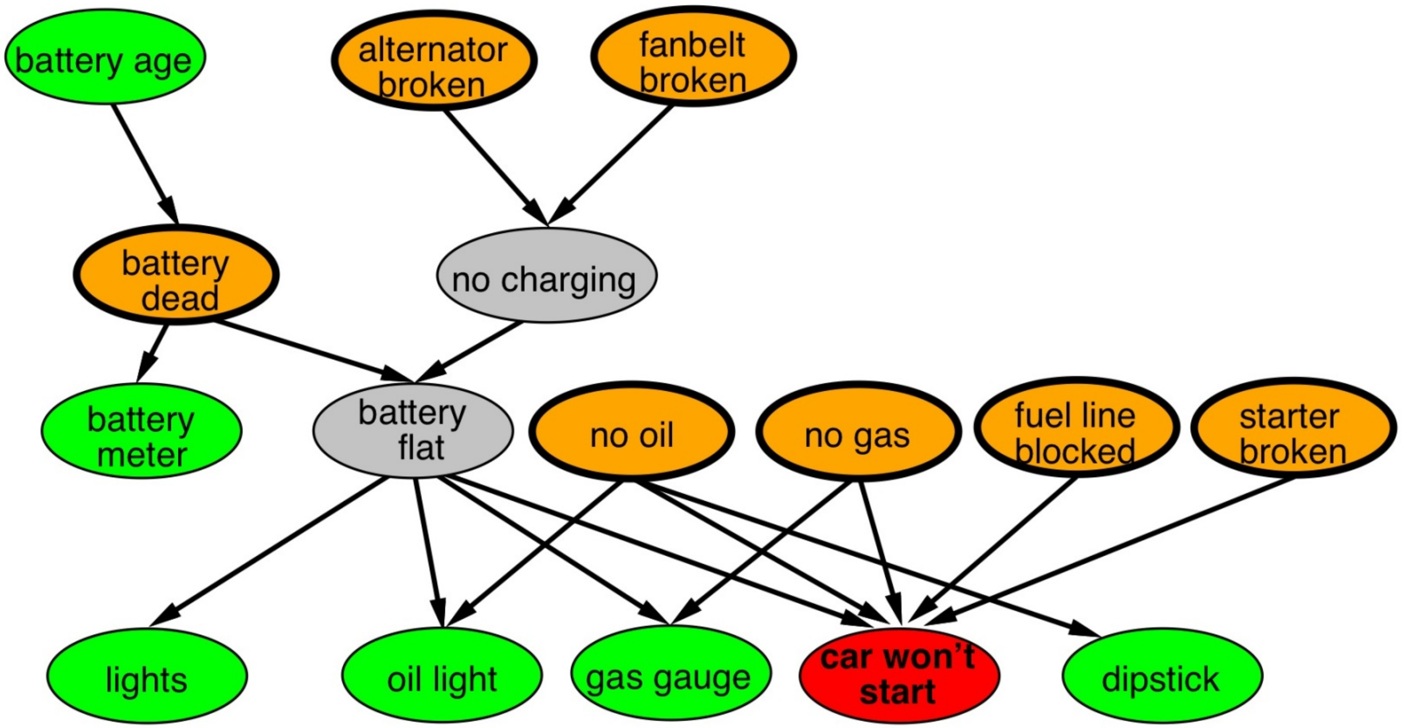
1. hidden state sequence is Markov
2. observation Y_t is conditionally independent of all other variables given S_t

Widely used in speech recognition, protein sequence models

Since this is a Bayesian network polytree, inference is linear in n

Example: Car diagnosis

Initial evidence: car won't start
Testable variables (green), "broken, so fix it" variables (orange)
Hidden variables (gray) ensure sparse structure, reduce parameters



