Artificial Intelligence

Bayes Nets

Nilsson - Chapter 19

Russell and Norvig - Chapter 14

print troubleshooter
(part of Windows 95)

joint probability distributions

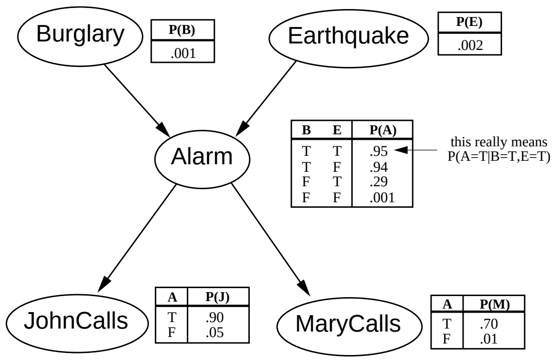
E1	E2	E3	E4	E5	D1	D2	D3	
T	T	T	T	T	T	T	T	0.05
T	T	T	T	T	T	T	F	0.03
...	...	...	...	...	...	...	...	...

P(D1 | E1, NOT E3)

- inefficient for reasoning
- hard to acquire the probabilities

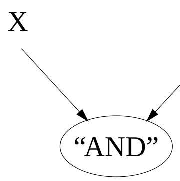
Bayes nets = belief nets
make use of independence inherent in the domain
expert systems: medicine, Microsoft

Bayes net
= directed acyclic graph with conditional probability tables



nodes = random variables, links = direct influences

conditional probability tables

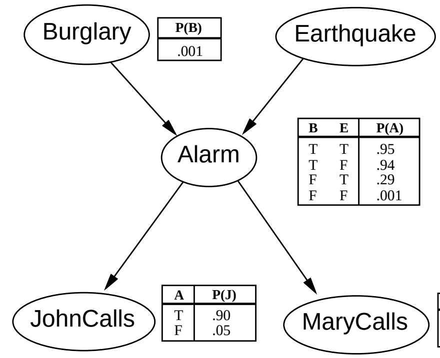


X	Y	P(“AND” | X, Y)
T	T	1.0
T	F	0.0
F	T	0.0
F	F	0.0

P(COIN)

0.5

from Bayes nets to joint probability distributions



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

Burglary	Earthquake	Alarm	JohnCalls	MaryCalls	
T	T	T	T	T	0.000001197

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

from joint probability distributions to Bayes nets (1)

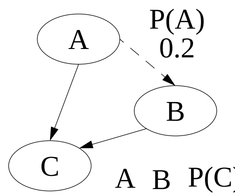
repeatedly:

- pick a variable
- condition it on the smallest possible set of variables picked previously

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

order: A, B, C

A	B	C	
T	T	T	0.054
T	T	F	0.126
T	F	T	0.002
T	F	F	0.018
F	T	T	0.432
F	T	F	0.288
F	F	T	0.032
F	F	F	0.048



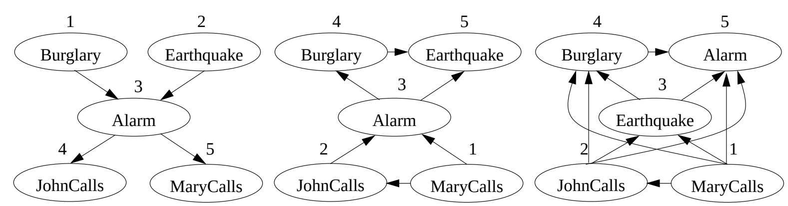
P(A)
0.2

A	P(B)
T	0.9
F	0.9

A	B	P(C)
T	T	0.3
T	F	0.1
F	T	0.6
F	F	0.4

from joint probability distributions to Bayes nets (2)

ordering does matter



sizes of the conditional probability tables

10	13	31

put causes before effects
- smaller network
- easier to make probability judgements

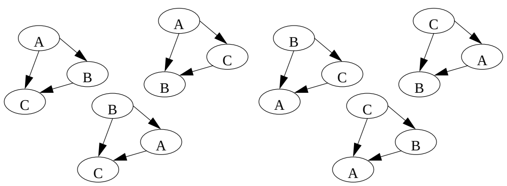
from joint probability distributions to Bayes nets (3)

warning

Bayes nets merely represent joint probability distributions
Bayes nets have nothing to do with causality