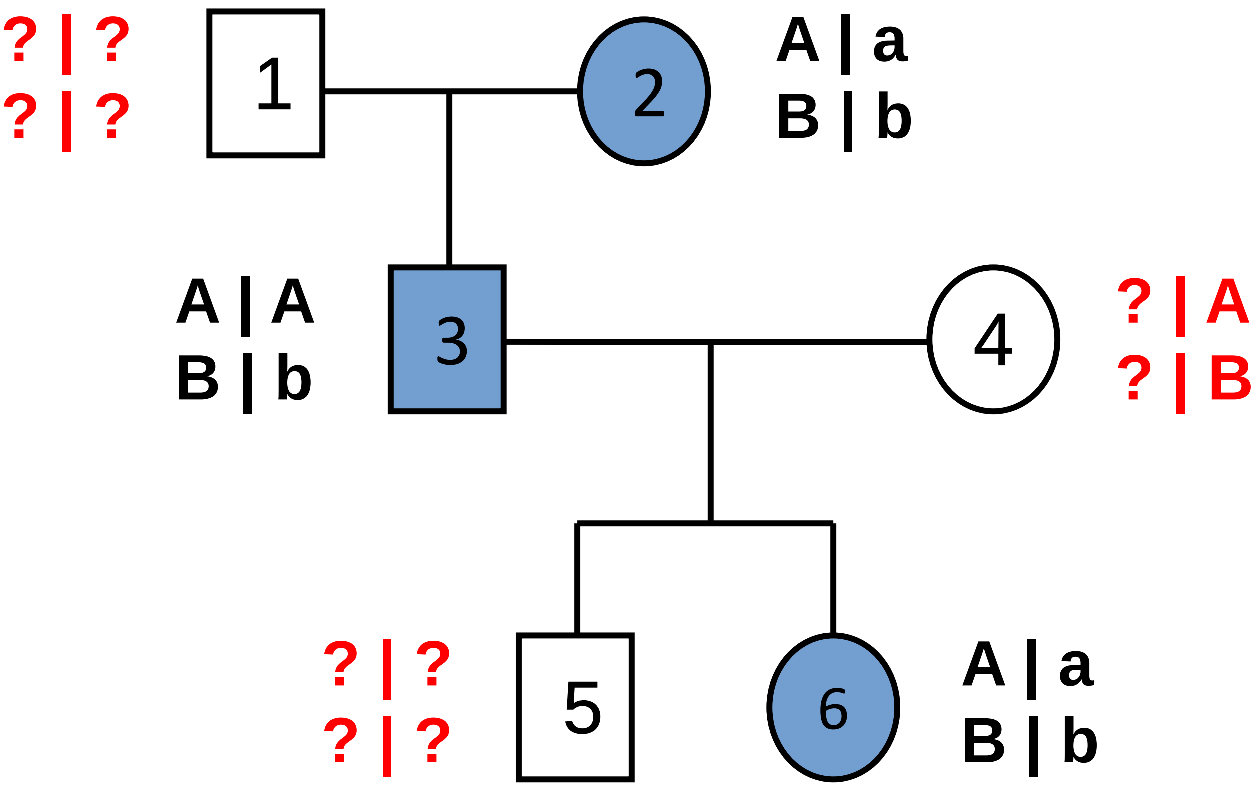
Compact conditional distributions contd.

Noisy-OR distributions model multiple noninteracting causes
1) Parents $U_1 \ldots U_k$ include all causes (can add leak node)
2) Independent failure probability q_i for each cause alone
$\Rightarrow\ P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \prod_{i=1}^{j} q_i$

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

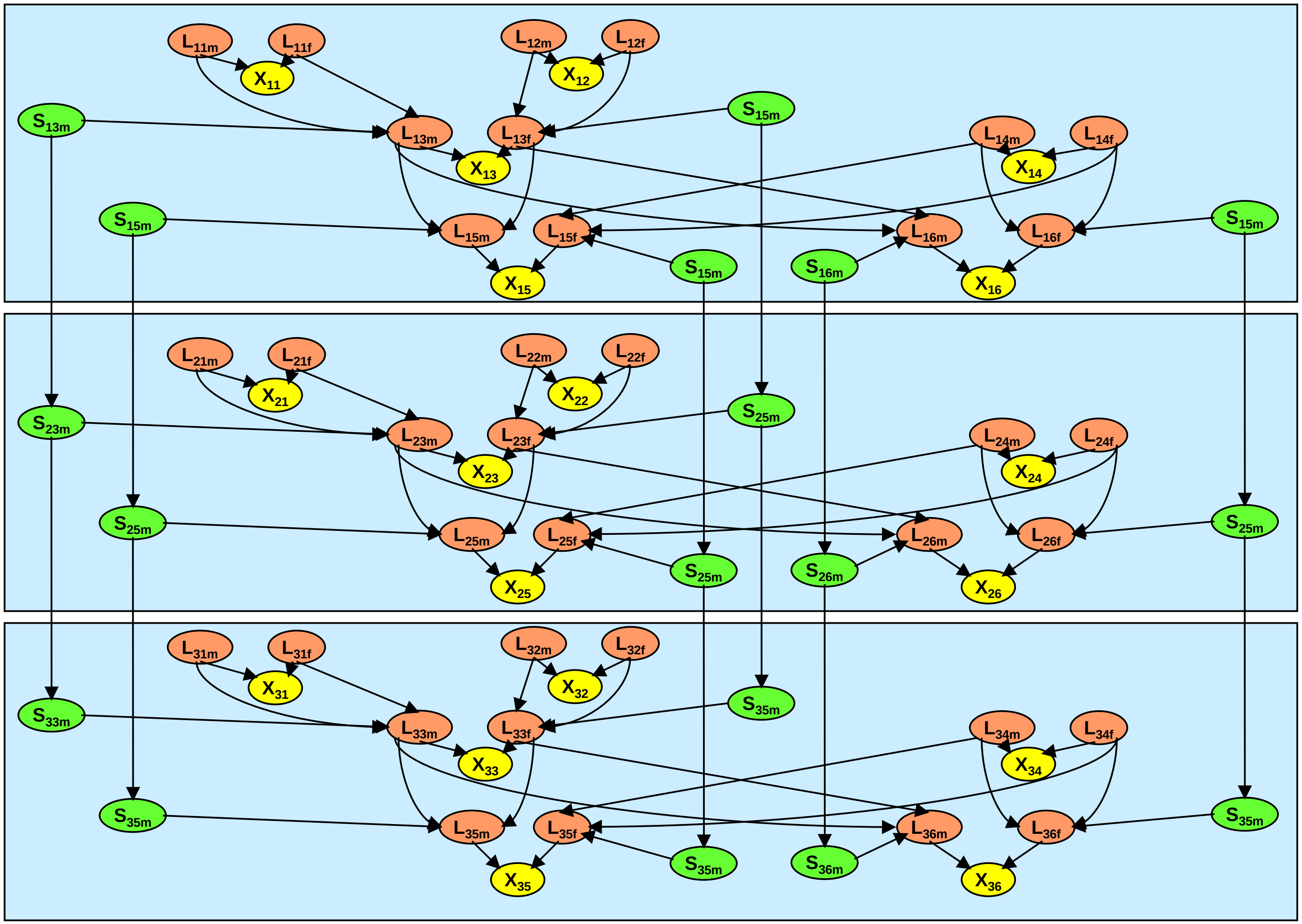
Number of parameters **linear** in number of parents

Examples of “real world” Bayesian Networks: Genetic linkage analysis



- 6 individuals
- Haplotype: {2, 3}
- Genotype: {6}
- Unknown

Examples of “real world” Bayesian Networks: Pedigree model: 6 people, 3 markers



Inference in Bayesian Networks

- **X** *= { X1, X2, …, Xk }* = **query variables** of interest
- **E** *= { E1, …, El }* = **evidence variables** that are observed
  - (**e**, an **event**)
- **Y** = *{ Y1, …, Ym }* = **hidden variables** (nonevidence, nonquery)