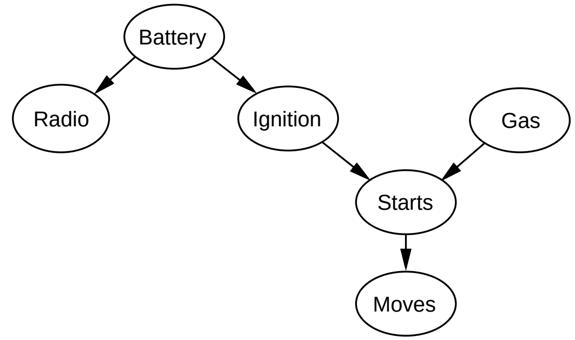
it is smart, but not necessary, to make the edges go from causes to effects

A	B	C	
T	T	T	...
T	T	F	...
T	F	T	...
T	F	F	...
F	T	T	...
F	T	F	...
F	F	T	...
F	F	F	...



all six of these Bayes nets are fine!

independence example

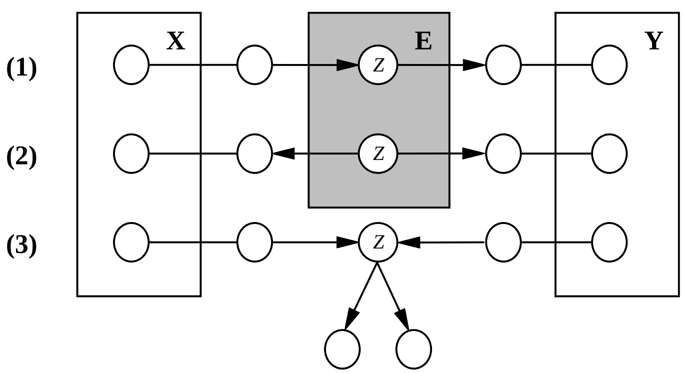


gas and radio
- independent given ignition
- independent given battery
- independent given nothing
- dependent given starts
- dependent given moves

direction-dependent separation

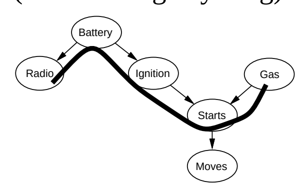
if every undirected path from a node in X to a node in Y is blocked by E,
then X and Y are conditionally independent given E

three ways in which a path from X to Y can be blocked by evidence E



example (1)

Are Radio and Gas are guaranteed to be independent
(not knowing anything)?

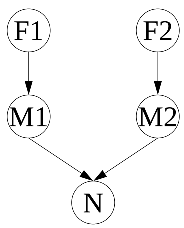


Yes, the structure guarantees it.
There is only one undirected path from Radio to Gas.
This path is blocked because Ignition -> Starts <- Gas is blocked.

example (2)

Two astonomers, in different parts of the world, make measurements M1 and M2 of the number of stars N in some small region of the sky, using their telescopes. Normally, there is a small possibility of error by up to one star. Each telescope can also (with a slightly smaller probability) be badly out of focus (event F1 and F2), in which case the scientist will undercount by three or more stars.

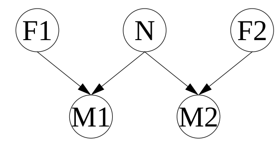
Does the following network correctly reflect these facts?



example (2)

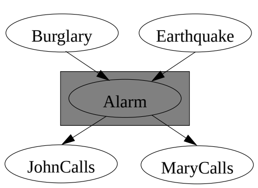
Two astonomers, in different parts of the world, make measurements M1 and M2 of the number of stars N in some small region of the sky, using their telescopes. Normally, there is a small possibility of error by up to one star. Each telescope can also (with a slightly smaller probability) be badly out of focus (event F1 and F2), in which case the scientist will undercount by three or more stars.

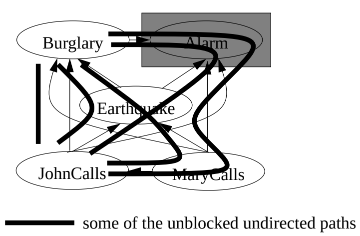
Does the following network correctly reflect these facts?



example (3)

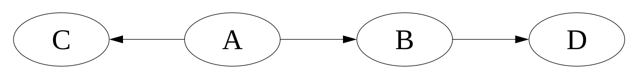
Are Burglary and JohnCalls are guaranteed to be conditionally independent given Alarm?