- **What is the posterior distribution of X, given E?**
- $\mathbf{P}(\mathbf{X} \mid \mathbf{e}) = \alpha \Sigma_{\mathbf{y}} \mathbf{P}(\mathbf{X}, \mathbf{y}, \mathbf{e})$

- **What is the most likely assignment of values to X, given E?**
- $\mathbf{argmax}_{\mathbf{x}} P(\mathbf{x} \mid \mathbf{e}) = \mathbf{argmax}_{\mathbf{x}} \Sigma_{\mathbf{y}} P(\mathbf{x}, \mathbf{y}, \mathbf{e})$

Inference in Bayesian Networks

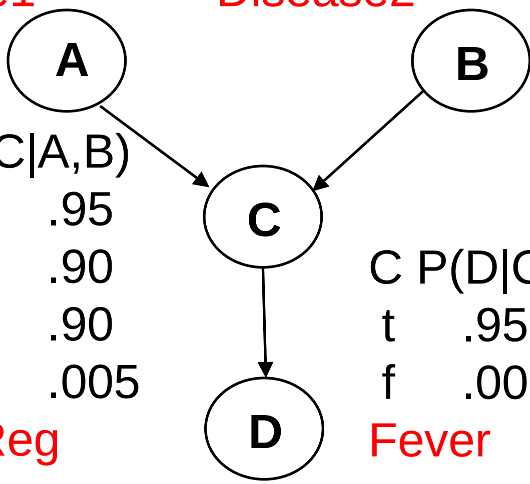
Simple Example

P(A)
.05

Disease1

P(B)
.02

Disease2



A	B	P(C|A,B)
t	t	.95
t	f	.90
f	t	.90
f	f	.005

TempReg

C	P(D|C)
t	.95
f	.002

Fever

} **Query Variables A, B**

} **Hidden Variable C**

} **Evidence Variable D**

Note: Not an anatomically correct model of how diseases cause fever!

Suppose that two different diseases influence some imaginary internal body temperature regulator, which in turn influences whether fever is present.

Inference in Bayesian Networks

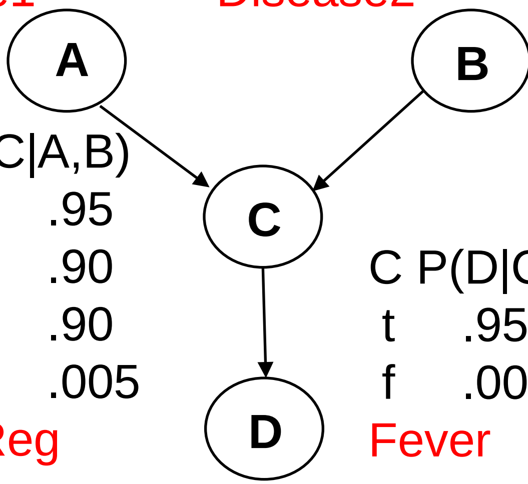
Simple Example

Disease1

P(A)
.05

Disease2

P(B)
.02



TempReg

A	B	P(C|A,B)
t	t	.95
t	f	.90
f	t	.90
f	f	.005

Fever

C	P(D|C)
t	.95
f	.002

What is the posterior conditional distribution of our query variables, given that fever was observed?

$$P(A,B|d) = \alpha \, \Sigma_c \; P(A,B,c,d)$$
$$= \alpha \, \Sigma_c \; P(A)P(B)P(c|A,B)P(d|c)$$
$$= \alpha \, P(A)P(B) \, \Sigma_c \; P(c|A,B)P(d|c)$$

$$P(a,b|d) = \alpha\, P(a)P(b)\, \Sigma_c \; P(c|a,b)P(d|c) = \alpha\, P(a)P(b)\{ P(c|a,b)P(d|c)+P(\neg c|a,b)P(d|\neg c) \}$$
$$= \alpha\, .05 \times .02 \times \{.95 \times .95 + .05 \times .002\} \approx \alpha\, .000903 \approx .014$$

$$P(\neg a,b|d) = \alpha\, P(\neg a)P(b)\, \Sigma_c \; P(c|\neg a,b)P(d|c) = \alpha\, P(\neg a)P(b)\{ P(c|\neg a,b)P(d|c)+P(\neg c|\neg a,b)P(d|\neg c) \}$$
$$= \alpha\, .95 \times .02 \times \{.90 \times .95 + .10 \times .002\} \approx \alpha\, .0162 \approx .248$$

$$P(a,\neg b|d) = \alpha\, P(a)P(\neg b)\, \Sigma_c \; P(c|a,\neg b)P(d|c) = \alpha\, P(a)P(\neg b)\{ P(c|a,\neg b)P(d|c)+P(\neg c|a,\neg b)P(d|\neg c) \}$$
$$= \alpha\, .05 \times .98 \times \{.90 \times .95 + .10 \times .002\} \approx \alpha\, .0419 \approx .642$$

$$P(\neg a,\neg b|d) = \alpha\, P(\neg a)P(\neg b)\, \Sigma_c \; P(c|\neg a,\neg b)P(d|c) = \alpha\, P(\neg a)P(\neg b)\{ P(c|\neg a,\neg b)P(d|c)+P(\neg c|\neg a,\neg b)P(d|\neg c) \}$$
$$= \alpha\, .95 \times .98 \times \{.005 \times .95 + .995 \times .002\} \approx \alpha\, .00627 \approx .096$$

$$\alpha \approx 1 / (.000903+.0162+.0419+.00627) \approx 1 / .06527 \approx 15.32$$

Inference in Bayesian Networks

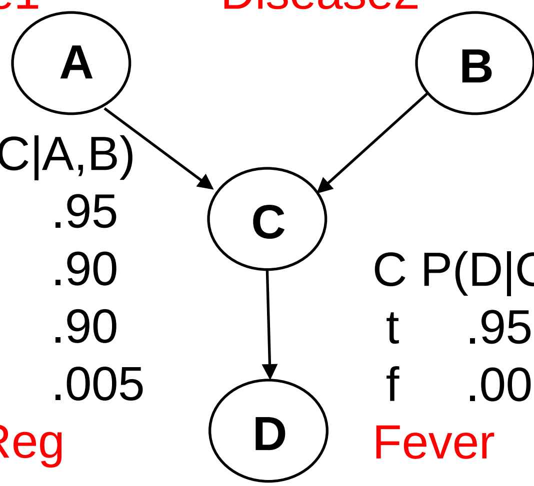
Simple Example

P(A)
.05

Disease1

P(B)
.02

Disease2



A	B	P(C|A,B)
t	t	.95
t	f	.90
f	t	.90
f	f	.005

TempReg

C	P(D|C)
t	.95
f	.002

Fever

What is the most likely posterior conditional assignment of values to our query variables, given that fever was observed?

$$\text{argmax}_{\{a,b\}} P(a, b \mid d)$$
$$= \text{argmax}_{\{a,b\}} \Sigma_c \; P(a,b,c,d)$$
$$= \{ a, \neg b \}$$

$P(a,b|d) = \alpha\, P(a)P(b)\, \Sigma_c \; P(c|a,b)P(d|c) = \alpha\, P(a)P(b)\{ P(c|a,b)P(d|c)+P(\neg c|a,b)P(d|\neg c) \}$
$= \alpha\ .05x.02x\{.95x.95+.05x.002\} \approx \alpha\ .000903 \approx .014$

$P(\neg a,b|d) = \alpha\, P(\neg a)P(b)\, \Sigma_c \; P(c|\neg a,b)P(d|c) = \alpha\, P(\neg a)P(b)\{ P(c|\neg a,b)P(d|c)+P(\neg c|\neg a,b)P(d|\neg c) \}$
$= \alpha\ .95x.02x\{.90x.95+.10x.002\} \approx \alpha\ .0162 \approx .248$