Yes, the structure guarantees it. No, the structure does not guarantee it.

some simple inferences



A	P(C)
T	0.8
F	0.3

P(A)
0.4

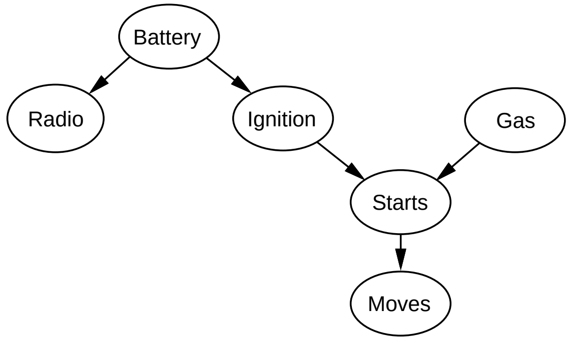
A	P(B)
T	0.8
F	0.3

B	P(D)
T	0.8
F	0.3

these do NOT need to sum to one

P(B | A) = 0.8
P(NOT B | A) = 1 - P(B | A) = 0.2
P(B | NOT A) = 0.3
P(NOT B | NOT A) = 1 - P(B | NOT A) = 0.7
P(C) = P(A) P(C | A) + P(NOT A) P(C | NOT A) = 0.4 0.8 + 0.6 0.3 = 0.5
P(A | C) = P(C | A) P(A) / P(C) = 0.8 0.4 / 0.5 = 0.64
P(B, C) = P(A) P(B | A) P(C | A) + P(NOT A) P(B | NOT A) P(C | NOT A)
= 0.4 0.8 0.8 + 0.6 0.3 0.3 = 0.31
P(D | A) = P(D | B) P(B | A) + P(D | NOT B) P(NOT B | A)
= 0.8 0.8 + 0.3 0.2 = 0.7

more complex inferences



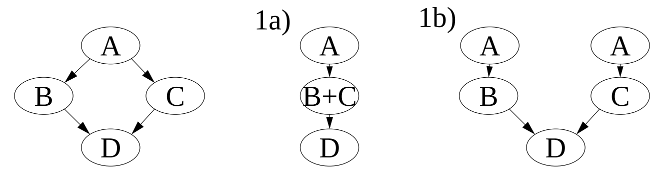
observe evidence; for example, symptoms
calculate the probability of the various diseases given the evidence

complexity

probabilistic inference on polytrees can be done in polynomial time

polytrees are DAGs
where there is at most one path between any two nodes

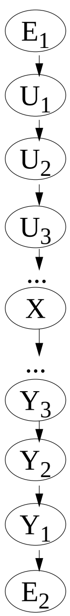
in general, probabilistic inference is NP hard



clustering

conditioning

2) stochastic simulation

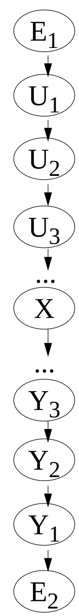


here: algorithms for causal chains

$P(X \mid E_1, E_2)$

$P(X \mid E_2, E_1) = P(E_2 \mid X, E_1)\, P(X \mid E_1) / P(E_2 \mid E_1)$

$P(X \mid E_2, E_1)$ is proportional to $P(E_2 \mid X)\, P(X \mid E_1)$

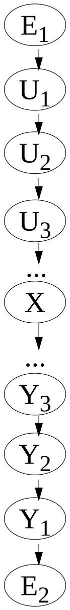


here: algorithms for causal chains

$P(X \mid E_1)$

for each node U_i from U_1 to X:

$$
\begin{aligned}
&P(U_i \mid E_1) \\
&= P(U_{i-1}, U_i \mid E_1) + P(\text{NOT } U_{i-1}, U_i \mid E_1) \\
&= P(U_{i-1} \mid E_1)\, P(U_i \mid U_{i-1}, E_1) \\
&\quad + P(\text{NOT } U_{i-1} \mid E_1)\, P(U_i \mid \text{NOT } U_{i-1}, E_1) \\
&= P(U_{i-1} \mid E_1)\, P(U_i \mid U_{i-1}) \\
&\quad + (1 - P(U_{i-1} \mid E_1))\, P(U_i \mid \text{NOT } U_{i-1})
\end{aligned}
$$



here: algorithms for causal chains

$$P(E_2 \mid X)$$

for each node Y_i from X to Y_1:

$$
\begin{aligned}
& P(E_2 \mid Y_i) \\
& = P(Y_{i-1}, E_2 \mid Y_i) + P(\text{NOT } Y_{i-1}, E_2 \mid Y_i) \\
& = P(Y_{i-1} \mid Y_i)\, P(E_2 \mid Y_i, Y_{i-1}) \\
& \quad + P(\text{NOT } Y_{i-1} \mid Y_i)\, P(E_2 \mid Y_i, \text{NOT } Y_{i-1}) \\
& = P(Y_{i-1} \mid Y_i)\, P(E_2 \mid Y_{i-1}) \\
& \quad + (1 - P(Y_{i-1} \mid Y_i))\, P(E_2 \mid \text{NOT } Y_{i-1})
\end{aligned}
$$