$P(a,\neg b|d) = \alpha\, P(a)P(\neg b)\, \Sigma_c \; P(c|a,\neg b)P(d|c) = \alpha\, P(a)P(\neg b)\{ P(c|a,\neg b)P(d|c)+P(\neg c|a,\neg b)P(d|\neg c) \}$
$= \alpha\ .05x.98x\{.90x.95+.10x.002\} \approx \alpha\ .0419 \approx .642$

$P(\neg a,\neg b|d) = \alpha\, P(\neg a)P(\neg b)\, \Sigma_c \; P(c|\neg a,\neg b)P(d|c) = \alpha\, P(\neg a)P(\neg b)\{ P(c|\neg a,\neg b)P(d|c)+P(\neg c|\neg a,\neg b)P(d|\neg c) \}$
$= \alpha\ .95x.98x\{.005x.95+.995x.002\} \approx \alpha\ .00627 \approx .096$

$\alpha \approx 1 / (.000903+.0162+.0419+.00627) \approx 1 / .06527 \approx 15.32$

Inference in Bayesian Networks

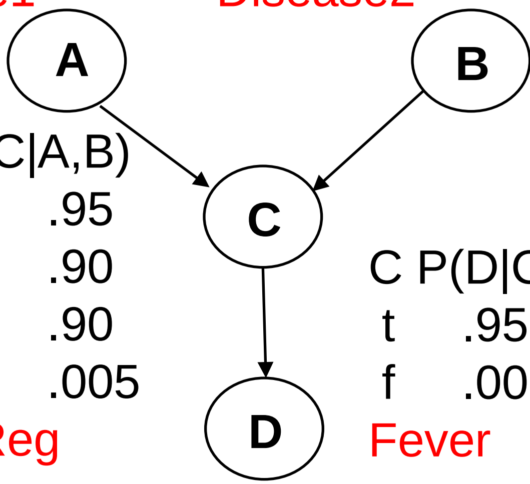
Simple Example

P(A)
.05

Disease1

P(B)
.02

Disease2



A	B	P(C|A,B)
t	t	.95
t	f	.90
f	t	.90
f	f	.005

TempReg

C	P(D|C)
t	.95
f	.002

Fever

What is the posterior conditional distribution of A, given that fever was observed? (I.e., temporarily make B into a hidden variable.)

We can use P(A,B|d) from above.

$P(A|d) = \alpha \, \Sigma_b \; P(A,b|d)$

$P(a|d) = \Sigma_b P(a,b|d) = P(a,b|d)+P(a,\neg b|d)$
$= (.014+.642) \approx .656$

$P(\neg a|d) = \Sigma_b P(\neg a,b|d) = P(\neg a,b|d)+P(\neg a,\neg b|d)$
$= (.248+.096) \approx .344$

This is a marginalization, so we expect from theory that α = 1; but check for round-off error.

A	B	P(A,B|d) from above
t	t	≈ .014
f	t	≈ .248
t	f	≈ .642
f	f	≈ .096

General Strategy for inference

- Want to compute P(q | e)

Step 1:

$P(q \mid e) = P(q,e)/P(e) = \alpha P(q,e)$, since P(e) is constant wrt Q

Step 2:

$P(q,e) = \Sigma_{a..z} P(q, e, a, b, \ldots. z)$, by the law of total probability

Step 3:

$\Sigma_{a..z} P(q, e, a, b, \ldots. z) = \Sigma_{a..z} \Pi_i P(\text{variable } i \mid \text{parents } i)$

(using Bayesian network factoring)

Step 4:

Distribute summations across product terms for efficient computation

Section 14.4 discusses exact inference in Bayesian Networks. The complexity depends strongly on the network structure. The general case is intractable, but there are things you can do. Section 14.5 discusses approximation by sampling.

Summary

- Bayesian networks represent a joint distribution using a graph
- The graph encodes a set of conditional independence assumptions
- Answering queries (or inference or reasoning) in a Bayesian network amounts to computation of appropriate conditional probabilities
- Probabilistic inference is intractable in the general case
  - Can be done in linear time for certain classes of Bayesian networks (polytrees: at most one directed path between any two nodes)
  - Usually faster and easier than manipulating the full joint distribution